线性代数讲义

Linear Algebra,LN

學院 数据科学学院（SDS）

2026 年 1 月 18 日

Abstract

这是关于 MAT2040/MAT2041/MAT2043/MAT3040 的课程笔记

《声明：此版本并非最新版本》

最新版本请见：https://williamsheltongoh.github.io/lecturenotes.html

线性代数及其应用

0.1 前言

《声明：此版本并非最新版本》

最新版本请见：https://williamsheltongoh.github.io/lecturenotes.html

线性代数作为数学领域的重要分支，核心聚焦于**线性关系**与**线性结构**的研究。在现代科学技术的诸多领域，从基础的物理化学，到计算机与电气自动化，再到站在潮头之上的人工智能，都离不开线性代数的理论支撑。线性代数不仅是需要掌握的知识与工具，更是必须熟练掌握、深刻理解的**方法与思想**。

一、线性代数的研究对象

线性代数的研究对象可概括为：

- 基于域的向量空间的结构（基、维数、子空间）；
- 线性变换的表示（矩阵）与运算（加法、乘法、逆变换）；
- 线性方程组的解的存在性与构造；
- 特征值、特征向量与对角化等核心问题。

通过**线性关系**揭示空间中的代数解构与几何直观。

二、线性代数的重要研究方法

数学归纳法是一种重要的研究方法。数学归纳法或许无法完整地展示底层逻辑，但其通过“多米诺骨牌”般的递推过程，使它在分析代数问题中具有重要的现实意义！

三、线性代数的重要研究思想

- **线性组合（Linear Combination）**：线性组合思想贯穿线性代数始终。
- **变换映射（Transformation）**：线性变换是线性代数的核心概念之一。它将一个（或 2 个）向量空间中的向量映射到另一个向量空间，保持向量的加法和数乘运算。通过研究线性变换的矩阵表示，可以深入理解变换的性质。
- **等价关系与代表元的思想（The Idea of Equivalence Relation and Representative Element）**：通过定义等价关系，可以将具有相同性质的对象归为一类，然后选取代表元来研究这类对象的共性。这种思想有助于简化问题，突出本质特征。

代数中“**分类**”的核心目标是判断“**两个对象是否本质相同**”，我们采用的逻辑链为：

- 明确研究对象（Object），刻画对象的本质特征与其核心代数结构（Algebra Construction）。
- 通过“保持结构的同构”定义等价关系（Equiv. Relation），确立某个分类的不变量（Invariant）。
- 构建数学工具（Tool）——用于刻画代数结构、保持等价关系、计算不变量且具象化对象的方法。

在正式介绍线性代数的具体知识前，我们将先介绍以下前备知识和前垫概念：

0.2 线性代数的前备知识

在介绍线性代数之前，让我们先从“封闭”这个概念讲起：

【定义】：对于数集 S，若对任意 $a, b \in S$，都有 $a \pm b, ab, \frac{a}{b} \in S$，则称 S 是一个“**封闭**”的数集。

Example

实数集（域）$\mathbb{R}$、有理数集（域）$\mathbb{Q}$ 都是封闭的。

Exercise

无理数集封闭吗？
整数集封闭吗？

【定义】：**未知数/变量（variable）**：表达式中待确定值的符号或可以改变的量，通常用 x, y, z 字母等表示。

Example

表达式 $3x + 5$ 中，x 是未知数，也是变量。选择不同的 x 会给出不同的值（result value）。

【定义】：**多项式（Polynomial）**：由变量（如 x）和系数通过有限次加法、减法、乘法和正整数次幂运算得到的表达式。

Example

$3x^2 + 2x + 1$ 就是关于变量 x 的多项式。

其中变量的最高次数，我们称之为多项式的“次数”。
在刚刚的例子中，次数为 2。

【定义】: **n 元一次方程（n-ary Linear Equation）**：含有 n 个未知数，且每个未知数次数均为 1 的方程。

其中只含一个未知数且未知数（变量）的最高次数为 1 的方程，称为一次方程，也称线性方程（Linear Equation）。

Example

例如：$2x+3=7$ 是一元一次方程。
例如：$2x+3y=5$ 是二元一次方程。

在实践中，我们常常将多个 n 元一次线性方程组合起来就可以得到线性方程组：

【定义】: **n 元一次线性方程组（n-ary Linear Equations System with m equations）**：由 m 个 n 元一次方程组成的方程组。

Example

例如：$\begin{cases} x+y=1 \\ 2x-y=3 \end{cases}$
就是二元一次线性方程组（$n=2$，$m=2$）。

更一般的，我们来考察 $m \times n$ 的情节：

即由 m 行含有 n 个未知数的一次方程联立而成的方程组：

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

其中：a_{ij} 是第 i 个方程中第 j 个未知数的系数；$b_1, b_2, \ldots, b_m$ 是常数项。

由此，我们将这些未知数前面的系数提取并且排列出来，就可以得到“系数矩阵”和“增广矩阵”的概念。

【定义】: **系数矩阵（Coefficient Matrix）**：由方程组的系数 a_{ij} 构成的 $m\times n$ 矩阵：

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

简记为 $A=(a_{ij})_{m\times n}$。

【定义】: **增广矩阵（Augmented Matrix）**：在系数矩阵 A 的右侧增加一列常数项（即原方程等式右边的所有元素）$b_1, b_2, \ldots, b_m$ 构成的 $m\times(n+1)$ 矩阵：

$$\tilde{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}$$

简记为 $\tilde{A} = [A \mid b]$，其中 $b = (b_1, b_2, \ldots, b_m)^T$ 是常数项列向量。

矩阵形式表示上述方程组可简洁地表示为矩阵方程：$Ax = b$

其中：A 是系数矩阵（$m \times n$）；$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ 是未知数向量（$n \times 1$）；$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$ 是常数项向量（$m \times 1$）。

在刚刚的例子：$\begin{cases} x + y = 1 \\ 2x - y = 3 \end{cases}$ 中

上述方程组的系数矩阵为 $\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}$。

上述方程组的增广矩阵为 $\begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 3 \end{pmatrix}$。

更一般的，我们通过这种 m×n 的形式，进一步地抽象出一般矩阵的概念：

【定义】：矩阵（Matrix）： 由 $m \times n$ 个数 a_{ij} 排成的 m 行 n 列的数表，记为 $A = (a_{ij})_{m \times n}$。

接着我们将讨论一种作用在矩阵上的常规操作：**转置**（Transpose）

【定义】矩阵的转置： 将矩阵的行与列按顺序互换，即把原矩阵的第 i 行变为转置后矩阵的第 i 列，第 j 列变为转置后的第 j 行。

用数学符号表示：若原矩阵为 A，其转置矩阵记为 A^T 或者 ${}^tA$

设 A 是一个 $m \times n$ 的矩阵（m 行 n 列），则 A^T 是一个 $n \times m$ 的矩阵，满足：$(A^T)_{i,j} = A_{j,i}$ 其中 $(A^T)_{i,j}$ 表示转置矩阵中第 i 行第 j 列的元素，$A_{j,i}$ 表示原矩阵中第 j 行第 i 列的元素。

Example

例如：设矩阵 A 为：$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$

这是一个 2×3 的矩阵（2 行 3 列）。

其转置矩阵 A^T（或 ${}^tA$ ）为：$A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$

接下来，我们将考察一些特殊的矩阵：

【定义】：方阵（Square Matrix）：行数 m 等于列数 n 的矩阵，即 $m \times m$ 矩阵。

例子：$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$。就是一个三阶方阵。

【定义】：对角阵（Diagonal Matrix）：主对角线以外元素全为 0 的方阵，记为 $\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$。

例子：$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$。

【定义】：单位矩阵（Identity Matrix）：主对角线元素全为 1，其余元素全为 0 的 n 阶方阵，记为 I_n 或 E_n。

例如二阶单位矩阵：主对角线元素全为 1，其余元素全为 0 的 2×2 方阵 $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$。

又如三阶单位矩阵：主对角线元素全为 1，其余元素全为 0 的 3×3 方阵，记为 I_3 或 E_3。$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

我们现在不加证明的给出：

Fact or Background

单位阵类似于“1”，对于任意的 n 阶方阵 A，我们有：$IA = AI = A$

我们后续将在矩阵乘法中详细解释。

【定义】：数量矩阵（Scalar Matrix）：主对角线元素相同，其余元素全为 0 的方阵。(即 kI，k 属于 F)

例子：$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$。

除了对矩阵元素的研究，我们也可以对矩阵元素的排布与结构进行进一步研究：

- 定义：$W_+ := \{A \in M_n(\mathbb{F}) : {}^tA = A\} \subset M_n(\mathbb{F})$（对称矩阵集合）：
  - 解释：W_+ 是域 $\mathbb{F}$ 上所有 n 阶对称矩阵（symmetric matrix）构成的集合，其中 ${}^tA$ 表示矩阵 A 的转置（这种方式也可以表达转置），对称矩阵满足转置后与自身相等。
  - 例如（取 $\mathbb{F} = \mathbb{R}$，$n = 2$）：矩阵 $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$，其转置 ${}^tA = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = A$，故

$A \in W_+$。

- 定义：$W_- := \{A \in M_n(\mathbb{F}) : {}^tA = -A\} \subset M_n(\mathbb{F})$（反对称矩阵集合）：
  - 解释：W_- 是域 $\mathbb{F}$ 上所有 n 阶反对称矩阵构成的集合，反对称矩阵满足转置后与自身的相反数相等，且**主对角线元素全为** 0（因 $(A)_{ii} = -(A)_{ii}$，故 $(A)_{ii} = 0$）。
  - 例如 (取 $\mathbb{F} = \mathbb{R}$, $n = 2$): 矩阵 $B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, 其转置 ${}^tB = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = -B$, 故 $B \in W_-$。

在介绍完矩阵后，我们可以进入对行列式的讨论：

行列式的概念最早源于对线性方程组的研究，但是在线性代数的发展历程中又起到了重要的作用。接下来，我们将研究一类特殊的线性方程组 (n 个未知数，n 个方程）与其构成的方阵，从而引入行列式的概念。

我们将在后续的讲义中解释行列式的性质及由来（其内禀定义)。

Chapter 1

向量空间与线性变换

1.1　向量与向量空间

这个世界上有许多不同的实体（如向量、函数等），而他们往往具有**两种运算（"加法"和"标量乘法"）的相同规则**，由此通过将它们抽象为向量空间，我们只需要研究这个通用框架即可了解他们的**共性**。

1.1.1　向量空间与向量的定义

向量空间（vector space）是线性代数中的核心概念。设 V 是一个非空集合，F 是一个数域。在 V 上定义以下两种运算：

- "加法 addition"：对于任意 $\alpha, \beta \in V$，存在唯一的 $\alpha + \beta \in V$。
- "数乘 scalar multiplication"：对于任意 $k \in F$ 和 $\alpha \in V$，存在唯一的 $k\alpha \in V$。

则称定义在数域 F 上的空间 V 封闭。

【Remark】:

该加法与数乘不一定是传统意义上的加法与数乘。

Definition

如果 V 在满足封闭的前提下，同时满足以下八条公理：

1. 加法交换律：$\alpha+\beta=\beta+\alpha$。
2. 加法结合律：$(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$。
3. 零元存在：存在 $0\in V$，使得对于任意 $\alpha\in V$，$\alpha+0=\alpha$。
4. 负元存在：对于任意 $\alpha\in V$，存在 $\beta\in V$，使得 $\alpha+\beta=0$。
5. 数乘单位元：$1\cdot\alpha=\alpha$。
6. 数乘结合律：$k(l\alpha)=(kl)\alpha$。
7. 数乘分配律 1：$k(\alpha+\beta)=k\alpha+k\beta$。
8. 数乘分配律 2：$(k+l)\alpha=k\alpha+l\alpha$。

满足上述条件时，称 V 是数域 F 上的向量空间，记作 $(V,F,+,\cdot)$。

向量空间的定义可简化记忆为“**存在零元，两种运算，八条公理**”，其中**加法和数乘需满足封闭性**。

接下来我们会举几个典型的向量空间的例子来促进理解：

- 实数域上的 n 维向量空间 $\mathbb{R}^n$ 定义：由 n 个实数组成的有序数组 $(x_1,x_2,\ldots,x_n)$ 的集合，其中加法和数乘按分量进行。

  Example

  - $\mathbb{R}^2$：二维平面上的所有向量（如坐标点 (x,y)）；
  - $\mathbb{R}^3$：三维空间中的所有向量（如坐标点 (x,y,z)）。

- 同样的，可以得到复数域上的 n 维向量空间 $\mathbb{C}^n$
- 零空间 $\{0\}$, 定义：仅包含零向量的空间，维数为 0。
- $M_{m\times n}(\mathbb{F})$：
  - 解释：域 $\mathbb{F}$ 上所有 m 行 n 列矩阵的集合。
  - 验证：
    * 加法封闭：两 $m\times n$ 矩阵相加，结果仍为 $m\times n$ 矩阵。
    * 数乘封闭：域中数乘 $m\times n$ 矩阵，结果仍为 $m\times n$ 矩阵。

    - 其他公理（交换律、结合律等）由矩阵运算规则保证。

- $\mathcal{P}_n(\mathbb{F})$：
  - 解释：域 $\mathbb{F}$ 上次数不超过 n 的一元多项式集合，形如 $\sum_{i=0}^{n} a_i x^i$（$a_i \in \mathbb{F}$）。
  - 验证：
    - 加法封闭：两次数不超过 n 的多项式相加，次数仍不超过 n。
    - 数乘封闭：域中数乘次数不超过 n 的多项式，次数仍不超过 n。
    - 其他公理（交换律、结合律等）由多项式运算规则保证。

Definition

向量（vector）： 给定一个向量空间，其空间中的元素，我们称之为向量。

向量空间的元素（即向量）的具体形式（如数组、函数、矩阵等）由具体的研究场景决定。

1.1.2 向量空间中的线性运算

什么是线性运算?

满足上述向量空间八条公理的加法和数乘称为线性运算。

线性运算包含加法和数乘运算:

- **加法：** 对于向量空间 V 中的向量 α, β，它们的和 $\alpha + \beta$ 是按照向量空间所定义的加法规则得到的 V 中的唯一向量。
- **数乘：** 对于数域 F 中的数 k 和向量空间 V 中的向量 α，数乘 $k\alpha$ 是按照向量空间所定义的数乘规则得到的 V 中的唯一向量。

凡是符合上述向量空间中加法和数乘这两种运算性质的运算，我们称之为线性运算。

所以，我们可以给出论断:

Claim

向量空间的本质上是由集合与满足特定公理的线性运算共同构成的一种代数结构，二者缺一不可。

若集合 V 上定义的加法或数乘运算不满足线性运算的 8 条公理，则构成的结构不再是向量空间；

若集合 V_1 和集合 V_2 定义的运算满足公理但具体的定义方式不同（如调整数乘的域或加法规则），则形成不同的向量空间。

Example

例如：集合 $\mathbb{R}^n$ 按标准加法与实数乘构成实向量空间 $\mathbb{R}^n$；
若将数乘的域改为复数域 $\mathbb{C}$，则构成复向量空间 $\mathbb{C}^n$，二者为不同代数结构。

1.1.3 向量空间的五条重要性质

在介绍完向量空间的基本定义，窥探其本质后，我们可以对其数学性质做进一步挖掘：得到其 5 条重要的性质与推论。

Property

在向量空间中成立 $(-1)\cdot\boldsymbol{v} = -\boldsymbol{v}$ 和 $0\cdot\boldsymbol{v} = \mathbf{0}$
其中，v 为向量空间中的任意向量。

1. 证明 $0\cdot\boldsymbol{v} = \mathbf{0}$:

$$\begin{aligned}\boldsymbol{v} + 0\cdot\boldsymbol{v} &= 1\cdot\boldsymbol{v} + 0\cdot\boldsymbol{v}\\ &= (1+0)\cdot\boldsymbol{v}\\ &= 1\cdot\boldsymbol{v} = \boldsymbol{v}.\end{aligned}$$

   Add $-\boldsymbol{v}$ to both sides:

$$(\boldsymbol{v} + 0\cdot\boldsymbol{v}) + (-\boldsymbol{v}) = \boldsymbol{v} + (-\boldsymbol{v}) \implies 0\cdot\boldsymbol{v} = \mathbf{0}.$$

2. 证明 $(-1)\cdot\boldsymbol{v} = -\boldsymbol{v}$:

$$\begin{aligned}\boldsymbol{v} + (-1)\cdot\boldsymbol{v} &= 1\cdot\boldsymbol{v} + (-1)\cdot\boldsymbol{v}\\ &= (1+(-1))\cdot\boldsymbol{v}\\ &= 0\cdot\boldsymbol{v} = \mathbf{0}.\end{aligned}$$

   因此，$(-1)\cdot\boldsymbol{v}$ 是 $\boldsymbol{v}$ 的加法逆元，所以 $(-1)\cdot\boldsymbol{v} = -\boldsymbol{v}$。

接着我们给出其它三条重要性质：

Property

1. **零向量唯一**: 对于任意 $\alpha \in V$，若 $\alpha + 0_1 = \alpha$ 且 $\alpha + 0_2 = \alpha$，令 $\alpha = 0_1$ 代入 $\alpha + 0_2 = \alpha$ 可得 $0_1 + 0_2 = 0_1$，再令 $\alpha = 0_2$ 代入 $\alpha + 0_1 = \alpha$ 可得 $0_2 + 0_1 = 0_2$，所以 $0_1 = 0_2$，即零向量是唯一的。

2. **任一向量其负向量唯一**: 若 $\alpha+\beta = 0$ 且 $\alpha+\gamma = 0$, 则 $\beta = \beta+0 = \beta+(\alpha+\gamma) = (\beta + \alpha) + \gamma = 0 + \gamma = \gamma$，所以任一向量的负向量是唯一的。

3. **若 $\lambda\alpha = 0$，则 $\lambda = 0$ 或 $\alpha = 0$**: 假设 $\lambda \neq 0$，因为数域 F 中数的运算性质，λ 存在逆元 λ^{-1}，在 $\lambda\alpha = 0$ 两边同时左乘 λ^{-1}，可得 $\lambda^{-1}(\lambda\alpha) = \lambda^{-1} \cdot 0$，根据数乘结合律和 $0 \cdot \alpha = 0$，得到 $(\lambda^{-1}\lambda)\alpha = 0$，即 $1 \cdot \alpha = \alpha = 0$，所以若 $\lambda\alpha = 0$，则 $\lambda = 0$ 或 $\alpha = 0$。

1.1.4 向量空间的判断

判断给定集合 V 是否为向量空间，一般按以下步骤：

1. 考察加法和数乘的封闭性，即检查运算结果是否仍在集合 V 内。
2. 考察零元是否存在于集合 V 中。
3. 逐一考察验证八条公理是否成立。

非向量空间的例子)：

Example

- 集合 $V = \{(x, y) \in \mathbb{R}^2 | x+y = 1\}$ 不是向量空间, 因为若 (x_1, y_1) 和 (x_2, y_2) 满足 $x_1+y_1 = 1$ 且 $x_2+y_2 = 1$, (x_1+x_2, y_1+y_2) 不一定满足 $(x_1+x_2)+(y_1+y_2) = 1$, 即对加法不封闭，并且该集合不包含零向量。
- 次数严格等于 n 的多项式集合不是向量空间, 因为对加法不封闭。例如：$(x^n + x) + (-x^n + 1) = x + 1$，结果不再是次数严格等于 n 的多项式。

【Remark】：

注意到，多项式空间 $P_n[x]$（次数不超过 n 的实系数多项式全体）是向量空间。

1.1.5 课后习题

Exercise

1. 设 V 是向量空间，$\alpha, \beta \in V$。证明：方程 $\alpha + X = \beta$ 在 V 中有唯一解。
2. 在 $\mathbb{R}^2$ 中，定义 $(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 b_2)$ 和 $k(a_1, a_2) = (ka_1, a_2)$，证明其不构成向量空间。
3. 设集合 $V = \{(x, y) | x \geq 0, y \in \mathbb{R}\}$，定义普通的向量加法和实数域上的数乘运算。证明 V 不是线性空间。

1.2 子空间

1.2.1 子空间的定义与判定

Definition

子空间（subspace）： 设 V 是数域 F 上的向量空间，若 V 的子集 W 对 V 中定义的加法和数乘运算也构成数域 F 上的向量空间，则称 W 是 V 的子空间。

在这里，沿用上节向量空间的知识，我们不加证明的的给出子空间的判定条件：

Method

【子空间的判定条件】：若定义在数域 F 上的向量空间 V 的子集 W 是子空间，当且仅当以下 2 个判定条件成立：

1. 对任意 $w_1, w_2 \in W$ 以及任意 $a, b \in \mathbb{F}$，都有 $aw_1 + bw_2 \in W$;
2. 包含零向量：W 中存在零向量 $\mathbf{0}$。

我们来考察以下几个例子：

- 对于 $\mathbb{R}^2$ 中 $W = \{(x, y) \mid x + y = 0\}$：
  - 取 $w_1 = (a, -a)$，$w_2 = (b, -b) \in W$，任取 $a, b \in \mathbb{R}$，计算 $aw_1 + bw_2 = (a^2 + b^2, -a^2 - b^2)$，满足 $x + y = 0$，故 $aw_1 + bw_2 \in W$。
  - W 包含零向量 $(0, 0)$（当 $x = y = 0$ 时，$x + y = 0$ 成立）。

  因此 W 是 $\mathbb{R}^2$ 的子空间。
- 对于 $\mathbb{R}^2$ 中 $W_1 = \{(x, 0) \mid x \in \mathbb{R}\}$：
  - 取 $w_1 = (x_1, 0)$，$w_2 = (x_2, 0) \in W_1$，任取 $a, b \in \mathbb{R}$，$aw_1 + bw_2 = (ax_1 + bx_2, 0) \in W_1$。

  - W_1 包含零向量 $(0,0)$。

    因此 W_1 是 $\mathbb{R}^2$ 的子空间。

- 对于 $\mathbb{R}^2$ 中 $W_2 = \{(x,x) \mid x \in \mathbb{R}\}$:

  - 取 $w_1 = (x_1,x_1)$，$w_2 = (x_2,x_2) \in W_2$，任取 $a,b \in \mathbb{R}$，$aw_1 + bw_2 = (ax_1 + bx_2, ax_1 + bx_2) \in W_2$。

  - W_2 包含零向量 $(0,0)$。

    因此 W_2 是 $\mathbb{R}^2$ 的子空间。

- 对于多项式空间 $P_n(x)$ 中 $W_3 = \{a_nx^n + \cdots + a_1x \mid a_i \in \mathbb{R}, i = 1,\ldots,n\}$:

  - 取 $p(x) = a_nx^n + \cdots + a_1x$，$q(x) = b_nx^n + \cdots + b_1x \in W_3$，任取 $a,b \in \mathbb{R}$，$ap(x) + bq(x) = (aa_n + bb_n)x^n + \cdots + (aa_1 + bb_1)x \in W_3$。

  - W_3 包含零多项式（所有系数为 0 的多项式）。

    因此 W_3 是 $P_n(x)$ 的子空间。

- 对于多项式空间 $P_n(x)$ 中 $W_4 = \{a_nx^n \mid a_n \in \mathbb{R}\}$:

  - 取 $f(x) = a_nx^n$，$g(x) = b_nx^n \in W_4$，任取 $a,b \in \mathbb{R}$，$af(x) + bg(x) = (aa_n + bb_n)x^n \in W_4$。

  - W_4 包含零多项式（当 $a_n = 0$ 时）。

    因此 W_4 是 $P_n(x)$ 的子空间。

- 对于矩阵空间 $M_{m\times n}(\mathbb{R})$ 中 W_5（第一行任意、其余行全为 0 的矩阵集合）:

  - 取 $A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix}$，$B = \begin{pmatrix} b_{11} & \ldots & b_{1n} \\ 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix} \in W_5$，任取 $a,b \in \mathbb{R}$，$aA + bB = \begin{pmatrix} aa_{11} + bb_{11} & \ldots & aa_{1n} + bb_{1n} \\ 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix} \in W_5$。

  - W_5 包含零矩阵（所有元素为 0 的矩阵）。

    因此 W_5 是矩阵空间 $M_{m\times n}(\mathbb{R})$ 的子空间。

接下来是一个经典的例子:

Example

$\mathbb{R}_+ \subset \mathbb{R}$ 不是 $\mathbb{R}$ 的子空间。
其中 $R_+ = \{x \in R\ x > 0\}$

$\mathbb{R}$ 作为定义在实数域上的向量空间，运算为普通实数加法和实数对向量的数乘。但是 $\mathbb{R}_+$ 不满足：

零向量的存在性：$\mathbb{R}$ 的零向量是 0，但 $\mathbb{R}_+ = \{x > 0\}$ 不包含 0。

数乘封闭性：取 $x = 1 \in \mathbb{R}_+$，数域中取 $a = -1 \in \mathbb{R}$，则数乘结果为 $a \cdot x = -1 \times 1 = -1 \notin \mathbb{R}_+$，说明数乘不封闭。

重新定义 $\mathbb{R}_+$ 上的加法和数乘（脱离 $\mathbb{R}$ 的普通运算），使它满足向量空间公理：定义新运算：

加法：对任意 $x, y \in \mathbb{R}_+$，定义：$x \oplus y := x \cdot y$（用实数乘法代替加法）。

数乘：对任意 $x \in \mathbb{R}_+$，$a \in \mathbb{R}$（数域），定义 $a \odot x := x^a$（用实数的幂运算代替数乘）。

即可满足。

1.2.2 课后练习

Exercise

设 $V = M_{n\times n}(\mathbb{C})$ 是复 $n \times n$ 矩阵构成的 $\mathbb{C}$-向量空间，我们定义

$$W_{*,+} := \{A \in V : {}^t\overline{A} = A\} \quad 和 \quad W_{*,-} := \{A \in V : {}^t\overline{A} = -A\}$$

这里 ${}^t\overline{A} = {}^t\overline{(a_{ij})}_{i,j} := (\overline{a}_{ji})_{i,j}$，其中 $\overline{a}_{ij} = \overline{x + \sqrt{-1}y} := x - \sqrt{-1}y \in \mathbb{C}$。
证明 $W_{*,+}$ 是 V 的 $\mathbb{R}$-向量子空间；
证否 $W_{*,-}$ 是 V 的 $\mathbb{C}$-向量子空间。

1.3 线性组合与张成

1.3.1 线性组合与线性表出

Definition

线性组合（linear combination）： 在向量空间 V 中，对于非空子集 S，若向量 $v \in V$ 能够表示为 S 中有限个向量 $u_1, u_2, \cdots, u_n$ 与数域 F 中系数 $a_1, a_2, \cdots, a_n$ 的组合形式，即

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n,$$

则称 v 是 S 中向量的线性组合，
其中 $a_1, a_2, \cdots, a_n$ 被称为组合系数。

同样的，我们称 v 可以有 S 中向量的线性表出（Linearly represented）。

根据向量空间的运算性质，对于任意 $v \in V$，都有 $0v = 0$，**这表明零向量是任何非空子集的线性组合**。

【Remark】:

- 线性组合体现了向量空间中向量之间通过数乘与加法构建的联系。这种组合方式基于向量空间所定义的数乘和加法运算规则。
- 组合系数来自于数域 F，数域的性质（如有理数域、实数域、复数域等不同数域）会影响向量线性组合的表现形式和性质。

接下来我们考察两个经典的例子:

在二维向量空间 $\mathbb{R}^2$ 中，对于子集 $S = \{(1,0),(0,1)\}$，向量 $\vec{v} = (3,4)$ 可表示为 $\vec{v} = 3 \times (1,0) + 4 \times (0,1)$，这里 3 和 4 作为组合系数，清晰地展示了 $\vec{v}$ 是 S 中向量的线性组合。

在多项式向量空间 $P_2(\mathbb{R})$（由次数不超过 2 的实系数多项式构成）中，对于子集 $S = \{1, x, x^2\}$，多项式 $p(x) = 2 - 3x + 5x^2$ 可表示为 $p(x) = 2 \times 1 + (-3) \times x + 5 \times x^2$。这表明在多项式向量空间中，通过基向量（这里的 $1, x, x^2$ 类似于向量空间的基）的线性组合可以得到该空间内的具体多项式。

1.3.2 张成（Spanning）

Definition

张成（spanning）:

- （定义 1）：设 S 是向量空间 V 的非空子集，S 的张成 $\text{span}(S)$ 是由 S 中所有向量的线性组合所构成的集合。即：若 $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n \in V$，则：

$$\text{Span}(\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n) = \{\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n \mid \alpha_1, \alpha_2, \cdots, \alpha_n \in F\}$$

  特别规定 $\text{span}(\varnothing) = \{0\}$，即空集的张成是仅包含零向量的集合。

- （定义 2）：$\text{span}(S)$ 也可被视作 V 中包含子集 S 的最小子空间。

  这意味着一方面 $\text{span}(S)$ 是 V 的子空间，满足子空间的判定条件；另一方面，任何其他包含 S 的 V 的子空间必然也包含 $\text{span}(S)$，体现了 $\text{span}(S)$ 在包含 S 的子空间中的最小性。

容易观察到，向量空间的向量集合张成的结果依然是一个向量空间，满足运算封闭性和 8 条公理。

接下来，我们考察几个典型的例子：

- 在 $P_2(\mathbb{R})$ 中，取 $S = \{1, x, x^2\}$，$\text{span}(S)$ 包含所有形如 $a + bx + cx^2$（$a, b, c \in \mathbb{R}$）的多项式，即 $\text{span}(S) = P_2(\mathbb{R})$，且 span（S）的结果二次多项式空间的全空间。

- 在 $\mathbb{R}^3$ 里，设 $\vec{v} = (1, 2, 3)$，$\text{span}(\{\vec{v}\}) = \{k(1, 2, 3) \mid k \in \mathbb{R}\}$，几何上是过原点且与 $\vec{v}$ 共线的直线，该直线是 $\mathbb{R}^3$ 的一个子空间（见前节）。

- 在 $\mathbb{R}^3$ 中，集合 $\{(1, 0, 0), (0, 1, 0)\}$ 的张成 $\text{span}\{(1, 0, 0), (0, 1, 0)\}$ 是 $\mathbb{R}^3$ 中形如 $a(1, 0, 0) + b(0, 1, 0) = (a, b, 0)$（$a, b \in \mathbb{R}$）的所有向量构成的集合，从几何角度看，它张成的是一个平面，对应于 xy-平面，是 $\mathbb{R}^3$ 的一个子空间。这一示例直观地展示了张成集合在三维空间中的几何形态以及作为子空间的性质。

- 在矩阵向量空间 $M_{2\times2}(\mathbb{R})$（所有 2×2 实矩阵构成的向量空间）中，对于子集 $S = \left\{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right\}$，其张成 $\text{span}(S)$ 是所有形如 $a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}$（$a, b \in \mathbb{R}$）的矩阵集合，它是 $M_{2\times2}(\mathbb{R})$ 的一个子空间。此例将张成的概念拓展到矩阵向量空间，展示了在不同类型向量空间中张成集合的构建方式和性质。

由此，我们还可以证明一条有用的结论：

Claim

当域 $\mathbb{F}$ 的特征 $\mathrm{char}(\mathbb{F}) \neq 2$ 时，n 阶全矩阵空间 $M_n(\mathbb{F})$ 等于 $\mathbb{F}$ 上由对称矩阵空间 W_+ 和反对称矩阵空间 W_- 张成的线性空间

证明.

- 对任意矩阵 $A \in M_n(\mathbb{F})$，构造两个矩阵：

$$A_+ = \frac{1}{2}(A + A^T), \quad A_- = \frac{1}{2}(A - A^T)$$

当 $\mathrm{char}(\mathbb{F}) \neq 2$ 时，2 在 $\mathbb{F}$ 中可逆（即 $\frac{1}{2}$ 存在），因此 A_+ 和 A_- 都是 $\mathbb{F}$ 上的合法矩阵。

- 验证 A_+ 是对称矩阵，A_- 是反对称矩阵：
  - 对 A_+，有：

$$(A_+)^T = \left(\frac{1}{2}(A + A^T)\right)^T = \frac{1}{2}(A^T + (A^T)^T) = \frac{1}{2}(A^T + A) = A_+$$

  - 对 A_-，有：

$$(A_-)^T = \left(\frac{1}{2}(A - A^T)\right)^T = \frac{1}{2}(A^T - (A^T)^T) = \frac{1}{2}(A^T - A) = -A_-$$

- **证明 $M_n(\mathbb{F}) = \mathbf{Span}_{\mathbb{F}}(W_+ \cup W_-)$：**
  - 由分解式 $A = A_+ + A_-$ 可知，任意矩阵都能表示为对称矩阵与反对称矩阵的和，因此 $M_n(\mathbb{F}) \subseteq \mathrm{Span}_{\mathbb{F}}(W_+ \cup W_-)$。
  - 另一方面，对称矩阵和反对称矩阵都是全矩阵空间的元素，它们的张成空间自然包含于全矩阵空间，即 $\mathrm{Span}_{\mathbb{F}}(W_+ \cup W_-) \subseteq M_n(\mathbb{F})$。
  - 综上，两者相等。

- **关于特征不能等于 2 的说明：**
  - 当 $\mathrm{char}(\mathbb{F}) = 2$ 时，域中元素满足 $1 + 1 = 0$，即 $2 = 0$，此时 $\frac{1}{2}$ 无意义，上述分解式 $A_+ = \frac{1}{2}(A + A^T)$ 和 $A_- = \frac{1}{2}(A - A^T)$ 无法定义。
  - 进一步，在特征为 2 的域中，$-1 = 1$，因此反对称矩阵与对称矩阵的定义变得一致（$A^T = -A$ 等价于 $A^T = A$），导致 $W_+ = W_-$，它们的张成空间无法覆盖整个全矩阵空间。
  - 因此，只有当 $\mathrm{char}(\mathbb{F}) \neq 2$ 时，上述分解和结论才成立。

□

1.3.3 课后练习

Exercise

证明 $\mathbb{F}[x]=\mathrm{Span}_{\mathbb{F}}(W_e\cup W_o)$，其中 $W_e:=\{f(x)\in\mathbb{F}[x]\mid f(x)=f(-x)\}$（偶函数多项式集合），$W_o:=\{f(x)\in\mathbb{F}[x]\mid f(x)=-f(-x)\}$（奇函数多项式集合）。（请特别注意 $\mathrm{char}(\mathbb{F})=2$ 的情形）

1.4 线性相关性

线性相关

Definition

线性相关（Linearly dependent）： 若向量空间 V 的子集 S 中存在**有限个**不同向量 $u_1,u_2,\cdots,u_n$ 以及**不全为零**的数域 F 中的系数 $a_1,a_2,\cdots,a_n$，使得

$$a_1u_1+a_2u_2+\cdots+a_nu_n=0,$$

则称子集 S 是**线性相关的**，也说 S 中的**向量线性相关**。

【判定方法】：对于给定向量集合，可通过建立线性方程组，看是否存在非零解来判断线性相关性。

接下来，我们将展示两个例子，以加深理解：

Example

- 在 $\mathbb{R}^2$ 中，向量集合 $S=\{(1,2),(2,4)\}$ 是线性相关的。设 $a(1,2)+b(2,4)=(0,0)$，即 $(a+2b,2a+4b)=(0,0)$，得到方程组 $\begin{cases}a+2b=0\\2a+4b=0\end{cases}$，取 $a=-2$，$b=1$ 就是一组非零解，说明存在零向量的非平凡表示，所以向量集合 S 线性相关。

- 在多项式向量空间 $P_2(\mathbb{R})$ 中，集合 $S=\{1+x,2+2x,x^2\}$ 是线性相关的。设 $a(1+x)+b(2+2x)+cx^2=0$（这里 0 是零多项式），整理得 $(a+2b)+(a+2b)x+cx^2=0$，取 $a=2$，$b=-1$，$c=0$ 就是一组非零解，表明该集合线性相关。

线性无关

Definition

线性无关（Linearly independent）： 若向量空间 V 的子集 S 不是线性相关的，则称 S 是线性无关的，也说 S 中的向量线性无关。

即：若 $a_1 = a_2 = \cdots = a_n = 0$ 时才有 $a_1u_1 + a_2u_2 + \cdots + a_nu_n = 0$，称这种情况为线性无关。

接下来我们对非零向量与空集进行辨析：

Property

- 空集是线性无关的，因为线性相关集合必须非空。（空集不等于 0 集）
- 由单个非零向量构成的集合是线性无关的。因为若 $\{u\}$ 线性相关，则存在非零标量 a 使 $au = 0$，那么 $u = a^{-1}(au) = a^{-1}0 = 0$，与 u 非零矛盾。

让我们来考察接下来两个例子，加深对线性无关的理解：

Example

- 在 $\mathbb{R}^3$ 中，向量集合 $S = \{(1,0,0),(0,1,0),(0,0,1)\}$ 是线性无关的。设 $a(1,0,0)+b(0,1,0)+c(0,0,1) = (0,0,0)$，即 $(a,b,c) = (0,0,0)$，只能得到 $a = b = c = 0$，所以该集合线性无关。
- 在定义在 $[0,1]$ 上的连续函数构成的向量空间 $C[0,1]$）里，集合 $S = \{1, x, x^2\}$ 是线性无关的。假设 $a \times 1 + b \times x + c \times x^2 = 0$（这里 0 是零函数）对所有 $x \in [0,1]$ 都成立，根据多项式函数的性质，只有 $a = b = c = 0$ 时等式才恒成立，所以集合 S 线性无关。

线性无关性在线性代数中具有相当重要的地位，我们接下来将讨论本节的两个主定理：

Theorem

定理 1： 设 V 是一个向量空间，且 $S_1 \subseteq S_2 \subseteq V$。若 S_1 线性相关，则 S_2 线性相关；且若 S_2 线性无关，则 S_1 线性无关。

Theorem

定理 2： 设 S 是向量空间 V 的一个线性无关子集，v 是 V 中一个不属于 S 的向量。那么 $S \cup \{v\}$ 线性相关当且仅当 $v \in \mathrm{span}(S)$。

1.5 基与维度

1.5.1 基的定义

Definition

基/基底（basis）： 向量空间 V 的一个基底 β 是 V 的一个**线性无关子集**，并且它**生成**（generates or spans）V。
基向量（Base vectors）： 如果 β 是 V 的一个基底，我们也说 β 里面的向量是 V 的基向量。

【Remark】:

“基”和“基底”是同一概念的不同中文表述，二者的英文原文均为“basis”（复数形式为“bases”）

因为涉及佐恩引理，极大元和塔的概念, 我们这里不加证明地给出：

Claim

每个向量空间都有一个基，这一论断。

因此，我们对基底应该有如下的认识：
基不仅仅是可以张成全空间的向量组合，同时这个组合线性无关，即判断条件为：
1. 张成全空间；
2. 线性无关性
用下图生动反映了基底与生成集与线性无关集的关系：

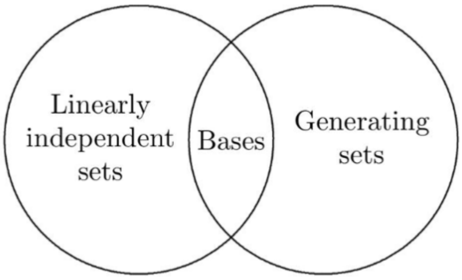


图 1.1: 基和生成集与线性无关集的关系

Example

在 $\mathbb{F}^n$ 中，令 $e_1 = (1,0,0,\dots,0)$，$e_2 = (0,1,0,\dots,0)$，$\dots$，$e_n = (0,0,\dots,0,1)$；$\{e_1,e_2,\dots,e_n\}$ 就是 $\mathbb{F}^n$ 的一个基底，被称为 $\mathbb{F}^n$ 的标准基底（standard basis）。
在 $P_n(F)$ 中，集合 $\{1,x,x^2,\dots,x^n\}$ 是一个基底。我们也称这个基底为 $P_n(F)$ 的标准基底。

【Remark】:

一个常见的误区是：假设 $S = \{s_1, s_2, \cdots, s_n\}$ 是一组基，但是不考虑 V 是否为有限维

接下来，我们给出另一个易错的论断：

Claim

子空间与全空间基的交集不一定是子空间的基。

Example

设 $V = \mathbb{R}^2$（二维实向量空间），子空间 $W = \{(x,-x) \mid x \in \mathbb{R}\}$。可验证 W 对线性组合封闭，因此是 V 的子空间。
定义全空间的基取 V 的一组基 $S = \{(1,0),(0,1)\}$
验证交集不是子空间的基计算 $W \cap S$，即找既属于 W 又属于 S 的向量。但 S 中的向量是 $(1,0)$（纵坐标为 0）和 $(0,1)$（横坐标为 0），都不满足 W 中“横坐标与纵坐标相反”的条件，因此 $W \cap S = \emptyset$（空集）。
而子空间的基必须是非空的线性无关组且能张成子空间，空集显然不满足，因此 $W \cap S$ 不是 W 的基。

1.5.2 基底的重要性质

接下来，我们将展示本节的 2 条主定理：

Theorem

设 V 是一个向量空间，且 $\beta = \{u_1, u_2, \ldots, u_n\}$ 是 V 的一个子集。那么 β 是 V 的一个基底**当且仅当**每个 $v \in V$ 都可以唯一地表示为 β 中向量的线性组合，即可以表示为

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n$$

其中 $a_1, a_2, \ldots, a_n$ 是唯一的标量且均属于数域 F。

证明. **正向证明：**

设 β 是 V 的一个基底。如果 $v \in V$，那么 $v \in \text{span}(\beta)$，因为 $\text{span}(\beta) = V$。因此 v 是 β 中向量的线性组合。假设

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n \quad 和 \quad v = b_1u_1 + b_2u_2 + \cdots + b_nu_n$$

是 v 的两个这样的表示。用第一个方程减去第二个方程得到

$$0 = (a_1 - b_1)u_1 + (a_2 - b_2)u_2 + \cdots + (a_n - b_n)u_n$$

由于 β 是线性无关的，所以 $a_1 - b_1 = a_2 - b_2 = \cdots = a_n - b_n = 0$。因此 $a_1 = b_1$，$a_2 = b_2$，…，$a_n = b_n$，所以 v 可以唯一地表示为 β 中向量的线性组合。

反向证明：

已知每个 $v \in V$ 都能唯一地表示为 $\beta = \{u_1, u_2, \ldots, u_n\}$ 中向量的线性组合，所以 $\text{span}(\beta) = V$，即 β 生成 V。

线性无关：设 $a_1u_1 + a_2u_2 + \cdots + a_nu_n = 0$，又 $0 = 0u_1 + 0u_2 + \cdots + 0u_n$，根据表示唯一性，可得 $a_1 = a_2 = \cdots = a_n = 0$，故 β 线性无关。

综上，β 线性无关且生成 V，所以 β 是 V 的基底。 □

接着，我们将讨论有限生成向量空间的基底：

Theorem

如果一个向量空间 V 由一个有限集 S 生成，那么 S 一定存在某个子集是 V 的一个基底。
因此 V 有一个有限基底。

证明. 如果 $S = \varnothing$ 或 $S = \{0\}$，那么 $V = \{0\}$ 且 $\varnothing$ 是 S 的一个子集，它是 V 的一个基底。

否则 S 包含一个非零向量 u_1。$\{u_1\}$ 是一个线性无关集。如果可能的话，继续在 S 中选择向量 $u_2, \ldots, u_k$，使得 $\{u_1, u_2, \ldots, u_k\}$ 是线性无关的。

由于 S 是一个有限集，我们最终一定会到达一个阶段，此时 $\beta = \{u_1, u_2, \ldots, u_k\}$ 是 S 的一个线性无关子集，但是将 S 中任何不在 β 中的向量添加到 β 中都会产生一个线性相关集。我们声称 β 是 V 的一个基底。

因为根据构造 β 是线性无关的，所以只需证明 β 生成 V。我们需要证明 $S \subseteq \mathrm{span}(\beta)$。令 $v \in S$。如果 $v \in \beta$，那么显然 $v \in \mathrm{span}(\beta)$。否则，如果 $v \notin \beta$，那么前面的构造表明 $\beta \cup \{v\}$ 是线性相关的。

因此 $v \in \mathrm{span}(\beta)$。

因此 $S \subseteq \mathrm{span}(\beta)$。 □

1.5.3 n 维向量及其坐标表示

思考：结合前言知识里面的向量加减法，我们已经有了向量坐标化的思想。现在又在向量空间规定了基的定义，我们就可以提出有序基，标准基底与坐标表示的概念了。

在向量空间中，基是一组线性无关且能生成整个空间的向量集合。而有序基则是为基中的向量赋予了特定顺序的基，即基中向量的排列顺序是固定且有意义的。

Definition

有序基（ordered basis）：给定的基底中向量的顺序 $\alpha_1, \alpha_2, \ldots, \alpha_n$ 是固定且被明确指定的。则称 $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ 为向量空间 V 的一个有序基。
标准基底（standard base）： F^n 的标准基底是 $\{e_1, e_2, \ldots, e_n\}$ ，e_i 第 i 个分量为 1 ，其余为 0 ，它线性无关且生成 F^n 。
坐标表示 (Coordinate representation)：对于 F^n 的有序基底 $\beta = \{v_1, v_2, \ldots, v_n\}$ ，向量 $\mathbf{u} \in F^n$ 可唯一表示为 $\mathbf{u} = \sum_{i=1}^{n} c_i v_i$ ，$(c_1, c_2, \ldots, c_n)$ 是 $\mathbf{u}$ 关于 β 的坐标，记为 $[\mathbf{u}]_\beta$ 。

1.5.4 有限维与维度的定义

Definition

维度 (dimension)：V 的每个基底中向量的唯一数量被称为 V 的维度 (dimension)，记为 $\dim(V)$。

一个向量空间如果有一个由有限个向量组成的基底，就被称为有限维的。**一个不是有限维的向量空间被称为无限维的**。

一般来说，本讲义的讨论都局限于有限维中。

从该定义我们可以观察到，若向量空间 V 存在有限基，则 V 的任意基所包含的向量个数必然相同。这一性质表明，基的向量数量是向量空间 V 的固有属性 (见后半部分的 lemma)，不依赖于基的具体选取。

联系前言部分，我们再通过几个例子，加深对维度的理解

Example

向量空间 $\{0\}$ 的维度为零。
向量空间 $\mathbb{F}^n$ 的维度为 n。
向量空间 $M_{m\times n}(F)$ 的维度为 mn。
向量空间 $P_n(F)$ 的维度为 $n+1$。

接下来，我们来证明：

Lemma

若向量空间 V 存在有限基，则 V 的任意基所包含的**向量个数必然相同**。

证明. **有限维情形：**

设 V 是有限维向量空间，$S_1=\{u_1,u_2,\dots,u_m\}$ 与 $S_2=\{v_1,v_2,\dots,v_n\}$ 是 V 的两个基。

因 S_1 张成 V 且 S_2 线性无关，我们现在需要证明：线性无关集的基数不超过张成集的基数，即 $n\le m$。

因为 $\mathrm{span}(S)=V$，所以 L 中第一个向量 v_1 可由 S 线性表出：$v_1=a_1u_1+a_2u_2+\cdots+a_mu_m$.

由于 $v_1\neq 0$（线性无关集中向量非零），至少有一个系数 $a_i\neq 0$。不妨设 $a_1\neq 0$，则 u_1 可由 $\{v_1,u_2,\dots,u_m\}$ 线性表出，因此：

$\mathrm{span}(\{v_1,u_2,\dots,u_m\})=\mathrm{span}(S)=V.$

重复该过程，假设 $n>m$（即 S_2 的向量个数比 S_1 多）：

S_1 只有 m 个向量，因此最多能替换 m 次。替换 m 次后，S_1 的所有向量 $u_1,u_2,\dots,u_m$ 会被完全替换成 $v_1,v_2,\dots,v_m$，此时张成组为 $\{v_1,v_2,\dots,v_m\}$，且 $\mathrm{span}(\{v_1,v_2,\dots,v_m\})=V$。

但 S_2 还有第 $m+1$ 个向量 v_{m+1}，由于 $v_{m+1}\in V$，它必须能由 $\{v_1,v_2,\dots,v_m\}$ 线性表出——这与"$S_2=\{v_1,v_2,\dots,v_n\}$ 线性无关"矛盾（线性无关组中任意向量都不能由其余向量线性表出）。

同理，交换 S_1 与 S_2 的角色，可得 $m\le n$。

综上，$m=n$，即有限维空间的基有相同元素个数。

无限维情形：

设 V 是无限维向量空间，S_1,S_2 为其两个基。

对任意 $u\in S_1$，因 S_2 张成 V，u 可表为 S_2 中有限个向量的组合。记所有这些有限子集的并为 $F\subseteq S_2$，则 $\mathrm{span}(F)=V$。

由 S_1 的线性无关性，得 $|S_1|\le|F|\le|S_2|$。

同理可得 $|S_2|\le|S_1|$。

由基数反对称性，$|S_1| = |S_2|$，即无限维空间的基有相同基数。

故，向量空间的任意两个基具有相同基数，故维数 $\dim(V)$ 是良定义的。 □

接下来，我们来讨论一个著名的公式：

Theorem

若 W_1 和 W_2 是向量空间 V 的有限维子空间，则子空间 $W_1 + W_2$ 是有限维的，且

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

证明. 设 W_1, W_2 是向量空间 V 的有限维子空间，令 $\dim(W_1 \cap W_2) = r$。

取 $W_1 \cap W_2$ 的一组基：$\alpha_1, \alpha_2, \ldots, \alpha_r$。

由于 $W_1 \cap W_2 \subseteq W_1$，将这组基扩展为 W_1 的基：$\alpha_1, \alpha_2, \ldots, \alpha_r, \beta_1, \beta_2, \ldots, \beta_s$，此时 $\dim(W_1) = r + s$。

同理，将 $W_1 \cap W_2$ 的基扩展为 W_2 的基：$\alpha_1, \alpha_2, \ldots, \alpha_r, \gamma_1, \gamma_2, \ldots, \gamma_t$，此时 $\dim(W_2) = r + t$。

需证明向量组 $\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s, \gamma_1, \ldots, \gamma_t$ 是 $W_1 + W_2$ 的基。

首先证明生成性：对任意 $\xi \in W_1 + W_2$，存在 $\xi_1 \in W_1$，$\xi_2 \in W_2$，使得 $\xi = \xi_1 + \xi_2$。

因 $\xi_1 \in W_1$，故 $\xi_1 = \sum_{i=1}^r a_i\alpha_i + \sum_{j=1}^s b_j\beta_j$。

因 $\xi_2 \in W_2$，故 $\xi_2 = \sum_{i=1}^r c_i\alpha_i + \sum_{k=1}^t d_k\gamma_k$。

因此：

$$\xi = \xi_1 + \xi_2 = \sum_{i=1}^r (a_i + c_i)\alpha_i + \sum_{j=1}^s b_j\beta_j + \sum_{k=1}^t d_k\gamma_k$$

即 ξ 可由该向量组线性表示。

再证明线性无关性：假设存在数 $k_1, \ldots, k_r, l_1, \ldots, l_s, m_1, \ldots, m_t$，使得：

$$\sum_{i=1}^r k_i\alpha_i + \sum_{j=1}^s l_j\beta_j + \sum_{k=1}^t m_k\gamma_k = 0 \tag{1}$$

令 $\eta = \sum_{i=1}^r k_i\alpha_i + \sum_{j=1}^s l_j\beta_j$，则由式 (1) 得 $\eta = -\sum_{k=1}^t m_k\gamma_k$。

一方面，$\eta \in W_1$；另一方面，$-\sum_{k=1}^t m_k\gamma_k \in W_2$，故 $\eta \in W_1 \cap W_2$。

因此 η 可由 $\alpha_1, \ldots, \alpha_r$ 线性表示，即存在 $n_1, \ldots, n_r$，使得 $\eta = \sum_{i=1}^r n_i\alpha_i$。

结合 $\eta = \sum_{i=1}^r k_i\alpha_i + \sum_{j=1}^s l_j\beta_j$，且 $\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s$ 线性无关，可得 $l_j = 0$ $(j = 1, \ldots, s)$，且 $k_i = n_i$ $(i = 1, \ldots, r)$。

此时 $\eta = \sum_{i=1}^r n_i\alpha_i = -\sum_{k=1}^t m_k\gamma_k$，即：

$$\sum_{i=1}^r n_i\alpha_i + \sum_{k=1}^t m_k\gamma_k = 0$$

因 $\alpha_1, \ldots, \alpha_r, \gamma_1, \ldots, \gamma_t$ 线性无关，故 $n_i = 0$ $(i = 1, \ldots, r)$，$m_k = 0$ $(k = 1, \ldots, t)$，进而 $k_i = 0$ $(i = 1, \ldots, r)$。

因此该向量组线性无关，确为 W_1+W_2 的基，含 $r+s+t$ 个向量，故 $\dim(W_1+W_2)=r+s+t$。

又 $\dim(W_1)=r+s$，$\dim(W_2)=r+t$，$\dim(W_1\cap W_2)=r$，因此：

$$\dim(W_1+W_2)=\dim(W_1)+\dim(W_2)-\dim(W_1\cap W_2)$$

综上，定理得证。

□

1.5.5 课后练习

Exercise

1. 对于固定的 $a\in\mathbb{R}$，确定由 $\{f\in P_n(\mathbb{R}):f(a)=0\}$ 定义的 $P_n(\mathbb{R})$ 的子空间的维度。
2. 设 W 是有限维向量空间 V 的一个子空间。那么 W 是有限维的，且 $\dim(W)\leq\dim(V)$。此外，如果 $\dim(W)=\dim(V)$，那么 $V=W$。
3. 考虑由所有迹为零的 $n\times n$ 矩阵构成的集合 V（若 $A=(a_{ij})$ 是一个 $n\times n$ 矩阵，A 的迹是和 $\sum_{i=1}^n a_{ii}$）。

   1. 证明它是一个向量空间。
   2. 对于所有迹为零的 2×2 矩阵，求该向量空间的一个基。
   3. 求 V 的一个基及其维数。

4. 设 $V=M_{n\times n}(\mathbb{C})$ 是复 $n\times n$ 矩阵构成的 $\mathbb{C}$-向量空间，我们定义

$$W_{*,+}:=\{A\in V:{}^t\overline{A}=A\}\quad 和\quad W_{*,-}:=\{A\in V:{}^t\overline{A}=-A\}$$

这里 ${}^t\overline{A}={}^t\overline{(a_{ij})}_{i,j}:=(\overline{a}_{ji})_{i,j}$，其中 $\overline{a}_{ij}=\overline{x+\sqrt{-1}y}:=x-\sqrt{-1}y\in\mathbb{C}$。
证明: 集合 $\{E_{ij}+E_{ji}:i\geq j\}\cup\{\sqrt{-1}(E_{ij}-E_{ji}):i>j\}$ 是作为 $\mathbb{R}$-向量空间的 $W_{*,+}$ 的一组基。这里 $E_{ij}:=(a_{kl})_{k,l}$ 表示第 i 行第 j 列元素为 1，其余元素为 0 的矩阵。

E_{ij} 的例子:

例子 1：E_{11}（以 2 阶矩阵为例）$E_{11}=\begin{pmatrix}1&0\\0&0\end{pmatrix}$，即第 1 行第 1 列元素为 1，其余元素为 0。

例子 2：E_{12}（以 2 阶矩阵为例）$E_{12}=\begin{pmatrix}0&1\\0&0\end{pmatrix}$，即第 1 行第 2 列元素为 1，其余元素为 0。

1.6 替换定理

接下来，我们将利用数学归纳法这一强大的归纳工具来证明线性代数中的一个重要定理——替换定理。该证明过程和归纳思想都相当重要，是后文许多思想所在的源泉。

Theorem

替换定理（Replacement Theorem）： 设 V 是一个由恰好含 n 个向量的集合 G 生成的向量空间，L 是 V 的一个恰好含 m 个向量的线性无关子集。那么 $m \leq n$ ，且存在 G 的一个恰好含 $n-m$ 个向量的子集 H ，使得 $L \cup H$ 生成 V 。

Example

例如：设向量空间 $V = \mathbb{R}^3$（三维实数向量空间），而线性无关子集 $L = \{\mathbf{v}_1, \mathbf{v}_2\}$，其中 $\mathbf{v}_1 = (1, 1, 0)$，$\mathbf{v}_2 = (1, -1, 1)$。我们可以找到 3-2=1 个向量，如 $\mathbf{e}_3 = (0, 0, 1)$，使得这三个向量可以生成（span）V。

该定理还有相当丰富的推论：

Corollary & Secondary Conclusion

推论 1：设 V 是一个具有有限基底的向量空间。那么 V 的每一个基底都包含相同数量的向量。(即上一节的重要定理)

Corollary & Secondary Conclusion

推论 2：设 V 是一个维度为 n 的向量空间。

(a)V 的任何有限生成集至少包含 n 个向量，且恰好包含 n 个向量的 V 的生成集是 V 的一个基底，故线性无关。

(b)V 的任何恰好包含 n 个向量的线性无关子集是 V 的一个基底。

(c)V 的每一个线性无关子集都可以扩展为 V 的一个基底。

1.7 子空间上的运算 1

1.7.1 子空间上的加法和减法

设 $W_1, W_2 \subset V$ 是两个子空间，它们的**和（sum）**定义为 $W_1 + W_2 := \{w_1 + w_2 : w_i \in W_i,\ i = 1, 2\}$。

可以证明：它仍然是一个子空间。

Example

例如：若 $S = \{(1,0)\}$，$W = \{(0,1)\}$，则 $S + W = \{(1,0) + (0,1)\} = \{(1,1)\}$，即 $S + W$ 为点 $(1,1)$；
若 S 为 x 轴（向量形如 $(a,0), a \in \mathbb{R}$），W 为 y 轴（向量形如 $(0,b), b \in \mathbb{R}$），则 $S + W = \{(a,0) + (0,b) \mid a, b \in \mathbb{R}\} = \{(a,b) \mid a, b \in \mathbb{R}\}$，即 $S + W$ 为全平面。

【Remark】：

$W_1, W_2 \subset V$ 其各自的基是 $S_1 \cup S_2$，那么这两组基是否为 $W_1 + W_2$ 的基？若 $S_1 \cap S_2 = \emptyset$ 呢？

子空间的和 $W_1 + W_2$ 的基需满足线性无关且张成 $W_1 + W_2$。

一般情况：$S_1 \cup S_2$ 不是 $W_1 + W_2$ 的基。即使 $S_1 \cap S_2 = \emptyset$：仍不一定是基。

原因：若 W_1, W_2 有公共非零向量，其基 S_1, S_2 中可能包含重复向量（或线性相关的向量），导致 $S_1 \cup S_2$ 线性相关。

Example

例如：设 $V = \mathbb{R}^2$，$W_1 = \text{span}\{(1,0)\}$（基 $S_1 = \{(1,0)\}$），$W_2 = \text{span}\{(2,0)\}$（基 $S_2 = \{(2,0)\}$）。
则 $W_1 + W_2 = W_1$，其基为 $\{(1,0)\}$，但 $S_1 \cup S_2 = \{(1,0), (2,0)\}$ 线性相关，不是基。

设 $W_1, W_2 \subset V$ 是两个子空间，它们的**差集（difference set）**定义为 $W_1 - W_2 := \{w_1 - w_2 : w_i \in W_i, i = 1, 2\}$。

Example

例子 1：不同子空间的差集为全空间设 $V=\mathbb{R}^2$，取子空间 $W_1=\{(a,0)\mid a\in\mathbb{R}\}$（x 轴），$W_2=\{(0,b)\mid b\in\mathbb{R}\}$（y 轴）。则 $W_1-W_2=\{(a-0,0-b)\mid a,b\in\mathbb{R}\}=\{(a,-b)\mid a,b\in\mathbb{R}\}=\mathbb{R}^2$（因为任意 $(x,y)\in\mathbb{R}^2$ 可表示为 $(x,0)-(0,-y)$，其中 $(x,0)\in W_1$，$(0,-y)\in W_2$）。
例子 2：相同子空间的差集为自身设 $V=\mathbb{R}^2$，取子空间 $W_1=W_2=\{(a,a)\mid a\in\mathbb{R}\}$（直线 $y=x$）。则 $W_1-W_2=\{(a-b,a-b)\mid a,b\in\mathbb{R}\}=\{(c,c)\mid c\in\mathbb{R}\}=W_1$（因为任意 $(c,c)\in W_1$ 可表示为 $(c,c)-(0,0)$，其中 $(0,0)\in W_2$）。

1.7.2 子空间上的交与并

两个子空间 $W_1,W_2\subset V$ 的交：

Definition

交 (Intersection)： $W_1\cap W_2:=\{w\in V: w\in W_1 \text{ 且} w\in W_2\}$，**它仍然是一个子空间**。

Property

交的性质：若 W_1 和 W_2 是 V 的子空间，则 $W_1\cap W_2$ 一定是 V 的子空间。

- 例如：在 $\mathbb{R}^3$ 中，设 $W_1=\{(x,y,z)\mid x+y=0\}$，$W_2=\{(x,y,z)\mid x-z=0\}$，则 $W_1\cap W_2=\{(x,-x,x)\mid x\in\mathbb{R}\}$，满足子空间判定条件。

【Remark】：

$S_1\cap S_2$ 是否为 $W_1\cap W_2$ 的基?

一般情况：$S_1\cap S_2$ 不是 $W_1\cap W_2$ 的基。虽这些向量线性无关，但可能“数量不足”，无法张成整个 $W_1\cap W_2$。

Example

设 $V=\mathbb{R}^2$，$W_1=\text{span}\{(1,0),(1,1)\}=\mathbb{R}^2$（基 $S_1=\{(1,0),(1,1)\}$），$W_2=\text{span}\{(1,0),(2,1)\}=\mathbb{R}^2$（基 $S_2=\{(1,0),(2,1)\}$）。则 $W_1\cap W_2=\mathbb{R}^2$，其基需 2 个线性无关向量，但 $S_1\cap S_2=\{(1,0)\}$，无法张成 $\mathbb{R}^2$。

两个子空间 $W_1,W_2\subset V$ 的并：

Definition

并 (Union)： $W_1\cup W_2:=\{w\in V: w\in W_1 \text{ 或} w\in W_2\}$。

【Remark】:

Union 一般来说并不是一个子空间。

Example

例如：在 $\mathbb{R}^2$ 中，W_1 为 x 轴上的向量集，W_2 为 y 轴上的向量集。$(1,0) \in W_1$，$(0,1) \in W_2$，但 $(1,0)+(0,1)=(1,1) \notin W_1 \cup W_2$，不满足加法封闭性，故 $W_1 \cup W_2$ 不是子空间。

Claim

$W_1 \cup W_2$ 是子空间当且仅当 $W_1 \subseteq W_2$ 或 $W_2 \subseteq W_1$。

证明. 充分性容易证明；

下面证明**必要性**：

假设 $W_1 \cup W_2$ 是子空间，但 $W_1 \nsubseteq W_2$ 且 $W_2 \nsubseteq W_1$

此时：因为 $W_1 \nsubseteq W_2$，所以存在 $w_1 \in W_1 \setminus W_2$（$w_1$ 在 W_1 中，但不在 W_2 中）；

因为 $W_2 \nsubseteq W_1$，所以存在 $w_2 \in W_2 \setminus W_1$（$w_2$ 在 W_2 中，但不在 W_1 中）。

考虑这两个向量的和 $w_1 + w_2$：由于 $W_1 \cup W_2$ 是子空间，它对加法封闭，因此 $w_1 + w_2$ 必须属于 $W_1 \cup W_2$，即要么 $w_1 + w_2 \in W_1$，要么 $w_1 + w_2 \in W_2$。

若 $w_1 + w_2 \in W_1$：

因为 W_1 是子空间，对减法封闭（子空间对数乘和加法封闭，自然对减法封闭：$a - b = a + (-1)b$），所以 $w_2 = (w_1 + w_2) - w_1 \in W_1$。但这与 $w_2 \in W_2 \setminus W_1$（$w_2$ 不在 W_1 中）矛盾。

若 $w_1 + w_2 \in W_2$：

同理，W_2 对减法封闭，所以 $w_1 = (w_1 + w_2) - w_2 \in W_2$。但这与 $w_1 \in W_1 \setminus W_2$（$w_1$ 不在 W_2 中）矛盾。 □

接下来我们来介绍一种特殊的和：

1.7.3 子空间上的直和

Definition

直和/直接和(direct sum)：若 W_1 和 W_2 是向量空间 V 的子空间，且满足 $W_1 \cap W_2 = \{0\}$ 以及 $W_1 + W_2 = V$，则称向量空间 V 是 W_1 和 W_2 的直和，记为 $V = W_1 \oplus W_2$。

接下来我们将介绍多个子空间的直和:

Definition

设 $V_1, V_2, \cdots, V_m$ 是线性空间 V 的子空间并且其和是全空间，且对任意 $2 \le i \le m$ ，$V_i \cap (V_1 + V_2 + \cdots + V_{i-1}) = \{0\}$ ，那么：$V_0 = V_1 \oplus V_2 \oplus \cdots \oplus V_m$ 是直和

【Remark】:

注意到：若 $V = \sum_{i=1}^{k} W_i$ 且对于 $i \neq j$ 有 $W_i \cap W_j = \{0\}$，但是不一定有 $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$。

我们可以找出反例：例如，取 $W_1 = \text{span}\{(1,0)\}$，$W_2 = \text{span}\{(0,1)\}$，以及 $W_3 = \text{span}\{(1,1)\}$，说明不满足条件时不能得到直和结论。

在介绍本节的主定理以前，我们将先介绍 2 个重要的引理：

Lemma

lemma1: 若 S_i 是 W_i 的基（$i = 1, 2$），则 $S_1 \cup S_2$ 就是 $W_1 \oplus W_2$ 的基；
lemma2: 若 $S_1 \cup S_2$ 是 $W_1 \oplus W_2$ 的基，且 S_1 是 W_1 的基，那么 S_2 就是 W_2 的基；反之亦然。

证明. 对于引理 1：

张成性：

对任意 $v \in W_1 \oplus W_2$，由直和定义，$v = w_1 + w_2$（其中 $w_1 \in W_1$，$w_2 \in W_2$）。

因 S_1 是 W_1 的基，故 w_1 可由 S_1 线性表出；同理，w_2 可由 S_2 线性表出。

因此，$v = w_1 + w_2$ 可由 $S_1 \cup S_2$ 线性表出，即 $S_1 \cup S_2$ 张成 $W_1 \oplus W_2$。

线性无关性：

假设 $\sum_{s \in S_1} a_s s + \sum_{s \in S_2} b_s s = 0$（$a_s, b_s$ 为系数）。令 $w_1 = \sum_{s \in S_1} a_s s \in W_1$，$w_2 = \sum_{s \in S_2} b_s s \in W_2$，则 $w_1 + w_2 = 0$。

将等式变形为：$w_1 = -w_2$。

此时：$w_1 \in W_1$（由 w_1 的定义）；$-w_2 \in W_2$（因为 W_2 是子空间，对“数乘”封闭，$w_2 \in W_2$ 则 $-w_2 \in W_2$）。

因此，w_1 同时属于 W_1 和 W_2，即 $w_1 \in W_1 \cap W_2$。

由直和的“交为零空间”（$W_1 \cap W_2 = \{0\}$），得 $w_1 = 0$ 且 $w_2 = 0$。

又因 S_1 是 W_1 的基（线性无关），故 $w_1 = 0$ 蕴含所有 $a_s = 0$；同理，S_2 线性无关，故 $w_2 = 0$ 蕴含所有 $b_s = 0$。

因此，$S_1 \cup S_2$ 线性无关。

对于引理 2：需验证 S_2 满足线性无关且张成 W_2。

张成性:

对任意 $w_2 \in W_2$, 因 $S_1 \cup S_2$ 张成 $W_1 \oplus W_2$, 故 w_2 可表为: $w_2 = \sum_{s\in S_1} a_s s + \sum_{s \in S_2} b_s s$

令 $w_1 = \sum_{s\in S_1} a_s s \in W_1$，则 $w_2 = w_1 + \sum_{s\in S_2} b_s s$。

由直和的"交为零空间"($W_1 \cap W_2 = \{0\}$)，若 $w_1 \neq 0$，则 $w_1 = w_2 - \sum_{s\in S_2} b_s s \in W_1 \cap W_2$ 且非零 (右边: $w_2 \in W_2$，且 $\sum_{s\in S_2} b_s s \in W_2$ (因为 S_2 张成 W_2，子空间对线性组合封闭)。因此，$w_2 - \sum_{s\in S_2} b_s s \in W_2$ (子空间对减法封闭)。)，与 $W_1 \cap W_2 = \{0\}$ 矛盾。

因此 $w_1 = 0$，即: $w_2 = \sum_{s\in S_2} b_s s$ 故 S_2 张成 W_2。

线性无关性:

假设 $\sum_{s\in S_2} b_s s = 0$。

因 S_1 是 W_1 的基(线性无关), 且 $S_1 \cup S_2$ 线性无关, 故组合 $\sum_{s\in S_1} 0 \cdot s + \sum_{s\in S_2} b_s s = 0$ 蕴含所有 $b_s = 0$。因此，S_2 线性无关。 □

接下来我们将讲解本节的主定理:

Theorem

设 $V_1, V_2, \cdots, V_m$ 是线性空间 V 的子空间，$V_0 = V_1 + V_2 + \cdots + V_m$ ，则以下命题**等价**:

1. $V_0 = V_1 \oplus V_2 \oplus \cdots \oplus V_m$ 是直和;
2. 对任意 $2 \leq i \leq m$ ，$V_i \cap (V_1 + V_2 + \cdots + V_{i-1}) = \{0\}$;
3. $\dim(V_1 + V_2 + \cdots + V_m) = \dim V_1 + \dim V_2 + \cdots + \dim V_m$;
4. $V_1, V_2, \cdots, V_m$ 的一组基可拼成 V_0 的一组基;
5. V_0 中向量表示为 $V_1, V_2, \cdots, V_m$ 中向量之和时表示唯一，即若 $\alpha \in V_0$ 且 $\alpha = v_1 + v_2 + \cdots + v_m = u_1 + u_2 + \cdots + u_m$ ，其中 $v_i, u_i \in V_i$ ，则 $u_i = v_i$ ($i = 1, 2, \cdots, m$)。

证明. □

1.8 等价关系、等价类与商集

1.8.1 等价关系 (Equivalence Relation)

Definition

设 A 是非空集合，R 是 A 上的二元关系。若 R 满足以下三条性质，则称 R 为 A 上的**等价关系**（equivalence relation）：

1. **自反性** (reflexivity)：对任意 $a \in A$，有 $(a,a) \in R$；（或 $a\ a$）
2. **对称性** (symmetry)：对任意 $a,b \in A$，若 $(a,b) \in R$，则 $(b,a) \in R$；
3. **传递性** (transitivity)：对任意 $a,b,c \in A$，若 $(a,b) \in R$ 且 $(b,c) \in R$，则 $(a,c) \in R$。

若 $(a,b) \in R$，则称 a 与 b 等价，记为 $a \sim b$。

Example

我们有以下几种常见的等价关系：三角形集合上的“相似（is similar to）”关系；整数集 $\mathbb{Z}$ 上的“模 n 同余（is congruent to modulo n）”关系; 函数 $f : X \to Y$ 下有相同的像的关系

相似三角形的定义是“对应角相等，对应边成比例”。
因此：

- 任意三角形与自身相似（角相等、边比例为 $1:1$），故 $a \sim a$。
- 对称性：若三角形 A 相似于 B，则 B 的角与 A 相等、边与 A 成比例，因此 B 也相似于 A，故 $a \sim b \iff b \sim a$。
- 若 A 相似于 B，B 相似于 C，则 A 与 C 的对应角都等于 B 的角（相等），对应边比例为“A 对 B 的比例”乘以“B 对 C 的比例”（仍为常数），因此 A 相似于 C，故 $a \sim b \wedge b \sim c \implies a \sim c$。

若整数 a,b 满足“n 整除 $a-b$”，则称 $a \equiv b \pmod n$。

在整数集 $\mathbb{Z}$ 上，定义关系 $R = \{(a,b) \mid a \equiv b \pmod 2\}$（即 a 与 b 奇偶性相同）。验证 R 是等价关系：

1. 自反性：$\forall a \in \mathbb{Z}$，$a - a = 0$ 是 2 的倍数，故 $a \sim a$；
2. 对称性：若 $a \sim b$，则 $a-b$ 是 2 的倍数，显然 $b-a$ 也是 2 的倍数，故 $b \sim a$；即：若 $a \equiv b \pmod n$，则 $n \mid (a-b)$，即存在整数 k 使得 $a-b=nk$。因此 $b-a=-nk=n(-k)$，故 $n \mid (b-a)$，即 $b \equiv a \pmod n$。

3. 传递性：若 $a \sim b$ 且 $b \sim c$，则 $a-b=2k$，$b-c=2m\,(k,m\in\mathbb{Z})$，故 $a-c=2(k+m)$，即 $a \sim c$。

因此，R 是 $\mathbb{Z}$ 上的等价关系。

至于函数：

一般情况：对 $x_1,x_2\in X$，定义 $x_1 \sim x_2 \iff f(x_1)=f(x_2)$。

- 自反性：对任意 $x\in X$，$f(x)=f(x)$，故 $x\sim x$。
- 对称性：若 $x_1\sim x_2$（即 $f(x_1)=f(x_2)$），则 $f(x_2)=f(x_1)$，故 $x_2\sim x_1$。
- 传递性：若 $x_1\sim x_2\,(f(x_1)=f(x_2))$ 且 $x_2\sim x_3\,(f(x_2)=f(x_3))$，则 $f(x_1)=f(x_3)$，故 $x_1\sim x_3$。

我们在这里还有 1 个例子：

Example

集合 $X=\{a,b,c\}$ 上的关系 $\sim$（由序对 $R=\{(a,a),(b,b),(c,c),(a,b),(b,a)\}$ 定义）——该关系 $\sim$ 的本质是：a 与 a、b 与 b、c 与 c 相关；a 与 b、b 与 a 相关；c 仅与自身相关。

其中：序对 $(x,y)\in R$ 就表示 “x 和 y 满足关系 $\sim$”（即 $x\sim y$）

容易验证其自反与对称性，只需验证传递性即可。

问题：找出那些因恰好违背了定义中的三个条件（自反性、对称性、传递性）之一而不是等价关系的二元关系的例子。

- 违反自反性：设集合 $X=\{a,b\}$，定义关系 $R=\{(a,a),(a,b),(b,a)\}$。此时 $b\not\sim b$，不满足自反性，而对称性、传递性成立。
- 违反对称性：设集合 $X=\{1,2\}$，定义关系 $R=\{(1,1),(2,2),(1,2)\}$。此时 $1\sim 2$ 但 $2\not\sim 1$，不满足对称性，而自反性、传递性成立。
- 违反传递性：设集合 $X=\{1,2,3\}$，定义关系 $R=\{(1,1),(2,2),(3,3),(1,2),(2,1),(2,3),(3,2)\}$。此时 $1\sim 2$ 且 $2\sim 3$，但 $1\not\sim 3$，不满足传递性，而自反性、对称性成立。

1.8.2 等价类

等价类是基于等价关系对集合元素的分类，其定义依赖于等价关系：

Definition

设 R 是集合 A 上的等价关系，对任意 $a \in A$，称集合

$$[a]_R = \{x \in A \mid x \sim a\}$$

为 a 关于 R 的**等价类**（equivalence class of a under ）或称为**陪集**（Coset），简记为 $[a]$, 或：$\bar{a}$。其中 a 称为该等价类的**代表元**。

用数学语言来说：

若定义 $v_1 \sim v_2 \iff v_1 - v_2 \in W$。

那么陪集（即等价类）记为：$\bar{v} = [v] = \{v + w \mid w \in W\}$（即 "$v$ 所在的陪集"，可理解为 "v 被 W 平移后的集合"）。

【Remark】：

每个陪集的 "内部容量"（元素个数）可能无限

Example

对上例中的等价关系 R（模 2 同余）：
- 取 $a = 0$，则 $[0] = \{x \in \mathbb{Z} \mid x \equiv 0 \pmod 2\} = \{\dots, -4, -2, 0, 2, 4, \dots\}$（所有偶数）；
- 取 $a = 1$，则 $[1] = \{x \in \mathbb{Z} \mid x \equiv 1 \pmod 2\} = \{\dots, -3, -1, 1, 3, 5, \dots\}$（所有奇数）。

可见，等价类是按 "等价关系" 对集合元素的**归类**。

等价类具有以下性质：

1. 对任意 $a \in A$，$a \in [a]$（非空性）；
2. 若 $a \sim b$，则 $[a] = [b]$（代表元等价则类相等）；
3. 若 $a \nsim b$，则 $[a] \cap [b] = \emptyset$（不同等价类互斥）。

1.8.3 划分的定义

集合的**划分**（**Partition**）是将一个集合分解为若干非空且互不相交的子集的集合，使得原集合中的每个元素恰好属于其中一个子集。更形式化地:

Definition

对于集合 A，其划分 $\mathcal{P}=\{A_i\}_{i\in I}$ 需满足以下条件：

- 对所有 $i\in I$，子集 A_i 是 A 的非空子集，即 $A_i\subset A$ 且 $A_i\neq\emptyset$；
- 所有子集的并集等于原集合 A，即 $\bigcup_{i\in I}A_i=A$；
- 任意两个不同的子集互不相交，即对所有 $i\neq j$，有 $A_i\cap A_j=\emptyset$。

划分中的每个子集 A_i 称为该划分的块（block）或部分（part）。

- 假设原集合 $A=\{1,2,3,4\}$。
- 例子 1：$\mathcal{P}=\{\{1,2\},\{3,4\}\}$ 是一个划分。
  - 验证：每个子集非空且是 A 的子集；并集为 $\{1,2,3,4\}=A$；两个子集交集为空。
- 例子 2：$\mathcal{P}=\{\{1\},\{2\},\{3,4\}\}$ 也是一个划分。
- 反例 1：$\mathcal{P}=\{\{1,2\},\{2,3\}\}$ 不是划分（因为元素 2 同时属于两个子集，违反“恰好一个”）。
- 反例 2：$\mathcal{P}=\{\{1,2\},\emptyset,\{3,4\}\}$ 不是划分（因为包含空集 $\emptyset$，违反“非空”要求）。

1.8.4 商集

商集是所有等价类构成的集合，是对原集合的“整体划分”：

Definition

设 R 是集合 A 上的等价关系，称所有等价类组成的集合

$$A/R=\{[a]\mid a\in A\}$$

为 A 关于 R 的**商集**。

Example

对整数域中的等价关系 R（模 2 同余），商集为：

$$\mathbb{Z}/R = \{[0], [1]\}$$

即整数集被划分为“偶数类”和“奇数类”两个部分。

再如，设 $A = \{1, 2, 3, 4\}$，R 是“模 3 同余”关系，则：
- $[1] = \{1, 4\}$，$[2] = \{2\}$，$[3] = \{3\}$，
- 商集 $A/R = \{\{1, 4\}, \{2\}, \{3\}\}$。

重要应用：有理数的构造

整数组成的分数 $\frac{m}{n}$（$n \neq 0$）有“不同形式但值相同”的情况（比如 $\frac{1}{2}$ 和 $\frac{2}{4}$ 应该表示同一个有理数）。

等价关系：定义 $\frac{a}{b} \sim \frac{c}{d} \iff ad = bc$（交叉相乘相等）。这个关系把“数值相等的分数”归为同一等价类。商集构造：有理数 $\mathbb{Q}$ 是“所有分数”模这个等价关系的商集，即 $\mathbb{Q} := \left\{ \frac{m}{n} \mid m \in \mathbb{Z}, n \in \mathbb{Z} \setminus \{0\} \right\} / \sim$。

核心思想：用等价关系“忽略分数的不同表示形式，只关注数值是否相等”，从而严格构造出有理数域。

重要应用：p 有限域的构造

Definition

$\mathbb{F}_p$（也记为 $\mathbb{Z}/p\mathbb{Z}$，其中 p 为素数）是模 p 的剩余类集合，元素为 $\{0, 1, 2, \ldots, p-1\}$。在其上定义模 p 加法和模 p 乘法，满足域的所有公理，因此是一个有限域（也称为素域，因为它无法由更小的域扩张得到）

例子:

Example

$\mathbb{F}_2$：元素为 $\{0, 1\}$
加法：$0 + 0 = 0$，$0 + 1 = 1$，$1 + 0 = 1$，$1 + 1 = 0$；
乘法：$0 \times 0 = 0$，$0 \times 1 = 0$，$1 \times 0 = 0$，$1 \times 1 = 1$。
$\mathbb{F}_3$：元素为 $\{0, 1, 2\}$
加法：$1 + 2 \equiv 0 \mod 3$，$2 + 2 \equiv 1 \mod 3$ 等；
乘法：$2 \times 2 \equiv 1 \mod 3$（即 2 的乘法逆元是自身）

代数结构性质:

Property

元素个数：恰好有 p 个元素，是素数幂阶有限域（幂次为 1）。

由此等价关系、等价类与商集的逻辑递进关系如下：

1. **基础**：等价关系是集合上满足自反、对称、传递性的二元关系，为元素分类提供“规则”；

2. **分类**：基于等价关系，将集合中相互等价的元素归为一类，形成等价类；

3. **整体划分**：所有等价类的集合构成商集，它是原集合的一个“划分”（各等价类非空、互斥且并集为原集合）。

三者的核心联系：等价关系 $\iff$ 集合的划分（商集本质是划分），而等价类是划分的基本单元。

1.8.5 等价关系的两条重要定理

Theorem

等价关系基本定理 (Fundamental Theorem of Equivalence Relations)：**集合上的等价关系**与该集合的**划分**之间存在自然的**一一对应**关系。

proof.

- **从等价关系到划分：**假设集合 X 上有等价关系 $\sim$。我们要证明所有等价类构成的集合 $\mathcal{P} = X/\sim$ 是 X 的划分。

  1. **等价类是 X 的非空子集：**对任意 $x \in X$，等价类 $[x] = \{y \in X \mid y \sim x\}$。由自反性，$x \sim x$，所以 $x \in [x]$，故 $[x] \neq \emptyset$；且等价类中元素都属于 X，故 $[x] \subset X$。

  2. **所有等价类的并集是 X：**对任意 $x \in X$，$x \in [x]$，而 $[x]$ 是 $\mathcal{P}$ 中的元素，因此 $X = \bigcup_{x \in X} [x]$。

  3. **不同等价类互不相交：**若等价类 $[a]$ 和 $[b]$ 有公共元素 c（即 $c \in [a] \cap [b]$），则 $a \sim c$ 且 $b \sim c$。由对称性，$c \sim b$，再由传递性，$a \sim b$。进而可证 $[a] \subset [b]$ 且 $[b] \subset [a]$，故 $[a] = [b]$。

- **从划分到等价关系：**假设 $\mathcal{P} = \{A_i\}_{i \in I}$ 是 X 的划分。定义二元关系 $\sim$：$a \sim b \iff a, b$ 属于同一个块 A_i。验证其为等价关系：

  1. **自反性：**对任意 $x \in X$，x 属于某个 A_i，故 $x \sim x$。

  2. **对称性：**若 $x \sim y$，则 x, y 属于 A_i，显然 y, x 也属于 A_i，故 $y \sim x$。

  3. **传递性：**若 $x \sim y$ 且 $y \sim z$，则 x, y 属于 A_i，y, z 属于 A_j。因 y 只能属于一个块，故 $A_i = A_j$，所以 x, z 属于 A_i，即 $x \sim z$。

Theorem

投影的泛性质定理（(Universal Property of the Projection)） 设 $f: X \to Y$ 是一个函数，且 $\sim$ 是 X 上的一个等价关系。
如果 f 在等价类上是常值的（即若 $a \sim b$ 蕴含 $f(a) = f(b)$），那么存在唯一的函数 $g: X/{\sim} \to Y$ 使得 $f = g \circ p$，其中 $p: X \to X/{\sim}$ 是投影映射。
且如果 f 是满射的，且等价关系由 $a \sim b \iff f(a) = f(b)$ 定义，那么诱导映射 g 是一个双射。

这通常用下述交换图来直观表示：

$$X[r, "f"][d, "p"']YX/{\sim}[ru, dashed, "g"'][ru, "\exists!"]$$

这个性质是基本的；它意味着与等价关系相容的函数能唯一地通过商映射分解。

proof.

- 证明存在唯一的 $g: X/\sim \to Y$ 使得 $f = g \circ p$

  1. 定义良定义的函数 g：

     因为 f 在每个等价类上是常值的（即若 $a \sim b$，则 $f(a) = f(b)$），对于商集 $X/\sim$ 中的任意等价类 $[x]$，它包含所有与 x 等价的元素 x'（满足 $x' \sim x$）。由于 f 在等价类上常值，所以这些 x' 的函数值 $f(x')$ 都等于 $f(x)$。

     因此，我们可以定义对于其中一个代表元 x，定义 $g([x]) = f(x)$，这个定义不依赖于从等价类 $[x]$ 中选择哪个具体的元素，所以 g 是从商集 $X/\sim$ 到 Y 的“良定义”函数。

  2. 证明 g 的唯一性：

     假设存在另一个函数 $h: X/\sim \to Y$，也满足 $f = h \circ p$。需证：$h = g$。

     对于商集 $X/\sim$ 中的任意等价类 $[x]$，取 $x \in X$（此时 $p(x) = [x]$）。由 $f = h \circ p$ 可得，$f(x) = (h \circ p)(x) = h(p(x)) = h([x])$。而根据 g 的定义，$g([x]) = f(x)$，所以 $h([x]) = g([x])$ 对所有 $[x] \in X/\sim$ 都成立，因此 $h = g$，即 g 是唯一的。

- 证明 f 是满射时，g 是双射

  双射需要同时满足单射（Injective）和满射（Surjective）。

  1. 证明 g 是单射：

     单射的定义是：若 $g([x]) = g([y])$，则 $[x] = [y]$。

由 g 的定义可知，$g([x])=f(x)$，$g([y])=f(y)$，所以如果 $g([x])=g([y])$，那么 $f(x)=f(y)$。

又因为等价关系 $\sim$ 定义为 $a\sim b \iff f(a)=f(b)$，所以 $f(x)=f(y)$ 意味着 $x\sim y$。而等价类相等的充要条件是其代表元等价，即 $x\sim y \implies [x]=[y]$，所以 g 是单射。

2. 证明 g 是满射

满射的定义是：对于任意 $y\in Y$，存在 $[x]\in X/\sim$ 使得 $g([x])=y$。

因为 f 是满射，所以对于任意 $y\in Y$，都存在 $x\in X$ 使得 $f(x)=y$。取这个 x 所在的等价类 $[x]\in X/\sim$，根据 g 的定义，$g([x])=f(x)=y$。因此，对于任意 $y\in Y$，都能找到对应的 $[x]\in X/\sim$，所以 g 也是满射。

1.9 子空间上的运算 2

1.9.1 子空间上的商与商空间

Definition

商空间 (Quotient space)： 所有陪集构成的集合 $V/W:=\{\bar{v}\mid v\in V\}$，并在其上定义线性运算：

$\bar{v}_1+\bar{v}_2=\overline{v_1+v_2}, \quad c\bar{v}=\overline{cv} \quad (\forall c\in\mathbb{F})$

可验证：这些运算良定义且使 V/W 成为域 $\mathbb{F}$ 上的向量空间，称为 V 对 W 的商空间。

用更加可以理解的语言说：

Fact or Background

(等价定义 1)：若 V 是数域 $\mathbb{F}$ 上的向量空间，W 是 V 的子空间，定义等价关系：$u\sim v \iff u-v\in W$。所有等价类的集合称为 V 关于 W 的**商空间**，记为 V/W。
(等价定义 2)：设 V 是数域 $\mathbb{F}$ 上的向量空间，W 是 V 的子空间，对任意 $u\in V$，其等价类定义为 $\bar{u}=u+W=\{u+w\mid w\in W\}$。而商空间 V/W 的元素就是等价类。

【Remark】:

商集： 仅为集合，无额外代数结构（如加法、数乘），仅由等价类的“归属”关系构成。
商空间： 是向量空间，继承了原空间的线性结构。对等价类定义加法和数乘：

在向量空间中，0 元是加法的单位元，即满足 $\bar{0}+\bar{u}=\bar{u}$ 的等价类。

根据等价类的加法定义：$\overline{0}+\overline{u}=\overline{0+u}=\overline{u}$ 而 $\overline{0}=0+W=W$（因为 $0\in V$，其等价类是所有 $0+w=w$，即子空间 W 本身）。

所以我们得到了：

Property

商空间 V/W 的零元就是子空间 W 本身。

对于商空间的基，我们有以下的方法：

Method

基的构造步骤第一步：取 W 的一组基 $\{w_1,w_2,\ldots,w_k\}$（其中 $k=\dim W$）。
第二步：将其扩充为 V 的一组基：$\{w_1,\ldots,w_k,v_{k+1},\ldots,v_n\}$（其中 $n=\dim V$）。
第三步：考虑这些基向量在商空间中的等价类 $\{\overline{v}_{k+1},\overline{v}_{k+2},\ldots,\overline{v}_n\}$，这组向量就是 V/W 的一组基。

证明. **线性无关：** 假设存在系数 $a_{k+1},\ldots,a_n\in\mathbb{F}$，使得 $a_{k+1}\overline{v}_{k+1}+\cdots+a_n\overline{v}_n=\overline{0}$ 根据等价类的数乘定义，这等价于 $a_{k+1}v_{k+1}+\cdots+a_nv_n\in W$ 但 W 的基是 $\{w_1,\ldots,w_k\}$，故可设 $a_{k+1}v_{k+1}+\cdots+a_nv_n=b_1w_1+\cdots+b_kw_k$。

由于 $\{w_1,\ldots,w_k,v_{k+1},\ldots,v_n\}$ 是 V 的基（线性无关），因此所有系数必须为 0，即 $a_{k+1}=\cdots=a_n=0$，$b_1=\cdots=b_k=0$。

故 $\{\overline{v}_{k+1},\ldots,\overline{v}_n\}$ 线性无关。

张成性： 对任意 $\overline{u}\in V/W$，$u\in V$ 可表示为 $u=b_1w_1+\cdots+b_kw_k+a_{k+1}v_{k+1}+\cdots+a_nv_n$ 则其等价类为 $\overline{u}=\overline{b_1w_1+\cdots+b_kw_k+a_{k+1}v_{k+1}+\cdots+a_nv_n}=\overline{b_1w_1}+\cdots+\overline{b_kw_k}+a_{k+1}\overline{v}_{k+1}+\cdots+a_n\overline{v}_n$ 由于 $b_iw_i\in W$，故 $\overline{b_iw_i}=\overline{0}$，因此 $\overline{u}=a_{k+1}\overline{v}_{k+1}+\cdots+a_n\overline{v}_n$ 即 $\{\overline{v}_{k+1},\ldots,\overline{v}_n\}$ 张成 V/W。 □

【Remark】:

基的核心要求是两个：线性无关 + 张成整个空间。
这两个要求只和“基元素（陪集）的个数”以及“它们的运算性质”有关，和“每个基元素（陪集）内部的元素个数”无关。
满足这两个条件，就是基——哪怕基里只有 1 个陪集，哪怕每个陪集里有无限个点（内部元素无限）。

1.9.2 子空间上的直积

Definition

两个 $\mathbb{F}$ - 向量空间 W_1 和 W_2 的**直积** (product) 定义为:
$W_1 \times W_2 := \{(w_1, w_2) : w_i \in W_i,\ i = 1, 2\}$。
在通常意义下，它是一个 $\mathbb{F}$-向量空间。

Example

欧氏空间的直积设 $\mathbb{F} = \mathbb{R}$，$W_1 = \mathbb{R}^2$（二维实向量空间，元素形如 (a, b)），$W_2 = \mathbb{R}^3$（三维实向量空间，元素形如 (c, d, e)）。
则它们的直积为：$W_1 \times W_2 = \{\big((a, b), (c, d, e)\big) \mid a, b, c, d, e \in \mathbb{R}\}$
其向量空间运算为分量：
加法：$\big((a, b), (c, d, e)\big) + \big((a', b'), (c', d', e')\big) = \big((a + a', b + b'), (c + c', d + d', e + e')\big)$
数乘：对 $\lambda \in \mathbb{R}$，$\lambda \cdot \big((a, b), (c, d, e)\big) = \big((\lambda a, \lambda b), (\lambda c, \lambda d, \lambda e)\big)$

1.9.3 双线性与张量积简述

为了照顾到子空间运算的完整性，我们不加前置条件的先提出：

Definition

双线性映射（bilinear mapping）：设 V, W, U 是域 $\mathbb{F}$ 上的向量空间，映射 $f : V \times W \to U$ 若满足：

- 对第一个变量线性：$\forall v_1, v_2 \in V, w \in W, a \in \mathbb{F}$，有 $f(v_1 + v_2, w) = f(v_1, w) + f(v_2, w)$ 且 $f(av, w) = af(v, w)$；
- 对第二个变量线性：$\forall v \in V, w_1, w_2 \in W, a \in \mathbb{F}$，有 $f(v, w_1 + w_2) = f(v, w_1) + f(v, w_2)$ 且 $f(v, aw) = af(v, w)$，

则称 f 为**双线性映射**。

之后在内积空间的章节中，我们会对此展开详细讨论。

Example

实数乘法: 取 $V = \mathbb{R}$，$W = \mathbb{R}$，$U = \mathbb{R}$（均为实向量空间），定义映射 $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ 为 $f(a, b) = a \times b$（普通实数乘法）。

验证双线性：

- 对第一个变量：$f(a_1 + a_2, b) = (a_1 + a_2)b = a_1b + a_2b = f(a_1, b) + f(a_2, b)$，且 $f(ka, b) = (ka)b = k(ab) = kf(a, b)$；

- 对第二个变量：$f(a, b_1+b_2) = a(b_1+b_2) = ab_1+ab_2 = f(a, b_1)+f(a, b_2)$，且 $f(a, kb) = a(kb) = k(ab) = kf(a, b)$。

故实数乘法是双线性映射。

但是双线性映射不一定是线性映射, 例如: 若 $f(a, b) = a\times b$; $f(3(a, b)) = f(3a, 3b) = 3a\times 3b \neq 3f(a, b) = 9ab$

为什么需要张量积：线性代数的核心工具（如矩阵、线性变换、特征值等）直接适用于**线性映射**（定义域为单个向量空间，对“整个输入”线性）。

双线性映射的定义域是直积空间 $V\times W$（其本身是向量空间，是有序对 (v, w) 的集合），但双线性映射不满足 $V\times W\to U$ 的线性映射定义（线性映射要求对‘有序对整体’的数乘齐次性，而双线性映射仅对单个变量满足线性），因此无法直接用线性代数工具（如基、矩阵）研究。

为了用线性代数工具研究双线性关联，需构造新向量空间 (张量积) $V\otimes_{\mathbb{F}} W$，使得所有 $V\times W$ 上的双线性映射，都能唯一对应到 $V\otimes_{\mathbb{F}} W$ 上的线性映射（即把“双线性问题”转化为“线性问题”）。

【Remark】:

张量积空间 $V\otimes_{\mathbb{F}} W$ 不是线性映射的空间（同态空间），它是一个独立的向量空间。它的元素是“张量”。

由此，给出**张量积的构造和定义**:

Method

设 V, W 是域 $\mathbb{F}$ 上的向量空间，构造步骤如下:

1. 定义基础集合 $S := \{(v, w) \mid v\in V, w\in W\}$;
2. 构造 $\mathbb{F}$ 上由 S 张成的向量空间 $\mathrm{Span}_{\mathbb{F}}(S)$（元素为 S 中元素的有限线性组合，形如 $\sum_{i=1}^{n} a_i(v_i, w_i)$，$a_i\in\mathbb{F}, v_i\in V, w_i\in W$）;
3. 定义 $\mathrm{Span}_{\mathbb{F}}(S)$ 的子集 S_0，包含四类元素以刻画双线性规则（如加法、数乘的线性性）:

$$\begin{aligned} S_0 := &\{(v_1+v_2, w)-(v_1, w)-(v_2, w) \mid v_1, v_2\in V, w\in W\} \\ &\cup\{(v, w_1+w_2)-(v, w_1)-(v, w_2) \mid v\in V, w_1, w_2\in W\} \\ &\cup\{a(v, w)-(v, aw) \mid v\in V, w\in W, a\in\mathbb{F}\} \\ &\cup\{a(v, w)-(av, w) \mid v\in V, w\in W, a\in\mathbb{F}\}\subset \mathrm{Span}_{\mathbb{F}}(S), \end{aligned}$$

于是，给出张量积（张量积空间）的定义:

Definition

定义：**张量积空间为商空间** $\text{Span}_{\mathbb{F}}(S)/\text{Span}_{\mathbb{F}}(S_0)$，即 $V \otimes_{\mathbb{F}} W := \text{Span}_{\mathbb{F}}(S)/\text{Span}_{\mathbb{F}}(S_0)$；

Definition

秩 1 张量： 对 $v \in V, w \in W$，记 $v \otimes w$ 为张量积空间 $V \otimes W$ 中的元素（称为‘秩 1 张量’）。

张量（tensor）： 是张量积空间 $U \otimes V$ 中的**任意**元素（形如 $\sum_i a_i u_i \otimes v_i$，$a_i \in \mathbb{F}, u_i \in U, v_i \in V$）

同时 $v \otimes w$ 也为 (v, w) 在商空间中的等价类 $\overline{(v,w)}$。

张量积空间的基底：

当 V, W **有限维时**，若 $\{e_i\}$ 是 V 的基、$\{f_j\}$ 是 W 的基，则 $\{e_i \otimes f_j\}$ 是 $V \otimes W$ 的基。

【Remark】：

$\{e_i \otimes f_j\}$ 不是一个数值（如 1、2、3），它是张量积空间的基向量，是线性空间中用于构建其他元素的“基本单元”，类似于 $\mathbb{R}^n$ 中标准基 $\mathbf{e}_i$（如 $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$）的角色，本身没有数值，仅作为线性空间的基元素存在。

注意到：在构造向量空间的时候，我们并不用“直接等式 $=0$”，即 $(v_1+v_2, w)-(v_1, w)-(v_2, w)=0$，而是通过“将差放入 S_0 + 商空间”的方式，是为了先建立无约束的基础空间 $\text{Span}_{\mathbb{F}}(S)$，再通过商空间严格施加双线性约束，确保张量积空间的构造在逻辑上是严谨且自洽的。

如果直接写 $(v_1+v_2, w)-(v_1, w)-(v_2, w)=0$，相当于“强行规定等式成立”，但向量空间的构造需要逻辑上的严谨性：我们需要先承认 $\text{Span}_{\mathbb{F}}(S)$ 中这些元素的**独立性**，再通过“商空间”的方式严格推导出“满足双线性规则的空间”。

下面将给出这个定义的完整证明：

证明. (v_1, w) 和 (v_2, w) 是 S 中两个不同的生成元（因为 $v_1 \neq v_2$），在“自由张成”下，它们是独立的（即无法用对方的线性组合表示）。

(v_1+v_2, w) 是 S 中另一个生成元（因为 v_1+v_2 是 V 中元素，w 是 W 中元素，所以 $(v_1+v_2, w) \in S$）。

在“自由张成”的规则下，(v_1+v_2, w) 并不天然等于 $(v_1, w)+(v_2, w)$——它只是一个独立的基元，和 (v_1, w)、(v_2, w) 之间没有预设的线性关系。

将 $(v_1+v_2,w)-(v_1,w)-(v_2,w)$ 放入 S_0，再构造 $\mathrm{Span}_{\mathbb{F}}(S_0)$（$S_0$ 中元素的线性组合），最后通过商空间来“消去”这些元素：在商空间中，若两个元素的差属于 $\mathrm{Span}_{\mathbb{F}}(S_0)$，则它们被视为“等价”（即“差为 0”）。

因此：(v_1+v_2,w) 与 $(v_1,w)+(v_2,w)$ 的差是 $(v_1+v_2,w)-(v_1,w)-(v_2,w)$（属于 S_0，进而属于 $\mathrm{Span}_{\mathbb{F}}(S_0)$）。

因此在商空间 $V\otimes_{\mathbb{F}}W$ 中，(v_1+v_2,w) 的等价类 等于 $(v_1,w)+(v_2,w)$ 的等价类，即 $(v_1+v_2)\otimes w=v_1\otimes w+v_2\otimes w$。

类似的，我们可以给出剩下的三个表达式：

第二类：$(v,w_1+w_2)-(v,w_1)-(v,w_2)$——要求“对第二个变量 w 加法线性”，同理。

第三类：$a(v,w)-(v,aw)$——要求“对第二个变量 w 数乘线性”，即数乘 a 可以从 w 外移到 w 内。

第四类：$a(v,w)-(av,w)$——要求“对第一个变量 v 数乘线性”，即数乘 a 可以从 v 外移到 v 内。

$\square$

我们可以用三个例子来演示：

1. **实直线与实直线的张量积**：取 $V=\mathbb{R}$，$W=\mathbb{R}$，则 $V\otimes_{\mathbb{R}}W\cong\mathbb{R}$。因为实数乘法是双线性的，张量积在此场景下可“还原”为实数自身，且 $a\otimes b$ 可对应到 ab（符合商空间中等价类的线性性）。

2. **二维与三维实空间的张量积**：设 $V=\mathbb{R}^2$（基为 $e_1=(1,0),e_2=(0,1)$），$W=\mathbb{R}^3$（基为 $f_1=(1,0,0),f_2=(0,1,0),f_3=(0,0,1)$），则 $V\otimes_{\mathbb{R}}W$ 是 6 维实向量空间，基为 $e_1\otimes f_1,e_1\otimes f_2,e_1\otimes f_3,e_2\otimes f_1,e_2\otimes f_2,e_2\otimes f_3$。

3. **多项式空间的张量积**：取 $V=\mathbb{R}[x]$（一元实多项式空间），$W=\mathbb{R}[y]$（另一元实多项式空间），则 $V\otimes_{\mathbb{R}}W$ 同构于二元实多项式空间 $\mathbb{R}[x,y]$。例如，$(a+bx)\otimes(c+dy)$ 对应二元多项式 $ac+adx+bcy+bdxy$，体现了“变量独立组合”的双线性性质。

接下来我们考察张量积的**维度**：

Theorem

设 U,V 是 $\mathbb{F}$-向量空间，分别以 S_1,S_2 为基。则 $U\otimes V$ 的基为 $S=\{\varepsilon\otimes\eta:\varepsilon\in S_1,\eta\in S_2\}$。

因此，$\dim_{\mathbb{F}}(U\otimes V)=\dim_{\mathbb{F}}(U)\cdot\dim_{\mathbb{F}}(V)$。

证明：我们证明集合 $S=\{\varepsilon\otimes\eta:\varepsilon\in S_1,\eta\in S_2\}$ 是 $U\otimes V$ 的基。

(i) $\mathrm{Span}_{\mathbb{F}}(S)=U\otimes V$。

$U\otimes V$ 中的任意元素都可写成有限和 $\sum_{i=1}^{n}u_i\otimes v_i$，其中 $u_i\in U$，$v_i\in V$。由于 S_1 是 U 的基，每个 u_i 可表示为 $u_i=\sum_{k=1}^{s}a_{ik}\varepsilon_k$，其中 $a_{ik}\in\mathbb{F}$。类似地，每个 v_i 可表示为 $v_i=\sum_{j=1}^{t}b_{ij}\eta_j$。由张量积的双线性性，我们有：

$$u_i\otimes v_i=\left(\sum_{k=1}^{s}a_{ik}\varepsilon_k\right)\otimes\left(\sum_{j=1}^{t}b_{ij}\eta_j\right)=\sum_{\substack{1\le k\le s\\1\le j\le t}}a_{ik}b_{ij}(\varepsilon_k\otimes\eta_j).$$

于是和式 $\sum_{i=1}^{n}u_i\otimes v_i$ 是 $\varepsilon_k\otimes\eta_j$ 型元素的线性组合。因此，S 张成 $U\otimes V$。

(ii) S 线性无关。

假设

$$\sum_{k,j}c_{kj}\varepsilon_k\otimes\eta_j=0,$$

其中仅有有限个 c_{kj} 非零。对每一对 (p,q)（其中 $\varepsilon_p\in S_1$，$\eta_q\in S_2$），考虑由对偶基定义的线性泛函 $\phi_p:U\to\mathbb{F}$ 和 $\psi_q:V\to\mathbb{F}$，即 $\phi_p(\varepsilon_k)=\delta_{pk}$，$\psi_q(\eta_j)=\delta_{qj}$（克罗内克 delta δ_{ij} 定义为 $\delta_{ij}=\begin{cases}1, & i=j,\\0, & i\neq j.\end{cases}$）。

定义双线性映射 $f_{pq}:U\times V\to\mathbb{F}$ 为 $f_{pq}(u,v)=\phi_p(u)\psi_q(v)$。由泛性质，这诱导出线性映射 $\widetilde{f}_{pq}:U\otimes V\to\mathbb{F}$，使得 $\widetilde{f}_{pq}(u\otimes v)=f_{pq}(u,v)$。

于是有

$$\widetilde{f}_{pq}\left(\sum_{k,j}c_{kj}\varepsilon_k\otimes\eta_j\right)=\sum_{k,j}c_{kj}\widetilde{f}_{pq}(\varepsilon_k\otimes\eta_j)=\sum_{k,j}c_{kj}\phi_p(\varepsilon_k)\psi_q(\eta_j)=\sum_{k,j}c_{kj}\delta_{pk}\delta_{qj}=c_{pq}.$$

由于该和为零，故 $c_{pq}=0$。这对所有 p,q 成立，因此所有系数均为零。故 S 线性无关。

综上，S 是 $U\otimes V$ 的基，且 $\dim_{\mathbb{F}}(U\otimes V)=\#(S_1)\cdot\#(S_2)=\dim_{\mathbb{F}}(U)\cdot\dim_{\mathbb{F}}(V)$。□

【Remark】:

张量是张量积空间 $U\otimes V$ 中的任意元素(形如 $\sum_i a_iu_i\otimes v_i, a_i\in\mathbb{F}, u_i\in U, v_i\in V$)，其中 $u_i\otimes v_i$ 称为‘秩 1 张量’；

而**张量积**包含两层含义：

一是‘张量积操作’ $\otimes:U\times V\to U\otimes V$（构造秩 1 张量的运算）；

二是‘张量积空间’ $U\otimes V$（所有张量构成的向量空间）；张量是张量积操作与空间的共同产物。

1.9.4 COMPOSITION

设 $\{W_i\}_{i\in I}$ 是向量空间 V 的子空间族，I 为指标集（可有限或无限）。

1. 和（Sum）: $\sum_{i\in I}W_i$

定义：由**有限个** W_i 中元素的和构成，即

$$\sum_{i\in I} W_i := \left\{ \sum_{i\in J} w_i \,\middle|\, w_i \in W_i,\ J \subset I,\ \#(J) < \infty \right\} \subset V.$$

（无限和无定义，故限制为有限和，元素可表示为有限个 W_i 中元素的和。）

例子：取 $V = \mathbb{R}^3$，$I = \{1, 2, 3\}$，$W_1 = \mathrm{span}\{(1, 0, 0)\}$（x 轴），$W_2 = \mathrm{span}\{(0, 1, 0)\}$（y 轴），$W_3 = \mathrm{span}\{(0, 0, 1)\}$（z 轴）。则 $\sum_{i=1}^{3} W_i = \mathbb{R}^3$，因任意 $(a, b, c) = a(1, 0, 0) + b(0, 1, 0) + c(0, 0, 1)$（有限和）。

2. 交（Intersection）： $\bigcap_{i\in I} W_i$

定义：所有 W_i 的**公共元素**组成的集合，即

$$\bigcap_{i\in I} W_i = \{v \in V \mid v \in W_i,\ \forall i \in I\},$$

是 V 的子空间。

例子：取 $V = \mathbb{R}^2$，$I = \{1, 2\}$，$W_1 = \mathrm{span}\{(1, 0)\}$（x 轴），$W_2 = \mathrm{span}\{(1, 1)\}$（直线 $y = x$）。

则 $\bigcap_{i=1}^{2} W_i = \{0\}$（仅零向量是公共元素）。

3. 并（Union）： $\bigcup_{i\in I} W_i$

定义：所有 W_i 的元素“合并”而成的集合，即

$$\bigcup_{i\in I} W_i = \{v \in V \mid \exists i \in I,\ v \in W_i\}.$$

例子：取 $V = \mathbb{R}^2$，$I = \{1, 2\}$，$W_1 = \mathrm{span}\{(1, 0)\}$（x 轴），$W_2 = \mathrm{span}\{(0, 1)\}$（y 轴）。则 $\bigcup_{i=1}^{2} W_i$ 是 x 轴和 y 轴上所有点的集合。

【Remark】：

但是该“并”**不是向量空间**（如 $(1, 0) + (0, 1) = (1, 1) \notin \bigcup_{i=1}^{2} W_i$，不满足加法封闭性）。

4. 内直和（Direct Sum）： $\bigoplus_{i\in I} W_i$（作为 V 的子集）

定义：由**有限个** W_i 的直和“扩展”到无限指标，即

$$\bigoplus_{i\in I} W_i := \bigcup_{\substack{J\subset I\\ \#(J)<\infty}} \left(\bigoplus_{i\in J} W_i \right) \subset V.$$

有限直和 $\bigoplus_{i\in J} W_i$ 满足“分解唯一性”，内直和要求这种唯一性推广到“无限和中仅有限项非零的分解”。

我们可以考察几个经典的例子：

Example

1. 取 $V=\mathbb{R}^\infty$，无限维实向量空间，元素为**有限非零**的实数列 $(a_1,a_2,\ldots)$。$I=\mathbb{N}$，$W_i=\operatorname{span}\{e_i\}$（$e_i$ 是第 i 个分量为 1、其余为 0 的向量）。
则 $\bigoplus_{i\in\mathbb{N}}W_i=V$，因任意 $v\in V$ 是有限个 e_i 的线性组合，且分解唯一（每个分量对应唯一的 W_i 中元素）。
2. 又如 $\bigoplus_{i\in I}\mathbb{F}$ 是"有限支撑"的元组空间：给指标集 I 中每个元素 i 分配一个 $\mathbb{F}$ 的拷贝，空间中的元素是"只有有限个位置非零的元组"。
2.1：例子 1（有限指标集）：取 $I=\{1,2,3\}$，则 $\bigoplus_{i=1}^3\mathbb{F}=\mathbb{F}\oplus\mathbb{F}\oplus\mathbb{F}=\mathbb{F}^3$（就是普通的 3 维向量空间，因为有限个分量自然"有限支撑"）。
2.2：例子 2（无限指标集）：取 $I=\mathbb{N}$（自然数集），则 $\bigoplus_{n\in\mathbb{N}}\mathbb{F}$ 中的元素是"只有有限个自然数位置非零的序列"。比如 $(a_0,a_1,0,0,\ldots)$（前两个分量非零，其余为 0）属于这个空间；但无限非零的序列（如 $(1,1,1,\ldots)$）不属于，因为"非零分量的位置集合（支撑）"不是有限的。

5. 直积（Product）：$\prod_{i\in I}W_i$

定义：由所有"序列"$(w_i)_{i\in I}$（其中 $w_i\in W_i$）组成的向量空间，运算为**逐分量**进行（和、数乘对每个分量单独操作）。

例子：取 $I=\mathbb{N}$，$W_i=\mathbb{R}$（对所有 $i\in\mathbb{N}$）。则 $\prod_{i\in\mathbb{N}}\mathbb{R}=\mathbb{R}^{\mathbb{N}}$（所有实数列的集合，包括**无限非零项**的序列，如 $(1,1,1,\ldots)$）。

6. 张量积（Tensor Product）：$\bigotimes_{i\in I}W_i$

定义：多重张量积的推广，先通过"有限张量积"（如 $W_1\otimes W_2$）的**泛性质**（将多线性映射转化为线性映射），再扩展到无限指标（核心是"捕捉多重线性关系"）。

例子 1：有限维向量空间的张量积（双线性性的直观体现）

取指标集 $I=\{1,2\}$，向量空间 $W_1=\mathbb{R}^2$（其标准基为 $\mathbf{e}_1=(1,0)$、$\mathbf{e}_2=(0,1)$），$W_2=\mathbb{R}^3$（其标准基为 $\mathbf{f}_1=(1,0,0)$、$\mathbf{f}_2=(0,1,0)$、$\mathbf{f}_3=(0,0,1)$）。

张量积 $W_1\otimes W_2$ 是 **6 维实向量空间**，它的一组基为 $\{\mathbf{e}_i\otimes\mathbf{f}_j\mid i=1,2;\ j=1,2,3\}$（共 $2\times3=6$ 个基元素）。

若取 W_1 中的元素 $\mathbf{v}=2\mathbf{e}_1+3\mathbf{e}_2$（即向量 $(2,3)$），W_2 中的元素 $\mathbf{w}=4\mathbf{f}_1+5\mathbf{f}_2+6\mathbf{f}_3$（即向量 $(4,5,6)$），则它们的张量积为：

$$\mathbf{v}\otimes\mathbf{w}=(2\mathbf{e}_1+3\mathbf{e}_2)\otimes(4\mathbf{f}_1+5\mathbf{f}_2+6\mathbf{f}_3)=8(\mathbf{e}_1\otimes\mathbf{f}_1)+10(\mathbf{e}_1\otimes\mathbf{f}_2)+12(\mathbf{e}_1\otimes\mathbf{f}_3)+12(\mathbf{e}_2\otimes\mathbf{f}_1)+15(\mathbf{e}_2\otimes\mathbf{f}_2)+18(\mathbf{e}_2\otimes\mathbf{f}_3).$$

这个例子直观体现了张量积的**双线性性**：对 W_1 和 W_2 中的元素分别保持线性，并且通过"基向量的张量积"生成了新空间的基。

例子 2：多项式空间的张量积（多重线性与变量组合）

取 $W_1=\mathbb{R}[x]$（一元实系数多项式空间，其基为 $\{1,x,x^2,\ldots\}$），$W_2=\mathbb{R}[y]$（另一元实系数多项式空间，其基为 $\{1,y,y^2,\ldots\}$）。

张量积 $W_1\otimes W_2$ 自然同构于**二元实系数多项式空间** $\mathbb{R}[x,y]$。

若 $p(x)=1+2x\in\mathbb{R}[x]$，$q(y)=3+4y+5y^2\in\mathbb{R}[y]$，则 $p(x)\otimes q(y)$ 对应二元多项式：

$$(1+2x)\otimes(3+4y+5y^2)=3(1\otimes1)+4(1\otimes y)+5(1\otimes y^2)+6(x\otimes1)+8(x\otimes y)+10(x\otimes y^2),$$

也就是二元多项式 $3+4y+5y^2+6x+8xy+10xy^2$。

该例子体现了张量积“捕捉变量间独立组合的多重线性关系”的特点——将单变量多项式的“线性组合”扩展为双变量多项式的“双线性组合”。

例子 3：线性变换空间的张量积（算子的张量表示）

设 $W_1=\mathcal{L}(V,V')$ (从向量空间 V 到向量空间 V' 的线性变换空间), $W_2=\mathcal{L}(W,W')$ (从向量空间 W 到向量空间 W' 的线性变换空间)。

张量积 $W_1\otimes W_2$ 可自然对应到 ** 双线性变换空间 ** $\mathcal{L}(V\times W,V'\times W')$（或者更精细地说，对应 $\mathcal{L}(V\otimes W,V'\otimes W')$）。

取 $V=\mathbb{R}^2$、$V'=\mathbb{R}^3$，定义线性变换 $T\in\mathcal{L}(\mathbb{R}^2,\mathbb{R}^3)$ 为 $T(\mathbf{e}_1)=(1,0,0)$、$T(\mathbf{e}_2)=(0,1,0)$；取 $W=\mathbb{R}^3$、$W'=\mathbb{R}^2$，定义线性变换 $S\in\mathcal{L}(\mathbb{R}^3,\mathbb{R}^2)$ 为 $S(\mathbf{f}_1)=(1,0)$、$S(\mathbf{f}_2)=(0,1)$、$S(\mathbf{f}_3)=(1,1)$；则张量积 $T\otimes S$ 作用于 $\mathbb{R}^2\otimes\mathbb{R}^3$ 中的元素 $\mathbf{v}\otimes\mathbf{w}$ 时，满足 $(T\otimes S)(\mathbf{v}\otimes\mathbf{w})=T(\mathbf{v})\otimes S(\mathbf{w})$。例如，元素 $\mathbf{e}_1\otimes\mathbf{f}_1$ 在 $T\otimes S$ 下的像为 $T(\mathbf{e}_1)\otimes S(\mathbf{f}_1)=(1,0,0)\otimes(1,0)$。

该例子体现了张量积在 ** 算子理论 ** 中的作用：将“单个空间上的线性变换”扩展为“两个空间上的双线性变换”，是研究多线性算子的复合与表示的核心工具。

1.10 线性变换及其核与像空间

1.10.1 映射

Definition

映射（mapping）：设 V 和 W 是向量空间，一个从 V 到 W 的**映射**（或**函数**）是一个规则 T，它为每个 $v\in V$ 指派唯一的元素 $T(v)\in W$。记作 $T:V\to W$。

映射可以看作函数这一概念的推广

然后我们可以自然地定义:**映射的定义域、值域与陪域**(Domain, Range and Codomain):

对于映射 $T:V\to W$,

- V 称为 T 的**定义域**;
- W 称为 T 的**陪域**;
- 集合 $\{T(v):v\in V\}$ 称为 T 的**值域**，记作 $\mathrm{R}(T)$。

1.10.2 线性变换的定义及相关术语

Definition

线性变换（linear transformation）：设 V 和 W 是数域 F 上的向量空间。一个映射 $T: V \to W$ 称为**线性变换**，如果对所有 $u, v \in V$ 和 $c \in F$，满足：

1. $T(u+v) = T(u) + T(v)$；
2. $T(cv) = cT(v)$。

一般来说，线性变换的记号为 T，当然也可记作 Φ（fai）。并且线性变换也称为**线性映射**

当 $V = W$ 时，线性变换也可以称为：**线性算子（Linear operator）**。

我们首先考察其中 2 种特殊的线性映射——**零变换与恒等变换**（Zero transformation and identity transformation）：

Definition

- **零变换** $O: V \to W$ 定义为对所有 $v \in V$，$O(v) = 0_W$。
- **恒等变换** $I_V: V \to V$ 定义为对所有 $v \in V$，$I_V(v) = v$。

【Remark】：

对于任意的线性变换 T，我们有 $T\,0 = 0$. 但是如果 $T(x) = 0$, 我们无法推出 T 为零变换或者 $x = 0$。（见下部分的核空间）

接下来我们来考察一下线性变换的实例——微分与积分的运算：

考虑向量空间 $P_n(F)$（次数不超过 n 的多项式），定义**微分算子** $D: P_n(F) \to P_n(F)$ 为

$$D(p(x)) = \frac{d}{dx}p(x).$$

D 是线性变换，因为：

1. $D(p(x)+q(x)) = D(p(x)) + D(q(x))$；
2. $D(cp(x)) = cD(p(x))$ 对所有 $c \in F$。

定义**积分算子** $T: P_n(F) \to P_{n+1}(F)$ 为

$$T(p(x)) = \int_0^x p(t)dt.$$

T 也是线性变换。

1.10.3 线性变换的性质

设 $T:V\to W$ 是线性变换，有以下几条常见的性质：

Property

- $T(0)=0$，即：$T(0_V)=0_W$；

  proof：$T(0)=T(0\cdot 0)=0\cdot T(0)=0$
- $T(-v)=-T(v)$ 对所有 $v\in V$；
- $T(c_1v_1+c_2v_2+\cdots+c_kv_k)=c_1T(v_1)+c_2T(v_2)+\cdots+c_kT(v_k)$ 对所有 $v_1,v_2,\ldots,v_k\in V$ 和 $c_1,c_2,\ldots,c_k\in F$；
- 如果 T 和 S 都是线性映射，那么：也是 $T+S$ 线性映射;

  proof:$(T+S)(av1+bv2)=T(av1+bv2)+S(av1+bv2)=aT(v1)+bT(v2)+aS(v1)+bS(v2)=a(T+S)(v1)+b(T+S)(v2)$.
- 如果 T 和 S 都是线性映射，那么：也是 $T\circ S$ 线性映射。

  proof:$(T\circ S)(au1+bu2)=T(S(au1+bu2))=T(aS(u1)+bS(u2))=aT(S(u1))+bT(S(u2))=a(T\circ S)(u1)+b(T\circ S)(u2)$.

由最后两条性质，我们可以定义线性变换的叠加与复合：

1.10.4 线性变换的叠加与复合

首先，我们介绍加法与数乘，即线性变换的叠加：

Definition

线性变换的加法与数乘：设 $T,U:V\to W$ 是线性变换，$c\in F$。定义：

- $(T+U)(v)=T(v)+U(v)$ 对所有 $v\in V$；
- $(cT)(v)=cT(v)$ 对所有 $v\in V$。

其中 $T+U$ 和 cT 都是线性变换，我们分别称之为：线性变换的加法与数乘。

Definition

线性变换的复合：设 $T: V \to W$ 和 $U: W \to Z$ 是线性变换。它们的**复合** $U \circ T: V \to Z$ 定义为

$$(U \circ T)(v) = U(T(v)) \quad 对所有v \in V.$$

$U \circ T$ 也是线性变换。我们称之为：线性变换的复合。

Property

线性变换的复合满足结合律：若 $T: V \to W$，$U: W \to Z$，$S: Z \to Y$，则

$$S \circ (U \circ T) = (S \circ U) \circ T.$$

证明．proof:

□

1.10.5 线性变换 T 的核与像

Definition

核空间/零空间（Kernel space /null space）：设 $T: V \to W$ 是线性变换。T 的**核空间**（或零空间）定义为

$$N_T = N\ T \ = ker(T) = \{v \in V : T(v) = 0_W\}.$$

换言之，核空间的几何意义就是记录**线性变换 T 把定义域空间里面的哪些向量信息搞丢了**（被压缩到陪域空间的 0 向量）。

Definition

值域/像空间（Range/Image space）：T 的**值域**（或**像空间**）定义为

$$\mathrm{R}(T) = \{T(v) : v \in V\}.$$

换言之，值域 $R\ T$ 则是记录了定义域空间经过线性变换 T，在**新的空间输出的全部信息**。但是注意到初学者可能将值域和陪域空间 W 相等。这是错误的，R 不一定就是 W，空间 W 中可能有不属于 R 的元素（即没有被线性变换 T 映射到的元素）

Theorem

设 $T: V \to W$ 是线性变换，则：$ker(T)$ 和 $R(T)$ 均为向量空间，进一步地，$\ker(T)$ 是 V 的子空间；$\mathrm{R}(T)$ 是 W 的子空间。

证明．proof: 要证明 $\mathrm{Ker}(T)$ 和 $\mathrm{Im}(T)$ 是子空间。

- 核（Kernel）：设 $v_1, v_2 \in \mathrm{Ker}(T)$，$a, b \in \mathbb{F}$。则 $T(av_1 + bv_2) = aT(v_1) + bT(v_2) = 0$，因此 $av_1 + bv_2 \in \mathrm{Ker}(T)$。

- 像（Image）：设 $w_1, w_2 \in \mathrm{Im}(T)$，故存在 $v_1, v_2 \in V$ 使得 $T(v_i) = w_i$。对于 $a, b \in \mathbb{F}$，有 $aw_1 + bw_2 = aT(v_1) + bT(v_2) = T(av_1 + bv_2) \in \mathrm{Im}(T)$。

□

接下来我们通过一个例子来更好地理解核与值域的概念：

Example

对于 $T : \mathbb{R}^3 \to \mathbb{R}^2$，$T(x, y, z) = (x - y, y + z)$

求 N_T。

令 $T(x, y, z) = (0, 0)$，即 $\begin{cases} x - y = 0 \\ y + z = 0 \end{cases}$，解得 $x = y, y = -z$。取 $(x, y, z) = (1, 1, -1)$，所以 $N_T = \mathrm{Span}\{(1, 1, -1)\}$。

求 $R(T)$

对于任意 $(a, b, c) \in \mathbb{R}^3$，$T(a, b, c) = (a - b, b + c)$。因为 $a - b$ 和 $b + c$ 都可以遍历所有实数 $\mathbb{R}$，所以 $R(T) = \mathbb{R}^2$。$R(T)$ 是 T 作用下 $\mathbb{R}^3$ 在 $\mathbb{R}^2$ 中的像集。

有意思的是，R 和 N 不一定是不同的。

Example

如线性变换 $T : R^2 \to R^2$ 的例子：$T(a, b) = (b, 0)$，我们有 $N(T) = R(T)$。

现在我们来考察一下线性变换对线性相关性的保持性：

Theorem

设 $T : V \to W$ 是线性变换，$B = \{v_1, v_2, \ldots, v_n\}$ 是 V 的基。那么：$\mathrm{R}(T) = \mathrm{span}(\{T(v_1), T(v_2), \ldots, T(v_n)\})$。

证明. proof:

要证明 $R(T) = \mathrm{span}\{T(v_1), T(v_2), \ldots, T(v_n)\}$，只需证明两边互相包含：

1. 任取 $w \in R(T)$，则存在 $v \in V$ 使 $T(v) = w$。因 $\{v_i\}$ 是基，$v = \sum a_i v_i$，故 $w = \sum a_i T(v_i) \in \mathrm{span}\{T(v_i)\}$。

2. 任取 $w \in \mathrm{span}\{T(v_i)\}$，则 $w = \sum b_i T(v_i) = T\left(\sum b_i v_i\right) \in R(T)$。

综上，$R(T) = \mathrm{span}\{T(v_1), \ldots, T(v_n)\}$。

□

【Remark】:

注意到：$\mathrm{R}(T)=\mathrm{span}(\{T(v_1),T(v_2),\ldots,T(v_n)\})$，只需张成，无需是基（线性无关）。我们可以证明：只有当 T 是单射时，才有线性无关这一条件。

Corollary & Secondary Conclusion

C1：对于任意的向量集 $S\in V$，若 S 线性相关，则向量集合 $T(S)$ 也线性相关。

C2：特别的，若 S 线性无关，$T(S)$ 无法保证依然线性无关。

1.10.6 线性算子幂的核与像

对于线性算子的幂，我们有：

Property

设 V 是 $\mathbb{C}$-向量空间，$T:V\to V$ 是 $\mathbb{C}$-线性算子。
那么对任意正整数 $i\in\mathbb{N}_+$，$\ker(T^i)$ 和 $\mathrm{Im}(T^i)$ 都是 V 的 $\mathbb{C}$-向量子空间。

证明. (i) 先证对任意 $i\in\mathbb{N}_+$，T^i 是 $\mathbb{C}$-线性映射。

我们对 i 用数学归纳法。当 $i=1$ 时，显然 T 是 $\mathbb{C}$-线性映射。假设当 $i=k-1$ 时 T^{k-1} 是 $\mathbb{C}$-线性映射，我们证明 $i=k$ 时结论成立。对任意标量 $a,b\in\mathbb{C}$ 和向量 $u,v\in V$，有

$$\begin{aligned}T^k(au+bv)&=T^{k-1}(T(au+bv))=T^{k-1}(aT(u)+bT(v))\\&=aT^{k-1}(T(u))+bT^{k-1}(T(v))=aT^k(u)+bT^k(v).\end{aligned}$$

由数学归纳法，对所有 $i\in\mathbb{N}_+$，T^i 是 $\mathbb{C}$-线性映射。

证明 $\ker(T^i)$ 是 $\mathbb{C}$-向量子空间。对任意标量 $a,b\in\mathbb{C}$ 和向量 $u,v\in\ker(T^i)$，有

$$T^i(au+bv)=aT^i(u)+bT^i(v)=a\cdot 0+b\cdot 0=0.$$

因此 $au+bv\in\ker(T^i)$，故 $\ker(T^i)$ 是 $\mathbb{C}$-向量子空间。

证明 $\mathrm{Im}(T^i)$ 是 $\mathbb{C}$-向量子空间。对任意标量 $a,b\in\mathbb{C}$ 和向量 $u,v\in\mathrm{Im}(T^i)$，即存在 $u',v'\in V$ 使得 $u=T^i(u')$，$v=T^i(v')$。则

$$T^i(au'+bv')=aT^i(u')+bT^i(v')=au+bv.$$

因此 $au+bv\in\mathrm{Im}(T^i)$，故 $\mathrm{Im}(T^i)$ 是 $\mathbb{C}$-向量子空间。 □

并且，我们还可以观察到，随着幂的次数逐渐增大：

Fact or Background

任意正整数 $i > j \in \mathbb{N}_+$，有 $\mathrm{Im}(T^i) \subset \mathrm{Im}(T^j)$：即即复合越多，值域空间越小；
而 $\ker(T^i) \supset \ker(T^j)$ 对任意正整数 $i > j$ 成立：即复合越多，kernal 空间越大；

Lemma

像的稳定性： $\mathrm{Im}(T^i) = \mathrm{Im}(T^j) \implies \mathrm{Im}(T^{j+n}) = \mathrm{Im}(T^j)$

证明. 设 V 是 $\mathbb{C}$-向量空间，$T : V \to V$ 是 $\mathbb{C}$-线性映射。已知存在正整数 $i > j$ 使得 $\mathrm{Im}(T^i) = \mathrm{Im}(T^j)$，需证对任意非负整数 n，有 $\mathrm{Im}(T^{j+n}) = \mathrm{Im}(T^j)$。

采用数学归纳法证明：

基例 $n = 1$：因 $i > j$，故 $i \geq j+1$。由像的包含性结论 $\mathrm{Im}(T^i) \subset \mathrm{Im}(T^{j+1}) \subset \mathrm{Im}(T^j)$，且已知 $\mathrm{Im}(T^i) = \mathrm{Im}(T^j)$，故 $\mathrm{Im}(T^{j+1}) = \mathrm{Im}(T^j)$。

归纳假设：假设对非负整数 $k-1$，有 $\mathrm{Im}(T^{j+k-1}) = \mathrm{Im}(T^j)$。

归纳步骤（$n = k$）：由像的定义，$\mathrm{Im}(T^{j+k}) = T^{j+k}(V) = T(T^{j+k-1}(V))$。

由归纳假设 $\mathrm{Im}(T^{j+k-1}) = \mathrm{Im}(T^j)$，即 $T^{j+k-1}(V) = T^j(V)$，故：

$$\mathrm{Im}(T^{j+k}) = T(T^j(V)) = T^{j+1}(V) = \mathrm{Im}(T^{j+1}) = \mathrm{Im}(T^j)$$

因此，由数学归纳法，对任意非负整数 n，$\mathrm{Im}(T^{j+n}) = \mathrm{Im}(T^j)$。 □

Lemma

核的稳定性： $\ker(T^i) = \ker(T^j) \implies \ker(T^{j+n}) = \ker(T^j)$

证明. 设 V 是 $\mathbb{C}$-向量空间，$T : V \to V$ 是 $\mathbb{C}$-线性映射。已知存在正整数 $i > j$ 使得 $\ker(T^i) = \ker(T^j)$，需证对任意非负整数 n，有 $\ker(T^{j+n}) = \ker(T^j)$。

采用数学归纳法证明：

基例 $n = 1$：因 $i > j$，故 $i \geq j+1$。由核的包含性结论 $\ker(T^j) = \ker(T^i) \supset \ker(T^{j+1}) \supset \ker(T^j)$，故 $\ker(T^{j+1}) = \ker(T^j)$。

归纳假设：假设对非负整数 $k-1$，有 $\ker(T^{j+k-1}) = \ker(T^j)$。

归纳步骤（$n = k$）：任取 $v \in \ker(T^{j+k})$，则 $T^{j+k}(v) = 0$。

分解得 $T^{j+k-1}(T(v)) = 0$，故 $T(v) \in \ker(T^{j+k-1})$。

由归纳假设 $\ker(T^{j+k-1}) = \ker(T^j)$，得 $T(v) \in \ker(T^j)$，即 $T^j(T(v)) = 0$，即 $T^{j+1}(v) = 0$，故 $v \in \ker(T^{j+1})$。由基例 $n = 1$ 知 $\ker(T^{j+1}) = \ker(T^j)$，故 $v \in \ker(T^j)$，即 $\ker(T^{j+k}) \subset \ker(T^j)$。

又由核的包含性结论 $\ker(T^{j+k}) \supset \ker(T^j)$，故 $\ker(T^{j+k}) = \ker(T^j)$。

因此，由数学归纳法，对任意非负整数 n，$\ker(T^{j+n}) = \ker(T^j)$。 □

Lemma

有限维下像与核的等价性: 设 $\dim_{\mathbb{C}}(V) < \infty$，那么对任意正整数 i, j:

1. $\mathrm{Im}(T^i) = \mathrm{Im}(T^j) \implies \ker(T^i) = \ker(T^j)$;
2. $\ker(T^i) = \ker(T^j) \implies \mathrm{Im}(T^i) = \mathrm{Im}(T^j)$。

证明. **证明 $\mathbf{Im}(T^i) = \mathbf{Im}(T^j) \implies \ker(T^i) = \ker(T^j)$**

若 $i = j$，结论显然成立。下设 $i > j$。

由 $\mathrm{Im}(T^i) = \mathrm{Im}(T^j)$，得 $\dim_{\mathbb{C}}(\mathrm{Im}(T^i)) = \dim_{\mathbb{C}}(\mathrm{Im}(T^j))$。根据秩-零化度定理（对任意正整数 k，$\dim_{\mathbb{C}}(V) = \dim_{\mathbb{C}}(\mathrm{Im}(T^k)) + \dim_{\mathbb{C}}(\ker(T^k))$），有:

$$\dim_{\mathbb{C}}(\ker(T^i)) = \dim_{\mathbb{C}}(V) - \dim_{\mathbb{C}}(\mathrm{Im}(T^i)) = \dim_{\mathbb{C}}(V) - \dim_{\mathbb{C}}(\mathrm{Im}(T^j)) = \dim_{\mathbb{C}}(\ker(T^j))$$

又由核的包含性结论 $\ker(T^i) \supset \ker(T^j)$，结合维数相等，得 $\ker(T^i) = \ker(T^j)$。

证明 $\ker(T^i) = \ker(T^j) \implies \mathbf{Im}(T^i) = \mathbf{Im}(T^j)$

若 $i = j$，结论显然成立。下设 $i > j$。

由 $\ker(T^i) = \ker(T^j)$，得 $\dim_{\mathbb{C}}(\ker(T^i)) = \dim_{\mathbb{C}}(\ker(T^j))$。根据秩-零化度定理，有:

$$\dim_{\mathbb{C}}(\mathrm{Im}(T^i)) = \dim_{\mathbb{C}}(V) - \dim_{\mathbb{C}}(\ker(T^i)) = \dim_{\mathbb{C}}(V) - \dim_{\mathbb{C}}(\ker(T^j)) = \dim_{\mathbb{C}}(\mathrm{Im}(T^j))$$

又由像的包含性结论 $\mathrm{Im}(T^i) \subset \mathrm{Im}(T^j)$，结合维数相等，得 $\mathrm{Im}(T^i) = \mathrm{Im}(T^j)$。 □

【Remark】:

在无限维的情况下，该引理不成立:

证明. 令 $V = \mathbb{C}^\infty$（所有复数列 $(a_1, a_2, \ldots, a_n, \ldots)$ 构成的复向量空间），定义两个线性映射:

T_1（左移映射）: $T_1(a_1, a_2, \ldots, a_n, \ldots) = (a_2, a_3, \ldots, a_n, \ldots)$;

T_2（右移 + 零填充映射）: $T_2(a_1, a_2, \ldots, a_n, \ldots) = (0, a_1, a_2, \ldots, a_n, \ldots)$。

分析 T_1: 像（Im）的分析: 对任意 $i, j \in \mathbb{N}_+$，$\mathrm{Im}(T_1^i) = \mathbb{C}^\infty$。

因为任给复数列 $(b_1, b_2, \ldots)$，总能找到原像 $(b_{i+1}, b_{i+2}, \ldots)$，使得 $T_1^i(b_{i+1}, b_{i+2}, \ldots) = (b_1, b_2, \ldots)$。

同理 $\mathrm{Im}(T_1^j) = \mathbb{C}^\infty$，故 $\mathrm{Im}(T_1^i) = \mathrm{Im}(T_1^j) = V$。

核（Ker）的分析: $\mathrm{Ker}(T_1^i)$ 是满足 $T_1^i(a_1, a_2, \ldots) = 0$ 的列，即 $(a_{i+1}, a_{i+2}, \ldots) = (0, 0, \ldots)$，因此: $\mathrm{Ker}(T_1^i) = \{(a_1, \ldots, a_i, 0, 0, \ldots) \mid a_1, \ldots, a_i \in \mathbb{C}\}$，

同理 $\mathrm{Ker}(T_1^j) = \{(a_1, \ldots, a_j, 0, 0, \ldots) \mid a_1, \ldots, a_j \in \mathbb{C}\}$。

当 $i \neq j$ 时（如 $i > j$），$\mathrm{Ker}(T_1^i)$ 与 $\mathrm{Ker}(T_1^j)$ 包含的序列长度不同（前 i 项 vs 前 j 项非零），故 $\mathrm{Ker}(T_1^i) \neq \mathrm{Ker}(T_1^j)$。

分析 T_2：说明 "$\mathrm{Ker}(T^i) = \mathrm{Ker}(T^j)$ 不蕴含 $\mathrm{Im}(T^i) = \mathrm{Im}(T^j)$"

核（Ker）的分析：$T_2^i(a_1, a_2, \dots) = (0, \dots, 0, a_1, a_2, \dots)$（前 i 个分量为 0），要使其为零向量，需 $a_1 = a_2 = \cdots = 0$，因此对任意 $i \in \mathbb{N}_+$，$\mathrm{Ker}(T_2^i) = \{0\}$，故 $\mathrm{Ker}(T_2^i) = \mathrm{Ker}(T_2^j) = \{0\}$。

像（Im）的分析：$\mathrm{Im}(T_2^i) = \{(0, \dots, 0, a_1, a_2, \dots) \mid a_1, a_2, \dots \in \mathbb{C}\}$（前 i 个分量为 0）。当 $i \neq j$ 时（如 $i > j$），$\mathrm{Im}(T_2^i)$ 中序列的"零前缀长度"为 i，$\mathrm{Im}(T_2^j)$ 为 j，显然不相等（例如 $(0, 1, 0, 0, \dots) \in \mathrm{Im}(T_2^2)$ 但不属于 $\mathrm{Im}(T_2^3)$），故 $\mathrm{Im}(T_2^i) \neq \mathrm{Im}(T_2^j)$。 □

Theorem

定义 $\mathrm{Ker}(T^\infty) := \bigcup_{i \geq 1} \mathrm{Ker}(T^i)$ 且 $\mathrm{Im}(T^\infty) := \bigcap_{i \geq 1} \mathrm{Im}(T^i)$。
那么当 $\dim_{\mathbb{C}}(V) < \infty$ 时，$V = \mathrm{Ker}(T^\infty) \oplus \mathrm{Im}(T^\infty)$。

证明. 有限维情形的证明：$V = \mathrm{Ker}(T^\infty) \oplus \mathrm{Im}(T^\infty)$ 我们分三步证明：

步骤 1：存在正整数 k，使得 $\mathrm{Ker}(T^\infty) = \mathrm{Ker}(T^k)$ 且 $\mathrm{Im}(T^\infty) = \mathrm{Im}(T^k)$ 线性映射的像列是递减的：$\mathrm{Im}(T) \supseteq \mathrm{Im}(T^2) \supseteq \mathrm{Im}(T^3) \supseteq \cdots$。

对应的维数序列 $\dim(\mathrm{Im}(T^i))$ 是非递增的，且下界为 0。由于 V 是有限维的，维数序列必然稳定，即存在正整数 k，使得对所有 $n \in \mathbb{N}_+$，有 $\dim(\mathrm{Im}(T^{k+n})) = \dim(\mathrm{Im}(T^k))$。

有限维子空间"维数相等则子空间相等"，故 $\mathrm{Im}(T^{k+n}) = \mathrm{Im}(T^k)$ 对所有 $n \in \mathbb{N}_+$ 成立。

结合上面的的结论（$\mathrm{Im}(T^i) = \mathrm{Im}(T^j) \implies \mathrm{Ker}(T^i) = \mathrm{Ker}(T^j)$），得 $\mathrm{Ker}(T^{k+n}) = \mathrm{Ker}(T^k)$ 对所有 $n \in \mathbb{N}_+$ 成立。

因此，$\mathrm{Ker}(T^\infty) = \bigcup_{i \geq 1} \mathrm{Ker}(T^i) = \mathrm{Ker}(T^k)$，$\mathrm{Im}(T^\infty) = \bigcap_{i \geq 1} \mathrm{Im}(T^i) = \mathrm{Im}(T^k)$。

步骤 2：$\mathrm{Ker}(T^k) + \mathrm{Im}(T^k)$ 是直和（即 $\mathrm{Ker}(T^k) \cap \mathrm{Im}(T^k) = \{0\}$）任取 $u \in \mathrm{Ker}(T^k) \cap \mathrm{Im}(T^k)$，由 $u \in \mathrm{Im}(T^k)$，存在 $v \in V$ 使得 $u = T^k(v)$。由 $u \in \mathrm{Ker}(T^k)$，得 $T^k(u) = 0$。

代入 $u = T^k(v)$，得 $T^k(T^k(v)) = T^{2k}(v) = 0$，即 $v \in \mathrm{Ker}(T^{2k})$。

由步骤 1，$\mathrm{Ker}(T^{2k}) = \mathrm{Ker}(T^k)$，故 $v \in \mathrm{Ker}(T^k)$，因此 $u = T^k(v) = 0$（因 $v \in \mathrm{Ker}(T^k)$，$T^k(v) = 0$）。故 $\mathrm{Ker}(T^k) \cap \mathrm{Im}(T^k) = \{0\}$，即 $\mathrm{Ker}(T^k) + \mathrm{Im}(T^k) = \mathrm{Ker}(T^k) \oplus \mathrm{Im}(T^k)$。

步骤 3：$V = \mathrm{Ker}(T^k) \oplus \mathrm{Im}(T^k)$

由秩 - 零化度定理，对 T^k 有 $\dim(V) = \dim(\mathrm{Ker}(T^k)) + \dim(\mathrm{Im}(T^k))$。

结合步骤 2, $\dim(\mathrm{Ker}(T^k)+\mathrm{Im}(T^k)) = \dim(\mathrm{Ker}(T^k))+\dim(\mathrm{Im}(T^k))-\dim(\mathrm{Ker}(T^k)\cap \mathrm{Im}(T^k)) = \dim(V)+0 = \dim(V)$。又因 $\mathrm{Ker}(T^k)+\mathrm{Im}(T^k) \subseteq V$，且 V 是有限维的，故 $\mathrm{Ker}(T^k)+\mathrm{Im}(T^k) = V$。

结合步骤 2 的直和，得 $V = \mathrm{Ker}(T^k) \oplus \mathrm{Im}(T^k)$。

再由步骤 1 的 $\mathrm{Ker}(T^\infty) = \mathrm{Ker}(T^k)$、$\mathrm{Im}(T^\infty) = \mathrm{Im}(T^k)$，最终得 $V = \mathrm{Ker}(T^\infty) \oplus \mathrm{Im}(T^\infty)$。

□

【Remark】:

该定理在无限维的时候不成立

证明. 构造无限维空间 $V = \mathbb{C}^\infty$（所有复数列构成的空间），并定义两个线性映射：

T_1（左移映射）：$T_1(a_1, a_2, \dots) = (a_2, a_3, \dots)$；

T_2（右移 + 零填充映射）：$T_2(a_1, a_2, \dots) = (0, a_1, a_2, \dots)$。

分析 T_1: $\mathrm{Ker}(T_1^\infty) = \bigcup_{i\geq 1} \mathrm{Ker}(T_1^i)$。其中 $\mathrm{Ker}(T_1^i) = \{(a_1, \dots, a_i, 0, 0, \dots) \mid a_1, \dots, a_i \in \mathbb{C}\}$（前 i 项任意，后续为 0）。

当 i 取遍所有正整数时，$\bigcup_{i\geq 1} \mathrm{Ker}(T_1^i) = \mathbb{C}^\infty = V$（任何序列都可被某个 i 包含）。$\mathrm{Im}(T_1^\infty) = \bigcap_{i\geq 1} \mathrm{Im}(T_1^i)$。因 $\mathrm{Im}(T_1^i) = \mathbb{C}^\infty$（任给序列都有原像），故交集为 $\mathbb{C}^\infty = V$。

此时 $\mathrm{Ker}(T_1^\infty) \oplus \mathrm{Im}(T_1^\infty) = V \oplus V \neq V$（直和要求交为 $\{0\}$，但此处交为 V）。

分析 T_2：$\mathrm{Ker}(T_2^\infty) = \bigcup_{i\geq 1} \mathrm{Ker}(T_2^i)$。

其中 $\mathrm{Ker}(T_2^i) = \{0\}$ (因 $T_2^i(a_1, a_2, \dots) = (0, \dots, 0, a_1, a_2, \dots)$, 要为 0 则所有 $a_j = 0$)，故并集为 $\{0\}$。$\mathrm{Im}(T_2^\infty) = \bigcap_{i\geq 1} \mathrm{Im}(T_2^i)$。其中 $\mathrm{Im}(T_2^i) = \{(0, \dots, 0, a_{i+1}, a_{i+2}, \dots) \mid a_{i+1}, \dots \in \mathbb{C}\}$（前 i 项为 0）。当 i 取遍所有正整数时，交集为 $\{0\}$（只有全零序列属于所有 $\mathrm{Im}(T_2^i)$）。

此时 $\mathrm{Ker}(T_2^\infty) \oplus \mathrm{Im}(T_2^\infty) = \{0\} \oplus \{0\} = \{0\} \neq V$。

综上，无限维空间中的结论不成立。 □

Method

在吗无限维空间中，如复数域的无限维空间，我们常常考察这两个线性映射：

T_1（左移映射）：$T_1(a_1, a_2, \dots) = (a_2, a_3, \dots)$；

T_2（右移 + 零填充映射）：$T_2(a_1, a_2, \dots) = (0, a_1, a_2, \dots)$。

Method

若需证明两个子空间（如 A, B）相等，常用路径如下：

1. **互为子集法**（最常用）：若能证明子空间满足 $A \subseteq B$ 且 $B \subseteq A$，则直接由集合相等的定义得 $A = B$

   - **【路径 1】**（经典方法）：任取子空间 A 中的元素 v，需证 $v \in B$；再任取子空间 B 中的元素 w，需证 $w \in A$。由双向包含 $A \subseteq B$ 且 $B \subseteq A$，得 $A = B$。
   - **【路径 2】**：双向包含夹逼法：若能证明子空间满足 $A \subseteq B \subseteq C \subseteq A$，则直接由集合相等的定义得 $A = B$（本质是“互为子集法”的简洁表述）。
   - **【路径 3】**（如涉及值域空间的相等）若存在连接 A, B 的线性映射 $T : A \to B$ 或 $T : B \to A$，可通过映射的“像”（记为 $\mathrm{Im}(T)$）分析：

     用 $T(A)$ 或 $T(B)$ 表示全部的映射在对应子空间上的像（值域子空间），结合像的性质（如线性封闭性）推导 A 与 B 的包含关系。

2. **维度夹逼法：** 如果 $A \subseteq B$ 同时：$dim(A) = dim(B)$，那么，我们可以认为 $A = B$。

3. **数学归纳法**：适用于涉及“整数指标的子空间序列”（如 A_n, B_n，$n \in \mathbb{N}$）的相等证明：先验证基例（如 $n = 1$ 时 $A_1 = B_1$），再假设 $n = k$ 时 $A_k = B_k$，递推证明 $n = k + 1$ 时 $A_{k+1} = B_{k+1}$。

1.10.7 课后练习

Exercise

1. 设 $T : V \to W$ 和 $U : W \to Z$ 是线性变换。证明：

   (a) $\mathrm{R}(U \circ T) \subseteq \mathrm{R}(U)$;

   (b) $\ker(T) \subseteq \ker(U \circ T)$;

2. 设 $T : V \to V$ 是线性变换，满足 $T^2 = T$（即 T 是幂等变换）。证明：

   (a) $V = \ker(T) \oplus \mathrm{R}(T)$;

   (b) 若 $v \in \ker(T) \cap \mathrm{R}(T)$，则 $v = 0$。

1.11 维度定理

这一节，我们将证明一个重要的定理.

但在此之前，我们将讨论两个核心概念：秩和零度，尤其是秩（rank），它将贯穿线

性代数的始终。

我们已经知道：核空间记录的是被线性变换 T “搞丢”的信息空间；而值域 $R\,T$ 则是记录了经过线性变换 T，在新的空间输出的全部信息。

我们有必要**考察核空间与值域的维度**，以**量化地**考察线性变换对信息传递的具体情况。

Definition

零度：核空间（$N(T)$）的维度，简记为 *nullity*。

秩：值域空间（$R(T)$）的维度，简记为 *rank*（或 r(T)）

现在我们可以论述**维度定理（秩-零度定理,***Rank-Nullity***）**；

Theorem

维度定理（Dimension Theorem）：设 $T: V \to W$ 是线性变换，且 V 是有限维的。则

$$\dim(V) = \dim(\ker(T)) + \dim(\mathrm{R}(T)).$$

其中 $\dim(\ker(T))$ 是 T 的**零度**，$\dim(\mathrm{R}(T))$ 是 T 的**秩**。

证明. 设 $n = \dim_{\mathbb{F}}(V)$，$k = T$ 的零度（Nullity）。设 $\{u_1, \dots, u_k\}$ 是 $\mathrm{Ker}(T)$ 的一组基。将其扩充为 V 的一组基 $\mathcal{B} = \{u_1, \dots, u_k, w_1, \dots, w_{n-k}\}$。我们断言 $\{T(w_1), \dots, T(w_{n-k})\}$ 是 $\mathrm{Im}(T)$ 的一组基。

- ** 张成性 **：对任意 $w \in \mathrm{Im}(T)$，存在 $v = \sum_{i=1}^{k} a_i u_i + \sum_{j=1}^{n-k} b_j w_j$ 使得 $w = T(v)$。由于 $T(u_i) = 0$，故 $w = \sum_{j=1}^{n-k} b_j T(w_j)$。
- ** 线性无关性 **：假设 $\sum_{j=1}^{n-k} b_j T(w_j) = 0$。则 $T(\sum_{j=1}^{n-k} b_j w_j) = 0$，故 $\sum_{j=1}^{n-k} b_j w_j \in \mathrm{Ker}(T)$。因此，存在 a_i 使得 $\sum_{j=1}^{n-k} b_j w_j = \sum_{i=1}^{k} a_i u_i$。由 $\mathcal{B}$ 的线性无关性，所有 $b_j = 0$（且所有 $a_i = 0$）。

因此，T 的秩（Rank）满足 $\mathrm{Rank}(T) = n - k = \dim_{\mathbb{F}}(V) - \text{零度}(T)$。 □

1.12 线性变换的矩阵表示

这一节，你将彻底理解矩阵的本质：**线性变换在两个向量空间的两组基底下的线性表示**。

首先我们先从“基决定”性质开始：

1.12.1 线性变换的基决定性质

我们先做出如下断言：线性变换 $T: V \to W$ 的作用可通过其在定义域向量空间 V 的一组基向量与陪域 W 上的一组向量上表现的行为完全确定。

详细的说，若向量空间 V 有一组有序基 $\{v_1, v_2, \ldots, v_n\}$，对于线性变换 $T: V \to W$，只要确定 T 对每个基向量 v_i 的作用（即 $T(v_i)$ 的值），那么 T 对 V 中任意向量 $x = \sum_{i=1}^n a_i v_i$ 的作用 $T(x) = \sum_{i=1}^n a_i T(v_i)$ 就被唯一确定。已知有序基向量的像，就可以通过线性组合即可得到任意向量的像。

Example

二维空间中，若已知 T 对基向量 $\mathbf{e}_1 = (1, 0)$ 和 $\mathbf{e}_2 = (0, 1)$ 的变换结果，则任意向量 $\mathbf{v} = a\mathbf{e}_1 + b\mathbf{e}_2$ 的变换结果为 $T(\mathbf{v}) = aT(\mathbf{e}_1) + bT(\mathbf{e}_2)$。

下面我们将为这段表述给出精准的表达与证明：

Theorem

设 V 和 W 是数域 F 上的向量空间，假设 $\{v_1, v_2, \ldots, v_n\}$ 是 V 的一组有序基。对于 W 中的 $w_1, w_2, \ldots, w_n$，存在唯一的线性变换 $T: V \to W$，使得对于 $i = 1, 2, \ldots, n$，有 $T(v_i) = w_i$。

【Remark】：

$w_1, w_2, \ldots, w_n$ 不一定是 W 的基，甚至它们张成的子空间 $\mathrm{span}\{w_1, w_2, \ldots, w_n\}$ 也可以不等于 W。线性变换的唯一性定理对 W 的维数没有要求，$\dim(W)$ 可以等于 n、大于 n 或小于 n，只要 $w_1, \ldots, w_n$ 是 W 中任意给定的向量即可。

证明. 任取 $x \in V$，因为 $\{v_1, v_2, \ldots, v_n\}$ 是 V 的基，所以 $x = \sum_{i=1}^n a_i v_i$，其中 $a_1, a_2, \ldots, a_n$ 是唯一的标量。定义 $T: V \to W$，$T(x) = \sum_{i=1}^n a_i w_i$。

验证 T 满足条件：当 x 取基向量 v_i 时，其系数 $a_i = 1$，其余 $a_j = 0$（$j \neq i$）。代入 T 的定义式，得 $T(v_i) = 1 \cdot w_i + 0 \cdot w_1 + \cdots + 0 \cdot w_{i-1} + 0 \cdot w_{i+1} + \cdots + 0 \cdot w_n = w_i$，即根据 T 的构造方式，直接得出 $T(v_i) = w_i$。

证明 T 的唯一性：假设存在另一个线性变换 $U: V \to W$ 满足 $U(v_i) = w_i$，则对于任意向量 $x = \sum_{i=1}^n a_i v_i$，有 $U(x) = U(\sum_{i=1}^n a_i v_i) = \sum_{i=1}^n a_i U(v_i) = \sum_{i=1}^n a_i w_i = T(x)$，故 $U = T$。 □

于是我们有以下的推论：

Corollary & Secondary Conclusion

设 V 和 W 是向量空间，假设 V 有一组有限基 $\{v_1, v_2, \ldots, v_n\}$。如果 $U, T: V \to W$ 是线性的，且对于 $i = 1, 2, \ldots, n$，$U(v_i) = T(v_i)$，那么 $U = T$。

我们将通过一个例子来理解上述定理与推论:

Example

设 $T:\mathbb{R}^2\to\mathbb{R}^2$ 是由 $T(a_1,a_2)=(2a_2-a_1,3a_1)$ 定义的线性变换，$U:\mathbb{R}^2\to\mathbb{R}^2$ 是线性的，已知 $U(1,2)=(3,3)$，$U(1,1)=(1,3)$，且 $\{(1,2),(1,1)\}$ 是 $\mathbb{R}^2$ 的一组基。

根据上述推论，因为 T 和 U 都是从 $\mathbb{R}^2$ 到 $\mathbb{R}^2$ 的线性变换，且在基 $\{(1,2),(1,1)\}$ 上取值相同（U 在基向量上的值已知，T 作用在基向量上也可计算得到相同结果），所以 $U=T$。

接下来，我们将把**矩阵表示与线性变换**联系起来:

1.12.2 坐标向量与线性变换的矩阵表示

Definition

设 $\beta=\{u_1,u_2,\ldots,u_n\}$ 是有限维向量空间 V 的一个有序基。对 $x\in V$，令 $a_1,a_2,\ldots,a_n$ 是唯一的标量，使得

$$x=\sum_{i=1}^{n}a_iu_i.$$

我们定义 x 相对于 β 的**坐标向量**，记为 $[x]_\beta$，为

$$[x]_\beta=\begin{pmatrix}a_1\\a_2\\\vdots\\a_n\end{pmatrix}.$$

Example

例如: 设 $V=P_2(R)$，且令 $\beta=\{1,x,x^2\}$ 是 V 的标准有序基。若 $f(x)=4+6x-7x^2$，则

$$[f]_\beta=\begin{pmatrix}4\\6\\-7\end{pmatrix}.$$

Definition

用上述记号，我们称由 $A_{ij}=a_{ij}$ 定义的 $m\times n$ 矩阵 A 为从向量空间 V→W 的线性变换 T 在 V 的有序基 β 和 W 的 γ 下的**矩阵表示**，并记为 $A=[T]_\beta^\gamma$。

其中 A 的第 j 列正是 $[T(v_j)]_\gamma$。即 V 的有序基 β 中的每个基向量 v，在作用 T 后

得到的在 W 中的向量 w，再把 w 用 W 中的有序基 γ 线性表示后找出坐标，竖直地排列起来。

即：

$$A = \begin{pmatrix} | & | & & | \\ [T(v_1)]_\gamma & [T(v_2)]_\gamma & \cdots & [T(v_n)]_\gamma \\ | & | & & | \end{pmatrix}.$$

若 $V = W$ 且 $\beta = \gamma$，则记为 $A = [T]_\beta$。

这个概念可能在理解上存在困难，所以我们将提供两个例子：

Example

例 1：设 $T : \mathbb{R}^2 \to \mathbb{R}^3$ 是由

$$T(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1 - 4a_2)$$

定义的线性变换。令 β 和 γ 分别是 $\mathbb{R}^2$ 和 $\mathbb{R}^3$ 的标准有序基。

故：

$$T(1, 0) = (1, 0, 2) = 1e_1 + 0e_2 + 2e_3$$

且

$$T(0, 1) = (3, 0, -4) = 3e_1 + 0e_2 - 4e_3.$$

因此

$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}.$$

Example

例 2：设 $T : P_3(R) \to P_2(R)$ 是由 $T(f(x)) = f'(x)$ 定义的线性变换。令 β 和 γ 分别是 $P_3(R)$ 和 $P_2(R)$ 的标准有序基。

则:

$$T(1) = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2,$$
$$T(x) = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2,$$
$$T(x^2) = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2,$$
$$T(x^3) = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2.$$

所以:

$$[T]_\beta^\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

1.12.3 矩阵对向量的作用

在刚刚的小节中，我们已经明白了将 $T : \mathbb{R}^n \to \mathbb{R}^m$ 用矩阵表达的形式，那么 T 对某个向量的作用具体是怎么样呢?

我们给出如下的公式:

$$[T(v)]_\gamma = [T]_\beta^\gamma \cdot [v]_\beta$$

这一等式将抽象的线性变换 T 转化为具体的矩阵乘法运算，实现了线性变换的数值化表示。

证明. 1. 对于系数矩阵 $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$（共 n 个，$\forall v \in V_n$ 须 n 个系数确定），与 V_n 中的标准有序基 $\{e_1, e_2, \ldots, e_n\}$ 线性组合。可得 V_n 中任意一个向量 v，即 $\forall v \in V_n$，

$$v = \sum_{i=1}^{n} x_i \cdot e_i$$

称 $x = [v]_\beta = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (x_1, x_2, \ldots, x_n)^T$（列向量）。

2. 注意到每一个 $e_j \in V_n$，在 W_m 中须 m 个参数确定（$\because \dim W_m = m$，$\therefore$ 每个 $T(e_j) = \sum_{i=1}^m a_{ij} \cdot w_i$，$a_{ij} \in \mathbb{R}$）。

3. 用 T 把 V_n 中的 v 映射到 W_m 并用 w_i 表示。

$$T(v) = T\left(\sum_{i=1}^{n} x_i e_i\right) = \sum_{i=1}^{n} x_i T(e_i) = x_1 T(e_1) + x_2 T(e_2) + \cdots + x_n T(e_n).$$

等价于以下式子:

$$x_1\begin{pmatrix}a_{11}\\a_{21}\\\vdots\\a_{m1}\end{pmatrix}+x_2\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{m2}\end{pmatrix}+\cdots+x_n\begin{pmatrix}a_{1n}\\a_{2n}\\\vdots\\a_{mn}\end{pmatrix}=\begin{pmatrix}x_1a_{11}+x_2a_{12}+\cdots+x_na_{1n}\\\vdots\\x_1a_{m1}+\cdots+x_na_{mn}\end{pmatrix}\in\mathbb{R}^m$$

进一步，可写为矩阵形式：

$$\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\begin{pmatrix}\sum_{j=1}^n a_{1j}x_j\\\sum_{j=1}^n a_{2j}x_j\\\vdots\\\sum_{j=1}^n a_{mj}x_j\end{pmatrix}\in\mathbb{R}^m.$$

其中 $m\times n$ 矩阵 $[T]_\beta^\gamma=([T(e_1)]_\gamma\ [T(e_2)]_\gamma\ \cdots\ [T(e_n)]_\gamma)$。

□

Example

设 $T:\mathbb{R}^2\to\mathbb{R}^3$ 是由

$$T(a_1,a_2)=(a_1+3a_2,0,2a_1-4a_2)$$

定义的线性变换。令

$$[T]_\beta^\gamma=\begin{pmatrix}1&3\\0&0\\2&-4\end{pmatrix}.$$

对于向量 $v=(4,5)\in\mathbb{R}^2$，其坐标向量为 $[v]_\beta=\begin{pmatrix}4\\5\end{pmatrix}$。根据矩阵表示公式：

$$[T(v)]_\gamma=[T]_\beta^\gamma\cdot[v]_\beta=\begin{pmatrix}1&3\\0&0\\2&-4\end{pmatrix}\begin{pmatrix}4\\5\end{pmatrix}=\begin{pmatrix}1\cdot4+3\cdot5\\0\cdot4+0\cdot5\\2\cdot4+(-4)\cdot5\end{pmatrix}=\begin{pmatrix}19\\0\\-12\end{pmatrix}.$$

因此，$T(4,5)=(19,0,-12)$。

对于向量 $u=(1,3)\in\mathbb{R}^2$，其坐标向量为 $[u]_\beta=\begin{pmatrix}1\\3\end{pmatrix}$。根据矩阵表示公式：

$$[T(u)]_\gamma=[T]_\beta^\gamma\cdot[u]_\beta=\begin{pmatrix}1&3\\0&0\\2&-4\end{pmatrix}\begin{pmatrix}1\\3\end{pmatrix}=\begin{pmatrix}1\cdot1+3\cdot3\\0\cdot1+0\cdot3\\2\cdot1+(-4)\cdot3\end{pmatrix}=\begin{pmatrix}10\\0\\-10\end{pmatrix}.$$

因此，$T(1,3)=(10,0,-10)$。

Example

设 $T: P_3(\mathbb{R})\to P_2(\mathbb{R})$ 是由 $T(f(x))=f'(x)$ 定义的线性变换。令 β 和 γ 分别是 $P_3(\mathbb{R})$ 和 $P_2(\mathbb{R})$ 的标准有序基

$$[T]_\beta^\gamma=\begin{pmatrix}0&1&0&0\\0&0&2&0\\0&0&0&3\end{pmatrix}.$$

对于多项式 $p(x)=3x+7\in P_3(\mathbb{R})$，其坐标向量为 $[p]_\beta=\begin{pmatrix}7\\3\\0\\0\end{pmatrix}$。

根据矩阵表示公式：

$$[T(p)]_\gamma=[T]_\beta^\gamma\cdot[p]_\beta=\begin{pmatrix}0&1&0&0\\0&0&2&0\\0&0&0&3\end{pmatrix}\begin{pmatrix}7\\3\\0\\0\end{pmatrix}=\begin{pmatrix}0\cdot7+1\cdot3+0\cdot0+0\cdot0\\0\cdot7+0\cdot3+2\cdot0+0\cdot0\\0\cdot7+0\cdot3+0\cdot0+3\cdot0\end{pmatrix}=\begin{pmatrix}3\\0\\0\end{pmatrix}.$$

因此，$T(3x+7)=3$。

对于多项式 $q(x)=2x^2+3x+10\in P_3(\mathbb{R})$，其坐标向量为 $[q]_\beta=\begin{pmatrix}10\\3\\2\\0\end{pmatrix}$。根据矩阵表示公式：

$$[T(q)]_\gamma=[T]_\beta^\gamma\cdot[q]_\beta=\begin{pmatrix}0&1&0&0\\0&0&2&0\\0&0&0&3\end{pmatrix}\begin{pmatrix}10\\3\\2\\0\end{pmatrix}=\begin{pmatrix}0\cdot10+1\cdot3+0\cdot2+0\cdot0\\0\cdot10+0\cdot3+2\cdot2+0\cdot0\\0\cdot10+0\cdot3+0\cdot2+3\cdot0\end{pmatrix}=\begin{pmatrix}3\\4\\0\end{pmatrix}.$$

因此，$T(2x^2+3x+10)=3+4x$。

1.13 线性变换的叠加和复合与矩阵的加法与乘法

1.13.1 矩阵的加法 · 线性变换的叠加

设 V 和 W 是有限维向量空间，β 和 γ 分别是它们的有序基，$T,U:V\to W$ 是线性变换。则：

(a) $[T+U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$

(b) 对于任意标量 a，$[aT]_\beta^\gamma = a[T]_\beta^\gamma$

证明：设 $\beta = \{v_1, v_2, \ldots, v_n\}$，$\gamma = \{w_1, w_2, \ldots, w_m\}$。存在唯一标量 a_{ij}，b_{ij}（$1 \le i \le m$，$1 \le j \le n$），使得

$$T(v_j) = \sum_{i=1}^m a_{ij} w_i, \quad U(v_j) = \sum_{i=1}^m b_{ij} w_i.$$

进而

$$(T+U)(v_j) = \sum_{i=1}^m (a_{ij} + b_{ij}) w_i,$$

可得

$$([T+U]_\beta^\gamma)_{ij} = a_{ij} + b_{ij} = ([T]_\beta^\gamma + [U]_\beta^\gamma)_{ij},$$

证明了 (a)，(b) 证明类似。

接下来，我们通过一个例子来理解一下矩阵的加法：$T : \mathbb{R}^2 \to \mathbb{R}^3$，$U : \mathbb{R}^2 \to \mathbb{R}^3$ 为线性变换，定义为

$$T(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1 - 4a_2), \quad U(a_1, a_2) = (a_1 - a_2, 2a_1, 3a_1 + 2a_2).$$

β，γ 为 $\mathbb{R}^2$，$\mathbb{R}^3$ 标准有序基。已知

$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}, \quad [U]_\beta^\gamma = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 2 \end{pmatrix}.$$

计算得

$$(T+U)(a_1, a_2) = (2a_1 + 2a_2, 2a_1, 5a_1 - 2a_2),$$

$$[T+U]_\beta^\gamma = \begin{pmatrix} 2 & 2 \\ 2 & 0 \\ 5 & -2 \end{pmatrix},$$

即 $[T+U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$，示例验证定理 2.8。

1.13.2 矩阵的乘法 · 线性变换的复合

有了上面的讨论与前一节的铺垫，我们自然可以引出矩阵的乘法

设 A 是一个 $m \times n$ 矩阵，B 是一个 $n \times p$ 矩阵。我们定义 A 与 B 的乘积，记为 AB，是一个 $m \times p$ 矩阵，满足

$$(AB)_{ij} = \sum_{k=1}^n A_{ik} B_{kj} \quad \text{对于} 1 \le i \le m,\ 1 \le j \le p.$$

我们将此线性变换的角度来进行解释与证明：

设 $T: V \to W$ 和 $U: W \to Z$ 为线性变换，且令 $A = [U]_\beta^\gamma$，$B = [T]_\alpha^\beta$，其中 $\alpha = \{v_1, v_2, \dots, v_n\}$，$\beta = \{w_1, w_2, \dots, w_m\}$，$\gamma = \{z_1, z_2, \dots, z_p\}$ 分别是 V、W 和 Z 的有序基。

而由矩阵乘法的几何意义我们知道：A 的列与 B 的行数量必须一致。

我们想要定义两个矩阵的乘积 AB，使得 $AB = [UT]_\alpha^\gamma$。考虑矩阵 $[UT]_\alpha^\gamma$。对于 $1 \le j \le n$，我们有：

$$\begin{aligned}(UT)(v_j) = U(T(v_j)) = U\left(\sum_{k=1}^{m} B_{kj} w_k\right) &= \sum_{k=1}^{m} B_{kj} U(w_k) \\ &= \sum_{k=1}^{m} B_{kj}\left(\sum_{i=1}^{p} A_{ik} z_i\right) = \sum_{i=1}^{p}\left(\sum_{k=1}^{m} A_{ik} B_{kj}\right) z_i \\ &= \sum_{i=1}^{p} C_{ij} z_i\end{aligned}$$

其中

$$C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}$$

所以，综合上面的讨论，我们总结出：V、W、Z 为有限维向量空间，有序基分别为 α、β、γ ，$T: V \to W$，$U: W \to Z$ 为线性变换，则

$$[UT]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta$$

下面简要介绍以下矩阵乘法的 2 条重要性质：

矩阵乘积的转置性质：

若 A 是 $m \times n$ 矩阵，B 是 $n \times p$ 矩阵，则 $(AB)^t = B^t A^t$ ，推导过程为：

$$(AB)_{ij}^t = (AB)_{ji} = \sum_{k=1}^{n} A_{jk} B_{ki}, \quad (B^t A^t)_{ij} = \sum_{k=1}^{n} (B^t)_{ik} (A^t)_{kj} = \sum_{k=1}^{n} B_{ki} A_{jk}$$

同阶对角阵的乘法：

设两个 n 阶对角阵 A 和 B,其主对角线元素分别为 $a_{11}, a_{22}, \dots, a_{nn}$ 和 $b_{11}, b_{22}, \dots, b_{nn}$,

即：$A = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$，$B = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & b_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{nn} \end{pmatrix}$

乘法规则乘积仍为对角阵，其主对角线元素为对应位置元素的乘积：

$$AB=\begin{pmatrix} a_{11}b_{11} & 0 & \cdots & 0 \\ 0 & a_{22}b_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn}b_{nn} \end{pmatrix}$$

但是，我们也要注意到矩阵乘法和一般乘法的**特殊性**：

以下是对矩阵乘法反直觉特点的提取及举例：

1. 没有交换律： 矩阵乘法一般不满足交换律，即 $AB \neq BA$，当 $AB=BA$ 时，矩阵 A 和 B 可交换。

例子：设 $A=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ ，$B=\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ 。$AB=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ ，$BA=\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ ，显然 $AB \neq BA$ 。

【Remark】：

同阶对角阵相乘满足交换性, 交换律成立：$AB=BA$,

2. 没有消去律: 若 $AC=BC$ ，不能得出 $A=B$ 。

例子: 设 $A=\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B=\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $C=\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ 。$AC=\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $BC=\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ ，此时 $AC=BC$ ，但 $A \neq B$ 。

3. 相乘得零但可能无零因子: 若 $AB=0$ ，不能得出 $A=0$ 或 $B=0$ 。

例子: 设 $A=\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, $B=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ 。$AB=\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, 这里 $AB=0$ ，但 $A \neq 0$ 且 $B \neq 0$ 。

1.13.3 矩阵乘法的向量理解

左乘 A**：** A 左乘列向量，即 $A\cdot$ 列向量 = 列向量。

用符号表示为：

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=x\cdot \mathrm{Col}_1+y\cdot \mathrm{Col}_2+z\cdot \mathrm{Col}_3$$

(Col_1、Col_2、Col_3 分别表示矩阵的列向量)。

我们做一点点变换：我们只取左边矩阵的其中的一行：

$$\begin{pmatrix} a_{i1} & a_{i2} & a_{i3} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = x \cdot \mathrm{Col}_1 + y \cdot \mathrm{Col}_2 + z \cdot \mathrm{Col}_3$$

在这里：$\mathrm{Col}_1 = a_{i1}$ 同样的：$\mathrm{Col}_2 = a_{i2}$，$\mathrm{Col}_3 = a_{i3}$.

同样的，我们有：

$$\begin{pmatrix} a_{i1} & a_{i2} & a_{i3} \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = x_2 \cdot a_{i1} + y_2 \cdot a_{i2} + z_2 \cdot a_{i2}$$

以此类推到每一个 x_j，当右边 j 个 x 排成一列时：就有了右乘：

右乘 A：A 右乘行向量，即行向量 $\cdot A =$ 行向量。

用符号表示为：

$$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} ---\mathrm{row}_1--- \\ ---\mathrm{row}_2--- \\ ---\mathrm{row}_3--- \end{pmatrix} = x_1 \cdot \mathrm{row}_1 + x_2 \cdot \mathrm{row}_2 + x_3 \cdot \mathrm{row}_3$$

我们将通过两个例题来加深理解（例子要修改!!!!!!!)：

- **左乘例题**：求矩阵 A，满足

$$A \cdot \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & -3 \\ 0 & 4 & 1 \end{pmatrix}$$

分析过程：因为 A 在左边，考虑 $\begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix}$ 的行向量 r_1、r_2、r_3，r_1 不变，$r_1 \times (-3) \to r_2$ 上，r_3 不变，得出

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

本质上就是排列起来的三个行向量形成矩阵，然后经过 LA 的变换，变成了另外三个排列起来的行向量。

- **右乘例题**：求矩阵 B，满足

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -3 \\ 0 & 4 & 1 \end{bmatrix} \cdot B = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{pmatrix}$$

分析过程：因为 B 在右边，考虑 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -3 \\ 0 & 4 & 1 \end{bmatrix}$ 的列向量 c_1、c_2、c_3。

- 观察目标矩阵的第一列：c_1 不变。
- 观察目标矩阵的第二列：$2c_2 + c_3$（即原矩阵第二列乘以 2 加上第三列）。
- 观察目标矩阵的第三列：$c_1 - c_3$（即原矩阵第一列减去第三列）。

得出

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

综合以上内容，我们不难得到向量层面的理解：

- 左乘矩阵：影响被乘矩阵的行向量，对应行变换。
- 右乘矩阵：影响被乘矩阵的列向量，对应列变换。
- 通过分析行/列向量的线性组合，可以直接构造出对应的变换矩阵。

1.14 可逆与同构

1.14.1 单射与满射

对于 $T : V \to W$：
(a) **单射（Injection）**，或称一对一（One-to-one）：若 $T(x) = T(y) \implies x = y$。
观察到：定义域里面除了 0，没有其它元素会被映射成 0，即 $N(T) = 0$。
我们可以进一步的论证：

Theorem

若 $T(v) = 0$，则 $v = 0$ 与单射等价。
即：T 是单射当且仅当 $\ker(T) = \{0\}$。

证明. **步骤 1：** 证明"若 $T(v) = 0$ 则 $v = 0 \implies T$ 是单射"设 $u, v \in V$ 满足 $T(u) = T(v)$。由线性映射的性质，$T(u - v) = T(u) - T(v) = 0$。

根据题设"若 $T(w) = 0$ 则 $w = 0$"，可得 $u - v = 0$，即 $u = v$。因此，T 是单射。

步骤 2： 证明"T 是单射 $\implies$ 若 $T(v) = 0$ 则 $v = 0$"设 T 是单射，且 $v \in V$ 满足 $T(v) = 0$。

注意到线性映射的基本性质：$T(0_V) = 0_W$（0_V 是 V 的零向量，0_W 是 W 的零向量）。

因此 $T(v)=T(0_V)$。又因为 T 是单射，故 $v=0_V$，即"若 $T(v)=0$ 则 $v=0$"。综上，"若 $T(v)=0$ 则 $v=0$" 与 "T 是单射" 互为充要条件，即二者等价。 □

同时，单射还有如下的性质：

Property

单射与线性无关性： 若 T 为单射，则任一线性无关组 $\{x_1,x_2,\dots,x_n\}\subset V$ 的像 $\{T(x_1),T(x_2),\dots,T(x_n)\}$ 仍线性无关。

证明. 证明：假设 $\sum a_iT(x_i)=0$，则 $T\left(\sum a_ix_i\right)=0$。由 T 单射，$\sum a_ix_i=0$，故 $a_i=0$，矛盾。 □

(b) **满射（Surjection）**，或称映上（Onto）：若对 W 中每个向量 w_i，至少存在 $\alpha\in V$ 使 $T(\alpha)=w_i$。

Example

$T(f(x))=f'(x)$ 是满射非单射；$T(f(x))=\int_0^x f(t)dt$ 是单射非满射。

1.14.2 逆与可逆

Definition

对于线性变换 $T:V\to W$，若存在另一个线性变换 $U:W\to V$，使得 $TU=I_W$ 且 $UT=I_V$，则称 U 为 T 的逆变换（或 T 的逆，记为 T^{-1}）。若 T 存在逆，则称 T 可逆（Invertible）。

例如：考虑二维向量空间 $\mathbb{R}^2$，定义线性变换 $T:\mathbb{R}^2\to\mathbb{R}^2$，对于任意向量 $\begin{pmatrix}x\\y\end{pmatrix}\in\mathbb{R}^2$，$T\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}2x+y\\x+y\end{pmatrix}$。

令 $U\begin{pmatrix}m\\n\end{pmatrix}=\begin{pmatrix}m-n\\-m+2n\end{pmatrix}$：容易验证：
$TU=I_{\mathbb{R}^2}$ 且 $UT=I_{\mathbb{R}^2}$。
所以 T 可逆，且逆为 U；
但是，我们又要介绍一对常见的误区。

【Remark】:

设 $T: P_2 \to P_3$，$T(f(x)) = \int_0^x f(t)dt$；$U: P_3 \to P_2$，$U(f(x)) = f'(x)$。
计算得 $(TU)(x) = f(x) - f(0) \neq I_{P_3}(f(x))$，$UT(f(x)) = f(x) = I_{P_2}(f(x))$，故 T、U 并非可逆关系。

Corollary & Secondary Conclusion

如果线性变换 T 可逆，那么**逆唯一**。

证明. □

Theorem

对于线性变换 $T: V \to W$，$\left([T]_\beta^\gamma\right)^{-1} = [T^{-1}]_\gamma^\beta$（反向）。

证明. □

接下来，我们将讨论：**逆的必要条件**

设 $T: V \to W$ 是有限维向量空间之间的线性映射，其矩阵表示为 $m \times n$ 矩阵 A。若 T 可逆，则：

1. V 和 W 的维数相等，即 $\dim(V) = \dim(W) = n$；
2. A 必为 $n \times n$ 方阵，且 $\mathrm{rank}(A) = n$（即 A 满秩）。

1.14.3 同构与同构映射

同构（Isomorphic）与同构映射（Isomorphism）是一种相当重要的映射，其背后的思想既有坐标化的思想，又有转换的思想，通过找到统一性与普遍性，连结特殊性。

Definition

若存在可逆的线性变换 $T: V \to W$，则称 V 和 W 同构，其中 T 称为为同构映射。

其具有以下是性质：

(a) 若 V 和 W 同构，则 $\dim(V) = \dim(W)$；反之，若 $\dim(V) = \dim(W)$，则 V 和 W 同构。

(b) 若 V_0 是 V 的子空间，则 $\dim(V_0) = \dim(T(V_0))$。

并且，我们可以证明以下论断：

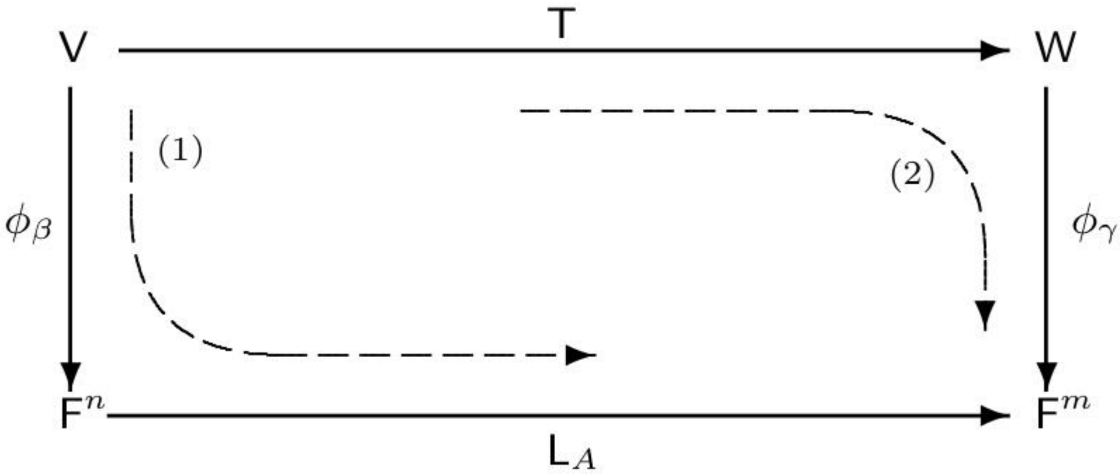


图 1.2: 同构映射概念图

Claim

若 V 和 W 同构（$\dim(V) = \dim(W)$），则以下三条等价：
T 是一对一的；
T 是满射的；
$\dim V = \dim W = \gamma(T)$。

1.14.4 向量空间的标准表示相关内容

设 β 是数域 F 上 n 维向量空间 V 的一个有序基。关于 β 的 V 的表示是函数 $\phi_\beta : V \to F^n$ ，对每个 $x \in V$，定义为 $\phi_\beta(x) = [x]_\beta$ 。

我们现在考察一个具体的例子来理解这个概念：

设 $\beta = \{(1,0),(0,1)\}$ 且 $\gamma = \{(1,2),(3,4)\}$。容易看出 β 和 γ 是 $\mathbb{R}^2$ 的有序基。对于 $x = (1,-2)$，我们有 $\phi_\beta(x) = [x]_\beta = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ ，$\phi_\gamma(x) = [x]_\gamma = \begin{pmatrix} -5 \\ 2 \end{pmatrix}$ 。

让我们再进一步，设 V 和 W 分别是维数为 n 和 m 的向量空间，且设 $T : V \to W$ 是一个线性变换。定义 $A = [T]_\beta^\gamma$ ，其中 β 和 γ 分别是 V 和 W 的任意有序基。我们现在能够使用 ϕ_β 和 ϕ_γ 来研究线性变换 T 和 $L_A : F^n \to F^m$ 之间的关系。

首先考虑图。注意到有两个将 V 映射到 F^m 的线性变换复合：

1. 用 ϕ_β 将 V 映射到 F^n ，然后用 L_A 进行变换；这产生复合 $L_A\phi_\beta$ 。

2. 用 T 将 V 映射到 W ，然后用 ϕ_γ 进行变换，得到复合 $\phi_\gamma T$ 。

这两个复合由图中的虚线箭头表示。显然，我们可以得出 $L_A\phi_\beta = \phi_\gamma T$ ，即该图是“可交换的”。直观地说，这种关系表明，在通过 ϕ_β 和 ϕ_γ 分别将 V 和 W 与 F^n 和 F^m

等同之后，我们可以将 T 与 L_A “等同”。**此图允许我们将抽象向量空间上的运算转换为可以量化研究的，结构清晰的数域 F^n 和 F^m 上的矩阵运算。**

因此矩阵与线性变换的联系可见下图：

$$\begin{array}{ccc} v & \xrightarrow{T} & w \\ \beta\downarrow & & m\downarrow\gamma \\ [v]_\beta & \xrightarrow{[T]_\beta^\gamma} & [T(v)]_\gamma \end{array}$$

在这张图中，ϕ_β 直接作用在向量 v 上，将其变成可用矩阵表示的：$[x]_\beta$；ϕ_γ 也是同理

我们再来考察一个例子，加深对上面讨论的理解：

Example

对于线性变换 $T: P_3(R) \to P_2(R)$（$T(f(x)) = f'(x)$）。设 β 和 γ 分别是 $P_3(R)$ 和 $P_2(R)$ 的标准有序基，且设 $\phi_\beta: P_3(R) \to \mathbb{R}^4$ 和 $\phi_\gamma: P_2(R) \to \mathbb{R}^3$ 分别是 $P_3(R)$ 和 $P_2(R)$ 对应的标准表示。若 $A = [T]_\beta^\gamma$.

则

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

。

考虑多项式 $p(x) = 2 + x - 3x^2 + 5x^3$ 。我们证明 $L_A\phi_\beta(p(x)) = \phi_\gamma T(p(x))$ 。现在

$$L_A\phi_\beta(p(x)) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ -3 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ -6 \\ 15 \end{pmatrix}$$

。

但由于 $T(p(x)) = p'(x) = 1 - 6x + 15x^2$ ，我们有

$$\phi_\gamma T(p(x)) = \begin{pmatrix} 1 \\ -6 \\ 15 \end{pmatrix}$$

。

所以 $L_A\phi_\beta(p(x)) = \phi_\gamma T(p(x))$ 。

1.14.5 同构的三大基本定理

Theorem

第一同构定理（first isomorphism）：给定域 $\mathbb{F}$ 上的线性映射 $T: V \to W$，存在 $\mathbb{F}$-线性的自然投影 (natural projection)$p: V \to V/\ker(T)$ （$p(v) = \bar{v} := v + \ker(T)$，即将 v 映射到其所在的陪集），使得 $T = \bar{T} \circ p$，其中 $\bar{T}: V/\ker(T) \to \mathrm{Im}(T)$ 是线性同构，定义为 $\bar{T}(\bar{v}) = T(v)$。

证明. 1. 映射 $p: V \to V/\ker(T)$，$v \mapsto \bar{v}$ 是线性且满射的。

首先验证线性性：对任意 $v, w \in V$，$p(v+w) = (v+w)+\ker(T) = (v+\ker(T))+(w+\ker(T)) = p(v)+p(w)$；对任意 $a \in \mathbb{F}$，$p(av) = av+\ker(T) = a(v+\ker(T)) = a \cdot p(v)$。

再验证满射性：对商空间 $V/\ker(T)$ 中任意陪集 $\bar{v} = v + \ker(T)$，存在 $v \in V$ 使得 $p(v) = \bar{v}$，故 p 是满射。

2. 定义 $\bar{T}: V/\ker(T) \to \mathrm{Im}(T)$ 为 $\bar{T}(\bar{v}) = T(v)$。下证其良定义：若 $\bar{v} = \bar{w}$，则 $v - w \in \ker(T)$，由核的定义知 $T(v-w) = 0$，即 $T(v) = T(w)$，故 $\bar{T}$ 良定义。

3. $\bar{T}$ 是线性的：对任意 $a, b \in \mathbb{F}$，$\bar{v}, \bar{w} \in V/\ker(T)$，$\bar{T}(a\bar{v} + b\bar{w}) = \bar{T}(\overline{av + bw}) = T(av + bw) = aT(v) + bT(w) = a\bar{T}(\bar{v}) + b\bar{T}(\bar{w})$。

4. $\bar{T}$ 是单射的：若 $\bar{T}(\bar{v}) = 0$，则 $T(v) = 0$，故 $v \in \ker(T)$，从而 $\bar{v} = \ker(T)$（商空间零元），即 $\bar{T}$ 的核仅含零元，故单射。

5. $\bar{T}$ 是满射的：对任意 $w \in \mathrm{Im}(T)$，存在 $v \in V$ 使 $T(v) = w$，则 $\bar{T}(\bar{v}) = T(v) = w$，故满射。

因此，$\bar{T}$ 是同构且 $T = \bar{T} \circ p$。

□

【Remark】：

因此，由该定理可知：线性映射的“像空间”与“核的商空间”代数结构完全一致——所以，维度定理的逻辑自洽。

Theorem

第二同构定理（second isomorphism）：设 V 和 W 是域 $\mathbb{F}$ 上的向量子空间，则自然 $\mathbb{F}$-线性映射 $T: V \to (V+W)/W$ 可通过投影 $p: V \to V/(V \cap W)$ 分解，即 $T = \bar{T} \circ p$，其中 $\bar{T}: V/(V \cap W) \xrightarrow{\cong} (V+W)/W$，并且 $\bar{T}$ 是同构。

即：对两个 $\mathbb{F}$- 子空间 V, W，有同构 $V/(V \cap W) \cong (V+W)/W$。

证明. 1. **定义映射 T 并验证线性性与满射性**：定义 $T: V \to (V+W)/W$ 为 $T(v) = \bar{v} = v + W$（因为 $v \in V \subset V + W$，所以 $v \in V + W$，满足商空间元素的定义）。

- **线性性**：对任意 $v_1, v_2 \in V$，$a, b \in \mathbb{F}$，

  $$T(av_1 + bv_2) = (av_1 + bv_2) + W = a(v_1 + W) + b(v_2 + W) = aT(v_1) + bT(v_2),$$

  故 T 是线性映射。

- **满射性**：对任意陪集 $\bar{x} = x + W \in (V+W)/W$，由 $V+W$ 的定义，$x = v + w$（其中 $v \in V$，$w \in W$）。此时 $\bar{x} = (v+w) + W = v + W = T(v)$，故存在 $v \in V$ 使得 $T(v) = \bar{x}$，即 T 是满射。

2. **计算 T 的核** $\ker(T)$：核的定义为 $\ker(T) = \{v \in V \mid T(v) = 0_{(V+W)/W}\}$，其中 $0_{(V+W)/W} = W$（商空间的零元是 W 本身）。因此，$T(v) = W \iff v + W = W \iff v \in W$。又因 $v \in V$，故 $\ker(T) = \{v \in V \mid v \in W\} = V \cap W$。

3. **应用第一同构定理**：第一同构定理指出：若 $T : V \to W'$ 是满射线性映射，则 $V/\ker(T) \cong \mathrm{Im}(T)$。此处 T 是满射，且 $\mathrm{Im}(T) = (V+W)/W$，$\ker(T) = V \cap W$，因此：

   $$V/(V \cap W) \cong (V+W)/W.$$

   记此同构为 $\bar{T} : V/(V \cap W) \to (V+W)/W$，则自然有 $T = \bar{T} \circ p$（其中 $p : V \to V/(V \cap W)$ 是投影映射 $v \mapsto v + (V \cap W)$）。

□

Theorem

第三同构定理 (third isomorphism)：设 $U \subset W \subset V$ 是域 $\mathbb{F}$ 上的向量子空间，则自然 $\mathbb{F}$-线性映射 $T : V \to (V/U)/(W/U)$ 可通过投影 $p : V \to V/W$ 分解，即 $T = \bar{T} \circ p$，其中 $\bar{T} : V/W \xrightarrow{\cong} (V/U)/(W/U)$，并且 $\bar{T}$ 是同构。

即：对嵌套子空间 $U \subset W \subset V$，有同构 $(V/U)/(W/U) \cong V/W$。

证明. 1. **定义辅助映射 ϕ 并验证线性性与满射性**：定义 $\phi : V/U \to (V/U)/(W/U)$ 为 $\phi(\bar{v}_U) = \overline{\bar{v}_U}_{W/U}$（即 $\bar{v}_U = v + U$ 在商空间 $(V/U)/(W/U)$ 中的陪集）。

- **线性性**：对任意 $\bar{v}_U, \bar{w}_U \in V/U$，$a, b \in \mathbb{F}$，

  $$\phi(a\bar{v}_U + b\bar{w}_U) = \phi(\overline{av+bw}_U) = \overline{\overline{av+bw}_U}_{W/U} = a\overline{\bar{v}_U}_{W/U} + b\overline{\bar{w}_U}_{W/U} = a\phi(\bar{v}_U) + b\phi(\bar{w}_U),$$

  故 ϕ 是线性映射。

- **满射性**：对任意陪集 $\bar{\bar{x}}_{W/U} \in (V/U)/(W/U)$，其中 $\bar{x}_U = x + U \in V/U$，显然 $\phi(\bar{x}_U) = \bar{\bar{x}}_{W/U}$，故 ϕ 是满射。

2. **计算 ϕ 的核** $\ker(\phi)$：核的定义为 $\ker(\phi) = \{\bar{v}_U \in V/U \mid \phi(\bar{v}_U) = 0_{(V/U)/(W/U)}\}$，其中 $0_{(V/U)/(W/U)} = W/U$（商空间的零元是 W/U 本身）。因此，$\phi(\bar{v}_U) = W/U \iff$

$\bar{v}_U + W/U = W/U \iff \bar{v}_U \in W/U$。即 $\ker(\phi) = \{\bar{v}_U \in V/U \mid \bar{v}_U \in W/U\} = W/U$。

3. **应用第一同构定理**：第一同构定理指出：若 $\phi : V/U \to (V/U)/(W/U)$ 是满射线性映射，则 $(V/U)/\ker(\phi) \cong \mathrm{Im}(\phi)$。此处 ϕ 是满射，且 $\mathrm{Im}(\phi) = (V/U)/(W/U)$，$\ker(\phi) = W/U$，因此：

$$(V/U)/(W/U) \cong (V/U)/(W/U) \implies (V/U)/(W/U) \cong V/W$$

（通过自然对应，商空间的“商”可转化为原空间的商 V/W）。记此同构为 $\bar{T} : V/W \to (V/U)/(W/U)$，则自然有 $T = \bar{T} \circ p$（其中 $p : V \to V/W$ 是投影映射 $v \mapsto v + W$）。

□

因此，三大定理可以由下面的逻辑表现清楚：

- 第一定理：用线性映射的“核”分解定义域，建立商空间到像空间的同构，有线性映射的“像空间”与“核的商空间”代数结构完全一致。
- 第二定理：联系“子空间的交”与“子空间的和的商空间”的结构。
- 第三定理：联系“商空间的商空间”与“大商空间”的结构。

1.14.6 同构空间与对偶空间

Definition

同态空间（Homomorphism space ）：设 V 和 W 是两个 $\mathbb{F}$-向量空间，从 V 到 W 的同态空间（简称 Hom space）定义为：

$$\mathrm{Hom}_{\mathbb{F}}(V, W) := \{\text{所有 } \mathbb{F}\text{-线性映射 } T : V \to W\}.$$

若 $V = W$，也记作 $\mathrm{End}_{\mathbb{F}}(V)$。

我们知道：从 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的线性映射 T，可唯一由一个 $m \times n$ 矩阵 A 表示（满足 $T(\mathbf{x}) = A\mathbf{x}$，$\mathbf{x} \in \mathbb{F}^n$）。所以我们可以考察以下同态空间及其性质：

Theorem

所有从 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的线性映射 T 构成的空间 $\mathrm{Hom}_{\mathbb{F}}(\mathbb{F}^n, \mathbb{F}^m)$，与“所有 $m \times n$ 矩阵”构成的空间 $M_{m\times n}(\mathbb{F})$ **同构**。即：$\mathrm{Hom}_{\mathbb{F}}(\mathbb{F}^n, \mathbb{F}^m) \cong M_{m\times n}(\mathbb{F})$；

证明. □

Corollary & Secondary Conclusion

当 V 是有限维 $\mathbb{F}$- 向量空间时，$\dim_{\mathbb{F}} \mathrm{Hom}_{\mathbb{F}}(V,W) = \dim_{\mathbb{F}}(V) \cdot \dim_{\mathbb{F}}(W)$。

证明. 线性映射的“自由度”由基的像唯一确定：

设 V 的基为 $\{v_1, v_2, \dots, v_n\}$ $(n = \dim V)$, W 的基为 $\{w_1, w_2, \dots, w_m\}$ $(m = \dim W)$。

对任意线性映射 $T : V \to W$，它的行为由 $T(v_1), T(v_2), \dots, T(v_n)$ 唯一决定（因为 V 中任意向量 $v = a_1v_1 + \cdots + a_nv_n$，都有 $T(v) = a_1T(v_1) + \cdots + a_nT(v_n)$）。每个 $T(v_i)$ 可表示为 W 基的线性组合：$T(v_i) = b_{i1}w_1 + b_{i2}w_2 + \cdots + b_{im}w_m$ （$b_{ij} \in \mathbb{F}$）。

这意味着，确定 T 需要选择 n 个“W 中的向量”，每个向量有 m 个独立的标量系数。

因此，所有线性映射的“自由度总数”是 $n \times m$（即 $\dim V \times \dim W$），对应矩阵空间 $M_{m\times n}(\mathbb{F})$ 的维数（$m \times n$ 个独立标量）。 □

Corollary & Secondary Conclusion

$\mathrm{Hom}_{\mathbb{F}}\left(\bigoplus_{i\in I} \mathbb{F}, \mathbb{F}^m\right) \cong \prod_{i\in I} \mathbb{F}^m$。

证明. 从 $\bigoplus_{i\in I} \mathbb{F}$ 到 $\mathbb{F}^m$ 的线性映射 T，由它在“直和的基元素”（每个分量的单位元，如第 i 个分量为 1、其余为 0 的元组）上的作用唯一确定。

对每个 $i \in I$，T 在第 i 个基元素上的像属于 $\mathbb{F}^m$；而所有这样的线性映射，等价于“为每个 $i \in I$ 选取一个 $\mathbb{F}^m$ 中的元素”——这正是直积 $\prod_{i\in I} \mathbb{F}^m$ 的定义（直积中元素是“对每个 $i \in I$ 取一个 $\mathbb{F}^m$ 元素”，无“有限非零”限制）。 □

同样的，我们可以得到：

Corollary & Secondary Conclusion

$\mathrm{Hom}_{\mathbb{F}}\left(\bigoplus_{i\in I} \mathbb{F}, \bigoplus_{j\in J} \mathbb{F}\right) \cong \prod_{i\in I}\left(\bigoplus_{j\in J} \mathbb{F}\right)$。

接下来我们讲介绍：**对偶空间**。

Definition

设 V 是 $\mathbb{F}$- 向量空间，**对偶空间** V^* 定义为“从 V 到数域 $\mathbb{F}$ 的所有线性映射构成的空间”，即：$V^* := \mathrm{Hom}_{\mathbb{F}}(V, \mathbb{F})$

Lemma

若 $\dim_{\mathbb{F}}(V) < \infty$（$V$ 是有限维向量空间），则：$\dim_{\mathbb{F}}(V^*) = \dim_{\mathbb{F}}(V)$

即：$V^* \simeq V$

对偶基（Dual Basis）：为了显式构造 V^* 的基，并建立 V^* 与 V 之间的同构，引入对偶基：

Definition

设 $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ 是 V 的一个基，定义 V^* 中的线性函数组 $\mathcal{B}^* = \{v_1^*, v_2^*, \ldots, v_n^*\}$，其中每个 $v_i^* : V \to \mathbb{F}$ 满足：$v_i^*(v_j) = \delta_{ij}$ （δ_{ij}为克罗内克函数，$i = j$时为 1，否则为 0）此时，$\mathcal{B}^*$ 是 V^* 的一个基（称为 $\mathcal{B}$ 的对偶基）。

通过线性延拓，可诱导出 V 与 V^* 之间的同构（例如 $v_i \mapsto v_i^* \mapsto v_i$）。

Example

$\mathbb{R}^n$ 是实数域上的 n 维向量空间（元素为 n 维实向量 $x = (x_1, x_2, \ldots, x_n)^T$）。其对偶空间 $(\mathbb{R}^n)^*$ 是所有从 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的线性泛函的集合。

原空间 $\mathbb{R}^n$ 的标准基 $\{e_1, e_2, \ldots, e_n\}$，就像 n 维空间里的坐标轴：

e_1 是 “x_1 轴” 的单位向量，e_2 是 “x_2 轴” 的单位向量，⋯，e_n 是 “x_n 轴” 的单位向量。

任意向量 $x = (x_1, x_2, \ldots, x_n)^T$，都能拆成“各坐标轴分量的组合”：$x = x_1e_1 + x_2e_2 + \cdots + x_ne_n$。

而对偶基可以理解为 “刻度读取器”。**对偶空间里的元素是线性泛函**（把向量变成数的 “函数”）。对偶基 $\{e_1^*, e_2^*, \ldots, e_n^*\}$，就像 “读取各坐标轴刻度的工具”：$e_i^*$ 是专门“读第 i 个坐标轴刻度” 的工具。当你用 e_i^* 去 “读” 向量 x 时，得到的就是 x 在第 i 个坐标轴上的分量 x_i，即 $e_i^*(x) = x_i$。

对偶基对 “精准读取坐标轴” 的验证：$e_i^*(e_j) = \delta_{ij}$，其实是说：用 “读第 i 个坐标轴” 的工具 e_i^*，去读 “第 j 个坐标轴的单位向量 e_j”，结果是什么？

如果 $i = j$：读的是 “第 i 个坐标轴的单位向量”，那它在第 i 个坐标轴上的刻度当然是 1，所以 $e_i^*(e_i) = 1$。

如果 $i \neq j$：读的是 “第 j 个坐标轴的单位向量”，它在第 i 个坐标轴上的刻度是 0，所以 $e_i^*(e_j) = 0$。

这就像 “读 x 轴刻度的工具，读 y 轴的单位向量时，得到的 x 轴刻度是 0；读 x 轴的单位向量时，得到的 x 轴刻度是 1”。

现在任取一个线性泛函 $f \in (\mathbb{R}^n)^*$，根据线性性，对 x 有：$f(x) = f(x_1e_1 + x_2e_2 + \cdots + x_ne_n)$ 由线性映射的 “可加性” 和 “齐次性”，展开得：$f(x) = x_1 \cdot f(e_1) + x_2 \cdot f(e_2) + \cdots + x_n \cdot f(e_n)$

那么对偶基如何表示这个线性泛函？对偶基 $\{e_1^*, e_2^*, \ldots, e_n^*\}$ 的定义是：对任意 $x = (x_1, \ldots, x_n)^T$，$e_i^*(x) = x_i$（即 e_i^* 是 “提取第 i 个分量” 的线性泛函）。令 $a_i = f(e_i)$ （a_i 是实数，因为 $f(e_i) \in \mathbb{R}$），则上式可改写为：$f(x) = a_1 \cdot e_1^*(x) + a_2 \cdot e_2^*(x) + \cdots + a_n \cdot e_n^*(x)$

由于上式对所有 $x \in \mathbb{R}^n$ 成立，因此作为线性泛函（映射），f 可表示为对偶基的线性组合：$f = a_1e_1^* + a_2e_2^* + \cdots + a_ne_n^*$

Example

【有限基】：$\mathbb{R}^2$：取 $n=2$，对偶基为 e_1^*, e_2^*，其中：$e_1^*(x_1,x_2)=x_1$（提取第一个分量），$e_2^*(x_1,x_2)=x_2$（提取第二个分量）。
现在有一个线性泛函 $f:\mathbb{R}^2\to\mathbb{R}$，满足：$f(1,0)=3$（即 $f(e_1)=3$），$f(0,1)=5$（即 $f(e_2)=5$）。根据上述推导，f 可表示为对偶基的线性组合：$f=3e_1^*+5e_2^*$.

【Remark】：

f 和 f（x）不一样！

验证：对任意 $x=(x_1,x_2)$，$f(x)=3e_1^*(x)+5e_2^*(x)=3x_1+5x_2$ 而由线性性，$f(x_1,x_2)=x_1f(1,0)+x_2f(0,1)=3x_1+5x_2$，完全一致。

Example

【无限基】：$\mathbb{F}[x]$ 是数域 $\mathbb{F}$ 上的所有多项式构成的向量空间（比如 $\mathbb{F}=\mathbb{R}$ 时，元素是 x^2+3x+5、$7x^3-2$ 等）。它的基为 $\mathcal{B}_\prime=\{1,x,x^2,x^3,\dots\}$

注意：这个基是**无限基**（因为多项式的次数可以任意高，基中包含无限多个单项式）。

对偶基的构造：“提取系数”的线性泛函对偶空间 $V^*=\mathrm{Hom}_{\mathbb{F}}(V,\mathbb{F})$ 是“所有从 $\mathbb{F}[x]$ 到 $\mathbb{F}$ 的线性映射（线性泛函）”的集合。我们为原基 $\mathcal{B}_\prime$ 中的每个单项式 x^i，构造一个线性泛函 $\delta_{x^i}\in V^*$，满足：$\delta_{x^i}(x^j)=\delta_{ij}$　直观来说，每个 δ_{x^i} 是“提取多项式某一项系数”的工具：δ_1（对应 $x^0=1$）：作用在多项式上，提取常数项。

-例如 $\delta_1(5x^3+2x+7)=7$，因为 $\delta_1(x^j)=\delta_{0j}$（$j=0$ 时为 1，$j\geq 1$ 时为 0）。δ_x（对应 x^1）：作用在多项式上，提取一次项系数。

-例如 $\delta_x(5x^3+2x+7)=2$，因为 $\delta_x(x^j)=\delta_{1j}$（$j=1$ 时为 1，其他为 0）。δ_{x^2}：提取二次项系数，δ_{x^3} 提取三次项系数，依此类推。

这样构造出的 $\mathcal{B}^*=\{\delta_1,\delta_x,\delta_{x^2},\delta_{x^3},\dots\}$ 就是原基 $\mathcal{B}_\prime$ 的对偶基。

Example

例如：组合 $3\delta_1+2\delta_x$：作用在多项式上，结果为“$3\times$ 常数项　$+2\times$ 一次项系数”。组合 $c_0\delta_1+c_1\delta_x+\cdots+c_n\delta_{x^n}$：仅涉及前 $n+1$ 个系数的线性组合。

无限维的特殊性：对偶基张成的空间是对偶空间的真子集

对偶空间 V^* 包含所有从 $\mathbb{F}[x]$ 到 $\mathbb{F}$ 的线性泛函，其中存在大量无法用“有限个对偶基元素组合”表示的泛函。

例子：“无限系数和”的线性泛函假设 $\mathbb{F}=\mathbb{R}$，定义线性泛函 $g:\mathbb{F}[x]\to\mathbb{R}$：对任意多项式 $p(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$，$g(p(x))=a_0+a_1+a_2+\cdots+a_n$（即“所有系数的和”）。g 是线性的：因为 $g(kp+q)=kg(p)+g(q)$（线性映射的可加性、齐次性）。

但 g 不在 $\mathrm{Span}_{\mathbb{F}}(\mathcal{B}^*)$ 中：因为 $\mathrm{Span}_{\mathbb{F}}(\mathcal{B}^*)$ 中的泛函只能涉及有限个系数的组合，而 g 需要"同时提取无限多个系数并求和"，无法用 $\mathcal{B}^*$ 中元素的有限线性组合表示。

因此，$\mathrm{Span}_{\mathbb{F}}(\mathcal{B}^*)$ 是 V^* 的真子集（即 $\mathrm{Span}_{\mathbb{F}}(\mathcal{B}^*) \subsetneq V^*$）。

有限维与无限维的核心差异: 在有限维空间（如 $\mathbb{R}^n$）中，对偶基的张成空间就是整个对偶空间（因为基是有限的，线性组合能覆盖所有线性泛函）。

但在无限维空间（如 $\mathbb{F}[x]$）中，对偶空间的"规模"远大于"原基对偶基的张成空间"——无限维空间的对偶空间无法被"原基的对偶基的有限组合"完全覆盖。

1.14.7 零化子及其性质

问题：哪些线性泛函 f 会'忽略' W 中的向量（$\forall w \in W, f(w) = 0$)?

其中**忽略**可以理解为：子空间 W 里的向量，在 f 这个"测量工具"下，测不出任何"数值"（结果都是 0）

Example

假设 V 是三维空间，W 是 xy 平面（所有 z 坐标为 0 的向量构成的子空间）。
如果有个泛函 $f(v)$ 是 v 的 z 坐标，那对 W 里的任意向量 w（z 坐标都是 0），都有 $f(w) = 0$。这说明：f 专门"测 z 方向"，而 W 里的向量"没 z 方向的分量"，所以被 f "忽略"了。

原空间 V 有"包含、直和、商空间"这些结构，对偶空间 V^* 的结构会和它们"反向对应"，相当于从"镜子里看原空间的结构"，用泛函的角度重新解释原空间。

于是我们可以导入零化子的概念及重要性质：

零化子的定义：

Definition

设 V 是域 $\mathbb{F}$ 上的向量空间，对子集 $S \subset V$，定义 S 在对偶空间 V^* 中的**零化子**（annihilator）为：
$\mathrm{Ann}(S) := \{f \in V^* \mid f(s) = 0,\ \forall s \in S\}$ 即对偶空间 V^* 中所有在 S 上恒为 0 的**线性泛函的集合**。

直和分解下的零化子性质:

Theorem

假设:$V = W \oplus W'$（W, W' 是 V 的子空间，且直和分解成立），则存在同构映射 $i_{W'} : W^* \oplus W'^* \xrightarrow{\sim} V^*$，其中对任意 $w^* \in W^*$、$w'^* \in W'^*$，以及 $v \in W$、$v' \in W'$，映射规则为：$w^* + w'^* \mapsto (v + v' \mapsto w^*(v) + w'^*(v'))$.

即：该线性同构将 W^* 与 W'^* 的直和 $W^* \oplus W'^*$ 映射到 V^*，将 W^* 和 W'^* 的泛函"拼接"为 V^* 中在 W, W' 上分别作用的泛函。

它通过"分解 - 拼接"的方式，反映了原空间直和结构在对偶空间上的对应关系，是研究对偶空间与原空间子空间关系的核心工具

【Remark】:

这里下标用 W'，是为了强调"以 W' 为参照"的拼接方式——尤其是后续讨论"零化子"时，W' 的角色更直接.

若将下标换为 W（定义 $i_W : W^* \oplus W'^* \to V^*$），纯数学上是可行的，但需重新定义映射规则（比如调整泛函作用的"分解方向"）。

但为了与后续性质（如 $\mathrm{Ann}(W)$ 与 W'^* 的关联）符号一致，通常选择 W' 作为下标——它能更自然地体现"W' 是 W 的补空间，且其对偶空间直接生成 W 的零化子"这一联系。

Example

$V = \mathbb{R}^2$ 的直和分解取 $V = \mathbb{R}^2$（实向量空间）；

并令：$W = \mathrm{span}\{(1,0)\}$（x 轴，一维子空间）；$W' = \mathrm{span}\{(0,1)\}$（y 轴，一维子空间）。

取 $w^* \in W^*$ 满足 $w^*((1,0)) = 2$（即 $w^*((x,0)) = 2x$），再取 $w'^* \in W'^*$ 满足 $w'^*((0,1)) = 3$（即 $w'^*((0,y)) = 3y$）。

W^* 是 W 的对偶空间，即"所有从 W 到数域 $\mathbb{F}$ 的线性泛函（线性映射）"构成的向量空间；

$w^* \in W^*$ 表示：w^* 是定义在 W 上的线性泛函，它满足"线性性"——对任意 $v, u \in W$，$a, b \in \mathbb{F}$，有 $w^*(av + bu) = aw^*(v) + bw^*(u)$，且作用是将 W 中的向量映射到数域 $\mathbb{F}$ 中。

W 的基为 $\{(1,0)\}$，因此 W 上的线性泛函 $w^* \in W^*$ 由"作用于基向量 $(1,0)$ 的值"唯一确定。

若 $w^*((1,0)) = a$（$a \in \mathbb{R}$），则对任意 $(x,0) \in W$，有 $w^*((x,0)) = ax$。

同理，W' 的基为 $\{(0,1)\}$，若 $w'^*((0,1)) = b$（$b \in \mathbb{R}$），则对任意 $(0,y) \in W'$，有 $w'^*((0,y)) = by$。

证明. **正向验证：**

根据同构映射 $i_{W'}$，拼接得到 V 上的线性泛函 $f = i_{W'}(w^* + w'^*)$。现在验证 f 对 V 中任意向量 (x,y) 的作用：将 (x,y) 分解为 $(x,0) + (0,y)$（$(x,0) \in W$，$(0,y) \in W'$），则：$f((x,y)) = w^*((x,0)) + w'^*((0,y)) = 2x + 3y$

显然，$f : \mathbb{R}^2 \to \mathbb{R}$ 是线性泛函（满足线性性：$f(kv) = kf(v)$，$f(v_1 + v_2) = f(v_1) + f(v_2)$），因此 $f \in V^*$

反向验证（泛函的“分解”）：

再取 V^* 中任意线性泛函，比如 $f((x,y)) = 5x - 4y$。我们可以将其“分解”为 W 和 W' 上的泛函：

对 W 中的 $(x,0)$，定义 $w^*((x,0)) = 5x$（即 $w^*((1,0)) = 5$，属于 W^*）；对 W' 中的 $(0,y)$，定义 $w'^*((0,y)) = -4y$（即 $w'^*((0,1)) = -4$，属于 W'^*）。

此时 $f = i_{W'}(w^* + w'^*)$，并且我们可以归纳出说明每个 V 上的泛函都能通过这种“拼接”得到，即映射 $i_{W'}$ 是**满射**；

同时，若两个不同的“拼接泛函”作用于 V 上的结果相同，则它们的 w^* 和 w'^* 必须分别相同（由直和分解的唯一性），因此 $i_{W'}$ 是**单射**。综上，$i_{W'}$ 是**双射且保持线性结构，故为同构映射**。

□

基于上述直和与同构，零化子具有以下关键性质：

Property

零化子与对偶空间的像相等：$\mathrm{Ann}(W) = i_{W'}(W'^*)$

特别地，商空间 V/W 的对偶空间与 $\mathrm{Ann}(W)$ 同构：$(V/W)^* \simeq \mathrm{Ann}(W)$

证明. **第一步：**$i_{W'}(W'^*) \subseteq$ **Ann**(W) 任取 $f \in i_{W'}(W'^*)$，则存在 $g \in W'^*$ 使得 $f = i_{W'}(g)$。对任意 $w \in W$，将 w 分解为 $w = w + 0$（其中 $0 \in W'$），则：$f(w) = i_{W'}(g)(w) = g(0) = 0$（线性泛函在零向量上的作用必为 0）。

因此 f 在 W 上恒为 0，即 $f \in \mathrm{Ann}(W)$。故 $i_{W'}(W'^*) \subseteq \mathrm{Ann}(W)$。

第二步：Ann$(W) \subseteq i_{W'}(W'^*)$ 任取 $f \in \mathrm{Ann}(W)$（即 $f \in V^*$ 且 $f(w) = 0,\ \forall w \in W$）。需要构造 $g \in W'^*$，使得 $f = i_{W'}(g)$：

对 $w' \in W'$，定义 $g(w') = f(w')$。由于 f 是线性泛函，g 也满足线性性（W' 中线性组合的作用保持线性），因此 $g \in W'^*$。

验证 $f = i_{W'}(g)$：对任意 $v = w + w' \in V$（$w \in W, w' \in W'$），有：$i_{W'}(g)(v) = g(w') = f(w')$，而 $f(v) = f(w + w') = f(w) + f(w') = 0 + f(w') = f(w')$（因 $f \in \mathrm{Ann}(W)$，故 $f(w) = 0$）。

因此 $f(v) = i_{W'}(g)(v)$ 对所有 $v \in V$ 成立，即 $f = i_{W'}(g)$。故 $f \in i_{W'}(W'^*)$，即 $\mathrm{Ann}(W) \subseteq i_{W'}(W'^*)$。

□

包含关系的“反向性”若子空间满足 $W_1 \subset W_2 \subset V$，则它们的零化子满足反向包含：$\mathrm{Ann}(W_1) \supset \mathrm{Ann}(W_2)$（子空间越大，能“零化”它的对偶泛函越少）

证明.

□

零化子的“自反性”(有限维情形) 若 W 是有限维子空间 ($\dim_{\mathbb{F}}(W) < \infty$), 则“零化子的零化子”等于 W 本身: $\mathrm{Ann}(\mathrm{Ann}(W)) = W$ 更精确地, $\mathrm{Ann}_V(\mathrm{Ann}(W)) = W$ (强调在 V 中取两次零化子)。

证明. □

直和分解的推广 $W \oplus (\mathrm{Ann}_V(i_{W'}(W^*))) = V$ (W 与由零化子定义的子空间也构成 V 的直和)

证明. □

这些性质体现了“零化子”在对偶空间与原空间的子空间之间的对应关系，以及对包含、直和等结构的刻画作用。

此外，我们还有两个推论:

Corollary & Secondary Conclusion

$\mathrm{Ann}(W_1 + W_2) = \mathrm{Ann}(W_1) \cap \mathrm{Ann}(W_2)$

证明. 第一步: $\mathrm{Ann}(W_1+W_2) \subseteq \mathrm{Ann}(W_1) \cap \mathrm{Ann}(W_2)$ 任取 $f \in \mathrm{Ann}(W_1+W_2)$, 根据零化子定义, 对所有 $v \in W_1+W_2$, 有 $f(v) = 0$。由于 $W_1 \subseteq W_1+W_2$ (任取 $w_1 \in W_1$, 可表为 $w_1 + 0 \in W_1 + W_2$), 因此对所有 $w_1 \in W_1$, $f(w_1) = 0$, 故 $f \in \mathrm{Ann}(W_1)$。

同理, $W_2 \subseteq W_1 + W_2$, 对所有 $w_2 \in W_2$, $f(w_2) = 0$, 故 $f \in \mathrm{Ann}(W_2)$。

因此 $f \in \mathrm{Ann}(W_1) \cap \mathrm{Ann}(W_2)$, 即 $\mathrm{Ann}(W_1 + W_2) \subseteq \mathrm{Ann}(W_1) \cap \mathrm{Ann}(W_2)$。

第二步: $\mathrm{Ann}(W_1) \cap \mathrm{Ann}(W_2) \subseteq \mathrm{Ann}(W_1+W_2)$ 任取 $f \in \mathrm{Ann}(W_1) \cap \mathrm{Ann}(W_2)$, 则 $f \in \mathrm{Ann}(W_1)$ 且 $f \in \mathrm{Ann}(W_2)$。

对任意 $v \in W_1 + W_2$, 由和空间定义, $v = w_1 + w_2$ (其中 $w_1 \in W_1$, $w_2 \in W_2$)。由线性泛函的线性性, $f(v) = f(w_1 + w_2) = f(w_1) + f(w_2)$。

因 $f \in \mathrm{Ann}(W_1)$, 故 $f(w_1) = 0$; 因 $f \in \mathrm{Ann}(W_2)$, 故 $f(w_2) = 0$。

因此 $f(v) = 0 + 0 = 0$。对所有 $v \in W_1 + W_2$, $f(v) = 0$, 故 $f \in \mathrm{Ann}(W_1 + W_2)$, 即 $\mathrm{Ann}(W_1) \cap \mathrm{Ann}(W_2) \subseteq \mathrm{Ann}(W_1 + W_2)$。

结合两步, $\mathrm{Ann}(W_1 + W_2) = \mathrm{Ann}(W_1) \cap \mathrm{Ann}(W_2)$, 命题成立。

□

【Remark】:

命题 $\mathrm{Ann}(W_1 \cap W_2) = \mathrm{Ann}(W_1) + \mathrm{Ann}(W_2)$ 不成立!

反例构造 (无限维情形) 设 $V = \mathbb{R}^\infty$ (基为 $\{e_n\}_{n=1}^\infty$), 取: $W_1 = \mathrm{span}\{e_{2k-1} \mid k \geq 1\}$ (奇数下标基向量生成的子空间), $W_2 = \mathrm{span}\{e_{2k} \mid k \geq 1\}$ (偶数下标基向量生成的子空间)。

则 $W_1 \cap W_2 = \{0\}$, 故 $\mathrm{Ann}(W_1 \cap W_2) = V^*$ (零子空间的零化子是整个对偶空间)。

分析 $\mathrm{Ann}(W_1)+\mathrm{Ann}(W_2)$

$\mathrm{Ann}(W_1)+\mathrm{Ann}(W_2)$ 中的元素形如 $f+g$（$f\in\mathrm{Ann}(W_1)$，$g\in\mathrm{Ann}(W_2)$），即：$f+g=\sum_{k=1}^{\infty}a_k e_{2k}^*+\sum_{k=1}^{\infty}b_k e_{2k-1}^*$

注意到 V^* 中的泛函可由任意实序列 $(c_1,c_2,\dots)$ 表示（因为 V 中向量是有限支撑的，线性泛函对 e_n 的作用 $c_n=f(e_n)$ 可任意选取）。但 $\mathrm{Ann}(W_1)+\mathrm{Ann}(W_2)$ 中的泛函具有“奇数位置系数来自 g、偶数位置系数来自 f”的限制，无法覆盖所有实序列。

1.14.8 伴随

有限维内积空间中，我们大致会讨论两种伴随，这两种伴随是同一概念的“代数版”与“内积版”，通过里斯同构自洽。

现在我们先来介绍代数版：从“向量映射”到“泛函映射”

设 $T:V\to W$ 是域 $\mathbb{F}$ 上的线性映射（向量空间间的线性映射）：W^* 是 W 的对偶空间：由所有 $W\to\mathbb{F}$ 的线性泛函（即“向量 $\to$ 数”的线性映射）构成的向量空间；V^* 是 V 的对偶空间，同理。

接着，我们给出伴随映射 $T^*:W^*\to V^*$ 的定义：

Definition

对任意 $f\in W^*$（W 上的泛函）和 $v\in V$，规定 $T^*(f)(v):=f(T(v))$..

逻辑：T 将 V 中向量 v 送到 W 中向量 $T(v)$；然后 f 将 W 中向量 $T(v)$ 送到数 $f(T(v))$。因此，“先作用 T，再作用 f”的复合，定义了一个 $V\to\mathbb{F}$ 的映射 $T^*(f)$。由于 T 和 f 都是线性的，复合后 $T^*(f)$ 仍为线性泛函，故 $T^*(f)\in V^*$。

即 T^* 把 W 上的泛函 f，转化为 V 上的泛函 $f\circ T$（先作用 T，再作用 f）。

Theorem

线性映射 $T:V\to W$ 与其伴随 $T^*:W^*\to V^*$ 的单射（injective）、满射（surjective）性质**反向等价**，即：

如果 T 是满射 $\iff$ T^* 是单射；

如果 T 是单射 $\iff$ T^* 是满射。

证明. **1. T 满射 $\implies T^*$ 单射**

目标：证明 T^* 的核为 $\{0\}$（单射的定义：核仅含零元）。

步骤：取 $f\in\ker(T^*)$（即 $T^*(f)=0$）。由伴随定义，对所有 $v\in V$，$T^*(f)(v)=0$，即 $f(T(v))=0$。

因 T 满射，对任意 $w\in W$，存在 $v\in V$ 使得 $T(v)=w$。因此，对任意 $w\in W$，$f(w)=f(T(v))=0$，即 f 是 W^* 中的零泛函（$f=0$）。

故 $\ker(T^*)=\{0\}$，即 T^* 是单射。

2.T^* 单射 $\implies$ T 满射

目标：证明 $\mathrm{Im}(T) = W$（满射的定义：像等于目标空间）。

步骤（反证法）：假设 T 不满射，则 $\mathrm{Im}(T) \subsetneq W$。由零化子性质（子空间的零化子非空）：

> *Claim*
>
> 根据零化子的基本性质：若 S 是 W 的真子空间，则 $\mathrm{Ann}(S)$ 中存在 * 非零泛函（子空间的零化子不会只有零泛函）

存在非零泛函 $f \in W^*$，使得 f 在 $\mathrm{Im}(T)$ 上恒为 0（即 $f(T(v)) = 0$ 对所有 $v \in V$ 成立）。

由伴随定义，$T^*(f)(v) = f(T(v)) = 0$ 对所有 $v \in V$ 成立，故 $T^*(f) = 0$，即 $f \in \ker(T^*)$。

但 $f \neq 0$，与 T^* 单射（$\ker(T^*) = \{0\}$）矛盾。因此假设不成立，T 必满射。

3. T 单射 $\implies$ T^* 满射

目标：证明对任意 $g \in V^*$，存在 $f \in W^*$ 使得 $T^*(f) = g$（满射的定义：像覆盖目标空间）。

步骤：因 T 单射，$T : V \to \mathrm{Im}(T)$ 是同构（单射且像为 $\mathrm{Im}(T)$），故存在逆映射 $T^{-1} : \mathrm{Im}(T) \to V$。对任意 $g \in V^*$，定义 $f : W \to \mathbb{F}$：若 $w \in \mathrm{Im}(T)$，则 $f(w) = g(T^{-1}(w))$；若 $w \in W \setminus \mathrm{Im}(T)$，利用“子空间补空间”（$W = \mathrm{Im}(T) \oplus U$），对 U 中元素任意赋值（不影响线性性）。易证 $f \in W^*$（线性泛函），且对任意 $v \in V$，$T^*(f)(v) = f(T(v)) = g(T^{-1}(T(v))) = g(v)$，故 $T^*(f) = g$。因此 T^* 是满射。

4. T^* 满射 $\implies$ T 单射

目标：证明 $\ker(T) = \{0\}$（单射的定义：核仅含零元）。

步骤（反证法）：假设 T 不单射，则 $\ker(T) \neq \{0\}$，故 $\dim \mathrm{Im}(T) < \dim V$。

由对偶空间维度性质（$\dim X^* = \dim X$），$\dim W^* = \dim W$，$\dim V^* = \dim V$。因 T^* 满射，$\dim \mathrm{Im}(T^*) = \dim V^* = \dim V$。但 $\mathrm{Im}(T^*) \subseteq V^*$，且 $\dim \mathrm{Im}(T^*) \leq \dim W^* = \dim W$，结合 $\dim \mathrm{Im}(T) < \dim V$，得 $\dim W \geq \dim V > \dim \mathrm{Im}(T)$。

由零化子性质,存在非零泛函 $f \in W^*$ 使 f 在 $\mathrm{Im}(T)$ 上恒为 0(即 $T^*(f) = 0$)，与 T^* 满射（$\mathrm{Im}(T^*) = V^*$，无“非零核”）矛盾。因此假设不成立，T 必单射。 □

> *Claim*
>
> $\ker(T^*) \simeq (W/\mathrm{Im}(T))^*$

解释：商空间 $W/\mathrm{Im}(T)$ 的元素是“W 中模 $\mathrm{Im}(T)$ 的等价类”。

其对偶空间 $(W/\mathrm{Im}(T))^*$ 中的泛函，必须“在 $\mathrm{Im}(T)$ 上恒为 0”（商空间泛函的“良定义性”要求）。

而 $\ker(T^*)$ 中的泛函 $f \in W^*$ 满足 $T^*(f) = 0$，即 $f(T(v)) = 0$ 对所有 $v \in V$ 成立，等价于 f 在 $\mathrm{Im}(T)$ 上恒为 0。

因此，$\ker(T^*)$ 与 $(W/\mathrm{Im}(T))^*$ 作为"在 $\mathrm{Im}(T)$ 上零化的 W 泛函"集合，是同构的。这也是 $W/\mathrm{Im}(T)$ 被称为 T 的余核 (cokernel) 的原因 (对偶于 T 的核)。

接下来我们讨论伴随映射的矩阵表示和转置的来源:

要找 T^* 在对偶基下的矩阵，与 T 在原基下的矩阵的关系，步骤如下：

1. 原映射 T 的矩阵: 设 V 的基 $\mathcal{A} = \{v_1, \ldots, v_n\}$，$W$ 的基 $\mathcal{B} = \{w_1, \ldots, w_m\}$。

T 在这组基下的矩阵为 $T_{\mathcal{B},\mathcal{A}} = (a_{ij})$, 即 $T(v_j) = \sum_{i=1}^m a_{ij} w_i$ (矩阵第 j 列是 $T(v_j)$ 在 $\mathcal{B}$ 下的坐标)。

2. 设伴随映射 T^* 的矩阵形式对偶基为 $\mathcal{A}^* = \{v_1^*, \ldots, v_n^*\}$ (V^* 的基)，$\mathcal{B}^* = \{w_1^*, \ldots, w_m^*\}$ (W^* 的基)。

要确定 $T^* : W^* \to V^*$ 在基 $\mathcal{B}^*$ (定义域基) 和 $\mathcal{A}^*$ (值域基) 下的矩阵 $T^*_{\mathcal{A}^*,\mathcal{B}^*}$，设：$T^*(w_k^*) = \sum_{l=1}^n b_{lk} v_l^*$ 需确定系数 b_{lk}。

用对偶基的性质求系数 b_{lk} 对 $v_j \in \mathcal{A}$，计算 $T^*(w_k^*)(v_j)$：由伴随定义：$T^*(w_k^*)(v_j) = (w_k^* \circ T)(v_j) = w_k^*(T(v_j))$。

分析 $T^*(w_k^*)(v_j)$ 的结构 $w_k^* \in W^*$：w_k^* 是 W 的对偶基 $\mathcal{B}^* = \{w_1^*, \ldots, w_m^*\}$ 中的元素 (因为 $\mathcal{B} = \{w_1, \ldots, w_m\}$ 是 W 的基)。$T^*(w_k^*) \in V^*$：由于 $T^* : W^* \to V^*$，将 W^* 中的 w_k^* 输入 T^*，输出是 V^* 中的一个线性泛函。$T^*(w_k^*)(v_j) \in \mathbb{F}$：将 V^* 中的泛函 $T^*(w_k^*)$ 作用在 V 的基向量 v_j 上，结果是域 $\mathbb{F}$ 中的一个标量。

回到主题，

代入 $T(v_j) = \sum_{i=1}^m a_{ij} w_i$，得到：$w_k^* \left(\sum_{i=1}^m a_{ij} w_i\right) = \sum_{i=1}^m a_{ij} \cdot w_k^*(w_i)$

由对偶基性质 $w_k^*(w_i) = \delta_{ki}$ (仅 $i = k$ 时为 1)，故上式化简为 a_{kj}。

另一方面，若 $T^*(w_k^*) = \sum_{l=1}^n b_{lk} v_l^*$，则作用在 v_j 上：$T^*(w_k^*)(v_j) = \sum_{l=1}^n b_{lk} v_l^*(v_j) = \sum_{l=1}^n b_{lk} \delta_{lj} = b_{jk}$

比较系数，得到转置关系由 $T^*(w_k^*)(v_j) = a_{kj}$ 且 $T^*(w_k^*)(v_j) = b_{jk}$，得 $b_{jk} = a_{kj}$。

这说明：T^* 在对偶基下的矩阵 (b_{jk})，恰好是 T 在原基下的矩阵 (a_{ij}) 的转置。即：$T^*_{\mathcal{A}^*,\mathcal{B}^*} = {}^t T_{\mathcal{B},\mathcal{A}}$

我们可以得到如下推论：

Corollary & Secondary Conclusion

${}^t(AB) = {}^tB\,{}^tA$ (矩阵转置的"乘积法则")

要证明这个"反转律"，需结合线性映射的复合与对偶映射的复合，再利用"对偶映射的矩阵是原矩阵的转置"。

证明. 步骤 1：线性映射的复合与矩阵乘法的关系设：U, V, W 是有限维向量空间；线性映射 $S : U \to V$，在基 $\mathcal{C}$ (U 的基) 和 $\mathcal{A}$ (V 的基) 下的矩阵为 $S_{\mathcal{A},\mathcal{C}}$；线性映射 $T : V \to W$，在基 $\mathcal{A}$ (V 的基) 和 $\mathcal{B}$ (W 的基) 下的矩阵为 $T_{\mathcal{B},\mathcal{A}}$。线性映射的复合

$T \circ S : U \to W$（先作用 S，再作用 T），在基 $\mathcal{C}$（U 的基）和 $\mathcal{B}$（W 的基）下的矩阵是矩阵乘法：$(T \circ S)_{\mathcal{B},\mathcal{C}} = T_{\mathcal{B},\mathcal{A}} \cdot S_{\mathcal{A},\mathcal{C}}$

步骤 2：对偶映射的复合与转置矩阵的关系对偶映射是反变的（即“复合的对偶 = 对偶的反向复合”）：对线性映射 S, T，有 $(T \circ S)^* = S^* \circ T^*$ 其中：$(T \circ S)^* : W^* \to U^*$ 是 $T \circ S$ 的对偶映射；$S^* : V^* \to U^*$ 是 S 的对偶映射，$T^* : W^* \to V^*$ 是 T 的对偶映射，故 $S^* \circ T^* : W^* \to U^*$。再结合对偶映射的矩阵是原矩阵的转置：$(T \circ S)^*$ 在对偶基 $\mathcal{C}^*$（U^* 的基）和 $\mathcal{B}^*$（W^* 的基）下的矩阵为 ${}^t(T_{\mathcal{B},\mathcal{A}} \cdot S_{\mathcal{A},\mathcal{C}})$（因为 $T \circ S$ 的矩阵是 $T_{\mathcal{B},\mathcal{A}} \cdot S_{\mathcal{A},\mathcal{C}}$，对偶映射的矩阵是其转置）；$S^*$ 在对偶基 $\mathcal{C}^*$（U^* 的基）和 $\mathcal{A}^*$（V^* 的基）下的矩阵为 ${}^tS_{\mathcal{A},\mathcal{C}}$；$T^*$ 在对偶基 $\mathcal{A}^*$（V^* 的基）和 $\mathcal{B}^*$（W^* 的基）下的矩阵为 ${}^tT_{\mathcal{B},\mathcal{A}}$。

步骤 3：复合对偶映射的矩阵 = 对偶映射矩阵的乘积因为 $(T \circ S)^* = S^* \circ T^*$，所以它们的矩阵也相等：${}^t(T_{\mathcal{B},\mathcal{A}} \cdot S_{\mathcal{A},\mathcal{C}}) = {}^tS_{\mathcal{A},\mathcal{C}} \cdot {}^tT_{\mathcal{B},\mathcal{A}}$ 若将线性映射对应的矩阵一般化为“任意矩阵”（令 $A = S_{\mathcal{A},\mathcal{C}}$，$B = T_{\mathcal{B},\mathcal{A}}$），则上式变为：${}^t(BA) = {}^tA\,{}^tB$ 对“任意性”稍作调整（交换 A, B 的角色，或重新命名映射），即可得到更常用的形式：对任意两个矩阵 A（尺寸 $n \times p$）和 B（尺寸 $p \times m$），有 ${}^t(AB) = {}^tB\,{}^tA$ 这就是矩阵转置的乘积法则（反转律）。 □

1.15 坐标系变换与过渡矩阵初步

坐标变换本质上是一种过渡矩阵的运算，也是一种“变换”的思想，我们将以一个例子详细展开。

问题情境：对于 $T : \mathbb{R}^2 \to \mathbb{R}^2$，寻找一个线性变换，使得点 (a, b) 与 $T(a, b)$ 关于直线 $l : y = 2x$ 对称。

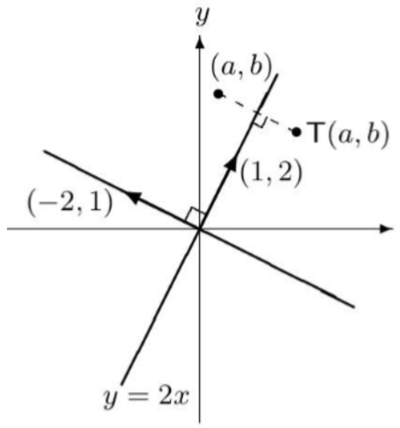
对以直角坐标为基的全空间，即线性空间 V ，其基为 $\beta = \{e_1, e_2\}$，$[e_1]_\beta = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$，$[e_2]_\beta = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$。

设 W 空间的基为 $\gamma = \{w_1, w_2\}$，其中 $[w_1]_\beta = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$，$[w_2]_\beta = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$。

容易证明，V 和 W 本质上就是同一空间，但是 β 和 γ 则是同一空间的不同的两组基底。

步骤 1：分析 W 空间中 T 的变换根据线性变换的性质：$[T(v)]_\gamma = [T]_\gamma [v]_\gamma$ 。

因为线性变换由基唯一确定，只需确定 T 对基的作用，就能确定 T 对 V 中所有向量的作用。因为 $[T(w_1)]_\gamma = (1, 0)$（保持不变），$[T(w_2)]_\gamma = (0, -1)$（反射变换）。由此得



到了在 W 空间中 γ 坐标系下 T 变换的矩阵 $[T]_\gamma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$。

步骤 2：构造过渡矩阵

根据公式推导：

$$\begin{aligned} [T(v)]_\beta &= [T]_\beta [v]_\beta \\ &= [I]_\gamma^\beta [T(v)]_\gamma \\ &= [I]_\gamma^\beta [T]_\gamma [v]_\gamma \\ &= [I]_\gamma^\beta [T]_\gamma [I]_\beta^\gamma [v]_\beta \end{aligned}$$

等价于 $[T]_\beta = [I]_\gamma^\beta [T]_\gamma [I]_\beta^\gamma$ 。又因为 $[I^{-1}]_\gamma^\beta = [I]_\beta^\gamma$ ，即 $([I]_\gamma^\beta)^{-1} = [I]_\beta^\gamma$ 。所以 $[T]_\beta = ([I]_\gamma^\beta)^{-1}[T]_\gamma [I]_\beta^\gamma$

【Remark】：

“从 $\beta \to \gamma$” 的 $[I]_\gamma^\beta$（恒等映射）$\neq I$（同一空间下的恒等映射）。

步骤 3：计算过渡矩阵：$[I]_\gamma^\beta$ ：由 $[I(v)]_\beta = [v]_\beta = [I]_\gamma^\beta [v]_\gamma$（从 $\gamma \to \beta$ ），可得 $[I]_\gamma^\beta = \Big([T(w_1)]_\beta \quad [T(w_2)]_\beta\Big) = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}$ 。

求 $[I]_\beta^\gamma$ ：由 $[v]_\gamma = [I(v)]_\gamma = [I]_\beta^\gamma [v]_\beta$ ，可得 $[I]_\beta^\gamma = \Big([T(e_1)]_\gamma \quad [T(e_2)]_\gamma\Big) = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{2}{5} & \frac{1}{5} \end{pmatrix}$ 。

观察：$\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{2}{5} & \frac{1}{5} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ，即 $[I]_\gamma^\beta \cdot [I]_\beta^\gamma = E_{2\times 2}$ 。

步骤 4：计算最终变换结果

$$
\begin{aligned}
[T(a,b)]_\beta &= [T]_\beta \begin{pmatrix} a \\ b \end{pmatrix} \\
&= \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \\
&= \frac{1}{5}\begin{pmatrix} -3a+4b \\ 4a+3b \end{pmatrix}
\end{aligned}
$$

其中 $[T]_\beta = ([I]_\gamma^\beta)^{-1}[T]_\gamma[I]_\beta^\gamma$ 。

用具体点测试：取直线 $y=2x$ 上的点 $(1,2)$（即 w_1），反射后应不变：$[T(1,2)]_\beta = \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} -3+8 \\ 4+6 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

取垂直于直线 $y=2x$ 的点 $(-2,1)$（即 w_2），反射后应反向为 $(2,-1)$：$[T(-2,1)]_\beta = \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix} \cdot \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 6+4 \\ -8+3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ 验证通过

Corollary & Secondary Conclusion

对于直线 $y=kx$ ，其反射矩阵为：$\begin{pmatrix} \frac{1-k^2}{1+k^2} & \frac{2k}{1+k^2} \\ \frac{2k}{1+k^2} & \frac{k^2-1}{1+k^2} \end{pmatrix}$ 。

当 $k=2$ 时，可据此推导出上述关于直线 $y=2x$ 对称变换的矩阵。

有兴趣的同学可以推导一下！

课后练习：

Exercise

完成对投影映射的坐标系变换的推导：

1.16 不变子空间与循环子空间

Definition

T-不变子空间（T-invariant subspace）：设 T 是线性空间 V 上的线性变换，W 是 V 的一个子空间。

若 $T(W) \subseteq W$，则称 W 是 T 的一个不变子空间（T-invariant subspace）。

Definition

受限线性变换/限制线性变换（Constrained Linear Transformation）：若 W 是 T-不变子空间，则 T 在 W 上的限制 $T|_W : W \to W$，定义为 $T|_W(\alpha) = T(\alpha)$ 对所有 $\alpha \in W$，是 W 上的线性变换。

接下来，我们讨论不变子空间与矩阵化简:

设 V 是有限维线性空间，W 是 T-不变子空间。取 W 的一组基 $\{\alpha_1,\dots,\alpha_k\}$，并将其扩充为 V 的基 $\{\alpha_1,\dots,\alpha_k,\alpha_{k+1},\dots,\alpha_n\}$。则 T 在这组基下的矩阵 $[T]$ 具有分块形式:

$$[T]=\begin{pmatrix}[A] & [B]\\ [0] & [C]\end{pmatrix}$$

其中 $[A]$ 是 $T|_W$ 在 $\{\alpha_1,\dots,\alpha_k\}$ 下的矩阵。

若 $V=W_1\oplus W_2\oplus\cdots\oplus W_s$，其中每个 W_i 都是 T-不变子空间，则 $[T]$ 可化为分块对角矩阵 $\mathrm{diag}([T|_{W_1}],[T|_{W_2}],\dots,[T|_{W_s}])$。

类似于子空间的交与和，不变子空间的和与交因其 **T-不变性**，有更多值得考察的性质。

Property

若 W_1,W_2 是 T-不变子空间，则:

1. $W_1+W_2=\{\alpha_1+\alpha_2\mid\alpha_1\in W_1,\alpha_2\in W_2\}$ 是 T-不变子空间;
2. $W_1\cap W_2$ 是 T-不变子空间。

证明. 证明：对于 W_1+W_2，任取 $\alpha=\alpha_1+\alpha_2\in W_1+W_2$，则 $T(\alpha)=T(\alpha_1)+T(\alpha_2)$。

因为 $T(\alpha_1)\in W_1$，$T(\alpha_2)\in W_2$，所以 $T(\alpha)\in W_1+W_2$。

对于 $W_1\cap W_2$，任取 $\alpha\in W_1\cap W_2$，则 $T(\alpha)\in W_1$ 且 $T(\alpha)\in W_2$，故 $T(\alpha)\in W_1\cap W_2$。 □

最后，我们给出循环子空间的定义，这在相似标准型中有重要的应用。

Definition

循环子空间（Cyclic subspace）：给定向量 $\alpha\in V$，定义由 α 生成的循环子空间 $Z(\alpha;T)=\mathrm{span}\{\alpha,T\alpha,T^2\alpha,\dots\}$。

Property

由 α 生成的循环子空间 $Z(\alpha;T)=\text{span}\{\alpha,T\alpha,T^2\alpha,\dots\}$ 是 T-不变子空间

证明. 显然 $T(T^k\alpha)=T^{k+1}\alpha\in Z(\alpha;T)$ 对任意非负整数 k 成立，故 $T(Z(\alpha;T))\subseteq Z(\alpha;T)$。 $\square$

Chapter 2

行列式

行列式是线性代数中的核心概念之一，它不仅是解线性方程组的重要工具，更是刻画线性变换、研究矩阵性质的有力手段。本章将系统地介绍行列式的定义、性质、计算方法及其应用，构建完整的行列式知识体系。

2.1 3 阶行列式

考虑三元线性方程组：$\begin{cases} a_{11}x + a_{12}y + a_{13}z = b_1 \\ a_{21}x + a_{22}y + a_{23}z = b_2 \\ a_{31}x + a_{32}y + a_{33}z = b_3 \end{cases}$

首先从方程(2.1)的前两个方程中消去 z：

$$\begin{aligned}
&\text{方程 } 1 \times a_{23}: && a_{11}a_{23}x + a_{12}a_{23}y + a_{13}a_{23}z = b_1a_{23} \\
&\text{方程 } 2 \times a_{13}: && a_{21}a_{13}x + a_{22}a_{13}y + a_{23}a_{13}z = b_2a_{13} \\
&\text{相减得:} && (a_{11}a_{23} - a_{21}a_{13})x + (a_{12}a_{23} - a_{22}a_{13})y = b_1a_{23} - b_2a_{13}
\end{aligned}$$

类似地，从方程(2.1)的第一个和第三个方程中消去 z：

$$\begin{aligned}
&\text{方程 } 1 \times a_{33}: && a_{11}a_{33}x + a_{12}a_{33}y + a_{13}a_{33}z = b_1a_{33} \\
&\text{方程 } 3 \times a_{13}: && a_{31}a_{13}x + a_{32}a_{13}y + a_{33}a_{13}z = b_3a_{13} \\
&\text{相减得:} && (a_{11}a_{33} - a_{31}a_{13})x + (a_{12}a_{33} - a_{32}a_{13})y = b_1a_{33} - b_3a_{13}
\end{aligned}$$

继续消元求解 x 的值，得到的两个新方程：

$$(a_{11}a_{23} - a_{21}a_{13})x + (a_{12}a_{23} - a_{22}a_{13})y = b_1a_{23} - b_2a_{13}$$
$$(a_{11}a_{33} - a_{31}a_{13})x + (a_{12}a_{33} - a_{32}a_{13})y = b_1a_{33} - b_3a_{13}$$

消去 y 后，可得：

$$x = \frac{(b_1a_{23} - b_2a_{13})(a_{12}a_{33} - a_{32}a_{13}) - (b_1a_{33} - b_3a_{13})(a_{12}a_{23} - a_{22}a_{13})}{(a_{11}a_{23} - a_{21}a_{13})(a_{12}a_{33} - a_{32}a_{13}) - (a_{11}a_{33} - a_{31}a_{13})(a_{12}a_{23} - a_{22}a_{13})}$$

观察 x 的表达式，我们可以设计三阶行列式如下：

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

进一步展开：

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

最后用行列式表示解

则 x 的值可表示为：

$$x = \frac{\begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}$$

类似地，y 和 z 的值为：

$$y = \frac{\begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}, \quad z = \frac{\begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}$$

2.1.1 3 阶行列式的定义

有了前面的讨论，我们直接引入 3 阶行列式的概念。

【定义】：3 阶行列式 (3 - order determinant)：由 3×3 矩阵 $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ 的元素构成的 3 阶行列式定义为：$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \mathrm{a}_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$

即第一行元素分别乘以其对应的代数余子式之和。其中，元素 a_{ij} 的代数余子式（Algebraic cofactor）是 $(-1)^{i+j}$ 乘以去掉第 i 行和第 j 列后剩下的二阶行列式。

我们以两个例子来说明三阶行列式的计算：

Example

例 1：计算 3 阶行列式：$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix}$

解：根据 3 阶行列式的定义：

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} &= 1 \cdot \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \cdot \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= 0 \end{aligned}$$

Example

例 2：计算 3 阶行列式：$\begin{vmatrix} 2 & 0 & 1 \\ 1 & -4 & -1 \\ -1 & 8 & 3 \end{vmatrix}$

解：按第一行展开：

$$\begin{aligned} \begin{vmatrix} 2 & 0 & 1 \\ 1 & -4 & -1 \\ -1 & 8 & 3 \end{vmatrix} &= 2 \cdot \begin{vmatrix} -4 & -1 \\ 8 & 3 \end{vmatrix} - 0 \cdot \begin{vmatrix} 1 & -1 \\ -1 & 3 \end{vmatrix} + 1 \cdot \begin{vmatrix} 1 & -4 \\ -1 & 8 \end{vmatrix} \\ &= 2 \cdot (-12 + 8) - 0 + 1 \cdot (8 - 4) \\ &= -4 \end{aligned}$$

2.2 n 阶行列式及其重要性质

2.2.1 n 阶行列式的定义

行列式（Determinant）定义：由 n 阶方阵 A 的元素按特定规则计算得到的一个数，记为 $\det(A)$ 或 $|A|$。

n 阶行列式的一般形式为：

$$D = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

其中：a_{ij} 表示行列式中第 i 行第 j 列的元素。

例如：二阶行列式 $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$。

又如：三阶行列式：$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 1\cdot(5\cdot9-6\cdot8)-2\cdot(4\cdot9-6\cdot7)+3\cdot(4\cdot8-5\cdot7)=0$

【行列式按行（列）展开】(Expansion of the determinant by row (column))

【定义】：余子式（Minor）：对于 n 阶方阵 $A=(a_{ij})$，元素 a_{ij} 的余子式 M_{ij} 是指删除 A 的第 i 行和第 j 列后剩余的 $(n-1)$ 阶子矩阵的行列式。

【定义】：代数余子式（Cofactor）：元素 a_{ij} 的代数余子式 C_{ij} 定义为：$C_{ij}=(-1)^{i+j}M_{ij}$ 其中，符号 $(-1)^{i+j}$ 决定了代数余子式的正负号。

于是我们可以得到 n 阶行列式按第 1 行展开的公式：

n 阶行列式可按第 1 行展开为：$\det(A)=a_{11}C_{11}+a_{12}C_{12}+\cdots+a_{1n}C_{1n}=\sum_{j=1}^{n}a_{1j}C_{1j}$

其中：a_{1j} 是第 1 行第 j 列的元素；$C_{1j}=(-1)^{1+j}M_{1j}$ 是对应的代数余子式。

当然，实际上我们可以更进一步：得到：

公式按行展开公式：对于 n 阶方阵 $A=(a_{ij})$，其行列式 $\det(A)$ 按第 i 行展开的公式为 $\det(A)=\sum_{j=1}^{n}a_{ij}A_{ij}=\sum_{j=1}^{n}a_{ij}(-1)^{i+j}M_{ij}$，其中 $i=1,2,\cdots,n$。这里 a_{ij} 是第 i 行第 j 列的元素，M_{ij} 是 a_{ij} 的余子式，即去掉 A 的第 i 行和第 j 列后剩下的 $(n-1)$ 阶子矩阵的行列式，$A_{ij}=(-1)^{i+j}M_{ij}$ 是 a_{ij} 的代数余子式。

按列展开公式：$\det(A)$ 按第 j 列展开的公式为 $\det(A)=\sum_{i=1}^{n}a_{ij}A_{ij}=\sum_{i=1}^{n}a_{ij}(-1)^{i+j}M_{ij}$，其中 $j=1,2,\cdots,n$。本质上按行和按列展开的原理是一致的，只是选取的方向不同。

接下来，我们将考察三类重要的也是特殊的行列式

2.2.2 三类重要 n 阶行列式及其性质

对角行列式：若方阵 $A=(a_{ij})$ 满足当 $i\neq j$ 时，$a_{ij}=0$，则 A 为对角矩阵，其行列式为对角行列式。

$$A=\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

$\det(A)=\prod_{i=1}^{n}a_{ii}$

上三角行列式：若方阵 $A=(a_{ij})$ 满足当 $i>j$ 时，$a_{ij}=0$，则 A 为上三角矩阵，其

行列式为上三角行列式。

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

$\det(A)=\prod_{i=1}^{n} a_{ii}$

下三角行列式：若方阵 $A=(a_{ij})$ 满足当 $i<j$ 时，$a_{ij}=0$，则 A 为下三角矩阵，其行列式为下三角行列式。

$$A=\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

$\det(A)=\prod_{i=1}^{n} a_{ii}$

2.2.3 n 阶行列式的重要性质

行列式具有许多重要的性质，这些性质不仅有助于行列式的计算，还揭示了行列式与矩阵的内在联系。

Theorem

【行列式的基本性质】设 A 为 n 阶矩阵，则：

1. $\det(A^T)=\det(A)$，即矩阵转置的行列式等于原矩阵的行列式。
2. 若交换矩阵 A 的两行（列），则行列式的值变号。
3. 若矩阵 A 的某一行（列）中所有元素都乘以同一个数 k，则行列式的值乘以 k。
4. 若矩阵 A 的某一行（列）是另外两行（列）的线性组合，则行列式的值为零。
5. 若矩阵 A 的某一行（列）加上另一行（列）的 k 倍，则行列式的值不变。

【Remark】:

这里我们就不加证明地给出性质 1 和 2，其证明与排序有关，因篇幅原因，这里不在赘述。
性质 1 表明行列式对行和列具有对称性，因此关于行的性质对列也成立，反之亦然。
性质 2 说明行列式的值与行（列）的顺序有关，而性质 3 则表明行列式是行（列）向量的多重线性函数。

Example

计算行列式：

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{vmatrix} \tag{2.1}$$

观察到第三行是第一行的 3 倍减去第二行，即：

$$(9, 10, 11, 12) = 3 \cdot (1, 2, 3, 4) - (5, 6, 7, 8) \tag{2.2}$$

根据行列式的性质 4，行列式的值为零。

2.3 行列式的计算：爪形行列式与范德蒙德行列式

2.3.1 爪形行列式

爪形行列式是一种特殊形式的行列式，其形状类似爪子，通常可以通过行列式的性质转化为上三角行列式进行计算。

Example

【爪形行列式的计算】: 计算行列式: D = $\begin{vmatrix} a_0 & 1 & 1 & \cdots & 1 \\ 1 & a_1 & 0 & \cdots & 0 \\ 1 & 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & a_n \end{vmatrix}$ 其中 $a_1a_2\cdots a_n \neq 0$。

将第 i 列的 $-\frac{1}{a_{i-1}}$ 倍加到第一列（$i = 2, 3, \ldots, n+1$），得到: D = $\begin{vmatrix} a_0 - \sum_{i=1}^{n} \frac{1}{a_i} & 1 & 1 & \cdots & 1 \\ 0 & a_1 & 0 & \cdots & 0 \\ 0 & 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n \end{vmatrix}$ 这是一个上三角行列式，其值为对角线上元素的乘积:

$$D = a_1a_2\cdots a_n \left(a_0 - \sum_{i=1}^{n} \frac{1}{a_i} \right) \tag{2.3}$$

Example

计算行列式:

D = $\begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{vmatrix}$

这是一个爪形行列式，将第 2、3、4 列的适当倍数加到第一列:

$$\begin{aligned} D &= \begin{vmatrix} 1-4-9-16 & 2 & 3 & 4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} \\ &= \begin{vmatrix} -28 & 2 & 3 & 4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} \\ &= -28 \end{aligned}$$

2.3.2 范德蒙德行列式

范德蒙德行列式是另一种重要的特殊行列式，它在多项式插值、信号处理等领域有广泛应用。

Definition

[范德蒙德行列式] n 阶范德蒙德行列式定义为：

$$\mathrm{V}(\mathrm{x}_1, x_2, \ldots, x_n) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix}$$

Theorem

[范德蒙德行列式的值] n 阶范德蒙德行列式的值为：
$\mathrm{V}(\mathrm{x}_1, x_2, \ldots, x_n) = \prod_{1\le i<j\le n}(x_j - x_i)$
即所有可能的差 $x_j - x_i$（其中 $j > i$）的乘积。

证明. 用数学归纳法证明：

1. 当 $n = 2$ 时，$V(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ x_1 & x_2 \end{vmatrix} = x_2 - x_1$，结论成立。

2. 假设对于 $n-1$ 阶范德蒙德行列式结论成立，考虑 n 阶范德蒙德行列式。从第 n 行开始，每行减去前一行的 x_1 倍，得到：

$$\mathrm{V}(\mathrm{x}_1, x_2, \ldots, x_n) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_n - x_1 \\ 0 & x_2(x_2 - x_1) & \cdots & x_n(x_n - x_1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_2^{n-2}(x_2 - x_1) & \cdots & x_n^{n-2}(x_n - x_1) \end{vmatrix}$$

按第一列展开，得到：$\mathrm{V}(\mathrm{x}_1, x_2, \ldots, x_n) = \prod_{i=2}^{n}(x_i - x_1) \cdot V(x_2, x_3, \ldots, x_n)$ $\quad \mathrm{V}(\mathrm{x}_2, x_3, \ldots, x_n) = \prod_{2\le i<j\le n}(x_j - x_i)$，因此：$\mathrm{V}(\mathrm{x}_1, x_2, \ldots, x_n) = \prod_{1\le i<j\le n}(x_j - x_i)$

□

Example

计算范德蒙德行列式：V(1, 2, 3, 4) = $\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{vmatrix}$

根据范德蒙德行列式的计算公式：V(1, 2, 3, 4) = (2-1)(3-1)(4-1)(3-2)(4-2)(4-3) = 1 ·2 · 3 · 1 · 2 · 1 = 12

2.4 行列式的计算 *：拉普拉斯定理

利用拉普拉斯定理，按前两行展开：

$$\begin{aligned} D = &\begin{vmatrix} 1 & 2 \\ 5 & 6 \end{vmatrix} \cdot (-1)^{1+2+3+4} \begin{vmatrix} 9 & 10 \\ 11 & 12 \end{vmatrix} + \begin{vmatrix} 1 & 3 \\ 5 & 7 \end{vmatrix} \cdot (-1)^{1+2+3+4} \begin{vmatrix} 0 & 10 \\ 0 & 12 \end{vmatrix} \\ &+ \begin{vmatrix} 1 & 4 \\ 5 & 8 \end{vmatrix} \cdot (-1)^{1+2+3+4} \begin{vmatrix} 0 & 9 \\ 0 & 11 \end{vmatrix} + \begin{vmatrix} 2 & 3 \\ 6 & 7 \end{vmatrix} \cdot (-1)^{1+2+3+4} \begin{vmatrix} 0 & 10 \\ 0 & 12 \end{vmatrix} \\ &+ \begin{vmatrix} 2 & 4 \\ 6 & 8 \end{vmatrix} \cdot (-1)^{1+2+3+4} \begin{vmatrix} 0 & 9 \\ 0 & 11 \end{vmatrix} + \begin{vmatrix} 3 & 4 \\ 7 & 8 \end{vmatrix} \cdot (-1)^{1+2+3+4} \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} \\ &= (-4) \cdot (-2) + (-8) \cdot 0 + (-12) \cdot 0 + (-4) \cdot 0 + (-8) \cdot 0 + (-4) \cdot 0 \\ &= 8 \end{aligned}$$

!

Chapter 3

矩阵性质的进一步考察

行列式是研究矩阵性质（如可逆性、秩）的重要工具，以此，我们将引出第二章内容:

3.1 矩阵的初等变换与初等矩阵

3.1.1 矩阵的初等变换

【定义】：矩阵的初等变换（Elementary Matrix Operations）：矩阵的初等变换包含行变换与列变换，其中矩阵的**初等行变换**是指以下三种操作：

1. **交换变换**：交换矩阵的两行，记为 $r_i \leftrightarrow r_j$；
2. **倍乘变换**：用非零常数 $k \neq 0$ 乘以某一行的所有元素，记为 $r_i \leftarrow kr_i$；
3. **倍加变换**：将某一行的 k 倍加到另一行，记为 $r_i \leftarrow r_i + kr_j$。

矩阵的**初等列变换**是指以下三种操作：

1. **交换变换**：交换矩阵的两列，记为 $c_i \leftrightarrow c_j$；
2. **倍乘变换**：用非零常数 $k \neq 0$ 乘以某一列的所有元素，记为 $c_i \leftarrow kc_i$；
3. **倍加变换**：将某一列的 k 倍加到另一列，记为 $c_i \leftarrow c_i + kc_j$。

!

3.1.2 初等矩阵

【定义】：初等矩阵（Elementary matrix）：由单位矩阵 I 经过一次初等变换得到的矩阵称为**初等矩阵**。对应三种初等变换，有三类初等矩阵：

1. **交换矩阵** E_{ij}：交换 I 的第 i 行和第 j 行；

2. **倍乘矩阵** $E_i(k)$：用非零常数 k 乘以 I 的第 i 行；

3. **倍加矩阵** $E_{ij}(k)$：将 I 的第 j 行的 k 倍加到第 i 行。

【初等矩阵的作用】设 A 是 $m \times n$ 矩阵，则：

1. 对 A 施行一次初等行变换，相当于在 A 的左边乘以相应的 m 阶初等矩阵；

2. 对 A 施行一次初等列变换，相当于在 A 的右边乘以相应的 n 阶初等矩阵。

接下来，我们将通过三个例子帮助你理解：
1. 交换两行（以交换第一行和第二行为例）
初等行变换操作：

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \Rightarrow \quad B = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

对应的初等矩阵及乘法操作：

$$E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$E_1 A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = B$$

2. 某一行乘以非零常数（以第一行乘以常数 $k \neq 0$ 为例）
初等行变换操作：

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \Rightarrow \quad C = \begin{pmatrix} ka_{11} & ka_{12} & ka_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

对应的初等矩阵及乘法操作：

$$E_2 = \begin{pmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$E_2 A = \begin{pmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} ka_{11} & ka_{12} & ka_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = C$$

3. 某一行的倍数加到另一行（以将第一行的 k 倍加到第二行为例）
初等行变换操作：

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \Rightarrow \quad D = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} + ka_{11} & a_{22} + ka_{12} & a_{23} + ka_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

对应的初等矩阵及乘法操作：

$$E_3 = \begin{pmatrix} 1 & 0 & 0 \\ k & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$E_3A = \begin{pmatrix} 1 & 0 & 0 \\ k & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} + ka_{11} & a_{22} + ka_{12} & a_{23} + ka_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = D$$

3.2 化阶梯形矩阵与行简化阶梯形矩阵

3.2.1 基本定义

【定义】：行阶梯形矩阵（Row echelon form matrix, REF）：矩阵 A 若满足以下条件：

1. 全零行位于矩阵底部；
2. 非零行的先导元素（首非零元）严格位于前一行先导元素的右侧；
3. 先导元素下方的元素全为零。

则矩阵 A 称为行阶梯形矩阵

Example

典型的行阶梯形矩阵形式：

$$\begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

其中 $*$ 表示非零元素。

行阶梯形矩阵示例：

例 1

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

例 2

$$\begin{pmatrix} 1 & 2 & 0 & 5 \\ 0 & 0 & 3 & -1 \\ 0 & 0 & 0 & 7 \end{pmatrix}$$

非行阶梯形矩阵示例（具有迷惑性）：

例 3（存在全零行不在底部）

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

例 4（主元列位置未严格递增）

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 5 & 6 \end{pmatrix}$$

【定义】：行简化阶梯形矩阵（Reduced Row Echelon Form, RREF）：若行阶梯形矩阵进一步满足：

1. 所有先导元素均为 1；
2. 先导元素所在列的其余元素全为 0，

则称为行简化阶梯形矩阵（Reduced Row Echelon Form, RREF）。

Example

典型的行简化阶梯形矩阵形式：

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

其中 $*$ 表示任意元素。

【定义】：主元位置与主元列（Pivot positions and pivot columns）：矩阵中的主元位置是其行简化阶梯形中先导元素 1 的位置，对应的列称为主元列。

3.2.2 行化简算法

【高斯消元法】(Gaussian elimination method) 每一列 j 选择当前列中第一个非零元素作为主元主元不在当前行 交换行使主元位于当前行用倍乘变换将主元化为 1 下方所有行 i 用倍加变换消去主元下方的元素

【高斯-约旦消元法】(Gauss-Jordan elimination method) 执行高斯消元法得到行阶梯形每一列 j 从后往前 找到主元所在行上方所有行 i 用倍加变换消去主元上方的元素

【Remark】:

高斯消元法通过向前消元得到行阶梯形，高斯-约旦消元法在此基础上通过向后消元得到行简化阶梯形。

3.2.3 化简实例

增广矩阵化简过程：

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 2 & 4 & 1 & 5 \\ -1 & -2 & 2 & 0 \end{array}\right) \xrightarrow{R_2-2R_1} \left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right) \xrightarrow{R_3-\frac{1}{3}R_2} \left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

3.2.4 重要性质

【存在性与唯一性】:

1. 任意矩阵均可通过初等行变换化为行阶梯形矩阵；
2. 每个矩阵行等价于唯一的行简化阶梯形矩阵。

3.3 向量组的秩与极大线性无关组，基扩张问题

在讨论这一章的重点：矩阵的秩之前，我们先简单介绍一下极大线性无关组与向量组的秩，这不仅是初等变换与 RREF 的应用，更有助于理解矩阵的秩的概念。

3.3.1 基本定义

【定义】：极大线性无关组 (maximal linearly independent set)：设向量组 S 的一个部分组 $\boldsymbol{\alpha}_{i_1}, \boldsymbol{\alpha}_{i_2}, \ldots, \boldsymbol{\alpha}_{i_r}$ 满足：

1. 该部分组线性无关；
2. 向量组 S 中任意向量均可由该部分组线性表示，

则称该部分组为向量组 S 的一个极大线性无关组。

【极大线性无关组的等价定义】：设向量组 S 的一个部分组 $\boldsymbol{\alpha}_{i_1}, \boldsymbol{\alpha}_{i_2}, \dots, \boldsymbol{\alpha}_{i_r}$ 线性无关，且将 S 中任一其他向量加入该部分组后所得向量组线性相关，则该部分组为 S 的一个极大线性无关组。

【定义】：向量组的秩（Rank of vector group）：向量组的极大线性无关组所含向量的个数称为该向量组的秩，记为 $r(S)$ 或 $\text{rank}(S)$。

3.3.2 主要性质与定理

Theorem

【向量组秩的性质】

1. 向量组线性无关的充要条件是其秩等于所含向量的个数；
2. 若向量组 A 可由向量组 B 线性表示，则 $r(A) \le r(B)$；

3.3.3 极大线性无关组的计算方法

我们将用矩阵初等行变换法求极大线性无关组

1. 将向量组中的向量按列构成矩阵 A；
2. 对矩阵 A 进行初等行变换化为行阶梯形矩阵；
3. 行阶梯形矩阵的主元所在列对应的原向量组中的向量即为一个极大线性无关组；
4. 非主元列对应的向量可由极大线性无关组线性表示，其系数可由行简化阶梯形矩阵直接读出。

例如求向量组 $\boldsymbol{\alpha}_1 = (1, 2, 3)^T$，$\boldsymbol{\alpha}_2 = (2, 4, 6)^T$，$\boldsymbol{\alpha}_3 = (1, 0, 1)^T$ 的极大线性无关组及秩。

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 0 \\ 3 & 6 & 1 \end{pmatrix} \xrightarrow{\text{初等行变换}} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

主元位于第 1、3 列，故 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_3$ 是一个极大线性无关组，向量组的秩为 2。

3.3.4 等价向量组

【定义】：等价向量组（Equivalent vector groups）：若向量组 $A = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_m\}$ 与 $B = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_n\}$ 满足：

1. A 可由 B 线性表示：

$$\forall \boldsymbol{\alpha}_i \in A, \quad \exists k_{i1}, k_{i2}, \ldots, k_{in} \text{ 使得} \boldsymbol{\alpha}_i = \sum_{j=1}^{n} k_{ij} \boldsymbol{\beta}_j$$

2. B 可由 A 线性表示：

$$\forall \boldsymbol{\beta}_j \in B, \quad \exists l_{j1}, l_{j2}, \ldots, l_{jm} \text{ 使得} \boldsymbol{\beta}_j = \sum_{i=1}^{m} l_{ji} \boldsymbol{\alpha}_i$$

则称 A 与 B 等价，记作 $A \sim B$。

接下来，我们将给一个例子辅助理解：设向量组：

$$A = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}, \quad B = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

验证等价性：

1. A 可由 B 表示：

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

2.B 可由 A 表示：

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

等价向量组具有以下三条基本性质（类比集合的等价关系）：

1. 自反性: 向量组自身与自身等价，即 $A \sim A$

2. 对称性: 若向量组 $A \sim B$，则 $B \sim A$。

3. 传递性: 若向量组 $A \sim B$ 且 $B \sim C$，则 $A \sim C$。

【定理】：等价的向量组具有相同的秩。

【注意】：逆命题不成立（秩相同的向量组不一定等价）

【思考】：其实不仅仅是等价的向量组具有相同的秩，等价矩阵也具有相同的秩，这体现了代数学科自洽的同一性与分类学与代表元的思想

3.3.5 应用：基扩张问题

设

$$V = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 : x_1 + 7x_2 + 5x_3 - 4x_4 + 2x_5 = 0\}.$$

很容易验证 V 是 $\mathbb{R}^5$ 的一个子空间，且

$$S=\{(-2,0,0,-1,-1),(1,1,-2,-1,-1),(-5,1,0,1,1)\}$$

是 V 的一个线性无关子集。我们要将 S 扩充为 V 的一组基

3.3.6 课后练习

!

!

3.4 矩阵的秩 1：基本概念与重要性质

【定义】：矩阵的秩（The rank of a matrix）：矩阵的秩一般定义为：非零子式的最高阶数。

但是这个定义比较抽象，我们实际上对矩阵的秩应该有更加本质的理解：

对于 $m \times n$ 型矩阵 $A \in M_{m\times n}(F)$，其秩 $\operatorname{rank}(A)$ 可基于由 A 诱导的线性变换 $L_A : F^n \to F^m$ 进行定义。具体而言，$\operatorname{rank}(A)$ 等于线性变换 L_A 的秩，即 $\operatorname{rank}(A) = \operatorname{rank}(L_A) = \dim(R(L_A))$。其中，$R(L_A)$ 是 A 的列向量在 F^m 空间中张成的子空间，它反映了线性变换 L_A 输出的“自由度”。

从矩阵视角来看，$\operatorname{rank}(A)$ 等于 A 的列向量组（或行向量组）中最大线性无关组所含向量的个数，这其实就是 A 的列向量在 F^m 空间中张成的子空间。由于矩阵的初等变换不改变矩阵的秩，所以实际上可以通过一系列初等行变换将矩阵化为行阶梯形，进而直观确定最大线性无关组的向量个数，完成矩阵秩的求解。

那么到底怎么求矩阵的秩呢?

3.4.1 求矩阵的秩

由后文我们知道：矩阵的秩和其转置的秩一样，且行（列）初等变换保秩，所以我们可以混用行/列变换求秩（这和以后求方程组只能用初等行变换有所不同）

例 1：设矩阵 $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$。

步骤 1：行化简为阶梯形矩阵

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix} \xrightarrow{R_2-2R_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 3 & 6 & 9 \end{pmatrix} \xrightarrow{R_3-3R_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

步骤 2：确定非零行数量阶梯形矩阵仅有 1 个非零行，故 $\operatorname{rank}(A) = 1$。

方阵的验证方法——行列式法例 1 中，$\det(A) = 0$，说明行向量线性相关，秩小于 3。进一步观察发现，所有行都是第一行的倍数，故秩为 1。

例 2：设矩阵 $B = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 5 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix}$。

步骤 1：行化简为阶梯形矩阵

$$\begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 5 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix} \xrightarrow{R_2-2R_1} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix} \xrightarrow{R_3-R_2} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 3 \end{pmatrix}$$

$$\xrightarrow{R_4-R_1} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_4-R_2} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

步骤 2：确定非零行数量阶梯形矩阵有 2 个非零行，故 $\mathrm{rank}(B) = 2$。

3.4.2 矩阵的秩的重要性质

若 $r(A) = 0 \Leftrightarrow A = 0$。
若 $r(A) \geq 1$，则 $A \neq 0$；
若 A 为 $M_{m\times n}$,则 $r(A) \in b$。 $\min(m, n)$
$r(kA) = r(A)$（$k \neq 0$），推论 $r(-A) = r(A)$。

若 A 为 n 阶方阵，A 可逆 $\Leftrightarrow$ A 中所有列向量线性无关 $\Leftrightarrow$ $|A| \neq 0$。（这个设计了下节逆矩阵的知识，但是为了完整性，我们依然再这里保留）

3.4.3 矩阵的等价与相似

【定义】：矩阵等价/相抵（Matrix equivalence）：对于两个 $m\times n$ 阶矩阵 A 和 B ，若存在 $m\times m$ 阶可逆矩阵 Q 和 $n\times n$ 阶可逆矩阵 P，使得 $B = QAP$ ，则称 A 和 B 等价。等价矩阵本质是可通过有限次初等行变换和初等列变换相互转化，核心性质是秩相等，即 $r(A) = r(B)$ 。比如 $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ ，$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ，$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ，$B = QAP = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ ，A 和 B 秩都为 1 ，二者等价。

矩阵等价/相抵需要满足：

1. **反身性**：$A \cong A$；
2. **对称性**：若 $A \cong B$，则 $B \cong A$；
3. **传递性**：若 $A \cong B$ 且 $B \cong C$，则 $A \cong C$。

【定义】：矩阵相似（Matrix similarity）：设 A、B 为 n 阶方阵，若存在 n 阶可逆矩阵 P，使得 $P^{-1}AP = B$ ，则称 A 与 B 相似。相似矩阵反映同一线性变换在不同基下的矩阵，具有相同特征多项式、特征值、行列式值、迹。例如 $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ ，$P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ ，$P^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ ，$B = P^{-1}AP = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ ，A 和 B 特征值都为 2（二重），二者相似。

二者联系：相似矩阵一定是等价矩阵。因为若 $P^{-1}AP = B$ ，此时 P^{-1} 和 P 相当于等价定义里的可逆矩阵 Q 和 P ，满足等价条件。但等价矩阵不一定是相似矩阵，等价只要求秩相等，对矩阵形状、特征值等没有像相似那样严格的要求。比如 $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ ，$B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ ，秩都为 1 ，二者等价，但特征值不同，不相似。

【定理】任意 $m \times n$ 矩阵 A 都等价于形如

$$\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

的矩阵，其中 I 是单位矩阵。该矩阵称为 A 的**等价标准形**。

【推论 1】：可逆矩阵可分解为有限个初等矩阵的乘积。即，若 A 可逆，则存在初等矩阵 $P_1, P_2, \ldots, P_s$，使得

$$A = P_1 P_2 \cdots P_s.$$

【推论 2】：两个 $m \times n$ 矩阵 A 和 B 等价的充分必要条件是存在可逆的 m 阶矩阵 P 和可逆的 n 阶矩阵 Q，使得 $B = PAQ$。

【定理】矩阵行（列）等价的条件：两个矩阵行（列）等价当且仅当它们具有相同的行（列）简化阶梯形矩阵。

3.5 矩阵的秩 2：重要推论与秩 1 矩阵

3.5.1 重要推论与二级结论

我们将通过一些重要的二级结论的证明，加深你对矩阵的秩，矩阵与线性变换的认识；并且在一定程度上加深你的对”变换”这一思想的理解程度，同时提高你的应试能力。

(1): 若 A 为 $m \times n$, B 为 $n \times s$, 令 $L_A = T$, $L_B = U$, $T: F^n \to F^m$, $U: F^s \to F^n$, 则：$r(UT) \le r(U)$；$r(UT) \le r(T)$；$r(AB) \le r(A)$；$r(AB) \le r(B)$。

推论：$r(A) + r(B) - n \leq r(AB) \leq \min(r(A), r(B))$。

(2)：若 A, B 同为 $m \times n$ 形矩阵，则 $r(A + B) \leq r(A) + r(B)$。(我们可以从两个角度来证明)

(3)：若 A 为 $m \times n$ 型，B 为 $n \times s$，且 $AB = 0$，则 $r(A) + r(B) \leq n$。

(4)：$r(A^TA)=r(AA^T)=r(A)=r(A^T)$（该证明设计可逆矩阵的知识点）

3.5.2 秩 1 矩阵分解

我们用上一节的例 1 作为例子，进行考察：
给定矩阵：

$$A=\begin{pmatrix}1&2&3\\2&4&6\\3&6&9\end{pmatrix}$$

对矩阵 A 进行初等行变换：

1. 第二行减去第一行的 2 倍（r_2-2r_1）：

$$\begin{pmatrix}1&2&3\\2-2\times1&4-2\times2&6-2\times3\\3&6&9\end{pmatrix}=\begin{pmatrix}1&2&3\\0&0&0\\3&6&9\end{pmatrix}$$

2. 第三行减去第一行的 3 倍（r_3-3r_1）：

$$\begin{pmatrix}1&2&3\\0&0&0\\3-3\times1&6-3\times2&9-3\times3\end{pmatrix}=\begin{pmatrix}1&2&3\\0&0&0\\0&0&0\end{pmatrix}$$

行阶梯形矩阵中非零行仅有 1 行，故 $\mathrm{rank}(A)=1$，即 A 是秩 1 矩阵。

由此我们可以得到秩 1 矩阵的分解原理：

若矩阵 B 是秩 1 矩阵，则：

- B 的所有行向量（或列向量）均线性相关。
- 存在非零行向量 $\boldsymbol{\alpha}=(a_1,a_2,\ldots,a_n)$，使得 B 的每一行 $\boldsymbol{\beta}_i$（$i=1,2,\ldots,m$）可表示为 $\boldsymbol{\beta}_i=k_i\boldsymbol{\alpha}$，其中 k_i 为常数。

此时，矩阵 B 可分解为：

$$B=\begin{pmatrix}k_1\\k_2\\\vdots\\k_m\end{pmatrix}\times(a_1,a_2,\ldots,a_n)$$

示例说明：设 $\boldsymbol{\alpha}=(1,2)$，$k_1=3$，$k_2=-1$，则：

$$\boldsymbol{\beta}_1=3\times(1,2)=(3,6),\quad \boldsymbol{\beta}_2=-1\times(1,2)=(-1,-2)$$

矩阵 B 为：

$$B=\begin{pmatrix}3&6\\-1&-2\end{pmatrix}=\begin{pmatrix}3\\-1\end{pmatrix}\times(1,2)$$

对矩阵 A 进行分解：对于矩阵 $A=\begin{pmatrix}1&2&3\\2&4&6\\3&6&9\end{pmatrix}$，观察可得：

- 第一行：$(1,2,3)$
- 第二行：$(2,4,6)=2\times(1,2,3)$
- 第三行：$(3,6,9)=3\times(1,2,3)$

根据分解原理，令 $\boldsymbol{\alpha}=(1,2,3)$，$k_1=1$，$k_2=2$，$k_3=3$，则：

$$A=\begin{pmatrix}1\\2\\3\end{pmatrix}\times(1,2,3)$$

【秩 1 矩阵分解的意义】：

1. **计算简便：**对于两个秩 1 矩阵 $A=\boldsymbol{u}\boldsymbol{v}^T$ 和 $B=\boldsymbol{s}\boldsymbol{t}^T$，其乘积可简化为：

   $$AB=\boldsymbol{u}(\boldsymbol{v}^T\boldsymbol{s})\boldsymbol{t}^T$$

   其中 $\boldsymbol{v}^T\boldsymbol{s}$ 为标量，将矩阵乘法转化为向量与标量的运算。

2. **理解矩阵结构：**通过分解可直观揭示秩 1 矩阵行向量或列向量间的线性相关关系，深入理解矩阵的内在结构。

3.6 可逆矩阵与逆矩阵的初等变换求法

3.6.1 矩阵可逆的定义与条件

【定义】：可逆矩阵（Invertible matrix）：设 A 是 n 阶方阵，若存在 n 阶方阵 B 使得

$$AB = BA = I_n,$$

则称 A 是可逆矩阵（或非奇异矩阵），称 B 为 A 的逆矩阵，记为 A^{-1}。

【定理】：对于 n 阶方阵 A，以下命题等价：

1. A 可逆；
2. A 的行（列）向量组线性无关；换言之：n 阶方阵 A 可逆的充要条件是其列向量组构成 $\mathbb{R}^n$ 的一组基。
3. $r(A) = n$；
4. $\det(A) \neq 0$；
5. A 可表示为有限个初等矩阵的乘积；
6. 齐次线性方程组 $Ax = 0$ 仅有零解；
7. 非齐次线性方程组 $Ax = b$ 有唯一解。

（4，6，7 将在后文详细介绍）

3.6.2 逆矩阵的基本性质

【逆矩阵的唯一性】：若矩阵 A 可逆，则其逆矩阵唯一。

Property

【逆矩阵的运算性质】设 A, B 为同阶可逆矩阵，$k \neq 0$，则：

1. $(A^{-1})^{-1} = A$;
2. $(kA)^{-1} = \frac{1}{k}A^{-1}$;
3. $(AB)^{-1} = B^{-1}A^{-1}$;
4. $(A^T)^{-1} = (A^{-1})^T$;
5. $\det(A^{-1}) = \frac{1}{\det(A)}$;
6. （了解即可）若 A 可逆，则 $A^* = \det(A) \cdot A^{-1}$，其中 A^* 为伴随矩阵（非“内积空间里面的伴随概念”）。

【分块对角矩阵的逆】设 $A = \mathrm{diag}(a_1, a_2, \ldots, a_s)$ 为对角矩阵，其中 a_i 数域 F，则

$$A^{-1} = \mathrm{diag}(a_1^{-1}, a_2^{-1}, \ldots, a_s^{-1}).$$

接着，我们讨论一下初等矩阵的可逆性：

3.6.3 初等矩阵的逆矩阵

初等矩阵都是可逆的，且其逆矩阵仍为同类型的初等矩阵：

1. $E_{ij}^{-1} = E_{ij}$；
2. $[E_i(k)]^{-1} = E_i\left(\frac{1}{k}\right)$；
3. $[E_{ij}(k)]^{-1} = E_{ij}(-k)$。

3.6.4 逆矩阵的初等行变换求法

Theorem

回顾初等矩阵与可逆性，我们知道：

1. 每个初等矩阵都是可逆的，且其逆矩阵仍为同类型的初等矩阵；
2. 方阵 A 可逆的充要条件是存在有限个初等矩阵 $P_1, P_2, \ldots, P_s$，使得 $A = P_1 P_2 \cdots P_s$。

于是我们提出初等行变换求逆矩阵的方法：
【初等行变换求逆矩阵】构造增广矩阵 $[A|I]$，对其施行初等行变换：

$$[A|I] \xrightarrow{\text{初等行变换}} [I|A^{-1}].$$

若在变换过程中 A 无法化为单位矩阵，则 A 不可逆。

Example

例如：求矩阵 $A=\begin{pmatrix}1&2&3\\2&2&1\\3&4&3\end{pmatrix}$ 的逆矩阵。

$$[A|I]=\begin{pmatrix}1&2&3&1&0&0\\2&2&1&0&1&0\\3&4&3&0&0&1\end{pmatrix}$$

经过一系列初等行变换：

$$\xrightarrow{\text{行变换}}\begin{pmatrix}1&0&0&1&3&-2\\0&1&0&-1.5&-3&2.5\\0&0&1&1&1&-1\end{pmatrix}$$

故 $A^{-1}=\begin{pmatrix}1&3&-2\\-1.5&-3&2.5\\1&1&-1\end{pmatrix}$。

【定理】：对于 2×2 矩阵：$A=\begin{pmatrix}a&b\\c&d\end{pmatrix}\implies A^{-1}=\frac{1}{ad-bc}\begin{pmatrix}d&-b\\-c&a\end{pmatrix}$

课后练习：

求矩阵 $B=\begin{pmatrix}1&2&3\\0&1&4\\5&6&0\end{pmatrix}$ 的逆矩阵。

答案：$B^{-1}=\frac{1}{\det(B)}\cdot\mathrm{adj}(B)=\begin{pmatrix}-24&20&-5\\18&-15&4\\5&-4&1\end{pmatrix}$

3.7 （应用）解可逆矩阵 A,B 构成矩阵方程的 $AX = B$

类比 kx=b 的一元一次方程，我们可以类似的提出矩阵方程的概念：对于矩阵方程 $AX = B$，其中 $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$，$B = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$。我们如何得到未知数（未知矩阵 X）的具体表达呢?

$$A^{-1} = \begin{pmatrix} -2 & 1 \\ 1.5 & -0.5 \end{pmatrix}, \quad X = A^{-1}B = \begin{pmatrix} -3 & -4 \\ 2 & 2 \end{pmatrix}.$$

接下来，我们将通过两个例题来增加对这个问题的理解，对其中易错点（是否可逆）的辨析：

例 1：已知 $Ax = 2x + B$，其中 $A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 4 \end{pmatrix}$，$B = \begin{pmatrix} 3 & 6 \\ 1 & 1 \\ 2 & -3 \end{pmatrix}$，求 x。

例 2：如果仅已知 A^{-1}，且 $ABA^{-1} = BA^{-1} + 3E$，求 B。

3.8 （应用）矩阵的 LU 分解与 PLU 分解

3.8.1 上三角矩阵与下三角矩阵

【定义】：上三角矩阵（Upper triangular matrix）：设 $A=(a_{ij})$ 是 n 阶方阵（行数和列数都为 n•），如果当 $i>j$•（i 表示行标，j 表示列标）时，$a_{ij}=0$•，即主对角线（从左上角到右下角这条对角线）下方的元素全部为 0•，则称矩阵 A 为上三角矩阵。例如：$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}$ 是一个三阶上三角矩阵。

【定义】：下三角矩阵（Lower triangular matrix）：设 $B=(b_{ij})$ 是 n 阶方阵，如果当 $i<j$ 时，$b_{ij}=0$•，即主对角线上方的元素全部为 0•，则称矩阵 B 为下三角矩阵。例如：$\begin{pmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$ 是一个三阶下三角矩阵。

【上/下三角矩阵的性质】：这里以上三角矩阵为例，下三角矩阵同理：数乘)：数乘上三角矩阵仍为上三角矩阵。设 A 是上三角矩阵，k 为常数，则 kA 也是上三角矩阵；

加减运算)：两个上三角矩阵的和与差仍是上三角矩阵。若 A、B 是上三角矩阵，则 $A\pm B$ 也是上三角矩阵；

乘法运算)：上三角矩阵的乘积仍是上三角矩阵。若 A、B 是上三角矩阵，则 AB 也是上三角矩阵，进一步：若 A 是上三角矩阵，m 为正整数，则 A^m 也是上三角矩阵；

可逆性)：上三角矩阵可逆当且仅当其主对角线上的所有元素都不为零。可逆上三角矩阵的逆矩阵仍然是上三角矩阵。

上/下三角在未来的舒尔引理（Schur's Lemma）中有相当重要的地位。

3.8.2 矩阵的 LU 分解

我们先来观察一个事实：从矩阵乘法的视角，高斯消元的每一步倍加操作对应左乘初等矩阵 $E_{ij}(m)$。这些初等矩阵的逆 $E_{ij}(-m)$ 仍为下三角，且下三角矩阵乘法封闭。如 $E_{21}(m)E_{31}(n)=\begin{pmatrix} 1 & 0 & 0 \\ m & 1 & 0 \\ n & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}=\begin{pmatrix} 1 & 0 & 0 \\ m & 1 & 0 \\ n & 0 & 1 \end{pmatrix}$，仍为下三角。

【定义与适用条件】若方阵 A 在高斯消元过程中无需行交换，可分解为 $A=LU$，其中 L 为下三角矩阵（lower triangular matrix），其主对角线元素均为 1（则为 Doolittle 分解）或非零（则为 Crout 分解）；U 为上三角矩阵（upper triangular matrix）。

适用条件：高斯消元法（Gaussian elimination）过程中，矩阵 A 的各阶顺序主子式均非零，即 $\det(A_{1:k,1:k})\neq 0$，$k=1,2,\ldots,n$。

【计算原理与步骤】设 A 为 $n\times n$ 矩阵，高斯消元的矩阵形式为：

$$E_{n-1}\cdots E_2E_1A=U$$

其中 E_i 为第二类初等矩阵(倍加矩阵), 表示第 i 步消元操作。则 $L=(E_{n-1}\cdots E_2E_1)^{-1}$, 因倍加矩阵的逆仍为倍加矩阵且下三角矩阵乘积为下三角矩阵, 故 L 为下三角矩阵, 从而 $A=LU$。具体步骤:

1. 消元过程: 对 A 进行行倍加变换, 将主对角线下方元素消为 0, 得到 U。例如, 对于 a_{ij} ($i>j$), 计算乘数 $m_{ij}=a_{ij}/a_{jj}$, 执行 $R_i \leftarrow R_i - m_{ij}R_j$。

2. 记录乘数: 将乘数 m_{ij} 填入 L 的对应位置 (L 主对角线为 1)。

3. 形成分解: 组装 L 和 U, 满足 $A=LU$。

为了更加形象化地理解, 我们将举一个例子详细介绍:

把下列矩阵进行 LU 分解:

$$\begin{bmatrix} 2 & 4 & 2 \\ 1 & 5 & 2 \\ 4 & -1 & 9 \end{bmatrix}$$

虽然 LU 分解在数值线性代数中是一种强大的工具, 但它存在一定的局限性: LU 分解要求矩阵 A 在高斯消元过程中无需进行行交换。而 PLU 分解正是为了克服这些局限性而产生的: PLU 分解通过引入置换矩阵 P, 对矩阵的行进行重新排列, 使得变换后的矩阵 PA 满足分解条件, 从而扩大了可分解矩阵的范围。

我们以这个矩阵为例:

$$E_2E_1A=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 3 \\ 3 & 5 & 9 \end{bmatrix}=\begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & -1 \\ 0 & 2 & 3 \end{bmatrix}$$

此时，不得不进行行变换。

3.8.3 矩阵的 PLU 分解

置换与置换矩阵

Definition

置换 π 是集合 $\{1,\ldots,n\}$ 的双射。置换矩阵 P 由单位矩阵 I 经行置换得到，即 $P=\begin{pmatrix} e_{\pi(1)}^T \\ \vdots \\ e_{\pi(n)}^T \end{pmatrix}$，其中 e_i^T 为 I 的第 i 行。

性质：P 为正交矩阵，$P^TP=I$，$\det(P)=\pm1$。

PLU 分解定理

Theorem

对任意 $n\times n$ 方阵 A，存在置换矩阵 P、单位下三角矩阵 L（主对角线为 1）和上三角矩阵 U，使得 $PA=LU$。

构造性证明与步骤

1. 选主元操作：高斯消元时，若需行交换，记录置换矩阵 P_i。例如，第 k 步选主元 $\max_{i\geq k}|a_{ik}|$，交换行得到 P_k。
2. 累积置换矩阵：令 $P=P_m\cdots P_2P_1$，则 PA 在消元时无需再行交换。
3. 执行 LU 分解：对 PA 进行常规 LU 分解，$PA=LU$，其中 L 由消元乘数构成，U 为上三角结果。

回顾上节遗留的问题，我们对它进行 PLU 分解：

$$A=\begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 3 \\ 3 & 5 & 9 \end{bmatrix}$$

Chapter 4

n 元一次线性方程组

线性方程组是线性代数的核心研究对象之一，对其解的结构的研究贯穿了整个线性代数理论。本章将系统地讨论 n 元一次线性方程组，包括齐次线性方程组和非齐次线性方程组，并通过代表性例题加深理解。

实现我们将提出三大议题，也就是线性方程组的基本问题：

个数问题: 有没有解? 有几个解?

解法问题: 能否给出求解的算法?

结构问题: 如果解不唯一, 解集结构如何?

我们会在接下来的两章中详细讨论。

4.1 齐次线性方程组解的结构

4.1.1 基本概念与定义

齐次线性方程组是指常数项全为零的线性方程组，其一般形式为：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0 \end{cases} \tag{4.1}$$

可以写成矩阵形式 $Ax = 0$，其中 $A = (a_{ij})$ 是系数矩阵，$x = (x_1, x_2, \ldots, x_n)^T$ 是未知向量。

齐次线性方程组总是有解的，因为 $x = 0$（零向量）一定是它的解，称为零解或平凡解。因此，对齐次线性方程组的研究主要集中在非零解的存在性及其结构上。

【定理】：齐次方程组 $Ax = 0$ 有非零解的等价条件是：

设 A 为 $m \times n$ 矩阵，$Ax = 0$ 有非零解 $\iff$ A 的核空间不为空 $\iff$ $\text{rank}(A) < n$（未知数个数）$\iff$ $\det(A) = 0$（如果 A 为方阵）$\iff$ A 的列向量组线性相关。

4.1.2 解的性质与结构定理

齐次线性方程组的解具有以下重要性质：解的线性组合仍为解若 ξ_1 和 ξ_2 是齐次线性方程组 $Ax=0$ 的解，则它们的任意线性组合 $k_1\xi_1+k_2\xi_2$ 也是该方程组的解，其中 k_1,k_2 是任意常数。

证明. 已知 $A\xi_1=0$ 和 $A\xi_2=0$，则对于任意常数 k_1,k_2，有：

$$A(k_1\xi_1+k_2\xi_2)=k_1A\xi_1+k_2A\xi_2=k_1\cdot 0+k_2\cdot 0=0$$

因此，$k_1\xi_1+k_2\xi_2$ 是 $Ax=0$ 的解。 □

这一性质表明，齐次线性方程组的全体解构成一个向量空间，称为该方程组的解空间，记作 $N(A)$（A 的零空间）。

【定理】：设齐次线性方程组 $Ax=0$ 的系数矩阵 A 的秩为 r，则其解空间 $N(A)$ 的维数为 $n-r$，即：

$$\dim(N(A))=n-r$$

其中 n 是未知量的个数。

证明. 对系数矩阵 A 进行初等行变换化为行最简形矩阵 R，则 $Ax=0$ 与 $Rx=0$ 同解。行最简形矩阵 R 有 r 个非零行，对应 r 个主元变量，其余 $n-r$ 个变量为自由变量。给每个自由变量赋一组值（例如，分别取 1 和 0），可以得到 $n-r$ 个线性无关的解向量，这些解向量构成解空间的一个基，因此解空间的维数为 $n-r$。

当然用维度定理理解就非常直观，这再一次显示了几何的重要性。

□

【推论】：n 个未知数至少需要 n 个方程确定，即如果行数 m <未知量个数 n，则必定有非零解。

【点击】：这本质上是一个自由度问题，n 个未知量，n 个自由度，要唯一确定这些自由度，必须有至少 n 个的限制。

Definition

【定义】：基础解系（fundamental system of solutions）：齐次线性方程组 $Ax=0$ 的解空间 $N(A)$ 的一个基称为该方程组的一个基础解系。具体地，若 $\xi_1,\xi_2,\ldots,\xi_{n-r}$ 是 $Ax=0$ 的解，且满足：

1. $\xi_1,\xi_2,\ldots,\xi_{n-r}$ 线性无关；

2. $Ax=0$ 的任意解都可以表示为 $\xi_1,\xi_2,\ldots,\xi_{n-r}$ 的线性组合，

其实就是可张成解空间的最大线性无关组

则称 $\xi_1,\xi_2,\ldots,\xi_{n-r}$ 是 $Ax=0$ 的一个基础解系。

4.1.3 求解齐次线性方程组的步骤

求解齐次线性方程组 $Ax = 0$ 的一般步骤如下:

1. 对系数矩阵 A 进行初等行变换，化为行最简形矩阵 R;
2. 确定主元变量和自由变量;
3. 对每个自由变量分别赋值，组合，提取参数，得到对应的解向量;
4. 这些解向量构成基础解系，方程组的通解即为基础解系的线性组合。

4.1.4 代表性例题

Example

求齐次线性方程组的通解：

$$\begin{cases} x_1 + x_2 - x_3 + 2x_4 = 0 \\ 2x_1 + 2x_2 - x_3 + 3x_4 = 0 \\ 3x_1 + 3x_2 - x_3 + 4x_4 = 0 \end{cases} \tag{4.2}$$

首先写出系数矩阵并进行初等行变换：

$$A = \begin{pmatrix} 1 & 1 & -1 & 2 \\ 2 & 2 & -1 & 3 \\ 3 & 3 & -1 & 4 \end{pmatrix} \xrightarrow{\text{行变换}} \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

行最简形矩阵对应的方程组为：

$$\begin{cases} x_1 + x_2 + x_4 = 0 \\ x_3 + x_4 = 0 \end{cases}$$

选取 x_2 和 x_4 为自由变量，令 $x_2 = s$，$x_4 = t$，则：

$$\begin{cases} x_1 = -s - t \\ x_2 = s \\ x_3 = -t \\ x_4 = t \end{cases}$$

写成向量形式：

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = s \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

因此，基础解系为：

$$\xi_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

通解为：

$$x = k_1\xi_1 + k_2\xi_2 \quad (k_1, k_2 \in \mathbb{R})$$

接下来，我们将研究一例 n 元 n 行的方程组：

Example

设齐次线性方程组：

$$\begin{cases} (\lambda+3)x_1 + x_2 + 2x_3 = 0 \\ \lambda x_1 + (\lambda-1)x_2 + x_3 = 0 \\ 3(\lambda+1)x_1 + \lambda x_2 + (\lambda+3)x_3 = 0 \end{cases} \tag{4.3}$$

问 λ 取何值时，方程组有非零解？并求其通解。
系数矩阵为：

$$A = \begin{pmatrix} \lambda+3 & 1 & 2 \\ \lambda & \lambda-1 & 1 \\ 3(\lambda+1) & \lambda & \lambda+3 \end{pmatrix}$$

计算行列式 $\det(A)$：

$$\det(A) = \lambda^2(\lambda-1)$$

当 $\det(A)=0$ 时，即 $\lambda=0$ 或 $\lambda=1$ 时，方程组有非零解。
当 $\lambda=0$ 时，系数矩阵变为：

$$A = \begin{pmatrix} 3 & 1 & 2 \\ 0 & -1 & 1 \\ 3 & 0 & 3 \end{pmatrix} \xrightarrow{\text{行变换}} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

通解为：

$$x = k\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \quad (k \in \mathbb{R})$$

当 $\lambda=1$ 时，系数矩阵变为：

$$A = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 0 & 1 \\ 6 & 1 & 4 \end{pmatrix} \xrightarrow{\text{行变换}} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$$

通解为：

$$x = k\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \quad (k \in \mathbb{R})$$

接下来，我们可以尝试考察一下线性方程组的几何意义（如有）考虑三维空间中的

齐次线性方程组：

$$\begin{cases} x+y+z=0 \\ x-y+z=0 \end{cases} \tag{4.4}$$

解释其解的几何意义。

对系数矩阵进行初等行变换：

$$A=\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \xrightarrow{\text{行变换}} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

通解为：

$$x=k\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \quad (k\in\mathbb{R})$$

几何上，每个方程表示一个过原点的平面，方程组的解是这两个平面的交线，即过原点且方向向量为 $(-1,0,1)$ 的直线。

!

4.2 非齐次线性方程组解的结构

4.2.1 基本概念与定义

非齐次线性方程组是指常数项不全为零的线性方程组，其一般形式为：

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1 \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2 \\ \vdots \\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m \end{cases} \tag{4.5}$$

可以写成矩阵形式 $Ax=b$，其中 $A=(a_{ij})$ 是系数矩阵，$x=(x_1,x_2,\dots,x_n)^T$ 是未知向量，$b=(b_1,b_2,\dots,b_m)^T\neq 0$。

对应的齐次线性方程组 $Ax=0$ 称为非齐次线性方程组 $Ax=b$ 的导出组。

4.2.2 解的存在性与判定定理

非齐次线性方程组 $Ax=b$ 的解的存在性由以下定理给出：

!

证明. 对增广矩阵 $\overline{A}$ 进行初等行变换化为行阶梯形矩阵。若 $\operatorname{rank}(A)<\operatorname{rank}(\overline{A})$，则行阶梯形矩阵中存在形如 $0=1$ 的矛盾方程，故方程组无解。反之，若 $\operatorname{rank}(A)=\operatorname{rank}(\overline{A})$，则行阶梯形矩阵中不存在矛盾方程，方程组有解。

当 $\text{rank}(A) = \text{rank}(\overline{A}) = n$ 时，行阶梯形矩阵对应的方程组为严格三角形方程组，有唯一解；当 $\text{rank}(A) = \text{rank}(\overline{A}) < n$ 时，存在自由变量，故有无穷多解。 □

4.2.3 解的结构定理

非齐次线性方程组的解与其导出组的解之间存在密切的关系：设 η 是非齐次线性方程组 $Ax = b$ 的一个特解，ξ 是其导出组 $Ax = 0$ 的通解，则 $Ax = b$ 的通解为：

$$x = \eta + \xi$$

即非齐次线性方程组的通解等于它的一个特解加上其导出组（对应的齐次方程组）的通解。

证明. 首先证明 $\eta + \xi$ 是 $Ax = b$ 的解：

$$A(\eta + \xi) = A\eta + A\xi = b + 0 = b$$

因此，$\eta + \xi$ 是 $Ax = b$ 的解。

其次证明 $Ax = b$ 的任意解都可以表示为 $\eta + \xi$ 的形式。设 x 是 $Ax = b$ 的任意解，则：

$$A(x - \eta) = Ax - A\eta = b - b = 0$$

因此，$x - \eta$ 是导出组 $Ax = 0$ 的解，即 $x - \eta = \xi$，从而 $x = \eta + \xi$。 □

4.2.4 求解非齐次线性方程组的步骤

求解非齐次线性方程组 $Ax = b$ 的一般步骤如下：

1. 对增广矩阵 $\overline{A} = b$ 行初等行变换，化为行最简形矩阵；

判断方程组是否有解（即 $\text{rank}(A) = \text{rank}(\overline{A})$）；

若有解，求出导出组 $Ax = 0$ 的通解 ξ；

求出非齐次方程组 $Ax = b$ 的一个特解 η（通常令自由变量全为 0）；

写出通解 $x = \eta + \xi$。

4.2.5 代表性例题

Example

求解非齐次线性方程组：

$$\begin{cases} x_1 + x_2 - x_3 + 2x_4 = 1 \\ 2x_1 + 2x_2 - x_3 + 3x_4 = 3 \\ 3x_1 + 3x_2 - x_3 + 4x_4 = 5 \end{cases} \tag{4.6}$$

增广矩阵为：

$$\overline{A} = \begin{pmatrix} 1 & 1 & -1 & 2 & 1 \\ 2 & 2 & -1 & 3 & 3 \\ 3 & 3 & -1 & 4 & 5 \end{pmatrix} \xrightarrow{\text{行变换}} \begin{pmatrix} 1 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

对应的方程组为：

$$\begin{cases} x_1 + x_2 + x_4 = 2 \\ x_3 + x_4 = 1 \end{cases}$$

选取 x_2 和 x_4 为自由变量，令 $x_2 = s$，$x_4 = t$，则：

$$\begin{cases} x_1 = 2 - s - t \\ x_2 = s \\ x_3 = 1 - t \\ x_4 = t \end{cases}$$

写成向量形式：

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + s \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

其中 $\begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ 是非齐次方程组的一个特解，后面两项是导出组的通解。

Example

讨论 λ 取何值时，非齐次线性方程组：

$$\begin{cases} \lambda x_1 + x_2 + x_3 = 1 \\ x_1 + \lambda x_2 + x_3 = \lambda \\ x_1 + x_2 + \lambda x_3 = \lambda^2 \end{cases} \tag{4.7}$$

有唯一解、无解或有无穷多解？并在有无穷多解时求其通解。
增广矩阵为：

$$\overline{A} = \begin{pmatrix} \lambda & 1 & 1 & 1 \\ 1 & \lambda & 1 & \lambda \\ 1 & 1 & \lambda & \lambda^2 \end{pmatrix}$$

计算系数矩阵的行列式：

$$\det(A) = (\lambda - 1)^2(\lambda + 2)$$

当 $\lambda \neq 1$ **且** $\lambda \neq -2$ **时**，$\det(A) \neq 0$，方程组有唯一解。
当 $\lambda = 1$ **时**，增广矩阵变为：

$$\overline{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \xrightarrow{\text{行变换}} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

对应的方程组为 $x_1 + x_2 + x_3 = 1$，有无穷多解。通解为：

$$x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + s\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \quad (s, t \in \mathbb{R})$$

当 $\lambda = -2$ **时**，增广矩阵变为：

$$\overline{A} = \begin{pmatrix} -2 & 1 & 1 & 1 \\ 1 & -2 & 1 & -2 \\ 1 & 1 & -2 & 4 \end{pmatrix} \xrightarrow{\text{行变换}} \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

此时 $\text{rank}(A) = 2$，$\text{rank}(\overline{A}) = 3$，方程组无解。

Example

[特解与通解的关系] 设四元非齐次线性方程组 $Ax = b$ 的系数矩阵 A 的秩为 3，已知它的三个解向量为 η_1, η_2, η_3，其中：

$$\eta_1 = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \quad \eta_2 + \eta_3 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

求该方程组的通解。

由于 $\operatorname{rank}(A) = 3$, 未知量个数 $n = 4$, 所以导出组 $Ax = 0$ 的解空间维数为 $4-3 = 1$。注意到 η_1 是非齐次方程组的一个特解，只需找到导出组的一个非零解。考虑：

$$\xi = 2\eta_1 - (\eta_2 + \eta_3) = 2\begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 5 \\ 6 \end{pmatrix}$$

容易验证 ξ 是导出组 $Ax = 0$ 的解，且 $\xi \neq 0$，因此 ξ 构成导出组解空间的一个基。于是，非齐次方程组的通解为：

$$x = \eta_1 + k\xi = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} + k\begin{pmatrix} 3 \\ 4 \\ 5 \\ 6 \end{pmatrix} \qquad (k \in \mathbb{R})$$

接下来，我们再展示一个例题，以加深理解的程度：

求解非齐次线性方程组：$\begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ x_2 + 4x_3 = 2 \\ 5x_1 + 6x_2 = 3 \end{cases}$

Chapter 5

对角化

在这一章，我们将此最简单的**线性算子**及其矩阵表示的**对角化**讲起，由前言所述：我们希望找到 V 的另一组基 β，使得 T 在 β 下的矩阵表示为对角阵 D。根据基变换的相似性原理，设从基 α 到 β 的过渡矩阵为 Q，则矩阵 A 与 D 满足相似关系：

$$D = Q^{-1}AQ$$

为了方便起见，假设我们已经拥有了一个可对角化的矩阵 A，和可逆阵 $Q = (\alpha_1, \alpha_2, \cdots, \alpha_n)$ 为其列分块。

由 $Q^{-1}AQ = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$，可得

$$AQ = Q\mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$$

即

$$(A\alpha_1, A\alpha_2, \cdots, A\alpha_n) = (\lambda_1\alpha_1, \lambda_2\alpha_2, \cdots, \lambda_n\alpha_n)$$

进而推出 $A\alpha_i = \lambda_i\alpha_i$，说明 α_i 是属于带有对角阵特征的数值 λ_i 的对应的特殊向量。

由此，我们受到启发，导引出**特征值与特征向量**的概念。

5.1 特征值与特征向量

5.1.1 线性变换的特征值与特征向量

Definition

线性变换的特征值与特征向量（Eigenvalues and eigenvectors of linear transformations）：设 V 是数域 $\mathbb{F}$ 上的线性空间，$T \in \mathcal{L}(V)$ 是定义在 V 上的线性变换（算子）。若存在非零向量 $v \in V$ 和标量 $\lambda \in \mathbb{F}$，使得

$$T(v) = \lambda v,$$

则称 λ 为 T 的**特征值**，v 为属于 λ 的**特征向量**。

【Remark】:

特征向量不为 0 向量;

特征向量在线性变换下**仅发生缩放**，方向保持不变（或反向）。几何上，特征向量定义了线性变换的" **不变方向**"。

Fact or Background

对于任意一个对应于特征值 λ 的特征向量，**其任意倍数的特征向量**还是一个对应于特征值 λ 的特征向量。即特征向量收缩变换后，依然为特征向量。

【Remark】:

但是特征值不是，**特征值是是由算子唯一确定的性质，具有" 不变性"**（特征值是算子在相似变换下的不变量）。

因此：同时一个特征值 λ 可以对应无穷的特征向量，但是一个特征向量只能对应一个特征值 λ。

同样的，我们可以定义矩阵的特征值与特征向量：

Definition

对于 n 阶方阵 A，若存在数 λ 和 n 维非零列向量 α，使得 $A\alpha = \lambda\alpha$ 成立，则称非零向量 α 是方阵 A 对应于特征值 λ 的一个特征向量。

接下来我们将通过对特征多项式的讨论，构建起**线性算子 T** 与**矩阵的特征值**，**特征向量**的内在联系，我们将看到数学概念优美的自洽性!

让我们先从上一个论述结尾的矩阵 A 起手，做一些分析与探讨。

经过移项，我们得到式子：$(A - \lambda I)v = 0$。而线性方程组 $(A - \lambda I)v = 0$ 有非零解的充要条件：

$$\det(\lambda I - A) = 0.$$

这一条件确保了线性变换 $A - \lambda I$ 不可逆，从而存在非零向量 v 使得 $(A - \lambda I)v = 0$。所以，我们将导入矩阵的特征多项式的概念：

Definition

（矩阵的）**特征多项式**（Characteristic polynomial）：矩阵 A 的**特征多项式**定义为

$$p_A(x) = \det(xI_{id} - A),$$

特征多项式根 x_i 的重数称为其**代数重数**（Algebraic multiplicity），记为 $AM(\lambda)$ 或 $m_g(\lambda)$。

我们有如下重要结论：

Fact or Background

矩阵 A 的特征值恰好是其特征多项式 $\mathfrak{p}_A(x)$ 在 $\mathbb{F}$ 中的根。

证明. 根据特征值定义，λ 是 A 的特征值 $\iff$ 存在非零向量 $\boldsymbol{x} \neq \mathbf{0}$，使得：

$$A\boldsymbol{x} = \lambda\boldsymbol{x}.$$

移项变形为齐次线性方程组的形式：

$$(A - \lambda I)\boldsymbol{x} = \mathbf{0},$$

其中 $A - \lambda I$ 是 $\mathbb{F}$ 上的 n 阶矩阵（将 A 的对角元减去 λ）。

必要性（$\Rightarrow$）：特征值 $\implies$ 特征多项式根假设 λ 是 A 的特征值，需证明 $\mathfrak{p}_A(\lambda) = 0$，步骤如下：

1. 由特征值定义，存在非零向量 $\boldsymbol{x} \neq \mathbf{0}$ 满足 $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$，即齐次线性方程组 $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$ 有非零解；

2. 对 n 阶矩阵，齐次线性方程组有非零解的充要条件是系数矩阵 $A - \lambda I$ 不可逆（否则方程组只有零解）；

3. 矩阵可逆的充要条件是其行列式非零（可逆矩阵 $\iff$ 满秩矩阵 $\iff$ 行列式 $\neq 0$），因此 $A - \lambda I$ 不可逆 $\implies \det(A - \lambda I) = 0$；

4. 结合特征多项式的定义 $\mathfrak{p}_A(x) = \det(xI - A)$，利用行列式性质 $\det(-M) = (-1)^n \det(M)$（$n$ 是矩阵阶数），可得：

$$\mathfrak{p}_A(\lambda) = \det(\lambda I - A) = \det(-(A - \lambda I)) = (-1)^n \det(A - \lambda I);$$

5. 由于 $\det(A - \lambda I) = 0$，因此 $\mathfrak{p}_A(\lambda) = 0$，即 λ 是特征多项式 $\mathfrak{p}_A(x)$ 的根。

充分性（$\Leftarrow$）：特征多项式根 $\implies$ 特征值假设 λ 是 $\mathfrak{p}_A(x)$ 的根（即 $\mathfrak{p}_A(\lambda) = 0$），需证明 λ 是 A 的特征值，步骤如下：

1. 由 $\mathfrak{p}_A(\lambda)=0$，结合必要性中的行列式变形：
$$0=\mathfrak{p}_A(\lambda)=(-1)^n\det(A-\lambda I),$$
因此 $\det(A-\lambda I)=0$；

2. 行列式为零的 n 阶矩阵不可逆，因此齐次线性方程组 $(A-\lambda I)\boldsymbol{x}=\mathbf{0}$ 有非零解（可逆矩阵的齐次方程组只有零解，不可逆则存在非零解）；

3. 取该方程组的一个非零解 $\boldsymbol{x}\neq\mathbf{0}$，则 $(A-\lambda I)\boldsymbol{x}=\mathbf{0}$，即 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，满足特征值的定义。

□

类似的，我们给出线性变换版本的特征多项式：

Definition

定义（线性变换的特征多项式与代数重数） 设 $T:V\to V$ 是域 $\mathbb{F}$ 上有限维向量空间 V 上的线性变换，取 V 的基 $\mathcal{B}$，$[T]_{\mathcal{B}}$ 为 T 在 $\mathcal{B}$ 下的 $n\times n$ 矩阵表示（$n=\dim_{\mathbb{F}}V$）。
- 特征多项式：$\mathfrak{p}_T(\lambda)=\det(\lambda I-[T]_{\mathcal{B}})$，其中 I 为同阶单位矩阵，$\det(\cdot)$ 表示行列式（注：相似矩阵特征多项式相同，定义唯一）；
- 代数重数：设 $\lambda_0\in\mathbb{F}$ 是 T 的特征值（即 $\mathfrak{p}_T(\lambda_0)=0$），则 λ_0 作为 $\mathfrak{p}_T(\lambda)$ 的根的重数，称为 λ_0 的代数重数，记为 $\mathrm{AM}(\lambda_0)$、$m_a(\lambda_0)$ 或 $m_g(\lambda_0)$。

【Remark】：

如果在复数域上，根据代数基本定理，特征多项式必定可分裂（*split*），故必有 n 个根。

因此，我们再拓展到代数闭域上，我们有以下推论：

Corollary & Secondary Conclusion

若 $\mathbb{F}$ 是代数闭域（例如 $\mathbb{C}$），则对任意有限维 $\mathbb{F}$-线性映射 $T:V\to V$（即 $\dim_{\mathbb{F}}(V)<\infty$），$T$ 必存在特征值和特征向量。

由此，我们可以介绍如下的求解特征值与特征向量的方法：

Method

1. 计算特征多项式 $p_A(\lambda)=\det(\lambda I-A)=det(A-\lambda I)$；
2. 求解特征方程 $p_A(\lambda)=0$，得到所有特征值 $\lambda_1,\lambda_2,\ldots,\lambda_n$；

例如：

Example

设线性变换 T 在基 β 下的矩阵表示为 A：

$$A = [T]_\beta = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$$

根据特征多项式的定义 $p_T(\lambda) = \det(\lambda I - A)$　I 是 3×3 单位矩阵：

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

计算矩阵 $\lambda I - A$：

$$\lambda I - A = \begin{pmatrix} \lambda - 1 & -2 & -3 \\ 0 & \lambda - 4 & -5 \\ 0 & 0 & \lambda - 6 \end{pmatrix}$$

行列式计算过程：

$$\begin{aligned} p_T(\lambda) &= \det \begin{pmatrix} \lambda - 1 & -2 & -3 \\ 0 & \lambda - 4 & -5 \\ 0 & 0 & \lambda - 6 \end{pmatrix} \\ &= (\lambda - 1)(\lambda - 4)(\lambda - 6) \\ &= \lambda^3 - 11\lambda^2 + 34\lambda - 24 \end{aligned}$$

特征值为特征方程 $p_T(\lambda) = 0$ 的根：

$$\lambda_1 = 1, \quad \lambda_2 = 4, \quad \lambda_3 = 6$$

由于三个特征值均为单根，其代数重数分别为：

$$AM(\lambda_1) = AM(\lambda_2) = AM(\lambda_3) = 1$$

通过这个例子，你应该对矩阵特征多项式有了一个基本的认识，当然如果仔细考察，我们发现特征多项式还有一个良好的性质：**特征多项式具有与矩阵表示的独立性.**

Property

特征多项式 $p_T(\lambda)$ 不依赖于基的选择，因此特征值是线性变换 T 本身的性质，与矩阵表示无关。

证明. 设 β 和 γ 是 V 的两组基，Q 是从 β 到 γ 的过渡矩阵，则

$$[T]_\gamma = Q^{-1}[T]_\beta Q.$$

因此

$$\det(\lambda I - [T]_\gamma) = \det(\lambda I - Q^{-1}[T]_\beta Q) = \det(Q^{-1}(\lambda I - [T]_\beta)Q) = \det(\lambda I - [T]_\beta).$$

即特征多项式与基的选择无关。 □

所以相似矩阵有一样的特征多项式，进而有一样的特征值。

故在同一个向量空间 V 内，虽然线性算子 T 在不同基底下的矩阵表示不同，但是因为矩阵特征多项式乃至特征值与基底选择无关，那么我们能不能定义线性算子 T 的特征多项式就是其在任意一组基底下矩阵表达的矩阵特征多项式？为了进一步地讲清楚，接下来我们将证明本节的主定理：

Theorem

设 $A \in M_n(\mathbb{F})$ 是线性变换 T 在基 β 下的矩阵表示，则 λ 是 T 的特征值当且仅当 λ 是矩阵 A 的特征值。
其中，判定条件为：$\det(\lambda I - A) = 0$，这是 A 与 T 共享的特征方程。

为了证明这个定理，我们需要从变换的视角来看待，为此，我们将结合下图，从两个方向分别证明充分性与必要性：

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, "\phi_\beta"'] & V \arrow[d, "\phi_\beta"] \\
\mathsf{F}^n \arrow[r, "\mathsf{L}_A"] & \mathsf{F}^n
\end{tikzcd}$$

证明. **从线性变换到矩阵：** 若 $v \in V$ 是 T 属于 λ 的特征向量（$T(v) = \lambda v$），设 $[v]_\beta$ 是 v 在基 β 下的坐标向量，则我们要证明：$[v]_\beta$ 是 $A = [T]_\beta$ 属于 λ 的特征向量。

推导：对 $T(v) = \lambda v$ 两边用过渡矩阵（同构变换）ϕ_β 作用，得 $\phi_\beta(T(v)) = \phi_\beta(\lambda v)$。由 ϕ_β 线性及 $A = [T]_\beta$，有 $A[v]_\beta = \lambda[v]_\beta$。

又 $v \neq 0$ 且 ϕ_β 是同构，故 $[v]_\beta \neq 0$，即 $[v]_\beta$ 是 A 的特征向量。

从矩阵到线性变换：若 $y \in \mathbb{F}^n$ 是 A 属于 λ 的特征向量（$Ay = \lambda y$），则我们要证明：$\phi_\beta^{-1}(y)$ 是 T 属于 λ 的特征向量。

推导：令 $v = \phi_\beta^{-1}(y)$，则 $[T(v)]_\beta = A[v]_\beta = Ay = \lambda y = \lambda[v]_\beta$。

又因 ϕ_β 是同构，故 $T(v) = \lambda v$，即 $v = \phi_\beta^{-1}(y)$ 是 T 的特征向量。

证明完毕。 □

接下来我们讲仔细考察特征值与特征向量的具体性质：

特征向量的线性无关性

Theorem

属于不同特征值的特征向量线性无关。即若 $\lambda_1, \lambda_2, \ldots, \lambda_k$ 是 A 的不同特征值，$v_1, v_2, \ldots, v_k$ 是对应的特征向量，则 $v_1, v_2, \ldots, v_k$ 线性无关。

归纳法证明. 1. 当 $k = 1$ 时，单个非零向量线性无关，命题成立。

2. 假设对于 $k = m$ 命题成立，即 $v_1, v_2, \ldots, v_m$ 线性无关。

3. 考虑 $k = m + 1$ 的情况，设存在标量 $c_1, c_2, \ldots, c_{m+1}$ 使得

$$c_1 v_1 + c_2 v_2 + \cdots + c_{m+1} v_{m+1} = 0.$$

用 A 作用两边，得

$$c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \cdots + c_{m+1} \lambda_{m+1} v_{m+1} = 0.$$

将第一个方程乘以 λ_{m+1} 并与第二个方程相减，得

$$c_1(\lambda_1 - \lambda_{m+1})v_1 + \cdots + c_m(\lambda_m - \lambda_{m+1})v_m = 0.$$

由归纳假设，$v_1, \ldots, v_m$ 线性无关，故

$$c_i(\lambda_i - \lambda_{m+1}) = 0, \quad i = 1, 2, \ldots, m.$$

由于 $\lambda_i \neq \lambda_{m+1}$，所以 $c_1 = \cdots = c_m = 0$，代入原方程得 $c_{m+1} = 0$。

由归纳法，命题对任意 k 成立。 □

我们还可以进一步窥探：

Property

设 $A \in M_n(\mathbb{F})$，$\lambda_1, \lambda_2, \ldots, \lambda_n$ 是其特征值（计重数），则：

1. $\sum_{i=1}^{n} \lambda_i = \operatorname{tr}(A)$；
2. $\prod_{i=1}^{n} \lambda_i = \det(A)$；
3. A 可逆当且仅当其所有特征值非零；
4. 若 λ 是 A 的特征值，则 λ^k 是 A^k 的特征值，对应的特征向量相同；（对任意的 $k = -1$（逆矩阵）或 k 为正整数）
5. 若 A 满足多项式 $f(A) = 0$，则 A 的每个特征值 λ 满足 $f(\lambda) = 0$。

5.1.2 特征子空间与特征值的几何重数

在这一节中，我们会先说明特征子空间这一重要的概念，同时，对特征多项式的性质做进一步窥探：

Definition

特征子空间（Eigenspace）： 对于特征值 λ，其对应的**特征子空间**定义为

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

这是 V 的一个子空间，其维数称为 λ 的**几何重数（geometric multiplicity）**，记为 $GM(\lambda)$ 或 $m_g(\lambda)$。

对于每个特征值 λ_i，求解齐次线性方程组 $(\lambda_i I - A)x = 0$，其非零解即为属于 λ_i 的特征向量。其基础解系构成特征子空间

我们以一个 2 阶矩阵的特征值与特征向量来作为例子，以更好地理解：

设矩阵 $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$，计算其特征值与特征向量：

$$p_A(\lambda) = \det \begin{pmatrix} \lambda - 2 & -1 \\ -1 & \lambda - 2 \end{pmatrix} = (\lambda - 3)(\lambda - 1),$$

特征值为 $\lambda_1 = 3$ 和 $\lambda_2 = 1$。

- 对于 $\lambda_1 = 3$，解方程组 $(3I - A)x = 0$，得基础解系 $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$，故 $E_3 = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$；
- 对于 $\lambda_2 = 1$，解方程组 $(I - A)x = 0$，得基础解系 $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$，故 $E_1 = \operatorname{span}\left\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\}$。

对于某个特征值，其代数重数与几何重数有如下关系：

Property

对于任意特征值 λ，恒有 $1 \leq GM(\lambda) \leq AM(\lambda)$。

证明. 设 $d = m_g(\lambda) = \dim\ker(T - \lambda I)$ 为 λ 的几何重数。

根据定义，这意味着特征子空间 $E_\lambda = \ker(T - \lambda I)$ 是 V 的一个 d- 维子空间。我们的目标是证明特征多项式 $\mathfrak{p}_T(x)$ 有一个因子 $(x-\lambda)^d$，这将蕴含 $d \leq m_a(\lambda)$。首先，选取一个合适的基设 $\{v_1, v_2, \ldots, v_d\}$ 为特征子空间 E_λ 的一组基。我们可以将这组基扩充为整个空间 V 的一组基 $\mathcal{B} = \{v_1, \ldots, v_d, w_{d+1}, \ldots, w_n\}$。

接下来，计算 T 关于基 $\mathcal{B}$ 的矩阵表示对于 $1 \leq i \leq d$，由于 v_i 是对应于 λ 的特征向量，我们有：$T(v_i) = \lambda v_i = \lambda v_i + 0 \cdot v_1 + \cdots + 0 \cdot v_{i-1} + 0 \cdot v_{i+1} + \cdots + 0 \cdot v_d + 0 \cdot w_{d+1} + \cdots + 0 \cdot w_n$

因此，$T(v_i)$ 关于基 $\mathcal{B}$ 的坐标向量为：$[T(v_i)]_{\mathcal{B}} = (\underbrace{0, \ldots, 0, \lambda}_{i\text{-th 位置}}, 0, \ldots, 0)^\top$ 对于 $d+1 \leq j \leq n$，像 $T(w_j)$ 是 V 中的某个向量，它可以在基 $\mathcal{B}$ 下表示为：$T(w_j) = \sum_{k=1}^{d} a_{kj} v_k + \sum_{k=d+1}^{n} b_{kj} w_k$ 其中 $a_{kj}, b_{kj} \in \mathbb{F}$ 是一些标量。

因此，T 关于基 $\mathcal{B}$ 的矩阵 $[T]_{\mathcal{B}}$ 具有如下分块结构：$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda I_d & A \\ 0 & B \end{pmatrix}$ 其中：λI_d 是 $d \times d$ 数量矩阵，A 是包含系数 a_{kj} 的 $d \times (n-d)$ 矩阵，0 是 $(n-d) \times d$ 零矩阵，B 是包含系数 b_{kj} 的 $(n-d) \times (n-d)$ 矩阵。然后，计算特征多项式 T 的特征多项式由 $\mathfrak{p}_T(x) = \det(xI_n - [T]_{\mathcal{B}})$ 给出。

利用我们找到的分块结构：$xI_n - [T]_{\mathcal{B}} = \begin{pmatrix} xI_d - \lambda I_d & -A \\ 0 & xI_{n-d} - B \end{pmatrix} = \begin{pmatrix} (x-\lambda)I_d & -A \\ 0 & xI_{n-d} - B \end{pmatrix}$ 这是一个分块上三角矩阵。行列式的一个基本性质是，对于分块三角矩阵，其行列式是对角块行列式的乘积。

因此：$\mathfrak{p}_T(x) = \det\begin{pmatrix} (x-\lambda)I_d & -A \\ 0 & xI_{n-d} - B \end{pmatrix} = \det((x-\lambda)I_d) \cdot \det(xI_{n-d} - B)$ 由于 $\det((x-\lambda)I_d) = (x-\lambda)^d$，我们得到：$\mathfrak{p}_T(x) = (x-\lambda)^d \cdot \det(xI_{n-d} - B) = (x-\lambda)^d \cdot q(x)$ 其中 $q(x) = \det(xI_{n-d} - B)$ 是关于 x 的多项式。最后，结论我们已经证明特征多项式 $\mathfrak{p}_T(x)$ 能被 $(x-\lambda)^d$ 整除。根据定义，代数重数 $m_a(\lambda)$ 是使得 $(x-\lambda)^k$ 整除 $\mathfrak{p}_T(x)$ 的最高次幂 k。

因此，我们必须有：$d \leq m_a(\lambda)$ 这恰好是结论：$m_g(\lambda) \leq m_a(\lambda)$ 证明完毕。 $\square$

5.1.3 特征多项式的结构与性质

Property

设 $A \in M_n(\mathbb{F})$，则其特征多项式 $p_A(\lambda)$ 是一个 n 次多项式，且：

1. 首项系数为 1；
2. λ^{n-1} 的系数为 $-\text{tr}(A)$；
3. 常数项为 $(-1)^n \det(A)$；
4. 特征多项式的根即为矩阵的特征值。

Theorem

因此对于 n 阶方阵 $A = (a_{ij})$，其特征多项式为

$$p_A(\lambda) = \lambda^n - \text{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A),$$

其中 $\text{tr}(A) = \sum_{i=1}^n a_{ii}$ 是矩阵的迹，且特征值之和等于迹，特征值之积等于行列式。

5.2 对角化

通过前言的讨论，我们已初步理解对角化的概念与相关工具。在本节中，我们将首先完整演示两个矩阵对角化的具体过程，帮助大家深入掌握对角化的操作步骤；随后，在后半部分我们将引入 3 条方阵可对角化的判定准则，并运用这些准则对前半部分的实例进行验证，从而加深对“**对角化**”（diagonalization）这一概念的理解。

5.2.1 可三角化

Definition

可三角化（triangularizable） 对 $\mathbb{F}$-线性映射 $T : V \to V$，若存在 V 的一组基 $\mathcal{B}$，使得 T 在基 $\mathcal{B}$ 下的矩阵表示 $T_{\mathcal{B},\mathcal{B}}$ 是三角矩阵（上三角或下三角）(upper or lower)，则称 T 是**可三角化的**。

Claim

雅可比三角化 (Jacobi’ s triangularization)): 若 $\mathbb{F}$ 是代数闭域，且 $T : V \to V$ 是有限维 $\mathbb{F}$-线性映射（即 $\dim_{\mathbb{F}}(V) < \infty$），则 T 是可三角化的。

证明. 雅可比三角化定理的证明（数学归纳法，列表形式）

我们采用**数学归纳法**。

1. **基例** $(n=1)$：当 $\dim_{\mathbb{F}} V = 1$ 时，线性变换 T 对应的矩阵是 1×1 矩阵 $[a]$ $(a\in\mathbb{F})$，其本身就是三角矩阵（上三角或下三角），故 T 显然可三角化。

2. **归纳假设**：假设对所有维数 **小于**n 的有限维 $\mathbb{F}$-线性空间，结论成立：即任意 $n-1$ 维空间上的线性变换，在代数闭域 $\mathbb{F}$ 下均可三角化。

3. **归纳步骤**（$n\geq 2$）：考虑 n 维空间 V 上的线性变换 T，按以下步骤推导：

   (a) **存在特征值与特征向量**因 $\mathbb{F}$ 是代数闭域，T 的特征多项式 $\mathfrak{p}_T(\lambda)=\det(\lambda I-[T]_{\mathcal{B}})$（$\mathcal{B}$ 是 V 的某组基）在 $\mathbb{F}$ 中必有根，记为 λ_1，对应特征向量 $v_1\in V$（满足 $T(v_1)=\lambda_1 v_1$ 且 $v_1\neq 0$）。

   (b) **构造不变子空间与商空间**令 $W=\operatorname{span}\{v_1\}$，则 W 是 T 的不变子空间（因 $T(W)\subseteq W$）。考虑商空间 V/W（维数为 $n-1$），诱导出线性变换 $\bar{T}: V/W\to V/W$（定义为 $\bar{T}(v+W)=T(v)+W$）。

   (c) **应用归纳假设**由归纳假设，$n-1$ 维空间上的 $\bar{T}$ 可三角化，即存在 V/W 的一组基 $\{\bar{v}_2,\dots,\bar{v}_n\}$，使得 $\bar{T}$ 在这组基下的矩阵是三角矩阵。

   (d) **提升为原空间的基**将 $\bar{v}_2,\dots,\bar{v}_n$ 提升为 V 中的向量 $v_2,\dots,v_n$（即 $\bar{v}_i=v_i+W$），则 $\{v_1,v_2,\dots,v_n\}$ 是 V 的一组基。

   (e) **验证三角化**在基 $\{v_1,v_2,\dots,v_n\}$ 下，T 的矩阵表示为上三角矩阵：

   $$[T]_{\{v_1,v_2,\dots,v_n\}}=\begin{pmatrix}\lambda_1 & * & \cdots & * \\ 0 & a_{22} & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn}\end{pmatrix}$$

   其中第一列除 λ_1 外均为 0，右下角的 $(n-1)\times(n-1)$ 子矩阵是 $\bar{T}$ 的三角矩阵表示。

由数学归纳法，对所有有限维 $\mathbb{F}$-线性空间 V，当 $\mathbb{F}$ 是代数闭域时，线性变换 T 可三角化。

□

5.2.2 对角化的两个典型例子

Example

例 1：求矩阵 $A=\begin{pmatrix}1 & 2\\ 3 & 2\end{pmatrix}$ 的特征值 λ_i、特征子空间 V_i 以及过渡矩阵 Q，使得 $Q^{-1}AQ$ 为对角矩阵。

Example

例 2：已知 $V = P_1(\mathbb{R})$，线性变换 $T : V \to V$ 定义为

$$T(ax+b) = (-6a+2b)x + (-6a+b)$$

求线性变换 T 在基 $\{1, x\}$ 下的矩阵表示 A。并且判断矩阵 A 是否可对角化。若可对角化，求出可逆矩阵 P，使得 $P^{-1}AP$ 为对角矩阵。

我们已经展示了两个可以对角化的例子，但是我们不禁要思考：**怎么样的方阵才可以对角化呢**?

5.2.3 可对角化的充要条件

从几何上看，若 n 维线性空间 V 上的线性变换 φ 在某组基 $\{e_1, e_2, \ldots, e_n\}$ 下的表示矩阵为对角阵：

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

则称 φ 为**可对角化 (diagonalizable)** 线性变换。此时 $\varphi(e_i) = \lambda_i e_i$，即 $e_1, e_2, \ldots, e_n$ 是 φ 的特征向量，于是 φ 有 n 个线性无关的特征向量。

反过来，若 n 维线性空间 V 上的线性变换 φ 有 n 个线性无关的特征向量 $e_1, e_2, \ldots, e_n$，则这组向量构成了 V 的一组基，且 φ 在这组基下的表示矩阵显然是一个对角阵。($\because$

$n = n$ ∴ 为基）

这样我们就证明了下述定理。

Theorem

可对角化的充要条件 1：n 阶方阵 A 可对角化的充要条件是 A 有 n 个线性无关的特征向量。

但是，我们注意到：

【Remark】：

给定一个有限维的 $\mathbb{F}$-线性映射 $T : V \to V$（即 $\dim_{\mathbb{F}}(V) < \infty$）。若其特征多项式 $\mathfrak{p}_T(x) = \prod_{i=1}^{s}(x - \lambda_i)$，其中 $\lambda_i \neq \lambda_j \in \mathbb{F}$，则 T 在 $\mathbb{F}$ 上可对角化。
但是其逆命题不一定成立。

证明. T 可对角化：由特征多项式有 s 个互异根 $\lambda_1, \ldots, \lambda_s \in \mathbb{F}$，知每个 λ_i 是特征值。互异特征值的特征向量线性无关，且 $\dim V = s$（特征多项式次数等于空间维数），故这些特征向量构成 V 的基，因此 T 可对角化。

反例（逆命题不成立）：取 $V = \mathbb{F}^2$，令 T 为恒等映射 I（即 $I(v) = v, \forall v \in V$）。T 可对角化（矩阵为单位矩阵 $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$），但特征多项式为 $(x-1)^2$，含重根，并非互异一次因式的乘积。故逆命题不成立。 □

现在我们来讨论不同特征值和它们的特征值子空间与全空间 V 的关系，我们会有以下定理：

Theorem

可对角化的充要条件 2：A 可对角化当且仅当 A 的特征子空间直和等于全空间，即

$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_s}$$

其中 $\lambda_1, \lambda_2, \ldots, \lambda_s$ 是 A 的所有不同特征值。

回顾：我们在上一节中，特征值 λ 在特征多项式中的根重数称为代数重数，对应特征子空间 V_λ 的维数称为几何重数。而对照几何重数与代数重数，受到启发，我们可以得到第三条定理：

Theorem

可对角化的充要条件 3：

$$A \text{ 可对角化} \iff \text{对每个特征值}\lambda_i \qquad \text{代数重数} = \text{几何重数}$$

即代数重数等于几何重数

当几何重数小于代数重数时，矩阵无法通过相似变换化为对角阵，此时需引入 Jordan 标准形理论（下一章节讨论）。

同时，本节隐去了利用极小多项式来判断可对角化的定理，我们将在下节介绍：

5.3 极小多项式与 Cayley-Hamilton 定理

5.3.1 极小多项式的定义

Definition

极小多项式（minimal polynomial）： 设 A 是数域 K 上的 n 阶方阵。在多项式环 $K[x]$ 中，存在唯一的首一多项式 $m(x)$，满足 $m(A)=0$ 且次数最低，称 $m(x)$ 为 A 的极小多项式。

即，对于任意的多项式 $f(x)\in K[x]$ 使得 $f(A)=0$，则 $\deg(m(x))\leq\deg(f(x))$，且 $m(x)$ 是**首一**的。

例如：

Example

设 $\boldsymbol{A}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$，则 $\boldsymbol{A}$ 定义了一个线性算子：

$$A: \quad \mathbb{F}^2\to\mathbb{F}^2$$

满足 $\boldsymbol{x}\mapsto\boldsymbol{A}\boldsymbol{x}$。

这里，特征多项式 $\chi_A(x)=(x-1)^2$，且 $\boldsymbol{A}-\boldsymbol{I}=\boldsymbol{0}$，故极小多项式 $m_A(x)=x-1$。

那么极小多项式一定存在吗？

Claim

对于有限维向量空间 V（dim=n），零化多项式存在。

证明. 显然，集合 $\{I, T, \ldots, T^n, T^{n+1}, \cdots, T^{n^2}\}$ 包含于线性变换空间 $\mathrm{Hom}(V, V)$。

由于 $\mathrm{Hom}(V, V)$ 的维数是 n^2，因此可推出集合 $\{I, T, \ldots, T^n, T^{n+1}, \cdots, T^{n^2}\}$ 是线性相关的，即存在不全为零的系数 a_i，使得 $a_0 I + a_1 T + \cdots + a_{n^2} T^{n^2} = 0$ 也就是说，存在次数小于 n^2 的多项式 $g(x)$，使得 $g(T) = 0$（即 T 满足一个次数小于 n^2 的多项式方程）。 □

但是在无限维上，极小多项式不一定存在：

考虑 $V = \mathbb{R}[x]$ 以及映射

$$\begin{aligned} T: \quad & V \to V \\ p(x) \mapsto & \int_0^x p(t)\,\mathrm{d}t \end{aligned}$$

特别地，$T(x^n) = \frac{1}{n+1} x^{n+1}$。假设 $m_T(x)$ 是次数为 n 的多项式，即

$$m_T(x) = x^n + \cdots + a_1 x + a_0,$$

那么

$$m_T(T) = T^n + \cdots + a_0 I$$

是零线性变换。

由此可得

$$[m_T(T)](x) = \frac{1}{n!} x^n + a_{n-1} \frac{1}{(n-1)!} x^{n-1} + \cdots + a_1 x + a_0 = 0_{\mathbb{F}},$$

这导致矛盾，因为左边对 $k = 1, \ldots, n$ 的 x^k 系数非零，而右边为零。

5.3.2 极小多项式的性质

Property

(1) **唯一性：** 对于给定的方阵 A，其极小多项式是唯一的首一多项式。

(2) **整除性：** 若 $f(x) \in K[x]$ 满足 $f(A) = 0$（称 $f(x)$ 为 A 的零化多项式），则 $m(x) \mid f(x)$。

(3) **相似不变性**：若 A 与 B 相似，即存在可逆矩阵 P 使得 $B = P^{-1}AP$，则 A 与 B 有相同的极小多项式。

证明. (1) 假设 f_1, f_2 是两个次数相同的不同极小多项式，即 $\deg(f_1) = \deg(f_2)$。由此可得：

- $\deg(f_1 - f_2) < \deg(f_1)$。
- $f_1 - f_2 \neq 0$
- $(f_1 - f_2)(T) = f_1(T) - f_2(T) = 0_{V\to V}$

对 $f_1 - f_2$ 进行缩放后，存在一个次数更低的首一多项式 g 满足 $g(T) = 0$，这与极小多项式的定义矛盾。

(3)：显然，$\deg(f) \geq \deg(m_T)$。由带余除法，我们有 $f(x) = q(x)m_T(x) + r(x)$.

因此，对任意 $v \in V$，$[r(T)](v) = [f(T)](v) - [q(T)m_T(T)](v) = 0_V - q(T)0_V = 0_V - 0_V = 0_V$ 因此，$r(T) = 0_{V\to V}$。由极小多项式的定义，可得 $r(x) \equiv 0$。

(4)：假设 $B = P^{-1}AP$，且 $m_A(x) = x^k + \cdots + a_1x + a_0$，$m_B(x) = x^\ell + \cdots + b_0$. 于是有

因此，因 $m_A(A) = 0$，故 $m_A(B) = 0$。

可得 $m_B(x) \mid m_A(x)$。类似地，$m_A(x) \mid m_B(x)$。由于 $m_A(x)$ 和 $m_B(x)$ 都是首一的，故可得 $m_A(x) = m_B(x)$

□

接下来，我们考察极小多项式与“维度”的关系

接下来，我们考察一下极小多项式与特征值的关系：

Claim

极小多项式与特征多项式有相同的根（重数可能不同）。

设 λ 是方阵 A 的特征值，对应的特征向量为 ξ，即 $A\xi = \lambda\xi$，$\xi \neq 0$。

- 若 $m(x)$ 是 A 的极小多项式，则 $m(\lambda) = 0$。即极小多项式的根必为矩阵 A 的特征值。

- 反之，若 λ 是 A 的特征值，则 $m(\lambda) = 0$。即矩阵 A 的特征值必为极小多项式的根。

证明. 我们需要证明：极小多项式的根等价于特征多项式的根（即两者根完全相同，仅重数可能不同）。

(1) “极小多项式的根必是特征值” 设 λ 是极小多项式 $m(x)$ 的根，即 $m(\lambda) = 0$。将 $m(x)$ 分解为 $m(x) = (x - \lambda)q(x)$（其中 $q(x)$ 是多项式），则：

$$m(A) = (A - \lambda I)q(A) = 0$$

若 $q(A)$ 不是零映射，取非零向量 ξ 使得 $q(A)\xi \neq 0$，则：

$$(A - \lambda I)(q(A)\xi) = 0 \implies A(q(A)\xi) = \lambda(q(A)\xi)$$

这说明 $q(A)\xi$ 是 A 属于 λ 的 ** 特征向量 **，故 λ 是 A 的特征值。

（2）“特征值必是极小多项式的根” 设 λ 是 A 的特征值，$\xi \neq 0$ 是对应的特征向量（即 $A\xi = \lambda\xi$）。由于 $m(A) = 0$，对特征向量 ξ 有：

$$m(A)\xi = 0$$

将 $m(x) = a_0 + a_1 x + \cdots + a_k x^k$ 代入，利用 $A^n\xi = \lambda^n\xi$（归纳可得：$A^2\xi = A(A\xi) = \lambda A\xi = \lambda^2\xi$，以此类推），则：

$$m(A)\xi = a_0\xi + a_1 A\xi + \cdots + a_k A^k\xi = (a_0 + a_1\lambda + \cdots + a_k\lambda^k)\xi = m(\lambda)\xi$$

因 $\xi \neq 0$ 且 $m(A)\xi = 0$，故 $m(\lambda) = 0$，即 λ 是 $m(x)$ 的根。

□

因此，我们有如下推论：

Corollary & Secondary Conclusion

方阵 A 的极小多项式 $m(x)$ 整除其特征多项式 $f_A(x)$。

同样的，我们可以拓展到线性变换的角度：

Corollary & Secondary Conclusion

线性算子 T 的极小多项式 $m_T(x)$ 整除其特征多项式 $f_T(x)$。

【Remark】：

两者的根完全相同，但重数可能不同。
例如：
- 若 $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$（Jordan 块），特征多项式 $\mathfrak{p}_A(x) = (x-\lambda)^2$，极小多项式 $m(x) = (x-\lambda)^2$（重数相同）。
- 若 $A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$（对角矩阵），特征多项式 $\mathfrak{p}_A(x) = (x-\lambda)^2$，极小多项式 $m(x) = x-\lambda$（重数不同）。

Claim

若有限维线性空间 V 上的线性变换 T 的极小多项式 $\mathfrak{m}_T(x)$ 在数域 $\mathbb{F}$ 上可以分解为互异的一次因式的乘积，即 $\mathfrak{m}_T(x) = \prod_{i=1}^{s}(x-\lambda_i)$, $\lambda_i \in \mathbb{F}, \lambda_i \neq \lambda_j$ for $i \neq j$,
则 V 可以分解为 T 的特征子空间的直和：$V = \ker(T-\lambda_1 I) \oplus \ker(T-\lambda_2 I) \oplus \cdots \oplus \ker(T-\lambda_s I)$. 其中，$\ker(T-\lambda_i I)$ 是 T 对应于特征值 λ_i 的特征子空间，由所有满足 $T(v) = \lambda_i v$ 的向量 $v \in V$ 构成。

这可以理解为**下节准素分解**第二条的特殊情况

证明. 证明第一点：**和的封闭性我们利用多项式的性质来构造这样的分解。**

考虑多项式 $p_i(x) = \prod_{\substack{j=1 \\ j\neq i}}^{s}(x-\lambda_j)$。即 $p_i(x)$ 是去掉了 $(x-\lambda_i)$ 后的极小多项式。由于 $\lambda_1, \lambda_2, \ldots, \lambda_s$ 互不相同，多项式 $p_1(x), p_2(x), \ldots, p_s(x)$ 是互素的（它们的最大公因子是 1）。

根据多项式互素的性质，存在多项式 $a_1(x), a_2(x), \ldots, a_s(x) \in \mathbb{F}[x]$，使得 $a_1(x)p_1(x) + a_2(x)p_2(x) + \cdots + a_s(x)p_s(x) = 1$.

现在，我们将这个多项式等式“作用”于线性变换 T，得到一个变换的等式：$a_1(T)p_1(T) + a_2(T)p_2(T) + \cdots + a_s(T)p_s(T) = I$, 其中 I 是恒等变换。（常数项 1 应被替换为 $1 \cdot I$，也就是 I，而非 T。我们是两边作用 T，而不是两边乘 T。）

对于任意一个向量 $v \in V$，我们将上面的等式两边同时作用于 v：$v = I(v) = \left(\sum_{i=1}^{s} a_i(T)p_i(T)\right)(v) = \sum_{i=1}^{s}\left(a_i(T)p_i(T)(v)\right)$. 令 $v_i = a_i(T)p_i(T)(v)$。我们只需证明 $v_i \in \ker(T-\lambda_i I)$。

注意到 $\mathfrak{m}_T(T) = 0$（极小多项式的定义），而 $\mathfrak{m}_T(x) = (x-\lambda_i)p_i(x)$。

因此：$(T-\lambda_i I)p_i(T) = \mathfrak{m}_T(T) = 0$. 所以，$(T-\lambda_i I)(v_i) = (T-\lambda_i I)\left(a_i(T)p_i(T)(v)\right) = a_i(T)\left((T-\lambda_i I)p_i(T)(v)\right) = a_i(T)(0) = 0$. 这就证明了 $v_i \in \ker(T-\lambda_i I)$。

因此，V 中任意向量 v 都可以表示为各特征子空间中向量的和。

证明第二点：**分解的唯一性**

假设 $0 = v_1 + v_2 + \cdots + v_s$，其中 $v_i \in \ker(T-\lambda_i I)$。我们来证明 $v_k = 0$ 对任意 k 成立。考虑多项式 $p_k(T) = \prod_{\substack{j=1 \\ j\neq k}}^{s}(T-\lambda_j I)$。

对等式 $0 = \sum_{i=1}^{s} v_i$ 两边同时作用 $p_k(T)$：$0 = p_k(T)(0) = p_k(T)\left(\sum_{i=1}^{s} v_i\right) = \sum_{i=1}^{s} p_k(T)(v_i)$. 现在分析每一项 $p_k(T)(v_i)$：当 $i \neq k$ 时，$p_k(x)$ 中含有因子 $(x-\lambda_i)$，所以 $p_k(T)$ 中含有因子 $(T-\lambda_i I)$。

因为 $v_i \in \ker(T-\lambda_i I)$，即 $(T-\lambda_i I)(v_i) = 0$，所以 $p_k(T)(v_i) = 0$。当 $i = k$ 时，$p_k(x) = \prod_{\substack{j=1 \\ j\neq k}}^{s}(x-\lambda_j)$。将 $x = \lambda_k$ 代入，得到 $p_k(\lambda_k) = \prod_{\substack{j=1 \\ j\neq k}}^{s}(\lambda_k-\lambda_j)$。由于所有 λ_i 互不相同，这个乘积不为零，是一个非零常数。

因此，$p_k(T)(v_k) = p_k(\lambda_k)\cdot v_k$（因为 $T(v_k) = \lambda_k v_k$，所以 $T^n(v_k) = \lambda_k^n v_k$）。因此，上面的求和式简化为：$0 = p_k(\lambda_k)\cdot v_k$. 因为 $p_k(\lambda_k) \neq 0$，我们可以在等式两边乘以其逆元，得到 $v_k = 0$。由 k 的任意性可知，所有 v_i 都必须是零向量。这就证明了分解的唯一性。结论综合以上两点，我们证明了 $V = \ker(T-\lambda_1 I)\oplus\ker(T-\lambda_2 I)\oplus\cdots\oplus\ker(T-\lambda_s I)$. □

一旦得到这个直和分解，我们就可以在每个特征子空间 $\ker(T-\lambda_i I)$ 中选取一组基。将这些基向量合并起来，就得到了线性空间 V 的一组基。这组基中的每一个向量都是 T 的特征向量。因此，T 在这组基下的矩阵就是一个对角矩阵，其对角线上的元素就是对应的特征值 λ_i。这就完成了“若 T 的极小多项式是互异一次因式的乘积，则 T 可对角化”的证明。

最后，给出可对角化的另一条判断准则：

Theorem

给定一个有限维的 $\mathbb{F}$-线性映射 $T: V \to V$（即 $\dim_{\mathbb{F}}(V) < \infty$）。证明 T 在 $\mathbb{F}$ 上可对角化当且仅当它的极小多项式 $\mathfrak{m}_T(x) = \prod_{i=1}^{s}(x-\lambda_i)$，其中 $\lambda_i \neq \lambda_j \in \mathbb{F}$。

证明. 我们分必要性和充分性两部分证明。

必要性（可对角化 ⇒ 极小多项式是互异一次因式的乘积）

若 T 可对角化，则存在 V 的一组基 $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$，使得 T 在这组基下的矩阵为对角矩阵：$A = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$ 其中 $\lambda_1, \lambda_2, \ldots, \lambda_n$ 是 T 的特征值（可能有重复）。

步骤 1：构造候选多项式设 $\lambda_1, \lambda_2, \ldots, \lambda_s$ 是 T 的所有互异特征值（即 $s \leq n$，重复的特征值只保留一个），构造多项式：$m(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_s)$

步骤 2：验证 $m(T) = 0$ 对对角矩阵 A，多项式作用于矩阵的结果是对角元素分别代入多项式：$m(A) = \begin{pmatrix} m(\lambda_1) & & & \\ & m(\lambda_2) & & \\ & & \ddots & \\ & & & m(\lambda_n) \end{pmatrix}$ 由于每个 $\lambda_i\,(i = 1, \ldots, n)$ 都是互异特征值中的一个（即存在 k 使得 $\lambda_i = \lambda_k$），因此：$m(\lambda_i) = (\lambda_i - \lambda_1)\cdots(\lambda_i - \lambda_k)\cdots(\lambda_i - \lambda_s) = 0$ 故 $m(A) = 0$，即 $m(T) = 0$，说明 $m(x)$ 是 T 的零化多项式。

步骤 3：证明 $m(x)$ 是极小多项式假设存在次数更低的首一零化多项式 $p(x)$，且 $\deg(p) < s$。由于 $p(T) = 0$，对每个互异特征值 λ_k，取对应的特征向量 α_k $(T\alpha_k = \lambda_k\alpha_k)$，则：$p(T)\alpha_k = p(\lambda_k)\alpha_k = 0$ 因 $\alpha_k \neq 0$，故 $p(\lambda_k) = 0$，即 $p(x)$ 有 s 个不同的根 $\lambda_1, \ldots, \lambda_s$。但 $\deg(p) < s$，与“次数为 d 的多项式最多有 d 个根（重根计入）”矛盾。因此 $m(x)$ 是次数最低的零化多项式，即极小多项式。

此时，T 的极小多项式是其所有互异特征值对应的一次因式的乘积，即

$$\mathfrak{m}_T(x) = \prod_{i=1}^{s}(x-\lambda_i),$$

且显然 $\lambda_i \neq \lambda_j$ 对任意 $i \neq j$ 成立。

充分性（极小多项式是互异一次因式的乘积 $\Rightarrow$ 可对角化）

设 $\mathfrak{m}_T(x) = \prod_{i=1}^{s}(x-\lambda_i)$，其中 $\lambda_i \neq \lambda_j \in \mathbb{F}$。根据线性空间的直和分解定理，若线性变换的极小多项式可分解为互异一次因式的乘积，则线性空间 V 可分解为其特征子空间的直和：

$$V = \ker(T-\lambda_1 I) \oplus \ker(T-\lambda_2 I) \oplus \cdots \oplus \ker(T-\lambda_s I).$$

对每个 i，$\ker(T-\lambda_i I)$ 是 T 对应于 λ_i 的特征子空间（由属于 λ_i 的特征向量张成）。

从每个特征子空间中选取一组基，合并后得到 V 的一组基（直和分解保证线性无关性），且这组基中每个向量都是 T 的特征向量。因此，T 在**这组基下**的矩阵是对角矩阵，即 T 在 $\mathbb{F}$ 上可对角化。

综上，T 在 $\mathbb{F}$ 上可对角化当且仅当它的极小多项式 $\mathfrak{m}_T(x)$ 是 $\mathbb{F}$ 上互异一次因式的乘积。 □

5.3.3 Cayley - Hamilton 定理

首先，我们有**矩阵**版本的 Cayley - Hamilton 定理：

Theorem

设 A 是数域 K 上的 n 阶方阵，其特征多项式为 $f_A(x) = \det(xI-A) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$，则 $f_A(A) = A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I = 0$。

即矩阵 A 满足它自身的特征方程。

其证明不仅仅可以使用刚刚提到的**极小多项式的概念**，而且可以使用“摄动法”这一思想。

于是，我们有如下推论：

Corollary & Secondary Conclusion

设 A 是 n 阶方阵，$f(x)$ 是 A 的特征多项式，对于任意多项式 $g(x)$，$g(A)$ 可表示为 $I, A, A^2, \cdots, A^{n-1}$ 的线性组合。

并且，我们可以得到**线性变换**的 Cayley - Hamilton 定理：

Theorem

设 φ 是数域 K 上 n 维线性空间 V 的线性变换，$f_\varphi(x)=\det(x\mathrm{id}_V-\varphi)$ 是 φ 的特征多项式（其中 id_V 是 V 上的恒等变换），则 $f_\varphi(\varphi)=0$。

即线性变换 φ 满足它自身的特征方程。

5.3.4 准素分解

Theorem

准素分解定理（Primary decomposition theorem）： 设 $T:V\to V$ 是有限维线性空间 V 上的线性算子，且其最小多项式为

$$m_T(x)=[p_1(x)]^{e_1}\cdots[p_k(x)]^{e_k}$$

其中 p_i 是两两不同的首一不可约多项式。令 $V_i=\ker(p_i(T)^{e_i})$，则

1. 每个 V_i 是 T-不变子空间（即 $T(V_i)\subseteq V_i$）
2. $V=V_1\oplus\cdots\oplus V_k$
3. $T|_{V_i}$ 的最小多项式为 $p_i(x)^{e_i}$。

证明. 结论 1：每个 $V_i=\ker(p_i(T)^{e_i})$ 是 T- 不变子空间对于任意 $v\in V_i$，由 V_i 的定义有 $p_i(T)^{e_i}(v)=0$。

计算 $p_i(T)^{e_i}(T(v))$：由于线性算子的多项式满足交换性（T 与 $p_i(T)$ 可交换），因此：$p_i(T)^{e_i}(T(v))=T(p_i(T)^{e_i}(v))=T(0)=0$ 这说明 $T(v)\in\ker(p_i(T)^{e_i})=V_i$，故 V_i 是 T- 不变子空间。

结论 2：

因 $p_1(x)^{e_1},\dots,p_k(x)^{e_k}$ 两两互质，由 Bezout 等式，存在 $f_i(x)$ 使 $\sum_{i=1}^k f_i(x)p_i(x)^{e_i}=1$。代入线性算子 T 得 $\sum_{i=1}^k f_i(T)p_i(T)^{e_i}=I$，故任意 $v\in V$ 可表为 $v=\sum_{i=1}^k f_i(T)p_i(T)^{e_i}(v)=\sum_{i=1}^k v_i$，其中 $v_i\in V_i$（因 $p_i(T)^{e_i}(v_i)=0$），即 $V=\sum V_i$。若 $\sum v_i=0$（$v_i\in V_i$），用 $f_i(T)p_i(T)^{e_i}$ 作用得 $v_i=0$，故为直和，即 $V=V_1\oplus\cdots\oplus V_k$。

结论 3：$T|_{V_i}$ 的最小多项式为 $p_i(x)^{e_i}$

记 $T_i=T|_{V_i}$，其最小多项式 $m_i(x)$。由 $V_i=\ker(p_i(T)^{e_i})$，得 $p_i(T_i)^{e_i}=0$，故 $m_i(x)\mid p_i(x)^{e_i}$。因 $p_i(x)$ 不可约，$m_i(x)=p_i(x)^{d_i}$（$d_i\le e_i$）。若 $d_i<e_i$，则 $p_i(T_i)^{d_i}=0$，结合 $V=\bigoplus V_j$，$p_i(T)^{d_i}$ 在 V 上恒零，与 T 的最小多项式 $m_T(x)=\prod p_j(x)^{e_j}$ 矛盾，故 $d_i=e_i$，即 $m_i(x)=p_i(x)^{e_i}$。

□

如何理解这个定理？

对每个 $p_i(x)^{e_i}$，定义子空间 $V_i = \ker(p_i(T)^{e_i})$（即被 $p_i(T)^{e_i}$ 映射到零的所有向量构成的子空间）。准素分解定理说明：**可以通过这些 V_i 把 V 拆解成结构更简单的"块"**。

而且我们有：

Property

1. 大空间 V 可以拆成 k 个"互不相干"的小空间的组合，每个小空间的结构由 $p_i(x)^{e_i}$ 主导。可以把 T 限制在每个 V_i 上单独研究，因为操作不会跳出该子空间。
2. 且可以把 T 限制在每个 V_i 上单独研究，因为操作不会跳出该子空间。
3. 除了多项式 $p_i(x)^{e_i}$，不存在次数更低的首一多项式能零化 $T|_{V_i}$（否则会与 $m_T(x)$ 是原空间极小多项式矛盾）。

该定理将线性空间按极小多项式的不可约因子幂分解为直和，每个直和项对算子 T 不变，且限制在每个项上的极小多项式就是对应的不可约因子幂。这是后续若尔当标准形、有理标准形等结构定理的核心基础。通过极小多项式的因式分解，直接对应空间的直和分解，深刻体现了线性代数中"代数性质（多项式）决定几何结构（空间分解）"的核心思想。

5.4 （应用）Lotka-Volterra 模型

5.4.1 简化的捕食者-猎物模型

假设兔子（猎物）数量为 x，狐狸（捕食者）数量为 y，建立离散模型：

$$x_{n+1} = 2x_n - y_n$$
$$y_{n+1} = 0.5x_n + y_n$$

写成矩阵形式：

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix}$$

令 $A = \begin{bmatrix} 2 & -1 \\ 0.5 & 1 \end{bmatrix}$，我们通过对角化求 A^k 来预测未来 k 步的种群变化。

【求特征值】：解特征方程 $\det(A - \lambda I) = 0$：

$$\begin{vmatrix} 2-\lambda & -1 \\ 0.5 & 1-\lambda \end{vmatrix} = (2-\lambda)(1-\lambda) + 0.5 = \lambda^2 - 3\lambda + 2.5 = 0$$

解得特征值：

$$\lambda_1 = \frac{3+\sqrt{9-10}}{2} = \frac{3+1}{2} = 2, \quad \lambda_2 = \frac{3-1}{2} = 1$$

【求特征向量】：- 对于 $\lambda_1 = 2$，解 $(A-2I)\mathbf{v} = 0$：

$$\begin{bmatrix} 0 & -1 \\ 0.5 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix} \implies \mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

- 对于 $\lambda_2 = 1$，解 $(A-1I)\mathbf{v} = 0$：

$$\begin{bmatrix} 1 & -1 \\ 0.5 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \implies \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

【构造矩阵 P 和对角矩阵 D】：令 $P = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$，则 $P^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$（通过伴随矩阵法求得）。对角矩阵 $D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$，满足 $A = PDP^{-1}$。

5.4.2 求矩阵 A 的 k 次方

根据对角化性质：

$$A^k = PD^kP^{-1}$$

其中 $D^k = \begin{bmatrix} 2^k & 0 \\ 0 & 1^k \end{bmatrix} = \begin{bmatrix} 2^k & 0 \\ 0 & 1 \end{bmatrix}$。

代入计算 A^k：

$$\begin{aligned}
A^k &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2^k & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\
&= \begin{bmatrix} 2 \cdot 2^k + 1 \cdot (-1) & 2 \cdot 0 + 1 \cdot 2 \\ 1 \cdot 2^k + 1 \cdot (-1) & 1 \cdot 0 + 1 \cdot 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\
&= \begin{bmatrix} 2^{k+1} - 1 & 2 \\ 2^k - 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\
&= \begin{bmatrix} (2^{k+1}-1) \cdot 1 + 2 \cdot (-1) & (2^{k+1}-1) \cdot (-1) + 2 \cdot 2 \\ (2^k-1) \cdot 1 + 2 \cdot (-1) & (2^k-1) \cdot (-1) + 2 \cdot 2 \end{bmatrix} \\
&= \begin{bmatrix} 2^{k+1} - 1 - 2 & -2^{k+1} + 1 + 4 \\ 2^k - 1 - 2 & -2^k + 1 + 4 \end{bmatrix} \\
&= \begin{bmatrix} 2^{k+1} - 3 & -2^{k+1} + 5 \\ 2^k - 3 & -2^k + 5 \end{bmatrix}
\end{aligned}$$

5.4.3 预测种群变化

假设初始状态 $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 100 \\ 50 \end{bmatrix}$，则 k 步后的种群数量为：

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = A^k \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 2^{k+1} - 3 & -2^{k+1} + 5 \\ 2^k - 3 & -2^k + 5 \end{bmatrix} \begin{bmatrix} 100 \\ 50 \end{bmatrix}$$

例如，当 $k = 3$ 时：

$$A^3 = \begin{bmatrix} 2^4 - 3 & -2^4 + 5 \\ 2^3 - 3 & -2^3 + 5 \end{bmatrix} = \begin{bmatrix} 13 & -11 \\ 5 & -3 \end{bmatrix}$$

$$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 13 & -11 \\ 5 & -3 \end{bmatrix} \begin{bmatrix} 100 \\ 50 \end{bmatrix} = \begin{bmatrix} 13 \cdot 100 + (-11) \cdot 50 \\ 5 \cdot 100 + (-3) \cdot 50 \end{bmatrix} = \begin{bmatrix} 750 \\ 350 \end{bmatrix}$$

Chapter 6

相似标准型

一般矩阵无法对角化的核心原因在于**特征向量的数量不足**。即使矩阵的特征多项式在数域上可分解，若存在某个特征值的几何重数（特征子空间的维数）小于其代数重数（在特征多项式中的重数），则无法找到足够多线性无关的特征向量来构成整个向量空间的基，从而不能对角化。

取矩阵 $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$。其特征方程 $\det(A - \lambda I) = (2-\lambda)^2 = 0$，特征值 $\lambda = 2$，代数重数为 2 。解 $(A - 2I)X = 0$ 得特征向量 $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$，几何重数为 1 。

因特征值 $\lambda = 2$ 的几何重数 1 小于代数重数 2，找不到足够线性无关特征向量构成 $\mathbb{R}^2$ 的基，所以 A 不可对角化。

那么对于这种不可对角化的问题，我们应该做什么“补偿”或退让来达到一种“近似的对角化”?

6.1 Jordan 标准型的讨论 1

6.1.1 Jordan 标准型与 Jordan 块的定义

设 T 是 $\mathbb{C}^8$ 上的线性算子，$\beta = \{v_1, v_2, \ldots, v_8\}$ 是 $\mathbb{C}^8$ 的有序基，且

$$J = [T]_\beta = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

是 T 的 Jordan 标准型。此时，T 的特征多项式为 $\det(J-tI)=(t-2)^4(t-3)^2t^2$，每个特征值的重数等于其在 J 对角线上出现的次数。同时，β 中仅 v_1, v_4, v_5, v_7 是 T 的特征向量，它们对应 J 中对角线上方无 1 的列。

我们将在接下来的讨论中证明，**每个特征多项式可分解的线性算子都有一个 Jordan 标准型**（在 Jordan 块顺序上唯一）。

再次考察矩阵 J 和有序基 β，由 $T(v_2)=v_1+2v_2$ 可得 $(T-2I)(v_2)=v_1$，类似地 $(T-2I)(v_3)=v_2$。因为 v_1 和 v_4 是 T 对应 $\lambda=2$ 的特征向量，所以 $(T-2I)^3(v_i)=0$ $(i=1,2,3,4)$；同理，$(T-3I)^2(v_i)=0$ $(i=5,6)$，$(T-0I)^2(v_i)=0$ $(i=7,8)$。

从 Jordan 标准型中每个 Jordan 块的结构可推广得出：若 v 位于线性算子 T 的 Jordan 标准基中，且与对角元素为 λ 的 Jordan 块相关联，则对于足够大的 p，$(T-\lambda I)^p(v)=0$。

而所谓的特征向量，就是满足 $p=1$ 时的特殊情景。

在上述例子中，矩阵 $J=[T]_\beta$ 呈现出特定的分块结构。例如，子矩阵 $\begin{pmatrix}2&1&0\\0&2&1\\0&0&2\end{pmatrix}$ 和 $\begin{pmatrix}3&1\\0&3\end{pmatrix}$ 等，具有如下显著特征：

Fact or Background

主对角线元素相同，主对角线上方的次对角线元素为 1，其余位置为 0。

这种形式的矩阵块便是 Jordan 块，它是构成 Jordan 标准型的基本单元，在描述线性算子结构中至关重要。

于是我们给出严谨定义：

Definition

一般地，对于特征值 λ，n 阶 *Jordan* 块（Jordan block）定义为：

$$J(\lambda,n)=\begin{pmatrix}\lambda&1&0&\cdots&0\\0&\lambda&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&\lambda\end{pmatrix}$$

通过研究 Jordan 块，我们能更深入理解线性算子的性质，如特征值、特征向量及算子在特定基下的表示等。

我们已经知道：形如 $J(\lambda, n) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}$ 的矩阵称为 Jordan 块，其中 λ 为特征值，n 为阶数。

以例子中的线性算子 T 为例，**其 Jordan 标准型 J 由多个 Jordan 块组成**，这些块的特征值与阶数反映了 T 的关键性质，如特征值的重数、广义特征向量的结构等。从这个例子出发，我们将系统探讨 Jordan 块的性质、由其构成的 Jordan 标准型，以及它们在线性算子分析中的重要应用。

Definition

Jordan 标准型（Jordan canonical form）：由若干个 Jordan 块组成的分块对角矩阵称为 Jordan 标准型，它是矩阵在相似变换下的一种标准形式。

更本质的，

设 $T : V \to V$ 的最小多项式为

$$m_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{e_i},$$

则存在基 $\mathcal{A}$，使得

$$[T]_{\mathcal{A},\mathcal{A}} = \operatorname{diag}(J_1, \ldots, J_\ell),$$

其中每个块 J_i 是形如

$$J_i = \begin{bmatrix} \mu_i & 1 & & \\ & \mu_i & \ddots & \\ & & \ddots & 1 \\ & & & \mu_i \end{bmatrix}$$

的方阵。

Fact or Background

设 V 是有限维向量空间，$T \in \mathcal{L}(V)$，若 T 的最小多项式 $m_T(x) = x^m$，则 T 可对角化为准 Jordan 块的直和（即 Jordan 标准形中所有特征值为 0 的 Jordan 块的直和）。

证明. 首先证明核空间的升链 $\{0\} = \ker(T^0) \subseteq \ker(T) \subseteq \ker(T^2) \subseteq \cdots \subseteq \ker(T^m) = V$ 是 ** 严格递增 ** 的，即对所有 $i = 1, 2, \ldots, m$，有 $\ker(T^{i-1}) \subsetneq \ker(T^i)$。

- 当 $i = m$ 时：因 $m_T(x) = x^m$ 是最小多项式，故 $T^m = 0$（即 $\ker(T^m) = V$），且 $T^{m-1} \neq 0$（否则最小多项式次数小于 m），因此 $\ker(T^{m-1}) \subsetneq V = \ker(T^m)$。

- 当 $i = m-1$ 时：假设 $\ker(T^{m-2}) = \ker(T^{m-1})$，则对任意 $\mathbf{x} \in \ker(T^m) = V$，有 $T^{m-1}(T\mathbf{x}) = T^m\mathbf{x} = 0 \implies T\mathbf{x} \in \ker(T^{m-1}) = \ker(T^{m-2})$，进而 $T^{m-2}(T\mathbf{x}) = T^{m-1}\mathbf{x} = 0 \implies \mathbf{x} \in \ker(T^{m-1})$。这导致 $\ker(T^m) \subseteq \ker(T^{m-1})$，与 $\ker(T^{m-1}) \subsetneq \ker(T^m)$ 矛盾，故 $\ker(T^{m-2}) \subsetneq \ker(T^{m-1})$。

- 一般地，对 $i = m, m-1, \ldots, 1$，用数学归纳法可证：假设 $\ker(T^i) \subsetneq \ker(T^{i+1})$，若 $\ker(T^{i-1}) = \ker(T^i)$，则对任意 $\mathbf{x} \in \ker(T^{i+1})$，有 $T^i(T\mathbf{x}) = T^{i+1}\mathbf{x} = 0 \implies T\mathbf{x} \in \ker(T^i) = \ker(T^{i-1}) \implies T^{i-1}(T\mathbf{x}) = T^i\mathbf{x} = 0 \implies \mathbf{x} \in \ker(T^i)$，即 $\ker(T^{i+1}) \subseteq \ker(T^i)$，矛盾。因此 $\ker(T^{i-1}) \subsetneq \ker(T^i)$ 对所有 $i = 1, 2, \ldots, m$ 成立。

定义商空间 $W_i = \ker(T^i)/\ker(T^{i-1})$ $(i = 1, 2, \ldots, m)$，设 $\mathcal{B}'_i = \{a^i_1 + \ker(T^{i-1}), a^i_2 + \ker(T^{i-1}), \ldots, a^i_{\ell_i} + \ker(T^{i-1})\}$ 是 W_i 的一组基，其中 $\ell_i = \dim W_i$。取基元素的代表元 $\mathcal{B}_i = \{a^i_1, a^i_2, \ldots, a^i_{\ell_i}\} \subseteq \ker(T^i)$，下面证明 $\mathcal{B} = \bigcup_{i=1}^m \mathcal{B}_i$ 是 V 的一组基。

先验证 $m = 2$ 的特殊情形：此时核升链为 $\{0\} = \ker(T^0) \subsetneq \ker(T) \subsetneq \ker(T^2) = V$。令 $U = \ker(T)$，其基为 $\mathcal{B}_1 = \{a^1_1, \ldots, a^1_{\ell_1}\}$；商空间 $W_2 = V/U$ 的基为 $\mathcal{B}'_2 = \{a^2_1 + U, \ldots, a^2_{\ell_2} + U\}$，对应代表元 $\mathcal{B}_2 = \{a^2_1, \ldots, a^2_{\ell_2}\}$。

- 张成性：对任意 $\mathbf{v} \in V$，由商空间的定义，存在唯一的 $\bar{\mathbf{v}} = \mathbf{v} + U \in W_2$，可表示为 $\bar{\mathbf{v}} = k_1(a^2_1+U)+\cdots+k_{\ell_2}(a^2_{\ell_2}+U) = (k_1a^2_1+\cdots+k_{\ell_2}a^2_{\ell_2})+U$，因此 $\mathbf{v}-(k_1a^2_1+\cdots+k_{\ell_2}a^2_{\ell_2}) \in U$。而 U 由 $\mathcal{B}_1$ 张成，故存在 $c_1, \ldots, c_{\ell_1}$ 使得 $\mathbf{v} = c_1a^1_1 + \cdots + c_{\ell_1}a^1_{\ell_1} + k_1a^2_1 + \cdots + k_{\ell_2}a^2_{\ell_2}$，即 $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ 张成 V。

- 线性无关性：设 $c_1a^1_1 + \cdots + c_{\ell_1}a^1_{\ell_1} + k_1a^2_1 + \cdots + k_{\ell_2}a^2_{\ell_2} = 0$，则 $k_1a^2_1 + \cdots + k_{\ell_2}a^2_{\ell_2} = -(c_1a^1_1 + \cdots + c_{\ell_1}a^1_{\ell_1}) \in U$，即 $k_1(a^2_1 + U) + \cdots + k_{\ell_2}(a^2_{\ell_2} + U) = 0 + U \in W_2$。因 $\mathcal{B}'_2$ 是 W_2 的基，故 $k_1 = \cdots = k_{\ell_2} = 0$，进而 $c_1a^1_1 + \cdots + c_{\ell_1}a^1_{\ell_1} = 0$。又 $\mathcal{B}_1$ 是 U 的基，故 $c_1 = \cdots = c_{\ell_1} = 0$，因此 $\mathcal{B}$ 线性无关。

- 维数匹配：$\dim V = \dim U + \dim(V/U) = \ell_1 + \ell_2 = |\mathcal{B}|$，故 $\mathcal{B}$ 是 V 的基。

对一般 m，用数学归纳法推广：假设对 $m = k$ 时结论成立，当 $m = k+1$ 时，令 $U = \ker(T^k)$，则 $T|_U$ 的最小多项式为 x^k，由归纳假设 U 有基 $\bigcup_{i=1}^k \mathcal{B}_i$；商空间 $W_{k+1} = V/U$ 的基为 $\mathcal{B}'_{k+1}$，对应代表元 $\mathcal{B}_{k+1}$，同理可证 $\bigcup_{i=1}^{k+1} \mathcal{B}_i$ 是 V 的基。因此 $\mathcal{B} = \bigcup_{i=1}^m \mathcal{B}_i$ 是 V 的基。

对 $i = 1, 2, \ldots, m-1$，考虑集合 $S_i = \{T(\mathbf{w}_j) + \ker(T^{i-1}) \mid \mathbf{w}_j \in \mathcal{B}_{i+1}\}$，下面证明 S_i 是 $W_i = \ker(T^i)/\ker(T^{i-1})$ 中的线性无关子集。

- 包含性：因 $\mathbf{w}_j \in \mathcal{B}_{i+1} \subseteq \ker(T^{i+1})$，故 $T^{i+1}(\mathbf{w}_j) = 0 \implies T^i(T\mathbf{w}_j) = 0 \implies T\mathbf{w}_j \in \ker(T^i)$，因此 $T(\mathbf{w}_j) + \ker(T^{i-1}) \in \ker(T^i)/\ker(T^{i-1}) = W_i$，即 $S_i \subseteq W_i$。

- 线性无关性：设 $\sum_{j=1}^{\ell_{i+1}} k_j(T(\mathbf{w}_j)+\ker(T^{i-1})) = 0+\ker(T^{i-1}) \in W_i$，则 $T\left(\sum_{j=1}^{\ell_{i+1}} k_j\mathbf{w}_j\right) \in \ker(T^{i-1}) \implies T^i\left(\sum_{j=1}^{\ell_{i+1}} k_j\mathbf{w}_j\right) = 0 \implies \sum_{j=1}^{\ell_{i+1}} k_j\mathbf{w}_j \in \ker(T^i)$。又 $\mathcal{B}'_{i+1} = \{\mathbf{w}_j + \ker(T^i) \mid j = 1, \ldots, \ell_{i+1}\}$ 是 $W_{i+1} = \ker(T^{i+1})/\ker(T^i)$ 的基，故 $\sum_{j=1}^{\ell_{i+1}} k_j(\mathbf{w}_j+\ker(T^i)) = 0 + \ker(T^i) \implies k_j = 0$ 对所有 j 成立。因此 S_i 线性无关。

由此可得 $|S_i| = |\mathcal{B}_{i+1}| = \ell_{i+1} \le \dim W_i = \ell_i$，即 $\ell_{i+1} \le \ell_i$ 对所有 $i = 1, \ldots, m-1$ 成立。

从 $\mathcal{B}_m$ 开始迭代构造 Jordan 基：

- 设 $\mathcal{B}'_m = \{u_1^m + \ker(T^{m-1}), u_2^m + \ker(T^{m-1}), \ldots, u_{\ell_m}^m + \ker(T^{m-1})\}$ 是 $W_m = \ker(T^m)/\ker(T^{m-1}) = V/\ker(T^{m-1})$ 的基，对应代表元 $\mathcal{B}_m = \{u_1^m, u_2^m, \ldots, u_{\ell_m}^m\}$。

- 考虑 $S_{m-1} = \{T(u_1^m) + \ker(T^{m-2}), \ldots, T(u_{\ell_m}^m) + \ker(T^{m-2})\} \subseteq W_{m-1} = \ker(T^{m-1})/\ker(T^{m-2})$，由步骤 3 知 S_{m-1} 线性无关。将其扩张为 W_{m-1} 的基 $\mathcal{B}'_{m-1} = S_{m-1} \cup \{u_1^{m-1} + \ker(T^{m-2}), \ldots, u_{\ell_{m-1}-\ell_m}^{m-1} + \ker(T^{m-2})\}$，对应代表元 $\mathcal{B}_{m-1} = \{T(u_1^m), \ldots, T(u_{\ell_m}^m), u_1^{m-1}, \ldots, u_{\ell_{m-1}-\ell_m}^{m-1}\}$。

- 重复上述过程：对 $i = m-2, \ldots, 1$，将 $S_i = \{T(\mathbf{w}) + \ker(T^{i-1}) \mid \mathbf{w} \in \mathcal{B}_{i+1}\}$ 扩张为 W_i 的基，得到 $\mathcal{B}_i$。最终 $\mathcal{B} = \bigcup_{i=1}^m \mathcal{B}_i$ 是 V 的基。

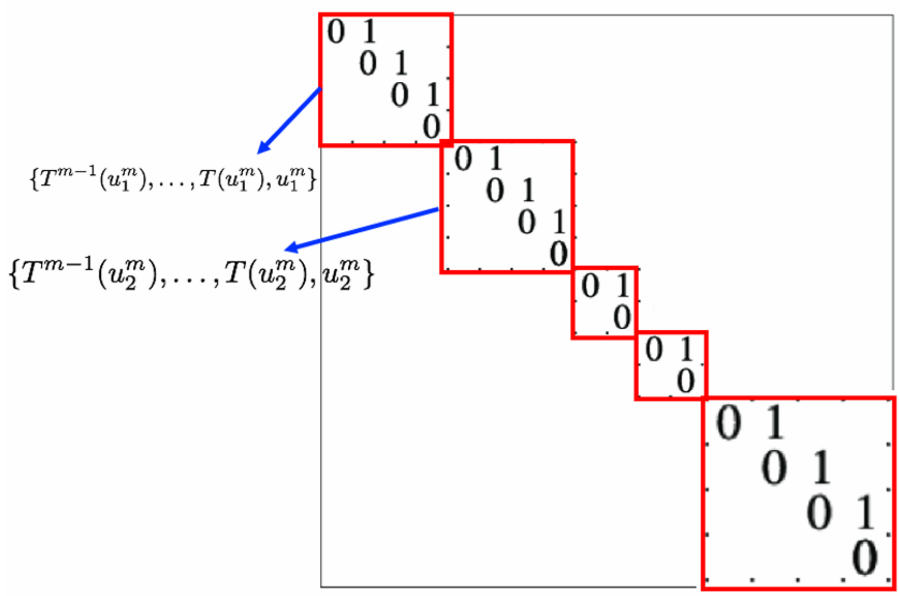
将 $\mathcal{B}$ 按 Jordan 链重新排列为 $\mathcal{A}$：对每个 u_j^m，形成链 $T^{m-1}(u_j^m), T^{m-2}(u_j^m), \ldots, T(u_j^m), u_j^m$；对扩张元 u_k^i $(i < m)$，形成链 $T^{i-1}(u_k^i), \ldots, T(u_k^i), u_k^i$。所有链按长度递减顺序排列，即构成 Jordan 基 $\mathcal{A}$。

验证 T 在基 $\mathcal{A}$ 下的矩阵为 Jordan 块的直和：

- 对角元：对任意链中的向量 $\mathbf{v} = T^k(u_j^i)$ $(k \geq 1)$，有 $T(\mathbf{v}) = T^{k+1}(u_j^i)$；当 $k = i$ 时，$T^i(u_j^i) = 0$（因 $u_j^i \in \ker(T^i)$），故 $T(T^{i-1}(u_j^i)) = 0$，即矩阵对角元全为 0。

- 超对角元：对链中向量 $\mathbf{v} = T^k(u_j^i)$ $(k \geq 1)$，若 $k < i$，则 $T(\mathbf{v}) = T^{k+1}(u_j^i)$ 是链中的下一个向量；若 $k = i$，则 $T(\mathbf{v}) = 0$。因此每个链对应一个 Jordan 块 $J_t = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$（$t$ 为链长），故 T 在 $\mathcal{A}$ 下的矩阵为 $\operatorname{diag}(J_1, J_2, \ldots, J_s)$，即 Jordan 块的直和。

综上，命题得证。 □



6.1.2 广义特征向量，广义特征值与广义特征子空间及其性质

Definition

广义特征向量与广义特征值（Generalized eigenvectors and Generalized eigenvalues）：设 T 是向量空间 V 上的线性算子，λ 是 T 的特征值。若存在正整数 p，使得 $(T-\lambda I)^p v=0$，则称 v 为 T 对应于 λ 的广义特征向量，λ 称为广义特征值。

广义特征向量是对特征向量概念的推广，用于解决一般矩阵无法对角化时的结构分析。

Exercise

证明对任意 $i\leq j$，有 $\ker(T-\lambda)^i\subset\ker(T-\lambda)^j$ 且 $\operatorname{im}(T-\lambda)^i\supset\operatorname{im}(T-\lambda)^j$。

类似我们在上一章的思维过程，我们也可以将特征子空间进行推广：

回顾广义特征向量的定义：存在某个正整数 m，使得 $(T-\lambda I)^m v=0$。

换句话说，只要有一个 m 能让 v 被 $(T-\lambda I)^m$ 映到 0，v 就是广义特征向量。因此，我们需要把所有这样的 v 收集起来，而这些 v 恰好分布在各个 $\ker((T-\lambda I)^p)$ 中（因为 $v\in\ker((T-\lambda I)^m)$ 就满足条件）。

Definition

广义特征子空间（Generalized eigenspace）：对于特征值 λ，广义特征子空间 $K_\lambda=\bigcup_{p=1}^{\infty}N((T-\lambda I)^p)$，即所有满足 $(T-\lambda I)^p v=0$（p 为正整数）的向量 v 构成的子空间。

当然，也有教材将其计为：$V_\lambda^a=\bigcup_{m\in\mathbb{N}_+}\ker((T-\lambda I)^m)$

接下来，我们将考察其性质：

Property

T - 不变性：$T(K_\lambda)\subseteq K_\lambda$，即 K_λ 在 T 的作用下保持不变。

证明. 任取 $v\in K_\lambda$，由广义特征子空间的定义，存在正整数 p 使得 $(T-\lambda I)^p v=0$。

对 $T(v)$ 计算 $(T-\lambda I)^p$ 的作用：$(T-\lambda I)^p(T(v))=T\left((T-\lambda I)^p v\right)=T(0)=0$

这说明 $T(v)\in N((T-\lambda I)^p)\subseteq K_\lambda$，故 $T(K_\lambda)\subseteq K_\lambda$，即 K_λ 是 T- 不变子空间。 □

Property

对于任意标量 $\mu\neq\lambda$，$T-\mu I$ 在 K_λ 上的限制是一一对应的。

证明. □

Property

若 λ 是 T 的特征值，重数为 m，则:
(a) $\dim(K_\lambda) \leq m$;
(b) $K_\lambda = N((T-\lambda I)^m)$。

证明. □

Lemma

（几何重数的代数刻画）设 V 是有限维 $\mathbb{F}$-向量空间，$T: V \to V$ 是 $\mathbb{F}$-线性映射。若 λ 是 T 的特征值，则 $m_a(\lambda) = \dim_{\mathbb{F}}(V_\lambda^a) = dim(K_\lambda)$。

证明. 设 V 是有限维 $\mathbb{F}$-向量空间，$T: V \to V$ 是线性映射，$\lambda \in \mathbb{F}$ 是 T 的特征值，$m_a(\lambda)$ 为 λ 的代数重数（即特征多项式 $\mathfrak{p}_T(x) = \det(xI - T)$ 中 $(x-\lambda)$ 的重数），$V_\lambda^a = \bigcup_{k\in\mathbb{N}_+} \ker((T-\lambda I)^k)$ 为广义 λ-特征子空间。我们需证明：$\dim_{\mathbb{F}}(V_\lambda^a) = m_a(\lambda)$。

采用 ** 对空间维数 $\dim V$ 的数学归纳法 **，步骤如下：

一、**基例：** $\dim V = 1$ 当 $\dim V = 1$ 时，V 的任意线性映射均可表示为 ** 数乘映射 **：存在 $\mu \in \mathbb{F}$，使得对所有 $v \in V$，$T(v) = \mu v$。分两种情况讨论：1. 若 $\lambda = \mu$：特征多项式 $\mathfrak{p}_T(x) = x - \lambda$，故代数重数 $m_a(\lambda) = 1$。对任意 $v \in V$，$(T-\lambda I)v = \mu v - \lambda v = 0$，即 $v \in \ker(T-\lambda I) \subset V_\lambda^a$，因此 $V_\lambda^a = V$，故 $\dim(V_\lambda^a) = \dim V = 1 = m_a(\lambda)$。2. 若 $\lambda \neq \mu$：特征多项式 $\mathfrak{p}_T(x) = x - \mu$，$\lambda$ 不是特征根，故 $m_a(\lambda) = 0$。此时 $T - \lambda I$ 是可逆映射（1 维空间上非零数乘映射可逆），故 $\ker(T-\lambda I) = \{0\}$。又因核的单调性（$\ker((T-\lambda I)^k) \subset \ker((T-\lambda I)^{k+1})$），所有 $\ker((T-\lambda I)^k) = \{0\}$，故 $V_\lambda^a = \{0\}$，$\dim(V_\lambda^a) = 0 = m_a(\lambda)$。

基例验证成立。

二、**归纳假设**假设：对所有维数 $\dim U < n$ 的有限维 $\mathbb{F}$-向量空间 U，以及任意线性映射 $S: U \to U$、任意特征值 $\lambda \in \mathbb{F}$，均有 $\dim(U_\lambda^a) = m_a^S(\lambda)$（其中 U_λ^a 是 S 的广义 λ-特征子空间，$m_a^S(\lambda)$ 是 S 的代数重数）。

三、**归纳步骤：** $\dim V = n$ 令 $A = T - \lambda I$（简化记号），定义 $W = \mathrm{Im}(A) = A(V) \subseteq V$（即 $T - \lambda I$ 的像空间）。分两种情况分析：

情况 1： $W = V$ 若 $W = V$，则 $A = T - \lambda I$ 是 ** 满射 **。由于 V 是有限维空间，根据“有限维空间上线性映射满射等价于单射”（秩-零度定理：$\dim V = \dim(\ker A) + \dim(\mathrm{Im} A)$，若 $\dim(\mathrm{Im} A) = n$，则 $\dim(\ker A) = 0$，即 A 单射），故 A 可逆。这意味着

$\ker A = \{0\}$，且对所有 $k \geq 1$，$\ker(A^k) = \{0\}$（可逆映射的幂仍可逆），因此 $V_\lambda^a = \{0\}$，即 λ 不是 T 的特征值，与题设矛盾。故情况 1 不可能发生。

情况 2： $W \subsetneq V$ **（必成立）** 此时 $\dim W < n$，且 W 是 T-不变子空间：- 证明：对任意 $w \in W$，存在 $v \in V$ 使得 $w = A(v) = (T - \lambda I)v$。对 w 应用 T，得 $T(w) = T(T - \lambda I)v = (T - \lambda I)T(v)$（线性映射可交换），而 $T(v) \in V$，故 $T(w) \in \mathrm{Im}(A) = W$，因此 W 对 T 封闭，即 T-不变。

基于 W 的 T-不变性，定义两个诱导映射：1. ** 限制映射 **：$T|_W : W \to W$，即 T 在子空间 W 上的限制（因 W 不变，$T|_W$ 是良定的线性映射）；2. ** 商映射 **：$\bar{T} : V/W \to V/W$，定义为 $\bar{T}(v + W) = T(v) + W$（因 W 不变，$T(v) + W$ 与代表元 v 的选择无关，良定）。

关键性质推导： 1. 特征多项式的分解：有限维空间上，若 W 是 T-不变子空间，则 T 的特征多项式等于限制映射 $T|_W$ 与商映射 $\bar{T}$ 的特征多项式的乘积：

$$\mathfrak{p}_T(x) = \mathfrak{p}_{T|_W}(x) \cdot \mathfrak{p}_{\bar{T}}(x)$$

（理由：取 W 的一组基 $\alpha_1, \ldots, \alpha_k$ 并扩张为 V 的基 $\alpha_1, \ldots, \alpha_k, \alpha_{k+1}, \ldots, \alpha_n$，则 T 在该基下的矩阵为分块上三角矩阵 $\begin{pmatrix} M_{T|_W} & * \\ 0 & M_{\bar{T}} \end{pmatrix}$，行列式为对角块行列式的乘积，即特征多项式分解。）

2. 商映射 $\bar{T}$ 的特征多项式：因 $W = \mathrm{Im}(A) = \mathrm{Im}(T - \lambda I)$，故对任意 $v \in V$，$A(v) = (T - \lambda I)v \in W$，即 $T(v) - \lambda v \in W$。对应商空间中，$T(v) + W = \lambda v + W$，即：

$$\bar{T}(v + W) = \lambda(v + W)$$

这说明 $\bar{T}$ 是商空间 V/W 上的 ** 数乘映射 λI**，因此其特征多项式为：

$$\mathfrak{p}_{\bar{T}}(x) = (x - \lambda)^{\dim(V/W)}$$

（$\dim(V/W) = n - \dim W$ 是商空间维数公式）。

3. 代数重数的分解：设 $m = m_a^{T|_W}(\lambda)$（λ 对 $T|_W$ 的代数重数），则由特征多项式分解：

$$\mathfrak{p}_T(x) = \mathfrak{p}_{T|_W}(x) \cdot (x - \lambda)^{\dim(V/W)}$$

对比两边 $(x - \lambda)$ 的重数，得代数重数的分解式：

$$m_a(\lambda) = m + \dim(V/W) \tag{1}$$

4. 限制映射的广义特征子空间：令 $W_\lambda^a = \bigcup_{k \in \mathbb{N}_+} \ker((T|_W - \lambda I)^k)$（$T|_W$ 的广义 λ-特征子空间）。因 $\dim W < n$，满足归纳假设，故：

$$\dim(W_\lambda^a) = m \tag{2}$$

构造辅助映射与秩-零度定理应用： 定义线性映射 $\varphi : V_\lambda^a \to W_\lambda^a$，满足 $\varphi(v) = A(v) = (T-\lambda I)v$ （$v \in V_\lambda^a$）。需验证 φ 良定（即 $\mathrm{Im}\varphi \subseteq W_\lambda^a$）：- 对任意 $v \in V_\lambda^a$，存在 $k \geq 1$ 使得 $A^k(v) = 0$。则 $A^{k-1}(\varphi(v)) = A^{k-1}(A(v)) = A^k(v) = 0$，故 $\varphi(v) \in \ker(A^{k-1}) \subseteq W_\lambda^a$（因 $\varphi(v) = A(v) \in W$，且 A^{k-1} 在 W 上的限制即 $(T|_W - \lambda I)^{k-1}$），因此 φ 良定。

接下来分析 φ 的核与像：- 核 $\ker\varphi$: $v \in \ker\varphi \iff A(v) = 0 \iff v \in \ker(A) = V_\lambda$（普通特征子空间），故 $\ker\varphi = V_\lambda$；- 像 $\mathrm{Im}\varphi \subseteq W_\lambda^a$（已验证）。

对 φ 应用 ** 秩-零度定理 **（dim(定义域) = dim(ker) + dim(像)）：

$$\dim(V_\lambda^a) = \dim(\ker\varphi) + \dim(\mathrm{Im}\varphi) = \dim(V_\lambda) + \dim(\mathrm{Im}\varphi) \tag{3}$$

关键等式推导： $\dim(\mathbf{Im}\varphi) = m$ 需证明 $\mathrm{Im}\varphi = W_\lambda^a$（即 φ 是满射）：- 对任意 $w \in W_\lambda^a$，存在 $k \geq 1$ 使得 $(T|_W - \lambda I)^k w = 0$，即 $A^k(w) = 0$。因 $w \in W = \mathrm{Im}(A)$，存在 $v \in V$ 使得 $w = A(v)$。此时：$A^{k+1}(v) = A^k(A(v)) = A^k(w) = 0$，故 $v \in \ker(A^{k+1}) \subseteq V_\lambda^a$，且 $\varphi(v) = A(v) = w$。因此 $w \in \mathrm{Im}\varphi$，即 $W_\lambda^a \subseteq \mathrm{Im}\varphi$。

结合 $\mathrm{Im}\varphi \subseteq W_\lambda^a$，得 $\mathrm{Im}\varphi = W_\lambda^a$，故 $\dim(\mathrm{Im}\varphi) = \dim(W_\lambda^a) = m$（由式 (2)）。

最终维度等式： 将 $\dim(\mathrm{Im}\varphi) = m$ 代入式 (3)：

$$\dim(V_\lambda^a) = \dim(V_\lambda) + m \tag{4}$$

再结合商空间维数公式与秩-零度定理（对 $A = T - \lambda I$）：

$$\dim(V/W) = \dim(\ker A) + \dim(\mathrm{Im}A) - \dim(\mathrm{Im}A) = \dim(\ker A) = \dim(V_\lambda)$$

（因 $\dim V = \dim(\ker A) + \dim(\mathrm{Im}A) = \dim(V_\lambda) + \dim W$，故 $\dim(V/W) = n - \dim W = \dim(V_\lambda)$）。

因此 $\dim(V_\lambda) = \dim(V/W)$，代入式 (4) 并结合式 (1)：

$$\dim(V_\lambda^a) = \dim(V/W) + m = m_a(\lambda)$$

四、结论 由数学归纳法，对所有有限维 $\mathbb{F}$-向量空间 V 及线性映射 $T : V \to V$，特征值 λ 的广义特征子空间维数等于其代数重数，即：

$$\dim_{\mathbb{F}}(V_\lambda^a) = m_a(\lambda)$$

□

接下来，我们将介绍本 section 中最重要的 2 个主定理：

6.1.3 主定理 1：Theorem 7.3

Theorem

设 T 是有限维向量空间 V 上的线性算子，且 T 的特征多项式可分解，$\lambda_1, \lambda_2, \ldots, \lambda_k$ 是 T 的不同特征值。则对于每一个 $x \in V$，存在向量 $v_i \in K_{\lambda_i}$ $(1 \leq i \leq k)$，使得 $x = v_1 + v_2 + \cdots + v_k$。

该定理表明，在特征多项式可分解的条件下，向量空间 V 的任意 v 可由各个广义特征子空间 K_{λ_i} 的向量线性表示。

6.1.4 主定理 2：Theorem 7.4

Theorem

设 T 是有限维向量空间 V 上的线性算子，T 的特征多项式可分解，$\lambda_1, \lambda_2, \ldots, \lambda_k$ 是 T 的不同特征值，对应的重数分别为 $m_1, m_2, \ldots, m_k$。对于 $1 \leq i \leq k$，设 β_i 是 K_{λ_i} 的有序基。则：
(a) 当 $i \neq j$ 时，$\beta_i \cap \beta_j = \varnothing$；
(b) $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ 是 V 的有序基；
(c) 对所有的 i，$\dim(K_{\lambda_i}) = m_i$。

此定理进一步阐述了广义特征子空间的基与整个向量空间基的关系，以及广义特征子空间的维数与特征值重数的关系。

6.1.5 广义循环子空间

设 T 是向量空间 V 上的线性算子，x 是 T 对应于特征值 λ 的广义特征向量，p 是使得 $(T-\lambda I)^p x = 0$ 的最小正整数。

Definition

广义循环子空间（Generalized cyclic subspace）：由 $\{(T-\lambda I)^{p-1}x, (T-\lambda I)^{p-2}x, \ldots, (T-\lambda I)x, x\}$ 张成的子空间 $\text{span}\{(T-\lambda I)^{p-1}x, (T-\lambda I)^{p-2}x, \ldots, (T-\lambda I)x, x\}$ 称为 T 对应于 λ 的广义循环子空间。

在将线性算子 T 化为 Jordan 标准型时，广义循环子空间起着关键作用。

Jordan 标准型是一种特殊的矩阵形式，它能反映线性算子的很多本质特征：

每个广义循环子空间对应 Jordan 标准型中的一个 Jordan 块。

一个 Jordan 块是一个上三角矩阵，主对角线上元素都相等（等于特征值 λ），主对角线上面一条次对角线上元素都为 1，其余元素为 0。

通过确定广义循环子空间，可以找到合适的基，使得线性算子 T 在这组基下的矩阵表示为 Jordan 标准型，从而更方便地研究线性算子的性质，如特征值、特征向量、幂次运算等。

为了理解上述理论，我们将举 2 个形象的例子进行讨论：

Example

设矩阵 A 为:

$$A = \begin{pmatrix} 3 & 1 & -2 \\ -1 & 0 & 5 \\ -1 & -1 & 4 \end{pmatrix} \in M_{3\times 3}(R)$$

求 A 的 Jordan 标准型。

Example

若线性算子 T 在基 β 下的矩阵已知

$$[\mathrm{T}]_\beta = A = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & -2 \\ 0 & 0 & -1 \end{pmatrix}$$

求 A 的 Jordan 标准型。

6.2 Jordan 标准型的讨论 2

假设对于某个特征值 λ_i，广义特征子空间 K_{λ_i} 的有序基 β_i 是四个循环子空间的并：$\beta_i = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$，对应的循环长度分别为 $p_1 = 3$，$p_2 = 3$，$p_3 = 2$，$p_4 = 1$。此时，限制算子 T_i 在基 β_i 下的矩阵 A_i 为：

$$A_i = \left(\begin{array}{ccccccccc}
\lambda_i & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \\
0 & \lambda_i & 1 & 0 & 0 & 0 & 0 & 0 & \\
0 & 0 & \lambda_i & 0 & 0 & 0 & 0 & 0 & \\
\hline
0 & 0 & 0 & \lambda_i & 1 & 0 & 0 & 0 & \\
0 & 0 & 0 & 0 & \lambda_i & 1 & 0 & 0 & \\
0 & 0 & 0 & 0 & 0 & \lambda_i & 0 & 0 & \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & \lambda_i & 1 & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_i & \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_i
\end{array}\right)$$

为了直观表示该结构，我们引入点图（Dot Diagram）的概念。

6.2.1 点图的定义与构造规则

【定义】：点图（Dot Diagram）：是描述线性算子 T_i 在广义特征子空间 K_{λ_i} 上结构的可视化工具。对于基 β_i 中的每个循环子空间 γ_j，其点图构造规则如下：

1. 每个循环子空间 γ_j 对应点图中的一列，列高为循环长度 p_j；
2. 列的顺序按循环长度降序排列（$p_1 \geq p_2 \geq \cdots \geq p_{n_i}$）；
3. 每列中的点表示循环中的广义特征向量，从上至下对应向量的幂次递增（例如，顶部为 $(T - \lambda_i I)^{p_j - 1} v$，底部为 v）。

对于导引中的矩阵 A_i，其点图表示为：

$$\begin{array}{cccc}
\bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \\
\bullet & \bullet & &
\end{array}$$

其中：

- 第一行点数 $r_1 = 4$，对应特征子空间 E_{λ_i} 的个数；
- 第二行点数 $r_2 = 3$；
- 第三行点数 $r_3 = 2$。

6.2.2 点图的性质与核心定理

接下来，我们将揭示点图与幂零空间的关系：

【定理】：对于任意正整数 r，点图中前 r 行的所有点对应的向量构成 $\mathrm{N}((T-\lambda_i I)^r)$ 的一组基。因此，前 r 行的点数等于 dim 行的点数等于 $\dim \mathrm{N}((T-\lambda_i I)^r)$。

证明概要： 设 β_i 是 K_{λ_i} 的循环基，将其分为两部分：

- $S_1 = \{x \in \beta_i \mid (T-\lambda_i I)^r(x) = 0\}$（对应点图前 r 行的点）；
- $S_2 = \{x \in \beta_i \mid (T-\lambda_i I)^r(x) \neq 0\}$。

显然 $\dim \mathrm{N}((T-\lambda_i I)^r) = |S_1|$。通过验证 S_1 线性无关且生成 $\mathrm{N}((T-\lambda_i I)^r)$，即可证明结论。

那么如何进行点图行数的计算呢?

【定理】：设 r_j 表示点图中第 j 行的点数，则：

(a) $r_1 = \dim V - \mathrm{rank}(T-\lambda_i I)$；

(b) 当 $j>1$ 时，$r_j = \mathrm{rank}((T-\lambda_i I)^{j-1}) - \mathrm{rank}((T-\lambda_i I)^j)$。

证明概要： 容易知道：前 j 行的点数之和为 $\dim \mathrm{N}((T-\lambda_i I)^j)$，即：

$$r_1 + r_2 + \cdots + r_j = \dim V - \mathrm{rank}((T-\lambda_i I)^j)$$

通过递推可得：

$$r_j = \left[\dim V - \mathrm{rank}((T-\lambda_i I)^j)\right] - \left[\dim V - \mathrm{rank}((T-\lambda_i I)^{j-1})\right]$$

【推论】：Jordan 标准型的唯一性

对于给定的线性算子 T，其点图结构由特征多项式和各幂零空间的维数唯一确定。因此，在特征值顺序固定的情况下，Jordan 标准型是唯一的。

综上所述，点图提供了一种直观分析线性算子结构的方法：

- 通过矩阵幂的秩计算点图行数 r_j；
- 行数决定 Jordan 块的大小分布；
- 列数等于特征子空间维数，即 Jordan 块数量。

通过例 1 中的矩阵 A_i，我们可以验证：

- $r_1 = 4$ 对应 4 个 Jordan 块；
- 最大块大小为 3（对应 3 行）；
- 点图完全确定了 Jordan 标准型的结构。

上述理论确实抽象，因此我们需要多个例子来加深对这一过程的理解：

例 1：

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

例 2：

$$A=\begin{pmatrix}2 & -4 & 2 & 2\\ -2 & 0 & 1 & 3\\ -2 & -2 & 3 & 3\\ -2 & -6 & 3 & 7\end{pmatrix}$$

Chapter 7

内积空间与双线性

7.1 内积空间与范数

7.1.1 内积与内积空间

Definition

内积 (inner product)： 设 V 是数域 F 上的向量空间，二元函数 $\langle\cdot,\cdot\rangle : V\times V\to F$，若满足：

1. **共轭对称性：** 当 $F=\mathbb{C}$ 时，$\langle x,y\rangle=\overline{\langle y,x\rangle}$；当 $F=\mathbb{R}$ 时，$\langle x,y\rangle=\langle y,x\rangle$。
2. **对第一变量线性：** $\langle \alpha x+\beta y,z\rangle=\alpha\langle x,z\rangle+\beta\langle y,z\rangle$，其中 $\alpha,\beta\in F$。
3. **正定性：** $\langle x,x\rangle\geq 0$，且 $\langle x,x\rangle=0\iff x=0$。

则称 $\langle\cdot,\cdot\rangle$ 为 V 上的内积。

例如，我们高中接触到向量的点积就是内积的一种。

Definition

内积空间（Inner product space）： 赋予内积 $\langle\cdot,\cdot\rangle$ 的向量空间 V，称为内积空间。

Example

1. 欧几里得空间 $\mathbb{R}^n$，内积 $\langle x,y\rangle=\sum_{i=1}^n x_iy_i$。
2. 复空间 $\mathbb{C}^n$，内积 $\langle x,y\rangle=\sum_{i=1}^n x_i\overline{y_i}$。

接下来考察 3 个在给定向量空间中不是内积的例子。

Example

二维实向量空间：$\mathbb{R}^2$ 上的双线性形式在 $\mathbb{R}^2$ 上定义双线性形式：

$$\langle (a,b),(c,d)\rangle = ac - bd$$

证明. 内积要求正定性：对任意非零向量 x，$\langle x,x\rangle > 0$。取非零向量 $x=(0,1)\in\mathbb{R}^2$，计算得：$\langle (0,1),(0,1)\rangle = 0\cdot 0 - 1\cdot 1 = -1 < 0$ □

Example

二阶实矩阵空间：$\mathrm{M}_{2\times 2}(\mathbb{R})$ 上的双线性形式在 $\mathrm{M}_{2\times 2}(\mathbb{R})$ 上定义双线性形式：

$$\langle A,B\rangle = \mathrm{tr}(A+B)$$

其中 tr 表示矩阵的迹。

证明. 取非零矩阵 $A=-I$，计算得：$\langle A,A\rangle = \mathrm{tr}(-I+(-I)) = \mathrm{tr}(-2I) = -4 < 0$ 不满足正定性，故不构成内积。 □

Example

实系数多项式空间：$\mathrm{P}(\mathbb{R})$ 上的双线性形式在 $\mathrm{P}(\mathbb{R})$ 上定义双线性形式：

$$\langle f(x),g(x)\rangle = \int_0^1 f'(t)g(t)\,dt$$

其中 $f'(t)$ 表示 $f(t)$ 的导数。

证明. 内积要求正定性：$\langle f,f\rangle = 0$ 当且仅当 $f=0$。取非零多项式 $f(t)=1$（常数多项式），其导数 $f'(t)=0$，计算得：$\langle f,f\rangle = \int_0^1 0\cdot 1\,dt = 0$ 但 $f(t)=1\neq 0$，不满足正定性，故不构成内积。 □

接下来，我们将导引出标准内积的定义，这在研究中有较大的意义：

Definition

标准内积（standard inner product）：

1. 在 $\mathbb{R}^n$ 中，标准内积为 $\langle x,y\rangle = \sum_{i=1}^n x_i y_i$。
2. 在 $\mathbb{C}^n$ 中，标准内积为 $\langle x,y\rangle = \sum_{i=1}^n x_i \overline{y_i}$。

并且定义两个重要的内积空间：

Definition

1. **实内积空间**（Real inner product space）：数域 $F=\mathbb{R}$ 的内积空间，其内积满足 $\langle x,y\rangle=\langle y,x\rangle$。

2. **酉空间**（Unitary space）：数域 $F=\mathbb{C}$ 的内积空间，其内积满足 $\langle x,y\rangle=\overline{\langle y,x\rangle}$。

接下来，我们将讨论内积的性质：

Property

设 V 为内积空间，对于 $x,y,z\in V$ 和 $c\in F$，有以下性质：

1. $\langle x,y+z\rangle=\langle x,y\rangle+\langle x,z\rangle$。

2. $\langle x,cy\rangle=\bar{c}\langle x,y\rangle$。

3. $\langle x,0\rangle=\langle 0,x\rangle=0$。

4. 若对所有 $x\in V$，$\langle x,y\rangle=\langle x,z\rangle$，则 $y=z$。

证明. 1. 证明 $\langle x,y+z\rangle=\langle x,y\rangle+\langle x,z\rangle$：由共轭对称性，$\langle x,y+z\rangle=\overline{\langle y+z,x\rangle}$；由内积对第一个变元的加性，$\langle y+z,x\rangle=\langle y,x\rangle+\langle z,x\rangle$；对等式取共轭（共轭的加性：$\overline{a+b}=\bar{a}+\bar{b}$），得：

$$\overline{\langle y+z,x\rangle}=\overline{\langle y,x\rangle}+\overline{\langle z,x\rangle}=\langle x,y\rangle+\langle x,z\rangle.$$

故 $\langle x,y+z\rangle=\langle x,y\rangle+\langle x,z\rangle$。

2. 证明 $\langle x,cy\rangle=\bar{c}\langle x,y\rangle$：由共轭对称性，$\langle x,cy\rangle=\overline{\langle cy,x\rangle}$；由内积对第一个变元的 ** 齐次性 **，$\langle cy,x\rangle=c\langle y,x\rangle$；对等式取共轭（共轭的齐次性：$\overline{ca}=\bar{c}\cdot\bar{a}$），得：

$$\overline{\langle cy,x\rangle}=\bar{c}\cdot\overline{\langle y,x\rangle}=\bar{c}\langle x,y\rangle.$$

故 $\langle x,cy\rangle=\bar{c}\langle x,y\rangle$。

3. 证明 $\langle x,0\rangle=\langle 0,x\rangle=0$：- 证 $\langle x,0\rangle=0$：因 $0=0+0$，结合性质 1 的加性，$\langle x,0\rangle=\langle x,0+0\rangle=\langle x,0\rangle+\langle x,0\rangle$；两边减 $\langle x,0\rangle$，得 $\langle x,0\rangle=0$。- 证 $\langle 0,x\rangle=0$：由共轭对称性，$\langle 0,x\rangle=\overline{\langle x,0\rangle}=\bar{0}=0$。故 $\langle x,0\rangle=\langle 0,x\rangle=0$。

4. 证明：若对所有 $x\in V$，$\langle x,y\rangle=\langle x,z\rangle$，则 $y=z$：将等式变形为 $\langle x,y\rangle-\langle x,z\rangle=0$；结合性质 1 的加性，得 $\langle x,y-z\rangle=0$ $(\forall x\in V)$；取 $x=y-z$，则 $\langle y-z,y-z\rangle=0$；由内积的正定性，$\langle a,a\rangle=0\iff a=0$，故 $y-z=0$，即 $y=z$。

□

7.1.2 内积空间的范数及其性质

Definition

范数（Norm）：由内积定义范数 $\|x\| = \sqrt{\langle x, x \rangle}$，它刻画了向量的“长度”。

Property

设 V 为内积空间，对于 $x, y \in V$ 和 $c \in F$，有以下性质：

1. $\|x\| \geq 0$
2. $\|cx\| = |c| \cdot \|x\|$。
3. $\|x\| = 0$ 当且仅当 $x = 0$。
4. 柯西–施瓦茨不等式（*Cauchy-Schwarz* inequality)：$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$。
5. 三角不等式（Triangle inequality：$\|x + y\| \leq \|x\| + \|y\|$。

证明. 设 V 是内积空间，$x, y \in V$，c 为标量，$\|x\| = \sqrt{\langle x, x \rangle}$ 是内积诱导的范数。

1. $\|x\| \geq 0$ 由内积的半正定性，$\langle x, x \rangle \geq 0$，故 $\|x\| = \sqrt{\langle x, x \rangle} \geq 0$。

2. $\|cx\| = |c| \cdot \|x\|$ 计算范数：

$$\|cx\| = \sqrt{\langle cx, cx \rangle} = \sqrt{c\bar{c}\langle x, x \rangle} = \sqrt{|c|^2\|x\|^2} = |c| \cdot \|x\|$$

3. $\|x\| = 0 \iff x = 0$ - 必要性：若 $x = 0$，则 $\langle 0, 0 \rangle = 0$，故 $\|0\| = \sqrt{0} = 0$；- 充分性：若 $\|x\| = 0$，则 $\langle x, x \rangle = \|x\|^2 = 0$，由内积正定性得 $x = 0$。

4. 柯西-施瓦茨不等式：$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$ 取实变量 t，构造二次函数：

$$f(t) = \langle x + ty, x + ty \rangle = \|x\|^2 + 2t\langle x, y \rangle + t^2\|y\|^2$$

因 $f(t) \geq 0$ 恒成立，判别式 $\Delta \leq 0$：

$$\Delta = (2\langle x, y \rangle)^2 - 4\|x\|^2\|y\|^2 \leq 0 \implies |\langle x, y \rangle| \leq \|x\| \cdot \|y\|$$

5. 三角不等式：$\|x + y\| \leq \|x\| + \|y\|$ 展开范数平方：

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2$$

由柯西-施瓦茨不等式，$2\langle x, y \rangle \leq 2\|x\| \cdot \|y\|$，故：

$$\|x + y\|^2 \leq (\|x\| + \|y\|)^2 \implies \|x + y\| \leq \|x\| + \|y\|$$

□

对于内积空间上的线性算子，我们来考察一种特殊情节：即保范

Theorem

设 T 是内积空间 V 上的一个线性算子，且假设对所有的 x 都有 $\|T(x)\| = \|x\|$。则 T 是单射。

证明. 要证 T 是单射，只需证：若 $T(x) = 0$，则 $x = 0$。

1. 由 T 是保范算子，对任意 $x \in V$，有 $\|T(x)\| = \|x\|$；

2. 若 $T(x) = 0$，则 $\|T(x)\| = \|0\| = 0$（零向量的范数为 0）；结合保范条件，得 $\|x\| = \|T(x)\| = 0$；

3. 根据范数的正定性（$\|x\| = 0 \iff x = 0$），故 $x = 0$。

因此，保范线性算子 T 的核空间仅含零向量，即 T 是单射。 □

【Remark】：

线性算子 T 作用在任意向量 x 上，像的范数与原向量的范数完全相等（这种算子也叫“等距线性算子”）。

接下来，我们将结合几何层面，来增加你对内积的理解。

7.1.3 平行四边形法则和极化恒等式

内积空间 V 上的平行四边形法则:

Theorem

对于所有 $x, y \in V$，有 $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$ 。

证明. □

接下来，我们将讨论另一个重要的结论：极化恒等式

Recall：在 $\triangle ABC$ 中，D 是 BC 的中点，$BC = 6$，$AD = 5$，如何求 $\overrightarrow{AB} \cdot \overrightarrow{AC}$ 的值?

根据极化恒等式：在三角形中，若 D 是 BC 边的中点，则 $\overrightarrow{AB} \cdot \overrightarrow{AC} = \overrightarrow{AD}^2 - \overrightarrow{BD}^2$。已知 D 是 BC 中点，$BC = 6$，则 $BD = \frac{1}{2}BC = 3$，又 $AD = 5$。将 $\overrightarrow{AD}^2 = 5^2 = 25$，$\overrightarrow{BD}^2 = 3^2 = 9$ 代入极化恒等式，可得 $\overrightarrow{AB} \cdot \overrightarrow{AC} = 25 - 9 = 16$。

那么在内积空间中呢?

Theorem

极化恒等式(Polarization identity): 设 V 是数域 F 上的内积空间, 对于所有 $x, y \in V$,
(a) 若 $F = R$，则 $\langle x, y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2)$;
(b) 若 $F = C$，则 $\langle x, y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2)$，其中 $i^2 = -1$。

7.1.4 范数的本质定义

设 V 是数域 F（F 为 $\mathbb{R}$ 或 $\mathbb{C}$）上的向量空间。无论 V 是否为内积空间，**范数** $\|\cdot\|$ 是 V 上的实值函数，对任意 $x, y \in V$ 和 $a \in F$，满足以下三个条件:

1. $\|x\| \geq 0$，且 $\|x\| = 0$ 当且仅当 $x = 0$。
2. $\|ax\| = |a| \cdot \|x\|$。
3. $\|x+y\| \leq \|x\| + \|y\|$（三角不等式）。

这在数学分析中的度量空间与机器学习中词嵌入中有非常大的应用。

所以我们可以得到**距离的定义**:

Definition

设 $\|\cdot\|$ 是向量空间 V 上的范数，定义 $x, y \in V$ 间的**距离 (distance)** 为:

$$d(x, y) = \|x - y\|$$

以下性质对任意 $x, y, z \in V$ 成立:

1. $d(x, y) \geq 0$（非负性）。
2. $d(x, y) = d(y, x)$（对称性）。
3. $d(x, y) \leq d(x, z) + d(z, y)$（三角不等式）。
4. $d(x, x) = 0$。
5. 若 $x \neq y$，则 $d(x, y) \neq 0$。

接下来，我们将考察矩阵空间的范数和连续函数空间的范数:

Definition

设 $V = \mathrm{M}_{m\times n}(F)$ 为 $m \times n$ 矩阵空间，对任意 $A \in V$，定义：

$$\|A\| = \max_{i,j} |A_{ij}|$$

证明此函数满足范数的三个条件。

证明.

1. **非负性：** 对任意 $A \in M_{m\times n}(F)$，$\|A\| = \max_{i,j} |A_{ij}| \geq 0$；且 $\|A\| = 0 \iff \max_{i,j} |A_{ij}| = 0 \iff A_{ij} = 0$（$\forall i, j$）$\iff A = 0$。
2. **齐次性：** 对任意 $\lambda \in F$，$\|\lambda A\| = \max_{i,j} |\lambda A_{ij}| = |\lambda| \cdot \max_{i,j} |A_{ij}| = |\lambda| \cdot \|A\|$。
3. **三角不等式：** 对任意 $A, B \in M_{m\times n}(F)$，$|A_{ij} + B_{ij}| \leq |A_{ij}| + |B_{ij}| \leq \|A\| + \|B\|$（$\forall i, j$），故 $\|A + B\| = \max_{i,j} |A_{ij} + B_{ij}| \leq \|A\| + \|B\|$。

$\square$

Definition

极大元素范数 (Maximum element norm)：设 $V = \mathrm{C}([0, 1])$ 为闭区间 $[0, 1]$ 上的连续函数空间，对任意 $f \in V$，定义：

$$\|A\|_{\max} = \max_{1\leq i\leq m, 1\leq j\leq n} |A_{ij}|.$$

证明此函数满足范数的三个条件。

证明.

1. **非负性：** 对任意 $f \in C([0, 1])$，$\|f\| = \max_t |f(t)| \geq 0$；且 $\|f\| = 0 \iff |f(t)| = 0$（$\forall t \in [0, 1]$）$\iff f$ 为零函数。
2. **齐次性：** 对任意 $\lambda \in \mathbb{R}$（或 $\mathbb{C}$），$\|\lambda f\| = \max_t |\lambda f(t)| = |\lambda| \cdot \max_t |f(t)| = |\lambda| \cdot \|f\|$。
3. **三角不等式：** 对任意 $f, g \in C([0, 1])$，$|f(t) + g(t)| \leq |f(t)| + |g(t)| \leq \|f\| + \|g\|$（$\forall t \in [0, 1]$），故 $\|f + g\| = \max_t |f(t) + g(t)| \leq \|f\| + \|g\|$。

$\square$

Definition

算子范数 (operator norm): 设 F 是数域，$M_{m\times n}(F)$ 为 $m\times n$ 矩阵空间，$\|\cdot\|_{\mathbb{V}}$ 是 n 维向量空间 F^n 上的向量范数。矩阵 $A\in M_{m\times n}(F)$ 的**算子范数**定义为:

$$\|A\| = \sup_{\|x\|_{\mathbb{V}}\neq 0} \frac{\|Ax\|_{\mathbb{V}}}{\|x\|_{\mathbb{V}}},$$

其中 $x\in F^n$ 是非零向量。

证明. 算子范数需满足范数的三个核心条件：非负性、齐次性、三角不等式。

1. 非负性对任意 $A\in M_{m\times n}(F)$，因向量范数 $\|\cdot\|_{\mathbb{V}}\geq 0$，故 $\frac{\|Ax\|_{\mathbb{V}}}{\|x\|_{\mathbb{V}}}\geq 0$（$\|x\|_{\mathbb{V}}\neq 0$），从而 $\|A\| = \sup_{\|x\|_{\mathbb{V}}\neq 0}\frac{\|Ax\|_{\mathbb{V}}}{\|x\|_{\mathbb{V}}}\geq 0$。

若 $A=0$，则 $Ax=0$，故 $\|A\|=0$；若 $\|A\|=0$，则 $\|Ax\|_{\mathbb{V}}=0$ 对所有 x 成立，由向量范数正定性得 $Ax=0$，故 $A=0$。

2. 齐次性对任意 $\lambda\in F$、$A\in M_{m\times n}(F)$，由向量范数齐次性 $\|\lambda y\|_{\mathbb{V}} = |\lambda|\|y\|_{\mathbb{V}}$，有:

$$\|\lambda A\| = \sup_{\|x\|_{\mathbb{V}}\neq 0}\frac{\|(\lambda A)x\|_{\mathbb{V}}}{\|x\|_{\mathbb{V}}} = |\lambda|\cdot\sup_{\|x\|_{\mathbb{V}}\neq 0}\frac{\|Ax\|_{\mathbb{V}}}{\|x\|_{\mathbb{V}}} = |\lambda|\|A\|.$$

3. 三角不等式对任意 $A,B\in M_{m\times n}(F)$，由向量范数三角不等式 $\|y+z\|_{\mathbb{V}}\leq\|y\|_{\mathbb{V}}+\|z\|_{\mathbb{V}}$，有:

$$\frac{\|(A+B)x\|_{\mathbb{V}}}{\|x\|_{\mathbb{V}}} = \frac{\|Ax+Bx\|_{\mathbb{V}}}{\|x\|_{\mathbb{V}}} \leq \frac{\|Ax\|_{\mathbb{V}}}{\|x\|_{\mathbb{V}}} + \frac{\|Bx\|_{\mathbb{V}}}{\|x\|_{\mathbb{V}}} \leq \|A\|+\|B\|.$$

取上确界得 $\|A+B\|\leq\|A\|+\|B\|$。

综上，算子范数满足范数的三个条件，故它是 $M_{m\times n}(F)$ 上的范数。 □

【Remark】:

极大元素范数 $\|\cdot\|_{\max}$ 不是算子范数。

Theorem

算子范数的相容性: 由向量范数诱导的算子范数满足矩阵乘法的相容性：对任意 $A\in M_{m\times n}(F)$、$B\in M_{n\times p}(F)$，有

$$\|AB\|\leq\|A\|\cdot\|B\|.$$

证明. 取任意非零向量 $x\in F^p$，令 $y=Bx$，则:

$$\|ABx\| = \|A(Bx)\| = \|Ay\| \leq \|A\|\cdot\|y\| = \|A\|\cdot\|Bx\|.$$

再由算子范数的性质，$\|Bx\|\leq\|B\|\cdot\|x\|$，因此:

$$\|ABx\|\leq\|A\|\cdot\|B\|\cdot\|x\|.$$

由于 $x \neq 0$，两边除以 $\|x\|$ 得：

$$\frac{\|ABx\|}{\|x\|} \leq \|A\| \cdot \|B\|.$$

对所有非零 x 取上确界，即得：

$$\|AB\| = \sup_{\|x\| \neq 0} \frac{\|ABx\|}{\|x\|} \leq \|A\| \cdot \|B\|.$$

证毕。 □

【Remark】:

$\|\cdot\|_{\max}$ 是 $M_{m\times n}(F)$ 上的范数（满足非负性、齐次性、三角不等式），但因不满足矩阵乘法的相容性，故不是算子范数。

证明. 算子范数需满足矩阵乘法的相容性（见上述定理），但 $\|\cdot\|_{\max}$ 不满足该性质：

取 $A = B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \in M_{2\times 2}(F)$，则

$$\|A\|_{\max} = \max\{|1|, |1|, |1|, |1|\} = 1.$$

计算矩阵乘积：

$$AB = \begin{pmatrix} 1\cdot 1 + 1\cdot 1 & 1\cdot 1 + 1\cdot 1 \\ 1\cdot 1 + 1\cdot 1 & 1\cdot 1 + 1\cdot 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix},$$

故

$$\|AB\|_{\max} = \max\{|2|, |2|, |2|, |2|\} = 2.$$

此时 $\|AB\|_{\max} = 2 > 1 = \|A\|_{\max} \cdot \|B\|_{\max}$，不满足相容性。因此 $\|\cdot\|_{\max}$ 不是算子范数。 □

7.2 正交基与 Gram-Schmidt 正交化

7.2.1 正交，正交基与标准正交基

Definition

正交（*Orthogonal*）：在内积空间 V 中，对于向量 $x, y \in V$，若 $\langle x, y\rangle = 0$，则称 x 与 y 正交。

这一概念是欧几里得空间中向量**垂直关系的推广**，基于内积定义，适用于一般内积空间。例如在连续函数空间 $C[a,b]$ 中，定义内积 $\langle f, g\rangle = \int_a^b f(x)g(x)dx$，若 $\int_a^b f(x)g(x)dx = 0$，则函数 f 与 g 正交。

Definition

正交基（Orthogonal basis）： 设 V 是内积空间，若向量组 $S=\{v_1,v_2,\cdots,v_n\}$ 满足 $v_i\neq 0$ 且 $\langle v_i,v_j\rangle=0$（$i\neq j$），则称 S 是 V 的一个正交基。

例如在三维欧几里得空间 $\mathbb{R}^3$ 中，$\{(1,0,0),(0,1.75,0),(0,0,5)\}$ 是正交基。

Definition

标准正交基（orthonormal basis）： 在正交基基础上，若每个向量 v_i 满足 $\|v_i\|=1$（$i=1,2,\cdots,n$），则称此正交基为标准正交基。

如 $\mathbb{R}^n$ 中的自然基 $\{e_1,e_2,\cdots,e_n\}$，e_i 第 i 分量为 1，其余为 0，$\langle e_i,e_j\rangle=\delta_{ij}$（克罗内科符号）。

7.2.2 克罗内科符号

Definition

克罗内科符号（Kronecker symbol）： δ_{ij} 定义为：$\delta_{ij}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$。

在标准正交基中，$\langle v_i,v_j\rangle=\delta_{ij}$，使公式推导更紧凑清晰，如向量坐标计算 $a_i=\langle y,v_i\rangle$ 等。

7.2.3 勾股定理

Theorem

勾股定理（Pythagorean theorem）： 设 V 是一个内积空间，假设 x 和 y 是 V 中的正交向量。$|x+y|^2=|x|^2+|y|^2$。

证明. **内积空间勾股定理：** 由内积诱导范数 $|z|^2=\langle z,z\rangle$，对正交向量 $x\perp y$（$\langle x,y\rangle=0$），展开得：

$$|x+y|^2=\langle x+y,x+y\rangle=\langle x,x\rangle+\langle x,y\rangle+\langle y,x\rangle+\langle y,y\rangle.$$

因 $\langle x,y\rangle=\langle y,x\rangle=0$，故 $|x+y|^2=|x|^2+|y|^2$。

□

由此推导出 $\mathbb{R}^2$ 中的勾股定理。

7.2.4 线性泛函初步理解

Definition

线性泛函（linear functional）：是从向量空间 V 到其标量域 F 的线性映射，即 $f: V \to F$，满足：

- 对任意 $x, y \in V$，$\alpha, \beta \in F$，有 $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$。

例如：矩阵的迹 $\mathrm{tr}: \mathrm{Mat}(n \times n, F) \to F$，满足 $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$，$\mathrm{tr}(cA) = c \cdot \mathrm{tr}(A)$，就是一种线性泛函。

系数确定): 设 $y = \sum_{i=1}^{k} a_i v_i$，对 $1 \leq j \leq k$，$\langle y, v_j \rangle = \sum_{i=1}^{k} a_i \langle v_i, v_j \rangle = a_j \|v_j\|^2$，故 $a_j = \frac{\langle y, v_j \rangle}{\|v_j\|^2}$。若 S 是标准正交基，$y = \sum_{i=1}^{k} \langle y, v_i \rangle v_i$，计算更简。

投影原理): 设 V 是内积空间，$S = \{v_1, v_2, \cdots, v_k\}$ 是由非零向量组成的正交子集，若 $y \in \mathrm{span}(S)$，$y = \sum_{i=1}^{k} \frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i$。$\frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i$ 是 y 在 v_i 方向的投影，体现向量按正交子集的分解。

投影与线性泛函的联系

1. 代数角度：对固定 $v \in V$，$f(x) = \langle x, v \rangle$ 是线性泛函，反映 x 与 v 的 “关联程度”。

2. 几何角度：投影是向量在子空间上的 “最佳逼近”，例如在 $\mathbb{R}^3$ 中向量向平面的投影。

例子：在 $\mathbb{R}^2$ 中，$f(x) = 2e_1 + 3e_2$ 可表示为 $\langle x, (2,3) \rangle$，体现 x 在 $(2,3)$ 方向的 “投影强度”。

7.2.5 标准正交基的好处

向量表示简洁，简化内积计算:

在标准正交基 $\{v_1, v_2, \cdots, v_n\}$ 下，任取 $y \in V$，设 $y = \sum_{i=1}^{n} a_i v_i$。根据内积性质，$\langle y, v_j \rangle = \langle \sum_{i=1}^{n} a_i v_i, v_j \rangle = a_j$，得 $a_i = \langle y, v_i \rangle$。无需解线性方程组，简化向量坐标表示。

在标准正交基 $\{v_1, v_2, \cdots, v_n\}$ 下，设 $x = \sum_{i=1}^{n} x_i v_i$，$y = \sum_{i=1}^{n} y_i v_i$，$\{v_1, \cdots, v_n\}$ 为标准正交基。则 $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$，形式如欧氏空间点积，方便计算与性质推导，如向量模长 $\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}$。

方便矩阵表示:

对于有限维内积空间 V，若有标准正交基 $\beta = \{v_1, v_2, \cdots, v_n\}$，线性算子 T 在该基下矩阵 $A = [\mathrm{T}]_\beta$。由推论，$A_{ij} = \langle \mathrm{T}(v_j), v_i \rangle$。证明源于 $\mathrm{T}(v_j) = \sum_{i=1}^{n} \langle \mathrm{T}(v_j), v_i \rangle v_i$，此表示使线性算子的矩阵与内积直接关联，便于研究算子性质。

既然正交基有这么多好处，那么我们怎么样才能把一组普通的基底化为正交基呢?

7.2.6 Gram-Schmidt 正交化

二或三个向量的正交化

我们可以简单的通过“投影”的思想来进行理解：

设向量组 $\{u_1, u_2\}$，令 $v_1 = u_1$。设 $v_2 = u_2 - \frac{\langle u_2, v_1\rangle}{\langle v_1, v_1\rangle}v_1$ 使 v_2 与 v_1 正交。

例如在 $\mathbb{R}^2$ 中，$u_1 = (1, 1)$，$u_2 = (2, 0)$，$v_2 = (2, 0) - \frac{2}{2}(1, 1) = (1, -1)$，$\langle v_1, v_2\rangle = 0$。

相似的，

对于 $\{u_1, u_2, u_3\}$，先令 $v_1 = u_1$；再令 $v_2 = u_2 - \frac{\langle u_2, v_1\rangle}{\langle v_1, v_1\rangle}v_1$；然后 $v_3 = u_3 - \frac{\langle u_3, v_1\rangle}{\langle v_1, v_1\rangle}v_1 - \frac{\langle u_3, v_2\rangle}{\langle v_2, v_2\rangle}v_2$。

Theorem

Gram-Schmidt 正交化： 设 $\{u_1, u_2, \cdots, u_n\}$ 是内积空间 V 的一个基，构造正交基 $\{v_1, v_2, \cdots, v_n\}$ 步骤：

- 令 $v_1 = u_1$；
- 对 $k = 2, \cdots, n$，令 $v_k = u_k - \sum_{i=1}^{k-1}\frac{\langle u_k, v_i\rangle}{\langle v_i, v_i\rangle}v_i$。

若再单位化 $e_i = \frac{v_i}{\|v_i\|}$，则得标准正交基 $\{e_1, e_2, \cdots, e_n\}$。

接下来，我们通过一个例子来更好地理解上述过程：

在 $\mathbb{R}^3$ 中，对基 $\{u_1 = (1, 1, 0), u_2 = (0, 1, 1), u_3 = (1, 0, 1)\}$ 进行 ***Gram - Schmidt*** 正交化：

$v_1 = u_1 = (1, 1, 0)$；$v_2 = u_2 - \frac{\langle u_2, v_1\rangle}{\langle v_1, v_1\rangle}v_1 = (0, 1, 1) - \frac{1}{2}(1, 1, 0) = (-\frac{1}{2}, \frac{1}{2}, 1)$；$v_3 = u_3 - \frac{\langle u_3, v_1\rangle}{\langle v_1, v_1\rangle}v_1 - \frac{\langle u_3, v_2\rangle}{\langle v_2, v_2\rangle}v_2 = (1, 0, 1) - \frac{1}{2}(1, 1, 0) - \frac{\frac{1}{2}}{\frac{3}{2}}(-\frac{1}{2}, \frac{1}{2}, 1) = (\frac{2}{3}, -\frac{2}{3}, \frac{2}{3})$。

单位化：$\|v_1\| = \sqrt{2}$，$e_1 = \frac{v_1}{\|v_1\|} = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0)$；$\|v_2\| = \sqrt{\frac{3}{2}}$，$e_2 = \frac{v_2}{\|v_2\|} = (-\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{3})$；$\|v_3\| = \sqrt{\frac{4}{3}}$，$e_3 = \frac{v_3}{\|v_3\|} = (\frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3})$。

得到标准正交基 $\{e_1, e_2, e_3\}$。

7.3 正交补与正交投影

7.3.1 正交补的定义与实例

Definition

正交补（Orthogonal complement）：设 S 为内积空间 V 的非空子集，定义 $S^\perp$ 为 V 中与 S 内每个向量都正交的所有向量的集合，即

$$S^\perp = \{\forall x \in V : \langle x, y\rangle = 0\ \forall y \in S\}.$$

$S^\perp$ 称为 S 的**正交补**。

Example

对于任何内积空间 V，有 $\{0\}^\perp = V$ 且 $V^\perp = \{0\}$。这是因为零向量与 V 中任意向量的内积为 0，而只有零向量与 V 中所有向量正交。
若 $V = \mathbb{R}^3$ 且 $S = \{e_3 = (0,0,1)\}$，则 $S^\perp$ 由所有满足 $\langle (x,y,z),(0,0,1)\rangle = z = 0$ 的向量组成，即 xy-平面。

7.3.2 正交分解定理与最佳逼近推论

Theorem

正交分解定理（Orthogonal Decomposition Theorem）：设 W 是内积空间 V 的有限维子空间，$y \in V$，则存在唯一的向量 $u \in W$ 和 $z \in W^\perp$，使得 $y = u + z$。进一步地，若 $\{v_1, v_2, \ldots, v_k\}$ 是 W 的标准正交基，则

$$u = \sum_{i=1}^{k} \langle y, v_i\rangle v_i.$$

这类似于我们在中学阶段中物理学研习的矢量的正交分解。

证明. 设 $\{v_1, v_2, \ldots, v_k\}$ 为 W 的标准正交基，按上述公式定义 u，令 $z = y - u$。显然 $u \in W$ 且 $y = u + z$。

要证 $z \in W^\perp$，对任意 v_j，计算

$$\begin{aligned}\langle z, v_j\rangle &= \langle y - \sum_{i=1}^{k} \langle y, v_i\rangle v_i, v_j\rangle \\ &= \langle y, v_j\rangle - \sum_{i=1}^{k} \langle y, v_i\rangle \langle v_i, v_j\rangle.\end{aligned}$$

由于 $\{v_1, \ldots, v_k\}$ 是标准正交基，$\langle v_i, v_j\rangle = \delta_{ij}$，故

$$\langle z, v_j\rangle = \langle y, v_j\rangle - \langle y, v_j\rangle = 0,$$

即 $z \in W^\perp$。

唯一性：若 $y = u' + z'$（$u' \in W$，$z' \in W^\perp$），则 $u - u' = z' - z \in W \cap W^\perp$。而 $W \cap W^\perp = \{0\}$（若 $w \in W \cap W^\perp$，则 $\langle w, w\rangle = 0$，由内积正定性得 $w = 0$），因此 $u = u'$，$z = z'$。 □

Corollary & Secondary Conclusion

最佳逼近推论：在上述定理的记号下，u 是 W 中**唯一"最接近"** y 的向量，即对任意 $x \in W$，有

$$\|y-x\| \geq \|y-u\|,$$

且等号成立当且仅当 $x=u$。

证明. $y=u+z$ （$z \in W^{\perp}$）。对任意 $x \in W$，$u-x \in W$，故 $u-x$ 与 z 正交。于是

$$\|y-x\|^2 = \|u+z-x\|^2 = \|(u-x)+z\|^2 = \|u-x\|^2 + \|z\|^2 \geq \|z\|^2 = \|y-u\|^2.$$

当 $\|y-x\| = \|y-u\|$ 时，$\|u-x\|^2 = 0$，即 $u-x=0$，所以 $x=u$；反之，若 $x=u$，显然 $\|y-x\| = \|y-u\|$。 □

上述推论中的向量 u 称为 y 在 W 上的**正交投影**，记为 $\mathrm{proj}_W(y)$。

从几何直观上看，在欧氏空间中，正交投影就是向量向子空间作垂直投影，使得投影误差向量 $y-u$ 与子空间 W 正交。这一性质保证了投影的"最佳逼近"特性，即正交投影是子空间中距离原向量最近的点。

在一般内积空间中，通过标准正交基 $\{v_1, \ldots, v_k\}$ 计算正交投影的公式

$$\mathrm{proj}_W(y) = \sum_{i=1}^{k} \langle y, v_i \rangle v_i$$

不仅明确了投影的具体形式，还揭示了内积空间中向量分解的本质。

例如在最小二乘法中，我们希望在由一组基向量张成的子空间 W 中找到对观测向量 y 的最佳拟合，本质上就是寻找 y 在 W 上的正交投影 $\mathrm{proj}_W(y)$，使得误差 $\|y - \mathrm{proj}_W(y)\|$ 最小。这正是正交投影最小距离性质的重要应用，它将抽象的内积空间理论与实际问题紧密结合，成为解决许多数学和工程问题的关键工具。

接下来，我们对本节的主定理做一个介绍：

Theorem

设 V 是一个内积空间，S 和 S° 是 V 的子集，W 是 V 的有限维子空间。我们有以下结论。

(a) $S^{\circ} \subseteq S$ 意味着 $S^{\perp} \subseteq (S^{\circ})^{\perp}$。

(b) $S \subseteq (S^{\perp})^{\perp}$；因此 $\mathrm{span}(S) \subseteq (S^{\perp})^{\perp}$。

(c) $W = (W^{\perp})^{\perp}$。

(d) $V = W \oplus W^{\perp}$。

Theorem

贝塞尔不等式（Bessel's inequality）：设 V 是一个内积空间，且设 $S = \{v_1, v_2, \ldots, v_n\}$ 是 V 的一个标准正交子集。对于任意 $x \in V$，我们有 $\|x\|^2 \geq \sum_{i=1}^{n} |\langle x, v_i\rangle|^2$。

即：把向量 x 向标准正交子集 S 中每个向量投影，所有投影长度的平方和，不会超过 x 自身的长度平方。

Theorem

正交补维数定理（Orthogonal Complement Dimension Theorem）：设 $S = \{v_1, v_2, \ldots, v_k\}$ 是 n-维内积空间 V 中的标准正交集，则：

1. S 可扩充为 V 的标准正交基 $\{v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n\}$；
2. 若 $W = \mathrm{span}(S)$，则 $S_1 = \{v_{k+1}, v_{k+2}, \ldots, v_n\}$ 是 $W^\perp$ 的标准正交基；
3. 若 W 是 V 的任意子空间，则 $\dim(V) = \dim(W) + \dim(W^\perp)$。

证明. 1. 由基的扩充定理，先将 S 扩充为 V 的一组基，再通过 Gram-Schmidt 正交化过程得到标准正交基。

2. 对任意 $x \in W^\perp$, x 与 S 中向量正交, 即 $\langle x, v_i\rangle = 0 (1 \leq i \leq k)$。而 $x = \sum_{i=1}^{n} \langle x, v_i\rangle v_i$, 故
$$x = \sum_{i=k+1}^{n} \langle x, v_i\rangle v_i \in \mathrm{span}(S_1),$$
说明 S_1 张成 $W^\perp$。又 S_1 是标准正交集，所以是 $W^\perp$ 的标准正交基。

3. 设 W 是 V 的子空间，V 有限维，故 W 有标准正交基 $\{v_1, v_2, \ldots, v_k\}$。由 (1) 和 (2)，$\dim(V) = n = k + (n-k) = \dim(W) + \dim(W^\perp)$。

□

Example

在 $\mathbb{F}^3$ 中，$W = \mathrm{span}(\{e_1, e_2\})$。对于 $x = (a, b, c) \in W^\perp$，需满足
$$\langle x, e_1\rangle = a = 0 \quad \text{且} \quad \langle x, e_2\rangle = b = 0,$$
即 $x = (0, 0, c)$，故 $W^\perp = \mathrm{span}(\{e_3\})$。由维数公式，$\dim(W) = 2$，$\dim(V) = 3$，则 $\dim(W^\perp) = 3 - 2 = 1$，与实际结果一致。

课后习题：

Exercise

1. 正交化下列向量组：

(a) V = R^3，S = (1, 1), (0, 1, 1), (0, 0, 1)，且 r = (1, 0, 1)

(b) V = P (R)，内积为 $(f(x), g(x)) = \int_0^1 f(t)g(t)dt$，S = 1, x, x^2，且 h(x) = 1 + x

Exercise

2. 设 $S = \{(1,0,i),(1,2,1)\}$ 是 $\mathbb{C}^3$ 中的集合。计算 $S^{\perp}$。

7.4 伴随算子与伴随矩阵

在介绍伴随之间，我们先来解释一下“伴随矩阵”（**注意，这和求逆矩阵的伴随矩阵没有直接关系**）

7.4.1 矩阵的共轭转置

Definition

定义矩阵 A 的**共轭转置**为 $A^* = \overline{A^T}$（或者记为：$A^* = \overline{A}'$，即先转置再取复共轭），其中经过共轭转置得到的矩阵我们称之为**伴随矩阵 (adjoint matrix)**。

设 V 是 n 维酉空间，$\{e_1, e_2, \ldots, e_n\}$ 为其标准正交基，并且标准正交基下，酉空间内积满足 $(\alpha, \beta) = \overline{x}'y$

若线性变换 φ 在该基下的表示矩阵为 A，向量 $\alpha = a_1e_1 + \cdots + a_ne_n$，$\beta = b_1e_1 + \cdots + b_ne_n$ 的坐标向量分别为 $x = (a_1, \ldots, a_n)'$，$y = (b_1, \ldots, b_n)'$，

则：$(\varphi(\alpha), \beta) = \overline{(Ax)}'y = \overline{x}'\overline{A}'y = \overline{x}'\overline{(A)}'y$；

这建立了内积运算与矩阵乘法的直接联系，为伴随概念的引入奠定了运算基础。

在上述内积表达式中，$\overline{A}'y$ 对应一个新的线性变换 ψ，其坐标运算体现了共轭转置矩阵的作用。通过定义 $\psi(\beta)$ 的坐标为 $\overline{A}'y$，初步展示了共轭转置矩阵在构造新线性变换中的角色，这是伴随矩阵的核心定义形式。

Claim

设 V 为有限维内积空间，φ 是 V 上的线性变换，则存在**唯一的线性变换** φ^*（称为 φ 的伴随算子），使得对任意 $\alpha, \beta \in V$，成立 $(\varphi(\alpha), \beta) = (\alpha, \varphi^*(\beta))$。

证明. **存在性证明**: 可通过构造坐标映射结合共轭转置矩阵实现, 与矩阵运算 $(\varphi(\alpha),\beta) = x'(\overline{A'}y)$ 呼应，定义 φ^* 在基下的作用由 $\overline{A'}$ 确定。

唯一性证明: 若存在 φ^2 也满足 $(\varphi(\alpha),\beta) = (\alpha,\varphi^2(\beta))$，则对任意 α，$(\alpha,\varphi^2(\beta) - \varphi^*(\beta)) = 0$。取 $\alpha = \varphi^2(\beta) - \varphi^*(\beta)$，由内积定义知 $\varphi^2(\beta) - \varphi^*(\beta) = 0$，故 $\varphi^2 = \varphi^*$，唯一性得证。

所以，对于任意的线性算子，都有唯一的伴随。 □

伴随算子的矩阵表示

Fact or Background

设 V 为有限维内积空间，β 为 V 的标准正交基，T 为 V 上的线性算子，则 $[T^*]_\beta = [T]_\beta^*$。即伴随算子 T^* 在 β 下的矩阵是 T 在 β 下矩阵的共轭转置。

证明. 设 $A = [T]_\beta$，$B = [T^*]_\beta$，$\beta = \{v_1,\ldots,v_n\}$，$B_{ij} = \langle T^*(v_j),v_i\rangle = \overline{\langle T(v_i),v_j\rangle} = \overline{A_{ji}} = (A^*)_{ij}$，故 $B = A^*$。 □

此 Fact 明确了伴随算子与矩阵共轭转置的直接对应，是连接算子与矩阵的关键桥梁。

7.4.2 伴随算子与伴随矩阵的性质

加法性质： $(T+U)^* = T^* + U^*$（T,U 为线性算子）；$(A+B)^* = A^* + B^*$（A,B 为矩阵）。

数乘性质： $(cT)^* = \bar{c}T^*$（$c \in \mathbb{C}$，T 为算子）；$(cA)^* = \bar{c}A^*$（$c \in \mathbb{C}$，A 为矩阵）。

乘法（复合）性质： $(TU)^* = U^*T^*$（T,U 为算子）；$(AB)^* = B^*A^*$（A,B 为矩阵）。

双重伴随性质： $T^{**} = T$（算子）；$A^{**} = A$（矩阵）。

单位元性质： $I^* = I$（I 为单位算子或单位矩阵）。

这些性质由伴随的定义与矩阵共轭转置的运算规则导出。

例如乘法性质 $(TU)^* = U^*T^*$，对任意 x,y，$\langle (TU)(x),y\rangle = \langle x,(TU)^*(y)\rangle$，同时 $\langle TU(x),y\rangle = \langle U(x),T^*(y)\rangle = \langle x,U^*(T^*(y))\rangle$，由伴随唯一性得 $(TU)^* = U^*T^*$。矩阵情形同理，通过坐标表示与算子性质对应可得。

综上，伴随算子与伴随矩阵通过**标准正交基**建立紧密联系，其性质在有限维内积空间的线性变换分析中具有重要地位，是深入研究算子结构与矩阵理论的重要工具。

7.4.3 本节主定理

Theorem

设 V 是一个内积空间，T 是 V 上的线性算子。则有以下结论：

(a) $R(T^*)^\perp = N(T)$。

(b) 若 V 是有限维的，则 $R(T^*) = N(T)^\perp$。

证明. (a) 证明 $R(T^*)^\perp = N(T)$（分两步证包含关系）：

$R(T^*)^\perp \subseteq N(T)$: 任取 $x \in R(T^*)^\perp$，则对任意 $z \in V$，$(x, T^*z) = 0$。由伴随算子的定义 $(x, T^*z) = (Tx, z)$，故对任意 $z \in V$，$(Tx, z) = 0$。根据内积的非退化性，得 $Tx = 0$，即 $x \in N(T)$。

$N(T) \subseteq R(T^*)^\perp$: 任取 $x \in N(T)$，则 $Tx = 0$。对任意 $z \in V$，由伴随算子定义：$(x, T^*z) = (Tx, z) = (0, z) = 0$，故 $x \perp R(T^*)$，即 $x \in R(T^*)^\perp$。

综上，$R(T^*)^\perp = N(T)$。

(b) 证明有限维时 $R(T^*) = N(T)^\perp$：

有限维内积空间中，子空间满足“正交补的正交补等于自身”，即对任意子空间 $W \subseteq V$，$(W^\perp)^\perp = W$。

由 (a) 得 $R(T^*)^\perp = N(T)$，两边取正交补：

$$(R(T^*)^\perp)^\perp = N(T)^\perp$$

左边即为 $R(T^*)$，故 $R(T^*) = N(T)^\perp$。 □

Theorem

如果 $V = W \oplus W^\perp$ 且 T 是沿着 $W^\perp$ 到 W 的投影（故：$N(T) = W^\perp$），那么 $T = T^*$。

证明. 已知 $V = W \oplus W^\perp$，T 是沿 $W^\perp$ 到 W 的投影，则对任意 $\alpha \in V$，存在唯一分解：

$$\alpha = \alpha_1 + \alpha_2, \quad \alpha_1 \in W,\ \alpha_2 \in W^\perp,$$

且 $T\alpha = \alpha_1$。对任意 $\beta \in V$，同理分解 $\beta = \beta_1 + \beta_2$（$\beta_1 \in W$, $\beta_2 \in W^\perp$）。

计算内积：

$$(T\alpha, \beta) = (\alpha_1, \beta_1 + \beta_2) = (\alpha_1, \beta_1) + (\alpha_1, \beta_2),$$

由 $W \perp W^\perp$，$(\alpha_1, \beta_2) = 0$，故 $(T\alpha, \beta) = (\alpha_1, \beta_1)$。

另一方面：

$$(\alpha, T\beta) = (\alpha_1 + \alpha_2, \beta_1) = (\alpha_1, \beta_1) + (\alpha_2, \beta_1),$$

同理 $(\alpha_2,\beta_1)=0$，故 $(\alpha,T\beta)=(\alpha_1,\beta_1)$。

因此对任意 $\alpha,\beta\in V$，$(T\alpha,\beta)=(\alpha,T\beta)$，由伴随算子的定义得 $T=T^*$。

□

当然，在之后的内容中，我们也可以定义“广义的线性算子”:

我们只需要抽取出伴随算子的定义核心: **内积对偶**:

所以，可以有 $\forall v\in V, w\in W$，有 $\langle Tv,w\rangle_W=\langle v,T^*w\rangle_V$。

我们发现: 左边 $Tv\in W$ 与 $w\in W$ 做内积，右边 $v\in V$ 必须与 V 中的元素 做内积——因此 T^*w 必须属于 V，即 T^* 的输出在 V，输入 w 在 W，故 $T^*:W\to V$。

7.4.4 课后习题

Exercise

设 T 是有限维向量空间 V 上的线性算子。证明以下结论:

(a) $N(T^*T)=N(T)$。由此推出 $\operatorname{rank}(T^*T)=\operatorname{rank}(T)$。

(b) $\operatorname{rank}(T)=\operatorname{rank}(T^*)$。由 (a) 推出 $\operatorname{rank}(TT^*)=\operatorname{rank}(T)$。

(c) 对于任意 $n\times n$ 矩阵 A，$\operatorname{rank}(A^*A)=\operatorname{rank}(AA^*)=\operatorname{rank}(A)$。

7.5 内积空间的同构 · 正交算子与酉算子

7.5.1 内积空间同构

在研究线性空间时，同构的概念帮助我们识别具有相同代数结构的空间。对于内积空间，我们不仅关注线性结构，还需保持内积运算的一致性。由此引出以下核心定义：

Definition

设 V, W 是内积空间，$\sigma : V \to W$ 是线性映射：保内积映射：对任意 $\alpha, \beta \in V$，都有 $\langle \sigma(\alpha), \sigma(\beta) \rangle = \langle \alpha, \beta \rangle$（内积完全“不变”）。

Property

设 $\sigma : V \to W$ 是线性映射, 则以下命题等价:

1. σ 是保内积映射;
2. $\|\sigma(\alpha)\| = \|\alpha\|$ 对所有 $\alpha \in V$ 成立;
3. 若 $\{\alpha_1, \dots, \alpha_n\}$ 是 V 的标准正交基, 则 $\{\sigma(\alpha_1), \dots, \sigma(\alpha_n)\}$ 是 W 的标准正交基.

概要. (1) ⇒ (2): 取 $\alpha = \beta$ 即得. (2) ⇒ (3): 利用内积与范数的关系验证基向量的正交性. (3) ⇒ (1): 任一向量可由标准正交基线性表出，结合线性性和内积运算验证. □

证明. **(1)** ⇒ **(2)**：由范数与内积的关系 $\|\gamma\| = \sqrt{\langle \gamma, \gamma \rangle}$，对任意 $\alpha \in V$，取 $\beta = \alpha$，则

$$\|\sigma(\alpha)\| = \sqrt{\langle \sigma(\alpha), \sigma(\alpha) \rangle}\text{因为保内积}\sqrt{\langle \alpha, \alpha \rangle} = \|\alpha\|.$$

(2) ⇒ **(3)**：设 $\{\alpha_1, \dots, \alpha_n\}$ 是 V 的标准正交基，即 $\langle \alpha_i, \alpha_j \rangle = \delta_{ij}$（$\delta_{ij}$ 为克罗内克符号）。

- **标准性：** 由 (2)，$\|\sigma(\alpha_i)\| = \|\alpha_i\| = 1$；

- 正交性:对 $i \neq j$,利用内积的极化恒等式(实内积:$\langle u, v \rangle = \frac{1}{2}\left(\|u+v\|^2 - \|u\|^2 - \|v\|^2\right)$),结合 σ 的线性性 $\sigma(\alpha_i + \alpha_j) = \sigma(\alpha_i) + \sigma(\alpha_j)$，得

$$\langle \sigma(\alpha_i), \sigma(\alpha_j) \rangle = \frac{1}{2}\left(\|\sigma(\alpha_i + \alpha_j)\|^2 - \|\sigma(\alpha_i)\|^2 - \|\sigma(\alpha_j)\|^2\right).$$

由 (2)，$\|\sigma(\alpha_i + \alpha_j)\| = \|\alpha_i + \alpha_j\| = \sqrt{\|\alpha_i\|^2 + \|\alpha_j\|^2} = \sqrt{2}$，代入得

$$\langle \sigma(\alpha_i), \sigma(\alpha_j) \rangle = \frac{1}{2}(2 - 1 - 1) = 0.$$

故 $\{\sigma(\alpha_1), \dots, \sigma(\alpha_n)\}$ 是 W 的标准正交基。

(3) ⇒ **(1)**：对任意 $\alpha, \beta \in V$，由标准正交基的线性表示性，设

$$\alpha = \sum_{i=1}^{n} k_i \alpha_i, \quad \beta = \sum_{j=1}^{n} l_j \alpha_j,$$

则内积 $\langle\alpha,\beta\rangle=\sum_{i=1}^{n}k_i\overline{l_i}$（复内积）。由 σ 的线性性，$\sigma(\alpha)=\sum_{i=1}^{n}k_i\sigma(\alpha_i)$，$\sigma(\beta)=\sum_{j=1}^{n}l_j\sigma(\alpha_j)$。因 $\{\sigma(\alpha_1),\ldots,\sigma(\alpha_n)\}$ 是 W 的标准正交基，故

$$\langle\sigma(\alpha),\sigma(\beta)\rangle=\sum_{i=1}^{n}k_i\overline{l_i}=\langle\alpha,\beta\rangle,$$

即 σ 是保内积映射。

综上，三个命题等价。 $\square$

Definition

设 V 和 W 是数域 $\mathbb{F}$ 上的内积空间. 若存在**双射的线性映射** $\sigma:V\to W$ 满足

$$(\sigma(\alpha),\sigma(\beta))_W=(\alpha,\beta)_V\quad\forall\alpha,\beta\in V,$$

则称 σ 为内积空间的**同构映射**, 此时称 V 与 W **同构**, 记作 $V\cong W$.

【Remark】:

该定义要求 σ 不仅是有一般同构映射的“双射”性质，还需保持内积结构（即保内积映射）. **这意味着 σ 将 V 中的标准正交基映射为 W 中的标准正交基，从而保持向量的长度、夹角等几何性质**。

7.5.2 正交算子的定义与基本性质

Definition

正交算子（Orthogonal operator）： 设 V 是实内积空间, $T\in\mathcal{L}(V)$ 是线性算子. 若 T 满足

$$(T(\alpha),T(\beta))=(\alpha,\beta)\quad\forall\alpha,\beta\in V,$$

则称 T 为 V 上的**正交算子**.

【Remark】:

正交算子可视为内积空间到自身的同构映射，其保持内积结构不变；
从几何角度看，正交算子对应刚体运动（旋转、反射等）.

Property

正交算子的等价刻画： 设 $T \in \mathcal{L}(V)$ 是线性算子, $\dim V = n$, 则以下命题等价:

1. T 是正交算子;
2. T 是双射且 $T^{-1} = T^*$;
3. $TT^* = T^*T = I$;
4. 若 $\{\alpha_1, \ldots, \alpha_n\}$ 是标准正交基, 则 $\{T(\alpha_1), \ldots, T(\alpha_n)\}$ 也是标准正交基;
5. T 在任意标准正交基下的矩阵表示是正交矩阵.

证明. 1. $1 \Rightarrow 2$：正交算子　双射且 $T^{-1} = T^*$

- 双射性： 若 $T\alpha = 0$，则 $\langle T\alpha, T\alpha \rangle = \langle \alpha, \alpha \rangle = 0$，故 $\alpha = 0$（内积非退化），即 T 是单射；有限维空间中，线性**算子**“单射就可以导出满射”，故 T 是双射，存在逆算子 T^{-1}。

- $T^{-1} = T^*$： 由伴随算子定义 $\langle T^*\alpha, \beta \rangle = \langle \alpha, T\beta \rangle$，结合正交算子的保内积性 $\langle T\alpha, T\beta \rangle = \langle \alpha, \beta \rangle$，得：

$$\langle T^*T\alpha, \beta \rangle = \langle \alpha, \beta \rangle \implies T^*T = I$$

同理 $TT^* = I$，故 $T^* = T^{-1}$。

2. $2 \Rightarrow 3$：双射且 $T^{-1} = T^*$　$TT^* = T^*T = I$ 由逆算子定义，若 $T^{-1} = T^*$，则：

$$TT^* = TT^{-1} = I, \quad T^*T = T^{-1}T = I$$

3. $3 \Rightarrow 4$：$TT^* = T^*T = I$　标准正交基映为标准正交基

设 $\{\alpha_1, \ldots, \alpha_n\}$ 是 V 的标准正交基（即 $\langle \alpha_i, \alpha_j \rangle = \delta_{ij}$，$\delta_{ij}$ 为克罗内克符号）。利用伴随算子性质 $\langle T\alpha_i, T\alpha_j \rangle = \langle T^*T\alpha_i, \alpha_j \rangle$，结合 $T^*T = I$，得：

$$\langle T\alpha_i, T\alpha_j \rangle = \langle I\alpha_i, \alpha_j \rangle = \langle \alpha_i, \alpha_j \rangle = \delta_{ij}$$

故 $\{T\alpha_1, \ldots, T\alpha_n\}$ 是 W 的标准正交基。

4. $4 \Rightarrow 5$：保标准正交基　矩阵是正交矩阵

设 T 在标准正交基 $\{\alpha_1, \ldots, \alpha_n\}$ 下的矩阵为 $A = (a_{ij})$（即 $T\alpha_j = \sum_{k=1}^n a_{kj}\alpha_k$）。由 $\{T\alpha_1, \ldots, T\alpha_n\}$ 是标准正交基，得：

$$\langle T\alpha_i, T\alpha_j \rangle = \sum_{k=1}^n a_{ki}a_{kj} = \delta_{ij}$$

这等价于 $A^TA = I$（正交矩阵定义，见下半部分）。

对任意标准正交基，设其与原基的过渡矩阵为正交矩阵 P，则 T 在新基下的矩阵为 $P^{-1}AP$；正交矩阵的乘积/逆仍为正交矩阵，故 $P^{-1}AP$ 是正交矩阵。

5. 5 $\Rightarrow$ 1：矩阵是正交矩阵　正交算子

设 T 在标准正交基下的矩阵 A 满足 $A^TA=I$。对任意 $\alpha,\beta\in V$，设其坐标为 x,y，则 $T\alpha,T\beta$ 的坐标为 Ax,Ay。由标准正交基下内积与坐标的关系：

$$\langle T\alpha,T\beta\rangle=(Ax)^T(Ay)=x^TA^TAy=x^Ty=\langle\alpha,\beta\rangle$$

故 T 是保内积的正交算子。

综上，5 个命题两两等价。 □

类似的，我们将考察**正交算子的矩阵表示**：正交矩阵是正交算子在**标准正交基**下的矩阵表示，具体对应关系如下：

从算子到矩阵： 若 T 是 $\mathbb{R}^n$ 上的正交算子，且取 $\mathbb{R}^n$ 的标准正交基（如自然基 $\{e_1,e_2,\dots,e_n\}$），则 T 在该基下的矩阵 A 是正交矩阵。

从矩阵到算子： 若 A 是正交矩阵，则 A 诱导的线性变换 $T(\alpha)=A\alpha$ 是 $\mathbb{R}^n$ 上的正交算子。

Theorem

正交矩阵（Orthogonal matrix）： 设 A 为 n 阶实矩阵，若满足 $A^TA=AA^T=I_n$ 则我们称 A 为正交矩阵:

Property

1. A 是正交矩阵**当且仅当**其列向量构成 $\mathbb{R}^n$ 的标准正交基;
2. A 是正交矩阵**当且仅当**其行向量构成 $\mathbb{R}^n$ 的标准正交基;
3. 正交矩阵的**行列式**为 ±1.

证明. 1. **证明：** A 是正交矩阵 $\iff$ 其列向量构成 $\mathbb{R}^n$ 的标准正交基。

设 $A=(\alpha_1,\alpha_2,\dots,\alpha_n)$，其中 $\alpha_i\in\mathbb{R}^n$ 是 A 的第 i 列向量，则：

$$A^TA=\begin{pmatrix}\alpha_1^T\\ \alpha_2^T\\ \vdots\\ \alpha_n^T\end{pmatrix}(\alpha_1,\alpha_2,\dots,\alpha_n)=\begin{pmatrix}\alpha_1^T\alpha_1 & \alpha_1^T\alpha_2 & \dots & \alpha_1^T\alpha_n\\ \alpha_2^T\alpha_1 & \alpha_2^T\alpha_2 & \dots & \alpha_2^T\alpha_n\\ \vdots & \vdots & \ddots & \vdots\\ \alpha_n^T\alpha_1 & \alpha_n^T\alpha_2 & \dots & \alpha_n^T\alpha_n\end{pmatrix}.$$

由正交矩阵定义，$A^TA=I_n \iff \alpha_i^T\alpha_j=\delta_{ij}$（克罗内克函数：$i=j$ 时为 1，$i\neq j$ 时为 0），即列向量两两正交且长度为 1（标准正交）。又 $\dim\mathbb{R}^n=n$，故 $\{\alpha_1,\dots,\alpha_n\}$ 构成 $\mathbb{R}^n$ 的标准正交基。

2. **证明：** A 是正交矩阵 $\iff$ 其行向量构成 $\mathbb{R}^n$ 的标准正交基。

正交矩阵满足 $A^T = A^{-1}$，故 $AA^T = I_n$。设 A 的第 i 行向量为 β_i^T（$\beta_i \in \mathbb{R}^n$），则：

$$AA^T = \begin{pmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_n^T \end{pmatrix} \begin{pmatrix} \beta_1 & \beta_2 & \dots & \beta_n \end{pmatrix} = I_n.$$

同理，$AA^T = I_n \iff \beta_i^T\beta_j = \delta_{ij}$，即行向量标准正交，且个数为 n，故构成 $\mathbb{R}^n$ 的标准正交基。

3. **证明：** 正交矩阵的行列式为 ± 1。

对 $A^TA = I_n$ 两边取行列式，由行列式乘法法则：

$$\det(A^TA) = \det(A^T) \cdot \det(A) = \det(I_n) = 1.$$

又 $\det(A^T) = \det(A)$，故 $\det(A)^2 = 1$，解得 $\det(A) = \pm 1$。

□

我们可以看到**正交算子与正交矩阵的核心性质的统一性：** 保内积与保范数：

正交算子 T 满足 $\langle T(\alpha), T(\beta)\rangle = \langle \alpha, \beta\rangle$;

特别地，若 $\alpha = \beta$, 则有保向量的模长（即 $\|T(\alpha)\| = \|\alpha\|$）。

正交矩阵 A 对应的线性变换同样满足 $\|A\alpha\| = \|\alpha\|$，因为 $\|A\alpha\|^2 = (A\alpha)^T(A\alpha) = \alpha^TA^TA\alpha = \alpha^T\alpha = \|\alpha\|^2$.

接下来，我们考察一下特征值的模（或绝对值）：

Property

正交算子和正交矩阵的特征值（若为实数）绝对值为 1；若为复数，则模长为 1（在实矩阵情形下，复数特征值成对共轭出现）。

证明. **1. 实数特征值的绝对值为 1**

设 A 为正交矩阵，$\lambda \in \mathbb{R}$ 是 A 的特征值，$\boldsymbol{x} \in \mathbb{R}^n$ 是对应非零特征向量，则 $A\boldsymbol{x} = \lambda\boldsymbol{x}$。对等式两边取转置得 $\boldsymbol{x}^TA^T = \lambda\boldsymbol{x}^T$，将两式相乘：

$$\boldsymbol{x}^TA^TA\boldsymbol{x} = \lambda^2\boldsymbol{x}^T\boldsymbol{x}.$$

由正交矩阵定义 $A^TA = I_n$，代入得 $\boldsymbol{x}^T\boldsymbol{x} = \lambda^2\boldsymbol{x}^T\boldsymbol{x}$。因 $\boldsymbol{x} \neq \boldsymbol{0}$，故 $\boldsymbol{x}^T\boldsymbol{x} = \|\boldsymbol{x}\|^2 > 0$，两边消去得 $\lambda^2 = 1$，即 $|\lambda| = 1$。

2. 复数特征值的模长为 1

将实正交矩阵 A 视为复矩阵（$\mathbb{C}^{n\times n}$ 中元素），设 $\lambda \in \mathbb{C}$ 是特征值，$\boldsymbol{x} \in \mathbb{C}^n$ 是对应非零特征向量，则 $A\boldsymbol{x} = \lambda\boldsymbol{x}$。对等式取共轭转置（记 $\overline{\lambda}$ 为 λ 的共轭，$\boldsymbol{x}^\dagger = \overline{\boldsymbol{x}}^T$）：

$$\boldsymbol{x}^\dagger A^T = \overline{\lambda}\boldsymbol{x}^\dagger.$$

结合 $A^T = A^{-1}$（正交矩阵性质），两边右乘 $A\boldsymbol{x}$：

$$\boldsymbol{x}^\dagger A^T A\boldsymbol{x} = \overline{\lambda}\boldsymbol{x}^\dagger(\lambda\boldsymbol{x}) \implies \boldsymbol{x}^\dagger\boldsymbol{x} = |\lambda|^2\boldsymbol{x}^\dagger\boldsymbol{x}.$$

因 $\boldsymbol{x} \neq \boldsymbol{0}$，故 $\boldsymbol{x}^\dagger\boldsymbol{x} = \|\boldsymbol{x}\|_{\mathbb{C}}^2 > 0$，消去得 $|\lambda|^2 = 1$，即 $|\lambda| = 1$。

3. 实矩阵的复数特征值成对共轭出现

设 A 为实矩阵，特征多项式 $f(\lambda) = \det(A - \lambda I_n)$ 是实系数多项式。若 $\lambda = a + bi$ $(b \neq 0)$ 是 $f(\lambda)$ 的根，则其共轭 $\overline{\lambda} = a - bi$ 也满足 $f(\overline{\lambda}) = \overline{f(\lambda)} = \overline{0} = 0$，即 $\overline{\lambda}$ 也是特征值。因此实正交矩阵的复数特征值必以共轭对形式存在（如 λ 与 $\overline{\lambda}$）。

□

接下来，我们来看两个正交算子的**应用**：**旋转与反射**。

旋转变换 R_θ

Definition

在二维欧氏空间 $\mathbb{R}^2$ 中，绕原点旋转 θ 角的变换，$R_\theta(x, y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$。

矩阵表示： 在标准有序（正交）基下的矩阵为

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

该矩阵是正交矩阵（行、列均构成 $\mathbb{R}^2$ 的标准正交基），且 $\det(R_\theta) = 1$。

特征性质： 特征多项式为 $t^2 - 2\cos\theta\, t + 1$。因此当 $\theta \neq k\pi$ 时，在 $\mathbb{R}$ 上不可对角化，在 $\mathbb{C}$ 上可对角化，特征值为 $e^{i\theta}$ 和 $e^{-i\theta}$。

反射变换 S_L（关于过原点直线 L）

Definition

若 $x \in L$，则 $S_L(x) = x$；若 $x \in L^\perp$，则 $S_L(x) = -x$。

矩阵推导： 设 α 是直线 L 与正 x-轴的夹角，取 $v_1 = (\cos\alpha, \sin\alpha) \in L$，$v_2 = (-\sin\alpha, \cos\alpha) \in L^\perp$，$\{v_1, v_2\}$ 为标准正交基。在该基下，反射矩阵为 $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$。

通过基变换(利用旋转矩阵 $Q = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ 及其逆 $Q^{-1} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$)，可得在标准基下的矩阵：

$$\begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix}$$

性质：
- 是正交算子，满足 $S_L^2 = I$;
- $\det(S_L) = -1$，在标准正交基下的矩阵是对称矩阵。

相似的，我们可以给出酉算子的定义与性质

Definition

酉算子（Unitary operator）： 设 V 是复内积空间, $T \in \mathcal{L}(V)$ 是线性算子. 若 T 满足

$$(T(\alpha), T(\beta)) = (\alpha, \beta) \quad \forall \alpha, \beta \in V,$$

则称 T 为 V 上的**酉算子**.

Property

酉算子的等价刻画： 设 $T \in \mathcal{L}(V)$ 是线性算子, 则以下命题等价:

1. T 是酉算子;
2. T 是双射且 $T^{-1} = T^*$;
3. $TT^* = T^*T = I$;
4. 若 $\{\alpha_1, \ldots, \alpha_n\}$ 是标准正交基, 则 $\{T(\alpha_1), \ldots, T(\alpha_n)\}$ 也是标准正交基.

类比于正交算子，我们也可以得到以下性质:

Property

设 U 是 n 阶复方阵, 则:

1. U 是酉矩阵**当且仅当**其列向量构成 $\mathbb{C}^n$ 的标准正交基;
2. U 是酉矩阵**当且仅当**其行向量构成 $\mathbb{C}^n$ 的标准正交基;
3. 酉矩阵的**行列式**的模为 1.

7.6 QR 分解

QR 分解是矩阵理论中的一个核心工具，其核心目标是将任意给定的矩阵分解为一**个正交矩阵与一个上三角矩阵的乘积**。具体而言:

Definition

QR 分解 (QR decomposition): 设 A 为 $m \times n$ 实矩阵（$m \geq n$），则存在 $m \times m$ 正交矩阵 Q 和 $m \times n$ 上三角矩阵 R，使得

$$A = QR$$

当处理复矩阵时，对应的分解称为酉三角分解，此时 Q 为酉矩阵。

若 A 是 $n \times n$ 可逆矩阵，则 Q 为 n 阶正交矩阵，R 为 n 阶可逆上三角矩阵。

那么我们不禁要问：QR 分解存在吗？唯一吗？

Theorem

(QR 分解)： 设 A 是 n 阶实（复）矩阵，则 A 可分解为 $A = QR$。其中 Q 是正交（酉）矩阵，R 是主对角线上元素均大于等于零的上三角阵。若 A 非异，则这种分解必唯一。

证明. 1. 对 A 进行列分块 $A = (u_1, u_2, \cdots, u_n)$，利用类似 Gram - Schmidt 正交化方法，通过数学归纳法构造两两正交的向量 $\{w_1, w_2, \cdots, w_n\}$，其中 w_k 为零向量或单位向量。

2. 定义 $v_k = u_k - \sum_{j=1}^{k-1}(u_k, w_j)w_j$(其中，$v_k$ 均互相正交)
若 $v_k = 0$，则 $w_k = 0$；若 $v_k \neq 0$，则 $w_k = \frac{v_k}{\|v_k\|}$。
由此可得 $u_k = \sum_{j=1}^{k-1}(u_k, w_j)w_j + \|v_k\|w_k$。

3. 进而 $A = (w_1, w_2, \cdots, w_n)R$,其中 R 是上三角阵,主对角线元素为 $\|v_1\|, \|v_2\|, \cdots, \|v_n\|$。若 $w_k = 0$，则 R 的第 k 行元素全为零。

4. 将非零的 w_i 扩张为 $\mathbb{R}^n$ 的标准正交基，构造正交矩阵 Q，最终得 $A = QR$。复矩阵情形证明类似。 □

QR 分解的核心意义在于**将矩阵的列空间结构显式化**，其中正交矩阵 Q 的列向量构成列空间的标准正交基，而上三角矩阵 R 则编码了原始列向量在该基下的**坐标**。

接下来，让我们通过一个例子来更好地理解：

Example

求矩阵 $A = \begin{pmatrix} 1 & 2 & 5 \\ 1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}$ 的 QR 分解。

解:

1. 计算 $v_1 = u_1 = (1, 1, 0)^T$，$w_1 = \frac{1}{\sqrt{2}}(1, 1, 0)^T$；
2. $v_2 = u_2 - \sqrt{2}w_1 = (1, -1, 1)^T$，$w_2 = \frac{1}{\sqrt{3}}(1, -1, 1)^T$；

3. $v_3 = u_3 - 3\sqrt{2}w_1 - 2\sqrt{3}w_2 = (0,0,0)^T$，$w_3 = (0,0,0)^T$。

4. 用 $\tilde{w}_3 = \frac{1}{\sqrt{6}}(-1,1,2)^T$ 代替 w_3，得：

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{pmatrix}, \quad R = \begin{pmatrix} \sqrt{2} & \sqrt{2} & 3\sqrt{2} \\ 0 & \sqrt{3} & 2\sqrt{3} \\ 0 & 0 & 0 \end{pmatrix}$$

其中，我们对被分解矩阵是否满秩还有不同的操作：

Example

满秩矩阵的 QR 分解： 考虑矩阵：

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

1. Gram-Schmidt 正交化过程：- 计算 β_1：

$$\beta_1 = \frac{\alpha_1}{\|\alpha_1\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

- 计算 γ_2：

$$\gamma_2 = \alpha_2 - (\alpha_2, \beta_1)\beta_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \\ -\frac{1}{2} \end{pmatrix}$$

- 计算 β_2：

$$\beta_2 = \frac{\gamma_2}{\|\gamma_2\|} = \frac{1}{\sqrt{\frac{3}{2}}} \begin{pmatrix} \frac{1}{2} \\ 1 \\ -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix}$$

2. 构造矩阵 Q 和 R：

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{pmatrix}, \quad R = \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{\frac{3}{2}} \end{pmatrix}$$

3. 验证分解结果：

$$QR = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{\frac{3}{2}} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = A$$

Example

列向量线性相关的 QR 分解考虑矩阵：

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 0 & 0 \end{pmatrix}$$

1. Gram-Schmidt 正交化过程：- 计算 β_1：

$$\beta_1 = \frac{\alpha_1}{\|\alpha_1\|} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{pmatrix}$$

- 计算 γ_2：

$$\gamma_2 = \alpha_2 - (\alpha_2, \beta_1)\beta_1 = \begin{pmatrix} 2 \\ 4 \\ 0 \end{pmatrix} - \frac{10}{\sqrt{5}} \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

由于 $\gamma_2 = 0$，说明列向量 α_1 和 α_2 线性相关。所以需要重新选择。

第二个列向量需要从列空间的正交补空间（与 β_1 正交的空间）中选一个单位向量，这里选了 $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$（它与 β_1 内积为 0，且是单位向量）。

2. 构造矩阵 Q 和 R：

$$Q = \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 \\ \frac{2}{\sqrt{5}} & 0 \\ 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} \sqrt{5} & 2\sqrt{5} \\ 0 & 0 \end{pmatrix}$$

3. 验证分解结果：

$$QR = \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 \\ \frac{2}{\sqrt{5}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{5} & 2\sqrt{5} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 0 & 0 \end{pmatrix} = A$$

【Remark】:

列空间的定义是"矩阵列向量的所有线性组合构成的子空间"，即 $\mathrm{Col}(A) = \mathrm{span}\{\alpha_1, \alpha_2\} = \mathrm{span}\{\alpha_1\}$（因为 $\alpha_2 = 2\alpha_1$）。

而 $A = QR$ 的展开式为：$A = \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 \\ \frac{2}{\sqrt{5}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{5} & 2\sqrt{5} \\ 0 & 0 \end{pmatrix} = \sqrt{5} \cdot \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ 可以看到：A 的列向量仍是 Q 第一列的线性组合（第二列的系数为 0），因此 A 的列空间还是 $\mathrm{span}\{\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}\}$，与原来完全一致。

补充的"0 0 1"在 A 的列组合中系数为 0，不会影响列空间。

QR 分解的意义

QR 分解在理论和应用中均具有重要价值：

理论价值 1. 矩阵结构分析：QR 分解揭示了矩阵列空间的标准正交基结构，正交矩阵 Q 的列向量构成列空间的标准正交基，而上三角矩阵 R 则反映了原始列向量在该基下的坐标。2. 线性方程组理论：对于线性方程组 $Ax = b$，通过 QR 分解可转化为 $QRx = b$，即 $Rx = Q^T b$。由于 Q 是正交矩阵，保持了向量的长度和夹角，从而简化了方程组的求解过程。

实际应用 1. 最小二乘问题：对于超定方程组 $Ax = b$，QR 分解提供了求解最小二乘解的有效方法。通过将 A 分解为 QR，最小二乘问题转化为求解 $Rx = Q^T b$，计算复杂度显著降低。2. 特征值计算：QR 算法是计算矩阵特征值的重要方法，其基本思想是通过反复进行 QR 分解和矩阵乘法来逼近矩阵的特征值。该算法具有收敛快、稳定性好等优点。3. 数值计算稳定性：在数值计算中，正交变换具有良好的数值稳定性。通过 QR 分解，许多数值问题可以转化为更稳定的形式进行求解。

综上所述，QR 分解是矩阵理论中不可或缺的工具，其不仅深化了我们对矩阵结构的理解，更为众多实际问题提供了高效、稳定的解决方案。

7.7 内积空间的对角化 · 正规算子与自伴随算子

7.7.1 Schur 引理：上三角化的基础

Schur 引理是内积空间对角化的重要基石，我们将围绕它进行详细讨论：

Lemma

(***Schur's Lemma***)**:** 设 T 是**有限维内积空间** V 上的**线性算子**，若其**特征多项式分裂**（splits），则存在 V 的一组**标准正交基** β，使得 $[T]_\beta$ 为**上三角矩阵**。

证明. 对有限维内积空间 V 的维度 $\dim V = n$ 归纳：

基例（$n = 1$）当 $\dim V = 1$ 时，V 的标准正交基为单个单位向量 $\beta = \{v_1\}$。线性算子 T 对应的矩阵 $[T]_\beta$ 是 1×1 矩阵（仅含特征值），显然是上三角矩阵。

归纳步骤（假设 $\dim V = k$ 成立，推 $\dim V = k + 1$ 成立）

设 $\dim V = k + 1$，因 T 的特征多项式分裂，故存在特征值 $\lambda \in \mathbb{C}$ 及对应的单位特征向量 $v_1 \in V$（内积空间中特征向量可单位化）。

令 $W = v_1^\perp$（v_1 的正交补空间），则 $\dim W = k$，且 $V = \mathrm{span}\{v_1\} \oplus W$（正交直和）。

取 W 的标准正交基 $\{w_1, w_2, \ldots, w_k\}$，则 $\beta = \{v_1, w_1, w_2, \ldots, w_k\}$ 是 V 的标准正交基（v_1 与 W 正交且单位，W 的基是标准正交的）。

对 T 在基 β 下的矩阵 $[T]_\beta$:因 $Tv_1 = \lambda v_1$,故 $[T]_\beta$ 的第一列是 $(\lambda, 0, 0, \ldots, 0)^\top$；对 $w_i \in W$，Tw_i 可表示为 $a_i v_1 + T' w_i$（$a_i \in \mathbb{C}$，T' 是 W 上的线性算子）。由归纳假设，T' 在 $\{w_1, \ldots, w_k\}$ 下的矩阵是上三角矩阵。因此，$[T]_\beta$ 是上三角矩阵（第一列下方为 0，右下角 $k \times k$ 子矩阵是上三角）。

由归纳法，结论对所有有限维内积空间成立。 □

【Remark】:

在一般有限维线性空间（只有线性结构，无内积）中：若线性算子的特征多项式分裂，确实存在**普通线性基**使其矩阵为上三角矩阵，但此时不一定存在“标准正交基”。

Example

取 $V = \mathbb{C}^2$，线性算子 T 在标准基下的矩阵为：$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

其特征多项式为 $(\lambda - 1)(\lambda - 3)$（分裂）。

特征值 $\lambda_1 = 1$ 对应的单位特征向量：$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$；

特征值 $\lambda_2 = 3$ 对应的单位特征向量：$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$。

这两个向量构成标准正交基 β，则 T 在 β 下的矩阵为对角矩阵（上三角的特例）：$[T]_\beta = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$

7.7.2 正规算子：复内积空间的对角化

Definition

正规算子（Normal operator）： 设 V 为内积空间，T 是 V 上的线性算子。若 $TT^* = T^*T$，则称 T 是**正规**的。

正规矩阵 (normal matrix): 对于 $n \times n$ 实矩阵或复矩阵 A，若 $AA^* = A^*A$，则称 A 是**正规**的。

Claim

T 是正规的**当且仅当** $[T]_\beta$ 是正规的，其中 β 是标准正交基。

证明. **前置结论**：$[T^*]_\beta = A^*$（A^* 为 A 的共轭转置）。

1. **必要性**：$T^*T = TT^* \implies [T^*T]_\beta = [TT^*]_\beta \implies A^*A = AA^*$；
2. **充分性**：$A^*A = AA^* \implies [T^*T]_\beta = [TT^*]_\beta \implies T^*T = TT^*$。

结论：T 正规 $\iff [T]_\beta$ 正规。 □

我们举一个例子，好更快的熟悉这个定理：

Example

设 $T : \mathbb{R}^2 \to \mathbb{R}^2$ 是旋转角度为 θ（$0 < \theta < \pi$）的旋转变换。T 在标准有序基下的矩阵表示为 $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$。注意到 $AA^* = I = A^*A$，所以 A 是正规的，进而 T 是正规的。

观察，再上述旋转的例子中：算子 T 甚至没有一个特征向量。所以在实内积空间的情形下，正规性并不足以保证存在由特征向量构成的标准正交基。

不过，若 V 是复内积空间，正规性就足够了。

在证明关于正规算子的对角化结论之前，我们需要了解正规算子的一些一般性质。

Property

设 V 为内积空间，T 是 V 上的正规算子，则以下命题成立：

(a) 对任意 $x \in V$，有 $\|T(x)\| = \|T^*(x)\|$ 。

(b) 对任意 $c \in F$，$T - cI$ 是正规的。

(c) 若 x 是 T 的特征向量，则 x 也是 T^* 的特征向量。事实上，若 $T(x) = \lambda x$，则 $T^*(x) = \overline{\lambda} x$ 。

(d) 若 λ_1 和 λ_2 是 T 的不同特征值，对应的特征向量分别为 x_1 和 x_2，则 x_1 与 x_2 正交。

证明. 1. 对任意 $x \in V$，$\|T(x)\| = \|T^*(x)\|$：由范数与内积的关系，

$$\|T(x)\|^2 = \langle T(x), T(x)\rangle = \langle T^*T(x), x\rangle = \langle TT^*(x), x\rangle = \langle T^*(x), T^*(x)\rangle = \|T^*(x)\|^2,$$

开方得 $\|T(x)\| = \|T^*(x)\|$。

2. 对任意 $c \in F$，$T - cI$ 是正规的：先计算伴随：$(T - cI)^* = T^* - \overline{c}I$。展开验证正规性：

$$(T - cI)^*(T - cI) = (T^* - \overline{c}I)(T - cI) = T^*T - cT^* - \overline{c}T + |c|^2 I,$$

$$(T - cI)(T - cI)^* = (T - cI)(T^* - \overline{c}I) = TT^* - \overline{c}T - cT^* + |c|^2 I.$$

因 T 正规（$T^*T = TT^*$），故两式相等，即 $T - cI$ 正规。

3. 若 $T(x) = \lambda x$，则 $T^*(x) = \overline{\lambda}x$：由 (b)，$T - \lambda I$ 是正规算子；又 $(T - \lambda I)x = 0$，故 $\|(T - \lambda I)x\| = 0$。由 (a)，$\|(T - \lambda I)^*(x)\| = 0$，而 $(T - \lambda I)^* = T^* - \overline{\lambda}I$，因此：

$$(T^* - \overline{\lambda}I)x = 0 \implies T^*(x) = \overline{\lambda}x.$$

4. 若 $\lambda_1 \neq \lambda_2$ 是 T 的特征值, 对应特征向量 x_1, x_2, 则 $\langle x_1, x_2\rangle = 0$: 计算 $\lambda_1\langle x_1, x_2\rangle = \langle \lambda_1 x_1, x_2\rangle = \langle T(x_1), x_2\rangle = \langle x_1, T^*(x_2)\rangle$。由 (c)，$T^*(x_2) = \overline{\lambda_2}x_2$，故：

$$\langle x_1, T^*(x_2)\rangle = \langle x_1, \overline{\lambda_2}x_2\rangle = \lambda_2\langle x_1, x_2\rangle.$$

整理得 $(\lambda_1 - \lambda_2)\langle x_1, x_2\rangle = 0$，因 $\lambda_1 \neq \lambda_2$，故 $\langle x_1, x_2\rangle = 0$。

□

接下来，我们将展现第一个主定理：

Theorem

设 T 是有限维复内积空间 V 上的线性算子。则 T 是**正规的**当且仅当存在由 T 的**特征向量**组成的 V 的**标准正交基**, 即 T 可以在一组标准正交基下**可对角化**。

证明. **1）必要性：若 T 正规，则 V 有由 T 的特征向量组成的标准正交基**

对 $\dim V = n$ 用数学归纳法：

- **基例** $n = 1$：取 V 中单位向量 u_1，则 $\{u_1\}$ 是标准正交基，且 u_1 是 T 的特征向量（复线性空间必有特征值）。

- **归纳假设**：若 $\dim V < n$，正规算子在其不变子空间上的限制满足"有标准正交特征向量基"。

- **归纳步骤**（$\dim V = n$）：复内积空间中 T 必有特征值 λ，取对应单位特征向量 u_1（即 $Tu_1 = \lambda u_1$ 且 $\|u_1\| = 1$）。记 $V_1 = \text{span}\{u_1\}^\perp$（$u_1$ 的正交补），需证 V_1 是 T 的不变子空间：对 $v \in V_1$，由 T 正规，$T^*u_1 = \overline{\lambda}u_1$（因 $\langle Tv, u_1\rangle = \langle v, T^*u_1\rangle$，且 $\langle Tu_1, u_1\rangle = \lambda = \langle u_1, T^*u_1\rangle$，故 $T^*u_1 = \overline{\lambda}u_1$），因此：

$$\langle Tv, u_1\rangle = \langle v, T^*u_1\rangle = \langle v, \overline{\lambda}u_1\rangle = \lambda\langle v, u_1\rangle = 0 \implies Tv \in V_1.$$

故 V_1 是 T 的不变子空间，且 $T|_{V_1}$ 仍正规（因 $(T|_{V_1})^* = T^*|_{V_1}$，故 $(T|_{V_1})^*(T|_{V_1}) = T^*T|_{V_1} = TT^*|_{V_1} = (T|_{V_1})(T|_{V_1})^*$）。由归纳假设，$V_1$ 有标准正交基 $\{u_2, \ldots, u_n\}$（均为 $T|_{V_1}$ 的特征向量），则 $\{u_1, u_2, \ldots, u_n\}$ 是 V 的标准正交基，且均为 T 的特征向量。

2）充分性：若 V 有由 T 的特征向量组成的标准正交基，则 T 正规

设 $\{u_1, \ldots, u_n\}$ 是 V 的标准正交基，且 $Tu_i = \lambda_i u_i$（$\lambda_i \in \mathbb{C}$）。

对伴随算子 T^*，标准正交基下 T^* 的矩阵是 T 矩阵的共轭转置，故 $T^*u_i = \overline{\lambda_i}u_i$。因此：

$$TT^*u_i = T(\overline{\lambda_i}u_i) = \lambda_i\overline{\lambda_i}u_i, \quad T^*Tu_i = T^*(\lambda_i u_i) = \overline{\lambda_i}\lambda_i u_i.$$

即 $TT^*u_i = T^*Tu_i$ 对所有基向量成立，故 $TT^* = T^*T$，即 T 正规。

综上，T 正规 $\iff$ V 有由 T 的特征向量组成的标准正交基（T 可在标准正交基下对角化）。

□

因此，有限维**复**内积空间 V 上的线性算子 T 正规 $\iff$ 存在由 T 的特征向量组成的标准正交基。证明利用 Schur 引理得 $[T]_\beta = A$ 上三角，再由正规性及性质推导出 A 这个上三角阵必须为对角矩阵，即所有基向量均为特征向量。

有趣的是，上述定理并不适用于**无限维复内积空间**，但具体原因我们不再讨论。

7.7.3 自伴随算子：实内积空间的对角化

为了在**实数域上也可以对角化**, 我们需要更强的条件, 即自伴随算子, 满足 $T = T^*$。

Definition

自伴随算子（Self - adjoint operator）： 设 T 是内积空间 V 上的线性算子。若 $T = T^*$，则称 T 是自伴的（厄米特（Hermite）的）。
Hermite 矩阵： 对于 $n \times n$ 的复矩阵 A，若 $A = A^*$，则称 A 是自伴的，或厄米特的（Hermite）。对于实矩阵，此条件简化为要求 A 是对称的。

Claim

若 β 是标准正交基，那么 T 是自伴的当且仅当 $[T]_\beta$ 是自伴的。

证明. (1) 必要性：若 T 自伴，则 $[T]_\beta$ 自伴因 $T = T^*$，由矩阵对应关系：

$$[T]_\beta = [T^*]_\beta = ([T]_\beta)^*$$

故 $[T]_\beta$ 是自伴矩阵。

(2) 充分性：若 $[T]_\beta$ 自伴，则 T 自伴因 $[T]_\beta = ([T]_\beta)^*$，结合伴随矩阵性质得：

$$[T]_\beta = [T^*]_\beta$$

同一标准正交基下矩阵相等的算子必相等，故 $T = T^*$，即 T 自伴。

实矩阵的简化若 V 是实内积空间，实矩阵的共轭转置等于转置，故自伴条件简化为 $A = A^T$（对称矩阵）。 □

在陈述自伴算子的主要结论之前，我们需要做一些预备工作。

实内积空间上的线性算子仅有实特征值。但是接下来的引理表明，复内积空间上的自伴算子也有同样的结论。

类似地，复内积空间上每个线性算子的特征多项式都可分解，实内积空间上的自伴算子也是如此。

Lemma

设 T 是有限维内积空间 V 上的自伴算子。
(a) T 的每个特征值都是实的。
(b) 假设 V 是实内积空间。那么 T 的特征多项式可分解。

证明. a) **自伴算子的每个特征值都是实的**

设 λ 是自伴算子 T 的任一特征值，$v \in V$ 是对应非零特征向量（即 $Tv = \lambda v$ 且 $v \neq 0$）。由自伴算子定义 $T = T^*$，内积满足 $\langle Tv, v\rangle = \langle v, T^*v\rangle = \langle v, Tv\rangle$。

代入 $Tv = \lambda v$：

- 左边：$\langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle$；
- 右边：$\langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle$。

因 $v \neq 0$，故 $\langle v, v\rangle = \|v\|^2 > 0$，消去得 $\lambda = \overline{\lambda}$，即 λ 为实数。

b) **实内积空间上自伴算子的特征多项式可分解**

设 V 是有限维实内积空间，T 自伴：

1. 取 V 的标准正交基 β，则 $[T]_\beta = A$ 为实对称矩阵；2. A 的特征多项式 $p(\lambda) = \det(\lambda I - A)$ 是实系数多项式；3. 由 (a)，T 的特征值全为实数，故 $p(\lambda)$ 的根均为实根；4. 实系数多项式的根全为实数时，可分解为实系数一次因式的乘积：

$$p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n),$$

其中 $\lambda_1, \dots, \lambda_n$ 是 T 的特征值（计重数）。

因此 T 的特征多项式可分解。

□

于是，我们引入本节第二个主定理：

Theorem

设 T 是有限维**实**内积空间 V 上的线性算子。那么 T 是自伴的当且仅当存在一组由 T 的特征向量构成的，可以张成 V 的标准正交基 β。

证明. **充分性：若存在由 T 的特征向量组成的 V 的标准正交基，则 T 是自伴的**

设 $\beta = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n\}$ 是 V 的标准正交基，且每个 $\boldsymbol{u}_i$ 是 T 的特征向量，即 $T(\boldsymbol{u}_i) = \lambda_i \boldsymbol{u}_i$ $(\lambda_i \in \mathbb{R})$。

此时，T 在基 β 下的矩阵 A 是对角矩阵：

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

在标准正交基下，伴随算子 T^* 的矩阵是原矩阵的转置（实内积空间的伴随算子性质）。**而对角矩阵的转置等于自身**（$A^T = A$），因此 T^* 在 β 下的矩阵也为 A，即 $T^* = T$。故 T 是自伴算子。

必要性：若 T 是自伴的，则存在由 T 的特征向量组成的 V 的标准正交基

我们对 V 的维数 $\dim V = n$ 用数学归纳法证明：

基例：$n = 1$ 若 $\dim V = 1$，任取 V 中单位向量 $\boldsymbol{u}_1$（即 $\|\boldsymbol{u}_1\| = 1$），则 $\{\boldsymbol{u}_1\}$ 是 V 的标准正交基。由于 1 维线性算子必为数乘算子，故 $T(\boldsymbol{u}_1) = \lambda_1 \boldsymbol{u}_1$ $(\lambda_1 \in \mathbb{R})$，即 $\boldsymbol{u}_1$ 是 T 的特征向量。基例成立。

归纳假设：假设对所有 $\dim V = n - 1$ 的有限维实内积空间，自伴算子都存在由其特征向量组成的标准正交基。

归纳步骤: $\dim V=n$ 自伴算子存在实特征值, 设 λ_1 是 T 的实特征值, 构造单位特征向量与不变子空间将 $\boldsymbol{v}_1$ 单位化, 得 $\boldsymbol{u}_1=\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}$, 则 $\|\boldsymbol{u}_1\|=1$, 且 $T(\boldsymbol{u}_1)=\lambda_1\boldsymbol{u}_1$。

令 $W=\operatorname{span}\{\boldsymbol{u}_1\}$, 则 W 是 T- 不变子空间（因 $T(\boldsymbol{u}_1)\in W$）。

证明 $W^{\perp}$ 是 T- 不变子空间，对任意 $\boldsymbol{v}\in W^{\perp}$（即 $\langle\boldsymbol{v},\boldsymbol{u}_1\rangle=0$），由 T 的自伴性: $\langle T\boldsymbol{v},\boldsymbol{u}_1\rangle=\langle\boldsymbol{v},T\boldsymbol{u}_1\rangle=\langle\boldsymbol{v},\lambda_1\boldsymbol{u}_1\rangle=\lambda_1\langle\boldsymbol{v},\boldsymbol{u}_1\rangle=0$ 故 $T\boldsymbol{v}\in W^{\perp}$，即 $W^{\perp}$ 是 T-不变子空间。

对 $W^{\perp}$ 应用归纳假设由有限维内积空间的正交分解, $V=W\oplus W^{\perp}$, 故 $\dim W^{\perp}=n-1$。考虑 T 在 $W^{\perp}$ 上的限制 $T|_{W^{\perp}}$: 它是 $W^{\perp}$ 上的自伴算子（因对任意 $\boldsymbol{x},\boldsymbol{y}\in W^{\perp}$，$\langle T|_{W^{\perp}}\boldsymbol{x},\boldsymbol{y}\rangle=\langle T\boldsymbol{x},\boldsymbol{y}\rangle=\langle\boldsymbol{x},T\boldsymbol{y}\rangle=\langle\boldsymbol{x},T|_{W^{\perp}}\boldsymbol{y}\rangle$）。

根据归纳假设，$W^{\perp}$ 存在由 $T|_{W^{\perp}}$ 的特征向量组成的标准正交基 $\{\boldsymbol{u}_2,\ldots,\boldsymbol{u}_n\}$，且每个 $\boldsymbol{u}_i$ $(i\geq 2)$ 也是 T 的特征向量 (因 $T|_{W^{\perp}}\boldsymbol{u}_i=T\boldsymbol{u}_i$)。合并得到 V 的标准正交基由于 $\boldsymbol{u}_1\in W$，$\boldsymbol{u}_2,\ldots,\boldsymbol{u}_n\in W^{\perp}$，故 $\langle\boldsymbol{u}_1,\boldsymbol{u}_i\rangle=0$ $(i\geq 2)$；且 $\{\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_n\}$ 中每个向量都是单位向量、两两正交，因此是 V 的标准正交基，且每个向量都是 T 的特征向量。

综上，归纳步骤成立。 □

【Remark】:

自伴算子在**标准正交基**（不可以是任意基）下的矩阵表示为实对称矩阵。
而且如我们之前所提到的，实对称矩阵是自伴的，且自伴矩阵是正规的。但是，仅仅适用于实数空间。

Example

下面的矩阵 A 是复对称矩阵:

$$A=\begin{pmatrix} i & i\\ i & 1\end{pmatrix}\quad \text{且}\quad A^*=\begin{pmatrix} -i & -i\\ -i & 1\end{pmatrix}$$

但 A 不是正规的，因为 $(AA^*)_{12}=1+i$ 且 $(A^*A)_{12}=1-i$。因此，复对称矩阵不一定是正规的。

Theorem

设 T 是有限维**实**内积空间 V 上的线性算子。那么 T 是自伴的当且仅当 T 在空间 V 的一组标准正交基（即它标准正交的特征向量组）下**可对角化**.

我们类比上小节的推演，容易得到：有限维实内积空间 V 上的线性算子 T 自伴随 $\Longleftrightarrow$ 存在由 T 的特征向量组成的标准正交基。通过 Schur 引理得 $[T]_\beta=A$ 上三角，由

自伴随性 $A=A^*$（实矩阵时为转置），上三角矩阵取共轭不变，推出 A 为对角矩阵，从而基向量均为特征向量。

从矩阵的角度：

对于任意实对称矩阵 A ，都可分解为 $A=VDV^T$ 。其中 V 是由 A 的标准正交特征向量组成的矩阵（$V^TV=I$ ，I 为单位矩阵），D 是对角矩阵，其对角线上的元素是 A 对应的特征值。

综上，Schur 引理为内积空间算子的上三角化提供了工具，正规算子（复空间）与自伴随算子（实空间）分别在各自内积空间中，基于特征向量标准正交基实现了对角化，构建了内积空间对角化的核心理论体系。

7.8 谱定理与谱分解

7.8.1 谱定理与谱分解

Theorem

谱定理（Spectral theorem）：设 T 是**有限维内积空间** V 上具有**不同**特征值 $\lambda_1,\lambda_2,\cdots,\lambda_k$ 的线性算子，且 $F=\mathbb{R}$ 时自伴，$F=\mathbb{C}$ 时正规，则有以下结论：

(a) 空间直和分解：$V=W_1\oplus W_2\oplus\cdots\oplus W_k$，即 V 可分解为特征子空间 $W_1,W_2,\cdots,W_k$ 的直和。

(b) 正交补关系：若 W_i' 表示 $j\neq i$ 时 W_j 的直和，则 $W_i^\perp=W_i'$。

(c) 投影算子的正交性：T_i 是到 W_i 的正交投影：有 T_i 的像为 W_i，核为 $W_i^\perp=W_i'$，且 $T_iT_j=\delta_{ij}T_i$（$1\leq i,j\leq k$），其中 δ_{ij} 为克罗内克符号。

(d) 恒等算子分解：$I=T_1+T_2+\cdots+T_k$，即恒等算子可表示为各正交投影算子之和。

(e) 算子的谱分解式：$T=\lambda_1T_1+\lambda_2T_2+\cdots+\lambda_kT_k$，实现了 T 的谱分解。

证明. (a): 易证

(b):

$W_i'\subseteq W_i^\perp$：自伴 / 正规算子的不同特征值的特征向量正交，故对任意 $j\neq i$，W_j 中向量与 W_i 中向量正交，因此 W_i' 中向量均与 W_i 正交，即 $W_i'\subseteq W_i^\perp$。

$W_i^\perp\subseteq W_i'$：任取 $\boldsymbol{v}\in W_i^\perp$，由 (a) 的直和分解，$\boldsymbol{v}=\boldsymbol{v}_1+\cdots+\boldsymbol{v}_k$（$\boldsymbol{v}_j\in W_j$）。因 $\boldsymbol{v}\perp W_i$，故 $\boldsymbol{v}_i=\langle\boldsymbol{v},\boldsymbol{e}_i\rangle=0$（$\boldsymbol{e}_i\in W_i$ 是单位特征向量），因此 $\boldsymbol{v}=\bigoplus_{j\neq i}\boldsymbol{v}_j\in W_i'$。

(c): 若 $i\neq j$: T_j 的像 $W_j\subseteq W_i^\perp$(即 W_j 是 T_i 的核), 故 $T_iT_j(\boldsymbol{v})=T_i(\boldsymbol{v}_j)=\mathbf{0}$（$\boldsymbol{v}_j\in W_j$）。若 $i=j$：投影算子满足 $T_i^2=T_i$（投影到自身空间）。

(d): 任取 $\boldsymbol{v}\in V$，由 (a) 的直和分解，$\boldsymbol{v}=\boldsymbol{v}_1+\boldsymbol{v}_2+\cdots+\boldsymbol{v}_k$（$\boldsymbol{v}_i\in W_i$）。而

$T_i(\boldsymbol{v}) = \boldsymbol{v}_i$ (T_i 是到 W_i 的投影), 故: $I(\boldsymbol{v}) = \boldsymbol{v} = \boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k = T_1(\boldsymbol{v}) + \cdots + T_k(\boldsymbol{v}) = (T_1 + \cdots + T_k)(\boldsymbol{v})$ 因此 $I = T_1 + \cdots + T_k$。

(e): 任取 $\boldsymbol{v} \in V$，由 $T(\boldsymbol{v}_i) = \lambda_i \boldsymbol{v}_i$（$\boldsymbol{v}_i \in W_i$），结合 (d) 的分解：$T(\boldsymbol{v}) = T(\boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k) = T(\boldsymbol{v}_1) + \cdots + T(\boldsymbol{v}_k) = \lambda_1 \boldsymbol{v}_1 + \cdots + \lambda_k \boldsymbol{v}_k$ 又 $\boldsymbol{v}_i = T_i(\boldsymbol{v})$，故：$T(\boldsymbol{v}) = \lambda_1 T_1(\boldsymbol{v}) + \cdots + \lambda_k T_k(\boldsymbol{v}) = (\lambda_1 T_1 + \cdots + \lambda_k T_k)(\boldsymbol{v})$

因此 $T = \lambda_1 T_1 + \cdots + \lambda_k T_k$。 □

其中最后一条就是著名的“谱分解”:

Definition

谱分解（spectral decomposition）：对有限维内积空间 V 上的线性算子 T：若 $F = \mathbb{R}$ 时，T 为自伴算子；当 $F = \mathbb{C}$ 时，T 为正规算子。此时，T 可分解为：

$$T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 是 T 的不同特征值，T_i 是 V 到 λ_i 对应特征子空间 W_i 的正交投影算子。

特征值集合 $\{\lambda_1, \lambda_2, \cdots, \lambda_k\}$ 称为 T 的谱;

$I = T_1 + T_2 + \cdots + T_k$ 称为由 T 诱导的恒等算子的分解;

$T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$ 称为 T 的谱分解，且在特征值排列顺序下唯一。

谱分解深刻揭示了自伴算子 (实数域) 或正规算子 (复数域) 的内在结构, 将复杂的线性算子分解为简单正交投影算子的组合，为研究此类算子的性质提供了清晰的框架。

并且，我们可以从算子层面推广到矩阵层面:

T 的算子**谱分解**为：$T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$

其中 $T_i : V \to V$ 是到 W_i 的正交投影算子（即对任意 $v \in V$，$T_i(v)$ 是 v 在 W_i 上的正交投影)。

在特征子空间中选标准正交基, 拼接成 V 的标准正交基。因为 $W_1, W_2, \ldots, W_k$ 是正交直和, 我们可以: 对每个特征子空间 W_i, 选取它的标准正交基 (比如用 Gram-Schmidt 正交化): 假设 $\dim W_i = m_i$, 则 W_i 的标准正交基为 $\mathcal{B}_i = \{e_{i,1}, e_{i,2}, \ldots, e_{i,m_i}\}$。把所有 $\mathcal{B}_i$ 拼接起来,得到 V 的整体标准正交基:$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k = \{e_{1,1}, \ldots, e_{1,m_1}, e_{2,1}, \ldots, e_{2,m_2}, \ldots, e_{k,1}, \ldots,$

接着求 T 在标准正交基 $\mathcal{B}$ 下的矩阵对基 $\mathcal{B}$ 中的任意向量 $e_{i,j}$（属于 W_i），由特征子空间的定义：$T(e_{i,j}) = \lambda_i e_{i,j}$ 因此，T 在基 $\mathcal{B}$ 下的矩阵是对角矩阵 D：$D = \begin{pmatrix} \lambda_1 I_{m_1} & & & \\ & \lambda_2 I_{m_2} & & \\ & & \ddots & \\ & & & \lambda_k I_{m_k} \end{pmatrix}$ 其中 I_{m_i} 是 m_i 阶单位矩阵，对角元就是 T 的特征值 $\lambda_1, \lambda_2, \ldots, \lambda_k$（每个 λ_i 重复 m_i 次）。

注意到：正交投影算子 T_i 的定义是：对 W_i 内的向量 $e_{i,j}$：$T_i(e_{i,j}) = e_{i,j}$；对其他特征子空间的向量 $e_{l,m}$ （$l \neq i$）：$T_i(e_{l,m}) = 0$。因此，T_i 在基 $\mathcal{B}$ 下的矩阵是对角矩阵 D_i：

$$D_i = \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & I_{m_i} & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix}$$（只有对应 W_i 的基位置是单位矩阵，其余位置是 0）。

算子层面的谱分解 $T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$，在基 $\mathcal{B}$ 下对应的矩阵等式为：$D = \lambda_1 D_1 + \lambda_2 D_2 + \cdots + \lambda_k D_k$ 这显然成立：右边是 "特征值 × 对应位置的单位矩阵"，最终拼接成对角矩阵 D。

我们通常研究的是 $V = \mathbb{R}^n$ （或 $\mathbb{C}^n$）的标准基 $\mathcal{E} = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ （ε_i 是第 i 个分量为 1、其余为 0 的向量）。

设：从标准基 $\mathcal{E}$ 到新基 $\mathcal{B}$ 的过渡矩阵为 **Q**（实空间）或 **U**（复空间）：Q 的列向量就是基 $\mathcal{B}$ 的向量（因为 $\mathcal{B}$ 是标准正交基，所以 Q 是正交矩阵，满足 $Q^{-1} = Q^T$；复空间中 U 是酉矩阵，满足 $U^{-1} = U^H$，$U^H = \overline{U}^T$）；

算子 T 在标准基 $\mathcal{E}$ 下的矩阵为 A (即 $T(v) = Av$ 对任意 $v \in \mathbb{R}^n$ 成立)。根据 "算子在不同基下的矩阵关系"：$A = QDQ^{-1}$（算子在标准基下的矩阵 = 过渡矩阵 × 算子在新基下的矩阵 × 过渡矩阵的逆）结合 Q 是正交矩阵（$Q^{-1} = Q^T$），代入得：$A = QDQ^T$ 这就是实对称矩阵的谱分解形式（对应实空间自伴算子）。

复空间中，酉矩阵 U 满足 $U^{-1} = U^H$，因此复正规矩阵的谱分解形式为：$A = UDU^H$

所以我们总结出矩阵角度下的谱分解定理：

Theorem

矩阵可通过正交 / 酉变换对角化，分解形式为：

实对称矩阵： $A = QDQ^T$（Q 为正交矩阵，$Q^T = Q^{-1}$；D 为对角矩阵，对角元为 A 的实特征值，Q 的列是对应标准正交特征向量）；

复正规矩阵： $A = UDU^H$（U 为酉矩阵，$U^H = U^{-1}$；D 为对角矩阵，对角元为 A 的复特征值，U 的列是对应标准正交特征向量）。

Example

从算子谱分解到实对称矩阵的 $A = QDQ^T$ 取实对称矩阵 $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$，对应的算子 $T : \mathbb{R}^2 \to \mathbb{R}^2$ 是 $T(v) = Av$。

算子谱分解： T 的特征值：$\lambda_1 = 4$，$\lambda_2 = 2$；特征子空间：$W_1 = \ker(T - 4I) = \text{span}\{(1,1)^T\}$，$W_2 = \ker(T - 2I) = \text{span}\{(1,-1)^T\}$；正交投影算子：$T_1$ 是到 W_1 的正交投影，T_2 是到 W_2 的正交投影，算子谱分解为 $T = 4T_1 + 2T_2$。

选标准正交基:W_1 的标准正交基:$e_1=\frac{1}{\sqrt{2}}(1,1)^T$;$W_2$ 的标准正交基:$e_2=\frac{1}{\sqrt{2}}(1,-1)^T$;整体标准正交基 $\mathcal{B}=\{e_1,e_2\}$。过渡矩阵 Q:$Q=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{pmatrix}$(正交矩阵)。对角矩阵 D:$D=\begin{pmatrix}4 & 0\\ 0 & 2\end{pmatrix}$。

验证矩阵谱分解:$QDQ^T=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{pmatrix}\begin{pmatrix}4 & 0\\ 0 & 2\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{pmatrix}=\begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix}=A$,得证。

7.8.2 谱定理的相关推论

Corollary & Secondary Conclusion

若 $F=\mathbb{C}$,则 T 正规当且仅当存在多项式 g,使得 $T^*=g(T)$。

证明. 设 $T=\lambda_1T_1+\lambda_2T_2+\cdots+\lambda_kT_k$ 为谱分解,因 T_i 自伴,故:

$$T^*=\overline{\lambda_1}T_1+\overline{\lambda_2}T_2+\cdots+\overline{\lambda_k}T_k$$

利用拉格朗日插值公式,选取多项式 g 使 $g(\lambda_i)=\overline{\lambda_i}$ $(1\le i\le k)$,则:

$$g(T)=\overline{\lambda_1}T_1+\overline{\lambda_2}T_2+\cdots+\overline{\lambda_k}T_k=T^*$$

反之,若 $T^*=g(T)$,则 T 与 T^* 可交换,故 T 正规。 □

Corollary & Secondary Conclusion

若 $F=\mathbb{C}$,则 T 酉当且仅当 T 正规且 T 的每个特征值 λ 满足 $|\lambda|=1$。

证明. 若 T 酉,则 T 正规且由定理 6.18 推论 2 知 $|\lambda|=1$。反之,设 $T=\lambda_1T_1+\lambda_2T_2+\cdots+\lambda_kT_k$,若 $|\lambda_i|=1$,则:

$$TT^*=(\lambda_1T_1+\cdots+\lambda_kT_k)(\overline{\lambda_1}T_1+\cdots+\overline{\lambda_k}T_k)=|\lambda_1|^2T_1+\cdots+|\lambda_k|^2T_k=T_1+\cdots+T_k=I$$

故 T 酉。 □

Corollary & Secondary Conclusion

若 $F=\mathbb{C}$ 且 T 正规,则 T 自伴当且仅当 T 的每个特征值为实数。

证明. 设 $T=\lambda_1T_1+\lambda_2T_2+\cdots+\lambda_kT_k$,若特征值均实,则:

$$T^*=\overline{\lambda_1}T_1+\cdots+\overline{\lambda_k}T_k=\lambda_1T_1+\cdots+\lambda_kT_k=T$$

反之,由定理 6.17 引理已证。 □

Corollary & Secondary Conclusion

设 $T=\lambda_1T_1+\lambda_2T_2+\cdots+\lambda_kT_k$ 为谱分解，则每个 T_j 是 T 的多项式。

证明. 选取多项式 g_j（$1\le j\le k$）使 $g_j(\lambda_i)=\delta_{ij}$，则：

$$g_j(T)=g_j(\lambda_1)T_1+g_j(\lambda_2)T_2+\cdots+g_j(\lambda_k)T_k=\delta_{1j}T_1+\delta_{2j}T_2+\cdots+\delta_{kj}T_k=T_j$$

□

7.9 （应用）最小二乘法

在数据拟合的背景下，考虑测量值 $(t_1,y_1),(t_2,y_2),\ldots,(t_m,y_m)$。我们旨在找到一条最适合数据的直线 $y=ct+d$。误差 E 定义为点到直线垂直距离的平方和，由下式给出：

$$E=\sum_{i=1}^{m}(y_i-ct_i-d)^2=\|y-Ax\|^2$$

其中 $A=\begin{pmatrix}t_1&1\\t_2&1\\\vdots&\vdots\\t_m&1\end{pmatrix}$，$x=\begin{pmatrix}c\\d\end{pmatrix}$，且 $y=\begin{pmatrix}y_1\\y_2\\\vdots\\y_m\end{pmatrix}$。

为了最小化 E，我们使用定理：对于 $A\in M_{m\times n}(F)$ 和 $y\in F^m$，存在 $x_0\in F^n$ 使得 $(A^*A)x_0=A^*y$ 且对于所有 $x\in F^n$ 有 $\|Ax_0-y\|\le\|Ax-y\|$。如果 $\operatorname{rank}(A)=n$，那么 $x_0=(A^*A)^{-1}A^*y$。

两个关键引理支持这一点：

引理 1：$\langle Ax,y\rangle_m=\langle x,A^*y\rangle_n$，建立了内积之间的关系。

引理 2：$\operatorname{rank}(A^*A)=\operatorname{rank}(A)$，推论为如果 $\operatorname{rank}(A)=n$，A^*A 是可逆的。

例如，对于数据 $(1,2),(2,3),(3,5),(4,7)$，$A=\begin{pmatrix}1&1\\2&1\\3&1\\4&1\end{pmatrix}$，$y=\begin{pmatrix}2\\3\\5\\7\end{pmatrix}$。计算 $A^*A=\begin{pmatrix}30&10\\10&4\end{pmatrix}$，其逆 $(A^*A)^{-1}=\frac{1}{20}\begin{pmatrix}4&-10\\-10&30\end{pmatrix}$，且 $x_0=(A^*A)^{-1}A^*y=\begin{pmatrix}1.7\\0\end{pmatrix}$，得到最小二乘直线 $y=1.7t$。

如果 $\sum_{i=1}^{m}t_i=0$，A 的列是正交的，将 A^*A 简化为对角矩阵。这种方法通过调整 A 为 $\begin{pmatrix}t_1^2&t_1&1\\\vdots&\vdots&\vdots\\t_m^2&t_m&1\end{pmatrix}$ 推广到多项式拟合（例如 $y=ct^2+dt+e$），将最小二乘原理扩展到更高次多项式。

7.10 矩阵的正定性

Definition

半定性: 我们称一个（对称）矩阵 A 为正定半定（PSD）或负定半定（NSD），当且仅当对所有向量 v，满足

$$v^\top Av \geq 0 \text{ (或 } v^\top Av \leq 0)$$

并且，我们有如下 2 条断言;

Claim

对称矩阵是 PSD 的当且仅当**所有特征值非负**

证明. 设 A 是对称矩阵,由谱分解定理,$A = Q\Lambda Q^\top$(Q 是正交矩阵,$\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ 是特征值构成的对角矩阵)。对任意向量 v，令 $w = Q^\top v$，则:

$$v^\top Av = w^\top \Lambda w = \sum_{i=1}^{n} \lambda_i w_i^2$$

若所有 $\lambda_i \geq 0$，则 $\sum_{i=1}^{n} \lambda_i w_i^2 \geq 0$，即 A 是 PSD 的;

反之,若 A 是 PSD 的,对每个特征向量 v_i(满足 $Av_i = \lambda_i v_i$),有 $v_i^\top Av_i = \lambda_i v_i^\top v_i \geq 0$,而 $v_i^\top v_i > 0$，故 $\lambda_i \geq 0$。 □

Claim

对**任意**矩阵 A，$A^\top A$ 是对称 PSD 矩阵

证明. 对称性：$(A^\top A)^\top = A^\top(A^\top)^\top = A^\top A$，故 $A^\top A$ 是对称矩阵。

PSD 性：对任意向量 v，令 $w = Av$，则:

$$v^\top(A^\top A)v = (Av)^\top(Av) = w^\top w = \|w\|^2 \geq 0$$

因此 $A^\top A$ 是 PSD 的。 □

【Remark】:

PSD 性质通常针对**对称矩阵**半定性的定义依赖于二次型 $v^\top Av$，只有对称矩阵的二次型能完整刻画其“半定”属性（非对称矩阵的二次型需通过对称化处理，见下一条），因此通常仅对**对称矩阵**讨论 PSD 性质。

但是，为了考虑到定义的完整性，我们对非对称矩阵也给出了相应的 PSD 定义:

首先，注意到:

对任意矩阵 A，二次型 $v^\top Av$ 可变形为:

$$v^\top Av = \frac{1}{2}v^\top(A + A^\top)v$$

其中 $\frac{1}{2}(A + A^\top)$ 是对称矩阵.

证明. 步骤 1：设 A 是任意 $n \times n$ 矩阵，v 是 n 维向量。

步骤 2：展开并相加两个二次型：将 $v^\top Av$ 和 $v^\top A^\top v$ 展开后相加：

$$v^\top Av + v^\top A^\top v = v^\top \left(A + A^\top\right) v$$

步骤 3：注意到 $v^\top A^\top v$ 是一个标量，而标量的转置等于自身，因此：

$$v^\top A^\top v = \left(v^\top A^\top v\right)^\top = v^\top Av$$

步骤 4：化简等式：将 $v^\top A^\top v = v^\top Av$ 代入步骤 2 的等式，得：

$$v^\top Av + v^\top Av = v^\top \left(A + A^\top\right) v$$

即：

$$2v^\top Av = v^\top \left(A + A^\top\right) v$$

步骤 5：两边同时除以 2，最终得到：

$$v^\top Av = \frac{1}{2}v^\top \left(A + A^\top\right) v$$

□

由于对称化矩阵是对称的，因此非对称矩阵的半定性可通过其对称化矩阵来定义。

Definition

因此，非对称矩阵 A 的 PSD 可通过其对称化矩阵 $\frac{1}{2}(A + A^\top)$ 来定义。

在半定的基础上，我们可以更进一步：

Definition

正定/负定矩阵（PD/ND）：我们称对称矩阵 A 为正定（PD）或负定（ND），当且仅当对所有 $v \neq 0$，有

$$v^\top Av > 0 \text{ (或 } v^\top Av < 0)$$

我们有：

- **PD 矩阵一定是半正定（PSD）的**，PD 是更强的概念；
- **对称矩阵是 PD** 的当且仅当它的所有**特征值都是正的**。

证明. **充分性（特征值全正 $\implies$ A 正定）：**
由实对称矩阵正交对角化定理，存在正交矩阵 $Q(Q^TQ=I)$ 和对角矩阵 $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$，使得 $A = Q\Lambda Q^T$，其中 $\lambda_i > 0$。对任意非零向量 $\mathbf{x} \in \mathbb{R}^n$，令 $\mathbf{y} = Q^T\mathbf{x} \neq \mathbf{0}$（因 Q 可逆）。则：

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T Q\Lambda Q^T \mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y} = \sum_{i=1}^{n} \lambda_i y_i^2 > 0,$$

故 A 正定。

必要性（A 正定 $\implies$ 特征值全正）：
设 λ 是 A 的特征值，$\mathbf{v}$ 为对应单位特征向量（$\mathbf{v}^T\mathbf{v} = 1$），满足 $A\mathbf{v} = \lambda\mathbf{v}$。则：

$$\lambda = \lambda\mathbf{v}^T\mathbf{v} = \mathbf{v}^T A\mathbf{v} > 0,$$

（因 A 正定，$\mathbf{v} \neq \mathbf{0}$ 故二次型为正）。因此所有特征值均正。 □

当然，我们还有一种考察正定性的方法：

Method

（西尔维斯特准则，Sylvester’s criterion） 设 $\mathbb{F}$ 为实数域 $\mathbb{R}$ 或复数域 $\mathbb{C}$，$A = (a_{ij})_{i,j} \in M_n(\mathbb{F})$ 是一个埃米特矩阵（即满足 ${}^t\overline{A} = A$）。记 $A_k := (a_{ij})_{1\le i,j\le k}$（其中 $1 \le k \le n$），则 A 是正定矩阵的充要条件是：对所有 k，均有 $\det(A_k) > 0$。

Example

设 $A \in M_n(\mathbb{C})$，若 A 是对角埃米特矩阵，即 $A = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 且对任意 i 有 $\lambda_i \in \mathbb{R}$，则 A 是正定矩阵当且仅当对任意 k，有 $\lambda_k = \det(A_k)\cdot\det(A_{k-1})^{-1} > 0$，当且仅当对任意 k，有 $\det(A_k) = \prod_{i=1}^{k} \lambda_i > 0$。

7.11 奇异值分解（SVD）

还是延续我们对角化的思路：对于 m×n 的矩阵 (m 不等于 n)，我们显然无法对其进行对角化。但是，参考 Jordan 标准型的讨论，我们依然可以做一些妥协来实践一种近似的对角化——矩形-对角阵。采取的策略就是奇异值。

我们已经考察了大量定义在同一个空间的线性算子，而接下来，我们将考察**不同空间之间的线性变换及其矩阵表示**，有趣的是其矩阵表示往往就是刚刚我们提到的 m×n 的矩阵 (m 不等于 n)。

奇异值定理就涉及**两个（通常不同的）内积空间以及两个（通常不同的）正交标准基**。如果这两个空间和两个基是相同的，那么这个变换实际上将是一个正规或自伴算子。

奇异值定理中的数值不变量，即奇异值，是非负的，而与之对应的特征值则没有这样的限制。这个性质对于保证奇异值的唯一性是必要的。

奇异值在图像处理，仿射变换具有重要的地位。

奇异值分解在实际应用中发挥着关键作用：

应用领域	**具体作用**
数据压缩	保留大奇异值部分，减少存储与计算量
图像处理	去除小奇异值降噪，提升图像质量
机器学习与数据分析	降维（如 PCA），避免维度灾难，提升算法效率
推荐系统	通过矩阵分解挖掘数据潜在模式

表 7.1: 奇异值分解的实际应用

7.11.1 线性变换的奇异值定理

Theorem

线性变换的奇异值定理（The Singular Value Theorem of Linear Transformations）： 设 V 和 W 为有限维内积空间，$T : V \to W$ 是一个秩为 r 的线性变换。则存在 V 的标准正交基 $\{v_1, v_2, \ldots, v_n\}$ 、W 的标准正交基 $\{u_1, u_2, \ldots, u_m\}$ 以及正标量 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ ，使得

$$T(v_i) = \begin{cases} \sigma_i u_i & 若 1 \leq i \leq r \\ 0 & 若 i > r \end{cases}$$

反之，假设上述条件成立。那么对于 $1 \leq i \leq n$ ，v_i 是 T^*T 的特征向量，当 $1 \leq i \leq r$ 时，对应的特征值为 σ_i^2 ，当 $i > r$ 时，特征值为 0 。因此，标量 $\sigma_1, \sigma_2, \ldots, \sigma_r$ 由 T 唯一确定。

证明. T^*T 是 V 上自伴正算子（因 $\forall v \in V, \langle T^*Tv, v\rangle = \langle Tv, Tv\rangle \geq 0$）。

由自伴算子的谱定理，V 存在标准正交基 $\{v_1, \ldots, v_n\}$，使得 $T^*Tv_i = \lambda_i v_i$，其中 $\lambda_i \geq 0$，且可排序为：$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$，$\lambda_{r+1} = \cdots = \lambda_n = 0$（因 $\operatorname{rank}(T^*T) = \operatorname{rank}(T) = r$）。

令 $\sigma_i = \sqrt{\lambda_i}$ $(1 \leq i \leq r)$，定义 $u_i = \frac{1}{\sigma_i} T v_i$ $(1 \leq i \leq r)$：

正交性：$\langle u_i, u_j\rangle = \frac{1}{\sigma_i\sigma_j}\langle Tv_i, Tv_j\rangle = \frac{1}{\sigma_i\sigma_j}\langle T^*Tv_i, v_j\rangle = \frac{\lambda_i}{\sigma_i\sigma_j}\delta_{ij} = \delta_{ij}$（$\delta_{ij}$ 为克罗内克函数），故 $\{u_1, \ldots, u_r\}$ 是 W 的正交单位组；

扩充：将 $\{u_1, \ldots, u_r\}$ 扩充为 W 的标准正交基 $\{u_1, \ldots, u_m\}$。

由于自伴算子 T^*T 的特征值唯一（不计重数），故 $\sigma_i = \sqrt{\lambda_i}$ 唯一，证毕。

□

Fact or Background

被分解的矩阵 A 的秩就等于奇异值的个数。

由此我们可以引出线性变换对应的矩阵表示的奇异值概念：

Definition

奇异值 (Singular value)： 对于 $m \times n$ 矩阵 A，其奇异值是 A^*A（或 AA^*）的非零特征值的平方根，按从大到小排列为 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$（$r$ 为 A 的秩）。

Fact or Background

从线性变换角度，奇异值是线性变换 $T = L_A$ 对应的特殊标量，**刻画了变换对向量长度的缩放因子**。

证明. 对 V 中 SVD 构造的标准正交基 $\{v_1, \ldots, v_n\}$ (称为 T 的 右奇异向量)：当 $i \leq r$：$T(v_i) = \sigma_i u_i$，其中 u_i 是 W 的标准正交基（左奇异向量）。

由于 $\|v_i\| = \|u_i\| = 1$（标准正交基），$\|T(v_i)\| = \|\sigma_i u_i\| = \sigma_i \cdot \|u_i\| = \sigma_i$——即 σ_i 是向量 v_i 经 T 变换后 长度的精确缩放比例。当 $i > r$：$T(v_i) = 0$，对应缩放比例为 0（这些向量属于 T 的零空间 $\ker(T)$）。 □

又因为单独的奇异值并不好求，因此我们需要架构一些线性变换的固定属性，如特征值与奇异值之间的联系：

$\langle T^*T(v), v\rangle = \|T(v)\|^2$，所以 T^*T 的物理意义是“向量经 T 变换后的长度平方”

于是，我们找到了奇异值与特征值的联系！

7.11.2 奇异值分解的原理与方法

Theorem

矩阵奇异值分解定理 (Matrix Singular Value Decomposition Theorem)： 对秩为 r 的 $m \times n$ 矩阵 A，存在 $m \times m$ 酉矩阵 U 和 $n \times n$ 酉矩阵 V，使得

$$A = U\Sigma V^*$$

其中 Σ 是 $m \times n$ 矩阵，主对角线前 r 个元素为奇异值 σ_i，其余为 0。

矩阵 A 的奇异值分解 $A = U\Sigma V^*$ 计算 A^*A 的特征值 λ_i 与对应的规范正交特征向量 v_i 取奇异值 $\sigma_i = \sqrt{\lambda_i}$，构造对角矩阵 Σ 对 $i \leq r$，计算 $u_i = \frac{1}{\sigma_i} A v_i$ 将 $\{u_i\}$ 扩展为

$\mathbb{F}^m$ 的规范正交基，构成矩阵 U 返回 U, Σ, V

其中：

V 是 $n \times n$ 正交矩阵，列向量 $\{v_1, \ldots, v_n\}$ 是输入空间 $\mathbb{R}^n$ 的正交基，对应线性变换 T^*T 的特征向量（输入方向的"主成分"）。几何上，V 表示对输入空间的旋转 / 反射（正交变换），将标准基转换为输入基。Σ（缩放矩阵）：

Σ 是 $m \times n$ 矩形-对角矩阵，对角元素 $\sigma_1 \geq \cdots \geq \sigma_r > 0$ 为奇异值，代表线性变换 A 在各输入基方向上的拉伸 / 压缩因子。几何上，Σ 对输入基向量进行缩放（仅前 r 个方向有效缩放，其余为零）。

U 是 $m \times m$ 正交矩阵，列向量 $\{u_1, \ldots, u_m\}$ 是输出空间 $\mathbb{R}^m$ 的正交基，对应线性变换 TT^* 的特征向量（输出方向的"主成分"）。几何上，U 表示对输出空间的旋转 / 反射（正交变换），将缩放后的向量转换为输出基。

接下来，我们来考察几个奇异值分解的例子:

Example

以矩阵 $A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix}$ 为例：

1. 计算 $A^*A = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{pmatrix}$，求得特征值 $\lambda_1 = 6$，$\lambda_2 = \lambda_3 = 0$。

2. 对应特征向量：

$$v_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad v_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

构成矩阵 V：

$$V = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}$$

3. 奇异值 $\sigma_1 = \sqrt{6}$，构造 Σ：

$$\Sigma = \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

4. 计算 $u_1 = \frac{1}{\sqrt{6}}Av_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$，选取 $u_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$ 构成矩阵 U：

$$U = \begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{pmatrix}$$

5. 最终 $A = U\Sigma V^*$，完成奇异值分解。

Example

求矩阵 $A = \begin{bmatrix}1 & 1\\1 & 1\\0 & 0\end{bmatrix}$ 的奇异值分解

7.11.3 极分解

矩阵的奇异值分解可用于分解方阵，其方式类似于将复数分解为模长为 1 的复数与非负数的乘积。在矩阵的情形中，模长为 1 的复数被替换为**酉矩阵**，非负数被替换为**半正定矩阵**。

Theorem

极分解（Polar decomposition）：对任意方阵 A，存在酉矩阵 W 和半正定矩阵 P，使得

$$A = WP$$

此外，若 A 可逆，则该分解是唯一的。

证明. 存在酉矩阵 U、V 和对角元非负的对角矩阵 Σ，使得 $A = U\Sigma V^*$。因此

$$A = U\Sigma V^* = UV^* \cdot V\Sigma V^* = WP$$

其中 $W = UV^*$，$P = V\Sigma V^*$。由于 W 是酉矩阵的乘积，故 W 是酉矩阵；又因 Σ 是半正定矩阵，且 P 与 Σ 酉等价，且容易证明 P 是半正定矩阵。

现假设 A 可逆，且可分解为 $A = WP = ZQ$（其中 W、Z 是酉矩阵，P、Q 是半正定矩阵）。由于 A 可逆，故 P、Q 是正定矩阵且可逆，因此 $Z^*W = QP^{-1}$。由此 QP^{-1} 是酉矩阵，故

$$I = (QP^{-1})^*(QP^{-1}) = P^{-1}Q^2P^{-1}$$

因此 $P^2 = Q^2$。由于 P、Q 均为正定矩阵，由 $P = Q$，进而 $W = Z$，故分解是唯一的。 □

将方阵 A 分解为 WP（其中 W 是酉矩阵，P 是半正定矩阵）的形式，称为 A 的**极分解**。

Example

求 $A=\begin{pmatrix}11 & -5\\ -2 & 10\end{pmatrix}$ 的极分解，首先求其奇异值分解 $U\Sigma V^*$。目标是找到 $\mathbb{R}^2$ 的标准正交基 β，使其由 A^*A 的特征向量组成。

可以验证：

$$v_1=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -1\end{pmatrix},\quad v_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 1\end{pmatrix}$$

是 A^*A 的标准正交特征向量，对应的特征值分别为 $\lambda_1=200$、$\lambda_2=50$。因此 $\beta=\{v_1,v_2\}$ 是合适的基。于是奇异值 $\sigma_1=\sqrt{200}=10\sqrt{2}$，$\sigma_2=\sqrt{50}=5\sqrt{2}$，故

$$V=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix},\quad \Sigma=\begin{pmatrix}10\sqrt{2} & 0\\ 0 & 5\sqrt{2}\end{pmatrix}$$

接下来求 U 的列 u_1 和 u_2：

$$u_1=\frac{1}{\sigma_1}Av_1=\frac{1}{5}\begin{pmatrix}4\\ -3\end{pmatrix},\quad u_2=\frac{1}{\sigma_2}Av_2=\frac{1}{5}\begin{pmatrix}3\\ 4\end{pmatrix}$$

因此

$$U=\begin{pmatrix}\frac{4}{5} & \frac{3}{5}\\ -\frac{3}{5} & \frac{4}{5}\end{pmatrix}$$

于是，按照刚刚的记号，有

$$W=UV^*=\begin{pmatrix}\frac{4}{5} & \frac{3}{5}\\ -\frac{3}{5} & \frac{4}{5}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}=\frac{1}{5\sqrt{2}}\begin{pmatrix}7 & -1\\ 1 & 7\end{pmatrix}$$

且

$$P=V\Sigma V^*=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}\begin{pmatrix}10\sqrt{2} & 0\\ 0 & 5\sqrt{2}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}=\frac{5}{\sqrt{2}}\begin{pmatrix}3 & -1\\ -1 & 3\end{pmatrix}$$

7.12 双线性映射与半双线性映射

回顾以往的知识，我们有：

Definition

双线性映射的定义：设 V,W,U 是域 $\mathbb{F}$ 上的向量空间，映射 $f:V\times W\to U$ 若满足：

- 对第一个变量线性：$\forall v_1,v_2\in V,w\in W,a\in\mathbb{F}$，有 $f(v_1+v_2,w)=f(v_1,w)+f(v_2,w)$ 且 $f(av,w)=af(v,w)$；
- 对第二个变量线性：$\forall v\in V,w_1,w_2\in W,a\in\mathbb{F}$，有 $f(v,w_1+w_2)=f(v,w_1)+f(v,w_2)$ 且 $f(v,aw)=af(v,w)$，

则称 f 为**双线性映射**。

比如，实数空间的内积就是一个典型的双线性映射。

比如在平面 $\mathbb{R}^2$ 里，取向量 $\vec{v}=(2,3)$、$\vec{w}=(4,5)$，点积是：$(\vec{v},\vec{w})=2\times4+3\times5=8+15=23$。

它满足“双线性”：比如 $(a\vec{v}_1+b\vec{v}_2,\vec{w})=a(\vec{v}_1,\vec{w})+b(\vec{v}_2,\vec{w})$——比如 $(2\times(1,0)+3\times(0,1),(4,5))=2\times(1,0)\cdot(4,5)+3\times(0,1)\cdot(4,5)=2\times4+3\times5=23$，和直接算结果一样。

并且取标准基 $\mathcal{B}=\{\vec{e}_1=(1,0),\vec{e}_2=(0,1)\}$。

先算基向量两两的点积：$(\vec{e}_1,\vec{e}_1)=1\times1+0\times0=1$，$(\vec{e}_1,\vec{e}_2)=1\times0+0\times1=0(\vec{e}_2,\vec{e}_1)=0\times1+1\times0=0$，$(\vec{e}_2,\vec{e}_2)=0\times0+1\times1=1$ 所以点积在基 $\mathcal{B}$ 下的矩阵 $M_{\mathcal{B}}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$（单位矩阵）。

现在用矩阵算 $\vec{v}=(2,3)$、$\vec{w}=(4,5)$ 的点积：$(\vec{v},\vec{w})=(4,5)\times M_{\mathcal{B}}\times\begin{pmatrix}2\\3\end{pmatrix}=(4,5)\times\begin{pmatrix}2\\3\end{pmatrix}=23$——和直接算点积结果一模一样！

再考虑一个高维度的双线性映射（$V_1=V_2=\mathbb{R}^2$，$W=\mathbb{R}^3$）

定义映射 $\phi:\mathbb{R}^2\times\mathbb{R}^2\to\mathbb{R}^3$：对 V_1 中向量 $\boldsymbol{v}=(a,b)$、V_2 中向量 $\boldsymbol{w}=(c,d)$，令：$\phi(\boldsymbol{v},\boldsymbol{w})=(ac,\ ad+bc,\ bd)$

验证双线性：

对第一个变量线性： 取 $\boldsymbol{v}_1=(a_1,b_1)$、$\boldsymbol{v}_2=(a_2,b_2)$，$k\in\mathbb{R}$，

先分别算两个映射结果，再向量相加（对应分量相加）：$\phi(\boldsymbol{v}_1,\boldsymbol{w})=(a_1c,\ a_1d+b_1c,\ b_1d)\phi(\boldsymbol{v}_2,\boldsymbol{w})=(a_2c,\ a_2d+b_2c,\ b_2d)$

则：$\phi(\boldsymbol{v}_1+\boldsymbol{v}_2,\boldsymbol{w})=((a_1+a_2)c,\ (a_1+a_2)d+(b_1+b_2)c,\ (b_1+b_2)d)=\phi(\boldsymbol{v}_1,\boldsymbol{w})+\phi(\boldsymbol{v}_2,\boldsymbol{w})$,$\phi(k\boldsymbol{v},\boldsymbol{w})=(kac,\ kad+kbc,\ kbd)=k\cdot\phi(\boldsymbol{v},\boldsymbol{w})$

对第二个变量线性： 同理，$\phi(\boldsymbol{v},\boldsymbol{w}_1+\boldsymbol{w}_2)=\phi(\boldsymbol{v},\boldsymbol{w}_1)+\phi(\boldsymbol{v},\boldsymbol{w}_2)$，$\phi(\boldsymbol{v},k\boldsymbol{w})=k\cdot\phi(\boldsymbol{v},\boldsymbol{w})$。

而仿照双线性映射的格式，突出“**其中一个变量的数乘要取域的共轭**”这个核心特点，我们可以构造：

Definition

半双线性映射的定义：设 $\mathbb{F}$ 是一个带有 **共轭自同构** $\bar{\cdot}:\mathbb{F}\to\mathbb{F}$（例如复数域 $\mathbb{C}$ 中 $\overline{a+bi}=a-bi$）的域，V,W,U 是 $\mathbb{F}$ 上的向量空间，映射 $f:V\times W\to U$ 若满足：

- 对第一个变量线性：$\forall v_1,v_2\in V,w\in W,a\in\mathbb{F}$，有 $f(v_1+v_2,w)=f(v_1,w)+f(v_2,w)$ 且 $f(av,w)=af(v,w)$；
- 对第二个变量共轭线性：$\forall v\in V,w_1,w_2\in W,a\in\mathbb{F}$，有 $f(v,w_1+w_2)=f(v,w_1)+f(v,w_2)$ 且 $f(v,aw)=\bar{a}f(v,w)$，

则称 f 为 **半双线性映射**。

复数空间里的内积是半双线性的（右边数乘要取共轭）。比如在 $\mathbb{C}^2$ 里，取 $\vec{v}=(1+i,2)$、$\vec{w}=(3,4-i)$，内积是：$(\vec{v},\vec{w})=(1+i)\times\overline{3}+2\times\overline{4-i}$（$\overline{x}$ 是取共轭）计算：$(1+i)\times 3+2\times(4+i)=3+3i+8+2i=11+5i$。

这里“半双线性”体现在：右边的系数 3 变成了 $\overline{3}=3$，右边的 $4-i$ 变成了 $\overline{4-i}=4+i$——如果是实数，共轭就是自己，所以实数内积其实是半双线性的“特例”。

半双线性映射（原实数域调整为 $V_1=V_2=\mathbb{C}^2$，$W=\mathbb{C}^3$）

定义映射 $\psi:\mathbb{C}^2\times\mathbb{C}^2\to\mathbb{C}^3$：对 V_1 中向量 $\boldsymbol{v}=(a,b)$（$a,b\in\mathbb{C}$）、V_2 中向量 $\boldsymbol{w}=(c,d)$（$c,d\in\mathbb{C}$），令：$\psi(\boldsymbol{v},\boldsymbol{w})=\left(a\bar{c},\ a\bar{d}+b\bar{c},\ b\bar{d}\right)$

验证半双线性：

对第一个变量线性：和双线性一样，$\psi(\boldsymbol{v}_1+\boldsymbol{v}_2,\boldsymbol{w})=\psi(\boldsymbol{v}_1,\boldsymbol{w})+\psi(\boldsymbol{v}_2,\boldsymbol{w})$，$\psi(k\boldsymbol{v},\boldsymbol{w})=k\cdot\psi(\boldsymbol{v},\boldsymbol{w})$（$k\in\mathbb{C}$）。

对第二个变量共轭线性(半双线性的核心)：取 $k\in\mathbb{C}$，则：$\psi(\boldsymbol{v},k\boldsymbol{w})=\left(a\overline{kc},\ a\overline{kd}+b\overline{kc},\ b\overline{kd}\right)=\bar{k}\cdot\left(a\bar{c},\ a\bar{d}+b\bar{c},\ b\bar{d}\right)=\bar{k}\cdot\psi(\boldsymbol{v},\boldsymbol{w})$（区别：双线性是 $k\cdot$，半双线性是 $\bar{k}\cdot$）

我们知道，内积不是线性变换，但是我们依然可以做一些操作，使得它也有所谓的“矩阵表示”：

内积是双线性/半双线性形式（输入两个向量，输出一个数），虽非线性变换（线性变换为“向量 $\to$ 向量”），但可通过基向量内积构造矩阵表示，将“双变量运算”转化为“矩阵乘法”计算。

设 V 为 n 维向量空间，$\mathcal{B}=\{v_1,v_2,\ldots,v_n\}$ 是 V 的一组基，$(\cdot,\cdot)$ 为 V 上的内积。

内积在基 $\mathcal{B}$ 下的矩阵表示 $M_{\mathcal{B}}$ 构造方式为：将基向量两两的内积结果排成 n 阶矩阵：

$$M_{\mathcal{B}} = \begin{pmatrix} (v_1,v_1) & (v_1,v_2) & \dots & (v_1,v_n) \\ (v_2,v_1) & (v_2,v_2) & \dots & (v_2,v_n) \\ \vdots & \vdots & \ddots & \vdots \\ (v_n,v_1) & (v_n,v_2) & \dots & (v_n,v_n) \end{pmatrix}.$$

对任意 $u,w \in V$，设 $u = a_1v_1 + a_2v_2 + \cdots + a_nv_n$，$w = b_1v_1 + b_2v_2 + \cdots + b_nv_n$，则 u,w 在基 $\mathcal{B}$ 下的坐标分别为：

$$[u]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad [w]_{\mathcal{B}} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

由内积的双线性/半双线性性质，展开得：

$$(u,w) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \cdot \overline{b_j} \cdot (v_i,v_j),$$

该双重求和等价于"w 坐标的共轭转置 $\times$ 内积矩阵 $M_{\mathcal{B}} \times u$ 的坐标"：

$$(u,w) = {}^t\overline{[w]_{\mathcal{B}}} \cdot M_{\mathcal{B}} \cdot [u]_{\mathcal{B}}.$$

（实数域下共轭可省略，即 $(u,w) = {}^t[w]_{\mathcal{B}} \cdot M_{\mathcal{B}} \cdot [u]_{\mathcal{B}}$）

证明. $M_{\mathcal{B}} \cdot [u]_{\mathcal{B}} = \begin{pmatrix} \sum_{i=1}^{n}(v_1,v_i)a_i \\ \sum_{i=1}^{n}(v_2,v_i)a_i \\ \vdots \\ \sum_{i=1}^{n}(v_n,v_i)a_i \end{pmatrix}$

将行向量的每个分量与列向量的对应分量相乘后求和：${}^t\overline{[w]_{\mathcal{B}}} \cdot (M_{\mathcal{B}} \cdot [u]_{\mathcal{B}}) = \sum_{k=1}^{n} \overline{b_k} \cdot (\sum_{i=1}^{n}(v_k,v_i)a_i)$

交换求和顺序将双重求和的顺序从"先对 i 求和，再对 k 求和"，改为"先对 k 求和，再对 i 求和"：$\sum_{k=1}^{n} \overline{b_k} \cdot (\sum_{i=1}^{n}(v_k,v_i)a_i) = \sum_{i=1}^{n} a_i \cdot \left(\sum_{k=1}^{n}(v_k,v_i)\overline{b_k}\right)$

利用内积的共轭对称性内积满足共轭对称性：$(v_k,v_i) = \overline{(v_i,v_k)}$，代入上式：$\sum_{i=1}^{n} a_i \cdot \left(\sum_{k=1}^{n} \overline{(v_i,v_k)} \cdot \overline{b_k}\right)$ 注意到"共轭的乘积 = 乘积的共轭"（即 $\overline{(v_i,v_k)} \cdot \overline{b_k} = \overline{(v_i,v_k) \cdot b_k}$），因此内层求和可写为：$\sum_{k=1}^{n} \overline{(v_i,v_k) \cdot b_k} = \overline{\sum_{k=1}^{n}(v_i,v_k) \cdot b_k}$

结合内积的半双线性性质根据内积对第二个变量的共轭线性，$\sum_{k=1}^{n}(v_i,v_k) \cdot b_k = (v_i, \sum_{k=1}^{n} b_kv_k) = (v_i,w)$，因此：$\overline{\sum_{k=1}^{n}(v_i,v_k) \cdot b_k} = \overline{(v_i,w)}$

最终化简将其代回原式，再利用内积对第一个变量的线性，展开整个求和：$\sum_{i=1}^{n} a_i \cdot \overline{(v_i,w)} = (\sum_{i=1}^{n} a_iv_i, w) = (u,w)$

□

示例：取 $\mathbb{R}^2$ 的标准基 $\mathcal{B}=\{\vec{e}_1=(1,0),\vec{e}_2=(0,1)\}$，点积的矩阵表示为：

$$M_{\mathcal{B}}=\begin{pmatrix}(\vec{e}_1,\vec{e}_1) & (\vec{e}_1,\vec{e}_2)\\ (\vec{e}_2,\vec{e}_1) & (\vec{e}_2,\vec{e}_2)\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

计算 $u=(2,3)$、$w=(4,5)$ 的点积：$[u]_{\mathcal{B}}=\begin{pmatrix}2\\3\end{pmatrix}$，$[w]_{\mathcal{B}}=\begin{pmatrix}4\\5\end{pmatrix}$，代入得：

$${}^t[w]_{\mathcal{B}}\cdot M_{\mathcal{B}}\cdot[u]_{\mathcal{B}}=(4,5)\cdot\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\cdot\begin{pmatrix}2\\3\end{pmatrix}=4\times2+5\times3=23,$$

与直接计算点积结果一致。

设 $\mathcal{A}=\{w_1,w_2,\cdots,w_n\}$ 和 $\mathcal{B}=\{v_1,v_2,\cdots,v_n\}$ 是 V 的两组基。记 $C_{\mathcal{B},\mathcal{A}}\in M_n(\mathbb{F})$ 为从基 $\mathcal{A}$ 到基 $\mathcal{B}$ 的基变换矩阵，即满足：

$$C_{\mathcal{B},\mathcal{A}}[v]_{\mathcal{A}}=[v]_{\mathcal{B}},\quad \forall v\in V.$$

设 $(\cdot,\cdot)$ 在基 $\mathcal{B}$ 和 $\mathcal{A}$ 下的矩阵表示分别为 $M_{\mathcal{B}}$ 和 $M_{\mathcal{A}}$，则对任意 $v,w\in V$，有：

$$(v,w)={}^t[w]_{\mathcal{A}}M_{\mathcal{A}}[v]_{\mathcal{A}}={}^t[w]_{\mathcal{B}}M_{\mathcal{B}}[v]_{\mathcal{B}}={}^t[w]_{\mathcal{A}}\,{}^t\overline{C_{\mathcal{B},\mathcal{A}}}M_{\mathcal{B}}C_{\mathcal{B},\mathcal{A}}[v]_{\mathcal{A}}.$$

因此得到**基变换公式**：

$$M_{\mathcal{A}}={}^t\overline{C_{\mathcal{B},\mathcal{A}}}M_{\mathcal{B}}C_{\mathcal{B},\mathcal{A}}.$$

受上述定义的启发，称域 $\mathbb{F}$ 上的两个矩阵 $A,B\in M_n(\mathbb{F})$ 是相合的（相应地，共轭相合的），若存在可逆矩阵 $P\in\mathrm{GL}_n(\mathbb{F})$，使得：

$$A=PB{}^tP\quad(\text{相应地，}\quad A=PB{}^t\overline{P}).$$

记为 $A\overset{\text{相合}}{\sim}B$（相应地，$A\overset{\text{共轭相合}}{\sim}B$）。