

线性代数 A 习题课讲义

原生生物

* 线性代数 A [赵玉凤老师班] 习题课讲义，前十三章在上册习题课讲义中，本讲义中不会引用只在上册习题课讲义里出现的知识，但可能会有逻辑上的类似参考。

* 配套主要教材为丘维声《高等代数》下册，除此以外课程还使用了王萼芳、石生明《高等代数》，个人还参考了王新茂《线性代数讲义》。

* 记号约定：上标 T 表示转置， H 表示取转置并对每个元素取共轭。 O 表示零矩阵， I_a 表示 a 阶单位阵 (\mathcal{I} 为恒等映射 $\mathcal{I}x = x$)，下标 $m \times n$ 表示矩阵的行列数， \det 表行列式， diag 表将后方元素拼到主对角线上形成的对角阵。 e_i 表示第 i 个分量为 1，其他为 0 的单位列向量。

目录

| | |
|--------------------------------|-----------|
| 十四 补充：写在正式开始之前 | 6 |
| §14.1 多项式的高观点 | 6 |
| 14.1.1 带余除法 | 6 |
| 14.1.2 等价性与序关系 | 7 |
| 14.1.3 辗转相除法 | 9 |
| 14.1.4 欧几里得环 | 11 |
| §14.2 高观点与低观点 | 15 |
| 14.2.1 教学方式的权衡 | 15 |
| 14.2.2 作为矩阵论高观点的线性空间 | 16 |
| 14.2.3 做题指南 | 18 |
| 14.2.4 记忆关键结论？ | 18 |
| 十五 从同余方程组说起 | 19 |
| §15.1 再谈数论 | 19 |
| 15.1.1 线性方程组？ | 19 |
| 15.1.2 模的构造尝试 | 20 |
| 15.1.3 行列变换的思路 | 23 |
| §15.2 整数方阵的相抵 | 24 |
| 15.2.1 可逆性 | 24 |
| 15.2.2 两种特殊情况 | 26 |
| 15.2.3 构造性证明 | 28 |
| 15.2.4 多项式矩阵 | 29 |
| §15.3 Smith 标准形 | 30 |
| 15.3.1 另一种构造方式 | 30 |
| 15.3.2 行列式因子 | 32 |
| 15.3.3 不变因子 | 34 |

| | |
|------------------------------|-----------|
| 十六 补充：复数域的相似标准形 | 35 |
| §16.1 特征方阵 | 35 |
| 16.1.1 特征方阵的相抵 | 35 |
| 16.1.2 带余除法出发的证明 | 36 |
| 16.1.3 计算系数出发的证明 | 39 |
| §16.2 Jordan 标准形 | 41 |
| 16.2.1 初等因子组 | 41 |
| 16.2.2 Jordan 块 | 44 |
| 16.2.3 矩阵的多项式 | 46 |
| 16.2.4 可对角化 IV | 48 |
| 十七 走向抽象 | 49 |
| §17.1 习题解答 | 49 |
| §17.2 有理标准形 | 55 |
| 17.2.1 一般数域上的相似 | 55 |
| 17.2.2 一般数域上的相似标准形 | 56 |
| 17.2.3 “表示” | 58 |
| §17.3 期末考试 | 59 |
| 17.3.1 试题 | 59 |
| 17.3.2 解答 | 60 |
| §17.4 期末的延伸 | 63 |
| 17.4.1 加法、数乘 | 63 |
| 17.4.2 整体性 | 67 |
| 17.4.3 最小二乘结构 | 70 |
| 十八 补充：直和与空间分解 | 72 |
| §18.1 期中前整体规划 | 72 |
| §18.2 维数与坐标 | 73 |
| 18.2.1 基本结论 | 73 |
| 18.2.2 例题 | 76 |
| 18.2.3 解答 | 78 |
| §18.3 直和 | 83 |
| 18.3.1 交空间与和空间 | 83 |
| 18.3.2 直和与补空间 | 86 |
| 18.3.3 例题 | 89 |
| 18.3.4 解答 | 90 |
| §18.4 商空间 | 94 |
| 18.4.1 基本结论 | 94 |
| 18.4.2 例题 | 96 |
| 18.4.3 解答 | 96 |
| 十九 线性映射 | 97 |
| §19.1 习题解答 | 97 |
| §19.2 同态与同构 | 101 |
| 19.2.1 集合的同态 | 101 |

| | |
|-----------------------------|------------|
| 19.2.2 线性映射 | 104 |
| 19.2.3 构造双射的方式 | 109 |
| 19.2.4 限制映射 I | 111 |
| 19.2.5 第一同构定理 | 112 |
| §19.3 矩阵表示 | 115 |
| 19.3.1 同构标准形 | 115 |
| 19.3.2 坐标的线性映射 | 116 |
| 19.3.3 线性映射的运算 | 120 |
| 19.3.4 基变换与相抵 | 123 |
| 二十 补充：从映射到变换 | 125 |
| §20.1 一般线性映射 | 125 |
| 20.1.1 例题 | 125 |
| 20.1.2 解答 | 126 |
| §20.2 有限维空间线性映射 | 131 |
| 20.2.1 限制映射 II | 131 |
| 20.2.2 投影映射 | 133 |
| 20.2.3 例题 | 135 |
| 20.2.4 解答 | 136 |
| §20.3 线性变换 | 142 |
| 20.3.1 多项式与特征系统 | 142 |
| 20.3.2 矩阵表示与基变换 | 143 |
| 20.3.3 化零多项式 | 145 |
| 二十一 相似标准形再探 | 148 |
| §21.1 习题解答 | 148 |
| §21.2 定义准备 | 153 |
| 21.2.1 限制映射 III | 153 |
| 21.2.2 根子空间与分解 | 156 |
| 21.2.3 循环子空间 | 161 |
| §21.3 标准形理论 | 161 |
| 21.3.1 可对角化 V | 161 |
| 21.3.2 循环子空间分解 | 161 |
| 21.3.3 Jordan 标准形 | 161 |
| 21.3.4 有理标准形 | 161 |
| 二十二 补充：期中复习 | 161 |
| §22.1 知识与技巧整理 | 161 |
| 22.1.1 上册知识 | 162 |
| 22.1.2 线性空间 | 164 |
| 22.1.3 线性映射 | 165 |
| 22.1.4 线性变换 | 166 |
| 22.1.5 多项式矩阵 | 168 |
| §22.2 复习题 | 169 |
| §22.3 解答 | 170 |

| | |
|-----------------------------|------------|
| 二十三 有关证明 | 185 |
| §23.1 期中考试 | 185 |
| 23.1.1 试题 | 185 |
| 23.1.2 解答 | 186 |
| §23.2 逻辑与证明 | 190 |
| 23.2.1 什么是证明 | 190 |
| 23.2.2 不完备性 | 191 |
| 23.2.3 形式化证明 | 191 |
| 二十四 补充：双线性函数简介 | 191 |
| §24.1 双线性性 | 191 |
| 24.1.1 动机与定义 | 191 |
| 24.1.2 非退化性 | 192 |
| 24.1.3 限制映射与同构定理 | 194 |
| 24.1.4 双线性函数空间 | 196 |
| §24.2 度量矩阵 | 199 |
| 24.2.1 向量空间的双线性函数 | 199 |
| 24.2.2 度量矩阵 | 201 |
| 24.2.3 基变换与相抵 | 203 |
| §24.3 对称与斜对称 | 204 |
| 24.3.1 转置 | 204 |
| 24.3.2 对称性与合同 | 205 |
| 24.3.3 补与补空间 | 210 |
| 二十五 实内积空间 | 214 |
| §25.1 习题解答 | 214 |
| §25.2 内积 | 218 |
| 25.2.1 实内积空间 | 218 |
| 25.2.2 标准正交基 | 222 |
| §25.3 实内积空间的同构 | 227 |
| 25.3.1 线性映射的矩阵表示 | 227 |
| 25.3.2 等距映射 | 227 |
| 25.3.3 同构 | 227 |
| §25.4 内积下的线性变换 | 227 |
| 25.4.1 伴随变换 | 227 |
| 25.4.2 正规变换 | 228 |
| 二十六 补充：一般内积空间 | 228 |
| §26.1 实内积空间综合 | 228 |
| 26.1.1 例题 | 228 |
| 26.1.2 解答 | 229 |
| §26.2 复空间 | 238 |
| 26.2.1 复内积 | 238 |
| 26.2.2 内积下的线性映射 | 238 |
| 26.2.3 内积下的线性变换 | 238 |

| | |
|-----------------------------|------------|
| 二十七 张量代数 | 238 |
| §27.1 多重线性函数 | 238 |
| 27.1.1 对偶空间 | 238 |
| 27.1.2 多重线性性 | 241 |
| 27.1.3 多重线性函数的张量积 | 244 |
| §27.2 线性空间的张量积 | 246 |
| 27.2.1 张量积定义 | 246 |
| 27.2.2 向量空间的张量积 | 247 |
| 27.2.3 张量积的一般性质 | 249 |
| 27.2.4 泛性质定义 | 251 |
| §27.3 线性映射的张量积 | 253 |
| 27.3.1 良好定义性 | 253 |
| 27.3.2 克罗内克积 | 254 |
| 二十八 补充：期末复习 | 258 |
| §28.1 知识与技巧整理 | 258 |
| 28.1.1 上册知识 | 258 |
| 28.1.2 内积空间 | 259 |
| 28.1.3 张量 | 261 |
| §28.2 复习题 | 261 |
| §28.3 解答 | 264 |

十四 补充：写在正式开始之前

§14.1 多项式的高观点

听了几节赵老师的课之后，我个人对赵老师上课风格的理解是，适合睡觉以低观点为主，伴有大量的计算进行推进。笼统来说，低观点是以更顺序性的结构安排课程，一条条写出定义、定理并逐步推进，而高观点则是以树状的结构进行安排，通过对某个问题（或本质概念）的不断分层、剖析、延伸引出知识。由于教材也基本是低观点的思路，我们先以高观点重新组织“多项式的性质”这一讲，来展示何为高观点。

* 看到看不下去的时候就可以直接跳转本讲义 14.2 了，这个作为高观点例子的讲解里几乎没有考试会涉及的部分，本讲义 14.2.1 开头也将所有本节需要了解的结论列举了。

14.1.1 带余除法

对于任何一个数域 \mathbb{K} ，我们定义 \mathbb{K} 上的 x 的多项式集合 $\mathbb{K}[x]$ 为

$$\left\{ \sum_{k=0}^m a_k x^k \mid a_k \in \mathbb{K} \right\}$$

这里 x 为一个形式变元——目前暂且理解为它与函数的自变量不完全等价即可，一个简单的例子是，在多项式除法的意义下 x^2 除以 x 等于 x ，而在函数意义下必须定义在非零元素上才满足。

在中学，我们已经学到了数域上的多项式如何进行加法、减法、乘法，且由于系数只进行相加/相减/相乘，两个 $\mathbb{K}[x]$ 中多项式的运算结果仍然在 $\mathbb{K}[x]$ 中。

然而，与数不同的是，两个多项式的除法往往是没有意义的——更准确来说， $\mathbb{K}[x]$ 对除法不封闭。例如，虽然以函数的视角， $\frac{x^2}{x+1}$ 是一个合理的函数，但由于它不再是多项式，讨论 $\mathbb{K}[x]$ 时并没有谈论的意义。

事实上，我们早在小学就已经碰到过这样无法进行除法的结构，那时我们的写法是

$$7 \div 3 = 2 \cdots \cdots 1$$

即使更之后学习分数后，这样的写法仍然有意义：因为它保证了问题的讨论限制在整数中，而不需要引入一套新的数。

当然，用省略号表示余数并不是一个好记号，不过我们姑且给这种方式一个更形式性的说法（这里我们允许被除数或除数为负数）：对整数 $a, b \neq 0$ ，记它们的带余除法

$$a \div b = q \cdots \cdots r$$

当且仅当 $a = bq + r$ 且 $0 \leq r \leq |b| - 1$, $q, r \in \mathbb{Z}$ ，其中 q 称为商， r 称为余数。

商和余数的存在性与唯一性是容易证明的，这里简单介绍：对于存在性， b 为正时记 q 为使得 $bq \leq a$ 的最大 q 、 b 为负时记 q 为使得 $bq \leq a$ 的最小 q ，并记 $r = a - bq$ 可验证满足要求；而 $bq_1 + r_1 = bq_2 + r_2$ 可以推出 $b(q_1 - q_2) = r_2 - r_1$ ，从而根据 r 的大小范围只能得到两侧均为 0。

从这段推理可以看出，带余除法定义里最关键的是余数大小的部分，它保证了余数的规模一定低于原问题中除数的规模。若 a 除以 b 余数为 0，我们称 b 整除 a ，也即 a 是 b 的倍数，记作

$$b \mid a$$

* 一点 LATEX 相关提示：这个整除符号并不是键盘上直接打竖线，而是通过 \mid 打出的。

* 根据定义，任何非零数都整除 0，这个性质我们在之后还将用到。

对于多项式，我们如果希望能类似定义这样的除法，首先需要的是进行大小的刻画。很显然，一个多项式的次数是一个不错的定义手段。我们将多项式 $f(x) \in \mathbb{K}[x]$ 的次数最高的非零项（称为首项）的次数定义为其次数 $\deg f(x)$ ，并规定零多项式次数为负无穷。这样，一个容易得到的计算结论是

$$\deg f(x) + \deg g(x) = \deg f(x)g(x)$$

由此也可以看出所谓的整环性质：若 $f(x)g(x) = 0$ ，则 $f(x) = 0$ 或 $g(x) = 0$ 。此外，我们还有

$$\deg(f(x) \pm g(x)) \leq \max\{\deg f(x), \deg g(x)\}$$

接下来，我们希望看到带余除法是否可以进行，仿照之前，我们需要证明，对多项式 $a(x) \in \mathbb{K}[x]$ 、 $0 \neq b(x) \in \mathbb{K}[x]$ ，存在唯一的商式 $q(x) \in \mathbb{K}[x]$ 与余式 $r(x) \in \mathbb{K}[x]$ 使得

$$a(x) = b(x)q(x) + r(x), \quad \deg r(x) < \deg b(x)$$

类似教材，我们可以直接通过长除法进行操作性的证明，我们这里给出一个更严谨一点的办法，注意和整数带余除法的存在唯一性证明类比，最核心的思路是利用余式的次数减小性：

证明：

先证明存在性。取 $q(x)$ 为使得 $a(x) - b(x)q(x)$ 次数最小的多项式（的其中一个），由于次数可能性有限，这样的 $q(x)$ 一定存在。下记 $r(x) = a(x) - b(x)q(x)$ ，我们证明 $\deg r(x) < \deg b(x)$ 即可。

若否，设 $r(x)$ 首项为 sx^m 、 $b(x)$ 首项为 tx^n ，由假设 $m \geq n$ 且 $st \neq 0$ 。取

$$q_0(x) = q(x) + \frac{s}{t}x^{m-n}$$

则直接计算可知

$$r_0(x) = a(x) - b(x)q_0(x) = r(x) - \frac{s}{t}x^{m-n}b(x)$$

由于 $\frac{s}{t}x^{m-n}b(x)$ 也为 m 次多项式且 m 次项系数与 $r(x)$ 相同， $r_0(x)$ 为至多 $m-1$ 次多项式，次数小于 $r(x)$ ，与 $q(x)$ 使得 $q(x) - b(x)q(x)$ 次数最小矛盾。

再证明唯一性。若有 $a(x) = b(x)q_1(x) + r_1(x) = b(x)q_2(x) + r_2(x)$ 均符合要求，有

$$b(x)(q_1(x) - q_2(x)) = r_2(x) - r_1(x)$$

若左侧非零，其次数至少为 $\deg b(x)$ ，与右侧次数低于 $\deg b(x)$ 矛盾。

当然，有了带余除法后，若余式是 0，我们仍然定义整除，也仍然记作

$$b(x) | a(x)$$

同样，任何非零多项式都整除 0。

* 一个更好的整除定义是存在 $q(x) \in \mathbb{K}[x]$ 使得 $a(x) = b(x)q(x)$ ，此定义与之前定义的区别是，由于不再利用除法，可以将 0 写在整除式的左端。从此定义可以直接得到， $0 | f$ 当且仅当 $f = 0$ ，对于整数，我们此后同样应用这个扩充的定义。

14.1.2 等价性与序关系

到此处，我们已经发现了整数和多项式的类似性质，之后无歧义时仍将 $f(x)$ 记为 f 。在进入公因数/公因式的讨论前，还有三件事需要处理：

- 如果我们只考虑整除相关的问题，整数与多项式都会存在某种等价性。具体来说，如果两个整数只相差符号，它们对应的整除性质应该是等价的（即，一个数是 3 的倍数/因数那么它一定是 -3 的倍数/因数，反之亦然）；对于两个多项式 $f, g \in \mathbb{K}[x]$ ，如果存在非零的 $t \in \mathbb{K}$ 使得 $f(x) = tg(x)$ ，则 f, g 的整除性质也是等价的。

证明：

我们以多项式为例进行证明：若 $g(x) | h(x)$ ，则 $h(x) = g(x)u(x)$ ，于是 $h(x) = f(x)(t^{-1}u(x))$ ，从而 $f(x) | h(x)$ ；若 $f(x) | h(x)$ ，则 $h(x) = f(x)v(x)$ ，于是 $h(x) = g(x)(tv(x))$ ，从而 $g(x) | h(x)$ ；若 $l(x) | f(x)$ 或 $l(x) | g(x)$ ，完全类似证明也可。

从证明过程可以看出，这样的等价性成立本质上是因为 t^{-1} 存在。整数或多项式（下统一记为 R ，注意与表示实数的双线体 \mathbb{R} 区分）中的可逆元素称为它的单位。具体来说，若 $a \in R$ 满足存在 $b \in R$ 使得 $ab = 1$ ，则称 a 为 R 的单位。

对整数而言，其与另一个整数乘积为 1 意味着它只能是 ± 1 ，称两个整数 a, b 等价当且仅当存在单位 t 使得 $a = tb$ ，即 $a = \pm b$ ，称 a 的等价类为所有与 a 等价的整数的集合，称等价类中的代表元为等价类中的非负整数（根据定义课直接看出存在唯一）；而对多项式集合 $\mathbb{K}[x]$ ，考虑次数可以发现其中可逆元只能是 \mathbb{K} 中的非零元素（作为零次多项式），称两个多项式 a, b 等价当且仅当存在单位 t 使得 $a = tb$ ，称 a 的等价类为所有与 a 等价的多项式的集合，称等价类中的代表元为等价类中的首项系数为 1 的多项式（称为首一多项式）——当然，任何非零多项式都能除以首项系数成为首一多项式，且结果唯一，对于零多项式，它的等价类中只有 0，因此 0 自身就是代表元。为了方便之后的讨论，我们也将 0 视为首一多项式。

由于 1 是单位、单位的逆是单位、两个单位的乘积还是单位，利用上学期知识可以验证这里定义的等价确实符合上学期所说的等价关系，因此，这里等价类和代表元的定义是合理的，代表元也符合我们“选出尽量形式简单的元素以表示”的要求。由于若 $b | a$ ，将 a, b 分别替换为等价类中的任何一个元素都不影响结果，整除性可以针对等价类进行讨论。

2. 对于整数或是多项式，就像之前已经定义的，我们可以用绝对值/次数去进行某种意义的大小比较。不过，这样的大小比较忽略了它们的数论意义（也就是带余数法相关的性质）。相对地，我们在之前定义了整除，如果它可以作为某种大小比较，问题就迎刃而解了。

回顾集合 A 上的关系是指 $A \times A$ 的一个子集 S ，若 $(x, y) \in S$ 则称 x 与 y 具有关系 S ，记为 xSy ，否则称不具有。我们定义一个关系为序关系当且仅当它满足：

- **自反性：** 对任何 $a \in A$ 有 aSa ；
- **斜对称性：** 对任何 $a, b \in A$ ，若 aSb 且 bSa ，则有 $a = b$ ；
- **传递性：** 对任何 $a, b, c \in A$ ，若 aSb 且 bSc ，则有 aSc 。

最简单的例子是，对于实数集，大于等于、小于等于都是序关系。对于整数或是多项式，整除非常接近序关系：可以直接验证自反性与传递性都满足，但若 $a | b$ 、 $b | a$ ，未必能直接得到 $a = b$ ，例如对整数而言取 $a = 2$ 、 $b = -2$ 。

好在，我们讨论的第一件事已经解决了这个问题。如果我们将整除定义在等价类上，从 $a | b$ 与 $b | a$ 确实能得到 a 与 b 等价：设 $a = ub$ 、 $b = va$ ，若 $a = 0$ ，则 $b = 0$ ，从而等价，否则 $a = uva$ ，于是 $uv = 1$ ，即得到 u, v 都是单位。

某种意义上来说，数论就是在研究整除性这样一个**定义在等价类上的序关系**。值得注意的是，并不是任何两个元素一定存在整除性，因此整除这个“比较”无法对任何两个元素进行，这与任何两个实数都可以分出大小有本质的不同。

3. 就像实数在有了 \leq, \geq 后可以比较大小，在集合 A 上定义了序关系 S 后，我们可以进一步给出如下的定义：

- 对 A 的子集 A_0 ，称它的一个上界为满足 $\forall a \in A_0, aSx$ 的 x 的集合；
- 对 A 的子集 A_0 ，称它的一个下界为满足 $\forall a \in A_0, xSa$ 的 x 的集合；
- 对 A 中的一些元素 a_1, \dots, a_k ，称它们的上/下界为集合 $\{a_1, \dots, a_k\}$ 的上/下界；
- 对 A 的子集 A_0 ，若存在 $a \in A_0$ 为 A_0 的上界，则称它为 A_0 中的最大元素（利用斜对称性，最大元素至多只有一个）；

- 对 A 的子集 A_0 , 若存在 $a \in A_0$ 为 A_0 的下界, 则称它为 A_0 中的**最小元素** (利用斜对称性, 最小元素至多只有一个);
- 对 A 的子集 A_0 , 若存在 $a \in A_0$ 使得对任何 A_0 中其他元素 a' 均没有 aSa' , 则称它为 A_0 中的**极大元素**;
- 对 A 的子集 A_0 , 若存在 $a \in A_0$ 使得对任何 A_0 中其他元素 a' 均没有 $a'Sa$, 则称它为 A_0 中的**极小元素**。

我们用整数中以整除作为序关系为例进行一些简单的研究 (注意将等价类视为整体):

- 由于 $b | a$ 也即 a 是 b 的倍数, 整数集合的上界事实上就是它的**公倍数集合**;
- 由于 $b | a$ 也即 b 是 a 的因数, 整数集合的上界事实上就是它的**公因数集合**;
- 当且仅当整数集合 A_0 满足其中所有数都是其中某个数的因数时, 其最大元素存在, 有 0 的整数集合**最大元素一定为 0**;
- 当且仅当整数集合 A_0 满足其中所有数都是其中某个数的倍数时, 其最小元素存在, 有 ± 1 的整数集合**最小元素一定为 ± 1** ;
- 只要整数集合中的某个元素任何倍数都不在其中, 其即为极大元素;
- 只要整数集合中的某个元素任何倍数都不在其中, 其即为极小元素。

根据定义可以发现, 一些整数可以有很多个上界或下界, 而为了刻画这些整数真正意义上的“范围”, 就像实数集合的上下确界一样, 我们希望找到上界中**最小的一个**、下界中**最大的一个**。根据定义, 这就是**最小公倍数与最大公因数**。完全类似地, 对于一些多项式, 我们将整除关系下的最小上界称为**最小公倍式**, 最大下界称为**最大公因式**。由于整除性的定义都是在等价类上, 最大公因数/式与最小公倍数/式都是一个等价类, 因此我们取出其中的**代表元**——非负整数或首一多项式——作为确定的结果。

* 将上述序关系的语言改为整除语言, 以多项式为例, 我们可以得到, 称 d 为 f, g 的**最大公因式** (所在等价类), 当且仅当 $d | f, d | g$, 且对任何 d_0 满足 $d_0 | f, d_0 | g$, 都有 $d_0 | d$ 。另一个常见定义是, d 是 f, g 所有公因式中次数最大的。我们之后将证明, 这两个定义事实上基本是等价的 (但在 0 的情况下有一定区别, 见后文)。第二个定义的好处是可以直接看出存在性, 但其实对于“最大”的含义有一定的误导作用, 因为这里的大小比较并不是指次数的大小。

* 由此, 最大公因数/式与最小公倍数/式是至关重要的, 因为它们刻画了整除关系下一个集合的“规模”。相比之下, 我们更喜欢研究最大公因数/式, 因为它在**更小的一端**, 可以缩减问题规模。

为了方便之后的讨论, 我们还需要定义另外一个关系: 对一些集合, 当且仅当 $A \subset B$ 时称 A 与 B 有关系, 可以验证这也构成一个序关系, 同样, 不是任何两个集合都存在包含关系。我们大部分时候对集合“大小”的谈论都是基于这个序关系, 例如以下两个命题:

- V 中的一个向量组 S 的**生成子空间**是指所有包含 S 的 V 的子空间中的**最小元素**;
- 一个向量集合 S 的**极大线性无关组**是指所有 S 的线性无关子集中的**极大元素**。

14.1.3 辗转相除法

在经历了一大堆看似复杂的定义和讨论后, 我们得到了一些“能用”的标准定义: 对于一些整数 a_1, \dots, a_m , 它们的**最大公因数**是指 (非负整数) 公因数 d 使得对任何公因数 d_0 都有 $d_0 | d$, 记作

$$\gcd(a_1, \dots, a_m)$$

最小公倍数是指 (非负整数) 公倍数 m 使得对任何公倍数 m_0 都有 $m \mid m_0$, 记作

$$\text{lcm}(a_1, \dots, a_m)$$

我们用以下的思路说明此定义的合理性 (注意唯一性已经由最大元素的定义通过序关系的斜对称性保证了, 因此只需说明存在性), 下文的字母无特殊说明时表示整数:

1. $\gcd(0, a) = |a|$

证明:

当 $a \neq 0$ 时, 由于任何数都为 0 的因数, 公因数集合即为 a 的因数集合, 而根据定义可知非负整数 $|a|$ 是 a 的因数, 且任何 a 的因数都是 $|a|$ 的因数, 于是 $|a|$ 是它们的最大公因数。

当 $a = 0$ 时, 公因数集合为 \mathbb{Z} , 其中包含 0, 且 0 是任何整数的倍数, 于是 $\gcd(0, 0) = 0$ 。

* 此处展现了最大公因数的“最大”与通常意义下大小关系的区别: **0** 是整除意义下的最大元素, 而模长意义上其应为最小元素。

2. a_1, a_2 的公因数集合与 $a_1, a_2 + ka_1$ 的公因数集合相同

证明:

若 $x \mid a_1, x \mid a_2$, 设 $a_1 = k_1x, a_2 = k_2x$, 可得 $a_2 + ka_1 = (k_2 + kk_1)x$, 于是是 x 的倍数; 由于 $a_2 = (a_2 + ka_1) - ka_1$, 在上一句证明中将 k 改为 $-k$ 可完全类似证出另一边。

3. $\gcd(a_1, a_2)$ 存在

证明:

我们用 $\text{cd}(a_1, a_2)$ 表示两者的公因数集合。 a_1, a_2 有 0 的情况已在之前解决, 此处不妨设两者均非零, 且由于乘单位不影响整除性, $\text{cd}(a_1, a_2) = \text{cd}(|a_1|, |a_2|)$, 由此可不妨设 a_1, a_2 为正整数。不妨设 $a_1 \geq a_2$, 则根据之前已证有 $\text{cd}(a_1, a_2) = \text{cd}(a_1 - ka_2, a_2)$, 取 k 为 a_1 除以 a_2 的商, 则得到的 $a_1 - ka_2$ 为 a_1 除以 a_2 的余数, 于是其小于 a_2 , 记 $a'_1 = a_2, a'_2 = a_1 - ka_2$ 。由此不断操作下去, 根据余数的性质可知每次较小数都在减小, 但过程中只会出现非负整数, 因此至多减少 a_2 次后较小数将成为 0, 由此可知一定存在某个非负整数 d 使得 $\text{cd}(a_1, a_2) = \text{cd}(d, 0)$ 。根据之前已证, d 就是两者的最大公因数。

4. $\gcd(a_1, a_2, a_3) = \gcd(\gcd(a_1, a_2), a_3)$

证明:

根据定义, $\gcd(\gcd(a_1, a_2), a_3)$ 一定是 $\gcd(a_1, a_2)$ 与 a_3 的公因数, 而根据定义可发现 $\gcd(a_1, a_2)$ 的因数一定还是 a_1, a_2 的公因数, 因此右侧确为 a_1, a_2, a_3 的公因数 (且根据 \gcd 定义为非负整数), 下证最大性。

考虑 a_1, a_2, a_3 的任何一个公因数 m , 由于其为 a_1, a_2 的公因数有 $m \mid \gcd(a_1, a_2)$, 且 $m \mid a_3$, 再次根据最大公因数的定义可得 $m \mid \gcd(\gcd(a_1, a_2), a_3)$, 从而得证。

5. 有限个元素的最大公因数存在

证明:

类似上方即可证明

$$\gcd(a_1, \dots, a_n) = \gcd(\gcd(a_1, \dots, a_{n-1}), a_n)$$

再对 n 归纳可得存在。

6. $\text{lcm}(0, a) = 0$

证明：

由于 0 的倍数只有 0，而 0 是 a 的倍数，两者的公倍数只有 0，从而其为最小公倍数。

7. $\text{lcm}(a_1, a_2)$ 存在

证明：

当 a_1, a_2 有 0 时已经证明，否则不妨设两者均为正整数。

由于 a_1, a_2 的公倍数集合非空，且至少有一个正整数 $|a_1 a_2|$ ，其中一定存在最小的正整数 m ，下面证明这即为 $\text{lcm}(a_1, a_2)$ 。若有另一个公倍数 l 不是 m 的倍数，类似第二部分可发现对任何整数 q ， $r = l - qm$ 都是 a_1, a_2 的公倍数，于是取 q 为 l 除以 m 的商，则 r 为余数，根据假设可知 $0 < r < m$ ，与 m 是其中的最小正整数矛盾。

8. $\text{lcm}(a_1, a_2, a_3) = \text{lcm}(\text{lcm}(a_1, a_2), a_3)$

证明：

与第四部分完全类似，根据定义可知右侧确为 a_1, a_2, a_3 的公倍数，而由于三者的公倍数一定是 $\text{lcm}(a_1, a_2)$ 与 a_3 的倍数，再由最小公倍数定义得成立。

9. 有限个元素的最小公倍数存在

证明：

与第五部分完全类似，利用

$$\text{lcm}(a_1, \dots, a_n) = \text{lcm}(\text{lcm}(a_1, \dots, a_{n-1}), a_n)$$

归纳即可。

注意到这里的绝对值事实上为在等价类中取代表元，对于多项式而言，将绝对值符号理解为取其中的首一多项式即可类似证明完全相同的结论。

此外，上述证明中不但说明了存在性，还给出了最大公因数的一个有效算法。然而，对最小公倍数，我们的证明方式无法给出有效的算法，这将在下一部分解决。

14.1.4 欧几里得环

在刚才的证明中，我们可以发现，整数和域上的多项式在数论性质上具有非常多的共通之处：从带余除法、整除性到最大公因数、最小公倍数，很多理论都可以类似应用。这启发我们，存在一个更本质的代数结构能同时刻画两者——而细究前面的所有定义，可以发现从带余除法出发能推出所有其他性质，因此定义的关键就是刻画何为带余除法。

观察可以发现，整数与多项式都能进行加、减、乘（乘法可交换），包含加法单位元 0 与乘法单位元 1，且两个元素乘积是 0 当且仅当其中某个是 0。这样的代数结构称为整环。更进一步地，可以给整数或多项式的每个元素 x 赋予一个非负整数 $\delta(x)$ ，使得对任何元素 a 与非零元素 b ，均存在元素 q, r 使得

$$a = bq + r, \quad \delta(r) < \delta(b)$$

对于整数，定义这里的 δ 为绝对值即可，而对多项式则稍微麻烦一点：定义 $\delta(0) = 0$ ，其他多项式的 δ 为次数加 1 即可（本质上仍然是对次数的刻画，只是要把 $\deg 0 = -\infty$ 这种特殊情况囊括到定义中）。

我们把满足上述要求的代数结构称为欧几里得环。当然，之后我们的讨论仍然只针对整数与多项式进行，且仍以整数为例。不过，感兴趣的同學可以自行将证明复刻到数域上的多项式中，乃至推广到所有欧几里得环中（注意复刻/推广的过程中要考虑乘单位对应的等价类）。

对于一个整环 D ，我们经常会关心它的理想。具体定义为，若 D 的子集 I 满足

$$\forall a, b \in I, \quad a + b \in I$$

$$\forall a \in I, r \in D, \quad ra \in I$$

则称它为一个理想。直观来看，这个定义与向量空间的定义有诸多相似处：都是要求对加法与“数乘”具有封闭性。在本讲义第十五章中，我们将对这个结构进行更多推广，成为所谓的模，并讨论一些性质。

很遗憾，我们无法直接从定义看出理想这个东西到底有多“理想”才让数学家们这么命名。不过，本部分剩下的内容里，我们将看到用理想表述最大公因数、最小公倍数这些概念有多么方便。这里，我们先证明一个很有意思的定理：对整数集合的任何一个理想 I ，存在非负整数 n 使得

$$I = \{nz \mid z \in \mathbb{Z}\}$$

我们经常将右侧的集合记作 $n\mathbb{Z}$ ，表示 n 的所有倍数——由于 n 的倍数求和或乘整数还是 n 的倍数， $n\mathbb{Z}$ 的确为一个理想。

证明：

若 I 中只包含 0，则其为 $\{0\} = 0\mathbb{Z}$ ，已经得证。否则，其中至少有一个非零整数 k ，取 r 为 k 的符号可知正整数 $|k| = rk \in I$ 。由此，取出 I 中的最小正整数 n ，若 I 中有某个不是 n 的倍数的数 l ，根据定义可知 $l - qn \in I$ 对任何 q 成立，而取 q 为 l 除以 n 的商可使余数 $l - qn$ 在 0 与 n 之间，与 n 是最小正整数矛盾。此外，由于 $n \in I$ ，其任何倍数都可以写为 rn ，于是 $n\mathbb{Z} \subset I$ ，两者结合即得到 $I = n\mathbb{Z}$ 。

* 我们给这个有趣但有点莫名其妙的定理添加一些注释：

- 首先，定理的证明与本讲义 14.1.3 的第七部分几乎相同，且这并不是偶然。直接根据定义验证可以发现，对任何 a_1, a_2 ，它们的公倍数集合构成了一个理想，因此确定这个理想到底是谁的倍数的过程应与找到 n 的过程相同。
- 其次，对数域上的多项式，这个定理的表述为： $\mathbb{K}[x]$ 的任何理想一定是 $\mathbb{K}[x]$ 中某个首一多项式的全部倍数的集合。同样，设这个多项式为 $f(x)$ ，可以记作 $f(x)\mathbb{K}[x]$ 。
- 事实上，整环中，某个元素的全部倍数集合一定是理想（直接验证即可），这样的理想称为主理想。上述定理意味着，欧几里得环中所有的理想都是主理想，这样的整环称为主理想整环。

利用理想的语言，最小公倍数的性质可改写为如下（这里的 \cap 即为通常意义的交集）：

$$a\mathbb{Z} \cap b\mathbb{Z} = \text{lcm}(a, b)\mathbb{Z}$$

注意到 $a\mathbb{Z}$ 是 a 的全部倍数集合， $b\mathbb{Z}$ 是 b 的全部倍数集合，左侧即代表 a 与 b 的公倍数集合，右侧则为最小公倍数的全部倍数集合。因此，这个式子的文字表述为： a, b 的公倍数等价于它们最小公倍数的倍数，这就是最小公倍数的定义。当然，讨论多项式时，对应的式子可以写作

$$f(x)\mathbb{K}[x] \cap g(x)\mathbb{K}[x] = \text{lcm}(f(x), g(x))\mathbb{K}[x]$$

值得注意的是，可以验证两个理想的交一定还是理想，称为交理想，于是最小公倍数实际上刻画了交理想。

不过，很容易发现，一个数全部的因数并不构成一个理想，因此最大公因数貌似没有这样的性质。然而，最大公因数满足如下更复杂的性质，称为裴蜀定理：

$$a\mathbb{Z} + b\mathbb{Z} = \text{gcd}(a, b)\mathbb{Z}$$

先解释左侧。 $a\mathbb{Z}$ 与 $b\mathbb{Z}$ 的定义已在之前说过，而它们的求和是指

$$\{x + y \mid x \in a\mathbb{Z}, y \in b\mathbb{Z}\}$$

更进一步即可写成

$$\{ua + vb \mid u, v \in \mathbb{Z}\}$$

由此，我们要证明左侧等于右侧，事实上是证明 $ua + vb = m$ 有整数解 u, v 当且仅当 m 是 $d = \gcd(a, b)$ 的倍数。证明分为三步：

1. 当 $m = d$ 时，若 a, b 有零可直接由定义构造解，否则，不妨设 a, b 均为正整数（否则 u, v 反号即可），利用辗转相除法的过程可以构造出对应的解 u, v 。

具体来说，我们归纳证明辗转相除法每一步的被除数、除数、余数都是 $ua + vb$ 的形式即可。不妨设 $a \geq b$ ，最初的被除数 a 、除数 b 、余数 $a - qb$ 满足要求，而被除数为 $u_1a + v_1b$ 、除数为 $u_2a + v_2b$ （它们若不为初始则都为某步的余数）时，余数

$$u_1a + v_1b - q_0(u_2a - v_2b) = (u_1 - q_0u_2)a + (v_1 + q_0v_2)b$$

也符合要求。最终得到的最大公因数一定是某步的余数，因此结论成立。

2. 当 $m = nd$ 时，取出 $ua + vb = d$ 的解 u_0, v_0 ，则 $(nu_0)a + (nv_0)b = nd$ ，从而可知 m 是 $\gcd(a, b)$ 的倍数时有整数解。
3. 由于 a, b 均为 d 的倍数， $ua + vb$ 一定为 d 的倍数，因此 m 不是 d 的倍数时没有整数解。

有了这三步证明，最大公因数的意义就明确了。正如线性空间可以定义和空间，两个理想可以定义和理想（可验证对整环两个理想的和还是理想），由此最大公因数实际上刻画了和理想。

当然，对多项式，我们也有

$$f(x)\mathbb{K}[x] + g(x)\mathbb{K}[x] = \gcd(f(x), g(x))\mathbb{K}[x]$$

此外，从这个式子中也可以进一步看出为什么我们要定义 $\gcd(0, 0) = 0$ ，此时式子两端为 $\{0\} + \{0\} = \{0\}$ 。

接下来，我们还要研究最大公因数与最小公倍数的一个重要性质：

$$\gcd(a, b) \operatorname{lcm}(a, b) = |ab|$$

同样，它有理想的表述

$$(I_1 + I_2)(I_1 \cap I_2) = I_1 I_2$$

这里理想 I_1, I_2 代入 $a\mathbb{Z}, b\mathbb{Z}$ 或 $f(x)\mathbb{K}[x], g(x)\mathbb{K}[x]$ 均成立。

当然，我们还是需要先定义两个集合的乘积，定义与加法完全类似：

$$I_1 I_2 = \{xy \mid x \in I_1, y \in I_2\}$$

事实上这个定义不严谨但反正我们碰不到反例所以无所谓了。

若 I_1 为 a 的所有倍数， I_2 为 b 的所有倍数，由定义可以得到 $I_1 I_2$ 表示 ab 的所有倍数。因此，上式右侧即为 $ab\mathbb{Z}$ ，而根据前面已证，左侧为 $\gcd(a, b)$ 所有倍数与 $\operatorname{lcm}(a, b)$ 所有倍数的集合乘积，也的确为 $\gcd(a, b) \operatorname{lcm}(a, b)\mathbb{Z}$ 。只要证明了 $\gcd(a, b) \operatorname{lcm}(a, b) = |ab|$ ，倍数集合自然相同。

当 a, b 有 0 时，由定义直接验证即可，其余情况不妨设它们均为正整数。记 $d = \gcd(a, b)$ ，即要证 $\operatorname{lcm}(a, b) = ab/d$ 。证明也自然分为两部分：

- 由公因数定义 b/d 是整数，于是 ab/d 为 a 的倍数，同理其为 b 的倍数，因此其为 a, b 的公倍数；

- 对 a, b 的任何公倍数 m , 设 $m = ka$, 则有 $b \mid ka$ 。由裴蜀定理, 可取出 u, v 使得 $ua + vb = d$, 由 $b \mid ka$ 可知 $b \mid kua$, 由定义得 $b \mid kvb$, 于是 $b \mid k(ua + vb)$, 即 $b \mid kd$ 。根据整除的定义可发现两边同乘整数仍然成立, 因此同乘 a/d 得到 $ab/d \mid ka$, 也就是 m 是 ab/d 的倍数, 又由其为正整数即得证其为最小公倍数。

将整数换成多项式可以得到完全类似的结论, 只是绝对值符号要改成 ab 乘非零倍数得到的首一多项式。在主理想整环中, 两个理想的乘积按我们的定义仍然是理想, 称为积理想, 由此最终得到, **最大公因数与最小公倍数的关系刻画了积理想的关系**。

最后, 我们来详细叙述一个颇麻烦但必要的定理: **唯一因子分解定理**。我们先给出整数中的版本: 称绝对值大于 1 的 (事实上意味着不为 0 或单位)、因数只有 1 与自身 (所在等价类, 即允许相差符号) 的整数为素数 (注意此定义下 $-2, -3$ 也是素数), 则任何一个绝对值大于 1 的整数 a 都可以表示为

$$a = \pm p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$$

其中 p_1, \dots, p_k 为不同正素数, a_1, \dots, a_k 为正整数。在改变 p_1 到 p_k 的顺序视为相同的情况下, 表示唯一。

多项式的版本则更为复杂: 称 $\mathbb{K}[x]$ 中至少一次的 (同样意味着不为 0 或单位)、因式只有 1 与自身 (所在等价类, 即允许乘 \mathbb{K} 中非零元素) 的多项式为不可约多项式, 则任何一个至少一次的多项式 $f(x)$ 都可以表示为

$$f(x) = tp_1^{a_1}(x)p_2^{a_2}(x) \cdots p_k^{a_k}(x)$$

其中 $p_1(x), \dots, p_k(x)$ 为不同首一不可约多项式, a_1, \dots, a_k 为正整数, $t \in \mathbb{K}$ 非零。在改变 $p_1(x)$ 到 $p_k(x)$ 的顺序视为相同的情况下, 表示唯一。

我们只对多项式进行证明, 毕竟这才是我们这学期关注的重点。表示的存在性是相对容易证明的: 取 t 为 $f(x)$ 的首项系数, 即可设 $f(x)$ 首一。只要 $f(x)$ 有 1 与自身外的其他因式 $g(x)$, 就必然存在首一的 $g(x)、h(x)$ 使得 $f(x) = g(x)h(x)$, 且 g, h 的次数都小于 f 的次数。对 g, h 进一步考虑, 由于多项式的次数有限, 能进行的分解次数同样有限, 必然某一步乘法的每一项都是不可约的, 再将相同因式整理到一起即可。

对于唯一性, 利用定义可发现 t 只能为 $f(x)$ 首项系数, 从而仍然不妨设 $f(x)$ 首一。若其有两种不同表示

$$p_1^{a_1}(x)p_2^{a_2}(x) \cdots p_k^{a_k}(x) = q_1^{b_1}(x)q_2^{b_2}(x) \cdots q_l^{b_l}(x)$$

将两边相同部分约去可不妨设 $p_1, \dots, p_k, q_1, \dots, q_l$ 互不相同。进一步地, 利用不可约可知 q_1 与 p_1, \dots, p_k 的最大公因式都为 1 (若否, 假设 $\gcd(q_1, p_1) = d \neq 1$, 则根据不可约定义只能 $d = p_1$ 且 $d = q_1$, 与两者不同矛盾)。

我们先证明引理: 若 $\gcd(a, b) = 1$ 、 $\gcd(a, c) = 1$, 则 $\gcd(a, bc) = 1$ 。利用裴蜀定理设 $u_1a + v_1b = 1$ 、 $u_2a + v_2b = 1$, 直接构造可得

$$1 = (u_1a + v_1b)(u_2a + v_2c) = (u_1u_2a + u_1v_2c + u_2v_1b)a + (v_1v_2)bc$$

从而 $ua + vbc = 1$ 有整数解 u, v , 再次根据裴蜀定理得结论。

利用引理, 归纳即得 $\gcd(q_1, p_1^{a_1}p_2^{a_2} \cdots p_k^{a_k}) = 1$, 但根据两式相等可得右侧应为 q_1 倍数, 矛盾。

对于最后这个重要的结论, 同样需要给出几点注释:

- 事实上可以证明, **主理想整环一定可以进行唯一因子分解**, 因此真正的推导顺序是欧几里得环 \Rightarrow 主理想整环 \Rightarrow 唯一分解整环。

- 与判断一个数是否是素数有简单的算法不同，判断一个 $\mathbb{K}[x]$ 中的多项式是否可约是一个非常困难的问题。更复杂的是，同一个多项式在 \mathbb{K} 不同时情况不同，例如 $x^2 - 2$ 作为 $\mathbb{Q}[x]$ 中多项式不可约，而作为 $\mathbb{R}[x]$ 中多项式可以约成 $(x - \sqrt{2})(x + \sqrt{2})$ ； $x^2 + 1$ 作为 $\mathbb{R}[x]$ 中多项式不可约，而作为 $\mathbb{C}[x]$ 中多项式可以约成 $(x + i)(x - i)$ 。
- 对于 $\mathbb{C}[x]$ 有如下代数学基本定理： $\mathbb{C}[x]$ 上的多项式不可约当且仅当其次数为 1。结合唯一分解定理，这就说明 $\mathbb{C}[x]$ 上的 n 次多项式可以唯一分解成 n 个一次因式。上学期我们默认了这个结论成立才能得到特征值的存在性与不计次序意义下的唯一性。
- 对于 $\mathbb{R}[x]$ 有如下定理： $\mathbb{R}[x]$ 上的多项式不可约当且仅当其次数为 1，或次数为 2 且无实根（即写成 $(x - a)^2 + b^2$ ， a 为实数， b 为正实数）。这将在我们之后讨论有理标准形时用到，那时我们将在假设代数学基本定理成立下证明。
- 根据唯一因子分解也可以计算最大公因式与最小公倍式。若有 $\mathbb{K}[x]$ 中

$$f(x) = tp_1^{a_1}(x)p_2^{a_2}(x) \dots p_k^{a_k}(x)$$

$$g(x) = sp_1^{b_1}(x)p_2^{b_2}(x) \dots p_k^{b_k}(x)$$

其中 t, s 是 \mathbb{K} 中非零元素， p_1, \dots, p_k 为不同的首一不可约多项式， $a_1, \dots, a_k, b_1, \dots, b_k$ 为非负整数（这样只在 f 或 g 中出现的因式在另一个中的次数为 0），则有

$$\gcd(f(x), g(x)) = p_1^{\min(a_1, b_1)}(x)p_2^{\min(a_2, b_2)}(x) \dots p_k^{\min(a_k, b_k)}(x)$$

$$\text{lcm}(f(x), g(x)) = p_1^{\max(a_1, b_1)}(x)p_2^{\max(a_2, b_2)}(x) \dots p_k^{\max(a_k, b_k)}(x)$$

定理的证明与分解唯一性的证明是类似的，虽然更加繁琐。

- 虽然上述方法对最大公因式与最小公倍式的计算与 $\mathbb{K}[x]$ 所取的 \mathbb{K} 有关（因为分解方式与 \mathbb{K} 有关，结果实际上是无关的，这是因为辗转相除法的过程与 \mathbb{K} 无关。由此，最大公因式与最小公倍式在 \mathbb{K} 取不同数域时是不变的）。

§14.2 高观点与低观点

14.2.1 教学方式的权衡

说到底，在考试范围内，本讲义 14.1 的全部讨论比起课上只是多得到了几个结论：

- 任何数都整除 0，0 也整除 0（因为可以只用乘法定义整除）。
- 在讨论多项式相关问题时，可以将 0 也视为首一多项式（因为要求首一本质上是在等价类中取出代表元）。
- 0 与任何其他多项式 f 的最大公因式都是 f 对应的首一多项式， $f = 0$ 时则有 $\gcd(0, 0) = 0$ （因为这里的最大是在整除性角度，0 是最大的）。
- 在多项式中，唯一因子分解定理仍然成立，也即 $\mathbb{K}[x]$ 中至少一次多项式可以唯一（交换顺序视为不变）分解为一些 $\mathbb{K}[x]$ 中的首一不可约多项式与某个 \mathbb{K} 中非零元素的乘积。此外，即使是同一个多项式，在 \mathbb{K} 不同时分解仍然可能不同。
- 利用唯一因子分解定理可以计算最大公因式与最小公倍式，但事实上最终结果与数域无关（因为可以通过辗转相除法直接得到）。

但是，为了得到这些结论，我们引入了一大堆新的概念，也进行了很多页的复杂讨论。如果这就是所谓的高观点，这一切，值得吗？

想解释这个问题，我们就必须知道何为更“高”的观点。就像之前所说，高观点的理解往往是需要对某个问题不断分层、剖析、延伸，补充清楚研究的动机与方法的思路。而很遗憾的是，很多问题如果想要解释清楚，都不得不引入更高级的概念。

例如，就 $\gcd(0, 0) = 0$ 这件事，我们事实上给出了两个理解：“0 作为序关系下的最大元素”与“最大公因数作为和理想的生成元”，无论是哪种，都需要引入一些更抽象的代数定义。然而，如果不这么做，我们的教学就只有两种选择：

- 模糊化这个问题，例如王、石教材中 8.3 节所做的，行列式因子只定义到存在非零子式的阶数，避免出现全是 0 时计算最大公因式的问题；
- 将这作为一种“规定”来呈现，例如王、石教材中 1.4 节所做的，直接给出最大性的定义并指出根据定义 $\gcd(0, 0) = 0$ 。

这两种选择各自存在问题：对第一种选择来说，这样的定义一定程度上破坏了一致性，无法体现行列式的最大公因式事实上可以在任意阶定义，从而只要通过各阶行列式的最大公因式就可以完全决定相抵标准形——由于允许 0 整除 0，相继整除也是可以一直得到满足的。对第二种选择来说，作为规定的定义显得不自然，因为 0 与通常意义上的最大是冲突的，而定义并没有解释这种反直觉性。

就我们所用的线代教材与课程而言，事实上是以低观点为主的，因为高观点有高观点的缺陷，也就是更加困难与需要更多篇幅。作为非数学专业的线性代数，将定义、定理顺序呈现并花时间给出充足的推导细节似乎是一个让更多人能够听懂的更好选择（而不是像习题课一样一段没跟上就再也不知道后面在干嘛了）。

但是，这里仍然有一个“但是”。正因为低观点教学的理解与动机是缺乏的，想要掌握知识所需的记忆与训练量更大——就像书上经常冒出很多莫名其妙不知道怎么想到的证明，如果想都能应用就只能记住全部的技巧。而众所周知，下册教材的厚度远超上册，想要全部记住更加不现实，于是也就更加考验高观点的整理。

一个 $\gcd(0, 0) = 0$ 自然是容易记忆的，但如果脑子里对它为什么成立没有概念，在学习的知识更多后迟早会忘掉，并在不知何时出来捅自己一刀。更何况，教材里还有无数这样“凭空冒出”的东西。下一部分我们将谈到，下册线性空间本来就可以视为矩阵论的高观点，因此就更需要理解了。

基于以上理由，既然大家已经听了足够多低观点的推导，无论是我自己的习题课还是这份讲义，都将尽量补充高观点的内容（但也有例外，详见下一部分）来帮助大家理解、记忆。在时间有限的情况下，高观点的东西难跟上、听懂是正常的——这当然并不意味着“能听懂”的课就一定更好——比起认真笔记过程细节，更好的听课思路是，尽量想清楚每一部分在做什么。遗憾的是，就算是这样，也还是可能有大量听了一遍不清楚的内容，这也就是这份足够详细的讲义存在的意义，能让大家通过阅读补全细节。

最后，如果有的部分实在掌握不了高观点，其实也没有任何问题：正如前面所说，对这门课程而言低观点的记忆可以部分替代高观点的理解，因此站在通过考试的角度，留一些不会的点死记硬背也是完全没有问题的。不过，除了人类的记忆力有限之外，如果想要这些知识真的留存下来，应用在之后的学习乃至科研中，高观点的理解仍然是重要的。

14.2.2 作为矩阵论高观点的线性空间

这一部分中，我们需要谈谈下册的核心，**线性空间**，是如何作为矩阵论的高观点而存在的。

本质上来说，高观点对问题的理解是为了推广到更广泛的结构或情况中。因此，说线性空间是高观点，恰恰是因为它除了能够完全表示矩阵理论之外，还可以处理矩阵理论无法处理的情况——虽然后者实际上已经超出了本学期有限维线性空间的讨论范畴。

这部分的更多内容将在本讲义第十六章对线性空间进行引入时讨论，这里只简单说明一些结论（其实有不少在上学期习题课讲义中已经写了，如本讲义第八章）：

- 为了将向量空间进行推广，我们将考虑一般的具有加法、数乘（也就是线性结构）的系统，称为线性空间；
- 由于有了加法、数乘后就可以描述线性相关性，我们仍然可以仿照向量空间定义基与维数——然而，一般线性空间的维数未必有限，这就会出现很多复杂的情况；
- 为了描述线性空间之间的“保结构映射”，我们仿照向量空间的线性映射可定义一般的线性映射；
- 由线性映射的单射/满射/双射可以定义线性同构，线性空间之间存在同构当且仅当维数相同，由此有限维线性空间都可以视为向量空间；
- 从一般的线性映射出发也可以构造一个同构，这需要先从等价类定义商空间，这样的同构构造称为第一同构定理；
- 利用矩阵乘法， \mathbb{K} 上的 $m \times n$ 矩阵可以自然看作 \mathbb{K}^n 到 \mathbb{K}^m (列向量) 的线性映射，可以证明所有 \mathbb{K}^n 到 \mathbb{K}^m 的线性映射都是矩阵乘法；
- 进一步推广到线性空间，利用同构，给定两边的基后，任何 n 维线性空间到 m 维线性空间的线性映射都可以唯一看作一个 $m \times n$ 矩阵，称为线性映射的矩阵表示；
- 若两边的基发生改变，即使映射本身不变，对应的矩阵表示也会改变，而改变前后的矩阵是相抵的，由此暗示相抵不变量是线性映射的不变量——也就是秩；
- 若前后的空间是同一个，此时的线性映射称为线性变换，可与线性映射完全相同定义矩阵表示；
- 要求前后空间取同一组基，则基发生改变时对应的矩阵表示也会改变，而改变前后的矩阵是相似的，由此暗示相似不变量是线性变换的不变量；
- 如上所述，相抵标准形可以看作寻找线性映射最简的矩阵表示，相似标准形可以看作寻找线性变换最简的矩阵表示；
- 一个 \mathbb{R} 或 \mathbb{C} 上的线性空间可以类似向量空间定义内积，这时其称为内积空间；
- 内积空间的“保结构映射”需要保内积，考虑到内积的几何意义，这称为等距映射；
- 等距映射的矩阵表示是正交矩阵，因此研究正交矩阵就是在研究等距映射；
- 另一方面，一组基互相之间的内积形成内积矩阵，在基进行变换时，内积矩阵是彼此合同的；
- 通过内积的某种对偶结构可以定义伴随映射，它对应的矩阵表示是原映射矩阵表示的转置，由此实对称阵（或 Hermite 阵）就是自伴映射——伴随是自身；
- 再次利用内积的几何意义，自伴映射对应的矩阵表示正定事实上意味着其某种意义上对方向的改变较小……

从上面的这些结论可以看出，我们上学期学习的全部结论事实上都有空间版本。除了正定阵以外，其他的空间版本应该都会在这学期学到。它们之所以展现了线性空间的高观点性，是因为推广为线性映射后往往都可以在无穷维讨论，而这是矩阵做不到的。

不过，想要详细解释这些高观点，事实上还是需要对低观点的讨论有足够的熟悉程度，实际难度是相对高的。因此，习题课讲义中也会尽量给空间问题提供矩阵论的低观点解法。这样，即使有高观点无法掌握的内容，也可以通过背熟定义转化为矩阵论进行求解。

14.2.3 做题指南

这一部分，我们来谈一谈这学期如何正确地做题。

由于知识量本身非常多，这学期做题不仅必要，也有着非常多的选择。然而，大家在这门课上分配的时间肯定有限（大概率没我写讲义多），因此需要聊聊怎么在有限的时间里让做题有最大的效果。细说下来，建议主要有三：

1. 少浪费时间思考。

这句话乍一看有点奇怪，思考为何是浪费时间呢？必须说明的是，这里“思考”指只用脑子想，而不动手操作——或是觉得无从下手。事实上，在笔没有动的时候，大部分的思考时间是浪费的，因为过于发散的思路往往很难有实际意义。

于是，一个更好的选择是，保持自己不停动笔尝试，而在有一段时间（或许是 5 到 10 分钟）无法尝试后，就该去寻求帮助了，因为这意味着短时间很难做出有效的进展：或许是因为遗忘了某个关键定理，又或者是没有见过某个特定操作，无论何种原因，此时继续硬做都不是一个明智的选择。

2. 少利用网络资料。

这个建议包含两个部分，无论是习题的搜集还是答案的寻找，个人都不建议利用网络资料（包括 AI 生成）。原因是一致的：网络资料无法保证知识体系和我们所学的相符。这学期的内容可能有很多不同的讲法（哪怕三个班都互不相同），题目自然也会根据知识顺序不同有完全不同的想法与做法。

就习题搜集而言，丘书的题目做一遍应该足够考到很不错的成绩了，如果真的还嫌不够可以考虑找其他数学系教材做对应章节。更重要的是关于寻找答案：非常推荐优先与一起选课的同学交流，其次询问助教，再次翻书后答案，最后才是网络搜索。

理论来说，与同学交流是最符合知识体系相同的，但一定需要确保交流后能有更正确的理解方向，否则可能会一起偏移；询问助教可以保证对知识体系有了解，但得到的思路或许未必是现在可以合理想到的；至于后两者，就只能祈祷看到的解答写得足够清晰明确了。

3. 一句句看答案。

最后，对于不会的题目，想要起到做题的效果，得到答案或提示后一定要一句句看。具体来说，只要看到某步觉得可以尝试就不要再看下去，试着自己动手将其他部分补全。

经验表明，一道题直接看答案理解其实与没做没什么两样，因为做题本就是在尝试中学会方法。哪怕是动手尝试的路是错的，去想清楚它为什么不可行（或是和同学/助教交流发现可行，只是自己没做出来）也是非常有价值的。尤其是在时间有限的情况下，就更需要让做题的时间真正达到“动手”的效果。

14.2.4 记忆关键结论？

最终，我们还是要聊到记忆。既然下册的内容如此之多（事实上很多上册中的结论也仍然重要），必须有一个选取何种内容值得记忆的方法。除了必须记忆的定义与大定理外，关于其他东西是否值得记忆，也还是有两点需要讨论：

1. 第一，判断一个结论是否是核心结论的常见方法是看它出现在了多少次其他题的引用中。一般来说，被引用越多就意味着越核心，也越值得记忆。

不过，我们不可能在所有题中统计一个结论被引用了多少次。因此，每看到答案中引用（或自己想到引用）一个结论时，有三个问题是必须思考的：这个结论是否显然？这个结论是否重要？此处的引用是否必要？

例如，上册书中有一道习题，已知 A, B 为同阶正定方阵，证明 AB 正定当且仅当 $AB = BA$ 。在右推左的证明中，它引用了一个结论：两个乘积可交换的同阶正定方阵可以同时正交相似对角化。

这个结论是否显然可以直接通过书上证明的长度看出：它完全不显然。事实上，有时教材确实会因为不想写显然后引用一些比较显然的结论，这些往往是自己可以比较轻易推出，于是不需要特别记忆。

对于这个结论是否重要，关键评判标准是它是否关联了核心结构。例如，正交相似对角化当然是处理对称阵的核心方法，而同时正交相似对角化意味着乘积的特征值可以直接看作特征值对应乘积，因此的确是重要的。不过，它的条件要求较高，意味着应用范围可能不广。

而所谓的必要性，则是指是否不能绕过这个结论证明。对于这题来说，答案是否定的：考虑 $A = P^T P$ ， P 可逆，则 $AB = P^T PB = P^T(PBP^T)P^{-T}$ 相似于 PBP^T ，特征值与 PBP^T 相同，而 PBP^T 与 B 合同，因此正定，从而 AB 特征值均为正，正定等价于对称，直接计算可发现对称即等价于 $AB = BA$ 。

由此，这个结论是一个**有重要性但应用范围不广、存在替代可能**的结论，于是，记忆它的优先级可以相对向后排，但准备时间充足时还是可以进行记忆的。

2. 第二，与永远不要指望能记得所有结论相对的是，哪怕对并不能记得的结论，也最好有一定印象，尤其是务必留意定理条件。

就改卷来说，只要某一步是错误的，基于它的证明一定也是错误的。因此，还是以正交相似对角化为例，期末考试凡是说两个正定阵可以同时正交相似对角化的同学都被扣了全分。

个人的建议是，如果没法记得正确的结论，至少也多读几遍加深一下印象，千万不要使用不确定条件的结论。

当然，讨论这点则又要说回之前所述的理解。一般来说，对一个知识点理解越深，就越不容易出现错误的记忆。我们也将在这学期之后的习题课里慢慢讨论如何进行**理解性的记忆**，如何对一个结论的正误产生一定的直觉。

* 虽然写下这一章时已经是第二周了，但总归这学期才刚刚开始，希望能和大家一起度过一个学有所得的学期——而做到这件事的基础是有问题及时解决，对于个人来说，则是建议及时来问。

十五 从同余方程组说起

* 建议先行阅读本讲义 14.1 以了解基本定义。

* 本章与上一章的内容大概率只有行列式因子、不变因子、初等因子组的计算会出现在考试中，介绍这么多主要是为了让大家了解更高视角下这一套理论引入的动机，也给后文的更多动机性的分析做铺垫。

§15.1 再谈数论

15.1.1 线性方程组？

从本次习题课开始的几次课中，我们将要研究任意数域 \mathbb{K} 上的相似标准形问题。而或许不可思议的是，为了研究相似，我们需要引入一个大量涉及数论的工具——多项式矩阵，也即每个元素都是多项式的矩阵。在上一章中，我们已经看到多项式与整数的诸多类似性质，因此，我们还是从相对简单的整数矩阵（每个元素都是整数的矩阵）谈起。

如果大家对上学期的内容还存在印象，应当能记得，我们引入数域上的矩阵是从求解线性方程组开始的。同样，想要自然给出整数矩阵的概念，也需要先观察一类特殊的方程组。

《孙子算经》中记载了一道著名的问题（有时也称韩信点兵问题）：今有物不知其数，三三数之剩二，五五数之剩三，七七数之剩二，问物几何？翻译一下也就是，如果一个整数除以 3 的余数是 2，除以 5 的余数是 3，除以 7 的余数是 2，求这个数的可能解。

再进行一点抽象，所有除以 3 余 2 的数可以用 $3k+2$ 来表示 (k 为整数)，因此上述问题可以写为，求整数 x 使得存在整数 k_1, k_2, k_3 满足

$$\begin{cases} x = 3k_1 + 2 \\ x = 5k_2 + 3 \\ x = 7k_3 + 2 \end{cases}$$

我们容易将它化成更熟悉的形式：求整数 x_1, x_2, x_3, x_4 满足方程组

$$\begin{cases} x_1 - 3x_2 = 2 \\ x_1 - 5x_3 = 3 \\ x_1 - 7x_4 = 2 \end{cases}$$

可以发现，这个问题完全就是一个系数均为整数的线性方程组（此后称为整线性方程组），只不过，我们需要关心的是它的整数解。由于任何线性同余方程组问题都可以类似上方转化成整线性方程组问题，而线性同余方程组又是数论的奠基性内容，就不难理解研究整线性方程组的必要性了。

仍然与之前学习线性方程组时相同，一个整线性方程组可以写为 $Ax = b$ 的形式，其中 A 为 $m \times n$ 阶整数矩阵， x 为 n 维未知列向量， b 为 m 维整数列向量。当 $b = 0$ 时，我们称这是一个齐次整线性方程组，否则称为非齐次整线性方程组，我们下面将 n 维整数列向量集合记作 \mathbb{Z}^n ， $m \times n$ 阶整数矩阵集合记作 $\mathbb{Z}^{m \times n}$ 。

事实上，整线性方程组的整数解与线性方程组的解有非常类似的性质，我们给出两条最重要的：

- 设 $A \in \mathbb{Z}^{m \times n}$ 、 $b \in \mathbb{Z}^m$ 。若 x_0 是 $Ax = b$ 的一个整数特解（即其为整数向量且 $Ax_0 = b$ ），则整线性方程组 $Ax = b$ 的整数解集可以写成

$$\{x_0 + y \mid Ay = 0, y \in \mathbb{Z}^n\}$$

这个定理的证明与上学期非齐次线性方程组的解集结构证明完全相同：若 $Ax_0 = b$ 、 $Ay = 0$ ，直接计算发现 $A(x_0 + y) = b$ ；若 $Ax_0 = b$ 、 $Ax_1 = b$ ，直接计算发现 $A(x_1 - x_0) = 0$ 。这告诉我们，非齐次整线性方程组的整数通解为其整数特解与对应齐次整线性方程组的整数通解之和，这个性质与一般线性方程组完全相同。

- 对于齐次整线性方程组，若 x, y 为其整数解，则 $x + y$ 为其整数解；对任何 $\lambda \in \mathbb{Z}$ ， λx 为其整数解。
这个定理同样直接计算就可以验证。回忆上学期，齐次线性方程组的类似性质可以推出齐次线性方程组的解集是一个线性空间。

出于以上两个性质，我们首先希望能在整数向量中定义一个类似线性空间的代数结构，再说明齐次整线性方程组的解集一定符合这个结构，如果一切顺利，下面我们就可以转而研究这个结构以确定整线性方程组的解了。

对 \mathbb{Z}^n 的一个子集 V ，若其满足 $x \in V, y \in V$ 则 $x + y \in V$ ，且对任何 $\lambda \in \mathbb{Z}$ ， $\lambda x \in V$ ，则称 V 为一个 \mathbb{Z} 模，此后无歧义时直接称为模。利用刚才已经证明的结论，齐次整线性方程组的解集是一个模。

* 若大家还记得本讲义 14.1.4 的概念，可以发现 \mathbb{Z} 的理想就是能作为 \mathbb{Z} 的子集的模。

15.1.2 模的构造尝试

至此，大家可能会产生一个非常直觉的疑问：为什么我们要把“模”与“线性空间”两个概念区分开？整线性方程组与线性方程组真的有很大差别吗？例如，我们并不会单独谈论“有理线性方程组的有理数解”，因为此结论已经被一般数域上的线性方程组解集结构所包含。本节的关键就是解决这个问题，这里先给出结论：整线性方程组的讨论方式与线性方程组有一定相似，但会存在重大区别。

* 必须注意，我们这里对模的所有定义和讨论都建立在模是 \mathbb{Z}^n 的子集上，与数学上的一般定义存在较大差别，因此这里的定义、定理并不对更一般的模适用。

在上学期，我们分析数域上的向量空间时，先定义了线性相关、线性无关、线性表出等概念，再从这些得到了极大线性无关组的概念，进而定义了线性空间的基与维数，最后回到线性方程组，完成了对齐次线性方程组解集的刻画。对于一些整数向量，我们仍然可以类似定义：

1. 对 $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$ ，若存在不全为 0 的 $\lambda_1, \dots, \lambda_k \in \mathbb{Z}$ 使得

$$\lambda_1\alpha_1 + \dots + \lambda_k\alpha_k = 0$$

则称它们整线性相关，否则称它们整线性无关。

2. 对 $\alpha_0, \alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$ ，若存在不全为 0 的 $\lambda_1, \dots, \lambda_k \in \mathbb{Z}$ 使得

$$\alpha_0 = \lambda_1\alpha_1 + \dots + \lambda_k\alpha_k$$

则称 α_0 可以被 $\alpha_1, \dots, \alpha_k$ 整线性表出。

3. 对 $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$ ，若它的一个子集互相整线性无关，且再添加任何一个后都整线性相关，则其称为 $\alpha_1, \dots, \alpha_k$ 的一个极大整线性无关组。

4. 对 $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$ ，称能被它们整线性表出的向量集合为它们的生成模，记作 $\langle \alpha_1, \dots, \alpha_k \rangle$ 。

* 这个定义的说法暗示了此集合是一个模，直接验证若 α, β 可被它们表出，则 $\alpha + \beta$ 、任意整数 λ 的 $\lambda\alpha$ 可被它们表出即可，与生成线性空间的证明完全相同。

5. 若一个模 V 是 $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$ 的生成模，且 $\alpha_1, \dots, \alpha_k$ 线性无关，则称它们是 V 的一个基。

用这些定义，我们马上就可以通过不算复杂的证明得到一些相同的结构：

1. 引理：对任意有限个有理数 q_1, \dots, q_n ，存在非零整数 m 使得 mq_1, \dots, mq_n 为整数。

证明：

若 q_1, \dots, q_n 全为 0，取 $m = 1$ 即可，否则考虑其中的非零有理数，将 m 取为所有非零有理数的分母乘积，可以直接验证 mq_1, \dots, mq_n 都是整数。

* 虽然这个定理的表述与证明都非常简单（本质是多个有理数可以通分），但我们在下面的证明中看到，它是沟通整线性方程组与之前的线性方程组理论的桥梁。

2. $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$ 整线性无关当且仅当它们看作 \mathbb{Q}^n 中的向量线性无关。

证明：

若 $\alpha_1, \dots, \alpha_k$ 整线性相关，存在不全为 0 的整数 $\lambda_1, \dots, \lambda_k$ 使得 $\lambda_1\alpha_1 + \dots + \lambda_k\alpha_k = 0$ ，由于不全为 0 的整数也是不全为 0 的有理数，它们看作 \mathbb{Q}^n 中的向量线性相关。

若 $\alpha_1, \dots, \alpha_k$ 看作 \mathbb{Q}^n 中的向量线性相关，存在不全为 0 的有理数 $\lambda_1, \dots, \lambda_k$ 使得 $\lambda_1\alpha_1 + \dots + \lambda_k\alpha_k = 0$ ，利用引理取非零整数 m 使得 $m\lambda_1, \dots, m\lambda_k$ 均为整数（由 m 非零它们仍不全为 0），则有

$$(m\lambda_1)\alpha_1 + \dots + (m\lambda_k)\alpha_k = 0$$

就已经找到了不全为 0 的整数 μ_1, \dots, μ_k 使得 $\mu_1\alpha_1 + \dots + \mu_k\alpha_k = 0$ ，于是它们整线性相关。由于一个向量组要么（整）线性无关，要么（整）线性相关，综合以上两部分即得到了证明。

3. $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$ 的极大整线性无关组就是它们看作 \mathbb{Q}^n 中的向量时的极大线性无关组。

证明:

由于线性无关性与线性相关性都与看作 \mathbb{Q}^n 中的向量时相同, 根据极大线性无关组的定义可得结论。

4. $\alpha_1, \dots, \alpha_k \in \mathbb{Z}^n$ 的任何极大整线性无关组所含的向量个数相同, 称为向量组的秩, 即为它们看作 \mathbb{Q}^n 中的向量时的秩。

证明:

由于极大线性无关组与看作 \mathbb{Q}^n 中的向量时相同, 利用看作 \mathbb{Q}^n 中的向量时所有极大线性无关组所含的向量个数相同可得结论。由此, 极大整线性无关组所含的向量个数也与看作 \mathbb{Q}^n 中的向量时相同, 也即秩同样可看作 \mathbb{Q}^n 中的向量计算。

5. $A \in \mathbb{Z}^{m \times n}$ 的行向量组、列向量组秩相同, 称为矩阵的秩, 记作 $\text{rank } A$, 即为它看作 $\mathbb{Q}^{m \times n}$ 中矩阵时的秩。秩仍然等于最大非零子式的阶数。

证明:

将 A 看作 \mathbb{Q} 上的矩阵, 行秩、列秩相等, 而这又等于作为整数向量组的行秩、列秩, 从而第一句结论成立。由于行列式计算方法不变, A 的非零子式与其看作 \mathbb{Q} 上的矩阵时的非零子式相同, 从而得证。

6. 对模 V , 记 $U = \{\mu v \mid \mu \in \mathbb{Q}, v \in V\}$, 则其构成 \mathbb{Q} 上的线性空间, 且 V 的基看作 \mathbb{Q}^n 中向量一定是 U 的基。

证明:

若 $x = \mu_1 v_1 \in U$ 、 $y = \mu_2 v_2 \in U$, 取非零整数 m 使得 $m\mu_1, m\mu_2$ 为整数, 则 $m\mu_1 v_1 \in V$ 、 $m\mu_2 v_2 \in V$, 从而

$$x + y = \frac{1}{m}((m\mu_1)v_1 + (m\mu_2)v_2) \in U$$

而直接由定义可发现对任何 $\mu \in \mathbb{Q}$ 有 $\mu x = (\mu\mu_1)v_1 \in U$, 于是其为线性空间。

对 V 的一组基 $\alpha_1, \dots, \alpha_k$, 之前已经证明了看作 \mathbb{Q} 中向量时线性无关, 由于 V 中任何元素都可以写为

$$v = \lambda_1 \alpha_1 + \dots + \lambda_k \alpha_k, \quad \lambda_1, \dots, \lambda_k \in \mathbb{Z}$$

U 中任何元素都可以写为

$$\mu v = (\mu\lambda_1)\alpha_1 + \dots + (\mu\lambda_k)\alpha_k$$

所有 $\mu\alpha_i$ 都是有理数, 从而根据定义 $\alpha_1, \dots, \alpha_k$ 确实构成 U 的一组基。

7. 模 V 的任何基个数相同, 称为 V 的维数, 记作 $\dim V$, 定义 $\dim\{0\} = 0$ 。

证明:

由于 V 的基一定是上方定义的 U 的基, 而由上学期知识 U 的任何基个数相同, 因此 V 的任何基个数相同。

* 由此也可以看出, V 只要存在一组基, 其维数就必然与 U 的维数相等。

8. 对 \mathbb{Q} 上的线性空间 U , 记 V 为 U 中所有整数向量构成的集合, 则 V 是模且 $U = \{\mu v \mid \mu \in \mathbb{Q}, v \in V\}$ 。

证明:

U 中两个整数向量的和仍然是 U 中的整数向量，其仍然在 V 中； U 中整数向量数乘整数后仍然是 U 中的整数向量，其仍然在 V 中。综合两者即得到 V 是一个模。

对 U 中任何向量 u ，可取非零整数 m 使得 mu 为整数向量，从而 $mu \in V$ ，于是 $u = \frac{1}{m}(mu) \in \{\mu v \mid \mu \in \mathbb{Q}, v \in V\}$ 。另一方面，当 $\mu \in \mathbb{Q}$, $v \in V \subset U$ 时，根据线性空间定义可知 $\mu v \in U$ ，因此 $U = \{\mu v \mid \mu \in \mathbb{Q}, v \in V\}$ 。

* 第二部分证明实际上是证明了两个集合相互包含，这是证明集合相等时最常用的办法。

很遗憾，能类似讨论的部分到这里就结束了。顺着这个思路，我们马上就会遇到一个非常困难的问题：当 V 不止包含零向量时，其基是否一定存在？我们知道，在一个线性空间中，任何数量等于维数的线性无关向量组都构成了一组基，而在模中，这一性质并不成立。例如，所有偶数的集合构成一个模， $\{4\}$ 作为向量组（里面只有一个一维向量）是整线性无关的，但它并不构成一组基，因为 2 并不是它的倍数。将上方构造的 U 的一组基乘一个非零整数，可以得到 V 中整线性无关的向量组，但它们未必是一组基。

在研究非齐次整线性方程组时，这个问题更加明显。在上一部分，我们事实上已经证明了，只要找到非齐次整线性方程组的一个整数特解，通解立刻就可以表示出来，但问题恰恰出现在特解上。我们先给出如下与普通非齐次线性方程组完全不同的结论： $\text{rank}(A, b) = \text{rank } A$ 不足以说明非齐次整线性方程组的整数解存在性。最简单的例子是考虑单个方程构成的方程组 $2x = 3$ ，它满足 $\text{rank}(2, 3) = \text{rank}(2) = 1$ ，但并不存在解。

上述问题的本质在于，向量组整线性相关并不能推出其中某个向量能被其他向量表出。两个整数（视为一维向量）一定是整线性相关的，但只要它们没有倍数关系，并不存在一个能被另一个表出。这就意味着，讨论整线性方程组不得不涉及数论（也就是倍数、因数）相关的问题。当然，从模的理论也是可以对这些问题给出解决方案的，但这需要一些更加抽象化的数学理论，因此，以目前的知识背景，我们必须换另一条路解决问题。

* 当然，这并不意味着上述的尝试就没有意义了。这些尝试代表我们在试着寻找代数结构间的共性，除了是重要的抽象以外，也提供了一个看待整数方阵的方式。事实上，就像矩阵理论有空间的解释，整数矩阵的相关理论也都有用模理解的版本，留给感兴趣的读者自己进行表述。

15.1.3 行列变换的思路

另一个解线性方程组的思路恰恰就是我们在上学期最开始学习的行列变换方法。由于我们已经知道进行行列变换相当于乘可逆方阵，比较形式化的方式是通过相抵标准形进行操作。

具体来说，对线性方程组 $Ax = b$ ，设 $A \in \mathbb{R}^{m \times n}$ 的相抵标准形

$$A = P \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} Q$$

则原方程可以写为

$$P \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} Qx = b$$

也即

$$\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} Qx = P^{-1}b$$

由于左侧相当于将 Qx 的后 $n - r$ 个分量变为 0，前 r 个分量不变，此方程组有解当且仅当 $P^{-1}b$ 的后

$n - r$ 个分量均为 0 (计算可发现其等价于 $\text{rank}(A, b) = \text{rank } b$)，此时直接计算得解为

$$x = Q^{-1} \begin{pmatrix} c_1 \\ \vdots \\ c_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix}$$

其中 c_1 到 c_r 为 Pb^{-1} 前 r 个分量， x_{r+1} 到 x_n 可以取任何实数。

回顾上学期得到相抵标准形的思路，我们又可以将得到相抵标准形的过程拆分为以下的步骤：

- 定义符合要求的行列变换，使得“能行列变换得到的矩阵”成为等价关系；
- 证明任何矩阵可以在行列变换下成为某个较简单的形式；
- 通过证明较简单形式的唯一性，最终确定分类；
- 定义可逆矩阵，使得“能左右乘可逆阵得到的矩阵”成为等价关系；
- 证明行列变换可以看作乘可逆阵，从而能行列变换得到的矩阵一定可以左右乘可逆阵得到；
- 证明可逆矩阵可以拆成行列变换对应矩阵的乘积，最终证明上述两个等价关系对应的分类是相同的，将这样的分类称为相抵标准形。

对于相抵来说，“较简单的形式”自然就是相抵标准形 $\text{diag}(I_r, O)$ ，证明唯一性的方式是通过秩的确定性，而最后则又从可逆阵的秩得到了可逆阵相抵标准形为 I ，于是其可以拆成初等变换阵的乘积。

自然，我们需要对较简单的形式有一定的期望。在相抵相关的问题中，我们至少希望它是一个对角阵——这里对角阵指除了 $i = j$ 的元素外全为 0 的矩阵——这样方程组 $Ax = b$ 仍然可以通过类似方法判定。

§15.2 整数方阵的相抵

15.2.1 可逆性

说到底，我们其实已经给出了用相抵解决问题的思路。想要这个思路被合理使用，我们至少还需要定义整数矩阵中的可逆矩阵与行列变换。

可逆矩阵仍然是易于定义的：对整数方阵 A ，若存在整数方阵 B 使得 $AB = BA = I$ ，则称 A 可逆 (也可称 A 为模方阵，此后将用这个词与通常的可逆区分)， B 是 A 的逆，记作 $B = A^{-1}$ 。若存在模方阵阵 P, Q 使得整数方阵 A, B 满足 $A = PBQ$ ，则称 A 与 B 相抵 (也可称模相抵，此后将用这个词与通常的相抵区分)。

* 直接验证可发现 I 是模方阵、 A 是模方阵则 $(A^{-1})^{-1} = A$ 、 A, B 是模方阵则 $(AB)^{-1} = B^{-1}A^{-1}$ ，由此可以得到模相抵是等价关系。

我们下面先说明，如果整数方阵 A 确实可以模相抵为对角阵，整线性方程组相关问题是易于解决的。对整线性方程组 $Ax = b$ ，设 $A = PDQ$ ，其中 P, Q 为模方阵， D 为对角阵，则方程组可以化为

$$PDQx = b$$

同时左乘 P^{-1} 得到

$$DQx = P^{-1}b$$

下面假设 D 的对角元满足 $d_1, \dots, d_r \neq 0$ ，其余为 0。由于 DQx 相当于把 Qx 的第 i 个元素乘 d_i 倍 (若 $m < n$ ，多出的元素乘 0)，有解当且仅当 $P^{-1}b$ 的第 $r + 1$ 到 n 个元素全为 0，且对 $i = 1, \dots, r$ ， $P^{-1}b$

第 i 个元素是 d_i 的倍数。此时整数通解可以写为

$$x = Q^{-1} \begin{pmatrix} c_1/d_1 \\ \vdots \\ c_r/d_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix}$$

其中 c_1 到 c_r 为 $P^{-1}b$ 前 r 个分量, x_{r+1} 到 x_n 可以取任何整数。

证明:

记 c_i 为 $P^{-1}b$ 的第 i 个分量, 并记 $y = Qx$, 则利用对角阵乘向量的结果, 删去 $0y_i = 0$ 的方程, 原方程组即 $d_1y_1 = c_1, \dots, d_r y_r = c_r$ 。由于 x 为整数解、 Q 为整数矩阵, y 一定为整数向量, 因此当 c_i 不为 d_i 倍数或 $i > r$ 且 c_i 非零时 y_i 无整数解, 从而原方程组无解; 反之, 若 $i > r$ 时 c_i 为 0, 且 c_i 是 d_i 的倍数, 则直接计算可发现如上方构造的 x 一定是整数解 (注意根据模方阵定义, P^{-1}, Q^{-1} 所有元素为整数)。

我们最后说明如上方构造的 x 是全部解。在 d_1, \dots, d_r 非零, 其他为 0 的假设下, $d_1y_1 = c_1, \dots, d_r y_r = c_r$ 都将确定 $y_i = d_i/c_i$ 。对于其他 y_i , 无论是 $0y_i = 0$ 还是未提到都可任取, 结合其为整数向量的要求, 即可取任何整数值, 因此上述形式就构成了 $Ax = b$ 的全部整数解。

这样的分析就说明了, 只要整数方阵的确可以模相抵为对角阵, 整线性方程组问题就将完全解决。比起之前定义抽象的模时的混沌, 这时我们已经有了一个明确的目标。但是, 如何找到 P 和 Q 呢? 相抵标准形的研究给了我们一个很不错的思路: 先进行行列变换, 再说明这相当于乘模方阵。据此, 我们需要先证明一些小结论 (关于上学期行列变换与初等方阵的结论见教材 4.2 节):

- 整数方阵是模方阵当且仅当其看作有理数方阵时可逆, 且逆矩阵的所有元素为整数。

证明:

若整数方阵 A 看作有理数方阵时可逆, 且逆矩阵 B 所有元素均为整数, 则由于矩阵乘法规则不变即得 $AB = BA = I$, 从而对于整数方阵仍有 B 为 A 的逆。

若存在整数方阵 B 使得 $AB = BA = I$, 根据逆的唯一性可得只能 B 也是 A 看作有理数方阵时的逆, 又由 B 全为整数可得证。

- 两个模相抵的整数方阵看作有理数方阵仍然相抵, 于是秩相等。

证明:

设整数方阵 $A = PBQ$, P, Q 为模方阵, B 为另一个整数方阵。由于 P, Q 看作有理数方阵仍然可逆, 就得到了 A, B 相抵。

- 整数方阵是模方阵当且仅当其行列式为 ± 1 。

证明:

利用上学期知识, “看作有理数方阵时可逆” 等价于行列式非零。由此, 我只需要说明, 对看作有理数方阵时可逆的整数方阵 A , 其逆 B 的所有元素为整数当且仅当 $\det A = \pm 1$ 。

若 $\det A = \pm 1$, 则有 $B = (\det A)^{-1}A^* = \pm A^*$ (这里 A^* 为伴随矩阵)。由于计算伴随矩阵只需要对 A 的元素进行加法、减法、乘法, A^* 的仍然为整数方阵, 从而 B 为整数方阵。

若 B 为整数方阵, 由于行列式计算方法不变, Binet-Cauchy 公式仍成立, 从而 $1 = \det(I) = \det(AB) = \det A \det B$ 。由于 A, B 为整数, 计算行列式只需要对元素进行加法、减法、乘法, $\det A, \det B$ 均为整数。两整数乘积为 1 即得到 $\det A = \pm 1$ 。

- 定义第一类整行列变换：将某行/列加上另一行/列的某整数倍，其可以看作乘模方阵（第一类初等方阵）。

证明：

根据上学期的初等方阵知识，第一类整行列变换可以看作左/右乘第一类初等方阵 $P(i, j(k))$ 。由于倍数为整数， k 为整数，从而其逆 $P(i, j(-k))$ 也为整数，从而得证其为模方阵。

- 定义第二类整行列变换：交换两行/两列，其可以看作乘模方阵（第二类初等方阵）。

证明：

根据上学期的初等方阵知识，第二类整行列变换可以看作左/右乘第二类初等方阵 $P(i, j)$ 。由于 $P(i, j)$ 与其逆 $P(i, j)$ 均为整数方阵，即得证其为模方阵。

- 定义第三类整行列变换：给某行/列乘 ± 1 ，其可以看作乘模方阵（第三类初等方阵）。

证明：

根据上学期的初等方阵知识，第三类整行列变换可以看作左/右乘第三类初等方阵 $P(i(c))$ 。由于 $c = \pm 1$ ， $P(i(c))$ 与其逆 $P(i(c^{-1}))$ 均为整数方阵，即得证其为模方阵。

* 上述讨论可以看出整行列变换只允许乘 ± 1 倍的原因：这是为了保证它可以看作乘模方阵。若允许乘任何整数倍，虽然在方程组意义下仍然进行了等价的变换，但无法看作乘模方阵，因此无法导向标准形。

下面，我们的任务就是利用这三类整行列变换将任何整数矩阵 A 化为对角阵——或者某个更简单的形式。

15.2.2 两种特殊情况

为了能将一般的整数方阵模相抵为对角阵，也为了给整行列变换提供操作范例，我们需要先考察两种特殊情况，一个是 $n \times 1$ 整数矩阵，另一个是 2×2 上三角整数矩阵。

* 之所以说这两种情况可以用于将一般整数方阵模相抵为与本讲义 10.4 证明奇异值分解的途径相同，都是为了先化作阶梯形再化作对角阵。

* 直接仿照普通行列变换的方式进行模相抵对角化是行不通的，因为在并非倍数的情况下，无法用非零元素消去其他元素。

1. 对于整数列向量 $(a_1, \dots, a_n)^T$ ，我们说明其可以整行列变换为 $(a, 0, \dots, 0)^T$ ，且 $a = \gcd(a_1, \dots, a_n)$ 。

证明可以利用归纳法，先考虑 $n = 2$ 时，若 a_1, a_2 有 0，将 0 换到第二个分量，并乘 ± 1 使得第一个分量非负则已经解决（利用 0 与任何整数最大公因数为其绝对值）。否则，可以发现辗转相除法的操作恰好符合行变换的要求：不妨设 $|a_2| \geq |a_1|$ ，否则交换两列。将第二行减第一行某倍数、再将第一行减第二行某倍数，重复此过程最终能得到某个位置为 0，再将其交换到第二个分量，并乘 ± 1 使得第一个分量非负，即得到结果。

由于辗转相除不改变最大公因数，当操作进行到其中一个为 0 时，另一个的绝对值一定为最大公因数，这就得到了结论。

若 $n = k$ 时结论成立，在 $n = k + 1$ 时，先通过前 k 行的行变换使得 $(a_1, \dots, a_{k+1})^T$ 变换为 $(\gcd(a_1, \dots, a_k), 0, \dots, 0, a_{k+1})^T$ ，再对第一行、最后一行按 $n = 2$ 的情况进行行变换，利用本讲义 14.1.3 的第五个结论即得到最终可变换为第一个分量是 $\gcd(a_1, \dots, a_{k+1})$ 、其他分量均为 0 的列向量。

* 从证明过程中，我们可以感受到最大公因数是行列变换操作的重要量。事实上，与辗转相除不改变最大公因数的证明完全相同，我们可以证明对 $(a_1, \dots, a_n)^T$ 进行行变换不会改变所有分量的最大公因数。因此，若其变换为了 $(a, 0, \dots, 0)^T$ 的形式，且 $a \geq 0$ ，则必然 $a = \gcd(a_1, \dots, a_n)$ 。

* 对于行向量的列变换有完全相同的结论，直接类似操作即可。

2. 对于二阶整数上三角阵 $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ ，我们说明其可以整行列变换为对角阵 $\text{diag}(x, y)$ 。

若 $b = 0$ ，原矩阵已经成为了对角阵；若 $a = 0$ ，与第一部分证明相同进行行变换即可将矩阵化为 $\text{diag}(0, \gcd(b, c))$ ；若 $c = 0$ ，完全同理（第一部分的证明也可以用在行向量的列变换中）可证能化为对角阵 $\text{diag}(\gcd(a, b), 0)$ 。由此，我们下面只考虑 $abc \neq 0$ 的情况。

将矩阵的两行按照第一部分证明进行行变换，可将第二列化为 $(\gcd(b, c), 0)^T$ 。由于这时第一列也进行了对应的行变换，设此时它变为了 $(s, t)^T$ ，则矩阵成为

$$\begin{pmatrix} s & \gcd(b, c) \\ t & 0 \end{pmatrix}$$

可以发现，通过交换两列，这又变回了一个上三角阵 $\begin{pmatrix} \gcd(b, c) & s \\ 0 & t \end{pmatrix}$ ，问题貌似并没有得到解决。但是，由于转化为最大公因式缩小了规模，我们可以想到从大小出发进行归纳：

- 当 $|a|, |b|, |c|$ 中最小值为 0 时，已经在之前讨论得到。
- 若 $|a|, |b|, |c|$ 中最小值小于 n 时结论成立，下面考虑最小值为 n 时。

在 $|c| = n$ 时，若 b 不为 c 的倍数，通过上方操作整行列变换为 $\begin{pmatrix} \gcd(b, c) & s \\ 0 & t \end{pmatrix}$ ，此时 $\gcd(b, c) < |c| = n$ ，于是化为了最小值小于 n 的情况，根据归纳假设可将其模相抵对角化。若 b 为 c 的倍数，第一行减第二行的 b/c 倍可直接化为对角阵。

在 $|a| = n$ 时，若 b 不为 a 的倍数，类似上方操作整行列变换为 $\begin{pmatrix} \gcd(a, b) & s \\ 0 & t \end{pmatrix}$ 同理化为了最小值小于 n 的情况，根据归纳假设可将其模相抵对角化。若 b 为 a 的倍数，第二列减第一列的 b/a 倍可直接化为对角阵。

在 $|b| = n$ 时，若 c 不为 b 的倍数或 a 不为 b 的倍数，仍然可以使得原矩阵整行列变换为 $\begin{pmatrix} \gcd(b, c) & s \\ 0 & t \end{pmatrix}$ 或 $\begin{pmatrix} \gcd(a, b) & s \\ 0 & t \end{pmatrix}$ ，从而最小值变得更小，由归纳假设成立。否则，设 $c = c_1 b$ 、 $a = a_1 b$ ，第一列减第二列的 a_1 倍、第二行减第一行的 c_1 倍，即可将原矩阵整行列变换为 $\begin{pmatrix} 0 & b \\ -a_1 c_1 b & 0 \end{pmatrix}$ ，再交换两行就成为了对角阵。

* 这个结论的证明里，我们可以发现模长估算对最大公因数相关问题的重要性。

理论来说，这两个定理已经足够将任何整数矩阵整行列变换为对角阵了。不过，为了之后讨论的方便，我们姑且开个“天眼”，证明另一个有用的结论：二阶整数对角阵 $\text{diag}(x, y)$ 可以整行列变换为 $\text{diag}(\gcd(x, y), \text{lcm}(x, y))$ 。

证明：

将第二列加上第一列可将原矩阵化为 $\begin{pmatrix} x & x \\ 0 & y \end{pmatrix}$ ，将两行按第一部分证明进行行变换则可以进一步化为 $\begin{pmatrix} s & \gcd(x, y) \\ t & 0 \end{pmatrix}$ ，这里 s 与 t 是第一列进行对应行变换后的结果。

在第一部分中我们已经证明，对第一列进行行变换不会改变第一列的最大公因数，因此 $(s, t) = x$ 成立，于是 s 是 x 的倍数，进而是 $\gcd(x, y)$ 的倍数。将第一列减第二列的 $s/\gcd(x, y)$ 倍（若 $\gcd(x, y) = 0$ 则必然 $s = 0$ ，无需再减）并交换两列即可将原矩阵化为 $\text{diag}(\gcd(x, y), t)$ 。

利用行列式知识，整行列变换只会使行列式乘 -1 或不变，因此必然有 $\gcd(x, y)t = xy$ 。若 $xy \neq 0$ ，即得 $t = \pm \text{lcm}(x, y)$ ，有负号再给第二行乘 -1 即可；若 $x = 0$ 或 $y = 0$ ， $\gcd(x, y) \neq 0$ 且 $xy = 0$ ，必有 $t = 0 = \text{lcm}(x, y)$ ；若 $x = y = 0$ ，原矩阵为零矩阵，直接符合要求。

* 本部分的三个结论在下一部分分别称为引理 1、2、3。可以发现，三个结论中，引理 2 事实上是最重要的，因为它克服了本质的困难：如何将不同行列的元素关联起来。正是因为引理 2 巧妙利用了模长进行归纳，我们才可以将一般的整数矩阵进行模相抵对角化。

15.2.3 构造性证明

准备到这里，我们终于可以开始对整数方阵的模相抵标准形进行刻画了。结论如下：任何整数矩阵 $A \in \mathbb{Z}^{m \times n}$ 可以整行列变换为

$$\begin{pmatrix} \text{diag}(d_1, \dots, d_r) & O \\ O & O \end{pmatrix}$$

其中 $r = \text{rank } A$ ，且正整数 d_1, \dots, d_r 满足 $d_1 | d_2 | d_3 | \dots | d_r$ （连续整除号表示每一个都整除后一个）。我们将此形式称为整数矩阵的相抵标准形/模相抵标准形，或称为 Smith 标准形。

* 感兴趣的同学可以参考本讲义 10.4.2 的奇异值分解存在性证明，观察两个证明的相似性。

证明：

不妨设 $m \geq n$ ，否则将下面证明中的行变换改为列变换，列变换改为行变换即可。分三部分进行证明。

1. A 可以整行列变换为上三角阵 B ，且 B 中除了满足 $i = j$ 与 $i = j - 1$ 的 b_{ij} 均为 0。

利用引理 1，进行行变换使得第一列中 a_{21} 到 a_{m1} 全部变换为 0，再对第 2 到 n 列进行列变换使得 a_{13} 到 a_{1n} 全部变换为 0。这时，矩阵 A 成为了（这里 * 表示未知元素，未写出元素为 0）

$$\begin{pmatrix} a_{11} & a_{12} & & & \\ * & * & \cdots & * & \\ \vdots & \vdots & \vdots & \vdots & \\ * & * & \cdots & * & \end{pmatrix}$$

接下来，对第 2 到 m 行进行行变换，使得第二列中 a_{32} 到 a_{m2} 全被变换为 0。由于第一列涉及的行已经全部为 0，变换不会改变它们为 0 的性质，从而 A 这时变为了

$$\begin{pmatrix} a_{11} & a_{12} & & & \\ a_{22} & * & \cdots & * & \\ \vdots & \vdots & \vdots & \vdots & \\ * & \cdots & * & & \end{pmatrix}$$

再对第 3 到 n 列进行列变换……重复这个过程（由于 $m \geq n$ 的假设，这个过程一定可以操作至 a_{nn} ），我们最终让 A 除了 a_{ii} 与 $a_{i,i+1}$ 外的元素全部为 0，这就是符合要求的矩阵 B 。

2. B 可以整行列变换为对角阵 C ，即 C 中除了 c_{11}, \dots, c_{nn} 外的元素全为 0。

本部分证明的思路类似引理 2（注意直接由引理 2 进行操作是不可行的，例如对 B 的前两行、前两列进行整行列变换，使得 $b_{12} = 0$ 会由于原本的 b_{23} 可能非零导致 b_{13} 可能变为非零）。对所有元素的绝对值最小值进行归纳，只要有元素不是绝对值最小元素的倍数，就能化为更小的情况，从而符合归纳假设。

3. C 可以整行列变换为符合 Smith 标准形要求的矩阵 D 。

首先，根据之前已证，有 $\text{rank } C = \text{rank } A$ ，因此 C 的非零对角元素个数应为 r ，利用第二类整行列变换将它们交换到 c_{11}, \dots, c_{rr} 。

由于引理 3 同样只涉及两行、两列，对于对角阵 C ，其应用在任何两个对角元上都不会影响其他元素。对 c_{11}, c_{22} 进行操作可使第一个对角元成为 $\gcd(c_{11}, c_{22})$ ，此时再对新的 c_{11} 与 c_{33} 进行操作可使第一个对角元成为 $\gcd(\gcd(c_{11}, c_{22}), c_{33}) = \gcd(c_{11}, c_{22}, c_{33})$ 。由此，最终可以将第一个对角元操作为所有对角元的最大公因数。

接下来，类似此方法，可使得第二个对角元成为第 2 到 r 个对角元的最大公因数。由于过程中只对首行首列外的行列进行了变换，由于 c_{22}, \dots, c_{rr} 都是 c_{11} 的倍数，可直接验证整行列变换后它们仍然是 c_{11} 的倍数。因此，此时第一个对角元仍然是第 1 到 r 个对角元的最大公因数。

以此类推，最终可以使得第 i 个对角元是第 i 到 r 个对角元的最大公因数。设此时对角元为 d_1, \dots, d_r ，根据最大公因数的定义（第 i 个对角元是所有后续对角元的最大公因数，因此是第 $i+1$ 个对角元的因数），这就说明了 $d_1 | d_2 | \dots | d_r$ 。

* 如同本讲义 10.4.2 中说的，证明第一部分里的矩阵 B 形式称为 **Hessenberg 阵**。

在上方的证明过程中，我们通过一个相对明确的构造性算法证明了任何整数矩阵都可以整行列变换为 Smith 标准形的形式。为了说明它的确是模相抵的标准形，我们还需要证明乘模方阵可以看作进行行列变换，且此标准形对任何矩阵唯一确定。

值得一提的是，哪怕这两件事并未证明，我们也足够进行具体整线性方程组的求解了：由于整行列变换可以看作乘模方阵，我们已经证明了对任何整数矩阵 A 都存在模方阵 P, Q 使得 PAQ 是对角阵。不过，就像线性方程组解的存在性可以归结为 $\text{rank}(A, b) = \text{rank } A$ ，我们希望能有一个简洁的判定定理说明整线性方程组是否有解，而不是必须每次都作出 A 的 Smith 标准形。因此，我们必须继续进行研究。

15.2.4 多项式矩阵

由于整数与多项式完全类似的性质，将对于模长的估算换成对次数的估算后，上方的所有结论都可以推广到多项式相关的讨论上。唯一需要注意的是，正如本讲义 14.1 介绍的“代表元”思想，整数中涉及绝对值的部分在多项式中应替换为首一多项式。本部分我们只列出定义与结论，证明几乎与前几部分完全相同：

- 若一个 m 行 n 列矩阵 A 的所有元素都是数域 \mathbb{K} 上关于 λ 的多项式，则称其为 \mathbb{K} 上的多项式矩阵（或 λ 矩阵），所有 m 行 n 列 \mathbb{K} 上的多项式矩阵构成的集合记作 $\mathbb{K}[\lambda]^{m \times n}$ 。
- 由于元素的加法、减法、乘法都可以进行， \mathbb{K} 上的多项式矩阵的加法、乘法、行列式的定义与常数矩阵时完全相同。
- 称 \mathbb{K} 上的多项式矩阵的秩为其最大非零子式的阶数。

* 当然，我们仍然可以从模与模上线性相关的语言得到秩的定义，此处省略。

- 对 \mathbb{K} 上的多项式方阵 A ，若存在 \mathbb{K} 上的多项式方阵 B 使得 $AB = BA = I$ （注意 \mathbb{K} 上的方阵都可以看作 \mathbb{K} 上的多项式方阵，只是元素没有超过零次的多项式），则称 A 可逆（或为模方阵）， B 是 A 的逆，记作 $B = A^{-1}$ 。
- 若存在模方阵 P, Q 使得 \mathbb{K} 上的多项式方阵 A, B 满足 $A = PBQ$ ，则称 A 与 B 相抵（或模相抵）。这是一个 \mathbb{K} 上同阶多项式方阵间的等价关系。
- \mathbb{K} 上的多项式方阵是模方阵当且仅当其行列式为 \mathbb{K} 中的非零元素。

* 右推左仍然是通过伴随方阵直接构造，左推右也仍然是先利用 Binet-Cauchy 公式，再根据次数计算得到两个多项式相乘为 1 则它们都为非零常数。

- \mathbb{K} 上的多项式方阵 A, B 模相抵，则它们的秩相等。

* 这个结论最简单的证明方法是像整数方阵一样“看作有理数方阵”讨论：我们考虑每个元素是 λ 的函数（允许在有限个点没有定义）的矩阵，这样所有的非零多项式就都可逆了。否则，对其的证明只能在学习行列式因子之后了，见下节。

- 定义第一类 \mathbb{K} 上多项式方阵的行列变换：将某行/列加上另一行/列的某 \mathbb{K} 上多项式倍，其可以看作乘模方阵（第一类初等方阵）。
 - 定义第二类 \mathbb{K} 上多项式方阵的行列变换：交换两行/两列，其可以看作乘模方阵（第二类初等方阵）。
 - 定义第三类 \mathbb{K} 上多项式方阵的行列变换：给某行/列乘 \mathbb{K} 上的某非零元素，其可以看作乘模方阵（第三类初等方阵）。
- * 这里要求乘 \mathbb{K} 上某非零元素的原因与整数时要求乘 ± 1 的原因相同，用本讲义 14.1 的语言，这是因为它们是环上的单位。
- 任何 \mathbb{K} 上的多项式矩阵 $A \in \mathbb{K}[\lambda]^{m \times n}$ 可以行列变换为

$$\begin{pmatrix} \text{diag}(d_1, \dots, d_r) & O \\ O & O \end{pmatrix}$$

其中非零首一多项式 d_1, \dots, d_r 满足 $d_1 | d_2 | \dots | d_r$ 。我们将此形式称为 \mathbb{K} 上多项式矩阵的相抵标准形/模相抵标准形，或称为 Smith 标准形。

* 由于模相抵的多项式方阵秩相等并不好证明，我们暂时规避了 r 就是 A 的秩这一相对自然的结论。这些细节将在介绍行列式因子时补充。

§15.3 Smith 标准形

15.3.1 另一种构造方式

本部分我们先对上一节内容做一点补充，提供一个操作性稍弱、理论性稍强的证明方式（也就是上课所学的证明方式）。这个证明中，对规模大小的控制仍然是核心要素，此外，为了方便讨论，我们将直接对 \mathbb{K} 上的多项式矩阵进行证明。

证明：

证明的核心思路归结于如下引理：对所有元素不全为 0 的 \mathbb{K} 上的多项式方阵 A ，可以进行行列变换使得它的某个元素 $d = \gcd(A)$ ，这里最大公因式表示对 A 所有元素取最大公因式。

记 $\min \deg(A)$ 为 A 中所有非零元素的最低次数，我们希望通过对其进行归纳证明这个结论。当 $\min \deg(A) = 0$ 时， A 中有元素是非零常数 c ，因此所有元素的最大公因式必然为 1 (c 的因式只有非零常数)，直接将 c 所在行乘 c^{-1} 倍就得到了 1（注意乘非零常数倍是多项式矩阵的第三类行列变换）。

若 $\min \deg A < n$ 时结论成立，下面考虑其等于 n 时。通过交换行列，可不妨设左上角元素 a_{11} 满足 $\deg a_{11} = n$ 。若所有其他元素已经是 a_{11} 的倍数，则第一行乘非零倍数使 a_{11} 成为首一多项式后结论已经成立。否则，分类讨论：

1. 若首行某元素 a_{1i} 不是 a_{11} 的倍数，类似上节引理 1 对 1、 i 两列列变换可将 a_{11} 变换为 $\gcd(a_{11}, a_{1i})$ ，其次数小于 a_{11} ，从而由归纳假设成立；
2. 若首列元素 a_{i1} 不是 a_{11} 的倍数，类似上节引理 1 对 1、 i 两行行变换可将 a_{11} 变换为 $\gcd(a_{11}, a_{i1})$ ，其次数小于 a_{11} ，从而由归纳假设成立；

3. 若首行首列之外的某元素 a_{ij} 不是 a_{11} 的倍数, 由于 a_{i1}, a_{1j} 是 a_{11} 的倍数, 可第 i 行减第一行的 a_{i1}/a_{11} 倍, 第 j 列减第一列的 a_{1j}/a_{11} 倍。这样操作后 (用上标 new 表示操作后的数), a_{ij}^{new} 为 $a_{ij} - a_{i1}a_{1j}/a_{11}$, 由于 a_{i1}, a_{1j} 是 a_{11} 的倍数, 第二项是 a_{11} 的倍数, 因此 a_{ij}^{new} 仍然不是 a_{11} 的倍数, 且 $a_{1j}^{new} = a_{i1}^{new} = 0, a_{11}^{new} = a_{11}$ 。类似上节引理 3 对 $1i$ 两行 $1j$ 两列进行变换即可使得左上角成为 $\gcd(a_{11}, a_{ij}^{new})$, 从而由归纳假设成立。

利用此引理, 若 A 全为 0, 已经是 Smith 标准形, 否则可以变换出 $\gcd(A)$, 将其通过交换行列移动到 a_{11} 的位置, 由于其他元素都是它的倍数, 通过第一类行变换可消去第一列的其他元素, 通过第一类列变换可消去第一行的其他元素, 这时 A 成为

$$\begin{pmatrix} \gcd(A) & 0 \\ 0 & A_2 \end{pmatrix}$$

对 A_2 重复上述操作, 直到没有剩下的部分或剩下的部分为零矩阵即可。这时 A 一定将成为

$$\begin{pmatrix} \text{diag}(d_1, \dots, d_r) & O \\ O & O \end{pmatrix}$$

最后我们证明它符合 Smith 标准形的要求。操作完成后, 由于所有 d_i 非零, 且它们是某些多项式的最大公因式, 它们一定是非零首一多项式。

此外, 直接验证可发现一些 d 的倍式在行列变换后仍然是 d 的倍式, 因此根据最大公因式定义 d_1 仍然是整个矩阵的最大公因式, 从而 $d_1 | d_2$ 。同理, 由于操作是在不断重复将左上角化为最大公因式的过程, 有 $d_2 | d_3, \dots, d_{r-1} | d_r$, 这就证明了它确实是 Smith 标准形。

* 此证明中的归纳法事实上完全类似之前证明引理 2 时所用的归纳, 都是从规模出发进行控制。

接下来, 我们证明上节欠缺的两件事之一: 任何模方阵都可以写为一系列能看作行列变换的模方阵的乘积, 这就得到了类似相抵标准形时的结论: A 可以行列变换成为 B 等价于 A 可以乘模方阵得到 B , 从而两种方式定义的等价关系事实上是同一种, 今后即可以混用。同样只针对 \mathbb{K} 上的多项式方阵进行证明, 整数方阵的情况可以类似证明。

证明:

先证明 \mathbb{K} 上的多项式模方阵 A 可以通过行列变换化为 I 。由于行列变换可以看作乘模方阵、模方阵的乘积是模方阵, 对 n 阶模方阵 A , 可设存在模方阵 P, Q (P, Q 都是一些行列变换对应模方阵的乘积) 使其满足

$$PAQ = \text{diag}(d_1, \dots, d_r, 0, \dots, 0)$$

这里我们使用了 $\text{diag}(d_1, \dots, d_r, 0, \dots, 0)$, 是因为对方阵而言, Smith 标准形必然是严格意义的对角阵。

利用模方阵当且仅当行列式为非零常数、模方阵乘积是模方阵, 左侧的行列式为 \mathbb{K} 中的非零常数 c , 从而首先有 $r = n$ (否则右侧行列式为 0), 且

$$d_1 \dots d_n = c$$

由于 d_1, \dots, d_n 都非零, 考虑次数可知只能它们都为零次, 而零次的首一多项式即为 1, 于是 $d_1 = \dots = d_n = 1$, 也即

$$PAQ = I$$

于是得证。

接下来, 由于行列变换可以看作乘模方阵, 将 P, Q 展开得到

$$P_1 \dots P_m A Q_1 \dots Q_n = I$$

于是 (这里每个 P_i, Q_i 都是能看做行列变换的模方阵)

$$A = P_m^{-1} \dots P_1^{-1} Q_n^{-1} \dots Q_1^{-1}$$

之前已验证行列变换阵的逆仍是行列变换阵，从而得到了结论。

结合这部分证明以后，我们最终可以得到模相抵标准形的确是模相抵下的某种标准形式：对任何 \mathbb{K} 上的多项式方阵 A ，存在模方阵 P, Q 使得

$$PAQ = \begin{pmatrix} \text{diag}(d_1, \dots, d_r) & O \\ O & O \end{pmatrix}$$

其中非零首一多项式 d_1, \dots, d_r 满足 $d_1 | d_2 | \dots | d_r$ 。

当然，由于模方阵的逆还是模方阵，利用消去可知也存在模方阵 P, Q 使得

$$A = P \begin{pmatrix} \text{diag}(d_1, \dots, d_r) & O \\ O & O \end{pmatrix} Q$$

其中非零首一多项式 d_1, \dots, d_r 满足 $d_1 | d_2 | \dots | d_r$ 。

15.3.2 行列式因子

在说清楚了“能行列变换得到”与“能左右乘可逆阵得到”对应的等价关系是同一种后，我们自然希望上述的标准形能够完全刻画模相抵性质，也即任何矩阵只能化作唯一一个 Smith 标准形。

由于模相抵是一个等价关系，上述结论又可以化为，两个不同的 Smith 标准形不模相抵。我们对这两个表述的等价性进行一个简单的证明，作为等价关系概念的复习。今后我们还会时常使用等价关系的性质，那时对标准形唯一性的证明都将直接从“不同标准形不等价”出发，忽略这段讨论过程。

证明：

若 Smith 标准形 P, Q 模相抵，则 P 既能化作自身（不变），也能化作 Q ，因此同一个矩阵会化成两个不同的 Smith 标准形。

若某矩阵 A 同时与 Smith 标准形 P, Q 模相抵，从 A 与 P 模相抵出发由等价关系的对称性可知 P 与 A 模相抵，而 A 与 Q 模相抵，由等价关系的传递性可知 P 与 Q 模相抵。

想要证明唯一性，最直接也是最基本的办法是，假设 S, T 都是 $\mathbb{K}[\lambda]^{m \times n}$ 中符合 Smith 标准形形式的多项式方阵，证明矩阵方程组 $PSQ = T$ 不存在模方阵 P, Q 为解。事实上，教材 6.2 节的定理 1（惯性定理）就是通过这样直接的方式证明了实对称阵的规范形唯一。

但是，这样的方法在此处会变得非常困难：两个模方阵提供了过多的可变化量，导致无解性的判定更加复杂。不过，证明仍然是可以进行的。我们的核心思路是，证明两个整数/多项式相等往往需要从相互整除出发。

证明：

设同阶 Smith 标准形

$$S = \begin{pmatrix} \text{diag}(d_1, \dots, d_s) & O \\ O & O \end{pmatrix}, \quad T = \begin{pmatrix} \text{diag}(c_1, \dots, c_t) & O \\ O & O \end{pmatrix}$$

模相抵，即存在模方阵 P, Q 使得 $S = PTQ$ 。

根据 Binet-Cauchy 公式，考虑 S 的左上角 k 阶主子式（若 $k > s$ ，记 $d_k = 0$ ， $k > t$ 时也记 $c_k = 0$ ）构成的行列式有

$$d_1 \dots d_k = \sum_{\substack{1 \leq i_1 < \dots < i_k \leq m \\ 1 \leq j_1 < \dots < j_k \leq n}} \det P \begin{pmatrix} 1 & \cdots & k \\ i_1 & \cdots & i_k \end{pmatrix} \det T \begin{pmatrix} i_1 & \cdots & i_k \\ j_1 & \cdots & j_k \end{pmatrix} \det Q \begin{pmatrix} j_1 & \cdots & j_k \\ 1 & \cdots & k \end{pmatrix}$$

这里第二项代表 T 的 i_1, \dots, i_k 行、 j_1, \dots, j_k 列交出的子矩阵的行列式，其他同理。由于 T 是对角阵，只要 i_1, \dots, i_k 、 j_1, \dots, j_k 不完全相同，此行列式必然为 0，其余情况行列式为对角元乘积，从而可化简为（注意求和中为多项式的乘法而非矩阵乘法，可以任意交换，将中间一项交换到了开头）

$$d_1 \dots d_k = \sum_{1 \leq i_1 < \dots < i_k \leq \min(m, n)} c_{i_1} \dots c_{i_k} \det P \begin{pmatrix} 1 & \cdots & k \\ i_1 & \cdots & i_k \end{pmatrix} \det Q \begin{pmatrix} i_1 & \cdots & i_k \\ 1 & \cdots & k \end{pmatrix}$$

利用 c_i 的互相整除性，任何 k 个对角元的乘积都是 $c_1 \dots c_k$ 的倍数（利用 0 是任何多项式的倍数），从而右侧为 $c_1 \dots c_k$ 的倍数，因此

$$c_1 \dots c_k \mid d_1 \dots d_k$$

利用模相抵的对称性，完全同理可以证明 $d_1 \dots d_k \mid c_1 \dots c_k$ 。这两个式子都对 $k = 1, \dots, \min(m, n)$ 成立。

由于 c_i 与 d_i 或为 0，或为非零首一多项式，它们的乘积也或为 0，或为非零首一多项式，因此相互整除即可推出相等（见本讲义 14.1.2），有

$$c_1 \dots c_k = d_1 \dots d_k, \quad k = 1, 2, \dots, \min(m, n)$$

若 $s \neq t$ ，取 $k = \min(s, t) + 1$ ，一侧为 0，一侧不为 0，矛盾，因此 $s = t$ ，记 $s = t = r$ 。由于非零性，直接归纳即可证明 $c_i = d_i$ 对 $i = 1, \dots, r$ 成立，这就证明了 $S = T$ 。

不过，上面的 Binet-Cauchy 计算还是过于麻烦了。仿照相抵标准形唯一性的证明，唯一性还有另一个更加“高观点”的方式，也就是不变量。我们提取相抵标准形唯一性的证明逻辑：

- 定义矩阵的秩；
- 证明行变换与列变换不改变秩；
- 证明可逆矩阵可以拆分成初等方阵的乘积；
- 综合二、三两部分得到乘可逆阵不改变秩；
- 证明每个相抵标准形对应唯一的秩；
- 综合四、五两部分得到不同的相抵标准形不相抵。

对于模相抵而言，我们已经说明了只谈论秩是不够的，且秩作为行/列向量极大线性无关组所含向量个数的定义也无法很好拓展（之前对模的讨论完全不充分）。不过，秩的最大非零子式定义可以给出一点启发，暗示我们需要定义的不变量和子式有关。另一方面，我们前面的证明中反复提及， A 的 Smith 标准形的首个对角元 d_1 必然为 $\gcd(A)$ （行列变换不改变所有元素的最大公因式），而所有元素也恰好是所有一阶子式。由此，我们希望进行如下的证明：

- 定义 \mathbb{K} 上的 $m \times n$ 多项式矩阵的行列式因子，第 k 个行列式因子为所有 k 阶子式的最大公因式， $k = 1, \dots, \min(m, n)$ ；
- 证明多项式矩阵的行变换与列变换不改变任何行列式因子；
- 证明模方阵可以拆分成初等方阵的乘积；
- 综合二、三两部分得到乘模方阵不改变行列式因子；
- 证明每个 Smith 标准形对应唯一一组行列式因子；
- 综合四、五两部分得到不同的 Smith 标准形不模相抵。

这六条过程里，只有第二、三、五部分是需要证明的。模方阵可以拆分成初等方阵的乘积已在之前证明，此处我们将证明最关键的一步，多项式矩阵的行列变换不改变任何行列式因子。

证明：

对三类行列变换分别说明即可，由于转置不影响行列式，我们只针对列变换说明：

1. 第二类列变换使得任何 k 阶子式不变或乘 -1 ，由于 -1 看作 \mathbb{K} 上的多项式可逆由定义因式不变，不影响整体最大公因式；

2. 第二类列变换使得任何 k 阶子式不变或乘 \mathbb{K} 中非零常数，由于 \mathbb{K} 中非零常数看作 \mathbb{K} 上的多项式可逆，由定义因式不变，不影响整体最大公因式。
3. 对第一类列变换，设它将第 i 列加上了第 j 列的 f 倍。若 k 阶子式同时包含这两列或不包含第 i 列，则不变（前者是因为行列式的行列变换性质），下面只考虑包含第 i 列而不包含第 j 列的子式，由于进行第二类列变换不影响整体最大公因式，我们可以不妨设考虑的子式是左上角 k 阶子式，第 k 列加上了第 $k+1$ 行的 f 倍。

设前 k 行构成子矩阵的前 $k+1$ 列为 $\alpha_1, \dots, \alpha_{k+1}$ ，则新的 k 阶子式为

$$\det(\alpha_1, \dots, \alpha_k + f\alpha_{k+1})$$

由行列式性质可知其为

$$\det(\alpha_1, \dots, \alpha_k) + \det(\alpha_1, \dots, \alpha_{k-1}, f\alpha_{k+1}) = \det(\alpha_1, \dots, \alpha_k) + f \det(\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+1})$$

由于 $\det(\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+1})$ 仍然为 A 的子式，利用最大公因式的性质可知

$$\gcd(\det(\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+1}), \det(\alpha_1, \dots, \alpha_k + f\alpha_{k+1}))$$

即为

$$\gcd(\det(\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+1}), \det(\alpha_1, \dots, \alpha_k))$$

对每个产生了变化的子式都如此考虑可发现整体最大公因式不变。

* 此定理也可以得到模相抵不改变秩：所有 k 阶子式全为 0 当且仅当他们的最大公因式为 0。这就进一步得到了 Smith 标准形中的 $r = \text{rank } A$ 。

* 大部分情况下，对于唯一性都可以采用直接方式证明，但对复杂的情况，就可能需要通过将等价关系看作操作、寻找操作中的不变量来间接证明。虽然两种方法本质上是等价的，但不变量能够更好刻画根本性质。

15.3.3 不变因子

为了完成不变量思路对相抵标准形唯一性的证明，我们最后说明 Smith 标准形可由行列式因子唯一确定。由于 Smith 标准形是对角的，将它的第 i 个对角元称为第 i 个不变因子， $i > r$ 时为 0。

证明：

记第 i 个行列式因子为 D_i ，第 i 个不变因子为 d_i ，利用 Smith 标准形对角元的互相整除性（注意 0 是任何多项式的倍数，几个 0 的最大公因式是 0），且对角阵除了主子式外的子式均为 0，可得其所有 k 阶行列式的最大公因式为 $d_1 \dots d_k$ 对应的首一多项式。而由于它们为 0 或非零首一多项式，对应的首一多项式为自身，从而

$$D_k = d_1 \dots d_k, \quad k = 1, 2, \dots, \min(m, n)$$

当所有 d_i 给定时，所有 D_i 自然给定。反之，若所有 D_i 给定，根据形式可发现，记使 D_i 非零的最大 i 为 r ，则 $d_r \neq 0, d_{r+1} = 0$ ，此后 $d_i = 0$ 。而再根据

$$d_1 = D_1, \quad d_i = \frac{D_i}{D_{i-1}}, \quad i = 2, \dots, r$$

即可确定所有 d_i 。这就说明同阶矩阵的不变因子与行列式因子有一一对应的关系，被 Smith 标准形唯一确定。

* 为什么 Smith 标准形的对角元叫不变因子？这是因为对于 Smith 标准形有结论：对矩阵 $A \in \mathbb{K}[\lambda]^{m \times n}$ ，考虑包含 \mathbb{K} 的数域 \mathbb{F} ，则也有 $A \in \mathbb{F}[\lambda]^{m \times n}$ 。但是，无论是在 $\mathbb{K}[\lambda]^{m \times n}$ 中还是在 $\mathbb{F}[\lambda]^{m \times n}$ 中， A 的不变因子是相等的，从而 Smith 标准形相同。

证明：

不变因子可以被行列式因子确定，只需说明行列式因子相等即可。在本讲义 14.1.4 的最后我们提到，由于辗转相除法过程的相同，两个 $\mathbb{K}[x]$ 上多项式的最大公因式在看作 $\mathbb{F}[x]$ 上的多项式时是不变的。而 A 看作不同数域上的多项式矩阵时任何子式的行列式不会改变，因此所有行列式的最大公因式不改变，得证。

有了不变因子后，我们最开始关于整线性方程组解的存在性的结论就可以归为如下定理：对矩阵 $A \in \mathbb{Z}^{m \times n}$ 与向量 $b \in \mathbb{Z}^m$ ， $Ax = b$ 有整数解当且仅当 (A, b) 与 A 有相同的不变因子组。

* 由于 (A, b) 与 A 不同阶，我们无法说它们的 Smith 标准形相同，但不变因子组仍然可以定义，只要将所有 $i > r$ 的不变因子视为 0，就可以谈论不变因子组相同。

读者可以先自行思考，我们将在前半学期复习时证明这个定理，并从 Smith 标准形出发彻底解决整线性方程组的可解性与解集结构问题。

十六 补充：复数域的相似标准形

* 为了保证同一模块内容的完整性，这几章的内容编排与习题课上课顺序稍有不同，如本讲义 15.3 的部分内容是与本章在同一次习题课介绍的，本章最后一部分是在下一次习题课介绍的。

§16.1 特征方阵

本节的三部分本质上来说都是在说明同一个定理：对于数域 \mathbb{K} 上的方阵 A, B ， A 与 B 相似当且仅当 $\lambda I - A$ 与 $\lambda I - B$ 作为 $\mathbb{K}[\lambda]$ 上的方阵相抵（我们仍然将后一种相抵称为模相抵）。此后我们称 $\lambda I - A$ 为 A 对应的特征方阵。本节中，为区分数量阵与多项式矩阵，用大写字母表示 \mathbb{K} 上的方阵， $U(\lambda), V(\lambda)$ 等表示 $\mathbb{K}[\lambda]$ 上的方阵，若可逆（仍然称为模方阵），将其逆记作 $U(\lambda)^{-1}$ 。

* 注意 $\lambda I - A$ 是对角元素为首——一次多项式 $\lambda - a_{ii}$ ，其他元素为常数的多项式方阵，它的很多特殊性质来源于此。此外，这里的 λ 在进行形式计算时可完全类似数乘进行处理，虽然表示的并不是数而是抽象变量。

16.1.1 特征方阵的相抵

首先，我们将给出“仅当”部分（也就是左推右）的证明，并试着理解这个结论——毕竟，直观来说相抵和相似应该是完全不同的概念，我们必须解释此处它们通过了何种方式联系到了一起。

证明：

若 \mathbb{K} 上方阵 A, B 相似，存在 \mathbb{K} 上可逆方阵 P 使得 $P^{-1}AP = B$ ，直接计算有

$$P^{-1}(\lambda I - A)P = \lambda P^{-1}P - P^{-1}AP = \lambda I - B$$

由于 \mathbb{K} 上的方阵可以看作 $\mathbb{K}[\lambda]$ 上的方阵， \mathbb{K} 上的可逆方阵符合 $\mathbb{K}[\lambda]$ 上模方阵的要求，因此也是 $\mathbb{K}[\lambda]$ 上的模方阵。模方阵的逆也是模方阵，这就说明了 $\lambda I - A$ 与 $\lambda I - B$ 模相抵。

为了看出多项式方阵的模相抵与通常相抵的区别，我们先给出一个结论：即使对任何 $x \in \mathbb{K}$ 有 $xI - A$ 、

$xI - B$ 作为 \mathbb{K} 上方阵相抵， A 与 B 也未必相似。反例需要在四阶方阵构造，考虑 \mathbb{C} 上的四阶方阵：

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

计算可得这两个方阵对应的 $xI - A$ 、 $xI - B$ 在 $x = 0$ 时秩均为 2，否则均可逆。至于它们的不相似性，读者可以自己列方程计算，或在学习 Jordan 标准形后直接得到。

这暗示着， $\lambda I - A$ 与 $\lambda I - B$ 作为多项式方阵的模相抵与取定 λ 后作为数量阵的相抵是有本质区别的。进一步观察，如果对任何 λ 都有 $\lambda I - A$ 与 $\lambda I - B$ 作为数量阵相抵，我们也能对每个 λ 找到 $P(\lambda)$ 与 $Q(\lambda)$ 使得 $P(\lambda)(\lambda I - A)Q(\lambda) = \lambda I - B$ ，但此时 $P(\lambda), Q(\lambda)$ 对应的函数未必是多项式。

对比多项式和一般函数的区别，从大家高等数学里学过的知识可以看出，多项式具有非常好的光滑性：它可以求任何阶导数，且在某阶导数之后一直为 0。所以，或许从相抵到相似就是需要构建某些更加光滑变化的过渡矩阵 P, Q 。

我们当然可以进一步追问，为什么光滑变化的过渡矩阵可以带来相似。这里提供一点纯个人的思考（完全不保证正确，只提供思路），考虑 \mathbb{R} 上的方阵与多项式方阵。假设 $P(\lambda), Q(\lambda)$ 每个分量都足够光滑（如为多项式函数，这样它们就是多项式方阵），且满足

$$P(\lambda)(\lambda I - A)Q(\lambda) = \lambda I - B$$

由于 λ 可以随意变化，当 λ 非常大（趋于正无穷）时， A, B 几乎可以忽略，此时应有

$$P(\lambda)(\lambda I)Q(\lambda) \approx \lambda I$$

也即 $P(\lambda)Q(\lambda) \approx I$ 。同理，当 λ 非常小（趋于负无穷）时，也有 $P(\lambda)Q(\lambda) \approx I$ 。

非常巧合的是，当 $\lambda \rightarrow -\infty$ 时，将精确方程写成

$$P(\lambda)(-\lambda I + A)Q(\lambda) = -\lambda I + B$$

可发现此时左右相比 $-\lambda I$ 的“误差”为 $+A, +B$ ，恰好与 $\lambda \rightarrow \infty$ 时的 $-A, -B$ 相反。因此，其中可能存在某个点使得误差能够“抵消”，真的达到 $P(\lambda_0)Q(\lambda_0) = I$ 。这时就得到了 $\lambda_0 I - A$ 与 $\lambda_0 I - B$ 相似，从而根据上学期知识可得 A 与 B 相似。

* 事实上，我们最后构造出的相似过渡矩阵形式确实接近 $P(\lambda_0)$ ，但还是有很大的区别，在从 $AX = XB$ 出发的证明中可以观察到

* 虽然上面整段可能都是在胡扯，但本质是希望大家对形式简洁、结果不太寻常的结论有一些自己的思考与理解，而不是单纯背诵。

下面，我们将通过两种方法对右推左给出证明。无论是哪种方法，都是希望从模相抵出发直接构造出 P 使得 $P^{-1}AP = B$ 。在接下来两部分中，我们假设 $A, B \in \mathbb{K}^{n \times n}$, $U(\lambda), V(\lambda) \in \mathbb{K}[\lambda]^{n \times n}$ 为模方阵，且

$$\lambda I - A = U(\lambda)(\lambda I - B)V(\lambda)$$

16.1.2 带余除法出发的证明

课上所学的证明是基于多项式矩阵的带余除法的，归结为存在如下的分解：

$$U(\lambda) = (\lambda I - A)Q(\lambda) + U_0$$

$$V(\lambda) = R(\lambda)(\lambda I - A) + V_0$$

这里 $P(\lambda), Q(\lambda) \in \mathbb{K}[\lambda]^{n \times n}$ 、 $U_0, V_0 \in \mathbb{K}^{n \times n}$ 。

我们证明一个更一般的结果：对所有 k 考虑每个位置的 λ^k 次项可将 \mathbb{K} 上的 n 阶多项式方阵写成

$$X(\lambda) = \sum_{i=0}^k \lambda^i X_i$$

且对非零多项式方阵，我们假设最高次项（称为首项）系数 $X_k \neq O$ ，并称其次数 $\deg X(\lambda) = k$ 。我们证明，若 X_k 可逆，则对任何 \mathbb{K} 上的 n 阶多项式方阵多项式方阵 $Y(\lambda)$ ，存在唯一 \mathbb{K} 上的 n 阶多项式方阵 $Q_L(\lambda)、R_L(\lambda)、Q_R(\lambda)、R_R(\lambda)$ 使得

$$Y(\lambda) = Q_L(\lambda)X(\lambda) + R_L(\lambda) = X(\lambda)Q_R(\lambda) + R_R(\lambda)$$

且满足 $\deg R_L(\lambda) < \deg X(\lambda)$ 、 $\deg R_R(\lambda) < \deg X(\lambda)$ 。

证明：

我们只证明 Q_R, R_R 的存在唯一性， Q_L, R_L 同理证明即可（注意对比和本讲义 14.1.1 完全类似的证明过程）。

先证明存在性。取 $Q_R(\lambda)$ 为使得 $Y(\lambda) - X(\lambda)Q_R(\lambda)$ 次数最小的多项式矩阵（的其中一个），由于 $Q_R(\lambda) = O$ 时次数与 $Y(\lambda)$ 相同，次数可能性有限，这样的 $Q_R(\lambda)$ 一定存在。下记 $R_R(\lambda) = Y(\lambda) - X(\lambda)Q_R(\lambda)$ ，我们证明 $\deg R_R(\lambda) < \deg X(\lambda)$ 即可。

若否，设 $R_R(\lambda)$ 首项为 $\lambda^m R_m$ ，根据条件 $X(\lambda)$ 首项为 $\lambda^k X_k$ ，由假设 $m \geq k$ 且 R_m, X_k 非零， X_k 可逆。取

$$Q_{R0}(\lambda) = Q_R(\lambda) + \lambda^{m-k} R_m X_k^{-1}$$

则直接计算可知

$$R_{R0}(\lambda) = Y(\lambda) - X(\lambda)Q_{R0}(\lambda) = R_R(\lambda) - \lambda^{m-k} R_m X_k^{-1} X(\lambda)$$

进一步计算发现 $\lambda^{m-k} R_m X_k^{-1} X(\lambda)$ 也为 m 次多项式矩阵且 m 次项系数与 $R_m(\lambda)$ 相同， $R_{R0}(\lambda)$ 为至多 $m-1$ 次多项式矩阵，次数小于 $R_R(\lambda)$ ，与 $Q_R(\lambda)$ 使得 $Y(\lambda) - X(\lambda)Q_R(\lambda)$ 次数最小矛盾。

再证明唯一性。若有 $Y(\lambda) = X(\lambda)Q_{R1}(\lambda) + R_{R1}(\lambda) = X(\lambda)Q_{R2}(\lambda) + R_{R2}(\lambda)$ 均符合要求，有

$$X(\lambda)(Q_{R1}(\lambda) - Q_{R2}(\lambda)) = R_{R2}(\lambda) - R_{R1}(\lambda)$$

若左侧非零，下面证明其次数至少为 $\deg X(\lambda)$ ，这就与右侧次数低于 $\deg X(\lambda)$ 矛盾。

若左侧非零，首先有 $Q_{R1}(\lambda) - Q_{R2}(\lambda)$ 非零，设其首项为 $\lambda^t Q_t$ ，则 $X(\lambda)(Q_{R1}(\lambda) - Q_{R2}(\lambda))$ 的首项为 $\lambda^{k+t} X_k Q_t$ 。由于 X_k 可逆，从 Q_t 非零可得 $X_k Q_t$ 非零，从而其次数为 $k+t \geq k$ 。

注意上述过程的矩阵乘法、求逆等操作都是在 \mathbb{K} 上矩阵中进行的，因此所有系数仍然是 \mathbb{K} 上矩阵，从而得到的仍然是 \mathbb{K} 上的多项式方阵。

* 本质上来说，这个证明和本讲义 14.1.1 证明多项式带余除法时有三点重要区别：

- 条件 X_k 可逆的要求是为了让 $X(\lambda)$ 一定可以左乘矩阵 $R_m X_k^{-1}$ 使得首项系数等于给定矩阵 R_m 。若仅要求 X_k 非零，这样的矩阵未必存在，也就未必可以做带余除法，最简单的例子是

$$Y(\lambda) = I_2, \quad X(\lambda) = \text{diag}(0, 1)$$

若能做带余除法，则由余项次数小于 0 可知必须 $R_R(\lambda) = O$ ，于是有 $X(\lambda)Q_R(\lambda) = Y(\lambda)$ ，但左侧秩至多为 1，右侧秩为 2，矛盾。

- 由于矩阵乘法的不可交换性，考虑左乘和右乘时的商和余项未必相同。根据计算过程事实上可以发现，当 $\deg Y(\lambda) \geq \deg X(\lambda)$ 时，上述得到的 $Q_L(\lambda)$ 和 $Q_R(\lambda)$ 首项系数分别为 X_k^{-1} 左/右乘 $Y(\lambda)$ 的首项系数，只要两者不可交换即不同。

- 由于两个非零矩阵乘积可能为 \mathbf{O} , 对一般的多项式方阵可能出现

$$\deg X(\lambda)Y(\lambda) < \deg X(\lambda) + \deg Y(\lambda)$$

但是, 从证明最后一部分能看出, 只要 $X(\lambda)$ 首项系数 X_k 可逆, 仍然能得到乘积次数为次数之和。

根据这三点我们可以得到, 首项系数可逆的多项式方阵几乎可以按照通常多项式进行处理, 否则性质会有很大差别。这个结论并不是偶然的, 因为多项式方阵可以看作系数在 $\mathbb{K}^{n \times n}$ 中取的多项式。这个矩阵集合可以进行加法、减法、乘法, 但不是一定可以对非零元素进行除法, 是一个环。当环上的元素可逆——用之前的语言也就是说它是环中的单位时, 会有很多与域类似的良好性质。

* 由于 $\lambda I - A$ 、 $\lambda I - B$ 的最高次项均为 I , 可逆, 由此上述的分解存在, 而它们的次数都为 1, 因此余项 R_L 、 R_R 的次数至多为 0, 这就证明了它为常矩阵。

* 至于为什么能想到利用带余除法进行定理证明, 可能是消去过程中带余除法可以自然起到抵消高次的作用, 这将在下方的证明里体现。

接下来, 我们从

$$U(\lambda) = (\lambda I - A)Q(\lambda) + U_0, \quad V(\lambda) = R(\lambda)(\lambda I - A) + V_0$$

出发进行构造。我们将证明 U_0 可逆, 且

$$V_0 = U_0^{-1}, \quad A = V_0 B U_0$$

也就是说, 特征方阵模相抵的过渡矩阵进行带余除法就可以直接得到相似的过渡矩阵。

证明:

在原式左右两侧同时左乘 $U(\lambda)^{-1}$, 并代入 V 的表达式、移项得到

$$(U(\lambda)^{-1} - (\lambda I - B)R(\lambda))(\lambda I - A) = (\lambda I - B)V_0$$

由于 $\lambda I - A$ 、 $\lambda I - B$ 首项系数 I 均可逆, 考虑两侧次数可知

$$\deg(U(\lambda)^{-1} - (\lambda I - B)R(\lambda)) + 1 = 1 + \deg(V_0)$$

由于 V_0 次数至多为 0, 这就得到了 $U(\lambda)^{-1} - (\lambda I - B)R(\lambda)$ 次数至多为 0, 为常数矩阵, 设为 T 。

代换 T 后展开, 得到 $\lambda T - TA = \lambda V_0 - BV_0$, 对比一次项、零次项系数可知

$$T = V_0, \quad TA = BV_0$$

只要再说明 T 可逆, 上式即可写成 $A = T^{-1}BT$, 得到证明。

在 T 的定义式两侧同时左乘 $U(\lambda)$, 可得

$$I - U(\lambda)(\lambda I - B)R(\lambda) = U(\lambda)T$$

由于我们已有了 $V_0 = U_0^{-1}$ 的期望, 我们希望 $U_0 T = I$, 由此展开右侧的 $U(\lambda)$ 并移项得到

$$I - U_0 T = U(\lambda)(\lambda I - B)R(\lambda) + (\lambda I - A)Q(\lambda)T$$

这里左侧出现了一个 $\lambda I - A$, 我们希望能将它提出, 这样就可以进行次数的估算了, 而注意我们尚未用过 $V(\lambda)$ 可逆, 且原式可以写为 $U(\lambda)(\lambda I - B) = (\lambda I - A)V(\lambda)^{-1}$, 于是

$$I - U_0 T = (\lambda I - A)V(\lambda)^{-1}R(\lambda) + (\lambda I - A)Q(\lambda)T$$

$$I - U_0 T = (\lambda I - A)(V(\lambda)^{-1}R(\lambda) + T)$$

仍然因为 $\lambda I - A$ 首项次数可逆, 右侧若非零则至少一次, 与左侧至多零次, 这就证明 $U_0 T = I$, 即 $T = U_0^{-1}$ 。

又由于 $T = V_0$ 、 $TA = BV_0$, 我们已经说明了 A 、 B 相似, 且恰好 $A = U_0^{-1}BU_0 = V_0BU_0$ 。

可以发现，这个证明的核心是利用特征方阵的首项系数可逆，从而相关的带余除法与次数估算都与多项式有类似的性质。本质上，它是作为系数为矩阵的多项式进行操作的，是一个相对高观点的办法——因为带余除法是不太可能凭空想到的。另一个证明方法就相对低观点一些，主要通过直接的对比系数计算得到。

16.1.3 计算系数出发的证明

回到最开始的模相抵式

$$\lambda I - A = U(\lambda)(\lambda I - B)V(\lambda)$$

如果想要对比系数，直接将右侧的三项乘法完全展开是不现实的，因此我们还是需要移项为

$$U(\lambda)^{-1}(\lambda I - A) = (\lambda I - B)V(\lambda)$$

取一个充分大的 m 使得 $U(\lambda)$ 、 $V(\lambda)$ 、 $U(\lambda)^{-1}$ 的次数都不超过 m ，并将它们写为

$$U(\lambda) = \sum_{i=0}^m \lambda^i U_i, \quad V(\lambda) = \sum_{i=0}^m \lambda^i V_i, \quad U(\lambda)^{-1} = \sum_{i=0}^m \lambda^i W_i$$

为了方便书写，在下标 i 小于 0 或大于 m 时， U_i 、 V_i 、 W_i 规定为 O 。

展开左侧的乘法得到

$$U(\lambda)^{-1}(\lambda I - A) = \sum_{i=0}^m \lambda^i W_i (\lambda I - A) = \sum_{i=0}^m (\lambda^{i+1} W_i + \lambda^i W_i A)$$

第二个等号应用了我们之前所说的， λ^i 可以形式上看作数乘。由于我们希望对比次数，按照次数归类可以改写为（注意之前的规定保证了 $W_{-1} = W_{m+1} = O$ ，于是这样的书写是合理的）

$$\sum_{i=0}^{m+1} (\lambda^i W_{i-1} + \lambda^i W_i A) = \sum_{i=0}^{m+1} \lambda^i (W_{i-1} + W_i A)$$

完全同理展开右侧（注意左右乘的区别），可以得到等式

$$\sum_{i=0}^{m+1} \lambda^i (W_{i-1} + W_i A) = \sum_{i=0}^{m+1} \lambda^i (V_{i-1} + B V_i)$$

由于两个多项式矩阵相等则它们的各项系数必然完全相等，可以得到 $m+2$ 个方程

$$W_{i-1} + W_i A = V_{i-1} + B V_i, \quad i = 0, \dots, m+1$$

注意到 V_0 到 V_m 只有 $m+1$ 个未知量，把它们都用 W_i 表示后，理应可以得到一个关于所有 W_i 的方程，这是我们想要的、能确定 $U(\lambda)^{-1}$ 的性质的中间结果：

$$\sum_{i=0}^m B^i (W_i A - B W_i) = O$$

证明：

代入 $i = m+1$ 时，有 $V_m = W_m$ （注意 $W_{m+1} = V_{m+1} = O$ ），而代入 $i = m$ 时，可以得到

$$W_{m-1} + W_m A = V_{m-1} + B V_m = V_{m-1} + B W_m$$

从而

$$V_{m-1} = W_{m-1} - W_m A + B W_m$$

同理可以得到 V_i 的递推

$$V_{i-1} = W_{i-1} - W_i A + B V_i$$

算几项后观察即可以归纳得到

$$V_i - W_i + BW_{i+1} - W_{i+1}A + B(BW_{i+2} - W_{i+2}A) + \cdots + B^{m-i-1}(BW_m - W_mA)$$

注意到，上述过程我们最多只用到了 $i=1$ 的方程 $W_0 + W_1A = V_0 + BV_1$ ，而 $i=0$ 的方程 $W_0A = BV_0$ （注意 $W_{-1} = V_{-1} = O$ ）可以提供最后的等式。将 $i=0$ 的 V_i 代入上方等式可知

$$V_0 = W_0 + BW_1 - W_1A + B(BW_2 - W_2A) + \cdots + B^{m-1}(BW_m - W_mA)$$

我们用更形式化的方式写出这个表达式

$$V_0 = W_0 + \sum_{i=1}^m B^{i-1}(BW_i - W_iA)$$

这样， $i=0$ 的方程 $W_0A - BV_0 = O$ 就可以写为

$$W_0A - BW_0 - B \sum_{i=1}^m B^{i-1}(BW_i - W_iA) = O$$

也即

$$\sum_{i=0}^m B^i(W_iA - BW_i) = O$$

虽然上述过程看起来有点复杂，但本质只是中学数学的代数式变形而已。为了接近相似的目标，我们可以发现，相似等价于寻找可逆的 X 使得 $XA = BX$ 。从这个中间结果里， X 的形式是可以变换出的，用分配律展开合并得到：

$$O = \sum_{i=0}^m (B^i W_i A - BB^i W_i) = \left(\sum_{i=0}^m B^i W_i \right) A - B \sum_{i=0}^m B^i W_i$$

由此，可以考虑取

$$X = \sum_{i=0}^m B^i W_i$$

下面，我们将来到这个证明最难的部分，找到 X 的逆。我们先给出逆的形式并证明，最后再来谈谈这个形式是如何想到的。事实上， X 的逆是

$$Y = \sum_{i=0}^m A^i U_i$$

证明：

只要证明 $XY = I$ 即可。由于 $XA = BX$ ，利用结合律有 $XA^2 = (XA)A = BXA = B(BX) = B^2X$ ，由此对任何 i 都有 $XA^i = B^i X$ ，从而利用乘法分配律可以计算出

$$XY = X \sum_{i=0}^m A^i U_i = \sum_{i=0}^m B^i X U_i = \sum_{i=0}^m B^i \sum_{j=0}^m B^j W_j U_i = \sum_{i=0}^m \sum_{j=0}^m B^{i+j} W_j U_i$$

我们将它按照 B 的次数重新整理为

$$XY = \sum_{k=0}^{2m} B^k \sum_{\substack{i+j=k \\ 0 \leq i \leq m \\ 0 \leq j \leq m}} W_j U_i$$

右侧究竟是不是我们想要的 I 呢？注意 $U(\lambda)^{-1}$ 与 $U(\lambda)$ 的互逆关系，有（由加法交换与结合律，求和次序可以交换）

$$I = U(\lambda)^{-1}U(\lambda) = \sum_{j=0}^m \lambda^j W_j \sum_{i=0}^m \lambda^i U_i = \sum_{j=0}^m \sum_{i=0}^m \lambda^{i+j} W_j U_i = \sum_{i=0}^m \sum_{j=0}^m \lambda^{i+j} W_j U_i$$

同样按照 λ^{i+j} 的次数整理为

$$I = \sum_{k=0}^{2m} \lambda^k \sum_{\substack{i+j=k \\ 0 \leq i \leq m \\ 0 \leq j \leq m}} W_j U_i$$

对比两侧系数，可以发现在 $k > 0$ 时， λ^k 后的系数应为 O 。而 $i + j = 0$ ，即 $i = j = 0$ 时，有 $W_0 U_0 = I$ 。因此，之前的求和可以化为（由任何矩阵零次方为 I ）

$$XY = B^0 I + \sum_{k=1}^{2m} B^k O = I$$

这就得到了证明。

* 关于如何想到逆的构造，一位热心同学提供了不错的思路：由于 $XA = BX$ ，其逆 $Y = X^{-1}$ 应满足 $AY = YB$ 。将原式重新改写为

$$(\lambda I - A)V(\lambda)^{-1} = U(\lambda)(\lambda I - B)$$

并设 $V(\lambda)^{-1}$ 的各项系数为 Z_i ，完全类似对比系数可以得到 Z_i 与 U_i 的关系，消去 Z_i 恰好可以发现

$$Y = \sum_{i=0}^m A^i U_i$$

满足 $AY = YB$ ，再结合 U_i 、 W_i 对应多项式方阵的互逆性，即可尝试证明这是逆的形式。

虽然这个证明的计算相对复杂，但思路比前一种证明更加自然。从这个证明中，我们更是可以得到一个重要的对比系数得到的方程组

$$V_{i-1} = W_{i-1} - W_i A + B V_i, \quad i = 0, \dots, m+1$$

注意到，寻找 $XA = BX$ 可以看作寻找 $SA = BT$ 且 $S = T$ 的 S 与 T 。当 $i = m+1$ 时，方程即

$$W_m = V_m$$

这时 W_i 、 V_i 符合第二个条件，但未必符合第一个。当 $i = 0$ 时，方程即

$$W_0 A = B V_0$$

这时 W_i 、 V_i 符合第一个条件，但未必符合第二个。

通过中间不同的次数的方程组递推，我们事实上在均衡两个条件的误差，并最终通过将每个 W_i 乘矩阵进行组合达到同时完成两个条件的效果。这样的递推、均衡思想在数学证明中是重要的。至于证明的主要手段，仍然是看作系数为多项式的矩阵进行同次数系数的对比。

§16.2 Jordan 标准形

16.2.1 初等因子组

上节的定理证明后，对相似标准形的讨论马上就可以化为对模相抵的讨论，具体来说，对 \mathbb{K} 上的矩阵 A ，相似标准形的构造方案变成了：

- 构造特征方阵 $\lambda I - A$;
- 计算 $\lambda I - A$ 的 Smith 标准形;
- 找固定简单形式的 \mathbb{K} 上方阵 B 使得 $\lambda I - B$ 的 Smith 标准形与 $\lambda I - A$ 相同;
- 将 B 定义为 A 的相似标准形。

虽然其他问题已经解决，这里的第三步仍然并不简单。本节，我们主要讨论相似的数域为复数域，即 $\mathbb{K} = \mathbb{C}$ ，并研究如何得到这样的矩阵 B 。不过，在进入 \mathbb{C} 的讨论前，我们还是需要看一看对一般的 \mathbb{K} 的结论。

我们已经知道，确定 Smith 标准形也就是确定不变因子组或行列式因子组。但无论是利用哪个，都具有一个十分麻烦的问题：分块对角阵的不变因子/行列式因子与每个对角块的不变因子/行列式因子关系并不明确。

我们考虑最简单的 $C(\lambda) = \text{diag}(A(\lambda), B(\lambda))$ 形式的 \mathbb{K} 上多项式矩阵。假设 $A(\lambda)$ 的不变因子为 $a_i(\lambda)$ 、 $B(\lambda)$ 的不变因子为 $b_i(\lambda)$ 、 $C(\lambda)$ 的不变因子为 $c_i(\lambda)$ 。

由于第一个不变因子为所有元素的最大公因式，有 $c_1(\lambda) = \gcd(a_1(\lambda), b_1(\lambda))$ 。但是，对于 $c_2(\lambda)$ 就没有如此简单的关系了。事实上，它满足的表达式为

$$c_1(\lambda)c_2(\lambda) = \gcd(a_1(\lambda)a_2(\lambda), b_1(\lambda)b_2(\lambda), a_1(\lambda)b_1(\lambda))$$

证明：

考虑 $C(\lambda)$ 的第二个行列式因子。对于其所有二阶子式，只包含在 $A(\lambda)$ 部分中时最大公因式应为 (利用不变因子与行列式因子关系) $a_1(\lambda)a_2(\lambda)$ 、只包含在 $B(\lambda)$ 部分中时最大公因式应为 $b_1(\lambda)b_2(\lambda)$ 。而在部分包含于 $A(\lambda)$ 中、部分包含于 $B(\lambda)$ 中时，由分块对角性其一定为 $A(\lambda)$ 中某元素与 $B(\lambda)$ 中某元素乘积。下面证明这种情况的最大公因式为 $a_1(\lambda)b_1(\lambda)$ 。若 $a_1(\lambda)b_1(\lambda) = 0$ ，由定义可知成立，以下假设两者均非零。

首先，由于 $A(\lambda)$ 中元素都是 $a_1(\lambda)$ 倍式、 $B(\lambda)$ 中元素都是 $b_1(\lambda)$ 倍式，最大公因式一定是 $a_1(\lambda)b_1(\lambda)$ 的倍式。

若最大公因式事实上为 $d(\lambda)a_1(\lambda)b_1(\lambda)$ ，且 $\deg d(\lambda) \geq 1$ ，将 $d(\lambda)$ 进行唯一因子分解 (见本讲义 14.1.4)，设某首一不可约多项式 $p(\lambda) | d(\lambda)$ 。

利用唯一因子分解定理，若 $p(\lambda) | x(\lambda)y(\lambda)$ ，说明 $x(\lambda)y(\lambda)$ 的分解式里 $p(\lambda)$ 次数至少为 1，从而 $x(\lambda)$ 与 $y(\lambda)$ 的分解式里 $p(\lambda)$ 的次数不能全为 0，也即 $p(\lambda) | x(\lambda)$ 或 $y(\lambda) | b(\lambda)$ 。由于任何 $A(\lambda)$ 中元素 $a(\lambda)$ 与 $B(\lambda)$ 中元素 $b(\lambda)$ 的乘积是 $p(\lambda)a_1(\lambda)b_1(\lambda)$ 的倍式，且 $a_1(\lambda) | a(\lambda)$ 、 $b_1(\lambda) | b(\lambda)$ ， $a(\lambda)/a_1(\lambda)$ 与 $b(\lambda)/b_1(\lambda)$ 中至少有一个是 $p(\lambda)$ 的倍式。

然而，若所有 $a(\lambda)/a_1(\lambda)$ 都是 $p(\lambda)$ 的倍式，可推出所有 $a(\lambda)$ 是 $a_1(\lambda)p(\lambda)$ 的倍数，与所有元素最大公因式 $a_1(\lambda)$ 矛盾；若有一个 $a(\lambda)/a_1(\lambda)$ 不是 $p(\lambda)$ 的倍式，取此 $a(\lambda)$ 与所有 $b(\lambda)$ 可得所有 $b(\lambda)/b_1(\lambda)$ 都是 $p(\lambda)$ 的倍式，类似上方仍然矛盾。

综合以上，即得到部分包含于 $A(\lambda)$ 中、部分包含于 $B(\lambda)$ 中的二阶子式最大公因式 $a_1(\lambda)b_1(\lambda)$ ，因此最终第二个行列式因子为

$$\gcd(a_1(\lambda)a_2(\lambda), b_1(\lambda)b_2(\lambda), a_1(\lambda)b_1(\lambda))$$

同理，我们最终可以证明 (省略 λ)

$$\prod_{i=1}^k c_i = \gcd \left(\prod_{i=1}^k a_i, \prod_{i=1}^{k-1} a_i b_1, \prod_{i=1}^{k-2} a_i b_1 b_2, \prod_{i=1}^{k-3} a_i \prod_{j=1}^3 b_j, \dots, \prod_{j=1}^k b_j \right)$$

对任何正整数 k 成立 (超出阶数的不变因子都记为 0)。若用 $C_i(\lambda)$ 、 $A_i(\lambda)$ 、 $B_i(\lambda)$ 表示 A 、 B 、 C 的第 i 个行列式因子，上式还可以写为

$$C_k(\lambda) = \gcd(A_k(\lambda), A_{k-1}(\lambda)B_1(\lambda), A_{k-2}(\lambda)B_2(\lambda) \dots, A_1(\lambda)B_{k-1}(\lambda), B_k(\lambda))$$

虽然这样看起来相对简单了一些，但还是要计算 k 个多项式的最大公因式，我们希望能有更直接的结论以方便进行构造。

在刚才的证明中，我们运用了唯一因子分解定理这个重要的结论。虽然它在不同数域中结果不同（见本讲义 14.1.4），破坏了不变性，但或许可以用来刻画一些性质。

基于这个观察，定义一个 \mathbb{K} 上的满秩多项式方阵 $A(\lambda)$ 的初等因子组为：设 $d_s(\lambda), \dots, d_n(\lambda)$ 为其所有不等于 1 的不变因子，且分别有分解

$$d_i(\lambda) = p_{i1}^{a_{i1}}(\lambda) \cdots p_{ik_i}^{a_{ik_i}}(\lambda), \quad i = s, s+1, \dots, n$$

其中 p_{i1} 到 p_{ik_i} 是不同的首一不可约多项式（注意由于上述 $d_i(\lambda)$ 都是至少一次的首一多项式，唯一因子分解定理中的常数项为 1），则将所有 $p_{ij}^{a_{ij}}(\lambda)$ 称为 $A(\lambda)$ 的初等因子组。

例如，考虑 \mathbb{C} 上的多项式方阵 $\text{diag}(1, 1, \lambda - 1, (\lambda - 1)\lambda, (\lambda - 1)^2\lambda^2(\lambda - 2))$ ，由于其已经符合 Smith 标准形的要求，所有对角元即为不变因子。其中，至少一次的不变因子有 $(\lambda - 1)^2$ 、 $(\lambda - 1)^2\lambda$ 与 $(\lambda - 1)^3\lambda^2(\lambda - 2)$ 。它们已经写成了 \mathbb{C} 上的唯一分解形式，因此其初等因子组为（注意重复的也要算上）

$$\lambda - 1, \quad \lambda - 1, \quad \lambda, \quad (\lambda - 1)^2, \quad \lambda^2, \quad \lambda - 2$$

* 初等因子组事实上是一个多重集合，也即每个元素可以出现不止一次的集合。在上方的例子中， $\lambda - 1$ 出现了两次，其他都只出现了一次。下面我们将提到把两个初等因子组放在一起，是指新的多重集合中每个元素的出现次数为原本的两个集合中对应元素的出现次数相加（当然，不出现视为出现 0 次）。

* 定义要求满秩多项式方阵的基本原因是，0 的分解是无法刻画的，因此希望所有有效的不变因子均非零。

根据定义，不变因子组可以完全确定初等因子组，而满秩多项式方阵，只要两个满秩多项式方阵的初等因子组相等，就有它们的不变因子组相等，从而模相抵。

证明：

为方便书写将多项式 $f(\lambda)$ 记为 f 。

假设多项式方阵 $A(\lambda)$ 的秩为 r ，且初等因子组为

$$\begin{aligned} & p_1^{k_{11}}, p_1^{k_{12}}, \dots, p_1^{k_{1m_1}}, \\ & p_2^{k_{21}}, p_2^{k_{22}}, \dots, p_2^{k_{2m_2}}, \\ & \dots, \\ & p_t^{k_{t1}}, p_t^{k_{t2}}, \dots, p_t^{k_{tm_t}} \end{aligned}$$

这里 p_1, \dots, p_t 为不同的不可约多项式，且 $l_1 < l_2$ 时 $k_{il_1} \geq k_{il_2}$ （也即次数从大到小排列）。

由于已知秩为 r ，非零的不变因子为 d_1, \dots, d_r 。我们下面证明

$$d_r = p_1^{k_{11}} \cdots p_t^{k_{t1}}$$

由于每个初等因子必然为某个不变因子分解出的，我们先证明 $p_1^{k_{11}}$ 到 $p_t^{k_{t1}}$ 都是 d_r 分解出的。若否，不妨假设 $p_1^{k_{11}}$ 是 $d_{r_0}, r_0 < r$ 分解出的，则根据唯一分解定理， d_r 中 p_1 的次数至多为 $p_1^{k_{12}}$ （根据定义，一个不变因子只能分解出一个以 p_1 为底数的初等因子）。由定义 $k_{12} \leq k_{11}$ ，但根据 $d_{r_0} | d_r$ 可知 $k_{12} \geq k_{11}$ ，综合得到 $k_{12} = k_{11}$ ，因此 $p_1^{k_{12}}$ 就是 $p_1^{k_{11}}$ ，从而得证。

考虑去掉 $p_1^{k_{11}}, \dots, p_t^{k_{t1}}$ 的初等因子组，它们必然是由 d_1, \dots, d_{r-1} 分解出的，重复上述过程，最终得到

$$d_{r-j+1} = p_1^{k_{1j}} \cdots p_t^{k_{tj}}, \quad j = 1, \dots, r$$

这里当 $j > m_i$ 时记 $k_{ij} = 0$ 。

至此， $A(\lambda)$ 的不变因子组完全确定，定理得证。

* 还有如下重要结论：所有初等因子乘积为行列式（对应的首一多项式）。这是因为满秩多项式方阵不变因子中没有 0，于是所有初等因子乘积就等于所有不变因子乘积，而这等于最后一个行列式因子，即其行列式对应的首一多项式。

另外，分块对角阵的初等因子组确实是容易确定的，我们可以证明， $C(\lambda) = \text{diag}(A(\lambda), B(\lambda))$ 的初等因子组就是 $A(\lambda)$ 的初等因子组与 $B(\lambda)$ 的初等因子组放在一起（假设 $A(\lambda)$ 、 $B(\lambda)$ 都是满秩多项式方阵）。

证明：

首先，利用秩的性质可知 $C(\lambda)$ 也满秩，从而初等因子组可以定义。

由于初等因子组对每个不可约多项式是分开的，我们只需要对每个 $A(\lambda)$ 或 $B(\lambda)$ 的初等因子组底数中出现的不可约多项式考虑即可。将不可约多项式 $p(\lambda)$ 记作 p ，假设它在 $A(\lambda)$ 的不变因子 a_1, \dots, a_r 中次数分别为 n_1, \dots, n_r 、在 $B(\lambda)$ 的不变因子 b_1, \dots, b_s 中次数分别为 m_1, \dots, m_s ，我们下面说明 $C(\lambda)$ 中以 p 为底数的初等因子是 $p^{n_1}, \dots, p^{n_r}, \dots, p^{m_1}$ 所有次数大于 0 的项。

根据不变因子的相互整除性， $n_1 \leq n_2 \leq \dots \leq n_r$, $m_1 \leq m_2 \leq \dots \leq m_s$ 。我们之前证明了（省略 λ ）对 $k = 1, \dots, r+s$ 有

$$\prod_{i=1}^k c_i = \gcd \left(\prod_{i=1}^k a_i, \prod_{i=1}^{k-1} a_i b_1, \prod_{i=1}^{k-2} a_i b_1 b_2, \prod_{i=1}^{k-3} a_i \prod_{j=1}^3 b_j, \dots, \prod_{j=1}^k b_j \right)$$

由此，假设 $c_i(\lambda)$ 中包含的 p 次数为 l_i ，应有（假设下标大于 r 时 n_i 为 $+\infty$ ，下标大于 s 时 m_j 为 $+\infty$ ，也即求最小值时忽略这些项）

$$\sum_{i=1}^k l_i = \min \left(\sum_{i=1}^k n_i, \sum_{i=1}^{k-1} n_i + m_1, \sum_{i=1}^{k-2} n_i + m_1 + m_2, \dots, \sum_{i=1}^k m_i \right)$$

利用归纳法即可证明 l_i 是所有 $n_1, \dots, n_r, m_1, \dots, m_s$ 中第 i 小的数（这看起来更像是计算机课程里会出现的问题，可参考归并排序的算法），大致思路为，由于 m_i 和 n_i 都是单调不减的，其中最小 k 个的和一定会出现在 \min 中，因此前 k 个 l_i 的和就是 $n_1, \dots, n_r, m_1, \dots, m_s$ 中最小 k 个的和，再简单归纳得到。

于是，由于 $c_i(\lambda)$ 中的底数为 p 的初等因子为 p^{l_i} （若 $l_i > 0$ ）， $c_1(\lambda), \dots, c_{r+s}(\lambda)$ 中能分解出的底数为 p 的初等因子一定为 $n_1, \dots, n_r, m_1, \dots, m_s$ 中所有非零次数对应的 p^{n_i} 或 p^{m_i} ，对 $a_i(\lambda)$ 、 $b_i(\lambda)$ 因式分解中出现的所有 p 考虑即得到证明（由于 $\det C(\lambda) = \det A(\lambda) \det B(\lambda)$ ， $c_i(\lambda)$ 因式分解中不会出现 $a_i(\lambda)$ 、 $b_i(\lambda)$ 中从未出现过的 p ）。

* 由此归纳即得到 $\text{diag}(A_1(\lambda), \dots, A_k(\lambda))$ 的初等因子组就是 $A_1(\lambda), \dots, A_k(\lambda)$ 的初等因子组放在一起。特别地，当 $A_1(\lambda), \dots, A_k(\lambda)$ 实际上是非零多项式 $a_1(\lambda), \dots, a_k(\lambda)$ 时，其看作一阶方阵唯一一个不变因子是自身（所化作的首一多项式），因此可得满秩对角多项式方阵的初等因子组可直接将每个对角元因式分解得到，这就规避了不变因子的计算。

有了以上的准备工作，我们就可以开始构建复方阵的相似标准形（称为 **Jordan 标准形**）了。下一部分的证明中，我们将看到上面两个定理是初等因子组发挥作用的核心。在本部分最后，我们需要指出，相同的多项式方阵在看作数域不同时初等因子组可能不同：将 $\lambda^2 + 1$ 看作一阶多项式方阵，若其为 \mathbb{C} 上的多项式方阵，则初等因子组为 $\lambda - i, \lambda + i$ ；若其为 \mathbb{R} 或 \mathbb{Q} 上的多项式方阵，则初等因子组为 $\lambda^2 + 1$ 。

16.2.2 Jordan 块

现在，让我们将目光回到复方阵 $A \in \mathbb{C}^{n \times n}$ 的特征方阵 $\lambda I - A$ 上。很容易发现， $\lambda I - A$ 一定是满秩的：由完全展开，其行列式 $\det(\lambda I - A)$ 是一个 λ 的首一 n 次多项式，从而非零。由此，根据上一部分已经证明的，对于 n 阶复方阵， $\lambda I - A$ 的初等因子组可以完全确定 $\lambda I - A$ 的不变因子组，进而确定 A 与怎样的方阵相似。

此外，由于所有初等因子乘积为行列式，复数域上的不可约多项式只有一次多项式， $\lambda I - A$ 的初等因子组一定可以写为

$$(\lambda - \lambda_1)^{k_1}, (\lambda - \lambda_2)^{k_2}, \dots, (\lambda - \lambda_s)^{k_s}$$

这里不同的 i 对应的 λ_i 与 k_i 均可能相同， $k_1 + \dots + k_s = n$ 。

现在，终于到了构造标准形式的时候。一个简单的想法是：如果我们能构造矩阵 B_i 使得 $\lambda I - B_i$ 的初等因子组只有 $(\lambda - \lambda_i)^{k_i}$ ，由于

$$\lambda I - \text{diag}(B_1, \dots, B_s) = \text{diag}(\lambda I - B_1, \dots, \lambda I - B_s)$$

根据上一部分已经证明的，分块对角阵的初等因子组是将每个对角块的初等因子组放在一起，此矩阵的初等因子组恰好为

$$(\lambda - \lambda_1)^{k_1}, (\lambda - \lambda_2)^{k_2}, \dots, (\lambda - \lambda_s)^{k_s}$$

因此与 A 相似。

* 由于两个复方阵对应的特征方阵初等因子组相同，它们的初等因子组次数之和也相同，而这就是方阵的阶数，因此特征方阵的初等因子组相同可以说明同阶，从而两个复方阵相似当且仅当对应的特征方阵初等因子组相同，此结论无需再叙述阶数相等。

那么，怎样的 B_i 满足要求呢？我们有如下的定理：设 k 阶复方阵 B 的各分量 b_{ij} 满足

$$\begin{cases} b_{ij} = 0 & i > j \\ b_{ij} = \lambda_0 & i = j \\ b_{ij} \neq 0 & i = j - 1 \end{cases}$$

或

$$\begin{cases} b_{ij} = 0 & i < j \\ b_{ij} = \lambda_0 & i = j \\ b_{ij} \neq 0 & i = j + 1 \end{cases}$$

则 $\lambda I - B$ 的初等因子组只有 $(\lambda - \lambda_0)^k$ 。

证明：

只对第一种情况，即 B 为上三角阵的情况证明。这时 $\lambda I - B$ 为

$$\begin{pmatrix} \lambda - \lambda_0 & -b_{12} & * & \dots & * \\ & \lambda - \lambda_0 & -b_{23} & \ddots & \vdots \\ & & \lambda - \lambda_0 & \ddots & * \\ & & & \ddots & -b_{n-1,n} \\ & & & & \lambda - \lambda_0 \end{pmatrix}$$

由于 $b_{12}, \dots, b_{n-1,n}$ 均非零，其右上角的 $n-1$ 阶行列式为某非零复数，因此第 k 个行列式因子为 1，且计算行列式得第 k 个行列式因子为 $(\lambda - \lambda_0)^k$ 。利用不变因子和行列式因子的关系， $\lambda I - B$ 前 $k-1$ 个不变因子只能都为 1，第 k 个为 $(\lambda - \lambda_0)^k$ ，再根据初等因子组定义得结论。

在如此多的选择里，我们希望为 0 的元素尽量多（这也是为了方便之后的多项式等计算），非零元素也

尽量简单（非零的最简单元素自然是 1），因此，一般有两种选择：

$$\begin{pmatrix} \lambda_0 & 1 & & \\ & \lambda_0 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_0 \end{pmatrix}, \quad \begin{pmatrix} \lambda_0 & & & \\ 1 & \lambda_0 & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_0 \end{pmatrix}$$

不同教材的选择有所不同，我们的讲义里采取前一种方式定义。将上述前一种方式定义的 k 阶矩阵称为一个 k 阶 **Jordan 块**，记作 $J_k(\lambda_0)$ 。在不得不使用后一种的时候，我们将把它称为下三角 **Jordan 块**，并记为 $J_k^{(L)}(\lambda_0)$ 。

用 Jordan 块代入之前得到的相似结论，我们可以证明：对任何复方阵 A ，存在一系列 Jordan 块 $J_{k_1}(\lambda_1), \dots, J_{k_s}(\lambda_s)$ 使得存在可逆矩阵 P 满足

$$A = P^{-1} \operatorname{diag}(J_{k_1}(\lambda_1), \dots, J_{k_s}(\lambda_s))P$$

其中 $k_1 + \dots + k_s = n$ ，这称为复方阵 A 的 **Jordan 标准形**。

那么，Jordan 标准形是否具有唯一性呢？事实上，它在交换 Jordan 块次序视为不变的意义下唯一。这里仍然是通过不变量对唯一性进行证明，而不变量就是已经给出的初等因子组。

证明：

由于每个 $J_k(\lambda_0)$ 的特征方阵对应唯一一个初等因子，不同的 $J_k(\lambda_0)$ 对应的初等因子不同，若两组 $J_{k_1}(\lambda_1), \dots, J_{k_s}(\lambda_s)$ 交换次序后仍然无法完全相同，则它们作为对角块的矩阵的特征方阵具有不同的初等因子组，从而根据特征方阵相似等价于初等因子组相同可知不相似。

* 注意 $J_{k_1}(\lambda_1)$ 的括号并不表示多项式方阵，只是对其对角元的标注。

16.2.3 矩阵的多项式

为了让大家感受到 Jordan 标准形的威力，我们将利用它彻底解决本讲义 7.4.2 的多项式问题，并用更简洁的证明给出可对角化的重要判定定理。

首先，我们需要计算 Jordan 块的多项式，它有如下性质：设 $f(\lambda) \in \mathbb{C}[\lambda]$ ， $A = f(J_n(x))$ 的各分量为 a_{ij} ，则

$$a_{ij} = \begin{cases} 0 & i > j \\ \frac{1}{(j-i)!} f^{(j-i)}(x) & i \leq j \end{cases}$$

这里上标 $j - i$ 表示进行 $j - i$ 次求导，进行 0 次求导即为 $f(x)$ 自身。

证明：

由于 Jordan 标准形可写为 $xI + N$ ，其中

$$N = \begin{pmatrix} 0 & 1 & & \\ 0 & \ddots & & \\ & \ddots & 1 & \\ & & & 0 \end{pmatrix}$$

计算可发现 N^k 当且仅当 $j - i = k$ 的元素为 1，其余为 0（当 $k \geq n$ 时即全为 0），由此，我们先考虑 $f(x) = x^m$ 的情况，此时利用 I, N 可交换，通过二项式定理展开有

$$(xI + N)^m = \sum_{k=0}^m C_m^k x^{m-k} N^k$$

由此，当 $k \leq \min(m, n-1)$ 时，其 $j - i = k$ 位置的元素为 $C_m^k x^{m-k}$ ，否则为 0。而对 x^m 求 k 阶导数的结果恰好在 $k \leq m$ 时为 $k! C_m^k x^{m-k}$ ，在 $k > m$ 时为 0，从而对 x^m 结论成立。

对一般的

$$f(x) = \sum_{k=0}^m c_k x^k$$

利用高等数学知识，对 $f(x)$ 求 l 次导相当于对每个 x^k 求 l 次导再乘对应 c_k 求和（事实上这意味着求导运算是线性的），而

$$f(J_n(x)) = \sum_{k=0}^m c_k J_n(x)^k$$

考虑每个分量可发现结论仍然成立。

下面，我们就可以利用 Jordan 标准形给出矩阵多项式的算法：由上学期知识，对多项式 f 有

$$f(P^{-1}AP) = P^{-1}f(A)P$$

因此，设 A 的 Jordan 标准形为

$$A = P^{-1} \operatorname{diag}(J_{k_1}(\lambda_1), \dots, J_{k_s}(\lambda_s))P$$

则（准对角阵的多项式是每个对角块分别作多项式，这可以直接计算得到）

$$f(A) = P^{-1} \operatorname{diag}(f(J_{k_1}(\lambda_1)), \dots, f(J_{k_s}(\lambda_s)))P$$

从上述过程可以看到，有了 Jordan 标准形后，矩阵的多项式问题可以通过具体的计算来确定。进一步地，我们来求解矩阵方程 $A^n = I$ ，假设 $n \geq 1$ 。

解答：

设 A 的 Jordan 标准形为

$$A = P^{-1} \operatorname{diag}(J_{k_1}(\lambda_1), \dots, J_{k_s}(\lambda_s))P$$

则

$$I = A^n = P^{-1} \operatorname{diag}(J_{k_1}(\lambda_1)^n, \dots, J_{k_s}(\lambda_s)^n)P$$

消去可得

$$I = \operatorname{diag}(J_{k_1}(\lambda_1)^n, \dots, J_{k_s}(\lambda_s)^n)$$

于是 $A^n = I$ 当且仅当其每个 Jordan 块的 n 次方都是 I 。

若 Jordan 块的阶数为 1， $J_1(\lambda_0)$ 即为 λ_0 ，其 n 次方为 I 当且仅当 $\lambda_0^n = 1$ ；若 Jordan 块的阶数大于 1，可发现 $J_k(\lambda_0)^n = I$ 需要对角元为 1、 $i - j = 1$ 位置的元素为 0，利用 Jordan 块多项式的结论也即

$$\lambda_0^n = 1, \quad n\lambda_0^{n-1} = 0$$

但这两个式子对复数 λ_0 不可能同时成立，矛盾。

综合以上，我们得到， $A^n = I$ 当且仅当 A 的 Jordan 标准形只有一阶 Jordan 块，且每个 Jordan 块的 λ_0 都满足 $\lambda_0^n = 1$ 。

再举一个例子，求解矩阵方程 $A^n = O$ 。**解答：**

设 A 的 Jordan 标准形为

$$A = P^{-1} \operatorname{diag}(J_{k_1}(\lambda_1), \dots, J_{k_s}(\lambda_s))P$$

则

$$O = A^n = P^{-1} \operatorname{diag}(J_{k_1}(\lambda_1)^n, \dots, J_{k_s}(\lambda_s)^n)P$$

消去可得

$$O = \operatorname{diag}(J_{k_1}(\lambda_1)^n, \dots, J_{k_s}(\lambda_s)^n)$$

于是 $A^n = O$ 当且仅当其每个 Jordan 块的 n 次方都是 O 。

利用 Jordan 块多项式的结论可知 Jordan 块 $J_k(\lambda_0)$ 的 n 次方对角元 λ_0^n , 由此其为 O 至少需要 $\lambda_0 = 0$ 。此时直接计算可发现 $k \leq n$ 时符合要求, 否则不符合。

综合以上, 我们得到, $A^n = I$ 当且仅当 A 的 Jordan 标准形中每个 Jordan 块的 $\lambda_0 = 0$, 且阶数不超过 n 。

* 事实上现实情况并没有这么美好, 因为所有涉及 Jordan 标准形的算法只能成为近似算法, 不可能精确: 由于五次及以上方程的不可解性, 我们无法精确得到任何不变因子分解出的初等因子组, 从而无法精确确定 Jordan 标准形。

* 另外, 可交换性即使有了 Jordan 标准形也并不容易确定, 之后有机会时我们将描述几个简单的结论, 并在期中复习题彻底解决可交换性问题。

16.2.4 可对角化 IV

接下来, 我们来看一看之前所学习的特征系统结论在 Jordan 标准形有什么对应。一个很重要的观察是, 相似不改变特征值、上三角阵的特征值是对角元, 因此每个 Jordan 块的 $\lambda_1, \dots, \lambda_k$ 一定是 A 的特征值。

此外, 由于单个对角元可以看作一阶 Jordan 块, 对角阵符合 Jordan 标准形的形式, 因此复方阵相似对角化是 Jordan 标准形的一种特殊情况。反过来, 只要 Jordan 标准形有二阶对角块, 它就不可能是对角阵了。综合这两个结论, 我们可以得到复方阵可以相似对角化当且仅当它所有 Jordan 块为一阶。利用 Jordan 块的定义, 我们可以进一步得到复方阵可以相似对角化当且仅当它的特征方阵的所有初等因子为一次。

* 由此, 上一部分 $A^n = I$ 的解也可以描述为, $A^n = I$ 当且仅当其可以相似对角化, 且所有特征值的 n 次方为 1。

不过, 这些内容比起有意义的推导都更像是纯粹的文字与逻辑游戏。为了让这部分内容不显得那么无聊, 我们来叙述一个更有意义的定理: 复方阵 A 可对角化当且仅当存在无重根多项式 f 使得 $f(A) = O$ 。这个定理在本讲义 9.2.3 已经证明了, 而通过 Jordan 标准形可以大大简化证明过程。

证明:

左推右: 若 A 可对角化, 设其不同特征值为 $\lambda_1, \dots, \lambda_k$, 代数重数分别为 n_1, \dots, n_k , 且

$$A = P^{-1} \operatorname{diag}(\lambda_1 I_{n_1}, \dots, \lambda_k I_{n_k}) P$$

令 $f(\lambda) = (\lambda - \lambda_1) \dots (\lambda - \lambda_k)$ 有

$$f(A) = P^{-1} \operatorname{diag}(f(\lambda_1 I_{n_1}), \dots, f(\lambda_k I_{n_k})) P$$

直接计算得其为

$$P^{-1} \operatorname{diag}(f(\lambda_1 I_{n_1}), \dots, f(\lambda_k I_{n_k})) P = P^{-1} O P = O$$

且 f 无重根, 从而得证。

右推左: 若有无重根多项式 f 使得 $f(A) = O$, 设 A 的 Jordan 标准形为

$$A = P^{-1} \operatorname{diag}(J_{k_1}(\lambda_1), \dots, J_{k_s}(\lambda_s)) P$$

则代入并消去可知

$$\operatorname{diag}(f(J_{k_1}(\lambda_1)), \dots, f(J_{k_s}(\lambda_s))) = O$$

从而每个对角块 $J_k(\lambda_0)$ 都满足 $f(J_k(\lambda_0)) = O$ 。我们下面只需要说明 $k \geq 2$ 时这件事不可能发生, 就足够得到证明了。

设 $f(x) = (x - x_1) \dots (x - x_r)$, 由于 $f(J_k(\lambda_0))$ 需要对角元为 0、 $i - j = 1$ 位置得元素为 0, 利用 Jordan 块多项式得结论也即

$$f(\lambda_0) = nf'(\lambda_0) = 0$$

从而 $f(\lambda_0) = f'(\lambda_0) = 0$ 。我们先证明 $\gcd(f(x), f'(x)) = 1$: 若否, 利用唯一因子分解定理, 必然存在 i_0 使得 $x - x_{i_0} \mid f'(x)$, 但利用乘积求导公式可知

$$f'(x) = \sum_{i=1}^r \prod_{j \neq i} (x - x_j)$$

求和中, 当 $i \neq i_0$ 时, 对应的项包含 $x - x_{i_0}$, 是 $x - x_{i_0}$ 的倍式, 但 $i = i_0$ 时对应的项为其他 $x - x_j$ 乘积, 不为 $x - x_{i_0}$ 倍式, 矛盾, 从而得到了结论。

利用裴蜀定理, 存在 $u(x)、v(x)$ 使得 $u(x)f(x) + v(x)f'(x) = 1$, 但根据条件

$$u(\lambda_0)f(\lambda_0) + v(\lambda_0)f'(\lambda_0) = 0$$

矛盾, 因此当 $k \geq 2$ 时不可能 $f(J_k(\lambda_0)) = O$ 。

* 更简单的证明是 $J_k(\lambda_0) - x_i I_k$ 在 $x_i \neq \lambda_0$ 时可逆, 从而乘积为 O 当且仅当 λ_0 为某个 x_i 且 $J_k(x_i) - x_i I_k = O$ (由乘可逆阵不改变秩), 这就得到了 $k = 1$ 。不过, 无重根多项式满足 $\gcd(f, f')$ 的性质较为重要, 因此选择了通过此性质进行证明。

* 此后, 我们很少会再讨论复方阵对角化的判定问题了, 因为它可以被 Jordan 标准形的理论囊括。不过, 可对角化矩阵的很多良好性质仍然需要掌握。

十七 走向抽象

§17.1 习题解答

1. (王书 习题 1.2)

(1) 求 m, p, q 使得 $x^2 + mx - 1 \mid x^3 + px + q$ 。

解答:

利用次数关系可设 $x^3 + px + q = (x^2 + mx - 1)(ax + b)$, 对比三次、零次项系数发现只能 $x^3 + px + q = (x^2 + mx - 1)(x - q)$, 于是二次项、一次项系数即为方程

$$\begin{cases} q = m \\ p = -qm - 1 \end{cases}$$

都用 m 表示也即 $(m, p, q) = (m, -m^2 - 1, m)$ 。

(2) 求 m, p, q 使得 $x^2 + mx + 1 \mid x^4 + px^2 + q$ 。

解答:

利用次数关系可设 $x^4 + px^2 + q = (x^2 + mx + 1)(ax^2 + bx + c)$, 对比四次、零次项系数发现只能

$$x^4 + px^2 + q = (x^2 + mx + 1)(x^2 + bx + q)$$

再考虑三次项发现 $b = -m$, 从而

$$x^4 + px^2 + q = (x^2 + mx + 1)(x^2 - mx + q)$$

二次项、一次项系数即为方程

$$\begin{cases} p = q + 1 - m^2 \\ 0 = mq - m \end{cases}$$

讨论 m 是否为 0 可知 (m, p, q) 为 $(0, q+1, q)$ 或 $(m, 2-m^2, 1)$ 。

* 注意多项式乘法后最高次与最低次项都是只进行相乘不进行相加得到的，可以有效减少方程数量。

2. (王书习题 1.7) 若 $x^3 + (1+t)x^2 + 2x + 2u$ 与 $x^3 + tx + u$ 最大公因式为二次，求 t, u 。

解答：

左减右得到 $(1+t)x^2 + (2-t)x + u$ ，由最大公因式性质可知此与 $x^3 + tx + u$ 的最大公因式仍为原最大公因式。若 $t = -1$ ，这一定是一次多项式，从而最大公因式不可能为二次。否则，利用辗转相除法性质可知必然有

$$(1+t)x^2 + (2-t)x + u \mid x^3 + tx + u$$

先假设 u 非零，类似上方对比三次、零次项系数可得

$$((1+t)x^2 + (2-t)x + u) \left(\frac{1}{1+t}x + 1 \right) = x^3 + tx + u$$

由二次项、一次项系数得到方程

$$\begin{cases} \frac{2-t}{1+t} + 1 + t = 0 \\ \frac{u}{1+t} + 2 - t = t \end{cases}$$

直接解出 (t, u) 为 $(\frac{-1+\sqrt{11}i}{2}, -7 - \sqrt{11}i)$ 或 $(\frac{-1-\sqrt{11}i}{2}, -7 + \sqrt{11}i)$ 。

当 u 为零时，也即

$$(1+t)x^2 + (2-t)x \mid x^3 + tx$$

同除以 x 不影响整除性得到

$$(1+t)x + (2-t) \mid x^2 + t$$

验证 $2-t=0$ 时不满足条件，从而类似有

$$((1+t)x + (2-t)) \left(\frac{x}{1+t} + \frac{t}{2-t} \right) = x^2 + t$$

对比一次项系数得到方程

$$\frac{2-t}{1+t} + \frac{t(1+t)}{2-t} = 0$$

这是三次方程，不过可尝试得到 $t = -4$ 为根，从而分解为 $(t+4)(t^2 - t + 1)$ ，最终解出 t 为 -4 或 $\frac{1 \pm \sqrt{3}i}{2}$ 。

综合以上， (t, u) 共有 5 个解

$$(-4, 0), \left(\frac{1+\sqrt{3}i}{2}, 0 \right), \left(\frac{1-\sqrt{3}i}{2}, 0 \right), \left(\frac{-1+\sqrt{11}i}{2}, -7 - \sqrt{11}i \right), \left(\frac{-1-\sqrt{11}i}{2}, -7 + \sqrt{11}i \right)$$

* 不建议大家对带未知数的式子用长除法，很容易漏讨论情况。

3. (王书习题 1.11) 若 $f(x), g(x)$ 不全为 0, $u(x)f(x) + v(x)g(x) = \gcd(f(x), g(x))$, 证明 $\gcd(u(x), v(x)) = 1$ 。

证明：

记 $d(x) = \gcd(f(x), g(x))$ ，由条件其非零，从而由整除性 $f_1(x) = \frac{f(x)}{d(x)}$ 与 $g_1(x) = \frac{g(x)}{d(x)}$ 都为多项式。等式两侧同除以 $d(x)$ 得到

$$f_1(x)u(x) + g_1(x)v(x) = 1$$

若 $m(x)$ 同时为 $u(x)$ 、 $v(x)$ 的因式，则其也为 1 的因式，从而只能为非零常数，因此

$$\gcd(u(x), v(x)) = 1$$

* 最后一步即为裴蜀定理的逆命题，一般可以直接使用。

4. (王书习题 8.1(3)) 将以下 λ 矩阵化为标准形：

$$\begin{pmatrix} \lambda^2 + \lambda & & \\ & \lambda & \\ & & (\lambda + 1)^2 \end{pmatrix}$$

解答：

这里介绍直接行列变换的做法。由本讲义 15.2.2 引理 3 的方式对右下角二阶子矩阵进行操作可变换为（具体流程：第三列加第二列 1 倍，第三行减第二行 $\lambda + 2$ 倍，第二行减第三行 λ 倍，第二列加第三列 $\lambda^2 + 2\lambda$ 倍）

$$\begin{pmatrix} \lambda^2 + \lambda & & \\ & \lambda(\lambda + 1)^2 & \\ & & 1 \end{pmatrix}$$

交换行列得到

$$\begin{pmatrix} 1 & & \\ & \lambda^2 + \lambda & \\ & & \lambda(\lambda + 1)^2 \end{pmatrix}$$

这时已经符合互相整除要求，从而为 Smith 标准形。

* 注意 Smith 标准形要求不变因子为首一多项式，最后可能需要某行/列乘常数。

5. (王书习题 8.2(5)) 求以下 λ 矩阵 $\text{diag}(\lambda + 1, \lambda + 2, \lambda - 1, \lambda - 2)$ 的不变因子。

解答：

同样介绍直接行列变换的做法。由本讲义 15.2.2 引理 3 的方式对左下角二阶子矩阵进行操作可变换为

$$\text{diag}(1, (\lambda + 1)(\lambda + 2), \lambda - 1, \lambda - 2)$$

再对中间两行、两列构成的二阶子矩阵进行操作得到

$$\text{diag}(1, 1, (\lambda + 1)(\lambda + 2)(\lambda - 1), \lambda - 2)$$

最后对右下角二阶子矩阵进行操作得到

$$\text{diag}(1, 1, 1, (\lambda + 1)(\lambda + 2)(\lambda - 1)(\lambda - 2))$$

这符合 Smith 标准形形式，于是其前三个不变因子为 1，最后一个为 $(\lambda + 1)(\lambda + 2)(\lambda - 1)(\lambda - 2)$ 。

* 注意说不变因子要说清楚第几个不变因子是什么。

6. (王书习题 8.3) 证明矩阵

$$\begin{pmatrix} \lambda & & a_0 \\ -1 & \lambda & a_1 \\ \ddots & \ddots & \vdots \\ & -1 & \lambda + a_{n-1} \end{pmatrix}$$

前 $n - 1$ 个不变因子为 1，最后一个为 $f(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0$ 。

证明：

通过行列式因子计算的方法见本讲义 17.2.1，仍然介绍行列变换的思路：最后一行加倒数第二行的 λ 倍、倒数第三行加倒数第二行的 λ 倍……第一行加第二行的 λ 倍，可得到

$$\begin{pmatrix} & f_0(\lambda) \\ -1 & f_1(\lambda) \\ \ddots & \vdots \\ -1 & f_{n-1}(\lambda) \end{pmatrix}$$

这里归纳可发现

$$f_i(\lambda) = \lambda^{n-i} + a_{n-1}\lambda^{n-i-1} + a_{n-2}\lambda^{n-i-2} + \cdots + a_{i+1}\lambda + a_i$$

从而即有 $f_0(\lambda) = f(\lambda)$ 。

最后一列加上第一列 $f_1(\lambda)$ 倍、第二列 $f_2(\lambda)$ 倍……第 $n-1$ 列 $f_{n-1}(\lambda)$ 倍得到

$$\begin{pmatrix} & f(\lambda) \\ -1 & \\ \ddots & \\ -1 & \end{pmatrix}$$

给前 $n-1$ 列乘 -1 ，并交换行列即得到

$$\text{diag}(1, \dots, 1, f(\lambda))$$

这符合 Smith 标准形形式，从而得证。

7. (王书习题 8.5) 求

$$\begin{pmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{pmatrix}^k$$

解答：

与本讲义 16.2.3 相同方法可得结果为

$$\begin{pmatrix} \lambda^k & & \\ k\lambda^{k-1} & \lambda^k & \\ \frac{k(k-1)}{2}\lambda^{k-2} & k\lambda^{k-1} & \lambda^k \end{pmatrix}$$

8. (王书习题 8.6(1,13,14)) 计算以下矩阵的 Jordan 标准形：

$$(1) \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ -2 & -2 & -1 \end{pmatrix}$$

解答：

计算其特征方阵为

$$\begin{pmatrix} \lambda - 1 & -2 & 0 \\ 0 & \lambda - 2 & 0 \\ 2 & 2 & \lambda + 1 \end{pmatrix}$$

其左下角二阶行列式为 $2(\lambda - 2)$ 、一三两行一三两列交出的行列式为 $(\lambda - 1)(\lambda + 1)$ ，最大公因式为 1，因此二阶行列式因子为 1。由此，前两个不变因子只能为 1，第三个不变因子为行

列式对应的首一多项式。直接计算可知行列式为 $(\lambda - 1)(\lambda - 2)(\lambda + 1)$, 因此不变因子分别为 1、1、 $(\lambda - 1)(\lambda - 2)(\lambda + 1)$, 初等因子组为 $\lambda - 1, \lambda - 2, \lambda + 1$ 。于是其 Jordan 标准形为

$$\text{diag}(J_1(1), J_1(2), J_1(-1)) = \text{diag}(1, 2, -1)$$

$$(2) \begin{pmatrix} 1 & -3 & 0 & 3 \\ -2 & 6 & 0 & 13 \\ 0 & -3 & 1 & 3 \\ -1 & 2 & 0 & 8 \end{pmatrix}$$

解答:

计算其特征方阵为

$$\begin{pmatrix} \lambda - 1 & 3 & 0 & -3 \\ 2 & \lambda - 6 & 0 & -13 \\ 0 & 3 & \lambda - 1 & -3 \\ 1 & -2 & 0 & \lambda - 8 \end{pmatrix}$$

第四列加第二列得到

$$\begin{pmatrix} \lambda - 1 & 3 & 0 & 0 \\ 2 & \lambda - 6 & 0 & \lambda - 19 \\ 0 & 3 & \lambda - 1 & 0 \\ 1 & -2 & 0 & \lambda - 10 \end{pmatrix}$$

第二行减第四行两倍得到

$$\begin{pmatrix} \lambda - 1 & 3 & 0 & 0 \\ 0 & \lambda - 2 & 0 & -(\lambda - 1) \\ 0 & 3 & \lambda - 1 & 0 \\ 1 & -2 & 0 & \lambda - 10 \end{pmatrix}$$

第一行减第三行得到

$$\begin{pmatrix} \lambda - 1 & 0 & -(\lambda - 1) & 0 \\ 0 & \lambda - 2 & 0 & -(\lambda - 1) \\ 0 & 3 & \lambda - 1 & 0 \\ 1 & -2 & 0 & \lambda - 10 \end{pmatrix}$$

由于左下角二阶行列式为 -3 , 第二个行列式因子为 1。直接计算行列式可知第四个行列式因子为 $(\lambda - 1)^2(\lambda^2 - 14\lambda + 19)$, 下面计算第三个行列式因子。

取出的三行三列只要包含第一行或第三行, 按第一行或第三行展开即得到其为 $\lambda - 1$ 的倍数。另一方面, 直接计算得 2、3、4 行与 1、2、4 列交出的子式为 $3(\lambda - 1)$, 因此第三个行列式因子即为 $\lambda - 1$, 从而所有不变因子为 $1, 1, \lambda - 1, (\lambda - 1)(\lambda^2 - 14\lambda + 19)$, 分解得到初等因子组, 进而算出 Jordan 标准形为

$$\text{diag}(1, 1, 7 + \sqrt{30}, 7 - \sqrt{30})$$

(3)

$$\begin{pmatrix} & & & 1 \\ 1 & & & \\ & \ddots & & \\ & & & 1 \end{pmatrix}$$

解答:

计算其特征方阵为

$$\begin{pmatrix} \lambda & & & -1 \\ -1 & \lambda & & \\ & \ddots & \lambda & \\ & & -1 & \lambda \end{pmatrix}$$

由于此方阵符合作业题第 6 题的形式，可知其前 $n-1$ 个不变因子为 1，最后一个为 $\lambda^n - 1$ 。直接因式分解可知其初等因子组为

$$\lambda - 1, \quad \lambda - \xi, \quad \dots, \quad \lambda - \xi^{n-1}$$

这里 $\xi = e^{2\pi i/n}$ 。由此可知其 Jordan 标准形

$$\text{diag}(1, \xi, \dots, \xi^{n-1})$$

* 行列变换到一定形式后再直接计算行列式因子往往是最快的方法。

9. (丘书 习题 8.1.23(1)) 已知

$$\alpha_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \alpha_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \alpha_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \alpha_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\beta_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \beta_2 = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 1 \end{pmatrix}, \beta_3 = \begin{pmatrix} 3 \\ 1 \\ -2 \\ 0 \end{pmatrix}, \beta_4 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 2 \end{pmatrix}$$

求 $\{\alpha_i\}$ 到 $\{\beta_i\}$ 的过渡矩阵与 $(x_1, x_2, x_3, x_4)^T$ 在 $\{\beta_i\}$ 下的坐标。

解答：

将所有 α 拼成矩阵 A ，将所有 β 拼成矩阵 B ，根据过渡矩阵 P 定义有 $B = AP$ ，由于 $A = I$ 即得 $P = B$ 。

由于 $(x_1, x_2, x_3, x_4)^T$ 即为基 $\{\alpha_i\}$ 下的坐标，利用坐标变换性质可知在 $\{\beta_i\}$ 下的坐标为

$$P^{-1} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -13x_1 + 17x_2 - 11x_3 - 14x_4 \\ x_1 - x_2 + x_3 + x_4 \\ 4x_1 - 5x_2 + 3x_3 + 4x_4 \\ 6x_1 - 8x_2 + 5x_3 + 7x_4 \end{pmatrix}$$

* 不熟悉性质可以考虑直接由定义构造方程组推导。

10. (丘书 习题 8.1.25) 在 \mathbb{K} 上线性空间 $\mathbb{K}[x]_n$ 中，求基 $1, x, \dots, x^{n-1}$ 到基 $1, x - a, \dots, (x - a)^{n-1}$ 的过渡矩阵， $a \in \mathbb{K}$ 为非零常数。

解答：

利用过渡矩阵 P 定义即

$$(1, x - a, \dots, (x - a)^{n-1}) = (1, x, \dots, x^{n-1})P$$

由此 p_{ij} 为 $(x - a)^{j-1}$ 中 x^{i-1} 的系数，当 $j < i$ 时为 0，否则为 $C_{j-1}^{i-1}(-a)^{j-i}$ 。

11. (丘书 习题 8.1.26)

(1) 证明 $\mathbb{K}[x]_n$ 中多项式组

$$f_i(x) = \prod_{\substack{1 \leq j \leq n \\ j \neq i}} (x - a_j)$$

构成其一组基，这里 $a_1, \dots, a_n \in \mathbb{K}$ 互不相同。

证明：

若 $\lambda_1 f_1(x) + \dots + \lambda_n f_n(x) = 0$ ，代入 $x = a_i$ 可以发现当 $j \neq i$ 时 $f_j(a_i) = 0$ ，而 $j = i$ 时 $f_i(a_i) \neq 0$ ，从而得到 $\lambda_i = 0$ 。对 $i = 1, \dots, n$ 考虑即得所有 λ_i 全为 0，因此 $f_1(x), \dots, f_n(x)$ 线性无关。

另一方面，此向量组的个数为 n ，与 $\mathbb{K}[x]_n$ 维数相等，因此它们即为一组基。

(2) 取 $\mathbb{K} = \mathbb{C}$ ， $a_i = \xi^{i-1}$ 时，求 $1, x, \dots, x^{n-1}$ 到基 $f_1(x), \dots, f_n(x)$ 的过渡矩阵。这里 $\xi = e^{2\pi i/n}$ 。

解答：

由于 $(x - 1)(x - \xi) \dots (x - \xi^{n-1}) = x^n - 1$ ，且 $\xi^n = 1$ ，可发现

$$f_j(x) = \frac{x^n - 1}{x - \xi^{j-1}} = \frac{x^n - (\xi^{j-1})^n}{x - \xi^{j-1}} = x^{n-1} + \xi^{j-1}x^{n-2} + \dots + \xi^{(n-1)(j-1)}$$

用过渡矩阵 P 定义有

$$(f_1(x), f_2(x), \dots, f_n(x)) = (1, x, \dots, x^{n-1})P$$

由此 p_{ij} 为 $f_j(x)$ 中 x^{i-1} 的系数，为 $\xi^{(n-i)(j-1)}$ 。

§17.2 有理标准形

17.2.1 一般数域上的相似

在解决了 \mathbb{C} 上的相似后，我们下面来讨论一般数域 \mathbb{K} 上的相似。我们现在终于可以证明相似最重要的定理之一了：两个 \mathbb{K} 上的方阵相似当且仅当它们看作 \mathbb{C} 上的方阵相似。换句话说，对 $A, B \in \mathbb{K}^{n \times n}$ ，若存在 \mathbb{C} 上的可逆阵 P 使得 $A = P^{-1}BP$ ，则一定存在 \mathbb{K} 上的可逆阵 Q 使得 $A = Q^{-1}BQ$ ，反之亦然。

证明：

若 \mathbb{K} 上方阵 A, B 相似，存在 \mathbb{K} 上的可逆阵 Q 使得 $A = Q^{-1}BQ$ ，由于 Q 也可以看成 \mathbb{C} 上可逆方阵，它们在 \mathbb{C} 上仍然相似。

若 \mathbb{K} 上方阵 A, B 看作 \mathbb{C} 上方阵相似， $\lambda I - A$ 与 $\lambda I - B$ 看作 \mathbb{C} 上的多项式方阵模相抵，因此具有相同的不变因子组。但是，由于 $\lambda I - A$ 与 $\lambda I - B$ 实际上是 \mathbb{K} 上的多项式方阵，根据不变因子组在不同域中的不变性（见本讲义 15.3.3），将它们看作 \mathbb{K} 上的多项式方阵时 Smith 标准形都不变，因此仍然相等，从而仍然模相抵，这就得到 A, B 作为 \mathbb{K} 上的方阵相似。

* 不变因子组扩域不变性质的重要性在这个证明中完全体现了。

一个很自然的问题是，既然任何数域 \mathbb{K} 上的相似都可以被 \mathbb{C} 上的相似给解决，那还有什么值得研究的呢？这里，我们必须再做一点抽象的说明：由于相似是一个等价关系，对某个 \mathbb{K} 上矩阵 A ，与它相似的矩阵事实上是一个等价类，不妨记作 $[A]_{\mathbb{K}}$ ，而 Jordan 标准形是其中的代表元。

两个 \mathbb{K} 上的方阵相似当且仅当它们看作 \mathbb{C} 上的方阵相似，可以归为如下式子：

$$[A]_{\mathbb{K}} = \mathbb{K}^{n \times n} \cap [A]_{\mathbb{C}}$$

这个式子的含义是， A 在 \mathbb{K} 上的相似等价类是 A 在 \mathbb{C} 上的相似等价类中所有 \mathbb{K} 上的方阵。

此处我们可以发现，虽然 \mathbb{K} 上的相似可以被 \mathbb{C} 上的相似刻画， $[A]_{\mathbb{K}}$ 与 $[A]_{\mathbb{C}}$ 仍然是不同的，一个很直观的影响是无法取到相同的代表元。例如，对没有实特征值的实方阵 $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ ，它的 Jordan 标准形 $\text{diag}(i, -i)$ 在 \mathbb{C} 上的相似等价类中，但不在 \mathbb{K} 上的相似等价类中。因此，对于更小的等价类 $[A]_{\mathbb{K}}$ ，我们至少希望重新取出合适的代表元，这也就是所谓的有理标准形（叫这个名字大概是因为有理数域 \mathbb{Q} 是最小的数域）。

要想构造出有理标准形，我们还是回到本讲义 16.2.1 的四步过程。此外，由于我们已经证明了 \mathbb{K} 上的多项式方阵的初等因子组的良好性质，我们可以直接从初等因子组构造，这里设 $A \in \mathbb{K}^{n \times n}$ ：

- 构造特征方阵 $\lambda I - A$ ；
- 计算 $\lambda I - A$ 的 Smith 标准形，从而得到全部不变因子；
- 计算 $\lambda I - A$ 的初等因子组 $p_1(\lambda)^{k_1}, \dots, p_s(\lambda)^{k_s}$ ；
- 找固定简单形式的 \mathbb{K} 上方阵 B_i 使得 $\lambda I - B_i$ 的初等因子组只有 $p_i(\lambda)^{k_i}$ ；
- 将 $\text{diag}(B_1, \dots, B_s)$ 定义为 A 的相似标准形。

由于最后一个行列式因子的次数仍然为 n ，我们有

$$k_1 \deg p_1(\lambda) + k_2 \deg p_2(\lambda) + \dots + k_s \deg p_s(\lambda) = n$$

于是与本讲义 16.2.2 完全相同得到两个 \mathbb{K} 上的方阵相似当且仅当对应的特征方阵初等因子组相同，无需再单独叙述阶数相等。此外，只要这样的 B_i 对每个 $p_i(\lambda)^{k_i}$ 是确定的，其相似标准形在交换对角块视为相同的意义下还是唯一的。

因此，我们最终需要解决的问题是，任给一个 $p(\lambda)^k$ ，其中 $p(\lambda)$ 是 $\mathbb{K}[\lambda]$ 中的不可约多项式，找一个 \mathbb{K} 上方阵 B （根据阶数计算，其应为 $k \deg p$ 阶）使得 $\lambda I - B$ 初等因子组只有 $p(\lambda)^k$ 。

17.2.2 一般数域上的相似标准形

为了构造这样的 B ，我们先给出一个引理：设 \mathbb{K} 上多项式 $f(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$ ，称 $\mathbb{K}^{n \times n}$ 中的矩阵

$$F(f) = \begin{pmatrix} & & -a_0 \\ 1 & & -a_1 \\ & \ddots & \vdots \\ & & 1 & -a_{k-1} \end{pmatrix}$$

为 f 的友方阵，其特征方阵前 $n-1$ 个不变因子为 1，最后一个为 $f(\lambda)$ 。

证明：

直接计算可知

$$\lambda I - F(f) = \begin{pmatrix} \lambda & & a_0 \\ -1 & \lambda & a_1 \\ \ddots & \ddots & \vdots \\ & -1 & \lambda + a_{k-1} \end{pmatrix}$$

由于其左下角 $n-1$ 阶子式为 ± 1 ，其第 $n-1$ 个行列式因子必然为 1，从而与本讲义 16.2.2 相同得其前 $n-1$ 个不变因子都为 1，第 n 个不变因子为其行列式 $\det(\lambda I - F(f))$ 。

记 $g(\lambda) = \lambda^{n-1} + a_{n-1}\lambda^{n-2} + \dots + a_2\lambda + a_1$ ，将原行列式按第一行展开可发现左上角对应的代数余子式是 $g(\lambda)$ 的友方阵的特征方阵的行列式， a_0 对应的代数余子式是对角阵的行列式乘 -1 的次方，从而

$$\det(\lambda I - F(f)) = \lambda \det(\lambda I - F(g)) + a_0(-1)^{n+1}(-1)^{n-1} = \lambda \det(\lambda I - F(g)) + a_0$$

由于一阶时其即为 $\lambda + a_0$ ，利用上式归纳得到结论。

由此, 只要 $f(\lambda)$ 是不可约多项式 $p(\lambda)$ 的 k 次方, $F(f)$ 的特征方阵的初等因子组就只有 $p^k(\lambda)$ 。另外, 不同的多项式 f 对应着不同的 $F(f)$, 从而可以得到 \mathbb{K} 上的**有理标准形**结论: 对任何 \mathbb{K} 上的方阵 A , 存在一系列 \mathbb{K} 上的首一不可约多项式 $p_1(\lambda), \dots, p_s(\lambda)$ 与对应次方 k_1, \dots, k_s 使得存在 \mathbb{K} 上的可逆矩阵 P 满足

$$A = P^{-1} \operatorname{diag}(F(p_1^{k_1}), \dots, F(p_s^{k_s}))P$$

其中 F 代表友方阵。此标准形在交换对角块视为相同的意义下唯一, 证明与 16.2.2 最后一段相同。

利用这个标准形, 我们可以完全解决任何数域 \mathbb{K} 上的相似问题, 由于友方阵对应的计算相对复杂, 我们将在学完空间相关知识后再给出它的一些应用。

* 此标准形的性质事实上与 \mathbb{K} 上的**线性空间理论**密切相关, 因此到这里明白它可以完全解决相似标准形问题并学会计算即可, 并不需要建立更深的理解。

在本节的最后, 我们需要指出, 上述的有理标准形构造 (也是教材中的构造) 确实解决了任意 \mathbb{K} 上矩阵的相似标准形问题, 但会出现一个很尴尬的情况: 当 $\mathbb{K} = \mathbb{C}$ 时, **有理标准形与 Jordan 标准形并不相同**, 例如 $(\lambda - 1)^2$ 对应的友方阵与 Jordan 块分别为

$$\begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

从直观上甚至无法一眼看出它们是相似的, 即使它们确定对应相同的单个初等因子。由于 \mathbb{C} 上的相似当然也是某个数域 \mathbb{K} 上的相似, 我们也希望能从 Jordan 标准形的推广出发定义有理标准形。这时的定义如下: 对任何 \mathbb{K} 上的方阵 A , 存在一系列 \mathbb{K} 上的首一不可约多项式 $p_1(\lambda), \dots, p_s(\lambda)$ 与对应次方 k_1, \dots, k_s 使得存在 \mathbb{K} 上的可逆矩阵 P 满足

$$A = P^{-1} \operatorname{diag}(F_{k_1}(p_1), \dots, F_{k_s}(p_s))P$$

其中有理块

$$F_k(p) = \begin{pmatrix} F(p) & & & \\ E_{1,\deg p} & F(p) & & \\ & \ddots & \ddots & \\ & & E_{1,\deg p} & F(p) \end{pmatrix}$$

这里 $F(p)$ 表示 p 的友方阵, 阶数为 $\deg p$, 其对角上排列了 k 个 $F(p)$, 因此总阶数为 $k \deg p$ 。此外, $E_{1,\deg p}$ 表示只有右上角元素为 1、其他为 0 的 $\deg p$ 阶方阵。此形式在交换对角块视为相同的意义下唯一。

证明:

我们事实上只需要证明 $F_k(p)$ 的特征方阵的初等因子组只有 $p^k(\lambda)$, 而观察可发现 $\lambda I - F_k(p)$ 所有 $i = j + 1$ 位置的元素都为 -1 , 因此左下角 $k \deg p - 1$ 阶子式为 1, 这就说明了前 $k \deg p - 1$ 个不变因子均为 1, 最后一个不变因子为行列式。利用已证 $\deg(\lambda I - F(p)) = p(\lambda)$, 由准三角阵行列式结论即得

$$\deg(\lambda I - F_k(p)) = \deg(\lambda I - F(p))^k = p^k(\lambda)$$

从而得证。

可以发现, 对这个标准形 (我们姑且称为 **Jordan-有理标准形**, 书上叫法是广义 Jordan 标准形), 当 $\mathbb{K} = \mathbb{C}$ 时, 由于 \mathbb{C} 上的不可约多项式只能是 $\lambda - \lambda_0$, 其对应的友方阵即 λ_0 , 因此 $F_k(\lambda - \lambda_0)$ 成为了下三角 **Jordan 块** $J_k^{(L)}(\lambda_0)$ 。这样, 我们就完成了一般 \mathbb{K} 上的相似与 \mathbb{C} 上的相似的统一。

* 事实上, 上述所有讨论都不止可以在数域进行。对抽象代数中定义的、更一般的域, 都可以谈论其中的有理标准形或 Jordan-有理标准形, 而能谈论 Jordan 标准形的要求是代数闭域, 也即满足其上的不可约

多项式只有一次的域。于是，从下半学期讲义开头到现在的讨论完全解决了域上相似的问题。接下来，我们就需要从空间的角度理解这些标准形到底意味着什么了。

* 哪怕只从矩阵论来看，这些标准形也是十分有用的，之前求解 $A^n = I$ 只是最基本的例子，之后我们也将面对更复杂的应用。

17.2.3 “表示”

引入空间前，我们最后来说一点“抽象废话”。

观察本讲义 17.2.1 开头对 \mathbb{K} 上相似等价于 \mathbb{C} 上相似或本讲义 16.2.2 结尾对 Jordan 标准形唯一性的证明，可以发现这些证明好像和我们习惯的数学证明有一定区别：比起推导和计算，它更多通过文字描述逻辑完成了证明。

细究它通过文字描述了怎样的逻辑，可以发现，除了核心的过程外，基本在进行语言的转换。就以 \mathbb{K} 上相似等价于 \mathbb{C} 上相似的证明为例，证明过程是基于以下六个命题依次等价：

- A, B 作为 \mathbb{K} 上的方阵相似；
- $\lambda I - A, \lambda I - B$ 作为 \mathbb{K} 上的多项式方阵模相抵；
- $\lambda I - A, \lambda I - B$ 作为 \mathbb{K} 上的多项式方阵不变因子相同；
- $\lambda I - A, \lambda I - B$ 作为 \mathbb{C} 上的多项式方阵不变因子相同；
- $\lambda I - A, \lambda I - B$ 作为 \mathbb{C} 上的多项式方阵模相抵；
- A, B 作为 \mathbb{C} 上的方阵相似。

证明的核心事实上是三、四两个命题等价这步，而这是之前讨论不变因子的性质时已经说明的，对于其他步骤，我们似乎都只是在对同一个命题改变表述。准确来说，在 \mathbb{K} 上方阵相似与 $\mathbb{K}[\lambda]$ 上方阵模相抵两套语言中，我们找到了同一件事的不同表述，而在 \mathbb{K} 上方阵相似这套语言中难以解决的问题，在 $\mathbb{K}[\lambda]$ 上方阵模相抵这套语言里可以找到解决方案——因为解决问题的关键不变因子在相似中几乎没有对应的概念。

这样的情况在之前的证明中也经常发生：一组向量线性相关等价于其中某个向量可被其他向量线性表示是在线性相关与线性表出两套语言中切换（之前谈到模时以往的方法失效就是因为整线性相关与整线性表出不再等价）、行秩等于列秩等于最大非零子式阶数是在向量组与矩阵两套语言中切换……即使证明等价的过程可能相对复杂，证明等价后我们可以掌握更多工具解决问题。

——以更高一点的观点来看，这样“不同语言之间的等价切换”在数学上往往称为表示，有一门名叫表示论的数学分支就专门研究这样的等价切换（主要是以线性结构表示）。

回到课程中来，如果说上学期我们研究的主题是矩阵是什么、矩阵有怎样的性质，这学期的研究核心则是矩阵能表示什么、怎么从矩阵说明被表示对象的性质与怎么从对被表示对象的研究中反推矩阵的性质。

因为矩阵是一个线性结构，矩阵能表示的东西往往都与一般的线性结构，线性空间，密切相关。例如， $A \in \mathbb{K}^{m \times n}$ 可以表示 n 维线性空间到 m 维线性空间的线性映射，矩阵的秩可以表示线性映射像空间的维数， $A \in \mathbb{K}^{n \times n}$ 可以表示一个 n 维线性空间上的线性变换，可逆的 n 阶的方阵可以表示 n 维线性空间两组基之间的变换与对应的坐标变换， n 阶对称实方阵可以表示 n 维欧氏空间的一个自伴的线性变换， n 阶正定矩阵可以表示 n 维欧氏空间上一组基的内积矩阵……

值得一提的是，矩阵能表示的对象局限于有限维线性空间，反之，有限维线性空间的理论也可以表示一切矩阵的理论。但当空间来到无限维时，不再能通过矩阵表示，也因此具有很多更复杂的性质——不过正如之前所说，这学期研究的核心是矩阵能表示什么，因此几乎不涉及无穷维的线性空间。出于严谨性，习题课讲义涉及的结论会标注只在有限维线性空间成立还是在无穷维也成立，虽然我们并不会直接要求考虑无穷维线性空间，但一个重要的建议是，对无穷维线性空间也成立的结论尽量不要采用矩阵理论相关的证明方法，因为这阻碍了结论的推广。

就像模相抵理论解决了矩阵的相似标准形问题，一些矩阵理论不好解决的问题在空间视角可能易于解决，本学期我们还将看到很多这样的例子，尤其是与线性变换，也就是方阵相关时。

* 某种意义上，本学期对有限维线性空间的学习只是为了获取解决矩阵问题的工具，所以一定要注意如何还原到矩阵去表述。

§17.3 期末考试

17.3.1 试题

1. 设三阶实对称矩阵

$$A = \begin{pmatrix} 1 & \lambda & -1 \\ \lambda & 4 & 2 \\ -1 & 2 & 4 \end{pmatrix}$$

为正定矩阵，求 λ 的取值范围。

2. 设 A 为数域 \mathbb{K} 上的三阶方阵， α, β, γ 分别为 A 的属于特征值 $0, 2, 3$ 的一个特征向量：

- (a) 求关于 x 的线性方程组 $Ax = \beta + \gamma$ 的全部解 (用 α, β, γ 的线性组合表示);
(b) 若关于 x 的线性方程组 $Ax = a\beta + b\beta + c\gamma$ 无解，求 a, b, c 满足的条件。

3. 设下标从 0 开始的数列 $\{a_k\}$ 满足 $a_0 = 0, a_1 = \frac{1}{2}$ ，且对任何 $n \geq 0$ 有

$$a_{n+2} = \frac{1}{2}(a_{n+1} + a_n)$$

求该数列的通项公式及 n 趋于无穷时的 a_n 。

4. 在 \mathbb{R}^4 中，设

$$\alpha_1 = (1, 0, 2, 1)^T, \quad \alpha_2 = (2, 0, 1, 1)^T, \quad \alpha_3 = (3, 0, 3, 0)^T$$

令 W 为 $\alpha_1, \alpha_2, \alpha_3$ 生成的子空间，求 $\beta = (1, 2, 3, -2)^T$ 在 W 上的正交投影向量 γ ，即 $\gamma \in W$ 且 $\beta - \gamma$ 与 W 中所有向量均正交。

5. 设实二次型

$$f(x) = 2x_1x_2 + 2x_1x_3 - 2x_1x_4 - 2x_2x_3 + 2x_2x_4 + 2x_3x_4, \quad x = (x_1, x_2, x_3, x_4)^T$$

- (a) 求二次型 $f(x)$ 的矩阵 A 及其特征值、特征向量；
(b) 求正交矩阵 P 与对角矩阵 D 使得 $P^TAP = D$ ；
(c) 求 $f(x)$ 在满足 $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$ 时的最大值和最小值，并确定在何处取到。

6. (a) 证明矩阵

$$A = \begin{pmatrix} 2 & 0 & 0 \\ u & 2 & 0 \\ v & w & -1 \end{pmatrix}$$

可对角化的充要条件是 $u = 0$ ；

- (b) 设 n 阶矩阵 A 满足 $A^2 - 5A + 6I = O$ ，判断 A 是否一定可以对角化，并说明理由或举出反例。

7. 设 n 阶方阵 A, B 满足 $\text{rank } A + \text{rank } B < n$ ，证明 A, B 有公共特征值和特征向量。

8. 设 A 是正定矩阵， B 是半正定矩阵，证明 $\det(A + B) \geq \det A + \det B$ 。

17.3.2 解答

1. 注意正定性最易于计算的判定定理：对称阵正定等价于所有顺序主子式为正。由此，直接计算其一、二、三阶顺序主子式分别为 1、 $4 - \lambda^2$ 、 $8 - 4\lambda - 4\lambda^2$ ，于是得到

$$\lambda \in (-2, 2) \cap (-2, 1) = (-2, 1)$$

2. 由于不同特征值对应的特征向量线性无关， α, β, γ 构成 \mathbb{K}^3 的一组基，也即任何三维 \mathbb{K} 中向量可以写为 $\lambda_1\alpha + \lambda_2\beta + \lambda_3\gamma$ ，这里 $\lambda_{1,2,3} \in \mathbb{K}$ 。

- (a) 设 $x = \lambda_1\alpha + \lambda_2\beta + \lambda_3\gamma$ ，则由特征向量定义

$$Ax = \lambda_1 A\alpha + \lambda_2 A\beta + \lambda_3 A\gamma = 2\lambda_2\beta + 3\lambda_3\gamma$$

利用线性无关性可知 $\lambda_2 = \frac{1}{2}$ 、 $\lambda_3 = \frac{1}{3}$ ， $\lambda_1 \in \mathbb{K}$ 可任取。

- (b) 设 $x = \lambda_1\alpha + \lambda_2\beta + \lambda_3\gamma$ ，同上计算得

$$Ax = 2\lambda_2\beta + 3\lambda_3\gamma$$

若 $a \neq 0$ ，利用基的线性无关性可知原方程组无解；若 $a = 0$ ，取 $\lambda_1 = 0$ 、 $\lambda_2 = \frac{1}{2}b$ 、 $\lambda_3 = \frac{1}{3}c$ 即为解。综合这两部分可知无解当且仅当 $a \neq 0$ 。

3. 事实上这题可以有很多不同的理解与做法，我们采用矩阵乘法的方案。递推式可以写为

$$\begin{pmatrix} a_{n+2} \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix}$$

由此可发现

$$\begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}^n \begin{pmatrix} a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}^n \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}$$

直接计算可发现 $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}$ 特征值为 1 与 $-\frac{1}{2}$ ，对应的特征向量分别是 $(1, 1)^T$ 与 $(-1, 2)^T$ ，由此将 $(\frac{1}{2}, 0)^T$ 表示为特征向量线性组合有

$$\begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}^n \left(\frac{1}{3} \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} -1 \\ 2 \end{pmatrix} \right)$$

利用乘法结合律与数乘可以交换位置可发现当 $A\alpha = \lambda\alpha$ 时 $A^n\alpha = \lambda^n\alpha$ ，从而

$$\begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} = \frac{1}{3} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}^n \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}^n \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{6} \frac{(-1)^n}{2^n} \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

对比第二个分量得到

$$a_n = \frac{1}{3} \left(1 - \frac{(-1)^n}{2^n} \right)$$

其在 $n \rightarrow +\infty$ 时极限为 $\frac{1}{3}$ 。

4. 利用最小二乘知识，此问题可以直接转化为找 $\lambda_1, \lambda_2, \lambda_3$ 使得 $\|\beta - \lambda_1\alpha_1 - \lambda_2\alpha_2 - \lambda_3\alpha_3\|^2$ 最小，然后套用最小二乘的公式。我们这里给出一个只需要定义，不需要任何特殊前置知识的做法。

首先，通过向量组的初等变换可以将基 $\alpha_1, \alpha_2, \alpha_3$ 化作更简单的形式（类似线性方程组化为简化阶梯形）

$$\gamma_1 = (1, 0, 0, 0)^T, \gamma_2 = (0, 0, 1, 0)^T, \gamma_3 = (0, 0, 0, 1)^T$$

由此, $W = \{(a, 0, b, c)^T \mid a, b, c \in \mathbb{R}\}$, 设所求向量 $\gamma = (a, 0, b, c)^T$, 根据条件可知

$$(\beta - \gamma)^T \gamma_1 = (\beta - \gamma)^T \gamma_2 = (\beta - \gamma)^T \gamma_3 = 0$$

直接计算可发现这三个方程就是 $1 - a = 0$ 、 $3 - b = 0$ 、 $-2 - c = 0$, 从而有

$$\gamma = (1, 0, 3, -2)^T$$

5. (a) 根据定义可知

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

计算得其有单特征值 -3 与三重特征值 $1, -3$ 对应的一个特征向量为 $(1, -1, -1, 1)^T$, 1 对应的特征子空间一组基为 $(-1, 0, 0, 1)^T, (1, 0, 1, 0)^T, (1, 1, 0, 0)^T$ 。

(b) 利用 Schmidt 正交化等方法可以构造出特征值 -3 的特征子空间的一组标准正交基

$$\left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2} \right)^T$$

特征值 1 的特征子空间的一组标准正交基 (不唯一, 有多种取法, 这个非常漂亮的构造来自同学)

$$\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right)^T, \quad \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2} \right)^T, \quad \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2} \right)^T$$

由此可取

$$P = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}, \quad D = \text{diag}(-3, 1, 1, 1)$$

(c) 由于 $P^T A P = D$, 设 $y = P^T x$ 有

$$f(x) = x^T A x = x^T P D P^T x = y^T D y = -3y_1^2 + y_2^2 + y_3^2 + y_4^2$$

而条件 $x^T x = 1$ 即为 $y^T y = y_1^2 + y_2^2 + y_3^2 + y_4^2 = 1$, 从而又有

$$f(x) = -4y_1^2 + 1$$

由条件 $y_1 \in [-1, 1]$, 从而 $f(x)$ 的最小值在 $y_1 = \pm 1$ 时 (此时 y 其他分量只能为 0, 由 $x = Py$ 即得 x 是特征值 -3 的特征向量) 取到, 为 -3 ; $f(x)$ 的最大值在 $y_1 = 0$ 时 (由 $x = Py$ 即得 x 是特征值 1 的特征向量) 取到, 为 1 。

* 更一般的情况可见本讲义 11.3.1。

6. (a) 直接计算特征多项式可得其有特征值 $2, -1$, 代数重数分别为 $2, 1$ 。由于可对角化等价于所有特征值代数重数等于几何重数, 且代数重数为 1 时几何重数只能为 1 , A 可对角化等价于特征值 2 的几何重数为 2 。

根据几何重数的定义, A 可对角化等价于 $\text{rank}(2I - A) = 1$, 直接写出此等式

$$\text{rank} \begin{pmatrix} 0 & 0 & 0 \\ -u & 0 & 0 \\ -v & -w & 3 \end{pmatrix} = 1$$

第一列加上第三列的 $v/3$ 倍、第二列加上第三列的 $w/3$ 倍可将此等式化为

$$\text{rank} \begin{pmatrix} 0 & 0 & 0 \\ -u & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix} = 1$$

考虑行秩可发现 $u = 0$ 时左侧秩为 1, 否则秩为 2, 从而等式成立当且仅当 $u = 0$, 得证,

(b) 我们来证明 A 一定可对角化。

利用 Sylvester 秩不等式, 由 $(A - 2I)(A - 3I) = O$ 可得

$$\text{rank}(A - 2I) + \text{rank}(A - 3I) \leq n + \text{rank}(A - 2I)(A - 3I) = n$$

另一方面有

$$\text{rank}(A - 2I) + \text{rank}(A - 3I) \geq \text{rank}((A - 2I) - (A - 3I)) = \text{rank } I = n$$

综合得到

$$\text{rank}(A - 2I) + \text{rank}(A - 3I) = n$$

将它改写成

$$(n - \text{rank}(A - 2I)) + (n - \text{rank}(A - 3I)) = n$$

若 $A = 2I$ 或 $A = 3I$, 它自然可以对角化, 否则 $n - \text{rank}(A - 2I)$ 与 $n - \text{rank}(A - 3I)$ 均非零, 也即 2、3 都是 A 的特征值, 且几何重数和为 n 。由于代数重数大于等于几何重数, 且特征值代数重数和为 n , 可知 A 的特征值必然只有 2、3, 且它们的代数重数等于几何重数, 这就得到了 A 可对角化。

* 这一问也有很操作性的证明方式, 与本讲义 4.2.2 的矩阵方法类似。

7. 将其改写成

$$(n - \text{rank } A) + (n - \text{rank } B) > n$$

由于 $n - \text{rank } A$ 与 $n - \text{rank } B$ 都至多为 n , 这至少说明两者都非零, 也即 A 、 B 中 0 的几何重数都非零, 它们有公共特征值 0。

另一方面, 设 A 中 0 对应的特征子空间 (即满足 $Ax = 0$ 的 x 构成集合) 为 W_A 、 B 中 0 对应的特征子空间 (即满足 $Bx = 0$ 的 x 构成集合) 为 W_B , 上方的改写也可以看作

$$\dim W_A + \dim W_B > n$$

由此, 取出 W_A 一组基 $\alpha_1, \dots, \alpha_r$ 、 W_B 一组基 β_1, \dots, β_s , 由于总个数大于 n , 它们一定线性相关, 也即存在不全为 0 的 $\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s$ 使得

$$\sum_{i=1}^r \lambda_i \alpha + \sum_{j=1}^s \mu_j \beta_j = 0$$

移项即

$$\sum_{i=1}^r \lambda_i \alpha = - \sum_{j=1}^s \mu_j \beta_j$$

等式左侧属于 W_A 、右侧属于 W_B , 因此属于 $W_A \cap W_B$, 且无论是某个 λ_i 非零还是某个 μ_j 非零, 都能由基的性质得到等式两边非零, 从而这就是 $W_A \cap W_B$ 中的非零向量, 也就是 A 、 B 属于特征值 0 的公共特征向量。

* 我们规避了上学期尚未学习的和空间维数公式, 这学期很快就将学到。

8. 由 A 正定, 利用合同标准形为 I 可知存在可逆阵 P 使得 $A = P^T P$, 从而 (最后一个等号通过 Binet-Cauchy 公式与转置不影响行列式)

$$\det(A + B) = \det(P^T P + B) = \det(P^T(I + P^{-T} B P^{-1})P) = (\det P)^2 \det(I + P^{-T} B P^{-1})$$

记 $C = P^{-T} B P^{-1}$, 左侧即为 $(\det P)^2 \det(I + C)$ 右侧可化为

$$\det(P^T P) + \det(P^T C P) = (\det P)^2(1 + \det C)$$

由于 P^{-1} 可逆, C 与 B 合同, 因此 C 仍半正定。而由 P 可逆知 $\det P \neq 0$, 从而 $(\det P)^2 > 0$, 只需证明

$$\det(I + C) \geq 1 + \det C$$

设 C 的正交相似标准形为 $C = Q^T D Q$, 其中 D 为对角阵, Q 为正交阵, 由半正定性 D 的对角元 d_{11}, \dots, d_{nn} 非负, 计算可知 (由正交阵定义 $Q^T Q = I$) 左侧化为

$$\det(Q^T Q + Q^T D Q) = (\det Q^T)(\det Q) \det(I + D) = \det(Q^T Q) \prod_{i=1}^n (1 + d_{ii}) = \prod_{i=1}^n (1 + d_{ii})$$

而右侧为

$$1 + \det(Q^T D Q) = 1 + \det(Q^T Q) \det D = 1 + \prod_{i=1}^n d_{ii}$$

由于所有 d_{ii} 非负, 右侧为左侧乘法全展开后的其中两项, 即得左侧大于等于右侧。

* 这题另一个做法是利用引理, 半正定阵与正定阵可以同时合同对角化 (而非正交相似对角化)。不过, 同时对角化的几个引理必须分清楚条件才能用, 两个正定阵一般不能同时正交相似对角化。

§17.4 期末的延伸

本节我们将从上学期的期末卷出发, 聊到本学期研究的内容, 并给出一些线性空间相关的基本定义。整节中, 我们所有讨论的概念都是针对无穷维线性空间也成立的, 因此维数、秩等概念将放在之后真正学习时进行研究。本节的几乎所有内容都只作为介绍, 除了线性空间、交空间、线性映射的定义外目前都不要求掌握。

17.4.1 加法、数乘

首先, 什么是线性?

我们知道, 平面是一个具有线性的结构。某种意义上, 线性就是这样的平直结构。再去思考平直结构的含义, 我们可以发现有两个要素: 对任何一个方向可以进行无限延长, 且对任何两点可以进行连线。在向量空间中, 无限延长自然就对应数乘, 而连线对应 $\lambda\alpha + (1 - \lambda)\beta$ 类型的计算, 在已经允许数乘的情况下, 连线就只需要允许加法。

因此, 只要能进行数乘与加法的结构, 都具有线性。我们已经学过无数具有线性的结构了: 除了向量空间以外, 实数、复数、多项式、一般的函数、矩阵, 都是符合要求的。

具体来说, 对于一个集合 V 与数域 \mathbb{K} , 加法是一个 $V \times V \rightarrow V$ 的映射 (也就是把 V 中两个元素操作为 V 中一个元素), 数乘是一个 $\mathbb{K} \times V \rightarrow V$ 的映射 (也就是把 \mathbb{K} 、 V 中各一个元素操作为 V 中一个元素)。为了让它们符合我们对于“通常”加法与数乘的期待, 我们额外要求

$$\forall \alpha, \beta, \gamma \in V, \quad (\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$$

$$\exists z \in V, \quad \forall \alpha \in V, \quad z + \alpha = \alpha + z = \alpha$$

* 此 z 称为零元，无歧义时通常直接记作 0。

$$\forall \alpha \in V, \exists \beta \in V, \alpha + \beta = \beta + \alpha = 0$$

* 此 β 称为 α 的逆元，通常记作 $-\alpha$ 。

$$\forall \alpha, \beta \in V, \alpha + \beta = \beta + \alpha$$

$$\forall \alpha \in V, 1\alpha = \alpha$$

* 注意这个 1 是 \mathbb{K} 中的，也就是通常的数字 1。

$$\forall \lambda, \mu \in \mathbb{K}, \forall \alpha \in V, \lambda(\mu\alpha) = (\lambda\mu)\alpha$$

$$\forall \lambda, \mu \in \mathbb{K}, \forall \alpha \in V, (\lambda + \mu)\alpha = \lambda\alpha + \mu\alpha$$

* 注意左侧的加法是数的加法，右侧的加法是我们自己定义的 V 上的加法。

$$\forall \lambda \in \mathbb{K}, \forall \alpha, \beta \in V, \lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$$

* 这里对零元、逆元的定义都未保证唯一，大家可以利用性质尝试验证其唯一性。

这八条中的前四条规定加法的性质、接下来两条规定数乘的性质（事实上数乘还自带 \mathbb{K} 作为数域的一些性质），最后两条分配律则代表相容性。我们这里只对相容性做一点解释：相容性事实上保证了通过加法和数乘可以生成有限线性组合。也就是说，用 $\alpha_1, \dots, \alpha_k$ 经过有限次加法、数乘得到的一定是它们的线性组合，而任何它们的线性组合都可以通过有限次加法、数乘得到。此结论证明相对简单，可参考本讲义 2.3.3，这里就不进行详细的叙述了。我们将这样的结构称为线性空间，其中的元素称为向量——虽然它可能已经不是我们熟悉的向量了。

* 加法是 $V \times V \rightarrow V$ 的映射事实上包含了重要的性质，封闭性，也就是对任意 $\alpha, \beta \in V$ 有 $\alpha + \beta \in V$ 。同理，数乘的封闭性即对任意 $\lambda \in \mathbb{K}, \alpha \in V$ 有 $\lambda\alpha \in V$ 。

* 虽然在习题里我们可能会碰到 \mathbb{R}^+ 上乘法当成“加法”、指数当成“数乘”这样的诡异情况，在现实研究中，线性空间的加法与数乘往往是自然的，也就是，在我们知道它是线性空间之前，我们已经学会了加法和数乘如何计算，无论是多项式还是矩阵都是如此。

就像向量空间一样，对一般的线性空间，我们也可以谈论线性相关与线性无关。为了方便之后的推广，我们需要谈论无穷多个向量线性相关或线性无关，而定义很简单，只要存在有限个向量线性相关，这无穷多个向量就线性相关，否则线性无关。形式化来写，对一族线性空间 V 中的向量 $\alpha_i, i \in I$ ， I 为某下标集合，它们线性相关当且仅当

$$\exists \{i_1, \dots, i_n\} \in I, \exists \text{不全为 } 0 \text{ 的 } \lambda_1, \dots, \lambda_n \in \mathbb{K}, \lambda_1\alpha_{i_1} + \dots + \lambda_n\alpha_{i_n} = 0$$

* 之后称 I 为指标集，例如 $I = \{1, \dots, n\}$ 就对应着有限个向量，而 $I = \mathbb{N}$ 就可以对应可数无穷个向量。

* 这里的写法必须非常谨慎：由于 α_i 可能会有相同的，所有 α_i 未必是集合，也就不能加大括号；而线性相关需要找 n 个不同下标对应向量，因此 i_1, \dots, i_n 必须构成集合。

* 由于线性无关的向量组不可能有重复向量（否则系数取为 1 与 -1 即矛盾），线性无关向量组一定是集合。

进一步地，向量 $\alpha \in V$ 能被 $\alpha_i, i \in I$ 线性表出是指能被其中有限个向量表出，也就是

$$\exists i_1, \dots, i_n \in I, \exists \lambda_1, \dots, \lambda_n \in \mathbb{K}, \alpha = \lambda_1\alpha_{i_1} + \dots + \lambda_n\alpha_{i_n}$$

而向量组 $\alpha_i, i \in I$ 称为线性空间 V 的一组基当且仅当它们线性无关，且任何 $\alpha \in V$ 可被它们线性表出——与有限维时很类似，利用线性无关性可以得到表出方式唯一。

* 若向量组 $\beta_i, i \in I$ 中有子集 S 线性无关，且任何 $\beta_i, i \in I$ 可被它们线性表出（利用线性表出的定义，这意味着 S 放入其他任何一个向量后都将线性相关），则称 S 是 $\beta_i, i \in I$ 的一个极大线性无关组。由此，基就是线性空间的极大线性无关组。

在介绍了这么多抽象概念之后，看更多例子之前，我们还是必须说明，即使直觉看来显然的结论，在抽象线性空间中的形式证明也可能并不简单。例如 $0\alpha = 0$ 与 $(-1)\alpha = -\alpha$ 。

证明：

由分配律可知

$$0\alpha = (0 + 0)\alpha = 0\alpha + 0\alpha$$

设 $\beta = -(0\alpha)$ ，两侧同加 β 得到

$$0\alpha = 0$$

再由分配律知

$$0 = 0\alpha = (1 + (-1))\alpha = \alpha + (-1)\alpha$$

从而由存在唯一性可知 $(-1)\alpha = -\alpha$ 。

大家或许会自然诞生一个疑问：既然我们往往是知道它是线性空间之前就已经学会了加法和数乘的计算方式，为何还要进行这些复杂的证明来得到“显然”的性质呢？但是，正是因为这些证明，我们才知道该如何抽象出线性结构。例如， $0\alpha = 0$ 是一个我们会觉得“显然对数乘成立”的性质，但是它不够本质（容易通过其他性质证明），因此无需放入线性空间的定义中。这样精简的八条定义能刻画出我们认为显然成立的性质，恰恰是这个定义十分优秀的体现。不过事实上，这八条定义中的加法交换律是可以通过其他七条推出的，放在定义中是因为交换的确刻画了加法的本质，强调本质的不可替代性。

证明：

证明的关键在于类似矩阵乘法中 $(AB)^{-1} = B^{-1}A^{-1}$ 的逆变结构。首先，直接利用结合律验证可发现

$$(-\alpha) + (-\beta) + \beta + \alpha = 0$$

从而根据唯一性有

$$(-\alpha) + (-\beta) = -(\beta + \alpha)$$

而由已证可知

$$(-1)\alpha + (-1)\beta = (-1)(\beta + \alpha)$$

再通过分配律可得

$$(-1)(\alpha + \beta) = (-1)(\beta + \alpha)$$

两侧同时数乘 -1 ，由数乘结合律得到结论

$$\alpha + \beta = \beta + \alpha$$

* 结合之前 $(-1)\alpha = -\alpha$ 的证明，我们可以发现，这个证明的确运用了所有剩下七条性质。大家可以自行确定每条性质用在了何处，加强对代数化的抽象证明的理解——之后，我们对线性空间相关定理的证明本质上都是从这八条出发的，无法再直接通过具体例子确定。

对于抽象的概念，最重要的学习方式就是通过具体例子进行把握。现在，我们来观察一些具体的线性空间例子（大家可以自行验证为线性空间，一般来说，当加法、数乘的性质符合通常感受时，较重要的是验证封闭性）：

- 定义 $V_1 = \mathbb{C}$, 则加法为通常的加法、数乘为通常的乘法时, V_1 是 \mathbb{R} 上的线性空间。

由于不存在不全为 0 的 $a, b \in \mathbb{R}$ 使得 $a + bi = 0$, 可知 1 和 i 是线性无关的, 而任何复数又都可以写为 $a + bi$, 于是 1、 i 构成 V_1 的一组基。

- 定义 $V_2 = \mathbb{R}$, 则加法为通常的加法、数乘为通常的乘法时, V_2 是 \mathbb{Q} 上的线性空间。

一个著名的结论是, e 是超越数, 也就是它不是任何非零整系数多项式的根。由此, 考虑向量组 $\{e^n \mid n \in \mathbb{N}\}$, 若存在不全为 0 的 $q_1, \dots, q_r \in \mathbb{Q}$ 使得

$$\sum_{k=1}^r q_k e^{m_k} = 0$$

将所有 m_k 从小到大排列并补全中间项可知存在不全为 0 的有理数 a_0, \dots, a_m 使得

$$a_0 + a_1 e + \cdots + a_m e^m = 0$$

通分可知存在非零整数 M 使得 Ma_i 均为整数, 由此有

$$(Ma_0) + (Ma_1)e + \cdots + (Ma_m)e^m = 0$$

由于此多项式非零, 与 e 是超越数矛盾。因此, 我们的确构造出了无穷多个线性无关的向量。

* 一般地, 若 \mathbb{K} 是一个数域, \mathbb{F} 是包含 \mathbb{K} 的数域, 在通常加法、乘法看作加法、数乘下, \mathbb{F} 一定是 \mathbb{K} 上的线性空间。

- 定义 $V_3 = \mathbb{R}[x]$, 即实系数多项式集合, 则加法为多项式加法、数乘为通常的多项式乘法(数看作零次多项式)时, V_3 是 \mathbb{R} 上的线性空间。

可以发现, $\{x^n \mid n \in \mathbb{N}\}$ 是线性无关的(否则能得到一个非零多项式与 0 相等), 且任何多项式都可以被它们的线性组合表出(这就是多项式的定义), 因此这无穷多个向量构成 V_3 的一组基。

- 定义 $V_4 = \mathbb{R}[x]_n$, 即小于 n 次的实系数多项式集合, 则加法为多项式加法、数乘为通常的多项式乘法(数看作零次多项式)时, V_4 是 \mathbb{R} 上的线性空间, 类似上个例子可以得到 $1, x, \dots, x^{n-1}$ 构成 V_4 的一组基。

- 定义 $V_5 = \mathbb{K}^{m \times n}$, 即 \mathbb{K} 上所有 m 行 n 列矩阵构成的集合, 则加法为矩阵加法、数乘为矩阵数乘时, V_5 是 \mathbb{K} 上的线性空间。

记 E_{ij} 为只有第 i 行第 j 列是 1, 其他是 0 的矩阵, 则 $\{E_{ij} \mid i = 1, \dots, m, j = 1, \dots, n\}$ 构成 V_5 的一组基: 它们的线性无关性可以直接通过定义得到, 而对任何第 i 行第 j 列为 a_{ij} 的矩阵 A , 它可以写成

$$A = \sum_{i=1}^m \sum_{j=1}^n a_{ij} E_{ij}$$

由此其符合基的定义。

下面我们尝试用线性空间的语言理解期末考试的第三题。定义 $V = \mathbb{C}^{\mathbb{N}}$, 也即所有下标为 0 的复数数列构成的集合, 我们将每个元素为 a_i 的数列即为 (a_k) 。其中的加法为对应元素相加 $((a_k) + (b_k))$ 的第 i 个元素为 $a_i + b_i$, 数乘为每个元素进行数乘 $(\lambda(a_k))$ 的第 i 个元素为 λa_i , 则可以验证其构成一个线性空间。

首先我们必须指出, 这个线性空间与多项式空间 $\mathbb{C}[x]$ 有本质不同: 记 e_i 为第 i 个分量是 1, 其余为 0 的数列, 则 $\{e_i \mid i \in \mathbb{N}\}$ 并不是这个空间的一组基: 所有元素全为 1 的数列无法被它们中有限个的线性组合表出(注意线性表出的定义要求)。

我们可以证明, 所有满足递推 $a_{n+2} = \frac{1}{2}(a_{n+1} + a_n)$ 的复数列也构成一个线性空间, 记为 U 。

证明:

由于其中的加法、数乘定义与 $\mathbb{C}^{\mathbb{N}}$ 中相同，无需再验证，只要证明封闭性。

若 $(a_k), (b_k)$ 满足递推

$$a_{n+2} = \frac{1}{2}(a_{n+1} + a_n), \quad b_{n+2} = \frac{1}{2}(b_{n+1} + b_n)$$

则计算可知其和 $(a_k + b_k)$ 也满足递推

$$a_{n+2} + b_{n+2} = \frac{1}{2}((a_{n+1} + b_{n+1}) + (a_n + b_n))$$

数乘 (λa_k) 也满足递推

$$\lambda a_{n+2} = \frac{1}{2}(\lambda a_{n+1} + \lambda a_n)$$

从而其为线性空间。

非常自然地，这样的线性空间 U 是 V 的子集，且与 V 具有相同的运算，因此称为 V 的子空间。

* 如同过程中那样，验证一个集合是给定线性空间的子空间无需再考虑加法、数乘的性质，只要验证封闭。

* 作为反例，虽然之前例子中 V_2 是 V_1 的子集，但并不构成子空间，因为数乘的定义不同，一个是针对有理数，一个是针对实数。

自然，我们希望能知道这个线性空间的一组基。可以证明，以 $a_0 = 0, a_1 = 1$ 开始的数列 (a_k) 与 $b_0 = 1, b_1 = 0$ 开始的数列 (b_k) 构成 U 的一组基。

证明：

先说明线性无关。若 $\lambda(a_k) + \mu(b_k) = 0$ ，考虑前两个分量可发现 $\lambda = \mu = 0$ 。

对任何满足 $c_0 = x, c_1 = y$ 的 U 中数列 (c_k) ，由于 $x(a_k) + y(b_k)$ 的前两个位置与 (c_k) 相同，且 U 中数列可被前两个位置决定，即得到 $x(a_k) + y(b_k) = (c_k)$ ，从而得证。

不过，从目前来看，似乎并没有办法确定这两个数列的具体表达式。这是因为这两组基并不具有很好的性质。为了得到更好的性质，我们将在下一部分引入线性变换的概念进一步刻画。

17.4.2 整体性

仿照我们对向量空间之间线性映射的定义，线性空间之间的线性映射定义也非常自然：对 \mathbb{K} 上的线性空间 U, V ，与 $U \rightarrow V$ 的映射 f ，若有

$$\forall \alpha, \beta \in U, \quad f(\alpha + \beta) = f(\alpha) + f(\beta)$$

$$\forall \alpha \in U, \lambda \in \mathbb{K}, \quad f(\lambda \alpha) = \lambda f(\alpha)$$

则称它为一个 U 到 V 的线性映射。当 $V = U$ 时，我们把线性映射称为 U 上的线性变换。

仍然沿用上一节的 V_1 到 V_5 的定义，我们给出一些线性映射与线性变换的例子，大家同样自行验证即可（这里写 \mathbb{K} 或 \mathbb{K}^n 时默认是 \mathbb{K} 上的向量空间）：

- $f(z) = \bar{z}$ 是一个 V_1 上的线性变换。
- \mathbb{C} 作为 \mathbb{C} 上的一维向量空间时， $f(z) = \bar{z}$ 并不是线性变换： $f(1) = 1$ ，但 $f(i1) = -i \neq i f(1)$ 。
* 思考这里与上一种情况的区别。
- $f(x) = \sqrt{2}x$ 是一个 V_2 上的线性变换。
- 对任何多项式 $f \in \mathbb{R}[x]$ ，定义 $L(f) = f(1)$ ，它是一个 V_3 到 \mathbb{R} 的线性映射。
- 对任何多项式 $f \in \mathbb{R}[x]$ ，定义

$$L(f) = \int_0^1 f(x) dx$$

它也是一个 V_3 到 \mathbb{R} 的线性映射。

- 对任何多项式 $f \in V_4$, 定义 $A(f) = f$, 它是一个 V_4 到 V_3 的线性映射。
- 对任何方阵 $A \in \mathbb{K}^{n \times n}$, 迹 $\text{tr}(A)$ 是一个 $\mathbb{K}^{n \times n} \rightarrow \mathbb{K}$ 的线性映射。
* 而 \det 、 rank 都不是线性映射。
- 给定列向量 $\beta_1, \dots, \beta_{n-1} \in \mathbb{K}^n$, 则映射

$$f(\alpha) = \det(\alpha, \beta_1, \dots, \beta_{n-1})$$

是一个 $\mathbb{K}^n \rightarrow \mathbb{K}$ 的线性映射。

当然, 向量空间的线性映射在我们的定义下还是线性映射。例如, 对 $A \in \mathbb{K}^{m \times n}$, $x \rightarrow Ax$ 是 $\mathbb{K}^n \rightarrow \mathbb{K}^m$ 的线性映射, 而 $m = n$ 时它可以看作 \mathbb{K}^n 上的线性变换。

为了能进行进一步的讨论, 我们必须思考一个问题: 当我们在谈论矩阵时, 我们在谈论什么?

最开始学习矩阵时, 它只不过是把一些数排成了行列。但现在, 当我们看到 $A^2 - 5A + 6I = O$ 这样的式子时, 我们想到的并不是各个分量展开形成的 n^2 个二次方程, 而是矩阵作为整体所展现出的性质。当我们逐渐不用去拆成分量看待矩阵时, 它就具有了更多的整体刻画。

期末试卷第五题中, 我们可以看到矩阵作为分量时与整体时考虑的对角化问题, 而为了把它在一般的线性变换(线性变换对应方阵)中推广, 我们则必须思考, 哪些性质是可以不依赖分量进行定义的。

一个例子是, 特征系统是可以不依赖分量定义: 设 \mathcal{A} 是一个 $U \rightarrow U$ 的线性变换, 若有非零向量 $x \in U$ 使得

$$\mathcal{A}(x) = \lambda x$$

则称 λ 为 \mathcal{A} 的一个特征值, x 为对应的特征向量。对某个 λ , 可直接验证其所有特征向量与 0 构成一个线性空间, 称为特征子空间。一个重要的性质是, \mathcal{A} 不同特征值的特征向量线性无关仍然成立。接下来, 不引起歧义时, 我们将把 $\mathcal{A}(x)$ 简记为 $\mathcal{A}x$ 。

证明:

若否, 存在 \mathcal{A} 的不同特征值 $\lambda_1, \dots, \lambda_n$ 与对应特征向量 x_1, \dots, x_n 使得存在不全为 0 的 μ_1, \dots, μ_n 满足

$$\mu_1 x_1 + \dots + \mu_n x_n = 0$$

在左侧作用 \mathcal{A} , 利用线性映射定义可知

$$\mu_1 \mathcal{A}x_1 + \dots + \mu_n \mathcal{A}x_n = \mathcal{A}0$$

由 $\mathcal{A}0 + \mathcal{A}0 = \mathcal{A}0$ 可知 $\mathcal{A}0 = 0$, 从而右侧仍为 0。而左侧根据特征值定义可知

$$\mu_1 \lambda_1 x_1 + \dots + \mu_n \lambda_n x_n = 0$$

重复此过程可得到

$$\mu_1 \lambda_1^2 x_1 + \dots + \mu_n \lambda_n^2 x_n = 0$$

直到

$$\mu_1 \lambda_1^{n-1} x_1 + \dots + \mu_n \lambda_n^{n-1} x_n = 0$$

设矩阵

$$P = \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \dots & \lambda_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{n-1} \end{pmatrix}$$

由于 λ_i 互不相同, Vandermonde 阵行列式非零, 从而 $P\alpha = b$ 对任何向量 b 有非零解。取 $b = e_1$ (第一个元素为 1, 其他为 0), 设解为 a_1, \dots, a_n , 对比 x_i 前的系数可发现

$$0 = \sum_{i=1}^n a_i (\mu_1 \lambda_1^{i-1} x_1 + \dots + \mu_n \lambda_n^{i-1} x_n) = \mu_1 x_1$$

由 $x_1 \neq 0$ 可知 $\mu_1 = 0$, 同理可得所有 μ_i 均为 0, 与不全为 0 矛盾。

由于更多对于线性变换的知识在之后才会涉及, 我们这里还是回到期末考试第三题的例子。接下来所有的分析基于如下的构造: 设 V 为上一部分中定义的所有复数列构成的线性空间, 则

$$\mathcal{B}(a_0, a_1, a_2, a_3, \dots) = (a_1, a_2, a_3, \dots)$$

是一个线性变换, 可以称为前移, 也即将数列的每个元素向前移动一位 (a_0 直接删除)。

这个线性变换的定义平平无奇, 但我们如果进行特征系统的分析, 就可以发现非常有趣的事。我们先来求解其全部特征值与特征向量。

解答:

设

$$\mathcal{B}(a_k) = \lambda(a_k)$$

展开可发现有

$$(a_1, a_2, a_3, \dots) = (\lambda a_0, \lambda a_1, \lambda a_2, \dots)$$

从而有

$$a_i = \lambda a_{i-1}, \quad \forall i \in \mathbb{N}^+$$

因此, \mathcal{B} 的特征值为全体复数, 特征值 c 对应的特征向量为以 c 为公比的等比数列。类似上一部分的做法可以验证, c 对应的特征子空间的基可以取为 $\{(c^k)\}$, 即一个第 k 个元素是 c^k 的数列。

* 利用不同特征值特征向量线性无关, 我们事实上构造了 V 中不可数个线性无关的向量。且它们仍然不构成 V 的一组基: 考虑 $a_n = e^{n^2}$, 利用无穷大阶数超过任何等比数列可发现其不可被这些向量线性表示。

下面, 我们用 \mathcal{B} 进行原问题的求解。若将数列记为 x , 数列满足

$$a_{n+2} = \frac{1}{2}(a_{n+1} + a_n)$$

事实上可验证其能写为

$$\mathcal{B}^2 x = \frac{1}{2} \mathcal{B}x + \frac{1}{2} x$$

也即上一部分定义的

$$U = \{x \in V \mid \mathcal{B}^2 x = \frac{1}{2} \mathcal{B}x + \frac{1}{2} x\}$$

这里 \mathcal{B}^2 表示 \mathcal{B} 的复合, 这里即重复作用两次, 也就是前移两位。此外, \mathcal{B} 左侧的数乘 $\frac{1}{2}$ 满足

$$\left(\frac{1}{2} \mathcal{B}\right)x = \frac{1}{2}(\mathcal{B}x)$$

记 \mathcal{I} 为所有数列映射到自身的单位映射, 可将上式进一步写成

$$\frac{1}{2}(2\mathcal{B}^2 - \mathcal{B} - \mathcal{I})x = 0$$

这里线性映射的加减满足

$$(2\mathcal{B}^2 - \mathcal{B})x = 2\mathcal{B}^2 x - \mathcal{B}x$$

可以发现, x 左侧已经成为了 \mathcal{B} 的多项式——由于乘法、数乘与求和都可以定义, 一个线性变换的多项式是完全可以定义的。这样完全类似矩阵乘法的形式让我们希望能做如下的分解

$$\left(\mathcal{B} + \frac{1}{2}\mathcal{I}\right)(\mathcal{B} - \mathcal{I})x = 0$$

而根据乘法看作映射复合的定义, 可以验证这确实是正确的。不仅如此, 我们还可以验证一个线性变换的多项式可交换, 从而改写成

$$(\mathcal{B} - \mathcal{I})(\mathcal{B} + \frac{1}{2}\mathcal{I})x = 0$$

由于线性变换一定将 0 映射到 0, 只要 $(\mathcal{B} - \mathcal{I})x = 0$ 或 $(\mathcal{B} + \frac{1}{2}\mathcal{I})x = 0$, 上式当然满足, 而这两个式子分别可以化为

$$\mathcal{B}x = x, \quad \mathcal{B}x = -\frac{1}{2}x$$

根据已经求解的特征方程, 前者的一个解为全为 1 的数列 (p_k) , 后者的一个解为满足 $q_k = (-\frac{1}{2})^k$ 的 (q_k) 。

由于前两个元素可以完全确定满足递推的数列, 与上一部分相同可验证 (p_k) 与 (q_k) 构成 U 的一组基, 因此可设所求数列为

$$\lambda(p_k) + \mu(q_k)$$

由前两个元素分别为 0、 $\frac{1}{2}$ 可以最终解出

$$\lambda = \frac{1}{3}, \quad \mu = -\frac{1}{3}, \quad a_k = \frac{1}{3} \left(1 - \frac{(-1)^k}{2^k}\right)$$

* 我们事实上证明了 $\text{Ker}(2\mathcal{B}^2 - \mathcal{B} - \mathcal{I}) = \text{Ker}(2\mathcal{B} + \mathcal{I}) \oplus \text{Ker}(\mathcal{B} - \mathcal{I})$, 在之后学到线性变换时, 我们将分析这件事是否一定成立。

17.4.3 最小二乘结构

最后, 我们再来看一看实方阵内积的推广。期末考试第四题中, 我们研究了如何通过正交投影构造空间里与给定向量最接近的向量。

很显然, 只要一个空间有距离的概念, 我们就可以研究这样的问题。考虑 $[0, 1]$ 上所有的实值连续函数构成的集合, 记为 $C[0, 1]$ 。以函数加法、数和函数的乘法作为加法和数乘, 可以验证这确实构成一个线性空间 (需要利用一点分析知识: 连续函数进行加法、数乘后还是连续函数)。在很多不同情况下, 我们都需要考虑用简单函数逼近复杂函数的问题, 因此在这个空间研究逼近是必要的。

那么, 如何度量这个空间里两个函数的距离呢? 就像欧氏空间那样, 我们希望通过内积解决这个问题。只要能定义合理的内积 $\langle f, g \rangle$, $\sqrt{\langle f - g, f - g \rangle}$ 就是距离的度量。

我们希望一个内积 (看作 $C[0, 1] \times C[0, 1]$ 映射到一个实数的函数) 具有如下三个性质, 下方 f, g 都代表函数, λ 表实数:

- 正定性, 也即 $\langle f, f \rangle \geq 0$, 当且仅当 $f = 0$ 时其为 0。很显然, 只有了这样的性质, $\sqrt{\langle f - g, f - g \rangle}$ 才是有意义的, 也真的能刻画距离 (只有两个函数一样时为 0)。
- 对称性, 也即 $\langle f, g \rangle = \langle g, f \rangle$ 。这是源于内积的夹角刻画性质, 即

$$\cos \theta = \frac{\langle a, b \rangle}{\sqrt{\langle a, a \rangle \langle b, b \rangle}}$$

我们不希望交换 a, b 位置后“夹角”的值会发生变化。

- 双线性性, 也即

$$\langle f_1 + f_2, g \rangle = \langle f_1, g \rangle + \langle f_2, g \rangle$$

$$\langle \lambda f_1, g \rangle = \lambda \langle f_1, g \rangle$$

作为线性空间上的映射，我们希望它对两个分量都是线性映射。这里保证了对第一个分量线性，再结合对称性即可得到对两个分量都线性。

最常用的内积是所谓的 L^2 内积，也就是

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx$$

除了正定性的验证需要一定的分析知识外，其他两个性质都是容易验证的。

回到开始的问题。一个简单的例子是，怎样的一次函数可以在 $C[0, 1]$ 中最好逼近 e^x ？

由定义，两个函数的距离是 $\sqrt{\langle f - g, f - g \rangle}$ ，我们就需要找到 a, b 使得

$$\int_0^1 (e^x - ax - b)^2 dx$$

最小（省略了根号，不影响最小性）。

验证可发现，所有的一次函数构成 $C[0, 1]$ 的一个子空间， $1, x$ 构成一组基，因此这个问题实际上也是子空间 U 中寻找对 V 中某元素 f 的最优逼近，符合我们熟悉的最小二乘问题的形式。

我们希望，与最小二乘问题一样，此问题的解 $f_h \in U$ 有性质

$$\forall g \in U, \quad \langle f - f_h, g \rangle = 0$$

这样，考虑一组基就可以把此问题改写为

$$\int_0^1 (e^x - ax - b)x dx = 0, \quad \int_0^1 (e^x - ax - b) dx = 0$$

直接计算积分可发现这是一个关于 a, b 的线性方程组，就易于求解了，事实上可以解出

$$\begin{cases} a = -6e + 18 \\ b = 4e - 10 \end{cases}$$

对这个问题，上述结论确实是成立的，直接在要最小化的式子中对 a, b 求导就可以得到。那么，一般情况是否成立呢？答案是肯定的。

证明：

若 f_h 使得 $\langle f - f_h, f - f_h \rangle$ 最小，假设还有 $g \in U$ 使得 $\langle f - f_h, g \rangle \neq 0$ ，则考虑 $g_h = f_h + \lambda g$ ，利用双线性性与对称性展开可得

$$\langle f - f_h - \lambda g, f - f_h - \lambda g \rangle = \langle f - f_h, f - f_h \rangle - 2\langle f - f_h, \lambda g \rangle + \langle \lambda g, \lambda g \rangle$$

进一步得到

$$\langle f - f_h - \lambda g, f - f_h - \lambda g \rangle - \langle f - f_h, f - f_h \rangle = -2\lambda \langle f - f_h, g \rangle + \lambda^2 \langle g, g \rangle$$

由于 $\langle f - f_h, g \rangle \neq 0$ 且 $\langle g, g \rangle > 0$ ($g = 0$ 时 $\langle f - f_h, g \rangle = 0$)，利用二次函数最小值理论可发现存在 λ 使右侧小于 0，从而

$$\langle f - f_h - \lambda g, f - f_h - \lambda g \rangle < \langle f - f_h, f - f_h \rangle$$

于是 g_h 的逼近误差更小，这就与最小性矛盾了。

反之，若 f_h 使得对任何 $g \in U$ 有 $\langle f - f_h, g \rangle = 0$ ，类似上方展开可发现对任何 $g \in U$ 有

$$\langle f - g, f - g \rangle = \langle f - f_h, f - f_h \rangle + 2\langle f - f_h, f_h - g \rangle + \langle f_h - g, f_h - g \rangle$$

由子空间封闭性 $f_h - g \in U$ ，从而右侧的中间一项为 0，根据正定性即得

$$\langle f - g, f - g \rangle = \langle f - f_h, f - f_h \rangle + \langle f_h - g, f_h - g \rangle \geq \langle f - f_h, f - f_h \rangle$$

从而 f_h 是一个最优的逼近。

事实上，我们还可以证明最优逼近若存在则唯一。

证明：

若 f_1, f_2 都是 f 在 U 中的最优逼近，由 $f_1 - f_2 \in U$ 与上方已证可知

$$\langle f - f_1, f_1 - f_2 \rangle = 0, \quad \langle f - f_2, f_1 - f_2 \rangle = 0$$

两式作差即得到

$$\langle f_1 - f_2, f_1 - f_2 \rangle = 0$$

因此 $f_1 = f_2$ 。

* 利用更深的理论可以证明，当 U 维数有限时，最优逼近一定存在唯一，但维数无限时可能并不存在。

* 一个与内积密切相关的概念是伴随，也就是，对给定的线性变换 \mathcal{A} ，满足 $\langle \mathcal{A}\alpha, \beta \rangle = \langle \alpha, \mathcal{B}\beta \rangle$ 的线性变换称为 \mathcal{A} 的伴随变换，这事实上是转置真正的来源。

* 和伴随密切相关的概念是对偶，对偶空间理论在多重线性代数中将发挥根本性的作用。

本节的内容是接下来要学习的知识在一般线性空间中版本的一个速览。可以发现，每一块内容事实上都是从上学期的矩阵理论中抽象出来，而正如本讲义 17.2.3 所说，我们将研究的有限维部分也可以完全由矩阵所表示——如何表示本身也是一块重要的学习内容。

在之后所有线性空间相关的学习中，希望大家能记住，我们本质还是在研究矩阵的理论，因此一切问题都可以转化成矩阵论来解决。习题课讲义中，也将补充更多空间和矩阵相关结论的相互表示（事实上，本讲义第四章等内容已经进行了一些基础的讨论了，感兴趣的同學可以参考）。

十八 补充：直和与空间分解

由于线性空间的很多基本的定义、性质与向量空间基本相同，解释定义只需要很简单的叙述，与向量空间相同的定理也无需再赘述证明。因此，本章主要以习题的形式给大家展现如何从抽象角度应用结论。在进行本章的学习前，请大家熟悉本讲义 17.4.1 有关线性空间、线性相关、线性无关、线性表出（注意这三个概念对无穷多个向量的定义，此后线性表出简称表出）、基、子空间的定义。

§18.1 期中前整体规划

从这次习题课开始，我们就要从空间定义一路讲到空间视角的 Jordan 与有理标准形了。由于连这次一共还有五次习题课，除去最后一次讲解复习题，我们还有四次课的机会进行这一套推导。实际上，推导总共分为三个重要的部分，我们仍然将以本讲义 17.2.3 讨论的表示思路进行叙述。

- **线性空间：**从什么是一个线性空间出发，解释抽象线性空间的定义，并给出从一个空间中构造其他空间的方法：子空间、生成空间、交空间、和空间（直和空间）、商空间。这些方法中，除了商空间是重新定义了运算，其他都是保持子空间性质，继承原有的运算。

在有限维情况，我们需要知道上述空间的维数关系，此时列向量可以用于表示一般有限维线性空间中向量的坐标，而可逆矩阵可以表示同一个有限维空间两组基的变换关系，与对应的坐标变换关系。

- **线性映射：**从如何建立两个线性空间的关系出发，解释抽象线性映射的定义，并对应定义线性单射、线性满射、线性同构，以刻画两个线性空间结构的关联。对于同构最重要的结论是第一同构定理，叙述了如何从同态诱导出一个同构——而为了介绍它又需要自然引入核空间与像空间的概念。此外，我们还将介绍之后用于线性映射与线性变换研究的最重要工具，限制映射，它可以通过控制问题的规模以局部简化问题。

在有限维情况，我们可以得到第一同构定理与它能推出的第二、第三同构定理的维数版本。给定一组基后，两个有限维空间之间的线性映射可以用矩阵表示，而线性单射、线性满射、线性同构则对应

列/行满秩与可逆。进行基变换时，线性映射的矩阵表示对应关系是相抵，于是相抵标准形可以看作给线性映射找一组最好的基。最后，限制映射本质上对应某种分块三角阵，因此用限制映射操作事实上与矩阵分块具有相同原理。

最后，同样两个空间之间的线性映射可以进行线性组合，而一个的陪域等于另一个的定义域的两个线性映射可以进行复合，这恰好对应矩阵的加法、数乘与乘法。

- **线性变换：**从一个空间到自身的线性映射出发，考察抽象线性变换能进行怎样的描述。由于线性变换可以与自身复合，我们可以谈论它的多项式，而所有一次多项式的核空间则对应了特征系统，也即特征值与特征向量（准确来说是相应的特征子空间）。

在有限维情况，线性变换对应的矩阵表示是方阵，而进行基变换时线性变换的矩阵表示对应关系是相似，于是相似标准形可以看作给线性变换找一组最好的基。

为了进行构造，我们需要进行两层分解，一层是根子空间分解，另一层是循环子空间分解，而这两层分解都需要基于根据限制映射定义出的不变子空间概念。利用这些工具，我们最终可以得到复方阵相似的 **Jordan 标准形** 与一般数域上方阵相似的有理标准形。

总的来说，对前半学期内容的一个理解是，我们对线性空间、线性映射与线性变换的讨论是为了推广矩阵理论，而在有限维的讨论则是通过语言转换以解决矩阵相似标准形问题。

§18.2 维数与坐标

18.2.1 基本结论

首先，仿照向量空间，我们可以谈论一个一般线性空间的维数。由于线性代数并不强制大家对集合的势（可数、不可数等）有概念，我们只叙述维数有限情况下的定义。

若一个线性空间 V 存在一组个数有限的基 $\alpha_1, \dots, \alpha_n$ ，则称它为一个有限维线性空间，它的维数为 n ，记作 $\dim V = n$ （若 V 只包含零向量，我们认为空集构成它的一组基，维数为 0，仍然算作有限维线性空间）。若不存在，则称为无穷维线性空间。很显然，想要此定义合理，我们需要证明任何一组基个数相同。

证明：

与向量空间完全相同，利用未知数个数比方程个数多的齐次线性方程组一定有非零解可以证明，若向量组 $\beta_i, i \in I$ 能被向量组 $\gamma_1, \dots, \gamma_n$ 表出，则 $\beta_i, i \in I$ 中不可能存在 $n+1$ 个线性无关的向量。

由此，由于 V 能被 $\alpha_1, \dots, \alpha_n$ 表出， V 中不可能存在 $n+1$ 个线性无关的向量，因此另一组基个数至多为 n ，设它为 $\kappa_1, \dots, \kappa_m$ ，有 $m \leq n$ 。由于 $\alpha_1, \dots, \alpha_n$ 能被 $\kappa_1, \dots, \kappa_m$ 表出，再次根据上方结论可知 $n \leq m$ ，综合 $m \leq n$ 得到 $m = n$ 。

* 一般的线性空间是否存在基与维数？如果认为选择公理成立，那么基一定存在，且仍然有“任何一组基个数相同”的结论，从而可以定义维数。不过，由于这些理论需要的数学背景过于抽象，即使真的在泛函分析这样关心无穷维线性空间的课程中，我们也很少会谈论无穷维空间的维数。本门课程中，我们默认线性空间一定存在一组基。

为了展示无穷维空间的确比有限维“更大”，我们有如下结论：对 \mathbb{K} 上的无穷维线性空间 V ，任意给定 $n \in \mathbb{N}$ ， V 存在一个 n 维的子空间。

证明：

当 $n = 0$ 时，取 $\{0\}$ 作为子空间即成立。下面假设存在 k 维子空间 U ，证明存在 $k+1$ 维子空间。取出 U 的一组基 $\alpha_1, \dots, \alpha_k$ ，由定义它们线性无关，若它们能表出 V 的所有向量，则根据定义它们也是 V 的一组基，与 V 不存在有限基矛盾，因此一定有向量 β 不能被它们表出，进一步与向量空间时完全相同可验证 $\alpha_1, \dots, \alpha_k, \beta$ 线性无关。

考虑集合

$$\{u + b\beta \mid b \in \mathbb{K}, u \in U\}$$

利用定义与子空间 U 的封闭性可验证此集合构成 V 的子空间，且其中元素由 $\alpha_1, \dots, \alpha_k, \beta$ 表出 (U 可由 $\alpha_1, \dots, \alpha_k$ 生成)，由此可知 $\alpha_1, \dots, \alpha_k, \beta$ 构成此空间一组基，这就是符合要求的子空间。

在刚才的讨论中，我们可以发现，直观来说，只要有一些向量，我们就可以用它们生成一个子空间，使得这个子空间的元素都可以被这些向量表出。基于这个感受，我们给出生成线性空间的两个等价定义。设 V 是一个 \mathbb{K} 上的线性空间， $\alpha_i, i \in I$ 是 V 中的向量组，我们称它们生成的线性空间为能被此向量组线性表出的所有向量集合，或所有包含此向量组的 V 的子空间中最小的一个，并记为

$$\langle \alpha_i, i \in I \rangle$$

* 这里包含向量组意味着每个元素都被包含，不考虑重复。

* 回顾本讲义 14.1.2，最小是指任何包含此向量组的 V 的子空间都包含 $\langle \alpha_i, i \in I \rangle$ 。

我们下面来证明第一个定义所得到的集合确实是一个 V 的子空间，且符合第二个定义。这样，由于第二个定义对应的子空间至多唯一（互相包含的集合是相等的），就能得到第二个定义得到的子空间也符合第一个定义，于是两个定义等价。

证明：

我们将第一个定义形式上写为 ($n = 0$ 时认为对应的元素为 0)

$$\{\lambda_1\alpha_{i_1} + \dots + \lambda_n\alpha_{i_n} \mid n \in \mathbb{N}, \quad i_1, \dots, i_n \in I, \quad \lambda_1, \dots, \lambda_n \in \mathbb{K}\}$$

先验证这的确是一个线性空间。假设两个元素分别写为 (λ_i, μ_i 是 \mathbb{K} 中元素， α_i 都是 V 中向量)

$$\lambda_1\alpha_{i_1} + \dots + \lambda_n\alpha_{i_n}, \quad \mu_1\alpha_{j_1} + \dots + \mu_m\alpha_{j_m}$$

它们的和即为

$$\lambda_1\alpha_{i_1} + \dots + \lambda_n\alpha_{i_n} + \mu_1\alpha_{j_1} + \dots + \mu_m\alpha_{j_m}$$

由于这仍然符合生成空间中元素的定义 ($n+m$ 个仍然为有限个元素)，因此仍然在生成空间中。

另一方面， $\lambda_1\alpha_{i_1} + \dots + \lambda_n\alpha_{i_n}$ 数乘 λ 的结果为

$$(\lambda\lambda_1)\alpha_{i_1} + \dots + (\lambda\lambda_n)\alpha_{i_n}$$

这也符合生成空间中元素的定义，由此得到生成空间确实是一个子空间。

接下来验证任何包含 $\alpha_i, i \in I$ 的线性空间都包含这个空间。利用对数乘的封闭性，对任何 $\alpha_{i_1}, \dots, \alpha_{i_n}$ 与 $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ ，包含 $\alpha_i, i \in I$ 的子空间都应包含

$$\lambda_1\alpha_{i_1}, \quad \dots, \quad \lambda_n\alpha_{i_n}$$

再利用对加法的封闭性，即可得到它包含

$$\lambda_1\alpha_{i_1} + \dots + \lambda_n\alpha_{i_n}$$

这就得到了包含关系。

* 当向量组元素个数有限时，生成空间可以写为

$$\langle \alpha_1, \dots, \alpha_n \rangle = \{\lambda_1\alpha_1 + \dots + \lambda_n\alpha_n \mid \lambda_1, \dots, \lambda_n \in \mathbb{K}\}$$

这是因为若选取的 α_{i_k} 少于 n 个，可以视为其他 λ_i 取 0。

* 利用定义可发现，原向量组的极大线性无关组构成生成空间的一组基。

接下来就该谈论坐标了。正如 \mathbb{R}^n 中一样，坐标是一种用数组表示向量的手段。必须强调，我们只在有限维线性空间中谈论坐标，不考虑无穷维的情况。给定 \mathbb{K} 上 n 维线性空间 V 的一组基 $S = \{\alpha_1, \dots, \alpha_n\}$ ，坐标是指如下 V 到 \mathbb{K}^n 的映射 π_S ：

$$\pi_S(\lambda_1\alpha_1 + \dots + \lambda_n\alpha_n) = (\lambda_1, \dots, \lambda_n)^T$$

根据基的定义，任何 V 中向量都可以写为 $\lambda_1\alpha_1 + \dots + \lambda_n\alpha_n$ ，而利用线性无关性可知此表示是唯一的（证明与向量空间中相同），因此 π_S 的确是 V 上的映射。事实上，它还是一个双射。

证明：

先证明满射。对任何 $(\lambda_1, \dots, \lambda_n)^T \in \mathbb{K}^n$ ，利用线性空间的封闭性可知 $\lambda_1\alpha_1 + \dots + \lambda_n\alpha_n$ 一定是 V 中元素，这就找到了原像。

再证明单射。若 $\pi_S(\beta) = \pi_S(\gamma) = (\lambda_1, \dots, \lambda_n)^T$ ，可知

$$\beta = \lambda_1\alpha_1 + \dots + \lambda_n\alpha_n = \gamma$$

从而即得单射。

因此，坐标是向量到数组的一一对应。我们将 $\pi_S(\beta)$ 记为 β_S （注意这是 \mathbb{K}^n 中元素），应有形式乘法

$$\beta = S\beta_S$$

这里 $S = (\alpha_1, \dots, \alpha_n)$ 是将基中每个向量作为元素拼成的行向量， $\beta_S = (\mu_1, \dots, \mu_n)^T$ 为列向量，根据矩阵乘法的形式定义即有右侧为

$$\mu_1\alpha_1 + \dots + \mu_n\alpha_n = \beta$$

我们必须补充一些说明：

- 之所以叫形式乘法而不是一般的矩阵乘法，是因为这里的 S 未必是矩阵，例如考虑 $\mathbb{R}[x]_3$ （三次以下实系数多项式）时， 1 、 x 、 x^2 构成一组基，于是 $3x^2 + 2x + 1$ 对应的坐标为 $(1, 2, 3)^T$ ，形式乘法为

$$3x^2 + 2x + 1 = \begin{pmatrix} 1 & x & x^2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

- 当 S 为向量空间的基，也即一些列向量时，上述的形式乘法即成为真正的矩阵乘法，例如考虑 \mathbb{R}^3 中 $(1, -1, 2)^T$ 与 $(1, 1, 1)^T$ 生成的子空间，并以此两向量作为一组基，则计算可发现 $(3, -1, 5)^T$ 对应的坐标为 $(2, 1)^T$ ，形式乘法为

$$\begin{pmatrix} 3 \\ -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

- 从现在起，我们谈论的基都是指 $S = (\alpha_1, \dots, \alpha_n)$ 的形式行向量形式，而不再是 $\{\alpha_1, \dots, \alpha_n\}$ 的集合形式，这是因为我们需要关心顺序。例如，在基 $(\alpha_1, \alpha_2, \alpha_3)$ 下坐标为 $(1, 2, 3)^T$ 的元素，在基 $(\alpha_1, \alpha_3, \alpha_2)$ 下坐标应为 $(1, 3, 2)^T$ 。改变基的顺序会导致坐标改变，而之后定义的基变换矩阵、坐标变换矩阵、线性映射的矩阵表示等也会对应改变。
- 形式乘法最重要的性质是结合律仍然满足，这与矩阵乘法的结合律验证方式完全相同。事实上，它们的本质都是映射复合具有结合律，而形式乘法只是定义了某种满足线性关系的映射而已，我们将在之后介绍矩阵表示时具体谈论。除此以外，其分配律、与数乘可交换等也仍然成立，与矩阵情况验证相同。

就如向量空间时的结论一样，一个线性空间往往可以有无穷多组不同的基。由定义任何一组基都可以表出全空间，因此若给定 \mathbb{K} 上 n 维线性空间 V 的一组基 $S = (\alpha_1, \dots, \alpha_n)$ ，对向量组 $T = (\beta_1, \dots, \beta_n)$ ，一定存在唯一 \mathbb{K} 上 n 阶方阵 C 使得

$$T = SC$$

更进一步地， T 也是 V 的一组基当且仅当 C 可逆。

证明：

将此形式乘法展开也即（设 C 第 i 行第 j 列即 c_{ij} ）

$$\forall i = 1, \dots, n, \quad \beta_i = \sum_{j=1}^n c_{ji} \alpha_j$$

而由于 β_i 一定可以写成 α_1 到 α_n 的线性组合，表示一定存在。另一方面，由基的线性无关性任何向量表示唯一，因此 c_{ji} 唯一确定。

我们下面用形式乘法来证明 T 是一组基当且仅当 C 可逆。考虑 n 维 \mathbb{K} 上列向量 x ，则由于 T 个数等于维数，与向量空间类似可证明它构成一组基当且仅当线性无关，也即 $Tx = 0$ 只有零解。

将其写为 $SCx = 0$ ，由形式乘法结合律即 $S(Cx) = 0$ ，利用 S 线性无关可知这当且仅当 $Cx = 0$ 。由此， T 构成一组基当且仅当 $Cx = 0$ 只有零解，利用 Cramer 法则即得这等价于 C 可逆（注意 C 是 \mathbb{K} 上的方阵， Cx 就是通常意义矩阵乘法）。

当 T 也是 V 的一组基时，我们将 C 称为 S 到 T 的过渡矩阵。如之前所证，它一定是一个可逆矩阵。假设 α 在 S 、 T 下的坐标分别为 α_S 、 α_T ，我们即得到了三个等式

$$T = SC, \quad \alpha = S\alpha_S = T\alpha_T$$

由此将 $T = SC$ 代入可得

$$S\alpha_S = SC\alpha_T$$

利用 S 的线性无关性，此式成立当且仅当

$$\alpha_S = C\alpha_T$$

因此

$$\alpha_T = C^{-1}\alpha_S$$

也即更换基时坐标的变换可看作左乘过渡矩阵的逆。

* 此过程里我们可以看出形式乘法的好处，它避免了复杂的中间计算，而是通过已经证明的矩阵乘法结合律直接得到了等量关系。

* 另一个与上学期相同过程可以得到的结论是，有限维线性空间中个数与维数相等的线性无关向量组一定为基。

18.2.2 例题

以下 $C(\mathbb{R})$ 表示 \mathbb{R} 上所有连续函数构成的 \mathbb{R} 上线性空间，其他记号如本讲义 17.4.1。

1. 证明对 $A \subset \mathbb{R}$ ， $C(\mathbb{R})$ 中的向量集合 (x 为函数自变量)

$$S = \{\sin(ax) \mid a \in A\}$$

线性无关当且仅当不存在 $a, b \in A$ 使得 $a + b = 0$ 。

* 由此可知 $C(\mathbb{R})$ 是无限维的。

2. 假设 $n > 1$ ，以下哪些矩阵集合是 n 阶实方阵空间 $\mathbb{R}^{n \times n}$ 的子空间？若是，给出它的基与维数。

- (a) 实对称方阵;
 - (b) 实斜对称方阵;
 - (c) 对角阵;
 - (d) 上三角方阵;
 - (e) 满足 $A^2 = O$ 的方阵;
 - (f) 满足 $\det A = a$ 的方阵 ($a \in \mathbb{R}$ 给定);
 - (g) 秩小于等于 r 的方阵 ($0 \leq r \leq n$ 给定);
 - (h) 满足 $\text{tr } A = a$ 的方阵 ($a \in \mathbb{R}$ 给定);
 - (i) 所有特征值都是实数的方阵;
 - (j) 可以相似对角化的方阵;
 - (k) 与 A 可交换的方阵 ($A \in \mathbb{R}^{n \times n}$ 给定, 本小问不要求给出基与维数);
 - (l) A 的多项式 ($A \in \mathbb{R}^{n \times n}$ 给定, 可假设其 Jordan 标准形已知)。
3. 已知 V 是 \mathbb{C} 上的线性空间, 并将其上的加法记为 $a + b$ 、数乘记为 $\lambda \cdot a$ 。在集合 V 上定义新的加法 \oplus 满足
- $$\forall a, b \in V, \quad a \oplus b = a + b$$
- \mathbb{R} 上数乘 \odot 满足
- $$\forall \lambda \in \mathbb{R}, a \in V, \quad \lambda \odot a = \lambda \cdot a$$
- 证明此时的 V (记作 $V_{\mathbb{R}}$) 构成 \mathbb{R} 上的线性空间。若 V 是 \mathbb{C} 上的 n 维线性空间, 一组基为 $\alpha_1, \dots, \alpha_n$, 确定 $V_{\mathbb{R}}$ 的维数与一组基。
4. 设 V 是 \mathbb{K} 上的线性空间, 证明当且仅当 V 是有限维线性空间时, U 是 V 的子空间且 U 的一组基到 V 的一组基存在双射可推出 $U = V$ 。
5. 设 V 是 \mathbb{K} 上的线性空间, 且 $\alpha_1, \dots, \alpha_n$ 线性无关, 对给定 $\lambda \in \mathbb{K}$, 求
- $$\alpha_1 + \lambda \alpha_2, \quad \alpha_2 + \lambda \alpha_3, \quad \dots, \quad \alpha_{n-1} + \lambda \alpha_n, \quad \alpha_n + \lambda \alpha_1$$
- 生成的线性空间的维数与一组基。
6. 在 $C(\mathbb{R})$ 中, 证明向量集合
- $$S = \{1, \sin x, \cos x, \sin 2x, \cos 2x, \sin 3x, \cos 3x\}$$
- $$T = \{\sin^2 x, \sin x \cos x, \cos^2 x, \sin^3 x, \sin^2 x \cos x, \sin x \cos^2 x, \cos^3 x\}$$
- 是同一个子空间的一组基, 并给出 S 到 T 的过渡矩阵与 $\frac{\sin 4x}{\sin x}, \frac{\sin 4x}{\cos x}$ 在 S, T 下的坐标 (这里分母为 0 处以极限定义)。
7. 在 $\mathbb{R}[x]_n$ 中, 对 $a \in \mathbb{R}$, 证明 $1, (x - a), \dots, (x - a)^{n-1}$ 为一组基, 并求其到 $1, x, \dots, x^{n-1}$ 的过渡矩阵。

18.2.3 解答

1. 先说明存在 $a, b \in A$ 使得 $a + b = 0$ 时线性相关。若 $a = b = 0$, $\sin 0x = 0 \in S$, 也即 S 中有零向量, 线性相关; 否则存在非零 a 使得 $\sin ax$ 、 $\sin(-a)x$ 都在 S 中, 而 $\sin ax + \sin(-a)x = 0$, 从而线性相关。

若不存在 $a, b \in A$ 使得 $a + b = 0$, 可发现 A 中无 0, 且对任何不同的 $a, b \in A$, $a^2 \neq b^2$, 下面以此出发得到结论。

要说明 S 线性相关, 只要说明任何有限子集线性相关。考虑写成形式行向量的子集 ($a_1, \dots, a_n \in A$ 互不相同)

$$T = (\sin a_1 x, \dots, \sin a_n x)$$

要证其线性无关只需证明 $T\alpha = 0$ 对 $\alpha \in \mathbb{R}^n$ 只有零解。

若 $T\alpha = 0$ 成立, 将此式求导成为

$$(a_1 \cos a_1 x, \dots, a_n \cos a_n x)\alpha = 0$$

再求导得到

$$(-a_1^2 \sin a_1 x, \dots, -a_n^2 \sin a_n x)\alpha = 0$$

上式又可以写成 (两侧同乘 -1 消去负号)

$$T \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} \begin{pmatrix} a_1^2 \\ \vdots \\ a_n^2 \end{pmatrix} = 0$$

类似将 $T\alpha = 0$ 写成

$$T \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 0$$

更进一步地, 对 $(-a_1^2 \sin a_1 x, \dots, -a_n^2 \sin a_n x)\alpha = 0$ 再求两阶导, 可以得到

$$T \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} \begin{pmatrix} a_1^4 \\ \vdots \\ a_n^4 \end{pmatrix} = 0$$

同理, 通过求导, 右侧的 4 次方可以换成任何 $2n$ 次方。我们考虑 0、2、4 直到 $2n-2$ 次方, 并按列拼成矩阵, 可得 (右侧是一个 n 个元素全为 0 的形式行向量)

$$T \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} \begin{pmatrix} 1 & a_1^2 & \dots & a_1^{2n-2} \\ 1 & a_2^2 & \dots & a_2^{2n-2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & a_n^2 & \dots & a_n^{2n-2} \end{pmatrix} = \mathbf{0}$$

由条件 a_1^2 到 a_n^2 互不相同, 因此利用 Vandermonde 行列式知第二个矩阵可逆, 从而同时右乘其逆得到

$$T \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} = \mathbf{0}$$

而这就意味着 $\alpha_1 \sin(a_1 x) = \alpha_2 \sin(a_2 x) = \cdots = \alpha_n \sin(a_n x) = 0$, 由于 a_1 到 a_n 非零即得所有 α_i 为 0, 得证。

* 这样的形式乘法写法可以让此题过程简洁很多, 但需要想清楚过程, 尤其是“拼成矩阵”一步, 本质是用了分块乘法 $AB = A(\beta_1, \dots, \beta_n) = (A\beta_1, \dots, A\beta_n)$, 由于分块乘法相关的结论只利用乘法定义就可以证明, 这一结论对形式矩阵乘法仍然成立。

* 为什么能想到求两阶导? 这里给出一个更本质的解释, 大家可以在学完线性变换后回看。考虑所有可以求任意多次导数的函数构成的集合 V , 可直接验证这是一个线性空间, 而求两阶导 $f \rightarrow f''$ 是其上的线性变换, 将此变换记为 C 。直接计算可以发现, a 非零时所有 $\sin ax$ 都是 C 的特征向量, 对应的特征值为 $-a^2$ 。由此, 若 A 中不存在 $a+b=0$, S 中所有元素是 C 不同特征值的特征向量, 因此线性无关 (证明可见本讲义 17.4.2, 大家可以思考如何用形式乘法重写本讲义 17.4.2 的证明)。

2. 我们以本题为例介绍子空间的验证方法与基、维数的求法。首先, 如本讲义 17.4.1 所说, 由于子空间的运算遵循原空间的运算, 八条性质自动满足, 验证子空间只需考虑封闭性, 也即加法、数乘仍在其中 (零元和逆元的存在性可以通过封闭性推得)。另外, 线性空间中一定有零元素, 因此若 O 不在子集中, 子集不可能构成子空间, 这就可以排除大部分情况 (需要证明不是子空间的题目往往都是对数乘封闭的, 一般需要通过加法构造反例)。

对于基和维数, 根据定义, 必然先找到基后确定维数, 因此核心难点在于确定一组基。一般来说, 如果子空间是由原空间进行一些限制得到的, 考虑坐标时其往往成为某个齐次线性方程组的解空间, 因此可以通过线性方程组思路确定基与维数。当然, 如果能直接看出一组能生成全空间的线性无关的向量, 问题自然也可以解决。以下用 E_{ij} 表示第 i 行第 j 列是 1, 其他是 0 的 n 阶方阵, 省略较直接的验证过程。

- (a) 是子空间。直接验证两个实对称方阵的和还是实对称方阵、实对称方阵数乘还是实对称方阵即可。实对称方阵的限制 $A = A^T$ 可以看作 $\frac{n(n-1)}{2}$ 个独立的线性方程 $a_{ij} = a_{ji}, 1 \leq i < j \leq n$, 由此可发现 E_{11}, \dots, E_{nn} 与 $E_{ij} + E_{ji}, 1 \leq i < j \leq n$ 构成一组基, 维数为 $\frac{n(n+1)}{2}$ 。
- (b) 是子空间。直接验证两个实斜对称方阵的和还是实斜对称方阵、实斜对称方阵数乘还是实斜对称方阵即可。实斜对称方阵的限制 $A = A^T$ 可以看作 $\frac{n(n+1)}{2}$ 个独立的线性方程 $a_{11} = \cdots = a_{nn} = 0, a_{ij} = -a_{ji}, 1 \leq i < j \leq n$, 由此可发现 $E_{ij} - E_{ji}, 1 \leq i < j \leq n$ 构成一组基, 维数为 $\frac{n(n-1)}{2}$ 。
- (c) 是子空间。直接验证两个对角阵的和还是对角阵、对角阵数乘还是对角阵即可。对角阵的限制也即要求对角元之外的元素全为 0, 由此可发现 E_{11}, \dots, E_{nn} 构成一组基, 维数为 n 。
- (d) 是子空间。直接验证两个上三角阵的和还是上三角阵、上三角阵数乘还是上三角阵即可。上三角阵的限制也即要求对角线以下的元素全为 0, 由此可发现 $E_{ij}, 1 \leq i \leq j \leq n$ 构成一组基, 维数为 $\frac{n(n+1)}{2}$ 。
- (e) 不是子空间。计算可发现 $E_{12}^2 = E_{21}^2 = O$, 但 $(E_{12} + E_{21})^2 = E_{11} + E_{22} \neq O$ 。
- (f) 不是子空间。当 a 不为 0 时, 通过 O 不在其中可知不是线性空间。当 $a = 0$ 时, 由于 $n > 2$, $\det E_{11} = \det(I - E_{11}) = 0$, 而 $\det I \neq 0$, 从而不构成子空间。
- (g) 当且仅当 $r = 0$ 或 n 时为子空间。当 $r = 0$ 时只有 O , 从而是子空间, 维数为 0; 当 $r = n$ 时包含全部 n 阶方阵, 从而是子空间, 基为所有 E_{ij} , 维数为 n^2 。对其他情况, 设 $A_r = E_{11} + \cdots + E_{rr}$, 则 $\text{rank } A_r = r \leq r$, $\text{rank } E_{r+1,r+1} = 1 \leq r$, $\text{rank}(A_r + E_{r+1,r+1}) = \text{rank } A_{r+1} = r+1 > r$, 从而不构成子空间。
- (h) 当且仅当 $a = 0$ 时为子空间。当 a 不为 0 时, 通过 O 不在其中可知不是线性空间。当 $a = 0$ 时, 限制是一个线性方程 $a_{11} + \cdots + a_{nn} = 0$, 由此可发现 $E_{ij}, i \neq j$ 与 $E_{11} - E_{kk}, k = 2, \dots, n$ 构成一组基, 维数为 $n^2 - 1$ 。

- (i) 不是子空间。计算特征多项式可发现, E_{12} 与 $-E_{21}$ 的特征值都全为 0, 而 $E_{21} - E_{12}$ 的特征值为 $n - 2$ 个 0 与 $\pm i$, 从而不构成子空间。
- (j) 不是子空间。由于 $E_{11} + E_{12}$ 与 $E_{22} + E_{12}$ 都满足 0 的代数重数、几何重数为 $n - 1$, 1 的代数重数、几何重数为 1, 它们都可以相似对角化, 而 $E_{11} + E_{22} + 2E_{12}$ 利用特征方阵计算可知 Jordan 标准形为 $n - 2$ 个特征值 0 的一阶 Jordan 块与 1 个特征值 1 的二阶 Jordan 块, 因此不可相似对角化, 从而不构成子空间。
- (k) 是子空间。 $AB = BA$ 、 $AC = CA$ 则 $A(B + C) = AB + AC = BA + CA = (B + C)A$, 且 $A(\lambda B) = \lambda AB = (\lambda B)A$ 。在 A 不可对角化时, 此子空间基和维数的确定往往非常困难。
- (l) 是子空间。直接验证 A 的多项式的和还是 A 的多项式、 A 的多项式数乘还是 A 的多项式即可。我们将在之后学习线性变换的最小多项式时解决其基与维数的问题。

3. 先说明封闭性, 加法的封闭性由 V 加法的封闭性得到, 而由于 $\mathbb{R} \subset \mathbb{C}$, $\lambda \odot a = \lambda \cdot a$ 一定还在 $V_{\mathbb{R}}$ 中。接着, $V_{\mathbb{R}}$ 加法与数乘的性质验证是直接的, 因为可以直接从 V 的加法、数乘性质得到, 这就得到了它构成一个线性空间。

* 虽然上面的验证看起来非常废话, 但仍然是必要的。数学上有时会称这种旧“东西”上通过等号直接定义出的新“东西”方式叫做诱导——这是一种叫法而非严谨的数学名词——出于等号, 诱导往往能保持原本“东西”的部分性质, 但有些性质会存在区别, 例如下面所说的基与维数。

由于 \mathbb{C} 作为 \mathbb{C} 上线性空间是一维, 基可以取 1, 作为 \mathbb{R} 上线性空间是二维, 基可以取 1、 i , 我们可以猜测维数会变为 $2n$ 。与之前类似, 我们仍然希望找到其一组基, 而上述构造方式则带来了对基的猜测(注意 $i \cdot \alpha$ 不能写为 $i \odot \alpha$, 因为 \odot 要求左边的数是实数):

$$\alpha_1, \dots, \alpha_n, \quad i \cdot \alpha_1, \dots, i \cdot \alpha_n$$

下面说明它们的确是 $V_{\mathbb{R}}$ 的一组基, 从而其维数为 $2n$ 。

- 线性无关性

若存在实数 $a_1, \dots, a_n, b_1, \dots, b_n$ 使得(由于加法完全相同, 我们不再区分 $+$ 与 \oplus , 这里实际上都是 \oplus)

$$\sum_{k=1}^n a_k \odot \alpha_k + \sum_{k=1}^n b_k \odot (i \cdot \alpha_k) = 0$$

利用 \odot 的定义, 我们可以将上式化为

$$\sum_{k=1}^n a_k \cdot \alpha_k + \sum_{k=1}^n b_k \cdot (i \cdot \alpha_k) = 0$$

进一步利用 V 的数乘结合律、分配律可得

$$\sum_{k=1}^n (a_k + b_k i) \cdot \alpha_k = 0$$

由 $\alpha_1, \dots, \alpha_n$ 为 V 的基可知所有 $a_k + b_k i$ 都为 0, 而由于 a_k, b_k 为实数即得所有 $a_k = b_k = 0$, 得证。

- 能表出所有向量

利用 $\alpha_1, \dots, \alpha_n$ 为 V 的基可知对任何向量 α , 存在复数 c_1, \dots, c_n 使得

$$\alpha = \sum_{k=1}^n c_k \cdot \alpha_k$$

设 $c_k = a_k + b_k i$, 其中 a_k, b_k 为实数, 与上方完全相同计算可知

$$\alpha = \sum_{k=1}^n a_k \odot \alpha_k + \sum_{k=1}^n b_k \odot (i \cdot \alpha_k)$$

这就说明了 $V_{\mathbb{R}}$ 任何向量都可被这 $2n$ 个向量表出。

4. 若 V 是有限维空间, 其基的个数即为维数, 而利用映射知识可知两个有限集存在双射当且仅当元素个数相等 (虽然这个结论很直观, 我们仍然将在下一次习题课证明它), 从而条件化为 $U \subset V$ 且 $\dim U = \dim V$ 。考虑 U 的一组基 $\alpha_1, \dots, \alpha_k$, 若它们不是 V 的基, 则 V 中还存在不能被它们线性表出的向量 β , 利用上学期知识可知 $\alpha_1, \dots, \alpha_k, \beta$ 线性无关, 与 $\dim V = k$ 矛盾。

利用类似希尔伯特旅店的思路 (每个元素映射到后一个元素, 事实上需要先取出可数子集, 这部分本课程不作要求) 可以证明无穷集合 S 去掉一个元素后与自身存在双射。由此, 设 V 的一组基为 $S = \{\alpha_i \mid i \in I\}$, 去掉其中任何一个 α_0 , 剩余的基生成子空间为 U , 则 $S \setminus \{\alpha_0\}$ 构成 U 的一组基, 它与 V 的一组基存在双射, 但根据基的线性无关性可知 α_0 不能被其他基生成, 从而 $U \neq V$ 。

* 这题揭示了无穷维线性空间与有限维的某种本质区别: 不再能通过某种“维数计算”确定相等, 讨论线性映射时, 我们还将再次谈论。

5. 这个向量组可用形式乘法写为

$$(\alpha_1, \dots, \alpha_n) \begin{pmatrix} 1 & \lambda & & \\ & \ddots & \ddots & \\ & & 1 & \lambda \\ \lambda & & & 1 \end{pmatrix}$$

由此, 根据过渡矩阵知识, 它们构成 V 的一组基当且仅当矩阵可逆。直接计算其行列式为 $(-\lambda)^n - 1$, 由此 $\lambda \in \mathbb{K}$ 且 $\lambda^n \neq (-1)^n$ 时它们生成的线性空间就是 V , 全部向量构成一组基。

若 $\lambda^n = (-1)^n$, 这 n 个向量线性相关。下面证明前 $n-1$ 个向量线性无关, 这样就得到了它们的一个极大线性无关组, 从而生成线性空间 $n-1$ 维, 前 $n-1$ 个向量构成其一组基。

考虑方程

$$\sum_{i=1}^{n-1} \lambda_i (\alpha_i + \lambda \alpha_{i+1}) = 0$$

由于 α_1 前的系数为 λ_1 , 由 $\alpha_1, \dots, \alpha_n$ 线性无关可知必然 $\lambda_1 = 0$, 再考虑 $\lambda_2, \lambda_3, \dots$, 以此类推可得到所有 λ_i 均为 0, 从而线性无关。

6. 我们先证明 S 的线性无关。这里引用两个分析中的结论, 首先, 对 S 中任何两个不同函数 $f(x), g(x)$ 有 (这个结论称为三角函数系的正交性)

$$\int_{-\pi}^{\pi} f(x)g(x)dx = 0$$

其次, 一个非负连续函数在 $[-\pi, \pi]$ 积分为 0 当且仅当其在 $[-\pi, \pi]$ 恒为 0。

将 S 中的向量记为 $f_1(x), \dots, f_7(x)$, 设 $\sum_{i=1}^7 \lambda_i f_i(x) = 0$, 则同平方并积分有

$$\int_{-\pi}^{\pi} \left(\sum_{i=1}^7 \lambda_i f_i(x) \right)^2 dx = 0$$

展开平方, 由于所有交叉项积分都为 0, 可以改写左侧为

$$\sum_{i=1}^7 \lambda_i^2 \int_{-\pi}^{\pi} f_i^2(x) dx = 0$$

而由于 $f_i(x)$ 均非零，平方积分均非零，因此只能所有 λ_i^2 为 0，也即所有 λ_i 为 0，从而得到线性无关。

利用三角函数知识直接计算可以得到

$$T = SP, \quad P = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} \end{pmatrix}$$

直接展开计算行列式可验证 P 可逆，从而 T 构成一组基， P 就是 S 到 T 的过渡矩阵。

利用三角函数知识得

$$\frac{\sin 4x}{\sin x} = 4 \cos x (\cos^2 x - \sin^2 x) = -4 \cos x \sin^2 x + 4 \cos^3 x$$

$$\frac{\sin 4x}{\cos x} = 4 \sin x (\cos^2 x - \sin^2 x) = -4 \sin^3 x + 4 \sin x \cos^2 x$$

从而二者在 T 下的坐标为 $(0, 0, 0, 0, -4, 0, 4)^T$ 、 $(0, 0, 0, -4, 0, 4, 0)^T$ ，再左乘 P 就得到 S 下的坐标为 $(0, 0, 2, 0, 0, 0, 2)^T$ 、 $(0, -2, 0, 0, 0, 2, 0)$

7. 我们先证明推广的结论：若 $\mathbb{R}[x]_n$ 中 n 个向量构成的向量组满足 $0, 1, \dots, n-1$ 次多项式各一个，则其构成一组基。

这里介绍两种证明方式（事实上它们本质是相同的）：

- 由于向量组个数等于空间维数，只要证明线性无关即构成一组基。设其中次数为 k 的多项式为 $f_k(x)$ ，考虑方程

$$\lambda_0 f_0(x) + \dots + \lambda_{n-1} f_{n-1}(x) = 0$$

若 λ_i 不全为 0，设下标最大的非零元素为 λ_k ，则可发现左侧的 k 次项只有 $\lambda_k f_k(x)$ 有，因此由 $f_k(x)$ 为 k 次多项式可知非零，矛盾，从而得证。

- 设其中次数为 k 的多项式为 $f_k(x)$ ，标准基为 $1, x, \dots, x^{n-1}$ ，可写出形式乘法

$$(f_0(x), \dots, f_{n-1}(x)) = (1, x, \dots, x^{n-1})P$$

由于次数要求，可发现 P 的对角元非零，且其下三角部分均为 0，由此其为对角元非零的上三角阵，从而知可逆，因此 $f_0(x), \dots, f_{n-1}(x)$ 构成一组基。

接下来计算过渡矩阵。设过渡矩阵为 Q ，由定义其应满足

$$(1, x, \dots, x^{n-1}) = (1, x-a, \dots, (x-a)^{n-1})Q$$

这里介绍一个相对技巧性（而不用直接计算矩阵逆的）的手段。由于两侧都为 x 的多项式，代入 $x = y+a$ 仍然成立，从而可以得到

$$(1, y+a, \dots, (y+a)^{n-1}) = (1, y, \dots, y^{n-1})Q$$

由此直接展开对比系数可以得到 Q 的第 i 行第 j 列元素 q_{ij} 满足

$$q_{ij} = \begin{cases} 0 & i > j \\ C_{j-1}^{i-1} a^{j-i} & i \leq j \end{cases}$$

* 补充一个与第五题证明方法类似可得到的重要结论：若 $S = (\alpha_1, \dots, \alpha_n)$ 线性无关，向量组 $T = (\beta_1, \dots, \beta_m)$ 可以用形式乘法写为

$$T = SP$$

这里 P 为 $n \times m$ 阶矩阵，则 T 线性无关当且仅当 P 列满秩。

证明：

T 线性无关当且仅当 $Tx = 0$ 对列向量 x 只有零解，这又等价于 $SPx = 0$ 只有零解，而根据 S 的线性无关性可知 $SPx = 0$ 等价于 $Px = 0$ ，从而最终等价为 $Px = 0$ 只有零解，由解空间维数定理即知这等价于 P 列满秩。

§18.3 直和

18.3.1 交空间与和空间

第一节中，我们介绍了线性空间、子空间与生成空间的概念，并给出了基的定义与有限维时维数的定义。接下来，我们的证明将遵循以下原则：对无限维线性空间也成立的结论不通过维数计算证明（例如第一同构定理等），无限维时无需取出一组基证明的结论不通过取基证明（这意味着此结论的成立不依赖选择公理）。

就像集合中可以考察交集、并集等“用两个子集构造出新的子集”概念，线性空间中也可以考察从给定的子空间中构造新的子空间。最简单的方式是所谓交空间，也就是，对 \mathbb{K} 上线性空间 U 与子空间 V 、 W ，其交集 $V \cap W$ 也是 \mathbb{K} 上的子空间，称为交空间。

证明：

注意到验证子空间只需验证封闭性即可，即可以进行简单的验证：若 $a, b \in V \cap W$ ，由 $a, b \in V$ 可知 $a + b \in V$ ，同理 $a + b \in W$ ，从而 $a + b \in V \cap W$ 。若 $a \in V \cap W$ ， $\lambda \in \mathbb{K}$ ，由 $a \in V$ 可知 $\lambda a \in V$ ，同理 $\lambda a \in W$ ，从而 $\lambda a \in V \cap W$ 。

不过，虽然它是线性空间的验证是简单的，它的计算往往是一个相对困难的问题。我们先考虑向量空间 \mathbb{K}^n 中，若已知子空间 V 的一组基 $\alpha_1, \dots, \alpha_s$ 、 W 的一组基 β_1, \dots, β_t ，如何计算 $V \cap W$ 的一组基呢？

解答：

根据定义，当且仅当存在 \mathbb{K} 中的 $\lambda_1, \dots, \lambda_s$ 与 μ_1, \dots, μ_t

$$\gamma = \sum_{i=1}^s \lambda_i \alpha_i = \sum_{j=1}^t \mu_j \beta_j$$

时， $\gamma \in V \cap W$ 。我们将中间的等号重新写成

$$\sum_{i=1}^s \lambda_i \alpha_i - \sum_{j=1}^t \mu_j \beta_j = 0$$

将 $\lambda_1, \dots, \lambda_s, \mu_1, \dots, \mu_t$ 看作未知数，这是一个 $t+s$ 个未知数、 n 个方程的线性方程组。

在上学期我们已经学到，可以通过消元构造出它的一个基础解系，假设其中有 r 个解，分别为

$$\lambda_1^{(1)}, \dots, \lambda_s^{(1)}, \mu_1^{(1)}, \dots, \mu_t^{(1)}$$

$$\lambda_1^{(2)}, \dots, \lambda_s^{(2)}, \mu_1^{(2)}, \dots, \mu_t^{(2)}$$

⋮

$$\lambda_1^{(r)}, \dots, \lambda_s^{(r)}, \mu_1^{(r)}, \dots, \mu_t^{(r)}$$

我们只考虑前 s 个分量，并用 V 的基将它们重组为

$$\sum_{i=1}^s \lambda_i^{(1)} \alpha_i, \sum_{i=1}^s \lambda_i^{(2)} \alpha_i, \dots, \sum_{i=1}^s \lambda_i^{(r)} \alpha_i$$

下面说明这 r 个向量就是交空间的一组基。

- 线性无关性

若它们线性相关，也即有非零的 c_1, \dots, c_r 使得

$$\sum_{k=1}^r c_k \sum_{i=1}^s \lambda_i^{(k)} \alpha_i = 0$$

利用分配律与结合律交换求和次序，将求和改写为

$$\sum_{i=1}^s \left(\sum_{k=1}^r c_k \lambda_i^{(k)} \right) \alpha_i = 0$$

由于所有 α_i 线性无关，这就可以得到

$$\forall i = 1, \dots, s, \quad \sum_{k=1}^r c_k \lambda_i^{(k)} = 0$$

另一方面，利用基础解系都为解可知

$$\sum_{i=1}^s \lambda_i^{(k)} \alpha_i = \sum_{j=1}^t \mu_j^{(k)} \beta_j$$

从而有

$$\sum_{k=1}^r c_k \sum_{j=1}^t \mu_j^{(k)} \beta_j = 0$$

同理交换求和次序并利用所有 β_j 线性无关得到

$$\forall j = 1, \dots, t, \quad \sum_{k=1}^r c_k \mu_j^{(k)} = 0$$

将两式结合可发现

$$\sum_{k=1}^r c_k (\lambda_1^{(k)}, \dots, \lambda_s^{(k)}, \mu_1^{(k)}, \dots, \mu_t^{(k)}) = 0$$

但根据基础解系的线性无关性可知只能所有 c_k 全为 0，矛盾。

- 能表出所有向量

对任何 $\gamma \in V \cap W$ ，上方已经说明有

$$\gamma = \sum_{i=1}^s \lambda_i \alpha_i = \sum_{j=1}^t \mu_j \beta_j$$

利用基础解系的定义，一定存在 $c_1, \dots, c_k \in \mathbb{K}$ 使得

$$(\lambda_1, \dots, \lambda_s, \mu_1, \dots, \mu_t) = \sum_{k=1}^r c_k (\lambda_1^{(k)}, \dots, \lambda_s^{(k)}, \mu_1^{(k)}, \dots, \mu_t^{(k)})$$

由此即有

$$\forall i = 1, \dots, s, \quad \sum_{k=1}^r c_k \lambda_i^{(k)} = \lambda_i$$

从而类似证明线性无关性时操作求和计算可发现

$$\gamma = \sum_{i=1}^s \lambda_i \alpha_i = \sum_{k=1}^r c_k \sum_{i=1}^s \lambda_i^{(k)} \alpha_i$$

虽然证明过程比较复杂，尤其是在证明线性无关时需要一些对求和的细节操作，但最终得到的算法还是非常明确的：以 $(\alpha_1, \dots, \alpha_s, -\beta_1, \dots, -\beta_t)$ 作为系数矩阵（由于 $-\beta_1, \dots, -\beta_t$ 也是 W 的基，事实上可以去掉负号）求解线性方程组，并将一个基础解系的前 s 个分量作为线性组合系数组合 $\alpha_1, \dots, \alpha_s$ ，得到的就是交空间的基。

对一般的有限维线性空间如何进行计算呢？最直观的做法是，取定 U 的一组基 S ，将 V 、 W 的基都写为 S 下的坐标，与上相同即可得到交空间的基的坐标，再用坐标的定义即可通过 S 还原出向量。不过，有的时候，在一般线性空间中，由于 V 、 W 的定义，它们的交集可能容易直接确定，这时算出交集再取基也是可行的做法。

与集合不同的是，直接将两个子空间取并往往得不到线性空间：考虑 \mathbb{R}^2 的子空间 $\langle(1, 0)^T\rangle, \langle(0, 1)^T\rangle$ ，可发现 $(1, 0), (0, 1)$ 都在并集中，但 $(1, 1)$ 不在，由此不能构成子空间。不过，这时我们可以考虑在并集之上利用之前提到的生成结构，也即，对 V 的子空间 U_1, U_2 ，我们定义它们的生成子空间为

$$U = \langle U_1 \cup U_2 \rangle$$

我们下面验证其可以有一个更简单的表述，和空间（第二个等号即为定义）：

$$U = U_1 + U_2 = \{u \in V \mid \exists u_1 \in U_1, u_2 \in U_2, u = u_1 + u_2\}$$

证明：

要证明两个集合相等的一般思路是证明相互包含，不过，由于生成子空间本身就是由包含性定义的，我们希望直接验证和空间符合生成子空间定义。

首先，由于封闭性要求，任何包含 $U_1 \cup U_2$ 的子空间必然会包含形如 $u_1 + u_2$, $u_1 \in U_1, u_2 \in U_2$ 的向量，从而其会包含 $U_1 + U_2$ 。

根据生成子空间的定义，只要证明了 $U_1 + U_2$ 构成一个子空间，我们即可以说明任何包含 $U_1 \cup U_2$ 的子空间包含它。而验证子空间只需说明封闭性：若 $u_1 + u_2 \in U_1 + U_2, v_1 + v_2 \in U_1 + U_2$ ，其中 $u_1, v_1 \in U_1, u_2, v_2 \in U_2$ ，利用子空间封闭性有 $u_1 + v_1 \in U_1, u_2 + v_2 \in U_2$ ，因此它们的和

$$u_1 + u_2 + v_1 + v_2 = (u_1 + v_1) + (u_2 + v_2) \in U_1 + U_2$$

同理可知数乘

$$\lambda(u_1 + u_2) = (\lambda u_1) + (\lambda u_2) \in U_1 + U_2$$

这就得到了证明。

虽然和空间的定义相对简单，我们还是有必要提及有限维时和空间的计算。仍然考虑向量空间 \mathbb{K}^n 中，已知子空间 V 的一组基 $\alpha_1, \dots, \alpha_s$ 、 W 的一组基 β_1, \dots, β_t ，下面来计算 $V + W$ 的一组基。不难证明， $\alpha_1, \dots, \alpha_s, \beta_1, \dots, \beta_t$ 的极大线性无关组就是 $V + W$ 的一组基。

证明：

线性无关性利用极大线性无关组定义即得。由于极大线性无关组可以表出 $\alpha_1, \dots, \alpha_s, \beta_1, \dots, \beta_t$ ，而这些向量可以表出 U, V 中的任何元素，从而根据定义可表出 $U + V$ 中任何元素，即得结果。

由此，我们通过上学期计算极大线性无关组的算法可以直接得到和空间。不过，我们还需要补充一些重要的注释：

- 计算性的习题往往需要同时计算交空间与和空间。可以发现，无论是计算极大线性无关组时，还是解线性方程组 $(\alpha_1, \dots, \alpha_s, \beta_1, \dots, \beta_t)x = 0$ 时，都需要对这个向量组进行行变换。因此，通过行变换化为简化阶梯形矩阵的结果可以同时在交空间、和空间计算时使用。
- 一般的有限维线性空间最直观的计算方式仍然是取定 U 的一组基 S ，将 V 、 W 的基都写为 S 下的坐标，与上相同利用极大线性无关组得到和空间的基的坐标，再用坐标的定义即可通过 S 还原出向量。

- 一个自然而然的问题是，上述做法真的严谨吗？我们可以发现，上述做法本质是把任何 \mathbb{K} 上的 n 维线性空间 U 当作 \mathbb{K}^n 来处理，因此严谨证明它的成立需要下一次习题课介绍的同构知识。到此处，至少希望大家能感受到这个做法是自然的。
- 从我们的交空间、和空间算法可以得到一个重要的结论：假设 \mathbb{K} 上线性空间 U 子空间 V 的一组基 $\alpha_1, \dots, \alpha_s$ 、 W 的一组基 β_1, \dots, β_t ，则交空间维数为 $(\alpha_1, \dots, \alpha_s, \beta_1, \dots, \beta_t)x = 0$ 的解空间维数，和空间维数为 $\text{rank}(\alpha_1, \dots, \alpha_s, \beta_1, \dots, \beta_t)$ ，利用解空间维数定理可以得到

$$\dim V \cap W + \dim(V + W) = s + t$$

也即著名的和空间维数公式

$$\dim(V + W) = \dim V + \dim W - \dim(V \cap W)$$

除了能用来计算维数外，它的更深含义将在之后介绍第二同构定理时讨论。

- 利用定义可发现

$$U_1 + U_2 = U_2 + U_1$$

$$U_1 + (U_2 + U_3) = (U_1 + U_2) + U_3 = \{u_1 + u_2 + u_3 \mid u_i \in U_i\}$$

由此可以直接写出多个空间求和，如 $U_1 + U_2 + U_3$ ，不会引起歧义。

- 虽然这门课程不会涉及，但还是可以简单介绍：无穷个空间的和空间定义为所有能从它们中取出任意有限个并在其中各取一个向量进行求和得到的向量。

和空间与集合的并集有一定类似的性质，但很多情况下完全不同，我们将在之后的例题里看到这种生成性质与直接取并本质差别。个人建议是，涉及判断空间中的等式是否成立，可以考虑 \mathbb{R}^2 中 $U_1 = \langle(0, 1)^T\rangle$ 、 $U_2 = \langle(1, 0)^T\rangle$ 、 $U_3 = \langle(1, 1)^T\rangle$ 作为例子，这样的三个彼此交为 $\{0\}$ 的子空间两两和为全空间在集合中是不可能出现的——集合中，若三个集合两两交为空，不可能任意两个并集都是全集。

18.3.2 直和与补空间

在介绍完和空间后，我们还需要介绍一种特殊的和结构，也就是直和。简单来说，若 $U_1 \cap U_2 = \{0\}$ ，我们就将 $U_1 + U_2$ 记为 $U_1 \oplus U_2$ 。

从直觉上，这个定义理应并不复杂，却有点莫名其妙：为什么我们需要把交为 $\{0\}$ 时候的和进行单独的定义呢？我们将在之后的讨论中看到，这是因为它对应着某种固定的空间分解方式。如无特殊说明，本节之后提到的空间都是某 \mathbb{K} 上线性空间 V 的子空间。

我们先证明，对于子空间 U_1 、 U_2 ，记 $W = U_1 + U_2$ ，以下命题互相等价：

1. $W = U_1 \oplus U_2$;
2. $U_1 \cap U_2 = \{0\}$;
3. 对任何 $w \in W$ ，有 $w = u_1 + u_2, u_1 \in U_1, u_2 \in U_2$ ，且 u_1, u_2 唯一确定；
4. 存在 $w \in W$ ，使得 $w = u_1 + u_2, u_1 \in U_1, u_2 \in U_2$ ，且 u_1, u_2 唯一确定；
5. U_1 、 U_2 的任意各一组基构成 W 一组基；
6. 存在 U_1 、 U_2 的各一组基构成 W 一组基。

证明：

1 等价于 2：这就是直和的定义。

2 推 3：由和空间定义知存在分解 $w = u_1 + u_2, u_1 \in U_1, u_2 \in U_2$ ，若有 u'_1, u'_2 使得 $w = u'_1 + u'_2, u'_1 \in U_1, u'_2 \in U_2$ ，移项可得 $u_1 - u'_1 = u'_2 - u_2$ ，由于左右分别属于 U_1, U_2 ，即可知 $u_1 - u'_1 = u'_2 - u_2 = 0$ ，从而 $u_1 = u'_1, u_2 = u'_2$ ，这就得到了分解唯一。

3 推 4：由于子空间 W 非空（至少包含 0），对 W 中任意元素都满足可以推出存在 w 满足。

4 推 2：若否，存在 $u \neq 0 \in U_1 \cap U_2$ ，则由封闭性 $-u \in U_1 \cap U_2$ 。对任何 $w \in W$ ，根据和空间定义存在分解 $w = u_1 + u_2, u_1 \in U_1, u_2 \in U_2$ ，同时也有 $w = (u_1 - u) + (u_2 + u)$ ，由之前假设 $u_1 - u \in U_1, u_2 + u \in U_2$ ，且由 $u \neq 0$ 此分解与 $u_1 + u_2$ 不同，矛盾。

2 推 5：考虑 U_1, U_2 的各一组基，若它们线性相关，类似上学期操作按照等号一边 U_1 、一边 U_2 移项即可得到 $U_1 \cap U_2$ 中的非零元素。另一方面，根据和空间定义可从它们能生成 U_1, U_2 得到它们能生成 W ，从而得证。

5 推 6：由任意各一组基都构成 W 一组基与基的存在性可知结论。

6 推 2：若存在非零 $u \in U_1 \cap U_2$ ，考虑其在 U_1, U_2 的各一组基下的表示可发现 U_1, U_2 的基必然线性相关，不可能构成 W 的基，矛盾。

当然，有限维时，利用和空间维数定理，我们还可以以维数给出等价条件

$$\dim U_1 + \dim U_2 = \dim W$$

从上述的等价命题中，我们已经可以看到直和的意义了：有了直和后， W 中的任何元素可以与 U_1, U_2 中各一个元素等价，这实际上成为了某种意义上的坐标，以此，一些问题可以分别在 U_1, U_2 进行讨论，再还原到 W 上。下一次习题课介绍完限制映射后，即可以讨论这个层面的应用。

$U_1 \oplus U_2 = U_2 \oplus U_1$ 仍然直接，我们下面要证明，若 $W = (U_1 \oplus U_2) \oplus U_3$ ，则 $W = U_1 \oplus (U_2 \oplus U_3)$ ，从而直和也可以交换、结合，进而 n 个子空间的直和 $U_1 \oplus \dots \oplus U_n$ 是不会引起歧义的。

证明：

由于 $(U_1 + U_2) + U_3 = U_1 + (U_2 + U_3)$ ，我们实际上只要证明 U_2 与 U_3 的和是直和、 U_1 与 $U_2 \oplus U_3$ 的和是直和即可。

若 $U_2 \cap U_3 \neq \{0\}$ ，则其中的元素一定在 $(U_1 \oplus U_2) \cap U_3$ 中，与 $U_1 \oplus U_2$ 与 U_3 的和是直和矛盾。

若 $U_1 \cap (U_2 \oplus U_3) \neq \{0\}$ ，设非零的 $u \in U_1$ 且 $u \in U_2 \oplus U_3$ 。根据和空间定义存在 $u = u_2 + u_3$ ，其中 $u_2 \in U_2, u_3 \in U_3$ ，于是 $u - u_2 \in U_1 \oplus U_2$ 且 $u - u_2 = u_3 \in U_3$ 。利用 $(U_1 \oplus U_2) \cap U_3 = \{0\}$ 可知 $u = u_2$ ，但这就得到 $u \in U_1 \cap U_2$ ，与 U_1, U_2 的和为直和矛盾，从而得证。

仿照上述命题的证明过程，我们可以说明，若 $W = U_1 + \dots + U_n$ ，则以下命题相互等价：

1. $W = U_1 \oplus \dots \oplus U_n$;
2. 对每个 U_i , $U_i \cap (U_1 + \dots + U_{i-1}) = \{0\}$;
3. 对每个 U_i , $U_i \cap (U_1 + \dots + U_{i-1} + U_{i+1} + \dots + U_n) = \{0\}$;
4. 对任何 $w \in W$, 有 $w = u_1 + \dots + u_n$, 其中 $u_i \in U_i$ 唯一确定;
5. 存在 $w \in W$, 使得 $w = u_1 + \dots + u_n$, 且 $u_i \in U_i$ 唯一确定;
6. U_1, \dots, U_n 的任意各一组基构成 W 一组基;
7. 存在 U_1, \dots, U_n 的各一组基构成 W 一组基。

证明：

1 等价于 2：这就是多个子空间和为直和的定义。

1 等价于 3：由多个子空间的直和可以交换、结合知 $W = U_1 \oplus \cdots \oplus U_{i-1} \oplus U_{i+1} \oplus \cdots \oplus U_n \oplus U_i$ ，此时再利用 2 即可得到 3。另一方面，由 $U_1 + \cdots + U_{i-1}$ 是 $U_1 + \cdots + U_{i-1} + U_{i+1} + \cdots + U_n$ 的子空间知 3 可以直接推出 2。

1 推 4：利用 $W = (U_1 \oplus \cdots \oplus U_{n-1}) \oplus U_n$ 可知其可以唯一分解为 $U_1 \oplus \cdots \oplus U_{n-1}$ 与 U_n 元素之和，由 $u_1 + \cdots + u_{n-1} \in U_1 \oplus \cdots \oplus U_{n-1}$ 即得 u_n 唯一，进一步归纳得证。

4 推 5：由于子空间 W 非空（至少包含 0），对 W 中任意元素都满足可以推出存在 w 满足。

5 推 3：记 $V_i = U_1 + \cdots + U_{i-1} + U_{i+1} + \cdots + U_n$ 、 $v_i = w - u_i$ ，若否，存在非零 $x \in U_i \cap V_i$ ，则 W 的分解 $w = u_i + v_i$ 还可以写为 $(u_i + x) + (v_i - x)$ ，于是 $u_i - x$ 也是符合要求的 u_i ，而 $v_i - x \in V_i$ 还是可以写为除 U_i 外各一个子空间的元素和，从而 W 的分解不唯一，矛盾。

1 推 6：利用两个空间情况可知 $U_1 \oplus \cdots \oplus U_{n-1}$ 、 U_n 的一组基构成 W 一组基，再归纳即可。

6 推 7：由任意各一组基都构成 W 一组基与基的存在性可知结论。

7 推 2：若否，取最小的 i 使得 $U_i \cap (U_1 + \cdots + U_{n-1}) \neq \{0\}$ ，则此前仍为直和，于是 U_1, \dots, U_{i-1} 的一组基构成 $U_1 + \cdots + U_{i-1}$ 的一组基，但根据两个空间和为直和的等价条件可知这组基与 U_i 的一组基线性相关，于是 U_1, \dots, U_i 的各一组基线性相关，再添加更多向量仍然线性相关，不可能构成 W 一组基，矛盾，从而得证。

有限维时仍然可以等价于维数等式

$$\sum_{i=1}^n \dim U_i = \dim W$$

不过此等式一般作为维数计算使用，很少作为判据，因为从中无法直接看出空间分解的性质。

就像对一个集合可以谈论补集，对一个空间也可以谈论补空间。具体来说，对 V 的子空间 U, W ，若 $V = U \oplus W$ ，则称 U, W 互为补空间。

自然的疑问是，对 V 的任何一个子空间 U ，补空间是否一定存在？我们先给出 $V = \mathbb{K}^n$ 时的一个算法。设 U 的一组基为 $\alpha_1, \dots, \alpha_r$ 。

解答：

由之前证明的等价性质，只需找到向量 $\beta_1, \dots, \beta_{n-r}$ 使得 $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{n-r}$ 构成 \mathbb{K}^n 的一组基。由线性无关性 $\alpha_1, \dots, \alpha_r$ 秩为 r ，将它们作为列拼成 n 行 r 列矩阵 A ，通过列变换可找到 A 的行向量的极大线性无关组，设为第 i_1, \dots, i_r 行。根据上学期知识， A 的 i_1, \dots, i_r 行构成的行列式非零。假设 1 到 n 除了 i_1, \dots, i_r 外的下标是 j_1, \dots, j_{n-r} ，取 $\beta_i = e_{j_i}$ ，这里 e_k 表示只有第 k 个分量为 1 的单位向量，则对后 $n-r$ 列依次展开可发现此时将所有 $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{n-r}$ 拼成的矩阵行列式为 A 的 i_1, \dots, i_r 行构成的行列式乘 -1 的某个次方，因此非零，从而这些向量构成一组基，得证。

* 由此，只要能看出 A 中某些行构成子式非零，即可取补空间一组基为剩下的行下标对应的单位向量，这是一个好用的“目测”补空间的方式。

我们以几个注释结束本部分：

- 对于一般的有限维线性空间 V ，可以通过坐标得到补空间的算法，与之前交空间、和空间类似。
- 利用直和的维数关系可发现有限维时 U 的补空间维数与 U 的维数和为全空间维数。
- 对无穷维线性空间，补空间的存在性依赖选择公理，本门课程中可以默认正确。
- 当 $U \neq \{0\}$ 或 V 时，补空间不唯一。我们仅说明有限维时为何成立：设 U 的一组基为 $\alpha_1, \dots, \alpha_r$ ，补空间一组基为 β_1, \dots, β_s ，则由于 s, r 都至少为 1，大家可以验证 $\alpha_1 + \beta_1, \beta_2, \dots, \beta_s$ 也可以成为补空间的一组基，且与 β_1, \dots, β_s 不等价。这与集合的补集唯一完全不同。

18.3.3 例题

1. 考虑多元实函数

$$f(x_1, \dots, x_n) = x_1^2 + \dots + x_p^2 - x_{p+1}^2 - \dots - x_r^2$$

其中 $p \geq 1$ 、 $r-p \geq 1$ 。 $f(x_1, \dots, x_n) = 0$ 的解集为 \mathbb{R}^n 的一个子集 A , 求

- (a) A 生成的 \mathbb{R}^n 子空间的维数与一组基;
- (b) A 包含的维数最大的向量空间的维数。

* 从此题中可以感受到生成和包含的本质差别。

2. 若 U 、 W_1 、 W_2 是某 \mathbb{K} 上线性空间 V 的子空间, 证明

$$(U + W_1) \cap (U + W_2) = U + (U + W_1) \cap W_2$$

并举出 U 、 W_1 、 W_2 使得

$$(U + W_1) \cap (U + W_2) \neq U + W_1 \cap W_2$$

* 由此集合的“容斥”性质对线性空间实际上不成立。

3. 给定 \mathbb{K} 上线性空间 V 中向量 $\alpha_1, \dots, \alpha_n$, 对 \mathbb{K}^n 子空间 U , 证明

$$V(U) = \{x_1\alpha_1 + \dots + x_n\alpha_n \mid x \in U\}$$

构成 V 的子空间。若 W 为 \mathbb{K}^n 另一个子空间, 证明

$$V(U + W) = V(U) + V(W)$$

$$V(U \cap W) \subset V(U) \cap V(W)$$

并给出 $V(U \cap W) \neq V(U) \cap V(W)$ 的例子。

4. 给定 $\mathbb{R}[x]$ 中的多项式 f 、 g , 证明 f 的倍式集合(可记为 (f))、 g 的倍式集合均为线性空间, 并计算它们的交空间与和空间, 给出一组基。

5. 给定 $a, b \in \mathbb{R}$ 、 $k \in \mathbb{N}^+$, 设 V_1 是 $\mathbb{R}[x]$ 中 $x-a, (x-a)^3, \dots, (x-a)^{2k-1}$ 生成的子空间, V_2 是 $\mathbb{R}[x]$ 中 $x-b, (x-b)^3, \dots, (x-b)^{2k-1}$ 生成的子空间, 若 $a \neq b$, 求证

$$V_1 \oplus V_2 = \mathbb{R}[x]_{2k}$$

6. 考虑 n 阶复方阵 A 与多项式 f 、 g , 回顾

$$\text{Ker } B = \{x \in \mathbb{C}^n \mid Bx = 0\}$$

为向量空间, 证明 $\gcd(f, g) = 1$ 时 $\text{Ker } f(A) \oplus \text{Ker } g(A) = \text{Ker } f(A)g(A)$, 并对逆命题举出反例。

* 此结论非常重要, 且事实上在无穷维仍然成立, 可对应之后学到的根子空间分解。

7. 对 \mathbb{K} 上的 n 维线性空间 V 与其 r 维子空间 V_1, \dots, V_s , 证明:

- (a) 当 $r \neq n$ 时存在元素 $v \in V$ 使得 $v \notin V_1 \cup V_2 \cup \dots \cup V_s$;
- (b) 存在空间 U 是 V_1, \dots, V_s 的共同补空间。

* 这样不同子空间的共同补空间是集合中的补集不具有的性质, 同样作为线性空间与集合本质差别的例子。

18.3.4 解答

1. 以 e_i 表示第 i 个单位向量。

(a) 我们证明 A 能够生成整个 \mathbb{R}^n , 从而基可取为 e_1, \dots, e_n 。

首先, 由于 f 只涉及前 r 个分量, 有 $e_{r+1}, \dots, e_n \in A$ 。此外, 直接计算可发现所有 $e_i \pm e_j$ 在 $1 \leq i \leq p < j \leq r$ 时在 A 中, 而 $e_i + e_j, e_i - e_j$ 可表出 e_i, e_j , 由此即得 e_1 到 e_r 也可由 A 中向量表出, 综合即得 A 能生成 \mathbb{R}^n 。

(b) 设 $m = \min(p, r - p)$, 考虑线性空间

$$V = \langle e_1 + e_{p+1}, e_2 + e_{p+2}, \dots, e_m + e_{p+m}, e_{r+1}, \dots, e_n \rangle$$

由于生成 V 的向量非零分量位置完全不同, 它们构成 V 的一组基, 从而 V 的维数为 $m + n - r$ 。另一方面, 对 V 中任何向量 x 都有 $x_1 = x_{p+1}, x_2 = x_{p+2}$, 直到 $x_m = x_{p+m}$, 且前 r 个分量中剩下的分量全为 0, 从而 $f(x) = 0$, 因此 $V \subset A$ 。

下面证明, A 中不存在 $m + n - r + 1$ 维子空间, 于是结果即为 $m + n - r$ 。先给出一个引理: \mathbb{R}^n 的任何一个 t 维子空间 U 中, 指定 $t - 1$ 个分量 i_1, \dots, i_{t-1} , 则 U 中存在这些分量均为 0 的非零向量 (这个引理曾经是某年线性代数 A1 的期中考题)。

证明:

设 U 的一组基为 $\alpha_1, \dots, \alpha_t$, 对 $\lambda_1, \dots, \lambda_t \in \mathbb{R}$, 考虑

$$\lambda_1 \alpha_1 + \dots + \lambda_t \alpha_t$$

可以发现, 要求其第 i_1, i_2, \dots, i_{t-1} 个分量为 0 构成了 $t - 1$ 个齐次线性方程, 而未知数 λ_i 共有 t 个, 因此一定存在非零解, 而由线性无关性, 代入非零解得到的 $\lambda_1 \alpha_1 + \dots + \lambda_t \alpha_t$ 非零, 从而即为 U 中第 i_1, \dots, i_{t-1} 是 0 的非零向量。

利用引理, 若 A 中存在 $m + n - r + 1$ 维子空间 U , 分类讨论。若 $p \leq r - p$, 有

$$m + n - r + 1 = n - (r - p) + 1$$

从而 U 中存在第 $1, 2, \dots, p, r + 1, r + 2, \dots, n$ 个分量为 0 的非零向量

$$x = (0, \dots, 0, x_{p+1}, \dots, x_r, 0, \dots, 0)^T$$

但根据 f 定义可发现 $f(x) < 0$, 矛盾。同理, 若 $r - p \leq p$, 指定后 $n - p$ 个分量为 0, U 中存在非零向量

$$x = (x_1, \dots, x_p, 0, \dots, 0)^T$$

于是 $f(x) > 0$, 也得到了矛盾。

2. 分为三个部分:

- $(U + W_1) \cap (U + W_2) \subset U + (U + W_1) \cap W_2$

由定义, 左侧任何元素 x 都可以写为 $x = u + w$, 其中 $u \in U, w \in W_2$, 且有 $x \in U + W_1$ 。

由此, 利用 $u \in U + W_1$ 可知 $w = x - u \in U + W_1$, 而 $w \in W_2$, 即得 $w \in (U + W_1) \cap W_2$, 从而 $u + w \in U + (U + W_1) \cap W_2$, 得证。

- $U + (U + W_1) \cap W_2 \subset (U + W_1) \cap (U + W_2)$

由于 $(U + W_1) \cap W_2 \subset U + W_1, U \subset U + W_1$ 可知 $U + (U + W_1) \cap W_2 \subset U + W_1$, 由于 $(U + W_1) \cap W_2 \subset W_2 \subset U + W_2, U \subset U + W_2$ 可知 $U + (U + W_1) \cap W_2 \subset U + W_2$, 综合两者得结论。

- 反例构造

考虑 \mathbb{R}^2 中 $U = \langle (1, 0)^T \rangle$, $W_1 = \langle (0, 1)^T \rangle$, $W_2 = \langle (1, 1)^T \rangle$, 则左侧为 \mathbb{R}^2 , 右侧为 U , 不相等。

* 如果觉得前两部分证明的直接构造方法没想到的话, 也可以考虑从基的角度出发, 如取出 U 的一组基, 扩充为 $U + W_1 \cap W_2$ 的一组基, 再分别扩充到 $U + W_1$ 、 $U + W_2$ …… 不过, 这类方法依赖选择公理, 而且只是把原本思考的难度转移到了基扩充顺序的构造上, 容易出现严谨性问题, 个人并不推荐。

3. 我们记 $S = (\alpha_1, \dots, \alpha_n)$, 从而可写出形式乘法

$$V(U) = \{Sx \mid x \in U\}$$

分为四个部分:

- 验证子空间

只需验证封闭性对 $V(U)$ 中元素 Sx 、 Sy , 其中 $x, y \in U$, 有 $Sx + Sy = S(x + y)$, 利用子空间封闭性 $x + y \in U$, 从而其仍在 $V(U)$ 中; 同理, 对 Sx 的数乘, $\lambda Sx = S(\lambda x) \in V(U)$, 从而得证。

- $V(U + W) = V(U) + V(W)$

直接利用定义得

$$V(U + W) = \{S(u + w) \mid u \in U, w \in W\} = \{Su + Sw \mid u \in U, w \in W\} = V(U) + V(W)$$

- $V(U \cap W) \subset V(U) \cap V(W)$

直接利用定义得

$$V(U \cap W) = \{Sx \mid x \in U \cap W\} \subset \{Sx \mid x \in U\} = V(U)$$

同理其在 $V(W)$ 中, 从而得证。

- 反例构造

考虑 $V = \mathbb{R}$, $\alpha_1 = 1$ 、 $\alpha_2 = 2$, $n = 2$, $U = \langle (1, 0) \rangle$ 、 $W = \langle (0, 1) \rangle$, 则 $V(U) = V(W) = \mathbb{R}$, $V(U \cap W) = \{0\}$ 。

4. • 线性空间验证

只需验证封闭性。由于 f 的倍式之和仍为 f 的倍式, 数乘仍为 f 的倍式即得为子空间。

- (f) 的一组基

当 $f = 0$ 时, 即为 $\{0\}$, 空集构成一组基, 否则我们证明

$$f(x), \quad xf(x), \quad x^2f(x), \quad \dots$$

构成其一组基。

先证明它们线性无关: 若否, 存在不全为 0 的 a_0, \dots, a_{n-1} 使得

$$\sum_{i=0}^{n-1} a_i x^i f(x) = \left(\sum_{i=0}^{n-1} a_i x^i \right) f(x) = 0$$

而由于 $f(x)$ 不为 0, 对比系数可知只能 $\sum_i a_i x^i = 0$, 矛盾。

再证明它们能表出全空间。 (f) 中任何元素可以写为 $f(x)g(x)$, 其中 g 为多项式, 由此设 $g(x) = \sum_{i=0}^{n-1} a_i x^i$ 即得

$$f(x)g(x) = \sum_{i=0}^{n-1} a_i x^i f(x)$$

从而得证。

- 交空间计算

根据最小公倍式定义，一个多项式既是 f 的倍式又是 g 的倍式当且仅当它是 $\text{lcm}(f, g)$ 的倍式，从而根据交空间定义有

$$(f) \cap (g) = (\text{lcm}(f, g))$$

再由第二部分证明给出一组基。

- 和空间计算

利用裴蜀定理可知一个多项式能写为 f 某倍式和 g 某倍式之和当且仅当其为 $\text{gcd}(f, g)$ 的倍式（可见本讲义 14.1.4），而 $(f) + (g)$ 即为所有能写为 f 某倍式和 g 某倍式之和的多项式，从而

$$(f) + (g) = (\text{gcd}(f, g))$$

再由第二部分证明给出一组基。

5. 利用直和的维数性质，由于生成向量的个数为 k 有 $\dim V_1 \leq k$ 、 $\dim V_2 \leq k$ ，从而只要证明了 $V_1 + V_2 = \mathbb{R}[x]_{2k}$ ，就可以直接得到

$$\dim V_1 = \dim V_2 = k, \quad V_1 \oplus V_2 = \mathbb{R}[x]_{2k}$$

记 $\alpha_i = (x - a)^{2i-1}$ 、 $\beta_i = (x - b)^{2i-1}$ ， $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k$ 可以生成向量组

$$\alpha_1 - \beta_1, \beta_1, \alpha_2 - \beta_2, \beta_2, \dots, \alpha_k - \beta_k, \beta_k$$

由于 $a \neq b$ ，直接通过二项式定理可展开计算得 $\alpha_i - \beta_i$ 是 $2i - 2$ 次多项式，由此利用本讲义 18.2.3 中第 7 题解答里证明的结论可知上方向量组构成 $\mathbb{R}[x]_{2k}$ 一组基，因此 $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k$ 的生成子空间包含 $\mathbb{R}[x]_{2k}$ ，又由它们都在 $\mathbb{R}[x]_{2k}$ 中即得到 $V_1 + V_2 = \mathbb{R}[x]_{2k}$ ，得证。

* 大部分情况下验证直和都是从交为 $\{0\}$ 的角度，但本题是一个从和空间维数验证直和的例子，提醒大家注意直和的不同等价形式可能适用不同情况。

6. • $\text{Ker } f(A) \cap \text{Ker } g(A)$ 和为直和

利用裴蜀定理，存在多项式 $u(x)$ 、 $v(x)$ 使得 $u(x)f(x) + v(x)g(x) = 1$ ，从而代入 A 即

$$u(A)f(A) + v(A)g(A) = I$$

右乘 x 得到

$$u(A)f(A)x + v(A)g(A)x = x$$

从而 $f(A)x = g(A)x = 0$ 可得 $x = 0$ ，即得证

$$\text{Ker } f(A) \cap \text{Ker } g(A) = \{0\}$$

从而根据直和等价定义可知结论。

- $\text{Ker } f(A) \oplus \text{Ker } g(A) \subset \text{Ker } f(A)g(A)$

利用上学期知识，两个 A 的多项式可交换，再利用矩阵乘法结合律可知无论是 $f(A)x = 0$ 还是 $g(A)x = 0$ 都可以推出 $g(A)f(A)x = f(A)g(A)x = 0$ ，因此

$$\text{Ker } f(A) \subset \text{Ker } f(A)g(A), \quad \text{Ker } g(A) \subset \text{Ker } f(A)g(A)$$

再利用 $\text{Ker } f(A)g(A)$ 对加法封闭性可知它们的和空间必然仍然包含于 $\text{Ker } f(A)g(A)$ ，从而得证。

- $\text{Ker } f(A)g(A) \subset \text{Ker } f(A) \oplus \text{Ker } g(A)$

若 $f(A)g(A)x = 0$, 利用上方的

$$x = u(A)f(A)x + v(A)g(A)x$$

通过 A 的多项式可交换发现

$$g(A)(u(A)f(A)x) = u(A)f(A)g(A)x = 0, \quad f(A)(v(A)g(A)x) = v(A)f(A)g(A)x = 0$$

从而 $u(A)f(A)x \in \text{Ker } g(A)$ 、 $v(A)g(A)x \in \text{Ker } f(A)$, 即得证。

- 逆命题反例

考虑 $A = I$ 、 $f(x) = g(x) = x$, 则

$$\text{Ker } f(A) = \text{Ker } g(A) = \text{Ker } f(A)g(A) = \{0\}$$

从而确实满足直和性质, 但 $\gcd(f, g) \neq 1$ 。

* 可以发现, 不成立的主要原因是出现了使得 $A - \lambda I$ 可逆的因式 $x - \lambda$ 。由此, 若限定 $f(x)$ 、 $g(x)$ 的根都在 A 的特征值当中, 此时即能从 $\text{Ker } f$ 的直和性质推出 $\gcd(f, g) = 1$ 。

* 本题中可以看到, 多项式作为整体进行处理时, 裴蜀定理是处理最大公因式条件的最常用手段。

7. (a) 我们先想办法建立直观: 当 $n = 2$ 、 $r = 1$ 时, 这即是在说有限条直线不能盖住平面, 而 $n = 3$ 、 $r = 2$ 时即是在说有限个平面不能盖住三维空间。对此的证明想法也很自然: 设法取定一条直线, 使得其中每个 V_i 与直线至多一个交点, 而直线上的点又有无穷多个, 这就得到了不可能盖住全空间。为此, 我们对 s 进行归纳。当 $s = 1$ 时, 由于 $\dim V_1 < \dim V$, 自然有 $V_1 \neq V$, 从而可取出 v , 下面假设结论对任意 $s - 1$ 个维数为 r 的空间成立。

- 直线选取

由归纳假设, 存在 $\alpha \notin V_1 \cup V_2 \cup \dots \cup V_{s-1}$, 若 $\alpha \notin V_s$, 则其已经符合要求, 下设 $\alpha \in V_s$ 。由于 $\dim V_s < \dim V$, 也可取出 V 中的 β 使得 $\beta \notin V_s$ 。同样, 若 β 不在 V_1 到 V_{s-1} 的任何中, β 已经符合要求, 因此下设 $\beta \in V_1 \cup \dots \cup V_{s-1}$ 。

* 注意这里无法直接得到 $\alpha + \beta$ 符合要求, 大家可以思考 $s = 3$ 时的反例。

考虑直线

$$L = \{k\alpha + \beta \mid k \in \mathbb{K}\}$$

由于 $\alpha \notin V_1$, 其不可能为零向量, 因此不同的 k 对应 L 中的不同元素, 这的确是一条包含无穷多个点的“直线”, 下面只要说明其与任何 V_i 交点至多一个即可。

- 交点个数

首先, 若 L 中有某个 $\lambda\alpha + \beta \in V_s$, 则由 $\alpha \in V_s$ 可知 $\beta \in V_s$, 矛盾, 因此其与 V_s 无交点。对 $i = 1, \dots, s - 1$, 若 L 中有不同的两点 $k_1\alpha + \beta$ 、 $k_2\alpha + \beta$ 都在 V_i 中, 则作差可知 $(k_1 - k_2)\alpha \in V_i$, 但 $k_1 \neq k_2$, 矛盾,

由此, L 在 $V_1 \cup \dots \cup V_s$ 中的点至多只有 $s - 1$ 个, 但是其又包含无穷多个点, 从而一定能选取出不在 $V_1 \cup \dots \cup V_s$ 中的点, 这就是符合要求的 v 。

- (b) 对维数 r 进行反向归纳。当 $r = n$ 时, 所有 V_i 都是全空间, 从而 $\{0\}$ 是共同补空间。下面假设 $r < n$, 结论对 $r + 1$ 成立。

利用 (a), 取出 v 使得 $v \notin V_1 \cup \dots \cup V_s$, 并记 $U_i = V_i + \langle v \rangle$, 这里 $\langle v \rangle$ 代表 v 生成的子空间。由于 $v \notin V_i$, 考虑 V_i 的一组基, 可发现其不能表出 v , 因此利用上学期知识知这组基添加 v 后仍然线性无关, 而其又可以生成 U_i , 从而 V_i 的一组基添加 v 成为 U_i 的基, 这就说明了对任何 U_i 有

$$\dim U_i = r + 1$$

利用归纳假设，设所有 U_i 的共同补空间为 W ，记 $U = W + \langle v \rangle$ ，下面证明 $V = U \oplus V_i$ 对任何 i 成立，这就说明了结论。

由补空间性质可知

$$V = W \oplus U_i = W \oplus (V_i + \langle v \rangle)$$

而根据维数可知 V_i 与 $\langle v \rangle$ 的和是直和，从而

$$V = W \oplus (V_i \oplus \langle v \rangle)$$

利用本讲义 18.3.2 已经证明的直和的交换、结合律，即得

$$V = (W \oplus \langle v \rangle) \oplus V_i = U \oplus V_i$$

从而得证。

* 此题的证明较为技巧性，但结论可以代表某种本质的和空间性质，也与并集有根本不同。利用不断构造补空间可以发现，对 \mathbb{K} 上的 $2k$ 维线性空间 V ，可以构造任意有限个子空间 V_1, \dots, V_s ，使得两两互为补空间。

§18.4 商空间

18.4.1 基本结论

最后，我们来介绍一个高中阶段并没有学习过的构造结构方式，也就是商，自然，我们要从集合的商开始介绍。复习上学期等价关系的定义：集合 A 上的关系是指 $A \times A$ 的子集 R ，若 $(a, b) \in R$ 可记为 aRb ，称为 a 与 b 具有关系 R 。等价关系是指满足 aRa 、 aRb 则 bRa 、 aRb , bRc 则 aRc 的关系。

有了等价关系 R 以后，就可以定义集合对等价关系的商集

$$A/R = \{[a] \mid a \in A\}, \quad [a]_R = \{x \in A \mid xRa\}$$

这里的 $[a]_R$ 称为 a 在关系 R 下的等价类，利用等价关系定义可发现不同的等价类不交，所有等价类并集为 A ，且 aRb 当且仅当 a 、 b 属于同一等价类中，或 $[a]_R = [b]_R$ 。

也就是说，一个集合对等价关系的商集是指其所有等价类的集合（注意要求为集合，因此不进行重复计算）。商集与交集、并集、补集有一个显著差别，也就是 A 对于某等价关系的商集并不是 A 的子集。

* 一个简单的例子是，考虑 \mathbb{Z} 上用除以 m 余数相同定义等价关系（记为 \equiv_m ），则得到的商集一共有 m 个元素，分别为 $[0]_{\equiv_m}, [1]_{\equiv_m}, \dots, [m-1]_{\equiv_m}$ ，也即写成

$$A/\equiv_m = \{[0]_{\equiv_m}, [1]_{\equiv_m}, \dots, [m-1]_{\equiv_m}\}$$

由于线性空间是集合，给定线性空间上的一个等价关系，自然也可以构造出对应的商集。但是，这样得到的商集往往无法自然定义成线性空间。我们当然希望线性空间的商还是线性空间，因此我们只考虑一类特殊的等价关系：设 \mathbb{K} 上的线性空间 V 有子空间 W ，定义 V 上的关系 \sim_W 为， $a \sim_W b$ 当且仅当 $a - b \in W$ 。我们先证明这是一个等价关系。

证明：

由子空间定义 $0 \in W$ （只要 W 非空，对 $a \in W$ ，可知 $(-1)a = -a \in W$ ，从而 0 在 W 中），于是 $a \sim_W a$ ；若 $a \sim_W b$ ，则 $b - a = (-1)(a - b) \in W$ ，于是 $b \sim_W a$ ；若 $a \sim_W b$ 、 $b \sim_W c$ ，可知 $a - c = (a - b) + (b - c) \in W$ ，从而 $a \sim_W c$ 。

由此，商集 V/\sim_W 存在。为了方便，将其记为 V/W ，下面我们说明等价类集合 $[a]_{\sim_W}$ 可以写为

$$a + W = \{a + w \mid w \in W\}$$

证明：

由 $a - b \in W$ 可知 $b = a - w, w \in W$, 从而 $b = a + (-w) \in a + W$; 若 $b \in a + W$, 设 $b = a + w, w \in W$ 有 $a - b = -w \in W$ 。综合两部分得到 $[a]_{\sim_W} = a + W$ 。

* 上方的证明中说明了商空间最常用的结论, $a + W = b + W$ 等价于 $a - b \in W$ 或 $b - a \in W$ 。

为了让其成为一个线性空间, 我们必须定义合适的加法和数乘, 简单的想法是对任何 $a, b \in V, \lambda \in \mathbb{K}$ 定义

$$(a + W) + (b + W) = (a + b) + W$$

$$\lambda(a + W) = (\lambda a) + W$$

但是, 这样的定义必须验证合理性: 这里 a, b 本质是在等价类集合中取出了代表元, 我们必须证明代表元更换而等价类不变时对应的计算结果不变。

证明:

加法的定义合理性也即要证明, $a + W = c + W, b + W = d + W$ 时

$$(a + b) + W = (c + d) + W$$

利用等价类的定义, $a + W = c + W$ 也即 $c \in a + W$, 或写成 $c - a \in W$, 第二个条件即可写为 $d - b \in W$, 设 $w_1 = c - a, w_2 = d - b$, 则利用加法交换、结合律

$$(c + d) = (a + b) + (w_1 + w_2)$$

由于 $w_1 + w_2 \in W$, 即得到 $(c + d) - (a + b) \in W$, 从而 $(a + b) + W = (c + d) + W$ 。

数乘的定义合理性也即要证明, $a + W = c + W$ 时

$$(\lambda a) + W = (\lambda c) + W$$

同样将条件写为 $w = c - a \in W$, 则 $\lambda c - \lambda a = \lambda w \in W$, 即得证结论。

* 对于线性空间的八条性质验证相对直接 (通过 V 中加法、数乘的性质验证即可), 这里不再赘述, 不过注意商空间中的零元为 $0 + W = W$ 。

至此, 我们的确构造出了 V 对等价关系 \sim_W 的商空间, 或称为 V 对子空间 W 的商。就像我们需要计算交空间与和空间一样, 对商空间也需要考虑如何计算。同样先考虑 \mathbb{K}^n 的情况, 设子空间 W 的一组基为 $\alpha_1, \dots, \alpha_r$, 考虑 \mathbb{K}^n/W 。

解答:

在本讲义 18.3.2 中, 我们已经给出了 W 的补空间 U 一组基的计算方法, 设它们为 $\beta_1, \dots, \beta_{n-r}$ 。我们下面证明, $\beta_1 + W, \dots, \beta_{n-r} + W$ 构成 \mathbb{K}^n/W 的一组基, 即给出了一个有效的算法。

- 线性无关性

注意商空间中零元素为 W , 若 $\sum_{i=1}^{n-r} \lambda_i (\beta_i + W) = W$, 利用商空间中加法、数乘的定义可得

$$\sum_{i=1}^{n-r} \lambda_i \beta_i + W = W$$

从而根据等价类性质可知

$$\sum_{i=1}^{n-r} \lambda_i \beta_i \in W$$

而又由于 $\sum_{i=1}^{n-r} \lambda_i \beta_i \in U$, 利用补空间定义可知只能和为 0, 再由 β_i 的线性无关性可知只能所有 λ_i 全为 0。

– 生成全空间

对任何 $a + W$, $a \in V$, 利用补空间性质设其为 $u + w$, $u \in U$ 、 $w \in W$, 则

$$a + W = (u + w) + W = u + (w + W) = u + W$$

而利用基的定义 u 可以由 $\beta_1, \dots, \beta_{n-r}$ 表出, 从而根据商空间加法、数乘定义 $u + W$ 可以由 $\beta_1 + W, \dots, \beta_{n-r} + W$ 表出, 得证。

最后提供一点注释:

- 如之前的所有计算一样, 一般的商空间计算可以通过取定全空间一组基后根据坐标构造。此外, 如果是对线性方程组 $Ax = 0$ 的解空间的商空间 $\mathbb{K}^m / \text{Ker } A$, 利用上学期知识可知 A 的行空间与 $\text{Ker } A$ 互为补空间, 因此也可以用行向量极大线性无关组加 $\text{Ker } A$ 成为商空间一组基。
- V 、 W 维数有限时, 根据上方的构造与补空间的维数关系即可得到商空间维数公式

$$\dim(V/W) = \dim V - \dim W$$

不过, 即使 V 、 W 均为无穷维, V/W 也可能维数有限, 我们将在下面的习题中看到例子。这时, 上方证明中的补空间一组基加 W 成为商空间一组基仍然成立, 证明过程几乎不变。

- 虽然我们已经给出了商空间的定义, 但目前它的意义还不明确。粗略地理解, 定义商空间是为了合并元素, 从而可以方便之后第一同构定理时的同构区别。
- 商空间也可以对应空间分解, 我们大致写为 $V = (V/W) \times W$ (事实上等号应为同构), 这里 $A \times B$ 是指所有向量对 (a, b) 使得 $a \in A$ 、 $b \in B$, 定义加法 $(a, b) + (c, d) = (a+c, b+d)$ 、数乘 $\lambda(a, b) = (\lambda a, \lambda b)$ 后形成的线性空间。
- 比起补空间, 商空间的空间分解具有显著的优势, 也即其唯一性: 虽然补空间可以有很多种不同构造, 商空间是唯一确定的, 补空间的不同构造只是对应商空间不同的基。

18.4.2 例题

1. 对 $\mathbb{R}[x]$ 中 f 的倍式构成的线性空间 (f) , 给出 $\mathbb{R}[x]/(f)$ 的维数与一组基。
2. 若 U 、 W 是 \mathbb{K} 上的线性空间 V 的有限维子空间, 证明

$$\dim(U + W)/W = \dim U/(U \cap W)$$

3. 若 U 是 \mathbb{K} 上有限维线性空间 V 的子空间、 W 是 U 的子空间, 证明 (注意 U/W 是 V/W 的子集)

$$\dim(V/W)/(U/W) = \dim V/U$$

18.4.3 解答

1. 对 $f(x) = 0$ 的平凡情况, $(f) = \{0\}$, 而根据商空间定义可发现 $\mathbb{R}[x]/\{0\}$ 每个元素的等价类只有自身, 因此本质与 $\mathbb{R}[x]$ 相同, 维数无穷, 基为所有 $\{x^i\}, i \in \mathbb{N}$ (注意商空间中元素为集合)。

当 $f(x)$ 不为零多项式时, 设 $n = \deg f$, 我们下面证明 $\mathbb{R}[x]_n$ 是 (f) 的补空间, 从而即得 $\mathbb{R}[x]/(f)$ 维数为 n , 一组基为 $1 + (f), x + (f), \dots, x^{n-1} + (f)$ 。

由于 (f) 中任何非零多项式次数都至少为 n 次, $\mathbb{R}[x]_n \cap (f) = \{0\}$ 。另一方面, 由于任何多项式都可以写为 $q(x)f(x) + r(x)$, 且 $\deg r(x) < n$, 根据 $q(x)f(x) \in (f)$ 、 $r(x) \in \mathbb{R}[x]_n$ 即得 $\mathbb{R}[x]_n + (f) = \mathbb{R}[x]$ 。综合两者得证。

2. 直接利用和空间维数公式与商空间维数公式计算

$$\dim(U + W)/W = \dim(U + W) - \dim W = \dim U + \dim W - \dim U \cap W + \dim W$$

化简得其等于 $\dim U - \dim U \cap W$, 即为右侧的维数。

3. 直接利用商空间维数公式计算

$$\dim(V/W)/(U/W) = \dim(V/W) - \dim(U/W) = (\dim V - \dim W) - (\dim U - \dim W)$$

化简得其等于 $\dim V - \dim U$, 即为右侧的维数。

十九 线性映射

§19.1 习题解答

1. (丘书 习题 8.2.9) 在 \mathbb{K}^4 中, 设

$$V_1 = \langle \alpha_1, \alpha_2, \alpha_3 \rangle, \quad \alpha_1 = (1, 1, -1, 2)^T, \quad \alpha_2 = (2, -1, 3, 0)^T, \quad \alpha_3 = (0, -3, 5, -4)^T$$

$$V_2 = \langle \beta_1, \beta_2 \rangle, \quad \beta_1 = (1, 2, 2, 1)^T, \quad \beta_2 = (4, -3, 3, 1)^T$$

求 $V_1 + V_2$ 与 $V_1 \cap V_2$ 的一个基和维数。

解答:

算法可参考本讲义 18.3.1 或教材, $V_1 + V_2$ 的一组基可以取为 $\alpha_1, \alpha_2, \beta_1$, 维数为 3; $V_1 \cap V_2$ 的一组基可以取为 $\beta_1 + \beta_2$, 维数为 1 (基的取法不唯一)。

2. (丘书 习题 8.2.17) 设 A, B 是数域 \mathbb{F} 上 $s \times n, m \times n$ 矩阵, 证明关于 x 的齐次线性方程组 $Ax = 0$ 与 $Bx = 0$ 解集相同当且仅当 A, B 行向量组等价。

证明:

* 本题最直接的做法是利用上学期正交补空间的知识: 解空间与行向量组生成子空间(行空间)互为正交补, 而正交补唯一, 因此解空间相同当且仅当行空间相同, 从而当且仅当行向量组等价。

* 未学正交时的做法可见本讲义第六章复习题 7。

由上学期记号将 $Ax = 0$ 解集记为 $\text{Ker } A$, 根据线性方程组定义有

$$\text{Ker } C = \text{Ker } A \cap \text{Ker } B, \quad C = \begin{pmatrix} A \\ B \end{pmatrix}$$

由此 $\text{Ker } A = \text{Ker } B$ 当且仅当

$$\text{Ker } A = \text{Ker } C = \text{Ker } B$$

我们下面证明第一个等号成立当且仅当 B 的行向量可被 A 的行向量线性表出, 同理第二个等号成立当且仅当 A 的行向量可被 B 的行向量线性表出, 这就得到了结论。

若 B 的行向量可被 A 的行向量线性表出, 由上学期知识(设出表出式后直接构造)存在矩阵 P 使得 $B = PA$, 从而 $Ax = 0$ 可推出 $PAx = 0$, 于是 $\text{Ker } A \subset \text{Ker } C$, 而由 $Cx = 0$ 中包含 $Ax = 0$ 的方程, 有 $\text{Ker } C \subset \text{Ker } A$, 因此 $\text{Ker } C = \text{Ker } A$ 。

若 B 的某个行向量不能被 A 的行向量线性表出, 则 A 的行向量组极大线性无关组增添 B 的对应行后仍然线性无关, 从而 $\text{rank } C > \text{rank } A$, 利用解空间维数定理可得 $\text{Ker } C < \text{Ker } A$, 不可能相等, 从而 $\text{Ker } C = \text{Ker } A$ 可推出 B 的行向量能被 A 的行向量线性表出。

3. (丘书 习题 8.4.2) 设 W 是 \mathbb{K} 上 n 维线性空间 V 的非平凡子空间, 从 W 中取一个基 $\delta_1, \dots, \delta_m$, 考虑两种不同扩充方式使得

$$S = (\delta_1, \dots, \delta_m, \alpha_{m+1}, \dots, \alpha_n)$$

$$T = (\delta_1, \dots, \delta_m, \beta_{m+1}, \dots, \beta_n)$$

都是 V 的一组基, 且 S 到 T 的过渡矩阵为 P 。求 V/W 的两组基

$$S_W = (\alpha_{m+1} + W, \dots, \alpha_n + W), \quad T_W = (\beta_{m+1} + W, \dots, \beta_n + W)$$

的过渡矩阵。

解答:

由过渡矩阵定义 $T = SP$, 从而设 P 第 i 行第 j 列为 p_{ij} 有

$$\forall j = m+1, \dots, n, \quad \beta_j = \sum_{i=1}^m p_{ij} \delta_i + \sum_{i=m+1}^n p_{ij} \alpha_i$$

由于所有 δ_i 构成 W 的一组基, 可发现

$$\forall j = m+1, \dots, n, \quad \beta_j - \sum_{i=m+1}^n p_{ij} \alpha_i \in W$$

利用商空间等价类定义也即

$$\forall j = m+1, \dots, n, \quad \beta_j + W = \sum_{i=m+1}^n p_{ij} \alpha_i + W$$

再利用商空间加法、数乘运算的性质可知

$$\forall j = m+1, \dots, n, \quad \beta_j + W = \sum_{i=m+1}^n p_{ij} (\alpha_i + W)$$

这已经符合过渡矩阵的定义式, 由此根据形式得 S_W 到 T_W 的过渡矩阵为 P 的第 $m+1$ 到第 n 行/列交成的子矩阵 (或表述为 P 右下角的 $n-m$ 阶方阵)。

4. (丘书习题 8.4.3) 设 \mathbb{K} 上矩阵

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

(1) 求 $Ax = 0$ 解空间 W 的一组基。

解答:

直接利用线性方程组知识得到 $\{(1, 3, 1)^T\}$ 可作为一组基 (其他正确结果为此向量的非零倍数)。

(2) 求 \mathbb{K}^3/W 的维数与一组基。

解答:

算法可参考本讲义 18.4.1 或教材, 维数为 2, 基可以取为 $\{e_1 + W, e_2 + W\}$ 或 $\{\alpha_1^T + W, \alpha_2^T + W\}$ 等 (取法不唯一), 其中 α_1, α_2 为 A 的两个行向量。

5. (丘书习题 8.3.8) 对正整数 n , 令

$$\mathbb{Q}(\sqrt[n]{3}) = \{a_0 + a_1 \sqrt[n]{3} + \dots + a_{n-1} \sqrt[n]{3^{n-1}} \mid a_0, \dots, a_{n-1} \in \mathbb{Q}\}$$

对不同正整数 m, n , $\mathbb{Q}(\sqrt[m]{3})$ 与 $\mathbb{Q}(\sqrt[n]{3})$ 是否同构?

解答:

我们下面证明 $1, \sqrt[n]{3}, \dots, \sqrt[n]{3^{n-1}}$ 线性无关，从而 $\mathbb{Q}(\sqrt[n]{3})$ 的维数为 n ，而有限维线性空间同构当且仅当维数相同，即可得到不同构。

* 下面这部分证明完全无需掌握，并不在这学期考察范围内，布置这题确实意味不明……

记 $t = \sqrt[n]{3}$ 。若它们线性相关，也即存在不全为 0 的有理数 a_0, \dots, a_{n-1} 使得 $a_0 + a_1t + \dots + a_{n-1}t^{n-1} = 0$ ，也即存在至多 $n-1$ 次的非零有理系数多项式 $f(x)$ 使得 $f(t) = 0$ 。

记多项式 $g(x) = x^n - 3$ ，则 $g(t) = 0$ 。由于 $g(x), f(x)$ 都有 $x-t$ 作为因式，两者最大公因式至少为一次，而根据最大公因式不随数域变化而变化（辗转相除法计算的结论），设有理系数多项式 $d(x)$ 满足 $d(x) = \gcd(f(x), g(x))$ ，则有

$$1 = \deg(x-t) \leq \deg d(x) \leq \deg f(x) \leq n-1$$

可得

$$1 \leq \deg d(x) \leq n-1$$

下设

$$x^n - 3 = d(x)h(x)$$

为了说明上式不可能成立，我们定义一个有理数 q 为 3 的“倍数”当且仅当化为最简分数后分子是 3 的倍数。有两个容易计算得到的结论：乘积为 3 的“倍数”的两个有理数至少有一个是 3 的“倍数”，反之亦然；3 的“倍数”和不是 3 的“倍数”的两数之和不是 3 的“倍数”（通分可证）。

由于 x^n 前系数为 1，通过 $d(x), h(x)$ 乘适当倍数可设 $d(x), h(x)$ 首项系数均为 1，从而可设

$$d(x) = a_0 + a_1x + \dots + a_{m-1}x^{m-1} + x^m, \quad h(x) = b_0 + b_1x + \dots + b_{n-m-1}x^{n-m-1} + x^{n-m}$$

由于 $a_0b_0 = 3$ ， a_0, b_0 中必然有一个是 3 的“倍数”，且若两个都是 3 的“倍数”，可发现乘积必然是 9 的“倍数”，但 3 不是 9 的“倍数”，矛盾，从而必然一个是 3 的“倍数”一个不是 3 的“倍数”。

设 a_0 是 3 的“倍数”而 b_0 不是。由于一次项 $0 = a_0b_1 + a_1b_0$ 是 3 的“倍数”，且 a_0b_1 是 3 的“倍数”，若 a_1 不是 3 的“倍数”，则 a_1b_0 不是 3 的“倍数”，从而两者求和不是 3 的“倍数”，矛盾，因此必须 a_1 是 3 的“倍数”。类似归纳，考察到第 m 次项（注意 $m < n$ ，由此乘积 1 到 m 次项均为 0）可发现， a_0 到 a_{m-1} 都是 3 的“倍数”，但 m 次项满足（记 $b_{n-m} = 1$ ，更高次项系数为 0）

$$0 = a_0b_m + \dots + a_{m-1}b_1 + b_0$$

从而左侧为 3 的“倍数”，右侧为 3 的“倍数”与 b_0 之和，不为 3 的“倍数”，矛盾。同理，若 b_0 是 3 的“倍数”而 a_0 不是，由于 $n-m < n$ ，考察到第 $n-m$ 次项可由 b_0 到 b_{n-m-1} 都是 3 的“倍数”推出矛盾。

* 上述方法可以推广为判别有理系数多项式不可约性的艾森斯坦因判别法。

6. (丘书习题 8.3.10) 设 n 阶循环移位方阵 $A = (e_n, e_1, \dots, e_{n-1})$ (e_i 为第 i 个单位向量)，求 $\mathcal{C}(A)$ 的维数与一组基，这里 $\mathcal{C}(A)$ 表示与 A 可交换的矩阵构成的子空间。

解答：

一个直接计算的方法：直接考虑方程组 $AX = XA$ ，可发现与 A 可交换当且仅当 $x_{ij} = x_{i+1,j+1}$ （并将 $n+1$ 视作 1）对任何 i, j 成立。由此简单分析可发现 X 的所有元素分成了相等的 n 组，于是其维数为 n ，一组基为 (E_{ij} 表示第 i 行第 j 列位置为 1，其他为 0 的矩阵>)

$$\{E_{1i} + E_{2,i+1} + \dots + E_{n-i+1,n} + E_{n-i,1} + \dots + E_{n,i-1} \mid i = 1, \dots, n\}$$

进一步分析可发现上述取 i 的矩阵恰好为 A^{i-1} , 由此一组基可写为

$$\{I, A, \dots, A^{n-1}\}$$

* 本质上这是由于 A 可对角化且特征值互不相同, 我们将在本讲义 20.2.2 的例题中对可对角化时的一般情况进行分析。更一般的情况见期中复习题。

7. (丘书 习题 8.3.11) 设 A, B 是特征多项式相同的实对称矩阵, 证明存在 \mathbb{R}^n 到自身的同构映射 σ 使得

$$\forall \alpha \in \mathbb{R}^n, \quad \sigma(\alpha)^T B \sigma(\alpha) = \alpha^T A \alpha$$

证明:

由本章讨论将得到, \mathbb{R}^n 到自身的同构映射当且仅当其是 $\sigma(\alpha) = P\alpha$, 其中 P 可逆。

由于 A, B 都是实对称阵, 特征多项式相同意味着特征值相同, 于是正交相似标准形相同, 从而它们正交相似, 也即存在正交阵 P 使得 $P^T B P = A$, 取 $\sigma(\alpha) = P\alpha$ 即可验证成立。

8. (丘书 习题 9.2.1) 已知 $\mathbb{K}^4 \rightarrow \mathbb{K}^5$ 的映射 \mathcal{A}

$$\mathcal{A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 - 3x_2 + x_3 - 2x_4 \\ 2x_1 + x_2 - x_3 + 3x_4 \\ -x_1 + 10x_2 - 4x_3 + 9x_4 \\ 3x_1 - 2x_2 + x_4 \\ 4x_1 + 9x_2 - 5x_3 + 13x_4 \end{pmatrix}$$

判断其是不是线性映射, 如果是, 求 $\text{Im } \mathcal{A}, \text{Ker } \mathcal{A}, \text{CoKer } \mathcal{A} = \mathbb{K}^5 / \text{Im } \mathcal{A}$ 。

解答:

直接验证可知其是线性映射。由于它即为左乘矩阵

$$\begin{pmatrix} 1 & -3 & 1 & -2 \\ 2 & 1 & -1 & 3 \\ -1 & 10 & -4 & 9 \\ 3 & -2 & 0 & 1 \\ 4 & 9 & -5 & 13 \end{pmatrix}$$

以下基的构造均不唯一。计算列向量极大线性无关组可知 $\text{Im } \mathcal{A}$ 一组基可取为

$$(1, 2, -1, 3, 4)^T, \quad (-3, 1, 10, -2, 9)^T$$

求解线性方程组可知 $\text{Ker } \mathcal{A}$ 一组基可取为

$$(2, 3, 7, 0)^T, \quad (1, 1, 0, -1)^T$$

由本讲义 18.4.1 算法或教材可知 $\text{CoKer } \mathcal{A}$ 一组基可取为

$$e_1 + \text{Im } \mathcal{A}, \quad e_2 + \text{Im } \mathcal{A}, \quad e_3 + \text{Im } \mathcal{A}$$

9. (丘书 习题 9.2.4) 求习题 9.2.1 中的映射 \mathcal{A} 对应的 $\mathbb{K}^4 / \text{Ker } \mathcal{A}$ 的一组基。

解答:

由本讲义 18.4.1 算法或教材可知 $\mathbb{K}^4 / \text{Ker } \mathcal{A}$ 一组基可取为 (取法不唯一)

$$e_1 + \text{Ker } \mathcal{A}, \quad e_2 + \text{Ker } \mathcal{A}$$

10. (丘书习题 9.2.8) 设 \mathcal{A} 是 \mathbb{K} 上 n 维线性空间 V 的线性变换, 证明存在正整数 m 使得

$$\mathcal{A}^m(V) = \mathcal{A}^{m+k}(V), \quad k \in \mathbb{N}$$

证明:

直接利用定义可知

$$\mathcal{A}^{s+t}(V) = \mathcal{A}^s(\mathcal{A}^t(V)) \subset \mathcal{A}^s(V)$$

由此记 $V_i = \mathcal{A}^i(V)$, 每个 V_i 包含在之前的 V_i 中, 因此 $i < j$ 时

$$V_j \subset V_i, \quad \dim V_i \geq \dim V_j$$

由于 $\dim V_i$ 的取值只能在 0 到 n 的整数中, $\dim V_i > \dim V_{i+1}$ 至多只能发生 n 次, 从而最后一次发生后 (设 $\dim V_{m-1} > \dim V_m$, 此后不减) 即有 $\dim V_i$ 不再变化, 而根据包含关系, $\dim V_i = \dim V_j$ 且 $i < j$ 即得 $V_i = V_j$, 从而满足题目要求。

* 一个更强的版本是, 只要 $\dim V_m = \dim V_{m+1}$, 即有 $V_m = V_{m+k}$, 这是因为由包含关系有 $V_m = V_{m+1}$, 且

$$V_{m+2} = \mathcal{A}(V_{m+1}) = \mathcal{A}(V_m) = V_{m+1} = V_m$$

再归纳可得证。由此可以得到取 $m = n + 1$ 必然符合要求 (至多下降 n 次后恒等)。

* 进一步加强, 若 $\dim V_1 = n$, 则 $\mathcal{A}(V) = V$, 从而可得 $\dim V_i = n$ 恒成立, 可取 $m = 1$ 。否则至多下降 $n - 1$ 次后恒等, 取 $m = n$ 必然符合要求。综合可得取 $m = n$ 必然符合要求。

* 设 V 的一组基为 $\alpha_1, \dots, \alpha_n$, 并定义 $\mathcal{A}(\alpha_1) = 0$, $i > 1$ 时 $\mathcal{A}(\alpha_i) = \alpha_{i-1}$, 可发现 $\mathcal{A}(V)$ 到 $\mathcal{A}^{n-1}(V)$ 互不相同, 因此不可能取到比 n 更小的 m 保证成立。

* 此结论的矩阵版本即为存在正整数 m 使得 $\text{rank } A^m = \text{rank } A^{m+k}$, $k \in \mathbb{N}$ 。

§19.2 同态与同构

19.2.1 集合的同态

本节的主要目的是考虑两个一般线性空间之间的“连接”: 线性映射。为了理解它的定义动机, 进行之后的分析, 我们必须先从高中集合论中就学过的, 大家更为熟悉的映射谈起。我们将从一个很基本的问题开始, 为什么要引入映射的概念?

我们就以两个集合 $\{a, b, c, d, e\}$ 与 $\{1, 2, 3, 4, 5\}$ 为例。我们会认为这两个集合本质上具有某种结构相同的特性。具体来说, 我们可以用 x_1, x_2, x_3, x_4, x_5 去称呼 a, b, c, d, e 或 $1, 2, 3, 4, 5$, 无论进行哪一种称呼, 都有

$$\{x_1, x_2, x_3\} \cup \{x_2, x_4\} = \{x_1, x_2, x_3, x_4\}, \quad \{x_1, x_5\} \cap \{x_2, x_5\} = \{x_5\}$$

等等。在集合的角度, 我们称这样的两个集合同构, 可以看出, 要证明两个集合同构, 我们只要找到一种合适的对应关系。而粗略理解下, 一般的映射即是一种结构的对应关系。

不过, 到底什么是结构与结构相同? 在数学上, 一个结构往往是一个对象与它所允许的操作构成的。具体来说, 对于一个集合, 我们允许对它的子集计算交、并、补, 那么这些运算的结果都是结构的一部分。因此, 当我们在说一个映射部分保持结构时, 事实上指的是映射可以部分保持交、并、补运算。我们下面假设 $f : X \rightarrow Y$ 是映射, A, B 是 X 的子集, 用 $f(A)$ 表示 $\{y \in Y \mid y = f(a), a \in A\}$, 称为 A 在 f 下的像, 则有如下两个结论刻画结构的对应关系:

$$f(A \cap B) \subset f(A) \cap f(B)$$

$$f(A \cup B) = f(A) \cup f(B)$$

证明:

对第一个结论，由定义可知

$$f(A \cap B) = \{y \in Y \mid y = f(a), a \in A \cap B\}$$

由于 $a \in A \cap B$ 满足 $a \in A$ 且 $a \in B$ ，根据定义 $f(A \cap B) \subset f(A)$ 且 $f(A \cap B) \subset f(B)$ ，从而成立。

对第二个结论，由定义可知

$$f(A \cap B) = \{y \in Y \mid y = f(a), a \in A \cup B\}$$

而满足 $y = f(a)$, $a \in A \cup B$ 的 y 一定满足 $y = f(a)$, $a \in A$ 或 $y = f(a)$, $a \in B$ 。反之，无论是 $y = f(a)$, $a \in A$ 还是 $y = f(a)$, $a \in B$ 都可以得到 $y = f(a)$, $a \in A \cup B$ 。综合两者可知集合相等。

而所谓的两个集合 X 、 Y 同构（事实上可以称为等势），是指存在映射 $f : X \rightarrow Y$ 满足对任何 X 的子集 A 、 B 有（上标 c 为补集，由此即看出结构相同即为存在保持交、并、补的对应关系）

$$f(A \cap B) = f(A) \cap f(B)$$

$$f(A \cup B) = f(A) \cup f(B)$$

$$f(A^c) = f(A)^c$$

我们将满足这三个条件的映射称为集合间的同构映射，或称为双射。有如下重要的判定定理： $f : X \rightarrow Y$ 是双射当且仅当 $f(X) = Y$ 且 $f(a) = f(b) \Rightarrow a = b$ 。

证明：

– 若 f 为双射，取 $A = \emptyset$ 可得 $f(X) = f(\emptyset^c) = f(\emptyset)^c = Y$ ，这里利用了空集的像集是空集，这可以由定义直接得到。

另一方面，由于当 $a \neq b$ 时 $\emptyset = f(\{a\} \cap \{b\}) = f(\{a\}) \cap f(\{b\}) = \{f(a)\} \cap \{f(b)\}$ ，即得到 $f(a) \neq f(b)$ 。

– 若 $f(X) = Y$ 且 $f(a) = f(b) \Rightarrow a = b$ ，下面验证 f 是双射。利用一般映射的性质，只需证明 $f(A) \cap f(B) \subset f(A \cap B)$ 且 $f(A^c) = f(A)^c$ 。

对第一条，假设存在 A 、 B 使得包含关系不成立，也即有元素在 $f(A) \cap f(B)$ 中，但不在 $f(A \cap B)$ 中。设其为 y ，存在 $a \in A$ 、 $b \in B$ 使得 $f(a) = f(b) = y$ ，但由于不存在 $A \cap B$ 中元素的像为 y ， $a \neq b$ ，从而与第二个条件矛盾。

对第二条，由于第一条已经成立，利用 $f(A) \cap f(A^c) = f(A \cap A^c) = \emptyset$ 可知 $f(A^c) \subset f(A)^c$ 。再由 $f(X) = Y$ ， $f(A) \cup f(A^c) = f(A \cup A^c) = Y$ ，因此即得到 $f(A)^c \subset f(A^c)$ 。综合两者得证。

* 非常建议大家在看这一段对映射的说明时自己举有限集合的例子观察。

* 利用双射的定义可直接验证双射的复合还是双射。

由上方的讨论，我们已经可以看出映射的意义了：保持结构就是保持允许的操作，这里为交、并、补，而双射——也可以称为集合间的同构——代表完全保持结构，一般的映射即代表部分保持结构。

我们将满足 $f(X) = Y$ 的映射称为满射，满足 $f(a) = f(b) \Rightarrow a = b$ 的映射称为单射。直观上理解，满射代表左侧的结构可以覆盖在右侧的结构上，某种意义上是左侧的结构比右侧的结构大，单射则代表左侧的结构可以嵌入右侧的结构，某种意义上是左侧的结构比右侧的结构小，根据上方证明的判定定理，双射等价于满射加单射，也就是说，如果左边的结构既比右边大，也比右边小，则两者结构相同。

* 上述说明中隐含了集合论中经典的康托伯恩斯坦定理，也即 X 到 Y 存在单射与满射则存在双射。这样的结论对一般的结构未必正确，但至少对集合、线性空间都是成立的。

我们给上面的大小关系一个直观的感受：当 X 、 Y 元素个数均有限时，用 $|X|$ 记 X 的元素个数，则 X 到 Y 存在单射当且仅当 $|X| \leq |Y|$ ，存在满射当且仅当 $|X| \geq |Y|$ ，存在双射当且仅当 $|X| = |Y|$ 。

证明：

设 $X = \{x_1, \dots, x_n\}$, $Y = \{y_1, \dots, y_m\}$, 则 $|X| = n$ 、 $|Y| = m$ 。

- 单射

若 X 到 Y 存在单射 f , 设 $f(x_i) = y_{k_i}$, 由单射定义, k_1 到 k_n 互不相同, 而它们又都在 1 到 m 中, 从而 $n \leq m$ 。

若 $n \leq m$, 考虑映射 $f(x_i) = y_i$, 其中 $i = 1, \dots, n$, 即可验证其为单射。

- 满射

若 X 到 Y 存在满射 g , 对任何 y_i 都存在 x_{k_i} (未必唯一, 取出其中一个) 使得 $g(x_{k_i}) = y_i$ 。由映射的定义同一个元素不可能映到两个不同元素, 因此 K_1 到 k_m 互不相同, 而它们又都在 1 到 n 中, 从而 $m \leq n$ 。

若 $m \leq n$, 考虑映射 $f(x_i) = y_i$, 其中 $i = 1, \dots, n$, 下标大于 n 的 x_i 都映射到 y_1 , 即可验证其为满射。

- 双射

若 X 到 Y 存在双射 h , 由于双射既是单射又是满射, 从上面两部分证明中可知 $m \geq n$ 且 $m \leq n$, 从而 $m = n$ 。

若 $m = n$, 考虑映射 $f(x_i) = y_i$, 其中 $i = 1, \dots, n$, 即可验证其为双射。

这样, 有限集合时的单射对应结构更小、满射对应结构更大、双射对应同构即得到了一个很好的展示。此结论将在之后讨论有限维线性空间的单同态、满同态、线性同构时用到。

接下来, 我们定义一个映射 $f : X \rightarrow Y$ 的逆映射 $g : Y \rightarrow X$ 为满足

$$\forall x \in X, \quad g(f(x)) = x$$

$$\forall y \in Y, \quad f(g(y)) = y$$

的映射 g , 记作 $g = f^{-1}$ 。我们可以证明, 一个映射是双射当且仅当它的逆映射存在。

证明:

若 f 是双射, 对任何 $y \in Y$, 考虑 $\{x \mid f(x) = y\}$, 由 $f(X) = Y$ 知其非空, 由 $f(a) = f(b) \Rightarrow a = b$ 可知其中最多只有一个元素, 因此可定义 $g(y)$ 为 $\{x \mid f(x) = y\}$ 的唯一元素 x 。根据定义可发现 $f(g(y)) = y$ 对任何 y 成立。对任何 $x \in X$, 代入 $y = f(x)$ 可知 $f(g(f(x))) = f(x)$, 再利用 f 为单射即得 $g(f(x)) = x$, 从而为逆映射。

若 f 存在逆映射 g , $\forall y \in Y$, 取 $x = g(y)$ 即得 $f(x) = y$, 从而 $f(X) = Y$ 。另一方面, 若 $f(a) = f(b)$, 有 $a = g(f(a)) = g(f(b)) = b$, 从而 f 为单射。综合两部分得 f 为双射。

* 某种意义上, 这是同构更本质的定义——映射存在说明 X 的结构可以复制到 Y 的结构, 逆映射存在则说明 Y 的结构可以复制到 X 的结构, 且这两种复制都是无损的 (可以还原为恒等映射)。利用双射等价于逆映射存在, 而逆映射是相互的, 可知双射的逆映射是双射。

最后, 我们可以证明集合的同构 (存在双射) 有如下性质:

- A 到 A 存在双射;
- 若 A 到 B 存在双射, 则 B 到 A 存在双射;
- 若 A 到 B 、 B 到 C 存在双射, 则 A 到 C 存在双射。

证明:

对第一个性质, 取 $\forall a \in A, f(a) = a$ 即得到符合要求的映射。

对第二个性质, 设 $f : A \rightarrow B$ 为双射, 已经证明 $f^{-1} : B \rightarrow A$ 也是双射, 从而成立。

对第三个性质, 设 $f : A \rightarrow B$ 、 $g : B \rightarrow C$ 是双射, 已经证明复合 $g \circ f : A \rightarrow C$ 也是双射, 从而成立。

可以发现，利用这三个性质，同构是集合上的等价关系（这里的等价关系与之前定义的有一个细微的差别，将在之后讨论），因此可以用同构划分出集合间的等价类。例如，根据之前已证，对于有限集合而言，等价类即可由元素个数确定。

* 还有一个相对抽象的映射的性质我们此处没有讨论，也即映射复合具有结合律，当 $f : A \rightarrow B, g : B \rightarrow C, h : C \rightarrow D$ 都为映射时，有

$$(h \circ g) \circ f = h \circ (g \circ f)$$

这里映射复合的定义是 $(g \circ f)(x) = g(f(x))$ ，由此可直接验证，某种意义上说，所有有结合律的地方都有映射存在，例如我们马上就要看到矩阵乘法的结合律对应线性映射的结合律。

19.2.2 线性映射

在讨论完两个集合之间结构的对应关系后，我们自然就希望刻画两个线性空间之间结构的对应关系。首先，线性空间至少是一个集合，因此对应关系至少是一个映射。此外，线性空间除了具有集合的所有结构外，还可以进行加法、和数乘，因此我们希望结构的对应关系也能保持这两点。具体来说，若 \mathbb{K} 上的线性空间 U, V 之间的映射 f 满足

$$\forall u_1, u_2 \in U, \quad f(u_1 + u_2) = f(u_1) + f(u_2)$$

$$\forall u \in U, \lambda \in \mathbb{K}, \quad f(\lambda u) = \lambda f(u)$$

则称它为一个线性映射，或线性同态（同态大致意味着保持部分结构的对应关系）。与映射时类似，若一个线性同态是单射则称为单同态，若一个线性同态是满射则称为满同态。

- 定义要求线性映射是在同一个数域上的线性空间讨论的，这是因为不同数域的线性空间不可能构造数乘的对应。
- 注意定义中括号内的加法、数乘是 U 上的加法、数乘，括号外则是 V 上的加法、数乘，对应的运算可能不同。
- 线性映射的例子可见本讲义 17.4.2。

不过，与单同态、满同态的定义略有区别的是，我们定义一个线性映射是同构（称为线性同构）当且仅当它是双射且逆映射也是线性映射。之所以要添加逆映射也是线性映射的要求，是为了让同构具有某种等价性：若 U 与 V 同构，则 V 与 U 也同构（逆映射作为同构映射）。幸运的是，我们将证明线性映射是双射则逆映射线性，由此，一个线性映射是线性同构当且仅当其为双射，无需再对逆进行单独验证。

证明：

设 U, V 为 \mathbb{K} 上线性空间， $f : U \rightarrow V$ 为可逆线性映射，下面证明 f^{-1} 是线性映射。

由 f 为线性映射且逆映射存在，在 $f(x+y) = f(x) + f(y)$ 两侧取逆可知

$$\forall x, y \in U, \quad x + y = f^{-1}(f(x) + f(y))$$

也即

$$f^{-1}(f(x)) + f^{-1}(f(y)) = f^{-1}(f(x) + f(y))$$

由 f 可逆，其为满射，因此对任何 $\alpha, \beta \in V$ ，存在 x, y 使得 $f(x) = \alpha, f(y) = \beta$ ，于是（事实上可直接代入 $x = f^{-1}(\alpha), y = f^{-1}(\beta)$ ）

$$\forall \alpha, \beta \in V, \quad f^{-1}(\alpha) + f^{-1}(\beta) = f^{-1}(\alpha + \beta)$$

同理，在 $f(\lambda x) = \lambda f(x)$ 两侧取逆可知

$$\forall x \in U, \lambda \in \mathbb{K} \quad \lambda f^{-1}(f(x)) = f^{-1}(\lambda f(x))$$

再由满射，对任何 α 取 $x = f^{-1}(\alpha)$ 得

$$\forall \alpha \in V, \lambda \in \mathbb{K} \quad \lambda f^{-1}(\alpha) = f^{-1}(\lambda\alpha)$$

综合得到的两个式子可知 f^{-1} 是线性映射。

此外，我们还需要证明线性映射的复合是线性映射，设 U, V, W 是 \mathbb{K} 上的线性空间， $f : U \rightarrow V, g : V \rightarrow W$ 都是线性映射，我们下面证明 $g \circ f$ 也是线性映射。

证明：

直接由线性性计算

$$\begin{aligned} g(f(x+y)) &= g(f(x)+f(y)) = g(f(x))+g(f(y)) \\ g(f(\lambda x)) &= g(\lambda f(x)) = \lambda g(f(x)) \end{aligned}$$

从而结论成立。

我们再对线性映射保持结构的性质作一点验证，大家可以对比这里的结论与上一部分结论的相似之处。仿照映射的性质 $f(A \cup B) = f(A) \cup f(B), f(A \cap B) \subset f(A) \cap f(B)$ ，我们有：若 U, V 是 \mathbb{K} 上的线性空间， $f : U \rightarrow V$ 是线性映射，则对 U 的子空间 W 有 $f(W)$ 是 V 的子空间，且对 U 的子空间 W_1, W_2 有

$$\begin{aligned} f(W_1 \cap W_2) &\subset f(W_1) \cap f(W_2) \\ f(W_1 + W_2) &= f(W_1) + f(W_2) \end{aligned}$$

证明：

- $f(W)$ 是子空间

只需验证封闭性。若 $x, y \in f(W)$ ，也即存在 $w_1, w_2 \in W$ 使得 $f(w_1) = x, f(w_2) = y$ ，从而由 W 封闭性

$$x + y = f(w_1) + f(w_2) = f(w_1 + w_2) \in W$$

同理，对 $\lambda \in \mathbb{K}$ 有

$$\lambda x = \lambda f(w_1) = f(\lambda w_1) \in W$$

于是结论成立。

- 交空间性质

由于这里交空间即定义为交集，利用上一部分证明的映射的性质即得成立。

- 和空间性质

由定义

$$f(W_1 + W_2) = \{f(w) \mid w \in W_1 + W_2\} = \{f(w_1 + w_2) \mid w_1 \in W_1, w_2 \in W_2\}$$

而

$$f(W_1) + f(W_2) = \{f(w_1) + f(w_2) \mid w_1 \in W_1, w_2 \in W_2\} = \{f(w_1 + w_2) \mid w_1 \in W_1, w_2 \in W_2\}$$

从而两者相等。

除此以外，若对 U 的任何子空间 W_1, W_2 与子空间 W 的补空间 W_0 有

$$f(W_1 \cap W_2) = f(W_1) \cap f(W_2)$$

$$f(W_1 + W_2) = f(W_1) + f(W_2)$$

$$f(W_0) \oplus f(W) = V$$

且 V 不为 $\{0\}$ ，则 f 是一个线性同构，反之，若 f 是线性同构，则上面三条也成立。

证明：

– 若三条性质成立, 取 $W_0 = \{0\}$, 其补空间只能是 U , 从而 $f(U) = V$, 为满射。

若 f 不为单射, 即存在 $f(a) = f(b)$ 、 $a \neq b$, 则由线性性 $f(a-b) = 0$, $a-b \neq 0$, 记 $u = a-b$, 下面证明 $f(x) = 0$ 对任何 $x \in U$ 成立, 这就说明 $f(U) = \{0\}$, 结合满射得 $V = \{0\}$, 矛盾, 从而结论成立。

首先, 若 x 能被 u 表出, 由线性性结论成立, 否则 x 、 u 线性无关, 而这就能说明 x 、 $x+u$ 线性无关, 从而

$$\{0\} = f(\langle x \rangle \cap \langle x+u \rangle) = f(\langle x \rangle) \cap f(\langle x+u \rangle)$$

但是, 由于 $f(u) = 0$, $f(x) = f(x+u)$, 因此若 $f(x) \neq 0$ 则矛盾, 于是得证。

– 若 f 是线性同构, 由其为双射, 根据映射时已经证明的结论, 第一条对任何子集都成立, 对子空间也自然成立。

已证第二条对任何线性映射都成立, 因此 f 为同构时成立。

对于第三条, 若有非零的 $a \in f(W_0) \cap f(W)$, 由于 $W_0 \cap W = \{0\}$, 存在 W_0 、 W 中的不同元素像都为 a , 矛盾, 从而 $f(W_0)$ 、 $f(W)$ 为直和。再由满射与第二条可知

$$V = f(U) = f(W_0 + W) = f(W_0) + f(W)$$

从而成立。

* 这里的结论比起映射时稍弱, 需要额外要求 $V \neq \{0\}$, 是因为第一条性质只要求对交空间成立, 而非一般交集, 因此无法得到单射。

定义了单同态、满同态、线性同构 (下无歧义时简称同构, 两个线性空间之间存在线性同构则称这两个线性空间同构, 可用 \cong 表示), 并验证了线性映射的复合与逆都一定是线性映射后, 线性映射确实是一个合理的保持结构的对应关系, 上面的证明中也可以看到, 有些结论与映射完全类似。不过, 线性空间和集合的最大区别是, 整个空间可以用基的线性组合来进行表示, 由此, 我们需要给出一些基相关的结论。以下假设 U 、 V 都是 \mathbb{K} 上的线性空间 (并假设 U 的基存在), f 是 U 到 V 的线性映射, 我们来研究单同态、满同态与同构如何用基来表示。接下里证明过程中, 我们将默认无穷维线性空间中在元素基下的线性表示仍然具有唯一性, 这种唯一性具体来说是每个基前的系数唯一 (注意至多有限多个系数非零), 证明与有限维时类似。有如下结论:

1. f 是单同态、 U 的任意一组基的像在 V 中线性无关、存在 U 的一组基的像在 V 中线性无关, 三者等价。

证明:

由于第二条可以推出第三条, 我们只需证明 1 推 2 与 3 推 1。

– 1 推 2

反证。设 U 的一组基为 $\{\alpha_i \mid i \in I\}$, 若其像线性相关, 存在 i_1, \dots, i_n 与不全为 0 的 $\lambda_1, \dots, \lambda_n$ 使得

$$\lambda_1 f(\alpha_{i_1}) + \lambda_2 f(\alpha_{i_2}) + \cdots + \lambda_n f(\alpha_{i_n}) = 0$$

而这就得到

$$f(\lambda_1 \alpha_{i_1} + \cdots + \lambda_n \alpha_{i_n}) = 0$$

由基的线性无关性, $\lambda_1 \alpha_{i_1} + \cdots + \lambda_n \alpha_{i_n}$ 非零, 但 $f(\lambda_1 \alpha_{i_1} + \cdots + \lambda_n \alpha_{i_n}) = f(0)$, 与单射矛盾。

– 3 推 1

设 U 的基 $\{\alpha_i \mid i \in I\}$ 的像在 V 中线性无关。若 f 不是单射，存在 $a \neq b$, $f(a) = f(b)$, 也即 $f(a - b) = 0$, 且 $a - b \neq 0$ 。设 $a - b$ 的为

$$\sum_{k=1}^n \lambda_i \alpha_{i_k}$$

则由 $f(a - b) = 0$ 与线性性可得

$$\sum_{k=1}^n \lambda_i f(\alpha_{i_k}) = 0$$

由于 $a - b \neq 0$, 所有 λ_i 不全为 0, 与所有 $f(\alpha_{i_k})$ 线性无关矛盾。

2. f 是满同态、 U 的任意一组基的像可以生成 V 、存在 U 的一组基的像可以生成 V , 三者等价。

证明:

由于第二条可以推出第三条, 我们只需证明 1 推 2 与 3 推 1。

- 1 推 2

设 U 的一组基为 $\{\alpha_i \mid i \in I\}$, 对任何 $v \in V$, 由于其有原像 u , 设

$$u = \sum_{k=1}^n \lambda_i \alpha_{i_k}$$

则利用 $f(u) = v$ 与线性性可得

$$\sum_{k=1}^n \lambda_i f(\alpha_{i_k}) = v$$

也即这组基的像可以生成 v 。从而结论成立。

- 3 推 1

设 U 的基 $\{\alpha_i \mid i \in I\}$ 的像可以生成 V , 也即对任何 $v \in V$ 存在 $\lambda_1, \dots, \lambda_n$ 使得

$$v = \sum_{k=1}^n \lambda_k f(\alpha_{i_k})$$

从而利用线性性, 设 $u = \sum_{k=1}^n \lambda_k \alpha_{i_k}$, 即得 $f(u) = v$, 从而 f 为满射。

3. f 是线性同构、 U 的任意一组基的像是 V 的一组基、存在 U 的一组基的像是 V 的一组基, 三者等价。

证明:

由于第二条可以推出第三条, 我们只需证明 1 推 2 与 3 推 1, 而这直接由前两部分证明可以得到: f 是线性同构可得其为单射、满射, 从而 U 任意一组基的像是能生成 V 的线性无关向量组, 即 V 一组基; 存在 U 的一组基的像是 V 的一组基, 即线性无关且可以生成 V , 从而 f 是单射、满射, 即其是线性同构。

4. 给定 U 的一组基, 任意指定它们的像都可以得到唯一一个线性映射 f 。我们接下来将称这种定义方式为用基映射定义线性映射。

证明:

假设 U 的一组基为 $\{\alpha_i \mid i \in I\}$, 且指定像为 $f(\alpha_i) = \beta_i \in V$ 。

- 存在性

由于对任何 $u \in U$, 利用基的定义可以唯一写成 (若 $u = 0$ 对应 $N = 0$ 的情况, 否则要求所有 λ_i 非零)

$$u = \sum_{k=1}^N \lambda_k \alpha_{j_k}$$

这里 $j_k \in I$ 。由于被基表出的存在唯一性，这确实对 f 在任何一点处的值给出了唯一定义。
由此可以定义 $f(0) = 0$ ，其他情况

$$f(u) = \sum_{k=1}^N \lambda_k f(\alpha_{j_k}) = \sum_{k=1}^N \lambda_k \beta_{j_k}$$

由于 α_i 的表示即为 $1\alpha_i$ ， f 的确满足 $f(\alpha_i) = \beta_i$ ，下面证明其为线性映射。

对 $u_1, u_2 \in U$ ，设（下方表出可以有系数为 0，这样能让 u_1, u_2 选取相同的一些基）

$$u_1 = \sum_{k=1}^n \mu_k \alpha_{i_k}, \quad u_2 = \sum_{k=1}^n \gamma_k \alpha_{i_k}$$

这里 $i_k \in I$ 。由于加 0 不改变元素，可以发现仍有

$$f(u_1) = \sum_{k=1}^n \mu_k \beta_{i_k}, \quad f(u_2) = \sum_{k=1}^n \gamma_k \beta_{i_k}$$

而利用结合律、分配律

$$u_1 + u_2 = \sum_{k=1}^n (\gamma_k + \mu_k) \alpha_{i_k}$$

于是

$$f(u_1 + u_2) = \sum_{k=1}^n (\gamma_k + \mu_k) \beta_{i_k} = f(u_1) + f(u_2)$$

对 $u \in U, \lambda \in \mathbb{K}$ ，设

$$u = \sum_{k=1}^n \lambda_k \alpha_{i_k}$$

则有

$$f(u) = \sum_{k=1}^n \lambda_k \beta_{i_k}$$

而利用结合律、分配律

$$\lambda u = \sum_{k=1}^n (\lambda \lambda_k) \alpha_{i_k}$$

于是

$$f(\lambda u) = \sum_{k=1}^n \lambda \lambda_k \beta_{i_k} = \lambda f(u)$$

从而得证 f 的确是线性映射。

- 唯一性

首先，对线性映射 f 有 $f(0) = f(0+0) = 2f(0)$ ，从而 $f(0) = 0$ 恒成立。对非零的

$$u = \sum_{k=1}^n \lambda_k \alpha_{i_k}$$

利用线性性可知必然

$$f(u) = \sum_{k=1}^n \lambda_k f(\alpha_{i_k}) = \sum_{k=1}^n \lambda_k \beta_{i_k}$$

从而不可能有上方定义之外的线性映射满足条件。

5. 若 U, V 维数有限， U 到 V 存在单同态当且仅当 $\dim U \leq \dim V$ 。

证明：

- 若 $\dim U \leq \dim V$, 考虑 U 的一组基 $\alpha_1, \dots, \alpha_n$, V 的一组基 β_1, \dots, β_m , 且 $n \leq m$, 用基映射定义线性映射

$$f(\alpha_i) = \beta_i, \quad i = 1, \dots, n$$

利用上方结论, $\alpha_1, \dots, \alpha_n$ 的像在 V 中是一组基的一部分, 因此线性无关, 从而 f 是单同态。

- 若 f 是单同态, 利用上方结论, U 的一组基 $\alpha_1, \dots, \alpha_n$ 的像在 V 中线性无关, 利用基等价于极大线性无关组可知 V 存在包含 $f(\alpha_1), \dots, f(\alpha_n)$ 的基, 从而 $\dim U \leq \dim V$ 。

6. 若 U 、 V 维数有限, U 到 V 存在满同态当且仅当 $\dim U \geq \dim V$ 。

证明:

- 若 $\dim U \geq \dim V$, 考虑 U 的一组基 $\alpha_1, \dots, \alpha_n$, V 的一组基 β_1, \dots, β_m , 且 $n \geq m$, 用基映射定义线性映射

$$f(\alpha_i) = \beta_i, \quad i = 1, \dots, m$$

$$f(\alpha_i) = 0, \quad i = m + 1, \dots, n$$

利用上方结论, $\alpha_1, \dots, \alpha_n$ 的像包含 V 的一组基, 因此能生成 V , 是满同态。

- 若 f 是满射, 利用上方结论, U 的一组基 $\alpha_1, \dots, \alpha_n$ 的像能生成 V 。设 $f(\alpha_1), \dots, f(\alpha_n)$ 的极大线性无关组为 $f(\alpha_{i_1}), \dots, f(\alpha_{i_r})$, 则 $f(\alpha_{i_1}), \dots, f(\alpha_{i_r})$ 线性无关且可生成 V , 从而是 V 的一组基, 其个数不会超过 n , 因此 $\dim V \leq \dim U$ 。

7. 若 U 、 V 维数有限, U 与 V 同构当且仅当 $\dim U = \dim V$ 。

证明:

- 若 $\dim U = \dim V$, 考虑 U 的一组基 $\alpha_1, \dots, \alpha_n$, V 的一组基 β_1, \dots, β_n , 用基映射定义线性映射

$$f(\alpha_i) = \beta_i, \quad i = 1, \dots, n$$

利用上方结论, $\alpha_1, \dots, \alpha_n$ 的像是 V 的一组基, 因此 f 是线性同构。

- 若 U 、 V 同构, 设同构映射为 f , U 的一组基 $\alpha_1, \dots, \alpha_n$ 的像 $f(\alpha_1), \dots, f(\alpha_n)$ 是 V 的一组基, 因此基个数相同, $\dim U = \dim V$ 。

* 于是有限维时证明同构只需证明维数相同。

第四条结论暗示了构造线性映射的方式: 只要提供一组基的像即可, 此结论的更完善版本将在之后的例题中介绍。利用它, 我们证明了后三条结论, 它们事实上与有限集之间存在单射、满射、双射的情况非常类似, 这意味着有限维线性空间的结构可以由维数完全确定, 而存在 U 到 V 的单同态/满同态也确实意味着 U 的结构更小/更大。

19.2.3 构造双射的方式

在进入线性映射的下一部分之前, 我们还是要回到映射, 研究从映射构造双射的方式。

我们已经知道, 映射代表结构的部分保持, 而双射代表结构的“迁移”。那么, 从映射中, 是否能看到一个相同的结构呢? 考虑如下的简单例子:

$$X = \{1, 2, 3, 4, 5\}, \quad Y = \{1, 2, 3, 4\}$$

$$f(1) = 1, \quad f(2) = 2, \quad f(3) = 1, \quad f(4) = 2, \quad f(5) = 4$$

一个简单的想法是, 先将没有原像的 3 去掉, 再在 1、2、4 的原像中各取一个, 例如, 记 $X' = \{1, 2, 5\}$ 、 $Y' = \{1, 2, 4\}$, 并定义 $\tilde{f}(1) = 1$ 、 $\tilde{f}(2) = 2$ 、 $\tilde{f}(5) = 4$ (即 $\tilde{f}(x) = f(x)$), 则 \tilde{f} 构成一个双射, 这意味着我

们确实从 f 中得到了一个双射。不过, \tilde{f} 映射还并不够好: 它不能完全刻画 f 的信息, 忽略了 $f(3) = 1$ 与 $f(4) = 2$ 的部分。想要解决这个问题, 更好完成构造双射的任务, 我们将先从单射、满射构造双射。

为了从单射构造双射, 我们要进行的过程类似上方的第一步, 也即将没有原像的元素去掉。若 $f : X \rightarrow Y$ 是一个单射, 我们下面证明, 定义

$$\tilde{f} : X \rightarrow f(X), \quad \tilde{f}(x) = f(x)$$

则其成为一个双射。

证明:

首先, 由于任何 $f(x)$ 当然在 $f(X)$ 中, 此映射定义合理。若 $\tilde{f}(a) = \tilde{f}(b)$, 由定义可知 $f(a) = f(b)$, 从而由 f 单射得 $a = b$, 这就说明了 \tilde{f} 为单射。对任何 $y \in f(X)$, 根据定义存在 $x \in X$ 使得 $f(x) = y$, 从而 $\tilde{f}(x) = y$, \tilde{f} 为满射。综合以上讨论得证 \tilde{f} 为双射。

那么, 若 $f : X \rightarrow Y$ 是一个满射, 如何定义对应的映射呢? 如果按照之前的想法, 我们的映射构造是这样的: 对每个 $y \in Y$, 在 $\{x \mid f(x) = y\}$ 中选择一个, 构成子集 S , 再在 S 上定义

$$\tilde{f} : S \rightarrow Y, \quad \tilde{f}(x) = f(x)$$

则其成为一个双射。

证明:

由于 $f(x) \in Y$, 此映射定义合理。若 $\tilde{f}(a) = \tilde{f}(b)$, 由于 \tilde{f} 的定义可知 $a, b \in \{x \mid f(x) = y\}$, 但根据 S 的定义 $S \cap \{x \mid f(x) = y\}$ 只有一个元素, 从而 $a = b$, \tilde{f} 为单射。同样, 由于对任何 $y \in Y$, $S \cap \{x \mid f(x) = y\}$ 有元素, 存在 $s \in S$ 使得 $\tilde{f}(s) = y$, 因此 \tilde{f} 为满射。

* 一个有趣的事情是, 上述的选择在集合有无穷多个元素时无法从集合论公理直接推出存在。于是, 选择的存在性成为了一条新的公理, 也就是选择公理。由于我们的课程默认其成立, 上述的证明是没有问题的。

不过, 刚才我们已经解释了, 这样的选择会损失原映射的信息。因此, 我们希望能有一个不损失信息的方式。对于之前举例的 f , 可以从它构造一个这样的映射 $\tilde{f} : X'' \rightarrow Y'$:

$$X' = \{\{1, 3\}, \{2, 4\}, \{5\}\}, \quad Y' = \{1, 2, 4\}$$

$$\tilde{f}(\{1, 3\}) = 1, \quad \tilde{f}(\{2, 4\}) = 2, \quad \tilde{f}(\{5\}) = 4$$

这里, 我们将原集合划分为了几个子集, 并在子集的集合上构造了双射。观察划分方式, 我们会发现, 每一类中都满足 $f(a) = f(b)$ 。由此, 若 $f : X \rightarrow Y$ 是一个满射, 我们定义 X 上的关系 \sim , $a \sim b$ 当且仅当 $f(a) = f(b)$ 。直接由定义可以验证这是一个等价关系, 从而可以存在商集 X / \sim (定义参考可本讲义 18.4.1)。接下来, 定义 X / \sim 到 Y 的映射

$$\tilde{f}([x]) = f(x)$$

下面说明这是一个双射。

证明:

等价关系验证: $f(a) = f(a)$, 因此 $a \sim a$; 若 $a \sim b$ 有 $f(a) = f(b)$, 从而 $f(b) = f(a)$, $b \sim a$; 若 $a \sim b$, $b \sim c$ 有 $f(a) = f(b) = f(c)$, 从而 $f(a) = f(c)$, $a \sim c$ 。

对 \tilde{f} , 首先验证定义合理性。若 $[x] = [y]$, 这意味着 $x \sim y$, 因此 $f(x) = f(y)$, 从而有 $\tilde{f}([x]) = \tilde{f}([y])$, 定义合理。

对任何 $y \in Y$, 由 f 为满射存在 x 使得 $f(x) = y$, 于是 $\tilde{f}([x]) = y$, 因此其为满射。对 $[x] \neq [y]$, 由等价关系定义可知 $f(x) \neq f(y)$, 从而 $\tilde{f}([x]) \neq \tilde{f}([y])$, 因此其为双射。综合两部分得双射。

最后，经过上面的单射定义双射与满射定义双射，一般映射定义双射的方式就很清晰了，它可以看成之前两步的结合。我们将定理叙述如下，这个定理可以称为集合之间的第一同构定理：

设 $f : X \rightarrow Y$ 是一个映射，在 X 上定义关系 \sim_f , $x \sim_f y$ 当且仅当 $f(x) = f(y)$, 则其构成等价关系，定义对应的商集到 f 像集的映射 $\tilde{f} : X / \sim_f \rightarrow f(X)$, $\tilde{f}([x]) = f(x)$ 是一个双射。

证明：

设 $\hat{f} : X \rightarrow f(X)$ 满足 $\hat{f}(x) = f(x)$, 则与单射构造双射的方式相同可以验证 \hat{f} 是一个满射。

由于 $\hat{f}(x) = f(x)$, 用 \hat{f} 仿照满射定义双射时定义的等价关系 $\sim_{\hat{f}}$ 事实上就是 \sim_f , 从而 \tilde{f} 即为满射 \hat{f} 按照上面满射定义双射的方式定义的双射，得证。

这个定理从一般映射得到了双射，事实上说明了一般映射保持了怎样的一个结构：它说明了左侧的某个商结构与右侧的某个子结构相同。

19.2.4 限制映射 I

在介绍线性空间之间的线性映射对应的第一同构定理前，我们先更深入讨论一个刚才讨论集合时涉及到的内容。

在上面的单射构造双射与第一种满射构造双射的方式（取子集 S ）中，我们都定义出了一个满足在定义域中 $g(x) = f(x)$ 的映射。我们将这样的映射称为限制映射，意为将 $f(x)$ “限制”到更小的集合中。再细分来看，事实上有两种限制的方式，一个是对陪域 ($f : X \rightarrow Y$ 的 Y) 进行的限制，一个是对定义域进行的限制。将它们整理到一起，我们可以如下定义限制映射：

对映射 $f : X \rightarrow Y$ 时，若 $X_0 \subset X$ 与 $Y_0 \subset Y$ 满足 $f(X_0) \subset Y_0$, 则可以定义限制映射 $g : X_0 \rightarrow Y_0$, 满足 $g(x) = f(x)$, 将这样的 g 记作

$$f|_{X_0 \rightarrow Y_0}$$

证明：

由于定义已经保证了 $f(X_0) \subset Y_0$, 这样的 $f|_{X_0 \rightarrow Y_0}$ 定义合理。

限制映射的意义很清晰：若 $X \rightarrow Y$ 之间保持了某种结构，那么 X 的任何一个子集到适当的 Y 的子集也对应保持了结构。利用限制映射，承认选择公理时我们可以给出另一种从映射构造双射的方式：

设 $f : X \rightarrow Y$ 是一个映射，对每个 $y \in f(X)$, 在 $\{x \mid f(x) = y\}$ 中选择一个，构成子集 S , 则 $f|_{S \rightarrow f(X)}$ 是一个双射。

证明：

由于 $f(x) \in f(X)$, 此映射定义合理。若 $f|_{S \rightarrow f(X)}(a) = f|_{S \rightarrow f(X)}(b)$, 由于 $f|_{S \rightarrow f(X)}$ 的定义可知 $a, b \in \{x \mid f(x) = y\}$, 但根据 S 的定义 $S \cap \{x \mid f(x) = y\}$ 至多只有一个元素，从而 $a = b$, $f|_{S \rightarrow f(X)}$ 为单射。同样，由于对任何 $y \in f(X)$, $S \cap \{x \mid f(x) = y\}$ 有元素，存在 $s \in S$ 使得 $f(s) = y$, 从而 $f|_{S \rightarrow f(X)}(s) = y$, 因此 $f|_{S \rightarrow f(X)}$ 为满射。

注意到，这种构造双射的方式说明了 X 存在子集 S 与 $f(X)$ 同构，而上一种方式则是 X / \sim_f 与 Y 同构。更一般地，可以说明，承认选择公理时，任何一个集合的商集一定和它某个子集同构。

对于线性映射，我们可以完全类似定义限制映射： $f : U \rightarrow V$ 是一个 \mathbb{K} 上线性空间之间的线性映射，若 U 的子空间 U_0 与 V 的子空间 V_0 满足 $f(U_0) \subset V_0$, 则可以定义限制映射 $g : U_0 \rightarrow V_0$, 满足 $g(x) = f(x)$, 将这样的 g 记作 $f|_{U_0 \rightarrow V_0}$, 它仍然是一个线性映射。

证明：

限制映射的定义合理性在之前对映射时已经说明，只需说明线性。验证是简单的：对 $x, y \in U_0$, 由 U_0 封闭性 $x + y \in U_0$, 且对 $\lambda \in \mathbb{K}$ 有 $\lambda x \in U_0$, 它们都在 $f|_{U_0 \rightarrow V_0}$ 定义域中，且由 f 线性性

$$f|_{U_0 \rightarrow V_0}(x + y) = f(x + y) = f(x) + f(y) = f|_{U_0 \rightarrow V_0}(x) + f|_{U_0 \rightarrow V_0}(y)$$

$$f|_{U_0 \rightarrow V_0}(\lambda x) = f(\lambda x) = \lambda f(x) = f|_{U_0 \rightarrow V_0}(\lambda x)$$

这就得到了证明。

* 这个证明看起来比较“废话”，恰恰是因为我们限制映射的定义就是为了不改变 f 的性质。

* 虽然下一节中不会涉及，但利用限制映射也可以从线性映射构造同构，见本讲义 20.1.1 的第一道例题。

这里我们只给出线性映射限制的概念，在之后的学习中，我们将慢慢看到线性映射限制映射的运用。某种意义上，它是比集合的限制映射更有用的，因为它意味着映射可以随空间一起分解。

19.2.5 第一同构定理

进行完上面对集合的讨论后，我们终于可以开始研究一般线性映射最重要的结论，也即如何从线性映射出发构造同构了。考虑 \mathbb{K} 上线性空间 $U \rightarrow V$ 的线性映射 f ，利用集合的第一同构定理，我们可以构造出 $\tilde{f}: U/\sim_f \rightarrow f(U)$ 的双射。下面的所有讨论事实上都在说明，这个双射就是一个线性同构。

首先，仿照集合，我们仍然希望先从单射、满射开始构造线性同构。不过，线性性会给单射、满射一些不同的刻画。具体来说，定义 $\text{Im } f = f(U)$ ，称为 f 的像空间，则满射的条件可以写为 $\text{Im } f = V$ ，而 $\text{Im } f$ 是 V 的子空间，因此有限维时可以直接通过维数计算确定。

证明：

在本讲义 19.2.2 已证明 W 是 U 子空间时 $f(W)$ 为线性空间，取 $W = U$ 得证 $\text{Im } f$ 线性。满射即为 $f(U) = V$ ，从而可写为 $\text{Im } f = V$ 。

对于单射，线性映射则有与集合截然不同的刻画：定义 $\text{Ker } f = \{x \in U \mid f(x) = 0\}$ ，则其构成 U 的子空间，称为 f 的核空间，且 f 是单射当且仅当 $\text{Ker } f = \{0\}$ ，这是对线性映射最常用的单射判定方式。

证明：

先证明其为线性空间。由于是子空间验证，只需证明封闭，若 $f(a) = 0, f(b) = 0$ ，有 $f(a + b) = f(a) + f(b) = 0$ ，从而 $a + b \in \text{Ker } f$ ；若 $f(a) = 0$ ，对任何 $\lambda \in \mathbb{K}$ ，有 $f(\lambda a) = \lambda f(a) = 0$ ，从而 $\lambda a \in \text{Ker } f$ 。因此 $\text{Ker } f$ 是 U 的子空间。

若 f 是单射，0 的原像至多一个，之前已证线性映射必然满足 $f(0) = 0$ ，从而 $\text{Ker } f = \{0\}$ ；若 $\text{Ker } f = \{0\}$ ，对任何 $v \in V$ ，假设 $f(x) = f(y) = v$ ，可发现 $f(x - y) = f(x) - f(y) = 0$ ，从而 $x - y \in \text{Ker } f$ ，即得到 $x = y$ ，符合单射条件。

进一步可以得到，按照之前定义的等价关系 $u \sim v$ 当且仅当 $f(u) = f(v)$ ， $u \in U$ 所在的等价类即为 $u + \text{Ker } f$ 。

证明：

对任何 $x \in u + \text{Ker } f$ ，设其为 $u + z, z \in \text{Ker } f$ ，则由 $\text{Ker } f$ 定义

$$f(x) = f(u + z) = f(u) + f(z) = f(u)$$

反之，若 $f(x) = f(u)$ ，则类似上方推导 $f(x - u) = 0$ ，从而 $x - u \in \text{Ker } f$ ，设 $z = x - u \in \text{Ker } f$ 有 $x = u + z$ ，即得 $x \in u + \text{Ker } f$ 。

由此， U/\sim_f 事实上就是商空间 $U/\text{Ker } f$ ，自然也可以看成线性空间。接下来，我们将快速走完之前集合构造双射的过程。值得注意的是，这里每一部分定义的映射都和集合时完全相同，只是额外增加了线性性验证：

1. 单同态构造同构：若 $f: U \rightarrow V$ 是 \mathbb{K} 上线性空间之间的单同态，则 $f|_{U \rightarrow \text{Im } f}$ 是一个同构。

证明：

由于这与映射时单射构造双射的方式完全相同，已验证得到的 $f|_{U \rightarrow \text{Im } f}$ 为双射。再根据子空间的限制映射保持线性性可知它是线性同构。

2. 满同态构造同构: 若 $f: U \rightarrow V$ 是 \mathbb{K} 上线性空间之间的满同态, 则定义 $\tilde{f}: U/\text{Ker } f \rightarrow V$, 使得 $\tilde{f}(x + \text{Ker } f) = f(x)$, 它是一个同构。

证明:

利用 $x + \text{Ker } f = [x]$, 这与映射时满射构造双射的方式完全相同, 已验证得到的 \tilde{f} 为双射 (包含定义合理性的验证都与映射时相同)。为验证线性性, 直接由定义与商空间的加法定义可知对 $x, y \in U$ 与 $\lambda \in \mathbb{K}$ 有

$$\begin{aligned}\tilde{f}((x + \text{Ker } f) + (y + \text{Ker } f)) &= \tilde{f}((x + y) + \text{Ker } f) \\ &= f(x + y) = f(x) + f(y) = \tilde{f}(x + \text{Ker } f) + \tilde{f}(y + \text{Ker } f) \\ \tilde{f}(\lambda(x + \text{Ker } f)) &= \tilde{f}(\lambda x + \text{Ker } f) = f(\lambda x) = \lambda f(x) = \lambda \tilde{f}(x + \text{Ker } f)\end{aligned}$$

从而其为线性映射, 再结合双射知为同构。

3. 线性映射构造同构: 若 $f: U \rightarrow V$ 是 \mathbb{K} 上线性空间之间的线性映射, 则定义 $\hat{f}: U/\text{Ker } f \rightarrow \text{Im } f$, 使得 $\hat{f}(x + \text{Ker } f) = f(x)$, 它是一个同构。

证明:

设 $\hat{f}: U \rightarrow \text{Im } f$ 满足 $\hat{f}(x) = f(x)$, 则与单同态构造同构的方式相同可以验证 \hat{f} 是一个满同态。由于 $\hat{f}(x) = f(x)$, 可知 $\text{Ker } \hat{f} = \text{Ker } f$, 用 \hat{f} 仿照满同态定义同构时定义的商空间 $U/\text{Ker } \hat{f}$ 事实上就是 $U/\text{Ker } f$, 从而 \hat{f} 即为满同态 \hat{f} 按照上面满同态定义同构的方式定义的同构, 得证。

* 换句话说, $U/\text{Ker } f$ 与 $\text{Im } f$ 同构。

这里映射构造双射的方式即称为第一同构定理, 它可以用于各种商空间同构相关的验证与维数计算。我们给出四个能用它推出的结论 (后三个结论的证明即为第一同构定理常见的使用方式):

1. 像与核维数公式: 若 $f: U \rightarrow V$ 是 \mathbb{K} 上线性空间之间的线性映射, 且 U, V 维数有限, 则

$$\dim \text{Ker } f + \dim \text{Im } f = \dim U$$

证明:

利用定义 $\dim U/\text{Ker } f = \dim U - \dim \text{Ker } f$, 而已证有限维线性空间同构等价于维数相等, 从而 $\dim U - \dim \text{Ker } f = \dim \text{Im } f$, 移项得证。

* V 维数无限时结论仍然成立, 这是因为利用同构将基映射到基可知有限维空间不可能与无穷维空间同构, 从而 $\text{Im } f$ 维数有限, 再结合之前的同构结论即得。此结论当 $f(x)$ 为矩阵乘法时即为上学期常用的解空间维数定理。

2. 商空间与子空间同构性: 对 \mathbb{K} 上线性空间 U 与其子空间 W , 设 W 对 U 的一个补空间是 V (若承认选择公理, 其一定存在), 则 U/W 与 V 同构。

证明:

由条件可知 $U = W \oplus V$, 从而任何 $u \in U$ 可以唯一写成 $w + v$, 其中 $w \in W, v \in V$ 。定义映射 $f: U \rightarrow U$ 满足 $f(w + v) = v$, 下面证明其是线性映射并构造第一同构定理。

- 定义合理性

由直和的性质, 对任何 u, v 是唯一确定的, 从而此映射的确是 $U \rightarrow U$ 的映射。

- 线性性

若 $u = w + v$ 、 $u' = w' + v'$, 利用 V 、 W 封闭性可知

$$u + u' = (w + w') + (v + v'), \quad w + w' \in W, \quad v + v' \in V$$

从而再根据分解唯一性

$$f(u + u') = v + v' = f(u) + f(u')$$

对任何 $\lambda \in \mathbb{K}$, 利用 V 、 W 封闭性可知

$$\lambda u = \lambda w + \lambda v, \quad \lambda w \in W, \quad \lambda v \in V$$

从而再根据分解唯一性

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

- 像空间

由定义可发现 $\text{Im } f \subset V$, 且对任何 $v \in V$ 有 $f(v) = v$, 从而有原像, 因此 $\text{Im } f = V$ 。

- 核空间

首先, 根据分解唯一性, 任何 $w \in W$ 等于 $w + 0$, 因此 $f(w) = 0$, 即 $W \in \text{Ker } f$ 。对不在 w 中的 u , 其分解 $u = w + v$ 中 v 不可能为 0, 否则 $u = w \in W$, 从而 $u \notin \text{Ker } f$ 。综合得到 $\text{Ker } f = W$ 。

利用第一同构定理即得 $U/W = U/\text{Ker } f$ 同构于 $\text{Im } f = V$ 。

3. 第二同构定理: 对 \mathbb{K} 上线性空间 V 的子空间 U 、 W , $(U + W)/U$ 同构于 $W/(U \cap W)$ 。

证明:

由定义, 任何 $x \in U + W$ 可以写为 $u + w$, 其中 $u \in U, w \in W$, 定义映射

$$f : (U + W) \rightarrow W/(U \cap W)$$

满足 $f(u + w) = w + U \cap W$, 下面证明其是线性映射并构造第一同构定理。

- 定义合理性

由于 $w \in W$, $w + U \cap W$ 的确是 $W/(U \cap W)$ 的元素。

对元素 v , 若有两种分解 $v = u + w = u' + w'$, 可得 $u - u' = w' - w$, 从而 $w' - w \in U$, 再由 $w, w' \in W$ 可知 $w' - w \in W$, 综合得到 $w' - w \in U \cap W$, 于是利用等价类定义 $w + U \cap W = w' + U \cap W$, 因此 v 映射到的像唯一, 即证此映射定义合理。

* 注意验证了定义合理性时, 对 v 我们可以选择想要的分解方式, 下面计算像与核的过程都应用了这点。

- 线性性

若 $v = u + w$ 、 $v' = u' + w'$, 利用 U 、 W 封闭性可知

$$v + v' = (u + u') + (w + w'), \quad u + u' \in U, \quad w + w' \in W$$

从而再根据分解唯一性与商空间加法定义

$$f(v + v') = (w + w') + U \cap W = w + U \cap W + w' + U \cap W = f(v) + f(v')$$

对任何 $\lambda \in \mathbb{K}$, 利用 U 、 W 封闭性可知

$$\lambda v = \lambda u + \lambda w, \quad \lambda u \in U, \quad \lambda w \in W$$

从而再根据分解唯一性与商空间加法定义

$$f(\lambda v) = \lambda w + U \cap W = \lambda(w + U \cap W) = \lambda f(v)$$

– 像空间

对任何 $w + U \cap W, w \in W$, 由于 $w \in U + W$, 由定义 $f(w) = f(0 + w) = w + U \cap W$, 于是其有原像, 从而 $\text{Im } f = W/(U \cap W)$ 。

– 核空间

对任何 $u \in U$, 有 $f(u) = f(u + 0) = U \cap W$, 从而 $u \in \text{Ker } f$ 。对不在 U 中的 v , 其分解 $v = u + w$ 中 w 不可能在 $U \cap W$ 中, 否则由 $u, w \in U$ 可知 $v \in U$, 矛盾, 从而根据等价类定义得到 $w + U \cap W \neq U \cap W, v \notin \text{Ker } f$ 。综合得 $\text{Ker } f = U$ 。

利用第一同构定理即得 $(U + W)/U = (U + W)/\text{Ker } f$ 同构于 $\text{Im } f = W/(U \cap W)$ 。

* 从此定理计算维数可以直接得到和空间维数公式, 也即核空间维数公式本质上是第二同构定理。

4. 第三同构定理: 对 \mathbb{K} 上线性空间 V 的子空间 U 、 U 的子空间 W , $(V/W)/(U/W)$ 同构于 V/U 。

* 注意 V/W 的元素为所有 $v + W, v \in V$, U/W 的元素为所有 $u + W, u \in U \subset V$, 因此后者为前者的子集, 且为线性空间, 从而为前者的子空间, 可以定义商空间, 商空间的元素应为 $(v + W) + U/W$ 。

证明:

定义映射 $f : (V/W) \rightarrow (V/U)$ 满足 $f(v + W) = v + U$, 下面证明其是线性映射并构造第一同构定理。

– 定义合理性

当 $v_1 + W = v_2 + W$ 时, 由等价类定义可知 $v_1 - v_2 \in W$, 而由 $W \subset U$ 可知 $v_1 - v_2 \in U$, 从而 $v_1 + U = v_2 + U$, 即同一个元素的像唯一, 映射定义合理。

– 线性性

利用商空间加法、数乘定义可知

$$\begin{aligned} f((v + W) + (v' + W)) &= f((v + v') + W) \\ &= (v + v') + U = (v + U) + (v' + U) = f(v + W) + f(v' + W) \\ f(\lambda(v + W)) &= f(\lambda v + W) = \lambda v + U = \lambda(v + U) = \lambda f(v) \end{aligned}$$

– 像空间

对任何 $v + U, v \in V$, 由定义 $f(v + W) = v + U$, 于是其有原像, 从而 $\text{Im } f = V/U$ 。

– 核空间

先说明 $v + W \in U/W$ 与 $v \in U$ 等价。利用商空间定义, $v \in U$ 时 $v + W$ 确实是 U/W 中元素。另一方面, 当 $v + W \in U/W$ 时, 意味着存在 $u \in U$ 使得 $u + W = v + W$, 也即存在 $u \in U$ 使得 $v - u \in W$, 而 $W \subset U$, 从而 $v \in U$, 得证。

根据商空间零元素定义 $v + W \in \text{Ker } f$ 当且仅当 $f(v + W) = v + U = U$, 利用等价类定义这当且仅当 $v \in U$, 再由上方已证这当且仅当 $v + W \in U/W$, 于是 $\text{Ker } f = U/W$ 。

这样, 我们就从第一同构定理出发证明出了对一般线性映射最重要的几个结论。读者可以自己思考第二与第三同构定理是否有集合论的版本。回顾对一般线性映射的讨论, 我们几乎都在研究它保持结构的具体方式 (最重要的结论是第一同构定理) 和与基的联系 (最重要的结论是用基映射定义线性映射)。接下来, 对于有限维线性空间, 我们将在基的个数有限时找到全部的线性映射, 并给出刻画。

§19.3 矩阵表示

19.3.1 同构标准形

既然我们这学期的研究主题仍然是有限维线性空间之间的映射, 我们还是希望能有更好的方式刻画出这些映射, 而不是只能通过抽象的 f 来代表。

为了之后的推理，我们先来证明线性空间的同构事实上是一种等价关系。

证明：

对线性空间 U ，定义 $f : U \rightarrow U$ 满足 $f(u) = u$ ，则可验证其为线性双射，从而为同构，于是 U 与 U 同构。

若 U 与 V 同构，设 f 为同构映射，则根据已证 f^{-1} 为 V 到 U 的同构映射，从而 V 与 U 同构。

若 U 与 V 同构、 V 与 W 同构，设 $f : U \rightarrow V$ 、 $g : V \rightarrow W$ ，由已证 $g \circ f$ 是 $U \rightarrow W$ 的线性映射，且根据双射的复合还是双射可知其为双射，从而得到 $g \circ f$ 是 U 到 W 的同构，从而 U 与 W 同构。

* 值得一提的是，出于某些集合论的规定（为了避免出现著名的理发师悖论），所有线性空间并不构成一个集合，而是构成一个类——可以简单理解为一个可以容纳比集合更多东西的“框子”。事实上，等价关系与等价类都可以定义在类上，具体可以参考范畴论的相关知识。我们将忽略这套抽象的理论，直接考虑每个线性空间所在的等价类的元素，而不考虑整个类。

在刚才的讨论中，我们证明了一个很重要的结论：两个 \mathbb{K} 上的有限维线性空间同构当且仅当有相同的维数。由此，所有 \mathbb{K} 上的 n 维线性空间构成一个等价类。毫无疑问，我们希望 \mathbb{K}^n 能成为这个等价类的代表元，因为它形式最简单。

更有意思的是，我们其实已经知道任何一个 \mathbb{K} 上 n 维线性空间 U 到 \mathbb{K}^n 的同构映射是什么。对写成形式行向量的 U 的一组基 $S = (\alpha_1, \dots, \alpha_n)$ ，设 $f : U \rightarrow \mathbb{K}^n$ 满足 $f(u) = u_S$ ，其中 u_S 为 u 的坐标，则其是一个同构。我们之后将把这个映射记为 π_S ，也即 $\alpha_S = \pi_S(\alpha)$ 。

证明：

首先，在之前介绍坐标时我们已经证明其唯一性，从而它确实是一个良好定义的映射。

先证明其为线性映射：对 $\alpha, \beta \in U$ 由定义有 $S\alpha_S = \alpha$ 、 $S\beta_S = \beta$ ，因此利用形式乘法分配律

$$\alpha + \beta = S\alpha_S + S\beta_S = S(\alpha_S + \beta_S)$$

从而由坐标唯一性 $\pi_S(\alpha + \beta) = \alpha_S + \beta_S$ 。同理利用形式乘法与数乘可交换，对任何 $\lambda \in \mathbb{K}$ 由定义有

$$\lambda\alpha = \lambda S\alpha_S = S(\lambda\alpha_S)$$

从而由坐标唯一性 $\pi_S(\lambda\alpha) = \lambda\alpha_S$ 。

根据坐标的定义， π_S 把 $\alpha_1, \dots, \alpha_n$ 映射到了 e_1, \dots, e_n ，这里 e_i 表示第 i 个单位向量。由于所有 e_i 构成 \mathbb{K}^n 的一组基， π_S 把一组基映射到一组基，因此为同构。

接下来，我们将要用这个同构——也就是说所有 \mathbb{K} 上 n 维线性空间都可以看成 \mathbb{K}^n ——刻画所有有限维线性空间的线性映射。

19.3.2 坐标的线性映射

就像我们之前对相抵标准形或合同标准形的应用，用标准形解决一般问题往往分为两步：解决标准形时的问题，并把一般问题化到标准形上。我们下面假设 U 、 V 分别是 \mathbb{K} 上的 n 维、 m 维线性空间，我们想解决的问题是，如何刻画全部 U 到 V 的线性映射。我们将所有 U 到 V 线性映射的集合记作 $\text{Hom}(U, V)$ 。

1. 标准形问题： $\text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ 的构成

首先，我们证明，对 $A \in \mathbb{K}^{m \times n}$ ， $\mathcal{A}(x) = Ax$ 是一个线性映射。

证明：

由矩阵乘法定义可知 $x \in \mathbb{K}^n$ 时 $Ax \in \mathbb{K}^m$ 。此外可直接验证 $x, y \in \mathbb{K}^n$ 、 $\lambda \in \mathbb{K}$ 时

$$\mathcal{A}(x + y) = A(x + y) = Ax + Ay = \mathcal{A}(x) + \mathcal{A}(y)$$

$$\mathcal{A}(\lambda x) = A(\lambda x) = \lambda Ax = \lambda \mathcal{A}(x)$$

从而结论成立。

其次, 对任何 $\mathcal{A} \in \text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$, 存在矩阵 $A \in \mathbb{K}^{m \times n}$ 使得 $\mathcal{A}(x) = Ax$ 对任何 $x \in \mathbb{K}^n$ 成立, 从而 \mathbb{K}^n 到 \mathbb{K}^m 的线性映射实际上就是矩阵乘法。

证明:

设 e_i 为第 i 个 n 维单位向量, 设 $\mathcal{A}(e_i) = \alpha_i \in \mathbb{K}^m$, 其中 $i = 1, \dots, n$, 定义

$$A = (\alpha_1, \dots, \alpha_n)$$

利用 α 为 m 维列向量可知 A 的确是 $m \times n$ 矩阵。且直接计算可发现

$$Ae_i = \alpha_i, \quad i = 1, \dots, n$$

根据之前已经证明, 线性映射可以通过一组基的像确定, 而 $x \rightarrow Ax$ 与 \mathcal{A} 都为线性映射, 且在 \mathbb{K}^n 的一组基 e_1, \dots, e_n 上的像都为 $\alpha_1, \dots, \alpha_n$, 从而即可知两映射相同, 即 $\mathcal{A}(x) = Ax$ 对任何 $x \in \mathbb{K}^n$ 成立。

最后这样的 A 是唯一的, 也即不同的线性映射对应的 A 不同。

证明:

考虑 $\mathcal{A}, \mathcal{B} \in \text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$, 由已证可知存在 $A, B \in \mathbb{K}^{m \times n}$ 使得 $\mathcal{A}(x) = Ax$ 、 $\mathcal{B}(x) = Bx$ 对任何 $x \in \mathbb{K}^n$ 成立。

若 $A = B$, 即有 $\mathcal{A}(x) = Ax = Bx = \mathcal{B}(x)$ 对任何 $x \in \mathbb{K}^n$ 成立。否则, A, B 至少有一列不同, 设为第 i 列, 则计算可发现 Ae_i 为 A 的第 i 列, 从而

$$\mathcal{A}(e_i) = Ae_i \neq Be_i = \mathcal{B}(e_i)$$

这就得到了 $\mathcal{A} \neq \mathcal{B}$ 。

结合两方面的讨论, 即得到 $A = B$ 当且仅当 $\mathcal{A} = \mathcal{B}$, 于是不同线性映射对应的 A 不同。

综合以上, $\text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ 到 $\mathbb{K}^{m \times n}$ 对应一个自然的双射, 即将满足 $\mathcal{A}(x) = Ax$ 的映射 \mathcal{A} 映射到 A (由于将每个 A 映射到 $\mathcal{A}(x) = Ax$ 的映射是其逆映射, 其必然双射)。更进一步来说, $\text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ 可以某种意义上“看作” $\mathbb{K}^{m \times n}$ 。

2. 一般的 $\text{Hom}(U, V)$ 转化为 $\text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$

在之前的讨论中, 我们已经知道了, 取定 U 的一组基 S 、 V 的一组基 T , 可以写出同构 $\pi_S : U \rightarrow \mathbb{K}^n$ 与 $\pi_T : V \rightarrow \mathbb{K}^m$ 。那么, 能不能用这两个同构将 U 到 V 的线性映射转移到 $\mathbb{K}^n \rightarrow \mathbb{K}^m$ 呢?

我们已经知道, 线性映射的复合还是线性映射, 同构的逆还是同构。因此, 如果我们有一个 U 到 V 的线性映射 f , 想象如下的过程(可以自行在 $U, V, \mathbb{K}^m, \mathbb{K}^n$ 之间画箭头表示下方过程):

- 为了能让映射的起点是 \mathbb{K}^n , 先将 \mathbb{K}^n 中的向量 x 同构到 U 上, 对应的同构映射是 π_S^{-1} , 也即 x 成为了 $u = \pi_S^{-1}(x)$;
- 接着, 应用我们已知的映射 f , 将 u 映射到 V 中元素 v ;
- 最后, 再利用同构将 V 中元素用坐标还原回 \mathbb{K}^m , 也即 $y = \pi_T(v)$ 。

经过这三步, 我们将 $f : U \rightarrow V$ “扩充”为了 \mathbb{K}^n 到 \mathbb{K}^m 的线性映射 \mathcal{A} , 使得 $\mathcal{A}(x) = \pi_T(f(\pi_S^{-1}(x)))$, 利用映射复合写为

$$\mathcal{A} = \pi_T \circ f \circ \pi_S^{-1}$$

更准确来说, 由于这个 \mathcal{A} 与 S, T 有关, 我们可以记作 \mathcal{A}_{ST} 。

由于求逆与复合保持线性性, \mathcal{A}_{ST} 的确是一个 \mathbb{K}^n 到 \mathbb{K}^m 的线性映射, 因此存在 A_{ST} 使得

$$\mathcal{A}_{ST}(x) = A_{ST}x, \quad \forall x \in \mathbb{K}^n$$

最后, 我们证明 $\text{Hom}(U, V)$ 到 $\text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ 的映射 (注意这是一个映射到映射的映射……但经过上面的讨论, 含义应该已经清晰了) $f \rightarrow \pi_T \circ f \circ \pi_S^{-1}$ 是一个双射, 从而上述的 f 与 \mathcal{A}_{ST} 是一一对应。

证明:

这里我们采取一个和以往不太一样的证明方法, 直接说明其逆映射存在。

将上述映射记为 φ , 考虑映射 $\psi: \mathcal{A} \rightarrow \pi_T^{-1} \circ \mathcal{A} \circ \pi_S$ 。从右往左看 (注意复合映射是从右往左计算的), 对 $\mathbb{K}^n \rightarrow \mathbb{K}^m$ 的线性映射 \mathcal{A} 得到的 $\pi_T^{-1} \circ \mathcal{A} \circ \pi_S$, 它先将 U 同构到 \mathbb{K}^n , 再将 \mathbb{K}^n 线性映射到 \mathbb{K}^m , 最后通过逆将 \mathbb{K}^m 同构到 V , 因此确实是一个 $U \rightarrow V$ 的线性映射, 从而 ψ 是一个 $\text{Hom}(\mathbb{K}^n, \mathbb{K}^m) \rightarrow \text{Hom}(U, V)$ 的映射。

对任何 $f \in \text{Hom}(U, V)$ 、 $\mathcal{A} \in \text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$, 利用映射复合的结合律和逆映射定义 (注意恒等映射复合任何其他映射为自身) 有

$$\psi(\varphi(f)) = \pi_T^{-1} \circ (\pi_T \circ f \circ \pi_S^{-1}) \circ \pi_S = (\pi_T^{-1} \circ \pi_T) \circ f \circ (\pi_S^{-1} \circ \pi_S) = f$$

$$\varphi(\psi(\mathcal{A})) = \pi_T \circ (\pi_T^{-1} \circ \mathcal{A} \circ \pi_S) \circ \pi_S^{-1} = (\pi_T \circ \pi_T^{-1}) \circ \mathcal{A} \circ (\pi_S \circ \pi_S^{-1}) = \mathcal{A}$$

由此可以得到 $\psi \circ \varphi$ 与 $\varphi \circ \psi$ 都是恒等映射, 从而二者互逆, 即得到 φ 是双射。

* 这个证明其实没有看起来那么可怕, 它某种意义上是在证明 (利用之后所说的线性映射复合对应矩阵表示乘法), 当 P, Q 为可逆矩阵时, 映射 $X \rightarrow PXQ^{-1}$ 是一个可逆映射, 由于可以直接左乘 P^{-1} 右乘 Q 还原回 X , 可以较自然想到直接构造逆的思路。

上面两部分中我们已经得到, 取定 U 的基 S 与 V 的基 T 后, $\text{Hom}(U, V)$ 中的线性映射 f 与 $\text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ 中的线性映射 \mathcal{A}_{ST} 是一一对应的, 而 $\text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ 中的线性映射 \mathcal{A}_{ST} 又是与 $\mathbb{K}^{m \times n}$ 中的矩阵 A_{ST} 一一对应的, 我们将矩阵 A_{ST} 称为 f 在基 S, T 下的矩阵表示,

由于在给定基 S, T 后存在唯一的 A_{ST} 可以表示 f , 利用上方证明中构造的 $\psi(\mathcal{A}) = \pi_T^{-1} \circ \mathcal{A} \circ \pi_S$, 我们最终得到, 对任何线性映射 $f \in \text{Hom}(U, V)$ 与 U 的基 S, V 的基 T , 存在唯一矩阵 A_{ST} 使得

$$f(u) = \pi_T^{-1}(A_{ST}\pi_S(u))$$

对任何 $u \in U$ 成立, 或设 $\mathcal{A}_{ST}(x) = A_{ST}x$ 写成

$$f = \pi_T^{-1} \circ \mathcal{A}_{ST} \circ \pi_S$$

此外, 对每个 A_{ST} , 上述构造都决定了一个 $f \in \text{Hom}(U, V)$ 。这样, 我们就给出了 $\text{Hom}(U, V)$ 中的全部线性映射。

不过, 上面这个表述终究还是太抽象了, 有没有更具体的写法呢? 首先, $\pi_S(u)$ 就是 u 的坐标 u_S , 而两边同时在左侧复合 π_T 得到 (右侧 π_T 与 π_T^{-1} 抵消)

$$\pi_T(f(u)) = A_{ST}u_S$$

这也就代表

$$f(u)_T = A_{ST}u_S$$

从而线性映射的矩阵表示就是坐标之间的矩阵乘法。

一个很简单的例子是, 如果 $S = (\alpha_1, \dots, \alpha_n)$ 、 $T = (\beta_1, \dots, \beta_m)$, $A_{ST} = \text{diag}(I_r, O)$, 其中 $r \leq \min(m, n)$, 我们来看看对应的 f 是什么。

解答:

由于 α_1 在 S 下的坐标即为 $e_1^{(n)}$, 这里 $e_1^{(n)}$ 代表 \mathbb{K}^n 的第 1 个单位向量, 而计算可发现当 $r \geq 1$ 时有

$$A_{ST}e_1^{(n)} = e_1^{(m)}$$

V 中在 T 下坐标为 $e_1^{(m)}$ 的元素恰恰就是 β_1 , 于是 $f(\alpha_1) = \beta_1$ 。

同理, 对 $i = 1, \dots, r$, α_i 在 S 下的坐标为 $e_i^{(n)}$, 直接计算得

$$A_{ST}e_i^{(n)} = e_i^{(m)}, \quad i = 1, \dots, r$$

V 中在 T 下坐标为 $e_i^{(m)}$ 的元素恰恰就是 β_i , 于是

$$f(\alpha_i) = \beta_i, \quad i = 1, \dots, r$$

对 $i > r$, α_i 在 S 下的坐标为 $e_i^{(n)}$, 但此时

$$A_{ST}e_i^{(n)} = 0, \quad i = r + 1, \dots, n$$

由坐标的定义, V 中坐标是 0 的元素当然是零向量 (也记为 0, 注意表示的和 \mathbb{K}^m 中的零向量不同), 于是

$$f(\alpha_i) = 0, \quad i = r + 1, \dots, n$$

由于基映射可以确定线性映射, 上面的所有 $f(\alpha_i)$ 的像已经可以确定 f 。

在刚才的例子中, 我们似乎发现, 矩阵表示还有另一种等价的看法, 也就是利用 $A_{ST}e_i^{(n)}$ 为 A_{ST} 的第 i 列, A_{ST} 的每一列是 $f(\alpha_i)$ 在 T 下的坐标。

由此, 记 A_{ST} 的第 i 列为 a_i , 利用坐标定义可以发现

$$f(\alpha_i) = Ta_i$$

再将它们拼成矩阵, 利用形式乘法仍然满足分块特性可发现

$$(f(\alpha_1), f(\alpha_2), \dots, f(\alpha_n)) = TA_{ST}$$

这就是教材中常见的矩阵表示定义。若将左侧记为 $f(S)$ (视为 f 作用在 S 每一个分量形成的形式行向量), 可以更简洁地写成

$$f(S) = TA_{ST}$$

此定义同样是矩阵表示的计算方法。

我们总共介绍了三种矩阵表示的定义: $f = \pi_T^{-1} \circ A_{ST} \circ \pi_S$ 、 $f(u)_T = A_{ST}u_S$ 与 $f(S) = TA_{ST}$ 。三种定义的侧重点不同。第一种定义最为抽象, 但是也是本质性的, 是代数上的定义; 第二种定义最为直接, 揭示了矩阵表示的真正含义: 看作坐标的矩阵乘法, 在已知矩阵表示计算向量的像时最为有效; 第三种定义虽然具体, 但从中其实很难看出矩阵表示的含义, 适合已知映射计算矩阵表示。

* 我们并未严谨证明三种定义完全等价, 留给感兴趣的同学们。至少必须掌握后两种定义, 用于进行两个方向的计算。

最后, 如果大家对基变换的过渡矩阵有印象, 它的定义其实和矩阵表示很像: \mathbb{K} 上 n 维线性空间 U 的基 S 到 T 的基变换过渡矩阵 P 定义为

$$S = TP$$

可以发现, 只要让 $f(S) = S$ (从中可看出 f 是恒等映射), 它确实就可以是 f 在基 S 、 T 下的矩阵表示, 因此它可以理解为恒等映射在两组不同基下的矩阵表示。不过, 这样我们对同一个空间取了两组不同基, 稍显混乱, 如果让 $f(T) = S$, 它可以成为 f 在基 T 、 T 下的矩阵表示, 似乎会更好。

真的能做这样的选择吗? 答案是肯定的。注意到 $f(T) = S$ 意味着 f 将 T 中的第 i 个基映射到了 S 中的第 i 个基, 可以由这样的基映射定义线性映射。更进一步地, 由于它将一组基映射到了一组基, 它还是一个线性同构。由于它是 $U \rightarrow U$, 也就是 U 到自身的线性同构, 我们称为 U 的自同构。于是, 过渡矩阵更合理的看法是看作自同构在给定基下的矩阵表示。

19.3.3 线性映射的运算

那么，已经知道了线性映射可以怎样刻画以后，我们还需要知道什么呢？答案是，我们还需要知道线性映射之间的结构。简单来说，定义了矩阵后，我们还希望知道矩阵能进行怎样的运算，那我们自然也会关心线性映射之间的运算。

既然线性映射可以“看成”矩阵，如果从仿照矩阵运算的角度，对线性映射之间应该可以定义加法、数乘与乘法，下面我们将看到，这三件事对线性映射来说意味着什么，又是如何与矩阵的运算对应的。

1. 加法

在之前的学习中，我们已经很熟悉两个函数的加法了，简单来说， $(f+g)(x) = f(x) + g(x)$ ，函数的和在某个点处的值应该是分别求值再求和。我们将这种简单的定义延伸到线性映射中，考虑 \mathbb{K} 上的线性空间 U, V ，对 $f, g \in \text{Hom}(U, V)$ ，定义 $f+g$ 为（这里右侧的加法为 V 中加法）

$$\forall u \in U, \quad (f+g)(u) = f(u) + g(u)$$

它是一个 $\text{Hom}(U, V)$ 中的线性映射。

证明：

由于线性空间对加法的封闭性， $f(u) + g(u) \in V$ ， $f+g$ 还是一个 U 到 V 的映射，下面验证其为线性映射。对任何 $u_1, u_2 \in U$ 有

$$(f+g)(u_1+u_2) = f(u_1+u_2) + g(u_1+u_2) = f(u_1) + f(u_2) + g(u_1) + g(u_2) = (f+g)(u_1) + (f+g)(u_2)$$

对任何 $u \in U, \lambda \in \mathbb{K}$ 有

$$(f+g)(\lambda u) = f(\lambda u) + g(\lambda u) = \lambda f(u) + \lambda g(u) = \lambda(f+g)(u)$$

从而其确实在 $\text{Hom}(U, V)$ 中。

假设 f 在基 S, T 下的矩阵表示为 A ， g 在基 S, T 下的矩阵表示为 B ，此处我们希望计算 $f+g$ 的矩阵表示，因此利用第三个定义得

$$(f+g)(S) = f(S) + g(S) = TA + TB = T(A+B)$$

因此 $f+g$ 的矩阵表示为 $A+B$ 。这确实与我们的期望相符。

2. 数乘

对于一个函数的数乘，最直接的定义当然就是将函数每个点的值乘上数，也就是 $(\lambda f)(x) = \lambda f(x)$ 。考虑 \mathbb{K} 上的线性空间 U, V ，对 $f \in \text{Hom}(U, V)$ 与 $\lambda \in \mathbb{K}$ ，可定义 λf 为（这里右侧的数乘为 V 中数乘）

$$\forall u \in U, \quad (\lambda f)(u) = \lambda f(u)$$

它是一个 $\text{Hom}(U, V)$ 中的线性映射。

证明：

由于线性空间对数乘的封闭性， $\lambda f(u) \in V$ ， λf 还是一个 U 到 V 的映射，下面验证其为线性映射。对任何 $u_1, u_2 \in U$ 有

$$(\lambda f)(u_1+u_2) = \lambda f(u_1+u_2) = \lambda f(u_1) + \lambda f(u_2) = (\lambda f)(u_1) + (\lambda f)(u_2)$$

对任何 $u \in U, \mu \in \mathbb{K}$ 有

$$(\lambda f)(\mu u) = \lambda f(\mu u) = \lambda \mu f(u) = \mu(\lambda f(u)) = \mu(\lambda f)(u)$$

从而其确实在 $\text{Hom}(U, V)$ 中。

假设 f 在基 S 、 T 下的矩阵表示为 A , 此处我们希望计算 λf 的矩阵表示, 因此利用第三个定义得

$$(\lambda f)(S) = \lambda f(S) = \lambda TA = T(\lambda A)$$

因此 λf 的矩阵表示 λA 。这也确实与我们的期望相符。

3. 乘法

最后是一个稍显复杂的问题, 矩阵乘法如何用映射表示?

对于映射来说, 唯一一个能将映射放在一起的方法是复合。根据已经证明的性质, 对 \mathbb{K} 上的线性空间 U, V, W , $f \in \text{Hom}(U, V)$, $g \in \text{Hom}(V, W)$ 时, 的确有 $g \circ f \in \text{Hom}(U, W)$ 。

我们下面证明, 若 U, V, W 的各一组基是 R, S, T , 且 f 在 R, S 下的矩阵表示为 B , g 在 S, T 下的矩阵表示为 A , 则 $g \circ f$ 在 R, T 下的矩阵表示为 AB 。

证明:

由矩阵表示第三个定义可知

$$f(R) = SB, \quad g(S) = TA$$

从而

$$(g \circ f)(R) = g(f(R)) = g(SB)$$

利用形式乘法与 g 的线性性可以发现

$$g(SB) = g(S)B$$

从而进一步计算得

$$g(S)B = TAB = T(AB)$$

由此最终得到

$$(g \circ f)(R) = T(AB)$$

这就符合了矩阵表示的定义。

因此, 我们也的确得到了线性映射的复合对应矩阵乘法。

至此我们已经可以看出, 有限维线性空间上的线性映射利用矩阵表示, 可以与矩阵有诸多类似的性质。最后, 在 $\text{Hom}(U, V)$ 上可以定义加法、数乘后, 我们也希望它构成一个线性空间, 而这确实是成立的。

证明:

在刚才加法、数乘的定义中已经验证了封闭性, 从而只需验证线性空间的八条性质。我们将默认 f 、 g, h 为 $\text{Hom}(U, V)$ 的任意线性映射, 注意验证 U 到 V 的映射相等只需说明任何元素像相同。

加法结合律:

$$\forall u \in U, \quad ((f+g)+h)(u) = (f+g)(u)+h(u) = f(u)+g(u)+h(u) = f(u)+(g+h)(u) = (f+(g+h))(u)$$

加法交换律:

$$\forall u \in U, \quad (f+g)(u) = f(u)+g(u) = g(u)+f(u) = (g+f)(u)$$

存在零元: 记 $\mathcal{O}(x)$ 为把所有向量映射到 0 的映射, 可验证其线性, 从而 $\mathcal{O}(x) \in \text{Hom}(U, V)$, 且

$$\forall u \in U, \quad (\mathcal{O}+f)(u) = 0 + f(u) = f(u) = f(u) + 0 = (f+\mathcal{O})(u)$$

存在逆元: 对 f , 记 $-f$ 满足 $(-f)(u) = -f(u)$, 可验证其线性, 且 $-f(u) \in V$, 从而 $(-f)(u) \in \text{Hom}(U, V)$, 且

$$\forall u \in U, \quad (-f+f)(u) = (-f)(u)+f(u) = 0 = \mathcal{O}(u) = f(u)+(-f)(u) = (f+(-f))(u)$$

数乘单位元：

$$\forall u \in U, (1f)(u) = 1f(u) = f(u)$$

数乘结合律：

$$\forall u \in U, \lambda, \mu \in \mathbb{K}, (\lambda(\mu f))(u) = \lambda(\mu f(u)) = \lambda\mu f(u) = (\lambda\mu)f(u) = ((\lambda\mu)f)u$$

加法分配数乘：

$$\forall u \in U, \lambda, \mu \in \mathbb{K}, ((\lambda + \mu)f)(u) = (\lambda + \mu)f(u) = \lambda f(u) + \mu f(u) = (\lambda f + \mu f)(u)$$

数乘分配加法：

$$\forall u \in U, \lambda \in \mathbb{K}, (\lambda(f + g))(u) = \lambda(f + g)(u) = \lambda(f(u) + g(u)) = \lambda f(u) + \lambda g(u) = (\lambda f + \lambda g)(u)$$

综合八条得证。

* 这里 $\text{Hom}(U, V)$ 构成线性空间的所有验证步骤都无需假设有限维，因此即使对无穷维也成立。

不仅如此，我们还可以发现，上一部分定义的 $\varphi : \text{Hom}(U, V) \rightarrow \text{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ 事实上就是两个空间之间的线性同构。

证明：

由于已经验证了双射，只需说明线性性。利用线性映射的线性性与映射加法定义，对任何 $f, g \in \text{Hom}(U, V), x \in \mathbb{K}^n$ 有

$$\varphi(f + g)(x) = \pi_T((f + g)(\pi_S^{-1}(x))) = \pi_T(f(\pi_S^{-1}(x))) + \pi_T(g(\pi_S^{-1}))(x) = \varphi(f)(x) + \varphi(g)(x)$$

从而再由映射加法定义可知 $\varphi(f + g) = \varphi(f) + \varphi(g)$ 。同理，利用线性映射的线性性与映射数乘定义，对任何 $f \in \text{Hom}(U, V), \lambda \in \mathbb{K}, x \in \mathbb{K}^n$ 有

$$\varphi(\lambda f)(x) = \pi_T((\lambda f)(\pi_S^{-1}(x))) = \lambda \pi_T(f(\pi_S^{-1})) = \lambda \varphi(f)(x)$$

从而再由映射数乘定义可知 $\varphi(\lambda f) = \lambda \varphi(f)$ 。

综合两部分得证为线性映射，即得其为同构。

* 如果对这部分证明的思路还有疑问，可以参考本讲义 20.1.1 的第二道例题。

不仅如此，将 $\text{Hom}(U, V)$ 中的元素 \mathcal{A} 映射到矩阵 $A \in \mathbb{K}^{m \times n}$ ，使得 $\mathcal{A}(x) = Ax$ ，此映射也是线性同构。

证明：

由于已经验证了双射，只需说明线性性。计算可发现这里的 A 即为 \mathcal{A} 在基 $e_1^{(n)}, \dots, e_n^{(n)}, e_1^{(m)}, \dots, e_m^{(m)}$ 下的矩阵表示，利用矩阵表示的性质可知 $\mathcal{A} + \mathcal{B}$ 的矩阵表示为 $A + B$ ， $\lambda \mathcal{A}$ 的矩阵表示为 λA ，由此其为线性映射，即得其为同构。

利用同构的复合还是同构，我们最终得到，将 \mathbb{K} 有限维线性空间 U, V 的线性映射 $f \in \text{Hom}(U, V)$ 映到其在基 S, T 下的矩阵表示 A_{ST} 的映射是线性同构。

这个结论最大的作用是，可以用来计算出 $\text{Hom}(U, V)$ 的一组基。我们下面设 U 的一组基为 $S = (\alpha_1, \dots, \alpha_n)$ ， V 的一组基 $T = (\beta_1, \dots, \beta_m)$ 。

解答：

由于线性同构将一组基映射到一组基，且其逆还是线性同构。考虑 $\mathbb{K}^{m \times n}$ 的一组基，将其从矩阵表示还原回一组基，即为 $\text{Hom}(U, V)$ 的一组基。

设 $E_{ij} \in \mathbb{K}^{m \times n}$ 为只有第 i 行第 j 列为 1，其他为 0 的矩阵，则所有

$$E_{ij}, i = 1, \dots, m, j = 1, \dots, n$$

构成 $\mathbb{K}^{m \times n}$ 的一组基。我们假设 E_{ij} 还原到的线性映射为 f_{ij} , 下面计算 f_{ij} 。

类似上一部分中对 $A_{ST} = \text{diag}(I_r, O)$ 的讨论, 利用矩阵表示的第三种定义, 由

$$f_{ij}(S) = TE_{ij}$$

直接计算右侧可发现 (这里 β_i 是形式行向量的第 j 个分量)

$$f_{ij}(S) = (0, \dots, 0, \beta_i, 0, \dots, 0)$$

由此进一步对比分量可发现

$$f_{ij}(\alpha_k) = \begin{cases} 0 & k \neq j \\ \beta_i & k = j \end{cases}$$

若用 δ_{kj} 表示 $k = j$ 时为 1, $k \neq j$ 时为 0 的量, 可以进一步写成

$$f_{ij}(\alpha_k) = \delta_{jk}\beta_i$$

利用基映射构造线性映射即可以确定 f_{ij} 。

* 这种类似 $f_{ij}(\alpha_k) = \delta_{jk}\beta_i$ 的构造往往是由同构得到的, 因此常出现在映射构成的线性空间的基的构造上——因为很难像之前那样直观找到它的基, 我们还将在之后的例题中看到。

19.3.4 基变换与相抵

值得注意的是, 不管之前在如何推导, 我们都是在固定 U, V 一组基的情况下得到的。那么, 非常自然地, 我们希望知道基进行变换时矩阵表示会如何变化。

设 \mathbb{K} 上 n 维线性空间 U 的两组基为 S, S' , V 的两组基为 T, T' , $f \in \text{Hom}(U, V)$ 在基 S, T 下的矩阵表示为 A_{ST} , 在基 S', T' 下的矩阵表示为 $A_{S'T'}$, 若假设 S 到 S' 的过渡矩阵为 P , T 到 T' 的过渡矩阵为 Q , A_{ST} 与 $A_{S'T'}$ 有何关系呢?

证明:

根据定义我们可以写出如下的四个形式乘法:

$$f(S) = TA_{ST}, \quad f(S') = T'A_{S'T'}, \quad S' = SP, \quad T' = TQ$$

用后两个式子代入第二式可得

$$f(SP) = TQA_{S'T'}$$

利用 f 的线性性, 左侧即为 $f(S)P$, 由过渡矩阵可逆性, 同时右乘 P^{-1} 得到

$$f(S) = TQA_{S'T'}P^{-1} = T(QA_{S'T'}P^{-1})$$

根据矩阵表示唯一性也即得到

$$A_{ST} = QA_{S'T'}P^{-1}$$

即

$$A_{S'T'} = Q^{-1}A_{ST}P$$

由此我们可以发现, 同一个线性映射 f 在不同基下的矩阵表示相抵。另一方面, 之前已经证明了过渡矩阵可以取为任何可逆阵, 从而若 f 在 S, T 下的矩阵表示为 A_{ST} , 任何与 A_{ST} 相抵的矩阵都可以看作 f 在某两组基下的矩阵表示。

于是, 利用上一部分已经计算的 $A_{ST} = \text{diag}(I_r, O)$ 对应的 f , 相抵标准形可以解释为: 对任何 \mathbb{K} 上 n 维线性空间 U 到 m 维线性空间 V 的线性映射 f , 存在 U 的一组基 $\alpha_1, \dots, \alpha_n$, V 的一组基 β_1, \dots, β_m , 使得存在自然数 r 满足

$$f(\alpha_i) = \beta_i, \quad i = 1, \dots, r$$

$$f(\alpha_i) = 0, \quad i = r+1, \dots, n$$

我们还可以用空间直接证明此结论，来展示有限维线性空间理论与矩阵理论的等价性。

证明：

设 $r = n - \dim \text{Ker } f$ ，考虑 $\dim \text{Ker } f$ 的一组基， $\alpha_{r+1}, \dots, \alpha_n$ ，则利用 $\text{Ker } f$ 定义

$$f(\alpha_i) = 0, \quad i = r+1, \dots, n$$

下面，将其扩充为 U 的一组基，假设扩充的向量为 $\alpha_1, \dots, \alpha_r$ ，记

$$\beta_i = f(\alpha_i), \quad i = 1, \dots, r$$

只要证明了 β_1, \dots, β_r 线性无关，即可以把它们扩充为 V 的一组基，而上述的定义已经符合要求的形式。

设 $\lambda_1, \dots, \lambda_r \in \mathbb{K}$ 使得

$$\sum_{i=1}^r \lambda_i \beta_i = 0$$

利用线性性可发现

$$\sum_{i=1}^r \lambda_i f(\alpha_i) = f\left(\sum_{i=1}^r \lambda_i \alpha_i\right) = 0$$

于是

$$\sum_{i=1}^r \lambda_i \alpha_i \in \text{Ker } f$$

但是，根据补空间的算法， $\alpha_1, \dots, \alpha_r$ 生成的子空间应为 $\text{Ker } f$ 的补空间，从而两者交只有 0，得到

$$\sum_{i=1}^r \lambda_i \alpha_i = 0$$

再由 $\alpha_1, \dots, \alpha_r$ 线性无关可知所有 λ_i 全为 0，得证 β_1, \dots, β_r 线性无关，从而结论成立。

* 此证明与上学期的操作性证明思路完全不同，大家可以体会这个思路的更加本质性。

最后，既然 f 的任何矩阵表示都相抵，它任何矩阵表示的秩应当相同，从而也是可以对 f 合理定义的量。事实上，可以证明， f 任何矩阵表示的秩等于 $\dim \text{Im } f$ 。

证明：

由于对任何矩阵表示秩相同，而矩阵表示不影响 f 本身，从而 $\dim \text{Im } f$ 也恒定，只需证明矩阵表示为相抵标准形 $\text{diag}(I_r, O)$ 时 $\dim \text{Im } f = r$ 即可。

仍按上个定理假设，由 $f(U) + f(W) = f(U + W)$ 可知

$$\text{Im } f = f\left(\sum_{i=1}^n \langle \alpha_i \rangle\right) = \sum_{i=1}^n f(\langle \alpha_i \rangle)$$

而由于 $f(\langle \alpha_i \rangle)$ 即为所有 $f(\lambda \alpha_i)$ ，与所有 $\lambda f(\alpha_i) = \lambda \beta_i$ 相同，而 $\langle 0 \rangle = 0$ ，上式可进一步化为

$$\text{Im } f = \sum_{i=1}^r \langle \beta_i \rangle = \langle \beta_1, \dots, \beta_r \rangle$$

由于 β_1, \dots, β_r 是一组基的一部分，它们线性无关，从而右侧维数为 r ，因此左侧维数也为 r ，得证。

* 这个证明虽然写起来有一点抽象，核心过程 $f(\langle \alpha_1, \dots, \alpha_n \rangle) = \langle f(\alpha_1), \dots, f(\alpha_n) \rangle$ 其实是非常直观且易于验证的。

由此，仍设 $\dim U = n$ 、 $\dim V = m$ ，能得到用矩阵表示刻画同构的结论：

1. f 是单同态当且仅当其任何矩阵表示列满秩。

证明：

由于矩阵表示为 $\mathbb{K}^{m \times n}$, 列满秩即等价于秩为 n , 从而等价于

$$\dim \text{Im } f = n$$

利用第一同构定理, 上式又等价于

$$\dim \text{Ker } f = n - \dim \text{Im } f = 0$$

从而等价于 $\text{Ker } f = \{0\}$, 即等价于单同态。

2. f 是满同态当且仅当其任何矩阵表示行满秩。

证明：

由于矩阵表示为 $\mathbb{K}^{m \times n}$, 列满秩即等价于秩为 m , 从而等价于

$$\dim \text{Im } f = m = \dim V$$

由于 $\text{Im } f$ 是 V 的子空间, 上式即等价于 $\text{Im } f = V$, 即等价于满同态。

3. f 是同构当且仅当其任何矩阵表示可逆。

证明：

由于同构既是单同态又是满同态, 由上面两部分证明, f 的任何矩阵表示既列满秩又行满秩, 由上学期知识即得其可逆。

* 方阵的行满秩、列满秩、可逆相互等价, 从而 U 、 V 维数相同时单同态、满同态、同构相互等价。

二十 补充：从映射到变换

§20.1 一般线性映射

20.1.1 例题

1. 若 U, V 是 \mathbb{K} 上线性空间, f 是 $U \rightarrow V$ 的线性映射, W 是 U 的子空间, 且 $\text{Ker } f \oplus W = U$, 证明 $f|_{W \rightarrow \text{Im } f}$ 是同构。

2. 设 U, V 均为 \mathbb{K} 上的线性空间, S 为 U 的一组基构成的集合, 对 $S \rightarrow V$ 的所有映射集合 $\text{Map}(S, V)$ 定义加法、数乘

$$(f + g)(x) = f(x) + g(x)$$

$$(\lambda f)(x) = \lambda f(x)$$

(a) 证明 $\text{Map}(S, V)$ 构成线性空间, 且与 $\text{Hom}(U, V)$ 同构。

(b) 对某有限集 $X = \{x_1, \dots, x_n\}$ 与 m 维 \mathbb{K} 线性空间 V , 与上相同将 $\text{Map}(X, V)$ 定义为线性空间, 计算其维数与一组基。

3. 设 U, V 均为 \mathbb{K} 上的线性空间, 给定非零向量 $\alpha \in U$, 考虑 $\text{Hom}(U, V)$ 到 V 的映射 $A_\alpha(f) = f(\alpha)$ 。

(a) 证明 A_α 为线性映射;

(b) 计算 A_α 的像与核, 并写出相应的第一同构定理;

(c) 若 U, V 分别为 m 维、 n 维，计算 $\text{Ker } A_\alpha$ 的维数与一组基。

4. 给定 $a \in \mathbb{R}$ ，考虑 $\mathbb{R}[x] \rightarrow \mathbb{R}[x]$ 的映射 S, D, T_a 使得

$$S(f) = \int_0^x f(t) dt, \quad D(f) = f', \quad T_a(f(x)) = f(x+a)$$

- (a) 验证 S, D, T_a 是线性映射；
 - (b) 计算 S, D, T_a 的像与核；
 - (c) 证明 $D \circ S, T_a$ 都是同构， $S \circ D$ 不是同构。
 - (d) 找最大的 $\mathbb{R}[x]$ 子空间 W 使得 $(S \circ D)|_{W \rightarrow W}$ 是同构，这里最大是指，只要 $(S \circ D)|_{V \rightarrow V}$ 是同构，就有 $V \subset W$ 。
5. 将 \mathbb{C} 看作 \mathbb{R} 上的二维线性空间（可见本讲义 17.4.1），通过第一同构定理证明 $\mathbb{R}[x]/(x^2 + 1)$ 同构于 \mathbb{C} ， $(x^2 + 1)$ 代表 $x^2 + 1$ 的所有倍式集合构成的 $\mathbb{R}[x]$ 的子空间。
6. 对 \mathbb{K} 上线性空间 U, V 之间的线性映射 f ，证明存在线性空间 W 使得 f 可以分解为单同态 $g : W \rightarrow V$ 与满同态 $h : U \rightarrow W$ 的复合，也即 $f = g \circ h$ 。给出此结论的矩阵版本。

20.1.2 解答

1. 由像空间定义 $f(W) \subset \text{Im } f$ ，因此限制映射的确存在。为证明同构只需说明是双射。

单射：若 $f|_{W \rightarrow \text{Im } f}(x) = 0$ ，由限制映射定义可知 $x \in W$ 且 $f(x) = 0$ ，于是根据核空间定义 $x \in \text{Ker } f \cap W$ ，由直和可知 $x = 0$ ，于是 $\text{Ker } f|_{W \rightarrow \text{Im } f} = 0$ ，为单射。

满射：对任何 $x \in \text{Im } f$ ，设 $f(y) = x$ ，由 U 为 $\text{Ker } f, W$ 直和可知存在 $z \in \text{Ker } f, w \in W$ 使得 $y = z + w$ ，由 $\text{Ker } f$ 定义 $f(w) = f(z) + f(w) = f(z + w) = x$ ，从而 $f|_{W \rightarrow \text{Im } f}(w) = x$ ，因此 x 有原像，为满射。

2. (a) 线性空间的验证与本讲义 19.3.3 过程基本相同，下面证明同构，我们采用直接构造同构映射的方式——此映射的构造事实上是基于基映射可以确定线性映射，而本问即是基映射可以确定线性映射这件事的形式化表述。

考虑 $\text{Hom}(U, V)$ 到 $\text{Map}(S, V)$ 的映射 L 满足 $L(\varphi) = \varphi|_{S \rightarrow V}$ ，我们下面来验证这是同构。为了书写清晰，我们将用中括号代表映射到映射的映射，即将 $L(\varphi)$ 记为 $L[\varphi]$ 。

* 注意此处由于 S 是 U 的子集， φ 在 $S \rightarrow V$ 的限制映射是可以定义的，也的确是 S 到 V 的映射，但由于 S 并非线性空间， $\varphi|_{S \rightarrow V}$ 并非线性映射。

- 线性性

为了说明线性性，我们需要先证明 $L[\varphi + \psi] = L[\varphi] + L[\psi]$ 对任何 $\varphi, \psi \in \text{Hom}(U, V)$ 成立，这里左右侧分别为 $\text{Hom}(U, V)$ 与 $\text{Map}(S, V)$ 中的加法。由于左右都是 $S \rightarrow V$ 的映射，要说明映射相等只需证明在任何元素下的像都相等。对任何 $s \in S$ 有

$$\begin{aligned} L[\varphi + \psi](s) &= (\varphi + \psi)|_{S \rightarrow V}(s) = (\varphi + \psi)(s) = \varphi(s) + \psi(s) \\ &= \varphi|_{S \rightarrow V}(s) + \psi|_{S \rightarrow V}(s) = (\varphi|_{S \rightarrow V} + \psi|_{S \rightarrow V})(s) = (L[\varphi] + L[\psi])(s) \end{aligned}$$

这里六个等号的依据依次是： L 的定义、限制映射定义、 $\text{Hom}(U, V)$ 加法定义、限制映射定义、 $\text{Map}(S, V)$ 加法定义、 L 的定义。

对数乘，我们也采用类似的方式。对 $\lambda \in \mathbb{K}$ ，考虑任何 $s \in S$ ，有

$$L[\lambda\varphi](s) = (\lambda\varphi)|_{S \rightarrow V}(s) = (\lambda\varphi)(s) = \lambda\varphi(s) = \lambda(\varphi|_{S \rightarrow V}(s)) = (\lambda\varphi|_{S \rightarrow V})(s) = (\lambda L[\varphi])(s)$$

这里六个等号的依据依次是： L 的定义、限制映射定义、 $\text{Hom}(U, V)$ 数乘定义、限制映射定义、 $\text{Map}(S, V)$ 数乘定义、 L 的定义。

综合两部分可得证线性。

- 单射

即要证明 $\text{Ker } L = \{0\}$ 。若 $\varphi \in \text{Ker } L$, 由定义可知

$$\varphi|_{S \rightarrow V} = 0$$

也即 φ 在一组基上均为 0, 不妨设这组基为 $\{\alpha_i \mid i \in I\}$ 。由于任何元素都可以写成基的线性组合, 利用 φ 的线性性可知

$$\varphi\left(\sum_{k=1}^n \lambda_k \alpha_{i_k}\right) = \sum_{k=1}^n \lambda_k \varphi(\alpha_{i_k}) = 0$$

从而 φ 是零映射, 得证。

- 满射

本讲义 19.2.2 中已经证明了可以用基映射定义线性映射, 而对任何 $f \in \text{Map}(S, V)$, 用 f 作为基映射定义的线性映射 φ 即满足 $\varphi|_{S \rightarrow V} = f$, 从而 $L[\varphi] = f$, 因此其为满射。

(b) 计算映射构成线性空间的基有两种方式, 根据同构(利用同构将一组基映射到一组基)或直接构造, 前者的困难在于找到同构的空间, 后者的困难在于写出合适的形式, 我们这里给出两种方法:

- 同构方法

由于已知了 $\text{Map}(X, V)$ 与 X 作为基的 $\text{Hom}(U, V)$ 同构, 我们可以不妨假设 x_1, \dots, x_n 是某个 \mathbb{K} 上的 n 维线性空间 U 的一组基(由于这里 x_i 是某种虚指标, 具体取何种元素并不影响映射的构造, 这个做法是合理的)。

在本讲义 19.3.3 中, 我们已经给出了 $\text{Hom}(U, V)$ 的一组基, 当 U 的基为 $\{x_1, \dots, x_n\}$ 、 V 的一组基为 $\{\beta_1, \dots, \beta_m\}$ 时, $\text{Hom}(U, V)$ 的一组基对应的线性映射 $\varphi_{ij} : U \rightarrow V$ 满足

$$\varphi_{ij}(x_k) = \begin{cases} 0 & k \neq j \\ \beta_i & k = j \end{cases}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

而根据 (a) 中构造的同构, 将它们作限制映射即得到对应的 $f : X \rightarrow V$ 为

$$f_{ij}(x_k) = \begin{cases} 0 & k \neq j \\ \beta_i & k = j \end{cases}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

这就是 $\text{Map}(X, V)$ 的一组基, 个数为 mn , 因此其维数为 mn 。

- 直接构造法

仍设 V 的一组基为 β_1, \dots, β_m , 我们尝试直接构造 $\text{Map}(X, V)$ 的一组基。

利用 V 的基, 每个映射 $f : X \rightarrow V$ 都可以写成

$$\forall i = 1, \dots, n \quad f(x_i) = \sum_{j=1}^m \lambda_{ij} \beta_j$$

这里坐标 $\lambda_{ij} \in \mathbb{K}$ 。

可以发现, 由于两个映射求和对应像求和, 利用坐标的性质也即每个 x_i 的像坐标求和, 而映射的数乘即对应 x_i 的像坐标数乘, 由此可以考虑只有一个 $\lambda_{ij} = 1$, 其他为 0 的映射

$$f_{ij}(x_i) = \beta_j, \quad f_{ij}(x_k) = 0, k \neq i$$

我们下面说明所有这些映射构成 $\text{Map}(X, V)$ 的一组基。即得到维数为 mn 。

- 表出所有向量

考虑某个其能表出的向量

$$f = \sum_{ij} \lambda_{ij} f_{ij}$$

直接计算

$$f(x_k) = \sum_{ij} \lambda_{ij} f_{ij}(x_k)$$

由于 $i \neq k$ 时 $f_{ij}(x_k) = 0$, 上式事实上只存在 $i = k$ 的项, 进一步根据映射加法定义

$$f(x_k) = \sum_{j=1}^m \lambda_{kj} f_{kj}(x_k) = \sum_{j=1}^m \lambda_{kj} \beta_j$$

由此, 对 $f \in \text{Map}(X, V)$, 设对每个 k , $f(x_k) = \sum_{j=1}^m \mu_{kj} \beta_j$, 有

- 线性无关性

若 $0 = \sum_{ij} \lambda_{ij} f_{ij}$, 根据上方计算可知

$$0 = \sum_{j=1}^m \lambda_{kj} \beta_j$$

对任何 $k = 1, \dots, n$ 成立, 而利用 β_1, \dots, β_m 的线性无关性即得必须所有 λ_{kj} 为 0。

* 这里的 f_{ij} 定义对应上一种方法中的 f_{ji} 。

* 映射空间的一组基往往先考虑映射只在某个元素/基为 $\mathbf{0}$ 的情况, 因为这些情况可以组合出全部情况。而这种情况下又可以考虑其取为右侧每个基, 这样得到的就是线性无关且能组合出全部的映射了。不过, 这种做法往往比同构更加麻烦。

3. (a) 根据映射加法、数乘定义, 对任何 $f, g \in \text{Hom}(U, V)$ 、 $\lambda \in \mathbb{K}$, 有

$$A_\alpha(f + g) = (f + g)(a) = f(a) + g(a) = A_\alpha(f) + A_\alpha(g)$$

$$A_\alpha(\lambda f) = (\lambda f)(a) = \lambda f(a) = \lambda A_\alpha(f)$$

从而得证。

(b) * 为了简单起见, 我们承认选择公理成立, α 可以扩充为 U 的一组基。

先证明 $\text{Im } A_\alpha = V$ 。将 α 扩充为 U 的一组基 $\alpha \cup \{u_i \mid i \in I\}$ 。对任何 $v \in V$, 指定 $f(\alpha) = v$ 、 $f(u_i) = 0, i \in I$, 并用基映射定义线性映射, 则得到的 f 由定义满足 $A_\alpha(f) = v$, 即得证成立。

对 $\text{Ker } A_\alpha$, 由定义可知

$$\text{Ker } A_\alpha = \{f \in \text{Hom}(U, V) \mid f(\alpha) = 0\} = \{f \in \text{Hom}(U, V) \mid \alpha \in \text{Ker } f\}$$

。

由此第一同构定理为 $\text{Hom}(U, V)/\{f \in \text{Hom}(U, V) \mid \alpha \in \text{Ker } f\}$ 与 V 同构。

(c) 利用第一同构定理可直接得到 $\dim \text{Ker } A_\alpha = \dim \text{Hom}(U, V) - \dim V = mn - n$ 。下面计算其一组基。

* 由于已知根据 U 、 V 的一组基可以通过同构构造出 $\text{Hom}(U, V)$ 的一组基, 只要从这组基进行研究即可。此外, 由于已知了某个特殊的非零向量 α , 我们当然希望取的 U 的基是包含 α 的, 这就有了下面的做法。

考虑将 α 扩充为 U 的一组基 $\alpha_1 = \alpha, \alpha_2, \dots, \alpha_m$, V 的一组基 β_1, \dots, β_n , 本讲义已构造出 $\text{Hom}(U, V)$ 的一组基

$$f_{ij}(\alpha_k) = \begin{cases} 0 & k \neq j \\ \beta_i & k = j \end{cases}, \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

利用定义可发现，这组基中满足 $j \neq 1$ 的 f_{ij} 都有 $f_{ij}(\alpha_1) = f_{ij}(\alpha) = 0$ ，从而在 $\text{Ker } A_\alpha$ 中。

上述向量共有 $n(m-1)$ 个，等于 $\dim \text{Ker } A_\alpha$ ，且线性无关，从而它们就是 $\text{Ker } A_\alpha$ 的一组基，写为

$$f_{ij}(\alpha_k) = \begin{cases} 0 & k \neq j \\ \beta_i & k = j \end{cases}, \quad i = 1, \dots, n, \quad j = 2, \dots, m$$

* 此题不太需要考虑同构或直接的思路，因为是在已知原空间基时构造子空间的一组基，只要取出原空间合适的基并组合即可。

4. (a) 利用分析知识可知多项式求导、不定积分或代入 $x+a$ 后仍为多项式，从而三者均为 $\mathbb{R}[x] \rightarrow \mathbb{R}[x]$ 的映射，只需验证线性性。

直接验证对任何 $f, g \in \mathbb{R}[x]$, $\lambda \in \mathbb{R}$, $x \in \mathbb{R}$ 有

$$S(f+g)(x) = \int_0^x (f(t)+g(t)) dt = \int_0^x f(t) dt + \int_0^x g(t) dt = S(f)(x) + S(g)(x) = (S(f) + S(g))(x)$$

$$S(\lambda f)(x) = \int_0^x (\lambda f(t)) dt = \lambda \int_0^x f(t) dt = \lambda S(f)(x) = (\lambda S(f))(x)$$

$$D(f+g) = (f+g)' = f' + g' = D(f) + D(g)$$

$$D(\lambda f) = (\lambda f)' = \lambda f' = \lambda D(f)$$

$$T_a(f+g)(x) = (f+g)(x+a) = f(x+a) + g(x+a) = T_a(f)(x) + T_a(g)(x) = (T_a(f) + T_a(g))(x)$$

$$T_a(\lambda f)(x) = (\lambda f)(x) = \lambda f(x) = \lambda T_a(f)(x) = (\lambda T_a(f))(x)$$

从而均为线性映射。

- (b) • $\text{Im } S$

利用分析知识可发现

$$\int_0^x \sum_{i=0}^n a_i t^i dt = \sum_{i=0}^n \frac{a_i}{i+1} x^{i+1}$$

从而 $\text{Im } S$ 中所有多项式常数项都为 0，且对常数项为 0 的多项式 f 计算有 $S(f') = f$ ，于是利用多项式零次项为 0 处值可知 $\text{Im } S = \{f \in \mathbb{R}[x] \mid f(0) = 0\}$ ，或写为 $\text{Im } S = (x)$ ，这里 (x) 为 x 的倍式集合。类似验证 $\mathbb{R}[x]$ 的基可知其一组基为 $\{x^i \mid i \in \mathbb{N}^+\}$ 。

- $\text{Ker } S$

利用上方的分析知识计算结果即知多项式在 0 到 x 积分为 0 当且仅当其为 0，从而 $\text{Ker } S = \{0\}$ 。

* 上述两部分讨论说明 S 是单射，但并不是满射，与有限维相同维数的线性空间上的线性映射单射、满射等价不同。

- $\text{Im } D$

利用分析知识直接计算可发现对任何 $f \in \mathbb{R}[x]$ ，取 $g = S(f)$ ，则 $D(g) = f$ ，于是 $\text{Im } D = \mathbb{R}[x]$ 。

- $\text{Ker } D$

利用分析知识即得多项式导数为 0 当且仅当其为常数，即 $\text{Ker } D = \mathbb{R}[x]_1$ (回顾之前的定义，这代表小于 1 次的多项式，即常数)。

* 同样，这说明 D 是满射，但并不是单射。

- $\text{Im } T_a$ 、 $\text{Ker } T_a$

见 (c) 中同构的验证，验证为同构后即得其为单射、满射，从而 $\text{Im } T_a = \mathbb{R}[x]$, $\text{Ker } T_a = \{0\}$ 。

- (c) • $D \circ S$ 是同构

利用分析知识直接计算验证可知 $D(S(f)) = f$ 对任何 $f \in \mathbb{R}[x]$ 成立，于是其即为恒等映射，由定义直接验证它是双射，因此是同构。

- T_a 是同构

这里我们采取直接构造逆映射的方式验证同构。由于 $a \in \mathbb{R}$, T_{-a} 也是 $\mathbb{R}[x] \rightarrow \mathbb{R}[x]$ 的线性映射, 且对任何 $f \in \mathbb{R}[x]$ 、 $x \in \mathbb{R}$ 有

$$T_{-a}(T_a(f))(x) = T_a(f)(x - a) = f(x - a + a) = f(x)$$

$$T_a(T_{-a}(f))(x) = T_{-a}(f)(x + a) = f(x + a - a) = f(x)$$

从而 T_{-a} 就是 T_a 的逆映射, 因此 T_a 为同构。

* 如果能看出逆映射的构造, 验证映射互逆往往比直接计算像与核更加简单。

- $S \circ D$ 不是同构

$S(D(1)) = S(0) = 0$, 从而 $1 \in \text{Ker } S \circ D$, 因此其不是单射, 不是同构。

* 本题说明, 无穷维线性空间单侧复合是恒等未必能推出互逆。

(d) 首先, 记 $W = \{f \in \mathbb{R}[x] \mid f(0) = 0\}$, 类似 (b) 中直接计算可得对任何 $g \in W$ 有 $S(D(g)) = g$, 从而 $S(D(W)) = W \subset W$, 限制映射可定义, 且 $(S \circ D)|_{W \rightarrow W}$ 是恒等映射, 因此是同构。下面说明 W 就是最大的符合要求的子空间。

考虑 V 使得 $(S \circ D)|_{V \rightarrow V}$ 是同构, 为说明 $V \subset W$, 只需说明 V 中不包含常数项非零的多项式, 这样由 W 是所有常数项为 0 的多项式即知 $V \subset W$ 。

若 V 中有多项式 $f(x) = \sum_{i=0}^n a_i x^i$, 且 $a_0 \neq 0$, 由限制映射可定义知 $S(D(V)) \subset V$, 于是 $S(D(f)) \in V$, 直接计算可知

$$g(x) = \sum_{i=1}^n a_i x^i \in V$$

再由子空间封闭性, $f(x) - g(x) \in V$, 这就得到了 $a_0 \in V$ 。但是, $S(D(a_0)) = S(0) = 0$, 再由 $a_0 \in V$ 与限制映射定义得

$$a_0 \in \text{Ker } (S \circ D)_{V \rightarrow V}$$

若其为同构, Ker 应只有 0, 矛盾。

5. 定义映射 $A : \mathbb{R}[x] \rightarrow \mathbb{C}$, 使得 $A(f) = f(i)$, 与 3(a) 完全类似可验证其为线性映射, 下面计算 $\text{Im } A$ 与 $\text{Ker } A$ 。

- $\text{Im } A$

对任何 $c \in \mathbb{C}$, 设 $c = a + bi$, 且 $a, b \in \mathbb{R}$, 则取 $f(x) = bx + a$ 即有 $f(x) = c$, 于是 $\text{Im } A = \mathbb{C}$ 。

- $\text{Ker } A$

若 $f(x) \in (x^2 + 1)$, 设 $f(x) = (x^2 + 1)g(x)$, g 为多项式, 则 $f(i) = (i^2 + 1)g(i) = 0$, 因此 $f \in \text{Ker } A$ 。

若 $f \in \text{Ker } A$, 由定义 $f(i) = 0$, 设 $f(x) = \sum_{k=0}^n a_k x^k$, $a_k \in \mathbb{R}$, 可知

$$\sum_{k=0}^n a_k i^k = 0$$

取共轭得到

$$\sum_{k=0}^n a_k (-i)^k = 0$$

从而 $f(-i) = 0$ 。

* 这是著名的实系数多项式虚根成对结论。

考虑 f 看作 $\mathbb{C}[x]$ 上的多项式进行因式分解, 由条件知必有 $x - i$ 、 $x + i$ 因式, 因此看作 $\mathbb{C}[x]$ 上的多项式

$$x^2 + 1 = (x - i)(x + i) \mid f(x)$$

此外，由于 $f(x)$ 对 $x^2 + 1$ 进行带余除法的商与余式都在 $\mathbb{R}[x]$ 中，看作 $\mathbb{R}[x]$ 上的多项式整除仍成立，从而 $f \in (x^2 + 1)$ 。

综合以上两部分得 $\text{Ker } A = (x^2 + 1)$ 。

由此，利用第一同构定理即得 $\mathbb{R}[x]/(x^2 + 1)$ 与 \mathbb{C} 同构。

6. 构造是直接的，取 $W = \text{Im } f \subset V$ ，并定义

$$h : U \rightarrow \text{Im } f, \quad h = f|_{U \rightarrow \text{Im } f}$$

$$g : \text{Im } f \rightarrow V, \quad \forall v \in \text{Im } f, \quad g(v) = v$$

下面进行验证：

- h 为满同态

由于 $f(U) \subset \text{Im } f$ ，限制映射定义合理，且左右均为子空间，从而为线性映射。根据 $\text{Im } f$ 的定义，对任何 $v \in \text{Im } f$ 存在 $u \in U$ 使得 $f(u) = v$ ，从而 $h(u) = v$ ， h 为满同态。

- g 为单同态

由 g 的定义可直接验证线性，且 $g(v) = 0$ 当且仅当 $v = 0$ ，从而 $\text{Ker } g = \{0\}$ ， g 为单同态。

- $f = g \circ h$

首先，由于 $g : \text{Im } f \rightarrow V$ 、 $h : U \rightarrow \text{Im } f$ ，的确有 $g \circ h \in \text{Hom}(U, V)$ 。

对任何 $u \in U$ ，有

$$g(h(u)) = g(f(u)) = f(u)$$

从而得证相等。

* 注意验证映射相等一定要先验证定义域与陪域相同。

最后，我们来考虑这个结论的矩阵版本。若考虑 U 、 V 、 W 的各一组基 R 、 S 、 T ， f 在 R 、 S 下的矩阵表示为 A ， g 在 T 、 S 下的矩阵表示为 P ， h 在 R 、 T 下的矩阵表示为 Q ，则利用矩阵表示与映射复合的关系可知

$$A = PQ$$

此外，由本讲义 19.3.4 节的结论， P 是列满秩的， Q 是行满秩的，这就是矩阵的满秩分解的存在性（对矩阵 A 存在列满秩矩阵 P 、行满秩矩阵 Q 使得 $A = PQ$ ），上学期我们用相抵标准形证明了这个结论。

* 在我们的构造中，由于 $W = \text{Im } f$ ，可知 $\dim W = \dim \text{Im } f = \text{rank } A$ 。此外，由于给定 A 相当于给定了 f 与 U 、 V 的基，而这里给定的 V 的基未必包含 $\text{Im } f$ 的基，于是 g 虽然定义简单，未必有很简单的矩阵表示。

§20.2 有限维空间线性映射

20.2.1 限制映射 II

上一节的习题主要是关于一般线性映射的，而本学期的主题仍然是在有限维线性空间之间的线性映射上。在进入本节的习题前，我们先补充一个上一章没来得及介绍的知识：将限制映射与矩阵表示结合能得到什么。

我们假设 U 、 V 分别是 \mathbb{K} 上的 n 、 m 维线性空间， $f \in \text{Hom}(U, V)$ ，且 U_0 是 U 的 $s \leq n$ 维子空间， V_0 是 V 的 $r \leq m$ 维子空间，满足 $f(U) \subset V$ ，从而限制映射 $f|_{U \rightarrow V}$ 存在。

为了让限制映射的矩阵表示存在意义，我们自然不能乱取 U 、 V 的基，而是应让它们与 U_0 、 V_0 的基存在关系。具体来说，设 U_0 一组基为 $S_0 = (\alpha_1, \dots, \alpha_s)$ ，将它们扩充为 U 一组基 $S = (\alpha_1, \dots, \alpha_n)$ ； V_0 一组基为 $T_0 = (\beta_1, \dots, \beta_r)$ ，将它们扩充为 V 一组基 $T = (\beta_1, \dots, \beta_m)$ 。我们用 S^c 、 T^c 表示扩充出的部分 $(\alpha_{s+1}, \dots, \alpha_n)$ 、 $(\beta_{r+1}, \dots, \beta_m)$ ，则 $f|_{U_0 \rightarrow V_0}$ 在 S_0 、 T_0 下的矩阵表示 $A_0 \in \mathbb{K}^{r \times s}$ 应满足

$$f|_{U_0 \rightarrow V_0}(S_0) = T_0 A_0$$

利用限制映射的定义与形式乘法的分块即有

$$f(S_0) = T_0 A_0 = (T_0, T^c) \begin{pmatrix} A_0 \\ O \end{pmatrix} = T \begin{pmatrix} A_0 \\ O \end{pmatrix}$$

而另一方面 f 的矩阵表示 $A \in \mathbb{K}^{m \times n}$ 满足

$$f(S) = (f(S_0), f(S^c)) = TA$$

再次利用形式乘法的分块即得到（这里 $B \in \mathbb{K}^{r \times (n-s)}$ ， $C \in \mathbb{K}^{(m-r) \times (n-s)}$ 为某个未知矩阵）

$$f(S) = T \begin{pmatrix} A_0 & B \\ O & C \end{pmatrix}$$

于是

$$A = \begin{pmatrix} A_0 & B \\ O & C \end{pmatrix}$$

* 对限制映射、分块等不熟的同学们可以写成求和的形式直接推导。

由此我们得到，若 $f|_{U_0 \rightarrow V_0}$ 存在，则 f 可以在适当的基下矩阵表示为分块三角阵，且左上角的对角块即为限制映射的矩阵表示。

上述结论的逆命题也成立，假设 f 在 U 的一组基 $\alpha_1, \dots, \alpha_m$ 、 V 的一组基 β_1, \dots, β_n 下矩阵表示为

$$A = \begin{pmatrix} A_0 & B \\ O & C \end{pmatrix}$$

且 A_0 是 $r \times s$ 阶矩阵，我们可以构造限制映射使得 A_0 为其矩阵表示。

* 这里我们用求和直接推导，同样，对求和操作不熟的同学们可以尝试用形式乘法操作。

设 $U_0 = \langle \alpha_1, \dots, \alpha_s \rangle$ 、 $V_0 = \langle \beta_1, \dots, \beta_r \rangle$ ，由于基的线性无关性可知 $\alpha_1, \dots, \alpha_s$ 构成 U_0 的基， β_1, \dots, β_s 构成 V_0 的基。我们下面证明 $f|_{U_0 \rightarrow V_0}$ 的矩阵表示就是 A_0 。

首先，设 A 的第 i 行第 j 列为 a_{ij} ， A_0 的第 i 行第 j 列为 a'_{ij} ，根据分块的形式可知

$$a'_{ij} = a_{ij}, \quad i = 1, \dots, r, \quad j = 1, \dots, s$$

利用矩阵表示的定义可知

$$\forall j = 1, \dots, m, \quad f(\alpha_j) = \sum_{i=1}^n a_{ij} \beta_i$$

对任何 $j = 1, \dots, s$ ，可发现 $a_{ij} = 0$ 对 $i = r+1, \dots, m$ 成立，于是实际上有

$$\forall j = 1, \dots, s, \quad f(\alpha_j) = \sum_{i=1}^r a_{ij} \beta_i$$

由于右端都是由 β_1, \dots, β_r 生成，可知对 $j = 1, \dots, s$ 都满足 $f(\alpha_j) \in V_0$ ，从而它们生成的空间也满足，即 $f(U_0) \subset V_0$ ，可以进行限制。此外，利用限制映射定义与 $a'_{ij} = a_{ij}$ 可知

$$\forall j = 1, \dots, s, \quad f|_{U_0 \rightarrow V_0}(\alpha_j) = \sum_{i=1}^r a'_{ij} \beta_i$$

这即是 $f|_{U_0 \rightarrow V_0}$ 的矩阵表示为 A_0 的定义。

经过以上两部分，我们实际上得到的结论是，有限维时限制映射与分块三角阵等价。由于限制映射不止可以在有限维线性空间定义，它可以看作某种分块三角阵的推广，在之后的推导中，能用限制映射操作时，我们往往就不再显式写出分块三角阵了。不过，这样的限制映射终究还是只刻画了原映射的部分信息，我们能否有机会用线性映射刻画原映射的全部信息呢？

答案是肯定的。考虑如下情况： U_1, U_2 是 U 的两个互补的子空间， V_1, V_2 是 V 的两个互补的子空间，且满足 $f(U_1) \subset V_1, f(U_2) \subset V_2$ 。

假设 U_1 的一组基为 $S_1 = (\alpha_1, \dots, \alpha_s)$ ， U_2 的一组基为 $S_2 = (\alpha_{s+1}, \dots, \alpha_n)$ ， V_1 的一组基为 $T_1 = (\beta_1, \dots, \beta_r)$ ， V_2 的一组基为 $T_2 = (\beta_{r+1}, \dots, \beta_m)$ ，设 $f|_{U_1 \rightarrow V_1}$ 在 S_1, T_1 下的矩阵表示为 A_1 ， $f|_{U_2 \rightarrow V_2}$ 在 S_2, T_2 下的矩阵表示为 A_2 ，下面我们来计算 f 在 $S = (S_1, S_2)$ 、 $T = (T_1, T_2)$ 下的矩阵表示。

解答：

我们用形式乘法的方式写出结果。与之前类似，我们可以写出

$$\begin{aligned} f(S_1) &= T_1 A_1 = (T_1, T_2) \begin{pmatrix} A_1 \\ O \end{pmatrix} = T \begin{pmatrix} A_1 \\ O \end{pmatrix} \\ f(S_2) &= T_2 A_2 = (T_1, T_2) \begin{pmatrix} O \\ A_2 \end{pmatrix} = T \begin{pmatrix} O \\ A_2 \end{pmatrix} \\ f(S) &= (f(S_1), f(S_2)) = T \begin{pmatrix} A_1 & O \\ O & A_2 \end{pmatrix} \end{aligned}$$

从而矩阵表示即为 $\text{diag}(A_1, A_2)$ ，这是一个分块对角阵。

同样的，若已知 f 在基 $S = (\alpha_1, \dots, \alpha_n)$ 、 $T = (\beta_1, \dots, \beta_m)$ 下的矩阵表示 $A = \text{diag}(A_1, A_2)$ ，其中 A_1 是 $r \times s$ 阶矩阵，我们也可以给出对应的限制映射。

解答：

设 $U_1 = \langle \alpha_1, \dots, \alpha_s \rangle$ 、 $U_2 = \langle \alpha_{s+1}, \dots, \alpha_n \rangle$ ， $V_1 = \langle \beta_1, \dots, \beta_r \rangle$ 、 $V_2 = \langle \beta_{r+1}, \dots, \beta_m \rangle$ 与之前类似，由于左下角为 O ，可以得到 $f(U_1) \subset V_1$ ，且 A_1 为 $f|_{U_1 \rightarrow V_1}$ 的矩阵表示；类似可推得 $f(U_2) \subset V_2$ ，且 A_2 为 $f|_{U_2 \rightarrow V_2}$ 的矩阵表示。利用基的关系可知 $U_1 \oplus U_2 = U$ 、 $V_1 \oplus V_2 = V$ ，从而这就得到了分解。

由此，有限维时分块对角等价于在子空间和补空间上有独立限制，这里独立是指 $f(U_1) \cap f(U_2) = \{0\}$ ，否则将无法分出符合要求的 V_1, V_2 。

20.2.2 投影映射

在刚才的讨论中，我们可以发现一个有趣的问题。在假设 U_1, U_2 互补， V_1, V_2 互补，且 $f(U_1) \subset V_1, f(U_2) \subset V_2$ 时，根据推导，我们似乎可以从 $f|_{U_1 \rightarrow V_1}, f|_{U_2 \rightarrow V_2}$ 确定 f 。这件事对无穷维也是成立的，我们可以写出如下的等式：

$$\forall u \in U, \quad f(u) = f|_{U_1 \rightarrow V_1}(P_{U_1}^{(U_1, U_2)}(u)) + f|_{U_2 \rightarrow V_2}(P_{U_2}^{(U_1, U_2)}(u))$$

这个等式中，我们首先需要解释 $P_{U_1}^{(U_1, U_2)}$ 与 $P_{U_2}^{(U_1, U_2)}$ 的含义。 $P_{U_1}^{(U_1, U_2)}$ 是指这样一个映射：由于任何 $u \in U$ 都可以唯一分解为 $u_1 + u_2$ ， $u_1 \in U_1, u_2 \in U_2$ ，定义 $U \rightarrow U$ 的映射

$$P_{U_1}^{(U_1, U_2)}(u) = u_1, \quad P_{U_2}^{(U_1, U_2)}(u) = u_2$$

为在 (U_1, U_2) 的分解下分别到 U_1 与 U_2 的投影。

我们证明，投影映射的确是线性映射，且等式成立。

证明：

本讲义 19.2.5 证明商空间与补空间同构时构造的映射即为投影 $P_V^{(W,V)}$ ，由此线性性验证完全相同。为说明结论成立，设 $u = u_1 + u_2$, $u_1 \in U_1$, $u_2 \in U_2$, 直接由限制映射与投影定义可得等式右端为

$$f|_{U_1 \rightarrow V_1}(u_1) + f|_{U_2 \rightarrow V_2}(u_2) = f(u_1) + f(u_2) = f(u_1 + u_2) = f(u)$$

从而成立。

* 必须强调，这里上标 (U_1, U_2) 是必要的，因为若 U_2 取 U_1 的另一个补空间 U'_2 ，则 $P_{U_1}^{(U_1, U'_2)} \neq P_{U_1}^{(U_1, U_2)}$ ，如考虑 $U = \mathbb{R}^2$, $U_1 = \langle (1, 0)^T \rangle$, $U_2 = \langle (0, 1)^T \rangle$, $U'_2 = \langle (1, 1)^T \rangle$, $u = (1, 1)^T$ ，大家可以计算两个映射在 u 上的结果。

* 此外，等式的成立并不意味着 $f = f|_{U_1 \rightarrow V_1} \circ P_{U_1}^{(U_1, U_2)} + f|_{U_2 \rightarrow V_2} \circ P_{U_2}^{(U_1, U_2)}$ ，因为这里的映射定义域与陪域导致了复合无法进行。一个合理的写法是

$$f = I_{V_1}^V \circ f|_{U_1 \rightarrow V_1} \circ P_{U_1}^{(U_1, U_2)}|_{U \rightarrow U_1} + I_{V_2}^V \circ f|_{U_2 \rightarrow V_2} \circ P_{U_2}^{(U_1, U_2)}|_{U \rightarrow U_2}$$

这里 I_W^V 代表 V 的子空间 W 到 V 的线性映射，满足 $I_W^V(w) = w$ 对任何 $w \in W$ 成立。不过，这样抽象的对映射的操作并不在我们的范围内，感兴趣的同學可以自行验证。

为了之后的讨论，我们还需要说明一些投影映射的性质：

1. 投影映射满足 $\text{Ker } P_{U_1}^{(U_1, U_2)} = U_2$ 、 $\text{Im } P_{U_1}^{(U_1, U_2)} = U_1$ 、 $\text{Ker } P_{U_2}^{(U_1, U_2)} = U_1$ 、 $\text{Im } P_{U_2}^{(U_1, U_2)} = U_2$ 。

证明：

本讲义 19.2.5 证明商空间与补空间同构时构造的映射即为投影 $P_V^{(W,V)}$ ，由此 Im 与 Ker 的计算完全相同可得结论。

2. 投影映射与自身的复合还是自身（这称为幂等性）。

证明：

由定义可知当 $u_1 \in U_1$ 时 $P_{U_1}^{(U_1, U_2)}(u_1) = u_1$ ，可知若 $u = u_1 + u_2$, $u_1 \in U_1$, $u_2 \in U_2$ ，有

$$P_{U_1}^{(U_1, U_2)}(P_{U_1}^{(U_1, U_2)}(u)) = P_{U_1}^{(U_1, U_2)}(u_1) = u_1 = P_{U_1}^{(U_1, U_2)}(u)$$

从而结论陈列馆 i。对 $P_{U_2}^{(U_1, U_2)}$ 同理。

3. 所有与自身复合还是自身的 $f \in \text{Hom}(U, U)$ 都是某个投影映射。

证明：

对满足条件的映射 f ，设 $\text{Im } f = U_1$ 、 $\text{Ker } f = U_2$ （注意它们都是 U 的子空间，可以计算交与和），下证

$$f = P_{U_1}^{(U_1, U_2)}$$

首先，对任何 $u_1 \in \text{Im } f$ ，设 $u_1 = f(u)$ 可发现 $f(u_1) = f(f(u)) = f(u) = u_1$ ，从而 $\text{Im } f$ 中映射到 0 的元素只能是 0，即 $\text{Ker } f \cap \text{Im } f = \{0\}$ ，和是直和。

此外，对任何 $u \in U$ 有 $u = f(u) + (u - f(u))$, $f(u) \in \text{Im } f$ ，而由条件可知 $f(u - f(u)) = f(u) - f(f(u)) = 0$ ，于是 $u - f(u) \in \text{Ker } f$ ，这就证明了 $\text{Ker } f \oplus \text{Im } f = U$ 。

最后，对任何 $u \in U$ ，设 $u = u_1 + u_2$, $u_1 \in \text{Im } f$, $u_2 \in \text{Ker } f$ ，根据已证有 $f(u) = f(u_1) = u_1$ ，这即符合投影映射 $P_{U_1}^{(U_1, U_2)}$ 的定义，得证。

4. $P_{U_1}^{(U_1, U_2)} + P_{U_2}^{(U_1, U_2)} = \mathcal{I}$ ，这里 \mathcal{I} 表示恒等映射。

证明：

直接由定义，对任何 $u \in U$ ，设 $u = u_1 + u_2$, $u_1 \in U_1$ 、 $u_2 \in U_2$ ，则

$$P_{U_1}^{(U_1, U_2)}(u) + P_{U_2}^{(U_1, U_2)}(u) = u_1 + u_2 = u$$

而 $P_{U_1}^{(U_1, U_2)}$ 、 $P_{U_2}^{(U_1, U_2)}$ 、 \mathcal{I} 都是 U 到 U 的线性映射，从而结论成立。

* 对于 $U_1 \oplus \cdots \oplus U_k = U$ 的情况，同样可以定义投影 $P_{U_i}^{(U_1, \dots, U_k)}$ ，定义为 u 的唯一分解中 U_i 里的分量。利用直和的性质，即有

$$P_{U_i}^{(U_1, \dots, U_k)} = P_{U_i}^{(U_i, \hat{U}_i)}, \quad \hat{U}_i = U_1 \oplus \cdots \oplus U_{i-1} \oplus U_{i+1} \oplus \cdots \oplus U_k$$

从而可以与分为两个子空间时的有类似性质。

20.2.3 例题

1. (a) 对 \mathbb{K} 上 n 阶方阵 A 定义线性映射 $\mathcal{A}(X) = AX - XA$ ，若 A 可对角化，计算 $\text{Ker } \mathcal{A}$ 与 $\text{Im } \mathcal{A}$ 的维数，并证明 $\text{Ker } \mathcal{A} = \langle I, A, \dots, A^{n-1} \rangle$ 当且仅当 A 的特征值互不相同。
 (b) 若 A 看作 \mathbb{C} 上方阵可对角化，而在 \mathbb{K} 上不可对角化，重新计算 $\text{Ker } \mathcal{A}$ 与 $\text{Im } \mathcal{A}$ 的维数，并讨论何时 $\text{Ker } \mathcal{A} = \langle I, A, \dots, A^{n-1} \rangle$ 。
2. 设 U 、 V 是 \mathbb{K} 上的有限维线性空间， f 是 $U \rightarrow V$ 的线性映射，用 $f(W)$ 表示 U 的子空间 W 的像集（本讲义 19.2.2 已验证其为线性空间）， $f^{-1}(X)$ 表示 V 的子空间 X 的原像集，证明
 - (a) $\dim W - \dim \text{Ker } f \leq \dim f(W) \leq \dim W$;
 - (b) $f^{-1}(X)$ 为 U 的子空间，且

$$\dim f^{-1}(X) \leq \dim X + \dim \text{Ker } f$$

- (c) 证明当 $X \subset \text{Im } f$ 时

$$\dim f^{-1}(X) \geq \dim X$$

并对逆命题举出反例。

3. 利用线性映射证明秩不等式（后两问需要限制映射）：

- (a) $\text{rank}(A + B) \leq \text{rank } A + \text{rank } B$;
- (b) $\text{rank}(AB) \leq \text{rank } A$;
- (c) $\text{rank}(AB) \leq \text{rank } B$;
- (d) $\text{rank} \begin{pmatrix} A & O \\ O & B \end{pmatrix} = \text{rank } A + \text{rank } B$;
- (e) $\text{rank} \begin{pmatrix} A & C \\ O & B \end{pmatrix} \geq \text{rank } A + \text{rank } B$ 。

4. 对 \mathbb{K} 上有限维线性空间 U 、 V 、 W 、 X 之间的线性映射

$$\mathcal{C}: U \rightarrow V, \quad \mathcal{B}: V \rightarrow W, \quad \mathcal{A}: W \rightarrow X$$

证明（这里省略了复合的记号）

$$\dim \text{Im } \mathcal{ABC} + \dim \text{Im } \mathcal{B} \geq \dim \text{Im } \mathcal{AB} + \dim \text{Im } \mathcal{BC}$$

并给出此结论的矩阵版本。

5. 对 \mathbb{K} 上线性空间 U, V 之间的线性映射 f , 若 $\dim \text{Im } f = r$, 证明存在 $f_1, \dots, f_r \in \text{Hom}(U, V)$ 使得

$$\forall i = 1, \dots, r, \quad \dim \text{Im } f_i = 1$$

$$f = \sum_{i=1}^r f_i$$

并给出此结论的矩阵版本。

20.2.4 解答

1. (a) 设 A 的特征多项式 $\varphi_A(x) = (x - \lambda_1)^{m_1} \dots (x - \lambda_k)^{m_k}$, 这里 $\lambda_1, \dots, \lambda_k$ 互不相同, 且 $m_1 + \dots + m_k = n$ 。

假设 A 的对角化为 $A = P^{-1}DP$, 其中 D 为对角阵

$$\text{diag}(\lambda_1 I_{m_1}, \lambda_2 I_{m_2}, \dots, \lambda_k I_{m_k})$$

- $\text{Ker } \mathcal{A}$ 计算

由条件 $X \in \text{Ker } \mathcal{A}$ 即 $AX = XA$, 从而同左乘 P 、右乘 P^{-1} (由于可以左乘 P^{-1} 右乘 P 从右推出左, 左右是等价的) 有

$$P^{-1}DPX = X P^{-1}DP \iff DPXP^{-1} = PX P^{-1}D$$

令 $Y = PX P^{-1}$, 则通过 $DY = YD$ 可直接对比分量 (这即得到 $(d_i - d_j)y_{ij} = 0$, d_i 为 D 的第 i 个对角元, 从而 $d_i = d_j$ 时 y_{ij} 可任取, 否则为 0) 算出所有的 Y 是

$$Y = \text{diag}(Y_1, Y_2, \dots, Y_k)$$

其中 Y_i 是 m_i 阶方阵。

由于 Y 的分量只有任取或为 0, 它构成一个线性空间, 记为 $\mathcal{C}(D)$ (与 D 可交换的方阵), 一组基为

$$\{E_{ij} \mid d_i = d_j\}$$

参考分块形式可看出总个数 (即维数) 为 $m_1^2 + m_2^2 + \dots + m_k^2$ 。

我们下面证明, 映射 $f(Y) = P^{-1}YP$ 是 $\mathbb{K}^{n \times n} \rightarrow \mathbb{K}^{n \times n}$ 的同构。直接验证可知其线性, 且记 $g(X) = PX P^{-1}$ 可直接计算得

$$f(g(X)) = X, \quad g(f(Y)) = Y$$

从而其存在逆映射, 是同构。

由此, 根据 $Y = PX P^{-1}$ 可知 $X = P^{-1}YP$, 从而所有满足要求的 X 即为 $f(\mathcal{C}(D))$ 。由 f 是同构可验证其将线性无关向量组映射到线性无关向量组, 从而将 $\{f(E_{ij}) \mid d_i = d_j\}$ 是线性无关的, 再由像空间定义可知它们构成 $f(\mathcal{C}(D))$ 一组基, 于是构成 $\text{Ker } \mathcal{A}$ 一组基, 这就得到了

$$\dim \text{Ker } \mathcal{A} = \dim \mathcal{C}(D) = m_1^2 + m_2^2 + \dots + m_k^2$$

- $\text{Im } \mathcal{A}$ 维数计算

利用第一同构定理可知

$$\dim \text{Im } \mathcal{A} = \dim \mathbb{K}^{n \times n} - \dim \text{Ker } \mathcal{A} = n^2 - m_1^2 - m_2^2 - \dots - m_k^2$$

事实上, 利用第一同构定理也可以得到 $\text{Im } \mathcal{A}$ 的一组基。

由于同构将补空间映射到补空间，而 $\mathcal{C}(D)$ 的补空间一组基可以取为 $\{E_{ij} \mid d_i \neq d_j\}$, $\text{Ker } \mathcal{A}$ 的补空间一组基可以取为 $\{f(E_{ij}) \mid d_i \neq d_j\}$ 。

由此，商空间 $\mathbb{K}^{n \times n}/\text{Ker } \mathcal{A}$ 的一组基可以取为

$$\{f(E_{ij}) + \text{Ker } \mathcal{A} \mid d_i \neq d_j\}$$

利用第一同构定理的构造，同构将基映射到基，从而 $\text{Im } \mathcal{A}$ 一组基为

$$\{\mathcal{A}(f(E_{ij})) \mid d_i \neq d_j\}$$

也即

$$\{AP^{-1}E_{ij}P - P^{-1}E_{ij}PA \mid d_i \neq d_j\}$$

- $\text{Ker } \mathcal{A} = \langle I, A, \dots, A^{n-1} \rangle$ 推特征值互不相同

若特征值有相同， m_1 到 m_k 中至少有一个大于 1，从而

$$\dim \text{Ker } \mathcal{A} = m_1^2 + \dots + m_k^2 > m_1 + \dots + m_k = n$$

但右侧为 n 个向量生成的子空间，维数至多为 n ，矛盾。

- 特征值互不相同推 $\text{Ker } \mathcal{A} = \langle I, A, \dots, A^{n-1} \rangle$

此时 $\dim \text{Ker } \mathcal{A} = n$ ，且由于 A 的多项式都与 A 可交换知

$$\langle I, A, \dots, A^{n-1} \rangle \subset \text{Ker } \mathcal{A}$$

从而只需证明左侧维数为 n ，即 I, A, \dots, A^{n-1} 线性无关即可。

直接计算可知，若 $a_0 + a_1A + \dots + a_{n-1}A_{n-1} = O$ ，有

$$O = a_0 + a_1A + \dots + a_{n-1}A_{n-1} = P^{-1}(a_0 + a_1D + \dots + a_{n-1}D^{n-1})P$$

从而 $a_0 + a_1D + \dots + a_{n-1}D^{n-1} = O$ 。由于 D 为对角阵， $a_0 + a_1D + \dots + a_{n-1}D^{n-1}$ 即为第 i 个对角元是 $a_0 + a_1d_i + \dots + a_{n-1}d_i^{n-1}$ 的对角阵。由此，根据其为 O ， A 的每个特征值都是 $a_0 + a_1x + \dots + a_{n-1}x^{n-1} = 0$ 的根，但一个非零 $n-1$ 次多项式至多 $n-1$ 个根，从而只能为 0，即得所有 a_i 为 0，线性无关。

* 也可以通过 Vandermonde 行列式直接从 $P^{-1}DP$ 的 1 到 $n-1$ 次方构造出 $\text{Ker } \mathcal{A}$ 。

- (b) 我们将证明，设 A 的特征多项式在 \mathbb{C} 上分解为 $\varphi_A(x) = (x - \lambda_1)^{m_1} \dots (x - \lambda_k)^{m_k}$ ，则 $\dim \text{Im } \mathcal{A}$ 与 $\dim \text{Ker } \mathcal{A}$ 与 $\text{Ker } \mathcal{A} = \langle I, A, \dots, A^{n-1} \rangle$ 都不变。

$X \in \text{Ker } \mathcal{A}$ 等价于 $AX - XA = O$ ，而展开分量可发现这是一个 n^2 个变量的线性方程组，从而看成不同数域上的解空间维数不变（可以考虑用简单阶梯形矩阵思考，这个结论本质是不同数域相抵标准形相同），于是 $\dim \text{Ker } \mathcal{A}$ 不变，再由第一同构定理得到 $\dim \text{Im } \mathcal{A}$ 不变。

由于 $\dim \text{Ker } \mathcal{A}$ 不变， $\text{Ker } \mathcal{A} = \langle I, A, \dots, A^{n-1} \rangle$ 仍能特征值互不相同，而若特征值互不相同，将 I, A, \dots, A^{n-1} 看作 \mathbb{C} 上方阵线性无关，但由于它们在 \mathbb{K} 上，看作 \mathbb{K} 上方阵仍线性无关（可看作 \mathbb{K} 上线性方程组），从而仍可得到结论。

2. (a) 考虑 f 的限制映射 $f|_{W \rightarrow V}$ ，利用第一同构定理可知

$$\dim W = \dim \text{Ker } f|_{W \rightarrow V} + \dim \text{Im } f|_{W \rightarrow V}$$

利用限制映射定义可知

$$\text{Ker } f|_{W \rightarrow V} = \{x \in W \mid f|_{W \rightarrow V}(x) = 0\} = \{x \in W \mid f(x) = 0\} = W \cap \text{Ker } f$$

$$\text{Im } f|_{W \rightarrow V} = \{v \in V \mid \exists w \in W, f|_{W \rightarrow V}(w) = v\} = \{v \in V \mid \exists w \in W, f(w) = v\} = f(W)$$

由此即得

$$\dim W = \dim(\text{Ker } f \cap W) + \dim f(W)$$

而由子空间维数不超过原空间维数可知

$$\dim f(W) \leq \dim(\text{Ker } f \cap W) + \dim f(W) \leq \dim \text{Ker } f + \dim f(W)$$

将中间替换为 $\dim W$ 并移项即得到要证的不等式。

* 若是不熟悉第一同构定理，构造 W 的基强做也是可以的，但会麻烦很多。以右侧不等号为例，先利用 f 线性说明线性相关向量组的像仍然线性相关，于是 $f(W)$ 中线性无关向量组的向量个数不可能超过 W 中的，这也可以说明结论。不过，由于之后还要大量应用限制映射的语言，还是推荐大家掌握限制技巧的。

- (b) 先验证 $f^{-1}(X)$ 为子空间。对 $u_1, u_2 \in f^{-1}(X)$ 、 $\lambda \in \mathbb{K}$ ，由条件 $f(u_1), f(u_2) \in X$ ，从而由 X 封闭性

$$f(u_1 + u_2) = f(u_1) + f(u_2) \in X, \quad f(\lambda u_1) = \lambda f(u_1) \in X$$

这就得证 $u_1 + u_2, \lambda u_1 \in f^{-1}(X)$ ，从而成立。

由 $f^{-1}(X)$ 的定义可知其上元素的像在 X 中，限制映射可定义。考虑 f 在 $f^{-1}(X) \rightarrow X$ 上的限制映射，利用第一同构定理有

$$\dim f^{-1}(X) = \dim \text{Ker } f|_{f^{-1}(X) \rightarrow X} + \dim \text{Im } f|_{f^{-1}(X) \rightarrow X}$$

与 (a) 相同分析得

$$\dim f^{-1}(X) = \dim(\text{Ker } f \cap f^{-1}(X)) + \dim f(f^{-1}(X))$$

我们下面证明 $f(f^{-1}(X)) = X \cap \text{Im } f$ 。根据定义有

$$f(f^{-1}(X)) = \{f(u) \mid u \in f^{-1}(X)\} = \{f(u) \mid f(u) \in X\}$$

令 $f(u) = u'$ ，则 u' 的取值范围为 $\text{Im } f$ ，由此

$$f(f^{-1}(X)) = \{u' \in \text{Im } f \mid u' \in X\} = X \cap \text{Im } f$$

从而

$$\dim f^{-1}(X) = \dim(\text{Ker } f \cap f^{-1}(X)) + \dim(\text{Im } f \cap X)$$

第二项不超过 $\dim \text{Ker } f$ ，第三项不超过 $\dim X$ ，即得结论。

- (c) 当 $X \subset \text{Im } f$ 时 (b) 中式子可改写为

$$\dim f^{-1}(X) = \dim(\text{Ker } f \cap f^{-1}(X)) + \dim X$$

从而结论成立。

考虑将 \mathbb{R}^2 所有元素映射到 $0 \in \mathbb{R}$ 的线性映射 \mathcal{O} ，记 $X = \mathbb{R}$ ，则 $\dim \mathcal{O}^{-1}(X) = \dim \mathbb{R}^2 = 2$ ，而 $\dim X = 1$ ，且 X 不包含在 $\text{Im } \mathcal{O} = \{0\}$ 中。

* 反例构造的方式是直接从等式中看出的：只要让 $\dim \text{Ker } f \cap f^{-1}(X)$ 充分大即可。

* 本题中有一个直接根据限制映射定义可得（从而可以直接使用）的重要的限制映射像与核结论：

$$\text{Ker } f|_{U_0 \rightarrow V_0} = \text{Ker } f \cap U_0, \quad \text{Im } f|_{U_0 \rightarrow V_0} = f(U_0)$$

此外，在 f 不可逆时一定要注意 $f(f^{-1}(W))$ 与 $f^{-1}(f(W))$ 都未必等于 W 。事实上我们已经证明了 $f(f^{-1}(W)) = W \cap \text{Im } f$ ，而 $f^{-1}(f(W))$ 可以是任何包含 W 的子空间，读者可以考虑如何构造符合要求的 f 。

3. 本题中我们用花体字母 $\mathcal{A}, \mathcal{B}, \mathcal{C}$ 表示线性映射 $x \rightarrow Ax, x \rightarrow Bx, x \rightarrow Cx$, 由此通过本讲义 19.3.4 秩与像空间维数关系可知

$$\dim \text{Im } \mathcal{A} = \text{rank } A$$

对 \mathcal{B}, \mathcal{C} 同理。

- (a) 由线性映射矩阵表示的加法是加法的矩阵表示可知也即要证明

$$\dim \text{Im } (\mathcal{A} + \mathcal{B}) \leq \dim \text{Im } \mathcal{A} + \dim \text{Im } \mathcal{B}$$

而对任何 $y \in \text{Im } (\mathcal{A} + \mathcal{B})$, 存在 x 使得 $y = (\mathcal{A} + \mathcal{B})(x) = \mathcal{A}(x) + \mathcal{B}(x)$, 由 $\mathcal{A}(x) \in \text{Im } \mathcal{A}, \mathcal{B}(x) \in \text{Im } \mathcal{B}$ 即可知

$$\text{Im } (\mathcal{A} + \mathcal{B}) \subset \text{Im } \mathcal{A} + \text{Im } \mathcal{B}$$

于是 (第二个不等号由和空间维数公式可得)

$$\dim \text{Im } (\mathcal{A} + \mathcal{B}) \leq \dim(\text{Im } \mathcal{A} + \text{Im } \mathcal{B}) \leq \dim \text{Im } \mathcal{A} + \dim \text{Im } \mathcal{B}$$

- (b) 由线性映射矩阵表示的加法是复合的矩阵表示可知也即要证明

$$\dim \text{Im } (\mathcal{A} \circ \mathcal{B}) \leq \dim \text{Im } \mathcal{A}$$

而利用定义

$$\text{Im } (\mathcal{A} \circ \mathcal{B}) = \{y \mid \exists x, y = \mathcal{A}(\mathcal{B}(x))\} = \{\mathcal{A}(z) \mid \exists x, z = \mathcal{B}(x)\} \subset \text{Im } \mathcal{A}$$

于是结论成立。

- (c) 由线性映射矩阵表示的加法是复合的矩阵表示可知也即要证明

$$\dim \text{Im } (\mathcal{A} \circ \mathcal{B}) \leq \dim \text{Im } \mathcal{B}$$

而类似 (b) 中推导可得 $\text{Im } (\mathcal{A} \circ \mathcal{B}) = \mathcal{A}(\text{Im } \mathcal{B})$, 于是利用第二题 (a) 右侧不等号可知成立。

- (d) 利用本讲义 20.2.1 的结论, 将左右视为矩阵表示可知只需证明对 \mathbb{K} 上的线性空间 $U, V, U_1 \oplus U_2 = U, V_1 \oplus V_2 = V$, 若 $f \in \text{Hom}(U, V)$ 满足 $f(U_1) \subset V_1, f(U_2) \subset V_2$, 则

$$\dim \text{Im } f = \dim \text{Im } f|_{U_1 \rightarrow V_1} + \dim \text{Im } f|_{U_2 \rightarrow V_2}$$

我们只需证明

$$\text{Im } f = \text{Im } f|_{U_1 \rightarrow V_1} \oplus \text{Im } f|_{U_2 \rightarrow V_2}$$

下记 $f_1 = f|_{U_1 \rightarrow V_1}, f_2 = f|_{U_2 \rightarrow V_2}$ 。

首先, 由条件 $V_1 \cap V_2 = \{0\}$, 从而根据限制映射定义可知 $\text{Im } f_1 \subset V_1, \text{Im } f_2 \subset V_2$, 因此两者交为 $\{0\}$, 和是直和。其次, 由限制映射定义 $\text{Im } f_1 \subset \text{Im } f, \text{Im } f_2 \subset \text{Im } f$, 右侧两个空间都是左侧的子空间, 直和也是左侧的子空间。最后, 由本讲义 20.2.2 证明的 $f(u) = f_1(P_{U_1}^{(U_1, U_2)}(u)) + f_2(P_{U_2}^{(U_1, U_2)}(u))$, 由于右侧是 $\text{Im } f_1$ 与 $\text{Im } f_2$ 中各一个元素求和, 其属于 $\text{Im } f_1 + \text{Im } f_2$, 即得 $\text{Im } f \subset \text{Im } f_1 + \text{Im } f_2$ 。综合三者得证。

- (e) 证明分为五个部分。

- 改写为映射语言

虽然左侧可以根据本讲义 20.2.1 视为能限制在 $U_1 \rightarrow V_1$ 的映射的矩阵表示, 但此时相对难以刻画右侧的 $\dim \text{Im } \mathcal{B}$, 因此我们将左侧看成

$$\begin{pmatrix} A & O \\ O & B \end{pmatrix} + \begin{pmatrix} O & C \\ O & O \end{pmatrix}$$

由此看成矩阵表示可成为两映射之和的矩阵表示。 f 、 U 、 V 、 $U_{1,2}$ 、 $V_{1,2}$ 、 $f_{1,2}$ 的定义都如 (d)，设 U_1 、 U_2 的一组基为 S_1 、 S_2 ， V_1 、 V_2 的一组基为 T_1 、 T_2 ，则 g 满足

$$(g(S_1), g(S_2)) = (T_1, T_2) \begin{pmatrix} O & C \\ O & O \end{pmatrix}$$

从而

$$g(S_1) = O, \quad g(S_2) = T_1 C$$

由此即得到 g 满足 $U_1 \in \text{Ker } g$ ， $g(U_2) \subset V_1$ 。需要证明的结论即

$$\dim \text{Im } (f + g) \geq \dim \text{Im } f_1 + \dim \text{Im } f_2$$

- $\text{Im } f_1$ 估算

对 $y \in \text{Im } f_1$ ，根据定义存在 $x \in U_1$ 使得 $f(x) = y$ ，则由 $U_1 \in \text{Ker } g$ 可得 $(f+g)(x) = f(x)$ ，从而 $\text{Im } f_1 \subset \text{Im } (f+g)$ ，再由限制映射定义可知

$$\text{Im } f_1 \subset \text{Im } (f+g) \cap V_1$$

- $\text{Im } f_2$ 估算

对任何 $v \in \text{Im } (f+g)$ ，设 $v = v_1 + v_2$ ， $v_1 \in V_1$ 、 $v_2 \in V_2$ ，将所有 v_2 构成的集合记为 W ，根据投影映射定义 $W = P_{V_2}^{(V_1, V_2)}(\text{Im } (f+g))$ ，从而其为子空间。可以证明 $\text{Im } f_2 \subset W$ ：对任何 $v \in \text{Im } f_2$ ，设 $f_2(u) = v$ ，由于 $u \in U_2$ 可知 $g(u) \in V_1$ ，进一步得 $(g(u)) \in V_1$ ，投影到 V_2 上为 0)

$$P_{V_2}^{(V_1, V_2)}((f+g)(u)) = P_{V_2}^{(V_1, V_2)}(v + g(u)) = v$$

* 虽然本题无需使用，不过事实上成立 $W = \text{Im } f_2$ 。另一边包含关系证明：若 $v_2 \in W$ ，设 $v_1 + v_2 = f(u) + g(u)$ ，这里 $v_1 \in V_1$ ，设 $u = u_1 + u_2$ ， $u_1 \in U_1$ 、 $u_2 \in U_2$ ，有 $f(u) + g(u) = (f(u_1) + g(u_1)) + f(u_2)$ ，可发现 $f(u_1) + g(u_1) \in V_1$ 、 $f(u_2) \in V_2$ ，由分解唯一性 $v_2 = f(u_2)$ ，再由限制映射定义得到 $v_2 = f_2(u_2)$ ，从而 $W \subset \text{Im } f_2$ 。

- 直和分解

综合以上，我们得到

$$\text{Im } f_1 \oplus \text{Im } f_2 \subset (\text{Im } (f+g) \cap V_1) \oplus P_{V_2}^{(V_1, V_2)}(\text{Im } (f+g))$$

左侧的直和在 (d) 中证明了，右侧的直和是因为第一个空间是 V_1 子空间，第二个空间是 V_2 子空间。

设 $X = \text{Im } (f+g)$ ，可知

$$\dim \text{Im } f_1 + \dim \text{Im } f_2 \leq \dim(X \cap V_1) + \dim P_{V_2}^{(V_1, V_2)}(X)$$

我们最后证明 $\dim(X \cap V_1) + \dim P_{V_2}^{(V_1, V_2)}(X) = X$ 对任何 V 的子空间 X 成立。

- 投影映射限制

对 $P_{V_2}^{(V_1, V_2)}|_{X \rightarrow V}$ 利用第一同构定理，与之前类似利用限制映射定义可发现

$$\dim X = \dim \text{Ker } P_{V_2}^{(V_1, V_2)}|_{X \rightarrow V} + \dim \text{Im } P_{V_2}^{(V_1, V_2)}|_{X \rightarrow V} = \dim(\text{Ker } P_{V_2}^{(V_1, V_2)} \cap X) + P_{V_2}^{(V_1, V_2)}(X)$$

利用 $\text{Ker } P_{V_2}^{(V_1, V_2)} = V_1$ 最终得到

$$\dim X = \dim(X \cap V_1) + \dim P_{V_2}^{(V_1, V_2)}(X)$$

* 利用商空间、补空间同构性， $(X \cap V_1) \oplus P_{V_2}^{(V_1, V_2)}(X)$ 事实上同构于 X 。

* 可能存在更简单的做法，但空间角度本质都是和上述讨论一样的。此外，上述讨论也可以得到等号成立当且仅当

$$\text{Im } f_1 = \text{Im } (f + g) \cap V_1$$

* 注意本题我们得到了很多 Im 之间的包含或相等关系，这些结论都是与维数无关的，可以视为秩不等式的推广。

4. 本题的核心思路是，既然讨论的是 Im 之间的关系，我们想将同一个空间的子空间放在等式同一边进行处理。

由于 Im 的定义， $\text{Im } \mathcal{AB}$ 、 $\text{Im } \mathcal{ABC}$ 都是 X 的子空间， $\text{Im } \mathcal{B}$ 、 $\text{Im } \mathcal{BC}$ 都是 W 的子空间，且第三题 (b) 已证 $\text{Im } \mathcal{BC} \subset \text{Im } \mathcal{B}$ 、 $\text{Im } \mathcal{ABC} \subset \text{Im } \mathcal{AB}$ ，从而将要证的不等式移项为

$$\dim \text{Im } \mathcal{B} - \dim \text{Im } \mathcal{BC} \geq \dim \text{Im } \mathcal{AB} - \dim \text{Im } \mathcal{ABC}$$

由于空间减子空间维数可以看作商空间或补空间的维数，但这里用商空间显得较为复杂，我们用补空间记。设 $\text{Im } \mathcal{BC}$ 对 $\text{Im } \mathcal{B}$ 的补空间为 W_0 ，即

$$W_0 \oplus \text{Im } \mathcal{BC} = \text{Im } \mathcal{B}$$

左侧即成为 $\dim W_0$ 。为了和右侧建立联系，在第三题 (c) 已证 $\text{Im } \mathcal{AB} = \mathcal{A}(\text{Im } \mathcal{B})$ ，从而两侧同时作用 \mathcal{A} 得到

$$\mathcal{A}(W_0 \oplus \text{Im } \mathcal{BC}) = \text{Im } \mathcal{AB}$$

本讲义 19.2.2 中证明了 $\mathcal{A}(W_0 + \text{Im } \mathcal{BC}) = \mathcal{A}(W_0) + \mathcal{A}(\text{Im } \mathcal{BC})$ ，从而

$$\mathcal{A}(W_0) + \text{Im } \mathcal{ABC} = \text{Im } \mathcal{AB}$$

两侧取 \dim ，利用和的维数不超过维数的和可知

$$\dim \mathcal{A}(W_0) + \dim \text{Im } \mathcal{ABC} \geq \dim \text{Im } \mathcal{AB}$$

移项，并利用第二题 (a) 已证的 $\dim \mathcal{A}(W_0) \leq \dim W_0$ ，即得到

$$\dim W_0 \geq \dim \mathcal{A}(W_0) \geq \dim \text{Im } \mathcal{AB} - \dim \text{Im } \mathcal{ABC}$$

即得证。

类似第三题可知矩阵版本即对 \mathbb{K} 上矩阵 A 、 B 、 C (假设乘法 AB 、 BC 可以进行) 有

$$\text{rank } B + \text{rank } ABC \geq \text{rank } AB + \text{rank } BC$$

即为 **Frobenius 秩不等式**。

5. 设 $\dim \text{Im } f$ 的一组基为 $\alpha_1, \dots, \alpha_r$ ，利用直和的等价条件可发现

$$\text{Im } f = \langle \alpha_1 \rangle \oplus \dots \oplus \langle \alpha_r \rangle$$

设 P_i 代表 $\text{Im } f$ 上的投影 $P_{\langle \alpha_i \rangle}^{(\langle \alpha_1 \rangle, \dots, \langle \alpha_r \rangle)}$ ，由投影映射性质可发现 $\text{Im } P_i = \langle \alpha_i \rangle$ ，且 $P_1 + \dots + P_r = \mathcal{I}$ (这里 \mathcal{I} 为 $\text{Im } f \rightarrow \text{Im } f$ 的恒等映射)。

记 $f_i = I_{\text{Im } f}^V \circ P_i \circ f|_{U \rightarrow \text{Im } f}$ ，我们下面来证明 f_1, \dots, f_r 满足条件，这里 $I_{\text{Im } f}^V$ 表示 $\text{Im } f$ 到 V 的满足 $I_{\text{Im } f}^V(x) = x$ 的映射，可由定义直接验证线性。首先，由于每个 f_i 从右到左的映射分别是 U 到 $\text{Im } f$ 、 $\text{Im } f$ 到 $\text{Im } f$ 、 $\text{Im } f$ 到 V 的线性映射，利用线性映射复合线性可知它的确是 U 到 V 的线性映射。

其次，利用映射加法和复合的定义可知

$$\sum_{i=1}^r f_i = \sum_{i=1}^r I_{\text{Im } f}^V \circ P_i \circ f|_{U \rightarrow \text{Im } f} = I_{\text{Im } f}^V \circ \sum_{i=1}^r P_i \circ f|_{U \rightarrow \text{Im } f} = I_{\text{Im } f}^V \circ f|_{U \rightarrow \text{Im } f}$$

而最后这个映射根据定义，对任何 $u \in U$ 有

$$I_{\text{Im } f}^V(f|_{U \rightarrow \text{Im } f}(u)) = I_{\text{Im } f}^V(f(u)) = f(u)$$

且也是 U 到 V 的映射 (U 到 $\text{Im } f$ 与 $\text{Im } f$ 到 V 的复合)，于是就是 f 。

* 最后这个映射其实就是本讲义 20.1.2 最后一题的满秩分解证明中定义的映射分解。

最后，由于 $I_{\text{Im } f}^V$ 不改变任何元素，可知

$$\text{Im } f_i = \text{Im } P_i \circ f|_{U \rightarrow \text{Im } f} \subset \text{Im } P_i = \langle \alpha_i \rangle$$

从而 $\dim \text{Im } f_i \leq 1$ 。另一方面，由 $f = \sum_{i=1}^r f_i$ 可知

$$\sum_{i=1}^r \dim \text{Im } f_i \geq \dim \text{Im } \sum_{i=1}^r f_i = r$$

于是只能所有 $\dim \text{Im } f_i$ 都为 1，得证。

类似第三题可知矩阵版本即对 \mathbb{K} 上矩阵 A ，若 $\text{rank } A = r$ ，存在矩阵 A_1, \dots, A_r 使得 $A = \sum_{i=1}^r A_i$ ，且 $\text{rank } A_i = 1$ 对每个 A_i 成立。

§20.3 线性变换

线性变换的定义非常简单。若一个线性映射的定义域和陪域相同，它就是一个线性变换。线性变换 $f \in \text{Hom}(U, U)$ 可以称为 U 上的线性变换。本节开始，如无特殊说明，省略线性变换复合的 \circ 记号。

20.3.1 多项式与特征系统

设 U 是 \mathbb{K} 上的线性空间。对于一个一般的线性变换 $\mathcal{A} \in \text{Hom}(U, U)$ ，它的很多讨论与线性映射是相同的，例如像、核、单同态/满同态/同构（此时称为自同构）条件、第一同构定理等等。不过，有一件事会存在一定的差别：由于它的定义域与陪域限制，它可以与自身进行复合。我们用 \mathcal{A}^n 记 \mathcal{A} 与自身复合 n 次的结果，且 \mathcal{A}^0 表示 \mathcal{I} ，即恒等变换 $\mathcal{I}(x) = x$ ，利用已经定义的加法、数乘，我们可以考察其多项式

$$f(\mathcal{A}) = \sum_{i=1}^k a_i \mathcal{A}^i$$

这里 $f \in \mathbb{K}[x]$ 是 \mathbb{K} 上的多项式。

事实上，所有 \mathcal{A} 的多项式构成 $\text{Hom}(U, U)$ 的子空间，记为 $\mathbb{K}[\mathcal{A}]$ ，且 $\mathbb{K}[\mathcal{A}]$ 中任何两个元素乘法可交换。

证明：

利用加法与数乘的交换、结合、分配性质，由于 \mathcal{A} 的多项式加法、数乘后还是 \mathcal{A} 的多项式，即得其为子空间。我们下面证明 $f(\mathcal{A})g(\mathcal{A}) = g(\mathcal{A})f(\mathcal{A}) = (fg)(\mathcal{A})$ 对任何 $f, g \in \mathbb{K}[x]$ 成立。

设 $f(x) = \sum_{i=0}^n a_i x^i$ 、 $g(x) = \sum_{i=0}^m b_i x^i$ ，对任何 $u \in U$ ，利用映射加法、数定义乘有

$$f(\mathcal{A})g(\mathcal{A})(u) = \sum_{i=0}^n a_i \mathcal{A}^i \left(\left(\sum_{j=0}^m b_j \mathcal{A}^j \right)(u) \right) = \sum_{i=0}^n a_i \mathcal{A}^i \left(\sum_{j=0}^m b_j \mathcal{A}^j(u) \right)$$

进一步由线性性与 \mathcal{A}^k 定义可得

$$f(\mathcal{A})g(\mathcal{A})(u) = \sum_{i=0}^n a_i \sum_{j=0}^m b_j \mathcal{A}^i(\mathcal{A}^j(u)) = \sum_{i=0}^n a_i \sum_{j=0}^m b_j \mathcal{A}^{i+j}(u)$$

而这就是 $(fg)(\mathcal{A})(u)$, 即得到 $f(\mathcal{A})g(\mathcal{A}) = (fg)(\mathcal{A})$, 而对多项式 f, g 有 $fg = gf$, 也即得证 $g(\mathcal{A})f(\mathcal{A}) = (gf)(\mathcal{A}) = (fg)(\mathcal{A})$ 。

* 当然, 对 $\mathbb{K}^{n \times n}$ 中的矩阵 A , 完全类似可定义 $\mathbb{K}^{n \times n}$ 的子空间 $\mathbb{K}[A]$, 其中任何两个元素也乘法可交换。

与矩阵时类似, $\mathcal{A}(x) = x$ 代表线性变换 \mathcal{A} 的不动点, 而如果我们放宽不动点的要求, 只要求不改变方向, 即得到 $\mathcal{A}(x) = \lambda x$ 其中 $\lambda \in \mathbb{K}$, 这称为 \mathcal{A} 的特征方程, 若存在非零解则称 $\lambda \in \mathbb{K}$ 为 \mathcal{A} 的特征值, 非零 x 称为 λ 的特征向量。利用定义, 所有 λ 的特征向量和 0 构成的集合为 $\text{Ker}(\lambda\mathcal{I} - \mathcal{A})$, 于是其为一个线性空间, 称为 λ 的特征子空间。

* 与一个 \mathbb{K} 上方阵可以直接看作 \mathbb{C} 上方阵不同, \mathbb{K} 上线性空间 U 上的线性变换想要“看成” \mathbb{C} 上线性空间的线性变换是难以简单定义的, 因此我们必须规避这样的说法, 只有化为矩阵论后才能放大数域。

就像一个 \mathbb{K} 上矩阵未必有特征值与特征向量, 一个 \mathbb{K} 上的线性变换也未必有特征值与特征向量, 例如 \mathbb{R}^2 上的线性变换 $(x, y) \rightarrow (y, -x)$ 。此外, 考虑无穷维线性空间时, 即使是 \mathbb{C} 上的线性变换也未必存在特征值与特征向量, 例如 $\mathbb{C}[x]$ 上的线性变换 $f(x) \rightarrow xf(x)$, 考虑次数可发现其特征方程不可能有非零解。在本讲义 17.4.2 中, 我们还给出了以所有 $c \in \mathbb{C}$ 为特征值的线性变换的例子。

在今后的讨论中, 我们主要关注的还是有限维空间上的线性变换, 这样至少可以保证性质与我们熟悉的矩阵相同。

20.3.2 矩阵表示与基变换

有限维时 (假设 $\dim U = n$), 我们自然需要考虑线性变换的矩阵表示。它的定义与线性映射的唯一不同是, 既然左右都是 U , 我们并不希望取出 U 的 S, T 两组不同基进行表示, 由此, 我们将左、右的基均取为 S 。一句话来说也就是, 线性变换 \mathcal{A} 在 U 的基 S 下的矩阵表示即是 \mathcal{A} 看作线性映射在基 S, S 下的矩阵表示。 \mathcal{A} 在 U 的基 S 下的矩阵表示 A_S 的三个等价定义式为 (第一行的 A_S 代表左乘 A_S 对应的线性变换)

$$\mathcal{A} = \pi_S^{-1} \circ A_S \circ \pi_S$$

$$\mathcal{A}(u)_S = A_S u_S$$

$$\mathcal{A}(S) = S A_S$$

利用映射复合与矩阵乘法的关系, 可知线性变换多项式的矩阵表示为矩阵表示的多项式, 即 \mathcal{A} 在基 S 下的矩阵表示为 A 时, $f(\mathcal{A})$ 在基 S 下的矩阵表示为 $f(A)$, 这里 f 为某个 \mathbb{K} 上多项式。

证明:

当基 S 变换为基 S' 时, 设 S 到 S' 的过渡矩阵为 P , 我们也希望知道 $A_S, A_{S'}$ 的关系。利用线性变换矩阵表示的定义, 从本讲义 19.3.4 的推导中即可发现, 这相当于 $T = S, T' = S'$ 的情况, 这时 T 到 T' 的过渡矩阵也为 P , 从而

$$A_{S'} = P^{-1} A_S P$$

我们直接复制线性映射时得到的结论: 同一个线性变换 \mathcal{A} 在不同基下的矩阵表示相似。另一方面, 之前已经证明了过渡矩阵可以取为任何可逆阵, 从而若 \mathcal{A} 在 S 下的矩阵表示为 A_S , 任何与 A_S 相似的方阵都可以看作 \mathcal{A} 在某组基下的矩阵表示——于是, 相似标准形问题可以转化为寻找线性变换的最好矩阵表示的问题。

由此, 所有相似等价的量都可以定义在有限维线性空间的线性变换上: 例如, 由于相似矩阵的迹相同, 我们可以定义 $\text{tr } \mathcal{A}$ 为 \mathcal{A} 的任何一个矩阵表示的 tr , 由于所有基下的矩阵表示相似, 这个定义下的值是唯一确定的, 从而定义合理。由此, 特征值、代数重数、几何重数、特征多项式、迹、秩、行列式, 都可以定义在线性变换上 (我们事实上主要关心与特征系统相关的部分)。

不过，我们之前已经单独定义它的特征值与特征向量，这与使用矩阵表示定义的特征值有何关系呢？我们用下面三个命题说明，这里仍假设 \mathcal{A} 是 \mathbb{K} 上 n 维线性空间 U 上的线性变换， U 的一组基为 S ，此基下 \mathcal{A} 的矩阵表示为 A_S ：

- \mathcal{A} 的特征值集合与 A_S 的特征值集合相同。

证明：

若 $\mathcal{A}(u) = \lambda u$ 对非零 u 成立，利用第二个等价定义可知

$$A_S u_S = (\mathcal{A}(u))_S = (\lambda u)_S = \lambda u_S$$

利用坐标定义 u 为非零向量时 u_S 也非零，从而 λ 是 A_S 特征值。

若 $A_S x = \lambda x$ 对非零 x 成立，利用第一个等价定义知

$$\mathcal{A}(\pi_S^{-1}(x)) = \pi_S^{-1}(A_S \pi_S(\pi_S^{-1}(x))) = \pi_S^{-1}(A_S x) = \pi_S^{-1}(\lambda x) = \lambda \pi_S^{-1}(x)$$

由 π_S^{-1} 为同构，考虑单射性质可知 x 非零时 $\pi_S^{-1}(x)$ 非零，即得 λ 为 \mathcal{A} 特征值。

- \mathcal{A} 与 A_S 的特征值 λ 的特征子空间满足 $\pi_S(\text{Ker}(\lambda I - \mathcal{A})) = \text{Ker}(\lambda I - A_S)$ 。

证明：

上方证明中事实上已经说明

$$u \in \text{Ker}(\lambda I - \mathcal{A}) \Rightarrow u_S \in \text{Ker}(\lambda I - A_S)$$

$$x \in \text{Ker}(\lambda I - A_S) \Rightarrow \pi_S^{-1}(x) \in \text{Ker}(\lambda I - \mathcal{A})$$

记 $W = \text{Ker}(\lambda I - \mathcal{A})$ 、 $V = \text{Ker}(\lambda I - A_S)$ ，由第一个式子根据定义可知

$$\pi_S(W) \subset V$$

对任何 $x \in V$ ，由于 $\pi_S^{-1}(x) \in W$ ，有 $x = \pi_S(\pi_S^{-1}(x)) \in \pi_S(W)$ ，从而 $V \subset \pi_S(W)$ 。综合两者得证。

- A_S 的特征值 λ 的几何重数（这也是 \mathcal{A} 的特征值 λ 的几何重数定义）为 $\dim \text{Ker}(\lambda I - \mathcal{A})$ 。

证明：

由于 π_S 是同构，它是单同态。我们可以从单同态推出 $\dim \pi_S(W) = \dim W$ (W 的定义如上个证明中)。

大家可以考虑使用基进行证明，这里提供一个利用限制映射的比较简单的证法：考虑 $\pi_S|_{W \rightarrow \mathbb{K}^n}$ ，利用第一同构定理

$$\dim W = \dim \text{Ker } \pi_S|_{W \rightarrow \mathbb{K}^n} + \dim \text{Im } \pi_S|_{W \rightarrow \mathbb{K}^n}$$

由已经证明的限制映射像与核性质，上式可改写为

$$\dim W = (\dim \text{Ker } \pi_S \cap W) + \dim \pi_S(W)$$

但由 π_S 是单同态， $\text{Ker } \pi_S = \{0\}$ ，从而第二项为 0，于是得证。

由于 λ 对 A_S 的几何重数定义为 $n - \text{rank}(\lambda I - A_S)$ ，利用解空间维数定理知即为 $\dim \text{Ker}(\lambda I - A_S)$ ，再由上个证明中已证即得

$$\dim \text{Ker}(\lambda I - A_S) = \dim \pi_S \text{Ker}(\lambda I - \mathcal{A}) = \dim \text{Ker}(\lambda I - \mathcal{A})$$

从而成立。

这样，我们就把方阵的特征系统理论对应上了线性变换的特征系统——至少特征值、特征向量与几何重数可以定义。而在之前，我们已经知道秩的对应关系

$$\text{rank } A_S = \dim \text{Im } \mathcal{A}$$

事实上，就如我们已经证明的，利用解空间维数定理可知 $\text{rank } A_S$ 也可以看作 n 减特征值 0 的几何重数，这对 \mathcal{A} 也相同。

很自然地，我们会开始思考代数重数与特征多项式对应着什么，不过，这个问题只有在下一章介绍根子空间时才能解决，而本章的最后，我们需要谈到另一个重要的矩阵相似不变量：化零多项式。

20.3.3 化零多项式

假设 \mathcal{A} 是 \mathbb{K} 上线性空间 U 上的线性变换，化零多项式的定义很简单：若 \mathbb{K} 上的多项式 f 满足 $f(\mathcal{A}) = \mathcal{O}$ (\mathcal{O} 表示所有元素映射到 0 的线性变换零变换)，它就称为线性变换 \mathcal{A} 的化零多项式。

当然， $f(x) = 0$ 是化零多项式，而对于一般的线性变换，未必存在非零的化零多项式。例如，考虑 $\mathbb{C}[x]$ 上的线性变换 $\varphi(f(x)) = xf(x)$ ，对非零多项式 g ，直接计算可发现 $g(\varphi)(f(x)) = g(x)f(x)$ ，考虑维数知当 $g(x)$ 为非零多项式时， $\text{Ker } g(\varphi) = \{0\}$ 恒成立，自然不可能 $g(\varphi) = \mathcal{O}$ 。

不过，当 U 为有限维时，线性变换的非零化零多项式一定存在。

证明：

设 $\dim U = n$ ，由于 $\text{Hom}(U, U)$ 为 n^2 维线性空间，其中的 $n^2 + 1$ 个向量 $I, \mathcal{A}, \mathcal{A}^2, \dots, \mathcal{A}^{n^2}$ 必然线性相关，于是存在 $n^2 + 1$ 个不全为 0 的系数 a_0, \dots, a_{n^2} 使得

$$\sum_{i=0}^{n^2} a_i \mathcal{A}^i = \mathcal{O}$$

这就得到了一个非零的至多 n^2 次的化零多项式。

完全类似，对于一个矩阵 $A \in \mathbb{K}^{n \times n}$ ，定义使得 $f(A) = O$ 的多项式 f 为 A 的化零多项式，则线性变换的化零多项式与其矩阵表示的化零多项式集合相同。

证明：

由于已经证明 $f(\mathcal{A})$ 的矩阵表示为 $f(A)$ ，且通过矩阵表示是线性映射到矩阵的同构可知一个变换是零变换当且仅当矩阵表示为 O ，从而得证。

* 由此结论可以自然得到相似的矩阵化零多项式相同，或由上学期知识通过 $P^{-1}f(A)P = f(P^{-1}AP)$ 并消去可逆矩阵得到 $f(A) = O$ 当且仅当 $f(P^{-1}AP) = O$ 。

不过，上述的化零多项式集合作为 $\mathbb{K}[x]$ 的一个子集，还是不具有清晰的结构。由此，我们需要引入下面的定理：对有限维 \mathbb{K} 上线性空间 U 上的线性变换 \mathcal{A} ，存在非零首一多项式 f ，使得所有化零多项式的集合为 (f) ，即 f 的所有倍式集合。称此多项式 f 为 \mathcal{A} 的最小多项式，记作 $m_{\mathcal{A}}$ 。

证明：

将化零多项式集合记作 N_A 。若 $f \in N_A$ ，利用零映射定义可知对任何多项式 h 有 $f(\mathcal{A})h(\mathcal{A}) = \mathcal{O}$ ，从而 $fh \in N_A$ ，若 $f, g \in N_A$ ，利用零映射加法零元性质可知 $f(\mathcal{A}) + g(\mathcal{A}) = \mathcal{O} + \mathcal{O} = \mathcal{O}$ ，从而 $f + g \in N_A$ 。

利用这两条性质，由于已知 N_A 中有非零多项式，可以取出其中次数最小的非零多项式，利用第一条性质可知乘非零数后仍属于 N_A ，于是乘非零数将其化为首一，记为 m ，下面证明 m 即为所求的多项式。

首先，根据第一条性质，的确有 $(m) \subset N_A$ ，从而只需证明第二条。若 N_A 有 (m) 外的元素 p ，利用带余除法可知存在多项式 q 与非零（若为 0 则 $p \in (m)$ ）且次数小于 m 的多项式 r 使得

$$p = qm + r$$

但利用两条性质，根据 $p \in N_A$ 、 $m \in N_A$ 可得 $r = p - qm \in N_A$ ，这就与 m 是次数最小的非零多项式矛盾。

* 参考本讲义 14.1.4，利用多项式集合是一个主理想整环，开始证明的两条性质即说明 N_A 是理想，于是必然是主理想，这就直接得到了证明。

完全类似，将矩阵 A 看成左乘 A 的线性变换，可发现矩阵的化零多项式集合具有相同的性质，将矩阵 A 的最小多项式记为 m_A 。

* 这里的最小性是本讲义 14.1.2 提到的整除作为序关系意义下的最小性，最好不要简单理解为非零且次数最低。

* 由于证明过程并未用到非零化零多项式的存在性或矩阵表示，上述最小多项式定义对无穷维线性变换 \mathcal{A} 也成立，只要要求 \mathcal{A} 没有非零化零多项式时最小多项式定义为 $\mathbf{0}$ （作为化零多项式集合中能整除其他所有的元素）。

我们还可以证明，对于 n 维 \mathbb{K} 上线性空间 U 上的线性变换 \mathcal{A} ，设其特征多项式（以任意矩阵表示的特征多项式定义）为 $\varphi_{\mathcal{A}}$ ，则 $\varphi_{\mathcal{A}}(\mathcal{A}) = \mathcal{O}$ ，这称为 **Hamilton-Caylay 定理**（可简写为 H-C 定理）。

证明：

此证明事实上在本讲义 9.3.3 已经写过。由于 $f(\mathcal{A}) = \mathcal{O}$ 当且仅当 $f(A) = O$ ，我们只需说明对 \mathbb{K} 上方阵 A 有 $\varphi_A(A) = O$ 。

由于特征多项式不会随 A 看作的数域改变，将 A 看作复方阵不影响结论。利用相似三角化，设 $A = P^{-1}UP$ ， U 为上三角阵，则 $\varphi_A(A) = P^{-1}\varphi_A(U)P$ 。为证 $\varphi_A(A) = O$ ，只需证明对于对角元为 $\lambda_1, \dots, \lambda_n$ 的上三角阵 U 有

$$\prod_{i=1}^n (\lambda_i I - U) = O$$

对 U 的阶数归纳，当 U 一阶时直接成立，若对 $n-1$ 阶成立，利用 U 的多项式的可交换性将结果改写为

$$(\lambda_1 - U) \prod_{i=2}^n (\lambda_i I - U)$$

由于上三角阵的乘积还是上三角阵，且 U 的右下角 $(n-1) \times (n-1)$ 阶子矩阵 U_0 对角元为 $\lambda_2, \dots, \lambda_n$ ，分块为

$$U = \begin{pmatrix} \lambda_1 & x \\ \mathbf{0} & U_0 \end{pmatrix}$$

后根据归纳假设可得第二项成为

$$\begin{pmatrix} \lambda_1 & y \\ \mathbf{0} & O \end{pmatrix}$$

而第一项由于是 $\lambda_1 I - U$ ，必然是第一个对角元为 0 的上三角阵，从而是

$$\begin{pmatrix} 0 & x \\ \mathbf{0} & X_{(n-1) \times (n-1)} \end{pmatrix}$$

直接计算可得乘积为 O 。

* 此结论的最简单证明方式就是直接相似三角化进行计算，不过注意需要先考虑更大的数域。

由此，根据最小多项式的性质有 $\varphi_{\mathcal{A}} \in (m_{\mathcal{A}})$ ，即 $m_{\mathcal{A}} | \varphi_{\mathcal{A}}$ ，由于特征多项式在 \mathbb{K} 中的根一定是特征值，可知最小多项式的根也为特征值，且每个根的重数不超过特征多项式中的重数。这个定理事实上给了最小多项式的次数一个上界：至多为 n 次。

最后，我们来算一算 Jordan 标准形（自然，它是 \mathbb{C} 上的方阵）的最小多项式。具体来说分为三步：

1. 若 A_1, A_2 为方阵, $\text{diag}(A_1, A_2)$ 的最小多项式为 $\text{lcm}(m_{A_1}, m_{A_2})$ 。

证明：

由于

$$f(\text{diag}(A_1, A_2)) = \text{diag}(f(A_1), f(A_2))$$

从而 $f(\text{diag}(A_1, A_2)) = O$ 当且仅当 $f(A_1) = O, f(A_2) = O$, 于是 $\text{diag}(A_1, A_2)$ 的化零多项式集合为 A_1, A_2 化零多项式集合的交集, 而由于 A_1, A_2 化零多项式集合分别为 $(m_{A_1}), (m_{(A_2)})$, 利用最小公倍式性质即得交集为 $(\text{lcm}(m_{A_1}, m_{A_2}))$, 再由最小公倍式首一即得证。

2. Jordan 块 $J_n(\lambda)$ 的最小多项式为 $m_{J_n(\lambda)}(x) = (x - \lambda)^n$ 。

证明：

直接计算可发现其特征多项式为 $\varphi_{J_n(\lambda)}(x) = (x - \lambda)^n$, 根据 C-H 定理得 $m_{J_n(\lambda)}(x)$ 一定是 $(x - \lambda)^n$ 的因式, 从而是 $(x - \lambda)^k, 0 \leq k \leq n$ 。

直接计算发现 $J_n(\lambda) - \lambda I = J_n(0)$, 且 $J_n(0), J_n(0)^2$ 到 $J_n(0)^{n-1}$ 均非零 (利用上学期知识计算即得 $J_n(0)^k$ 在 $j-i=k$ 的位置为 1, 其他为 0), 于是 $x - \lambda, (x - \lambda)^2$ 到 $(x - \lambda)^{n-1}$ 均不为化零多项式, 最小多项式只能为 $(x - \lambda)^n$ 。

3. Jordan 标准形 $J = \text{diag}(J_{n_1}(\lambda_1), \dots, J_{n_k}(\lambda_k))$ 的最小多项式计算。

解答：

假设特征值 λ_i 对应的 Jordan 块阶数为 $a_{i1} \geq a_{i2} \geq \dots \geq a_{ir_i}$, 不同特征值为 $\lambda_1, \dots, \lambda_k$, 我们下面说明其最小多项式为

$$m_J(x) = \prod_{i=1}^k (x - \lambda_i)^{a_{i1}}$$

利用第一部分证明归纳即得此对多个分块的对角阵也成立, 再利用第二部分证明可知每个对角块的最小多项式, 从而最终得到的最小多项式为以下所有多项式的最小公倍式:

$$(x - \lambda_i)^{a_{i1}}, (x - \lambda_i)^{a_{i2}}, (x - \lambda_i)^{a_{ir_i}}, \quad i = 1, \dots, k$$

我们首先计算对每个 i 的这些多项式的最小公倍式, 因为它们是同一个多项式的次方, 可以得到最小公倍式为最大的次方, 于是问题化为以下所有多项式的最小公倍式:

$$(x - \lambda_1)^{a_{11}}, (x - \lambda_2)^{a_{21}}, \dots, (x - \lambda_k)^{a_{k1}}$$

由于这些多项式两两根不相同, 利用唯一因子分解定理计算最小公倍式即得到必然为它们的乘积。

标准形的最小多项式有几个重要推论:

- 将 A 看作不同数域上的方阵, 最小多项式不变。

证明：

可以发现, 利用特征方阵理论, 上述的证明过程构造出的最小多项式 $m_0(x)$ 是 $xI - A$ 看作 $\mathbb{C}[x]$ 上方阵的最后一个不变因子, 而不变因子不随数域改变, 因此若 $A \in \mathbb{K}^{n \times n}$, 必然 $m_0 \in \mathbb{K}[x]$ 。

另一方面, $\mathbb{K}[x]$ 是 $\mathbb{C}[x]$ 的子集, $\mathbb{C}[x]$ 中不存在次数更小的非零化零多项式, $\mathbb{K}[x]$ 也自然不存在, 再由首一性, m_0 也是 $\mathbb{K}[x]$ 上的最小多项式。

- 复方阵 A 可对角化当且仅当最小多项式无重根。

证明：

A 可对角化当且仅当标准形中所有 Jordan 块都是一阶, 这又当且仅当 $a_{11} = a_{21} = \dots = a_{k1} = 1$, 从而等价于最小多项式没有二次项。

- 复方阵 A 满足 $m_A = \varphi_A$ 当且仅当其每个特征值 Jordan 块唯一。

证明：

利用 Jordan 标准形上三角，特征多项式为以下所有多项式的乘积：

$$(x - \lambda_i)^{a_{i1}}, (x - \lambda_i)^{a_{i2}}, (x - \lambda_i)^{a_{ir_i}}, \quad i = 1, \dots, k$$

而最小多项式相当于对每个 i 只保留第一项进行乘积，从而两者相等价于对每个 i 只有一项。

* 这样的方阵可以称为**循环变换**，具体含义会在下一章说明，之所以不叫循环方阵是因为教材中用循环方阵称呼别的方阵了。

* 直接计算 Jordan 标准形的特征向量可以发现特征值的几何重数等于 Jordan 块数量，因此上述结论可以等价于每个特征值几何重数为 1。

二十一 相似标准形再探

§21.1 习题解答

这里将给出每题的书写注意事项，以作为考试书写的提示。

- (丘书 习题 9.4.10) 设 \mathbb{C} 上 n 维线性空间 V 上的线性变换 \mathcal{A} 在一组基 $\alpha_1, \dots, \alpha_n$ 下的矩阵 A 为

$$A = \begin{pmatrix} & & 1 & & \\ & & \ddots & & \\ & & & & 1 \\ -a_0 & -a_1 & \dots & & -a_{n-1} \end{pmatrix}$$

求 \mathcal{A} 的特征多项式和属于每个特征值的全部特征向量，并判断它是否可对角化。

解答：

直接利用上学期行列式知识计算可知（可发现 A 为本讲义 17.2.2 已经计算过特征多项式的友方阵的转置） A 的特征多项式为 $\varphi(\lambda) = \lambda^m + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$ 。这就是 \mathcal{A} 的特征多项式定义。

对此多项式的某个根 λ_i ，直接求解

$$Ax = \lambda_i x$$

可发现对应方程组（设 x_j 为 x 的第 j 个分量）

$$x_{j+1} = \lambda_i x_j, \quad j = 1, \dots, n-1$$

$$\lambda_i x_n = -a_0 x_1 - a_1 x_2 - \dots - a_{n-1} x_n$$

将前 $n-1$ 个方程依次代入可得到通项

$$x_j = \lambda_i^{j-1} x_1, \quad j = 1, \dots, n$$

将这些结果代入最后一个方程，发现最后得到的方程为 $\varphi(\lambda_i)x_1 = 0$ ，由于已知 $\varphi(\lambda_i) = 0$ ，此式必然满足，而从通项中即可得到 λ_i 对应的 A 特征向量为

$$\{(x_1, \lambda_i x_1, \dots, \lambda_i^{n-1} x_1)^T \mid x_1 \neq 0\}$$

将它们看作坐标即得 λ_i 对应的 \mathcal{A} 特征向量（可参考本讲义 20.3.2）为

$$\left\{ x_1 \sum_{k=1}^n \lambda_i^{k-1} \alpha_i \mid x_1 \neq 0 \right\}$$

由此, A 的每个特征值对应的特征子空间都是一维的, 几何重数为 1, 利用可对角化等价于每个特征值代数重数等于几何重数, 若每个特征值代数重数为 1, 即 $\varphi(\lambda)$ 无重根, A 即可对角化, 否则 A 不可对角化。根据 \mathcal{A} 可对角化的定义即得当且仅当 $\varphi(\lambda)$ 无重根时 \mathcal{A} 可对角化。

* 本题由于问的内容较多, 计算特征多项式的过程不是本学期重点, 可以略写 (比如直接写一个归纳得特征多项式为, 但不许写显然)。不过, 书写特征向量时注意不要忘了从坐标还原回 V 中向量 (很多题都需要类似的最后一步, 不写会适当扣分)。最后, 进行对角化判定时尽量说清楚所用的判定定理。

2. (丘书习题 9.4.11) 设 \mathcal{A} 是数域 \mathbb{K} 上 n 维线性空间 V 上的线性变换, 证明:

(1) \mathcal{A} 的特征多项式的所有复根和等于 $\text{tr}(\mathcal{A})$ 。

证明:

由特征多项式与 tr 的定义只需对方阵 $A \in \mathbb{K}^{n \times n}$ 说明成立, 设其第 i 行第 j 列 a_{ij} 。考虑完全展开式, 直接计算可发现 $\varphi_A(\lambda) = \det(\lambda I - A)$ 的 $n-1$ 次项只会在对角元乘积 $\prod_{i=1}^n (\lambda - a_{ii})$ 中出现 (否则至多 $n-2$ 次), 而此式中的 $n-1$ 次项需要选择 $n-1$ 个 λ 与剩下元素, 于是为

$$-a_{11} - a_{22} - \cdots - a_{nn} = -\text{tr } A$$

另一方面, 利用 $\varphi_A(\lambda) = \prod_{i=1}^n (\lambda - \lambda_i)$ 与上类似可知其 $n-1$ 次项为 $-\lambda_1 - \cdots - \lambda_n$, 从而得到

$$\lambda_1 + \cdots + \lambda_n = \text{tr } A$$

结论成立。

(2) \mathcal{A} 的特征多项式的所有复根乘积等于 $\text{tr}(\mathcal{A})$ 。

证明:

由特征多项式与 \det 的定义只需对方阵 $A \in \mathbb{K}^{n \times n}$ 说明成立, 设其第 i 行第 j 列 a_{ij} 。考虑完全展开式, 直接计算可发现 $\varphi_A(\lambda) = \det(\lambda I - A)$ 的常数项必须在完全展开式中每一项都选取常数项, 从而其为 $\det(-A)$ 的完全展开式结果, 等于

$$\det(-A) = (-1)^n \det A$$

另一方面, 利用 $\varphi_A(\lambda) = \prod_{i=1}^n (\lambda - \lambda_i)$ 可知常数项为 $\prod_{i=1}^n (-\lambda_i) = (-1)^n \prod_{i=1}^n \lambda_i$, 从而得到

$$\prod_{i=1}^n \lambda_i = \det A$$

结论成立。

* 本题千万不说由矩阵论知识 $\text{tr } A$ 为 A 特征多项式所有复根之和, 因为直接使用此定理显然跳过了本题实际考察的部分 (真不确定一个结论是否能用时可以在考场上向助教提问)。两问的最后一部分说明也可以直接写为由韦达定理得, 这类熟知的定理一般是允许使用的。

3. (丘书习题 9.5.1) 对于复数域上的三阶方阵 A , 定义 \mathbb{C}^3 上的线性变换 $\mathcal{A}(\alpha) = A\alpha$, 当

$$A = \begin{pmatrix} 4 & 7 & -3 \\ -2 & -4 & 2 \\ -4 & -10 & 4 \end{pmatrix}$$

时, 求 \mathcal{A} 的所有不变子空间。

解答:

直接计算求解可知 A 的特征值为 $2, 1+i, 1-i$, 特征子空间分别为

$$V_1 = \langle (2, -1, -1)^T \rangle, \quad V_2 = \langle (1-2i, i-1, -2)^T \rangle, \quad V_3 = \langle (1+2i, -i-1, -2)^T \rangle$$

利用教材 9.5 节例 12 特征值互不相同的有限维复线性变换的所有不变子空间结论, 可知 \mathcal{A} 共有八个不变子空间

$$\{0\}, \quad V_1, \quad V_2, \quad V_3, \quad V_1 \oplus V_2, \quad V_1 \oplus V_3, \quad V_2 \oplus V_3, \quad \mathbb{C}^3$$

* 本章中将给出不变子空间个数有限时所有不变子空间的一般性结论。

* 计算为主的题目基本可以直接运用已经证明的结论, 即使是例题中的。不过, 要尽量说清楚结论的来源和内容, 其中结论内容是必须要说的, 结论来源实在不记得也可以不写。

4. (丘书习题 9.5.7) 求

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

的一个非零的零化多项式。

解答:

直接计算可得其特征多项式为 λ^n , 从而由 H-C 定理 $\varphi(\lambda) = \lambda^n$ 即为所求。

* 考试大概没有这么简单的证明题, 但哪怕真出现了也尽量编点过程, 比如直接计算的话最好写出 A^k 的具体形式, 不能直接说显然 $A^n = O$ 。

5. (丘书习题 9.6.12) 设 \mathbb{K} 上 n 维线性空间 V 上的线性变换 \mathcal{A} 在一组基 $\alpha_1, \dots, \alpha_n$ 下的矩阵 A 为

$$A = \begin{pmatrix} & & -a_0 \\ 1 & & -a_1 \\ & \ddots & \vdots \\ & 1 & -a_{n-1} \end{pmatrix}$$

证明 $\mathcal{C}(\mathcal{A}) = \mathbb{K}[\mathcal{A}]$, 且 $\dim \mathcal{C}(\mathcal{A}) = n$ 。

证明:

利用矩阵表示的定义直接计算可发现条件等价于

$$\mathcal{A}(\alpha_i) = \alpha_{i+1}, \quad i = 1, 2, \dots, n-1$$

$$\mathcal{A}(\alpha_n) = -a_0\alpha_1 - \cdots - a_{n-1}\alpha_n$$

从前 $n-1$ 个式子可以得出

$$\alpha_k = \mathcal{A}^{k-1}(\alpha_1), \quad k = 1, \dots, n$$

由于 $\alpha_1, \dots, \alpha_n$ 为一组基, 这也就得到了 $\alpha_1, \mathcal{A}(\alpha_1), \dots, \mathcal{A}^{n-1}(\alpha_1)$ 线性无关, 由此, 若 \mathcal{A} 存在次数小于 n 的非零化零多项式, 设其为 $f(\mathcal{A})$, 则 $f(\mathcal{A})(\alpha_1) = 0$, 展开可发现与线性无关矛盾。又由于 H-C 定理, 特征多项式是化零多项式, 即得 \mathcal{A} 的特征多项式等于最小多项式。

上方已经证明了 $\mathcal{I}, \mathcal{A}, \dots, \mathcal{A}^{n-1}$ 线性无关 (否则仍与 $\alpha_1, \mathcal{A}(\alpha_1), \dots, \mathcal{A}^{n-1}(\alpha_1)$ 线性无关矛盾), 而再添加 \mathcal{A}^n 后由特征多项式为 n 次可知将线性相关, 于是它们是 $\mathbb{K}[\mathcal{A}]$ 的极大线性无关组, 即得到 $\dim \mathbb{K}[\mathcal{A}] = n$ 。

下面说明 $\mathcal{C}(\mathcal{A}) = \mathbb{K}[\mathcal{A}]$ 。若 $\mathcal{B} \in \mathcal{C}(\mathcal{A})$, 由于 $\mathcal{B}(\alpha_1) \in V = \langle \alpha_1, \dots, \mathcal{A}^{n-1}(\alpha_1) \rangle$, 存在一个至多 $n-1$ 次的多项式 g 使得 $\mathcal{B}(\alpha_1) = g(\mathcal{A})(\alpha_1)$ 。由此, 利用可交换得

$$\mathcal{B}(\alpha_2) = \mathcal{B}\mathcal{A}(\alpha_1) = \mathcal{A}\mathcal{B}(\alpha_1) = \mathcal{A}g(\mathcal{A})(\alpha_1) = g(\mathcal{A})\mathcal{A}(\alpha_1) = g(\mathcal{A})(\alpha_2)$$

同理可归纳得 \mathcal{B} 与 $g(\mathcal{A})$ 在 $\alpha_1, \dots, \alpha_n$ 上的像均相同, 由基映射确定线性映射可知 $\mathcal{B} = g(\mathcal{A})$, 于是 $\mathcal{C}(\mathcal{A}) \subset \mathbb{K}[\mathcal{A}]$, 由线性变换的多项式可交换知另一边包含关系成立, 综合得证。

* 本题到底能不能直接写 A 的特征多项式等于最小多项式是需要确认的, 这里给出了一个直接证明的方法, 核心思路是将全空间看作向量 α_1 生成的循环子空间, 这其实就是本讲义 20.3.3 所说的循环变换的含义。满足 $m_{\mathcal{A}} = \varphi_{\mathcal{A}}$ 的循环变换的等价条件与性质可以进行简单整理。

6. (丘书习题 9.6.13) 设 \mathbb{K} 上 n 维线性空间 V 上的线性变换 \mathcal{A} 满足特征多项式在 \mathbb{K} 上可分解成一次因式乘积, 假设 λ 是 \mathcal{A} 的代数重数为 r 的特征值, 证明

$$\text{rank}(\mathcal{A} - \lambda I)^r = n - r$$

证明:

利用代数重数与秩定义可知只需对 \mathbb{K} 上方阵 A 证明成立。由特征多项式可分解为一次因式, A 可以在 \mathbb{K} 上相似为上三角方阵 U 。由此, 设 U 满足前 r 个对角元为 λ , 此后均不为 λ , 由相似不改变多项式的秩只需证明

$$\text{rank}(U - \lambda I)^r = n - r$$

将 U 分块

$$U = \begin{pmatrix} X & Y \\ O & Z \end{pmatrix}$$

其中 X 为对角元均为 λ 的 r 阶上三角方阵, Z 为对角元均非 λ 的 r 阶上三角方阵。于是

$$\lambda I - U = \begin{pmatrix} \lambda I_r - X & -Y \\ O & \lambda I_{n-r} - Z \end{pmatrix}$$

直接计算可知其 r 次方为 (Y' 为某未知 $r \times (n-r)$ 阶矩阵)

$$(\lambda I - U)^r = \begin{pmatrix} (\lambda I_r - X)^r & Y' \\ O & (\lambda I_{n-r} - Z)^r \end{pmatrix}$$

由于 $\lambda I_r - X$ 为对角元均 0 的 r 阶上三角阵, 计算得 r 次方为 0, 从而 $\lambda I - U$ 只有后 $n-r$ 列可能非 0。另一方面, 右下角 $\lambda I_{n-r} - Z$ 是对角元均非零的 $n-r$ 阶上三角阵, 其可逆, 因此 r 次方仍然可逆。于是, 我们在 $(\lambda I - U)^r$ 中找到了 $n-r$ 阶可逆子式, 但由于只有 $n-r$ 列可能非零, 不存在 $n-r+1$ 阶可逆子式, 这就说明了其秩为 $n-r$ 。

* 虽然按理说学到这节的时候尚未证明 Jordan 标准形, 如果想不起相似三角化的方法, 考试时候遇到这种题目往往是允许直接从 Jordan 标准形计算的。此外, 由于本题步骤较多, 证明对角元均为 0 的 r 阶上三角方阵 r 次方为 O 就不需要细写了。

* 当然, 本题也可以从空间角度由根子空间分解得到, 不过个人觉得上述的矩阵论方法反而是更加清晰直接的。

7. (王书) 阅读 8.7 节并写出定理 14 的完整证明。

证明:

见本讲义 17.2 的推导。

8. (丘书习题 9.6.2) 设 $A \in \mathbb{K}^{n \times n}$ 满足 $A^3 = A^2 + 4A - 4I$, 判断 A 是否可对角化。

解答:

由条件因式分解可得 $f(x) = (x-1)(x+2)(x-2)$ 是 A 的一个化零多项式, 其最小多项式应是其因式, 而由 f 无重根可知最小多项式无重根, 从而 A 可对角化。

* 本题除了使用关键的最小多项式可对角化判定(同样, 理应将这个判定依据直接写出)以外, 还一定要注意 f 未必是最小多项式。

9. (丘书习题 9.6.4) 设 \mathcal{A} 是 \mathbb{K} 上 5 维线性空间 V 上的线性变换, 在某组基下的矩阵

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \end{pmatrix}$$

判断 \mathcal{A} 是否可对角化。

解答:

由 \mathcal{A} 可对角化定义只需判断 A 是否可对角化。直接计算特征多项式可发现其为 $\lambda^2(\lambda^2-3)(\lambda-2)$, 从而 0 的代数重数为 2, 而由形式可得 $\text{rank } A = 4$, 从而 0 几何重数为 1, 几何重数小于代数重数, 不可对角化。

* 本题非常不推荐直接使用例题结论, 因为直接由定义的验证也是简单的(上述只是其中一种方法, 考虑最小多项式、限制映射等都可以)。另一个书写建议是, 由于 $\sqrt{3}$ 未必在 \mathbb{K} 中, 最好不要对 λ^2-3 进行分解。

10. (丘书习题 9.6.7) 定义 \mathbb{R}^3 上的线性变换 P_1 满足 $P_1((x, y, z)^T) = (x, y, 0)^T$, 求其最小多项式。

解答:

首先, 不存在 $\lambda \in \mathbb{R}$ 使得 $P_1((x, y, z)^T) = \lambda(x, y, z)^T$ 恒成立, 从而其最小多项式不为一次, 而直接计算可发现对任何 $x, y, z \in \mathbb{R}$ 有

$$(P_1^2 - P_1)((x, y, z)^T) = P_1((x, y, 0)^T) - (x, y, 0)^T = (x, y, 0)^T - (x, y, 0)^T = 0$$

从而利用最小多项式是次数最小的首一化零多项式, $m(x) = x^2 - x$ 为其最小多项式。

* 本题无论是直接用映射还是矩阵表示计算, 都必须先说明最小多项式不是一次。另外, 由于这是简单计算题, 也非常不推荐使用投影映射的结论, 而应该直接从定义验证。

11. (丘书习题 9.6.9) 设 $A \in \mathbb{K}^{n \times n}$ 、 $B \in \mathbb{K}^{m \times m}$, 证明若 A 的最小多项式 m_1 与 B 的最小多项式 m_2 互素, 则矩阵方程

$$XA = BX$$

对 $X \in \mathbb{K}^{m \times n}$ 成立当且仅当 $X = O$ 。

证明:

由 $XA = BX$ 有 $XA^2 = BXA = B^2X$, 同理归纳可知 $XA^k = B^kX$ 对任何自然数 k 成立, 对这些方程两端进行数乘并相加后可得对任何多项式 $f \in \mathbb{K}[x]$ 有

$$Xf(A) = f(B)X$$

取 $f = m_1$, 由 $m_1(A) = O$ 可得 $m_1(B)X = O$ 。

由于 m_1, m_2 互素, 由裴蜀定义可知存在多项式 $u, v \in \mathbb{K}[x]$ 使得

$$u(x)m_1(x) + v(x)m_2(x) = 1$$

代入 B , 利用 $m_2(B) = O$ 得到

$$u(B)m_1(B) = I$$

从而对 $m_1(B)X = O$ 左乘 $u(B)$ 可知 $X = O$, 得证。

* 对有一定难度的证明题, 过程不会有太细节的步骤要求, 但务必写清楚关键步骤。例如这里从 $XA = BX$ 推出 $Xf(A) = f(B)X$ 再代入最小多项式的过程, 和用裴蜀定理证明 $m_1(B)$ 可逆的过程。

* 由本题证明过程可推出的结论: 若 f 和 A 的最小多项式互素, 则 $f(A)$ 可逆。事实上此命题逆命题也成立, 考虑看成 \mathbb{C} 上方阵相似三角化即可。

* 一位同学给出了非常有趣的相抵标准形证法, 这是很不错的从上学期矩阵论操作中得到的思路:

证明:

若有非零解 X , 设 $X = P \text{diag}(I_r, O)Q$, 其中 P, Q 可逆, 由非零可知 $r > 0$ 。代入并同左乘 P^{-1} 、右乘 Q^{-1} 可得

$$\text{diag}(I_r, O)QAQ^{-1} = P^{-1}BP \text{diag}(I_r, O)$$

由相似不改变最小多项式, 记 $C = QAQ^{-1}$ 、 $D = P^{-1}BP$, 两者最小多项式也应互素。但由于

$$\text{diag}(I_r, O)C = D \text{diag}(I_r, O)$$

将 C, D 分块为

$$C = \begin{pmatrix} C_1 & C_2 \\ C_3 & C_4 \end{pmatrix}, \quad D = \begin{pmatrix} D_1 & D_2 \\ D_3 & D_4 \end{pmatrix}$$

使得 C_1, D_1 为 r 阶方阵, 直接计算并对比可发现

$$C_1 = D_1, \quad C_2 = O, \quad D_3 = O$$

从而设 $F = C_1 = D_1$ 有

$$C = \begin{pmatrix} F & O \\ C_3 & C_4 \end{pmatrix}, \quad D = \begin{pmatrix} F & D_2 \\ O & D_4 \end{pmatrix}$$

从而利用分块三角阵行列式知识

$$\det(\lambda I - C) = \det(\lambda I_r - F) \det(\lambda I_{n-r} - C_4), \quad \det(\lambda I - D) = \det(\lambda I_r - F) \det(\lambda I_{m-r} - D_4)$$

由于 F 至少一阶, C, D 有公共特征值 (看作复方阵), 再由所有特征值是最小多项式的根可知 C, D 最小多项式不互素 (注意看作不同数域最小多项式相同), 矛盾。

§21.2 定义准备

21.2.1 限制映射 III

之前, 我们已经讨论了线性变换的基本知识。非常自然, 任何一个对于线性映射可以定义的概念对线性变换都可以相同定义, 限制映射也不例外。不过, 对于线性变换, 我们也希望它的限制映射还能是线性变换。具体来说, 若 \mathcal{A} 是 \mathbb{K} 上线性空间 V 上的线性变换, W 是 V 的子空间, 且 $\mathcal{A}(W) \subset W$, 则我们将限制映射 $\mathcal{A}|_{W \rightarrow W}$ 简记为 $\mathcal{A}|_W$, 称为 \mathcal{A} 在 W 上的限制变换。我们将这样的 W 称为 \mathcal{A} -不变子空间。

* 直接利用定义可知, W 是 \mathcal{A} -不变子空间也等价于对任何 $w \in W$ 有 $\mathcal{A}(w) \in W$ 。

参考之前线性映射时的讨论 (主要见本讲义 20.2), 我们可以直接写出如下的性质:

- $\text{Ker } \mathcal{A}|_W = \text{Ker } \mathcal{A} \cap W$
- $\text{Im } \mathcal{A}|_W = \mathcal{A}(W)$
- 限制映射的第一同构定理, W 维数有限时

$$\dim W = \dim \text{Im } \mathcal{A}(W) + \dim \text{Ker } \mathcal{A} \cap W$$

- V 维数有限时, 取 W 的一组基扩充为 V 一组基, \mathcal{A} 在这组基下的矩阵表示为分块上三角阵, 且两个对角块都为方阵。
*之所以能要求方阵, 是因为左上对角块即为 $\mathcal{A}|_W$ 的矩阵表示, 为方阵, 而计算阶数可知右下对角块也是方阵。
- V 维数有限时, 若 \mathcal{A} 在某组基下的矩阵表示为两个对角块都是方阵的分块上三角阵, 若第一个对角块为 r 阶, 则这组基的前 r 个生成的子空间为 \mathcal{A} -不变子空间。
- V 维数有限时, 若 $V = W \oplus U$, 且 W, U 都是 \mathcal{A} -不变子空间, 取 W, U 的一组基组成 V 的一组基, \mathcal{A} 在这组基下的矩阵表示为分块对角阵, 且两个对角块都是方阵。
*两个对角块分别是 $\mathcal{A}|_W$ 与 $\mathcal{A}|_U$ 的矩阵表示。
- V 维数有限时, 若 \mathcal{A} 在某组基下的矩阵表示为两个对角块都是方阵的分块对角阵, 若第一个对角块为 r 阶, 则这组基的前 r 个与后 $n - r$ 个生成的子空间均为 \mathcal{A} -不变子空间。

线性变换与线性映射的一个很重要的差异是, 如果忽略多项式矩阵角度的讨论(这与我们现在讨论的内容不属于同一套理论), 我们目前对相似下的标准形并没有好的结果。而上述的性质则意味着, 若能将原空间分解为若干不变子空间的直和, 线性变换的矩阵表示就可以成为若干个分块的分块对角阵。由此, 不变子空间分解对于线性变换是格外重要的, 从标准形理论的角度, 利用不变子空间分解可以将一般的 \mathbb{K} 上方阵相似为分块对角阵(注意不同基下的矩阵表示相似), 从而大幅化简问题。不过, 在进行更进一步的讨论前, 我们需要先对不变子空间的性质与例子有基本的了解。

我们将用五个定理说明不变子空间的一些基本性质, 下仍假设 \mathcal{A} 是 \mathbb{K} 上线性空间 V 上的线性变换:

1. $\{0\}$ 与 V 是 \mathcal{A} -不变子空间。

证明:

由于线性映射一定将 0 映射至 0 , $\mathcal{A}(\{0\}) = \{0\} \subset \{0\}$, 从而零空间是 \mathcal{A} -不变子空间。

由映射定义必然 $\mathcal{A}(V) \subset V$, 从而 V 是 \mathcal{A} -不变子空间。

2. V 的一维子空间 W 是 \mathcal{A} -不变子空间当且仅当其基为 \mathcal{A} 的特征向量。

证明:

设其基为 α (由线性无关性其非零), 则 $W = \langle \alpha \rangle$, 由不变子空间定义, W 是 \mathcal{A} -不变子空间当且仅当

$$\mathcal{A}(\alpha) \in \langle \alpha \rangle$$

这也即等价于存在 λ 使得

$$\mathcal{A}(\alpha) = \lambda\alpha$$

由于 $\alpha \neq 0$, 上式即为特征向量定义。

3. 若 U, W 是 \mathcal{A} -不变子空间, $U \cap W, U + W$ 都是 \mathcal{A} -不变子空间。

证明:

对 $x \in U \cap W$, 由 $x \in U$ 有 $\mathcal{A}(x) \in U$, 由 $x \in W$ 有 $\mathcal{A}(x) \in W$, 从而 $\mathcal{A}(x) \in U \cap W$, 于是 $U \cap W$ 是 \mathcal{A} -不变子空间。

对 $x \in U + W$, 设 $x = u + w$, 且 $u \in U, w \in W$, 有 $\mathcal{A}(x) = \mathcal{A}(u + w) = \mathcal{A}(u) + \mathcal{A}(w)$, 利用不变子空间定义 $\mathcal{A}(u) \in U, \mathcal{A}(w) \in W$, 于是 $\mathcal{A}(x) \in U + W$, 于是 $U + W$ 是 \mathcal{A} -不变子空间。

4. 对任何 \mathbb{K} 上多项式 f , $\text{Im } f(\mathcal{A})$ 、 $\text{Ker } f(\mathcal{A})$ 是 \mathcal{A} -不变子空间。

证明:

对 $x \in \text{Im } f(\mathcal{A})$, 设 $x = f(\mathcal{A})(y)$, 则利用线性变换的多项式可交换 $\mathcal{A}(x) = \mathcal{A}f(\mathcal{A})(y) = f(\mathcal{A})(\mathcal{A}y) \in \text{Im } f(\mathcal{A})$, 于是 $\text{Im } f(\mathcal{A})$ 是 \mathcal{A} -不变子空间。

对 $x \in \text{Ker } f(\mathcal{A})$, 由于 $f(\mathcal{A})(x) = 0$, 利用线性变换的多项式可交换 $f(\mathcal{A})(\mathcal{A}x) = \mathcal{A}f(\mathcal{A})(x) = \mathcal{A}(0) = 0$, 于是 $\mathcal{A}(x) \in \text{Ker } f(\mathcal{A})$, $\text{Ker } f(\mathcal{A})$ 是 \mathcal{A} -不变子空间。

5. 对任何 \mathbb{K} 上多项式 f , \mathcal{A} -不变子空间 W 也是 $f(\mathcal{A})$ 的不变子空间, 且 $f(\mathcal{A})|_W = f(\mathcal{A}|_W)$ 。

证明:

设 $f(\lambda) = \sum_{i=0}^k a_i \lambda^i$ 。对任何 $x \in W$, 利用不变子空间定义有 $\mathcal{A}(x) \in W$, 再利用定义可知 $\mathcal{A}^2(x) \in W$, 同理可得 $\mathcal{A}^3(x), \dots, \mathcal{A}^k(x) \in W$, 而由线性变换多项式的定义

$$f(\mathcal{A})(x) = \sum_{i=0}^k a_i \mathcal{A}^i(x) = a_0 x + a_1 \mathcal{A}(x) + \cdots + a_k \mathcal{A}^k(x)$$

从而利用子空间对线性组合封闭可知 $f(\mathcal{A})(x) \in W$, 即得到 W 是 $f(\mathcal{A})$ 的不变子空间。由此, 根据限制映射定义可知对任何 $x \in W$ 有 $f(\mathcal{A})|_W(x) = f(\mathcal{A})(x)$, 只需证明这等于 $f(\mathcal{A}|_W)(x)$ 。

由于 $x \in W$ 时 $\mathcal{A}|_W(x) = \mathcal{A}(x)$, 可得 $(\mathcal{A}|_W)^2(x) = \mathcal{A}|_W(\mathcal{A}x) = \mathcal{A}^2(x)$, 于是同理归纳可知 $(\mathcal{A}|_W)^i(x) = \mathcal{A}^i(x)$ 对 $i = 0, 1, \dots, k$ 成立。进一步利用线性变换多项式定义得

$$f(\mathcal{A})(x) = \sum_{i=0}^k a_i \mathcal{A}^i(x) = \sum_{i=0}^k a_i (\mathcal{A}|_W)^i(x) = f(\mathcal{A}|_W)(x)$$

* 由此 \mathcal{A} 的化零多项式是 $\mathcal{A}|_W$ 的化零多项式, 因此 $\mathcal{A}|_W$ 的最小多项式一定是 \mathcal{A} 的最小多项式的因式。

* 从后两个定理中可以看出, 不变子空间和线性变换的多项式关系密切, 之后的构造也将从这个思路出发。

* 另一个重要性质(即传递性)是, W 是 \mathcal{A} -不变子空间, 则 $U \subset W$ 是 \mathcal{A} -不变子空间当且仅当 U 是 $\mathcal{A}|_W$ -不变子空间。这可以直接通过定义验证。

在本部分的最后, 我们接着前面讨论不变子空间分解。我们先引入一个概念: 设 \mathcal{A} 是 \mathbb{K} 上线性空间 V 上的线性变换, V 的一个 \mathcal{A} -不变子空间分解是指一些 V 的子空间 V_1, \dots, V_k 使得

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$$

且每个 V_i 都是非零 \mathcal{A} -不变子空间。对非零 \mathcal{A} -不变子空间 W , 我们也可以定义 W 的 \mathcal{A} -不变子空间分解, 形式与上方完全相同。

* 由于本讲义中并未定义无穷多个子空间的直和, 我们只讨论有限情况, 事实上不变子空间分解应允许分解为无穷多个的直和。

这个定义可以直接与我们之前讨论的投影映射对应(见本讲义 20.2.2): 如果 V_1, \dots, V_k 是 V 的 \mathcal{A} -不变子空间分解, 对任何 $v \in V$, 应有

$$\mathcal{A}(v) = \sum_{i=1}^k \mathcal{A}|_{V_i}(P_{V_i}^{(V_1, \dots, V_k)}(v))$$

如果用更简单的矩阵论角度看, 假设 V 为 n 维, $\dim V_i = n_i$, 这事实上也即是说, 在 V_1, \dots, V_k 的各一组基构成的 V 的基下, \mathcal{A} 可以矩阵表示为一个分块对角阵, 其对角块为 n_1, \dots, n_k 阶方阵, 对应 $\mathcal{A}|_{V_1}, \dots, \mathcal{A}|_{V_k}$ 的矩阵表示(证明与本讲义 20.2.1 完全类似)。

* 因此，不变子空间分解可以看作一个线性变换的“可对角化程度”（根据之前的讨论，相似不变量可以定义在线性变换上，是否可对角化也包含在其中）的刻画，分解出越多就越接近真正的对角阵，而可对角化意味着分解为了 n 个一维不变子空间的直和。

若 \mathcal{A} -不变子空间分解中 $k = 1$ ，也即分解就是 $V = V$ ，则称这个分解为平凡的，否则称为非平凡的；若一个 \mathcal{A} -不变子空间 V 不存在非平凡的 \mathcal{A} -不变子空间分解，则称它为不可约 \mathcal{A} -不变子空间；若 V 的 \mathcal{A} -不变子空间分解中每个 V_i 都是不可约 \mathcal{A} -不变子空间，则称这个分解为不可约 \mathcal{A} -不变子空间分解。

从“可对角化程度”的视角，这个定义也是自然的：若一个 \mathcal{A} -不变子空间分解 V_1, \dots, V_k 不是不可约的，那么一定存在某个 $V_i = V_{i1} \oplus V_{i2}$ ，使得 V_{i1} 与 V_{i2} 都是非零 \mathcal{A} -不变子空间（若 V_i 能分解为更多个不变子空间的直和，我们取第一个为 V_{i1} ，其余的直和为 V_{i2} ，利用不变子空间的和仍是不变子空间即得到）。这样，我们就得到了 V 的另一种分解

$$V = V_1 \oplus \cdots \oplus V_{i-1} \oplus V_{i1} \oplus V_{i2} \oplus \cdots \oplus V_k$$

这是一个比之前更精细的 \mathcal{A} -不变子空间分解，意味着 \mathcal{A} 的矩阵表示具有更多的对角块，更接近对角阵。于是，不可约不变子空间分解事实上意味着某种程度的不可进一步化简。

至此，我们终于得到了寻找线性变换“最好”矩阵表示的一个基本方案：我们希望对一个线性变换先进行不可约不变子空间分解，再寻找其在每个不可约不变子空间上的矩阵表示。可以想象，这样得到的标准形应当是一个分块对角阵，且每个对角块具有类似的形式。

不过，正如之前讨论过很多次的，标准形应当具有唯一性，但不可约不变子空间分解往往并不是唯一的：考虑 \mathbb{R}^n 上的恒等变换 I ，取 \mathbb{R}^n 的任何一组基 $\alpha_1, \dots, \alpha_n$ ，则 $\langle \alpha_1 \rangle, \dots, \langle \alpha_n \rangle$ 都构成 I -不变子空间分解，且一维不变子空间根据定义不可约。因此，如果上述过程合理，我们至少还需要不可约不变子空间分解具有某种意义上的唯一性。

如果 \mathcal{A} 的某个矩阵表示具有唯一的相似标准形，其作为分块对角阵，对角块数与每个对角块的阶数应唯一。由此，一个合理的期望是，我们希望不同的 \mathcal{A} -不可约不变子空间分解所分解出的不可约不变子空间个数唯一，且维数可以对应相等。这也是我们之后将证明的。

21.2.2 根子空间与分解

虽然之前讨论了很多关于不可约不变子空间分解的理论内容，但我们目前对于如何得到一个不可约不变子空间分解还没有任何思路。根据不可约不变子空间分解的定义，我们从任何一个不变子空间分解出发，不断细分下去，总能使每个不变子空间都不可约。因此，可以考虑先自上而下地得到一个不变子空间分解，再看如何进行细分。

事实上，从之前已经介绍过的内容，我们已经可以构造出一个不变子空间分解了。考虑 \mathbb{K} 上 n 维线性空间 V 上的线性变换 \mathcal{A} ，有如下结论：

- 与本讲义 18.3.3 第 6 题完全相同可以证明，若 f, g 都是 \mathbb{K} 上的多项式，且 $\gcd(f, g) = 1$ ，则

$$\text{Ker } f(\mathcal{A}) \oplus \text{Ker } g(\mathcal{A}) = \text{Ker } f(\mathcal{A})g(\mathcal{A})$$

- 在本讲义 20.3.3 中已经证明了 H-C 定理：设 $\varphi_{\mathcal{A}}$ 为 \mathcal{A} 的特征多项式，则 $\varphi_{\mathcal{A}}(\mathcal{A}) = \mathcal{O}$ ，利用 Ker 的定义可知

$$V = \text{Ker } \varphi_{\mathcal{A}}(\mathcal{A})$$

- 在本讲义 21.2.1 中，我们知道，对任何多项式 f ， $\text{Ker } f(\mathcal{A})$ 都是 \mathcal{A} -不变子空间。

综合上述三个结论，即得到根子空间分解：设 $\varphi_{\mathcal{A}}$ 为 \mathcal{A} 的特征多项式，其唯一因子分解为（见本讲义 14.1.4，由于 $\varphi_{\mathcal{A}}$ 首一且次数至少为 1，可假设因子均首一）

$$\varphi_{\mathcal{A}}(x) = p_1^{a_1}(x)p_2^{a_2}(x) \cdots p_k^{a_k}(x)$$

其中 p_1, \dots, p_k 为不同的首一不可约多项式。则有

$$V = \text{Ker } p_1^{a_1}(\mathcal{A}) \oplus \text{Ker } p_2^{a_2}(\mathcal{A}) \oplus \cdots \oplus \text{Ker } p_k^{a_k}(\mathcal{A})$$

每个 $\text{Ker } p_i^{a_i}(\mathcal{A})$ 都称为 \mathcal{A} 的根子空间。

证明：

本讲义 14.1.4 中已经证明了最大公因式可通过唯一因子分解计算：当两个多项式的因子分解中没有共同不可约因子时，最大公因式必然为 1，由此可知 $(p_1^{a_1}, p_2^{a_2}) = 1$ ，从而根据本讲义 18.3.3 第 6 题可知

$$\text{Ker } p_1^{a_1}(\mathcal{A}) \oplus \text{Ker } p_2^{a_2}(\mathcal{A}) = \text{Ker } (p_1^{a_1} p_2^{a_2})(\mathcal{A})$$

进一步地，由于 $p_1^{a_1} p_2^{a_2}$ 与 $p_3^{a_3}$ 也没有共同不可约因子，可知 $(p_1^{a_1} p_2^{a_2}, p_3^{a_3}) = 1$ ，从而再利用本讲义 18.3.3 第 6 题结论有

$$\text{Ker } (p_1^{a_1} p_2^{a_2})(\mathcal{A}) \oplus \text{Ker } p_3^{a_3}(\mathcal{A}) = \text{Ker } (p_1^{a_1} p_2^{a_2} p_3^{a_3})(\mathcal{A})$$

结合上一个式子展开得到

$$\text{Ker } p_1^{a_1}(\mathcal{A}) \oplus \text{Ker } p_2^{a_2}(\mathcal{A}) \oplus \text{Ker } p_3^{a_3}(\mathcal{A}) = \text{Ker } (p_1^{a_1} p_2^{a_2} p_3^{a_3})(\mathcal{A})$$

重复此过程，可证明等式右侧的确都为直和，且直和的结果为

$$\text{Ker } (p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k})(\mathcal{A})$$

而根据定义这即是 $\text{Ker } \varphi_{\mathcal{A}}(\mathcal{A})$ ，由 H-C 定理此为 $\text{Ker } \mathcal{O}$ ，于是是全空间 V ，得证。

为了方便讨论，我们下面先考虑 $\mathbb{K} = \mathbb{C}$ 的情况，这时，由于 \mathbb{C} 上的首一不可约多项式只能为 $p_i(x) = x - \lambda_i$ ，根子空间分解定义可以表述为：设 \mathcal{A} 的特征值为 $\lambda_1, \dots, \lambda_k$ ，对应的代数重数为 m_1, \dots, m_k ，则

$$V = \text{Ker } (\mathcal{A} - \lambda_1 \mathcal{I})^{m_1} \oplus \text{Ker } (\mathcal{A} - \lambda_2 \mathcal{I})^{m_2} \oplus \cdots \oplus \text{Ker } (\mathcal{A} - \lambda_k \mathcal{I})^{m_k}$$

每个 $\text{Ker } (\mathcal{A} - \lambda_i \mathcal{I})^{m_i}$ 都称为 \mathcal{A} 的根子空间。

证明：

由代数重数定义，复方阵的特征多项式 $\varphi_{\mathcal{A}}$ 分解为

$$\varphi_{\mathcal{A}}(x) = (x - \lambda_1)^{m_1} (x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$$

其中 m_i 为 λ_i 对应的代数重数。

而根据定义，多项式 $(x - \lambda_i)^{m_i}$ 代入 \mathcal{A} 得到的即是

$$(\mathcal{A} - \lambda_i \mathcal{I})^{m_i}$$

从而得证。

* 由于对任何多项式 f ， $\text{Ker } f(\mathcal{A})$ 都是 \mathcal{A} -不变子空间，上述根子空间分解的确是不变子空间分解。

下面，我们来介绍根子空间最重要的性质：不变子空间分解到根子空间上。若 V 的 \mathcal{A} -根子空间分解是

$$V = \text{Ker } p_1^{a_1}(\mathcal{A}) \oplus \text{Ker } p_2^{a_2}(\mathcal{A}) \oplus \cdots \oplus \text{Ker } p_k^{a_k}(\mathcal{A})$$

且 U 是一个 \mathcal{A} -不变子空间，则（由于不变子空间的交为不变子空间，下方直和中每一项都是 \mathcal{A} -不变子空间）

$$U = (\text{Ker } p_1^{a_1}(\mathcal{A}) \cap U) \oplus (\text{Ker } p_2^{a_2}(\mathcal{A}) \cap U) \oplus \cdots \oplus (\text{Ker } p_k^{a_k}(\mathcal{A}) \cap U)$$

证明：

为了方便，我们记 $V_i = \text{Ker } p_i^{a_i}(\mathcal{A})$ 。

设 \mathcal{A} 的最小多项式为 p ，由于 $p(\mathcal{A}) = \mathcal{O}$ ，利用限制映射多项式性质（见上一部分）可知

$$p(\mathcal{A}|_U) = p(\mathcal{A})|_U = \mathcal{O}$$

于是， $\mathcal{A}|_U$ 上的最小多项式一定是 $p(x)$ 的因式，设其为

$$p_U(x) = \prod_{i=1}^k p_i^{s_i}(x)$$

且 $s_i \leq a_i$ 。由此可对 U 进行 $\mathcal{A}|_U$ 根子空间分解

$$U = \text{Ker } p_1^{s_1}(\mathcal{A}|_U) \oplus \text{Ker } p_2^{s_2}(\mathcal{A}|_U) \oplus \cdots \oplus \text{Ker } p_k^{s_k}(\mathcal{A}|_U)$$

再由限制映射多项式性质可得

$$\text{Ker } p_i^{s_i}(\mathcal{A}|_U) = \text{Ker } (p_i^{s_i}(\mathcal{A}))|_U$$

再由限制映射的像与核结论可知其为

$$\text{Ker } p_i^{s_i}(\mathcal{A}) \cap U$$

于是其包含在 $U \cap V_i$ 中（由 $s_i \leq a_i$ 利用 $\text{Ker } (\mathcal{BC}) \subset \text{Ker } \mathcal{C}$ 得第一项包含在 V_i 中）。

另一方面，由于 V_1 到 V_k 为直和，利用直和等价于每个与其他所有求和交为 $\{0\}$ 可知 $U \cap V_1$ 到 $U \cap V_k$ 也是直和，而它们为 U 的子空间，因此直和包含在 U 中，由此利用每个的包含关系得到得到

$$U = \text{Ker } p_1^{s_1}(\mathcal{A}|_U) \oplus \cdots \oplus \text{Ker } p_k^{s_k}(\mathcal{A}|_U) \subset (U \cap V_1) \oplus \cdots \oplus (U \cap V_k) \subset U$$

由于左侧等于右侧，每个包含关系只能取等，即得结论。

* 事实上计算维数可进一步得到 $\text{Ker } p_i^{s_i}(\mathcal{A}|_U) = V_i \cap U$ 。这个证明中**限制映射的最小多项式是原映射最小多项式因式**结论是常用于分析限制映射的。

* 注意对一般的线性空间 V 与子空间 $U, U_1, U_2, V = U_1 \oplus U_2$ 不能推出 $U = (U_1 \cap U) \oplus (U_2 \cap U)$ （考虑 $V = \mathbb{R}^2, U_1 = \langle e_1 \rangle, U_2 = \langle e_2 \rangle, U = \langle e_1 + e_2 \rangle$ 的经典反例），由此可以看出根子空间讨论不变子空间问题的本质不同。

之所以说这是最重要的性质，是因为它说明了根子空间分解是得到不可约不变子空间分解的必经之路： V 的任何一个不可约 \mathcal{A} -不变子空间一定是某个根子空间的子空间，由此， V 的任何一个不可约 \mathcal{A} -不变子空间分解一定需要先进行根子空间分解。

证明：

我们仍记 \mathcal{A} -根子空间为 V_1 到 V_k 。设 U 是不可约 \mathcal{A} -不变子空间，利用之前结论有

$$U = (U \cap V_1) \oplus \cdots \oplus (U \cap V_k)$$

由于不变子空间的交为不变子空间，直和中每一项都是 \mathcal{A} -不变子空间，又由不可约性，直和中至多有一项不是零空间，否则即对 U 进行了分解。设 $U \cap V_i$ 外的项全为零空间，即得到

$$U = U \cap V_i$$

从而

$$U \subset V_i$$

由此，对 V 的任何一个不可约 \mathcal{A} -不变子空间分解（注意其中没有零空间，因此不可能有一个不变子空间包含在不同的 V_i 中），我们假设 $U_{i1}, U_{i2}, \dots, U_{it_i}$ 是包含于 V_i 的不变子空间，有

$$V = V_1 \oplus \cdots \oplus V_k = (U_{11} \oplus \cdots \oplus U_{1t_1}) \oplus \cdots \oplus (U_{k1} \oplus \cdots \oplus U_{kt_k})$$

由子空间的和还是子空间可知

$$U_{11} \oplus \cdots \oplus U_{1t_1} \subset V_1, \dots, U_{k1} \oplus \cdots \oplus U_{kt_k} \subset V_k$$

但由于直和的结果相等，考虑维数即得必须（否则第三项维数将小于第二项维数）

$$U_{11} \oplus \cdots \oplus U_{1t_1} = V_1, \dots, U_{k1} \oplus \cdots \oplus U_{kt_k} = V_k$$

这就得到所有不可约 \mathcal{A} -不变子空间分解都是从根子空间分解进一步分解得到。

利用此结论，我们还可以解决一些不变子空间相关的问题，例如确定一个 \mathbb{C} 上 n 维线性空间 V 上的线性变换 \mathcal{A} 的全部不变子空间，见下一章的复习题（复习题中给出了矩阵版本，对于线性变换，用矩阵表示可完全类似证明）。

最后，我们再证明一些根子空间的性质，简单起见，先考虑 \mathbb{C} 上 n 维线性空间 V 上的线性变换 \mathcal{A} ，这时我们可以称 $\text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{m_i}$ 为特征值 λ_i 对应的根子空间：

- 特征值 λ_i 对应的根子空间维数为其代数重数 m_i 。

证明：

上学期，利用相似三角化，我们已经证明了若 A 的特征值为 $\lambda_1, \dots, \lambda_n$ ，对复多项式 f ， $f(A)$ 的特征值为 $f(\lambda_1), \dots, f(\lambda_n)$ 。利用矩阵表示的性质，此结论对线性变换 \mathcal{A} 仍然成立。由此，记 $\mathcal{B} = (\mathcal{A} - \lambda_i \mathcal{I})^{m_i}$ ， \mathcal{B} 特征值为 \mathcal{A} 的特征值减 λ_i 后作 m_i 次方。

由于 \mathcal{A} 的特征值有 m_i 个 λ_i ，其余不为 λ_i ，可知 \mathcal{B} 的特征值有 m_i 个 0，其余不为 0，于是 \mathcal{B} 的特征值 0 的代数重数为 m_i 。另一方面， $\text{Ker } \mathcal{B} = \text{Ker}(\mathcal{B} - 0\mathcal{I})$ 为 0 对应的特征子空间， $\dim \text{Ker } \mathcal{B}$ 即为 \mathcal{B} 的特征值 0 的几何重数。利用几何重数不超过代数重数，有

$$\dim \text{Ker } \mathcal{B} = \dim \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{m_i} \leq m_i$$

此式对每个 i 均成立。但是，根据根子空间分解定理与代数维数定义有

$$\sum_{i=1}^k \dim \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{m_i} = n = \sum_{i=1}^k m_i$$

因此考虑维数关系可知必然 $\dim \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{m_i} = m_i$ ，从而得证。

* 这里我们通过将问题转化为比较几何重数与代数重数规避了更复杂的计算，事实上直接相似三角化后计算也可得到结论。

- 特征值 λ_i 对应的根子空间可以写为 $\text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{d_i}$ ，其中 d_i 为 λ_i 在最小多项式中的重数（即 $x - \lambda_i$ 的次数）。

证明：

在根子空间分解的证明中，可以发现用满足 $p(\mathcal{A}) = \mathcal{O}$ 的多项式 p 替代 $\varphi_{\mathcal{A}}$ ，结论仍然成立。由此，取 p 为 \mathcal{A} 的最小多项式 $d_{\mathcal{A}}$ ，利用最小多项式的性质可知其分解为

$$d_{\mathcal{A}}(x) = (x - \lambda_1)^{d_1} \cdots (x - \lambda_k)^{d_k}$$

从而进一步得到

$$V = \text{Ker}(\mathcal{A} - \lambda_1 \mathcal{I})^{d_1} \oplus \text{Ker}(\mathcal{A} - \lambda_2 \mathcal{I})^{d_2} \oplus \cdots \oplus \text{Ker}(\mathcal{A} - \lambda_k \mathcal{I})^{d_k}$$

另一方面，由最小多项式性质 $d_i \leq m_i$ ，于是 $\text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{d_i} \subset \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{m_i}$ ，对比维数可知必然有

$$\text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{d_i} = \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{m_i}$$

从而将 m_i 替换为 d_i 后仍为根子空间。

3. 记 $c_t = \dim \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^t$, 则

$$c_0 < c_1 < \cdots < c_{d_i} = c_{d_i+1} = c_{d_i+2} = \dots$$

证明:

上学期我们已经证明了, 对方阵 B , 记 $r_k = \text{rank } B^k$, 则 r_k 单调不增, 且若 $r_i = r_{i+1}$, 则此后的 r_i 都相等。

利用第一同构定理可知

$$c_t = n - \dim \text{Im}(\mathcal{A} - \lambda_i \mathcal{I})^t$$

设 B 为 $\mathcal{A} - \lambda_i I$ 在某组基下的矩阵表示, 利用秩与维数的关系可知

$$c_t = n - \text{rank } B^t$$

由此 c_t 单调不减, 且在某次不变后不再改变。

记 s 为使得 $c_s = c_{s+1}$ 的最小自然数, 我们只需证明 $s = d_i$ 即得到了结论。

分为两部分说明:

- $s \leq d_i$

利用最小多项式性质, 设 \mathcal{A} 的最小多项式为 $d_{\mathcal{A}}$, 则 $d_{\mathcal{A}}(x)(x - \lambda_i)$ 也是 \mathcal{A} 的化零多项式, 而其分解中其他 λ_j 的次数不变, λ_i 的次数增加 1, 因此有

$$V = \text{Ker}(\mathcal{A} - \lambda_1 \mathcal{I})^{d_1} \oplus \cdots \oplus \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{d_i+1} \oplus \cdots \oplus \text{Ker}(\mathcal{A} - \lambda_k \mathcal{I})^{d_k}$$

由于 $\text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{d_i} \subset \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{d_i+1}$, 完全类似对比维数可知两者必然相等, 从而只能

$$c_{d_i} = c_{d_i+1}$$

这就证明了 s 不会超过 d_i 的位置。

- $s \geq d_i$

若否, 可知必然有 $c_{d_i-1} = c_{d_i}$ (注意根据最小多项式的性质, 必然 $d_i \geq 1$, 因此 c_{d_i-1} 一定存在), 即

$$\dim \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{d_i-1} = \dim \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})_i^d$$

由于左包含于右, 通过维数相等即得

$$\text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{d_i-1} = \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})_i^d$$

将其代入最小多项式对应的分解, 可发现

$$V = \text{Ker}(\mathcal{A} - \lambda_1 \mathcal{I})^{d_1} \oplus \cdots \oplus \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{d_i-1} \oplus \cdots \oplus \text{Ker}(\mathcal{A} - \lambda_k \mathcal{I})^{d_k}$$

完全类似之前的计算可知右侧为

$$\text{Ker } p(\mathcal{A})$$

其中 $p(x) = (x - \lambda_1)^{d_1} \cdots (x - \lambda_i)^{d_i-1} \cdots (x - \lambda_k)^{d_k}$ 。

根据定义, $V = \text{Ker } p(\mathcal{A})$ 即说明 $p(\mathcal{A}) = \mathcal{O}$, 但 p 的次数比最小多项式小 1, 且并非零多项式, 与最小多项式的定义矛盾。

* 由此可以看出另一个计算最小多项式的方法: 解出 \mathcal{A} 的所有特征值 $\lambda_1, \dots, \lambda_k$ 后, 对每个 λ_i 计算上述的 c_t , 首次不再改变时得到的就是最小多项式中 λ_i 的次数。事实上, 计算上述的 c_t 可以直接得到 Jordan 标准形, 我们将在之后进行说明。

* 这部分的证明中, 我们反复利用了包含关系与维数关系结合的技巧, 这对有限维线性空间的分析十分重要, 值得熟练。

21.2.3 循环子空间

生成不变子空间-自下而上
循环向量
向量的最小多项式
存在向量最小多项式 = 最小多项式

§21.3 标准形理论

21.3.1 可对角化 V

可对角化与根子空间、特征子空间

21.3.2 循环子空间分解

循环子空间上的矩阵表示
存在不变补空间
一般空间分解为循环子空间
根子空间分解为循环子空间的不可约性

21.3.3 Jordan 标准形

根子空间分解循环子空间的矩阵表示
用秩的计算方法与 Jordan 形唯一性
空间分解唯一性-循环子空间同构
 A 与 A^T 相似
循环变换

21.3.4 有理标准形

一般数域上的循环子空间
一般数域的循环子空间分解
一般数域循环子空间分解矩阵表示
有理标准形

二十二 补充：期中复习

§22.1 知识与技巧整理

- * 按照丘书的顺序将多项式放在最后，务必至少掌握下方列出的算法，此外，由于每节的内容、结论较多，这里将顺序列举关键结论。
- * 所有算法和标注无需证明的结论都可以不用掌握证明。
- * 所有标注了解即可的结论看一下结论就行。
- * 标注简单推论的结论可以熟悉，不用强记，理应可以通过定义、定理简单推出。
- * 加粗的结论表示必须记忆的核心结论。

22.1.1 上册知识

主要包含期中需要用到的线性空间与矩阵论的基础知识，不再整理例题，尤其注意由于前半学期主要讨论的内容为线性变换，矩阵多项式与相似相关的结论非常重要。

1.1 矩阵、系数矩阵、增广矩阵定义

线性方程组通解算法

2.2 行列式定义

2.3 矩阵转置定义

行列式上三角化求值算法

2.4 行列式按行/列展开算法

Vandermonde 行列式结论 (无需证明)

3.1 \mathbb{K}^n 与其子空间定义

3.2 线性相关、线性无关定义

线性相关、线性无关的线性组合表述 (简单推论)

线性相关、线性无关的线性方程组表述 (简单推论)

线性相关、线性无关的基本性质 (简单推论)

3.3 极大线性无关组、向量组等价定义

极大线性无关组的行变换算法

向量组 S 能表出向量组 T ，且 T 个数多于 S ，则 T 线性相关 (无需证明)

极大线性无关组向量个数相同 (简单推论)

秩定义

3.4 向量空间基、维数定义

满足个数等于维数、能表出全空间、线性无关任意两个条件的向量组都是基 (简单推论)

子空间维数小于等于原空间，且维数相等可推相同 (简单推论)

矩阵行、列空间定义

3.5 行秩、列秩、矩阵秩定义

行秩等于列秩等于最大非零子式阶数 (无需证明)

矩阵秩的行列变换算法

方阵满秩等价于行列式非零 (简单推论)

一些基本的秩等式与不等式，如典型例题 7、8、9 与习题 14 (大部分简单推论)

3.6 线性方程组有解等价于系数矩阵秩等于增广矩阵秩 (简单推论)

3.7 齐次线性方程组解空间与基础解系算法

解空间维数定理 (无需证明)

* 最重要的特例：未知数个数大于方程个数的齐次线性方程组有非零解。

3.8 非齐次线性方程组解集结构 (简单推论)

* 最重要的特例：未知数个数等于方程个数的线性方程组存在唯一解当且仅当系数矩阵可逆。

4.1 矩阵加法、数乘、乘法、次方，单位阵定义

矩阵运算、转置与单位阵的基本性质 (大部分简单推论，其他无需证明，如 $(AB)' = B'A'$)

Jordan 块的次方，即典型例题 9 与习题 7

矩阵的多项式可相互交换，即典型例题 12 (无需证明)

4.2 对角矩阵、基本矩阵 E_{ij} 、三角矩阵、初等矩阵、对称矩阵、斜对称矩阵定义

与基本矩阵乘法结果 (简单推论)

与初等矩阵相乘等价于行列变换, 左乘行变换右乘列变换 (简单推论)

幂零阵、幂零指数的定义与基本性质, 即典型例题 9

4.3 对两个方阵 $\det(AB) = \det A \det B$ (无需证明) $\text{rank}(AB) \leq \max(\text{rank } A, \text{rank } B)$ (无需证明) $\text{rank}(A + B) \leq \text{rank } A + \text{rank } B$ (简单推论)

4.4 可逆矩阵定义

伴随方阵、行列变换计算逆算法

可逆的等价条件 (简单推论)

逆矩阵基本性质, 即性质 1 到 7 (无需证明)

4.5 分块矩阵在阶数相符时可以类似通常矩阵乘法计算乘法 (无需证明)

* 最重要的特例: AB 的每列等于 A 乘 B 的每列、分块为 2×2 方阵的情况。

分块初等矩阵定义与左右乘的基本性质

* 虽然期中考试不会考必须初等变换的证明, 但如果能掌握, 即相当于多一种解决问题的手段。

分块对角阵、分块三角阵定义

 $\text{diag}(I_m, I_n - AB)$ 到 $\text{diag}(I_n, I_m - AB)$ 的初等行列变换* 可以推出 $I - AB$ 、 $I - BA$ 的秩、行列式、逆矩阵关系, 本学期仍然有用。

Frobenius 秩不等式, 即习题 3

* 此结论常用于秩的估计, 如习题 4 这类题目。对它的空间/矩阵证明至少要掌握一个。

* 重要推论是 $\text{rank } B = \text{rank}(AB)$ 时 $\text{rank}(BC) = \text{rank}(ABC)$ 。* 这又可以推出 $\text{rank } A^m = \text{rank } A^{m+1}$ 则 $\text{rank } A^m = \text{rank } A^{m+k}$ 对任何自然数 k 成立。

4.7 * 本节上学期时不重要, 这学期必须熟练掌握。

映射、像、原像、单射、满射、双射、变换、函数、恒等映射、复合、可逆定义

映射的基本性质, 如可逆等价于双射、例题 1 (简单推论)

线性映射、像空间、核空间定义

像空间等于列空间, 核空间等于解空间 (简单推论)

5.1 * 这学期由于有商空间等概念, 必须熟悉等价类语言的叙述。

等价关系、等价类、商集、代表元定义

等价类基本性质 (简单推论)

5.2 相抵定义

相抵标准形定义、任何矩阵可以唯一相抵为标准形 (无需证明)

* 用相抵标准形可以证明很多秩相关的结论, 这学期同样是作为可选方法而非必须掌握。

相抵为等价关系, 相抵等价即秩相等 (简单推论)

满秩分解, 即例 3 (简单推论)

5.4 迹定义

迹的基本性质, 尤其 $\text{tr}(AB) = \text{tr}(BA)$ (简单推论)

相似、可对角化定义

相似为等价关系 (简单推论)

相似的基本性质, 尤其性质 1 可推出相似矩阵的多项式计算 (简单推论)

5.5 特特征值、特征向量、特征子空间、特征多项式、代数重数、几何重数定义

特征系统是相似不变量 (简单推论)

几何重数至少为 1, 至多为代数重数 (无需证明)

可逆等价于无 0 特特征值 (简单推论)

5.6 不同特征值的特征向量线性无关 (无需证明)

可对角化等价于存在一组都是特征向量的基 (简单推论)

可对角化等价于所有特征子空间维数和为 n (无需证明) \mathbb{K} 上可对角化等价于任何特征值代数重数等于几何重数且特征值在 \mathbb{K} 中 (简单推论)* 结合秩估算可以得到重要结论: 若 $f(A) = O$, 且 f 无重根, 则 \mathbb{C} 上方阵 A 可对角化。可对角化时 $P^{-1}AP = D$ 的 D 、 P 计算

5.7 复方阵相似三角化, 即例 6 (无需证明)

* 重要推论: 若 A 的特征值为 $\lambda_1, \dots, \lambda_n$, 则 $f(A)$ 特特征值为 $f(\lambda_1), \dots, f(\lambda_n)$, f 为多项式。* 由此可以直接计算证明 Cayley-Hamilton 定理, 设 A 特特征多项式为 φ_A , 则 $\varphi_A(A) = O$ 。

22.1.2 线性空间

注意我们只需要了解定义在数域上的线性空间, 无需掌握一般域的情况。

8.1 数域上线性空间定义

线性空间基本性质, 包含例 6 (简单推论)

一般向量组 (可能无穷个) 线性相关、线性无关、线性表出定义

一般向量组线性相关、线性无关、线性表出基本性质 (简单推论, 基本同 3.2)

有限个向量的极大线性无关组、向量组等价、秩定义与基本性质 (基本同 3.2)

线性空间的基、有限维线性空间、维数定义与基本性质 (基本同 3.2)

线性空间基与维数求法 (无固定算法, 基本思路是找到能表出全空间的较简单向量组, 需多看例题)

任何线性空间存在基 (无需证明)

形式乘法的基本性质与应用 (更多运用可参考本讲义 19.3)

* 推荐掌握原因: 可以很大程度简化书写, 且省略不少复杂的求和符号操作。

基变换过渡矩阵定义

* 由于算法即为直接由定义计算, 不单独列出算法。

基变换后的坐标变换方式是左乘基变换过渡矩阵的逆 (简单推论)

 \mathbb{K} 上方阵 A 的所有多项式构成线性空间, 记为 $\mathbb{K}[A]$ (简单推论)

重要例题:

- 线性空间验证 (习题 5)
- 基与维数算法 (例 16、18、23、25, 习题 3、14)
- 基变换与坐标变换 (习题 24、25)

8.2 子空间、生成子空间、交空间、和空间定义

交、和具有交换律、结合律 (简单推论)

交对和或和对交不存在分配律, 即命题 1 与其另一边包含关系的反例 (简单推论)

有限维子空间维数定理, 即定理 4 (无需证明)

向量空间交空间、和空间基与维数算法, 即例 11 (也可参考本讲义 18.3.1)

直和、补空间定义

两个有限维子空间和是直和的等价条件, 即定理 5、定理 6 (简单推论)

有限个有限维子空间和是直和的等价条件, 即定理 8、定理 9

所有与 \mathbb{K} 上方阵 A 可交换的方阵构成线性空间，记为 $\mathcal{C}(A)$ (简单推论)

数域上线性空间有限个真子空间并集不为全空间，即例 10 (无需证明)

可对角化等价于特征子空间直和为全空间，即例 15 (5.6 节结论的简单推论)

$(f, g) = 1$ 时 $\text{Ker } (f(A)g(A)) = \text{Ker } f(A) \oplus \text{Ker } g(A)$ ，即例 16

* 可以应用在如习题 13、14 上。

和空间的生成定义，即补充题八.1、八.2(无需证明)

重要例题：

- 子空间验证与计算 (例 1、3、5)
- 直和验证与证明 (习题 16、20，补充题八.3)
- 抽象说明 (例 23、24，习题 19)

8.4 商空间、余维数定义

商空间定义合理性 (无需证明)

商空间的基扩充算法 (也可参考本讲义 18.4.1)

有限维商空间维数公式 (简单推论)

重要例题：

- 维数无穷但余维数有限的情况 (习题 4，大概率不考不过有空可以看一下)
- 商空间的过渡矩阵 (习题 2，注意扩充对应的补空间不同，但商空间仍然相同，只是取了不同基)

22.1.3 线性映射

书里把线性映射、线性变换混在一起了，个人非常不推荐，因此仍然拆开叙述。

9.1 线性映射、线性函数、零映射定义

线性映射基本性质，尤其将线性相关组映射到线性相关组 (简单推论)

确定一组基的像就能确定线性映射 (简单推论)

直和分解上的投影映射定义

$\text{Hom}(U, V)$ 构成线性空间

重要例题：

- 线性映射验证 (例 4，习题 4、5、6)

8.3 线性同构定义

线性同构基本性质，尤其是同构当且仅当一组基映射到一组基 (简单推论)

同构的逆与复合还是同构，即习题 4、5 (简单推论)

有限维线性空间同构当且仅当维数相等 (简单推论)

同构可迁移子空间，即命题 1 (简单推论)

* 有效的应用如例 4，不过这类题目需要掌握不那么技巧性的硬做方法。

A, B 相似则 $\mathcal{C}(A), \mathcal{C}(B)$ 同构，即例 7

* 应用如例 8，这是解决一般 $\mathcal{C}(A)$ 求法的关键。

商空间与补空间同构，即 8.4.2 例 1

重要例题：

- 映射相关线性空间的基与坐标求法 (例 5，同样要掌握硬做方法，参考本讲义 20.1.1 的例题)
- 通过基的映射构造同构 (习题 1、3)
- 有限维时计算维数证明同构 (8.4.2 例 7，至少要会直接计算维数证明有限维情况)

9.2 一般线性映射像空间、核空间、余核定义

用像与核确定单射、满射 (简单推论)

线性映射可迁移子空间，即例 1(1) (简单推论)

第一同构定理，即定理 1、2 (无需证明，不建议使用映射秩、零度的概念，容易与矩阵混淆)

有限维线性空间单射满射等价 (简单推论)

补充：限制映射 (可参考本讲义 19.2.4、20.2.1)

* 我们学习范围的书上并未明确介绍限制映射，但大量用到了，因此建议整体学习理论。

重要例题：

- 像与核相关的抽象说明 (例 6)

- 维数公式联系限制映射证明 (例 1、4，习题 7，注意例 4 和上册证法完全不同)

9.3 矩阵表示定义

矩阵表示的含义：坐标进行矩阵乘法，即第三部分

像空间维数即为矩阵表示秩，即例 8 (简单推论)

映射的加法、数乘、复合对应矩阵表示的加法、数乘、乘法，即定理 1、例 20 (无需证明)

$\text{Hom}(U, V)$ 与 $\mathbb{K}^{\dim U, \dim V}$ 同构

* 由此可构造 $\text{Hom}(U, V)$ 一组基。

补充：一些关于线性映射矩阵表示的基本结论 (可参考本讲义 19.3，例 9、10)

* 上方很多结论教材讲解都只对线性变换说明，破坏了其通用性，非常建议先弄明白线性映射。

基变换后的矩阵表示相抵，即习题 12

重要例题：

- 矩阵表示计算 (例 19、习题 4)

- 线性映射抽象说明 (例 21、习题 7)

22.1.4 线性变换

重中之重，但很多结论其实是无需会证明的，也必须清楚前几部分所有前置知识才能更好理解。

9.1 线性变换、幂等变换、恒等变换、数乘变换定义

线性变换的幂与多项式定义

投影是幂等变换 (简单推论)

空间的投影分解，即例 11

重要例题：

- 整体思路证明 (例 10、习题 11，注意对一般线性变换只能采用整体思路，不再有矩阵的局部)

9.2 幂等变换是投影，即命题 3

重要例题：

- 有限维线性变换的单射满射等价 (例 11、12)

- 线性变换多项式与维数 (习题 8)

9.3 线性变换矩阵表示的定义

线性变换可逆等价于矩阵表示可逆 (简单推论)

线性变换的多项式的矩阵表示是矩阵表示的多项式 (线性映射情况的简单推论)

线性变换不同基下矩阵相似 (线性映射情况的简单推论)

* 从而矩阵的相似不变量可以定义为线性变换的量。

相似的矩阵可以看作同一个线性变换的矩阵表示，即习题 11 (由此相似与取基完全等价)
有限维化零多项式的存在性、最小多项式的思路，即例 7 (之后将完善此题结论)

重要例题：

- 矩阵表示计算 (例 16、习题 10，尤其注意例 16 中利用坐标的计算方法)
- 利用矩阵表示进行证明 (例 11、15、习题 6)

9.4 线性变换的特征值、特征向量、特征多项式、代数重数、几何重数、可对角化定义

不同特征值的特征向量线性无关，即习题 2 (矩阵情况简单推论)

A 的多项式的特征值是特征值的多项式 (矩阵情况简单推论)

可对角化等价条件，即定理 1，推论 1、2、3，命题 1 (5.6 节结论的简单推论)

* 仍然注意 \mathbb{K} 上的线性变换要考虑特征值不在 \mathbb{K} 中。

特征多项式基本性质，即习题 11 (矩阵情况的简单推论)

重要例题：

- 整体性的特征系统说明 (例 5、6、7，习题 8)
- 可对角化判定 (例 8，习题 7、14，习题 7 注意矩阵论技巧的使用)

9.5 不变子空间、诱导商空间映射定义

不变子空间分解上限制映射的矩阵表示为分块对角阵 (简单推论)

不变子空间与其补空间上分解的矩阵表示，即例 8 (无需证明)

Jordan 块、特征值互不相同时的所有不变子空间，即例 10、12 (无需证明，一般情况见本讲义 21.2.2)

A 的多项式的核空间分解，即定理 2、3 (8.2 节结论的简单推论)

Hamilton-Caylay 定理 (无需证明，即使要证也建议用相似三角化证明，即例 22)

根子空间定义与根子空间分解 (简单推论)

$AX = XB$ 只有零解当且仅当 A, B 无公共特征值，即习题 10、11 (解决一般可交换性问题的引理)

* 利用下节知识，这也等价于特征多项式/最小多项式互素。

可交换矩阵可以同时相似三角化，即例 5 (无需证明，注意此时 $A + B$ 特征值为 A, B 特征值求和)

重要例题：

- 线性变换的限制技巧 (例 2、4、19)
- 不变子空间操作 (例 15)
- 多项式相关操作 (例 21、23、24，习题 8，注意例 21 证明过程已经不再需要矩阵论)

9.6 最小多项式定义与唯一性 (简单推论)

零化多项式当且仅当是最小多项式倍式 (简单推论)

最小多项式的根与特征多项式相同，且不随扩域改变，相似不改变 (无需证明)

分块对角阵的最小多项式是每个对角块最小多项式的最小公倍式 (简单推论)

不变子空间分解后的最小多项式是每个不变子空间上最小多项式的最小公倍式 (矩阵情况简单推论)

Jordan 块最小多项式为特征多项式

* 上述结论已经足以得到 **Jordan 标准形**的最小多项式。

可对角化当且仅当最小多项式能分解为一次因式乘积

可对角化映射的任何不变子空间上的限制可对角化 (简单推论)

可对角化映射的任何不变子空间存在补空间是不变子空间 (了解即可)

根子空间分解中每个空间对应的次数可写为最小多项式中次数 (简单推论)

根子空间维数等于代数重数，即例 19 (无需证明)

$\mathbb{K}[A]$ 维数与基，即例 10

不变子空间可以分解到根子空间上，即例 22 (了解即可，解决一般求所有不变子空间问题的引理)

重要例题：

- 最小多项式、根子空间计算 (例 6、7、14, 习题 12)
- 最小多项式与可对角化 (例 16、习题 2、11)

9.7 强循环子空间定义

强循环子空间分解与幂零变换 Jordan 形，即定理 1、2 (无需证明)

幂零变换 Jordan 形利用 $\text{rank } A^k$ 的算法

重要例题：

- 幂零变换的整体操作 (习题 2、10)
- 幂零阵 Jordan 标准形应用 (例 9、12、13, 习题 7)
- Jordan 形可交换矩阵计算 (例 14、16, 本质上都是直接硬算)

9.8 C 上 Jordan 标准形定义与利用 $\text{rank}(\lambda I - A)^k$ 的算法

Jordan 标准形与特征值代数重数、几何重数、最小多项式 (简单推论)

最小多项式等于特征多项式等价于不同 Jordan 块特征值不同 (简单推论)

* 根据之前的结论可发现，此时不变子空间个数有限，且 $\mathcal{C}(A) = \mathbb{K}[A]$ 。

A 与 A^T 相似，即例 5 (无需证明)

Jordan 基定义与算法，即例 2

分块三角阵 Jordan 标准形，即例 18 (无需证明)

Jordan 块次方的 Jordan 标准形，即例 19、21，习题 12

重要例题：

- Jordan 标准形计算 (习题 3、4、8)
- Jordan 标准形应用 (例 24, 习题 6、7)

9.9 一般数域上的根子空间分解，即 9.6 节例 17、例 20 (无需证明)

循环子空间定义

有理标准形的定义与利用秩的算法

* 没考过，但在范围内，所以会一下。用秩和下面介绍的用初等因子算会一个就行。

另一种有理标准形，广义 Jordan 标准形，即例 3 (了解即可)

22.1.5 多项式矩阵

往年从未考过必须会多项式矩阵才能做的题，但对不少结论，多项式矩阵是很好的方法。此外，虽然多项式矩阵 (即 λ -矩阵) 可能不考，多项式的基本性质与算法是必须掌握的。

7.1 多项式与加法、数乘、乘法、次数定义

次数的基本性质 (简单推论)

7.2 整除、带余除法定义与算法

整除与带余除法的基本性质 (简单推论)

λ -矩阵与相抵定义

λ -矩阵的相抵标准形算法

7.3 最大公因式定义、基本性质与算法

裴蜀定理，即定理 1 (无需证明，几乎是多项式这块最重要的结论)

互素定义与基本性质 (简单推论)

λ -矩阵的不变因子、行列式因子定义

- λ -矩阵的不变因子随扩域不变 (基本推论)
- λ -矩阵的不变因子与行列式因子关系 (无需证明)
- λ -矩阵相抵与不变因子相同、行列式因子相同等价 (简单推论)

7.4 不可约多项式定义

- 唯一因子分解定理 (无需证明)

7.5 根的重数定义

7.6 复多项式的唯一因子分解定理 (无需证明)

9.8 λ -矩阵的初等因子定义

- 特征方阵 $\lambda I - A$ 与其初等因子定义
- 特征方阵相抵等价于初等因子相同 (无需证明)
- 矩阵相似等价于对应的特征方阵相抵 (无需证明)
- * 此结论很重要, 如证明 A 与 A^T 相似时这是最好的方法。
- 矩阵相似随扩域不变 (不变因子随扩域不变的基本推论)
- 用初等因子计算 Jordan 标准形的算法
- 最后一个不变因子是矩阵的最小多项式 (无需证明)
- * 这可以证明最小多项式随扩域不变。

9.9 用初等因子计算有理标准形的算法

§22.2 复习题

除 5(d) 与 9 外都是从教材/往年题中选取、改编的。

1. 设 V_1, V_2 是 V 的子空间, 且和为 V , 若 W 是 $V_1 \cap V_2$ 对 V_1 的补空间, 证明它也是 V_2 对 V 的补空间, 并对逆命题举出反例。
2. 对 \mathbb{K} 上的有限维线性空间 U, V 之间的线性映射 f, g , 若 $\text{Ker } f = \text{Ker } g$, 证明存在 V 上的可逆线性变换 h 使得 $f = h \circ g$ 。给出此结论的矩阵版本。
3. 若 \mathcal{A} 是线性空间 V 上的线性变换:
 - (a) 证明 $V = \text{Im } \mathcal{A} + \text{Ker } \mathcal{A}$ 当且仅当 $\mathcal{A}|_{\text{Im } \mathcal{A}}$ 为满射;
 - (b) 证明 $V = \text{Im } \mathcal{A} \oplus \text{Ker } \mathcal{A}$ 当且仅当 $\mathcal{A}|_{\text{Im } \mathcal{A}}$ 为双射;
 - (c) 若 V 维数有限, 证明 $V = \text{Im } \mathcal{A} \oplus \text{Ker } \mathcal{A}$ 当且仅当 $\mathcal{A}|_{\text{Im } \mathcal{A}}$ 为单射。
4. 定义 $\mathbb{K}^{n \times n}$ 上的线性变换 $\mathcal{A}(X) = AX$, 其中 $A \in \mathbb{K}^{n \times n}$.
 - (a) 给出 $\text{Im } \mathcal{A}$ 、 $\text{Ker } \mathcal{A}$ 的维数与一组基;
 - (b) 证明 \mathcal{A} 可对角化当且仅当 A 可对角化。
5. 定义 $\mathbb{C}^{n \times n}$ 上的线性变换 $\mathcal{A}(X) = AX - XA$, 其中 $A \in \mathbb{C}^{n \times n}$.
 - (a) 若 A 可对角化, 证明 \mathcal{A} 可对角化;
 - (b) 若 $B \in \text{Ker } \mathcal{A}$, 证明 A, B 有公共特征向量;
 - (c) 证明 $\text{Ker } \mathcal{A} = \langle I, A, \dots, A^{n-1} \rangle$ 当且仅当 A 的特征多项式等于最小多项式;
 - (d) 对一般的 A , 计算 $\text{Ker } \mathcal{A}$ 的维数与一组基, 并给出 $\text{Im } \mathcal{A}$ 的维数。

6. 已知复方阵 A 的特征多项式为 $f(x) = \prod_{i=1}^k (x - \lambda_i)^{m_i}$, 最小多项式为 $p(x) = \prod_{i=1}^k (x - \lambda_i)^{n_i}$, 其中 $\lambda_1, \dots, \lambda_k$ 互不相同。
- 证明 A 的 Jordan 标准形能通过 $f(x)$ 与 $p(x)$ 确定当且仅当 n_i 为 1 或 $m_i - 1$ 或 m_i 对每个 n_i 成立;
 - 写出 A 对应的根子空间分解，并求 A 的多项式 $f(A)$ 使得它是到第 i 个根子空间的投影变换，且使得其他根子空间变为 0;
 - 证明 $\deg p \leq \operatorname{rank} A + 1$;
 - 证明存在 α 使得 $\alpha, A\alpha, \dots, A^{\deg p-1}\alpha$ 线性无关;
 - 若每个 $n_i = m_i$, 计算 A -不变子空间的个数;
 - 对复多项式 h , 证明 $h(A)$ 可逆当且仅当 $(h, f) = (h, p) = 1$, 并证明此时存在复多项式 g 使得 $h(A)^{-1} = g(A)$ 。

7. 设 \mathbb{R}^4 中向量

$$\alpha_1 = (3, 4, 0, 1)^T, \quad \alpha_2 = (2, 5, 0, 1)^T, \quad \alpha_3 = (2, 4, 1, 1)^T, \quad \alpha_4 = (6, 7, 1, 2)^T$$

设 \mathbb{R}^4 上的线性映射 \mathcal{A} 满足

$$\mathcal{A}(\alpha_1) = \alpha_2, \quad \operatorname{Im} \mathcal{A} = \langle \alpha_1, \alpha_2 \rangle, \quad \operatorname{Ker} \mathcal{A} = \langle \alpha_3, \alpha_4 \rangle$$

证明存在一组基使得 \mathcal{A} 在这组基下的矩阵表示为 Jordan 标准形的形式，并求此标准形。

8. 对 \mathbb{K} 上的 n 维线性空间 V , 已知 V 上的非零线性变换 \mathcal{A} 满足 $\mathcal{A}^m = O$ 、 $\mathcal{A}^{m-1} \neq O$, m 为正整数。
- 证明 $\operatorname{Im} \mathcal{A}^{m-1} \subset \operatorname{Ker} \mathcal{A} \cap \operatorname{Im} \mathcal{A}$;
 - 给出 $\operatorname{Im} \mathcal{A}^{m-1} \neq \operatorname{Ker} \mathcal{A} \cap \operatorname{Im} \mathcal{A}$ 的例子;
 - 当 $n = m + 1$ 时, 证明 $\operatorname{Im} \mathcal{A}^{m-1} = \operatorname{Ker} \mathcal{A} \cap \operatorname{Im} \mathcal{A}$ 。

9. 对方程组 $Ax = b$, 其中 A 为 \mathbb{K} 上的 $m \times n$ 阶多项式矩阵, b 为 \mathbb{K} 上的 m 维多项式列向量, 证明存在 \mathbb{K} 上的 n 维多项式列向量 x 为解的充分必要条件是 A 与 (A, b) 的所有不变因子对应相等 (定义所有超过阶数的不变因子都为 0)。

§22.3 解答

1. * 这类题目的通用思路: 先找一些例子, 取基感受结论大致是怎么回事。

也即已知 $V_1 + V_2 = V$ 、 $W \oplus (V_1 \cap V_2) = V_1$, 求证 $W \oplus V_2 = V$ 。

- $W \cap V_2 = \{0\}$

若有 $w \in W \cap V_2$, 由于 $W \oplus (V_1 \cap V_2) = V_1$ 可知 $W \subset V_1$, 于是 $W \cap V_2 \subset V_1 \cap V_2$, 即 w 既在 W 中也在 $V_1 \cap V_2$ 中, 由直和可知 $w = 0$, 从而得证。

另证: 由 W 与 $V_1 \cap V_2$ 为直和可知 $W \cap (V_1 \cap V_2) = \{0\}$, 利用交集运算可以结合, 由 $W \subset V_1$ 可知

$$\{0\} = W \cap (V_1 \cap V_2) = (W \cap V_1) \cap V_2 = W \cap V_2$$

* 这两个做法分别是从结论与条件入手的。

- $W + V_2 = V$

对任何 $v \in V$, 存在 $v_1 \in V_1$ 、 $v_2 \in V_2$ 使得 $v = v_1 + v_2$; 对任何 $v_1 \in V_1$, 存在 $w \in W$ 、 $v_{12} \in V_1 \cap V_2$ 使得 $v_1 = w + v_{12}$, 从而

$$v = v_1 + v_2 = w + v_{12} + v_2$$

由于 $v_{12}, v_2 \in V_2$ 可知 $v_{12} + v_2 \in V_2$, 这就得到了 $V \subset W + V_2$, 而 W 、 V_2 都为 V 的子空间, 于是 $W + V_2 = V$ 。

* 不要忘记子空间的和还是子空间。

- 逆命题反例

取 $V = \mathbb{R}^2$, $V_1 = \langle (1, 0)^T \rangle$ 、 $V_2 = \langle (0, 1)^T \rangle$ 、 $W = \langle (1, 1)^T \rangle$, 可验证 $V_1 + V_2 = V$ 、 $W \oplus V_2 = V$, 但 W 不包含在 V_1 中, 自然不可能是 $V_1 \cap V_2$ 对 V_1 的补空间。

* 记住经典反例三个两两互补的空间。

* 本题如果出成有限维版本其实也可以硬算维数 (但不会有人想算吧……)

2. 记 $n = \dim U$, $m = \dim V$ 。我们把证明分成三个部分, U 上的基扩充、 V 上的基扩充与映射性质验证:

- 设 $\text{Ker } f = \text{Ker } g$ 的一组基为 k_1, \dots, k_{n-r} , 将它们扩充 u_1, \dots, u_r 成为 U 的一组基, 先证明 $f(u_1), \dots, f(u_r)$ 线性无关。

若否, 存在不全为 0 的 $\lambda_1, \dots, \lambda_r$ 使得 $\sum_{i=1}^r \lambda_i f(u_i) = 0$, 则 $f(\sum_{i=1}^r \lambda_i u_i) = 0$, 于是 $\sum_{i=1}^r \lambda_i u_i \in \text{Ker } f$, 但由于 $\langle u_1, \dots, u_r \rangle$ 构成 $\text{Ker } f$ 的补空间, 它们的交只有 0, 从而 $\sum_{i=1}^r \lambda_i u_i = 0$, 再由线性无关性即得所有 λ_i 全为 0。

同理, $g(u_1), \dots, g(u_r)$ 线性无关。

* 这部分的结论非常经典, 是线性映射相关问题常用的。证明叙述方式有不止一种, 这里采取了补空间的方法, 也可以直接从基线性无关推导。

- 将 $f(u_1), \dots, f(u_r)$ 扩充 v_1, \dots, v_{m-r} 成为 V 的一组基, $g(u_1), \dots, g(u_r)$ 扩充 v'_1, \dots, v'_{m-r} 成为 V 的一组基, 这样即得到了 V 的两组不同的基。

- 定义 $h(f(u_i)) = g(u_i)$ 对 $i = 1, \dots, r$ 成立, 且 $h(v_j) = v'_j$ 对 $j = 1, \dots, m-r$ 成立。由于这指定了一组基的像, 线性映射 h 已经确定; 由于 h 将一组基映射到了一组基, h 必然是同构。最后, 定义保证了对所有 u_i 有 $h(f(u_i)) = g(u_i)$, 而对所有 k_j 由定义有

$$h(f(k_j)) = h(0) = 0 = g(k_j)$$

从而 $h \circ f$ 与 g 在一组基上的像相同, 且它们都是 $\text{Hom}(U, V)$ 中的线性映射, 从而相等。

* 有限维线性映射最重要的操作是取适当基与基扩充。只要有了基映射, 线性映射就容易构造了。

矩阵版本: 考虑 U 的一组基 S , V 的一组基 T (和我们证明中构造的基可以完全不相干), 设 f 在 S 、 T 下的矩阵表示为 F , g 在 S 、 T 下的矩阵表示为 G , h 在 T 下的矩阵表示为 H 。

条件可以化为 $\text{Ker } F = \text{Ker } G$, 而利用同构的矩阵表示可逆与线性映射复合对应矩阵乘积, 结论即存在可逆阵 H 使得 $F = HG$ 。

* 本题直接用矩阵论证明是并不简单的, 因为未必能构造出可逆的 H 。空间取基虽然有些技巧性, 但思路是明确的。

3. * 这类抽象证明题的思路来源: 若要证明和相关则从配凑符合要求的部分出发, 证明交相关则往往需要利用反证推导。只要思路清晰, 具体的过程细节总是可以推理出的。

(a) 利用限制映射的定义, $\text{Im } \mathcal{A}|_{\text{Im } \mathcal{A}} = \mathcal{A}(\text{Im } \mathcal{A})$, 于是其为满射等价于 $\mathcal{A}(\text{Im } \mathcal{A}) = \text{Im } \mathcal{A}$ 。

左推右: 若 $V = \text{Im } \mathcal{A} + \text{Ker } \mathcal{A}$, 对任何 $v \in V$ 存在 $x \in \text{Im } \mathcal{A}$ 、 $y \in \text{Ker } \mathcal{A}$ 使得 $v = x + y$, 从而 $\mathcal{A}(v) = \mathcal{A}(x + y) = \mathcal{A}(x) + \mathcal{A}(y) = \mathcal{A}(x)$ (利用 $\text{Ker } \mathcal{A}$ 定义)。当 v 取任何元素时, $\mathcal{A}(v)$ 可为任何 $\text{Im } \mathcal{A}$ 中元素, 而右侧由于 $x \in \text{Im } \mathcal{A}$, 有 $\mathcal{A}(x) \in \mathcal{A}(\text{Im } \mathcal{A})$, 从而得到 $\text{Im } \mathcal{A} \subset \mathcal{A}(\text{Im } \mathcal{A})$ 。又由定义可知 $\mathcal{A}(\text{Im } \mathcal{A}) \subset \mathcal{A}(V) = \text{Im } \mathcal{A}$, 于是等式成立。

右推左: 若 $\mathcal{A}(\text{Im } \mathcal{A}) = \text{Im } \mathcal{A}$, 对任何 $v \in V$ 有 $\mathcal{A}(v) \in \text{Im } \mathcal{A}$, 从而由条件存在 $x \in \text{Im } \mathcal{A}$ 使得 $\mathcal{A}(x) = \mathcal{A}(v)$ 。进一步地, 有 $\mathcal{A}(v - x) = \mathcal{A}(v) - \mathcal{A}(x) = 0$, 于是 $v - x \in \text{Ker } \mathcal{A}$, 从而 $v = x + (v - x) \in \text{Ker } \mathcal{A} + \text{Im } \mathcal{A}$, 这就说明了 $V \subset \text{Ker } \mathcal{A} + \text{Im } \mathcal{A}$, 再由右侧都为 V 子空间可知相等。

(b) 利用限制映射的定义, $\text{Ker } \mathcal{A}|_{\text{Im } \mathcal{A}} = \text{Ker } \mathcal{A} \cap \text{Im } \mathcal{A}$ (任何 $\mathcal{A}|_{\text{Im } \mathcal{A}}(x) = 0$ 的 x 至少需要 $\mathcal{A}(x) = 0$, 且 x 在限制映射定义域 $\text{Im } \mathcal{A}$ 中)。由此, $\mathcal{A}|_{\text{Im } \mathcal{A}}$ 是单射等价于 $\text{Ker } \mathcal{A}|_{\text{Im } \mathcal{A}} = \{0\}$, 这又等价于 $\text{Ker } \mathcal{A} \cap \text{Im } \mathcal{A} = \{0\}$ 。

结合 (a) 的结论, $\mathcal{A}|_{\text{Im } \mathcal{A}}$ 是双射等价于它是单射、满射, 即等价于 $\text{Ker } \mathcal{A} \cap \text{Im } \mathcal{A} = \{0\}$ 、 $\text{Ker } \mathcal{A} + \text{Im } \mathcal{A} = V$, 与 $\text{Ker } \mathcal{A} \oplus \text{Im } \mathcal{A} = V$ 等价。

(c) 对于有限维线性空间上的线性变换, 其单射、满射、双射相互等价。

* 此结论一般可以直接使用, 证明思路: 若其为单射, 将基映射到线性无关向量组, 而个数等于维数的线性无关向量组必然为基, 从而其将一组基映射到一组基, 是双射; 若其为满射, 将基映射到能表出全空间的向量组, 而个数等于维数的能表出全空间的向量组必然为基, 从而其将一组基映射到一组基, 是双射; 反之, 双射自然可以推出单射与满射。

由此, $\mathcal{A}|_{\text{Im } \mathcal{A}}$ 为单射等价于它为双射, 根据 (b) 得结论成立。

* 需要熟悉限制映射像与核的表达 $\text{Im } f|_{U_0 \rightarrow V_0} = f(U_0)$ 、 $\text{Ker } f|_{U_0 \rightarrow V_0} = \text{Ker } f \cap U_0$, 这样往往可以将限制映射条件转化为原映射条件。

4. (a) 这类题目的最通用方法是考虑矩阵表示: 考虑 $\mathbb{K}^{n \times n}$ 的一组基

$$S = (E_{11}, \dots, E_{n1}, E_{12}, \dots, E_{n2}, \dots, E_{1n}, \dots, E_{nn})$$

这里 E_{ij} 为第 i 行第 j 列位置为 1, 其他为 0 的 n 阶方阵。

计算可发现 $\mathcal{A}(E_{ij})$ 得到的矩阵是将 A 的第 i 列放在第 j 列, 其他列为 0 的方阵, 从而直接计算可得

$$\mathcal{A}(S) = S \text{diag}(A, A, \dots, A)$$

于是 \mathcal{A} 在 S 下的矩阵表示为 $\text{diag}(A, A, \dots, A)$, 也即对角块共有 n 个, 每个都是 A 的分块对角阵。

对任何一个矩阵 X , 设其坐标为 x , 将 x 自上而下分为 n 个 n 阶列向量 $x^{(1)}, x^{(2)}, \dots, x^{(n)}$, 计算可发现

$$\text{diag}(A, \dots, A) \begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(n)} \end{pmatrix} = \begin{pmatrix} Ax^{(1)} \\ \vdots \\ Ax^{(n)} \end{pmatrix}$$

由此, $X \in \text{Ker } \mathcal{A}$ 当且仅当每个 $x^{(i)} \in \text{Ker } A$, $X \in \text{Im } \mathcal{A}$ 当且仅当每个 $x^{(i)} \in \text{Im } A$ 。进一步观察可发现 $x^{(i)}$ 即为 X 的第 i 列。

* 事实上直接计算 $A(x^{(1)}, \dots, x^{(n)}) = (Ax^{(1)}, \dots, Ax^{(n)})$ 的效果是完全相同的, 只是矩阵表示方法更加通用。

为了构造一组基, 我们设 $\alpha_1, \dots, \alpha_r$ 为 $\text{Im } A$ 一组基, $\beta_1, \dots, \beta_{n-r}$ 为 $\text{Ker } A$ 一组基 (由解空间维数定理可确定个数)。

设 X_{ij}^I 为第 i 列取 α_j , 其他列取 0 的矩阵, 这里 $i = 1, \dots, n$ 、 $j = 1, \dots, r$ 。这些矩阵线性无关: $\sum_{i=1}^n \sum_{j=1}^r \lambda_{ij} X_{ij}^I$ 的第 i 列为 $\sum_{j=1}^r \lambda_{ij} \alpha_j$, 由此若线性组合为 O , 当且仅当每列都为 0, 即可由 α_j 之间的线性无关性推出所有系数为 0。此外, 上方计算中已经可以说明 X_{ij}^I 的每一列都可取为 $\text{Im } \mathcal{A}$ 中任何元素, 因此可表出全空间。由此, X_{ij}^I 构成 $\text{Im } \mathcal{A}$ 的一组基, 其维数为 nr , 根据矩阵论结论可知 $r = \text{rank } A$, 因此 $\dim \text{Im } \mathcal{A} = n \text{rank } A$ 。

同理, 设 X_{ij}^K 为第 i 列取 β_j , 其他列取 0 的矩阵, 这里 $i = 1, \dots, n$ 、 $j = 1, \dots, n-r$, 可验证它们构成 $\text{Ker } \mathcal{A}$ 一组基, 从而维数为 $n(n - \text{rank } A)$ 。

* 这里为一组基的构造看起来有些复杂, 本质上是由于这样构造后非零分量位置不重合时可以独立组合, 本题也即对每一列组合。

- (b) 本题当然可以从特征值与特征子空间出发证明, 但我们这里要介绍一种更好的做法: 直接从最小多项式说明。

直接计算可发现 $\mathcal{A}^2(X) = A(AX) = A^2X$, 同理 $\mathcal{A}^n(X) = A^nX$, 进一步计算得到 $f(\mathcal{A})(X) = f(A)X$ 对任何多项式 $f \in \mathbb{K}[x]$ 成立。

由此, $f(\mathcal{A}) = O$ 当且仅当 $f(A)X = O$ 对任何 X 成立, 取 $X = I$ 即得这当且仅当 $f(A) = O$ ($f(A) = O$ 推 $f(A)X = O$ 是直接的)。

由此 $f(\mathcal{A}) = O$ 等价于 $f(A) = O$, 二者的化零多项式集合相同, 由定义最小多项式也相同。可对角化当且仅当最小多项式无重根且能在 \mathbb{K} 上分解为一次因式, 这对 A 与 \mathcal{A} 同时成立或不成立, 从而得证。

* 当线性变换的多项式可以直接计算时, 用最小多项式说明可对角化往往是最简单的办法。

5. (a) * 本题直接计算 \mathcal{A} 的多项式会发现没有明显规律, 矩阵表示形式相对复杂, 特征子空间也难以确定, 因此只能回到找一组特征向量作为全空间基的原始定义。定义方法与特征子空间方法的差别在于, 此时无需考虑对应的特征值是多少。

考虑特征向量满足的方程, 即存在 $\lambda \in \mathbb{C}$ 使得

$$AX - XA = \lambda X$$

由条件, 设 $A = P^{-1}DP$, $D = \text{diag}(\lambda_1, \dots, \lambda_n)$, 代入方程可得

$$P^{-1}DPX - X P^{-1}DP = \lambda X$$

同时左乘 P 右乘 P^{-1} 得到

$$DPXP^{-1} - PXP^{-1}D = \lambda PXP^{-1}$$

设 $Y = PXP^{-1}$, 方程化为 $DY - YD = \lambda Y$, 直接计算并对比分量可知这当且仅当对每个 i, j 满足

$$(\lambda_i - \lambda_j - \lambda)y_{ij} = 0$$

只要某个 $y_{ij} = 1$, 其他 y_{ij} 全为 0, $\lambda = \lambda_i - \lambda_j$, 此方程即能满足, 于是所有 E_{ij} 都可以满足此方程, 这就得到所有 $P^{-1}E_{ij}P$ 都是 \mathcal{A} 的特征值。

由于 $Y \rightarrow P^{-1}YP$ 是线性映射, 且计算验证可知它与 $X \rightarrow PXP^{-1}$ 互逆, 从而是同构, 这就能说明它将 $\mathbb{C}^{n \times n}$ 的一组基映射到一组基, 因此所有 $P^{-1}E_{ij}P$ 是 $\mathbb{C}^{n \times n}$ 一组基, 得证可对角化。

* 这里证明同构利用了直接构造逆映射的技巧, 因为双射等价于存在逆映射。

- (b) * 本题的核心思路是将 A, B 有公共特征向量转化为 B 有包含在 A 的某个特征子空间上的特征向量。我们将复方阵看作复线性变换。由于复线性空间上的线性变换一定有特征值与特征向量, 只要 B 能限制在某个特征子空间上即可。

由条件可知 $AB = BA$ 。

设 A 的某个特征子空间为 $U = \text{Ker}(\lambda I - A)$, 对任何 $x \in U$, 有

$$(\lambda I - A)Bx = \lambda Bx - ABx = \lambda Bx - BAx = B(\lambda I - A)x = 0$$

从而 $Bx \in U$ 。考虑复线性变换 $\mathcal{B}: x \rightarrow Bx$, 则 $\mathcal{B}|_U$ 必然有特征值 μ 与属于它的特征向量, 设其为 y , 则 $y \in U$, 且可发现

$$Ay = \lambda y, \quad By = \mathcal{B}(y) = \mathcal{B}|_U(y) = \mu y$$

从而 y 即为符合要求的公共特征向量。

* 这里最重要的过程是将 B 看作线性变换后进行限制以得到子空间上的特征向量作为原本的特征向量。

(c) 我们将给出一个基本可以直接使用的结论: 特征多项式等于最小多项式的方阵可以相似为友方阵, 反之亦然。

证明:

我们用特征方阵进行考虑。若 n 阶方阵 A 的特征多项式等于最小多项式, 其初等因子组中每个不可约因式只能出现一次, 从而可发现 $\lambda I - A$ 不变因子组只能为 $n - 1$ 个 1 与特征多项式 $\varphi_A(\lambda)$ 。

另一方面, 考虑 $\varphi_A(\lambda)$ 的友方阵 (形式与性质可见本讲义 17.2.2 开头), 其特征方阵不变因子组也为 $n - 1$ 个 1 与 $\varphi_A(\lambda)$, 利用相似等价于特征方阵相抵即得到此友方阵与 A 相似。反之, 由于 f 的友方阵的不变因子组为 $n - 1$ 个 1 与 $f(\lambda)$, 其分解的初等因子组中每个不可约因式只出现一次, 从而也满足特征多项式等于最小多项式。

* 若利用结论最小多项式就是最后一个不变因子, 可以更快说明。

- 右推左

根据矩阵的多项式可交换性, $\langle I, A, \dots, A^{n-1} \rangle \subset \text{Ker } \mathcal{A}$, 于是只需证明另一边包含关系。

设 $A = P^{-1}FP$, 其中

$$F = \begin{pmatrix} & & -a_{n-1} \\ 1 & & -a_{n-2} \\ & \ddots & \vdots \\ & & 1 & -a_0 \end{pmatrix}$$

为 A 的特征多项式的友方阵。

与 (a) 类似, $X \in \text{Ker } \mathcal{A}$ 等价于 $AX = XA$, 记 $Y = PXP^{-1}$ 可得 $YF = FY$ 。

计算可发现 F 有性质

$$Fe_i = e_{i+1}, \quad i = 1, \dots, n-1$$

于是有

$$e_i = F^{i-1}e_1, \quad i = 1, \dots, n$$

由此, 全空间一组基即为 $e_1, Fe_1, \dots, F^{n-1}e_1$ 。由于 Ye_1 是空间中向量, 存在多项式 $g \in \mathbb{C}[x]$ 使得

$$Ye_1 = g(F)e_1$$

利用可交换性有

$$YFe_1 = FYe_1 = Fg(F)e_1 = g(F)Fe_1$$

同理归纳可得 $YF^k e_1 = g(F)F^k e_1$ 对任何自然数 k 成立, 从而 Y 与 $g(F)$ 看作 $x \rightarrow Yx$ 、 $x \rightarrow g(F)x$ 的线性变换在一组基下的像相同, 从而得到两个变换相等, 即得 $Y = g(F)$ 。

利用相似不改变多项式直接计算即得

$$X = P^{-1}YP = P^{-1}g(F)P = g(P^{-1}FP) = g(A)$$

利用 H-C 定理, A 的任何 n 次以上的多项式对特征多项式利用带余除法都可以化为不超过 $n - 1$ 次的多项式, 于是 $\langle I, A, \dots, A^{n-1} \rangle$ 可以表示 A 的所有多项式, 从而 $X \in \langle I, A, \dots, A^{n-1} \rangle$, 这就得到了证明。

* 称最小多项式等于特征多项式的线性变换为循环变换, 能相似为友方阵从而得到全空间一组基是它的核心性质。我们上面的一些推导对更一般的循环变换也成立。下一道复习题中, 我们还将更多考虑其性质。

- 左推右

类似上方的情况, 设 A 的 Jordan 标准形为 J , 且 $A = P^{-1}JP$, 只需找到 $YJ = JY$ 且 $Y \neq g(J)$ 对任何多项式 $g \in \mathbb{C}[x]$ 成立, 取 $X = P^{-1}YP$ 即可发现 X 不是 A 的多项式, 从而矛盾。

若 A 的每个特征值只有一个 Jordan 块, 利用 **Jordan 标准形的最小多项式结论**可知 A 的特征多项式等于最小多项式, 于是 A 必然存在特征值 λ 有 $J_a(\lambda), J_b(\lambda)$ 两个 Jordan 块, 我们写出它的 Jordan 标准形(注意 Jordan 块可任意排列)为

$$\text{diag}(J_a(\lambda), J_b(\lambda), \dots, J_{n_k}(\lambda_k))$$

计算可以发现, 分块对角阵的多项式相当于每个对角块变为多项式, 由此, 只要找到这些对角块以外有元素且与 J 可交换的 Y , 即推出了结论。

由于特殊性在左上角地两个块, 我们设 Y 与 J 相同分块, 且(未写出元素均为 0)

$$Y = \begin{pmatrix} O & Y_0 \\ & \ddots \\ & & O \end{pmatrix}$$

也即只有第一行第二列的分块为 $Y_0 \in \mathbb{C}^{a \times b}$, 其他为 O 。直接计算可知 $JY = YJ$ 等价于

$$J_a(\lambda)Y_0 = Y_0J_b(\lambda)$$

两侧同减 λY_0 , 可以进一步化简为

$$J_a(0)Y_0 = (J_a(\lambda) - \lambda I_a)Y_0 = J_a(\lambda)Y_0 - \lambda Y_0 = Y_0J_b(\lambda) - \lambda Y_0 = Y_0(J_b(\lambda) - \lambda I_b) = Y_0J_b(0)$$

由此只需构造非零的 Y_0 使得 $J_a(0)Y_0 = J_b(0)$, 直接计算可发现 $Y_0 = E_{1b}$ 则两侧都为 O , 符合要求, 结论成立。

* 复线性空间的循环变换等价于 **Jordan 标准形**每个特征值 **Jordan 块**只有一个。

(d) 在 (c) 中已经说明了, 设 A 的 Jordan 标准形为 J , 且 $A = P^{-1}JP$, 只需找到 $YJ = JY$ 的所有 Y , 令 $X = P^{-1}YP$ 即得到 $AX = XA$ 的所有 X 。此外, 与 (a) 类似可证明, 只要找到了符合要求的 Y 的一组基, $P^{-1}YP$ 即可得到 X 的一组基。

对一般的 Jordan 标准形 $J = \text{diag}(J_{n_1}(\lambda_1), \dots, J_{n_k}(\lambda_k))$ (这里 λ_i 未必互不相同), 受到 (c) 的启发, 我们将 Y 相同分块为

$$Y = \begin{pmatrix} Y_{11} & \dots & Y_{1k} \\ \vdots & \ddots & \vdots \\ Y_{k1} & \dots & Y_{kk} \end{pmatrix}$$

直接分块计算可发现 $YJ = JY$ 当且仅当

$$J_{n_i}(\lambda_i)Y_{ij} = Y_{ij}J_{n_j}(\lambda_j)$$

- 若 $\lambda_i \neq \lambda_j$, 我们下面说明无非零解。

首先, 类似 (c) 两侧同减 λ_i 得到

$$J_{n_i}(0)Y_{ij} = Y_{ij}J_{n_j}(\lambda_j - \lambda_i)$$

直接计算有

$$J_{n_i}^2(0)Y_{ij} = J_{n_i}(0)Y_{ij}J_{n_j}(\lambda_j - \lambda_i) = Y_{ij}J_{n_j}^2(\lambda_j - \lambda_i)$$

同理归纳得

$$J_{n_i}^k(0)Y_{ij} = Y_{ij}J_{n_j}^k(\lambda_j - \lambda_i)$$

对任何自然数 k 成立。取 $k = n_i$, 则计算得左侧为 O , 右侧 $J_{n_j}^{n_i}(\lambda_j - \lambda_i)$ 为对角元 $(\lambda_j - \lambda_i)^{n_i}$ 非零的上三角阵, 必然可逆, 因此只能 $Y_{ij} = O$ 。

* 这个结论的推广是, 只要复方阵 A 、 B (未必同阶) 无公共特征值, $AX = XB$ 都只有零解, 见本讲义 21.1 习题解答。

- 若 $\lambda_i = \lambda_j$, 与 (c) 相同可得这转化为

$$J_{n_i}(0)Y_{ij} = Y_{ij}J_{n_j}(0)$$

设 Y_{ij} 的各个分量为 y_{st} , 直接计算可发现左侧相当于把 Y_{ij} 所有元素上移一行, 然后最下一行变为 0; 右侧相当于把所有元素右移一列, 然后最左一列变成 0, 从而得到的方程事实上是, 对所有 $s = 1, \dots, n_i$ 、 $t = 1, \dots, n_j$ 、且 $s+k = 1, \dots, n_i$ 、 $t+k = 1, \dots, n_j$ 的 s 、 k 、 t 有

$$y_{st} = y_{s+k, t+k}$$

也即每条从左上到右下的对角线上元素相同, 且第一列除了 y_{11} 、最后一行除了 $y_{n_i n_j}$ 外全为 0。

* 上面的过程虽然写起来比较复杂, 计算一下会发现是非常直观的。

首先, 由于 $i-j$ 可以从 $1-n_j$ 取到 n_i-1 , 共有 n_i+n_j-1 条对角线。第一列除了 y_{11} 、最后一行除了 $y_{n_i n_j}$ 外的元素一共包含 $\max(n_i, n_j) - 1$ 条对角线 (可以自己画一下确定), 于是作差得到还剩下 $\min(n_i, n_j)$ 条对角线可任取。

由此, Y_{ij} 的解空间维数为 $\min(n_i, n_j)$, 对应的一组基为 $i-j = n_i-1$ 一直到 $i-j = n_i-\min(n_i, n_j)$ (从右上角往下数 $\min(n_i, n_j)$ 条) 的每条对角线取 1, 其他对角线取 0 的矩阵。

* 同样, 计算一下会发现结果非常直观, 我们举两个简单的例子: $J_4(0)Y = YJ_3(0)$ 的解空间一组基为

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$J_2(0)Y = YJ_3(0)$ 的解空间一组基为

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

最后, 所有的 Y 相当于把每个符合要求的 Y_{ij} 拼在一起, 由此所有 Y 的一组基为: 对每个 Y_{ij} , 在 Y_{ij} 位置取 Y_{ij} 的一组基, 其他位置取 O , 并将所有这些方阵拼在一起 (能这么做是类似复习题第三题构造基的思路, Y_{ij} 的位置互相不重叠), 总维数为所有 Y_{ij} 维数和, 即

$$\sum_{i=1}^k \sum_{j=1}^k \delta_{\lambda_i \lambda_j} \min(n_i, n_j)$$

这里 $\delta_{\lambda_i \lambda_j}$ 表示一个当 $\lambda_i = \lambda_j$ 时取 1，否则取 0 的量，也即加法事实上只对 $\lambda_i = \lambda_j$ 时进行。由上方论述，对每个基取 $P^{-1}YP$ 即可以从所有 Y 的一组基得到所有 X 的一组基，且由第一同构定理 $\text{Im } \mathcal{A}$ 的维数是

$$n - \sum_{i=1}^k \sum_{j=1}^k \delta_{\lambda_i \lambda_j} \min(n_i, n_j)$$

* 若设特征值 λ_i 的 Jordan 块为 $J_{n_{i1}}(\lambda_i), \dots, J_{n_{it_i}}(\lambda_i)$ ，且 $n_{i1} \leq \dots \leq n_{it_i}$ ，特征值 $\lambda_1, \dots, \lambda_s$ 互不相同，则考虑每个 n_{ij} 计算次数可发现求和还可以化为（注意前一个求和中 $n_i = n_j$ 的项）

$$\sum_{i=1}^s \sum_{j=1}^{t_i} (2(t_i - j) + 1)n_{ij}$$

* 对这个相对复杂的大问题的解决其实可以代表用标准形分析相似不变问题的方式。利用标准形后，往往可以将复杂的问题化为直接的计算。

* 期中考试最重要的技巧是，遇到不会分析的复方阵问题转化为 **Jordan 标准形** 考虑，一般数域上的问题只要对扩域不变或 **Jordan 标准形** 存在，仍然可以转化为复方阵 **Jordan 标准形** 问题。

6. 本题中我们进一步观察怎么将复杂的问题化为直接的计算，以及在这之上作进一步的处理。

(a) 确定 **Jordan 标准形** 只需确定每个特征值对应的 **Jordan 块阶数**，于是 **Jordan 标准形** 唯一当且仅当每个特征值对应的 **Jordan 块阶数都能够确定**。根据 **Jordan 标准形最小多项式的结论**， m_i 表示每个特征值对应的 **Jordan 块总阶数**， n_i 表示此特征值对应的最大的 **Jordan 块阶数**（且由 H-C 定理 $n_i \leq m_i$ ）。

- 若 $n_i = 1$ ，由于最大 **Jordan 块** 只有一阶，此特征值有 m_i 个一阶 **Jordan 块**。
- 若 $m_i = n_i$ ，由于最大 **Jordan 块** 即为 $n_i = m_i$ 阶，此特征值只有一个 m_i 阶 **Jordan 块**。
- 若 $m_i = n_i - 1$ ，由于最大 **Jordan 块** 为 $m_i - 1$ 阶，剩下只有一阶了，此特征值有一个 $m_i - 1$ 阶 **Jordan 块** 和一个一阶 **Jordan 块**。
- 若 $n_i > 1$ 且 $n_i < m_i - 1$ ，假设 $m_i = kn_i + r$ ，其中 $r < n_i$ 、 $k \geq 1$ 。若 $k = 1$ ，则由条件 $r \geq 2$ ，可能出现一个 n_i 阶、一个 r 阶与一个 n_i 阶、 r 个 1 阶，根据条件这两种情况不同；若 $k > 1$ ，可能出现 k 个 n_i 阶、一个 r 阶与 $k - 1$ 个 n_i 阶、一个 r 阶、 n_i 个一阶，根据条件这两种情况不同。

综合得证。

* 本题的结论值得熟悉。当所有 $n_i = m_i$ 时即为 **循环变换**，当所有 $n_i = 1$ 时即为 **可对角化变换**。这两种特殊的线性变换的各种性质是常用的。

(b) 由条件可知根子空间分解可用最小多项式写为

$$\mathbb{C}^n = \text{Ker}(A - \lambda_1 I)^{n_1} \oplus \text{Ker}(A - \lambda_2 I)^{n_2} \cdots \oplus \text{Ker}(A - \lambda_k I)^{n_k}$$

我们证明如下结论：

- $f(A)$ 满足对任何 $v \in \text{Ker}(A - \lambda_i I)^{n_i}$ 有 $f(A)v = 0$ 当且仅当 $(x - \lambda_i)^{n_i} \mid f(x)$ 。
首先，由 $\text{Ker}(A - \lambda_i I)^{n_i}$ 的定义，只要 $(x - \lambda_i)^{n_i} \mid f(x)$ ，设 $f(x) = (x - \lambda_i)^{n_i} g(x)$ 即有

$$f(A)v = (A - \lambda_i I)^{n_i} g(A)v = g(A)(A - \lambda_i I)^{n_i} v = A0 = 0$$

反之，若存在 $(x - \lambda_i)^{n_i} \nmid f(x)$ 使得 $f(A)v$ 恒为 0，由定义可知

$$(g(A)f(A) + h(A)(A - \lambda_i I)^{n_i})v = 0$$

对任何多项式 g, h 恒成立，从而根据裴蜀定理可知能使得左侧为 $\gcd(f, (x - \lambda_i)^{n_i})(A)$ 。由于 $(x - \lambda_i)^{n_i} \nmid f$ ，只能存在比 n_i 次数更低的 p_i 使得 $(A - \lambda_i I)^{p_i} v$ 恒为 0，也即存在 $p_i < n_i$ 使得

$$\text{Ker}(A - \lambda_i I)^{n_i} \subset \text{Ker}(A - \lambda_i I)^{p_i}$$

此时类似另一边证明直接计算可发现 $\prod_{j \neq i} (x - \lambda_j)^{n_j} (x - \lambda_i)^{p_i}$ 也是 A 的非零化零多项式，但其次数比最小多项式更低，矛盾。

* 这也是最小多项式的常见性质： λ_i 对应的根子空间至少需要 $A - \lambda_i I$ 的 n_i 次方才能表示出。

- $f(A)$ 满足对任何 $v \in \text{Ker}(A - \lambda_i I)^{n_i}$ 有 $f(A)v = v$ 当且仅当 $(x - \lambda_i)^{n_i} \mid f(x) - 1$ 。

由条件 $f(x) - 1$ 满足对任何 $v \in \text{Ker}(A - \lambda_i I)^{n_i}$ 有 $(f(A) - 1)v = 0$ ，由上一种情况成立。

下面进行构造，关键如下：只要 $f(A)$ 满足对任何 $v \in \text{Ker}(A - \lambda_i I)^{n_i}$ 有 $f(A)v = v$ ，且对任何 $v \in \text{Ker}(A - \lambda_j I)^{n_j}, j \neq i$ 有 $f(A)v = 0$ ，它就符合要求。证明同样直接：利用直和的性质，任何 v 都可以唯一分解为 $v_1 + \dots + v_k$ ，从而有

$$f(A)(v_1 + \dots + v_k) = f(A)v_1 + \dots + f(A)v_k = f(A)v_i = v_i$$

因此这就是符合要求的投影。

* 这里利用了映射随空间分解的特点，本质是由于对任何多项式 g ， $\text{Ker } g(A)$ 都是 $f(A)$ 的不变子空间。

根据之前证明的结论，只需

$$(x - \lambda_1)^{n_1} \mid f(x), \dots, (x - \lambda_{i-1})^{n_{i-1}} \mid f(x), (x - \lambda_{i+1})^{n_{i+1}} \mid f(x), \dots, (x - \lambda_k)^{n_k} \mid f(x)$$

且 $(x - \lambda_i)^{n_i} \mid f(x) - 1$ 即可。

由于所有 λ_i 互不相同，第一行其实可以推出

$$\prod_{j \neq i} (x - \lambda_j)^{n_j} \mid f(x)$$

从而设

$$f(x) = p(x) \prod_{j \neq i} (x - \lambda_j)^{n_j}$$

由第二行可设

$$f(x) = q(x)(x - \lambda_i)^{n_i} + 1$$

根据裴蜀定理，由于 $\prod_{j \neq i} (x - \lambda_j)^{n_j}$ 与 $-(x - \lambda_i)^{n_i}$ 互素，一定存在 $p(x)$ 、 $q(x)$ 使得

$$p(x) \prod_{j \neq i} (x - \lambda_j)^{n_j} - q(x)(x - \lambda_i)^{n_i} = 1$$

移项就得到了符合要求的 f 。

* 考试范围内多项式相关的技巧往往归结于裴蜀定理。

- (c) 利用相似不改变几何重数，利用 Jordan 标准形直接计算可发现特征值的几何重数等于它的 Jordan 块个数。

由此，利用解空间维数定理，右侧 $\text{rank } A + 1 = n - \dim \text{Ker } A + 1$ 。设 $s = \dim \text{Ker } A$ 为 0 的 Jordan 块个数，则右侧为 $n - (s - 1)$ 。另一方面，左侧为每个特征值最大 Jordan 块的阶数，可以看成每个特征值只保留一个最大 Jordan 块、去掉其他 Jordan 块后的剩余阶数。由于 0 有 s 个 Jordan 块，至少去掉了 $s - 1$ 个，而去掉的每个又至少一阶，因此至少去掉了 $s - 1$ 阶，即得左侧不超过 $n - (s - 1)$ ，得证。

* 由于看作不同数域不影响最小多项式与秩，此结论对任何数域上的方阵都成立。

* 由此，最小多项式等于特征多项式当且仅当所有特征值几何重数为 1。一定要熟悉这些 Jordan 标准形相关的计算性质。

(d) * 此结论非常经典，有不同方法说明，我们这里结合 Jordan 标准形与循环变换性质证明，希望大家掌握这个方法。

由最小多项式的每个特征值对应次数等于最大 Jordan 块阶数，设 $Q^{-1}AQ = J$, A 的 Jordan 标准形 J 一定可以写为

$$\text{diag}(J_{n_1}(\lambda_1), \dots, J_{n_k}(\lambda_k), J_0)$$

这里 J_0 为一些其他 Jordan 块。由于左上角的 $\text{diag}(J_{n_1}(\lambda_1), \dots, J_{n_k}(\lambda_k))$ 特征多项式等于最小多项式，存在可逆阵 P 将它相似为友方阵，即 (注意此时两侧都为 $\deg p$ 阶方阵)

$$P^{-1} \text{diag}(J_{n_1}(\lambda_1), \dots, J_{n_k}(\lambda_k)) P = \begin{pmatrix} & & -a_{n-1} \\ 1 & & -a_{n-2} \\ & \ddots & \vdots \\ & & 1 & -a_0 \end{pmatrix}$$

因此进一步计算得

$$\begin{pmatrix} P^{-1} & \\ & I \end{pmatrix} Q^{-1} A Q \begin{pmatrix} P & \\ & I \end{pmatrix} = \left(\begin{pmatrix} & & -a_{n-1} \\ 1 & & -a_{n-2} \\ & \ddots & \vdots \\ & & 1 & -a_0 \end{pmatrix} \quad J_0 \right)$$

记 $R = Q \text{diag}(P, I)$, 并将右侧方阵记为 B , 即得 $R^{-1}AR = B$ 。

直接计算可知 $Be_1 = e_2$ 、 $Be_2 = e_3$, 直到 $Be_{\deg p-1} = e_{\deg p}$, 从而 $e_1, Be_1, \dots, B^{\deg p-1}e_1$ 线性无关。记 $\alpha = R^{-1}e_1$, 利用相似性质有

$$A^k \alpha = R^{-1}(RA^kR^{-1})R\alpha = R^{-1}(RAR^{-1})^k R\alpha = R^{-1}B^k e_1$$

由于 R^{-1} 是可逆矩阵，可验证 R^{-1} 乘线性无关向量组后仍然线性无关 (可逆矩阵视为同构，同构将一组基映射为一组基，从而满足)，由此即得 $\alpha, A\alpha, \dots, A^{\deg p-1}\alpha$ 线性无关，得证。

* 这个证明方法的核心思路是从 A 中构造出一个子方阵使得其为循环变换，这样就可以用友方阵直接构造对应的向量 (友方阵最重要的性质就是可以用它的次方乘 e_1 生成全空间一组基)。

(e) 由于所有 $n_i = m_i$, 由 (a) 中讨论可设

$$A = P^{-1}JP, \quad J = \text{diag}(J_{n_1}(\lambda_1), \dots, J_{n_k}(\lambda_k))$$

其中 $\lambda_1, \dots, \lambda_k$ 互不相同。证明分为五个部分：

- $J_n(0)$ -不变子空间计算

对 $J_n(0)$ 的不变子空间 U , 我们下面证明必须

$$U = \langle e_1, \dots, e_i \rangle$$

这里 $i = 0, 1, \dots, n$, i 取 0 即 $U = \{0\}$ 。

证明：

若 $U \neq \{0\}$, 对任何非零向量 $x \in U$, 都存在 i 使得 $x_i \neq 0$ 、 $x_{i+1} = \dots = x_n = 0$ 。取 x 使得对应的 i 最大, 直接计算可发现, 若 $x = (x_1, \dots, x_i, 0, \dots, 0)^T$, 有

$$J_n(0)x = (x_2, \dots, x_i, 0, 0, \dots, 0)^T$$

$$J_n(0)^2x = (x_3, \dots, x_i, 0, 0, 0, \dots, 0)^T$$

直到 $J_n(0)^{i-1}x = (x_i, 0, \dots, 0)^T$ 。由于 U 包含 x , 利用不变子空间定义可知上述所有向量都在 U 中, 且考虑这些列向量拼成的矩阵, 其前 i 行构成的子矩阵是一个对角线均为 x_i (非零) 的三角阵, 从而可逆, 因此它们线性无关。由于它们只有前 i 个分量非零, 可知它们生成的子空间为

$$\langle e_1, \dots, e_i \rangle$$

* 也可以直接用它们列变换出 e_1, \dots, e_i 以说明等价。

而由于子空间的封闭性, $\langle e_1, \dots, e_i \rangle \subset U$ 。

另一方面, 由于 x 使得 i 最大, U 中没有后 $n-i$ 个分量非零的元素, 于是

$$U \subset \langle e_1, \dots, e_i \rangle$$

综合两方面包含得证。

此外, 当 $U = \langle e_1, \dots, e_i \rangle$ 时, 利用 $J_n(0)e_1 = 0$, $J_n(0)e_i = e_{i-1}, i > 1$, 可通过一组基的像验证其为不变子空间, 从而得证。

- $J_n(\lambda)$ -不变子空间计算

我们下面证明, 它的不变子空间仍为

$$U = \langle e_1, \dots, e_i \rangle$$

这里 $i = 0, 1, \dots, n$, i 取 0 即 $U = \{0\}$ 。

证明:

对 $J_n(\lambda)$ -不变子空间 U , 若 $u \in U$, 由于 $J_n(\lambda)u \in U$, 利用空间封闭性可知

$$J_n(\lambda) - \lambda u \in U$$

而这即说明 $J_n(0)u \in U$, 由于这对任何 $u \in U$ 成立, U 也是 $J_n(0)$ -不变子空间。

完全同理, 对 $J_n(0)$ -不变子空间 U , 若 $u \in U$, 由于 $J_n(0)u \in U$, 利用空间封闭性可知

$$J_n(0) + \lambda u \in U$$

而这即说明 $J_n(\lambda)u \in U$, 由于这对任何 $u \in U$ 成立, U 也是 $J_n(\lambda)$ -不变子空间。

于是, $J_n(\lambda)$ 与 $J_n(0)$ 的不变子空间完全相同, 即得证。

- 根子空间分解引理

对于 J -不变子空间的计算关键在于如下引理: 设 \mathbb{C} 上线性空间 V 上的线性变换 A 的根子空间分解为 (这里假设 n_i 为 A 的最小多项式中 λ_i 次数)

$$V = V_1 \oplus V_2 \oplus \dots \oplus V_k, \quad V_k = \text{Ker}(A - \lambda_i I)^{n_i}$$

则对 A -不变子空间 U , 有

$$U = (U \cap V_1) \oplus (U \cap V_2) \oplus \dots \oplus (U \cap V_k)$$

* 直接验证可知不变子空间的交还是不变子空间, 又由于根子空间是不变子空间, 可得 $U \cap V_1, \dots, U \cap V_k$, 都是 A -不变子空间。

证明:

设 \mathcal{A} 的最小多项式为 p , 由于 $p(\mathcal{A}) = \mathcal{O}$, 利用限制映射定义可知

$$p(\mathcal{A}|_U) = \mathcal{O}$$

于是, \mathcal{A} 在 U 上的最小多项式一定是 $p(x)$ 的因式, 设其为

$$p_U(x) = \prod_{i=1}^k (x - \lambda_i)^{s_i}$$

且 $s_i \leq m_i$ 。由此可对 U 进行根子空间分解

$$U = \text{Ker}(\mathcal{A}|_U - \lambda_1 \mathcal{I}|_U)^{s_1} \oplus \text{Ker}(\mathcal{A}|_U - \lambda_2 \mathcal{I}|_U)^{s_2} \oplus \cdots \oplus \text{Ker}(\mathcal{A}|_U - \lambda_k \mathcal{I}|_U)^{s_k}$$

利用限制映射与映射的指数(复合)的定义可知

$$(\mathcal{A}|_U)^k = \mathcal{A}^k|_U$$

从而再作加法、数乘可知

$$\text{Ker}(\mathcal{A}|_U - \lambda_i \mathcal{I}|_U)^{s_i} = \text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{s_i}|_U$$

再由限制映射的像与核结论可知其为

$$\text{Ker}(\mathcal{A} - \lambda_i \mathcal{I})^{s_i} \cap U$$

于是其包含在 $U \cap V_i$ 中(由 $s_i \leq n_i$ 可知第一项包含在 V_i 中)。

另一方面, 由于 V_1 到 V_k 为直和, 利用直和等价于每个与其他所有求和交为 $\{0\}$ 可知 $U \cap V_1$ 到 $U \cap V_k$ 也是直和, 而它们为 U 的子空间, 因此直和包含在 U 中, 由此利用每个的包含关系得到得到

$$U = \text{Ker}(\mathcal{A}|_U - \lambda_1 \mathcal{I}|_U)^{s_1} \oplus \cdots \oplus \text{Ker}(\mathcal{A}|_U - \lambda_k \mathcal{I}|_U)^{s_k} \subset U \cap V_1 \oplus \cdots \oplus U \cap V_k \subset U$$

由于左侧等于右侧, 每个包含关系只能取等, 即得结论。

* 此结论对一般的空间分解或非不变子空间的 U 未必成立, 是根子空间的本质性质。

- J -不变子空间计算

设

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$$

这里 x_i 为 n_i 阶列向量。直接计算可发现

$$Jx = \begin{pmatrix} J_{n_1}(\lambda_1)x_1 \\ J_{n_2}(\lambda_2)x_2 \\ \vdots \\ J_{n_k}(\lambda_k)x_k \end{pmatrix}$$

由此直接计算可发现 $\text{Ker}(J - \lambda_i I)^{n_i}$ 为 x_i 可任取, 其他分量为 0 的列向量。对任何一个 J -不变子空间 U , $U_i = U \cap \text{Ker}(J - \lambda_i I)^{n_i}$ 也即 U 中除 x_i 外全为 0 的向量。

由于 U_i 是 J 不变子空间, 假设其中所有 x_i 集合为 U'_i (由于其他分量为 0, U_i 即为 U'_i 中所有向量增添零分量), 利用上方的计算结果可知 U'_i 是一个 $J_{n_i}(\lambda_i)$ -不变子空间。根据已证, 必须 $U'_i = \langle e_1, \dots, e_j \rangle$, $j = 0, 1, \dots, n_i$ 。

综合以上，由于 J -不变子空间 U 一定为

$$U_1 \oplus U_2 \oplus \cdots \oplus U_k$$

而每个 U_i 完全由 U'_i 确定，因此有 $n_i + 1$ 种取法，故总取法数为

$$\prod_{i=1}^k (n_i + 1)$$

* 一些说明细节：由不变子空间定义可验证不变子空间的和是不变子空间，从而构造出每个 U_i 后一定能得到不变子空间 U ，且 U_i 不同时对应的非零分量位置不同，得到的 U 不同。

- A -不变子空间计算

我们最后证明 A -不变子空间与 J -不变子空间个数相同。

设 U 为 J 的不变子空间，记 $P^{-1}U = \{P^{-1}u \mid u \in U\}$ ，可直接验证其为子空间。由于 $A = P^{-1}JP$ ，对任何 $u \in U$ ， $Ju \in U$ 可以推出对任何 $P^{-1}u \in P^{-1}U$ 有

$$P^{-1}JP(P^{-1}u) = P^{-1}(Ju) \in P^{-1}U$$

于是 $P^{-1}U$ 是 A -不变子空间。

完全同理，若 V 是 A -不变子空间，则 PV 是 J -不变子空间，因此 A -不变子空间与 J -不变子空间一一对应，即得到个数相同，从而最终结论为

$$\prod_{i=1}^k (n_i + 1) = \prod_{i=1}^k (m_i + 1)$$

* 这里第三部分的引理是核心结论，不变子空间在根子空间上的分解，再根据根子空间的具体情况可确定所有不变子空间。的确考过求所有不变子空间的题目，因此至少这题结论值得记忆。

* 若某个 $n_i < m_i$ ，利用几何重数等于 Jordan 块个数可知 λ_i 的几何重数至少为 2。设其两个线性无关的特征向量为 α_1, α_2 ，则对任何 $k \in \mathbb{C}$ ， $\alpha_1 + k\alpha_2$ 都是 λ_i 的特征向量，于是 $\langle \alpha_1 + k\alpha_2 \rangle$ 是 A 的不变子空间，可验证这无穷多个不变子空间彼此不同。由此，本题的题干即为不变子空间个数有限的充要条件。

- (f) 利用裴蜀定理， $(h, f) = 1$ 可推出存在多项式 u, v 使得 $uh + vf = 1$ ，代入可得 $u(A)h(A) + v(A)f(A) = I$ ，由 H-C 定理可知 $f(A) = O$ ，从而 $u(A)h(A) = I$ ，这就说明了 h 可逆且存在 A 的多项式为 h 的逆。

若 $(h, f) = d \neq 1$ ，可知 d 的次数至少为 1，从而可取出其某一次因式 $x - \lambda$ ，由于 $d \mid f$ 可知 λ 为 A 的特征值，从而 $\text{rank}(A - \lambda I) < n$ ，其不可逆，利用唯一因子分解定理可知 $h(A)$ 可以分解为包含 $A - \lambda I$ 的一系列矩阵的乘积，因此不可逆。

* 这里的做法是较为形式化的，更直接的计算做法是，由于相似不改变多项式结论，可以不妨设 A 已经是 Jordan 标准形了，这时由于 A 是对角元为特征值 λ_i 的上三角阵，判断 $h(A)$ 可逆性只需计算对角元，而 $h(A)$ 的对角元即为所有 $h(\lambda_i)$ ，从而均非零等价于任何 λ_i 都不是 h 的因式，也能得证。

7. * 这类具体计算的问题一定要先分析所给向量的结构。计算可以发现 $\alpha_1, \alpha_2, \alpha_3$ 线性无关，且

$$\alpha_4 = 2\alpha_1 - \alpha_2 + \alpha_3$$

由此，设 $\alpha_1, \alpha_2, \alpha_3, \beta$ 构成全空间一组基，条件可以化为基映射的条件，即 $\text{Im } \mathcal{A} = \langle \alpha_1, \alpha_2 \rangle$ 能推出

$$\mathcal{A}(\alpha_1), \mathcal{A}(\alpha_2), \mathcal{A}(\alpha_3), \mathcal{A}(\beta) \in \langle \alpha_1, \alpha_2 \rangle$$

且 $\text{rank}(\mathcal{A}(\alpha_1), \mathcal{A}(\alpha_2), \mathcal{A}(\alpha_3), \mathcal{A}(\beta)) = 2$ (否则 $\text{Im } \mathcal{A}$ 维数至多 1)。

$\text{Ker } \mathcal{A} = \langle \alpha_3, \alpha_4 \rangle$ 能推出

$$\mathcal{A}(\alpha_3) = 0, \quad \mathcal{A}(2\alpha_1 - \alpha_2 + \alpha_3) = 0$$

* 事实上上述已经条件与原条件等价了，这是因为 rank 确定了 $\text{Im } \mathcal{A}$ 与 $\text{Ker } \mathcal{A}$ 的维数。

回到原题的目标，由于已知 $\mathcal{A}(\alpha_1) = \alpha_2$ 、 $\mathcal{A}(\alpha_3) = 0$ ，我们还需要确定 $\mathcal{A}(\alpha_2)$ 和 $\mathcal{A}(\beta)$ 的情况。利用 $\text{Ker } \mathcal{A}$ 的第二个条件可知

$$2\mathcal{A}(\alpha_1) - \mathcal{A}(\alpha_2) + \mathcal{A}(\alpha_3) = 0$$

从而代入得到 $\mathcal{A}(\alpha_2) = 2\alpha_2$ 。对于 $\mathcal{A}(\beta)$ ，至少可知其能写成 $x\alpha_1 + y\beta_2$ (由 $\text{Im } \mathcal{A} = \langle \alpha_1, \alpha_2 \rangle$)，从而由此 \mathcal{A} 在基 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 下的矩阵表示为

$$\begin{pmatrix} 0 & 0 & 0 & x \\ 1 & 2 & 0 & y \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

直接计算可发现此矩阵特征多项式为 $\varphi(\lambda) = \lambda^3(\lambda - 2)$ ，从而有 3 重 0 特征值、1 重 2 特征值。

若 $x = 0$ ，则 $\text{Im } \mathcal{A} = \langle \alpha_2 \rangle$ ，矛盾，由此 $x \neq 0$ ，直接计算得此矩阵秩为 2 (事实上也可从 $\dim \text{Im } \mathcal{A} = 2$ 直接得到)，于是 0 的几何重数为 2，而 3 重 0 分为两个 Jordan 块 (直接计算秩可发现几何重数是 Jordan 块个数) 只能为一个一阶、一个二阶，从而 Jordan 标准形为 (未写出元素均为 0)

$$\begin{pmatrix} 0 & 1 & & \\ 0 & 0 & & \\ & & 0 & \\ & & & 2 \end{pmatrix}$$

8. (a) 由 \mathcal{A} 非零， $m > 1$ ，从而利用定义

$$\text{Im } \mathcal{A}^{m-1} = \mathcal{A}(\text{Im } \mathcal{A}^{m-2}) \subset \mathcal{A}(V) = \text{Im } \mathcal{A}$$

* 这类直接的式子不用写过多细节，如果对证明还有疑问，可以取元素进行详细讨论。

另一方面，对任何 $v \in \text{Im } \mathcal{A}^{m-1}$ ，设 $v = \mathcal{A}^{m-1}(u)$ ，则 $\mathcal{A}(v) = \mathcal{A}^m(u) = 0$ ，于是 $\text{Im } \mathcal{A}^{m-1} \subset \text{Ker } \mathcal{A}$ 。

综合两式得证。

(b) 反例构造为，设 $V = \mathbb{R}^5$ ， \mathcal{A} 满足

$$\mathcal{A}(e_1) = \mathcal{A}(e_2) = 0, \quad \mathcal{A}(e_3) = e_1, \quad \mathcal{A}(e_4) = e_2, \quad \mathcal{A}(e_5) = e_3$$

可验证 $m = 3$ ，且 $\text{Im } \mathcal{A}^2 = \langle e_1 \rangle$ 、 $\text{Ker } \mathcal{A} \cap \text{Im } \mathcal{A} = \langle e_1, e_2 \rangle$ 。

* 映射的反例往往需要从基映射去想，这里想完成真包含就需要 $\text{Ker } \mathcal{A} \cap \text{Im } \mathcal{A}$ 至少为二维，以此出发至少需要四个基，再添加一个可导出矛盾。

(c) 利用条件可说明 λ^m 是 \mathcal{A} 的化零多项式，而 λ^{m-1} 不是，利用化零多项式当且仅当是最小多项式的倍数可知 \mathcal{A} 的最小多项式为 $\lambda^m = \lambda^{n-1}$ 。由于此最小多项式在 \mathbb{K} 上可以分解为一次因式，应有 \mathcal{A} 可以化为 Jordan 标准形。类似复习题第六题 (a) 的讨论知其 Jordan 标准形只能为 $\text{diag}(J_{n-1}(0), 0)$ 。

设 \mathcal{A} 在基 $\alpha_1, \dots, \alpha_n$ 下的矩阵表示 Jordan 标准形，直接利用 Jordan 标准形计算可发现

$$\text{Im } \mathcal{A} = \langle \alpha_1, \dots, \alpha_{n-2} \rangle$$

$$\text{Ker } \mathcal{A} = \langle \alpha_1, \alpha_n \rangle$$

$$\text{Im } \mathcal{A}^{m-1} = \text{Im } \mathcal{A}^{n-2} = \langle \alpha_1 \rangle$$

从而直接计算得两侧都为 $\langle \alpha_1 \rangle$, 结论成立。

* 遇到莫名其妙的题目一定要尝试用标准形计算。

9. • 左推右

由 $Ax = b$ 可知

$$(A, b) = (A, Ax)$$

设 A 的每列为 $\alpha_1, \dots, \alpha_n$, 则 $Ax = x_1\alpha_1 + \dots + x_n\alpha_n$, 从而将 (A, Ax) 最后一列减去前面第 i 列的 x_i 倍 (这的确符合多项式矩阵初等变换的要求), 可得到

$$(A, 0)$$

而添加零列不影响任何非零子式, $(A, 0)$ 和 A 的各个行列式因子相同, 从而不变因子对应相等, 又由 $(A, 0)$ 与 (A, b) 作为多项式方阵相抵即得 (A, b) 与 $(A, 0)$, 它们的不变因子对应相等, 即得证。

• 右推左

设 A 的 Smith 标准形为

$$PAQ = \begin{pmatrix} \text{diag}(a_1, \dots, a_r) & O \\ O & O \end{pmatrix}$$

这里 P, Q 为可逆多项式方阵, a_i 为 A 的第 i 个不变因子。

由此可直接计算得到

$$P(A, b) \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} = \left(\begin{pmatrix} \text{diag}(a_1, \dots, a_r) & O \\ O & O \end{pmatrix}, Pb \right)$$

由于 $\text{diag}(Q, 1)$ 也是可逆多项式方阵 (可直接构造逆为 $\text{diag}(Q^{-1}, 1)$), 我们只需要证明, 若右侧矩阵的第 i 个不变因子是 a_i , $i > r$ 时不变因子为 0, 则原方程有解。我们设 $Pb = c$, 这是一个 m 阶列向量。

首先, 若 c_{r+1} 到 c_m 有非零元素, 考虑右侧矩阵的前 r 行 r 列的子式添加非零元素所在行列, 可发现对应子式非零, 矛盾。

其次, 我们可以归纳证明 $a_i|c_i$: 由行列式因子性质, $a_1 = \gcd(a_1, c_1)$, 从而 $a_1 | c_1$ 。接下来, 考虑前两行、第一列与最后一列构成的二阶子式, 其行列式为 $a_1 c_2$, 由行列式因子性质 $a_1 a_2 | a_1 c_2$, 由 a_1 非零 $a_2 | c_2$, 以此类推, 对 $i = 1, \dots, r$ 有 $a_i | c_i$ 。

最后, 我们来构造对应的 x 。设

$$x = Q \begin{pmatrix} c_1/a_1 \\ \vdots \\ c_r/a_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

由已经证明的整除性, 这确实是一个多项式列向量, 且有 (第三个等号是由于 c_{r+1} 到 c_m 均为 0)

$$Ax = P^{-1} \begin{pmatrix} \text{diag}(a_1, \dots, a_r) & O \\ O & O \end{pmatrix} \begin{pmatrix} c_1/a_1 \\ \vdots \\ c_r/a_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = P^{-1} \begin{pmatrix} c_1 \\ \vdots \\ c_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = P^{-1}c = P^{-1}Pb = b$$

从而得证。

* 应用到整数上完全类似, 这就完全解决了同余方程组的可解性问题。

二十三 有关证明

* 这两周作业只有一题, 见上章复习题 6(d)。

§23.1 期中考试

23.1.1 试题

1. 求矩阵的 Jordan 标准形:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2. 设复方阵

$$A = \begin{pmatrix} -2 & 1 & 0 \\ -4 & 2 & 0 \\ -2 & 1 & 0 \end{pmatrix}$$

问是否存在复方阵 B 使得 $A = B^2$, 若存在则给出构造, 不存在则给出证明。

3. 设 n 阶复方阵 A 的元素均为 1, 求最小多项式并说明它是否可对角化。

4. 设 ϕ, ψ 是实数域奇数维线性空间 V 上的线性变换, 且 $\psi\phi = \phi\psi$, 证明二者有公共特征向量。

5. 设 A 是数域 Π 上的 n 阶幂零矩阵, 这里 $n > 1$, 证明若 $A^n = O$ 、 $A^{n-1} \neq O$, 则不存在 Π 上的 n 阶矩阵 B 使得 $B^2 = A$ 。

6. 设 V 为次数不超过 n 的实系数多项式构成的线性空间。

(a) 证明 $1, (x+1), (x+1)^2, \dots, (x+1)^{n-1}$ 构成 V 一组基;

(b) 将求导看作 V 上的线性变换 D , 求 D 在 (a) 中基下的矩阵;

(c) 求 D 所有不变子空间。

7. 设 n 阶复方阵 A 有 n 个不同特征值 $\lambda_1, \dots, \lambda_n$, 并设 M_n 为所有 n 阶复方阵构成的线性空间。考虑 M_n 上的线性变换

$$L(B) = ABA^T$$

这里 A^T 为 A 的转置。

- (a) 设 $x, y \in \mathbb{C}^n$ 为 A 的特征值 λ_1, λ_2 的特征向量, 利用 x, y 构造 L 的一个特征向量;
(b) 求 L 的一个非零化零多项式。
8. 设 n 阶复方阵 A 特征值均为 1, 证明 A 与 A^k 相似, 这里 k 为正整数。

23.1.2 解答

1. 直接计算可得 A 的特征多项式 $(\lambda - 1)^4$, 从而特征值为 4 重 1, 且 $\text{rank}(A - I) = 3$, 从而 1 的几何重数为 1, Jordan 标准形只有一块, 得到 Jordan 标准形为

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

2. • Jordan 标准形计算

直接计算可发现 $\det(\lambda I - A) = \lambda^3$ 、 $\text{rank } A = 1$, 因此特征值 0 的代数重数为 3, 几何重数为 $3 - 1 = 2$, A 的 Jordan 标准形只能为

$$J = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

下面求过渡矩阵 P 使得 $P^{-1}AP = J$ 。由于直接计算可知

$$\text{Im } A = \langle (1, 2, 1)^T \rangle$$

$$\text{Ker } A = \langle (0, 0, 1)^T, (1, 2, 0)^T \rangle$$

我们取出一个有原像的 0 的特征向量 $(1, 2, 1)^T$, 并找到它的一个原像 $(0, 1, 0)^T$, 另一个特征向量直接有用 $(0, 0, 1)^T$ 即可, 从而得到可取

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

- 构造

我们先尝试构造 H 使得 $H^2 = J$ 。可发现

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

而 J 相当于右侧交换行列的结果 (事实上是交换 23 两列、23 两行, 看作初等变换阵相似应将左侧也交换 23 两列、23 两行), 因此对左侧也尝试交换行列可算出

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

由此可取左侧为 H , 而已知 $A = PJP^{-1}$ 、 $J = H^2$, 可发现取 $B = PHP^{-1}$ 即有 $B^2 = A$

- 结果计算

综合以上结果并计算可得

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

从而有

$$B = PHP^{-1} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -3 & 1 & 1 \end{pmatrix}$$

3. 给出两种做法:

- 利用上学期知识, 秩为 1 的矩阵可以分解为列向量乘行向量, 设 $\mathbf{1}$ 为全为 1 的列向量可得分解 $A = \mathbf{1}\mathbf{1}^T$, 从而 $A^2 = \mathbf{1}(\mathbf{1}^T\mathbf{1})\mathbf{1}^T = n\mathbf{1}\mathbf{1}^T = nA$, 而 $A - \lambda I$ 对任何 λ 都非零, 因此 $f(x) = x^2 - nx$ 就是 A 的最小多项式。由于其无重根, A 可对角化。
- 直接利用添行等方法计算行列式可发现 A 的特征多项式为 $\lambda^{n-1}(\lambda - n)$, 而 $\text{rank } A = 1$, 于是特征值 0 的几何重数为 $n - 1$ 等于代数重数, 而特征值 n 代数重数为 1, 几何重数必然等于代数重数, 由此 A 可对角化, 最小多项式为 $\lambda(\lambda - n)$ 。

4. * 从更基础出发的推导方法可见 9.6 节例 21。

这里不加证明地使用如下结论: 奇数维实线性空间上的线性变换一定存在实特征值与对应的实特征向量。实特征值的存在性可由奇数次实系数多项式存实根证明, 对应特征向量的存在性则由域上线性方程组的理论证明。

归纳。当 $\dim V = 1$ 时, 由于上述结论, V 的基一定为 ϕ, ψ 的公共特征向量。下面假设 $\dim V \leq 2k - 1$ 时命题成立, 考虑 $\dim V = 2k + 1$ 时。

我们先证明以下四个命题, 下方 λ 为任意实数。

- $\text{Ker}(\phi - \lambda\mathcal{I})$ 是 ϕ -不变子空间。

若 $x \in \text{Ker}(\phi - \lambda\mathcal{I})$, 由定义可知 $\phi(x) = \lambda x$, 由于子空间封闭性即得 $\phi(x) \in \text{Ker}(\phi - \lambda\mathcal{I})$ 。

- $\text{Ker}(\phi - \lambda\mathcal{I})$ 是 ψ -不变子空间。

若 $x \in \text{Ker}(\phi - \lambda\mathcal{I})$, 由定义可知 $\phi(x) = \lambda x$, 于是 $\phi\psi(x) = \psi\phi(x) = \lambda\psi(x)$, 从而 $\psi(x) \in \text{Ker}(\phi - \lambda\mathcal{I})$ 。

- $\text{Im}(\phi - \lambda\mathcal{I})$ 是 ϕ -不变子空间。

若 $x \in \text{Im}(\phi - \lambda\mathcal{I})$, 设 $(\phi - \lambda\mathcal{I})y = x$, 则利用 ϕ 多项式的可交换性

$$\phi(x) = \phi(\phi - \lambda\mathcal{I})y = (\phi - \lambda\mathcal{I})\phi(y)$$

从而 $\phi(x) \in \text{Im}(\phi - \lambda\mathcal{I})$ 。

- $\text{Im}(\phi - \lambda\mathcal{I})$ 是 ψ -不变子空间。

若 $x \in \text{Im}(\phi - \lambda\mathcal{I})$, 设 $(\phi - \lambda\mathcal{I})y = x$, 则利用 ψ 与 ϕ 可交换

$$\psi(x) = \psi(\phi - \lambda\mathcal{I})y = \phi\psi(y) - \lambda\psi(y) = (\phi - \lambda\mathcal{I})\psi(y)$$

从而 $\psi(x) \in \text{Im}(\phi - \lambda\mathcal{I})$ 。

若存在 $\lambda \in \mathbb{R}$ 使得 $\phi = \lambda I$, 所有向量都是 ϕ 的特征向量, 而 ψ 存在实特征向量, 从而已经得证。否则, 取 λ 为 ϕ 的实特征值, 根据解空间维数定理

$$\dim \text{Ker}(\phi - \lambda I) + \dim \text{Im}(\phi - \lambda I) = 2k + 1$$

且 $0 < \text{Ker}(\phi - \lambda I) < 2k + 1$ (第一个不等号由特征向量定义, 第二个不等号由 $\phi \neq \lambda I$)。由此也有 $\dim \text{Im}(\phi - \lambda I) < 2k + 1$ 。由于维数和为奇数, $\text{Ker}(\phi - \lambda I)$ 与 $\text{Im}(\phi - \lambda I)$ 中必有一个维数为奇数, 设其为 V_1 , 考虑 $\phi|_{V_1}$ 与 $\psi|_{V_1}$ 。

利用限制映射定义, $\phi|_{V_1}\psi|_{V_1} = \psi|_{V_1}\phi|_{V_1}$, 且 V_1 是维数小于 $2k + 1$ 的奇数维实线性空间, 归纳假设即得 $\phi|_{V_1}$ 与 $\psi|_{V_1}$ 存在公共特征向量。利用特征向量定义可发现限制映射的特征向量必然为原映射的特征向量, 这就找到了 ϕ 与 ψ 的公共特征向量。

5. 我们证明更强的结论: 不存在复方阵 B 使得 B^2 与 A 相似, 这样自然就不存在 Π 上的 B 使得 $B^2 = A$ 了。

* 下方推理其实无需 A 是复方阵, 因为幂零阵的 Jordan 标准形一定存在, 这里叙述复方阵是为了便于推广结论。

将 A 看作复方阵。由条件可知 A 的化零多项式集合包含 x^n 、不包含 x^{n-1} , 利用所有化零多项式都是最小多项式的倍式可知最小多项式为 $f(x) = x^n$, 从而可确定 A 的 Jordan 标准形为一个特征值 0 的 n 阶 Jordan 块 $J_n(0)$ 。

* 事实上利用第二题类似构造, 存在 B 使得 B^2 与 A 相似等价于存在 B 使得 $B^2 = A$ 。

若存在, 由于 $B^{2n} = O$, B 也是幂零方阵, 从而特征值只有 0, 进一步从 0 的重数非零得到 $\text{rank } B < n$ 。另一方面, 由于 $\text{rank } A = \text{rank } J_n(0) = n - 1$, 利用 $\text{rank } A = \text{rank } B^2 \leq \text{rank } B$ 可知 $\text{rank } B \geq n - 1$, 从而只能 $\text{rank } B = n - 1$, 于是 0 的几何重数为 1, 利用几何重数等于 Jordan 块个数可知 B 的 Jordan 标准形只能为 $J_n(0)$ 。

但是, 利用相似不改变秩, 且 B^2 与 $J_n^2(0)$ 相似, 直接计算可知此时 $\text{rank } B^2 = n - 2 < \text{rank } A = n - 1$, 从而 B^2 与 A 不可能相似, 矛盾。

* 更简单 (但不本质) 的做法: 由于 $B^{2n-2} \neq O$ 、 $B^{2n} = O$, B 的幂零指数为 $2n - 1$ 或 $2n$, 而幂零阵幂零指数不超过 n , 从而 $n > 1$ 时不可能成立

6. (a) 分两部分:

- 线性无关

考虑方程 $\lambda_1 + \lambda_2(x+1) + \cdots + \lambda_{n+1}(x+1)^n = 0$ 。若其成立, 其对任何 $x \in \mathbb{R}$ 成立, 从而对任何 $y = x+1$ 成立, 于是

$$\lambda_1 + \lambda_2y + \cdots + \lambda_{n+1}y^n = 0$$

当 x 取遍所有实数时 y 也取遍所有实数, 从而上式恒成立可推出所有系数为 0, 得证。

* 细致一点证明可以说上述至多 n 次的多项式的根超过 n 个, 所以只能为 0。

- 生成全空间

对任何多项式 $f(x) \in V$, 设

$$f(x) = \sum_{i=0}^n a_i x^i$$

将其写为 $\sum_{i=0}^n a_i((x+1)-1)^i$ 用二项式定理展开即得到其能被 $1, (x+1), \dots, (x+1)^n$ 表出, 从而能生成全空间。

* 证明上面两部分之一后说此空间维数 $n+1$ 所以 $n+1$ 个线性无关/生成全空间的向量构成基也算对。

(b) 直接计算可发现 $D(x+1)^k = k(x+1)^{k-1}$, 从而由定义矩阵表示为 (未写出的元素为 0)

$$\begin{pmatrix} 0 & 1 & & & \\ & 0 & 2 & & \\ & & 0 & \ddots & \\ & & & \ddots & n \\ & & & & 0 \end{pmatrix}$$

(c) * 其实与复习题 6(e) 第一步完全相同。

我们下面证明, D 的不变子空间 U 只能是 $\{0\}$ 或

$$\langle 1, x, \dots, x^i \rangle, \quad i = 0, 1, \dots, n$$

也即所有不超过 i 次的多项式构成的子空间。

若 $U \neq \{0\}$, 对任何非零多项式 $f(x) \in U$, 取 $f(x)$ 使得 $k = \deg f$ 最大, 直接计算可发现, $\deg D(f) = k-1$ 、 $\deg D^2(f) = k-2$, 直到 $\deg D^k(f) = 0$ 。由于 U 包含 f , 利用不变子空间定义可知上述所有向量都在 U 中, 且若

$$\sum_{j=0}^k \lambda_j D^j(f) = 0$$

假设 λ 不全为 0, 考虑下标最小的非零 λ_j , 可发现上式左侧应为 $k-j$ 次, 不为 0, 矛盾, 因此它们线性无关。又由于它们的次数都不超过 k , 且共 $k+1$ 个, 可知它们生成的子空间为 (所有不超过 k 次的多项式构成的子空间维数 $k+1$, 于是它们生成的必然为所有不超过 k 次的多项式)

$$\langle 1, x, \dots, x^k \rangle$$

而由于子空间 U 的封闭性, $\langle 1, x, \dots, x^k \rangle \subset U$ 。

另一方面, 由于 f 使得 $\deg f$ 最大, U 中没有大于 k 次的多项式, 于是

$$U \subset \langle 1, x, \dots, x^k \rangle$$

综合两方面包含得证。

此外, 当 $U = \langle 1, x, \dots, x^k \rangle$ 时, 由于不超过 k 次的多项式求导仍然不超过 k 次, 可验证的确为不变子空间, 综合即得这给出了全部 D -不变子空间。

7. (a) 设 $B = xy^T$, 可发现

$$L(xy^T) = Axxy^TA^T = Ax(Ay)^T = \lambda_1x(\lambda_2y)^T = \lambda_1\lambda_2xy^T$$

由于向量 x 、 y 均非零, 计算得 xy^T 不是零矩阵, 从而 xy^T 是 L 属于特征值 $\lambda_1\lambda_2$ 的特征向量。

(b) 以 (a) 作为提示, 可猜测 L 的所有特征值为 $\lambda_i\lambda_j$, $i, j = 1, \dots, n$, 于是猜测下面的 f 是非平凡零化多项式:

$$f(x) = \prod_{i=1}^n \prod_{j=1}^n (x - \lambda_i\lambda_j)^{n^2}$$

我们下面证明的确如此。

设

$$ABA^T = \lambda B$$

由特征值互不相同, A 可对角化, 设 $A = P^{-1}DP$, $D = \text{diag}(\lambda_1, \dots, \lambda_n)$, 则计算得

$$P^{-1}DPBP^TDP^{-T} = \lambda B$$

从而同时左乘 P 右乘 P^T 得到

$$DPBP^TD = \lambda PBP^T$$

设 $C = PBP^T$, 设此式成立, 可发现对任何 $i, j = 1, \dots, n$ 有

$$(\lambda_i \lambda_j - \lambda) c_{ij} = 0$$

由此, 若 λ 与一切 $\lambda_i \lambda_j$ 不同, 只能所有 c_{ij} 为 0, 从而已经得证 L 的特征值只能为所有 $\lambda_i \lambda_j$, 利用 H-C 定理将每个特征值作阶数次方一定是化零多项式。

* 事实上进一步可得所有 $P^{-1}E_{ij}P^{-T}$ 是 L 线性无关的特征向量 (这即是 (a) 的构造), 构成全空间一组基, 从而 L 可对角化, f 所有项次数为 1 即可。

8. • $J_m(1)$ 与 $J_m^k(1)$ 相似

直接归纳计算可知 $J_m(1)^k$ 是一个对角元素为 1、第 i 行第 $i+1$ 列 ($i = 1, \dots, m-1$) 元素为 k 的上三角方阵, 从而与第一题同理计算不变因子得 $J_m^k(1)$ 的相似标准形是 $J_m(1)$, 即存在可逆阵 P_m 使得

$$J_m(1) = P_m^{-1} J_m^k(1) P_m$$

- A 与 A^k 相似

由于 A 的特征值全为 1, 一定存在可逆阵 P 使得 $A = P^{-1}JP$, 其中

$$J = \text{diag}(J_{m_1}(1), J_{m_2}(1), \dots, J_{m_t}(1))$$

直接计算可知 $A^k = P^{-1}J^kP$, 且由已证

$$\begin{pmatrix} P_{m_1} & & \\ & \ddots & \\ & & P_{m_t} \end{pmatrix} \begin{pmatrix} J_{m_1}^k(1) & & \\ & \ddots & \\ & & J_{m_t}^k(1) \end{pmatrix} \begin{pmatrix} P_{m_1}^{-1} & & \\ & \ddots & \\ & & P_{m_t}^{-1} \end{pmatrix} = \begin{pmatrix} J_{m_1}(1) & & \\ & \ddots & \\ & & J_{m_t}(1) \end{pmatrix}$$

于是记 $Q = \text{diag}(P_{m_1}, \dots, P_{m_t})$ 即有 $J = QJ^kQ^{-1}$, 也即 $A \sim J \sim J^k \sim A^k$ (这里 \sim 表示相似), 利用相似是等价关系可知 A 与 A^k 相似。

§23.2 逻辑与证明

23.2.1 什么是证明

证明-看作序列

命题逻辑与公理

谓词逻辑

数学证明是代码?

顺序证明-搜索

分类与归纳/条件与循环

23.2.2 不完备性

最简单的代码语言-图灵完备
停机问题不可解
不可计算数
数学证明可以看作某种意义上的代码停机
哥德尔不完备性

23.2.3 形式化证明

纯函数-数学上没有储存
函数式语言
算法与形式化证明

二十四 补充：双线性函数简介

§24.1 双线性性

24.1.1 动机与定义

* 本部分所有内容都仅作介绍，但希望大家了解本节所述的认知新概念的方式。

从本章开始，我们将介绍下半学期的内容：双线性函数、实内积空间、复内积空间与张量初步。而事实上，接下来的每一件事都是以双线性函数为基础，因此必须先对这个概念做一点简单的引入。

在上学期我们已经学过， \mathbb{K}^n 上的二次型可以定义为

$$f(x) = x^T Ax$$

其中 $A \in \mathbb{K}^{n \times n}$ 为对称阵。有了这样的二次函数后，我们即可以讨论最值等问题。那么，既然学习了一般线性空间，一个自然的问题是，能否在一般线性空间上定义二次函数呢？有限维时，这个问题还是易于解决的，只要将其转化为坐标就可以类似上方进行定义。但是，这终究只是一个权宜之计。我们到底有没有办法给出更一般的定义呢？更进一步来说，通过什么手段能在一般线性空间定义一般的多项式？

在数学上，一个重要的技巧是，通过对已有工具的重复应用处理新工具的问题。就像解决一元多次线性常微分方程时可将其化为多元一次线性常微分方程组，我们将二次型看作如下的函数 φ

$$\varphi(x, y) = x^T A y$$

在 $x = y$ 时的值。可以验证，固定 x 时 $\varphi(x, y)$ 对 y 是线性的，固定 y 时 $\varphi(x, y)$ 对 x 是线性的，因此这样的 $\varphi(x, y)$ 称为双线性函数。

很明显，这样的定义可以推广到任何线性空间上：对数域 \mathbb{K} 上的线性空间 V, W ，定义 $V \times W \rightarrow \mathbb{K}$ 的映射 φ （即 φ 将一对 (v, w) 映射到 \mathbb{K} 中一个数），若其对两个分量分别线性，即对任何 $v_1, v_2, v \in V, w_1, w_2, w \in W, \lambda \in \mathbb{K}$ 有

$$\begin{aligned}\varphi(v_1 + v_2, w) &= \varphi(v_1, w) + \varphi(v_2, w), & \varphi(\lambda v, w) &= \lambda \varphi(v, w) \\ \varphi(v, w_1 + w_2) &= \varphi(v, w_1) + \varphi(v, w_2), & \varphi(v, \lambda w) &= \lambda \varphi(v, w)\end{aligned}$$

则称其为 V, W 上的双线性函数。

* 定义也可以等价写为，对任何 $v_1, v_2, v \in V, w_1, w_2, w \in W, \lambda, \mu \in \mathbb{K}$ 有

$$\varphi(\lambda v_1 + \mu v_2, w) = \lambda \varphi(v_1, w) + \mu \varphi(v_2, w)$$

$$\varphi(v, \lambda w_1 + \mu w_2) = \lambda\varphi(v, w_1) + \mu\varphi(v, w_2)$$

从原定义直接验证可得此定义成立，反之，分别取 $\lambda = \mu = 1, \mu = 0$ 可得原定义。

* 若 $V = W$ ，我们简称其为 V 上的双线性函数。

从双线性函数的定义，我们即可以得到二次函数的定义：对 $V \rightarrow \mathbb{K}$ 的函数 q ，若存在 V 上的双线性函数 φ 使得

$$\forall \alpha \in V, q(\alpha) = \varphi(\alpha, \alpha)$$

则称 q 是 V 上的二次型（二次型与二次函数的差别是没有一次、零次项）。

* 我们将在后面的部分验证这样的二次型定义在 \mathbb{K}^n 上与通常定义一致，从而这确实是二次型对一般线性空间的推广。

* 很明显，完全类似双线性函数，我们可以定义 V 上的三线性函数、四线性函数……且类似定义三次型、四次型，这就得到了多项式的推广。这部分内容将在介绍张量时作为补充。

最后，为了说明双线性函数的定义确实是新的，我们将说明非平凡双线性函数不是线性映射，无法通过之前的理论刻画。

对于 \mathbb{K} 上的线性空间 V, W ， $V \times W$ 可以自然定义成一个 \mathbb{K} 上的线性空间，其中的运算为：

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2)$$

$$\lambda(v, w) = (\lambda v, \lambda w)$$

验证其为线性空间的方式和验证向量空间为线性空间几乎完全相同，这称为 V 与 W 的积空间。

由此， $\text{Hom}(V \times W, \mathbb{K})$ 可以表示 $V \times W$ 到 \mathbb{K} 的线性映射。我们下面证明，若一个双线性函数属于 $\text{Hom}(V \times W, \mathbb{K})$ ，则它只能是零映射（即任何 (v, w) 都映射到 0），从而结论成立。

证明：

若 V, W 上的双线性函数 φ 不是零映射，至少存在 $v_0 \in V, w_0 \in W$ 使得

$$\varphi(v_0, w_0) = c \neq 0$$

由于其为线性映射，有

$$\varphi(2v_0, 2w_0) = \varphi(2(v_0, w_0)) = 2\varphi(v_0, w_0) = 2c$$

但由于其为双线性函数，又有

$$\varphi(2v_0, 2w_0) = 2\varphi(v_0, 2w_0) = 4\varphi(v_0, w_0) = 4c$$

于是 $2c = 4c$ ，与 $c \neq 0$ 矛盾。

* 虽然教学范围并没有积空间的概念，但将双线性函数当作线性函数处理确实是常犯的错误，尤其注意

$$\varphi(\lambda v, \lambda w) = \lambda^2 \varphi(v, w)$$

24.1.2 非退化性

在有个这个新概念以后，我们就需要开始思考可以研究它的何种性质。对于一个线性映射，我们会考虑它的 Ker 与 Im ，也即零点与像空间。对双线性函数，像空间的意义并不大：只要一个双线性函数不是零映射，设 $\varphi(v, w) = c \neq 0$ ，有 $\varphi(\lambda v, w) = \lambda c$ ，考虑 λ 取遍 \mathbb{K} 可得其值域必然为 \mathbb{K} 。不过，它的零点集合仍然是重要的，事实上有结论：若两个双线性函数的零点集合相同，则它们一定相差倍数，在复习题中我们将证明这个结论。

遗憾的是，双线性函数的零点集合往往并不构成线性空间 ($V \times W$ 的子空间意义下)，例如 \mathbb{R} 上的双线性函数 $f(x, y) = xy$ ，其零点集合为 $x = 0$ 或 $y = 0$ ，这自然不是 \mathbb{R}^2 的子空间。不过，仔细观察可以发

现， $x = 0$ 或 $y = 0$ 对这个双线性函数是具有特殊性的， $x = 0$ 时无论 y 如何取，都有 $f(x, y) = 0$ 。我们将满足这样性质的集合称为双线性函数的根。

具体来说，对 V, W 上的双线性函数 f ，定义（严格来说这里的 \perp 应当为 \perp_f ，我们将在不会有歧义时省略下标 f 以简化书写）

$${}^\perp W = \{v \in V \mid \forall w \in W, f(v, w) = 0\}$$

$$V^\perp = \{w \in W \mid \forall v \in V, f(v, w) = 0\}$$

分别称为 f 的左根与右根。

很明显，这样的定义可以推广到任何 V 的子集 S 与 W 的子集 T ：若 $T \subset W$ ，定义其左补为

$${}^\perp T = \{v \in V \mid \forall w \in T, f(v, w) = 0\}$$

若 $S \subset V$ ，定义其右补为

$$S^\perp = \{w \in W \mid \forall v \in S, f(v, w) = 0\}$$

这样，左根就是 W 的左补，右根就是 V 的右补。

有结论： ${}^\perp T = {}^\perp \langle T \rangle$ 是 V 的子空间， $S^\perp = \langle S \rangle^\perp$ 是 W 的子空间。

证明：

此处只证明第一句，第二句完全类似即可得到。

首先，利用生成子空间定义，对任何 $\langle T \rangle$ 中元素 w ，设

$$w = \sum_{i=1}^n \lambda_i w_i$$

其中 $\lambda_i \in \mathbb{K}$ 、 $w_i \in T$ ，则利用双线性性，对任何 $v \in {}^\perp T$ 由定义有

$$f(v, w) = \sum_{i=1}^n \lambda_i f(v, w_i) = \sum_{i=1}^n \lambda_i 0 = 0$$

于是 $v \in {}^\perp \langle T \rangle$ ，这就说明了 ${}^\perp \langle T \rangle \subset {}^\perp T$ 。另一方面，由于 $T \subset \langle T \rangle$ ，从定义可以直接看出

$$\{v \in V \mid \forall w \in \langle T \rangle, f(v, w) = 0\} \subset \{v \in V \mid \forall w \in T, f(v, w) = 0\}$$

于是 ${}^\perp T \subset {}^\perp \langle T \rangle$ ，综合得证 ${}^\perp T = {}^\perp \langle T \rangle$ 。

我们下面说明它是 V 的子空间。对任何 $v_1, v_2 \in {}^\perp T$ 、任何 $\lambda \in \mathbb{K}$ 与任何 $w \in T$ ，有

$$f(v_1 + v_2, w) = f(v_1, w) + f(v_2, w) = 0, \quad f(\lambda v_1, w) = \lambda f(v_1, w) = 0$$

从而 $v_1 + v_2$ 与 λv_1 也在 ${}^\perp T$ 中，得证。

至此，我们成功用线性空间刻画出了 f 的零点集合： f 的所有零点可以描述成

$$\{(v, w) \in V \times W \mid v \in {}^\perp \{w\}\} = \{(v, w) \in V \times W \mid w \in \{v\}^\perp\}$$

就像有了 Ker 与 Im 后，我们可以单射与满射，有了左根、右根后，我们可以定义双线性函数的非退化性：若 ${}^\perp W = \{0\}$ ，则称 f 为左非退化；若 $V^\perp = \{0\}$ ，则称 f 为右非退化。若 f 左非退化且右非退化，则称它非退化。

* 此定义暗含了 $f(0, w) = f(v, 0) = 0$ ，这是因为固定右侧为 w 或左侧为 v 时， f 对另一个分量均为线性映射，因此利用线性映射性质必然在 0 处为 0。

那么，到底什么是非退化呢？一个直观的理解是，**非退化意味着可区分**。例如，当且仅当 f 左非退化时，对任何不同的 $v_1, v_2 \in V$ ，存在 $w \in W$ 使得

$$f(v_1, w) \neq f(v_2, w)$$

从而 v_1 与 v_2 在双线性函数 f 下可区分。类似地，当且仅当 f 右非退化时，对任何不同的 $w_1, w_2 \in W$ ，存在 $v \in V$ 使得 $f(v, w_1) \neq f(v, w_2)$ 。

证明：

同样只证明第一句，第二句完全类似即可得到。

若 f 存在非零左根 v ，对任何 $w \in W$ 有 $f(0, w) = f(v, w) = 0$ ，从而可区分性不成立。

若存在不同的 $v_1, v_2 \in V$ 不可区分，即对任何 $w \in W$ 有 $f(v_1, w) = f(v_2, w)$ ，作差利用线性性得 $f(v_1 - v_2, w) = 0$ 恒成立，于是 $v_1 - v_2$ 是 f 的左根，且由 $v_1 \neq v_2$ 其非零。

综合以上两部分得证。

* 可以发现，非退化的定义与单射有一定的相似处。在之后讨论矩阵表示时将看到，若 V, W 维数均有限，右非退化/左非退化/非退化与单射/满射/双射有微妙的对应关系。

最后，我们证明关于和空间、交空间补的性质。这里假设 f 是 \mathbb{K} 上的线性空间 V, W 上的双线性函数，并设 U_1, U_2 为 V 的子空间：

$$1. (U_1 + U_2)^\perp = U_1^\perp \cap U_2^\perp$$

证明：

$$-(U_1 + U_2)^\perp \subset U_1^\perp \cap U_2^\perp$$

设 $w \in (U_1 + U_2)^\perp$ ，也即对任何 $v \in U_1 + U_2$ 都有 $f(v, w) = 0$ 从而对任意 $v \in U_1$ 或 $v \in U_2$ 都有 $f(v, w) = 0$ ，即得到 $w \in U_1^\perp$ 且 $w \in U_2^\perp$ 。

$$- U_1^\perp \cap U_2^\perp \subset (U_1 + U_2)^\perp$$

设 $w \in U_1^\perp \cap U_2^\perp$ ，也即对任何 $v \in U_1$ 或 $v \in U_2$ 都有 $f(v, w) = 0$ 。对任何 $v \in U_1 + U_2$ ，设 $v = v_1 + v_2$ ，且 $v_1 \in V_1, v_2 \in V_2$ ，则

$$f(v, w) = f(v_1, w) + f(v_2, w) = 0$$

从而得证 $w \in (U_1 + U_2)^\perp$ 。

$$2. U_1^\perp + U_2^\perp \subset (U_1 \cap U_2)^\perp$$

证明：

只需证明 $U_1^\perp \subset (U_1 \cap U_2)^\perp$ 即可，对称可证 $U_2^\perp \subset (U_1 \cap U_2)^\perp$ ，再由子空间的和还是子空间得证。

根据定义，若 $w \in U_1^\perp$ ，对任何 $v \in U_1$ 都有 $f(v, w) = 0$ ，从而由 $U_1 \cap U_2 \subset U_1$ 对任何 $v \in (U_1 \cap U_2)$ 有 $f(v, w) = 0$ ，得证。

同理，对左补有对称的性质：

- ${}^\perp(U + W) = {}^\perp U \cap {}^\perp W$ ；
- ${}^\perp U + {}^\perp W \subset {}^\perp(U \cap W)$ 。

24.1.3 限制映射与同构定理

接下来，就像线性映射可以定义限制映射，对于双线性函数也有自然的限制：对 \mathbb{K} 上的线性空间 V, W ，设 V_0 为 V 的子空间， W_0 为 W 的子空间，给定 V, W 上的双线性函数 f ，可以定义 V_0, W_0 上的函

数 $f|_{V_0, W_0}$ 满足

$$\forall v \in V_0, w \in W_0, \quad f|_{V_0, W_0}(v, w) = f(v, w)$$

我们将其称为 f 在 V_0, W_0 上的限制映射。若 $V = W, V_0 = W_0$, 我们简称其为 f 在 V_0 上的限制映射，并直接记为 $f|_{V_0}$ 。类似线性映射的情况，直接由定义即可验证双线性函数的限制映射仍为双线性函数。

对于线性映射的限制映射，我们可以从原映射的 Ker 与 Im 简单得到限制映射的 Ker 与 Im，然而，这点对双线性函数并不成立。考虑如下例子（写成行向量以方便书写）：

$$V = W = \mathbb{R}^2, \quad f((x_1, y_1), (x_2, y_2)) = x_1 x_2 - y_1 y_2$$

可以直接验证其为非退化的双线性函数。然而，考虑 $V_0 = \langle(1, 1)\rangle$ ，可计算得对任何 $v, w \in V_0$ 有 $f(v, w) = 0$ ，从而

$$f|_{V_0} = 0$$

于是 V_0 为限制映射的左根与右根，从而即使原映射非退化，限制映射也无法保持非退化性。

* 之后学习实内积空间时可以发现，若 f 满足正定性（这可以推出非退化性），则其限制映射仍然是正定的，这就保证了实内积空间的子空间仍然是实内积空间。

* 由于一般双线性函数的限制映射无法保持太多性质，我们这里只引入定义，之后不再进行更进一步的讨论。事实上，学习矩阵表示后，限制映射的矩阵表示同样可以看作原映射矩阵表示的分块。

另一个值得考虑的问题是，就像第一同构定理可以从任何一个线性映射出发构造同构，从任何一个双线性函数出发，能否构造一个非退化的双线性函数呢？此处的结论与第一同构定理的形式也有很多类似处，分为商空间与补空间两个版本：

1. 商空间版本

记 $V_0 = {}^\perp W, W_0 = V^\perp$ 。

设 $\tilde{V} = V/V_0, \tilde{W} = W/W_0$ ，定义 \tilde{V}, \tilde{W} 上的函数 \tilde{f} 满足

$$\forall v \in V, w \in W, \quad \tilde{f}(v + V_0, w + W_0) = f(v, w)$$

则其为非退化的双线性函数。

证明：

- 定义合理性

若 $v + V_0 = v' + V_0, w + W_0 = w' + W_0$ ，利用商空间性质可知 $v - v' \in V_0, w - w' \in W_0$ ，从而利用左右根定义可知 $f(v - v', w') = 0, f(v, w - w') = 0$ ，从而再由双线性性

$$f(v', w') = f(v', w') + f(v - v', w') + f(v, w - w') = f(v, w') + f(v, w - w') = f(v, w)$$

由此， \tilde{f} 确实将每对等价类映射到了同一个元素，定义合理。

- 双线性性

我们只对第一个分量验证线性，第二个分量完全类似。

对任何 $\lambda, \mu \in \mathbb{K}$ 与 $v_1, v_2 \in V, w \in W$ ，有（第一个等号来自商空间运算定义）

$$\begin{aligned} \tilde{f}(\lambda(v_1 + V_0) + \mu(v_2 + V_0), w + W_0) &= \tilde{f}((\lambda v_1 + \mu v_2) + V_0, w + W_0) \\ &= f(\lambda v_1 + \mu v_2, w) = \lambda f(v_1, w) + \mu f(v_2, w) \\ &= \lambda \tilde{f}(v_1 + V_0, w + W_0) + \mu \tilde{f}(v_2 + V_0, w + W_0) \end{aligned}$$

从而符合线性性定义。

– 非退化性

我们只验证左非退化性，右非退化性完全类似。

由左根定义，对任何 $v \in V$ 且 $v \notin V_0$ ，存在 $w \in W$ 使得 $f(v, w) \neq 0$ ，因此

$$\tilde{f}(v + V_0, w + W_0) = f(v, w) \neq 0$$

于是， $v + V_0$ 是 \tilde{f} 的左根当且仅当 $v \in V_0$ ，从而由商空间等价类性质其即等于 V_0 ，即商空间中的零元，这就证明了左非退化性。

2. 补空间版本

记 $V_0 = {}^\perp W$ 、 $W_0 = V^\perp$ 。

设 V_0 对 V 的某个补空间为 V_1 ， W_0 对 W 的某个补空间为 W_1 ，则限制映射 $f|_{V_1, W_1}$ 是非退化的双线性函数。

证明：

之前已说明限制映射是双线性函数，只需证明非退化性。我们同样只验证左非退化性，右非退化性完全类似。

由 $V_1 \cap V_0 = \{0\}$ ，若 $v_1 \in V_1$ 非零，存在 $w \in W$ 使得 $f(v_1, w) \neq 0$ 。由于 $W = W_0 \oplus W_1$ ，设 $w = w_0 + w_1$ ，其中 $w_0 \in W_0$ ， $w_1 \in W_1$ ，由右根性质可得

$$f(v_1, w_1) = f(v_1, w_1) + f(v_1, w_0) = f(v_1, w) \neq 0$$

也即对任何非零的 $v_1 \in V_1$ ，存在 $w_1 \in W_1$ 使得 $f(v_1, w_1) \neq 0$ ，这就说明了 $f|_{V_1, W_1}$ 的左根只有 $\{0\}$ ，从而其有左非退化性。

24.1.4 双线性函数空间

最后，就像线性映射可以由基映射构造得到，我们希望知道双线性函数与基的关系，例如能否通过基映射构造双线性函数、能否通过基映射结果确定非退化性等。后者对于双线性函数是非常困难的，但前者仍然可行。

我们接下来证明，对于 \mathbb{K} 上的线性空间 V 与 W ，给定 V 的一组基 $S = \{v_i \mid i \in I\}$ ， W 的一组基 $T = \{w_j \mid j \in J\}$ ，则对任何一组 $c_{ij}, i \in I, j \in J$ ，存在唯一双线性函数 f 满足

$$\forall i \in I, j \in J, \quad f(v_i, w_j) = c_{ij}$$

证明：

– 存在性

由于对任何 $v \in V$ 、 $w \in W$ ，利用基的定义可以唯一写成 ($v = 0$ 对应 $n = 0$ ， $w = 0$ 对应 $m = 0$ ，否则要求所有系数非零)

$$v = \sum_{k=1}^n \lambda_k v_{i_k}, \quad w = \sum_{l=1}^m \mu_l w_{j_l}$$

这里 $i_k \in I$ 、 $j_l \in J$ 。由此可以定义 $f(v, w)$ 当 v 或 w 为 0 时为 0，否则满足

$$f(v, w) = \sum_{k=1}^n \sum_{l=1}^m \lambda_k \mu_l c_{i_k j_l}$$

由于被基表出的存在唯一性，这确实对 f 在任何一点处的值给出了唯一定义。由于 v_i 的唯一表示即为 $1v_i$ 、 w_j 的唯一表示即为 $1w_j$ ， f 的确满足 $f(v_i, w_j) = c_{ij}$ ，下面证明其为双线性函数。

我们只对第一个分量验证线性，对第二个分量完全类似即可。对 $v', v'' \in V$ 、 $a, b \in \mathbb{K}$ ，设（下方表出可以有系数为 0，这样能让 v' 、 v'' 选取相同的一些基）

$$v' = \sum_{k=1}^N a_k v_{i'_k}, \quad v'' = \sum_{k=1}^N b_k v_{i'_k}$$

这里 $i'_k \in I$ 。由于加 0 不改变元素，可以发现仍有

$$f(v', w) = \sum_{k=1}^N \sum_{l=1}^m a_k \mu_l c_{i'_k j_l}$$

$$f(v'', w) = \sum_{k=1}^N \sum_{l=1}^m b_k \mu_l c_{i'_k j_l}$$

由此，直接利用结合律、分配律计算可知（注意第二个等号开始所有的符号都表示 \mathbb{K} 中数，可任意进行交换、结合、分配）

$$af(v', w) + bf(v'', w) = a \sum_{k=1}^N \sum_{l=1}^m a_k \mu_l c_{i_k j_l} + b \sum_{k=1}^N \sum_{l=1}^m b_k \mu_l c_{i'_k j_l} = \sum_{k=1}^N \sum_{l=1}^m (aa_k + bb_k) \mu_l c_{i'_k j_l}$$

另一方面，利用线性空间性质计算可知

$$av' + bv'' = \sum_{k=1}^N (aa_k + bb_k) v_{i'_k}$$

从而根据 f 的定义，由坐标唯一性有

$$f(av' + bv'', w) = \sum_{k=1}^N \sum_{l=1}^m (aa_k + bb_k) \mu_l c_{i'_k j_l}$$

于是即得到 $af(v', w) + bf(v'', w) = f(av' + bv'', w)$ ，对第一个分量的线性成立。

- 唯一性

若 f 为满足条件的双线性函数，对任何 $v \in V$ 、 $w \in W$ ，设

$$v = \sum_{k=1}^n \lambda_k v_{i_k}, \quad w = \sum_{l=1}^m \mu_l w_{j_l}$$

这里 $i_k \in I$ 、 $j_l \in J$ ，则利用双线性性可得

$$f(v, w) = \sum_{k=1}^n \lambda_k f(v_{i_k}, w) = \sum_{k=1}^n \lambda_k \sum_{l=1}^m \mu_l f(v_{i_k}, w_{j_l}) = \sum_{k=1}^n \lambda_k \sum_{l=1}^m \mu_l c_{i_k j_l}$$

由此，再通过一步乘法分配律可得 f 必须满足存在性证明中给出的定义，从而唯一。

* 注意对比此处证明与本讲义 19.2.2 的类似性，这也可以称为从基映射构造双线性函数。

事实上，某种意义上，从给定的 c_{ij} 构造 f 的过程是线性的。首先，对于两个 $X \rightarrow \mathbb{K}$ 的映射 ϕ_1, ϕ_2 ，我们定义它们的和为

$$(\phi_1 + \phi_2)(x) = \phi_1(x) + \phi_2(x)$$

一个映射数乘 $\lambda \in \mathbb{K}$ 为

$$(\lambda \phi_1)(x) = \lambda \phi_1(x)$$

由此，双线性函数也可以定义和与数乘。对于 \mathbb{K} 上的线性空间 V 与 W ，给定 V 的一组基 $S = \{v_i \mid i \in I\}$ ， W 的一组基 $T = \{w_j \mid j \in J\}$ ，我们有如下结论：

1. 所有 $S \times T \rightarrow \mathbb{K}$ 的映射构成线性空间，记为 $\text{Map}(S \times T, \mathbb{K})$ 。

证明：

这里定义映射的运算与本讲义 20.1.1 第 2 题完全相同，根据该题结论，对任何集合 X 与线性空间 V ， $\text{Map}(X, V)$ 都构成线性空间，此处取 $X = S \times T$ 、 $V = \mathbb{K}$ 为其中特例。

2. 所有 V, W 上的双线性函数构成线性空间，记为 $T(V, W)$ 。

证明：

与上方类似可知 $\text{Map}(V \times W, \mathbb{K})$ 为线性空间，由定义 $T(V, W)$ 为其子集，为说明其为线性空间只需证明封闭性。

若 $f, g \in T(V, W)$ ，对任何 $\lambda, \mu, a, b \in \mathbb{K}$ 、 $v, v_1, v_2 \in V$ 、 $w, w_1, w_2 \in W$ ，根据映射的加法、数乘定义与双线性性有

$$\begin{aligned} (\lambda f + \mu g)(av_1 + bv_2, w) &= \lambda f(av_1 + bv_2, w) + \mu g(av_1 + bv_2, w) \\ &= \lambda af(v_1, w) + \lambda bf(v_2, w) + \mu ag(v_1, w) + \mu bg(v_2, w) \\ &= a(\lambda f(v_1, w) + \mu g(v_1, w)) + b(\lambda f(v_2, w) + \mu g(v_2, w)) \\ &= a(\lambda f + \mu g)(v_1, w) + b(\lambda f + \mu g)(v_2, w) \end{aligned}$$

同理

$$(\lambda f + \mu g)(v, aw_1 + bw_2) = a(\lambda f + \mu g)(v, w_1) + b(\lambda f + \mu g)(v, w_2)$$

由此证明了双线性函数的线性组合仍然是双线性函数，从而得到了封闭性。

3. $T(V, W)$ 与 $\text{Map}(S \times T, \mathbb{K})$ 同构。

证明：

在从基映射构造双线性函数的过程中，我们事实上已经从任何一个 $\varphi \in \text{Map}(S \times T, \mathbb{K})$ (满足对任何 $i \in I, j \in J$ 有 $\varphi(v_i, w_j) = c_{ij}$) 得到了唯一的 $f \in T(V, W)$ 。我们将这个过程记作映射 T ，即 $T(\varphi) = f$ 。为了书写清晰，我们用中括号表示映射到映射的映射，将其写作

$$f = T[\varphi]$$

由于存在唯一性，我们已经证明了 T 的良好定义性，下面证明其为线性同构。

- 线性性

对于 $\varphi_1, \varphi_2 \in \text{Map}(S \times T, \mathbb{K})$ ， $\lambda, \mu \in \mathbb{K}$ ，记 $\varphi = \lambda\varphi_1 + \mu\varphi_2$ ，根据映射加法、数乘定义可知

$$\forall i \in I, j \in J, \quad \varphi(v_i, w_j) = \lambda\varphi_1(v_i, w_j) + \mu\varphi_2(v_i, w_j)$$

由此，设 $f = T[\varphi]$ 、 $f_1 = T[\varphi_1]$ 、 $f_2 = T[\varphi_2]$ ，对任何 $v \in V$ 、 $w \in W$ ，设

$$v = \sum_{k=1}^n \lambda_k v_{i_k}, \quad w = \sum_{l=1}^m \mu_l w_{j_l}$$

这里 $i_k \in I$ 、 $j_l \in J$ ，则有

$$\begin{aligned} f(v, w) &= \sum_{k=1}^n \sum_{l=1}^m \lambda_k \mu_l \varphi(v_{i_k}, w_{j_l}) \\ &= \sum_{k=1}^n \sum_{l=1}^m \lambda_k \mu_l (\lambda\varphi_1(v_{i_k}, w_{j_l}) + \mu\varphi_2(v_{i_k}, w_{j_l})) \\ &= \lambda \sum_{k=1}^n \sum_{l=1}^m \lambda_k \mu_l \varphi_1(v_{i_k}, w_{j_l}) + \mu \sum_{k=1}^n \sum_{l=1}^m \lambda_k \mu_l \varphi_2(v_{i_k}, w_{j_l}) \end{aligned}$$

而根据定义，这即是 $\lambda f_1(v, w) + \mu f_2(v, w)$ ，从而

$$T[\lambda\varphi_1 + \mu\varphi_2] = \lambda T[\varphi_1] + \mu T[\varphi_2]$$

T 的确是线性映射。

- 单射

由于 f 在 $S \times T$ 中的值与 φ 相同，不同的 φ 对应的 f 至少在 $S \times T$ 中某点不同，从而不可能相同，因此其为单射。

- 满射

对任何 f ，定义 φ 为 f 在 $S \times T$ 上的限制映射（这里只取映射的限制映射，忽略线性性），则 $T[\varphi]$ 与 f 在 $S \times T$ 上的值完全相同。由于指定了 f 在 $S \times T$ 上的值后得到的双线性函数是唯一的，必然有 $T[\varphi] = f$ ，这就说明 φ 是 f 的原像，其为满射。

* 事实上，整段的分析都与本讲义 20.1.1 第 2 题非常类似。

至此，我们已经将所有可以与一般线性映射类似研究的内容研究完成了。下面我们将先讨论可以与有限维空间线性映射类似研究的内容，再讨论无法类似研究的性质。

§24.2 度量矩阵

24.2.1 向量空间的双线性函数

本节，我们将类似有限维空间的线性映射，给出有限维空间双线性函数的度量矩阵，并用度量矩阵刻画其性质。我们已经知道，矩阵表示是利用有限维线性空间的“同构标准形”——向量空间来“替代”一般空间，由此，仍然需要先研究向量空间中的双线性函数。

首先，直接由定义可验证，对任何 $A \in \mathbb{K}^{m \times n}$ ， $\mathbb{K}^m \times \mathbb{K}^n$ 上的函数 $f(x, y) = x^T A y$ 是一个双线性函数。

本节的核心目的即是证明，对任何 $\mathbb{K}^m, \mathbb{K}^n$ 上的双线性函数 f ，存在唯一矩阵 $A \in \mathbb{K}^{m \times n}$ ，使得

$$\forall x \in \mathbb{K}^m, y \in \mathbb{K}^n, \quad f(x, y) = x^T A y$$

证明：

- 存在性

设 $e_i^{(m)}$ 表示 \mathbb{K}^m 中的第 i 个单位向量， $e_j^{(n)}$ 表示 \mathbb{K}^n 中的第 j 个单位向量，记

$$a_{ij} = f(e_i^{(m)}, e_j^{(n)}), \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

将所有 a_{ij} 拼成矩阵 A ，我们下面证明 $f(x, y) = x^T A y$ 对任何 $x \in \mathbb{K}^m, y \in \mathbb{K}^n$ 成立。

直接展开计算可知

$$x^T A y = \sum_{i=1}^m \sum_{j=1}^n x_i y_j a_{ij}$$

另一方面，由于

$$x = \sum_{i=1}^m x_i e_i^{(m)}, \quad y = \sum_{j=1}^n y_j e_j^{(n)}$$

由双线性性

$$f(x, y) = \sum_{i=1}^m x_i f(e_i^{(m)}, y) = \sum_{i=1}^m x_i \sum_{j=1}^n y_j f(e_i^{(m)}, e_j^{(n)}) = \sum_{i=1}^m x_i \sum_{j=1}^n y_j a_{ij}$$

再通过一步乘法分配律可知两者相等。

— 唯一性

设 $f_A(x, y) = x^T A y$ 、 $f_B(x, y) = x^T B y$, 若 $A \neq B$, 设 i, j 满足 $a_{ij} \neq b_{ij}$, 则计算可得

$$f_A(e_i^{(m)}, e_j^{(n)}) = a_{ij} \neq b_{ij} = f_B(e_i^{(m)}, e_j^{(n)})$$

从而 $f_A \neq f_B$, 由此可知表示唯一。

此外, 上述确定 A 的过程是 $T(\mathbb{K}^m, \mathbb{K}^n)$ 到 $\mathbb{K}^{m \times n}$ 的同构, 从而 $\dim T(\mathbb{K}^m, \mathbb{K}^n) = mn$ 。

证明:

设从双线性函数 $T(\mathbb{K}^m, \mathbb{K}^n)$ 构造 $\mathbb{K}^{m \times n}$ 的过程为映射 L , 即当 $f(x, y) = x^T A y$ 时 $L(f) = A$ 。由于存在唯一性, 我们已经证明了 L 的良好定义性, 下面证明其为线性同构。

— 线性性

对 $f_1, f_2 \in T(\mathbb{K}^m, \mathbb{K}^n)$ 、 $\lambda, \mu \in \mathbb{K}$, 设 $L(f_1) = A_1$ 、 $L(f_2) = A_2$, 有 (最后一步利用了矩阵乘法的分配律与数乘可以任意移动位置)

$$(\lambda f_1 + \mu f_2)(x, y) = \lambda f_1(x, y) + \mu f_2(x, y) = \lambda x^T A_1 y + \mu x^T A_2 y = x^T (\lambda A_1 + \mu A_2) y$$

由此

$$L(\lambda f_1 + \mu f_2) = \lambda A_1 + \mu A_2 = \lambda L(f_1) + \mu L(f_2)$$

这就说明了线性性。

— 双射

我们记映射 T 满足 $T(A) = f_A$, 这里 $f_A(x, y) = x^T A y$, 则直接由定义验证可发现

$$\forall f \in T(\mathbb{K}^m, \mathbb{K}^n), \quad T(L(f)) = f$$

$$\forall A \in \mathbb{K}^{m \times n}, \quad L(T(A)) = A$$

从而 T 是 L 的逆映射, 因此 L 为双射。

在这一部分的最后, 我们利用这样的同构给出 $T(\mathbb{K}^m, \mathbb{K}^n)$ 的一组基。

解答:

在刚才的证明中, 我们已经构造了 L 的逆映射 T 。由于同构的逆映射也是同构, T 是 $\mathbb{K}^{m \times n} \rightarrow T(\mathbb{K}^m, \mathbb{K}^n)$ 的线性同构, 因此其将一组基映射到一组基。

设 $E_{ij} \in \mathbb{K}^{m \times n}$ 为只有第 i 行第 j 列为 1, 其他为 0 的矩阵, 则所有

$$\{E_{ij} \mid i = 1, \dots, m, \quad j = 1, \dots, n\}$$

构成 $\mathbb{K}^{m \times n}$ 一组基, 于是所有

$$\{T(E_{ij}) \mid i = 1, \dots, m, \quad j = 1, \dots, n\}$$

构成 $T(\mathbb{K}^m, \mathbb{K}^n)$ 一组基。我们下面设 $f_{ij} = T(E_{ij})$, 更进一步计算得:

$$f_{ij}(x, y) = x^T E_{ij} y = x_i y_j$$

从而所有满足 $f_{ij}(x, y) = x_i y_j$ 的 f_{ij} 构成 $T(\mathbb{K}^m, \mathbb{K}^n)$ 的一组基, 这里 $i = 1, \dots, m$ 、 $j = 1, \dots, n$ 。

24.2.2 度量矩阵

解决了向量空间的问题后，自然就应该开始解决一般的有限维线性空间。对 \mathbb{K} 上的 m 维线性空间 V 与 n 维线性空间 W ，给定 V 的一组基 S 、 W 的一组基 T ，用下标 S 、 T 表示元素在基下的坐标，我们先证明对任何 $f \in T(V, W)$ ，存在唯一矩阵 $A \in \mathbb{K}^{m \times n}$ 使得

$$\forall v \in V, w \in W, \quad f(v, w) = v_S^T A w_T$$

证明：

设 π_S 为坐标映射 $\pi_S(v) = v_S$ ， π_T 为坐标映射 $\pi_T(w) = w_T$ ，之前已证明它们均为线性同构。

- 存在性

对任何 $f \in T(V, W)$ ，定义 $\varphi : \mathbb{K}^m \times \mathbb{K}^n \rightarrow \mathbb{K}$ 使得

$$\forall x \in \mathbb{K}^m, y \in \mathbb{K}^n, \quad \varphi(x, y) = f(\pi_S^{-1}(x), \pi_T^{-1}(y))$$

由于线性映射的复合是线性映射，分别考虑 φ 的两个分量可知其的确为双线性函数。由此，利用前一部分的证明，存在唯一矩阵 $A \in \mathbb{K}^{m \times n}$

$$\varphi(x, y) = x^T A y$$

由 π_S 、 π_T 为同构可知 $\pi^{-1}(x)$ 取遍所有 v 、 $\pi^{-1}(y)$ 取遍所有 w ，从而可进一步得到

$$f(v, w) = \varphi(\pi_S(v), \pi_T(w)) = v_S^T A w_T$$

这已经符合要求。

- 唯一性

若存在 $A_1, A_2 \in \mathbb{K}^{m \times n}$ 使得

$$f(v, w) = v_S^T A_1 w_T = v_S^T A_2 w_T$$

由坐标映射为同构， v_S 可取遍 \mathbb{K}^m 、 w_T 可取遍 \mathbb{K}^n ，从而对任何 $x \in \mathbb{K}^m$ 、 $y \in \mathbb{K}^n$ 有

$$x^T A_1 y = x^T A_2 y$$

这时左右都是 $\mathbb{K}^m, \mathbb{K}^n$ 上的双线性函数，利用上一部分已证可知对应的矩阵唯一，因此 $A_1 = A_2$ ，得证。

我们将上述的 A 称为 f 在基 S, T 下的度量矩阵——可以看作另一种称呼的矩阵表示。事实上，度量矩阵的元素有一种更简单的计算方式：设 $S = (v_1, \dots, v_m)$ 、 $T = (w_1, \dots, w_n)$ ，则 A 的第 i 行第 j 列元素 a_{ij} 应满足

$$a_{ij} = f(v_i, w_j)$$

证明：

与上一部分相同计算可以发现，当 $v_S = e_i^{(m)}$ 、 $w_T = e_j^{(n)}$ 时

$$f(v, w) = v_S^T A w_T = a_{ij}$$

而由坐标定义， $v_S = e_i^{(m)}$ 即意味着 $v = v_i$ ， $w_T = e_j^{(n)}$ 即意味着 $w = w_j$ ，从而得证。

* 我们可以将其简单记作 $A = f(S \times T)$ ，这也是教材上定义度量矩阵的方式。

与上一部分类似，我们也可以证明度量矩阵是 $T(V, W)$ 到 $\mathbb{K}^{m \times n}$ 的同构，从而 $\dim T(V, W) = mn$ 。

证明：

在之前的证明中，我们对任何 $f \in T(V, W)$ 构造出了 $\varphi \in T(\mathbb{K}^m, \mathbb{K}^n)$ 使得

$$\forall x \in \mathbb{K}^m, y \in \mathbb{K}^n, \quad \varphi(x, y) = f(\pi_S^{-1}(x), \pi_T^{-1}(y))$$

由于这对任何点给出了定义，我们对任何 f 都定义出了唯一的 φ ，由此可记上述关系为满足 $\varphi = P(f)$ 的映射 P 。我们先证明 P 是 $T(V, W)$ 到 $T(\mathbb{K}^m, \mathbb{K}^n)$ 的同构。

- 线性性

对任何 $f_1, f_2 \in T(V, W)$ 、 $\lambda, \mu \in \mathbb{K}$ ，设

$$\varphi_1 = P(f_1), \quad \varphi_2 = P(f_2), \quad f = \lambda f_1 + \mu f_2, \quad \varphi = P(f)$$

则根据 P 与映射加法、数乘定义可知

$$\begin{aligned} \varphi(x, y) &= (\lambda f_1 + \mu f_2)(\pi_S^{-1}(x), \pi_T^{-1}(y)) \\ &= \lambda f_1(\pi_S^{-1}(x), \pi_T^{-1}(y)) + \mu f_2(\pi_S^{-1}(x), \pi_T^{-1}(y)) \\ &= \lambda \varphi_1(x, y) + \mu \varphi_2(x, y) \end{aligned}$$

再代入 φ 的定义即得线性性。

- 双射

对任何 $\varphi \in T(\mathbb{K}^m, \mathbb{K}^n)$ ，构造 $f \in \text{Map}(V \times W, \mathbb{K})$ 使得

$$f(v, w) = \varphi(v_S, w_T)$$

由于线性映射的复合是线性映射，分别考虑 f 的两个分量可知其的确为双线性函数。由于这对任何点给出了定义，我们对任何 φ 都定义出了唯一的 Q ，由此可记上述关系为满足 $f = Q(\varphi)$ 的映射 Q 。

利用定义可直接验证对任何 $f \in T(V, W)$ 有

$$\forall v \in V, w \in W, \quad Q(P(f))(v, w) = P(f)(v_S, w_T) = f(\pi_S^{-1}(v_S), \pi_T^{-1}(w_T)) = f(v, w)$$

从而 $Q \circ P$ 为恒等映射，类似计算得 $P \circ Q$ 为恒等映射，于是 Q 是 P 的逆映射，因此 P 为双射。

可发现，之前证明中从 φ 得到 A 的方式即是前一部分构造的同构 L ，由此利用同构的复合还是同构可知 $L \circ P$ 就是 $T(V, W)$ 到 $\mathbb{K}^{m \times n}$ 的同构。

最后，我们还是从 $\mathbb{K}^{m \times n}$ 的一组基出发给出 $T(V, W)$ 的一组基。

解答：

与上一部分类似，考虑

$$\{E_{ij} \mid i = 1, \dots, m, j = 1, \dots, n\}$$

利用同构性，通过这组 $\mathbb{K}^{m \times n}$ 的基可以构造 $T(V, W)$ 的一组基。

具体来说，设 f_{ij} 是以 E_{ij} 为基 S, T 下度量矩阵的 $T(V, W)$ 中双线性函数，即有

$$f_{ij}(v, w) = v_S^T E_{ij} w_T$$

则根据同构将一组基映射到一组基，所有

$$\{f_{ij} \mid i = 1, \dots, m, j = 1, \dots, n\}$$

构成 $T(V, W)$ 的一组基。

更进一步计算可知

$$f_{ij}(v, w) = v_S^T E_{ij} w_T = (v_S)_i (w_T)_j$$

即其在 (v, w) 的值为 v 中 v_i 的坐标分量乘 w 中 w_j 的坐标分量。

* 这两部分的操作与本讲义 19.3.2 到 19.3.3 本质上相同。

24.2.3 基变换与相抵

既然我们定义了基下的度量矩阵，仍然与线性映射类似，我们希望知道进行基变换后度量矩阵会进行怎样的改变。

设 V, W 为 \mathbb{K} 上的 m, n 维线性空间，且 S, S' 为 V 的基， S 到 S' 的过渡矩阵为 P ， T, T' 为 W 的基， T 到 T' 的过渡矩阵为 Q 。假设 $f \in T(V, W)$ 在基 S, T 下的度量矩阵为 A 、在基 S', T' 下的度量矩阵为 A' ，我们来计算 A 与 A' 的关系。

解答：

由度量矩阵定义，对任何 $v \in V, w \in W$ 有

$$f(v, w) = v_S^T A w_T = v_{S'}^T A' w_{T'}$$

本讲义 18.2.1 中已经给出了坐标变换与过渡矩阵的关系，即

$$v_{S'} = P^{-1} v_S, \quad w_{T'} = Q^{-1} w_T$$

代入并利用 $(XY)^T = Y^T X^T$ 可得

$$v_S^T A w_T = v_S^T P^{-T} A' Q^{-1} w_T$$

与之前的分析类似，由于 v_S 可取遍 \mathbb{K}^m 、 w_T 可取遍 \mathbb{K}^n ，左右均可看作关于 v_S, w_T 的双线性函数，利用度量矩阵唯一性可知

$$A = P^{-T} A' Q^{-1}$$

同时左乘 P^T 、右乘 Q 改写为

$$A' = P^T A Q$$

由此，双线性函数在不同基下的度量矩阵相抵，因此其度量矩阵的秩保持不变，称为其矩阵秩。另一方面，由于过渡矩阵可以取为任何可逆阵，任何秩等于 f 的矩阵秩的 $\mathbb{K}^{m \times n}$ 中矩阵都可以看作 f 在某两组基下的度量矩阵。

有了相抵与矩阵秩后，我们自然可以得到一些相关的结论，例如从相抵标准形可以得到，设 V, W 为 \mathbb{K} 上的 m, n 维线性空间，设 $f \in T(V, W)$ 的矩阵秩为 r ，则存在 V 的一组基 S 、 W 的一组基 T 使得 f 在基 S, T 下的度量矩阵为

$$\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$$

最后，正如单射、满射可以用秩来刻画，非退化性事实上也可以用矩阵秩来刻画。具体来说， f 左非退化当且仅当其度量矩阵行满秩，右非退化当且仅当其度量矩阵列满秩，非退化当且仅当其度量矩阵可逆。

证明：

考虑 V 的任何一组基 S ， W 的任何一组基 T ， f 的度量矩阵 A 应满足

$$f(v, w) = v_S^T A w_T$$

先说明左非退化的等价条件。对 $v \in {}^\perp W$ ，由定义其对任何 $w \in W$ 都有

$$f(v, w) = v_S^T A w_T = (A^T v_S)^T w_T = 0$$

由坐标的同构性， w_T 可任取，从而取 $w_T = A^T v_S$ 可知只能 $A^T v_S = 0$ 。

由此, ${}^\perp W = \{0\}$ 当且仅当 $A^T v_S = 0$ 没有非零的 v_S 作为解 (之前已证明 $v_S = 0 \Leftrightarrow v = 0$), 利用解空间维数定理可知这当且仅当 A^T 列满秩, 即 A 行满秩。

完全类似, 右非退化可等价于 $Aw_T = 0$ 没有非零的 w_T 作为解, 利用解空间维数定理可知这当且仅当 A 列满秩。

综合以上两部分结论, 非退化等价于行满秩且列满秩, 即可逆。

最后补充几个注释:

- 考虑矩阵的阶数可发现, 当 V, W 维数有限时, 当且仅当 $\dim V \leq \dim W$ 时其可能左非退化, $\dim V \geq \dim W$ 时其可能右非退化, $\dim V = \dim W$ 时其可能非退化。
- 利用行秩等于列秩, $\dim V = \dim W$ 时, 左非退化、右非退化、非退化相互等价。
- 对上述过程利用解空间维数定理可进一步得到

$$\dim {}^\perp W = m - \text{rank } A, \quad \dim V^\perp = n - \text{rank } A$$

当 $\dim V = \dim W$ 时二者维数相等。

* 可以将这部分的讨论与本讲义 19.3.4 进行对比, 观察结论的相同与不同处。

§24.3 对称与斜对称

前两节中的所有讨论事实上都与对线性映射的研究完全类似。不过, 既然双线性函数的确是一个新概念, 其应当有不能被之前研究涵盖的内容, 而所有这些内容都指向了一个共同的根源: 对称性, 或者说, 用矩阵论的语言, 我们可以研究其转置。

24.3.1 转置

在前半学期讨论线性变换时, 我们一直回避了一个上学期经常出现的概念, 也就是矩阵的转置。既然双线性函数的度量矩阵与基变换下的变换都与转置密切相关, 可以想象它将成为我们研究转置的空间工具。

事实上, 对双线性函数来说, 其“转置”的定义非常简单: 设 V, W 是 \mathbb{K} 上的线性空间, 对 $f \in T(V, W)$, 定义

$$\forall v \in V, w \in W, \quad g(w, v) = f(v, w)$$

利用定义可直接验证 g 是 W, V 上的双线性函数。我们将这样的 g 记作 f^T 。利用定义, 转置的一些性质是可以直接验证的, 例如:

- $(f^T)^T = f$;
- f^T 左非退化当且仅当 f 右非退化;
- f^T 右非退化当且仅当 f 左非退化;
- f^T 非退化当且仅当 f 非退化。

为了说明它真的是某种“转置”, 我们需要介绍如下结论: 设 V, W 为 \mathbb{K} 上的 m, n 维线性空间, 给定 V 的一组基 S , W 的一组基 T , 设 f 在基 S, T 下的度量矩阵为 A , 则 f^T 在基 T, S 下的度量矩阵为 A^T 。

证明:

由度量矩阵定义可知

$$f(v, w) = v_S^T A w_T$$

从而 (第三个等号是 1×1 矩阵等于其转置)

$$f^T(w, v) = f(v, w) = v_S^T A w_T = (v_S^T A w_T)^T = w_T^T A^T v_S$$

这符合度量矩阵的定义式，由度量矩阵唯一性即得 f^T 在基 T, S 下的度量矩阵是 A^T 。
因此，双线性函数的转置可以看作矩阵转置的某种推广。

24.3.2 对称性与合同

可以想到，当 f 是 V 上的双线性函数时， f^T 也成为了 V 上的双线性函数，由此即可以对两者进行比较：若 $f = f^T$ ，我们称 f 对称；若 $f = -f^T$ ，我们称 f 斜对称。根据定义 f 对称当且仅当

$$\forall v, w \in V, \quad f(v, w) = f(w, v)$$

f 斜对称当且仅当

$$\forall v, w \in V, \quad f(v, w) = -f(w, v)$$

本节剩下的内容里，我们都将讨论 V 上的双线性函数与它具有的特殊性质。

我们先列举几个一般线性空间上对称或斜对称双线性函数的基本性质，以下假设 V 是 \mathbb{K} 上的线性空间：

1. 若 f 是 V 上的对称或斜对称双线性函数，则对任何子集 $S \subset V$ ， ${}^\perp S = S^\perp$ (这里 \perp 指 \perp_f)。

证明：

当 f 对称或斜对称时，由于 $f(x, y) = \pm f(y, x)$ 对任何 $x, y \in V$ 成立，它们同时为 0 或非零，因此 $f(x, y) = 0$ 当且仅当 $f(y, x) = 0$ 。

由此，根据定义有

$${}^\perp S = \{w \in V \mid \forall v \in S, f(w, v) = 0\} = \{w \in V \mid \forall v \in S, f(v, w) = 0\} = S^\perp$$

从而得证。

2. 设 V 上的双线性函数空间为 $T_2(V)$ ，则对称双线性函数集合 $T_2^S(V)$ 与斜对称双线性函数集合 $T_2^I(V)$ 都是其子空间，且

$$T_2(V) = T_2^S(V) \oplus T_2^I(V)$$

证明：

- 子空间

直接验证可知，若 f, g 均为对称/斜对称的 V 上双线性函数，对任何 $\lambda, \mu \in \mathbb{K}$ ， $\lambda f + \mu g$ 对任何 $\alpha, \beta \in V$ 满足

$$(\lambda f + \mu g)(\alpha, \beta) = \lambda f(\alpha, \beta) + \mu g(\alpha, \beta) = \pm \lambda f(\beta, \alpha) \pm \mu g(\beta, \alpha) = \pm (\lambda f + \mu g)(\beta, \alpha)$$

从而对称/斜对称双线性函数均具有封闭性，为子空间。

- 交为 $\{0\}$

若 f 既为对称双线性函数又为斜对称双线性函数，对任何 $\alpha, \beta \in V$ 有

$$f(\alpha, \beta) = -f(\beta, \alpha) = f(\beta, \alpha)$$

而等于自己相反数的数只有 0，从而必然 $f(\alpha, \beta) = 0$ ，即 f 是零映射。

- 和为全空间

对任何 $f \in T_2(V)$ ，设

$$f_S = \frac{1}{2}(f + f^T), \quad f_I = \frac{1}{2}(f - f^T)$$

由定义可发现 $f = f_S + f_I$, 另一方面, 对任何 $\alpha, \beta \in V$ 有

$$f_S(\alpha, \beta) = \frac{1}{2}(f(\alpha, \beta) + f(\beta, \alpha)) = f_S(\beta, \alpha)$$

$$f_I(\alpha, \beta) = \frac{1}{2}(f(\alpha, \beta) - f(\beta, \alpha)) = -\frac{1}{2}(f(\beta, \alpha) - f(\alpha, \beta)) = -f_I(\beta, \alpha)$$

由此可得 $f_S \in T_2^S(V)$ 、 $f_I \in T_2^I(V)$, 得证。

3. 回顾二次型定义: 对 $V \rightarrow \mathbb{K}$ 的函数 q , 若存在 V 上的双线性函数 φ 使得对任何 $\alpha \in V$ 有 $q(\alpha) = \varphi(\alpha, \alpha)$, 则称其为 V 上的二次型。

对任何 V 上的二次型 q , 存在唯一 V 上的对称双线性函数 f 使得

$$\forall \alpha \in V, \quad q(\alpha) = f(\alpha, \alpha)$$

证明:

- 存在性

设 φ 是 q 对应的双线性函数, 记 $f = \frac{1}{2}(\varphi + \varphi^T)$, 即

$$\forall \alpha, \beta \in V, \quad f(\alpha, \beta) = \frac{1}{2}(\varphi(\alpha, \beta) + \varphi(\beta, \alpha))$$

在上一个证明中我们已经说明了 f 对称, 另一方面有

$$\forall \alpha \in V, \quad f(\alpha, \beta) = \frac{1}{2}(\varphi(\alpha, \alpha) + \varphi(\alpha, \alpha)) = \varphi(\alpha, \alpha)$$

因此有 $q(\alpha) = f(\alpha, \alpha)$ 对任何 α 成立, 得证。

- 唯一性

也即要证明, 若 f, g 都是 V 上的对称双线性函数, 且对任何 $\alpha \in V$ 有 $f(\alpha, \alpha) = g(\alpha, \alpha)$, 则 $f = g$ 。直接计算可知(将右侧利用双线性性展开得到 $\frac{1}{2}(f(\alpha, \beta) + f(\beta, \alpha))$, 再利用对称性可知即为 $f(\alpha, \beta)$):

$$\forall \alpha, \beta \in V, \quad f(\alpha, \beta) = \frac{1}{2}(f(\alpha + \beta, \alpha + \beta) - f(\alpha, \alpha) - f(\beta, \beta))$$

同理

$$\forall \alpha, \beta \in V, \quad g(\alpha, \beta) = \frac{1}{2}(g(\alpha + \beta, \alpha + \beta) - g(\alpha, \alpha) - g(\beta, \beta))$$

由条件

$$f(\alpha, \alpha) = g(\alpha, \alpha), \quad f(\beta, \beta) = g(\beta, \beta), \quad f(\alpha + \beta, \alpha + \beta) = g(\alpha + \beta, \alpha + \beta)$$

* 这可以看作对称双线性函数的某种本质性质: 可以被所有 $f(\alpha, \alpha)$ 的值确定。

* 这即对应上学期结论: 对任何二次型, 存在唯一的对称阵 A 使其表示为 $f(x) = x^T A x$ 。

4. 若 f 是 V 上的双线性函数, 其斜对称当且仅当

$$\forall \alpha \in V, \quad f(\alpha, \alpha) = 0$$

证明:

若 f 斜对称, 有

$$f(\alpha, \alpha) = -f(\alpha, \alpha)$$

从而只能为 0。

反之，若 $g(\alpha, \alpha) = 0$ 对任何 α 成立，直接计算有（这里第一个等号左右为直接展开得到，建议熟悉这种展开方式）

$$\forall \alpha, \beta \in V, \quad g(\beta, \alpha) + g(\alpha, \beta) = g(\alpha + \beta, \alpha + \beta) - g(\alpha, \alpha) - g(\beta, \beta) = 0$$

从而即得 $g(\beta, \alpha) = -g(\alpha, \beta)$ ，得证。

为了对有限维情况进行更深入的讨论，我们需要先从有限维情况双线性函数的度量矩阵说起。

首先，就像线性映射矩阵表示与线性变换矩阵表示的差异，既然左右是同一个空间，我们也希望取同一组基。由此，若 \mathbb{K} 上的线性空间 V 维数有限，设其一组基为 S ，我们将 V 上的双线性函数 f 在基 S, S 下的度量矩阵简称为其在基 S 下的度量矩阵。

更进一步，设 f 在基 S 下度量矩阵为 A ，在基 S' 下度量矩阵为 A' ，且 S 到 S' 的过渡矩阵为 P ，利用之前的计算结果有

$$A' = P^T A P$$

由此，与线性映射时类似可以得到两个结论： f 在不同基下的度量矩阵合同，任何与 A 合同的矩阵都可以看作 f 在某组基下的度量矩阵。

* 这样将相抵加强为合同的过程事实上也与线性变换将相抵加强为相似的过程一致。

由此，研究合同下标准形的问题可以等价为取合适的基使得有限维线性空间上的双线性函数度量矩阵最简。我们将先对对称与斜对称的情况给出解答。

一个很基本的结论是，有限维线性空间上的双线性函数对称/斜对称当且仅当其度量矩阵对称/斜对称。

证明：

考虑有限维线性空间 V ，给定一组基 S ，假设 f 在 S 下的度量矩阵为 A 。

根据上一部分的最后， f^T 在 S 下的度量矩阵为 A^T ，利用度量矩阵是同构可知双线性函数相等当且仅当同一组基下度量矩阵相等，于是 $f = f^T$ 当且仅当 $A = A^T$ ，对称情况得证。

由于度量矩阵是线性同构， $-f^T$ 的度量矩阵为 $-A^T$ ，于是 $f = -f^T$ 当且仅当 $A = -A^T$ ，斜对称情况得证。

对 \mathbb{K} 上的 n 维线性空间 V ，若 f 是 V 上的对称双线性函数，存在一组基 S 使得 f 在 S 下的度量矩阵是对角阵。

证明：

– 先证明存在一组基 v_1, \dots, v_n 使得 $f(v_1, v_i) = 0$ 对 $i \neq 1$ 成立。

若 f 为零映射，任取一组基即成立，否则，利用之前已证可知存在向量 v_1 使得 $f(v_1, v_1) \neq 0$ ：

若所有 $f(\alpha, \alpha)$ 全为 0 则 f 斜对称，既对称又斜对称只能为 0。

由此，将 v_1 扩充为一组基 v_1, u_2, \dots, u_n ，并取

$$v_i = u_i - \frac{f(v_1, u_i)}{f(v_1, v_1)} v_1, \quad i = 2, \dots, n$$

则可发现对任何 $i = 2, \dots, n$ 都有

$$f(v_1, v_i) = f\left(v_1, u_i - \frac{f(v_1, u_i)}{f(v_1, v_1)} v_1\right) = f(v_1, u_i) - \frac{f(v_1, u_i)}{f(v_1, v_1)} f(v_1, v_1) = 0$$

* 这步的技巧非常常用，也即利用双线性性从已知的非零项构造零点。事实上，这可以看作上学期间的成对初等行列变换的推广。

另一方面，由于

$$u_i = v_i + \frac{f(v_1, u_i)}{f(v_1, v_1)} v_1$$

v_1, \dots, v_n 与 v_1, u_2, \dots, u_n 可以相互表示，从而等价， v_1, \dots, v_n 也构成全空间一组基。

- 利用度量矩阵的定义，直接计算可发现 f 在这组基下的度量矩阵为

$$\begin{pmatrix} f(v_1, v_1) & \mathbf{0}^T \\ \mathbf{0} & A_0 \end{pmatrix}$$

这里 A_0 为某个 $\mathbb{K}^{(n-1) \times (n-1)}$ 中方阵。

至此，我们已经可以用矩阵论证明完成：上述证明在矩阵论中对应，对任何 \mathbb{K} 上 n 阶对称阵 A ，存在 \mathbb{K} 上 n 阶可逆阵 P 使得

$$P^T A P = \begin{pmatrix} a_1 & \mathbf{0}^T \\ \mathbf{0} & A_0 \end{pmatrix}$$

由于合同不改变对称性， A' 仍然对称，从而存在可逆阵 P_0 使得 ($A_1 \in \mathbb{K}^{(n-2) \times (n-2)}$)

$$(P_0)^T A_0 P_0 = \begin{pmatrix} a_2 & \mathbf{0}^T \\ \mathbf{0} & A_1 \end{pmatrix}$$

由此取

$$Q_0 = P \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & P_0 \end{pmatrix}$$

即可计算得

$$Q_0^T A Q_0 = \begin{pmatrix} \text{diag}(a_1, a_2) & O \\ O & A_1 \end{pmatrix}$$

重复此过程即得到对角阵。

* 这个递减阶数归纳过程在上学期证明相似三角化、正交相似对角化等过程时十分常用。

- 我们再用空间方法给出证明。考虑 v_2, \dots, v_n 生成的子空间，记为 V_0 ，利用左/右补的线性性可知 $f(v_1, v) = 0$ 对任何 $v \in V_0$ 成立。若 $f|_{V_0}$ 为零映射，则 f 在这组基下的度量矩阵已经为 $\text{diag}(f(v_1, v_1), 0, \dots, 0)$ ，从而是对角阵。

否则，直接验证可发现 $f|_{V_0}$ 仍然是对称的，从而可构造 V_0 的一组基 v'_2, \dots, v'_n 使得 $f(v'_2, v'_i) = 0$ 对任何 $i = 3, \dots, n$ 成立。

利用 V_0 的基的等价性， v_1, v'_2, \dots, v'_n 也构成 V 的一组基，再由 $f(v_1, v) = 0$ 对任何 $v \in V_0$ 成立可知 f 在这组基下的度量矩阵应为

$$\text{diag}(f(v_1, v_1), f(v'_2, v'_2), A_1)$$

的形式，这里 $A_1 \in \mathbb{K}^{(n-2) \times (n-2)}$ 。固定 v'_2 ，重复此过程，即构造出一组基使得 f 在其下的度量矩阵为对角阵。

当 $\mathbb{K} = \mathbb{R}$ 时，存在一组基 S 使得 f 在 S 下的度量矩阵是对角元只有 0、±1 的对角阵；当 $\mathbb{K} = \mathbb{C}$ 时，存在一组基 S 使得 f 在 S 下的度量矩阵是对角元只有 0、1 的对角阵。

证明：

在刚才的证明中，我们已经构造了一组基 v_1, \dots, v_n 使得 f 在这组基下的度量矩阵为 $\text{diag}(a_1, \dots, a_n)$ ，这里 $a_i = f(v_i, v_i)$ 。

若 $\mathbb{K} = \mathbb{R}$ ，我们记

$$u_i = \begin{cases} v_i & a_i = 0 \\ \frac{v_i}{\sqrt{|a_i|}} & a_i \neq 0 \end{cases}$$

直接利用双线性性计算可发现 $f(u_i, u_j)$ 是 $f(v_i, v_j)$ 的倍数，从而 $i \neq j$ 或 $i = j$ 且 $a_i = 0$ 时 $f(u_i, u_j) = 0$ 。其他情况有

$$f(u_i, u_i) = f\left(\frac{v_i}{\sqrt{|a_i|}}, \frac{v_i}{\sqrt{|a_i|}}\right) = \frac{1}{|a_i|} f(v_i, v_i) = \frac{a_i}{|a_i|}$$

从而为 ± 1 。

若 $\mathbb{K} = \mathbb{C}$ ，我们记

$$u_i = \begin{cases} v_i & a_i = 0 \\ \frac{v_i}{b_i} & a_i \neq 0 \end{cases}, \quad b_i^2 = a_i$$

利用复数的性质，当 a_i 非零时存在非零的 b_i 使得 $b_i^2 = a_i$ ，与上类似代入验证可知 $i \neq j$ 或 $i = j$ 且 $a_i = 0$ 时 $f(u_i, u_j) = 0$ ，其他情况 $f(u_i, u_i) = 1$ 。

无论上述哪种构造，每个 u_i 均为对应 v_i 乘非零倍数，除以此倍数即可还原，由此 u_1, \dots, u_n 与 v_1, \dots, v_n 等价，它们仍然构成 V 一组基，得证。

* 更一般的情况无法化简，例如考虑 \mathbb{Q} 上的双线性函数 $f(x, y) = 2xy$ ，其在基 $\{1\}$ 下的度量表示为 (2) ，由于 $\sqrt{2}$ 不是有理数，不可能存在基使得它的度量矩阵为 (0) 或 (± 1) 。

* 由于合同不改变矩阵的秩，可知 $\mathbb{K} = \mathbb{C}$ 时 1 的个数唯一，即为 f 的矩阵秩；利用上学期的矩阵论知识，我们还可以证明 $\mathbb{K} = \mathbb{R}$ 时 1、-1 的个数均唯一，称为 f 的正惯性指数与负惯性指数。

对 \mathbb{K} 上的 n 维线性空间 V ，若 f 是 V 上的斜对称双线性函数，存在一组基 S 使得 f 在基 S 下的度量矩阵为

$$\text{diag}\left(\left(\begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array}\right), \dots, \left(\begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array}\right), 0, \dots, 0\right)$$

* 也即其为分块对角阵，每个对角块或为上述形式的二阶块，或为 0。

证明：

* 此证明与上学期的矩阵论论证法几乎一致。

- 与对称阵类似，我们只需要证明以下结论：当 f 不为零映射时，存在 V 的一组基 v_1, \dots, v_n 使得 $f(v_1, v_2) = 1$ ，且 $f(v_i, v_j) = 0$ 对 $i \leq 2, j > 3$ 成立。此时，再利用斜对称性可发现 $f(v_1, v_1) = f(v_2, v_2) = 0, f(v_2, v_1) = -1$ ，从而其在这组基下的度量矩阵为

$$\text{diag}\left(\left(\begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array}\right), A_0\right)$$

这里 $A_0 \in \mathbb{K}^{(n-2) \times (n-2)}$ ，再由矩阵论或空间理论重复上述过程即可得到结果。

- 仍然与对称阵类似，我们需要构造性质较一般的一组基 u_1, \dots, u_n ，以此为起点进行更进一步刻画。

首先，由于 f 为斜对称双线性函数，对任何线性相关的 α, β （也即它们构成倍数关系），不妨设 $\beta = \lambda\alpha$ ，有

$$f(\alpha, \beta) = f(\alpha, \lambda\alpha) = \lambda f(\alpha, \alpha) = 0$$

由此，若 f 不为零映射，一定存在线性无关的 α, β 使得 $f(\alpha, \beta) \neq 0$ （由于任何两个向量或线性无关或线性相关，若否可得 f 在任何两个向量上均为 0）。

设 $u_1 = \alpha, u_2 = \beta$ ，扩充为 V 的一组基 u_1, \dots, u_n 。

- 最后，我们 u_1, \dots, u_n 出发利用线性组合构造 v_1, \dots, v_n 。

我们记

$$v_1 = \frac{1}{f(u_1, u_2)} u_1, \quad v_2 = u_2$$

由于 $f(v_1, v_2)$ 非零，此定义合理，且类似之前考虑等价性可知 $v_1, v_2, u_3, \dots, u_n$ 构成 V 的一组基，且计算可知

$$f(v_1, v_2) = 1$$

接下来，类似对称阵，我们希望从 u_3, \dots, u_n 中去除 v_1, v_2 的部分，使得正交性（即为 0 的要求）成立。

设

$$\forall i = 3, \dots, n, \quad v_i = u_i + f(v_2, u_i)v_1 - f(v_1, u_i)v_2$$

此时，利用斜对称性 $f(v_1, v_1) = f(v_2, v_2) = 0, f(v_2, v_1) = -1$ ，可进一步计算得

$$f(v_1, v_i) = f(v_1, u_i) + f(v_2, u_i)f(v_1, v_1) - f(v_1, u_i)f(v_1, v_2) = 0$$

$$f(v_2, v_i) = f(v_2, u_i) + f(v_2, u_i)f(v_2, v_1) - f(v_1, u_i)f(v_2, v_2) = 0$$

利用等价性可知 v_1, \dots, v_n 为 V 一组基，于是符合要求，结合之前讨论得证。

* 这里 v_3, \dots, v_n 的构造事实上是先待定系数 $v_i = u_i + av_1 + bv_2$ ，再配凑出来的。

* 对于更一般的情况，事实上只要 A 不是斜对称方阵，一定存在可逆矩阵 P 使得 $P^T AP$ 是上三角阵，从而有对应的取基方式。不过，这个合同三角化结论并不像相似三角化那么重要（根本没见任何地方用过），这里仅作介绍。

24.3.3 补与补空间

最后，我们讨论一些关于左补、右补——在本讲义 24.1.2 已经证明了其为子空间——的定理。有两个较为自然的关注点：左补何时等于右补，左/右补何时构成补空间。此外，子空间交与和的左/右补性质也值得关注。我们仍然假设 V 是 \mathbb{K} 上的线性空间， f 是 V 上的双线性函数，下方的 \perp 均表示 \perp_f 。

首先，关于左补等于右补，有结论： f 是对称或斜对称双线性函数当且仅当对 V 的任何子空间 W ，有 ${}^\perp W = W^\perp$ 。

证明：

我们已经证明了 f 对称或斜对称时 ${}^\perp W = W^\perp$ 对任何子集 W 成立，从而只需从对任何子空间 W 有 ${}^\perp W = W^\perp$ 推出 f 对称或斜对称。证明分为四步。

- 对任何子空间 W 有 ${}^\perp W = W^\perp$ 可推出 f 与 f^T 零点集合相同。

* 利用 f^T 的定义，这当且仅当对任何 $x, y \in V$ 有 $f(x, y) = 0 \Leftrightarrow f(y, x) = 0$ 。

若 $f(\alpha, \beta) = 0$ ，将其写为 $\beta \in \{\alpha\}^\perp$ ，利用右补的性质可知 $\beta \in \langle \alpha \rangle^\perp$ 。由于 $\langle \alpha \rangle$ 为子空间可得 $\langle \alpha \rangle^\perp = {}^\perp \langle \alpha \rangle$ 可知 $\beta \in {}^\perp \langle \alpha \rangle$ ，从而根据定义

$$f(\beta, \alpha) = 0$$

同理，从 $f(\beta, \alpha) = 0$ 可推出 $f(\alpha, \beta) = 0$ ，这即说明

$$\forall x, y \in V, \quad f(x, y) = 0 \Leftrightarrow f(y, x) = 0$$

- 若 $f(x, x) = 0$ 对任何 $x \in V$ 成立，则利用之前已证可知 f 斜对称，从而已经符合要求，下面设存在 z 使得 $f(z, z) \neq 0$ ，我们证明 f 对称。

- 先证明对任何 $y \in V$ 有 $f(y, z) = f(z, y)$ 。

我们与之前思路类似，利用 $f(z, z) \neq 0$ 设法消除 y 中 z 的部分，定义

$$y' = y - \frac{f(y, z)}{f(z, z)}z$$

则

$$f(y', z) = f(y, z) - \frac{f(y, z)}{f(z, z)} f(z, z) = 0$$

于是由条件 $f(z, y') = 0$, 将 y' 展开利用双线性性得到

$$f(z, y) - \frac{f(y, z)}{f(z, z)} f(z, z) = 0$$

这即证明了 $f(z, y) = f(y, z)$ 。

- 最后证明对任何 $x, y \in V$ 有 $f(x, y) = f(y, x)$ 。我们进一步分为三种情况：

1. 若 $f(x, z) \neq 0$, 定义

$$y' = y - \frac{f(x, y)}{f(x, z)} z$$

则

$$f(x, y') = f(x, y) - \frac{f(x, y)}{f(x, z)} f(x, z) = 0$$

于是由条件 $f(y', x) = 0$, 将 y' 展开利用双线性性得到

$$f(y, x) - \frac{f(x, y)}{f(x, z)} f(z, x) = 0$$

再结合 $f(x, z) = f(z, x)$ 即证明了 $f(y, x) = f(x, y)$ 。

2. 若 $f(z, y) \neq 0$, 同理定义

$$x' = x - \frac{f(x, y)}{f(z, y)} z$$

证明过程与上方相同。

3. 若 $f(x, z) = f(z, y) = 0$, 定义

$$x' = x + z, \quad y' = y - \frac{f(x, y)}{f(z, z)} z$$

则利用 $f(x, z) = f(z, y) = 0$ 直接展开可得

$$f(x', y') = f(x, y) - \frac{f(x, y)}{f(z, z)} f(z, z) = 0$$

于是由条件 $f(y', x') = 0$ 。进一步由条件可知 $f(y, z) = f(z, x) = 0$, 从而展开 x' 、 y' 利用双线性性得到

$$f(y, x) - \frac{f(x, y)}{f(z, z)} f(z, z) = 0$$

这即证明了 $f(x, y) = f(y, x)$ 。

* 这个证明的思路很值得细究：我们先通过等价定义排除一种情况，再反复通过线性组合将非零点移至零点进行说明，这是零点相关条件的一般处理方式。具体将哪个变量进行何种平移具有一定的技巧性，一般情况将在之后复习题中详细研究。

接下来，我们假设 V 维数有限， $\dim V = n$ 。我们来研究何时对 V 的任何子空间 W , W^\perp 是 W 的补空间。对于有限维且无特殊条件的问题，我们往往需要借助度量矩阵进行操作。我们分为以下部分进行研究：

1. 当且仅当 f 非退化时，对 V 的任何子空间 W 有

$$\dim W + \dim W^\perp = n$$

证明：

- 设 V 的一组基 S 下 f 的度量矩阵为 A , 则根据度量矩阵定义对任何 $\alpha, \beta \in V$ 有

$$f(\alpha, \beta) = \alpha_S^T A \beta_S$$

利用坐标的同构性质, 假设 W 的一组基为 $\alpha_1, \dots, \alpha_r$, 则它们的坐标 w_1, \dots, w_r 构成 W 中所有向量坐标构成的 \mathbb{K}^n 子空间的一组基。

由此, $\beta \in W^\perp$ 当且仅当

$$w_1^T A \beta_S = w_2^T A \beta_S = \dots = w_r^T A \beta_S = 0$$

将这些方程拼成整体的关于 β_S 线性方程组 $B^T A \beta_S = 0$, 这里 $B = (w_1, \dots, w_r)$ 。再次利用同构性, 其解空间维数应等于 W^\perp 的维数, 从而

$$\dim W^\perp = n - \text{rank}(W^T A)$$

若 f 非退化, 之前已证明 A 可逆, 由乘可逆阵不改变秩可知 $\text{rank}(W^T A) = \text{rank } W^T = \text{rank } W$, 而根据 B 的定义考虑列秩有 $\text{rank } B = \dim W$, 从而得证。

* 证明中我们大量运用了坐标的同构性, 一个更简单的考虑方法是, 坐标的同构性即是说, **直接假设 V 为 \mathbb{K}^n** , 得到的结论并无差别。

- 反之, 若 f 退化, 取 $W = V$, 利用退化定义可知 $\dim V^\perp \neq 0$, 而 $\dim V = n$, 于是 $\dim V + \dim V^\perp > n$, 矛盾。

同理, 当且仅当 f 非退化时, ${}^\perp W$ 与 W 维数和等于全空间维数。

- 在 f 非退化时, 我们只考虑 V 是 \mathbb{R} 上的线性空间的情况。设 f 在 V 的某组基下的度量矩阵为 A 。当且仅当 $A + A^T$ 正定或负定时, 对 V 的任何子空间 W 有 $W \cap W^\perp = \{0\}$, 从而结合上一部分有

$$W \oplus W^\perp = V$$

证明:

- 先证明对 V 的任何子空间 W 有 $W \cap W^\perp = \{0\}$ 等价于对任何 $x \neq 0$ 有 $f(x, x) \neq 0$ 。
 - 若 $x \neq 0$ 且 $f(x, x) = 0$, 取 $W = \langle x \rangle$, 由条件 $x \in W^\perp$, 从而 $x \in W \cap W^\perp$ 。
 - 若 $x \neq 0$ 且存在 W 使得 $x \in W \cap W^\perp$, 由 $x \in W^\perp$ 可知对任何 $w \in W$ 有 $f(w, x) = 0$, 再由 $x \in W$ 即得 $f(x, x) = 0$ 。
- 再证明对任何 $x \neq 0$ 有 $f(x, x) \neq 0$ 当且仅当其恒大于 0 或小于 0。
 - 若 $x, y \in V$ 使得 $f(x, x) > 0$ 、 $f(y, y) < 0$, 我们来证明存在非零的 $z \in V$ 使得 $f(z, z) = 0$, 从而矛盾。
 - 首先, 若 x, y 线性相关, 由于它们都非零可设 $y = \lambda x$, 于是 $f(y, y) = \lambda^2 f(x, x)$, 两者不可能异号, 矛盾。
 - 设 $z = \lambda x + (1 - \lambda)y$, 展开计算可得

$$f(z, z) = \lambda^2 f(x, x) + (1 - \lambda)^2 f(y, y) + \lambda(1 - \lambda)(f(x, y) + f(y, x))$$

由于其是关于 λ 的二次函数, 必然连续, 且 $\lambda = 0$ 时为 $f(y, y)$, $\lambda = 1$ 时为 $f(x, x)$, 在其中存在零点。由于 x, y 线性无关, 且 λ 与 $1 - \lambda$ 不可能同时为 0, 这样得到的 z 不为零向量, 即得证。

- 我们证明恒大于 0 的情况对应 $A + A^T$ 正定, 恒小于 0 对于负定可同理证明。

利用度量矩阵的定义, 设这组基为 S , 对任何 $x \in V$ 且 $x \neq 0$ 有

$$f(x, x) = x_S^T A x_S > 0$$

由于 x 为 0 当且仅当 x_S 为零向量，再利用同构性质，这即是对任何 $w \in \mathbb{R}^n$ 且 $w \neq 0$ 有

$$w^T A w > 0$$

由于 $w^T A w$ 是一阶方阵，等于其转置，上式成立当且仅当任意非零 $w \in \mathbb{R}^n$ 有

$$w^T (A + A^T) w = w^T A w + (w^T A w)^T > 0$$

直接验证得 $A + A^T$ 对称，从而这即是 $A + A^T$ 正定的定义。

* 我们将满足 $f(x, x) > 0$ 对非零 x 恒成立的 f 称为正定，上文的讨论中已经得到有限维实线性空间中 f 正定当且仅当 $A + A^T$ 正定。下一章将讨论其更多性质。

* 对一般情况未必成立，例如在 \mathbb{Q}^2 上定义

$$f((x_1, y_1)^T, (x_2, y_2)^T) = x_1 x_2 - 2y_1 y_2$$

其在标准基下的度量矩阵并不正定，但不存在有理向量 q 使得 $f(q, q) = 0$ ，否则 $\sqrt{2}$ 将为有理数。进一步可验证其对任何 \mathbb{Q}^2 的子空间 W 的确满足 $W \oplus W^\perp = \mathbb{Q}^2$ 。

* 当 V 是 \mathbb{C} 上至少二维的线性空间时，总存在子空间 W 使得 W^\perp 不是 W 的补空间，证明可利用本讲义 25.1 第 3 题 (2)。

最后，利用上面的讨论还可以证明，设 $\dim V = n$ ，当 f 非退化时， $U_1^\perp + U_2^\perp = (U_1 \cap U_2)^\perp$ 对任何 V 的子空间 U_1, U_2 成立，左补有完全对称的结论。

证明：

由于本讲义 24.1.2 已经证明了 $U_1^\perp + U_2^\perp \subset (U_1 \cap U_2)^\perp$ ，只需证明维数相等即可。

利用和空间维数定理直接计算，并由非退化时补的维数性质有

$$\dim(U_1^\perp + U_2^\perp) = \dim U_1^\perp + \dim U_2^\perp - \dim(U_1^\perp \cap U_2^\perp) = 2n - \dim U_1 - \dim U_2 - \dim(U_1^\perp \cap U_2^\perp)$$

再由本讲义 24.1.2 已证的 $U_1^\perp \cap U_2^\perp = (U_1 + U_2)^\perp$ ，进一步计算有

$$\dim(U_1^\perp + U_2^\perp) = n - \dim U_1 - \dim U_2 + \dim(U_1 + U_2) = n - \dim U_1 \cap U_2$$

从而左侧维数与右侧相等，得证。

当 f 退化时，上式未必不成立，如考虑 f 为零映射，则 U^\perp 恒为 V ，结论仍然成立。不过，只要 f 退化且不为零映射，上式的确不可能对任何 V 的子空间都成立。

证明：

由 f 不为零映射，存在 $x \in V$ 使得 $\{x\}^\perp \neq V$ ，设 $U_1 = \langle x \rangle$ 。另一方面，存在非零的 $z \in {}^\perp V$ ，设 $U_2 = \langle x + z \rangle$ 。

利用定义可知 $f(z, v) = 0$ 对任何 $v \in V$ 成立，从而 $f(x + z, v) = 0$ 当且仅当 $f(x, v) = 0$ ，根据补的性质与和空间定义即有

$$U_2^\perp = U_1^\perp = \{x\}^\perp \neq V$$

$$U_2^\perp + U_1^\perp = \{x\}^\perp \neq V$$

另一方面， x 与 z 不可能线性相关，否则根据补为子空间必然有 $x \in {}^\perp V$ ，从而 $\{x\}^\perp = V$ ，矛盾。因此， $U_1 \cap U_2 = \{0\}$ ，进一步得到

$$(U_1 \cap U_2)^\perp = V$$

这就得到了矛盾。

* 这个反例构造相对技巧性，可以从特例中尝试思路。

二十五 实内积空间

§25.1 习题解答

1. (丘书 习题 10.1.1) 在 \mathbb{K}^4 上, 定义双线性函数

$$f((x_1, x_2, x_3, x_4)^T, (y_1, y_2, y_3, y_4)^T) = x_1y_2 - 2x_2y_1 + x_3y_4 - 3x_4y_2$$

求 f 在基

$$\alpha_1 = (1, 2, 1, 1)^T, \quad \alpha_2 = (2, 3, 1, 0)^T, \quad \alpha_3 = (3, 1, 1, -2)^T, \quad \alpha_4 = (4, 2, -1, -6)^T$$

下的度量矩阵。

解答:

可以直接由定义计算所有 $f(\alpha_i, \alpha_j)$, 另一个做法(个人觉得并没有更简单)是先由形式直接得到 f 在基 e_1, e_2, e_3, e_4 下的度量矩阵是

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

再由标准基到基 $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ 的过渡矩阵即为 $P = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ (将每个基作为列拼成的矩阵), 得到结果为

$$P^T AP = \begin{pmatrix} -7 & -14 & -16 & -26 \\ -1 & -6 & -18 & -26 \\ 17 & 23 & 1 & 4 \\ 39 & 58 & 12 & 34 \end{pmatrix}$$

2. (丘书 习题 10.1.3) 设 $A \in \mathbb{F}^{m \times m}$, 对 $V = \mathbb{F}^{m \times n}$, 定义 $V \times V$ 到 \mathbb{F} 的映射 f 满足

$$f(G, H) = \text{tr}(G^T AH)$$

(1) 证明 f 是 V 上的双线性函数。

证明:

对 $G_1, G_2, H_1, H_2, G, H \in V$ 、 $\lambda, \mu \in \mathbb{F}$, 利用 tr 的线性性与 $(\lambda G_1 + \mu G_2)^T = \lambda G_1^T + \mu G_2^T$ 直接计算可知

$$f(\lambda G_1 + \mu G_2, H) = \lambda \text{tr}(G_1^T AH) + \mu \text{tr}(G_2^T AH) = \lambda f(G_1, H) + \mu f(G_2, H)$$

$$f(G, \lambda H_1 + \mu H_2) = \lambda \text{tr}(G^T AH_1) + \mu \text{tr}(G^T AH_2) = \lambda f(G, H_1) + \mu f(G, H_2)$$

从而得证双线性性。

(2) 求 f 在基 $E_{11}, E_{12}, \dots, E_{1n}, \dots, E_{m1}, \dots, E_{mn}$ 下的度量矩阵。

解答:

直接利用矩阵乘法定义计算可知 $E_{ij}^T AE_{kl}$ 只有第 j 行第 l 列为 a_{ik} , 其他元素为 0, 从而可知

$$\text{tr}(E_{ij}^T AE_{kl}) = \begin{cases} 0 & j \neq l \\ a_{ik} & j = l \end{cases}$$

取较小的 m, n 尝试计算即可发现结果形式为

$$\begin{pmatrix} a_{11}I_n & \cdots & a_{1m}I_n \\ \vdots & \ddots & \vdots \\ a_{m1}I_n & \cdots & a_{mm}I_n \end{pmatrix}$$

或利用之后定义的矩阵克罗内克积写为 $A \otimes I_n$ 。

3. (丘书习题 10.1.4) 设 V 是 \mathbb{C} 上的 n 维线性空间, $n \geq 2$, f 是 V 上的对称双线性函数。

- (1) 证明存在 V 的一组基 $\delta_1, \dots, \delta_n$, 使得这组基下 f 的度量矩阵为

$$\text{diag}(1, \dots, 1, 0, \dots, 0)$$

证明:

见本讲义 24.3.2 倒数第二个证明。

- (2) 证明 V 中存在非零向量 ξ 使得 $f(\xi, \xi) = 0$ 。

证明:

设 f 的矩阵秩为 r 。当 $r = 1$ 时, 取 $\xi = \delta_2$ 利用度量矩阵定义即得 $f(\delta_2, \delta_2) = 0$; 否则, 取 $\xi = \delta_1 + i\delta_2$, 有

$$f(\xi, \xi) = f(\delta_1, \delta_1) - f(\delta_2, \delta_2) + i(f(\delta_1, \delta_2) + f(\delta_2, \delta_1))$$

利用度量矩阵定义可知其为 $1 - 1 + i(0 + 0) = 0$ 。

- (3) 若 f 非退化, 存在线性无关的向量 ξ, η , 使得

$$f(\xi, \eta) = 1, \quad f(\xi, \xi) = f(\eta, \eta) = 0$$

证明:

由非退化性, 矩阵秩 $r = n > 1$ 。取

$$\xi = \frac{\sqrt{2}}{2}(\delta_1 + i\delta_2), \quad \eta = \frac{\sqrt{2}}{2}(\delta_1 - i\delta_2)$$

与前一问完全相同计算可知 $f(\xi, \eta) = 1$, 且 $f(\xi, \xi) = f(\eta, \eta) = 0$ 。

由于

$$\delta_1 = \frac{\sqrt{2}}{2}(\xi + \eta), \quad \delta_2 = \frac{\sqrt{2}}{2i}(\xi - \eta)$$

δ_1, δ_2 与 η, ξ 等价, 从而由 δ_1, δ_2 线性无关可知 η, ξ 线性无关, 得证。

4. 对于 \mathbb{K} 上线性空间 V 上的双线性函数 f , 若对任何 $\alpha, \beta \in V$, $f(\alpha, \beta) = 0$ 当且仅当 $f(\beta, \alpha) = 0$, 则 f 对称或斜对称。

证明:

见本讲义 24.3.3 第一个证明的后三步。

5. (丘书习题 10.1.9) 设 V 是 n 维实线性空间, Q_1, Q_2 是 V 上两个二次函数, 证明, 若 Q_1, Q_2 分别再 V 的一个基下的表达式的正惯性指数都小于 $\frac{n}{2}$, 那么 $Q_1 + Q_2$ 在 V 的一个基下的表达式不是正定的。

证明:

我们假设 Q_1 (对应的对称双线性函数 f_1) 在基 $\alpha_1, \dots, \alpha_n$ 下取到合同规范形

$$\text{diag}(I_{p_1}, -I_{q_1}, O_{n-p_1-q_1})$$

Q_2 (对应的对称双线性函数 f_2) 在基 β_1, \dots, β_n 下取到合同规范形

$$\text{diag}(I_{p_2}, -I_{q_2}, O_{n-p_2-q_2})$$

设 $W_1 = \langle \alpha_{p_1+1}, \dots, \alpha_n \rangle$, 我们先证明对任何 $w \in W_1$ 有 $Q_1(w) \leq 0$ 。

设

$$w = \sum_{i=p_1+1}^n \lambda_i \alpha_i$$

则根据 Q_1 的表达式即为 f_1 的度量矩阵, 进一步由定义计算得 (最后一个等号是由于 $f(\alpha_i, \alpha_j)$ 在 $i \neq j$ 时为 0, 否则为度量矩阵对应对角元, 即 -1 或 0)

$$Q_1(w) = f_1(w, w) = \sum_{i=p_1+1}^n \sum_{j=p_1+1}^n \lambda_i \lambda_j f(\alpha_i, \alpha_j) = - \sum_{i=p_1+1}^{p_1+q_1} \lambda_i^2 \leq 0$$

同理, 设 $W_2 = \langle \beta_{p_2+1}, \dots, \beta_n \rangle$, 对任何 $w \in W_2$ 有 $Q_2(w) \leq 0$ 。

由于 $\dim W_1 = n - p_1$ 、 $\dim W_2 = n - p_2$, 利用条件 $p_1 < \frac{n}{2}$ 、 $p_2 < \frac{n}{2}$ 可得

$$\dim W_1 \cap W_2 = \dim W_1 + \dim W_2 - \dim W_1 + W_2 \geq \dim W_1 + \dim W_2 - n > 0$$

从而存在非零向量 $w \in W_1 \cap W_2$, 此时

$$Q_1(w) + Q_2(w) \leq 0$$

与正定定义矛盾, 即得证。

6. (丘书习题 10.1.10) 判断 \mathbb{K} 上下列两个斜对称矩阵是否合同:

$$A = \begin{pmatrix} 0 & 2 & 1 & -3 \\ -2 & 0 & 4 & 5 \\ -1 & -4 & 0 & -1 \\ 3 & -5 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & -4 & -1 \\ -1 & 0 & 3 & -2 \\ 4 & -3 & 0 & 11 \\ 1 & 2 & -11 & 0 \end{pmatrix}$$

解答:

直接计算可知 $\text{rank } A = 4$ 、 $\text{rank } B = 2$, 从而不可能合同。

* 注意利用斜对称阵的合同规范形结论, 斜对称阵秩为偶数, 且合同等价于秩相等。

7. (丘书习题 10.1.12) 判断以下两个实对称矩阵是否可以同时合同对角化:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

解答:

利用教材 10.1 节例 20 的结论, 直接计算

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

其特征值为 $\pm i$, 无实特征值, 因此看作实方阵不可对角化, 于是 A 、 B 不可同时合同对角化。

8. (丘书习题 10.2.17) 在实内积空间 $C[0, 2\pi]$ ($[0, 2\pi]$ 上所有连续函数) 中, 指定的内积为

$$(f, g) = \int_0^{2\pi} f(x)g(x)dx$$

证明 $C[0, 2\pi]$ 的子集

$$S = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos(nx), \frac{1}{\sqrt{\pi}} \sin(nx) \mid n \in \mathbb{N}^+ \right\}$$

是正交规范集 (即每个向量均为单位向量且相互正交)。

证明:

利用三角函数的和差化积公式可以得到

$$\begin{aligned} \cos(mx)\cos(nx) &= \frac{1}{2}(\cos((m-n)x) + \cos((m+n)x)) \\ \sin(mx)\sin(nx) &= \frac{1}{2}(\cos((m-n)x) - \cos((m+n)x)) \\ \sin(mx)\cos(nx) &= \frac{1}{2}(\sin((m+n)x) + \sin((m-n)x)) \end{aligned}$$

利用三角函数的周期性, 在 k 为自然数时, $\cos(kx)$ 与 $\sin(kx)$ 当且仅当 $k = 0$ 时在 $[0, 2\pi]$ 上积分分为 2π , 否则为 0, 从而利用上述公式得到 (这里 δ_{mn} 当 $m = n$ 时为 1, 否则为 0)

$$\begin{aligned} \forall m \in \mathbb{N}^+, \quad \forall n \in \mathbb{N}, \quad \int_0^{2\pi} \sin(mx)\cos(nx)dx &= 0 \\ \forall m, n \in \mathbb{N}^+, \quad \int_0^{2\pi} \cos(mx)\cos(nx)dx &= \delta_{mn}\pi \\ \forall m, n \in \mathbb{N}^+, \quad \int_0^{2\pi} \sin(mx)\sin(nx)dx &= \delta_{mn}\pi \\ \int_0^{2\pi} 1dx &= 2\pi \end{aligned}$$

再对应乘倍数即可验证其为正交规范集。

9. (丘书习题 10.2.18) 设 $\alpha_1, \dots, \alpha_n$ 是 n 维欧氏空间 V 中的非零向量组, 证明

$$|\mathbf{G}(\alpha_1, \dots, \alpha_n)|^2 \leq \|\alpha_1\|^2 \|\alpha_2\|^2 \dots \|\alpha_n\|^2$$

且等号成立当且仅当 $\alpha_1, \dots, \alpha_n$ 相互正交。

证明:

由教材 10.2 节例 10, 当且仅当 $\alpha_1, \dots, \alpha_n$ 线性相关时左侧为 0, 此时由此向量组非零可知右侧非零, 小于号严格成立。

下面考虑 $\alpha_1, \dots, \alpha_n$ 线性无关的情况。此时利用教材 10.2 节例 11, 设 β_1, \dots, β_n 为 $\alpha_1, \dots, \alpha_n$ 的正交化 (不进行标准化) 结果, 有

$$|\mathbf{G}(\alpha_1, \dots, \alpha_n)|^2 = \|\beta_1\|^2 \|\beta_2\|^2 \dots \|\beta_n\|^2$$

另一方面, 直接由定义与 β_1, \dots, β_n 正交性计算可得

$$(\beta_i, \beta_i) = \left(\alpha_i - \sum_{j=1}^{i-1} \frac{(\alpha_i, \beta_j)}{(\beta_j, \beta_j)} \beta_j, \beta_i \right) = (\alpha_i, \beta_i) = (\alpha_i, \alpha_i) - \sum_{j=1}^{i-1} \frac{(\alpha_i, \beta_j)^2}{(\beta_j, \beta_j)}$$

由此可发现 $\|\beta_i\|^2 \leq \|\alpha_i\|^2$, 从而不等号成立。

另一方面, 由于已知所有 $\|\alpha_i\|$ 非零, 等号成立当且仅当 $\alpha_1, \dots, \alpha_n$ 线性无关且每个 $\|\beta_i\| = \|\alpha_i\|$, 根据上式也即得到

$$\forall j < i, \quad (\alpha_i, \beta_j) = 0$$

由于 $\langle \beta_1, \dots, \beta_{i-1} \rangle = \langle \alpha_1, \dots, \alpha_{i-1} \rangle$, 此式等价于所有 α_i 相互正交, 即得证。

10. (丘书 习题 10.5.1) 在指定标准内积的酉空间 \mathbb{C}^3 中, 设

$$\alpha = (1, -1, 1)^T, \quad \beta = (1, 0, i)^T$$

求 $\|\alpha\|$ 、 $\|\beta\|$, α 与 β 的夹角 $\langle \alpha, \beta \rangle$ 。

解答:

$$\begin{aligned}\|\alpha\| &= \sqrt{\alpha^H \alpha} = \sqrt{3}, \quad \|\beta\| = \sqrt{\beta^H \beta} = \sqrt{2} \\ \langle \alpha, \beta \rangle &= \arccos \frac{|\beta^H \alpha|}{\|\alpha\| \|\beta\|} = \arccos \frac{\sqrt{3}}{3}\end{aligned}$$

11. (丘书 习题 10.5.4) 在指定标准内积的酉空间 \mathbb{C}^3 中, 设

$$\alpha_1 = (1, -1, i)^T, \quad \alpha_2 = (1, 0, i)^T, \quad \alpha_3 = (1, 1, 1)^T$$

求其一组标准正交基。

解答:

对 $\alpha_1, \alpha_2, \alpha_3$ 进行 Schmidt 正交化可得到标准正交基

$$\eta_1 = \frac{1}{\sqrt{3}}(1, -1, i)^T, \quad \eta_2 = \frac{\sqrt{6}}{6}(1, 2, i)^T, \quad \eta_3 = \frac{1}{2}(1 + i, 0, 1 - i)^T$$

* 由于本题并未要求用正交化, 其实直接写 e_1, e_2, e_3 也是可以的...

12. (丘书 习题 10.5.6) 在 $\mathbb{C}^{n \times n}$ 中, 指定内积为

$$(A, B) = \text{tr}(AB^H)$$

设 W 是所有 n 阶复对角矩阵构成的子空间, 求 W^\perp 及其标准正交基。

解答:

这事实上是 $\mathbb{C}^{n \times n}$ 中的标准内积定义, 直接计算验证可发现

$$\{E_{ij} \mid i = 1, \dots, n, \quad j = 1, \dots, n\}$$

相互正交, 且每个模长都是 1, 考虑到个数等于维数即得其构成此内积下的标准正交基。

进一步计算可得 $W = \langle E_{11}, \dots, E_{nn} \rangle$, 从而利用正交补的算法 (可见本讲义 25.2.2) 即得

$$W^\perp = \langle E_{ij} \mid i \neq j \rangle$$

所有满足 $i \neq j$ 的 E_{ij} 即构成 W^\perp 的标准正交基。

§25.2 内积

25.2.1 实内积空间

从本章起, 我们将开始学习内积空间的知识。首先需要关注的即是实内积空间: 在实线性空间 V 中, 若对称的双线性函数 f 满足正定性要求

$$\forall x \in V, x \neq 0, \quad f(x, x) > 0$$

则称其为一个内积。带有内积的实线性空间称为实内积空间, 我们将 $f(x, y)$ 简记为 (x, y) 。

* 我们将在本节的最后通过长度与夹角展示为何要如此定义内积。

对称性已经在之前的章节中研究过, 因此这里我们需要给出正定的判定与性质, 这里列举如下, 我们假设 f 是 V 上的对称双线性函数:

1. 若 f 正定，则其非退化，且其在 V 任何子空间上的限制都对称正定（从而非退化）。

证明：

由于对任何非零的 x , $f(x, x) > 0$, 不可能对任何 $y \in V$ 有 $f(x, y) = 0$ 或对任何 $y \in V$ 有 $f(y, x) = 0$, 于是其左根、右根都只有零向量，从而非退化。

对任何 V 的子空间 W 上，根据定义有

$$\forall a, b \in W, \quad f|_W(a, b) = f(a, b) = f(b, a) = f|_W(b, a)$$

$$\forall a \in W, a \neq 0, \quad f|_W(a, a) = f(a, a) > 0$$

从而对称正定。

2. 若 f 正定，对 V 的任何子空间 W , ${}^\perp W = W^\perp$, 且 $W \cap W^\perp = \{0\}$ 。

证明：

由对称性可知 ${}^\perp W = W^\perp$, 对任何非零 $x \in W$, 由于 $f(x, x) \neq 0$, 根据定义 $x \notin W^\perp$, 于是

$$W \cap W^\perp = \{0\}$$

3. 若 V 维数有限， f 正定当且仅当其在任意一组基下的度量矩阵正定。

证明：

本讲义 24.3.3 的 $W \cap W^\perp = \{0\}$ 相关讨论中，已经证明了 f 正定当且仅当 $A + A^T$ 正定，又由于 f 对称， $A = A^T$ ，这等价于 $2A$ 正定，利用矩阵论知识也即 A 正定。

4. 若 V 维数有限， f 正定当且仅当其在某组基下的度量矩阵为单位阵 I 。

证明：

利用矩阵论知识， A 正定当且仅当其合同规范形为 I 。由此，利用上个结论，若 f 正定，先取其任何一组基下的度量矩阵（设为 A ），再对应根据本讲义 24.3.2 的讨论进行基变换即可构造一组基使得其度量矩阵为 I 。

另一方面，由于 I 是正定阵，根据上个结论可知若 f 在某组基下度量矩阵为 I ，则其正定。

5. 若 V 维数有限且 f 正定，对 V 的任何子空间 W , W^\perp 是 W 的补空间，即 $W \oplus W^\perp = V$ 。

证明：

由于已经证明了 f 非退化，且满足度量矩阵 A 正定，再由对称性 $A + A^T = 2A$ 正定，根据本讲义 24.3.3 的讨论即得结论。

* V 维数有限时，计算维数可发现，只要 V 的子空间 W_0 与 W 正交（即 $W_0 \subset W^\perp$ ），且 $\dim W_0 + \dim W = \dim V$ ，必有 $W_0 = W^\perp$ 。这也是 W^\perp 的常用判定。

6. 若 V 维数有限且 f 正定，对 V 的任何子空间 W , $(W^\perp)^\perp = W$ 。

证明：

首先，根据 W^\perp 的定义， W 中任何向量都与 W 中任何向量相互正交，于是

$$W \subset (W^\perp)^\perp$$

另一方面，根据上一个性质可知

$$\dim(W^\perp)^\perp = \dim V - \dim W^\perp = \dim V - (\dim V - \dim W) = \dim W$$

从而只能相等。

* 若 f 未必对称, 由于正定性条件只与 $f(x, x)$ 相关, 之前已经介绍过, $g = \frac{f+f^T}{2}$ 是满足 $g(x, x) = f(x, x)$ 处处成立的对称双线性函数, 因此上述讨论可以归为 g 的性质。

* 此后, 在实内积空间中, 我们谈论 \perp 均指内积 f 下的 \perp_f 。由于对称性, 我们只保留 S^\perp 的记号, 不再单独写出 $\perp S$ 。

既然我们将赋予了内积的实线性空间定义为了实内积空间, 我们自然需要像上半学期对线性空间那样研究它作为代数结构的性质。首先是关于子结构的结论: 在实内积空间 V 的子空间 W 上, 定义内积为 V 的内积在 W 上的限制, 则 W 也是实内积空间。

* 此后, 当我们谈论实内积空间的子空间时, 均默认以这种方式定义为实内积空间。

证明:

根据正定对称双线性函数的性质, 内积在 W 上的限制正定对称 (之前已证明双线性函数在子空间上的限制仍为双线性函数), 从而符合内积定义, W 仍构成实内积空间。

由此, 无论是交空间、和空间还是生成子空间, 由于都是子空间的运算, 在实内积空间中都可以正常讨论。

对于商空间, 有如下结论: 对 V 的子空间 W , 若 $V = W \oplus W^\perp$ (之前已经证明, 至少这在有限维时恒成立), 则在 V/W 中定义内积 (注意左侧代表商空间中内积, 右侧代表 V 中内积, 由于不会引起歧义, 用相同记号表示)

$$\forall v_1, v_2 \in W^\perp, \quad (v_1 + W, v_2 + W) = (v_1, v_2)$$

其构成一个实内积空间。

证明:

我们先证明对任何商空间中的元素 $x + W$, 存在唯一 $v \in W^\perp$ 使得 $x + W = v + W$ 。

- 存在性

由直和的性质, 设 $x = w + v$, 其中 $w \in W$ 、 $v \in W^\perp$, 则

$$x + W = w + v + W = v + (w + W) = v + W$$

最后一个等号是由于商空间的等价类性质。

- 唯一性

由于 $W \cap W^\perp = \{0\}$, 若 $v_1, v_2 \in W^\perp$ 且 $v_1 \neq v_2$, 进一步通过 $v_1 - v_2 \neq 0$ 、 $v_1 - v_2 \in W^\perp$ 可知 $v_1 - v_2 \notin W$, 从而由商空间等价类性质 $v_1 + W \neq v_2 + W$, 这就证明了唯一性。

由此, 上述的内积的确对任何两个元素给出了唯一定义, 是 $V/W \times V/W$ 上的函数。可直接利用商空间运算验证其对第一个分量的线性性, 对任何 $v_1, v_2, v \in W^\perp$ 、 $\lambda, \mu \in \mathbb{R}$ 有

$$\begin{aligned} (\lambda(v_1 + W) + \mu(v_2 + W), v + W) &= (\lambda v_1 + \mu v_2 + W, v + W) \\ &= (\lambda v_1 + \mu v_2, v) \\ &= \lambda(v_1, v) + \mu(v_2, v) \\ &= \lambda(v_1 + W, v + W) + \mu(v_2 + W, v + W) \end{aligned}$$

同理对第二个分量也线性, 于是其是双线性函数。

其对称性也可直接验证, 对任何 $v_1, v_2 \in W^\perp$:

$$(v_1 + W, v_2 + W) = (v_1, v_2) = (v_2, v_1) = (v_2 + W, v_1 + W)$$

最后, 由商空间元素与 W^\perp 的唯一确定性, $v \in W^\perp$ 使得 $v + W = W$ 当且仅当 $v = 0$, 于是对任何商空间中的非零元素 $v + W$, $v \in W^\perp$ 、 $v \neq 0$ 有

$$(v + W, v + W) = (v, v) > 0$$

这就证明了正定性。

另一个较有趣的事是实内积空间的积空间(积空间定义见本讲义 24.1.1), 若 V, W 都是实内积空间, 内积分别为 ϕ_V, ϕ_W , 在 $V \times W$ 上定义内积 φ :

$$\varphi((v_1, w_1), (v_2, w_2)) = \phi_V(v_1, v_2) + \phi_W(w_1, w_2)$$

则 $V \times W$ 构成实内积空间。

证明:

- 线性性(只验证第一个分量, 第二个分量类似)

设 $v_1, v_2, v \in V, w_1, w_2, w \in W, \lambda, \mu \in \mathbb{R}$, 有

$$\begin{aligned} \varphi(\lambda(v_1, w_1) + \mu(v_2, w_2), (v, w)) &= \varphi((\lambda v_1 + \mu v_2, \lambda w_1 + \mu w_2), (v, w)) \\ &= \phi_V(\lambda v_1 + \mu v_2, v) + \phi_W(\lambda w_1 + \mu w_2, w) \\ &= \lambda \phi_V(v_1, v) + \mu \phi_V(v_2, v) + \lambda \phi_W(w_1, w) + \mu \phi_W(w_2, w) \\ &= \lambda \varphi((v_1, w_1), (v, w)) + \mu \varphi((v_2, w_2), (v, w)) \end{aligned}$$

从而其为双线性函数。

- 对称性

对任何 $v_1, v_2 \in V, w_1, w_2 \in W$ 有

$$\varphi((v_1, w_1), (v_2, w_2)) = \phi_V(v_1, v_2) + \phi_W(w_1, w_2) = \phi_V(v_2, v_1) + \phi_W(w_2, w_1) = \varphi((v_2, w_2), (v_1, w_1))$$

于是对称性满足。

- 正定性

$V \times W$ 中, 若 (v, w) 为零元, 当且仅当 $v = 0, w = 0$, 于是对任何非零的 (v, w) 有

$$\varphi((v, w), (v, w)) = \phi_V(v, v) + \phi_W(w, w)$$

由于 v 与 w 至多一个为 0, 右侧两项必然一个大于等于 0、一个大于 0, 从而得证。

* 商空间与积空间成为实内积空间的过程不要求掌握, 但可以此作为内积验证和操作的练习。

最后, 我们将看到内积定义的意义。仿照高中学过有了内积以后, 我们希望定义向量的模长为 $\|x\| = \sqrt{(x, x)}$, 正定性即要求所有向量可以度量长度, 且只有零向量长度为 0。两个非零向量的夹角为

$$\Theta(x, y) = \arccos \frac{(x, y)}{\|x\| \|y\|}$$

不过, 这自然产生了一个问题: 是否真的有 $|(x, y)| \leq \|x\| \|y\|$, 使得夹角可以定义? 答案是肯定的, 这称为内积空间的 Cauchy 不等式。

证明:

* 下面这个配凑二次函数的方法是柯西不等式的常用证明思路。

若 $y = 0$, 左右均为 0, 已经得证, 下不妨设 $y \neq 0$, 从而 $(y, y) > 0$ 。对任何 $\lambda \in \mathbb{R}$, 由内积正定性有

$$(x - \lambda y, x - \lambda y) \geq 0$$

展开并利用对称性得到

$$\lambda^2(y, y) - 2\lambda(x, y) + (x, x) \geq 0$$

由于这是一个关于 λ 的二次项非零的二次函数, 直接利用二次函数知识可知其最小值在 $\lambda = \frac{(x, y)}{(y, y)}$ 取到, 代入可得

$$-\frac{(x, y)^2}{(y, y)} + (x, x) \geq 0$$

由于 $(y, y) > 0$ 再两侧同乘 (y, y) 即得

$$(x, x)(y, y) \geq (x, y)^2$$

两侧同作平方根得证原命题。

有了夹角以后，我们自然可以定义两个向量 x, y 正交（若均非零即指夹角为 $\frac{\pi}{2}$ ）当且仅当 $(x, y) = 0$ 。若对 V 的子集（或子空间） S, T ，有

$$\forall s \in S, t \in T, \quad (s, t) = 0$$

则称这两个子集（或子空间）正交。根据补的定义（注意内积下的左右补相等），两个子集正交即等价于 $S \subset T^\perp$ 或 $T \subset S^\perp$ 。此外，对 V 的子空间 W ，根据 W^\perp 的定义，它是所有与 W 中所有向量都正交的向量的集合，在实内积空间中称为 W 的正交补。

事实上，正交是一种更强的无关性：若实内积空间 V 中的一些非零向量 $\alpha_i, i \in I$ （这里 I 为指标集）两两正交，则它们线性无关。

证明：

假设它们线性相关，即存在一些下标 $i_1, \dots, i_n \in I$ 与不全为 0 的 $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ 使得

$$\lambda_1\alpha_{i_1} + \lambda_2\alpha_{i_2} + \dots + \lambda_n\alpha_{i_n} = 0$$

由此有

$$(\lambda_1\alpha_{i_1} + \lambda_2\alpha_{i_2} + \dots + \lambda_n\alpha_{i_n}, \lambda_1\alpha_{i_1} + \lambda_2\alpha_{i_2} + \dots + \lambda_n\alpha_{i_n}) = 0$$

利用 α_i 的两两正交性，将左右侧全部展开，所有下标不同的内积 $(\alpha_{i_j}, \alpha_{i_k}), j \neq k$ 均为 0，于是只剩下

$$\lambda_1^2(\alpha_{i_1}, \alpha_{i_1}) + \lambda_2^2(\alpha_{i_2}, \alpha_{i_2}) + \dots + \lambda_n^2(\alpha_{i_n}, \alpha_{i_n}) = 0$$

但是，由于 α_i 均非零，每项 $(\alpha_{i_k}, \alpha_{i_k}) > 0$ ，若所有 λ_i 不全为 0 可知左侧大于 0，矛盾，这就得到了证明。

* 这里的展开后只保留下标相同项是利用正交性的常见化简。

* 由此，正交性只需要两两成立，即能保证整体的线性无关，但向量组两两线性无关仍然可能整体线性相关，这就体现了正交比一般的线性无关更强。

* 事实上，有了模长后，还可以定义两个向量的距离 $d(x, y) = \|x - y\|$ ，自此就可以研究收敛性、函数极限等分析问题了。

25.2.2 标准正交基

既然正交是更强的线性无关，一个自然的问题是，我们是否可以找到实内积空间的一组两两正交的基？设 V 是实内积空间，先引入定义：两两正交的非零向量组称为 V 中的正交集，两两正交的一组基称为 V 的正交基；两两正交且模长均为 1 的向量组称为 V 中的正交规范集，两两正交且模长均为 1 的一组基称为 V 的标准正交基。

我们可以直接得到以下结论：

1. V 存在正交基与 V 存在标准正交基相互等价。

证明：

若 $\alpha_i, i \in I$ 是 V 的一组正交基，由线性无关性可知均非 0，从而 $\|\alpha_i\| > 0$ ，考虑

$$\frac{1}{\|\alpha_i\|}\alpha_i, i \in I$$

由于这是将每个基放缩了非零倍数，通过乘倍数可还原，由此这与 $\alpha_i, i \in I$ 等价，为 V 的一组基。另一方面， $i, j \in I$ 且 $i \neq j$ 时有

$$\left(\frac{1}{\|\alpha_i\|} \alpha_i, \frac{1}{\|\alpha_j\|} \alpha_j \right) = \frac{1}{\|\alpha_i\| \|\alpha_j\|} (\alpha_i, \alpha_j) = 0$$

且对任何 $i \in I$ 有

$$\left\| \frac{\alpha_i}{\|\alpha_i\|} \right\| = \frac{1}{\|\alpha_i\|^2} (\alpha_i, \alpha_i) = 1$$

从而这的确构成标准正交基。

另一方面，标准正交基是正交基，从而若存在标准正交基也存在正交基。

* 从正交基构造标准正交基的过程称为规范化，由于仅进行了放缩，相对简单，我们之后基本只讨论正交基的构造，需要标准正交基时进行一步规范化即可。

2. 若 V 无限维，其可能不存在标准正交基。

证明：

我们只给出例子，忽略所有证明细节。这里仅作为科普，无需掌握。

考虑所有下标为 0 开始的实数列 (a_k) 构成的线性空间 V ，定义两个数列的和为逐项求和，数列的数乘为逐项相乘。

定义 W 为所有满足平方和极限收敛，即

$$\sum_{k=0}^{\infty} a_k^2 < +\infty$$

的数列构成的集合，可验证它是 V 的一个子空间（也可记作 l^2 ）。

在 W 上定义

$$((a_k), (b_k)) = \sum_{k=0}^{\infty} a_k b_k$$

可验证其构成一个内积，从而 W 成为实内积空间。

W 中的任何一个正交集至多可数，但其维数不可数，从而不存在标准正交基。

* 因此，我们之后基本只讨论有限维实内积空间，忽略无穷维的情况，以保证性质良好。

3. 若 V 维数有限，其一定存在标准正交基。

证明：

之前讨论正定对称双线性函数的性质时已经证明，存在一组基使得内积在这组基下的度量矩阵为 I ，而根据度量矩阵的定义，这即代表这组基相互正交且每个的模长都为 1，从而得证。

4. 若 V 维数有限，其正交基等价于极大正交集。

* 这里极大指正交集 S 满足不存在正交集 T 使得 $S \subsetneq T$ ，即 T 真包含 S 。

证明：

设 $\dim V = n$ 。证明分为两个方向：

- 正交基是极大正交集

若 S 是 V 的正交基，由其为基可知个数为 n ，但由于之前已经证明了正交集是线性无关向量组，不可能有个数大于 n 的正交集，从而不会有真包含 S 的正交集。

- 极大正交集是正交基

若 S 是 V 中的正交集，且不为正交基，由于线性无关性，其个数必然小于 n ，设其为 $\alpha_1, \dots, \alpha_r$ ，且 $r < n$ 。

由于其不为基，存在向量 β 与 $\alpha_1, \dots, \alpha_r$ 线性无关，定义

$$\gamma = \beta - \sum_{i=1}^r \frac{(\alpha_i, \beta)}{(\alpha_i, \alpha_i)} \alpha_i$$

计算验证可发现

$$(\alpha_j, \gamma) = (\alpha_j, \beta) - \sum_{i=1}^r \frac{(\alpha_i, \beta)}{(\alpha_i, \alpha_i)} (\alpha_j, \alpha_i)$$

由于 α_i 之间的正交性， (α_j, α_i) 当且仅当 $i = j$ 时非零，从而上式化简为

$$(\alpha_j, \gamma) = (\alpha_j, \beta) - \frac{(\alpha_j, \beta)}{(\alpha_j, \alpha_j)} (\alpha_j, \alpha_j) = 0$$

由此可知 $\alpha_1, \dots, \alpha_r, \gamma$ 也相互正交。

另一方面，由于 γ 由 β 减掉 $\alpha_1, \dots, \alpha_r$ 的线性组合得到， β 可以由 γ 加 $\alpha_1, \dots, \alpha_r$ 的线性组合得到，因此 $\alpha_1, \dots, \alpha_r, \beta$ 与 $\alpha_1, \dots, \alpha_r, \gamma$ 等价，从而 $\alpha_1, \dots, \alpha_r, \gamma$ 也线性无关，这就可得到 γ 非零，从而 $\alpha_1, \dots, \alpha_r, \gamma$ 是一个正交集。

综合以上讨论， $\alpha_1, \dots, \alpha_r, \gamma$ 是真包含 $\alpha_1, \dots, \alpha_r$ 的正交集，从而 $\alpha_1, \dots, \alpha_r$ 不是极大正交集。考虑逆否命题可得要证的结论。

* 注意我们证明中又用到了去除指定分量的操作与等价性证明技巧。

5. 若 V 维数有限，其标准正交基等价于极大正交规范集。

证明：

利用第一个结论中的规范化操作，任何正交集均可规范化得到正交规范集，任何正交基均可规范化得到标准正交基，从而结合上一个结论得证。

从最后两个命题中可以看到，对有限维实内积空间 V (可以称为欧氏空间)，任给 V 中的正交集，我们都可以扩充出 V 的一组正交基。现在我们给出具体的算法。假设 $\dim V = n$ ，已知 V 中的正交集 $\alpha_1, \dots, \alpha_r$ (可以 $r = 0$ ，这时初始的正交集是空集，相当于从头构造 V 的正交基)。

解答：

在第四个性质的证明中，我们已经给出了部分的算法，现在我们将其完善。

由于 $\alpha_1, \dots, \alpha_r$ 相互正交，它们线性无关，从而记

$$W = \langle \alpha_1, \dots, \alpha_r \rangle$$

它们构成 W 的正交基。

利用本讲义 18.3.2 的算法，可以构造 W 的补空间的一组基 $\beta_{r+1}, \dots, \beta_n$ ，则 $\alpha_1, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_n$ 构成 V 的一组基。

接下来，我们从这组基出发构造正交基：我们对 $k = r + 1$ 到 n 依次进行操作：

$$\alpha_k = \beta_k - \sum_{i=1}^{k-1} \frac{(\alpha_i, \beta_k)}{(\alpha_i, \alpha_i)} \alpha_i$$

每次操作后，我们用 α_k 替换对应 β_k 。根据之前的证明，每次操作后得到的 α_k 都与之前的 $\alpha_1, \dots, \alpha_{k-1}$ 相互正交，且 $\alpha_1, \dots, \alpha_k$ 与 $\alpha_1, \dots, \alpha_{k-1}, \beta_k$ 等价。利用等价性的定义，等价的向量组增添一些相同的向量也等价，于是 $\alpha_1, \dots, \alpha_k, \beta_{k+1}, \dots, \beta_n$ 与 $\alpha_1, \dots, \alpha_{k-1}, \beta_k, \dots, \beta_n$ 等价。由此，进行完 $n - r$ 次操作后，即可得到最终的 $\alpha_1, \dots, \alpha_n$ 与 $\alpha_1, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_n$ 等价，为 V 的一组基，且相互正交。这就构造出了 V 的一组正交基。

* 此算法必须掌握, 称为 Schmidt 正交化。若需要标准正交基, 再进行规范化即可。有时候我们也将 Schmidt 正交化与规范化的过程合称为 Schmidt 正交化。

此算法事实上也给出了正交补除了定义外的另一种计算方式: 若上述 $\alpha_1, \dots, \alpha_r$ 是 V 的子空间 W 的正交基, 扩充出的基为 $\alpha_{r+1}, \dots, \alpha_n$, 则 $\alpha_{r+1}, \dots, \alpha_n$ 构成 W^\perp 的正交基。

证明:

之前已经证明了 W^\perp 为 W 的补空间, 由此从 $\dim W = r$ 可知 $\dim W^\perp = n - r$ 。

由于 $\alpha_{r+1}, \dots, \alpha_n$ 是正交基的一部分, 它们线性无关且彼此正交。根据 $\alpha_1, \dots, \alpha_n$ 的两两正交性, 应有

$$\forall i = 1, \dots, r, \quad \forall j = r + 1, \dots, n, \quad (\alpha_i, \alpha_j) = 0$$

从而可知

$$\forall j = 1, \dots, n, \quad \alpha_j \in \{\alpha_1, \dots, \alpha_r\}^\perp$$

利用之前证明的补的性质即有

$$\forall j = 1, \dots, n, \quad \alpha_j \in \langle \alpha_1, \dots, \alpha_r \rangle^\perp = W^\perp$$

由于它们的个数等于 W^\perp 的维数, 且线性无关, 它们即构成 W^\perp 的一组基, 又由正交性知为正交基。

标准正交基在实内积空间中具有重要的性质, 以下假设 V 是有限维实内积空间, 一组标准正交基 S 为 $\alpha_1, \dots, \alpha_n$:

1. 对 V 中任何向量 x 有

$$x = \sum_{i=1}^n (\alpha_i, x) \alpha_i$$

证明:

由于 $\alpha_1, \dots, \alpha_n$ 构成 V 一组基, 可设

$$x = \sum_{i=1}^n \lambda_i \alpha_i$$

由于坐标的唯一性, 只需证明 $\lambda_i = (\alpha_i, x)$ 即可。对任何 $k = 1, \dots, n$, 两侧同时与 α_k 作内积, 利用双线性性得到

$$(x, \alpha_k) = \sum_{i=1}^n \lambda_i (\alpha_i, \alpha_k)$$

再利用 S 的标准正交性, 右侧只有 $i = k$ 时为 λ_i , 否则为 0, 因此右侧即为 λ_k , 这就得到了

$$\forall k = 1, \dots, n, \quad \lambda_k = (x, \alpha_k) = (\alpha_k, x)$$

从而得证。

2. 对 V 中任何向量 x , 设其坐标为 x_S , 则 $(x, y) = x_S^T y_S$ (右侧即相当于 \mathbb{R}^n 中的内积)。

证明:

根据标准正交基的定义, 其下内积的度量矩阵为 I , 而根据度量矩阵的定义可知

$$(x, y) = x_S^T I y_S = x_S^T y_S$$

即得证。

3. 若 T 是 V 的一组基, T 是标准正交基当且仅当 S 到 T 的过渡矩阵 P 为正交阵。

证明:

由之前证明的讨论, T 为标准正交基当且仅当其下内积的度量矩阵为 I , 也当且仅当

$$\forall x, y \in V, \quad (x, y) = x_T^T y_T$$

又由于 $(x, y) = x_S^T y_S$, 上式等价于

$$\forall x, y \in V, \quad x_T^T y_T = x_S^T y_S$$

设 S 到 T 的过渡矩阵为 P , 利用坐标变换的性质有 $x_S = P x_T$, 于是可进一步等价于

$$\forall x, y \in V, \quad x_T^T y_T = (P x_T)^T P x_T$$

利用坐标变换的同构性质, 其是满射, 于是上式可等价为

$$\forall \alpha, \beta \in \mathbb{R}^n, \quad \alpha^T \beta = \alpha^T P^T P \beta$$

由于左右都是 α, β 的双线性函数, 利用度量矩阵唯一性可知上式恒成立等价于 $P^T P = I$, 这即是方阵 P 为正交阵的定义。

最后, 由于 V 维数有限, 之前已证任何子空间 W 均满足 $V = W \oplus W^\perp$ 。我们将投影 $P_W^{(W, W^\perp)}$ (定义见本讲义 20.2.2) 简记为 P_W , 称为到 W 的正交投影。除了投影本身具有的幂等性外, 其还有性质:

1. 对任何 $x \in V$ 有

$$\forall w \in W, \quad \|x - P_W x\| \leq \|x - w\|$$

* 这说明 $P_W x$ 是 W 中离 x 最近的点。

证明:

两侧同平方, 只需证明

$$\forall w \in W, \quad (x - P_W x, x - P_W x) \leq (x - w, x - w)$$

设 $x = y + z$, 其中 $y \in W$ 、 $z \in W^\perp$, 则 $P_W x = y$, 于是左侧即为 (z, z) , 右侧可展开并利用对称性写为

$$(y + z - w, y + z - w) = (z, z) + (y - w, y - w) + 2(z, y - w)$$

根据定义, w, y 均在 W 中, 从而 $y - w \in W$, 而 $z \in W^\perp$, 第三项为 0, 这样右侧就化为了

$$(z, z) + (y - w, y - w)$$

由正定性, $(y - w, y - w) \geq 0$, 于是左侧不超过右侧恒成立。

2. 对任何 $x, y \in V$ 有

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

证明:

设 $x = x_1 + x_2$, 其中 $x_1 \in W$ 、 $x_2 \in W^\perp$; $y = y_1 + y_2$, 其中 $y_1 \in W$ 、 $y_2 \in W^\perp$ 。

第一项即为 (x_1, y_1) , 而第二项利用正交性为

$$(x_1, y_1 + y_2) = (x_1, y_1) + (x_1, y_2) = (x_1, y_1)$$

于是前两项相等。同理展开 x 可得一、三两项相等, 于是得证。

3. 设 W 的一组标准正交基为 $\alpha_1, \dots, \alpha_r$, 则

$$P_W x = \sum_{i=1}^r (\alpha_i, x) \alpha_i$$

证明:

将 $\alpha_1, \dots, \alpha_r$ 扩充为全空间的标准正交基 $\alpha_1, \dots, \alpha_n$, 已证明 $\alpha_{r+1}, \dots, \alpha_n$ 构成 W^\perp 的标准正交基。

利用标准正交基的性质

$$x = \sum_{i=1}^n (\alpha_i, x) \alpha_i = \sum_{i=1}^r (\alpha_i, x) \alpha_i + \sum_{i=r+1}^n (\alpha_i, x) \alpha_i$$

将两项分别记作 x_1, x_2 。利用子空间的封闭性, 由于 $\alpha_1, \dots, \alpha_r \in W, \alpha_{r+1}, \dots, \alpha_n \in W^\perp$ 即得 $x_1 \in W, x_2 \in W^\perp$, 于是即有

$$P_W x = x_1 = \sum_{i=1}^r (\alpha_i, x) \alpha_i$$

* 事实上, 利用投影映射的第一条性质, 之前的最小二乘问题可以直接表述为投影以解决。第三条性质则给出了投影的具体算法: 从标准正交基构造进行计算。

§25.3 实内积空间的同构

25.3.1 线性映射的矩阵表示

矩阵表示

标准正交基下的矩阵

基变换-正交相抵

奇异值分解, 详见下章例题

25.3.2 等距映射

一定是线性单射

矩阵表示?

25.3.3 同构

矩阵表示-正交阵

自同构-等距映射

§25.4 内积下的线性变换

25.4.1 伴随变换

转置

自伴(对称)、斜自伴(斜对称)

可正交相似对角化当且仅当对称

25.4.2 正规变换

定义

判定

正规变换的相似标准形

由正规标准形推自伴、斜自伴、正交的相似标准形

性质与标准形详见期末复习题

二十六 补充：一般内积空间

§26.1 实内积空间综合

26.1.1 例题

1. (a) 对 n 维实内积空间 V 与 $V \rightarrow \mathbb{R}$ 的线性映射 f , 证明存在 $\alpha \in V$ 使得

$$\forall \beta \in V, \quad f(\beta) = (\alpha, \beta)$$

- (b) 对 \mathbb{F} 上的有限维线性空间 V 与 V 上的双线性函数 ϕ , 若 ϕ 退化, 构造线性映射 f 使得不存在 $\alpha \in V$ 满足

$$\forall \beta \in V, \quad f(\beta) = \phi(\alpha, \beta)$$

- (c) 对 \mathbb{F} 上的有限维线性空间 V 与 V 上的双线性函数 ϕ , 若 ϕ 非退化, 是否对任何线性映射 f 都存在 $\alpha \in V$ 使得

$$\forall \beta \in V, \quad f(\beta) = \phi(\alpha, \beta)$$

给出证明或反例。

2. 对 n 维实内积空间 V 中的两个向量 x, y , 定义它们的距离为

$$d(x, y) = \sqrt{(x - y, x - y)}$$

对两个集合 S, T , 定义它们的距离为

$$d(S, T) = \inf_{x \in S, t \in T} d(x, t)$$

也即其中各一点距离的最小值。接下来的每问结果可用正交投影表示, 到子空间 V_0 的正交投影记为 P_{V_0} 。

- (a) 设 $S = \{x\}$, T 为 V 的子空间 W , 计算 $d(S, T)$;

- (b) 设 $S = x + U$, 其中 U 为 V 的子空间, T 为 V 的子空间 W , 计算 $d(S, T)$ 。

- (c) 设 $S = x + U$, 其中 U 为 V 的子空间, T 为 $y + W$, 其中 W 为 V 的子空间, 计算 $d(S, T)$ 。

3. 对 n 维实内积空间 V 中的两个子空间 W_1, W_2 , 设到它们的正交投影为 P_1, P_2 。若 $\lambda P_1 + \mu P_2$ 为到子空间 W_0 的正交投影, 求所有可能的 λ, μ, W_0 。

4. 设 U, V 为 n, m 维实内积空间, \mathcal{A} 是 U, V 之间的非零线性映射。

- (a) 求定义在 $\alpha \neq 0$ 上的函数

$$f(\alpha) = \frac{\|\mathcal{A}(\alpha)\|}{\|\alpha\|}$$

的最大值与取到最大值的一个 α 。

(b) 证明存在向量 $\alpha \in U$ 使得

$$\mathcal{A}(\alpha) \neq 0, \quad \mathcal{A}(\langle \alpha \rangle^\perp) \subset \langle \mathcal{A}(\alpha) \rangle^\perp$$

(c) 证明存在 U, V 的一组标准正交基使得 \mathcal{A} 在这组基下的矩阵表示为

$$\begin{pmatrix} \Sigma & O \\ O & O \end{pmatrix}, \quad \Sigma = \text{diag}(\sigma_1, \dots, \sigma_r), \quad \sigma_1 \geq \dots \geq \sigma_r > 0$$

5. 设 \mathcal{A} 为 n 维实内积空间 U 上的线性变换，用 \mathcal{A}^* 表示其伴随变换。

(a) 若 $(\alpha, \beta) = 0$ 等价于 $(\mathcal{A}\alpha, \mathcal{A}\beta) = 0$ 对任何 $\alpha, \beta \in U$ 成立，证明存在正数 t 使得 $\mathcal{A}^{-1} = t\mathcal{A}^*$ 。

(b) 若 $\mathcal{A}^* + \mathcal{A} = \mathcal{O}$ ，证明 $\text{Ker } \mathcal{A} = (\text{Im } \mathcal{A})^\perp$ ，其逆命题是否成立？给出证明或反例。

(c) 若 $\mathcal{A} = \mathcal{A}^*$ ，证明 \mathcal{A} 的不同特征子空间正交。若 \mathcal{A} 可对角化且不同特征子空间正交，是否有 $\mathcal{A} = \mathcal{A}^*$ ？给出证明或反例。

26.1.2 解答

1. (a) 设 V 的一组标准正交基是 β_1, \dots, β_n ，并取

$$\alpha = \sum_{i=1}^n f(\beta_i) \beta_i$$

则对任何 $\beta \in V$ ，利用标准正交基的性质可知

$$\beta = \sum_{i=1}^n (\beta, \beta_i) \beta_i$$

从而利用 f 与内积的线性性有

$$f(\beta) = \sum_{i=1}^n (\beta, \beta_i) f(\beta_i) = \left(\sum_{i=1}^n f(\beta_i) \beta_i, \beta \right) = (\alpha, \beta)$$

即得证。

* 事实上从上述过程中还可看出对任何 $\beta \in V$ 均有

$$\left(\sum_{i=1}^n f(\beta_i) \beta_i, \beta \right) = (\alpha, \beta)$$

从而考虑 β 为一组标准正交基可知 α 唯一确定。这个定理称为 **Riesz 表示定理**。

(b) 由 ϕ 退化，其存在非零右根 γ 。将 γ 扩充为 V 的一组基 $\gamma, \gamma_2, \dots, \gamma_n$ ，并用基映射定义线性映射

$$f(\gamma) = 1, \quad f(\gamma_2) = \dots = f(\gamma_n) = 0$$

此时，由右根定义，对任何 $\alpha \in V$ 有 $\phi(\alpha, \gamma) = 0$ ，但 $f(\gamma) = 1$ ，等号不可能成立。

(c) 答案是肯定的，我们下面给出证明，设 $\dim V = n$ 。

考虑 V 的一组基 β_1, \dots, β_n ，由 ϕ 非退化， ϕ 在这组基下的度量矩阵 A 应可逆，设 A 各分量为 a_{ij} 。

对任何 $\alpha, \beta \in V$ ，设

$$\alpha = \sum_{i=1}^n \lambda_i \beta_i, \quad \beta = \sum_{j=1}^n \mu_j \beta_j$$

这里所有 $\lambda_i, \mu_j \in \mathbb{F}$, 则利用线性性与双线性性直接计算可知

$$f(\beta) = \sum_{j=1}^n \mu_j f(\beta_j)$$

$$\phi(\alpha, \beta) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j \phi(\beta_i, \beta_j) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j a_{ij}$$

要想上式对任何 β 成立, 也即对一切 μ_j 成立, 对比 μ_j 前的系数可发现也即等价于

$$\sum_{i=1}^n \lambda_i a_{ij} = f(\beta_j), \quad j = 1, \dots, n$$

将所有 λ_i 拼成列向量 x , 所有 $f(\beta_j)$ 拼成列向量 y , 可发现上式即方程组

$$A^T x = y$$

从而由 A 可逆即存在唯一解

$$x = A^{-T} y$$

以解出的 x 作为坐标得到的 α 即为满足要求的元素。

* 同样可发现 α 具有唯一性。

2. 用 $\|x\|$ 表示 $\sqrt{x, x}$, 则 $d(x, y) = \|x - y\|$.

(a) 本讲义 25.2.2 中已经证明了 x 到 W 的正交投影 $P_W x$ 满足

$$\forall w \in W, \quad \|x - P_W x\| \leq \|x - w\|$$

且 $P_W x \in W$, 可取到, 于是根据定义可知

$$\|x - P_W x\| = \inf_{w \in W} d(x, w) = d(S, T)$$

(b) 我们下面证明

$$d(S, T) = \|x - P_{U+W} x\|$$

根据 S 、 T 的定义有

$$d(S, T) = \inf_{s \in S, w \in W} \|s - w\| = \inf_{u \in U, w \in W} \|(x + u) - w\|$$

分两部分说明:

- 存在 $s \in S$ 、 $t \in T$ 使得 $d(s, t) = \|x - P_{U+W} x\|$

由于 $P_{U+W} x \in U + W$, 设其为 $u_0 + w_0$, 其中 $u_0 \in U$ 、 $w_0 \in W$, 即可发现取 $u = -u_0$ 、 $w = w_0$ 得到 (由子空间封闭性 $-u_0 \in U$)

$$\|(x + u) - w\| = \|x - u_0 - w_0\| = \|x - P_{U+W} x\|$$

从而其确实可取到。

- 对任意 $s \in S$ 、 $t \in T$, $d(s, t) \geq \|x - P_{U+W} x\|$

若 $u \in U$ 、 $w \in W$, 由子空间封闭性与和空间定义可知 $-u + w \in U + W$, 根据正交投影性质有

$$\forall y \in U + W, \quad \|x - P_{U+W} x\| \leq \|x - y\|$$

取 $y = -u + w$ 即得

$$\|x - P_{U+W} x\| \leq \|(x + u) - w\|$$

这就证明了其的确最小值。

(c) 利用定义有

$$d(S, T) = \inf_{s \in S, t \in T} \|s - t\| = \inf_{u \in U, w \in W} \|(x + u) - (y + w)\|$$

利用加法结合律直接得到上式改写为

$$d(S, T) = \inf_{u \in U, w \in W} \|(x - y + u) - w\| = d(x - y + U, W)$$

从而由 (b) 得结果为 $\|x - y - P_{U+W}(x - y)\|$ 。

* 此题说明了正交投影在距离相关问题中的处理方式。

3. 我们先证明引理： n 维实内积空间中的线性变换 \mathcal{A} 是正交投影当且仅当 $\mathcal{A}^2 = \mathcal{A} = \mathcal{A}^*$ 。

* 此结论非常建议掌握证明。

证明：

- 左推右

若 \mathcal{A} 是正交投影，则由投影映射性质可知 $\mathcal{A}^2 = \mathcal{A}$ ，只需证明 $\mathcal{A}^* = \mathcal{A}$ 。

在本讲义 25.2.2 中已经证明了 $(\mathcal{A}x, y) = (x, \mathcal{A}y)$ 对任何 $x, y \in V$ 成立，而根据定义伴随映射是满足 $(\mathcal{A}x, y) = (x, \mathcal{A}^*y)$ 恒成立的唯一线性映射，由唯一性得结论。

- 右推左

若 $\mathcal{A}^2 = \mathcal{A}$ ，由本讲义 20.2.2 的证明可知 \mathcal{A} 为直和分解 $V = \text{Im } \mathcal{A} \oplus \text{Ker } \mathcal{A}$ 时到 $\text{Im } \mathcal{A}$ 的投影。为了证明其为正交投影，只需再证明 $\text{Ker } \mathcal{A} = (\text{Im } \mathcal{A})^\perp$ 。

首先，对任何 $x \in \text{Ker } \mathcal{A}$ 、 $y \in \text{Im } \mathcal{A}$ ，设 $y = \mathcal{A}z$ ，由 $\mathcal{A}^* = \mathcal{A}$ 可知

$$(x, y) = (x, \mathcal{A}z) = (\mathcal{A}^*x, z) = (\mathcal{A}x, z) = (0, z) = 0$$

从而由定义 $\text{Ker } \mathcal{A} \subset (\text{Im } \mathcal{A})^\perp$ 。由于 $(\text{Im } \mathcal{A})^\perp$ 与 $\text{Ker } \mathcal{A}$ 均为 $\text{Im } \mathcal{A}$ 的补空间，其维数相等，从而根据包含关系即得 $\text{Ker } \mathcal{A} = (\text{Im } \mathcal{A})^\perp$ ，得证。

由此，由于 $P_1^* = P_1$ 、 $P_2^* = P_2$ ，利用伴随的线性性可知

$$(\lambda P_1 + \mu P_2)^* = \lambda P_1^* + \mu P_2^* = \lambda P_1 + \mu P_2$$

从而其为正交投影当且仅当其幂等，即

$$(\lambda P_1 + \mu P_2)^2 = \lambda P_1 + \mu P_2$$

展开并利用 $P_1^2 = P_1$ 、 $P_2^2 = P_2$ 整理得到 (右侧代表零映射)

$$(\lambda^2 - \lambda)P_1 + (\mu^2 - \mu)P_2 + \lambda\mu(P_1P_2 + P_2P_1) = O$$

我们记左侧的映射为 $\mathcal{A}_{\lambda, \mu}$ 。先讨论一些平凡情况，可直接通过上式代入特殊值验证：

情况 1: $W_1 = W_2 = \{0\}$

此时 P_1 、 P_2 为零映射，无论如何取 λ 、 μ 均为到 $\{0\}$ 的正交投影，即零映射。

情况 2: 若 $W_1 = \{0\}$ 、 $W_2 \neq \{0\}$

此时 P_1 为零映射， λ 可任取， μ 只能为 0 或 1， μ 取为 0 代表到 $\{0\}$ 的正交投影， μ 取为 1 代表到 W_2 的正交投影。

情况 3: $W_2 = \{0\}$ 、 $W_1 \neq \{0\}$

此时 P_2 为零映射， μ 可任取， λ 只能为 0 或 1， λ 取为 0 代表到 $\{0\}$ 的正交投影， λ 取为 1 代表到 W_1 的正交投影。

情况 4： W_1 、 W_2 均不为零空间、 μ 或 λ 有 0

当 $\mu = 0$ 时 λ 只能取为 0 或 1，当 $\lambda = 0$ 时 μ 只能取为 0 或 1，对应正交投影到的子空间与前两种情况相同。

我们下面假设 W_1 、 W_2 均不为零空间，并考虑 λ 、 μ 均非 $\mathbf{0}$ 的解。仍然先说明引理， $\mathcal{A}_{\lambda,\mu}$ 为零映射当且仅当 $\mathcal{A}_{\lambda,\mu}(W_1) = \{0\}$ 、 $\mathcal{A}_{\lambda,\mu}(W_2) = \{0\}$ 。

证明：

左推右由零映射定义可证，只需证明右推左。

根据本讲义 24.1.2 的证明， $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$ ，对任何 $x \in (W_1 + W_2)^\perp$ ，由 $x \in W_1^\perp$ 利用正交投影定义可知 $P_1x = 0$ ，同理 $P_2x = 0$ ，于是根据 $\mathcal{A}_{\lambda,\mu}x$ 的每一项都进行了投影 P_1 或 P_2 可知 $\mathcal{A}_{\lambda,\mu}x = 0$ ，从而

$$\mathcal{A}((W_1 + W_2)^\perp) = \{0\}$$

若 $\mathcal{A}_{\lambda,\mu}(W_1) = \{0\}$ 、 $\mathcal{A}_{\lambda,\mu}(W_2) = \{0\}$ ，利用线性性可验证

$$\mathcal{A}_{\lambda,\mu}(W_1 + W_2 + (W_1 + W_2)^\perp) = \{0\} + \{0\} + \{0\} = \{0\}$$

而利用正交补是补空间，左侧即 $\mathcal{A}_{\lambda,\mu}(V)$ ，从而得证 $\mathcal{A}_{\lambda,\mu}$ 为零映射。

由此，我们将问题化为了 $x \in W_1$ 与 $x \in W_2$ 的情况。可以发现，当 $x \in W_1$ 时，利用正交投影定义 $P_1x = x$ ，从而整理得原式即化为

$$\forall x \in W_1, \quad (\lambda^2 - \lambda)x + (\mu^2 - \mu + \lambda\mu)P_2x + \lambda\mu P_1P_2x = 0$$

同理得 $x \in W_2$ 时可化为

$$\forall x \in W_2, \quad (\lambda^2 - \lambda + \lambda\mu)P_1x + (\mu^2 - \mu)x + \lambda\mu P_2P_1x = 0$$

以上两式即与 $\lambda P_1 + \mu P_2$ 是正交投影等价。

我们先解决一个特殊情况，即 $W_1 \cap W_2 \neq \{0\}$ 时。此时，取 $x \in W_1 \cap W_2$ ，应有 $P_1x = P_2x = x$ ，从而可发现

$$(\lambda^2 - \lambda + \mu^2 - \mu + 2\lambda\mu)x = 0$$

由于 x 非零，可知前面系数必然为 0，将其因式分解为 $(\lambda + \mu - 1)(\lambda + \mu)$ 即得到 $\lambda + \mu = 1$ 或 $\mu = -\lambda$ 。

情况 5： $W_1 \cap W_2 \neq \{0\}$ 且 $\lambda + \mu = 1$

此时代入 $\mu = 1 - \lambda$ ，两式可化为

$$\forall x \in W_1, \quad (\lambda^2 - \lambda)x + (\lambda - \lambda^2)P_1P_2x = 0$$

$$\forall x \in W_2, \quad (\lambda^2 - \lambda)x + (\lambda - \lambda^2)P_2P_1x = 0$$

由于已经假设了 λ 、 μ 均非零，即得

$$\forall x \in W_1, \quad x = P_1P_2x$$

$$\forall x \in W_2, \quad x = P_2P_1x$$

先考虑第一式，利用正交投影的性质可知 $\|x\| \geq \|P_2x\| \geq \|P_1P_2x\|$ ，且等号成立当且仅当 $x \in W_2$ 、 $P_1x \in W_1$ ，由此可直接得到 $W_1 \subset W_2$ ，而第二式可以得出 $W_2 \subset W_1$ ，从而此时存在非平凡解当且仅当 $W_1 = W_2$ 。这时可直接验证任何满足 $\lambda + \mu = 1$ 的 λ 、 μ 都对应到 W_1 （或 W_2 ）的正交投影。

情况 6: $W_1 \cap W_2 \neq \{0\}$ 且 $\mu = -\lambda$

此时代入得两式可化为

$$\forall x \in W_1, (\lambda^2 - \lambda)x + \lambda P_2x - \lambda^2 P_1P_2x = 0$$

$$\forall x \in W_2, -\lambda P_1x + (\lambda^2 + \lambda)x - \lambda^2 P_2P_1x = 0$$

由 $\lambda \neq 0$ 进一步消去并整理可得 (\mathcal{I} 表示恒等映射)

$$\forall x \in W_1, (\lambda - 1)x = (\lambda P_1 - \mathcal{I})(P_2x)$$

$$\forall x \in W_2, (\lambda + 1)x = (\mathcal{I} + \lambda P_2)(P_1x)$$

由于 $\lambda \neq 0$, 讨论其大于或小于 0 即可。若 $\lambda > 0$, 根据第二式有

$$(\lambda + 1)\|x\| = \|(\mathcal{I} + \lambda P_2)(P_1x)\|$$

但由范数三角不等式可知

$$\|(\mathcal{I} + \lambda P_2)(P_1x)\| \leq \|P_1x\| + \lambda\|P_2P_1x\|$$

进一步由正交投影性质 (类似情况 5 中的讨论)

$$\|P_1x\| + \lambda\|P_2P_1x\| \leq \|x\| + \lambda\|x\| = (1 + \lambda)\|x\|$$

由上式对任意 $x \in W_2$ 取等即得必然有 $P_1x = x$ 对任何 $x \in W_2$ 成立, 从而 $W_2 \subset W_1$, 此时代入可发现第二式恒成立, 第一式由于 $P_1P_2x = P_2x$ 成为

$$\forall x \in W_1, (\lambda - 1)x = (\lambda - 1)P_2x$$

由此可得 $\lambda = 1$ 或 $W_2 = W_1$, $W_2 = W_1$ 时任何 $\mu = -\lambda$ 都对应到 $\{0\}$ 的正交投影, $W_2 \neq W_1$ 、 $\lambda = 1$ 、 $\mu = -1$ 时对应到 W_2 在 W_1 中正交补的正交投影。

* 即 W_0 满足 $W_0 \oplus W_2 = W_1$ 且 W_0 与 W_2 正交。

若 $\lambda = -1$, 同理讨论, 最终此情况综合为:

- $W_1 = W_2$, 此时任何满足 $\lambda = -\mu$ 的 λ 、 μ 都对应到 $\{0\}$ 的正交投影;
- $W_2 \subset W_1$ 且不相等, 此时必然 $\lambda = 1$ 、 $\mu = -1$, 对应到 W_2 在 W_1 中正交补的正交投影;
- $W_1 \subset W_2$ 且不相等, 此时必然 $\lambda = -1$ 、 $\mu = 1$, 对应到 W_1 在 W_2 中正交补的正交投影。

最后, 我们来研究 $W_1 \cap W_2 = \{0\}$ 的情况。此时若 $x \in W_1$, 由正交投影性质可知 $P_2x \in W_2$ 、 $P_1P_2x \in W_1$, 根据

$$\forall x \in W_1, (\lambda^2 - \lambda)x + (\mu^2 - \mu + \lambda\mu)P_2x + \lambda\mu P_1P_2x = 0$$

利用交为 0 即得

$$\forall x \in W_1, (\mu^2 - \mu + \lambda\mu)P_2x = 0$$

$$\forall x \in W_1, (\lambda^2 - \lambda)x + \lambda\mu P_1P_2x = 0$$

由 λ 、 μ 非零进一步化简为

$$\forall x \in W_1, (\lambda + \mu - 1)P_2x = 0$$

$$\forall x \in W_1, (\lambda - 1)x + \mu P_1P_2x = 0$$

同理

$$\forall x \in W_2, (\lambda + \mu - 1)P_1x = 0$$

$$\forall x \in W_2, (\mu - 1)x + \lambda P_2P_1x = 0$$

至此, 须讨论 $\lambda + \mu$ 是否为 1。

情况 7: $W_1 \cap W_2 = \{0\}$ 且 $\lambda + \mu = 1$

此时代入 $\mu = 1 - \lambda$, 条件即

$$\forall x \in W_1, \quad \mu(x - P_1 P_2 x) = 0$$

$$\forall x \in W_2, \quad \lambda(x - P_2 P_1 x) = 0$$

由 λ 、 μ 均不为 0, 上式化为

$$\forall x \in W_1, \quad x - P_1 P_2 x = 0$$

$$\forall x \in W_2, \quad x - P_2 P_1 x = 0$$

与情况 5 相同, 此两式成立当且仅当 $W_1 = W_2$, 但又由交为 0 知只能全为零空间, 为之前讨论过的平凡情况。

情况 8: $W_1 \cap W_2 = \{0\}$ 且 $\lambda + \mu \neq 1$

此时可得到

$$\forall x \in W_1, \quad P_2 x = 0$$

$$\forall x \in W_2, \quad P_1 x = 0$$

这两个式子利用正交投影定义即得 $W_1 \subset W_2^\perp$ 、 $W_2 \subset W_1^\perp$, 这说明两子空间相互垂直, 即

$$\forall x \in W_1, y \in W_2, \quad (x, y) = 0$$

而将其代入另外两个条件可得

$$\forall x \in W_1, \quad (\lambda - 1)x = 0$$

$$\forall x \in W_2, \quad (\mu - 1)x = 0$$

由于已经假设两者均不为零空间, 只能 $\lambda = \mu = 1$, 此时可直接验证对应到 $W_1 + W_2$ 的正交投影。

综合情况 1 到情况 8 即给出了这个问题的完整解答。由此可见, W_0 除了 $\{0\}$ 、 W_1 、 W_2 这三种平凡情况外, 只可能是 $W_1 + W_2$ (此时 $\lambda = \mu = 1$) 或 W_1 对 W_2 的正交补 (此时 $\lambda = -1$ 、 $\mu = 1$) 或 W_2 对 W_1 的正交补 (此时 $\lambda = 1$ 、 $\mu = -1$)。

* 教材中的例题包含 $\lambda = \mu = 1$ 与 $\lambda = 1$ 、 $\mu = -1$ 的特例。

4. (a) 设 U 的一组标准正交基是 S , V 的一组标准正交基是 T , α 在 S 下的坐标为 x , \mathcal{A} 在基 S 、 T 下的矩阵表示为 A , 则根据标准正交基的性质 (本讲义 25.2.2) 与矩阵表示的定义可知

$$f(\alpha) = \sqrt{\frac{(Ax)^T(Ax)}{x^T x}} = \sqrt{\frac{x^T A^T A x}{x^T x}}$$

由上学期知识 $A^T A$ 是半正定对称阵, 从而根据上学期知识存在正交阵 P 使得 $A^T A = P^T D P$, P 为正交阵, D 为对角元非负的对角阵。记 $y = Px$, 并设 D 的对角元为 d_1, \dots, d_n , 由正交阵性质可知 $x^T x = x^T P^T P x = y^T y$, 从而直接计算可知 (根据乘可逆阵与坐标映射的同构性可知 $\alpha \neq 0$ 时分母非零)

$$f(\alpha) = \sqrt{\frac{y^T D y}{y^T y}} = \sqrt{\frac{\sum_{i=1}^n d_i y_i^2}{\sum_{i=1}^n y_i^2}}$$

由于乘可逆阵为同构, 坐标映射为同构, α 由 y 唯一确定, 只需找到 y 使得右侧最大即可。不妨设 d_1 是 D 最大的对角元, 分母可直接放大为

$$\sum_{i=1}^n d_i y_i^2 \leq \sum_{i=1}^n d_1 y_i^2 = d_1 \sum_{i=1}^n y_i^2$$

由此可知 $f(\alpha) \leq \sqrt{d_1}$ 。另一方面，当 $y = e_1$ 时，等号成立，从而这就对应了最大值与取到最大值的 α 。

* 由于 D 的对角元是 $A^T A$ 的特征值， $\sqrt{d_1}$ 是其最大特征值的平方根，即 $f(\alpha)$ 最大值是 $A^T A$ 最大特征值的平方根。直接验证可发现 x 取为 $A^T A$ 最大特征值对应的特征向量。

* 此结论事实上是上学期的 **Rayleigh 商**结论，可见本讲义 11.3.1。

(b) 我们说明 (a) 中的 α 即符合要求。

首先，利用定义可发现对任何 $\alpha \neq 0$ 有 $f(\alpha) \geq 0$ ，且等号成立当且仅当 $\mathcal{A}(\alpha) = 0$ ，于是由 \mathcal{A} 不为零映射可知 $f(\alpha)$ 最大值大于 0，也即此时 $\mathcal{A}(\alpha) \neq 0$ 。

为证明第二个式子，由子空间封闭性可知 $\beta \in \langle \alpha \rangle^\perp$ 当且仅当 $(\alpha, \beta) = 0$ 、 $\mathcal{A}\beta \in \langle \mathcal{A}\alpha \rangle^\perp$ 当且仅当 $(\mathcal{A}\alpha, \mathcal{A}\beta) = 0$ ，从而利用正交补定义只需证对任何满足 $(\alpha, \beta) = 0$ 的 β 有 $(\mathcal{A}\alpha, \mathcal{A}\beta) = 0$ 。

若存在 β 使得 $(\alpha, \beta) = 0$ 且 $(\mathcal{A}\alpha, \mathcal{A}\beta) \neq 0$ ，对 $\lambda \in \mathbb{R}$ ，利用线性性展开计算并代入 $(\alpha, \beta) = 0$ 可知

$$f^2(\alpha + \lambda\beta) = \frac{\|\mathcal{A}\alpha + \lambda\mathcal{A}\beta\|^2}{\|\alpha + \lambda\beta\|^2} = \frac{\|\mathcal{A}\alpha\|^2 + 2\lambda(\mathcal{A}\alpha, \mathcal{A}\beta) + \lambda^2\|\mathcal{A}\beta\|^2}{\|\alpha\|^2 + \lambda^2\|\beta\|^2}$$

从而直接计算可知

$$f^2(\alpha + \lambda\beta) - f^2(\alpha) = \lambda \frac{2(\mathcal{A}\alpha, \mathcal{A}\beta)\|\alpha\|^2 + \lambda(\|\mathcal{A}\beta\|^2\|\alpha\|^2 - \|\beta\|^2\|\mathcal{A}\alpha\|^2)}{(\|\alpha\|^2 + \lambda^2\|\beta\|^2)\|\alpha\|^2}$$

由于 $(\mathcal{A}\alpha, \mathcal{A}\beta) \neq 0$ ，只要 $|\lambda|$ 充分小即可使分母与 $(\mathcal{A}\alpha, \mathcal{A}\beta)$ 同号，再取 λ 与 $(\mathcal{A}\alpha, \mathcal{A}\beta)$ 同号即得

$$f^2(\alpha + \lambda\beta) - f^2(\alpha) > 0$$

由 f 非负性可知 $f(\alpha + \lambda\beta) > f(\alpha)$ ，与最大值矛盾，从而得证。

(c) 对 U 的维数进行归纳。当 $\dim U = 1$ 时，取 U 的一组标准正交基 $\{\alpha\}$ ，由 \mathcal{A} 不是零映射可知 $\mathcal{A}(\alpha) \neq 0$ ，由此可将

$$\gamma_1 = \frac{\mathcal{A}\alpha}{\|\mathcal{A}\alpha\|}$$

扩充为 V 的标准正交基 $\gamma_1, \dots, \gamma_m$ ，直接由定义可知 \mathcal{A} 在 $\{\alpha\}$ 与 $\gamma_1, \dots, \gamma_m$ 下的矩阵表示为

$$(\|\mathcal{A}\alpha\|, 0, \dots, 0)^T$$

符合要求的形式。

若 $\dim U = n - 1$ 时结论成立，考虑 $\dim U = n$ 时。由于 \mathcal{A} 非零，可按照 (b) 取出符合要求的 α 。记

$$u = \frac{\alpha}{\|\alpha\|}, \quad v = \frac{\mathcal{A}\alpha}{\|\mathcal{A}\alpha\|}, \quad U_1 = \langle u \rangle, \quad V_1 = \langle v \rangle, \quad U_2 = U_1^\perp, \quad V_2 = V_1^\perp$$

由定义可知 $\mathcal{A}(U_1) = V_1$ ，根据 (b) 中证明有 $\mathcal{A}(U_2) \subset \mathcal{A}(V_2)$ 。由于 U_1, U_2 互补， V_1, V_2 互补，根据本讲义 20.2.1 可知 \mathcal{A} 的矩阵表示可由 $\mathcal{A}|_{U_1 \rightarrow V_1}$ 与 $\mathcal{A}|_{U_2 \rightarrow V_2}$ 的矩阵表示作为分块对角阵拼成。

由于 $\dim U_2 = n - 1$ ，且根据 $\mathcal{A}\alpha \neq 0$ 知 $\dim V_2 = n - \dim V_1 = n - 1$ ，存在 U_2 的一组标准正交基 u_2, \dots, u_n 与 V_2 的一组标准正交基 v_2, \dots, v_m 使得 $\mathcal{A}|_{U_2 \rightarrow V_2}$ 在这组基下的矩阵表示为 O (对应零映射情况) 或

$$\begin{pmatrix} \Sigma_2 & O \\ O & O \end{pmatrix}, \quad \Sigma_2 = \text{diag}(\sigma_2, \dots, \sigma_r), \quad \sigma_2 \geq \dots \geq \sigma_r > 0$$

利用正交补的性质直接验证可知 u, u_2, \dots, u_n 为 U 的标准正交基， v, v_2, \dots, v_m 为 V 的标准正交基，又由于

$$\mathcal{A}(u) = \frac{1}{\|\alpha\|} \mathcal{A}\alpha = \frac{\|\mathcal{A}\alpha\|}{\|\alpha\|} v = f(\alpha)v$$

即得 $\mathcal{A}|_{U_1 \rightarrow V_1}$ 的矩阵表示为 $f(\alpha)$, 从而根据本讲义 20.2.1, \mathcal{A} 在上述标准正交基下的矩阵表示为 (第一种情况对应 $\mathcal{A}|_{U_2 \rightarrow V_2}$ 为零映射, 未写出元素为 0)

$$\begin{pmatrix} f(\alpha) & \\ & O \end{pmatrix}$$

或

$$\begin{pmatrix} \Sigma & O \\ O & O \end{pmatrix}, \quad \Sigma = \text{diag}(f(\alpha), \sigma_2, \dots, \sigma_r), \quad \sigma_2 \geq \dots \geq \sigma_r > 0$$

第一种情况已经符合要求, 对第二种情况, 我们只需再证明 $f(\alpha) \geq \sigma_2$ 。

直接利用矩阵表示定义可发现 σ_2 满足

$$\mathcal{A}(u_2) = \sigma_2 v_2$$

从而 $\|\mathcal{A}(u_2)\| = \sigma_2 \|v_2\|$, 再根据 u_2, v_2 为标准正交基一部分, 模长均为 1, 即可知

$$\sigma_2 = \frac{\|\mathcal{A}u_2\|}{\|v_2\|} = \frac{\|\mathcal{A}u_2\|}{\|u_2\|} = f(u_2)$$

由此利用 α 满足 $f(\alpha)$ 最大即得 $f(\alpha) \geq f(u_2) = \sigma_2$, 得证。

* 这就是奇异值分解的空间版本证明, 注意 (b) 中构造 λ 的步骤本质上在利用极限技巧, 这是在有距离的线性空间中可以进行的操作。

5. (a) * 注意此题过程与本讲义 24.3.3 第一个证明的类似之处, 在期末复习题中我们将看到更本质的推广。

首先, 由条件代入 $\beta = \alpha$ 可知 (α, α) 非零时 $(\mathcal{A}\alpha, \mathcal{A}\alpha)$ 非零, 从而利用内积正定性 $\text{Ker } \mathcal{A} = \{0\}$, 又由其为有限维线性空间上的线性变换可知其为同构, 逆的确存在。

为证明 $\mathcal{A}^{-1} = t\mathcal{A}^*$, 只需证明对任何 $x \in V$ 有

$$\mathcal{A}^{-1}x = t\mathcal{A}^*x$$

而这又等价于对任何 $y \in V$ 有 (取 $y = \mathcal{A}^{-1}x - t\mathcal{A}^*x$ 即可从下式成立由正定性得到上式成立)

$$(\mathcal{A}^{-1}x, y) = t(\mathcal{A}^*x, y)$$

由 \mathcal{A} 为同构, 记 z 为 $\mathcal{A}^{-1}x$, 其可取遍 V 。由此再利用伴随映射的定义, 我们最终需要证明存在 $t > 0$ 使得对任何 $z, y \in V$ 有

$$(z, y) = t(\mathcal{A}z, \mathcal{A}y)$$

证明:

任取非零的 $x_0 \in V$, 由 \mathcal{A} 为同构 $\mathcal{A}x_0 \neq 0$, 从而可记

$$t = \frac{(x_0, x_0)}{(\mathcal{A}x_0, \mathcal{A}x_0)}$$

由正定性其大于 0。

我们先证明对任何 $z \in V$ 有 $(z, x_0) = t(\mathcal{A}z, \mathcal{A}x_0)$ 。记 $\lambda = \frac{(z, x_0)}{(x_0, x_0)}$, 则直接计算可知

$$(z - \lambda x_0, x_0) = (z, x_0) - \lambda(x_0, x_0) = 0$$

由此根据条件有 $(\mathcal{A}(z - \lambda x_0), \mathcal{A}x_0) = 0$, 由线性性展开可知

$$(\mathcal{A}z, \mathcal{A}x_0) = \lambda(\mathcal{A}x_0, \mathcal{A}x_0) = \frac{(z, x_0)(\mathcal{A}x_0, \mathcal{A}x_0)}{(x_0, x_0)} = \frac{1}{t}(z, x_0)$$

从而得证。再利用对称性, 对任何 $z \in V$ 也有 $(x_0, z) = t(\mathcal{A}x_0, \mathcal{A}z)$ 。

下面对任何 y, z 证明 $(y, z) = t(\mathcal{A}y, \mathcal{A}z)$ 。我们需要分类讨论:

- 若 $(y, x_0) \neq 0$, 记 $\lambda = \frac{(y, z)}{(y, x_0)}$, 直接计算可知

$$(y, z - \lambda x_0) = (y, z) - \lambda(y, x_0) = 0$$

由此根据条件有 $(\mathcal{A}y, \mathcal{A}(z - \lambda x_0)) = 0$, 由线性性展开可知

$$(\mathcal{A}y, \mathcal{A}z) = \lambda(\mathcal{A}y, \mathcal{A}x_0) = \frac{(y, z)}{t} \mathcal{A}y, \mathcal{A}x_0)(y, x_0) = \frac{1}{t}(y, z)$$

最后一个等号是利用之前已证, 从而结论成立。

- 若 $(x_0, z) \neq 0$, 与上同理记 $\lambda = \frac{(y, z)}{(x_0, z)}$ 并考虑 $(y - \lambda x_0, z)$ 可得结论。
- 若 $(y, x_0) = (x_0, z) = 0$, 对任何 $\lambda, \mu \in \mathbb{R}$ 由条件展开有

$$(y - \lambda x_0, z - \mu x_0) = (y, z) + \lambda\mu(x_0, x_0)$$

由此, 取 $\lambda = -1, \mu = \frac{(y, z)}{(x_0, x_0)}$ 即可使上式为 0, 此时有

$$(\mathcal{A}y - \lambda \mathcal{A}x_0, \mathcal{A}z - \mu \mathcal{A}x_0) = 0$$

利用条件可知 $(\mathcal{A}y, \mathcal{A}x_0) = (\mathcal{A}x_0, \mathcal{A}z) = 0$, 从而上式展开得

$$(\mathcal{A}y, \mathcal{A}z) + \lambda\mu(\mathcal{A}x_0, \mathcal{A}x_0) = 0$$

代入即最终得到

$$(\mathcal{A}y, \mathcal{A}z) = \frac{(y, z)(\mathcal{A}x_0, \mathcal{A}x_0)}{(x_0, x_0)} = \frac{1}{t}(y, z)$$

综合三种情况得证。

- (b) 对任何 $x \in \text{Ker } \mathcal{A}, y \in \text{Im } \mathcal{A}$, 设 $y = \mathcal{A}z$, 由伴随映射定义可知

$$(x, y) = (x, \mathcal{A}z) = (\mathcal{A}^*x, z) = -(\mathcal{A}x, z) = 0$$

从而 $\text{Ker } \mathcal{A} \subset (\text{Im } \mathcal{A})^\perp$ 。另一方面, 根据第一同构定理可知 $\dim \text{Ker } \mathcal{A} + \dim \text{Im } \mathcal{A} = n$, 由正交补性质可知 $\dim(\text{Im } \mathcal{A})^\perp + \dim \text{Im } \mathcal{A} = n$, 于是 $\text{Ker } \mathcal{A}$ 与 $(\text{Im } \mathcal{A})^\perp$ 维数相等, 可得只能相等。

其逆命题不成立, 考虑 \mathcal{A} 为恒等变换 \mathcal{I} 即可, 此时由定义 $\mathcal{A}^* = \mathcal{I}$, $\text{Ker } \mathcal{A} = \{0\}$, $\text{Im } \mathcal{A} = V$, 符合要求。

- (c) 设 x 是 \mathcal{A} 的特征值 λ 的特征向量, y 是 \mathcal{A} 的特征值 μ 的特征向量, 利用双线性性与伴随变换定义 $((\mathcal{A}x, y) = (x, \mathcal{A}^*y) = (x, \mathcal{A}y))$ 有

$$(\lambda - \mu)(x, y) = (\lambda x, y) - (x, \mu y) = (\mathcal{A}x, y) - (x, \mathcal{A}y) = (x, \mathcal{A}y) - (x, \mathcal{A}y) = 0$$

若 $\lambda \neq \mu$ 即得 $(x, y) = 0$, 从而得证。

反之, 若 \mathcal{A} 可对角化且不同特征子空间正交, 的确能推出 $\mathcal{A} = \mathcal{A}^*$ 。

设 \mathcal{A} 的不同特征子空间为 V_1, \dots, V_r , 设各自的维数为 n_1, \dots, n_r 。由于它们相互正交, 设 V_1 的一组标准正交基为 v_1, \dots, v_{n_1} , V_2 的一组标准正交基为 $v_{n_1+1}, \dots, v_{n_1+n_2}$, 直到 V_n 的一组标准正交基为 $v_{n_1+\dots+n_{r-1}+1}, \dots, v_n$ 。我们先证明 v_1, \dots, v_n 是 V 的一组标准正交基。

* 由此可见相互正交能推出各自的(标准)正交基拼成直和的(标准)正交基, 这是关于正交基的重要结论。

证明:

由于 \mathcal{A} 可对角化，其不同特征子空间直和为全空间，从而各自的一组基可以拼成全空间一组基，也即 v_1, \dots, v_n 是 V 的一组基。

另一方面，由于每个都是单位向量，规范性满足。对任何不同的 i, j ，若 v_i 与 v_j 在同一个 V_k 中，利用标准正交基定义可知 $(v_i, v_j) = 0$ ，否则利用特征子空间两两正交性可知 $(v_i, v_j) = 0$ 。综合以上得这是一组标准正交基。

由于 v_1, \dots, v_n 都是 \mathcal{A} 的特征向量，利用矩阵表示定义可知 \mathcal{A} 在这组标准正交基下的矩阵表示为对角阵，而实对角阵是实对称阵，利用 $\mathcal{A} = \mathcal{A}^*$ 当且仅当其在某组标准正交基下的矩阵表示对称可知成立。

§26.2 西空间

26.2.1 复内积

半线性性与共轭转置

复内积定义

标准正交基构造

酉方阵基本性质

标准正交基过渡矩阵为酉方阵

正交补、正交投影

26.2.2 内积下的线性映射

等距映射对应标准正交基下矩阵表示的列向量标准正交

一般线性映射-奇异值分解

26.2.3 内积下的线性变换

可酉相似对角化当且仅当为正规变换

二十七 张量代数

本章中，我们将按照如下的顺序介绍有限维线性空间相关的张量基本知识：首先，我们类比双线性函数定义一般的多重线性函数，并从单线性函数（对偶空间）出发构造多重线性函数的一组基，以此定义单线性函数的张量积，即对偶空间张量积。接下来，我们将保持两组基的张量积得到张量积空间一组基的思路，定义一般有限维线性空间的张量积。为了能够利用“同构标准形”进行研究，我们给出向量空间的张量积的两种具体表示方式。最后，我们可以在线性空间张量积的基础上定义线性映射的张量积，并对应在向量空间版本得到矩阵张量积及其性质，在线性空间版本得到其矩阵表示。

§27.1 多重线性函数

27.1.1 对偶空间

在本讲义 24.1.1 中，我们提到，类似双线性函数，我们可以定义任意重线性函数，而本节，我们将解决一个看起来平凡的情况，即单线性函数，或线性函数。

对于 \mathbb{K} 上的线性空间 V ， V 上的线性函数即为 V 到 \mathbb{K} 的线性映射，我们称所有线性函数构成的线性空间 $\text{Hom}(V, \mathbb{K})$ （此处 Hom 即是前半学期定义的所有线性映射构成的线性空间）为 V 的对偶空间，记作 V^* 。

我们将本讲义 19 章得到的结论直接应用到 V^* 上，并给出一些推广：

1. V^* 是一个线性空间, 且 $\dim V = n$ 时 $\dim V^* = n$ 。

证明:

线性空间验证见本讲义 19.3.3, 在 $\dim V = n$ 时, 由于 $\dim \mathbb{K} = 1$, 在本讲义 19.3.3 中我们证明了 $\text{Hom}(V, \mathbb{K})$ 与 $\mathbb{K}^{n \times 1}$ 同构, 于是其维数为 n 。

* 当 V 维数无限时, V^* 可能与 V 不同构, 如 V 为所有实系数多项式构成的线性空间时, V^* 同构于所有实数列构成的线性空间, 前者维数可数后者维数不可数。

2. 给定 V 的一组基 $S = (\alpha_1, \dots, \alpha_n)$, 通过基映射确定线性映射

$$f_i(\alpha_j) = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

则 f_1, \dots, f_n 构成 V^* 的一组基, 称为 S 的对偶基。

证明:

在本讲义 19.3.3 中, 我们已经给出了一般的 $\text{Hom}(U, V)$ 的一组基的构造方式。此处, 由于考虑的是 $\text{Hom}(V, \mathbb{K})$, 我们直接取 \mathbb{K} 的基为 $\{1\}$, 对比可以发现按本讲义 19.3.3 的构造所构造出的就是上述 f_1, \dots, f_n , 从而它们构成一组基。

3. 若 $\dim V = n$, 对 V^* 的任何一组基 f_1, \dots, f_n , 存在 V 的唯一一组基 $\alpha_1, \dots, \alpha_n$ 使得 f_1, \dots, f_n 是 $\alpha_1, \dots, \alpha_n$ 的对偶基。

证明:

任取 V 的一组基 S , 并取 \mathbb{K} 的基为 $T = \{1\}$ 。在这两组基下, 设 f_1, \dots, f_n 的矩阵表示为 β_1, \dots, β_n 。

利用矩阵表示的性质, 所有 β_i 均为 $1 \times n$ 行向量, 且根据矩阵表示为同构, β_1, \dots, β_n 线性无关。

对任何 $v \in V$, 设其在基 S 下坐标为 v_S , 应有

$$(f_i(v))_T = \beta_i v_S$$

由于 T 下任何数的坐标为自身, 上式可以进一步化简为

$$f_i(v) = \beta_i v_S$$

f_1, \dots, f_n 是 $\alpha_1, \dots, \alpha_n$ 的对偶基意味着

$$f_i(\alpha_j) = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

从而有

$$\beta_i(\alpha_j)_S = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

我们将所有 β_i 作为行拼成矩阵 B , 所有 $(\alpha_j)_S$ 作为列拼成矩阵 A , 计算可发现上式等价于

$$BA = I_n$$

由于已知 β_1, \dots, β_n 线性无关, B 可逆, 即得存在唯一解 $A = B^{-1}$, 从而 $(\alpha_1)_S, \dots, (\alpha_n)_S$ 存在唯一。又由于每个坐标对应唯一的向量, $\alpha_1, \dots, \alpha_n$ 存在唯一, 得证。

* 上述过程也给出了给出 f_1, \dots, f_n 求 $\alpha_1, \dots, \alpha_n$ 使得 f_1, \dots, f_n 为其对偶基的算法：从任何一组基给出矩阵表示并拼接，求逆后还原出 $\alpha_1, \dots, \alpha_n$ 。

* 由此， V 的一组基与 V^* 的一组基一一对应。

4. 若 $\alpha_1, \dots, \alpha_n$ 是 V 的一组基， f_1, \dots, f_n 它们的对偶基，则任何映射 $f \in V^*$ 有

$$f = \sum_{i=1}^n f(\alpha_i) f_i$$

对任何向量 $\alpha \in V$ 有

$$\alpha = \sum_{i=1}^n f_i(\alpha) \alpha_i$$

证明：

- 第一式

由于 f_1, \dots, f_n 构成 V^* 一组基，可设

$$f = \sum_{i=1}^n \lambda_i f_i$$

由于坐标的唯一性，只需证明 $\lambda_i = f(\alpha_i)$ 即可。对任何 $k = 1, \dots, n$ ，两侧同时代入 α_k ，利用映射加法、数乘定义得到

$$f(\alpha_k) = \sum_{i=1}^n \lambda_i f_i(\alpha_k)$$

再利用对偶基的定义，右侧只有 $i = k$ 时为 λ_i ，否则为 0，因此右侧即为 λ_k ，这就得到了

$$\forall k = 1, \dots, n, \quad \lambda_k = f(\alpha_k)$$

从而得证。

- 第二式

由于 $\alpha_1, \dots, \alpha_n$ 构成 V^* 一组基，可设

$$\alpha = \sum_{i=1}^n \lambda_i \alpha_i$$

由于坐标的唯一性，只需证明 $\lambda_i = f_i(\alpha)$ 即可。对任何 $k = 1, \dots, n$ 两侧同时代入 f_k ，利用线性性得到

$$f_k(\alpha) = \sum_{i=1}^n \lambda_i f_k(\alpha_i)$$

再利用对偶基的定义，右侧只有 $i = k$ 时为 λ_i ，否则为 0，因此右侧即为 λ_k ，这就得到了

$$\forall k = 1, \dots, n, \quad \lambda_k = f_k(\alpha)$$

从而得证。

* 此定理两式的相似形式与相似证明中可以看出对偶基的对偶体现在何处。此外，可以发现此定理的形式与之前内积空间时 $x = \sum_{i=1}^n (\alpha_i, x) \alpha_i$ 存在相似，事实上二者的确相同，对应内积空间的对偶，不过这不在课程范围内。

5. 若 V 的基 $\alpha_1, \dots, \alpha_n$ 到基 β_1, \dots, β_n 的过渡矩阵为 P ， $\alpha_1, \dots, \alpha_n$ 的对偶基到 β_1, \dots, β_n 的对偶基的过渡矩阵为 P^{-T} 。

证明：

设 $\alpha_1, \dots, \alpha_n$ 的对偶基为 f_1, \dots, f_n , β_1, \dots, β_n 的对偶基为 g_1, \dots, g_n , 由上个性质有

$$\forall j = 1, \dots, n, \quad g_j = \sum_{i=1}^n g_j(\alpha_i) f_i$$

由此, 设 f_1, \dots, f_n 到 g_1, \dots, g_n 的过渡矩阵为 Q , 利用过渡矩阵定义可发现 Q 第 i 行第 j 列的元素 q_{ij} 满足

$$q_{ij} = g_j(\alpha_i)$$

为证明 $Q = P^{-T}$, 两侧同取转置并右乘 P 可得只需证明 $Q^T P = I$ 。而 Q^T 第 i 行第 j 列即为 $g_i(\alpha_j)$, 从而可整体写成

$$Q^T = \begin{pmatrix} g_1(S) \\ g_2(S) \\ \vdots \\ g_n(S) \end{pmatrix}$$

这里 $g_i(S)$ 表示将 S 看作形式行向量, 再将 g 作用在其每个分量上得到新的行向量, 定义与矩阵表示时一致。利用分块矩阵性质可得 (Q 分块为 $n \times 1$, P 看作整体 1×1 , 可乘)

$$Q^T P = \begin{pmatrix} g_1(S)P \\ g_2(S)P \\ \vdots \\ g_n(S)P \end{pmatrix}$$

由 g_i 的线性性, $g_i(S)P = g_i(SP)$, 再根据过渡矩阵的定义 $SP = T$, 从而得到

$$Q^T P = \begin{pmatrix} g_1(T) \\ g_2(T) \\ \vdots \\ g_n(T) \end{pmatrix}$$

考虑其每个分量可发现 $Q^T P$ 的第 i 行第 j 列为 $g_i(\beta_j)$, 由对偶基定义即得 $i = j$ 时为 1, 否则为 0, 从而 $Q^T P = I$, 得证。

于是, V^* 比起一般的 $\text{Hom}(U, V)$ 的特殊性在于对偶基的构造, 且其任何一组基都可看作 V 某组基的对偶基, 以此容易计算 f 在给定基下的坐标。我们之后将见到, 这样的性质将为构造多重线性函数的基提供便利, 这就是为什么对偶空间值得单独引入。

* 另一个较特殊的性质是 V^* 中的函数依靠 Ker 就几乎可以确定, 见期末复习题第 3 题。

27.1.2 多重线性性

我们首先仿照双线性函数的情况, 定义 \mathbb{K} 上线性空间 V_1, V_2, \dots, V_m 上的一个多重线性函数为 $V_1 \times V_2 \times \dots \times V_m$ 到 \mathbb{K} 的映射, 且对每个分量都是线性的。也即, φ 对 1 到 m 中的任何一个 j 满足

$$\varphi(v_1, \dots, \lambda v_{j1} + \mu v_{j2}, \dots, v_m) = \lambda \varphi(v_1, \dots, v_{j1}, \dots, v_m) + \mu \varphi(v_1, \dots, v_{j2}, \dots, v_m)$$

这里 $v_i \in V_i$ 对 $i \neq j$ 成立, $v_{j1}, v_{j2} \in V_j$, $\lambda, \mu \in \mathbb{K}$ 。

* 多重线性会自然与高次函数相关, 这是由于 $\varphi(\lambda v_1, \dots, \lambda v_m) = \lambda^m \varphi(v_1, \dots, v_m)$, 因此 m 重线性中可以构造出 m 次函数。在之前, 我们已经定义出了二重线性与二次型的关系。

* 多重线性函数可以与双线性函数类似定义非退化性: 它对第 k 个分量非退化是指不存在非零的 $z_k \in V_k$ 使得

$$\forall v_1 \in V_1, \dots, v_{k-1} \in V_{k-1}, v_{k+1} \in V_{k+1}, \dots, v_m \in V_m, \quad \varphi(v_1, \dots, v_{k-1}, z_k, v_{k+1}, \dots, v_m) = 0$$

这个定义也可以用之前的“可区分性”进行表示，证明过程完全类似。

与双线性函数类似（见本讲义 24.1.4），我们可以直接验证 V_1, \dots, V_m 上的多重线性函数构成线性空间，记作 $T(V_1, \dots, V_m)$ 。由此，假设每个 V_i 都是有限维的，设其维数 n_i ，一组基为 S_i ，我们本部分的目标是确定其维数与一组基。不过，双线性函数时我们尚能用矩阵给出对应关系，面对多重线性函数，我们就不得不采用其他策略了。

首先，我们还是需要回到 \mathbb{K} 上线性空间 V, W 上的双线性函数，构造其一组基。假设 $\dim V = m$ 、 $\dim W = n$ ，这次，我们将采用对偶空间的策略进行构造：设 f_1, \dots, f_m 是 V^* 的一组基， g_1, \dots, g_n 是 W^* 的一组基，设（这里 $f_i(v)g_j(w)$ 即为 $f_i(v)$ 乘 $g_j(w)$ ）

$$\varphi_{ij}(v, w) = f_i(v)g_j(w), \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

则所有 φ_{ij} 构成 $T(V, W)$ 的一组基。

证明：

首先需要证明 φ_{ij} 的确是 V, W 上的双线性函数。对 $v_1, v_2 \in V, w \in W, \lambda, \mu \in \mathbb{K}$ ，有

$$\varphi_{ij}(\lambda v_1 + \mu v_2, w) = f_i(\lambda v_1 + \mu v_2)g_j(w) = \lambda f_i(v_1)g_j(w) + \mu f_i(v_2)g_j(w) = \lambda \varphi_{ij}(v_1, w) + \mu \varphi_{ij}(v_2, w)$$

对第二个分量可同理验证线性性。

由于所有 φ_{ij} 共有 mn 个，只需证明它们的线性无关性，即得到它们构成 $T(V, W)$ 一组基。

对一组 $\lambda_{ij} \in \mathbb{K}$ ，若 $\sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} \varphi_{ij} = 0$ ，由零映射定义可知

$$\forall v \in V, w \in W, \quad \sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} \varphi_{ij}(v, w) = 0$$

即

$$\forall v \in V, w \in W, \quad \sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} f_i(v)g_j(w) = 0$$

我们先固定 v 整理求和，这意味着（注意由于加法的交换、结合，求和符号可交换顺序，由于分配律可任意加括号）

$$\forall v \in V, w \in W, \quad \sum_{j=1}^n \left(\sum_{i=1}^m \lambda_{ij} f_i(v) \right) g_j(w) = 0$$

当 v 固定时，上式右侧括号中的项可看作固定的系数，从而将其看作映射 g_j 的线性组合可得

$$\forall v \in V, \quad \sum_{j=1}^n \left(\sum_{i=1}^m \lambda_{ij} f_i(v) \right) g_j = 0$$

由于 g_1, \dots, g_n 的线性无关性，这即能得到

$$\forall v \in V, \quad \forall j = 1, \dots, n, \quad \sum_{i=1}^m \lambda_{ij} f_i(v) = 0$$

接下来，将其看作映射 f_i 的线性组合，可得

$$\forall j = 1, \dots, n, \quad \sum_{i=1}^m \lambda_{ij} f_i = 0$$

由于 f_1, \dots, f_n 的线性无关性，对每个 j 考虑即可进一步得到

$$\forall j = 1, \dots, n, \quad \forall i = 1, \dots, m, \quad \lambda_{ij} = 0$$

从而得证。

* 注意证明的思路中，我们在不断进行**映射为 0** 与**映射在每点处都为 0** 两个等价概念的转化，这是由于 g_1, \dots, g_n 或 f_1, \dots, f_n 作为基的性质是定义在**映射**上的，而我们真正能处理的部分是考虑**映射**在每个点处的情况。

仿照此过程，我们也可以给出多重线性函数的一组基。对 $T(V_1, \dots, V_m)$ ，假设 $\dim V_i = n_i$, V_i^* 的一组基为 $f_i^{(1)}, \dots, f_i^{(n_i)}$, 定义

$$\varphi_{j_1 j_2 \dots j_m}(v_1, \dots, v_m) = f_1^{(j_1)}(v_1) f_2^{(j_2)}(v_2) \dots f_m^{(j_m)}(v_m)$$

这里每个 j_i 分别取遍 1 到 n_i , 则得到的所有 φ 构成 $T(V_1, \dots, V_m)$ 的一组基，从而其维数为 $n_1 n_2 \dots n_m$ 。

证明：

* 这个证明更严谨的写法是利用归纳，为了展示更真实的化简过程，此处用自顶向下递推的方式书写证明。

首先，与双线性函数时完全类似可验证这样定义的 $\varphi_{j_1 j_2 \dots j_m}$ 的确是多重线性函数。我们接下来说明它构成一组基。

- 线性无关性

设 $\lambda_{j_1 j_2 \dots j_m} \in \mathbb{R}$ 满足 (这里求和表示对 j_1 从 1 到 n_1 、 j_2 从 1 到 n_2 , 直到 j_m 从 1 到 n_m)

$$\sum_{j_1, j_2, \dots, j_m} \lambda_{j_1 j_2 \dots j_m} \varphi_{j_1 j_2 \dots j_m} = 0$$

则有

$$\forall v_1 \in V_1, \dots, v_m \in V_m, \quad \sum_{j_1, j_2, \dots, j_m} \lambda_{j_1 j_2 \dots j_m} \varphi_{j_1 j_2 \dots j_m}(v_1, \dots, v_m) = 0$$

用定义展开 $\varphi_{j_1 j_2 \dots j_m}(v_1, \dots, v_m)$, 我们先提取出 j_m 来，将其写成 (与之前类似，求和符号可交换，且由分配律可任意加括号)

$$\forall v_1 \in V_1, \dots, v_m \in V_m, \quad \sum_{j_m=1}^{n_m} \left(\sum_{j_1, j_2, \dots, j_{m-1}} \lambda_{j_1 j_2 \dots j_m} f_1^{(j_1)}(v_1) \dots f_{m-1}^{(j_{m-1})}(v_{m-1}) \right) f_m^{(j_m)}(v_m) = 0$$

固定 v_1, \dots, v_{m-1} , 将其看成所有 $f_m^{(j_m)}$ 进行线性组合的结果，可发现

$$\forall v_1 \in V_1, \dots, v_{m-1} \in V_{m-1}, \quad \sum_{j_m=1}^{n_m} \left(\sum_{j_1, j_2, \dots, j_{m-1}} \lambda_{j_1 j_2 \dots j_m} f_1^{(j_1)}(v_1) \dots f_{m-1}^{(j_{m-1})}(v_{m-1}) \right) f_m^{(j_m)} = 0$$

由于所有 $f_m^{(j_m)}$ 构成 V_m^* 的一组基，它们线性无关，因此所有系数应全为 0 才能保证映射为 0，也即 (这里 j_m 取值为 1 到 n_m)

$$\forall v_1 \in V_1, \dots, v_{m-1} \in V_{m-1}, \quad \forall j_m, \quad \sum_{j_1, j_2, \dots, j_{m-1}} \lambda_{j_1 j_2 \dots j_m} f_1^{(j_1)}(v_1) \dots f_{m-1}^{(j_{m-1})}(v_{m-1}) = 0$$

可以发现，固定每个 j_m , 上式的结构与开始的式子完全一致，只是从 m 个变量变为了 $m-1$ 个变量。由此，不断重复此过程，类似双线性函数时，最终可以得到

$$\forall j_1, \dots, j_m, \quad \lambda_{j_1 \dots j_m} = 0$$

这就是要证的线性无关性结论。

- 生成全空间

也即要证明，对任何 $\varphi \in T(V_1, \dots, V_m)$, 存在系数 $\mu_{j_1 j_2 \dots j_m}$ 使得

$$\varphi = \sum_{j_1, j_2, \dots, j_m} \mu_{j_1 j_2 \dots j_m} \varphi_{j_1 j_2 \dots j_m}$$

此系数的构造依赖之前证明的重要定理： V_i^* 的任何一组基一定可以看成 V_i 某组基的对偶基。由此，我们假设 $f_i^{(1)}, \dots, f_i^{(n_i)}$ 是 $\alpha_i^{(1)}, \dots, \alpha_i^{(n_i)}$ 的对偶基。

我们下面证明

$$\mu_{j_1 j_2 \dots j_m} = \varphi(\alpha_1^{(j_1)}, \alpha_2^{(j_2)}, \dots, \alpha_m^{(j_m)})$$

由于 φ 是多重线性的，对任何 v_1, \dots, v_m ，假设每个 v_i 在基 $\alpha_i^{(1)}, \dots, \alpha_i^{(n_i)}$ 下的表示为

$$v_i = \sum_{k_i=1}^{n_i} \xi_i^{(k_i)} \alpha_i^{(k_i)}$$

可以将 $\varphi(v_1, v_2, \dots, v_m)$ 展开为（这里求和表示对 k_1 从 1 到 n_1 、 k_2 从 1 到 n_2 ，直到 k_m 从 1 到 n_m ）

$$\begin{aligned} \varphi(v_1, \dots, v_m) &= \varphi\left(\sum_{k_1=1}^{n_1} \xi_1^{(k_1)} \alpha_1^{(k_1)}, \dots, \sum_{k_m=1}^{n_m} \xi_m^{(k_m)} \alpha_m^{(k_m)}\right) \\ &= \sum_{k_1, \dots, k_m} \xi_1^{(k_1)} \dots \xi_m^{(k_m)} \varphi(\alpha_1^{(k_1)}, \dots, \alpha_m^{(k_m)}) \end{aligned}$$

另一方面，利用已经证明的对偶基性质有

$$v_i = \sum_{k_i=1}^{n_i} f_i^{(k_i)}(v_i) \alpha_i^{(k_i)}$$

由此对比系数可知

$$\xi_i^{(k_i)} = f_i^{(k_i)}(v_i)$$

从而最终改写为

$$\varphi(v_1, \dots, v_m) = \sum_{k_1, \dots, k_m} f_1^{(k_1)}(v_1) \dots f_m^{(k_m)}(v_m) \varphi(\alpha_1^{(k_1)}, \dots, \alpha_m^{(k_m)})$$

由于这里的乘法都是数的乘法，可以任意交换次序，我们将最后一项交换到开头，并利用 $\varphi_{k_1 \dots k_m}$ 的定义即得

$$\varphi(v_1, \dots, v_m) = \sum_{k_1, \dots, k_m} \varphi(\alpha_1^{(k_1)}, \dots, \alpha_m^{(k_m)}) \varphi_{k_1 \dots k_m}(v_1, \dots, v_m)$$

由于此等式对任何 v_1, \dots, v_m 都成立，即得映射相等，于是

$$\varphi = \sum_{k_1, \dots, k_m} \varphi(\alpha_1^{(k_1)}, \dots, \alpha_m^{(k_m)}) \varphi_{k_1 \dots k_m}$$

这就证明了结论。

* 非常建议用 $m = 2$ 的特殊情况进行操作以理解这个非常复杂的证明过程。

到这里，我们得到了 $T(V_1, \dots, V_m)$ 的基与维数，事实上，上方的构造过程就是张量积的过程。

27.1.3 多重线性函数的张量积

我们将这种各自一组基两两匹配构成整体一组基的过程称为张量积。例如，在双线性函数的例子中，对于 $f \in V^*$, $g \in W^*$ ，我们定义 f 与 g 的张量积为 $\varphi \in T(V, W)$ ，使得

$$\varphi(v, w) = f(v)g(w)$$

记作 $\varphi = f \otimes g$ ，并记 $T(V, W) = V^* \otimes W^*$ 。这样，我们就可以将双线性函数的基构造过程写为，对 V^* 的一组基 f_1, \dots, f_m , W^* 的一组基 g_1, \dots, g_n ，所有 $f_i \otimes g_j$ 构成 $V^* \otimes W^*$ 的一组基。

为了将构造多重线性函数一组基的过程用张量积表示，我们还需要推广上面的定义。根据多重线性函数的定义，对偶空间 V^* 也可以写为 $T(V)$ ，由此可以定义：对 $f \in T(V_1, \dots, V_m)$, $g \in T(W_1, \dots, W_n)$ ，我们定义 f 与 g 的张量积为 $\varphi \in T(V_1, \dots, V_m, W_1, \dots, W_n)$ ，使得

$$\varphi(v_1, \dots, v_m, w_1, \dots, w_n) = f(v_1, \dots, v_m)g(w_1, \dots, w_n)$$

记作 $\varphi = f \otimes g$, 并记 $T(V_1, \dots, V_m, W_1, \dots, W_n) = T(V_1, \dots, V_m) \otimes T(W_1, \dots, W_n)$ 。我们需要证明, 若 $\dim V_i = r_i$ 、 $\dim W_j = s_j$, 当 f_1, \dots, f_M 为 $T(V_1, \dots, V_m)$ 一组基, g_1, \dots, g_N 为 $T(W_1, \dots, W_n)$ 一组基时, 所有 $f_i \otimes g_j$ 构成 $T(V_1, \dots, V_m) \otimes T(W_1, \dots, W_n)$ 的一组基。

证明:

与双线性函数时完全类似可验证这样定义的 $f_i \otimes g_j$ 的确是多重线性函数。之前已经得到了多重线性函数空间的维数结论, 即

$$M = \dim T(V_1, \dots, V_m) = r_1 \dots r_m$$

$$N = \dim T(W_1, \dots, W_n) = s_1 \dots s_n$$

$$\dim T(V_1, \dots, V_m, W_1, \dots, W_n) = r_1 \dots r_m s_1 \dots s_n = MN$$

由此, 所有 $f_i \otimes g_j$ 的确与 $T(V_1, \dots, V_m, W_1, \dots, W_n)$ 维数相同, 只需证明线性无关性就可得到它们构成其一组基。证明过程与双线性函数是类似的: 对一组 $\lambda_{ij} \in \mathbb{K}$, 若 $\sum_{i=1}^M \sum_{j=1}^N \lambda_{ij} f_i \otimes g_j = 0$, 由零映射定义可知 (这里任意中 i 取 1 到 m , j 取 1 到 n)

$$\forall v_i \in V_i, w_j \in W_j, \quad \sum_{i=1}^M \sum_{j=1}^N \lambda_{ij} f_i(v_1, \dots, v_m) g_j(w_1, \dots, w_n) = 0$$

固定 v_1, \dots, v_m 整理可得

$$\forall v_i \in V_i, w_j \in W_j, \quad \sum_{j=1}^N \left(\sum_{i=1}^M \lambda_{ij} f_i(v_1, \dots, v_m) \right) g_j(w_1, \dots, w_n) = 0$$

与之前类似将其上方看作 w_1, \dots, w_n 上的多重线性函数, 利用所有 g_j 的线性无关性即得

$$\forall v_i \in V_i, \quad \forall j = 1, \dots, N, \quad \sum_{i=1}^M \lambda_{ij} f_i(v_1, \dots, v_m) = 0$$

进一步固定每个 j , 将上方看作 v_1, \dots, v_m 上的多重线性函数, 利用所有 f_i 的线性无关性即得

$$\forall j = 1, \dots, N, \quad \forall i = 1, \dots, M, \quad \lambda_{ij} = 0$$

这就得到了证明。

* 事实上此定理中的“生成全空间”可以不用维数结论, 而是直接类似之前构造证明, 这样即可以通过更简单的方式直接证明之前对 $T(V_1, \dots, V_m)$ 的基的构造: 通过归纳可知 (根据上方定义可直接发现张量积具有结合律, 从而可以写出连乘)

$$T(V_1, \dots, V_m) = V_1^* \otimes V_2^* \otimes \cdots \otimes V_m^*$$

且 V_1^* 一组基与 V_2^* 一组基作张量积 (这里“作张量积”指对第一组基的任何一个与第二组的任何一个作) 可构成 $V_1^* \otimes V_2^*$ 的一组基, 再与 V_3^* 一组基作张量积可构成 $V_1^* \otimes V_2^* \otimes V_3^*$ 的一组基, 以此再进行归纳即得到结论。

最后, 我们来聊一聊何为“张量”。粗略来说, 一个 m 阶张量就是一个 m 维的数组, 具有 m 个下标 (忽略更高级的协变/反变等内容)。从这个角度来说, 矩阵就是一个二阶的张量。

在上节构造 $T(V_1, \dots, V_m)$ 基的过程中, 我们得到了等式

$$\varphi = \sum_{k_1, \dots, k_m} \mu_{k_1 k_2 \dots k_m} \varphi_{k_1 \dots k_m}, \quad \mu_{k_1 k_2 \dots k_m} = \varphi(\alpha_1^{(k_1)}, \dots, \alpha_m^{(k_m)})$$

这里组合的系数 $\mu_{k_1 k_2 \dots k_m}$ 就是一个 m 阶张量, 各个维度分别为 n_1, n_2, \dots, n_m 。可以发现, 在 $m = 2$ 时, $\mu_{k_1 k_2}$ 的定义即为 $\varphi(\alpha_1^{(k_1)}, \alpha_2^{(k_2)})$, 从而是双线性函数的度量矩阵。由此, 我们可以称这个 m 阶张量为多重线性函数的度量张量。

由此等式，我们可以发现 φ 由所有 $\varphi(\alpha_1^{(k_1)}, \dots, \alpha_m^{(k_m)})$ 确定，也即度量张量可以确定多重线性函数，这就对应之前说的基映射确定双线性函数。

与矩阵构成线性空间类似，定义加法为分量对应相加，数乘为每个分量数乘，所有 \mathbb{K} 上每个维度为 n_1, n_2, \dots, n_m 的 m 阶张量也构成线性空间，记为 $\mathbb{K}^{n_1 \times n_2 \times \dots \times n_m}$ 。由于每个 φ 可以定义唯一一个度量张量，每个度量张量也可以确定 φ ， $\mathbb{K}^{n_1 \times n_2 \times \dots \times n_m}$ 与 $T(V_1, \dots, V_m)$ 可以建立双射。事实上，正如度量矩阵是线性同构，通过更复杂的线性性验证可以证明这个双射也是一个线性同构。

* 从这条路径出发，也可以从 $\mathbb{K}^{n_1 \times n_2 \times \dots \times n_m}$ 的一组基（所有分量中仅有一个为 1，其他为 0 的张量）构造 $T(V_1, \dots, V_m)$ 的一组基，得到的基事实上完全相同。

至此，我们得到了多重线性函数空间张量积的定义。不过，并不是所有的线性空间都容易看作某个多重线性函数空间，因此，我们需要其他策略对一般线性空间的张量积进行定义。

§27.2 线性空间的张量积

27.2.1 张量积定义

对于一般的有限维线性空间，其上的张量积需要满足什么要求呢？假设我们需要定义 \mathbb{K} 上的 m 维线性空间 V 与 n 维线性空间 W 的张量积，从之前的过程来看，我们首先需要一个 \mathbb{K} 上的线性空间 U 来“接收”向量的张量积结果，并对所有 $x \in V, y \in W$ 定义

$$x \otimes y \in U$$

* 由此张量积 \otimes 是 $V \times W \rightarrow U$ 的映射，下方要求了它满足双线性性，因此可以称为一个双线性映射。

此外，由于是线性空间上的运算，我们希望张量积具有双线性性，也即对任何 $x_1, x_2, x \in V, y_1, y_2, y \in W, \lambda, \mu \in \mathbb{K}$ 有（仍然注意左右的加法、数乘为不同线性空间中的）

$$(\lambda x_1 + \mu x_2) \otimes y = \lambda(x_1 \otimes y) + \mu(x_2 \otimes y)$$

$$x \otimes (\lambda y_1 + \mu y_2) = \lambda x \otimes y_1 + \mu x \otimes y_2$$

最后，也是最重要的要求，我们希望张量积能够构造基，也即若 x_1, \dots, x_m 是 V 的一组基， y_1, \dots, y_n 是 W 的一组基，所有

$$x_i \otimes y_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

应构成 U 的一组基（之后我们将其称为张量积的基性质）。

若线性空间 U 与运算 \otimes 满足上述的性质，我们即记 $U = V \otimes W$ ，称 U 为 V 与 W 的一个张量积空间。

* 上一节定义的张量积并未验证线性性，实际上的确满足：对 V^*, W^* 上的张量积，由映射数乘定义验证对任何 $f_1, f_2 \in V^*, g \in W^*, \lambda, \mu \in \mathbb{K}$ 有

$$\forall v \in V, w \in W, \quad (\lambda f_1 + \mu f_2)(v)g(w) = \lambda f_1(v)g(w) + \mu f_2(v)g(w)$$

另一边对称可证，从而线性性得证。对 $T(V_1, \dots, V_m), T(W_1, \dots, W_n)$ 上的张量积，证明完全类似。

张量积空间并不唯一，所有与 U 同构的线性空间都可以成为 V 与 W 的张量积空间。

证明：

若 $U = V \otimes W$ ， U' 与 U 同构， U 到 U' 的同构映射为 ψ ，我们定义映射 $\otimes' : V \times W \rightarrow U'$ ：

$$x \otimes' y = \psi(x \otimes y)$$

利用同构映射的线性性可以验证 \otimes' 的线性性，且由于同构将一组基映射到一组基，若 x_1, \dots, x_m 是 V 一组基， y_1, \dots, y_n 是 W 的一组基，由于所有 $x_i \otimes y_j$ 构成 U 一组基，所有 $\psi(x_i \otimes y_j)$ 应构成 U' 一组基，从而得证 \otimes' 的确为张量积，于是 $U' = V \otimes' W$ 。

不过，由于要求了基的个数为 mn ，必然有所有 V 与 W 的张量积空间彼此同构。

27.2.2 向量空间的张量积

既然我们讨论的是有限维线性空间，利用同构标准形的思想，我们还是希望对向量空间进行处理，不妨考虑 \mathbb{K}^m 与 \mathbb{K}^n 。

利用矩阵论知识，我们可以给出一种合理的定义方式：对任何 $x \in \mathbb{K}^m$ 、 $y \in \mathbb{K}^n$ ，定义

$$x \otimes y = xy^T \in \mathbb{K}^{m \times n}$$

则此定义下 $\mathbb{K}^{m \times n} = \mathbb{K}^m \otimes \mathbb{K}^n$ 。

证明：

直接计算知对任何 $x_1, x_2, x \in \mathbb{K}^m$ 、 $y_1, y_2, y \in \mathbb{K}^n$ 、 $\lambda, \mu \in \mathbb{K}$ 有

$$(\lambda x_1 + \mu x_2)y^T = \lambda x_1 y^T + \mu x_2 y^T$$

$$x(\lambda y_1 + \mu y_2)^T = \lambda xy_1^T + \mu xy_2^T$$

从而双线性性成立。

若 x_1, \dots, x_m 为 \mathbb{K}^m 一组基， y_1, \dots, y_n 为 \mathbb{K}^n 一组基，由于 $\dim \mathbb{K}^{m \times n} = mn$ ，只需证明

$$x_i y_j^T, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

线性无关即得到它们为一组基。

设 $\lambda_{ij} \in \mathbb{K}$ 使得

$$\sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} x_i y_j^T = O$$

交换求和次序并利用分配律得到

$$\sum_{j=1}^n \left(\sum_{i=1}^m \lambda_{ij} x_i \right) y_j^T = O$$

直接计算可发现左侧矩阵的第 k 行为 ($x_{i,k}$ 表示 x_i 的第 k 个分量)

$$\sum_{j=1}^n \left(\sum_{i=1}^m \lambda_{ij} x_{i,k} \right) y_j^T$$

由于其为 0，作转置得到

$$\sum_{j=1}^n \left(\sum_{i=1}^m \lambda_{ij} x_{i,k} \right) y_j = 0$$

由 y_j 之间的线性无关性，此式为 0 当且仅当对任何 k 系数都为 0，重新拼合得到

$$\forall j = 1, \dots, n, \quad \sum_{i=1}^m \lambda_{ij} x_i = 0$$

对每个 j ，由 x_i 之间的线性无关性即得到所有 λ_{ij} 为 0，得证。

不过，上述定义存在一个很大的麻烦：向量空间张量积的结果不再是向量空间，这样，如果我们希望考虑 $x \otimes y \otimes z$ 这样的式子，还需要重新定义矩阵与向量的张量积。为了让结果仍然是向量空间，我们需要一个看起来略显复杂的定义方式：对任何 $x \in \mathbb{K}^m$ 、 $y \in \mathbb{K}^n$ ，定义

$$x \otimes y = \begin{pmatrix} x_1 y \\ x_2 y \\ \vdots \\ x_m y \end{pmatrix} \in \mathbb{K}^{mn}$$

我们也可以证明，此定义下 $\mathbb{K}^{mn} = \mathbb{K}^m \otimes \mathbb{K}^n$ 。我们将此定义称为两个列向量的克罗内克积，之后大部分情况下谈论列向量的张量积都遵循此定义。

证明：

我们只需要构造 $\mathbb{K}^{m \times n}$ 到 \mathbb{K}^{mn} 的同构 ϕ 使得 $\phi(xy^T) = x \otimes y$ ，即可与本讲义 27.2.1 最后相同证明 $x \otimes y$ 为张量积。

直接构造 ϕ 满足（设 A 各个分量为 a_{ij} ）

$$\phi(A) = (a_{11}, a_{12}, \dots, a_{1n}, a_{21}, \dots, a_{2n}, \dots, a_{m1}, \dots, a_{mn})^T$$

可直接计算得到 $\phi(xy^T) = x \otimes y$ ，且由于 $\phi(A)$ 每个分量都直接为 A 某个分量，利用矩阵加法、数乘定义可知这的确是线性映射。此外， $\phi(A)$ 出现了 A 所有分量，因此 $\phi(A) = 0$ 必须 $A = O$ ，这就说明了 $\text{Ker } \phi = \{O\}$ ，由两侧维数相同可知其为同构，得证。

* 此证明的直观理解是，克罗内克积的定义只是在张量积为矩阵后重新排列了顺序，因此对应的线性相关、线性无关性不变，仍可以保持基的性质。

从形式中可以看出，克罗内克积不可交换，即 $x \otimes y$ 一般不等于 $y \otimes x$ 。不过，它具有重要的结合律性质，即对 $x \in \mathbb{K}^m$ 、 $y \in \mathbb{K}^n$ 、 $z \in \mathbb{K}^p$ 有

$$(x \otimes y) \otimes z = x \otimes (y \otimes z)$$

证明：

直接计算验证可发现 $x \in \mathbb{K}^m$ 、 $y \in \mathbb{K}^n$ 时 $x \otimes y$ 可写为所有

$$x_i y_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

按照先 j 后 i 的顺序排列成的列向量，也即

$$(x_1 y_1, x_1 y_2, \dots, x_1 y_n, x_2 y_1, \dots, x_2 y_n, \dots, x_m y_1, \dots, x_m y_n)^T$$

由此进一步计算得到无论 $(x \otimes y) \otimes z$ 还是 $x \otimes (y \otimes z)$ 都是所有

$$x_i y_j z_k, \quad i = 1, \dots, m, \quad j = 1, \dots, n, \quad k = 1, \dots, p$$

按照先 k 、然后 j 、最后 i 的顺序排列成的列向量，从而相等。

我们之后将利用克罗内克积定义一般线性空间的张量积。此外，它的很多性质也能代表张量积空间的性质，例如：

- 只要 $m > 1$ 且 $n > 1$ ，即得并非所有 \mathbb{K}^{mn} 中的向量都能写为 $x \otimes y$ ， $x \in \mathbb{K}^m$ 、 $y \in \mathbb{K}^n$ 的形式。

证明：

我们考虑 \mathbb{K}^m 的一组基 x_1, \dots, x_m 、 \mathbb{K}^n 的一组基 y_1, \dots, y_n , 由条件可考虑 $x_1 \otimes y_1 + x_2 \otimes y_2$, 若存在 $u \in \mathbb{K}^m$ 、 $v \in \mathbb{K}^n$ 使得

$$x_1 \otimes y_1 + x_2 \otimes y_2 = u \otimes v$$

设 $u = \sum_{i=1}^m u_i x_i$ 、 $v = \sum_{j=1}^n v_j y_j$, 这里 $u_i, v_j \in \mathbb{K}$, 利用线性性可展开得

$$u \otimes v = \sum_{i=1}^m \sum_{j=1}^n u_i v_j x_i \otimes y_j$$

由于所有 $x_i \otimes y_j$ 构成一组基, 左右系数应相同, 左侧 $x_1 \otimes y_1$ 、 $x_2 \otimes y_2$ 前的系数为 1, $x_1 \otimes y_2$ 前的系数为 0, 因此 $u_1 v_1 = u_2 v_2 = 1$, $u_1 v_2 = 0$, 但 $u_1 v_1 = u_2 v_2 = 1$ 能推出这四个数均不是 0, 矛盾, 从而得证不存在。

2. 若 $x \otimes y = 0$, $x \in \mathbb{K}^m$ 、 $y \in \mathbb{K}^n$, 则 x 与 y 至少一个为零向量。

证明:

若 x, y 均不为 0, 将 x 扩充为 \mathbb{K}^m 一组基、 y 扩充为 \mathbb{K}^n 一组基, 则 $x \otimes y$ 根据张量积定义是 \mathbb{K}^{mn} 一组基的一部分, 矛盾, 从而只能 x 或 y 为零向量。

* 从中可以发现讨论线性无关向量的张量积常采用扩充为基的方式进行讨论。

* 上述两个结论的证明过程事实上只用到了一组基与一组基张量积得到一组基, 因此对一般线性空间的张量积也成立。更多向量张量积的性质可见期末复习题最后一题。

27.2.3 张量积的一般性质

至此, 向量空间的张量积已经定义完成, 而对于一般的 \mathbb{K} 上有限维线性空间 V 与 W , 我们还是设法将其转移到向量空间讨论: 若 V, W 分别为 m, n 维, V 的一组基为 S 、 W 的一组基为 T , 我们可以对任何 $v \in V, w \in W$ 定义 (下标 S, T 代表坐标)

$$v \otimes w = v_S \otimes w_T \in \mathbb{K}^{mn}$$

这里第二个 \otimes 为列向量的克罗内克积。在此定义下, 的确有 $\mathbb{K}^{mn} = V \otimes W$, 这就定义了任何两个有限维线性空间的一个张量积 (虽然张量积并不唯一, 我们如此定义至少说明了其存在性)。

证明:

利用坐标映射的线性性与线性映射的复合还是线性映射可知这样定义的 $v \otimes w$ 是双线性映射。

由于坐标映射的同构性, 若 v_1, \dots, v_m 构成 V 的一组基, $(v_1)_S, \dots, (v_m)_S$ 构成 \mathbb{K}^m 的一组基; 同理, 若 w_1, \dots, w_n 构成 W 的一组基, $(w_1)_T, \dots, (w_n)_T$ 构成 \mathbb{K}^n 的一组基, 从而再由克罗内克积的基性质可知所有 $v_i \otimes w_j$ 构成 \mathbb{K}^{mn} 的一组基, 即得证。

事实上, 哪怕不这么定义, 张量积空间也与向量克罗内克积密切相关: 设 $U = V \otimes W$, 并假设 V 一组基为 $S = (\alpha_1, \dots, \alpha_m)$, W 一组基为 $T = (\beta_1, \dots, \beta_n)$, 考虑 U 的一组基 (可记作 $S \otimes T$)

$$(\alpha_1 \otimes \beta_1, \alpha_1 \otimes \beta_2, \dots, \alpha_1 \otimes \beta_n, \alpha_2 \otimes \beta_1, \dots, \alpha_2 \otimes \beta_n, \dots, \alpha_m \otimes \beta_1, \dots, \alpha_m \otimes \beta_n)$$

对任何 $v \in V, w \in W$, 应有 $v \otimes w$ 在 $S \otimes T$ 下的坐标为 $v_S \otimes w_T$, 这里坐标的 \otimes 为列向量的克罗内克积。

证明:

由基的性质可设 $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_n \in \mathbb{K}$ 使得

$$v = \sum_{i=1}^m \lambda_i \alpha_i, \quad w = \sum_{j=1}^n \mu_j \beta_j$$

于是由坐标定义

$$v_S = (\lambda_1, \dots, \lambda_m)^T, \quad w_T = (\mu_1, \dots, \mu_n)^T$$

利用张量积的双线性性可以直接展开计算得到

$$v \otimes w = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \alpha_i \otimes \beta_j$$

由此利用坐标定义可以直接得到它在基 $S \otimes T$ 下的坐标为

$$(\lambda_1 \mu_1, \lambda_1 \mu_2, \dots, \lambda_1 \mu_n, \lambda_2 \mu_1, \dots, \lambda_2 \mu_n, \dots, \lambda_m \mu_1, \dots, \lambda_m \mu_n)^T$$

可发现其恰好是

$$\begin{pmatrix} \lambda_1 \mu \\ \lambda_2 \mu \\ \vdots \\ \lambda_m \mu \end{pmatrix} = v_S \otimes w_T$$

从而得证。

最后，我们来聊聊子空间与多个线性空间张量积的情况。我们给出如下两个定理以介绍其性质（以下假设 U, V, W, X 为 \mathbb{K} 上的有限维线性空间，当我们写出 $U = V \otimes W$ 时，默认运算 \otimes 已经定义）：

- 若 $U = V \otimes W$ ，对 V 的子空间 V_0 与 W 的子空间 W_0 ，记 U_0 为所有 V_0 中元素与 W_0 中元素张量积生成的线性空间，即

$$U_0 = \langle v_0 \otimes w_0 \mid v_0 \in V_0, w_0 \in W_0 \rangle$$

定义 \otimes_0 满足对任何 $v_0 \in V_0, w_0 \in W_0$ 有

$$v_0 \otimes_0 w_0 = v_0 \otimes w_0 \in U_0$$

则 $U_0 = V_0 \otimes_0 W_0$ 。

证明：

首先， U_0 作为生成子空间，的确是一个线性空间，且包含所有 $v_0 \otimes_0 w_0 = v_0 \otimes w_0$ 的结果。此外，由于 \otimes_0 的定义保持了 \otimes 的定义，类似限制映射时验证可知其满足线性性，下面证明其基性质：设 v_1, \dots, v_r 为 V_0 的一组基， w_1, \dots, w_s 是 W_0 的一组基，我们证明

$$v_i \otimes_0 w_j, \quad i = 1, \dots, r, \quad j = 1, \dots, s$$

构成 U_0 的一组基。

- 线性无关性

将 v_1, \dots, v_r 扩充为 V 的一组基 v_1, \dots, v_m 、 w_1, \dots, w_s 扩充为 W 的一组基 w_1, \dots, w_n ，则

$$v_i \otimes w_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

构成 U 的一组基。由于

$$v_i \otimes w_j, \quad i = 1, \dots, r, \quad j = 1, \dots, s$$

是这组基的一部分，它们必然线性无关。而 $v_i \otimes_0 w_j = v_i \otimes w_j$ 对 $i = 1, \dots, r, j = 1, \dots, s$ 成立，从而线性无关性得证。

- 生成全空间

利用生成关系的传递性，只需证明此向量组能生成全部 $\{v_0 \otimes w_0 \mid v_0 \in V_0, w_0 \in W_0\}$ ，即可证明它们能生成 U_0 。对任何 $v_0 \in V_0, w_0 \in W_0$ 设（所有 $\lambda_i, \mu_j \in \mathbb{K}$ ）

$$v_0 = \sum_{i=1}^r \lambda_i v_i, \quad w_0 = \sum_{j=1}^s \mu_j w_j$$

直接由 \otimes_0 的双线性性可知

$$v_0 \otimes w_0 = v_0 \otimes_0 w_0 = \sum_{i=1}^r \sum_{j=1}^s \lambda_i \mu_j v_i \otimes_0 w_j$$

这就证明了 $v_0 \otimes w_0$ 可由此向量组表出，从而得证。

* 此定义中的 \otimes_0 本质上是 \otimes 的限制映射，于是张量积的限制仍然是张量积。

2. 若 $U = (V \otimes W) \otimes X$ (注意这里左右的 \otimes 不相同)，任取 V 的一组基 v_1, \dots, v_m 、 W 的一组基 w_1, \dots, w_n 、 X 的一组基 x_1, \dots, x_p ，则所有

$$(v_i \otimes w_j) \otimes x_k, \quad i = 1, \dots, m, \quad j = 1, \dots, n, \quad k = 1, \dots, p$$

构成 U 的一组基。

证明：

利用 V 与 W 张量积的基性质，所有

$$v_i \otimes w_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

构成 $V \otimes W$ 的一组基。将 $v_i \otimes w_j$ 记为 y_{ij} ，利用 $V \otimes W$ 与 X 张量积的基性质，对 U 的一组基 y_{ij} 与 X 的一组基 x_1, \dots, x_p ，所有

$$y_{ij} \otimes x_k, \quad i = 1, \dots, m, \quad j = 1, \dots, n, \quad k = 1, \dots, p$$

构成 U 的一组基。再代入 y_{ij} 的表达式得证。

* 注意由于运算可能有不同定义，不考虑同构时一般的张量积不具有结合律，但事实上由于彼此同构仍然可视为结合。

* 此结论对于更多个线性空间也成立，这与多重线性函数的基构造方式一致。

27.2.4 泛性质定义

* 本部分仅作为介绍，不要求掌握。

本节中介绍的张量积定义依赖取基，从而在无穷维依赖选择公理，并不利于在无穷维进行推广。为了能在更一般的情况下定义张量积，我们这里给出其泛性质定义。

首先定义双线性映射：给定 \mathbb{K} 上线性空间 V, W, U ， $V \otimes W$ 到 U 的映射 τ 满足对任何 $v_1, v_2, v \in V$ 、 $w_1, w_2, w \in W$ 、 $\lambda, \mu \in \mathbb{K}$ 有

$$\tau(\lambda v_1 + \mu v_2, w) = \lambda \tau(v_1, w) + \mu \tau(v_2, w)$$

$$\tau(v, \lambda w_1 + \mu w_2) = \lambda \tau(v, w_1) + \mu \tau(v, w_2)$$

则称其为 V, W 到 U 的双线性映射。

* 与双线性函数的唯一区别是将右侧从数集 \mathbb{K} 变为了一般的线性空间 U 。

给定 \mathbb{K} 上的线性空间 V, W ，若 \mathbb{K} 上的线性空间 U 与 V, W 到 U 的双线性映射 \otimes (我们仍然记为 $v \otimes w$ 而非 $\otimes(v, w)$ 以保持与之前形式一致) 满足，对任何 \mathbb{K} 上线性空间 X ，给定任何 V, W 到 X 的双线性映射 τ ，都存在唯一 U 到 X 的线性映射 \mathcal{A} 使得

$$\forall v \in V, w \in W, \quad \tau(v, w) = \mathcal{A}(v \otimes w)$$

则记 $U = V \otimes W$ ，称 U 为 V 与 W 的一个张量积空间。

上述关于双线性映射的性质称为张量积的泛性质，对无穷维也可如此定义。我们最后来证明对有限维线性空间，泛性质定义与之前给出的定义等价。

证明：

由于双线性性在两个定义中都有要求，我们事实上只需要证明泛性质与对基的要求相互等价。假设 V 为 m 维线性空间， W 为 n 维线性空间。

- 原定义推泛性质定义

我们只需对给定的 τ 证明存在唯一的 \mathcal{A} 即可。设 V 的一组基为 v_1, \dots, v_m 、 W 的一组基为 w_1, \dots, w_n 。对任何 $v \in V$ 、 $w \in W$ ，假设它们在对应基下表示为（所有 $\lambda_i, \mu_j \in \mathbb{K}$ ）

$$v = \sum_{i=1}^m \lambda_i v_i, \quad w = \sum_{j=1}^n \mu_j w_j$$

利用 τ 的双线性性可直接展开得到

$$\tau(v, w) = \tau\left(\sum_{i=1}^m \lambda_i v_i, \sum_{j=1}^n \mu_j w_j\right) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \tau(v_i, w_j)$$

另一方面，假设 \mathcal{A} 满足要求，利用张量积的双线性性与 \mathcal{A} 的线性性可知

$$\mathcal{A}(v, w) = \mathcal{A}\left(\sum_{i=1}^m \lambda_i v_i \otimes \sum_{j=1}^n \mu_j w_j\right) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \mathcal{A}(v_i \otimes w_j)$$

由于两式相等对任何 λ_i, μ_j 成立，考虑所有 λ_i 只有 λ_{i_0} 取 1、其他取 0；所有 μ_j 只有 μ_{j_0} 取 1、其他取 0，计算可得

$$\tau(v_{i_0}, w_{j_0}) = \mathcal{A}(v_{i_0} \otimes w_{j_0})$$

由此将 i_0 取遍 1 到 m 、 j_0 取遍 1 到 n 可知上式等价于

$$\mathcal{A}(v_i \otimes w_j) = \tau(v_i, w_j), \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

由于所有 $v_i \otimes w_j$ 构成 U 的一组基，上方即给出了基映射，利用基映射确定线性映射可知 \mathcal{A} 存在唯一。

- 泛性质定义推原定义

也即需要证明，任取 V 的一组基 $S = (v_1, \dots, v_m)$ 、 W 的一组基 $T = (w_1, \dots, w_n)$ ，则

$$v_i \otimes w_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

构成 U 的一组基。

* 这里的证明思路是通过泛性质直接由同构性证明。

考虑 V, W 到 \mathbb{K}^{mn} 的双线性映射

$$\phi(v, w) = v_S \otimes w_T$$

这里坐标的 \otimes 为列向量的克罗内克积。

由定义，存在唯一的线性映射 $\mathcal{A}: U \rightarrow \mathbb{K}^{m \times n}$ 使得

$$\forall v \in V, w \in W, \quad \phi(v, w) = \mathcal{A}(v \otimes w)$$

另一方面，由于已经证明了 ϕ 是一个符合我们定义的张量积，且原定义可以推出泛性质定义，将 $v \otimes w$ 看作 V, W 到 U 的线性映射，则存在唯一的线性映射 $\mathcal{B}: \mathbb{K}^{m \times n} \rightarrow \mathbb{K}^{mn}$ 使得

$$\forall v \in V, w \in W, \quad v \otimes w = \mathcal{B}(\phi(v, w))$$

由此可知

$$\forall v \in V, w \in W, \quad v \otimes w = \mathcal{B}\mathcal{A}(v \otimes w)$$

接下来的操作非常具有技巧性：在泛性质定义中设 $\tau(v, w) = v \otimes w$ ，则存在唯一的线性映射 $\mathcal{C} : U \rightarrow U$ 使得

$$\forall v \in V, w \in W, \quad v \otimes w = \mathcal{C}(v \otimes w)$$

利用唯一性，必然 $\mathcal{B}\mathcal{A} = \mathcal{C}$ 。而 $U \rightarrow U$ 的恒等映射 \mathcal{I} 满足上式，从而再由唯一性 $\mathcal{C} = \mathcal{I}$ ，最终得到

$$\mathcal{B}\mathcal{A} = \mathcal{I}$$

同理，由于 $\phi(v, w) = v_S \otimes w_T$ 也满足张量积的泛性质，也可以证明 $\mathcal{A}\mathcal{B}$ 是 \mathbb{K}^{mn} 上的恒等映射，由此即得

$$\mathcal{B} = \mathcal{A}^{-1}$$

从而 \mathcal{A} 、 \mathcal{B} 均为线性同构。

最后，我们从同构出发证明它们目标向量组确实构成一组基。由于 $\phi(v, w) = \mathcal{A}(v \otimes w)$ ，直接计算可发现（这里上标 (m) 表示 m 维列向量， (n) 同理）

$$v_i \otimes w_j = \mathcal{B}(\phi(v_i, w_j)) = \mathcal{B}(e_i^{(m)} \otimes e_j^{(n)}), \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

由于向量的克罗内克积是向量空间的张量积，所有

$$e_i^{(m)} \otimes e_j^{(n)}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

构成 \mathbb{K}^{mn} 一组基，再由同构将一组基映射为一组基即得证。

*“泛性质”大致可以理解为某种“更本质的性质”，例如这里的本质就体现在它不依赖基进行定义。

§27.3 线性映射的张量积

27.3.1 良好定义性

正如我们之前研究的过程，既然单个线性空间可以进行张量积，如果线性空间之间进行了结构的关联，也即定义了线性映射，对应的关联能否传递到张量积空间中呢？

具体来说，对 \mathbb{K} 上的有限维线性空间 V_1, W_1, V_2, W_2 ，若 $\mathcal{A} \in \text{Hom}(V_1, W_1)$ 、 $\mathcal{B} \in \text{Hom}(V_2, W_2)$ ，若线性映射 $\mathcal{C} \in \text{Hom}(V_1 \otimes V_2, W_1 \otimes W_2)$ 满足

$$\forall v_1 \in V_1, v_2 \in V_2, \quad \mathcal{C}(v_1 \otimes v_2) = (\mathcal{A}v_1) \otimes (\mathcal{B}v_2)$$

则称其为 \mathcal{A} 与 \mathcal{B} 的张量积映射，记作 $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$ 。

不过，正如之前在向量空间张量积时已经证明的，所有 $v_1 \otimes v_2$ 往往并不是 $V_1 \otimes V_2$ 中的全部向量，因此，我们需要单独验证这样的 \mathcal{C} 的存在唯一性。

证明：

设 $V_1 \times V_2 \rightarrow W_1 \otimes W_2$ 的映射 τ 满足 $\tau(v_1, v_2) = (\mathcal{A}v_1) \otimes (\mathcal{B}v_2)$ ，利用线性映射的复合是线性映射与张量积是双线性映射可知 τ 是双线性映射，从而由张量积的泛性质可得存在唯一的 \mathcal{C} 满足要求。

* 由此可以看出泛性质定义某种意义上是为了线性映射张量积可以定义。

* 当 \mathcal{A} 是 V_1 上的线性变换、 \mathcal{B} 是 V_2 上的线性变换时，根据定义， $\mathcal{A} \otimes \mathcal{B}$ 即为 $V_1 \otimes V_2$ 上的线性变换，这称为线性变换的张量积。

此外，既然是有限维线性空间的线性映射，我们也要关心 $\mathcal{A} \otimes \mathcal{B}$ 的矩阵表示。不过，由于张量积可能有多种不同定义，只有与具体定义无关时才能谈论矩阵表示。我们考虑如下的情况：

设 V_1, W_1, V_2, W_2 的一组基分别为 S_1, T_1, S_2, T_2 ，且映射 \mathcal{A} 在 S_1, T_1 下的矩阵表示为 A ，映射 \mathcal{B} 在 S_2, T_2 下的矩阵表示为 B 。我们来计算 $\mathcal{A} \otimes \mathcal{B}$ 在 $S_1 \otimes S_2, T_1 \otimes T_2$ 下的矩阵表示 C ($S_1 \otimes S_2$ 与 $T_1 \otimes T_2$ 的定义见本讲义 27.2.3)。

解答：

设 V_1, W_1, V_2, W_2 的维数分别为 n, m, q, p , 且

$$S_1 = (v_1, \dots, v_n), \quad T_1 = (w_1, \dots, w_m), \quad S_2 = (u_1, \dots, u_q), \quad T_2 = (x_1, \dots, x_p)$$

并设 A 的各分量为 a_{ij} 、 B 的各分量为 b_{kl} , 则由定义有

$$\mathcal{A}(v_j) = \sum_{i=1}^m a_{ij} w_i$$

$$\mathcal{B}(u_l) = \sum_{k=1}^p b_{kl} x_k$$

从而利用双线性性展开有

$$(\mathcal{A} \otimes \mathcal{B})(v_j \otimes u_l) = \mathcal{A}(v_j) \otimes \mathcal{B}(u_l) = \sum_{i=1}^m \sum_{k=1}^p a_{ij} b_{kl} w_i \otimes x_k$$

由于基 $T_1 \otimes T_2$ 为

$$(w_1 \otimes x_1, w_1 \otimes x_2, \dots, w_1 \otimes x_p, w_2 \otimes x_1, \dots, w_2 \otimes x_p, \dots, w_m \otimes x_1, \dots, w_m \otimes x_p)$$

$(\mathcal{A} \otimes \mathcal{B})(v_j \otimes u_l)$ 在这组基下的坐标为

$$(a_{1j} b_{1l}, a_{1j} b_{2l}, \dots, a_{1j} b_{pl}, a_{2j} b_{1l}, \dots, a_{2j} b_{pl}, \dots, a_{mj} b_{1l}, \dots, a_{mj} b_{pl})^T$$

可发现其即为 $(a_{1j}, \dots, a_{mj})^T \otimes (b_{1l}, \dots, b_{pl})^T$, 再按照 $S_1 \otimes S_2$ 的顺序拼成矩阵可发现即为

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

也即分块为 $m \times n$ 块, 每个分块为 $a_{ij}B$ 的 $(mp) \times (nq)$ 阶矩阵。

我们将这样形式的 C 记为 $A \otimes B$, 称为矩阵的克罗内克积。可以发现, 之前定义的列向量的克罗内克积是矩阵克罗内克积的特殊情况。

27.3.2 克罗内克积

本章的最后, 我们将给出矩阵克罗内克积的一些性质, 以此说明线性映射与线性变换张量积的相抵/相似性质。

首先, 最重要的是结合律与乘法性质:

$$\forall A \in \mathbb{K}^{m \times n}, B \in \mathbb{K}^{p \times q}, C \in \mathbb{K}^{r \times s}, \quad (A \otimes B) \otimes C = A \otimes (B \otimes C)$$

$$\forall A \in \mathbb{K}^{m \times n}, B \in \mathbb{K}^{n \times p}, C \in \mathbb{K}^{q \times r}, D \in \mathbb{K}^{r \times s}, \quad (A \otimes C)(B \otimes D) = (AB) \otimes (CD)$$

证明:

- 结合律

设 A 的列为 a_1, \dots, a_n 、 B 的列为 b_1, \dots, b_q 、 C 的列为 c_1, \dots, c_s , 上方得到克罗内克积定义的过程中已经证明了 $A \otimes B$ 是将所有 $a_i \otimes b_j$ 按照先 j 后 i 顺序作为列拼成的矩阵。

由此, 进一步计算可知 $(A \otimes B) \otimes C$ 是将所有 $(a_i \otimes b_j) \otimes c_k$ 按照先 k 、然后 j 、最后 i 的顺序作为列拼成的矩阵, $A \otimes (B \otimes C)$ 是将所有 $a_i \otimes (b_j \otimes c_k)$ 按照先 k 、然后 j 、最后 i 的顺序作为列拼成的矩阵。之前已经证明了向量的克罗内克积具有结合律, 从而矩阵的克罗内克积也具有结合律。

- 乘法性质

这里我们采用线性映射的思路，以下所有 \otimes 均指克罗内克积。

将 A, C 看作对应向量空间中的线性映射，由线性映射张量积定义与本讲义 27.2.3 坐标在对应基下（这里事实上是 \mathbb{K}^n 与 \mathbb{K}^q 标准基的张量积）为克罗内克积，可知

$$\forall \alpha \in \mathbb{K}^n, \beta \in \mathbb{K}^q, (A \otimes C)(\alpha \otimes \beta) = (A\alpha) \otimes (C\beta)$$

同理有

$$\forall \gamma \in \mathbb{K}^p, \xi \in \mathbb{K}^s, (B \otimes D)(\gamma \otimes \xi) = (B\gamma) \otimes (D\xi)$$

而由于上方的 $B\gamma \in \mathbb{K}^n, C\xi \in \mathbb{K}^m$ ，即得

$$\forall \gamma \in \mathbb{K}^p, \xi \in \mathbb{K}^s, (A \otimes C)(B \otimes D)(\gamma \otimes \xi) = (AB\gamma) \otimes (CD\xi)$$

另一方面，根据 AB 与 CD 张量积的定义，应有

$$\forall \gamma \in \mathbb{K}^p, \xi \in \mathbb{K}^s, ((AB) \otimes (CD))(\gamma \otimes \xi) = (AB\gamma) \otimes (CD\xi)$$

最后，再由张量积定义，存在唯一的 \mathbb{K}^{ps} 到 \mathbb{K}^{mq} 线性映射 ϕ 使得

$$\forall \gamma \in \mathbb{K}^p, \xi \in \mathbb{K}^s, \phi(\gamma \otimes \xi) = (AB\gamma) \otimes (CD\xi)$$

利用向量空间之间线性映射表示为矩阵乘法的唯一性即得必然

$$(A \otimes C)(B \otimes D) = (AB) \otimes (CD)$$

从线性映射张量积的角度来说，前者对应多个线性映射张量积某种意义上的结合律，后者对应复合映射的张量积拆分为张量积映射的复合。

此外，还有一些不变量相关的性质：

- 对任何 \mathbb{K} 上方阵 A, B ，若 A, B 均可逆，则 $A \otimes B$ 可逆，且 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ 。

证明：

设 A 为 m 阶、 B 为 n 阶，若 A, B 均可逆，直接由乘法性质计算可知

$$(A \otimes B)(A^{-1} \otimes B^{-1}) = I_m \otimes I_n$$

直接计算可发现 $I_m \otimes I_n = I_{mn}$ ，从而 $(A^{-1} \otimes B^{-1})$ 即为 $A \otimes B$ 的逆。

- 对任何 \mathbb{K} 上矩阵 A, B 有 $\text{rank}(A \otimes B) = (\text{rank } A)(\text{rank } B)$ 。

证明：

考虑相抵标准形，设 $\text{rank } A = a, \text{rank } B = b$ ，可逆矩阵 P_A, Q_A, P_B, Q_B 使得

$$A = P_A \begin{pmatrix} I_a & O \\ O & O \end{pmatrix} Q_A$$

$$B = P_B \begin{pmatrix} I_b & O \\ O & O \end{pmatrix} Q_B$$

利用张量积的乘法性质可知

$$A \otimes B = (P_A \otimes P_B) \left(\begin{pmatrix} I_a & O \\ O & O \end{pmatrix} \otimes \begin{pmatrix} I_b & O \\ O & O \end{pmatrix} \right) (Q_A \otimes Q_B)$$

已经证明 $P_A \otimes P_B$ 与 $Q_A \otimes Q_B$ 可逆，从而只需计算中间部分的秩，直接计算可发现

$$\begin{pmatrix} I_a & O \\ O & O \end{pmatrix} \otimes \begin{pmatrix} I_b & O \\ O & O \end{pmatrix} = \text{diag} \left(\left(\begin{pmatrix} I_a & O \\ O & O \end{pmatrix}, \dots, \begin{pmatrix} I_a & O \\ O & O \end{pmatrix}, O \right) \right)$$

且非零的对角块共 a 个，于是其有 ab 个不同行不同列的 1，其余元素为 0，考虑列秩可知秩为 ab ，得证。

* 这即代表张量积映射的像空间维数为两映射像空间维数乘积。接下来所有性质都可直接对应为线性变换张量积的性质。

* 由此也可利用可逆等价于满秩直接得到 $A \otimes B$ 可逆当且仅当 A, B 均可逆。

3. 对任何 \mathbb{K} 上方阵 A, B ，若 A, B 均可对角化，则 $A \otimes B$ 可对角化。

证明：

设 $A = P^{-1}DP$ 、 $B = Q^{-1}FQ$ ， D, F 为对角阵， P, Q 可逆，则利用张量积的乘法性质可知

$$A \otimes B = (P^{-1} \otimes Q^{-1})(D \otimes F)(P \otimes Q)$$

已经证明 $P^{-1} \otimes Q^{-1} = (P \otimes Q)^{-1}$ ，直接计算可发现对角阵的张量积仍为对角阵（只有对角分块可能非零，且每个对角分块都是对角阵倍数），从而这已经给出了 $A \otimes B$ 的相似对角化。

4. 对 $A \in \mathbb{C}^{m \times m}$ 与 $B \in \mathbb{C}^{n \times n}$ ，设 A 的特征值为 $\lambda_1, \dots, \lambda_m$ ， B 的特征值为 μ_1, \dots, μ_n ，则 $A \otimes B$ 的特征值为

$$\lambda_i \mu_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

证明：

我们需要利用相似三角化进行证明。设 $A = P^{-1}SP$ 、 $B = Q^{-1}TQ$ ， S, T 为上三角阵， P, Q 可逆。

利用张量积乘法性质得到

$$A \otimes B = (P^{-1} \otimes Q^{-1})(S \otimes T)(P \otimes Q)$$

与之前类似得到 $A \otimes B$ 相似于 $S \otimes T$ 。设 S 各分量为 s_{ij} 、 T 各分量为 t_{ij} ，利用 S 上三角将其展开写为（未写出部分为 0）

$$\begin{pmatrix} s_{11}T & \cdots & s_{1m}T \\ \ddots & \ddots & \vdots \\ & & s_{mm}T \end{pmatrix}$$

进一步由于 T 上三角，可发现其每个对角块均为上三角阵，从而 $S \otimes T$ 也为上三角阵，且所有对角元为

$$s_{ii}t_{jj}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

按先 j 后 i 的顺序排列而成。

利用上三角矩阵的对角元即为特征值，且相似不改变特征值，即得 s_{11}, \dots, s_{mm} 为 A 的特征值， t_{11}, \dots, t_{nn} 为 B 的特征值，且

$$s_{ii}t_{jj}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

为 $A \otimes B$ 的特征值。这就是结论的形式。

5. 对任何 \mathbb{K} 上方阵 A, B 有 $\text{tr}(A \otimes B) = (\text{tr } A)(\text{tr } B)$ 。

证明：

沿用性质 5 的假设，利用迹为特征值之和，并运用乘法结合律可得

$$\text{tr}(A \otimes B) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j = \sum_{i=1}^m \lambda_i \left(\sum_{j=1}^n \mu_j \right) = \left(\sum_{i=1}^m \lambda_i \right) \left(\sum_{j=1}^n \mu_j \right)$$

再次运用迹为特征值之和即得乘积的两项分别为 $\text{tr}(A)$ 、 $\text{tr}(B)$ ，从而得证。

6. 对 $A \in \mathbb{K}^{m \times m}$ 与 $B \in \mathbb{K}^{n \times n}$ 有 $\det(A \otimes B) = (\det A)^n (\det B)^m$ 。

证明：

沿用性质 5 的假设，利用行列式为特征值之积可得

$$\det(A \otimes B) = \prod_{i=1}^m \prod_{j=1}^n \lambda_i \mu_j$$

可以发现，每个 λ_i 在乘积中出现了 n 次（对应右侧为 μ_1, \dots, μ_n 时），每个 μ_j 在乘积中出现了 m 次（对应左侧为 $\lambda_1, \dots, \lambda_m$ 时），从此乘积可写为

$$\det(A \otimes B) = (\lambda_1 \dots \lambda_m)^n (\mu_1 \dots \mu_n)^m$$

再次运用行列式为特征值之积即得两个括号中分别为 $\det A$ 、 $\det B$ ，从而得证。

至此，我们对有限维线性空间的张量积与对应的线性映射、线性变换张量积都有了基本的研究。仍然注意我们用到的张量积最重要性质来自其定义：将左右的一组基两两配对形成整体的一组基。对于线性无关向量组，可以扩充为基进行考虑。

二十八 补充：期末复习

* 按照丘书的顺序，务必至少掌握下方列出的算法，此外，由于每节的内容、结论较多，这里将顺序列举关键结论。期中复习提纲已经列出的结论不再列举。期末考试会考期中内容，因此必要的算法与证明技巧仍需复习。

* 所有算法和标注无需证明的结论都可以不用掌握证明。

* 所有标注了解即可的结论看一下结论就行。

* 标注简单推论的结论可以熟悉，不用强记，理应可以通过定义、定理简单推出。

* 加粗的结论表示必须记忆的核心结论。

§28.1 知识与技巧整理

28.1.1 上册知识

主要包含期末需要用到的正交矩阵与对称矩阵相关基础知识。

4.6 * 本节这学期**必须熟练掌握**，可以考虑将例题与习题做一遍。

正交矩阵定义

正交矩阵基本性质 (简单推论)

\mathbb{R}^n 内积、长度、正交、正交向量组、正交单位向量组、正交基、标准正交基定义

内积与长度的基本性质 (简单推论)

施密特正交化算法

正交补与正交投影定义与性质，即例 17、例 18

QR 分解、最小二乘解，即例 5、例 6、例 19

酉矩阵定义与性质，即习题 16 到 18

5.5 正交矩阵特征值与特征向量基本性质，即例 8

5.7 实对称矩阵、正交相似定义

正交相似是等价关系 (简单推论)

实对称矩阵特征值、特征向量性质

实对称矩阵正交相似对角化 (无需证明)

正交相似对角化算法

实对称矩阵相似性质，即命题 1、命题 2、习题 2 (简单推论)

只有实特征值的实正规矩阵是实对称矩阵，即例 5

斜对称矩阵、Hermite 阵、斜 Hermite 阵特征值性质，即例 7、习题 7、习题 8

* 最好掌握例 6 相似三角化结论，此结论证明一些命题较为方便。

6.1 二次型、合同定义

斜对称阵等价性质，即例 7

斜对称阵合同标准形，即例 10 (无需证明)

Rayleigh 商基本性质，即例 11

同时正交相似对角化，即例 14

6.2 实对称阵的合同规范形 (无需证明)

复对称阵的合同规范形 (无需证明)

规范形出发的向量构造，即习题 5

6.3 正定、半正定、负定、半负定定义

* 有时间可以看看性质与相关技巧，不过本学期不会应用太多。

28.1.2 内积空间

期末考试的重点考察内容，上学期正交阵、对称阵相关知识与前半学期线性空间、线性映射知识的通过新概念“双线性函数”进行结合。

10.1 双线性函数、度量矩阵定义

双线性函数换基对应度量矩阵合同

左根、右根、非退化定义

非退化对应度量矩阵性质（简单推论）

对称与斜对称双线性函数定义

对称与斜对称双线性函数的最简度量矩阵（无需证明）

* 注意上册我们只讨论了实对称方阵、复对称方阵与实斜对称方阵，这里考虑的是一般数域。

二次函数定义

二次函数唯一对应对称双线性函数（简单推论）

惯性定理（无需证明）

* 惯性定理无需证明，因此本学期和上学期的形式无区别。

双线性函数构成线性空间（简单推论）

双线性函数下的正交、正交补定义，即例 6、例 7

* 有时间可以学习例 6 的构造性证明思路。

同时对角化结论，即例 20（无需证明）

重要例题：

- 双线性函数的基本计算（习题 1、习题 3、习题 8）
- 利用抽象性质的证明（例 5、例 8、习题 5）
- 利用标准形的证明（习题 4、习题 9）

10.2 正定、内积、实内积空间、欧几里得空间、长度、单位向量、夹角、正交、距离定义

柯西不等式（无需证明）

内积与长度的基本性质（简单推论）

正交基、标准正交基定义

施密特正交化算法

* 最好看一下此算法的证明，此算法也可用于将一些正交向量组扩充为标准正交基，或计算正交补空间。

标准正交基下的坐标，即命题 6

标准正交基之间的过渡矩阵是正交阵，即命题 7

保距同构定义

保距同构对应的基性质，即定理 2、推论 7

通过内积确定向量，即例 8

重要例题：

- 内积的基本性质（例 2、例 3、例 6、例 7）
- 构造性问题（习题 13、习题 16）
- 不等式问题（习题 2、习题 18）

10.3 正交补、子空间正交、正交投影定义

正交补与正交投影基本性质，即例 2、例 3、例 5、例 7、例 14、例 15（简单推论）

正交投影的最佳逼近性，即定理 2、例 6

最小二乘解定义与等价性质

重要例题：

- 投影与坐标 (例 10、例 12、例 16)
- 正交补计算与应用 (例 4、习题 6、习题 10)

10.4 * 本节的很多结论对更一般的实正规变换成立，因此可先跳过，在 10.5 学习一般情况。

正交变换、镜面反射定义

正交变换基本性质 (简单推论)

正交变换等价定义, 即命题 3、命题 4

对称变换定义

对称变换基本性质 (简单推论)

正交投影等价定义, 即例 24

* 更多讨论见本讲义 26.1.1 第 3 题。

重要例题：

- 构造性问题 (例 6、例 10)
- 整体操作 (例 20、例 21、例 27、习题 12)
- 不等式问题 (例 30、习题 10)

10.5 * 以下为与实内积空间相同部分，几乎可类似证明：

复内积、酉空间、长度、单位向量、夹角、正交、距离定义与性质

正交规范集、标准正交基、Gram 矩阵定义与性质

酉矩阵、保距同构、正交补、正交投影定义与性质

酉变换、Hermite 变换、斜 Hermite 变换 (例 14) 定义与性质

正交补、正交投影、最佳逼近相关性质, 即例 22 到 29 (只需会有限维版本)

* 以下为额外部分，注意有些结论的实内积与复内积版本：

伴随变换定义

用伴随变换定义正交变换、对称变换等，即定义 11 推论、例 14(2)、例 16、例 17、例 48 (简单推论)

伴随变换存在唯一性与矩阵表示

正规变换定义

正规变换基本性质 (简单推论)

复正规矩阵的酉相似对角化 (注意算法，无需证明，但有时间建议看一下证明)

复正规矩阵酉相似对角化推论, 即定理 14、定理 15、例 14(4) (简单推论)

实正规矩阵的正交相似结论, 即习题 25 (无需证明，但有时间建议看一下证明)

实正规矩阵的正交相似推论, 包括实正交阵、实对称阵、实斜对称阵的正交相似标准形 (简单推论)

Hermite 型、正定性定义与性质 (无需证明)

* 以下为建议掌握的二级结论：

QR 分解, 即例 7 (注意实方阵版本)

酉相似三角化, 即例 39 (注意实方阵版本, 形式类似习题 25)

同时相似对角化, 即例 42 (注意实方阵版本)

极分解, 即例 54 (注意实方阵版本)

重要例题：

- 整体操作 (例 5、例 9、例 19、例 41、例 58(1)、习题 11 到 13、习题 23, 注意例 9 不使用标准形的做法)
- 标准形操作 (例 45、习题 20)
- 计算性问题 (例 11、例 37)
- 不等式问题 (例 55、习题 15)

* 习题 30 等正定性相关问题从往年来看不是重点，因此建议有时间再看。

28.1.3 张量

不如内积空间重点，但仍然会考，更重要的是学会基本定义后掌握偏抽象的形式推理方式，并适当了解技巧。注意我们只需掌握有限维版本。

9.10 线性函数、对偶空间、对偶基定义

对偶基下的坐标与过渡矩阵（简单推论）

V^{**} 到 V 的自然同构（对书上的讲法非常重要，用本讲义 27 章的讲法可部分规避）

重要例题：

- 对偶相关计算（例 5、习题 5）
- 实内积空间对偶（例 10）

* 例题中出现的一些更高级的内容往年来看没有考察过，可先只了解基础，有时间再看。

10.1 线性函数张量积定义

张量积构造双线性函数空间基

双线性函数秩与秩空间定义

双线性函数秩大于等于其矩阵秩，且对称或斜对称时相等（无需证明，有时间可以看看）

11.1 多重线性映射、多重线性函数定义

基映射定义多重线性映射，即定理 1

多重线性映射两种基构造方式，即定理 2 与定理 3

* 可参考本讲义 27.1。

重要例题：

- 双线性性相关抽象说明（例 3、例 4）

11.2 张量积定义与特征性质（无需证明）

张量积空间的同构唯一性与基的构造（无需证明，基的构造是对考试最重要的性质）

张量积可交换、结合（无需证明）

线性变换张量积定义与存在唯一性（无需证明）

线性变换张量积的基本性质（简单推论）

张量积的矩阵表示对应矩阵克罗内克积（无需证明，以此可通过矩阵论证明性质）

* 可参考本讲义 27 章相关内容，个人认为有限维情况较书上讲法更简洁清晰。

重要例题：

- 向量张量积的性质证明（例 1、例 3）
- 矩阵克罗内克积的性质（例 6 到例 8）

* 讲解与例题中出现的一些更高级的内容往年来看没有考察过，可先只了解基础，有时间再看。

§28.2 复习题

去年期末的证明题基本是 1、2、3(c)(i)、4(b)(d)、8(c)，5、6 取自往年期末。

1. 在 n 阶实方阵空间 $V = \mathbb{R}^{n \times n}$ 中定义函数 $f(A, B) = \text{tr}(A^T B)$ 。

(a) 求证 f 为 V 上的内积。

(b) 求 f 下的一组标准正交基，使得基中每个向量都为对称阵或斜对称阵。

- (c) 对任何矩阵 A , 求 A 到 V 的子空间 $W = \{X \mid \text{tr}(X) = 0\}$ 的正交投影。
2. 若 f 是 \mathbb{K} 上的 n 维线性空间 V 上的非退化双线性函数, 对 V 的一组基 $\alpha_1, \dots, \alpha_n$, 证明存在唯一一组 V 的基 β_1, \dots, β_n 使得

$$f(\alpha_i, \beta_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

3. 设 V 为数域 \mathbb{K} 上的线性空间。

- (a) 若 $f, g \in V^*$ 满足 $\text{Ker } f = \text{Ker } g$, 证明存在非零常数 $\lambda \in \mathbb{K}$ 使得 $g = \lambda f$ 。
- (b) 若 f, g 是两个 V 上的双线性函数, 且零点集合相同, 即

$$\forall \alpha, \beta \in V, \quad f(\alpha, \beta) = 0 \Leftrightarrow g(\alpha, \beta) = 0$$

证明存在非零常数 $\lambda \in \mathbb{K}$ 使得 $g = \lambda f$ 。

- (c) (i) 证明若 V 上双线性函数 f 保持正交性质对称 (即 $f(\alpha, \beta) = 0$ 当且仅当 $f(\beta, \alpha) = 0$), 则其一定为对称或斜对称。
- (ii) 若 V 是 n 维实内积空间, 其上线性变换 \mathcal{A} 保持正交性 (即 $(\mathcal{A}\alpha, \mathcal{A}\beta) = 0$ 当且仅当 $(\alpha, \beta) = 0$), 则一定存在正交变换 \mathcal{B} 与非零常数 $\lambda \in \mathbb{R}$ 使得 $\mathcal{A} = \lambda \mathcal{B}$ 。
- (iii) 若 V 是 3 维实内积空间, 其上斜对称变换 \mathcal{A}, \mathcal{B} 满足 $\text{Ker } \mathcal{A} = \text{Ker } \mathcal{B}$, 证明存在非零常数 $\lambda \in \mathbb{R}$ 使得 $\mathcal{A} = \lambda \mathcal{B}$ 。
- (iv) 若 f 为 V 上的对称或斜对称双线性函数, 且存在线性函数 $g, h \in V^*$ 使得

$$\forall \alpha, \beta \in V, \quad f(\alpha, \beta) = g(\alpha)h(\beta)$$

证明存在线性函数 $l \in V^*$ 与非零常数 $\lambda \in \mathbb{K}$ 使得

$$\forall \alpha, \beta \in V, \quad f(\alpha, \beta) = \lambda l(\alpha)l(\beta)$$

- (d) 若 f, g 是两个 V 上的 n 重线性函数, 且零点集合相同, 即

$$\forall \alpha_1, \dots, \alpha_n \in V, \quad f(\alpha_1, \dots, \alpha_n) = 0 \Leftrightarrow g(\alpha_1, \dots, \alpha_n) = 0$$

证明存在非零常数 $\lambda \in \mathbb{K}$ 使得 $g = \lambda f$ 。

4. 设 \mathcal{A} 为实内积空间 V 上的正规变换, 且存在伴随变换 \mathcal{A}^* 。

- (a) 证明 \mathcal{A} 的特征向量是 \mathcal{A}^* 属于相同特征值的特征向量。
- (b) 证明 \mathcal{A} 属于不同特征值的特征向量正交。
- (c) 若 V 为有限维, 证明 $\text{Ker } \mathcal{A} = (\text{Im } \mathcal{A})^\perp$ 。
- (d) 若 V 为有限维, 证明当且仅当存在 $k \in \mathbb{N}$ 使得 $\mathcal{A}^k = \mathcal{A}^{k+1}$ 时 \mathcal{A} 是正交投影。
- (e) 若 V 为有限维, 证明若 U 为 \mathcal{A} -不变子空间, 则其为 \mathcal{A}^* -不变子空间, 且 U^\perp 也为 \mathcal{A} -不变子空间。
- (f) 若 V 为有限维, 证明若 U 为 \mathcal{A} -不变子空间, 则 $(\mathcal{A}|_U)^* = (\mathcal{A}^*)|_U$, 进而说明 $\mathcal{A}|_U$ 为正规变换
- (g) 若 V 为有限维, 证明 V 存在一维或二维的 \mathcal{A} -不变子空间, 进一步证明存在一组标准正交基使得 \mathcal{A} 在其上的矩阵表示为分块对角阵, 每个对角块一阶或二阶。
- (h) 若 V 为有限维, 证明 \mathcal{A} 与 \mathcal{A}^* 在任意一组标准正交基下的矩阵表示正交相似。

5. 若 n 阶实方阵 A 所有复特征值为 $\lambda_1, \dots, \lambda_n$ 。

(a) 证明存在正交阵 Q 使得 $Q^T A Q$ 为分块上三角阵，每个对角块为一阶或二阶。

(b) 证明

$$\text{tr}(AA^T) \geq \sum_{i=1}^n |\lambda_i|^2$$

且等号成立当且仅当 A 正规。

6. 若酉方阵 A 满足 -1 不是其特征值，证明

$$\text{i}(I - A)(I + A)^{-1}$$

是 Hermite 阵。

7. 对 $A, B \in \mathbb{C}^{n \times n}$ ，考虑 $\mathbb{C}^{n \times n}$ 的线性变换

$$\mathcal{C}(X) = AX - XB$$

(a) 证明其在基 $E_{11}, E_{21}, \dots, E_{n1}, \dots, E_{1n}, E_{2n}, \dots, E_{nn}$ 下的矩阵表示为 (I 指 n 阶单位阵)

$$C = I \otimes A - B^T \otimes I$$

(b) 证明 $A \otimes I - I \otimes B$ 与 $I \otimes A - B \otimes I$ 相似。

(c) 若 A, B 均可对角化，证明 \mathcal{C} 可对角化。

(d) 证明 \mathcal{C} 是同构当且仅当 A, B 无公共特征值。

(e) 若 \mathcal{C} 可对角化， A, B 是否均可对角化？给出证明或反例。

8. 设 U, V 为 \mathbb{K} 上 n, m 维线性空间。

(a) 若 $\alpha_1, \alpha_2 \in U$ 线性无关、 $\beta_1, \beta_2 \in V$ 线性无关，证明不存在 $v \in U, w \in V$ 使得

$$\alpha_1 \otimes \beta_1 + \alpha_2 \otimes \beta_2 = v \otimes w$$

若去掉 α_1, α_2 线性无关的条件，何时存在 $v \in U, w \in V$ 使得上式成立？

(b) 若 $\alpha \in U, \beta \in U$ ，且

$$\alpha \otimes \beta + \beta \otimes \alpha = 0$$

证明 α, β 中有零向量。

(c) 若 $\alpha, \beta, \gamma \in U$ ，且

$$\alpha \otimes \beta \otimes \gamma + \gamma \otimes \alpha \otimes \beta + \beta \otimes \gamma \otimes \alpha = 0$$

证明 α, β, γ 中有零向量。

(d) 若 $\alpha, \beta, \gamma, x, y, z \in U$ ，且

$$\alpha \otimes \beta \otimes \gamma + \gamma \otimes \alpha \otimes \beta + \beta \otimes \gamma \otimes \alpha = x \otimes y \otimes z \neq 0$$

是否能推出 α, β, γ 两两线性相关？给出证明或反例。

§28.3 解答

1. 以下我们用 a_{ij} 表示 A 的各个分量, b_{ij} 表示 B 的各个分量。

(a) 利用 tr 与转置的线性性可计算得, 对任何 $A_1, A_2, A, B_1, B_2, B \in \mathbb{R}^{n \times n}$ 、 $\lambda, \mu \in \mathbb{R}$ 有

$$f(\lambda A_1 + \mu A_2, B) = \text{tr}(\lambda A_1^T B + \mu A_2^T B) = \lambda \text{tr}(A_1^T B) + \mu \text{tr}(A_2^T B) = \lambda f(A_1, B) + \mu f(A_2, B)$$

$$f(A, \lambda B_1 + \mu B_2) = \text{tr}(\lambda A^T B_1 + \mu A^T B_2) = \lambda \text{tr}(A^T B_1) + \mu \text{tr}(A^T B_2) = \lambda f(A, B_1) + \mu f(A, B_2)$$

从而双线性性成立。

由 $\text{tr}(X) = \text{tr}(X^T)$ 可得

$$f(A, B) = \text{tr}(A^T B) = \text{tr}((A^T B)^T) = \text{tr}(B^T A) = f(B, A)$$

从而对称性成立。

直接计算可发现 (此计算是上学期矩阵论内容, 可参考上学期教材)

$$f(A, A) = \text{tr}(A^T A) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \geq 0$$

由于 $a_{ij} \in \mathbb{R}$, 当且仅当所有 a_{ij} 为 0 时等号成立, 从而正定性成立。

(b) * 这类不容易直接观察出结果的题目一般先设法取出基, 再进行正交化与规范化。

用 E_{ij} 表示只有第 i 行第 j 列为 1, 其余为 0 的方阵。我们先利用前半学期知识取出所有实对称阵的一组基:

$$E_{ii}, \quad i = 1, \dots, n$$

$$E_{ij} + E_{ji}, \quad 1 \leq i < j \leq n$$

所有斜对称阵的一组基:

$$E_{ij} - E_{ji}, \quad i \leq i < j \leq n$$

然后利用以上的基进行构造。

- 我们先证明上面取出的 n^2 个矩阵在 f 下相互正交, 从而根据两两正交的非零向量线性无关可知它们线性无关。

首先, 直接计算可得

$$f(A, B) = \text{tr}(A^T B) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ij}$$

由此, 若 A 与 B 非零元素位置不同, 必然 $f(A, B) = 0$ 。

根据形式可发现, 上述 n^2 个矩阵中, 只有 i, j 固定时的 $E_{ij} + E_{ji}$ 与 $E_{ij} - E_{ji}$ 存在相同位置的非零元素, 从而其他情况必然相互正交, 而直接计算有 (注意 $i < j$)

$$\text{tr}((E_{ij} + E_{ji})^T (E_{ij} - E_{ji})) = \text{tr}(E_{jj} - E_{ii}) = 0$$

于是 n^2 个矩阵中任意两个不同矩阵在 f 下相互正交。

- 由于这些矩阵的个数等于 V 的维数, 且线性无关, 它们已经构成 V 的一组基, 于是再由相互正交可知它们构成 V 的正交基。要得到标准正交基只需再进行规范化。

直接计算可知 $\text{tr}(E_{ii}^T E_{ii}) = 1$, 且 $i < j$ 时

$$\text{tr}((E_{ij} + E_{ji})^T (E_{ij} + E_{ji})) = \text{tr}(E_{ii} + E_{jj}) = 2$$

$$\text{tr}((E_{ij} - E_{ji})^T (E_{ij} - E_{ji})) = \text{tr}(E_{ii} + E_{jj}) = 2$$

于是除以模长得到规范化结果

$$\begin{aligned} E_{ii}, \quad i = 1, \dots, n \\ \frac{\sqrt{2}}{2}(E_{ij} + E_{ji}), \quad 1 \leq i < j \leq n \\ \frac{\sqrt{2}}{2}(E_{ij} - E_{ji}), \quad i \leq i < j \leq n \end{aligned}$$

综合以上讨论即得到了原空间每个都是对称或斜对称阵的标准正交基。

(c) 我们介绍利用最短距离与正交补的两种不同计算方式。

- 最短距离思路

假设要求的正交投影为 X , 其各个分量为 x_{ij} , 根据正交投影性质有

$$\forall B \in W, \quad f(A - X, A - X) \leq f(A - B, A - B)$$

也即, 我们要找到使得 $f(A - B, A - B)$ 最小的 $B \in W$ 。根据定义可直接计算得

$$f(A - B, A - B) = \text{tr}((A - B)^T (A - B)) = \sum_{i=1}^n \sum_{j=1}^n (a_{ij} - b_{ij})^2$$

对任何矩阵 $B \in W$, 由于有条件 $b_{11} + \dots + b_{nn} = 0$, 我们代入 $b_{nn} = -b_{11} - \dots - b_{n-1,n-1}$, 将上式重新改写为

$$f(A - B, A - B) = \sum_{i \neq j} (a_{ij} - b_{ij})^2 + \sum_{i=1}^{n-1} (a_{ii} - b_{ii})^2 + (a_{nn} + b_{11} + \dots + b_{n-1,n-1})^2$$

由于 $i \neq j$ 时 b_{ij} 相互独立, 取 $b_{ij} = a_{ij}$ 即可使得平方项为 0, 最小, 接下来观察剩下的项。可以发现, 剩下 n 项括号内求和为定值, 由此利用 Cauchy 不等式可得

$$\begin{aligned} & \sum_{i=1}^{n-1} (a_{ii} - b_{ii})^2 + (a_{nn} + b_{11} + \dots + b_{n-1,n-1})^2 \\ &= \frac{1}{n} \left(\sum_{i=1}^{n-1} (a_{ii} - b_{ii})^2 + (a_{nn} + b_{11} + \dots + b_{n-1,n-1})^2 \right) (1^2 + \dots + 1^2) \\ &\geq \frac{1}{n} ((a_{11} - b_{11}) + \dots + (a_{n-1,n-1} - b_{n-1,n-1}) + (a_{nn} + b_{11} + \dots + b_{n-1,n-1}))^2 \\ &= \frac{1}{n} (\text{tr } A)^2 \end{aligned}$$

且等号成立当且仅当 $a_{11} - b_{11}, \dots, a_{n-1,n-1} - b_{n-1,n-1}, a_{nn} + b_{11} + \dots + b_{n-1,n-1}$ 与各分量全为 1 的向量成正比, 也即它们全部相等, 由此即可解出

$$b_{ii} = a_{ii} - \frac{1}{n} \text{tr}(A), \quad i = 1, \dots, n-1$$

可计算验证上式对 $i = n$ 也成立, 因此最终得到 X 的各分量为

$$x_{ij} = \begin{cases} a_{ij} & i \neq j \\ a_{ij} - \frac{1}{n} \text{tr}(A) & i = j \end{cases}$$

或整体写为

$$X = A - \frac{1}{n} \text{tr}(A)$$

* 求最小值点也可以使用条件极值或多元函数最小值等思路, 总体来说这个做法相对分析化(适合高数学得好的同学使用)。

- 正交补思路

根据定义，只要计算出 W^\perp ，即可通过直和分解将 A 分解到 W 上得到正交投影。

利用上半学期知识（可见本讲义 18.2.2 第 2 题）可给出 W 的一组基

$$E_{ij}, \quad i \neq j$$

$$E_{11} - E_{kk}, \quad k = 2, \dots, n$$

利用正交的性质，与 W 所有矩阵正交等价于与其一组基正交。若 B 与 W 中所有矩阵正交，直接利用 $f(A, B) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ij}$ 计算可知

$$f(E_{ij}, B) = b_{ij} = 0, \quad i \neq j$$

$$f(E_{11} - E_{kk}, B) = b_{11} - b_{kk} = 0, \quad k = 2, \dots, n$$

这即是说 $B = b_{11}I$ (I 指 n 阶单位阵)，从而 $W^\perp = \langle I \rangle$ 。

有了 W^\perp 的表达式后，为计算 A 在 W 下的正交投影 X ，我们只需要找到方阵 X, Y 使得

$$A = X + Y, \quad X \in W, \quad Y \in W^\perp$$

设 $y \in \mathbb{R}$ 使得 $Y = yI$ ，上式即是说 $X = A - yI$ ，于是 $\text{tr}(A - yI) = 0$ ，直接计算可知

$$y = \frac{1}{n} \text{tr}(A)$$

从而最终得到

$$X = A - \frac{1}{n} \text{tr}(A)$$

2. * 对于有限维问题往往需要通过度量矩阵进行处理，转化为矩阵论后即可通过方程组、逆矩阵等知识得到存在唯一性。

设 f 在基 $\alpha_1, \dots, \alpha_n$ 下的度量矩阵为 A ，其各分量为 a_{ij} ，利用非退化性定义可知 A 可逆。

另一方面，将 β_1, \dots, β_n 在基 $\alpha_1, \dots, \alpha_n$ 下展开，也即设 $c_{ij} \in \mathbb{R}$ 使得

$$\forall k = 1, \dots, n, \quad \beta_j = \sum_{k=1}^n c_{kj} \alpha_k$$

可利用双线性性与度量矩阵的定义计算得到

$$f(\alpha_i, \beta_j) = \sum_{k=1}^n c_{kj} f(\alpha_i, \alpha_k) = \sum_{k=1}^n c_{kj} a_{ik} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

假设所有 c_{ij} 拼成矩阵 C ，可以发现（数的乘法可以交换） $\sum_{k=1}^n c_{kj} a_{ik} = \sum_{k=1}^n a_{ik} c_{kj}$ 即为 AC 的第 i 行第 j 列元素，于是上式等价于

$$AC = I$$

由于 A 可逆，存在唯一 C 使得 $AC = I$ ，而既然每个 β_j 在基 $\alpha_1, \dots, \alpha_n$ 下的坐标存在唯一，每个 β_j 也存在唯一。

最后，由于 C 也可逆，由 C 的列线性无关利用坐标的同构性可知 β_1, \dots, β_n 线性无关，又由 $\dim V = n$ ，上述的唯一一组向量组是 V 的基。

3. 零点相关的问题往往基于如下的核心思路：利用线性性将非零点“移动”到零点进行处理。

(a) 若 f 为零映射, 则由于 $\text{Ker } g = \text{Ker } f = V$ 可知 g 为零映射, 从而 $g = f$, 得证, 下假设 f 不为零映射。

由于存在 $x \in V$ 使得 $f(x) \neq 0$, 根据条件有 $x \notin \text{Ker } g$, 从而 $g(x) \neq 0$, 记 $\lambda = \frac{g(x)}{f(x)} \in \mathbb{K}$, 其非零。

对任何 $y \in V$, 有

$$f\left(y - \frac{f(y)}{f(x)}x\right) = f(y) - \frac{f(y)}{f(x)}f(x) = 0$$

于是根据 $\text{Ker } f = \text{Ker } g$ 可知

$$g\left(y - \frac{f(y)}{f(x)}x\right) = 0$$

展开得到

$$g(y) - \frac{f(y)}{f(x)}g(x) = 0$$

也即

$$g(y) = \frac{f(x)}{g(x)}f(y) = \lambda f(y)$$

由于此式对任何 $y \in V$ 成立, 即得 $g = \lambda f$ 。

(b) 利用双线性性, 固定 $\alpha \in V$, $\varphi(\beta) = f(\alpha, \beta)$ 与 $\psi(\beta) = g(\alpha, \beta)$ 为 V^* 中的线性函数。由于 f 与 g 的零点集合相同, 应有 $\text{Ker } \varphi = \text{Ker } \psi$, 从而由 (a) 可知存在 \mathbb{K} 中的 $\lambda_\alpha \neq 0$ (一定要注意这里的 λ 与 α 有关) 使得 $\psi = \lambda_\alpha \varphi$, 也即

$$\forall \alpha \in V, \exists \lambda_\alpha \in \mathbb{K}, \forall \beta \in V, g(\alpha, \beta) = \lambda_\alpha f(\alpha, \beta)$$

我们接下来证明对所有 α, λ_α 应相同。

若 f 为零映射, 由零点集合相同可知 g 为零映射, 取 $\lambda = 1$ 即可, 否则, 存在某个 $\alpha_0, \beta_0 \in V$ 使得 $f(\alpha_0, \beta_0) \neq 0$, 从而可记 $\lambda = \lambda_{\alpha_0}$, 其满足

$$\forall \beta \in V, g(\alpha_0, \beta) = \lambda f(\alpha_0, \beta)$$

同理, 固定 $\beta \in V$, $f(\alpha, \beta)$ 与 $g(\alpha, \beta)$ 也为关于 α 的 V^* 中线性函数, 从而重复上述过程可知

$$\forall \alpha \in V, g(\alpha, \beta_0) = \lambda f(\alpha, \beta_0)$$

最后对一切 $\alpha, \beta \in V$ 考虑, 分为三类讨论:

- 若 $f(\alpha_0, \beta) \neq 0$, 可知

$$f\left(\alpha - \frac{f(\alpha, \beta)}{f(\alpha_0, \beta)}\alpha_0, \beta\right) = f(\alpha, \beta) - \frac{f(\alpha, \beta)}{f(\alpha_0, \beta)}f(\alpha_0, \beta) = 0$$

从而根据零点集合相同可知

$$g\left(\alpha - \frac{f(\alpha, \beta)}{f(\alpha_0, \beta)}\alpha_0, \beta\right) = 0$$

展开并利用 $g(\alpha_0, \beta) = \lambda f(\alpha_0, \beta)$ 得到

$$g(\alpha, \beta) - \frac{\lambda f(\alpha_0, \beta)}{f(\alpha_0, \beta)}f(\alpha, \beta) = 0$$

即得 $g(\alpha, \beta) = \lambda f(\alpha, \beta)$ 。

- 若 $f(\alpha, \beta_0) \neq 0$, 利用 $g(\alpha, \beta_0) = \lambda f(\alpha, \beta_0)$ 考虑

$$f\left(\alpha, \beta - \frac{f(\alpha, \beta)}{f(\alpha, \beta_0)}\beta_0\right)$$

完全类似上一种情况可得结论。

- 若 $f(\alpha_0, \beta) = f(\alpha, \beta_0) = 0$, 构造更加具有技巧性: 先待定系数 $\mu, \gamma \in \mathbb{K}$, 考虑

$$f(\alpha - \mu\alpha_0, \beta - \gamma\beta_0)$$

利用条件可将其展开为

$$f(\alpha - \mu\alpha_0, \beta - \gamma\beta_0) = f(\alpha, \beta) - \mu f(\alpha_0, \beta) - \gamma f(\alpha, \beta_0) + \mu\gamma f(\alpha_0, \beta_0) = f(\alpha, \beta) + \mu\gamma f(\alpha_0, \beta_0)$$

为了将其移动到零点, 我们取 $\mu = -1$ 、 $\gamma = \frac{f(\alpha, \beta)}{f(\alpha_0, \beta_0)}$, 即有 $f(\alpha - \mu\alpha_0, \beta - \gamma\beta_0) = 0$ 。

由于 f 与 g 的零点相同, 可知 $g(\alpha_0, \beta) = g(\beta, \alpha_0) = 0$, 且 $g(\alpha - \mu\alpha_0, \beta - \gamma\beta_0) = 0$, 从而进一步展开得到

$$g(\alpha, \beta) + \mu\gamma g(\alpha_0, \beta_0) = 0$$

代入即得

$$g(\alpha, \beta) - \frac{f(\alpha, \beta)}{f(\alpha_0, \beta_0)} g(\alpha_0, \beta_0) = 0$$

再利用 $g(\alpha_0, \beta_0) = \lambda f(\alpha_0, \beta_0)$ 即得到

$$g(\alpha, \beta) = \lambda f(\alpha, \beta)$$

由于定义保证了 λ 不为 0, 综合以上三种情况得证。

- (c) * 这部分的题目均为教材习题或往年题, 可以发现上述零点相同推出相差倍数的性质对于解决一些问题非常有用, 不过, 解决问题的基础是构造合适的线性/双线性函数。

- (i) 定义 $V \times V$ 上的函数 g 满足

$$\forall \alpha, \beta \in V, \quad g(\alpha, \beta) = f(\beta, \alpha)$$

利用 f 对两个分量均线性可发现 g 也是双线性函数。条件即成为 g 与 f 零点集合相同, 从而根据 (b) 存在 $\lambda \neq 0$ 使得 $g = \lambda f$, 即

$$\forall \alpha, \beta \in V, \quad f(\beta, \alpha) = \lambda f(\alpha, \beta)$$

若 f 为零映射, 其满足对称性, 符合要求。否则存在 $\alpha_0, \beta_0 \in V$ 使得 $f(\alpha_0, \beta_0) \neq 0$, 且

$$f(\alpha_0, \beta_0) = \lambda f(\beta_0, \alpha_0) = \lambda^2 f(\alpha_0, \beta_0)$$

由此即得 $\lambda^2 = 1$, 于是 $\lambda = \pm 1$, $\lambda = 1$ 满足对称性定义, $\lambda = -1$ 满足斜对称性定义。

- (ii) 根据内积的定义, 内积应为 V 上的双线性函数。定义 $V \times V$ 函数 φ 满足

$$\forall \alpha, \beta \in V, \quad \varphi(\alpha, \beta) = (\mathcal{A}\alpha, \mathcal{A}\beta)$$

利用线性映射的复合仍为线性映射可知 φ 也是双线性函数。条件即成为 φ 与内积零点集合相同, 从而根据 (b) 存在 $\mu \neq 0$ 使得

$$\forall \alpha, \beta \in V, \quad (\mathcal{A}\alpha, \mathcal{A}\beta) = \mu(\alpha, \beta)$$

我们先证明 $\mu > 0$: 任取 $\alpha \neq 0$, 可得 $(\mathcal{A}\alpha, \mathcal{A}\alpha) = \mu(\alpha, \alpha)$, 由于内积的正定性可知 $(\alpha, \alpha) > 0$ 、 $(\mathcal{A}\alpha, \mathcal{A}\alpha) \geq 0$, 从而 $\mu \geq 0$ 。由于 μ 非零, 其必然大于 0。

由此, 记 $\lambda = \sqrt{\mu}$, $\mathcal{B} = \frac{1}{\lambda} \mathcal{A}$, 利用内积双线性直接计算得到

$$\forall \alpha, \beta \in V, \quad (\mathcal{B}\alpha, \mathcal{B}\beta) = (\alpha, \beta)$$

这即是有限维线性空间正交变换的等价定义, 从而 \mathcal{B} 是正交变换, 且 $\mathcal{A} = \lambda \mathcal{B}$ 。

(iii) 首先，利用斜对称矩阵的秩一定为偶数，可知 $\dim \text{Im } \mathcal{A}$ 为偶数，从而由 $\dim V = 3$ 可得其为 0 或 2。 $\dim \text{Im } \mathcal{A} = 0$ 即代表 \mathcal{A} 为零映射，此时 \mathcal{B} 也为零映射，取 $\lambda = 1$ 即可。以下假设 \mathcal{A}, \mathcal{B} 不为零映射。

此时，根据第一同构定理有

$$\dim \text{Ker } \mathcal{A} = \dim V - \dim \text{Im } \mathcal{A} = 3 - 2 = 1$$

设其一组基为 v_1 ，由 $\text{Ker } \mathcal{A} = \text{Ker } \mathcal{B}$ 知 v_1 也是 $\text{Ker } \mathcal{B}$ 一组基。将 v_1 扩充为 V 的一组基 v_1, v_2, v_3 。

考虑 $V \times V$ 上的函数 f, g 满足对任何 $\alpha, \beta \in V$ 有

$$f(\alpha, \beta) = (\mathcal{A}\alpha, \beta), \quad g(\alpha, \beta) = (\mathcal{B}\alpha, \beta)$$

利用线性映射的复合仍为线性映射可知 f, g 是双线性函数。

设 $\alpha = a_1v_1 + a_2v_2 + a_3v_3, \beta = b_1v_1 + b_2v_2 + b_3v_3$ ，由于 $\mathcal{A}v_1 = 0$ ，利用斜对称性可发现

$$\forall \beta \in V, \quad (v_1, \mathcal{A}\beta) = -(\mathcal{A}v_1, \beta) = 0$$

且由于 $(\beta, \mathcal{A}\beta) = -(\mathcal{A}\beta, \beta)$ ，由内积对称性可知其只能为 0，也即

$$\forall \beta \in V, \quad (\mathcal{A}\beta, \beta) = 0$$

由此，直接展开计算，消去 $(\mathcal{A}v_1, v_i)$ 、 $(v_1, \mathcal{A}v_i)$ 与 $(v_i, \mathcal{A}v_i)$ 项可得

$$f(\alpha, \beta) = a_2b_3(\mathcal{A}v_2, v_3) + a_3b_2(\mathcal{A}v_3, v_2) = (a_2b_3 - a_3b_2)(\mathcal{A}v_2, v_3)$$

同理得到

$$g(\alpha, \beta) = a_2b_3(\mathcal{B}v_2, v_3) + a_3b_2(\mathcal{B}v_3, v_2) = (a_2b_3 - a_3b_2)(\mathcal{B}v_2, v_3)$$

若 $(\mathcal{A}v_3, v_2) = 0$ ，可发现对任何 α, β 有 $f(\alpha, \beta) = 0$ ，于是 $(\mathcal{A}\alpha, \beta) = 0$ 恒成立，固定 α ，利用内积的非退化性可知 $\mathcal{A}\alpha$ 必须为 0，由于这对任何 α 成立，即得 \mathcal{A} 为零映射，矛盾。同理， $(\mathcal{B}v_3, v_2)$ 也非零。

综合以上讨论， $f(\alpha, \beta) = 0$ 当且仅当 $a_2b_3 = a_3b_2$ ， $g(\alpha, \beta) = 0$ 也当且仅当 $a_2b_3 = a_3b_2$ ，于是两者零点集合相同，由 (b) 存在非零的 λ 使得 $f = \lambda g$ ，将其写为

$$\forall \alpha, \beta \in V, \quad (\mathcal{A}\alpha, \beta) = \lambda(\mathcal{B}\alpha, \beta)$$

* 若不用 (b) 中结论，也可直接令 $\lambda = \frac{(\mathcal{A}v_2, v_3)}{(\mathcal{B}v_2, v_3)}$ 后计算得 $f = \lambda g$ 。

利用线性性改写为

$$\forall \alpha, \beta \in V, \quad ((\mathcal{A} - \lambda\mathcal{B})\alpha, \beta) = 0$$

与之前讨论类似，由内积的非退化性可知 $(\mathcal{A} - \lambda\mathcal{B})\alpha$ 对任何 α 为 0，由此其为零映射，这就得到了

$$\mathcal{A} = \lambda\mathcal{B}$$

* 本题并不一定需要直接利用 (b) 中结论，但核心思路构造双线性函数考虑是从其中得到的。

(iv) 若 g 或 h 为零映射，由定义可发现 f 为零映射，从而取 l 为零映射、 $\lambda = 1$ 即符合要求，下假设 g 与 h 均不为零映射。记 $U = \text{Ker } g, W = \text{Ker } h$ ，则它们都是 V 的真子空间。我们下面证明 $U = W$ 。

记 f 的零点集合为 $K \subset V \times V$ 。首先，根据定义可知 $f(\alpha, \beta) = 0$ 当且仅当 $g(\alpha) = 0$ 或 $h(\beta) = 0$ ，从而有 ($U \times V$ 代表 $g(\alpha) = 0$ 时 β 可任取， $V \times W$ 代表 $h(\beta) = 0$ 时 α 可任取)

$$K = (U \times V) \cup (V \times W)$$

另一方面，由于 f 对称或斜对称，可知 $f(\beta, \alpha)$ 的零点集合与 f 相同，而 $f(\beta, \alpha) = g(\beta)h(\alpha)$ ，类似讨论即得

$$K = (W \times V) \cup (V \times U)$$

由此，对任何 $u \in U$ ，由于 U 为真子空间可取 $x \notin u$ ，由 $(u, x) \in U \times V$ 根据 K 的第一个表达式可知 $(u, x) \in K$ ，但由于 $x \notin U$ ， $(u, x) \notin V \times U$ ，根据 K 的第二个表达式只能 $(u, x) \in W \times V$ ，从而 $u \in W$ 。同理，对任何 $w \in W$ 可推出 $w \in U$ ，这就得到了 $U = W$ 。利用 (a)，存在非零的 λ 使得 $g = \lambda h$ ，于是再取 $l = h$ 即可发现符合要求。

* 另一个值得注意的地方是，我们的所有证明过程几乎都单独讨论了零映射，这本质上是由于平移零点必须依赖非零值。

(d) 对 n 进行归纳。我们已经证明了 n 为 1、2 时成立，下假设结论对 $n - 1$ 成立，考虑 n 的情况。

类似 $n = 2$ 时的证明，固定第一个分量时 f, g 均为 $n - 1$ 重线性函数，从而根据 $n - 1$ 的情况可知

$$\forall \alpha \in V, \quad \exists \lambda_\alpha \in \mathbb{K}, \lambda_\alpha \neq 0, \quad \forall \alpha_2, \dots, \alpha_n \in V, \quad g(\alpha, \alpha_2, \dots, \alpha_n) = \lambda_\alpha f(\alpha, \alpha_2, \dots, \alpha_n)$$

若 f 为零映射，由零点集合相同可知 g 为零映射，取 $\lambda = 1$ 即可，否则存在某组 $\beta_1, \dots, \beta_n \in V$ 使得 $f(\beta_1, \dots, \beta_n) \neq 0$ ，从而可记 $\lambda = \lambda_{\beta_1}$ ，其满足

$$\forall \alpha_2, \dots, \alpha_n \in V, \quad g(\beta_1, \alpha_2, \dots, \alpha_n) = \lambda f(\beta_1, \alpha_2, \dots, \alpha_n)$$

同理，固定第二个分量类似讨论得到

$$\forall \alpha_1, \alpha_3, \dots, \alpha_n \in V, \quad g(\alpha_1, \beta_2, \alpha_3, \dots, \alpha_n) = \lambda f(\alpha_1, \beta_2, \alpha_3, \dots, \alpha_n)$$

固定每个分量都有类似的结论，直到固定最后一个分量得到

$$\forall \alpha_1, \dots, \alpha_{n-1} \in V, \quad g(\alpha_1, \dots, \alpha_{n-1}, \beta_n) = \lambda f(\alpha_1, \dots, \alpha_{n-1}, \beta_n)$$

下面考虑一般的 $f(\alpha_1, \dots, \alpha_n)$ 。我们的核心思路是，利用 α_i 与 β_i 组合出零点。

具体来说，我们需要证明如下引理：对任何 $\alpha_1, \dots, \alpha_n \in V$ ，存在 $\mu_1, \dots, \mu_n \in \mathbb{K}$ ，使得

$$f(\alpha_1 - \mu_1 \beta_1, \dots, \alpha_n - \mu_n \beta_n) = 0$$

证明：

上式可展开成

$$f(\alpha_1 - \mu_1 \beta_1, \dots, \alpha_{n-1} - \mu_{n-1} \beta_{n-1}, \alpha_n) - \mu_n f(\alpha_1 - \mu_1 \beta_1, \dots, \alpha_{n-1} - \mu_{n-1} \beta_{n-1}, \beta_n) = 0$$

只要存在 μ_1, \dots, μ_{n-1} 使得 μ_n 前的系数不为 0，取

$$\mu_n = \frac{f(\alpha_1 - \mu_1 \beta_1, \dots, \alpha_{n-1} - \mu_{n-1} \beta_{n-1}, \alpha_n)}{f(\alpha_1 - \mu_1 \beta_1, \dots, \alpha_{n-1} - \mu_{n-1} \beta_{n-1}, \beta_n)}$$

即可。下面证明一定存在这样的 μ_1, \dots, μ_{n-1} 。

若否，对任何 μ_1, \dots, μ_{n-1} 均有

$$f(\alpha_1 - \mu_1 \beta_1, \dots, \alpha_{n-1} - \mu_{n-1} \beta_{n-1}, \beta_n) = 0$$

展开倒数第二项得到

$$f(\alpha_1 - \mu_1\beta_1, \dots, \beta_{n-2}, \alpha_{n-1}, \beta_n) - \mu_{n-1}f(\alpha_1 - \mu_1\beta_1, \dots, \beta_{n-1}, \beta_n) = 0$$

只要 $f(\alpha_1 - \mu_1\beta_1, \dots, \alpha_{n-2} - \mu_{n-2}\beta_{n-2}, \beta_{n-1}, \beta_n) \neq 0$, 即可取 μ_{n-1} 使得上式不为 0, 矛盾。由此可推出对任何 μ_1, \dots, μ_{n-2} 必然有

$$f(\alpha_1 - \mu_1\beta_1, \dots, \alpha_{n-2} - \mu_{n-2}\beta_{n-2}, \beta_{n-1}, \beta_n) = 0$$

继续展开倒数第三项, 重复此过程, 可发现最终可以得到对任何 μ_1 有

$$f(\alpha_1 - \mu_1\beta_1, \beta_2, \dots, \beta_n) = 0$$

但上式左即

$$f(\alpha_1, \beta_2, \dots, \beta_n) - \mu_1 f(\beta_1, \beta_2, \dots, \beta_n)$$

由于 μ_1 前的系数不为 0, 总可取到 μ_1 使得整体不为 0, 矛盾, 于是得证。

由此, 利用零点相同可得

$$g(\alpha_1 - \mu_1\beta_1, \dots, \alpha_n - \mu_n\beta_n) = 0 = \lambda f(\alpha_1 - \mu_1\beta_1, \dots, \alpha_n - \mu_n\beta_n)$$

将左右两侧完全展开, 可发现除了 $g(\alpha_1, \dots, \alpha_n)$ 与 $f(\alpha_1, \dots, \alpha_n)$ 外的项至少有一个位置是 β_i , 因此根据固定每一个分量的结果可知左侧与右侧的 λ 倍相互抵消, 最后只剩下

$$g(\alpha_1, \dots, \alpha_n) = \lambda f(\alpha_1, \dots, \alpha_n)$$

由于定义保证了 λ 不为 0, 这就得到了结论的证明。

4. * 大部分自伴、斜自伴、正交变换的性质都可被正规包含, 也即此题的讨论。由此, 在此题之外需要特别注意一些特殊性质, 如对称阵当且仅当可正交相似对角化、正交阵当且仅当所有行/列向量相互正交。

(a) 设 x 是 \mathcal{A} 关于特征值 λ 的特征向量, 则我们只需要证明 $\mathcal{A}^*x - \lambda x = 0$ 。利用内积对称性与双线性性直接展开计算可得

$$(\mathcal{A}^*x - \lambda x, \mathcal{A}^*x - \lambda x) = (\mathcal{A}^*x, \mathcal{A}^*x) - 2\lambda(\mathcal{A}^*x, x) + \lambda^2(x, x)$$

由伴随变换定义

$$(\mathcal{A}^*x, x) = (x, \mathcal{A}x) = (\mathcal{A}x, x)$$

由伴随变换定义与正规变换性质

$$(\mathcal{A}^*x, \mathcal{A}^*x) = (\mathcal{A}\mathcal{A}^*x, x) = (\mathcal{A}^*\mathcal{A}x, x) = (\mathcal{A}x, \mathcal{A}x)$$

从而代入可得

$$(\mathcal{A}^*x - \lambda x, \mathcal{A}^*x - \lambda x) = (\mathcal{A}x, \mathcal{A}x) - 2\lambda(\mathcal{A}x, x) + \lambda^2(x, x)$$

而右侧即可合并为 $(\mathcal{A}x - \lambda x, \mathcal{A}x - \lambda x)$, 由特征向量定义可知其为 0, 得证。

* 需要熟悉正规变换的等价定义 $(\mathcal{A}x, \mathcal{A}y) = (\mathcal{A}^*x, \mathcal{A}^*y)$ 。

* 在内积空间中, 证明向量为 0 常通过模长为 $\mathbf{0}$ 进行。

(b) 设 x, y 是 \mathcal{A} 关于特征值 λ, μ 的特征向量, 且 $\lambda \neq \mu$, 利用 (a) 有

$$(\mathcal{A}x, y) - (x, \mathcal{A}y) = (x, \mathcal{A}^*y) - (x, \mathcal{A}y) = \mu(x, y) - \mu(x, y) = 0$$

另一方面

$$(\mathcal{A}x, y) - (x, \mathcal{A}y) = \lambda(x, y) - \mu(x, y) = (\lambda - \mu)(x, y)$$

从而由 $\lambda \neq \mu$ 即得 $(x, y) = 0$, 得证。

(c) 我们先证明 $\text{Ker } \mathcal{A}$ 与 $\text{Im } \mathcal{A}$ 正交。对任何 $x \in \text{Ker } \mathcal{A}$ 、 $y \in \text{Im } \mathcal{A}$, 设 $y = \mathcal{A}z$ 、 $z \in V$, 有

$$(x, y) = (x, \mathcal{A}z) = (\mathcal{A}^*x, z)$$

由于 $x \in \text{Ker } \mathcal{A}$ 也即 x 是 \mathcal{A} 关于特征值 0 的特征向量, 由 (a) 得 x 也是 \mathcal{A}^* 关于特征值 0 的特征向量, 于是 $\mathcal{A}^*x = 0$, 这就得到了 $(x, y) = 0$ 。

利用正交补的定义, $\text{Ker } \mathcal{A} \subset (\text{Im } \mathcal{A})^\perp$, 另一方面, 由于 V 维数有限, 利用第一同构定理可知

$$\dim V = \dim \text{Ker } \mathcal{A} + \dim \text{Im } \mathcal{A}$$

从而 $\dim \text{Ker } \mathcal{A} = \dim(\text{Im } \mathcal{A})^\perp$ 。综合两方面即得 $\text{Ker } \mathcal{A} = (\text{Im } \mathcal{A})^\perp$ 。

(d) 若 \mathcal{A} 是正交投影, 其为投影映射, 从而取 $k = 1$ 即可, 下面证明另一边, 也即从 $k \in \mathbb{N}, \mathcal{A}^k = \mathcal{A}^{k+1}$ 推出 \mathcal{A} 是正交投影。

当 $k = 0$ 时, 即代表 $\mathcal{A} = \mathcal{I}$, 为对全空间的正交投影, 下设 $k \geq 1$ 。移项可得 $\mathcal{A}^k(\mathcal{A} - \mathcal{I}) = \mathcal{O}$, 从而 $x^k(x - 1)$ 是 \mathcal{A} 的一个化零多项式, \mathcal{A} 的特征值只有 0 或 1。由此可写出根子空间分解

$$V = \text{Ker } \mathcal{A}^k \oplus \text{Ker } (\mathcal{A} - \mathcal{I})$$

我们先证明 $\text{Ker } \mathcal{A} = \text{Ker } \mathcal{A}^k$ 。

证明:

* 此证明基本与期中复习题第 3 题 (b) 等价。

由 Ker 定义可得

$$\text{Ker } \mathcal{A} \subset \text{Ker } \mathcal{A}^k$$

另一方面

$$\dim \text{Ker } \mathcal{A} = \dim V - \dim \text{Im } \mathcal{A}, \quad \dim \text{Ker } \mathcal{A}^k = \dim V - \dim \text{Im } \mathcal{A}^k$$

由此只需证明 $\dim \text{Im } \mathcal{A} = \dim \text{Im } \mathcal{A}^k$ 。

由 (c) 中已证, $V = \text{Ker } \mathcal{A} \oplus \text{Im } \mathcal{A}$ 。对任何 $y \in \text{Im } \mathcal{A}$, 设 $y = \mathcal{A}x$ 、 $x \in V$, 将 x 分解为 $z + w$ 使得 $z \in \text{Ker } \mathcal{A}$ 、 $w \in \text{Im } \mathcal{A}$, 有

$$\mathcal{A}(w) = \mathcal{A}(w + z) = \mathcal{A}(x) = y$$

由此, $\text{Im } \mathcal{A}$ 中的任何元素在 $\text{Im } \mathcal{A}$ 中有原像, 即 $\text{Im } \mathcal{A} \subset \mathcal{A}(\text{Im } \mathcal{A})$ 。根据 $\text{Im } \mathcal{A}$ 的定义, 另一边包含关系成立, 于是

$$\mathcal{A}(\text{Im } \mathcal{A}) = \text{Im } \mathcal{A}$$

再次利用像空间定义可得

$$\text{Im } \mathcal{A}^k = \mathcal{A}(\text{Im } \mathcal{A}^{k-1})$$

从而利用归纳法即得 $\text{Im } \mathcal{A}^k = \mathcal{A}(\text{Im } \mathcal{A}^{k-1}) = \mathcal{A}(\text{Im } \mathcal{A}) = \text{Im } \mathcal{A}$ 对任何正整数 k 成立, 得证。

由此, 我们得到了分解

$$V = \text{Ker } \mathcal{A} \oplus \text{Ker } (\mathcal{A} - \mathcal{I})$$

另一方面, 由于 $\text{Ker } (\mathcal{A} - \mathcal{I})$ 代表所有满足 $\mathcal{A}x = x$ 的向量 x , 其原像为自身, 有 $\text{Ker } (\mathcal{A} - \mathcal{I}) \subset \text{Im } \mathcal{A}$ 。利用 (c) 可知

$$V = \text{Ker } \mathcal{A} \oplus \text{Im } \mathcal{A}$$

于是考虑维数得只能 $\text{Ker } (\mathcal{A} - \mathcal{I}) = \text{Im } \mathcal{A}$ 。

记 $U = \text{Im } \mathcal{A}$, 利用 $\text{Ker}(\mathcal{A} - \mathcal{I}) = \text{Im } \mathcal{A}$ 可知 \mathcal{A} 在 U 上为恒等映射, 利用 $\text{Ker } \mathcal{A} = U^\perp$ 可知 \mathcal{A} 在 U^\perp 上为零映射, 这即代表 \mathcal{A} 是到 U 的正交投影, 从而得证。

* 这题事实上可以直接通过正规变换的正交相似标准形得到, 不过, 为了展现逻辑顺序, 我们采取了更基本的方式证明。

(e) • U 是 \mathcal{A}^* -不变子空间

设 U 的一组标准正交基是 u_1, \dots, u_r , 扩充为 V 的标准正交基 $u_1, \dots, u_r, v_1, \dots, v_{n-r}$ 。

对任何 u_i , $i = 1, \dots, r$, 利用标准正交基的性质可知

$$\mathcal{A}u_i = \sum_{j=1}^r (\mathcal{A}u_i, u_j)u_j + \sum_{k=1}^{n-r} (\mathcal{A}u_i, v_k)v_k$$

由不变子空间定义可知 $\mathcal{A}u_i \in U$, 从而后 $n-r$ 个分量为 0, 进一步写成

$$\mathcal{A}u_i = \sum_{j=1}^r (\mathcal{A}u_i, u_j)u_j$$

直接展开计算, 利用 u_i 两两正交可发现

$$(\mathcal{A}u_i, \mathcal{A}u_i) = \left(\sum_{j=1}^r (\mathcal{A}u_i, u_j)u_j, \sum_{j=1}^r (\mathcal{A}u_i, u_j)u_j \right) = \sum_{j=1}^r (\mathcal{A}u_i, u_j)^2$$

另一方面, 同理可将 \mathcal{A}^*u_i 写成

$$\mathcal{A}^*u_i = \sum_{j=1}^r (\mathcal{A}^*u_i, u_j)u_j + \sum_{k=1}^{n-r} (\mathcal{A}^*u_i, v_k)v_k$$

也同样可以直接利用正交性展开计算得

$$(\mathcal{A}^*u_i, \mathcal{A}^*u_i) = \sum_{j=1}^r (\mathcal{A}^*u_i, u_j)^2 + \sum_{k=1}^{n-r} (\mathcal{A}^*u_i, v_k)^2$$

进一步由伴随变换性质写为

$$(\mathcal{A}^*u_i, \mathcal{A}^*u_i) = \sum_{j=1}^r (\mathcal{A}u_j, u_i)^2 + \sum_{k=1}^{n-r} (\mathcal{A}^*u_i, v_k)^2$$

由正规变换性质 $(\mathcal{A}^*u_i, \mathcal{A}^*u_i) = (\mathcal{A}u_i, \mathcal{A}u_i)$, 从而对任何 $i = 1, \dots, r$ 可最终写出

$$\sum_{j=1}^r (\mathcal{A}u_i, u_j)^2 = \sum_{j=1}^r (\mathcal{A}u_j, u_i)^2 + \sum_{k=1}^{n-r} (\mathcal{A}^*u_i, v_k)^2$$

将两侧对所有 i 求和, 可以发现, 左侧第一项都是对所有 i, j 的 $(\mathcal{A}u_i, u_j)^2$ 求和, 右侧第一项是对所有 i, j 的 $(\mathcal{A}u_j, u_i)^2$ 求和, 只相差次序, 因此可以相互消去, 最终得到

$$\sum_{i=1}^r \sum_{k=1}^{n-r} (\mathcal{A}^*u_i, v_k)^2 = 0$$

因此对任何 $i = 1, \dots, r$ 、 $k = 1, \dots, n-r$ 有 $(\mathcal{A}^*u_i, v_k) = 0$, 代回之前的展开即得

$$\mathcal{A}^*u_i = \sum_{j=1}^r (\mathcal{A}^*u_i, u_j)u_j$$

于是 $\mathcal{A}^*u_i \in U$ 对 $i = 1, \dots, r$ 成立, 再由子空间封闭性即得对任何 $u \in U$ 有 $\mathcal{A}^*u \in U$, 从而 U 是 \mathcal{A}^* -不变子空间。

- U^\perp 是 \mathcal{A} -不变子空间

由于 U 是 \mathcal{A}^* -不变子空间，对任何 $\alpha \in U$ 有 $\mathcal{A}^*\alpha \in U$ ，于是对任何 $\beta \in U^\perp$ 有

$$(\mathcal{A}^*\alpha, \beta) = 0$$

另一方面， $(\mathcal{A}^*\alpha, \beta) = (\alpha, \mathcal{A}\beta)$ 。上式证明了 $\mathcal{A}\beta$ 满足对任何 $\alpha \in U$ 都有 $(\alpha, \mathcal{A}\beta) = 0$ ，因此 $\mathcal{A}\beta \in U^\perp$ ，这就得到了证明。

(f) 在 (e) 中已经证明了 U 是 \mathcal{A}^* -不变子空间，从而限制映射 $(\mathcal{A}^*)|_U$ 的确存在。

根据定义， $(\mathcal{A}|_U)^*$ 是满足

$$\forall u, v \in U, \quad (\mathcal{A}|_U u, v) = (u, \mathcal{B}v)$$

的唯一 U 上线性变换 \mathcal{B} ，而由限制映射定义可发现

$$\forall u, v \in U, \quad (\mathcal{A}|_U u, v) = (\mathcal{A}u, v) = (u, \mathcal{A}^*v) = (u, \mathcal{A}^*|_U v)$$

从而 $\mathcal{A}^*|_U$ 是满足此性质的线性变换，根据唯一性得必然 $(\mathcal{A}|_U)^* = \mathcal{A}^*|_U$ 。

由此，利用 $\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}$ ，直接计算可知

$$\mathcal{A}|_U(\mathcal{A}|_U)^* = \mathcal{A}|_U(\mathcal{A}^*)|_U$$

根据限制映射的定义可发现对任何 $u \in U$ 有（第二个等号是由于不变子空间定义， $\mathcal{A}^*(u) \in U$ ）

$$\mathcal{A}|_U(\mathcal{A}^*)|_U(u) = \mathcal{A}|_U(\mathcal{A}^*(u)) = \mathcal{A}(\mathcal{A}^*(u)) = \mathcal{A}^*(\mathcal{A}(u)) = \mathcal{A}^*(\mathcal{A}|_U(u)) = (\mathcal{A}^*)|_U \mathcal{A}|_U(u)$$

从而即得到 $\mathcal{A}|_U$ 也正规。

(g) • 不变子空间存在性

我们证明更一般的结论： n 维实线性空间上的线性变换在一维或二维不变子空间（可参考教材 9.5 节例 15）。

设 \mathcal{A} 在某组基下的矩阵表示为 A 。利用上学期结论，实方阵 A 看作复方阵时所有特征值成对出现，也即所有复特征值可以写为

$$\lambda_1, \dots, \lambda_r, a_1 + b_1i, a_1 - b_1i, \dots, a_k + b_ki, a_k - b_ki$$

这里所有 λ_i 、 a_i 为实数， b_i 为正实数。

若 $r > 0$ ， \mathcal{A} 存在实特征值 λ_1 与对应的实特征向量 α_1 ，由定义可知 $\langle \alpha_1 \rangle$ 是 \mathcal{A} 的一维不变子空间，已经得证。

否则，由于 $r + 2k$ 应等于 A 的阶数，必有 $k > 0$ 。考虑 A 看作复方阵的特征值 $a_1 + b_1i$ 。假设其一个（复）特征向量是 $x + yi$ ，这里 x 、 y 为实向量。由于 A 为实方阵，利用

$$A(x + yi) = (a_1 + b_1i)(x + yi)$$

对比实部、虚部可计算得

$$Ax = a_1x - b_1y, \quad Ay = b_1x + a_1y$$

由此， $\langle x, y \rangle$ 为 A 的一个不变子空间，于是考虑坐标为 x 的向量 α 与坐标为 y 的向量 β ， $\langle \alpha, \beta \rangle$ 为 A 的一个不变子空间。由于其被两个向量生成，必然为一维或二维。

为了进行接下来的证明，我们再补充一个性质：若 A 是实正规变换 \mathcal{A} 在标准正交基下的矩阵表示，且其有特征值 $a_1 + b_1i$ ，上面构造的不变子空间 $\langle \alpha, \beta \rangle$ 是二维的，且 α 、 β 模长相等且正交， $\mathcal{A}|_{\langle \alpha, \beta \rangle}$ 在基 α, β 下的矩阵表示为

$$\begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix}$$

证明：

我们用上标 H 表共轭转置。

由于坐标映射的同构性，只要证明 x 与 y 线性无关即有 α 与 β 线性无关。

首先，根据 $Ax = a_1x - b_1y$ 、 $Ay = b_1x + a_1y$ ，且 $b_1 > 0$ ，若 x 或 y 为 0 可推出 $x = y = 0$ ，与特征向量矛盾。另一方面，计算可知

$$A(x - yi) = (a_1 - b_1i)(x - yi)$$

由此 $x - yi$ 是属于特征值 $a_1 - b_1i$ 的特征向量。将 A 看作复正规变换（由于 $A^T A = AA^T$ 且 A 为实方阵，也有 $A^H A = AA^H$ ），仍可与 (b) 相同证明其不同特征值的特征向量正交，从而

$$(x - yi)^H(x + yi) = 0$$

由于 $x - yi$ 的共轭即为 $x + yi$ ，左侧事实上是 $x + yi$ 的转置，从而展开并对比实部、虚部得到

$$x^T x - y^T y = 0, \quad x^T y + y^T x = 0$$

由一维方阵等于其转置可知第二个式子得到 $x^T y = 0$ ，第一个式子得到 $x^T x = y^T y$ 。利用标准正交基的性质，这即得到 α 与 β 正交、 $(\alpha, \alpha) = (\beta, \beta)$ ，从而得证。

最后，利用坐标与矩阵表示的性质可知

$$\mathcal{A}(\alpha) = a_1\alpha - b_1\beta, \quad \mathcal{A}(\beta) = b_1\alpha + a_2\beta$$

这就直接由矩阵表示的定义得到了 $\mathcal{A}|_{\langle \alpha, \beta \rangle}$ 在基 α, β 下的矩阵表示符合要求。

- 标准形构造

我们下面证明，假设 \mathcal{A} 在某组标准正交基下的矩阵表示为 A ，且 A 看作复方阵的特征值为

$$\lambda_1, \dots, \lambda_r, a_1 + b_1i, a_1 - b_1i, \dots, a_k + b_ki, a_k - b_ki$$

其中所有 λ_i 、 a_i 为实数， b_i 为正实数，则存在一组标准正交基使得 \mathcal{A} 在其下的矩阵表示为

$$\text{diag}\left(\lambda_1, \dots, \lambda_r, \begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix}, \dots, \begin{pmatrix} a_k & b_k \\ -b_k & a_k \end{pmatrix}\right)$$

我们利用类似归纳的思路证明。先分两种情况构造不变子空间（注意根据矩阵表示的性质， \mathcal{A} 的实特征值与矩阵表示 A 的实特征值相同）：

- \mathcal{A} 有实特征值

此时，取出 \mathcal{A} 的实特征值 λ 与对应的特征向量 α 。由于特征向量乘非零倍数还是特征向量，记 $U = \langle \alpha \rangle$ ，单位化的特征向量

$$\gamma = \frac{1}{\sqrt{\alpha, \alpha}}\alpha$$

由内积双线性性可直接得到 γ 为单位向量，进一步根据特征向量定义可知 $\mathcal{A}|_U$ 在 γ 下的矩阵表示即为 (λ) 。

- \mathcal{A} 无实特征值

此时， A 必然有复特征值，假设其中一个为 $a + bi$ 、 $b > 0$ 。根据上一部分的证明，存在模长相等且正交的 α 、 β 使得 $U = \langle \alpha, \beta \rangle$ 是二维不变子空间，且 $\mathcal{A}|_U$ 在基 α, β 下的矩阵表示为

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

设 α, β 的模长均为 c , 我们将 α, β 规范化成为 U 的标准正交基:

$$\gamma_1 = \frac{1}{c}\alpha, \quad \gamma_2 = \frac{1}{c}\beta$$

利用矩阵表示的定义, 由于

$$\mathcal{A}(\alpha) = a\alpha - b\beta, \quad \mathcal{A}(\beta) = b_1\alpha + a_2\beta$$

且 \mathcal{A} 为线性映射, 两侧同除以 c 可得

$$\mathcal{A}(\gamma_1) = a\gamma_1 - b\gamma_2, \quad \mathcal{A}(\gamma_2) = b\gamma_1 + a\gamma_2$$

于是 $\mathcal{A}|_U$ 在基 γ_1, γ_2 下的矩阵表示不变。

再给出归纳过程:

- 基础情况

若 A 只有一个实特征值, 则上述第一种方法构造的 U 即为全空间, 从而已经符合要求。

若 A 无实特征值且有一对非实特征值 $a \pm bi$, 则上述第二种方法构造的 U 即为全空间。

我们下面假设 A 的实特征值个数少于 r 个或非实特征值对数少于 k 对时结论成立, 考虑实特征值 r 个、非实特征值 k 对的情况。此时假设空间维数为 n , A 的所有特征值为

$$\lambda_1, \dots, \lambda_r, a_1 + b_1i, a_1 - b_1i, \dots, a_k + b_ki, a_k - b_ki$$

其中所有 λ_i 、 a_i 为实数, b_i 为正实数。

- 有实特征值时的降阶

若 $r > 0$, 设其有一个实特征值 λ_1 , 利用上述第一种方法构造 U , 并设构造的基为 γ 。

将 γ 扩充为全空间的标准正交基 $\gamma, \gamma_2, \dots, \gamma_n$ 。利用正交补性质 (可见本讲义 25.2.2) 可知 $U^\perp = \langle \gamma_2, \dots, \gamma_n \rangle$ 。由于 U 为 \mathcal{A} 的不变子空间, U^\perp 也为 \mathcal{A} 的不变子空间, 从而利用不变子空间的矩阵表示性质 (可见本讲义 21.2.1 开头) 可得 \mathcal{A} 在这组基下的矩阵表示为 (未写出元素为 0)

$$\begin{pmatrix} \lambda_1 \\ & A_0 \end{pmatrix}$$

这里 A_0 为 $\mathcal{A}|_{U^\perp}$ 在 $\gamma_2, \dots, \gamma_n$ 下的矩阵表示。

由于 \mathcal{A} 在不同基下的矩阵表示相似, 实方阵相似则它们看作复方阵也相似 (实可逆阵也是复可逆阵), 此矩阵的特征值应与 A 完全相同, 从而考虑特征多项式可知 A_0 的所有复特征值为

$$\lambda_2, \dots, \lambda_r, a_1 + b_1i, a_1 - b_1i, \dots, a_k + b_ki, a_k - b_ki$$

由于 A_0 实特征值少于 r 个, 且利用 (f) 可知 $\mathcal{A}|_{U^\perp}$ 也正规, 符合归纳假设, 存在 U^\perp 的一组标准正交基 η_2, \dots, η_n 使得 $\mathcal{A}|_{U^\perp}$ 在这组基下的矩阵表示为

$$\text{diag} \left(\lambda_2, \dots, \lambda_r, \begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix}, \dots, \begin{pmatrix} a_k & b_k \\ -b_k & a_k \end{pmatrix} \right)$$

考虑 $\gamma_1, \eta_2, \dots, \eta_n$, 由于 U 与 U^\perp 中任何向量均正交, 这仍然是一组 V 的标准正交基, 且再次利用不变子空间的矩阵表示性质可知 \mathcal{A} 在这组基下的矩阵表示为

$$\text{diag} \left(\lambda_1, \dots, \lambda_r, \begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix}, \dots, \begin{pmatrix} a_k & b_k \\ -b_k & a_k \end{pmatrix} \right)$$

这就得到了证明。

- 无实特征值时的降阶

若 $r = 0$, 必然 k 至少为 1, 这时 A 的特征值为

$$a_1 + b_1i, \quad a_1 - b_1i, \quad \dots, \quad a_k + b_ki, \quad a_k - b_ki$$

对 $a_1 \pm b_1i$ 利用上述第二种方法构造 U , 并设构造的基为 γ_1, γ_2 。将它们扩充为全空间的标准正交基 $\gamma_1, \gamma_2, \dots, \gamma_n$ 。利用正交补性质可知 $U^\perp = \langle \gamma_3, \dots, \gamma_n \rangle$ 。由于 U 为 \mathcal{A} 的不变子空间, U^\perp 也为 \mathcal{A} 的不变子空间, 从而利用不变子空间的矩阵表示性质可得 \mathcal{A} 在这组基下的矩阵表示为 (未写出元素为 0)

$$\begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \\ & A_0 \end{pmatrix}$$

这里 A_0 为 $\mathcal{A}|_{U^\perp}$ 在 $\gamma_3, \dots, \gamma_n$ 下的矩阵表示。

由于 \mathcal{A} 在不同基下的矩阵表示相似, 实方阵相似则它们看作复方阵也相似 (实可逆阵也是复可逆阵), 此矩阵的特征值应与 A 完全相同, 从而考虑特征多项式可知 A_0 的所有复特征值为

$$a_2 + b_2i, \quad a_2 - b_2i, \quad \dots, \quad a_k + b_ki, \quad a_k - b_ki$$

由于 A_0 非实特征值少于 k 对, 且利用 (f) 可知 $\mathcal{A}|_{U^\perp}$ 也正规, 符合归纳假设, 存在 U^\perp 的一组标准正交基 η_3, \dots, η_n 使得 $\mathcal{A}|_{U^\perp}$ 在这组基下的矩阵表示为

$$\text{diag} \left(\begin{pmatrix} a_2 & b_2 \\ -b_2 & a_2 \end{pmatrix}, \dots, \begin{pmatrix} a_k & b_k \\ -b_k & a_k \end{pmatrix} \right)$$

考虑 $\gamma_1, \gamma_2, \eta_3, \dots, \eta_n$, 由于 U 与 U^\perp 中任何向量均正交, 这仍然是一组 V 的标准正交基, 且再次利用不变子空间的矩阵表示性质可知 \mathcal{A} 在这组基下的矩阵表示为

$$\text{diag} \left(\begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix}, \dots, \begin{pmatrix} a_k & b_k \\ -b_k & a_k \end{pmatrix} \right)$$

这就得到了证明。

* 综合以上讨论, 我们最终得出了 \mathcal{A} 在正交相似下能表为的最简形式。在矩阵论中, 这称为实正规方阵的正交相似标准形。

* 这道题虽然非常复杂, 但希望大家能耐心阅读证明, 尤其是掌握正交等内容如何与前半学期的不变子空间知识结合来构造矩阵表示, 这大概是期末范围内可能最难的部分了。

(h) 我们将 (f) 中构造的矩阵称为 \mathcal{A} 的矩阵表示标准形。由于此标准形只与 \mathcal{A} 的某组标准正交基下矩阵表示 A 的所有复特征值有关, 而 \mathcal{A}^* 在相同基下的矩阵表示为 A^T , 由上学期结论, A^T 看作复方阵的所有特征值与 A 相同, 于是 \mathcal{A}^* 的矩阵表示标准形与 \mathcal{A} 相同。由于不同标准正交基下的矩阵表示正交相似, 任意一组基下 \mathcal{A} 与 \mathcal{A}^* 的矩阵表示均正交相似。

5. * 这题真的是往年原题... 这就是为什么希望大家仔细阅读第 4 题 (f) 的证明, 因为真不是不可能考...

(a) 我们将仿照第 4 题 (f) 进行证明, 不过采用更矩阵论的技巧。我们证明更强的结论, 也即若 A 的所有特征值为

$$\lambda_1, \quad \dots, \quad \lambda_r, \quad a_1 + b_1i, \quad a_1 - b_1i, \quad \dots, \quad a_k + b_ki, \quad a_k - b_ki$$

其中所有 λ_i, a_i 为实数, b_i 为正实数, 则存在正交阵 Q 使得 $Q^T A Q$ 是对角块为 $\lambda_1, \dots, \lambda_r, A_1, \dots, A_k$ 的分块上三角阵, 其中每个 A_i 是特征值为 $a_i \pm b_i i$ 的二阶实方阵。

完全类似对 A 的特征值进行分类:

- A 若为一阶且存实特征值或二阶且不存实特征值，均已经符合题目要求，下面假设 A 的实特征值个数小于 r 个或非实特征值对数小于 k 对时已经成立，考虑实特征值 r 个、非实特征值 k 对的情况。此时假设 A 为 n 阶方阵， A 的特征值为

$$\lambda_1, \dots, \lambda_r, a_1 + b_1i, a_1 - b_1i, \dots, a_k + b_ki, a_k - b_ki$$

其中所有 λ_i 、 a_i 为实数， b_i 为正实数。

- 存实特征值情况

若 A 存在实特征值 λ_1 ，设其一个特征向量为 α ，考虑单位化的特征向量

$$\gamma = \frac{1}{\sqrt{\alpha^T \alpha}} \alpha$$

可发现 $\gamma^T \gamma = 1$ 。将 γ 扩充为 \mathbb{R}^n 的一组标准正交基 $\gamma, \gamma_2, \dots, \gamma_n$ ，并将它们作为列拼成矩阵 P ，由上学期知识，方阵为正交阵当且仅当其列向量标准正交，由此 P 是正交阵。

利用分块矩阵知识可知 AP 的第一列是 $A\gamma = \lambda_1\gamma$ ，从而计算得 (x 与 A_0 为未知行向量/矩阵，左下角分块为零向量)

$$AP = P \begin{pmatrix} \lambda_1 & x \\ & A_0 \end{pmatrix}$$

同左乘 P^T 也即

$$P^T AP = \begin{pmatrix} \lambda_1 & x \\ & A_0 \end{pmatrix}$$

由于相似不改变特征值，右侧所有特征值应与 A 完全相同，从而考虑特征多项式可知 A_0 的所有复特征值为

$$\lambda_2, \dots, \lambda_r, a_1 + b_1i, a_1 - b_1i, \dots, a_k + b_ki, a_k - b_ki$$

由于 A_0 实特征值少于 r 个，符合归纳假设，存在 $n-1$ 阶正交阵 R 使得 $R^T A_0 R$ 是对角块为 $\lambda_2, \dots, \lambda_r$ 、 A_1, \dots, A_k 的分块上三角阵，其中每个 A_i 是特征值为 $a_i \pm b_i i$ 的二阶实方阵。进一步计算可得

$$\begin{pmatrix} 1 & \\ & R^T \end{pmatrix} P^T AP \begin{pmatrix} 1 & \\ & R \end{pmatrix}$$

是对角块为 $\lambda_1, \dots, \lambda_r$ 、 A_1, \dots, A_k 的分块上三角阵，其中每个 A_i 是特征值为 $a_i \pm b_i i$ 的二阶实方阵。

此外，利用正交阵定义与分块矩阵运算可直接验证

$$\begin{pmatrix} 1 & \\ & R^T \end{pmatrix} P^T = \left(P \begin{pmatrix} 1 & \\ & R \end{pmatrix} \right)^T$$

且

$$\begin{pmatrix} 1 & \\ & R^T \end{pmatrix} P^T P \begin{pmatrix} 1 & \\ & R \end{pmatrix} = I$$

于是记

$$Q = P \begin{pmatrix} 1 & \\ & R \end{pmatrix}$$

即得到了结论。

- 不存实特征值情况

此时 $r = 0$ ， A 的所有特征值为

$$a_1 + b_1i, a_1 - b_1i, \dots, a_k + b_ki, a_k - b_ki$$

与第 4 题 (f) 第一部分证明完全相同可知存在实向量 x, y 使得

$$A(x + yi) = (a_1 + b_1i)(x + yi)$$

且对比实部、虚部得到

$$Ax = a_1x - b_1y, \quad Ay = b_1x + a_1y$$

于是计算得

$$A(x - yi) = (a_1 - b_1i)(x - yi)$$

由于不同特征值的特征向量线性无关， $x + yi$ 与 $x - yi$ 看作复向量应线性无关，从而可知 x 与 y 必然线性无关，否则 $x + yi, x - yi$ 均能写成其中某个的复数倍。

将 x, y 进行正交化、规范化得到两个正交且模长为 1 的向量 γ_1, γ_2 ，将它们扩充为 \mathbb{R}^n 的一组标准正交基 $\gamma_1, \gamma_2, \dots, \gamma_n$ ，并将它们作为列拼成矩阵 P ，由上学期知识，方阵为正交阵当且仅当其列向量标准正交，由此 P 是正交阵。我们下面证明

$$AP = P \begin{pmatrix} A_1 & X \\ & A_0 \end{pmatrix}$$

这里 X 与 A_0 为未知矩阵，左下角分块为零矩阵， A_1 是特征值为 $a_1 \pm b_1i$ 的二阶实方阵。

证明：

由于 γ_1, γ_2 与 x, y 等价，可知 γ_1, γ_2 能被 x, y 表出，从而由线性性 $A\gamma_1, A\gamma_2$ 能被 Ax, Ay 表出。根据之前的等式可知 Ax, Ay 能被 x, y 表出，再次利用等价性可得 x, y 能被 γ_1, γ_2 表出，我们即最终得到了 $A\gamma_1, A\gamma_2$ 能被 γ_1, γ_2 表出。于是， A_1 下方的部分的确可以保证全为 0。

接下来，由定义可发现 A_1 是 A 作为线性变换限制在 $\langle \gamma_1, \gamma_2 \rangle$ 上，并在基 γ_1, γ_2 下的矩阵表示，而 $\langle \gamma_1, \gamma_2 \rangle = \langle x, y \rangle$ ，且根据之前的等式，此限制映射在基 x, y 下的矩阵表示是

$$\begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \\ & A_0 \end{pmatrix}$$

利用更换基时线性变换的矩阵表示相似，从而不改变特征值， A_1 的特征值应与上方矩阵完全相同，计算可知为 $a_1 \pm b_1i$ 。

同左乘 P^T 得到

$$P^T AP = \begin{pmatrix} A_1 & X \\ & A_0 \end{pmatrix}$$

由于相似不改变特征值，右侧所有特征值应与 A 完全相同，从而考虑特征多项式可知 A_0 的所有复特征值为

$$a_2 + b_2i, \quad a_2 - b_2i, \quad \dots, \quad a_k + b_ki, \quad a_k - b_ki$$

由于 A_0 非实特征值少于 k 个，符合归纳假设，存在 $n - 2$ 阶正交阵 R 使得 $R^T A_0 R$ 是对角块为 A_2, \dots, A_k 的分块上三角阵，其中每个 A_i 是特征值为 $a_i \pm b_i i$ 的二阶实方阵。进一步计算可得 (I_2 为二阶单位阵)

$$\begin{pmatrix} I_2 & \\ & R^T \end{pmatrix} P^T AP \begin{pmatrix} I_2 & \\ & R \end{pmatrix}$$

是对角块为 A_1, \dots, A_k 的分块上三角阵，其中每个 A_i 是特征值为 $a_i \pm b_i i$ 的二阶实方阵。与上一部分类似取

$$Q = P \begin{pmatrix} 1 & \\ & R \end{pmatrix}$$

即得到了结论。

(b) 保持 Q 为 (a) 中的定义, 利用 Q 的正交性与 tr 可交换得到

$$\text{tr}(AA^T) = \text{tr}(AQQ^TA^TQQ^T) = \text{tr}(Q^TAQQ^TA^TQ) = \text{tr}(Q^TAQ(Q^TAQ)^T)$$

设 $B = Q^TAQ$, 其各分量为 b_{ij} , 直接计算可发现

$$\text{tr}(BB^T) = \sum_{i=1}^n \sum_{j=1}^n b_{ij}^2 \geq \sum_{i=1}^r \lambda_i^2 + \sum_{j=1}^k s_j$$

这里 s_j 表示 A_j 各元素的平方和, 不等号是由于所有 λ_i 与所有 A_j 都是 B 的元素的一部分。

我们先证明, 若二阶实方阵 A 特征值为 $a \pm bi$, 这里 a 为实数, $b > 0$, 则 A 的四个元素平方和至少为 $2a^2 + 2b^2$, 且等号成立当且仅当以下二者之一成立:

$$A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \quad A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

证明:

设

$$A = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}$$

其特征多项式为 (左侧为直接计算, 右侧为利用特征值推出)

$$(\lambda - a_1)(\lambda - a_3) - a_2a_4 = (\lambda - a - bi)(\lambda - a + bi)$$

对比各项系数可知

$$a_1a_3 - a_2a_4 = a^2 + b^2, \quad a_1 + a_3 = 2a$$

利用第一个等式可知

$$2a^2 + 2b^2 = 2a_1a_3 - 2a_2a_4$$

利用 $(a_1 - a_3)^2 + (a_2 + a_4)^2 \geq 0$ 展开并移项即得到

$$2a_1a_3 - 2a_2a_4 \leq a_1^2 + a_3^2 + a_2^2 + a_4^2$$

从而不等式得证。若等号成立, 利用 $(a_1 - a_3)^2 + (a_2 + a_4)^2 = 0$ 可知须 $a_1 = a_3$ 、 $a_2 = -a_4$,

由 $a_1 + a_3 = 2a$ 知 $a_1 = a_3 = a$, 于是再由此时 $a_2a_4 = -b^2$ 可知 $a_4 = \pm b$ 。

由于 $2a^2 + 2b^2 = |a + bi|^2 + |a - bi|^2$, 即得

$$s_j \geq |a_j + b_j i|^2 + |a_j - b_j i|^2$$

从而我们已经证明了 $\text{tr}(BB^T)$ 大于等于 A 所有特征值模长平方和, 由 $\text{tr}(AA^T) = \text{tr}(BB^T)$ 即得证原不等式。

若等号成立, 根据上方两步不等号, 所有 λ_i 、 A_j 外的元素应当全为 0, 且每个 A_j 的形式也已经固定, 从而有

$$B = \text{diag}(\lambda_1, \dots, \lambda_r, A_1, \dots, A_k)$$

每个 A_j 为以下两者之一:

$$\begin{pmatrix} a_j & b_j \\ -b_j & a_j \end{pmatrix}, \quad \begin{pmatrix} a_j & b_j \\ b_j & a_j \end{pmatrix}$$

直接计算可验证 A_j 正规, 从而分块计算可发现 B 也正规。

此时，由 Q 正交性可知 $A = QBQ^T$ ，进一步计算得

$$AA^T = QBQ^TQB^TQ^T = QBB^TQ^T = QB^TBQ^T = QB^TQ^TQBQ^T = A^TA$$

从而 A 正规，这就证明了等号成立时 A 正规。

反之，从 A 正规可推出 A 能正交相似第 4 题 (f) 的标准形形式，设标准形为 B ，仍然利用正交性即证得 $\text{tr}(AA^T) = \text{tr}(BB^T)$ 。而这里的 B 已经符合上面分析的等号成立情况要求，从而 A 正规时等号成立。

6. 用上标 H 表示共轭转置，我们介绍两种完全不同的方法：

- 酉相似标准形

* 回顾酉相似标准形相关的知识：复正规阵当且仅当可以酉相似对角化，Hermite 阵当且仅当可以酉相似对角化且对角元全为实数，斜 Hermite 阵当且仅当可以酉相似对角化且对角元实部全为 0，酉方阵当且仅当可以酉相似对角化且对角元模长全为 1。

利用酉相似对角化的结论，由于 A 为酉方阵，存在酉方阵 P 使得 $A = P^H DP$ ，其中 D 为对角元模长均为 1 的对角阵。

由于 -1 不为 A 的特征值， $I + A$ 可逆，利用 P 为酉方阵直接计算有

$$\text{i}(I - A)(I + A)^{-1} = \text{i}(I - P^H DP)(I + P^H DP)^{-1} = \text{i}(P^H IP - P^H DP)(P^H IP + P^H DP)^{-1}$$

从而可以展开成

$$\text{i}P^H(I - D)P(P^H(I + D)P)^{-1}$$

由于 D 的对角元为 A 的特征值，其中没有 -1 ，于是 $I + D$ 也可逆，进一步写为

$$\text{i}P^H(I - D)PP^{-1}(I + D)^{-1}(P^H)^{-1} = P^H\text{i}(I - D)(I + D)^{-1}P$$

设 D 的每个对角元为 $d_j = a_j + b_j\text{i}$ ，这里 a_j, b_j 为实数，可发现 $\text{i}(I - D)(I + D)^{-1}$ 是每个对角元为 (这里利用了分子分母同乘共轭的复数除法计算技巧，最后一个等号直接展开分子即得)

$$\text{i}\frac{1 - d_j}{1 + d_j} = \text{i}\frac{1 - a_j - b_j\text{i}}{1 + a_j + b_j\text{i}} = \text{i}\frac{(1 - a_j - b_j\text{i})(1 + a_j - b_j\text{i})}{(1 + a_j)^2 + b_j^2} = \frac{2b_j + (1 - b_j^2 - a_j^2)\text{i}}{(1 + a_j)^2 + b_j^2}$$

的对角阵。由于 D 的对角元模长均为 1， $a_j^2 + b_j^2 = 1$ ，从而 $\text{i}(I - D)(I + D)^{-1}$ 每个对角元均为实数。

由此，上式给出了 $\text{i}(I - A)(I + A)^{-1}$ 的酉相似对角化，且每个对角元均为实数，利用酉相似对角化性质即得其为 Hermite 阵。

- 整体法操作

* 回顾共轭转置的基本性质 (几乎与转置类似)，如 $(A + B)^H = A^H + B^H$ 、 $(\lambda A)^H = \bar{\lambda}A^H$ 、 $(AB)^H = B^H A^H$ 、 A 可逆时 $(A^H)^{-1} = (A^{-1})^H$ 等。

由于 -1 不为 A 的特征值， $I + A$ 可逆。我们需要证明

$$(\text{i}(I - A)(I + A)^{-1})^H = \text{i}(I - A)(I + A)^{-1}$$

左减右并利用共轭转置的性质可知我们需要证明

$$-\text{i}((I + A)^{-1})^H(I - A)^H - \text{i}(I - A)(I + A)^{-1} = O$$

利用求逆与共轭转置可交换顺序，结合 $I^H = I$ 可得要证的式子等价于

$$-\text{i}(I + A^H)^{-1}(I - A^H) - \text{i}(I - A)(I + A)^{-1} = O$$

由于乘可逆阵不影响等价性 (上述利用共轭转置性质的化简过程已经暗含了 $I + A^H$ 可逆性的证明)，我们对上式同时左乘 $I + A^H$ 、同时右乘 $I + A$ ，即可等价地化为

$$-\text{i}(I - A^H)(I + A) - \text{i}(I + A^H)(I - A) = O$$

将其展开化简为

$$-2\text{i}(I - A^H A) = O$$

由于 A 为酉方阵， $A^H A = I$ ，即可得到上式成立。其可等价变形为原命题，即得 $\text{i}(I - A)(I + A)^{-1}$ 的 Hermite 性。

* 与期中类似，利用标准形的思路进行计算是需要熟练的操作。不过，整体法如果能正确应用，往往可以大幅节省时间。在遇到实际问题时，两者都是需要优先尝试的方法，由此每种方法开头的注里提到的重要结论需要熟悉。

7. (a) 利用矩阵表示的线性性，只需证明 $\mathcal{A}(X) = AX$ 在这组基下的矩阵表示为 $I \otimes A$ ， $\mathcal{B}(X) = XB$ 在这组基下的矩阵表示为 $B^T \otimes I$ ，即可得到结论。

- $\mathcal{A}(X) = AX$

设 A 各分量为 a_{ij} 。直接计算可知 AE_{ij} 是将 A 的第 i 列作为第 j 列，其他列是 0 的方阵，也即

$$AE_{ij} = \sum_{k=1}^n a_{ki} E_{kj}$$

进一步考虑基的排列顺序 (可先取 $n = 2$ 由定义从上式写出矩阵表示，再观察一般情况) 可得矩阵表示即 $\text{diag}(A, A, \dots, A)$ ，从而为 $I \otimes A$ 。

- $\mathcal{B}(X) = XB$

设 A 各分量为 b_{ij} 。直接计算可知 $E_{ij}B$ 是将 B 的第 j 行作为第 i 行，其他行是 0 的方阵，也即

$$E_{ij}B = \sum_{k=1}^n b_{jk} E_{ik}$$

进一步考虑基的排列顺序 (可先取 $n = 2$ 由定义从上式写出矩阵表示，再观察一般情况) 可得矩阵表示为

$$\begin{pmatrix} b_{11}I & \cdots & b_{n1}I \\ \vdots & \ddots & \vdots \\ b_{1n}I & \cdots & b_{nn}I \end{pmatrix}$$

从而为 $B^T \otimes I$ 。

- (b) 沿用上一问记号，考虑先行后列顺序的基 $E_{11}, E_{12}, \dots, E_{1n}, \dots, E_{n1}, E_{n2}, \dots, E_{nn}$ ，利用两个求和式可知 $\mathcal{A}(X)、\mathcal{B}(X)$ 在这组基下的矩阵表示分别为 $A \otimes I、I \otimes B^T$ ，从而 C 在这组基下的矩阵表示为

$$A \otimes I - I \otimes B^T$$

由于线性变换不同基下的矩阵表示相似，对任何 n 阶复方阵 $A、B$ 有 $A \otimes I - I \otimes B^T$ 与 $I \otimes A - B^T \otimes I$ 相似。

由于 B^T 可取遍一切复方阵，将 F 看作某个方阵 B 的转置即得对任何 n 阶复方阵 $A、F$ 有 $A \otimes I - I \otimes F$ 与 $I \otimes A - F \otimes I$ 相似。这就是题目要证明的结论。

- (c) 设 $A = P^{-1}DP$ 、 $B = Q^{-1}FQ$ ，且 $D、F$ 为对角阵，直接计算可知 (a) 中的 C 满足 (注意对角阵 F 的转置仍为 F)

$$C = I \otimes (P^{-1}DP) - (Q^T F Q^{-T}) \otimes I$$

利用 $I = Q^T IQ^{-T}$, 利用克罗内克积乘法性质可知

$$I \otimes (P^{-1}DP) = (Q^T IQ^{-T}) \otimes (P^{-1}DP) = (Q^T \otimes P^{-1})(I \otimes D)(Q^{-T} \otimes P)$$

同理对第二项将 I 展开为 $P^{-1}IP$, 类似展开并运用乘法分配律可得

$$C = (Q^T \otimes P^{-1})(I \otimes D - F \otimes I)(Q^{-T} \otimes P)$$

直接计算可知对角阵的克罗内克积仍为对角阵, 从而中间部分为对角阵, 此外再次利用乘法性质得到

$$(Q^T \otimes P^{-1})(Q^{-T} \otimes P) = (Q^T Q^{-T}) \otimes (P^{-1}P) = I \otimes I$$

而计算得单位阵得克罗内克积仍为单位阵, 从而上述 C 的表达式左右部分互逆, 且中间为对角阵, 已经是一个相似对角化。

- (d) 利用复方阵的相似上三角化, 设 $A = P^{-1}SP$ 、 $B^T = Q^{-1}TQ$, 且 S 、 T 为上三角阵, 完全类似
(c) 可计算得

$$C = (Q^{-1} \otimes P^{-1})(I \otimes S - T \otimes I)(Q \otimes P)$$

假设 S 的对角元为 $\lambda_1, \dots, \lambda_n$ 、 T 的对角元为 μ_1, \dots, μ_n , 可发现 $I \otimes S$ 左上角的分块是 S , 而 $T \otimes I$ 左上角的分块是 $\mu_1 I$, 于是两者作差的左上角是对角元为 $\lambda_1 - \mu_1, \dots, \lambda_n - \mu_1$ 的上三角阵, 对其他对角线上的分块同理。此外, 无论是 $I \otimes S$ 还是 $T \otimes I$, 下三角部分的分块都恒为 0, 由此进一步计算可得 $I \otimes S - T \otimes I$ 是所有对角元为

$$\lambda_i - \mu_j, \quad i = 1, \dots, n, \quad j = 1, \dots, n$$

的上三角阵。

与 (c) 相同可验证 $Q^{-1} \otimes P^{-1} = (Q \otimes P)^{-1}$, 从而上述 C 的表达式是其相似上三角化, 由相似不改变特征值、上三角阵特征值为对角元可知 C 的特征值为

$$\lambda_i - \mu_j, \quad i = 1, \dots, n, \quad j = 1, \dots, n$$

同理, $\lambda_1, \dots, \lambda_n$ 是 A 的特征值, μ_1, \dots, μ_n 是 B^T 的特征值, 由 B 与 B^T 相似可知它们也是 B 的特征值。

最后, 有限维线性空间上的线性变换 C 是同构当且仅当它是单同态, 即 $\text{Ker } C = \{O\}$, 而这又等价于 0 不是 C 的特征值 (C 与矩阵表示 C 特征值相同), 也即等价于没有相等的 λ_i 与 μ_j , 得证。

- (e) 一定均可对角化。

设 A 的 Jordan 标准形为 J 、 B 的 Jordan 标准形为 K , 由于 B^T 与 B 相似, B^T 的 Jordan 标准形也为 K , 从而存在可逆阵 P 、 Q 使得

$$A = P^{-1}JP, \quad B^T = Q^{-1}KQ$$

与 (c) 相同得

$$C = (Q^{-1} \otimes P^{-1})(I \otimes J - K \otimes I)(Q \otimes P)$$

于是仍可得到 C 与 $I \otimes J - K \otimes I$ 相似, 且由于 J 、 K 为上三角阵, 这仍然是一个上三角阵, 记为 T 。由于 C 可对角化, C 应可对角化, 于是 T 可对角化。

我们先证明 J 实际上是对角阵, 这就证明了 A 可对角化。

证明:

考虑 J 的某个特征值 λ 与 K 的某个特征值 μ 。

由 (d) 考虑 C 的特征值 $\lambda - \mu$ 。由于 T 是上三角阵，其特征值 $\lambda - \mu$ 的代数重数即为对角元个数。我们假设 J 中特征值 λ 为 c 重、 T 中 $\lambda - \mu$ 为 s 重。利用张量积的定义，只考虑左上角第一个分块，将 T 写成

$$\begin{pmatrix} J - \mu I & X \\ O & Y \end{pmatrix}$$

由于 $J - \mu I$ 对角元中的 $\lambda - \mu$ 个数为 J 中 λ 重数 c , Y 为一个 $(n-1)n$ 阶且 $\lambda - \mu$ 为 $s-c$ 重的上三角阵。记 $N = n(n-1)$ 。

直接计算可知 $\text{rank}((\lambda - \mu)I_{n^2} - T)$ 为

$$\text{rank} \begin{pmatrix} \lambda I - J & -X \\ O & (\lambda - \mu)I_N - Y \end{pmatrix} \geq \text{rank}(\lambda I - J) + \text{rank}((\lambda - \mu)I_N - Y)$$

利用几何重数不超过代数重数与解空间维数定理， $\text{rank}(\lambda I - J)$ 至少为 $n-c$, $\text{rank}((\lambda - \mu)I_N - Y)$ 至少为 $s-c$ 。但由于 C 可对角化，特征值 $\lambda - \mu$ 的代数重数 s 等于几何重数 $\text{rank}((\lambda - \mu)I_{n^2} - T)$ 。因此，上述两个大于等于号必须取等，也即

$$\text{rank}(\lambda I - J) = n - c$$

而这就说明 J 中特征值 c 的几何重数等于代数重数，由于这样的讨论对每个 J 的特征值 λ 都成立，应有 J 可对角化，从而再由其为 Jordan 形可知其为对角阵，得证。

最后，我们用类似 (b) 的对称性的思路说明 B 也可对角化，即得证。由 (b) 中已经计算， C 在另一组基下的矩阵表示为

$$A \otimes I - I \otimes B^T$$

于是， C 可对角化可推出上方矩阵可对角化，将其改写为

$$I \otimes (-B^T) - (-A^T)^T \otimes I$$

这里 $-B^T$ 替换了上方证明中 A 的位置，从而由 A 的 Jordan 标准形 J 为对角阵可知 $-B^T$ 可对角化。设其对角化为

$$-B^T = P_0^{-1} D_0 P_0$$

可发现

$$B = P_0^T (-D_0) P_0^{-T}$$

从而 B 也可对角化，得证。

* 克罗内克积相关问题最重要的是利用其乘法性质进行拆分。

8. (a) 将 α_1, α_2 扩充为 U 的基 $\alpha_1, \dots, \alpha_n$ 、 β_1, β_2 扩充为 V 的基 β_1, \dots, β_m ，利用张量积的基性质可知

$$\alpha_i \otimes \beta_j, \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

构成 $U \otimes V$ 的一组基。若存在 v 与 w ，设 (所有 v_i, w_j 为 \mathbb{K} 中数)

$$v = \sum_{i=1}^n v_i \alpha_i, \quad w = \sum_{j=1}^m w_j \beta_j$$

则利用内积双线性性直接计算得

$$v \otimes w = \sum_{i=1}^n \sum_{j=1}^m v_i w_j \alpha_i \otimes \beta_j$$

利用基下坐标的唯一性，考虑两侧 $\alpha_1 \otimes \beta_1$ 、 $\alpha_1 \otimes \beta_2$ 、 $\alpha_2 \otimes \beta_2$ 前的系数可发现

$$v_1 w_1 = 1, \quad v_1 w_2 = 0, \quad v_2 w_2 = 1$$

但第二式意味着 v_1 、 w_2 中有 0，一三两式意味着这四个数都不是 0，矛盾。

若 α_1, α_2 不再线性无关， α_2 能被 α_1 表出时设 $\alpha_2 = \lambda \alpha_1$ ， $\lambda \in \mathbb{K}$ ，则由内积双线性性可得

$$\alpha_1 \otimes \beta_1 + \alpha_2 \otimes \beta_2 = \alpha_1 \otimes \beta_1 + \lambda(\alpha_1 \otimes \beta_2) = \alpha_1 \otimes \beta_1 + \alpha_1 \otimes (\lambda \beta_2) = \alpha_1 \otimes (\beta_1 + \lambda \beta_2)$$

当 α_1 能被 α_2 表出时同理，从而只要 α_1, α_2 线性相关即存在 v, w 使得等式成立。

- (b) 由于右侧为 $0 \otimes 0$ ，若 α, β 线性无关，利用 (a) 可直接得到不可能成立。由此它们线性相关，由对称性可不妨设 β 能被 α 表出，即 $\beta = \lambda \alpha$ ， $\lambda \in \mathbb{K}$ 。

此时，利用双线性性直接计算可知

$$\alpha \otimes \beta + \beta \otimes \alpha = 2\lambda(\alpha \otimes \alpha)$$

若 $\lambda = 0$ ，则 $\beta = 0$ ，结论已经成立，否则应有 $\alpha \otimes \alpha = 0$ 。若 α 非零，将 α 扩充为 V 的一组基 $\alpha, \alpha_2, \dots, \alpha_n$ ，这组基两两张量积构成 $V \otimes V$ 的一组基，但 $\alpha \otimes \alpha$ 是这组基的一部分，其不可能为 0，矛盾，这就得到了 $\alpha = 0$ 。

- (c) 我们分为三步证明：

- α, β, γ 线性相关

若否，将 α, β, γ 扩充为 V 的一组基 $\eta_1 = \alpha, \eta_2 = \beta, \eta_3 = \gamma, \eta_4, \dots, \eta_n$ 。利用张量积的基性质可知所有

$$\eta_i \otimes \eta_j \otimes \eta_k, \quad i, j, k \in \{1, 2, \dots, n\}$$

构成 $V \otimes V \otimes V$ 的一组基。由此，左侧求和的三个部分是三个不同的基，利用线性无关性不可能为 0，矛盾。

利用对称性可不妨设 $\gamma = a\alpha + b\beta$ ，这里 $a, b \in \mathbb{K}$ 。

- 若 γ 可被 α, β 表出，则 $\gamma = 0$ 或 α, β 线性相关

在上方的假设下，直接展开求和并利用双线性性可得

$$a(\alpha \otimes \alpha \otimes \beta + \alpha \otimes \beta \otimes \alpha + \beta \otimes \alpha \otimes \alpha) + b(\beta \otimes \beta \otimes \alpha + \beta \otimes \alpha \otimes \beta + \alpha \otimes \beta \otimes \beta) = 0$$

若 α, β 线性无关，将它们扩充为 V 的一组基，可发现上方的张量积项是不同的 6 个基，从而利用基的线性无关性，上式成立当且仅当 $a = b = 0$ 。于是 $\gamma = 0$ 。

由此即证明了 $\gamma = 0$ 或 α, β 线性相关。由于 $\gamma = 0$ 时已经得证，我们只需讨论最后一种情况。

- α, β, γ 存在零向量

若 γ 能被 α, β 表出，且 α, β 线性相关。先假设 β 能被 α 表出，此时 γ 也能被 α 表出，从而可设 $\beta = x\alpha$ 、 $\gamma = y\alpha$ ， $x, y \in \mathbb{K}$ 。直接展开计算可知原式化为

$$3xy\alpha \otimes \alpha \otimes \alpha = 0$$

与 (b) 的最后一步讨论类似，可知 $xy = 0$ 或 $\alpha = 0$ 必然成立，而 $xy = 0$ 也能推出 β 或 γ 为零向量，这就得到了证明。

若 α 能被 β 表出，情况完全类似。综合以上得结论。

- (d) 我们证明结论的正确性。仍然进行讨论：

- α, β, γ 线性相关

若否，将 α, β, γ 扩充为 V 的一组基 $\eta_1 = \alpha, \eta_2 = \beta, \eta_3 = \gamma, \eta_4, \dots, \eta_n$ 。利用张量积的基性质可知所有

$$\eta_i \otimes \eta_j \otimes \eta_k, \quad i, j, k \in \{1, 2, \dots, n\}$$

构成 $V \otimes V \otimes V$ 的一组基。

设 (所有 x_i, y_i, z_i 为 \mathbb{K} 中数)

$$x = \sum_{i=1}^n x_i \eta_i, \quad y = \sum_{i=1}^n y_i \eta_i, \quad z = \sum_{i=1}^n z_i \eta_i$$

则利用内积双线性直接计算得

$$x \otimes y \otimes z = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n x_i y_j z_k \eta_i \otimes \eta_j \otimes \eta_k$$

利用基下坐标的唯一性，考虑两侧 $\alpha \otimes \alpha \otimes \alpha, \alpha \otimes \beta \otimes \gamma, \gamma \otimes \alpha \otimes \beta, \beta \otimes \gamma \otimes \alpha$ 前的系数可知

$$x_1 y_1 z_1 = 0, \quad x_1 y_2 z_3 = 1, \quad x_3 y_1 z_2 = 1, \quad x_2 y_3 z_1 = 1$$

但第一式意味着 x_1, y_1, z_1 中有 0，后三个式子意味着这九个数都不是 0，矛盾。

利用对称性可不妨设 $\gamma = a\alpha + b\beta$ ，这里 $a, b \in \mathbb{K}$ 。

- 若 γ 可被 α, β 表出，则 α, β 线性相关

与之前完全类似，将左侧展开为

$$a(\alpha \otimes \alpha \otimes \beta + \alpha \otimes \beta \otimes \alpha + \beta \otimes \alpha \otimes \alpha) + b(\beta \otimes \beta \otimes \alpha + \beta \otimes \alpha \otimes \beta + \alpha \otimes \beta \otimes \beta)$$

若 α, β 线性无关，将它们扩充 V 的一组基 $\eta_1 = \alpha, \eta_2 = \beta, \eta_3, \dots, \eta_n$ 。仍设

$$x = \sum_{i=1}^n x_i \eta_i, \quad y = \sum_{i=1}^n y_i \eta_i, \quad z = \sum_{i=1}^n z_i \eta_i$$

对比

$$x \otimes y \otimes z = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n x_i y_j z_k \eta_i \otimes \eta_j \otimes \eta_k$$

在张量积空间的基下坐标，考虑 i, j, k 取 1, 2 的情况，可得

$$x_1 y_1 z_1 = x_2 y_2 z_2 = 0$$

$$x_1 y_1 z_2 = x_1 y_2 z_1 = x_2 y_1 z_1 = a$$

$$x_2 y_2 z_1 = x_2 y_1 z_2 = x_1 y_2 z_2 = b$$

从第一个式子中可看出 x_1, y_1, z_1 里有 0，而无论哪个是 0，第二个式子左侧都至少有一项是 0，从而 $a = 0$ ，同理 $b = 0$ ，由此即得到 $\gamma = 0$ ，但这与条件中的非零矛盾，从而只能 α, β 线性相关。

- 综合

由第一部分证明，利用线性无关等价定义可知三个向量中有一个能被其他两个表出。与第二部分证明相同，无论三个向量哪个能被其他两个表出，都能推出剩下两个向量线性相关，从而 $\dim \langle \alpha, \beta, \gamma \rangle = 1$ ，这就说明了它们两两线性相关。

* 对于张量积问题，核心操作是利用其基性质进行处理（尤其注意不可交换性， $v_i \otimes v_j$ 与 $v_j \otimes v_i$ 在 v_i, v_j 是一组基中的两个时代表张量积空间不同的基），对于线性无关向量组可扩充为一组基，对于未知向量组则先讨论其是否线性无关。