高等代数：从入门到精通（下册）

胡　勇
huy@sustech.edu.cn

2022 年 6 月 21 日

目录

关于记号和术语的约定

本书将一直采用如下约定的记号和术语.

(0.1) 关于集合

1. 若 A, B 都是集合 X 的子集, $A \setminus B$ 表示二者的差集, 即, $A \setminus B = \{x \in X \,|\, x \in A \text{ 但 } x \notin B\}$.
2. 自然数集记为$\mathbb{N}$, 约定其中包含0.
   非零自然数构成的集合记为$\mathbb{N}^* = \mathbb{N} \setminus \{\,0\,\} = \{\,1, 2, \ldots\,\}$.
3. 用 $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ 和 $\mathbb{C}$ 分别表示整数集, 有理数集, 实数集和复数集.
4. 记 $\mathbb{R}^* = \mathbb{R} \setminus \{\,0\,\}$. 类似地, $\mathbb{Q}^* = \mathbb{Q} \setminus \{\,0\,\}$; $\mathbb{C}^* = \mathbb{C} \setminus \{\,0\,\}$.
5. 记 $\mathbb{R}^*_+ = \mathbb{R}_{>0} := \{\,x \in \mathbb{R} \,|\, x > 0\,\}$. 类似地, $\mathbb{Q}^*_+ = \mathbb{Q}_{>0} := \{\,x \in \mathbb{Q} \,|\, x > 0\,\}$.
6. 若 $a \le b$ 均为实数, 记:
   $]a,\, b[\, = \{x \in \mathbb{R} \,|\, a < x < b\}$;
   (数学分析或高等数学教材中的开区间常记为 $(a,\, b)$, 但为了区别于平面中点的坐标之记号, 本书改用上面这种开区间记号. 这是法语数学文献中常用的记号. 以下半开半闭区间和闭区间的记号也是类似定义.)
   $[a,\, b[\, = \{x \in \mathbb{R} \,|\, a \le x < b\}$;
   $]a,\, b] = \{x \in \mathbb{R} \,|\, a < x \le b\}$;
   $[a,\, b] = \{x \in \mathbb{R} \,|\, a \le x \le b\}$;
   另外, 我们为整数构成的区间引入一个特别的记号如下:
   $[\![a,\, b]\!] = \{k \in \mathbb{Z} \,|\, a \le k \le b\}$.
   例如 $[\![1,\, 5]\!] = \{1,\, 2,\, 3,\, 4,\, 5\}$.
7. 一个集合 S 的基数 (对于有限集而言即为元素个数) 记为 $\#S$ 或 $\mathrm{Card}(S)$. 在不会与其他记号 (如行列式、绝对值等) 混淆时也常记为 $|S|$.

(0.2) 关于矩阵

以下 m, n 表示正整数, K 表示复数域 $\mathbb{C}$ 中的任一子域 (或者更一般地, K 可以是任意一个环).

1. $\mathbf{M}_{m\times n}(K)$: 矩阵系数 (即矩阵中列出的元素) 取自于集合 K 的所有 $m \times n$ 矩阵全体.
2. $\mathbf{M}_n(K)$: 矩阵系数取自于 K 的所有 $n \times n$ 矩阵全体.
3. $\mathbf{GL}_n(K)$: 在 $\mathbf{M}_n(K)$ 中所有可逆的矩阵全体.

4. $\mathbf{SL}_n(K)$: 在 $\mathbf{M}_n(K)$ 中所有行列式为 1 的矩阵全体.
5. $\mathbf{O}_n(K)$: 在 $\mathbf{M}_n(K)$ 中的所有正交矩阵全体.
6. $\mathbf{SO}_n(K)$: 在 $\mathbf{M}_n(K)$ 中的所有特殊正交矩阵 (行列式为 1 的正交矩阵) 全体.
7. $\mathbf{Sp}_n(K)$: 在 $\mathbf{M}_{2n}(K)$ 中的所有辛矩阵全体.
8. $\mathbf{U}_n(\mathbb{C})$: 在 $\mathbf{M}_n(\mathbb{C})$ 中的所有酉矩阵全体.
9. $\mathbf{SU}_n(\mathbb{C})$: 在 $\mathbf{M}_n(\mathbb{C})$ 中的所有特殊酉矩阵 (行列式为 1 的酉矩阵) 全体.

(0.3) 其他

1. $A := B$ 意思是 A 的定义为 B.
2. $A =: B$ 意思是将 A 记为 B.

更多内容以后适时补充.

知乎 首页 发现 等你来答

法律 思考 法律常识 违法

在中国生活，有哪些行为可能涉嫌违法却鲜为人知？

写回答 查看全部 229 个回答 ›

Zhaoyang
面向梯度编程
＋ 关注

大学生学习不刻苦是违法的。

——《中华人民共和国高等教育法》第五十三条第一款规定，高等学校的学生应当刻苦学习。

发布于 昨天 05:26

中华人民共和国教育部
Ministry of Education of the People's Republic of China

热词：2019年教育1+1系列发布采访

当前位置：首页 › 教育部司局机构 › 政策法规司

中华人民共和国高等教育法

（1998年8月29日第九届全国人民代表大会常务委员会第四次会议通过　根据2015年12月27日第十二届全国人民代表大会常务委员会第十八次会议《关于修改〈中华人民共和国高等教育法〉的决定》修正）

目录

第一章 总则

第二章 高等教育基本制度

............................省略内容..

第六章 高等学校的学生

第五十三条 高等学校的学生应当遵守法律、法规，遵守学生行为规范和学校的各项管理制度，尊敬师长，刻苦学习，增强体质，树立爱国主义、集体主义和社会主义思想，努力学习马克思列宁主义、毛泽东思想、邓小平理论，具有良好的思想品德，掌握较高的科学文化知识和专业技能。

第五章　Jordan 标准形

我们假定读者已经学习过本书上册的内容, 掌握了线性方程组和矩阵的基本理论, 并熟悉抽象向量空间及其子空间的概念. 为方便读者, 我们会在 §5.0 节回顾一些主要的概念和结论.

本章中我们以 K 表示复数域 $\mathbb{C}$ 的一个子域*. 若非有明确的相反说明, 向量空间总是指 K 上的向量空间. 字母 n 用来表示一个正整数.

5.0　前情回顾

本节的主要目的是复习回顾本书上册中的一些主要内容, 为后续内容的展开做好准备.

5.0.1　向量空间与线性映射

(5.0.1) 所谓一个 K 上的*向量空间* 或者 K-*向量空间*†是指一个附带有两种运算的集合 V, 其中一种运算在 V 的元素之间两两进行, 通常称为*加法*, 它可以视为 $V\times V$ 到 V 的一个映射, 记为

$$V\times V\longrightarrow V\,;\quad (u,\,v)\longmapsto u+v\,.$$

而另一种运算在 K 的元素和 V 的元素之间进行, 通常称为*数乘*, 它可以视为 $K\times V$ 到 V 的一个映射, 记作

$$K\times V\longrightarrow V\,;\quad (\lambda\,,\,v)\longmapsto \lambda\cdot v\,.$$

(我们常将 $\lambda\cdot v$ 简记为 λv.)

在定义中, 向量空间 V 上的加法和数乘需要满足 8 条公理. 这些公理有许多推论. 例如, 这些公理保证了 V 中加法恒等元的存在性和唯一性, 该元素通常记为 $\mathbf{0}$, 或者直接写成 0 (普通印刷体而非粗体的样子), 称为 V 中的*零元*或*零向量*.

当 V 是 K-向量空间时, 我们可以将 V 中的元素称为*向量*; 而相对于 V 而言, K 的元素可被称为*常数*、*纯量*或*标量*. 如果一个常数 $\lambda\in K$ 和一个向量 $v\in V$ 的数乘结果 λv 是 $\mathbf{0}$, 则要么 v 等于零向量 $\mathbf{0}$, 要么常数 λ 等于 0.

设 U 是 K-向量空间 V 的子集. 如果它满足以下三个条件

- $\mathbf{0}\in U$;
- U 关于加法封闭, 即, 对于任意 $u,\,v\in U$, 均有 $u+v\in U$;
- U 关于数乘封闭, 即, 对于任意 $u\in U$ 和任意 $\lambda\in K$, 均有 $\lambda u\in U$;

*很多中文教材称 K 为一个**数域**. 这种说法可能是起源于 [10] (英文版 [11]). 但读者需要留意, 在现代的代数数论文献中“数域”一词通常是指更为特殊的一类域.

†有很多书也用“线性空间”作为“向量空间”的同义词.

那么我们说 U 是 V 的一个*子空间*. 根据以上封闭性的假设, 我们可以把向量空间 V 上的加法和数乘运算限制于 U 的元素, 使得运算结果始终落在 U 之内. 通过这种操作使 U 获得加法和数乘运算, 而且关于这样的加法和数乘 U 也成为一个向量空间. 我们常说子空间 U 从 V 那里自然地*继承*了加法和数乘运算, 或者说 V 的加法和数乘运算自然地*诱导*出子空间 U 上的加法和数乘运算.

最小的向量空间仅由零向量构成. 我们称之为*零向量空间*, 并记之为 0. 对于任意向量空间而言, 零向量空间都可以视作它的一个子空间. ■

(5.0.2) 我们谈论“向量组”时默认总是指由有限多个 (可以是零个) 向量构成的组. (当需要讨论无穷多向量的情况时, 我们换用“向量族”这个词.)

设 $v_1, \cdots, v_n$ 是 K-向量空间 V 中的向量组. 它们的一个 *K-线性组合*是指可以写成

$$v = a_1v_1 + \cdots + a_nv_n \quad \text{其中每个 } a_i \in K$$

这种形式的一个向量 $v \in V$. 以一个向量组 S 的所有线性组合为元素形成的集合记为 $\mathrm{span}(S)$, 它是 V 的一个子空间, 称为向量组 S *张成*或*生成*的子空间. 空向量组生成的子空间被认为是零向量空间.

如果向量组 $v_1, \cdots, v_n$ 的线性组合 $a_1v_1 + \cdots + a_nv_n$ 只有在每个系数 a_i 都等于 0 时才会等于零向量, 我们说该向量组是*线性无关*的, 否则称该向量组是*线性相关*的. 空向量组被认为是*线性无关*的.

如果存在有限多个向量 $v_1, \cdots, v_n \in V$ 使 $V = \mathrm{span}(v_1, \cdots, v_n)$, 则称 V 是*有限生成*的, 而此时 $v_1, \cdots, v_n$ 称为 V 的一个*生成组*或*张成组*. 如果 V 的一个生成组 $v_1, \cdots, v_n$ 是线性无关的, 则称 $v_1, \cdots, v_n$ 是 V 的一组*基*.

当一个 K-向量空间 V 以某一 (有限多向量构成的) 向量组为一组基时, 我们说 V 是*有限维*的. 可以证明此时 V 的每一组基所含的向量数目都相同, 这个数目被定义为 V 的*维数*, 记为 $\dim V$ 或者更准确地, $\dim_K V$. 如果 V 不是有限维的, 我们说 V 是*无限维*的, 并记 $\dim V = +\infty$. 可以证明: 一个向量空间是有限维的当且仅当它是有限生成的. ■

(5.0.3) 现在来回忆向量空间的一些重要例子.

对于任意正整数 m, n, 由系数取自 K 的所有 $m \times n$ 矩阵构成的集合 $\mathbf{M}_{m\times n}(K)$ 按照通常的矩阵加法和数乘运算构成一个 K-向量空间. 这是个有限维向量空间, 其维数等于 mn. 特别地, 所有 n 阶方阵构成的集合 $\mathbf{M}_n(K)$ 是一个 n^2 维的 K-向量空间. 向量空间 $K^{n\times 1}$ 和 $K^{1\times n}$ 分别是我们通常的 n 维列向量空间和行向量空间. 在不必要对二者进行区分时, 我们允许以 K^n 表示 $K^{n\times 1}$ 或 $K^{1\times n}$.

任取 K-向量空间 V 和非空集合 A, 令 $U = \mathrm{Map}(A, V)$ 为所有从 A 到 V 的映射构成的集合. 对于任意 $f, g \in U$, 通过逐点取值相加的方式可以定义加法 $f + g$ 的运算结果为如下映射

$$A \longrightarrow V\,; \quad x \longmapsto f(x) + g(x)\,.$$

类似地, 对于 $\lambda \in K$ 和 $f \in U$, 数乘 λf 的运算结果可以定义为如下映射

$$A \longrightarrow V\,; \quad x \longmapsto \lambda f(x)\,.$$

通过以上加法和数乘运算我们可以将 $U = \mathrm{Map}(A, V)$ 视作 K-向量空间. 特别地, 对于 $V = K$ 的情况我们得到一个向量空间 $\mathrm{Map}(A, K)$.

以 X 表示一个形式符号. 一个*以 X 为未定元、系数取自 K 的多项式 (形式)* 是指一个形式表达式

$$f(X) := \sum_{i=0}^{+\infty} a_iX^i \quad \text{其中每个 } a_i \in K \text{ 且满足 } a_i \neq 0 \text{ 的指标 } i \text{ 只有有限多个.}$$

我们常以 f 简记之. 如果每个 a_i 都等于 0, 相应的多项式 f 被称为*零多项式*, 通常记为 $f=0$. 对于非零多项式我们可以按通常熟知的方式定义多项式 f 的*次数*和*首项系数*. 首项系数为 1 的多项式称为*首一多项式*. 零多项式的次数规定为 $-\infty$, 首项系数则规定为 0.

我们用 $K[X]$ 表示所有以 X 为未定元且系数在 K 中的多项式构成的集合. 对于任意 $m\in\mathbb{N}$, 用 $K[X]_{\le m}$ 表示 $K[X]$ 中次数不超过 m 的那些多项式构成的子集. 读者应当熟悉通常的多项式加法、乘法和数乘运算. 不难看出, (关于通常的加法和数乘运算) $K[X]$ 是个 K-向量空间, 它是无限维的. 而对于每个 $m\in\mathbb{N}$, $K[X]_{\le m}$ 是 $K[X]$ 的子空间, 它是有限维的, 且 $\dim_K K[X]_{\le m}=m+1$. ■

(5.0.4) 设 F 和 G 是向量空间 V 中的非空子集. 我们可以定义新的集合 $F+G$ 为

$$F+G:=\{x+y\,|\,x\in F\,,\,y\in G\}\,.$$

更一般地, 若 $F_1,\cdots,F_r$ 都是 V 的非空子集, 则可定义

$$F_1+\cdots+F_r:=\{x_1+\cdots+x_r\,|\ \text{对每个 } i\in[\![1,\,r]\!]\,,\ x_i\in F_i\}\,.$$

设 U 和 W 都是 V 的子空间, 则 $U+W$ 也是 V 的子空间, 称为子空间 U 与 W 的*和*. 按照定义, $U+W$ 中的每个元素 v 都可以写成如下形式:

$$v=u+w\,,\ \ \text{其中 } u\in U\,,\,w\in W\,.$$

如果对于每个 $v\in U+W$ 来说, 以上表达式都是唯一的, 则称 $U+W$ 是个*直和*. 此时可以记之为 $U\oplus W$. 能够保证 $U+W$ 是直和的一个充分必要条件是 $U\cap W=0$. 如果已知 U 和 W 都是有限维的, 那么有维数公式

$$\dim(U+W)=\dim U+\dim W-\dim(U\cap W)\,.$$

因此, 在有限维的假设下, $U+W$ 是直和的另一个充分必要条件是 $\dim(U+W)=\dim U+\dim W$.

更一般地, 我们可以考虑 V 的多个子空间 $U_1,\cdots,U_r$. 和空间 $U_1+\cdots+U_r$ 中的元素 v 总是具有 $u_1+\cdots+u_r$ 的形式, 其中每个 $u_i\in U_i$. 如果这样的表达式对于任意选定的 $v\in U_1+\cdots+U_r$ 来说都是唯一确定的, 我们就说 $U_1+\cdots+U_r$ 是直和. 在每个 U_i 都是有限维的情况下,

$$U_1+\cdots+U_r\ \text{是直和}\quad\Longleftrightarrow\quad \dim(U_1+\cdots+U_r)=\sum_{i=1}^{r}\dim U_i\,.$$

需要注意的是, 当 $r\ge 3$ 时, U_i 之间两两相交为 0 是不足以保证 $U_1+\cdots+U_r$ 为直和的. ■

(5.0.5) 设 U 和 V 是 K-向量空间. 如果一个映射 $f:U\to V$ 满足以下两个条件:

- f 保持向量加法, 即, 对任意 $u,\,w\in U$ 均有 $f(u+w)=f(u)+f(w)$;
- f 保持向量数乘, 即, 对任意 $\lambda\in K$ 及任意 $u\in U$ 均有 $f(\lambda u)=\lambda f(u)$;

那么我们说 f 是个 *K-线性映射*, 或简称线性映射. 将从 U 到 V 的所有线性映射构成的集合记为 $\mathrm{Hom}(U,\,V)$, 或者更确切地, $\mathrm{Hom}_K(U,\,V)$. 不难看出, $\mathrm{Hom}(U,\,V)$ 是向量空间 $\mathrm{Map}(U,\,V)$ 的一个子空间.

对于 $U=V$ 的情况, 我们记 $\mathrm{End}(V)=\mathrm{Hom}(V,\,V)$. 向量空间 $\mathrm{End}(V)$ 中的元素也称为 V 上的*线性变换*或*线性算子*. 我们常将 V 上的恒等变换 $v\mapsto v$ 记为 Id_V 或者更简单地记为 I.

如果一个线性映射 $f:U\to V$ 是一个双射 (亦即可逆映射), 我们称 f 是从 U 到 V 的一个*同构映射*, 或简称*同构*. 若 f 是个同构, 则 f 的逆映射 $f^{-1}:V\to U$ 也是线性映射, 故而也是同构. 若已知 U 和 V 都是有限维的, 那么对任何线性映射 $f:U\to V$ 来说下列三个条件等价:

- f 是同构;
- f 是单射, 并且 $\dim U = \dim V$;
- f 是满射, 并且 $\dim U = \dim V$.

若 $f : U \to V$ 是向量空间之间的线性映射, 则

$$\mathrm{Ker}(f) := \{u \in U \mid f(u) = 0\} \text{ 和 } \mathrm{Im}(f) := f(U) = \{f(u) \mid u \in U\}$$

分别称为 f 的核与像. 有的书也将 $\mathrm{Ker}(f)$ 称为 f 的*零化空间*, 记为 $\mathrm{null}(f)$; 而 $\mathrm{Im}(f)$ 也被称为 f 的值域, 记为 $\mathrm{range}(f)$. 容易看出, $\mathrm{Ker}(f)$ 和 $\mathrm{Im}(f)$ 分别是 U 和 V 的子空间.

对于线性映射 $f : U \to V$, 可以用 $\mathrm{Ker}(f) = 0$ 作为 f 是单射的充分必要条件, 而 $\mathrm{Im}(f) = V$ 则是 f 为满射的符号描述.

有时我们将 $\mathrm{Ker}(f)$ 的维数称为 f 的*零化度*, $\mathrm{Im}(f)$ 的维数则称为 f 的*秩*, 记为 $\mathrm{rank}(f)$. 以下结论是常用的*秩–零化度定理*: 若 $f : U \to V$ 是有限维向量空间之间的线性映射, 则

$$\dim \mathrm{Ker}(f) + \mathrm{rank}(f) = \dim U .$$

任给矩阵 $A \in \mathbf{M}_{m\times n}(K)$, 通过矩阵乘法可以定义出线性映射

$$f_A \,:\, K^{n\times 1} \longrightarrow K^{m\times 1} \,; \quad X \longmapsto AX .$$

根据相关定义可知, $\mathrm{Ker}(f_A)$ 就是矩阵 A 的零化空间 $\mathcal{N}(A)$, 亦即齐次线性方程组 $AX = 0$ 的解空间; 而 $\mathrm{Im}(f_A)$ 则等于矩阵 A 的列空间 $\mathscr{C}(A)$. 于是, $\mathrm{rank}(f_A) = \dim \mathrm{Im}(f_A)$ 也就是矩阵 A 的秩. 以像空间维数的观点来理解矩阵的秩, 这在研究与矩阵秩相关的问题时可能会提供一些新颖的思路. ■

5.0.2 矩阵表示与基变换

(5.0.6) 设 V 是 n 维 K-向量空间. 当 V 的一组基 $v_1, \cdots, v_n$ 被当作一个有序排列的向量组时, 我们称之为 V 的一组*有序基*, 并且换用带括号的记号 $(v_1, \cdots, v_n)$ 来强调这种观念. 例如, 我们熟知的列向量空间 K^n 有一组标准基为

$$e_1 = (1, 0, \cdots, 0)^T, \cdots, e_n = (0, \cdots, 0, 1)^T .$$

以 $n = 3$ 的情况为例. 当我们把标准基中的向量按不同方式排序时, 可以得到 K^3 的 6 组不同的有序基, 即

$$(e_1, e_2, e_3), \ (e_1, e_3, e_2), \ (e_2, e_1, e_3) \quad \text{等}.$$

不过, 为了术语的简洁, 当我们需要以有序基的观点来描述 K^n 的标准基时, 通常将 $(e_1, \cdots, e_n)$ 这组有序基作为 K^n 的*有序标准基*, 在不至于引起歧义时也简称 $(e_1, \cdots, e_n)$ 为 K^n 的标准基.

若已取定向量空间 V 的一组有序基 $\mathcal{B} = (v_1, \cdots, v_n)$, 则 V 中任意向量 v 可以唯一地表示为

$$v = a_1v_1 + \cdots + a_nv_n = (v_1, \cdots, v_n) \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

的形式, 其中列向量 $(a_1, \cdots, a_n)^T$ 属于 K^n, 它依赖于向量 v 以及有序基 $\mathcal{B}$ 的选取, 我们记之为 $\mathcal{M}_{\mathcal{B}}(v)$, 并称之为向量 v 在有序基 $\mathcal{B}$ 下的*坐标向量*或简称*坐标*. 映射

$$\mathcal{M}_{\mathcal{B}} \,:\, V \longrightarrow K^n \,; \quad v \longmapsto \mathcal{M}_{\mathcal{B}}(v)$$

是向量空间 V 到 K^n 的一个同构映射.

如果 $\mathcal{B}'=(v_1',\cdots,v_n')$ 是 V 的另一组有序基, 则存在唯一的可逆矩阵 $P=P_{\mathcal{B},\mathcal{B}'}\in\mathbf{M}_n(K)$ 使得

$$\mathcal{B}'=\mathcal{B}\cdot P_{\mathcal{B},\mathcal{B}'}\,,\qquad \text{也即是}\ \ (v_1',\cdots,v_n')=(v_1,\cdots,v_n)\cdot P\,.$$

具体来说, 如果 $\begin{pmatrix}a_{1j}\\ \vdots\\ a_{nj}\end{pmatrix}$ 表示矩阵 $P=P_{\mathcal{B},\mathcal{B}'}$ 的第 j 列, 那么 $v_j'=(v_1,\cdots,v_n)\cdot\begin{pmatrix}a_{1j}\\ \vdots\\ a_{nj}\end{pmatrix}$. 换言之, $P=P_{\mathcal{B},\mathcal{B}'}$ 的第 j 列就是新的有序基 $\mathcal{B}'$ 中第 j 个向量 v_j' 在原来的有序基 $\mathcal{B}$ 下的坐标向量. 我们将矩阵 $P_{\mathcal{B},\mathcal{B}'}$ 称为有序基 $\mathcal{B}$ 到 $\mathcal{B}'$ 的*过渡矩阵*或*基变换矩阵*.

对于任意 $v\in V$, 我们有如下坐标变换公式:

$$\mathcal{M}_{\mathcal{B}}(v)=P_{\mathcal{B},\mathcal{B}'}\cdot\mathcal{M}_{\mathcal{B}'}(v)\,.$$

如果以交换图表的语言来描述, 以上公式的含义是如下图表可交换:

$$\begin{tikzcd} & V \arrow[dl, "\mathcal{M}_{\mathcal{B}'}"'] \arrow[dr, "\mathcal{M}_{\mathcal{B}}"] & \\ K^n \arrow[rr, "P_{\mathcal{B},\mathcal{B}'}\,\bullet"'] & & K^n \end{tikzcd}$$

这里, "$P_{\mathcal{B},\mathcal{B}'}\,\bullet$" 表示左乘矩阵 $P_{\mathcal{B},\mathcal{B}'}$ 给出的线性映射 $X\mapsto P_{\mathcal{B},\mathcal{B}'}X$. ■

(5.0.7) 现在设 V 和 W 是两个有限维 K-向量空间. 设 $\mathcal{B}=(v_1,\cdots,v_n)$ 和 $\mathcal{C}=(\eta_1,\cdots,\eta_m)$ 分别为 V 和 W 的有序基. 设 $T:V\to W$ 为线性映射. 对任意 $v\in V$, 我们常把 $T(v)$ 简记为 Tv.

对于 $\mathcal{B}$ 中的每个向量 v_j, 经过映射 T 作用之后得到 W 中的向量 Tv_j. 该向量可以在有序基 $\mathcal{C}$ 下写出坐标:

$$Tv_j=(\eta_1,\cdots,\eta_m)\cdot\begin{pmatrix}a_{1j}\\ \vdots\\ a_{mj}\end{pmatrix}\,.$$

我们将矩阵

$$\mathcal{M}_{\mathcal{B},\mathcal{C}}(T):=\Big(a_{ij}\Big)_{\substack{1\le i\le m\\ 1\le j\le n}}$$

称为线性映射 T 在有序基 $\mathcal{B}$ 和 $\mathcal{C}$ 下的*矩阵*或*矩阵表示*. 这个矩阵的第 j 列是 $\mathcal{B}$ 中第 j 个向量 v_j 经过 T 映射所得向量 Tv_j 在有序基 $\mathcal{C}$ 下的坐标列.

对于 V 中一般的向量 v (不一定是 $\mathcal{B}$ 中的向量), 将 v 在 $\mathcal{B}$ 下的坐标列左乘 T 的矩阵就可以得到 Tv 在 $\mathcal{C}$ 下的坐标列. 换成抽象但简洁的符号表述即为

$$\mathcal{M}_{\mathcal{C}}(Tv)=\mathcal{M}_{\mathcal{B},\mathcal{C}}(T)\cdot\mathcal{M}_{\mathcal{B}}(v)\,.$$

我们也可以用交换图表

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\mathcal{M}_{\mathcal{B}}"'] & W \arrow[d, "\mathcal{M}_{\mathcal{C}}"] \\ K^n \arrow[r, "\mathcal{M}_{\mathcal{B},\mathcal{C}}(T)\bullet"'] & K^m \end{tikzcd}$$

来表述上面的公式. 其中, "$\mathcal{M}_{\mathcal{B},\mathcal{C}}(T)\bullet$" 表示左乘矩阵 $\mathcal{M}_{\mathcal{B},\mathcal{C}}(T)$ 给出的线性映射 $X\mapsto\mathcal{M}_{\mathcal{B},\mathcal{C}}(T)X$.

以上我们看到, 取定了空间 V 和 W 的有序基 $\mathcal{B}$ 和 $\mathcal{C}$ 之后, 可以将每个线性映射 $T \in \mathrm{Hom}(V, W)$ 唯一确定地对应到一个矩阵 $\mathcal{M}_{\mathcal{B},\mathcal{C}}(T) \in \mathbf{M}_{m\times n}(K)$. 于是我们得到一个映射

$$\mathcal{M}_{\mathcal{B},\mathcal{C}} \;:\; \mathrm{Hom}(V, W) \longrightarrow \mathbf{M}_{m\times n}(K)\,.$$

这个映射实际上是向量空间 $\mathrm{Hom}(V, W)$ 和 $\mathbf{M}_{m\times n}(K)$ 之间的一个同构映射. ■

(5.0.8) 再来考虑有限维向量空间 V 到自身的线性映射, 即 V 上的线性变换. 此时我们可以只取 V 的一组有序基 $\mathcal{B}$, 将每个线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 对应到矩阵 $\mathcal{M}_{\mathcal{B},\mathcal{B}}(\mathscr{A})$. 我们将记号 $\mathcal{M}_{\mathcal{B},\mathcal{B}}(\mathscr{A})$ 简化为 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$, 并称之为线性变换 $\mathscr{A}$ 在有序基 $\mathcal{B}$ 下的矩阵 (表示).

若 $f(X) = a_rX^r + \cdots + a_1X + a_0$ 是 $K[X]$ 中一个多项式, 我们可以定义

$$f(\mathscr{A}) := a_r\mathscr{A}^r + \cdots + a_1\mathscr{A} + a_0I$$

其中 I 表示 V 上的恒等变换. 通常我们认为 $\mathscr{A}^0 = I$.

类似地, 若 $A \in \mathbf{M}_n(K)$ 是方阵, 可以定义

$$f(A) = a_rA^r + \cdots + a_1A + a_0I$$

这里 I 表示 n 阶单位矩阵. 我们约定 $A^0 = I$.

如果 $A = \mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 在 V 的有序基 $\mathcal{B}$ 下的矩阵, 那么在同一组有序基下 $f(\mathscr{A})$ 的矩阵就是 $f(A)$. ■

注记 5.0.9. 若 $\mathcal{B}$ 和 $\mathcal{C}$ 是同一向量空间 V 的两组有序基, $\mathscr{A} \in \mathrm{End}(V) = \mathrm{Hom}(V, V)$, 读者需要留意

$$\mathcal{M}_{\mathcal{B}}(\mathscr{A})\,,\;\; \mathcal{M}_{\mathcal{C}}(\mathscr{A}) \quad \text{和} \quad \mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$$

这些记号的含义差别. ■

(5.0.10) 设 $\mathcal{B}$ 和 $\mathcal{B}'$ 是 n 维 K-向量空间 V 的两组 (可能相同的) 有序基, $\mathscr{A} \in \mathrm{End}(V)$. 设 $P = P_{\mathcal{B},\mathcal{B}'}$ 是有序基 $\mathcal{B}$ 到 $\mathcal{B}'$ 的过渡矩阵 (故 $\mathcal{B}' = \mathcal{B}\cdot P$). 则矩阵 $A := \mathcal{M}_{\mathcal{B}}(\mathscr{A})$, $A' := \mathcal{M}_{\mathcal{B}'}(\mathscr{A})$ 和 $P = P_{\mathcal{B},\mathcal{B}'}$ 之间满足如下关系:

$$A' = P^{-1}AP\,, \quad \text{即}\,, \quad \mathcal{M}_{\mathcal{B}'}(\mathscr{A}) = P_{\mathcal{B},\mathcal{B}'}^{-1}\mathcal{M}_{\mathcal{B}}(\mathscr{A})P_{\mathcal{B},\mathcal{B}'}\,.$$

这表明: 同一线性变换在不同基下的矩阵满足*相似*的关系. (我们说两个方阵 $M, N \in \mathbf{M}_n(K)$ 在 K 上相似, 是指存在可逆矩阵 $Q \in \mathbf{M}_n(K)$ 使得 $M = Q^{-1}NQ$.) ■

(5.0.11) 设 V 为有限维 K-向量空间. 我们称线性变换 $\mathscr{A} \in \mathrm{End}(V)$ *可对角化* 是指存在 V 的一组有序基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是对角阵. 类似地, $\mathscr{A}$ *可上三角化*是指存在有序基 $\mathcal{B}$ 使得 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是上三角阵.

不难证明: 以下各论断是等价的:

- $\mathscr{A}$ 可对角化;
- 对于 V 的任意有序基 $\mathcal{C}$, 矩阵 $\mathcal{M}_{\mathcal{C}}(\mathscr{A})$ 在 K 上相似于一个对角阵;
- 存在 V 的一组有序基 $\mathcal{C}$, 矩阵 $\mathcal{M}_{\mathcal{C}}(\mathscr{A})$ 在 K 上相似于一个对角阵.

关于上三角化也有类似地结论.

如果 A 是 K 上的 n 阶方阵, 则 A 在 K 上可对角化 (可上三角化) 是指存在 $\mathbf{M}_n(K)$ 中的可逆矩阵 P 使得 $P^{-1}AP$ 是对角阵 (上三角阵). 若考虑线性变换

$$f_A : K^{n\times 1} \longrightarrow K^{n\times 1}\,; \quad X \longmapsto AX$$

则矩阵 A 在 K 上可对角化 (可上三角化) 与 K-向量空间 $K^{n\times 1}$ 的线性变换 f_A 可对角化 (可上三角化) 是相互等价的条件. ■

5.0.3 特征值与特征多项式

(5.0.12) 设 V 为 K-向量空间, $\mathscr{A} \in \mathrm{End}(V)$, $\lambda \in K$. 以 I 表示 V 上的恒等变换. 集合

$$E(\lambda, \mathscr{A}) := \mathrm{Ker}(\lambda I - \mathscr{A}) = \{v \in V \mid \mathscr{A}v = \lambda v\}$$

称为 $\mathscr{A}$ 的属于 λ 的特征子空间. 这是 V 的一个子空间.

如果 $E(\lambda, \mathscr{A}) \neq 0$, 则称 λ 是 $\mathscr{A}$ 的一个特征值, 同时将 $E(\lambda, \mathscr{A})$ 中的非零向量称为 $\mathscr{A}$ 的属于 λ 的特征向量. (注意: 当 V 视作 K-向量空间时, 线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 的特征值默认要求是 K 中元素.)

对于一个方阵 $A \in \mathbf{M}_n(K)$, 线性变换

$$f_A : K^{n\times 1} \to K^{n\times 1}\,; \quad X \longmapsto AX$$

的特征值和特征向量分别称为 A 在 K 中的特征值和在 K 上的特征向量. ■

注记 5.0.13. 设 $A \in \mathbf{M}_n(K)$, L 是 $\mathbb{C}$ 的子域且 $K \subseteq L$ (我们称 L 是 K 在 $\mathbb{C}$ 中的一个扩域). 我们当然可以将 A 视作 $\mathbf{M}_n(L)$ 中的矩阵. 此时, 线性变换

$$f_{A,\,L} : L^{n\times 1} \to L^{n\times 1}\,; \quad X \longmapsto AX$$

的特征值和特征向量分别是 A 在 L 中的特征值和在 L 上的特征向量.

例如, 对于任意实方阵我们既可以讨论它的实特征值也可以讨论它的复特征值. 这与线性变换的情况有一些细微的差别.

类似地, 我们既可以讨论 A 在 K 上是否可对角化 (可上三角化) 也可以讨论 A 在 L 上是否可对角化 (可上三角化). 如果用线性变换的语言, 那么讨论 A 在 K 上的对角化时要考虑的是 K-向量空间 $K^n = K^{n\times 1}$ 上的线性变换 f_A; 而讨论 A 在 L 上的对角化时要关心的则是 L-向量空间 $L^n = L^{n\times 1}$ 上的线性变换 $f_{A,\,L}$. 虽然 f_A 和 $f_{A,\,L}$ 的定义十分相似, 但在有必要区分二者的时候, 读者应当对二者的差别十分清楚. ■

(5.0.14) 对于任意 K-向量空间 V 上的线性变换 $\mathscr{A}$, 属于不同特征值的特征向量构成的向量组一定是线性无关的. 特别地, 若 $\lambda_1, \cdots, \lambda_r \in K$ 是 $\mathscr{A}$ 的一些不同特征值, 则特征子空间的和

$$E(\lambda_1, \mathscr{A}) + \cdots + E(\lambda_r, \mathscr{A})$$

是个直和. 若 $\lambda_{r+1}, \cdots, \lambda_s \in K$ 是另外一些两两不同的常数, 且都不是 $\mathscr{A}$ 的特征值, 则

$$E(\lambda_{r+1}, \mathscr{A}) = \cdots = E(\lambda_s, \mathscr{A}) = 0\,.$$

此时 (按照有关的定义)

$$E(\lambda_1, \mathscr{A}) + \cdots + E(\lambda_r, \mathscr{A}) + E(\lambda_{r+1}, \mathscr{A}) + \cdots + E(\lambda_s\,, \mathscr{A})$$

仍然是直和.

假如 V 是有限维的, 那么线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 可对角化的一个充分必要条件是: 如果以 $\lambda_1, \cdots, \lambda_r \in K$ 表示 $\mathscr{A}$ 的所有不同的特征值, 那么

$$V = E(\lambda_1, \mathscr{A}) \oplus \cdots \oplus E(\lambda_r, \mathscr{A})\,.$$

如果 $\mathscr{A}$ 没有特征值但 $V \neq 0$, 则 $\mathscr{A}$ 不能对角化. ■

思考题 5.1. 设 $\mathscr{A}$ 为 K-向量空间 V 上的线性变换, $\lambda \in K$ 是 $\mathscr{A}$ 的一个特征值, $v \in V$ 是 $\mathscr{A}$ 的属于 λ 的一个特征向量.

证明: 对于任意多项式 $f \in K[X]$, $f(\lambda)$ 是线性变换 $f(\mathscr{A})$ 的一个特征值, v 是 $f(\mathscr{A})$ 的属于特征值 $f(\lambda)$ 的一个特征向量.

(5.0.15) 我们假定读者已经熟悉行列式的基本概念和性质 (参见本书上册第四章).

设 $A \in \mathbf{M}_n(K)$, $n \geq 1$. 若以 X 作为多项式的变元, 则 A 的*特征多项式* $P_A(X)$ 是由行列式 $|XI_n - A|$ 计算出的 (以 X 为变元的) 多项式. 它是系数全部在 K 中、次数为 n 的首一多项式. 如果另一矩阵 $B \in \mathbf{M}_n(K)$ 与 A 在 $\mathbb{C}$ 上相似, 则 A 和 B 的特征多项式相同.

现在设 V 是有限维 K-向量空间, $\dim V \geq 1$. 设 $\mathscr{A} \in \mathrm{End}(V)$. 对于 V 的任意一组有序基 $\mathcal{B}$, 矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 的特征多项式总是相同的. 我们将其定义为线性变换 $\mathscr{A}$ 的特征多项式, 记为 $P_{\mathscr{A}}(X)$. 对于任意 $\lambda \in K$, 下面的论断互相等价:

- λ 是 $\mathscr{A}$ 的特征值;
- λ 是 $P_{\mathscr{A}}(X)$ 在 K 中的一个根;
- 线性变换 $\lambda I - \mathscr{A}$ 不可逆;
- 线性变换 $\lambda I - \mathscr{A}$ 不是单射.

假设 λ 是 $\mathscr{A}$ 的特征值. 我们将 λ 作为 $P_{\mathscr{A}}(X)$ 根的重数称为 λ 作为 $\mathscr{A}$ 的特征值的*代数重数*或者简称*重数* (参见本书上册 (4.3.13)). 根据多项式的相关理论, 特征值 λ 的重数为 m 意味着特征多项式 $P_{\mathscr{A}}(X)$ 具有如下分解式

$$P_{\mathscr{A}}(X) = (X-\lambda)^m h(X)\,, \quad \text{其中 } h \in K[X] \text{ 且 } h(\lambda) \neq 0\,.$$

(参见本书上册 (A.4.10).)

若 $\lambda_1, \cdots, \lambda_r$ 是 $P_{\mathscr{A}}(X)$ 在 $\mathbb{C}$ 中的所有不同的根, 重数分别为 $m_1, \cdots, m_r$. 则在 $\mathbb{C}[X]$ 中我们有多项式分解式

$$P_{\mathscr{A}}(X) = (X-\lambda_1)^{m_1} \cdots (X-\lambda_r)^{m_r}\,.$$

特别地

$$m_1 + \cdots + m_r = \deg(P_{\mathscr{A}}(X)) = \dim V\,.$$

因此, 当 $K = \mathbb{C}$ 时, 我们可以说*按重数计入的方式计算时* $\mathscr{A}$ 共有 n 个特征值.

矩阵 $A \in \mathbf{M}_n(K)$ 在 K 中的特征值正是线性变换

$$\mathscr{A} \,:\, K^n \longrightarrow K^n\,; \quad X \longmapsto AX$$

的特征值. 由此我们可以对方阵的特征值定义代数重数, 而且以上相应的结论对矩阵的情况也适用.

关于特征多项式的一个重要定理是 Cayley–Hamilton *定理*:

若 V 是有限维 K-向量空间, $\dim V \geq 1$, $\mathscr{A} \in \mathrm{End}(V)$, $f(X) = P_{\mathscr{A}}(X)$ 为 $\mathscr{A}$ 的特征多项式, 则 $f(\mathscr{A}) = 0$.

若以矩阵方式叙述, 以上定理的内容是: 对任意正整数 n 和任意 $A \in \mathbf{M}_n(K)$, 若 $f(X) = P_A(X)$ 是 A 的特征多项式, 则 $f(A) = 0$. ■

5.0.4 不变子空间

(5.0.16) 设 V 为 K-向量空间, $\mathscr{A} \in \mathrm{End}(V)$. 所谓 $\mathscr{A}$ 的一个不变子空间是指满足以下条件的一个子空间 $U \subseteq V$:

$$\text{对任意 } u \in U\,,\quad \text{向量 } \mathscr{A}u \text{ 仍属于 } U\,.$$

当 U 为 $\mathscr{A}$ 的不变子空间时, 我们可以定义映射

$$\mathscr{A}' \;:\; U \longrightarrow U\,,\; u \longmapsto \mathscr{A}'(u) := \mathscr{A}(u)\,.$$

这是 U 上的一个线性变换, 称为 $\mathscr{A}$ 在 U 上的限制, 记为 $\mathscr{A}|_U$.

根据定义, $\mathscr{A}$ 的 1 维不变子空间实际上就是由 $\mathscr{A}$ 的一个特征向量生成的子空间. 另外, 对于任意 $\lambda \in K$, 特征子空间 $E(\lambda, \mathscr{A}) = \mathrm{Ker}(\lambda I - \mathscr{A})$ 一定是 $\mathscr{A}$ 的不变子空间. ■

(5.0.17) 假设 V 是有限维向量空间, $\mathscr{A} \in \mathrm{End}(V)$.

如果 V 可以分解为 $\mathscr{A}$ 的一些不变子空间 $U_1, \cdots, U_r$ 的直和: $V = U_1 \oplus \cdots \oplus U_r$, 且 $d_i = \dim U_i$, 则存在 V 的一组有序基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有准对角形 $\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_r \end{pmatrix}$ 的形式, 其中每个 A_i 是 d_i 阶方阵.

反过来, 如果在 V 的某一组有序基 $\mathcal{B}$ 下 $\mathscr{A}$ 的矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有上述准对角阵的形式, 则 V 可以分解为 $\mathscr{A}$ 的一些不变子空间 $U_1, \cdots, U_r$ 的直和, 其中 $\dim U_i = d_i$. ■

思考题 5.2. 设 $\mathscr{A}$ 为 K-向量空间 V 上的线性变换, U 是 $\mathscr{A}$ 的不变子空间.

证明:

1. 对于任意多项式 $f \in K[X]$, U 也是 $f(\mathscr{A})$ 的不变子空间.
2. 若 $\mathscr{B} \in \mathrm{End}(V)$ 与 $\mathscr{A}$ 可交换, 则 $\mathrm{Ker}(\mathscr{B})$ 和 $\mathrm{Im}(\mathscr{B})$ 都是 $\mathscr{A}$ 的不变子空间. 特别地, 对于任意多项式 $f \in K[X]$, $\mathrm{Ker}(f(\mathscr{A}))$ 和 $\mathrm{Im}(f(\mathscr{A}))$ 都是 $\mathscr{A}$ 的不变子空间.

思考题 5.3. 设 $\mathscr{A}$ 为 K-向量空间 V 上的线性变换, U 是 $\mathscr{A}$ 的不变子空间, $\mathscr{A}' = \mathscr{A}|_U \in \mathrm{End}(U)$ 为 $\mathscr{A}$ 在 U 上的限制.

证明:

1. 对于任意多项式 $f \in K[X]$, $f(\mathscr{A})|_U = f(\mathscr{A}')$ 且 $\mathrm{Ker}(f(\mathscr{A})) \cap U = \mathrm{Ker}(f(\mathscr{A}'))$.
2. 对任意 $\lambda \in K$, $E(\lambda\,, \mathscr{A}) \cap U = E(\lambda\,, \mathscr{A}')$.
3. 假设 W 是 $\mathscr{A}$ 的不变子空间且 $V = U \oplus W$. 再假设 V 是有限维的. 设 $f, g, h \in K[X]$ 分别为 $\mathscr{A} \in \mathrm{End}(V)$, $\mathscr{A}' = \mathscr{A}|_U \in \mathrm{End}(U)$ 和 $\mathscr{A}'' = \mathscr{A}|_W \in \mathrm{End}(W)$ 的特征多项式.

   则 $f(X) = g(X)h(X)$.

(5.0.18) 使用不变子空间的概念还可以很方便地描述线性变换可上三角化的条件. 详言之, 若 V 是 n 维向量空间, $(v_1, \cdots, v_n)$ 是 V 的一组有序基, $\mathscr{A} \in \mathrm{End}(V)$, 则下列条件等价:

- $\mathscr{A}$ 在有序基 $(v_1, \cdots, v_n)$ 下的矩阵是上三角阵;
- 对于每个 $j \in [\![1, n]\!]$, 均有 $\mathscr{A}v_j \in \mathrm{span}(v_1, \cdots, v_j)$;
- 对于每个 $j \in [\![1, n]\!]$, $\mathrm{span}(v_1, \cdots, v_j)$ 都是 $\mathscr{A}$ 的不变子空间.

可以证明 (参见本书上册中的命题3.3.32): 有限维复向量空间 (即 $\mathbb{C}$ 上的向量空间) 的任何线性变换都可以上三角化. 换成矩阵语言来说, 任何复方阵在 $\mathbb{C}$ 上都可以上三角化. ■

5.0.5 习题

若非特别说明, 以下设 m, n 为正整数.

习题 5.0.1. 设 $f: A \to B$ 为集合间的映射. 证明:

1. f 有左逆当且仅当 f 是单射.
2. f 有右逆当且仅当 f 是满射.
3. 若 f 既有左逆又有右逆, 则 f 有唯一的左逆和唯一的右逆, 并且它的左逆和右逆相等.

习题 5.0.2. 设

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

1. 求线性方程组 $AX = 0$ 的一个基础解系.
2. A 是否有左逆? 若有, 求出 A 的所有左逆, 若无, 请解释为什么.
3. A 是否有右逆? 若有, 求出 A 的所有右逆, 若无, 请解释为什么.

习题 5.0.3. 设 $U = \{(x, y, z, t) \in \mathbb{R}^4 \mid x - 3y = 2t\,, \quad y = z - t\}$.

1. 求 U 的一组基.
2. 找出 $\mathbb{R}^4$ 的一个子空间 W 使 $\mathbb{R}^4 = U \oplus W$.

习题 5.0.4. 设 $v_1, \cdots, v_m$ 和 $u_1, \cdots, u_n$ 是同一向量空间中的两个向量组. 假设 $v_1, \cdots, v_m$ 线性无关, 且每个 v_i, $i = 1, \cdots, m$ 都可以被向量组 $u_1, \cdots, u_n$ 线性表出.

证明 $m \leq n$.

习题 5.0.5. 设 U 是 $\mathbf{M}_n(K)$ 中所有的上三角阵构成的子集. 证明 U 是 $\mathbf{M}_n(K)$ 的子空间, 并求出 $\dim U$.

习题 5.0.6. 考虑列向量空间 K^4 中的向量

$$\alpha_1 = (1, 1, 0, 0)^T\,,\ \alpha_2 = (0, 1, 1, 0)^T \quad \text{和} \quad \beta_1 = (1, 2, 3, 4)^T\,,\ \beta_2 = (0, 1, 2, 2)^T\,.$$

令 $U = \mathrm{span}(\alpha_1, \alpha_2)$, $W = \mathrm{span}(\beta_1, \beta_2)$.

分别求 $U \cap W$ 和 $U + W$ 的一组基.

习题 5.0.7. 设 U, W 是向量空间 V 的子空间. 证明: 并集 $U \cup W$ 仍是 V 的子空间当且仅当包含关系 $U \subseteq W$ 和 $W \subseteq U$ 至少有一个成立.

习题 5.0.8. 设 $v_1, \cdots, v_n$ 是向量空间 V 中的线性无关组, $w \in V$. 证明:

$$\dim \mathrm{span}(v_1 + w\,, \cdots, v_n + w) \geq n - 1\,.$$

再举例说明以上不等式可能取等号也可能取严格的不等号.

习题 5.0.9. 假设 v_1, v_2, v_3, v_4 是向量空间 V 的一个生成组. 证明: $v_1 - v_2\,, v_2 - v_3, v_3 - v_4\,, v_4$ 也是 V 的生成组.

习题 5.0.10. 设 $A, B \in \mathbf{M}_n(K)$. 证明: $\operatorname{rank}(AB) \geq \operatorname{rank}(A) + \operatorname{rank}(B) - n$.

习题 5.0.11. 设 $A \in \mathbf{M}_{m\times n}(K)$.

1. 证明: 若 $K \subseteq \mathbb{R}$, 则 $\operatorname{rank}(A) = \operatorname{rank}(AA^T) = \operatorname{rank}(A^T A)$.

2. 若 $K = \mathbb{C}$, 等式 $\operatorname{rank}(A) = \operatorname{rank}(AA^T)$ 是否一定成立?

习题 5.0.12. 写出实向量空间 $\mathbb{R}^2$ 上的一个线性变换 $\mathscr{A}$ 满足 $\mathscr{A}^4 = -I$. (这里 I 表示恒等变换.)

习题 5.0.13. 对于一个复矩阵 $A = (a_{ij}) \in \mathbf{M}_{m\times n}(\mathbb{C})$, 以 $\overline{A}^T = (\overline{a}_{ji}) \in \mathbf{M}_{n\times m}(\mathbb{C})$ 表示它的共轭转置. 即, $\overline{A}^T$ 的第 (i, j) 为 a_{ji} 的共轭, 其中 a_{ji} 是 A 的第 (j, i) 位元素.

1. 按照通常的矩阵加法和数乘, 集合 $\mathbf{M}_{m\times n}(\mathbb{C})$ 作为 $\mathbb{C}$ 上向量空间的维数 $\dim_{\mathbb{C}} \mathbf{M}_{m\times n}(\mathbb{C})$ 是多少？若将 $\mathbf{M}_{m\times n}(\mathbb{C})$ 视为 $\mathbb{R}$ 上的向量空间, 其维数 $\dim_{\mathbb{R}} \mathbf{M}_{m\times n}(\mathbb{C})$ 是多少？

   如果一个方阵 $A \in \mathbf{M}_n(\mathbb{C})$ 满足 $A = \overline{A}^T$, 则称 A 是一个 **Hermite 矩阵**. 令 H 表示 $\mathbf{M}_n(\mathbb{C})$ 中所有 Hermite 矩阵构成的子集.

2. 若将 $\mathbf{M}_n(\mathbb{C})$ 视为复向量空间, H 是否是其子空间？若是, $\dim_{\mathbb{C}} H$ 是多少？若否, 原因是什么？

3. 若将 $\mathbf{M}_n(\mathbb{C})$ 视为实向量空间, H 是否是其子空间？若是, $\dim_{\mathbb{R}} H$ 是多少？若否, 原因是什么？

习题 5.0.14. 设 U, W, L 是向量空间 V 的子空间. 假设 $U \subseteq W$ 且 $V = U \oplus L$.
证明: $W = U \oplus (W \cap L)$.

习题 5.0.15. 设 $V = K[X]_{\leq 4}$, $U = \{f \in V \mid f(0) = f(1) = f(-1)\}$.

1. 找出 U 的一组基.

2. 将前一小题找到的 U 的基扩充为 V 的一组基.

习题 5.0.16. 设 $n \geq 1$, $V = K[X]_{\leq n}$.

1. 设 $f_0, f_1, \cdots, f_n$ 为 V 中向量组, 且对每个 $i \in [\![0, n]\!]$ 均有 $f_i(2) = 0$. 证明: $f_0, \cdots, f_n$ 是 V 中的线性相关组.

2. 设 $g_0, g_1, \cdots, g_n$ 为 V 中向量组, 且对每个 $i \in [\![0, n]\!]$ 均有 $g_i'(2) = 0$. 问: $g_0, \cdots, g_n$ 是否有可能是 V 中的线性无关组？

习题 5.0.17. 设 V 是向量空间. 回忆**仿射集**的定义: V 中的一个仿射(子)集是指具有如下形式的一个子集

$$v + U := \{v + u \mid u \in U\}$$

其中 $v \in V$, U 是 V 中的一个子空间. 该仿射集的维数定义为子空间 U 的维数.

1. 设 A_1, A_2 是 V 中的仿射集. 证明: $A_1 \cap A_2$ 要么是空集, 要么是仿射集.

2. 设 $v_1, \cdots, v_m \in V$. 令

$$A = \{a_1 v_1 + \cdots + a_m v_m \mid a_1, \cdots, a_m \in K \text{ 且 } a_1 + \cdots + a_m = 1\}.$$

   (a) 证明 A 是 V 中的仿射集, 且每个 v_i, $i = 1, \cdots, m$ 都属于 A.

(b) 证明 A 是 V 中包含 $v_1, \cdots, v_m$ 的最小仿射集, 即, 如果 F 是 V 中的仿射集且 $v_i, i = 1, \cdots, m$ 都包含于 F, 则 $A \subseteq F$.

我们称仿射集 A 可以由向量组 $v_1, \cdots, v_m$ 生成, 或者说 $v_1, \cdots, v_m$ 是仿射集 A 的一个生成组.

(c) 证明 $\dim A \leq m-1$.

(d) 试给出 $\dim A = m-1$ 的一个充分必要条件.

3. 假设 V 是有限维的. 证明: V 中的每个仿射子集都是有限生成的, 即, 对于 V 的任何仿射子集一定可以像问题 (b) 中的集合 A 一样由有限多个向量生成.

习题 5.0.18. 考虑线性变换

$$\mathscr{A} \;:\; K^3 \longrightarrow K^3\;; \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \longmapsto \begin{pmatrix} x-z \\ -x+2y \\ y+z \end{pmatrix}$$

1. 求 $\mathscr{A}$ 在有序标准基 $\mathcal{B} = (e_1, e_2, e_3)$ 下的矩阵.

2. 求 $\mathscr{A}$ 在有序基 (η_1, η_2, η_3) 下的矩阵, 其中

$$\eta_1 = (2,\, 0,\, -1)^T\,,\ \eta_2 = (1,\, 1,\, 1)^T\,,\ \eta_3 = (1,\, 0,\, 1)^T\,.$$

习题 5.0.19. 在 K^4 中考虑向量

$$v_1 = (1,\, 1,\, 1,\, 1)^T,\, v_2 = (1,\, 1,\, -1,\, -1)^T,\, v_3 = (1,\, -1,\, 1,\, -1)^T,\, v_4 = (1,\, -1,\, -1,\, 1)^T$$

及

$$\eta_1 = (1,\, 1,\, 0,\, 1)^T,\, \eta_2 = (2,\, 1,\, 3,\, 1)^T,\, \eta_3 = (1,\, 1,\, 0,\, 0)^T,\, \eta_4 = (0,\, 1,\, -1,\, 0)^T\,.$$

1. 证明 v_1, v_2, v_3, v_4 和 $\eta_1, \eta_2, \eta_3, \eta_4$ 是 K^4 的两组基.

2. 令 $\mathcal{B} = (v_1, v_2, v_3, v_4)$, $\mathcal{C} = (\eta_1, \eta_2, \eta_3, \eta_4)$.

   求向量 $u = (1, 0, 0, -1)^T$ 在这两组有序基下的坐标.

3. 求出一个非零向量 w 使得它在 $\mathcal{B}$ 和 $\mathcal{C}$ 这两组有序基下的坐标列相同.

4. 令 $\mathcal{B}' = (v_2, v_4, v_1, v_3)$. 求有序基 $\mathcal{B}$ 到 $\mathcal{B}'$ 的过渡矩阵.

习题 5.0.20. 设 V, W 分别是 n 和 m 维的 K-向量空间, $\mathscr{A} : V \to W$ 是线性映射. 设 $r = \operatorname{rank}(\mathscr{A})$ (即 $r = \dim \operatorname{Im}(\mathscr{A})$).

证明: 存在 V 的一组有序基 $\mathcal{B}$ 和 W 的一组有序基 $\mathcal{C}$ 使得

$$\mathcal{M}_{\mathcal{B},\,\mathcal{C}}(\mathscr{A}) = \begin{pmatrix} I_r & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}\,.$$

(这里 $\mathbf{0}_{s\times t}$ 表示大小为 $s \times t$ 的零矩阵.)

习题 5.0.21. 设 V 是有限维向量空间, $\mathscr{A} \in \operatorname{End}(V)$ 满足 $\mathscr{A}^2 = \mathscr{A}$.

证明 $V = \operatorname{Ker}(\mathscr{A}) \oplus \operatorname{Im}(\mathscr{A})$.

习题 5.0.22. 设 $(x_k)_{k\geq 0}$, $(y_k)_{k\geq 0}$ 和 $(z_k)_{k\geq 0}$ 三个数列满足

$$\begin{cases} x_{k+1} = 2x_k + 2y_k + 2z_k \\ y_{k+1} = 2x_k + 5y_k - 4z_k \\ z_{k+1} = -2x_k - 4y_k + 5z_k \end{cases} \qquad k \in \mathbb{N}.$$

假设 $x_0 = 0$, $y_0 = -1$, $z_0 = 2$. 求 (x_k), (y_k), (z_k) 三个数列的通项公式.

习题 5.0.23. 设 V 为有限维向量空间, $U \subseteq V$ 是个子空间.

证明: 如果对于每个 $\mathscr{A} \in \mathrm{End}(V)$ 来说 U 都是 $\mathscr{A}$ 的不变子空间, 则 $U = 0$ 或 $U = V$.

习题 5.0.24. 设 $\mathscr{A}$ 为向量空间 V 上的线性变换.

1. 假设存在非零向量 $v, w \in V$ 满足 $\mathscr{A}v = 3w$, $\mathscr{A}w = 3v$. 证明: 3 或 -3 是 $\mathscr{A}$ 的特征值.
2. 证明: 如果 V 中的非零向量都是 $\mathscr{A}$ 的特征向量, 那么 $\mathscr{A}$ 一定是恒等变换 I 的常数倍.
3. 假设 V 是有限维的, $n = \dim V \geq 1$. 证明: 如果 V 的每个 $n-1$ 维子空间都是 $\mathscr{A}$ 的不变子空间, 那么 $\mathscr{A}$ 一定是恒等变换 I 的常数倍.

习题 5.0.25. 设 V 是有限维 K-向量空间, $\mathscr{A} \in \mathrm{End}(V)$, $\lambda \in K$. 证明: 存在 $\alpha \in K$ 满足 $|\alpha - \lambda| < 10^{-5}$ 且线性变换 $\alpha I - \mathscr{A}$ 可逆.

习题 5.0.26. 设 $\mathscr{A}$ 是向量空间 V 上的线性变换. 假设 $\dim \mathrm{Im}(\mathscr{A}) = m$. 证明: $\mathscr{A}$ 最多有 $m+1$ 个不同的特征值.

习题 5.0.27. 令 $V = K[X]_{\leq 4}$. 考虑 V 的有序基 $\mathcal{B} = (1, X-1, (X-1)^2, (X-1)^3, (X-1)^4)$.

1. 设 $\alpha = X^2 - 2 \in V$. 求 α 在有序基 $\mathcal{B}$ 下的坐标.
2. 定义线性变换
$$\mathscr{A} : V \to V\,;\ f(X) \mapsto Xf'(X)$$
其中 $f'(X)$ 表示多项式 $f(X)$ 的形式导数.
求 $\mathscr{A}$ 在有序基 $\mathcal{B}$ 下的矩阵.
3. 求上个小题中线性变换 $\mathscr{A}$ 的特征多项式.
4. 令
$$U = \{f \in V \mid f(X) = f(-X)\,,\ W = \{f \in V \mid f(-X) = -f(X)\}\}\,.$$
分别求 U 和 W 的一组基, 并证明 $V = U \oplus W$.
5. 证明: (d) 中的子空间 U 和 W 都是 (b) 中线性变换 $\mathscr{A}$ 的不变子空间.
6. 设 $\mathscr{A}$ 如问题 (b), U 和 W 如问题 (d). 求出 $\mathscr{A}|_U$ 和 $\mathscr{A}|_W$ 的特征多项式.

习题 5.0.28. 设 V 是 n 维复向量空间, $\mathscr{A} \in \mathrm{End}(V)$. 证明: 对任意 $k \in [\![1, n]\!]$, $\mathscr{A}$ 一定有 k 维的不变子空间.

5.1 广义特征子空间和幂零变换

5.1.1 广义特征子空间

定义 5.1.1. 设 U 是 K-向量空间, $\mathscr{A} \in \mathrm{End}(U)$ 是其上的线性变换, $\lambda \in K$. 我们将集合

$$G(\lambda, \mathscr{A}) := \bigcup_{k \geq 1} \mathrm{Ker}(\lambda I - \mathscr{A})^k = \left\{ u \in U \,\middle|\, \text{存在正整数 } k \text{ 使 } (\lambda I - \mathscr{A})^k u = 0 \right\}$$

称为 $\mathscr{A}$ 的属于 λ 的**广义特征子空间** (generalized eigenspace) 或**根子空间** (root space). (当然, $G(\lambda, \mathscr{A})$ 的确是 U 的子空间. 参见命题 5.1.2 (1).)

若 v 是 $G(\lambda, \mathscr{A})$ 中的非零向量, 则称 v 是 $\mathscr{A}$ 的属于 λ 的一个**广义特征向量** (generalized eigenvector) 或**根向量** (root vector). ■

命题 5.1.2: 记号如定义 5.1.1.

1. $G(\lambda, \mathscr{A})$ 是 U 的子空间, 而且它包含特征子空间 $E(\lambda, \mathscr{A})$.

2. 若 $G(\lambda, \mathscr{A}) \neq 0$, 则 $E(\lambda, \mathscr{A}) \neq 0$, 因此此时 λ 是 $\mathscr{A}$ 的特征值.

3. $G(\lambda, \mathscr{A})$ 是 $\mathscr{A}$ 的不变子空间.

   根据思考题 5.2 (1), 对于任意 $f \in K[X]$, $G(\lambda, \mathscr{A})$ 也是 $f(\mathscr{A})$ 的不变子空间.

   特别地, 对于任意 $m \in \mathbb{N}$, $G(\lambda, \mathscr{A})$ 是 $(\lambda I - \mathscr{A})^m$ 的不变子空间.

证明. (1) 请读者自己按照"子空间"的定义来验证.

(2) 假设 $G(\lambda, \mathscr{A})$ 中含有一个非零向量 u. 我们希望证明: 特征子空间 $E(\lambda, \mathscr{A}) = \mathrm{Ker}(\lambda I - \mathscr{A})$ 也一定包含某个非零向量, 从而 λ 是 $\mathscr{A}$ 的一个特征值.

根据广义特征子空间的定义, 存在正整数 k 使 $(\lambda I - \mathscr{A})^k u = 0$. 根据自然数集的良序性质 (参见本书上册 (A.1.1))[‡], 我们可以选取 k 为满足此条件的最小正整数. 这意味着 $v := (\lambda I - \mathscr{A})^{k-1} u$ 不再等于零 (向量). 而 $(\lambda I - \mathscr{A})^k u = 0$ 这个条件恰好是指

$$(\lambda I - \mathscr{A}) v = (\lambda I - \mathscr{A})\big((\lambda I - \mathscr{A})^{k-1} u\big) = (\lambda I - \mathscr{A})^k u = 0\,,$$

亦即, $v \in E(\lambda, \mathscr{A}) = \mathrm{Ker}(\lambda I - \mathscr{A})$. 前面已经说明了 $v \neq 0$, 所以我们证明了需要的结论.

(3) 需要验证的结论是: 对于任意 $u \in G(\lambda, \mathscr{A})$, 均有 $\mathscr{A} u \in G(\lambda, \mathscr{A})$. 事实上, $u \in G(\lambda, \mathscr{A})$ 意味着存在正整数 k 满足 $(\lambda I - \mathscr{A})^k u = 0$. 由于 $(\lambda I - \mathscr{A})^k$ 和 $\mathscr{A}$ 这两个线性变换可交换, 故

$$(\lambda I - \mathscr{A})^k \mathscr{A} u = \mathscr{A}\big((\lambda I - \mathscr{A})^k u\big) = \mathscr{A}(0) = 0\,.$$

所以 $\mathscr{A} u \in \mathrm{Ker}(\lambda I - \mathscr{A})^k \subseteq G(\lambda, \mathscr{A})$. □

思考题 5.4. 记号如定义 5.1.1. 假设 $W \subseteq U$ 是 $\mathscr{A}$ 的不变子空间, $\mathscr{A}' = \mathscr{A}|_W \in \mathrm{End}(W)$, 从而可以对 $\lambda \in K$ 定义 $\mathscr{A}'$ 在 W 中的广义特征子空间 $G(\lambda, \mathscr{A}')$.

证明: $G(\lambda, \mathscr{A}') = G(\lambda, \mathscr{A}) \cap W$.

在本书上册引理 3.3.18 中我们曾证明: 属于不同特征值的特征向量构成的向量组一定是线性无关的. 现在我们来证明同样的结论对于广义特征向量也成立.

[‡]即, 自然数集的任何一个非空子集都含有最小的元素. 根据这个性质可以知道, 如果存在某个自然数满足某个条件, 那么一定可以找到满足同一条件的最小自然数.

命题 5.1.3: 设 U 是 K-向量空间, $\mathscr{A} \in \operatorname{End}(U)$, $\lambda_1, \cdots, \lambda_r \in K$ 是 $\mathscr{A}$ 的不同特征值, $v_1, \cdots, v_r$ 分别是属于 $\lambda_1, \cdots, \lambda_r$ 的广义特征向量.

则向量组 $v_1, \cdots, v_r$ 是线性无关的.

证明. 假设 $a_1, \cdots, a_r \in K$ 满足

$$a_1 v_1 + \cdots + a_r v_r = 0\,. \tag{5.1.3.1}$$

根据广义特征向量的定义, 对每个 $i \in [\![1, r]\!]$, 存在正整数 m_i 使得 $(\lambda_i I - \mathscr{A})^{m_i} v_i = 0$. 我们可以假设所选取的 m_i 已经达到最小可能的取值, 从而 $(\lambda_i I - \mathscr{A})^{m_i - 1} v_i \neq 0$. 于是, 若令 $k = m_1 - 1$, 则 $w := (\lambda_1 I - \mathscr{A})^{m_1 - 1} v_1$ 是非零向量, 且

$$(\lambda_1 I - \mathscr{A}) w = (\lambda_1 I - \mathscr{A})^{m_1} v_1 = 0\,.$$

后面这个条件说明 $\mathscr{A} w = \lambda_1 w$. 由此可以进一步说明 (参见思考题 5.1)

$$\text{对任意 } f \in K[X]\,, \quad f(\mathscr{A}) w = f(\lambda_1) w\,. \tag{5.1.3.2}$$

现在考虑线性变换

$$S := (\lambda_1 I - \mathscr{A})^k (\lambda_2 I - \mathscr{A})^{m_2} \cdots (\lambda_r I - \mathscr{A})^{m_r}\,.$$

因为 $(\lambda_1 I - \mathscr{A})^k$, $(\lambda_2 I - \mathscr{A})^{m_2}, \cdots, (\lambda_r I - \mathscr{A})^{m_r}$ 这些线性变换都是关于 $\mathscr{A}$ 的多项式, 所以它们两两可交换. 又因为 $(\lambda_i I - \mathscr{A})^{m_i} v_i = 0$, 故

$$S(v_2) = \cdots = S(v_r) = 0\,.$$

另一方面, 计算可知

$$\begin{aligned}
S(v_1) &= (\lambda_2 I - \mathscr{A})^{m_2} \cdots (\lambda_r I - \mathscr{A})^{m_r} \big((\lambda_1 I - \mathscr{A})^k v_1 \big) \\
&= (\lambda_2 I - \mathscr{A})^{m_2} \cdots (\lambda_r I - \mathscr{A})^{m_r} w \\
\text{根据 (5.1.3.2)} \quad &= (\lambda_2 - \lambda_1)^{m_2} \cdots (\lambda_r - \lambda_1)^{m_r} w
\end{aligned}$$

于是, 结合 (5.1.3.1) 式可知

$$0 = S(0) = a_1 S(v_1) + \cdots + a_r S(v_r) = a_1 S(v_1) = a_1 (\lambda_2 - \lambda_1)^{m_2} \cdots (\lambda_r - \lambda_1)^{m_r} w\,.$$

因为 λ_i 两两不同且 $w \neq 0$, 所以上式表明 $a_1 = 0$.

同理可证, (5.1.3.1) 式中所有系数 a_i 都必须等于 0. 这就证明了向量组 $v_1, \cdots, v_r$ 的线性无关性. □

今后我们主要研究有限维向量空间上的线性变换. 现在我们希望在有限维的情况给出广义特征子空间的更简洁描述.

为此我们先证明一个引理.

引理 5.1.4: 设 T 是向量空间 U 上的线性变换.

1. 对任意 $k \in \mathbb{N}$ 均有 $\operatorname{Ker}(T^k) \subseteq \operatorname{Ker}(T^{k+1})$. (按通常的约定, $T^0 = I$ 是恒等变换.)
2. 若存在 $m \in \mathbb{N}$ 使得 $\operatorname{Ker}(T^m) = \operatorname{Ker}(T^{m+1})$, 则对所有自然数 $k \geq m$ 均有 $\operatorname{Ker}(T^k) = \operatorname{Ker}(T^{k+1})$.
3. 若 U 是有限维向量空间, $n = \dim U$, 则

$$\operatorname{Ker}(T^n) = \operatorname{Ker}(T^{n+1}) = \operatorname{Ker}(T^{n+2}) = \cdots$$

证明. (1) 按定义验证即可.

(2) 设 $k \geq m$. 若 $v \in \operatorname{Ker}(T^{k+1})$, 则 $u := T^{k-m}v \in \operatorname{Ker}(T^{m+1})$. 根据 $\operatorname{Ker}(T^{m+1}) = \operatorname{Ker}(T^m)$ 这一假设, $u = T^{k-m}v \in \operatorname{Ker}(T^m)$. 这就是说

$$T^k v = T^m(T^{k-m}v) = T^m u = 0\,,$$

即 $v \in \operatorname{Ker}(T^k)$. 这样我们证明了 $\operatorname{Ker}(T^{k+1}) \subseteq \operatorname{Ker}(T^k)$. 因为另一方向的包含关系也成立 (如结论 (1) 所述), 所以 $\operatorname{Ker}(T^k) = \operatorname{Ker}(T^{k+1})$.

(3) 若存在 $k \leq n$ 使得 $\operatorname{Ker}(T^k) = \operatorname{Ker}(T^{k+1})$, 则结论由 (2) 即得. 若不然, 则

$$0 = \operatorname{Ker}(T^0) \subsetneq \operatorname{Ker}(T) \subsetneq \cdots \subsetneq \operatorname{Ker}(T^{n-1}) \subsetneq \operatorname{Ker}(T^n) \subsetneq \operatorname{Ker}(T^{n+1})$$

这一串包含关系都是真包含. 则

$$\dim \operatorname{Ker}(T^{n+1}) \geq 1 + \dim \operatorname{Ker}(T^n) \geq 1 + (1 + \dim \operatorname{Ker}(T^{n-1})) \geq \cdots \geq n + 1 + \dim \operatorname{Ker}(T^0) = n + 1\,.$$

但 $\operatorname{Ker}(T^{n+1})$ 是 U 的子空间, 故上式与 $\dim U = n$ 这个假设矛盾. □

推论 5.1.5: 设 V 是 n 维 K-向量空间. 则对于任意 $\mathscr{A} \in \operatorname{End}(V)$ 和任意 $\lambda \in K$ 均有

$$G(\lambda\,,\,\mathscr{A}) = \operatorname{Ker}(\lambda I - \mathscr{A})^n\,.$$

证明. 将引理 5.1.4 中的结论 (1) 和 (3) 应用于线性变换 $T := \lambda I - \mathscr{A}$ 即可. □

例 5.1.6. 设 $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $\mathscr{A} \in \operatorname{End}(K^3)$ 是由对应法则 $X \mapsto AX$ 给出的线性变换. 则

$$G(0\,,\,\mathscr{A}) = \operatorname{Ker}(\mathscr{A}^2) = \operatorname{Ker}((0 \cdot I - \mathscr{A})^2) = \operatorname{Ker}((0 \cdot I - \mathscr{A})^3)\,.$$

(这个例子表明: 在推论 5.1.5 的条件下有可能存在 $m < \dim V$ 使得 $G(\lambda\,,\,\mathscr{A}) = \operatorname{Ker}(\lambda I - \mathscr{A})^m$.) ■

例 5.1.7. 设 $A = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} \in \mathbf{M}_n(K)$, $\mathscr{A} \in \operatorname{End}(K^n)$ 是由对应法则 $X \mapsto AX$ 给出的线性变换. 则 $K^n = G(0\,,\,\mathscr{A}) = \operatorname{Ker}(\mathscr{A}^n) \neq \operatorname{Ker}(\mathscr{A}^{n-1})$. ■

例 5.1.8. 考虑向量空间 K^3 上的线性变换

$$\mathscr{A}\ :\ K^3 \longrightarrow K^3\,;\quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \longmapsto \begin{pmatrix} 4y \\ 0 \\ 5z \end{pmatrix}\,.$$

不难求出 $\mathscr{A}$ 的特征多项式为 $f(\lambda) = \lambda^2(\lambda - 5)$, 特征值为 0 和 5, 相应的特征子空间为

$$E(0\,,\,\mathscr{A}) = \operatorname{span}((1,\,0,\,0)^T)\,,\quad E(5\,,\,\mathscr{A}) = \operatorname{span}((0,\,0,\,1)^T)$$

而相应的广义特征子空间为

$$G(0\,,\,\mathscr{A}) = \operatorname{Ker}(\mathscr{A}^2) = \operatorname{Ker}(\mathscr{A}^3) = \operatorname{span}((0,\,1,\,0)^T\,,\,(1,\,0,\,0)^T)\,,$$
$$G(5\,,\,\mathscr{A}) = E(5\,,\,\mathscr{A}) = \operatorname{Ker}(5I - \mathscr{A}) = \operatorname{span}((0,\,0,\,1)^T)\,.$$

不难看出, $K^3 = G(0\,,\,\mathscr{A}) \oplus G(5\,,\,\mathscr{A})$ 但 $K^3 \neq E(0\,,\,\mathscr{A}) \oplus E(5\,,\,\mathscr{A})$. ■

接下来我们想说明一个重要的事实: 在有限维的情况下, 当一个线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 有足够多的特征值时, 整个空间 V 可以分解为 $\mathscr{A}$ 的所有广义特征子空间的直和.

为此我们先证明一个引理.

引理 5.1.9: 设 V 是 n 维向量空间, $T \in \mathrm{End}(V)$. 则 $V = \mathrm{Ker}(T^n) \oplus \mathrm{Im}(T^n)$.

证明. 首先来说明 $\mathrm{Ker}(T^n) + \mathrm{Im}(T^n)$ 是一个直和, 即 $\mathrm{Ker}(T^n) \cap \mathrm{Im}(T^n) = 0$. 为此, 假设 $v \in \mathrm{Ker}(T^n) \cap \mathrm{Im}(T^n)$. 由 $v \in \mathrm{Im}(T^n)$ 可知, 存在 $u \in V$ 使得 $v = T^n u$. 再由 $v \in \mathrm{Ker}(T^n)$ 可得 $T^{2n}u = T^n(T^n u) = T^n v = 0$. 所以, $u \in \mathrm{Ker}(T^{2n})$. 根据引理 5.1.4 (3), $\mathrm{Ker}(T^n) = \mathrm{Ker}(T^{2n})$. 所以 $u \in \mathrm{Ker}(T^n)$, 故 $v = T^n u = 0$.

根据以上结果我们可以得到

$$
\begin{aligned}
& \dim(\mathrm{Ker}(T^n) + \mathrm{Im}(T^n)) \\
& \text{(根据本书上册定理 2.3.30 (2) 中的维数公式)} \\
&= \dim \mathrm{Ker}(T^n) + \dim \mathrm{Im}(T^n) - \dim(\mathrm{Ker}(T^n) \cap \mathrm{Im}(T^n)) \\
&= \dim \mathrm{Ker}(T^n) + \dim \mathrm{Im}(T^n) + 0 \\
\text{(根据秩–零化度定理)} \quad &= \dim V\,.
\end{aligned}
$$

这说明 $\mathrm{Ker}(T^n) + \mathrm{Im}(T^n) = V$. 因为已经说明了这里的 $\mathrm{Ker}(T^n) + \mathrm{Im}(T^n)$ 是直和, 引理至此得证. □

定理 5.1.10: 设 V 是有限维 K-向量空间, $\dim V \geq 1$, $\mathscr{A} \in \mathrm{End}(V)$, $f(X) = P_{\mathscr{A}}(X) \in K[X]$ 为 $\mathscr{A}$ 的特征多项式. 假设 f 在 $\mathbb{C}$ 中所有不同的根 $\lambda_1, \cdots, \lambda_r$ 都属于 K. (这个假设在 $K = \mathbb{C}$ 时显然成立.)

则

$$V = G(\lambda_1\,, \mathscr{A}) \oplus \cdots \oplus G(\lambda_r\,, \mathscr{A})\,,$$

即, V 可以分解为 $\mathscr{A}$ 的所有不同特征值对应的广义特征子空间之直和.

证明. 我们通过对 V 的维数 $n = \dim V$ 做归纳来完成证明.

若 $n = 1$, 则特征多项式 f 的次数为 $\deg(f) = n = 1$, 故 $r = 1$ 且对于特征值 λ_1 我们知道 $G(\lambda_1\,, \mathscr{A}) \neq 0$. 作为 1 维向量空间 V 的非零子空间, $G(\lambda_1, \mathscr{A})$ 必然等于 V. 因此命题的结论在 $n = \dim V$ 等于 1 时成立.

下面我们假设 $n = \dim V > 1$. 按照第二数学归纳原理 (本书上册定理 A.1.4), 我们假设以下结论成立: 对于任意维数小于 n 的非零向量空间 U 以及任意线性变换 $\mathscr{A}' \in \mathrm{End}(U)$, 若 $\mathscr{A}'$ 的特征多项式在 $\mathbb{C}$ 中的根全部属于 K, 则 U 可以分解 $\mathscr{A}'$ 的所有不同特征值对应的广义特征子空间之直和.

对于 V 上的线性变换 $\mathscr{A}$, 我们现在考虑引理 5.1.9 给出的直和分解

$$V = U \oplus W\,, \quad \text{其中 } U := \mathrm{Im}(\lambda_1 I - \mathscr{A})^n\,, \quad W := \mathrm{Ker}(\lambda_1 I - \mathscr{A})^n = G(\lambda_1\,, \mathscr{A})\,. \tag{5.1.10.1}$$

(当然, 这里 $W = G(\lambda_1\,, \mathscr{A})$ 这个等式是基于推论 5.1.5.) 根据思考题 5.2 (2), 这里的 U 和 W 都是 $\mathscr{A}$ 的不变子空间. 又由于 $W = G(\lambda_1\,, \mathscr{A}) \neq 0$, 故 $\dim U < n = \dim V$.

设 $g(X) \in K[X]$ 为 $\mathscr{A}' := \mathscr{A}|_U$ 的特征多项式. 则思考题 5.3 (2) 告诉我们 $g(X)$ 整除 $f(X)$. 特别地, $g(X)$ 的复数根包含在 $f(X)$ 的复数根集合 $\{\lambda_1, \cdots, \lambda_r\}$ 之中. 此外, 注意到 (思考题 5.4)

$$G(\lambda_1\,, \mathscr{A}') = G(\lambda_1\,, \mathscr{A}) \cap U = W \cap U = 0$$

所以 λ_1 不是 $\mathscr{A}'$ 的特征值 (命题 5.1.2 (3)), 即, λ_1 不是 $g(X)$ 的根.

不妨设 $\lambda_s, \cdots, \lambda_r$ (其中 $2 \leq s \leq r$) 是 g 的所有不同复数根. 根据题设, 它们全都属于 K. 于是, 根据归纳假设,

$$U = G(\lambda_s\,, \mathscr{A}') \oplus \cdots \oplus G(\lambda_r\,, \mathscr{A}')\,.$$

对于任意常数 $\lambda\in K$, 当 λ 不是 $\mathscr{A}'$ 的特征值时 $G(\lambda', \mathscr{A}')=0$ (命题 5.1.2 (3)). 此时我们可以在以上直和分解式中添加 $G(\lambda, \mathscr{A}')$ 这一项. 由此可知, 以上直和分解式可以改写为

$$U=G(\lambda_2\,, \mathscr{A}')\oplus\cdots\oplus G(\lambda_r\,, \mathscr{A}')\,. \tag{5.1.10.2}$$

综合 (5.1.10.1) 和 (5.1.10.2) 可以看出, 现在我们只需要验证

$$\text{对于每个 } i\geq 2\,,\quad G(\lambda_i\,, \mathscr{A}')=G(\lambda_i\,, \mathscr{A})\,.$$

事实上, 因为 $G(\lambda_i\,, \mathscr{A}')=G(\lambda_i\,, \mathscr{A})\cap U$ (思考题 5.4), 所以实际上要做的事情是验证

$$\text{对于每个 } i\geq 2\,,\quad G(\lambda_i\,, \mathscr{A})\subseteq U=\operatorname{Im}(\lambda_1 I-\mathscr{A})^n\,. \tag{5.1.10.3}$$

为此, 设 $v\in G(\lambda_i\,, \mathscr{A})$, 其中 $i\geq 2$. 将 v 视为 V 中元素, 利用 (5.1.10.1) 式可得分解式

$$v=w+u\,,\qquad \text{其中 } w\in W=G(\lambda_1\,, \mathscr{A})\,,\ u\in U\,.$$

再根据 (5.1.10.2) 式, $u\in U$ 可以写成

$$u=u_2+\cdots+u_r\,,\qquad \text{其中对每个 } j\in[\![2\,,r]\!]\,,\ u_j\in G(\lambda_j\,, \mathscr{A}')\subseteq G(\lambda_j\,, \mathscr{A})\,.$$

于是

$$v=w+u=w+u_2+\cdots+u_r\,.$$

由上式整理可得

$$\begin{aligned}0&=w+u_2+\cdots+u_{i-1}+(u_i-v)+u_{i+1}+\cdots+u_r\\&=v_1+v_2\cdots+v_{i-1}+v_i+v_{i+1}+\cdots+v_r\end{aligned} \tag{5.1.10.4}$$

其中

$$v_1:=w\,,\ v_i:=u_i-v\,,\quad \text{而对于 } j\in[\![2\,,r]\!]\setminus\{i\}\,,\ v_j:=u_j\,.$$

这里, 对每个 $j\in[\![1\,,r]\!]$, 均有 $v_j\in G(\lambda_j\,, \mathscr{A})$ 成立. (注意: $v\in G(\lambda_i\,, \mathscr{A})$ 和 $w\in G(\lambda_1\,, \mathscr{A})$ 是我们之前的假设.) 所以, 结合命题 5.1.3 与 (5.1.10.4) 式可知每个 $v_j=0$. 因此 $w=v_1=0$, 而 $v=w+u=u\in U$. 这样我们就证明了 (5.1.10.3). 至此定理证毕. □

推论 5.1.11: 设 V 是有限维 K-向量空间, $\dim V\geq 1$, $\mathscr{A}\in\operatorname{End}(V)$. 设 $\lambda_1,\cdots,\lambda_r$ 是特征多项式 $f(X)=P_{\mathscr{A}}(X)$ 的所有不同复数根.

则以下条件等价:

1. $\mathscr{A}$ 可对角化.
2. 对于每个 $i\in[\![1, r]\!]$, 均有 $\lambda_i\in K$ 且 $E(\lambda_i, \mathscr{A})=G(\lambda_i, \mathscr{A})$.

证明. 假设 $\mathscr{A}$ 可对角化. 则在 V 的某一组有序基下 $\mathscr{A}$ 的矩阵 A 是系数取自 K 的对角阵. 此时 A 的对角线上元素就是特征多项式 $f(X)$ 的所有复数根 (重数计入). 因此 f 的所有复数根属于 K. 根据 $\mathscr{A}$ 可对角化的假设, 再结合定理 5.1.10 可知

$$E(\lambda_1, \mathscr{A})\oplus\cdots\oplus E(\lambda_r, \mathscr{A})=V=G(\lambda_1, \mathscr{A})\oplus\cdots\oplus G(\lambda_r, \mathscr{A})\,.$$

因此,

$$\sum_{i=1}^r\dim E(\lambda_i, \mathscr{A})=\dim V=\sum_{i=1}^r\dim G(\lambda_i, \mathscr{A})\,.$$

我们又知道 $E(\lambda_i, \mathscr{A}) \subseteq G(\lambda_i, \mathscr{A})$, 故由以上等式可知 $\dim E(\lambda_i, \mathscr{A}) = \dim G(\lambda_i, \mathscr{A})$ 对每个 $i \in [\![1, r]\!]$ 成立. 于是我们证明了 (1)⇒(2).

反过来, 若条件 (2) 成立, 则根据定理 5.1.10 可得

$$V = E(\lambda_1, \mathscr{A}) \oplus \cdots \oplus E(\lambda_r, \mathscr{A})\,.$$

而这正是 $\mathscr{A}$ 可对角化的一个充分必要条件 (参见 (5.0.14) 一段或本书上册定理 3.3.21). 这就证明了 (2)⇒(1). □

5.1.2 幂零变换及其 Jordan 标准形

设 $\mathscr{A}$ 是有限维 K-向量空间 V 上的线性变换. 假设 $\mathscr{A}$ 的特征多形式 $P_{\mathscr{A}}(X)$ 的复数根全部属于 K. 在推论 5.1.11 中, 我们通过比较 $\mathscr{A}$ 的特征子空间和广义特征子空间得到了判断 $\mathscr{A}$ 是否可对角化的又一个判别法. 一般来说, 对于 $\mathscr{A}$ 的任一特征值 $\lambda \in K$, 我们已经知道 $U := E(\lambda, \mathscr{A})$ 和 $W := G(\lambda, \mathscr{A})$ 都是线性变换 $T := \lambda I - \mathscr{A}$ 的不变子空间. 显然 $T|_U = 0$, 即 T 在 U 上的限制恒等于 0. 然而, 当 $U \neq W$ 时, $T|_W$ 则不是零变换. 若想在一般情况下更方便地描述 $T|_W$ 的性质, 我们本小节将要讨论的幂零变换这个概念会显得十分有用.

定义 5.1.12. 设 $\mathscr{N}$ 是 K-向量空间 W 上的线性变换. 如果存在正整数 r 使得 $\mathscr{N}^r = 0$, 则称 $\mathscr{N}$ 是一个**幂零变换** (nilpotent transformation) 或**幂零算子** (nilpotent operator). 当 $\mathscr{N}$ 是幂零变换时, 满足条件 $\mathscr{N}^r = 0$ 的正整数 r 有一个最小值, 这个值称为 $\mathscr{N}$ 的**幂零阶** (nilpotent order).

设 $m \in \mathbb{N}^*$, N 为 m 阶方阵. 如果存在正整数 r 使得 $N^r = 0$, 则称 N 是一个**幂零矩阵** (nilpotent matrix). 当 N 是幂零矩阵时, 满足条件 $N^r = 0$ 的正整数 r 有一个最小值, 这个值称为 N 的**幂零阶** (nilpotent order). ■

思考题 5.5. 设 $\mathscr{N}$ 是有限维向量空间 W 上的线性变换, N 是 $\mathscr{N}$ 在 W 的某一组有序基下的矩阵.

证明: $\mathscr{N}$ 是幂零变换当且仅当 N 是幂零矩阵, 而且在条件成立时 $\mathscr{N}$ 的幂零阶等于 N 的幂零阶.

例 5.1.13. 设 $\mathscr{A}$ 是有限维 K-向量空间 V 上的线性变换, $\lambda \in K$ 为 $\mathscr{A}$ 的特征值, $W = G(\lambda, \mathscr{A})$ 为相应的广义特征子空间. 则线性变换 $\lambda I - \mathscr{A}$ 在 W 上的限制 $\mathscr{N} := (\lambda I - \mathscr{A})|_W$ 是 W 上的一个幂零变换. (请读者解释理由.) ■

推论 5.1.14: 设 $\mathscr{N}$ 为有限维 K-向量空间 W 上的幂零变换.

则 $\mathscr{N}$ 可对角化当且仅当 $\mathscr{N} = 0$. 换言之, 非零的幂零变换均不可对角化.

若以矩阵语言表述, 任何非零的幂零矩阵在 $\mathbb{C}$ 上都不可对角化.

证明. 若 $\mathscr{N}$ 可对角化, 则在 W 的某一组基下 $\mathscr{N}$ 的矩阵 N 是对角阵. 而思考题 5.5 说明 N 又是幂零矩阵. 幂零的对角矩阵只能是零矩阵, 因此 $N = 0$, 从而 $\mathscr{N} = 0$. □

引理 5.1.15: 设 $m \in \mathbb{N}^*$, $\mathscr{N}$ 为 m 维 K-向量空间 W 上的线性变换.

则下列陈述等价:

1. $\mathscr{N}^m = 0$.
2. $\mathscr{N}$ *是幂零变换.*
3. *(在不计重根的意义下)* $\mathscr{N}$ *的特征多项式以 0 为唯一的复数根.*
4. $\mathscr{N}$ *的特征多项式为* X^m.

证明. 由幂零变换的定义可知 (1) ⇒ (2). 而 Cayley–Hamilton 定理说明 (4) ⇒ (1).

现在设 $f(X) = P_{\mathscr{N}}(X)$ 为 $\mathscr{N}$ 的特征多项式. 我们知道在 $\mathbb{C}[X]$ 中存在分解式

$$f(X) = (X - \lambda_1)^{m_1} \cdots (X - \lambda_r)^{m_r} \,,$$

其中 $\lambda_1, \cdots, \lambda_r$ 是 $f(X)$ 所有不同的复数根, $m_1, \cdots, m_r$ 为相应的重数, 且

$$m_1 + \cdots + m_r = \deg(f) = \dim W = m \,.$$

因此, 如果 0 是 $f(X)$ 的唯一复数根, 即上述讨论中 $r = 1$, 那么 $m_1 = m$, $f(X) = (X - 0)^{m_1} = X^m$. 反之, 如果 $f(X) = X^m$, 则显然 f 的唯一复数根是 0. 这就说明了 (3) 和 (4) 等价.

下面只需证明 (2) ⇒ (3).

我们假设 $\mathscr{N}$ 是幂零变换, 我们希望证明 0 是特征多项式 $f(X) = P_{\mathscr{N}}(X)$ 的唯一复数根.

为此, 假设 $\lambda \in \mathbb{C}$ 是 $f(X)$ 的根. 设 $N \in \mathbf{M}_m(K)$ 是 $\mathscr{N}$ 在 W 的某一组有序基下的矩阵. 则 f 也是矩阵 N 的特征多项式. 因此, f 也是复线性变换

$$\varphi \,:\, \mathbb{C}^m \longrightarrow \mathbb{C}^m \,; \quad v \longmapsto Nv$$

的特征多项式. 故 λ 是 φ 的一个特征值. 由于 $\mathscr{N}$ 是幂零变换, 故 N 是幂零矩阵 (思考题 5.5). 因此, 若取正整数 r 使 $N^r = 0$, 则 $\varphi^r = 0$.

设 $v \in \mathbb{C}^m$ 为 φ 的属于 λ 的特征向量. 则 $0 = \varphi^r v = \lambda^r v$ (前一个等号是因为 $\varphi^r = 0$). 由此可知 $\lambda^r = 0$, 故 $\lambda = 0$. 引理至此证毕. □

推论 5.1.16: *设 $m \in \mathbb{N}^*$. 对于任意 m 阶幂零矩阵 N, 必有 $N^m = 0$.*

证明. 将引理 5.1.15 中条件 (1) 和 (2) 的等价性换成矩阵语言表述即可. □

例 5.1.17. 设 N 为 m 阶严格上三角阵 (即 N 的对角线和对角线以下的矩阵元素都是 0). 则 N 是幂零矩阵.

要说明这一点, 我们可以用行列式的基本性质知道 N 的特征多项式等于 X^m, 从而引理 5.1.15 中的条件 (4) 被验证.

事实上, 不用行列式和特征多项式的概念也可以直接说明 N 是幂零的. 为此, 我们可以考虑线性变换 $\mathscr{N} : K^m \to K^m$; $x \mapsto Nx$, 它在 K^m 的标准基 $(e_1, \cdots, e_m)$ 下的矩阵就是 N. 根据 N 是严格上三角阵的假设可以知道

$$\begin{aligned}
&\mathscr{N} e_1 = 0 \,, \\
&\mathscr{N} e_2 \in \operatorname{span}(e_1) \,, \\
&\mathscr{N} e_3 \in \operatorname{span}(e_1,\, e_2) \,, \\
&\cdots\cdots\cdots \\
&\mathscr{N} e_m \in \operatorname{span}(e_1, \cdots,\, e_{m-1}) \,.
\end{aligned}$$

于是, 若令

$$W_0 = 0 \subseteq W_1 = \operatorname{span}(e_1) \subseteq \cdots \subseteq W_{m-1} = \operatorname{span}(e_1, \cdots, e_{m-1}) \subseteq W_m = K^m \,,$$

则对于每个 $i \in [\![1,\, m]\!]$ 均有 $\mathscr{N}(W_i) \subseteq W_{i-1}$. 于是

$$\mathscr{N}^m(K^m) = \mathscr{N}^m(W_m) \subseteq \mathscr{N}^{m-1}(W_{m-1}) \subseteq \cdots \subseteq \mathscr{N}(W_1) \subseteq W_0 = 0 \,.$$

所以 $\mathscr{N}^m = 0$. 这就说明 $N^m = 0$.

根据推论 5.1.14, 我们现在知道: 任何非零的严格上三角阵都不可对角化. ■

思考题 5.6. 设 $A \in \mathbf{M}_m(K)$. 假设存在 $\lambda \in K$ 使得 $A - \lambda I_m$ 是严格上三角阵但 $A \neq \lambda I_m$.

证明: A 不可以对角化.

利用例 5.1.17 和思考题 5.6 我们可以轻易写出许多不能对角化的线性变换. 但我们可以接下来问这样的问题: 尽管一个线性变换可能无法对角化, 但是否可以找到一组基使该变换在这组基下的矩阵具有尽可能简单的形式? 如果可以, 这样的最简形式是否具有某种唯一性?

要回答这类问题就需要使用 Jordan 标准形的概念.

定义 5.1.18. 设 $n \in \mathbb{N}^*$. 我们将形如

$$\begin{pmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \lambda & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{pmatrix}, \quad \text{其中 } \lambda \in K$$

的 n 阶方阵称为 K 上的一个 n 阶 Jordan **块** (**矩阵**) (Jordan[‡] block (matrix)), 并记之为 $J_n(\lambda)$. 这里, 如果 $n = 1$, 我们认为 $J_1(\lambda)$ 就是 1 阶矩阵 (λ).

具体来说, Jordan 块 $J_n(\lambda)$ 的矩阵元素 a_{ij} 要求满足

$$a_{ij} = \begin{cases} \lambda & \text{当 } i = j\,, \\ 1 & \text{当 } i = j - 1\,, \\ 0 & \text{其他情况.} \end{cases}$$

如果矩阵 $J \in \mathbf{M}_n(K)$ 具有如下准对角形的形式

$$J = \begin{pmatrix} J_{m_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{m_r}(\lambda_r) \end{pmatrix}, \quad \text{其中 } J_{m_i}(\lambda_i) \in \mathbf{M}_{m_i}(K) \text{ 均为 Jordan 块},$$

则称 J 是 K 上的 Jordan **形矩阵** (Jordan matrix, matrix of Jordan type). (注意: 这里的 λ_i 不要求两两不同.) 显然, 这里的 λ_i 都是 J 的特征值, Jordan 块的数量 r、特征值序列 $\lambda_1, \cdots, \lambda_r$ 的排列顺序和各个 Jordan 块的大小完全决定了 Jordan 形矩阵 J 的形状.

对于任意矩阵 $A \in \mathbf{M}_n(K)$, 如果存在 Jordan 形矩阵 J 与 A 在 K 上相似, 则称 J 是 A 在 K 上的一个 Jordan **标准形** (Jordan canonical form). (注意: 在这个定义中 J 的存在性只是一个假设, 不是必然发生的.)

在有可能的情况下, 找到可逆矩阵 $P \in \mathbf{M}_n(K)$ 使 $J = P^{-1}AP$ 为 Jordan 形矩阵并写出 J 的具体形式, 这个 (论证及计算) 全过程通常被概括为"将 A 化为 Jordan 标准形"的过程. ∎

显然, 如果矩阵 $A \in \mathbf{M}_n(K)$ 在 K 上有 Jordan 标准形, 则 A 的所有复特征值必须都属于 K.

思考题 5.7. 设 J 是 n 阶的幂零 Jordan 块矩阵, A 为 $m \times n$ 矩阵, B 为 $n \times s$ 矩阵.

通过计算 AJ 和 JB 你能发现什么现象? J 的幂零阶是多少?

思考题 5.8. 假设 J 是 6 阶的 Jordan 形幂零矩阵. 请问: J 一共有几种可能的形式? 它们分别是什么样子? 它们各自的幂零阶是多少?

[‡]Camille Jordan (1838–1922), 法国数学家. 他于 1870 年陈述了 Jordan 分解定理. 在大学数学专业课中, 这位 Jordan 的名字还会出现在"Jordan 测度"、"Jordan 闭曲线定理"、(群论中的)"Jordan–Hölder 定理"等名词当中.

思考题 5.9. 假设 J 是 Jordan 形幂零矩阵. 证明: J 中的各个 Jordan 块的阶数最大值等于 J (作为幂零矩阵) 的幂零阶.

定义 5.1.19. 设 $n \in \mathbb{N}^*$, V 是 n 维 K-向量空间, $\mathscr{A} \in \mathrm{End}(V)$.

如果在 V 的一组有序基 $\mathcal{B}$ 下 $\mathscr{A}$ 的矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是 Jordan 形矩阵, 我们就称 $\mathcal{B}$ 是 $\mathscr{A}$ 的一组 Jordan **基** (Jordan basis), 而矩阵 $J = \mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 称为 $\mathscr{A}$ 的一个 Jordan **标准形** (Jordan canonical form). 在有可能的情况下, 找到一组 Jordan 基 $\mathcal{B}$ 并写出矩阵 $J = \mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 的具体形式, 这个 (论证及计算) 全过程通常被概括为"将 $\mathscr{A}$ 化为 Jordan 标准形"的过程. ■

我们在本节剩下的任务是: 证明有限维向量空间上的幂零变换都有 Jordan 标准形, 并解释如何将其化为 Jordan 标准形.

为此, 我们先来进一步研究幂零变换的 Jordan 基应该具备什么性质.

命题 5.1.20: 设 W 是有限维 K-向量空间, $\mathscr{N} \in \mathrm{End}(W)$ 是个幂零变换. 假设 $\mathcal{B}$ 是 $\mathscr{N}$ 的一组 Jordan 基, r 为 Jordan 形矩阵 $J = \mathcal{M}_{\mathcal{B}}(\mathscr{N})$ 中 Jordan 块的数目, $m_1, \cdots, m_r \in \mathbb{N}^*$ 是 J 中各个 Jordan 块的大小. 即,

$$J = \mathcal{M}_{\mathcal{B}}(\mathscr{N}) = \begin{pmatrix} J_{m_1}(0) & & \\ & \ddots & \\ & & J_{m_r}(0) \end{pmatrix}.$$

将 $\mathcal{B}$ 按如下方式拆分为 r 个互不相交的向量组 $\mathcal{B}_1, \cdots, \mathcal{B}_r$: 其中第一组 $\mathcal{B}_1$ 由 $\mathcal{B}$ 中的前 m_1 个向量组成, 第二组 $\mathcal{B}_2$ 由 $\mathcal{B}$ 中的第 m_1+1 至第 m_1+m_2 个向量组成, 如此类推 (这样每个 $\mathcal{B}_j$ 恰有 m_j 个向量, 且按照 $(\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_r)$ 这样依次排序后与有序基 $\mathcal{B}$ 吻合).

1. 若令 $C_j = \mathrm{span}(\mathcal{B}_j)$ 为向量组 $\mathcal{B}_j$ 生成的子空间, 则 C_j 是 $\mathscr{N}$ 的 m_j 维不变子空间, $\mathcal{B}_j$ 是 C_j 的一组基, $\mathscr{N}|_{C_j}$ 在 $\mathcal{B}_j$ 这组基下的矩阵是 Jordan 块 $J_{m_j}(0)$, 而且有如下直和分解式:

   (5.1.20.1) $$W = C_1 \oplus \cdots \oplus C_r\,.$$

2. 对于每个 $j \in [\![1, r]\!]$, 将向量组 $\mathcal{B}_j$ 中的第一个向量记为 w_{j1}, 则

   (5.1.20.2) $$w_{11}, w_{21}, \cdots, w_{r1}$$

   构成 $\mathrm{Ker}(\mathscr{N})$ 的一组基.

   因此, r 等于 $\mathrm{Ker}(\mathscr{N})$ (也就是特征子空间 $E(0, \mathscr{N})$) 的维数. 特别地, r 由 $\mathscr{N}$ 完全确定, 而不依赖于 Jordan 基 $\mathcal{B}$ 的选取.

3. 对于每个 $j \in [\![1, r]\!]$, 向量组 $\mathcal{B}_j$ 中的向量按照顺序可以写成如下形式:

   (5.1.20.3) $$\mathscr{N}^{m_j-1}v_j\,,\, \mathscr{N}^{m_j-2}v_j\,,\, \cdots, \mathscr{N}v_j\,,\, v_j\,.$$

   而且, 向量组

   (5.1.20.4) $$\mathscr{N}^{m_1-1}v_1, \cdots, \mathscr{N}v_1\,;\, \mathscr{N}^{m_2-1}v_2, \cdots, \mathscr{N}v_2\,; \cdots;\, \mathscr{N}^{m_r-1}v_r\,, \cdots, \mathscr{N}v_r$$

   构成 $\mathrm{Im}(\mathscr{N})$ 的一组基. (也就是说, 将每个组 $\mathcal{B}_j$ 内的最后一个向量从 $\mathcal{B}$ 中去掉, 余下的向量组是 $\mathrm{Im}(\mathscr{N})$ 的一组基.)

注意, 命题 5.1.20 中每个子空间 C_j 具有形如 (5.1.20.3) 式的一组基. 这组基的特点是可以从一个非零向量 v 出发, 反复用线性变换 $\mathscr{N}$ 作用得到新的向量并添加到向量组中, 直到找出向量序列 $v, \mathscr{N}v, \mathscr{N}^2v, \cdots$ 中的一个极大线性无关组, 这样就得到 C_j 的一组基. 按照我们之后将要给出的定义 (参见定义 5.2.29), 子空间 C_j 是 $\mathscr{N}$ 的一个循环不变子空间, 而 (5.1.20.3) 式给出的是 C_j 的一组循环基.

命题 5.1.20 的证明. 将 $\mathcal{B}_j$ 中的向量组依次记为

$$w_{j1}\,,\, w_{j2}\,, \cdots,\, w_{jm_j}\,.$$

因为这个向量组是线性无关组 $\mathcal{B}$ 的一部分, 所以子空间

$$C_j = \operatorname{span}(\mathcal{B}_j) = \operatorname{span}\left(w_{j1}, \cdots, w_{jm_j}\right)$$

的维数是 m_j, 而 $\mathcal{B}_j$ 是它的一组基. 由于 $\mathcal{B}_1, \cdots, \mathcal{B}_r$ 合并起来就是 W 的基 $\mathcal{B}$, 所以 (5.1.20.1) 中的分解式成立. 而且, 根据矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{N})$ 的定义和相关的题设条件可知

$$\begin{aligned} \mathscr{N}w_{j1} &= 0\,,\\ \mathscr{N}w_{j2} &= w_{j1}\,,\\ &\cdots\cdots\cdots\\ \mathscr{N}w_{jm_j} &= w_{j,\,m_j-1}\,. \end{aligned} \tag{5.1.20.5}$$

这说明 $C_j = \operatorname{span}(w_{j1}, \cdots, w_{jm_j})$ 是 $\mathscr{N}$ 的不变子空间, 且 $\mathscr{N}|_{C_j}$ 在 $\mathcal{B}_j = (w_{j1}, \cdots, w_{jm_j})$ 这组有序基下的矩阵就是

$$J_{m_j}(0) = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} \in \mathbf{M}_{m_j}(K)\,.$$

至此命题中的论断 (1) 已被完全证明.

(5.1.20.5) 式的第一行还说明了 (5.1.20.2) 式中的向量组包含于 $\operatorname{Ker}(\mathscr{N})$. 因为是 $\mathcal{B}$ 的一部分, 所以该向量组线性无关. 为了完成 (2) 的证明, 只需要再证明: $\operatorname{Ker}(\mathscr{N})$ 中的每个向量都可以写成 $w_{11}, w_{21}, \cdots, w_{r1}$ 的线性组合.

为此, 任取 $v \in \operatorname{Ker}(\mathscr{N})$. 我们先利用 W 的有序基 $\mathcal{B}$ 将 v 写成如下形式:

$$\begin{aligned} v &= \sum_{j=1}^{r}\sum_{k=1}^{m_j} a_{jk}w_{jk}\\ &= (a_{11}w_{11} + a_{12}w_{12}\cdots + a_{1m_1}w_{1m_1})\\ &\quad + \cdots\cdots\cdots\\ &\quad + (a_{r1}w_{r1} + a_{r2}w_{r2}\cdots + a_{rm_r}w_{rm_r}) \end{aligned} \tag{5.1.20.6}$$

其中 $a_{jk} \in K$. 将 $\mathscr{N}$ 作用于 (5.1.20.6), 由 $v \in \operatorname{Ker}(\mathscr{N})$ 这个假设以及 (5.1.20.5) 式列出的结果我们得到

$$\begin{aligned} 0 = \mathscr{N}v &= (a_{11}\cdot 0 + a_{12}w_{11}\cdots + a_{1m_1}w_{1,\,m_1-1})\\ &\quad + \cdots\cdots\cdots\\ &\quad + (a_{r1}\cdot 0 + a_{r2}w_{r1}\cdots + a_{rm_r}w_{r,\,m_r-1})\\ &= (a_{12}w_{11}\cdots + a_{1m_1}w_{1,\,m_1-1})\\ &\quad + \cdots\cdots\cdots\\ &\quad + (a_{r2}w_{r1}\cdots + a_{rm_r}w_{r,\,m_r-1})\,. \end{aligned}$$

由此可以推出

$$0 = a_{12} = \cdots = a_{1,\,m_1} = \cdots\cdots = a_{r2} = \cdots = a_{rm_r}\,,$$

因为向量组

$$w_{11}, \cdots, w_{1,\,m_1-1}\,;\ \cdots\cdots\,;\ w_{r1}, \cdots, w_{r,\,m_r-1}$$

(作为 $\mathcal{B}$ 的一部分) 必然线性无关.

现在回到 (5.1.20.6) 式即得

$$v = a_{11}w_{11} + a_{21}w_{21} + \cdots + a_{r1}w_{r1}\,.$$

这就证明了所期待的结论: $\operatorname{Ker}(\mathscr{N})$ 中的任意向量 v 都是 (5.1.20.2) 中向量组的线性组合.

最后再来证明命题中的断言 (3).

首先, 若令 $v_j = w_{jm_j}$ (即向量组 $\mathcal{B}_j$ 中的最后一个向量), 则 (5.1.20.5) 式恰好说明 $\mathcal{B}_j = (w_{j1}, \cdots, w_{jm_j})$ 可以写成 (5.1.20.3) 式中的样子. 于是, (5.1.20.4) 式中的向量组就是 $\mathcal{B}$ 的一部分, 因而是线性无关组. 另外, (5.1.20.4) 式中的向量显然都属于 $\operatorname{Im}(\mathscr{N})$. 根据前面已经证明的结论 $r = \operatorname{Ker}(\mathscr{N})$, 结合秩–零化度定理可知 (5.1.20.4) 式中向量的数目恰好等于 $\dim W - r = \dim \operatorname{Im}(\mathscr{N})$. 因此 (5.1.20.4) 必然是 $\operatorname{Im}(\mathscr{N})$ 的一组基.

命题至此证毕. □

命题 5.1.20 的陈述略显冗长, 但其中包含很多简单而有用的信息. 以下歌诀希望能对读者理解记忆这些内容有所帮助.

幂零变换若当基,
按块分组排整齐.
每组第一充核内,
末位淘汰生像集.
直和各项皆不变,
基底循环隐玄机.
计算过程虽辛苦,
熟能生巧化神奇.

接下来我们再来考察 Jordan 标准形的一些更精细结构信息.

命题 5.1.21: 设 $\mathscr{N}$ 是有限维向量空间 W 上的幂零变换. 假设 J 是 $\mathscr{N}$ 的一个 Jordan 标准形. 设 r 为 J 中 Jordan 块的数量, M 和 m 分别为 J 中 Jordan 块阶数的最大值和最小值.

1. M 等于 $\mathscr{N}$ 的幂零阶.

2. 对于任意 $j \in [\![1,\, m]\!]$, 均有等式 $\operatorname{rank}(\mathscr{N}^j) = \dim W - j \dim \operatorname{Ker}(\mathscr{N})$ 成立.

   (注意: 根据秩–零化度定理, 等式 $\operatorname{rank}(\mathscr{N}^j) = \dim W - j \dim \operatorname{Ker}(\mathscr{N})$ 与 $\dim \operatorname{Ker}(\mathscr{N}^j) = j \dim \operatorname{Ker}(\mathscr{N})$ 等价.)

3. $m+1$ 是满足不等式 $\operatorname{rank}(\mathscr{N}^j) \neq \dim W - j \dim \operatorname{Ker}(\mathscr{N})$ 的最小正整数 j.

   (根据前一条结论, $m+1$ 也是满足 $\dim \operatorname{Ker}(\mathscr{N}^j) \neq j \dim \operatorname{Ker}(\mathscr{N})$ 的最小正整数 j.)

   事实上, $\operatorname{rank}(\mathscr{N}^{m+1}) > \dim W - (m+1) \dim \operatorname{Ker}(\mathscr{N})$, $\dim \operatorname{Ker}(\mathscr{N}^{m+1}) < (m+1) \dim \operatorname{Ker}(\mathscr{N})$.

4. 对于每个 $k \in [\![m,\, M]\!]$, 若记 s_k 为 J 中阶为 k 的 Jordan 块数量, 则

$$\begin{aligned} s_k &= \big(\operatorname{rank}(\mathscr{N}^{k-1}) - \operatorname{rank}(\mathscr{N}^k)\big) - \big(\operatorname{rank}(\mathscr{N}^k) - \operatorname{rank}(\mathscr{N}^{k+1})\big) \\ &= \operatorname{rank}(\mathscr{N}^{k-1}) + \operatorname{rank}(\mathscr{N}^{k+1}) - 2\operatorname{rank}(\mathscr{N}^k)\,. \end{aligned}$$

证明. (1) 由思考题 5.9 可得.

为证明 (2)–(4), 设 $\mathcal{B}$ 为 J 对应的 Jordan 基, 即 $J = \mathcal{M}_{\mathcal{B}}(\mathscr{N})$. 设 $m_1, \cdots, m_r$ 为各 Jordan 块的阶数. 不妨设

$$m_1 \le m_2 \le \cdots \le m_r\,.$$

根据命题 5.1.20 所述, 可以将 $\mathcal{B}$ 中的向量按顺序写成如下 (分组的) 形式:

(5.1.21.1) $$\mathscr{N}^{m_1-1}v_1\,, \cdots, \mathscr{N}v_1,\, v_1\,; \cdots\cdots\cdots\,;\, \mathscr{N}^{m_r-1}v_r\,, \cdots, \mathscr{N}v_r,\, v_r$$

其中向量 v_i, $1 \le i \le r$ 满足

(5.1.21.2) $$\mathscr{N}^{m_i}v_i = 0\,.$$

对任意正整数 j, 将 $\mathscr{N}^j$ 作用于 (5.1.21.1) 中的向量组得到

(5.1.21.3) $$\mathscr{N}^{m_1+j-1}v_1\,, \cdots,\, \mathscr{N}^{j+1}v_1,\, \mathscr{N}^jv_1\,; \cdots\cdots\cdots\,;\, \mathscr{N}^{m_r+j-1}v_r\,, \cdots,\, \mathscr{N}^{j+1}v_r,\, \mathscr{N}^jv_r\,.$$

这是 $\mathrm{Im}(\mathscr{N}^j)$ 的一个生成组. 根据 (5.1.21.2) 式可知, (5.1.21.3) 中的非零向量一定包含在下面的向量组中:

(5.1.21.4) $$\mathscr{N}^{m_1-1}v_1\,, \cdots, \mathscr{N}^jv_1\,; \cdots\cdots\cdots\,;\, \mathscr{N}^{m_r-1}v_r\,, \cdots,\, \mathscr{N}^jv_r\,.$$

这仍是 $\mathrm{Im}(\mathscr{N}^j)$ 的一个生成组. (这里, 如果 $j \ge m_i$, 则认为上式中出现的向量组 $\mathscr{N}^{m_i-1}v_i\,, \cdots,\, \mathscr{N}^jv_i$ 是空向量组.) 在 (5.1.21.4) 中可能还有一些向量是零向量. 不过, 再将这些零向量去掉之后, 剩余的向量组是有序基 $\mathcal{B}$ 的一部分 (参见 (5.1.21.1)), 从而是线性无关的, 由此可以得到 $\mathrm{Im}(\mathscr{N}^j)$ 的一组基.

当 $j \le m = \min_{1\le i\le r}\{m_i\}$ 时, (5.1.21.4) 恰好含有 $\dim W - jr$ 个向量, 而且都是非零向量 (验证 $j = m$ 的情况可能需要更细心一些, 但结论仍然是对的). 而 $r = \dim\mathrm{Ker}(\mathscr{N})$ (参见命题 5.1.20 (2)), 所以

$$\mathrm{rank}(\mathscr{N}^j) = \dim\mathrm{Im}(\mathscr{N}^j) = \dim W - j\,\dim\mathrm{Ker}(\mathscr{N})\,.$$

这样就证明了命题中的结论 (2).

当 $j = m+1$ 时, (5.1.21.4) 式中所含的非零向量数目大于 $\dim W - jr$. 所以此时不等式

$$\mathrm{rank}(\mathscr{N}^j) > \dim W - j\,\dim\mathrm{Ker}(\mathscr{N})$$

成立. 这就证明了 (3).

为了证明 (4), 固定 $k \in [\![m,\, M]\!]$, 并记 $s = s_k$. 不妨设

$$m_1 \le \cdots \le m_t < k = m_{t+1} = \cdots = m_{t+s} < m_{t+s+1} \le \cdots \le m_r\,,$$

即, 阶为 k 的 Jordan 块恰好是第 $t+1$ 到第 $t+s$ 个 Jordan 块. (若 $s = 0$, 上式就是指 k 介于 m_t 和 m_{t+1} 之间, 并且和 m_t, m_{t+1} 都不相等.)

分别在 (5.1.21.4) 式中令 $j = k-1$ 和 $j = k$, 考虑其中的非零向量可知

$$\begin{aligned}
&\mathscr{N}^{k-1}v_{t+1},\, \mathscr{N}^{k-1}v_{t+2},\, \cdots,\, \mathscr{N}^{k-1}v_{t+s}\,,\\
&\mathscr{N}^{m_{t+s+1}-1}v_{t+s+1}\,, \cdots, \mathscr{N}^{k+1}v_{t+s+1},\, \mathscr{N}^kv_{t+s+1},\, \mathscr{N}^{k-1}v_{t+s+1}\,;\\
&\cdots\cdots\cdots\cdots\cdots\\
&\mathscr{N}^{m_r-1}v_r\,, \cdots, \mathscr{N}^{k+1}v_r\,,\, \mathscr{N}^kv_r,\, \mathscr{N}^{k-1}v_r
\end{aligned}$$

是 $\mathrm{Im}(\mathscr{N}^{k-1})$ 的一组基; 而

$$\begin{aligned}
&\mathscr{N}^{m_{t+s+1}-1}v_{t+s+1}\,, \cdots, \mathscr{N}^{k+1}v_{t+s+1},\, \mathscr{N}^kv_{t+s+1}\,;\\
&\cdots\cdots\cdots\cdots\cdots\\
&\mathscr{N}^{m_r-1}v_r\,, \cdots, \mathscr{N}^{k+1}\,,\, \mathscr{N}^kv_r
\end{aligned}$$

是 $\mathrm{Im}(\mathscr{N}^k)$ 的一组基. 所以,

$$\mathrm{rank}(\mathscr{N}^{k-1})-\mathrm{rank}(\mathscr{N}^k)=r-t\,.$$

同理可知,

$$\begin{aligned}&\mathscr{N}^{m_{t+s+1}-1}v_{t+s+1}\,,\cdots,\mathscr{N}^{k+1}v_{t+s+1}\,;\\&\cdots\cdots\cdots\cdots\cdots\\&\mathscr{N}^{m_r-1}v_r\,,\cdots,\mathscr{N}^{k+1}v_r\end{aligned}$$

是 $\mathrm{Im}(\mathscr{N}^{k+1})$ 的一组基. (这里, 如果 $k+1=m_i$, 则认为向量组 $\mathscr{N}^{m_i-1}v_i,\cdots,\mathscr{N}^{k+1}v_i$ 是空向量组.) 所以,

$$\mathrm{rank}(\mathscr{N}^k)-\mathrm{rank}(\mathscr{N}^{k+1})=r-(t+s)\,.$$

于是,

$$\begin{aligned}&\big(\mathrm{rank}(\mathscr{N}^{k-1})-\mathrm{rank}(\mathscr{N}^k)\big)-\big(\mathrm{rank}(\mathscr{N}^k)-\mathrm{rank}(\mathscr{N}^{k+1})\big)\\&=(r-t)-(r-(t+s))=s\,.\end{aligned}$$

至此命题完全得证. □

(5.1.22) 设 $\mathscr{N}$ 是有限维向量空间 W 上的幂零变换. 综合命题 5.1.20 和 5.1.21 的结论可以知道: 对于 $\mathscr{N}$ 的任意一个 Jordan 标准形 J (我们现在是假设 J 存在, 不过随后我们将会证明它的确存在), 以下数据完全由 $\mathscr{N}$ 本身决定:

- J 所含的 Jordan 块数量;
- J 中 Jordan 块的最大阶数 M;
- J 中 Jordan 块的最小阶数 m;
- 对于区间 $[\![m,\,M]\!]$ 内的任意正整数 k, J 中的 k 阶 Jordan 块数目 s_k.

这说明, J 中出现的 Jordan 块数量和每个 Jordan 块的大小都完全由 $\mathscr{N}$ 确定. 如果不考虑 Jordan 块的排列顺序, 以上这些信息就完全确定了幂零 Jordan 形矩阵 J 的形状. 所以, 如果不计 Jordan 块的排列顺序, 那么 $\mathscr{N}$ 的 Jordan 标准形只有一个. 换言之, 对于 $\mathscr{N}$ 的任意两个 Jordan 标准形 J_1 和 J_2, 一定可以仅通过调整 J_1 中 Jordan 块的顺序把 J_1 变成 J_2.

事实上, 对于幂零的 Jordan 块来说, 知道了其大小就等于完全确定了其形状. 所以命题 5.1.20 和 5.1.21 已经告诉了我们如何计算出 $\mathscr{N}$ 的 Jordan 标准形是什么样子. 只不过, 现在我们还没有解释怎样找到与 Jordan 标准形对应的 Jordan 基. 或者说, 从 $\mathscr{N}$ 在某一组基下的矩阵 N 出发, 我们已经知道 N 可以相似到哪个 Jordan 形矩阵 J, 但还没有解释如何找到可逆矩阵 P 使 $P^{-1}NP=J$. 这个任务我们留到 § 5.1.3 小节来完成. ■

现在我们证明幂零变换的 Jordan 标准形总是存在的. 这个证明过程读者不必强记, 只要在阅读时能够理解无障碍即可.

定理 5.1.23: 设 W 是有限维 K-向量空间, $\dim W\geq 1$, $\mathscr{N}$ 为 W 上的幂零变换.

则 W 中一定存在 $\mathscr{N}$ 的一组 Jordan 基, 即, 一组具有如下 (分组排列) 形式的有序基 (参见命题 5.1.20):

$$\mathscr{N}^{m_1-1}v_1,\cdots,\mathscr{N}v_1,\,v_1\,;\,\cdots\cdots\,;\,\mathscr{N}^{m_r-1}v_r\,,\cdots,\,\mathscr{N}v_r\,,\,v_r\,,\tag{5.1.23.1}$$

其中 $r,\,m_1,\cdots,m_r\in\mathbb{N}^*$ 且 $\mathscr{N}^{m_1}v_1=\cdots=\mathscr{N}^{m_r}v_r=0$.

证明. 我们使用对 $\dim W$ 的归纳法.

若 $\dim W = 1$, 则幂零变换 $\mathscr{N}$ 满足 $\mathscr{N} = \mathscr{N}^{\dim W} = 0$. 此时 W 的任意一组基都是 $\mathscr{N} = 0$ 的 Jordan 基.

以下假设 $\dim W > 1$, 来证 W 上的幂零变换 $\mathscr{N}$ 一定有 Jordan 基. 显然, 若 $\mathscr{N} = 0$, 则任取 W 的一组基即可 (这相当于在 (5.1.23.1) 中取 $r = \dim W$, $m_1 = \cdots = m_r = 1$). 所以, 我们可以假设 $\mathscr{N} \neq 0$, 即 $\dim \operatorname{Im}(\mathscr{N}) \geq 1$. 另一方面, 由于幂零变换 $\mathscr{N}$ 不可能是单射 (请读者回答为什么), 故由秩–零化度定理可知 $\operatorname{Im}(\mathscr{N})$ 的维数严格小于 $\dim W$. 我们还知道 $\operatorname{Im}(\mathscr{N})$ 是 $\mathscr{N}$ 的不变子空间, 因而可以考虑限制映射 $\mathscr{N}' := \mathscr{N}|_{\operatorname{Im}(\mathscr{N})}$. 将归纳假设应用于幂零变换 $\mathscr{N}'$ 可知: $\operatorname{Im}(\mathscr{N})$ 有一组 $\mathscr{N}'$ 的 Jordan 基:

$$\mathscr{N}^{e_1-1}x_1, \cdots, \mathscr{N}x_1, x_1\,;\ \cdots\cdots\,;\ \mathscr{N}^{e_t-1}x_t\,, \cdots, \mathscr{N}x_t, x_t\,. \tag{5.1.23.2}$$

这里的向量 x_j 和正整数 e_j 满足

$$\mathscr{N}^{e_j}x_j = 0\,, \quad j \in [\![1, t]\!]\,. \tag{5.1.23.3}$$

接下来我们希望通过向量组 (5.1.23.2) 来找到期待的向量组 (5.1.23.1). 正如我们在命题 5.1.20 (3) 中所见, (5.1.23.2) 应该可以取为 (5.1.23.1) 式中的向量组经过 $\mathscr{N}$ 作用之后剩余下来的非零向量. 因此, 如果 (5.1.23.1) 中某个 v_i 满足 $\mathscr{N}v_i \neq 0$ (即 $v_i \notin \operatorname{Ker}(\mathscr{N})$), 那么 $\mathscr{N}v_i$ 应该是 (5.1.23.2) 式中的某个 x_j. 所以这样的 v_i 可以通过求 x_j (在 $\mathscr{N}$ 作用下) 的原像来找出. 也就是说, 对于每个 $j \in [\![1, t]\!]$, 我们希望取出一个向量 $v_j \in W$ 使 $x_j = \mathscr{N}v_j$. 由于 $x_i \in \operatorname{Im}(\mathscr{N})$, 这样的向量 v_j 一定存在.

将这样找到的向量 $v_1, \cdots, v_t$ 添加到 (5.1.23.2) 中得到一个新的向量组

$$\mathscr{N}^{e_1}v_1, \cdots, \mathscr{N}v_1, v_1\,;\ \cdots\cdots\,;\ \mathscr{N}^{e_t}v_t\,, \cdots, \mathscr{N}v_t, v_t\,. \tag{5.1.23.4}$$

注意, 根据 x_j 和 v_j 之间的关系我们有

$$\text{对于任何 } k \in \mathbb{N}^*\,, \quad \mathscr{N}^{k-1}x_j = \mathscr{N}^k v_j\,.$$

(5.1.23.4) 和 (5.1.23.1) 的差距是可能还缺少一些满足 $\mathscr{N}v_{t+1} = \cdots = \mathscr{N}v_r = 0$ 的向量 $v_{t+1}, \cdots, v_r$. 这些向量添加到 (5.1.23.4) 之后应该是 W 的一组基. 为了达成最终的目标, 我们现在来依次完成以下两个步骤:

第一步: 先验证 (5.1.23.4) 是个线性无关组, 从而可以扩充为 W 的一组基

$$\mathscr{N}^{e_1}v_1, \cdots, \mathscr{N}v_1, v_1\,;\ \cdots\,;\ \mathscr{N}^{e_t}v_t, \cdots, \mathscr{N}v_t\,, v_t\,;\ w_1, \cdots, w_p\,. \tag{5.1.23.5}$$

第二步: 将 (5.1.23.5) 中的向量组 $w_1, \cdots, , w_p$ 作适当的调整, 换成新的一组向量 $v_{t+1}, \cdots, v_{t+p}$, 使得 $v_{t+1}, \cdots, v_{t+p}$ 均属于 $\operatorname{Ker}(\mathscr{N})$, 并且

$$\mathscr{N}^{e_1}v_1, \cdots, \mathscr{N}v_1, v_1\,;\ \cdots\,;\ \mathscr{N}^{e_t}v_t, \cdots, \mathscr{N}v_t\,, v_t\,;\ v_{t+1}, \cdots, v_{t+p} \tag{5.1.23.6}$$

仍然构成 W 的一组基. 此时, 令

$$r = t + p\,,\ m_1 = e_1 + 1, \cdots, m_t = e_t + 1\,,\ m_{t+1} = \cdots = m_r = 1$$

即可将 (5.1.23.6) 改写成 (5.1.23.1) 的形式, 从而完成定理的证明.

现在我们证明 (5.1.23.4) 线性无关. 假设存在常数 a_{ij} 满足

$$\begin{aligned} 0 = {} & a_{e_1,1}\mathscr{N}^{e_1}v_1 + a_{e_1-1,1}\mathscr{N}^{e_1-1}v_1 + \cdots + a_{11}\mathscr{N}v_1 + a_{01}v_1 \\ & + \cdots\cdots\cdots\cdots \\ & + a_{e_t,t}\mathscr{N}^{e_t}v_t + a_{e_t-1,t}\mathscr{N}^{e_t-1}v_t + \cdots + a_{1t}\mathscr{N}v_t + a_{0t}v_t\,. \end{aligned} \tag{5.1.23.7}$$

将 $\mathscr{N}$ 作用于上式, 再注意到 $\mathscr{N}^{e_j+1}v_j = \mathscr{N}^{e_j}x_j = 0$ (参见 (5.1.23.3) 式) 可以得到

$$\begin{aligned}0 &= a_{e_1-1,\,1}\mathscr{N}^{e_1}v_1 + \cdots + a_{11}\mathscr{N}^2 v_1 + a_{01}\mathscr{N}v_1\\ &\quad + \cdots\cdots\cdots\cdots\\ &\quad + a_{e_t-1,\,t}\mathscr{N}^{e_t}v_t + \cdots + a_{1t}\mathscr{N}^2 v_t + a_{0t}\mathscr{N}v_t\\ &\quad = a_{e_1-1,\,1}\mathscr{N}^{e_1-1}x_1 + \cdots + a_{11}\mathscr{N}x_1 + a_{01}x_1\\ &\quad + \cdots\cdots\cdots\cdots\\ &\quad + a_{e_t-1,\,t}\mathscr{N}^{e_t-1}x_t + \cdots + a_{1t}\mathscr{N}x_t + a_{0t}x_t\,.\end{aligned}$$

观察最后这个线性组合, 再利用 (5.1.23.2) 的线性无关性, 我们得出

$$\begin{aligned}&a_{e_1-1,\,1} = \cdots = a_{11} = a_{01} = 0\,,\\ &\cdots\cdots\cdots\cdots\\ &a_{e_t-1,\,1} = \cdots = a_{1t} = a_{0t} = 0\,.\end{aligned}$$

由此返回 (5.1.23.7) 式可得

$$\begin{aligned}0 &= a_{e_1,\,1}\mathscr{N}^{e_1}v_1 + a_{e_2,\,2}\mathscr{N}^{e_2}v_2 + \cdots + a_{e_t,\,t}\mathscr{N}^{e_t}v_t\\ &= a_{e_1,\,1}\mathscr{N}^{e_1-1}x_1 + a_{e_2,\,2}\mathscr{N}^{e_2-1}x_2 + \cdots + a_{e_t,\,t}\mathscr{N}^{e_t-1}x_t\,.\end{aligned}$$

由于向量组 (5.1.23.2) 的子向量组也都线性无关, 所以上式表明 $a_{e_1,\,1} = a_{e_2,\,2} = \cdots = a_{e_t,\,t} = 0$. 于是 (5.1.23.7) 式中的所有常数系数等于 0. 这就完成了之前列举的第一个步骤. 因而可以通过扩充得到 W 的一组基如 (5.1.23.5) 中形式.

最后, 我们要将 (5.1.23.5) 中的向量 $w_1, \cdots, w_p$ 适当调整为 $\mathrm{Ker}(\mathscr{N})$ 中的一些向量 $v_{t+1}, \cdots, v_{t+p}$ 使 (5.1.23.6) 构成 W 的基. 因为 (5.1.23.5) 和 (5.1.23.6) 所含的向量数目相同, 所以要使后者也成为 W 的基, 只需要确保 $w_1, \cdots, w_p$ 都可以被向量组 (5.1.23.6) 线性表出. 而这一点成立的充分条件是: 每个 $y_j := w_j - v_{t+j}$ 都属于 (5.1.23.4) 生成的子空间, 因为 (5.1.23.4) 是 (5.1.23.6) 的一部分. 当然, 我们不能忘记 v_{t+j}, $1 \le j \le p$ 还需要来自 $\mathrm{Ker}(\mathscr{N})$.

假如这样的 v_{t+j} 存在, 那么 $\mathscr{N}(y_j) = \mathscr{N}(w_j - v_{t+j}) = \mathscr{N}(w_j) - 0 = \mathscr{N}(w_j)$. 反过来, 如果我们可以先找到

$$y_1, \cdots, y_p \;\in\; \mathrm{span}((5.1.23.4)) \text{ 使得 } \mathscr{N}(y_j) = \mathscr{N}(w_j),\;\; j = 1, \cdots, p\,,$$

那么可以定义 $v_{t+j} := w_j - y_j$ 使以上所有要求被满足.

要找出所求的向量 y_j, 需要注意 (5.1.23.2) 和 (5.1.23.4) 两个向量组的联系. 事实上, 根据 $v_1, \cdots, v_t$ 这些向量的选取, (5.1.23.2) 中每个向量都可以通过 $\mathscr{N}$ 作用于 (5.1.23.4) 中一个向量得到. 这说明

$$\mathrm{span}((5.1.23.2)) \;\subseteq\; \mathscr{N}\left(\mathrm{span}((5.1.23.4))\right)\,.$$

再注意到 (5.1.23.2) 式给出 $\mathrm{Im}(\mathscr{N})$ 的一组基, 所以上式左边等于 $\mathrm{Im}(\mathscr{N})$. 也就是说, $\mathrm{Im}(\mathscr{N})$ 中每个向量都可以写成 $\mathscr{N}(y)$ 的形式, 其中 $y \in \mathrm{span}((5.1.23.4))$. 特别地, 每个 $\mathscr{N}(w_j)$ 都可以写成 $\mathscr{N}(y_j)$ 的形式, 其中 $y_j \in \mathrm{span}((5.1.23.4))$. 这正是欲证的结论.

至此, 定理的证明终于解释完毕. □

5.1.3 幂零变换的 Jordan 基计算

在这一小节中, 总设 $\mathscr{N}$ 是有限维 K-向量空间 W 上的幂零变换.

现在我们讨论将幂零变换 $\mathscr{N}$ 化为 Jordan 标准形的具体计算方法. 虽然 Jordan 标准形的矩阵中各个 Jordan 块的大小可能出现各种排序, 但这里为了方便, 在可以自行选择的情况下, 我们一般采取从小到大的顺序排列 Jordan 标准形中的 Jordan 块.

(5.1.24) 我们先回忆一下如何计算出 $\mathscr{N}$ 的一个 Jordan 标准形 J. 根据前面命题 5.1.20 和 5.1.21 中提供的方法, 我们可以依次执行以下操作:

1. 求出 $\operatorname{Ker}(\mathscr{N})$ 并算出其维数 r.

   这样得到的 r 是 Jordan 标准形 J 所含的 Jordan 块个数.

2. 计算 $\mathscr{N}$ 的幂 $\mathscr{N}^2$, $\mathscr{N}^3$ 等. 由此找出 $\mathscr{N}$ 的幂零阶 M.

   这样得到的 M 是 Jordan 标准形 J 中 Jordan 块的最大阶数.

3. 依次对 $j=1,2,\cdots$ 等检验不等式 $\dim\operatorname{Ker}(\mathscr{N}^j)\neq j\dim\operatorname{Ker}(\mathscr{N})$ 是否成立. 若将第一个满足此不等式的 j 写成 $m+1$ 的形式, 则 m 是 Jordan 标准形 J 中 Jordan 块的最小阶数.

   当然, 在这里也可以用 $\dim\operatorname{Im}(\mathscr{N}^j)\neq\dim W-j\dim\operatorname{Ker}(\mathscr{N})$ 代替 $\dim\operatorname{Ker}(\mathscr{N}^j)\neq j\dim\operatorname{Ker}(\mathscr{N})$ 来考虑.

4. 对于每个 $k\in[\![m,M]\!]$, 计算

$$\begin{aligned}s_k&=\operatorname{rank}(\mathscr{N}^{k-1})+\operatorname{rank}(\mathscr{N}^{k+1})-2\operatorname{rank}(\mathscr{N}^k)\\&=2\dim\operatorname{Ker}(\mathscr{N}^k)-\dim\operatorname{Ker}(\mathscr{N}^{k-1})-\dim\operatorname{Ker}(\mathscr{N}^{k+1})\,.\end{aligned}$$

   这样算出的 s_k 是 Jordan 标准形 J 中 k 阶 Jordan 块的个数.

经过以上步骤我们便掌握了足够的信息可以将 Jordan 形矩阵 J 具体写出. 例如, 假如以上计算结果是 $r=4$, $M=5$, $m=2$, $s_2=2$, $s_3=0$, $s_4=s_5=1$, 那我们就知道

$$J=\begin{pmatrix}J_2(0)&&&\\&J_2(0)&&\\&&J_4(0)&\\&&&J_5(0)\end{pmatrix}.$$

(按照本小节开头所述, 我们约定 Jordan 块是按从左上到右下逐渐增大的方式排序.)

值得注意的是, 以上总结的方法是一般的方法, 其具体步骤感觉有点多. 然而在实际操作中, 有时候我们不一定要完全拘泥于上述程序, 而是可以通过综合各方面考虑, 适当地简化求解过程. 例如, 通过 $m_1+\cdots+m_r=\dim W$ 这个限制条件就可以将 r, M, m 的许多不可能取值排除掉. 这种思路在空间维数比较低的情况下经常是非常有效的.

例如, 如果是 3 维空间上的幂零变换, 那么只要知道 Jordan 块个数是 2 就可以确定 J 的形状. 即 (按照我们约定的 Jordan 块排序方式) $J=\begin{pmatrix}J_1(0)&\\&J_2(0)\end{pmatrix}=\begin{pmatrix}0&&\\&0&1\\&&0\end{pmatrix}$.

再比如, 在 4 维的情况下, 在知道 Jordan 块个数为 2 的情况下, (不计 Jordan 块排序的话) J 只有两种可能:

$$J=\begin{pmatrix}J_1(0)&\\&J_3(0)\end{pmatrix}=\begin{pmatrix}0&&&\\&0&1&\\&&0&1\\&&&0\end{pmatrix}\quad\text{或者}\quad J=\begin{pmatrix}J_2(0)&\\&J_2(0)\end{pmatrix}=\begin{pmatrix}0&1&&\\&0&&\\&&0&1\\&&&0\end{pmatrix}.$$

而这两种情况只需要考虑 $\mathscr{N}$ 的幂零阶即可进行区分. 由此可见, 在低维的情况下我们常常不需要完全按部就班地去完成之前列出的 4 个步骤, 而是有可能一两个步骤就够了. 希望读者通过实例的演算和分析体会这一点. ■

(5.1.25) 前面一段我们已经阐明了如何求出 Jordan 标准形的矩阵 J. 我们设 $m_1,\cdots,m_r$ 是 J 的各个 Jordan 块阶数, 并假设

$$m_1 \le m_2 \le \cdots \le m_r\,.$$

现在我们来讨论如何求出与 J 对应的一组 Jordan 基 $\mathcal{B}$ (即满足 $\mathcal{M}_{\mathcal{B}}(\mathscr{N}) = J$ 的一组基 $\mathcal{B}$).

根据前一小节的讨论 (命题 5.1.20 和 5.1.21 及其证明过程), 我们知道 $\mathcal{B}$ 具有如下形式:

(5.1.25.1) $$\mathscr{N}^{m_1-1}v_1\,,\cdots,\mathscr{N}v_1,\,v_1\,;\cdots\cdots\cdots\,;\;\mathscr{N}^{m_r-1}v_r\,,\cdots,\mathscr{N}v_r,\,v_r\,.$$

寻求 Jordan 基实际上就是要找出合适的向量组 $v_1,\cdots,v_r$ 使得上式列出的向量组构成 W 的一组基. 不过问题是, 在实际计算过程中 v_i 这些向量通常不容易直接找到. 我们实际操作的方法是先找出

$$w_1 := \mathscr{N}^{m_1-1}v_1,\, w_2 := \mathscr{N}^{m_2-1}v_2,\cdots,w_r := \mathscr{N}^{m_r-1}v_r$$

这组向量. 这些 w_i 应该满足的条件包括

(5.1.25.2) $$w_1,\,w_2,\cdots,w_r \quad \text{构成 } \mathrm{Ker}(\mathscr{N}) \text{ 的一组基}$$

且

(5.1.25.3) $$\text{对于每个 } i=1,\dots,m\,,\quad \text{均有}\;\; w_i \in \mathrm{Im}(\mathscr{N}^{m_i-1})\,.$$

(这是因为我们希望将 w_i 写成 $w_i = \mathscr{N}^{m_i-1}v_i$ 的形式). 事实上, 它们还需要满足

(5.1.25.4) $$\text{对于每个 } i=1,\dots,m\,,\quad \text{均有}\;\; w_i \notin \mathrm{Im}(\mathscr{N}^{m_i})\,.$$

这是因为, 我们知道 (参见命题 5.1.21 的证明): 如果我们用 $\mathscr{N}^{m_i}$ 作用于向量组 (5.1.25.1) 并且删去其中所得的零向量, 那么剩余的非零向量构成的向量组

(5.1.25.5) $$\mathscr{N}^{m_t-1}v_t,\cdots,\,\mathscr{N}^{m_i}v_t\,;\;\cdots\cdots\cdots\,;\;\mathscr{N}^{m_r-1}v_r,\cdots,\,\mathscr{N}^{m_i}v_r$$

构成 $\mathrm{Im}(\mathscr{N}^{m_i})$ 的一组基. 这里的下标 t 按照

$$m_1 \le \cdots \le m_i = \cdots = m_{t-1} < m_t \le \cdots \le m_r$$

这一要求选取. (注意我们之前已假设 $m_1,\cdots,m_r$ 是从小到大排列的.) 注意到 (5.1.25.5) 是 (5.1.25.1) 的一部分, 而我们希望的 w_i 应该是 $\mathscr{N}^{m_i-1}v_i$, 它不在向量组 (5.1.25.5) 之内, 因而不能写成 (5.1.25.5) 的线性组合. 也就是说, $w_i \notin \mathrm{Im}(\mathscr{N}^{m_i})$. 这就是我们要求 (5.1.25.4) 成立的理由.

将 (5.1.25.2)–(5.1.25.4) 综合起来, 我们可以设计如下递归式的算法来求出一组合适的 w_i:

- 第一步, 在 $\mathrm{Im}(\mathscr{N}^{m_r-1})$ 中任意选取一个非零向量 w_r.

  因为 m_r 是 $\mathscr{N}$ 的幂零阶 (命题 5.1.21 (1)), 所以自然有 $\mathrm{Im}(\mathscr{N}^{m_r}) = 0$, 于是 $w_r \notin \mathrm{Im}(\mathscr{N}^{m_r})$ 且 $\mathscr{N}w_r \in \mathrm{Im}(\mathscr{N}^{m_r}) = 0$, 即 $w_r \in \mathrm{Ker}(\mathscr{N})$.

- 第二步, 在 $\mathrm{Ker}(\mathscr{N})$ 中再任意选取一个向量 w_{r-1} 满足

  $$w_{r-1} \in \mathrm{Im}(\mathscr{N}^{m_{r-1}-1}) \setminus \mathrm{Im}(\mathscr{N}^{m_{r-1}}) \quad\text{且}\quad w_{r-1} \notin \mathrm{span}(w_r)\,.$$

  (后面这个条件是为了使 $w_{r-1},\,w_r$ 构成 $\mathrm{Ker}(\mathscr{N})$ 中的线性无关组, 从而可以逐步扩充为 $\mathrm{Ker}(\mathscr{N})$ 的基.)

当然, 我们执行这一步的前提是 $r>1$, 否则我们的算法在第一步就可以停止了. 只要 $r>1$, 满足以上要求的 w_{r-1} 一定是可以找到的. 因为根据 (5.1.25.1), $\operatorname{Im}(\mathscr{N}^{m_{r-1}-1})$ 应该至少包含 $\mathscr{N}^{m_{r-1}-1}v_{r-1}$ 和 $\mathscr{N}^{m_{r-1}-1}v_r$ 这两个线性无关的向量. 而 $\operatorname{span}(w_r)$ 只是 $\operatorname{Im}(\mathscr{N}^{m_r-1})\subseteq\operatorname{Im}(\mathscr{N}^{m_{r-1}-1})$ 中一个 1 维子空间. 此外, $\operatorname{Im}(\mathscr{N}^{m_{r-1}})$ 也是 $\operatorname{Im}(\mathscr{N}^{m_{r-1}-1})$ 的真子空间. 所以 $\operatorname{Im}(\mathscr{N}^{m_{r-1}-1})\cup\operatorname{span}(w_r)$ 这个并集一定也是 $\operatorname{Im}(\mathscr{N}^{m_{r-1}})$ 的真子集 (请读者思考为什么). 这就说明: 无论之前的 w_r 是如何选取的, 这一步需要的 w_{r-1} 按照上面的做法一定可以找到.

- 假设已经找到了向量 $w_{j+1},\cdots,w_r$, 那么继续找出前一个向量 w_j 要做的是: 在 $\operatorname{Ker}(\mathscr{N})$ 中任意选取一个向量 w_j 满足
$$w_j\in\operatorname{Im}(\mathscr{N}^{m_j-1})\setminus\operatorname{Im}(\mathscr{N}^{m_j})\quad\text{且}\quad w_j\notin\operatorname{span}(w_{j+1},\cdots,w_r)\,.$$
和前一步一样, 这里需要的 w_j (只要 $j+1>1$) 一定能够取到. (这一论断的证明如下: 首先, 通过观察 (5.1.21.1) 式可知 $\dim\operatorname{Im}(\mathscr{N}^{m_j-1})\geq r-j+1>r-j=\dim\operatorname{span}(w_{j+1},\cdots,w_r)$. 另外, $\operatorname{Im}(\mathscr{N}^{m_j})$ 也是 $\operatorname{Im}(\mathscr{N}^{m_j-1})$ 的真子空间. 所以 $\operatorname{span}(w_{j+1},\cdots,w_r)\cup\operatorname{Im}(\mathscr{N}^{m_j})$ 一定是 $\operatorname{Im}(\mathscr{N}^{m_j-1})$ 的真子集.)

以上步骤完成之后可以得到合适的向量组 $w_1,\cdots,w_r$, 然后通过条件 $\mathscr{N}^{m_i-1}v_i=w_i$ 找出 v_i (当 w_i 和 $\mathscr{N}$ 已知, 求 v_i 的过程本质上就是求解一个线性方程组). 这样 (5.1.25.1) 式中的向量组就可以完全确定了, 因此也就得到了所需要的一组 Jordan 基 $\mathcal{B}$.

如果最初我们给定了 $\mathscr{N}$ 在某一组基 $\mathcal{E}$ 下的矩阵 N, $P=P_{\mathcal{E},\mathcal{B}}$ 为 $\mathcal{E}$ 到 $\mathcal{B}$ 的过渡矩阵, 则 $P^{-1}NP=J$. 这样就找到了一种方式将矩阵 N 化为 Jordan 标准形 J.

有些读者可能会好奇: 为什么我们上面介绍的算法是按照先找 w_r, 最后找 w_1 的顺序进行的, 而不是按照从 w_1 到 w_r 的顺序? 理由是如果按照从 w_1 到 w_r 的顺序操作, 那么上面所述的算法有可能行不通. 这主要是因为我们在每一步寻找 w_j 时只施加了最少的必要条件, 按照这些条件任意去选择的话, 未必能够保证后面步骤的可行性[§].

例如, 假设待确定的 Jordan 基形如 $v_1,v_2,\mathscr{N}v_3,v_3$ (即, $r=3$, $m_1=m_2=1$, $m_3=2$). 按照与上面类似的算法, 如果我们先去找 w_1, 那么我们所提的要求仅仅是 $w_1\in\operatorname{Ker}(\mathscr{N})$ 但 $w_1\notin\operatorname{Im}(\mathscr{N}^{m_1})=\operatorname{Im}(\mathscr{N})$. 可是, 按照这个要求, 如果我们任意选取的话, 有可能取到 $w_1=v_1+\mathscr{N}v_3$. 接下来, 第二步寻找 w_2, 我们只要求了 $w_2\in\operatorname{Ker}(\mathscr{N})\setminus\operatorname{Im}(\mathscr{N})$ 且 w_1,w_2 线性无关. 这样的话有可能会取到 $w_2=v_1-\mathscr{N}v_3$. 假如不幸遇到这种情况, 那么在第三步我们就找不到满足需要的 w_3 了. 因为我们一方面要求 $w_3\in\operatorname{Im}(\mathscr{N})=\operatorname{Im}(\mathscr{N}^{m_3-1})$, 另一方面又要求 $w_3\notin\operatorname{span}(w_1,w_2)$. 可是刚才选取的 w_1,w_2 满足 $\operatorname{span}(w_1,w_2)\supseteq\operatorname{span}(\mathscr{N}v_3)=\operatorname{Im}(\mathscr{N})$. 所以按照上面的算法, w_3 在前面两个向量 w_1,w_2 选择不慎的情况下就找不到了. ■

我们上面总结的是一套理论上适用于一般情况的方法. 不过, 在处理具体例子时, 我们可以通过实际的计算结果进行合理的观察和猜测, 这样有时可以通过更简便的方法找到所需的 Jordan 基. 比如, 在一些实例中, 也许我们不必按照 (5.1.25) 一段所说的算法来先找 w_r, 最后找 w_1. 在比较顺利的情况下 (正如我们后面遇到的不少例子), 也许我们通过观察和综合分析, 按照从 w_1 到 w_r 的顺序 (甚至在 r 很小时, 可能可以不分顺序直接同时考虑 $w_1,\cdots,w_r$) 也能适当地选出这些向量.

下面我们举出一些计算实例, 希望读者能够在具体例子中理解掌握我们介绍的方法.

例 5.1.26. 考虑如下线性变换
$$\mathscr{N}\;:\;K^3\longrightarrow K^3\,;\quad\begin{pmatrix}x\\y\\z\end{pmatrix}\longmapsto\begin{pmatrix}2y-z\\0\\0\end{pmatrix}.$$

[§]本书作者之前就犯过这个错误, 后来经过南科大邬龙挺教授的指正才修改过来. 在此郑重对邬教授表示感谢.

简单的计算和观察说明

$$\mathscr{N}^2=0\,,\ \mathrm{Ker}(\mathscr{N})=\mathrm{span}(e_1,\,e_2+2e_3)\,,\ \mathrm{Im}(\mathscr{N})=\mathrm{span}(e_1)\,.$$

所以 $\mathscr{N}$ 的 Jordan 标准形 J 含有两个 Jordan 块, 最大的 Jordan 块是 2 阶. 所以 $J=\begin{pmatrix}0&&\\&0&1\\&&0\end{pmatrix}$.

若要找出对应的 Jordan 基, 就需要找到向量 v_1,v_2 使得 $v_1,\mathscr{N}v_2$ 构成 $\mathrm{Ker}(\mathscr{N})$ 的一组基, 而 $v_1,\mathscr{N}v_2,v_2$ 构成 K^3 的一组基. 按照 (5.1.25) 一段采用的记号, 这里的 $w_1=v_1$ 应该满足 $w_1\in\mathrm{Ker}(\mathscr{N})\setminus\mathrm{Im}(\mathscr{N})$, 而 $w_2=\mathscr{N}v_2$ 应该满足 $w_2\in\mathrm{Im}(\mathscr{N})$ 且 w_1,w_2 构成 $\mathrm{Ker}(\mathscr{N})$ 的一组基.

观察上面关于 $\mathrm{Ker}(\mathscr{N})$ 和 $\mathrm{Im}(\mathscr{N})$ 的计算结果, 不难发现 $w_1=v_1$ 可以取为 $v_1=e_2+2e_3=(0,1,2)^T$, 而 w_2 的一个好的选择是 $w_2=e_1=\mathscr{N}(-e_3)$, 于是 v_2 可取为 $v_2=-e_3$.

综上所述, $\mathcal{B}:=(e_2+2e_3,\,e_1,\,-e_3)$ 是 $\mathscr{N}$ 的一组 Jordan 基, 在这组基下 $\mathscr{N}$ 的矩阵是 Jordan 矩阵 $J=\begin{pmatrix}0&&\\&0&1\\&&0\end{pmatrix}$. 如果要考察 $\mathscr{N}$ 的矩阵, 那么在标准基 $\mathcal{E}=(e_1,e_2,e_3)$ 下 $\mathscr{N}$ 的矩阵是 $N=\begin{pmatrix}0&2&-1\\0&0&0\\0&0&0\end{pmatrix}$. 从 $\mathcal{E}$ 到 $\mathcal{B}$ 的过渡矩阵是 $P=\begin{pmatrix}0&1&0\\1&0&0\\2&0&-1\end{pmatrix}$. 读者可以验证 $P^{-1}NP=J$ 的确成立.

当然, 满足要求的向量 w_i 和 v_i 不是仅有以上一种选择. 例如, 我们还可以将 $w_1=v_1$ 选为 $e_1+e_2+2e_3=(1,1,2)^T$ (注意要在 $\mathrm{Ker}(\mathscr{N})\setminus\mathrm{Im}(\mathscr{N})$ 中选择) 而 w_2 可以选为 $-2e_1$ (注意 $w_2\in\mathrm{Im}(\mathscr{N}^{m_2-1})=\mathrm{Im}(\mathscr{N})$), v_2 可以选为 $-e_2$. 此时得到的 Jordan 基为 $\mathcal{B}'=(e_1+e_2+2e_3,\,-2e_1,\,-e_2)$, $\mathscr{N}$ 在 $\mathcal{B}'$ 下的矩阵仍是 $J=\begin{pmatrix}0&&\\&0&1\\&&0\end{pmatrix}$. 从标准基到 $\mathcal{B}'$ 的过渡矩阵为 $P'=\begin{pmatrix}1&-2&0\\1&0&-1\\2&0&0\end{pmatrix}$. 读者可以验证 $(P')^{-1}NP'=J$ 仍然成立. ■

例 5.1.27. 将幂零矩阵 $N=\begin{pmatrix}0&-1&1&0&-1&0\\&0&0&-2&0&1\\&&0&1&0&0\\&&&0&0&1\\&&&&0&0\\&&&&&0\end{pmatrix}$ 化为 Jordan 标准形.

我们将 N 视为线性变换 $\mathscr{N}:K^6\to K^6;\ X\mapsto NX$ 在标准基 $\mathcal{E}=(e_1,\cdots,e_6)$ 下的矩阵. 计算可得

$$\mathrm{Ker}(\mathscr{N})=\mathrm{span}(e_1,\,e_2+e_3,\,e_3+e_5)\,,\ \ \mathrm{Im}(\mathscr{N})=\mathrm{span}(e_1,\,-2e_2+e_3,\,e_2+e_4)\,,$$

而

$$N^2=\begin{pmatrix}0&0&0&3&0&-1\\&0&0&0&0&-2\\&&0&0&0&1\\&&&0&0&0\\&&&&0&0\\&&&&&0\end{pmatrix},\quad N^3=\begin{pmatrix}0&0&0&0&0&3\\&0&0&0&0&0\\&&0&0&0&0\\&&&0&0&0\\&&&&0&0\\&&&&&0\end{pmatrix},\quad N^4=0\,,$$

$$\mathrm{Im}(\mathscr{N}^2)=\mathrm{span}(e_1,\,-e_1-2e_2+e_3)\,,\quad \mathrm{Im}(\mathscr{N}^3)=\mathrm{span}(e_1)\,.$$

所以 N 的 Jordan 标准形 J 应该由 3 个 Jordan 块构成, 最大的 Jordan 块是 4 阶的 (即 $m_3 = 4$). 由于 $m_1 + m_2 + m_3 = 6$, 所以必然 $m_1 = m_2 = 1$. 所以

$$J = \begin{pmatrix} J_1(0) & & \\ & J_1(0) & \\ & & J_4(0) \end{pmatrix} = \begin{pmatrix} 0 & & & & & \\ & 0 & & & & \\ & & 0 & 1 & 0 & 0 \\ & & & 0 & 1 & 0 \\ & & & & 0 & 1 \\ & & & & & 0 \end{pmatrix}.$$

希望寻找的 Jordan 基形如 $v_1,\, v_2,\, \mathscr{N}^3 v_3,\, \mathscr{N}^2 v_3,\, \mathscr{N} v_3,\, v_3$.

对于 $w_1 = v_1$, $w_2 = v_2$ 和 $w_3 = \mathscr{N}^3 v_3$, 我们要求

$$v_1,\, v_2 \in \mathrm{Ker}(\mathscr{N}) \setminus \mathrm{Im}(\mathscr{N})\,,\; w_3 \in \mathrm{Im}(\mathscr{N}^3),\, w_3 \neq 0 \quad \text{且}\; v_1,\, v_2,\, w_3 \text{ 线性无关}.$$

通过观察 N^3 和 $\mathrm{Im}(\mathscr{N}^3)$ 的表达式可知, $3e_1 = \mathscr{N}^3 e_6$. 所以 $w_3 = 3e_1$ 是一个不错的选择, 因为这样随之找到 $v_3 = e_6$. 此时

$$\mathscr{N}^2 v_3 = -e_1 - 2e_2 + e_3\,,\; \mathscr{N} v_3 = e_2 + e_4\,.$$

再观察 $\mathrm{Ker}(\mathscr{N})$ 和 $\mathrm{Im}(\mathscr{N})$ 的表达式, 不难发现 v_1, v_2 可以取为 $v_1 = e_2 + e_3$, $v_2 = e_3 + e_5$.

综上所述,

$$\mathcal{B} := (e_2 + e_3\,,\; e_3 + e_5\,,\; 3e_1\,,\; -e_1 - 2e_2 + e_3\,,\; e_2 + e_4\,,\; e_6)$$

是 $\mathscr{N}$ 的一组 Jordan 基, 从标准基 $\mathcal{E}$ 到 $\mathcal{B}$ 的过渡矩阵是

$$P = \begin{pmatrix} 0 & 0 & 3 & -1 & 0 & 0 \\ 1 & 0 & 0 & -2 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

读者可以验证 $P^{-1}NP = J$ 成立. ■

5.1.4 习题

以下 K 总是表示 $\mathbb{C}$ 的一个子域, V 表示 K-向量空间, 且 $\dim V \geq 1$. 设 $\mathscr{A}, \mathscr{B} \in \mathrm{End}(V)$.

习题 5.1.1. 定义复向量空间 $\mathbb{C}^3$ 上的线性变换 $\mathscr{A}$ 如下:

$$\mathscr{A}\ :\ \mathbb{C}^3 \longrightarrow \mathbb{C}^3\,; \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \longmapsto \begin{pmatrix} 3x + y - 2z \\ -x + 5z \\ -x - y + 4z \end{pmatrix}.$$

求 $\mathscr{A}$ 的所有特征值及相应的广义特征子空间.

习题 5.1.2. 设 $\mathscr{A}$ 是 K-向量空间 V 上的可逆线性变换, $\mathscr{A}^{-1}$ 为其逆变换. 证明: 对于任意 $0 \neq \lambda \in K$, $G(\lambda, \mathscr{A}) = G(\lambda^{-1}, \mathscr{A}^{-1})$.

习题 5.1.3. 假设 $\dim V = 3$, $\mathscr{A} \in \mathrm{End}(V)$ 在某一组有序基下的矩阵为 A. 对以下两种情况分别讨论 $\mathscr{A}$ 是否是幂零变换. 如是, 它的幂零阶是多少?

$$A = \begin{pmatrix} 0 & -3 & 3 \\ -2 & -7 & 13 \\ -1 & -4 & 7 \end{pmatrix} \qquad A = \begin{pmatrix} 3 & 6 & -15 \\ 1 & 2 & -5 \\ 1 & 2 & -5 \end{pmatrix}$$

习题 5.1.4. 证明: 若 $\mathscr{A}\mathscr{B}$ 是幂零变换, 则 $\mathscr{B}\mathscr{A}$ 也是幂零变换.

习题 5.1.5. 假设 $\mathscr{A}$, $\mathscr{B}$ 都是幂零变换.

1. 证明: 若 $\mathscr{A}$, $\mathscr{B}$ 可交换, 则 $\mathscr{A}+\mathscr{B}$ 一定是幂零变换.
2. 举例说明: 若 $\mathscr{A}$, $\mathscr{B}$ 不可交换, 则 $\mathscr{A}+\mathscr{B}$ 可能不是幂零变换.
3. $\mathscr{A}\mathscr{B}$ 是否一定是幂零变换? 若是, 请给出证明; 若否, 请举出反例.

习题 5.1.6. 设 $\mathscr{A}$ 是有限维向量空间 V 上的幂零变换. 证明: 对任意非零常数 $\alpha\in K$, $\alpha I+\mathscr{A}$ 是可逆变换, 并求出 $(\alpha I+\mathscr{A})^{-1}$.

习题 5.1.7. 设 $m\in\mathbb{N}$ 满足 $\mathrm{Im}(\mathscr{A}^m)=\mathrm{Im}(\mathscr{A}^{m+1})$.
证明: 对于所有自然数 $k\geq m$ 均有 $\mathrm{Im}(\mathscr{A}^k)=\mathrm{Im}(\mathscr{A}^{k+1})$.

习题 5.1.8. 设 $n=\dim V$. 证明: 对于自然数 $k\geq n$ 均有 $\mathrm{Im}(\mathscr{A}^k)=\mathrm{Im}(\mathscr{A}^{k+1})$.

习题 5.1.9. 设 $n=\dim V\geq 1$, $\mathscr{A}\in\mathrm{End}(V)$ 为幂零变换. 设 $\dim E(0,\mathscr{A})=k$. 证明: $\mathscr{A}^{n-k+1}=0$.

习题 5.1.10. 设 $n=\dim V\geq 1$. 假设 $\mathrm{Ker}(\mathscr{A}^n)\neq\mathrm{Ker}(\mathscr{A}^{n-1})$.
证明 $\mathscr{A}$ 是幂零变换并对每个 $j\in\mathbb{N}$ 求出 $\dim\mathrm{Ker}(\mathscr{A}^j)$.

习题 5.1.11. 设 $n=\dim V\geq 1$. 假设 $\mathscr{A}$ 不是幂零变换.
证明: $V=\mathrm{Ker}(\mathscr{A}^{n-1})\oplus\mathrm{Im}(\mathscr{A}^{n-1})$.

习题 5.1.12. 设 A 为 $\mathbf{M}_n(K)$ 中的幂零矩阵. 定义 $V=\mathbf{M}_n(K)$ 上的线性变换

$$\mathscr{A}\ :\ V\longrightarrow V\ ;\quad X\longmapsto AX-XA\,.$$

证明 $\mathscr{A}$ 是个幂零变换.

习题 5.1.13. 假设 $\dim V=3$, $\mathscr{A}\in\mathrm{End}(V)$, $\mathcal{B}=(\varepsilon_1,\varepsilon_2,\varepsilon_3)$ 为 V 的一组有序基. 令 $A=\mathcal{M}_{\mathcal{B}}(\mathscr{A})$. 对以下两种情况分别求出 $\mathscr{A}$ 的一组 Jordan 基:

$$A=\begin{pmatrix}0&-3&3\\-2&-7&13\\-1&-4&7\end{pmatrix}\qquad A=\begin{pmatrix}3&6&-15\\1&2&-5\\1&2&-5\end{pmatrix}$$

习题 5.1.14. 设 $A\in\mathbf{M}_n(K)$, $V=\mathbf{M}_n(K)$. 定义线性变换 $\mathscr{A}\in\mathrm{End}(V)$ 为

$$\forall\,X\in V=\mathbf{M}_n(K)\,,\quad \mathscr{A}(X):=A^TXA\,.$$

1. 证明: 若 A 是幂零矩阵, 则 $\mathscr{A}$ 是幂零变换.
2. 假设 $n=2$, $A=\begin{pmatrix}0&1\\0&0\end{pmatrix}$. 找出 $\mathscr{A}$ 在 V 中的一组 Jordan 基.
3. 证明: 若 $\mathscr{A}$ 是幂零变换, 则 A 是幂零矩阵.

习题 5.1.15. 证明或给出反例: 若 $K=\mathbb{C}$, $n=\dim V$, 则 $\mathscr{A}^n$ 一定可对角化.

习题 5.1.16. 验证以下矩阵 A 为幂零矩阵, 并将其化为 Jordan 标准形:

$$(1)\ \ A=\begin{pmatrix}0&2&3&4\\0&0&2&3\\0&0&0&2\\0&0&0&0\end{pmatrix}\ ;\quad (2)\ \ A=\begin{pmatrix}2&1&0&0\\-4&-2&0&0\\7&1&1&1\\-17&-6&-1&-1\end{pmatrix}\ ;\quad (3)\ \ A=\begin{pmatrix}0&3&0&-3\\2&7&0&-13\\0&3&0&-3\\1&4&0&-7\end{pmatrix}$$

习题 5.1.17. 设 $A \in \mathbf{M}_2(K)$. 假设存在矩阵 $B \in \mathbf{M}_2(K)$ 使得 $AB - BA = A$.

1. 证明 $\mathrm{Tr}(A) = \mathrm{Tr}(A^2) = 0$.
2. 证明 A 是幂零矩阵.

习题 5.1.18. 设 $\mathscr{A} \in \mathrm{End}(V)$ 在 V 的一组有序基 $\varepsilon_1, \cdots, \varepsilon_n$ 下的矩阵是

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

求 $\mathscr{A}$ 在 V 中的一组 Jordan 基.

5.2 一般线性变换的 Jordan 标准形

本节中我们总是假设 V 是有限维 K-向量空间, 且 $\dim V \geq 1$.

5.2.1 Jordan 标准形的存在唯一性

有了 5.1 节关于广义特征子空间和幂零变换的细致分析, 我们现在可以轻松自如地讨论一般线性变换的 Jordan 标准形.

定理 5.2.1: 设 V 是有限维 K-向量空间, $\dim V \geq 1$, $\mathscr{A} \in \mathrm{End}(V)$, $f(X) = P_{\mathscr{A}}(X) \in K[X]$ 为 $\mathscr{A}$ 的特征多项式.

假设 $f(X)$ 的所有复数根都属于 K (若 $K = \mathbb{C}$, 这一假设自动成立).

则一定存在 V 的一组基 $\mathcal{B}$ 使 $\mathscr{A}$ 在 $\mathcal{B}$ 下的矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是 Jordan 形矩阵. 也就是说, $\mathscr{A}$ 在 K 上可以化为 Jordan 标准形.

证明. 设 $\lambda_1, \cdots, \lambda_r$ 是 $f(X)$ 所有不同的复数根. 根据题设条件, 它们都属于 K. 根据定理 5.1.10, 可以将空间 V 做如下直和分解:

$$V = W_1 \oplus \cdots \oplus W_r \quad \text{其中} \quad W_i = G(\lambda_i, \mathscr{A}).$$

这里的广义特征子空间 $W_i = G(\lambda_i, \mathscr{A})$ 都是 $\mathscr{A}$ 的不变子空间. 对于任意 $\lambda \in K$, W_i 也是 $\mathscr{A} - \lambda I$ 的不变子空间. 所以我们可以定义限制映射 $\mathscr{N}_i := (\mathscr{A} - \lambda_i I)|_{W_i}$. 则 $\mathscr{N}_i$ 是 W_i 上的幂零变换. 所以, 由定理 5.1.23 可知, 每个 W_i 内存在 $\mathscr{N}_i$ 的一组 Jordan 基 $\mathcal{B}_i$. 在 $\mathcal{B}_i$ 下 $\mathscr{N}_i = (\mathscr{A} - \lambda_i I)|_{W_i}$ 的矩阵 M_i 是一个 Jordan 形幂零矩阵. 而 $\mathscr{A}|_{W_i} = \mathscr{N}_i + \lambda_i I$ 在 $\mathcal{B}_i$ 下的矩阵是 $J_i := M_i + \lambda_i I$. 易见 J_i 也是 Jordan 形矩阵 (它的 Jordan 块大小及分布情况和 Jordan 形矩阵 M_i 相同, 只不过特征值从 0 变为 λ_i, 或者说对角线元素从 0 变成了 λ_i).

由于 $V = W_1 \oplus \cdots \oplus W_r$, 故 $\mathcal{B} := (\mathcal{B}_1, \cdots, \mathcal{B}_r)$ 构成 V 的一组有序基, 在这组基下 $\mathscr{A}$ 的矩阵为如下分块矩阵

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{pmatrix}.$$

这是一个 Jordan 形矩阵. 因此我们证明了 $\mathscr{A}$ 的 Jordan 标准形存在性. □

思考题 5.10. 考虑 Jordan 形矩阵

$$J=\begin{pmatrix} J_2(1) & & & \\ & J_2(2) & & \\ & & J_1(3) & \\ & & & J_4(4) \end{pmatrix} \quad \text{和} \quad J'=\begin{pmatrix} J_1(3) & & & \\ & J_4(4) & & \\ & & J_2(2) & \\ & & & J_2(1) \end{pmatrix}.$$

写出一个可逆矩阵 $P\in\mathbf{M}_9(K)$ 使 $P^{-1}JP=J'$.

定理 5.2.1 说明了复线性变换的 Jordan 标准形的存在性. 事实上, 在不计 Jordan 块排序变化的意义下, Jordan 标准形还具有唯一性. 为了方便说明这一点. 我们回忆一下特征值重数的概念.

(5.2.2) 设 $\mathscr{A}\in\operatorname{End}(V)$, $f(X)=P_{\mathscr{A}}(X)\in K[X]$ 为 $\mathscr{A}$ 的特征多项式. 我们知道, 如果 $\lambda\in K$ 是 $\mathscr{A}$ 的特征值, 则 λ 是 $f(X)$ 的一个根. 我们曾将 λ 作为 $f(X)$ 根的重数定义为 λ 作为 $\mathscr{A}$ 的特征值的代数重数 (或者简称重数) (参见 (5.0.15) 或本书上册 (4.3.13)).

若 $A\in\mathbf{M}_n(K)$, $n\geq 1$ 是系数取自 K 的方阵, 则 A 在 K 中的特征值正是线性变换

$$\mathscr{A}\ :\ K^n\longrightarrow K^n\,;\quad X\longmapsto AX$$

的特征值. 由此我们可以对方阵的特征值定义代数重数.

若 L 是 K 在 $\mathbb{C}$ 中的一个扩域, 那么 A 在 K 中的任何特征值 λ 都是 A 在 L 中的特征值, 也即是线性变换

$$\mathscr{A}_L\ :\ L^n\longrightarrow L^n\,;\quad X\longmapsto AX$$

的特征值. 事实上, λ 作为 $\mathscr{A}\in\operatorname{End}(K^n)$ 的特征值和作为 $\mathscr{A}_L\in\operatorname{End}(L^n)$ 的特征值对应的代数重数是相同的. 这一事实成立的原因主要是 $\mathscr{A}$ 和 $\mathscr{A}_L$ 的特征多项式都是矩阵 A 的特征多项式 $f(X)$. 而多项式 $f(X)$ 在 K 中的根在考虑重数时, 在 K 上考虑和在任何其他扩域中考虑都是一样的. 具体来说, 若 λ 作为 $\mathscr{A}$ 的特征值代数重数为 m, 则在 $K[X]$ 中存在分解式 $f(X)=(X-\lambda)^m h(X)$, 其中 $h(X)\in K[X]$ 满足 $h(\lambda)\neq 0$ (参见本书上册 (A.4.10) 一段). 因为同样的分解式也在 $L[X]$ 中成立, 所以同样的分解式表明 λ 作为 $\mathscr{A}_L$ 的特征值代数重数也是 m. ■

我们将使用下面的引理来证明关于特征值重数的一些重要结论.

引理 5.2.3: *设 V 是有限维 K-向量空间, $\mathscr{A}\in\operatorname{End}(V)$, $\lambda\in K$ 是 $\mathscr{A}$ 的一个特征值, $W=G(\lambda,\mathscr{A})$, $\mathscr{A}'=\mathscr{A}|_W\in\operatorname{End}(W)$.*

则 $\mathscr{A}'$ 的特征多项式为 $P_{\mathscr{A}'}(X)=(X-\lambda)^{\dim W}$.

证明. 设 $m=\dim W$, $\mathscr{N}=\mathscr{A}'-\lambda I$. 则 $\mathscr{N}$ 是 $W=G(\lambda,\mathscr{A})$ 上的幂零变换 (例 5.1.13). 由引理 5.1.15 可知 $\mathscr{N}$ 的特征多项式为 X^m. 这就是说, 如果 N 是 $\mathscr{N}$ 在 W 的某一组基下的矩阵, 则 $\det(XI_m-N)=X^m$. 在同一组基下 $\mathscr{A}'$ 的矩阵是 $A:=N+\lambda I_m$, 所以 $\mathscr{A}'$ 的特征多项式为

$$\det(XI_m-A)=\det(XI_m-\lambda I_m-N)=\det((X-\lambda)I_m-N)=(X-\lambda)^m\,.$$

这就是所欲证的结论. □

命题 5.2.4: *设 V 是有限维 K-向量空间, $\mathscr{A}\in\operatorname{End}(V)$. 假设特征多项式 $P_{\mathscr{A}}(X)$ 的复数根 $\lambda_1,\cdots,\lambda_r$ 都属于 K. 设 m_i 是特征值 λ_i 的代数重数.*

则对每个 $i\in[\![1,r]\!]$, 均有 $m_i=\dim G(\lambda_i\,,\mathscr{A})$, 即每个特征值的代数重数等于相应的广义特征子空间维数.

证明. 记 $W_i = G(\lambda_i, \mathscr{A})$, $\mathscr{A}_i \in \mathrm{End}(W_i)$ 为 $\mathscr{A}$ 在广义特征子空间 W_i 上的限制. 因为 V 是 W_i 这些不变子空间的直和 (命题 5.1.10), 所以由思考题 5.3 (2) 的结论即可知道, $\mathscr{A}$ 的特征多项式等于 $\mathscr{A}_i$ 的特征多项式之乘积. 而 $\mathscr{A}_i$ 的特征多项式由引理 5.2.3 给出. 综上可知 $\mathscr{A}$ 的特征多项式为

$$\prod_{i=1}^{r}(X-\lambda_i)^{\dim G(\lambda_i\,,\,\mathscr{A})}\,.$$

由此结合特征值代数重数的定义即得所需结论. □

注记 5.2.5. 对于线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 的任意特征值 $\lambda \in K$, 很多书把特征子空间 $E(\lambda, \mathscr{A})$ 的维数称为特征值 λ 的**几何重数** (geometric multiplicity).

若 $A \in \mathbf{M}_n(K)$, $n \geq 1$ 是系数取自 K 的方阵, 则 A 在 K 中的特征值正是线性变换

$$\mathscr{A}\ :\ K^n \longrightarrow K^n\ ;\quad X \longmapsto AX$$

的特征值. 由此我们可以对方阵的特征值定义几何重数.

若 L 是 K 在 $\mathbb{C}$ 中的一个扩域, 那么 A 在 K 中的任何特征值 λ 都是 A 在 L 中的特征值, 也即是线性变换

$$\mathscr{A}_L\ :\ L^n \longrightarrow L^n\ ;\quad X \longmapsto AX$$

的特征值. 和代数重数的情形类似, λ 作为 $\mathscr{A} \in \mathrm{End}(K^n)$ 的特征值和作为 $\mathscr{A}_L \in \mathrm{End}(L^n)$ 的特征值对应的几何重数是相同的.

要理解这个事实, 我们只需要注意到 $\lambda I - \mathscr{A}$ 和 $\lambda I - \mathscr{A}_L$ 各自在 K^n 和 L^n 的标准基下的矩阵都是 $\lambda I - A$. 而几何重数 $\dim_K E(\lambda, \mathscr{A}) = \dim_K \mathrm{Ker}(\lambda I - \mathscr{A})$ 和矩阵 $\lambda I - A$ 在 K 上的零化度相等. 根据秩–零化度定理, $\dim_K E(\lambda, \mathscr{A}) = n - \mathrm{rank}(\lambda I - A)$. 矩阵的秩由它的相抵标准形确定 (本书上册定理 1.2.35). 当一个矩阵 $B \in \mathbf{M}_n(K)$ 经过 K 上的一系列初等行列变换化为相抵标准形时, 也可以认为是经过了 L 上的初等变换化成了相抵标准形. 而如果 $B = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ (相抵标准形的形式), 则在 K 上计算 $\mathrm{rank}(B)$ 时是计算 K-向量空间

$$\mathrm{span}_K(e_1, \cdots, e_r) = \{x_1e_1 + \cdots + x_re_r \,|\, x_i \in K\}$$

的维数, 在 L 上计算 $\mathrm{rank}(B)$ 时是计算 L-向量空间

$$\mathrm{span}_L(e_1, \cdots, e_r) = \{x_1e_1 + \cdots + x_re_r \,|\, x_i \in L\}\,.$$

以上两个维数显然都是 r. 所以, $\mathbf{M}_n(K)$ 中的矩阵在计算零化度和秩时, 在 K 上考虑和在 L 上考虑时结果是一致的. ■

思考题 5.11. 题设如命题 5.2.4. 对于每个特征值 λ_i, 记 d_i 为其几何重数.

证明:

1. 对每个 $i \in [\![1, r]\!]$, 均有 $m_i \geq d_i$, 即每个特征值的代数重数大于等于其几何重数.

2. $\mathscr{A}$ 可对角化的充分必要条件是对每个 $i \in [\![1, r]\!]$ 均有 $m_i = d_i$.

现在我们可以证明 Jordan 标准形的唯一性了.

命题 5.2.6: 设 $\mathscr{A} \in \mathrm{End}(V)$. 假设 $\mathscr{A}$ 在 K 上可以化为 Jordan 标准形 (根据定理 5.2.1, 这等价于说特征多项式 $P_{\mathscr{A}}(X) \in K[X]$ 的所有复数根都属于 K).

则 $\mathscr{A}$ 的任意两个 Jordan 标准形所含的 Jordan 块数量以及每个 Jordan 块大小一定相互吻合. 因此, 在不计 Jordan 块排序变化的意义下, $\mathscr{A}$ 在 K 上的 Jordan 标准形具有唯一性.

证明. 设 $\lambda_1, \cdots, \lambda_r$ 是特征多项式 $P_{\mathscr{A}}(X)$ 的所有复数根 (根据题设, 它们全部属于 K). 通过适当调整 Jordan 块的排序不妨设 $\mathscr{A}$ 的一个 Jordan 标准形 J 具有如下形式

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{pmatrix}$$

其中每个 J_i 也都是 Jordan 形矩阵, 且 J_i 中 Jordan 块的特征值 (亦即对角线元素) 都是 λ_i. 容易看出, 这里每个 J_i 的阶 m_i 实际上就是特征值 λ_i 的代数重数.

设 $\mathcal{B}$ 是与 J 对应的 Jordan 基, $\mathcal{B}_1, \cdots, \mathcal{B}_r$ 是将 $\mathcal{B}$ 按顺序分成 r 组所得的子向量组, 其中每个 $\mathcal{B}_i$ 所含的向量数目为 m_i. 根据以上 J 的分块情况可知, $G_i := \operatorname{span}(\mathcal{B}_i)$ 是 $\mathscr{A}$ 的不变子空间, 且 $\mathscr{A}_i := \mathscr{A}|_{G_i}$ 在 $\mathcal{B}_i$ 这组基下的矩阵为 J_i. 注意到 $J_i - \lambda_i I_{m_i}$ 是严格上三角阵, 因而是幂零矩阵 (例 5.1.17), 所以 $\mathscr{A}_i - \lambda_i I$ 是 G_i 上的幂零变换. 因此由推论 5.1.5 知道

$$G_i = \operatorname{Ker}(\mathscr{A}_i - \lambda I)^{\dim G_i} = \operatorname{Ker}(\mathscr{A}_i - \lambda_i I)^{m_i} \subseteq \operatorname{Ker}(\mathscr{A} - \lambda_i I)^{m_i} \subseteq G(\lambda_i, \mathscr{A}).$$

也就是说 G_i 包含在 $\mathscr{A}$ 的广义特征子空间 $W_i := G(\lambda_i, \mathscr{A})$ 之内. 我们前面说过 m_i 等于 λ_i 的代数重数, 根据命题 5.2.4 可知 $m_i = \dim G(\lambda_i, \mathscr{A})$, 即 $\dim G_i = \dim W_i$. 所以 $G_i = \operatorname{span}(\mathcal{B}_i)$ 就等于广义特征子空间 $W_i := G(\lambda_i, \mathscr{A})$. 由此可知 J_i 是 $\mathscr{A}|_{W_i}$ 的 Jordan 标准形, 而 $J_i - \lambda_i I_{m_i}$ 是幂零变换 $(\mathscr{A} - \lambda_i I)|_{W_i}$ 的 Jordan 标准形. 注意 $W_i = G(\lambda_i, \mathscr{A})$ 与 Jordan 基和 Jordan 标准形的选取是无关的. 而根据幂零变换的 Jordan 标准形唯一性 (参见 (5.1.22) 一段), 在不计 Jordan 块排序的意义下, $J_i - \lambda_i I_{m_i}$ 完全由 $(\mathscr{A} - \lambda_i I)|_{W_i}$ (进而完全由 $\mathscr{A}$) 决定. 于是, $J_i = (J_i - \lambda_i I_{m_i}) + \lambda_i I_{m_i}$ 完全由 $\mathscr{A}$ 确定. 这样就最终证明了 $J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{pmatrix}$ 完全由 $\mathscr{A}$ 确定. 命题至此证毕. □

我们提醒读者注意, 以上我们只是证明了 Jordan 标准形的唯一性, 但是将线性变换化为 Jordan 标准型的 Jordan 基并不是唯一的. 事实上, 对于同一个 Jordan 形矩阵 J, 我们可以找到不同的基 $\mathcal{B}_1, \mathcal{B}_2$ 满足 $\mathcal{M}_{\mathcal{B}_1}(\mathscr{A}) = J = \mathcal{M}_{\mathcal{B}_2}(\mathscr{A})$. 这一现象其实在例 5.1.26 中我们就已经遇到了.

推论 5.2.7: *设 $\mathscr{A}$ 是有限维 K-向量空间 V 上的线性变换.*

则 $\mathscr{A}$ 可对角化当且仅当以下两个条件成立:

1. *特征多项式 $P_{\mathscr{A}}(X)$ 的所有复数根都属于 K;*
2. *$\mathscr{A}$ 的 Jordan 标准形是对角阵.*

换成矩阵语言表述, 矩阵 $A \in \mathbf{M}_n(K)$ 在 K 上可对角化的充分必要条件是: A 的所有复特征值属于 K 且 A 在 K 上的 Jordan 标准形是对角阵.

证明. 若 (1) 和 (2) 两个条件成立, 那么存在 $\mathscr{A}$ 的 Jordan 基 $\mathcal{B}$ 使得 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是 Jordan 标准形, 而且是对角阵. 于是按照定义, $\mathscr{A}$ 可以对角化.

反过来, 若 $\mathscr{A}$ 可对角化, 则在某一组基下 $\mathscr{A}$ 的矩阵 A 是系数取自 K 的对角阵, 此时 $\mathscr{A}$ 的特征多项式 $P_{\mathscr{A}}(X)$ 的所有复数根就是 A 的对角线元素, 因此 (1) 成立. 因为对角矩阵 A 本身就是 Jordan 形矩阵 (其每个 Jordan 块都是 1 阶的), 所以由 Jordan 标准形的唯一性可知 A 就是 $\mathscr{A}$ 的一个Jordan 标准形, 而 $\mathscr{A}$ 的其它 Jordan 标准形都可以通过调整 A 的对角线元素排序得到, 所以 $\mathscr{A}$ 的 Jordan 标准形都是对角阵, 即条件 (2) 成立. □

推论 5.2.8: *矩阵 $A \in \mathbf{M}_n(K)$ 在 K 上可对角化当且仅当 A 在 $\mathbb{C}$ 上可对角化且 A 的所有复特征值都属于 K.*

证明. 设 $f(X)$ 为 A 的特征多项式. 根据推论 5.2.7, A 在 K 上可对角化当且仅当 $f(X)$ 的复数根都属于 K 且 A 在 K 上的 Jordan 标准形是对角阵. 显然, A 在 K 上的任意 Jordan 标准形 J 也是 A 在 $\mathbb{C}$ 上的一个 Jordan 标准形. 根据 Jordan 标准形的唯一性, A 在 $\mathbb{C}$ 上的 Jordan 标准形都是 J 调整 Jordan 块顺序得到的. 所以, J 是对角阵当且仅当 A 在 $\mathbb{C}$ 上的 Jordan 标准形是对角阵. 最后这一条件正好等价于 A 在 $\mathbb{C}$ 上可对角化. □

5.2.2 计算实例

现在我们讨论如何将一般的线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 化为 Jordan 标准形.

从定理 5.2.1 的证明可以看出, 具体的计算过程可以分成两个阶段. 第一个阶段是求出 $\mathscr{A}$ 的所有特征值 λ_i 以及对应的广义特征子空间 $W_i = G(\lambda_i, \mathscr{A})$. 第二个阶段考虑 $\mathscr{A}$ 在 W_i 上的限制 $\mathscr{A}_i := \mathscr{A}|_{W_i}$. 此时 $\mathscr{A}_i - \lambda_i I$ 是 W_i 上的幂零变换. 所以, 按照 5.1.3 节介绍的方法求出 $\mathscr{A}_i - \lambda_i I$ 的 Jordan 基 $\mathcal{B}_i$. 在这组基下 $\mathscr{A}_i - \lambda_i I$ 的矩阵 M_i 是幂零的 Jordan 形矩阵, 而 $\mathscr{A}_i$ 的矩阵则是 $J_i = M_i + \lambda_i I$, 它就是 $\mathscr{A}_i$ 的 Jordan 标准形. 将各个 $\mathcal{B}_i$ 合并起来就得到 $\mathscr{A}$ 的一组 Jordan 基, 相应的 Jordan 标准形就是以 J_i 为对角块的分块矩阵.

下面我们通过几个例子来帮助读者体验一下上述的算法具体如何实施.

例 5.2.9. 将矩阵 $A = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 1 \\ & 1 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 \\ & & & 1 & 1 & 0 \\ & & & & 2 & -1 \\ & & & & & 2 \end{pmatrix}$ 化为 Jordan 标准形.

将线性变换 $K^6 \to K^6$, $X \mapsto AX$ 记为 $\mathscr{A}$. 容易观察到 A 有两个复特征值 $\lambda_1 = 1$ 和 $\lambda_2 = 2$. 相应的广义特征子空间 $W_i = G(\lambda_i, \mathscr{A})$ 维数是相应的代数重数, 即 $\dim W_1 = 4$, $\dim W_2 = 2$.

接下来考虑幂零变换 $\mathscr{N}_i := (\mathscr{A} - \lambda_i I)|_{W_i}$.

对于 W_1 和 $\mathscr{N}_1$, 我们先计算

$$A - \lambda_1 I = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 1 \\ & 0 & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 & 0 \\ & & & 0 & 1 & 0 \\ & & & & 1 & -1 \\ & & & & & 1 \end{pmatrix} \quad \text{故} \quad \mathrm{Ker}(\mathscr{A} - \lambda_1 I) = \mathrm{span}(e_1, e_3, e_4)$$

而

$$(A - \lambda_1 I)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ & 0 & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 & 0 \\ & & & 0 & 1 & -1 \\ & & & & 1 & -2 \\ & & & & & 1 \end{pmatrix}, \quad \mathrm{Ker}(\mathscr{A} - \lambda_1 I)^2 = \mathrm{span}(e_1, e_2, e_3, e_4)\,.$$

注意到 $G(\lambda_1, \mathscr{A}) = W_1$ 的维数是 4, 这与上面得到的 $\dim \mathrm{Ker}(\mathscr{A} - \lambda_1 I)^2$ 相等. 所以至此我们已经可以得到 $W_1 = \mathrm{Ker}(\mathscr{A} - \lambda_1 I)^2$ (而不必再计算 $\mathrm{Ker}(\mathscr{A} - \lambda_1 I)^3$, $\mathrm{Ker}(\mathscr{A} - \lambda_1 I)^4$). 这说明 $\mathscr{N}_1$ 的幂零阶为 2. (请读者思考: 这里 $\mathscr{N}_1^2 = 0$, 可是上面计算的矩阵 $(A - \lambda_1 I)^2$ 并不等于 0. 这两个事实为什么没有矛盾呢?)

同时, 我们又看到 $\dim \operatorname{Ker}(\mathscr{N}_1) = \dim(\operatorname{Ker}(\mathscr{A} - \lambda_1 I) \cap W_1) = \dim \operatorname{Ker}(\mathscr{A} - \lambda_1 I) = 3$. 所以 $\mathscr{N}_1$ 的 Jordan 标准形含 3 个 Jordan 块, 最大的 Jordan 块是 2 阶的, 所有 Jordan 块阶数加起来是 4. 所以, $\mathscr{N}_1$ 的 Jordan 标准形是

$$M_1 = \begin{pmatrix} J_1(0) & & \\ & J_1(0) & \\ & & J_2(0) \end{pmatrix} = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 0 & 1 \\ & & & 0 \end{pmatrix}.$$

与 M_1 对应的 Jordan 基形如 v_1, v_2, $\mathscr{N}_1 v_3$, v_3, 其中 v_1, v_2, $w_3 := \mathscr{N}_1 v_3$ 构成 $\operatorname{Ker}(\mathscr{N}_1)$ 的一组基, 且 $w_3 \in \operatorname{Im}(\mathscr{N}_1) \setminus \operatorname{Im}(\mathscr{N}_1{}^2) = \operatorname{Im}(\mathscr{N}_1) \setminus \{0\}$. 由于

$$\operatorname{Im}(\mathscr{N}_1) = (\mathscr{A} - \lambda_1 I)(W_1) = (\mathscr{A} - \lambda_1 I)\big(\operatorname{span}(e_1, e_2, e_3, e_4)\big) = \operatorname{span}(-e_1)$$

而 $\operatorname{Ker}(\mathscr{N}_1) = \operatorname{span}(e_1, e_3, e_4)$, 所以可以想到的一个合适选择是

$$v_1 = e_3\,,\ v_2 = e_4,\ w_3 = -e_1\,,\ v_3 = e_2\,.$$

(注意 $\mathscr{N}_1 e_2 = (\mathscr{A} - \lambda_1 I)e_2 = -e_1$). 也就是说, $\mathcal{B}_1 = (e_3, e_4, -e_1, e_2)$ 是 $\mathscr{N}_1 = (\mathscr{A} - \lambda_1 I)|_{W_1}$ 的一组 Jordan 基. 它同时也是 $\mathscr{A}_1 = \mathscr{A}|_{W_1}$ 的一组 Jordan 基, 而

$$\mathcal{M}_{\mathcal{B}_1}(\mathscr{A}_1) = M_1 + \lambda_1 I = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}.$$

再来考虑第二个广义特征子空间 $W_2 = G(\lambda_2, \mathscr{A})$. 计算可得

$$A - \lambda_2 I = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 & 1 \\ & -1 & 0 & 0 & 0 & 0 \\ & & -1 & 0 & 0 & 0 \\ & & & -1 & 1 & 0 \\ & & & & 0 & -1 \\ & & & & & 0 \end{pmatrix} \quad 故 \quad \operatorname{Ker}(\mathscr{A} - \lambda_2 I) = \operatorname{span}(e_4 + e_5)$$

而

$$(A - \lambda_2 I)^2 = \begin{pmatrix} 1 & 2 & 0 & 0 & 0 & -1 \\ & 1 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 \\ & & & 1 & -1 & -1 \\ & & & & 0 & 0 \\ & & & & & 0 \end{pmatrix}, \quad \operatorname{Ker}(\mathscr{A} - \lambda_2 I)^2 = \operatorname{span}(e_1 + e_4 + e_6,\, e_4 + e_5)\,.$$

于是得到 $W_2 = \operatorname{Ker}(\mathscr{A} - \lambda_2 I)^2 = \operatorname{span}(e_1 + e_4 + e_6, e_4 + e_5)$, $\dim \operatorname{Ker}(\mathscr{N}_2) = \dim \operatorname{Ker}(\mathscr{A} - \lambda_2 I) = 1$. 所以, $\mathscr{N}_2$ 的 Jordan 标准形 M_2 只有 1 个 Jordan 块, 其阶数是 2. 即, $M_2 = \begin{pmatrix} 0 & 1 \\ & 0 \end{pmatrix}$. 与之对应的 Jordan 基形如 $\mathscr{N}_2 v_4$, v_4, 其中 $w_4 := \mathscr{N}_2 v_4$ 属于 $\operatorname{Im}(\mathscr{N}_2)$. 由于

$$\operatorname{Im}(\mathscr{N}_2) = (\mathscr{A} - \lambda_2 I)(W_2) = \operatorname{span}\big((\mathscr{A} - \lambda_2 I)(e_1 + e_4 + e_6)\,,\,(\mathscr{A} - \lambda_2 I)(e_4 + e_5)\big) = \operatorname{span}(-e_4 - e_5)\,,$$

可以想到 w_4 和 v_4 的一种合理选择是 $w_4 = -e_4 - e_5$, $v_4 = e_1 + e_4 + e_6$. 也就是说, $\mathcal{B}_2 = (-e_4 - e_5, e_1 + e_4 + e_6)$ 是 $\mathscr{A}_2$ 的一组 Jordan 基, 在这组基下 $\mathcal{M}_{\mathcal{B}_2}(\mathscr{A}_2) = M_2 + \lambda_2 I = \begin{pmatrix} 2 & 1 \\ & 2 \end{pmatrix}$.

综上所述, $\mathcal{B} = (e_3, e_4, -e_1, e_2, -e_4 - e_5, e_1 + e_4 + e_6)$ 是 $\mathscr{A}$ 的一组 Jordan 基, 在这组基下 $\mathscr{A}$ 的矩阵是

$$J = \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & 1 & & \\ & & & 1 & & \\ & & & & 2 & 1 \\ & & & & & 2 \end{pmatrix}$$

从标准基到 $\mathcal{B}$ 的过渡矩阵为

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

该矩阵满足 $P^{-1}AP = J$. ■

例 5.2.10. 将矩阵 $A = \begin{pmatrix} 4 & -1 & 0 & 0 \\ 5 & -2 & 0 & 0 \\ 1 & 2 & -3 & 2 \\ 4 & 3 & -6 & 5 \end{pmatrix}$ 化为 Jordan 标准形.

将线性变换 $K^4 \to K^4$, $X \mapsto AX$ 记为 $\mathscr{A}$. 计算 $\mathscr{A}$ 的特征多形式得到 $f(X) = (X+1)^2(X-3)^2$. 故 A 恰有两个复特征值 $\lambda_1 = -1$ 和 $\lambda_2 = 3$, 它们的代数重数均为 2.

计算可得

$$A - \lambda_1 I = \begin{pmatrix} 5 & -1 & 0 & 0 \\ 5 & -1 & 0 & 0 \\ 1 & 2 & -2 & 2 \\ 4 & 3 & -6 & 6 \end{pmatrix}, \quad \operatorname{Ker}(\mathscr{A} - \lambda_1 I) = \operatorname{span}(e_3 + e_4),$$

$$(A - \lambda_1 I)^2 = \begin{pmatrix} 20 & -4 & 0 & 0 \\ 20 & -4 & 0 & 0 \\ 21 & -1 & -8 & 8 \\ 53 & -1 & -24 & 24 \end{pmatrix}, \quad \operatorname{Ker}(\mathscr{A} - \lambda_1 I)^2 = \operatorname{span}(e_3 + e_4, e_1 + 5e_2 + 2e_3).$$

于是 $W_1 := G(\lambda_1, \mathscr{A}) = \operatorname{span}(e_3 + e_4, e_1 + 5e_2 + 2e_3)$, $\mathscr{N}_1 := (\mathscr{A} - \lambda_1 I)|_{W_1}$ 的 Jordan 标准形仅有 1 个 Jordan 块. 所以 $\mathscr{A}_1 := \mathscr{A}|_{W_1}$ 的 Jordan 标准形为 $J_1 = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$. 相应的 Jordan 基形如 $(\mathscr{A} - \lambda_1 I)v_1, v_1$, 其中 $w_1 := (\mathscr{A} - \lambda_1 I)v_1 = \mathscr{N}_1 v_1$ 属于 $\operatorname{Im}(\mathscr{N}_1)$. 由于

$$\operatorname{Im}(\mathscr{N}_1) = \operatorname{span}\big((\mathscr{A} - \lambda_1 I)(e_3 + e_4), (\mathscr{A} - \lambda_1 I)(e_1 + 5e_2 + 2e_3)\big) = \operatorname{span}(7e_3 + 7e_4)$$

可取 $w_1 = 7e_3 + 7e_4$, $v_1 = e_1 + 5e_2 + 2e_3$.

同样通过计算可得

$$A-\lambda_2 I=\begin{pmatrix}1&-1&0&0\\5&-5&0&0\\1&2&-6&2\\4&3&-6&2\end{pmatrix},\quad \operatorname{Ker}(\mathscr{A}-\lambda_2 I)=\operatorname{span}(e_3+3e_4),$$

$$(A-\lambda_2 I)^2=\begin{pmatrix}-4&4&0&0\\-20&20&0&0\\13&-17&24&-8\\21&-25&24&-8\end{pmatrix},\quad \operatorname{Ker}(\mathscr{A}-\lambda_2 I)^2=\operatorname{span}(e_3+3e_4,\,-2e_1-2e_2+e_4).$$

于是 $W_2:=G(\lambda_1,\mathscr{A})=\operatorname{span}(e_3+3e_4,\,-2e_1-2e_2+e_4)$, $\mathscr{N}_2:=(\mathscr{A}-\lambda_2 I)|_{W_2}$ 的 Jordan 标准形仅有 1 个 Jordan 块. 所以 $\mathscr{A}_2:=\mathscr{A}|_{W_2}$ 的 Jordan 标准形为 $J_2=\begin{pmatrix}3&1\\0&3\end{pmatrix}$. 相应的 Jordan 基形如 $(\mathscr{A}-\lambda_2 I)v_2,\,v_2$, 其中 $w_2:=(\mathscr{A}-\lambda_2 I)v_2=\mathscr{N}_2 v_2$ 属于 $\operatorname{Im}(\mathscr{N}_2)$. 由于

$$\operatorname{Im}(\mathscr{N}_2)=\operatorname{span}\big((\mathscr{A}-\lambda_2 I)(e_3+3e_4),\,(\mathscr{A}-\lambda_2 I)(-2e_1-2e_2+e_4)\big)=\operatorname{span}(-4e_3-12e_4)$$

可取 $w_2=-4e_3-12e_4$, $v_2=-2e_1-2e_2+e_4$.

综上所述, $\mathcal{B}=(7e_3+7e_4\,,\,e_1+5e_2+2e_3\,,\,-4e_3-12e_4\,,\,-2e_1-2e_2+e_4)$ 是 $\mathscr{A}$ 的一组 Jordan 基, 在这组基下 $\mathscr{A}$ 的矩阵是 $J=\begin{pmatrix}-1&1&&\\&-1&&\\&&3&1\\&&&3\end{pmatrix}$. 从标准基到 $\mathcal{B}$ 的过渡矩阵为 $P=\begin{pmatrix}0&1&0&-2\\0&5&0&-2\\7&2&-4&0\\7&0&-12&1\end{pmatrix}$, 该矩阵满足 $P^{-1}AP=J$. ■

5.2.3 Jordan–Chevalley 分解

本小节的目的是给出 Jordan 标准形理论的一些简单而重要的应用.

定理 5.2.11: 设 $\mathscr{A}$ 是有限维 K-向量空间 V 上的线性变换. 假设 $\mathscr{A}$ 的特征多项式的所有复数根属于 K.

则存在一对线性变换 $\mathscr{D},\mathscr{N}\in\operatorname{End}(V)$ 满足以下条件:

$$\mathscr{A}=\mathscr{D}+\mathscr{N}\,,\ \mathscr{D}\mathscr{N}=\mathscr{N}\mathscr{D}\quad \text{且 } \mathscr{D} \text{ 可对角化, } \mathscr{N} \text{ 是幂零变换.}$$

满足上述条件的表达式 $\mathscr{A}=\mathscr{D}+\mathscr{N}$ 称为 $\mathscr{A}$ 的一个 (**加法式**) Jordan–Chevalley **分解** (additive Jordan–Chevalley* decomposition).

证明. 根据定理 5.2.1, V 中存在 $\mathscr{A}$ 的一组 Jordan 基 $\mathcal{B}$, 使得 $\mathscr{A}$ 在 $\mathcal{B}$ 这组有序基下的矩阵为 Jordan 形矩阵 $J=\begin{pmatrix}J_{m_1}(\lambda_1)&&\\&\ddots&\\&&J_{m_r}(\lambda_r)\end{pmatrix}$. 这里的每个 Jordan 块可以写成 $J_{m_i}(\lambda_i)=\lambda_i I_{m_i}+N_i$ 的形式, 其中 N_i 为严格上三角阵 (因此是幂零矩阵). 令

$$D=\begin{pmatrix}\lambda_1 I_{m_1}&&\\&\ddots&\\&&\lambda_r I_{m_r}\end{pmatrix},\quad N=\begin{pmatrix}N_1&&\\&\ddots&\\&&N_r\end{pmatrix}.$$

*Claude Chevalley (1909–1984), 法国数学家.

容易看出 $DN = ND$ 且 $J = D + N$. 令 $\mathscr{D}$ 和 $\mathscr{N}$ 分别为在有序基 $\mathcal{B}$ 下矩阵表示为 D 和 N 的线性变换, 则 $\mathscr{A} = \mathscr{D} + \mathscr{N}$ 即为满足需求的一种分解式. □

可以对角化的线性变换有时也被称为**半单** (semisimple) **变换**[¶]. 所以定理 5.2.11 说明: 有限维复向量空间上的任何线性变换都可以分解为可交换的两个线性变换之和, 其中一个是半单变换而另一个是幂零变换.

注记 5.2.12. 读者可能会问定理 5.2.11 中的 Jordan–Chevalley 分解是否具有唯一性.

答案是肯定的. 一种最常见的证明方法需要用到一些多项式的理论. 具体的情况我们将在 5.3 节中阐明.

简言之, 我们可以通过多项式的理论来构造出一个特殊的 Jordan–Chevalley 分解 $\mathscr{A} = \mathscr{D} + \mathscr{N}$, 该分解还额外满足一个条件: 其中的 $\mathscr{D}$ 和 $\mathscr{N}$ 都可以写成 $\mathscr{A}$ 的多项式. 这意味着任何与 $\mathscr{A}$ 可交换的线性变换都和 $\mathscr{D}$ 和 $\mathscr{N}$ 可交换. 这样一来, 如果 $\mathscr{A} = \mathscr{D}' + \mathscr{N}'$ 是另外一种 Jordan–Chevalley 分解式, 那么由于 $\mathscr{N}'$ 和 $\mathscr{D}'$ 以及自身都可交换, 所以 $\mathscr{N}'$ 和 $\mathscr{A} = \mathscr{D}' + \mathscr{N}'$ 可交换. 于是 $\mathscr{N}'$ 和之前那个特殊的分解式中的 $\mathscr{N}$ 可交换. 这个性质可以保证 $\mathscr{N} - \mathscr{N}'$ 仍是幂零变换. 同理, $\mathscr{D}'$ 和 $\mathscr{D}$ 也可交换. 读者应该之前做过这样的习题 (参见本书上册习题 3.3.32): 两个可交换的可对角化线性变换 $\mathscr{D}$ 和 $\mathscr{D}'$ 可以同时对角化 (或者说可以找到公共的一组基使 $\mathscr{D}$ 和 $\mathscr{D}$ 在该组基下的矩阵都是对角阵). 因此, $\mathscr{D}' - \mathscr{D}$ 也可以对角化.

现在, 由 $\mathscr{D} + \mathscr{N} = \mathscr{A} = \mathscr{D}' + \mathscr{N}'$ 可得 $\mathscr{D}' - \mathscr{D} = \mathscr{N} - \mathscr{N}'$. 我们上面的讨论说明, 这个等式的左边是可对角化的线性变换, 而右边是幂零变换. 根据推论 5.1.14, 必然左右两边都等于 0, 即 $\mathscr{D}' = \mathscr{D}$, $\mathscr{N}' = \mathscr{N}$. 这就给出了 Jordan–Chevalley 分解具有唯一性的一个解释.

以上讨论涉及到一个有趣的结论: 在加法式 Jordan Chevalley 分解 $\mathscr{A} = \mathscr{D} + \mathscr{N}$ 中出现的 $\mathscr{D}$, $\mathscr{N}$ 实际上都可以写成 $\mathscr{A}$ 的多项式. 如果忽略这个结论, 只去论证分解式中 $\mathscr{D}$, $\mathscr{N}$ 的唯一性, 实际上也可以用其他的方法, 该方法只用到我们之前讲到的广义特征子空间理论. 具体细节我们留作习题 5.2.20. 读者亦可参看 `https://math.stackexchange.com/questions/3379483` (Stack Exchange 网站上的数学问答). ■

(5.2.13) 显然, 如果以矩阵的语言表述定理 5.2.11 的结论, 则可得到以下结论: 设 $n \geq 1$, $A \in \mathbf{M}_n(K)$. 假设 A 的复特征值都属于 K, 则存在一对矩阵 $D, N \in \mathbf{M}_n(K)$ 满足

$$A = D + N\,,\ DN = ND \quad \text{且 } D \text{ 可对角化, } N \text{ 是幂零矩阵.}$$

由此我们可以定义矩阵的 Jordan–Chevalley 分解. ■

思考题 5.12. 令 $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

1. 写出 A 的 Jordan–Chevalley 分解.
2. 找出 A 的 Jordan 标准形 J, 并写出 J 的 Jordan–Chevalley 分解.
3. 记 $\mathscr{A}$ 为线性变换 $\mathbb{R}^3 \to \mathbb{R}^3$; $X \mapsto AX$. 以上两个小题的结果是否和 $\mathscr{A}$ 的 Jordan–Chevalley 分解唯一性矛盾？为什么？

定义 5.2.14. 设 $\mathscr{A}$ 是有限维 K-向量空间 V 上的线性变换. 如果 $\mathscr{A} - I$ 是幂零变换, 我们称 $\mathscr{A}$ 是一个**幂幺变换** (unipotent transformation). ■

[¶]这个术语在在李群或代数群理论中相当常用.

定理 5.2.15: 设 $\mathscr{A}$ 是有限维 K-向量空间 V 上的可逆线性变换. 假设 $\mathscr{A}$ 的特征多项式的所有复数根属于 K.

则存在一对线性变换 $\mathscr{D}, \mathscr{U} \in \mathrm{End}(V)$ 满足以下条件:

$$\mathscr{A} = \mathscr{D}\mathscr{U} = \mathscr{U}\mathscr{D} \quad \text{且 } \mathscr{D} \text{ 可对角化, } \mathscr{U} \text{ 是幂幺变换.}$$

满足上述条件的表达式 $\mathscr{A} = \mathscr{D}\mathscr{U}$ 称为 $\mathscr{A}$ 的 **(乘法式)** Jordan–Chevally **分解** (multiplicative Jordan–Chevellay decomposition).

证明. 根据定理 5.2.11, 不妨设已经找到一个加法式的 Jordan–Chevalley 分解 $\mathscr{A} = \mathscr{D} + \mathscr{N}$. 因为 $\mathscr{N}$ 和 $\mathscr{A}$ 可交换 (为什么?), 不难证明 $\mathscr{A} - \mathscr{N}$ 是可逆变换. 于是, $\mathscr{U} := I + \mathscr{D}^{-1}\mathscr{N}$ 为幂幺变换. 由 $\mathscr{D}$ 和 $\mathscr{N}$ 的可交换性可以知道 $\mathscr{U}\mathscr{N} = \mathscr{N}\mathscr{U}$. 此外, $\mathscr{D}\mathscr{U} = \mathscr{D}(I + \mathscr{D}^{-1}\mathscr{N}) = \mathscr{D} + \mathscr{N} = \mathscr{A}$. 结论证毕. □

作为练习, 请读者自己以矩阵的形式叙述定理 5.2.15 中的结论.

思考题 5.13. 设 T 是有限维 K-向量空间 V 上的可逆线性变换. 证明:

1. T 的特征多项式常数项一定不等于 0.
2. 存在多项式 $g \in K[X]$ 使得 $T^{-1} = g(T)$. (因此, 如果 T 可以表示为某线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 的多项式, 则 T^{-1} 也可以写成 $\mathscr{A}$ 的多项式.)

思考题 5.14. 仿照注记 5.2.12 中的讨论, 通过多项式理论提供的一些额外结论, 解释乘法式 Jordan–Chevalley 分解的唯一性. (提示: 思考题 5.13 可能会有用处.)

读者还可以通过加法式 Jordan–Chevalley 分解的唯一性来论证乘法式 Jordan–Chevalley 分解的唯一性 (习题 5.2.21).

Jordan 标准形的另一个重要应用是矩阵幂的计算. 具体来说, 如果我们能将矩阵 $A \in \mathbf{M}_n(K)$ 化为 Jordan 标准形 J, 并找到了可逆矩阵 P 使得 $P^{-1}AP = J$, 那么对于任意 $k \in \mathbb{N}$ 我们得到 $A^k = PJ^kP^{-1}$. 所以只要能计算出 J^k 就能算出 A^k. 于是问题归结为 Jordan 形矩阵幂的计算. 而这又化为 Jordan 块幂的计算. 事实上, 若 $J = J_m(\lambda)$ 为 m 阶 Jordan 块, 其特征值为 λ, 则 $N := J - \lambda I_m = J_m(0)$ 是幂零阶为 m 的幂零 Jordan 块. 对于任意 $k \in \mathbb{N}$, 我们有

$$J^k = (\lambda I + N)^k = \sum_{i=0}^{k} \lambda^{k-i} \binom{k}{i} N^i \,.$$

因为 $N^m = 0$, 所以上式右边的求和式中通项在 $i \geq m$ 时一定为 0. 这一点经常会给实际计算带来很大的便利.

例 5.2.16. 令 $A = \begin{pmatrix} 2 & 6 & -15 \\ 1 & 1 & -5 \\ 1 & 2 & -6 \end{pmatrix}$. 可以算出 A 的特征多项式为 $f(X) = (X+1)^3$, 从而知道 $\lambda = -1$ 是 A 的唯一特征值, 而 $A - \lambda I = \begin{pmatrix} 3 & 6 & -15 \\ 1 & 2 & -5 \\ 1 & 2 & -5 \end{pmatrix}$. 若记 $\mathscr{A} : K^3 \to K^3$ 为线性变换 $X \mapsto AX$, 则

$$\mathrm{Ker}(\mathscr{A}-\lambda I) = \mathrm{span}(5e_1+e_3,\ -2e_1+e_2) = \mathrm{span}(5e_1+e_3,\ 3e_1+e_2+e_3)\ ,\ \mathrm{Im}(\mathscr{A}-\lambda I) = \mathrm{span}(3e_1+e_2+e_3)\ .$$

所以 A 的 Jordan 标准形有两个 Jordan 块, 即 $J = \begin{pmatrix} -1 & & \\ & -1 & 1 \\ & & -1 \end{pmatrix}$. 欲求的 Jordan 基形如 v_1, $(\mathscr{A} - \lambda I)v_2$, v_2. 其中一种选择是

$$v_1 = 5e_1 + e_3\ ,\ (\mathscr{A} - \lambda I)v_2 = 3e_1 + e_2 + e_3\,,\ v_2 = e_1\,.$$

于是, 若令 $P=\begin{pmatrix}5&3&1\\0&1&0\\1&1&0\end{pmatrix}$, 则 $P^{-1}=\begin{pmatrix}0&-1&1\\0&1&0\\1&2&-5\end{pmatrix}$, $P^{-1}AP=J$.

于是

$$J^2=\lambda^2 I+2\lambda\begin{pmatrix}0&&\\&0&1\\&&0\end{pmatrix}+\begin{pmatrix}0&&\\&0&1\\&&0\end{pmatrix}^2=\begin{pmatrix}1&&\\&1&-2\\&&1\end{pmatrix}$$

$$\cdots\cdots\cdots$$

$$J^k=\begin{pmatrix}(-1)^k&&\\&(-1)^k&(-1)^{k-1}k\\&&(-1)^k\end{pmatrix}.$$

因此

$$\begin{aligned}A^k=PJ^kP^{-1}&=\begin{pmatrix}5&3&1\\0&1&0\\1&1&0\end{pmatrix}\begin{pmatrix}(-1)^k&&\\&(-1)^k&(-1)^{k-1}k\\&&(-1)^k\end{pmatrix}\begin{pmatrix}0&-1&1\\0&1&0\\1&2&-5\end{pmatrix}\\&=\begin{pmatrix}(-1)^{k-1}3k+(-1)^k&(-1)^{k-1}6k&(-1)^k15k\\(-1)^{k-1}k&(-1)^k+(-1)^{k-1}2k&(-1)^k5k\\(-1)^{k-1}k&(-1)^{k-1}2k&(-1)^k+(-1)^k5k\end{pmatrix}.\end{aligned}$$

以上这个例子计算过程应该不算复杂, 但很具有代表性. 希望读者认真体会其中的思想方法. ■

5.2.4 最小多项式

我们已经知道, 特征多项式是研究线性变换的一个重要工具. 在这一小节中, 我们将再介绍一种与线性变换密切相关的多项式, 并特别关注它与 Jordan 标准形之间的联系.

在正式介绍相关的理论之前, 我们先来温习一个熟悉的例子. 如本节开头所述, 我们总是假定 V 是有限维 K-向量空间, 且 $\dim V\geq 1$.

(5.2.17) 设 $n=\dim V$. 根据引理 5.1.15, V 上的任何幂零变换的特征多项式都相同, 都等于 X^n. 但显然, V 上不同的幂零变换可能具有不同的 Jordan 标准形. 例如, K^3 上存在两个幂零变换分别以下面两个矩阵为 Jordan 标准形:

$$J_1=\begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix},\quad J_2=\begin{pmatrix}0&0&0\\0&0&1\\0&0&0\end{pmatrix}$$

这里, J_1 是一个 Jordan 块, 而 J_2 则含有两个 Jordan 块. 如果从多项式的角度去观察二者的区别, 可以发现 $J_1^2\neq 0$, 但 $J_2^2=0$. 也就是说, 在多项式 $g(X)=X^2$ 中以 $X=J_2$ 代入时会得到零矩阵, 但是以 $X=J_1$ 代入时则不然.

事实上, 对于一般的两个线性变换 $\mathscr{A}$ 和 $\mathscr{B}$, 一旦存在多项式 f 使得 $f(\mathscr{A})\neq 0$ 而 $f(\mathscr{B})=0$, 那么 $\mathscr{A}$ 和 $\mathscr{B}$ 不可能具有相同的 Jordan 标准形 (参见下面的思考题 5.16). ■

前面的讨论引导我们做下面的定义.

定义 5.2.18. 设 $\mathscr{A} \in \mathrm{End}(V)$. 如果一个多项式 $f(X) \in K[X]$ 满足 $f(\mathscr{A}) = 0$, 我们就说 f 是 $\mathscr{A}$ 的一个**零化多项式或化零多项式** (annihilating polynomial). 我们将 $\mathscr{A}$ 的所有零化多项式构成的集合记为 $\mathbf{Ann}(\mathscr{A})$.[‖]

根据 Cayley–Hamilton 定理, $\mathscr{A}$ 的特征多项式是 $\mathscr{A}$ 的一个零化多项式, 而且它是个首一 (故非零) 多项式. 特别地, $\mathbf{Ann}(\mathscr{A})$ 一定存在非零多项式.

若 $n \geq 1$, $A \in \mathbf{M}_n(K)$. 我们将 $K[X]$ 中满足 $f(A) = 0$ 的多项式称为 A 在 K 上的**零化多项式**. 所有这些多项式构成的集合记为 $\mathbf{Ann}_K(A)$ 或简记为 $\mathbf{Ann}(A)$. ■

思考题 5.15. 记号如定义 5.2.18. 不使用 Cayley–Hamilton 定理, 你能否证明线性变换 $\mathscr{A}$ (或矩阵 $A \in \mathbf{M}_n(K)$) 总有非零的零化多项式?

定义 5.2.19. 设 $\mathscr{A} \in \mathrm{End}(V)$. 若 g 是 $\mathbf{Ann}(\mathscr{A})$ 中的首一多项式, 并且在 $\mathbf{Ann}(\mathscr{A})$ 所含的所有非零多项式中 g 的次数达到最小值, 则称 g 是 $\mathscr{A}$ 的**最小多项式或极小多项式** (minimal polynomial). 根据正整数集的良序性质 (参见第 14 页脚注或本书上册 (A.1.1)), $\mathscr{A}$ 的最小多项式总存在. 后面的推论 5.2.21 会说明其唯一性.

类似地, 对于任意 $A \in \mathbf{M}_n(K)$, 我们可以定义 A 在 K 上的**最小多项式**. ■

思考题 5.16. 设 $A, B \in \mathbf{M}_n(K)$. 证明: 如果 A 和 B 在 $\mathbb{C}$ 上相似, 则 $\mathbf{Ann}(A) = \mathbf{Ann}(B)$. 特别地, 相似的矩阵具有相同的最小多项式.

按照定义, 最小多项式是一个首一的零化多项式. 它的"最小性"其实不仅仅体现在多项式次数的大小比较方面, 还体现在多项式的整除关系上. (帮助读者回忆一下: 对于任意两个多项式 $f, g \in K[X]$, 我们说 g 整除 f 是指存在 $h \in K[X]$ 使得 $f = gh$. 此时我们也说 g 是 f 的一个因式, f 是 g 的一个倍式, 或者 f 能被 g 整除, 并记 $g \mid f$.)

命题 5.2.20: *设 $\mathscr{A}$ 是有限维 K-向量空间 V 上的线性变换, $g \in K[X]$ 是 $\mathscr{A}$ 的一个最小多项式.*

则对于任意 $f \in K[X]$, f 是 $\mathscr{A}$ 的零化多项式当且仅当 g 整除 f. 也就是说, 集合 $\mathbf{Ann}(\mathscr{A})$ 可以写成如下形式

$$\mathbf{Ann}(\mathscr{A}) = g.K[X] := \{gh \mid h \in K[X]\}.$$

证明. 如果 $f \in K[X]$ 能被 g 整除, 即存在 $h \in K[X]$ 使 $f = gh$, 则 $f(\mathscr{A}) = g(\mathscr{A})h(\mathscr{A}) = 0 \cdot h(\mathscr{A}) = 0$. 故 $f \in \mathbf{Ann}(\mathscr{A})$.

反过来, 假设 f 是 $\mathscr{A}$ 的一个零化多项式. 我们希望证明整除关系 $g \mid f$ 成立. 为此我们先将 f 除以 g, 得到带余除法式 (参见本书上册引理 A.4.5)

$$f = gq + r \quad 其中\ q, r \in K[X] \quad 且 \quad \deg(r) < \deg(g).$$

由于 f 和 g 都是 $\mathscr{A}$ 的零化多项式, 故 $r(\mathscr{A}) = f(\mathscr{A}) - q(\mathscr{A})g(\mathscr{A}) = 0 - 0 = 0$. 因此 r 也是 $\mathscr{A}$ 的零化多项式. 但是 $\deg(r) < \deg(g)$. 所以, 根据最小多项式的定义, r 只能是零多项式. 于是, 由 $f = qg + r = qg$ 即得 $g \mid f$. □

推论 5.2.21: *设 $\mathscr{A}$ 是有限维 K-向量空间 V 上的线性变换. 则 $\mathscr{A}$ 具有唯一的最小多项式.*

换成矩阵语言表述即为: $\mathbf{M}_n(K)$ 中任何矩阵的最小多项式都是唯一的.

证明. 若 f 和 g 都是 $\mathscr{A}$ 的最小多项式, 则命题 5.2.20 说明 f 和 g 互相整除. 由于最小多项式还要求是首一多项式, 因此必有 $f = g$. □

[‖]有的书在定义零化多项式时要求是非零多项式. 我们不做这一要求. 这样可以使得 $\mathbf{Ann}(\mathscr{A})$ 具有更好的代数结构. 例如, 我们定义的 $\mathbf{Ann}(\mathscr{A})$ 是 $K[X]$ 的一个线性子空间. (如果用抽象代数的语言, 可以更精确地说 $\mathbf{Ann}(\mathscr{A})$ 是多项式代数 $K[X]$ 中的一个理想 (ideal).)

思考题 5.17. 设 $A \in \mathbf{M}_n(K)$, $r \in \mathbb{N}^*$.

1. 证明: 如果在 $\mathbb{C}[X]$ 中 A 有一个次数不超过 r 的零化多项式, 那么在 $K[X]$ 中 A 也有一个次数不超过 r 的零化多项式. (提示: 考虑某个系数来自于矩阵 $A, A^2, \cdots, A^r$ 的线性方程组.)

2. 证明: A 在 K 上的最小多项式和 A 在 $\mathbb{C}$ 上的最小多项式是相同的.

例 5.2.22. 设 $N = J_m(0)$ 是个 m 阶的幂零 Jordan 块矩阵. 则对于任意常数 $a_0, \cdots, a_{m-1}$, 计算可得

$$a_0 I + a_1 N + \cdots + a_{m-1} N^{m-1} = \begin{pmatrix} a_0 & a_1 & \cdots & \cdots & a_{m-1} \\ & a_0 & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & a_1 \\ & & & & a_0 \end{pmatrix}.$$

这说明, 对于任意次数 $< m$ 的非零多项式 $P(X) = a_0 + a_1 X + \cdots + a_{m-1} X^{m-1}$ 均有 $P(N) \neq 0$. 所以 N 的最小多项式次数 $\geq m$. 另一方面, 多项式 X^m 在代入 N 时为零 (并且是首一多项式). 所以, 幂零 Jordan 块 $J_m(0)$ 的最小多项式为 X^m.

通过简单的变元代换可以看出, 对于一般的常数 λ, 以 λ 为特征值的 Jordan 块 $J_m(\lambda)$ 的最小多项式为 $(X-\lambda)^m$.

下面设 $J = \begin{pmatrix} J_{m_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{m_r}(\lambda_r) \end{pmatrix}$, 其中 $J_{m_i}(\lambda_i)$ 是以 λ_i 为特征值的 m_i 阶 Jordan 块.

假如 λ_i 两两不同, 则 J 的最小多项式为

$$g(X) = (X-\lambda_1)^{m_1} \cdots (X-\lambda_r)^{m_r}.$$

若不假设 λ_i 互异, 读者觉得以上结论仍成立吗? ■

(5.2.23) 假设 $A = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_r \end{pmatrix} \in \mathbf{M}_n(K)$ 是准对角阵, 其中 A_i 的最小多项式是 $g_i \in K[X]$. 对于任意 $f \in K[X]$, 容易看出 $f(A) = \begin{pmatrix} f(A_1) & & \\ & \ddots & \\ & & f(A_r) \end{pmatrix}$. 因此 $f(A) = 0$ 的充分必要条件是对于每个 $i \in [\![1, r]\!]$ 均有 $f(A_i) = 0$. 由命题 5.2.20 我们知道, $\mathbf{Ann}(A_i)$ 恰好由 g_i 的所有倍式构成. 因此, $\mathbf{Ann}(A)$ 恰好由同时能被每个 g_i 整除的多项式构成. 除了零多项式之外, 这些多项式中次数达到最低的多项式称为 $g_1, \cdots, g_r$ 这组多项式的一个**最小公倍式** (least common multiple). 以上讨论表明: 分块对角阵 $A = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_r \end{pmatrix}$ 的最小多项式就是 $g_1, \cdots, g_r$ 这组多项式的首一最小公倍式, 其中 g_i 是对角块 A_i 的最小多项式. ■

思考题 5.18. 利用 (5.2.23) 一段的结论来说明: 如果 J 是一般的 Jordan 形矩阵, $\lambda_1, \cdots, \lambda_r$ 是其所有不同的复特征值, 而 J 所含的以 λ_i 为特征值的 Jordan 块阶数最大值为 M_i, 则 J 的最小多项式为 $g(X) = (X-\lambda_1)^{M_1} \cdots (X-\lambda_r)^{M_r}$.

对于一般的矩阵 $A \in \mathbf{M}_n(K)$, 我们可以将其视为复矩阵, 然后在 $\mathbb{C}$ 上找出其 Jordan 标准形, 然后通过 Jordan 标准形算出 A 在 $\mathbb{C}$ 上的最小多项式. 根据思考题 5.17, 这样求出的也是 A 在 K 上的最小多项式. 所以, 我们通过先求 Jordan 标准形可以求出一般矩阵的最小多项式.

思考题 5.19. 设 J_1 和 J_2 为 Jordan 形矩阵, 其中的 Jordan 块都按照从左上到右下逐渐增大的方式排列. 假设 J_1 和 J_2 的特征多项式和最小多项式都相同. 是否可以断定 $J_1 = J_2$? 若是, 请解释为什么. 若否, 请举出反例.

接下来我们研究最小多项式和特征多项式之间的关系.

命题 5.2.24: 设 $f, g \in K[X]$ 分别为线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 的特征多项式和最小多项式.

1. 整除关系 $g \,|\, f$ 一定成立.

2. 对于任意复数 λ, λ 是 f 的根当且仅当 λ 是 g 的根. 也就是说, 在不计零点重数的意义下, 最小多项式和特征多项式在复数域中的零点集相同.

3. 存在正整数 m 使整除关系 $f \,|\, g^m$ 在 $K[X]$ 中成立. 其中一种可能的选择是 $m = \dim V$.

证明. (1) 因为特征多项式 f 是 $\mathscr{A}$ 的零化多项式 (Cayley–Hamilton 定理), 所以此结论是命题 5.2.20 的一个特殊情况.

(2) 由 (1) 可知, g 的复数根一定是 f 的根. 现在假设 λ 是 f 的一个复数根, 来证明 λ 也一定是 g 的复数根.

设 A 是 $\mathscr{A}$ 在某一组有序基下的矩阵. 在 $\mathbb{C}$ 上可以将 A 化为 Jordan 标准形 J. 则 f 和 g 分别是 J 的特征多项式和最小多项式. 根据假设, λ 是 f 的一个根, 也即是 J 的一个特征值. 因此它是 J 中某些 Jordan 块的特征值. 根据思考题 5.18, $X - \lambda$ 一定是 g 在 $\mathbb{C}[X]$ 中的一个因式, 因而 λ 必然是 g 的根.

(3) 假设特征多项式 f 在 $\mathbb{C}[X]$ 中分解为 $f(X) = (X-\lambda_1)^{m_1} \cdots (X-\lambda_r)^{m_r}$, 其中 $\lambda_1, \cdots, \lambda_r \in \mathbb{C}$ 互异. 则由 (2) 可知 $h(X) := (X-\lambda_1)\cdots(X-\lambda_r)$ 是 g 在 $\mathbb{C}[X]$ 中的一个因式. 而只要取正整数 $m \geq \max_{1\leq i\leq r}\{m_i\}$, 整除关系 $f \,|\, h^m$ 就会在 $\mathbb{C}[X]$ 中成立. 于是 $f \,|\, g^m$ 在 $\mathbb{C}[X]$ 中成立. 因为 f 和 g^m 都属于 $K[X]$, 所以 $f \,|\, g^m$ 在 $K[X]$ 中也成立 (参见本书上册注记 A.4.6). □

注记 5.2.25. 在学过足够多的多项式分解理论之后 (参见 5.3 节), 可以由命题 5.2.24 的结论 (1) 和 (3) 推知 f 和 g 具有相同的不可约因式 (只是不可约因式在二者的分解式中出现的幂次可以不同). 这一点可以有助于我们找出一个矩阵的最小多项式.

例如, 假设一个实矩阵 A 的特征多项式 f 在 $\mathbb{R}[X]$ 中具有如下形式的因式分解

$$f(X) = (X-\lambda_1)^{m_1} \cdots (X-\lambda_r)^{m_r} q_1(X)^{e_1} \cdots q_s(X)^{e_s}$$

其中 $\lambda_1, \cdots, \lambda_r$ 是互异的实数, $q_1, \cdots, q_s$ 是两两不同的二次首一多项式且都没有实数根. 则 A 的最小多项式 $g(X)$ 一定具有如下形式:

$$g(X) = (X-\lambda_1)^{k_1} \cdots (X-\lambda_r)^{k_r} q_1(X)^{d_1} \cdots q_s(X)^{d_s}$$

其中 $1 \leq k_i \leq m_i$, $1 \leq d_j \leq e_j$. 当然, 要确定每个 k_i 和 d_j 的值还需要具体实例具体分析. 当 m_i 和 e_j 的值都很小且数量不多时, 以上信息已经可以帮助我们将最小多项式锁定在少数几个可能性之内了. 余下只要再结合一些其它信息或经过几次计算检验即可求出最小多项式. ■

例 5.2.26. 设 $A = \begin{pmatrix} 4 & -5 & 7 \\ 1 & -4 & 9 \\ -4 & 0 & 5 \end{pmatrix}$.

若想求出 A 的最小多项式, 可以先算出其特征多项式 $f(X)=(X-1)(X^2-4X+13)$. 根据命题 5.2.24 和注记 5.2.25 中的讨论, A 的最小多项式必然也是 f.

再考虑矩阵 $B=\begin{pmatrix} 1 & -3 & 0 & 3 \\ -2 & -6 & 0 & 13 \\ 0 & -3 & 1 & 3 \\ -1 & -4 & 0 & 8 \end{pmatrix}$.

计算可知 B 的特征多项式为 $P(X)=(X-1)^4$. 故其最小多项式 $g(X)$ 必然形如 $g(X)=(X-1)^m$, 其中 $1\le m\le 4$. 逐次计算 $B-I$ 的幂可知 $g(X)=(X-1)^3$. ■

现在我们来说明: 从最小多项式的形状可以很容易判断线性变换或矩阵能否对角化.

命题 5.2.27: *设 $\mathscr{A}$ 是有限维 K-向量空间 V 上的线性变换, g 为 $\mathscr{A}$ 的最小多项式. 则以下陈述等价:*

1. *$\mathscr{A}$ 可对角化.*
2. *g 的所有复数根都属于 K 且 g 在 $\mathbb{C}$ 中没有重根.*

证明. 由命题 5.2.24 我们知道 g 的复数根与特征多项式的复数根相同. 于是, 根据推论 5.2.7, 我们只需要再说明 $\mathscr{A}$ 的 Jordan 标准形 J 是对角阵当且仅当最小多项式 $g(X)$ 无重根. 而这由思考题 5.18 的结论立即可得. □

例 5.2.28. 通过命题 5.2.27 来判别线性变换或矩阵是否可对角化有时是很方便的. 例如, 如果知道某个矩阵 A 满足 $A^3=A$, 则 $f=X^3-X=X(X+1)(X-1)$ 是 A 的一个零化多项式. 于是其最小多项式 g 是 f 的因式. 显然 f 没有重根, 因此 g 也没有重根. 于是由命题 5.2.27 我们可以轻易判断 A 一定可以对角化. 注意, 这个例子中的矩阵 A 可能是超过 3 阶的, 其特征多项式有可能是有重根的. (请读者尝试举出具体的例子.) ■

思考题 5.20. 设 $A\in \mathbf{M}_n(K)$. 请读者总结一下迄今为止学到的 A 在 K 上可对角化的充分必要条件有哪些.

最后我们来讨论一下最小多项式和特征多项式吻合的情况. 这其实和我们曾经提到的“循环不变子空间”概念有关. 事实上, 我们之前提到循环不变子空间是和 Jordan 标准形中的 Jordan 块相联系的 (参见第 23 页). 而一个 Jordan 块的最小多项式和特征多项式总是相同的.

定义 5.2.29. 设 $\mathscr{A}$ 是有限维向量空间 V 上的线性变换. 如果 V 的一个子空间 C 满足以下条件, 我们就说 C 是 $\mathscr{A}$ 的一个**循环不变子空间** (cyclic invariant subspace) 或者简称**循环子空间** (cyclic subspace):

存在非零向量 $v\in C$ 使得 C 可以由向量族 $\mathscr{A}^r v$, $r\in\mathbb{N}$ 生成, 即 $C=\operatorname{span}(v,\mathscr{A}v,\mathscr{A}^2v,\cdots)$.

此时我们也称 C 是由向量 v **循环生成的** (cyclically generated) 子空间. 该条件成立时我们也记 $C=K[\mathscr{A}]v$. 这是因为, C 中的元素恰好是所有具有以下形式的向量:

$$f(\mathscr{A})v=a_0v+a_1\mathscr{A}v+\cdots+a_N\mathscr{A}^Nv,$$

其中 $f(X)=a_0+\cdots+a_NX^N$ 是 $K[X]$ 中任意多项式.

如果 V 是 $\mathscr{A}$ 的一个循环不变子空间, 我们称线性变换 $\mathscr{A}$ 是**循环的**. ■

注意, 按照以上定义, 循环子空间要求是非零的, 而且很容易验证 $\mathscr{A}$ 的循环子空间都是 $\mathscr{A}$ 的不变子空间.

命题 5.2.30: 设 $\mathscr{A}$ 是 n 维向量空间 V 上的线性变换, C 是 $\mathscr{A}$ 的一个循环不变子空间, $m=\dim C$. 则一定存在 $v\in C$ 使得 $v, \mathscr{A}v, \cdots, \mathscr{A}^{m-1}v$ 构成 C 的一组基. 这样的基称为循环不变子空间 C 的一组**循环基** (cyclic basis).

特别地, $C=\mathrm{span}(v, \mathscr{A}v, \cdots, \mathscr{A}^{n-1}v)$.

证明. 由定义, 存在非零向量 $v\in C$ 使得向量族 $v, \mathscr{A}v, \cdots$ 生成 C. 我们来证 $v, \mathscr{A}v, \cdots, \mathscr{A}^{m-1}v$ 构成 C 的一组基. 因为该向量组所含的向量个数等于 C 的维数, 故只需证明它是线性无关的.

为此我们采用反证法. 假设该向量组线性相关, 即, 存在不全为零的常数 $a_0, \cdots, a_{m-1}$ 使得

$$a_0 v + a_1\mathscr{A}v + \cdots + a_{m-1}\mathscr{A}^{m-1}v = 0\,. \tag{5.2.30.1}$$

设 $s\in [\![0,\, m-1]\!]$ 是满足 $a_j\neq 0$ 的最大下标 j, 即

$$a_s\neq 0 \quad \text{但} \quad a_{s+1}=\cdots=a_{m-1}=0\,.$$

则 (5.2.30.1) 式表明

$$\mathscr{A}^s v = -a_s^{-1}(a_0 v+\cdots+a_{s-1}\mathscr{A}^{s-1}v) \;\in\; \mathrm{span}(v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{s-1}v)\,. \tag{5.2.30.2}$$

一方面, 若记 $C_0=\mathrm{span}(v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{s-1}v)$, 则 (5.2.30.2) 式蕴含着

$$\mathrm{span}(v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{s-1}v\,,\, \mathscr{A}^s v) = \mathrm{span}(v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{s-1}v) = C_0\,. \tag{5.2.30.3}$$

另一方面, 将 $\mathscr{A}$ 作用于 (5.2.30.2) 式再结合 (5.2.30.3) 可得

$$\begin{aligned}\mathscr{A}^{s+1}v \in \mathscr{A}\big(\mathrm{span}(v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{s-1}v)\big) &= \mathrm{span}(\mathscr{A}v\,, \cdots, \mathscr{A}^{s-1}v\,,\, \mathscr{A}^s v)\\ &\subseteq \mathrm{span}(v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{s-1}v\,,\, \mathscr{A}^s v) = \mathrm{span}(v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{s-1}v) = C_0\,.\end{aligned}$$

依此类推不难发现, 对于任意自然数 r 均有 $\mathscr{A}^r v\in \mathrm{span}(v,\, \mathscr{A}\,, \cdots, \mathscr{A}^{s-1}v)=C_0$. 也就是说, 无论用 $\mathscr{A}$ 反复作用于 v 多少次, 最终结果永远逃不出 C_0 这个子空间. 唐代张九龄《感遇十二首 · 其七》诗云: "运命唯所遇, 循环不可寻". 世间运命因果之循环, 大抵如此.

以上讨论表明

$$C=\mathrm{span}(v,\, \mathscr{A}v\,, \cdots) = \mathrm{span}\big((\mathscr{A}^r v)_{r\in\mathbb{N}}\big) \subseteq \mathrm{span}(v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{s-1}v) = C_0\,.$$

于是 $m=\dim C\leq \dim C_0\leq s$. 但 $s\leq m-1$, 所以这是一个矛盾. 命题至此证毕. □

引理 5.2.31: 设 $\mathscr{A}$ 为有限维 K-向量空间 V 上的线性变换. 如果 $\mathscr{A}$ 是循环的 (即, V 是 $\mathscr{A}$ 的循环子空间), 则 $\mathscr{A}$ 的特征多项式和最小多项式相等.

证明. 设 $n=\dim V$. 我们只需要证明 $\mathscr{A}$ 的最小多项式次数为 n (请读者回答为什么). 若不然, 则存在不全为零的常数 $a_0, \cdots, a_{n-1}$ 使得

$$a_0 I + a_1\mathscr{A} + \cdots + a_{n-1}\mathscr{A}^{n-1} = 0\,.$$

但命题 5.2.30 告诉我们 V 有一组 $v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{n-1}v$ 形式的循环基. 将以上等式作用于向量 v 可得

$$a_0 v + a_1\mathscr{A}v + \cdots + a_{n-1}\mathscr{A}^{n-1}v = 0\,.$$

这与向量组 $v,\, \mathscr{A}v\,, \cdots, \mathscr{A}^{n-1}v$ 的线性无关性矛盾. □

在 Jordan 标准形存在的情况下, 特征多项式与最小多项式吻合的一些等价条件列举在下面的命题中.

命题 5.2.32: 设 $\mathscr{A}$ 为 n 维 K-向量空间 V 上的线性变换. 假设 $\mathscr{A}$ 的特征多项式的所有复数根属于 K. 则下列陈述等价:

1. $\mathscr{A}$ 的特征多项式和最小多项式相同.

2. $\mathscr{A}$ 的最小多项式次数为 n.

3. 对于 $\mathscr{A}$ 的每个特征值 λ, 在 $\mathscr{A}$ 的 Jordan 标准形中仅有一个 Jordan 块以 λ 为特征值. (这相当于说 λ 的几何重数一定是 1.)

4. 对于 $\mathscr{A}$ 的每个特征值 λ, 广义特征子空间 $G(\lambda, \mathscr{A})$ 是 $\mathscr{A}$ 的循环子空间.

证明. 作为练习, 请读者证明前三个论断等价. 我们只需再证明 (3) 可以推出 (4), (4) 可以推出 (2).

设 $\lambda \in K$ 为 $\mathscr{A}$ 的一个特征值, $W = G(\lambda, \mathscr{A})$, $\mathscr{N} := (\mathscr{A} - \lambda I)|_W$. 若条件 (3) 成立, 则由命题 5.1.20 可知 $\mathscr{N}$ 的任意一组 Jordan 基形如 $\mathscr{N}^{m-1}v, \cdots, \mathscr{N}v, v$. 我们只需证明 $v, \mathscr{A}v, \cdots, \mathscr{A}^{m-1}v$ 构成 W 的一组基. 而这只需要 (请读者解释为什么) 验证子空间 $U := \mathrm{span}(v, \mathscr{A}v, \cdots, \mathscr{A}^{m-1}v)$ 包含每个 $\mathscr{N}^j v$, 其中 $j = 0, 1, \cdots, m-1$.

简单的观察可知 $\mathscr{N}^0 v = v \in \mathrm{span}(v) \subseteq U$, $\mathscr{N}v = \mathscr{A}v - \lambda v \in \mathrm{span}(v, \mathscr{A}v) \subseteq U$. 假如已知

$$\mathscr{N}^{j-1}v \ \in \ \mathrm{span}(v, \mathscr{A}v, \cdots, \mathscr{A}^{j-1}v)$$

则

$$\begin{aligned}
\mathscr{N}^j v = \mathscr{N}(\mathscr{N}^{j-1}v) \ &\in \mathscr{N}\big(\mathrm{span}(v, \mathscr{A}v, \cdots, \mathscr{A}^{j-1}v)\big) \\
&= \mathrm{span}(\mathscr{N}v, \mathscr{N}\mathscr{A}v, \cdots, \mathscr{N}\mathscr{A}^{j-1}v) = \mathrm{span}(\mathscr{N}v, \mathscr{A}\mathscr{N}v, \cdots, \mathscr{A}^{j-1}\mathscr{N}v) \\
&= \mathrm{span}((\mathscr{A} - \lambda I)v, \mathscr{A}(\mathscr{A} - \lambda I)v, \cdots, \mathscr{A}^{j-1}(\mathscr{A} - \lambda I)v) \\
&= \mathrm{span}(\mathscr{A}v - \lambda v, \mathscr{A}^2 v - \lambda\mathscr{A}v, \cdots, \mathscr{A}^j v - \lambda\mathscr{A}^{j-1}v) \\
&\subseteq \mathrm{span}(v, \mathscr{A}v, \cdots, \mathscr{A}^{j-1}v, \mathscr{A}^j v).
\end{aligned}$$

因此由归纳法得证 $\mathscr{N}^j v \in \mathrm{span}(v, \mathscr{A}v, \cdots, \mathscr{A}^j v)$ 对于每个 j 成立. 因此, $v, \mathscr{N}v, \cdots, \mathscr{N}^{m-1}v$ 均包含于 $U = \mathrm{span}(v, \mathscr{A}v, \cdots, \mathscr{A}^{m-1}v)$. 至此我们证明了 (3) $\Rightarrow$ (4).

反过来, 假设 (4) 成立. 记 $\lambda_1, \cdots, \lambda_r$ 为 $\mathscr{A}$ 的所有特征值, $W_i = G(\lambda_i, \mathscr{A})$, $\mathscr{A}_i = \mathscr{A}|_{W_i}$. 根据引理 5.2.31, 每个 $\mathscr{A}_i$ 的最小多项式 $g_i(X)$ 次数等于 λ_i 的代数重数 $m_i = \dim W_i$. 由于 λ_i 是 $\mathscr{A}_i$ 唯一的特征值, 故 $g_i(X) = (X - \lambda_i)^{m_i}$. 这说明对于不同的下标 $i, j \in [\![1, r]\!]$, g_i 和 g_j 之间没有公共的不可约因式. 于是由 (5.2.23) 一段的结论可以知道 $\mathscr{A}$ 的最小多项式为 $g_1 \cdots g_r$, 因此其次数等于所有特征值代数重数之和, 即 n. 所以我们证明了 (4) $\Rightarrow$ (2). 命题的证明至此完成. □

思考题 5.21. 证明命题 5.2.32 中的 (1)–(3) 三个论断等价.

在实际应用中, 我们基本上都是通过矩阵来计算最小多项式和特征多项式. 而对于一个矩阵 $A \in \mathbf{M}_n(K)$, 可以将其当作 $\mathbb{C}$ 上的矩阵来计算. 因此, 以上命题 5.2.32 中的条件对于矩阵 A 来说实际上都可以在 $\mathbb{C}$ 上进行验证 (参见思考题 5.17). 所以, Jordan 标准形的存在性本质上不会制造任何麻烦. 我们在 $\mathbb{C}$ 上计算出 Jordan 标准形后就可以判断最小多项式是什么样子了.

注记 5.2.33. 引理 5.2.31 的逆命题也成立. 即, 如果线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 的特征多项式和最小多项式相同, 则 $\mathscr{A}$ 是循环变换.

不过, 以上这个结论的证明似乎并不容易. 有兴趣的读者可以参看 [35, 第 192 页, 定理 7.8 的系 1]. 如果额外假设 $\mathscr{A}$ 在 K 上存在 Jordan 标准形 (即 $\mathscr{A}$ 的特征多项式复数根均属于 K), 那么习题 5.2.22 会给出一个较为简单的证明. ■

5.2.5 习题

习题 5.2.1. 证明: 如果 $\mathscr{A}$ 是有限维向量空间 V 上的幂零变换, 则 $\mathscr{A}$ 是循环的当且仅当 $\mathscr{A}$ 的幂零阶等于 $\dim V$.

习题 5.2.2. 令 $V = K[X]_{\leq n}$. 通过求多项式的形式导数定义线性变换 $\mathscr{D} : V \to V$, 即,

$$\mathscr{D}(1) = 0\,,\quad \text{对 } k \in [\![1,\, n]\!],\ \ \mathscr{D}(X^k) = kX^{k-1}\,.$$

证明 $\mathscr{D}$ 是一个循环的幂零变换, 并写出一组循环基.

习题 5.2.3. 设 $\mathscr{A}$ 是 n 维向量空间 V 上的循环幂零变换, $\varepsilon_1, \cdots, \varepsilon_n$ 是它的一组循环基. 试求 $\mathscr{A}$ 的所有不变子空间.

习题 5.2.4. 设 $\mathscr{A}$ 是有限维向量空间 V 上的幂零变换. 假设 $\mathscr{A}$ 有两个线性无关的特征向量. 证明 $\mathscr{A}$ 不是循环的.

习题 5.2.5. 将以下矩阵 A 化为 Jordan 标准形:

$$(1)\ \ A = \begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix}\ ;\quad (2)\ \ A = \begin{pmatrix} 1 & -3 & 3 \\ -2 & -6 & 13 \\ -1 & -4 & 8 \end{pmatrix}\ ;\quad (3)\ \ A = \begin{pmatrix} 4 & 6 & -15 \\ 1 & 3 & -5 \\ 1 & 2 & -4 \end{pmatrix}\ ;$$

$$(4)\ \ A = \begin{pmatrix} 3 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 3 & 0 & 5 & -3 \\ 4 & -1 & 3 & -1 \end{pmatrix}\ ;\quad (5)\ \ A = \begin{pmatrix} 3 & -4 & 0 & 2 \\ 4 & -5 & -2 & 4 \\ 0 & 0 & 3 & -2 \\ 0 & 0 & 2 & -1 \end{pmatrix}\ ;\quad (6)\ \ A = \begin{pmatrix} 1 & -3 & 0 & 3 \\ -2 & -6 & 0 & 13 \\ 0 & -3 & 1 & 3 \\ -1 & -4 & 0 & 8 \end{pmatrix}$$

习题 5.2.6. 将以下矩阵 A 化为 Jordan 标准形:

$$(1)\ \ A = \begin{pmatrix} 2 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}\ ;\quad (2)\ \ A = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 0 \\ 1 & 1 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

习题 5.2.7. 令 $A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 0 & 3 \\ 0 & 1 & 0 \end{pmatrix}$. 对每个 $n \in \mathbb{N}$ 求出 A^n.

习题 5.2.8. 设 5 阶方阵 A 满足下列条件:

$$\operatorname{rank}(A) = 3\,,\ \operatorname{rank}(A^2) = 2\,,\ \operatorname{rank}(A + I) = 4\,,\ \operatorname{rank}(A + I)^2 = 3\,.$$

求 A 的 Jordan 标准形.

习题 5.2.9. 设 $\mathscr{A}$ 是有限维复向量空间 V 上的线性变换, J 为 $\mathscr{A}$ 的 Jordan 标准形. 设 λ_0 是 $\mathscr{A}$ 的一个特征值, $\mathscr{B} := A - \lambda_0 I$.

1. 对每个 $i \in \mathbb{N}$, 令 $M_i = \operatorname{Ker}(\mathscr{B}^i)$, $k := \min\{i \in \mathbb{N} \mid M_i = M_{i+1}\}$. 证明: k 等于 J 中以 λ_0 为特征值的 Jordan 块的最大阶数.

2. 设 k 如前一小题. 令 $N_k = \operatorname{Im}(\mathscr{B}^k)$. 证明 λ_0 不是 $\mathscr{A}|_{N_k}$ 的特征值, 因此 $\mathscr{B}|_{N_k}$ 是可逆变换.

3. 证明 $\dim M_k$ 等于特征值 λ_0 的 (代数) 重数.

4. 设 λ_1 也是 $\mathscr{A}$ 的特征值, $\lambda_1 \neq \lambda_0$. 证明 $G(\lambda_1, \mathscr{A}) \subseteq \mathrm{Im}(\mathscr{B}^k) = N_k$.

习题 5.2.10. 设 $A \in \mathbf{M}_n(\mathbb{C})$ 且存在正整数 k 使得 $A^k = I_n$. 证明 A 在 $\mathbb{C}$ 上可对角化.

习题 5.2.11. 证明: 任意复方阵 A 与它的转置 A^T 在 $\mathbb{C}$ 上相似.

习题 5.2.12. 求以下 n 阶矩阵在 $\mathbb{C}$ 上的 Jordan 标准形:

$$(1)\ \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & -1 \\ & & & & 1 \end{pmatrix};\quad (2)\ \begin{pmatrix} 0 & a & & & \\ 0 & 0 & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \ddots & \ddots & 0 & a \\ a & 0 & \cdots & \cdots & 0 \end{pmatrix} \text{ 其中 } a \text{ 是非零常数.}$$

习题 5.2.13. 设 $a_{12}, a_{23}, \cdots, a_{n-1,n}$ 是非零复数. 求复矩阵

$$\begin{pmatrix} a & a_{12} & a_{13} & \cdots & a_{1n} \\ & a & a_{23} & \cdots & a_{2n} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & a_{n-1,n} \\ & & & & a \end{pmatrix}$$

的 Jordan 标准形.

习题 5.2.14. 设 J 是 n 阶幂零的 Jordan 块. 对每个 $k \in [\![1, n]\!]$ 求出 J^k 的 Jordan 标准形.

习题 5.2.15. 设 $a \in K$. 求 $\mathbf{M}_n(K)$ 中最小多项式为 $X - a$ 的所有矩阵.

习题 5.2.16. 设 $J = J_n(0)$, $V = \mathbf{M}_n(K)$. 通过矩阵乘法定义线性变换 $\mathscr{A}: V \to V$, $M \mapsto JM$. 求 $\mathscr{A}$ 的最小多项式和特征多项式.

习题 5.2.17. 求以下矩阵 A 的最小多项式.

$$A = \begin{pmatrix} 0 & \cdots & 0 & -a_n \\ 1 & \ddots & \vdots & \vdots \\ & \ddots & 0 & -a_2 \\ & & 1 & -a_1 \end{pmatrix}$$

习题 5.2.18. 设 $A = (a_{ij}) \in \mathbf{M}_n(K)$, 其中对每个 $i, j \in [\![1, n]\!]$, $a_{ij} = 1$. 求 A 的最小多项式.

习题 5.2.19. 设 $\mathscr{A} \in \mathrm{End}(V)$ 在 V 的一组有序基 $\varepsilon_1, \cdots, \varepsilon_n$ 下的矩阵是 $A = \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}$, 其中

$$B = \begin{pmatrix} \lambda_1 & -1 & & \\ & \lambda_1 & \ddots & \\ & & \ddots & -1 \\ & & & \lambda_1 \end{pmatrix} \in \mathbf{M}_r(K), \quad C = \begin{pmatrix} \lambda_2 & 0 & 1 & & \\ & \lambda_2 & 0 & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & \ddots & 0 \\ & & & & \lambda_2 \end{pmatrix} \in \mathbf{M}_{n-r}(K).$$

1. 求 $\mathscr{A}$ 在 V 中的一组 Jordan 基.

2. 求 $\mathscr{A}$ 的最小多项式.

习题 5.2.20. 设 V 是有限维 K-向量空间.

1. 假设线性变换 $\mathscr{D}, \mathscr{D}' \in \mathrm{End}(V)$ 均可对角化, 并且二者的特征值 (不计重数) 集合均为 $\{\lambda_1, \cdots, \lambda_r\}$.

   证明: $\mathscr{D} = \mathscr{D}'$ 当且仅当对每个 $i \in [\![1, r]\!]$ 均有 $E(\lambda_i, \mathscr{D}) = E(\lambda_i, \mathscr{D}')$.

2. 假设 $\mathscr{A} \in \mathrm{End}(V)$ 有一种 Jordan–Chevalley 分解式 $\mathscr{A} = \mathscr{D} + \mathscr{N}$, 即, $\mathscr{D} \in \mathrm{End}(V)$ 可对角化, $\mathscr{N} \in \mathrm{End}(V)$ 是幂零变换, 且 $\mathscr{D}, \mathscr{N}$ 可交换. 设 $\lambda_1, \cdots, \lambda_r$ 是 $\mathscr{D}$ 的所有不同特征值.

   (a) 对每个 $i \in [\![1, r]\!]$, 令 $T_i := \mathscr{A} - \lambda_i I$. 证明: $E(\lambda_i, \mathscr{D})$ 是 $\mathscr{N}$ 和 T_i 的不变子空间, $\mathscr{N}|_{E(\lambda_i, \mathscr{D})} = T_i|_{E(\lambda_i, \mathscr{D})}$ 且 $E(\lambda_i, \mathscr{D}) \subseteq G(\lambda_i, \mathscr{A})$.

   (b) 证明: $\lambda_1, \cdots, \lambda_r$ 恰为 $\mathscr{A}$ 的特征多项式的所有不同复数根, 而且 $E(\lambda_i, \mathscr{D}) = G(\lambda_i, \mathscr{A})$.

3. 证明定理 5.2.11 中加法式 Jordan–Chevalley 分解式的唯一性[**].

习题 5.2.21. 利用加法式 Jordan–Chevalley 分解的唯一性证明乘法式 Jordan–Chevalley 分解的唯一性.

习题 5.2.22. 设 V 是 n 维 K-向量空间. 假设线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 的特征多项式 $P_A(X)$ 的所有复数根 $\lambda_1, \cdots, \lambda_r$ 都属于 K. 对于每个 $i \in [\![1, r]\!]$, 记 $m_i = \dim G(\lambda_i, \mathscr{A})$. 假设 $\mathscr{A}$ 的最小多项式等于其特征多项式 $P_A(X)$. 根据命题 5.2.32, 可以取到向量 $v_i \in G(\lambda_i, \mathscr{A})$ 使得 $\mathscr{A}^{m_i-1}v_i, \cdots, \mathscr{A}v_i, v_i$ 构成 $G(\lambda_i, \mathscr{A})$ 的一组基.

1. 对任意多项式 $f \in K[X]$, 证明: 如果 $f(\mathscr{A})v_i = 0$, 则对任何 $u \in G(\lambda_i, \mathscr{A})$ 均有 $f(\mathscr{A})u = 0$.

2. 令 $v = v_1 + \cdots + v_r$.

   证明向量组 $v, \mathscr{A}v, \cdots, \mathscr{A}^{n-1}v$ 是线性无关的. 由此证明 $\mathscr{A}$ 是循环变换.[††].

   (提示: 对任意多项式 $f \in K[X]$, $G(\lambda_i, \mathscr{A})$ 是 $f(\mathscr{A})$ 的不变子空间, $f(\mathscr{A})v = f(\mathscr{A})v_1 + \cdots + f(\mathscr{A})v_r$.)

5.3 * 多项式在 Jordan–Chevalley 分解中的应用

本节为选学内容, 以后适时补充.

[**]这个证明是作者从南科大 2020 级本科生章芃栩同学那里获知的.

[††]这个证明也是作者从南科大 2020 级本科生章芃栩同学那里获知的.

第六章　双线性型与二次型

在之前的章节中, 我们的主要精力集中于研究向量空间的自身线性结构以及向量空间上定义的 (一元) 线性映射. 接下来我们关心的问题是, 在一般的抽象向量空间上是否还有其他有意思的结构值得研究？当然, 从线性代数的角度来说, 我们所指的其他结构应该和向量空间本身的线性结构具有紧密联系并且有适当的协调性. 如果我们考虑最熟悉的向量空间 $\mathbb{R}^n$, 特别是坐标几何中常用的 $\mathbb{R}^2$ 和 $\mathbb{R}^3$, 那么很容易想到这类其他结构的例子. 例如, $\mathbb{R}^n$ 中的任意两个向量之间有一种内积运算 (当 $n = 2$ 或 3 时该运算还有很明确的几何直观), 它的定义其实和抽象向量空间自带的两种基本运算 (加法和数乘) 没有直接的依赖性, 但内积的实用性是众所周知的. 如何在一般的抽象向量空间上引入类似于 $\mathbb{R}^n$ 上内积的运算, 以及如何利用这样的运算进一步研究向量空间及其线性变换的结构, 这将是我们在本章和下一章要讨论的中心课题.

本章中总是以 K 表示 $\mathbb{C}$ 的一个子域.

6.1　双线性型的基础理论

6.1.1　双线性型及其矩阵

对于本章引言部分提到的 $\mathbb{R}^n$ 空间上的内积, 我们熟知它关于两个向量变元分别都是线性的. 也就是说, 如果将 $\mathbb{R}^n$ 上的内积运算视为一个映射

$$\varphi\,:\,\mathbb{R}^n\times\mathbb{R}^n\longrightarrow\mathbb{R}\,;\quad (u\,,\,v)\longmapsto\varphi(u\,,\,v):=u\cdot v\,,$$

那么对于任意 $u\,,\,v\,,\,w\in\mathbb{R}^n$ 和任意常数 $c\in\mathbb{R}$, 以下性质成立:

$$\begin{aligned}&\varphi(u+v\,,\,w)=\varphi(u\,,\,w)+\varphi(v\,,\,w)\,,\ \varphi(w\,,\,u+v)=\varphi(w\,,\,u)+\varphi(w\,,\,v)\,,\\&\varphi(cu\,,\,v)=c\,\varphi(u\,,\,v)=\varphi(u\,,\,cv)\,.\end{aligned}$$

此外, 内积还具有对称性. 即上述映射 φ 满足:

$$\text{对于所有 } u\,,\,v\in\mathbb{R}^n\,,\quad \varphi(u\,,\,v)=\varphi(v\,,\,u)\,.$$

以上表述的性质可以很容易地做抽象的推广.

定义 6.1.1. 设 V 为任意 K-向量空间. 考虑一个映射 $\varphi: V\times V\longrightarrow K$.

1. 如果 φ 满足以下两个性质:

   (a) 对任意 $u, v, w\in V$, 均有 $\varphi(u+v\,,\,w)=\varphi(u\,,\,w)+\varphi(v\,,\,w)$ 及 $\varphi(w\,,\,u+v)=\varphi(w\,,\,u)+\varphi(w\,,\,v)$.

   (b) 对任意 $u, v\in V$ 和任意 $c\in K$, 均有 $\varphi(cu\,,\,v)=c\varphi(u\,,\,v)=\varphi(u\,,\,cv)$.

那么我们说 φ 是 V 上的一个**双线性型** (bilinear form) 或者**双线性函数** (bilinear function).

文献中关于双线性型的一个常用记号是将 φ 写成 $\langle \cdot\,,\,\cdot \rangle$ 的样子, 即对任意向量 $u, v \in V$, 记 $\langle u\,,\,v \rangle = \varphi(u, v)$.

2. 如果对于任意 $u, v \in V$ 均有 $\varphi(u, v) = \varphi(v, u)$, 则称 φ 是**对称的** (symmetric).

3. 如果对于任意 $u, v \in V$ 均有 $\varphi(u, v) = -\varphi(v, u)$, 则称 φ 是**反对称的** (anti-symmetric) 或 **斜对称的** (skew-symmetric).

4. 如果对于任意 $v \in V$ 均有 $\varphi(v, v) = 0$, 则称 φ 是**交错的**或者**交错型的** (alternating).

当 φ 是 V 上的 (对称、反对称或交错) 双线性型时, 也称序对 (V, φ) 为一个 (对称、反对称或交错) **双线性空间** (bilinear space), 将 $\dim V$ 称为双线性空间 (V, φ) 的**维数**. 通常我们可以忽略 $\dim V = 0$ 的情况.

我们分别以 $\mathrm{Bil}(V)$, $\mathrm{Sym}(V)$ 和 $\mathrm{Alt}(V)$ 表示 V 上所有的双线性型、对称双线性型和交错双线性型构成的集合. (因为我们即将在命题 6.1.3 中证明斜对称的双线性型和交错的双线性型是一回事, 所以我们不再为斜对称的双线性型集合专门固定记号.) ■

细心的读者可能会发现, 以上我们定义的双线性型其实只是本书上册定义 4.2.1 的一个特殊情况.

(6.1.2) 对于任意向量空间 V, 我们知道集合 $V \times V$ 本身具有一种自然的向量空间结构, 即按照向量空间直积的构造方式 (参见本书上册 (2.3.38)) 将 $V \times V$ 上的加法和数乘按照逐个分量做加法和数乘的方式定义. 但在这里我们提醒读者留意, 在定义 V 上的双线性函数时, 我们并不把 $V \times V$ 整体上当作一个向量空间, 而仅仅看作集合意义上的笛卡尔积. 如果将 $V \times V$ 视为直积方式得到的向量空间, 一个双线性函数 $\varphi : V \times V \to K$ 并不是从 $V \times V$ 到 K 的线性函数. 这是因为, 对于任意 $u_1, u_2, v_1, v_2 \in V$, 双线性函数 φ 会满足

$$\begin{aligned}\varphi(u_1 + u_2\,,\, v_1 + v_2) &= \varphi(u_1, v_1 + v_2) + \varphi(u_2, v_1 + v_2) \\ &= \varphi(u_1, v_1) + \varphi(u_1, v_2) + \varphi(u_2, v_1) + \varphi(u_2, v_2)\,.\end{aligned}$$

而一个线性函数 $f : V \times V \to K$ 则应满足

$$f(u_1 + u_2, v_1 + v_2) = f\big((u_1, v_1) + (u_2, v_2)\big) = f(u_1, v_1) + f(u_2, v_2)\,.$$

举个最简单的例子来看, 若将 $\mathbb{R}$ 视为一个实向量空间, 则映射

$$\varphi : \ \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}\,; \quad (x, y) \longmapsto \varphi(x, y) := xy$$

是一个双线性函数. 但如果将 $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ 视为实向量空间, 这个映射 φ 显然不是 $\mathbb{R}^2$ 到 $\mathbb{R}$ 的线性函数. ■

命题 6.1.3: 设 φ 为向量空间 V 上的双线性型.

1. φ 是斜对称的当且仅当它是交错的.

2. 如果 φ 既是对称的又是交错的, 则必然 $\varphi = 0$, 即对于任意 $u, v \in V$ 均有 $\varphi(u, v) = 0$.

证明. (1) 映射 φ 的双线性说明

(6.1.3.1) $$\varphi(u, v) + \varphi(v, u) = \varphi(u + v, u + v) - \varphi(u, u) - \varphi(v, v)$$

对所有 $u, v \in V$ 成立.

若 φ 是交错的, 则按定义可知 (6.1.3.1) 式右端恒等于 0. 而左端恒等于 0 意味着 φ 是斜对称的.

反之, 若 φ 是斜对称的, 则 (6.1.3.1) 式左端恒等于 0. 而在 (6.1.3.1) 式右端取 $u = v$ 并利用定义 6.1.1 (1) 中的条件 (b) 可知

$$0 = \varphi(2v, 2v) - 2\varphi(v, v) = 2\varphi(v, 2v) - 2\varphi(v, v) = 2 \cdot 2\varphi(v, v) - 2\varphi(v, v) = 2\varphi(v, v)\,.$$

因此 $\varphi(v, v)$ 总是等于 0, 即 φ 是交错的.

(2) 留给读者练习. □

例 6.1.4. 来举几个简单的双线性型的例子.

1. 令 $V = K^{2\times 1}$ 为 2 维列向量空间. 则行列式映射

$$\det\,:\,V \times V \longrightarrow K\,;\quad \left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right) \longmapsto \det\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

   是 V 的交错双线性型.

2. 令 $V = K[X]$. 映射

$$V \times V \longrightarrow K\,;\quad (f, g) \longmapsto \langle\, f, g\,\rangle := f(0)g(0)$$

   是 V 上的对称双线性型.

3. 设 $A \in \mathbf{M}_n(K)$, $V = K^{n\times 1}$. 则映射

$$\langle\,\cdot\,,\,\cdot\,\rangle_A\,:\,V \times V \longrightarrow K\,;\quad (x, y) \longmapsto x^T A y$$

   是 V 上的双线性型.

   如果 A 是对称矩阵, 则以上双线性型 $\langle\,\cdot\,,\,\cdot\,\rangle_A$ 是对称的. 如果 A 是斜对称矩阵 (即 $A^T = -A$), 则该双线性型是斜对称的. ■

思考题 6.1. 对于例 6.1.4 中的第 3 个例子, 证明: 如果双线性型 $\langle\,\cdot\,,\,\cdot\,\rangle_A$ 是对称的 (斜对称的), 则矩阵 A 必然是对称的 (斜对称的).

思考题 6.2. 我们知道, 对于任意集合 A, 从 A 到 K 的所有映射构成的集合 $\mathrm{Map}(A, K)$ 是一个 K-向量空间, 其中元素的加法和数乘按照函数逐点取值相加和数乘的方式定义.

对于任意 K-向量空间 V, 证明:

1. $\mathrm{Bil}(V)$, $\mathrm{Sym}(V)$ 和 $\mathrm{Alt}(V)$ 都是 $\mathrm{Map}(V \times V, K)$ 的子空间.

2. $\mathrm{Bil}(V) = \mathrm{Sym}(V) \oplus \mathrm{Alt}(V)$.

   (提示: 若 $\varphi \in \mathrm{Map}(V \times V, K)$, 可以定义其**转置** φ^T 为 $\varphi^T(u, v) := \varphi(v, u)$. 当 $\varphi \in \mathrm{Bil}(V)$ 时可以验证 $\varphi^T \in \mathrm{Bil}(V)$.)

在过去研究线性映射的时候我们看到: 在有限维的情况下, 选定了向量空间的有序基之后, 线性映射可以和矩阵建立起非常好的对应 (参见本书上册 (3.2.8) 或 (5.0.7)). 现在我们来说明对于双线性型来说也有类似的现象.

(6.1.5) 设 V 是有限维 K-向量空间, $n=\dim V\geq 1$. 取定 V 的一组有序基 $\mathcal{B}=(v_1,\cdots,v_n)$. 若 φ 是 V 上的双线性型, 则对于任意 $u,\,v\in V$, 我们有如下方法确定 $\varphi(u,\,v)$ 的值: 先将 $u,\,v$ 在有序基 $\mathcal{B}$ 下写成坐标形式

$$u=(v_1,\cdots,v_n)\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix},\quad v=(v_1,\cdots,v_n)\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}$$

即 $\mathcal{M}_{\mathcal{B}}(u)=(x_1,\cdots,x_n)^T$ 和 $\mathcal{M}_{\mathcal{B}}(v)=(y_1,\cdots,y_n)^T$ 是 u 和 v 在 $\mathcal{B}$ 下的坐标列 (参见 (5.0.6)). 根据 φ 的双线性可得

(6.1.5.1)
$$\begin{aligned}\varphi(u,\,v)&=\varphi\Big(\sum_{i=1}^n x_iv_i\,,\,\sum_{j=1}^n y_jv_j\Big)=\sum_i\varphi\Big(x_iv_i\,,\,\sum_j y_jv_j\Big)=\sum_i\sum_j x_iy_j\varphi(v_i\,,\,v_j)\\ &=(x_1\,,\cdots,\,x_n)\cdot\mathcal{M}_{\mathcal{B}}(\varphi)\cdot\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}=\mathcal{M}_{\mathcal{B}}(u)^T\cdot\mathcal{M}_{\mathcal{B}}(\varphi)\cdot\mathcal{M}_{\mathcal{B}}(v)\,.\end{aligned}$$

其中矩阵

(6.1.5.2)
$$\mathcal{M}_{\mathcal{B}}(\varphi):=\big(\varphi(v_i,\,v_j)\big)_{1\leq i,\,j\leq n}$$

被称为 φ 在有序基 $\mathcal{B}$ 下的 Gram **矩阵** (Gram* matrix) 或者简称 φ 在 $\mathcal{B}$ 下的矩阵. 注意, (6.1.5.2) 中的 Gram 矩阵由 φ 和有序基 $\mathcal{B}$ 完全确定, 在它确定之后就可以根据 (6.1.5.1) 给出的公式

$$\varphi(u,\,v)=\mathcal{M}_{\mathcal{B}}(u)^T\mathcal{M}_{\mathcal{B}}(\varphi)\mathcal{M}_{\mathcal{B}}(v)$$

计算出 φ 在任意元素 $(u,\,v)\in V\times V$ 处的取值. 如果按照例 6.1.4 (3) 中的方式定义双线性型

$$\langle\cdot\,,\,\cdot\rangle_{\mathcal{M}_{\mathcal{B}}(\varphi)}\ :\ K^n\times K^n\longrightarrow K\,;\quad (x,\,y)\longmapsto x^T\mathcal{M}_{\mathcal{B}}(\varphi)y\,,$$

则 (6.1.5.1) 式的含义就是以下图表可交换:

(6.1.5.3)
$$\begin{tikzcd} V\times V \arrow[r, "\varphi"] \arrow[d, "\mathcal{M}_{\mathcal{B}}(\cdot)\times\mathcal{M}_{\mathcal{B}}(\cdot)"'] & K \arrow[d, "\mathrm{Id}"] \\ K^n\times K^n \arrow[r, "\langle\cdot\,,\,\cdot\rangle_{\mathcal{M}_{\mathcal{B}}(\varphi)}"'] & K \end{tikzcd}$$

这里 $\mathcal{M}_{\mathcal{B}}(\cdot):V\to K^n$ 表示将 V 中向量在有序基 $\mathcal{B}$ 下取坐标得到的映射 (其实是个同构).

如果将 $\mathcal{B}$ 固定, 而认为 φ 可以在 $\mathrm{Bil}(V)$ 中变化, 那么以上讨论说明: 通过取 Gram 矩阵我们得到了一个映射

(6.1.5.4)
$$\mathcal{M}_{\mathcal{B}}\ :\ \mathrm{Bil}(V)\longrightarrow\mathbf{M}_n(K)\,;\quad \varphi\longmapsto\mathcal{M}_{\mathcal{B}}(\varphi)\,.$$

和矩阵空间 $\mathbf{M}_n(K)$ 一样, $\mathrm{Bil}(V)$ 也是 K 上的向量空间 (思考题 6.2 (1)). 请读者验证: (6.1.5.4) 中的映射是一个线性映射. ■

命题 6.1.6: 设 V 为 n 维 K-向量空间, $\mathcal{B}$ 是 V 的一组有序基.

则 (6.1.5.4) 中的线性映射是个同构. 而且, 在这个映射建立的对应下, 对称双线性型和对称矩阵一一对应, 斜对称双线性型和斜对称阵一一对应.

*Jørgen Pedersen Gram (1850–1916) 丹麦数学家.

证明. 命题中后一论断从前一论断结合思考题 6.1 可得. 以下我们只要解释前一个断言即可.

观察 (6.1.5.3) 中的交换图表, 其中左右两个竖直方向的映射是随着 $\mathcal{B}$ 的选取固定下来的. 如果有 $\varphi, \psi \in \mathrm{Bil}(V)$ 对应的 Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(\varphi)$ 和 $\mathcal{M}_{\mathcal{B}}(\psi)$ 是相同的, 则 (6.1.5.3) 的底下一个水平方向的映射对应相同. 由图表的交换性即知 $\varphi = \psi$. 这说明 (6.1.5.4) 中的映射是单射.

为说明该映射是满射, 只需要注意到: 对于任意给定的 $A \in \mathbf{M}_n(K)$, 通过下图中的方式复合

$$\begin{CD} V \times V @. K \\ @V{\mathcal{M}_{\mathcal{B}}(\cdot)\times\mathcal{M}_{\mathcal{B}}(\cdot)}VV @AA{\mathrm{Id}}A \\ K^n \times K^n @>>{\langle\cdot\,,\,\cdot\rangle_A}> K \end{CD}$$

即可得到一个映射

$$\text{(6.1.6.1)} \qquad \Phi_{\mathcal{B}}(A) \;:\; V \times V \longrightarrow K\,; \quad (u,\, v) \longmapsto \mathcal{M}_{\mathcal{B}}(u)^T \cdot A \cdot \mathcal{M}_{\mathcal{B}}(v)\,.$$

不难验证这是一个双线性型, 而且它在 $\mathcal{B}$ 下的 Gram 矩阵就是 A.

至此我们证明了 (6.1.5.4) 是双射, 它的逆映射为

$$\text{(6.1.6.2)} \qquad \Phi_{\mathcal{B}} \;:\; \mathbf{M}_n(K) \longrightarrow \mathrm{Bil}(V)\,; \quad A \longmapsto \Phi_{\mathcal{B}}(A)\,,$$

其中 $\Phi_{\mathcal{B}}(A)$ 按照 (6.1.6.1) 式定义. □

(6.1.7) 为了方便读者, 我们不厌其烦地复述一下命题 6.1.6 的意义.

该结论是说, 如果选定了 V 的一组有序基 $\mathcal{B} = (v_1, \cdots, v_n)$, 那么任何双线性型 $\varphi \in \mathrm{Bil}(V)$ 被其 Gram 矩阵中的元素 $\varphi(v_i,\, v_j)$, $1 \le i,\, j \le n$ 唯一地决定. 反过来, 任意给定一组数 $a_{ij} \in K$, $1 \le i,\, j \le n$, 存在唯一的双线性型 $\varphi \in \mathrm{Bil}(V)$ 使 $\varphi(v_i,\, v_j) = a_{ij}$.

这类似于以前我们对线性映射得到的如下结论 (参见本书上册命题 3.1.32): 对于任意线性映射 $f : V \to W$, f 被 $f(v_i)$, $1 \le i \le n$ 这组向量唯一决定; 反过来, 任意给定一组向量 $w_1, \cdots, w_n \in W$, 存在唯一的线性映射 $f \in \mathrm{Hom}(V,\, W)$ 满足 $f(v_i) = w_i$, $1 \le i \le n$. ■

例 6.1.8. 若设 $V = K^n$ 为 n 维列向量空间, $\mathcal{B} = (e_1, \cdots, e_n)$ 为其有序标准基, 则对于任意 $A \in \mathbf{M}_n(K)$, (6.1.6.1) 式给出的 $\Phi_{\mathcal{B}}(A)$ 就是例 6.1.4 (3) 中给出的双线性型 $\langle\cdot\,,\,\cdot\rangle_A$. 结合命题 6.1.6 可以得到如下结论:

如果矩阵 $A,\, B \in \mathbf{M}_n(K)$ 能使 $x^TAy = x^TBy$ 对所有列向量 $x, y \in K^n$ 成立, 那么必然 $A = B$.

这和我们直接通过矩阵运算可以验证的结果一致. 事实上, 矩阵计算表明 $e_i^TAe_j$ 就是矩阵 A 在 $(i,\, j)$ 位置的元素. ■

6.1.2 矩阵相合与对称双线性型的对角化

我们知道, 线性变换的矩阵在做基变换时会发生相应的变化, 与之对应的概念是矩阵的相似. 现在我们来看双线性型的矩阵在做基变换时会发生什么变换.

在这一小节中, 我们一直假设 $n \ge 1$, V 是 n 维 K-向量空间.

(6.1.9) 取 V 的两组有序基 $\mathcal{B} = (v_1, \cdots, v_n)$ 和 $\mathcal{B}' = (v_1', \cdots, v_n')$. 设 P 是从 $\mathcal{B}$ 到 $\mathcal{B}'$ 的过渡矩阵, 即,

$$(v_1', \cdots, v_n') = (v_1, \cdots, v_n) \cdot P \quad \text{或者简写为} \;\; \mathcal{B}' = \mathcal{B} \cdot P\,.$$

则对于任意 $v \in V$, 它在 $\mathcal{B}$ 和 $\mathcal{B}'$ 两组基下的坐标 $\mathcal{M}_{\mathcal{B}}(v)$ 和 $\mathcal{M}_{\mathcal{B}'}(v)$ 应该满足 $\mathcal{M}_{\mathcal{B}}(v) = P\mathcal{M}_{\mathcal{B}'}(v)$.

现在设 $\varphi\in \mathrm{Bil}(V)$. 按照双线性型 Gram 矩阵的定义, $\mathcal{M}_{\mathcal{B}}(\varphi)$ 和 $\mathcal{M}_{\mathcal{B}'}(\varphi)$ 是由以下公式唯一决定的:

$$(6.1.9.1)\qquad \varphi(u,\,v)=\mathcal{M}_{\mathcal{B}}(u)^T\cdot\mathcal{M}_{\mathcal{B}}(\varphi)\cdot\mathcal{M}_{\mathcal{B}}(v)\,,\quad \varphi(u,\,v)=\mathcal{M}_{\mathcal{B}'}(u)^T\cdot\mathcal{M}_{\mathcal{B}'}(\varphi)\cdot\mathcal{M}_{\mathcal{B}'}(v)\,,$$

其中 $u,\,v\in V$ 可以任意选取. 将关系式 $\mathcal{M}_{\mathcal{B}}(u)=P\mathcal{M}_{\mathcal{B}'}(u)$ 和 $\mathcal{M}_{\mathcal{B}}(v)=P\mathcal{M}_{\mathcal{B}'}(v)$ 代入到 (6.1.9.1) 中的第一个公式可得

$$\begin{aligned}\varphi(u,\,v)&=\mathcal{M}_{\mathcal{B}}(u)^T\cdot\mathcal{M}_{\mathcal{B}}(\varphi)\cdot\mathcal{M}_{\mathcal{B}}(v)\\&=\big(P\mathcal{M}_{\mathcal{B}'}(u)\big)^T\cdot\mathcal{M}_{\mathcal{B}}(\varphi)\cdot\big(P\mathcal{M}_{\mathcal{B}'}(v)\big)\\&=\mathcal{M}_{\mathcal{B}'}(u)^T\cdot\big(P^T\cdot\mathcal{M}_{\mathcal{B}}(\varphi)\cdot P\big)\cdot\mathcal{M}_{\mathcal{B}'}(v)\,.\end{aligned}$$

由此再和 (6.1.9.1) 中第二个公式对比可得

$$(6.1.9.2)\qquad \mathcal{M}_{\mathcal{B}'}(\varphi)=P^T\cdot\mathcal{M}_{\mathcal{B}}(\varphi)\cdot P\,.$$

这就是基变换下 Gram 矩阵的变化规律. ■

为便于表述 (6.1.9.2) 式揭示的规律, 我们引入下面的定义:

定义 6.1.10. 对于任意矩阵 $A,\,A'\in\mathbf{M}_n(K)$, 如果存在可逆矩阵 $P\in\mathbf{M}_n(K)$ 使得 $A'=P^TAP$, 则称 A 和 A' 在 K 上 **相合或合同** (congruent). ■

利用以上定义, (6.1.9.2) 式可以解释为: 双线性型在不同基下的 Gram 矩阵满足相合关系.

请读者自行验证, 相合是集合 $\mathbf{M}_n(K)$ 上的等价关系. 即, 对于任意 $A,\,B,\,C\in\mathbf{M}_n(K)$, 以下论断成立:

- (**自反性**) A 和 A 自己是相合的.
- (**对称性**) 若 A 与 B 相合, 则 B 也与 A 相合.
- (**传递性**) 若 A 与 B 相合, B 与 C 相合, 则 A 与 C 相合.

矩阵在相似意义下的许多不变量 (如行列式、迹、特征多项式等) 在相合意义下都可能产生变化. 因此, 在一般情况下, 相合关系和相似关系是由很大区别的. 这一点希望读者能够留意.

思考题 6.3. 设 $\varphi\in\mathrm{Bil}(V)$.

1. 证明下列陈述等价:
   (a) φ 是对称的.
   (b) 存在 V 的一组有序基 $\mathcal{B}$, 使得 Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(\varphi)$ 是对称阵.
   (c) 对于 V 的任意一组有序基 $\mathcal{B}$, Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(\varphi)$ 是对称阵.
2. 对于斜对称的双线性型, 叙述并证明和前一个问题类似的结论.

过去我们关心矩阵在相似意义下能够化简成什么样的最简形式, 现在我们自然想知道矩阵在相合意义下可以化简成什么样的简单形式. 最重要的情况当然是可以相合于对角阵的情况. 显然, 如果 D 是个对角阵, 那么 D 是对称矩阵. 于是, 对于任何相同大小的矩阵 P, P^TDP 仍然是对称阵. 因此, 一个矩阵能够相合于对角阵的必要条件是该矩阵必须是对称阵. 接下来我们将逐步证明, 这个条件也是充分条件. 即, 任何对称矩阵都可以相合到某个对角阵.

(6.1.11) 正如研究矩阵相似的恰当方法是转化为相应的线性变换来研究, 我们现在讨论矩阵的相合应该转化为双线性型的问题来探讨.

设 $A \in \mathbf{M}_n(K)$ 为对称阵. 我们在列向量空间 K^n 上考虑对称双线性型

$$\langle \cdot\,,\,\cdot \rangle_A \;:\; K^n \times K^n \longrightarrow K\,; \quad (x,\,y) \longmapsto x^T A y\,.$$

则 A 是该双线性型在有序标准基 $\mathcal{E} = (e_1, \cdots, e_n)$ 下的 Gram 矩阵. 根据 (6.1.9.2) 式, 要找出可逆矩阵 P 使 $P^T AP$ 为对角阵, 实际上就是要找出新的一组有序基 $\mathcal{B}$, 使新的 Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(\langle \cdot\,,\,\cdot \rangle_A)$ 是对角阵. 这相当于要求新的有序基 $\mathcal{B} = (v_1, \cdots, v_n)$ 中向量满足如下条件:

$$\text{当 } i,\, j \in [\![1,\, n]\!] \text{ 不同时, 总有 } \langle\, v_i\,,\, v_j\,\rangle_A = 0\,.$$

这样的一组基称为 $\langle \cdot\,,\,\cdot \rangle_A$ 的一组正交基. 所以, 问题转化为寻找对称双线性型的正交基问题. ■

鉴于其特殊的重要性, 我们把上面提到的正交基以及相关的"正交"概念做如下简单的推广:

定义 6.1.12. 设 $\varphi \in \mathrm{Bil}(V)$.

1. 若 $u\,, v$ 满足 $\varphi(u\,,\,v) = \varphi(v\,,\,u) = 0$, 则称 u 和 v **关于** φ **正交** (orthogonal with respect to φ), 记为 $u \perp_\varphi v$, 或者更简单地, $u \perp v$.

2. 如果 V 的一组有序基 $\mathcal{B}$ 中每一对不同的向量关于 φ 都正交, 则称 $\mathcal{B}$ 是 φ 的一组**正交基** (orthogonal basis).

   注意, 如果 $\dim V = 1$, 即 $\mathcal{B}$ 只含有一个向量, 那么自动认为 $\mathcal{B}$ 是一组正交基.

   因为正交基对应的 Gram 矩阵是对角阵, 通常我们把找出 φ 的一组正交基的过程, 称为 φ 的**对角化** (digonalization) 过程.

3. 若 S 是 V 的子集, 我们将集合

   $$S^\perp := \{v \in V \,|\, v \text{ 和 } S \text{ 中的所有向量都关于 } \varphi \text{ 正交}\}$$

   称为 S **关于** φ **的正交补** (orthogonal complement with respect to φ). (若 S 为空集, 约定其正交补为 V.)

   请读者注意, 为了简化记号, $S^\perp$ 这个记号中没有体现出双线性型 φ 的信息, 但其定义是依赖于 φ 的.

4. 对于 V 的两个子集 S 和 S', 如果 S 中的每个向量都和 S' 中的所有向量都关于 φ 正交, 我们就说 S 和 S' 关于 φ **正交**, 记为 $S \perp_\varphi S'$ 或简记为 $S \perp S'$. 这等价于说 $S \subseteq (S')^\perp$, 也等价于说 $S' \subseteq S^\perp$. ■

思考题 6.4. 设 $\varphi \in \mathrm{Bil}(V)$, $S\,, S'$ 为 V 的子集. 证明:

1. S 关于 φ 的正交补 $S^\perp$ 是 V 的一个子空间.

2. 若 $S \subseteq S'$, 则 $(S')^\perp \subseteq S^\perp$.

3. 若 S 与 S' 关于 φ 正交, 则 $\mathrm{span}(S) \perp_\varphi \mathrm{span}(S')$, 即 S 生成的子空间 $\mathrm{span}(S)$ 和 S' 生成的子空间关于 φ 正交.

(6.1.13) 我们再继续之前的讨论, 考虑将一个对称双线性型 $\varphi \in \mathrm{Sym}(V)$ 对角化的问题. 现在问题已经被转化为寻找 φ 的一组正交基 $\mathcal{B} = (v_1, \cdots, v_n)$. 一个自然的想法是先找出第一个向量 v_1, 然后通过适当的归纳来说明其余的向量 $v_2, \cdots, v_n$ 也可以找到. 在找到 $v_2, \cdots, v_n$ 之前, 实际上我们能够确定的信息主要是: 它们应该都取自于 V 的某个子空间 U, 这个子空间和 $V_1 := \mathrm{span}(v_1)$ 是正交的, 而且 $V = V_1 \oplus U$. (因为若正交基 $(v_1, \cdots, v_n)$ 可以找到的话, 那么 $U = \mathrm{span}(v_2, \cdots, v_n)$ 一定会满足这样的条件.) 接下来, U 的维数已经严格小于 V 的维数了. 不难想到, 通过归纳法继续找到 $v_2, \cdots, v_n$ 应该就不成问题了.

具体来说, 如果 U 是 V 的一个子空间, 我们用 $\varphi|_U$ 表示将 $\varphi : V \times V \to K$ 的定义域限制到 $U \times U$ 上得到的映射, 称为 φ 在 U 上的**限制**. 即

$$\varphi|_U \ : \ U \times U \longrightarrow K\,; \quad (u\,,\, v) \longmapsto \varphi(u,\, v)\,.$$

也就是说, $\varphi|_U$ 的取值和 φ 是一样的, 只是人为地缩小了定义域而已. 显然, $\varphi|_U$ 是 U 上的双线性型. 因为 φ 是对称的, 故 $\varphi|_U$ 亦如此.

于是, 当 $\dim U = n - 1 < \dim V = n$ 时, 通过归纳法可以知道 U 中存在 $\varphi|_U$ 的一组正交基 $v_2, \cdots, v_n$. 由于 $\varphi|_U$ 只是 φ 的限制, 所以 $v_2, \cdots, v_n$ 两两之间关于 φ 也是正交的. 另一方面, 因为 $U = \mathrm{span}(v_2, \cdots, v_n)$ 整个都与 $\mathrm{span}(v_1)$ 关于 φ 正交, 所以 $v_2, \cdots, v_n$ 中的每一个都与 v_1 关于 φ 正交. 所以 $(v_1, v_2, \cdots, v_n)$ 就给出了 φ 的一组正交基. ■

通过 (6.1.13) 一段的讨论, 我们看到: 将 $\varphi \in \mathrm{Sym}(V)$ 正交化的关键步骤就是先找到空间 V 的一个直和分解 $V = V_1 \oplus U$, 其中 V_1 和 U 关于 φ 正交, 并且 $\dim U < \dim V$. 由此我们就明白下面的定义 6.1.14 意义何在了.

定义 6.1.14. 设 $\varphi \in \mathrm{Bil}(V)$, $V_1, \cdots, V_r$ 都是 V 的子空间.

1. 如果对于每两个不同的指标 $i, j \in [\![1, r]\!]$ 均有 $V_i \perp_\varphi V_j$, 而且 $V_1 + \cdots + V_r$ 是一个直和, 那么我们说 $V_1 + \cdots + V_r$ 是关于 φ 的一个**正交和** (orthogonal sum).

2. 如果 $V = V_1 + \cdots + V_r$ 并且 $V_1 + \cdots + V_r$ 是关于 φ 的正交和, 我们记

$$V = V_1 \perp_\varphi \cdots \perp_\varphi V_r \qquad \text{或简记为} \qquad V = V_1 \perp \cdots \perp V_r$$

并将此表达式称为 V 关于 φ 的一个**正交分解** (orthogonal decomposition), 或者称为双线性空间 (V, φ) 的一个正交分解. ■

现在我们可以证明之前承诺的结论了.

定理 6.1.15: *对于有限维向量空间 V 上的任意对称双线性型 φ, V 中一定存在 φ 的一组正交基.*

换成矩阵语言来说, 任意对称矩阵 $A \in \mathbf{M}_n(K)$ 都可以在 K 上相合于某个对角阵.

证明. 根据我们之前在 (6.1.13) 一段的讨论, 通过对 V 的维数做归纳法, 我们只需要证明: 存在非零向量 $v_1 \in V$ 使得 $V_1 = \mathrm{span}(v_1)$ 能够和另外某个子空间 $U \subseteq V$ 构成 V 的一个正交分解 $V = V_1 \perp U$.

要选取 U 与 V_1 正交, 一个理想的选择是 $U = V_1^\perp$. 不过, 此处我们还需要这一理想化的选择能够使正交分解 $V = V_1 \perp V_1^\perp$ 成立. 读者不难验证, 这一性质的一个必要条件是 $\varphi(v_1, v_1) \neq 0$ (参见下面的思考题 6.5). 现在我们来说明这个必要条件也是充分的. 即, 只要 v_1 满足 $\varphi(v_1, v_1) \neq 0$, 那么 $V_1 = \mathrm{span}(v_1)$ 一定和它的正交补 $V_1^\perp$ 形成 V 的一个正交分解.

首先, 由 $\varphi(v_1, v_1) \neq 0$ 可以轻易推出 $V_1 \cap V_1^\perp = 0$, 从而 $V_1 + V_1^\perp$ 是个直和. 然后我们需要证明: 对于任意确定的 $v \in V$, 一定可以找到常数 λ 使得 $v - \lambda v_1 \in V_1^\perp$, 即 $\varphi(v - \lambda v_1, v_1) = 0$. 事实上, 由 φ 的双线性容易看出, 这一点只要取 $\lambda = \varphi(v, v_1)/\varphi(v_1, v_1)$ 即可满足.

以上我们说明了, 只要能够找到 $v_1 \in V$ 使 $\varphi(v_1, v_1) \neq 0$, 那么通过归纳法就完成了定理的证明. 剩下的问题是如果这样的 v_1 不存在怎么办. 其实这种情况更为简单. 因为如果所有的 $v \in V$ 都使 $\varphi(v, v) = 0$, 那么双线性型 φ 是交错的. 可是题设中又说了 φ 是对称的. 根据命题 6.1.3 (2), 此时 φ 必然恒等于 0, 于是 V 的任意一组基都是 φ 的正交基. □

思考题 6.5. 设 $\varphi \in \mathrm{Bil}(V)$, $v_1 \in V$ 为非零向量, $V_1 = \mathrm{span}(v_1)$.

证明: 如果 $V_1 \cap V_1^{\perp} = 0$, 则 $\varphi(v_1, v_1) \neq 0$.

(6.1.16) 设 $A \in \mathbf{M}_n(K)$ 为对称矩阵. 如果对角矩阵 $D \in \mathbf{M}_n(K)$ 在 K 上与 A 相合, 则称 D 是 A 的一个**相合标准形** (congruence canonical form). 定理 6.1.15 说明了 A 的相合标准形总是存在的. 不过, 相合标准形不是唯一的. 例如, 实矩阵 $A = \begin{pmatrix} 1 & \\ & 4 \end{pmatrix}$ 与单位矩阵 $I = I_2$ 相合, 因此 I 和 A 本身都是 A 的相合标准形.

定理 6.1.15 的证明过程实际上为我们提供了一种化对称阵为相合标准型的计算方法. 例如, 如果想找出矩阵 $A = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & -2 \\ -3 & -2 & 1 \end{pmatrix}$ 的一个相合标准型 D, 并找到一个可逆矩阵 P 使 $P^TAP = D$, 我们可以考虑相应的双线性型

$$\varphi_A \;:\; K^3 \times K^3 \longrightarrow K\,; \quad \left((x_1, y_1, z_1)^T, (x_2, y_2, z_2)^T\right) \longmapsto (x_1, y_1, z_1)A \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}.$$

我们先寻找一个向量 $v = (x, y, z)^T \in K^3$ 使得

$$0 \neq \varphi_A(v, v) = x^2 + y^2 + z^2 + 4xy - 6xz - 4yz\,.$$

例如可取 $v_1 = e_1 = (1, 0, 0)^T$. 然后我们计算 $V_1 = \mathrm{span}(v_1)$ 关于 φ_A 的正交补, 即求出满足

$$0 = \varphi_A(e_1, (x, y, z)^T) = x + 2y - 3z$$

的所有向量 $u = (x, y, z)^T$ 构成的空间 U. 易见

$$U = \mathrm{span}((-2, 1, 0)^T, (3, 0, 1)^T) = \{(-2y + 3z, y, z)^T \,|\, y, z \in K\}\,.$$

接下来我们考虑 $\varphi_A|_U$, 在 U 中取 v_2 满足 $\varphi_A(v_2, v_2) \neq 0$. 例如可取 $v_2 = (-2, 1, 0)^T$. 下一步找出 $\mathrm{span}(v_2)$ 在 U 中的正交补, 即求 $(-2y + 3z, y, z)^T$ 形式的向量 v_3 与 v_2 关于 φ_A 正交. 这需要求解

$$(-2, 1, 0) \begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & -2 \\ -3 & -2 & 1 \end{pmatrix} \begin{pmatrix} -2y + 3z \\ y \\ z \end{pmatrix} = 0\,.$$

简单的计算发现一个解为 $v_3 = (1, 4, 3)^T$.

综上可知

$$v_1 = (1, 0, 0)^T\,, \; v_2 = (-2, 1, 0)^T\,, \; v_3 = (1, 4, 3)^T$$

对应的矩阵 $P = \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{pmatrix}$ 满足 $P^TAP = \begin{pmatrix} 1 & & \\ & -3 & \\ & & -24 \end{pmatrix}$.

以上就是采用定理 6.1.15 的证明思路将一个对称矩阵化成相合标准形的计算实例. 这个过程应该不算很困难, 但是过程显得较为拖沓. 今后我们采用二次型的观点再考虑化对称阵为相合标准形的问题时, 会有更加直观简洁的方法, 因此我们到时候会再详细讨论这个问题. ■

基于矩阵相似不改变秩的性质, 我们可以通过相应矩阵的秩来定义线性变换的秩. 对于双线性型, 我们可以做类似的定义. 其理论基础是矩阵的相合变换也不改变矩阵的秩.

思考题 6.6. 证明: 如果两个 n 阶方阵 A 和 B 相合, 则 $\operatorname{rank}(A) = \operatorname{rank}(B)$.

定义 6.1.17. 设 φ 是有限维 K-向量空间 V 上的双线性型, A 是 φ 在 V 的某一组有序基 $\mathcal{B}$ 下的 Gram 矩阵. 我们将矩阵 A 的秩称为双线性型 φ 的**秩** (rank), 记为 $\operatorname{rank}(\varphi)$ 或者更精确地, $\operatorname{rank}(V, \varphi)$. 根据思考题 6.6 和 (6.1.9.2) 式, $\operatorname{rank}(\varphi)$ 不依赖于 $\mathcal{B}$ 的选取. 也就是说, 对于任意一组有序基 $\mathcal{B}$, Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(\varphi)$ 的秩总是相同的.

如果 $\operatorname{rank}(V, \varphi) = \dim V$, 我们就说双线性型 φ 是**满秩的** (of full rank). 此时 φ 在任意一组基下的 Gram 矩阵都是满秩的. 因为一个方阵是满秩的当且仅当它是非奇异的, 所以 V 上满秩的双线性型也被称为是**非奇异的** (nonsingular). ■

6.1.3 非退化双线性型与对偶空间

设 φ 为 K-向量空间 V 上的双线性型. 根据定义我们知道, 如果取定 V 中一个向量 u, 而令 φ 的第二个变量变动, 那么我们得到一个线性映射

$$\varphi_u \;:\; V \longrightarrow K\,; \quad v \longmapsto \varphi_u(v) := \varphi(u,\, v)\,.$$

类似地, 如果固定一个向量 $w \in V$ 为 φ 的第二个变量, 而令第一个变量变动, 那么我们得到另一个线性映射

$$\varphi'_w \;:\; V \longrightarrow K\,; \quad v \longmapsto \varphi'_w(v) := \varphi(v,\, w)\,.$$

这一小节我们的主要目标就是深入研究这种现象.

为便于讨论我们引入一些概念.

定义 6.1.18. 设 φ 为 K-向量空间 V 上的双线性型.

1. 我们将集合

   $$\operatorname{Rad}_1(V, \varphi) = \operatorname{Rad}_1(\varphi) := \{u \in V \mid \text{对于所有 } v \in V \text{ 均有 } \varphi(u, v) = 0\}$$

   称为双线性型 φ 或双线性空间 (V, φ) 的**左根** (left radical). 如果 $\operatorname{Rad}_1(\varphi) = 0$, 则称双线性型 φ 或双线性空间 (V, φ) 是**左非退化的** (nondegenerate on the left).

   类似地, 将集合

   $$\operatorname{Rad}_2(V, \varphi) = \operatorname{Rad}_2(\varphi) := \{w \in V \mid \text{对于所有 } v \in V \text{ 均有 } \varphi(v, w) = 0\}$$

   称为双线性型 φ 或双线性空间 (V, φ) 的**右根** (right radical). 如果 $\operatorname{Rad}_2(\varphi) = 0$, 则称双线性型 φ 或双线性空间 (V, φ) 是**右非退化的** (nondegenerate on the right).

   如果 $\operatorname{Rad}_1(\varphi) = \operatorname{Rad}_2(\varphi) = 0$, 则称双线性型 φ 或双线性空间 (V, φ) 是**非退化的** (nondegenerate).

   根据这个定义, 如果 $\dim V = 0$, 则 (V, φ) 是非退化的. 不过, 这种平凡的情况我们通常都可以直接忽略.

2. 假设 φ 是对称的或者交错的. 此时 $\operatorname{Rad}_1(\varphi) = \operatorname{Rad}_2(\varphi)$, 故可将其简记为 $\operatorname{Rad}(\varphi)$, 称为 φ 的**根** (radical). ■

引理 6.1.19: 设 φ 为 K-向量空间 V 上的双线性型.

1. 左根 $\mathrm{Rad}_1(\varphi)$ 和右根 $\mathrm{Rad}_2(\varphi)$ 都是 V 的子空间.
2. 对于任意 $u \in V$, $u \in \mathrm{Rad}_1(\varphi)$ 的充分必要条件是线性映射

$$\varphi_u \,:\, V \longrightarrow K\,; \quad v \longmapsto \varphi_u(v) := \varphi(u,\, v)$$

   为零映射.

   类似地, 对于任意 $w \in V$, $w \in \mathrm{Rad}_2(\varphi)$ 的充分必要条件是线性映射

$$\varphi'_w \,:\, V \longrightarrow K\,; \quad v \longmapsto \varphi'_w(v) := \varphi(v,\, w)$$

   为零映射.

3. 假设 V 是有限维的, $n = \dim V \geq 1$. 设 $\mathcal{B}$ 为 V 的一组有序基, $A = \mathcal{M}_{\mathcal{B}}(\varphi) \in \mathbf{M}_n(K)$.

   则向量 $u \in V$ 属于 $\mathrm{Rad}_1(\varphi)$ 的充分必要条件是其坐标列向量 $\mathcal{M}_{\mathcal{B}}(u)$ 属于转置矩阵 A^T 的零化空间 $\mathcal{N}(A^T)$, 即, $\mathcal{M}_{\mathcal{B}}(u)^T A = 0$.

   类似地, 向量 $w \in V$ 属于 $\mathrm{Rad}_2(\varphi)$ 的充分必要条件是其坐标列向量 $\mathcal{M}_{\mathcal{B}}(w)$ 属于矩阵 A 的零化空间 $\mathcal{N}(A)$, 即, $A\mathcal{M}_{\mathcal{B}}(w) = 0$.

证明. (1) 留给读者练习. (2) 此证明属于抽象废话, 请读者认真确认自己可以独立解释清楚个中缘由.

(3) 对于任意 $v \in V$, $\varphi_u(u,\, v)$ 的计算公式为

$$\varphi_u(v) = \varphi(u,\, v) = \mathcal{M}_{\mathcal{B}}(u)^T A \mathcal{M}_{\mathcal{B}}(v)\,.$$

当 v 取遍 V 中向量时, 其坐标列 $\mathcal{M}_{\mathcal{B}}(v)$ 可以取遍列向量空间 K^n 中的所有元素. 因此, 若记 $e_1, \cdots, e_n$ 为 K^n 的标准基, 则

$$\begin{aligned}
u \;\in\; \mathrm{Rad}_1(\varphi) &\iff \mathcal{M}_{\mathcal{B}}(u)^T A y = 0 \text{ 对所有 } y \in K^n \text{ 成立} \\
&\iff \mathcal{M}_{\mathcal{B}}(u)^T A e_i = 0 \text{ 对每个 } i = 1, \cdots, n \text{ 都成立} \\
&\iff \text{ 矩阵 } \mathcal{M}_{\mathcal{B}}(u)^T A \text{ 的每一列均为零} \\
&\iff \mathcal{M}_{\mathcal{B}}(u)^T A = 0 \iff A^T \mathcal{M}_{\mathcal{B}}(u) = 0\,.
\end{aligned}$$

关于右根的证明留给读者自行补充. □

推论 6.1.20: 设 φ 是有限维 K-向量空间 V 上的双线性型, $n = \dim V \geq 1$. 则下列陈述等价:

1. φ 是非退化的, 即, $\mathrm{Rad}_1(\varphi) = \mathrm{Rad}_2(\varphi) = 0$.
2. φ 是左非退化的, 即, $\mathrm{Rad}_1(\varphi) = 0$.
3. φ 是右非退化的, 即, $\mathrm{Rad}_2(\varphi) = 0$.
4. φ 是非奇异的 (亦即满秩的), 即, φ 在任意一组基下的 Gram 矩阵是非奇异的 (亦即满秩的).

证明. 留给读者. □

下面我们将采用线性映射的观点再来考察双线性型. 为此我们先将以前熟悉的一个空间赋予特殊的名字.

定义 6.1.21. 设 V 为 K-向量空间. 我们知道, 如果将 K 自身也视为一个 (1 维的) K-向量空间, 那么从 V 到 K 的所有线性映射构成的集合 $\mathrm{Hom}(V, K)$ 也是一个 K-向量空间.

今后我们常将 $\mathrm{Hom}(V, K)$ 记为 V°, 并称之为 V 的**对偶空间** (dual space)[†]. 空间 V° 中的元素就是从 V 到 K 的**线性函数** (linear function), 有时也称为**线性泛函** (linear functional). ■

关于对偶空间有许多课题值得细致讨论, 但此时我们只想研究它与双线性型之间的联系.

(6.1.22) 设 φ 为 K-向量空间 V 上的双线性型. 根据之前的讨论, 任意取定 $u \in V$ 后, 我们可以定义一个相应的线性泛函

$$\varphi_u \ : \ V \longrightarrow K\,; \quad v \longmapsto \varphi_u(v) := \varphi(u,\, v)\,. \tag{6.1.22.1}$$

换言之, 通过选定 $\varphi \in \mathrm{Bil}(V)$ 和 $u \in V$ 可以得到对偶空间 V° 中的一个元素 φ_u. 如果我们固定 φ, 那么从 $u \in V$ 到 $\varphi_u \in V^\circ$ 的对应关系可以认为是一个映射

$$\hat{\varphi} \ : \ V \longrightarrow V^\circ\,; \quad u \longmapsto \varphi_u\,. \tag{6.1.22.2}$$

读者不难验证这是一个线性映射.

类似地, 任意取定 $w \in V$ 后, 我们可以定义一个相应的线性泛函

$$\varphi'_w \ : \ V \longrightarrow K\,; \quad v \longmapsto \varphi_w(v) := \varphi(v,\, w)\,. \tag{6.1.22.3}$$

对应关系 $w \mapsto \varphi'_w$ 定义了一个映射

$$\hat{\varphi}' \ : \ V \longrightarrow V^\circ\,; \quad w \longmapsto \varphi'_w\,. \tag{6.1.22.4}$$

这也是一个线性映射. ■

思考题 6.7. 记号如 (6.1.22), 证明:

1. (6.1.22.2) 和 (6.1.22.4) 中的映射都是线性映射.
2. $\mathrm{Ker}(\hat{\varphi}) = \mathrm{Rad}_1(\varphi)$, $\mathrm{Ker}(\hat{\varphi}') = \mathrm{Rad}_2(\varphi)$.

   因此, φ 是左非退化的当且仅当 $\hat{\varphi}$ 是单射 (φ 是右非退化的当且仅当 $\hat{\varphi}'$ 是单射).

定理 6.1.23 (Riesz **表示定理** Riesz[‡] representation theorem): *设 V 是有限维 K-向量空间, $\dim V \geq 1$. 设 φ 是 V 上的非退化双线性型.*

则 (6.1.22.2) 和 (6.1.22.4) 中的映射都是双射. 换言之, 对于任意线性泛函 $f \in V^\circ$, 一定存在唯一的向量 $u \in V$ 使得 $f = \varphi_u$, 也存在唯一的向量 $w \in V$ 使得 $f = \varphi'_w$.

证明. 根据思考题 6.7, $\hat{\varphi}$ 和 $\hat{\varphi}'$ 都是线性的单射. 因此, 要证明它们是双射只需要知道 $\dim V$ 和 $\dim V^\circ$ 相等即可. 而这一点只要考虑线性映射对应的矩阵即可理解 (参见 (5.0.7) 或本书上册思考题 3.6 或 (3.2.9)). □

注记 6.1.24. 为了帮助读者加深对 Riesz 表示定理的理解, 我们提醒读者注意以下事实:

(1) 当 V 为有限维线性空间时, 我们可以简单的通过比较维数知道 V 和 V° 之间一定存在着同构映射. 但是, 这并不意味着任何方式得到的线性映射 $V \to V^\circ$ 都是同构, 而且有一些从 V 到 V° 的同构映射需要人为地选定基之后生硬地构造出来. Riesz 表示定理的意义在于, 当 V 被赋予一个自然的非退化双线性型时, 可以通过该双线性型建立起 V 到 V° 的两个自然的线性同构.

[†]关于对偶空间的另一个常见记号是 V^*. 我们不采用这一记号的主要原因是: 当我们考虑线性映射的对偶时, 相应的记号应该与空间的对偶一致, 但是对于一个线性变换 $\mathscr{A}$ 来说, $\mathscr{A}^*$ 这个记号似乎更加常见的是表示伴随变换 (参见 § 7.2.2).

[‡]Frigyes Riesz (1880–1956), 匈牙利数学家.

(2) 在定理 6.1.23 中, 除了 φ 要求是非退化的, 还有一个关键的假设是 V 是有限维空间. 这使得我们的证明变得非常简洁. 而如果 V 是无限维空间, 那么定理 6.1.23 不能在上述的最质朴形式下成立. 事实上, 对于一个无限维空间 V, 有可能在 V 和它的对偶空间 V° 之间不存在任何线性同构, 甚至不存在任何双射.

我们尝试为读者简单介绍一个这样的例子: 例如取 $K = \mathbb{Q}$, $V = K[X]$. 从集合论的观点来看, V 是一个可数集. 而它的对偶空间 $V^\circ = \mathrm{Hom}(V, K)$ 是一个不可数集: 对于任意一个通项为有理数的无穷数列 $c = (c_n)_{n\geq 0}$, 可以通过

$$a_0 + a_1 X + \cdots + a_m X^m \longmapsto a_0 c_0 + a_1 c_1 + \cdots + a_m c_m$$

这个法则得到一个线性泛函 $f_c \in V^\circ$. 不难看出 $c \mapsto f_c$ 这个对应是个单射, 而所有的有理数通项的无穷数列构成一个不可数集, 因此 V° 也是个不可数集.

因此, 在这个例子中 V 和 V° 之间不可能建立双射.

(3) 虽然上面说了无限维的情况下定理 6.1.23 不再成立, 但是在分析学领域人们可以考虑具有其他额外条件和附加假设的 Riesz 表示定理 (例如关于 Hilbert 空间的 Riesz 表示定理). 这是泛函分析学科的一个重要议题, 读者可以在今后更深入的数学学习中留意. ■

非退化的性质还可以帮助我们理解正交分解.

命题 6.1.25: *设 φ 是向量空间 V 上的双线性型. 假设 φ 是对称的或交错的.*

如果一个有限维子空间 $W \subseteq V$ 能使双线性空间 $(W, \varphi|_W)$ 是非退化的, 那么 $V = W \perp_\varphi W^\perp$, 即, W 和它的正交补构成 V 的一个正交分解.

证明. 记 $\phi := \varphi|_W$. 根据抽象废话, $W \cap W^\perp = \mathrm{Rad}(\phi)$. 所以关于 ϕ 非退化的假设说明 $W + W^\perp$ 是一个直和. 以下只需证明 $V = W + W^\perp$. 为此我们还需要进一步使用 ϕ 的非退化性质. 这个假设表明 Riesz 表示定理适用于 ϕ. 也就是说, 对于 W 上的任意线性泛函 $f \in W^\circ$, 一定存在 $w \in W$ 使得

(6.1.25.1) $$f(x) = \phi(w, x) = \varphi(w, x) \quad \text{对所有 } x \in W \text{ 成立}.$$

现在任取 $v \in V$, 我们希望证明: 存在 $w \in W$ 使得 $v - w \in W^\perp$, 即 $\varphi(v - w, x) = 0$ 对于所有 $x \in W$ 成立. 这个条件等价于

(6.1.25.2) $$\varphi(v, x) = \varphi(w, x) \quad \text{对所有 } x \in W \text{ 成立}.$$

对比 (6.1.25.1) 和 (6.1.25.2) 式容易看出, 只要将 (6.1.25.1) 中的线性泛函 $f \in W^\circ$ 定义为

$$f(x) := \varphi(v, x), \quad \text{对任意} \ \ x \in W$$

那么所需的结论立即得证. 事实上, 这个 f 是 V 上的线性泛函 $\hat{\varphi}(v) \in V^\circ$ 定义域限制于 W 得到的. □

在上述命题中, 如果取 V 中一个非零向量 v, $W = \mathrm{span}(v)$, 那么 $\varphi|_W$ 非退化的充分必要条件是 $\varphi(v, v) \neq 0$. 因此, 当 $\varphi(v, v) \neq 0$ 时我们可以得到正交分解 $V = W \perp W^\perp$. 这其实正是定理 6.1.15 的证明中一个关键步骤.

我们还想提醒读者注意: 命题 6.1.25 并未要求 φ 是非退化的, 而是要求 $\varphi|_W$ 非退化. 一般来说, 即使 φ 是非退化的, 也不能保证 $\varphi|_W$ 一定非退化 (参见以下思考题 6.8 (3)).

思考题 6.8. 设 φ 是有限维向量空间 V 上的对称或交错双线性型. 如果一个非零向量 $v \in V$ 满足 $\varphi(v, v) = 0$, 则称 v 是 φ 的一个**迷向** (isotropic) 向量. 如果这样的向量存在, 我们说 φ 是**迷向的**, 否则称 φ 是**非迷向的** (anisotropic).

1. 举出一个迷向的对称双线性型的例子和一个非迷向的对称双线性型的例子.

2. 证明: 如果 φ 是非迷向的, 那么对于任意子空间 $W \subseteq V$, $\varphi|_W$ 是非退化的.

3. 证明: 如果 φ 是迷向的, 则 (无论 φ 是否退化) 一定存在子空间 $W \subseteq V$ 使 $\varphi|_W$ 是退化的.

与命题 6.1.25 相似、但适用性更广的一个重要结论是如下定理:

定理 6.1.26 (**正交补维数定理** dimension theorem for orthogonal complements): *设 φ 是有限维向量空间 V 上的双线性型, $W \subseteq V$ 是一个子空间. 假设 φ 是对称的或交错的.*

如果 φ 或 $\varphi|_W$ 是非退化的, 那么 $\dim W + \dim W^{\perp} = \dim V$.

证明. 如果 $\varphi|_W$ 是非退化的, 那么结论可以从命题 6.1.25 直接看出.

现在假设 φ 是非退化的. 我们希望确定 W 和 $W^{\perp}$ 的维数之间的关系. 为此, 我们选取 W 的一组基 $v_1, \cdots, v_r$ 并扩充至 V 的一组基 $v_1, \cdots, v_r, v_{r+1}, \cdots, v_n$. 来看 $W^{\perp}$ 的元素如何描述.

事实上, 对于任意 $v \in V$, 可以将其唯一地写成 $v = \sum_{j=1}^{n} x_j v_j$, 其中 x_j 为常数. 则 $v \in W^{\perp}$ 当且仅当 $\varphi(v_1, v) = \cdots = \varphi(v_r, v) = 0$. 而对于每个 $i \in [\![1, r]\!]$,

$$\varphi(v_i, v) = \sum_{j=1}^{n} \varphi(v_i, v_j) x_j \,.$$

因此, $v \in W^{\perp}$ 的充分必要条件是其坐标向量 $(x_1, \cdots, x_n)^T$ 是线性方程组 $AX = 0$ 的解, 其中

$$A = \big(\varphi(v_i, v_j)\big)_{\substack{1 \le i \le r \\ 1 \le j \le n}} = \begin{pmatrix} \varphi(v_1, v_1) & \cdots & \cdots & \varphi(v_1, v_n) \\ \varphi(v_2, v_1) & \cdots & \cdots & \varphi(v_2, v_n) \\ \vdots & \ddots & \ddots & \vdots \\ \varphi(v_r, v_1) & \cdots & \cdots & \varphi(v_r, v_n) \end{pmatrix} .$$

由此可见, $W^{\perp}$ 与 A 的零化空间 $\mathcal{N}(A)$ 维数相同. 即 $\dim W^{\perp} = n - \operatorname{rank}(A)$. 而若以 M 表示 φ 在 $(v_1, \cdots, v_n)$ 这组有序基下的 Gram 矩阵, 则以上矩阵 A 恰好是由 M 的前 r 行构成的. 因为 φ 非退化, 故 M 是满秩的 (推论 6.1.20). 所以它的前 r 行线性无关. 这就说明 $\operatorname{rank}(A) = r = \dim W$. 所以我们得到 $\dim W^{\perp} = n - r = \dim V - \dim W$. 这就完成了定理的证明. □

读者需要留意: 定理 6.1.26 和命题 6.1.25 的结论是有区别的. 当 φ 非退化但 $\varphi|_W$ 退化时, 定理 6.1.26 只能保证维数公式 $\dim V = \dim W + \dim W^{\perp}$ 成立, 并不能说明 $W + W^{\perp}$ 是直和! 当然, 根据思考题 6.8 (2), 如果 φ 既是非退化的又是非迷向的, 那么 φ 在任何子空间上的限制仍是非退化的. 今后我们要讨论的内积就是这样一种性质非常好的双线性型.

6.1.4 习题

习题 6.1.1. 按如下方式定义一个函数 $\varphi : K^4 \times K^4 \to K$: 若

$$x = (x_1, x_2, x_3, x_4)^T \,, \quad y = (y_1, y_2, y_3, y_4)^T \,,$$

定义

$$\varphi(x, y) := 3x_1y_2 - 5x_2y_1 + x_3y_4 - 4x_4y_3 \,.$$

1. 证明 φ 是 K^4 上的一个双线性型.

2. 令

$$\varepsilon_1=\begin{pmatrix}1\\2\\-1\\0\end{pmatrix},\quad \varepsilon_2=\begin{pmatrix}1\\-1\\1\\1\end{pmatrix},\quad \varepsilon_3=\begin{pmatrix}-1\\2\\1\\1\end{pmatrix},\quad \varepsilon_4=\begin{pmatrix}-1\\-1\\0\\1\end{pmatrix}.$$

求 φ 在有序基 $\mathcal{B}=(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)$ 下的 Gram 矩阵.

3. 设 $\mathcal{C}=(\eta_1,\eta_2,\eta_3,\eta_4)$ 是 K^4 的另一组有序基. 假设

$$(\eta_1,\eta_2,\eta_3,\eta_4)=(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)P\quad 其中\quad P=\begin{pmatrix}1&1&1&1\\1&1&-1&-1\\1&-1&1&-1\\1&-1&-1&1\end{pmatrix}.$$

求 φ 在有序基 $\mathcal{C}$ 下的 Gram 矩阵.

习题 6.1.2. 设 $V=\mathbf{M}_n(K)$. 定义

$$\varphi:V\times V\longrightarrow K;\quad (A,B)\longmapsto \mathrm{Tr}(AB).$$

1. 证明 φ 是 V 上的对称双线性型.

2. 令 $n=2$, 取 V 的一组基 $\mathcal{B}=(\varepsilon_{11},\varepsilon_{12},\varepsilon_{21},\varepsilon_{22})$ 如下:

$$\varepsilon_{11}=\begin{pmatrix}1&0\\0&0\end{pmatrix},\quad \varepsilon_{12}=\begin{pmatrix}0&1\\0&0\end{pmatrix},\quad \varepsilon_{21}=\begin{pmatrix}0&0\\1&0\end{pmatrix},\quad \varepsilon_{22}=\begin{pmatrix}0&0\\0&1\end{pmatrix}.$$

求 φ 在 $\mathcal{B}$ 下的 Gram 矩阵 $A_1=\mathcal{M}_{\mathcal{B}}(\varphi)$.

3. 仍设 $n=2$. 另取 V 的 $\mathcal{C}=(\eta_1,\eta_2,\eta_3,\eta_4)$ 为

$$\eta_1=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \eta_2=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad \eta_3=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \eta_4=\begin{pmatrix}0&1\\-1&0\end{pmatrix}.$$

求 φ 在 $\mathcal{C}$ 下的 Gram 矩阵 $A_2=\mathcal{M}_{\mathcal{C}}(\varphi)$.

4. 对于前两个小题中的矩阵 A_1, A_2, 找出一个可逆矩阵 $P\in\mathbf{M}_4(K)$ 使得 $P^TA_1P=A_2$.

习题 6.1.3. 考虑分块对角阵 $A=\begin{pmatrix}A_1&&&\\&A_2&&\\&&A_3&\\&&&A_4\end{pmatrix}$, 其中 A_i 是大小相同的方阵. 令 $A'=\begin{pmatrix}A_4&&&\\&A_3&&\\&&A_1&\\&&&A_2\end{pmatrix}$. 证明 A 与 A' 相合.

习题 6.1.4. 在列向量空间 K^4 上定义双线性型 φ 如下: 对于 $x=(x_1,\cdots,x_4)^T$, $y=(y_1,\cdots,y_4)^T$, 令

$$\varphi(x,y):=-x_1y_3+x_1y_4+x_3y_1+x_3y_4-x_4y_1-x_4y_3.$$

请判断 φ 是否是对称的？是否是满秩的？

习题 6.1.5. (* 本题有一定的技巧性.)

设 $\langle\cdot\,,\,\cdot\rangle$ 是 K-向量空间 V 上的双线性型. 假设对于任意 $u, v \in V$, $\langle u, v\rangle = 0$ 成立时 $\langle v, u\rangle = 0$ 也成立. 证明:

1. 对任意 $x, y, z \in V$ 均有 $\langle x, y\rangle\langle z, x\rangle - \langle x, z\rangle\langle y, x\rangle = 0$.
2. 对任意 $x, y \in V$ 均有 $\langle x, x\rangle\big(\langle x, y\rangle - \langle y, x\rangle\big) = 0$.
3. 假设有向量 $u, v, w \in V$ 满足
$$\langle u, v\rangle \neq \langle v, u\rangle \quad \text{但} \quad \langle u, w\rangle = \langle w, u\rangle\,,\ \langle v, w\rangle = \langle w, v\rangle\,.$$
则
$$\langle u, w\rangle = \langle v, w\rangle = 0\,,\ \langle u, u\rangle = \langle u+w, u+w\rangle = \langle w, w\rangle = 0\,.$$
4. 作为双线性型 $\langle\cdot\,,\,\cdot\rangle$ 要么是对称的要么交错的.

习题 6.1.6. 举例说明: 一个双线性型的左根和右根可以不相等.

习题 6.1.7. 设 $A \in \mathbf{M}_n(K)$. 证明:

1. A 是斜对称阵当且仅当对任意列向量 $x \in K^n$ 均有 $x^TAx = 0$.
2. 若 A 是对称阵, 且对于任意 $x \in K^n$ 均有 $x^TAx = 0$, 则 $A = 0$.

习题 6.1.8. 设 $\langle\cdot\,,\,\cdot\rangle$ 是 n 维 K 向量空间 V 上的非退化双线性型, $(v_1, \cdots, v_n)$ 是 V 的一组有序基.

1. 证明: 对任意 $k \in [\![1, n]\!]$ 及任意 $a_1, \cdots, a_k \in K$, 一定存在向量 $w \in V$ 使得
$$\text{对每个 } i \in [\![1, k]\!] \text{ 均有 } \langle v_i\,,\,w\rangle = a_i\,.$$
2. 证明: 若 $k = n$, 则上一小题中的向量 w 是唯一的, 而且只要 $a_1, \cdots, a_n$ 不全为零, 必有 $w \neq 0$.

习题 6.1.9. 设 φ 是 K-向量空间 V 上的对称或交错双线性型, U, W 是 V 的子空间.

1. 证明: 如果 $V = U + W$ 并且 U 和 W 关于 φ 正交, 那么 $\mathrm{Rad}(\varphi) = \mathrm{Rad}(\varphi|_U) + \mathrm{Rad}(\varphi|_W)$.
2. 假设 $V = U + U^\perp$. 证明: φ 是非退化的当且仅当 $\varphi|_U$ 和 $\varphi|_{U^\perp}$ 都是非退化的.

习题 6.1.10. 设 φ 是由如下表达式给出的 K^4 上的双线性型:
$$\varphi(u, v) = -x_1y_1 + 2x_1y_2 + 2x_2y_1 - 3x_2y_2 + x_2y_4 - 2x_1y_4 + x_4y_2 - 2x_4y_1 + 2x_4y_4\,,$$
其中 $u = (x_1, x_2, x_3, x_4)^T$, $v = (y_1, y_2, y_3, y_4)^T$. 令 $W = \mathrm{span}(e_1, e_2) \subseteq K^4$.

1. 求 $W^\perp$.
2. 证明 $\varphi|_W$ 是非退化的.
3. 将 $v = (1, 2, 3, 4)^T$ 分解为 $v = w + w'$, 其中 $w \in W$, $w' \in W^\perp$.

习题 6.1.11. 设 φ 为 K-向量空间 V 上的双线性型, M 是 V 的子空间. 定义
$$\mathcal{L}(M) = \{v \in V \mid \forall\, x \in M\,,\ \varphi(v, x) = 0\}\,;\quad \mathcal{R}(M) = \{v \in V \mid \forall\, x \in M\,,\ \varphi(x, v) = 0\}\,.$$

1. 证明 $\mathcal{L}(M)$ 和 $\mathcal{R}(M)$ 都是 V 的子空间.

2. 设 $N \subseteq V$ 也是一个子空间. 证明:

$$\mathcal{L}(M+N) = \mathcal{L}(M) \cap \mathcal{L}(N)\,, \quad \mathcal{R}(M+N) = \mathcal{R}(M) \cap \mathcal{R}(N)\,.$$

3. 假设 V 是有限维的, 并且 φ 是非退化的. 证明:

$$\dim \mathcal{L}(M) = \dim \mathcal{R}(M) = \dim V - \dim M\,, \quad \mathcal{R}(\mathcal{L}(M)) = \mathcal{L}(\mathcal{R}(M)) = M\,.$$

4. 假设 V 是有限维的, 并且 φ 是非退化的. 证明: 对任意子空间 $M,\, N \subseteq V$,

$$\mathcal{L}(M \cap N) = \mathcal{L}(M) + \mathcal{L}(N)\,, \quad \mathcal{R}(M \cap N) = \mathcal{R}(M) + \mathcal{R}(N)\,.$$

习题 6.1.12. 设 $f,\, g$ 是 K-向量空间 V 上的线性函数. 假设对于任意 $v \in V$ 均有 $f(v)g(v) = 0$.

证明 $f = 0$ 或 $g = 0$.

习题 6.1.13. 设 φ 是 K-向量空间 V 上的对称双线性型. 假设存在 V 上的线性函数 $f,\, g : V \to K$ 使得

$$\text{对任意 } u,\, v \in V\,,\ \varphi(u,\, v) = f(u)g(v)\,.$$

证明: 存在线性函数 $\ell : V \to K$ 以及非零常数 $\lambda \in K$ 使得

$$\text{对任意 } u,\, v \in V\,,\ \varphi(u,\, v) = \lambda \ell(u)\ell(v)\,.$$

习题 6.1.14. 考虑分块对角阵 $A = \begin{pmatrix} 0 & 0 \\ 0 & A_1 \end{pmatrix} \in \mathbf{M}_n(K)$, $B = \begin{pmatrix} 0 & 0 \\ 0 & B_1 \end{pmatrix} \in \mathbf{M}_n(K)$, 其中 A_1, B_1 均为可逆方阵.

证明: A 与 B 相合的充分必要条件是 A_1 与 B_1 相合.

6.2 二次型的基础理论

本节我们讨论与对称双线性型密切相关的一种数学结构——二次型. 从函数的观点来说, 二次型还与多元二次函数有天然的联系, 这种联系对于多元微积分等分析学的内容也意义重大.

本节中总是以 V 表示域 K 上的一个向量空间. 通常我们可以忽略 $V = 0$ 的情况.

6.2.1 二次型的定义及其等价刻画

定义 6.2.1. 所谓 K-向量空间 V 上的一个**二次型** (quadratic form) 是指满足以下条件的一个函数 $q : V \to K$:

1. **齐次性** (homogeneity): 对于任意 $v \in V$ 和 $c \in K$, 均有 $q(cv) = c^2 q(v)$.

2. **极化** (polarization): 映射

$$b_q \,:\, V \times V \longrightarrow K\,; \quad (u,\, v) \longmapsto \frac{1}{2}\big(q(u+v) - q(u) - q(v)\big)$$

是 V 上的双线性型.

易见上面的双线性型 b_q 是对称的. 我们称之为二次型 q 的**极化型** (polar form).

我们将 V 上的所有二次型构成的集合记为 $\mathrm{Quad}(V)$. 读者可以验证 $\mathrm{Quad}(V)$ 是 $\mathrm{Map}(V, K)$ 的一个子空间. (但是要注意 $\mathrm{Quad}(V)$ 和 $\mathrm{Hom}(V, K)$ 的区别.)

当 q 是 V 上二次型时, 有序对 (V, q) 称为一个**二次型空间** (quadratic space), 以 V 的维数作为它的维数. ■

(6.2.2) 根据定义, 每个二次型 $q \in \mathrm{Quad}(V)$ 可以通过极化得到一个对称双线性型 $b_q \in \mathrm{Sym}(V)$. 换言之, 我们有一个自然的映射

(6.2.2.1) $$B\ :\ \mathrm{Quad}(V) \longrightarrow \mathrm{Sym}(V)\,;\quad q \longmapsto b_q\,.$$

反过来, 如果先给定一个双线性型 $\varphi \in \mathrm{Bil}(V)$, 则映射

(6.2.2.2) $$q_\varphi\ :\ V \longrightarrow K\,;\quad v \longmapsto q_\varphi(v) := \varphi(v, v)$$

是 V 上的一个二次型 (思考题 6.9 (1)), 称为双线性型 φ 的**迹型**[§] (trace form). 于是我们又得到一个映射

(6.2.2.3) $$\widetilde{Q}\ :\ \mathrm{Bil}(V) \longrightarrow \mathrm{Quad}(V)\,;\quad \varphi \longmapsto q_\varphi\,.$$

将 $\widetilde{Q}$ 的定义域限制于子空间 $\mathrm{Sym}(V) \subseteq \mathrm{Bil}(V)$, 我们得到映射

(6.2.2.4) $$Q\ :\ \mathrm{Sym}(V) \longrightarrow \mathrm{Quad}(V)\,;\quad \varphi \longmapsto q_\varphi\,.$$

或者说, 如果用 $\iota : \mathrm{Sym}(V) \to \mathrm{Bil}(V)$ 表示子空间 $\mathrm{Sym}(V)$ 到 $\mathrm{Bil}(V)$ 的自然含入映射, 那么 $Q = \widetilde{Q} \circ \iota$. 换言之, 如下图表交换:

(6.2.2.5) $$\begin{tikzcd} \mathrm{Sym}(V) \arrow[rr, hook, "\iota"] \arrow[dr, "Q"'] & & \mathrm{Bil}(V) \arrow[dl, "\widetilde{Q}"] \\ & \mathrm{Quad}(V) & \end{tikzcd}$$

(形如 $\hookrightarrow$ 的箭头常常用来表示一个单射.) ■

思考题 6.9. 记号如 (6.2.2), 证明:

1. (6.2.2.2) 式中的映射 q_φ 的确是 V 上的二次型.
2. (6.2.2.1), (6.2.2.3) 和 (6.2.2.4) 中的映射都是线性映射.

命题 6.2.3: *记号如 (6.2.2). 则 (6.2.2.1) 和 (6.2.2.4) 中的映射 B 和 Q 互为逆映射. 特别地, 它们都是双射, 而 (6.2.2.3) 中的映射 $\widetilde{Q}$ 为满射.*

证明. 按照定义验证即可. 细节留给读者. (或者课堂授课时讲授.) □

以上命题可以认为是总结了双线性型和二次型之间的联系: 对称双线性型和二次型之间通过取迹型和极化型可以建立双射; 而如果不局限于对称的双线性型, 那么任何双线性型都可以通过取迹型来对应到一个二次型, 只不过这次的对应不再是一一对应.

[§]这个名词在中文文献中似乎并不常见, 不过在有关二次型的外文专著中还是偶有出现. 如果读者不太习惯这个称呼, 不妨直接就把 q_φ 称为**双线性型 φ 对应的二次型**.

思考题 6.10. 记号如 (6.2.2), 对于任意 $\varphi \in \mathrm{Bil}(V)$, 令 φ^T 表示它的转置 (参见思考题 6.2 (2)), 即 $\varphi^T(u, v) = \varphi(v, u)$. 定义映射

$$\sigma\ :\ \mathrm{Bil}(V) \longrightarrow \mathrm{Sym}(V)\ ;\quad \varphi \longmapsto \frac{1}{2}(\varphi + \varphi^T)\,.$$

证明:

1. 复合映射 $\sigma \circ \iota$ 等于集合 $\mathrm{Sym}(V)$ 上的恒等映射, 其中 $\iota : \mathrm{Sym}(V) \hookrightarrow \mathrm{Bil}(V)$ 为自然含入映射.
2. 下面的图表可交换:
$$\begin{tikzcd} \mathrm{Bil}(V) \arrow[rr, "\sigma"] \arrow[dr, "\widetilde{Q}"'] & & \mathrm{Sym}(V) \\ & \mathrm{Quad}(V) \arrow[ur, "B"'] & \end{tikzcd}$$
其中 B 和 $\widetilde{Q}$ 的定义如 (6.2.2.1) 和 (6.2.2.3) 式.

例 6.2.4. 既然明白了二次型与双线性型之间的对应关系, 便不难举出二次型的例子.

1. 映射
$$q\ :\ V \longrightarrow K\ ;\quad f \longmapsto f(0)^2$$
是向量空间 $V = K[X]$ 上的一个二次型. 它实际上是例 6.1.4 (2) 中双线性型的迹型.
2. 设 V 为 n 维列向量空间 K^n, $A \in \mathbf{M}_n(K)$. 则映射
$$q_A\ :\ K^n \longrightarrow K\ ;\quad x \longmapsto x^T A x$$
是 V 上的二次型. 它的一个常用记号是 $\langle A \rangle$. 容易看出, 这个二次型是例 6.1.4 (3) 中双线性型 $\langle \cdot\,,\,\cdot \rangle_A$ 的迹型. ■

(6.2.5) 鉴于其重要性, 我们来继续考查例 (6.2.4) 中的第二个例子.

当我们令 A 在集合 $\mathbf{M}_n(K)$ 中变动时, 通过 $A \mapsto \langle A \rangle$ 这个法则可以得到一个映射

(6.2.5.1) $$\mathbf{M}_n(K) \longrightarrow \mathrm{Quad}(K^n)\ ;\quad A \longmapsto \langle A \rangle\,.$$

这是一个满的线性映射, 但一般来说不是单射. 不过, 如果暂时用 $\mathbf{SymM}_n(K)$ 表示 $\mathbf{M}_n(K)$ 中对称矩阵构成的子集, 那么将 (6.2.5.1) 限制于 $\mathbf{SymM}_n(K)$ 得到的映射

(6.2.5.2) $$\mathbf{SymM}_n(K) \longrightarrow \mathrm{Quad}(K^n)\ ;\quad A \longmapsto \langle A \rangle\,.$$

却是一个双射. 事实上, 若令 $\tau : \mathbf{SymM}_n(K) \hookrightarrow \mathbf{M}_n(K)$ 表示自然含入映射, 那么以下图表是交换的:

(6.2.5.3) $$\begin{tikzcd} \mathbf{SymM}_n(K) \arrow[r, hook, "\tau"] \arrow[d, "A \mapsto \langle \cdot\,,\,\cdot \rangle_A"'] & \mathbf{M}_n(K) \arrow[r, "A \mapsto \langle A \rangle"] \arrow[d, "A \mapsto \langle \cdot\,,\,\cdot \rangle_A"] & \mathrm{Quad}(K^n) \arrow[d, equal] \\ \mathrm{Sym}(K^n) \arrow[r, hook, "\iota"'] & \mathrm{Bil}(K^n) \arrow[r, "\widetilde{Q}\,:\,\varphi \mapsto q_\varphi"'] & \mathrm{Quad}(K^n) \end{tikzcd}$$

根据命题 6.1.6 (及其证明过程), 这里竖直方向的前两个映射都是同构.

思考题 6.11. 证明: (6.2.5.1) 中的映射是满射, 而 (6.2.5.2) 中的映射是双射.

然后对于 $n = 2$ 的情形, 具体举出两个不同的矩阵 $A, B \in \mathbf{M}_n(K)$ 使得 $\langle A \rangle$ 和 $\langle B \rangle$ 两个二次型相等.

利用二次型与对称双线性型的一一对应, 我们还可以将双线性型的相关概念转移到二次型的理论中来.

定义 6.2.6. 设 $q \in \mathrm{Quad}(V)$, $b_q \in \mathrm{Sym}(V)$ 为 q 的极化型.

1. 我们将二次型 q 的**根**定义为对称双线性型 b_q 的根, 记之为 $\mathrm{Rad}(q)$.

   如果 $\mathrm{Rad}(q) = 0$ (亦即 b_q 是非退化的), 则称二次型 q 是**非退化的**.

2. 假设 V 是有限维的, $\mathcal{B} = (v_1, \cdots, v_n)$ 是 V 的一组有序基.

   我们将 b_q 的 Gram 矩阵称为二次型 q 在有序基 $\mathcal{B}$ 下的 Gram **矩阵**, 记为 $\mathcal{M}_{\mathcal{B}}(q)$. 将 $\mathcal{M}_{\mathcal{B}}(q)$ 的秩称为 q 的**秩**, 记为 $\mathrm{rank}(q)$.

   当 q 在某一组有序基下的 Gram 矩阵是满秩的 (亦即非奇异的), 我们就说 q 是**满秩的**或**非奇异的**. 当然, 根据推论 6.1.20 可知, 这等价于说 q 是非退化的. ■

(6.2.7) 根据 (6.2.5.2) 式建立的双射, 我们可以从对称矩阵的角度完整地描述列向量空间 K^n 上的所有二次型. 事实上, (6.2.5.2) 的逆映射可以通过在标准基 $\mathcal{E} = (e_1, \cdots, e_n)$ 下取 Gram 矩阵得到.

现在我们换用函数的观点再来看 K^n 上二次型的描述.

任取 $q \in \mathrm{Quad}(K^n)$. 根据前面的讨论我们知道, 存在唯一的对称矩阵 $A = (a_{ij}) \in \mathbf{M}_n(K)$ 使得 $q = \langle A \rangle$. 通过将 K^n 中向量在标准基下的坐标形式 $x = (x_1, \cdots, x_n)^T$, 可以代入定义式计算得到

$$q(x) = \langle A \rangle(x) = x^T A x = (x_1, \cdots, x_n) \cdot (a_{ij}) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \ .$$

注意到 $A = (a_{ij})$ 是对称矩阵, 所以, 若令

$$c_{ij} = \begin{cases} a_{ii} & \text{若 } i = j \\ 2a_{ij} & \text{若 } i < j \end{cases} \tag{6.2.7.1}$$

(当 $i > j$ 时我们不需要定义 c_{ij}), 则上面关于 $q(x)$ 的表达式可以改写为

$$q(x) = q(x_1, \cdots, x_n) = \sum_{1 \le i \le j \le n} c_{ij} x_i x_j \ . \tag{6.2.7.2}$$

所以, 如果我们将形如

$$\sum_{1 \le i \le j \le n} c_{ij} x_i x_j \quad \text{其中 } c_{ij} \in K$$

的表达式称为 K 上的一个关于变量 $x_1, \cdots, x_n$ 的 n 元**二次齐次函数** (quadratic homogeneous function) 或**二次型函数** (quadratic function), 那么 (6.2.7.2) 式意味着 K^n 上的一个二次型和 K 上的 n 二次齐次函数本质上是一回事. 事实上, 如果不是预先给定二次型 q 或对称矩阵 A, 而是先给定一个 $\sum_{1 \le i \le j \le n} c_{ij} x_i x_j$ 形式的表达式, 或者说先给定一组常数 c_{ij}, $1 \le i \le j \le n$, 那么我们可以通过计算验证: 按照 (6.2.7.2) 式定义的映射 $q : K^n \to K$ 的确是 K^n 上的一个二次型.

我们还可以再稍微变换一下视角. 若以 $X_1, \cdots, X_n$ 为一些形式符号, 那么任何形如

$$f = \sum_{1 \le i \le j \le n} c_{ij} X_i X_j \quad \text{其中 } c_{ij} \in K$$

的表达式称为一个**系数**在 K 中的以 $X_1, \cdots, X_n$ 为**未定元**的 n 元**二次齐次多项式** (quadratic homogeneous polynomial). 我们可以直观的感觉到, 通过把二次齐次多项式视为 K 上的 n 元函数,

可以得到二次齐次多项式和二次齐次函数之间的一一对应.[¶] 我们把所有上述形式的二次齐次多项式构成的集合记为 $K[X_1,\cdots,X_n]_2$. 即

$$K[X_1,\cdots,X_n]_2 := \left\{ \sum_{1\le i\le j\le n} c_{ij}X_iX_j \;\middle|\; \text{其中每个 } c_{ij}\in K \right\}.$$

于是, 通过 (6.2.7.1) 式我们得到了一个双射

$$\begin{aligned} K[X_1,\cdots,X_n]_2 &\xrightarrow{\sim} \mathrm{Quad}(K^n) \\ f=\sum_{1\le i\le j\le n} c_{ij}X_iX_j &\longmapsto q(x_1,\cdots,x_n) := \sum_{1\le i\le j\le n} c_{ij}x_ix_j\ . \end{aligned} \tag{6.2.7.3}$$

也就是说, K^n 上的二次型可以和系数取自 K 的 n 元二次齐次多项式等同看待.

综上所述, K^n 上的二次型至少可以从以下四个实质等价的角度来理解: (1) 系数在 K 中的 n 元二次齐次多项式; (2) 如定义 6.2.1 所描述的具有二次齐次性和可极化性质的映射 $q : K^n \to K$; (3) K^n 上的对称双线性型; (4) 系数取自 K 的 n 阶对称矩阵.

具体来说, 我们可以通过以下图表理解四者间的相互关系:

(6.2.7.4)
$$\begin{tikzcd}
K[X_1,\cdots,X_n]_2 \arrow[rr, leftrightarrow, "\sim", "\text{多项式视为函数}"'] \arrow[d, shift left, "(6.2.7.6)"] & & \mathrm{Quad}(K^n) \arrow[d, shift left, "\text{取极化型}"] \\
\mathbf{SymM}_n(K) \arrow[u, shift left, "(6.2.7.5)"] \arrow[rr, shift left, "A\mapsto\langle\cdot\,,\,\cdot\rangle_A"] & & \mathrm{Sym}(K^n) \arrow[u, shift left, "\text{取迹型}"] \arrow[ll, shift left, "\text{标准基下 Gram 矩阵}"]
\end{tikzcd}$$

这里最左边两个箭头的含义在下面解释: 事实上, 从前面的 (6.2.7.1) 式可以看出从对称矩阵到二次齐次多项式的对应法则为

$$\mathbf{SymM}_n(K) \longrightarrow K[X_1,\cdots,X_n]_2\ ; \quad A=(a_{ij}) \longmapsto \sum_{i=1}^n a_{ii}X_i^2 + \sum_{1\le i<j\le n} 2a_{ij}X_iX_j\ . \tag{6.2.7.5}$$

容易看出, 这一映射的逆映射为

$$\begin{aligned} K[X_1,\cdots,X_n]_2 &\longrightarrow \mathbf{SymM}_n(K) \\ f=\sum_{1\le i\le j\le n} c_{ij}X_iX_j &\longmapsto \begin{pmatrix} c_{11} & \frac{1}{2}c_{12} & \cdots & \frac{1}{2}c_{1n} \\ \frac{1}{2}c_{12} & c_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{2}c_{n-1,n} \\ \frac{1}{2}c_{1n} & \cdots & \frac{1}{2}c_{n-1,n} & c_{nn} \end{pmatrix} \end{aligned} \tag{6.2.7.6}$$

我们请读者认真确认自己理解图表 (6.2.7.4) 的交换性. 这个图表概括了描述 K^n 上二次型的四种最基本语言. 今后我们可能会随时根据论述的方便在这四种语言之间自由切换, 希望读者对此能够逐渐习惯. ■

[¶]严格来说, 我们目前还不方便十分严谨地给出这个论断的理论证明. 不过, 我们不建议读者在这里钻这个牛角尖, 只需承认这个直观而正确的结论即可.

例 6.2.8. 根据 (6.2.7) 一段的讨论, 三元二次型 $q(X_1, X_2, X_3) = X_1^2 - 2X_1X_2 + 3X_1X_3 - X_2X_3 + 4X_3^2$ 的 Gram 矩阵是 $A = \begin{pmatrix} 1 & -1 & \frac{3}{2} \\ -1 & 0 & -\frac{1}{2} \\ \frac{3}{2} & -\frac{1}{2} & 4 \end{pmatrix}$. 它对应的双线性型为

$$\begin{aligned} \langle \cdot\,, \cdot \rangle_A \;:\; & K^3 \times K^3 \longrightarrow K \\ \left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \right) & \longmapsto (x_1, x_2, x_3) A \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\ & = x_1y_1 - (x_1y_2 + x_2y_1) + \frac{3}{2}(x_1y_3 + x_3y_1) - \frac{1}{2}(x_2y_3 + x_3y_2) + 4x_3y_3 \,. \end{aligned}$$

请读者自己留意观察以上双线性型和二次型 q 表达式的区别和联系. ■

(6.2.9) 对于一般的 n 维 K-向量空间 V, 我们知道通过选定 V 的一组有序基 $\mathcal{B}$ 可以建立一个同构

$$\mathcal{M}_{\mathcal{B}} \,:\, V \xrightarrow{\sim} K^n \,; \quad v \longmapsto \mathcal{M}_{\mathcal{B}}(v) \,.$$

于是, 对于任意 $q \in \mathrm{Quad}(K^n)$, 复合映射 $q \circ \mathcal{M}_{\mathcal{B}} : V \to K$ 给出 V 上一个二次型. 类似地, 将一个 K^n 上的对称双线性型 $\varphi \in \mathrm{Sym}(K^n)$ 与 $\mathcal{M}_{\mathcal{B}} \times \mathcal{M}_{\mathcal{B}}$ 复合可以得到 V 上一个对称双线性型

$$V \times V \xrightarrow{\mathcal{M}_{\mathcal{B}} \times \mathcal{M}_{\mathcal{B}}} K^n \times K^n \xrightarrow{\varphi} K \,.$$

此时我们有如下交换图表

(6.2.9.1)

$$\begin{tikzcd}
\mathrm{Quad}(V) \arrow[r, shift left, "取极化型"] & \mathrm{Sym}(V) \arrow[l, shift left, "取迹型"] \arrow[dr, equal] & & & \\
& & \mathrm{Sym}(V) \arrow[r, hook, "\iota"] \arrow[dd, "取\ \mathrm{Gram}\ 矩阵"] & \mathrm{Bil}(V) \arrow[r, "\varphi \mapsto \frac{1}{2}(\varphi + \varphi^T)"] \arrow[dd, "取\ \mathrm{Gram}\ 矩阵"] & \mathrm{Sym}(V) \arrow[dd, "取\ \mathrm{Gram}\ 矩阵"] \\
\mathrm{Quad}(K^n) \arrow[uu, "q \mapsto q \circ \mathcal{M}_{\mathcal{B}}"] \arrow[r, shift left, "取极化型"] & \mathrm{Sym}(K^n) \arrow[uu, "\varphi \mapsto \varphi \circ \mathcal{M}_{\mathcal{B}} \times \mathcal{M}_{\mathcal{B}}"] \arrow[l, shift left, "取迹型"] \arrow[dr, shift left, "标准基下矩阵"] & & & \\
& & \mathbf{SymM}_n(K) \arrow[r, hook, "\tau"] \arrow[ul, shift left, "A \mapsto \langle \cdot\,, \cdot \rangle_A"] & \mathbf{M}_n(K) \arrow[r, "M \mapsto \frac{1}{2}(M + M^T)"] & \mathbf{SymM}_n(K)
\end{tikzcd}$$

这里每当出现双向的箭头时, 其实我们可以任意选取一个方向的箭头都不会影响图表的交换性. 而且, 竖直方向的箭头都是同构. ■

6.2.2 配方法化二次型为标准形

(6.2.10) 设 V 是 n 维 K-向量空间, $q \in \mathrm{Quad}(V)$, $\varphi \in \mathrm{Sym}(V)$ 是 q 的极化型. 对于任意 $u, v \in V$, 我们有

$$q(v) = \varphi(v, v) \,, \quad \varphi(u, v) = \frac{q(u+v) - q(u) - q(v)}{2} \,.$$

根据定理 6.1.15 我们知道: 存在 V 的一组基 $\mathcal{B}$ 使得 φ 的 Gram 矩阵 (亦即 q 的 Gram 矩阵) 为对角阵. 假设 $\mathcal{M}_{\mathcal{B}}(\varphi) = \mathcal{M}_{\mathcal{B}}(q) = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}$. 对于任意 $v \in V$, 若记 $\mathcal{M}_{\mathcal{B}}(v) = (x_1, \cdots, x_n)^T$, 则

$$q(v) = \varphi(v,\, v) = (x_1, \cdots, x_n)^T \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = d_1 x_1^2 + \cdots + d_n x_n^2\,.$$

也就是说, Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(q)$ 化成相合标准形时, q 关于向量坐标的表达式是形如 $d_1 x_1^2 + \cdots + d_n x_n^2$ 的只含 (带系数) 平方项的二次齐次多项式形式. 我们通常把这样的表达式也称为二次型 q 的一个**标准形** (canonical form). 找出这样一种表达式的过程就称为“将 q 化为标准形”的过程. ■

至于从具体计算的角度如何将给定的二次形化为标准形, 可用的方法有不少. 现在我们只想介绍一种非常简明实用且思路上容易掌握的方法, 即所谓的**配方法** (method of completing squares).

我们先通过例子让读者对配方法有直观的感觉, 然后再从理论上证明这种方法的合理性. 我们只讨论列向量 K^n 上的二次型 (一般的情况可以通过图表 (6.2.9.1) 化归到这种情况), 并且主要采用二次齐次多项式 (或函数) 的观点来理解二次型.

例 6.2.11. 将二次型 $q(x,\, y,\, z) = x^2 + y^2 + z^2 + 4xy - 6xz - 4yz$ 化成标准形.

如果将 q 写成矩阵形式, 那么我们看到

$$q(x,\, y,\, z) = (x,\, y,\, z)^T A \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad \text{其中} \qquad A = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & -2 \\ -3 & -2 & 1 \end{pmatrix}.$$

我们的目标是找到可逆矩阵 P 使得 $P^TAP = D$ 是个对角阵 $\begin{pmatrix} d_1 & & \\ & d_2 & \\ & & d_3 \end{pmatrix}$.

将 $A = (P^T)^{-1} D P^{-1} = (P^{-1})^T D P^{-1}$ 代入上式可得

$$q(x,\, y,\, z) = (x,\, y,\, z)^T (P^{-1})^T D P^{-1} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \left[P^{-1} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right]^T D \left[P^{-1} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right].$$

因此, 如果通过

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = P^{-1} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{6.2.11.1}$$

这种方式引入三个新的变量 $x',\, y',\, z'$, 那么 q 用新的变量可以表达为

$$q(x',\, y',\, z') = (x',\, y',\, z') D \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = d_1 (x')^2 + d_2 (y')^2 + d_3 (z')^2\,.$$

这就是我们需要的一种标准形.

所以, 问题的本质就是找到一个 (6.2.11.1) 形式的变量代换, 使得 q 在使用新的变量表达时可以写成各个变量平方的线性组合. 因为 (6.2.11.1) 式中 P 是个可逆变换, 而且新的变量和旧的变量之

间显然是线性依赖的关系, 所以我们通常把 (6.2.11.1) 式称为一个**可逆的线性变量代换** (invertible linear change of variables) 或者简称**线性变量代换** (linear change of variables). (当然, "变量"、"代换"这两个词可以分别用同义词"变元"、"替换"等代替.)

前面写出的矩阵 A 是二次型 q 在 K^3 的标准基 $\mathcal{E} = (e_1, e_2, e_3)$ 下的矩阵, 而 $(x, y, z)^T$ 可以认为是 K^3 中向量在标准基下的坐标. 那么 $(x', y', z')^T$ 可以认为是同一向量在新的一组有序基 $\mathcal{B}$ 下的坐标. 根据新旧坐标之间的转换关系, (6.2.11.1) 式中的矩阵 P 恰好应该是标准基 $\mathcal{E}$ 到 $\mathcal{B}$ 的过渡矩阵. 换言之, 如果 v_1, v_2, v_3 表示矩阵 P 的三个列, 那么 q 在有序基 $\mathcal{B} = (v_1, v_2, v_3)$ 下的 Gram 矩阵就是 $D = \begin{pmatrix} d_1 & & \\ & d_2 & \\ & & d_3 \end{pmatrix}$.

接下来我们具体来看如何找到合适的变量替换. 我们要讲的"配方法"核心思想就是从原来的表达式

$$q(x, y, z) = x^2 + y^2 + z^2 + 4xy - 6xz - 4yz$$

中将各个变量 x, y, z 逐渐放入凑出来的"平方"之中. 以当前给的二次型为例, 我们先集中注意力在第一个变元 x 上, 将以上 q 的表达式中其他项合并起来看成 x 的系数. 于是得到

$$q = x^2 + x(4y - 6z) + (y^2 + z^2 - 4yz)\,.$$

如果希望把含 x 的项全部放入一个 (可能带系数的) 平方项中, 那么我们要给 $x^2 + x(4y - 6z) = x^2 + 2x(2y - 3z)$ 补上一项 $(2y - 3z)^2$. 因此我们继续做以下恒等变形

$$\begin{aligned} q &= x^2 + x(4y - 6z) + (y^2 + z^2 - 4yz) \\ &= \left[x^2 + 2x(2y - 3z) + (2y - 3z)^2\right] - (2y - 3z)^2 + (y^2 + z^2 - 4yz) \\ &= (x + 2y - 3z)^2 - 3y^2 - 8z^2 + 8yz\,. \end{aligned}$$

以上最后得到的平方项 $(x + 2y - 3z)^2$ 之后就不需要再理睬了. 我们只需要对 $-3y^2 - 8z^2 + 8yz$ 再采取之前的对策——将所有含 y 的项凑到一个带系数的平方项之内. 为此, 只要按如下方式做恒等变形

$$\begin{aligned} -3y^2 - 8z^2 + 8yz &= -3\left(y^2 - \frac{8}{3}yz\right) - 8z^2 \\ &= -3\left(y - \frac{4}{3}z\right)^2 + 3\left(\frac{4}{3}z\right)^2 - 8z^2 \\ &= -3\left(y - \frac{4}{3}z\right)^2 - \frac{8}{3}z^2\,. \end{aligned}$$

综上可得

$$q = (x + 2y - 3z)^2 - 3\left(y - \frac{4}{3}z\right)^2 - \frac{8}{3}z^2\,.$$

也就是说, 如果令

$$\begin{cases} x' = x + 2y - 3z \\ y' = y - \frac{4}{3}z \\ z' = z \end{cases} \qquad \text{亦即} \qquad \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 2 & -3 \\ & 1 & -\frac{4}{3} \\ & & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

那么 q 就被化成了标准形: $q = (x')^2 - 3(y')^2 - \frac{8}{3}(z')^2$.

注意, 这里的矩阵 $\begin{pmatrix} 1 & 2 & -3 \\ & 1 & -\frac{4}{3} \\ & & 1 \end{pmatrix}$ 在 (6.2.11.1) 式中对应于矩阵 P^{-1}, 而不是 P! ∎

为了帮助读者理解消化上面例子中的配方法操作过程, 我们再看一个例子.

例 6.2.12. 将二次型 $q(x, y, z) = (x-y)^2 + (y-z)^2 + (z-x)^2$ 化为标准形.

乍一看这个例子好像十分简单, 只要做变量替换

$$\begin{cases} X = x - y \\ Y = y - z \\ Z = z - x \end{cases} \tag{6.2.12.1}$$

即可立即将 q 的表达式变成 $q = X^2 + Y^2 + Z^2$.

但是, 这种做法实际上有个致命错误: 新的三个变量 X, Y, Z 之间不是独立的, 比如 $Z = -X - Y$ 就是三者之间一个非平凡的线性关系. 如果考查矩阵表达式

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

我们会发现这里的系数矩阵 $\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}$ 不是可逆阵! 因此 (6.2.12.1) 式展示的变量替换是不可行的, 初学者应当谨慎小心这一点.

正确的做法是先做变量替换

$$\begin{cases} X_1 = x - y \\ Y_1 = y - z \end{cases} \tag{6.2.12.2}$$

由此将 q 的表达式整理为

$$q = X_1^2 + Y_1^2 + (X_1 + Y_1)^2 = 2X_1^2 + 2Y_1^2 + 2X_1Y_1 = 2\left(X_1 + \frac{1}{2}Y_1\right)^2 + \frac{3}{2}Y_1^2 .$$

于是, 接着做变量替换

$$\begin{cases} X = X_1 + \frac{1}{2}Y_1 \\ Y = Y_1 \end{cases} \tag{6.2.12.3}$$

即可得到

$$q = 2X^2 + \frac{3}{2}Y^2 .$$

以上操作当然没有任何问题, 但一件稍显奇怪的事情是: 原来 q 的表达式用到了 x, y, z 三个变量, 而以上操作只引入两个新的变量就把 q 化成了对角形. 那么第三个新变量该如何选择呢?

其实这个现象的解释也不难. 它只是说明了第三个新变量无论如何选择, 在最终的表达式中并不出现. 因此, 第三个变量的选择只要能够保证从原来三个变量到新的三个变量之间是个可逆的变量替换即可. 在这个例子中, 综合 (6.2.12.2) 和 (6.2.12.3) 式可以看出, 最终的变量 X, Y 与原来三个变量 x, y, z 的关系为

$$\begin{cases} X = x - \frac{1}{2}y - \frac{1}{2}z \\ Y = y - z \end{cases}$$

现在只要补充第三个变量 Z 的选取, 使得 x, y, z 到 X, Y, Z 之间的变换是可逆线性变换即可. 不难看出, $Z=z$ 就是一种可行的选择. 由此得到 q 的一个标准形为:

$$q=2X^2+\frac{3}{2}Y^2+0\cdot Z^2\,.$$

综上所述, 我们最终选择的变量替换为

$$\begin{cases} X=x-\frac{1}{2}y-\frac{1}{2}z \\ Y=y-z \\ Z=z \end{cases} \quad \text{亦即} \quad \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}=\begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ & 1 & -1 \\ & & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

此处的系数矩阵 $\begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ & 1 & -1 \\ & & 1 \end{pmatrix}$ 显然是可逆阵. ■

接下来的例子为我们展示了实施配方法时需要特别注意的一种情况.

例 6.2.13. 将二次型 $q(x, y, z)=xy+yz+xz$ 化为标准形.

在这个例子中, x, y, z 三个变量在表达式中都没有以平方的形式出现, 要将它们直接凑到某个平方项里面似乎是办不到的. 因此, 我们此时需要一个额外的技巧. 我们做如下线性变量替换

$$X_1:=\frac{x+y}{2}\,,\ Y_1=\frac{x-y}{2}\,,\ Z_1=z\,.$$

(这相当于令 $x=X_1+Y_1$, $y=X_1-Y_1$.) 这样可以将 q 的表达式化为

$$q=(X_1+Y_1)(X_1-Y_1)+(X_1-Y_1)Z_1+(X_1+Y_1)Z_1=X_1^2-Y_1^2+2X_1Z_1\,.$$

现在我们可以继续施行以前的“凑平方”方案了. 于是得到如下变形

$$q=X_1^2+2X_1Z_1+Z_1^2-Y_1^2-Z_1^2=(X_1+Z_1)^2-Y_1^2-Z_1^2\,.$$

综上可知, 变量替换

$$\begin{cases} X=X_1+Z_1=\frac{x+y}{2}+z \\ Y=Y_1=\frac{x-y}{2} \\ Z=Z_1=z \end{cases}$$

可以将二次型 q 化为标准形 $q=X^2-Y^2-Z^2$.

(6.2.14) 现在我们从理论上总结一下前面介绍的配方法, 并论证其算法合理性 (即任意 n 元二次型总是可以用这种方法化为标准形). 我们采用二次齐次多项式的观点来看待二次型.

设 $q=\sum_{1\le i\le j\le n}^{n} a_{ij}X_iX_j$ 是一个 n 元二次型. 我们所谓的配方法实际上就是分情况执行以下操作:

首先观察平方项 $a_{11}X_1^2, a_{22}X_2^2, \cdots, a_{nn}X_n^2$ 的系数 $a_{11}, \cdots, a_{nn}$ 是否全部为 0.

情况 1: 如果 $a_{11}, \cdots, a_{nn}$ 不全为零, 不妨设 $a_{11}\ne 0$.

这种情况下我们可以做如下变形

$$q=a_{11}X_1^2+X_1(a_{12}X_2+\cdots+a_{1n}X_n)+r(X_2,\cdots,X_n)$$

其中 $r(X_2,\cdots,X_n)$ 是只含 $n-1$ 个变元 $X_2,\cdots,X_n$ 的二次型. 由此我们可以将出现 X_1 的项全部凑入一个 (带系数的) 平方项之中, 即做如下变形

$$\begin{aligned}
q &= a_{11}\left(X_1^2 + 2X_1\cdot\frac{a_{12}X_2+\cdots+a_{1n}X_n}{2a_{11}} + \left(\frac{a_{12}X_2+\cdots+a_{1n}X_n}{2a_{11}}\right)^2\right)\\
&\quad - a_{11}\left(\frac{a_{12}X_2+\cdots+a_{1n}X_n}{2a_{11}}\right)^2 + r(X_2,\cdots,X_n)\\
&= a_{11}\left(X_1 + \frac{a_{12}X_2+\cdots+a_{1n}X_n}{2a_{11}}\right)^2\\
&\quad - a_{11}\left(\frac{a_{12}X_2+\cdots+a_{1n}X_n}{2a_{11}}\right)^2 + r(X_2,\cdots,X_n)
\end{aligned}$$

此时

$$q_1 := -a_{11}\left(\frac{a_{12}X_2+\cdots+a_{1n}X_n}{2a_{11}}\right)^2 + r(X_2,\cdots,X_n)$$

是只含 $X_2,\cdots,X_n$ 的二次型. 而变量替换

$$Y_1 := X_1 + \cdot\frac{a_{12}X_2+\cdots+a_{1n}X_n}{2a_{11}}\ ,\ Y_2 = X_2\,,\cdots,Y_n = X_n \tag{6.2.14.1}$$

可将 q 的表达式化为

$$q = a_{11}Y_1^2 + q_1(Y_2,\cdots,Y_n)\ . \tag{6.2.14.2}$$

很明显, (6.2.14.1) 式展示的变量替换是可逆的.

情况 2: 现在假设 q 的表达式中平方项的系数 $a_{11},\cdots,a_{nn}$ 全部为零.

因为我们可以忽略 q 的系数全部为零的情况, 所以此时存在一对指标 $i,\,j\in[\![1,\,n]\!]$, $i<j$ 使得 $a_{ij}\neq 0$. 不妨设 $a_{12}\neq 0$.

此时 q 的表达式为

$$q = a_{12}X_1X_2 + X_1\left(\sum_{3\le j\le n} a_{1j}X_j\right) + X_2\left(\sum_{3\le j\le n} a_{2j}X_j\right) + \sum_{3\le i<j\le n} a_{ij}X_iX_j\,.$$

现在我们做变量替换

$$Y_1 = \frac{X_1+X_2}{2}\ ,\ Y_2 = \frac{X_1-X_2}{2}\ ,\ Y_3 = X_3\,,\cdots,Y_n = X_n \tag{6.2.14.3}$$

可以将 q 的表达式化为

$$q = a_{12}(Y_1^2 - Y_2^2) + (Y_1+Y_2)\left(\sum_{3\le j\le n} a_{1j}Y_j\right) + (Y_1-Y_2)\left(\sum_{3\le j\le n} a_{2j}Y_j\right) + \sum_{3\le i<j\le n} a_{ij}Y_iY_j$$

观察现在的表达式可以看出 Y_1^2 的系数为 $a_{12}\neq 0$. 因此, 接下来我们可以按照情况 1 的方式继续做可逆的线性变量替换. 因为 (6.2.14.1) 和 (6.2.14.3) 两种方式给出的变量替换都是可逆的, 所以最终综合起来的效果仍是可逆的变量替换.

综上可知, 我们总是可以通过一个可逆的线性变量替换将 q 写成一个带系数的平方项与一个变元数量严格减少的二次型 q_1 之和 (如 (6.2.14.2) 式). 接下来对 q_1 继续实施前面所述的操作, 最终可由归纳法知道一定能把 q 化为仅含带系数平方项的表达式之和, 即把 q 化成了标准形. ■

6.2.3　实二次型和复二次型的规范形

我们现在将之前关于二次型的一般性讨论应用到两种最常用的情况: K 为实数域 $\mathbb{R}$ 或者复数域 $\mathbb{C}$ 的情况.

复数域的情况应当是最简单的.

定理 6.2.15:　复向量空间 $\mathbb{C}^n$ 上的任意二次型 q 都可以经过可逆的线性变量替换化成如下形式:

$$q = X_1^2 + \cdots + X_r^2$$

而且 (在不考虑表示变元的字母可任意替换的意义下) 这种形式是由 q 唯一确定的, 称为 q 的**规范形** (normal form).

换用矩阵的语言来说, 任意一个对称复方阵 $A \in \mathbf{M}_n(\mathbb{C})$ 都在 $\mathbb{C}$ 上相合于唯一一个 $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ 形式的 (分块) 矩阵.

证明. 所涉及的表达式实际上只依赖于其中出现变元的数量 r. 而这里的 r 等于二次型 q 的秩, 所以它是由 q 唯一决定的.

关于存在性, 我们首先根据二次型标准形的存在性知道, 一定可以经过变量替换将 q 的表达式化为 $q = d_1x_1^2 + \cdots + d_rx_r^2$, 其中每个 $d_i \in \mathbb{C}$ 都不为 0. 然后做变量替换

$$X_i = \sqrt{d_i}\, x_i\ ,\ i = 1, \cdots\ ,\ n$$

即可将 q 改写为 $q = X_1^2 + \cdots + X_r^2$. □

注意, 以上证明过程中用到 $\mathbb{C}$ 的一个重要性质: 任何复数 $d \in \mathbb{C}$ 都在 $\mathbb{C}$ 中有平方根. 这个性质当然换成其它的域 K 可能就不再成立. (在代数学领域, 一般我们允许采用 $\sqrt{d}$ 这个记号来表示 d 的一个平方根. 例如, $\sqrt{-1}$ 这个记号是允许使用的. 虽然 -1 在 $\mathbb{C}$ 中有两个平方根, $\sqrt{-1}$ 这个记号不能明确表示出究竟是二者中的哪一个, 但在我们的讨论中通常这一点不重要, 所以使用 $\sqrt{-1}$ 这个记号是没有关系的.)

定理 6.2.15 表明: 若想区分复数域上的两个二次型是否等价 (即通过变量替换可以将其中一个表达式化为另一个), 只需要考查维数和秩两个不变量即可. 或者说, 对于相同大小的两个对称复方阵, 它们在 $\mathbb{C}$ 上相合的充分必要条件是二者的秩相等. (我们提醒读者注意, 这里我们要求是对称矩阵. 如果两个 n 阶复方阵的秩相同, 但是一个对称一个不对称, 那么显然二者不能相合.)

思考题 6.12. 写出一个可逆矩阵 $P \in \mathbf{M}_3(\mathbb{C})$ 使得 $P^T \begin{pmatrix} -1 & & \\ & 4 & \\ & & -9 \end{pmatrix} P = I_3$.

定理 6.2.16 (Sylvester **惯性定律** Sylvester's law of inertia):　实向量空间 $\mathbb{R}^n$ 上的任意二次型 q 都可以经过可逆变量替换化为如下形式:

$$q = X_1^2 + \cdots + X_p^2 - X_{p+1}^2 - \cdots - X_r^2$$

而且这种形式是唯一的, 称为 q 的**规范形** (normal form). 这里, 系数为 1 的平方项个数 p 称为 q 的**正惯性指数** (positive index of inertia), 系数为 -1 的平方项个数 $s := r - p$ 称为 q 的**负惯性指数** (negative index of inertia), 二者的差 $p - s = 2p - r$ 称为 q 的**符号差** (signature).

换用矩阵语言表述: 任意实对称阵 $A \in \mathbf{M}_n(\mathbb{R})$ 在 $\mathbb{R}$ 上相合于唯一一个 $\begin{pmatrix} I_p & & \\ & -I_s & \\ & & 0 \end{pmatrix}$ 形式的 (分块) 矩阵.

证明. 存在性是容易证明的. 先将 q 化成标准形 $q = a_1x_1^2 + \cdots + a_rx_r^2$, 其中 $a_i \in \mathbb{R}$ 均不为零. 不妨设 $a_1, \cdots, a_p$ 大于零, 而其他系数 $a_{p+1}, \cdots, a_r$ 小于零. 于是 q 可改写为

$$q = d_1x_1^2 + \cdots + d_px_p^2 - d_{p+1}x_{p+1}^2 - \cdots - d_rx_r^2$$

其中每个 d_i 都是正实数. 只需再做变量替换 $X_i = \sqrt{d_i}\,x_i$ 即可将 q 化成定理中陈述的形式.

为证明唯一性, 首先注意到非零系数的总数 r 等于 q 的秩, 它是由 q 唯一确定的. 我们只需要再证明系数为 1 的平方项数目也是固定的.

假设有两种不同的变量替换将 q 化成以下两种形式

$$q = X_1^2 + \cdots + X_p^2 - X_{p+1}^2 - \cdots - X_r^2 \quad \text{及} \quad q = Y_1^2 + \cdots + Y_l^2 - Y_{l+1}^2 - \cdots - Y_r^2\,.$$

这相当于说, 在 $\mathbb{R}^n$ 的一组有序基 $\mathcal{B} = (v_1, \cdots, v_p, \cdots, v_n)$ 下, 若将向量 $v \in \mathbb{R}^n$ 的坐标写成 $\mathcal{M}_{\mathcal{B}}(v) = (x_1, \cdots, x_n)^T$, 则 $q(v)$ 的值为 $x_1^2+\cdots+x_p^2-x_{p+1}^2-\cdots-x_r^2$. 而在另一组有序基 $\mathcal{C} = (u_1, \cdots, u_l, \cdots, u_n)$ 下, 若将 $v \in \mathbb{R}^n$ 的坐标写成 $\mathcal{M}_{\mathcal{C}}(v) = (y_1, \cdots, y_n)^T$, 则 $q(v)$ 的值为 $y_1^2 + \cdots + y_l^2 - y_{l+1}^2 - \cdots - y_r^2$.

现在考虑 $\mathbb{R}^n$ 的如下两个子空间: $U = \operatorname{span}(v_1, \cdots, v_p)$, $W = \operatorname{span}(u_{l+1}, \cdots, u_n)$. 我们希望证明 $U \cap W = 0$. 由此可得

$$p + (n - l) = \dim U + \dim W = \dim(U + W) \leq n\,,$$

即 $p \leq l$. 因为 $\mathcal{B}$ 和 $\mathcal{C}$ 两组有序基的地位是对等的, 所以将它们互换一下进行上面相同的论证可知 $l \leq p$. 所以 $p = l$. 这就完成了证明.

唯一剩下的工作就是验证上面定义的子空间 U 和 W 交集为 0. 为此, 假设 $v \in U \cap W$. 根据 U 和 W 的选取, v 在 $\mathcal{B}$ 和 $\mathcal{C}$ 两组有序基的坐标分别为

$$\mathcal{M}_{\mathcal{B}}(v) = (x_1, \cdots, x_p, 0, \cdots, 0)^T \quad \text{和} \quad \mathcal{M}_{\mathcal{C}}(v) = (0, \cdots, 0, y_{l+1}, \cdots, y_n)^T$$

的形式. 代入 q 的相应表达式可知

$$x_1^2 + \cdots + x_p^2 = q(v) = -y_{l+1}^2 - \cdots - y_r^2\,.$$

此等式左边的值非负, 而右边的值则小于或等于 0. 因此左右两边的值均为 0. 而在 x_i 均为实数时 $x_1^2 + \cdots + x_p^2 = 0$ 只有在 x_i 均为 0 时才能成立. 所以 v 在 $\mathcal{B}$ 下的坐标为 $(0, \cdots, 0)$, 即 $v = 0$. 这就证明了所需的结论. □

定理 6.2.16 只叙述了向量空间 $\mathbb{R}^n$ 的情形. 对于一般的 n 维实向量空间 V, 通过选定一组有序基 $\mathcal{B}$ 即可将 V 上的任意二次型 q 转化为 $\mathbb{R}^n$ 上的一个二次型 q' (参见图表 (6.2.9.1)). 而且 q' 的规范形不依赖于 $\mathcal{B}$ 的选取. 因此, 对于 V 上的二次型我们仍然可以定义正负惯性指数、符号差等概念.

定理 6.2.16 表明: 若想区分实数域上的两个二次型是否等价, 除了要考查维数和秩两个不变量之外, 还需要再比较正惯性指数 (或者负惯性指数, 又或者符号差) 是否相同. 或者说, 对于相同大小的两个对称实方阵, 它们在 $\mathbb{R}$ 上相合的充分必要条件是二者的秩和正惯性指数都相等 (或者秩和负惯性指数都相等, 又或者秩和符号差都相等).

为了避免读者误解, 关于定理 6.2.15 和 6.2.16 中的唯一性我们通过一个简单的例子来强调一下其含义. 考虑二元二次型 $q(x_1, x_2) = x_1^2 + 2x_1x_2 + 2x_2^2$. 容易看出

$$q = (x_1 + x_2)^2 + x_2^2 = \left(\frac{x_1}{\sqrt{2}}\right)^2 + \left(\frac{x_1}{\sqrt{2}} + \sqrt{2}x_2\right)^2.$$

所以, 至少有以下两种变量替换可以将 q 化为规范形, 即, $q = X_1^2 + X_2^2 = Y_1^2 + Y_2^2$, 其中

$$\begin{cases} X_1 = x_1 + x_2 \\ X_2 = x_2 \end{cases} \quad \text{而} \quad \begin{cases} Y_1 = \frac{x_1}{\sqrt{2}} \\ Y_2 = \frac{x_1}{\sqrt{2}} + \sqrt{2}x_2 \end{cases} .$$

这里的表达式 $X_1^2 + X_2^2$ 和 $Y_1^2 + Y_2^2$ 我们认为形式上是一样的, 在这种意义下我们才有定理 6.2.15 和 6.2.16 中唯一性的结论. 不过这个例子表明, 唯一性并不意味着化为规范形所执行的线性变量替换只有唯一的方式, 或者说在规范形中新的变量与原始变量之间的代换关系一般是不唯一的. 如果换成矩阵的方式解释, 这相当于说, 对于给定的对称矩阵 A, 和 A 相合的那个规范形矩阵是唯一确定的, 但是能够使得 P^TAP 是规范形矩阵的那个可逆矩阵 P 不是唯一的.

对于一般的域 K, 要区分 K 上两个二次型是否等价并不是件容易的事情. 如果 K 是代数数论意义下的**代数数域** (algebraic number field), 即 K 作为 $\mathbb{Q}$-向量空间是有限维的 (例如 $K = \mathbb{Q}(\sqrt{3})$), 那么代数数论会给我们提供一套行之有效的方法区分 K 上的二次型等价类. 有兴趣的读者将来可以在这方面继续深入学习.

6.2.4 二次型和对称矩阵的正定性

本小节我们讨论一种专门针对实二次型的性质.

定义 6.2.17. 设 V 是实向量空间, q 是 V 上的二次型.

如果对于 V 中的所有非零向量 v 均有 $q(v) > 0$ (或相应地, $q(v) \geq 0$), 则称 q 是**正定的** (positive definite) (或相应地, **半正定的** (positive semidefinite)).

如果 $-q$ 是正定的 (或相应地, 半正定的), 我们就说 q 是**负定的** (negative definite) (或相应地, **半负定的** (negative semidefinite)). 也就是说, 如果对于 V 中的所有非零向量 v 均有 $q(v) < 0$ (或相应地, $q(v) \leq 0$), 我们称 q 是**负定的** (negative definite) (或相应地, **半负定的** (negative semidefinite)).

如果 q 不满足正定、半正定、负定、半负定四个性质中的任何一个, 那么我们说 q 是**不定的** (indefinite).

若 φ 是 q 的极化型, 那么根据 q 是正定、半正定、负定、半负定或不定的情况, 我们也相应地称对称双线性型 φ 是**正定、半正定、负定、半负定或不定的**.

如果 A 是一个 n 阶对称实矩阵, 并且相应的二次型

$$\langle A \rangle : \mathbb{R}^n \longrightarrow \mathbb{R}; \quad x \longmapsto x^T A x$$

是正定、半正定、负定、半负定或不定的, 那么我们对应地将 A 称为**正定、半正定、负定、半负定或不定的**. ■

思考题 6.13. 设 A, B 是相同大小的实方阵. 如果 A, B 在 $\mathbb{R}$ 上相合, 那么 A 正定 (半正定) 当且仅当 B 正定 (半正定).

命题 6.2.18: 设 q 是 n 维实向量空间 V 上的二次型, $r = \mathrm{rank}(q)$, p 为 q 的正惯性指数. 则以下断言等价:

1. 二次型 q 是正定的.
2. $p = n$.
3. $p = r = n$.
4. 在 V 的某一组有序基 $\mathcal{B}$ 下 Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(q)$ 是正定矩阵.

5. 在 V 的任意一组有序基 $\mathcal{B}$ 下 Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(q)$ 总是正定矩阵.

6. 对于 V 的某一组有序基 $\mathcal{B}$, Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(q)$ 与单位矩阵 I_n 在 $\mathbb{R}$ 上相合, 即, 存在可逆矩阵 $P \in \mathbf{M}_n(\mathbb{R})$ 使得 $\mathcal{M}_{\mathcal{B}}(q) = P^T P$.

7. 对于 V 的任何一组有序基 $\mathcal{B}$, Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(q)$ 与单位矩阵 I_n 在 $\mathbb{R}$ 上相合.

证明. 对于 V 的任意有序基 $\mathcal{B}$, 二次型 q 在任何向量 $v \in V$ 处的取值可由公式

$$q(v) = \mathcal{M}_{\mathcal{B}}(v)^T \mathcal{M}_{\mathcal{B}}(q) \mathcal{M}_{\mathcal{B}}(v)$$

计算. 当 v 取遍 V 中向量时, 其坐标列 $\mathcal{M}_{\mathcal{B}}(v)$ 取遍 $\mathbb{R}^n$ 中元素, 反之亦然. 由此即可按定义得到 (1), (4), (5) 之间的等价性.

因为不等式 $p \leq r \leq n$ 总成立, 所以 (2) 和 (3) 等价. 正惯性指数等于 n 的意思就是 q 的 Gram 矩阵以单位阵 I_n 为一个相合标准形, 所以 (2) 和 (6) 等价. 与单位矩阵相合的矩阵一定是正定矩阵, 所以 (6) 蕴含 (4). 如果正惯性指数 p 小于 n, 那么 q 可以在某些非零向量处取值为 0 或负数, 这样 q 不可能正定. 所以 (1) 蕴含 (2). (6) 和 (7) 的等价性留给读者自行说明. 命题至此证毕. □

思考题 6.14. 设 q 是 n 维实向量空间 V 上的二次型, $r = \operatorname{rank}(q)$, p 为 q 的正惯性指数. 则以下断言等价:

1. 二次型 q 是半正定的.

2. $p = r$.

3. 在 V 的某一组有序基 $\mathcal{B}$ 下 Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(q)$ 是半正定矩阵.

4. 在 V 的任意一组有序基 $\mathcal{B}$ 下 Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(q)$ 总是半正定矩阵.

5. 对于 V 的某一组有序基 $\mathcal{B}$, Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(q)$ 与矩阵 $\begin{pmatrix} I_r & \\ & 0 \end{pmatrix}$ 在 $\mathbb{R}$ 上相合.

6. 对于 V 的任何一组有序基 $\mathcal{B}$, Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(q)$ 与矩阵 $\begin{pmatrix} I_r & \\ & 0 \end{pmatrix}$ 在 $\mathbb{R}$ 上相合.

7. 对于 V 的某一组有序基 $\mathcal{B}$, 存在矩阵 $M \in \mathbf{M}_n(\mathbb{R})$ 使得 $\mathcal{M}_{\mathcal{B}}(q) = M^T M$ 且 $\operatorname{rank}(M) = r$.

8. 对于 V 的任何一组有序基 $\mathcal{B}$, 总存在矩阵 $M \in \mathbf{M}_n(\mathbb{R})$ 使得 $\mathcal{M}_{\mathcal{B}}(q) = M^T M$ 且 $\operatorname{rank}(M) = r$.

9. 对于 V 的某一组有序基 $\mathcal{B}$, 存在矩阵 $M \in \mathbf{M}_n(\mathbb{R})$ 使得 $\mathcal{M}_{\mathcal{B}}(q) = M^T M$.

10. 对于 V 的任何一组有序基 $\mathcal{B}$, 总存在矩阵 $M \in \mathbf{M}_n(\mathbb{R})$ 使得 $\mathcal{M}_{\mathcal{B}}(q) = M^T M$.

在判别正定性时使用矩阵语言有一个很特殊的便利之处, 就是可以通过一些行列式的正负号来判别矩阵的正定性.

(6.2.19) 我们需要回忆几个概念 (参见本书上册 (4.3.4) 一段).

设 $A = (a_{ij}) \in \mathbf{M}_n(K)$, I 是 $[\![1, n]\!]$ 的一个非空子集, $|I| = r$. 将 I 中元数按从小到大顺序排列为 $i_1, \cdots, i_r$. 以下行列式

$$A\begin{Bmatrix} I \\ I \end{Bmatrix} = A\begin{Bmatrix} i_1\ i_2\ \cdots\ i_r \\ i_1\ i_2\ \cdots\ i_r \end{Bmatrix} := \begin{vmatrix} a_{i_1i_1} & a_{i_1i_2} & \cdots & a_{i_1i_r} \\ a_{i_2i_1} & a_{i_2i_2} & \cdots & a_{i_2i_r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_ri_1} & a_{i_ri_2} & \cdots & a_{i_ri_r} \end{vmatrix}.$$

被称为 A 的一个 r 阶主子式 (principal minor). 如果 $I = [\![1,\, r]\!]$, 则称 $A\begin{Bmatrix} I \\ I \end{Bmatrix}$ 为 A 的 r 阶顺序主子式或领衔主子式 (leading principal minor). ■

命题 6.2.20: 设 A 为 n 阶实对称阵. 以下论断等价:

1. A 正定.
2. 存在可逆矩阵 $P \in \mathbf{M}_n(\mathbb{R})$ 使 $A = P^T P$.
3. A 的所有顺序主子式大于 0.
4. A 的所有主子式大于 0.

证明. 考虑 $\mathbb{R}^n$ 上的二次型 $q := \langle A \rangle : x \mapsto x^T Ax$. 由命题 6.2.18 的断言 (4)–(7) 即可知道 (1) 和 (2) 等价.

现在证明 (1) 与 (3) 等价. 先假设 A 正定. 对于 $r \in [\![1,\, n]\!]$, 顺序主子式 $A\begin{Bmatrix} 1,2,\cdots,r \\ 1,2,\cdots,r \end{Bmatrix}$ 实际上是 A 的如下子矩阵 A_r 的行列式:

$$A_r = \begin{pmatrix} a_{11} & \cdots & a_{1r} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rr} \end{pmatrix}$$

(即 A_r 是位于 A 前 r 行和前 r 列交叉处的子矩阵.) 注意到 A 是二次型 $q = \langle A \rangle$ 在标准基 $e_1, \cdots, e_n$ 下的 Gram 矩阵, 而若令 $W := \operatorname{span}(e_1, \cdots, e_r)$ 则 A_r 是二次型 $q|_W$ 的 Gram 矩阵. 由 A 正定可知 q 正定, 从而 $q|_W$ 也正定. 根据命题 6.2.18 可知, A_r 作为 $q|_W$ 的 Gram 矩阵必然可以写成 $Q^T Q$ 的形式, 其中 Q 是 r 阶可逆实矩阵. 于是 $|A_r| = |Q|^2 > 0$. 这样我们证明了 (1) $\Rightarrow$ (3).

再反过来假设条件 (3) 成立, 我们来证明 A 是正定阵. 我们采用对 n 的归纳法. 当 $n = 1$ 时结论显然成立. 以下我们假设: 如果一个 $n-1$ 阶实对称阵的顺序主子式都大于零, 那么该矩阵正定. 特别地, 位于 A 的左上角的 $n-1$ 阶子矩阵 A_{n-1} 是正定矩阵.

令 $U = \operatorname{span}(e_1, \cdots, e_{n-1})$, 则 A_{n-1} 是 U 上二次型 $q|_U$ 在有序基 $e_1, \cdots, e_{n-1}$ 下的 Gram 矩阵. 将命题 6.2.18 应用于二次型 $q|_U$ 可知, 存在 U 的一组有序基 $\eta_1, \cdots, \eta_{n-1}$ 使得 $(\varphi(\eta_i,\, \eta_j)) = I_{n-1}$, 这里 $\varphi = b_q$ 是 q 的对应的 (极化) 二次型. 因为 A_{n-1} 是 $\varphi|_U$ 在 U 的一组基下的矩阵, 而且作为 A 的 $n-1$ 阶顺序主子式, 我们的假设条件包含 $|A_{n-1}| > 0$. 所以 $\varphi|_U$ 是非退化的. 根据命题 6.1.25, 我们有正交分解 $\mathbb{R}^n = U \perp_\varphi U^\perp$ 成立. 因为 $\dim U = n-1$, 所以 $U^\perp$ 是 1 维的. 设 w 是 $U^\perp$ 的一组基, 则 $\eta_1, \cdots, \eta_{n-1}, w$ 构成 φ 在 V 中的一组正交基. 根据前面得到的事实 $(\varphi(\eta_i,\, \eta_j)) = I_{n-1}$ 可知, φ 在 $\eta_1, \cdots, \eta_{n-1}, w$ 这组基下的矩阵形如 $A' = \begin{pmatrix} I_{n-1} & \\ & d \end{pmatrix}$, 其中 $d = \varphi(w,\, w)$. 这说明矩阵 A 相合于 A'. 作为 A 的 n 阶顺序主子式, 行列式 $|A|$ 是正值. 所以 $d = |A'|$ 也是正数. 由此容易看出 A' 与 I_n 相合, 从而 A 与 I_n 相合. 因此 A 是正定矩阵. 所以我们证明了 (1) 和 (3) 等价.

条件 (4) 显然蕴含 (3). 而 (1) 推出 (4) 的证明可以采用上面 (1) 推出 (3) 的同样方式加以论证. 或者, 也可以用以下两个事实来论证 (1) $\Rightarrow$ (4). 第一是和一个正定矩阵相合的矩阵仍然正定 (思考题 6.13). 第二, 对于任意 n 阶方阵 A 和任意两个不同的指标 $i, j \in [\![1,\, n]\!]$, 如果将 A 的第 i, j 行对调、第 i, j 列也对调得到的矩阵是 B, 那么 B 与 A 相合. 事实上, 如果 P 是将单位阵 I_n 的第 i 行和第 j 行对换得到的初等矩阵, 那么 $P^T AP = B$. 反复应用第二个事实可知, 对于一般的主子式 $A\begin{Bmatrix} i_1,\cdots,i_r \\ i_1,\cdots,i_r \end{Bmatrix}$, 经过适当的相合变换后可以将 A 化为一个新的矩阵 B, 使得 $A\begin{Bmatrix} i_1,\cdots,i_r \\ i_1,\cdots,i_r \end{Bmatrix}$ 成为 B 的 r 阶顺序主子式 $B\begin{Bmatrix} 1,\cdots,r \\ 1,\cdots,r \end{Bmatrix}$. 根据上面说的第一个性质, A 的正定性表明 B 也正定, 故

$$A\begin{Bmatrix} i_1,\cdots,i_r \\ i_1,\cdots,i_r \end{Bmatrix} = B\begin{Bmatrix} 1,\cdots,r \\ 1,\cdots,r \end{Bmatrix} > 0\,.$$

(这里的不等号是将 (1) 和 (3) 的等价性应用于正定矩阵 B 得到的.) □

命题 6.2.20 中的条件 (4) 实际上对正定矩阵的样貌有很强的限制. 例如, 正定矩阵的对角线上元素必须都是正数 (因为它们都是 1 阶主子式).

思考题 6.15. 设 $A=(a_{ij})\in\mathbf{M}_n(\mathbb{R})$ 是正定矩阵. 令 $M=\max_{1\le i,\,j\le n}\{|a_{ij}|\}$ 为矩阵 A 中元素绝对值的最大值.

证明: 对任意两个不同的指标 $k,\,l\in[\![1,\,n]\!]$, 必有 $|a_{kl}|<M$. (也就是说, 只有 A 的对角线上元素才可能达到绝对值的最大值.)

思考题 6.16. 能否通过顺序主子式的正负号来判别对称矩阵是否负定？如何判别？

举个例子来说明如何使用命题 6.2.20 来对矩阵的正定性做判定.

例 6.2.21. 试问实数 t 取什么值时, 下面的二次型是正定的:

$$q(x_1,\,x_2,\,x_3,\,x_4)=x_1^2+4x_2^2+4x_3^2+3x_4^2+2tx_1x_2-2x_1x_3+4x_2x_3\,.$$

首先, q (在标准基下) 的 Gram 矩阵为 $A=\begin{pmatrix}1&t&-1&0\\t&4&2&0\\-1&2&4&0\\0&0&0&3\end{pmatrix}$. 依次计算 A 的四个顺序主子式可得

$$|A_1|=1>0\,,\quad |A_2|=\begin{vmatrix}1&t\\t&4\end{vmatrix}=4-t^2\,,$$

$$|A_3|=\begin{vmatrix}1&t&-1\\t&4&2\\-1&2&4\end{vmatrix}=-4(t-1)(t+2)\,,\quad |A_4|=3|A_3|=-12(t-1)(t+2)\,.$$

根据命题 6.2.20 中的条件 (3), q 正定的充分必要条件是以下两个不等式成立:

$$4-t^2>0\,,\ -4(t-1)(t+2)>0\,.$$

因此, t 的取值范围是 $-2<t<1$. ∎

还有一点需要提醒读者注意, 我们定义正定矩阵时首先要求该矩阵是对称的. 如果不是对称矩阵, 通过顺序主子式大于零不能得到类似正定的性质, 如下面的思考题所示.

思考题 6.17. 举例说明: 存在一个上三角矩阵 $A\in\mathbf{M}_2(\mathbb{R})$, 它的非零元素都是正数 (因此 A 的顺序主子式都是正的), 但二次型 $q=\langle A\rangle:\mathbb{R}^2\to\mathbb{R},\ \alpha\mapsto\alpha^TA\alpha$ 可以在某些 $\alpha\in\mathbb{R}^2$ 处取到负值.

读者可能会根据命题 6.2.20 中的条件 (3) 来猜测: 如果一个对称矩阵的顺序主子式都是非负的, 那么该矩阵是半正定的. 但这个结论是不对的. 例如, 对角矩阵 $A=\begin{pmatrix}0&&\\&1&\\&&-1\end{pmatrix}$ 的顺序主子式都大于等于 0, 但显然 A 不是半正定的.

不过, 命题 6.2.20 中的条件 (4) 有针对半正定矩阵的版本: 如果一个实对称矩阵的所有主子式都非负, 那么它一定是半正定的 (参见下面的注记 6.2.23).

关于矩阵半正定性的判别, 我们可以使用下面的命题.

命题 6.2.22: 设 A 是 n 阶实对称阵. 则下列断言等价:

1. A 半正定.
2. 存在矩阵 $M \in \mathbf{M}_n(\mathbb{R})$ 使 $A = M^T M$.
3. 对于任意实数 $t > 0$, 矩阵 $A + tI_n$ 正定.

证明. (1) 和 (2) 的等价性留给读者.

如果 A 半正定, 那么对于任意 $\alpha \in \mathbb{R}^n$, 均有 $\alpha^T A\alpha \geq 0$. 于是, 对于任何正实数 t,

$$\alpha^T(A + tI_n)\alpha = \alpha^T A\alpha + t\alpha^T\alpha \geq t\alpha^T\alpha\,.$$

当 $\alpha \neq 0$ 时, $\alpha^T\alpha > 0$. 所以上式最右端严格大于 0. 按定义, 这说明 $A + tI_n$ 正定.

反过来, 假设对所有正实数 t, $A + tI_n$ 都是正定矩阵. 那么对于任意确定的 $\alpha \in \mathbb{R}^n$, 可以考虑一元函数

$$f_\alpha \,:\, \mathbb{R} \longrightarrow \mathbb{R}\,; \quad t \longmapsto \alpha^T(A + tI_n)\alpha = \alpha^T A\alpha + t\alpha^T\alpha\,.$$

前面的假设说明 $f_\alpha(t) > 0$ 在 $t > 0$ 时恒成立. 令 t 从大于 0 的一侧趋于 0 来求极限 $\lim_{t\to 0+} f_\alpha(t)$, 可以得到 $\alpha^T A\alpha = \lim_{t\to 0+} f_\alpha(t) \geq 0$. 因为这对任何 $\alpha \in \mathbb{R}^n$ 都成立, 因此按照定义, A 是半正定的. 至此我们证明了 (1) 和 (3) 等价. □

注记 6.2.23. 设 A 是 n 阶实对称阵. 可以证明: A 半正定的另一个充分必要条件是 A 的所有主子式非负[†].

事实上, 如果 $A\left\{ {I \atop I} \right\} = A\left\{ {i_1\cdots i_r \atop i_1\cdots i_r} \right\}$ 是 A 的一个主子式, 我们可以考虑二次型 $q = \langle\, A\,\rangle \in \mathrm{Quad}(\mathbb{R}^n)$ 在子空间 $W = \mathrm{span}(e_{i_1}, \cdots, e_{i_r})$ 上的限制 $q|_W$. 容易看出, $q|_W$ 在 $e_{i_1}, \cdots, e_{i_r}$ 这组基下的矩阵行列式就等于 $A\left\{ {I \atop I} \right\}$. 因为 q 的半正定性质蕴含 $q|_W$ 的半正定性质. 而根据命题 6.2.22 中 (1) 和 (2) 的等价性容易看出, 半正定矩阵的行列式非负. 因此我们得到 $A\left\{ {I \atop I} \right\} \geq 0$. 这就完成了必要性的证明.

充分性的证明需要用到如下结论: 如果将含参数 t 的行列式 $|A + tI_n|$ 当作关于 t 的多项式, 那么

(6.2.23.1) $|A + tI_n| = t^n + c_1 t^{n-1} + \cdots + c_{n-1}t + c_n\,,$ 其中每个 c_i 等于 A 所有 i 阶主子式之和.

根据上面这个公式, 如果 A 的所有主子式非负, 那么行列式 $|A + tI_n|$ 在 $t > 0$ 时一定是正的. 如果用 $A_1, \cdots, A_n$ 表示位于 A 左上角的 n 个子矩阵, 其中 A_r 是 r 阶方阵, 那么 A_r 的所有主子式也都是 A 的主子式, 因此都是非负的. 如果在上面的讨论中将 A 替换为 A_r, 那么可以推知: 当 $t > 0$ 时, $|A_r + tI_r| > 0$ 对每个 $r \in [\![1,\, n]\!]$ 成立. 这说明当 $t > 0$ 时, 矩阵 $A + tI_n$ 的所有顺序主子式是正的. 根据命题 6.2.20 即知 $A + tI_n$ 正定. 再结合命题 6.2.22 即知矩阵 A 半正定.

公式 (6.2.23.1) 的证明不是此处论题的重点, 因此我们将此证明略过. 有兴趣的读者可以作为练习自我挑战一下. (提示: 利用行列式关于列的多重线性, 可以将行列式 $|A + tI_n|$ 写成 2^n 个行列式的和. 对于这 2^n 个行列式中的每一个, 其每一列要么来自于 A 要么来自于 tI_n. 如果该行列式中来自于 A 的列恰好是第 $i_1, \cdots, i_r$ 列, 则该行列式的值是 $A\left\{ {i_1\cdots i_r \atop i_1\cdots i_r} \right\} t^{n-r}$.)

事实上, 通过检查所有主子式的符号来判别矩阵是否半正定并不是高效实惠的算法, 因为它所涉及的行列式计算量通常很大: 一个 n 阶矩阵的主子式共有 $2^n - 1$ 个. 与其这样似乎还不如引入一个参数 t, 计算 $A + tI_n$ 的 n 个顺序主子式得到 n 个关于 t 的多项式, 然后考查一下这些多项式是否在正实轴上恒取正值 (即检验 $A + tI_n$ 的正定性). ■

注记 6.2.24. 今后我们会看到, 实对称矩阵都可以和某个对角阵既相合又相似 (参见定理 7.2.23). 因此, 通过特征值的正负号来判定对称矩阵是否正定或半正定是另一个简明而有用的判别法. ■

[†]实际上另一个等价条件是: 对于每个 $k \in [\![1,\, n]\!]$, A 的所有 k 阶主子式之和非负.

二次型的半正定性质和多元函数的极值问题密切相关. 不过, 这个问题一般是在多元微积分的课程中重点讨论的. 我们这里仅通过二元二次函数的特例来简要解释其中的主要思想.

(6.2.25) 考虑 $\mathbb{R}^2$ 上定义的一个二元二次函数

$$Q\ :\ \mathbb{R}^2\longrightarrow\mathbb{R}\,;\quad (x,\,y)\longmapsto Q(x,\,y)=ax^2+2bxy+cy^2+2dx+2ey+f$$

其中 $a,\,b,\,c,\,d,\,e,\,f\in\mathbb{R}$ 是常数.

对于平面内取定的一点 $(x_0,\,y_0)$, 我们想判定这个点是否是函数 Q 的极小值点. 根据定义, $(x_0,\,y_0)$ 是 Q 的极小值点意味着存在 $\varepsilon>0$ 使得对开区间 $]-\varepsilon,\,\varepsilon[$ 内的任意实数 $h,\,k$, 均有

$$Q(x_0+h,\,y_0+k)-Q(x_0,\,y_0)\geq 0\,.$$

通过以上 Q 的表达式计算可得

$$Q(x_0+h,\,y_0+k)-Q(x_0,\,y_0)=(ah^2+2bhk+ck^2)+(2ax_0+2by_0+2d)h+(2bx_0+2cy_0+2e)k\,.$$

如果这里的系数 $D:=2ax_0+2by_0+2d$ 不等于 0, 例如, $D>0$, 那么令 $k=0$, 而 h 为绝对值充分小的负数时,

$$Q(x_0+h,\,y_0)-Q(x_0,\,y_0)=ah^2+Dh=h(ah+D)<0\,.$$

因此, 这种情况下 $(x_0,\,y_0)$ 不可能是 Q 的极小值点. 同理, 如果 $2bx_0+2cy_0+2e\neq 0$, 则 $(x_0,\,y_0)$ 也不可能是极小值点. 因此, 不妨设 $2ax_0+2by_0+2d=2bx_0+2cy_0+2e=0$. (这相当于说函数 $Q(x,\,y)$ 的两个偏导数 $\frac{\partial Q}{\partial x}$ 和 $\frac{\partial Q}{\partial y}$ 在 $(x_0,\,y_0)$ 处的取值均为 0.) 于是

$$Q(x_0+h,\,y_0+k)-Q(x_0,\,y_0)=ah^2+2bhk+ck^2\,.$$

如果以 $h,\,k$ 为变量, 那么上式右边就是一个二元二次型. 我们希望它的取值总是非负的, 这就等价于要求该二次型是半正定的, 或者说矩阵 $A=\begin{pmatrix}a&b\\b&c\end{pmatrix}$ 是半正定的.

对于一般的多元函数 $f(x_1,\cdots,x_n)$, 如果假定 f 具有良好的可微性质, 那么可以通过多元 Taylor 展开式的二次项部分得到一个 n 元二次型 q. 函数 f 的极小值点判定会归结为二次型 q 是否半正定的问题. 关于这个问题的详细讨论读者可以在任何标准的多元微积分教材中找到. ■

6.2.5 习题

习题 6.2.1. 设 q 为 K-向量空间 V 上的二次型, $\varphi=b_q$ 是它的极化型.

证明以下恒等式 (通常称为**极化恒等式** (polarization identity)):

$$\text{对任意 } u,\,v\in V\,,\quad \varphi(u,\,v)=\frac{q(u+v)-q(u-v)}{4}\,.$$

习题 6.2.2. 对于下列二次型 $q\in\mathrm{Quad}(K^4)$, 求出 q 在 K^4 的标准基下的 Gram 矩阵以及极化型 b_q 的表达式:

(1) $q=-x_1x_3-2x_1x_4+x_3^2-5x_3x_4$;

(2) $q=-x_2^2-x_3^2-x_1x_4$;

(3) $q=x_1x_2+x_1x_3+x_1x_4+x_2x_3+x_2x_4+x_3x_4$.

习题 6.2.3. 找出适当的可逆线性变量替换将下列二次型化为标准形:

(1) $q=-4x_1x_2+2x_1x_3+2x_2x_3$;

(2) $q=x_1^2+2x_1x_2+2x_2^2+4x_2x_3+4x_3^2$;

(3) $q=x_1^2-3x_2^2-2x_1x_2+2x_1x_3-6x_2x_3$;

(4) $q=8x_1x_4+2x_3x_4+2x_2x_3+8x_2x_4$.

习题 6.2.4. 找出适当的可逆线性变量替换将下列二次型化为标准形:

$$(1)\quad q = x_1x_{2n} + x_2x_{2n-1} + \cdots + x_nx_{n+1}\ ;$$
$$(2)\quad q = x_1x_2 + x_2x_3 + \cdots + x_{n-1}x_n\ .$$

习题 6.2.5. 找出适当的可逆线性变量替换将下列二次型化为标准形:

$$(1)\quad q = \sum_{i=1}^{n} x_i^2 + \sum_{1\le i<j\le n}^{n} x_ix_j\ ;$$
$$(2)\quad q = \sum_{i=1}^{n} (x_i - \bar{x})^2\ ,\qquad \text{其中 } \bar{x} := \frac{x_1 + x_2 + \cdots + x_n}{n}\ .$$

习题 6.2.6. 证明: 秩等于 r 的对称矩阵可以写成 r 个秩为 1 的对称矩阵之和.

习题 6.2.7. 设 $A = (a_{ij})$ 为 $s \times n$ 实矩阵. 考虑实二次型

$$q = \sum_{i=1}^{s} (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n)^2\ .$$

证明 $\operatorname{rank}(q) = \operatorname{rank}(A)$.

习题 6.2.8. 设 $A, P \in \mathbf{M}_n(K)$. 假设 P 是对角线元素全为 1 的上三角阵, $B = P^TAP$. 证明: A 和 B 的各阶顺序主子式都相等, 即,

$$\text{对每个 } r \in [\![1,\, n]\!],\quad A\left\{\begin{smallmatrix}1\ 2\ \cdots\ r\\ 1\ 2\ \cdots\ r\end{smallmatrix}\right\} = B\left\{\begin{smallmatrix}1\ 2\ \cdots\ r\\ 1\ 2\ \cdots\ r\end{smallmatrix}\right\}\ .$$

习题 6.2.9. 设 $A \in \mathbf{M}_n(K)$ 为对称矩阵. 对每个 $k \in [\![1,\, n]\!]$, 记 $d_k = A\left\{\begin{smallmatrix}1\ 2\ \cdots\ k\\ 1\ 2\ \cdots\ k\end{smallmatrix}\right\}$ 为 A 的 k 阶顺序主子式.

1. 证明下列陈述等价:

   (a) 存在对角线元素全为 1 的上三角阵 P 使得

   $$P^TAP = \begin{pmatrix}\lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ & & & 0\end{pmatrix},\qquad \text{其中每个 } \lambda_i \neq 0\,.$$

   (b) $\operatorname{rank}(A) = r$, $d_1, \cdots, d_r$ 均不为 0, 而 $d_{r+1} = \cdots = d_n = 0$.

2. 假设上一小题中的等价条件成立. 证明其中出现的 λ_i 满足: $\lambda_1 = d_1$, $\lambda_2 = d_2/d_1\,, \cdots, \lambda_r = d_r/d_{r-1}$.

习题 6.2.10. 找出 $\mathbb{C}$ 上适当的可逆线性变量替换将下列二次型化为复规范形:

$$(1)\quad q = 2x_1x_2 - 6x_2x_3 + 2x_1x_3\,,$$
$$(2)\quad q = (-1-i)x_1x_2 + 2ix_2^2\,,$$
$$(3)\quad q = (1+i)x_1^2 - (\sqrt{2}+2i)x_2^2 - 3ix_3^2\,.$$

习题 6.2.11. 找出 $\mathbb{R}$ 上适当的可逆线性变量替换将下列二次型化为实规范形, 并求出每个二次型的秩和正、负惯性指数:

$$\begin{aligned}
&(1) \quad q = 2x_1x_2 - 6x_2x_3 + 2x_1x_3\,,\\
&(2) \quad q = -5x_1^2 + 3x_2^2 - 2x_4^2\,,\\
&(3) \quad q = -(x_1 - 2x_2 + 3x_3 - x_4)^2 + (2x_1 - x_3 + 3x_4)^2 + (-2x_1 + 4x_2 - 6x_3 + 2x_4)^2\,.
\end{aligned}$$

习题 6.2.12. 系数在 K 中的一个有限维二次型称为在 K 上是**可约的**是指作为多项式它可以写成两个系数取自 K 的非零一次齐次多项式的乘积.

证明: 一个有限维实二次型 q 在 $\mathbb{R}$ 上是可约的当且仅当它满足以下两个条件之一:

(1) q 的秩等于 1;

(2) q 的秩等于 2, 符号差为 0.

习题 6.2.13. 设 q 为有限维实向量空间 V 上的二次型. 假设存在向量 $v_1, v_2 \in V$ 使得 $q(v_1) < 0 < q(v_2)$. 证明:

1. 一定存在非零向量 v_0 使得 $q(v_0) = 0$.

   这样的向量 v_0 称为 q 的一个**迷向向量**. (参见思考题 6.8.)

2. 存在 V 的一组基 $\varepsilon_1, \cdots, \varepsilon_n$ 其中每个 ε_i 均为 q 的迷向向量.

习题 6.2.14. 设 $1 \le p \le r$ 为正整数, 对每个 $i \in [\![1, r]\!]$, 假设 L_i 是一个实系数的一次齐次多项式. 令 $q = L_1^2 + \cdots + L_p^2 - L_{p+1}^2 - \cdots - L_r^2$.

证明: 实二次型 q 的正惯性指数 $\le p$, 负惯性指数 $\le r - p$.

习题 6.2.15. 设 q 是 n 维实向量空间 V 上的二次型. 定义 $C(q) := \{v \in V \mid q(v) = 0\}$.

1. 证明: $C(q)$ 是 V 的子空间当且仅当 q 是半正定或半负定的.

2. 假设 $C(q)$ 是个子空间. 求它的维数.

3. 对于一般的情况, 设 q 的秩为 r, 正、负惯性指数分别为 p, s.

   证明 $C(q)$ 中能够包含的子空间维数最大值是 $n - \max\{p, s\} = \min\{p, s\} + n - r$.

习题 6.2.16. 求能使以下三元二次型正定的实数 t 的取值范围:

$$q = x_1^2 + x_2^2 + 5x_3^2 + 2tx_1x_2 - 2x_1x_3 + 4x_2x_3\,.$$

习题 6.2.17. 对任意实数 t, 讨论以下三元二次型是否半正定或半负定:

$$q = x_1^2 + 4x_2^2 + x_3^2 + 2tx_1x_2 + 10x_1x_3 + 6x_2x_3\,.$$

习题 6.2.18. 设 $A \in \mathbf{M}_n(\mathbb{R})$ 是对称矩阵. 证明: 对于任何充分大的实数 t, 矩阵 $tI_n + A$ 是正定矩阵.

习题 6.2.19. 判断以下 n 元实二次型是否正定

$$q = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n-1} x_i x_{i+1}\,.$$

习题 6.2.20. 证明: 正定矩阵的逆矩阵也正定, 两个正定矩阵的和也正定.

习题 6.2.21. 设 $A \in \mathbf{M}_n(\mathbb{R})$ 为对称矩阵. 证明: 存在正实数 c 使得

$$\text{对任何} \quad x \in \mathbb{R}^n \quad \text{均有} \ |x^T A x| \le c x^T x \ .$$

习题 6.2.22. 假设 q_1, q_2 是两个 n 元实二次型.

1. 证明: 如果 q_1, q_2 的正惯性指数都小于 $n/2$, 则 $q_1 + q_2$ 不可能是正定的.
2. 举例说明: 即使 q_1, q_2 都不是正定的, $q_1 + q_2$ 仍有可能是正定的.

习题 6.2.23. 假设 $A = (a_{ij})$ 为 n 阶正定实矩阵. 定义

$$q : \mathbb{R}^n \longrightarrow \mathbb{R} \ ; \quad y = (y_1, \cdots, y_n)^T \longmapsto \det \begin{pmatrix} A & y \\ y^T & 0 \end{pmatrix} .$$

1. 证明: q 是 $\mathbb{R}^n$ 上的二次型, 并且 q 是负定的.
2. 设 P_{n-1} 是 A 的 $n-1$ 阶顺序主子式. 证明: $|A| \le a_{nn} P_{n-1}$.
3. 证明: $|A| \le a_{11} a_{22} \cdots a_{nn}$.
4. 证明: 对于任意可逆矩阵 $T = (t_{ij}) \in \mathbf{M}_n(\mathbb{R})$ 均有

$$|T|^2 \le \prod_{i=1}^{n} (t_{1i}^2 + t_{2i}^2 + \cdots + t_{ni}^2) \ .$$

6.3 二次曲线和二次曲面的仿射分类

这一节我们讨论二次型理论在空间坐标几何中的应用.

6.3.1 仿射空间与仿射变换

现在似乎是时候为读者比较精确地引入一些几何概念了.

(6.3.1) 设 $n \in \mathbb{N}^*$. 所谓 n 维 **(实) 仿射空间** (real affine space) 是指集合 $\mathbb{R}^n$. 为了便于讨论几何, 我们可能随时需要考虑 $\mathbb{R}^n$ 除了线性空间结构之外的其他结构, 或者说我们需要综合考虑 $\mathbb{R}^n$ 这个集合上所附带的一切有用信息. 因此, 当我们把集合 $\mathbb{R}^n$ 当作仿射空间来看待时, 我们改用记号 $\mathbb{A}^n(\mathbb{R})$ 或 $\mathbb{A}^n$ 来表示它. (更一般地, 我们也可以将任意域 K 上的 n 维仿射空间定义为集合 K^n, 并以记号 $\mathbb{A}^n(K)$ 记之.) 通常将 $\mathbb{A}^n$ 中的元素称为**点** (point), 以坐标行 $(x_1, \cdots, x_n)$ 或坐标列 $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ 的方式表示. 将 $\mathbb{A}^n$ 中的子集称为**图形** (figure).

最常见且较为容易研究的图形是所谓的仿射超曲面. 这个定义用到有关多元多项式的一些基本概念.

一般地, 对于一个域 K 以及一些形式符号 $X_1, \cdots, X_n$, 一个**系数取自** K、**以** $X_1, \cdots, X_n$ **为未定元**的**单项式** (monomial in the indeterminates $X_1, \cdots, X_n$ with coefficients in K) 是指一个形如 $cX_1^{d_1} \cdots X_n^{d_n}$ 的形式表达式, 其中 $c \in K$, $d_1, \cdots, d_n \in \mathbb{N}$. (如果 $d_1 = \cdots = d_n = 0$, 也将这个表达式等同于常数 c.) 将常数 c 称为该单项式的**系数** (coefficient). 如果 $c \neq 0$, 则将自然数 $d_1 + \cdots + d_n$ 称为该单项式的**次数** (degree). 如果 $c = 0$, 则将此单项式的次数规定为 $-\infty$.

一个系数取自 K、以 $X_1, \cdots, X_n$ 为未定元的**多项式形式** (formal polynomial) 或简称**多项式** (polynomial) 是指由有限多个单项式相加得到的形式表达式. 所有这些多项式构成的集合记为

$K[X_1,\cdots,X_n]$. 一个多项式 $f\in K[X_1,\cdots,X_n]$ 的**次数**定义为其中所含单项式的次数最大值. 如果 f 中系数非零的单项式次数都等于 d, 则称 f 是一个 d 次**齐次多项式** (homogeneous polynomial).

如果 $f\in\mathbb{R}[X_1,\cdots,X_n]$, 那么以下集合

$$V(f):=\{(a_1,\cdots,a_n)\in\mathbb{A}^n(\mathbb{R})\,|\,f(a_1,\cdots,a_n)=0\}$$

称为 f 在仿射空间 $\mathbb{A}^n(\mathbb{R})$ 中的**零点集** (zero locus). 如果 f 是非零多项式, 则将图形 $V(f)$ 称为仿射空间 $\mathbb{A}^n$ 中的一个**仿射超曲面** (affine hypersurface) 或简称**超曲面** (hypersurface).

如果一个图形 F 可以写成 $F=V(f)$ 的形式, 其中 f 是一个二次多项式, 我们就说 F 是一个**二次超曲面** (quadratic hypersurface, quadric). 二维仿射空间 $\mathbb{A}^2$ 也称为**仿射平面** (affine plane), 其中的二次超曲面称为**二次曲线** (quadratic curve). 三维仿射空间 $\mathbb{A}^3$ 中的二次超曲面则称为**二次曲面** (quadratic surface). ■

注记 6.3.2. 关于二次超曲面的定义有一些注意事项.

第一, "超曲面"这个词仅仅是为了术语方便采用的一个称号, 它并不意味着几何上对应的图形确实具有理所应当的维数. 或者说, 我们的定义允许"超曲面"是退化的. 例如, 三元二次多项式 $X_1^2+X_2^2+X_3^2$ 在 $\mathbb{A}^3$ 中的零点集虽然也被称为超曲面, 但实际上该集合只含一个点.

第二, 我们说一个图形 F 是个二次超曲面时只要求存在一个二次多项式 f 使它的零点集等于 F. 这并不意味着其它次数的多项式零点集不会等于 F. (例如 $X_1^2+X_2^2$ 和 $X_1^4+X_2^4$ 在仿射平面内的零点集都是只含 (0, 0) 这一个点.) 所以, 虽然我们有"二次超曲面"这样的称呼, 但这并不意味着这样的几何图形 F 有一个确定的 *次数* 概念. (关于"次数"的概念, 精确合理的定义可能需要在后续的代数几何课程中才能给出.) ■

定义 6.3.3. 仿射空间 $\mathbb{A}^n$ 上的一个**仿射变换** (affine transformation) 是指如下形式的一个映射:

$$\alpha\ :\ \mathbb{A}^n\longrightarrow\mathbb{A}^n\ ;\quad \begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}\longmapsto A\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}+b$$

其中 $A\in\mathbf{M}_n(\mathbb{R})$ 是固定的矩阵, $b\in\mathbb{R}^n$ 是固定的列向量. 如果其中的 b 等于 0, 则称 α 是一个**线性变换**. 如果 $A=I_n$, 则称 α 是个**平移变换** (translation).

两个图形 $F_1,F_2\subseteq\mathbb{A}^n$ 称为**仿射等价的** (affine equivalent), 是指存在可逆的仿射变换 $\alpha:\mathbb{A}^n\to\mathbb{A}^n$ 使 $\alpha(F_1)=F_2$. ■

思考题 6.18. 以下只考虑 $\mathbb{A}^n$ 上的仿射变换和 $\mathbb{A}^n$ 中的图形. 证明:

1. 假设仿射变换 $\alpha:X\mapsto AX+b$ 由矩阵 A 和向量 b 给出. 则下列陈述等价:

   (a) α 是可逆映射.

   (b) 矩阵 A 是可逆矩阵.

   (c) α 是单射.

   (d) α 是满射.

2. 恒等变换是一个仿射变换. 两个仿射变换的复合仍是仿射变换. 如果一个仿射变换是可逆映射, 那么其逆映射也是仿射变换.

3. 对于同一个仿射空间内的图形而言, 仿射等价是一个等价关系.

4. 对于任意取定的仿射变换 α, 存在唯一一对仿射变换 $\beta:\mathbb{A}^n\to\mathbb{A}^n$ 和 $\theta:\mathbb{A}^n\to\mathbb{A}^n$ 满足: $\alpha=\beta\circ\theta$, 且 β 是平移变换, θ 是线性变换;

   如果 α 是可逆的, 则又存在唯一一对仿射变换 $\beta':\mathbb{A}^n\to\mathbb{A}^n$ 和 $\theta':\mathbb{A}^n\to\mathbb{A}^n$ 满足: $\alpha=\theta'\circ\beta'$, 且 β' 是平移变换, θ' 是线性变换.

例 6.3.4. 设 U_1, U_2 是 $\mathbb{R}^n$ 中的两个子空间, v_1, v_2 是 $\mathbb{R}^n$ 中两个向量. (通过将集合 $\mathbb{R}^n$ 等同于 $\mathbb{A}^n$) 我们可以把 (线性) 仿射集 $F_i:=v_i+U_i$ 视为 $\mathbb{A}^n$ 中的图形. 则 F_1 与 F_2 仿射等价的充分必要条件是 $\dim U_1=\dim U_2$ (也就是 $\dim F_1=\dim F_2$, 因为我们曾将 (线性) 仿射集的维数定义为相应的向量空间维数).

事实上, 如果存在可逆的仿射变换 $\alpha: X\mapsto AX+b$ 使得 $\alpha(F_1)=F_2$, 则有

$$\mathscr{A}v_1+\mathscr{A}(U_1)+b=v_2+U_2$$

其中 $\mathscr{A}\in\mathrm{End}(\mathbb{R}^n)$ 表示 $\mathbb{R}^n$ 上的线性变换 $X\mapsto AX$. 按照本书上册 § 2.2.3 节的术语, 上式左边是刚性平行于子空间 $\mathscr{A}(U_1)$ 的一个仿射集, 而右边是刚性平行于子空间 U_2 的一个仿射集. 因此, 根据本书上册的命题 2.2.24 (3), 必然 $\mathscr{A}(U_1)=U_2$. 另一方面, 由思考题 6.18 (1) 可知, $\mathscr{A}$ 是可逆的线性变换. 所以 $\dim U_1=\dim\mathscr{A}(U_1)=\dim U_2$.

反过来, 假设 $\dim U_1=\dim U_2$, 那么一定存在可逆的线性变换 $\mathscr{A}\in\mathrm{End}(\mathbb{R}^n)$ 使得 $\mathscr{A}(U_1)=U_2$ (请读者解释为什么). 令 $b=v_2-\mathscr{A}v_1$, α 为 $X\mapsto\mathscr{A}(X)+b$ 给出的仿射变换, 则 $\alpha(F_1)=F_2$. ■

例 6.3.5. 设 F_1 为仿射平面 $\mathbb{A}^2$ 内的单位圆, 即二元多项式 x^2+y^2-1 的零点集. 设 F_2 是方程 $\frac{(x-1)^2}{4}+\frac{(y-3)^2}{9}=1$ 给出的椭圆. 则仿射变换

$$\alpha\ :\ \mathbb{A}^2\longrightarrow\mathbb{A}^2\ ;\quad (x,\,y)\longmapsto(2x+1,\,3y+3)=(2x,\,3y)+(1,\,3)$$

将 F_1 映射为 F_2. 事实上, 若令

$$(x',\,y')=\alpha(x,\,y)=(2x+1,\,3y+3)$$

则 $x=\frac{x'-1}{2}$, $y=\frac{y'-3}{3}$. 所以, 当 $(x,\,y)$ 满足 $x^2+y^2-1=0$ 时,

$$\frac{(x'-1)^2}{4}+\frac{(y'-3)^2}{9}=1$$

即 $(x',\,y')=\alpha(x,\,y)\in F_2$. 反之, 不难验证: 若 $(x',\,y')$ 是 F_2 上的点, 则一定存在 $(x,\,y)\in F_1$ 使得 $(x',\,y')=\alpha(x,\,y)$.

从这个例子可以看出, 平面内的圆和椭圆全都仿射等价. ■

6.3.2 平面二次曲线

(6.3.6) 现在我们正式开始完整地讨论平面二次曲线的仿射分类.

我们知道, 最一般形式的二元二次多项式形如

$$f(x,\,y)=a_{11}x^2+2a_{12}xy+a_{22}y^2+2b_1x+2b_2y+C \tag{6.3.6.1}$$

其中 $a_{11}, a_{12}, a_{22}, b_1, b_2, C\in\mathbb{R}$ 均为常数, 且 a_{11}, a_{12}, a_{13} 三者不全为 0. 令

$$A=\begin{pmatrix}a_{11}&a_{12}\\a_{12}&a_{22}\end{pmatrix}\ ,\ B=(b_1,\,b_2)$$

则 (6.3.6.1) 式可改写为如下矩阵形式

$$f(x,\,y)=(x,\,y)A\begin{pmatrix}x\\y\end{pmatrix}+2B\begin{pmatrix}x\\y\end{pmatrix}+C\,. \tag{6.3.6.2}$$

设

$$\alpha\ :\ \mathbb{A}^2\longrightarrow\mathbb{A}^2\,;\quad \begin{pmatrix}x\\y\end{pmatrix}\longmapsto M\begin{pmatrix}x\\y\end{pmatrix}+t \tag{6.3.6.3}$$

为一个可逆的仿射变换, 其中 $M=(m_{ij})$ 为可逆矩阵, $t=(t_1,\,t_2)^T\in\mathbb{R}^2$. 若令

$$\begin{cases}X:=m_{11}x+m_{12}y+t_1\\ Y:=m_{21}x+m_{22}y+t_2\end{cases}\qquad \text{亦即}\qquad \begin{pmatrix}X\\Y\end{pmatrix}:=M\begin{pmatrix}x\\y\end{pmatrix}+t\,, \tag{6.3.6.4}$$

则

$$\begin{pmatrix}x\\y\end{pmatrix}=P\begin{pmatrix}X\\Y\end{pmatrix}+u\qquad \text{其中}\ \ P=M^{-1}\,,\ u=-Pt\,. \tag{6.3.6.5}$$

记 $F=V(f)$ 为 (6.3.6.1) 式中多项式 f 定义的平面二次曲线. 将 (6.3.6.5) 式代入 (6.3.6.2) 式可知, 当 $(x,\,y)\in F=V(f)$ 时, 点 $(X,\,Y):=\alpha(x,\,y)$ 满足方程 $g(X,\,Y)=0$, 其中

$$\begin{aligned}g(X,\,Y)&=\big((X,\,Y)P^T+u^T\big)\,A\left(P\begin{pmatrix}X\\Y\end{pmatrix}+u\right)+2B\left(P\begin{pmatrix}X\\Y\end{pmatrix}+u\right)+C\\&=(X,\,Y)(P^TAP)\begin{pmatrix}X\\Y\end{pmatrix}+(X,\,Y)P^TAu+u^TAP\begin{pmatrix}X\\Y\end{pmatrix}+u^TAu+2BP\begin{pmatrix}X\\Y\end{pmatrix}+2Bu+C\\&=(X,\,Y)(P^TAP)\begin{pmatrix}X\\Y\end{pmatrix}+2(u^TAP+BP)\begin{pmatrix}X\\Y\end{pmatrix}+(u^TAu+2Bu+C)\end{aligned}$$

注意这里的 $(X,\,Y)P^TAu$ 是 1×1 矩阵, 故而等于其转置 $u^TAP\begin{pmatrix}X\\Y\end{pmatrix}$.

因此, 若令

$$A'=P^TAP\,,\ B'=u^TAP+BP\,,\ C'=u^TAu+2Bu+C \tag{6.3.6.6}$$

则

$$g(X,\,Y)=(X,\,Y)A'\begin{pmatrix}X\\Y\end{pmatrix}+2B'\begin{pmatrix}X\\Y\end{pmatrix}+C'\,. \tag{6.3.6.7}$$

所以, 如果用字母 X,Y 作为新的坐标, 那么 (6.3.6.7) 中多项式方程 $g=0$ 就是原来二次曲线 $F=V(f)$ 在新坐标下的方程, 或者说, g 的零点集 $G=V(g)$ 就是原来二次曲线 $F=V(f)$ 在 (6.3.6.3) 式所给的仿射变换映射下的像.

以上讨论说明, 如果想做仿射变换将二次曲线 F 的方程化成更简单的形式, 那么实际上就是找出一个适当的如 (6.3.6.4) 式所示的变量替换, 使得转换之后的多项式 g 具有更简单的形式. 我们将 (6.3.6.4) 式称为仿射平面的一个 (可逆) **仿射变量替换** (affine change of variables) 或**仿射坐标变换** (affine coordinate change), (6.3.6.5) 式展示的是它的逆变换.

我们还可以换一个角度来理解上面的讨论. 假设先给出 (6.3.6.2) 和 (6.3.6.7) 所示的两个多项式, 其中 $A,\,A'$ 是二阶非零对称阵, $B,\,B'\in\mathbb{R}^{1\times 2}$, $C,\,C'\in\mathbb{R}$. 如果存在可逆矩阵 $P\in\mathbf{M}_2(\mathbb{R})$ 和列向量 $u\in\mathbb{R}^2$ 使得 (6.3.6.6) 式中的关系成立, 那么二次曲线 $F=V(f)$ 和 $G=V(g)$ 仿射等价. (注意: (6.3.6.6) 中的第一个关系式说明对称矩阵 A 和 A' 相合.) 因为多项式 f 和它的一个非零常数倍 λf

决定出的二次曲线 $V(f)$ 和 $V(\lambda f)$ 是相同的, 所以我们可以把刚才所述的条件减弱一点, 即要求存在可逆矩阵 $P \in \mathbf{M}_2(\mathbb{R})$, 列向量 $u \in \mathbb{R}^2$ 和非零常数 λ 使得

$$\lambda A' = P^T AP\,,\ \lambda B' = u^T AP + BP\,,\ \lambda C' = u^T Au + 2Bu + C\,. \tag{6.3.6.8}$$

当 (6.3.6.8) 式成立时, 我们称 (6.3.6.2) 和 (6.3.6.7) 中所给的两个二次多项式 f 和 g **仿射等价**, 或者说 $f = 0$ 和 $g = 0$ 这两个二次方程**仿射等价**. ■

(6.3.7) 我们继续设 $F = V(f)$ 是由 (6.3.6.1) 式 (或 (6.3.6.2) 式) 所示多项式定义的平面二次曲线. 以下我们主要通过研究二次方程的仿射等价来对二次曲线进行分类.

首先我们知道, 非零对称阵 A 相合于以下五种规范形之一

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

因为 $V(f) = V(-f)$, 在不计非零常数倍时, 我们只需考虑以下 3 种情况

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

也就是说, 经过适当的可逆仿射变换, 我们可以设 (6.3.6.2) 式中的矩阵 A 具有以上 3 种形式之一.

情况 1: $A = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$.

此时 (6.3.6.1) 式形如 $f(x, y) = x^2 + y^2 + 2b_1x + 2b_2y + C$. 显然, 再做坐标变换

$$X = x + b_1\,,\ Y = y + b_2$$

(几何上说, 这种坐标变换对应坐标轴的平移) 即可将 $f(x, y)$ 化为

$$g(X, Y) = X^2 + Y^2 - d\,, \qquad \text{其中 } d := b_1^2 + b_2^2 - C\,.$$

情况 1.1: 若 $d > 0$, 则 $V(g)$ 是个圆 (circle), 而原来的 $V(f)$ 一般来说是个椭圆 (ellipse)‖

情况 1.2: 若 $d = 0$, 则 $V(g)$ 是一个点, 原来的 $V(f)$ 也是一个点.

情况 1.3: 若 $d < 0$, 则 $V(g)$ 空集, 故原来的 $V(f)$ 也是空集. 此时我们可以说二次曲线 $V(f)$ **没有实轨迹** (no real locus) 或**没有实图像** (no real image). 不过, 同样的方程如果允许在 $(X, Y) \in \mathbb{C}^2$ 范围内求解, 那么经过 $X' = iX$, $Y' = iY$ 这样的坐标变换又可以将方程化为一个椭圆的方程. 所以, 此时 $V(f)$ 可以想象成一个椭圆上点的坐标全都乘以虚数单位 i 后得到的图像. 人们把这种情况下的二次曲线 $V(f)$ 称为**虚椭圆** (imaginary ellipse).

情况 2: $A = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$.

此时 (6.3.6.1) 式形如 $f(x, y) = x^2 - y^2 + 2b_1x + 2b_2y + C$. 显然, 再做坐标变换

$$X = x + b_1\,,\ Y = y - b_2$$

即可将 $f(x, y)$ 化为

$$g(X, Y) = X^2 - Y^2 - d\,, \qquad \text{其中 } d := b_1^2 - b_2^2 - C\,.$$

情况 2.1: 若 $d \neq 0$, 则 $V(g)$ 是双曲线 (hyperbola).

‖不过这一点我们需要用到今后介绍的正交变换才能加以说明.

情况 2.2: 若 $d=0$, 则 $V(g)$ 是两条相交直线 (two intersecting lines).

情况 3: $A=\begin{pmatrix}1 & \\ & 0\end{pmatrix}$.

此时 (6.3.6.1) 式形如 $f(x,y)=x^2+2b_1x+2b_2y+C$. 令 $X=x+b_1$, $Y=y$ 即可将 $f(x,y)$ 化为

$$g(X,Y)=X^2+2b_2Y-d\,,\quad \text{其中 } d:=b_1^2-C\,.$$

情况 3.1: 若 $b_2\neq 0$, 则 $V(g)$ 是抛物线 (parabola).

情况 3.2: 若 $b_2=0$, $d>0$, 则 $V(g)$ 是两条平行直线 (two parallel lines).

情况 3.3: 若 $b_2=d=0$, 则 $V(g)$ 是两条重合直线 (two coincident lines).

情况 3.4: 若 $b_2=0$, $d<0$, 则 $V(g)$ 是空集. 不过, 和虚椭圆的情况类似, 通过在复数范围内做变量替换 $X'=iX$ 可以将方程变为两条平行直线的方程. 因此此时我们称 $V(g)$ 是两条**虚平行直线** (two imaginary parallel lines).

综上所述, 二元二次方程按照仿射等价来分类共可分成 9 类. 我们将其分成三大类型, 分别称为**椭圆型** (elliptic type)、**双曲型** (hyperbolic type) 和**抛物型** (parabolic type), 详细列举如下:

- 椭圆型三类: 椭圆、一个点、虚椭圆;
- 双曲型两类: 双曲线、两条相交直线;
- 抛物型四类: 抛物线、两条平行直线、两条重合直线、两条虚平行直线.

若从相应二次型的类型来看, 椭圆型对应的是正定的二次型, 抛物型对应的是半正定但不正定的二次型, 而双曲型对应的是不定的二次型. 这种通过二次型的特征来区分几何对象的方法实际上在微分几何学中还可以做更一般的推广.

如果去掉没有实轨迹的两种情况 (虚椭圆和两条虚平行直线), 那么二次曲线实际上只有 7 种类别. 两条具有实轨迹的二次曲线属于这 7 种类别中的同一种当且仅当它们仿射等价. 不过, 由于按照二次方程进行分类可以获得更多的信息量, 许多教材都会说平面二次曲线有 9 类. 这种说法虽然和实际图像上看到的现象不完全吻合, 但至今似乎已经成为约定俗成的说法了.

椭圆、双曲线和抛物线这三种二次曲线也合称**圆锥曲线** (conic section), 因为它们都可以实现为三维空间中圆锥面 (conical surface) 和平面的交线 (平面与圆锥面的不同位置关系区分了椭圆、双曲线和抛物线这三种不同情况). 圆锥曲线是真正非退化意义下的二次曲线, 它们的确在几何上具有 1 维的特征, 而且图像上不是由一次曲线拼成的. ■

注记 6.3.8. 准确来说, 前面的讨论只说明了任何平面二次曲线 F 可以通过仿射变换变成上面说的 9 种类型之一, 但目前的论证还不足以说明 F 本身的图形形貌和仿射变换之后的形貌一致. 例如, 即使知道了 F 经过仿射变换变成椭圆 (或双曲线等), 我们当前还无法说明 F 本身就是椭圆 (或双曲线等), 尽管这样的结论的确是对的. 这里的主要原因是仿射变换不能保持距离和角度的不变性, 因此变换之后的图形对称性不能保证变换之前的图形具有同样的对称性. 这一缺陷正是人们研究正交变换—即保持距离和角度的仿射变换—的动因. 我们今后会说明, 当二次曲线 F 可以通过仿射变换变成以上 9 种类型之一时, 它也可以通过正交变换变成同样类型的曲线. 所以, 这就可以保证从图像上看二次曲线确实只有 9 种 (或者说 7 种) 形状. ■

6.3.3 空间二次曲面

有了上一小节分类二次曲线的经验, 我们不难发现对空间二次曲面可以用类似的方法处理.

(6.3.9) 空间二次曲面由三元二次多项式的零点集定义. 采用矩阵方式书写的话, 三元二次多项式的一般形式为

$$f(x,\,y,\,z)=(x,\,y,\,z)A\begin{pmatrix}x\\y\\z\end{pmatrix}+2B\begin{pmatrix}x\\y\\z\end{pmatrix}+C \tag{6.3.9.1}$$

其中 A 为 3 阶对称阵, $B\in\mathbb{R}^{1\times 3}$, $C\in\mathbb{R}$. 通过考虑 A 的相合规范形, 并且忽略非零常数倍带来的差别, 我们只需考虑 A 是以下矩阵之一的情况:

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

情况 1: $A=\begin{pmatrix}1&&\\&1&\\&&1\end{pmatrix}$.

这种情况下 (6.3.9.1) 式可以改写为

$$f(x,\,y,\,z)=x^2+y^2+z^2+2b_1x+2b_2y+2b_3z+C\,.$$

于是, 通过坐标轴的如下平移

$$X=x+b_1\,,\ Y=y+b_2\,,\ Z=z+b_3$$

可将 f 化为

$$g(X,\,Y,\,Z)=X^2+Y^2+Z^2-d\quad \text{其中 } d:=b_1^2+b_2^2+b_3^2-C\,.$$

接下来很容易再区分出三种情况:

情况 1.1: 若 $d>0$, 则 $V(g)$ 是个球面 (sphere), 而原来的 $V(f)$ 一般来说是个椭球面 (ellipsoid).

情况 1.2: 若 $d=0$, 则 $V(g)$ 是一个点, 原来的 $V(f)$ 也是一个点.

情况 1.3: 若 $d<0$, 则 $V(g)$ 和原来的 $V(f)$ 都是空集. 不过, 和虚椭圆的情况类似, 我们说此时 $V(g)$ 是个**虚椭球面** (imaginary ellipsoid).

情况 2: $A=\begin{pmatrix}1&&\\&1&\\&&-1\end{pmatrix}$.

这种情况下 (6.3.9.1) 式可以改写为

$$f(x,\,y,\,z)=x^2+y^2-z^2+2b_1x+2b_2y+2b_3z+C\,.$$

于是, 通过坐标轴的如下平移

$$X=x+b_1\,,\ Y=y+b_2\,,\ Z=z-b_3$$

可将 f 化为

$$g(X,\,Y,\,Z)=X^2+Y^2-Z^2-d\quad \text{其中 } d:=b_1^2+b_2^2-b_3^2-C\,.$$

情况 2.1: 若 $d>0$, 图形 $V(g)$ 称为一个**单叶双曲面** (hyperboloid of one sheet). 它是一个连通的曲面, 用平行于 z 轴的平面截取交线时得到的是双曲线, 而用任何平行于 xy 平面的平面取截线时则得到椭圆.

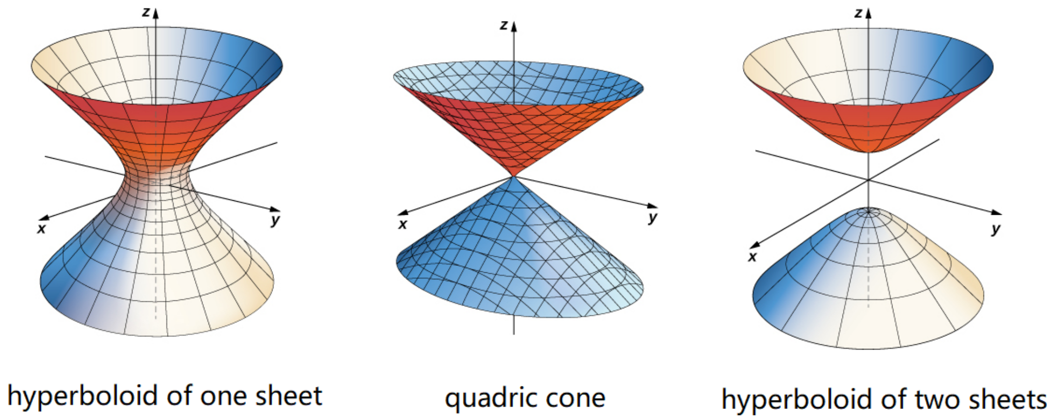

图 6.1: 单叶双曲面、二次锥面和双叶双曲面

情况 2.2: 若 $d=0$, 则图形 $V(g)$ 称为一个**二次锥面** (quadric cone) 或**双锥面** (double cone). 它由两个顶点重合且对称放置的圆锥面组成.

情况 2.3: 若 $d<0$, 图形 $V(g)$ 称为一个**双叶双曲面** (hyperboloid of two sheets). 它是一个不连通的曲面, 用平行于 xy 平面的平面取截线时, 若平面远离原点可以得到椭圆, 但若平面很靠近原点, 则会得到空集.

情况 3: $A=\begin{pmatrix}1 & & \\ & 1 & \\ & & 0\end{pmatrix}$.

此时 f 可化为

$$g(X,Y,Z)=X^2+Y^2+2b_3Z-d \quad \text{其中 } d:=b_1^2+b_2^2-C\,.$$

然后共分以下 4 种情况:

情况 3.1: 若 $b_3\neq 0$, 则 $V(g)$ 称为一个**椭圆抛物面** (elliptic paraboloid). 以平行于 Z 轴的平面截取交线时会得到抛物线, 而以平行于 XY 平面的平面取截线时会得到椭圆 (或一个点, 或虚椭圆).

情况 3.2: 若 $b_3=0$ 且 $d>0$, 则 $V(g)$ 是一个**椭圆柱面** (elliptic cylinder).

情况 3.3: 若 $b_3=0$ 且 $d=0$, 则 $V(g)$ 是一条直线.

情况 3.4: 若 $b_3=0$ 且 $d<0$, 则 $V(g)$ 是空集, 但我们称之为一个**虚椭圆柱面** (imaginary elliptic cylinder).

情况 4: $A=\begin{pmatrix}1 & & \\ & -1 & \\ & & 0\end{pmatrix}$

此时 f 可化为

$$g(X,Y,Z)=X^2-Y^2+2b_3Z-d \quad \text{其中 } d:=b_1^2-b_2^2-C\,.$$

然后共分以下 3 种情况:

情况 4.1: 若 $b_3\neq 0$, 则 $V(g)$ 称为一个**双曲抛物面** (hyperbolic paraboloid) 或**马鞍面** (saddle surface). 以平行于 Z 轴的平面截取交线时会得到抛物线, 而以平行于 XY 平面的平面取截线时会得到双曲线.

情况 4.2: 若 $b_3=0$ 且 $d\neq 0$, 则 $V(g)$ 是一个**双曲柱面** (hyperbolic cylinder).

情况 4.3: 若 $b_3=0$ 且 $d=0$, 则 $V(g)$ 是两个相交平面.

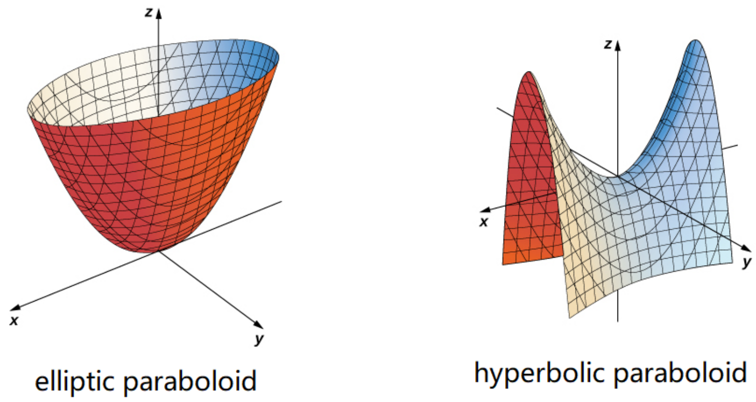

图 6.2: 椭圆抛物面和双曲抛物面

情况 5: $A = \begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \end{pmatrix}$.

此时 f 可化为

$$g(X, Y, Z) = X^2 + 2b_2Y + 2b_3Z - d \quad \text{其中 } d := b_1^2 - C\,.$$

然后共分以下 4 种情况:

情况 5.1: 若 b_2, b_3 不全为 0, 例如 $b_2 \neq 0$. 此时通过进一步的坐标变换 $Y' = Y + b_2^{-1}b_3Z$, $Z' = Z$ 不妨设 $b_3 = 0$. 于是不难看出此时 $V(g)$ 是一个**抛物柱面** (parabolic cylinder).

情况 5.2: 若 $b_2 = b_3 = 0$ 且 $d > 0$, 则 $V(g)$ 是两个平行平面.

情况 5.3: 若 $b_2 = b_3 = 0$ 且 $d = 0$, 则 $V(g)$ 是两个重合平面.

情况 5.4: 若 $b_2 = b_3 = 0$ 且 $d < 0$, 则 $V(g)$ 是空集, 但我们称之为一对**虚平行平面** (imaginary parallel planes).

综上所述, 空间中二次曲面的方程按仿射等价来分类共有 17 个类型, 其中有实际图像的有 14 类.

和二次曲线的情况一样, 以上的讨论只能说明二次曲面可以经过仿射变换变成以上 17 种 (或 14 种) 类型, 但当前我们还不能说明它们原来的样貌就是这些类型之一. 要说明这一点要使用今后介绍的正交变换的概念. ∎

6.3.4 习题

习题 6.3.1. 确定如下方程定义的二次曲面类型:

1. $(2x + y + z)^2 - (x - y - z)^2 = y - z$;
2. $9x^2 - 25y^2 + 9z^2 - 24xz + 80x - 60z = 0$.

习题 6.3.2. 假设如下方程定义的二次曲线是椭圆 (或双曲线、抛物线)

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2b_1x + 2b_2y + c = 0\,,$$

其中 $a_{11}, a_{12}, a_{22}, b_1, b_2, c \in \mathbb{R}$. 试问二次曲面

$$z = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2b_1x + 2b_2y + c$$

是哪种类型的曲面？

习题 6.3.3. 假设 $a_i, b_i, c_i \in \mathbb{R}$, $i = 1, 2, 3$ 满足 $\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \neq 0$. 试问二次曲面

$$(a_1x + b_1y + c_1z)^2 + (a_2x + b_2y + c_2z)^2 + (a_3x + b_3y + c_3z)^2 = 1$$

是哪种类型的曲面？

习题 6.3.4. 假设含有实参数 λ 的二次方程

$$\lambda x^2 + 4xy + y^2 - 4x - 2y - 3 = 0$$

定义的平面曲线是一对 (平行、相交或重合) 直线. 求出参数 λ 的可能取值.

习题 6.3.5. 按照实参数 λ 的取值讨论下列二次曲线的仿射等价类型:

1. $(1 + \lambda^2)(x^2 + y^2) - 4\lambda xy + 2\lambda(x + y) + 2 = 0$.
2. $x^2 - 4xy - 4y^2 + 8y - 2\lambda(2xy + x + 1) + 3 = 0$.

习题 6.3.6. 假设实数 a, b, c 满足 $(a^2 + b^2)c \neq 0$. 试判断二次曲线

$$ab(x^2 - y^2) - (a^2 - b^2)xy = c$$

的仿射等价类型.

习题 6.3.7. 通过适当的仿射坐标变换确定下列二次曲面按仿射等价分类属于 (17 种类型中的) 哪种类别?

1. $x^2 + 7y^2 + z^2 + 10xy + 2xz + 10yz + 8x + 4y + 8z - 6 = 0$.
2. $y^2 + 2xy - 4xz - 2yz - 2y + 8z + 3 = 0$.

第七章　实内积空间

这一章和下一章中我们将展开对向量空间度量结构的研究. 所谓一个向量空间具备了度量, 就是指在其中的元素 (即向量) 之间有距离的概念. 实向量空间上定义“距离”的最常见办法是用内积来定义, 这是从我们最熟悉的空间 $\mathbb{R}^n$ 的情况抽象推广而来.

若非有相反的声明, 本章中的向量空间总是指实向量空间.

7.1　内积空间与相关概念

7.1.1　内积的概念及常用不等式

定义 7.1.1. 设 V 是实向量空间. 所谓 V 上的一个**内积** (inner product) 是指 V 上一个正定的对称双线性型. 也就说, V 上的内积是指一个映射

$$\langle\,\cdot\,,\,\cdot\,\rangle\;:\;V\times V\longrightarrow\mathbb{R}\,;\quad (u,\,v)\longmapsto\langle\,u,\,v\,\rangle$$

它满足以下性质:

- 双线性: 对于任意 $u,\,v,\,w\in V$ 和任意 $a,\,b\in\mathbb{R}$,

$$\begin{aligned}\langle\,au+bv,\,w\,\rangle&=a\langle\,u,\,w\,\rangle+b\langle\,v,\,w\,\rangle\,,\\ \langle\,w,\,au+bv\,\rangle&=a\langle\,w,\,u\,\rangle+b\langle\,w,\,v\,\rangle\,.\end{aligned}$$

- 对称性: 对任意 $u,\,v\in V$, $\langle\,u,\,v\,\rangle=\langle\,v,\,u\,\rangle$.
- 正定性: 对任意 $v\in V$, $\langle\,v,\,v\,\rangle\geq 0$ 且只有当 $v=0$ 时才会有 $\langle\,v,\,v\,\rangle=0$.

一个 (实) **内积空间** (inner product space) 是指一个有序对 $(V\,,\,\langle\,\cdot\,,\,\cdot\,\rangle)$, 其中 V 是个实向量空间, $\langle\,\cdot\,,\,\cdot\,\rangle$ 是 V 上的一个内积. 通常在不致引起歧义的时候我们简称 V 是个内积空间. 实内积空间也称为**欧几里得空间**或**欧氏空间** (Euclidean space).

假设 $(V,\langle\,\cdot\,,\,\cdot\,\rangle)$ 是个内积空间. 根据正定性, 对于每个向量 $v\in V$, 可以定义其**长度** (length)、**模长** (modulus) 或**范数** (norm) 为 $\|v\|:=\sqrt{\langle\,v,\,v\,\rangle}$. 只有零向量的长度是 0, 其它向量的长度都是正实数. 此外, 对于 V 中每两个向量 $u,\,v$, 可以定义二者的**距离** (distance) 为 $\|u-v\|$. 读者需要留意, 长度和距离等相关的概念依赖于内积的选取. 同一向量相对于不同的内积长度可以是不同的.

当 $\mathcal{B}=(v_1,\cdots,v_n)$ 是 V 的一组有序基, $\langle\,\cdot\,,\,\cdot\,\rangle$ 是 V 上的一个内积, 相应的 Gram 矩阵 $(\langle\,v_i,\,v_j\,\rangle)$ 也称为内积 $\langle\,\cdot\,,\,\cdot\,\rangle$ 在 $\mathcal{B}$ 下的**度量矩阵** (metric matrix). ■

思考题 7.1. 设 $(V,\langle\,\cdot\,,\,\cdot\,\rangle)$ 是 (实) 内积空间. 证明:

1. 对于任意 $v\in V$ 和任意 $c\in\mathbb{R}$, $\|cv\|=|c|\cdot\|v\|$.
2. (**极化恒等式** polarization identity) 对于任意 $u,\,v\in V$, $\langle\,u\,,\,v\,\rangle=\frac{1}{4}\big(\|u+v\|^2-\|u-v\|^2\big)$.

3. (**平行四边形恒等式** parallelogram identity) 对于任意 $u, v \in V$,

$$\|u+v\|^2+\|u-v\|^2=2(\|u\|^2+\|v\|^2)\,.$$

(此恒等式的一个几何解释为: 平行四边形的两条对角线长度的平方和等于其四条边长度的平方和.)

例 7.1.2. 来举几个内积空间的例子.

(1) 向量空间 $\mathbb{R}^n$ 上熟知的一个内积为

$$\langle\, x,\, y\,\rangle := x_1y_1+\cdots+x_ny_n$$

其中 $x=(x_1,\cdots,x_n)^T$, $y=(y_1,\cdots,y_n)^T$. 我们称之为 $\mathbb{R}^n$ 上的**标准内积** (standard inner product). 今后若无相反的说明, $\mathbb{R}^n$ 作为内积空间时总是认为配上标准内积.

(2) 对于任意正定矩阵 $A\in\mathbf{M}_n(\mathbb{R})$, 对称双线性型

$$\langle\cdot\,,\,\cdot\rangle_A\ :\ \mathbb{R}^n\times\mathbb{R}^n\longrightarrow\mathbb{R}\,;\quad \langle\, x,\, y\,\rangle_A=x^TAy$$

是 $\mathbb{R}^n$ 上的内积. 当 $A=I_n$ 时, 这就是上面所说的标准内积. 再比如, 对于任意正实数 $c_1,\cdots,c_n$,

$$\mathbb{R}^n\times\mathbb{R}^n\longrightarrow\mathbb{R}\,;\quad (x,\,y)\longmapsto c_1x_1y_1+\cdots+c_nx_ny_n$$

是 $\mathbb{R}^n$ 上的内积.

(3) 设 $[a,\,b]$ 是 $\mathbb{R}$ 中的有界闭区间, $V=\mathcal{C}[a,\,b]$ 是从 $[a,\,b]$ 到 $\mathbb{R}$ 的所有连续函数构成的空间. 则

$$V\times V\longrightarrow\mathbb{R}\,;\quad (f,\,g)\longmapsto\langle\, f,\, g\,\rangle:=\int_a^b f(x)g(x)\mathrm{d}x$$

是 V 上的一个内积.

(4) 考虑次数 $\le n$ 的实系数多项式空间 $V=\mathbb{R}[X]_{\le n}$. 则

$$V\times V\longrightarrow\mathbb{R}\,;\quad (f,\,g)\longmapsto\langle\, f,\, g\,\rangle:=\sum_{k=0}^n f(k)g(k)$$

是 V 上的一个内积. ■

(7.1.3) 按照定义, 向量空间 V 上的内积 $\langle\cdot\,,\,\cdot\rangle$ 是一个对称双线性型. 因此, 我们可以对 V 中向量定义正交的概念. 即, 若 $u,\,v\in V$ 满足 $\langle\, u,\, v\,\rangle=0$, 则称 u 与 v 正交, 记为 $u\perp v$. 对于 V 中的任意子集 S, 可以像定义 6.1.12 那样定义其正交补 $S^\perp$. 因为内积是正定的二次型, 所以它是个非退化的双线性型, 因此 V 的正交补 (即双线性型 $\langle\cdot\,,\,\cdot\rangle$ 的根) 等于 0. 也就是说, 能够和 V 中所有向量都正交的有且仅有零向量.

若 W 是 V 的一个子空间, 那么将 $\langle\cdot\,,\,\cdot\rangle$ 限制于 W 所得的双线性型 $\langle\cdot\,,\,\cdot\rangle_W$ 是 W 上的正定双线性型, 也就是 W 上的内积. 特别地, $\langle\cdot\,,\,\cdot\rangle_W$ 仍是非退化的.

此外, 如果子空间 W 是有限维空间的, 则 Riesz 表示定理对 W 成立: 即对于任意线性泛函 $f\in W^\circ$, 一定存在唯一的向量 $w\in W$ 使得

$$f=\langle\cdot\,,\,w\,\rangle\ :\ x\longmapsto f(x)=\langle\, x,\, w\,\rangle\,.$$

根据命题 6.1.25 可知: 对于内积空间 V 中的每个有限维子空间 W, 总有正交分解 $V=W\perp W^\perp$ 成立. (注意 V 不需要是有限维的.) 作为一个特例, 可以考虑任意一个非零向量 $v\in V$. 此时正交分解式 $V=\mathrm{span}(v)\perp(\mathrm{span}(v))^\perp$ 表明

(7.1.3.1) 对于任意 $u\in V$, 存在唯一的实数 $c\in\mathbb{R}$ 及向量 $w\in V$, 使得 $u=cv+w$ 且 $w\perp v$,.

事实上, (7.1.3.1) 式中系数 c 很容易求出, 它应该等于 $c=\frac{\langle\, u,\, v\,\rangle}{\|v\|^2}$. ■

命题 7.1.4 (毕达哥拉斯定理 Pythagoras theorem): 设 $(V, \langle\cdot\,,\,\cdot\rangle)$ 是 (实) 内积空间. 则对于任意 $u, v \in V$, 等式 $\|u+v\|^2 = \|u\|^2 + \|v\|^2$ 成立的充分必要条件是 $u \perp v$.

证明. 展开计算 $\|u+v\|^2 = \langle\, u+v\,,\, u+v\,\rangle$ 即可. □

定理 7.1.5: 设 $(V, \langle\cdot\,,\,\cdot\rangle)$ 是 (实) 内积空间, $u, v \in V$.

1. Cauchy–Schwarz **不等式**, 也称 Cauchy–Bunyakovsky–Schwarz **不等式** (Cauchy–Schwarz inequality, Cauchy*–Bunyakovsky†–Schwarz‡ inequality):

   $|\langle\, u\,,\, v\,\rangle| \le \|u\| \cdot \|v\|$. 等号成立的充分必要条件是 u, v 线性相关.

2. **三角不等式** (triangle inequality): $\|u+v\| \le \|u\| + \|v\|$. 等号成立的充分必要条件是存在实数 $c \ge 0$ 满足 $u = cv$ 或 $v = cu$.

证明. (1) 鉴于其重要性, 我们介绍两种最常见的证明方法.

证法一: 不妨设 $v \ne 0$. 欲证不等式两边都是非负实数, 因此可以取平方后再比较二者的大小关系. 根据 (7.1.3.1), 存在 $c \in \mathbb{R}$ 和向量 w 与 v 正交, 使得 $u = cv + w$. 于是,

$$\begin{aligned}
(\|u\| \cdot \|v\|)^2 &= \|cv + w\|^2 \cdot \|v\|^2 \\
\text{(根据毕达哥拉斯定理)} \qquad &= (\|cv\|^2 + \|w\|^2) \cdot \|v\|^2 \\
&\ge \|cv\|^2 \cdot \|v\|^2 = |c|^2 \cdot \|v\|^4 = (c\|v\|^2)^2 \\
&= (c\langle\, v\,,\, v\,\rangle)^2 = (\langle\, cv\,,\, v\,\rangle)^2 = (\langle\, cv + w\,,\, v\,\rangle)^2 \\
&= (\langle\, u\,,\, v\,\rangle)^2 .
\end{aligned}$$

对于以上推理过程中出现的唯一不等号, 等号能够取到的充分必要条件是 $\|w\|^2 = 0$, 即 $w = 0$, 也就是说 $u = cv$. 这就证明了等号成立的充分必要条件是 u, v 线性相关.

证法二: 仍不妨设 $v \ne 0$. 于是实数 $a := \langle\, v,\, v\,\rangle = \|v\|^2 > 0$. 再令 $b = \langle\, u\,,\, v\,\rangle$, $c = \langle\, u\,,\, u\,\rangle = \|u\|^2$. 此时对所有实数 $t \in \mathbb{R}$ 均有

$$0 \le \|u + tv\|^2 = \langle\, u + tv\,,\, u + tv\,\rangle = t^2\langle\, v\,,\, v\,\rangle + 2t\langle\, u\,,\, v\,\rangle + \langle\, u,\, u\,\rangle = at^2 + 2bt + c\,.$$

这说明一元二次函数 $f(t) = at^2 + 2bt + c$ 始终取非负值. 我们已经知道 $a > 0$. 因此 f 恒取非负值的充分必要条件是方程 $f(t) = 0$ 最多只有一个实根. 而通过初等数学我们知道这个性质成立的充分必要条件是 $(2b)^2 - 4ac \le 0$, 此即

$$(\langle\, u\,,\, v\,\rangle)^2 \le \|u\|^2 \cdot \|v\|^2\,.$$

等号成立 (即 $(2b)^2 - 4ac = 0$) 的充分必要条件是二次方程 $at^2 + 2bt + c = 0$ 有一个实根. 也就是说, 存在 $t \in \mathbb{R}$ 使得

$$at^2 + 2bt + c = \|u + tv\|^2 = 0\,.$$

这就是说 $u + tv = 0$, 即 u, v 线性相关. 因此我们说明了等号成立的充分必要条件是 u, v 线性相关.

(2) 利用 Cauchy–Bunyakovsky–Schwarz 不等式可得

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

---

*Augustin Louis Cauchy (1789–1857), 法国数学家.

†Viktor Yakovlevich Bunyakovsky (俄文: Виктор Яковлевич Буняковскии) (1804–1889), 俄国数学家.

‡Hermann Amandus Schwarz (1843–1921), 德国数学家.

以上等号成立的充分必要条件是 $\langle u, v\rangle = \|u\| \cdot \|v\|$. 根据 (1) 中所述等号成立的条件, 三角不等式取到等号的充分必要条件是 u, v 线性相关且 $\langle u\,, v\rangle \geq 0$. 读者不难验证, 这和定理中所述的条件是等价的. $\square$

注记 7.1.6. 我们已经介绍了与内积有关的两个概念: 范数 (或曰长度) 和距离. 这两个概念其实可以在更一般的情况下定义.

首先, 范数的概念可以不依赖于内积来定义, 而且可以不仅限于实向量空间. 事实上, 对于 $\mathbb{C}$ 的任一子域 K 和任意 K-向量空间 V, 所谓 V 上的一个**范数** (norm) 是指一个映射

$$\|\cdot\| \;:\; V \longrightarrow \mathbb{R}\,; \quad v \longmapsto \|v\|\,,$$

它满足以下条件:

(N 1) 对于任意 $v \in V$ 均有 $\|v\| \geq 0$, 而且 $\|v\| = 0$ 当且仅当 $v = 0$.

(N 2) 对于任意 $v \in V$ 和 $c \in K$, 均有 $\|cv\| = |c| \cdot \|v\|$. 这里 $|c|$ 是 c 作为复数计算出的模 (或称绝对值).

(N 3) 三角不等式成立: 即对于任意 $u, v \in V$, 均有 $\|u + v\| \leq \|u\| + \|v\|$.

向量空间 V 配上一个范数得到的有序对 $(V\,, \|\cdot\|)$ 称为一个**赋范向量空间** (normed vector space). 这个概念在后续深入的现代分析学中 (甚至在现代的数论中) 非常有用.

读者应该很容易明白: 如果 V 是一个实内积空间, 那么定义 7.1.1 中由 $v \mapsto \|v\| := \sqrt{\langle v\,, v\rangle}$ 给出的映射满足上述要求. 因此我们可以说实内积空间是赋范向量空间的一种.

至于距离的概念, 实际上可以对一般的集合来定义. 设 X 为非空集合. 所谓 X 上的一个**距离** (distance) 或**度量** (metric) 是指一个映射 $d : X \times X \longrightarrow \mathbb{R}$, 它满足以下性质:

(M 1) 对于任意 $x, y \in X$, $d(x, y) \geq 0$, 且 $d(x, y) = 0$ 的充分必要条件是 $x = y$.

(M 2) 对称性成立: 即对任意 $x, y \in X$, $d(x, y) = d(y, x)$.

(M 3) 三角不等式成立: 即对任意 $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.

集合 X 配上一个度量 d 构成的有序对 (X, d) 称为一个**度量空间** (metric space). 这是现代几何学 (包括拓扑学) 中非常重要的一个概念.

有兴趣的读者可以验证: 如果 $(V, \|\cdot\|)$ 是一个赋范向量空间, 那么通过 $d(x, y) := \|x - y\|$ 定义出的映射是 V 上的一个度量. 因此, 赋范向量空间可以认为是度量空间的一个特例. 特别地, 实内积空间都自然具有度量空间的结构. $\blacksquare$

例 7.1.7. 以下是 Cauchy–Bunyakovsky–Schwarz 不等式最常见的两个例子:

(1) 对于任意实数 $a_1, \cdots, a_n$ 和 $b_1, \cdots, b_n$,

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right) \cdot \left(\sum_{i=1}^{n} b_i^2\right)\,.$$

这是 $\mathbb{R}^n$ 的标准内积对应的不等式.

(2) 对于闭区间 $[a, b]$ 上的任意连续函数 f, g,

$$\left(\int_a^b f(x)g(x)\mathrm{d}x\right)^2 \leq \left(\int_a^b f(x)^2\mathrm{d}x\right) \cdot \left(\int_a^b g(x)^2\mathrm{d}x\right)\,.$$

这是例 7.1.2 (3) 中内积对应的不等式. $\blacksquare$

(7.1.8) 对于内积空间 V 中的任意两个非零向量 u, v, 根据 Cauchy–Bunyakovsky–Schwarz 不等式可知 $\frac{\langle u\,, v\rangle}{\|u\|\cdot\|v\|} \in [-1, 1]$. 因此反余弦函数值

$$\arccos\left(\frac{\langle u\,, v\rangle}{\|u\| \cdot \|v\|}\right)$$

是可以定义的, 我们称之为 u, v 之间的**夹角** (angle between u and v), 记为 $\angle uv$, 它的取值范围是 $[0, \pi]$ 这一区间. 因为内积具有对称性, 所以这个夹角是不带方向的, 即 $\angle uv = \angle vu$. 而且, $\angle uv = \frac{\pi}{2}$ 恰好对应于 $\langle u, v \rangle = 0$ (即 u 和 v 正交) 的情形. 注意, 如果 u 或 v 是零向量, 一般认为它们的夹角是无法定义的.

对于熟悉的几何空间 $\mathbb{R}^2$ 或 $\mathbb{R}^3$, 以上这个夹角的概念和我们直观理解的夹角概念吻合. ■

7.1.2 规范正交基和 Gram–Schmidt 正交化

之前的定理 6.1.15 已经告诉我们, 有限维空间上任意对称双线性型都可以找到一组正交基. 内积也是一种对称双线性型, 因此内积空间中总是可以找到对应于给定内积的正交基. 与之前的一般情况相比, 内积可以额外地为向量带来长度的概念. 因此, 在内积空间中我们可以讨论更为特殊的一种基——规范正交基.

定义 7.1.9. 设 V 为实内积空间.

(1) 如果向量 $v \in V$ 的长度为 1, 即, $\|v\| = 1$, 我们就称 v 是一个**单位向量** (unit vector).

(2) 如果 $v_1, \cdots, v_r$ 是 V 中的一组单位向量, 并且它们两两之间相互正交, 那么我们称 $v_1, \cdots, v_r$ 是 V 中的一个**规范正交组** (orthonormal system).

注意, 由一个单位向量 v_1 构成的向量组 v_1 自动认为是规范正交组.

(3) 如果一个规范正交组 $v_1, \cdots, v_n$ 构成 V 的一组基, 我们就称之为一组**规范正交基** (orthonormal basis). (也有很多书将我们所说的规范正交基称为"**标准正交基**".) ■

以下命题告诉我们: 在一个内积空间中, 两两正交的非零向量构成的向量组一定是线性无关的.

命题 7.1.10: *设 V 是 (有限或无限维) 实内积空间, $\mathscr{F}$ 是 V 中的一族两两正交的非零向量.*

则向量族 $\mathscr{F}$ 是线性无关的, 也就是说, 对于任意有限多个取自 $\mathscr{F}$ 的两两不同向量 $v_1, \cdots, v_m$, 向量组 $v_1, \cdots, v_m$ 总是线性无关的.

证明. 我们用反证法. 假设存在有限多个 $\mathscr{F}$ 中取出的互异向量 $v_1, \cdots, v_m$, 它们线性相关. 不妨设 v_m 可以写成向量组 $v_1, \cdots, v_{m-1}$ 的线性组合, 即 v_m 属于子空间 $W := \operatorname{span}(v_1, \cdots, v_{m-1})$. 因为 v_m 和 $v_1, \cdots, v_{m-1}$ 中每个向量正交, 因此它和空间 W 中所有向量正交. 刚才我们假设了 $v_m \in W$, 故 v_m 与自身正交. 但在 $v_m \neq 0$ 的前提下 $\langle v_m, v_m \rangle = \|v_m\|^2$ 不会等于 0. 因此我们找到了矛盾. □

思考题 7.2. 假设 φ 是向量空间 V 上一般的非退化对称双线性型, $v_1, v_2 \in V$ 是非零向量且关于 φ 正交. 是否 v_1, v_2 一定线性无关? 若是, 请解释理由. 若否, 请举出反例.

推论 7.1.11: *设 V 是 n 维实内积空间, $v_1, \cdots, v_n$ 是一组单位向量.*

则 $v_1, \cdots, v_n$ 是一组规范正交基当且仅当 $v_1, \cdots, v_n$ 两两正交.

证明. 根据命题 7.1.10, 若 $v_1, \cdots, v_n$ 两两正交, 则它们一定线性无关. 由于该向量组所含向量数量等于 $\dim V$, 因此一定构成 V 的一组基. □

使用规范正交基的意义在下面的结论中可以体现.

命题 7.1.12: *设 $\mathcal{E} = (e_1, \cdots, e_n)$ 是内积空间 V 的一组规范正交基.*

则对于任意向量 $v \in V$, 它在 $\mathcal{E}$ 下的坐标由下式给出:

$$\mathcal{M}_{\mathcal{E}}(v)^T = \left(\langle v, e_1 \rangle, \cdots, \langle v, e_n \rangle\right),$$

而 v 的长度可以按下式计算:

$$\|v\| = \sqrt{|\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2}.$$

事实上, 对任意 $u, v \in V$ 均有

$$\langle u, v\rangle = \mathcal{M}_{\mathcal{E}}(u)^T \mathcal{M}_{\mathcal{E}}(v)\,.$$

也就是说, 以下图表可交换

$$\begin{tikzcd} V \times V \arrow[r, "\langle\cdot,\cdot\rangle"] \arrow[d, "\mathcal{M}_{\mathcal{E}}(\cdot)\times\mathcal{M}_{\mathcal{E}}(\cdot)"'] & \mathbb{R} \arrow[d, equal] \\ \mathbb{R}^n \times \mathbb{R}^n \arrow[r, "\text{标准内积}"'] & \mathbb{R} \end{tikzcd}$$

证明. 设 $\mathcal{M}_{\mathcal{E}}(v)^T = (a_1, \cdots, a_n)$. 根据规范正交基的定义, 计算即得 $a_i = \langle v, e_i\rangle$. 至于 $\|v\|$ 和 $\langle u, v\rangle$ 的计算公式, 只需要注意到内积 $\langle\cdot\,,\,\cdot\rangle$ 在规范正交基 $\mathcal{E}$ 下的 Gram 矩阵是单位矩阵, 然后使用公式 (6.1.5.1) 即可. (计算 $\|v\|$ 的公式也可以使用毕达哥拉斯定理来得到.) □

例 7.1.13. 在内积空间 $\mathbb{R}^n$ 中 (相对于标准内积), 标准基 $e_1, \cdots, e_n$ 是一组规范正交基. 其它的规范正交基其实也有很多. 例如, 在 $\mathbb{R}^4$ 中,

$$\begin{pmatrix} \frac12 \\ \frac12 \\ \frac12 \\ \frac12 \end{pmatrix}, \begin{pmatrix} \frac12 \\ \frac12 \\ -\frac12 \\ -\frac12 \end{pmatrix}, \begin{pmatrix} \frac12 \\ -\frac12 \\ -\frac12 \\ \frac12 \end{pmatrix}, \begin{pmatrix} -\frac12 \\ \frac12 \\ -\frac12 \\ \frac12 \end{pmatrix}$$

也是一组规范正交基. ■

关于规范正交基的存在性证明, 主要有两种方法. 其中一种是先利用定理 6.1.15 得到一组正交基 $v_1, \cdots, v_n$. 这里的向量 v_i 都是非零向量, 故 $\|v_i\| \neq 0$. 令 $e_i = \|v_i\|^{-1} v_i$ 即可得到单位向量, 而 $e_1, \cdots, e_n$ 就构成规范正交基. 下面的命题表明, 在构造规范正交基时可以要求包含事先给定的一个规范正交组. (这类似于构造一组基时可以要求包含事先给定的一个线性无关组.)

命题 7.1.14: 设 V 是有限维实内积空间, $n = \dim V \geq 1$.

则对于任意给定的规范正交组 $e_1, \cdots, e_r$, 一定可以找到向量 $e_{r+1}, \cdots, e_n$ 使得

$$e_1, \cdots, e_r, e_{r+1}, \cdots, e_n$$

构成一组规范正交基. 特别地, V 中一定存在规范正交基.

证明. 先证明第一个论断. 假设已经给定规范正交向量组 $e_1, \cdots, e_r$. 令 $W = \mathrm{span}(e_1, \cdots, e_r)$. 若 $W = V$, 则无需再做任何事情. 故可设 $W \neq V$. 因为有正交分解 $V = W \perp W^\perp$, 在 $W^\perp$ 中任取非零向量 w, 再取 $e_{r+1} = \|w\|^{-1}.w$ 就得到 $W^\perp$ 中一个单位向量. 此时

$$e_1, \cdots, e_r\,, e_{r+1}$$

是 V 中一个新的规范正交组. 反复这样操作可以最终得到一个规范正交组 $e_1, \cdots, e_r, \cdots, e_n$. 根据推论 7.1.11, 这必然是 V 的一组规范正交基.

以上我们证明了, 从 V 中任意规范正交组出发一定可以扩充得到一组规范正交基. 而最初我们可以先任意选取一个非零向量 v, 通过令 $e_1 = \|v\|^{-1}.v$ 就得到一个单位向量. 从 e_1 单独构成的规范正交组出发, 可以将其扩充为 V 的一组规范正交基. 因此, 规范正交基必然存在. □

例 7.1.15. 命题 7.1.14 的证明当然也给出了一种找到规范正交基的算法. 我们通过一个例子来展示这一操作过程.

设 $V=\mathbb{R}[X]_{\leq 2}$. 考虑其上的内积

$$\langle f\,,\,g\rangle := \int_{-1}^{1} f(x)g(x)\mathrm{d}x\ , \quad f\,,\,g\ \in\ V=\mathbb{R}[X]_{\leq 2}\ .$$

为了找出一组规范正交基, 我们可以先将向量 $v_1=1\in V$ 调整为一个单位向量 e_1, 即令

$$e_1 := \frac{1}{\|1\|}\cdot 1 = \frac{1}{\left(\int_{-1}^{1} 1^2\mathrm{d}x\right)^{1/2}}\cdot 1 = \frac{1}{\sqrt{2}}\cdot 1\ .$$

计算子空间 $W_1=\operatorname{span}(v_1)$ 的正交补可得

$$W_1^{\perp} = \{aX^2+bX+c\in V\mid a+3c=0\}\ .$$

从中可以取出非零向量 X, 以及单位向量

$$e_2 = \frac{1}{\|X\|}\cdot X = \frac{1}{\left(\int_{-1}^{1} x^2\mathrm{d}x\right)^{1/2}}\cdot X = \frac{\sqrt{3}}{\sqrt{2}}X\ .$$

继续计算 $W_2=\operatorname{span}(1,\,X)$ 的正交补可得

$$W_2^{\perp} = \{aX^2+bX+c\in V\mid a+3c=b=0\}\ .$$

从中可以取出非零向量 $3X^2-1$, 以及单位向量

$$e_3 = \frac{1}{\|3X^2-1\|}\cdot(3X^2-1) = \frac{1}{\left(\int_{-1}^{1}(3x^2-1)^2\mathrm{d}x\right)^{1/2}}\cdot(3X^2-1) = \frac{\sqrt{5}}{\sqrt{8}}(3X^2-1)\ .$$

综上可知,

$$e_1=\frac{1}{\sqrt{2}}\ ,\ e_2=\frac{\sqrt{3}}{\sqrt{2}}X\ ,\ e_3=\frac{\sqrt{5}}{\sqrt{8}}(3X^2-1)$$

构成以上内积空间 $(\mathbb{R}[X]_{\leq 2}\,,\langle\cdot\,,\,\cdot\rangle)$ 的一组规范正交基. ■

下面我们再介绍一种构造规范正交基的重要方法. 当我们可以先找到空间的一组基 (不一定是正交基) 时, 使用这种方法非常有效. 这一方法不必多次计算正交补空间, 而且有统一的简明计算公式 (尽管在实际运算过程中计算量可能较大), 还可以给出额外的有用信息.

定理 7.1.16 (Gram–Schmidt **正交化过程** Gram–Schmidt[§] orthogonalization process): 设 V 是内积空间, $v_1,\cdots,v_m$ 是 V 中一个线性无关的向量组.

则必然存在 V 中的规范正交组 $e_1,\cdots,e_m$ 使得对于每个 $j\in[\![1,\,m]\!]$ 均有

(7.1.16.1) $$\operatorname{span}(v_1,\cdots,v_j)=\operatorname{span}(e_1,\cdots,e_j)\,.$$

事实上, 这一规范正交组必然满足

(7.1.16.2) $$\begin{aligned} & e_1 = \pm\frac{v_1}{\|v_1\|}\ ,\\ \text{对每个 } j=2,\cdots,m,\quad & e_j=\pm\frac{v_j-\langle e_1,\,v_j\rangle e_1-\cdots-\langle e_{j-1}\,,\,v_j\rangle e_{j-1}}{\|v_j-\langle e_1,\,v_j\rangle e_1-\cdots-\langle e_{j-1}\,,\,v_j\rangle e_{j-1}\|}\ .\end{aligned}$$

[§] Erhard Schmidt (1876–1959), 德国数学家.

证明. 我们可以用归纳的方式逐步构造出 $e_1, \cdots, e_m$ 这组向量. 因为 e_1 必须是单位向量, 且满足 $\mathrm{span}(e_1) = \mathrm{span}(v_1)$, 因此必然 $e_1 = \pm\|v_1\|^{-1}v_1$. 下面假设对于某个 $k \in [\![2, m]\!]$ 已经构造出 $e_1, \cdots, e_{k-1}$ 满足所需条件, 即对于每个 $j \in [\![1, k-1]\!]$ 均有 (7.1.16.1) 和 (7.1.16.2) 成立. 现在只要证明存在向量 e_k 满足

(7.1.16.3) $$\|e_k\| = 1\,,\ \mathrm{span}(e_1, \cdots, e_k) = \mathrm{span}(v_1, \cdots, v_k) \quad \text{且 } e_k \text{ 与 } e_1, \cdots, e_{k-1} \text{ 都正交}$$

而且满足这些要求的向量只能是

$$e_k = \pm\frac{v_k - \langle\, e_1,\, v_k \,\rangle e_1 - \cdots - \langle\, e_{k-1}\,,\, v_k \,\rangle e_{k-1}}{\|v_k - \langle\, e_1,\, v_k \,\rangle e_1 - \cdots - \langle\, e_{k-1}\,,\, v_k \,\rangle e_{k-1}\|}\,.$$

因为已经知道 $\mathrm{span}(e_1, \cdots, e_{k-1}) = \mathrm{span}(v_1, \cdots, v_{k-1})$, 所以

$$\begin{aligned}\mathrm{span}(v_1, \cdots, v_k) &= \mathrm{span}(v_1, \cdots, v_{k-1}) + \mathrm{span}(v_k) \\ &= \mathrm{span}(e_1, \cdots, e_{k-1}) + \mathrm{span}(v_k) = \mathrm{span}(e_1, \cdots, e_{k-1},\, v_k)\,.\end{aligned}$$

由此易见, 等式 $\mathrm{span}(e_1, \cdots, e_k) = \mathrm{span}(v_1, \cdots, v_k)$ 成立的充分必要条件是 $e_k \in \mathrm{span}(e_1, \cdots, e_{k-1},\, v_k)$. 即, 存在常数 $a_1, \cdots, a_k$ 满足

(7.1.16.4) $$e_k = a_k v_k - a_1 e_1 - \cdots - a_{k-1}e_{k-1}\,.$$

我们已经知道 $e_1, \cdots, e_{k-1}$ 是两两正交的单位向量, 而 e_k 又要求和 $e_1, \cdots, e_{k-1}$ 都正交. 通过 (7.1.16.4) 式逐个计算内积 $\langle\, e_j,\, e_k \,\rangle$, $j = 1, \cdots, k-1$ 可得

$$0 = a_k\langle\, e_1,\, v_k \,\rangle - a_1 = a_k\langle\, e_2,\, v_k \,\rangle - a_2 = \cdots = a_k\langle\, e_{k-1},\, v_k \,\rangle - a_{k-1}\,.$$

将这些结果代回 (7.1.16.4) 式可得

$$e_k = a_k\big(v_k - \langle\, e_1,\, v_k \,\rangle e_1 - \cdots - \langle\, e_{k-1}\,,\, v_k \,\rangle e_{k-1}\big)\,.$$

这里的向量 $u := v_k - \langle\, e_1,\, v_k \,\rangle e_1 - \cdots - \langle\, e_{k-1}\,,\, v_k \,\rangle e_{k-1}$ 一定是非零向量, 因为向量组 $v_1, \cdots, v_m$ 的线性无关性保证

$$v_k \notin \mathrm{span}(v_1, \cdots, v_{k-1}) = \mathrm{span}(e_1, \cdots, e_{k-1})\,.$$

从 $e_k = a_k u$ 和 $\|e_k\| = 1$ 可知, $a_k = \pm\|u\|^{-1}$. 也就是说, e_k 只能是 (7.1.16.2) 式限定的形式. 当然, 当 e_k 具有这种形式时, 以上讨论还可以说明 (7.1.16.3) 中的要求一定被满足. 至此定理证毕. □

推论 7.1.17: *设 V 是有限维内积空间, $v_1, \cdots, v_n$ 是 V 的任意一组基. 则存在规范正交基 $e_1, \cdots, e_n$ 使得*

$$\text{对每个 } j \in [\![1,\, n]\!]\,, \quad \mathrm{span}(v_1, \cdots, v_j) = \mathrm{span}(e_1, \cdots, e_j)\,.$$

(特别地, V 中一定存在规范正交基.)

证明. 由定理 7.1.16 显然. □

(7.1.18) 如果从几何的角度理解 Gram–Schmidt 正交化, 可以将其与 (7.1.3) 一段提到的正交分解联系起来. 事实上, (7.1.16.1) 式表明: e_1 的构造无非是将第一个向量 v_1 单位化, 而如果对每个 $j \in [\![2, m]\!]$ 令

$$W_{j-1} = \mathrm{span}(v_1, \cdots, v_{j-1}) = \mathrm{span}(e_1, \cdots, e_{j-1})$$

那么

$$u_j := \langle\, e_1,\, v_j \,\rangle e_1 + \cdots + \langle\, e_{j-1},\, v_j \,\rangle e_{j-1} \ \in\ W_{j-1}$$

而

$$f_j := v_j - u_j = v_j - \langle e_1, v_j \rangle e_1 - \cdots - \langle e_{j-1}, v_j \rangle e_{j-1} \in W_{j-1}^{\perp}.$$

因此 $v_j = u_j + f_j$ 是将 v_j 按照正交分解 $V = W_{j-1} \perp W_{j-1}^{\perp}$ 来分拆向量 v_j 所得的结果. 这里的 u_j 称为 v_j 在子空间 W_{j-1} 上的**正交投影** (orthogonal projection). 而根据构造, $f_j = v_j - u_j$ 与向量 $e_1, \cdots, e_{j-1}$ 全都正交. 向量 e_j 则只不过是 f_j 的单位化, 即 $e_j = \pm\|f_j\|^{-1} f_j$. 有兴趣的读者不妨自己画个草图来帮助理解上述内容. ■

例 7.1.19. 再来看例 7.1.15 中的内积空间. 这里 $V = \mathbb{R}[X]_{\leq 2}$, 其上的内积取为 $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\mathrm{d}x$.

现在我们用 Gram–Schmidt 正交化方法来从 V 的基 $v_1 = 1$, $v_2 = X$, $v_3 = X^2$ 构造出一组规范正交基. 我们可以直接套用 (7.1.16.2) 中的公式. 首先

$$e_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\left(\int_{-1}^{1} 1^2 \mathrm{d}x\right)^{1/2}} v_1 = \frac{1}{\sqrt{2}} v_1 = \frac{1}{\sqrt{2}}.$$

其次

$$v_2 - \langle e_1, v_2 \rangle e_1 = X - \left(\int_{-1}^{1} x \frac{1}{\sqrt{2}} \mathrm{d}x\right) \cdot \frac{1}{\sqrt{2}} = X$$

故

$$e_2 = \frac{v_2 - \langle e_1, v_2 \rangle e_1}{\|v_2 - \langle e_1, v_2 \rangle e_1\|} = \frac{X}{\|X\|} = \frac{1}{\left(\int_{-1}^{1} x^2 \mathrm{d}x\right)^{1/2}} X = \frac{\sqrt{3}}{\sqrt{2}} X.$$

接下来计算

$$\begin{aligned} v_3 - \langle e_1, v_3 \rangle e_1 - \langle e_2, v_3 \rangle e_2 &= v_3 - \left(\int_{-1}^{1} x^2 \frac{1}{\sqrt{2}} \mathrm{d}x\right) e_1 - \left(\int_{-1}^{1} x^2 \frac{\sqrt{3}}{\sqrt{2}} \mathrm{d}x\right) e_2 \\ &= v_3 - \frac{2}{3\sqrt{2}} e_1 = X^2 - \frac{1}{3}. \end{aligned}$$

于是

$$\|v_3 - \langle e_1, v_3 \rangle e_1 - \langle e_2, v_3 \rangle e_2\| = \left(\int_{-1}^{1} \left(x^2 - \frac{1}{3}\right)^2 \mathrm{d}x\right)^{1/2} = \frac{\sqrt{8}}{\sqrt{45}}$$

而

$$e_3 = \frac{v_3 - \langle e_1, v_3 \rangle e_1 - \langle e_2, v_3 \rangle e_2}{\|v_3 - \langle e_1, v_3 \rangle e_1 - \langle e_2, v_3 \rangle e_2\|} = \frac{\sqrt{45}}{\sqrt{8}} \left(X^2 - \frac{1}{3}\right) = \frac{\sqrt{5}}{\sqrt{8}} (3X^2 - 1).$$

以上得到的 e_1, e_2, e_3 就是将 $1, X, X^2$ 这组基正交化得到的规范正交基. 虽然这个结果和我们在例 7.1.15 中得到的结果恰好相同, 但这绝不意味着当前内积空间的规范正交基仅有这一种可能. ■

Gram–Schmidt 正交化方法的意义不仅在于给出了一种构造规范正交基的方法, 更在于它还给出了额外的有用信息. 事实上, 我们只需要考虑 $\mathbb{R}^n$ 的标准内积就可以得到以下关于矩阵的有用结论:

定理 7.1.20 (**QR 分解** QR decomposition/factorization): *对于任意列满秩的实矩阵 $A \in \mathbf{M}_{n\times m}(\mathbb{R})$, 存在唯一一对矩阵 $Q \in \mathbf{M}_{n\times m}(\mathbb{R})$ 和 $R \in \mathbf{M}_m(\mathbb{R})$ 满足*

$$A = QR, \quad \text{且 } Q \text{ 的列向量组是 } \mathbb{R}^n \text{ 中的规范正交组, } R \text{ 是对角线元素均为正数的上三角阵.}$$

这样的表达式称为 A 的 QR **分解**.¶

证明. 设 A 的列向量组为 $v_1, \cdots, v_m \in \mathbb{R}^n$. 根据定理 7.1.16 可知, 存在规范正交组 $e_1, \cdots, e_m$ 满足 (7.1.16.2) 式. 这说明存在常数 r_{ij}, $1 \le i \le j \le m$ 使得

$$\begin{aligned} v_1 &= r_{11}e_1 \\ v_2 &= r_{12}e_1 + r_{22}e_2 \\ &\cdots\cdots\cdots\cdots\cdots \\ v_m &= r_{1m}e_1 + \cdots + r_{mm}e_m \end{aligned} \tag{7.1.20.1}$$

此外, 根据 (7.1.16.2) 式, 这里的每个系数 $r_{11}, \cdots, r_{mm}$ 都只有绝对值相同的一正一负两种选择. 如果要求 r_{ii}, $1 \le i \le m$ 都是正数, 那么 r_{ii} 的选择就是唯一的. 因此, 若以 $e_1, \cdots, e_m \in \mathbb{R}^n$ 为列向量组构成矩阵 $Q \in \mathbf{M}_{n\times m}(\mathbb{R})$, 以 R 表示以上常数构成的上三角阵 $(r_{ij}) \in \mathbf{M}_n(\mathbb{R})$ (当 $i > j$ 时认为 $r_{ij} = 0$), 则

$$A = (v_1, \cdots, v_m) = (e_1, \cdots, e_m)R = QR.$$

这就证明了 QR 分解的存在性.

当 $A = QR$ 是一个 QR 分解时, 矩阵 Q 的列向量组 $e_1, \cdots, e_m$ 满足 (7.1.20.1) 式. 而根据定理 7.1.16 中 (7.1.16.2) 式给出的限制, 由 r_{ii} 均大于零这个条件可知 Q 的列向量组 $e_1, \cdots, e_m$ 是唯一的. 换言之, 矩阵 Q 是由 A 唯一确定的. 也就是说, 如果 $A = Q_1R_1$ 是 A 的另一种 QR 分解, 那么 $Q_1 = Q$. 由于 A 有左逆, R 可逆, 故 $Q = AR^{-1}$ 有左逆. 取定 Q 的一个左逆 M 即可由 $QR = A = QR_1$ 推出 $R = MQR = MA = MQR_1 = R_1$. 所以 R 也是由 A 唯一确定的. 这就证明了 QR 分解的唯一性. □

注意, 如果 A 是一个列满秩的方阵, 那么其 QR 分解中的矩阵 Q 也是可逆阵.

例 7.1.21. 举一个具体的例子来看矩阵的 QR 分解如何计算. 令 $A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & -1 \end{pmatrix}$. 即 A 的列向量组由 $\mathbb{R}^4$ 中的如下三个向量给出

$$v_1 = (0, 1, 0, 1)^T, \ v_2 = (0, -1, 1, -1)^T, \ v_3 = (1, 0, 0, -1)^T.$$

¶这一分解被称为 QR 分解的理由大概如下: 首先这里的矩阵 Q 如果是方阵, 那么它是所谓的正交矩阵. 而 orthogonal 这一英文单词的首字母 O 不太适合在代数学中用作首字母缩写, 因为它和 0 太相似. 与 O 临近的字母 P 又常常用于矩阵相似时使用的过渡矩阵, 因此人们常用 Q 表示一个正交矩阵. 此外, 此处的矩阵 R 是个上三角矩阵, 但是 upper triangular 的首字母 U 在 LU 分解这个概念中已经用过了. 一个矩阵 A 可能既有 LU 分解又有 QR 分解, 这两个分解中出现的上三角阵 U 和 R 可能完全不同. 因此, 此处我们改用 right 的首字母 R 表示一个上三角阵. 这样做的一个理由是上三角阵的非零元素也可以认为是处于对角线的右边.

根据 Gram–Schmidt 正交化方法, 令

$$
\begin{aligned}
e_1 &= \frac{1}{\|v_1\|}v_1 = \frac{1}{\sqrt{2}}v_1 = \left(0\,,\,\frac{1}{\sqrt{2}}\,,\,0\,,\,\frac{1}{\sqrt{2}}\right)^T\,,\\
f_2 &= v_2 - \langle\, e_1,\, v_2\,\rangle e_1 = v_2 - \frac{1}{\sqrt{2}}\langle\, v_2,\, v_1\,\rangle e_1 = v_2 + v_1 = (0,\,0,\,1,\,0)^T\,,\\
e_2 &= \frac{1}{\|f_2\|}f_2 = v_2 + v_1 = (0,\,0,\,1,\,0)^T\,,\\
f_3 &= v_3 - \langle\, e_1,\, v_3\,\rangle e_1 - \langle\, e_2,\, v_3\,\rangle e_2 = v_3 + \frac{1}{2}v_1 = \left(1\,,\,\frac{1}{2}\,,\,0\,,\,-\frac{1}{2}\right)^T\,,\\
e_3 &= \frac{1}{\|f_3\|}f_3 = \left(\frac{\sqrt{2}}{\sqrt{3}}\,,\,\frac{1}{\sqrt{6}}\,,\,0\,,\,-\frac{1}{\sqrt{6}}\right)^T = \frac{\sqrt{2}}{\sqrt{3}}v_3 + \frac{\sqrt{1}}{\sqrt{6}}v_1\,.
\end{aligned}
$$

根据以上计算结果总结出 $v_1,\,v_2,\,v_3$ 与 $e_1,\,e_2,\,e_3$ 的关系为

$$
(v_1,\,v_2,\,v_3) = (e_1,\,e_2,\,e_3)\begin{pmatrix} \sqrt{2} & -\sqrt{2} & -\frac{\sqrt{2}}{2} \\ & 1 & 0 \\ & & \frac{\sqrt{3}}{\sqrt{2}} \end{pmatrix}\,.
$$

因此, 若令

$$
Q = (e_1,\,e_2,\,e_3) = \begin{pmatrix} 0 & 0 & \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{6}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{6}} \end{pmatrix}\,,\quad R = \begin{pmatrix} \sqrt{2} & -\sqrt{2} & -\frac{\sqrt{2}}{2} \\ & 1 & 0 \\ & & \frac{\sqrt{3}}{\sqrt{2}} \end{pmatrix}
$$

则 $A = QR$ 即为所求的 QR 分解. ■

注记 7.1.22. QR 分解在数值线性代数 (计算数学的重要内容之一) 中具有非常重要的应用. 事实上, 矩阵的特征值计算是数值计算的一个重要课题. 数值计算专家们发现, 使用 QR 分解可以得到一种非常有效的特征值计算方法, 称为 QR **算法** (QR algorithm). 这种算法的基本思路如下:

设 A 为给定的实方阵. 若 A 不可逆, 可以适当选取实数 λ 使 $A' := A + \lambda I$ 可逆. 如果能求出 A' 的所有特征值, 那么这些特征值减去 λ 就可以得到 A 的特征值. 因此我们不妨设 A 可逆. 这样我们可以执行以下算法: 先令 $A_0 = A$, 将 A_0 做 QR 分解 $A_0 = Q_0R_0$, 再令 $A_1 = R_0Q_0$. 因为 $Q_0^{-1}A_0Q_0 = Q_0^{-1}Q_0R_0Q_0 = A_1$, 所以 A_0 与 A_1 相似. 因此 A_0 和 A_1 的特征值连同重数都对应相同. 接下来再以

$$
A_k = Q_kR_k \Longrightarrow A_{k+1} := R_kQ_k
$$

的方式迭代产生矩阵序列 $(A_k)_{k\geq 0}$. 在理想情况下, 矩阵序列 A_k 会越来越接近于 (或者准确来说, 收敛于) 一个上三角阵, 而上三角阵的特征值可以直接从对角线上元素得到. 这就是 QR 算法的基本思路.

当然在实际计算中, 计算数学家们还要设计改进的算法, 使得算法的收敛速度更快. 这是计算数学的核心思想之一. 有兴趣的读者未来可以在数值计算类的课程中深入学习这些内容. ■

7.1.3 正交矩阵与正交相似

与规范正交基密切联系的一个矩阵概念是所谓的正交矩阵.

以列向量空间 $\mathbb{R}^{n\times 1}$ 的标准内积为例. 假设 $v_1,\cdots,v_n$ 是 $\mathbb{R}^{n\times 1}$ 中的一组向量, Q 是以 $v_1,\cdots,v_n$ 为列向量组得到的 n 阶方阵. 那么

$$Q^TQ=\begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix}\cdot(v_1,\cdots,v_n)=\begin{pmatrix} v_1^Tv_1 & v_1^Tv_2 & \cdots & v_1^Tv_n \\ v_2^Tv_1 & v_2^Tv_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ v_n^Tv_1 & \cdots & \cdots & v_n^Tv_n \end{pmatrix}$$
$$=\begin{pmatrix} \langle v_1, v_1\rangle & \langle v_1, v_2\rangle & \cdots & \langle v_1, v_n\rangle \\ \langle v_2, v_1\rangle & \langle v_2, v_2\rangle & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \langle v_n, v_1\rangle & \cdots & \cdots & \langle v_n, v_n\rangle \end{pmatrix}.$$

因此,

$$Q^TQ=I_n \iff Q \text{ 的列向量组 } v_1,\cdots,v_n \text{ 是 } \mathbb{R}^{n\times 1} \text{ 中的规范正交基.}$$

注意到 $Q^TQ=I_n$ 和 $QQ^T=I_n$ 是等价的. 通过类似的讨论可以看出, 这一性质也等价于说 Q 的行向量组 $\alpha_1,\cdots,\alpha_n\in\mathbb{R}^{1\times n}$ 构成行向量空间 $\mathbb{R}^{1\times n}$ (关于标准内积) 的一组规范正交基.

定义 7.1.23. 设 Q 为 n 阶方阵. 如果 $Q^TQ=I_n$, 则称 Q 是一个**正交矩阵** (orthogonal matrix). 对于一个域 K, 我们用 $\mathbf{O}_n(K)$ 表示 $\mathbf{M}_n(K)$ 中所有正交矩阵构成的集合. ■

根据前面的讨论, 我们已经得到如下结论:

命题 7.1.24: 对于任意 n 阶方阵 Q, 以下论断等价:

1. Q 可逆且 $Q^{-1}=Q^T$.
2. $Q^TQ=I_n$.
3. $QQ^T=I_n$.

若 Q 为实方阵, 则以上条件也和下面的条件等价:

4. Q 的列向量组构成列向量空间 $\mathbb{R}^{n\times 1}$ 关于标准内积的一组规范正交基.
5. Q 的行向量组构成行向量空间 $\mathbb{R}^{1\times n}$ 关于标准内积的一组规范正交基.

思考题 7.3. 证明正交方阵集合 $\mathbf{O}_n(K)$ 具有以下性质:

1. $I_n\in\mathbf{O}_n(K)$.
2. 若 $A\in\mathbf{O}_n(K)$, 则 $\det(A)\in\{\pm1\}$, A 可逆并且 $A^{-1}\in\mathbf{O}_n(K)$.
   集合 $\{A\in\mathbf{O}_n(K)\,|\,\det(A)=1\}$ 通常记为 $\mathbf{SO}_n(K)$, 其中元素称为**特殊正交矩阵** (special orthogonal matrix).
3. 若 $A, B\in\mathbf{O}_n(K)$, 则 $AB\in\mathbf{O}_n(K)$.

思考题 7.4. 考虑二阶实正交方阵集合 $\mathbf{O}_2(\mathbb{R})$, 其中的子集 $\mathbf{SO}_2(\mathbb{R})$ 如思考题 7.3 中定义.
证明:

1. 对于任意 $A\in\mathbf{M}_2(\mathbb{R})$, 下列陈述等价:

(a) $A \in \mathbf{SO}_2(\mathbb{R})$.

(b) 存在唯一的 $\theta \in \mathbb{R}$, $0 \leq \theta < 2\pi$ 使得 $A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$.

2. 令 $J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. 则对任意 $A \in \mathbf{M}_2(\mathbb{R})$ 下列陈述等价:

(a) $A \in \mathbf{O}_2(\mathbb{R})$ 且 $\det(A) = -1$.

(b) 存在唯一的 $\theta \in [0, 2\pi[:= \{x \in \mathbb{R} \,|\, 0 \leq x < 2\pi\}$ 使得 $A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$.

(c) $AJ \in \mathbf{SO}_2(\mathbb{R})$.

(d) $JA \in \mathbf{SO}_2(\mathbb{R})$.

3. 对于任意 $A \in \mathbf{SO}_2(\mathbb{R})$, 均有 $J^{-1}AJ \in \mathbf{SO}_2(\mathbb{R})$. 而且, 映射

$$\mathbf{SO}_2(\mathbb{R}) \longrightarrow \mathbf{SO}_2(\mathbb{R})\,; \quad M \longmapsto J^{-1}MJ$$

是双射.

上面我们提到实正交矩阵的列向量组构成 $\mathbb{R}^n$ 中的规范正交基. 对于一般的实内积空间, 正交矩阵和规范正交基的对应关系可以用下面的命题描述.

命题 7.1.25: 取定 n 维实内积空间 $(V, \langle\cdot\,,\,\cdot\rangle)$ 的一组有序规范正交基 $\mathcal{E} = (e_1, \cdots, e_n)$.

1. 对于 V 的任意一组有序基 $\mathcal{B} = (v_1, \cdots, v_n)$, 下列陈述等价:

(a) $\mathcal{B}$ 仍是一组规范正交基.

(b) 从 $\mathcal{E}$ 到 $\mathcal{B}$ 的过渡矩阵 $P_{\mathcal{E},\mathcal{B}}$ 是正交矩阵.

(c) 从 $\mathcal{B}$ 到 $\mathcal{E}$ 的过渡矩阵 $P_{\mathcal{B},\mathcal{E}}$ 是正交矩阵.

2. 将 V 的每一组有序规范正交基作为一个元素, 将所有这些元素构成的集合记为 $\mathrm{OB}(V)$. 则通过对应法则

$$\psi_{\mathcal{E}} \ :\ \mathbf{O}_n(\mathbb{R}) \longrightarrow \mathrm{OB}(V)\,; \quad Q \longmapsto \mathcal{E}\cdot Q$$

可以得到一个双射, 其逆映射是通过求过渡矩阵得到的如下映射

$$\phi_{\mathcal{E}} \ :\ \mathrm{OB}(V) \longrightarrow \mathbf{O}_n(\mathbb{R})\,; \quad \mathcal{B} \longmapsto P_{\mathcal{E},\mathcal{B}}\,.$$

证明. (1) 设 P 为 $\mathcal{E}$ 到 $\mathcal{B}$ 的过渡矩阵. 根据不同基下 Gram 矩阵的相合关系 (参见 (6.1.9.2)),

$$\mathcal{M}_{\mathcal{B}}(\langle\cdot\,,\,\cdot\rangle) = P^T\mathcal{M}_{\mathcal{E}}(\langle\cdot\,,\,\cdot\rangle)P\,.$$

根据规范正交基的定义可知, $\mathcal{B}$ 是规范正交基当且仅当 Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(\langle\cdot\,,\,\cdot\rangle)$ 是单位矩阵. (故 $\mathcal{M}_{\mathcal{E}}(\langle\cdot\,,\,\cdot\rangle) = I_n$.) 因此上式表明

$$\begin{aligned} \mathcal{B}\text{ 是规范正交基} &\iff \mathcal{M}_{\mathcal{B}}(\langle\cdot\,,\,\cdot\rangle)\text{ 是单位矩阵} \\ &\iff P^TP\text{ 是单位矩阵} \iff P\text{ 是正交矩阵.} \end{aligned}$$

这就证明了 (a)$\Leftrightarrow$(b). 条件 (c) 和 (a), (b) 的等价性留给读者自证.

(2) 这只不过是 (1) 中 (a) 和 (b) 等价性的抽象废话表述. 或者说, 由 (1) 中结论可知所给的两个映射 $\psi_{\mathcal{E}}$ 和 $\phi_{\mathcal{E}}$ 是合理定义的. 而根据过渡矩阵的定义即知二者互为逆映射. □

定义 7.1.26. 设 $A, A' \in \mathbf{M}_n(K)$. 如果存在正交矩阵 $Q \in \mathbf{O}_n(K)$ 使得 $Q^{-1}AQ = A'$ 则称 A 和 A' 在 K 上**正交相似** (orthogonally similar). 因为对于正交矩阵而言, 逆矩阵和转置矩阵是相同的, 所以正交相似也被称为**正交相合** (orthogonally congruent). (一般我们在讨论内积或更一般的双线性型的 Gram 矩阵时使用相合的说法, 而讨论线性变换在给定基下的矩阵时使用相似的说法.) ■

推论 7.1.27: 设 V 是 n 维实内积空间, $\mathscr{A} \in \mathrm{End}(V)$, $\mathcal{E}$ 是 V 的一组规范正交基, $A = \mathcal{M}_{\mathcal{E}}(\mathscr{A}) \in \mathbf{M}_n(\mathbb{R})$.

则对于任意 $A' \in \mathbf{M}_n(\mathbb{R})$, A 和 A' (在 $\mathbb{R}$ 上) 正交相似的充分必要条件是存在规范正交基 $\mathcal{B}$ 使得 $A' = \mathcal{M}_{\mathcal{B}}(\mathscr{A})$.

证明. 留给读者练习. □

命题 7.1.28: 设 V 是 n 维实内积空间, $\mathscr{A} \in \mathrm{End}(V)$. 如果存在 V 的一组有序基 $\mathcal{B}$ 使得 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是上三角阵, 那么一定存在一组规范正交基 $\mathcal{E}$ 使得 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 是上三角阵.

换用矩阵的语言来说, 如果一个方阵 $A \in \mathbf{M}_n(\mathbb{R})$ 实相似于一个上三角阵, 那么 A 也可以正交相似于一个 (可能不同的) 上三角阵.

证明. 为了帮助读者深入领会之前学过的内容, 我们以两种方式叙述证明. 其中一种采用线性变换的语言, 另一种采用矩阵的语言.

以线性变换的语言叙述: 在一组有序基 $\mathcal{B} = (v_1, \cdots, v_n)$ 下线性变换 $\mathscr{A}$ 的矩阵为上三角阵, 这等价于说 $\mathcal{B}$ 这组基满足如下条件:

$$\text{对于每个 } j \in [\![1, n]\!],\ \mathrm{span}(v_1, \cdots, v_j) \text{ 都是 } \mathscr{A} \text{ 的不变子空间.}$$

根据 Gram–Schmidt 正交化 (定理 7.1.16), 有序基 $\mathcal{B}$ 总是可以转化为一组规范正交基 $\mathcal{E} = (e_1, \cdots, e_n)$ 使得

$$\mathrm{span}(v_1, \cdots, v_j) = \mathrm{span}(e_1, \cdots, e_j) \quad \text{对每个 } j \in [\![1, n]\!] \text{ 成立.}$$

因此, 当 $\mathrm{span}(v_1, \cdots, v_j)$ 都是 $\mathscr{A}$ 的不变子空间时, 以上等式当然说明每个 $\mathrm{span}(e_1, \cdots, e_j)$ 也都是 $\mathscr{A}$ 的不变子空间, 即 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 也是上三角阵.

以矩阵语言叙述: 假设 $P \in \mathbf{M}_n(\mathbb{R})$ 是可逆矩阵使得 $P^{-1}AP =: A'$ 是上三角阵. 根据矩阵的 QR 分解 (定理 7.1.20), 存在正交矩阵 Q 和可逆上三角阵 R 使得 $P = QR$. 于是 $A' = P^{-1}AP = R^{-1}(Q^{-1}AQ)R$, 亦即 $Q^{-1}AQ = RA'R^{-1}$. 注意 R 的逆矩阵 R^{-1} 也是上三角阵. 乘积 $RA'R^{-1}$ 中三个因子矩阵 R, A', R^{-1} 都是上三角阵, 故 $RA'R^{-1}$ 也是上三角阵. 结论证毕. □

再次提醒读者注意: 命题 7.1.28 并不是说如果 $A \in \mathbf{M}_n(\mathbb{R})$ 和一个上三角阵 $U \in \mathbf{M}_n(\mathbb{R})$ 相似, 那么 A 和 U 也正交相似. 该命题只是说 A 可以正交相似于某个上三角阵 (但该上三角阵可能不是刚才选取的 U).

7.1.4 习题

习题 7.1.1. 设 $V = \mathbb{R}[X]_{\leq 2}$. 定义 V 上的双线性型

$$\langle f, g \rangle := f(0)g(0)\,, \quad f, g \in V = \mathbb{R}[X]_{\leq 2}\,.$$

该双线性型是否是 V 上的内积？为什么？

习题 7.1.2. 设 $(V_1, \langle \cdot\,, \cdot \rangle_1)$ 和 $(V_2, \langle \cdot\,, \cdot \rangle_2)$ 均为内积空间. 设 $V = V_1 \times V_2$. 定义

$$\langle \cdot\,, \cdot \rangle\ :\ V \times V \longrightarrow \mathbb{R}$$

为

$$\big\langle (u_1,\,u_2)\,,\,(v_1,\,v_2)\big\rangle := \langle\, u_1,\,v_1\,\rangle_1 + \langle\, u_2\,,\,v_2\,\rangle_2\,.$$

证明 $\langle\cdot\,,\,\cdot\rangle$ 是 V 上的一个内积.

习题 7.1.3. 设 $(V,\langle\cdot\,,\,\cdot\rangle)$ 是实内积空间, $u,\,v\in V$.

1. 证明: 若 u 和 v 长度相同, 则 $u+v$ 与 $u-v$ 正交.

2. 证明菱形的两条对角线互相垂直.

习题 7.1.4. 设 $(V,\langle\cdot\,,\,\cdot\rangle)$ 是实内积空间, $u,\,v\in V$. 证明下列条件等价:

1. u 与 v 正交.

2. 对所有 $a\in\mathbb{R}$ 均有 $\|u\|\le\|u+av\|$.

习题 7.1.5. 设 $(V,\langle\cdot\,,\,\cdot\rangle)$ 是实内积空间, $u,\,v\in V$. 证明: 若 $\|u\|\le 1$, $\|v\|\le 1$, 则

$$\sqrt{1-\|u\|^2}\cdot\sqrt{1-\|v\|^2}\le 1-|\langle\, u\,,\,v\,\rangle|\,.$$

习题 7.1.6. 设 $n\in\mathbb{N}^*$. 证明:

1. 对于任意实数 $a_1,\cdots,a_n$, 均有

$$\left(\frac{a_1+\cdots+a_n}{n}\right)^2\le\frac{a_1^2+\cdots+a_n^2}{n}\,.$$

2. n 元实二次型 $q=n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2$ 是半正定的.

习题 7.1.7. 设 $V=\mathbf{M}_n(\mathbb{R})$. 定义 V 上的双线性型如下:

$$\langle\cdot,\cdot\rangle\,:\,V\times V\longrightarrow\mathbb{R}\,;\quad (A,\,B)\longmapsto\langle\, A,\,B\,\rangle := \mathrm{Tr}(A^TB)\,.$$

证明这个双线性型是 V 上的一个内积. 该内积称为 $\mathbf{M}_n(\mathbb{R})$ 上的 Frobenius[‖]**内积**.

习题 7.1.8. 设 $u,\,v$ 是内积空间 V 中的向量.
证明: $\|u\|=\|v\|$ 当且仅当对任意 $a,\,b\in\mathbb{R}$ 均有 $\|au+bv\|=\|bu+av\|$.

习题 7.1.9. 在内积空间 $\mathbb{R}^4$ 中求向量 $\alpha,\,\beta$ 的夹角, 其中

$$\alpha=(1,\,2,\,2,\,3)^T\;,\;\beta=(3,\,1,\,5,\,1)^T\;.$$

习题 7.1.10. 在内积空间 $\mathbb{R}^4$ 中求一个单位向量与 $\alpha,\,\beta,\,\gamma$ 均正交, 其中

$$\alpha=(1,\,1,\,-1,\,1)^T\;,\;\beta=(1,\,-1,\,-1,\,1)^T\,,\;\gamma=(2,\,1,\,1,\,3)^T\;.$$

习题 7.1.11. 假设 $\varepsilon_1,\,\varepsilon_2,\,\varepsilon_3$ 是三维内积空间 V 中的一组规范正交基. 令

$$\eta_1=\frac{1}{3}(2\varepsilon_1+2\varepsilon_2-\varepsilon_3)\;,\;\eta_2=\frac{1}{3}(2\varepsilon_1-\varepsilon_2+2\varepsilon_3)\;,\;\eta_3=\frac{1}{3}(\varepsilon_1-2\varepsilon_2-2\varepsilon_3)\;.$$

证明: $\eta_1,\,\eta_2,\,\eta_3$ 也是 V 的一组规范正交基.

[‖]Ferdinand Georg Frobenius (1849–1917), 德国数学家.

习题 7.1.12. 假设 $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5$ 是五维内积空间 V 中的一组规范正交基. 令

$$\alpha_1 = \varepsilon_1 + \varepsilon_3 \ , \ \alpha_2 = \varepsilon_1 - \varepsilon_2 + \varepsilon_4 \ , \ \alpha_3 = 2\varepsilon_1 + \varepsilon_2 + \varepsilon_3 \ .$$

求子空间 $U := \operatorname{span}(\alpha_1, \alpha_2, \alpha_3)$ 的一组规范正交基.

习题 7.1.13. 设 V 是线性方程组

$$\begin{aligned} 2x_1 + x_2 - x_3 + x_4 - 3x_5 &= 0 \\ x_1 + x_2 - x_3 + x_5 &= 0 \end{aligned}$$

在 $\mathbb{R}^5$ 中的解空间. 将 $\mathbb{R}^5$ 的标准内积限制于 V 使之成为内积空间. 求该内积空间的一组规范正交基.

习题 7.1.14. 设 $\alpha_1, \cdots, \alpha_r$ 是内积空间 V 中的一组向量. 令 $A = (a_{ij}) \in \mathbf{M}_r(\mathbb{R})$, 其中 $a_{ij} = \langle \alpha_i, \alpha_j \rangle$.

证明: 向量组 $\alpha_1, \cdots, \alpha_r$ 线性无关当且仅当 $\det(A) \neq 0$.

习题 7.1.15. 设 $\mathcal{C}[-\pi, \pi]$ 为闭区间 $[-\pi, \pi]$ 上的所有实值连续函数构成的向量空间, 在其上定义内积

$$\langle f, g \rangle := \int_{-\pi}^{\pi} f(x) g(x) \, \mathrm{d}x \ .$$

将函数组

$$1 \, , \cos(x) \, , \sin(x) \, , \cos(2x) \, , \sin(2x) \, , \cdots \, , \cos(nx) \, , \sin(nx) \, .$$

视为 $\mathcal{C}[-\pi, \pi]$ 中的一组向量.

1. 证明以上函数组作为 $\mathcal{C}[-\pi, \pi]$ 中的向量是两两正交的.
2. 设 V 是以上函数组在 $\mathcal{C}[-\pi, \pi]$ 中张成的子空间. 求 V 的一组规范正交基.

习题 7.1.16. 给定 4 维列向量

$$\alpha_1 = \left(\frac{1}{3} \, , -\frac{2}{3} \, , 0 \, , \frac{2}{3} \right)^T \ , \quad \alpha_2 = \left(-\frac{2}{\sqrt{6}} \, , 0 \, , \frac{1}{\sqrt{6}} \, , \frac{1}{\sqrt{6}} \right)^T \ .$$

请给出一个正交矩阵 $Q \in \mathbf{M}_4(\mathbb{R})$, 它的前两列为 α_1, α_2.

习题 7.1.17. 假设 $A \in \mathbf{M}_n(\mathbb{R})$ 既是正交矩阵又是上三角阵. 证明: A 必然是对角阵, 而且它的对角线元素均为 ± 1.

习题 7.1.18. 设 $A \in \mathbf{M}_n(\mathbb{R})$ 是正定矩阵. 证明: 存在上三角阵 R 使得 $A = R^T R$.

习题 7.1.19. 设 V 是线性方程

$$x_1 - x_2 + x_3 - x_4 + \cdots + x_{2n-1} - x_{2n} = 0$$

在 $\mathbb{R}^{2n}$ 中的解空间. 将 $\mathbb{R}^{2n}$ 的标准内积限制于 V 给出 V 上的内积.

求 V (相对于所给内积)的一组规范正交基.

习题 7.1.20. 设

$$J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ , \ A_1 = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \ , \ A_2 = \begin{pmatrix} \cos\theta_2 & \sin\theta_2 \\ \sin\theta_2 & -\cos\theta_2 \end{pmatrix}$$

其中 $\theta_1, \theta_2 \in \mathbb{R}$. 记 $B := J A_1 A_2$.

证明 $B \in \mathbf{SO}_2(\mathbb{R})$, 并求 $\theta \in \mathbb{R}$ 使得 $B = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$.

习题 7.1.21. 求以下矩阵的 QR 分解:

$$\begin{pmatrix} 3 & -1 & 2 \\ 0 & 0 & 9 \\ 4 & 7 & 11 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 2 & 4 \\ 0 & 1 & 1 & 3 \end{pmatrix}.$$

习题 7.1.22. 假设内积空间 V 上的一个线性变换 $\mathscr{A}$ 满足: 对于所有 $v \in V$, 均有 $\|\mathscr{A}v\| \leq \|v\|$.

证明 $\mathscr{A} + 3I$ 是单射.

7.2 内积空间上的线性变换

本节中 V 和 W 总是表示非零的有限维实内积空间, 它们的内积记为 $\langle \cdot\,,\,\cdot \rangle_V$ 和 $\langle \cdot\,,\,\cdot \rangle_W$, 或者在不至于引起歧义时都简单记作 $\langle \cdot\,,\,\cdot \rangle$. 类似地, V 中向量 v 的长度在需要强调时记为 $\|v\|_V$, 否则可简记为 $\|v\|$.

7.2.1 等距映射和正交变换

我们曾经提过, 内积为向量空间带来距离和长度的概念. 保持映射前后距离和长度不变的线性映射因此也值得特别注意.

定义 7.2.1. 设 $\mathscr{A} : V \to W$ 为 (有限维内积空间之间的) 线性映射. (对于 $v \in V$, 将 $\mathscr{A}(v)$ 简记为 $\mathscr{A}v$.)

如果对于所有的 $u, v \in V$ 均有 $\|\mathscr{A}u - \mathscr{A}v\|_W = \|u - v\|_V$, 则称 $\mathscr{A}$ 是一个**等距映射** (isometric map) 或**等距嵌入** (isometric embedding).

如果线性映射 $\mathscr{A}$ 既是等距映射又是双射, 则称 $\mathscr{A}$ 是一个**等距同构** (isometry). 当这样的线性映射存在时, 我们称内积空间 V 和 W **等距同构** (be isometric). ■

思考题 7.5. 证明: 如果线性映射 $\mathscr{A} : V \to W$ 是等距映射, 则 $\mathscr{A}$ 是单射. 如果 $\mathscr{A}$ 是等距同构, 则逆映射 $\mathscr{A}^{-1} : W \to V$ 也是等距同构.

命题 7.2.2: 设 $\mathscr{A} : V \to W$ 为 (有限维内积空间之间的) 线性映射. 则下列陈述等价:

1. $\mathscr{A}$ 是等距映射.
2. $\mathscr{A}$ 保持向量长度不变, 即, 对于任意 $v \in V$ 均有 $\|\mathscr{A}v\|_W = \|v\|_V$.
3. $\mathscr{A}$ 保持向量内积不变, 即, 对于任意 $u, v \in V$ 均有 $\langle \mathscr{A}u\,,\, \mathscr{A}v \rangle_W = \langle u\,,\, v \rangle_V$.
4. $\mathscr{A}$ 将任意规范正交向量组映射为规范正交向量组, 即, 对于 V 中任意的规范正交向量组 $e_1, \cdots, e_m$, W 中的向量组 $\mathscr{A}e_1, \cdots, \mathscr{A}e_m$ 也是规范正交组.
5. $\mathscr{A}$ 将 V 的任意一组规范正交基映射为规范正交向量组,
6. 存在 V 的一组规范正交基 $e_1, \cdots, e_n$ 使得 $\mathscr{A}e_1, \cdots, \mathscr{A}e_n$ 是 W 中的规范正交组.
7. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 和 W 的一组有序规范正交基 $\mathcal{C}$, 使得 $\mathscr{A}$ 在 $\mathcal{B}$ 和 $\mathcal{C}$ 下的矩阵 $\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 具有规范正交的列向量组.
8. 对于 V 的任意一组有序规范正交基 $\mathcal{B}$ 和 W 的任意一组有序规范正交基 $\mathcal{C}$, $\mathscr{A}$ 在 $\mathcal{B}$ 和 $\mathcal{C}$ 下的矩阵 $\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 具有规范正交的列向量组.

9. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 和 W 的一组有序规范正交基 $\mathcal{C}$, 使得 $\mathscr{A}$ 在 $\mathcal{B}$ 和 $\mathcal{C}$ 下的矩阵 $A=\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 满足 $A^TA=I_n$, 其中 $n=\dim V$.

10. 对于 V 的任意一组有序规范正交基 $\mathcal{B}$ 和 W 的任意一组有序规范正交基 $\mathcal{C}$, $\mathscr{A}$ 在 $\mathcal{B}$ 和 $\mathcal{C}$ 下的矩阵 $A=\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 满足 $A^TA=I_n$, 其中 $n=\dim V$.

证明. (1) $\Rightarrow$ (2). 在定义 7.2.1 中取 $u=0$ 即可.

(2) $\Rightarrow$ (1). 对于任意 $u,v\in V$, 令 $w=u-v$ 可得 $\mathscr{A}u-\mathscr{A}v=\mathscr{A}w$. 由此易见, (2) $\Rightarrow$ (1).

(2) $\Rightarrow$ (3). 利用极化恒等式 (思考题 7.1 (2)) 易得.

(3) $\Rightarrow$ (4). 一个向量组 $e_1,\cdots,e_m$ 是规范正交组的等价描述是矩阵 $(\langle e_i\,,e_j\rangle)$ 为单位矩阵. 条件 (3) 说明 $(\langle e_i\,,e_j\rangle)=(\langle \mathscr{A}e_i\,,\mathscr{A}e_j\rangle)$. 因此 (3) $\Rightarrow$ (4).

(4) $\Rightarrow$ (5). 显然.

(5) $\Rightarrow$ (6). 由于规范正交基总存在, 故显然 (5) $\Rightarrow$ (6).

类似地, (8) $\Rightarrow$ (7), (10) $\Rightarrow$ (9).

(6) $\Rightarrow$ (7). 设 $\mathcal{B}=(e_1,\cdots,e_n)$ 为 V 中的规范正交组, 使得 $f_1:=\mathscr{A}e_1,\cdots,f_n:=\mathscr{A}e_n$ 构成 W 中的规范正交组. 根据命题 7.1.14, 可以将 $f_1,\cdots,f_n$ 扩充为 W 的一组规范正交基 $\mathcal{C}:=(f_1,\cdots,f_n,\cdots,f_m)$. 此时 $\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})=\binom{I_n}{0}_{m\times n}$. 故显然 $\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 的各列构成规范正交组.

一个矩阵 $M\in\mathbf{M}_{m\times n}(\mathbb{R})$ 各列规范正交的等价描述就是 $M^TM=I_n$. 因此 (7) $\Leftrightarrow$ (9), (8) $\Leftrightarrow$ (10).

(9) $\Rightarrow$ (10). 设 $\mathcal{B}$ 和 $\mathcal{C}$ 分别是 V 和 W 的一组规范正交基, 使得 $A:=\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 满足 $A^TA=I_n$ (亦即, A 的列向量组为规范正交组). 现在考虑 V 和 W 的任意规范正交基 $\mathcal{B}'$ 和 $\mathcal{C}'$. 根据命题 7.1.25, 相应的过渡矩阵 $P=P_{\mathcal{B},\mathcal{B}'}$ 和 $Q=P_{\mathcal{C},\mathcal{C}'}$ 都是正交矩阵. (只不过 P 的阶是 $n=\dim V$, 而 Q 的阶是 $m=\dim W$.) 在 $\mathcal{B}'$ 和 $\mathcal{C}'$ 下 $\mathscr{A}$ 的矩阵 $A'=\mathcal{M}_{\mathcal{B}',\mathcal{C}'}(\mathscr{A})$ 和 $A=\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 之间的关系为 $A'=Q^{-1}AP=Q^TAP$ (参见本书上册命题 3.3.1). 于是,

$$(A')^TA'=(P^TA^TQ)(Q^{-1}AP)=P^TA^T(QQ^{-1})AP=P^T(A^TA)P=P^TP=I_n\,.$$

(9) $\Rightarrow$ (2). 取定 V 的规范正交基 $\mathcal{B}$ 和 W 的规范正交基 $\mathcal{C}$ 使得 $A=\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 满足 $A^TA=I_n$. 则对于任意 $v\in V$,

$$\begin{aligned}\|\mathscr{A}v\|_W^2&=\mathcal{M}_{\mathcal{C}}(\mathscr{A}v)^T\cdot\mathcal{M}_{\mathcal{C}}(\mathscr{A}v)=\big(A\cdot\mathcal{M}_{\mathcal{B}}(v)\big)^T\cdot(A\cdot\mathcal{M}_{\mathcal{B}}(v))\\&=\mathcal{M}_{\mathcal{B}}(v)^T(A^TA)\mathcal{M}_{\mathcal{B}}(v)=\mathcal{M}_{\mathcal{B}}(v)^T\mathcal{M}_{\mathcal{B}}(v)=\|v\|_V^2\,.\end{aligned}$$

命题至此证毕. □

推论 7.2.3: 维数相同的有限维实内积空间之间一定存在等距同构.

证明. 设 V 和 W 都是 n 维实内积空间. 分别为 V 和 W 取定一组规范正交基 $e_1,\cdots,e_n$ 和 $f_1,\cdots,f_n$. 则存在 (唯一的) 线性 (同构) 映射 $\mathscr{A}:V\to W$ 满足 $\mathscr{A}e_i=f_i$, $i=1,\cdots,n$. 根据命题 7.2.2 中条件 (1) 和 (6) 的等价性, $\mathscr{A}$ 是一个等距映射. 结论得证. □

定义 7.2.4. 设 $\mathscr{A}\in\mathrm{End}(V)$ 是 (有限维内积空间) V 上的线性变换. 如果 $\mathscr{A}$ 是个等距映射, 我们也称 $\mathscr{A}$ 是 V 上的一个 (线性) **正交算子** (orthogonal operator) 或 (线性) **正交变换** (orthogonal transformation). 根据思考题 7.5, 此时 $\mathscr{A}$ 一定是单射, 故而是双射. 因此, V 上的正交变换就是 V 到自身的等距同构.

我们用 $\mathbf{O}(V)$ 来表示 V 上所有正交变换构成的集合. ■

使用以上定义即可从命题 7.2.2 得到以下推论:

推论 7.2.5: 设 $\mathscr{A}\in\mathrm{End}(V)$. 则下列陈述等价:

1. $\mathscr{A}$ 是正交变换.
2. $\mathscr{A}$ 保持向量长度不变.
3. $\mathscr{A}$ 保持向量内积不变.
4. $\mathscr{A}$ 将任意规范正交向量组映射为规范正交向量组.
5. $\mathscr{A}$ 将 V 的任意一组规范正交基映射为规范正交基.
6. $\mathscr{A}$ 将 V 的某一组规范正交基映射为规范正交基.
7. 存在 V 的一组有序规范正交基 $\mathcal{B}$, 使得 $\mathscr{A}$ 在 $\mathcal{B}$ 下的矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是正交矩阵.
8. 对于 V 的任意一组有序规范正交基 $\mathcal{B}$, $\mathscr{A}$ 在 $\mathcal{B}$ 下的矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是正交矩阵.
9. 存在 V 的两组 (相同或不同的) 有序规范正交基 $\mathcal{B}$ 和 $\mathcal{C}$, 使得 $\mathscr{A}$ 在 $\mathcal{B}$ 和 $\mathcal{C}$ 下的矩阵 $A = \mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 是正交矩阵.
10. 对于 V 的任意两组 (相同或不同的) 有序规范正交基 $\mathcal{B}$ 和 $\mathcal{C}$, $\mathscr{A}$ 在 $\mathcal{B}$ 和 $\mathcal{C}$ 下的矩阵 $A = \mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 是正交矩阵.

思考题 7.6. 证明 V 上的正交变换满足以下性质:

1. 恒等变换是正交变换.
2. 若 $\mathscr{A} \in \mathbf{O}(V)$, 则 $\det(\mathscr{A}) \in \{\pm 1\}$ (故 $\mathscr{A}$ 可逆) 且 $\mathscr{A}^{-1} \in \mathbf{O}(V)$.

   以 $\mathbf{SO}(V)$ 表示 V 上行列式为 1 的正交变换集合. 该集合中的元素也称为**特殊正交变换** (special orthogonal transformation) 或者**第一类正交变换** (orthogonal transformation of the first kind). 行列式为 -1 的正交变换也称为**第二类正交变换** (orthogonal transformation of the second kind).
3. 若 $\mathscr{A}, \mathscr{B} \in \mathbf{O}(V)$, 则 $\mathscr{A}\mathscr{B} \in \mathbf{O}(V)$.

7.2.2 线性映射的伴随

回忆我们在本节的记号约定: V 和 W 总是表示非零的有限维实内积空间.

(7.2.6) 上一小节我们介绍了正交变换的概念. 我们知道, 如果 $\mathscr{A} \in \mathrm{End}(V)$ 是一个正交变换, 那么对于 V 中任意一对向量 u, v, 将 $\mathscr{A}$ 作用之后得到的两个向量 $\mathscr{A}u, \mathscr{A}v$ 之间的内积和之前 u, v 之间的内积相等 (推论 7.2.5 (3)). 也就是说, 我们有如下交换图表

(7.2.6.1)
$$
\begin{tikzcd}
 & V \times V \arrow[dr, "\langle \cdot\,,\cdot \rangle"] & \\
V \times V \arrow[ur, "\mathscr{A}\times\mathscr{A}"] \arrow[dr, "\mathrm{Id}\times\mathrm{Id}"'] & & \mathbb{R} \\
 & V \times V \arrow[ur, "\langle \cdot\,,\cdot \rangle"'] &
\end{tikzcd}
$$

这个图表说明, 当我们只考虑每一对向量之间内积的时候, 线性变换 $\mathscr{A}$ 的作用效果和恒等映射是一样的, 或者线性变换 $\mathscr{A}$ 的作用在这里可以忽略不计.

不知道读者是否想过这样的问题: 如果对于任意一对向量 $u, v \in V$ 我们只将 $\mathscr{A}$ 作用于第一个向量, 然后与另一个向量做内积, 这样运算的效果是怎样的呢? 有没有可能存在一个适当的映射 $V \times V \longrightarrow V \times V$ 使得图表

(7.2.6.2)
$$\begin{tikzcd} & V \times V \arrow[dr, "\langle \cdot , \cdot \rangle"] & \\ V \times V \arrow[ur, "\mathscr{A} \times \mathrm{Id}"] \arrow[dr, "?"'] & & \mathbb{R} \\ & V \times V \arrow[ur, "\langle \cdot , \cdot \rangle"'] & \end{tikzcd}$$

可交换呢?

为了回答这个问题, 我们采用这样的思路: 先任意取定一个向量 $v \in V$, 我们考虑

$$u \in V \longmapsto \langle \mathscr{A} u, v \rangle \in \mathbb{R}$$

给出的映射. 这是 V 上的一个线性泛函. 事实上, 如果采用之前 (6.1.22) 一段引入的记号, 以 φ'_v 表示线性泛函 $V \to \mathbb{R} : u \mapsto \langle u, v \rangle$, 那么上式给出的映射其实就是以下复合映射

$$V \xrightarrow{\mathscr{A}} V \xrightarrow{\varphi'_v} \mathbb{R}.$$

根据 Riesz 表示定理, 这个线性泛函具有 φ'_w 的形式. 即, 存在唯一的向量 w 使得 $\langle \mathscr{A} u, v \rangle = \langle u, w \rangle$ 对于所有的 $u \in V$ 成立. 显然, 这个向量 w 依赖于 v 和线性映射 $\mathscr{A}$, 我们把它记成 $\mathscr{A}^* v$. 于是,

(7.2.6.3)
$$\langle \mathscr{A} u, v \rangle = \langle u, \mathscr{A}^* v \rangle \quad \text{对于所有 } u \in V \text{ 成立.}$$

接下来, 我们再让 v 在 V 中任意变动, 通过 $v \mapsto \mathscr{A}^* v$ 得到一个映射 $\mathscr{A}^* : V \to V$. 则 (7.2.6.3) 式表明, 在图表 (7.2.6.2) 中问号代表的映射可以取为 $\mathrm{Id} \times \mathscr{A}^*$. 换言之, 我们有如下交换图表:

(7.2.6.4)
$$\begin{tikzcd} & V \times V \arrow[dr, "\langle \cdot , \cdot \rangle"] & \\ V \times V \arrow[ur, "\mathscr{A} \times \mathrm{Id}"] \arrow[dr, "\mathrm{Id} \times \mathscr{A}^*"'] & & \mathbb{R} \\ & V \times V \arrow[ur, "\langle \cdot , \cdot \rangle"'] & \end{tikzcd}$$

以上得到的映射 $\mathscr{A}^*$ 称为 $\mathscr{A}$ 的**伴随** (adjoint). ■

现在将上面介绍的伴随概念做一个推广, 并证明它仍是线性映射.

命题 7.2.7: 设 $\mathscr{A} : V \to W$ 为 (有限维内积空间之间的) 线性映射. 则存在唯一一个从 W 到 V 的映射, 通常记为 $\mathscr{A}^* : W \to V$, 使得

(7.2.7.1)
$$\langle \mathscr{A} v, w \rangle = \langle v, \mathscr{A}^* w \rangle \quad \text{对于所有 } v \in V \text{ 和所有 } w \in W \text{ 成立.}$$

这一映射 $\mathscr{A}^*$ 是线性映射, 称为 $\mathscr{A}$ 的**伴随**.

证明. 和 (7.2.6) 一段中的讨论一样, $\mathscr{A}^*$ 的存在唯一性由 Riesz 表示定理保证. 只需再证明 $\mathscr{A}^*$ 是线性的.

设 $w_1, w_2 \in W$. 对于任意 $v \in V$, 我们有

$$\begin{aligned}\langle\, v\,,\, \mathscr{A}^*(w_1+w_2)\,\rangle &= \langle\, \mathscr{A}v\,,\, w_1+w_2\,\rangle = \langle\, \mathscr{A}v\,,\, w_1\,\rangle + \langle\, \mathscr{A}v\,,\, w_2\,\rangle \\ &= \langle\, v\,,\, \mathscr{A}^*w_1\,\rangle + \langle\, v\,,\, \mathscr{A}^*w_2\,\rangle = \langle\, v\,,\, \mathscr{A}^*w_1 + \mathscr{A}^*w_2\,\rangle\,.\end{aligned}$$

这说明 $\mathscr{A}^*(w_1+w_2) - \mathscr{A}^*w_1 - \mathscr{A}^*w_2$ 这个向量能够和所有的 $v \in V$ 正交. 根据内积的非退化性质, 必然 $\mathscr{A}^*(w_1+w_2) = \mathscr{A}^*w_1 + \mathscr{A}^*w_2$.

类似地, 可以证明: 对于任意 $w \in W$ 和 $c \in \mathbb{R}$, $\mathscr{A}^*(cw) = c\mathscr{A}^*w$. 所以 $\mathscr{A}^*$ 是线性映射. □

例 7.2.8. 考虑线性映射

$$\mathscr{A}\ :\ \mathbb{R}^3 \longrightarrow \mathbb{R}^2\,; \quad (x,\, y,\, z) \longmapsto (y+3z\,,\, 2x)\,.$$

求 $\mathscr{A}^*$.

为此, 任意取定 $(a,\, b) \in \mathbb{R}^2$, 则对于任意 $(x,\, y,\, z) \in \mathbb{R}^3$,

$$\begin{aligned}\langle\,(x,\,y,\,z)\,,\, \mathscr{A}^*(a,\,b)\,\rangle &= \langle\, \mathscr{A}(x,\,y,\,z)\,,\,(a,\,b)\,\rangle = \langle\,(y+3z,\,2x)\,,\,(a,\,b)\,\rangle \\ &= a(y+3z)+2bx = \langle\,(x,\,y,\,z)\,,\,(2b,\,a,\,3a)\,\rangle\end{aligned}$$

因此 $\mathscr{A}^*(a,\,b) = (2b,\,a,\,3a)$. ■

例 7.2.9. 取定 $u \in V$ 和 $z \in W$. 定义

$$\mathscr{A}\ :\ V \longrightarrow W\,; \quad v \longmapsto \langle v,\, u\rangle z\,.$$

则对于任意取定的 $w \in W$, 我们有

$$\begin{aligned}\text{对任意 } v \in V\,,\ \langle\, v,\, \mathscr{A}^*w\,\rangle &= \langle\, \mathscr{A}v\,,\, w\,\rangle = \Big\langle\, \langle\, v,\, u\,\rangle z\,,\, w\,\Big\rangle \\ &= \langle\, v,\, u\,\rangle\langle\, z,\, w\,\rangle = \Big\langle\, v\,,\, \langle\, z,\, w\,\rangle u\,\Big\rangle.\end{aligned}$$

因此 $\mathscr{A}^*$ 的表达式为 $\mathscr{A}^*w = \langle\, z,\, w\,\rangle u$. ■

命题 7.2.10: 设 U, V, W 为有限维内积空间, $\mathscr{A}, \mathscr{B} \in \mathrm{Hom}(V,\, W)$, $\mathscr{C} \in \mathrm{Hom}(W,\, U)$, $c \in \mathbb{R}$. 则以下性质成立:

$$(\mathscr{A}+\mathscr{B})^* = \mathscr{A}^* + \mathscr{B}^*\,,\ (\mathscr{C}\mathscr{A})^* = \mathscr{A}^*\mathscr{C}^*\,,\ (c\mathscr{A})^* = c\mathscr{A}^*\,,\ (\mathscr{A}^*)^* = \mathscr{A}\,,\ 0^* = 0\,,\ I^* = I\,.$$

(这里 0 表示零变换, $I : V \to V$ 表示恒等变换.)

证明. 这里仅给出等式 $(\mathscr{C}\mathscr{A})^* = \mathscr{A}^*\mathscr{C}^*$ 的验证过程, 其余留给读者练习. 任意取定 $u \in U$, 则对于任意 $v \in V$,

$$\langle\, v,\, (\mathscr{C}\mathscr{A})^*u\,\rangle = \langle\, \mathscr{C}\mathscr{A}v\,,\, u\,\rangle = \langle\, \mathscr{A}v\,,\, \mathscr{C}^*u\,\rangle = \langle\, v\,,\, \mathscr{A}^*\mathscr{C}^*u\,\rangle\,.$$

因此 $(\mathscr{C}\mathscr{A})^* = \mathscr{A}^*\mathscr{C}^*$. □

接下来我们描述线性映射及其伴随的核与像之间的关系.

命题 7.2.11: 对任意 $\mathscr{A} \in \mathrm{Hom}(V,\, W)$ 以下关系成立:

(7.2.11.1) 在 W 中, $\mathrm{Ker}(\mathscr{A}^*) = \big(\mathrm{Im}(\mathscr{A})\big)^{\perp}$; $\mathrm{Im}(\mathscr{A}) = \big(\mathrm{Ker}(\mathscr{A}^*)\big)^{\perp}$.

(7.2.11.2) 在 V 中, $\mathrm{Ker}(\mathscr{A}) = \big(\mathrm{Im}(\mathscr{A}^*)\big)^{\perp}$; $\mathrm{Im}(\mathscr{A}^*) = \big(\mathrm{Ker}(\mathscr{A})\big)^{\perp}$.

(附图示意.)

证明. 这里只证明 (7.2.11.1) 式, (7.2.11.2) 式的证明留给读者练习.

对任意 $w \in W$,

$$\begin{aligned} w \in \operatorname{Ker}(\mathscr{A}^*) &\iff \mathscr{A}^* w = 0 \\ \text{根据内积的非退化性} &\iff \forall\, v \in V\,,\ \langle\, v,\, \mathscr{A}^* w\,\rangle = 0 \\ &\iff \forall\, v \in V\,,\ \langle\, \mathscr{A} v\,,\, w\,\rangle = 0 \\ &\iff \forall\, u \in \operatorname{Im}\mathscr{A}\,,\ \langle\, u\,,\, w\,\rangle = 0 \\ &\iff w \in \big(\operatorname{Im}(\mathscr{A})\big)^{\perp}. \end{aligned}$$

这就证明了 (7.2.11.1) 中的第一个等式. 第二个等式可以从第一个等式以及以下更一般的结论得到: 对于 W 的任意子空间 U,

$$(U^{\perp})^{\perp} = U\,. \tag{7.2.11.3}$$

事实上, 如 (7.1.3) 一段所说, 我们有正交分解 $W = U \perp U^{\perp}$ 成立. 因此 $\dim U^{\perp} = \dim W - \dim U$. 同理, 对于 $U' := U^{\perp}$, 也有 $\dim(U')^{\perp} = \dim W - \dim U'$. 于是

$$\dim U = \dim W - \dim U^{\perp} = \dim W - \dim U' = \dim(U')^{\perp} = \dim(U^{\perp})^{\perp}\,.$$

而包含关系 $U \subseteq (U^{\perp})^{\perp}$ 由定义可得, 结合以上维数间的等式就可以得到 (7.2.11.3). □

还有一个自然的问题需要回答: 线性映射及其伴随的矩阵有什么联系呢？

命题 7.2.12: 设 $\mathscr{A} \in \operatorname{Hom}(V, W)$, $\mathcal{B}$ 为 V 的一组有序规范正交基, $\mathcal{C}$ 为 W 的一组有序规范正交基. 则

$$\mathcal{M}_{\mathcal{C},\,\mathcal{B}}(\mathscr{A}^*) = \mathcal{M}_{\mathcal{B},\,\mathcal{C}}(\mathscr{A})^T\,.$$

也就是说, 在选择同一对有序规范正交基时, $\mathscr{A}$ 的矩阵和 $\mathscr{A}^*$ 的矩阵互为转置.

证明. 设 $\mathcal{B} = (e_1, \cdots, e_n)$, $\mathcal{C} = (f_1, \cdots, f_m)$. 对于每个 $i \in [\![1, n]\!]$, 我们希望求出 $\mathscr{A} e_i$ 在 $f_1, \cdots, f_m$ 这组基下的坐标列. 因为 $\mathcal{C} = (f_1, \cdots, f_m)$ 是一组规范正交基, 所以由命题 7.1.12 可知

$$\mathcal{M}_{\mathcal{C}}(\mathscr{A} e_i) = (\langle\, \mathscr{A} e_i\,,\, f_1\,\rangle\,, \cdots\ \langle\, \mathscr{A} e_i\,,\, f_m\,\rangle)^T\,.$$

由此可知

$$\mathcal{M}_{\mathcal{B},\,\mathcal{C}}(\mathscr{A}) = \Big(\mathcal{M}_{\mathcal{C}}(\mathscr{A} e_1)\,, \cdots, \mathcal{M}_{\mathcal{C}}(\mathscr{A} e_n)\Big) = \begin{pmatrix} \langle\, \mathscr{A} e_1,\, f_1\,\rangle & \langle\, \mathscr{A} e_2,\, f_1\,\rangle & \cdots & \langle\, \mathscr{A} e_n,\, f_1\,\rangle \\ \langle\, \mathscr{A} e_1,\, f_2\,\rangle & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \langle\, \mathscr{A} e_1,\, f_m\,\rangle & \cdots & \cdots & \langle\, \mathscr{A} e_n,\, f_m\,\rangle \end{pmatrix}.$$

类似可得

$$\mathcal{M}_{\mathcal{C},\,\mathcal{B}}(\mathscr{A}^*) = \Big(\mathcal{M}_{\mathcal{B}}(\mathscr{A}^* f_1)\,, \cdots, \mathcal{M}_{\mathcal{B}}(\mathscr{A}^* f_m)\Big) = \begin{pmatrix} \langle\, \mathscr{A}^* f_1,\, e_1\,\rangle & \langle\, \mathscr{A}^* f_2,\, e_1\,\rangle & \cdots & \langle\, \mathscr{A}^* f_m,\, e_1\,\rangle \\ \langle\, \mathscr{A}^* f_1,\, e_2\,\rangle & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \langle\, \mathscr{A}^* f_1,\, e_n\,\rangle & \cdots & \cdots & \langle\, \mathscr{A}^* f_m,\, e_n\,\rangle \end{pmatrix}.$$

由于 $\langle\, \mathscr{A} e_i,\, f_j\,\rangle = \langle\, e_i,\, \mathscr{A}^* f_j\,\rangle = \langle\, \mathscr{A}^* f_j,\, e_i\,\rangle$, 所以对比以上 $\mathcal{M}_{\mathcal{B},\,\mathcal{C}}(\mathscr{A})$ 和 $\mathcal{M}_{\mathcal{C},\,\mathcal{B}}(\mathscr{A}^*)$ 的表达式即知, 二者的关系是互为转置矩阵. □

读者不难验证, 在例 7.2.8 中, 映射 $\mathscr{A}$ 在 $\mathbb{R}^3$ 和 $\mathbb{R}^2$ 的标准基下的矩阵为 $\begin{pmatrix} 0 & 1 & 3 \\ 2 & 0 & 0 \end{pmatrix}$, 而 $\mathscr{A}^*$ 在标准基下的矩阵为 $\begin{pmatrix} 0 & 2 \\ 1 & 0 \\ 3 & 0 \end{pmatrix}$.

我们提醒读者特别注意, 在命题 7.2.12 中所取的基 $\mathcal{B}$ 和 $\mathcal{C}$ 都要求是规范正交基, 否则命题的结论可能是不对的.

例 7.2.13. 设 $V = W = \mathbb{R}^2$, $\mathscr{A} = I$ 为恒等映射, $\mathcal{B} = (e_1, e_2)$ 为标准基, $\mathcal{C} = (v_1, v_2)$, 其中

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \; v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

则

$$\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A}) = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathcal{M}_{\mathcal{C},\mathcal{B}}(\mathscr{A}) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

这两个矩阵不是互为转置的关系. 这里 $\mathcal{C}$ 不是规范正交基, 所以此现象并没有和命题 7.2.12 产生矛盾. ■

下面我们专门讨论线性变换的情况: 若 $\mathscr{A}$ 是 V 到自身的线性变换, 则 $\mathscr{A}^*$ 也是 V 上的线性变换. 故此时可以考虑 $\mathscr{A}^*\mathscr{A}$ 和 $\mathscr{A}\mathscr{A}^*$ 两种复合映射. 以下是一个十分有用的概念.

定义 7.2.14. 如果线性变换 $\mathscr{A} \in \operatorname{End}(V)$ 和它的伴随 $\mathscr{A}^* \in \operatorname{End}(V)$ 可交换, 即 $\mathscr{A}\mathscr{A}^* = \mathscr{A}^*\mathscr{A}$, 那么我们称 $\mathscr{A}$ 是一个**正规变换** (normal transformation) 或**正规算子** (normal operator). 也有的中文书使用“**规范变换**”或“**规范算子**”的名称.

如果一个矩阵 $A \in \mathbf{M}_n(\mathbb{R})$ 满足 $A^TA = AA^T$, 则称 A 是个**正规矩阵** (normal matrix) 或者**规范矩阵**.

如果 $\mathscr{A} \in \operatorname{End}(V)$ 满足 $\mathscr{A} = \mathscr{A}^*$, 则称 $\mathscr{A}$ 是个**对称变换/算子**(symmetric transformation/operator) 或者**自伴变换/算子** (self-adjoint transformation/operator).

如果 $\mathscr{A} = -\mathscr{A}^*$, 则称 $\mathscr{A}$ 是个**斜对称变换/算子**(skew-symmetric transformation/operator). 形容词**交错** (alternating) 和**反对称** (anti-symmetric) 可以认为是“斜对称”的同义词.

显然, 自伴变换 (亦即对称变换) 和斜对称变换 (亦即交错变换) 都是正规变换. ■

思考题 7.7. 设 $\mathscr{A} \in \operatorname{End}(V)$. 则下列陈述等价:

1. $\mathscr{A}$ 是正规变换 (对称变换、斜对称变换).
2. 对于 V 的任意一组规范正交基 $\mathcal{B}$, 矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是正规 (对称、斜对称) 矩阵.
3. 存在 V 的一组规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是正规 (对称、斜对称) 矩阵.

例 7.2.15. 假设 $\mathscr{A} \in \operatorname{End}(V)$ 是个正交变换. 则对于任意 $u, v \in V$,

$$\langle u, v \rangle = \langle \mathscr{A}u, \mathscr{A}v \rangle = \langle u, \mathscr{A}^*\mathscr{A}v \rangle.$$

所以 $\mathscr{A}^*\mathscr{A} = \mathrm{Id}$. 因为 $\mathscr{A}$ 是线性变换, 这可以推出 $\mathscr{A}\mathscr{A}^* = \mathrm{Id}$. 因此, 正交变换都是正规变换. 当然, 我们也可以用矩阵的观点解释这个事实. ■

例 7.2.16. 除了正交变换、对称变换和斜对称变换之外, 还有其它很多正规变换.

例如, 以矩阵的方式来看, 读者不难验证 $A = \begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}$ 就是一个正规矩阵, 但 A 不是正交矩阵、对称矩阵或斜对称矩阵. ■

思考题 7.8. 设 $A \in \mathbf{M}_n(\mathbb{R})$ 为正规矩阵.

1. 证明: 与 A 正交相似的矩阵也是正规矩阵.
2. 举例说明: 与 A 相似的矩阵不一定是正规矩阵.
3. 举例说明: 与 A 相合的矩阵不一定是正规矩阵.

接下来我们研究正规算子——特别是自伴算子——的一些特殊性质.

引理 7.2.17: 设 $\mathscr{A} \in \mathrm{End}(V)$ 是自伴算子. 如果对于所有 $v \in V$ 均有 $\langle \mathscr{A}v\,,\, v\rangle = 0$, 则必有 $\mathscr{A} = 0$.

证明. 取定 V 的一组规范正交基 $\mathcal{B}$. 则 $\mathscr{A}$ 自伴意味着 $A = \mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是对称矩阵 (思考题 7.7). 根据题设条件, 对于所有 $v \in V$, 我们有

$$0 = \langle \mathscr{A}v\,,\, v\rangle = \langle v\,,\, \mathscr{A}v\rangle = \mathcal{M}_{\mathcal{B}}(v)^T \cdot \mathcal{M}_{\mathcal{B}}(\mathscr{A}v) = \mathcal{M}_{\mathcal{B}}(v)^T A \mathcal{M}_{\mathcal{B}}(v)\,.$$

这说明矩阵 A 能够使所有 $\alpha \in \mathbb{R}^n$ (其中 $n = \dim V$) 都满足 $\alpha^T A\alpha = 0$. 也就是说, 二次型

$$\langle A\rangle \,:\, \mathbb{R}^n \longrightarrow \mathbb{R}\,;\, \alpha \longmapsto \alpha^T A\alpha$$

恒等于 0. 根据二次型和对称矩阵之间的一一对应关系, 此时必有 $A = 0$. 所以 $\mathscr{A} = 0$. □

注记 7.2.18. 注意, 引理 7.2.17 中 $\mathscr{A}$ 自伴的条件不能去掉. 例如, 线性变换

$$\mathscr{A} \,:\, \mathbb{R}^2 \longrightarrow \mathbb{R}^2\,;\quad \begin{pmatrix} x \\ y \end{pmatrix} \longmapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}$$

满足 $\langle \mathscr{A}v, v\rangle = 0$ 对所有 $v \in \mathbb{R}^2$ 成立 (读者能否对这个事实给出几何解释?), 但 $\mathscr{A} \neq 0$. ■

命题 7.2.19: 线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 是正规变换的充分必要条件是对于所有 $v \in V$, $\|\mathscr{A}v\| = \|\mathscr{A}^*v\|$.

证明. 注意到 $\mathscr{A}^*\mathscr{A} - \mathscr{A}\mathscr{A}^*$ 是自伴算子, 所以

$$\begin{aligned}
\forall\, v \in V\,,\ \|\mathscr{A}v\|^2 = \|\mathscr{A}^*v\|^2 \ &\Longleftrightarrow\ \forall\, v \in V\,,\ \langle \mathscr{A}v\,,\, \mathscr{A}v\rangle = \langle \mathscr{A}^*v\,,\, \mathscr{A}^*v\rangle \\
&\Longleftrightarrow\ \forall\, v \in V\,,\ \langle v\,,\, \mathscr{A}^*\mathscr{A}v\rangle = \langle \mathscr{A}\mathscr{A}^*v\,,\, v\rangle \\
&\Longleftrightarrow\ \forall\, v \in V\,,\ \langle \mathscr{A}^*\mathscr{A}v\,,\, v\rangle = \langle \mathscr{A}\mathscr{A}^*v\,,\, v\rangle \\
&\Longleftrightarrow\ \forall\, v \in V\,,\ \langle (\mathscr{A}^*\mathscr{A} - \mathscr{A}\mathscr{A}^*)v\,,\, v\rangle = 0 \\
\text{根据引理 7.2.17}\ &\Longleftrightarrow\ \mathscr{A}^*\mathscr{A} - \mathscr{A}\mathscr{A}^* = 0\,.
\end{aligned}$$

这就证明了所需的结论. □

由上述命题可以推出正规变换的特征值和特征向量的一些重要性质.

命题 7.2.20: 设 $\mathscr{A} \in \mathrm{End}(V)$ 为正规变换, $\lambda \in \mathbb{R}$ 是 $\mathscr{A}$ 的一个特征值, $v \in V$ 是 $\mathscr{A}$ 的属于 λ 的一个特征向量.

则 λ 也是 $\mathscr{A}^*$ 的特征值, 并且 v 也是 $\mathscr{A}^*$ 的属于 λ 的特征向量.

证明. 事实上, 读者不难验证 $\mathscr{A} - \lambda I$ 也是正规变换. 根据命题 7.2.19 可得

$$0 = \|(\mathscr{A} - \lambda I)v\| = \|(\mathscr{A} - \lambda I)^*v\| = \|(\mathscr{A}^* - \lambda I)v\|\,.$$

所以 $\mathscr{A}^*v = \lambda v$. 结论得证. □

思考题 7.9. 设 $\mathscr{A} \in \mathrm{End}(V)$.

1. 证明: 实数 λ 是 $\mathscr{A}$ 的特征值当且仅当 λ 是 $\mathscr{A}^*$ 的特征值.

   (提示: 若 λ 不是 $\mathscr{A}^*$ 的特征值, 则 $\mathscr{A}^* - \lambda I$ 是满射.)

2. 举例说明: $\mathscr{A}$ 的特征向量不一定是 $\mathscr{A}^*$ 的特征向量.

   (因此命题 7.2.20 中 $\mathscr{A}$ 正规的假设是不能去掉的.)

命题 7.2.21: 设 $\mathscr{A} \in \mathrm{End}(V)$ 为正规变换, λ, μ 是 $\mathscr{A}$ 的两个不同特征值, $v, w \in V$ 是分别属于 λ, μ 的特征向量.

则 v 与 w 正交. 所以, 特征子空间 $E(\lambda, \mathscr{A})$ 和 $E(\mu, \mathscr{A})$ 相互正交.

证明. 根据命题 7.2.20, w 也是 $\mathscr{A}^*$ 的特征向量, 对应的特征值仍是 μ. 于是

$$\lambda\langle v, w\rangle = \langle \lambda v, w\rangle = \langle \mathscr{A} v, w\rangle = \langle v, \mathscr{A}^* w\rangle = \langle v, \mu w\rangle = \mu\langle v, w\rangle .$$

由于 $\lambda \neq \mu$, 上式可以推出 $\langle v, w\rangle = 0$. □

7.2.3 实自伴算子的谱定理和对称矩阵的正交相似标准形

这一小节的主要目标是讨论实自伴算子在规范正交基下的矩阵最简单形式是什么样子, 或者说实对称矩阵在正交相似产生的变化之下能化成什么最简单的样子.

以前我们讨论线性算子或矩阵的对角化时, 关键步骤是将整个空间尽可能地分解为不变子空间的直和. 现在我们需要处理的问题有些类似, 但又有所加强: 我们希望将整个空间尽可能地分解为不变子空间的正交和.

引理 7.2.22: 设 $\mathscr{A} \in \mathrm{End}(V)$ 是自伴算子, U 是 $\mathscr{A}$ 的不变子空间. 则以下结论成立:

1. 正交补空间 $U^\perp$ 也是 $\mathscr{A}$ 的不变子空间.

   (因此正交分解式 $V = U \perp U^\perp$ 可以认为是一种由 $\mathscr{A}$ 的两个不变子空间给出的直和分解式.)

2. 限制映射 $\mathscr{A}|_U \in \mathrm{End}(U)$ 和 $\mathscr{A}|_{U^\perp} \in \mathrm{End}(U^\perp)$ 也都是自伴算子.

证明. (1) 设 $v \in U^\perp$, 要验证 $\mathscr{A} v \in U^\perp$, 只要验证 $\mathscr{A} v$ 和 U 中所有向量正交. 事实上, 对任意 $u \in U$, 因为 U 是 $\mathscr{A}$ 的不变子空间, 故 $\mathscr{A} u \in U$. 所以,

$$\begin{aligned}\langle \mathscr{A} v, u\rangle &= \langle v, \mathscr{A}^* u\rangle \\ &= \langle v, \mathscr{A} u\rangle \quad (\text{因为 } \mathscr{A} \text{ 自伴}) \\ &= 0 \quad (\text{因为 } \mathscr{A} u \in U \text{ 且 } v \in U^\perp) .\end{aligned}$$

(2) 要证明 $\mathscr{A}|_U$ 是自伴的, 就是要证明对于任意 $u \in U$, 均有 $\mathscr{A}|_U u = (\mathscr{A}|_U)^* u$. 其证明方法是验证: 对于所有 $v \in U$, $\langle v, \mathscr{A}|_U u\rangle = \langle v, (\mathscr{A}|_U)^* u\rangle$. 事实上,

$$\begin{aligned}\langle v, \mathscr{A}|_U u\rangle &= \langle v, \mathscr{A} u\rangle \\ (\text{因为 } \mathscr{A} \text{ 自伴}) \quad &= \langle v, \mathscr{A}^* u\rangle = \langle \mathscr{A} v, u\rangle \\ (\text{因为 } v \in U) \quad &= \langle \mathscr{A}|_U v, u\rangle = \langle v, (\mathscr{A}|_U)^* u\rangle .\end{aligned}$$

根据结论 (1), 以上讨论对于 $\mathscr{A}$ 的不变子空间 $U' := U^\perp$ 也成立, 即, $\mathscr{A}|_{U'} = \mathscr{A}|_{U^\perp}$ 也是自伴算子. □

现在我们想证明: 自伴算子总会在某一组规范正交基下的矩阵为对角阵. 也就是说, 定理 6.1.15 具有如下加强版:

定理 7.2.23 (**自伴算子的谱定理** (spectral theorem for self-adjoint operators)): 设 $(V, \langle\cdot\,,\,\cdot\rangle)$ 是有限维非零实内积空间, $\mathscr{A} \in \mathrm{End}(V)$ 是自伴算子.

则一定存在内积 $\langle\cdot\,,\,\cdot\rangle$ 的一组规范正交基 $\mathcal{E}$ 使得矩阵 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 是对角阵.

以矩阵语言言之, 任意实对称方阵可以在 $\mathbb{R}$ 上正交相似 (也是正交相合) 于一个实对角矩阵.

在物理学或力学中, 以上定理经常被称为**主轴定理** (principal axis theorem), 这可能和此定理的几何解释有关. 我们之后会回到这个话题 (参见例 7.3.5).

为证明以上定理, 我们先要证明自伴算子一定有实特征值. 我们希望从线性变换的角度解决这个问题. 为此我们先证明一个引理.

引理 7.2.24: 设 $\mathscr{A}$ 是 (非零实内积空间) V 上的自伴算子, $b, c \in \mathbb{R}$ 满足 $b^2 - 4c < 0$.

则线性变换 $\mathscr{A}^2 + b\mathscr{A} + cI$ 是可逆的.

证明. 只需要证明 $\mathscr{A}^2 + b\mathscr{A} + cI$ 是单射即可. 假如不是, 那么存在非零向量 $v \in V$ 使得 $(\mathscr{A}^2 + b\mathscr{A} + cI)v = 0$. 于是 $\langle (\mathscr{A}^2 + b\mathscr{A} + cI)v\,,\,v\rangle = 0$. 但另一方面,

$$\begin{aligned}
0 = \langle (\mathscr{A}^2 + b\mathscr{A} + cI)v\,,\,v\rangle &= \langle \mathscr{A}^2 v\,,\,v\rangle + b\langle \mathscr{A}v\,,\,v\rangle + c\langle v\,,\,v\rangle \\
&= \langle \mathscr{A}v\,,\,\mathscr{A}^* v\rangle + b\langle \mathscr{A}v\,,\,v\rangle + c\langle v\,,\,v\rangle \\
(\text{因为 } \mathscr{A} \text{ 自伴}) \quad &= \langle \mathscr{A}v\,,\,\mathscr{A}v\rangle + b\langle \mathscr{A}v\,,\,v\rangle + c\langle v\,,\,v\rangle \\
(\text{根据 Cauchy–Schwarz 不等式}) \quad &\geq \|\mathscr{A}v\|^2 - |b|\cdot\|\mathscr{A}v\|\cdot\|v\| + c\|v\|^2 \\
(\text{令 } x = \|\mathscr{A}v\|\,,\, y = \|v\|) \quad &= x^2 - |b|xy + cy^2 \\
&= (x,\, y)\begin{pmatrix} 1 & -\frac{1}{2}|b| \\ -\frac{1}{2}|b| & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.
\end{aligned}$$

由题设条件 $b^2 - 4c < 0$ 可知矩阵 $\begin{pmatrix} 1 & -\frac{1}{2}|b| \\ -\frac{1}{2}|b| & c \end{pmatrix}$ 正定. 而 $y = \|v\| \neq 0$. 所以上式末端应该严格大于 0. 这就产生了矛盾. □

命题 7.2.25: 设 $\mathscr{A}$ 是 (有限维非零实内积空间) V 上的自伴算子. 则 $\mathscr{A}$ (在 $\mathbb{R}$ 内) 一定有特征值.

以矩阵语言表述: 任意实对称矩阵一定有实的特征值.

证明. 设 $f(X) \in \mathbb{R}[X]$ 为 $\mathscr{A}$ 的特征多项式. 我们知道, 在 $\mathbb{R}[X]$ 中可以将 f 做如下因式分解 (参见本书上册命题 A.4.12):

$$f(X) = (X^2 + b_1X + c_1)\cdots(X^2 + b_MX + c_M)(X - \lambda_1)\cdots(X - \lambda_m)$$

其中 $M, m \in \mathbb{N}$ 都可能为 0, $\lambda_1, \cdots, \lambda_m \in \mathbb{R}$, 而 $b_j, c_j \in \mathbb{R}$ 均满足 $b_j^2 - 4c_j < 0$. 由 Cayley–Hamilton 定理可知

$$0 = f(\mathscr{A}) = (\mathscr{A}^2 + b_1\mathscr{A} + c_1I)\cdots(\mathscr{A}^2 + b_M\mathscr{A} + c_MI)(\mathscr{A} - \lambda_1I)\cdots(\mathscr{A} - \lambda_mI)\,.$$

而引理 7.2.24 说明 $(\mathscr{A}^2 + b_1\mathscr{A} + c_1I)\cdots(\mathscr{A}^2 + b_M\mathscr{A} + c_MI)$ 是可逆变换. 于是上式说明 $m \geq 1$, 而且

$$(\mathscr{A} - \lambda_1I)\cdots(\mathscr{A} - \lambda_mI) = 0\,.$$

于是, $\mathscr{A} - \lambda_1I\,, \cdots, \mathscr{A} - \lambda_mI$ 这些线性变换中至少有一个是不可逆的. 与之对应的 λ_i 一定是 $\mathscr{A}$ 的特征值. □

定理 7.2.23 的证明. 我们采用对 $\dim V$ 的归纳法来完成证明. 设 $n = \dim V$. 若 $n = 1$, 则结论显然成立.

以下设 $n > 1$. 根据命题 7.2.25, 存在 $\lambda_1 \in \mathbb{R}$ 是 $\mathscr{A}$ 的特征值. 取 $v \in V$ 为 $\mathscr{A}$ 的属于 λ_1 的特征向量. 于是 $U = \operatorname{span}(v)$ 是 $\mathscr{A}$ 的一维不变子空间. 根据引理 7.2.22, $U^\perp$ 是 $\mathscr{A}$ 的 $n-1$ 维不变子空间, 且 $\mathscr{A}' := \mathscr{A}|_{U^\perp}$ 仍是自伴算子. 根据归纳假设, $U^\perp$ 中存在一组规范正交基 $e_2, \cdots, e_n$, 使得 $\mathscr{A}'$ 在这组基下的矩阵为对角阵 $\begin{pmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$. 取向量 e_1 构成 U 的规范正交基 (例如 $e_1 = \|v\|^{-1}v$), 则由正交分解式 $V = U \perp U^\perp$ 可知, $e_1, e_2, \cdots, e_n$ 构成 V 中的一组规范正交基, 而且在这组基下 $\mathscr{A}$ 的矩阵为 $\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$. 这就证明了定理. □

推论 7.2.26: *若 $A \in \mathbf{M}_n(\mathbb{R})$ 为对称矩阵, 则 A 的所有复特征值都是实数.*

证明. 根据定理 7.2.23, A 正交相似于一个实对角阵 A'. 显然 A' 的复特征值都是实数, 而 A 和 A' 的复特征值连同重数都对应相同, 所以推论成立. □

注记 7.2.27. 推论 7.2.26 有另外一种常见的矩阵形式证明如下:

对于任意复矩阵 $M = (\alpha_{ij}) \in \mathbf{M}_{m\times n}(\mathbb{C})$, 可以定义其**共轭** (conjugate) 为 $\overline{M} = (\overline{\alpha}_{ij})$. 特别地, 对于列向量 $v \in \mathbb{C}^n$ 可以定义其共轭 $\overline{v}$.

设 $A \in \mathbf{M}_n(\mathbb{R})$ 为实对称阵. 将其视为复向量空间 $\mathbb{C}^n$ 上的线性变换 $v \mapsto Av$ 在标准基下的矩阵. 设 $\lambda \in \mathbb{C}$ 为它的一个复特征值, $x \in \mathbb{C}^n$ 为属于 λ 的一个特征向量. 则由 $Ax = \lambda x$ 两边取共轭可得

$$A\overline{x} = \overline{Ax} = \overline{\lambda}\overline{x}.$$

再同时左乘 x^T 可得

$$x^T A\overline{x} = \overline{\lambda} x^T \overline{x}. \tag{7.2.27.1}$$

另一方面, 由等式 $Ax = \lambda x$ 先两边取转置, 再从右边乘以 $\overline{x}$ 可得

$$x^T A\overline{x} = (Ax)^T\overline{x} = (\lambda x)^T\overline{x} = \lambda x^T\overline{x}. \tag{7.2.27.2}$$

比较 (7.2.27.1) 和 (7.2.27.2) 式可得 $\overline{\lambda} x^T\overline{x} = \lambda x^T\overline{x}$. 因为这里的 $x^T\overline{x}$ 是个非零实数, 所以我们得到 $\overline{\lambda} = \lambda$. 即, λ 是实数. ■

思考题 7.10. 设 $A \in \mathbf{M}_n(\mathbb{R})$ 为对称矩阵. 证明: A 正定 (半正定) 的充分必要条件是 A 的所有特征值为正实数 (非负实数).

(7.2.28) 定理 7.2.23 已经说明: 任意实对称阵 A 总是正交相似于一个对角阵 D, 这样的 D 称为 A 的一个**正交相似标准形** (orthogonal similarity canonical form). 显然, D 的对角形元素全由 A 的特征值 (连同重数) 决定. 因此, 在不计对角线元素排序的意义下, D 是由 A 完全确定的.

现在来讨论将实对称阵化为正交相似标准形的具体计算方法. 事实上, 过去我们知道, 一个线性变换 $\mathscr{A} \in \operatorname{End}(V)$ 可对角化的等价条件是空间 V 中存在一组由 $\mathscr{A}$ 的特征向量构成的基, 也等价于说 V 可以分解为 $\mathscr{A}$ 的所有特征子空间之直和. 现在定理 7.2.23 的结论实际上可以换成以下两种方式中的任意一种来叙述:

- 对于任意自伴算子 $\mathscr{A} \in \mathrm{End}(V)$, 一定存在由 $\mathscr{A}$ 的特征向量构成的规范正交基.
- 对于任意自伴算子 $\mathscr{A} \in \mathrm{End}(V)$, 若 $\lambda_1, \cdots, \lambda_r$ 为 $\mathscr{A}$ 的所有不同特征值, 则 V 可以分解为特征子空间 $E(\lambda_1, \mathscr{A}), \cdots, E(\lambda_r, \mathscr{A})$ 的正交和.

据此, 我们可以通过以下步骤找到一组规范正交基将自伴算子的矩阵化为对角形:

首先, 计算出自伴算子 $\mathscr{A} \in \mathrm{End}(V)$ 的所有不同特征值 $\lambda_1, \cdots, \lambda_r$.

然后, 对每个 λ_i, 在特征子空间 $U_i = E(\lambda_i, \mathscr{A})$ 内找出一组规范正交基 $\mathcal{E}_i$ (例如可以先任取 U_i 的一组基, 通过 Gram–Schmidt 正交化得到一组规范正交基).

最后, 将 $\mathcal{E}_1, \cdots, \mathcal{E}_r$ 合并起来就得到 V 的一组规范正交基 $\mathcal{E}$. 根据构造, 这组规范正交基全部由 $\mathscr{A}$ 的特征向量构成, 因此 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 必定是对角阵. ■

例 7.2.29. 设 $A = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix}$. 求正交矩阵 Q 使 $Q^{-1}AQ$ 成为对角阵.

考虑线性变换 $\mathscr{A} : \mathbb{R}^4 \to \mathbb{R}^4;\ X \mapsto AX$.

首先, 计算特征多项式得 $|\lambda I - A| = (\lambda - 1)^3(\lambda + 3)$. 故 $\mathscr{A}$ 共有两个不同的特征值: $\lambda_1 = 1$ 和 $\lambda_2 = -3$.

计算特征子空间得

$$\begin{aligned} E(\lambda_1, \mathscr{A}) &= \mathrm{Ker}(\lambda_1 I - \mathscr{A}) = \{(x_1, x_2, x_3, x_4)^T \in \mathbb{R}^4 \mid x_1 - x_2 - x_3 + x_4 = 0\}\,. \\ E(\lambda_2, \mathscr{A}) &= \mathrm{Ker}(\lambda_2 I - \mathscr{A}) = \{(x_1, x_2, x_3, x_4)^T \in \mathbb{R}^4 \mid x_1 - x_2 - x_3 - 3x_4 = 0\,, \\ &\qquad\qquad x_2 + x_4 = 0\,,\ x_3 + x_4 = 0\}\,. \end{aligned}$$

于是 $E(\lambda_1, \mathscr{A})$ 有一组基为

$$v_1 = (1, 1, 0, 0)^T\,,\ v_2 = (1, 0, 1, 0)^T\,,\ \ v_3 = (-1, 0, 0, 1)^T\,.$$

$E(\lambda_2, \mathscr{A})$ 有一组基为 $v_4 = (1, -1, -1, 1)^T$.

通过 Gram–Schmidt 正交化可以得到 $E(\lambda_1, \mathscr{A})$ 的一组规范正交基

$$\begin{aligned} f_1 &= \frac{v_1}{\|v_1\|} = \left(\frac{1}{\sqrt{2}}\,, \frac{1}{\sqrt{2}}\,, 0\,, 0\right)^T\,, \\ f_2 &= \frac{v_2 - \langle v_2, f_1 \rangle f_1}{\|v_2 - \langle v_2, f_1 \rangle f_1\|} = \frac{\left(\frac{1}{2}\,, -\frac{1}{2}\,, 1\,, 0\right)^T}{\|\left(\frac{1}{2}\,, -\frac{1}{2}\,, 1\,, 0\right)^T\|} = \left(\frac{1}{\sqrt{6}}\,, -\frac{1}{\sqrt{6}}\,, \frac{2}{\sqrt{6}}\,, 0\right)^T\,, \\ f_3 &= \frac{v_3 - \langle v_3, f_1 \rangle f_1 - \langle v_3, f_2 \rangle f_2}{\|v_3 - \langle v_3, f_1 \rangle f_1 - \langle v_3, f_2 \rangle f_2\|} = \frac{\left(-\frac{1}{3}\,, \frac{1}{3}\,, \frac{1}{3}\,, 1\right)^T}{\|\left(-\frac{1}{3}\,, \frac{1}{3}\,, \frac{1}{3}\,, 1\right)^T\|} = \left(-\frac{1}{\sqrt{12}}\,, \frac{1}{\sqrt{12}}\,, \frac{1}{\sqrt{12}}\,, \frac{3}{\sqrt{12}}\right)^T\,. \end{aligned}$$

而 $E(\lambda_2, \mathscr{A})$ 的一组规范正交基为

$$f_4 = \frac{v_4}{\|v_4\|} = \left(\frac{1}{2}\,, -\frac{1}{2}\,, -\frac{1}{2}\,, \frac{1}{2}\right)^T\,.$$

综上可知, 在 f_1, f_2, f_3, f_4 这组规范正交基下, 线性变换 $X \mapsto AX$ 的矩阵为对角阵 $D = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & -3 \end{pmatrix}$. 而从标准基到这组基的过渡矩阵 $Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{12}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{12}} & -\frac{1}{2} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{12}} & -\frac{1}{2} \\ 0 & 0 & \frac{3}{\sqrt{12}} & \frac{1}{2} \end{pmatrix}$ 是正交矩阵, 它满足 $Q^{-1}AQ = D$. ■

7.2.4 实正规算子的谱定理和正规矩阵的正交相似标准形

上一小节我们讨论了自伴算子 (亦即对称变换) 的谱定理. 现在我们希望进一步了解对于其他类型的正规算子——例如正交算子——情况又会如何.

和之前的讨论一样, 我们希望将内积空间分解为不变子空间的正交和. 为此我们先证明如下结论:

引理 7.2.30: 设 $\mathscr{A} \in \mathrm{End}(V)$ 是正规变换, U 是 $\mathscr{A}$ 的不变子空间. 则以下结论成立:

1. 正交补空间 $U^\perp$ 也是 $\mathscr{A}$ 的不变子空间, 而 U 也是伴随变换 $\mathscr{A}^*$ 的不变子空间.
2. $(\mathscr{A}|_U)^* = (\mathscr{A}^*)|_U$, 即, $\mathscr{A}^*$ 在 U 上的限制等于 $\mathscr{A}|_U$ 的伴随.
3. 限制映射 $\mathscr{A}|_U \in \mathrm{End}(U)$ 和 $\mathscr{A}|_{U^\perp} \in \mathrm{End}(U^\perp)$ 也都是正规变换.

证明. (1) 我们知道 V 可以做正交和分解 $V = U \perp U^\perp$. 取 U 的一组规范正交基 $e_1, \cdots, e_m$ 及 $U^\perp$ 的一组规范正交基 $e_{m+1}, \cdots, e_{m+r}$, 则 $\mathcal{B} = (e_1, \cdots, e_m, \cdots, e_{m+r})$ 构成 V 的一组规范正交基. 由于 U 是 $\mathscr{A}$ 的不变子空间, 故 $\mathscr{A}$ 在 $\mathcal{B}$ 下的矩阵形如

$$\mathcal{M}_{\mathcal{B}}(\mathscr{A}) = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}, \quad \text{其中}\ A \in \mathbf{M}_m(\mathbb{R}),\ B \in \mathbf{M}_{m\times r}(\mathbb{R}),\ C \in \mathbf{M}_r(\mathbb{R}). \tag{7.2.30.1}$$

注意到 $\mathcal{B}$ 是规范正交基, 故由命题 7.2.12 可知, $\mathscr{A}^*$ 在 $\mathcal{B}$ 下的矩阵为 $\mathcal{M}_{\mathcal{B}}(\mathscr{A}^*) = \mathcal{M}_{\mathcal{B}}(\mathscr{A})^T = \begin{pmatrix} A^T & 0 \\ B^T & C^T \end{pmatrix}$.

我们只需要证明 $B = 0$.

为此我们将使用命题 7.2.19 所给出的正规算子的性质, 即, 对于任意 $v \in V$, $\|\mathscr{A}v\|^2 = \|\mathscr{A}^* v\|^2$. 现在对 U 的规范正交基 $e_1, \cdots, e_m$ 中每个向量考虑这个性质. 对于每个 $j \in [\![1, m]\!]$, $\|\mathscr{A}e_j\|^2$ 可以通过向量 $\mathscr{A}e_j$ 在规范正交基 $\mathcal{B}$ 下的坐标列以各坐标平方和的形式给出 (命题 7.1.12). 也就是说

$$\|\mathscr{A}e_j\|^2 = \text{矩阵 } A \text{ 的第 } j \text{ 列所有元素的平方和.}$$

于是

$$\sum_{j=1}^m \|\mathscr{A}e_j\|^2 = \text{矩阵 } A \text{ 所有元素的平方和.} \tag{7.2.30.2}$$

类似地, 对于每个 $j \in [\![1, m]\!]$,

$$\begin{aligned} \|\mathscr{A}^* e_j\|^2 &= \text{矩阵 } \begin{pmatrix} A^T \\ B^T \end{pmatrix} \text{ 的第 } j \text{ 列所有元素的平方和 .} \\ &= \text{矩阵 } A^T \text{ 的第 } j \text{ 列所有元素的平方和 } + \text{ 矩阵 } B^T \text{ 的第 } j \text{ 列所有元素的平方和.} \end{aligned}$$

故

$$\begin{aligned} \sum_{j=1}^m \|\mathscr{A}^* e_j\|^2 &= \text{矩阵 } A^T \text{ 中所有元素的平方和 } + \text{ 矩阵 } B^T \text{ 中所有元素的平方和} \\ &= \text{矩阵 } A \text{ 中所有元素的平方和 } + \text{ 矩阵 } B \text{ 中所有元素的平方和.} \end{aligned} \tag{7.2.30.3}$$

命题 7.2.19 表明 (7.2.30.2) 和 (7.2.30.3) 两式中的等号左边是相等的. 因此, 矩阵 B 的所有元素平方和等于 0. 这就是说 $B = 0$ (因为 B 是实矩阵).

(2) 欲证结论为: 对于任意 $u \in U$, $(\mathscr{A}|_U)^* u = \mathscr{A}^*|_U u$. 为此我们设 $u \in U$ 任意取定, 再考虑 U 中所有向量 $w \in U$. 只要能验证 $\langle w, (\mathscr{A}|_U)^* u\rangle$ 和 $\langle w, \mathscr{A}^*|_U u\rangle$ 这两个内积总相等即可. 事实上,

$$\langle w, (\mathscr{A}|_U)^* u\rangle = \langle \mathscr{A}|_U w, u\rangle = \langle \mathscr{A} w, u\rangle = \langle w, \mathscr{A}^* u\rangle = \langle w, \mathscr{A}^*|_U u\rangle .$$

另外一种证明方法是使用 (1) 的证明过程中的记号, 来验证在 U 的规范正交基 $\mathcal{E} = (e_1, \cdots, e_m)$ 下 $(\mathscr{A}|_U)^*$ 和 $(\mathscr{A}^*)|_U$ 的矩阵相同. 事实上, 若采用 (7.2.30.1) 式中的记号, 那么 $\mathcal{M}_{\mathcal{E}}(\mathscr{A}|_U) = A$, 从而 $\mathcal{M}_{\mathcal{E}}((\mathscr{A}|_U)^*) = A^T$. 而从 $\mathcal{M}_{\mathcal{B}}(\mathscr{A}^*) = \begin{pmatrix} A^T & 0 \\ 0 & C^T \end{pmatrix}$ 可知, $\mathcal{M}_{\mathcal{E}}(\mathscr{A}^*|_U) = A^T$.

(3) 根据 (2),

$$(\mathscr{A}|_U)\cdot(\mathscr{A}|_U)^* = \mathscr{A}|_U \cdot \mathscr{A}^*|_U = (\mathscr{A}\mathscr{A}^*)|_U = (\mathscr{A}^*\mathscr{A})|_U = \mathscr{A}^*|_U \cdot \mathscr{A}|_U = (\mathscr{A}|_U)^* \cdot (\mathscr{A}|_U).$$

所以由定义即知 $\mathscr{A}|_U$ 为正规变换.

由于 $U' = U^\perp$ 也是 $\mathscr{A}$ 的不变子空间, 将刚才的结论用于 U' 即知 $\mathscr{A}|_{U'} = \mathscr{A}|_{U^\perp}$ 是正规变换.

若用矩阵的方式来解释, 则根据 (1) 中的证明, 在 $\mathcal{B} = (e_1, \cdots, e_m, e_{m+1}, \cdots, e_{m+r})$ 这组规范正交基下, $\mathcal{M}_{\mathcal{B}}(\mathscr{A}) = \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}$. 而在 U 的规范正交基 $\mathcal{E} = (e_1, \cdots, e_m)$ 下, $\mathcal{M}_{\mathcal{E}}(\mathscr{A}|_U) = A$, 在 $U^\perp$ 的规范正交基 $\mathcal{F} = (e_{m+1}, \cdots, e_{m+r})$ 下, $\mathcal{M}_{\mathcal{E}}(\mathscr{A}|_{U^\perp}) = C$. 由于 $\mathscr{A}$ 是正规变换, 故 $\begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}$ 是正规矩阵. 由此易知 A 和 C 均为正规矩阵. 所以 $\mathscr{A}|_U$ 和 $\mathscr{A}|_{U^\perp}$ 都是正规变换 (思考题 7.7). □

引理 7.2.31: *对于 (有限维非零实向量空间) V 上的任意线性变换 $\mathscr{A}$, 一定存在 $\mathscr{A}$ 的不变子空间 $U \subseteq V$ 满足 $\dim U = 1$ 或 $\dim U = 2$.*

证明. 如果 $\mathscr{A}$ 有 (实) 特征值, 则任一特征向量张成 $\mathscr{A}$ 的一个 1 维不变子空间. 以下不妨设 $\mathscr{A}$ 没有特征值. 于是 $\mathscr{A}$ 的特征多项式 $f(X) = \det(XI - A)$ 在 $\mathbb{R}[X]$ 中具有如下分解式

$$f(X) = (X^2 + b_1 X + c_1)\cdots(X^2 + b_M X + c_M),$$

其中 $M \ge 1$, b_i, $c_i \in \mathbb{R}$ 满足 $b_i^2 - 4c < 0$. 由 Cayley–Hamilton 定理,

$$(\mathscr{A}^2 + b_1\mathscr{A} + c_1 I)\cdots(\mathscr{A}^2 + b_M\mathscr{A} + c_M I) = f(\mathscr{A}) = 0.$$

由此不妨设 $\mathscr{A}^2 + b_1\mathscr{A} + c_1 I$ 不可逆. 于是存在非零向量 $v \in V$ 使得 $(\mathscr{A}^2 + b_1\mathscr{A} + c_1 I)v = 0$. 此时可以轻易验证 $U = \mathrm{span}(v, \mathscr{A}v)$ 是 $\mathscr{A}$ 的一个 (循环) 不变子空间. □

引理 7.2.32: *假设 V 是 2 维实内积空间, $\mathscr{A} \in \mathrm{End}(V)$. 则下列条件等价:*

1. *$\mathscr{A}$ 是正规算子但不是自伴算子.*
2. *对 V 的任意一组有序规范正交基 $\mathcal{B}$, 矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 形如 $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, 其中 $a, b \in \mathbb{R}$ 且 $b \ne 0$.*
3. *存在 V 的一组有序规范正交基 $\mathcal{B}$, 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 形如 $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, 其中 $a, b \in \mathbb{R}$ 且 $b \ne 0$.*
4. *存在 V 的一组有序规范正交基 $\mathcal{B}$, 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 形如 $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, 其中 $a, b \in \mathbb{R}$ 且 $b > 0$.*
5. *存在 V 的一组有序规范正交基 $\mathcal{B}$, 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 形如 $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, 其中 $a, b \in \mathbb{R}$ 且 $b < 0$.*

证明. 通过分析 2 阶实正规矩阵的矩阵元素特点, 不难证明 (1), (2), (3) 之间的等价性, 细节留给读者练习.

若 $\mathcal{B}=(e_1, e_2)$ 是一组有序规范正交基使得 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})=\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, 则在 $\mathcal{C}=(e_2, e_1)$ 这组新的有序规范正交基下 $\mathcal{M}_{\mathcal{C}}(\mathscr{A})=\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. 由此可得 (3), (4), (5) 互相等价. □

定理 7.2.33 (**实正规算子的谱定理** (spectral theorem for real normal operators)): 设 $\mathscr{A}$ 是 (有限维非零实内积空间) V 上的线性算子. 则下列陈述等价:

1. $\mathscr{A}$ 是正规算子.
2. 存在正交和分解 $V=U_1\perp\cdots\perp U_m\perp W_1\perp\cdots\perp W_r$, 其中每个 U_i 是 $\mathscr{A}$ 的 2 维不变子空间, 且 $\mathscr{A}|_{U_i}$ 均为非自伴的正规算子, 而 W_j 都是 $\mathscr{A}$ 的 1 维不变子空间. (这里 m 或 r 可能为 0.)
3. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有如下分块对角形式
$$\mathcal{M}_{\mathcal{B}}(\mathscr{A})=\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$
其中 $\lambda_1,\cdots,\lambda_r\in\mathbb{R}$, $A_i=\begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix}\in\mathbf{M}_2(\mathbb{R})$ 且 $b_i\neq 0$. (这里 m 或 r 可能为 0.)
4. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有如下分块对角形式
$$\mathcal{M}_{\mathcal{B}}(\mathscr{A})=\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$
其中 $\lambda_1,\cdots,\lambda_r\in\mathbb{R}$, $A_i=\begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix}\in\mathbf{M}_2(\mathbb{R})$ 且 $b_i>0$. (这里 m 或 r 可能为 0.)
5. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有如下分块对角形式
$$\mathcal{M}_{\mathcal{B}}(\mathscr{A})=\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$
其中 $\lambda_1,\cdots,\lambda_r\in\mathbb{R}$, $A_i=\begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix}\in\mathbf{M}_2(\mathbb{R})$ 且 $b_i<0$. (这里 m 或 r 可能为 0.)

证明. (1) $\Rightarrow$ (2). 我们采用对 $\dim V$ 的归纳法. 若 $\dim V = 1$, 则结论显然成立. 若 $\dim V = 2$, 则 $\mathscr{A}$ 自伴时可由定理 7.2.23 (又可参见 (7.2.28) 一段) 知道 V 是两个 1 维不变子空间的正交和, 而 $\mathscr{A}$ 不自伴时结论由引理 7.2.32 保证.

以下设 $\dim V > 2$. 根据引理 7.2.31, 存在 $\mathscr{A}$ 的 1 维或 2 维不变子空间 $U \subseteq V$. 如果 $\dim U = 2$, 我们可以假设 $\mathscr{A}|_U$ 不是自伴的 (否则根据命题 7.2.25 可知 U 中还存在 $\mathscr{A}$ 的 1 维不变子空间). 考虑正交和分解 $V = U \perp U^\perp$. 由引理 7.2.30 可知, $\mathscr{A}_1 := \mathscr{A}|_U$ 和 $\mathscr{A}_2 := \mathscr{A}|_{U^\perp}$ 均为正规算子. 根据归纳假设, $U^\perp$ 关于 $\mathscr{A}_2$ 可以分解为一些 1 维或 2 维不变子空间的正交和, 而且 $\mathscr{A}_2$ 在其中的 2 维不变子空间上限制为非自伴的正规算子. 由于 $\dim U \leq 2$, 所以 U 关于 $\mathscr{A}_1$ 也有这样的正交和分解. 将 U 和 $U^\perp$ 分别所得的正交和分解合并写出, 就得到 V 所需的正交和分解.

(2) $\Rightarrow$ (3). 取 U_i 和 W_j 的规范正交基合并成 V 的一组规范正交基 $\mathcal{B}$, 结合引理 7.2.32 即得.

(3) $\Rightarrow$ (1). 根据思考题 7.7, 只要验证 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是正规矩阵即可.

(3), (4) 和 (5) 的等价性证明类似于引理 7.2.32 中相应结论的证明. □

将定理 7.2.33 以矩阵形式重述, 可以得到:

定理 7.2.34 (**实正规矩阵的正交相似标准形**): 设 $A \in \mathbf{M}_n(\mathbb{R})$. 则下列陈述等价:

1. A 是正规矩阵.

2. A 正交相似于一个如下形式的分块对角阵

$$\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$

其中 $\lambda_1, \cdots, \lambda_r \in \mathbb{R}$, $A_i = \begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix} \in \mathbf{M}_2(\mathbb{R})$ 且 $b_i \neq 0$. (这里 m 或 r 可能为 0.)

3. A 正交相似于一个如下形式的分块对角阵

$$\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$

其中 $\lambda_1, \cdots, \lambda_r \in \mathbb{R}$, $A_i = \begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix} \in \mathbf{M}_2(\mathbb{R})$ 且 $b_i > 0$. (这里 m 或 r 可能为 0.)

4. A 正交相似于一个如下形式的分块对角阵

$$\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$

其中 $\lambda_1,\cdots,\lambda_r\in\mathbb{R}$, $A_i=\begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix}\in\mathbf{M}_2(\mathbb{R})$ 且 $b_i<0$. (这里 m 或 r 可能为 0.)

思考题 7.11. 设 $A\in\mathbf{M}_n(\mathbb{R})$ 为正规矩阵. 证明: 定理 7.2.34 论断 (3) 中的分块对角阵在不计对角块排序差异的情况下是由 A 唯一确定的. 类似地, 该定理论断 (4) 中的分块对角阵也具有这种唯一性. (提示: 考虑特征多项式.)

因此, 我们可以将定理 7.2.34 (3) 或 (4) 中的矩阵称为正规矩阵 A 的**正交相似标准形**.

有了一般正规算子的矩阵标准形之后, 我们可以很容易地得出关于斜对称算子和正交算子的相应结论.

命题 7.2.35: 设 $\mathscr{A}$ 是 (有限维非零实内积空间) V 上的线性算子. 则下列陈述等价:

1. $\mathscr{A}$ 是斜对称算子.
2. 存在正交和分解 $V=U_1\perp\cdots\perp U_m\perp W_1\perp\cdots\perp W_r$, 其中每个 U_i 是 $\mathscr{A}$ 的 2 维不变子空间, 且 $\mathscr{A}|_{U_i}$ 均为非零斜对称算子, 而 W_j 都是 $\mathscr{A}$ 的 1 维不变子空间, 且 $\mathscr{A}|_{W_j}=0$.

   (这里 m 或 r 可能为 0.)
3. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有如下分块对角形式

$$\mathcal{M}_{\mathcal{B}}(\mathscr{A})=\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

   其中 $A_i=\begin{pmatrix} 0 & -b_i \\ b_i & 0 \end{pmatrix}\in\mathbf{M}_2(\mathbb{R})$ 且 $b_i\neq 0$. (这里可能 $m=0$ 或 $2m=\dim V$.)
4. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有如下分块对角形式

$$\mathcal{M}_{\mathcal{B}}(\mathscr{A})=\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

其中 $A_i = \begin{pmatrix} 0 & -b_i \\ b_i & 0 \end{pmatrix} \in \mathbf{M}_2(\mathbb{R})$ 且 $b_i > 0$. (这里可能 $m = 0$ 或 $2m = \dim V$.)

5. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有如下分块对角形式

$$\mathcal{M}_{\mathcal{B}}(\mathscr{A}) = \begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

其中 $A_i = \begin{pmatrix} 0 & -b_i \\ b_i & 0 \end{pmatrix} \in \mathbf{M}_2(\mathbb{R})$ 且 $b_i < 0$. (这里可能 $m = 0$ 或 $2m = \dim V$.)

证明. 留给读者练习. □

思考题 7.12. 证明命题 7.2.35, 并且陈述它的矩阵形式.

推论 7.2.36: 若 $\mathscr{A}$ 为 (有限维非零实内积空间) V 上的斜对称算子, 则 $\operatorname{rank}(\mathscr{A})$ 必为偶数.

以矩阵语言言之, 若 $A \in \mathbf{M}_n(\mathbb{R})$ 是斜对称矩阵, 则 $\operatorname{rank}(A)$ 必为偶数.

证明. 由命题 7.2.35 中的论断 (3), (4) 或 (5) 可得. □

定理 7.2.37 (**实正交算子的谱定理** (spectral theorem for real orthogonal operators)): 设 $\mathscr{A}$ 是 (有限维非零实内积空间) V 上的线性算子. 则下列陈述等价:

1. $\mathscr{A}$ 是正交算子.

2. 存在正交和分解 $V = U_1 \perp \cdots \perp U_m \perp W_1 \perp \cdots \perp W_r$, 其中每个 U_i 是 $\mathscr{A}$ 的 2 维不变子空间, 且 $\mathscr{A}|_{U_i}$ 均为非自伴的正交算子, 而 W_j 都是 $\mathscr{A}$ 的 1 维不变子空间, 且 $\mathscr{A}|_{W_j} = \pm I$.

   (这里 m 或 r 可能为 0.)

3. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有如下分块对角形式

$$\mathcal{M}_{\mathcal{B}}(\mathscr{A}) = \begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$

其中 $\lambda_1, \cdots, \lambda_r \in \{\pm 1\}$, $A_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$ 且 $\theta_i \in \mathbb{R}$ 不是 π 的整数倍 (即 $\sin\theta_i \neq 0$). (这里 m 或 r 可能为 0.)

4. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有如下分块对角形式

$$\mathcal{M}_{\mathcal{B}}(\mathscr{A})=\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$

其中 $\lambda_1,\cdots,\lambda_r\in\{\pm1\}$, $A_i=\begin{pmatrix}\cos\theta_i & -\sin\theta_i\\ \sin\theta_i & \cos\theta_i\end{pmatrix}$ 且 $0<\theta_i<\pi$. (这里 m 或 r 可能为 0.)

5. 存在 V 的一组有序规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 具有如下分块对角形式

$$\mathcal{M}_{\mathcal{B}}(\mathscr{A})=\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$

其中 $\lambda_1,\cdots,\lambda_r\in\{\pm1\}$, $A_i=\begin{pmatrix}\cos\theta_i & \sin\theta_i\\ -\sin\theta_i & \cos\theta_i\end{pmatrix}$ 且 $0<\theta_i<\pi$. (这里 m 或 r 可能为 0.)

证明. 留给读者练习. 或课堂授课时讲授. $\square$

定理 7.2.38 (实正交矩阵的正交相似标准形): 设 $A\in\mathbf{M}_n(\mathbb{R})$. 则下列陈述等价:

1. A 是正交矩阵.
2. A 正交相似于一个如下形式的分块对角阵

$$\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$

其中 $\lambda_1,\cdots,\lambda_r\in\{\pm1\}$, $A_i=\begin{pmatrix}\cos\theta_i & -\sin\theta_i\\ \sin\theta_i & \cos\theta_i\end{pmatrix}$ 且 $\theta_i\in\mathbb{R}$ 不是 π 的整数倍 (即 $\sin\theta_i\neq0$). (这里 m 或 r 可能为 0.)

3. A 正交相似于一个如下形式的分块对角阵

$$\begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_m & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$

其中 $\lambda_1,\cdots,\lambda_r\in\{\pm1\}$, $A_i=\begin{pmatrix}\cos\theta_i & -\sin\theta_i\\ \sin\theta_i & \cos\theta_i\end{pmatrix}$ 且 $0<\theta_i<\pi$. (这里 m 或 r 可能为 0.)

4. A 正交相似于一个如下形式的分块对角阵

$$\begin{pmatrix}A_1 &&&&&\\ &\ddots&&&&\\ &&A_m&&&\\ &&&\lambda_1&&\\ &&&&\ddots&\\ &&&&&\lambda_r\end{pmatrix}$$

其中 $\lambda_1,\cdots,\lambda_r\in\{\pm1\}$, $A_i=\begin{pmatrix}\cos\theta_i & \sin\theta_i\\ -\sin\theta_i & \cos\theta_i\end{pmatrix}$ 且 $0<\theta_i<\pi$. (这里 m 或 r 可能为 0.)

证明. 留给读者练习. □

推论 7.2.39: 设 $A\in\mathbf{O}_n(\mathbb{R})$ (即 A 为 n 阶实正交阵). 则对于 A 在 $\mathbb{C}$ 中的任意特征值 λ, 均有 $|\lambda|=1$. 特别地, A 的实特征值只能是 ±1.

证明. 通过定理 7.2.38 中的正交相似标准形计算 A 的特征多项式即可. □

(7.2.40) 我们现在来说明正交矩阵的正交相似标准形以及相应的 (正交) 过渡矩阵具体如何计算.

设 $A\in\mathbf{O}_n(\mathbb{R})$, $f(X)$ 为 A 的特征多项式. 按照定理 7.2.38 (4) 中的正交相似标准形计算特征多项式可得

$$f(X)=(X-\lambda_1)\cdots(X-\lambda_r)\prod_{i=1}^{s}\left(X^2-2\cos\theta_i\cdot X+1\right)^{m_i} \tag{7.2.40.1}$$

其中 $\lambda_1,\cdots,\lambda_r\in\{\pm1\}$, $m_i\in\mathbb{N}^*$, 而 $\theta_i\in]0,\pi[$ 两两不同.

如果正交矩阵 $Q\in\mathbf{O}_n(\mathbb{R})$ 能使 $Q^{-1}AQ$ 形如定理 7.2.38 (4) 中的标准形, 我们可以将 Q 的列向量组分组排列如下:

$$\begin{aligned}&(u_{11},\,v_{11}),(u_{12},\,v_{12}),\cdots,(u_{1m_1},\,v_{1m_1});\\&(u_{21},\,v_{21}),(u_{22},\,v_{22}),\cdots,(u_{2m_2},\,v_{2m_2});\\&\cdots\cdots\cdots\cdots\cdots\\&(u_{s1},\,v_{s1}),(u_{s2},\,v_{s2}),\cdots,(u_{sm_s},\,v_{sm_s});\\&f_1,\cdots,f_r\end{aligned}$$

其中 $f_1,\cdots,f_r$ 是分别属于特征值 $\lambda_1,\cdots,\lambda_r$ 的单位特征向量, 而每一对向量 $(u_{ij},\,v_{ij})$ 满足

$$(Au_{ij}\,,\,Av_{ij})=(u_{ij},\,v_{ij})\begin{pmatrix}\cos\theta_i & \sin\theta_i\\ -\sin\theta_i & \cos\theta_i\end{pmatrix}. \tag{7.2.40.2}$$

特征向量 $f_1,\cdots,f_r$ 的计算方法无需赘述. 以下解释 $(u_{ij},\,v_{ij})$ 这些向量对如何找出即可.

我们可以取定一个 i, 记 $M=m_i$, 并将记号 u_{ij}, v_{ij} 简化为 u_j, v_j, 其中 j 取遍 $1,\cdots,M$. 它们对应特征多项式中的一个二次因子 $X^2-2\cos\theta\cdot X+1$, 该二次因子出现的重数为 M.

为了找出这 M 对向量 $(u_j,\,v_j)$, 首先注意到从 (7.2.40.2) 式可以得到

$$(A^2-(2\cos\theta)A+I)u_j=0\,. \tag{7.2.40.3}$$

因此, 我们可以先在 $A^2-(2\cos\theta)A+I$ 的零化空间中找出一个单位向量 u_1, 然后通过公式

$$Au_1=(\cos\theta)u_1+(-\sin\theta)v_1 \tag{7.2.40.4}$$

即可求出相应的 v_1. 注意这里 $\sin\theta\neq 0$, 故 v_1 可以由 u_1 通过 (7.2.40.4) 式唯一确定. 而且这样求出的一对 $(u_1,\,v_1)$ 确实满足 (7.2.40.2) 式, 因为由

$$\begin{aligned}
(\text{根据 (7.2.40.4)})\ \ (\sin\theta)Av_1 &= (\cos\theta)Au_1-A^2u_1\\
(\text{根据 (7.2.40.3)})\ \ &= (\cos\theta)Au_1-(2\cos\theta)Au_1+u_1=-(\cos\theta)Au_1+u_1\\
(\text{再根据 (7.2.40.4)})\ \ &= -(\cos\theta)^2u_1+(\cos\theta\sin\theta)v_1+u_1=\sin\theta((\sin\theta)u_1+(\cos\theta)v_1)
\end{aligned}$$

可以得到 $Av_1=(\sin\theta)u_1+(\cos\theta)v_1$. 由于

$$\begin{aligned}
\|v_1\|^2=\|Av_1\|^2=\|(\sin\theta)u_1+(\cos\theta)v_1\|^2&=(\sin\theta)^2\|u_1\|^2+(\cos\theta)^2\|v_1\|^2+2\sin\theta\cos\theta\langle\, u_1,\,v_1\,\rangle\,,\\
\langle\, u_1,\,v_1\,\rangle=\langle\, Au_1,\,Av_1\,\rangle&=\langle\,(\cos\theta)u_1+(-\sin\theta)v_1\,,\,(\sin\theta)u_1+(\cos\theta)v_1\,\rangle\\
&=\sin\theta\cos\theta(\|u_1\|^2-\|v_1\|^2)-(\sin\theta)^2\langle\, u_1,\,v_1\,\rangle+(\cos\theta)^2\langle\, u_1,\,v_1\,\rangle
\end{aligned}$$

(这里 $\|u_1\|^2=1$), 可以解出 $\|v_1\|^2=1$, $\langle\, u_1,\,v_1\,\rangle=0$. 也就是说, 以上找到的 u_1, v_1 的确构成规范正交组, 并且满足 (7.2.40.2).

接下来, (如果 $M>1$) 我们再从 $A^2-(2\cos\theta)A+I$ 的零化空间中找出另一个单位向量 u_2, 使得 u_2 和前面的 u_1, v_1 都正交. 然后再按刚才的方法, 由公式 $Au_2=(\cos\theta)u_2+(-\sin\theta)v_2$ 找出 v_2. 照此反复操作, 最终就可以逐一找到 $(u_1,\,v_1),\cdots,(u_M,\,v_M)$ 这些向量对. ■

例 7.2.41. 设 $A=\begin{pmatrix}\frac12&\frac12&\frac12&\frac12\\ \frac12&\frac12&-\frac12&-\frac12\\ -\frac12&\frac12&-\frac12&\frac12\\ -\frac12&\frac12&\frac12&-\frac12\end{pmatrix}$. 试求正交矩阵 $Q\in\mathbf{O}_4(\mathbb{R})$ 使 $Q^{-1}AQ$ 成为标准形.

首先计算 A 的特征多项式得到 $f(X)=X^4-1=(X-1)(X+1)(X^2+1)$.

对于实特征值 $\lambda_1=1$ 和 $\lambda_2=-1$ 分别在特征子空间中找出单位特征向量, 得到

$$f_1=\left(\frac{1}{\sqrt2}\,,\,\frac{1}{\sqrt2}\,,\,0\,,\,0\right)^T\,,\quad f_2=\left(0\,,\,0\,,\,-\frac{1}{\sqrt2}\,,\,\frac{1}{\sqrt2}\right)^T.$$

接下来我们在 A^2+I 的零化空间中找出一个单位向量 u. 计算可得

$$A^2+I=\begin{pmatrix}1&1&&\\1&1&&\\&&1&-1\\&&-1&1\end{pmatrix}$$

于是 u 可取为 $\left(-\frac12\,,\,\frac12\,,\,\frac12\,,\,\frac12\right)^T$.

因为此处对应的 $\theta=\frac{\pi}{2}$, 故由 (7.2.40.4) 式得到向量 $v=-Au=\left(\frac12\,,\,-\frac12\,,\,\frac12\,,\,\frac12\right)^T$.

以 u, v, f_1, f_2 为列向量得到的矩阵 $Q=\begin{pmatrix}-\frac12&\frac12&\frac{1}{\sqrt2}&0\\ \frac12&-\frac12&\frac{1}{\sqrt2}&0\\ \frac12&\frac12&0&-\frac{1}{\sqrt2}\\ \frac12&\frac12&0&\frac{1}{\sqrt2}\end{pmatrix}$ 满足所求. ■

7.2.5　习题

若非特别说明, 以下 V, W 总是表示非零的有限维内积空间.

习题 7.2.1. 在定义 7.2.1 中我们定义等距映射时预先要求是线性映射. 这个习题中我们证明: 将零向量映射为零向量、且保持距离不变的映射一定是线性映射.

设映射 $\mathscr{A}: V \to W$ 满足 $\mathscr{A}(0) = 0$ 且

$$\text{对于任意 } u\,,\, v \in V\,,\ \|\mathscr{A}u - \mathscr{A}v\| = \|u - v\|\,.$$

证明:

1. $\mathscr{A}$ 保持向量长度不变, 即, 对任意 $v \in V$, $\|\mathscr{A}v\| = \|v\|$.
2. $\mathscr{A}$ 保持向量内积不变, 即, 对于任意 $u, v \in V$, $\langle \mathscr{A}u\,,\, \mathscr{A}v \rangle = \langle u\,,\, v \rangle$. (*注意: 在不知道 $\mathscr{A}$ 线性的时候, 是否可以从极化恒等式得到想要的结论?*)
3. $\mathscr{A}$ 一定是线性映射. (*提示: 考虑 $\|\mathscr{A}(u+v) - \mathscr{A}u - \mathscr{A}v\|$.*)

习题 7.2.2. 设 V 是 n 维内积空间.

1. 对任意非零向量 $x \in V$, 定义映射
$$\tau_x \;:\; V \longrightarrow V\,; \quad v \longmapsto v - 2\frac{\langle v,\, x \rangle}{\|x\|^2}x\,.$$
证明: τ_x 是个线性变换, 并且是第二类正交变换. 能够写成 τ_x 这种形式的线性变换称为**镜面反射** (mirror reflection) 或**镜像变换** (mirror transformation).
2. 证明: 若 $\mathscr{A} \in \mathrm{End}(V)$ 是个镜面反射, 则 $\mathscr{A}^2 = I$.
3. 假设 $\mathscr{A} \in \mathrm{End}(V)$ 是正交变换, 1 是它的一个特征值, 并且满足 $\dim E(1, \mathscr{A}) = n - 1$. 证明: $\mathscr{A}$ 是个镜面反射.

习题 7.2.3. 假设 $\mathscr{A} \in \mathrm{End}(V)$ 是个镜面反射.

1. 证明: 映射
$$\varphi \;:\; V \times V \longrightarrow \mathbb{R}\,; \quad (u,\, v) \longmapsto \langle \mathscr{A}u\,,\, v \rangle$$
是 V 上的对称双线性型.
2. 证明: 若 $\mathscr{B} \in \mathrm{End}(V)$ 是正交变换, 则 $\mathscr{B}^{-1}\mathscr{A}\mathscr{B}$ 仍是镜面反射.

习题 7.2.4. 设 u, v 是内积空间 V 中两个不同的向量, 且 $\|u\| = \|v\|$.

证明: 存在一个镜面反射 $\mathscr{A}$ 使得 $\mathscr{A}u = v$.

习题 7.2.5. (Cartan–Dieudonné 定理) 设 V 是 n 维内积空间. 证明: 对任意正交变换 $\mathscr{A} \in \mathrm{End}(V)$, 存在镜面反射 $\tau_1, \cdots, \tau_m$, 其中 $0 \le m \le n$, 使得 $\mathscr{A} = \tau_m \cdots \tau_2\tau_1$.

(当 $m = 0$ 时我们认为 $\tau_m \cdots \tau_2\tau_1 = \mathrm{Id}$.)

习题 7.2.6. 设 $M \in \mathbf{M}_n(\mathbb{R})$. 如果存在长度为 1 的列向量 $\alpha \in \mathbb{R}^n$ 使得 $M = I_n - 2\alpha\alpha^T$, 则称 M 是个**镜像矩阵** (mirror matrix).

证明: 对于 n 维内积空间 V 上的任意线性变换 $\mathscr{A}$, 下列条件等价:

1. $\mathscr{A}$ 是镜像变换 (参见习题 7.2.2).
2. 对于 V 的任意一组规范正交基 $\mathcal{E}$, $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 是镜像矩阵.
3. 存在 V 的一组规范正交基 $\mathcal{E}$ 使得 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 是镜像矩阵.

习题 7.2.7. 设 $u_1, \cdots, u_r$ 和 $v_1, \cdots, v_r$ 是内积空间 V 中的两组向量. 证明下列陈述等价:

1. 存在 $\mathscr{A} \in \mathbf{O}(V)$ 使得 $\mathscr{A} u_i = v_i$ 对每个 $i \in [\![1, r]\!]$ 成立.
2. 对任意 $i, j \in [\![1, r]\!]$ 均有 $\langle u_i, u_j \rangle = \langle v_i, v_j \rangle$.

习题 7.2.8. 假设 $\mathscr{A} \in \mathrm{End}(V)$ 是可逆变换. 证明: $\mathscr{A}^*$ 也是可逆变换, 并且 $(\mathscr{A}^*)^{-1} = (\mathscr{A}^{-1})^*$.

习题 7.2.9. 设 $\mathscr{A} \in \mathrm{End}(V)$, U 是 V 的子空间. 证明: U 是 $\mathscr{A}$ 的不变子空间当且仅当 $U^\perp$ 是 $\mathscr{A}^*$ 的不变子空间.

习题 7.2.10. 设 $\mathcal{E} = (e_1, e_2, \cdots, e_n)$ 是内积空间 V 中的一组规范正交基. 定义线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 为 $v \mapsto \mathscr{A} v = \langle v, e_1 \rangle e_2$ (参见例 7.2.9).

求矩阵 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 和 $\mathcal{M}_{\mathcal{E}}(\mathscr{A}^*)$.

习题 7.2.11. 设 $V = \mathbf{M}_n(\mathbb{R})$. 考虑 V 上的 Frobenius 内积 (参见习题 7.1.7):

$$(A, B) \in V \times V \longmapsto \langle A, B \rangle := \mathrm{Tr}(A^T B)\,.$$

给定矩阵 $M \in \mathbf{M}_n(\mathbb{R})$, 定义 $\varphi \in \mathrm{End}(V)$ 为 $\varphi(A) = MA$. 求伴随变换 φ^* 的表达式.

习题 7.2.12. 设 $V = \mathbb{R}[X]_{\le 2}$. 按照如下方式定义 V 上的内积:

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

定义线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 为 $\mathscr{A}(a_0 + a_1 X + a_2 X^2) = a_1 X$.

1. 求 $\mathscr{A}$ 在有序基 $\mathcal{B} = (1, X, X^2)$ 下的矩阵.
2. $\mathscr{A}$ 是否是自伴变换? 为什么?

习题 7.2.13. 对下列对称阵 A, 求正交矩阵 Q 使 $Q^{-1}AQ$ 成为对角阵:

$$A = \begin{pmatrix} 2 & -2 & 0 \\ -2 & 1 & -2 \\ 0 & -2 & 0 \end{pmatrix}\ ;\ A = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 5 & -4 \\ -2 & -4 & 5 \end{pmatrix}\ ;$$
$$A = \begin{pmatrix} 0 & 0 & 4 & 1 \\ 0 & 0 & 1 & 4 \\ 4 & 1 & 0 & 0 \\ 1 & 4 & 0 & 0 \end{pmatrix}\ ;\ A = \begin{pmatrix} -1 & -3 & 3 & -3 \\ -3 & -1 & -3 & 3 \\ 3 & -3 & -1 & -3 \\ -3 & 3 & -3 & -1 \end{pmatrix}\ ;\ A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}\,.$$

习题 7.2.14. 设 $\mathscr{A}, \mathscr{B} \in \mathrm{End}(V)$. 证明或举出反例:

1. 若 $\mathscr{A}, \mathscr{B}$ 都是自伴算子, 则 $\mathscr{A} + \mathscr{B}$ 仍是自伴算子.
2. 若 $\mathscr{A}, \mathscr{B}$ 都是斜对称算子, 则 $\mathscr{A} + \mathscr{B}$ 仍是斜对称算子.

3. 若 $\mathscr{A}$, $\mathscr{B}$ 都是正规算子, 则 $\mathscr{A}+\mathscr{B}$ 仍是正规算子.

4. 若 $\mathscr{A}$, $\mathscr{B}$ 都是自伴算子, 则 $\mathscr{A}\mathscr{B}$ 仍是自伴算子.

习题 7.2.15. 设 $\mathscr{A}\in\mathrm{End}(V)$. 证明或举出反例:

如果存在 V 的一组规范正交基 $e_1,\cdots,e_n$ 使得对于每个 $j\in[\![1,n]\!]$ 均有 $\|\mathscr{A}e_j\|=\|\mathscr{A}^*e_j\|$, 则 $\mathscr{A}$ 是正规的.

习题 7.2.16. 取定 $u, x\in V$. 定义线性变换 $\mathscr{A}\in\mathrm{End}(V)$ 为 $v\mapsto\mathscr{A}v=\langle v,u\rangle x$. 证明下列陈述等价:

1. $\mathscr{A}$ 是自伴的.

2. $\mathscr{A}$ 是正规的.

3. 向量组 u, x 线性相关.

习题 7.2.17. 举例说明: 如果不假设 $\mathscr{A}\in\mathrm{End}(V)$ 是正规算子, 那么有可能存在子空间 U 是 $\mathscr{A}$ 的不变子空间, 但 $U^\perp$ 不是 $\mathscr{A}$ 的不变子空间.

习题 7.2.18. 设 $\mathscr{A}$ 是 $\mathbb{R}^n$ 上的正规变换, 且 $\mathscr{A}(e_1+\cdots+e_n)=2(e_1+\cdots+e_n)$ (其中 $e_1,\cdots,e_n$ 为 $\mathbb{R}^n$ 的标准基).

证明: 对于任意 $x=(x_1,\cdots,x_n)^T\in\mathrm{Ker}(\mathscr{A})$, 均有 $x_1+\cdots+x_n=0$.

习题 7.2.19. 设 $\mathscr{A}\in\mathrm{End}(V)$ 为正规变换. 证明: $\mathrm{Im}(\mathscr{A})=\mathrm{Im}(\mathscr{A}^*)$.

习题 7.2.20. 设 $A, B\in\mathbf{M}_n(\mathbb{R})$ 均为对称阵. 证明: A 和 B 正交相似的充分必要条件是 A 和 B 的特征多项式相同.

习题 7.2.21. 设 $A, B\in\mathbf{M}_n(\mathbb{R})$ 均为对称阵, 其中 A 正定. 证明: 存在可逆矩阵 $P\in\mathbf{M}_n(\mathbb{R})$ 使得 P^TAP 和 P^TBP 同时为对角阵. (提示: 先考虑 $A=I_n$ 的情况.)

习题 7.2.22. 证明: 如果 $A\in\mathbf{M}_n(\mathbb{R})$ 是斜对称阵, 那么 A 在 $\mathbb{C}$ 中的特征值都是纯虚数 (即, 可以写成 bi 的形式, 其中 $b\in\mathbb{R}$).

习题 7.2.23. 设 $b, c\in\mathbb{R}$ 且 $b^2-4c<0$. 举例说明: 存在线性变换 $\mathscr{A}\in\mathrm{End}(V)$ 使得 $\mathscr{A}^2+b\mathscr{A}+cI$ 不可逆.

(这说明: 引理 7.2.24 如果不假设 $\mathscr{A}$ 自伴, 那么结论不再成立.)

习题 7.2.24. 设 $\mathscr{A}$, $\mathscr{B}$ 是 n 维内积空间 V 上的两个自伴算子. 证明下列陈述等价:

1. 存在 V 的一组规范正交基 $\mathcal{E}$ 使得 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 和 $\mathcal{M}_{\mathcal{E}}(\mathscr{B})$ 同时为对角阵.

2. $\mathscr{A}$ 和 $\mathscr{B}$ 可交换.

习题 7.2.25. 设 $\mathscr{A}$ 为 $\mathbb{R}^3$ 上的 (线性) 正交变换. 证明: 存在非零向量 $x\in\mathbb{R}^3$ 使得 $\mathscr{A}^2x=x$.

习题 7.2.26. 设 A 为 n 阶实正规矩阵. 假设 $\lambda\in\mathbb{C}$ 是 A 的一个复特征值, 列向量 $\alpha\in\mathbb{C}^n$ 满足 $A\alpha=\lambda\alpha$. 证明: $A^T\alpha=\bar{\lambda}\alpha$.

习题 7.2.27. 设 $\mathscr{A}\in\mathrm{End}(V)$ 是正规变换且最小多项式为 $g(X)=(X-a)^2+b^2$, 其中 $a, b\in\mathbb{R}$, $b\neq 0$.

证明: $\mathscr{A}$ 是可逆变换, 且 $\mathscr{A}^*=(a^2+b^2)\mathscr{A}^{-1}$.

习题 7.2.28. 设 $A=\begin{pmatrix}\frac{2}{3} & -\frac{1}{3} & \frac{2}{3}\\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3}\\ -\frac{1}{3} & \frac{2}{3} & \frac{2}{3}\end{pmatrix}$. 试求正交矩阵 $Q\in\mathbf{O}_3(\mathbb{R})$ 使 $Q^{-1}AQ$ 成为标准形.

习题 7.2.29. 设 $\mathcal{C}[-\pi, \pi]$ 是闭区间 $[-\pi, \pi]$ 上所有的实值连续函数构成的空间, 在其上定义内积如下:

$$\langle f, g \rangle := \int_{-\pi}^{\pi} f(x)g(x)\mathrm{d}x \ .$$

取

$$V = \mathrm{span}(1, \cos(x), \cos(2x), \cos(3x), \sin(x), \sin(2x), \sin(3x)).$$

1. 证明: 对于任意 $f \in V$, 其导函数 f' 也属于 V.
2. 定义 V 上的线性变换 $\mathscr{A} : V \to V$, $f \mapsto f'$. 证明 $\mathscr{A}$ 是斜对称变换.
3. 求 V 的一组规范正交基 $\mathcal{E}$, 使得矩阵 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 具有正交相似标准形 (即命题 7.2.35 的结论 (3), (4) 或 (5) 中的形式).

习题 7.2.30. 设 $A, B \in \mathbf{M}_n(\mathbb{R})$, 其中 A 正定. 证明:

1. 对任意正整数 k, A^k 也正定.
2. 如果存在正整数 r 使得 B 与 A^r 可以交换, 则 B 和 A 也可交换.

(提示: 利用正交相似标准形.)

习题 7.2.31. 设 U, V 分别为 n 维、m 维内积空间, $W = \mathrm{Hom}(U, V)$. 取定 U 的一组规范正交基 $\varepsilon_1, \cdots, \varepsilon_n$.

1. 证明: 映射

$$\langle \cdot, \cdot \rangle_W \ : \ W \times W \ ; \quad (f, g) \longmapsto \sum_{i=1}^{n} \langle f(\varepsilon_i), g(\varepsilon_i) \rangle_V$$

是 W 上的一个内积.

2. 按照上一小题定义的内积将 W 视为内积空间. 对任意 $\mathscr{A} \in \mathrm{End}(U)$, 定义映射

$$T(\mathscr{A}) \ : \ W \longrightarrow W \ ; \quad f \longmapsto \big(T(\mathscr{A})(f) : x \in U \longmapsto f(\mathscr{A}x) \in V\big) \ .$$

证明: $T(\mathscr{A})$ 是 W 上的线性变换.

再证明: $T(\mathscr{A})$ 是 W 上的正交变换当且仅当 $\mathscr{A}$ 是 U 上的正交变换.

7.3 一些应用和补充

7.3.1 二次曲线和二次曲面的度量分类

过去我们知道实对称矩阵都可以相合到对角阵之后, 就可以对平面二次曲线和空间二次曲面进行仿射分类. 现在我们想利用实对称矩阵可以正交相合于对角阵的性质, 来对二次曲线和二次曲面进行更细致的分类.

我们先引入以下定义.

定义 7.3.1. 设 α 是仿射空间 $\mathbb{A}^n$ 到自身的映射. 如果存在正交矩阵 $A \in \mathbf{O}_n(\mathbb{R})$ 和向量 $b \in \mathbb{R}^n$ 使得 α 由以下公式给出

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \longmapsto A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + b$$

则称 α 是仿射空间 $\mathbb{A}^n$ 的一个**等距变换** (isometry) 或**正交变换** (orthogonal transformation), 物理学中也常称之为一个**刚体运动** (rigid motion).

如果上式中 $b=0$, 则称 α 是一个**线性等距** (linear isometry). 如果上式中 $A=I_n$, 则 α 是定义 6.3.3 意义下的平移变换.

根据定义, 等距变换都是定义 6.3.3 意义下的仿射变换, 恒等变换是一个等距变换. 而且由思考题 6.18 (1) 可知, 等距变换都是可逆变换, 其逆变换仍是等距变换.

对于 $\mathbb{A}^n$ 中的两个图形 F 和 F', 如果存在等距变换 $\alpha:\mathbb{A}^n\to\mathbb{A}^n$ 使得 $\alpha(F)=F'$, 则称 F 和 F' **度量等价** (isometric) 或者**全等** (congruent). ■

命题 7.3.2: 设 α 是仿射空间 $\mathbb{A}^n$ 到自身的映射, O 为 $\mathbb{A}^n$ 中的原点, 即坐标为 $(0,\cdots,0)$ 的点. 通过将 $\mathbb{A}^n$ 等同于 $\mathbb{R}^n$ 并且使用 $\mathbb{R}^n$ 上的标准内积, 我们可以谈论 $\mathbb{A}^n$ 中任意两点的距离.

下列条件是等价的:

1. α 是等距变换, 且 $\alpha(O)=O$.

2. α 是线性等距.

3. $\alpha(O)=O$, 且 α 保持任意两点间的距离不变, 即, 对于任意 $P,Q\in\mathbb{A}^n$, $\alpha(P)$ 和 $\alpha(Q)$ 之间的距离等于 P 和 Q 之间的距离.

证明. (1) $\Rightarrow$ (2). 根据定义, $\alpha(O)=b$. 所以 $\alpha(O)=O$ 意味着 $b=0$. (2) $\Rightarrow$ (1). 由定义显然.

(2) $\Rightarrow$ (3). 将 $\mathbb{A}^n$ 中的点等同于内积空间 $\mathbb{R}^n$ 中的向量, 则根据定义, α 可以写成 $x\mapsto Ax$ 的形式, 其中 $A\in\mathbf{O}_n(\mathbb{R})$. 由此显然 $\alpha(O)=O$, 且对于任意 $u,v\in\mathbb{A}^n=\mathbb{R}^n$,

$$\begin{aligned}\|\alpha(u)-\alpha(v)\|^2&=\|Au-Av\|^2=\langle A(u-v)\,,\,A(u-v)\,\rangle\\&=(u-v)^TA^TA(u-v)=(u-v)^T(u-v)=\|u-v\|^2\,.\end{aligned}$$

(3) $\Rightarrow$ (2). 仍将 $\mathbb{A}^n$ 中的点等同于内积空间 $\mathbb{R}^n$ 中的向量. 根据假设条件, 对于任意 $w\in\mathbb{R}^n$,

$$\|\alpha(w)\|=\|\alpha(w)-0\|=\|\alpha(w)-\alpha(0)\|=\|w-0\|=\|w\|\,.$$

即, α 保持向量长度不变.

任意 $u,v\in\mathbb{R}^n$, 根据以上性质和内积的对称双线性可得

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

根据 (3) 的假设, 以上两式的左边是相等的. 所以通过对比右边可知 $\langle\,\alpha(u),\,\alpha(v)\,\rangle=\langle\,u,\,v\,\rangle$. 这说明 al 保持向量内积不变.

接下来, 对于任意 $u,v\in\mathbb{R}^n$,

$$\begin{aligned}&\|\alpha(u+v)-\alpha(u)-\alpha(v)\|^2\\&=\langle\,\alpha(u+v)-\alpha(u)-\alpha(v)\,,\,\alpha(u+v)-\alpha(u)-\alpha(v)\,\rangle\\&=\|\alpha(u+v)\|^2-2\langle\,\alpha(u+v),\,\alpha(u)+\alpha(v)\,\rangle+\|\alpha(u)+\alpha(v)\|^2\\&=\|\alpha(u+v)\|^2-2\langle\,\alpha(u+v)\,,\,\alpha(u)\,\rangle-2\langle\,\alpha(u+v),\,\alpha(v)\,\rangle\\&\quad+\|\alpha(u)\|^2+2\langle\,\alpha(u),\,\alpha(v)\,\rangle+\|\alpha(v)\|^2\\\text{(因为 }\alpha\text{ 保持长度和内积)}\ &=\|u+v\|^2-2\langle\,u+v\,,\,u\,\rangle-2\langle\,u+v,\,v\,\rangle\\&\quad+\|u\|^2+2\langle\,u,\,v\,\rangle+\|v\|^2\\\text{(将之前的演算逆向进行)}\ &=\|(u+v)-u-v\|^2=0\,.\end{aligned}$$

所以, $\alpha(u+v)=\alpha(u)+\alpha(v)$. 类似可证 $\alpha(cv)=c\alpha(v)$ 对所有 $v\in\mathbb{R}^n$ 和 $c\in\mathbb{R}$ 成立. 因此 α 是线性映射. 根据命题 7.2.2 可知 α 在标准基下的矩阵 A 是正交矩阵. 于是 α 的表达式为 $x\mapsto Ax$, 所以 α 是定义 7.3.1 意义下的线性等距. □

思考题 7.13. 以下只考虑 $\mathbb{A}^n$ 上的变换和 $\mathbb{A}^n$ 中的图形. 证明:

1. 两个等距变换的复合仍是等距变换.
2. 对于 $\mathbb{A}^n$ 中的图形而言, 度量等价是一个等价关系.
3. 对于任意取定的等距变换 α, 存在唯一一对等距变换 $\beta:\mathbb{A}^n\to\mathbb{A}^n$ 和 $\theta:\mathbb{A}^n\to\mathbb{A}^n$ 满足: $\alpha=\beta\circ\theta$, 且 β 是平移变换, θ 是线性等距; 又存在唯一一对等距变换 $\beta':\mathbb{A}^n\to\mathbb{A}^n$ 和 $\theta':\mathbb{A}^n\to\mathbb{A}^n$ 满足: $\alpha=\theta'\circ\beta'$, 且 β' 是平移变换, θ' 是线性等距.

思考题 7.14. 设 F_1 和 F_2 是 $\mathbb{A}^n$ 中的两个线性仿射集, 即, 存在 $\mathbb{R}^n$ 的子空间 U_1, U_2 及向量 v_1, v_2 使得 $F_i=v_i+U_i$. 证明以下断言等价:

1. F_1 和 F_2 维数相等, 即, $\dim U_1=\dim U_2$.
2. F_1 和 F_2 度量等价.
3. F_1 和 F_2 仿射等价.

根据思考题 7.13 (3), 仿射空间上的等距变换总是可以分解为两个等距变换的复合, 其中一个是平移变换, 另一个是线性等距 (恒等变换既可以认为是平移变换 (按零向量来做平移) 又可以认为是线性等距). 平移变换是相对容易理解的. 因此我们重点是理解线性的等距变换. 而这实际上也已经在§7.2.1 节讨论过了.

现在我们想对二维和三维仿射空间的情况进一步明确线性等距变换的几何意义.

(7.3.3) 先来看仿射平面 $\mathbb{A}^2$ 的情况. 首先, $\mathbb{A}^2$ 上的线性等距变换可以视为内积空间 $\mathbb{R}^2$ 上 (关于标准内积) 的一个正交变换 $\mathscr{A}$. 根据定理 7.2.37, 一定存在 $\mathbb{R}^2$ 的一组规范正交基 $\mathcal{F}=(f_1, f_2)$ 使得矩阵 $\mathcal{M}_{\mathcal{F}}(\mathscr{A})$ 具有以下形式之一:

$$\begin{pmatrix}1&\\&1\end{pmatrix},\ \begin{pmatrix}1&\\&-1\end{pmatrix},\ \begin{pmatrix}-1&\\&-1\end{pmatrix},\ \text{或}\quad \begin{pmatrix}\cos\theta&-\sin\theta\\ \sin\theta&\cos\theta\end{pmatrix},\quad \text{其中}\quad 0<\theta<\pi\,.$$

若 $\mathcal{M}_{\mathcal{F}}(\mathscr{A})=\begin{pmatrix}1&\\&1\end{pmatrix}$, 则显然 $\mathscr{A}=I$ 是恒等变换.

若 $\mathcal{M}_{\mathcal{F}}(\mathscr{A})=\begin{pmatrix}1&\\&-1\end{pmatrix}$, 则对于 $\mathbb{R}^2$ 中的任意向量 v, 当我们写出 $\mathcal{M}_{\mathcal{F}}(v)=(x, y)^T$, 即 $v=xf_1+yf_2$ 时, $\mathcal{M}_{\mathcal{F}}(\mathscr{A}v)=(x, -y)^T$, 即 $\mathscr{A}v=xf_1+(-y)f_2$. 因此, 如果我们将 $\mathbb{R}^2$ 的新坐标轴建立在 f_1, f_2 这组规范正交基上, 那么 $\mathscr{A}$ 的作用效果正好是关于 f_1 轴的反射 (或者称镜像对称、镜面反射等, 参见本书上册例 3.3.26 及习题 7.2.2).

若 $\mathcal{M}_{\mathcal{F}}(\mathscr{A})=\begin{pmatrix}-1&\\&-1\end{pmatrix}$, 那么 $\mathscr{A}(xf_1+yf_2)=(-x)f_1+(-y)f_2$. 此时从 (f_1, f_2) 为坐标轴的坐标系上看, $\mathscr{A}$ 的作用效果是关于原点的**中心对称** (central symmetry), 也相当于旋转 $180°$.

若 $\mathcal{M}_{\mathcal{F}}(\mathscr{A})=\begin{pmatrix}\cos\theta&-\sin\theta\\ \sin\theta&\cos\theta\end{pmatrix}$, 则

$$\begin{aligned}\mathscr{A}f_1&=\cos\theta\cdot f_1+\sin\theta\cdot f_2\,,\\ \mathscr{A}f_2&=(-\sin\theta)\cdot f_1+\cos\theta\cdot f_2\,.\end{aligned}$$

以 (f_1, f_2) 为坐标轴来观察的话, $\mathscr{A}$ 的作用效果是旋转角度 θ 对应的旋转 (若 f_1 到 f_2 的有向角是逆时针的 90 度, 则该旋转沿逆时针方向).

以上无论哪种情况, 我们可以明显地观察到 $\mathscr{A}$ 只会将平面图形移动位置 (包括旋转, 或者关于原点做中心对称, 又或者关于某直线做轴对称). 因此, 线性等距变换把任意平面图形变换为一个形状外观上没有任何差别的图形, 这也就是我们直观感觉上的"图形全等". 由于平移变换也不改变图形的样貌, 所以我们可以知道: 仿射平面到自身的任何等距变换 (无论是否线性) 都保持图形形状"全等". ■

(7.3.4) 现在考虑仿射平面 $\mathbb{A}^2$ 中的一条二次曲线 $F = V(f)$, 其中二元二次多项式 f 形如

$$\begin{aligned} f(x, y) &= a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2b_1x + 2b_2y + C \\ &= (x, y)A\begin{pmatrix} x \\ y \end{pmatrix} + 2B\begin{pmatrix} x \\ y \end{pmatrix} + C\,, \quad \text{其中 } A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, \ B = (b_1, b_2)\,. \end{aligned} \tag{7.3.4.1}$$

这里我们假设 a_{11}, a_{12}, a_{13} 三者不全为 0 (即矩阵 A 非零). 和过去仿射变换的情况 (参见 § 6.3.2) 类似, 讨论 F 在等距变换下的像基本相当于讨论多项式 f 的表达式在直角坐标变换 (见下文定义) 下产生什么变化.

设

$$\alpha \ : \ \mathbb{A}^2 \longrightarrow \mathbb{A}^2\,; \quad \begin{pmatrix} x \\ y \end{pmatrix} \longmapsto M\begin{pmatrix} x \\ y \end{pmatrix} + t \tag{7.3.4.2}$$

为一个等距变换, 其中 $M = (m_{ij})$ 是正交矩阵, $t = (t_1, t_2)^T \in \mathbb{R}^2$. 若令

$$\begin{cases} X := m_{11}x + m_{12}y + t_1 \\ Y := m_{21}x + m_{22}y + t_2 \end{cases} \quad \text{亦即} \quad \begin{pmatrix} X \\ Y \end{pmatrix} := M\begin{pmatrix} x \\ y \end{pmatrix} + t\,,$$

则

$$\begin{pmatrix} x \\ y \end{pmatrix} = P\begin{pmatrix} X \\ Y \end{pmatrix} + u \quad \text{其中} \ \ P = M^{-1}\,, \ u = -Pt\,. \tag{7.3.4.3}$$

以 (7.3.4.3) 式代入 f 的表达式, 化简可得

$$g(X, Y) = (X, Y)A'\begin{pmatrix} X \\ Y \end{pmatrix} + 2B'\begin{pmatrix} X \\ Y \end{pmatrix} + C' \tag{7.3.4.4}$$

其中

$$A' = P^TAP\,, \ B' = u^TAP + BP\,, \ C' = u^TAu + 2Bu + C\,.$$

以上讨论和过去仿射变换的情况完全类似. 主要的差别在于, 现在我们必须强调: 在 (7.3.4.3) 式中, P 必须是正交矩阵. 因此 (7.3.4.3) 式也被称为一个**正交变量替换** (orthogonal change of variables) 或者**直角坐标变换** (rectangular coordinate change).

根据定理 7.2.23, 非零对称阵 A 总是可以正交相似于一个对角矩阵, 即以下五种情况之一:

$$\begin{pmatrix} \delta_1 & \\ & \delta_2 \end{pmatrix}, \ \begin{pmatrix} -\delta_1 & \\ & -\delta_2 \end{pmatrix}, \ \begin{pmatrix} \delta_1 & \\ & -\delta_2 \end{pmatrix}, \ \begin{pmatrix} \delta & \\ & 0 \end{pmatrix}, \ \begin{pmatrix} -\delta & \\ & 0 \end{pmatrix}$$

其中 $\delta_1, \delta_2, \delta$ 均为正实数. 因为 $V(f) = V(-f)$, 在不计非零常数倍时, 我们只需考虑以下三种情况

$$\begin{pmatrix} \delta_1 & \\ & \delta_2 \end{pmatrix}, \ \begin{pmatrix} \delta_1 & \\ & -\delta_2 \end{pmatrix}, \ \begin{pmatrix} \delta & \\ & 0 \end{pmatrix} \quad \text{其中} \ \ \delta_1, \delta_2, \delta > 0\,.$$

也就是说, 经过适当的直角坐标变换, 不妨设 (7.3.4.1) 式中的矩阵 A 具有以上三种形式之一.

接下来的讨论几乎是 (6.3.7) 一段的依样画葫芦. 为了减少重复, 我们适当地省略其中的部分细节.

情况 1: $A = \begin{pmatrix} \delta_1 & \\ & \delta_2 \end{pmatrix}$, 其中 $\delta_1, \delta_2 > 0$.

此时 $f(x, y) = \delta_1 x^2 + \delta_2 y^2 + 2b_1 x + 2b_2 y + C$. 通过平移变换 $X = x + b_1/\delta_1$, $Y = y + b_2/\delta_2$ 可将 $f(x, y)$ 化为

$$g(X, Y) = \delta_1 X^2 + \delta_2 Y^2 - d\,, \qquad \text{其中 } d := \frac{b_1^2}{\delta_1} + \frac{b_2^2}{\delta_2} - C\,.$$

情况 1.1: 若 $d > 0$, 则 $V(g)$ 是个椭圆. 故根据 (7.3.3) 一段的论述, 原来的 $V(f)$ 也是个椭圆.

实际上, 通过适当的中心对称、镜面反射或平移操作之后, $F = V(f)$ 的图形可以与 $V(g)$ 的图形完全重合.

情况 1.2: 若 $d = 0$, 则 $V(g)$ 是一个点, 原来的 $V(f)$ 也是一个点.

情况 1.3: 若 $d < 0$, 则 $V(g)$ 和 $V(f)$ 都是虚椭圆.

情况 2: $A = \begin{pmatrix} \delta_1 & \\ & -\delta_2 \end{pmatrix}$, 其中 $\delta_1, \delta_2 > 0$.

此时 $f(x, y) = \delta_1 x^2 - \delta_2 y^2 + 2b_1 x + 2b_2 y + C$. 经过平移后可将 $f(x, y)$ 化为

$$g(X, Y) = \delta_1 X^2 - \delta_2 Y^2 - d\,, \qquad \text{其中 } d := \frac{b_1^2}{\delta_1} - \frac{b_2^2}{\delta_2} - C\,.$$

情况 2.1: 若 $d \neq 0$, 则 $V(g)$ 和 $V(f)$ 是双曲线.

情况 2.2: 若 $d = 0$, 则 $V(g)$ 和 $V(f)$ 是两条相交直线.

情况 3: $A = \begin{pmatrix} \delta & \\ & 0 \end{pmatrix}$, 其中 $\delta > 0$.

此时 $f(x, y) = \delta x^2 + 2b_1 x + 2b_2 y + C$. 经过平移可将 $f(x, y)$ 化为

$$g(X, Y) = \delta X^2 + 2b_2 Y - d\,, \qquad \text{其中 } d := \frac{b_1^2}{\delta} - C\,.$$

情况 3.1: 若 $b_2 \neq 0$, 则 $V(g)$ 和 $V(f)$ 都是抛物线.

情况 3.2: 若 $b_2 = 0$, $d > 0$, 则 $V(g)$ 和 $V(f)$ 都是两条平行直线.

情况 3.3: 若 $b_2 = d = 0$, 则 $V(g)$ 和 $V(f)$ 都是两条重合直线.

情况 3.4: 若 $b_2 = 0$, $d < 0$, 则 $V(g)$ 和 $V(f)$ 都是两条虚平行直线.

读者可能会问: 以上讨论不是说明了平面二次曲线在度量等价意义下也是分成 9 个大类吗？这和之前仿射分类的结果有什么区别呢？

实际上, 区别在于仿射等价意义下, 以上 9 类中同一类的二次曲线全都相互等价. 比如, 所有的椭圆之间都相互仿射等价. 但在度量分类时, 情况并非如此. 不同的椭圆之间显然不会都全等. 因此, 即使同属于椭圆这一类, 它们两两之间仍然有无穷多个不动的度量等价类型. 因此, 以上关于二次曲线度量等价的分类只是分成了 9 个大的形状类别, 并不意味着真正图形的度量类型只有 9 种. 九大类中同一类的二次曲线是否全等, 还需要看对应的 δ_1, δ_2, d 等参数是否按照合适的比例搭配. ■

例 7.3.5. 考虑二元二次多项式

$$f(x, y) = x^2 - 4xy + y^2 + 30x - 30y + 30$$

定义的平面二次曲线 $F = V(f)$. 试通过直角坐标变换将 f 化为标准形并确定图形 F 的形状.

首先将 f 写成矩阵的形式

$$f(x,\,y)=(x,\,y)\begin{pmatrix}1&-2\\-2&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+(30,\,-30)\begin{pmatrix}x\\y\end{pmatrix}+30\,.$$

计算矩阵 $A=\begin{pmatrix}1&-2\\-2&1\end{pmatrix}$ 的特征值得到 $\lambda_1=3,\ \lambda_2=-1$. 分别在 $\lambda_1,\,\lambda_2$ 的特征子空间中找出规范正交基得到 $f_1=\left(\frac{1}{\sqrt2},-\frac{1}{\sqrt2}\right)^T$ 和 $f_2=\left(\frac{1}{\sqrt2},\frac{1}{\sqrt2}\right)^T$. 因此,

$$\text{若令 } P=\begin{pmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\-\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{pmatrix},\ \text{则 } P^TAP=A':=\begin{pmatrix}3&\\&-1\end{pmatrix}.$$

于是, 当我们令

$$\begin{pmatrix}x'\\y'\end{pmatrix}=P^T\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}\frac{x-y}{\sqrt2}\\\frac{x+y}{\sqrt2}\end{pmatrix} \tag{7.3.5.1}$$

化简 f 的表达式可得

$$\begin{aligned}f(x,\,y)&=(x,\,y)A\begin{pmatrix}x\\y\end{pmatrix}+(30\,,\,-30)\begin{pmatrix}x\\y\end{pmatrix}+30\\&=(x',\,y')P^TAP\begin{pmatrix}x'\\y'\end{pmatrix}+(30,\,-30)P\begin{pmatrix}x'\\y'\end{pmatrix}+30\\&=3(x')^2-(y')^2+30\sqrt2x'+30\\&=3(x'+5\sqrt2)^2-(y')^2-120\,.\end{aligned}$$

再经过平移

$$\begin{pmatrix}X\\Y\end{pmatrix}=\begin{pmatrix}x'+5\sqrt2\\y'\end{pmatrix} \tag{7.3.5.2}$$

可将 f 的表达式化为 $g(X,\,Y)=3X^2-Y^2-120$.

于是 $G=V(g)$ 是由方程 $\frac{X^2}{40}-\frac{Y^2}{120}=1$ 给出的双曲线. 也就是说, 如果以字母 X 和 Y 表示 $\mathbb{A}^2$ 中的点在新的直角坐标系中的坐标, 那么旧的坐标下 $f(x,\,y)=0$ 对应的方程就转化为新坐标下 $g(X,\,Y)=0$ 给出的方程. 而后者 (不计常数倍的差别) 给出双曲线的标准方程.

考察之前使用的坐标变换 (7.3.5.1) 和 (7.3.5.2) 可以看出, 综合起来的坐标变换为

$$\begin{pmatrix}X\\Y\end{pmatrix}=P^T\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}5\sqrt2\\0\end{pmatrix}=\begin{pmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}5\sqrt2\\0\end{pmatrix}. \tag{7.3.5.3}$$

它对应的等距变换为

$$\alpha\ :\ \mathbb{A}^2\longrightarrow\mathbb{A}^2\,;\quad\begin{pmatrix}x\\y\end{pmatrix}\longmapsto\begin{pmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}5\sqrt2\\0\end{pmatrix} \tag{7.3.5.4}$$

其几何上的作用效果是先逆时针旋转 45 度, 然后沿水平向右的方向平移 $5\sqrt2$ 倍的单位向量.

因为 $g(X,\,Y)=0$ 等价于 $\frac{X^2}{40}-\frac{Y^2}{120}=1$, 从 (7.3.5.3) 式可以看出

$$f(x,\,y)=0\iff\frac{\left(\frac{x-y}{\sqrt2}+5\sqrt2\right)^2}{40}-\frac{\left(\frac{x+y}{\sqrt2}\right)^2}{120}=1\,.$$

注意 α 的线性部分将 $\mathbb{R}^2$ 的规范正交基 $u_1=\left(\frac{1}{\sqrt{2}},\,-\frac{1}{\sqrt{2}}\right)^T$, $u_2=\left(\frac{1}{\sqrt{2}},\,\frac{1}{\sqrt{2}}\right)^T$ 映射为标准基 $e_1=(1,\,0)^T$, $e_2=(0,\,1)^T$. 因为变换之后的双曲线 $V(g)$ 以 e_1 和 e_2 的方向为两条主轴 (实轴和虚轴) 所在的方向, 所以原来的图形 $F=V(f)$ 以向量 u_1 和 u_2 所在的方向为主轴方向. 观察 (7.3.5.3) 式可知, u_1 和 u_2 实际上来自于矩阵 P 的列向量. 这里的矩阵 P 就是将原来 f 的系数矩阵 A 正交相似到对角形所用到的正交 (过渡) 矩阵. 也许正是由于这个原因, 之前的定理 7.2.23 才会有"主轴定理"的名称吧. ■

(7.3.6) 现在讨论空间 $\mathbb{A}^3$ 上等距变换的几何意义. 和平面的情况一样, 平移变换是容易理解的. 因此我们只讨论线性的等距变换. 这样的变换可以视为内积空间 $\mathbb{R}^3$ 上的一个正交变换 $\mathscr{A}$. 根据定理 7.2.37, 存在 $\mathbb{R}^3$ 的一组规范正交基 $\mathcal{F}=(f_1,\,f_2,\,f_3)$ 使得矩阵 $\mathcal{M}_{\mathcal{F}}(\mathscr{A})$ 具有以下形式之一:

$$I\,,\,-I\,,\begin{pmatrix}1&&\\&1&\\&&-1\end{pmatrix},\ \begin{pmatrix}1&&\\&-1&\\&&-1\end{pmatrix},\ \begin{pmatrix}\cos\theta&-\sin\theta&\\\sin\theta&\cos\theta&\\&&1\end{pmatrix},\ \begin{pmatrix}\cos\theta&-\sin\theta&\\\sin\theta&\cos\theta&\\&&-1\end{pmatrix}.$$

通过适当调整 f_3 的正负号, 不妨设 $(f_1,\,f_2,\,f_3)$ 构成右手系.

若 $\mathcal{M}_{\mathcal{F}}(\mathscr{A})=I$, $\mathscr{A}=I$ 为恒等变换.

若 $\mathcal{M}_{\mathcal{F}}(\mathscr{A})=-I$, 则在 $f_1,\,f_2,\,f_3$ 为坐标轴的坐标系下 $\mathscr{A}$ 是关于原点的中心对称.

若 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})=\begin{pmatrix}1&&\\&1&\\&&-1\end{pmatrix}$, 则在 $f_1,\,f_2,\,f_3$ 为坐标轴的坐标系下 $\mathscr{A}$ 是关于平面 $\mathrm{span}(f_1,\,f_2)$ 的镜面反射, 即, 每一点被映射为关于平面 $\mathrm{span}(f_1,\,f_2)$ 的对称点.

若 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})=\begin{pmatrix}1&&\\&-1&\\&&-1\end{pmatrix}$, 则在 $f_1,\,f_2,\,f_3$ 为坐标轴的坐标系下 $\mathscr{A}$ 是关于 f_1 轴的轴对称.

若 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})=\begin{pmatrix}\cos\theta&-\sin\theta&\\\sin\theta&\cos\theta&\\&&1\end{pmatrix}$, 则在 $f_1,\,f_2,\,f_3$ 为坐标轴的坐标系下 $\mathscr{A}$ 是以 f_3 轴为旋转轴, 按右手系逆时针方向 (即, 以右手拇指指向为 f_3 轴的正方向) 旋转角度 θ 的旋转变换.

若 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})=\begin{pmatrix}\cos\theta&-\sin\theta&\\\sin\theta&\cos\theta&\\&&-1\end{pmatrix}$, 则在 $f_1,\,f_2,\,f_3$ 为坐标轴的坐标系下 $\mathscr{A}$ 的作用效果是一次旋转和一次镜面对称的复合, 而且这两种操作可以交换. 这里的旋转变换按右手系逆时针方向绕着 f_3 轴所做的旋转角度为 θ 的旋转变换, 镜面反射则是关于平面 $\mathrm{span}(f_1,\,f_2)$ 的镜面反射.

和 (7.3.2) 一段的讨论一样, 以上分类说明: $\mathbb{A}^3$ 上的等距变换的确在几何上将任何图形都做全等变换. ■

关于空间二次曲面的度量分类, 读者应该可以仿照 (6.3.9) 和 (7.3.4) 两段的论述来自行完成, 我们在此就不再重复了.

我们仅举两个具体的计算实例.

例 7.3.7. 考虑 $\mathbb{A}^3$ 中由以下多项式定义的二次曲面 $F=V(f)$:

$$f(x,\,y,\,z)=x^2+2y^2+3z^2-4xy-4yz+12x-6z-11\,.$$

试找出适当的直角坐标变换将 f 化成标准形式, 并确定曲面 $F=V(f)$ 的形状.

首先, 考虑矩阵 $A = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 2 & -2 \\ 0 & -2 & 3 \end{pmatrix}$ 的正交相似标准形. 计算 A 的特征多项式得到 $f(X) = (X-2)(X-5)(X+1)$. 于是 A 有三个特征值 $\lambda_1 = 2$, $\lambda_2 = 5$, $\lambda_3 = -1$. 分别在特征子空间中找出单位向量可得如下规范正交基

$$u_1 = \begin{pmatrix} -\frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{pmatrix},\ u_2 = \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ \frac{2}{3} \end{pmatrix},\ u_3 = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}.$$

于是正交矩阵 $Q = \begin{pmatrix} -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$, 使得 $Q^TAQ = \begin{pmatrix} 2 & & \\ & 5 & \\ & & -1 \end{pmatrix}$.

若令 $\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = Q^T \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, 则

$$\begin{aligned} f(x, y, z) &= (x', y', z')Q^TAQ \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + (12, 0, -6)Q \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} - 11 \\ &= 2(x')^2 + 5(y')^2 - (z')^2 + (-12, 0, 6) \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} - 11 \\ &= 2(x'-3)^2 + 5(y')^2 - (z'-3)^2 - 20\,. \end{aligned}$$

所以, 若令 $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} x'-3 \\ y' \\ z'-3 \end{pmatrix}$, 则 f 可化为 $g(X, Y, Z) = 2X^2 + 5Y^2 - Z^2 - 20$. 于是 $V(g)$ 是

$$\frac{X^2}{10} + \frac{Y^2}{4} - \frac{Z^2}{20} = 1$$

给出的单叶双曲面. 原来的图形 $F = V(f)$ 也是单叶双曲面. ■

(7.3.7') 也许我们应该特别指出, 在 (6.3.9) 一段中讨论的第 5 种情况, 当我们做正交变换时需要注意一点差别. 这种情况下, 对应的正交相似标准形是 $A = \begin{pmatrix} \delta & & \\ & 0 & \\ & & 0 \end{pmatrix}$, 其中 $\delta > 0$. 此时的二次曲面方程形如 $\delta X^2 + 2b_2Y + 2b_3Z - d = 0$.

如果 $b_2 = b_3 = 0$, 那么和过去一样讨论即可. 现在假设 b_2, b_3 不全为 0, 例如 $b_2 \neq 0$. 过去我们做仿射坐标变换时, 可以做类似

$$\begin{pmatrix} Y' \\ Z' \end{pmatrix} = \begin{pmatrix} Y + b_2^{-1}b_3Z \\ Z \end{pmatrix} = \begin{pmatrix} 1 & b_2^{-1}b_3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Y \\ Z \end{pmatrix}$$

这种形式的变换. 但是现在我们需要做正交变换, 上面的表达式给出的变换不是正交变换. 正确的做法是可以令

$$\begin{pmatrix} Y' \\ Z' \end{pmatrix} = \begin{pmatrix} \frac{b_2Y+b_3Z}{D} \\ \frac{-b_3Y+b_2Z}{D} \end{pmatrix} = \begin{pmatrix} \frac{b_2}{D} & \frac{b_3}{D} \\ \frac{-b_3}{D} & \frac{b_2}{D} \end{pmatrix} \begin{pmatrix} Y \\ Z \end{pmatrix} \qquad \text{其中}\ \ D := \sqrt{b_2^2 + b_3^2}$$

这样将方程 $\delta X^2 + 2b_2Y + 2b_3Z - d = 0$ 变成 $\delta X^2 + 2DY' - d = 0$ 的形式, 并由此可以看出曲面是一个抛物柱面. ■

7.3.2 正交投影和最小二乘问题

设 V 是实内积空间. 我们曾在 (7.1.3) 一段中提到, 根据命题 6.1.25 可知: 对于 V 的任何有限维子空间 U, 无论 $\dim V$ 是否有限, 总可以将 V 做正交和分解 $V = U \perp U^{\perp}$. 我们现在讨论与此有关的一个重要概念——正交投影, 以及它的一个重要应用.

定义 7.3.8. 设 V 是实内积空间, $U \subseteq V$ 是一个有限维子空间. 因为有正交和分解 $V = U \perp U^{\perp}$, 我们可以将任何 $v \in V$ 唯一地写成 $v = u + w$ 的形式, 其中 $u \in U$, $w \in U^{\perp}$. 通过将 v 对应到 u (并仍把 u 当作 V) 中的元素, 我们得到一个映射

$$P_U \,:\, V \longrightarrow V\,;\, v \longmapsto u\,, \quad \text{其中 } u \text{ 是 } U \text{ 中唯一能使 } v - u \in U^{\perp} \text{ 的那个向量},$$

称之为 V 到 U 上的**正交投影** (orthogonal projection). 向量 $P_U v$ 也称为 v 在子空间 U 上的正交投影. ■

例 7.3.9. 设 x 为内积空间 V 中的非零向量, $U = \mathrm{span}(x)$. 则对于任意 $v \in V$, $P_U(v) = \frac{\langle x, v \rangle}{\|x\|^2} x$. 事实上, 这只不过是 (7.1.3.1) 式的一个重述. ■

命题 7.3.10: 设 V 是实内积空间, $U \subseteq V$ 是一个有限维子空间. 则正交投影 P_U 满足以下性质:

1. $P_U \in \mathrm{End}(V)$, 即, P_U 是 V 上的线性变换.
2. $P_U^2 = P_U$.
3. $\mathrm{Ker}(P_U) = \mathrm{Im}(P_U - I) = U^{\perp}$, $\mathrm{Im}(P_U) = \mathrm{Ker}(P_U - I) = U$.
4. 对于任意 $v \in V$, $\|P_U v\| \le \|v\|$.
5. 对于 U 的任意一组规范正交基 $e_1, \cdots, e_m$ 均有
$$P_U v = \langle e_1, v \rangle e_1 + \cdots + \langle e_m, v \rangle e_m\,.$$

证明. 留作练习. □

思考题 7.15. 设 V 是实内积空间, $U \subseteq V$ 是一个有限维子空间. 证明: $(U^{\perp})^{\perp} = U$. (注意: $U^{\perp}$ 可能是无限维的.)

正交投影在实际问题中的应用主要归因于下面一条非常容易证明的结果:

命题 7.3.11: 设 V 是实内积空间, $U \subseteq V$ 是一个有限维子空间. 则对于任意 $v \in V$,

$$\|v - P_U v\| = v \text{ 到子空间 } U \text{ 的距离} := \min\{\|v - u\| \,:\, u \in U\}\,.$$

事实上, 子空间 U 中能够满足等式 $\|v - P_U v\| = \|v - u\|$ 的向量 $u \in U$ 只有 $u = P_U v$.

证明. 对于任意 $u \in U$, 注意到 $v - P_U v \in U^{\perp}$, 而 $P_U v - u \in U$. 于是

$$\begin{aligned} \|v - P_U v\|^2 &\le \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ \text{(根据毕达哥拉斯定理)} \quad &= \|(v - P_U v) + (P_U v - u)\|^2 = \|v - u\|^2\,. \end{aligned}$$

(附图示意) 显然, 上式中的不等号取到等号当且仅当 $P_U v - u = 0$, 即, $u = P_U v$. □

将命题 7.3.11 和命题 7.3.10 (5) 中的结论结合起来, 通常可以用来找到一些最小值问题的显式解.

(7.3.12) 我们这里展示一个来自教材 [3] 的一个数值实例.

考虑闭区间 $[-\pi, \pi]$ 上的所有实值连续函数构成的无限维向量空间 V. 在其上考虑如下内积:

$$\langle f, g \rangle := \int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}\,x\,.$$

我们希望找一个次数不超过 5 的多项式 $u(X) \in \mathbb{R}[X]$ 使其在区间 $[-\pi, \pi]$ 上尽可能好地逼近函数 $g(x) := \sin(x)$. 我们以 $\|u - g\|$ 是否达到最小作为标准来衡量 u 与 g 的近似是否达到最优.

通过计算机辅助计算, 可以找到所求多项式 u 的表达式近似为

$$u(x) = 0.987862x - 0.155271x^3 + 0.00564312x^5$$

(这里我们把精确解中的系数换成了适当的十进制近似值).

通过对比函数图像可以得到图 7.1.

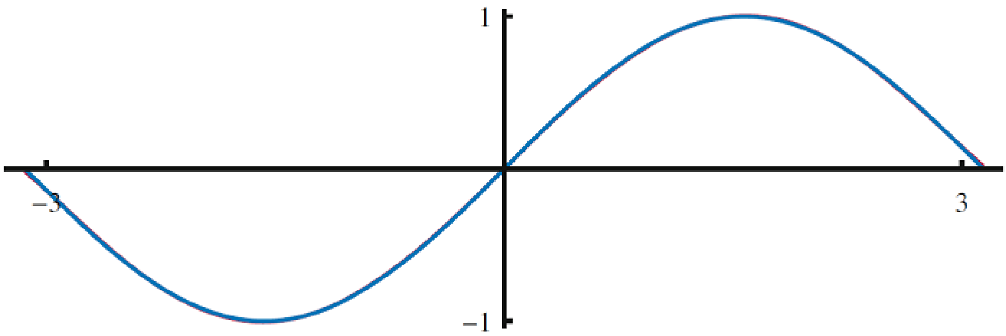

图 7.1: 图片来自 [3]

据说, 图 7.1 中绘制了两条函数曲线, 一条蓝色的是 $\sin(x)$ 的真实图像, 另一条红色的是它的近似多项式 $u(x)$ 的图像. 然而, 读者如果不使用电子设备放大观察此图的话, 恐怕很难发现真的有一条红色曲线. 它几乎完全被蓝色曲线覆盖了 (只有在 x 轴 3 和 -3 的附近才能观察到两条曲线的细微差别).

如果我们用 $\sin(x)$ 的 5 次 Taylor 多项式 $x - \frac{x^3}{3!} + \frac{x^5}{5!}$ 做近似的话, 那么得到的图像如图 7.2. 这

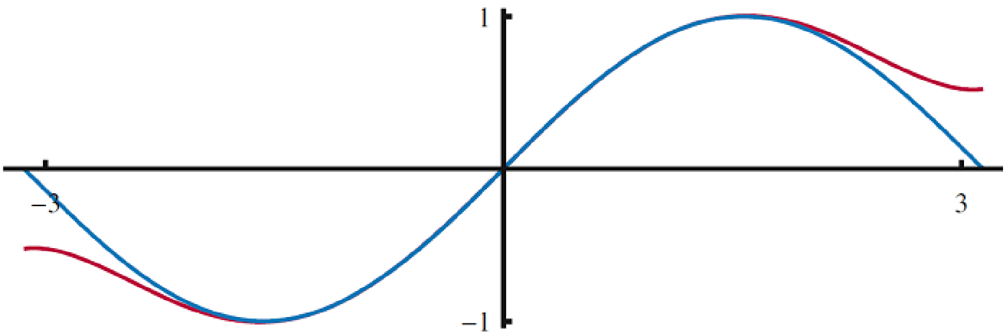

图 7.2: 图片来自 [3]

次红色和蓝色两条曲线的差别就明显许多了.

这说明, Taylor 多项式虽然在 0 点附近对 $\sin(x)$ 有很好的近似, 但是稍微远离原点之后它对原来函数的近似程度就大打折扣了.

这个例子很好地说明了线性代数应用于函数分析时, 有时候甚至可以得到比分析学方法更好的结果! ■

以上是无限维情况的一个例子. 下面再来看有限维情况的一个重要应用.

(7.3.13) 经常对实验数据进行分析或者做数学建模的人都会有这样的经历, 实验取得的数据标注在坐标平面上之后形成了一些离散的点. 我们通常需要将这些点用一条平滑的曲线连接起来, 然后尽量分析出该曲线对应的函数关系. 这其中一种简单但不失重要性的情况就是线性的关系.

假设实验采集到一组数据 (a_1, b_1), $\cdots$, (a_n, b_n) 从图像上看非常接近于同一条直线上的点, 但是由于实验误差或其他各种可能的原因，我们可能无法找到一条直线精确地使这些点都位于该直线上. 此时我们希望找到适当的直线方程 $y = \alpha x + \beta$ 使得

(7.3.13.1)
$$\sum_{i=1}^{n} \big((a_i\alpha + \beta) - b_i\big)^2$$

尽可能小. 如果考虑关于 α, β 的二元函数

$$f(\alpha, \beta) = \sum_{i=1}^{n} \big((a_i\alpha + \beta) - b_i\big)^2$$

那么问题可以转述为: 求函数 f 的最小值点. 于是我们可以采用多元微积分的方法来研究这个问题. 但如果换一个角度, 考虑

$$A = \begin{pmatrix} a_1 & 1 \\ a_2 & 1 \\ \vdots & \vdots \\ a_n & 1 \end{pmatrix} \quad \text{及} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

那么表达式 (7.3.13.1) 可以改写为 $\|A(\alpha, \beta)^T - b\|^2$. 这样问题可以描述为: 求适当的 $\alpha, \beta \in \mathbb{R}$ 使得 $\|A(\alpha, \beta)^T - b\|^2$ 达到最小.

由此我们可以把问题再做个一般化.

考虑一般的矩阵 $A \in \mathbf{M}_{n\times m}(\mathbb{R})$ 和列向量 $b \in \mathbb{R}^n$, 其中 $m \le n$. 欲求解的问题是:

(7.3.13.2)
$$\text{求向量 } x \in \mathbb{R}^m \ \text{ 使得 } \|Ax - b\| \text{ 取最小值.}$$

这类问题通常被称为**最小二乘问题** (least squares problem). 事实上, 如果用 $u_1, \cdots, u_m$ 表示矩阵 $A \in \mathbf{M}_{n\times m}(\mathbb{R})$ 的列向量组, 那么 x 取遍 $\mathbb{R}^m$ 中向量时, Ax 取遍 $\mathbb{R}^n$ 的子空间 $U = \mathrm{span}(u_1, \cdots, u_m)$ 中元素. 所以问题就是, 求出 U 中的向量 u 使得 $\|b - u\|$ 达到最小值, 然后通过求解线性方程组 $Ax = u$ 得到最小值点 x. 根据命题 7.3.11, 这里的 u 就是向量 b 在空间 U 上做正交投影得到的像. 它可以通过先求 U 的一组规范正交基, 再用命题 7.3.10 (5) 中的公式得到.

注意: 虽然 u 是唯一的, 但满足 $Ax = u$ 的向量 x 不一定唯一. 也就是说, 如果矩阵 A 不是列满秩的, 问题 (7.3.13.2) 的解可能不唯一.

这种标准的求解最小二乘问题的方法就是人们常说的**最小二乘法** (least squares method). 这一方法最早是由 Legendre** 发明的. 稍后但几乎同时地, 高斯也独立发现并使用了这个方法. 在 [30] 一书中读者可以找到关于“最小二乘法”的一些故事和有趣讨论. ■

** Adrien-Marie Legendre (1752–1833), 法国数学家.

(7.3.14) 前面介绍的先利用命题 7.3.10 (5) 求正交投影, 再求解一个线性方程组的方法, 虽然是最小二乘法的通用思路, 但在实际问题中我们还希望能够有更简洁明了的求解公式. 为此, 我们仍像之前一样设 $u_1, \cdots, u_m$ 是矩阵 $A \in \mathbf{M}_{n\times m}(\mathbb{R})$ 的列向量组.

如果 x 是问题 (7.3.13.2) 的解, 则根据命题 7.1.11, $u := x_1u_1 + \cdots + x_mu_m = Ax$ 应该是向量 b 在 U 上的正交投影. 于是 $b - u \in U^{\perp}$. 换言之,

$$\langle b-u, u_1\rangle = \cdots = \langle b-u, u_m\rangle = 0\,. \tag{7.3.14.1}$$

计算 $\langle b-u, u_i\rangle$ 可得

$$\begin{aligned}\langle b-u, u_i\rangle &= \langle b, u_i\rangle - x_1\langle u_1, u_i\rangle - \cdots - x_m\langle u_m, u_i\rangle \\ &= -(u_i^Tu_1)x_1 - \cdots - u_i^Tu_mx_m + u_i^Tb\,.\end{aligned}$$

于是, (7.3.14.1) 式可以重新整理为

$$\begin{cases}(u_1^Tu_1)x_1 + \cdots + (u_1^Tu_m)x_m = u_1^Tb\\ (u_2^Tu_1)x_1 + \cdots + (u_2^Tu_m)x_m = u_2^Tb\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ (u_m^Tu_1)x_1 + \cdots + (u_m^Tu_m)x_m = u_m^Tb\end{cases}$$

亦即

$$(A^TA)x = A^Tb\,. \tag{7.3.14.2}$$

如果 A 是列满秩的, 则 $A^TA \in \mathbf{M}_m(\mathbb{R})$ 也是满秩的 (这个结论的证明留给读者练习). 所以此时 (7.3.14.2) 有且仅有唯一的解

$$x = (A^TA)^{-1}A^Tb\,. \tag{7.3.14.3}$$

有不少书 (例如 [18]) 喜欢把 (7.3.14.3) 式写成 $x = \frac{A^Tb}{A^TA}$.

在一般情况下, 即使 A 不是列满秩的, 仍然可以证明方程 (7.3.14.2) 一定有解 (思考题 7.16), 而且从上面论证过程可以看出, 它的每个解 x 都是最小二乘问题 (7.3.13.2) 的解. ■

思考题 7.16. 证明: 对任意矩阵 $A \in \mathbf{M}_{n\times m}(\mathbb{R})$ 和任意列向量 $b \in \mathbb{R}^n$, 分块矩阵 $(A^TA \;\; A^Tb)$ 的秩等于 A^TA 的秩. (因此方程 (7.3.14.2) 总会有解.)

7.3.3 习题

习题 7.3.1. 找出适当的正交变量替换, 将下列实二次型化成标准形 (即对角形):

1. $q = x_1^2 + 2x_2^2 + 3x_3^2 - 4x_1x_2 - 4x_2x_3$.
2. $q = x_1^2 - 2x_2^2 - 2x_3^2 - 4x_1x_2 + 4x_1x_3 + 8x_2x_3$.
3. $q = 2x_1x_2 + 2x_3x_4$.
4. $q = x_1^2 + x_2^2 + x_3^2 + x_4^2 - 2x_1x_2 + 6x_1x_3 - 4x_1x_4 = 4x_2x_3 + 6x_2x_4 - 2x_3x_4$.

习题 7.3.2. 找出适当的正交变量替换, 将以下 n 元实二次型化成标准形 (即对角形):

$$q = \sum_{i=1}^{n}(x_i - \bar{x})^2\,, \quad 其中 \;\; \bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)\,.$$

习题 7.3.3. 假设 $a_i,\, b_i,\, c_i \in \mathbb{R},\, i=1,\,2,\,3$ 满足

$$a_1b_2 - a_2b_1 \neq 0\ ,\quad a_1a_2 + b_1b_2 = 0\ .$$

试问二次曲线

$$(a_1x + b_1y + c_1)^2 + (a_2x + b_2y + c_2)^2 = 1$$

是哪种类型的曲线？

习题 7.3.4. 考虑 $\mathbb{A}^3$ 中由以下多项式定义的二次曲面 $F = V(f)$:

$$f(x,\, y,\, z) = x^2 + 7y^2 + z^2 + 10xy + 2xz + 10yz + 8x + 4y + 8z - 6$$

试找出适当的直角坐标变换将 f 化成标准形式, 并确定曲面 $F = V(f)$ 的形状.

习题 7.3.5. 通过直角坐标变换确定下列二次曲线的类型、形状和位置, 并画出草图:

1. $11x^2 + 6xy + 3y^2 - 12x - 12y - 12 = 0$;
2. $5x^2 + 12xy - 22x - 12y - 19 = 0$;
3. $x^2 - 2xy + y^2 - 10x - 6y + 25 = 0$;
4. $6x^2 + 12xy + y^2 - 36x - 6y = 0$;
5. $4x^2 - 4xy + y^2 - 2x - 14y + 7 = 0$.

习题 7.3.6. 判断下列二次曲线的类型和形状:

1. $x^2 - 3xy + y^2 + 10x - 10y + 21 = 0$;
2. $6xy + 8y^2 - 12x - 26y + 11 = 0$;
3. $x^2 + 4xy + 4y^2 - 20x + 10y - 50 = 0$;
4. $x^2 + 2xy + y^2 - 8x + 4 = 0$;
5. $5x^2 + 8xy + 5y^2 - 18x - 18y + 9 = 0$;
6. $5x^2 - 16xy + 296^2 + 10x - 34y + 10 = 0$;
7. $x^2 + xy - 2y^2 - 11x - y + 28 = 0$;
8. $8x^2 + 8xy + 2y^2 - 6x - 3y - 5 = 0$;
9. $9x^2 - 8xy + 24y^2 - 32x - 16y + 138 = 0$.

习题 7.3.7. 根据实参数 λ 的取值讨论下列二次曲线的类型:

1. $x^2 - 4xy + 4y^2 + 8y + 3 + 2\lambda(-2xy - x - 1) = 0$;
2. $\lambda x^2 - 2xy + \lambda y^2 - 2x + 2y + 5 = 0$;
3. $(1+\lambda^2)(x^2 + y^2) - 4\lambda xy + 2\lambda(x + y) + 2 = 0$.

习题 7.3.8. 实数 λ 取什么值时, 方程

$$\lambda x^2+4xy+y^2-4x-2y-3=0$$

表示两条 (不重合的) 直线？

习题 7.3.9. 下列二次曲线是否有对称轴？若有, 请求出对称轴的方程.

1. $2xy-4x+2y+3=0$;

2. $x^2-4xy+4y^2-5x+6=0$.

习题 7.3.10. 设 a_{11}, a_{12}, a_{22} 和 c 均为实数, 且 a_{12} 和 $a_{11}-a_{22}$ 不全为 0. 假设方程 $a_{11}x^2+2a_{12}xy+a_{22}y^2+c=0$ 表示的二次曲线 C 是椭圆或双曲线.

证明: 方程

$$a_{12}(x^2-y^2)-(a_{11}-a_{22})xy=0$$

表示的曲线是两条相交直线, 并且这两条直线是曲线 C 的对称轴.

习题 7.3.11. 假设方程

$$a_{11}x^2+2a_{12}xy+a_{22}y^2+2b_1x+2b_2y+c=0$$

定义的曲线是一条抛物线. 证明: 该抛物线的顶点是原点的充分必要条件是:

$$a_{11}b_1^2+a_{22}b_2^2+2a_{12}b_1b_2=c=0\,.$$

习题 7.3.12. 假设方程

$$a_{11}x^2+2a_{12}xy+a_{22}y^2+2b_1x+2b_2y+c=0$$

定义的曲线 C 是双曲线.

证明:

1. C 的两条渐近线分别和方程 $a_{11}x^2+2a_{12}xy+a_{22}y^2=0$ 所表示的两条相交直线平行.

2. 进一步假设 $b_1=b_2=0$. 则双曲线 C 的两条渐近线恰好是方程 $a_{11}x^2+2a_{12}xy+a_{22}y^2=0$ 所表示的两条相交直线.

习题 7.3.13. 设 a, b, $c\in\mathbb{R}$, 其中 $c\neq 0$, a, b 不全为 0. 证明方程

$$ab(x^2-y^2)-(a^2-b^2)xy=C$$

表示双曲线, 并求出该双曲线的渐近线方程.

习题 7.3.14. 假设 a_i, b_i, $c_i\in\mathbb{R}$, $i=1,\,2,\,3$ 满足 $a_1b_2-a_2b_1\neq 0$.

证明二次曲线

$$(a_1x+b_1y+c_1)^2-(a_2x+b_2y+c_2)^2=1$$

是双曲线, 并求出该双曲线的渐近线方程.

习题 7.3.15. 设 V 是实内积空间, U, W 是两个子空间, α, $\beta\in V$. 将

$$d(\alpha,\,\beta+W):=\inf\{\|\alpha-v\|\,:\,v\in\beta+W\}\,,$$
$$d(\alpha+U\,,\,\beta+W):=\inf\{\|\gamma-v\|\,:\,\gamma\in\alpha+U\,,\,v\in\beta+W\}$$

分别称为 α 到 (线性) 仿射集 $\beta+W$ (参见本书上册 2.2.3 节) 的距离, 以及 (线性) 仿射集 $\alpha+U$ 和 $\beta+W$ 之间的距离.

1. 假设 $\beta-\alpha=\beta_1+\beta_2$, 其中 $\beta_1\in W$, $\beta_2\in W^{\perp}$. 证明: $d(\alpha,\,\beta+W)=\|\beta_2\|$.

2. 假设 $\beta-\alpha=\delta_1+\delta_2$, 其中 $\delta_1\in U+W$, $\beta_2\in (U+W)^{\perp}$. 证明: $d(\alpha+U,\,\beta+W)=\|\delta_2\|$.

习题 7.3.16. 在 $\mathbb{R}^4$ 中考虑

$$v_1=(1,\,1,\,0,\,0)^T\,,\ v_2=(1,\,1,\,1,\,2)^T\,,\ U=\operatorname{span}(v_1,\,v_2)\,.$$

求 $u\in U$ 使得 $\|u-(1,\,2,\,3,\,4)^T\|$ 达到最小.

习题 7.3.17. 考虑平面内坐标为

$$(a_1,\,b_1)=(-6,\,-1)\,,\ (a_2,\,b_2)=(-2,\,2)\,,\ (a_3,\,b_3)=(1,\,1)\,,\ (a_4,\,b_4)=(7,\,6)$$

的四个点. 求直线方程 $y=\alpha x+\beta$ 使得

$$\sum_{i=1}^{4}(\alpha a_i+\beta-b_i)^2$$

达到最小.

习题 7.3.18. 设 V 是有限维内积空间, $\mathscr{A}\in\operatorname{End}(V)$. 证明以下条件等价:

1. $\operatorname{Ker}(\mathscr{A})$ 与 $\operatorname{Im}(\mathscr{A})$ 正交, 且 $\mathscr{A}^2=\mathscr{A}$.

2. 存在子空间 $U\subseteq V$ 使 $\mathscr{A}=P_U$.

习题 7.3.19. 设 V 是有限维内积空间, $U\subseteq V$ 是一个子空间.

1. 证明 $P_{U^{\perp}}=I-P_U$.

2. 设 $\mathscr{A}\in\operatorname{End}(V)$. 证明以下条件等价:

   (i) U 和 $U^{\perp}$ 都是 $\mathscr{A}$ 的不变子空间.

   (ii) P_U 与 $\mathscr{A}$ 可交换.

习题 7.3.20. 设 V 是有限维内积空间, $\mathscr{A}\in\operatorname{End}(V)$. 证明以下条件等价:

1. $\mathscr{A}^2=\mathscr{A}$ 且对于所有 $v\in V$ 均有 $\|\mathscr{A}v\|\le\|v\|$.

2. 存在子空间 $U\subseteq V$ 使 $\mathscr{A}=P_U$.

习题 7.3.21. 对每个 $k\in\mathbb{N}$, 按照如下方式定义多项式 $P_k(X)\in\mathbb{R}[X]$:

$$P_0(X):=1\,,$$
$$\text{当 } k\ge 1 \text{ 时, } P_k(X):=\frac{1}{2^k k!}\big((X^2-1)^k\big)^{(k)}\,.$$

(这里 $f^{(k)}$ 表示多项式 f 的 k 阶导数.) 称 $P_k(X)$ 为 k 次 Legendre **多项式**.

1. 证明: 对任意 $k\in\mathbb{N}^*$, $(k+1)P_{k+1}(X)-(2k+1)XP_k(X)+kP_{k-1}(X)=0$.

2. 令 $n\in\mathbb{N}^*$, $V=\mathbb{R}[X]_{\le n}$. 通过以下内积将 V 视为内积空间:

$$\langle\, f,\,g\,\rangle:=\int_{-1}^{1}f(x)g(x)\,\mathrm{d}x\,.$$

   证明: Legendre 多项式 $P_0(X),\,P_1(X),\cdots,P_n(X)$ 是 V 的一组正交基.

习题 7.3.22. 设 V 为闭区间 $[-1, 1]$ 上的所有连续函数构成的实向量空间, 在其上取定内积为

$$\langle f, g \rangle := \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x\ .$$

令 $W = \{f \in V \mid f(0) = 0\}$.

证明: $W^{\perp} = 0$, $V \neq W + W^{\perp}$, $(W^{\perp})^{\perp} \neq W$.

第八章　酉空间

如果只考虑度量结构, 那么 n 维复向量空间可以视为 $2n$ 维实向量空间, 其度量就像实内积空间那样定义就可以了. 但是我们还希望度量结构能和复线性变换有好的协调关系, 所以还是需要把度量结构和复线性的结构综合起来考虑. 这样就需要研究本章所关注的酉空间 (也称复内积空间). 与实内积的情况不同, 复内积不是双线性型. 为此我们需要一个和"双线性"类似的概念, 这个概念和复数的共轭运算有关.

对任意复数 z, 用 $\bar{z}$ 表示它的共轭, 用 $|z|$ 表示它的模 (也称绝对值). 即, 如果 $z = x + yi$, 其中 $x, y \in \mathbb{R}$, 则 $\bar{z} = x - yi$, $|z| = \sqrt{x^2 + y^2}$.

8.1　酉空间与相关概念

若非有相反的声明, 本节中 V 总是表示 $\mathbb{C}$ 上的一个向量空间.

8.1.1　Hermite 型与 Hermite 内积

定义 8.1.1. 复向量空间 V 上的一个 **Hermite 型** (Hermitian form) 是指一个映射

$$\langle \cdot\,,\, \cdot \rangle \;:\; V \times V \longrightarrow \mathbb{C}\,; \quad (u, v) \longmapsto \langle\, u, v \,\rangle$$

它满足以下性质:

- **半双线性** (sesquilinearity)* 或**共轭双线性** (conjugate bilinearity): 对于任意 $u, v, w \in V$ 和任意 $a, b \in \mathbb{C}$,

$$\text{(关于第一个位置的向量满足“共轭线性”)}\ \langle\, au + bv, w \,\rangle = \overline{a}\langle\, u, w \,\rangle + \overline{b}\langle\, v, w \,\rangle\,,$$
$$\text{(关于第二个位置的向量满足通常的“线性”)}\ \langle\, w, au + bv \,\rangle = a\langle\, w, u \,\rangle + b\langle\, w, v \,\rangle\,.$$

- **共轭对称性** (conjugate symmetry): 对任意 $u, v \in V$, $\langle\, u, v \,\rangle = \overline{\langle\, v, u \,\rangle}$.

以上 Hermite 型被称为是**正定的**, 是指对任意 $v \in V$, $\langle\, v, v \,\rangle$ 总是非负实数, 而且只有当 $v = 0$ 时才会有 $\langle\, v, v \,\rangle = 0$. 正定的 Hermite 型也被称为 Hermite **内积** (Hermitian inner product) 或**复内积** (complex inner product).

一个**酉空间** (unitary space) 或者**复内积空间** (complex inner product space) 是指一个有序对 $(V, \langle \cdot\,, \cdot \rangle)$, 其中 V 是个复向量空间, $\langle \cdot\,, \cdot \rangle$ 是 V 上的一个 Hermite 内积 (复内积). 通常在不致引起歧义的时候我们简称 V 是个酉空间.

与实内积空间的情况类似, 现在假设 $(V, \langle \cdot\,, \cdot \rangle)$ 是个酉空间. 根据正定性, 对于每个向量 $v \in V$, 可以定义其**长度** (length)、 **模长** (modulus) 或**范数** (norm) 为 $\|v\| := \sqrt{\langle\, v, v \,\rangle}$. 只有零向量的长度是

*单词 sesquilinearity 中前缀 sesqui 的意思是"一个半、一又二分之一", 将 sesquilinearity 翻译为"半双线性" 似乎并不准确, 但在中文文献中这种译法好像比较常见.

0, 其它向量的长度都是正实数. 此外, 对于 V 中每两个向量 u, v, 可以定义二者的**距离** (distance) 为 $\|u-v\|$. 长度和距离等相关的概念依赖于 Hermite 内积的选取. 同一向量相对于不同的 Hermite 内积长度可以是不同的. ■

有些书 (例如 [3], [21]) 在定义 Hermite 型时使用的"半双线性" 定义是关于第一个位置的向量线性, 而关于第二个位置的向量满足共轭线性. 这种定义和我们的定义没有本质的差别. 当我们考虑 Hermite 型时, 如果 $\langle \cdot\,,\,\cdot \rangle$ 是定义 8.1.1 意义下的 Hermite 型, 那么映射 $\langle \cdot\,,\,\cdot \rangle' : (u,\,v) \mapsto \langle v,\,u \rangle$ 就满足那些书中的"半双线性" (及相同的"共轭对称性").

思考题 8.1. 设 $(V, \langle \cdot\,,\,\cdot \rangle)$ 是酉空间. 证明:

1. 对于任意 $v \in V$ 和任意 $c \in \mathbb{C}$, $\|cv\| = |c| \cdot \|v\|$.

2. (**极化恒等式** polarization identity) 对于任意 u, $v \in V$,

$$\langle\, u\,,\, v\,\rangle = \frac{1}{4}\left(\|u+v\|^2 - \|u-v\|^2 - \|u+iv\|^2 i + \|u-iv\|^2 i\right) .$$

3. (**平行四边形恒等式** parallelogram identity) 对于任意 u, $v \in V$,

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2) .$$

例 8.1.2. 来举几个酉空间的例子.

1. 向量空间 $\mathbb{C}^n$ 上熟知的一个 Hermite 内积为

$$\langle\, x,\, y\,\rangle := \overline{x}^T y = \bar{x}_1 y_1 + \cdots + \bar{x}_n y_n$$

   其中 $x = (x_1, \cdots, x_n)^T$, $y = (y_1, \cdots, y_n)^T$, 我们称之为 $\mathbb{C}^n$ 上的**标准** Hermite **内积** (standard Hermitian inner product). 今后若无相反的说明, $\mathbb{C}^n$ 作为酉空间时总是认为配上标准 Hermite 内积.

   显然, 对于这个标准 Hermite 内积来说, $x = (x_1, \ldots, x_n)^T \in \mathbb{C}^n$ 的长度由下式给出:

$$\|x\| = \sqrt{|x_1|^2 + \cdots + |x_n|^2}\,.$$

2. 对于任意正实数 $c_1, \cdots, c_n$,

$$\mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{C}\,; \quad (x,\, y) \longmapsto c_1\overline{x}_1 y_1 + \cdots + c_n \overline{x}_n y_n$$

   也是 $\mathbb{C}^n$ 上的 Hermite 内积.

3. 考虑次数 $\leq n$ 的复系数多项式空间 $V = \mathbb{C}[X]_{\leq n}$. 则

$$V \times V \longrightarrow \mathbb{C}\,; \quad (f,\, g) \longmapsto \langle\, f,\, g\,\rangle := \sum_{k=0}^{n} \overline{f(k)} g(k)$$

   是 V 上的一个 Hermite 内积. ■

定义 8.1.3. 对任意矩阵 $A = (a_{ij}) \in \mathbf{M}_{m\times n}(\mathbb{C})$, 通过取每个矩阵元素 $a_{ij} \in \mathbb{C}$ 的共轭可以定义相同大小的矩阵 $\overline{A} = (\bar{a}_{ij})$, 称为 A 的 (**复**) **共轭** (complex conjugate) (参见注记 7.2.27). 矩阵 $\overline{A}$ 的转置 $\overline{A}^T$ 则称为 A 的**共轭转置** (conjugate transpose). 易见, $\overline{A}^T$ 也等于 A^T 的共轭.

一个 Hermite **矩阵** (Hermitian matrix) 是指满足 $\overline{A}^T = A$ 的一个方阵 $A \in \mathbf{M}_n(\mathbb{C})$. ■

思考题 8.2. 设 $n \in \mathbb{N}^*$, $A \in \mathbf{M}_n(\mathbb{C})$. 定义映射[†]

$$\langle A \rangle_h \ : \ \mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{C}\,; \quad (x\,,\,y) \longmapsto \overline{x}^T A y\,.$$

证明: $\langle A \rangle_h$ 是 $\mathbb{C}^n$ 上的 Hermite 型当且仅当 A 是 Hermite 阵.

尽管 Hermite 型不是双线性型, 但是很多概念是可以类似定义的, 而许多关于对称双线性型的结论在 Hermite 型的情况也有类似的版本. 我们下面列举一些这样的定义和结果.

定义 8.1.4. 设 $\langle \cdot\,,\,\cdot \rangle$ 是 V 上的 Hermite 型. 若 $\mathcal{B} = (v_1, \cdots, v_n)$ 是 V 的一组有序基, 相应的矩阵 $\mathcal{M}_{\mathcal{B}}(\langle \cdot, \cdot \rangle) = (\langle v_i, v_j \rangle)$ 称为 Hermite 型 $\langle \cdot\,,\,\cdot \rangle$ 在 $\mathcal{B}$ 下的 Gram **矩阵**. 当 $\langle \cdot\,,\,\cdot \rangle$ 是 Hermite 内积时, 该矩阵也常被称为**度量矩阵** (metric matrix). ■

引理 8.1.5: *设 V 是 n 维复向量空间, $n \in \mathbb{N}^*$, $\mathcal{B} = (v_1, \cdots, v_n)$ 是 V 的一组基, $\langle \cdot\,,\,\cdot \rangle$ 是 V 上的 Hermite 型, $G = \mathcal{M}_{\mathcal{B}}(\langle \cdot\,,\,\cdot \rangle) = (\langle v_i, v_j \rangle) \in \mathbf{M}_n(\mathbb{C})$ 为相应的 Gram 矩阵.*

则 G 是 Hermite 阵, 且对于任意 $u, v \in V$,

$$\langle u, v \rangle = \overline{\mathcal{M}_{\mathcal{B}}(u)}^T \cdot G \cdot \mathcal{M}_{\mathcal{B}}(v)\,.$$

证明. 留作练习. □

定义 8.1.6. 设 $\langle \cdot\,,\,\cdot \rangle$ 是复向量空间 V 上的 Hermite 型, $u, v \in V$. 若 $\langle u, v \rangle = 0$, 则称 u, v 关于 $\langle \cdot\,,\,\cdot \rangle$ **正交** 或简称 u, v 正交, 记为 $u \perp v$. (因为 $\langle v, u \rangle = \overline{\langle u, v \rangle}$, 所以 $\langle u, v \rangle = 0$ 蕴含 $\langle v, u \rangle = 0$).

对于 V 的任意子集 S, 可以定义 S 关于 $\langle \cdot\,,\,\cdot \rangle$ 的**正交补**

$$S^{\perp} := \{v \in V \,|\, v \text{ 和 } S \text{ 中所有元素正交}\}\,.$$

不难验证: $S^{\perp}$ 总是 V 的子空间.

若 $V^{\perp} = 0$, 则称 $\langle \cdot\,,\,\cdot \rangle$ 是非退化的. 显然, 若 $\langle \cdot\,,\,\cdot \rangle$ 是个 Hermite 内积 (即, 正定的 Hermite 型), 那么 $\langle \cdot\,,\,\cdot \rangle$ 是非退化的. ■

思考题 8.3. 设 V 是有限维复向量空间, $n = \dim V \geq 1$, $\langle \cdot\,,\,\cdot \rangle$ 是 V 上的 Hermite 型.

则下列陈述等价:

1. $\langle \cdot\,,\,\cdot \rangle$ 是非退化的.
2. 对于 V 的任意一组有序基 $\mathcal{B} = (v_1, \cdots, v_n)$, Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(\langle \cdot\,,\,\cdot \rangle) = (\langle v_i, v_j \rangle)$ 是非奇异阵.
3. 存在 V 的一组有序基 $\mathcal{B} = (v_1, \cdots, v_n)$ 使得 Gram 矩阵 $\mathcal{M}_{\mathcal{B}}(\langle \cdot\,,\,\cdot \rangle) = (\langle v_i, v_j \rangle)$ 是非奇异阵.

命题 8.1.7 (**毕达哥拉斯定理**): *设 V 是酉空间, $u, v \in V$. 若 $u \perp v$, 则 $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.*

证明. 留作练习. □

思考题 8.4. 与实内积空间的情况略有不同的是, 毕达哥拉斯定理的逆定理在酉空间的情况不成立. 即, 在酉空间中可能存在两个向量 u, v 满足 $\|u+v\|^2 = \|u\|^2 + \|v\|^2$ 但 u 和 v 不正交.

请举出这样的具体例子.

引理 8.1.8: *设 V 是酉空间, $v \in V$ 是非零向量. 则对于任意 $u \in V$, 存在唯一的 $c \in \mathbb{C}$ 及 $w \in V$ 使得*

$$u = cv + w \quad \text{且} \quad v \perp w\,.$$

事实上, $c = \frac{\langle v, u \rangle}{\|v\|^2}$.

[†]这里的记号 $\langle A \rangle_h$ 带有下标 h 是为了区别例 6.2.4 (2) 中出现的记号 $\langle A \rangle$ (和例 6.1.4 (3) 中出现的记号 $\langle \cdot\,,\,\cdot \rangle_A$).

证明. 留作练习. □

定理 8.1.9: 设 V 是酉空间, $u, v \in V$.

1. Cauchy–Bunyakovsky–Schwarz 不等式:
$$|\langle u, v \rangle| \leq \|u\| \cdot \|v\| .$$
等号成立的充分必要条件是 u, v 线性相关.

2. 三角不等式:
$$\|u + v\| \leq \|u\| + \|v\| .$$
等号成立的充分必要条件是存在实数 $c \geq 0$ 满足 $u = cv$ 或 $v = cu$.

证明. 留作练习. □

8.1.2 酉空间中的规范正交基

(8.1.10) 设 V 是酉空间. 完全照搬实内积空间的情况, 我们可以做以下定义:

(1) 如果向量 $v \in V$ 的长度为 1, 即, $\|v\| = 1$, 我们就称 v 是一个**单位向量** (unit vector).

(2) 如果 $v_1, \cdots, v_r$ 是 V 中的一组单位向量, 并且它们两两之间相互正交, 那么我们称 $v_1, \cdots, v_r$ 是 V 中的一个**规范正交组** (orthonormal system).

(注意, 由一个单位向量 v_1 构成的向量组 v_1 自动认为是规范正交组.)

(3) 如果一个规范正交组 $v_1, \cdots, v_n$ 构成 V 的一组基, 我们就称之为一组**规范正交基** (orthonormal basis) 或者**标准正交基**.

不难证明: 如果 $\mathscr{F}$ 是 V 中的一族两两正交的非零向量, 那么向量族 $\mathscr{F}$ 是线性无关的 (参见命题 7.1.10).

因此, 若 $\dim V = n$, 那么 V 中的向量组 $e_1, \cdots, e_n$ 构成一组规范正交基当且仅当它们是两两正交的单位向量 (可对照推论 7.1.11).

若 $\mathcal{E} = (e_1, \cdots, e_n)$ 的确是 V 的一组规范正交基, 则对于任意 $u, v \in V$,

$$\begin{aligned} \mathcal{M}_{\mathcal{E}}(v) &= (\langle e_1, v \rangle, \cdots, \langle e_n, v \rangle)^T \\ \|v\| &= \sqrt{|\langle e_1, v \rangle|^2 + \cdots + |\langle e_n, v \rangle|^2} \\ \langle u, v \rangle &= \overline{\mathcal{M}_{\mathcal{E}}(u)}^T \mathcal{M}_{\mathcal{E}}(v) . \end{aligned} \tag{8.1.10.1}$$

由此可见, 通过一组规范正交基的选取可以将一般的 n 维酉空间和标准的酉空间 $\mathbb{C}^n$ (参见例 8.1.2 (1)) 等同起来. ■

定理 8.1.11 (Gram–Schmidt **正交化过程**): 设 V 是酉空间, $v_1, \cdots, v_m$ 是 V 中一个线性无关的向量组.

则必然存在 V 中的规范正交组 $e_1, \cdots, e_m$ 使得对于每个 $j \in [\![1, m]\!]$ 均有

$$\operatorname{span}(v_1, \cdots, v_j) = \operatorname{span}(e_1, \cdots, e_j) . \tag{8.1.11.1}$$

事实上, 这一规范正交组一定可以满足以下关系:

$$\begin{aligned} e_1 &= \lambda_1 \frac{v_1}{\|v_1\|} , \\ \text{对每个 } j = 2, \cdots, m, \quad e_j &= \lambda_j \frac{v_j - \langle e_1, v_j \rangle e_1 - \cdots - \langle e_{j-1}, v_j \rangle e_{j-1}}{\|v_j - \langle e_1, v_j \rangle e_1 - \cdots - \langle e_{j-1}, v_j \rangle e_{j-1}\|} \end{aligned} \tag{8.1.11.2}$$

其中 $\lambda_1, \cdots, \lambda_m$ 可以是任意模长为 1 的复数.

证明. 完全模仿定理 7.1.16 的证明即可. □

推论 8.1.12: 设 V 是有限维酉空间.

1. 对于 V 的任意一组基 $v_1, \cdots, v_n$, 一定存在 V 的规范正交基 $e_1, \cdots, e_n$ 使得
$$\text{对于每个 } j \in [\![1, n]\!], \quad \text{均有 } \operatorname{span}(v_1, \cdots, v_j) = \operatorname{span}(e_1, \cdots, e_j)\,.$$
2. V 中的任意规范正交组一定可以将其扩充为一组规范正交基.

证明. (1) 直接由定理 8.1.11 可得. (2) 若 $e_1, \cdots, e_m$ 是 V 的一个规范正交组. 取 $v_1 = e_1, \cdots, v_m = e_m$, 再将 $v_1, \cdots, v_m$ 扩充为 V 的一组基 $v_1, \cdots, v_m, v_{m+1}, \cdots, v_n$. 在 (8.1.11.2) 式中取 $\lambda_1 = \cdots = \lambda_n = 1$, 由此得到的规范正交基就是 $e_1, \cdots, e_m$ 的扩充. □

推论 8.1.12 中的结论 2 是命题 7.1.14 的类似结论. 不过, 我们这里采用 Gram–Schmidt 正交化的方法来证明. 此方法与之前的证明略有不同.

定理 8.1.13 (QR **分解**): 对于任意列满秩的复矩阵 $A \in \mathbf{M}_{n\times m}(\mathbb{C})$, 存在唯一一对矩阵 $Q \in \mathbf{M}_{n\times m}(\mathbb{C})$ 和 $R \in \mathbf{M}_m(\mathbb{C})$ 满足
$$A = QR, \quad \text{且 } Q \text{ 的列向量组是 } \mathbb{C}^n \text{ 中的规范正交组, } R \text{ 是对角线元素均为正实数的上三角阵.}$$
这样的表达式称为 A 的 QR **分解**.

证明. 完全模仿定理 7.1.20 的证明即可. □

8.1.3 酉矩阵和酉相似

定义 8.1.14. 设 Q 为 n 阶复方阵. 如果 $\overline{Q}^T Q = I_n$, 则称 Q 是一个**酉矩阵** (unitary matrix). 我们用 $\mathbf{U}_n(\mathbb{C})$ 表示 $\mathbf{M}_n(\mathbb{C})$ 中所有酉矩阵构成的集合.

显然, $\mathbf{U}_n(\mathbb{C}) \cap \mathbf{M}_n(\mathbb{R}) = \mathbf{O}_n(\mathbb{R})$. 也就是说, 一个实矩阵是酉矩阵当且仅当它是正交矩阵. ■

命题 8.1.15: 对于任意 n 阶复方阵 Q, 以下论断等价:

1. Q 可逆且 $Q^{-1} = \overline{Q}^T$.
2. Q 是酉矩阵.
3. Q 的列向量组构成列向量空间 $\mathbb{C}^{n\times 1}$ 关于标准 Hermite 内积的一组规范正交基.
4. Q 的行向量组构成行向量空间 $\mathbb{C}^{1\times n}$ 关于标准 Hermite 内积的一组规范正交基.

证明. 留作练习. □

像命题 7.1.25 那样, 不难证明以下命题:

命题 8.1.16: 设 V 是 n 维酉空间. 取定 V 的一组有序规范正交基 $\mathcal{E} = (e_1, \cdots, e_n)$.

1. 对于 V 的任意一组有序基 $\mathcal{B} = (v_1, \cdots, v_n)$, 下列陈述等价:
   (a) $\mathcal{B}$ 仍是一组规范正交基.
   (b) 从 $\mathcal{E}$ 到 $\mathcal{B}$ 的过渡矩阵 $P_{\mathcal{E},\mathcal{B}}$ 是酉矩阵.
   (c) 从 $\mathcal{B}$ 到 $\mathcal{E}$ 的过渡矩阵 $P_{\mathcal{B},\mathcal{E}}$ 是酉矩阵.

2. 将 V 的每一组有序规范正交基作为一个元素, 将所有这些元素构成的集合记为 $\mathrm{OB}(V)$. 则通过对应法则
$$\psi_{\mathcal{E}} \ : \ \mathbf{U}_n(\mathbb{C}) \longrightarrow \mathrm{OB}(V)\,; \quad Q \longmapsto \mathcal{E}\cdot Q$$
可以得到一个双射, 其逆映射是通过求过渡矩阵得到的如下映射
$$\phi_{\mathcal{E}} \ : \ \mathrm{OB}(V) \longrightarrow \mathbf{U}_n(\mathbb{C})\,; \quad \mathcal{B} \longmapsto P_{\mathcal{E},\mathcal{B}}\,.$$

定义 8.1.17. 设 $A, A' \in \mathbf{M}_n(\mathbb{C})$. 如果存在酉矩阵 $Q \in \mathbf{U}_n(\mathbb{C})$ 使得 $Q^{-1}AQ = A'$ 则称 A 和 A' **酉相似** (unitarily similar). ■

推论 8.1.18: 设 V 是 n 维酉空间, $\mathscr{A} \in \mathrm{End}(V)$, $\mathcal{E}$ 是 V 的一组规范正交基, $A = \mathcal{M}_{\mathcal{E}}(\mathscr{A}) \in \mathbf{M}_n(\mathbb{C})$.

则对于任意 $A' \in \mathbf{M}_n(\mathbb{C})$, A 和 A' 酉相似的充分必要条件是存在 V 的规范正交基 $\mathcal{B}$ 使得 $A' = \mathcal{M}_{\mathcal{B}}(\mathscr{A})$.

证明. 留给读者练习. □

以下定理是命题 7.1.28 的类似结论:

定理 8.1.19 (Schur[‡] **定理**): 设 V 是 n 维酉空间, $\mathscr{A} \in \mathrm{End}(V)$. 那么一定存在 V 的一组规范正交基 $\mathcal{E}$ 使得 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 是上三角阵.

换用矩阵的语言来说, 任意复方阵 $A \in \mathbf{M}_n(\mathbb{C})$ 都可以酉相似于一个上三角阵.

证明. 我们曾经证明 (参见本书上册命题 3.3.32): 任意复方阵均可上三角化. 将这个结论与 Gram–Schmidt 正交化 (定理 8.1.11) 结合起来即可. □

8.1.4 正交和分解与正交投影

(8.1.20) 设 V 为酉空间. 作为复向量空间, 其对偶空间 V° 是由从 V 到 $\mathbb{C}$ 的所有复线性泛函 (参见定义 6.1.21) 构成的, 即
$$V^\circ = \mathrm{Hom}_{\mathbb{C}}(V, C) = \{\mathbb{C}\text{-线性映射 } f : V \longrightarrow \mathbb{C}\}\,.$$
由 Hermite 型的半双线性可知: 对于每个 $u \in V$, 映射

(8.1.20.1) $$\psi_u \ : \ V \longrightarrow \mathbb{C}\,; \quad v \longmapsto \langle u\,,\, v\rangle$$

是 V 上的线性泛函. 我们可以考虑映射

(8.1.20.2) $$\Psi \ : \ V \longrightarrow V^\circ = \mathrm{Hom}_{\mathbb{C}}(V\,,\,\mathbb{C})\,; \quad u \longmapsto \psi_u\,.$$

不难验证, 映射 Ψ 具有以下性质:

(8.1.20.3) $$\begin{aligned} \text{对任意 } u_1, u_2 \in V\,, \quad & \Psi(u_1 + u_2) = \Psi(u_1) + \Psi(u_2)\,, \\ \text{对任意 } u \in V, c \in \mathbb{C}\,, \quad & \Psi(cu) = \overline{c}\,\Psi(u)\,. \end{aligned}$$

注意, (8.1.20.3) 式中的第二个公式表明, 一般情况下 Ψ 并不是复向量空间之间的线性映射! ■

定理 8.1.21 (Riesz **表示定理**): 设 V 为有限维酉空间. 则 (8.1.20.2) 中的映射 Ψ 是双射.

换言之, 对于任意线性泛函 $f : V \to \mathbb{C}$, 存在唯一的向量 $u \in V$ 使得 $f(v) = \langle u\,,\, v\rangle$ 对所有 $v \in V$ 成立.

[‡] Issai Schur (1875–1941) 德国数学家.

证明. 取定 V 的一组规范正交基 $e_1,\cdots,e_n$. 则任意线性泛函 $f\in V^\circ$ 由它在 $e_1,\cdots,e_n$ 这组向量上的取值唯一决定. 因此, 只要证明: 存在唯一的 $u\in V$ 使得 f 和 $\Psi(u)$ 满足

$$\text{(8.1.21.1)}\qquad \text{对每个 } i\in[\![1,\,n]\!]\,,\quad f(e_i)=\Psi(u)(e_i)=\langle\, u\,,\, e_i\,\rangle\,.$$

为此, 我们设待定的 u 可以写成 $u=a_1e_1+\cdots+a_ne_n$. 则 $\langle\, u\,,\, e_i\,\rangle=\langle\, a_ie_i\,,\, e_i\,\rangle=\overline{a}_i$. 因此, (在 $f\in V^\circ$ 给定之后) 满足 (8.1.21.1) 的 u 是存在唯一的, 即,

$$\text{(8.1.21.2)}\qquad u=\overline{f(e_1)}e_1+\cdots+\overline{f(e_n)}e_n\,.$$

注意, 以上论证过程还说明: (8.1.21.2) 的等号右边虽然表达式用到了选取的规范正交基 $e_1,\cdots,e_n$, 但实际上该表达式只依赖于线性泛函 f (也就是说, 无论规范正交基 $e_1,\cdots,e_n$ 如何选取, (8.1.21.2) 式右边表达式给出的向量总是一样的). □

注记 8.1.22. 细心的读者可能会发现, 我们证明定理 8.1.21 的方法和之前证明定理 6.1.23 (即关于非退化双线性型的 Riesz 表示定理) 的方法稍微是有些差别的. 过去在证明定理 6.1.23 时, 我们只用到了双线性型的“非退化”性质, 并不需要假设更强的“正定性”条件, 在证明过程中也不需要用到规范正交基. 而与那个定理的情况相比, 定理 8.1.21 的情况有一个重要的不同, 就是 (8.1.21.2) 中的映射 Ψ 并不是复向量空间之间的线性映射, 因此不能直接通过比较维数来说明它是同构.

不过, 定理 6.1.23 的证明思路也不是完全不可模仿的. 我们实际上需要稍微转变一下视角来考查 (8.1.21.2) 式中的映射 Ψ.

事实上, 对于复向量空间 V, 同一个集合 V 也可以视为一个实向量空间, 因为既然 V 和 $\mathbb{C}$ 中的元素之间都有数乘, 那么当然 V 和子集 $\mathbb{R}\subseteq\mathbb{C}$ 中的元素也都有数乘. (更一般地, 若 U 是域 K 上的向量空间, F 是 K 的一个子域, 那么 U 也具有 F-向量空间的结构.) 只不过, V 作为复向量空间和实向量空间的维数会有变化: $2\dim_{\mathbb{C}}V=\dim_{\mathbb{R}}V$.

如果在 (8.1.21.2) 式中把 V 和 $V^\circ=\mathrm{Hom}_{\mathbb{C}}(V\,,\,\mathbb{C})$ 都当作 $\mathbb{R}$ 上的线性空间, 那么 (8.1.20.3) 式说明 Ψ 是一个 $\mathbb{R}$-线性映射! 于是, 只要 Ψ 是单射, 那么可以断定 Ψ 是两个相同维数的实向量空间之间的同构, 因此 Ψ 是双射.

用这种思路的话, 实际上只需要假设 $\langle\cdot\,,\,\cdot\rangle$ 是 V 上的非退化 Hermite 型. ■

(8.1.23) 我们可以和过去双线性型的情况一样对 Hermite 型定义正交和分解的概念. 即, 假设复向量空间 V 上已经给定 Hermite 型 $\langle\cdot\,,\,\cdot\rangle$ 以及一些子空间 $V_1\,,\cdots,\,V_r$. 如果 $V_1+\cdots+V_r$ 是直和, 并且 V_i 之间关于 $\langle\cdot\,,\,\cdot\rangle$ 两两正交, 那么我们说 $V_1+\cdots+V_r$ 是一个**正交和**. 如果 $V=V_1+\cdots+V_r$ 且 $V_1+\cdots+V_r$ 是正交和, 则记 $V=V_1\perp\cdots\perp V_r$, 称之为 V (关于 $\langle\cdot\,,\,\cdot\rangle$) 的一个**正交和分解**. ■

命题 8.1.24: *设 U 是酉空间 V 中的有限维子空间. 则正交分解式 $V=U\perp U^\perp$ 成立.*

证明. 因为有了相应版本的 Riesz 表示定理 (定理 8.1.21), 此命题的证明可模仿命题 6.1.25 的证明来进行. 或者, 我们也可以像定理 8.1.21 的证明那样, 使用规范正交基来直接证明.

首先, Hermite 内积的非退化性质说明 $U\cap U^\perp=0$.

接下来只要再说明 $V=U+U^\perp$ 即可. 为此, 可以取定有限维子空间 U 的一组规范正交基 $e_1,\cdots,e_m$. 对任意 $v\in V$, 要将 v 分解为 $v=u+w$, 其中 $u\in U=\mathrm{span}(e_1,\cdots,e_m)$, $w\in U^\perp$, 就是要找出合适的系数 $a_1,\cdots,a_m$ 使得

$$\langle\, e_i\,,\, v-a_1e_1-\cdots-a_me_m\,\rangle=0\qquad \text{对每个 } i=1,\cdots,m \text{ 成立.}$$

事实上, 只要取 $a_i=\langle\, e_i,\, v\,\rangle$ 即可. □

和实内积空间的情况 (参见思考题 7.15) 一样, 对于酉空间 V 中的任意有限维子空间 U, 总有 $(U^\perp)^\perp=U$.

推论 8.1.25: 若 V 是有限维酉空间, 则对于任意子空间 $U \subseteq V$, $\dim U + \dim U^{\perp} = \dim V$.

证明. 由命题 8.1.24 显然. □

(8.1.26) 设 U 是酉空间 V 中的有限维子空间. 根据命题 8.1.24, 我们可以模仿定义 7.3.8 来定义 V 到 U 上的**正交投影** $P_U : V \to V$, 即, 对于任意 $v \in V$,

$$P_U v = \text{ 子空间 } U \text{ 中唯一能使 } v - u \in U^{\perp} \text{ 的那个向量 } u\,.$$

可以验证: 与命题 7.3.10 完全类似的性质仍然成立. 例如, 若 $e_1, \cdots, e_m$ 是 U 的一组规范正交基, 则

$$P_U v = \langle e_1,\, v \rangle e_1 + \cdots + \langle e_m,\, v \rangle e_m\,.$$

而且, 和命题 7.3.11 的结论一样, 对于给定的 $v \in V$,

$$\|v - P_U v\| = v \text{ 到子空间 } U \text{ 的距离 } := \min\{\|v-u\| \,:\, u \in U\}\,.$$

事实上, 子空间 U 中能够满足等式 $\|v - P_U v\| = \|v-u\|$ 的向量 $u \in U$ 只有 $u = P_U v$. ■

8.1.5 习题

习题 8.1.1. 在 $V = \mathbb{C}[X]_{\le 2}$ 上取定 Hermite 内积

$$\langle f\,,\, g \rangle := \sum_{k=0}^{2} \overline{f(k)} g(k)\,.$$

求此酉空间 $(V, \langle\cdot\,,\,\cdot\rangle)$ 的一组规范正交基.

习题 8.1.2. 求矩阵 $A = \begin{pmatrix} 2 & i \\ 1 & 0 \\ i & 1 \end{pmatrix}$ 的 QR 分解.

习题 8.1.3. 证明 Schur 不等式: 设 $\lambda_1, \cdots, \lambda_n$ 是复方阵 $A \in \mathbf{M}_n(\mathbb{C})$ 的所有特征值. 则

$$\mathrm{Tr}(A\overline{A}^T) \ge \sum_{i=1}^{n} |\lambda_i|^2$$

并且等号成立的充分必要条件是 A 为正规矩阵. (提示: 使用 Schur 定理.)

习题 8.1.4. 设 V 是有限维酉空间, $\mathscr{A} \in \mathrm{End}(V)$. 证明: $\mathscr{A}$ 是正规变换的充分必要条件是 $\mathscr{A}$ 的特征向量都是伴随变换 $\mathscr{A}^*$ 的特征向量.

习题 8.1.5. 设 $V = \mathbf{M}_n(\mathbb{C})$.

1. 证明: 映射
$$V \times V \longrightarrow \mathbb{C}\,; \quad (A,\, B) \longmapsto \langle A,\, B \rangle := \mathrm{Tr}(\overline{A}^T B)$$
是 V 上的 Hermite 内积.

2. 设 U 是 V 中的对角矩阵构成的子空间. 求 U 关于以上 Hermite 内积的正交补 $U^{\perp}$.

习题 8.1.6. 设 V 是 n 维复向量空间. 设 $\sigma : V \to V$ 是满足以下条件的一个映射 (称为一个**共轭映射**):

- 对任意 $u, v \in V$, $\sigma(u+v) = \sigma(u) + \sigma(v)$.

- 对任意 $c \in \mathbb{C}$, $v \in V$, $\sigma(cv) = \bar{c}\,\sigma(v)$.
- $\sigma^2 = I$.

令 $R_\sigma(V) := \{v \in V \mid \sigma(v) = v\}$.
证明:

1. $R_\sigma(V)$ 是 n 维实向量空间.
2. 每个 $v \in V$ 可以唯一地表示为 $v = u + iw$, 其中 $u, w \in R_\sigma(V)$.
3. 假设 $\langle\cdot\,,\,\cdot\rangle$ 是 $R_\sigma(V)$ 上的一个内积. 则以下映射

$$\begin{aligned}\langle\cdot\,,\,\cdot\rangle' :\quad & V \times V \longrightarrow V\\ & \langle v_1\,,\,v_2\rangle' := \langle u_1, u_2\rangle + \langle w_1, w_2\rangle + i\big(\langle u_1\,,\,w_2\rangle - \langle u_2\,,\,w_1\rangle\big)\end{aligned}$$

是 V 上的 Hermite 内积, 其中 $v_1 = u_1 + iw_1$ 和 $v_2 = u_2 + iw_2$ 是上面所说的唯一分解式, 其中 $u_1, u_2, w_1, w_2 \in R_\sigma(V)$.

习题 8.1.7. 设 V 为内积空间或酉空间, U, W 是 V 的有限维子空间. 证明 U 和 W 正交的充分必要条件是 $P_U P_W = 0$.

习题 8.1.8. 设 V 为内积空间或酉空间, $U \subseteq V$ 是个有限维子空间, $\mathscr{A} \in \mathrm{End}(V)$. 证明: U 是 $\mathscr{A}$ 的不变子空间当且仅当 $P_U \mathscr{A} P_U = \mathscr{A} P_U$.

习题 8.1.9. 设 V 是有限维酉空间, $\dim V \geq 1$. 设 $S \subseteq \mathrm{End}(V)$ 是由 V 上的自伴算子构成的子集. 证明: S 一定不是 $\mathrm{End}(V)$ 的线性子空间.

习题 8.1.10. 设 V 是内积空间或酉空间, $W \subseteq V$ 是有限维子空间, $\mathscr{A} \in \mathrm{End}(V)$ 满足 $\langle \mathscr{A}u, \mathscr{A}v\rangle = \langle u, v\rangle$ 对所有 $u, v \in V$ 成立.

则当 W 是 $\mathscr{A}$ 的不变子空间时, $W^\perp$ 也是 $\mathscr{A}$ 的不变子空间. (提示: $\mathscr{A}|_W : W \to W$ 是满射.)

8.2 酉空间上的线性变换

若非有相反的说明, 本节中 V 和 W 总是表示非零的有限维酉空间, 它们的 Hermite 内积记为 $\langle\cdot\,,\,\cdot\rangle_V$ 和 $\langle\cdot\,,\,\cdot\rangle_W$, 或者在不至于引起歧义时简单记为 $\langle\cdot\,,\,\cdot\rangle$. 类似地, V 中向量 v 的长度在需要强调时记为 $\|v\|_V$, 否则可简记为 $\|v\|$.

8.2.1 伴随映射与复正规变换

命题 8.2.1: 设 $\mathscr{A} : V \to W$ 是 (有限维酉空间之间的) 线性映射. 则存在唯一一个从 W 到 V 的映射, 通常记为 $\mathscr{A}^* : W \to V$, 使得

(8.2.1.1) $$\langle \mathscr{A}v\,,\,w\rangle = \langle v\,,\,\mathscr{A}^*w\rangle \quad \text{对于所有 } v \in V\,,\ w \in W \text{ 成立.}$$

这一映射 $\mathscr{A}^*$ 是线性映射, 称为 $\mathscr{A}$ 的**伴随**.

证明. 通过取复共轭, 可以将 (8.2.1.1) 式改写为如下等价形式:

(8.2.1.2) $$\langle w\,,\,\mathscr{A}v\rangle = \langle \mathscr{A}^*w\,,\,v\rangle \quad \text{对于所有 } v \in V\,,\ w \in W \text{ 成立.}$$

然后读者可以仿照命题 7.2.7 的证明来完成证明. 过程细节留给读者练习. □

读者不难验证, 与命题 7.2.10 类似的结论在我们现在讨论的酉空间的情况仍然成立. (事实上, 对于任意 $\mathscr{A} \in \mathrm{Hom}(V, W)$ 和任意 $c \in \mathbb{C}$, $(c\mathscr{A})^* = \overline{c}\mathscr{A}^*$. 除此之外命题 7.2.10 中的其他结论没有任何变化.)

思考题 8.5. 对酉空间的情况证明命题 7.2.11.

命题 8.2.2: *设 $\mathscr{A} \in \mathrm{Hom}(V, W)$, $\mathcal{B}$ 为 V 的一组有序规范正交基, $\mathcal{C}$ 为 W 的一组有序规范正交基. 则*

$$\mathcal{M}_{\mathcal{C},\mathcal{B}}(\mathscr{A}^*) = \overline{\mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})}^T.$$

也就是说, 在选择同一对有序规范正交基时, $\mathscr{A}$ 的矩阵和 $\mathscr{A}^*$ 的矩阵互为共轭转置.

证明. 留给读者练习或课堂授课时讲授. (提示: 会用到 (8.1.10.1) 中的第一个公式.) □

定义 8.2.3. 设 $\mathscr{A}$ 为 (有限维酉空间) V 上的线性变换.

1. 如果 $\mathscr{A}\mathscr{A}^* = \mathscr{A}^*\mathscr{A}$, 则称 $\mathscr{A}$ 是一个**正规变换**或**规范变换**.
2. 如果 $\mathscr{A} = \mathscr{A}^*$, 则称 $\mathscr{A}$ 是一个 Hermite **变换** (Hermitian transformation) 或**自伴变换**.
3. 如果 $\mathscr{A} = -\mathscr{A}^*$, 则称 $\mathscr{A}$ 是一个**斜** Hermite **变换** (skew-hermitian transformation).
4. 如果 $\mathscr{A}\mathscr{A}^* = I$, 则从 $\mathscr{A}$ 是一个**酉变换** (unitary transformation).

以上名词中“变换”一词也常被替换为“算子”.

之前在定义 8.1.3 和 8.1.14 中我们已经定义了 Hermite 矩阵和酉矩阵. 现在补充另外两个相关的定义.

如果矩阵 $A \in \mathbf{M}_n(\mathbb{C})$ 满足 $A\overline{A}^T = \overline{A}^T A$, 则称 A 是**正规矩阵**. (当 A 是实矩阵时, 此定义和定义 7.2.14 吻合.)

如果矩阵 $A \in \mathbf{M}_n(\mathbb{C})$ 满足 $A = -\overline{A}^T$, 则称 A 是**斜** Hermite **矩阵** (skew-hermitian matrix). (一个实矩阵是斜 Hermite 矩阵当且仅当它是斜对称的实矩阵.) ■

思考题 8.6. 设 $\mathscr{A}$ 为 (有限维酉空间) V 上的线性变换. 则下列陈述等价:

1. $\mathscr{A}$ 是正规变换 (Hermite 变换、斜 Hermite 变换、酉变换).
2. 对于 V 的任意一组规范正交基 $\mathcal{B}$, 矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是正规 (Hermite、斜 Hermite、酉) 矩阵.
3. 存在 V 的一组规范正交基 $\mathcal{B}$ 使得矩阵 $\mathcal{M}_{\mathcal{B}}(\mathscr{A})$ 是正规 (Hermite、斜 Hermite、酉) 矩阵.

思考题 8.7. 仿照定义 7.2.1 给出有限维酉空间之间的 (线性) **等距映射**和**等距同构**的定义. 然后叙述命题 7.2.2 在酉空间情况下的对应版本.

再说明有限维酉空间上的线性变换是等距同构当且仅当该变换是酉变换.

下面的引理和引理 7.2.17 的结论很相似. 不同的一点在于, 现在我们讨论的是复向量空间上的线性变换, 这种情况相对于实向量空间的情况有一些额外的性质可以利用. 由此我们发现, 下面的引理不再需要引理 7.2.17 中“自伴”的假设.

引理 8.2.4: *设 T 是 (有限维酉空间) V 上的线性变换.*

如果对于所有 $v \in V$ 均有 $\langle Tv, v\rangle = 0$, 则必有 $T = 0$.

证明. 根据 Schur 定理 (定理 8.1.19), 存在 V 的一组规范正交基 $\mathcal{E}=(e_1,\cdots,e_n)$ 使得 $A:=\mathcal{M}_{\mathcal{E}}(T)$ 是上三角阵. 注意到 A 的第 (i,j) 位置元素为 $\langle e_i, Te_j\rangle$. 我们只需证明对于所有的 $i,j\in[\![1,n]\!]$, 均有 $\langle e_i, Te_j\rangle=0$.

若 $i>j$, 则结论由 A 是上三角阵可知. 若 $i=j$, 则结论由题设条件保证. 因此不妨设 $i<j$.

我们将题设条件用于 $v=e_i+e_j$ 得到

$$\begin{aligned}0&=\langle e_i+e_j\,,\,T(e_i+e_j)\rangle\\&=\langle e_i, Te_i\rangle+\langle e_j, Te_i\rangle+\langle e_i\,,Te_j\rangle+\langle e_j\,,Te_j\rangle\\&=0+\langle e_j, Te_i\rangle+\langle e_i\,,Te_j\rangle+0\,.\end{aligned}$$

因为 $i<j$ 且 A 为上三角阵, 所以 $\langle e_j, Te_i\rangle=0$. 所以由上式可得 $\langle e_i, Te_j\rangle=0$. □

思考题 8.8. 证明: 对于酉空间 V 中的任意向量 u,w 和任意线性变换 $T\in\mathrm{End}(V)$, 以下恒等式成立:

$$\begin{aligned}4\langle Tu,\,w\rangle=\langle T(u+w)\,,\,u+w\rangle-\langle T(u-w)\,,\,u-w\rangle\\-i\langle T(u+iw)\,,\,u+iw\rangle+i\langle T(u-iw)\,,\,u-iw\rangle\,.\end{aligned}$$

再利用以上恒等式给出引理 8.2.4 的另一个证明.

命题 8.2.5: 设 $\mathscr{A}$ 是 (有限维酉空间) V 上的线性变换.

则 $\mathscr{A}$ 是正规变换的充分必要条件是对于所有 $v\in V$, $\|\mathscr{A}v\|=\|\mathscr{A}^*v\|$.

证明. 模仿命题 7.2.19 的证明即可. □

命题 8.2.6: 设 $\mathscr{A}$ 是 (有限维酉空间) V 上的正规变换, $\lambda\in\mathbb{C}$ 是 $\mathscr{A}$ 的一个特征值, $v\in V$ 是 $\mathscr{A}$ 的属于 λ 的一个特征向量.

则 $\overline{\lambda}$ 是 $\mathscr{A}^*$ 的特征值, 并且 v 是 $\mathscr{A}^*$ 的属于 $\overline{\lambda}$ 的特征向量.

证明. 留给读者练习. □

命题 8.2.7: 设 $\mathscr{A}$ 是 (有限维酉空间) V 上的正规变换, λ,μ 是 $\mathscr{A}$ 的两个不同特征值, $v,w\in V$ 是分别属于 λ,μ 的特征向量.

则 v 与 w 正交. 所以, 特征子空间 $E(\lambda,\mathscr{A})$ 和 $E(\mu,\mathscr{A})$ 相互正交.

证明. 留给读者练习. □

思考题 8.9. 设 $\mathscr{A}$ 是 (有限维酉空间) V 上的线性变换.

1. 证明: 复数 λ 是 $\mathscr{A}$ 的特征值当且仅当 $\overline{\lambda}$ 是 $\mathscr{A}^*$ 的特征值.

2. 证明: 若 $\mathscr{A}$ 是自伴变换 (亦即 Hermite 变换), 则 $\mathscr{A}$ 的特征值都是实数.

8.2.2 复正规算子的谱定理和复正规矩阵的酉相似标准形

定理 8.2.8 (**复正规算子的谱定理**): 设 $\mathscr{A}$ 是 (有限维酉空间) V 上的线性变换. 则下列陈述等价.

1. $\mathscr{A}$ 是正规算子.

2. 存在 V 的一组规范正交基, 其中向量全都是 $\mathscr{A}$ 的特征向量.

3. 存在 V 的一组规范正交基 $\mathcal{E}$ 使矩阵 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 是对角阵.

证明. (2) 和 (3) 的等价性是一种抽象废话. (3) 蕴含 (1) 也是容易的.

以下我们来证明 (1) 蕴含 (3). 读者可以模仿定理 7.2.23 的证明, 采用对维数的归纳法来完成证明 (参见下面的思考题 8.10). 不过, 此处我们还可以使用 Schur 定理给出一个简洁的证明.

根据 Schur 定理 (定理 8.1.19), 可以找到 V 的一组规范正交基 $\mathcal{E}=(e_1,\cdots,e_n)$ 使得矩阵 $A=(a_{ij})=\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 是上三角阵. 我们的目标是利用 $\mathscr{A}$ 是正规变换的条件, 证明矩阵 A 一定是对角阵. 这里的一个重要工具是命题 8.2.5 给出的正规变换的性质.

首先, 由于 $\mathcal{E}$ 是规范正交基, 所以由 $A=\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 可知 $\overline{A}^T=\mathcal{M}_{\mathcal{E}}(\mathscr{A}^*)$. 于是

$$\|\mathscr{A}e_1\|^2=|a_{11}|^2\,,\quad \|\mathscr{A}^*e_1\|^2=|a_{11}|^2+|a_{12}|^2+\cdots+|a_{1n}|^2\,.$$

由命题 8.2.5 可得 $\|\mathscr{A}e_1\|=\|\mathscr{A}^*e_1\|$. 所以以上两个等式说明 $a_{12}=\cdots=a_{1n}=0$. 接下来,

$$\|\mathscr{A}e_2\|^2=|a_{12}|^2+|a_{22}|^2=|a_{22}|^2\,,\quad \|\mathscr{A}^*e_2\|^2=|a_{22}|^2+|a_{23}|^2+\cdots+|a_{2n}|^2$$

通过 $\|\mathscr{A}e_2\|=\|\mathscr{A}^*e_2\|$ 来对比以上两个等式可得 $a_{23}=\cdots=a_{2n}=0$. 如此继续下去可以证明上三角阵 $A=(a_{ij})$ 的确是对角阵. □

思考题 8.10. 设 $\mathscr{A}$ 是 (有限维酉空间) V 上的正规变换, $U\subseteq V$ 是 $\mathscr{A}$ 的不变子空间.

1. 证明引理 7.2.30 在酉空间的情况也成立. 即,
   (a) $U^{\perp}$ 也是 $\mathscr{A}$ 的不变子空间, U 也是 $\mathscr{A}^*$ 的不变子空间.
   (b) $(\mathscr{A}|_U)^*=(\mathscr{A}^*)|_U$.
   (c) $\mathscr{A}|_U\in\mathrm{End}(U)$ 和 $\mathscr{A}|_{U^\perp}\in\mathrm{End}(U^\perp)$ 也都是正规变换.
2. 由此再给出定理 8.2.8 的一个证明.

将定理 8.2.8 换成矩阵语言表述, 可以得到以下结论:

定理 8.2.9 (**复正规矩阵的酉相似标准形**): *设 $A\in\mathbf{M}_n(\mathbb{C})$. 则 A 是正规矩阵的充分必要条件是 A 可以酉相似于一个对角矩阵.*

将定理 8.2.9 分别应用于几类特殊的复正规矩阵可以得到以下一些推论.

推论 8.2.10: *复方阵 $A\in\mathbf{M}_n(\mathbb{C})$ 是 Hermite 矩阵的充分必要条件是 A 可以酉相似于一个实对角阵.*

证明. 留作练习. □

推论 8.2.11: *复方阵 $A\in\mathbf{M}_n(\mathbb{C})$ 是斜 Hermite 矩阵的充分必要条件是 A 可以酉相似于一个 iD 形式的矩阵, 其中 $i=\sqrt{-1}\in\mathbb{C}$ 是虚数单位, D 是一个实对角阵.*

证明. 留作练习. □

推论 8.2.12: *复方阵 $A\in\mathbf{M}_n(\mathbb{C})$ 是酉矩阵的充分必要条件是 A 可以酉相似于一个如下形式的矩阵*

$$\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}\qquad \text{其中 } \lambda_1,\cdots,\lambda_n\in\mathbb{C} \text{ 的模均为 } 1\,.$$

证明. 留作练习. □

8.2.3 习题

以下总设 V 是非零的有限维酉空间, $n \in \mathbb{N}^*$.

习题 8.2.1. 设 V 是 n 维酉空间, $\mathscr{A} \in \mathrm{End}(V)$ 是酉变换. 假设 $\lambda_1, \cdots, \lambda_n$ 是 $\mathscr{A}$ 的所有特征值 (重数计入). 求 $\mathscr{A}^{-1}$ 的所有特征值.

习题 8.2.2. 将一个复方阵 $A \in \mathbf{M}_n(\mathbb{C})$ 分解为 $A = P + iQ$, 其中 $P, Q \in \mathbf{M}_n(\mathbb{R})$. 证明: A 是酉矩阵当且仅当以下两个条件成立: P^TQ 是对称阵, $P^TP + Q^TQ = I_n$.

习题 8.2.3. 记 $\xi = \exp\left(\frac{2\pi i}{n}\right) = \cos(\frac{2\pi}{n}) + i\sin(\frac{2\pi}{n}) \in \mathbb{C}$. 证明复矩阵

$$U := \frac{1}{\sqrt{n}}\begin{pmatrix} 1 & 1 & \cdots & \cdots & 1 \\ 1 & \xi & \xi^2 & \cdots & \xi^{n-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & \xi^{n-2} & \xi^{2(n-2)} & \ddots & \xi^{(n-2)(n-1)} \\ 1 & \xi^{n-1} & \xi^{2(n-1)} & \cdots & \xi^{(n-1)^2} \end{pmatrix}$$

是酉矩阵.

习题 8.2.4. 证明: 任意一个二阶酉矩阵 U 可以分解为

$$U = \begin{pmatrix} e^{i\theta_1} & 0 \\ 0 & e^{i\theta_2} \end{pmatrix} \cdot \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} e^{i\theta_3} & 0 \\ 0 & e^{i\theta_4} \end{pmatrix}$$

其中 $\theta, \theta_i \in \mathbb{R}$.

习题 8.2.5. 设 $\mathscr{A}, \mathscr{B} \in \mathrm{End}(V)$ 均为自伴算子. 证明: $\mathscr{A}\mathscr{B}$ 是自伴算子的充分必要条件是 $\mathscr{A}\mathscr{B} = \mathscr{B}\mathscr{A}$.

习题 8.2.6. 设 $\mathscr{A}, \mathscr{B} \in \mathrm{End}(V)$ 均为自伴算子. 证明: $\mathscr{A}\mathscr{B} + \mathscr{B}\mathscr{A}$ 和 $i(\mathscr{A}\mathscr{B} - \mathscr{B}\mathscr{A})$ 都是自伴算子.

习题 8.2.7. 对任意 $\mathscr{A} \in \mathrm{End}(V)$, 证明: 如果 $\mathscr{A}$ 满足以下三个条件中的两个, 则 $\mathscr{A}$ 也满足第三个条件: (1) $\mathscr{A}$ 自伴; (2) $\mathscr{A}$ 是酉变换; (3) $\mathscr{A}^2 = I$.

8.3 补充专题

本节我们设 $(V, \langle\cdot\,,\,\cdot\rangle)$ 是有限维的实内积空间或酉空间, $\dim V \geq 1$.

8.3.1 半正定算子

定义 8.3.1. 设 $\mathscr{A}$ 是 V 上的线性算子, $A \in \mathbf{M}_n(\mathbb{C})$.

1. 如果 $\mathscr{A}$ 是自伴的并且对于任意非零向量 $v \in V$, $\langle \mathscr{A}v, v\rangle$ 总是非负实数, 则称 $\mathscr{A}$ 是个**半正定算子** (positive semi-definite operator). 有的书 (例如 [3]) 使用的称呼是**正算子** (positive operator).

2. 如果 $\mathscr{A}$ 是自伴的并且对于任意非零向量 $v \in V$, $\langle \mathscr{A}v, v\rangle$ 总是正实数, 则称 $\mathscr{A}$ 是个**正定算子** (positive definite operator).

3. 如果 A 是 Hermite 矩阵, 并且对于任意非零列向量 $x \in \mathbb{C}^n$, $\overline{x}^TAx$ 总是非负 (正) 实数, 则称 A 是**半正定的** (**正定的**). (如果 A 是实矩阵, 此处的定义和之前的定义吻合.)

不难看出, $\mathscr{A} \in \operatorname{End}(V)$ 是半正定 (正定) 的, 当且仅当存在 V 的一组规范正交基 $\mathcal{E}$ 使得矩阵 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 是半正定 (正定) 的, 当且仅当对于 V 的任意一组规范正交基 $\mathcal{E}$ 矩阵 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 总是半正定 (正定) 的.

当然, 我们可以定义**半负定** (**负定**) 算子为满足 $-\mathscr{B}$ 半正定 (正定) 的线性算子 $\mathscr{B} \in \operatorname{End}(V)$. ■

思考题 8.11. 设 V 是有限维酉空间, $\mathscr{A} \in \operatorname{End}(V)$. 证明下列条件等价:

1. $\mathscr{A}$ 是自伴算子 (亦即 Hermite 变换).
2. 对任意 $v \in V$, $\langle \mathscr{A}v\,, v\rangle$ 是实数.

(这说明: 对于酉空间的情况, 定义半正定 (正定) 算子时, "自伴"的条件可以由另一个条件推出.)

(提示: 利用引理 8.2.4.)

命题 8.3.2: *对任意 $\mathscr{A} \in \operatorname{End}(V)$, 下列陈述等价:*

1. *$\mathscr{A}$ 是半正定算子.*
2. *$\mathscr{A}$ 是自伴算子, 且 $\mathscr{A}$ 的所有特征值都是非负实数.*
3. *$\mathscr{A}$ 有半正定的平方根, 即, 存在半正定算子 $\mathscr{P} \in \operatorname{End}(V)$ 使 $\mathscr{A} = \mathscr{P}^2$.*
4. *$\mathscr{A}$ 有自伴的平方根, 即, 存在自伴算子 $\mathscr{P} \in \operatorname{End}(V)$ 使 $\mathscr{A} = \mathscr{P}^2$.*
5. *存在 $\mathscr{B} \in \operatorname{End}(V)$ 使得 $\mathscr{A} = \mathscr{B}^*\mathscr{B}$.*

证明. (1) ⇒ (2). 若 $\mathscr{A}$ 半正定, 则按照定义 $\mathscr{A}$ 是自伴的. 对于 $\mathscr{A}$ 的任意特征值 λ, 取一个对应的特征向量 v. 则

$$\lambda\langle v, v\rangle = \langle v\,, \lambda v\rangle = \langle v\,, \mathscr{A}v\rangle = \overline{\langle \mathscr{A}v\,, v\rangle}\,.$$

根据 $\mathscr{A}$ 半正定的假设, $\langle \mathscr{A}v\,, v\rangle$ 是非负实数, 因此上式表明 $\lambda\langle v\,, v\rangle$ 是非负实数. 因为 $v \neq 0$, 故 $\langle v\,, v\rangle = \|v\|^2$ 是正实数. 所以, λ 必然是非负实数.

(2) ⇒ (3). 根据谱定理 (定理 7.2.23 和 8.2.8), 存在 V 的一组规范正交基 $\mathcal{E} = (e_1, \cdots, e_n)$ 使得 $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 是对角阵. 根据 (2) 的假设, $\mathcal{M}_{\mathcal{E}}(\mathscr{A})$ 的对角线元素 λ_i 均为非负实数. 定义 $\mathscr{P} \in \operatorname{End}(V)$ 使得 $\mathscr{P}e_i = \sqrt{\lambda_i}e_i$. 很容易验证 $\mathscr{P}$ 是半正定算子, 而且 $\mathscr{P}^2 = \mathscr{A}$.

(3) ⇒ (4). 因为按定义半正定算子是自伴的, 故结论显然.

(4) ⇒ (5). 取 $\mathscr{B} = \mathscr{P}$ 即可.

(5) ⇒ (1). 若 $\mathscr{A} = \mathscr{B}^*\mathscr{B}$, 直接验证可知 $\mathscr{A}$ 是自伴的. 对于任意 $v \in V$,

$$\langle \mathscr{A}v\,, v\rangle = \langle \mathscr{B}^*\mathscr{B}v\,, v\rangle = \overline{\langle v\,, \mathscr{B}^*\mathscr{B}v\rangle} = \overline{\langle \mathscr{B}v\,, \mathscr{B}v\rangle} = \overline{\|\mathscr{B}v\|^2} = \|\mathscr{B}v\|^2\,.$$

因为向量的长度总是非负实数, 所以上式可以说明 $\langle \mathscr{A}v, v\rangle$ 总是非负实数. 按定义, 这就是说 $\mathscr{A}$ 半正定. □

细心的读者不难发现, 命题 8.3.2 中的条件 5 与命题 6.2.22 中的条件 2 之间有明显的相似性.

思考题 8.12. 对正定算子叙述并证明与命题 8.3.2 类似的结论.

命题 8.3.3: *若 $\mathscr{A} \in \operatorname{End}(V)$ 是半正定算子, 则存在唯一的半正定算子 $\mathscr{P} \in \operatorname{End}(V)$ 使 $\mathscr{A} = \mathscr{P}^2$. 我们将这个半正定算子记为 $\sqrt{\mathscr{A}}$.*

以矩阵语言言之, 若 $A \in \mathbf{M}_n(\mathbb{C})$ 是半正定 Hermite 矩阵, 则存在唯一的半正定 Hermite 阵是 A 的平方根, 我们将其记为 $\sqrt{A}$.

证明. 存在性的部分已经在命题 8.3.2 中讨论过了. 具体来说, 因为 $\mathscr{A}$ 是自伴的, 所以存在 V 的规范正交基 $\mathcal{E} = (e_1, \cdots, e_n)$ 完全由 $\mathscr{A}$ 的特征向量构成. 记 λ_i 为 e_i 对应的特征值, 则 $\mathscr{A}$ 的半正定性保证 λ_i 都是非负实数. 通过定义 $\mathscr{P}e_i = \sqrt{\lambda_i}e_i$ 即可得到 $\mathscr{A}$ 的一个半正定平方根.

现在设 $\mathscr{R} \in \operatorname{End}(V)$ 也是半正定算子, 且 $\mathscr{A} = \mathscr{R}^2$. 为了说明 $\mathscr{R} = \mathscr{P}$, 只需要证明 $\mathscr{R}e_i = \sqrt{\lambda_i}e_i$ 对每个 $i \in [\![1, n]\!]$ 成立. 不妨取定指标 i, 令 $v = e_i$, $\lambda = \lambda_i$. 下面来证明 $\mathscr{R}v = \sqrt{\lambda}v$.

因为 $\mathscr{R}$ 是半正定的, 所以 $\mathscr{R}$ 是自伴的, 且所有特征值都是非负实数. 我们将 $\mathscr{R}$ 的特征值写成 $\sqrt{a_i}$ 的形式, 其中每个 a_i 都是非负实数. 和先前对 $\mathscr{A}$ 的讨论一样, 可以选取 V 的一组规范正交基 $\mathscr{F} = (f_1, \cdots, f_n)$ 完全由 $\mathscr{R}$ 的特征向量构成. 于是不妨设对于每个 $i \in [\![1, n]\!]$, $\mathscr{R}f_i = \sqrt{a_i}f_i$.

现在将 v 写成 $f_1, \cdots, f_n$ 的线性组合: $v = c_1f_1 + \cdots + c_nf_n$. 则

$$\begin{aligned}
\mathscr{R}v &= c_1\mathscr{R}f_1 + \cdots + c_n\mathscr{R}f_n = c_1\sqrt{a_1}f_1 + \cdots + c_n\sqrt{a_n}f_n \\
\mathscr{R}^2v &= \mathscr{R}(c_1\sqrt{a_1}f_1 + \cdots + c_n\sqrt{a_n}f_n) \\
&= c_1\sqrt{a_1}\mathscr{R}f_1 + \cdots + c_n\sqrt{a_n}f_n = c_1a_1f_1 + \cdots + c_na_nf_n\,.
\end{aligned}$$

因为 $\mathscr{R}^2 = \mathscr{A}$ 且 $\mathscr{A}v = \lambda v$, 所以综上可得

$$\lambda(c_1f_1 + \cdots + c_nf_n) = \lambda v = \mathscr{R}^2v = c_1a_1f_1 + \cdots + c_na_nf_n\,.$$

于是, 对于每个 $i \in [\![1, n]\!]$, 均有 $c_i(\lambda - a_i) = 0$. 容易看出, 这蕴含着 $c_i\sqrt{\lambda} = c_i\sqrt{a_i}$ 对每个 $i \in [\![1, n]\!]$ 成立. 因此

$$\mathscr{R}v = \sum_{i=1}^{n} c_i\sqrt{a_i}f_i = \sum_{i=1}^{n} c_i\sqrt{\lambda}f_i = \sqrt{\lambda}\left(\sum_{i=1}^{n} c_if_i\right) = \sqrt{\lambda}v\,.$$

这就证明了所需的结论. □

思考题 8.13. 证明: 若 $\mathscr{A} \in \operatorname{End}(V)$ 是正定算子, 则 $\sqrt{\mathscr{A}}$ 也是正定的.

8.3.2 奇异值分解与极分解

我们知道, 即使在 $\mathbb{C}$ 上考虑, 一般的方阵也不一定可以相似到对角矩阵. 而如果我们把相似放宽为相抵, 那么任意域 K 上的任意矩阵 $A \in \mathbf{M}_{m\times n}(K)$ 总是可以相抵于一个 $\begin{pmatrix} I_r & \\ & \mathbf{0} \end{pmatrix}$ 形式的矩阵 (这里的 r 是 A 的秩, $\mathbf{0}$ 是大小为 $(m-r)\times(n-r)$ 的零矩阵). 该矩阵称为 A 的相抵标准形 (参见本书上册定理 1.2.35). 前面一节的谱定理告诉了我们复方阵什么时候能够酉相似于对角阵. 那么如果考虑用酉矩阵做相抵, 是不是所有的复矩阵都可以有一种"酉相抵标准形"呢?

这一小节的第一个目标就是给出这个问题的答案. 事实上, 这个问题的解答和数值线性代数中非常有用的奇异值分解有关.

先来为即将出现的几个重要概念给出定义.

定义 8.3.4. 设 $A, A' \in \mathbf{M}_{m\times n}(\mathbb{C})$. 如果存在酉矩阵 $U \in \mathbf{U}_m(\mathbb{C})$ 和 $U' \in \mathbf{U}_n(\mathbb{C})$ 使得 $UAU' = A'$, 那么我们称 A 和 A' **酉相抵** (unitarily equivalent).

类似地, 对于两个实矩阵 $B, B' \in \mathbf{M}_{m\times n}(\mathbb{R})$, 如果存在实正交矩阵 $Q \in \mathbf{O}_m(\mathbb{R})$ 和 $Q' \in \mathbf{O}_n(\mathbb{R})$ 使得 $QBQ' = B'$, 则称 B 和 B' (实) **正交相抵** (orthogonally equivalent). ■

思考题 8.14. 设 $\mathscr{A} : V \to W$ 是有限维酉空间 (实内积空间) 之间的线性映射, $\mathcal{B}, \mathcal{C}$ 分别为 V, W 的一组规范正交基, $A = \mathcal{M}_{\mathcal{B},\mathcal{C}}(\mathscr{A})$ 是相应的矩阵.

则对于任意大小与 A 相同的矩阵 A', 证明: A 和 A' 酉相抵 (正交相抵) 的充分必要条件是存在 V, W 的规范正交基 $\mathcal{B}', \mathcal{C}'$ 使得 $A' = \mathcal{M}_{\mathcal{B}',\mathcal{C}'}(\mathscr{A})$.

定义 8.3.5. 设 V, W 均为非零的有限维实内积空间或酉空间, $\mathscr{A} \in \mathrm{Hom}(V, W)$, $\mathscr{A}^* \in \mathrm{Hom}(W, V)$ 是 $\mathscr{A}$ 的伴随. (容易验证 $\mathscr{A}^*\mathscr{A} \in \mathrm{End}(V)$ 和 $\mathscr{A}\mathscr{A}^* \in \mathrm{End}(W)$ 都是半正定算子, 因此它们的特征值都是非负实数.)

设 $\lambda_1, \cdots, \lambda_r$ 是 $\mathscr{A}^*\mathscr{A} \in \mathrm{End}(V)$ 的所有非零特征值 (重数计入). 则 $\sigma_i = \sqrt{\lambda_i}$ 称为 $\mathscr{A}$ 的**奇异值** (singular value[§]), 而 λ_i 作为 $\mathscr{A}^*\mathscr{A}$ 特征值的重数也称为 σ_i 作为 $\mathscr{A}$ 的奇异值的**重数**.

通常我们列举奇异值时都计入重数. 即, 当我们说 $\sigma_1, \cdots, \sigma_r$ 是 $\mathscr{A}$ 的所有特征值时, 其中的 σ_i 可以有某些相同的值重复出现, 重复的次数就是相应的奇异值重数. 读者不难验证: $\mathscr{A} \in \mathrm{Hom}(V, W)$ 的所有奇异值 (重数计入) 恰好是半正定算子 $\sqrt{\mathscr{A}^*\mathscr{A}} \in \mathrm{End}(V)$ 的所有正特征值 (重数计入).

一个矩阵 $A \in \mathbf{M}_{m\times n}(\mathbb{C})$ 的奇异值定义为 $\overline{A}^T A$ 的正特征值的平方根. 等价地说, A 的奇异值定义为线性映射 $\mathscr{A} : \mathbb{C}^n \to \mathbb{C}^m$, $x \mapsto Ax$ 的奇异值. ■

注记 8.3.6. 对于任意 $m \times n$ 矩阵 A 和 $n \times m$ 矩阵 B, 通过分块矩阵和行列式的技巧可以证明 (参见本书上册例 4.3.17):

$$X^n \det(XI_m - AB) = X^m \det(XI_n - BA)\,.$$

由此可知, 对于定义 8.3.5 中的线性映射 $\mathscr{A}$, 线性变换 $\mathscr{A}^*\mathscr{A} \in \mathrm{End}(V)$ 和 $\mathscr{A}\mathscr{A}^* \in \mathrm{End}(W)$ 的非零特征值 (连同重数) 全都对应相等. 所以, $\mathscr{A}$ 的奇异值也可以定义为 $\mathscr{A}\mathscr{A}^*$ 的非零特征值的平方根. 换言之, $\mathscr{A}$ 和 $\mathscr{A}^*$ 的奇异值 (连同重数) 完全相同. ■

思考题 8.15. 设 V, W 均为非零的有限维实内积空间或酉空间, $\mathscr{A} \in \mathrm{Hom}(V, W)$. 证明: $\mathrm{rank}(\mathscr{A}) = \mathrm{rank}(\mathscr{A}\mathscr{A}^*) = \mathrm{rank}(\mathscr{A}^*\mathscr{A})$.

(因此, 只要 $\mathscr{A} \neq 0$, 线性变换 $\mathscr{A}^*\mathscr{A}$ 一定有非零的特征值, 从而 $\mathscr{A}$ 一定有奇异值.)

定理 8.3.7 (**奇异值分解** Singular Value Decomposition): *设 V, W 均为非零的有限维实内积空间或酉空间, $\mathscr{A} \in \mathrm{Hom}(V, W)$.*

则一定存在 V 的一组规范正交基 $\mathcal{E}$ 和 W 的一组规范正交基 $\mathcal{F}$ 使得

$$\mathcal{M}_{\mathcal{E},\mathcal{F}}(\mathscr{A}) = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}$$

其中 $\sigma_1 \geq \cdots \geq \sigma_r > 0$ 是 $\mathscr{A}$ 的所有奇异值 (重数计入), $n = \dim V$, $m = \dim W$.

证明. 为找出满足条件的规范正交基 $\mathcal{E} = (e_1, \cdots, e_r, \cdots, e_n)$ 和 $\mathcal{F} = (f_1, \cdots, f_r, \cdots, f_m)$, 首先来看它们需要满足的一些必要条件. 事实上, 若记 $A = \mathcal{M}_{\mathcal{E},\mathcal{F}}(\mathscr{A})$, 则 A 具有定理所说的形式意味着

$$\begin{cases} \mathscr{A}e_j = \sigma_j f_j\,, & \text{当 } j \in [\![1, r]\!]\,, \\ \mathscr{A}e_j = 0\,, & \text{当 } j > r\,. \end{cases} \tag{8.3.7.1}$$

此外, 因为 $\mathcal{E}$, $\mathcal{F}$ 是规范正交基, 所以

$$\mathcal{M}_{\mathcal{F},\mathcal{E}}(\mathscr{A}^*) = \overline{A}^T = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0}_{(n-r)\times(m-r)} \end{pmatrix}$$

[§]有些书 (例如 [3]) 中也允许 0 为奇异值. 我们不采用这个约定, 而是像 [18] 等书一样, 在定义中要求奇异值非零. 这样做的一个好处是: 即使 $\dim V$ 和 $\dim W$ 不相等, $\mathscr{A}$ 和 $\mathscr{A}^*$ 的奇异值 (连同重数) 总是完全相同的 (参见注记 8.3.6).

于是

$$\mathcal{M}_{\mathcal{E}}(\mathscr{A}^*\mathscr{A}) = \overline{A}^T A = \begin{pmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_r^2 & \\ & & & \mathbf{0}_{(n-r)\times(n-r)} \end{pmatrix}.$$

也就是说, $\mathcal{E}$ 是 V 的一组规范正交基使得 $\mathscr{A}^*\mathscr{A}$ 的矩阵是对角阵. 更确切地说,

$$\begin{cases} \mathscr{A}^*\mathscr{A}e_j = \sigma_j^2 e_j\,, & \text{当 } j \in [\![1,\, r]\!]\,, \\ \mathscr{A}^*\mathscr{A}e_j = 0\,, & \text{当 } j > r\,. \end{cases} \tag{8.3.7.2}$$

来分析一下 (8.3.7.1) 和 (8.3.7.2) 给我们什么信息.

首先, (8.3.7.1) 式的第一行说明: 当 $1 \le j \le r$ 时, f_j 只能等于 $\sigma_j^{-1}\mathscr{A}e_j$. 也就是说, $\mathcal{F}$ 中的前 r 个向量由 $\mathcal{E}$ 中的前 r 个向量 (和 $\mathscr{A}$) 确定, 而后面的向量 $f_{r+1}, \cdots, f_m$ 只要能和前面的 $f_1, \cdots, f_r$ 构成 W 的规范正交基就行了, 没有其他要求. 此外, 当 (8.3.7.1) 式的第一行成立时, 对任意 $i, j \in [\![1,\, r]\!]$,

$$\begin{aligned} \langle\, f_i,\, f_j \,\rangle &= \sigma_i^{-1}\sigma_j^{-1}\langle\, \mathscr{A}e_i\,,\, \mathscr{A}e_j \,\rangle = \sigma_i^{-1}\sigma_j^{-1}\langle\, e_i,\, \mathscr{A}^*\mathscr{A}e_j \,\rangle \\ \text{根据 (8.3.7.2) 第一行 } &= \sigma_i^{-1}\sigma_j^{-1}\langle\, e_i,\, \sigma_j^2 e_j \,\rangle = (\sigma_j/\sigma_i)\langle\, e_i,\, e_j \,\rangle \\ &= \delta_{ij} \text{ (Kronecker delta 符号)}\,. \end{aligned}$$

这说明, 只要按照 $f_j = \sigma_j^{-1}\mathscr{A}e_j$ 来选取 $f_1, \cdots, f_r$, 那么它们自动构成规范正交组.

因此, 要找出合适的 $\mathcal{E}$ 和 $\mathcal{F}$, 只要先找到合适的 $\mathcal{E}$, 然后令

$$f_1 = \sigma_1^{-1}\mathscr{A}e_1\,, \cdots, f_r = \sigma_r^{-1}\mathscr{A}e_r$$

再将 $f_1, \cdots, f_r$ 扩充为 W 的一组规范正交基 $\mathcal{F}$ 即可.

要寻找的规范正交基 $\mathcal{E}$ 需要满足 (8.3.7.2) 式. 因为 $\mathscr{A}^*\mathscr{A}$ 是半正定算子 (故而是自伴算子), 由谱定理可知存在 V 的规范正交基 $\mathcal{E} = (e_1, \cdots, e_r, \cdots, e_n)$ 满足 (8.3.7.2) 式. 然后我们按照刚才说的方法找出 W 的一组规范正交基 $\mathcal{F}$. 剩下只要验证一下在 $j > r$ 时 $\mathscr{A}e_j = 0$ (即 (8.3.7.1) 式的第二行) 也成立即可.

事实上, 当 $j > r$ 时

$$\|\mathscr{A}e_j\|^2 = \langle\, \mathscr{A}e_j\,,\, \mathscr{A}e_j \,\rangle = \langle\, e_j,\, \mathscr{A}^*\mathscr{A}e_j \,\rangle\,.$$

由 (8.3.7.2) 式的第二行可知上式最右端等于 0. 定理至此证毕. □

将定理 8.3.7 以矩阵形式表述, 可以得到:

定理 8.3.8 (**矩阵的酉** (**正交**) **相抵标准形**): *设 $m, n \in \mathbb{N}^*$, $A \in \mathbf{M}_{m\times n}(\mathbb{C})$. 设 $\sigma_1 \ge \cdots \ge \sigma_r > 0$ 为 A 的所有奇异值.*

1. *存在酉矩阵 $U_1 \in \mathbf{U}_m(\mathbb{C})$ 和 $U_2 \in \mathbf{U}_n(\mathbb{C})$ 使得*

$$U_1 A U_2 = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}.$$

2. 若 $A \in \mathbf{M}_{m\times n}(\mathbb{R})$, 则存在实正交矩阵 $Q_1 \in \mathbf{O}_m(\mathbb{R})$ 和 $Q_2 \in \mathbf{O}_n(\mathbb{R})$ 使得

$$Q_1AQ_2 = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}.$$

证明. (1) 可以将 A 视为线性映射 $\mathscr{A}: \mathbb{C}^n \to \mathbb{C}^m;\ x \mapsto Ax$ 在 $\mathbb{C}^n$ 和 $\mathbb{C}^m$ 的标准基下的矩阵. 则结合定理 8.3.7 和思考题 8.14 即得所需结论. (2) 的证明类似. □

下面是与奇异值分解有关的一个重要结论.

定理 8.3.9 (**极分解** Polar decomposition/factorization): 设 V 为非零的有限维实内积空间或酉空间, $\mathscr{A}$ 是 V 上的线性算子.

则存在半正定算子 S, S_1 和等距同构 (即酉变换或正交变换) Q, Q_1 使得 $\mathscr{A} = QS = S_1Q_1$. 这两个分解式中的 S, S_1 是由 $\mathscr{A}$ 唯一确定的, 事实上, $S = \sqrt{\mathscr{A}^*\mathscr{A}}$, $S_1 = \sqrt{\mathscr{A}\mathscr{A}^*}$.

证明. 我们只证明前一个分解式有关的结论. 另一个分解式的证明留给读者练习.

事实上, 若 $\mathscr{A} = QS$, 其中 Q 是等距同构, S 是半正定算子, 则 $\mathscr{A}^*\mathscr{A} = S^*Q^*QS = S^*S = S^2$. 故 $S = \sqrt{\mathscr{A}^*\mathscr{A}}$. 这就证明了关于唯一性的结论.

为证存在性, 首先注意到奇异值分解可以保证 V 中存在两组规范正交基 $\mathcal{E} = (e_1, \cdots, e_n)$ 和 $\mathcal{F} = (f_1, \cdots, f_n)$ 使得

$$\mathscr{A}e_1 = \sigma_1 f_1\,, \cdots, \mathscr{A}e_r = \sigma_r f_r\,,\ \mathscr{A}e_{r+1} = \cdots = \mathscr{A}e_n = 0$$

其中 $\sigma_1 \geq \cdots \geq \sigma_r > 0$ 是 $\mathscr{A}$ 的所有奇异值. 现在只要按照

$$\begin{cases} Se_1 = \sigma_1 e_1\,, \cdots, Se_r = \sigma_r e_r\,,\ Se_{r+1} = \cdots = Se_n = 0 \\ Qe_1 = f_1\,, \cdots, Qe_n = f_n \end{cases}$$

来定义 S 和 Q 即可, 因为按照上述定义可以轻易验证 S 半正定, Q 是等距, 且 $\mathscr{A} = QS$. □

如果将 $\mathrm{End}(V)$ 和 $\mathbb{C}$ 做一个类比, 线性变换 $\mathscr{A}$ 的伴随 $\mathscr{A}^*$ 类比于复数 z 的共轭 $\bar{z}$, 那么酉变换的定义 $\mathscr{A}\mathscr{A}^* = 1$ 对应于 $|z|^2 = z\bar{z} = 1$, 半正定算子对应于非负实数. 于是线性变换的极分解 $\mathscr{A} = QS = S_1Q_1$ 恰好对应于复数 z 的分解式 $z = rz_0$, 其中 r 是非负实数, $|z_0| = 1$. 后者正是复数的极坐标表示. 这可能是定理 8.3.9 中结论被冠以"极分解"之名称的原因吧.

思考题 8.16. 叙述定理 8.3.9 的矩阵形式, 并用矩阵形式的奇异值分解 (即定理 8.3.8) 证明之.

8.3.3 * Rayleigh 商及 Hermite 矩阵的谱

在许多应用问题中, 矩阵的特征值都起到关键的作用. 作为以往所学知识的复习, 我们在本小节讨论关于矩阵特征值的一些有趣结论.

定义 8.3.10. 设 $A \in \mathbf{M}_n(\mathbb{C})$.

我们将 A 在 $\mathbb{C}$ 中所有不同的特征值构成的集合记为 $\sigma(A)$, 称之为 A 的**谱** (spectrum). 再定义 A 的**谱半径** (spectral radius) $\rho(A) := \max\{|\lambda| : \lambda \in \sigma(A)\}$. 即, 谱半径 $\rho(A)$ 是 A 的特征值的模长最大值. ■

如果 A 是个 Hermite 矩阵, 那么 A 的复特征值都是实数 (推论 8.2.10 或思考题 8.9 (2)), 此时 $\sigma(A)$ 中的元素可以按大小顺序排列. 我们现在来介绍一种决定 Hermite 矩阵特征值的重要方法.

定义 8.3.11. 设 $A \in \mathbf{M}_n(\mathbb{C})$ 为 Hermite 矩阵. 对于任意非零列向量 $x \in \mathbb{C}^n$, 定义

$$\mathcal{R}_A(x) := \frac{\overline{x}^T A x}{\overline{x}^T x},$$

称为 A 在 x 处的 Rayleigh **商** (Rayleigh[¶] quotient) 或者 Rayleigh–Ritz **商** (Rayleigh–Ritz[‖] quotient).

因为 A 是 Hermite 矩阵, 所以 $\mathcal{R}_A(x)$ 总是实数. 我们可以定义函数

$$\mathcal{R}_A \ : \ \mathbb{C}^n \setminus \{0\} \longrightarrow \mathbb{R}\,; \quad x \longmapsto \mathcal{R}_A(x)\,.$$

容易看出

$$\text{对于任意非零常数 } c,\ \mathcal{R}_A(cx) = \mathcal{R}_A(x)\,.$$

所以, 如果要考虑 Rayleigh 商 $\mathcal{R}_A$ 的取值范围, 我们可以只考虑满足 $\|x\| = 1$ 的向量 x. (这里我们按照 $\mathbb{C}^n$ 的标准 Hermite 内积来计算向量 x 的长度.) ■

命题 8.3.12 (Rayleigh–Ritz **定理** Rayleigh–Ritz theorem): *设 $A \in \mathbf{M}_n(\mathbb{C})$ 为 Hermite 矩阵, $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ 为 A 的所有特征值 (重数计入). 考虑由 Rayleigh 商定义的函数 $\mathcal{R}_A : \mathbb{C}^n \setminus \{0\} \longrightarrow \mathbb{R}$.*

则

1. *λ_1 是 $\mathcal{R}_A$ 的最小值, 当 v_1 是属于 λ_1 的特征向量时, $\mathcal{R}_A(v_1) = \lambda_1$;*
2. *λ_n 是 $\mathcal{R}_A$ 的最大值, 当 v_n 是属于 λ_n 的特征向量时, $\mathcal{R}_A(v_n) = \lambda_n$.*

证明. 取酉矩阵 $U \in \mathbf{U}_n(\mathbb{C})$ 使得 $UA\overline{U}^T = A' := \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$. 容易看出 $x \mapsto y := Ux$ 是 $\mathbb{C}^n \setminus \{0\}$ 到自身的一个双射. 而

$$\mathcal{R}_A(x) = \mathcal{R}_{A'}(y) = \frac{\sum_{i=1}^n \lambda_i |y_i|^2}{\sum_{i=1}^n |y_i|^2}\,.$$

容易看出

$$\lambda_1 \left(\sum_{i=1}^n |y_i|^2 \right) = \sum_{i=1}^n \lambda_1 |y_i|^2 \le \sum_{i=1}^n \lambda_i |y_i|^2 \le \sum_{i=1}^n \lambda_n |y_i|^2 = \lambda_n \left(\sum_{i=1}^n |y_i|^2 \right)$$

所以 $R_A(x) = R_{A'}(y)$ 的值总是在闭区间 $[\lambda_1\,,\,\lambda_n]$ 之内. 只需要再说明 $\mathcal{R}_A$ 可以取到 λ_1, λ_n 这两个值即可. 事实上, 当 v_i 是属于 λ_i 的特征向量时, 直接验证可得 $\mathcal{R}_A(v_i) = \lambda_i$. □

注记 8.3.13. 设 $A \in \mathbf{M}_n(\mathbb{R})$ 为实对称矩阵. 我们可以类似地考虑函数

$$\mathcal{R}'_A \ : \ \mathbb{R}^n \longrightarrow \mathbb{R}\,; \quad x \longmapsto \frac{x^T A x}{x^T x}\,.$$

和命题 8.3.12 的证明同理, 读者不难证明: $\mathcal{R}'_A$ 的最大值和最小值也分别是 A 的最大特征值和最小特征值, 而且在 x 是相应的特征向量时可以取到最值.

事实上, 在现在的情况下, 读者还可以利用多元微积分中关于函数极值的基本定理证明: 能使 $\mathcal{R}'_A$ 取到最大值 λ_n 的向量 x 只能是属于 λ_n 的特征向量 (最小值的情况也类似). ■

[¶]Lord Rayleigh (1842–1919), 英国物理学家、数学家. 本名 John William Strutt, 因为世袭成为第三代 Rayleigh 男爵 (Baron). 按照英国的习惯, 他的名字随后经常以 Rayleigh, John William Strutt, Baron 或 Rayleigh, John William Strutt, Lord 等等形式出现. 最后 Lord Rayleigh 成为人们对他的一种习惯称呼. 因为和化学家 William Ramsay 一起发现了氩等惰性气体元素, Rayleigh 获得了 1904 年的诺贝尔物理学奖 (Ramsay 则获得当年的诺贝尔化学奖).

[‖]Walther Heinrich Wilhelm Ritz (1878–1909), 瑞士物理学家、数学家.

思考题 8.17. 设 $A \in \mathbf{M}_n(\mathbb{C})$ 为 Hermite 矩阵, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ 为 A 的所有特征值 (重数计入). 假设 $(v_1, \cdots, v_n)$ 为 $\mathbb{C}^n$ 的一组规范正交基, 且对于每个 i, v_i 是属于 λ_i 的特征向量. 令

$$V_i = \operatorname{span}(v_1, \cdots, v_i), \quad U_i = \operatorname{span}(v_i, \cdots, v_n), \quad i \in [\![1, n]\!].$$

证明: 对于每个 $i \in [\![1, n]\!]$,

$$\lambda_i = \max_{0 \neq x \in V_i} \mathcal{R}_A(x) = \min_{0 \neq x \in U_i} \mathcal{R}_A(x).$$

即, λ_i 是 Rayleigh 商 $\mathcal{R}_A$ 在子空间 V_i 中的非零向量处能取到的最大值, 也是 $\mathcal{R}_A$ 在 U_i 中的非零向量处能取到的最小值.

下面介绍一则重要的定理, 它可以为我们提供一种通过 Rayleigh 商来求出 Hermite 矩阵所有特征值的方法.

定理 8.3.14 (**极小化极大值原理** Minimax principle、Courant–Fischer **定理** Courant**–Fischer†† theorem): 设 $A \in \mathbf{M}_n(\mathbb{C})$ 为 Hermite 矩阵, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ 为 A 的所有特征值 (重数计入). 对每个 $i \in [\![1, n]\!]$, 以 $\mathbb{C}^n$ 的每个 i 维子空间作为元素, 将这些元素构成的集合记为 $\mathscr{V}_i$.

则对于每个 $i \in [\![1, n]\!]$,

$$\lambda_i = \min_{V \in \mathscr{V}_i} \max_{0 \neq x \in V} \mathcal{R}_A(x).$$

证明. 首先, 根据思考题 8.17, 存在 $V \in \mathscr{V}_i$ 使得 $\lambda_i = \max_{0 \neq x \in V} \mathcal{R}_A(x)$. 因此,

$$\lambda_i \geq \min_{V \in \mathscr{V}_i} \max_{0 \neq x \in V} \mathcal{R}_A(x).$$

以下只需再证明: 对于 $\mathbb{C}^n$ 中的任意 i 维子空间 V, 总有 $\lambda_i \leq \max_{0 \neq x \in V} \mathcal{R}_A(x)$.

取酉矩阵 $U \in \mathbf{U}_n(\mathbb{C})$ 使得 $UA\overline{U}^T = A' := \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$. 令

$$V' := \{Ux \mid x \in V\}.$$

则 V' 也是 $\mathbb{C}^n$ 中的一个 i 维子空间, 而且

$$\max_{0 \neq x \in V} \mathcal{R}_A(x) = \max_{0 \neq x \in V} \mathcal{R}_{A'}(Ux) = \max_{0 \neq y \in V'} \mathcal{R}_{A'}(y).$$

于是, 问题转化为: 在子空间 V' 中找到非零向量 y, 使得 $\mathcal{R}_{A'}(y) \geq \lambda_i$. 根据 $\mathcal{R}_{A'}(y)$ 的计算公式

$$\mathcal{R}_{A'}(y) = \frac{\sum_{i=1}^n \lambda_i |y_i|^2}{\sum_{i=1}^n |y_i|^2}$$

以及特征值的排序 $\lambda_1 \leq \cdots \leq \lambda_n$, 当 y 来自如下子空间 W 时, $\mathcal{R}_{A'}(y) \geq \lambda_i$. 这里

$$W := \left\{(a_1, \cdots, a_{i-1}, a_i, \cdots, a_n)^T \in \mathbb{C}^n \mid a_1 = \cdots = a_{i-1} = 0\right\}.$$

所以, 最后只要说明, 在 $V' \cap W$ 中一定能够找到非零向量 y 即可. 因为 $\dim V' = \dim V = i$, $\dim W = n - i + 1$, 由维数公式 (参见本书上册定理 2.3.30)

$$\dim(V' \cap W) + \dim(V' + W) = \dim V' + \dim W$$

以及显然的不等式 $\dim(V' + W) \leq \dim \mathbb{C}^n = n$ 可得 $\dim(V' \cap W) \geq 1$. 所以, 所需的向量 y 一定存在. 定理至此证毕. □

**Richard Courant (1888–1972), 德裔美籍数学家.
††Ernst Sigismund Fischer (1875–1954), 奥地利数学家.

下面结论的证明和定理 8.3.14 的证明类似.

定理 8.3.15 (**极大化极小值原理** Maxmin principle): 设 $A \in \mathbf{M}_n(\mathbb{C})$ 为 Hermite 矩阵, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ 为 A 的所有特征值 (重数计入). 对每个 $i \in [\![1, n]\!]$, 以 $\mathbb{C}^n$ 的每个 i 维子空间作为元素, 将这些元素构成的集合记为 $\mathscr{V}_i$.

则对于每个 $i \in [\![1, n]\!]$,

$$\lambda_{n+1-i} = \max_{V \in \mathscr{V}_i} \min_{0 \neq x \in V} \mathcal{R}_A(x) .$$

证明. 留给读者练习. □

8.3.4 习题

以下 V 总是表示非零的有限维实内积空间或酉空间.

习题 8.3.1. 设 U 是 V 的子空间. 证明正交投影 $P_U \in \mathrm{End}(V)$ 是半正定算子.

习题 8.3.2. 设 V 是有限维酉空间, $\mathscr{A} \in \mathrm{End}(V)$ 是正规变换. 证明 $\mathscr{A}$ 必有平方根, 即, 存在 $\mathscr{B} \in \mathrm{End}(V)$ 使得 $\mathscr{A} = \mathscr{B}^2$.

习题 8.3.3. 假设 $\dim V > 1$. 证明恒等算子 $I \in \mathrm{End}(V)$ 有无穷多个平方根.

习题 8.3.4. 设 W 是有限维复向量空间, $\mathscr{A} \in \mathrm{End}(W)$ 是可逆线性变换. 证明 $\mathscr{A}$ 必有平方根. (提示: 利用 Jordan 标准形或 Jordan–Chevalley 分解.)

习题 8.3.5. 证明或举出反例: 设 V 是有限维酉空间或实内积空间, $\mathscr{A} \in \mathrm{End}(V)$ 是自伴算子, $(e_1, \cdots, e_n)$ 是 V 的一组规范正交基. 如果对每个 $i \in [\![1, n]\!]$ 均有 $\langle \mathscr{A} e_i, e_i \rangle$ 为非负实数, 那么 $\mathscr{A}$ 是半正定算子.

习题 8.3.6. 取定非零向量 $u, x \in V$. 定义线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 为

$$\mathscr{A} v = \langle u, v \rangle x .$$

求 $\sqrt{\mathscr{A}^* \mathscr{A}}$ 的表达式.

习题 8.3.7. 定义 $\mathscr{A} \in \mathrm{End}(\mathbb{R}^3)$ 为

$$(x, y, z)^T \longmapsto (z, 2x, 3y)^T .$$

找出一个等距同构 $Q \in \mathrm{End}(\mathbb{R}^3)$ 使得 $\mathscr{A} = Q\sqrt{\mathscr{A}^* \mathscr{A}}$.

习题 8.3.8. 设 $\mathscr{A} \in \mathrm{End}(V)$.

1. 证明: 若 $\mathscr{A}$ 是可逆变换, 则在极分解式 $\mathscr{A} = QS = S_1 Q_1$ 中, 等距同构 Q 和 Q_1 也是由 $\mathscr{A}$ 唯一确定的.
2. 举例说明: 若 $\mathscr{A}$ 不可逆, 则极分解 $\mathscr{A} = QS$ 中的等距同构 Q 可能不是唯一的 (S 为半正定算子).

习题 8.3.9. 设 $\mathscr{A}, \mathscr{B} \in \mathrm{End}(V)$ 为正定算子, $\mathscr{U} \in \mathrm{End}(V)$ 是酉变换. 证明: 如果 $\mathscr{A} = \mathscr{B}\mathscr{U}$, 则必然 $\mathscr{A} = \mathscr{B}$, $\mathscr{U} = I$.

习题 8.3.10. 设 $A \in \mathbf{M}_n(\mathbb{C})$ 是正定的 Hermite 矩阵, $B \in \mathbf{M}_n(\mathbb{C})$ 是半正定的 Hermite 矩阵.

证明: AB 相似于一个对角阵 $\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}$, 其中 d_i 都是非负实数.

(提示: 将 A, B 写成 $A = A_1^2$, $B = B_1^2$, 其中 A_1, B_2 是 Hermite 阵. 令 $C = A_1 B_1$. 考虑 $C\overline{C}^T$ 和 AB 的关系.)

习题 8.3.11. 求矩阵 $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ 的所有奇异值, 并找出正交矩阵 Q_1, Q_2 使 $Q_1 A Q_2$ 称为正交相抵标准形.

习题 8.3.12. 找出一个 $\mathscr{A} \in \mathrm{End}(\mathbb{C}^2)$ 使得 0 是 $\mathscr{A}$ 的唯一特征值, 而 5 是 $\mathscr{A}$ 的唯一奇异值.

习题 8.3.13. 设 σ 是线性变换 $\mathscr{A} \in \mathrm{End}(V)$ 的一个奇异值. 证明: 存在单位向量 $v \in V$ 使得 $\|\mathscr{A} v\| = \sigma$.

习题 8.3.14. 设 $\mathscr{A} \in \mathrm{End}(V)$ 为自伴算子, σ 为正实数. 证明: σ 是 $\mathscr{A}$ 的奇异值当且仅当存在 $\mathscr{A}$ 的特征值 λ 使得 $\sigma = |\lambda|$.

习题 8.3.15. 设 $\mathscr{A} \in \mathrm{End}(V)$. 证明或举出反例: $\mathscr{A}^2$ 的奇异值一定是 $\mathscr{A}$ 的某个奇异值的平方.

习题 8.3.16. 设 $\mathscr{A} \in \mathrm{End}(V)$. 证明: $\mathscr{A}$ 是等距同构当且仅当 $\mathscr{A}$ 可逆并且 $\mathscr{A}$ 的所有奇异值等于 1.

习题 8.3.17. 设 $\mathscr{A} \in \mathrm{End}(V)$ 是可逆的线性变换. 设 σ_r 是 $\mathscr{A}$ 的最小奇异值, σ_1 是 $\mathscr{A}$ 的最大奇异值. 证明:

1. 对任意 $v \in V$, $\sigma_r \|v\| \le \|\mathscr{A} v\| \le \sigma_1 \|v\|$.
2. 对于 $\mathscr{A}$ 的任意特征值 λ, 必有 $\sigma_r \le |\lambda| \le \sigma_1$.

第九章　多重线性代数初步

本章为选讲内容, 以后适时补充.
参考: [26, 第九章], [22, 第二、三章].

附录 A 测验与考试试题选辑

A.1 小测验 I

测验日期: 2021 年 3 月 5 日
时间: 50 分钟; 满分: 50 分

以下总设 K 是 $\mathbb{C}$ 的子域, $n \in \mathbb{N}^*$, V 是 K 上的 n 维向量空间.

1. (12 分) 判断正误, 正确的请解释理由, 错误的请举出反例.

   (a) 设 U, W 是 V 的子空间且 $V = U \oplus W$. 则对于任意 $x \in V$, $x \notin U$ 时必有 $x \in W$.

   (b) 设 U, W 是 V 的子空间 $V = U \oplus W$. 设 $\mathscr{A} \in \mathrm{End}(V)$. 若 U 是 $\mathscr{A}$ 的不变子空间, 则 W 也是.

   (c) 设 $\mathscr{A} \in \mathrm{End}(V)$. 如果 u, v 都是 $\mathscr{A}$ 的特征向量且 $u + v \neq 0$, 则 $u + v$ 也是 $\mathscr{A}$ 的特征向量.

   (d) 设 $\mathscr{A} \in \mathrm{End}(V)$ 是可逆变换, W 是 $\mathscr{A}$ 的不变子空间, 则 W 也是 $\mathscr{A}^{-1}$ 的不变子空间.

2. (15 分) 设 $V = K[X]_{\leq 3}$, 定义线性变换 $\mathscr{A} \in \mathrm{End}(V)$:

$$\mathscr{A} \,:\, f(X) \longmapsto (X-1)f'(X)\,.$$

   (a) 令 $\mathcal{B} = (1,\, X,\, X^2,\, X^3)$. 写出 $\mathscr{A}$ 在有序基 $\mathcal{B}$ 下的矩阵.

   (b) 求 $\mathscr{A}$ (在 K 中)的所有特征值及相应的特征子空间.

3. (15 分) 设 U 为 $\mathbf{M}_n(K)$ 中所有的上三角阵构成的子空间, A 为 $\mathbf{M}_n(K)$ 中所有反对称阵构成的子空间. (说矩阵 $M \in \mathbf{M}_n(K)$ 反对称是指 $M + M^T = 0$.)

   (a) 求 $\dim U$ 和 $\dim A$.

   (b) 直和分解式 $\mathbf{M}_n(K) = U \oplus A$ 是否成立？为什么？

4. (8 分) 设 $\mathscr{A} \in \mathrm{End}(V)$ 满足 $\mathscr{A}^2 = \mathscr{A}$. 证明: $V = \mathrm{Ker}(I - \mathscr{A}) \oplus \mathrm{Ker}(\mathscr{A})$.

A.2 小测验 II

测验日期: 2021 年 4 月 2 日
时间: 50 分钟; 满分: 50 分

以下总设 K 是 $\mathbb{C}$ 的子域, $n \in \mathbb{N}^*$, V 是 K 上的 n 维向量空间.

1. (9 分) 判断正误. (不需要解释理由.)

(a) 设 $A \in \mathbf{M}_6(K)$ 满足 $A^5 = A$. 则 A 在 K 上一定可以对角化.

(b) 假设复方阵 A 可对角化, λ 是 A 的一个特征值. 则 λ 的代数重数与几何重数相等.

(c) 设 $A, B \in \mathbf{M}_2(\mathbb{C})$ 具有相同的特征多项式和最小多项式, 则 A, B 在 $\mathbb{C}$ 上相似.

2. (3+4+3=10 分) 填空. (不需要解释理由.)

(a) 假设 $A \in \mathbf{M}_5(K)$ 是幂零阶为 2 的幂零矩阵. 则 $\operatorname{rank}(A)$ 的可能取值有 ____

(b) 矩阵 $-I_n$ 的特征多项式是 ____, 最小多项式是 ____

(c) 设 $\dim V = 5$, $\mathscr{N} \in \operatorname{End}(V)$ 是循环的幂零变换. 则 0 作为 $\mathscr{N}$ 的特征值几何重数是 ____

3. (16 分) 设 $A \in \mathbf{M}_n(K)$ 为幂零矩阵, $P \in \mathbf{M}_n(K)$ 为可逆矩阵.

(a) $P^{-1}AP$ 是否一定是幂零矩阵？若是, 请解释理由. 若否, 请给出反例.

(b) AP 是否一定是幂零矩阵？若是, 请解释理由. 若否, 请给出反例.

(c) 证明 $A + I_n$ 是可逆矩阵.

(d) $A + P$ 是否一定是可逆矩阵？若是, 请解释理由. 若否, 请给出反例.

4. (10 分) 将矩阵 $A = \begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix}$ 化为 Jordan 标准形.

5. (5 分) 设 $A \in \mathbf{M}_n(K)$ 满足 $A^2 = A$. 证明 $\operatorname{rank}(A) = \operatorname{Tr}(A)$.

A.3 小测验 III

测验日期: 2021 年 5 月 7 日
时间: 50 分钟; 满分: 50 分

以下总设 $n \in \mathbb{N}^*$, K 是 $\mathbb{C}$ 的子域.

1. (16 分) 简答题. (直接写出答案, 不需解释理由.)

(a) 写出一个可逆矩阵 $P \in \mathrm{M}_4(\mathbb{R})$ 使得 $P^T \begin{pmatrix} 1 & 1 & & \\ & 6 & & \\ & & 8 & \\ & & & 9 \end{pmatrix} P = \begin{pmatrix} 9 & & & \\ & 8 & & \\ & & 1 & 1 \\ & & & 6 \end{pmatrix}$.

(b) 考虑 K^3 上的三元二次型

$$q(x) = x_1^2 + 2x_1x_2 - 2x_2^2 + 4x_2x_3 - 4x_3^2, \quad \text{其中 } x = (x_1, x_2, x_3)^T.$$

设 $b = b_q$ 是 q 的极化型. 通过 $x = (x_1, x_2, x_3)^T$ 和 $y = (y_1, y_2, y_3)^T$ 的坐标写出双线性型 $b(x, y)$ 的明显表达式.

(c) 写出两个可逆的 3 阶实矩阵 A, B 使得 A, B 和 I_3 三个矩阵在 $\mathbb{R}$ 上两两之间互不相合.

(d) 假设实数 t 能使矩阵 $\begin{pmatrix} t & 3 & 0 \\ 3 & t & 4 \\ 0 & 4 & t \end{pmatrix}$ 半正定. 求 t 的取值范围.

2. (15 分) 判断正误. (不需解释理由.)

   (a) 设 $A \in \mathbf{M}_n(K)$. 则映射
   $$q : K^n \longrightarrow K ; \quad x \longmapsto x^T A x$$
   是 K^n 上的二次型当且仅当 A 是对称矩阵.

   (b) 设 $A \in \mathbf{M}_n(\mathbb{R})$. 则 A 是负定矩阵的充分必要条件是 A 的所有顺序主子式小于 0.

   (c) 若 $A, B \in \mathbf{M}_n(\mathbb{R})$ 都是可逆矩阵且 A, B 相合, 则 A^{-1}, B^{-1} 也相合.

   (d) 设 $A \in \mathbf{M}_n(\mathbb{R})$ 是对称阵. 如果 $\det(A) < 0$, 一定存在列向量 $\alpha \in \mathbb{R}^n$ 使得 $\alpha^T A\alpha < 0$.

   (e) 设 $f(x, y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2b_1x + 2b_2y + c$ 为实系数二元二次多项式, $X = V(f)$ 是 f 在仿射平面 $\mathbb{A}^2$ 中的零点集. 如果 X 是一条抛物线, 则矩阵 $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$ 的秩为 2.

3. (6 分) 找出一种可逆的线性变量替换将下面的三元二次型 q 化为标准形:
   $$q(x, y, z) = x^2 - 7y^2 + 3z^2 - 6xy + 4xz - 4yz\,.$$

4. (8 分) 考虑列向量空间 K^2 上的双线性型
   $$\varphi : K^2 \times K^2 \longrightarrow K ; \quad (\alpha\,, \beta) \longmapsto \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \text{ 其中 } \alpha = \begin{pmatrix} a \\ c \end{pmatrix}, \beta = \begin{pmatrix} b \\ d \end{pmatrix}.$$

   (a) 求 φ 在 K^2 的标准基下的 Gram 矩阵.

   (b) 定义线性函数 $f : K^2 \to K : (x, y)^T \mapsto 2x - y$. 是否存在 $v \in K^2$ 使得对所有 $\alpha \in K^2$ 均有 $f(\alpha) = \varphi(v\,, \alpha)$? 若是, 请找出一个这样的 v. 若否, 请解释为何这样的 v 不存在.

5. (5 分) 设 V 是有限维 K-向量空间, φ 是 V 上一个非退化的对称双线性型. 假设 φ 是迷向的, 即, 存在非零向量 $v \in V$ 使得 $\varphi(v, v) = 0$.

   证明: 对应任意 $\lambda \in K$, 均存在 $u \in V$ 使得 $\varphi(u, u) = \lambda$.

   (提示: 考虑 φ 在适当的 2 维子空间上的限制.)

A.4 小测验 IV

测验日期: 2021 年 5 月 28 日
时间: 50 分钟; 满分: 50 分

以下总设 $n \in \mathbb{N}^*$, K 是 $\mathbb{C}$ 的子域.

1. (16 分) 简答题. (直接写出答案, 不需解释理由.)

   (a) 写出矩阵 $A = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}$ 的 QR 分解.

   (b) 设 u, v 是实内积空间 V 中互相正交的向量. 假设 $\|u\| = 1$, $\|v\| = 4$. 求 $\|2u - v\|$.

   (c) 设 $A \in \mathbf{M}_3(\mathbb{R})$ 是行列式为 -1 的正交矩阵. 请写出 A 的正交相似标准形所有可能的形式. (如有二阶的对角块, 请将其置于左上角, 并写成 $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, $b > 0$ 的形式.)

(d) 写出一个 $\mathbb{R}^2$ 到自身的等距(线性)变换 $\mathscr{A}$ 满足 $\mathscr{A}e_2 = \left(-\frac{4}{5}, \frac{3}{5}\right)^T$ (其中 $e_2 = (0, 1)^T$).

2. (12 分) 判断正误. (不需解释理由.)

   (a) 若 $A, B \in \mathbf{M}_n(\mathbb{R})$ 都是对称矩阵, 则 A 和 B 正交相似的充分必要条件是它们的特征多项式相同.

   (b) 设 $\mathcal{E} = (e_1, \cdots, e_n)$ 为 V 的一组规范正交基, $\mathscr{A} \in \mathrm{End}(V)$, $A = (a_{ij})$ 为 $\mathscr{A}$ 在有序基 $\mathcal{E}$ 下的矩阵. 则 $a_{ij} = \langle \mathscr{A}^* e_i, e_j \rangle$.

   (c) 假设如下方程定义的二次曲线是椭圆

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2b_1x + 2b_2y + c = 0,$$

   其中 $a_{11}, a_{12}, a_{22}, b_1, b_2, c \in \mathbb{R}$. 则二次曲面

$$z = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2b_1x + 2b_2y + c$$

   是椭圆柱面.

   (d) 若 A 是实正规矩阵, 并且在 $\mathbb{R}$ 上可以对角化, 那么 A 一定是对称矩阵.

3. (7 分) 设 V 是有限维实内积空间, $\mathscr{A} \in \mathrm{End}(V)$. 则 $\mathscr{A}$ 是斜对称变换的充分必要条件是对任意 $v \in V$ 均有 $\langle \mathscr{A}v, v \rangle = 0$.

4. (15 分) 设 $V = \mathbf{M}_n(\mathbb{R})$.

   (a) 证明: 映射

$$\mathbf{M}_n(\mathbb{R}) \times \mathbf{M}_n(\mathbb{R}) \longrightarrow \mathbb{R}; \quad (A, B) \longmapsto \mathrm{tr}(A^T B)$$

   是 V 上的内积.

   (b) 证明: 对于任意 $A, B \in \mathbf{M}_n(\mathbb{R})$,

$$\mathrm{tr}((A^T + B^T)(A + B)) \leq 2\left(\mathrm{tr}(A^T A) + \mathrm{tr}(B^T B)\right).$$

   再给出等号成立的充分必要条件.

   (c) 假设 $n \geq 2$. 证明不等式:

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{ij}{i+j}\right)^2 < \left(\sum_{i=1}^{n} i^2\right)^2 \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{(i+j)^2}\right)$$

参考文献

[1] Tom M. Apostol, *Linear Algebra: A First Course with Applications to Differential Equations*, John Wiley & Sons, Inc., 1997. 中文翻译版书名《线性代数及其应用导论》, 沈灏、沈佳辰 译, 人民邮电出版社, 2010.

[2] Michael Artin, *Algebra, (Second edition)*, Pearson, 2010. 国内有机械工业出版社的影印版.

[3] Sheldon Axler, *Linear Algebra Done Right, (Third edition)*, Springer, 2015. 中文翻译版书名《线性代数应该这样学》, 杜现昆、刘大艳、马晶 译, 人民邮电出版社, 2016.

[4] Sheldon Axler, *Down with determinants!*, American Mathematical Monthly, **102**, 1995, no. 2, pp. 139–154.

[5] Marcel Berger, *Geometry I*, Translated from the 1977 French original by M. Cole and S. Levy, Springer-Verlag, Berlin, 2009.

[6] Nicolas Bourbaki, *Algèbre, Chapitres 1–3*, Springer–Verlag, 2007.

[7] Alexander Chubarev and Iosif Pinelis, *Fundamental theorem of geometry without the* 1*-to-*1 *assumption*, Proc. Amer. Math. Soc., **127**, 1999, no. 9, pp. 2735–2744.

[8] Gerald Farin and Dianne Hansford, *Practical Linear Algebra*, CRC Press, 2014. 国内有机械工业出版社的翻译版 (董晓波 等译) 和注释影印版 (李红玲 注释).

[9] Roger Godement, *Algebra*, Translated from the French, Hermann, Paris, 1968.

[10] Erich Hecke, *Vorlesungen über die Theorie der algebraischen Zahlen*, Akademische Verlagsgesellschaft, Leipzig, 1923.

[11] Erich Hecke, *Lectures on the Theory of Algebraic Numbers*, Graduate Texts in Mathematics 77, Springer, 1981.

[12] Serge Lang, *Algebra (Revised third edition)*, Graduate Texts in Mathematics 211, Spring, 2002.

[13] Peter D. Lax, *Linear Algebra and Its Applications (Second edition)*, John Wiley & Sons, Inc., 2007.

[14] David C. Lay, Steven R. Lay and Judi J. McDonald, *Linear Algebra and Its Applications (Fifth edition)*, Pearson Eduction, Inc., 2016. 国内有机械工业出版社的影印版.

[15] Steven J. Leon, *Linear Algebra with Applications (Ninth edition)*, Pearson Eduction, Inc., 2015. 国内有机械工业出版社的影印版.

[16] Steven Roman, *Advanced Linear Algebra (Third edition)*, Graduate Texts in Mathematics 135, Springer, 2008.

[17] Kenneth H. Rosen, *Elementary Number Theory and Its Applications (Sixth edition)*, Pearson, 2011. 国内有机械工业出版社的影印版.

[18] Gilbert Strang, *Linear Algebra and Its Applications (Fourth edition)*, Thomson Learning, Inc., 2006.

[19] Terence Tao, *Analysis I, (Third edition)*, Springer, 2016. 中文翻译版书名《陶哲轩实分析 (第 3 版)》, 李馨 译, 人民邮电出版社, 2018.

[20] 陈伏兵, 应用线性代数; 科学出版社, 2011.

[21] 蓝以中, 高等代数简明教程, 上、下册 (第二版); 北京大学出版社, 2007.

[22] 黎景辉, 白正简, 周国晖, 高等线性代数学, 高等教育出版社, 2014.

[23] 李尚志, 线性代数, 高等教育出版社, 2011.

[24] 李尚志, 线性代数 (数学专业用), 高等教育出版社, 2006.

[25] 李文威; 代数学方法 (卷一: 基础架构); 高等教育出版社; 2019. 网络版本可以于如下网址下载: http://www.wwli.url.tw/index.php/zh-CN/

[26] 聂灵沼、丁石孙; 代数学引论 (第二版); 高等教育出版社, 2000.

[27] 丘维声, 简明线性代数; 北京大学出版社, 2002.

[28] 丘维声, 解析几何 (第三版); 北京大学出版社, 2015.

[29] 任广千, 谢聪, 胡翠芳, 线性代数的几何意义; 西安电子科技大学出版社, 2015.

[30] 唐小谦 (笔名), 欢乐品数学: 爱上你眼中的魔鬼学科, 清华大学出版社, 2018.

[31] 王殿军、张贺春、胡冠章, 大学代数与几何, 清华大学出版社, 2012.

[32] 许甫华、张贤科, 高等代数解题方法 (第 2 版), 清华大学出版社, 2005.

[33] 俞正光、鲁自群、林润亮, 线性代数与几何, 上、下册 (第 2 版), 清华大学出版社, 2014.

[34] 张苍 (西汉) 等 辑撰, 曾海龙 译解, 九章算术; 江苏人民出版社, 2011.

[35] 张贤科、许甫华, 高等代数学 (第 2 版), 清华大学出版社, 2004.