线性代数复习笔记

作者：吕浩哲（Lucas Shen）

时间：July 14, 2025

封面：https://www.pixiv.net/artworks/100631860

悟已往之不谏，知来者之可追。——陶渊明

目录

第 1 章 线性空间与线性变换

线性空间是线性代数的起点，线性空间中的元素称为向量（也因此线性空间也被称为向量空间）. 如果从初学者的视角，由线性空间开始不是值得推荐的顺序，很多优秀的教材选择通过更直观的行列式/解线性方程组等开始，而后从矩阵/线性空间切入主题，在此基础上研究有限维线性空间.

但本文的最大目的在于复习，从复习者的角度，由于已经进行过一遍线性代数的学习，因此并不需要以直观起步，也能够接受从线性空间开始对整个过程的回顾.

1.1 线性空间

线性空间定义在某个域 $\mathbb{F}$ 上，我们大多使用的域是 $\mathbb{R}$ 或 $\mathbb{C}$.

定义 1.1 (线性空间/向量空间)

称一个带有加法与数乘的集合 $(V/\mathbb{F},+,\cdot)$ 为域 $\mathbb{F}$ 线性空间（向量空间），要求：

1. $(V,+)$ 构成 Abel 群：
   (a). 单位元：$0 \in V$.
   (b). 逆元：$\forall v \in V, \exists(-v) \in V : v+(-v)=0$.
   (c). 封闭性：$\forall u,v \in V, u+v \in V$.
   (d). 结合律：$\forall u,v,w \in V, (u+v)+w=u+(v+w)$.
   (e). 交换律：$\forall u,v \in V, u+v=v+u$.
2. 数乘定义为 $\mathbb{F}\times V \to V$，满足
   (a). 乘法单位元：$1 \in \mathbb{F}, \forall v \in V, 1v=v$.
   (b). 数乘结合律：$\forall a,b \in \mathbb{F}, v \in V : \lambda(\mu v)=(\lambda\mu)v$.
   (c). 分配律：

$$\forall a,b \in \mathbb{F}, v \in V : (a+b)v = av+bv \tag{1.1}$$

$$\forall a \in \mathbb{F}, u,v \in V : a(u+v) = au+av \tag{1.2}$$

♣

注 当 $\mathbb{F}=\mathbb{R}$ 或 $\mathbb{F}=\mathbb{C}$ 时，对应的向量空间称为**实向量空间**或**复向量空间**.

在数学中，对于某个代数对象，我们总会定义对应的子对象，比如子群，子环，子拓扑等，对于线性空间，也可以定义子空间.

定义 1.2 (子（向量）空间)

设线性空间 $(V/\mathbb{F},+,\cdot)$，若集合 $W \subset V$，且 $(W/\mathbb{F},+,\cdot)$ 构成线性空间，则称 W 为 V 的子空间，记为 $W \leqslant V$.

♣

由于子集 W 会继承原集合 V 中的运算性质，因此从子集到子空间并没必要一一验证定义，我们有如下命题.

命题 1.1

设 $V/\mathbb{F}$ 为子空间，$W \subset V$，则 $W \leqslant V$ 只需要满足：

1. 对于 $0 \in V$，有 $0 \in W$.
2. W 关于两种运算封闭（V 上的加法与数乘）.

♠

1.1.1 子空间的和

对于某个线性空间 V 的子空间，可以作和运算，和运算之于线性空间类似并运算之于集合，但不同的是，除了将元素相并，也要将其“组合”并入其中（如果没有这个要求，我们得到的和空间可能会不满足封闭性）.

定义 1.3 (子空间的和)

设 V 为线性空间，$U_1,\cdots,U_m \leqslant V$，则 $U_1,\cdots,U_m$ 的和定义为

$$U_1+\cdots+U_m=\{u_1+\cdots+u_m : u_i\in U_i, i=1,\cdots,m\}, \tag{1.3}$$

即其中元素所有可能的和构成的集合.

♣

命题 1.2

设 V 为线性空间，$U_1,\cdots,U_m \leqslant V$，则 $U_1+\cdots+U_m$ 是 V 的包含 $U_1,\cdots,U_m$ 的最小子空间.

♠

证明 首先证明 $0\in U_1+\cdots+U_m$ 是 V 的子空间，由于 $0=0+\cdots+0\in U_1+\cdots+U_m$，假设 $v=u_1+\cdots+u_m, w=u_1'+\cdots+u_m'\in U_1+\cdots+U_m$，则对任意 $a,b\in\mathbb{F}$，都有

$$au+bv=a(u_1+\cdots+u_m)+b(u_1'+\cdots+u_m') \tag{1.4}$$

$$=(au_1+bu_1')+\cdots+(au_m+bu_m')\in U_1+\cdots+U_m \tag{1.5}$$

这说明 $U_1+\cdots+U_m\leqslant V$.

为了说明最小性，只要证明对任意 $W\leqslant V$，若 $U_1,\cdots,U_m\subset W$，则 $U_1+\cdots+U_m\subset W$ 即可. 由于 $U_1,\cdots,U_m\subset W$，根据线性空间的封闭性，对任意 $u_1\in U_1,\cdots,u_m\in U_m$，都有

$$u_1+\cdots+u_m\in W \tag{1.6}$$

因此 $U_1+\cdots+U_m\subset W$，得证.

因此，我们也可以将 $U_1,\cdots,U_m\leqslant V$ 的和空间定义为包含 $U_1,\cdots,U_m$ 的最小子空间.

对集合取并时，若 $A\cap B=\emptyset$，则它们的并也可以写作无交并，即 $A\cup B=A\sqcup B$；对于线性空间的和，也在某种“独立”情况下可写作直和，只不过这时不可能无交，因为根据定义任意 V 的子空间都要包含加法单位元 0.

定义 1.4 (直和)

设 V 为线性空间，$U_1,\cdots,U_m\leqslant V$，若 $U_1+\cdots+U_m$ 中的元素都能唯一表示为 $u_1+\cdots+u_m$，其中每个 $u_i\in U_i$，则称这样的和是直和，可表示为

$$U_1\oplus\cdots\oplus U_m. \tag{1.7}$$

♣

根据定义直接验证直和是一件很麻烦的事，不过使用数学中常用的做差思想可以得到直和的等价命题.

命题 1.3

设 V 为线性空间，$U_1,\cdots,U_m\leqslant V$，则 $U_1+\cdots+U_m$ 是直和当且仅当 0 表示为 $u_1+\cdots+u_m, u_i\in U_i$ 的唯一方式是每个 $u_i=0$.

♠

证明 易知，若 $U_1+\cdots+U_m$ 是直和，由于 0 显然有表示 $0=0+\cdots+0$，因此的这种表示方式唯一. 反之，假设 0 的唯一表示方法为 $0=0+\cdots+0$，则假设 $v\in U_1+\cdots+U_m$ 有表示方法 $v=u_1+\cdots+u_m=u_1'+\cdots+u_m'$，则 0 有表示方法

$$0=(u_1-u_1')+\cdots+(u_m-u_m') \tag{1.8}$$

根据假设，每个 $u_i=u_i'$，这说明每个 $v\in U_1+\cdots+U_m$ 的表示方法是唯一的，即和是直和.

对于直和，容易验证如下命题成立.

命题 1.4

设 V 为线性空间，$U, W \leqslant V$，则 $U+W$ 为直和当且仅当 $U \cap W = \{0\}$.

证明 假设 $U+W$ 为直和，则对任意 $v \in U \cap W$，$0 = v + (-v)$，由于这种表示唯一，且 $v = -v = 0$，这说明 $U \cap W = \{0\}$.

反之，若 $U \cap W = \{0\}$，由于 $0 \in U \cap W$，故存在 $u \in U, w \in W$，使得 $0 = u + w$，这表明 $u = -w$，即 $u, w \in U \cap W$，因此 $u = w = 0$，得证.

类似地，我们可以证明更一般的结论.

命题 1.5

设 V 为线性空间，$U_1, \cdots, U_m \leqslant V$，则 $U_1 + \cdots + U_m$ 为直和等价于下面任意一条：

- 对任意 $1 < i \leqslant m$，$(U_1 \cup \cdots \cup U_{i-1}) \cap U_i = \{0\}$.
- $U_1, \cdots, U_m$ 中任意个集合的交均为 $\{0\}$.

1.1.2 线性组合与线性无关

我们在前面讨论了子空间的和与直和，其中涉及了线性空间中元素的某种“组合”，我们来进一步讨论这件事，首先给出线性组合的概念.

定义 1.5 (线性组合)

线性空间 V 中的向量 $v_1, \cdots, v_m$ 的线性组合是指形如

$$a_1 v_1 + \cdots + a_m v_m \tag{1.9}$$

的向量，其中 $a_1, \cdots, a_m \in \mathbb{F}$.s

线性组合实际上是线性空间中两种运算的复合——域 $\mathbb{F}$ 中的数乘与 V 中的加法，并且根据线性空间的封闭性，可知 V 中任意向量线性组合的结果仍然在 V 中[1].

有了线性组合的概念，给定一组向量 $v_1, \cdots, v_m \in V$，我们自然会考虑它们线性组合的全体，事实上，这也是一个子空间，称为 $v_1, \cdots, v_m$ 的张成空间.

定义 1.6 (张成空间)

设 $V/\mathbb{F}$ 为线性空间，则 $v_1, \cdots, v_m \in V$ 的所有线性组合的全体构成的集合称为 $v_1, \cdots, v_m$ 的张成空间，即

$$\mathrm{Span}(v_1, \cdots, v_m) = \{a_1 v_1 + \cdots + a_m v_m : a_1, \cdots, a_m \in \mathbb{F}\} \tag{1.10}$$

张成空间也可表示为 $\mathbb{F}\langle v_1, \cdots, v_m \rangle$，空向量组的张成空间定义为 $\{0\}$.

命题 1.6

线性空间 V 中一组向量 $v_1, \cdots, v_m$ 的张成空间是包含这组向量的最小子空间.

证明 首先容易验证 $\mathrm{Span}(v_1, \cdots, v_m) \leqslant V$. 对于任意 $W \leqslant V$，$v_1, \cdots, v_m \in W$，根据线性空间的封闭性，$v_1, \cdots, v_m$ 的线性组合仍然在 W 中，因此

$$\mathrm{Span}(v_1, \cdots, v_m) \subset W, \tag{1.11}$$

得证.

容易发现，向量越多，得到的张成空间也会越大，即若 $n \leqslant m$，则

$$\mathrm{Span}(v_1, \cdots, v_n) \leqslant \mathrm{Span}(v_1, \cdots, v_m). \tag{1.12}$$

[1] 虽然无限的线性组合很容易形式地定义出来，但为了防止过多讨论，这里所讨论的都是有限线性组合.

在子空间的直和中，我们讨论了表示的“唯一性”，对于某些向量组的张成空间也有类似的性质，我们称为线性无关性，即 $\mathrm{Span}(v_1,\cdots,v_m)$ 中的元素都可以唯一表示为 $v_1,\cdots,v_m$ 的线性组合，仿照我们在前面进行过的操作（做差），可以得到如下定义

定义 1.7 (线性无关)

设 $v_1,\cdots,v_m\in V/\mathbb{F}$，若

$$a_1v_1+\cdots+a_mv_m=0\Longleftrightarrow a_1=\cdots=a_m=0, \tag{1.13}$$

则称 $v_1,\cdots,v_m$ 线性无关，反之称其线性相关.

反过来，线性相关的向量组有一种“浪费”的感觉，假设 $v_1,\cdots,v_m$ 线性相关，则根据定义，存在不全为 0 的 $a_1,\cdots,a_m$ 使得 $a_1v_1+\cdots+a_mv_m=0$，假设 $j=\max\{1\leqslant k\leqslant m:a_k\neq 0\}$，移项可得

$$v_j=-\frac{a_1}{a_j}v_1-\cdots-\frac{a_{j-1}}{a_j}v_{j-1} \tag{1.14}$$

这表明 $v_j\in\mathrm{Span}(v_1,\cdots,\widehat{v_j},\cdots,v_m)$，其中 $\widehat{v_j}$ 表示去掉 v_j，也就是说对于任意线性组合 $b_1v_1+\cdots+b_1v_m$，我们可以将其中的 v_j 用上面的式子代换，得到新的不含 v_j 的表示，即

$$\mathrm{Span}(v_1,\cdots,v_j,\cdots,v_m)=\mathrm{Span}(v_1,\cdots,\widehat{v_j},\cdots,v_m) \tag{1.15}$$

这说明向量组 $v_1,\cdots,v_m$ 中 v_j 是“多余”的，将 v_j 去掉所得到的张成空间与原有空间相同，这可以概括为（更好用的）如下定理.

定理 1.1 (线性相关性引理)

设 $v_1,\cdots,v_m$ 为 V 中一个线性相关的向量组，则存在 $j\in\{1,\cdots,m\}$，使得

1. $v_j\in\mathrm{Span}(v_1,\cdots,v_{j-1})$.
2. $\mathrm{Span}(v_1,\cdots,v_j,\cdots,v_m)=\mathrm{Span}(v_1,\cdots,\widehat{v_j},\cdots,v_m)$.

对于一个线性相关的向量组 $v_1,\cdots,v_m$ 张成的空间，重复上述过程，删去“多余”的向量，我们总能从中找到线性无关的子向量组（不妨记为 $v_1,\cdots,v_n$，$1\leqslant n\leqslant m$），使得

$$\mathrm{Span}(v_1,\cdots,v_n)=\mathrm{Span}(v_1,\cdots,v_m). \tag{1.16}$$

这其中蕴含了很深刻的思想，上面的讨论告诉我们，任意向量组的张成空间都可以被一个线性无关组张成，也就是说，**只需要有限的向量就可以表示一个向量组的张成空间**，并且根据前面讨论过的线性无关表示的唯一性，在确定这样无关张成组的前提下，张成空间中的每个向量的表示也是唯一的！对于这样的无关张成组，我们可以将其数量定义为秩.

定义 1.8 (向量组的秩)

对于 V 中的一个向量组 $v_1,\cdots,v_n$，若从中可以找到 m 个线性无关的向量（不妨设为 $v_1,\cdots,v_m$），使得

$$\mathrm{Span}(v_1,\cdots,v_m)=\mathrm{Span}(v_1,\cdots,v_n) \tag{1.17}$$

秩是一个很有趣的概念，它描述了向量组中真正“有效”的向量的个数，不过上面的定义也有一些问题没有说明白，比如这样定义的 m 是唯一的吗？是否存在不同的 m_1,m_2 都满足条件？既然秩体现的是有效的向量个数，那么我们自然会猜想上面的定义是合理的，下面我们简单说明这一点.

注：这里如果将后面关于维数唯一性的讨论整过来会更自然些，日后再改.

由于张成空间也是线性空间，因此我们很自然会思考：上面讨论的张成、相关、秩等概念能否被推广到一般的线性空间中？问题的关键在于，线性空间 V 是否能被有限张成，即是否存在 $v_1,\cdots,v_m\in V$，使得 $V=\mathrm{Span}(v_1,\cdots,v_m)$，若存在这样的向量组，那么上面的讨论都可以移植过来，问题也就得到了解决.

不巧的是，很多空间都不满足这样的性质，比如考虑 $\mathbb{R}$ 上的多项式集合 $\mathbb{R}[x]$，则 $\mathbb{R}[x]$ 在多项式的加法与实数的数乘下构成线性空间，但必定不存在这样的有限张成组，因为对于有限个多项式 $f_1,\cdots,f_m$，我们总能找到比它们次数都高的多项式，这显然不能被其线性表示.

不过也没必要泄气，在线性代数中，我们的主要研究对象是可以被有限张成的对象，称之为**有限维线性空间**，下面正式给出定义.

定义 1.9 (有限维向量空间)

如果一个向量空间 V 可以由其中的某个向量组张成，则称 V 是有限维向量空间，这组向量称为 V 的一个张成组，反之为无限维.

♣

上面的例子 $\mathbb{R}[x]/\mathbb{R}$ 就是一个无限维向量空间，而我们所熟悉的 $\mathbb{R}^2,\mathbb{R}^3$ 都是有限维向量空间，根据定义，张成空间也是有限维向量空间.

1.1.3 基与维数

线性代数的主要研究对象是有限维向量空间，但其定义终究只是一种定性的描述，为了更精确描述有限维向量空间，我们定义其中最重要的属性，即其名称中的“维数”，根据维数我们可以选择一定数量的向量来描述整个空间，这样选择的向量被称为基，通过前面对张成空间的讨论，有限维向量空间的基与维数的概念已经呼之欲出了.

定义 1.10 (基与维数)

设 $V/\mathbb{F}$ 为有限维向量空间，则存在线性无关的向量组 $v_1,\cdots,v_n$，使得

$$V=\operatorname{Span}(v_1,\cdots,v_n), \tag{1.18}$$

这样的张成组称为 V 的一组基，其中向量的个数称为 V 的维数，记为 $n=\dim V$.

♣

上面的定义中还存在一些问题，比如选择不同的基，维数是否相同？这在数学中称为良定性（well-defined），可以理解为某种“无矛盾性”，我们下面来讨论这件事，其核心在于讨论线性无关组与张成组之间的联系.

我们在前面讨论过，线性相关代表了某种“多余”，那么作为其反面的线性无关性也会代表某种“不足”；线性相关组中可以删去的多余向量进行缩减，且缩减到一定程度，它就能跨过边界变成无关组，那么对于无关组，也应该可以加入新的向量进行扩张，在加入扩张到一定程度，它也能跨过边界变成相关组.

将这种讨论应用到无关组与张成组上：若线性空间 V 的一个张成组恰好线性无关，则根据定义它是 V 的一组基；若不然，则根据线性相关性引理，可以将其约化到一组基；反之，V 的一个线性无关组恰好能张成 V，则根据定义它是一组基，若不然，则也应该能扩张，直到成为一个张成组.

出于这种思考，我们可以证明如下命题.

命题 1.7

在有限维向量空间中，线性无关向量组的长度小于等于线性空间每个张成组的长度.

♠

证明 任取 V 中的线性无关向量组 $A=\{u_1,\cdots,u_m\}$，张成组 $B=\{v_1,\cdots,v_n\}$，需证 $m\leqslant n$，我们证明，可以通过有限步操作，每一步将某个 v 替换为 u，并保证新的向量组仍然张成 V.

首先将 u_1 加入 B，得到 $B_1'=\{u_1,v_1,\cdots,v_n\}$，则 B_1' 是线性相关组（因为 B 为张成组，故 $u_1\in V$ 可刷 B 中向量线性表示），根据线性相关引理，可以删去某个 v（不妨设为 v_m），使得新的组 B_1 仍然张成 V.

再将 u_2 加入 B_1，得到 $B_2'=\{u_1,u_2,v_1,\cdots,v_{m-1}\}$，则 B_2' 是线性相关组，根据线性相关引理，可以删去某个 v（不妨设为 v_{m-1}），使得新组 B_2 仍然张成 V.

将上述操作进行 m 步，则所有的 u 都被添加到了组 B 中，这说明 $m\leqslant n$.

注 对上面的证明做一些说明，根据线性相关引理的描述，对于线性相关组 $v_1,\cdots,v_m$，删去的 v_j 要满足 $v_j\in\operatorname{Span}(v_1,\cdots,v_{j-1})$，则根据 $u_1,\cdots,u_m$ 的线性无关性，使用引理删去的元素必定为某个 v，而不会是 u，根据这

种性质，若 $m>n$，则必有某一步删去了 u，这与 $u_1,\cdots,u_m$ 的线性相关性是矛盾的.

上述命题实际上说明了线性无关组与张成组存在某种“边界”，无关组可以通过扩张达到边界，张成组可以通过缩减达到边界，这个边界，同时也是上面命题的取等条件，就是**基**，线性空间中的**线性无关张成组**. 由此可以给出一个基的判定准则.

命题 1.8 (基的判定准则)

$V/\mathbb{F}$ 中向量组 $v_1,\cdots,v_n$ 是 V 的基当且仅当每个（由张成保证）$v\in V$ 都能唯一（由线性无关保证）表示为

$$v=a_1v_1+\cdots+a_nv_n, \tag{1.19}$$

其中 $a_1,\cdots,a_n\in\mathbb{F}$.

前面大量的讨论还可以得到一个重要的副产物.

定理 1.2

有限维线性空间中，任何张成组都可以化简成一组基；任何线性无关组都可以扩充为一组基.

做了大量的铺垫，最后我们使用命题1.7证明我们所期望的命题.

定理 1.3 (维数的良定性)

有限维向量空间 V 的任意两个基的长度都相同，即维数不依赖基的选取.

证明 任取 V 的两个基 B_1,B_2，则 B_1 是 V 中的线性无关组，B_2 是 V 中的张成组，因此 B_1 的长度不超过 B_2 的长度；交换 B_1,B_2 的角色，即 B_2 是 V 中的线性无关组，B_1 是 V 中的张成组，因此 B_2 的长度不超过 B_1 的长度，这说明 B_1,B_2 的长度相等.

可以说，**有限维线性空间最良好的性质就是存在一组名为基的有限向量组，它给了抽象以具体的刻画，由此可以窥探线性空间的一切性质**. 在线性代数后面的学习中，我们在很多地方会用基的角度思考问题：用基得到核与像的维数关系（线性代数基本定理）、选择不同的基使得线性映射在其下的矩阵有更好的形式（相抵、相似、相合、Jordan 标准型...）、甚至借助特殊的基进行估计（内积空间、标准正交基）……

在此，我们从基和维数的角度出发，重新看待子空间以及子空间的和空间.

例 1.1 假设 $V/\mathbb{F}$ 为有限维向量空间，$W\leqslant V$，我们取 V 的一组基 $v_1,\cdots,v_n$，则由于 $W\leqslant V=\mathrm{Span}(v_1,\cdots,v_n)$，因此 W 的张成组至多有 n 个元素，这便得到了第一个结论：

$$\dim W\leqslant\dim V. \tag{1.20}$$

二者不仅有维数上的关系，在基的角度上还有更精确的关系：借助 V 的一组基，我们可以得到 W 的一组基；同样借助 W 一组基，也可将其扩张为 V 的一组基.

先考虑前一种情况，由多到少，自然要删去一些“无用”的向量，比如若某个 $v\in V\backslash W$，那么它对于张成 W 毫无帮助，因此我们保留 $W\cap\{v_1,v_2,\cdots,v_n\}$ 中元素即可，不妨设为 $v_1,\cdots,v_m$，由于线性相关组的子向量组仍然线性相关，因此我们只需证明，$W=\mathrm{Span}(v_1,\cdots,v_m)$.

由于对任意 $w\in W\leqslant v$，存在唯一的 $a_1,\cdots,a_n\in\mathbb{F}$，使得

$$w=\sum_{k=1}^{n}a_kv_k=\sum_{k=1}^{m}a_kv_k+\sum_{k=m+1}^{n}a_kv_k \tag{1.21}$$

只需证明 $a_{m+1}=\cdots=a_n=0$ 即可，若不然，则由于 $w'=-\sum\limits_{k=1}^{m}a_kv_k\in W$（封闭性），则 $w+w'\in W$，但是

$$w+w'=\sum_{k=m+1}^{n}a_kv_k\in W\in\mathrm{Span}(v_{m+1},\cdots,v_n, \tag{1.22}$$

$$\mathrm{Span}(v_1,\cdots,v_m)\cap\mathrm{Span}(v_{m+1},\cdots,v_n)=\{0\} \tag{1.23}$$

因此必有 $w+w'=0$，即证.

考虑另一方面，假设给出了 w 的一组基 $w_1,\cdots,w_m$，只要添加 $n-m$ 个处于 $V\backslash W$ 中的线性无关向量，即可得到 V 中的一组基（根据定义）.

上面的过程中，由于

$$V=\mathrm{Span}(v_1,\cdots,v_n)=\mathrm{Span}(v_1,\cdots,v_m)\oplus\mathrm{Span}(v_{m+1},\cdots,v_n):=W\oplus W' \tag{1.24}$$

因此给定 V 的子空间 W，我们可以找到 V 的另一个子空间，使得 V 可以表示为两个子空间直和的形式，由此自然引出了补空间的概念.

定义 1.11 (补空间)

设 V 为有限维向量空间，对于 $W\leqslant V$，若 $W'\leqslant V$，且 $V=W\oplus W'$，则称 W' 是 W 的一个补空间. ♣

上面的过程实际上在暗示补空间是不唯一的，实际上考虑 $\mathbb{R}^3$，很容易得到例子（考虑两个向量张成的平面与平面外的向量）.

上面的过程实际上还反映出一个结果：若 $V=W\oplus W'$，那么 $\dim V=\dim W+\dim W'$. 重复使用这个结论，我们可以得到如下命题.

命题 1.9

设 V 为有限维向量空间，$U_1,\cdots,U_m\leqslant V$，并且

$$V=U_1\oplus\cdots\oplus U_m \tag{1.25}$$

则

$$\dim V=\dim U_1+\cdots+\dim U_m. \tag{1.26}$$

♠

如果对于 V 中两个一般的子空间，对于它们的和空间的维数，也有一般的结果，它与集合中的容斥原理类似.

命题 1.10

设 V 为有限维向量空间，$U_1,U_2\leqslant V$，则

$$\dim(U_1+U_2)=\dim U_1+\dim U_2-\dim(U_1\cap U_2) \tag{1.27}$$

♠

证明 证明有空再补.

维数给了我们衡量有限维线性空间与其子空间“大小”的一个偏序关系，根据上面的讨论也很容易看出，不存在与 V 维数相等的，真包含于 V 的子空间.

经过漫长的探索，我们终于建立起了对（有限维）线性空间的基本认识，从一开始的抽象定义到对基、维数的剖析，眼前的迷雾在逐渐散去. 数学的学习就是这样一个由祛魅的过程.

1.2 线性映射

有了线性空间，可以定义线性空间之间的一种特殊映射：线性映射. 简单来说，线性映射就是“保加法与数乘”的映射.

定义 1.12 (线性映射)

设有域 $\mathbb{F}$ 上的线性空间 V,W，称 $f:V\to W$ 为线性映射，若其满足

1. $\forall u,v\in V, f(u+v)=f(u)+f(v)$.
2. $\forall\lambda\in\mathbb{F},v\in V,f(\lambda v)=\lambda f(v)$.

特别的，若 $W=V$，则称此线性映射为 V 上的线性变换. 记 V 到 W 的线性映射的集合为 $\mathcal{L}(V,W)$，当

$W = V$ 时，可简写为 $\mathcal{L}(V, V) = \mathcal{L}(V)$.

我们最早接触的线性映射就是一次函数，线性映射的例子无处不在，比如微分就是一个线性映射，如果考虑 $\mathbb{R}^n$ 上的微分算子，它将一个 $\mathbb{R}^n$ 中的元素映为一个线性映射.

事实上，线性映射的集合也具有线性空间结构. 如果在线性映射的集合中考虑映射的加法，以及域中元素对映射的数乘，那么可以验证在这两种运算下，线性映射也会构成一个线性空间.

命题 1.11 (线性映射构成的线性空间)

对于域 $\mathbb{F}$ 上的有限维线性空间 V, W，定义 $\mathcal{L}(V, W)$ 中的加法与数乘：

1. 加法：$\forall f, g \in \mathcal{L}(V, W) : (f+g)(v) = f(v) + g(v), \forall v \in V$.
2. 数乘：$\forall f \in L(V, W), \lambda \in \mathbb{F} : (\lambda f)(v) = \lambda(f(v)), \forall v \in V$.

那么 $(\mathcal{L}(V, W), +, \cdot)$ 也是 $\mathbb{F}$ 上的线性空间.

证明 依次验证定义即可.

1.2.1 线性映射的核与像

不论是一般的映射，还是群同态、环同态，我们都会考虑两个与映射有关的重要对象：核与像. 群同态中核与像都是群，环同态的核是理想、像是环，而在线性空间中，我们将证明核与像都是线性空间，因此前者也被称为零空间，后者也被称为像空间[2].

定义 1.13 (核与像)

对于 $f \in \mathcal{L}(V, W)$，定义线性映射 f 的核与像分别为

$$\operatorname{Ker} f = \{v \in V : f(v) = 0\} \tag{1.28}$$

$$\operatorname{Im} f = \{f(v) \in W : v \in V\} = \{w \in W : \exists v \in V, w = f(v)\} \tag{1.29}$$

核空间包含了 V 中所有会被 f 零化的向量，像空间包含了 W 中所有有原像的向量.

命题 1.12

对于 $f \in \mathcal{L}(V, W)$，有 $\operatorname{Ker} f \leqslant V$，$\operatorname{Im} f \leqslant W$ 成立.

证明 首先证明 $\operatorname{Ker} f \leqslant V$. 显然有 $\operatorname{Ker} f \subset V$，并且 $f(0) = 0$，因此 $0 \in \operatorname{Ker} f$，假设 $v_1, v_2 \in \operatorname{Ker} f$，则对任意 $a, b \in \mathbb{F}$，都有

$$f(av_1 + bv_2) = af(v_1) + bf(v_2) = 0 \tag{1.30}$$

故 $av_1 + bv_2 \in \operatorname{Ker} f$，综上可得 $\operatorname{Ker} f \leqslant V$.

再来证明 $\operatorname{Im} f \leqslant W$. 显然有 $\operatorname{Im} f \subset W$，并且 $f(0) = 0$，因此 $0 \in \operatorname{Im} f$，对任意 $w_1, w_2 \in \operatorname{Im} f$，存在 $v_1, v_2 \in V$，使得 $w_i = f(v_i)$，则对任意 $a, b \in \mathbb{F}$，有

$$aw_1 + bw_2 = af(v_1) + bf(v_2) = f(av_1 + bv_2) \in \operatorname{Im} f \tag{1.31}$$

综上可得 $\operatorname{Im} f \leqslant W$.

从上面的证明过程也可以看出，由于 $\operatorname{Ker} f, \operatorname{Im} f$ 都会包含 0（虽然这样说，但二者所包含的分别是 V, W 中的零元），因此它们非空. 其次，如果从维数的角度理解上述结论，可以得到

$$\dim \operatorname{Ker} f \leqslant \dim V, \quad \dim \operatorname{Im} f \leqslant \dim W \tag{1.32}$$

[2]核与像有许多不同称呼，以及不同表示：核 (Ker)/零空间 (null)；像空间 (Im)/值域 (range).

对于线性映射，其核与像和其单性与满性有非常紧密的联系，像方面与一般映射相同，但是在核方面有更强的结论.

命题 1.13

设 $f \in \mathcal{L}(V, W)$，则

1. f 为单射当且仅当 $\operatorname{Ker} f = \{0\}$.
2. f 为满射当且仅当 $\operatorname{Im} f = W$.

证明

1. 对任意 $u, v \in V$，若 $f(u) = f(v)$，则 $f(u-v) = 0$. 假设 $\operatorname{Ker} f = \{0\}$，则说明 $f(u-v) = 0$ 必然有 $u - v = 0, u = v$，这说明 f 为单射.
   反之，假设 f 为单射，由于 $0 \in \operatorname{Ker} f$，若 $\operatorname{Ker} f$ 含有非 0 元素，则存在 $v \neq 0$，$f(v) = f(0) = 0$，这与 f 的单性矛盾，故 $\operatorname{Ker} f = \{0\}$.
2. 这是显然的. 根据定义，$\operatorname{Im} f = W$ 等价于对任意 $w \in W$，存在 $v \in V$ 使得 $f(v) = w$，即 f 为满射.

前面我们都是从宏观角度讨论线性映射，下面我们利用有限维线性空间的重要工具：基来讨论这件事.

对于 $f \in \mathcal{L}(V, W)$，我们取 V 的一组基 $v_1, \cdots, v_n$，由于 $\operatorname{Ker} f \leqslant V$，因此其中必有一些向量能张成 $\operatorname{Ker} f$，不妨设

$$\operatorname{Ker} f = \operatorname{Span}(v_1, \cdots, v_m), \tag{1.33}$$

则根据 $\operatorname{Ker} f$ 的定义，必有 $f(v_i) = 0, i = 1, \cdots, m$，$f(v_j) \neq 0, j = m+1, \cdots, n$. 这些都是我们在子空间维数部分讨论过的结果，但核的特性使得我们可以由此理解像的结构. 假设 $v = \sum\limits_{k=1}^{n} a_k v_k$，则

$$f(v) = f\left(\sum_{k=1}^{n} a_k v_k\right) = \sum_{k=m+1}^{n} a_k f(v_k) = f\left(\sum_{k=m+1}^{n} a_k v_k\right) \tag{1.34}$$

由于 $v_{m+1}, \cdots, v_n$ 线性无关，因此 $a_{m+1}v_{m+1} + \cdots + a_n v_n = 0$ 当且仅当 $a_{m+1} = \cdots = a_n = 0$，并且 $f(v_j) \neq 0$，因此

$$0 = \sum_{k=m+1}^{n} a_k f(v_k) = f\left(\sum_{k=m+1}^{n} a_k v_k\right) \iff \sum_{k=m+1}^{n} a_k v_k \in \operatorname{Ker} f \tag{1.35}$$

$$\iff a_{m+1} = \cdots = a_n = 0. \tag{1.36}$$

这说明 $f(v_{m+1}), \cdots, f(v_n) \in \operatorname{Im} f \subset W$ 线性无关. 并且由于核空间中的向量都被 f 零化，因此其像不会在张成 $\operatorname{Im} f$ 中做出贡献，因此我们可以断言：$\operatorname{Im} f = \operatorname{Span}(f(v_{m+1}), \cdots, f(v_n))$. 实际证明也很简单，对于任意 $w \in \operatorname{Im} f$，都存在 $v = \sum\limits_{k=1}^{n} a_k v_k$，使得

$$w = f(v) = f\left(\sum_{k=1}^{n} a_k v_k\right) = \sum_{k=m+1}^{n} a_k f(v_k) \tag{1.37}$$

因此 $w \in \operatorname{Span}(f(v_{m+1}), \cdots, f(v_n))$，这表明 $\operatorname{Im} f \subset \operatorname{Span}(f(v_{m+1}), \cdots, f(v_n))$，根据互相包含性，我们有（同时将核空间的结果对比）

$$\operatorname{Ker} f = \operatorname{Span}(v_1, \cdots, v_m) \tag{1.38}$$

$$\operatorname{Im} f = \operatorname{Span}(f(v_{m+1}), \cdots, f(v_n)) \tag{1.39}$$

这里顺便证明了一个结论：有限维线性空间 V 在线性映射 f 下的像仍然是有限维线性空间. 但不止于此，从上面的对比我们可以看出很明显的数量关系，这是线性代数中一个非常重要的结论：

定理 1.4 (线性映射基本定理 (维数定理))

设 V 为有限维向量空间，$f \in \mathcal{L}(V, W)$，则 $\operatorname{Im} f \leqslant W$ 是有限维的，且

$$\dim V = \dim \operatorname{Ker} f + \dim \operatorname{Im} f. \tag{1.40}$$

借助这个定理，可以得到两个关于单性与满性的结论.

命题 1.14

设 V, W 为有限维线性空间，$f \in \mathcal{L}(V, W)$，则

1. 若 $\dim V > \dim W$，则 f 不是单射.
2. 若 $\dim V < \dim W$，则 f 不是满射.

证明

1. $\dim \operatorname{Ker} f = \dim V - \dim \operatorname{Im} f \geqslant \dim V - \dim W \geqslant 1$.
2. $\dim \operatorname{Im} f = \dim V - \dim \operatorname{Im} f \leqslant \dim V < \dim W$.

借助线性映射，可以对齐次线性方程组进行一些讨论.

例 1.2 关于齐次线性方程组的讨论.

1.2.2 线性空间的同构

有限维线性空间看似非常非常庞杂，但可以借助维数将其转化为一个个等价类，这之间的等价关系被称为同构（容易证明，可逆线性映射的逆映射仍然是可逆映射）.

定义 1.14 (同构)

称两个有限维线性空间 V, W 是同构的，当且仅当存在从 V 到 W 的可逆线性映射 f，这时 f 称为两个空间之间的同构映射，简称同构.

下面的命题告诉我们，线性映射可以沟通两个线性空间中向量的线性无关性.

命题 1.15

设 $f \in \mathcal{L}(V, W)$，假设对于 $v_1, \cdots, v_m \in V$，$f(v_1), \cdots, f(v_m)$ 线性无关，则 $v_1, \cdots, v_m \in V$ 线性无关.

证明 假设 $\sum\limits_{k=1}^{m} a_k v_k = 0$，则

$$0 = f\left(\sum_{k=1}^{m} a_k v_k\right) = \sum_{k=1}^{m} a_k f(v_k) \Longrightarrow a_1 = \cdots = a_m = 0. \tag{1.41}$$

因此 $v_1, \cdots, v_m$ 线性无关.

下面的定理揭示了维数与同构之间的关系.

定理 1.5 (同构与维数)

有限维线性空间 V, W 同构当且仅当 $\dim V = \dim W$.

证明 $V \cong W$ 当且仅当存在双射 $f \in \mathcal{L}(V, W)$，这表明 $\operatorname{Ker} f = \{0\}, \operatorname{Im} f = W$，而根据维数定理有

$$\dim V = \dim \operatorname{Ker} f + \dim \operatorname{Im} f = \dim W \tag{1.42}$$

得证.

事实上，上面的证明中有多余的条件，这也是线性映射的良好性质之一，在特定情形下单性与满性等价，这一事实作为如下推论给出.

推论 1.1

设 V,W 为有限维向量空间，若 $\dim V=\dim W$，则 $f\in\mathcal{L}(V,W)$ 为单射当且仅当 f 为满射.

证明 维数定理表明 $\dim V=\dim\operatorname{Ker}f+\dim\operatorname{Im}f$，因此

$$\operatorname{Ker}f=\{0\}\Longleftrightarrow\dim\operatorname{Ker}f=0\Longleftrightarrow\dim\operatorname{Im}f=\dim V=\dim W\Longleftrightarrow\operatorname{Im}f=W \tag{1.43}$$

注 乍看来上述推论的条件要求过高，但这一结论当 $W=V$ 时有很好的应用，这一点在讨论线性变换（线性算子）时会涉及到.

根据维数定理可以很容易证明定理，但从基的角度可以更好理解“同构”的概念.$\dim V=\dim W$ 告诉我们 V,W 中基的数量是相同的，更进一步，借助上面的命题1.15可知，若 $w_1,\cdots,w_n$ 为 W 的一组基，那么设 $v_i=f^{-1}(w_i),i=1,\cdots,n$，则 $v_1,\cdots,v_n$ 为 V 的一组基.

甚至更好地，对于同构的线性空间 V,W 中的基 $v_1,\cdots,v_n$，$w_1,\cdots,w_n$，我们必定存在一个同构映射 f，使得 $f(v_i)=w_i,i=1,\cdots,n$，并且这种映射是唯一的.

命题 1.16

设有限维线性空间 $V\cong W$，$v_1,\cdots,v_n$，$w_1,\cdots,w_n$ 分别为 V,W 中的一组基，则存在唯一双射 $f\in\mathcal{L}(V,W)$，使得 $f(v_i)=w_i,i=1,\cdots,n$.

证明 定义映射

$$f:V\longrightarrow W \tag{1.44}$$

$$\sum_{k=1}^{n}a_kv_k\longmapsto\sum_{k=1}^{n}a_kw_k \tag{1.45}$$

下面我们依次证明 f 线性、双射、唯一.

线性：对于任意 $v_1=\sum a_kv_k,v_2=\sum b_kv_k\in V$，$\lambda,\mu\in\mathbb{F}$，有

$$f(\lambda v_1+\mu v_2)=f\left(\sum(\lambda a_k+\mu b_k)v_k\right)=\sum(\lambda a_k+\mu b_k)w_k=\lambda f(v_1)+\mu f(v_2) \tag{1.46}$$

双射：$f(v)=0$ 等价于存在 $a_1,\cdots,a_n$ 使得

$$f(v)=f\left(\sum_{k=1}^{n}a_kv_k\right)=\sum_{k=1}^{n}a_kw_k=0 \tag{1.47}$$

根据 w 的线性无关性可知 $a_1=\cdots=a_n=0$，因此 $\operatorname{Ker}f=\{0\}$，故 f 单. 对于任意 $w=\sum a_kw_k\in W$，都存在 $v=\sum a_kv_k\in V$ 使得 $f(v)=w$，故 f 满.

唯一性：假设存在线性映射 $f_1,f_2\in\mathcal{L}(V,W)$ 满足条件，则令 $g=f_1-f_2\in\mathcal{L}(V,W)$，对于任意 $v=\sum a_kv_k$ 都有 $g(v)=\sum a_kg(v_k)=0$，故 $g=0,f_1=f_2$.

1.2.3 线性空间的积

我们前面讨论过同一个线性空间中的子空间的和，下面我们讨论同一个域上线性空间的积（也称外直和）.

定义 1.15 (积空间)

设 $V_1,\cdots,V_m$ 均为域 $\mathbb{F}$ 上的线性空间，则定义其积空间为

$$\prod_{k=1}^{m}V_k=V_1\times\cdots\times V_m=\{(v_1,\cdots,v_m):v_k\in V_k\}, \tag{1.48}$$

其中加法、数乘定义为各个分量上的加法、数乘.

容易验证，上面定义的确实是一个线性空间. 线性空间的积就是将一系列线性空间并置得到的新空间，比如通过域 $\mathbb{R}$ 可以构造出空间 $\mathbb{R}^n$，这是我们最熟悉的一类线性空间.

构造出了积空间，我们很自然会考虑它是否为有限维，若是则进一步考虑维数与基. 事实上，积空间可以对于每个分量作如下分解（虽然都使用了记号 0，但它们代表不同线性空间中的零元）

$$V_1\times\cdots\times V_m=(V_1\times\{0\}\times\cdots\times\{0\})\oplus(\{0\}\times V_2\times\cdots\times\{0\})\oplus(\{0\}\times\{0\}\times\cdots V_m)\tag{1.49}$$

因此很容易得到如下结论

命题 1.17

设 $V_1,\cdots,V_m$ 均为域 $\mathbb{F}$ 上的有限维向量空间，则

$$\dim(V_1\times\cdots\times V_m)=\dim V_1+\cdots+\dim V_m.\tag{1.50}$$

♠

1.2.4 商空间与限制映射

本节讨论线性空间与其子空间之间的商运算，首先介绍向量与子空间的和：仿射子集的概念.

定义 1.16 (仿射子集)

对于线性空间 V 以及其子空间 U，以及 $v\in V$，可以定义 V 的仿射子集

$$v+U=\{v+u:u\in U\}.\tag{1.51}$$

我们称 $v+U$ 平行于 U.

♣

仿射子集的概念很好理解，比如考虑 $V=\mathbb{R}^2$，U 为过原点的一条直线，则 $v+U$ 就代表了与 U 平行，且包含 v 的直线. 如果考虑 $V=\mathbb{R}^3$，也可以用空间中的平面类比这一点，或者说，这就是对 U 在 V 中的一种“平移”.

根据这种几何直观，可以立即发现一个事实：对于不同的向量 $v_1,v_2\in V$，平移的结果——仿射子集 v_1+U,v_2+U 也有可能是相同的，由此可以定义一个等价关系，但我们先给出如下命题.

引理 1.1

设线性空间 $U\leqslant V$，$v_1,v_2\in V$，则 $v_1+U=v_2+U$ 等价于 $v_1-v_2\in U$.

♡

证明 假设 $v_1+U=v_2+U$，则任取 $v_1+u\in v_1+U$，存在 $v_2+u'\in v_2+U$ 使得 $v_1+u=v_2+u'$，这表明 $v_1-v_2=u'-u\in U$. 反之，若 $v_1-v_2\in U$，则存在 $u_0\in U$ 使得 $v_1=v_2+u_0$，根据 U 的封闭性

$$v_1+U=\{v_1+u:u\in U\}=\{v_2+u_0+u:u\in U\}=\{v_2+u:u\in U\}=v_2+U\tag{1.52}$$

得证.

下面给出这种等价关系的定义，这里借用数论中的同余语言来描述

定义 1.17

设线性空间 $U\leqslant V$，则称 $v_1,v_2\in V$ 模 U 同余，若 $v_1+U=v_2+U$，或者说 $v_1-v_2\in U$，记为 $v_1\equiv v_2(\operatorname{mod} U)$.

♣

容易验证，仿射子集并不是线性空间（通常不满足封闭性），但如果考虑仿射子集的全体，那么它们在良定义的加法与数乘下构成线性空间，即商空间.

定义 1.18 (商空间)

设 $U\leqslant V$，则定义商空间为 V 的所有与 U 平行的仿射子集的集合

$$V/U=\{v+U:v\in V\}.\tag{1.53}$$

♣

还需要验证商空间确实是线性空间，我们定义加法与数乘为

$$(v_1+U)+(v_2+U)=(v_1+v_2)+U,\quad a(v_1+U)=av_1+U\tag{1.54}$$

这在集合的角度来看是自然的，只需证明其是良定义的，即对任意 $v_1 \equiv v_1'(\mathrm{mod}\, U), v_2 \equiv v_2'(\mathrm{mod}\, U)$，有

$$(v_1+v_2)+U=(v_1'+v_2')+U, \quad av_1+U=av_1'+U \tag{1.55}$$

由于 $v_1-v_1', v_2-v_2' \in U$，因此根据 U 的封闭性可知

$$(v_1+v_2)-(v_1'+v_2')=(v_1-v_1')+(v_2-v_2')\in U, \quad a(v_1-v_1')\in U. \tag{1.56}$$

其余验证过程是容易的.

事实上，线性空间可以如此轻易作商运算并不是一件自然的事，上面的过程如此容易的原因在于线性空间的优良性质. 商群需要正规子群，商环需要理想，而商空间只需要子空间，并没有其它特殊的要求，可以简单理解为，线性空间自带加法 Abel 群以及两种运算的封闭性.

对于上面的讨论，可以很容易定义出商映射.

定义 1.19 (商映射)

设 $U\leqslant V$，则定义商映射 $\pi: V\to V/U$，对任意 $v\in V$，$\pi(v)=v+U$.

借助商映射，可以求出商空间的维数. 易知 π 是一个线性映射，并且是满的，使用维数定理可得

$$\dim V=\dim\operatorname{Ker}\pi+\dim V/U \tag{1.57}$$

考虑 $\operatorname{Ker}\pi=\{v\in V: v\equiv 0\}$，可知 $\operatorname{Ker} f\subset U$，因此取 V 的一组基 $v_1,\cdots,v_m,\cdots,v_n$，不妨设其中仅 $v_1,\cdots,v_m\in U$，则

$$\operatorname{Ker}\pi=\operatorname{Span}(v_1,\cdots,v_m) \tag{1.58}$$

这表明 $\dim\operatorname{Ker}\pi=\dim U$，因此我们有如下结论

命题 1.18

设 V 为有限维向量空间，$U\leqslant V$，则

$$\dim V/U=\dim V-\dim U \tag{1.59}$$

如果对于 $f\in\mathcal{L}(V,W)$，令 $U=\operatorname{Ker} f$，则 $\dim V/\operatorname{Ker} f=\dim V-\dim\operatorname{Im} f$

命题 1.19

设 V,W 为有限维向量空间，$f\in\mathcal{L}(V,W)$，则 $V/\operatorname{Ker} f\cong\operatorname{Im} f$.

上面的结果很像同态基本定理，直接构造映射 $g: V/\operatorname{Ker} f\to\operatorname{Im} f$，$g(v+\operatorname{Ker} f)=f(v)$ 也容易证明.

1.3 矩阵与线性映射

矩阵是线性代数的重点之一，可以说线性映射与矩阵分别是线性代数“飘渺”与“实在”的两种体现，很多问题都有两种思考方向上的不同形式，有时需要直击本质，有时又需要脚踏实地；但另一方面线性映射与矩阵相伴而生，二者的联系在于同构，因为**有限维线性空间之间的线性映射构成的空间，在一定条件下存在与之同构的矩阵空间**，这种“重贴标签”使得二者可以互相处理一些繁琐的问题，甚至相互揭示本质，首先从定义矩阵开始.

1.3.1 矩阵的定义

定义 1.20 (域上的矩阵)

称域 $\mathbb{F}$ 上的 $m\times n$ 矩阵为 $\mathbb{F}$ 中的元素构成的 m 行 n 列的矩形阵列

$$A=\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \tag{1.60}$$

$\mathbb{F}$ 上全体 $m\times n$ 矩阵构成的集合记作 $\mathbb{F}^{m\times n}$.

仿照 $\mathbb{F}^n$，可以定义矩阵中的加法与数乘，这实际上是每个分量上的加法与数乘（这也说明加法只能在同阶矩阵间进行）

$$(a_{ij})+(b_{ij})=(a_{ij}+b_{ij}),\quad \lambda(a_{ij})=(\lambda a_{ij}) \tag{1.61}$$

仿照 $\mathbb{F}^n$，容易验证 $\mathbb{F}^m/\mathbb{F}$ 是一个线性空间. 更进一步，如果将矩阵的所有元素排成一列，可以发现同构关系 $\mathbb{F}^{m\times n}\cong\mathbb{F}^{mn}$，由此可以得到 $\mathbb{F}^{m\times n}$ 的一组基 $E_{ij},1\leqslant i\leqslant m,1\leqslant j\leqslant n$，每个 E_{ij} 的 ij 位置为 1，其余位置为 0，称为矩阵空间的标准基.

这些讨论可以归结为如下命题

命题 1.20

$\mathbb{F}^{m\times n}/\mathbb{F}$ 是一个线性空间，并且 $\dim\mathbb{F}^{m\times n}=mn$.

1.3.2 线性映射的矩阵

本节正式讨论“矩阵与线性映射相伴而生”这一观点，我们已经定义了域上的矩阵，而域中的元素是进行线性空间中线性组合的关键，因此可以定义出线性映射的矩阵.

定义 1.21 (线性映射的矩阵)

设 V,W 为有限维向量空间，$f\in\mathcal{L}(V,W)$，固定 V,W 的基 $v_1,\cdots,v_n$，$w_1,\cdots,w_m$，则定义 f 关于这些基的矩阵为 $\mathcal{M}(f)=(a_{ij})\in\mathbb{F}^{m\times n}$，其中

$$\begin{cases} f(v_1)=a_{11}w_1+\cdots+a_{m1}w_m \\ \cdots\cdots \\ f(v_n)=a_{1n}w_1+\cdots+a_{mn}w_m \end{cases} \tag{1.62}$$

注 上面的定义中分别固定了 V,W 中的一组基，关于不同基下矩阵的关系留与后文讨论.

容易验证，在固定的基下，若映射 $f,g\in\mathcal{L}(V,W)$ 对应的矩阵为 A,B，则 $\lambda f+\mu g$ 对应的矩阵恰好是 $\lambda A+\mu B$，或者说

$$\mathcal{M}(\lambda f+\mu g)=\lambda\mathcal{M}(f)+\mu\mathcal{M}(g). \tag{1.63}$$

由此可以看出，即使不参照 $\mathbb{F}^n$，依旧可以定义出矩阵的加法与数乘.

从上面的关系中已经可以看出

$$\mathcal{L}(V,W)\cong\mathbb{F}^{m\times n},\quad n=\dim V,m=\dim W \tag{1.64}$$

这是线性空间角度的同构，由此可以看出，矩阵可以承载线性映射所蕴含的信息，进一步思考，映射在某些情况下可以进行复合，那么能否定义矩阵的运算，使得映射复合的矩阵恰好等于映射对应矩阵进行运算的结果？出于这种思路，可以定义出矩阵乘法，首先考虑一个例子.

例 1.3 设 $f\in\mathcal{L}(V,W),g\in\mathcal{L}(W,U)$，$\mathcal{M}(f)=(a_{ij})\in\mathbb{F}^{m\times n},\mathcal{M}(g)=(b_{ij})\in\mathbb{F}^{p\times m}$，取三个空间的基 $v_1,\cdots,v_m$，

$w_1, \cdots, w_n$，$u_1, \cdots, u_p$，则

$$\begin{cases} f(v_1) = a_{11}w_1 + \cdots + a_{m1}w_m \\ \cdots\cdots \\ f(v_n) = a_{1n}w_1 + \cdots + a_{mn}w_m \end{cases} \qquad \begin{cases} g(w_1) = b_{11}u_1 + \cdots + b_{p1}u_p \\ \cdots\cdots \\ g(w_m) = b_{1m}u_1 + \cdots + b_{pm}u_p \end{cases} \tag{1.65}$$

设 $\mathcal{M}(g\circ f) = (c_{ij})$，则

$$\begin{cases} g\circ f(v_1) = \sum\limits_{l=1}^{m} a_{l1}g(w_l) = \sum\limits_{k=1}^{p}\sum\limits_{l=1}^{n} b_{kl}a_{l1}u_k = \sum\limits_{k=1}^{p} c_{k1}u_k \\ \cdots\cdots \\ g\circ f(v_n) = \sum\limits_{l=1}^{m} a_{ln}g(w_l) = \sum\limits_{k=1}^{p}\sum\limits_{l=1}^{n} b_{kl}a_{ln}u_k = \sum\limits_{k=1}^{p} c_{kn}u_k \end{cases} \tag{1.66}$$

其中 $c_{ij} = \sum\limits_{k=1}^{n} b_{ik}a_{kj}$，根据下面的定义，可将 $\mathcal{M}(g\circ f)$ 定义为 $\mathcal{M}(g)\mathcal{M}(f)$，即矩阵乘法.

定义 1.22 (矩阵乘法)

设 $A \in \mathbb{F}^{m\times n}, B \in \mathbb{F}^{n\times p}$，则定义

$$AB = \left(\sum_{k=1}^{n} a_{ik}b_{kj}\right) = (c_{ij}) \in \mathbb{F}^{m\times p}. \tag{1.67}$$

♣

矩阵乘法定义看似麻烦，但将两个矩阵写出则会非常直观

$$AB = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{np} \end{pmatrix} = \begin{pmatrix} \sum\limits_{k=1}^{n} a_{1k}b_{k1} & \cdots & \sum\limits_{k=1}^{n} a_{1k}b_{kp} \\ \vdots & & \vdots \\ \sum\limits_{k=1}^{n} a_{mk}b_{k1} & \cdots & \sum\limits_{k=1}^{n} a_{mk}b_{kp} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1p} \\ \vdots & & \vdots \\ c_{m1} & \cdots & c_{mp} \end{pmatrix} = C \tag{1.68}$$

并且因为线性映射有结合律，由此矩阵乘法也自然满足结合律（交换律则不然），并且易知，矩阵乘法的单位元为单位阵（它实际上对应了线性空间到自身的恒定映射）

$$I_n = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \tag{1.69}$$

即对于 $A \in \mathbb{F}^{m\times n}$，$AI_n = I_mA = A$，可以看出左单位与右单位不同.

综上，我们证明了矩阵空间与线性映射空间的同构关系，以及映射复合与矩阵乘法之间的联系.

1.3.3 线性映射的矩阵表示

对于线性映射 $f \in \mathcal{L}(V, W)$，假设它在基 $v_1, \cdots, v_n$，$w_1, \cdots, w_m$ 下的矩阵为 $A = (a_{ij})$，则

$$\begin{cases} f(v_1) = a_{11}w_1 + \cdots + a_{m1}w_m \\ \cdots\cdots \\ f(v_n) = a_{1n}w_1 + \cdots + a_{mn}w_m \end{cases} \tag{1.70}$$

如果换一种写法，将 $v_1, \cdots, v_n$ 并置，将结果也并置，并且写成矩阵乘法的形式，则有

$$f(v_1, \cdots, v_n) = \left(\sum_{k=1}^{n} a_{k1}w_k, \cdots, \sum_{k=1}^{n} a_{kn}w_k\right) = (w_1, \cdots, w_k) \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \tag{1.71}$$

这种写法就是线性映射的矩阵表示，这种形式非常直观表示了“映射在何基下有何矩阵”，并且只需改变不同的基和矩阵，就可以得到同一映射在不同基下的矩阵. 在此基础上，它也可以契合线性映射的复合，比如延续讨论矩阵乘法前的例子中的记号，则

$$g\circ f(v_1,\cdots,v_n)=g(w_1,\cdots,w_m)\begin{pmatrix}a_{11}&\cdots&a_{1n}\\\vdots&&\vdots\\a_{m1}&\cdots&a_{mn}\end{pmatrix}=(u_1,\cdots,u_p)\begin{pmatrix}b_{11}&\cdots&b_{1m}\\\vdots&&\vdots\\b_{p1}&\cdots&b_{pm}\end{pmatrix}\begin{pmatrix}a_{11}&\cdots&a_{1n}\\\vdots&&\vdots\\a_{m1}&\cdots&a_{mn}\end{pmatrix}\tag{1.72}$$

$$=(u_1,\cdots,u_p)BA\tag{1.73}$$

上面的每一步都用到了结合律，但这种结合律的本质仍然是映射的复合. 注意，线性映射的矩阵表示只是只是一种契合矩阵规则的表示方法，并没有定义多少新的东西.

下面定义坐标的概念，坐标有助于矩阵研究线性映射，同时也能充分展示矩阵表示的优势.

定义 1.23 (坐标)

设 $v_1,\cdots,v_n$ 为 V 的一组基，则对于任意 $v\in V$，存在唯一 $a_1,\cdots,a_n$ 使得 $v=a_1v_1+\cdots+a_nv_n$，这里记 $(a_1,\cdots,a_n)^T\in\mathbb{F}^{n\times1}$ 为 v 在基 $v_1,\cdots,v_n$ 下的坐标. ♣

如果使用矩阵表示，那么上面的事实可以写作

$$v=(v_1,\cdots,v_n)\begin{pmatrix}a_1\\\vdots\\a_n\end{pmatrix}\tag{1.74}$$

延续前面的记号，若要求 $f(v)$，则

$$f(v)=f(v_1,\cdots,v_n)\begin{pmatrix}a_1\\\vdots\\a_n\end{pmatrix}=(w_1,\cdots,w_m)\begin{pmatrix}a_{11}&\cdots&a_{1n}\\\vdots&&\vdots\\a_{m1}&\cdots&a_{mn}\end{pmatrix}\begin{pmatrix}a_1\\\vdots\\a_n\end{pmatrix}\tag{1.75}$$

这里依然满足结合律，并且它告诉我们，$f(v)$ 的坐标正好是矩阵 $\mathcal{M}(f)$ 左乘 v 的坐标后，得到的矩阵与基 $w_1,\cdots,w_m$ 作线性组合的结果.

除此之外，矩阵表示也很方便进行基变换，这也是后面要讨论的主题之一，它可以归结到这样的问题：

问题 1.1 研究线性映射 $f\in\mathcal{L}(V,W)$，可以分别取 V,W 的基，研究这组基下的矩阵，那么是否存在合适的基，使得 f 在这对基下的矩阵尽可能简单?

这个问题的重点在于如何定义“简单”，以及如何联系不同基之间的矩阵，最后得到的结果被称为相抵标准型.

第 2 章 线性映射与矩阵理论

本章将在前一章基础上进一步讨论矩阵与线性变换之间的关系，首先对于矩阵的基本理论进行完善，并介绍行列式，再讨论相抵标准型理论，最后讨论相似标准型.

讨论线性映射的矩阵表示方法，然后是矩阵分块相关理论，以及行列式，最后仿照相抵标准型的思路研究相似变换.

2.1 行列式

2.1.1 行列式的定义

定义 2.1 (n 阶行列式)

对于域 $\mathbb{F}$，定义行列式函数 $\det : \underbrace{\mathbb{F}^{n\times 1} \times \cdots \times \mathbb{F}^{n\times 1}}_{n} \to \mathbb{F}$

$$\det(a_{\cdot 1}, \cdots, a_{\cdot n}) = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \sum_{(j_1,\cdots,j_n)\in S_n} (-1)^{\tau(j_1,\cdots,j_n)} a_{1j_1} \cdots a_{nj_n} \tag{2.1}$$

其中 S_n 表示 n 阶置换群，τ 表示置换的逆序数.

♣

命题 2.1

行列式函数的性质：反对称，多重线性，标准性

♠

行列式的唯一性.

2.1.2 行列式的展开

按行、列展开.

定理 2.1 (Laplace 展开)

♡

2.1.3 Cauchy-Binet 公式

定理 2.2 (Cauchy-Binet 公式)

♡

2.2 矩阵理论

2.2.1 矩阵的分块

矩阵的分块是一种很有用的操作方法，它表示将矩阵中某些部分看作一块，整体进行操作，这对应线性映射的限制与直积，这方面内容从映射角度就不如矩阵角度直观，而后者也易于操作.

2.2.2 行/列空间与秩

如果说第一章是在讨论线性映射与向量间的作用，那么本节就是在对应讨论矩阵与对应列向量之间的乘法，即 $\mathbb{F}^{m\times n}\times\mathbb{F}^{n\times 1}\to\mathbb{F}^{m\times 1}$. 首先定义 $\mathbb{F}^n$ 空间.

定义 2.2 (列向量空间)

设 $\mathbb{F}$ 为域，定义 $\mathbb{F}^n$ 为 $\mathbb{F}^{n\times 1}$ 在矩阵的加法、数乘下构成的线性空间. ♣

之所以称为列向量，是因为 $\mathbb{F}^{n\times 1}$ 矩阵具有 n 行 1 列的形式，类似也可将 $\mathbb{F}^{1\times m}$ 矩阵称为行向量，并且定义行向量空间. 而若取标准基 $e_i=(\delta_{i1},\cdots,\delta_{in})^T, i=1,\cdots,n$，则容易验证 $\mathcal{L}(\mathbb{F}^m,\mathbb{F}^n)$ 中映射的作用恰好与该映射在标准基下的矩阵左乘行向量的结果相同.

从列向量的角度看，$\mathbb{F}^{m\times n}$ 矩阵可以按照列向量进行分块，即 $A=(\alpha_1,\cdots,\alpha_n)$，其中每个 $\alpha_i\in\mathbb{F}^{m\times 1}$ 为 m 维列向量空间中的元素.

定义 2.3 (矩阵的列空间)

设矩阵 $A=(\alpha_1,\cdots,\alpha_n)\in\mathbb{F}^{m\times n}$，定义其列向量的张成空间为矩阵的列空间，即

$$\operatorname{col}A=\operatorname{Span}(\alpha_1,\cdots,\alpha_n)=\{Au:u\in\mathbb{F}^{n\times 1}\}\leqslant\mathbb{F}^{m\times 1}\tag{2.2}$$

并且称列空间的维数 $\dim\operatorname{col}A$ 为矩阵的列秩，记为 $\mathrm{rk}A$. 特别地，若 $\mathrm{rk}A=n$，则称矩阵 A 是列满秩的. ♣

注 第一章中定义过向量组的秩，这里的列秩可以看作是一种特殊的向量组的秩.

如果将矩阵 A 看作 $\mathcal{L}(\mathbb{F}^n,\mathbb{F}^m)$ 中元素，即左乘作用 $A(u)=Au$，那么 $\operatorname{col}A=\operatorname{Im}A$. 在这种考虑下，可以定义出矩阵的零空间

$$A=\operatorname{Ker}A=\{u\in\mathbb{F}^{n\times 1}:Au=0\}.\tag{2.3}$$

上面的一切都可以类似定义在行空间中，记为 $\operatorname{row}A$，并且定义中的列向量右乘需要改为行向量左乘. 借用新的记号，可以在矩阵的行/列空间中表述维数定理，注意到 $\operatorname{row}A=\operatorname{col}A^T$（这从定义或直观都是显然的），因此可以避免多余的记号.

定理 2.3 (维数定理)

设 $A\in\mathbb{F}^{m\times n}$，则

$$n=\dim\operatorname{Ker}A+\dim\operatorname{Im}A=\dim\operatorname{null}A+\mathrm{rk}A\tag{2.4}$$

$$m=\dim\operatorname{Ker}A^T+\dim\operatorname{Im}A^T=\dim\operatorname{null}A^T+\mathrm{rk}A^T\tag{2.5}$$

♡

有限维线性空间可以找到基，这呈现在矩阵中会更加直观. 对于行空间与列空间，设 $A\in\mathbb{F}^{m\times n}$，$\mathrm{rk}A=r\leqslant n$，设 $S=(s_1,\cdots,s_r)\in\mathbb{F}^{m\times r}$ 为 $\operatorname{col}A$ 的一组基，则存在 $c_1,\cdots,c_n\in\mathbb{F}^{r\times 1}$，使得每个 $Sc_i=\alpha_i$，因此令 $C=(c_1,\cdots,c_n)\in\mathbb{F}^{r\times n}$，则 $A=SC$.

定理 2.4

设 $A\in\mathbb{F}^{m\times n}$，$\mathrm{rk}A=r\leqslant n$，则存在矩阵列满秩矩阵 $S=(s_1,\cdots,s_r)\in\mathbb{F}^{m\times r},C\in\mathbb{F}^{r\times n}$ 使得

$$A=SC\tag{2.6}$$

♡

证明 如上所述.

特别的，张成空间的基可以直接从张成组中找到，因此有如下推论.

推论 2.1

设 $A\in\mathbb{F}^{m\times n}$，$\mathrm{rk}A=r\leqslant n$，则存在列满秩矩阵 $S=(\alpha_{i_1},\cdots,\alpha_{i_r})\in\mathbb{F}^{m\times r},C\in\mathbb{F}^{r\times n}$ 使得

$$A=SC \tag{2.7}$$

♡

为了定义矩阵的秩，需要证明如下引理

引理 2.1

矩阵的行秩等于列秩，若设 $A\in\mathbb{F}^{m\times n}$，则 $\dim\operatorname{col}A=\dim\operatorname{col}A^T$.

♡

证明 不妨设 $A=(\alpha_1,\cdots,\alpha_n)\neq 0$（否则命题显然成立），$\dim\operatorname{col}A=r\leqslant n$，则使用上面的引理，不妨设 $S=(\alpha_1,\cdots,\alpha_r)$ 张成 $\operatorname{col}A$，则存在 $C\in\mathbb{F}^{r\times n}$ 使得 $A=SC$，两边同时作转置，并分块 $C^T=(x_1,\cdots,x_r)$ 可得

$$A^T=C^TS^T=\begin{pmatrix}x_1&\cdots&x_r\end{pmatrix}\begin{pmatrix}\alpha_1^T\\ \vdots\\ \alpha_r^T\end{pmatrix}=x_1\alpha_1^T+\cdots+x_r\alpha_r^T \tag{2.8}$$

由于

$$\operatorname{col}A^T=\operatorname{Im}A^T=\{A^Tu:u\in\mathbb{F}^{m\times 1}\}=\{x_1(\alpha_1^Tu)+\cdots+x_r(\alpha_r^Tu):u\in\mathbb{F}^{m\times 1}\}\leqslant\operatorname{Span}(x_1,\cdots,x_r) \tag{2.9}$$

因此 $\dim\operatorname{col}A^T\leqslant r=\dim\operatorname{col}A$，将上述结论应用到 A^T 中，可得 $\dim\operatorname{col}A\leqslant\dim\operatorname{col}A^T$，命题得证.

定义 2.4 (矩阵的秩)

定义矩阵 $A\in\mathbb{F}^{m\times n}$ 的秩为其列秩，记为 $\mathrm{rk}A$. 特别地，若 $\mathrm{rk}A=n$ 则称 A 列满秩；若 $\mathrm{rk}A=m$ 则称 A 行满秩.

♣

2.2.3 矩阵的行列式

前面讨论了行列式，它与矩阵有相近的形式，因此我们可以对矩阵（更准确说是方阵）定义行列式.

定义 2.5 (矩阵的行列式)

设 $A=(a_{ij})\in\mathbb{F}^{n\times n}$，则定义矩阵的行列式为

$$\det A=\begin{vmatrix}a_{11}&\cdots&a_{1n}\\ \vdots&\ddots&\vdots\\ a_{n1}&\cdots&a_{nn}\end{vmatrix}. \tag{2.10}$$

♣

也就是说，矩阵的行列式就是将矩阵的每个元素放到行列式的对应位置，进行计算得到的结果. 由于行列式是一个多重线性函数，因此将矩阵列分块 $A=(\alpha_1,\cdots,\alpha_n)\in\mathbb{F}^{n\times n}$，则可得矩阵的行列式的另一种形式

$$\det A=\det(\alpha_1,\cdots,\alpha_n) \tag{2.11}$$

行列式的许多性质都可以追溯到最根本的三条性质，因此在这种写法下可以挖掘出行列式之于矩阵的许多性质.

。。。。。行列式与列向量的线性相关性，。。。。。

逆的定义与大多代数结构中的逆相似，即对于矩阵 A，若存在 B 使得 $AB=BA=I$，则称 B 为 A 的逆，但对于这个描述还有很多要补充的. 首先要保证 A,B 可以作乘法，因此若 $A\in\mathbb{F}^{m\times n}$，则 $B\in\mathbb{F}^{n\times p}$，同理，$B$ 也需要左乘 A，这表示 $p=m$. 其次，由于 $AB=BA=I$，根据 B-C 公式可知，若 $m\neq n$，则 $\det(AB),\det(BA)$ 中必有一者为 0，这说明逆矩阵存在的必要条件是矩阵为方阵.

定义 2.6 (逆矩阵)

对于 $A \in \mathbb{F}^{n\times n}$，若存在 $B \in \mathbb{F}^{n\times n}$ 使得 $AB = BA = I_n$，则称 B 为 A 的逆矩阵，记为 A^{-1}.

定理 2.5

设 $A \in \mathbb{F}^{n\times n}$，则 A 可逆当且仅当 $\det A \neq 0$.

2.3 相抵标准型

2.3.1 初等方阵与初等变换

本节讨论矩阵的初等变换，“初等”代表简单，并且通过组合可以得到很多变换. 三种初等变换为:

定义 2.7 (矩阵的初等变换)

定义如下三种变换为矩阵的初等变换:

1. 交换：交换矩阵的某两行（或某两列）.
2. 倍乘：给矩阵的某一行（或某一列）乘一个 $\mathbb{F}$ 中的非零倍数.
3. 乘加：将矩阵的某行倍乘后加到另一行（或将某一列倍乘后加到另一列）.

一个自然的问题是，上面三种变换能否通过矩阵乘法进行表示（下面先仅考虑列变换），为了讨论这个问题，可以将矩阵进行列分块.

首先考虑交换，不妨设交换第一列与第二列，即 $A = (a_1, a_2, \cdots, a_n) \mapsto (a_2, a_1, \cdots, a_n)$，其中 $A \in \mathbb{F}^{m\times n}$. 为了实现这个变换，需要有一个 $n \times n$ 方阵右乘 A，不妨设为 P，由于该变换中仅有前两列改变，因此可以进一步分块得到

$$(a_1, a_2, \alpha)\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = (a_1p_{11} + a_2p_{21} + \alpha p_{31}, a_1p_{12} + a_2p_{22} + \alpha p_{32}, a_1p_{13} + a_2p_{23} + \alpha p_{33}) \tag{2.12}$$

$$= (a_2, a_1, \alpha) \tag{2.13}$$

由此可得

$$P = \begin{pmatrix} 0 & 1 & \\ 1 & 0 & \\ & & I_{n-2} \end{pmatrix} \in \mathbb{F}^{n\times n} \tag{2.14}$$

如果设 $e_i = (\delta_{1i}, \cdots \delta_{ni})^T$，则更一般的交换 i, j 列的矩阵 $P_{ij} = (e_1, \cdots, e_j, \cdots, e_i, \cdots, e_n) \in \mathbb{F}^{n\times n}$（这里不妨设 $i < j$）.

仿照上面的过程，可以得到倍乘 $D_i(\lambda)$ 矩阵与乘加矩阵 $T_{ij}(\lambda)$（表示将 i 列倍乘加到 j 列）为

$$D_i(\lambda) = \begin{pmatrix} I_{i-1} & & \\ & \lambda & \\ & & I_{n-i} \end{pmatrix}, \quad T_{ij}(\lambda) = \mathrm{diag}(e_1, \cdots, e_j + \lambda e_i, \cdots, e_n), \quad \lambda \neq 0 \tag{2.15}$$

将上述矩阵对矩阵 A 右乘，即可进行初等列变换. 根据对称性，可以得到初等行变换矩阵，这实际上就是将上面叙述中的行于列互换，并且将右乘改为左乘.

对象在一个变换下的不变性往往值得关注，对于上面的初等变换，一个很重要的性质就是秩不变性.

定理 2.6

初等（行、列）变换不改变矩阵的秩.

证明 初等变换实际上是对列向量进行线性组合，而这一过程不会改变张成空间，因此不会改变张成组的维数，因此不会改变秩.

有了这样的性质，自然会考虑其反面：对于两个同秩的同阶矩阵，能否通过不断作初等变换使之相等？这就是矩阵的相抵问题.

2.3.2 矩阵的相抵

为了更方便表述，先利用初等变换定义矩阵的相抵.

定义 2.8 (相抵)

称同阶矩阵 A, B 相抵，若 A 可以通过一系列初等变换变成 B，记为 $A \sim B$.

由于初等变换的逆变换也是初等变换（初等矩阵的逆也是初等矩阵），因此相抵是一个等价关系. 如此我们前面的问题转化为探索相抵与秩之间的关系. 下面的命题，有助于进一步讨论相抵.

命题 2.2

可逆矩阵 $A \in \mathbb{F}^{n\times n}$ 相抵于单位阵 I_n，换句话说，A 可以写作初等矩阵的乘积.

证明 对 n 进行归纳，当 $n=1$ 时命题显然成立. 假设 n 的情况成立，考虑 $n+1$ 的情况，设

$$A = \begin{pmatrix} a_{11} & \alpha^T \\ \beta & A' \end{pmatrix} \in \mathbb{F}^{(n+1)\times(n+1)} \tag{2.16}$$

不妨设 $a_{11} \neq 0$（否则可以通过交换使得 a_{11} 位置非 0），进一步也可以不妨设 $a_{11} = 1$（否则可以通过对第一列倍乘 a_{11}^{-1} 使之为 1），则可以通过依次加乘变换，使得

$$A = \begin{pmatrix} 1 & \alpha^T \\ \beta & A' \end{pmatrix} \sim \begin{pmatrix} 1 & \\ -\beta & I_n \end{pmatrix} \begin{pmatrix} a_{11} & \alpha^T \\ \beta & A' \end{pmatrix} \begin{pmatrix} 1 & -\alpha^T \\ & I_n \end{pmatrix} = \begin{pmatrix} 1 & \\ & A' - \beta\alpha^T \end{pmatrix} \tag{2.17}$$

根据条件，$A' - \beta\alpha^T \in \mathbb{F}^{N\times n}$ 可逆，且与 I_n 相抵，这说明 $A' - \beta\alpha^T = P_1 \cdots P_s$，其中 P_i 为初等矩阵，这表明

$$A \sim \begin{pmatrix} 1 & \\ & P_1 \end{pmatrix} \cdots \begin{pmatrix} 1 & \\ & P_s \end{pmatrix} \tag{2.18}$$

命题得证.

根据上面的命题可知，任何可逆矩阵都可写作初等阵的乘积，这说明可逆阵的左乘、右乘等效于数个初等阵进行左乘、右乘，因此可以对相抵进行非常重要的推广.

推论 2.2

A, B 相抵，当且仅当存在可逆矩阵 P, Q，使得 $PAQ = B$.

有了前面的铺垫，下面正式向相抵标准型推进，我们希望证明如下命题

定理 2.7 (相抵标准型)

设 $A \in \mathbb{F}^{m\times n}$，$\mathrm{rk}A = r$，则

$$A \sim \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}. \tag{2.19}$$

上面的矩阵被称为相抵标准型.

证明 证明可类比可逆情形进行归纳，或从向量组的考虑先构造 $r \times r$ 可逆块进行打洞.

可以看出，相抵标准型仅与矩阵的阶与秩有关，这也回答了前面的问题：同阶同秩矩阵必相抵.

定理 2.8

同阶矩阵相抵当且仅当其同秩.

上面结论也可以从基变换的角度解释，基变换顾名思义，就是线性空间中两组基之间的变换，如下定义.

定义 2.9 (基变换)

设 $V/\mathbb{F}$ 为有限维向量空间，$v_1,\cdots,v_n;w_1,\cdots,w_n$ 为 V 的两组基，则存在唯一可逆矩阵 $P\in\mathbb{F}^{n\times n}$ 使得

$$(w_1,\cdots,w_n)=(v_1,\cdots,v_n)P \tag{2.20}$$

其中 P 称为由 v 到 w 的基变换矩阵.

上面定义中的一些补充证明是简单的. 下面回到线性映射在的话题，考虑线性映射

$$\mathscr{A}(v_1,\cdots,v_n)=(w_1,\cdots,w_m)A \tag{2.21}$$

其中 $rkA=r$，则存在可逆矩阵 P,Q 使得 $PAQ=\begin{pmatrix}I_r & O\\ O & O\end{pmatrix}$，如果将 P,Q 分别作为 $\mathscr{A}$ 值域与定义域中的基变换矩阵，即设

$$(v_1,\cdots,v_n)Q=(v_1',\cdots,v_1'),\quad (w_1,\cdots,w_n)=(w_1',\cdots,w_m')P \tag{2.22}$$

则

$$\mathscr{A}(v_1',\cdots,v_n')Q^{-1}=(w_1',\cdots,w_m')PA \tag{2.23}$$

$$\mathscr{A}(v_1',\cdots,v_n')=(w_1',\cdots,w_m')PAQ \tag{2.24}$$

$$=(w_1',\cdots,w_m')\begin{pmatrix}I_r & O\\ O & O\end{pmatrix} \tag{2.25}$$

这说明，如果选择基 $v_1',\cdots,v_n';w_1',\cdots,w_m'$，则 $\mathscr{A}$ 在这对基下的矩阵恰好为 $\begin{pmatrix}I_r & O\\ O & O\end{pmatrix}$，这是一个形式非常简单的矩阵，此外还有更强的结论：对任何与 A 相抵的矩阵 B，都可以选择适当的基使得 $\mathscr{A}$ 在这组基下的矩阵为 B，并且不论如何如何选择基，$\mathscr{A}$ 在不同基下的矩阵总是相抵的，因此可以同一定义线性映射的秩

定义 2.10 (线性映射的秩)

定义 $\mathscr{A}\in\mathcal{L}(V,W)$ 的秩为 $\mathscr{A}$ 在任一对基下的矩阵的秩.

2.3.3 方阵的性质

方阵、变换的环、多项式代入同态，

2.3.4 线性变换与相似

前面讨论对象为一般的线性映射，如果考虑线性空间到自身的线性变换（当值域与定义域相等时称线性映射为线性变换），由于值域与定义域相同，因此只需取一组基 $v_1,\cdots,v_n$，$\mathscr{A}\in\mathcal{L}(V)$ 在这组基下有

$$\mathscr{A}(v_1,\cdots,v_n)=(v_1,\cdots,v_n)A \tag{2.26}$$

如果作基变换 $(w_1,\cdots,w_n)P=(v_1,\cdots,v_n)$，则

$$\mathscr{A}(w_1,\cdots,w_n)=(v_1,\cdots,w_n)PAP^{-1} \tag{2.27}$$

这里出现了一个相抵之上的，具有更高要求的矩阵关系：相似.

定义 2.11 (矩阵的相似)

设 $A,B\in\mathbb{F}^{n\times n}$，则称 A,B 相似，若存在可逆矩阵 $P\in\mathbb{F}^{n\times n}$ 使得 $PAP^{-1}=B$，记作 $A\overset{s}{\sim}B$.

很容易验证相似也是一个等价关系. 类比相抵标准型，自然会产生问题：相似变换下有哪些不变量？能否找到相似标准型？相似涉及到的问题要比相抵复杂得多，但也正是因为其复杂，使得我们可以从许多角度看待一些问题，也有许多工具来解决这些问题. 在此仅先讨论一些简单的性质.

命题 2.3

若 $A\overset{s}{\sim}B$，则 $\mathrm{rk}A=\mathrm{rk}B, \det A=\det B, \mathrm{tr}A=\mathrm{tr}B$.

根据上面的命题，可以定义线性变换的行列式/迹定义为其在任一组基下矩阵的行列式/迹，这也启发我们：若矩阵的某属性 P 是相似不变的，那么可以对线性变换定义类似的属性 P.

2.3.5 不变子空间

作为最后的铺垫，介绍不变子空间.

定义 2.12 (不变子空间)

设 $\mathscr{A}\in\mathcal{L}(V)$，$U\leqslant V$，若 $\mathscr{A}(U)\subset U$，即对任意 $u\in U$，$\mathscr{A}u\in U$，则称 U 为 $\mathscr{A}$-不变子空间.

顾名思义，不变子空间就是在某个线性变换下保持不变的子空间，比如 $V,\{0\}$ 就是两个平凡不变子空间，再比如 $\mathrm{Ker}\,\mathscr{A},\mathrm{Im}\,\mathscr{A}$ 也是两个不变子空间.

不变子空间具有较好的性质，比如

命题 2.4

取 V 基 $v_1,\cdots,v_n$，并设其中 $v_1,\cdots,v_k$ 张成了一个 $\mathscr{A}$-不变子空间，则 $\mathscr{A}$ 在这组基上的矩阵是分块上三角的，即

$$\mathscr{A}(v_1,\cdots,v_n)=(v_1,\cdots,v_n)\begin{pmatrix}A_{11} & A_{12}\\ & A_{22}\end{pmatrix}. \tag{2.28}$$

证明 计算每个 $\mathscr{A}v_i$ 易得.

上面的结论在映射的限制下，有更进一步的推论.

定义 2.13 (线性变换的限制)

设 $\mathscr{A}\in\mathcal{L}(V)$，$U\leqslant V$，则定义 $\mathscr{A}$ 在 U 上的限制映射为 $\mathscr{A}|_U\in\mathcal{L}(U)$，对任意 $u\in U\leqslant V$，$\mathscr{A}|_U u=\mathscr{A}u$.

注 上面的定义可以推广到一般的线性映射 (甚至对于一般的映射也都可以定义限制)，只不过在线性变换中使用较多，因此这里从变换的角度定义.

延续前面的记号，可以得到如下推论

推论 2.3

若 $U=\mathrm{Span}(v_1,\cdots,v_k)\leqslant V$ 为 $\mathscr{A}$-不变子空间，则 $\mathscr{A}|_U$ 在基 $v_1,\cdots,v_k$ 下的矩阵为 A_{11}.

若考虑商空间，又有如下推论

推论 2.4

若 $U=\mathrm{Span}(v_1,\cdots,v_k)\leqslant V$ 为 $\mathscr{A}$-不变子空间，定义 V/U 上的商变换 $\tilde{\mathscr{A}}\in\mathcal{L}(V/U)$，对任一 $\tilde{v}=v+U\in V/U$，$\tilde{\mathscr{A}}(\tilde{v})=\mathscr{A}v+U$，则商变换在基 $v_{k+1},\tilde{\cdots},v_n$ 下的矩阵为 A_{22}.

命题 2.5

取 V 基 $v_1, \cdots, v_n$，并设其中 $v_1, \cdots, v_k; v_{k+1}, \cdots, v_m$ 分别张成了一个 $\mathscr{A}$-不变子空间 U, W，即 $V = U \oplus W$，则 $\mathscr{A}$ 在这组基上的矩阵是分块对角的，即

$$\mathscr{A}(v_1, \cdots, v_n) = (v_1, \cdots, v_n) \begin{pmatrix} A_{11} & \\ & A_{22} \end{pmatrix}. \tag{2.29}$$

第 3 章 特征值理论

本节主要讨论线性代数中非常重要的特征值理论，我们首先讨论复向量空间中，最后一部分特别讨论实向量空间（因为 $\mathbb{C}$ 为代数闭域，具有更好的性质）.

3.1 线性变换与方阵的特征值

3.1.1 特征值与特征空间

上一章的最后提到了不变子空间，这里从最简单的不变子空间：一维不变子空间开始. 假设 $U = \mathrm{Span}(v) = \{kv : k \in \mathbb{F}\}$ 为 $\mathscr{A}$-不变子空间，则对任意 $kv \in U$，都有

$$\mathscr{A}(kv) = k\mathscr{A}(v) = \lambda(kv) \tag{3.1}$$

上述结果表明 $\mathscr{A}$ 在任何 $u \in U$ 上的变换都可以看作一个伸缩变换，具有这种性质的向量 u 与 λ 称为线性变换 $\mathscr{A}$ 的特征向量与特征值.

从矩阵的角度也可以看待特征值，延续上述记号取基 $v_1, \cdots, v_n$，设 v 在这组基下的坐标为 $x \in \mathbb{F}^{n\times 1}$，变换在这组基下的矩阵为 A，则 $\mathscr{A}v = \lambda v$ 等价于 $Ax = \lambda x$，这说明矩阵在这个列向量上左乘的作用等价于伸缩变换，或者说单位阵的倍数 λI 的左乘作用. 根据上面的讨论，可以定义线性变换与矩阵的特征值.

定义 3.1 (线性变换的特征值)

设 $T \in \mathcal{L}(V)$，若存在 $0 \neq v \in V, \lambda \in \mathbb{F}$ 使得 $Tv = \lambda v$，则称 λ 为 T 的特征值，v 是相应于 λ 的特征向量. 此外，定义线性变换 T 的特征值的全体称为 T 的谱，记为 $\mathrm{spec}\, T$. ♣

定义 3.2 (矩阵的特征值)

设 $A \in \mathbb{F}^{n\times n}$，若存在 $0 \neq x \in \mathbb{F}^{n\times 1}, \lambda \in \mathbb{F}$ 使得 $Av = \lambda x$，则称 λ 为 A 的特征值，v 是相应于 λ 的特征向量. 此外，定义矩阵 A 的特征值的全体称为 A 的谱，记为 $\mathrm{spec}\, A$. ♣

可以通过变换与矩阵两个角度看待特征值，但它们本质实际是相同的，考虑线性变换的特征值也可以等价于考虑其在某组基下矩阵的特征值，或者说可以考虑其在任一组基下矩阵的特征值.

命题 3.1

设 λ, v 为 $A \in \mathbb{F}^{n\times n}$ 的特征值与特征向量，若 $B \overset{s}{\sim} A$，则 λ, v 也是 B 的特征向量与特征值. ♠

上面的命题给出了相似变换新的不变量：特征值与特征向量.

从上面的讨论可以看出，相应于同一特征值的特征向量可以构成一个线性空间，并且这是一个不变子空间，由此可以定义将相应于同一特征值的特征向量合并的空间：特征空间.

定义 3.3 (特征空间)

设 $T \in \mathcal{L}(V), \lambda \in \mathrm{spec}\, T$，则称全体相应于 λ 的特征向量的全体为相应于 λ 的特征空间，记作 $E(\lambda, T)$. ♣

注

1. 对于矩阵也可以定义类似的特征空间，在此不做重复.
2. 由于零向量可以看作任何变换的，相应于任何 $\lambda \in \mathbb{F}$ 的特征值，但若 λ 不是特征值，则显然 $E(\lambda, T) = \{0\}$，但这样的特征空间是平凡的，因此可以不认为这是线性空间.

命题 3.2

设 $T \in \mathcal{L}(V), \lambda \in \operatorname{spec} T$，则 $E(\lambda, T)$ 为一个 T-不变子空间.

特征值是讨论“相似标准型”的一个切入点，考虑下例.

例 3.1 假设对于 $T \in \mathcal{L}(V)$，v_1, v_2 线性无关，且均为特征向量（不妨设各自相应于 λ_1, λ_2），则考虑 $T' = T|_{\operatorname{Span}(v_1,v_2)}$，可知

$$T'(v_1, v_2) = (v_1, v_2)\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \tag{3.2}$$

这时 T' 在基 v_1, v_2 下的矩阵为对角阵 $\operatorname{diag}(\lambda_1, \lambda_2)$.

上面的例子说明，以特征向量作为基，可以得到形式非常简单的对角阵，那么对任意 $T \in \mathcal{L}(V)$，是否存在 V 中一组由 T 的特征向量组成的基？这个问题等价于：V 是否能表示为特征空间的直和的形式？

命题 3.3

设 $T \in \mathcal{L}(V)$，$v_1, \cdots, v_m$ 是相应于不同特征值的特征向量，则 $v_1, \cdots, v_m$ 线性无关.

证明 设对应于每个 v_i 的特征值为 λ_i，并且 $f_i(\lambda) = (\lambda - \lambda_1) \cdots \widehat{(\lambda - \lambda_i)} \cdots (\lambda - \lambda_m)$. 假设 $a_1 v_1 + \cdots + a_m v_m = 0$，则用 $f_i(T)$ 依次作用于等式两边，可得 $a_1 = \cdots = a_m = 0$.

上述命题表明，所有特征空间之和确实是直和，但这样的直和是否等于 V，仍有待讨论，这个问题可以等同于讨论矩阵的相似型，也称可对角化问题.

为了更精确描述特征空间，可以采取另一种方法来描述一系列的特征 XX. 由于 $\mathscr{A} v = \lambda v$ 等价于 $(\mathscr{A} - \lambda I)v = 0$，即 $v \in \operatorname{Ker}(\mathscr{A} - \lambda I)$. 因此 $E(\lambda, \mathscr{A}) = \operatorname{Ker}(\mathscr{A} - \lambda I)$，在这种视角下，$\lambda$ 为特征值当且仅当 $\operatorname{Ker}(\mathscr{A} - \lambda I) \neq \{0\}$；类似的，$v$ 为相应于 λ 的特征值当且仅当 $0 \neq v \in \operatorname{Ker}(\mathscr{A} - \lambda I) \neq \{0\}$.

由此可以定义出特征空间的一个重要属性：几何重数

定义 3.4 (几何重数)

设 $T \in \mathcal{L}(V), \lambda \in \operatorname{spec} T$，则定义 λ_i 的几何重数 $m_i = \dim E(\lambda_i, T) = \dim \operatorname{Ker}(T - \lambda_i I)$.

使用维数定理，还能得到几何重数的一个计算方法：

$$m_i = \dim \operatorname{Ker}(T - \lambda_i I) = n - \dim \operatorname{rk}(T - \lambda_i I) \tag{3.3}$$

命题 3.4

设 $T \in \mathcal{L}(V), \lambda \in \operatorname{spec} T$，则对于任意多项式 $f \in \mathbb{F}[x]$，$f(\lambda)$ 为 $f(T)$ 的特征值.

3.1.2 特征多项式与代数重数

介绍特征值后，一个很自然的问题就是如何求出特征值，这种“实际”的问题就需要从“实际”的角度考虑，即矩阵. 首先考虑：对于 $\lambda \in \mathbb{F}$，如何判定它是否为矩阵的特征值？

仿照上一节的移项操作可知

$$\lambda \in \operatorname{spec} A \iff \exists x \neq 0 : (A - \lambda I)x = 0 \iff \operatorname{Ker}(A - \lambda I) \neq \{0\} \tag{3.4}$$

而这又等价于矩阵 $A - \lambda I$ 不可逆，即 $\det(A - \lambda I) = 0$，因此可以总结为如下命题.

命题 3.5

设 $A \in \mathbb{F}^{n \times n}$，则 $\lambda \in \operatorname{spec} A$ 当且仅当 $\det(A - \lambda I) = 0$.

也就是说，只需要根据计算行列式，就可以进行特征值的判定. 以此为基础，讨论 A 的特征值实际上就是在讨论 $g(\lambda) = \det(A - \lambda I)$ 的零点，通过行列式的定义可知，这个函数实际上是一个关于 λ 的多项式，这就引出

了特征多项式的定义.

定义 3.5 (特征多项式)

设 $A \in \mathbb{F}^{n\times n}$，则定义 A 的特征多项式为

$$\varphi_A(\lambda) = \det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0. \tag{3.5}$$

特征多项式是一个首一多项式，由于一开始就假设了 $\mathbb{F}$ 为代数闭域，因此它可以因式分解为一次多项式的乘积，在这种形式下，可以定义出每一项的重数：代数重数.

定义 3.6 (代数重数)

设 $A \in \mathbb{F}^{n\times n}$，$A$ 的特征多项式为

$$\varphi_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0 = (\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}\cdots(\lambda-\lambda_s)^{n_s}, \tag{3.6}$$

其中 $\lambda_1,\cdots,\lambda_s$ 为 A 的所有互异特征值，即 $\operatorname{spec} A = \{\lambda_1,\cdots,\lambda_s\}$，上面的 n_k 称为 λ_k 的代数重数.

特征值理论中有两种重数：几何重数与代数重数，它们的定义并不相同，而这两种重数可以看作是连接后面的广义特征空间、Schur 三角化理论的纽带，因此这里先仅考虑一些简单性质，二者的辨析留与后文说.

命题 3.6

设 $A \in \mathbb{F}^{n\times n}$，$\varphi_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0$，$\operatorname{spec} A = \{\lambda_1,\cdots,\lambda_s\}$，则

$$\operatorname{tr} A = \lambda_1 + \cdots + \lambda_s = a_{n-1} \tag{3.7}$$

$$\det A = \lambda_1\cdots\lambda_s = (-1)^n a_0 \tag{3.8}$$

下面的定理告诉了我们几何重数与代数重数的数量关系，同时也表明：借助特征空间来分解线性空间 V 的想法是不现实的.

定理 3.1

设 $A \in \mathbb{F}^{n\times n}$，则对任意 $\lambda_i \in \operatorname{spec} A$，都有 $m_i \leqslant n_i$，其中 m_i, n_i 分别是 λ_i 的几何重数与代数重数.

证明

3.1.3 Schur 三角化

直接讨论对角阵或许会有些突兀，暂且退而求其次，讨论上三角矩阵. 从计算角度考虑，上三角矩阵很容易计算行列式、迹、秩甚至特征值（从特征多项式来看，其对角元恰好是所有特征值）；从形式上来说，上三角矩阵具有明显的"递归"特性，若线性变换在某组基下成上三角阵，那么这种递归特性会使得这组基也有很好的性质.

命题 3.7

设 $T \in \mathcal{L}(V)$，则 T 在基 $v_1,\cdots,v_n$ 下为上三角矩阵与下述任一条等价.

1. 对任意 $j = 1,\cdots,n$，都有 $T(v_j) \in \operatorname{Span}(v_1,\cdots,v_j)$.
2. 对任意 $j = 1,\cdots,n$，$\operatorname{Span}(v_1,\cdots,v_j)$ 都是 T-不变子空间.

事实上，上面的结论可以直接将上三角矩阵分块成准上三角阵得到.

很容易证明，对于任一线性变换 $T \in \mathcal{L}(V)$，存在 V 的一组基使得 T 在这组基下矩阵为上三角矩阵.

定理 3.2

设 $T \in \mathcal{L}(V)$，则存在 V 的基 $v_1,\cdots,v_n$，使得 T 在这组基下的矩阵为上三角阵.

证明 对维数进行归纳，当 $\dim V = 1$ 时显然成立，假设命题对 $\dim V < n$ 的维线性空间成立，考虑 n 维. 取 T 的特征值 λ 与特征向量 v，则有 $T(v) = \lambda v$，设 $U = \mathrm{Im}\,(T - \lambda I)$，则 $\dim U < n$，并且 U 在 T 下不变，因此 $T|_U$ 是 U 上的线性变换，根据归纳假设，存在 U 的基 $u_1, \cdots, u_m$ 使得 $T|_U$ 在该基下为上三角阵，将其扩充为 V 的基 $u_1, \cdots, u_m, u_{m+1}, \cdots, u_n$，则对任意 $k = m+1, \cdots, n$ 有

$$Tu_k = (T - \lambda I)u_k + \lambda u_k \in U + \mathrm{Span}(u_k) \subset \mathrm{Span}(u_1, \cdots, u_k), \tag{3.9}$$

根据前面的命题可知，T 在这组基下矩阵为上三角矩阵.

上面的结果也可以用矩阵语言表述，这称为 Schur 三角化.

定理 3.3 (Schur 三角化)

设 $A \in \mathbb{F}^{n\times n}$，则存在可逆矩阵 P，使得 $A = PTP^{-1}$，其中 T 为上三角矩阵. 更准确来说，设 A 的可重特征值为 $\lambda_1, \cdots, \lambda_n$，则 T 的对角元为 $\lambda_1, \cdots, \lambda_n$. 特别地，如果将相同特征值排列在一起（设有 d 个相异特征值），则

$$T = \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1d} \\ 0 & T_{22} & \cdots & T_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_{dd} \end{pmatrix} \tag{3.10}$$

其中 $T_{ii} \in \mathbb{F}^{n_i\times n_i}$ 均为上三角矩阵，且对角元为（重新排列后的）λ_i，n_i 为对应的代数重数. ♡

注 事实上，上面的结论是弱化版的 Schur 三角化，其中的相似可以加强到酉相似（复正交相似），涉及一些内积理论，因此下面贴出更强版本的证明，待日后调整.

证明 对 n 归纳，当 $n = 1$ 命题显然成立，假设命题对 $n \leqslant k$ 阶矩阵都成立，考虑 $n = k+1$ 情形.

对于 λ_1，取 $\boldsymbol{x} \in E(\lambda_1, A)$，且 $|\boldsymbol{x}| = 1$，则存在酉矩阵 $V = (\boldsymbol{x}, V_2)$，计算得

$$AV = \big(A\boldsymbol{x}, AV_2\big) = \big(\lambda_1\boldsymbol{x}, AV_2\big) \tag{3.11}$$

$$V^*AV = \begin{pmatrix} \boldsymbol{x}^* \\ v_2^* \end{pmatrix} \big(\lambda_1\boldsymbol{x}, AV_2\big) = \begin{pmatrix} \lambda_1 & * \\ 0 & A' \end{pmatrix}, \tag{3.12}$$

其中 A' 特征值与 A 的对应块相同，根据归纳假设，存在酉矩阵 W，使得 $A' = WT'W^*$，T 对角元为 $\lambda_2, \cdots, \lambda_n$. 因此取

$$U = V\begin{pmatrix} 1 & 0 \\ 0 & W^* \end{pmatrix}, \tag{3.13}$$

易验证 $A = UTU^*$，上三角阵 T 对角元为 $\lambda_1, \cdots, \lambda_n$，命题得证.

3.1.4 Caylay-Hamilton 定理

前面定义了特征多项式，因此可以顺便定义两个与之类似的概念，它们可以看作特征多项式“向上”与“向下”两个方向的拓展.

定义 3.7 (零化多项式)

设 $A \in \mathbb{F}^{n\times n}$，若首一多项式 $f \in \mathbb{F}[x]$ 使得 $f(A) = O$，则称 f 为 A 的零化多项式. ♣

定义 3.8 (极小多项式)

设 $A \in \mathbb{F}^{n\times n}$，定义 A 的次数最低的零化多项式为 A 的极小多项式，记为 d_A. ♣

注 由于 A 的零化多项式的全体是 $\mathbb{F}$ 上多项式环的一个理想，而 $\mathbb{F}[x]$ 为主理想整环，因此存在唯一 $p(x) \in \mathbb{F}[x]$，使得 A 的零化多项式恰好为 $p(x)$ 生成的理想 $(p(x))$，若要求 p 首一，则这时的 p 就是极小多项式；或者从唯一分解整环的角度看，极小多项式可以定义为全体零化多项式的最大公因子.

极小多项式的唯一性可以通过带余除法保证，因此重要的是考虑其存在性. 一种基本方法是从矩阵空间的张成考虑，存在零化多项式等价于存在 $k \in \mathbb{N}$，$I, A, \cdots, A^k$ 线性相关，由于 $\mathbb{F}^{n\times n}$ 是一个 $\mathbb{F}$ 上的线性空间，并且 $\mathrm{Span}(I, A, \cdots, A^k, \cdots) \leqslant \mathbb{F}^{n\times n}$，因此取 $k = \dim \mathbb{F}^{n\times n} = n^2$，则必有 $A^{n^2} \in \mathrm{Span}(I, A, \cdots, A^{n^2-1})$（否则与 $\mathbb{F}^{n\times n}$ 维数矛盾），故存在性得证.

根据上面的观点，对于矩阵的幂次张成的空间也可用多项式表示为

$$\mathrm{Span}(I, A, \cdots, A^k, \cdots) = \{f(A) : f \in \mathbb{F}[x]\}. \tag{3.14}$$

上面的讨论表明矩阵 A 存在一个次数至多为 n^2 的零化多项式，但下面的定理告诉我们，n^2 可以被降到 n，因为特征多项式本身就是一个零化多项式.

定理 3.4 (Cayley-Hamilton 定理)

设 $A \in \mathbb{F}^{n\times n}$，$\varphi_A$ 为 A 的特征多项式，则 $\varphi_A(A) = O$.

证明 根据矩阵的三角化，存在可逆矩阵 $P = (p_1, \cdots, p_n)$，使得 PAP^{-1} 为上三角阵，设对角元为 $\lambda_1, \cdots, \lambda_n$.

设 $V_k = \mathrm{Span}(p_1, \cdots, p_{n-k}), k = 0, \cdots, n$，则可知 $\mathbb{R}^{n\times 1} = V_0 \supset \cdots \supset V_n = \{0\}$，并且根据上三角阵的结构有

$$(A - \lambda_{n-k} I)p_{n-k} \in \mathrm{Span}(p_1, \cdots, p_{n-k-1}) = V_{k+1}, \quad (A - \lambda_{n-k} I)V_k \subset V_{k+1} \tag{3.15}$$

由此可得

$$\begin{aligned}
\varphi_A(A)(\mathbb{F}^{n\times 1}) &= (A - \lambda_1 I) \cdots (A - \lambda_n I)V_0 &(3.16)\\
&\subset (A - \lambda_1 I) \cdots (A - \lambda_{n-1} I)V_1 &(3.17)\\
&\subset \cdots &(3.18)\\
&\subset (A - \lambda_1 I)V_{n-1} \subset V_n = \{0\} &(3.19)
\end{aligned}$$

命题得证.

事实上，极小多项式也包含了许多信息，容易验证，特征值必然是极小多项式的根，在此基础上根据极小多项式的整除关系，可知极小多项式的零点集就是特征值的集合（不计重数）.

命题 3.8

设 $A \in \mathbb{F}^{n\times n}$，则对 A 的任何特征值 λ_k 都有 $d_A(\lambda_k) = 0$.

证明 假设 v 为相应于 λ 的特征向量，则有 $T^j v = \lambda^j v$，因此

$$0 = d_A(A)v = d(\lambda)v \tag{3.20}$$

故 $d_A(\lambda) = 0$，得证.

3.2 广义特征空间

本节延续前面对空间分解的讨论，即希望将 V 分解为特征空间的想法，首先介绍特征空间的推广：广义特征空间.

3.2.1 幂的零空间链

线性变换可以进行复合，设 $T \in \mathcal{L}(V)$，由于 $T^k v = 0$ 蕴含 $T^{k+1} v = 0$，因此可以发现线性变换幂的零空间有包含关系，但由于 V 为有限维的，因此这个链条必然会在某一刻中止，因此有如下结论.

命题 3.9

设 $T \in \mathcal{L}(V)$，则

1. $\{0\} = \operatorname{Ker} T^0 \subseteq \operatorname{Ker} T^1 \subseteq \cdots \subseteq \operatorname{Ker} T^k \subseteq \operatorname{Ker} T^{k+1} \subseteq \cdots$.
2. 若存在 $m \in \mathbb{N}$，满足 $\operatorname{Ker} T^m = \operatorname{Ker} T^{m+1}$，则对于任意 $k \in \mathbb{N}$，都有 $\operatorname{Ker} T^m = \operatorname{Ker} T^{m+k}$.
3. 上面提到的 m 必定存在，且 $m \leqslant n = \dim V$.

证明

1. 若 $v \in \operatorname{Ker} T^m$，则 $T^m v = 0$，$T^{m+1} v = T(T^m v) = 0$，故 $v \in \operatorname{Ker} T^{m+1}$，即 $\operatorname{Ker} T^m \subseteq \operatorname{Ker} T^{m+1}$.
2. 只需证明 $\operatorname{Ker} T^{m+k} = \operatorname{Ker} T^{m+k+1}$，而根据 1，只需证明 $\operatorname{Ker} T^{m+k} \supseteq \operatorname{Ker} T^{m+k+1}$.
   设 $v \in \operatorname{Ker} T^{m+k+1}$，即 $T^{m+k+1} v = T^{m+1}(T^k v) = 0$，因此有 $T^k v \in \operatorname{Ker} T^{m+1} = \operatorname{Ker} T^m$，故 $T^m(T^k v) = T^{m+k} v = 0$，因此 $\operatorname{Ker} T^{m+k} \supseteq \operatorname{Ker} T^{m+k+1}$，得证.
3. 只需证明，必有 $\operatorname{Ker} T^n = \operatorname{Ker} T^{n+1}$. 若不然，则对某个 V, T 如题所述，有
$$\{0\} = \operatorname{Ker} T^0 \subsetneq \operatorname{Ker} T^1 \subsetneq \cdots \subsetneq \operatorname{Ker} T^{n-1} \subsetneq \operatorname{Ker} T^n \subsetneq \operatorname{Ker} T^{n+1}. \tag{3.21}$$
   这表明
$$\begin{aligned} \dim \operatorname{Ker} T^{n+1} &\geqslant \dim \operatorname{Ker} T^n + 1 \geqslant \dim \operatorname{Ker} T^{n-1} + 2 && (3.22)\\ &\geqslant \cdots && (3.23)\\ &\geqslant \dim \operatorname{Ker} T + n \geqslant n + 1 && (3.24) \end{aligned}$$
   这与 $\operatorname{Ker} T^{n+1} \subseteq V$，$\dim \operatorname{Ker} T^{n+1} \leqslant n$，矛盾.

对于线性变换使用维数定理可以得到
$$\dim V = \dim \operatorname{Ker} T + \dim \operatorname{Im} T, \tag{3.25}$$
因此借助维数定理，可以对上面的结论给出一个 $\operatorname{Im} T$ 的降链版本.

如果对维数定理敏感，就会猜测：能否通过 $\operatorname{Ker} T$ 与 $\operatorname{Im} T$ 对空间进行分解？答案是否定的，虽然有维数上对应，但这两个空间的和不是直和（或者说，可能存在 $v \neq 0$，使得 $T^2 v = 0$）. 但借助根据上面的命题，可以得到 Fitting 引理，它给出了空间 V 的一个粗糙分解.

定理 3.5 (Fitting 引理)

设 $V/\mathbb{F}$，$\dim V = n < \infty$，$T \in \mathcal{L}(V)$，则
$$V = \operatorname{Ker} T^n \oplus \operatorname{Im} T^n. \tag{3.26}$$

证明 根据维数定理，只需证明 $\operatorname{Ker} T^n \cap \operatorname{Im} T^n = \{0\}$.

设 $v \in \operatorname{Ker} T^n \cap \operatorname{Im} T^n$，即
$$T^n v = 0 \tag{3.27}$$
$$\exists u \in V, v = T^n u \tag{3.28}$$
根据上面两点，可知 $T^n v = T^{2n} u = 0$，即 $u \in \operatorname{Ker} T^{2n} = \operatorname{Ker} T^n$，故 $v = T^n u = 0$，命题得证.

下面介绍一种特殊的变换，它也具有迭代一定次数后“停止”的性质.

定义 3.9 (幂零变换)

设 $N \in \mathcal{L}(V)$，若存在 $k \in \mathbb{N}$，使得 $N^k = O$，则称 T 为幂零变换.

注 同理可定义幂零矩阵.

同样称为幂零，容易验证幂零算子与幂零矩阵直接存在直接的联系.

命题 3.10

设 $N \in \mathcal{L}(V)$ 为幂零的，当且仅当存在 V 的一组基，使得 N 在这组基下的矩阵为幂零阵.

设 $N \in \mathcal{L}(V)$ 幂零，则最小的使 $N^k = O$ 的正整数称为幂零指数，根据上面的结论就有

$$\{0\} = \operatorname{Ker} N^0 \subsetneq \operatorname{Ker} N^1 \subsetneq \cdots \subsetneq \operatorname{Ker} N^k = V \tag{3.29}$$

也就是说，$\operatorname{Ker} N^p$ 在 N 增长到 k 时就可以包含空间 V，根据 Fitting 引理也可以得到 $V = \operatorname{Ker} N^k$.

幂零变换还有一些重要的性质，易知，$f(x) = x^p, p \geqslant k$ 都是 N 的零化多项式，因此幂零变换仅有零特征值. 假设 $\dim V = n$，则 $\varphi_N(\lambda) = \lambda^n, d_N(\lambda) = \lambda^k$

在此基础上，借助 Schur 三角化可以得到如下推论

推论 3.1

设 $N \in \mathcal{L}(V)$ 幂零，则存在 V 的一组基，使得 N 在这组基下的矩阵为上三角阵，且对角元均为 0.

3.2.2 广义特征空间

一方面，代数重数与几何重数的关系否定了使用特征空间分解一个线性空间 V 的想法；另一方面，Fitting 引理告诉我们线性变换幂的核与像可以分解一个空间，因此我们自然会考虑线性变换的幂，推广一般特征空间.

定义 3.10 (广义特征向量/空间)

设 $T \in \mathcal{L}(V)$，λ 为 T 的特征值，向量 $0 \neq v \in V$ 称为相应于特征值 λ 的广义特征向量，若存在 $j \in \mathbb{N}$ 使得

$$(T - \lambda I)^j v = 0. \tag{3.30}$$

所有相应于 λ 的广义特征向量张成的空间称为 T 的相应于 λ 的广义特征空间，记为 $G(\lambda, T)$.

根据定义，$v \in G(\lambda, T)$ 当且仅当存在 $j \in \mathbb{N}$，使得 $v \in \operatorname{Ker}(T - \lambda I)^j$，因此广义特征空间可写作

$$G(\lambda, T) = \bigcup_{j=1}^{\infty} \operatorname{Ker}(T - \lambda I)^j \tag{3.31}$$

根据零空间链的"单调有界"性，可以得到如下命题

命题 3.11

设 $T \in \mathcal{L}(V), \dim V = n < \infty$，$\lambda$ 为 T 特征值，则

$$G(\lambda, T) = \operatorname{Ker}(T - \lambda I)^n. \tag{3.32}$$

证明 显然有 $G(\lambda, T) \supset \operatorname{Ker}(T - \lambda I)^n$. 由于对任意 $k \in \mathbb{N}$ 都有 $\operatorname{Ker}(T - \lambda I)^{n+k} = \operatorname{Ker}(T - \lambda I)^{n+k+1}$，故

$$G(\lambda, T) = \bigcup_{j=1}^{\infty} \operatorname{Ker}(T - \lambda I)^j = \bigcup_{j=1}^{n} \operatorname{Ker}(T - \lambda I)^j \tag{3.33}$$

$$\subseteq \operatorname{Ker}(T - \lambda I)^n. \tag{3.34}$$

得证.

通过这种表示，可以看出特征空间与广义特征空间之间的关系

$$E(\lambda, T) = \operatorname{Ker}(T - \lambda I) \subset \quad G(\lambda, T) = \operatorname{Ker}(T - \lambda I)^n \tag{3.35}$$

这直接表明了从特征空间到广义特征空间的推广.

定义广义特征空间后，可以正式开始研究其性质. 首先可以看出，广义特征空间继承了特征空间的一些性质.

命题 3.12

设 $T \in \mathcal{L}(V)$，$\operatorname{spec} T = \{\lambda_1, \cdots, \lambda_s\}$，则

1. 对任意 $\lambda \in \operatorname{spec} T$，$G(\lambda, T)$ 在 T 下是不变的.
2. 设 $v_1, \cdots, v_s$ 为相应的广义特征向量，则 $v_1, \cdots, v_s$ 线性无关.

证明

1.
$$T(G(\lambda, T)) = (T - \lambda I)(G(\lambda, T)) + \lambda G(\lambda, T) \leqslant G(\lambda, T). \tag{3.36}$$

2. 设 $v_i \in G(\lambda_i, T)(i = 1, \cdots, m)$，$0 = a_1 v_1 + \cdots + a_m v_m$，设 k_1 是使得 $(T - \lambda_1 I)^k \neq 0$ 的最大 k，考虑多项式 $f_1(\lambda) = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^n \cdots (\lambda - \lambda_s)^n$，则
$$0 = f_1(T)(a_1 v_1 + \cdots + a_m v_m) \tag{3.37}$$
$$= a_1 f_1(T)(v_1) + \cdots + a_m f_1(T)(v_m) \tag{3.38}$$
$$= a_1 f_1(T)(v_1) \tag{3.39}$$
可得 $a_1 = 0$. 重复上述操作，则有 $a_1 = \cdots = a_m = 0$，得证.

进一步，很容易得到这样的推论.

命题 3.13

设 $T \in \mathcal{L}(V)$，则 $G(\lambda, T)$ 是 $(T - \lambda I)$-不变子空间，并且 $(T - \lambda I)|_{G(\lambda,T)}$ 是幂零变换.

有了这些准备，下面进入正片.

3.2.3 广义特征分解与代数重数

前面的结果告诉我们 $G(\lambda, T) = \operatorname{Ker}(T - \lambda I)^n$，根据 Fitting 引理可得
$$V = \operatorname{Ker}(T - \lambda I)^n \oplus \operatorname{Im}(T - \lambda I) = G(\lambda, T) \oplus U, \tag{3.40}$$
注意到，T 在 U 中的特征值为 $\operatorname{spec} T \backslash \{\lambda\}$，因此可以不断分拆得到 V 的一个广义特征空间的分解，即如下定理.

定理 3.6

设 $T \in \mathcal{L}(V)$，$\lambda_1, \cdots, \lambda_m$ 为 T 的所有相异特征值，则
$$V = G(\lambda_1, T) \oplus G(\lambda_2, T) \oplus \cdots \oplus G(\lambda_m, T). \tag{3.41}$$

证明 对维数归纳，当 $\dim V = 1$ 时命题显然成立. 假设命题对 $\dim V < n$ 都成立，则考虑维数为 $n+1$ 的情况，根据 Fitting 引理，有
$$V = \operatorname{Ker}(T - \lambda_1 I)^n \oplus \operatorname{Im}(T - \lambda_1 I)^n \tag{3.42}$$
$$= G(\lambda_1, T) \oplus U \tag{3.43}$$
考虑限制映射 $T|_U$，可知 $\operatorname{spec} T|_U = \{\lambda_2, \lambda_3, \cdots, \lambda_m\}$，根据归纳假设，有
$$U = G(\lambda_2, T|_U) \oplus \cdots \oplus G(\lambda_2, T|_U) \tag{3.44}$$
因此只需证明对每个 $\lambda_i(i = 2, \cdots, m)$，有
$$G(\lambda_i, T|_U) = G(\lambda_i, T). \tag{3.45}$$
显然有 $G(\lambda_i, T|_U) \subseteq G(\lambda_i, T)$，故只需证 $G(\lambda_i, T|_U) \supseteq G(\lambda_i, T)$，不失一般性，仅考虑 λ_2.

任取 $v \in G(\lambda_2, T)$，根据 Fitting 引理可设 $v = v_1 + u$, 其中 $v_1 \in G(\lambda_1, T), u \in U$，再设

$$u = v_2 + v_3 + \cdots + v_m. \tag{3.46}$$

其中每个 $v_i \in G(\lambda_i, T|_U)$，故

$$v = v_1 + v_2 + \cdots + v_m. \tag{3.47}$$

根据广义特征向量的线性无关性，上式中除 v_2 外均有 $v_i = 0$，因此 $u = 0, v = v_2 \in G(\lambda_2, T)$，包含关系得证.

综上，命题得证.

上面的定理非常重要，它表明借助推广后的广义特征空间，就可以恰好将 V 分解. 从矩阵角度看，它给出了一个比上三角化更强的结论——分块对角化.

定理 3.7 (分块对角化)

设 $T \in \mathcal{L}(V)$，$\operatorname{spec} A = \{\lambda_1, \cdots, \lambda_m\}$，则存在 V 的一组由 T 的广义特征向量构成的基，使得 T 在这组基下的矩阵为分块对角矩阵，即

$$A = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_m \end{pmatrix} \tag{3.48}$$

且其中每个 A_i 均为上三角矩阵，对角元全为 λ_i.

借助上面的定理，可以揭示几何重数与代数重数之间的美妙联系. 上面的定理告诉我们，任何矩阵都可以相似到分块对角矩阵，并且每个块都是以特征值为对角元的上三角阵，如果我们设 $A_i \in \mathbb{F}^{n_i \times n_i}$，可以计算出特征多项式

$$\varphi_A(\lambda) = (\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_m)^{n_m} \tag{3.49}$$

如果与特征多项式的定义进行对比，可以发现 A_i 块的边长 n_i 恰好是 λ_i 的代数重数. 不止于此，上面分块对角化的本质实际上是广义特征空间的分解，因此可以得到

$$n_i = \dim G(\lambda_i, T). \tag{3.50}$$

将几何重数与代数重数进行对比可得

命题 3.14

设 $T \in \mathcal{L}(V)$，$\operatorname{spec} T = \{\lambda_1, \cdots, \lambda_m\}$，设 m_i, n_i 分别对应 λ_i 的几何、代数重数，则

- λ_i 的几何重数对应特征空间的维数，即 $m_i \dim E(\lambda_i, T)$.
- λ_i 的代数重数对应广义特征空间的维数，即 $n_i \dim E(\lambda_i, T)$.

这也给出了另一种更简洁、不借助特征多项式定义代数重数的方法，特别在这种定义下，几何重数小于代数重数是显然的.

在前面的定义中 $G(\lambda_i, T) = \operatorname{Ker}(T - \lambda I)^n$，利用代数重数可以简化为 $G(\lambda_i, T) = \operatorname{Ker}(T - \lambda I)^{n_i}$，但这并不代表 n_i 就是使得链条终止的最小正整数.

3.3 矩阵的对角化

本节从 Schur 三角化开始，侧重矩阵相似来讨论分块对角化.

3.3.1 分块对角化

许多内涵在上一节中已经讨论足够了，因此我们主要测重新的证明，所使用的方法还是“打洞”，首先证明一个引理.

引理 3.1 (Sylvester 定理)

设 $A \in \mathbb{F}^{m\times m}, B \in \mathbb{F}^{n\times n}$，且 $\operatorname{spec} A \cap \operatorname{spec} B = \emptyset$，则对任意 $C \in \mathbb{F}^{m\times n}$，矩阵方程

$$AX - XB = C \tag{3.51}$$

有唯一解 $X \in \mathbb{F}^{m\times n}$.

该引理有两种证明方法，都从考虑线性算子出发，一种是常规讨论其单性与满性，另一种是直接写出该算子在一组基下的矩阵，根据其特征值来判断（也是我个人偏爱的证法），这里只考虑前者.

证明 定义线性算子

$$T : \mathbb{F}^{m\times n} \longrightarrow \mathbb{F}^{m\times n} \tag{3.52}$$

$$X \longmapsto AX - XB, \tag{3.53}$$

则原命题等价于证明 T 是双射，由于线性变换的单满等价，故只需证明 T 是单的.

若 $AX = XB$，则对任意 $n \in \mathbb{N}$，都有 $A^n X = XB^n$，同理也有 $\varphi_B(\lambda)(A)X = X\varphi_B(\lambda)(B) = 0$. 根据条件，由于 $\operatorname{spec} A \cap \operatorname{spec} B = \emptyset$，故 $\varphi_B(A) \neq O$，因此必有 $X = 0$，即 $\operatorname{Ker} T = \{0\}$，单性得证，命题得证.

下面证明矩阵的分块对角化定理.

定理 3.8

设 $A \in \mathbb{F}^{n\times n}$，$\operatorname{spec} A = \{\lambda_1, \cdots, \lambda_d\}$，$\lambda_i$ 对应代数重数为 n_i，则即存在可逆矩阵 $P \in GL_n(\mathbb{F})$，准对角阵 $T = \operatorname{diag}(T_{11}, \cdots, T_{dd})$，满足

$$A = PTP^{-1}. \tag{3.54}$$

其中 $T_{ii} \in \mathbb{F}^{n_i\times n_i}$ 为上三角阵，且对角元均为 λ_i.

证明 我们基于 Schur 三角化得到的上三角阵讨论该定理. 对 d 归纳，当 $d = 1$ 时命题显然成立，设对 $d \geqslant 2, d-1$ 命题成立，考虑 d. 设

$$T = \begin{pmatrix} T_{11} & C \\ 0 & T' \end{pmatrix}, C = \begin{pmatrix} T_{12} & \cdots & T_{1d} \end{pmatrix}, \tag{3.55}$$

根据 Sylvester 定理，存在唯一 X，使得 $T_{11}X - XT' = C$，则

$$T \overset{s}{\sim} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} T_{11} & C \\ 0 & T' \end{pmatrix} \begin{pmatrix} I & -X \\ 0 & I \end{pmatrix} = \begin{pmatrix} T_{11} & C - T_{11}X + XT' \\ 0 & T' \end{pmatrix} = \begin{pmatrix} T_{11} & 0 \\ 0 & T' \end{pmatrix}, \tag{3.56}$$

根据归纳假设，有可逆阵 S，使得 $ST'S^{-1} = \operatorname{diag}(T_{22}, \cdots, T_{dd})$，因此使用矩阵 $\operatorname{diag}(I, S)$，可将 T 相似到所需形式，命题得证.

我们证明了，并非所有的方阵都能对角化，但所有的方阵都能分块对角化，自然会有进一步的问题：附加哪些条件，可以使得矩阵可对角化？下面简单介绍一个条件.

命题 3.15

设 $A \in \mathbb{F}^{n\times n}$，$\operatorname{spec} A = \{\lambda_1, \cdots, \lambda_d\}$，则 A 可对角化的充要条件是极小多项式 $d_A(\lambda)$ 无重根，即

$$d_A(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_d). \tag{3.57}$$

证明 必要性显然，下面只证明必要性. 设 $d_A(\lambda)$ 无重根，并不妨设 $A = \operatorname{diag}(A_1, \cdots, A_d)$，则

$$0 = d_A(A) = \operatorname{diag}(d_A(A_1), \cdots, d_A(A_d)), \tag{3.58}$$

即对每个 A_i，都有

$$0 = d_A(A_i) = (A_i - \lambda_i I)\prod_{j\neq i}(A_i - \lambda_j I) = (A_i - \lambda_i I)B \tag{3.59}$$

由于 B 对角元都不为 0，故可逆，即得 $A_i = \lambda_i I$，这表明 A 可对角化.
可对角化的讨论

3.3.2 对角化与交换性

抄 second course 内容

第 4 章 Jordan 标准型及其应用

本章延续前面的话题，讨论“相似标准型”问题的一个顶点：Jordan 标准型理论，以及一些使用的例子（比如线性微分方程组）.

首先延续上一章的讨论，下面从空间分解与矩阵相似型的路径给出 Jordan 标准型.

4.1 广义特征空间的进一步分解

上一节已经得到了一个重要结果

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T) \tag{4.1}$$

这已经是一个令人满意的结果，唯一美中不足的是，在我们看来每个 $G(\lambda_i, T)$ 仍然是一个“黑箱”. 若希望寻找更进一步的结论，则需要对每个广义子空间继续分解. 这时我们的研究对象可以得到约化，即考虑在 V 上具有唯一特征值 λ 的线性变换（也称单谱算子）.

4.1.1 循环子空间

在许多代数结构中，都会涉及到生成的概念，比如由一个元素生成的群、理想等. 对于线性空间也有类似的考虑，借助 V 中一个向量以及 V 上的一个线性变换可以定义出这种生成空间，由于这种生成往往具有循环的特性，因此也被称为循环子空间.

定义 4.1 (循环子空间)

设 $v \in V, T \in \mathcal{L}(V)$，则称 v 生成的 T-循环子空间为

$$\langle v \rangle_T = \mathrm{Span}(v, Tv, T^2v, \cdots, T^kv, \cdots). \tag{4.2}$$

♣

对于循环子空间，很容易验证下面的命题成立

命题 4.1

设 $v \in V, T \in \mathcal{L}(V)$，则

1. $\langle v \rangle_T$ 是 V 的子空间.
2. $\langle v \rangle_T$ 是 T-不变子空间.

♠

可以看出 $\langle v \rangle_T$ 中的向量形如

$$a_0v + a_1Tv + \cdots + a_kT^kv = (a_0 + a_1T + \cdots + a_kT^k)v \tag{4.3}$$

这实际上等价于一个代入 T 的多项式作用在 v 上的结果，因此可以证明

命题 4.2

设 $v \in V, T \in \mathcal{L}(V)$，则

$$\langle v \rangle_T = \{f(T)v : f \in \mathbb{F}[x]\}. \tag{4.4}$$

♠

上面的操作在上一章中讨论极小多项式存在性时遇到过，因此很容易看出这里也可以定义与零化多项式、极小多项式类似的概念.

定义 4.2 (向量的极小多项式)

设 $v \in V, T \in \mathcal{L}(V)$，对于多项式 $f \in \mathbb{F}$，若 $f(T)v = 0$，则称 f 为 v 关于 T 的零化多项式；称次数最小的首一零化多项式为 v 关于 T 的极小多项式，记为 $d_{T,v}$.

同样有极小多项式的名字，容易发现向量关于变换 T 的极小多项式与变换 T 的极小多项式有一定联系.

命题 4.3

设 $v \in V, T \in \mathcal{L}(V)$，则 $d_{T,v} | d_T$，并且对任意 v 关于 T 的零化多项式 f，都有 $d_{T,v} | f$.

易知，若 V 是有限维的，那么显然 $\langle v \rangle_T$ 也是有限维的，并且

$$v, Tv, \cdots, T^{m-1}v \tag{4.5}$$

是 $\langle v \rangle_T$ 的一组基，其中 $m = \dim\langle v \rangle_T$. 由此也可得 $\deg d_{T,v} = \dim\langle v \rangle_T$.

有了不变子空间的一组基，下一步自然会考虑算子 T 在其上的矩阵，则有

命题 4.4

设 $v \in V, T \in \mathcal{L}(V), \dim\langle v \rangle_T = n, d_{T,v}(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0$. 则 $T|_{\langle v \rangle_T}$ 在基 $v, Tv, T^2v, \cdots, T^{n-1}v$ 下的矩阵为

$$\begin{pmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & \ddots & & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{pmatrix} \tag{4.6}$$

注 可以看出，上面的矩阵仅与 $d_{T,v}$ 的各项系数有关，由此这样的矩阵可以对于任何首一多项式 f 定义，称为 f 的**友矩阵**.

反过来，如果计算友矩阵 A 的特征多项式与极小多项式，可以发现 $\varphi_A = d_A = d_{T,v}$.

命题 4.5

设 $v \in V, T \in \mathcal{L}(V)$，并设 $S = T|_{\langle v \rangle_T}$，则 $\varphi_S(\lambda) = d_S(\lambda) = d_{S,v}(\lambda)$.

证明 计算可得 $\varphi_S(\lambda) = d_{S,v}(\lambda)$，再根据 $d_{S,v}(\lambda) | d_{\mathscr{A}}(\lambda) | \varphi_S(\lambda)$，得证.

特别的，如果限制映射 $T|_{\langle v \rangle_T}$ 幂零，或者说 $\varphi_{T,v}(\lambda) = \lambda^n$，那么 T 的友矩阵恰好为

$$\begin{pmatrix} 0 & & & & \\ 1 & 0 & & & \\ & 1 & \ddots & & \\ & & \ddots & 0 & \\ & & & 1 & 0 \end{pmatrix} \tag{4.7}$$

回到一开始的问题，考虑空间 V 以及幂零算子 $T \in (V)$，如果 V 可以分解为 T-循环子空间的直和，那么 T 在对应的一组基下的矩阵恰好是多个友矩阵的直和. 在此基础上，对于一般的单谱算子 $T \in \mathcal{L}(V)$，设特征值为 λ_1，则 $T = (T - \lambda_1 I) + \lambda_1 I$，其中 $T - \lambda_1 I$ 幂零，那么 T 在同样的基下的矩阵同样是许多矩阵的直和，通过改

变基的顺序，其中每个矩阵都有形式

$$\begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} \tag{4.8}$$

这样的矩阵被称为一个 Jordan 块.

定义 4.3 (Jordan 块)

定义具有如下形式的矩阵 $J_m(\lambda) \in \mathbb{F}^{m\times m}$ 为一个 Jordan 块

$$J_m(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} \tag{4.9}$$

♣

其名称已经暗示，这与 Jordan 标准型有关.

4.1.2 Jordan 标准型

仿照一开始借助广义特征空间分解线性空间的过程，我们希望证明这样的引理

引理 4.1

设 $T \in \mathcal{L}(V)$ 为 m 次幂零变换，则

1. 则存在 $v \in V$，使得 $d_{T,v}(\lambda) = \lambda^m$.
2. 在此基础上，存在 T-不变子空间 V_1，满足
$$V = \langle v\rangle_T \oplus V_1. \tag{4.10}$$

♡

证明

1. 根据指数的定义可知，$T^{m-1} \neq O, T^m = O$，因此存在 $v \in \operatorname{Ker} T^m \backslash \operatorname{Ker} T^{m-1}$，使得 $T^{m-1}v \neq 0, \quad T^m v = 0$，得证.
2.

因此根据上一节末的推理，可得如下定理.

定理 4.1

设 $T \in \mathcal{L}(V)$，则对任意广义特征空间 $G(\lambda, T)$，存在向量 $v_1, v_2, \cdots, v_s \in G(\lambda, T)$，使得

$$G(\lambda, T) = \langle v_1\rangle_T \oplus \cdots \oplus \langle v_s\rangle_T. \tag{4.11}$$

♡

命题 4.6

设 $N \in \mathcal{L}(V)$ 幂零，则存在向量，$v_1, \cdots, v_n \in V$ 和非负整数 $m_1, \cdots, m_n$，使得

$$N^{m_1}v_1, \cdots, Nv_1, v_1, \cdots, N^{m_n}v_1, \cdots, Nv_n, v_n, \tag{4.12}$$

是 V 的一组基，并且

$$N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0. \tag{4.13}$$

♠

证明 对维数归纳，当 $\dim V = 1$ 时，N 为零算子，取任意非零向量 v_1 满足条件. 假设命题对所有维数小于 $\dim V$ 的空间成立，由于 N 幂零，故不是满的，因此 $\operatorname{Im} N$ 为 V 的非零子空间，且维数严格小于 V，对限制算子 $N|_{\operatorname{Im} N}$

使用归纳假设，即存在向量 $v_1,\cdots,v_n\in\operatorname{Im}N$ 以及非负整数 $m_1,\cdots,m_n$ 使得

$$N^{m_1}v_1,\cdots,Nv_1,v_1,\cdots,N^{m_n}v_1,\cdots,Nv_n,v_n, \tag{4.14}$$

是 $\operatorname{Im}N$ 的一组基，并且

$$N^{m_1+1}v_1=\cdots=N^{m_n+1}v_n=0. \tag{4.15}$$

由于 $v_j\in\operatorname{Im}N$，故对应每个 v_j，存在 $u_j\in V$ 使得 $v_j=Nu_j$，上面的向量组中添加这些 u_j 得到的向量组为

$$N^{m_1+1}u_1,\cdots,Nu_1,u_1,\cdots,N^{m_n+1}u_1,\cdots,Nu_n,u_n, \tag{4.16}$$

下面说明它们线性无关. 假设它们有某个线性组合为 0，则对两边同时作用 N，则 $N^{m_k+1}u_k$ 项被消去，剩余项恰好为所取的 $\operatorname{Im}N$ 一组基，可知它们对应的系数全为 0. 而对于被消去的

$$N^{m_1+1}u_1,\cdots,N^{m_n+1}u_n\Longleftrightarrow N^{m_1}v_1,\cdots,N^{m_n}v_n \tag{4.17}$$

可知它们也是线性无关的（部分基），故得证.

接下来将上面的无关组扩充为 V 的一组基

$$N^{m_1+1}u_1,\cdots,Nu_1,u_1,\cdots,N^{m_n+1}u_1,\cdots,Nu_n,u_n,w_1,\cdots,w_p, \tag{4.18}$$

由于每个 $Nw_k\in\operatorname{Im}N$，因此存在 $x_k\in V$，使得 $Nx_j=Nw_j$，设 $u_{n+j}=w_j-x_j$，可知 $Nu_{n+j}=0$. 并且注意到 $x_j\in\operatorname{Span}N^{m_1+1}u_1,\cdots,Nu_1,\cdots,N^{m_n+1}u_1,\cdots,Nu_n$，故

$$N^{m_1+1}u_1,\cdots,Nu_1,u_1,\cdots,N^{m_n+1}u_1,\cdots,Nu_n,u_n,u_{n+1},\cdots,u_{n+p}, \tag{4.19}$$

同样是 V 的基（线性无关 + 个数相同），这个基具有我们所要的形式.

注 从上面的过程也可以看出，向量 v 的个数为从 $\operatorname{Im}N$ 到 V 扩充的向量数，即

$$\dim V-\dim\operatorname{Im}N=\dim\operatorname{Ker}N, \tag{4.20}$$

这恰好是 0 特征的几何重数.

如果考虑在对应基下的矩阵，则立刻可以得到 Jordan 标准型.

定理 4.2 (Jordan 标准型)

设 $T\in\mathcal{L}(V)$，$\operatorname{spec}T=\{\lambda_1,\cdots,\lambda_m\}$，则存在 $v_{11},\cdots,v_{1k_1}\in G(\lambda_1,T);\cdots;v_{m1},\cdots,v_{mk_m}\in G(\lambda_m,T)$，使得

$$\begin{aligned}V&=G(\lambda_1,T)\oplus G(\lambda_2,T)\oplus\cdots\oplus G(\lambda_m,T) &(4.21)\\&=(\langle v_{11}\rangle_T\oplus\cdots\oplus\langle v_{1k_1}\rangle_T)\oplus\cdots\oplus(\langle v_{m1}\rangle_T\oplus\cdots\oplus\langle v_{mk_m}\rangle_T) &(4.22)\end{aligned}$$

且 T 在基

$$\begin{aligned}&T^{n_{11}-1}v_{11},\cdots,v_{11},\cdots,T^{n_{1k_1}-1}v_{1k_1}\cdots,v_{1k_1}, &(4.23)\\&\cdots\cdots &(4.24)\\&T^{n_{m1}-1}v_{m1},\cdots,v_{m1},\cdots,T^{n_{mk_m}-1}v_{mk_m},\cdots,v_{mk_m}. &(4.25)\end{aligned}$$

下的矩阵为

$$J=\operatorname{diag}(J_{n_{11}}(\lambda_1),\cdots,J_{n_{1k_1}}(\lambda_1),\cdots,J_{n_{m1}}(\lambda_m),\cdots,J_{n_{mk_m}}(\lambda_m)) \tag{4.26}$$

其中

$$J_{n_{ij}}(\lambda_i)=\begin{pmatrix}\lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}_{n_{ij}\times n_{ij}} \tag{4.27}$$

为一个 Jordan 块，矩阵 J 称为 Jordan 标准型.

♡

4.1.3 幂零上三角阵的进一步相似

根据上一章的讨论，不妨设 $A = \mathrm{diag}(A_1, \cdots, A_d)$，其中 A_k 为以 λ_k 为对角元的上三角矩阵. 因此只需考虑每个块 $A_i = \lambda_i I + A_i'$，而

$$PA_iP^{-1} = P(\lambda_i I + A_i')P^{-1} = \lambda_i I + PA_i'P^{-1}, \tag{4.28}$$

故可不妨设 $\lambda_i = 0$，上一节的讨论已经引出了 Jordan 块与 Jordan 标准型，因此这里仅通过打洞的方法证明一下.

引理 4.2

设 $A \in \mathbb{F}^{n\times n}$，且为幂零上三角阵（也即 $\varphi_A = \lambda^n$，A 仅有特征值 0），则 $A \overset{s}{\sim} J = \mathrm{diag}(J_{m_1}, \cdots, J_{m_t})$，其中

$$J_{m_k} = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}_{k\times k} \tag{4.29}$$

♡

证明 归纳证明，当 $n = 1$ 时命题显然成立，设对 $n-1$ 阶的矩阵 A 成立，也即不妨设

$$A = \begin{pmatrix} J_{m_1} & & & X_1 \\ & \ddots & & \vdots \\ & & J_{m_t} & X_t \\ & & & 0 \end{pmatrix} \tag{4.30}$$

记

$$X_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{im_i} \end{pmatrix}, \quad X_i' = \begin{pmatrix} 0 \\ x_{i1} \\ \vdots \\ x_{im_{i-1}} \end{pmatrix}_{m_i\times 1}, \quad Y_i = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ x_{im_i} \end{pmatrix} \tag{4.31}$$

则

$$A \overset{s}{\sim} \begin{pmatrix} I_{m_1} & & & X_1' \\ & \ddots & & \vdots \\ & & I_{m_t} & X_t' \\ & & & 1 \end{pmatrix} \begin{pmatrix} J_{m_1} & & & X_1 - J_{m_1}X_1' \\ & \ddots & & \vdots \\ & & J_{m_t} & X_t - J_{m_t}X_t' \\ & & & 0 \end{pmatrix} \begin{pmatrix} I_{m_1} & & & -X_1' \\ & \ddots & & \vdots \\ & & I_{m_t} & -X_t' \\ & & & 1 \end{pmatrix} \tag{4.32}$$

$$= \begin{pmatrix} J_{m_1} & & & Y_1 \\ & \ddots & & \vdots \\ & & J_{m_t} & Y_t \\ & & & 0 \end{pmatrix} \tag{4.33}$$

此时若每个 Y_i 均为 0，则命题已然得证，故不妨设 $Y_1 \neq 0$，$x_{1m_1} = 1$，再设

$$\Lambda_i = \begin{pmatrix} 0 & x_{im_i} I_{m_i} \end{pmatrix}_{m_i\times m_1}, \quad i = 2, 3, \cdots, t, \tag{4.34}$$

则

$$A \overset{s}{\sim} \begin{pmatrix} I_{m_1} & & & & \\ -\Lambda_2 & I_{m_2} & & & \\ \vdots & & \ddots & & \\ -\Lambda_t & & & I_{m_t} & \\ & & & & 1 \end{pmatrix} \begin{pmatrix} J_{m_1} & & & & y_1 \\ & J_{m_2} & & & Y_2 \\ & & \ddots & & \vdots \\ & & & J_{m_t} & Y_t \\ & & & & 0 \end{pmatrix} \begin{pmatrix} I_{m_1} & & & & \\ \Lambda_2 & I_{m_2} & & & \\ \vdots & & \ddots & & \\ \Lambda_t & & & I_{m_t} & \\ & & & & 0 \end{pmatrix} \tag{4.35}$$

$$= \begin{pmatrix} J_{m_1} & & & & Y_1 \\ -\Lambda_2 J_{m_1} + J_{m_2}\Lambda_2 & J_{m_2} & & & \\ \vdots & & \ddots & & \\ -\Lambda_t J_{m_1} + J_{m_t}\Lambda_t & & & J_{m_t} & \\ & & & & 0 \end{pmatrix} = \begin{pmatrix} J_{m_1} & & & & Y_1 \\ & J_{m_2} & & & \\ & & \ddots & & \\ & & & J_{m_t} & \\ & & & & 0 \end{pmatrix} \tag{4.36}$$

$$\overset{s}{\sim} \begin{pmatrix} J_{m_1} & Y_1 & & & \\ & 0 & & & \\ & & J_{m_2} & & \\ & & & \ddots & \\ & & & & J_{m_t} \end{pmatrix} = \begin{pmatrix} J'_{m_1} & & & \\ & J_{m_2} & & \\ & & \ddots & \\ & & & J_{m_t} \end{pmatrix} \tag{4.37}$$

故命题对 n 阶矩阵也成立，原命题得证.

注

1. 上面不妨设 $x_{1m_1} = 1$ 是类似于这样的相似过程
$$\begin{pmatrix} 1 & \\ & a \end{pmatrix} \begin{pmatrix} 1 & a \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & \\ & \frac{1}{a} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}, \tag{4.38}$$
2. 第二步中，两个 J 矩阵分别将 Λ 所有元素右移、上移，而这二者结果是一致的.
3. 最后一步交换过程如下
$$\begin{pmatrix} I_{m_1} & & \\ & & 1 \\ & I & \end{pmatrix} \begin{pmatrix} J_{m_1} & & Y_1 \\ & J & \\ & & 0 \end{pmatrix} \begin{pmatrix} I_{m_1} & & \\ & & I \\ & 1 & \end{pmatrix} = \begin{pmatrix} J_{m_1} & Y_1 & \\ & 0 & \\ & & J \end{pmatrix} \tag{4.39}$$

通过对每个分块使用上面的引理，即可得到最终的 Jordan 标准型.

定理 4.3 (Jordan 标准型)

设 $A \in \mathbb{F}^{n\times n}$，$\operatorname{spec} A = \{\lambda_1, \cdots, \lambda_d\}$ 为 A 的所有相异特征值，则 $A \overset{s}{\sim} J = \operatorname{diag}(J_1(\lambda_1), \cdots, J_d(\lambda_d))$ 其中

$$J_i(\lambda_i) = \operatorname{diag}(J_{n_{i1}}(\lambda_i), \cdots, J_{n_{i,k_i}}(\lambda_i)) \tag{4.40}$$

$$J_{n_{ik_i}}(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{n_{ik_i}\times n_{ik_i}} \tag{4.41}$$

矩阵 J 称为 Jordan 标准型.

事实上，这样的 Jordan 标准型是唯一的，关于这一点需要更细致地刻画 Jordan 型.

4.1.4 Weyr 特征

本节通过引入 Weyr 特征来说明 Jordan 型的唯一性. 由于相似变换可以对换 Jordan 块的顺序，因此我们讨论的是在置换下的唯一性.

对于矩阵 A，若将其 Jordan 型块按照特征值的不同进行分块，则 J 的唯一性实际上是每个 $\lambda \in \operatorname{spec} A$ 对应分块的唯一性. 而这样的分块中又有多个不同大小的 Jordan 块，因此每个单谱矩阵（只有一个特征值的矩阵）的唯一性实质上是指**不同大小的 Jordan 块数量上的唯一性**.

比如考虑单谱零特征矩阵

$$J = \operatorname{diag}(J_3, J_3, J_2, J_2, J_2, J_1), \tag{4.42}$$

要说明 J 的唯一性，实质上是说明与 J 相似的 Jordan 型 J' 同样含有 2 个三阶 Jordan 块、3 个二阶 Jordan 块、1 个 1 阶 Jordan 块. 为此，我们首先要看到 Jordan 阵本身具有的良好性质.

设 A 为单谱的 Jordan 型，特征值为 λ，则 A 可以分解为

$$A = \lambda I + J, \tag{4.43}$$

其中 J 为单谱零特征矩阵. 而这个单谱幂零矩阵，具有“平方降秩”的特性.

问题 4.1 设

$$J = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{4.44}$$

试对任意 $p \in \mathbb{N}$，讨论 $\mathrm{rk} J^p$ 的变化.

解 计算得

$$J^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad J^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad J^4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{4.45}$$

因此

$$\mathrm{rk} J^p = \begin{cases} 4 - p & 0 \leqslant p < 4, \\ 0 & p \geqslant 4. \end{cases} \tag{4.46}$$

此外，也可以看出 $\mathrm{rk} J^p$ 随 p 线性变化，即

$$\mathrm{rk} J^{p-1} - \mathrm{rk} J^p = \begin{cases} 1 & 1 \leqslant p \leqslant k, \\ 0 & p > k. \end{cases} \tag{4.47}$$

上面的性质并不是特例，更一般地，我们有

命题 4.7

设 $J_k \in \mathbb{F}^{k\times k}$ 为幂零 Jordan 块，则

$$\mathrm{rk} J_k^p = \begin{cases} k - p & 0 \leqslant p < k, \\ 0 & p \geqslant k. \end{cases} \tag{4.48}$$

$$\mathrm{rk} J_k^{p-1} - \mathrm{rk} J_k^p = \begin{cases} 1 & 1 \leqslant p \leqslant k, \\ 0 & p > k. \end{cases} \tag{4.49}$$

可以看出，幂零 Jordan 块的秩随幂增加而减小，并且当幂等于阶数后被零化，也就是说，幂零 Jordan 块的

阶可以被其幂次的秩刻画.

推论 4.1

1. 设 J 为幂零 Jordan 块，则 J 的阶大于 p，当且仅当 $\mathrm{rk}J^{p-1}-\mathrm{rk}J^{p}=1$.
2. 设 J 为幂零 Jordan 块，则 J 的阶等于 p，当且仅当 $\mathrm{rk}J^{p-1}-\mathrm{rk}J^{p}=1$ 且 $\mathrm{rk}J^{p}-\mathrm{rk}J^{p+1}=0$，也即 $(\mathrm{rk}J^{p-1}-\mathrm{rk}J^{p})-(\mathrm{rk}J^{p}-\mathrm{rk}J^{p+1})=1$.

证明 从已有结论来看，证明是显然的.

上面的结论是对局部的一个 Jordan 块考虑的，将之应用到一般幂零 Jordan 型的每个块上，可以得到

命题 4.8

设 $J\in\mathbb{F}^{n\times n}$ 为幂零 Jordan 型，设

$$w_p=\mathrm{rk}J^{p-1}-\mathrm{rk}J^{p}, \tag{4.50}$$

则其中阶至少为 p 的 Jordan 块的个数为 w_p，阶恰好为 p 的 Jordan 块的个数为 w_p-w_{p+1}.

证明 设 $J=\mathrm{diag}(J_{n_1},\cdots,J_{n_s})$，则

$$w_p=\mathrm{rk}J^{p-1}-\mathrm{rk}J^{p} \tag{4.51}$$

$$=\sum_{i=1}^{s}(\mathrm{rk}J_{n_i}^{p-1}-\mathrm{rk}J_{n_i}^{p}) \tag{4.52}$$

上式告诉我们，其中一个求和项为 1，当且仅当其对应的 Jordan 块的阶至少为 p，因此上式可用于计数阶至少为 p 的 Jordan 块的个数，而阶恰好为 p 的块的个数同理.

上面的命题给出了幂零 Jordan 型唯一性的刻画，是时候回到一般情况了，我们首先根据上面的推理，定义一般方阵 A 的 Weyr 特征.

定义 4.4 (Weyr 特征)

设 $A\in\mathbb{F}^{n\times n}$，$\lambda\in\mathrm{spec}\,A$，令 q 为 $A-\lambda I$ 的指数，则定义

$$w_p(\lambda):=\mathrm{rk}(A-\lambda I)^{p-1}-\mathrm{rk}(A-\lambda I)^{p} \tag{4.53}$$

为 A 的相伴于特征值 λ 的 Weyr 特征.

由此，不难将前面的结论推广到一般形式.

定理 4.4 (Weyr)

设 $A\in\mathbb{F}^{n\times n}$，$\lambda\in\mathrm{spec}\,A$，则 A 的 Jordan 块中相应于 λ 的，大小为 p 的 Jordan 块的个数为

$$w_p(\lambda)-w_{p+1}(\lambda) \tag{4.54}$$

其中 $w_p(\lambda)$ 为 A 的相伴于特征值 λ 的 Weyr 特征.

证明 设 $A\overset{s}{\sim}J=\mathrm{diag}(J_{n_1}(\lambda_1),J')$，$\lambda_1\notin\mathrm{spec}\,J'$，则

$$w_p(\lambda_1)=\mathrm{rk}(A-\lambda_1 I)^{p-1}-\mathrm{rk}(A-\lambda_1 I)^{p} \tag{4.55}$$

$$=\mathrm{rk}(J-\lambda_1 I)^{p-1}-\mathrm{rk}(J-\lambda_1 I)^{p} \tag{4.56}$$

$$=(\mathrm{rk}(J_{n_1}(\lambda_1)-\lambda_1 I)^{p-1}+\mathrm{rk}(J'-\lambda_1 I)^{p-1}) \tag{4.57}$$

$$-(\mathrm{rk}(J_{n_1}(\lambda_1)-\lambda_1 I)^{p}+\mathrm{rk}(J'-\lambda_1 I)^{p}) \tag{4.58}$$

$$=\mathrm{rk}J_{n_1}^{p-1}-\mathrm{rk}J_{n_1}^{p}, \tag{4.59}$$

根据已有结论，可知命题对 λ_1 成立，同理可知，命题对每个 $\lambda\in\mathrm{spec}\,A$ 成立.

注 事实上，Weyr 特征中的 λ 可以不局限于 A 的特征值，只是 $\lambda\notin\mathrm{spec}\,A$ 时它没有意义，只要注意到：若

$\lambda \notin \operatorname{spec} A$，则

$$\operatorname{rk}(A-\lambda I)^p \equiv \operatorname{rk}(A-\lambda I) \quad \forall p \in \mathbb{N}. \tag{4.60}$$

也就是说 $w_p(\lambda) \equiv 0$，这可以理解为 A 相应于 λ 的 Jordan 块个数为 0，同样表示 $\lambda \notin \operatorname{spec} A$.

根据上面的讨论，可以得到我们的目标定理

定理 4.5

设 $A \in \mathbb{F}^{n\times n}$，则 A 的 Jordan 标准型在置换下是唯一的.

证明 只要注意到 Wery 特征是相似不变量即可.

最后再来讨论 Weyr 特征的其它性质.

命题 4.9

设 $A \in \mathbb{F}^{n\times n}$，对应 Jordan 型为 J，则关于 Weyr 特征，对任意 $\lambda_i \in \operatorname{spec} A$，有如下命题成立

1. 对给定 $\lambda \in \operatorname{spec} A$，$\{w_p\}$ 为正项递减序列.
2. 设 q_i 为 $A-\lambda_i I$ 的指数，则 $w_{q_i} > 0, w_{q_i+1} = 0$，且由此得 q_i 为 J 中相应于 λ_i 的最大块的阶数.
3. $w_1(\lambda_i) = m_i$，m_i 为 λ_i 对应的几何重数.
4. $w_1(\lambda_i)$ 表示 J 中相应于 λ_i 的 Jordan 块的个数.
5. 设 q_i 为 $A-\lambda_i I$ 的指数，则

$$\sum_{k=1}^{\infty} w_k(\lambda_i) = \sum_{k=1}^{q_i} w_k(\lambda_i) = n_i, \tag{4.61}$$

   n_i 为 λ_i 对应的代数重数.
6. A 的所有 Weyr 特征之和为 A 的阶数 n.

证明

根据 Weyr 特征，也可以给出另一个看待相似的视角.

推论 4.2

设 $A, B \in \mathbb{F}^{n\times n}$，且 $\operatorname{spec} A = \operatorname{spec} B$，则 $A \overset{s}{\sim} B$ 当且仅当 A, B 相应于相同特征值的 Weyr 特征相同.

4.2 lambda 矩阵

这一部分讨论 λ 矩阵（多项式矩阵），首先从一般环上的矩阵入手.

4.2.1 交换环上的矩阵环

设 R 为交换环[1]，则自然引出交换环上的 $m\times n$ 阶矩阵

$$R^{m\times n} = \{(a_{ij})_{m\times n} : a_{ij} \in R\}. \tag{4.62}$$

且

$$(a_{ij})_{m\times n} = (b_{ij})_{p\times q} \Longleftrightarrow m = p, n = q, a_{ij} = b_{ij}, \forall i, j. \tag{4.63}$$

仿照过去主要研究的域 $\mathbb{F}$ 上矩阵的性质，可以定义加法、数乘与乘法

[1] 默认交换环含幺.

定义 4.5

设 R 为交换环，$R^{m\times n}$ 为 R 上的 $m\times n$ 矩阵集合，则定义运算

$$+ : R^{m\times n}\times R^{m\times n}\longrightarrow R^{m\times n} \tag{4.64}$$

$$(a_{ij})_{m\times n}+(b_{ij})_{m\times n}\longmapsto (a_{ij}+b_{ij})_{m\times n} \tag{4.65}$$

$$\cdot : R\times R^{m\times n}\longrightarrow R^{m\times n} \tag{4.66}$$

$$r(a_{ij})_{m\times n}\longmapsto (ra_{ij})_{m\times n} \tag{4.67}$$

$$\cdot : R^{m\times n}\times R^{n\times p}\longrightarrow R^{m\times p} \tag{4.68}$$

$$(a_{ij})_{m\times n}\cdot (b_{ij})_{n\times p}\longmapsto (c_{ij})_{m\times p} \tag{4.69}$$

其中 $c_{ij}=\sum\limits_{k=1}^{n}a_{ik}b_{kj}$.

♣

矩阵乘法满足结合律（易验证）. 同样地，可以定义 R 上的零矩阵与单位阵，其中的 “0，1” 分别为 R 中的加法、乘法单位元，于是

命题 4.10

设 R 为交换环，则 $(R^{m\times m},+)$ 是一个阿贝尔群，$(R^{m\times m},\cdot)$ 是一个幺半群，并且 $(R^{m\times m},+,)$ 是一个交换环.

♠

由于 R 为交换环，因此可以顺理成章定义其上的行列式.

定义 4.6

设 R 为交换环，定义

$$\det : R^{n\times n}\longrightarrow R \tag{4.70}$$

$$\begin{pmatrix} a_{11} & \cdots & a_{1n}\\ \vdots & \ddots & \vdots\\ a_{n1} & \cdots & a_{nn}\end{pmatrix}\longmapsto \sum_{\sigma\in S_n}(-1)^{\tau(\sigma)}a_{1\sigma(1)}\cdots a_{n\sigma(n)}. \tag{4.71}$$

♣

行列式也可看作是从 n 个 R 中的列向量（或行向量）到 R 的映射，并且也是满足多重线性性，反对称性，规范性的唯一函数，这些过去已经讨论过，在此不多赘述. 有了行列式，也自然就有了 Laplace 展开与 Binet-Cauchy 公式.

命题 4.11

设 R 为交换环，$A\in R^{n\times n}$，$1\leqslant i_1<\cdots<i_r\leqslant n$，则

$$\det A=\sum_{1\leqslant i_1<\cdots<i_r\leqslant n}(-1)^{(i_1+\cdots+i_r)+(j_1+\cdots+j_r)}A\begin{pmatrix} i_1,\cdots,i_r\\ j_1,\cdots,j_r\end{pmatrix}A\begin{pmatrix} \hat{i}_1,\cdots,\hat{i}_r\\ \hat{j}_1,\cdots,\hat{j}_r\end{pmatrix}. \tag{4.72}$$

♠

命题 4.12

设 R 为交换环，$A\in R^{m\times n},B\in R^{n\times p}$，$1\leqslant i_1<\cdots<i_r\leqslant m$，$1\leqslant k_1<\cdots<k_r\leqslant p$，则

$$(AB)\begin{pmatrix} i_1,\cdots,i_r\\ k_1,\cdots,k_r\end{pmatrix}=\sum_{1\leqslant j_1<\cdots<j_r\leqslant n}A\begin{pmatrix} i_1,\cdots,i_r\\ j_1,\cdots,j_r\end{pmatrix}B\begin{pmatrix} j_1,\cdots,j_r\\ k_1,\cdots,k_r\end{pmatrix}. \tag{4.73}$$

♠

据此，可定义逆矩阵，并证明矩阵 $A\in R^{m\times n}$ 可逆的充要条件是 $m=n$ 且 $\det A$ 为 R 中的可逆元（单位）. 以及伴随矩阵 A^*（行列式伴随），且有

$$AA^*=A^*A=\operatorname{diag}(\det A,\cdots,\det A)=(\det A)I_n. \tag{4.74}$$

特别地，若 A 可逆，则 $A^{-1} = (\det A)^{-1}A^*$.

由此，可以定义矩阵 A 的秩为非零子式阶的最大值，只不过

$$A \in \mathbb{F}^{n\times n}\text{可逆} \Longleftrightarrow \mathrm{rk}(A) = n \tag{4.75}$$

$$A \in R^{n\times n}\text{可逆} \Longrightarrow \mathrm{rk}(A) = n \tag{4.76}$$

其本质在于环中非零元未必可逆.

接下来仿照 $\mathbb{F}^{n\times n}$ 给出初等方阵的定义，

1. P_{ij} 交换 I_n 的 i, j 两列.
2. $T_{ij}(a)$ 将 I_n 的 j 行（i 列）乘 a 加到 i 行（j 列）.
3. $D_i(a)$ I_n 的第 i 行（i 列）乘 a 倍，a 为 R 中可逆元.

且易知 $\det P_{ij} = -1, \det T_{ij}(a) = 1, \det D_i(a) = a$. 左乘初等阵等价于对矩阵作初等行变换，右乘初等阵等价于作初等列变换.

例 4.1 设 $V/\mathbb{F}$，则 $(\mathcal{L}(V), +, \circ)$ 为一个环，取 $\mathscr{A} \in \mathcal{L}(V)$，则 $R = \mathbb{F}[\mathscr{A}]$ 为由 $\mathscr{A}$ 生成的（交换）子环.

上面的例子在 Caylay-Hamilton 定理的一种证法中，我们曾使用过.

准备工作与回顾到此为止，下面开始正题.

4.2.2 λ 矩阵的定义

定义 4.7

设 $R = \mathbb{F}[\lambda]$ 为域 $\mathbb{F}$ 上的多项式环，则定义其上矩阵（λ 矩阵）为

$$A(\lambda) = \begin{pmatrix} a_{11}(\lambda) & \cdots & a_{1n}(\lambda) \\ \vdots & \ddots & \vdots \\ a_{m1}(\lambda) & \cdots & a_{mn}(\lambda) \end{pmatrix} \tag{4.77}$$

所有 $m \times n$ 阶 λ 矩阵组成的集合为

$$\mathbb{F}[\lambda]^{m\times n} = \{(a_{ij}(\lambda))_{m\times n} : a_{ij}(\lambda) \in \mathbb{F}[\lambda]\}. \tag{4.78}$$

♣

值得注意的是，多项式环上的 λ 矩阵，可与矩阵环上的多项式对应，即

$$\mathbb{F}[\lambda]^{n\times n} \longleftrightarrow \mathbb{F}^{n\times n}[\lambda] \tag{4.79}$$

只不过，上面的性质只适用于方阵，否则 $\mathbb{F}^{m\times n}$ 不是一个矩阵环. 这个有趣的性质我们在 Caylay-Hamilton 定理的一个证明中也使用过.

关于 λ 矩阵的基本运算（加法、乘法等）在前面已经介绍过更一般的情况，因此可以直接从初等变换说起.

定义 4.8

定义 $\mathbb{F}[\lambda]$ 上的初等 λ 阵为

1. P_{ij}　交换 I_n 的 i, j 两列.
2. $T_{ij}(f(\lambda))$　将 I_n 的 j 行（i 列）乘 $f(\lambda)$ 加到 i 行（j 列）.
3. $D_i(a)$　I_n 的第 i 行（i 列）乘 a 倍，$a \in \mathbb{F}^\times$.

并且有 $\det P_{ij} = -1, \det T_{ij}(f(\lambda)) = 1, \det D_i(a) = a$.

♣

按照学习数阵时的节奏，下一步是对 λ 阵作初等变换，得到相对简单的形式.

4.2.3 λ 矩阵的相抵

定义 4.9

设 $A(\lambda), B(\lambda) \in \mathbb{F}[\lambda]^{m\times n}$, 则称 $A(\lambda)$ 与 $B(\lambda)$ 相抵, 记作 $A(\lambda) \sim B(\lambda)$, 若存在初等矩阵 $P_1(\lambda), \cdots, P_s(\lambda) \in GL_m(\mathbb{F}[\lambda]), Q_1(\lambda), \cdots, Q_r(\lambda) \in GL_n(\mathbb{F}[\lambda])$，使得

$$A(\lambda) = P_s(\lambda)\cdots P_1(\lambda)B(\lambda)Q_1(\lambda)\cdots Q_r(\lambda). \tag{4.80}$$

♣

类比于数阵的相抵，易验证 λ 矩阵的相抵也是一种等价关系. 在数阵中，我们证明过任意矩阵都能通过相抵变换变化为相抵标准型，对应于 λ 矩阵，也有类似的结果，只不过并不那样简单如“1”.

定理 4.6

设 $A(\lambda) \in \mathbb{F}[\lambda]^{m\times n}$，$r = \mathrm{rk}(A(\lambda))$，则

$$A(\lambda) \sim \begin{pmatrix} d_1(\lambda) & & & & & \\ & \ddots & & & & \\ & & d_r(\lambda) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \tag{4.81}$$

其中 $d_1(\lambda), \cdots, d_r(\lambda)$ 都为首一多项式，且

$$d_1(\lambda)\big|d_2(\lambda)\big|\cdots\big|d_r(\lambda). \tag{4.82}$$

称为 $A(\lambda)$ 的不变因子.

♡

证明 记

$$\mathscr{S} = \{B(\lambda) : B(\lambda) \sim A(\lambda)\}. \tag{4.83}$$

任取 $B(\lambda) \in \mathscr{S}$ 满足

- $b_{11}(\lambda) \neq 0$ 且首一.
- $\forall C(\lambda) \in \mathscr{S}$，$c_{11}(\lambda) = 0$ 或 $\deg b_{11}(\lambda) \leqslant \deg c_{11}(\lambda)$.

第一条很容易满足，而由于 11 位置元素次数有限，故这样的 $B(\lambda)$ 必然存在，下面断言：$b_{11}(\lambda)\big|b_{ij}(\lambda), \forall i, j$.

1. $b_{11}(\lambda)\big|b_{i1}(\lambda)$.
   若不然，作带余除法 $b_{i1}(\lambda) = q(\lambda)b_{11}(\lambda) + r(\lambda)$，通过左乘 $P_{1i}(\lambda)T_{1i}(-q(\lambda))$，可得新的 $B'(\lambda)$，使得 $\deg b'_{11}(\lambda) < \deg b_{11}(\lambda)$，这与 $B(\lambda)$ 的选取矛盾.
2. $b_{11}(\lambda)\big|b_{1j}(\lambda)$.
   与上一条同理.
3. $b_{11}(\lambda)\big|b_{ij}(\lambda)$.

$$\begin{pmatrix} b_{11}(\lambda) & b_{11}(\lambda)q(\lambda) \\ b_{i1}(\lambda) & b_{ij}(\lambda) \end{pmatrix} \sim \begin{pmatrix} b_{11}(\lambda) & 0 \\ b_{i1}(\lambda) & b_{ij}(\lambda) - q(\lambda)b_{i1}(\lambda) \end{pmatrix} \tag{4.84}$$

$$\sim \begin{pmatrix} b_{11}(\lambda) & 0 \\ (1-q(\lambda))b_{i1}(\lambda) + b_{ij}(\lambda)(\lambda) & b_{ij}(\lambda) - q(\lambda)b_{i1}(\lambda) \end{pmatrix} \tag{4.85}$$

$$\sim \begin{pmatrix} b_{11}(\lambda) & 0 \\ b_{ij}(\lambda) & b_{ij}(\lambda) - q(\lambda)b_{i1}(\lambda) \end{pmatrix} \tag{4.86}$$

上面最后一步相抵到的矩阵 $B'(\lambda)$ 满足我们选取 $B(\lambda)$ 时的要求，故也满足已经证明的两条性质（11 元素

整除第一列的元素），即得

$$b'_{11}(\lambda)|b'_{i1}(\lambda) \Longleftrightarrow b_{11}(\lambda)|b_{ij}(\lambda) \tag{4.87}$$

综上可知

$$B(\lambda) \sim \begin{pmatrix} b_{11}(\lambda) & 0 \\ & b_{11}(\lambda)\tilde{B(\lambda)} \end{pmatrix} \tag{4.88}$$

故根据归纳法可知，命题成立.

注 上述过程中并未明确提及对首一的处理，事实上，有了最后的对角形式后首一是易得的.

根据上面的定理，再加上相抵 λ 矩阵的行列式仅相差一个单位倍，因此可以得到以下推论.

命题 4.13

$A(\lambda) \in \mathbb{F}[\lambda]^{n\times n}$ 可逆当且仅当 $A(\lambda)$ 能表示成若干初等方阵的乘积.

在这里顺便对前面提到的一个小结论举个例子

$$A \in \mathbb{F}^{n\times n}\text{可逆} \Longleftrightarrow \mathrm{rk}(A) = n \tag{4.89}$$

$$A \in R^{n\times n}\text{可逆} \Longrightarrow \mathrm{rk}(A) = n \tag{4.90}$$

例 4.2

$$\begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \not\sim \begin{pmatrix} 1 & \\ & \lambda \end{pmatrix} \tag{4.91}$$

数阵的相抵标准型是唯一的，我们自然希望 λ 阵相抵到的上面的形式也是唯一的，这就是下一步要研究的事.

4.2.4 Smith 标准型

为研究最后的标准型，首先需要关注相抵变换中的不变量.

定义 4.10

设 $A(\lambda) \in \mathbb{F}[\lambda]^{m\times n}$，定义 $A(\lambda)$ 的所有 k 阶非 0 子式的首一最大公约式为 $D_k(\lambda)$，特别地，若 k 阶子式全为 0，则约定 $D_k(\lambda) = 0$.

命题 4.14

设 $A(\lambda) \sim B(\lambda)$，则 $A(\lambda)$ 与 $B(\lambda)$ 具有相同的行列式因子.

证明 根据 Binet-Cauchy 公式，可得二者相互整除，命题得证.

可以看出，行列式因子确实是相抵不变量，由于我们已经证明了每个 λ 矩阵都可以化为对角形式（这里忽略该名词在非方阵的不适性），且对角元有整除的链式关系，因此可得对 $A(\lambda)$

$$D_k(\lambda) = d_1(\lambda)\cdots d_k(\lambda) \Longleftrightarrow d_k(\lambda) = \frac{D_k(\lambda)}{D_{k-1}(\lambda)}. \tag{4.92}$$

这里约定 $D_0(\lambda) = 1$.

由此可以看出，我们求得的形式确实是唯一的标准型，我们概括为如下定理（下面为追求简便，将记号 diag 用在非方阵上）

定理 4.7 (Smith 标准型)

设 $A(\lambda) \in \mathbb{F}[\lambda]^{m\times n}$，$r = \mathrm{rk}(A(\lambda))$，则 $A(\lambda)$ 相似于唯一的 Smith 标准型

$$A(\lambda) \sim \mathrm{diag}(d_1(\lambda), \cdots, d_r(\lambda), 0 \cdots, 0), \tag{4.93}$$

其中 $d_k(\lambda)=\frac{D_k(\lambda)}{D_{k-1}(\lambda)}$（$D_k(\lambda)$ 为 $A(\lambda)$ 的 k 阶行列式因子），均为都为首一多项式，且

$$d_1(\lambda)\big|d_2(\lambda)\big|\cdots\big|d_r(\lambda). \tag{4.94}$$

每个 $d_i(\lambda)$ 称为 $A(\lambda)$ 的不变因子，而它们的全体称为 $A(\lambda)$ 的不变因子组.

接下来要做什么？回想一下，得到相抵标准型后，我们发现只要知道秩就可以直接得到矩阵的相抵标准型；得到相似标准型（Jordan 标准型）后，我们发现可以通过特征值、Weyr 特征来直接得到相似标准型.

所以，下一步就是要研究如何通过尽可能少的条件，得到 λ 矩阵的 Smith 标准型.

定义 4.11 (多项式的初等因子组)

设多项式 $f(\lambda)\in\mathbb{F}[\lambda]$，且

$$f(\lambda)=a\prod_{k=1}^{s}p_k(\lambda)^{n_k}, \tag{4.95}$$

其中 $a\in\mathbb{F}^\times, p_1,\cdots,p_s$ 两两互素，$n_1+\cdots+n_s=\deg f$，则称 $p_1^{n_1},\cdots,p_s^{n_s}$ 为 $f(\lambda)$ 的初等因子组.

定义 4.12 (λ 矩阵的初等因子组)

设 $A(\lambda)\in\mathbb{F}^{m\times n}$，则 $A(\lambda)$ 的初等因子组定义为所有不变因子 $d_1(\lambda),\cdots,d_r(\lambda)$ 的初等因子组的并（记重数）.

将上面的内容汇总，可以得到如下定理

定理 4.8

设 $A(\lambda),B(\lambda)\in\mathbb{F}[\lambda]^{m\times n}$，则下述命题等价

1. $A(\lambda)\sim B(\lambda)$.
2. $A(\lambda),B(\lambda)$ 有相同的秩与行列式因子组.
3. $A(\lambda),B(\lambda)$ 有相同的秩与不变因子组.
4. $A(\lambda),B(\lambda)$ 有相同的秩与初等因子组.

证明 事实上，只需证明 $4\Longrightarrow 3$.

设 $\mathrm{rk}(A(\lambda))=r$，且其相异素因子个数为 s. 下面我们将“1”也视为初等因子，即若 $d_i(\lambda)$ 不含因子 p_k，则记其有初等因子 p_k^0，事实上，添加这样的定义后的初等因子组，并不会产生矛盾，即原本相同的初等因子组不会因添加这些“1”而变得不同，反之亦然.

在这样的定义下，可知 $A(\lambda)$ 的每个不变因子都有 s 个初等因子. 根据不变因子的整除关系，可知 $A(\lambda)$ 的初等因子组有如下唯一的表示形式

$$p_1^{a_{11}},p_2^{a_{12}},\cdots,p_s^{a_{2s}} \tag{4.96}$$
$$p_1^{a_{21}},p_2^{a_{22}},\cdots,p_s^{a_{2s}} \tag{4.97}$$
$$\cdots\cdots\cdots \tag{4.98}$$
$$p_1^{a_{r1}},p_2^{a_{r2}},\cdots,p_s^{a_{rs}} \tag{4.99}$$

满足

1. $r=\mathrm{rk}(A)$.
2. $a_{ij}\geqslant 0$.
3. $d_r(\lambda)=p_1^{n_1}\cdots p_s^{n_s}, n_k>0, p_k\neq 1$.
4. $d_i(\lambda)=p_1^{a_{i1}}p_2^{a_{i2}}\cdots p_s^{a_{is}}$，也即第 i 行为 $d_i(\lambda)$ 的初等因子组.
5. 数列 $\{a_{i\cdot}\}$ 单调递增，也即每列的幂次单调递增.

到这里，可以看出对任意具有相同初等因子组的 $A(\lambda), B(\lambda)$，它们的初等因子组有如上的表示形式，并且根据上述形式，每行都能唯一确定一个不变因子，即 $A(\lambda), B(\lambda)$ 具有相同的不变因子组，得证.

注 上面的证明过程同时也给出了一种从初等因子组求 Smith 标准型的算法：

1. 找出所有素因子（s 个）.
2. 按列依次构建 $r \times s$ 表，即对每个素因子，由含该素因子的初等因子的次数由高到低，由下至上标明次数.
3. 将每行相乘，得到 $d_1, \cdots, d_r$.

例 4.3 设矩阵 $A(\lambda)$ 秩为 3，初等因子组为

$$\lambda, \lambda+1, \lambda, \lambda-4, (\lambda-4)^2, \tag{4.100}$$

则可知 $A(\lambda)$ 的初等因子组

$$\lambda^0, (\lambda+1)^0, (\lambda-4)^0 \to 1 \tag{4.101}$$

$$\lambda^1, (\lambda+1)^0, (\lambda-4)^1 \to \lambda(\lambda-4) \tag{4.102}$$

$$\lambda^1, (\lambda+1)^1, (\lambda-4)^2 \to \lambda(\lambda+1)(\lambda-4)^2 \tag{4.103}$$

即 $A(\lambda)$ 的相抵标准型为 $\mathrm{diag}(1, \lambda(\lambda-4), \lambda(\lambda+1)(\lambda-4)^2)$

上述命题说明，初等因子也是一种相抵不变量，并且很多时候，初等因子比行列式因子更好求得，由此通过初等因子组得到矩阵要更方便. 为了完善这一方法，给出如下命题

命题 4.15

1. $A(\lambda) = \mathrm{diag}(f_1(\lambda), \cdots, f_r(\lambda), 0, \cdots, 0)$ 的初等因子组为所有 $f_i(\lambda)$ 的初等因子组的并.
2. 准对角阵 $A(\lambda) = \mathrm{diag}(A_1(\lambda), \cdots, A_s(\lambda))$ 的初等因子组为所有 $A_i(\lambda)$ 的初等因子组的并.

证明

1. 若 $f_1|f_2|\cdots|f_r$，则这就是 Smith 标准型，命题显然成立. 若不然，不妨设 $f_1 \nmid f_2$，即 $f_1(\lambda) = g(\lambda)^{m_1}q_1(\lambda)$，$f_2(\lambda) = g(\lambda)^{m_2}q_2(\lambda)$，其中 $(g, p_1) = (g, p_2) = 1$，且 $m_1 > m_2 > 0$，由于行列式因子相同与相抵等价，因此有

$$\begin{pmatrix} f_1(\lambda) & \\ & f_2(\lambda) \end{pmatrix} = \begin{pmatrix} g(\lambda)^{m_1}q_1(\lambda) & \\ & g(\lambda)^{m_2}q_2(\lambda) \end{pmatrix} \tag{4.104}$$

$$\sim \begin{pmatrix} g(\lambda)^{m_2}q_1(\lambda) & \\ & g(\lambda)^{m_1}q_2(\lambda) \end{pmatrix} \tag{4.105}$$

$$= \begin{pmatrix} f_1'(\lambda) & \\ & f_2'(\lambda) \end{pmatrix} \tag{4.106}$$

并且 f_1, f_2 初等因子组的并与新的 f_1', f_2' 的初等因子组的并相同. 因此，可以通过重复上述变换，将 $A(\lambda)$ 相抵到 Smith 标注型，且每一步都保证初等因子组的并不变，即命题得证.

2. $A_i(\lambda) \sim \mathrm{diag}(d_{i1}(\lambda), \cdots, d_{ir_i}(\lambda))$，则利用 1 中结论可知，命题得证.

接下来，我们把话题回到 Jordan 标准型上，以这样的问题作为过渡

问题 4.2 设 $J(\lambda) = \lambda I - J$，其中 $J = \mathrm{diag}(A_1, \cdots, A_s)$ 为 Jordan 型矩阵，求 $J(\lambda)$ 的初等因子组.

解 根据上面的理论，由于 $J(\lambda)$ 是准对角阵，故只需考虑每个 Jordan 块的初等因子组，以 $A_1 = \mathrm{diag}(J_{n_{11}}(\lambda_1), \cdots, J_{n_{1m_1}}(\lambda_1))$ 为例（不妨设 $n_{11} \leqslant \cdots \leqslant n_{1m_1}$）. 通过计算行列式因子，可得每个 $J_{n_{1k}}$ 不变因子为

$$d_1(\lambda) = \cdots = d_{n_{1k-1}}(\lambda) = 1, d_{n_{1k}}(\lambda) = (\lambda-\lambda_1)^{n_{1k}} \tag{4.107}$$

因此其初等因子组为

$$(\lambda-\lambda_1)^{n_{11}}, \cdots, (\lambda-\lambda_1)^{n_{1m_1}} \tag{4.108}$$

回到矩阵 $J(\lambda)$ 的情形，其初等因子组为每个 A 的初等因子组的并

$$(\lambda-\lambda_1)^{n_{11}}, \cdots, (\lambda-\lambda_1)^{n_{1m_1}} \tag{4.109}$$

$$\cdots\cdots\cdots \tag{4.110}$$

$$(\lambda-\lambda_s)^{n_{s1}},\cdots,(\lambda-\lambda_s)^{n_{sm_s}} \tag{4.111}$$

可以发现，Jordan 型矩阵的初等因子组就是其每个 Jordan 块的特征多项式的集合（记重数），而初等因子组是相抵不变量，因此每个矩阵 A 的 Jordan 型 J 的形式可由 $\lambda I-A$ 的初等因子组确定. 接下来要做的就是用 λ 矩阵研究方阵的相似标准型.

4.2.5 有理标准型

定义 4.13

设 $A\in\mathbb{F}^{n\times n}$，则称 $A(\lambda)=\lambda I-A\in\mathbb{F}[\lambda]^{n\times n}$ 为 A 的特征方阵. ♣

从形式来看，这里的 “特征” 方阵与过去接触的 “特征” 多项式很相似，事实上，根据定义，特征方阵 $A(\lambda)$ 的行列式就是 A 的特征多项式，并且，若 $A\overset{s}{\sim}B$，则显然有 $A(\lambda)\sim B(\lambda)$. 反之，若两个 λ 矩阵相抵，则二者都具有同一个 Smith 标准型，那么通过相同的 Smith 标准型能否得到 $A\overset{s}{\sim}B$，这是接下来要考虑的问题.

引理 4.3

设 $V/\mathbb{F}$，$\dim V=n<\infty$，$\mathscr{A}\in\mathcal{L}(V,\mathbb{F})$，$A$ 为 $\mathscr{A}$ 在基 $M=(\alpha_1,\cdots,\alpha_n)$ 下的矩阵. 设 $A(\lambda)$ 相抵于

$$D(\lambda)=\operatorname{diag}(1,\cdots,1,f_{s+1}(\lambda),\cdots,f_n(\lambda)), \tag{4.112}$$

其中每个 $f_i(\lambda)$ 为首一多项式，次数为 $d_i\geqslant 1$，则

1. V 可以分解为 $n-s$ 个循环子空间 $\langle\beta_i\rangle_{\mathscr{A}}$ 的直和，其中每个 β_i 的极小多项式为 $f_i(\lambda)$，且 $\dim\langle\beta_i\rangle_{\mathscr{A}}=d_i$.
2. 对每个 $s+1\leqslant i\leqslant n$，$B_i=(\beta_i,\cdots,\mathscr{A}^{d_i-1}\beta_i)$ 是 $\langle\beta_i\rangle_{\mathscr{A}}$ 的一组基，且 $B=\bigcup_{i=s+1}^n B_i$ 是 V 的一组基，若设

$$f_i(\lambda)=\lambda^{d_i}+a_{i,d_i-1}\lambda^{d_i-1}+\cdots+a_{i1}\lambda+a_{i0}. \tag{4.113}$$

则 $\mathscr{A}$ 在基 B 下的矩阵为 $B'=\operatorname{diag}(B_{s+1},\cdots,B_n)$，其中 B_i 为 $f_i(\lambda)$ 对应的友方阵. ♡

证明 设 $A(\lambda)=\lambda I-A$，则有

$$(\alpha_1,\cdots,\alpha_n)A(\mathscr{A})=(0,\cdots,0). \tag{4.114}$$

由于 $A(\lambda)\sim\operatorname{diag}(1,\cdots,1,f_{s+1}(\lambda),\cdots,f_n(\lambda))$，故存在可逆 λ 矩阵 $P(\lambda),Q(\lambda)$，使得

$$P(\lambda)A(\lambda)Q(\lambda)=\operatorname{diag}(1,\cdots,1,f_{s+1}(\lambda),\cdots,f_n(\lambda)). \tag{4.115}$$

在等式两边取行列式

$$\det(P(\lambda))\det(A(\lambda))\det(Q(\lambda))=f_{s+1}(\lambda)\cdots f_n(\lambda). \tag{4.116}$$

由于 $\det(P(\lambda))\det(Q(\lambda))\in\mathbb{F}^\times$，且 $\det(A(\lambda))=\varphi_A(\lambda)$，比较首项系数可知 $\det(P(\lambda)),\det(Q(\lambda))=1$，故

$$\varphi_A(\lambda)=f_{s+1}(\lambda)\cdots f_n(\lambda). \tag{4.117}$$

将 λ 用 $\mathscr{A}$ 替换（交换环的代入同态），回到最开始的式子，有

$$(\alpha_1,\cdots,\alpha_n)P(\mathscr{A})^{-1}P(\mathscr{A})A(\mathscr{A})Q(\mathscr{A})=(0,\cdots,0) \tag{4.118}$$

$$\Longrightarrow(\beta_1,\cdots,\beta_n)\operatorname{diag}(1,\cdots,1,f_{s+1}(\mathscr{A}),\cdots,f_n(\mathscr{A}))=(0,\cdots,0) \tag{4.119}$$

其中 $(\beta_1,\cdots,\beta_n)P(\mathscr{A})=(\alpha_1,\cdots,\alpha_n)$，对比两边可以发现

$$\begin{cases}\beta_i=0,1\leqslant i\leqslant s\\ f_i(\mathscr{A})\beta_i=0,s+1\leqslant i\leqslant n\end{cases} \tag{4.120}$$

即

$$V = \text{Span}(\alpha_1, \cdots, \alpha_n) \subset \langle \beta_1, \cdots, \beta_n \rangle_{\mathscr{A}} \tag{4.121}$$
$$= \langle \beta_{s+1} \rangle_{\mathscr{A}} + \cdots + \langle \beta_n \rangle_{\mathscr{A}} \tag{4.122}$$
$$\Longrightarrow V = \langle \beta_{s+1} \rangle_{\mathscr{A}} + \cdots + \langle \beta_n \rangle_{\mathscr{A}} \tag{4.123}$$

由于 d_i 次多项式 $f_i(\lambda)$ 零化 β_i，故 $\dim\langle\beta\rangle_{\mathscr{A}} \leqslant d_i$，根据上面和的关系，可知必然取等（同时可知上面的和实际上是直和），即 $f_i(\lambda)$ 是 β_i 的极小多项式，因此可以取基 $B_i = (\beta_i, \cdots, \mathscr{A}^{d_i-1}\beta_i)$，$B = \bigcup_{i=s+1}^n B_i$ 是 V 的一组基，且 $\mathscr{A}$ 在这一组基下的矩阵为友方阵组成的准对角阵.

注 $D(\lambda)$ 的对角元中没有 0 是由 $\det(\lambda I - A) = \varphi_{\mathscr{A}}(\lambda) \neq 0$ 保证的.

事实上，如果

定理 4.9

设 $V/\mathbb{F}$，$\dim V = n < \infty$，$\mathscr{A} \in \mathcal{L}(V, \mathbb{F})$，$A$ 为 $\mathscr{A}$ 在基 $M = (\alpha_1, \cdots, \alpha_n)$ 下的矩阵. 设 $A(\lambda)$ 的 Smith 标准型为

$$D(\lambda) = \text{diag}(1, \cdots, 1, d_{s+1}(\lambda), \cdots, d_n(\lambda)), \tag{4.124}$$

其中每个 $d_i(\lambda)$ 为首一多项式，次数为 $d_i \geqslant 1$，则

1. $V = \bigoplus_{i=s+1}^{n} U_i$，其中 $U_i = \langle\beta_i\rangle_{\mathscr{A}}$，$d_i(\lambda) = d_{\mathscr{A},\beta_i}(\lambda)$.
2. A 相似于矩阵 $B = \text{diag}(B_{s+1}, \cdots, B_n)$，其中 B_i 为友方阵，由 A 唯一确定，B 称为 A 的有理标准型.

上面的定理是引理的特殊情况. 通过该定理可以发现，数阵的特征方阵唯一确定了有理标准型. 作为上面定理的推论，可以得到命题

命题 4.16

设 $A, B \in \mathbb{F}^{n\times n}$，则

$$A \overset{s}{\sim} B \Longleftrightarrow \lambda I - A \sim \lambda I - B. \tag{4.125}$$

证明 只需证明"$\Leftarrow$".

$$\lambda I - A \sim \lambda I - B \tag{4.126}$$
$$\Longleftrightarrow A(\lambda), B(\lambda)\text{相抵于同一 Smith 标准型} \tag{4.127}$$
$$\Longleftrightarrow A(\lambda), B(\lambda)\text{具有相同的不变因子} \tag{4.128}$$
$$\Longrightarrow A, B\text{具有相同的有理标准型} \tag{4.129}$$
$$\Longleftrightarrow A \overset{s}{\sim} B. \tag{4.130}$$

同样作为标准型，似乎 Jordan 标准型要优于有理标准型，但是在前面讨论 Jordan 标准型时，我们默认 $\mathbb{F}$ 为代数闭域（类比复数域 $\mathbb{C}$），因此对一般的域上的矩阵，未必有 Jordan 型，但必然有有理标准型.

关于矩阵的相似，有如下有趣的命题

命题 4.17

设 $A, B \in \mathbb{F}^{n\times n}$，且数域 $K \supset F$，则

$$A, B\text{在 F 上相似} \Longleftrightarrow A, B\text{在 K 上相似} \tag{4.131}$$

特别地，如果考虑实数域与复数域，可得

$$A, B\text{实相似} \Longleftrightarrow A, B\text{复相似}. \tag{4.132}$$

证明

$$
\begin{aligned}
&A \overset{s}{\sim} B(\mathbb{F}) &(4.133)\\
\Longleftrightarrow &A(\lambda) \sim B(\lambda)(\mathbb{F}) &(4.134)\\
\Longleftrightarrow &A(\lambda), B(\lambda)\text{行列式因子相同}(\mathbb{F}) &(4.135)\\
\Longleftrightarrow &A(\lambda), B(\lambda)\text{行列式因子相同}(\mathbb{K}) &(4.136)\\
\Longleftrightarrow &A(\lambda) \sim B(\lambda)(\mathbb{K}) &(4.137)\\
\Longleftrightarrow &A \overset{s}{\sim} B(\mathbb{K}) &(4.138)
\end{aligned}
$$

注 $\mathbb{F}$ 上的一组多项式的最大公因式与将它们看作扩域 $\mathbb{K}$ 上的多项式的最大公因式相同（都通过有限次辗转相除求出），从而中间行列式因子在不同域上等价.

最后以一个命题结尾，它带给了我们另一种看待特征多项式与极小多项式的视角.

命题 4.18

设 $V/\mathbb{F}$，$\mathscr{A} \in \mathcal{L}(V)$，$A$ 为 $\mathscr{A}$ 在某组基下的矩阵，设 $A(\lambda) = \lambda I - A$ 的 Smith 标准型为

$$\operatorname{diag}(d_1(\lambda), \cdots, d_n(\lambda)), \tag{4.139}$$

其中 $d_i(\lambda)$ 为 $A(\lambda)$ 的不变因子，则

1. $\varphi_{\mathscr{A}}(\lambda) = d_1(\lambda)\cdots d_n(\lambda)$，$d_{\mathscr{A}}(\lambda) = d_n(\lambda)$.
2. $\varphi_{\mathscr{A}}(\lambda) = d_{\mathscr{A}}(\lambda) \Longleftrightarrow V$ 为循环子空间.

♠

证明

1. $\varphi_{\mathscr{A}}(\lambda) = D_n(\lambda) = d_1(\lambda)\cdots d_n(\lambda)$. 另一方面，由于 $A \overset{s}{\sim} \operatorname{diag}(A_1, \cdots, A_s)$ 为 A 的有理标准型，故每个 A_i 为 $d_i(\lambda)$ 的友方阵，因此
$$d_{\mathscr{A}}(\lambda) = \operatorname{lcm}(d_1(\lambda), \cdots, d_n(\lambda)) = d_n(\lambda). \tag{4.140}$$
2. $\varphi_{\mathscr{A}} = d_{\mathscr{A}}$，等价于 A 相似于友方阵，这是 $\mathscr{A}$ 在基
$$(\beta, \mathscr{A}\beta, \cdots, \mathscr{A}^{n-1}\beta) \tag{4.141}$$
下的矩阵，因此 V 为循环子空间.

注 如果从 Smith 标准型角度考虑，则 2 等价于 $A(\lambda)$ 的 Smith 标准型为

$$\operatorname{diag}(1, 1, \cdots, 1, d_n(\lambda)), \tag{4.142}$$

这等价于 A 的有理标准型仅有一个友方阵分块，即 A 为友方阵，V 为循环子空间.

4.3 实向量空间中的变换

4.3.1 复化与诸多不变

复化是将实向量空间嵌入复向量空间的过程，由此可以将二者联系起来.

定义 4.14 (实向量空间的复化)

实向量空间 V 的复化记为 $V_{\mathbb{C}} = V^2 = \{(u, v) : u, v \in V/\mathbb{R}\} = \{u + iv : u, v \in V/\mathbb{R}\}$，容易验证它在

1. 加法：$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$.
2. 数乘：$(a + bi)(u + iv) = (au - bv) + i(av + bu)$.

下为复向量空间.

♣

在后文中，除特别指出外，向量空间均视为实向量空间.

命题 4.19

设 V 为实向量空间，V_C 为其复化，若 $v_1,\cdots,v_n$ 为 V 的一组基，则它也是 $V\mathbb{C}$ 的一组基，并且 $\dim V_C=\dim V(:=\dim V/\mathbb{R})$.

证明 设 $v_1,\cdots,v_n$ 为 V 的一组基，那么它们在 V_C 中的张成组包含 $v_1,\cdots,v_n,iv_1,\cdots,iv_n$，由此 $v_1,\cdots,v_n$ 能够张成 V_C；另一方面，设 $a_1v_1+\cdots+a_nv_n=0$，这里 $a_k\in\mathbb{C}$，则分别取实部与虚部可知

$$(\operatorname{Re}a_1)v_1+\cdots+(\operatorname{Re}a_n)v_n=(\operatorname{Im}a_1)v_1+\cdots+(\operatorname{Im}a_n)v_n=0 \tag{4.143}$$

但 $v_1,\cdots,v_n$ 在 V 中线性无关，由此 $a_1=\cdots=a_n=0$，得证.

除了空间的复化，线性变换也可以复化.

定义 4.15 (线性变换的复化)

设 V 为实向量空间，$T\in\mathcal{L}(V)$，则定义 T 的复化 $T_C\in\mathcal{L}(V_C)$，使得

$$T_C(u+iv)=T(u)+iT(v). \tag{4.144}$$

可以看出，复化并不是多么神秘的概念，只是将实线性变换嵌入到复线性空间中考虑罢了.

命题 4.20

设 $v_1,\cdots,v_n$ 为 V 的一组基，$T\in\mathcal{L}(V)$，则 T,T_C 在这组基下的矩阵相同.

有了基本的定义，可以进一步讨论实变换的特征值理论. 由于

$$(T_C)^n(u+iv)=T^nu+iT^nv, \tag{4.145}$$

因此对任意 $f\in\mathbb{R}[x]$，$f(T_C)=f(T)$，由此可以得到两个推论

推论 4.3

设 V 为实向量空间，$T\in\mathcal{L}(V)$，则 $d_T(\lambda)=d_{T_C}(\lambda)$，即二者的极小多项式相同.

推论 4.4

设 V 为实向量空间，$T\in\mathcal{L}(V)$，$\lambda\in\mathbb{R}$，则 $\lambda\in\operatorname{spec}T$ 当且仅当 $\lambda\in\operatorname{spec}T_C$.

命题 4.21

设 V 为实向量空间，$T\in\mathcal{L}(V),\lambda\in\mathbb{C},j\geqslant 0,u,v\in V$，则

$$(T_C-\lambda I)^j(u+iv)=0\iff(T_C-\bar\lambda I)^j(u-iv)=0 \tag{4.146}$$

证明 只需归纳证明一个方面.$j=1$ 时显然成立，假设对 $j-1$ 成立，考虑 j 的情形，设

$$(T_C-\lambda I)^j(u+iv)=(T_C-\lambda I)^{j-1}((T_C-\lambda I)(u+iv))=0 \tag{4.147}$$

记 $\lambda=a+bi$，则

$$(T_C-\lambda I)(u+iv)=(Tu-au+bv)+i(Tv-av-bu) \tag{4.148}$$

$$(T_C-\bar\lambda I)(u+iv)=(Tu-au+bv)-i(Tv-av-bu) \tag{4.149}$$

根据 $(T_C-\lambda I)^j(u+iv)=0$ 可知 $Tu-au+bv=Tv-av-bu=0$，因此 $(T_C-\bar\lambda I)^j(u-iv)=0$.

推论 4.5

设 V 为实向量空间，$T\in\mathcal{L}(V)$，则 $\lambda\in\mathbb{C}$ 为 T_C 的特征值当且仅当 $\bar\lambda$ 是 T_C 的特征值. 进一步，二者作为 T_C 特征值的重数相同.

推论 4.6

设 V 为实向量空间，$T \in \mathcal{L}(V)$，则 $\varphi_T(\lambda) = \varphi_{T_C}(\lambda)$.

由此可知，实线性变换的特征多项式为实系数多项式，次数也为 $\dim V$，并且特征值恰好为 φ_T 的实根. 在此基础上也容易验证 Caylay-Hamilton 定理成立.

由于实化变换的特征值成对出现，这体现在多项式上就有

$$(x-\lambda)(x-\bar{\lambda}) = x^2 - 2\mathrm{Re}\,(\lambda) + |\lambda|^2 \in \mathbb{R}[x] \tag{4.150}$$

这实际上说明了实向量空间的一个重要特点：最小的不变子空间未必是一维的，有可能是二维的.

例 4.4 设 V 为实向量空间，$T \in \mathcal{L}(V)$，$T_C(u+iv) = (a+ib)(u+iv)$，则对比实部与虚部可知

$$T(u,v) = (au-bv, av+bu) = (u,v)\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \tag{4.151}$$

因此 $W = \mathrm{Span}(u,v)$ 是 T-不变子空间，并且在 W 中不存在维数更小的不变子空间. 这相应到 T_C 实际上是 $E(a+bi,T_C) \oplus E(a-bi,T_C) = \mathrm{Span}(u+iv,u-iv) = \mathrm{Span}(u,v)$.

下面讨论实方阵的实相似，并给出两种实相似标准型.

4.3.2 实相似

在 λ 矩阵部分已经证明过如下结论，这里采用另一种方法也可以得到：实方阵实相似当且仅当其复相似.

命题 4.22

设 $A,B \in \mathbb{R}^{n\times n}$，则 $A \sim B(\mathbb{C})$ 当且仅当 $A \sim B(\mathbb{R})$.

证明 设 $AP = PB$，$P = P_1 + iP_2$，$P_1, P_2 \in \mathbb{R}^{n\times n}$，则有 $AP_1 = P_1B, AP_2 = P_2B$，进一步对任意 $x \in \mathbb{R}$，都有

$$A(P_1 + xP_2) = (P_1 + xP_2)B \tag{4.152}$$

设 $f(x) = \det(P_1 + xP_2) \in \mathbb{R}[x]$，则必然存在 x_0 使得 $f(x_0) \neq 0$（因为将 f 在 $\mathbb{C}[x]$ 中考虑时，有 $f(i) \neq 0$，故 f 不是零多项式，因此零点仅有限个），因此有 $A = (P_1 + xP_2)B(P_1 + xP_2)^{-1}$，得证.

下面讨论一个重要问题：实相似标准型. 前文的内容给出了一种复相似标准型——Jordan 型，实方阵往往不具有这样简洁的标准型，但可以得到与 Jordan 型类似的结论，首先考虑下例.

例 4.5 设 $z = a+bi \in \mathbb{C}$，$a,b \in \mathbb{R}, b \neq 0$，考虑与

$$A = \begin{pmatrix} z & \\ & \bar{z} \end{pmatrix} = \begin{pmatrix} a+bi & \\ & a-bi \end{pmatrix} \tag{4.153}$$

相似的实方阵.

4.4 一些例子

命题 4.23

任意 $A \in \mathbb{C}^{n\times n}$ 都可以分解为 $A = B + C$，其中 B 可对角化，C 幂零，$BC = CB$，进一步地，这种分解是唯一的.

证明 不妨设 A 为 Jordan 标准型，则令 $B = \mathrm{diag}(\lambda_1 I_{n_1}, \cdots, \lambda_s I_{n_s})$，显然 B 可对角化（事实上这就是对角阵），C 幂零，且计算可知 $BC = CB$（二者都是分块对角阵，因此只需计算每个分块，而根据单位阵的可交换性易知每个分块都是可交换的），因此只需要证明唯一性. 设 $A = B + C = B' + C'$，B, C 如上构造，则

$$A - B = C = (B' - B) + C' \tag{4.154}$$

由于 $A-B=C$ 仅有零特征值，因此其幂零，而 B',C' 可交换，$A=B'+C'$，因此 A,B',C' 两两可交换，进而 $A-B=C$ 与 C' 可交换且均幂零，这说明 $B'-B=C-C'$ 幂零（设 $C^m=C'^n=0$，则 $(C-C')^{m+n}=0$），但 $B'-B$ 可对角化，因此存在可逆阵 Q 使得 $Q(B-B')Q^{-1}=0$，即 $B'=B$，易得 $C'=C$，故这种分解是唯一的.

推论 4.7

任意 $A\in\mathbb{C}^{n\times n}$ 可逆都能分解为 $A=BC$，其中 B 可对角化，C 特征值均为 1，$BC=CB$，进一步地，这种分解是唯一的.

证明 由上一条命题可知 $A=B+C$，并且由 A 可逆能推出 B 可逆，即有 $A=B+C=B(I+B^{-1}C)$，这里 B 可对角化，$B,I+B^{-1}C$ 可交换并且 $B^{-1}C$ 幂零，即 $I+B^{-1}C$ 特征值均为 1，由 B,C 的唯一性可知 $B,I+B^{-1}C$ 也是唯一的，命题得证.

关于上面的加法分解，还有一个更强的版本.

命题 4.24

设 $A\in\mathbb{C}^{n\times n}$，则存在 $f_S,f_N\in\mathbb{C}[x]$，使得 $A=f_S(A)+f_N(A)$，并且 $f_S(A)$ 可对角化，$f_N(A)$ 幂零，并且这种分解是唯一的.

证明 不妨设 $A=\operatorname{diag}(A_1,\cdots,A_s)$ 为 Jordan 型矩阵，其中每个分块依次为相应于互异特征值 $\lambda_1,\cdots,\lambda_s$ 的所有 Jordan 块，设 $A_1,\cdots,A_s$ 的极小多项式分别为 $d_1,\cdots,d_s$，这 s 个多项式两两互素，根据中国剩余定理，存在 $f_S\in\mathbb{C}[x]$，使得 $f_S(\lambda)\equiv\lambda_i\lambda(\operatorname{mod} d_i(\lambda))$，因此

$$\begin{aligned} f_S(A)&=\operatorname{diag}(f_S(A_1),\cdots,f_S(A_s)) &(4.155)\\ &=\operatorname{diag}(\lambda_1 I_{n_1},\cdots,\lambda_s I_{n_s}) &(4.156)\end{aligned}$$

即 $f_S(A)$ 可对角化，并且取 $f_N(\lambda)=\lambda-f_S(\lambda)$，可知 $f_N(A)$ 为幂零矩阵，而唯一性的证明方法是相同的（或者利用中国剩余定理构造的唯一性直接得到）.

命题 4.25

设 $T\in\mathcal{L}(V)$，则存在 $\beta\in V$ 使得 $d_{T,\beta}(\lambda)=d_T(\lambda)$.

定理 4.10

设 $T\in\mathcal{L}(V)$，则 V 可以写作 T-循环子空间的直和.

证明 根据引理，存在 $\alpha\in V$ 使得 $d_{T,\alpha}(\lambda)=d_T(\lambda)$，设 $U_0=\langle\alpha\rangle_T=\mathbb{F}[T](\alpha)$，并且令

$$\mathcal{S}=\{U\leqslant V:T(U)\subset U,U\cap U_0=\{0\}\},\tag{4.157}$$

取 $\mathcal{S}$ 中极大元 U_1，令 $W=U_1\oplus U_0$，只需证明 $W=V$（根据定义，这里的直和是显然的）. 下面采用反证法，尝试构造出更大的 $U_1\supsetneq U\in\mathcal{S}$，推出矛盾.

若不然，任取 $\beta\in V\backslash W$，存在多项式 $f\in\mathbb{F}[x]$ 使得 $f(T)(\beta)\in W$（否则可知

$$\langle\beta\rangle_T\cap U_0\subset\langle\beta\rangle_T\cap W=\{0\},\quad U_1<U_1\oplus\langle\beta\rangle_T\tag{4.158}$$

这与 U_1 的极大性矛盾），令 $\tilde{d}$ 为满足上述条件的次数最小的首一多项式，则显然 $\tilde{d}(\lambda)\big|f(\lambda)$，特别地，由于 $d_T(T)(\beta)=0\in U_0$，因此 $\tilde{d}(\lambda)\big|d_T(\lambda)$，故设 $d_T(\lambda)=\tilde{d}(\lambda)p(\lambda)$.

由于 $\tilde{d}(T)(\beta)\in W$，因此存在唯一的 $u_1\in U_1,u_0\in U_0=\langle\alpha\rangle_T$，使得

$$\tilde{d}(T)(\beta)=u_1+u_0=u_1+g(T)(\alpha),\tag{4.159}$$

两边同时作用 $p(T)$，可得

$$0=\underbrace{p(T)u_1}_{\in U_1}+\underbrace{p(T)g(T)(\alpha)}_{\in U_0}\tag{4.160}$$

因此 $p(T)g(T)(\alpha)=0$（直和），这说明 $d_T=d_{T,\alpha}\big|p(\lambda)g(\lambda)$，即 $\tilde{d}p\big|pg$，$\tilde{d}\big|g$，再令 $g(\lambda)=\tilde{d}(\lambda)q(\lambda)$，以及（$\beta_1$ 删去了 β 中的 U_0 部分）

$$\beta_1=\beta-q(T)(\alpha), \tag{4.161}$$

则

$$\tilde{d}(T)(\beta_1)=\tilde{d}(T)(\beta)-\tilde{d}(T)q(T)(\alpha)=\tilde{d}(T)(\beta)-g(T)(\alpha)=u_1\in U_1, \tag{4.162}$$

令

$$U_2=U_1+\langle\beta_1\rangle_T=U_1+\mathbb{F}[T](\beta_1) \tag{4.163}$$

则 $U_2\subseteq U_1$（否则 $\beta_1\in W$ 导致 $\beta\in W$，矛盾），再证明 $U_2\cap U_0=\{0\}$，对任意 $u\in U_2\cap U_0$，设

$$u=v_1+h(T)(\beta_1)\in U_2\quad(v_1\in U_1) \tag{4.164}$$

由于 $u\in U_0$，因此 $h(T)(\beta_1)=u-v_1\in U_0\oplus U_1=W$，同时有

$$h(T)(\beta)=h(T)(\beta_1)+h(T)q(T)(\alpha)\in W \tag{4.165}$$

这说明 $\tilde{d}\big|h$，即 $h(T)(\beta_1)\in U_1$，上面的结果说明 $u\in U_1\cap U_0\cap U_2$，因此 $u=0$.

可以看出，$U_2\in\mathcal{S}$，并且 $U_2\subsetneq U_1$，这与 U_1 的极大性矛盾，因此不存在 $0\neq v\in V\backslash W$，也即 $V=W$，得证.

第 5 章 内积空间

本节讨论线性空间上定义了内积之后的结构：内积空间. 主要分为实内积空间和复内积空间. 我们首先讨论内积空间的一般理论，然后讨论 Euclid 空间与酉空间中的独特理论.

本章中，若无特殊说明，$\mathbb{F}$ 均表示 $\mathbb{R}$ 或 $\mathbb{C}$.

5.1 内积空间

5.1.1 内积与范数

在向量空间定义内积，可以将其变成内积空间.

定义 5.1 (内积)

称 $V/\mathbb{F}$ 上的二元函数 $\langle\cdot,\cdot\rangle : V^2 \to \mathbb{F}$ 为内积，若满足

1. 正定性：对任意 $v \in V : \langle v, v\rangle \geqslant 0$，取等当且仅当 $v = 0$.
2. 第一个位置的线性：对任意 $u_1, u_2, v \in V, a, b \in \mathbb{F} : \langle au_1 + bu_2, v\rangle = a\langle u_1, v\rangle + b\langle u_2, v\rangle$.
3. 共轭对称性：对任意 $u, v \in V : \langle u, v\rangle = \overline{\langle v, u\rangle}$.

♣

由共轭对称性与第一个位置的线性可以推出第二个位置的共轭线性，特别当 $\mathbb{F} = \mathbb{R}$ 时，由于实数的共轭仍然为自身，此时内积变为对称双线性函数，称为实内积. 有了内积，就可以给出内积空间的定义.

定义 5.2 (内积空间)

称带有内积的向量空间为内积空间.

♣

内积会自然诱导出范数，范数是内积空间中衡量“距离”的一个指标，它是一个满足三条定义的函数.

定义 5.3 (范数)

称线性空间 V 上的函数 $||\cdot|| : V \to \mathbb{R}^{\geqslant 0}$ 为范数，若其满足

1. 正定性：对任意 $v \in V : ||v|| \geqslant 0$，取等当且仅当 $v = 0$.
2. 齐次性：对任意 $c \in \mathbb{F}, v \in V : ||cv|| = |c|\,||v||$.
3. 三角不等式：对任意 $u, v \in V : ||u + v|| \leqslant ||u|| + ||v||$.

♣

注 由于范数的齐次性蕴含正定性，因此上面定义中的正定性可改为非负性，不过为了方便以及与内积的定义对比，故直接写为正定性.

内积可以自然诱导出范数，即范数 $||v|| = \sqrt{\langle v, v\rangle}$，容易验证这确实是一个范数，在此只证其三角不等式.

命题 5.1 (三角不等式)

设 $V/\mathbb{F}$ 为内积空间，$||\cdot||$ 为内积诱导的范数，则对任意 $x, y \in V$

$$||x + y|| \leqslant ||x|| + ||y||. \tag{5.1}$$

♠

证明

$$\begin{aligned}
||u + v||^2 &= ||u||^2 + ||v||^2 + 2\mathrm{Re}\,\langle u, v\rangle && (5.2)\\
&\leqslant ||u||^2 + ||v||^2 + 2|\langle u, v\rangle| && (5.3)\\
&\leqslant ||u||^2 + ||v||^2 + 2||u||\,||v|| && (5.4)\\
&= (||u| + ||v||)^2 && (5.5)
\end{aligned}$$

得证，同时可知等号成立当且仅当 u,v 共线且同向.

举个例子，对于 $V=\mathbb{F}^n$ 的情形，可以定义出我们熟悉的 Euclid 内积与 Euclid 范数.

定义 5.4 (Euclid 内积/范数)

设 $V=\mathbb{F}^n$ 为 $\mathbb{F}$ 向量空间，则对于 $x=(x_1,\cdots,x_n),y=(y_1,\cdots,y_n)\in V$，定义 Euclid 内积为

$$\langle x,y\rangle=\sum_{k=1}^{n}x_k\overline{y_k} \tag{5.6}$$

由之诱导的 Euclid 范数为

$$||x||=\sqrt{\sum_{k=1}^{n}|x_k|^2} \tag{5.7}$$

♣

通常也使用 $|\cdot|$ 代替 $||\cdot||$ 表示欧式范数.

定义内积之后，就可以讨论一个重要概念：正交性. 如果从 $\mathbb{R}^2,\mathbb{R}^3$ 的几何直观来看，正交代表直角、垂直，而更广义的正交实际上也是这种不相干性、无关性的推广，后面我们会证明，正交性与勾股定理的联系.

定义 5.5 (正交性)

称内积空间 V 中两个 u,v 正交，若 $\langle u,v\rangle=0$，记为 $u\perp v$.

♣

利用内积的共轭线性性可知，若 u 与 v 正交，则 v 与 u 正交，这表明在正交性的叙述中向量的次序无关紧要；并且 0 与任何向量正交，同时也是唯一与任何向量都正交的向量（根据内积的正定性，0 也是 V 中唯一与自身正交的向量）. 类比几何中的勾股定理，内积空间中对于正交向量也存在类似的结论，可以看作勾股定理的抽象推广.

定理 5.1 (勾股定理)

对任意正交的 $x,y\in V$，都有 $||x+y||^2=||x||^2+||y||^2$.

♡

证明 正交性告诉我们 $\langle x,y\rangle=\langle y,x\rangle=0$，因此直接计算可得

$$||x+y||^2=\langle x+y,x+y\rangle=\langle x,x\rangle+\langle x,y\rangle+\langle y,x\rangle+\langle y,y\rangle=||x||^2+||y||^2. \tag{5.8}$$

注 如果条件改为 k 个两两正交的向量 $x_1,\cdots,x_k$，则使用相同的方法可以得到进一步的勾股定理.

从上面的命题可以看出，如果一个向量可以分解为相互正交的向量之和，那么计算其范数只需要单独考虑各个分量. 如果考虑两个非零向量 u,v，则可以将向量 u 分解为沿向量 $v\neq 0$ 方向以及与 v 正交的方向上的向量之和，这称为正交分解. 如果在 $\mathbb{R}^2,\mathbb{R}^3$ 中，正交分解具有非常明显的几何直观，我们总可以作垂线来构造这样的分解.

命题 5.2 (正交分解)

设 $u,v\in V/\mathbb{F},v\neq 0$，则存在 $c\in\mathbb{F},w\perp v$ 使得

$$u=cv+w \tag{5.9}$$

这称为 u 在 v 方向上的正交分解，其中 $w\perp v$.

♠

证明 最直接的证明方法就是构造出满足条件的 c,w，再来证明唯一性. 假设存在这样的分解，那么必有

$$\langle u,v\rangle=\langle cv+w,v\rangle=c||v||^2, \tag{5.10}$$

反过来 $c=\langle u,v\rangle/||v||^2$，$w=v-cu$，则显然 $\langle w,v\rangle=\langle v,v\rangle-c\langle u,v\rangle=0$，即 $w\perp v$. 再来证明唯一性，假设 $u=c_1v+w_1=c_2v+w_2$，则 $(c_1-c_2)v=w_2-w_1$，两边同时与 v 做内积，可知 $c_1=c_2$，继而得到 $w_1=w_2$，唯一性得证.

正交分解有很大用处，在后面的正交基、最优化等内容中都有体现，不过在此，我们可以借助正交分解很容

易证明内积空间中的 Cauchy-Schwarz 不等式.

命题 5.3 (Cauchy-Schwarz 不等式)

设 $V/\mathbb{F}$ 为内积空间，$||\cdot||$ 为内积诱导的范数，则对任意 $x,y\in V$ 都有 $|\langle x,y\rangle|\leqslant ||x||\,||y||$.

证明 当 $y=0$ 时，显然成立，若不然，则令

$$x=\frac{\langle x,y\rangle}{||y||^2}y+z \tag{5.11}$$

两边同时与 y 做内积可知 $z\perp y$，因此

$$\begin{aligned}||x||^2&=\left|\left|\frac{\langle x,y\rangle}{||y||^2}y\right|\right|^2+||z||^2 &(5.12)\\&=\frac{|\langle x,y\rangle|^2}{||y||^2}+||z||^2 &(5.13)\\&\geqslant\frac{|\langle x,y\rangle|^2}{||y||^2} &(5.14)\end{aligned}$$

移项得证，同时可知等号成立当且仅当 u,v 共线.

例 5.1 Cauchy-Schwarz 不等式在具体例子中有许多不同的形式：

1. 对于任意 $x_1,\cdots,x_n,y_1,\cdots,y_n\in\mathbb{R}$，有
$$\left(\sum_{k=1}^n x_ky_k\right)^2\leqslant\left(\sum_{k=1}^n x_k^2\right)\left(\sum_{k=1}^n y_k^2\right). \tag{5.15}$$
2. 对于闭区间 $[a,b]$ 上的 Riemann 可积函数 f,g，有
$$\left(\int_a^b f(x)g(x)\mathrm{d}x\right)^2\leqslant\left(\int_a^b f^2(x)\mathrm{d}x\right)\left(\int_a^b g^2(x)\mathrm{d}x\right). \tag{5.16}$$

5.1.2 正交基与正交化

基是研究线性空间的神器，因此本节我们通过讨论内积空间中的一组特殊的基：正交基来进一步揭示内积空间的结构，为简单起见，无特殊说明的情况下本节讨论空间都是有限维线性空间.

在上一节中，我们看到了正交向量的良好性质，特别是正交分解. 由于有限维线性空间中的任何向量都可以唯一表示为基的线性组合（也可以理解为分解到这些向量的方向上），线性无关性本身包含了“无关性”，若要进一步得到正交意义下的无关性则自然希望寻找一组基，其中向量两两正交，这样就能将每个向量分解到相互正交的方向上，进而简化内积意义下许多问题的讨论. 在讨论之前，作为猜测合理性的保证，可以证明

命题 5.4

设非 0 向量组 $v_1,\cdots,v_m\in V$ 两两正交，则它们线性无关.

证明 设 $0=\sum\limits_{k=1}^m a_kv_k$，则对等式两边分别与 $v_1,\cdots,v_m$ 作内积，可得 $a_1=\cdots=a_m=0$，得证.

对于一个正交的向量组，可称之为正交组. 上面的命题表明正交组一定是无关组，而一定数量的无关组就是张成组，因此一定数量的正交组就是张成组，即一组基，这样的基就是正交基.

定义 5.6 (正交基)

设 $V/\mathbb{F}$ 为有限维内积空间，若 V 的一组基 B 中向量两两正交，则称 B 为 V 的一组正交基，特别地，若任意 $v\in B$ 都有 $||v||=1$，则称其为标准正交基（或规范正交基）.

虽然定义了正交基，但我们还没有证明正交基的存在性，下面我们给出一个非常美妙的定理，但定理的证明过程依赖一个由无关组构造正交组的过程（算法），称为 Gram-Schmidt 过程.

定理 5.2 (Gram-Schmidt 过程)

设 $v_1,\cdots,v_m$ 为 V 中的一个线性无关组，设 $e_1 = v_1/||v_1||$，对于 $j=2,\cdots,m$ 定义

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{||v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}||} \tag{5.17}$$

则 $e_1,\cdots,e_m$ 是 V 中的标准正交组，并且对任意 $j=1,\cdots,m$ 都有

$$\mathrm{Span}(v_1,\cdots,v_j) = \mathrm{Span}(e_1,\cdots,e_j). \tag{5.18}$$

♡

证明

对 V 中一组基进行 Gram-Schmidt 过程，即可得到 V 的一个基，因此有如下美妙的定理.

定理 5.3

任何有限维内积空间 V 都存在标准正交基. ♡

对于正交基，可以仿照勾股定理得到如下结论，它们在 Fourier 分析中使用更广泛.

定理 5.4 (Parseval 等式)

设 $e_1,\cdots,e_n$ 为 V 中一组标准正交基，则对任意 $v,w\in V$ 都有

$$\langle v, w\rangle = \sum_{k=1}^{n}\langle v, e_k\rangle\langle e_k, w\rangle. \tag{5.19}$$

特别的，当 $v=w$ 时，上面的等式可以看作勾股定理的抽象推广. ♡

定理 5.5 (Bessel 不等式)

设 $u_1,\cdots,u_m$ 为 V 中一组标准正交组，则对任意 $v\in V$ 都有

$$||v||^2 \geqslant \sum_{k=1}^{m}|\langle v, u_i\rangle|^2. \tag{5.20}$$

等号成立当且仅当 $u_1,\cdots,u_m$ 恰好为一组正交基. ♡

最后讨论一个关于线性泛函的问题. 内积具有 Hermite 对称性，但若固定第二个位置上的向量，那么它就变成了一个线性泛函，更精妙的是下面的结论，它表明借助与某一个向量的内积 $\langle -, u\rangle$ 可以表示任意线性泛函.

定理 5.6 (Riesz 表示定理)

设 V 为有限维内积空间，$\varphi\in V^*$，则存在唯一的向量 $u\in V$，使得对任意 $v\in V$ 都有 $\varphi(v)=\langle v, u\rangle$. ♡

证明 首先证明存在性，设 $e_1,\cdots,e_n$ 为 V 的标准正交基，则对任意 $v\in V$ 都有

$$\begin{aligned}
\varphi(v) &= \varphi\left(\sum_{k=1}^{n}\langle v, e_k\rangle e_k\right) &(5.21)\\
&= \sum_{k=1}^{n}\langle v, e_k\rangle\varphi(e_k) &(5.22)\\
&= \left\langle v, \sum_{k=1}^{n}\overline{\varphi(e_k)}e_k\right\rangle &(5.23)\\
&= \langle v, u\rangle &(5.24)
\end{aligned}$$

再证明唯一性，若存在 u_1,u_2 满足条件，则对任意 $v\in V$ 都有 $\varphi(v)=\langle v,u_1\rangle=\langle v,u_2\rangle$，因此 $\langle v,u_1-u_2\rangle=0$，这表明 $u_1=u_2$，得证.

5.1.3 正交矩阵与正交对角化

内积空间给相似也带来了新的生命. 我们在过去总讨论一般基之间的变换，如果考虑两组标准正交基之间的变换，比如二者的过渡矩阵，可以得到许多有趣的结果. 设 $(u_1,\cdots,u_n)=(v_1,\cdots,v_n)A$，其中 u,v 均为标准正交基，由于 $u_m=\sum\limits_{k=1}^{n}a_{km}v_k$，因此根据正交性可知

$$\sum_{k=1}^{n}|a_{km}|^2=1,\cdots,\sum_{k=1}^{n}a_{km}\overline{a_{kp}}=0 \tag{5.25}$$

如果设 $A=(\alpha_1,\cdots,\alpha_n)$，则上面的结果表明 $\alpha_1,\cdots,\alpha_n$ 是 $\mathbb{F}^{n\times 1}$ 中的一个正交基，这种矩阵就被称为正交矩阵. 借助共轭转置可以更好描述这种关系

$$\begin{pmatrix}\alpha_1^*\\ \vdots\\ \alpha_n^*\end{pmatrix}(\alpha_1,\cdots,\alpha_n)=I_n \tag{5.26}$$

借此给出正交矩阵的定义.

定义 5.7 (正交矩阵)

设 $P\in\mathbb{F}^{n\times n}$，若 $P^*P=I_n$，则称 P 为正交矩阵. ♣

正交矩阵有许多很好的，容易验证的性质.

命题 5.5

设 $P\in\mathbb{F}^{n\times n}$ 为正交阵，则

1. P 可逆，并且 $P^{-1}=P^*$.
2. $PP^*=I_n$.
3. 对任意 $x\in\mathbb{F}^{n\times 1}$，都有 $||Px||=||x||$. ♠

上一节中的 Gram-Schmidt 过程也有一个矩阵结果，如果将 $\mathbb{F}^{n\times 1}/\mathbb{F}$ 中的一组基 $(\alpha_1,\cdots,\alpha_n)$ 标准正交化为 $(e_1,\cdots,e_n)$，则显然二者之间的过渡矩阵为一个上三角阵，这说明每个可逆矩阵都可以化为一个正交阵与一个上三角阵的乘积，即

定理 5.7 (QR 分解)

设 $A\in\mathbb{F}^{n\times n}$ 可逆，则存在正交矩阵 $Q\in\mathbb{F}^{n\times n}$，上三角矩阵 $R\in\mathbb{F}^{n\times n}$，使得 $A=QR$，并且这种分解是唯一的. ♡

正交阵的良好性质使得其在相似理论中有一席之地.

定义 5.8 (正交相似)

设 $A,B\in\mathbb{F}^{n\times n}$，若存在正交阵 $P\in\mathbb{F}^{n\times n}$ 使得 $A=PBP^{-1}PBP^*=$，则称 A,B 正交相似. ♣

利用正交相似，可以得到真正的 Schur 三角化定理（将相似加强为正交相似）.

定理 5.8 (Schur 三角化)

设 $A\in\mathbb{F}^{n\times n}$，则存在正交阵 P，使得 $A=PTP^*$，其中 T 为上三角矩阵. 更准确来说，设 A 的可重特征值为 $\lambda_1,\cdots,\lambda_n$，则 T 的对角元为 $\lambda_1,\cdots,\lambda_n$. 特别地，如果将相同特征值排列在一起（设有 d 个相异特

征值)，则

$$T=\begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1d} \\ 0 & T_{22} & \cdots & T_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_{dd} \end{pmatrix} \tag{5.27}$$

其中 $T_{ii}\in\mathbb{F}^{n_i\times n_i}$ 均为上三角矩阵，且对角元为（重新排列后的）λ_i，n_i 为对应的代数重数.

5.1.4 正交变换

在内积空间中有一种特殊的线性变换，它保持内积不变，即对于任意 $v_1,v_2\in V$，都有 $\langle S(v_1),S(v_2)\rangle=\langle v_1,v_2\rangle$，取 $v_1=v_2$，则可得 $||S(v)||=||v||$，反过来借助极化恒等式可知等距与保内积是等价的.

命题 5.6 (极化恒等式)

设 V 为内积空间，则对于任意 $u,v\in V$

1. 若为实内积空间，则

$$\langle u,v\rangle=\frac{1}{4}(||u+v||^2-||u-v||^2). \tag{5.28}$$

2. 若为复内积空间，则

$$\langle u,v\rangle=\frac{1}{4}(||u+v||^2-||u-v||^2+i||u+iv||^2-i||x-iy||^2). \tag{5.29}$$

由此定义等距变换

定义 5.9 (等距变换/正交变换)

设 $S\in\mathcal{L}(V)$，若对任意 $v\in V$ 都有 $||S(v)||=||v||$，则称 S 是一个等距变换/正交变换.

等距变换最典型的例子就是旋转变换与反射变换，由此也可以看出，等距变换不会对向量进行拉伸或压缩，也就是说 $S(v)=0$ 当且仅当 $v=0$，这一方面表明等距变换必然是可逆的，另一方面从特征值角度考虑，这说明等距变换的特征值的模长均为 1.

命题 5.7

设 $S\in\mathcal{L}(V)$ 为等距变换，则 $\operatorname{spec}S\subset\{|\lambda|=1:\lambda\in\mathbb{F}\}$.

另外，等距变换的保内积性使得它在一组标准正交基下的像仍然是一组标准正交基，这种观察给出了等距变换与正交矩阵之间的联系，因此等距变换也被称为正交变换.

定理 5.9

设 $S\in\mathcal{L}(V)$ 为等距变换，则它在任一标准正交基下的矩阵为正交阵.

因此前面对于正交变换特征值的讨论同样适用于正交阵.

此时还有一个问题：既然通过取特殊的基可以简化映射的矩阵，那么能否选取特殊的正交基简化正交变换的矩阵？这个问题在实内积空间与复内积空间上有不同的答案，因此留到下一部分讨论.

5.1.5 正交补与正交投影

借助正交的概念，可以定义一种具有更好性质的补空间：正交补.

定义 5.10 (正交补)

与补空间不同，容易验证正交补是**唯**一的，在此基础上有许多良好的性质.

命题 5.8

正交补、各种维数、正交投影

5.1.6 极小化问题

下面讨论一些舒适的例子，这是内积、范数、正交等性质的重要应用之一.

正弦函数的逼近，最小二乘法

5.1.7 伴随变换

定义 5.11 (伴随变换)

设 $T \in \mathcal{L}(V,W)$，则定义 T 的伴随映射 $T^* \in \mathcal{L}(W,V)$，对任意 $v \in V, w \in W$ 有

$$\langle T(v), w\rangle = \langle v, T^*(w)\rangle. \tag{5.30}$$

注 上标 * 的记号在线性代数中共有三处用到：行列式中的伴随矩阵、伴随变换与伴随矩阵、对偶空间，其中第一个用到较少，后两个明显可以区分，因此通常不会矛盾.

容易验证，伴随变换是线性变换，另一个重要问题是伴随变换的良定性，假设有 $f, g \in \mathcal{L}(W,V)$ 都是 $T \in \mathcal{L}(V,W)$ 的伴随，则对任意 $v \in V, w \in W$ 有

$$\langle T(v), w\rangle = \langle v, f(w)\rangle = \langle v, g(w)\rangle. \tag{5.31}$$

则 $\langle v, (f-g)(w)\rangle = 0$，这说明 $f = g$，即若伴随变换存在，那么必然是唯一的. 除了这种方法，还可以直接利用 Riesz 表示定理，构造映射 $f_u(v) = \langle Tv, u\rangle$，可知存在 $\tilde{u}$ 使得 $f_u(v) = \langle v, \tilde{u}\rangle = \langle Tv, u\rangle$，这里的 $\tilde{u} = T^*u$.

容易验证伴随变换的一些性质

命题 5.9

设 $S, T \in \mathcal{L}(V,W)$，$G \in \mathcal{L}(W,V)$ 则

1. $(T^*)^* = T$.
2. $(S+T)^* = S^* + T^*$.
3. $(\lambda T)^* = \bar{\lambda} T^*$.
4. $(GT)^* = T^* G^*$.

事实上，伴随变换最重要的性质就是它在一组基下的矩阵恰好为原变换的共轭转置.

定理 5.10

设 $T \in \mathcal{L}(V,W)$，T 在 V, W 中基 $v_1, \cdots, v_n$，$w_1, \cdots, w_m$ 下的矩阵为 A，则 $T^* \in \mathcal{L}(V,W)$ 在相同基下的矩阵为 A^*，即 A 的共轭转置.

根据上述定理，也可以发现伴随变换的性质实际上就是共轭转置的性质.

如果从正交补的角度考虑，可以得到伴随变换的另一种刻画.

命题 5.10

设 $T \in L(V,W)$，则

$$\operatorname{Ker} T = (\operatorname{Im} T^*)^{\perp}, \quad \operatorname{Im} T = (\operatorname{Ker} T^*)^{\perp} \tag{5.32}$$

证明

$$\begin{aligned} v \in \operatorname{Ker} T &\Longleftrightarrow T(v) = 0 && (5.33)\\ &\Longleftrightarrow \langle T(v), w\rangle = 0, \forall w \in W && (5.34)\\ &\Longleftrightarrow \langle v, T^*(w)\rangle = 0, \forall w \in W && (5.35)\\ &\Longleftrightarrow v \in (\operatorname{Im} T^*)^{\perp} && (5.36)\end{aligned}$$

5.2 实/复内积空间的差别

上面讨论的大多是内积空间的共性，下面对实内积空间与复内积空间中的一些差别做一些讨论.

5.2.1 正规变换与自伴变换

内积空间中有两种特殊的变换，它们具有非常良好的性质.

定义 5.12 (正规变换/规范变换)

设 $T \in \mathcal{L}(V)$，若 $TT^* = T^*T$，则称 T 为正规变换（或规范变换），定义正规变换在一组标准正交基下的矩阵为正规阵，即满足 $AA^* = A^*A$ 的矩阵.

正规变换还有一种刻画. 首先考虑引理

引理 5.1

若 $T \in \mathcal{L}(V)$，且对任意 $v \in V$ 都有 $\langle Tv, v\rangle = 0$，则 $T = 0$.

证明 对任意 $u, v \in V$ 都有

$$\begin{aligned} \langle Tu, v\rangle &= \frac{\langle T(u+v), u+v\rangle - \langle T(u-v), u-v\rangle}{4} + i\frac{\langle T(u+iv), u+iv\rangle - \langle T(u-iv), u-iv\rangle}{4} && (5.37)\\ &= 0 && (5.38)\end{aligned}$$

取 $v = Tu$，可知 $T = 0$.

命题 5.11

$T \in \mathcal{L}(V)$ 是正规变换当且仅当对任意 $v \in V$ 都有 $||T(v)|| = ||T^*(v)||$.

证明

$$\begin{aligned} T^*T - TT^* = 0 &\Longleftrightarrow \langle (T^*T - TT^*)v, v\rangle = 0, \forall v \in V && (5.39)\\ &\Longleftrightarrow \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle, \forall v \in V && (5.40)\\ &\Longleftrightarrow \langle Tv, Tv\rangle = \langle T^*v, T^*v\rangle, \forall v \in V && (5.41)\\ &\Longleftrightarrow ||Tv|| = ||T^*v||, \forall v \in V && (5.42)\end{aligned}$$

得证.

下面考虑正规变换的一些简单性质.

命题 5.12

设 $T \in \mathcal{L}(V)$ 正规，则

1. T^* 也是正规变换.
2. 设 $f \in \mathbb{F}[x]$，则 $f(T)$ 也是正规变换.
3. $\lambda \in \operatorname{spec} T$ 当且仅当 $\bar{\lambda} \in \operatorname{spec} T^*$，并且 T 与 T^* 有相同的特征向量.
4. $\operatorname{Ker} T = \operatorname{Ker} T^*, \operatorname{Im} T = \operatorname{Im} T^*$.

下面考虑正规阵的一个性质，它为正规阵的可对角化性质埋下了伏笔.

命题 5.13

设 $A = \begin{pmatrix} A_1 & A_2 \\ & A_4 \end{pmatrix} \in \mathbb{F}^{n\times n}$，则 A 为正规阵当且仅当 A_1, A_4 为规范阵，$A_2 = 0$.

证明 一边是显然的. 设 A 为正规阵，则 $AA^* = A^*A$，这说明

$$AA^* = \begin{pmatrix} A_1A_1^* + A_2A_2^* & A_2A_4^* \\ A_4A_2^* & A_4A_4^* \end{pmatrix} = \begin{pmatrix} A_1A_1^* & A_1^*A_2 \\ A_2^*A_1 & A_2^*A_2 + A_4^*A_4 \end{pmatrix} = A^*A \tag{5.43}$$

可得 $A_2A_2^* = 0$，两边取 trace（或者考虑 Frobenius 范数）可知 $A_2 = 0$，命题得证.

推论 5.1

设 $T \in \mathcal{L}(V)$ 正规，$W \leqslant V$ 为 T-不变子空间，则

1. W 是 T^*-不变子空间.
2. $W^\perp$ 是 T-不变子空间.

证明 取 W 一组标准正交基 $v_1, \cdots, v_m$，并扩充为 V 的标准正交基 $v_1, \cdots, v_m, \cdots, v_n$，可知 T 在这一组基下的矩阵为

$$A = \begin{pmatrix} A_1 & A_2 \\ & A_4 \end{pmatrix} \tag{5.44}$$

而 T 正规，说明 A 为正规阵，根据引理可知 $A_2 = 0$，故 $W^\perp$ 为 T-不变，考虑 T^* 在同组基下的矩阵可知 $W, W^\perp$ 也是 T^* 不变的.

还有一种比伴随变换更特殊的变换，自伴变换.

定义 5.13 (自伴变换/对称变换)

设 $T \in \mathcal{L}(V)$，若 $T = T^*$，则称 T 为自伴变换（或对称变换），定义自伴变换在一组标准正交基下的矩阵为对称阵，即满足 $A = A^*$ 的矩阵.

注 与对称形式相似，将 $T = -T^*$ 的变换称为反对称变换，对应矩阵称为反对称阵.

显然的，自伴变换也是正规变换，因此自伴变换满足上面介绍的许多性质，这里仅讨论一些特殊的性质.

命题 5.14

$T \in \mathcal{L}(V)$ 自伴，则 T 在 V 中的每个特征值都是实数.

证明 若 T 自伴，$Tv = \lambda v$，则

$$\lambda||v||^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \bar{\lambda}||v||^2 \tag{5.45}$$

得证.

命题 5.15

设 $T\in\mathcal{L}(V)$，则 T 自伴当且仅当对任意 $v\in V$ 都有 $\langle Tv,v\rangle\in\mathbb{R}$.

证明 设 $v\in V$，则

$$\langle Tv,v\rangle-\overline{\langle Tv,v\rangle}=\langle Tv,v\rangle-\langle v,Tv\rangle=\langle Tv,v\rangle-\langle T^*v,v\rangle=\langle (T-T^*)v,v\rangle \tag{5.46}$$

若 T 自伴，显然成立；反之若 $\langle Tv,v\rangle\in\mathbb{R}$，则根据前面的引理可知 $T-T^*=0$，得证.

5.2.2 谱定理

从上一节可以看出正规变换与自伴变换的许多良好性质，事实上，它们还蕴含了可对角化的性质.

定理 5.11 (复谱定理)

设 $\mathbb{F}=\mathbb{C},T\in\mathcal{L}(V)$，则 T 为正规变换（即 $TT^*=T^*T$）当且仅当 V 有一个由 T 的特征向量构成的标准正交基（这显然等价于 T 在某组标准正交基下的矩阵可正交对角化）.

证明 根据 Schur 三角化定理，存在标准正交基 $e_1,\cdots,e_n$，使得 T 在其下的矩阵 A 为上三角阵，若 T 正规，则 A 为正规矩阵，根据正规矩阵的性质可知 A 为对角阵，即这组基是 T 的特征向量. 反之，若 V 有由 T 的特征向量构成的一组基，则 T 在这组基下的矩阵 B 为对角阵（也为正规阵），故 T 为正规变换.

定理 5.12 (实谱定理)

设 $\mathbb{F}=\mathbb{R},T\in\mathcal{L}(V)$，则 T 为自伴变换（即 $T=T^*$）当且仅当 V 有一个由 T 的特征向量构成的标准正交基（这显然等价于 T 在某组标准正交基下的矩阵可正交对角化）.

事实上，复谱定理可以推出实谱定理，但这里还是给出另一种较独立的证法.

证明 若 V 有由 T 的特征向量构成的一组基，则 T 在这组基下的矩阵 B 为对角阵（也为对称阵），故 T 为自伴变换.

反之归纳证明，对于 $\dim V=1$ 显然成立，设对 $\dim V<n$ 成立，考虑 $n+1$ 的情形. 由于 T 自伴，特征值均为实数，因此必然存在单位特征向量 u，由于 T 在 $W=\mathrm{Span}(u)$ 不变，因此它在 $W^\perp=\mathrm{Span}(u)^\perp$ 也不变，考虑 $T'=T|_{W^\perp}$，根据归纳假设，在 $W^\perp$ 中存在由 T' 的特征向量构成的基，将这组基与 u 合并，可得 V 中一组由 T 的特征向量构成的基，命题得证.

5.2.3 正规矩阵的进一步讨论

正规矩阵还有许多性质，在此多做一些讨论.

定理 5.13 (Schur 不等式)

设 $A\in\mathbb{F}^{n\times n}$，全部特征值为 $\lambda_1,\cdots,\lambda_n$，则

$$\sum_{k=1}^{n}|\lambda_i|^2\leqslant||A||_F^2 \tag{5.47}$$

当且仅当 A 正规时取等.

定义 5.14 (亏量)

设 $A\in\mathbb{F}^{n\times n}$，全部特征值为 $\lambda_1,\cdots,\lambda_n$，则定义

$$\Delta(A)=||A||_F^2-\sum_{k=1}^{n}|\lambda_i|^2=\mathrm{tr}A^*A-\sum_{k=1}^{n}|\lambda_i|^2 \tag{5.48}$$

为 A 偏离正规性的亏量.

根据 Schu 不等式，对任意方阵 A 都有 $\Delta(A) \geqslant 0$，当且仅当 A 为正规矩阵时亏量为 0.

引理 5.2

设 $A_1, \cdots, A_n$ 均为正规矩阵，则存在多项式 $p \in \mathbb{F}[x]$，对每个 $1 \leqslant k \leqslant n$ 都有 $A_k^* = p(A_k)$ ♡

定理 5.14 (Fuglede-Putnam 定理)

设 $A \in \mathbb{F}^{m\times m}, B \in \mathbb{F}^{n\times n}$ 均为正规矩阵，则对任意 $X \in \mathbb{F}^{m\times n}$，$AX = XB$ 当且仅当 $A^*X = XB^*$. ♡

5.2.4 正规 & 正交变换

特征值；
正规矩阵：相似到分块分块对角的形式（正交为特例）

5.2.5 实内积空间中的变换

5.3 二次型

本节将主要以实二次型讨论.

5.3.1 双线性函数

定义 5.15 (双线性函数)

称 $f \in \mathcal{L}(V^2, \mathbb{F})$ 为 V 上的一个双线性函数. 特别地，若 $f(u, v) = f(v, u)$，则称 f 是一个对称双线性函数，记为 $f \in S(V^2, \mathbb{F})$. ♣

仿照线性映射，我们希望找到矩阵能与双线性函数一一对应，因此定义独立矩阵的概念.

定义 5.16 (度量矩阵)

设 $f \in \mathcal{L}(V^2, \mathbb{F})$，取 V 中的基 $v_1, \cdots, v_n$，称矩阵

$$F = (f_{ij})_{n\times n} := (f(\xi_i, \xi_j))_{n\times n} = \begin{pmatrix} f_{11} & \cdots & f_{1,n} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nn} \end{pmatrix}. \tag{5.49}$$

为 f 在上述基下的度量矩阵. ♣

设 $u, v \in V$，取 V 的一组基，并设 $X, Y \in \mathbb{F}^{n\times 1}$ 表示 u, v 在这组基下的坐标，则

$$\begin{aligned} f(u, v) &= f\left(\sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j\right) && (5.50) \\ &= \sum_{i,j=1}^n x_i f(v_i, v_j) y_j = \sum_{i,j=1}^n x_i f_{ij} y_j && (5.51) \\ &= \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} f_{11} & \cdots & f_{1,n} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} && (5.52) \\ &= X^T F Y. && (5.53) \end{aligned}$$

如果考虑基变换

$$(v_1, \cdots, v_n)P = (v'_1, \cdots, v'_n) \tag{5.54}$$

设 f 在新基下的度量阵为 F'，则

$$\begin{aligned} F' = (f'_{ij})_{n\times n} &= (f(v'_i, v'_j))_{n\times n} && (5.55)\\ &= f\left(\sum_{s=1}^{n} a_{si}v_s, \sum_{t=1}^{n} a_{tj}v_t\right) && (5.56)\\ &= \sum_{s,t=1}^{n} a_{si}f_{st}a_{vj} && (5.57)\\ &= P^TFP. && (5.58)\end{aligned}$$

由此，得到了线性代数中继相抵、相似后的又一个重要等价关系：相合.

定义 5.17 (相合)

设 $F, G \in \mathbb{F}^{n\times n}$，若存在可逆矩阵 P 使得 $F = P^TGP$，则称 F, G 相合.

♣

相合也可以看作是一种特殊的相抵，具有相抵的一切性质，比如保秩，因此可以定义双线性函数的秩.

定义 5.18 (双线性函数的秩)

设 $f \in \mathcal{L}(V^2, \mathbb{F})$，则定义其在 V 中任何一组基下的度量矩阵的秩为 f 的秩，记为 $\mathrm{rk}f$. 若度量矩阵满秩，则称 f 为非退化的，反之称为退化的.

♣

根据上面的讨论，易得如下命题

命题 5.16

设 $f \in \mathcal{L}(V^2, \mathbb{F})$，则 f 对称当且仅当其在一组基下的度量矩阵 F 是对称的.

♠

在前面给出度量矩阵的定义后，自然有如下事实

定理 5.15

设 $f \in \mathcal{L}(V^2, \mathbb{F})$，则 $\mathcal{L}(V^2, \mathbb{F}) \cong \mathbb{F}^{n\times n}$，若设 V 的一组基 $B = (v_1, \cdots, v_n)$，则存在线性空间同构

$$\begin{aligned} \Phi_B : \mathcal{L}(V^2) &\longrightarrow \mathbb{F}^{n\times n} && (5.59)\\ f &\longmapsto A_f && (5.60)\end{aligned}$$

♡

也就是说，给定 V 上的一组基，则对任意 V 上的双线性函数 f，可以唯一确定 $\mathbb{F}^{n\times n}$ 中的一个矩阵，反之亦然，这个矩阵就是 f 在这组基下的度量矩阵.

证明 只需证明 Φ_B 即单又满. 由于 $\mathrm{Ker}\,\Phi_B = 0$，故单性是显见的，下面证明满性.

对任意 $A \in \mathbb{F}^{n\times n}$，存在映射 f_A，对任意 $u = \sum x_i\xi_i, v = \sum y_j\xi_j$，有

$$f_A(u, v) = X^TAY. \tag{5.61}$$

由此依次取基 B 中的每个向量，可得

$$f_A(\xi_i, \xi_j) = a_{ij}. \tag{5.62}$$

因此必然存在这样的 f_A，使得 $\Phi_B(f_A) = A$，满性即证.

推论 5.2

设 $f \in S(V^2, \mathbb{F})$，则 $\mathcal{L}(V^2, \mathbb{F}) \cong S_n(\mathbb{R})$，若设 V 的一组基 $B = (v_1, \cdots, v_n)$，则存在线性空间同构

$$\Phi_B : \mathcal{L}(V^2) \longrightarrow \mathbb{F}^{n\times n} \tag{5.63}$$

$$f \longmapsto A_f \tag{5.64}$$

由于双线性函数的自变量有两个向量，因此对任意 $v \in V$，都可以诱导出两个线性泛函，即视为固定其中一个位置的向量，对另一个位置的作用

$$L_f : V \longrightarrow V^* = \mathcal{L}(V,\mathbb{F}) \tag{5.65}$$
$$v \longmapsto f(v,-) \tag{5.66}$$
$$R_f : V \longrightarrow V^* = \mathcal{L}(V,\mathbb{F}) \tag{5.67}$$
$$v \longmapsto f(-,v) \tag{5.68}$$

考虑这两个线性算子，可以自然引出左/右正交补的概念

定义 5.19 (左/右正交补)

设 $v \in V, f \in \mathcal{L}(V^2,\mathbb{F})$，则定义 v 的左/右正交补为

$${}^{\perp}v := \operatorname{Ker} R_f(v) = \{u \in V : f(u,v) = 0\}, \tag{5.69}$$
$$v^{\perp} := \operatorname{Ker} L_f(v) = \{u \in V : f(v,u) = 0\}. \tag{5.70}$$

进一步可以定义集合 $S \subset V$ 的左/右正交补为

$${}^{\perp}S := \operatorname{Ker} R_f(S) = \{u \in V : f(u,S) = 0\}, \tag{5.71}$$
$$S^{\perp} := \operatorname{Ker} L_f(S) = \{u \in V : f(S,u) = 0\}. \tag{5.72}$$

特别的，$\operatorname{Ker} R_f = {}^{\perp}V, \operatorname{Ker} L_f = V^{\perp}$.

关于正交补，有以下有趣的性质：

命题 5.17

设集合 $S \subseteq V$，则

1. $S^{\perp} \leqslant V, {}^{\perp}S \leqslant V$.
2. $S \subseteq T$，则 $S^{\perp} \geqslant T^{\perp}, {}^{\perp}S \geqslant {}^{\perp}T$.
3. $({}^{\perp}(S^{\perp}))^{\perp} = S^{\perp}$.

下面回到对 L_f, R_f 本身的讨论. 它们是从 V 到 V 的对偶空间 V^* 的映射，这个映射本身依赖于 f 的选取，但下面的命题说明这种依赖是一一对应的.

命题 5.18

设 $V/\mathbb{F}$，$f \in \mathcal{L}(V^2,\mathbb{F})$，则

$$\mathcal{L}(V^2,\mathbb{F}) \cong \mathcal{L}(V,V^*), \tag{5.73}$$

或者说对任意 $\mathscr{A} \in \mathcal{L}(V,V^*)$，都有唯一 $f_{\mathscr{A}} \in \mathcal{L}(V^2,\mathbb{F})$ 与之对应.

$$\begin{tikzcd}
\mathcal{L}(V^2,\mathbb{F}) \arrow[r, leftrightarrow, "\sim"] \arrow[d, "\text{取定基}"'] & \mathcal{L}(V,V^*) \arrow[d, "\text{取定对偶基}"] \\
\mathbb{F}^{n\times n} & \mathbb{F}^{n\times n}
\end{tikzcd} \tag{5.74}$$

命题 5.19

设 $B=(v_1,\cdots,v_n)$ 为 V 的一组基，$B^*=(v^1,\cdots,v^n)$ 为 B 的对偶基，V 上双线性函数 f 在基 B 下的度量阵为 A，则 $R_f(L_f)$ 在基 B 以及 B^* 下的矩阵为 $A(A^T)$，即

$$R_f(v_1,\cdots,v_n)=(v^1,\cdots,v^n)A \tag{5.75}$$

$$L_f(v_1,\cdots,v_n)=(v^1,\cdots,v^n)A^T \tag{5.76}$$

证明 $(R_f(\xi_j))(\xi_i)=f(\xi_i,\xi_j)=a_{ij}$，故

$$R_f(\xi_j)=\sum_{i=1}^{n}a_{ij}\xi^i. \tag{5.77}$$

对于 L_f,R_f，根据线性映射基本定理可知

$$\dim V=\dim\operatorname{Ker}L_f+\dim\operatorname{Im}L_f=\dim\operatorname{Ker}L_f+\dim\operatorname{Im}L_f \tag{5.78}$$

$$=\dim V^{\perp}+\operatorname{rk}f=\dim{}^{\perp}V+\operatorname{rk}f \tag{5.79}$$

故 $\dim V^{\perp}=\dim{}^{\perp}V=\dim V-\operatorname{rk}f$.

下面以最后一个定理结尾.

定理 5.16

设 V 为有限维向量空间，$f\in\mathcal{L}(V^2,\mathbb{F})$，则 f 非退化当且仅当对任意 $\varphi\in V^*$，都存在唯一 $v\in V$，使得 $\varphi=L_f(v)=f(v,-)$.

证明 f 非退化 $\Longleftrightarrow$ L_f 是双射 $\forall\varphi\in V^*,\exists v\in V:L_f(v)=\varphi$，并且这样的 v 是唯一的.

上面定理的 $\Rightarrow$ 方向即为 Riesz 表示定理.

注

1. f 非退化等价于 L_f(或R_f) 为同构，也等价于 $V^{\perp}={}^{\perp}V=\{0\}$.
2. $\dim V^{\perp}=\dim{}^{\perp}V=\dim V-\operatorname{rk}f$.

5.3.2 二次型

定义 5.20 (二次型)

设 $f\in\mathcal{L}(V^2,\mathbb{F})$，则函数 $Q\in\mathcal{L}(V,\mathbb{F})$，对任意 $v\in V$ 都有 $Q(v)=f(v,v)$ 称为 V 上的一个二次型. 特别地，若 $\mathbb{F}=\mathbb{R}$，则称 $Q(V)$ 为实二次型.

取 V 基 $v_1,\cdots,v_n$，记 f 在该基下的度量矩阵为 A，则

$$Q:\mathbb{F}^n\longrightarrow\mathbb{F} \tag{5.80}$$

$$Q(x_1,\cdots,x_n)=Q(x_1\xi_1+\cdots+x_n\xi_n) \tag{5.81}$$

$$=(x_1,\cdots,x_n)A\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix} \tag{5.82}$$

$$=X^TAX=\sum_{i,j=1}^{n}f_{ij}x_ix_j \tag{5.83}$$

这里不区分 $Q:\mathbb{F}^n\to\mathbb{F}$ 与 $Q:V\to\mathbb{F}$，也不会引起歧义. 借助度量矩阵，二次型实际上等价于一个 $\mathbb{F}$ 上的二次多项式函数. 事实上，从上面的结果可以看出矩阵 A 并没有非常严格的要求，甚至可以令

$$A=(a_{ij})_{n\times n}=((f_{ij}+f_{ji})/2)_{n\times n} \tag{5.84}$$

使之成为一个对称阵，根据矩阵与双线性函数间的对应关系，存在 $f \in \mathcal{L}(V^2, \mathbb{F})$ 能做到这件事，即 $\tilde{f} = (f + f^{\mathrm{op}})/2 \in S(V^2, \mathbb{F})$，这里 $f^{\mathrm{op}}(u, v) = f(v, u)$.

命题 5.20

设 $\operatorname{char}\mathbb{F} \neq 2$，则 $S(V^2, \mathbb{F}) \cong \{V\text{上的二次型}\}$，即存在双射

$$\Phi : S(V^2, \mathbb{F}) \longrightarrow \{V\text{上的二次型}\} \tag{5.85}$$

$$f \longmapsto Q_f \tag{5.86}$$

其中 $Q_f(v) = f(v, v)$.

研究二次型的一个动机就是求一些二次多项式的值域、极值，如

$$Q(x, y, z) = x^2 + y^2 + z^2 s + 4xy - 6xz - 4yz \tag{5.87}$$

$$Q(x, y, z) = 2x^2 + 3y^2 + 5z^2 s + 4xy - 6xz - 7yz \tag{5.88}$$

如果 $Q(x_1, \cdots, x_n) = \lambda x_1^2 + \cdots + \lambda x_n^2$，那么问题会变得非常简单，因此称这种形式为标准型（或对角型），这种形式反映到度量矩阵上，可知 A 恰好为对角阵，这时出现了问题：是否存在一组基，使得其上的二次型为标准二次型？这也等价于：对任意对称阵 A，是否能相合到对角阵？事实上，二次型可以通过配方、换元化为标准型，因此答案是肯定的.

定理 5.17 (相合标准型)

设 $f \in S(V^2, \mathbb{F})$，则存在一组基，使得 f 在该基下的度量矩阵为

$$F = \operatorname{diag}(I_r, -I_s, 0_{n-r-s}) \tag{5.89}$$

定理的证明有两种方法：1. 归纳 + 打洞；2. 在谱定理的基础上“微调”.

5.3.3 正定二次型与正变换

二次型的一个经典问题是判定其是否正定.

定义 5.21 (正定/负定)

设 Q 为 V 中的二次型，则称 Q

1. 正定：若 $Q(v) \geqslant 0$，当且仅当 $v = 0$ 时取等. 记为 $Q > 0$.
2. 半正定：若 $Q(v) \geqslant 0$. 记为 $Q \geqslant 0$.
3. 负定：若 $Q(v) \leqslant 0$，当且仅当 $v = 0$ 时取等. 记为 $Q < 0$.
4. 半正定：若 $Q(v) \leqslant 0$. 记为 $Q \leqslant 0$.

由定义可知，$Q > 0$ 等价于 $-Q < 0$，$Q \geqslant 0$ 等价于 $-Q \leqslant 0$.

定义 5.22 (正变换)

设 $T \in \mathcal{L}(V)$，则称 T 为正变换，若 $\langle Tv, v\rangle \geqslant 0$，当且仅当 $v = 0$ 时取等.

讨论二次型的正定，也可以等价于讨论一个对称变换是否为正变换，而仿照正定等概念，也可以定义出矩阵的正定、负定等概念，因此二次型的正定问题实际上是对称矩阵的正定问题. 根据实谱定理，$\mathbb{R}$ 中的对称矩阵有非常良好的性质，即可正交对角化，许多讨论将会变得方便.

命题 5.21

设 $S \in \mathbb{R}^{n\times n}$ 对称，则 $S > 0$ 当且仅当 S 的特征值均为正数.

证明 设 $\lambda \in \operatorname{spec} S$，假设有 $\lambda \leqslant 0$，则存在 $0 \neq x \in \mathbb{F}^{n\times 1}$ 使得 $Q(x) = x^T S x = \lambda |x|^2 \leqslant 0$，矛盾.

推论 5.3

设 $S \in \mathbb{R}^{n\times n}$ 对称，则 $S>0$ 当且仅当 S 可逆且 $S^{-1}>0$.

命题 5.22

设 $S \in \mathbb{R}^{n\times n}$ 对称，则 $S>0$ 当且仅当存在可逆阵 $P \in \mathbb{F}^{n\times n}$，使得 $S=P^TP$.

证明 设 $S=P^TP$，则对任意 $x\in\mathbb{F}^{n\times 1}$ 都有 $Q(x)=x^TP^TPx=(Px)^T(Px)\geqslant 0$，并且 $Q(x)=0$ 当且仅当 $Px=0$，当且仅当 $x=0$. 反之若 S 正定，则其特征值均为正，根据谱定理，存在正交矩阵 $P_1\in\mathbb{F}^{n\times n}$ 使得

$$S=P_1^T\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}P_1 \tag{5.90}$$

$$=P_1^T\begin{pmatrix}\sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n}\end{pmatrix}\begin{pmatrix}\sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n}\end{pmatrix}P_1 \tag{5.91}$$

$$=P^TP \tag{5.92}$$

其中 P 可逆.

推论 5.4

设 $S\in\mathbb{R}^{n\times n}$，$S>0$，则 $\det S>0$.

从上面的推论可以看出，正定可以反映行列式的正性，反过来，通过行列式的推广：主子式也可以反映正定性.

命题 5.23

设 $S \in \mathbb{R}^{n\times n}$ 对称，则 $S>0$ 当且仅当其所有顺序主子式 >0.

证明 若 S 正定，则 $S=P^TP$，P 可逆，根据 C-B 公式可知对 $1\leqslant i_1<\cdots<i_k\leqslant n$ 有

$$S\begin{pmatrix}i_1,\cdots,i_k\\ i_1,\cdots,i_k\end{pmatrix}=\sum_{1\leqslant j_1<\cdots<j_n\leqslant n}P\begin{pmatrix}j_1,\cdots,j_k\\ i_1,\cdots,i_k\end{pmatrix}^2\geqslant 0 \tag{5.93}$$

由于 P 可逆，因此上面的求和式中必有一项为正，否则将 $\det P$ 按照 $i_1,\cdots,i_k$ 列展开可知 $\det P=0$，矛盾.

反之，若 S 的所有顺序主子式为正，对阶归纳：1 阶显然成立，假设 $n-1$ 阶成立，考虑 n 阶情况，设 $S=\begin{pmatrix}S_1 & \alpha\\ \alpha^T & s_{nn}\end{pmatrix}$，由条件可知 $S_1>0, s_{nn}>0$，故可逆，打洞得到

$$D=\begin{pmatrix}I_{n-1} & \\ -\alpha^TS_1^{-1} & 1\end{pmatrix}\begin{pmatrix}S_1 & \alpha\\ \alpha^T & s_{nn}\end{pmatrix}\begin{pmatrix}I_{n-1} & -S_1^{-1}\alpha\\ & 1\end{pmatrix}=\begin{pmatrix}S_1 & \\ & s_{nn}-\alpha^TS_1\alpha\end{pmatrix}>0 \tag{5.94}$$

根据归纳假设，设 $S_1=P_1^TP_1$，并且注意到 $S^{-1}=(S^{-1})^T$，则

$$S=\begin{pmatrix}I_{n-1} & -S_1^{-1}\alpha\\ & 1\end{pmatrix}\begin{pmatrix}P_1^TP_1 & \\ & s_{nn}-\alpha^TS_1\alpha\end{pmatrix}\begin{pmatrix}I_{n-1} & \\ -\alpha^TS_1^{-1} & 1\end{pmatrix} \tag{5.95}$$

$$=\left[\begin{pmatrix}P_1 & \\ & \sqrt{s_{nn}-\alpha^TS_1\alpha}\end{pmatrix}\begin{pmatrix}I_{n-1} & \\ -\alpha^TS_1^{-1} & 1\end{pmatrix}\right]^T\begin{pmatrix}P_1 & \\ & \sqrt{s_{nn}-\alpha^TS_1\alpha}\end{pmatrix}\begin{pmatrix}I_{n-1} & \\ -\alpha^TS_1^{-1} & 1\end{pmatrix} \tag{5.96}$$

$$=P^TP \tag{5.97}$$

这说明 S 正定.

上面的条件可以做一些调整，即

命题 5.24

设 $S \in \mathbb{R}^{n\times n}$ 对称，则 $S > 0$ 当且仅当其可逆，并且所有主子式非负.

证明 若 S 正定，则上面的命题中的过程已经证明了所有主子式非负，并且 S 可逆. 反之，设 S 可逆且所有主子式非负，同样对阶归纳，1 阶显然成立，设 $n-1$ 阶成立，考虑 n 阶情形.

首先有 $s_{11} > 0$，否则对任意 j 有 $\begin{vmatrix} s_{11} & s_{1j} \\ s_{j1} & s_{jj} \end{vmatrix} = -s_{1j}^2 = 0$，说明 $s_{1j} = 0$，这导致 $\det S = 0$，与可逆矛盾. 打洞可得

$$S' = \begin{pmatrix} 1 & \\ -S_{21}/s_{11} & I_{n-1} \end{pmatrix} \begin{pmatrix} s_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} 1 & \\ -S_{12}/s_{11} & I_{n-1} \end{pmatrix} = \begin{pmatrix} s_{11} & \\ & S_{22} - s_{11}^{-1} S_{21} S_{12} \end{pmatrix} = \begin{pmatrix} s_{11} & \\ & S'_{22} \end{pmatrix} \tag{5.98}$$

而显然 $S'\binom{1,i_2,\cdots,i_k}{1,i_2,\cdots,i_k} = S\binom{1,i_2,\cdots,i_k}{1,i_2,\cdots,i_k}$，因此 $\det S'_{22} > 0$，根据归纳假设可知 $S'_{22} > 0$，因此 $S' > 0, S > 0$，得证.

上面的结论可以整合为如下定理

定理 5.18

设 $S \in \mathbb{F}^{n\times n}$ 对称，则 $S > 0$ 与下述命题等价

1. S 的特征值均为正数.
2. 存在可逆矩阵 $P \in \mathbb{F}^{n\times n}$ 使得 $S = P^T P$.
3. S 可逆且所有主子式非负.
4. S 的所有顺序主子式为正.

类似的，对于半正定矩阵也有一些等价刻画.

定理 5.19

设 $S \in \mathbb{F}^{n\times n}$ 对称，则 $S \geqslant 0$ 与下述命题等价

1. S 的特征值均非负.
2. 存在行满秩矩阵 P，使得 $S = P^T P$.
3. 存在方阵 P，使得 $S = P^T P$.
4. S 的所有主子式非负.

证明

（半）正定矩阵还有一个重要的结论：平方根[1]，一般的矩阵是不能随意开根的，因为随便就能找到两个不相等，但平方相等的矩阵.

引理 5.3

设方阵 $A, B \geqslant 0$，若 $A^2 = B^2$，则 $A = B$.

证明 若 A, B 均为对角矩阵，则 $a_{ii}^2 = b_{ii}^2$，$a_{ii}, b_{ii} \geqslant 0$，因此 $a_{ii} = b_{ii}$，得证.

若不然，则根据实谱定理，不妨设 $A^2 = B^2 = D = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ 为对角阵，且 B 为对角阵，A 不为对角阵，则根据实谱定理（以及特征值关系），A, B 正交相似，设 $A = PBP^T$，可知 $A^2 = PB^2P^T = B^2$，即 $PD = DP$，计算可知，对于 $1 \leqslant i, j \leqslant n$，$p_{ij} = 0$ 与 $\lambda_i = \lambda_j$ 必有一者成立，这也表明 $p_{ij} = 0$ 与 $\lambda_i^{1/2} = \lambda_j^{1/2}$，必有一者成立，由此反过来得到 $PB = BP$，但根据假设有 $PB = AP$，P 可逆，故 $A = B$，因此 A 为对角阵，得证.

[1] 对正变换也有类似的结论

定义 5.23 (半正定的平方根)

设 $A \in \mathbb{F}^{n\times n}$，若 $A \geqslant 0$，则定义 A 的（唯一）平方根 $A^{1/2}$ 为唯一的半正定矩阵 B 使得 $B^2 = A$.

注 同理可以定义正变换的平方根.

推论 5.5

设 $A \in \mathbb{F}^{n\times n}, A \geqslant 0$，则

1. $A > 0$ 当且仅当 $A^{1/2} > 0$.
2. 存在多项式 $f \in \mathbb{R}[x]$，使得 $A^{1/2} = p(A)$.
3. $\operatorname{Ker} A = \operatorname{Ker} A^{1/2}, \operatorname{Im} A = \operatorname{Im} A^{1/2}$.

证明

1. 特征值.
2. Lagrange 插值.
3. 显然 $\operatorname{Ker} A \supset \operatorname{Ker} A^{1/2}$，设 $B = A^{1/2}$，则若 $Ax = B^2x = 0$，则 $x^TB^2x = (Bx)^T(Bx) = 0$，这说明 $Bx = 0$，$x \in \operatorname{Ker} A^{1/2}$，即 $\operatorname{Ker} A \subset \operatorname{Ker} A^{1/2}$. 另一条类似（或借助包含 + 维数）.

例 5.2 设 $A > 0, B \geqslant 0$，则 $\det(A + B) \geqslant \det A$，当且仅当 $B = 0$ 时取等.

定理 5.20 (Cholesky 分解)

设 $A \in \mathbb{F}^{n\times n}, A \geqslant 0$，则存在对角元非负的下三角阵 L，使得 $A = LL^T$，并且若 $A > 0$，则 L 唯一.

证明 对 A 的平方根作分解 $A^{1/2} = QR$，则 $A^2 = (A^{1/2})^T(A^{1/2}) = R^TR$，得证.

5.3.4 相合关系

本节进一步讨论相合关系，比如其不变量. 前面以及介绍过相合标准型，但并未考虑其唯一性，事实上这只需要证明 p, q 的唯一性.

定理 5.21

设 $S \in \mathbb{F}^{n\times n}$ 对称，则 S 相合于 $F = \operatorname{diag}(I_r, -I_s, 0_{n-r-s})$，这里 r, s 由 S 唯一确定.

证明 令 Q 为 S 对应的二次型，则其相合型等价价于 $V = \mathbb{R}^{n\times 1} = V_+ \oplus V_- \oplus V_0$，使得 $Q\Big|_{V_+} > 0, Q\Big|_{V_-} < 0, Q\Big|_{V_0} = 0$. 这里 $\dim V_+ = r, \dim V_- = s$. 假设还有另一分解 $V = \mathbb{R}^{n\times 1} = W_+ \oplus W_- \oplus W_0$，不妨设 $\dim W_+ > \dim V_+$，则

$$\dim W_+ + \dim V_- + \dim V_0 > \dim V_+ + \dim V_- + \dim V_0 = n \tag{5.99}$$

因此 $W_+ \cap (V_- \oplus V_0) \neq \{0\}$，任取其中向量 v，则 $v \in W_+$ 说明 $Q(v) > 0$，$v \in V_+ \oplus V_0$ 说明 $Q(v) \leqslant 0$，矛盾.

注 证明中的 $\oplus$ 视为集合的无交分解，而非空间的直和分解.

定义 5.24 (惯性指数)

设 $S \in \mathbb{F}^{n\times n}$ 对称，并且其相合标准型为 $F = \operatorname{diag}(I_r, -I_s, 0_{n-r-s})$，则定义 S 的正、负惯性指数为 r, s. 这是一个相合不变量.

命题 5.25

设 Q 为 V 上的二次型，则 $Q \geqslant 0$ 或 $Q \leqslant 0$，当且仅当 $\{v \in V : Q(v) = 0\}$ 为线性空间.

证明 若 $Q \geqslant 0$ 或 $Q \leqslant 0$，则根据相合标准型（分块）可知命题成立. 反之，略.

5.3.5 奇异值分解与极分解

定义 5.25 (奇异值)

设 $T \in \mathcal{L}(V, W)$，则称 $\sqrt{T^*T}$ 的非零特征值为 T 的奇异值.

注 由定义可知，T 与 T^* 的奇异值相同（对应的重数也相同）.

定理 5.22 (奇异值分解)

设 $T \in \mathcal{L}(V, W)$，所有奇异值为 $\mu_1, \cdots, \mu_r$，则存在 V, W 的标准正交基 $v_1, \cdots, v_n; w_1, \cdots, w_m$，使得 T 在这组基下的矩阵为

$$\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{F}^{m \times n}, \tag{5.100}$$

其中 $\Sigma_r = \mathrm{diag}(\mu_1, \cdots, \mu_r)$.

注 矩阵形式：任意矩阵 $A \in \mathbb{F}^{m \times n}$ 都可以写作 $P_1 \Sigma P_2^*$，这里 $P_1 \in O_m, P_2 \in O_n$，Σ 如上所述.

证明 由于 T^*T 正规，因此存在 V 标准正交基 $\alpha_1, \cdots, \alpha_n$，使得 T^*T 在这组基下的矩阵为 $\mathrm{diag}(\mu_1^2, \cdots, \mu_n^2)$ (这里约定对 $k > r$ 有 $\mu_k = 0$)，这说明

$$\langle T\alpha_i, T\alpha_j \rangle = \langle T^*T\alpha_i, \alpha_j \rangle = \begin{cases} \delta_{ij}\mu_i^2, & i \leqslant r \\ 0, & i > r \end{cases} \tag{5.101}$$

因此 $T\alpha_1, \cdots, T\alpha_r \in W$ 为正交向量组，$T\alpha_{r+1} = \cdots = T\alpha_n = 0$. 将 $T\alpha_1, \cdots, T\alpha_r$ 标准化为 $\beta_k = \mu_k^{-1}T\alpha_k, 1 \leqslant k \leqslant r$，再扩充为 W 的一组基 $\beta_1, \cdots, \beta_m$，则可知

$$T\alpha_k = \begin{cases} \mu_k\beta_k, & k \leqslant r \\ 0, & k > r \end{cases} \tag{5.102}$$

因此

$$T(\alpha_1, \cdots, \alpha_n) = (\beta_1, \cdots, \beta_m) \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \tag{5.103}$$

得证.

定理 5.23 (极分解)

设 $T \in \mathcal{L}(V)$，则存在正变换 P，正交变换 U 使得 $T = PU$，并且这种分解是唯一的.

注

1. 矩阵形式：任意矩阵 $A \in \mathbb{F}^{n \times n}$ 可唯一分解为正定矩阵与正交矩阵之积.
2. 可以得知 $P = \sqrt{TT^*}$.
3. 极分解的另一种形式为 $T = UP$，此时 $P = \sqrt{T^*T}$.

证明 根据奇异值分解，存在 V 的两组标准正交基 $\alpha_1, \cdots, \alpha_n; \beta_1, \cdots, \beta_n$，使得

$$Tv = \sum_{i=1}^{n} \langle v, \alpha_i \rangle T\alpha_i \tag{5.104}$$

$$= \sum_{i=1}^{n} \langle v, \alpha_i \rangle \mu_i \beta_i \tag{5.105}$$

$$\tag{5.106}$$

定义变换 $P, U \in \mathcal{L}(V)$，使得 $P(\alpha_i) = \mu_i\alpha_i, U(\alpha_i) = \beta_i$，则显然 $T = PU$，并且 P 为正变换，U 为正交变换（将标准正交基变为标准正交基）.

5.4 特征值的分布

下面讨论一些特征值分布的问题. 首先考虑下面的定理.

定理 5.24 (Hirsh)

设 $A \in \mathbb{F}^{n\times n}$，记 $B = \frac{1}{2}(A + A^*), C = \frac{1}{2i}(A - A^*)$，以及

$$M_A = \max\{|a_{kl}| : 1 \leqslant k, l \leqslant n\}, \tag{5.107}$$

设 $\lambda \in \operatorname{spec} A$，则

$$|\lambda| \leqslant nM_A, \quad |\operatorname{Re}\lambda| \leqslant nM_B, \quad |\operatorname{Im}\lambda| \leqslant nM_C. \tag{5.108}$$

并且若 $\mathbb{F} = \mathbb{R}$，则

$$|\operatorname{Im}\lambda| \leqslant M_C \left(\frac{n(n-1)}{2}\right)^{1/2} \tag{5.109}$$

♡

证明 设 $x \in E(\lambda, A)$，则 $Ax = \lambda x$，因此

$$x^*Ax = \sum_{k,l=1}^{n} a_{kl}\bar{x}_k x_l \tag{5.110}$$

$$= \lambda x^* x = \lambda\left(\sum_{k=1}^{n} |x_k|^2\right) \tag{5.111}$$

即得

$$|\lambda|\left(\sum_{k=1}^{n} |x_k|^2\right) \leqslant \sum_{k,l=1}^{n} |a_{kl}|\,|x_k|\,|x_l| \tag{5.112}$$

$$\leqslant M_A\left(\sum_{k=1}^{n} |x_k|\right)^2 \tag{5.113}$$

$$\leqslant nM_A\left(\sum_{k=1}^{n} |x_k|^2\right) \tag{5.114}$$

最后一步用到了 Cauchy 不等式，因此 $|\lambda| \leqslant nM_A$.

分别取实部与虚部可得

$$(\operatorname{Re}\lambda)x^*x = x^*Bx, \quad (\operatorname{Im}\lambda)x^*x = x^*Cx \tag{5.115}$$

因此分别对 $\operatorname{Re}\lambda, B; \operatorname{Im}\lambda, C$ 使用已证结论，可得

$$|\operatorname{Re}\lambda| \leqslant nM_B, \quad |\operatorname{Im}\lambda| \leqslant nM_C \tag{5.116}$$

当 $\mathbb{F} = \mathbb{R}$ 时，$C^T = -C$，因此 $c_{kk} = 0$，既有

$$|\operatorname{Im}\lambda|\left(\sum_{k=1}^{n} |x_k|^2\right) = \left|\sum_{1\leqslant k\neq l\leqslant n} c_{kl}\bar{x}_k x_l\right| \tag{5.117}$$

$$= \left|\sum_{1\leqslant k\neq l\leqslant n} c_{kl}(\bar{x}_k x_l - x_k\bar{x}_l)\right| \tag{5.118}$$

$$\leqslant \sum_{1\leqslant k<l\leqslant n} |c_{kl}|\,|\bar{x}_k x_l - x_k\bar{x}_l| \tag{5.119}$$

$$\leqslant M_C \sum_{1\leqslant k<l\leqslant n} \left|\frac{\bar{x}_k x_l - x_k\bar{x}_l}{i}\right| \tag{5.120}$$

由于 $\frac{\bar{x}_k x_l - x_k \bar{x}_l}{i}$ 为实数，因此根据 Cauchy 不等式有

$$\sum_{1\leqslant k<l\leqslant n}\left|\frac{\bar{x}_k x_l - x_k \bar{x}_l}{i}\right| \leqslant \left(\frac{n(n-1)}{2}\right)^{1/2}\left(\sum_{1\leqslant k<l\leqslant n}\left|\frac{\bar{x}_k x_l - x_k \bar{x}_l}{i}\right|^2\right) \tag{5.121}$$

并且由于

$$-\left(\sum_{1\leqslant k<l\leqslant n}(\bar{x}_k x_l - x_k \bar{x}_l)^2\right) = \left(\sum_{k=1}^n x_k\bar{x}_k\right)^2 - \left(\sum_{k=1}^n x_k^2\right)\left(\sum_{k=1}^n \bar{x}_k^2\right) \leqslant \left(\sum_{k=1}^n x_k\bar{x}_k\right)^2 \tag{5.122}$$

因此仿照前面的过程可证

$$|\mathrm{Im}\,\lambda| \leqslant M_C\left(\frac{n(n-1)}{2}\right)^{1/2} \tag{5.123}$$

定义 5.26 (行/列占优矩阵)

设 $A\in\mathbb{C}^{n\times n}$，记第 i 行元素模之和为 $R_i=\sum_{j=1}^n|a_{ij}|$，第 j 列元素模之和为 $T_j=\sum_{i=1}^n|a_{ij}|$，再令 $P_i=R_i-|a_{ii}|$，$Q_j=T_j-|a_{jj}|$，则

1. 若 A 满足 $|a_{ii}|>P_i$，则称 A 行对角占优.
2. 若 A 满足 $|a_{jj}|>T_j$，则称 A 列对角占优.

♣

定理 5.25 (Levy-Desplanques)

设 $A\in\mathbb{C}^{n\times n}$ 行对角占优或列对角占优，则 $\det A\neq 0$. ♡

证明 若不然，则存在 $x\neq 0$ 使得 $Ax=0$，设 $x=(x_1,\cdots,x_n)^T$，以及 $\max\{|x_1|,\cdots,|x_n|\}=|x_k|>0$，因此 $\sum_{j=1}^n a_{kj}x_j=0$，这说明

$$a_{kk}x_k = -\sum_{\substack{j=1\\j\neq k}}^n a_{kj}x_j \tag{5.124}$$

取模可得

$$|a_{kk}| \leqslant \sum_{\substack{j=1\\j\neq k}}^n |a_{kj}|\frac{|x_j|}{|x_k|} \tag{5.125}$$

$$\leqslant \sum_{\substack{j=1\\j\neq k}}^n |a_{kj}| \tag{5.126}$$

$$= R_k - a_{kk} = P_k \tag{5.127}$$

这与行占优矛盾（对列的情况考虑 A^T，同理）.

推论 5.6 (Gersgorin 圆盘定理)

设 $A\in\mathbb{C}^{n\times n}$，圆盘 $G_k(A)=\{z\in\mathbb{C}:|z-a_{kk}|\leqslant P_k\}$，则

$$\mathrm{spec}\,A\subset\bigcap_{k=1}^n G_k(A), \tag{5.128}$$

即特征值位于 n 个圆盘的并集之中. ♡

证明 设 $\lambda_0\in\mathrm{spec}\,A$，则 $\varphi_A(\lambda_0)=\det(\lambda_0 I-A)=0$，根据 L-D 定理可证 $\lambda_0 I-A$ 不是行占优的，因此必定存在 $1\leqslant i\leqslant n$ 使得 $|\lambda_0-a_{ii}|\leqslant P_i$，得证.

事实上，根据 A^T, A 的特征值相同，上面的定理可以作一些加强，即将 P_k 换为 $\min\{P_k,T_k\}$. 此外，定理还有一种推广，首先考虑 L-D 定理的推广.

定理 5.26

设 $A \in \mathbb{C}^{n\times n}$，并且 $|a_{ii}||a_{jj}| > P_iP_j$，则 $\det A \neq 0$.

证明 若不然，则设 $0 \neq x = (x_1, \cdots, x_n)^T \in \operatorname{Ker} A$，并且 x_r, x_s 满足对任意 $i = 1, 2, \cdots, r-1, r+1, \cdots, n$ 都有

$$|x_r| \geqslant |x_s| \geqslant |x_i| \tag{5.129}$$

若 $x_s = 0$，则 $Ax = 0$ 可推出 $a_{rr}x_r = 0$，这说明 $a_{rr} = 0$，与假设矛盾. 故 $x_s \neq 0$. 仿照前面的过程有

$$|a_{rr}||x_r| = \left|\sum_{\substack{j=1\\ j\neq r}}^{n} a_{rj}x_j\right| \leqslant \sum_{\substack{j=1\\ j\neq r}}^{n} |a_{rj}||x_j| \leqslant |x_s|P_r \tag{5.130}$$

$$|a_{ss}||x_s| = \left|\sum_{\substack{j=1\\ j\neq s}}^{n} a_{sj}x_j\right| \leqslant \sum_{\substack{j=1\\ j\neq r}}^{n} |a_{sj}||x_j| \leqslant |x_r|P_s \tag{5.131}$$

因此 $|a_{rr}||a_{ss}| \leqslant P_rP_s$，矛盾，说明 $\det A \neq 0$.

推论 5.7

设 $A \in \mathbb{C}^{n\times n}$，则 A 的特征值一定位于复平面上 $\frac{n(n-1)}{2}$ 个 Cassini 卵形区域

$$|z - a_{ii}||z - a_{jj}| \leqslant P_iP_j \tag{5.132}$$

的并集中.

最后考虑一个定理，它说明特征值会确定特征向量的某些属性.

定理 5.27

设 $A \in \mathbb{C}^{n\times n}$ 自伴，U 为正交矩阵且

$$U^*AU = \operatorname{diag}(\lambda_1, \cdots, \lambda_n) \tag{5.133}$$

令 M_j 为 A 中删去第 j 行、第 j 列所得的 $n-1$ 阶子方阵，设 M_j 的特征值为 $\lambda_1^{(j)}, \cdots, \lambda_{n-1}^{(j)}$，则

$$|u_{ji}|^2 \prod_{\substack{k=1\\ k\neq i}}^{n} (\lambda_i - \lambda_k) = \prod_{\substack{k=1\\ k\neq i}}^{n-1} (\lambda_i - \lambda_k^{(j)}) \tag{5.134}$$

证明 令 $\lambda_j^\dagger = \lambda_1 \cdots \widehat{\lambda_j} \cdots \lambda_n$，则

$$U^*A^\dagger U = \operatorname{diag}(\lambda_1^*, \cdots, \lambda_n^*) \tag{5.135}$$

其中 $A^\dagger$ 表示 A 的行列式伴随阵，并且 $U_j = (u_{1j}, \cdots, u_{nj})^T \in E(\lambda_j, A)$，$U_j \in E(\lambda_j^\dagger, A^\dagger)$，同理可得

$$U^*(\lambda_iI - A)^\dagger U = \operatorname{diag}(\prod_{k=2}^{n} (\lambda_i - \lambda_k), \cdots, \prod_{\substack{k=1\\ k\neq i}}^{n} (\lambda_i - \lambda_k), \cdots, \prod_{k=1}^{n-1} (\lambda_i - \lambda_k)) \tag{5.136}$$

因此

$$(\lambda_iI - A)^\dagger = U\operatorname{diag}(0, \cdots, \prod_{\substack{k=1\\ k\neq i}}^{n} (\lambda_i - \lambda_k), \cdots, 0)U^* \tag{5.137}$$

$$= \prod_{\substack{k=1\\ k\neq i}}^{n} (\lambda_i - \lambda_k)U_iU_i^* \tag{5.138}$$

比较两边 j, j 分量，得证.

附录 A 补充内容

A.1 Cayley-Hamilton 定理的几种证明

定理 A.1 (Cayley-Hamilton 定理)

设 n 阶方阵 A，其特征多项式为 $\varphi_A(\lambda)=\det(\lambda I-A)$，则 $\varphi_A(A)=0$.

♡

证明 【证法一】

考虑 A 对应的线性映射 $\mathscr{A}$，设有一组基 $(\alpha_1,\cdots,\alpha_n)$，使得 $\mathscr{A}$ 在该基下的矩阵为上三角阵. 设 $V_k=\langle\alpha_1,\alpha_2,\cdots,\alpha_{n-k}\rangle$，则有不变子空间链

$$V=V_0\supset V_1\supset\cdots\supset V_{n-1}\supset V_n=\mathbf{0}. \tag{A.1}$$

由于 $(\mathscr{A}-\lambda_{n-k}\mathscr{I})\alpha_{n-k}\in V_{k+1}$，故

$$(\mathscr{A}-\lambda_{n-k}\mathscr{I})V_k\subset V_{k+1} \tag{A.2}$$

即有

$$\begin{aligned}\varphi_A(\mathscr{A})V&=(\mathscr{A}-\lambda_1\mathscr{I})\cdots(\mathscr{A}-\lambda_n\mathscr{I})V_0 &\text{(A.3)}\\&\subseteq(\mathscr{A}-\lambda_1\mathscr{I})\cdots(\mathscr{A}-\lambda_{n-1}\mathscr{I})V_1 &\text{(A.4)}\\&\subseteq\cdots\subseteq(\mathscr{A}-\lambda_1\mathscr{I})V_{n-1}\subseteq V_n=\mathbf{0}. &\text{(A.5)}\end{aligned}$$

命题得证.

【证法二】

利用广义特征向量分解（根子空间分解）.

设

$$\varphi_A(\lambda)=(\lambda-\lambda_1)^{n_1}\cdots(\lambda-\lambda_s)^{n_s}, \tag{A.6}$$

则

$$V=\bigoplus_{k=1}^{s}G(\lambda_k,A)=\bigoplus_{k=1}^{s}\operatorname{Ker}(A-\lambda_kI)^{n_k}. \tag{A.7}$$

则对任意 $v\in V$，都存在 $v_1\in G(\lambda_1,A),\cdots,v_s\in G(\lambda_s,A)$，使得

$$v=v_1+\cdots+v_s. \tag{A.8}$$

因此

$$\begin{aligned}\varphi_A(A)(v)&=\varphi_A(A)\left(\sum_{k=1}^{s}v_k\right)=\sum_{k=1}^{s}\varphi_A(A)(v_k) &\text{(A.9)}\\&=\sum_{k=1}^{s}\frac{\varphi_A(A)}{(A-\lambda_kI)^{n_k}}(A-\lambda_kI)^{n_k}(v_k) &\text{(A.10)}\\&=0 &\text{(A.11)}\end{aligned}$$

由于上式对所有 $v\in V$ 成立，故定理得证.

【证法三】

设有域 $\mathbb{F}$ 上的 n 维向量空间 V，则 $(\mathcal{L}(V),+,\circ)$ 为一个含幺交换环，取 $\mathscr{A}\in\mathcal{L}(V)$，$R=\mathscr{A}$ 为 $\mathscr{A}$ 生成的子环. 取 V 的一组基 $(\alpha_1,\cdots,\alpha_n)$，矩阵 $A\in\mathbb{F}^{n\times n}\subseteq\mathbb{F}[\mathscr{A}]^{n\times n}$ 满足

$$\mathscr{A}(\alpha_1,\cdots,\alpha_n)=(\alpha_1,\cdots,\alpha_n)A \tag{A.12}$$

定义运算

$$V^{1\times p}\times\mathbb{F}^{p\times q}\longrightarrow V^{1\times q} \tag{A.13}$$

$$(\beta_1,\cdots,\beta_p)\begin{pmatrix}\mathscr{A}_{11} & \cdots & \mathscr{A}_{1q}\\ \vdots & \ddots & \vdots\\ \mathscr{A}_{p1} & \cdots & \mathscr{A}_{pq}\end{pmatrix}=\left(\sum_{i=1}^{p}\mathscr{A}_{i1}(\beta_i)\quad\cdots\quad\sum_{i=1}^{p}\mathscr{A}_{iq}(\beta_q)\right) \tag{A.14}$$

$$\mathbb{F}[\mathscr{A}]^{p\times q}\times\mathbb{F}[\mathscr{A}]^{q\times r}\longrightarrow\mathbb{F}[\mathscr{A}]^{p\times r} \tag{A.15}$$

$$\begin{pmatrix}\mathscr{A}_{11} & \cdots & \mathscr{A}_{1q}\\ \vdots & \ddots & \vdots\\ \mathscr{A}_{p1} & \cdots & \mathscr{A}_{pq}\end{pmatrix}\begin{pmatrix}\mathscr{B}_{11} & \cdots & \mathscr{B}_{1r}\\ \vdots & \ddots & \vdots\\ \mathscr{B}_{q1} & \cdots & \mathscr{B}_{qr}\end{pmatrix}=\begin{pmatrix}\mathscr{C}_{11} & \cdots & \mathscr{C}_{1r}\\ \vdots & \ddots & \vdots\\ \mathscr{C}_{p1} & \cdots & \mathscr{C}_{pr}\end{pmatrix} \tag{A.16}$$

$$\mathscr{C}_{ij}=\sum_{k=1}^{q}\mathscr{A}_{ik}\mathscr{B}_{kj} \tag{A.17}$$

且易验证

$$[(\beta_1,\cdots,\beta_p)(\mathscr{A}_{ij})_{p\times q}](\mathscr{B}_{ij})_{q\times r}=(\beta_1,\cdots,\beta_p)[(\mathscr{A}_{ij})_{p\times q}](\mathscr{B}_{ij})_{q\times r}] \tag{A.18}$$

设

$$\begin{aligned}\mathbb{A}=\varphi_A(\mathscr{A})&=\mathscr{A}I_n-A\in\mathbb{F}[\mathscr{A}]^{n\times n} &\text{(A.19)}\\ &=\begin{pmatrix}\mathscr{A}-a_{11}1_R & \cdots & -a_{1n}\\ \vdots & \ddots & \vdots\\ -a_{n1} & \cdots & \mathscr{A}-a_{nn}1_R\end{pmatrix} &\text{(A.20)}\end{aligned}$$

$$\begin{aligned}\mathbb{A}\mathbb{A}^*&=\det\mathbb{A}I_n &\text{(A.21)}\\ &=\operatorname{diag}(\mathbb{A}1_R,\cdots,\mathbb{A}1_R) &\text{(A.22)}\end{aligned}$$

$$\begin{aligned}(\alpha_1,\cdots,\alpha_n)\mathbb{A}&=(0,\cdots,0) &\text{(A.23)}\\ \Rightarrow(\alpha_1,\cdots,\alpha_n)(\mathbb{A}\mathbb{A}^*)&=(0,\cdots,0) &\text{(A.24)}\\ \Rightarrow\varphi_A(\mathscr{A})&=\det\mathbb{A}=0 &\text{(A.25)}\end{aligned}$$

【证法四】

考虑 λ 矩阵 $B(\lambda)=\lambda I-A$，设

$$B^*(\lambda)=\lambda^nB_n+\cdots+\lambda B_1+B_0. \tag{A.26}$$

则根据定义有

$$\begin{aligned}B(\lambda)B^*(\lambda)&=\operatorname{diag}(\varphi_A(\lambda),\cdots,\varphi_A(\lambda)) &\text{(A.27)}\\ &=\lambda^nI+\lambda^{n-1}a_{n-1}I+\cdots+a_0I. &\text{(A.28)}\end{aligned}$$

如果将左边展开，可以得到

$$\begin{aligned}B(\lambda)B^*(\lambda)&=(\lambda I-A)(\lambda^nB_n+\cdots+\lambda B_1+B_0) &\text{(A.29)}\\ &=\lambda^{n+1}B_n+\lambda^n(B_{n-1}-AB_n)+\cdots+\lambda(B_0-AB_1)-AB_0. &\text{(A.30)}\end{aligned}$$

对比两边多项式的系数，可以得到

$$\begin{aligned}B_n&=0 &\text{(A.31)}\\ B_{n-1}-AB_n&=I &\text{(A.32)}\\ B_{n-2}-AB_{n-1}&=a_{n-1}I &\text{(A.33)}\\ &\cdots &\text{(A.34)}\end{aligned}$$

$$B_0 - AB_1 = a_1 I \tag{A.35}$$
$$-AB_0 = a_0 I \tag{A.36}$$

易得

$$B_0 = AB_1 + a_1 I \tag{A.37}$$
$$= A^2 B_2 + a_2 A + a_1 I \tag{A.38}$$
$$= \cdots \tag{A.39}$$
$$= A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2 A + a_1 I. \tag{A.40}$$

而根据上面的最后一条，有

$$AB_0 = -A(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2 A + a_1 I) \tag{A.41}$$
$$= -A^n - a_{n-1}A^{n-1} - \cdots - a_2 A^2 - a_1 A \tag{A.42}$$
$$= a_0 I. \tag{A.43}$$

移项，得到

$$\varphi_A(A) = A^n + a_{n-1}A^{n-1} + \cdots + a_2 A^2 + a_1 A + a_0 I = 0 \tag{A.44}$$

命题得证.

A.2 双线性函数

定义 A.1

称 $f \in \mathcal{L}(V^2, \mathbb{F})$ 为 V 上的一个双线性函数.
特别地，若 $f(u,v) = f(v,u)$，则称 f 是一个对称双线性函数，记为 $f \in S(V^2, \mathbb{F})$.

仿照单线性函数中的讨论，若 $\dim V = n$，则取 V 中的基

$$B = (\xi_1, \cdots, \xi_n), \tag{A.45}$$

则矩阵

$$(f_{ij})_{n\times n} := (f(\xi_i, \xi_j))_{n\times n} = \begin{pmatrix} f_{11} & \cdots & f_{1,n} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nn} \end{pmatrix}. \tag{A.46}$$

称为 f 在基 B 下的**度量矩阵**. 度量矩阵有助于我们从矩阵角度看待双线性函数对应的线性变换.
设 $u, v \in V$，且

$$u = \sum_{i=1}^n x_i \xi_i, \quad v = \sum_{i=1}^n y_i \xi_i \tag{A.47}$$

则

$$f(u,v) = f\left(\sum_{i=1}^n x_i \xi_i, \sum_{j=1}^n y_j \xi_j\right) \tag{A.48}$$
$$= \sum_{i,j=1}^n x_i f(\xi_i, \xi j) y_j \tag{A.49}$$
$$= \sum_{i,j=1}^n x_i f_{ij} y_j \tag{A.50}$$

$$= \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} f_{11} & \cdots & f_{1,n} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{A.51}$$

$$= X^T FY. \tag{A.52}$$

如果考虑基变换，则可以揭示继相抵、相似后的又一个重要等价关系——**相合**.

$$(\xi_1', \cdots, \xi_n') = (\xi_1', \cdots, \xi_n')P. \tag{A.53}$$

则有

$$(f_{ij}')_{n\times n} = (f(\xi_i', \xi_j'))_{n\times n} \tag{A.54}$$

$$= f\left(\sum_{u=1}^{n} a_{ui}\xi_u, \sum_{v=1}^{n} a_{vj}\xi_v\right) \tag{A.55}$$

$$= \sum_{u,v=1}^{n} a_{ui} f_{uv} a_{vj} \tag{A.56}$$

$$= P^T FP. \tag{A.57}$$

对上面的关系做如下定义

定义 A.2

设 $F', F \in \mathbb{F}^{n\times n}$，则称 F', F 相合，若存在可逆阵 P 使得

$$F' = P^T FP. \tag{A.58}$$

♣

定义 A.3

定义双线性函数 f 的秩（$\operatorname{rk}(f)$）为其在一组基下的度量矩阵的秩，若 $\operatorname{rk}(f) = n = \dim V$，则称 f 是非退化的.

♣

通过上面的分析，f 在不同基下的度量矩阵之间有相合关系，从而秩是一个相合不变量.

根据上面的讨论，易得如下命题

命题 A.1

设 f 为 V 上的双线性函数，则 f 是对称的当且仅当其在一组基下的度量矩阵 F 是对称的.

♠

¶ 有趣的东西

在前面给出度量矩阵的定义后，自然有如下事实

命题 A.2

设有限维向量空间 $V/\mathbb{F}$，$\dim V = n < \infty$，B 为 V 上一组基，则

$$\mathcal{L}(V^2, \mathbb{F}) \cong \mathbb{F}^{n\times n}. \tag{A.59}$$

即存在线性空间同构

$$\Phi_B : \mathcal{L}(V^2) \longrightarrow \mathbb{F}^{n\times n} \tag{A.60}$$

$$f \longmapsto A_f \tag{A.61}$$

♠

也就是说，给定 V 上的一组基，则对任意 V 上的双线性函数 f，可以唯一确定 $\mathbb{F}^{n\times n}$ 中的一个矩阵，反之亦然，这个矩阵就是 f 在这组基下的度量矩阵.

证明 只需证明 Φ_B 即单又满. 由于 $\operatorname{Ker} f = 0$，故单性是显见的，下面证明满性.

对任意 $A \in \mathbb{F}^{n\times n}$，存在映射 f_A，对任意 $u = \sum x_i\xi_i, v = \sum y_j\xi_j$，有

$$f_A(u,v) = X^T AY. \tag{A.62}$$

由此依次取基 B 中的每个向量，可得

$$f_A(\xi_i,\xi_j) = a_{ij}. \tag{A.63}$$

因此必然存在这样的 f_A，使得 $\Phi_B(f_A) = A$，满性即证.

上面的关系可以用下面的图理解

$$\begin{tikzcd} V^2 \arrow[r, "f"] \arrow[d] & \mathbb{F} \\ (\mathbb{F}^{n\times 1})^2 \arrow[r, "A_f"] & \mathbb{F} \end{tikzcd} \tag{A.64}$$

由于双线性函数的自变量有两个向量，因此对任意 $v \in V$，都可以诱导出两个线性泛函（视为固定其中一个位置的向量），即

$$L_f : V \longrightarrow V^* = \mathcal{L}(V,\mathbb{F}) \tag{A.65}$$

$$v \longmapsto f(v,-) \tag{A.66}$$

$$R_f : V \longrightarrow V^* = \mathcal{L}(V,\mathbb{F}) \tag{A.67}$$

$$v \longmapsto f(-,v) \tag{A.68}$$

考虑这两个线性算子，可以自然引出正交补

定义 A.4

设 $v \in V$，则定义 v 的右正交补为

$$v^\perp := \operatorname{Ker} L_f(v) = \{u \in V : f(v,u) = 0\}. \tag{A.69}$$

同理定义 v 的右左交补为

$$^\perp v := \operatorname{Ker} R_f(v) = \{u \in V : f(u,v) = 0\}. \tag{A.70}$$

♣

由此可以直接得到，左右正交补都是 V 的子空间. 仿照上述定义，可以定义集合的正交补

定义 A.5

设集合 $S \subseteq V$，则定义 S 的右正交补为

$$S^\perp := \operatorname{Ker} L_f(S) = \{u \in V : f(S,u) = 0\}. \tag{A.71}$$

同理定义 S 的右左交补为

$$^\perp S := \operatorname{Ker} R_f(S) = \{u \in V : f(u,S) = 0\}. \tag{A.72}$$

特别地，$\operatorname{Ker} L_f = V^\perp$，$\operatorname{Ker} R_f = {}^\perp V$.

♣

两个定义从几号上就能看出明显的对称关系. 关于正交补，有以下有趣的性质：

命题 A.3

设集合 $S \subseteq V$，则

1. $S^\perp \leqslant V, ^\perp S \leqslant V$.
2. $S \subseteq T$，则 $S^\perp \geqslant T^\perp, ^\perp S \geqslant ^\perp T$.
3. $(^\perp(S^\perp))^\perp = S^\perp$.

♠

下面回到对 L_f, R_f 本身的讨论. 它们是从 V 到 V 的对偶空间 V^* 的映射，这个映射本身依赖于 f 的选取，然由下面的命题，可以说明这种依赖是一一对应的.

命题 A.4

设 $V/\mathbb{F}$，$f \in \mathcal{L}(V^2, \mathbb{F})$，则

$$\mathcal{L}(V^2, \mathbb{F}) \cong \mathcal{L}(V, V^*), \tag{A.73}$$

或者说对任意 $\mathscr{A} \in \mathcal{L}(V, V^*)$，都有唯一 $f_{\mathscr{A}} \in \mathcal{L}(V^2, \mathbb{F})$ 与之对应.

$$\begin{tikzcd} \mathcal{L}(V^2,\mathbb{F}) \arrow[r, leftrightarrow, "\sim"] \arrow[d, "\text{取定基}"'] & \mathcal{L}(V^2,\mathbb{F}) \arrow[d, "\text{取定对偶基}"] \\ \mathbb{F}^{n\times n} & \mathbb{F}^{\ltimes\times\ltimes} \end{tikzcd} \tag{A.74}$$

注 事实上，有下面的事实对无限维线性空间成立：

命题 A.5

设 $B = (\xi_1, \cdots, \xi_n)$ 为 V 的一组基，$B^* = (\xi^1, \cdots, \xi^n)$ 为 B 的对偶基，V 上双线性函数 f 在基 B 下的度量阵为 A，则 $R_f(L_f)$ 在基 B 以及 B^* 下的矩阵为 $A(A^T)$，即

$$R_f(\xi_1, \cdots, \xi_n) = (\xi^1, \cdots, \xi^n)A \tag{A.75}$$

$$L_f(\xi_1, \cdots, \xi_n) = (\xi^1, \cdots, \xi^n)A^T \tag{A.76}$$

证明 $(R_f(\xi_j))(\xi_i) = f(\xi_i, \xi_j) = a_{ij}$，故

$$R_f(\xi_j) = \sum_{i=1}^{n} a_{ij}\xi^i. \tag{A.77}$$

这里用到了 V^* 中的线性泛函可以由一组对偶基张成的性质，以及通过一组对偶基确定它的方法，简单来说就是

$$\varphi = \sum_{i=1}^{n} \varphi(e_i)e^i. \tag{A.78}$$

这部分可以参考 Linear Algebra Done Right 中关于对偶的部分.

对于 L_f, R_f，根据线性映射基本定理（秩-零化度定理）可知，

$$\dim V = \dim \operatorname{Ker} L_f + \dim \operatorname{Im} L_f = \dim \operatorname{Ker} L_f + \dim \operatorname{Im} L_f \tag{A.79}$$

$$= \dim V^{\perp} + \operatorname{rk} f = \dim {}^{\perp}V + \operatorname{rk} f \tag{A.80}$$

故 $\dim V^{\perp} = \dim {}^{\perp}V = \dim V - \operatorname{rk} f$.

下面以最后一个定理结尾.

定理 A.2

设 V 为有限维向量空间，$f \in \mathcal{L}(V^2, \mathbb{F})$，则 f 非退化当且仅当对任意 $\varphi \in V^*$，都存在唯一 $v \in V$，使得 $\varphi = L_f(v) = f(v, -)$.

证明 f 非退化 $\iff L_f$ 是双射 $\forall \varphi \in V^*, \exists v \in V : L_f(v) = \varphi$，并且这样的 v 是唯一的.

上面定理的 $\Rightarrow$ 方向即为 Riesz 表示定理.

注

1. f 非退化等价于 $L_f(R_f)$ 为同构，也等价于 $V^{\perp} = {}^{\perp}V = \{0\}$.
2. $\dim V^{\perp} = \dim {}^{\perp}V = \dim V - \operatorname{rk} f$.