

## 5A 不变子空间

### 特征值

#### 5.1 算子

称从一个向量空间到其本身的线性映射为算子，记作  $T \in \mathcal{L}(V)$

$$V = V_1 \oplus \cdots \oplus V_m$$

其中各  $V_k$  都是  $V$  的非零子空间，那么要理解  $T$  的作用，我们只需要理解各  $T|_{V_k}$  的作用即可。此处  $T|_{V_k}$  表示将  $T$  限于更小的定义空间  $V_k$  上。因为  $V_k$  是比  $V$  小的向量空间，所以讨论  $T|_{V_k}$  要比讨论  $T$  简单些。

然而如果想应用算子研究中一些有用的工具（例如取算子的幂），那么就会遇到这个问题： $T|_{V_k}$  可能不会将  $V_k$  映射至它本身；换句话说， $T|_{V_k}$  可能不是  $V_k$  上的算子。由此，在  $V$  的形式如上的分解式中，我们仅考虑满足  $T$  将各  $V_k$  都映射到本身的情形。因此，现在我们给  $V$  的被  $T$  映射到本身的子空间取个名称：

#### 5.2 不变子空间

$T \in \mathcal{L}(V)$ 。对于  $V$  的子空间  $U$ ，若对每个  $u \in U$  均有  $Tu \in U$ ，则称  $U$  在  $T$  下是不变的。于是，如果  $T|_U$  是  $U$  上的算子，则  $U$  在  $T$  下是不变的。

#### 5.3 微分算子下不变的子空间

$T \in \mathcal{L}(\mathcal{P}(\mathbf{R}))$  定义为  $Tp = p'$ 。那么  $\mathcal{P}(\mathbf{R})$  的子空间  $\mathcal{P}_4(\mathbf{R})$  在  $T$  下是不变的，因为若  $p \in \mathcal{P}(\mathbf{R})$  的次数不高于 4，那么  $p'$  的次数亦不高于 4。

#### 5.4 四个不变子空间（并不一定全都不同）

$T \in \mathcal{L}(V)$ ，那么  $V$  的以下子空间在  $T$  下都是不变的：①  $\{0\}$  子空间  $\{0\}$  在  $T$  下是不变的，因为若  $u \in \{0\}$ ，那么  $u = 0$ ，因此  $Tu = 0 \in \{0\}$ ；②  $V$  在  $T$  下是不变的，因为若  $u \in V$ ，那么  $Tu \in V$ ；③  $\text{null } T$  子空间  $\text{null } T$  在  $T$  下是不变的，因为若  $u \in \text{null } T$ ，那么  $Tu = 0$ ，因此  $Tu \in \text{null } T$ ；④  $\text{range } T$  子空间  $\text{range } T$  在  $T$  下是不变的，因为若  $u \in \text{range } T$ ，那么  $Tu \in \text{range } T$ 。

一算子  $T \in \mathcal{L}(V)$  一定会有除了  $\{0\}$  和  $V$  以外的不变子空间吗？稍后我们会看到，如果  $V$  是有限维的，且  $\dim V > 1$  ( $\mathbf{F} = \mathbf{C}$ ) 或  $\dim V > 2$  ( $\mathbf{F} = \mathbf{R}$ )，那么这个问题的答案是肯定的。可参看 5.19 与 5B 节的习题 29。

上例中提到， $\text{null } T$  和  $\text{range } T$  在  $T$  下是不变的。然而，不能仅凭这两个不变子空间存在，就轻率地回答上述有关除了  $\{0\}$  和  $V$  以外的不变子空间存在的问题，因为有可能  $\text{null } T$  等于  $\{0\}$  而  $\text{range } T$  等于  $V$  ( $T$  可逆时就会出现这种情况)。

我们稍后再回过头来更深入地研究不变子空间。现在我们转而研究最简单的非平凡不变子空间——维数为 1 的不变子空间。

任取  $v \in V$  且  $v \neq 0$ ，并令  $U$  等于  $\sigma_v$  的所有标量倍构成的集合：

$$U = \{\lambda v : \lambda \in \mathbf{F}\} = \text{span}(v)$$

那么， $U$  是  $V$  的一维子空间（且只要适当选取  $v$ ， $V$  的每个一维子空间都可写成这种形式）。若  $U$  在算子  $T \in \mathcal{L}(V)$  下是不变的，那么  $Tv \in U$ ，因此存在标量  $\lambda \in \mathbf{F}$  使得

$$Tv = \lambda v$$

反之，若对某  $\lambda \in \mathbf{F}$  有  $Tv = \lambda v$ ，那么  $\text{span}(v)$  是  $V$  的在  $T$  下不变的一维子空间。

刚才看到，等式  $Tv = \lambda v$  与一维不变子空间紧密联系在一起。这个等式很重要，因此满足该式的标量  $\lambda$  和向量  $v$  都有特别的名称。

#### 5.5 特征值

$T \in \mathcal{L}(V)$ . 对于数  $\lambda \in \mathbf{F}$ , 若存在  $\nu \in V$  使得  $\nu \neq 0$  且  $T\nu = \lambda\nu$ , 则称  $\lambda$  为  $T$  的特征值

要求  $\nu \neq 0$  因每个标量  $\lambda \in \mathbf{F}$  都满足  $T0 = \lambda 0$

上面的讨论表明,  $V$  有在  $T$  下不变的一维子空间, 当且仅当  $T$  有特征值

### 5.6 特征值

定义算子  $T \in \mathcal{L}(\mathbf{F}^3)$  为: 对  $(x, y, z) \in \mathbf{F}^3$ ,

$$T(x, y, z) = (7x + 3z, 3x + 6y + 9z, -6y)$$

那么  $T(3, 1, -1) = (18, 6, -6) = 6(3, 1, -1)$ . 所以 6 是  $T$  的特征值

下面结论中的诸等价关系, 乃至线性代数中许多深刻的结论, 都只在有限维向量空间中成立

### 5.7 成为特征值的等价条件

$V$  是有限维的,  $T \in \mathcal{L}(V)$  且  $\lambda \in \mathbf{F}$  TFAE( $I \in \mathcal{L}(V)$  是恒等算子, 对所有  $\nu \in V$  有  $I\nu = \nu$ ):

- (a)  $\lambda$  不是  $T$  的特征值
- (b)  $T - \lambda I$  不是单射
- (c)  $T - \lambda I$  不是满射
- (d)  $T - \lambda I$  不可逆

### 三 证

(a) 与(b) 等价是因为等式  $T\nu = \lambda\nu$  等价于等式  $(T - \lambda I)\nu = 0$ . (b)、(c) 和 (d) 等价是由于 3.65

设  $T \in \mathcal{L}(V)$  且  $\lambda \in \mathbf{F}$  是  $T$  的特征值. 若向量  $\nu \in V$  满足  $\nu \neq 0$  且  $T\nu = \lambda\nu$ , 则称该向量是  $T$  对应于  $\lambda$  的特征向量

换句话说, 非零向量  $\nu \in V$  是算子  $T \in \mathcal{L}(V)$  的特征向量, 当且仅当  $T\nu$  是  $\nu$  的标量倍. 因为  $T\nu = \lambda\nu$  当且仅当  $(T - \lambda I)\nu = 0$ , 所以, 向量  $\nu \in V$  ( $\nu \neq 0$ ) 是  $T$  对应于  $\lambda$  的特征向量, 当且仅当  $\nu \in \text{null}(T - \lambda I)$

### 5.9 特征值和特征向量

$T \in \mathcal{L}(\mathbf{R}^2)$  定义为  $T(w, z) = (-z, w)$ .

(a)  $\mathbf{F} = \mathbf{R}$  的情况. 那么  $T$  的作用就是将向量绕  $\mathbf{R}^2$  的原点逆时针旋转  $90^\circ$ . 一个算子有特征值, 当且仅当其定义空间内存在可被其映射为自身标量倍的非零向量. 将  $\mathbf{R}^2$  中的非零向量逆时针旋转  $90^\circ$  不会得到原向量的标量倍. 所以: 若  $\mathbf{F} = \mathbf{R}$ , 那么  $T$  没有特征值 (于是也没有特征向量)

(b)  $\mathbf{F} = \mathbf{C}$  的情况. 为求出  $T$  的特征值, 我们必须求出标量  $\lambda$ , 使得  $T(w, z) = \lambda(w, z)$  有除  $w = z = 0$  之外的解. 方程  $T(w, z) = \lambda(w, z)$  等价于联立方程

$$-z = \lambda w, \quad w = \lambda z$$

将第二个方程所给出的  $w$  的表达式代入第一个方程, 得

$$-z = \lambda^2 z$$

由于  $z$  不能等于 0 【否则由式 (5.10) 得  $w = 0$ , 而我们要找的式 (5.10) 的解须满足  $(w, z)$  不是向量 0】，所以上述方程就可化为

$$-1 = \lambda^2$$

此方程的解是  $\lambda = i$  和  $\lambda = -i$

可验证,  $i$  和  $-i$  都是  $T$  的特征值. 的确, 对应于特征值  $i$  的特征向量是形如  $(w, -wi)$  的向量, 其中  $w \in \mathbf{C}$  且  $w \neq 0$ ; 对应于特征值  $-i$  的特征向量是形如  $(w, wi)$  的向量, 其中  $w \in \mathbf{C}$  且  $w \neq 0$

在下面证明中, 再次使用了如下等价关系:

$$T\nu = \lambda\nu \longleftrightarrow (T - \lambda I)\nu = 0$$

### 5.11 线性无关的特征向量

$T \in \mathcal{L}(V)$ . 那么分别对应于  $T$  的不同特征值的特征向量构成的每个组都线性无关

#### 证明

假设欲证结论不成立. 那么存在最小的正整数  $m$ , 使得  $T$  对应于其互异特征值  $\lambda_1, \dots, \lambda_m$  的特征向量  $\nu_1, \dots, \nu_m$  构成线性相关向量组 (注意  $m \geq 2$ , 因为根据定义, 特征向量是非零的)

于是, 存在  $a_1, \dots, a_m \in \mathbf{F}$ , 其中各数均非零 (因为  $m$  最小 $1$ ), 使得

$$a_1\nu_1 + \dots + a_m\nu_m = 0$$

将  $T - \lambda_m I$  作用于上式两侧, 得

$$a_1(\lambda_1 - \lambda_m)\nu_1 + \dots + a_{m-1}(\lambda_{m-1} - \lambda_m)\nu_{m-1} = 0$$

因为特征值  $\lambda_1, \dots, \lambda_m$  互异, 所以上式中各系数均不等于0. 于是  $\nu_1, \dots, \nu_{m-1}$  就是由对应于  $T$  的互异特征值的  $m-1$  个特征向量构成的线性相关向量组, 这和  $m$  最小的假设相矛盾. 此矛盾就说明欲证命题成立

上面结论可引出下面结论的一个简短证明. 下面结论给一个算子的互异特征值个数设置了上界

### 5.12 算子的特征值个数不多于向量空间的维数

设  $V$  是有限维的. 那么  $V$  上的每个算子最多有  $\dim V$  个互异特征值

#### 证

令  $T \in \mathcal{L}(V)$ . 设  $\lambda_1, \dots, \lambda_m$  是  $T$  的互异特征值. 令  $\nu_1, \dots, \nu_m$  是与之对应的特征向量. 那么 5.11 表明  $\nu_1, \dots, \nu_m$  线性无关. 于是  $m \leq \dim V$  (见2.22), 证毕

## 将多项式作用于算子

算子 (它们将向量空间映射到其自身) 的理论比一般的线性映射的理论更丰富的主要原因, 就在于算子可以自乘为幂. 在本小节中, 我们给出算子的幂的定义, 并提出将多项式作用于算子的概念. 这个概念, 将是我们在下一节中证明“非零有限维复向量空间上的每个算子都有特征值”时, 所使用的关键工具. 若  $T$  是一个算子, 那么  $TT$  是有意义的 (见3.7), 且它也是  $T$  的定义空间上的算子. 我们通常把  $TT$  写成  $T^2$ . 更一般地说来, 我们定义  $T^m$  如下

### 5.13 记号

设  $T \in \mathcal{L}(V)$ ,  $m$  是正整数

定义  $T^m \in \mathcal{L}(V)$  为  $T^m = T \cdots T$ .  $\lambda_m \{\uparrow zT\}$

定义  $T^0$  为  $V$  上的恒等算子  $I$

若  $T$  是可逆的, 且其逆为  $T^{-1}$ , 那么  $T^{-m} \in \mathcal{L}(V)$  的定义是

$$T^{-m} = (T^{-1})^m$$

你应自行验证, 若  $T$  为算子, 那么

其中, 若  $T$  是可逆的, 则  $m$  和  $n$  是任意整数; 若  $T$  不可逆, 则  $m$  和  $n$  为非负整数

定义了算子的幂, 我们就可以定义将多项式作用于算子是什么含义了

### 5.14 记号: $p(T)$

$T \in \mathcal{L}(V)$  且  $p \in \mathcal{P}(\mathbf{F})$  是由下式给出的多项式: 对所有  $z \in \mathbf{F}$ ,

$$p(z) = a_0 + a_1z + a_2z^2 + \dots + a_mz^m$$

那么  $p(T)$  是  $V$  上的算子

$$p(T) = a_0I + a_1T + a_2T^2 + \cdots + a_mT^m$$

我们把  $p$  应用在算子上而不仅是  $\mathbf{F}$  中的数上, 给符号  $p$  增添了一种新用法. 我们的想法是, 要得到  $p(T)$ , 将  $p$  的定义式中的  $z$  替换成  $T$  即可. 注意,  $p(z)$  中的常数项  $a_0$  变成了算子  $a_0I$  (这写法很合理: 因为  $a_0 = a_0z^0$ , 所以我们应将  $a_0$  替换成  $a_0T^0$ , 也就是  $a_0I$ )

### 5.15 将多项式作用于微分算子

$D \in \mathcal{L}(\mathcal{P}(\mathbf{R}))$  是由  $Dq = q'$  定义的微分算子,  $p$  是定义为  $p(x) = 7 - 3x + 5x^2$  的多项式. 那么  $p(D) = 7I - 3D + 5D^2$ . 于是对每个  $q \in \mathcal{P}(\mathbf{R})$ , 都有

$$(p(D))q = 7q - 3q' + 5q''$$

若取定一算子  $T \in \mathcal{L}(V)$ , 则从  $\mathcal{P}(\mathbf{F})$  到  $\mathcal{L}(V)$  的函数  $p \rightarrow p(T)$  是线性的. 你应自行验证这一点

### 5.16 多项式的乘积

$p, q \in \mathcal{P}(\mathbf{F})$ , 那么  $pq \in \mathcal{P}(\mathbf{F})$  是按下式定义的多项式: 对所有  $z \in \mathbf{F}$ ,

$$(pq)(z) = p(z)q(z)$$

如下面结论(b) 所示, 对单个算子的多项式取乘积时, 顺序无关紧要

### 5.17 乘积性质

$p, q \in \mathcal{P}(\mathbf{F})$  且  $T \in \mathcal{L}(V)$ . 那么

- (a)  $(pq)(T) = p(T)q(T)$
- (b)  $p(T)q(T) = q(T)p(T)$

### 证

不太正式的证明: 利用分配性质展开多项式之乘积时, 跟符号用  $z$  还是  $T$  没有关系

正式证明

- (a) 设  $p(z) = \sum_{j=0}^m a_j z^j$  且  $q(z) = \sum_{k=0}^n b_k z^k$  对所有  $z \in \mathbf{F}$  都成立. 那么

$$(pq)(z) = \sum_{j=0}^m \sum_{k=0}^n a_j b_k z^{j+k}$$

于是

$$\begin{aligned} (pq)(T) &= \sum_{j=0}^m \sum_{k=0}^n a_j b_k T^{j+k} \\ &= \left( \sum_{j=0}^m a_j T^j \right) \left( \sum_{k=0}^n b_k T^k \right) \\ &= p(T)q(T) \end{aligned}$$

- (b) 运用(a) 两次, 我们可得  $p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T)$

之前我们已经看到, 若  $T \in \mathcal{L}(V)$ , 那么子空间  $\text{null } T$  和  $\text{range } T$  在  $T$  下是不变的 (见例5.4). 现在我们证明,  $T$  的每个多项式的零空间和值域在  $T$  下也都是不变的

### 5.18 的零空间和值域在 $T$ 下不变

$T \in \mathcal{L}(V)$  且  $p \in \mathcal{P}(\mathbf{F})$  那么  $\text{null } p(T)$  和  $\text{range } p(T)$  在  $T$  下不变

### 三 证

设  $u \in \text{null}_{p(T)}$  . 那么  $p(T)u = 0$  . 于是

$$(p(T))(Tu) = (p(T)T)(u) = (Tp(T))(u) = T(p(T)u) = T(0) = 0$$

因此  $Tu \in \text{null}(T)$  . 于是  $\text{null}_{p(T)}$  在  $T$  下是不变的

设  $u \in \text{range } p(T)$  , 那么存在  $v \in V$  使得  $u = p(T)v$  . 于是

$$Tu = T(p(T)v) = p(T)(Tv)$$

因此  $Tu \in \text{range } p(T)$  . 于是  $\text{range } p(T)$  在  $T$  下是不变的

## 5B 最小多项式

### 复向量空间上特征值的存在性

现在我们给出关于有限维复向量空间上算子的一个核心结论

#### 5.19 特征值的存在性

非零有限维复向量空间上的每个算子都有特征值

### 三 证

设  $V$  是有限维复向量空间, 其维数  $n > 0$  ,  $T \in \mathcal{L}(V)$  . 取  $v \in V$  且  $v \neq 0$  . 那么

$$v, T\bar{v}, T^2\bar{v}, \dots, T^n\bar{v}$$

不是线性无关的, 因为  $V$  的维数是  $n$  , 该组的长度却是  $n+1$  . 因此上述向量的某个线性组合 (其中各系数不全为0) 会等于向量0. 于是, 存在次数最小的非常值多项式  $p$  使得

$$p(T)\bar{v} = 0$$

由代数基本定理的版本一 (见 4.12), 存在  $\lambda \in \mathbf{C}$  使得  $p(\lambda) = 0$  , 故存在多项式  $q \in \mathcal{P}(\mathbf{C})$  使得

$$p(z) = (z - \lambda)q(z)$$

对每个  $z \in \mathbf{C}$  都成立 (见4.6). 由此可得 (利用 5.17)

$$0 = p(T)\bar{v} = (T - \lambda I)(q(T)\bar{v})$$

因为  $q$  的次数小于  $p$  的次数, 所以我们可知  $q(T)\bar{v} \neq 0$  . 于是上述等式就表明  $\lambda$  是  $T$  的一个特征值,  $q(T)\bar{v}$  是对应于它的特征向量

上述证明的关键是运用代数基本定理. 习题 16 后的注记有助于解释为何代数基本定理与上述结论结合得如此紧密

上述结论中的  $\mathbf{F} = \mathbf{C}$  这个假设不能替换为  $\mathbf{F} = \mathbf{R}$ . ——从例 5.9 可看出. 下例则说明, 上述结论中的“有限维”这个假设也不能去掉



#### 5.20 复向量空间上的一个无特征值的算子

$T \in \mathcal{L}(\mathcal{P}(\mathbf{C}))$  定义为  $(Tp)(z) = zp(z)$  . 若  $p \in \mathcal{P}(\mathbf{C})$  是非零多项式, 那么  $Tp$  的次数比  $p$  的次数多1, 从而  $Tp$  不会等于  $p$  的标量倍. 因此  $T$  没有特征值

因为  $\mathcal{P}(\mathbf{C})$  是无限维的, 所以本例并不与上述结论矛盾

## 特征值与最小多项式

在本小节中，我们引入与每个算子都相关联的一个重要的多项式。我们从下面定义出发

### 5.21 首一多项式

首一多项式是最高次项系数等于1的多项式

#### 例

如多项式  $2 + 9z^2 + z^7$  是次数为7的首一多项式

### 5.22 最小多项式的存在性、唯一性和次数

设  $V$  是有限维的， $T \in \mathcal{L}(V)$ 。那么存在唯一的次数最小的首一多项式  $p \in \mathcal{P}(\mathbf{F})$ ，使得  $p(T) = 0$ 。此外， $\deg p \leq \dim V$

#### 证明

(存在性)

若  $\dim V = 0$ ，那么  $I$  就是  $V$  上的零算子，从而我们取  $p$  为常值多项式1即可

现在对  $\dim V$  用归纳法，于是假设  $\dim V > 0$  且欲证结论对于所有维数更小的向量空间上的所有算子都成立。设  $u \in V$  ( $u \neq 0$ )。组  $u, Tu, \dots, T^{\dim V} u$  的长度为  $1 + \dim V$ ，因而是线性相关的。根据线性相关性引理 (2.19)，存在最小正整数  $m \leq \dim V$  使得  $T^m u$  是  $u, Tu, \dots, T^{m-1} u$  的线性组合。于是存在标量  $c_0, c_1, c_2, \dots, c_{m-1} \in \mathbf{F}$  使得

$$c_0 u + c_1 Tu + \dots + c_{m-1} T^{m-1} u + T^m u = 0$$

定义首一多项式  $q \in \mathcal{P}_m(\mathbf{F})$  为

$$q(z) = c_0 + c_1 z + \dots + c_{m-1} z^{m-1} + z^m$$

那么式 (5.23) 就表明  $q(T)u = 0$

若  $k$  为非负整数，那么

$$q(T)(T^k u) = T^k(q(T)u) = T^k(0) = 0$$

线性相关性引理 (2.19) 表明， $u, Tu, \dots, T^{m-1} u$  线性无关，于是由上式可得  $\dim \text{null } q(T) \geq m$  因此

$$\dim \text{range } q(T) = \dim V - \dim \text{null } q(T) \leq \dim V - m$$

因为  $\text{range } q(T)$  在  $T$  下是不变的（由 5.18），所以我们可以将归纳假设应用于  $\text{range } q(T)$  上的算子  $T|_{\text{range } q(T)}$ ，从而存在首一多项式  $s \in \mathcal{P}(\mathbf{F})$  使得

因此，对于所有  $\nu \in V$ ，我们有

$$((sq)(T))(\nu) = s(T)(q(T)\nu) = 0$$

这是因为  $q(T)\nu \in \text{range } q(T)$  且  $s(T)|_{\text{range } q(T)} = s(T|_{\text{range } q(T)}) = 0$ 。于是  $sq$  是满足  $\deg sq \leq \dim V$  和  $(sq)(T) = 0$  的首一多项式

上段表明，存在次数不高于  $\dim V$  的首一多项式，其作用于  $T$  可得算子0。于是，作用于  $T$  得到算子0的次数最小的首一多项式存在（唯一性）

令  $p \in \mathcal{P}(\mathbf{F})$  是使  $p(T) = 0$  成立的次数最小的首一多项式，设  $r \in \mathcal{P}(\mathbf{F})$  是次数与  $p$  相同的首一多项式，且  $r(T) = 0$ 。那么  $(p-r)(T) = 0$  且  $\deg(p-r) < \deg p$  若  $p-r$  不等于0，那么将  $p-r$  除以其最高次项的系数，可得一个作用于  $T$  会得到算子0的首一多项式（且其次数小于  $p$  的次数）。于是  $p-r = 0$

上述结论为下面的定义提供了依据

### 5.24 最小多项式

$V$  是有限维的，且  $T \in \mathcal{L}(V)$ 。那么  $T$  的最小多项式是唯一使得  $p(T) = 0$  成立的次数最小的首一多项式  $p \in \mathcal{P}(\mathbf{F})$

为计算一个算子  $T \in \mathcal{L}(V)$  的最小多项式，我们需要求出使得  $c_0I + c_1T + \cdots + c_{m-1}T^{m-1} = -T^m$  有解  $c_0, c_1, \dots, c_{m-1} \in \mathbf{F}$  的最小正整数  $\nabla_m$ 。如果我们选取  $V$  的一个基，再将上述方程中的  $T$  换成  $T$  的矩阵，那么可以将上述方程看成有  $(\dim V)^2$  个方程、 $m$  个未知数  $c_0, c_1, \dots, c_{m-1} \in \mathbf{F}$  的线性方程组。要知道解是否存在，我们用高斯消元法或者线性方程组的其他快速解法，对  $m = 1, 2, \dots$  诸值依次作检验，直到发现有解即可。由 5.22，对于某个最小的正整数  $m \leq \dim V$ ，上述方程组有解。那么  $T$  的最小多项式就是  $c_0 + c_1z + \cdots + c_{m-1}z^{m-1} + z^m$ 。更快（且更常用）的做法是，选取  $\nu \in V \setminus \{0\}$  并考虑方程  $c_0\nu + c_1T\nu + \cdots + c_{\dim V-1}T^{\dim V-1}\nu = -T^{\dim V}\nu$ 。利用  $V$  的一个基，将上述方程转化为有  $\dim V$  个方程、 $\dim V$  个未知数  $c_0, c_1, \dots, c_{\dim V-1}$  的线性方程组。若该方程组有唯一解  $c_0, c_1, \dots, c_{\dim V-1}$ （这是最常出现的情况），那么标量  $c_0, c_1, \dots, c_{\dim V-1}, 1$  就是  $T$  的最小多项式的各系数（因为 5.22 指出最小多项式的次数最多是  $\dim V$ ）。

考虑  $\mathbf{R}^4$  上的算子（可看成关于标准基的  $4 \times 4$  矩阵），运用上段方法并取  $\nu = (1, 0, 0, 0)$

此处的百分比估计值基于对数百万个随机矩阵的实测

对于由区间  $[10, 10]$  上的整数构成的  $4 \times 4$  矩阵，此法适用于其中超过 99.8% 的矩阵；而对于由区间  $[-100, 100]$  上的整数构成的  $4 \times 4$  矩阵，此法适用于其中超过 99.999% 的矩阵。

下例展示了上面讨论的快速方法

### 5.26 $\mathbf{F}^5$ 上一算子的最小多项式

设  $T \in \mathcal{L}(\mathbf{F}^5)$ ，其关于标准基  $e_1, e_2, e_3, e_4, e_5$  的矩阵为

$$M(T) = \begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

在式 (5.25) 中取  $\nu = e_1$ ，有

$$\begin{aligned} Te_1 &= e_2 & T^4e_1 &= T(T^3e_1) = Te_4 = e_5, \\ T^2e_1 &= T(Te_1) = Te_2 = e_3 & T^5e_1 &= T(T^4e_1) = Te_5 = -3e_1 + 6e_2 \\ T^3e_1 &= T(T^2e_1) = Te_3 = e_4 & T^3e_1 &= T(T^2e_1) = Te_3 = e_4 \end{aligned}$$

于是  $3e_1 - 6Te_1 = -T^5e_1$ 。组  $e_1, Te_1, T^2e_1, T^3e_1, T^4e_1$  就等于组  $e_1, e_2, e_3, e_4, e_5$ ，是线性无关的，所以该组的其他线性组合都不会等于  $-T^5e_1$ 。因此  $T$  的最小多项式就是  $3 - 6z + z^5$ 。

回忆一下，根据定义， $V$  上算子的特征值和  $\mathcal{P}(\mathbf{F})$  中多项式的零点都必为  $\mathbf{F}$  中的元素。特别地，若  $\mathbf{F} = \mathbf{R}$ ，那么特征值和零点都必为实数。

### 5.27 特征值即最小多项式的零点

$V$  是有限维的且  $T \in \mathcal{L}(V)$

- (a)  $T$  的最小多项式的零点即  $T$  的特征值
- (b) 若  $V$  是复向量空间，那么  $T$  的最小多项式具有形式  $(z - \lambda_1) \cdots (z - \lambda_m)$ ，其中  $\lambda_1, \dots, \lambda_m$  是  $T$  的所有特征值（可能有重复）

### 证

令  $p$  是  $T$  的最小多项式

- (a) 首先假设  $\lambda \in \mathbf{F}$  是  $p$  的零点之一。那么  $p$  可以写成如下形式

$$p(z) = (z - \lambda)q(z)$$

其中  $q$  是系数在  $\mathbf{F}$  中的首一多项式（见 4.6）。因为  $p(T) = 0$ ，所以对全体  $\nu \in V$  我们都有

$$0 = (T - \lambda I)(q(T)\nu)$$

因为  $\deg q = (\deg p) - 1$  且  $p$  是  $T$  的最小多项式，所以至少存在一个向量  $\nu \in V$  使得  $q(T)\nu \neq 0$ 。因而，上述等式表明  $\lambda$  是  $T$  的一个特征值，即  $p$  的零点都为  $T$  的特征值。为证明  $T$  的每个特征值都是  $p$  的一个零点，现设  $\lambda \in \mathbf{F}$  是  $T$  的一个特征值。于是，存在  $\nu \in V$  ( $\nu \neq 0$ ) 使得  $T\nu = \lambda\nu$ 。将  $T$  反复作用在此式两端，即得对每个非负整数  $k$  均有  $T^k\nu = \lambda^k\nu$ 。于是

$$p(T)\nu = p(\lambda)\nu$$

因为  $p$  是  $T$  的最小多项式，所以我们有  $p(T)\nu = 0$ 。所以上述等式就表明  $p(\lambda) = 0$ 。于是  $\lambda$  是  $p$  的一个零点，即  $T$  的特征值都为  $p$  的零点。

- (b) 利用(a) 和代数基本定理的版本二（见 4.13）即可证明。

非零多项式的互异零点个数最多不超过其次数（见4.8）。因此，将上个结论的(a)结合“ $V$ 上的算子的最小多项式的次数不超过  $\dim V$ ”这个结论，就能以另一种方式证明5.12（该定理内容为，“ $V$ 上的算子至多有  $\dim V$  个互异的特征值”）

每个首一多项式都是某个算子的最小多项式（该命题在习题 16 中证明），这是对例 5.26 的推广。因而，5.27 (a) 就表明，求出算子的特征值的确切表达式，就等价于求出多项式的零点的确切表达式（因而有些算子的特征值的确切值是求不出来的）

### 5.28 无法确切求出特征值的算子

$$T \in \mathcal{L}(\mathbb{C}^5)$$

$$T(z_1, z_2, z_3, z_4, z_5) = (-3z_5, z_1 + 6z_5, z_2, z_3, z_4).$$

关于  $\mathbb{C}^5$  的标准基的矩阵如例 5.26 所示。在那个例子中，我们求得  $T$  的最小多项式为

$$3 - 6z + z^5$$

上述多项式的零点都无法用有理数及其根式与通常的算数规则来表达（这个结论的证明远远超出了线性代数的范围）。又由于上述多项式的零点也就是  $T$  的特征值【由 5.27 (a)】，所以我们就无法用我们熟悉的任何形式来确切表达  $T$  的任何特征值。

用数值计算方法（此处我们不做讨论）可得上述多项式的零点，也即  $T$  的特征值，近似为下面这五个复数：注意，该多项式的两个非实数零点互成复共轭。这个结果在我们意料之中，毕竟上述多项式是实系数多项式（见 4.14）

下面结论完整地刻画了作用于算子时会得到算子 0 的多项式

### 5.29 $q(T) = 0 \iff q$ 是最小多项式的多项式倍

$V$  是有限维的， $T \in \mathcal{L}(V)$ ，且  $q \in \mathcal{P}(\mathbb{F})$ 。那么  $q(T) = 0$  当且仅当  $q$  是  $T$  的最小多项式的多项式倍

#### 三 证

令  $p$  表示  $T$  的最小多项式。

首先假设  $g(T) = 0$ 。由多项式的带余除法（4.9），存在多项式  $s, r \in \mathcal{P}(\mathbb{F})$  使得

$$q = ps + r$$

且  $\deg r < \deg p$ ，我们有

$$0 = q(T) = p(T)s(T) + r(T) = r(T)$$

上式表明  $r = 0$ （否则，将  $r$  除以其最高次项的系数可得一个首一多项式，将该多项式作用于  $T$  能得到 0，而该多项式的次数比最小多项式的次数还低，这就产生了矛盾）。于是式(5.30)化为  $q = ps$ 。因此， $q$  是  $p$  的多项式倍，此即待证命题的一个方向。

为了证明另一方向，现假设  $q$  是  $p$  的多项式倍。那么存在多项式  $s \in \mathcal{P}(\mathbb{F})$  使得  $q = ps$ 。我们有

$$q(T) = p(T)s(T) = 0s(T) = 0$$

则此方向得证。

下面结论是上述结论的一个很漂亮的推论

### 5.31 受限算子的最小多项式

设  $V$  是有限维的， $T \in \mathcal{L}(V)$ ，且  $U$  是  $V$  的在  $T$  下不变的子空间。那么  $T$  的最小多项式是  $T|_U$  的最小多项式的多项式倍

#### 三 证

设  $p$  是  $T$  的最小多项式。于是对于所有  $v \in V$ ，有  $p(T)v = 0$ 。特别地，对于所有  $u \in U$ ， $p(T)u = 0$ 。于是  $p(T|_U) = 0$ 。现在将 5.29 中的  $T$  换成  $T|_U$ （即将该结论作用于  $T|_U$  而不是  $T$  上），即可说明  $p$  是  $T|_U$  的最小多项式的多项式倍。

有关商算子的类似结论详见习题25

下面结论表明，算子的最小多项式的常数项决定了这个该算子是否可逆

### 5.32 $T$ 不可逆 $T$ 的最小多项式的常数项为0

设  $V$  是有限维的且  $T \in \mathcal{L}(V)$ 。那么  $T$  不可逆，当且仅当  $T$  的最小多项式的常数项为0

#### 证

设  $T \in \mathcal{L}(V)$  且  $p$  是  $T$  的最小多项式。那么  $T$  不可逆  $\iff 0$  是  $T$  的特征值  $\iff 0$  是  $p$  的零点  $\iff p$  的常数项是0，其中，第一个等价成立是根据5.7，第二个等价成立是根据5.27 (a)，而最后一个等价成立是因为  $p$  的常数项就等于  $p(0)$

## 奇数维的实向量空间上的特征值

下面的结论，将是我们用于证明“奇数维的实向量空间上的算子都有特征值”的关键工具

### 5.33 偶数维的零空间

设  $\mathbf{F} = \mathbf{R}$  且  $V$  是有限维的，并设  $T \in \mathcal{L}(V), b, c \in \mathbf{R}$  使得  $b^2 < 4c$ 。那么  $\dim \text{null}(T^2 + bT + cI)$  是偶数

#### 证

回忆一下由5.18， $\text{null}(T^2 + bT + cI)$  在  $T$  下是不变的。将  $V$  替换成  $\text{null}(T^2 + bT + cI)$  并将  $T$  替换成限制在  $\text{null}(T^2 + bT + cI)$  上的  $T$ ，我们便可认为  $T^2 + bT + cI = 0$ ，这样一来，我们需证明的就是  $\dim V$  为偶数

设  $\lambda \in \mathbf{R}$  及  $\nu \in V$  满足  $T\nu = \lambda\nu$ ，则

$$0 = (T^2 + bT + cI)\nu = (\lambda^2 + b\lambda + c)\nu = \left(\left(\lambda + \frac{b}{2}\right)^2 + c - \frac{b^2}{4}\right)\nu$$

上式最外层括号中的数是个正数。于是上式表明  $\nu = 0$ 。因此我们证明了  $T$  没有特征向量

令  $U$  是  $V$  的一个子空间，它在  $T$  下是不变的，并且，所有在  $T$  下不变且维数为偶的  $V$  的子空间中，它具有最大的维数。若  $U = V$ ，那么我们的证明完成了；否则，设存在某个  $w \in V$  满足  $w \notin U$

令  $W = (w, Tw)$ 。那么  $W$  在  $T$  下不变，因为  $T(Tw) = -bTw - cw$ 。另外， $\dim W = 2$ ，否则  $w$  就会成为  $T$  的特征向量。这样一来

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W) = \dim U + 2$$

其中  $U \cap W = \{0\}$ ，否则  $U \cap W$  就会成为  $V$  的一个1维子空间且在  $T$  下不变（这是不可能的，因为  $T$  没有特征向量）

因为  $U + W$  在  $T$  下不变，所以上述等式表明，存在在  $T$  下不变、维数为偶且大于  $\dim U$  的  $V$  的子空间。于是， $U \neq V$  的假设不正确。因此， $V$  的维数为偶数

接下来这条结果指出，奇数维的向量空间上的每个算子都有特征值。我们已知此结论对于复向量空间成立（且无需奇数维的假设）。于是在下面的证明中，我们将假设  $\mathbf{F} = \mathbf{R}$

### 5.34 奇数维向量空间上的算子总有特征值

奇数维向量空间上的每个算子都有特征值

#### 证

设  $\mathbf{F} = \mathbf{R}$  且  $V$  是有限维的。令  $n = \dim V$ ，并设  $n$  为奇数。令  $T \in \mathcal{L}(V)$ 。为证明  $T$  有特征值，我们对  $n$  用步长为 2 的归纳法。首先，注意到若  $\dim V = 1$ ，那么欲证结论成立，因为这样的话  $V$  中的每个向量都是  $T$  的特征向量

现在，假设  $n \geq 3$  且欲证结论对于维数小于  $n$  的所有奇数维向量空间上的所有算子都成立令  $p$  表示  $T$  的最小多项式。若对于某  $\lambda \in \mathbf{R}$ ， $p$  是  $x - \lambda$  的多项式倍，那么  $\lambda$  是  $T$  的一个特征值【由 5.27 (a)】，证明就完成了。于是我们可设存在  $b, c \in \mathbf{R}$  使得  $b^2 < 4c$  且  $p$  是  $x^2 + bx + c$  的多项

式倍 (见4.16)

从而存在首一多项式  $q \in \mathcal{P}(\mathbf{R})$  使得  $p(x) = q(x)(x^2 + bx + c)$  对所有  $x \in \mathbf{R}$  成立, 这样

$$0 = p(T) = (q(T))(T^2 + bT + cI)$$

由此,  $q(T)$  等于  $\text{range}(T^2 + bT + cI)$  上的算子 0. 因为  $\deg q < \deg p$  且  $p$  是  $T$  的最小多项式, 所以上式表明  $\text{range}(T^2 + bT + cI) \neq V$  由线性映射基本定理 (3.21)

$$\dim V = \dim \text{null}(T^2 + bT + cI) + \dim \text{range}(T^2 + bT + cI)$$

因为  $\dim V$  是奇数 (由命题假设) 且  $\dim \text{null}(T^2 + bT + cI)$  是偶数 (由5.33), 所以上式就说明  $\dim \text{range}(T^2 + bT + cI)$  是奇数 所以  $\text{range}(T^2 + bT + cI)$  是  $V$  的在  $T$  下不变的子空间 (由5.18), 其维数为奇且小于  $\dim V$  这样一来, 由我们的归纳假设可得, 限制于  $\text{range}(T^2 + bT + cI)$  的算子  $T$  有特征值, 也即  $T$  有特征值

上述结果的其他证明方法见8B 节的习题23 和9C 节的习题10

## 5C 上三角矩阵

在第 3 章中, 我们定义了从一个有限维向量空间到另一个有限维向量空间的线性映射的矩阵. 该矩阵依赖于这两个向量空间的基的选取. 我们现在研究将一向量空间映射至自身的算子, 需要着重考虑的就是仅用一个基来描述它

### 5.35 算子的矩阵

设  $T \in \mathcal{L}(V)$ .  $T$  关于  $V$  的基  $\nu_1, \dots, \nu_n$  的矩阵是  $n \times n$  矩阵

$$M(T) = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \cdots & A_{n,n} \end{pmatrix}$$

其中的元素  $A_{j,k}$  定义为

$$T\nu_k = A_{1,k}\nu_1 + \cdots + A_{n,k}\nu_n$$

若根据上下文无法明确看出选取哪个基, 就用  $M(T, (\nu_1, \dots, \nu_n))$  这个记号

### 5.36 一算子关于标准基的矩阵

算子的矩阵是方阵 (其行数等于列数), 与前面我们讨论的一般线性映射的长方形矩阵不同

若  $T$  是  $\mathbf{F}^n$  上的算子, 且未明确基的选取, 矩阵  $\nabla M(T)$  的第  $k$  列, 是由将  $T\nu_k$  写成基那么就假定所取的基是标准基 (其中第  $k$   $\nu_1, \dots, \nu_n$  的线性组合时所用的系数构成的基向量除第  $k$  个坐标为 1 外, 其余坐标均为 0). 此时你可以认为  $\nabla M(T)$  的第  $k$  列是将  $T$  作用于第  $k$  个基向量所得的结果 (这里把  $\mathbf{F}^n$  中的元素与  $n \times 1$  列向量等同起来看)

$T \in \mathcal{L}(\mathbf{F}^3)$ ,  $T(x, y, z) = (2x + y, 5y + 3z, 8z)$  关于  $\mathbf{F}^3$  的标准基的矩阵为

$$M(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}$$

你应自行验证

线性代数的中心目标之一, 就是证明对于有限维向量空间  $V$  上的算子  $T$ , 存在  $V$  的一个基, 使得  $T$  关于该基有个相当简单的矩阵. 这个说法比较含糊, 说得再确切些, 就是我们尝试选取  $V$  的一个基, 使得  $\nabla M(T)$  有很多 0

若  $V$  是有限维向量空间, 那么我们凭已学过的知识, 足以证明存在  $V$  的一个基, 使得  $T$  关于该基的矩阵的第一列中, 除了第一个元素可能非零外, 其余元素均是 0. 换言之, 存在

$V$  的一个基, 使得  $T$  关于该基的矩阵形如

此处 \* 代表除了第一列以外的所有元素\*. 为证明以上命题, 令  $\lambda$  是  $T$  的一个特征值 (由 5.19 可知其存在性) 并令  $\nu$  是对应于它的特征向量. 将  $\nu$  扩充为  $V$  的一个基, 那么  $T$  关于该基的矩阵就有上述形式. 很快我们就会看到, 可以选取  $V$  的一个基, 使  $T$  关于该基的矩阵有更多的 0

### 5.37 矩阵的对角线

方阵的对角线由从它的左上角到右下角的直线上的元素所构成

### 例

如例5.36 中的矩阵

$$M(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}$$

的对角线是由2、5、8 所构成，在上式中以红色标出

### 5.38 上三角矩阵

称一个方阵为上三角矩阵，若其中所有在对角线之下的元素都是0

### 例

如上面所示的  $3 \times 3$  矩阵就是上三角矩阵我们一般将上三角矩阵表示成下面的形式

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

上述矩阵中的0 表示在这个  $n \times n$  矩阵中，在对角线之下的元素都等于0

常用 \* 来表示矩阵中那些我们未知的，或与正在讨论的问题无关的元素

若  $n$  很大，那么在  $n \times n$  上三角矩阵中至少有将近一半的元素都是0下面结论在上三角矩阵和不变子空间之间建立了有用的联系

### 5.39 上三角矩阵的条件

设  $T \in \mathcal{L}(V)$  且  $\nu_1, \dots, \nu_n$  是  $V$  的一个基。那么下面几条结论等价

- (a)  $T$  关于  $\nu_1, \dots, \nu_n$  的矩阵是上三角矩阵
- (b) 对每个  $k = 1, \dots, n$ ，均有  $\text{span}(\nu_1, \dots, \nu_k)$  在  $T$  下不变
- (c) 对每个  $k = 1, \dots, n$ ，均有  $T\nu_k \in \text{span}(\nu_1, \dots, \nu_k)$

### 证

设(a) 成立。为证明(b) 成立，设  $k \in \{1, \dots, n\}$ 。若  $j \in \{1, \dots, n\}$ ，那么

$$T\nu_j \in \text{span}(\nu_1, \dots, \nu_j),$$

这是因为  $T$  关于  $\nu_1, \dots, \nu_n$  的矩阵是上三角矩阵。因为若  $j \leq k$  则  $\text{span}(\nu_1, \dots, \nu_j) \subseteq \text{span}(\nu_1, \dots, \nu_k)$ ，所以我们可知，对每个  $j \in \{1, \dots, k\}$ ，有

$$T\nu_j \in \text{span}(\nu_1, \dots, \nu_k).$$

于是， $\text{span}(\nu_1, \dots, \nu_k)$  在  $T$  下不变，这就证明了(a) 蕴涵(b)

现在设(b) 成立，因此对每个  $k = 1, \dots, n$ ，均有  $\text{span}(\nu_1, \dots, \nu_k)$  在  $T$  下不变。特别地，对每个  $k = 1, \dots, n$ ，均有  $T\nu_k \in \text{span}(\nu_1, \dots, \nu_k)$ 。于是(b) 蕴涵(c)

现在设(c) 成立，因此对每个  $k = 1, \dots, n$ ，均有  $T\nu_k \in \text{span}(\nu_1, \dots, \nu_k)$ 。这就意味着，当用基向量  $\nu_1, \dots, \nu_n$  的线性组合来表达  $T\nu_k$  时，我们只需用到向量  $\nu_1, \dots, \nu_k$ 。因此， $\nabla M(T)$  对角线之下所有元素都是0。于是， $\nabla M(T)$  是个上三角矩阵，这就证明了(c) 蕴涵(a)

我们证明了(a) () (c) (a)，也就证明了(a)、(b) 和 (c) 是等价的

下面结论告诉我们，若  $T \in \mathcal{L}(V)$ ，且关于  $V$  的某个基的矩阵是

$$M(T) = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

那么  $T$  满足一个依赖于  $\lambda_1, \dots, \lambda_n$  的简单等式

### 5.40 具有上三角矩阵的算子满足的等式

$T \in \mathcal{L}(V)$  且存在  $V$  的一个基，使得  $T$  关于该基有上三角矩阵，且该矩阵的对角线元素是  $\lambda_1, \dots, \lambda_n$ ，那么

$$(T - \lambda_1 I) \cdots (T - \lambda_n I) = 0$$

#### 证

令  $\nu_1, \dots, \nu_n$  是  $V$  的一个基，且  $T$  关于该基有对角线元素为  $\lambda_1, \dots, \lambda_n$  的上三角矩阵那么  $T\nu_1 = \lambda_1\nu_1$ ，这意味着  $(T - \lambda_1 I)\nu_1 = 0$ ，由此可推出  $(T - \lambda_1 I) \cdots (T - \lambda_m I)\nu_1 = 0$  对于  $m = 1, \dots, n$  成立（利用各  $T - \lambda_j I$  和各  $T - \lambda_k I$  的可交换性可得）

注意到  $(T - \lambda_2 I)\nu_2 \in (\nu_1)$ 。于是  $(T - \lambda_1 I)(T - \lambda_2 I)\nu_2 = 0$ （根据上段可得），并可由此推出  $(T - \lambda_1 I) \cdots (T - \lambda_m I)\nu_2 = 0$  对于  $m = 2, \dots, n$  成立（利用各  $T - \lambda_j I$  和各  $T - \lambda_k I$  的可交换性可得）

注意到  $(T - \lambda_3 I)\nu_3 \in \text{span}(\nu_1, \nu_2)$ 。于是由上段可得， $(T - \lambda_1 I)(T - \lambda_2 I)(T - \lambda_3 I)\nu_3 = 0$ ，并可由此推出  $(T - \lambda_1 I) \cdots (T - \lambda_m I)\nu_3 = 0$  对于  $m = 3, \dots, n$  成立（利用各  $T - \lambda_j I$  和各  $T - \lambda_k I$  的可交换性可得）

按照这种方式继续下去，我们就会发现对每个  $k = 1, \dots, n$  都有  $(T - \lambda_1 I) \cdots (T - \lambda_n I)\nu_k = 0$  于是  $(T - \lambda_1 I) \cdots (T - \lambda_n I)$  是 0 算子，因为它把  $V$  的一个基中所有向量都对应至 0

由算子的矩阵确切计算出算子的特征值的方法并不存在，这实为憾事。然而，如果我们有幸找到一个基，使得算子关于该基的矩阵是上三角的，那么特征值的计算问题就变得很平凡，如下面结论所示

### 5.41 由上三角矩阵确定特征值

设  $T \in \mathcal{L}(V)$  关于  $V$  的某个基有上三角矩阵。那么  $T$  的特征值恰为该上三角矩阵对角线上的各元素

#### 证

设  $\nu_1, \dots, \nu_n$  是  $V$  的一个基，且  $T$  关于该基有上三角矩阵

$$M(T) = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

因为  $T\nu_1 = \lambda_1\nu_1$ ，我们可见  $\lambda_1$  是  $T$  的一个特征值

设  $k \in \{2, \dots, n\}$ 。那么  $(T - \lambda_k I)\nu_k \in \text{span}(\nu_1, \dots, \nu_{k-1})$ 。于是  $T - \lambda_k I$  将  $\text{span}(\nu_1, \dots, \nu_k)$  映射至  $\text{span}(\nu_1, \dots, \nu_{k-1})$ 。又因为

$$\dim \text{span}(\nu_1, \dots, \nu_{k-1}) = k - 1$$

所以限制于  $\text{span}(\nu_1, \dots, \nu_k)$  的算子  $T - \lambda_k I$  不是单射（由3.22）。于是，存在  $\nu \in \text{span}(\nu_1, \dots, \nu_k)$  满足  $\nu \neq 0$  且  $(T - \lambda_k I)\nu = 0$ 。于是  $\lambda_k$  是  $T$  的特征值。因此我们证明了  $\nabla M(T)$  对角线上各元素均为  $T$  的特征值

为证明  $T$  无其他特征值，令  $q$  是定义为  $q(z) = (z - \lambda_1) \cdots (z - \lambda_n)$  的多项式。那么  $q(T) = 0$ （由5.40）。因此  $q$  是  $T$  的最小多项式的多项式倍

（由5.29）。于是  $T$  的最小多项式的每个零点都是  $q$  的零点。因为  $T$  的最小多项式的零点即  $T$  的特征值（由5.27），所以可推出  $T$  的每个特征值都是  $q$  的零点。因此， $T$  的特征值都包含在  $\lambda_1, \dots, \lambda_n$  这组数中

### 5.42 由上三角矩阵得特征值

定义  $T \in \mathcal{L}(\mathbf{F}^3)$  为  $T(x, y, z) = (2x + y, 5y + 3z, 8z)$ .  $T$  关于标准基的矩阵为

$$M(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}$$

由 5.41 可得  $T$  的特征值为 2、5 和 8

### 5.43 $T$ 是否有上三角矩阵可能取决于 $\mathbf{F}$

下面例子形象阐释了 5.44: 一算子关于某个基有上三角矩阵, 当且仅当该算子的最小多项式是一次多项式的乘积

$T \in \mathcal{L}(\mathbf{F}^4)$

$$T(z_1, z_2, z_3, z_4) = (-z_2, z_1, 2z_1 + 3z_3, z_3 + 3z_4)$$

于是  $T$  关于  $\mathbf{F}^4$  的标准基的矩阵是

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

$T$  的最小多项式是定义如下多项式

$$p(z) = 9 - 6z + 10z^2 - 6z^3 + z^4$$

先考虑  $\mathbf{F} = \mathbf{R}$  的情况. 那么多项式  $p$  可因式分解成

$$p(z) = (z^2 + 1)(z - 3)(z - 3)$$

其中  $z^2 + 1$  无法被进一步分解成两个一次实系数多项式的乘积. 于是 5.44 表明, 不存在  $\mathbf{R}^4$  的一个基, 使得  $T$  关于该基有上三角矩阵  
现在考虑  $\mathbf{F} = \mathbf{C}$  的情况. 那么多项式  $p$  可因式分解成

$$p(z) = (z - i)(z + i)(z - 3)(z - 3)$$

其中各因子都具有  $z - \lambda_k$  的形式. 于是 5.44 表明, 存在  $\mathbf{C}^4$  的一个基, 使得  $T$  关于该基有上三角矩阵. 的确如此, 你可以验证一下, 算子  $T$  关于  $\mathbf{C}^4$  的基  $(4 - 3i, -3 - 4i, -3 + i, 1), (4 + 3i, -3 + 4i, -3 - i, 1), (0, 0, 0, 1), (0, 0, 1, 0)$  有上三角矩阵

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

### 5.44 存在上三角矩阵的充要条件

$V$  是有限维的,  $T \in \mathcal{L}(V)$ . 那么  $T$  关于  $V$  的某个基有上三角矩阵, 当且仅当  $T$  的最小多项式等于  $(z - \lambda_1) \cdots (z - \lambda_m)$  (其中  $\lambda_1, \dots, \lambda_m \in \mathbf{F}$ )

#### 证

先设  $T$  关于  $V$  的某个基具有上三角矩阵. 令  $\alpha_1, \dots, \alpha_n$  代表该矩阵对角线上各元素定义多项式  $q \in \mathcal{P}(\mathbf{F})$

$$q(z) = (z - \alpha_1) \cdots (z - \alpha_n)$$

那么由 5.40,  $q(T) = 0$ . 因此由 5.29 得,  $q$  是  $T$  的最小多项式的多项式倍. 于是,  $T$  的最小多项式等于  $(z - \lambda_1) \cdots (z - \lambda_m)$ , 其中  $\lambda_1, \dots, \lambda_m \in \mathbf{F}$  且  $\{\lambda_1, \dots, \lambda_m\} \subseteq \{\alpha_1, \dots, \alpha_n\}$ .

为了证明另一方向的蕴涵关系, 现在设  $T$  的最小多项式等于  $(z - \lambda_1) \cdots (z - \lambda_m)$  (其中  $\lambda_1, \dots, \lambda_m \in \mathbf{F}$ ). 我们对  $m$  用归纳法. 首先, 若  $m = 1$ , 则  $z - \lambda_1$  是  $T$  的最小多项式, 由此可得  $T = \lambda_1 I$ , 进而可推出  $T$  (关于  $V$  的任何基) 的矩阵都是上三角的. 现设  $m > 1$  且欲证结论当  $\nabla_m$  为更小的正整数值时都成立. 令

$$U = \text{range}(T - \lambda_m I)$$

那么  $U$  在  $T$  下是不变的【这是 5.18 在  $p(z) = z - \lambda_m$  时的特殊情况】，那么  $T|_U$  就是  $U$  上的算子.

若  $u \in U$ , 那么对某  $\nu \in V$  有  $= (T - \lambda_m I)\nu$ , 并且

$$(T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (T - \lambda_1 I) \cdots (T - \lambda_m I)\nu = 0$$

因此由 5.29,  $(z - \lambda_1) \cdots (z - \lambda_{m-1})$  就是  $T|_U$  的最小多项式的多项式倍. 于是  $T|_U$  的最小多项式就是至多  $m-1$  个形如  $z - \lambda_k$  的项的乘积  
由归纳假设, 存在  $U$  的基  $u_1, \dots, u_M$ , 使得  $T|_U$  关于该基有上三角矩阵. 于是对于每个  $k \in \{1, \dots, M\}$ , 利用 5.39

$$Tu_k = (T|_U)(u_k) \in \text{span}(u_1, \dots, u_k)$$

将  $u_1, \dots, u_M$  扩充为  $V$  的一个基  $u_1, \dots, u_M, v_1, \dots, v_n$ . 对每个  $k \in \{1, \dots, n\}$  有

$$Tv_k = (T - \lambda_m I)v_k + \lambda_m v_k$$

由  $U$  的定义可知,  $(T - \lambda_m I)v_k \in U = \text{span}(u_1, \dots, u_M)$ . 于是上式表明

$$Tv_k \in \text{span}(u_1, \dots, u_M, v_1, \dots, v_k).$$

由式 (5.45) 和式 (5.46), 并利用 5.39, 我们可得出结论:  $T$  关于  $V$  的基  $u_1, \dots, u_M, v_1, \dots, v_n$  有上三角矩阵, 命题得证

上个结果中得到的数集  $\{\lambda_1, \dots, \lambda_m\}$  就等于  $T$  的特征值所成的集合 (因为, 由 5.27 可知  $T$  的最小多项式的零点构成的集合就等于  $T$  的特征值构成的集合), 尽管  $\lambda_1, \dots, \lambda_m$  这组数中可能有重复的数

下面结论是个很好的结论, 但即便如此, 在第 8 章中, 我们还可将其改进得更好, 参看 8.37 和 8.46

### 5.47 有限维复向量空间上的每个算子都有上三角矩阵

设  $V$  是有限维复向量空间且  $T \in \mathcal{L}(V)$ . 那么  $T$  关于  $V$  的某个基具有上三角矩阵

**证**

由 5.44 和代数基本定理的版本二 (4.13) 可得

对于满足  $ST = TS$  的两个算子  $s$  和  $T$ , 上述结论可进一步拓展, 参见 5.80. 另外, 上述结论还可拓展到多于两个算子的情形, 可参看 5E 节的习题 9 (b)  
注意若算子  $T \in \mathcal{L}(V)$  关于  $V$  的某个基  $v_1, \dots, v_n$  有上三角矩阵, 那么如 5.41 所示,  $T$  的特征值恰为  $\nabla \mathcal{M}(T)$  对角线上的元素, 此外,  $v_1$  是  $T$  的一个特征向量. 然而,  $v_2, \dots, v_n$  并不一定是  $T$  的特征向量. 事实上, 基向量  $v_k$  是  $T$  的特征向量, 当且仅当  $T$  的矩阵的第  $k$  列中除了第  $k$  个元素可能非零外, 其余各元素都是 0

回忆一下, 每个矩阵都可以被化成行阶梯形矩阵. 如果是对方阵作变换, 那么所得的行阶梯形矩阵是个上三角矩阵. 但不要把这个上三角矩阵和算子关于某基的上三角矩阵 (由 5.47, 若  $F = C$  则其一定存在) 混淆起来. 这两个上三角矩阵之间是没有联系的

由算子的行阶梯形矩阵并不能得到该算子的特征值. 与之不同的是, 算子关于某个基的上三角矩阵则可以给出该算子的所有特征值. 然而, 这样的上三角矩阵是无法确切计算得到的, 即便 5.47 确保了  $F = C$  时它一定存在

## 5D 可对角化算子

### 对角矩阵

#### 5.48 对角矩阵

对角矩阵是对角线之外元素均为 0 的方阵

#### 5.49 对角矩阵

$$\begin{pmatrix} 8 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

是对角矩阵

若一算子关于某基具有对角矩阵, 那么该矩阵对角线上的元素就是该算子的特征值由 5.41 即可证明这个结论 (也可专为对角矩阵构造个更简单直接的证法)

每个对角矩阵都是上三角的. 一般而言, 对角矩阵比大多数相同大小的上三角矩阵含有更多的 0

## 5.50 可对角化

若  $V$  上的一个算子关于  $V$  的某个基具有对角矩阵，则称该算子是可对角化的

## 5.51 对角化可能需要不同的基

定义  $T \in \mathcal{L}(\mathbf{R}^2)$  为

$$T(x, y) = (41x + 7y, -20x + 74y)$$

$T$  关于  $\mathbf{R}^2$  的标准基的矩阵

$$\begin{pmatrix} 41 & 7 \\ -20 & 74 \end{pmatrix}$$

这不是对角矩阵。然而， $T$  是可对角化的。具体而言， $T$  关于基  $(1, 4), 7, 5$  的矩阵是

$$\begin{pmatrix} 69 & 0 \\ 0 & 46 \end{pmatrix}$$

因为  $T(1, 4) = (69, 276) = 69(1, 4)$ ,  $T(7, 5) = (322, 230) = 46(7, 5)$

下面，我们给被算子  $T$  映射至其  $\lambda$  ( $\lambda \in \mathbf{F}$ ) 倍的向量所构成的集合起个名字，并用个记号来表示它，这将便于我们进行讨论

## 5.52 特征空间

$T \in \mathcal{L}(V)$  且  $\lambda \in \mathbf{F}$ 。 $T$  对应于  $\lambda$  的特征空间记作  $(\lambda, T)$ ，是定义如下的  $V$  的子空间

$$(\lambda, T) = \text{null}(T - \lambda I) = \{\nu \in V : T\nu = \lambda\nu\}$$

因此  $(\lambda, T)$  是  $T$  对应于  $\lambda$  的所有特征向量以及 0 向量所构成的集合

## 5.53 一算子的特征空间

对  $T \in \mathcal{L}(V)$  且  $\lambda \in \mathbf{F}$ ，集合  $(\lambda, T)$  是  $V$  的子空间，因为  $V$  上任何线性映射的零空间都是  $V$  的子空间。由定义可得， $\lambda$  是  $T$  的特征值当且仅当  $(\lambda, T) \neq \{0\}$

设算子  $T \in \mathcal{L}(V)$  关于  $V$  的基  $\nu_1, \nu_2, \nu_3$  的矩阵是例 5.49 中的矩阵。那么

$$(8, T) = \text{span}(\nu_1), \quad (5, T) = \text{span}(\nu_2, \nu_3)$$

若  $\lambda$  是算子  $T \in \mathcal{L}(V)$  的特征值，那么限制于  $(\lambda, T)$  的  $T$  就是将向量乘以  $\lambda$  的算子

## 5.54 特征空间之和是直和

设  $T \in \mathcal{L}(V)$  且  $\lambda_1, \dots, \lambda_m$  是  $T$  的互异特征值。那么

$$(\lambda_1, T) + \dots + (\lambda_m, T)$$

是直和。此外，若  $V$  是有限维的，那么

$$\dim(\lambda_1, T) + \dots + \dim(\lambda_m, T) \leq \dim V$$

## 证

为证明  $(\lambda_1, T) + \dots + (\lambda_m, T)$  是直和，设

$$\nu_1 + \dots + \nu_m = 0$$

其中各  $\nu_k$  分别属于  $(\lambda_k, T)$ . 因为对应于互异特征值的特征向量是线性无关的 (由5.11), 故可得各  $\nu_k$  都等于0. 因此,  $(\lambda_1, T) + \cdots + (\lambda_m, T)$  是直和 (由1.45)  
现设  $V$  是有限维的. 那么

$$\dim(\lambda_1, T) + \cdots + \dim(\lambda_m, T) = \dim((\lambda_1, T) \oplus \cdots \oplus (\lambda_m, T)) \leq \dim V$$

其中第一行源于3.94 而第二行源于2.37

## 可对角化的条件

对可对角化算子的如下刻画很有用处

### 5.55 可对角化的等价条件

设  $V$  是有限维的,  $T \in \mathcal{L}(V)$ . 令  $\lambda_1, \dots, \lambda_m$  表示  $T$  的所有互异特征值. 则下列命题等价(a)  $T$  是可对角化的(b)  $V$  有由  $T$  的特征向量构成的基(c)  $V = (\lambda_1, T) \oplus \cdots \oplus (\lambda_m, T)$ . (d)  $\dim V = \dim(\lambda_1, T) + \cdots + \dim(\lambda_m, T)$

### 证

算子  $T \in \mathcal{L}(V)$  关于  $V$  的某个基  $\nu_1, \dots, \nu_n$  具有对角矩阵

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

当且仅当对每个  $k$  均有  $T\nu_k = \lambda_k\nu_k$ . 于是 (a) 与 (b) 等价

设 (b) 成立, 于是  $V$  有由  $T$  的特征向量组成的基. 因此  $V$  中每个向量都是  $T$  的特征向量的线性组合, 这意味着

$$V = (\lambda_1, T) + \cdots + (\lambda_m, T).$$

将此式结合 5.54 可知(c) 成立, 也就证明了(b) 蕴涵(c)

由 3.94 立得 (c) 蕴涵 (d)

设(d) 成立, 于是

$$\dim V = \dim(\lambda_1, T) + \cdots + \dim(\lambda_m, T)$$

为各  $(\lambda_k, T)$  选取一个基, 再将这些基合在一起, 就形成由  $T$  的特征向量构成的组  $\nu_1, \dots, \nu_n$ , 其中  $n = \dim V$  【由式(5.56)】. 为证明这个组是线性无关的, 设

$$a_1\nu_1 + \cdots + a_n\nu_n = 0$$

其中  $a_1, \dots, a_n \in \mathbf{F}$ . 对每个  $k = 1, \dots, m$ , 令  $u_k$  表示所有  $a_j\nu_j$  之和, 其中各  $\nu_j \in (\lambda_k, T)$ . 于是各  $u_k$  分别属于  $(\lambda_k, T)$ , 且

$$u_1 + \cdots + u_m = 0$$

因为对应于互异特征值的特征向量是线性无关的 (见5.11), 故可得各  $u_k$  都等于0. 因为各  $u_k$  是由若干  $a_j\nu_j$  求和得到, 其中这些  $\nu_j$  组成  $(\lambda_k, T)$  的基, 故可知各  $a_j$  都等于0. 于是  $\nu_1, \dots, \nu_n$  是线性无关的, 因此是  $V$  的一个基 (由 2.38). 所以(d) 蕴涵(b)

与可对角化等价的其他条件见于 5.62, 本节习题 5 和习题 15, 7B 节的习题 24, 以及 8A 节的习题 15

我们知道, 非零有限维复向量空间上的每个算子都有特征值. 然而, 非零有限维复向量空间上, 不是每个算子都有足够多的特征向量使其可对角化, 如下例所示

### 5.57 不可对角化的算子

$T \in \mathcal{L}(\mathbf{F}^3)$  为  $T(a, b, c) = (b, c, 0)$ , 其关于  $\mathbf{F}^3$  的标准基的矩阵为

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

它是上三角矩阵, 但不是对角矩阵

你应自行验证, 0 是  $T$  的唯一特征值, 并且

$$(0, T) = \{(a, 0, 0) \in \mathbf{F}^3 : a \in \mathbf{F}\}$$

因此 5.55 中 (b)、(c) 和 (d) 这三条都不成立 (当然, 由于这几条是等价的, 所以只要检验其中一条不成立就够了). 于是 5.55 中的 (a) 这条也不成立. 因此不管  $\mathbf{F} = \mathbf{R}$  还是  $\mathbf{F} = \mathbf{C}$ ,  $T$  都不可对角化

下面结论表明, 如果一个算子的互异特征值数目与其定义空间的维数一样, 那么该算子就是可对角化的

### 5.58 特征值足够多意味着可对角化

设  $V$  是有限维的且  $T \in \mathcal{L}(V)$  有  $\dim V$  个互不相同的特征值. 那么  $T$  是可对角化的

#### 证

设  $T$  有互不相同的特征值  $\lambda_1, \dots, \lambda_{\dim V}$ . 对每个  $k$ , 令  $v_k \in V$  是对应于特征值  $\lambda_k$  的一个特征向量. 因为对应于不同特征值的特征向量线性无关 (见 5.11), 所以  $v_1, \dots, v_{\dim V}$  是线性无关的

$V$  中包含  $\dim V$  个向量的线性无关组是  $V$  的基 (见 2.38). 由此,  $v_1, \dots, v_{\dim V}$  是  $V$  的基关于这个由特征向量构成的基,  $T$  具有对角矩阵

在后面的章节中我们还会得到更多能推出某个算子可对角化的条件, 例如实谱定理 (7.29) 和复谱定理 (7.31)

上述结论给出了算子可对角化的充分条件. 然而, 这个条件不是必要的. 例如, 定义为  $T(x, y, z) = (6x, 6y, 7z)$  的  $\mathbf{F}^3$  上的算子  $T$  只有两个特征值 (6 和 7) 而  $\dim \mathbf{F}^3 = 3$ , 但  $T$  也是可对角化的 (用  $\mathbf{F}^3$  的标准基即可)

下例展示了对角化的重要作用, 即可以借助它来计算算子的高次幂——利用式  $T^k v = \lambda^k v$  ( $v$  是  $T$  对应于特征值  $\lambda$  的特征向量) 即可. 该技巧的一个更精彩的应用示于习题 21 中, 该题展示了如何利用对角化来得出斐波那契数列的第  $n$  项的确切表达式



### 5.59 利用对角化来计算 $T^{100}$

$T \in \mathcal{L}(\mathbf{F}^3)$ ,  $T(x, y, z) = (2x + y, 5y + 3z, 8z)$ ,  $T$  关于标准基的矩阵为

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}$$

上面矩阵是上三角矩阵, 但不是对角矩阵. 由 5.4-1 可知  $T$  的特征值是 2、5 和 8. 因为  $T$  是三维向量空间上的算子且  $T$  有三个互不相同的特征值, 5.58 就确保存在  $\mathbf{F}^3$  的一个基使得  $T$  关于该基具有对角矩阵

为了求出这个基, 我们只需求出对应于每个特征值的特征向量即可. 换言之, 我们需要求出方程

$$T(x, y, z) = \lambda(x, y, z)$$

在  $\lambda = 2$ 、 $\lambda = 5$ 、 $\lambda = 8$  这三种情形下的非零解. 求解这些简单的方程即得对应于  $\lambda = 2$  的特征向量  $(1, 0, 0)$ 、对应于  $\lambda = 5$  的特征向量  $(1, 3, 0)$  和对应于  $\lambda = 8$  的特征向量  $(1, 6, 6)$

于是,  $(1, 0, 0), (1, 3, 0), (1, 6, 6)$  是由  $T$  的特征向量构成的  $\mathbf{F}^3$  的基, 且  $T$  关于这个基的矩阵即为对角矩阵

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 8 \end{pmatrix}.$$

下面以计算  $T^{100}(0, 0, 1)$  为例说明如何计算  $T^{100}$ . 我们先把  $(0, 0, 1)$  用特征向量构成的基的线性组合表示出来

$$(0, 0, 1) = \frac{1}{6}(1, 0, 0) - \frac{1}{3}(1, 3, 0) + \frac{1}{6}(1, 6, 6)$$

再将  $T^{100}$  作用于上式两端, 可得

$$\begin{aligned} T^{100}(0, 0, 1) &= \frac{1}{6}(T^{100}(1, 0, 0)) - \frac{1}{3}(T^{100}(1, 3, 0)) + \frac{1}{6}(T^{100}(1, 6, 6)) \\ &= \frac{1}{6}(2^{100}(1, 0, 0) - 2 \cdot 5^{100}(1, 3, 0) + 8^{100}(1, 6, 6)) \\ &= \frac{1}{6}(2^{100} - 2 \cdot 5^{100} + 8^{100}, 6 \cdot 8^{100} - 6 \cdot 5^{100}, 6 \cdot 8^{100}). \end{aligned}$$

之前我们已经看到，有限维向量空间  $V$  上的算子  $T$  关于  $V$  的某个基有上三角矩阵，当且仅当  $T$  的最小多项式等于  $(z - \lambda_1) \cdots (z - \lambda_m)$ ，其中  $\lambda_1, \dots, \lambda_m \in \mathbf{F}$ （见 5.44）。前面也提到过（见 5.47），这个结论当  $\mathbf{F} = \mathbf{C}$  时总是成立的

接下来的结果 5.62 的内容是，算子  $T \in \mathcal{L}(V)$  关于  $V$  的某个基有对角矩阵，当且仅当  $T$  的最小多项式等于  $(z - \lambda_1) \cdots (z - \lambda_m)$ ，其中  $\lambda_1, \dots, \lambda_m \in \mathbf{F}$  且互不相同。在正式地陈述这个结论之前，我们先给出两个应用它的实例

### 5.60 可对角化，但无法求出确切特征值

定义  $T \in \mathcal{L}(\mathbf{C}^5)$  为

$$T(z_1, z_2, z_3, z_4, z_5) = (-3z_5, z_1 + 6z_5, z_2, z_3, z_4).$$

$T$  的矩阵如例 5.26 所示，在此例中，我们求得  $T$  的最小多项式为  $3 - 6z + z^5$

在例 5.28 中曾提及，无法求出上述多项式的零点的确切表达式，但是用数值计算方法可得该多项式的零点近似为  $-1.67, 0.51, 1.40, -0.12 + 1.59i, -0.12 - 1.59i$ 。

算出这些近似值的软件可以给出不止三位数字的精确值。于是，上述近似结果足以显示五个特征值是互异的。而  $T$  的最小多项式就是将以上各数作为零点的五次首一多项式。则由 5.62 可知  $T$  是可对角化的

### 5.61 说明一算子不可对角化

定义  $T \in \mathcal{L}(\mathbf{F}^3)$  为

$$T(z_1, z_2, z_3) = (6z_1 + 3z_2 + 4z_3, 6z_2 + 2z_3, 7z_3).$$

$T$  关于  $\mathbf{F}^3$  的标准基的矩阵是

$$\begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix}.$$

上述矩阵是上三角矩阵，但不是对角矩阵。 $T$  关于  $\mathbf{F}^3$  的其他基是否可能具有对角矩阵呢？

为了回答这个问题，我们需要求出  $T$  的最小多项式。首先，注意到  $T$  的特征值是上述矩阵对角线上的项（由 5.41）。于是  $T$  的最小多项式的零点是 6 和 7【由 5.27 (a)】。由上述矩阵的对角线又可得  $(T - 6I)^2(T - 7I) = 0$ （由 5.40）。 $T$  的最小多项式的次数最高为 3（由 5.22）。综合所有这些结论，我们即得  $T$  的最小多项式要么是  $(z - 6)(z - 7)$ ，要么是  $(z - 6)^2(z - 7)$ 。

简单计算即知  $(T - 6I)(T - 7I) \neq 0$ 。因而  $T$  的最小多项式是  $(z - 6)^2(z - 7)$

则由 5.62 可知  $T$  不可对角化

### 5.62 可对角化的充要条件

设  $V$  是有限维的，且  $T \in \mathcal{L}(V)$ 。那么  $T$  是可对角化的，当且仅当  $T$  的最小多项式等于  $(z - \lambda_1) \cdots (z - \lambda_m)$ ，其中  $\lambda_1, \dots, \lambda_m \in \mathbf{F}$  为互不相同的数

#### 证

先设  $T$  是可对角化的。那么存在由  $T$  的特征向量构成的  $V$  的基  $\nu_1, \dots, \nu_n$ 。令  $\lambda_1, \dots, \lambda_m$  是  $T$  的所有互异特征值。那么对每个  $\nu_j$ ，存在  $\lambda_k$  使得  $(T - \lambda_k I)\nu_j = 0$ 。于是

$$(T - \lambda_1 I) \cdots (T - \lambda_m I)\nu_j = 0,$$

意味着  $T$  的最小多项式等于  $(z - \lambda_1) \cdots (z - \lambda_m)$

为了证明另一方向的蕴涵关系，现设  $T$  的最小多项式等于  $(z - \lambda_1) \cdots (z - \lambda_m)$ ，其中  $\lambda_1, \dots, \lambda_m \in \mathbf{F}$  为互不相同的数。于是

$$(T - \lambda_1 I) \cdots (T - \lambda_m I) = 0.$$

我们将通过对  $m$  用归纳法来证明  $T$  是可对角化的。首先设  $m = 1$ 。那么  $T - \lambda_1 I = 0$ ，这意味着  $T$  是恒等算子的标量倍，也就说明了  $T$  是可对角化的。

现在设  $m > 1$  且欲证结论在  $\nabla m$  为更小的值时都成立。 $\text{range}(T - \lambda_m I)$  这个子空间在  $T$  下是不变的【这是 5.18 在  $p(z) = z - \lambda_m$  时的特殊情况】。那么限制在  $\text{range}(T - \lambda_m I)$  上的  $T$  就是  $\text{range}(T - \lambda_m I)$  上的算子

若  $u \in \text{range}(T - \lambda_m I)$ ，那么对某个  $\nu \in V$  有  $= (T - \lambda_m I)\nu$ ，结合式 (5.63) 可得

$$(T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (T - \lambda_1 I) \cdots (T - \lambda_m I)\nu = 0.$$

因此， $(z - \lambda_1) \cdots (z - \lambda_{m-1})$  是限制于  $\text{range}(T - \lambda_m I)$  上的  $T$  的最小多项式的多项式倍（由 5.29）。于是，由归纳假设， $\text{range}(T - \lambda_m I)$  中存在由  $T$  的特征向量组成的基。

设  $u \in \text{range}(T - \lambda_m I) \cap \text{null}(T - \lambda_m I)$ 。那么  $Tu = \lambda_m u$ 。结合式 (5.64) 可得

$$\begin{aligned} 0 &= (T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u \\ &\quad 0 \\ &= (\lambda_m - \lambda_1) \cdots (\lambda_m - \lambda_{m-1})u. \end{aligned}$$

因为  $\lambda_1, \dots, \lambda_m$  是互异的，所以上式表明  $u = 0$ 。因此  $\text{range}(T - \lambda_m I) \cap \text{null}(T - \lambda_m I) = \{0\}$ 。

于是  $\text{range}(T - \lambda_m I) + \text{null}(T - \lambda_m I)$  是直和（由 1.46），且其维数是  $\dim V$ （由 3.94 和 3.21）。因此  $\text{range}(T - \lambda_m I) \oplus \text{null}(T - \lambda_m I) = V$ 。  
 $\text{null}(T - \lambda_m I)$  中的每个非零向量都是  $T$  对应于特征值  $\lambda_m$  的特征向量。在该证明的靠前部分，我们已证明  $\text{range}(T - \lambda_m I)$  存在由  $T$  的特征向量组成的基。将该基与  $\text{null}(T - \lambda_m I)$  的基合并，就得到了由  $T$  的特征向量组成的  $V$  的基。 $T$  关于这个基的矩阵就是对角矩阵，命题得证。

5 次以上多项式没有求根公式。然而，利用上述结论，连最小多项式零点的近似值都不用求，就能判断复向量空间上一算子是否可对角化，详见习题 15。

下面这条结论将是我们证明两算子的同时对角化时所运用的关键工具，详见 5.76。由 5.62，可从最小多项式角度刻画可对角化算子，请留意我们如何利用这一点来引出下面这条结论的简短证明。

### 5.65 将可对角化算子限制于不变子空间

设  $T \in \mathcal{L}(V)$  是可对角化的， $U$  是  $V$  的子空间且在  $T$  下不变。那么  $T|_U$  是  $U$  上的可对角化算子。

### 证

因为算子  $T$  是可对角化的，所以  $T$  的最小多项式等于  $(z - \lambda_1) \cdots (z - \lambda_m)$ ，其中  $\lambda_1, \dots, \lambda_m \in \mathbf{F}$  为互不相同的数（由 5.62）。 $T$  的最小多项式是  $T|_U$  的最小多项式的多项式倍（由 5.31）。因此  $T|_U$  的最小多项式具有 5.62 所需的形式，由此可说明  $T|_U$  是可对角化的。

## 格什戈林圆盘定理

### 5.66 格什戈林圆盘

设  $T \in \mathcal{L}(V)$  且  $\nu_1, \dots, \nu_n$  是  $V$  的一个基。令  $A$  表示  $T$  关于该基的矩阵。 $T$  关于基  $\nu_1, \dots, \nu_n$  的格什戈林圆盘是形如

$$\left\{ z \in \mathbf{F} : |z - A_{j,j}| \leq \sum_{\substack{k=1 \\ k \neq j}}^n |A_{j,k}| \right\}$$

的集合，其中  $j \in \{1, \dots, n\}$

因为上面定义中的  $j$  有  $n$  种取值，所以  $T$  有  $n$  个格什戈林圆盘。若  $\mathbf{F} = \mathbf{C}$ ，那么对每个  $j \in \{1, \dots, n\}$ ，与之对应的格什戈林圆盘就是 中以  $A_{j,j}$  为圆心的闭圆盘，其中  $A_{j,j}$  是  $A$  的对角线上的第  $j$  个元素。这个闭圆盘的半径，等于  $A$  的第  $j$  行除对角线上的元素外各元素绝对值之和。若  $\mathbf{F} = \mathbf{R}$ ，那么格什戈林圆盘是  $\mathbf{R}$  中的闭区间。

在上述方阵  $A$  为对角阵的特殊情况下，每个格什戈林圆盘都仅包含一个点，即  $A$  的对角线上的元素（进而  $T$  的每个特征值都是这样的点，这也是下面结论所保证的）。下面的结论有个推论：如果  $A$  的非对角线元素很小，那么  $T$  的每个特征值都与  $A$  的对角线上的一个元素很接近。

### 5.67 格什戈林圆盘定理

设  $T \in \mathcal{L}(V)$  且  $\nu_1, \dots, \nu_n$  是  $V$  的一个基。那么  $T$  的每个特征值都被包含在  $T$  关于基  $\nu_1, \dots, \nu_n$  的某个格什戈林圆盘中。

#### 证

设  $\lambda \in \mathbf{F}$  是  $T$  的一个特征值。令  $w \in V$  是与之对应的特征向量。存在  $c_1, \dots, c_n \in \mathbf{F}$ ，使得

$$w = c_1\nu_1 + \dots + c_n\nu_n$$

令  $A$  表示  $T$  关于基  $\nu_1, \dots, \nu_n$  的矩阵。将  $T$  作用于上式两侧，可得

$$\begin{aligned}\lambda w &= \sum_{k=1}^n c_k T \nu_k \\ &= \sum_{k=1}^n c_k \sum_{j=1}^n A_{j,k} \nu_j \\ &= \sum_{j=1}^n \left( \sum_{k=1}^n A_{j,k} c_k \right) \nu_j\end{aligned}$$

令  $j \in \{1, \dots, n\}$  使得

$$|c_j| = \max\{|c_1|, \dots, |c_n|\}$$

利用式(5.68)，我们发现，在式(5.69)左侧，各  $|\nu_j$  的系数等于  $\lambda c_j$ ，而这些系数必须与式(5.70)右侧中各  $\nu_j$  的系数相等。换言之

$$\lambda c_j = \sum_{k=1}^n A_{j,k} c_k.$$

同减  $A_{j,j} c_j$  同除  $c_j$

$$\begin{aligned}|\lambda - A_{j,j}| &= \sum_{\substack{k=1 \\ k \neq j}}^n A_{j,k} \frac{c_k}{c_j} \\ &\leq \sum_{\substack{k=1 \\ k \neq j}}^n A_{j,k}\end{aligned}$$

于是  $\lambda$  处在关于基  $\nu_1, \dots, \nu_n$  的第  $j$  个格什戈林圆盘中。

习题22给出了格什戈林圆盘定理的一个很漂亮的应用实例，格什戈林圆盘定理是用谢苗·阿罗诺维奇·格什戈林 (Semyon Aronovich Gershgorin) 的名字命名的，他于1931年发表了该结论。

习题23提出可将每个格什戈林圆盘的半径改为等于其对应列（而不是行）中除对角线元素外各元素之和，而格什戈林圆盘定理仍然成立。

## 5E 可交换算子

### 5.71 可交换

对同一向量空间上的两个算子  $S$  和  $T$ ，若  $ST = TS$ ，则它们可交换；对于两个大小相同的方阵  $A$  和  $B$ ，若  $AB = BA$ ，则它们可交换。

#### 例

如  $T$  是算子,  $p, q \in \mathcal{P}(\mathbf{F})$ ,  $p(T)$  和  $q(T)$  可交换【见 5.17 (b)】

如  $I$  是  $V$  上的恒等算子,  $I$  与  $V$  上每个算子都可交换

### 5.72 偏微分算子可交换

设  $m$  是非负整数. 令  $\mathcal{P}_m(\mathbf{R}^2)$  表示具有两个实自变量且次数最高为  $m$  的实系数多项式构成的实向量空间, 其带有实值函数的一般加法和标量乘法运算. 于是  $\mathcal{P}_m(\mathbf{R}^2)$  中的元素, 是  $\mathbf{R}^2$  上如下函数

$$p = \sum_{j+k \leq m} a_{j,k} x^j y^k$$

其中下标  $j$  和  $k$  可取遍每个满足  $j + k \leq m$  的非负整数值, 每个  $a_{j,k}$  都属于  $\mathbf{R}$ ,  $x^j y^k$  表示定义为  $(x, y) \mapsto x^j y^k$  的  $\mathbf{R}^2$  上的函数. 定义算子  $D_x, D_y \in \mathcal{L}(\mathcal{P}_m(\mathbf{R}^2))$  为

$$D_x p = \frac{p}{x} = \sum_{j+k \leq m} j a_{j,k} x^{j-1} y^k$$

$$D_y p = \frac{p}{y} = \sum_{j+k \leq m} k a_{j,k} x^j y^{k-1}$$

其中  $p$  定义如式(5.73). 因为算子  $D_x$  和  $D_y$  都只对两个变量之一求微分, 而将另一个变量当作常数, 所以它们得名为偏微分算子. 算子  $D_x$  和  $D_y$  可交换, 因为若  $p$  定义如式 (5.73), 则

$$(D_x D_y)p = \sum_{j+k \leq m} j k a_{j,k} x^{j-1} y^{k-1} = (D_y D_x)p$$

等式  $D_x D_y = D_y D_x$  在  $\mathcal{P}_m(\mathbf{R}^2)$  上成立, 还说明了另一个更为一般的结论, 那就是对于性质良好的函数, 偏微分运算的顺序是无关紧要的.

#### 注

可交换矩阵并不很常见. 例如, 各元素均为区间  $[5, 5]$  内整数的  $2 \times 2$  矩阵, 两两共可凑出 214,358,881 对 (考虑顺序), 但如此多对矩阵中仅有约 0.3% 是可交换的.

用计算机检验得, 讨论范围内的 214,358,881 (等于  $11^8$ ) 对  $2 \times 2$  矩阵中, 仅有 674,609 对矩阵是可交换的.

下结论表明两算子可交换当且仅当它们关于同一基的矩阵可交换

### 5.74 可交换算子对应可交换矩阵

$S, T \in \mathcal{L}(V)$  且  $\nu_1, \dots, \nu_n$  是  $V$  的一个基. 那么  $S$  和  $T$  可交换, 当且仅当  $M(S, (\nu_1, \dots, \nu_n))$  和  $M(T, (\nu_1, \dots, \nu_n))$  可交换

#### 证

$$\iff M(S)M(T) = M(T)M(S)$$

接下来这条结果表明, 若两个算子可交换, 那么其中一个算子的每个特征空间必在另一算子下不变. 一对可交换算子比起一对不可交换的算子具有更好的性质, 其主要原因之一就在于此, 我们也会多次用到这一结果

### 5.75 特征空间在可交换算子下不变

$S, T \in \mathcal{L}(V)$  可交换且  $\lambda \in \mathbf{F}$ . 那么  $(\lambda, S)$  在  $T$  下不变

#### 证

设  $\nu \in (\lambda, S)$

$$S(T\nu) = (ST)\nu = (TS)\nu = T(S\nu) = T(\lambda\nu) = \lambda T\nu$$

故  $T\nu \in (\lambda, S)$ ,  $(\lambda, S)$  在  $T$  下不变

设我们有两个算子并且它们都可对角化. 如果我们想进行的计算同时涉及它们两者 (例如, 涉及它们的和), 那么我们就希望这两个算子可关于相同的基作对角化. 根据下面的结论, 这种操作在这两个算子可交换时是可行的

### 5.76 可同时对角化 可交换性

同一向量空间上的两个可对角化算子关于相同的基都有对角矩阵, 当且仅当这两个算子可交换

#### 证

先设  $S, T \in \mathcal{L}(V)$  关于同一个基有对角矩阵. 两个大小相同的对角矩阵的乘积, 等于将这两个矩阵对角线上的元素对应相乘所得的对角矩阵. 因此任意两个大小相同的对角矩阵都可交换. 于是  $s$  和  $T$  可交换 (由5.74)

为了证明另一方向, 现设  $S, T \in \mathcal{L}(V)$  是可对角化算子且可交换. 令  $\lambda_1, \dots, \lambda_m$  代表  $s$  的所有互异特征值. 因为  $s$  可对角化, 所以5.55 (c) 就表明

$$V = (\lambda_1, S) \oplus \dots \oplus (\lambda_m, S)$$

子空间  $(\lambda_k, S)$  ( $k = 1, \dots, m$ ) 在  $T$  下不变 (由5.75). 因为  $T$  是可对角化的, 所以5.65表明, 对于每个  $k$ ,  $T|_{(\lambda_k, S)}$  均可对角化. 所以对每个  $k = 1, \dots, m$ , 都存在由  $T$  的特征向量组成的  $(\lambda_k, S)$  的基. 将这些基合并起来, 就得到了  $V$  的一个基【由式 (5.77)】, 且该基中每个向量既既是  $s$  的特征向量, 又是  $T$  的特征向量. 于是  $s$  和  $T$  关于这个基均具有对角矩阵, 命题得证

习题2 将上述结论推广到多于两个算子的情况

设  $V$  是非零有限维复向量空间. 那么  $V$  上的每个算子都有特征向量 (见5.19). 下面结论说明, 如果  $V$  上的两个算子可交换, 那么  $V$  中存在这两个算子共有的特征向量 (但这两个可交换算子并不一定有共同的特征值). 习题9 (a) 将该结论推广到多于两个算子的情况

### 5.78 可交换算子的公共特征向量

非零有限维复向量空间上的每对可交换算子都有公共的特征向量

#### 证

设  $V$  是非零有限维复向量空间且  $S, T \in \mathcal{L}(V)$  可交换. 令  $\lambda$  是  $s$  的特征值 (5.19 告诉我们  $s$  肯定有特征值). 于是  $(\lambda, S) \neq \{0\}$ . 且  $(\lambda, S)$  在  $T$  下不变 (由5.75)

于是, 再次利用5.19 得,  $T|_{(\lambda, S)}$  具有特征向量, 且该向量既是  $s$  的特征向量又是  $T$  的特征向量, 证毕

### 5.79 偏微分算子的公共特征向量

令  $\mathcal{P}_m(\mathbf{R}^2)$  定义如例 5.72 所示, 并令  $D_x, D_y \in \mathcal{L}(\mathcal{P}_m(\mathbf{R}^2))$  是该例中的可交换偏微分算子. 你可验证, 这两个算子的唯一特征值是0. 并且

$$(0, D_x) = \left\{ \sum_{k=0}^m a_k y^k : a_0, \dots, a_m \in \mathbf{R} \right\}$$
$$(0, D_y) = \left\{ \sum_{j=0}^m c_j x^j : c_0, \dots, c_m \in \mathbf{R} \right\}$$

这两个特征空间的交集是由这两个算子共有的特征向量所构成的集合. 因为  $(0, D_x) \cap (0, D_y)$  是由常值函数构成的集合, 所以我们可以看出  $D_x$  和  $D_y$  确实有公共的特征向量, 正如5.78 所言

下面结论将 5.47 (使算子具有上三角矩阵的基的存在性) 拓展至两个可交换算子的情形中

### 5.80 可交换算子可同时上三角化

设  $V$  是有限维复向量空间,  $s, T$  是  $V$  上的可交换算子. 那么存在  $V$  的一个基, 使得  $s$  和  $T$  关于该基均有上三角矩阵

#### 证

令  $n = \dim V$ . 我们对  $n$  用归纳法. 欲证结论对  $n = 1$  是成立的, 因为所有的  $1 \times 1$  矩阵都是上三角矩阵. 现在假设  $n > 1$ , 且欲证结论对于所有维数是  $n - 1$  的复向量空间都成立

令  $\nu_1$  为  $s$  和  $T$  共有的特征向量 (利用5.78). 因此  $S\nu_1 \in \text{span}(\nu_1)$  且  $T\nu_1 \in \text{span}(\nu_1)$ . 令  $W$  为  $V$  的子空间且满足

$$V = \text{span}(\nu_1) \oplus W$$

关于  $W$  这一子空间的存在性, 请看2.33. 定义线性映射  $: V \rightarrow W$  为: 对各  $a \in \mathbf{C}$  和各  $w \in W$  有

$$(a\nu_1 + w) = w$$

定义  $S, T \in \mathcal{L}(W)$  为: 对每个  $w \in W$ ,

$$Sw = (Sw) \quad Tw = (Tw)$$

为将归纳假设应用于  $S$  和  $T$ , 我们首先必须说明  $W$  上的这两个算子可交换. 为此, 设  $w \in W$ . 那么存在  $a \in \mathbf{C}$ , 使得

$$(ST)w = S((Tw)) = S(Tw - a\nu_1) = (S(Tw - a\nu_1)) = ((ST)w)$$

其中最后一个等号成立是因为  $\nu_1$  是  $s$  的特征向量且  $\nu_1 = 0$ . 类似有

$$(TS)w = ((TS)w)$$

因为算子  $S$  和  $T$  可交换, 所以上述两式就可说明  $(ST)w = (TS)w$ . 因此  $S$  和  $T$  可交换

于是我们可以利用归纳假设得出, 存在  $W$  的一个基  $\nu_2, \dots, \nu_n$  使得  $S$  和  $T$  关于该基都有上三角矩阵. 组  $\nu_1, \dots, \nu_n$  就是  $V$  的一个基. 若  $k \in \{2, \dots, n\}$ , 那么存在  $a_k, b_k \in \mathbf{C}$ , 使得

因为  $S$  和  $T$  关于  $\nu_2, \dots, \nu_n$  有上三角矩阵, 所以我们可知,  $S\nu_k \in \text{span}(\nu_2, \dots, \nu_k)$  且  $T\nu_k \in \text{span}(\nu_2, \dots, \nu_k)$ . 因此由上述等式得

$$S\nu_k \in \text{span}(\nu_1, \dots, \nu_k) \quad T\nu_k \in \text{span}(\nu_1, \dots, \nu_k)$$

于是,  $s$  和  $T$  关于基  $\nu_1, \dots, \nu_n$  有上三角矩阵, 命题得证

### 习题9 (b) 将上述结论推广到多于两个算子的情形

一般仅凭两个算子的特征值, 无法确定这两个算子的和或积的特征值. 然而当可交换时, 我们就有下面这个很棒的结论

### 5.81 可交换算子的和与积的特征值

设  $V$  是有限维复向量空间,  $s, T$  是  $V$  上的可交换算子. 那么  $s + T$  的每个特征值都等于  $s$  的特征值加上  $T$  的特征值  $ST$  的每个特征值都等于  $s$  的特征值乘以  $T$  的特征值

#### 证

存在  $V$  的一个基, 使得  $s$  和  $T$  关于该基都有上三角矩阵 (由5.80). 由3.35 和3.43, 关于该基的矩阵满足

矩阵加法的定义表明,  $\mathcal{M}(S + T)$  对角线上的每个元素都等于  $\mathcal{M}(S)$  的对角线和  $\nabla \mathcal{M}(T)$  对角线上对应元素之和. 类似地, 由于  $\mathcal{M}(S)$  和  $\nabla \mathcal{M}(T)$  都是上三角矩阵, 所以矩阵乘法的定义表明,  $\mathcal{M}(ST)$  对角线上的每个元素等于  $\mathcal{M}(S)$  的对角线和  $\nabla \mathcal{M}(T)$  对角线上对应元素之积. 此外,  $\mathcal{M}(S + T)$  和  $\mathcal{M}(ST)$  都是上三角矩阵 (见 5C 节的习题2)

$\mathcal{M}(S)$  对角线上的每个元素都是  $s$  的特征值,  $\nabla \mathcal{M}(T)$  对角线上的每个元素都是  $T$  的特征值 (由 5.41).  $s + T$  的每个特征值都在  $\mathcal{M}(S + T)$  对角线上,  $ST$  的每个特征值都在  $\mathcal{M}(ST)$  对角线上 (这两条结论同样源于5.41). 综上所述, 我们即可得出结论,  $s + T$  的每个特征值都等于  $s$  的特征值加上  $T$  的特征值,  $ST$  的每个特征值都等于  $s$  的特征值乘以  $T$  的特征值