目录

1 线性空间

1.1 数域, $\mathbf{R}^n$ 与 $\mathbf{C}^n$

定义 (域): 设 $\mathbf{F}$ 是一个元素个数大于等于 2 的集合, 其上定义了两个运算, 满足:

(1) (交换律) $\alpha+\beta=\beta+\alpha$, $\alpha\beta=\beta\alpha$.

(2) (结合律) $(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$, $(\alpha\beta)\gamma=\alpha(\beta\gamma)$.

(3) (单位元) $\alpha+0=\alpha$, $\alpha 1=\alpha$.

(4) (加法逆元) 对任意的 $\alpha\in\mathbf{F}$, 存在 $\beta\in\mathbf{F}$, 使得 $\alpha+\beta=0$.

(5) (乘法逆元) 对任意的 $0\neq\alpha\in\mathbf{F}$, 存在 $\beta\in\mathbf{F}$, 使得 $\alpha\beta=1$.

(6) (分配律) $\lambda(\alpha+\beta)=\lambda\alpha+\lambda\beta$.

则称 $\mathbf{F}$ 关于这两个运算作成一个域.

注: 以后, 记号 $\mathbf{F}$ 总是表示实数域 $\mathbf{R}$ 或复数域 $\mathbf{C}$.

一个域是一个集合, 它至少包含两个不同的元素 0 与 1, 带有加法与乘法运算, 并且这些运算满足上面提到的所有性质, 因此 $\mathbf{R}$ 与 $\mathbf{C}$ 都是域. 有理数集 $\mathbf{Q}$ 连同通常的加法与乘法也是域. 域的另一个例子是集合 $\{0,1\}$, 带有通常的加法与乘法运算, 但是规定 $1+1=0$.

在本书中, 我们无需处理 $\mathbf{R}$ 与 $\mathbf{C}$ 以外的域. 然而, 线性代数中许多对 $\mathbf{R}$ 与 $\mathbf{C}$ 有效的定义、定理、证明都可以照搬到任意域上. 如果想这这么做, 在第 $1\sim 4$ 章中我们尽可以把 $\mathbf{F}$ 当做任意的域, 而不仅仅是 $\mathbf{R}$ 和 $\mathbf{C}$, 只有个别例子和习题要求对任意的正整数 n, 有 $\underbrace{1+1+\cdots+1}_{n\text{个}}\neq 0$.

1.2 线性空间的定义

定义 (线性空间): 设集合 V 上有一个代数运算叫做加法, 数域 $\mathbf{F}$ 与 V 之间有一个运算叫做数量乘法, 且满足

(1) 对任意的 $\alpha,\beta\in V$, $\alpha+\beta=\beta+\alpha$;

(2) 对任意的 $\alpha,\beta,\gamma\in V$, $(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$;

(3) 存在 $0\in V$, 对任意的 $\alpha\in V$, $0+\alpha=\alpha$;

(4) 对任意的 $\alpha\in V$, 存在 $\beta\in V$, $\alpha+\beta=0$;

(5) 对任意的 $k\in\mathbf{F},\alpha,\beta\in V$, $k(\alpha+\beta)=k\alpha+k\beta$;

(6) 对任意的 $k,l\in\mathbf{F},\alpha\in V$, $(k+l)\alpha=k\alpha+l\alpha$;

(7) 对任意的 $k,l\in\mathbf{F},\alpha\in V$, $(kl)\alpha=k(l\alpha)$;

(8) $1\alpha=\alpha$,

则称 $(V,+,\cdot,\mathbf{F})$ 为数域 $\mathbf{F}$ 上的线性空间 (或向量空间). 当不至于引起混淆时, 简记为 V.

注: 在线性空间的定义中, 我们并没有要求集合 V 非空. 实际上, 空集不是线性空间, 因为它不满足定义中的 (3).

我们把 $\mathbf{R}$ 上的线性空间称为实线性空间, 把 $\mathbf{C}$ 上的线性空间称为复线性空间. 最简单的线性空间是 $\{0\}$, 只含有一个点. 此外, 不难验证 $\mathbf{F}^n$ 关于通常的加法与数量乘法是 $\mathbf{F}$ 上的线性空间. 正是这一例子启发我们定义了线性空间.

下面我们来看一下线性空间的一些例子.

例 1 设 $\mathbf{F}^\infty$ 是 $\mathbf{F}$ 中的无穷序列所成的集合:

$$\mathbf{F}^\infty = \{(x_1, x_2, \cdots) : x_j \in \mathbf{F}, j = 1, 2, \cdots\}.$$

定义 $\mathbf{F}^\infty$ 中的加法与数量乘法为

$$\begin{gathered}(x_1, x_2, \cdots) + (y_1, y_2, \cdots) = (x_1 + y_1, x_2 + y_2, \cdots),\\ \lambda(x_1, x_2, \cdots) = (\lambda x_1, \lambda x_2, \cdots).\end{gathered}$$

易知 $\mathbf{F}^\infty$ 按这样定义的加法与数乘构成一个线性空间.

设 S 为一个集合, 我们把从 S 到 $\mathbf{F}$ 的全体函数所成集合记为 $\mathbf{F}^S$. 对于函数 $f, g \in \mathbf{F}^S$, 我们规定它们的和函数 $f + g : S \to \mathbf{F}$ 为

$$(f + g)(x) = f(x) + g(x), \quad x \in S.$$

对于标量 $\lambda \in \mathbf{F}$, 规定函数 $\lambda f : S \to \mathbf{F}$ 为

$$(\lambda f) = \lambda f(x), \quad x \in S.$$

例 2 $\mathbf{F}^S$ 关于如上定义的加法与数乘构成一个线性空间.

注: 实际上, $\mathbf{F}^n$ 与 $\mathbf{F}^\infty$ 都可以看成 $\mathbf{F}^S$ 的特例. 也就是说, 我们可以将 $\mathbf{F}^n$ 看做是 $\mathbf{F}^{\{1,\cdots,n\}}$, 将 $\mathbf{F}^\infty$ 看做是 $\mathbf{F}^{\{1,2,\cdots\}}$.

1.3 线性空间的性质

性质：设 V 为数域 $\mathbf{F}$ 上的线性空间, 那么

(1) 零元唯一.

(2) 负元唯一.

(由此, 我们可以用记号 $-v$ 表示 v 的负元了. 此外, 我们定义 $w-v \xlongequal{\Delta} w+(-v)$.)

(3) $0v=0$.

(4) $\lambda 0=0$.

(5) $(-1)v=-v$.

证明 (1) 设 $0_1, 0_2$ 均为零元, 则

$$0_1=0_1+0_2=0_2+0_1=0_2.$$

故零元唯一.

(2) 设 β_1, β_2 均为 α 的负元, 则

$$\beta_2=0+\beta_2=(\alpha+\beta_1)+\beta_2=(\alpha+\beta_2)+\beta_1=0+\beta_1=\beta_1.$$

故负元唯一.

(3)

$$0v=(0+0)v=0v+0v,$$

故 $0v$ 是零元, 即 $0v=0$.

(4)

$$\lambda 0=\lambda(0+0)=\lambda 0+\lambda 0.$$

故 $\lambda 0$ 是零元, 即 $\lambda 0=0$.

(5)

$$0=0v=(1+(-1))v=1v+(-1)v=v+(-1)v.$$

故 $(-1)v$ 是 v 的负元, 即 $(-1)v=-v$. □

1.4 子空间

通过考虑子空间, 我们可以大量地扩充线性空间的例子.

定义 (子空间)：设 V 为数域 $\mathbf{F}$ 上的线性空间, U 为 V 的子集. 如果 U 关于同样的加法与数量乘法也构成线性空间, 则称 U 是 V 的子空间.

例 1 $\{(x_1,x_2,0):x_1,x_2\in\mathbf{F}\}$ 是 $\mathbf{F}^3$ 的子空间.

下面的结论给出了判断子集是否为子空间的方法 (甚至可以说是最简单的一种方法), 它让我们不必把线性空间定义中的每一条都验证一番.

命题 (子空间的判定)：设 V 为数域 $\mathbf{F}$ 上的线性空间, U 为 V 的子集, 如果

(1) U 非空.

(2) U 关于加法封闭.

(3) U 关于数乘封闭.

那么 U 是 V 的子空间.

显然, $\{0\}$ 是 V 的最小子空间, V 是 V 的最大子空间, 它们称为 V 的平凡子空间.

$\mathbf{R}^2$ 的子空间恰为 $\{0\}$、$\mathbf{R}^2$ 中所有过原点的直线与 $\mathbf{R}^2$ 自身. $\mathbf{R}^3$ 的子空间恰好为 $\{0\}$、$\mathbf{R}^3$ 中所有过原点的直线、$\mathbf{R}^3$ 中所有过原点的平面与 $\mathbf{R}^3$ 自身. 它们都是子空间这一点不难验证, 难点在于证明它们是 $\mathbf{R}^2$ 或 $\mathbf{R}^3$ 仅有的子空间. 当我们在下一章引入一些工具之后, 证明这些结论会变得容易一些.

1.5 子空间的和

在研究线性空间时, 我们通常只对子空间感兴趣, 而不是任意子集. 这时, 子空间的和这一概念非常有用.

定义 (子集的和)：设 V 为数域 $\mathbf{F}$ 上的线性空间, $U_1,\cdots,U_m$ 是 V 的子集, 我们称

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

为子集 $U_1,\cdots,U_m$ 的和.

例 1 设 U,W 是 $\mathbf{F}^3$ 的子集, $U=\{(x,0,0):x\in\mathbf{F}\}$, $W=\{(0,y,0):y\in\mathbf{F}\}$, 则

$$U+W=\{(x,y,0):x,y\in\mathbf{F}\}.$$

例 2 设 U,W 是 $\mathbf{F}^4$ 的子集, $U=\{(x,x,y,y):x,y\in\mathbf{F}\}$, $W=\{(x,x,x,y):x,y\in\mathbf{F}\}$, 则

$$U+W=\{(x,x,y,z):x,y,z\in\mathbf{F}\}.$$

下面的结果说明子空间的和仍是子空间, 并且是包含这些子空间的最小子空间.

命题 (子空间之和仍是子空间)：设 V 为数域 $\mathbf{F}$ 上的线性空间, $U_1, \cdots, U_m$ 是 V 的子空间, 那么 $U_1 + \cdots + U_m$ 也是 V 的子空间, 并且是包含 $U_1, \cdots, U_m$ 的最小子空间.

证明 先证 $U_1 + \cdots + U_m$ 是子空间. 因为 $0 = 0 + \cdots + 0 \in U_1 + \cdots + U_m$, 故 $U_1 + \cdots + U_m$ 非空. 又

$$(\alpha_1 + \cdots + \alpha_m) + (\beta_1 + \cdots + \beta_m) = (\alpha_1 + \beta_1) + \cdots + (\alpha_m + \beta_m) \in U_1 + \cdots + U_m,$$
$$k(\alpha_1 + \cdots + \alpha_m) = k\alpha_1 + \cdots + k\alpha_m \in U_1 + \cdots + U_m,$$

故 $U_1 + \cdots + U_m$ 是子空间.

再证 $U_1 + \cdots + U_m$ 是包含 $U_1, \cdots, U_m$ 的最小子空间. 设 U 是包含 $U_1, \cdots, U_m$ 的任一子空间. 任取 $\alpha_1 \in U_1, \cdots, \alpha_m \in U_m$, 则 $\alpha_1, \cdots, \alpha_m \in U$, 故 $\alpha_1 + \cdots + \alpha_m \in U$, 即 $U_1 + \cdots + U_m \subseteq U$. □

注：子空间的并一般不是子空间, 这就是我们通常讨论子空间的和, 而不讨论子空间的并的原因.

线性空间中子空间的和类似于集合论中子集的并集. 给定一个线性空间的两个子空间, 它们的和是包含它们的最小子空间. 类似地, 给定一个集合的两个子集, 它们的并是包含它们的最小子集.

定义 (子空间的直和)：设 V 为数域 $\mathbf{F}$ 上的线性空间, $U_1, \cdots, U_m$ 是 V 的子空间, 若 $U_1 + \cdots + U_m$ 中每一个元素都能唯一地表示为 $u_1 + \cdots + u_m$, 其中 $u_j \in U_j$, 则称子空间的和 $U_1 + \cdots + U_m$ 是直和 (direct sum), 记为

$$U_1 \oplus \cdots \oplus U_m.$$

例 3 设有 $\mathbf{F}^3$ 的子空间

$$U = \{(x, y, 0) : x, y \in \mathbf{F}\}, \quad W = \{(0, 0, z) : z \in \mathbf{F}\},$$

则 $\mathbf{F}^3 = U \oplus W$.

例 4 设 U_j 是 $\mathbf{F}^n$ 中除第 j 个分量外, 其余分量均为 0 的那些元素所成的子空间, 例如 $U_2 = \{(0, x, 0, \cdots, 0) \in \mathbf{F}^n : x \in \mathbf{F}\}$, 那么

$$\mathbf{F}^n = U_1 \oplus \cdots \oplus U_n.$$

反面的例子有时能像证明正面的例子一样加深我们的理解.

例 5 设有 $\mathbf{F}^3$ 的三个子空间

$$\begin{aligned}U_1&=\{(x,y,0):x,y\in\mathbf{F}\},\\U_2&=\{(0,0,z):z\in\mathbf{F}\},\\U_3&=\{(0,y,y):y\in\mathbf{F}\}.\end{aligned}$$

试证明 $U_1+U_2+U_3$ 不是直和.

证明 因为

$$(x,y,z)=(x,y,0)+(0,0,z)+(0,0,0),$$

因此 $\mathbf{F}^3=U_1+U_2+U_3$. 下面证明 $\mathbf{F}^3$ 中存在表示不唯一的元素. 因为

$$\begin{aligned}0&=(0,0,0)+(0,0,0)+(0,0,0)\\&=(0,1,0)+(0,0,1)+(0,-1,-1),\end{aligned}$$

故 0 的表示方法不唯一, $U_1+U_2+U_3$ 不是直和. □

直和的定义要求每一个元素都能唯一地表示为个空间中的元素之和. 以下的结果表明, 我们只需要验证 0 的表示方法是否唯一.

命题 (**用零元的表示方法来判定直和**): $U_1+\cdots+U_m$ 是直和的充要条件为 0 的表示方法唯一, 即 0 表示成 $u_1+\cdots+u_m$ 的唯一方式是 $u_1=\cdots=u_m=0$.

证明 必要性显然, 下面证明充分性. 任取 $\alpha\in V$, 设 α 有两种表示

$$\begin{aligned}\alpha&=\alpha_1+\cdots+\alpha_m,\\\alpha&=\beta_1+\cdots+\beta_m.\end{aligned}$$

从而

$$0=\alpha-\alpha=(\alpha_1-\beta_1)+\cdots+(\alpha_m-\beta_m).$$

因此

$$\alpha_1-\beta_1=0,\quad\cdots,\quad\alpha_m-\beta_m=0,$$

即 α 的表示方法唯一. 由 α 的任意性知 $U_1+\cdots+U_m$ 是直和. □

下面的结果给出了验证两个子空间的和是否为直和的简单方法.

命题 (**利用交集判定直和, 只适用于两个子空间**): 设 U,W 是 V 的子空间, 则 $U+W$ 为直和的充要条件为 $U\cap W=\{0\}$.

证明 ($\Longrightarrow$) 任取 $\alpha \in U \cap W$, 注意到 $0 \in U \cap W$, 因为

$$\alpha = \alpha + 0 = 0 + \alpha,$$

因此由分解的唯一性知 $\alpha = 0$, 即 $V \cap W = \{0\}$.

($\Longleftarrow$) 我们来证明 0 的分解是唯一的. 设

$$0 = \alpha + \beta, \quad \alpha \in U, \beta \in W,$$

则 $\alpha = -\beta \in W$. 因此 $\alpha \in U \cap W = \{0\}$, $\alpha = 0$, 进而 $\beta = 0$, 也即 0 的分解唯一. □

上面的结果只考虑了两个子空间的情形. 遗憾的是, 在考虑多个子空间的和是否为直和时, 仅仅验证任意两个子空间的交为 $\{0\}$ 是不够的. 对于这一点, 你可以考察一下前面的例5. 但还好, 利用两个子空间的结论, 我们不难得到多个子空间的相应结论.

命题 (利用交集判定直和, 多个子空间的情形): 设 V 为数域 $\mathbf{F}$ 上的线性空间, $U_1, \cdots, U_m$ 是 V 的子空间, 则 $U_1 + \cdots + U_m$ 是直和的充要条件为

$$U_j \cap (U_1 + \cdots + U_{j-1} + U_{j+1} + \cdots + U_m) = \{0\}, \quad j = 1, \cdots, n.$$

证明 容易验证子空间的加法具有交换律与结合律, 因此, 要验证 $U_1 \oplus \cdots \oplus U_m$, 只需验证对任意的 $1 \leqslant j \leqslant m$, 下面两个子空间的和

$$U_j + (U_1 + \cdots + U_{j-1} + U_{j+1} + \cdots + U_m).$$

是直和, 而这等价于

$$U_j \cap (U_1 + \cdots + U_{j-1} + U_{j+1} + \cdots + U_m) = \{0\}.$$

从而结论得证. □

1.6 习题

1. 若存在正数 p, 对任意的 $x \in \mathbf{R}$, 都有 $f(x) = f(x+p)$, 则称 f 是周期函数. 试问: $\mathbf{R}$ 到 $\mathbf{R}$ 的全体周期函数构成 $\mathbf{R}^{\mathbf{R}}$ 的子空间吗?

解 不构成. 例如取周期函数 $f(x) = \sin\sqrt{2}x$ 与 $g(x) = \cos x$, 它们的和函数 $f(x) + g(x) = \sin\sqrt{2}x + \cos x$ 不是周期函数.

2. 设 U_1, U_2 为 V 的子空间, 证明 $U_1 \cap U_2$ 为 V 的子空间.

3. 设 $\{U_j\}_{j\in J}$ 是 V 的一族子空间, 证明 $\bigcap\limits_{j\in J} U_j$ 也是 V 的子空间.

4. 证明子空间的加法具有交换律与结合律. 即是说, 若 U_1, U_2, U_3 都是 V 的子空间, 那么

$$U_1 + U_2 = U_2 + U_1,$$
$$(U_1 + U_2) + U_3 = U_1 + (U_2 + U_3).$$

5. 举出反例说明: 子空间的加法不具有消去律, 且当子空间的加法加强为直和时, 消去律仍然不成立.

解 只需说明第二点. 取 $V = \mathbf{R}^2$, $U_1 = \{(x, 0) : x \in \mathbf{R}\}$, $U_2 = \{(0, y) : y \in \mathbf{R}\}$ 与 $W = \{(z, z) : z \in \mathbf{R}\}$, 则

$$V = U_1 \oplus W = U_2 \oplus W,$$

但 $U_1 \neq U_2$.

6. V 的子空间的加法运算有单位元吗? 哪些子空间有加法逆元?

解 有加法单位元 $\{0\}$. 只有子空间 $\{0\}$ 有加法逆元.

7. 设函数 $f : \mathbf{R} \to \mathbf{R}$. 若对任意的 $x \in \mathbf{R}$, 有 $f(-x) = f(x)$, 则称 f 为偶函数; 若对任意的 $x \in \mathbf{R}$, 有 $f(-x) = -f(x)$, 则称 f 为奇函数. 令 U_e 为 $\mathbf{R}$ 上的实值偶函数的集合, U_o 为 $\mathbf{R}$ 上的实值奇函数的集合. 试证明:

$$\mathbf{R}^{\mathbf{R}} = U_e \oplus U_o.$$

证明 显然 U_e, U_o 均为 $\mathbf{R}^{\mathbf{R}}$ 的子空间. 任取 $f \in \mathbf{R}^{\mathbf{R}}$, 令

$$f_e(x) = \frac{f(x) + f(-x)}{2},$$
$$f_o(x) = \frac{f(x) - f(-x)}{2},$$

则 $f_e \in U_e$, $f_o \in U_o$, 且 $f = f_e + f_o$. 因此 $\mathbf{R}^{\mathbf{R}} = U_e + U_o$. 下证 $\mathbf{R}^{\mathbf{R}} = U_e \oplus U_o$. 任取 $f \in U_e \cap U_o$, 则对任意的 $x \in \mathbf{R}$, 有

$$-f(x) = f(-x) = f(x).$$

因此 $f(x) = 0$. 由 x 的任意性知 $f = 0$, 因此 $U_e \cap U_o = \{0\}$, $\mathbf{R}^{\mathbf{R}} = U_e \oplus U_o$. □

2 有限维线性空间

上一章我们介绍了线性空间. 但线性代数所关注的并不是任意的线性空间, 而是本章介绍的有限维线性空间. 回忆一下记号: 我们总用 $\mathbf{F}$ 代表 $\mathbf{R}$ 或 $\mathbf{C}$, 用 V 代表数域 $\mathbf{F}$ 上的线性空间.

2.1 张成空间

定义 (线性组合)：我们称形如

$$k_1v_1+\cdots+k_mv_m$$

的元素为 V 中向量 $v_1,\cdots,v_m$ 的线性组合，其中 $k_1,\cdots,k_m\in\mathbf{F}$.

定义 (张成空间)：V 中一组向量 $v_1,\cdots,v_m$ 的线性组合全体所成集合

$$\{k_1v_1+\cdots+k_mv_m : k_i\in\mathbf{F}, i=1,\cdots,m\}.$$

称为 $v_1,\cdots,v_m$ 的张成空间，记为 $\operatorname{span}\{v_1,\cdots,v_m\}$. 空向量组的张成空间定义为 $\{0\}$.

命题 (张成空间是包含这组向量的最小子空间)：V 中向量 $v_1,\cdots,v_m$ 的张成空间 $\operatorname{span}\{v_1,\cdots,v_m\}$ 是包含 $v_1,\cdots,v_m$ 的最小子空间.

定义 (张成)：若 $\operatorname{span}\{v_1,\cdots,v_m\}=V$，则称向量组 $v_1,\cdots,v_m$ 张成了 V.

例 1 向量组

$$e_1=(1,0,\cdots,0),\quad e_2=(0,1,0,\cdots,0),\quad \cdots\quad ,e_n=(0,0,\cdots,0,1)$$

张成了 $\mathbf{F}^n$.

现在我们给出线性代数中的一个关键定义.

定义 (有限维线性空间)：若 V 可被有限个向量张成，则称 V 是有限维线性空间.

例 2 对任意的正整数 n, $\mathbf{F}^n$ 都是有限维线性空间. (由例 1 即得)

我们记 $\mathcal{P}(\mathbf{F})$ 是系数属于 $\mathbf{F}$ 的全体多项式所成集合. 不难验证 $\mathcal{P}(\mathbf{F})$ 是 $\mathbf{F}^{\mathbf{F}}$ 的子空间. 此外，我们规定零多项式 0 的次数为 $-\infty$. 另外，我们把系数在 $\mathbf{F}$ 中的次数不超过 m 的全体多项式所成集合记为 $\mathcal{P}_m(\mathbf{F})$. 这样一来，零多项式属于 $\mathcal{P}_m(\mathbf{F})$.

例 3 对任意的非负整数 m, $\mathcal{P}_m(\mathbf{F})$ 是有限维线性空间.

证明 只需注意到 $\operatorname{span}\{1,z,\cdots,z^m\}=\mathcal{P}_m(\mathbf{F})$. $\square$

定义 (无限维线性空间)：若 V 不是有限维线性空间，则称 V 是无限维线性空间.

例 4 $\mathcal{P}(\mathbf{F})$ 是无限维线性空间.

证明 假设 $\mathcal{P}(\mathbf{F})$ 是有限维的. 设多项式 $p_1,\cdots,p_m$ 张成了它，且 $p_1,\cdots,p_m$ 中的最高次数为 k，那么 $z^{k+1}\notin\operatorname{span}\{p_1,\cdots,p_m\}=\mathcal{P}(\mathbf{F})$，矛盾. $\square$

2.2 线性无关

设 $v \in \mathrm{span}\{v_1, \cdots, v_m\}$, 那么存在 $k_1, \cdots, k_m \in \mathbf{F}$, 使得

$$v = k_1 v_1 + \cdots + k_m v_m.$$

下面我们考虑系数 $k_1, \cdots, k_m$ 选取的唯一性问题. 假设还有 $l_1, \cdots, l_m \in \mathbf{F}$, 使得

$$v = l_1 v_1 + \cdots + l_m v_m,$$

那么

$$0 = (k_1 - l_1)v_1 + \cdots + (k_m - l_m)v_m.$$

由此可以看出, 若 v 表示成 $v_1, \cdots, v_m$ 的线性组合的方法是唯一的, 那么 0 表示成 $v_1, \cdots, v_m$ 的线性组合的方法是唯一的, 只能是每个 v_i 的系数都取 0. 受此启发, 我们有下面的线性无关的定义.

定义 (线性无关): 若使得 $a_1 v_1 + \cdots + a_m v_m = 0$ 成立的系数 $a_1, \cdots, a_m \in \mathbf{F}$ 只有

$$a_1 = \cdots = a_m = 0,$$

那么我们就称向量组 $v_1, \cdots, v_m$ 是线性无关的. 规定空向量组是线性无关的.

由本节开始的推导可知, 若 $v_1, \cdots, v_m$ 是线性无关的, 那么 $\mathrm{span}\{v_1, \cdots, v_m\}$ 中每个向量都可以唯一地表示为 $v_1, \cdots, v_m$ 的线性组合.

例 1 (线性无关组)

(1) 由一个向量组成的向量组 v 线性无关当且仅当 $v \neq 0$.

(2) 由两个向量组成的向量组线性无关当且仅当每个向量都不能写成另一个向量的倍数.

(3) 对每个非负整数 m, $\mathcal{P}(\mathbf{F})$ 中的向量组 $1, z, \cdots, z^m$ 都是线性无关的.

定义 (线性相关): 若向量组 $v_1, \cdots, v_m$ 不是线性无关的, 则称它们是线性相关的.

不难发现, 向量组 $v_1, \cdots, v_m$ 线性相关等价于存在不全为 0 的数 $k_1, \cdots, k_m \in \mathbf{F}$, 使得

$$k_1 v_1 + \cdots + k_m v_m = 0.$$

下面的线性相关引理告诉我们, 如果一组向量是线性相关的, 那么其中必然有某个向量可以被它前面的诸向量线性表出, 并且扔掉该向量后, 余下的向量所张成的空间和原来是一样的.

引理 (线性相关引理)：设 $v_1, \cdots, v_m$ 是一组线性相关的向量, 则有 $j \in \{1, \cdots, m\}$, 使得

(1) $v_j \in \operatorname{span}\{v_1, \cdots, v_{j-1}\}$.

(2) $\operatorname{span}\{v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_m\} = \operatorname{span}\{v_1, \cdots, v_m\}$.

证明 存在不全为 0 的数 $k_1, \cdots, k_m$, 使得 $k_1 v_1 + \cdots + k_m v_m = 0$. 设其中下标最大的非零系数是 k_j, 则 $k_1 v_1 + \cdots + k_j v_j = 0$. 当 $j = 1$ 时结论显然成立, 下设 $j > 1$.

(1) $v_j = -\dfrac{k_1}{k_j} v_1 - \cdots - \dfrac{k_{j-1}}{k_j} v_{j-1} \in \operatorname{span}\{v_1, \cdots, v_{j-1}\}$.

(2) 设 $b = c_1 v_1 + \cdots + c_m v_m \in \operatorname{span}\{v_1, \cdots, v_m\}$, 由上式知 $b \in \operatorname{span}\{v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_m\}$, 故 $\operatorname{span}\{v_1, \cdots, v_m\} \subseteq \operatorname{span}\{v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_m\}$. $\square$

注：由证明过程可以看出, 上述引理中的 j 其实就是非零系数的最大下标. 这一点对于下面定理的证明是至关重要的.

定理 (线性无关组的长度 $\leqslant$ 张成组的长度)：在有限维线性空间中, 线性无关向量组的长度小于等于该向量空间每一个张成组的长度.

证明 设 $u_1, \cdots, u_m$ 是 V 中的一组线性无关的向量, $w_1, \cdots, w_n$ 是 V 的张成组. 下证 $m \leqslant n$. 下面的每一步都是添加了某个 u, 而去掉了某个 w.

第 1 步: 因为 u_1 可由 $w_1, \cdots, w_n$ 线性表出, 故

$$u_1, w_1, \cdots, w_n$$

是线性相关的. 因为 u_1 线性无关, 故根据线性相关引理及其注记, 存在个某 w, 去掉该 w 后, 余下向量所张成的空间仍为 V. 不妨设去掉了 w_1. 我们得到了向量组 $u_1, w_2, \cdots, w_n$, 且它们张成了 V.

第 j 步: 因为 $u_1, \cdots, u_{j-1}, w_j, \cdots, w_n$ 张成了 V, 故添加了 u_j 进去之后, 向量组

$$u_1, \cdots, u_j, w_j, \cdots, w_n$$

是线性相关的. 因为 $u_1, \cdots, u_j$ 是线性无关的, 故由线性相关引理可知存在某个 w, 去掉该 w 之后, 余下的 n 个向量仍然张成了 V. 不妨设去掉的是 w_j.

经过 m 步之后, 我们将所有的 u 都添加到了 V 的张成组中, 且保持张成组的向量个数不变, 因此 $m \leqslant n$. $\square$

定理 (有限维空间的子空间是有限维的)：有限维线性空间的子空间一定是有限维的.

证明 设 V 是有限维线性空间, U 是 V 的子空间. 注意下面每一步取的 v 都是 U 中的元素.

第 1 步: 若 $U = \{0\}$, 则 U 是有限维的. 若 $U \neq \{0\}$, 则可取 $0 \neq v_1 \in U$.

第 j 步: 若 $U=\text{span}\{v_1,\cdots,v_{j-1}\}$, 则 U 是有限维的. 若 $U\neq\text{span}\{v_1,\cdots,v_{j-1}\}$(说明它是 U 的真子集), 故可取到 $v_j\in U$, 但 $v_j\notin\text{span}\{v_1,\cdots,v_{j-1}\}$.

该程序每执行一次, 线性无关组的长度就加 1. 而线性无关组的长度不超过张成组的长度, 因此此程序必然在有限步内停止, 故 U 是有限维的. □

2.3 基

前面我们讨论了线性空间的张成组和线性无关组, 下面我们将这两个概念结合起来.

定义 (基): 若 V 中的一个向量组 $v_1,\cdots,v_m$ 既线性无关, 又张成 V, 则称向量组 $v_1,\cdots,v_m$ 是 V 的基 (basis).

例 1

(1) $(1,0,\cdots,0),(0,1,0,\cdots,0),\cdots,(0,\cdots,0,1)$ 是 $\mathbf{F}^n$ 的基, 称为 $\mathbf{F}^n$ 的**标准基**.

(2) $1,z,\cdots,z^m$ 是 $\mathcal{P}_m(\mathbf{F})$ 的基.

下面的性质说明了为什么基非常有用.

命题 (基的作用): V 中的向量组 $v_1,\cdots,v_n$ 是 V 的基当且仅当每一个 $v\in V$ 都可以唯一地写成

$$v=a_1v_n+\cdots+a_nv_n.$$

证明 ($\Longrightarrow$) 因为 $\text{span}\{v_1,\cdots,v_n\}=V$, 故 $v\in V$ 可表示为 $v_1,\cdots,v_n$ 的线性组合. 下证表示的唯一性. 设有两种表示方式

$$v=a_1v_n+\cdots+a_nv_n=b_1v_n+\cdots+b_nv_n.$$

那么

$$(a_1-b_1)v_1+\cdots+(a_n-b_n)v_n=0.$$

由 $v_1,\cdots,v_n$ 线性无关知 $a_1=b_1,\cdots,a_n=b_n$.

($\Longleftarrow$) 由条件知 $\text{span}\{v_1,\cdots,v_n\}=V$. 下证 $v_1,\cdots,v_n$ 线性无关. 设

$$k_1v_1+\cdots+k_nv_n=0.$$

因为

$$0v_1+\cdots+0v_n=0,$$

故由表示的唯一性知

$$k_1=\cdots=k_n=0,$$

即 $v_1,\cdots,v_n$ 线性无关, 它们是 V 的基. □

引理 (张成组含有基): 有限维向量空间的任何一个张成组可以化简成一个基.

证明 设 $v_1,\cdots,v_n$ 是 V 的张成组, 将其记为 B.

第 1 步: 若 $v_1=0$, 则从 B 中去掉 v_1. 若 $v_1\neq 0$, 则 B 保持不变.

第 j 步: 若 $v_j\in\operatorname{span}\{v_1,\cdots,v_{j-1}\}$, 则从 B 中去掉 v_j. 若 $v_j\notin\operatorname{span}\{v_1,\cdots,v_{j-1}\}$, 则 B 保持不变.

重复程序 n 步, 最后得到的向量组就是 V 的基. 这是因为: (1) 去掉的向量已经包含在它前面诸向量的张成空间里面, 故最后 B 仍然张成 V. (2) 留下的向量都不属于它前面诸向量的张成空间, 因此最后 B 中的向量都是线性无关的 (这一点可以用反证法证明) . □

注: 从证明过程我们可以看出,"有限维"这一条件是不可或缺的. 下面我们即将看到的两个重要结论, 它们的证明用到了这一引理, 因此也必须要求"有限维".

定理 (有限维线性空间基的存在性): 每个有限维线性空间都有基.

证明 根据定义, 有限维线性空间都有张成组, 而前面的结论告诉我们张成组可以化简为基. □

下面的结论可以看成是前面引理的对偶. 前面的引理说的是任何一个张成组都可以化简为基, 而下面的命题告诉我们任何一个线性无关组都可以扩充为基.

定理 (基扩充定理): 有限维线性空间 V 的任一线性无关组都可以扩充为 V 的一个基.

证明 设 $u_1,\cdots,u_m$ 是 V 中的一个线性无关组, $w_1,\cdots,w_n$ 是 V 的一个基. 因为向量组

$$u_1,\cdots,u_m,w_1,\cdots,w_n$$

张成了 V, 因此我们可以按照前述引理的方法, 将其化简为一个基. 这个基包含了 $u_1,\cdots,u_m$ 和某些 w (因为 $u_1,\cdots,u_m$ 是线性无关的, 故它们满足被保留下来的条件) . □

2.4 维数

虽然我们一直在讨论有限维线性空间, 但还不曾定义过有限维线性空间的维数. 注意到 $\mathbf{F}^n$ 的标准基

$$(1,0,\cdots,0),(0,1,0,\cdots,0),\cdots,(0,\cdots,0,1)$$

的长度为 n, 因此我们希望用基的长度来定义线性空间的维数. 不过一般来说, 一个线性空间可以有很多不同的基. 这样的话, 只有保证每个基都具有相同的长度, 我们的定义才是合理的. 幸运的是, 情况确实如此.

命题 (**基的长度不依赖基的选取**)：有限维线性空间的任意两个基的长度都相等.

证明 设 B 与 C 是 V 的两个基. 因为 B 是线性无关组, C 是张成组, 因此 B 的长度小于等于 C. 同理可得 C 的长度小于等于 B. 因此二者长度相等. □

现在我们可以正式地定义有限维线性空间的维数了.

定义 (**线性空间的维数**)：我们把有限维线性空间 V 的基的长度称为 V 的维数 (dimension), 记为 $\dim V$.

例 1

(1) $\dim \mathbf{F}^n = n$, 因为 $\mathbf{F}^n$ 的标准基的长度为 n.

(2) $\dim \mathcal{P}_m(\mathbf{F}) = m+1$, 因为基 $1, z, \cdots, z^m$ 的长度为 $m+1$.

命题 (**子空间的维数**)：设 V 是有限维线性空间, U 是 V 的子空间, 那么 $\dim U \leqslant \dim V$.

证明 前面我们已经证明了有限维线性空间的子空间是有限维的, 因此具有维数. 注意到 U 的基是 V 中的线性无关组, 而 V 的基是张成组, 因此 $\dim U \leqslant \dim V$. □

要验证一个向量组是否为基, 我们必须验证两点: 是否张成 V, 是否线性无关. 但是, 对于具有适当长度的向量组来说, 只需要验证它满足两个性质之一. 这就是下面的两个命题.

命题 (**具有适当长度的线性无关组是基**)：设 V 是有限维线性空间, 则 V 中每个长度为 $\dim V$ 的线性无关组都是 V 的基.

证明 设 $\dim V = n$ 且 $v_1, \cdots, v_n$ 是 V 的一个线性无关组. 根据基扩充定理, 我们可以将 $v_1, \cdots, v_n$ 扩充为 V 的基. 由于 V 的每个基的长度都是 n, 因此我们无需给 $v_1, \cdots, v_n$ 添加向量. 这说明 $v_1, \cdots, v_n$ 就是 V 的基. □

命题 (**具有适当长度的张成组是基**)：设 V 是有限维线性空间, 则 V 中每个长度为 $\dim V$ 的张成组都是 V 的基.

证明 设 $\dim V = n$ 且 $v_1, \cdots, v_n$ 是 V 的一个张成组. 因为张成组可以化简为基, 而 V 的每个基的长度都是 n, 因此我们不必去掉 $v_1, \cdots, v_n$ 中的任何向量. 这说明 $v_1, \cdots, v_n$ 是 V 的基. □

下面的命题给出了有限维线性空间的两个子空间之和的维数公式, 它类似于有限集合的计数公式.

定理 (子空间和的维数)：设 U_1 和 U_2 是有限维线性空间 V 的两个子空间, 那么

$$\dim(U_1+U_2)=\dim U_1+\dim U_2-\dim(U_1\cap U_2).$$

证明 设 $u_1,\cdots,u_m$ 是 $U_1\cap U_2$ 的基. 我们将其扩充为 U_1 的基

$$u_1,\cdots,u_m,v_1,\cdots,v_j.$$

又将 $u_1,\cdots,u_m$ 扩充为 U_2 的基

$$u_1,\cdots,u_m,w_1,\cdots,w_k.$$

下面证明

$$u_1,\cdots,u_m,v_1,\cdots,v_j,w_1,\cdots,w_k$$

是 U_1+U_2 的基即可. 设

$$a_1u_1+\cdots+a_mu_m+b_1v_1+\cdots+b_jv_j+c_1w_1+\cdots+c_kw_k=0. \tag{1}$$

那么

$$c_1w_1+\cdots+c_kw_k=-a_1u_1-\cdots-a_mu_m-b_1v_1-\cdots-b_jv_j.$$

因为上述等式右边属于 U_1, 左边属于 U_2, 因此 $c_1w_1+\cdots+c_kw_k\in U_1\cap U_2$. 故可设

$$c_1w_1+\cdots+c_kw_k=d_1u_1+\cdots+d_mu_m.$$

因为 $u_1,\cdots,u_m,w_1,\cdots,w_k$ 是 U_2 的基, 所以所有的 c 与 d 都为 0. 从而由 (1) 式可得

$$a_1u_1+\cdots+a_mu_m+b_1v_1+\cdots+b_jv_j=0.$$

由于 $u_1,\cdots,u_m,v_1,\cdots,v_j$ 是 U_1 的基, 因此所有的 a 与 b 全都是 0.

综上, $u_1,\cdots,u_m,v_1,\cdots,v_j,w_1,\cdots,w_k$ 是 U_1+U_2 的线性无关组. 显然它们是 U_1+U_2 的张成组, 因此它们是 U_1+U_2 的基. □

注：你可能会猜测, 对于三个有限维子空间的情形, 也有与集合的计数公式相似的结论:

$$\begin{aligned}\dim(U_1+U_2+U_3)=&\dim U_1+\dim U_2+\dim U_3\\&-\dim(U_1\cap U_2)-\dim(U_1\cap U_3)-\dim(U_2\cap U_3)\\&+\dim(U_1\cap U_2\cap U_3).\end{aligned}$$

但令人遗憾的是, 这一公式并不成立. 关于此, 你可以看看下一节最后一个习题.

2.5 习题

1. 设 $U_1, \cdots, U_m$ 是 V 的有限维子空间, 证明 $U_1 + \cdots + U_m$ 也是有限维的, 并且

$$\dim(U_1 + \cdots + U_m) \leqslant \dim U_1 + \cdots + \dim U_m.$$

证明 设 $u_{i_1}, \cdots, u_{i_k}$ 是 U_i 的基, 那么

$$U_1 + \cdots + U_m = \operatorname{span}\{u_{1_1}, \cdots, u_{1_k}, \cdots, u_{m_1}, \cdots, u_{m_k}\}.$$

因此 $U_1 + \cdots + U_m$ 是有限维的. 注意到基的长度小于等于张成组的长度, 故题中不等式成立. □

2. 设 V 是有限维线性空间, $\dim V = n \geqslant 1$. 证明存在 V 的 n 个 1 维子空间 $U_1, \cdots, U_n$, 使得

$$V = U_1 \oplus \cdots \oplus U_n.$$

证明 设 $u_1, \cdots, u_n$ 是 V 的基. 令 $U_i = \operatorname{span}\{u_i\}$, 则 $\dim U_i = 1$ 且 $V = U_1 + \cdots + U_n$. 由 $u_1, \cdots, u_n$ 线性无关知 $U_1 + \cdots + U_n$ 是直和. □

3. 设 $U_1, \cdots, U_m$ 是 V 的有限维子空间, 且 $U_1 + \cdots + U_m$ 是直和. 证明

$$\dim(U_1 \oplus \cdots \oplus U_m) = \dim U_1 + \cdots + \dim U_m.$$

证明 由第 1 题知 $U_1 \oplus \cdots \oplus U_m$ 是有限维子空间. 下面对 m 使用数学归纳法.

当 $m = 2$ 时, 由 $U_1 + U_2$ 是直和, 我们得到 $U_1 \cap U_2 = \{0\}$, 故

$$\dim(U_1 \oplus U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2.$$

假设结论对 $m-1$ 成立, 下面我们考察 m 个子空间的情况. 已知 $U_1 + \cdots + U_m$ 是直和, 易证 $U_1 + \cdots + U_{m-1}$ 也是直和 (可以利用反证法). 因此

$$\dim(U_1 \oplus \cdots \oplus U_{m-1}) = \dim U_1 + \cdots + \dim U_{m-1}.$$

记 $W=U_1\oplus\cdots\oplus U_{m-1}$, 那么

$$\begin{aligned}\dim(U_1\oplus\cdots\oplus U_m)&=\dim(W\oplus U_m)=\dim W+\dim U_m\\&=\dim U_1+\cdots+\dim U_{m-1}+\dim U_m.\end{aligned}$$

结论得证. □

4. 两个有限维子空间和的维数公式与有限集合的计数公式类似. 你可能会猜测, 对于三个有限维子空间, 有结论:

$$\begin{aligned}\dim(U_1+U_2+U_3)=&\dim U_1+\dim U_2+\dim U_3\\&-\dim(U_1\cap U_2)-\dim(U_1\cap U_3)-\dim(U_2\cap U_3)\\&+\dim(U_1\cap U_2\cap U_3).\end{aligned}$$

但遗憾的是, 此公式并不成立, 试举出反例.

解 取 $V=\mathbf{R}^2$, 令

$$U_1=\{(x,0):x\in\mathbf{R}\},\quad U_2=\{(0,y):y\in\mathbf{R}\},\quad U_3=\{(x,x):x\in\mathbf{R}\}.$$

那么 $V=U_1+U_2+U_3$, 故

$$\dim V=\dim(U_1+U_2+U_3)=2.$$

但是

$$\begin{gathered}\dim U_1=\dim U_2=\dim U_3=1,\\ \dim(U_1\cap U_2)=\dim(U_1\cap U_3)=\dim(U_2\cap U_3)=0,\\ \dim(U_1\cap U_2\cap U_3)=0.\end{gathered}$$

因此题中所给等式不成立.

3 线性映射

现在我们要来讨论线性代数真正令人感兴趣的部分——线性映射了. 我们用 $\mathbf{F}$ 代表 $\mathbf{R}$ 或 $\mathbf{C}$, 用 V 和 W 代表 $\mathbf{F}$ 上的线性空间.

3.1 线性映射

定义 (线性映射)：设 V 和 W 是数域 $\mathbf{F}$ 上的线性空间，若映射 $T: V \to W$ 满足:

(1) $T(u+v) = Tu + Tv$;

(2) $T(\lambda v) = \lambda(Tv)$;

则称 T 是从 V 到 W 的线性映射.

我们把从 V 到 W 的全体线性映射构成的集合记作

$$\mathcal{L}(V, W).$$

例 1 (**常见的线性映射**)

(1) (**零映射**) 我们称把一个空间中的任意元素映为另一个空间的零元的映射 $0 \in \mathcal{L}(V, W)$ 为零映射, 它具体定义为

$$0v = 0.$$

这里等号左边的 0 是映射, 右边的 0 是 W 中的零元.

(2) (**恒等映射**) 我们把从 V 到自身的映射 $I \in \mathcal{L}(V, V)$ 称为恒等映射, 它的具体定义为

$$Iv = v.$$

(3) (**微分**) 我们定义 $D \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ 如下:

$$Dp = p'.$$

(4) (**积分**) 我们定义 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathbf{R})$ 为

$$Tp = \int_0^1 p(x)\,\mathrm{d}x.$$

(5) (**向前位移**) 我们定义 $T \in \mathcal{L}(\mathbf{F}^\infty, \mathbf{F}^\infty)$ 为

$$T(x_1, x_2, x_3, \cdots) = (x_2, x_3, \cdots).$$

(6) (**乘以 x^2**) 我们定义 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ 为对任意的 $x \in \mathbf{R}$,

$$(Tp)(x) = x^2 p(x).$$

(7) (从 $\mathbf{R}^3$ 到 $\mathbf{R}^2$) 我们定义 $T \in \mathcal{L}(\mathbf{R}^3, \mathbf{R}^2)$ 为

$$T(x, y, z) = (2x - y + 3z, 7x + 5y - 6z).$$

(8) (从 $\mathbf{F}^n$ 到 $\mathbf{F}^m$) 将上一例推广一下. 设 $a_{ij} \in \mathbf{F}$, 我们定义 $T \in \mathcal{L}(\mathbf{F}^n, \mathbf{F}^m)$ 为

$$T(x_1, \cdots, x_n) = (a_{11}x_1 + \cdots + a_{1n}x_n, \cdots, a_{m1}x_1 + \cdots + a_{mn}x_n).$$

事实上, 从 $\mathbf{F}^n$ 到 $\mathbf{F}^m$ 的每个线性映射都是这种形式的.

以下命题的存在性部分表明线性映射可以根据其在一个基上的取值来构造, 而唯一性部分则表明一个线性映射完全由其在基上的取值来确定.

定理 (线性映射的存在唯一性定理)：设 $v_1, \cdots, v_n$ 是 V 的基, $w_1, \cdots, w_n \in W$, 则存在唯一的 $T \in \mathcal{L}(V, W)$, 使得对每一个 $j = 1, \cdots, n$, 都有

$$Tv_j = w_j.$$

证明 (存在性) 我们定义映射 $T : V \to W$ 为

$$T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n.$$

因为 $v_1, \cdots, v_n$ 是基, 所以上述映射的定义是合理的. 容易验证 T 是保加法和保数乘的, 故 $T \in \mathcal{L}(V, W)$. 取 $c_j = 1$, 其余 c 全取为 0, 我们得到 $Tv_j = w_j$.

(唯一性) 设另有 $T' \in \mathcal{L}(V, W)$, 使得 $T'v_j = w_j$, 那么对任意的 $v = c_1v_1 + \cdots + c_nv_n \in V$, 有

$$T'v = c_1T'v_1 + \cdots + c_nT'v_n = c_1w_1 + \cdots + c_nw_n = Tv.$$

由 v 的任意性知 $T' = T$. □

注：这里我们并没有要求 W 的维数和 V 相等, 因为上述取自 W 的元素 $w_1, \cdots, w_n$ 没有任何限制.

3.2 线性映射的运算

定义 (**线性映射的加法与数乘**)：设有线性映射 $S, T \in \mathcal{L}(V, W)$ 与标量 $\lambda \in \mathbf{F}$, 我们将 S 与 T 的和 $S + T : V \to W$ 以及 λ 对 T 的数乘 $\lambda T : V \to W$ 分别定义为: 对任意的 $v \in V$,

$$(S + T)(v) = S(v) + T(v),$$
$$(\lambda T)(v) = \lambda(Tv).$$

请自行验证如上定义的 $S + T$ 与 λT 的确是线性映射.

性质 ($\mathcal{L}(V, W)$ **构成线性空间**)：$\mathcal{L}(V, W)$ 按上面定义的加法与数乘构成 $\mathbf{F}$ 上的线性空间.

一般来说, 线性空间中的元素相乘是没有意义的, 但是对于一对适当的线性映射, 却存在一种有用的乘积.

定义 (**线性映射的乘积**)：设 U, V, W 是 $\mathbf{F}$ 上的线性空间, $T \in \mathcal{L}(U, V)$, $S \in \mathcal{L}(V, W)$, 我们定义 S 与 T 的乘积 $ST : U \to W$ 为: 对任意的 $u \in U$,

$$(ST)(u) = S(Tu).$$

实际上这样定义的乘积 ST 就是通常的映射复合 $S \circ T$. 请自行验证 ST 的确是线性映射. 需要注意的是, 只有当 T 映到 S 的定义域内时, 乘积 ST 才有定义. 请读者自行验证线性映射的乘积具有下面所列的性质.

性质 (**线性映射乘积的性质**)：

(1) (结合律) $(T_1T_2)T_3 = T_1(T_2T_3)$.

(2) (单位元) $IT = TI = T$.

(3) (分配律) $(T_1 + T_2)T_3 = T_1T_3 + T_2T_3$, $T_1(T_2 + T_3) = T_1T_2 + T_1T_3$.

上述性质 (2) 的两个 I 不是同一个映射. 设 $T : V \to W$, 那么第一个 $I : W \to W$ 是 W 上的恒等映射, 第二个 $I : V \to V$ 是 V 上的恒等映射. 为了以示区别, 有时把它们分别记为 I_W 与 I_V.

注：线性映射的乘法不具有交换律. 也就是说, 即使 ST 与 TS 都有意义, 也不一定有

$$ST = TS.$$

3.3 核与像

定义 (线性映射的核)：设 $T \in \mathcal{L}(V, W)$, 我们称集合

$$\{x \in V : Tx = 0\}$$

为线性映射 T 的核, 记为 $\operatorname{Ker} T$.

有些教材把核称为零空间, 并记为 $\operatorname{null} T$.

命题 (核是子空间)：设 $T \in \mathcal{L}(V, W)$, 则 $\operatorname{Ker} T$ 是 V 的子空间.

定义 (单射)：设有集合 A, B 与映射 $T : A \to B$. 若从 $Tu = Tv$ 能够得出 $u = v$, 则称 T 是单射.

有些数学家采用术语"一对一的", 意思与单的相同. 上述定义等价于若 $u \neq v$, 则必有 $Tu \neq Tv$. 即是说, 单射将不同的输入映为不同的输出. 下面的命题说明, 对于线性映射, 我们要验证它是单的, 只需要验证 0 的原像只有 0.

命题 (线性映射为单射的判定)：设 $T \in \mathcal{L}(V, W)$, 则 T 是单射当且仅当 $\operatorname{Ker} T = \{0\}$.

定义 (线性映射的像)：设 $T \in \mathcal{L}(V, W)$, 我们称集合

$$\{Tx : x \in V\}$$

为线性映射 T 的像, 记为 $\operatorname{Im} T$.

命题 (像是子空间)：设 $T \in \mathcal{L}(V, W)$, 则 $\operatorname{Im} T$ 是 W 的子空间.

定义 (满射)：设有集合 A, B 与映射 $T : A \to B$. 若 B 中每个元素都有原像, 则称 T 是满射.

有些数学家采用"映上"这个术语, 意思与满的相同. 一个映射是不是满的, 与其陪域有关. 线性映射作为一种特殊的映射亦是如此.

例 1 定义为 $D_1 p = p'$ 的微分映射 $D_1 \in \mathcal{L}(\mathcal{P}_5(\mathbf{R}), \mathcal{P}_5(\mathbf{R}))$ 不是满的. 但定义为 $D_2 p = p'$ 的微分映射 $D_2 \in \mathcal{L}(\mathcal{P}_5(\mathbf{R}), \mathcal{P}_4(\mathbf{R}))$ 是满的.

3.4 线性映射基本定理

下面的结果非常重要, 所以它有一个响亮的名字——线性映射基本定理.

定理（**线性映射基本定理**）：设 $T \in \mathcal{L}(V, W)$，其中 V 是有限维的，那么 $\operatorname{Im} T$ 也是有限维的，并且

$$\dim V = \dim \operatorname{Ker} T + \dim \operatorname{Im} T.$$

证明 设 $u_1, \cdots, u_m$ 是 $\operatorname{Ker} T$ 的基，将其扩充为 V 的基 $u_1, \cdots, u_m, v_1, \cdots, v_n$。下面我们只需证明 $Tv_1, \cdots, Tv_n$ 是 $\operatorname{Im} T$ 的基即可。

任取 $v = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n \in V$，则

$$Tv = b_1Tv_1 + \cdots + b_nTv_n.$$

因此 $\operatorname{Im} T = \operatorname{span}\{Tv_1, \cdots, Tv_n\}$，故 $\operatorname{Im} T$ 是有限维的。下证 $Tv_1, \cdots, Tv_n$ 线性无关。设

$$c_1Tv_1 + \cdots + c_nTv_n = 0,$$

即

$$T(c_1v_1 + \cdots + c_nv_n) = 0.$$

这说明 $c_1v_1 + \cdots + c_nv_n \in \operatorname{Ker} T$，因此可设

$$c_1v_1 + \cdots + c_nv_n = d_1u_1 + \cdots + d_mu_m.$$

因为 $u_1, \cdots, u_m, v_1, \cdots, v_n$ 线性无关，故所有的 c 与 d 都是 0，因此 $Tv_1, \cdots, Tv_n$ 线性无关。□

3.5 线性映射基本定理的推论

有了线性映射基本定理，现在我们可以证明从一个有限维线性空间到更小的线性空间的线性映射不可能是单的，这里的“更小”是用维数来衡量的。

推论（**到更小维数线性空间的线性映射不是单的**）：设 V 和 W 都是有限维线性空间，且 $\dim V > \dim W$，那么从 V 到 W 的线性映射一定不是单的。

证明 因为

$$\dim \operatorname{Ker} T = \dim V - \dim \operatorname{Im} T \geqslant \dim V - \dim W > 0,$$

故 $\operatorname{Ker} T \neq \{0\}$，$T$ 不是单的。□

下面的命题表明，从有限维线性空间到一个更大的有限维线性空间的线性映射不可能是满的，这里的“更大”是用维数来衡量的。

推论（到更大维数线性空间的线性映射不是满的）：设 V 和 W 都是有限维线性空间，且 $\dim V < \dim W$，那么从 V 到 W 的线性映射一定不是满的.

证明 因为

$$\dim \operatorname{Im} T = \dim V - \dim \operatorname{Ker} T < \dim W - \dim \operatorname{Ker} T \leqslant \dim W,$$

故 T 不是满的. □

注：本节的两个推论可以表述为：设 V 和 W 均为有限维线性空间，
(1) 若 $\dim V > \dim W$，那么 $\mathcal{L}(V, W)$ 中没有单射.
(2) 若 $\dim V < \dim W$，那么 $\mathcal{L}(V, W)$ 中没有满射.

我们即将看到上述两个推论在线性方程组理论的重要应用，其想法是用线性映射来表述线性方程组的问题.

例 1 用线性映射重述齐次线性方程组是否有非零解的问题.

解 定义线性映射 $T : \mathbf{F}^n \to \mathbf{F}^m$ 为

$$T(x_1, \cdots, x_n) = (a_{11}x_1 + \cdots + a_{1n}x_n, \cdots, a_{m1}x_1 + \cdots + a_{mn}x_n).$$

那么齐次线性方程组

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = 0, \\ \cdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = 0 \end{cases} \tag{1}$$

等价于方程

$$T(x_1, \cdots, x_n) = 0. \tag{2}$$

因此方程组 (1) 有非零解当且仅当 T 不是单射. 由线性映射基本定理的推论可知当 $n > m$ 时，T 不可能是单的，此时方程组 (1) 必有非零解.

命题（齐次线性方程组有非零解的充分条件）：当变量个数多于方程个数时，齐次线性方程组必有非零解.

例 2 考虑是否可以选取常数使得非齐次线性方程组无解的问题，并用线性映射重述这一问题.

解 定义线性映射 $T : \mathbf{F}^n \to \mathbf{F}^m$ 为

$$T(x_1, \cdots, x_n) = (a_{11}x_1 + \cdots + a_{1n}x_n, \cdots, a_{m1}x_1 + \cdots + a_{mn}x_n).$$

那么非齐次线性方程组

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = c_1, \\ \cdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = c_m \end{cases} \tag{3}$$

等价于方程

$$T(x_1, \cdots, x_n) = (c_1, \cdots, c_m). \tag{4}$$

因此当即 T 不是满射时, 我们就可以取到 $(c_1, \cdots, c_m)$, 使得方程组 (3) 无解. 由线性映射基本定理的推论知当 $n < m$ 时, 我们一定可以取到适当的常数使得方程组 (3) 无解.

命题 (**非齐次线性方程组无解的充分条件**): 当方程个数多于变量个数时, 必有一组常数项使得相应的非齐次线性方程组无解.

注: 变量多于方程的齐次线性方程组和方程多于变量的非齐次性线性方程组的这些结果通常都是用高斯消元法来证明的. 我们这里采用的抽象处理方法使得证明更加简洁.

3.6 线性映射的矩阵

定义 (**线性映射的矩阵**): 设 $T \in \mathcal{L}(V, W)$, $v_1, \cdots, v_n$ 是 V 的基, $w_1, \cdots, w_m$ 是 W 的基. 我们定义 T 关于这些基的矩阵为 $\mathcal{M}(T) = (a_{ij})_{m\times n}$, 其中 a_{ij} 满足

$$Tv_j = a_{1j}w_1 + \cdots + a_{mj}w_m.$$

如果这些基不是上下文自明的, 我们将矩阵 $\mathcal{M}(T)$ 记为 $\mathcal{M}(T, (v_1, \cdots, v_n), (w_1, \cdots, w_m))$.

将上述定义用矩阵形式写出来就是 $(Tv_1, \cdots, Tv_n) = (w_1, \cdots, w_m)\mathcal{M}(T)$, 也即

$$(Tv_1, \cdots, Tv_n) = (w_1, \cdots, w_m)\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

如果 T 是从 $\mathbf{F}^n$ 到 $\mathbf{F}^m$ 的线性映射, 那么除非特别说明, 我们总设所考虑的基是标准基. 在考虑 $\mathcal{P}_m(\mathbf{F})$ 时, 除非特别声明, 我们总是使用标准基 $1, z, \cdots, z^m$.

例 1 求下列线性映射关于标准基的矩阵.

(1) $T \in \mathcal{L}(\mathbf{F}^2, \mathbf{F}^3)$ 定义为 $T(x, y) = (x + 3y, 2x + 5y, 7x + 9y)$.

(2) $D \in \mathcal{L}(\mathcal{P}_3(\mathbf{R}), \mathcal{P}_2(\mathbf{R}))$ 是微分映射 $Dp = p'$.

解 (1) $\mathcal{M}(T)=\begin{pmatrix}1&3\\2&5\\7&9\end{pmatrix}$. (2) $\mathcal{M}(D)=\begin{pmatrix}0&1&0&0\\0&0&2&0\\0&0&0&3\end{pmatrix}$.

性质 (线性映射的和与数乘的矩阵): 设 V 和 W 均为 $\mathbf{F}$ 上的有限维线性空间, 取定它们的基. 设 $S,T\in\mathcal{L}(V,W)$, $\lambda\in\mathbf{F}$, 那么
(1) $\mathcal{M}(S+T)=\mathcal{M}(S)+\mathcal{M}(T)$.
(2) $\mathcal{M}(\lambda T)=\lambda\mathcal{M}(T)$.

上述性质的证明留给读者.

性质 (线性映射的乘积的矩阵): 设 U,V,W 均为 $\mathbf{F}$ 上的有限维线性空间, 取定它们的基. 设 $T\in\mathcal{L}(U,V)$, $S\in\mathcal{L}(V,W)$, 则 $\mathcal{M}(ST)=\mathcal{M}(S)\mathcal{M}(T)$.

证明 设 $\dim U=p,\dim V=n,\dim W=m$. 又设 $\mathcal{M}(ST)=C\in\mathbf{F}^{m\times p}$, $\mathcal{M}(S)=A\in\mathbf{F}^{m\times n}$, $\mathcal{M}(T)=B\in\mathbf{F}^{n\times p}$, 则

$$((ST)u_1,\cdots,(ST)u_p)=(w_1,\cdots,w_m)C,$$
$$(Sv_1,\cdots,Sv_n)=(w_1,\cdots,w_m)A,$$
$$(Tu_1,\cdots,Tu_p)=(v_1,\cdots,v_n)B.$$

由线性映射矩阵的定义知

$$\begin{aligned}\sum_{s=1}^m c_{sk}w_s=c_{1k}w_1+\cdots+c_{mk}w_m&=(ST)u_k=S(Tu_k)\\&=S\left(\sum_{r=1}^n b_{rk}v_r\right)=\sum_{r=1}^n b_{rk}Sv_r\\&=\sum_{r=1}^n b_{rk}\left(\sum_{s=1}^m a_{sr}w_s\right)=\sum_{s=1}^m\left(\sum_{r=1}^n a_{sr}b_{rk}\right)w_s.\end{aligned}$$

因此

$$c_{sk}=\sum_{r=1}^n a_{sr}b_{rk},$$

故 $C=AB$, 即 $\mathcal{M}(ST)=\mathcal{M}(S)\mathcal{M}(T)$. □

注: 实际上我们定义矩阵乘法规则的动机就来自线性映射的乘积.

3.7 可逆线性映射与同构

定义 (可逆线性映射)：设 V, W 为 $\mathbf{F}$ 上的线性空间, $T \in \mathcal{L}(V, W)$. 若存在 $S \in \mathcal{L}(W, V)$, 使得 $ST = I_V$ 且 $TS = I_W$, 则称 T 是可逆的, 并且把 S 称为 T 的逆映射.

性质 (逆映射的唯一性)：若线性映射 T 是可逆的, 那么它的逆映射是唯一的.

证明 设 $S_1, S_2 \in \mathcal{L}(W, V)$ 均为 $T \in \mathcal{L}(V, W)$ 的逆映射, 那么对任意的 $w \in W$, 有

$$S_1 w = S_1(I_W w) = S_1(TS_2)w = (S_1T)S_2w = I_V(S_2w) = S_2w.$$

故 $S_1 = S_2$. $\square$

由逆映射的唯一性, 现在我们可以使用记号 T^{-1} 来表示 T 的逆映射了.

命题 (逆映射的判定)：线性映射 T 可逆当且仅当 T 既是单射又是满射.

证明 $(\Longrightarrow)$ 由 $T^{-1}T$ 是单射不难验证 T 是单射, 由 TT^{-1} 是满射不难验证 T 是满射.

$(\Longleftarrow)$ 设 $T: V \to W$. 因为 T 满, 故对任意的 $w \in W$, 存在 $v \in V$, 使得 $Tv = w$. 因为 T 是单射, 故这样的 v 是唯一的. 由此我们可以定义一个满足 $Sw = v$ 的映射 $S: W \to V$, 使得 $Tv = w$. 由此易知 $TS = I_W$. 又因为对任意的 $v \in V$, 有

$$(ST)v = S(Tv) = Sw = v,$$

故 $ST = I_V$. 下证 S 为线性映射. 由于 T 为单射且

$$T(S(w_1 + w_2)) = (TS)(w_1 + w_2) = (TS)w_1 + (TS)w_2 = T(Sw_1) + T(Sw_2) = T(Sw_1 + Sw_2),$$

故 $S(w_1 + w_2) = Sw_1 + Sw_2$. 同理可得 $S(\lambda w) = \lambda Sw$, 故 S 为线性映射, 因此 T 可逆. $\square$

注：实际上映射可逆的充要条件就是该映射为双射. 上述结论告诉我们, 将映射加强为线性映射, 充要条件不变. 这是因为一旦线性映射作为映射可逆了, 那么它的逆映射一定是线性映射.

定义 (同构)：我们称可逆的线性映射为同构 (isomorphism). 如果两个线性空间之间存在同构, 那么我们就称这两个线性空间是同构的 (isomorphic).

命题 (同构等价于维数相同)：$\mathbf{F}$ 上的两个有限维线性空间同构当且仅当它们的维数相同.

证明 $(\Longrightarrow)$ 设 T 为线性空间 V 到 W 的同构, 那么 T 既是单的又是满的. 由线性映射基本定理的推论可知此时 $\dim V \leqslant \dim W$ 且 $\dim V \geqslant \dim W$, 因此 $\dim V = \dim W$.

$(\Longleftarrow)$ 作映射 $T : V \to W$ 为

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n v_n.$$

易知 T 为线性映射且既单又满, 故 T 为同构, 从而 V 与 W 同构. □

以上的定理表明 $\mathbf{F}$ 上的 n 维线性空间均同构于 $\mathbf{F}^n$.

设 V 和 W 是 $\mathbf{F}$ 上的两个有限维线性空间, 取定 V 的基 $v_1, \cdots, v_n$ 与 W 的基 $w_1, \cdots, w_m$, 则

$$\begin{aligned} \mathcal{M} : \mathcal{L}(V, M) &\longrightarrow \mathbf{F}^{m\times n} \\ T \quad &\longmapsto \mathcal{M}(T) \end{aligned}$$

就是从 $\mathcal{L}(V, M)$ 到 $\mathbf{F}^{m\times n}$ 的映射了. 又因为 $\mathcal{M}(S+T) = \mathcal{M}(S) + \mathcal{M}(T)$, $\mathcal{M}(\lambda T) = \lambda\mathcal{M}(T)$, 故 $\mathcal{M}$ 还是一个线性映射. 下面我们将会证明, 这个线性映射还是可逆的. 即是说, $\mathcal{L}(V, M)$ 同构于 $\mathbf{F}^{m\times n}$.

引理 ($\mathcal{L}(V, M)$ **与** $\mathbf{F}^{m\times n}$ **同构**): 取定取定 V 的基 $v_1, \cdots, v_n$ 与 W 的基 $w_1, \cdots, w_m$, 则 $\mathcal{L}(V, W)$ 与 $\mathbf{F}^{m\times n}$ 是同构的, 且 $\mathcal{M}$ 就是它们之间的一个同构.

证明 只需证明 $\mathcal{M}$ 既是单射又是满射. 设 $\mathcal{M}(T) = A = 0$, 则

$$Tv_j = a_{1j} w_1 + \cdots + a_{mj} w_m = 0.$$

由线性映射的存在唯一性知 $T = 0$, 故 $\mathcal{M}$ 是单射. 任取 $A \in \mathbf{F}^{m\times n}$, 令

$$Tv_j = a_{1j} w_1 + \cdots + a_{mj} w_m,$$

那么 $\mathcal{M}(T) = A$, 故 $\mathcal{M}$ 是满射. 综上, $\mathcal{M}$ 是同构. □

现在我们可以确定从一个线性空间到另一个线性空间的所有线性映射构成的空间的维数了.

定理 (线性映射空间的维数): 设 V, W 为有限维线性空间, 则 $\mathcal{L}(V, W)$ 也是有限维线性空间, 且

$$\dim \mathcal{L}(V, W) = \dim V \cdot \dim W.$$

证明 由前面两个命题和引理的结论即得. □

3.8 将线性映射视为矩阵乘

定义 (向量的矩阵)：取定线性空间 V 的基 $v_1, \cdots, v_n$, 我们称 $n \times 1$ 矩阵

$$\mathcal{M}(v) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

为向量 $v = c_1 v_1 + \cdots + c_n v_n$ 的矩阵.

向量 v 的矩阵 $\mathcal{M}(v)$ 与基的选取有关, 但基的选取通常是上下文自明的, 所以就不把基包括在记号里面了.

例 1

(1) 多项式 $p = 2 - 7x + 5x^3 \in \mathcal{P}_3(\mathbf{R})$ 在标准基下的矩阵为

$$\mathcal{M}(p) = \begin{pmatrix} 2 \\ -7 \\ 0 \\ 5 \end{pmatrix}.$$

(2) 向量 $x = (x_1, \cdots, x_n) \in \mathbf{F}^n$ 在标准基下的矩阵为

$$\mathcal{M}(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

一旦我们把 n 为线性空间 V 的基取定, 那么 V 中的元素就通过映射

$$\begin{aligned} \mathcal{M} : V &\longrightarrow \mathbf{F}^{n \times 1} \\ v &\longmapsto \mathcal{M}(v) \end{aligned}$$

建立起与 $\mathbf{F}^{n \times 1}$ 的同构了. 也就是说, 基取定以后, 我们可以把 n 维线性空间 V 中的向量与 n 维列向量等同起来.

性质：设 $T \in \mathcal{L}(V, W)$, $v_1, \cdots, v_n$ 是 V 的基, $w_1, \cdots, w_m$ 是 W 的基, 那么

$$\mathcal{M}(Tv_j) = \mathcal{M}(T)_j.$$

这里 $\mathcal{M}(T)_j$ 是矩阵 $\mathcal{M}(T)$ 的第 j 列.

证明 由线性映射的矩阵 ($\mathcal{M}(T)$) 与向量的矩阵 ($\mathcal{M}(w)$ 的定义即得. □

下面的命题表明线性映射的矩阵、向量的矩阵与矩阵乘法是如何联系到一起的.

命题 (线性映射视为矩阵乘)：设 $T \in \mathcal{L}(V, W)$, $v_1, \cdots, v_n$ 是 V 的基, $w_1, \cdots, w_m$ 是 W 的基, $v \in V$, 则

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v).$$

证明 设 $v = c_1 v_1 + \cdots + c_n v_n$, 则

$$\begin{aligned}
\mathcal{M}(Tv) &= \mathcal{M}(c_1 T v_1 + \cdots + c_n T v_n) \\
&= c_1 \mathcal{M}(T v_1) + \cdots + c_n \mathcal{M}(T v_n) \\
&= c_1 \mathcal{M}(T)_1 + \cdots + c_n \mathcal{M}(T)_n \\
&= \mathcal{M}(T)\mathcal{M}(v).
\end{aligned}$$

这里第二个等号利用的是向量的矩阵的线性性, 第三个等号利用的是前面的性质, 最后一个等号利用的是 $m \times n$ 矩阵与 $n \times 1$ 矩阵的乘法等于列向量的线性组合. □

每个矩阵 $A \in \mathbf{F}^{m \times n}$ 都诱导了一个从 $\mathbf{F}^{n \times 1}$ 到 $\mathbf{F}^{m \times 1}$ 的线性映射

$$\begin{aligned}
T : \mathbf{F}^{n \times 1} &\longrightarrow \mathbf{F}^{m \times 1} \\
x \quad &\longmapsto Ax
\end{aligned}$$

反过来, 上述命题说明取定基以后, 我们可以 (通过同构) 将线性映射当做矩阵乘映射. 即是说, 我们可以将一个线性映射视为 $\mathbf{F}^{n \times 1}$ 上某个矩阵 A 的矩阵乘法. 具体而言, 若将 v 等同于 $\mathcal{M}(v)$, 则我们可将 Tv 等同于 $\mathcal{M}(T)\mathcal{M}(v)$.

注：注意到 $\mathcal{M}(T)$ 依赖于基的选取, 后续章节很多最重要结果的主旨之一就是探究如何选取基使得矩阵 $\mathcal{M}(T)$ 尽可能简单.

3.9 线性变换

在线性代数中, 也在我们接下来的学习中, 最重要最深刻的内容就是研究线性变换.

定义 (线性变换)：我们把线性空间到自身的线性映射称为线性变换. $\mathcal{L}(V)$ 表示 V 上的线性变换全体所成集合, 即 $\mathcal{L}(V) = \mathcal{L}(V, V)$.

线性映射既单又满则可逆, 对于一个线性变换, 我们想知道仅由单性或满性之一能否推出可逆? 对于无限维线性空间, 这两个条件中的任何一个都不能单独推出可逆性. 请看下面的例子.

例 1 (**无限维线性空间单射与满射没有必然联系**)

(1) $\mathcal{P}(\mathbf{R})$ 上乘以 x^2 的线性变换是单的, 但不是满的. $((Tp)(x)=x^2p(x))$

(2) $\mathbf{F}^\infty$ 上的向前位移变换是满的, 但不是单的. $(T(x_1,x_2,x_3,\cdots)=(x_2,x_3,\cdots))$

与无限维的情形不同, 对于有限维线性空间, 从单性或满性中的一个就能得到另一个. 即是说, 有限维线性空间上的线性变换是单的等价于是满的.

性质 (**有限维线性空间上单射与满射等价**): 设 V 为有限维线性空间, $T\in\mathcal{L}(V)$, 则下面各条是等价的:

(1) T 是可逆的.

(2) T 是单的.

(3) T 是满的.

证明 我们按照 (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (1) 的顺序证明.

(1) $\Longrightarrow$ (2) 显然.

(2) $\Longrightarrow$ (3) 因为 T 是单的, 故 $\operatorname{Ker}T=\{0\}$. 由线性映射基本定理知 $\dim V=\dim\operatorname{Ker}T+\dim\operatorname{Im}T=\dim\operatorname{Im}T$, 从而 T 为满射.

(3) $\Longrightarrow$ (1) $\dim\operatorname{Im}T=\dim V$, 故由线性映射基本定理知 $\dim\operatorname{Ker}T=0$, 故 T 是单的, 又因为 T 是满的, 故 T 可逆. $\square$

例 2 证明对每个对多项式 $q\in\mathcal{P}(\mathbf{R})$, 都存在一个多项式 $p\in\mathcal{P}(\mathbf{R})$, 使得

$$((x^2+5x+7)p)''=q.$$

证明 $\mathcal{P}(\mathbf{R})$ 是无限维线性空间, 不能直接应用上述精妙的结论, 但是注意到每个非零多项式都有次数 m, 因此我们仅考虑 $\mathcal{P}_m(\mathbf{R})$ 即可.

设 $q\in\mathcal{P}_m(\mathbf{R})$, 定义线性变换 $T:\mathcal{P}_m(\mathbf{R})\to\mathcal{P}_m(\mathbf{R})$ 为 $Tp=((x^2+5x+7)p)''$(注意到用 x^2+5x+7 去乘多项式 p 会使得 p 的次数增加 2, 做两次微分次数又降低 2). 设 $Tp=0$, 那么

$$((x^2+5x+7)p)''=0,$$

从而

$$(x^2+5x+7)p=ax+b,\quad \text{其中 } a,b\in\mathbf{R}.$$

因此 $p=0$, 即 $\operatorname{Ker}T=\{0\}$, T 为单射. 因为 $\mathcal{P}_m(\mathbf{R})$ 是有限维线性空间, 故 T 是满射, 从而对任意

的 $q \in \mathcal{P}_m(\mathbf{R})$, 存在 $p \in \mathcal{P}_m(\mathbf{R})$, 使得

$$Tp = ((x^2 + 5x + 7)p)'' = q.$$

3.10 习题

1. (**像线性无关, 则原像线性无关**) 设 $T \in \mathcal{L}(V, W)$, $v_1, \cdots, v_m$ 是 V 中的向量, $Tv_1, \cdots, Tv_m$ 在 W 中线性无关. 证明 $v_1, \cdots, v_m$ 线性无关.

2. (**线性映射的扩张**) 设 V 是有限维的, 证明 V 的子空间上的线性映射可以扩张成 V 上的线性映射. 也就是说, 若 U 是 V 的子空间, $S \in \mathcal{L}(U, W)$, 那么存在 $T \in \mathcal{L}(V, W)$, 使得对任意的 $n \in U$, 都有 $Tu = Su$.

3. 设 V 是有限维的且 $\dim V \geqslant 2$, 证明存在 $S, T \in \mathcal{L}(V, V)$, 使得 $ST \neq TS$.

4. 设 φ_1, φ_2 都是 V 到 $\mathbf{F}$ 的线性映射, 且有相同的零空间. 证明存在常数 $c \in \mathbf{F}$, 使得 $\varphi_1 = c\varphi_2$.

5. 设 n 为正整数, $(a_{ij}) \in \mathbf{F}^{n \times n}$, 证明下面两个陈述等价:

(1) 下列齐次方程组只有零解.

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = 0, \\ \qquad\qquad \cdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n = 0. \end{cases}$$

(2) 对任意的 $c_1, \cdots, c_n \in \mathbf{F}$, 下列方程组都有解.

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = c_1, \\ \qquad\qquad \cdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n = c_n. \end{cases}$$

6. 设 V 是有限维的, $S, T \in \mathcal{L}(V)$. 证明: $ST = I$ 当且仅当 $TS = I$.

4 积空间, 商空间, 对偶空间

本章我们将会利用已知的线性空间构造许多新的线性空间, 而这些构造的手法在数学的各个领域都是很常见的, 你应当掌握它.

4.1 积空间

定义 (线性空间的乘积)：设 $V_1, \cdots, V_m$ 均为 $\mathbf{F}$ 上的线性空间, 定义它们的乘积为

$$V_1 \times \cdots \times V_m = \{(v_1, \cdots, v_m) : v_1 \in V_1, \cdots, v_m \in V_m\}.$$

对于上面定义的 $V_1 \times \cdots \times V_m$, 我们规定其上的加法与数乘为

$$(u_1, \cdots, u_m) + (v_1, \cdots, v_m) = (u_1 + v_1, \cdots, u_m + v_m),$$
$$\lambda(v_1, \cdots, v_m) = (\lambda v_1, \cdots, \lambda v_m).$$

性质 (线性空间的乘积也是线性空间)：设 $V_1, \cdots, V_m$ 均为 $\mathbf{F}$ 上的线性空间, 则它们的乘积 $V_1 \times \cdots \times V_m$ 也是 $\mathbf{F}$ 上的线性空间.

证明 请读者自行验证. □

例 1 $\mathbf{R}^2 \times \mathbf{R}^3$ 等于 $\mathbf{R}^5$ 吗? $\mathbf{R}^2 \times \mathbf{R}^3$ 同构于 $\mathbf{R}^5$ 吗?

解 $\mathbf{R}^2 \times \mathbf{R}^3$ 中的元素为二元组 $((x_1, x_2), (x_3, x_4, x_5))$, 而 $\mathbf{R}^5$ 中的元素 $(x_1, x_2, x_3, x_4, x_5)$ 为五元组, 故 $\mathbf{R}^2 \times \mathbf{R}^3$ 不等于 $\mathbf{R}^5$. 但若定义 $T(((x_1, x_2), (x_3, x_4, x_5))) = (x_1, x_2, x_3, x_4, x_5)$, 则 T 是 $\mathbf{R}^2 \times \mathbf{R}^3$ 与 $\mathbf{R}^5$ 之间的同构, 即 $\mathbf{R}^2 \times \mathbf{R}^3$ 与 $\mathbf{R}^5$ 是同构的.

例 2 求 $\mathcal{P}_2(\mathbf{R}) \times \mathbf{R}^2$ 的一个基.

解 考虑向量组

$$(1, (0, 0)), (x, (0, 0)), (x^2, (0, 0)), (0, (1, 0)), (0, (0, 1)).$$

上述向量组线性无关, 且张成了 $\mathcal{P}_2(\mathbf{R}) \times \mathbf{R}^2$, 故它们是 $\mathcal{P}_2(\mathbf{R}) \times \mathbf{R}^2$ 的基.

定理 (积的维数等于维数之和)：设 $V_1, \cdots, V_m$ 为 $\mathbf{F}$ 上的有限维线性空间, 则 $V_1 \times \cdots \times V_m$ 也是有限维的, 且

$$\dim(V_1 \times \cdots \times V_m) = \dim V_1 + \cdots + \dim V_m.$$

证明 参考例 2 中基的构造即可. □

设 $U_1, \cdots, U_m$ 为 V 的子空间. 对于乘积 $U_1 \times \cdots \times U_m$ 与和 $U_1 + \cdots + U_m$, 我们可以构造映射

$$\Gamma : U_1 \times \cdots \times U_m \longrightarrow U_1 + \cdots + U_m$$
$$(u_1, \cdots, u_m) \longmapsto u_1 + \cdots + u_m$$

易知 $U_1 + \cdots + U_m$ 是直和当且仅当 Γ 是单射.

命题（**子空间的和为直和当且仅当维数相加**）：设 $U_1, \cdots, U_m$ 为 V 的有限维子空间，则 $U_1 + \cdots + U_m$ 是直和当且仅当

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m.$$

证明 由前面的讨论知 $U_1 + \cdots + U_m$ 是直和当且仅当 Γ 是单射. 注意到 Γ 是满射, 由线性映射基本定理知 $U_1 + \cdots + U_m$ 是直和当且仅当

$$\dim(U_1 \times \cdots \times U_m) = \dim(U_1 + \cdots + U_m).$$

由于 $\dim(U_1 \times \cdots \times U_m) = \dim U_1 + \cdots + \dim U_m$, 故 $U_1 + \cdots + U_m$ 是直和当且仅当

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m.$$

结论得证. □

4.2 商空间

我们先定义向量与子空间的和, 以便引入商空间.

定义（**陪集**）：设 U 是 V 的子空间, $v \in V$, 我们称集合

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

为向量 v 关于子空间 U 的陪集, 它是 V 的子集.

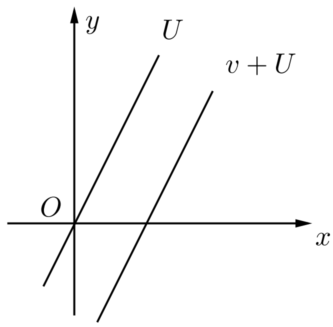
例 1 过原点的斜率为 2 的直线 $U = \{(x, 2x) \in \mathbf{R}^2 : x \in \mathbf{R}\}$ 是 $V = \mathbf{R}^2$ 的子空间, $(17, 20) \in \mathbf{R}^2$, 那么 $v = (17, 20)$ 关于 U 的陪集

$$v + U = \{(17, 20) + (x, 2x) : x \in \mathbf{R}\}$$

就是 $\mathbf{R}^2$ 中过点 $(17, 20)$ 且斜率为 2 的直线.

设 U 为 V 的子空间, $v \in V$, 我们称陪集 $v + U$ 平行于子空间 U.

例 2 $U = \{(x, y, 0) \in \mathbf{R}^3 : x, y \in \mathbf{R}\}$ 是 $\mathbf{R}^3$ 的子空间, 则 $\mathbf{R}^3$ 的平行于 U 的陪集就是 $\mathbf{R}^3$ 中通常意义下平行于平面 U 的那些平面.



性质 (陪集相等的等价条件)：设 U 是 V 的子空间, $v_1, v_2 \in V$, 则下面各条是等价的:

(1) $v_1 - v_2 \in U$.

(2) $v_1 + U = v_2 + U$.

(3) $(v_1 + U) \cap (v_2 + U) \neq \varnothing$.

证明 (1) $\Longrightarrow$ (2) 证明 $v_1 + U$ 与 $v_2 + U$ 互相包含即可. 任取 $v_2 + u \in v_2 + U$, 则

$$v_2 + u = v_1 + u + (v_2 - v_1) \in v_1 + U.$$

故 $v_2 + U \subseteq v_1 + U$. 同理 $v_1 + U \subseteq v_2 + U$, 故 $v_1 + U = v_2 + U$.

(2) $\Longrightarrow$ (3) 显然.

(3) $\Longrightarrow$ (1) 任取 $w \in (v_1 + U) \cap (v_2 + U)$, 则 $w \in v_1 + U$ 且 $w \in v_2 + U$. 因此存在 $u_1, u_2 \in U$, 使得 $w = v_1 + u_1, w = v_2 + u_2$, 两式相减得 $(v_1 - v_2) + (u_1 - u_2) = 0$, 从而 $v_1 - v_2 = u_2 - u_1 \in U$. □

注：上述性质说明两个陪集要么不相交, 要么相等, 因此陪集可以将整个线性空间的元素划分为互不相交的若干份.

定义 (商空间)：设 U 是 V 的子空间, 我们称 V 中向量关于子空间 U 的全体陪集所成集合为 V 关于子空间 U 的商空间, 记作 V/U, 即

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

例 3 (商空间)

(1) 设 $U = \{(x, 2x) \in \mathbf{R}^2 : x \in \mathbf{R}\}$, 则商空间 $\mathbf{R}^2/U$ 就是 $\mathbf{R}^2$ 中斜率为 2 的全体直线的集合.

(2) 设 U 是 $\mathbf{R}^3$ 中包含原点的直线, 则 $\mathbf{R}^3/U$ 就是 $\mathbf{R}^3$ 中平行于直线 U 的全体直线所成集合.

(3) 设 U 是 $\mathbf{R}^3$ 中包含原点的平面, 则 $\mathbf{R}^3/U$ 就是 $\mathbf{R}^3$ 中平行于平面 U 的全体平面所成集合.

我们的目标是使 V/U 成为线性空间, 因此需要在 V/U 上定义加法与数乘.

定义 (**陪集的加法与数乘**)：设 U 是 V 的子空间，对任意的 $v,w\in V$ 与 $\lambda\in\mathbf{F}$，我们定义陪集的加法与数乘为

$$(v+U)+(w+U)=(v+w)+U,$$
$$\lambda(v+U)=\lambda v+U.$$

下面我们证明这样定义的加法与数乘是有意义的，即是说，运算的结果不会因为陪集代表元的不同而改变.

证明 设 $v_1+U=v_2+U$, $w_1+U=w_2+U$, 则 $v_1-v_2\in U, w_1-w_2\in U$, 从而

$$(v_1-v_2)+(w_1-w_2)=(v_1+w_1)-(v_2+w_2)\in U,$$

因此

$$(v_1+w_1)+U=(v_2+w_2)+U.$$

同理可证明数乘的定义是合理的. □

性质 (**商空间是线性空间**)：设 U 是 V 的子空间，则商空间 V/U 是线性空间.

证明 请读者自行验证. □

容易知道商空间 V/U 的零元是 $0+U$(等于 U), $v+U$ 的负元是 $(-v)+U$.

设 U 是 V 的子空间，那么从 V 到商空间 V/U 有一个非常自然的映射

$$\begin{aligned}\pi: V&\longrightarrow V/U\\ v&\longmapsto v+U\end{aligned}$$

我们称这样的映射 π 为商映射，它是一个满射. 请读者自行验证商映射是一个线性映射.

定理 (**商空间的维数**)：设 V 是有限维线性空间，U 是 V 的子空间，则

$$\dim(V/U)=\dim V-\dim U.$$

证明 (**方法一：基扩充**) 设 $u_1,\cdots,u_m$ 为 U 的基，将其扩充为 V 的基 $u_1,\cdots,u_m,v_1,\cdots,v_n$. 下面我们证明 $v_1+U,\cdots,v_n+U$ 是 V/U 的基. 任取 $v=a_1u_1+\cdots+a_mu_m+b_1v_1+\cdots+b_nv_n\in V$，则

$$v+U=b_1(v_1+U)+\cdots+b_n(v_n+U).$$

因此 $V/U=\operatorname{span}\{v_1+U,\cdots,v_n+U\}$. 下面只需证明 $v_1+U,\cdots,v_n+U$ 线性无关. 设

$$c_1(v_1+U)+\cdots+c_n(v_n+U)=0+U,$$

那么

$$c_1v_1+\cdots+c_nv_n\in U,$$

故可设

$$c_1v_1+\cdots+c_nv_n=d_1u_1+\cdots+d_mu_m.$$

由于 $u_1,\cdots,u_m,v_1,\cdots,v_n$ 线性无关, 故所有 c 和 d 均为 0, $v_1+U,\cdots,v_n+U$ 是 V/U 的基. □

(**方法二: 商映射**) 由于商映射 π 满足 $\operatorname{Ker}\pi=U$, $\operatorname{Im}\pi=V/U$, 故由线性映射基本定理知

$$\dim V=\dim U+\dim(V/U).$$

结论得证. □

每一个 $T\in\mathcal{L}(V,W)$ 都诱导了一个从 $V/\operatorname{Ker}T$ 到 W 的映射

$$\begin{aligned}\tilde{T}:V/\operatorname{Ker}T&\longrightarrow W\\ v+\operatorname{Ker}T&\longmapsto Tv\end{aligned}$$

请读者自行验证这是一个定义合理的线性映射.

性质：设 $T\in\mathcal{L}(V,W)$, $\tilde{T}$ 按如上方式定义, 则

(1) $\operatorname{Ker}\tilde{T}=\{\operatorname{Ker}T\}$, 即 $\tilde{T}$ 是单的 (因为 $\operatorname{Ker}T$ 是 $V/\operatorname{Ker}T$ 的零元).

(2) $\operatorname{Im}\tilde{T}=\operatorname{Im}T$.

(3) $V/\operatorname{Ker}T\cong\operatorname{Im}T$.

4.3 对偶空间, 对偶基, 对偶映射

映到 $\mathbf{F}$ 的线性映射在线性代数中扮演了重要的角色, 故它们有一个特别的名字——线性泛函.

定义 (线性泛函)：设 V 是 $\mathbf{F}$ 上的线性空间, 称从 V 到 $\mathbf{F}$ 的线性映射, 也即 $\mathcal{L}(V,\mathbf{F})$ 中的元素为线性泛函 (linear functional).

例 1 (**线性泛函**)

(1) 定义 $\varphi:\mathbf{R}^3\to\mathbf{R}$ 为 $\varphi(x,y,z)=4x-5y+2z$, 则 φ 是 $\mathbf{R}^3$ 上的线性泛函.

(2) 取定 $(c_1,\cdots,c_n)\in\mathbf{F}^n$, 定义 $\varphi:\mathbf{F}^n\to\mathbf{F}$ 为 $\varphi(x_1,\cdots,x_n)=c_1x_1+\cdots+c_nx_n$, 则 φ 是 $\mathbf{F}^n$ 上的线性泛函.

(3) 定义 $\varphi:\mathcal{P}(\mathbf{R})\to\mathbf{R}$ 为 $\varphi(p)=3p''(5)+7p(4)$, 则 φ 为 $\mathcal{P}(\mathbf{R})$ 上的线性泛函.

(4) 定义 $\varphi:\mathcal{P}(\mathbf{R})\to\mathbf{R}$ 为 $\varphi(p)=\displaystyle\int_0^1 p(x)\,\mathrm{d}x$, 则 φ 为 $\mathcal{P}(\mathbf{R})$ 上的线性泛函.

定义 (对偶空间)：设 V 为 $\mathbf{F}$ 上的线性空间, 我们称线性空间 $\mathcal{L}(V,\mathbf{F})$ 为 V 的对偶空间 (dual space), 记作 V'.

性质 (对偶空间的维数)：设 V 是有限维线性空间, V' 为 V 的对偶空间, 则

$$\dim V' = \dim V.$$

证明 由 $\dim V' = \dim \mathcal{L}(V,\mathbf{F}) = \dim V \cdot \dim \mathbf{F}$ 即得. □

根据线性映射的存在唯一性定理, 下面的定义是合理的.

定义 (对偶基)：设 $v_1,\cdots,v_n$ 是 V 的基, 定义 V' 中的元素

$$\varphi_i(v_j) = \begin{cases} 1, & j = i, \\ 0, & j \neq i. \end{cases}$$

我们把 $\varphi_1,\cdots,\varphi_n$ 称为 $v_1,\cdots,v_n$ 的对偶基 (dual basis).

例 2 求 $\mathbf{F}^n$ 的标准基的对偶基.

解 定义 $\varphi_i : \mathbf{F}^n \to \mathbf{F}$ 为

$$\varphi_i(x_1,\cdots,x_n) = x_i,$$

那么

$$\varphi_i(e_j) = \begin{cases} 1, & j = i, \\ 0, & j \neq i. \end{cases}$$

因此如上定义的 $\varphi_1,\cdots,\varphi_n$ 为标准基 $e_1,\cdots,e_n$ 的对偶基.

命题 (对偶基的作用)：设 $v_1,\cdots,v_n$ 是 V 的基, $\varphi_1,\cdots,\varphi_n$ 是 $v_1,\cdots,v_n$ 的对偶基, 则对任意的 $v \in V$, 有

$$v = \varphi_1(v)v_1 + \cdots + \varphi_n(v)v_n.$$

下面的命题说明对偶基的确是对偶空间的基, 因此它的名字“对偶基”是名正言顺的.

性质 (对偶基是对偶空间的基)：设 $v_1,\cdots,v_n$ 是 V 的基, $\varphi_1,\cdots,\varphi_n$ 是 $v_1,\cdots,v_n$ 的对偶基, 那么 $\varphi_1,\cdots,\varphi_n$ 是对偶空间 V' 的基.

证明 因为 $\varphi_1,\cdots,\varphi_n$ 的长度恰好为 $\dim V'$, 因此我们只需证明 $\varphi_1,\cdots,\varphi_n$ 线性无关. 设

$$a_1\varphi_1 + \cdots + a_n\varphi_n = 0,$$

则

$$(a_1\varphi_1+\cdots+a_n\varphi_n)(v_j)=0.$$

由于 $\varphi_j(v_j)=1$, 故 $a_j=0$, $j=1,\cdots,n$. 因此 $\varphi_1,\cdots,\varphi_n$ 线性无关, 是 V' 的基. □

定义 (对偶映射): 设 $T\in\mathcal{L}(V,W)$, 定义 $T'\in\mathcal{L}(W',V')$ 为

$$\begin{aligned} T': W' &\longrightarrow V' \\ \varphi &\longmapsto \varphi T \end{aligned}$$

我们称 T' 为 T 的对偶映射 (dual map).

容易验证 $T':W'\to V'$ 的确是线性映射. 在下面的例题中 $'$ 有两种完全不相关的用法: D' 表示 D 的对偶映射, 而 p' 表示多项式 p 的导数.

例 3 设 $D:\mathcal{P}(\mathbf{R})\to\mathcal{P}(\mathbf{R})$ 为微分映射 $Dp=p'$.

(1) 设线性泛函 $\varphi:\mathcal{P}(\mathbf{R})\to\mathbf{R}$ 定义为 $\varphi(p)=p(3)$, 那么线性泛函 $D'(\varphi):\mathcal{P}(\mathbf{R})\to\mathbf{R}$ 为

$$(D'(\varphi))(p)=(\varphi D)(p)=\varphi(Dp)=\varphi(p')=p'(3).$$

(2) 设线性泛函 $\varphi:\mathcal{P}(\mathbf{R})\to\mathbf{R}$ 定义为 $\varphi(p)=\displaystyle\int_0^1 p(x)\,\mathrm{d}x$, 那么线性泛函 $D'(\varphi):\mathcal{P}(\mathbf{R})\to\mathbf{R}$ 为

$$(D'(\varphi))(p)=(\varphi D)(p)=\varphi(Dp)=\varphi(p')=\int_0^1 p'(x)\,\mathrm{d}x=p(1)-p(0).$$

下面结果的前两条说明将 T 变为 T' 的映射是从 $\mathcal{L}(V,W)$ 到 $\mathcal{L}(W',V')$ 的线性映射.

性质 (对偶映射的线性性与乘积): 设 U,V,W 均为 $\mathbf{F}$ 上的线性空间, 则

(1) 对任意的 $S,T\in\mathcal{L}(V,W)$, 有 $(S+T)'=S'+T'$.

(2) 对任意的 $T\in\mathcal{L}(V,W)$ 与 $\lambda\in\mathbf{F}$, 有 $(\lambda T)'=\lambda T'$.

(3) 对任意的 $T\in\mathcal{L}(U,V)$ 与 $S\in\mathcal{L}(V,W)$, 有 $(ST)'=T'S'$.

证明 由下面三行中 φ 的任意性可知结论成立.

(1) $(S+T)'(\varphi)=\varphi(S+T)=\varphi S+\varphi T=S'(\varphi)+T'(\varphi)$.

(2) $(\lambda T)'(\varphi)=\varphi(\lambda T)=\lambda(\varphi T)=\lambda T'(\varphi)$.

(3) $(ST)'(\varphi)=\varphi(ST)=(\varphi S)T=T'(\varphi S)=T'(S'(\varphi))=(T'S')(\varphi)$. □

注：有些书用 V^* 和 T^* 来表示对偶空间 V' 与对偶映射 T', 但是我们将在学习后面内积空间的线性映射时用 T^* 表示伴随, 因此这里我们不使用记号 T^* 表示对偶.

4.4 零化子

我们想要利用 $\operatorname{Im} T$ 与 $\operatorname{Ker} T$ 去描述 $\operatorname{Ker} T'$ 与 $\operatorname{Im} T'$. 本节的主要结果如下: 对于有限维线性空间 V, W 及线性映射 $T \in \mathcal{L}(V, W)$, 有

(1) $\operatorname{Ker} T' = (\operatorname{Im} T)^0$.

(2) $\operatorname{Im} T' = (\operatorname{Ker} T)^0$.

(3) T 是满的当且仅当 T' 是单的.

(4) T 是单的当且仅当 T' 是满的.

定义 (零化子)：设 U 是 V 的子集, 我们称

$$U^0 = \{\varphi \in V' : \varphi|_U = 0\}$$

为子集 U 的零化子 (annihilator).

定义中的 $\varphi|_U = 0$ 就是对任意的 $u \in U$, $\varphi(u) = 0$. 另外, φ 是对偶空间 V' 中的元素, 因此 U^0 其实是与 V 有关的, 所以记号 U_V^0 更准确, 但是包含 U 的线性空间在上下文中往往是明确的, 因此我们采用更简单的记号 U^0.

性质 (零化子是子空间)：设 U 是 V 的子集, 那么 U^0 是 V' 的子空间.

证明 首先 $0 \in U^0$, 故 U^0 非空. 其次 $(\varphi + \psi)(u) = \varphi(u) + \psi(u) = 0 + 0 = 0$, 故 $\varphi + \psi \in U^0$. 类似地, U^0 关于数乘封闭. □

例 1 设 $\varphi_1, \cdots, \varphi_5$ 是 $\mathbf{R}^5$ 的标准基 $e_1, \cdots, e_5$ 的对偶基, 设

$$U = \operatorname{span}\{e_1, e_2\},$$

证明 $U^0 = \operatorname{span}\{\varphi_3, \varphi_4, \varphi_5\}$.

证明 任取 $\varphi \in U^0$, 则对任意的 $u = c_1 e_1 + c_2 e_2 \in U$,

$$\varphi(c_1 e_1 + c_2 e_2) = c_1 \varphi(e_1) + c_2 \varphi(e_2) = 0.$$

由 c_1, c_2 的任意性知 $\varphi(e_1)=0$, $\varphi(e_2)=0$. 设 $\varphi=d_1\varphi_1+d_2\varphi_2+d_3\varphi_3+d_4\varphi_4+d_5\varphi_5$, 则

$$(d_1\varphi_1+d_2\varphi_2+d_3\varphi_3+d_4\varphi_4+d_5\varphi_5)(e_1)=0,$$
$$(d_1\varphi_1+d_2\varphi_2+d_3\varphi_3+d_4\varphi_4+d_5\varphi_5)(e_2)=0.$$

因此 $d_1=d_2=0$, 故

$$\varphi=d_3\varphi_3+d_4\varphi_4+d_5\varphi_5\subseteq\operatorname{span}\{\varphi_3,\varphi_4,\varphi_5\}.$$

此外, 不难验证 $\varphi_3,\varphi_4,\varphi_5\in U^0$, 故 $\operatorname{span}\{\varphi_3,\varphi_4,\varphi_5\}\subseteq U^0$. 综上, $U^0=\operatorname{span}\{\varphi_3,\varphi_4,\varphi_5\}$. □

下面的命题表明 $\dim U^0$ 是 $\dim V$ 与 $\dim U$ 的差. 例如, 若 U 是 $\mathbf{R}^5$ 的二维子空间, 那么 U^0 是 $(\mathbf{R}^5)'$ 的三维子空间, 这就是例 1 的结果.

定理 (**零化子的维数**)：设 V 是有限维线性空间, U 是 V 的子空间, 那么

$$\dim U^0=\dim V-\dim U.$$

证明 我们介绍两种证明方法, 其中第一种方法更容易想到, 我们只写出思路, 第二种方法更为简洁.

(**方法一: 基扩张**) 设 $u_1,\cdots,u_m$ 是 U 的基, 将其扩张为 V 的基 $u_1,\cdots,u_m,v_1,\cdots,v_n$. 设 $u_1,\cdots,u_m,v_1,\cdots,v_n$ 的对偶基为 $\psi_1,\cdots,\psi_m,\varphi_1,\cdots,\varphi_n$, 证明 $\varphi_1,\cdots,\varphi_n$ 是 U^0 的基即可. □

(**方法二: 嵌入映射**) 设 $i\in\mathcal{L}(U,V)$ 为嵌入映射: 对任意的 $u\in U$, $i(u)=u$. 因为 $i'\in\mathcal{L}(V',U')$, 故由线性映射基本定理知

$$\dim V'=\dim\operatorname{Ker} i'+\dim\operatorname{Im} i'.$$

注意到 $\dim V'=\dim V$, $\operatorname{Ker} i'=U^0$, 因此上式变为

$$\dim V=\dim U^0+\dim\operatorname{Im} i'.$$

下面证明 $\dim\operatorname{Im} i'=\dim U$ 即可. 根据上一章习题 2. 知任意的 $\varphi\in U'$ 可以扩张为 $\psi\in V'$. 由 i 的定义知

$$i'(\psi)=\psi i=\varphi.$$

因此 $\varphi\in\operatorname{Im} i'$, 从而 $\operatorname{Im} i'=U'$, 故

$$\dim\operatorname{Im} i'=\dim U'=\dim U.$$

结论得证. □

注：上述结论也常常写为

$$\dim V=\dim U+\dim U^0.$$

它说明线性空间的维数等于子空间的维数与该子空间的零化子的维数之和.

注意下面这一结论不需要 V 和 W 是有限维的.

命题 (**对偶映射的核与原映射的像**)：设 $T\in\mathcal{L}(V,W)$, 则

$$\operatorname{Ker}T'=(\operatorname{Im}T)^0.$$

证明 $(\operatorname{Im}T)^0=\{\varphi\in W':$ 对任意的 $w\in\operatorname{Im}T,\varphi(w)=0\}$.

先证 $\operatorname{Ker}T'\subseteq(\operatorname{Im}T)^0$. 任取 $\varphi\in\operatorname{Ker}T'$, 则 $T'(\varphi)=\varphi T=0$. 因此对任意的 $v\in T$, $(\varphi T)(v)=\varphi(Tv)=0$, 故 $\varphi\in(\operatorname{Im}T)^0$.

再证 $(\operatorname{Im}T)^0\subseteq\operatorname{Ker}T'$. 任取 $\varphi\in(\operatorname{Im}T)^0$, 则对任意的 $v\in V$, 有 $(T'(\varphi))(v)=(\varphi T)(v)=\varphi(Tv)=0$, 即 $T'(\varphi)=0$, 故 $\varphi\in\operatorname{Ker}T'$. □

注：当限定 V 和 W 是有限维的时候, 我们还能利用线性映射基本定理得到

$$\dim\operatorname{Ker}T'=\dim\operatorname{Ker}T+\dim W-\dim V.$$

下面的命题很有用, 因为有时候证明 T' 是单的比证明 T 是满的更容易.

命题 (**T 是满的等价于 T' 是单的**)：设 V 和 W 都是有限维的, $T\in\mathcal{L}(V,W)$, 则 T 是满的当且仅当 T' 是单的.

证明 T 满 $\iff\operatorname{Im}T=W\iff(\operatorname{Im}T)^0=\{0\}\iff\operatorname{Ker}T'=\{0\}\iff T'$ 单. 其中第二个等价符号是因为线性空间的维数等于其子空间的维数与对应零化子的维数之和, 第三个等价符号利用了前一结论. □

引理 (**T' 的像与 T 的像的维数相同**)：设 V 和 W 都是有限维的, $T\in\mathcal{L}(V,W)$, 则

$$\dim\operatorname{Im}T'=\dim\operatorname{Im}T.$$

证明 由线性映射基本定理与上一注记知

$$\begin{aligned}\dim\operatorname{Im}T' &= \dim W' - \dim\operatorname{Ker}T' \\ &= \dim W - \dim\operatorname{Ker}T' \\ &= \dim W - (\dim\operatorname{Ker}T + \dim W - \dim V) \\ &= \dim V - \dim\operatorname{Ker}T \\ &= \dim\operatorname{Im}T.\end{aligned}$$

结论得证. □

命题 (对偶映射的像与原映射的核): 设 V 和 W 都是有限维的, $T \in \mathcal{L}(V, W)$, 则

$$\operatorname{Im}T' = (\operatorname{Ker}T)^0.$$

证明 先证 $\operatorname{Im}T' \subseteq (\operatorname{Ker}T)^0$. 任取 $\varphi \in \operatorname{Im}T'$, 则存在 $\psi \in W'$, 使得 $\varphi = T'(\psi) = \psi T$, 从而对任意的 $v \in \operatorname{Ker}T$, 有 $\varphi(v) = (\psi T)(v) = \psi(Tv) = 0$, 故 $\varphi \in (\operatorname{Ker}T)^0$.

再证 $\dim\operatorname{Im}T' = \dim(\operatorname{Ker}T)^0$. 由上述引理与线性映射基本定理知

$$\dim\operatorname{Im}T' = \dim\operatorname{Im}T = \dim V - \dim\operatorname{Ker}T = \dim(\operatorname{Ker}T)^0.$$

故 $\operatorname{Im}T' = (\operatorname{Ker}T)^0$. □

命题 (T 是单的等价于 T' 是满的): 设 V 和 W 都是有限维的, $T \in \mathcal{L}(V, W)$, 则 T 是单的当且仅当 T' 是满的.

证明 T 单 $\iff \operatorname{Ker}T = \{0\} \iff (\operatorname{Ker}T)^0 = V' \iff \operatorname{Im}T' = V' \iff T'$ 满. □

4.5 对偶映射的矩阵, 矩阵的转置

定义 (矩阵的转置): 设 $A = (a_{ij}) \in \mathbf{F}^{m\times n}$, 我们称

$$A^{\mathrm{T}} = (a_{ji}) \in \mathbf{F}^{n\times m}$$

为矩阵 A 的转置 (transpose).

性质 (转置的线性性)：设 $A, C \in \mathbf{F}^{m\times n}$, $\lambda \in \mathbf{F}$, 则

(1) $(A+C)^{\mathrm{T}} = A^{\mathrm{T}} + C^{\mathrm{T}}$.

(2) $(\lambda A)^{\mathrm{T}} = \lambda A^{\mathrm{T}}$.

证明 请读者自行验证. □

性质 (矩阵相乘的转置)：设 $A \in \mathbf{F}^{m\times n}$, $C \in \mathbf{F}^{n\times p}$, 则 $(AC)^{\mathrm{T}} = C^{\mathrm{T}}A^{\mathrm{T}}$.

证明 设 $B = AC$, $D = C^{\mathrm{T}}A^{\mathrm{T}}$, 因为

$$d_{ji} = \sum_{r=1}^{n} c_{rj}a_{ir} = \sum_{r=1}^{n} a_{ir}c_{rj} = b_{ij}.$$

因此 $B = D^{\mathrm{T}}$, 即 $(AC)^{\mathrm{T}} = C^{\mathrm{T}}A^{\mathrm{T}}$. □

性质 (对偶映射与转置)：设 $T \in \mathcal{L}(V, W)$. 又设 $v_1, \cdots, v_n$ 是 V 的基, $\varphi_1, \cdots, \varphi_n$ 为其对偶基; $w_1, \cdots, w_m$ 是 W 的基, $\psi_1, \cdots, \psi_m$ 为其对偶基, 则

$$\mathcal{M}(T') = (\mathcal{M}(T))^{\mathrm{T}}.$$

证明 记 $A = \mathcal{M}(T')$, $C = \mathcal{M}(T)$, 下面证明 $A = C^{\mathrm{T}}$. 由于

$$(T'(\psi_1), \cdots, T'(\psi_m)) = (\varphi_1, \cdots, \varphi_n)A,$$
$$(Tv_1, \cdots, Tv_n) = (w_1, \cdots, w_m)C.$$

故

$$T'(\psi_i) = \psi_i T = a_{1i}\varphi_1 + \cdots + a_{ni}\varphi_n,$$
$$Tv_j = c_{1j}w_1 + \cdots + c_{mj}w_m.$$

上述第一个等式用两边作用于向量 v_j, 第二个等式用映射 ψ_i 作用, 利用对偶基的定义我们得到

$$a_{ji} = (\psi_i T)(v_j) = \psi_i(Tv_j) = c_{ij}.$$

故 $A = C^{\mathrm{T}}$, 即 $\mathcal{M}(T') = (\mathcal{M}(T))^{\mathrm{T}}$. □

4.6 矩阵的秩

我们首先定义两个与矩阵有关的非负整数.

定义 (**行秩, 列秩**): 设 $A \in \mathbf{F}^{m\times n}$, 我们称 A 的各行在 $\mathbf{F}^{1\times n}$ 中张成空间的维数为 A 的行秩 (row rank), 称 A 的各列在 $\mathbf{F}^{m\times 1}$ 中张成空间的维数为 A 的列秩 (column rank).

例 1 求 $A = \begin{pmatrix} 4 & 7 & 1 & 8 \\ 3 & 5 & 2 & 9 \end{pmatrix}$ 的行秩与列秩.

解 A 的行秩等于 $\mathbf{F}^{1\times 4}$ 中

$$\operatorname{span}\left\{\begin{pmatrix} 4 & 7 & 1 & 8 \end{pmatrix}, \begin{pmatrix} 3 & 5 & 2 & 9 \end{pmatrix}\right\}$$

的维数. 因为这两个向量中的任何一个都不是另一个的标量倍, 因此行秩为 2.

A 的列秩等于 $\mathbf{F}^{2\times 1}$ 中

$$\operatorname{span}\left\{\begin{pmatrix} 4 \\ 3 \end{pmatrix}, \begin{pmatrix} 7 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 8 \\ 9 \end{pmatrix}\right\}$$

的维数. 因为前两个向量是线性无关的, 因此 A 的列秩大于等于 2. 又因为 $\dim \mathbf{F}^{2\times 1} = 2$, 因此 $\mathbf{F}^{2\times 1}$ 中线性无关组的最大长度为 2, 故 A 的列秩为 2.

注意在下面结果的陈述中并没有出现基. 虽然 $\mathcal{M}(T)$ 依赖于 V 与 W 的基的选取, 但下面的结果表明 $\mathcal{M}(T)$ 的列秩对任意选定的基都是相同的, 因为 $\operatorname{Im} T$ 并不依赖基的选取.

引理 (**列秩等于像的维数**): 设 V, W 是有限维线性空间, $T \in \mathcal{L}(V, W)$, 那么

$$\mathcal{M}(T) \text{ 的列秩} = \dim \operatorname{Im} T.$$

证明 设 $v_1, \cdots, v_n$ 是 V 的基, $w_1, \cdots, w_m$ 是 W 的基. 因为将 $\operatorname{span}\{Tv_1, \cdots, Tv_n\}$ 中的 w 映为 $\operatorname{span}\{\mathcal{M}(Tv_1), \cdots, \mathcal{M}(Tv_n)\}$ 中的 $\mathcal{M}(w)$ 的映射是一个同构, 因此

$$\begin{aligned} \dim \operatorname{span}\{Tv_1, \cdots, Tv_n\} &= \dim \operatorname{span}\{\mathcal{M}(Tv_1), \cdots, \mathcal{M}(Tv_n)\} \\ &= \dim \operatorname{span}\{\mathcal{M}(T)_1, \cdots, \mathcal{M}(T)_n\} \\ &= \mathcal{M}(T) \text{ 的列秩}. \end{aligned}$$

又因为 $\operatorname{Im} T = \operatorname{span}\{Tv_1, \cdots, Tv_n\}$, 故

$$\dim \operatorname{Im} T = \mathcal{M}(T) \text{ 的列秩}.$$

结论得证. $\square$

例 1 中矩阵 A 的行秩与列秩相等, 下面我们证明这个结论是普遍成立的.

定理 (**行秩与列秩相等**): 设 $A \in \mathbf{F}^{m\times n}$, 则 A 的行秩等于 A 的列秩.

证明 定义线性映射

$$\begin{aligned} T: \mathbf{F}^{n\times 1} &\longrightarrow \mathbf{F}^{m\times 1} \\ x &\longmapsto Ax \end{aligned}$$

取 $\mathbf{F}^{n\times 1}$ 与 $\mathbf{F}^{m\times 1}$ 的标准基, 则 $\mathcal{M}(T)=A$, 于是

$$\begin{aligned} A\,\text{的列秩} &= \mathcal{M}(T)\,\text{的列秩} \\ &= \dim \operatorname{Im} T \\ &= \dim \operatorname{Im} T' \\ &= \mathcal{M}(T')\,\text{的列秩} \\ &= (\mathcal{M}(T))^{\mathrm{T}}\,\text{的列秩} \\ &= \mathcal{M}(T)\,\text{的行秩} \\ &= A\,\text{的行秩}. \end{aligned}$$

这里第三个等号利用了 §4.4 的引理, 且 $\mathcal{M}(T')$ 按照标准基的对偶基计算. □

以上命题使得我们可以不区分行秩与列秩这两个术语, 而直接使用更简单的术语 “秩”.

定义 (秩): 我们把矩阵 $A\in\mathbf{F}^{m\times n}$ 的列秩称为 A 的秩 (rank).

5 多项式

这简短的一章包含多项式的内容, 我们在讨论线性变换时要用到这些材料, 本章的很多结果你可能已经从其他课程熟悉了, 把它们写在这里只是为了完整性.

因为这一章不是关于线性代数的, 所以老师可能会讲得很快. 你无需把所有的证明都仔细看一遍, 但至少要把本章所有结论的陈述看一遍, 并且理解它们——本书的后续章节要用到这些结论.

在讨论复系数或实系数多项式之前, 我们需要多了解一点复数知识. 下面是复数的一些重要的性质.

性质 (复数): 设 $z,w\in\mathbf{C}$, 则

(1) $z+\bar{z}=2\operatorname{Re} z$.

(2) $z-\bar{z}=2\mathrm{i}\operatorname{Im} z$.

(3) $z\bar{z}=|z|^2$.

(4) $\overline{z+w}=\bar{z}+\bar{w}$, $\overline{zw}=\bar{z}\bar{w}$.

(5) $\bar{\bar{z}}=z$.

(6) $\operatorname{Re} z \leqslant |z|$, $\operatorname{Im} z \leqslant |z|$.

(7) $|\bar{z}| = |z|$.

(8) $|zw| = |z||w|$.

(9) $|z+w| \leqslant |z| + |w|$.

回忆一下, 设函数 $p: \mathbf{F} \to \mathbf{F}$, 若存在 $a_0, a_1, \cdots, a_m \in \mathbf{F}$, 使得对任意的 $z \in \mathbf{F}$, 都有

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m,$$

则称函数 p 是系数在 $\mathbf{F}$ 中的多项式.

命题 (**若多项式为零函数, 则系数全为** 0) : 设 $a_0, \cdots, a_m \in \mathbf{F}$, 若对任意的 $z \in \mathbf{F}$, 都有

$$a_0 + a_1 z + \cdots + a_m z^m = 0,$$

则 $a_0 = a_1 = \cdots = a_m = 0$.

证明 假设并非所有的系数都是 0, 那么我们总可以设 a_m 不为 0. 取

$$z = \frac{|a_0| + |a_1| + \cdots + |a_{m-1}|}{|a_m|} + 1,$$

那么 $z \geqslant 1$(注意这里所取的 z 是一个实数), 于是

$$|a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}| \leqslant (|a_0| + |a_1| + \cdots + |a_{m-1}|) z^{m-1} < |a_m z^m|.$$

因此 $a_0 + a_1 z + \cdots + a_m z^m \neq 0$, 矛盾. □

命题 (**多项式系数的唯一性**) : 设 $p: \mathbf{F} \to \mathbf{F}$ 是多项式, 那么存在唯一的 $a_0, a_1, \cdots, a_m \in \mathbf{F}$, 使得对任意的 $z \in \mathbf{F}$, 都有

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m.$$

证明 假设多项式 p 有两组不同的系数, 那么它们相减之后得到 0, 由前一命题可知对应次数的系数是相等的. □

由上可知将多项式的次数定义为最高非零系数的幂是合理的. 多项式 0 的次数规定为 $-\infty$, 必要时可以使用关于 $-\infty$ 的显然运算. 例如对于任意的整数 m, 有 $-\infty < m$, $-\infty + m = -\infty$.

注：规定零多项式的次数为 $-\infty$, 很多结果就不再有例外. 例如即使 $p=0$, 也有

$$\deg(pq)=\deg p+\deg q.$$

5.1 带余除法

命题（多项式的带余除法）：设 $p,s\in\mathcal{P}(\mathbf{F})$, 那么存在唯一的 $q,r\in\mathcal{P}(\mathbf{F})$ 使得

$$p=qs+r,$$

其中 $\deg r<\deg s$.

定义（零点）：设 $p\in\mathcal{P}(\mathbf{F})$, 若存在 $\lambda\in\mathbf{F}$ 使得 $p(\lambda)=0$, 则称 λ 为多项式 p 的零点 (zero of a polynomial) 或根.

定义（因式）：设 $p,q\in\mathcal{P}(\mathbf{F})$, 若存在 $s\in\mathcal{P}(\mathbf{F})$, 使得 $p=sq$, 则称 q 为 p 的因式 (factor).

命题（多项式的每个零点对应一个一次因式）：设多项式 $p\in\mathcal{P}(\mathbf{F})$, $\lambda\in\mathbf{F}$, 则 $p(\lambda)=0$ 当且仅当存在多项式 $q\in\mathcal{P}(\mathbf{F})$, 使得对任意的 $z\in\mathbf{F}$, 有

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

命题（多项式的零点个数不超过它的次数）：设 $p\in\mathcal{P}(\mathbf{F})$ 的次数为 $m(\geqslant 0)$, 则 p 在 $\mathbf{F}$ 中最多有 m 个互不相同的零点.

5.2 C 上多项式的分解

定理（代数学基本定理）：任意一个非常数复系数多项式必有零点.

代数学基本定理给出了下面的复系数多项式分解定理.

定理（C 上多项式的分解）：若 $p\in\mathcal{P}(\mathbf{C})$ 是非常数多项式, 那么 p 可以唯一分解（不计因式次序）为

$$p(z)=c(z-\lambda_1)\cdots(z-\lambda_m),$$

其中 $c, \lambda_1, \cdots, \lambda_m \in \mathbf{C}$.

5.3 R 上多项式的分解

实系数多项式可能没有实的零点, 例如 $1+x^2$. 为了得到 $\mathbf{R}$ 上的分解定理, 我们要利用 $\mathbf{C}$ 上的分解定理.

首先我们有如下结论.

定理 (**实系数多项式的非实零点是成对出现的**): 设 $p \in \mathcal{P}(\mathbf{C})$ 为实系数多项式. 若 λ 是 p 的零点, 那么 $\bar{\lambda}$ 也是 p 的零点.

证明 设 $p(z) = a_0 + a_1 z + \cdots + a_m z^m$, 其中 $a_0, \cdots, a_m$ 均为实数. 设 $\lambda \in \mathbf{C}$ 是 p 的零点, 则

$$p(\lambda) = a_0 + a_1 \lambda + \cdots + a_m \lambda^m = 0.$$

等式两端取共轭得

$$\overline{p(\lambda)} = a_0 + a_1 \bar{\lambda} + \cdots + a_m \bar{\lambda}^m = 0.$$

因此 $\bar{\lambda}$ 也是 p 的零点. □

我们想要得到实系数多项式的分解定理. 为此需要先来描述可以写成两个一次实系数多项式之积的二次实系数多项式.

命题 (**二次多项式的分解**): 设 $b, c \in \mathbf{R}$, 则存在 $\lambda_1, \lambda_2 \in \mathbf{R}$ 使得分解式

$$x^2 + bx + c = (x - \lambda_1)(x - \lambda_2)$$

成立当且仅当 $b^2 - 4c \geqslant 0$.

在下面的定理中, m 或 M 可能等于 0.

定理 (**R 上多项式的分解**): 设 $p \in \mathcal{P}(\mathbf{R})$ 是非常数多项式, 则 p 可以唯一分解 (不计因子次序) 为

$$p(x) = c\,\underbrace{(x-\lambda_1)\cdots(x-\lambda_m)}_{\text{一次多项式}}\,\underbrace{(x^2+b_1x+c_1)\cdots(x^2+b_Mx+c_M)}_{\text{二次多项式}}.$$

其中 $c, \lambda_1, \cdots, \lambda_m, b_1, \cdots, b_M, c_1, \cdots, c_M$ 均为实数, 且对每个 j 都有 $b_j^2 - 4c_j < 0$.

6 特征值与特征向量

现在我们开始研究线性空间到自身的线性映射. 对这种线性映射的研究构成了线性代数最重要的部分. 我们仍然用 $\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$, 用 V 表示 $\mathbf{F}$ 上的线性空间.

6.1 不变子空间

我们来看看如何能更好地理解线性变换. 设 $T \in \mathcal{L}(V)$, 如果 V 有直和分解

$$V = U_1 \oplus \cdots \oplus U_m,$$

其中每个 U_j 都是 V 的真子空间, 那么要想了解 T 的特性, 只需要了解每个 $T|_{U_j}$ 的特性. 因为 U_j 是比 V 更小的线性空间, 因此处理 $T|_{U_j}$ 应该比处理 T 更容易.

但是, 要想使用算子研究中一些有用的工具 (比如取幂), 就会有一个问题: $T|_{U_j}$ 可能不把 U_j 映到自身, 换句话说, $T|_{U_j}$ 可能不是 U_j 上的线性变换. 因此我们只考虑 V 的这种直和分解——T 把每个 U_j 都映到自身. 我们为这种特殊的子空间取了一个形象的名字——不变子空间.

定义 (不变子空间)：设 U 是 V 的子空间, 若对任意的 $u \in U$, 都有 $Tu \in U$, 即 $TU \subseteq U$, 则称 U 是 V 的 T-不变子空间 (invariant subspace) , 或称 U 在 T 下不变.

也就是说, U 在 T 下不变当且仅当 $T|_U$ 是 U 上的线性变换.

例 1 (**四个最常见的不变子空间**) 设 $T \in \mathcal{L}(V)$, 则

$$\{0\}, \quad V, \quad \operatorname{Ker} T, \quad \operatorname{Im} T$$

都是 V 在 T 下的不变子空间.

例 2 设 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}))$ 定义为 $Tp = p'$, 则 $\mathcal{P}(\mathbf{R})$ 的子空间 $\mathcal{P}_4(\mathbf{R})$ 在 T 下是不变的, 因为若 $p \in \mathcal{P}(\mathbf{R})$ 的次数不超过 4, 那么 p' 的次数也不超过 4.

一个线性变换 $T \in \mathcal{L}(V)$ 是不是一定有不同于 $\{0\}$ 与 V 的不变子空间呢? 以后我们会知道, 如果 V 是有限维的, 且 $\dim V > 1$ (对于 $\mathbf{F} = \mathbf{C}$) 或 $\dim V > 2$ (对于 $\mathbf{F} = \mathbf{R}$), 这个问题的答案是肯定的.

6.2 特征值与特征向量

我们以后还会回过头来更深入的研究不变子空间. 现在我们先来研究最简单的非平凡不变子空间——一维不变子空间.

任取 $0 \neq v \in V$, 则

$$U = \{\lambda v : \lambda \in \mathbf{F}\} = \operatorname{span}\{v\}$$

是 V 的一维子空间. 若 U 在 T 下不变, 则对任意的 $v \in U$, 有 $Tv \in U$, 于是必有 $\lambda \in \mathbf{F}$, 使得 $Tv = \lambda v$. 反之, 设 $v \in V$, 若有某 $\lambda \in \mathbf{F}$ 使得 $Tv = \lambda v$, 那么 $\operatorname{span}\{v\}$ 就是 V 在 T 下的不变子空间. 由以上讨论我们看到方程

$$Tv = \lambda v$$

与一维不变子空间密切相关, 于是我们为其取了特别的名字.

定义 (**特征值、特征向量**) : 设 V 为 $\mathbf{F}$ 上的线性空间, $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{F}$, 若存在非零向量 $0 \neq v \in V$, 使得

$$Tv = \lambda v,$$

则称 λ 为 T 的特征值 (eigenvalue), 称 v 为 T 的相应于特征值 λ 的特征向量 (eigenvector).

在上面的定义中, 我们要求 $v \neq 0$, 是因为对任意的 $\lambda \in \mathbf{F}$, 都有 $T0 = \lambda 0$.

注 : eigenvalue 这个词一半是德文, 一半是英文. 德文形容词 eigen 的意思是 "特有的".

由前面的讨论我们得到特征值与一维不变子空间具有如下关系.

性质 (**特征值与一维不变子空间**) : T 有一维不变子空间当且仅当 T 有特征值.

命题 (**特征值的等价条件**) : 设 V 是有限维线性空间, $T \in \mathcal{L}(V)$, 则下列各条是等价的:

(1) λ 是 T 的特征值.

(2) $T - \lambda I$ 不是单的.

(3) $T - \lambda I$ 不是满的.

(4) $T - \lambda I$ 不是可逆的.

证明 因为 $Tv = \lambda v$ 有非零解 v 当且仅当 $(T - \lambda I)v = 0$ 有非零解 v, 当且仅当 $T - \lambda I$ 不是单的, 故 (1) 与 (2) 是等价的. 又因为有限维线性空间中的线性变换是单的当且仅当是满的, 当且仅当是可逆的, 故 (2)(3)(4) 是等价的. □

我们不难看出, $0 \neq v \in V$ 是 T 的相应于特征值 λ 的特征向量, 当且仅当

$$0 \neq v \in \operatorname{Ker}(T - \lambda I).$$

例 1 设 $T \in \mathcal{L}(\mathbf{F}^2)$ 定义为 $T(w, z) = (-z, w)$.

(1) 当 $\mathbf{F} = \mathbf{C}$ 时, 求 T 的特征值与特征向量.

(2) 当 $\mathbf{F} = \mathbf{R}$ 时, 求 T 的特征值与特征向量.

解 当 $\mathbf{F}=\mathbf{C}$ 时, 特征值为 $\lambda_1=\mathrm{i}$ 与 $\lambda_2=-\mathrm{i}$, 相应的特征向量为 $v_1=k(\mathrm{i},1)$ 与 $v_2=k(-\mathrm{i},1)$, 其中 $0\neq k\in\mathbf{C}$. 当 $\mathbf{F}=\mathbf{R}$ 时, T 没有特征值. 实际上, 此时的 T 是 $\mathbf{R}^2$ 中逆时针 90° 的旋转, 因此它不可能将一个非零向量变为此向量的标量倍.

性质 (相应于不同特征值的特征向量线性无关): 设 $T\in\mathcal{L}(V)$, $\lambda_1,\cdots,\lambda_m$ 是 T 的不同特征值, $v_1,\cdots,v_m$ 是相应的特征向量, 则 $v_1,\cdots,v_m$ 是线性无关的.

证明 对 m 使用数学归纳法. 当 $m=1$ 时, 由于特征向量非零, 故必线性无关. 假设对 m 个互异特征值的情形成立, 下面证明对 $m+1$ 个互异特征值的情形也成立. 设

$$k_1v_1+\cdots+k_mv_m+k_{m+1}v_{m+1}=0. \tag{$*$}$$

用 T 作用于 $(*)$ 式两端, 则

$$k_1\lambda_1v_1+\cdots+k_m\lambda_mv_m+k_{m+1}\lambda_{m+1}v_{m+1}=0. \tag{1}$$

用 λ_{m+1} 乘以 $(*)$ 式, 得

$$k_1\lambda_{m+1}v_1+\cdots+k_m\lambda_{m+1}v_m+k_{m+1}\lambda_{m+1}v_{m+1}=0. \tag{2}$$

(1)−(2) 得

$$k_1(\lambda_1-\lambda_{m+1})v_1+\cdots+k_m(\lambda_m-\lambda_{m+1})v_m=0.$$

由假设知 $v_1,\cdots,v_m$ 线性无关, 又因为各个 λ_i 互异, 故

$$k_1=\cdots=k_m=0,$$

从而 $k_{m+1}=0$, $v_1,\cdots,v_m,v_{m+1}$ 线性无关. 由数学归纳法知结论得证. □

推论 (互异特征值的个数不超过线性空间的维数): 设 V 是有限维线性空间, $T\in\mathcal{L}(V)$, 则 T 最多有 $\dim V$ 个互异特征值.

证明 由前一结论即得. □

6.3 限制变换与商变换

若 $T\in\mathcal{L}(V)$ 且 U 是 V 的 T-不变子空间, 那么 U 以自然的方式确定了另外两个线性变换——$T|_U\in\mathcal{L}(U)$ 与 $T/U\in\mathcal{L}(V/U)$. 限制变换 $T|_U$ 就是我们早已熟知的映射的限制, 而因为 U 在 T 下不变, 因此 $T|_U$ 确实是 U 上的变换. 下面我们看看商变换是如何定义的.

定义 (商变换)：设 U 为 V 的 T-不变子空间，我们把

$$\begin{aligned} T/U:\ V/U &\longrightarrow V/U \\ v+U &\longmapsto Tv+U \end{aligned}$$

称为 T 关于不变子空间 U 的商变换.

我们要证明商变换是一个定义合理的映射.

证明 设 $v+U=w+U$, 则 $v-w\in U$, 从而 $T(v-w)\in U$, $Tv+U=Tw+U$. □

6.4 线性变换的多项式

线性变换的理论要比线性映射的理论更为丰富，主要原因是线性变换能自乘为幂. 我们从线性变换的幂以及多项式作用于线性变换这一关键概念的定义开始说起.

若 $T\in\mathcal{L}(V)$, 那么 TT 有意义，且 TT 也属于 $\mathcal{L}(V)$. 通常用 T^2 表示 TT. 更一般地，我们有下面的定义.

定义 (线性变换的幂)：设 $T\in\mathcal{L}(V)$, m 为正整数，定义

$$T^m=\underbrace{T\cdots T}_{m\text{个}}.$$

此外，定义 $T^0=I$ 为 V 上的恒等变换. 若 T 是可逆的，定义

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

性质 (线性变换幂的运算)：设 $T\in\mathcal{L}(V)$, 则

$$T^mT^n=T^{m+n},\quad (T^m)^n=T^{mn}.$$

其中当 T 可逆时，m,n 为任意整数，当 T 不可逆时，m,n 为非负整数.

证明 请读者自行验证. □

定义 (线性变换的多项式)：设 $T\in\mathcal{L}(V)$, $p=a_0+a_1z+\cdots+a_mz^m\in\mathcal{P}(\mathbf{F})$, 则定义线性变换 $p(T)$ 为

$$p(T)=a_0I+a_1T+\cdots+a_mT^m\in\mathcal{L}(V).$$

注：$p(T)$ 是符号 p 的新用法，因为我们把它作用于线性变换，而不仅仅是 $\mathbf{F}$ 里面的元素.

例 1 设 $D \in \mathcal{L}(\mathcal{P}(\mathbf{R}))$ 定义为 $Dq = q'$, 多项式 p 为 $p(x) = 7 - 3x + 5x^2$, 那么 $p(D) = 7I - 3D + 5D^2$, 且对任意的 $q \in \mathcal{P}(\mathbf{R})$, 有 $(p(D))(q) = (7I - 3D + 5D^2)(q) = 7q - 3q' + 5q''$.

请自行验证, 对固定的 T, 则由 $p \mapsto p(T)$ 定义的从 $\mathcal{P}(\mathbf{F})$ 到 $\mathcal{L}(V)$ 的映射是线性的.

定义 (多项式的乘积): 设 $p, q \in \mathcal{P}(\mathbf{F})$, 则定义 $pq \in \mathcal{P}(\mathbf{F})$ 为对任意的 $z \in \mathbf{F}$,

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

引理 (多项式的乘积作用于线性变换): 设 $T \in \mathcal{L}(V)$, $p, q \in \mathcal{P}(\mathbf{F})$, 则 $(pq)(T) = p(T)q(T)$.

证明 设 $p(z) = \sum_{i=0}^{m} a_i z^i$, $q(z) = \sum_{j=0}^{n} b_j z^j$, 则 $(pq)(z) = p(z)q(z) = \sum_{i=0}^{m}\sum_{j=0}^{n} a_i b_j z^{i+j}$, 从而

$$(pq)(T) = \sum_{i=0}^{m}\sum_{j=0}^{n} a_i b_j T^{i+j} = \left(\sum_{i=0}^{m} a_i T^i\right)\left(\sum_{j=0}^{n} b_j T^j\right) = p(T)q(T).$$

结论得证. □

下面我们将要证明线性变换的多项式是交换的.

性质 (线性变换多项式满足交换律): 设 $T \in \mathcal{L}(V)$, $p, q \in \mathcal{P}(\mathbf{F})$, 则 $p(T)q(T) = q(T)p(T)$.

证明 由上述引理知 $p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T)$. □

6.5 特征值的存在性

现在给出复线性空间上线性变换的中心结果之一.

定理 (特征值的存在性): 复有限维线性空间上每个线性变换都有特征值.

证明 设 $\dim V = n$, 任取 $0 \neq v \in V$, 则 $n+1$ 个向量

$$v, Tv, \cdots, T^n v$$

是线性相关的, 因此存在不全为 0 的数 $a_0, a_1, \cdots, a_n \in \mathbf{C}$, 使得

$$a_0 v + a_1 Tv + \cdots + a_n T^n v = 0.$$

并且 $a_1, \cdots, a_n$ 也一定不全为 0(否则 $a_0 = 0$). 根据 $\mathbf{C}$ 上多项式的分解, 存在 $\lambda_1, \cdots, \lambda_m \in \mathbf{C}$, 使得

$$a_0 + a_1 z + \cdots + a_n z^n = c(z - \lambda_1)\cdots(z - \lambda_m).$$

其中 c 为非零复数. 注意 m 不必等于 n, 因为 a_n 可能为 0. 由此可得

$$
\begin{aligned}
0 &= a_0 v + a_1 T v + \cdots + a_n T^n v \\
&= (a_0 I + a_1 T + \cdots + a_n T^n) v \\
&= c(T - \lambda_1 I) \cdots (T - \lambda_m I) v.
\end{aligned}
$$

于是至少有一个 $T - \lambda_i I$ 不是单的 (可用反证法). 也就是说, T 有特征值. □

6.6 上三角矩阵

前面我们讨论了从一个线性空间到另一个线性空间的线性映射的矩阵, 该矩阵依赖于两个空间基的选取. 现在我们研究同一个空间上的线性变换, 自然地只需要使用一个基了.

定义 (线性变换的矩阵): 设 $T \in \mathcal{L}(V)$, $v_1, \cdots, v_n$ 是 V 的基, 设

$$
T v_j = a_{1j} v_1 + \cdots + a_{nj} v_n,
$$

我们称 $\mathcal{M}(T) = (a_{ij})$ 为线性变换 T 关于基 $v_1, \cdots, v_n$ 的矩阵. 当基不是上下文自明的时候, 记为 $\mathcal{M}(T, (v_1, \cdots, v_n))$.

若 T 是 $\mathbf{F}^n$ 的线性变换, 且没有指明基, 则默认为标准基.

例 1 设 $T \in \mathcal{L}(\mathbf{F}^3)$ 定义为 $T(x, y, z) = (2x + y, 5y + 3z, 8z)$, 则

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

线性代数的中心目标就是证明对于一个给定的算子 $T \in \mathcal{L}(V)$, 必然存在 V 的一个基, 使得 T 关于该基有一个相当简单的矩阵. 更具体一点, 我们要选择 V 的基, 使得 $\mathcal{M}(T)$ 有很多的 0.

若 V 是有限维复线性空间, 那么由于 $T \in \mathcal{L}(V)$ 必有一个特征值, 因此 V 有一个基使得 T 关于该基的矩阵形如

$$
\begin{pmatrix} \lambda & & \\ 0 & & \\ \vdots & & * \\ 0 & & \end{pmatrix}.
$$

为此只需要取 T 的一个特征向量, 再将其扩充为 V 的一个基.

马上就会看到, 我们可以选取 V 的一个基, 使得 T 关于这个基的矩阵有更多的 0.

定义 (对角线)：方阵的对角线由位于左上角到右下角的直线上的元素组成.

定义 (上三角矩阵)：若方阵的对角线下方的元素均为 0, 则称该矩阵为上三角矩阵 (upper-triangular matrix).

我们可以认为上三角矩阵是相当简单的, 因为当 n 比较大时, 几乎有一半的元素都等于 0. 上三角矩阵具有如下的典型形式:

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

注：我们经常用 $*$ 表示矩阵的未知元素或者其取值对讨论问题无关紧要的元素.

下面的命题揭示了上三角矩阵和不变子空间之间的很有用的联系.

命题 (上三角矩阵的等价条件)：设 $T \in \mathcal{L}(V)$, $v_1, \cdots, v_n$ 是 V 的基, 则下列各条等价:

(1) $\mathcal{M}(T)$ 是上三角矩阵.

(2) 对每个 j, $Tv_j \in \operatorname{span}\{v_1, \cdots, v_j\}$.

(3) 对每个 j, $\operatorname{span}\{v_1, \cdots, v_j\}$ 在 T 下不变.

证明 考虑

$$(Tv_1, \cdots, Tv_n) = (v_1, \cdots, v_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

不难得到结论. □

我们将在下一节证明, 对于复线性空间上的每个线性变换, 都存在一个基, 使得该线性变换关于这个基的矩阵是一个上三角矩阵.

6.7 复线性变换必有上三角矩阵

深刻的见解往往源于一个定理的多种证明. 对于以下结论, 我们给出两种证明方式, 你可以选择自己喜欢的那一种.

定理 ($\mathbf{C}$ 上的每个线性变换都有上三角矩阵)：设 V 是有限维复线性空间, $T \in \mathcal{L}(V)$, 则一定有 V 的某个基, 使得 T 关于这个基的矩阵 $\mathcal{M}(T)$ 为上三角矩阵.

证明 下面的两种方法都是对 V 的维数进行归纳.

(**方法一: 基扩张**) 当 $\dim V = 1$ 时结论显然成立. 假设结论对所有维数小于 $\dim V$ 的复线性空间均成立. 取 T 的一个特征值 λ, 那么 $T - \lambda I$ 不是满的, 因此对于

$$U = \operatorname{Im}(T - \lambda I)$$

有 $\dim U < \dim V$. 此外, 对任意的 $u \in U$, 有

$$Tu = (T - \lambda I)u + \lambda u \in U,$$

故 U 在 T 下不变. 设 $T|_U$ 关于 U 的基 $u_1, \cdots, u_m$ 的矩阵是上三角矩阵, 那么对每个 $j = 1, \cdots, m$, 都有

$$Tu_j = (T|_U)u_j \in \operatorname{span}\{u_1, \cdots, u_j\}. \tag{1}$$

将 $u_1, \cdots, u_m$ 扩充为 V 的基 $u_1, \cdots, u_m, v_1, \cdots, v_n$, 则

$$Tv_k = (T - \lambda I)v_k + \lambda v_k.$$

注意到 $(T - \lambda I)v_k \in U = \operatorname{span}\{u_1, \cdots, u_m\}$, 故对每个 $k = 1, \cdots, n$, 都有

$$Tv_k \in \operatorname{span}\{u_1, \cdots, u_m, v_1, \cdots, v_k\}. \tag{2}$$

由 (1)(2) 两式与上一节最后一个命题即得 T 关于基 $u_1, \cdots, u_m, v_1, \cdots, v_n$ 的矩阵是上三角矩阵. □

(**方法二: 商变换**) 当 $\dim V = 1$ 时结论显然成立. 设 $\dim V = n > 1$, 假设结论对所有维数小于 n 的复线性空间均成立. 取 T 的一个特征向量 v_1, 则

$$U = \operatorname{span}\{v_1\}$$

是 V 的 T-不变子空间, 且 $\dim U = 1$.

由于 $\dim V/U = n-1$, 我们可以对 V/U 使用归纳法的假设. 设 T/U 关于基 $v_2+U, \cdots, v_n+U$ 的矩阵是上三角矩阵, 于是对每个 $j = 2, \cdots, n$, 都有

$$(T/U)(v_j + U) \in \operatorname{span}\{v_2 + U, \cdots, v_j + U\}.$$

由此不难验证对每个 $j = 1, \cdots, n$, 都有

$$Tv_j \in \operatorname{span}\{v_1, \cdots, v_j\}.$$

从而 T 关于基 $v_1, \cdots, v_n$ 的矩阵是上三角矩阵. (容易验证 $v_1, \cdots, v_n$ 的确是 V 的基) □

注： 上述结论对实线性空间不成立, 因为由证明过程可以看出"存在特征值"这一条件是必不可少的, 而实线性空间是不一定有特征值的. 实际上, 如果一个实线性空间上的线性变换没有特征值, 那么它在任何基下的矩阵都不可能是上三角矩阵.

如何通过观察线性变换的矩阵来确定线性变换的是否可逆呢? 如果我们很幸运地使得该线性变换关于这个基的矩阵是上三角的, 那么问题就会变得方程简单.

命题 (由上三角矩阵确定线性变换的可逆性)： 设 T 关于某个基有上三角矩阵, 则 T 是可逆的当且仅当上三角矩阵 $\mathcal{M}(T)$ 对角线上的元素都不是 0.

证明 设 T 关于基 $v_1, \cdots, v_n$ 有上三角矩阵

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & & * \\ 0 & \lambda_2 & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

($\Longrightarrow$) 设 T 可逆. 首先可以肯定 $\lambda_1 \neq 0$, 否则 $Tv_1 = \lambda_1 v_1 = 0$, 于是 T 不单, 矛盾. 假设有某个 $2 \leqslant j \leqslant n$ 使得 $\lambda_j = 0$, 那么

$$T(\operatorname{span}\{v_1, \cdots, v_j\}) \subseteq \operatorname{span}\{v_1, \cdots, v_{j-1}\}.$$

因为 $\dim \operatorname{span}\{v_1, \cdots, v_j\} = j$, 而 $\dim \operatorname{span}\{v_1, \cdots, v_{j-1}\} = j-1$, 故 T 限制在 $\operatorname{span}\{v_1, \cdots, v_j\}$ 上不可能是单的, 于是 T 不是单的, 矛盾. 因此对任意的 j, 都有 $\lambda_j \neq 0$.

($\Longleftarrow$) 设对每个 j 都有 $\lambda_j \neq 0$. 将对角线上方的元素记为 a_{ij}, 设 $v = c_1 v_1 + \cdots + c_n v_n$ 满足 $Tv = 0$, 下证 $v = 0$. 由 $\mathcal{M}(T)$ 的定义得

$$\begin{aligned} 0 = Tv &= c_1 Tv_1 + \cdots + c_n Tv_n \\ &= (c_1\lambda_1 + c_2 a_{12} + \cdots + c_n a_{1n})v_1 + \cdots + (c_{n-1}\lambda_{n-1} + c_n a_{n-1,n})v_{n-1} + c_n \lambda_n v_n. \end{aligned}$$

由 $v_1, \cdots, v_n$ 线性无关知 v_n 的系数为 0, 又 $\lambda_n \neq 0$, 故 $c_n = 0$. 进而 $c_{n-1} = 0, \cdots, c_1 = 0$. 因此 $v = 0$, 即 T 是单的, 从而 T 是可逆的. □

令人遗憾的是, 目前为止我们还不能利用线性变换的矩阵精确地计算线性变换的特征值. 但是, 如果我们有幸找到了一个基, 使得线性变换关于这个基的矩阵是上三角的, 那么特征值的计算问题就变得非常平凡了.

命题（由上三角矩阵确定特征值）：设 T 关于某个基有上三角矩阵，那么 λ 是 T 的特征值当且仅当 λ 是该上三角矩阵对角线上的元素.

证明　设 $v_1,\cdots,v_n$ 是 V 的基，且 T 关于这个基的矩阵为上三角矩阵

$$\mathcal{M}(T)=\begin{pmatrix}\lambda_1 & & & * \\ 0 & \lambda_2 & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \lambda_n\end{pmatrix}.$$

设 $\lambda\in\mathbf{F}$，则

$$\mathcal{M}(T-\lambda I)=\begin{pmatrix}\lambda_1-\lambda & & & * \\ 0 & \lambda_2-\lambda & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \lambda_n-\lambda\end{pmatrix}.$$

从而由前一命题可知 $T-\lambda I$ 不可逆当且仅当且仅当 λ 等于 $\lambda_1,\cdots,\lambda_n$ 中的某一个，于是 λ 是 T 的特征值当且仅当 λ 等于 $\lambda_1,\cdots,\lambda_n$ 中的某一个. □

例 1　定义 $T\in\mathcal{L}(\mathbf{F}^3)$ 为 $T(x,y,z)=(2x+y,5y+3z,8z)$，则 T 关于标准基的矩阵为

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

因为 $\mathcal{M}(T)$ 是上三角的，故 T 的特征值为 $2,5,8$.

一旦知道 $\mathbf{F}^n$ 上线性变换的特征值了，利用高斯消元法就不难求得特征向量.

6.8　特征子空间与对角矩阵

定义（特征子空间）：设 $T\in\mathcal{L}(V)$，$\lambda\in\mathbf{F}$，我们称

$$\operatorname{Ker}(T-\lambda I)$$

为 T 的相应于 λ 的特征子空间 (eigen-subspace)，记作 $E(\lambda,T)$.

注：我们并没有要求上述定义中的 λ 是 T 的特征值. 实际上, 当 λ 不是 T 的特征值时,

$$E(\lambda, T) = \{0\}.$$

不难看出, $E(\lambda, T)$ 就是 λ 的全体特征向量加上 0 所成的集合, 并且它确实是子空间 (因为核一定是子空间). 此外, 由定义容易验证 λ 是 T 的特征值当且仅当 $E(\lambda, T) \neq \{0\}$.

例 1 设 $T \in \mathcal{L}(V)$ 关于基 v_1, v_2, v_3 的矩阵是

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

那么 $E(8, T) = \operatorname{span}\{v_1\}, E(5, T) = \operatorname{span}\{v_2, v_3\}$.

若 λ 是 T 的特征值, 那么 T 限制到 $E(\lambda, T)$ 上恰好是由 λ 确定的数乘变换.

命题 (特征子空间之和是直和)：设 V 是有限维线性空间, $T \in \mathcal{L}(V)$, $\lambda_1, \cdots, \lambda_m$ 是 T 的互异特征值, 那么

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

是直和. 此外还有

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leqslant \dim V.$$

证明 设

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

其中每个 $u_i \in E(\lambda_i, T)$. 因为属于不同特征值的特征向量线性无关, 因此必有每个 u_i 都为 0, 故 $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ 是直和, 进一步有

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) = \dim(E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)) \leqslant \dim V.$$

结论得证. □

定义 (对角矩阵)：对角线以外全是 0 的方阵称为对角矩阵 (diagonal matrix).

显然每个对角矩阵都是上三角矩阵, 但对角阵一般来说比上三角阵有更多的 0. 若线性变换 T 关于某个基有对角矩阵, 那么该对角矩阵对角线上的元素就是 T 的特征值.

定义 (可对角化)：若线性变换 T 关于某个基有对角矩阵, 则称 T 是可对角化的.

命题 (可对角化的充要条件)：设 V 是有限维线性空间, $T\in\mathcal{L}(V)$, $\lambda_1,\cdots,\lambda_m$ 为 T 的所有互异特征值, 则下面各条等价:

(1) T 可对角化.

(2) V 有由 T 的特征向量构成的基.

(3) V 有在 T 下不变的一维子空间 $U_1,\cdots,U_n$, 使得 $V=U_1\oplus\cdots\oplus U_n$.

(4) $V=E(\lambda_1,T)\oplus\cdots\oplus E(\lambda_m,T)$.

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

证明 先证 (1)(2)(3) 等价.

(1) $\Longrightarrow$ (2) 设 T 关于基 $v_1,\cdots,v_n$ 有对角矩阵

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

则 $v_1,\cdots,v_n$ 就是相应于 $\lambda_1,\cdots,\lambda_n$ 的特征向量. 注意到 $v_1,\cdots,v_n$ 是 V 的基, 结论得证.

(2) $\Longrightarrow$ (3) 设 $v_1,\cdots,v_n$ 是 V 的基, 且是 T 的特征向量. 令 $U_i=\operatorname{span}\{v_i\}$. 由 $v_1,\cdots,v_n$ 张成 V 知 $V=U_1+\cdots+U_n$, 由 $v_1,\cdots,v_n$ 线性无关知 $V=U_1\oplus\cdots\oplus U_n$.

(3) $\Longrightarrow$ (1) 任取 $0\neq v_i\in U_i$. 因为 U_i 是一维子空间, 故 $U_i=\operatorname{span}\{v_i\}$. 因为 U_i 关于 T 不变, 故存在 $\lambda_i\in\mathbf{F}$ 使得 $Tv_i=\lambda_i v_i$. 由 $V=U_1\oplus\cdots\oplus U_n$ 知 $v_1,\cdots,v_n$ 是 V 的基. 因为 T 关于 $v_1,\cdots,v_n$ 的矩阵为

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

故 T 可对角化.

下面证明 (2) $\Longrightarrow$ (4) $\Longrightarrow$ (5) $\Longrightarrow$ (2).

(2) $\Longrightarrow$ (4) 设 $v_1,\cdots,v_n$ 是特征向量也是基, 那么 $V=E(\lambda_1,T)+\cdots+E(\lambda_m,T)$. 因为互异特征子空间之和为直和, 故结论得证.

(4) $\Longrightarrow$ (5) 由 §4.1 的结论即得.

(5) $\Longrightarrow$ (2) 设 $v_{j,1},\cdots,v_{j,k_j}$ 为 $E(\lambda_j,T)$ 的基. 将它们合在一起得到一组特征向量. 设

$$\underbrace{a_{1,1}v_{1,1}+\cdots+a_{1,k_1}v_{1,k_1}}_{E(\lambda_1,T)}+\cdots+\underbrace{a_{m,1}v_{m,1}+\cdots+a_{m,k_m}v_{1,k_m}}_{E(\lambda_m,T)}=0.$$

又设

$$u_j=a_{j,1}v_{j,1}+a_{j,k_j}v_{j,k_j}\in E(\lambda_j,T),$$

则

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

因为互异特征子空间之和为直和, 故 $u_1 = \cdots = u_m = 0$. 又因为 $v_{j,1}, \cdots, v_{j,k_j}$ 是 $E(\lambda_j, T)$ 的基, 故 $a_{j,1} = \cdots = a_{j,k_j} = 0$. 因此

$$v_{1,1}, \cdots, v_{1,k_1}, \cdots, v_{m,1}, \cdots, v_{m,k_m}$$

线性无关, 是 V 的基. □

令人遗憾的是, 并不是每个线性变换都可对角化, 即使在复线性空间中也是如此.

例 2 考虑复线性空间 $\mathbf{C}^2$. 设 $T \in \mathcal{L}(\mathbf{C}^2)$ 定义为 $T(w, z) = (z, 0)$, 试证明 T 不可对角化.

证明 易知 T 有唯一特征值 0, 且 $E(0, T) = \{(w, 0) \in \mathbf{C}^2 : w \in \mathbf{C}\}$. 因 $\dim E(0, T) = 1 < 2 = \dim \mathbf{C}^2$, 故 T 不可对角化. □

但是, 如果一个线性变换的互异特征值的个数与线性空间的维数相同, 那么该线性变换就一定可对角化.

推论 (可对角化的充分条件): 设 V 是有限维线性空间, $T \in \mathcal{L}(V)$. 若 T 有 $\dim V$ 个互异的特征值, 则 T 可对角化.

证明 设 $\dim V = n$, $v_1, \cdots, v_n$ 分别为相应于不同特征值 $\lambda_1, \cdots, \lambda_n$ 的特征向量, 于是 $v_1, \cdots, v_n$ 线性无关. 注意到 $v_1, \cdots, v_n$ 的长度为 V 的维数, 因此 $v_1, \cdots, v_n$ 是 V 的基, T 关于这个基的矩阵为对角阵. □

例 3 定义 $T \in \mathcal{L}(\mathbf{F}^3)$ 为 $T(x, y, z) = (2x + y, 5y + 3z, 8z)$. 求 $\mathbf{F}^3$ 的一个基, 使得 T 关于这个基有对角矩阵.

解 因为 T 在标准基下的矩阵为

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

这是一个上三角矩阵, 故 T 有三个互异特征值 $2, 5, 8$, 可对角化. 直接求解方程

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

可得相应的特征向量分别为 $v_1 = (1, 0, 0), v_2 = (1, 3, 0), v_3 = (1, 6, 6)$, 它们构成 $\mathbf{F}^3$ 的一个基, T 关于这个基有对角矩阵

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

注：即使线性变换没有与线性空间维数一样多的特征值, 也可能可对角化, 只要它有足够多线性无关的特征向量即可. (参见前面可对角化的充要条件 (2))

7 内积空间

我们在定义线性空间时推广了 $\mathbf{R}^2$ 和 $\mathbf{R}^3$ 的线性结构 (加法与数乘), 而忽略了其他的重要特征, 例如长度与角度的概念, 这些思想蕴含于我们现在要研究的内积中.

本章我们仍然用 $\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$, 用 V 表示 $\mathbf{F}$ 上的线性空间.

7.1 内积与范数

关于下面的定义中记号的两点说明:

(1) 若 λ 是复数, 那么 $\lambda \geqslant 0$ 表示 λ 是实数, 且非负.

(2) 我们使用通用的记号 $\langle u, v\rangle$ 来表示内积. 有些教材使用圆括号, 但圆括号这个记号比较含糊, 可以表示多种东西.

定义 (内积)：设 V 是 $\mathbf{F}$ 上的线性空间. V 上的内积就是一个函数, 它将有序对 (u, v) 映到一个数 $\langle u, v\rangle \in \mathbf{F}$, 且满足:

(1) (正定性) $\langle u, u\rangle \geqslant 0$. $\langle u, u\rangle = 0$ 当且仅当 $u = 0$.

(2) (对第一个变元的线性性) $\langle \alpha u + \beta v, w\rangle = \alpha \langle u, w\rangle + \beta \langle v, w\rangle$. 其中 $\alpha, \beta \in \mathbf{F}$.

(3) (共轭对称性) $\langle u, v\rangle = \overline{\langle v, u\rangle}$.

每个实数都等于自己的复共轭, 因此若 $\mathbf{F} = \mathbf{R}$, 共轭对称性就直接变为对称性: $\langle u, v\rangle = \langle v, u\rangle$. 此外, 将 (2)(3) 结合起来, 我们得到内积关于第二个变元具有共轭线性性:

$$\langle u, \alpha v + \beta w\rangle = \bar{\alpha} \langle u, v\rangle + \bar{\beta} \langle u, w\rangle .$$

例 1 (一些常见的内积)

(1) (**欧几里得内积**) $\mathbf{F}^n$ 上的欧几里得内积定义为

$$\langle (w_1, \cdots, w_n), (z_1, \cdots, z_n)\rangle = w_1 \overline{z_1} + \cdots + w_n \overline{z_n}.$$

(2) (**变形的欧几里得内积**) 若 $c_1, \cdots, c_n$ 均为正数, 则可定义 $\mathbf{F}^n$ 上的内积为

$$\langle (w_1, \cdots, w_n), (z_1, \cdots, z_n)\rangle = c_1 w_1 \overline{z_1} + \cdots + c_n w_n \overline{z_n}.$$

注意这里的正数不能减弱为非负数, 因为内积要满足定性: $\langle u, u\rangle = 0$ 当且仅当 $u = 0$.

(3) 定义 $[-1,1]$ 上实值连续函数全体所成线性空间上的内积为

$$\langle f,g\rangle=\int_{-1}^{1}f(x)g(x)\,\mathrm{d}x.$$

(4) 定义 $\mathcal{P}(\mathbf{R})$ 上的内积为

$$\langle p,q\rangle=\int_{0}^{+\infty}p(x)q(x)\mathrm{e}^{-x}\,\mathrm{d}x.$$

定义 (内积空间)：带有内积的线性空间称为内积空间 (inner product space).

内积空间最重要的例子就是带有欧几里得内积的 $\mathbf{F}^n$. 以后当我们提到内积空间 $\mathbf{F}^n$ 时, 若无特别说明, 总假定使用的是欧几里得内积.

注：为避免不断重复说明, 在本章余下的部分, 我们总假定 V 是 $\mathbf{F}$ 上的内积空间.

性质 (零元的内积性质)：

(1) 对任意的 $v\in V$, 有 $\langle 0,v\rangle=0$.

(2) 对任意的 $u\in V$, 有 $\langle u,0\rangle=0$.

证明 (1) $\langle 0,v\rangle=\langle 0+0,v\rangle=\langle 0,v\rangle+\langle 0,v\rangle$.

(2) 对 (1) 的结论取共轭即可. □

每个内积都确定了一种范数.

定义 (范数)：设 V 是 $\mathbf{F}$ 上的内积空间, 我们定义 $v\in V$ 的范数 (norm) 为

$$\|v\|=\sqrt{\langle v,v\rangle}.$$

例 2 (**一些常见的范数**)

(1) (**欧几里得范数**) $\mathbf{F}^n$ 上的欧几里得范数为

$$\|(z_1,\cdots,z_n)\|=\sqrt{|z_1|^2+\cdots+|z_n|^2}.$$

(2) 取 $[-1,1]$ 上实值连续函数全体所成的内积空间, 并取例 1 中定义的内积, 则

$$\|f\|=\left(\int_{-1}^{1}(f(x))^2\,\mathrm{d}x\right)^{\frac{1}{2}}.$$

性质：范数具有如下性质:

(1) (正定性) $\|v\| \geqslant 0$. $\|v\| = 0$ 当且仅当 $v = 0$.

(2) (齐次性) $\|\lambda v\| = |\lambda| \, \|v\|$.

证明 (1) 由 $\|v\| = \sqrt{\langle v, v\rangle}$ 即得.

(2) $\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v\rangle} = \sqrt{\lambda \bar{\lambda} \langle v, v\rangle} = |\lambda| \sqrt{\langle v, v\rangle} = |\lambda| \, \|v\|$. □

7.2 正交性

现在我们给出一个关键的定义.

定义 (正交)：若 $\langle u, v\rangle = 0$, 则称向量 u 和 v 是正交的 (orthogonal).

上述定义中向量的次序是无关紧要的, 因为 $\langle u, v\rangle = 0$ 当且仅当 $\langle v, u\rangle = 0$. 考察 $\mathbf{R}^2$ 中的内积我们会发现向量 u 与 v 正交当且仅当 u 与 v 在通常的几何意义下是垂直的. 因此, 正交实际上就是直观几何里面垂直的推广.

我们从下面的简单结果开始研究正交性.

性质 (正交性与零向量)：

(1) 0 正交于 V 中的任何向量.

(2) 0 是 V 中唯一一个与自身正交的向量.

证明 (1) 前一节已经证明 $\langle 0, v\rangle = 0$.

(2) 由内积的定性: $\langle v, v\rangle = 0$ 当且仅当 $v = 0$ 即得. □

下面的定理是平面几何中关于直角三角形勾股定理的推广.

定理 (勾股定理)：若 u, v 正交, 则 $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.

证明 因为 $\langle u, v\rangle = 0$, 故

$$\begin{aligned} \|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \overline{\langle u, v\rangle} \\ &= \|u\|^2 + \|v\|^2 + 2\operatorname{Re}\langle u, v\rangle \\ &= \|u\|^2 + \|v\|^2 . \end{aligned}$$

结论得证. □

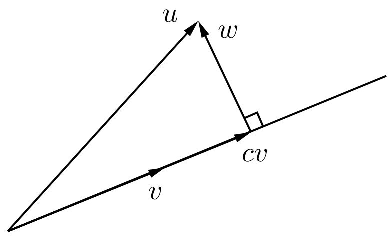
注：由上述证明不难看出，在实内积空间中，若 $\|u+v\|^2=\|u\|^2+\|v\|^2$，那么 u,v 正交. 因此勾股定理的逆命题对于实内积空间成立.

我们想要将 u 分解为 v 的标量倍与一个正交于 v 的向量之和，这样的分解称之为正交分解. 为了得到 u 的正交分解，我们将 u 写为 $u=cv+(u-cv)$. 如果 $u-cv$ 与 v 正交，那么

$$0=\langle u-cv,v\rangle=\langle u,v\rangle-c\|v\|^2.$$

因此取 $c=\dfrac{\langle u,v\rangle}{\|v\|^2}$，我们就得到了 u 关于 v 的正交分解. 换言之，我们得到了如下结论.

命题（正交分解）：设 $u,v\in V$ 且 $v\neq 0$，令 $c=\dfrac{\langle u,v\rangle}{\|v\|^2}$，$w=u-cv$，那么 $w\perp v$，且 $u=w+cv$.



我们将使用正交分解来证明下面的柯西-施瓦茨不等式，它是数学中最重要的不等式之一.

定理（柯西-施瓦茨不等式）：设 $u,v\in V$，则

$$|\langle u,v\rangle|\leqslant\|u\|\,\|v\|.$$

其中等号成立当且仅当 u,v 之一是另一个的标量倍.

证明 若 $v=0$，则结论是显然的. 下面假设 $v\neq 0$. 由正交分解，我们取 $w=u-\dfrac{\langle u,v\rangle}{\|v\|^2}v$，那么 $u=w+\dfrac{\langle u,v\rangle}{\|v\|^2}v$，且 $w\perp v$. 由勾股定理得

$$\begin{aligned}\|u\|^2&=\|w\|^2+\left\|\frac{\langle u,v\rangle}{\|v\|^2}v\right\|^2\\&=\|w\|^2+\frac{|\langle u,v\rangle|^2}{\|v\|^4}\|v\|^2\\&=\|w\|^2+\frac{|\langle u,v\rangle|^2}{\|v\|^2}\\&\geqslant\frac{|\langle u,v\rangle|^2}{\|v\|^2}.\end{aligned}$$

两边同时乘以 $\|v\|^2$ 再开方即得结论. 此时等号成立当且仅当 $w=0$. 综上, 等号成立当且仅当 u,v 之一是另一个的标量倍. □

例 1 下面是一些柯西-施瓦茨不等式的例子, 其中法国数学家柯西证明了 (1), 德国数学家施瓦茨证明了 (2).

(1) 设 $x_1,\cdots,x_n,y_1,\cdots,y_n\in\mathbf{R}$, 则

$$(x_1y_1+\cdots+x_ny_n)^2\leqslant(x_1^2+\cdots+x_n^2)(y_1^2+\cdots+y_n^2).$$

(2) 若 $f,g\in C[-1,1]$, 则

$$\left(\int_{-1}^{1}f(x)g(x)\,\mathrm{d}x\right)^2\leqslant\left(\int_{-1}^{1}(f(x))^2\,\mathrm{d}x\right)\left(\int_{-1}^{1}(g(x))^2\,\mathrm{d}x\right).$$

下面的命题称为三角不等式, 它有如下的几何解释: 三角形任意一边的长度小于另外两边的长度之和. 三角不等式表明两点之间的最短路线是直线.

命题 (三角不等式): 设 $u,v\in V$, 则

$$\|u+v\|\leqslant\|u\|+\|v\|,$$

其中等号成立当且仅当 u,v 其中一个是另一个的非负标量倍.

证明 由 $\operatorname{Re}z\leqslant|z|$ 与柯西-施瓦茨不等式得

$$\begin{aligned}\|u+v\|^2&=\|u\|^2+\|v\|^2+2\operatorname{Re}\langle u,v\rangle\\&\leqslant\|u\|^2+\|v\|^2+2|\langle u,v\rangle|\\&\leqslant\|u\|^2+\|v\|^2+2\|u\|\,\|v\|\\&=(\|u\|+\|v\|)^2.\end{aligned}$$

开方即得三角不等式. 容易验证等号当且仅当 u,v 之一是另一个的非负标量倍. □

下面的命题称为平行四边形公式, 因为它的几何解释为: 平行四边形的两条对角线的长度的平方和等于四条边的长度的平方和.

命题 (平行四边形公式): 设 $u,v\in V$, 则

$$\|u+v\|^2+\|u-v\|^2=2(\|u\|^2+\|v\|^2).$$

证明 直接计算得

$$\|u+v\|^2 = \langle u+v, u+v\rangle = \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle,$$
$$\|u-v\|^2 = \langle u-v, u-v\rangle = \|u\|^2 + \|v\|^2 - \langle u, v\rangle - \langle v, u\rangle.$$

两式相加即得结论. □

7.3 规范正交基与格拉姆-施密特正交化

定义 (规范正交组)：设 $e_1, \cdots, e_m$ 是 V 中的向量组. 如果每个 e_i 的范数为 1, 且与其余 e_j 均正交, 则称 $e_1, \cdots, e_m$ 是规范正交的 (orthonormal).

上述定义等价于

$$\langle e_i, e_j\rangle = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

例 1 $\mathbf{F}^n$ 的标准基是规范正交组.

例 2 $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)$ 是 $\mathbf{F}^3$ 中的规范正交组.

下面的命题表明规范正交组特别容易处理.

命题 (规范正交组线性组合的范数)：设 $e_1, \cdots, e_m$ 是 V 中的规范正交组, $a_1, \cdots, a_m \in \mathbf{F}$, 则

$$\|a_1e_1 + \cdots + a_me_m\|^2 = |a_1|^2 + \cdots + |a_m|^2.$$

证明 直接计算即得. □

上述命题有如下的重要推论.

推论 (规范正交组必线性无关)：规范正交组一定是线性无关组.

证明 设 $e_1, \cdots, e_m$ 是 V 中的规范正交组, 设

$$a_1e_1 + \cdots + a_me_m = 0.$$

由前述命题知

$$|a_1|^2 + \cdots + |a_m|^2 = 0,$$

从而 $a_1 = \cdots = a_m = 0$, $e_1, \cdots, e_m$ 线性无关. □

定义 (规范正交基)：规范正交组构成的基称为规范正交基 (orthonormal basis).

例 3 $\mathbf{F}^n$ 的标准基是规范正交基.

命题 (适当长度的规范正交组是规范正交基): V 中每个长度为 $\dim V$ 的规范正交组都是规范正交基.

证明 因为规范正交组是线性无关组, 而长度为 $\dim V$ 的线性无关组是基, 结论得证. □

一般的, 给定 V 的基 $v_1, \cdots, v_n$ 与向量 $v \in V$, 我们知道由标量 $a_1, \cdots, a_n \in \mathbf{F}$, 使得

$$v = a_1 v_1 + \cdots + a_n v_n.$$

对任意的一个基求出系数 $a_1, \cdots, a_n$ 不是一件容易的事. 但幸运的是, 对于规范正交基来说, 系数 a_j 就等于 $\langle v, e_j \rangle$.

命题 (规范正交基的作用): 设 $e_1, \cdots, e_n$ 是 V 的规范正交基, 那么对任意的 $v \in V$, 有

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n.$$

证明 设 $v = a_1 e_1 + \cdots + a_n e_n$. 两边用 e_j 作内积得 $a_j = \langle v, e_j \rangle$. □

现在我们知道了规范正交基的用处, 那我们如何找到它们呢? 例如, 对于由

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x) \, \mathrm{d}x$$

给出的内积, $\mathcal{P}_m(\mathbf{R})$ 有规范正交基吗? 下面的命题将会给出答案. 证明过程中用到的算法称为格拉姆-施密特正交化, 它可以把一个线性无关组转化为一个规范正交组, 且两个组有相同的张成空间.

命题 (格拉姆-施密特正交化): 设 $v_1, \cdots, v_m$ 是 V 中的线性无关组, 令 $e_1 = \dfrac{v_1}{\|v_1\|}$, 对 $j = 2, \cdots, m$, 定义

$$e_j = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{\|v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}\|},$$

则 $e_1, \cdots, e_m$ 是 V 中的规范正交组, 且每一个 $j = 1, \cdots, m$, 有

$$\operatorname{span}\{e_1, \cdots, e_j\} = \operatorname{span}\{v_1, \cdots, v_j\}.$$

例 4 求 $\mathcal{P}_2(\mathbf{R})$ 的一个规范正交基, 这里的内积是 $\langle p, q \rangle = \displaystyle\int_{-1}^{1} p(x) q(x) \, \mathrm{d}x$.

解 取 $\mathcal{P}_2(\mathbf{R})$ 的基 $1, x, x^2$, 分别记为 p_1, p_2, p_3.

第一步 因为 $\|p_1\| = \sqrt{2}$, 故 $e_1 = \dfrac{1}{\sqrt{2}}$.

第二步 因为 $p_2 - \langle p_2, e_1 \rangle e_1 = p_2$, 而 $\|p_2\| = \sqrt{\dfrac{2}{3}}$, 故 $e_2 = \sqrt{\dfrac{3}{2}} x$.

第三步 因为 $p_3-\langle p_2,e_1\rangle e_1-\langle p_2,e_2\rangle e_2=x^2-\frac{1}{3}$, 而 $\left\|x^2-\frac{1}{3}\right\|=\sqrt{\frac{8}{45}}$, 故 $e_3=\sqrt{\frac{45}{8}}\left(x^2-\frac{1}{3}\right)$. 因此

$$\frac{1}{\sqrt{2}},\quad \sqrt{\frac{3}{2}}x,\quad \sqrt{\frac{45}{8}}\left(x^2-\frac{1}{3}\right)$$

是 $\mathcal{P}_2(\mathbf{R})$ 的规范正交基.

现在我们可以回答规范正交基的存在性问题了.

命题 (**规范正交基的存在性**): 有限维内积空间必有规范正交基.

证明 设 V 是有限维的, 取 V 的一个基, 对其进行格拉姆-施密特正交化, 得到一个长度为 $\dim V$ 的规范正交组, 它就是 V 的一个规范正交基. $\square$

命题 (**规范正交组可扩充为规范正交基**): 设 V 是有限维内积空间, 则 V 的任何一个规范正交组都可扩充为规范正交基.

证明 设 $e_1,\cdots,e_m$ 为规范正交组. 因为它们线性无关, 故可将其扩充为 V 的基

$$e_1,\cdots,e_m,v_1,\cdots,v_n.$$

对这个基进行格拉姆-施密特正交化, 得到 V 的一个规范正交基. 由于前 m 个向量 $e_1,\cdots,e_m$ 已经是规范正交的了, 故它们会被保留下来. 于是最终得到的规范正交组为

$$e_1,\cdots,e_m,f_1,\cdots,f_n.$$

结论得证. $\square$

下面的命题表明, 如果 T 关于某个基有上三角矩阵, 那么 T 一定也关于某个规范正交基有上三角矩阵.

命题 (**关于规范正交基的上三角矩阵**): 设 $T\in\mathcal{L}(V)$. 若 T 关于 V 的某个基有上三角矩阵, 那么 T 关于 V 的某个规范正交基也有上三角矩阵.

证明 设 T 关于基 $v_1,\cdots,v_n$ 有上三角矩阵. 由上三角矩阵的等价条件 (3) 知对每个 j 都有 $\operatorname{span}\{v_1,\cdots,v_j\}$ 在 T 下不变. 对 $v_1,\cdots,v_n$ 进行格拉姆-施密特正交化得到规范正交基 $e_1,\cdots,e_n$, 则

$$\operatorname{span}\{e_1,\cdots,e_j\}=\operatorname{span}\{v_1,\cdots,v_j\}.$$

故 $\operatorname{span}\{e_1,\cdots,e_j\}$ 在 T 下不变. 因此 T 关于 $e_1,\cdots,e_n$ 也有上三角矩阵. $\square$

7.4 内积空间上的线性泛函与里斯表示定理

回忆一下, $\mathcal{L}(V,\mathbf{F})$ 中的元素称为线性泛函.

例 1 如下定义的映射 $\varphi:\mathbf{F}^3\to\mathbf{F}$

$$\varphi(z_1,z_2,z_3)=2z_1-5z_2+z_3$$

是 $\mathbf{F}^3$ 上的线性泛函. 我们可以将这个泛函写成如下形式: 对每个 $z\in\mathbf{F}^3$,

$$\varphi(z)=\langle z,u\rangle\,,$$

其中 $u=(2,-5,1)$.

例 2 如下定义的映射 $\varphi:\mathcal{P}_2(\mathbf{R})\to\mathbf{F}$

$$\varphi(p)=\int_{-1}^{1}p(t)\cos(\pi t)\,\mathrm{d}t$$

是 $\mathcal{P}_2(\mathbf{R})$ 上的线性泛函, 这里的内积是 $\langle p,q\rangle=\displaystyle\int_{-1}^{1}p(t)q(t)\,\mathrm{d}t$. 这里有一个并不显然的事实: 存在 $u\in\mathcal{P}_2(\mathbf{R})$, 使得对任意的 $p\in\mathcal{P}_2(\mathbf{R})$, 都有

$$\varphi(p)=\langle p,u\rangle\,.$$

注意我们不能取 $u(t)=\cos(\pi t)$, 因为它不是 $\mathcal{P}_2(\mathbf{R})$ 中的元素.

对于取定的 $u\in V$, 把 v 映到 $\langle v,u\rangle$ 的映射是 V 上的线性泛函. 下述定理表明, V 上的每个线性泛函都具有这种形式. 上面的例 2 显示了它的威力, 因为对这个例子而言, u 的取法并不明显.

> **定理 (里斯表示定理)**: 设 V 是有限维内积空间, $\varphi\in V'$, 那么存在唯一的 $u\in V$, 使得对任意的 $v\in V$, 都有
>
> $$\varphi(v)=\langle v,u\rangle\,.$$

证明 先证 u 的存在性. 设 $e_1,\cdots,e_n$ 是 V 的规范正交基, 那么对任意的 $v\in V$, 有

$$\begin{aligned}\varphi(v)&=\varphi(\langle v,e_1\rangle e_1+\cdots+\langle v,e_n\rangle e_n)\\&=\langle v,e_1\rangle\varphi(e_1)+\cdots+\langle v,e_n\rangle\varphi(e_n)\\&=\langle v,\overline{\varphi(e_1)}e_1+\cdots+\overline{\varphi(e_n)}e_n\rangle.\end{aligned}$$

取 $u=\overline{\varphi(e_1)}e_1+\cdots+\overline{\varphi(e_n)}e_n$, 则 $\varphi(v)=\langle v,u\rangle$.

再证 u 的唯一性. 设另有 $u' \in V$ 满足对任意的 $v \in V$, 有 $\varphi(v) = \langle v, u' \rangle$, 那么

$$\langle v, u' - u \rangle = 0.$$

取 $v = u' - u$, 得 $u' - u = 0$, 即 $u' = u$. □

例 3 求 $u \in \mathcal{P}_2(\mathbf{R})$, 使得对任意的 $p \in \mathcal{P}_2(\mathbf{R})$, 都有

$$\int_{-1}^{1} p(t)u(t)\,\mathrm{d}t = \int_{-1}^{1} p(t)\cos(\pi t)\,\mathrm{d}t$$

解 上式等价于 $\varphi(p) = \langle p, u \rangle$, 其中线性泛函 φ 为 $\varphi(p) = \int_{-1}^{1} p(t)\cos(\pi t)\,\mathrm{d}t$. 我们使用 §7.3 例 4 中的规范正交基

$$\frac{1}{\sqrt{2}}, \quad \sqrt{\frac{3}{2}}x, \quad \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right).$$

将它们分别记为 e_1, e_2, e_3. 由里斯表示定理的证明, 我们取

$$u = \overline{\varphi(e_1)}e_1 + \overline{\varphi(e_2)}e_2 + \overline{\varphi(e_3)}e_3$$

即可. 直接计算得

$$\varphi(e_1) = \varphi(e_2) = 0, \quad \varphi(e_3) = -\frac{4}{\pi^2}\sqrt{\frac{45}{8}},$$

从而

$$u = -\frac{45}{2\pi^2}\left(x^2 - \frac{1}{3}\right).$$

注: 我们可以通过公式

$$u = \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n$$

找到满足 $\varphi(v) = \langle v, u \rangle$ 的 u. 看起来等式右端依赖于线性泛函 φ 与规范正交基 $e_1, \cdots, e_n$, 但由里斯表示定理我们知道 u 是由 φ 唯一确定的, 它与规范正交基的选取无关.

7.5 正交补与投影定理

定义 (正交补): 设 U 是 V 的子集, 我们称

$$U^{\perp} = \{v \in V : v \perp U\}$$

为子集 U 的正交补 (orthogonal complement).

例 1 U 是 $\mathbf{R}^3$ 中过原点的直线, 则 $U^{\perp}$ 为过原点且垂直于 U 的平面.

例 2 U 是 $\mathbf{R}^3$ 中过原点的平面, 则 $U^{\perp}$ 为过原点且垂直于 U 的直线.

性质 (正交补的基本性质)：设 V 是 $\mathbf{F}$ 上的内积空间, 则

(1) 若 U 是 V 的子集, 则 $U^{\perp}$ 是 V 的子空间.

(2) $\{0\}^{\perp}=V$.

(3) $V^{\perp}=\{0\}$.

(4) 若 U 是 V 的子集, 则 $U\cap U^{\perp}\subseteq\{0\}$.

(5) 若 $U\subseteq W$ 均为 V 的子集, 则 $W^{\perp}\subseteq U^{\perp}$.

(6) 若 U 为 V 的子集, 则 $U\subseteq(U^{\perp})^{\perp}$.

证明 $U^{\perp}=\{v\in V:v\perp U\}=\{v\in V:\text{对任意的}\,u\in U,\langle v,u\rangle=0\}$.

(1) 显然 $0\in U^{\perp}$. 任取 $v_1,v_2\in U^{\perp}$, 则对任意的 $u\in U$, $\langle v_1+v_2,u\rangle=\langle v_1,u\rangle+\langle v_2,u\rangle=0$, 故 $v_1+v_2\in U^{\perp}$. 任取 $\lambda\in\mathbf{F}$ 与 $v\in U^{\perp}$, 那么 $\langle\lambda v,u\rangle=\lambda\langle v,u\rangle=0$, 故 $\lambda v\in U^{\perp}$. 即 $U^{\perp}$ 为子空间.

(2) V 中的一切元素都与 0 正交.

(3) 与 V 中所有元素都正交的元素只有 0.

(4) 当 $U\cap U^{\perp}=\varnothing$ 时, 结论成立. 当 $U\cap U^{\perp}\neq\varnothing$ 时, 任取 $v\in U\cap U^{\perp}$, 则 $v\in U$ 且 $v\in U^{\perp}$. 因此 $\langle v,v\rangle=0$, $v=0$.

(5) 任取 $v\in W^{\perp}$, 则对任意的 $w\in W$, 都有 $\langle v,w\rangle=0$. 从而对任意的 $u\in U\subseteq W$, 都有 $\langle v,u\rangle=0$. 因此 $v\in U^{\perp}$, $W^{\perp}\subseteq U^{\perp}$.

(6) 任取 $u\in U$. 对任意的 $v\in U^{\perp}=\{v\in V:v\perp U\}$, 都有 $\langle u,v\rangle=0$, 故 $u\in(U^{\perp})^{\perp}$. $\square$

以下定理表明 V 的每个有限维子空间都导致了 V 的一个自然的直和分解. 注意这里我们对 V 的维数没有要求.

定理 (投影定理)：设 U 是 V 的有限维子空间, 则 $V=U\oplus U^{\perp}$.

证明 设 $e_1,\cdots,e_m$ 是 U 的规范正交基, 则

$$v=\underbrace{\langle v,e_1\rangle e_1+\cdots+\langle v,e_m\rangle e_m}_{u}+\underbrace{v-\langle v,e_1\rangle e_1-\cdots-\langle v,e_m\rangle e_m}_{w}.$$

设 u,w 如上式, 则 $u\in U$, 下证 $w\in U^{\perp}$. 对每个 $j=1,\cdots,m$, 有

$$\langle w,e_j\rangle=\langle v,e_j\rangle-\langle v,e_j\rangle=0.$$

故 $w\perp\operatorname{span}\{e_1,\cdots,e_m\}=U$, 即 $w\in U^{\perp}$. 因此 $V=U+U^{\perp}$. 由正交补的基本性质 (4) 可知 $U\cap U^{\perp}=\{0\}$, 故 $V=U\oplus U^{\perp}$. $\square$

推论 (正交补的维数)：设 V 是有限维的, U 是 V 的子空间, 则 $\dim U^{\perp} = \dim V - \dim U$.

证明 由投影定理知 $\dim V = \dim(U \oplus U^{\perp}) = \dim U + \dim U^{\perp}$. □

下面的命题是投影定理的重要推论.

推论 (有限维子空间的二重正交补等于自身)：设 U 是 V 的有限维子空间, 则 $U = (U^{\perp})^{\perp}$.

证明 由正交补的基本性质 (6), 我们只需证明 $(U^{\perp})^{\perp} \subseteq U$. 任取 $v \in (U^{\perp})^{\perp}$, 由投影定理知存在 $u \in U$ 与 $w \in U^{\perp}$, 使得 $v = u + w$. 从而 $v - u = w \in U^{\perp}$. 注意到 $v \in (U^{\perp})^{\perp}$, $u \in U \subseteq (U^{\perp})^{\perp}$, 故

$$v - u \in (U^{\perp})^{\perp}.$$

因为 $(U^{\perp}) \cap (U^{\perp})^{\perp} = \{0\}$, 故 $v - u = 0$, $v = u \in U$, 从而 $(U^{\perp})^{\perp} \subseteq U$. □

7.6 正交投影

现在我们对 V 的每一个有限维子空间都定义一个线性变换 P_U.

定义 (正交投影)：设 U 是 V 的有限维子空间, 定义 V 到 U 上的正交投影 (orthogonal projection) 为线性变换 $P_U \in \mathcal{L}(V)$: 对 $v \in V$, 将其写为 $v = u + w$, 其中 $u \in U$, $w \in U^{\perp}$, 那么定义 $P_U(v) = u$. 当不至于引起混淆时, 将 P_U 简记为 P.

由投影定理可知对每个 $v \in V$, 存在唯一的 $u \in U$ 与 $w \in U^{\perp}$, 使得 $v = u + w$, 因此 P_U 的定义是合理的.

例 1 设 $0 \neq x \in V$, $U = \operatorname{span}\{x\}$, 证明: 对任意的 $v \in V$, v 在 U 上的正交投影为

$$Pv = \frac{\langle v, x\rangle}{\|x\|^2} x.$$

证明 U 是 V 的一维子空间. 由于

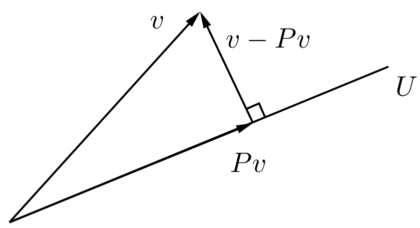
$$v = \frac{\langle v, x\rangle}{\|x\|^2} x + \left(v - \frac{\langle v, x\rangle}{\|x\|^2} x\right),$$

其中右端第一项属于 U, 第二项属于 $U^{\perp}$, 因此 Pv 等于右边第一项. □

性质 (正交投影的性质)：设 U 是 V 的有限维子空间, $v \in V$, 则

(1) $P \in \mathcal{L}(V)$.

(2) 对任意的 $u \in U$, 有 $Pu = u$.



(3) 对任意的 $w \in U^{\perp}$, 有 $Pw = 0$.

(4) $\operatorname{Im} P = U$.

(5) $\operatorname{Ker} P = U^{\perp}$.

(6) $v - Pv \in U^{\perp}$.

(7) $P^2 = P$.

(8) $\|Pv\| \leqslant \|v\|$.

(9) 对 U 的任一规范正交基 $e_1, \cdots, e_m$, 有 $Pv = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$.

证明 由投影定理知 $V = U \oplus U^{\perp}$. 我们约定下面出现的 u 总是属于 U, w 总是属于 $U^{\perp}$.

(1) 设 $v_1 = u_1 + w_1, v_2 = u_2 + w_2$, 则 $v_1 + v_2 = (u_1 + u_2) + (w_1 + w_2)$, 故 $P(v_1 + v_2) = u_1 + u_2 = Pv_1 + Pv_2$. 设 $v = u + w$, 则 $kv = ku + kw$, 故 $P(kv) = kPv$. 因此 P 是线性的.

(2) $u = u + 0$, 故 $Pu = u$.

(3) 设 $w = u + w'$, 则 $w - w' = u \in U$, 又 $w - w' \in U^{\perp}$, 故 $w - w' = 0$, $w = w'$, 故 $u = 0$, 即 $Pw = 0$.

(4) 由定义知 $\operatorname{Im} P \subseteq U$, 故只需证明 $U \subseteq \operatorname{Im} P$. 任取 $u \in U$, 由 (2) 知 $Pu = u$, 故 $u \in \operatorname{Im} P$, 从而 $U \subseteq \operatorname{Im} P$.

(5) 由 (3) 知 $U^{\perp} \subseteq \operatorname{Ker} P$, 只需证 $\operatorname{Ker} P \subseteq U^{\perp}$. 任取 $v \in \operatorname{Ker} P$, 则 $Pv = 0$, 故 $v = 0 + w$, 从而 $v = w \in U^{\perp}$, 故 $\operatorname{Ker} P \subseteq U^{\perp}$.

(6) 由定义知 $v = Pv + w$, 故 $v - Pv = w \in U^{\perp}$.

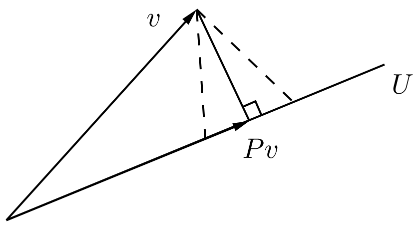
(7) 由 (2) 知对任意的 $u \in U$, 有 $P(Pu) = Pu$, 故 $P^2 = P$.

(8) 因为 $v = Pv + w$, 且 $Pv \perp w$, 故 $\|v\|^2 = \|Pv + w\|^2 = \|Pv\|^2 + \|w\|^2 \geqslant \|Pv\|^2$.

(9) $v = \underbrace{\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m}_{Pv} + \underbrace{v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m}_{w}$. □

7.7 极小化问题

我们经常会遇到这样的问题: 给定 V 的子空间 U 和点 $v \in V$, 求点 $u \in U$, 使得 $\|u - v\|$ 最小. 下面的命题表明通过取 $u = Pv$ 就可以解决这一问题.



Pv 是 U 中离 v 最近的点

命题 (向量到子空间的最小距离)：设 U 是 V 的有限维子空间，$v \in V$，那么对任意的 $u \in U$，有

$$\|v - Pv\| \leqslant \|v - u\|,$$

其中等号成立当且仅当 $u = Pv$.

证明 由 $v - Pv \in U^{\perp}$ 与勾股定理得

$$\|v-u\|^2 = \|\underbrace{(v-Pv)}_{U^{\perp}} + \underbrace{(Pv-u)}_{U}\|^2 = \|v-Pv\|^2 + \|Pv-u\|^2 \geqslant \|v-Pv\|^2.$$

等号成立当且仅当 $\|Pv-u\| = 0$, 也即 $u = Pv$. $\square$

上述命题通常与正交投影的性质 (9): 对 U 的任一规范正交基 $e_1, \cdots, e_m$, 有

$$Pv = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$$

结合起来计算极小化问题的显式解.

例 1 求一个次数不超过 5 的实系数多项式 u, 使其在区间 $[-\pi, \pi]$ 上尽量好地逼近 $\sin x$, 即使得

$$\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \, \mathrm{d}x$$

最小, 并将你的结果与泰勒级数逼近进行比较.

解 设 $C_{\mathbf{R}}[-\pi, \pi]$ 是区间 $[-\pi, \pi]$ 是上全体实值连续函数所成的内积空间, 其内积定义为

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x) \, \mathrm{d}x.$$

又设 $v \in C_{\mathbf{R}}[-\pi, \pi]$ 是函数 $\sin x$. 令 U 表示区间 $[-\pi, \pi]$ 是上全体次数不超过 5 的实系数多项式所成的内积空间, 它是 $C_{\mathbf{R}}[-\pi, \pi]$ 的子空间. 于是问题可以重新表述为: 寻找多项式 $u \in U$, 使得

$$\|v - u\|$$

最小. 要解决这一问题, 首先对基 $1, x, x^2, x^2, x^4, x^5$ 进行格拉姆-施密特正交化得到规范正交基

$e_1, e_2, e_3, e_4, e_5, e_6$. 然后利用公式

$$Pv = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$$

求出 Pv, 取 $u = Pv$ 即可. 计算可得近似值

$$u(x) = 0.987862x - 0.155271x^3 + 0.00564312x^5.$$

通过作图可以发现在 $[-\pi, \pi]$ 上 $u(x)$ 与 $\sin x$ 的图像几乎完全重合, 肉眼难以看出差别. 与之相比, 熟知的泰勒多项式 $x - \frac{x^3}{3!} + \frac{x^5}{5!}$ 在 $|x| > 2$ 时并不怎么精确. 线性代数帮助我们发现了 $\sin x$ 的一个极好的逼近, 这个逼近改进了我们在微积分中学到的逼近!

8 内积空间上的算子

我们现在要讨论内积空间上的线性映射, 这方面的研究成果在内积空间的理论中最为深刻. 我们将利用伴随的性质详细描述内积空间上的几类重要线性映射.

本章使用 $\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$, 用 V 与 W 表示 $\mathbf{F}$ 上的有限维内积空间. 现在开始, 我们把线性映射也叫做线性算子.

8.1 伴随与共轭转置

再次提醒一下, 根据我们约定的记号, 本章中出现的 V 与 W 都是有限维的.

定义 (伴随): 设 $T \in \mathcal{L}(V, W)$, 我们称满足如下条件的算子 $T^* : W \to V$ 为 T 的伴随 (adjoint): 对任意的 $v \in V$ 与 $w \in W$, 都有

$$\langle Tv, w \rangle = \langle v, T^* w \rangle.$$

注: 伴随这个词在线性代数中还有另一个意思. 在本书中我们不需要第二种意思. 如果你在别的地方遇到伴随的另一种意思, 要注意伴随的这两种意思之间没有任何联系.

我们使用里斯表示定理来验证上述定义的合理性, 即 T^* 是存在且唯一的. 为此, 取定 $w \in W$, 定义线性泛函 $\varphi_w : V \to \mathbf{F}$ 为 $\langle Tv, w \rangle$. 由里斯表示定理知 V 中存在唯一的向量 (记为 $T^* w$), 满足

$$\langle Tv, w \rangle = \varphi_w(v) = \langle v, T^* w \rangle.$$

也就是说对取定的 $w \in W$, $T^* w$ 是 V 中唯一一个满足对任意的 $v \in V$, $\langle Tv, w \rangle = \langle v, T^* w \rangle$ 的向量.

例 1 定义 $T : \mathbf{R}^3 \to \mathbf{R}^2$ 为 $T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1)$, 求 T^*.

解　任取 $(x_1, x_2, x_3) \in \mathbf{R}^3, (y_1, y_2) \in \mathbf{R}^2$, 则

$$
\begin{aligned}
\langle (x_1, x_2, x_3), T^*(y_1, y_2) \rangle &= \langle T(x_1, x_2, x_3), (y_1, y_2) \rangle \\
&= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2) \rangle \\
&= x_2 y_1 + 3x_3 y_1 + 2x_1 y_2 \\
&= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1) \rangle .
\end{aligned}
$$

因此 $T^*(y_1, y_2) = (2y_2, y_1, 3y_1)$.

例 2　取定 $u \in V$ 与 $x \in W$, 定义 $T \in \mathcal{L}(V, W)$ 为 $Tv = \langle v, u \rangle x$, 求 T^*.

解

$$
\begin{aligned}
\langle v, T^* w \rangle &= \langle Tv, w \rangle \\
&= \langle \langle v, u \rangle x, w \rangle \\
&= \langle v, u \rangle \langle x, w \rangle \\
&= \langle v, \overline{\langle x, w \rangle} u \rangle \\
&= \langle v, \langle w, x \rangle u \rangle .
\end{aligned}
$$

因此 $T^* w = \langle w, x \rangle u$.

上面两个例题中的 T^* 都是线性的, 我们将要证明这是一个普遍成立的结论. 注意下面两个命题的证明都用到了一个技巧: 将 T^* 从内积的一个向量转移到另一个向量上就变为 T.

命题 (伴随是线性映射)：设 $T \in \mathcal{L}(V, W)$, 那么 $T^* \in \mathcal{L}(W, V)$.

性质 (伴随的基本性质)：设 U, V, W 均为 $\mathbf{F}$ 上的有限维内积空间, 则

(1) 对任意的 $S, T \in \mathcal{L}(V, W)$, 有 $(S + T)^* = S^* + T^*$.

(2) 对任意的 $T \in \mathcal{L}(V, W)$ 与 $\lambda \in \mathbf{F}$, 有 $(\lambda T)^* = \bar{\lambda} T^*$.

(3) 对任意的 $T \in \mathcal{L}(V, W)$, 有 $(T^*)^* = T$.

(4) $I^* = I$, 这里 I 是 V 上的恒等映射.

(5) 对任意的 $T \in \mathcal{L}(V, W)$ 与 $S \in \mathcal{L}(W, U)$, 有 $(ST)^* = T^* S^*$.

以下命题描述了线性映射与其伴随的核与像之间的关系.

命题 (伴随的核与像)：设 $T \in \mathcal{L}(V, W)$, 则

(1) $\operatorname{Ker} T^* = (\operatorname{Im} T)^\perp$.

(2) $\operatorname{Im} T^* = (\operatorname{Ker} T)^\perp$.

(3) $\operatorname{Ker} T = (\operatorname{Im} T^*)^\perp$.

(4) $\operatorname{Im} T = (\operatorname{Ker} T^*)^{\perp}$.

证明 (1) $T^* \in \mathcal{L}(W, V)$,

$$\begin{aligned}
\operatorname{Ker} T^* &= \{w \in W : T^*w = 0\} \\
&= \{w \in W : \text{对任意的 } v \in V, \langle v, T^*w\rangle = 0\} \\
&= \{w \in W : \text{对任意的 } v \in V, \langle Tv, w\rangle = 0\} \\
&= \{w \in W : w \perp \operatorname{Im} T\} \\
&= (\operatorname{Im} T)^{\perp}.
\end{aligned}$$

将 (1) 中的 T 换为 T^*, 并利用 $(T^*)^* = T$ 即得 (3). 将 (1) 的等式两边取正交补, 并利用 $((\operatorname{Im} T)^{\perp})^{\perp} = \operatorname{Im} T$ 即得 (4). 将 (4) 中的 T 换为 T^*, 并利用 $(T^*)^* = T$ 即得 (2). □

定义 (**矩阵的共轭转置**)：将 $m \times n$ 矩阵 A 转置, 再对所得矩阵的每个元素取复共轭得到的 $n \times m$ 矩阵称为 A 的共轭转置 (conjugate transpose), 记为 A^H.

若 $\mathbf{F} = \mathbf{R}$, 则 A 的共轭转置就等于 A 的转置.

以下命题说明了怎样通过 T 的矩阵来计算 T^* 的矩阵. 注意下面的结果只能对规范正交基使用, 对于非规范正交基, T^* 的矩阵未必等于 T 的矩阵的共轭转置.

命题 (**伴随与共轭转置**)：设 $T \in \mathcal{L}(V, W)$, $e_1, \cdots, e_n$ 是 V 的规范正交基, $f_1, \cdots, f_m$ 是 W 的规范正交基, 则

$$\mathcal{M}(T^*) = (\mathcal{M}(T))^H.$$

证明 记 $\mathcal{M}(T)$ 为 A, $\mathcal{M}(T^*)$ 为 B, 那么

$$(Te_1, \cdots, Te_n) = (f_1, \cdots, f_m)A,$$
$$(T^*f_1, \cdots, T^*f_m) = (e_1, \cdots, e_n)B,$$

从而

$$Te_j = a_{1j}f_1 + \cdots + a_{mj}f_m,$$
$$T^*f_i = b_{1i}e_1 + \cdots + b_{ni}e_n.$$

由规范正交基的性质可知 $a_{ij} = \langle Te_j, f_i\rangle$, $b_{ji} = \langle T^*f_i, e_j\rangle$, 故

$$a_{ij} = \langle Te_j, f_i\rangle = \langle e_j, T^*f_i\rangle = \overline{\langle T^*f_i, e_j\rangle} = \overline{b_{ji}}.$$

故 $b_{ji} = \overline{a_{ij}}$, $B = A^H$, 结论得证. □

8.2 自伴算子

现在我们考察 V 到 V 自身的线性映射.

定义 (自伴算子)：设 $T \in \mathcal{L}(V)$, 如果 $T = T^*$, 那么我们称 T 是自伴的 (self-adjoint).

根据定义, T 是自伴的当且仅当对任意的 $v, w \in V$, 都有

$$\langle Tv, w\rangle = \langle v, Tw\rangle .$$

例 1 设 T 是 $\mathbf{F}^2$ 上的算子, 它关于标准基的矩阵为

$$\begin{pmatrix} 2 & b \\ 3 & 7 \end{pmatrix}.$$

求数 b 使得 T 是自伴的.

解 因为 $T^* = T$, 故 $\mathcal{M}(T^*) = \mathcal{M}(T)$, 又因为标准基是规范正交基, 故 $\mathcal{M}(T^*) = (\mathcal{M}(T))^H$. 因此 $\mathcal{M}(T) = (\mathcal{M}(T))^H$, 解得 $b = 3$.

请自行验证, 自伴算子的和是自伴算子, 实数与自伴算子的数乘是自伴算子.

伴随在 $\mathcal{L}(V)$ 上所起的作用类似于复共轭在 $\mathbf{C}$ 上所起的作用. 复数 z 是实数当且仅当 $z = \bar{z}$, 因此自伴算子可与实数类比.

我们将看到, 这种类比也反映在自伴算子的某些重要性质上, 先来看特征值.

若 $\mathbf{F} = \mathbf{R}$, 那么特征值自然是实数, 因此下面的命题只有当 $\mathbf{F} = \mathbf{C}$ 时才有意思.

命题 (自伴算子的特征值是实的)：$\mathbf{C}$ 上自伴算子的特征值都是实数.

证明 设 T 是 V 上的自伴算子, λ 是 T 的特征值, $v \neq 0$ 是相应于 λ 的特征向量, 则

$$\lambda \|v\|^2 = \lambda \langle v, v\rangle = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \bar{\lambda} \langle v, v\rangle = \bar{\lambda} \|v\|^2 .$$

故 $\bar{\lambda} = \lambda$, 因此 λ 是实数. □

下面的命题对实内积空间不成立. 例如, 考虑 $T \in \mathcal{L}(\mathbf{R}^2)$ 是 $\mathbf{R}^2$ 中绕原点逆时针旋转 90° 的变换 $T(x, y) = (-y, x)$. 显然对每个 $v \in \mathbf{R}^2$ 都有 Tv 正交于 v, 但 $T \neq 0$.

命题 (在 $\mathbf{C}$ 上只有零算子才能使得 Tv 总正交于 v)：设 V 是复内积空间, $T \in \mathcal{L}(V)$, 如果对任意的 $v \in V$, 都有 $\langle Tv, v\rangle = 0$, 那么 $T = 0$.

证明 任取 $v, w \in V$ 与 $\lambda \in \mathbf{C}$, 则

$$
\begin{aligned}
0 &= \langle T(\lambda v + w), \lambda v + w\rangle \\
&= \lambda\bar{\lambda}\langle Tv, v\rangle + \lambda\langle Tv, w\rangle + \bar{\lambda}\langle Tw, v\rangle + \langle Tw, w\rangle \\
&= \lambda\langle Tv, w\rangle + \bar{\lambda}\langle Tw, v\rangle .
\end{aligned}
$$

取 $\lambda = \mathrm{i}$, 得

$$\langle Tv, w\rangle - \langle Tw, v\rangle = 0.$$

取 $\lambda = 1$, 得

$$\langle Tv, w\rangle + \langle Tw, v\rangle = 0.$$

上述两式相加即得 $\langle Tv, w\rangle = 0$. 取 $w = Tv$ 得 $Tv = 0$, 从而 $T = 0$. □

尽管在实内积空间上, 非零算子可能会使得对任意的 $v \in V$, 都有 $\langle Tv, v\rangle = 0$, 但是以下命题表明对于非零自伴算子, 这种情况不会发生.

命题 (**$\mathbf{R}$ 上满足 $\langle Tv, v\rangle = 0$ 的自伴算子一定是零算子**) : 设 V 是实内积空间, $T \in \mathcal{L}(V)$ 是自伴算子. 若对任意的 $v \in V$, 都有 $\langle Tv, v\rangle = 0$, 那么 $T = 0$.

证明 任取 $v, w \in V$ 与 $\lambda \in \mathbf{R}$, 则

$$
\begin{aligned}
0 &= \langle T(\lambda v + w), \lambda v + w\rangle \\
&= \lambda^2\langle Tv, v\rangle + \lambda\langle Tv, w\rangle + \lambda\langle Tw, v\rangle + \langle Tw, w\rangle \\
&= \lambda\langle Tv, w\rangle + \lambda\langle Tw, v\rangle .
\end{aligned}
$$

取 $\lambda = 1$, 得

$$\langle Tv, w\rangle + \langle Tw, v\rangle = 0.$$

由于 T 是自伴的, 故

$$\langle v, Tw\rangle + \langle Tw, v\rangle = 2\langle Tw, v\rangle = 0.$$

因此 $\langle Tv, w\rangle = 0$. 取 $w = Tv$ 得 $Tv = 0$, 从而 $T = 0$. □

注 : 上面两个结论告诉我们: $\mathbf{C}$ 上满足 $\langle Tv, v\rangle = 0$ 的算子一定是零, $\mathbf{R}$ 上满足 $\langle Tv, v\rangle = 0$ 的自伴算子一定是零. 它们的证明过程非常类似.

命题 (**自伴算子的刻画**) : 设 V 是复内积空间, $T \in \mathcal{L}(V)$, 则 T 是自伴算子当且仅当对任意的 $v \in V$ 都有 $\langle Tv, v\rangle \in \mathbf{R}$.

证明 $(\Longrightarrow)$ 对任意的 $v \in V$, 有

$$\langle Tv, v\rangle = \langle v, Tv\rangle = \overline{\langle Tv, v\rangle}.$$

故 $\langle Tv, v\rangle$ 是实数.

$(\Longleftarrow)$ 因为 $\langle Tv, v\rangle$ 是实数, 故

$$0 = \langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = \langle Tv, v\rangle - \langle v, Tv\rangle = \langle Tv, v\rangle - \langle T^*v, v\rangle = \langle (T - T^*)v, v\rangle .$$

由前一命题可知 $T - T^* = 0$, 故 $T = T^*$, 即 T 是自伴的. □

8.3 正规算子

定义 (正规算子): 设 $T \in \mathcal{L}(V)$, 若 T 满足满足 $TT^* = T^*T$, 则称 T 是正规的 (normal).

自伴算子显然是正规的, 因为对自伴算子而言 $T^* = T$, 从而 $TT^* = T^2 = T^*T$.

例 1 设 $T \in \mathcal{L}(\mathbf{F}^2)$ 关于标准基的矩阵为

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

证明 T 是正规的但 T 不是自伴的.

证明 注意到标准基是规范正交基, 故 $\mathcal{M}(T^*) = (\mathcal{M}(T))^H$, 故

$$\mathcal{M}(T^*) = \begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}.$$

从而

$$\mathcal{M}(TT^*) = \mathcal{M}(T)\mathcal{M}(T^*) = \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix} = \mathcal{M}(T^*)\mathcal{M}(T) = \mathcal{M}(T^*T).$$

故 $TT^* = T^*T$. 但是由于 $\mathcal{M}(T) \neq \mathcal{M}(T^*)$, 故 $T \neq T^*$, 即 T 不是自伴的. □

在下一节我们会看到为什么正规算子值得特别关注.

以下命题给出了正规算子的简单刻画, 它说明对每个正规算子均有 $\operatorname{Ker} T = \operatorname{Ker} T^*$.

命题 (正规算子的刻画): 设 $T \in \mathcal{L}(V)$, 则 T 是正规算子当且仅当对任意的 $v \in V$, 均有

$$\|Tv\| = \|T^*v\|.$$

证明　我们将同时证明这个结论的两个方面.

$$
\begin{aligned}
T\text{是正规算子} &\iff TT^* - T^*T = 0 \\
&\iff \text{对任意的 } v \in V,\text{ 都有 } \langle (TT^* - T^*T)v, v\rangle = 0 \\
&\iff \text{对任意的 } v \in V,\text{ 都有 } \langle TT^*v, v\rangle = \langle T^*Tv, v\rangle \\
&\iff \text{对任意的 } v \in V,\text{ 都有 } \langle T^*v, T^*v\rangle = \langle Tv, Tv\rangle \\
&\iff \text{对任意的 } v \in V,\text{ 都有 } \|T^*v\| = \|Tv\|.
\end{aligned}
$$

其中第二个等价符号是因为 $TT^* - T^*T$ 是自伴的. □

把下面的命题与习题 1. 相比较, 那个习题说得是算子的伴随的全体特征值 (作为集合) 等于该算子全体特征值的复共轭. 但是, 那个习题没有说特征向量的事, 这是因为算子与其伴随可以有不同的特征向量. 然而, 由下面的命题可知, 正规算子与其伴随有相同的特征向量.

命题 (正规算子有相同的特征向量): 设 $T \in \mathcal{L}(V)$ 为正规算子. 若 v 是 T 的相应于特征值 λ 的特征向量, 那么 v 也是 T^* 的相应于特征值 $\bar{\lambda}$ 的特征向量.

证明　设 v 是 T 的相应于特征值 λ 的特征向量. 因为 T 正规, 所以 $T-\lambda I$ 也正规 (请自行验证), 故

$$
0 = \|(T-\lambda I)v\| = \|(T-\lambda I)^*v\| = \left\|(T^* - \bar{\lambda} I)v\right\|,
$$

从而 v 是 T^* 的相应于特征值 $\bar{\lambda}$ 的特征向量. 这里第二个等号利用了正规算子的特点. □

因为自伴算子是正规的, 所以下面的命题也适用于自伴算子.

命题 (正规算子属于不同特征值的特征向量彼此正交): 设 $T \in \mathcal{L}(V)$ 是正规的, 则 T 的相应于不同特征值的特征向量是正交的.

证明　设 α, β 是 T 的不同特征值, u, v 分别为 T 的相应于 α, β 的特征向量, 则 $Tu = \alpha u$ 且 $Tv = \beta v$. 由上一命题知 $T^*v = \bar{\beta} v$, 故

$$
\begin{aligned}
(\alpha - \beta)\langle u, v\rangle &= \alpha\langle u, v\rangle - \beta\langle u, v\rangle \\
&= \langle \alpha u, v\rangle - \left\langle u, \bar{\beta} v\right\rangle \\
&= \langle Tu, v\rangle - \langle u, T^*v\rangle \\
&= 0.
\end{aligned}
$$

这里最后一个等号利用的是伴随的定义. 因为 $\alpha \neq \beta$, 故 $\langle u, v\rangle = 0$. □

8.4 复谱定理

回忆一下, V 上的某个算子关于某个基有对角矩阵当且仅当这个基是由算子的特征向量构成的.

关于 V 上的某个规范正交基有对角矩阵的算子是 V 上最好的算子, 因为它们具有如下性质: V 有一个由 T 的特征向量构成的规范正交基. 我们后面的目标是证明谱定理. 谱定理表明: 具有上述性质的算子当 $\mathbf{F}=\mathbf{C}$ 时恰为正规算子, 当 $\mathbf{F}=\mathbf{R}$ 时恰为自伴算子. 我们可以认为谱定理是研究内积空间上算子的最有用的工具.

因为谱定理的结论依赖于 $\mathbf{F}$, 所以我们把谱定理分成两部分——复谱定理与实谱定理. 同线性代数中的大多数情形一样, 处理复线性空间要比处理实线性空间容易, 因此我们先给出复谱定理.

复谱定理主要是说, 在 $\mathbf{C}$ 上若 $T\in\mathcal{L}(V)$ 是正规的, 那么 T 就关于 V 的某个规范正交基有对角矩阵.

例 1 设 $T\in\mathcal{L}(\mathbf{C}^2)$ 关于标准基的矩阵为

$$\begin{pmatrix}2 & -3\\ 3 & 2\end{pmatrix}.$$

请自行验证 T 有特征向量 $\dfrac{1}{\sqrt{2}}(\mathrm{i},1),\dfrac{1}{\sqrt{2}}(-\mathrm{i},1)$, 它们构成 $\mathbf{C}^2$ 的规范正交基, 并且 T 关于这个基有对角矩阵

$$\begin{pmatrix}2+3\mathrm{i} & 0\\ 0 & 2-3\mathrm{i}\end{pmatrix}.$$

在下面的定理中, (2)(3) 的等价性是容易的, 因此我们只需证明 (1) 与 (3) 的等价性即可.

定理 (复谱定理): 设 $\mathbf{F}=\mathbf{C}$, $T\in\mathcal{L}(V)$, 那么下面各条等价:

(1) T 是正规的.

(2) V 有由 T 的特征向量构成的规范正交基.

(3) T 关于某个规范正交基有对角矩阵.

证明 (3) $\Longrightarrow$ (1) 设 T 在某个规范正交基下有对角矩阵, 那么由 $\mathcal{M}(T^*)=(\mathcal{M}(T))^H$ 知 T^* 在该规范正交基下也有对角矩阵. 因为对角矩阵是可交换的, 因此 $\mathcal{M}(T)\mathcal{M}(T^*)=\mathcal{M}(T^*)\mathcal{M}(T)$, 从而 $\mathcal{M}(TT^*)=\mathcal{M}(T^*T)$, 进而 $TT^*=T^*T$, 即 T 是正规的.

(1) $\Longrightarrow$ (3) $\mathbf{C}$ 上的线性算子必关于某个基有上三角矩阵, 从而由 §7.3 最后一个命题可知 T 关

于某个规范正交基 $e_1, \cdots, e_n$ 有上三角矩阵, 即

$$(Te_1, \cdots, Te_n) = (e_1, \cdots, e_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix},$$

进而

$$(T^*e_1, \cdots, T^*e_n) = (e_1, \cdots, e_n) \begin{pmatrix} \overline{a_{11}} & 0 & \cdots & 0 \\ \overline{a_{12}} & \overline{a_{22}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \overline{a_{1n}} & \overline{a_{2n}} & \cdots & \overline{a_{nn}} \end{pmatrix}.$$

我们将证明上三角矩阵 $A = (a_{ij})$ 实际上是个对角矩阵. 注意到

$$\begin{aligned} \|Te_1\|^2 &= |a_{11}|^2, \\ \|T^*e_1\|^2 &= |a_{11}|^2 + |a_{12}|^2 + \cdots + |a_{1n}|^2, \end{aligned}$$

由正规算子的特点 $\|Te_1\| = \|T^*e_1\|$ 可得 A 的第一行中 a_{11} 之外的元素均为 0. 进一步, 由 $\|Te_2\| = \|T^*e_2\|$ 以及

$$\begin{aligned} \|Te_2\|^2 &= |a_{22}|^2, \\ \|T^*e_2\|^2 &= |a_{22}|^2 + |a_{23}|^2 + \cdots + |a_{1n}|^2 \end{aligned}$$

可得 A 的第二行 a_{22} 之外的元素全为 0. 如此不断重复下去, A 的对角线之外的元素均为 0, A 为对角矩阵. □

8.5 实谱定理

为了证明实谱定理, 我们需要几个引理.

你可能会猜到下面的引理, 甚至可能通过考虑实系数的二次多项式给出其证明具体来说, 设 $b, c \in \mathbf{R}$ 且 $b^2 - 4c < 0$, x 为实数, 那么

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right) > 0.$$

特别地, $x^2 + bx + c$ 是 "可逆实数". 用自伴算子代替实数 x(回想一下自伴算子和实数之间的类比), 可以得出下面的引理.

引理 (可逆的二次式): 设 $T \in \mathcal{L}(V)$ 是自伴的, $b, c \in \mathbf{R}$ 且 $b^2 - 4c < 0$, 则

$$T^2 + bT + cI$$

是可逆的.

证明 任取 $0 \neq v \in V$, 则

$$\begin{aligned}\left\langle (T^2 + bT + cI)v, v \right\rangle &= \left\langle T^2 v, v \right\rangle + b\left\langle Tv, v\right\rangle + c\left\langle v, v\right\rangle \\ &= \|Tv\|^2 + b\left\langle Tv, v\right\rangle + c\|v\|^2 \\ &\geqslant \|Tv\|^2 - |b|\,\|Tv\|\,\|v\| + c\|v\|^2 \\ &= \left(\|Tv\| - \frac{|b|\,\|v\|}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)\|v\|^2 \\ &> 0.\end{aligned}$$

其中大于等于号那里利用了 $\langle Tv, v\rangle$ 是实数. 因此 $(T^2 + bT + cI)v \neq 0$, 故 $T^2 + bT + cI$ 是单的, 因而是可逆的. □

我们已经知道, 有限维复线性空间上的线性算子必有特征值, 因此下面的引理仅对实内积空间是新的.

引理 (自伴算子必有特征值): 设 $V \neq \{0\}$, $T \in \mathcal{L}(V)$ 为自伴算子, 则 T 有特征值.

证明 设 V 是实内积空间, $\dim V = n$, 任取 $0 \neq v \in V$, 则 $n+1$ 个向量

$$v, Tv, \cdots, T^n v$$

是线性相关的, 从而存在不全为 0 的实数 $a_0, a_1, \cdots, a_n$, 使得

$$0 = a_0 v + a_1 Tv + \cdots + a_n T^n v.$$

因为 $v \neq 0$, 故 $a_1, \cdots, a_n$ 一定不全为 0. 由 $\mathbf{R}$ 上多项式的分解, 可设

$$a_0 + a_1 x + \cdots + a_n x^n = c(x^2 + b_1 x + c_1)\cdots(x^2 + b_M x + c_M)(x - \lambda_1)\cdots(x - \lambda_m).$$

其中 $c \neq 0$, b_j, c_j, λ_j 均为实数, 且 $b_j^2 - 4c_j < 0$. 从而

$$\begin{aligned}0 &= (a_0 I + a_1 T + \cdots + a_n T^n)v \\ &= c(T^2 + b_1 T + c_1 I)\cdots(T^2 + b_M T + c_M I)(T - \lambda_1 I)\cdots(T - \lambda_m I)v\end{aligned}$$

由前述引理知每个 $T^2+b_jT+c_jI$ 都是可逆的, 又 $c\neq 0$, 故 $m>0$, 且

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

从而至少有一个 λ_j 使得 $T-\lambda_j I$ 不是单的, 故 T 有特征值. □

下面的引理表明, 若 U 是 V 在自伴算子 T 下的不变子空间, 那么 $U^\perp$ 也在 T 下不变. 以后我们会证明条件 " T 是自伴的" 可以换为更弱的 " T 是正规的".

引理：设 $T\in\mathcal{L}(V)$ 是自伴算子, U 是 V 的在 T 下不变的子空间, 那么
(1) $U^\perp$ 也在 T 下不变.
(2) $T|_U\in\mathcal{L}(U)$ 是自伴的.
(3) $T|_{U^\perp}\in\mathcal{L}(U^\perp)$ 是自伴的.

证明 (1) 任取 $v\in U^\perp$. 对任意的 $u\in U$, 由于 $Tu\in U$, 故

$$\langle Tv,u\rangle=\langle v,Tu\rangle=0,$$

因此 $Tv\in U^\perp$, 即 $U^\perp$ 在 T 下不变.

(2) 对任意的 $u,v\in U$, 有

$$\langle (T|_U)u,v\rangle=\langle Tu,v\rangle=\langle u,Tv\rangle=\langle u,(T|_U)v\rangle,$$

故 $T|_U\in\mathcal{L}(U)$ 是自伴的.

(3) 由 (1) 我们可以将 (2) 中的 U 换为 $U^\perp$, 从而 (3) 得证. □

定理 (实谱定理)：设 $\mathbf{F}=\mathbf{R}$, $T\in\mathcal{L}(V)$, 则下面各条等价:
(1) T 是自伴的.
(2) V 有由 T 的特征向量构成的规范正交基.
(3) T 关于某个规范正交基有对角矩阵.

证明 我们将按照 (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (1) 的顺序证明结论.

(1) $\Longrightarrow$ (2) 我们对 $\dim V$ 使用数学归纳法. 若 $\dim V=1$, 结论显然成立. 现在设 $\dim V>1$, 且结论对所有维数小于 $\dim V$ 的实内积空间均成立. 下面证明结论对 V 也成立.

因为 T 是自伴的, 故 T 有特征值. 设 u 是 T 的一个特征向量且 $\|u\|=1$. 令 $U=\operatorname{span}\{u\}$, 则 U 是 V 的一维 T-不变子空间. 由上一引理, $T|_{U^\perp}\in\mathcal{L}(U^\perp)$ 是自伴的.

由归纳法假设, $U^\perp$ 有一个由 $T|_{U^\perp}$ 的特征向量构成的规范正交基. 把 u 添加到 $U^\perp$ 的规范正交基, 就得到了 V 的一个由 T 的特征向量构成的规范正交基.

(2) $\Longrightarrow$ (3) 显然.

(3) $\Longrightarrow$ (1) 设 T 关于规范正交基 $e_1, \cdots, e_n$ 有对角矩阵, 则 $\mathcal{M}(T^*) = (\mathcal{M}(T))^H = \mathcal{M}(T)$, 因此 $T = T^*$, 即 T 是自伴的. □

例 1 考虑 $\mathbf{R}^3$ 上的自伴算子 T, 它关于标准基的矩阵为

$$\begin{pmatrix} 14 & -13 & 8 \\ -13 & 14 & 8 \\ 8 & 8 & -7 \end{pmatrix}.$$

请自行验证

$$\frac{1}{\sqrt{2}}(1,-1,0), \quad \frac{1}{\sqrt{3}}(1,1,1), \quad \frac{1}{\sqrt{6}}(1,1,-2)$$

是 $\mathbf{R}^3$ 中由 T 的特征向量构成的规范正交基, 且 T 关于这个基的矩阵是对角矩阵

$$\begin{pmatrix} 27 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & -15 \end{pmatrix}.$$

若 $\mathbf{F} = \mathbf{C}$, 则复谱定理给出了正规算子的完全描述. 同时, 由于自伴算子是正规算子, 因此我们也得到了此时自伴算子的完全描述.

若 $\mathbf{F} = \mathbf{R}$, 则实谱定理给出了自伴算子的完全描述. 在后面我们将给出 $\mathbf{F} = \mathbf{R}$ 时正规算子的完全描述.

8.6 半正定算子

定义 (半正定算子): 设 $T \in \mathcal{L}(V)$. 若 T 是自伴的, 且对任意的 $v \in V$, 都有 $\langle Tv, v\rangle \geqslant 0$, 则称 T 是半正定算子.

若 V 是复内积空间, 则 T 是自伴的这一条件可以从上面的定义中去掉, 因为复内积空间上满足 $\langle Tv, v\rangle \in \mathbf{R}$ 的算子一定是自伴的.

注: 有些数学家采用 "正算子"(positive operator) 这个术语, 它的意思与半正定算子一样. 实际上, 半正定算子对应于 $[0, +\infty)$ 中的数, 因此在术语上更应该称为非负的, 而不是称为正的. 但是算子理论学家始终称之为正算子, 因此正算子这一称谓其实是一种传统.

例 1 (**半正定算子**)

(1) (**正交投影**) 设 U 是 V 的有限维子空间, 则正交投影 P_U 是半正定算子. (请自行验证)

(2) (**可逆的二次式**) 设 $T \in \mathcal{L}(V)$ 是自伴的, $b, c \in \mathbf{R}$ 且 $b^2 - 4c < 0$, 则 $T^2 + bT + cI$ 是半正定算子. 证明过程见 §8.5 的第一个引理.

定义 (**平方根**) : 若算子 R 满足 $R^2 = T$, 则称 R 为算子 T 的平方根 (square root).

例 2 设 $T \in \mathcal{L}(\mathbf{F}^3)$ 定义为 $T(z_1, z_2, z_3) = (z_3, 0, 0)$, 那么定义为 $R(z_1, z_2, z_3) = (z_2, z_3, 0)$ 的 $R \in \mathcal{L}(\mathbf{F}^3)$ 就是 T 的平方根.

下面的定理对半正定算子的刻画与 $\mathbf{C}$ 中非负数的刻画是相对应的. 具体来说, 复数 z 非负当且仅当它有非负平方根, 这对应于条件 (3). 此外, z 非负当且仅当它有实的平方根, 这对应于条件 (4). 最后, z 是非负的当且仅当有复数 w 使得 $z = \bar{w}w$, 这对应于条件 (5).

定理 (**半正定算子的刻画**) : 设 $T \in \mathcal{L}(V)$, 则下面各条等价:

(1) T 是半正定算子.

(2) T 自伴, 且所有特征值非负.

(3) T 有半正定的平方根.

(4) T 有自伴的平方根.

(5) 存在 $R \in \mathcal{L}(V)$ 使得 $T = R^*R$.

证明 我们将证明 (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (5) $\Longrightarrow$ (1).

(1) $\Longrightarrow$ (2) 由定义知 T 是自伴的. 又

$$\lambda \|v\|^2 = \lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle Tv, v \rangle \geqslant 0,$$

故 $\lambda \geqslant 0$, 这里最后一个等号是因为 T 是半正定的.

(2) $\Longrightarrow$ (3) 因 T 自伴, 故由谱定理可知 V 有由 T 的特征向量构成的规范正交基 $e_1, \cdots, e_n$. 设它们对应的特征值分别为 $\lambda_1, \cdots, \lambda_n$, 且所有的 λ_j 均非负. 定义 $R \in \mathcal{L}(V)$ 为

$$Re_j = \sqrt{\lambda_j} e_j,$$

那么 R 是半正定的, 且 $R^2 e_j = \lambda_j e_j = Te_j$, 从而 $R^2 = T$, 故 R 为 T 的平方根.

(3) $\Longrightarrow$ (4) 由定义知半正定算子必为自伴算子.

(4) $\Longrightarrow$ (5) 设 R 为 T 的自伴的平方根, 那么 $R^* = R$, $R^2 = T$, 从而 $R^*R = T$.

(5) $\Longrightarrow$ (1) 由于 $T^* = (R^*R)^* = R^*R = T$, 故 T 是自伴的. 又

$$\langle Tv, v \rangle = \langle (R^*R)v, v \rangle = \langle Rv, Rv \rangle \geqslant 0,$$

故 T 是半正定的. □

每个非负数都有唯一的非负平方根, 以下命题表明正算子也有类似的性质.

定理 (半正定算子的半正定平方根具有唯一性)：半正定算子有唯一的半正定平方根.

注：半正定算子可能有无穷多个平方根, 但其中只有一个是半正定的. 例如, 若 $\dim V > 1$, 则恒等算子 I_V 就有无穷多个平方根, 详情见习题 2. .

证明 设 $T \in \mathcal{L}(V)$ 为半正定算子, R 为 T 的半正定平方根. 因为 T 是半正定的, 故有非负的特征值, 设 v 是 T 的相应于特征值 $\lambda \geqslant 0$ 的特征向量, $Tv = \lambda v$. 下面证明 $Rv = \sqrt{\lambda} v$. (由谱定理, V 有由 T 的特征向量构成的基, 因此只要 R 在 T 的特征向量上是唯一确定的, 那么 R 就是唯一确定的.)

因为 R 自伴, 故 V 有由 R 的特征向量构成的规范正交基 $e_1, \cdots, e_n$. 因为 R 是半正定的, 故 R 的特征值都是非负的. 因此可设 $e_1, \cdots, e_n$ 相应的特征值为 $\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n}$. 设

$$v = a_1 e_1 + \cdots + a_n e_n,$$

那么

$$Rv = a_1 \sqrt{\lambda_1} e_1 + \cdots + a_n \sqrt{\lambda_n} e_n.$$

从而

$$R^2 v = a_1 \lambda_1 e_1 + \cdots + a_n \lambda_n e_n.$$

因为 $R^2 = T$ 且 $Tv = \lambda v$, 故

$$a_1 \lambda_1 e_1 + \cdots + a_n \lambda_n e_n = a_1 \lambda e_1 + \cdots + a_n \lambda e_n.$$

从而

$$a_j (\lambda - \lambda_j) = 0.$$

因此当 $\lambda_j \neq \lambda$ 时, $a_j = 0$, 故

$$v = \sum_{\{j : \lambda_j = \lambda\}} a_j e_j.$$

从而

$$Rv = \sum_{\{j : \lambda_j = \lambda\}} a_j \sqrt{\lambda_j} e_j = \sum_{\{j : \lambda_j = \lambda\}} a_j \sqrt{\lambda} e_j = \sqrt{\lambda} v.$$

故 R 在 T 的特征向量上是唯一确定的, 进而在 V 的由 T 的特征向量构成的基上是唯一确定的. $\square$

8.7 等距同构

保持范数的算子十分重要, 它们应该有一个名字.

定义 (等距同构): 设 $T \in \mathcal{L}(V)$, 若对任意的 $v \in V$, 都有

$$\|Tv\| = \|v\|,$$

则称 T 为等距同构 (isometry).

例如, 当 $\lambda \in \mathbf{F}$ 满足 $|\lambda| = 1$ 时, λI 就是等距同构, 这是等距同构最平凡的情形.

注: 实内积空间上的等距同构通常称为正交算子, 复内积空间上的等距同构通常称为酉算子. 我们采用等距同构这个术语, 以使我们的结果对实内积空间与复内积空间都适用.

例 1 设 $\lambda_1, \cdots, \lambda_n \in \mathbf{F}$ 且 $|\lambda_j| = 1$, $e_1, \cdots, e_n$ 是 V 的规范正交基. $T \in \mathcal{L}(V)$ 满足 $Te_j = \lambda_j e_j$. 证明 T 是等价同构.

证明 任取 $v \in V$, 则

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n,$$
$$Tv = \langle v, e_1 \rangle \lambda_1 e_1 + \cdots + \langle v, e_n \rangle \lambda_n e_n.$$

从而

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2 = \|Tv\|^2.$$

故 T 为等距同构. □

以下的定理给出了等距同构的若干等价条件. (1) 与 (2) 的等价性表明一个算子是等距同构当且仅当它保持内积. (1) 和 (3)(或者 (4)) 的等价性表明一个算子是等距同构当且仅当关于每一个 (某一个) 规范正交基, 其矩阵的列是规范正交的. 后面我们会知道这里的 “列” 也可以换为 “行”.

定理 (等距同构的刻画): 设 $T \in \mathcal{L}(V)$, 则下面各条等价:

(1) T 是等距同构.

(2) 对任意的 $u, v \in V$, 有 $\langle Tu, Tv \rangle = \langle u, v \rangle$.

(3) 对 V 的任一规范正交基 $e_1, \cdots, e_n$, 向量组 $Te_1, \cdots, Te_n$ 也是 V 的规范正交基.

(4) 存在 V 的规范正交基 $e_1, \cdots, e_n$, 使得向量组 $Te_1, \cdots, Te_n$ 也是 V 的规范正交基.

(5) $T^*T = I$.

(6) $TT^* = I$.

(7) T^* 是等距同构.

(8) T 是可逆的, 且 $T^{-1} = T^*$.

证明 我们将证明 (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (5) $\Longrightarrow$ (6) $\Longrightarrow$ (7) $\Longrightarrow$ (8) $\Longrightarrow$ (1).

(1) $\Longrightarrow$ (2) 等距同构保范数, 而内积可用范数来表示:

$$\text{当 } \mathbf{F}=\mathbf{R} \text{ 时,} \quad \langle u, v\rangle = \frac{1}{4}\left(\|u+v\|^2 - \|u-v\|^2\right),$$

$$\text{当 } \mathbf{F}=\mathbf{C} \text{ 时,} \quad \langle u, v\rangle = \frac{1}{4}\sum_{k=1}^{4} \mathrm{i}^k \left\|u+\mathrm{i}^k v\right\|^2.$$

因此不论实内积空间还是复内积空间, 总有 $\langle Tu, Tv\rangle = \langle u, v\rangle$.

(2) $\Longrightarrow$ (3) $\langle Te_i, Te_j\rangle = \langle e_i, e_j\rangle$, 故 $Te_1, \cdots, Te_n$ 也是规范正交的. 又它的长度等于基的长度, 故它是规范正交基.

(3) $\Longrightarrow$ (4) 显然.

(4) $\Longrightarrow$ (5)

$$\langle T^*Te_i, e_j\rangle = \langle Te_i, Te_j\rangle = \langle e_i, e_j\rangle,$$

其中第一个等号利用了伴随的定义与 $(T^*)^* = T$, 第二个等号是因为 $Te_1, \cdots, Te_n$ 是规范正交的. 因为 $e_1, \cdots, e_n$ 是 V 的基, 故对任意的 $u, v \in V$, 均有 $\langle T^*Tu, v\rangle = \langle u, v\rangle$, 也即

$$\langle (T^*T - I)u, v\rangle = 0.$$

注意到 $T^*T - I$ 是自伴的, 从而 $T^*T - I = 0$, 故 $T^*T = I$.

(5) $\Longrightarrow$ (6) 一般情况下算子乘法不满足交换律, 但由第三章习题知 $S_1S_2 = I$ 当且仅当 $S_2S_1 = I$, 故由 $T^*T = I$ 可得 $TT^* = I$.

(6) $\Longrightarrow$ (7) 由 $TT^* = I$ 知

$$\|T^*v\|^2 = \langle T^*v, T^*v\rangle = \langle TT^*v, v\rangle = \langle v, v\rangle = \|v\|^2,$$

故 T^* 为等距同构.

(7) $\Longrightarrow$ (8) 已知 T^* 是等距同构. 由 (1)$\Longrightarrow$(5) 以及 (1)$\Longrightarrow$(6) 可知 $TT^* = I$ 且 $T^*T = I$, 故 T 可逆, 且 $T^{-1} = T^*$.

(8) $\Longrightarrow$ (1) 由 $T^{-1} = T^*$ 可知

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle = \left\langle T^{-1}Tv, v\right\rangle = \langle v, v\rangle = \|v\|^2,$$

故 T 为等距同构. □

由等距同构的刻画 (5)(6) 知若 T 为等距同构, 则 $T^*T = TT^* = I$. 因此, 等距同构必为正规算子. 于是, 正规算子的刻画可以用来给出等距同构的描述. 复的情形见下面的定理, 实的情形在后面的章节中.

定理 (**C** 上等距同构的描述) : 设 $T \in \mathcal{L}(V)$, 则下面各条等价:

(1) T 是等距同构.

(2) V 有由 T 的特征向量构成的规范正交基, 且相应特征值的模长均为 1.

证明 (1) $\Longrightarrow$ (2) 由于 T 是等距同构, 故 T 是正规的. 由复谱定理知 V 有一个由 T 的特征向量构成的规范正交基 $e_1, \cdots, e_n$. 设 e_j 相应的特征值为 λ_j, 则由 T 为等距同构知

$$\|e_j\| = \|Te_j\| = \|\lambda_j e_j\| = |\lambda_j| \, \|e_j\| .$$

故 $|\lambda_j| = 1$.

(2) $\Longrightarrow$ (1) 见例 1 . □

注 : 从上述定理的证明我们看到, 如果 $\mathbf{F} = \mathbf{C}$, 那么例 1 包括了复内积空间上所有的等距同构.

8.8 极分解

回忆一下我们在 **C** 与 $\mathcal{L}(V)$ 之间做的类比. 一个复数 z 相应于一个算子 T, 而 $\bar{z}$ 相应于 T^*. 实数 $(z = \bar{z})$ 相应于自伴算子 $(T = T^*)$, 而非负实数相应于半正定算子.

此外, **C** 还有一个重要的子集——单位圆 $|z| = 1$, 它等价于 $\bar{z}z = 1$. 按照我们的类比, 这相应于 $T^*T = I$, 它等价于 T 是等距同构. 也就是说, **C** 中的单位圆相应于 $\mathcal{L}(V)$ 中的全体等距同构.

继续我们的类比. 注意到每一个非零复数 z 都可以写成

$$z = \left(\frac{z}{|z|}\right) \sqrt{\bar{z}z}$$

的形式, 其中 $\dfrac{z}{|z|}$ 是单位圆中的元素. 据此我们猜测任何一个算子 $T \in \mathcal{L}(V)$ 都可以写成等距同构乘以 $\sqrt{T^*T}$ 的形式. 下面我们就来证明这个猜测确实是对的.

设 T 是半正定算子, 由 §8.6 可知 T 有唯一的半正定平方根, 我们用记号 $\sqrt{T}$ 表示这个唯一的半正定平方根.

注意到对每个 $T \in \mathcal{L}(V)$, 算子 T^*T 都是半正定的, 因此记号 $\sqrt{T^*T}$ 是有意义的.

引理 : 设 $T \in \mathcal{L}(V)$, 则对任意的 $v \in V$, 有 $\|Tv\| = \|\sqrt{T^*T}v\|$.

证明 注意到 $\sqrt{T^*T}$ 是自伴的, 故对任意的 $v\in V$, 有

$$\begin{aligned}\|Tv\|^2 &= \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle\\ &= \langle \sqrt{T^*T}\sqrt{T^*T}v, v\rangle\\ &= \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle\\ &= \|\sqrt{T^*T}v\|^2.\end{aligned}$$

结论得证. $\square$

现在我们来陈述并证明极分解定理, 它给出 V 上任意算子的一个漂亮的描述.

定理 (极分解定理): 设 $T\in\mathcal{L}(V)$, 则存在等距同构 $S\in\mathcal{L}(V)$, 使得 $T=S\sqrt{T^*T}$.

证明 定义

$$\begin{aligned}S_1: \operatorname{Im}\sqrt{T^*T} &\longrightarrow \operatorname{Im}T\\ \sqrt{T^*T}v &\longmapsto Tv\end{aligned}$$

若 $\sqrt{T^*T}v_1=\sqrt{T^*T}v_2$, 则由前述引理知 $\|T(v_1-v_2)\|=\|\sqrt{T^*T}(v_1-v_2)\|=0$, 故 $Tv_1=Tv_2$, 因此 S_1 是合理定义的. 易证 S_1 为线性映射. 我们的主要思路是将 S_1 扩张为 V 上的等距同构, 使得 $T=S\sqrt{T^*T}$.

由 S_1 的定义易知 S_1 是满的. 任取 $u=\sqrt{T^*T}v\in\operatorname{Im}\sqrt{T^*T}$, 由前述引理知

$$\|S_1u\|=\|S_1\sqrt{T^*T}v\|=\|Tv\|=\|\sqrt{T^*T}v\|=\|u\|,$$

故 S_1 保持范数. 由此可知 S_1 是单的, 故由线性映射基本定理得

$$\dim\operatorname{Im}\sqrt{T^*T}=\dim\operatorname{Im}T.$$

从而由投影定理知

$$\dim(\operatorname{Im}\sqrt{T^*T})^\perp=\dim(\operatorname{Im}T)^\perp.$$

取 $(\operatorname{Im}\sqrt{T^*T})^\perp$ 的规范正交基 $e_1,\cdots,e_m$ 与 $(\operatorname{Im}T)^\perp$ 的规范正交基 $f_1,\cdots,f_m$. 定义线性映射

$$\begin{aligned}S_2: \quad (\operatorname{Im}\sqrt{T^*T})^\perp &\longrightarrow (\operatorname{Im}T)^\perp\\ a_1e_1+\cdots+a_me_m &\longmapsto a_1f_1+\cdots+a_mf_m\end{aligned}$$

由规范正交组的性质知 $\|S_2w\|=\|w\|$. 由投影定理知对任意的 $v\in V$, 存在唯一的 $u\in\operatorname{Im}\sqrt{T^*T}$ 与

$w \in (\operatorname{Im}\sqrt{T^*T})^{\perp}$, 使得 $v = u + w$, 据此我们可以定义 $S \in \mathcal{L}(V)$ 为

$$Sv = S_1u + S_2w.$$

容易知道 S 在 $\operatorname{Im}\sqrt{T^*T}$ 上与 S_1 相等, 在 $(\operatorname{Im}\sqrt{T^*T})^{\perp}$ 上与 S_2 相等. 由于

$$S\sqrt{T^*T}v = S_1\sqrt{T^*T}v = Tv,$$

故 $T = S\sqrt{T^*T}$. 下面证明 S 为等距同构. 由勾股定理与 S_1, S_2 保持范数知

$$\|Sv\|^2 = \|S_1u + S_2w\|^2 = \|S_1u\|^2 + \|S_2w\|^2 = \|u\|^2 + \|w\|^2 = \|v\|^2.$$

结论得证. □

极分解定理说的是 V 上的每一个算子都是一个等距同构与一个半正定算子的乘积, 并且它们都是我们已经完全描述, 同时能够比较好地理解的算子类.

特别地, 考虑 $\mathbf{F} = \mathbf{C}$ 的情形, $T = S\sqrt{T^*T}$ 是 $T \in \mathcal{L}(V)$ 的极分解, 其中 S 为等距同构. 则 V 有一个规范正交基使得 S 关于这个基有对角矩阵, 并且 V 还有一个规范正交基使得 $\sqrt{T^*T}$ 关于这个基有对角矩阵. 但是请注意, 这两个规范正交基可能是不同的, 我们未必能找到一个规范正交基使得 S 与 $\sqrt{T^*T}$ 同时有对角矩阵.

8.9 奇异值

定义 (奇异值)：设 $T \in \mathcal{L}(V)$, 我们称 $\sqrt{T^*T}$ 的特征值为 T 的奇异值 (singular values), 其中每个特征值 λ 要重复 $\dim E(\lambda, \sqrt{T^*T})$ 次.

注：因为 T 的奇异值是半正定算子 $\sqrt{T^*T}$ 的特征值, 因此它们都是非负的.

例 1 定义 $T \in \mathcal{L}(\mathbf{F}^4)$ 为

$$T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4).$$

求 T 的奇异值.

解 取标准基, 直接计算得

$$\mathcal{M}(\sqrt{T^*T}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

因此 $\sqrt{T^*T}$ 有特征值 0, 3 (二重), 2, 故 T 的奇异值为 $0, 2, 3, 3$.

把谱定理用于半正定算子 $\sqrt{T^*T}$, 并利用线性变换可对角化的充要条件 (5) 可知每个 $T \in \mathcal{L}(V)$ 都有 $\dim V$ 个奇异值, 正如我们在例 1 所看到的那样.

以下命题表明, 利用奇异值和 V 的两个规范正交基 $e_1, \cdots, e_n$ 与 $f_1, \cdots, f_n$, 每一个算子 $T \in \mathcal{L}(V)$ 都有一个简洁的描述.

定理 (奇异值分解): 设 $T \in \mathcal{L}(V)$ 的全部奇异值为 $s_1, \cdots, s_n$, 那么存在 V 的两个规范正交基 $e_1, \cdots, e_n$ 与 $f_1, \cdots, f_n$, 使得对任意的 $v \in V$, 都有

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n.$$

证明 奇异值 $s_1, \cdots, s_n$ 是半正定算子 $\sqrt{T^*T}$ 的特征值. 由谱定理知存在 V 的规范正交基 $e_1, \cdots, e_n$ 使得 $\sqrt{T^*T}e_j = s_j e_j$. 由于

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n,$$

两边用 $\sqrt{T^*T}$ 作用得

$$\sqrt{T^*T}v = s_1 \langle v, e_1 \rangle e_1 + \cdots + s_n \langle v, e_n \rangle e_n.$$

由极分解定理知存在等距同构 $S \in \mathcal{L}(V)$ 使得 $T = S\sqrt{T^*T}$. 用 S 作用于上式得

$$Tv = s_1 \langle v, e_1 \rangle Se_1 + \cdots + s_n \langle v, e_n \rangle Se_n.$$

令 $f_j = Se_j$, 由 S 是等距同构知 $f_1, \cdots, f_n$ 也是规范正交基. 结论得证. □

回忆一下, 在讨论 $\mathcal{L}(V, W)$ 时, 我们取定了 V 与 W 的基, 而在讨论 $\mathcal{L}(V)$ 时, 我们几乎总是让同一个基扮演这两个角色. 奇异值分解给了我们一个难得的机会——对线性变换的矩阵使用两个不同的基. 为此, 设 $T \in \mathcal{L}(V)$, $s_1, \cdots, s_n$ 为 T 的所有奇异值, $e_1, \cdots, e_n$ 与 $f_1, \cdots, f_n$ 是 V 的满足 T

的奇异值分解的两个规范正交基, 那么由 $Te_j = s_j f_j$ 可得

$$\mathcal{M}(T,(e_1,\cdots,e_n),(f_1,\cdots,f_n)) = \begin{pmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_n \end{pmatrix}.$$

也就是说, 只要允许我们在处理线性变换时使用两个不同的基, 而不是像通常那样只使用一个单独的基, 那么每个 $T \in \mathcal{L}(V)$ 关于 V 的某些规范正交基都有对角矩阵.

引理 ($\sqrt{T}$ **的特征值与** T **的特征值的关系**): 设 $T \in \mathcal{L}(V)$ 半正定, 则 $\lambda_1,\cdots,\lambda_n$ 为 T 的特征值当且仅当 $\sqrt{\lambda_1},\cdots,\sqrt{\lambda_n}$ 为 $\sqrt{T}$ 的特征值.

证明 ($\Longleftarrow$) 设 v_j 为 $\sqrt{T}$ 的相应于 $\sqrt{\lambda_j}$ 的特征向量, 则 $\sqrt{T}v_j = \sqrt{\lambda_j}v_j$, 从而

$$Tv_j = \sqrt{T}\sqrt{T}v = \sqrt{T}(\sqrt{\lambda_j}v_j) = \sqrt{\lambda_j}(\sqrt{T}v_j) = \sqrt{\lambda_j}(\sqrt{\lambda_j}v_j) = \lambda_j v_j.$$

因此 λ_j 是 T 的特征值.

($\Longrightarrow$) 设 $\sqrt{T}$ 关于规范正交基 $e_1,\cdots,e_n$ 的矩阵为 A, 那么

$$\sqrt{T}e_j = a_{1j}e_1 + \cdots + a_{nj}e_n,$$

又因为

$$\begin{aligned}\lambda_j \langle e_j, e_i\rangle = \langle \lambda_j e_j, e_i\rangle = \langle Te_j, e_i\rangle = \langle \sqrt{T}e_j, \sqrt{T}e_i\rangle \\ = \langle a_{1j}e_1 + \cdots + a_{nj}e_n, a_{1i}e_1 + \cdots + a_{ni}e_n\rangle, \\ = a_{1j}\overline{a_{1i}} + \cdots + a_{nj}\overline{a_{ni}}.\end{aligned}$$

故

$$a_{1j}\overline{a_{1i}} + \cdots + a_{nj}\overline{a_{ni}} = \begin{cases} \lambda_j, & i = j, \\ 0, & i \neq j.\end{cases}$$

因此

$$A^H A = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

由于 $\sqrt{T}$ 是自伴的, 故 $A^H = A$, 从而

$$A = \begin{pmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_n} \end{pmatrix}.$$

因此

$$\sqrt{T}e_j = \sqrt{\lambda_j}e_j.$$

结论得证. □

命题 (不对算子开平方描述奇异值): 设 $T \in \mathcal{L}(V)$, 则 T 的奇异值就是 T^*T 的特征值的非负平方根, 且每个特征值 λ 要重复 $\dim E(\lambda, T^*T)$ 次.

证明 对算子 T^*T 使用上述引理即得结论. □

8.10 习题

1. 设 $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{F}$, 则 λ 是 T 的特征值当且仅当 $\bar{\lambda}$ 是 T^* 的特征值.

证明

$$\begin{aligned}
\lambda \text{不是 } T \text{ 的特征值} &\Longleftrightarrow T - \lambda I \text{是可逆的} \\
&\Longleftrightarrow (T - \lambda I)S = S(T - \lambda I) = I \\
&\Longleftrightarrow S^*(T - \lambda I)^* = (T - \lambda I)^* S^* = I \\
&\Longleftrightarrow (T - \lambda I)^* \text{是可逆的} \\
&\Longleftrightarrow T^* - \bar{\lambda} I \text{是可逆的} \\
&\Longleftrightarrow \bar{\lambda} \text{不是 } T^* \text{ 的特征值.}
\end{aligned}$$

结论得证. □

2. 证明或反驳: $\mathbf{F}^2$ 上的恒等算子有无穷多个自伴的平方根.

9 复线性空间上的算子

本章将更深入地研究复线性空间上算子的结构, 这里用不到内积, 因此我们回到了一般的有限维线性空间的情形. 为了避免某些平凡性, 我们将假设 $V \neq \{0\}$.

本章我们用 $\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$, 用 V 表示 $\mathbf{F}$ 上的有限维非零线性空间.

9.1 算子幂的核

本章先来讨论算子幂的核.

命题 (递增的核序列)：设 $T \in \mathcal{L}(V)$, 则

$$\{0\} = \operatorname{Ker} T^0 \subseteq \operatorname{Ker} T^1 \subseteq \operatorname{Ker} T^2 \subseteq \cdots \subseteq \operatorname{Ker} T^k \subseteq \operatorname{Ker} T^{k+1} \subseteq \cdots .$$

证明 请读者自行验证. □

下面的命题是说, 如果在这个子空间序列中有相邻两项相等, 那么此后的所有项都相等.

命题 (相等的核序列)：设 $T \in \mathcal{L}(V)$, 若有非负整数 m 使得 $\operatorname{Ker} T^m = \operatorname{Ker} T^{m+1}$, 则

$$\operatorname{Ker} T^m = \operatorname{Ker} T^{m+1} = \operatorname{Ker} T^{m+2} = \operatorname{Ker} T^{m+3} = \cdots .$$

证明 下证对任意的正整数 k, 都有

$$\operatorname{Ker} T^{m+k} = \operatorname{Ker} T^{m+k+1}.$$

由前一命题知我们只需证明 $\operatorname{Ker} T^{m+k+1} \subseteq \operatorname{Ker} T^{m+k}$. 任取 $v \in \operatorname{Ker} T^{m+k+1}$, 则 $T^{m+1}(T^k v) = 0$, 从而 $T^k v \in \operatorname{Ker} T^{m+1} = \operatorname{Ker} T^m$, 故

$$T^{m+k} v = T^m (T^k v) = 0.$$

因此 $v \in \operatorname{Ker} T^{m+k}$, 从而 $\operatorname{Ker} T^{m+k+1} \subseteq \operatorname{Ker} T^{m+k}$. □

上面的命题引出了一个问题——是否存在非负整数 m, 使得 $\operatorname{Ker} T^m = \operatorname{Ker} T^{m+1}$ ？下面的结果告诉我们, 这个等式至少在 m 等于 V 的维数时成立.

命题 (核停止增长)：设 $T \in \mathcal{L}(V)$, $\dim V = n$, 则

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

证明 假设 $\operatorname{Ker} T^n \neq \operatorname{Ker} T^{n+1}$, 则由前述两个命题可知

$$\{0\} = \operatorname{Ker} T^0 \subsetneqq \operatorname{Ker} T^1 \subsetneqq \operatorname{Ker} T^2 \subsetneqq \cdots \subsetneqq \operatorname{Ker} T^n \subsetneqq \operatorname{Ker} T^{n+1}.$$

上述关系式每个真包含关系处的维数至少增加 1, 因此 $\dim T^{n+1} \geqslant n + 1$, 矛盾. □

尽管 $V = \operatorname{Ker} T \oplus \operatorname{Im} T$ 并不是对每个 $T \in \mathcal{L}(V)$ 都成立, 但是我们有以下命题作为它的替补.

命题 (线性空间分解为核与像的直和)：设 $T \in \mathcal{L}(V)$, $\dim V = n$, 则

$$V = \operatorname{Ker} T^n \oplus \operatorname{Im} T^n.$$

证明 先证 $\operatorname{Ker} T^n \cap \operatorname{Im} T^n = \{0\}$. 任取 $v \in \operatorname{Ker} T^n \cap \operatorname{Im} T^n$, 则 $T^n v = 0$, 且存在 $u \in V$ 使得 $v = T^n u$, 因此

$$0 = T^n v = T^{2n} u.$$

故 $u \in \operatorname{Ker} T^{2n} = \operatorname{Ker} T^n$(由前述命题), 从而 $v = T^n u = 0$.

又因为

$$\dim(\operatorname{Ker} T^n \oplus \operatorname{Im} T^n) = \dim \operatorname{Ker} T^n + \dim \operatorname{Im} T^n = \dim V,$$

故 $V = \operatorname{Ker} T^n \oplus \operatorname{Im} T^n$, 这里第二个等号利用了线性映射基本定理. □

例 1 设 $T \in \mathcal{L}(\mathbf{F}^3)$ 定义为

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3).$$

易知 $\operatorname{Ker} T = \{(z_1, 0, 0) : z_1 \in \mathbf{F}\}$, $\operatorname{Im} T = \{(z_1, 0, z_3) : z_1, z_3 \in \mathbf{F}\}$, 故 $\operatorname{Ker} T^n \cap \operatorname{Im} T^n \neq \{0\}$, 且 $\operatorname{Ker} T + \operatorname{Im} T \neq \mathbf{F}^3$. 但是

$$T^3(z_1, z_2, z_3) = (0, 0, 125z_3),$$

故 $\operatorname{Ker} T^3 = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{F}\}$, $\operatorname{Im} T^3 = \{(0, 0, z_3) : z_3 \in \mathbf{F}\}$, 因此

$$\mathbf{F}^3 = \operatorname{Ker} T^3 \oplus \operatorname{Im} T^3.$$

9.2 广义特征向量

有些线性变换因为没有足够多线性无关的特征向量而没有一个好的描述. 因此, 本节将引入广义特征向量的概念, 这一概念对线性变换结构的描述起着重要作用.

取定线性变换 $T \in \mathcal{L}(V)$, 为了描述 T, 我们想要找到一个 "好的" 直和分解

$$V = U_1 \oplus \cdots \oplus U_m,$$

其中每个 U_j 都是在 T 下不变的子空间. 一维不变子空间是最简单的不变子空间, 而上述分解中每个 U_j 均为一维不变子空间当且仅当 V 有一个由 T 的特征向量构成的基, 也当且仅当 V 有如下的特征子空间直和分解

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T), \tag{1}$$

其中 $\lambda_1, \cdots, \lambda_m$ 是 T 的全部互异特征值.

谱定理证明了, 当 $\mathbf{F} = \mathbf{C}$ 时, 形如 (1) 式的分解对每个正规算子都成立, 当 $\mathbf{F} = \mathbf{R}$ 时, 形如 (1) 式的分解对每个自伴算子都成立, 因为它们有足够多的特征向量构成 V 的一个基.

令人遗憾的是, 即使是在复线性空间上, 形如 (1) 式的分解对更一般的线性变换也可能不成立. 我们现在要引入的广义特征向量与广义特征子空间将会改善这种局面.

定义 (广义特征向量)：设 $T \in \mathcal{L}(V)$, $0 \neq v \in V$. 若存在正整数 j 使得

$$(T - \lambda I)^j v = 0,$$

则称 v 为 T 的相应于 λ 的广义特征向量 (generalized eigenvector) .

尽管上述定义中的 j 可以是任意正整数, 但我们将会证明, 当 $j = \dim V$ 时, 每个广义特征向量都满足 $(T - \lambda I)^j v = 0$.

定义 (广义特征子空间)：设 $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{F}$, 我们把 T 的相应于 λ 的所有广义特征向量加上零向量所成的集合称为 T 的相应于 λ 的广义特征子空间 (generalized eigen-subspace), 记为 $G(\lambda, T)$. 注意这里我们并没有要求 λ 为特征值.

注：我们并没有定义广义特征值, 这是因为若 $(T - \lambda I)^j$ 不是单的, 那么 $T - \lambda I$ 也不是单的, 从而 λ 为 T 的特征值.

在广义特征向量的定义中取 $j = 1$ 可知, 每个特征向量都是广义特征向量, 因此

$$E(\lambda, T) \subseteq G(\lambda, T).$$

以下命题表明广义特征子空间确实是子空间 (因为核一定是子空间) .

命题 (广义特征子空间的刻画)：设 $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{F}$, 那么 $G(\lambda, T) = \operatorname{Ker}(T - \lambda I)^{\dim V}$.

证明 任取 $v \in \operatorname{Ker}(T - \lambda I)^{\dim V}$. 由广义特征向量的定义可知 $v \in G(\lambda, T)$ (取 $j = \dim V$ 即可), 故 $\operatorname{Ker}(T - \lambda I)^{\dim V} \subseteq G(\lambda, T)$.

反之, 任取 $v \in G(\lambda, T)$, 则存在正整数 j 使得 $(T - \lambda I)^j v = 0$, 故

$$v \in \operatorname{Ker}(T - \lambda I)^j \subseteq \operatorname{Ker}(T - \lambda I)^{\dim V}.$$

这里的 $\subseteq$ 利用了 §9.1 的第一个与第三个命题. 因此 $G(\lambda, T) \subseteq \operatorname{Ker}(T - \lambda I)^{\dim V}$. □

例 1 设 $T \in \mathcal{L}(\mathbf{C}^3)$ 定义为

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3).$$

(1) 求 T 的所有特征值以及相应的特征子空间与广义特征子空间.

(2) 证明 $\mathbf{C}^3$ 等于 T 的所有互异广义特征子空间的直和.

解 (1) 由 $T(z_1, z_2, z_3) = \lambda(z_1, z_2, z_3)$ 可得 T 有特征值 0 与 5, 特征子空间

$$E(0, T) = \{(z_1, 0, 0) : z_1 \in \mathbf{C}\},$$
$$E(5, T) = \{(0, 0, z_3) : z_1 \in \mathbf{C}\}.$$

又因为

$$T^3(z_1, z_2, z_3) = (0, 0, 125z_3),$$
$$(T - 5I)^3(z_1, z_2, z_3) = (300z_2 - 125z_1, -125z_2, 0),$$

从而

$$G(0, T) = \{(z_1, z_2, 0) : z_1 \in \mathbf{C}\},$$
$$G(5, T) = \{(0, 0, z_3) : z_1 \in \mathbf{C}\}.$$

(2) 由 (1) 知 $\mathbf{C}^3 = G(0, T) \oplus G(5, T)$.

本章的主要目标之一是证明上面例 1 中 (2) 的结果对有限维复线性空间上的算子总是成立的. 我们将在后面给出证明.

我们已经证明过相应于不同特征值的特征向量是线性无关的, 现在我们对广义特征向量证明类似的结果.

性质 (相应于不同特征值的广义特征向量线性无关): 设 $T \in \mathcal{L}(V)$, $\lambda_1, \cdots, \lambda_m$ 是 T 的特征值, $v_1, \cdots, v_m$ 分别为相应的广义特征向量, 则 $v_1, \cdots, v_m$ 线性无关.

证明 设

$$0 = a_1 v_1 + \cdots + a_m v_m. \tag{1}$$

先证 $a_1 = 0$. 设 k 是使得 $(T - \lambda_1 I)^k v_1 \neq 0$ 的最大非负整数, 令

$$w = (T - \lambda_1 I)^k v_1.$$

则

$$(T - \lambda_1 I)w = (T - \lambda_1 I)^{k+1} v_1 = 0,$$

故 $Tw = \lambda_1 w$, 从而对任意的 $\lambda \in \mathbf{F}$, 有 $(T - \lambda I)w = (\lambda_1 - \lambda)w$, 从而

$$(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w, \tag{2}$$

其中 $n = \dim V$. 用 $(T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n$ 作用于 (1) 式得

$$\begin{aligned} 0 &= a_1 (T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n v_1 \\ &= a_1 (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n w \\ &= a_1 (\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_m)^n w. \end{aligned}$$

其中第二个等式利用了线性变换多项式的交换性, 第三个等式利用了 (2) 式. 因 λ_j 互异, $w \neq 0$, 故 $a_1 = 0$. 同理可得 $a_2 = \cdots = a_m = 0$, 故 $v_1, \cdots, v_m$ 线性无关. □

9.3 幂零算子

定义 (幂零算子): 若一个算子的某个幂等于 0, 则称该算子是幂零的 (nilpotent).

注: 拉丁词 nil 的意思是无或者零, 拉丁词 potent 的意思是幂, 故 nilpotent 的字面意思是幂零.

例 1 (幂零算子)

(1) 定义为 $N(z_1, z_2, z_3, z_4) = (z_3, z_4, 0, 0)$ 的算子 $N \in \mathcal{L}(\mathbf{F}^4)$ 是幂零的, 因为 $N^2 = 0$.

(2) $\mathcal{P}_m(\mathbf{R})$ 上的微分算子 D 是幂零的, 因为任何一个次数不超过 m 的多项式的 $m+1$ 阶导数都等于 0, 即 $D^{m+1} = 0$.

以下的命题表明对于幂零算子, 我们不必考虑比空间的维数更高的幂.

命题 (幂零算子的 $\dim V$ 次幂必为零算子): 设 $N \in \mathcal{L}(V)$ 是幂零的, 则 $N^{\dim V} = 0$.

证明 任取 $v \in V$, 则存在正整数 k 使得 $N^k v = 0$, 即 $(N - 0I)^k v = 0$, 故 $v \in G(0, N)$, 从而

$$V = G(0, N) = \operatorname{Ker} (N - 0I)^{\dim V} = \operatorname{Ker} N^{\dim V}.$$

因此 $N^{\dim V} = 0$. □

给定一个算子 T, 我们想要找到一个基, 使得 T 关于这个基的矩阵尽可能简单, 即这个矩阵包含很多的 0.

下面的结果表明, 若 N 是幂零的, 那么可以取到 V 的一个基, 使得 N 关于这个基的矩阵有一半以上的元素都为 0. 在本章的后面会有更好的结果.

命题（幂零算子的矩阵）：设 $N \in \mathcal{L}(V)$ 是幂零算子，那么存在 V 的一个基使得 N 关于这个基的矩阵形如

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix},$$

其中对角线以及对角线下方的元素均为 0.

证明 设 $\dim V = n$. 取 $\operatorname{Ker} N$ 的一个基，将其扩充为 $\operatorname{Ker} N^2$ 的基，再将其扩充为 $\operatorname{Ker} N^3$ 的基. 如此重复下去，可在有限步内得到 V 的一个基（因为 $\operatorname{Ker} N^{\dim V} = V$）. 设最终得到的基为 $v_1, \cdots, v_n$. 下面我们证明

$$(Nv_1, Nv_2, \cdots, Nv_n) = (v_1, v_2, \cdots, v_n) \begin{pmatrix} 0 & & & * \\ 0 & 0 & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

由于 $v_1 \in \operatorname{Ker} N$，故 $Nv_1 = 0$，从而矩阵第一列全为 0.

$v_2 \in \operatorname{Ker} N$ 或者 $v_2 \in \operatorname{Ker} N^2$. 当 $v_2 \in \operatorname{Ker} N$ 时，第二列全为 0. 当 $v_2 \in \operatorname{Ker} N^2$ 时，$Nv_2 \in \operatorname{Ker} N$，故第二列对角线及其下方全为 0.

以此类推，每一列对角线及其下方的元素全为 0，结论得证. □

推论（幂零算子的特征值只有 0）：设 $N \in \mathcal{L}(V)$ 为幂零算子，则 0 是 N 仅有的特征值.

证明 N 关于某个基有对角线全为 0 的上三角矩阵，而上三角矩阵对角线上的元素就是 N 的特征值，故幂零算子 N 的特征值只有 0. □

9.4 复线性空间上算子的刻画，复线性空间分解为广义特征子空间的直和

前面我们看到，即使在有限维复线性空间上，线性空间也未必能分解为特征子空间的直和. 本节我们将会看到，有限维复线性空间上的每一个线性变换都有足够的广义特征向量来给出一个分解.

我们已经知道对每个 $T \in \mathcal{L}(V)$，$\operatorname{Ker} T$ 与 $\operatorname{Im} T$ 都在 T 下不变. 现在我们要证明，T 的每个多项式的核与像也在 T 下不变，它的证明主要是利用了线性变换多项式的可交换性.

引理（线性变换 T 的多项式的核与像在 T 下不变）：设 $T \in \mathcal{L}(V)$，$p \in \mathcal{P}(\mathbf{F})$，则 $\operatorname{Ker} p(T)$ 与 $\operatorname{Im} p(T)$ 在 T 下不变.

证明 任取 $v \in \operatorname{Ker} p(T)$, 则 $p(T)v = 0$, 从而

$$p(T)(Tv) = T(p(T)v) = T0 = 0.$$

故 $Tv \in \operatorname{Ker} p(T)$, 即 $\operatorname{Ker} p(T)$ 在 T 下不变.

任取 $v \in \operatorname{Im} p(T)$, 则存在 $u \in V$ 使得 $v = p(T)u$, 从而

$$Tv = T(p(T)u) = p(T)(Tu) \in \operatorname{Im} p(T).$$

因此 $\operatorname{Im} p(T)$ 在 T 下不变. □

下面的定理给出了复线性空间上算子的刻画.

定理 (复线性空间上算子的刻画): 设 V 为复线性空间, $T \in \mathcal{L}(V)$, $\lambda_1, \cdots, \lambda_m$ 是 T 的全部互异特征值, 则

(1) 每个 $G(\lambda_j, T)$ 在 T 下不变.

(2) 每个 $(T - \lambda_j I)|_{G(\lambda_j, T)}$ 是幂零算子.

证明 (1) $G(\lambda_j, T) = \operatorname{Ker}(T - \lambda_j I)^{\dim V}$, 由前述引理知它是不变子空间 (取多项式 $p(z) = (z - \lambda_j)^{\dim V}$).

(2) 任取 $v \in G(\lambda_j, T) = \operatorname{Ker}(T - \lambda_j I)^{\dim V}$, 则 $(T - \lambda_j I)^{\dim V} v = 0$, 故结论得证. □

注: 结合下面的定理我们可以看到, 复线性空间上的每个算子 $T \in \mathcal{L}(V)$ 都可以看成是由几部分组成的, 其中每一部分 $T|_{G(\lambda_j, T)}$ 都是一个幂零算子加上恒等算子的标量倍:

$$T|_{G(\lambda_j,T)} = (T - \lambda_j I)|_{G(\lambda_j,T)} + \lambda_j I|_{G(\lambda_j,T)}.$$

下面我们证明复线性空间一定能够分解为互异的广义特征子空间的直和. 注意到实线性空间上的线性变换可能连特征值都没有, 因此下述定理自然对实线性空间不成立.

定理 (复线性空间分解为广义特征子空间的直和): 设 V 是复线性空间, $T \in \mathcal{L}(V)$, $\lambda_1, \cdots, \lambda_m$ 是 T 的全部互异特征值, 则

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

证明 设 $\dim V = n$, 我们将对 n 使用数学归纳法来证明结论.

$n = 1$ 时结论显然成立. 设 $n > 1$, 假设结论对任意维数小于 n 的复线性空间都成立, 下证结论

对 n 为复线性空间成立. 复线性空间上的线性变换必有特征值, 设 λ_1 为 T 的特征值, 那么

$$\begin{aligned} V &= \operatorname{Ker}(T-\lambda_1 I)^n \oplus \operatorname{Im}(T-\lambda_1 I)^n \\ &= G(\lambda_1, T) \oplus U. \end{aligned} \tag{1}$$

其中 $U = \operatorname{Im}(T-\lambda_1 I)^n$. 由前述引理知 U 在 T 下不变. 因为 $G(\lambda_1, T)$ 包含了 T 的相应于 λ_1 的所有广义特征向量, 于是 $T|_U$ 没有相应于 λ_1 的广义特征向量, 故 $T|_U$ 的全部特征值为 $\lambda_2, \cdots, \lambda_m$. 由于 $G(\lambda_1, T) \neq \{0\}$, 故 $\dim U < n$, 因此可对 U 使用归纳假设, 从而

$$U = G(\lambda_2, T|_U) \oplus \cdots \oplus G(\lambda_m, T|_U).$$

由 (1) 式知

$$V = G(\lambda_1, T) \oplus G(\lambda_2, T|_U) \oplus \cdots \oplus G(\lambda_m, T|_U).$$

下面证明对每个 $k = 2, \cdots, m$, 都有 $G(\lambda_k, T|_U) = G(\lambda_k, T)$.

显然 $G(\lambda_k, T|_U) \subseteq G(\lambda_k, T)$, 故只需证明 $G(\lambda_k, T) \subseteq G(\lambda_k, T|_U)$. 任取 $v \in G(\lambda_k, T)$, 则有分解

$$v = v_1 + v_2 + \cdots + v_m,$$

其中 $v_1 \in G(\lambda_1, T)$, $v_2 \in G(\lambda_2, T|_U), \cdots, v_m \in G(\lambda_m, T|_U)$. 因为每个 v_j 都是相应于 λ_j 的广义特征向量, 而相应于不同特征值的广义特征向量线性无关, 故除 v_k 外每个 v_j 都为 0, 从而 $v = v_k \in G(\lambda_k, T|_U)$. 结论得证. $\square$

我们知道复线性空间可能没有足够多线性无关的特征向量构成基. 以下命题表明复线性空间一定有足够多线性无关的广义特征向量构成基.

推论 (复线性空间必可取到广义特征向量作为基): 设 V 是复线性空间, $T \in \mathcal{L}(V)$, 则 V 有一个由 T 的广义特征向量构成的基.

证明 设 $\lambda_1, \cdots, \lambda_m$ 为 T 的全部互异特征值. 取每个 $G(\lambda_j, T)$ 的基, 由上述定理, 将它们合在一起就得到了 V 的一个由 T 的广义特征向量构成的基. $\square$

9.5 特征值的重数

设 V 是复线性空间, $T \in \mathcal{L}(V)$, 则由

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$$

给出的分解是一个强大的工具. 包含在这个分解中的子空间的维数非常重要, 我们为其取了一个特别的名字.

定义 (特征值的代数重数)：设 $T \in \mathcal{L}(V)$, λ 为 T 的特征值, 我们称广义特征子空间 $G(\lambda, T)$ 的维数为 λ 的代数重数 (algebraic multiplicity)，简称重数 (multiplicity)．

例 1 定义 $T \in \mathcal{L}(\mathbf{C}^3)$ 为

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

易得 T 关于标准基有上三角矩阵

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

故 T 有特征值 6 与 7. 请自行验证

$$G(6, T) = \operatorname{Ker}(T - 6I)^3 = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{C}\},$$
$$G(7, T) = \operatorname{Ker}(T - 7I)^3 = \{c(10, 2, 1) : c \in \mathbf{C}\}.$$

故特征值 6 的代数重数为 2, 特征值 7 的代数重数为 1. 此外, $\mathbf{C}^3 = G(6, T) \oplus G(7, T)$, 并且 $\mathbf{C}^3$ 有一个由 T 的广义特征向量构成的基 $(1, 0, 0), (0, 1, 0), (10, 2, 1)$.

上述例题中特征值的代数重数之和恰好为空间的维数. 以下命题表明在复线性空间上这是一个普遍成立的结论.

命题 (复线性空间上代数重数之和为线性空间的维数)：设 V 是复线性空间, $T \in \mathcal{L}(V)$, 则 T 的所有特征值的代数重数之和等于 $\dim V$.

证明 利用代数重数的定义, $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$ 以及直和的维数公式即得. □

定义 (几何重数)：设 $T \in \mathcal{L}(V)$, λ 为 T 的特征值, 我们称特征子空间 $E(\lambda, T)$ 的维数为 λ 的几何重数 (geometric multiplicity)．

注：设 $T \in \mathcal{L}(V)$, λ 为 T 的特征值. 按照我们的定义,

$$\lambda \text{ 的代数重数} = \dim G(\lambda, T) = \dim \operatorname{Ker}(T - \lambda I)^{\dim V},$$
$$\lambda \text{ 的几何重数} = \dim E(\lambda, T) = \dim \operatorname{Ker}(T - \lambda I).$$

这里给出的代数重数的定义比涉及行列式的传统定义更加简洁. 后面我们会看到二者是等价的.

9.6 分块对角矩阵

为了用矩阵形式来解释我们的结果, 我们提出以下定义, 它推广了对角矩阵的概念.

定义 (分块对角矩阵): 若方阵 A 形如

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

其中 $A_1, \cdots, A_m$ 位于对角线上且为方阵, 其余元素均为 0, 则称 A 为分块对角矩阵 (block diagonal matrix).

若每个 A_j 都是 1×1 矩阵, 则分块对角矩阵就是对角矩阵.

例 1 5×5 矩阵

$$A = \begin{pmatrix} \begin{pmatrix} 4 \end{pmatrix} & 0 \quad 0 & 0 \quad 0 \\ 0 & \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix} \end{pmatrix}$$

是分块对角矩阵.

命题 (复线性变换具有上三角块的分块对角矩阵): 设 V 是复线性空间, $T \in \mathcal{L}(V)$, $\lambda_1, \cdots, \lambda_m$ 是 T 的全部互异特征值, 重数分别为 $d_1, \cdots, d_m$, 则 V 有一个基使得 T 关于这个基有分块对角矩阵

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

其中每个 A_j 都是如下所示的 $d_j \times d_j$ 上三角矩阵:

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}.$$

证明 由 §9.4 知每个 $(T - \lambda_j I)|_{G(\lambda_j, T)}$ 都是幂零的. $G(\lambda_j, T)$ 是 d_j 维线性空间. 由 §9.3 幂零

算子的矩阵知存在 $G(\lambda_j, T)$ 的某个基, 使得 $(T-\lambda_j I)|_{G(\lambda_j,T)}$ 关于这个基的矩阵形如

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix}.$$

又因为

$$T|_{G(\lambda_j,T)} = (T-\lambda_j I)|_{G(\lambda_j,T)} + \lambda_j I|_{G(\lambda_j,T)},$$

故 $T|_{G(\lambda_j,T)}$ 关于上面这个基的矩阵就是

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}.$$

将这些基放在一起就得到了 V 的一个基, T 关于这个基的矩阵就具有我们想要的形式. □

例 2 定义 $T \in \mathcal{L}(\mathbf{C}^3)$ 为

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

T 关于标准基的矩阵为

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

这是一个上三角阵. $\mathbf{C}^3$ 有一个由 T 的广义特征向量构成的基 $(1,0,0), (0,1,0), (10,2,1)$, T 关于这个基的矩阵为

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

这是前述命题中的具有上三角块的分块对角矩阵.

在后面讨论若尔当标准型时将会看到, 可以找到一个基使得算子关于这个基的矩阵有更多的 0. 尽管如此, 上述定理实际上已经非常有用了.

9.7 平方根

回忆一下, 算子 $T \in \mathcal{L}(V)$ 的平方根是指满足 $R^2 = T$ 的算子 $R \in \mathcal{L}(V)$. 每个复数都有平方根, 但复线性空间上的算子不一定都有平方根. 我们很快就会看到, 若复线性空间上的算子可逆, 那么它就一定有平方根.

我们首先证明恒等算子加上一个幂零算子总有平方根. 注意下面的引理对实线性空间与复线性空间都成立.

引理 (恒等算子加幂零算子总有平方根): 设 $N \in \mathcal{L}(V)$ 是幂零的, 则 $I+N$ 必有平方根.

证明 考虑 $\sqrt{1+x}$ 的泰勒展式

$$\sqrt{1+x} = 1 + a_1 x + a_2 x^2 + \cdots . \tag{1}$$

我们无需找出系数的显式公式, 也不必担心级数是否收敛, 因为我们只想从中收到一些启发.

因为 N 是幂零的, 故存在正整数 m 使得 $N^m = 0$. 由 (1) 式我们可以猜测 $I+N$ 具有形如

$$I + a_1 N + a_2 N^2 + \cdots + a_{m-1} N^{m-1}$$

的平方根. 假设

$$(I + a_1 N + a_2 N^2 + \cdots + a_{m-1} N^{m-1})^2 = I + N,$$

那么

$$I + 2a_1 N + (2a_2 + a_1^2) N^2 + \cdots + (2a_{m-1} + \text{包含 } a_1, \cdots, a_{m-2} \text{ 的项}) N^{m-1} = I + N.$$

于是 $2a_1 = 1$, 从中可解出 a_1. 由 $2a_2 + a_1^2 = 0$ 又可解出 a_2. 如此进行下去, 可以解出每一个 a_j, 并且它们都是实数. 我们不关心这些解的具体值, 只需要知道它们给出了 $I+N$ 的一个平方根即可. □

尽管上述引理对实线性空间与复线性空间都成立, 但下面的结论只对复线性空间成立. 例如实线性空间 $\mathbf{R}$ 上乘以 -1 的算子就没有平方根.

命题 ($\mathbf{C}$ 上的可逆算子都有平方根): 设 V 为复线性空间, $T \in \mathcal{L}(V)$ 是可逆的, 那么 T 有平方根.

证明 设 $\lambda_1, \cdots, \lambda_m$ 为 T 的全部互异特征值. 由 §9.4 定理知 $T|_{G(\lambda_j,T)} = \lambda_j I + N_j$, 其中 $N_j \in \mathcal{L}(G(\lambda_j, T))$ 为幂零算子. 因为 T 可逆, 故每个 λ_j 均不为 0, 从而

$$T|_{G(\lambda_j,T)} = \lambda_j \left(I + \frac{N_j}{\lambda_j} \right).$$

易知 $\dfrac{N_j}{\lambda_j}$ 是幂零的, 从而由前述引理知 $I + \dfrac{N_j}{\lambda_j}$ 有平方根. 取 λ_j 的一个平方根乘以 $I + \dfrac{N_j}{\lambda_j}$ 的平方根即得 $T|_{G(\lambda_j,T)}$ 的一个平方根 R_j.

由于 $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$, 故对任意的 $v \in V$, 存在唯一的 $u_j \in G(\lambda_j, T)$ 使得

$$v = u_1 + \cdots + u_m.$$

定义线性映射 $R \in \mathcal{L}(V)$ 为

$$\begin{aligned} R: \quad & V \longrightarrow V \\ v = u_1 + \cdots + u_m & \longmapsto R_1 u_1 + \cdots + R_m u_m \end{aligned}$$

请自行验证 $R^2 = T$, 故 T 有平方根. □

仿照本节的方法可以证明: 若 V 是复线性空间, $T \in \mathcal{L}(V)$ 是可逆的, 那么对每个正整数 k, T 都有 k 次方根.

9.8 特征多项式

若 $\mathbf{F} = \mathbf{C}$, 则下面的定义把每个 $T \in \mathcal{L}(V)$ 和一个多项式联系在一起. 对于 $\mathbf{F} = \mathbf{R}$, 相应的的定义将在下一章给出.

定义 (特征多项式): 设 V 是复线性空间, $T \in \mathcal{L}(V)$, $\lambda_1, \cdots, \lambda_m$ 为 T 的全部互异特征值, 重数分别为 $d_1, \cdots, d_m$, 则称

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

为 T 的特征多项式 (characteristic polynomial) .

例 1 定义 $T \in \mathcal{L}(\mathbf{C}^3)$ 为

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

由于 T 的特征值是 6 和 7, 重数是 2 和 1, 故 T 的特征多项式为 $(z - 6)^2(z - 7)$.

性质 (特征多项式的次数与零点): 设 V 是复线性空间, $T \in \mathcal{L}(V)$, 则
(1) T 的特征多项式的次数为 $\dim V$.
(2) T 的特征多项式的零点恰好是 T 的特征值.

证明 因为特征值的重数之和为 $\dim V$, 故 (1) 成立. (2) 由特征多项式的定义即得. □

大部分教材利用行列式来定义特征多项式, 后面我们会看到二者是等价的. 这里采用的方法更简单, 并由此得到凯莱-哈密顿定理的一个简单证明. 在下一章我们会看到这个定理对实线性空间也成立.

定理 (凯莱-哈密顿定理)：设 V 是复线性空间, $T \in \mathcal{L}(V)$, q 为 T 的特征多项式, 则 $q(T)=0$.

证明 设 $\lambda_1, \cdots, \lambda_m$ 为 T 的全部互异特征值, 重数分别为 $d_1, \cdots, d_m$, 则

$$q(z)=(z-\lambda_1)^{d_1}\cdots(z-\lambda_m)^{d_m}.$$

因为 $V=G(\lambda_1,T)\oplus\cdots\oplus G(\lambda_m,T)$, 故只需验证 $q(T)|_{G(\lambda_j,T)}=0$ 即可. 因为 $\dim G(\lambda_j,T)=d_j$, $(T-\lambda_j I)|_{G(\lambda_j,T)}$ 是幂零的, 故

$$((T-\lambda_j I)|_{G(\lambda_j,T)})^{d_j}=(T-\lambda_j I)^{d_j}|_{G(\lambda_j,T)}=0.$$

因为

$$q(T)=(T-\lambda_1 I)^{d_1}\cdots(T-\lambda_m I)^{d_m},$$

上式右端的算子是交换的, 因此可以把 $(T-\lambda_j I)^{d_j}$ 移到最后, 从而由 $(T-\lambda_j I)^{d_j}|_{G(\lambda_j,T)}=0$ 得 $q(T)|_{G(\lambda_j,T)}=0$. □

9.9 极小多项式

本节引入与每个 $T \in \mathcal{L}(V)$ 相联系的另一个重要的多项式. 先从以下定义开始.

定义 (首一多项式)：最高次数的项的系数为 1 的多项式称为首一多项式 (monic polynomial) .

例 1 $2+5z^2+z^7$ 是 7 次首一多项式.

引理 (极小多项式的存在唯一性)：设 $T \in \mathcal{L}(V)$, 则存在唯一一个次数最低的首一多项式 p 使得 $p(T)=0$.

证明 先证存在性. 设 $\dim V=n$. 因为 $\dim \mathcal{L}(V)=n^2$, 故 $I, T, T^2, \cdots, T^{n^2}$ 必线性相关. 设 m 是使得 $I, T, \cdots, T^m$ 线性相关的最小正整数. 根据线性相关引理, 上述算子中有一个算子是它前面诸算子的线性组合. 由 m 的定义知 T^m 是 $I, T, \cdots, T^{m-1}$ 的线性组合, 故存在 $a_0, a_1, \cdots, a_{m-1} \in \mathbf{F}$ 使得

$$a_0 I+a_1 T+\cdots+a_{m-1}T^{m-1}+T^m=0.$$

定义 $p \in \mathcal{P}(\mathbf{F})$ 为 $p(z)=a_0+a_1 z+\cdots+a_{m-1}z^{m-1}+z^m$, 则 $p(T)=0$.

再证唯一性. 由 m 的最小性, 设另有次数为 m 的首一多项式 q 满足 $q(T)=0$, 则 $(p-q)(T)=0$, $\deg(p-q)<m$. 由 m 的最小性知 $q=p$. □

定义 (极小多项式)：设 $T \in \mathcal{L}(V)$, 称 T 的唯一一个次数最低的满足 $p(T)=0$ 的首一多项式为 T 的极小多项式 (minimal polynomial) .

设 $\dim V = n$. 上述证明过程告诉我们极小多项式的次数最多为 n^2. 由凯莱-哈密顿定理, 复线性空间上极小多项式的次数最多为 n. 后面将会看到, 这一显著的改进对实线性空间也成立.

定义 (零化多项式)：设 $T \in \mathcal{L}(V)$, $p \in \mathcal{P}(\mathbf{F})$. 若 $p(T)=0$, 则称 p 为 T 的零化多项式.

性质 (零化多项式是极小多项式的倍式)：设 $T \in \mathcal{L}(V)$, $q \in \mathcal{P}(\mathbf{F})$, 则 q 是 T 的零化多项式当且仅当 q 是 T 的极小多项式的多项式倍.

证明 设 T 的极小多项式为 p.

($\Longleftarrow$) 设 $q = ps$, 则 $q(T) = p(T)s(T) = 0$, 故 q 为零化多项式.

($\Longrightarrow$) 由带余除法, 设 $q = ps + r$, $\deg r < \deg p$ 或 $r = 0$. 由于

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

故 $r = 0$ (否则将与 p 为极小多项式矛盾) , 从而 $q = ps$. □

命题 (特征多项式是极小多项式的多项式倍)：设 $T \in \mathcal{L}(V)$, 则 T 的特征多项式是 T 的极小多项式的多项式倍.

证明 由凯莱-哈密顿定理知特征多项式是零化多项式, 再由前述性质即得结论. □

我们知道 (至少当 $\mathbf{F} = \mathbf{C}$ 时), 特征多项式的零点恰好是特征值. 现在我们要证明极小多项式有相同的零点 (尽管这些零点的重数可能不同) .

定理 (特征值是极小多项式的零点)：设 $T \in \mathcal{L}(V)$, 则 λ 是 T 的特征值当且仅当 λ 是 T 的极小多项式的零点.

证明 设 T 的极小多项式为 p.

($\Longleftarrow$) 设 q 为 T 的特征多项式, $q = ps$, 则

$$q(\lambda) = p(\lambda)s(\lambda) = 0 \cdot s(\lambda) = 0,$$

故 λ 为特征多项式的零点, 即 λ 为特征值.

($\Longrightarrow$) 设极小多项式为 $p(z) = a_0 + a_1 z + \cdots + a_m z^m \in \mathcal{P}(\mathbf{F})$. 设 $0 \neq v \in V$ 是 T 的相应于 λ

的特征向量, 则

$$
\begin{aligned}
0 &= p(T)v \\
&= (a_0 I + a_1 T + \cdots + a_m T^m)v \\
&= (a_0 + a_1 \lambda + \cdots + a_m \lambda^m)v \\
&= p(\lambda)v.
\end{aligned}
$$

因为 $v \neq 0$, 故 $p(\lambda) = 0$. $\square$

推论 (特征多项式与极小多项式有相同的根集): 设 $T \in \mathcal{L}(V)$, 则 T 的特征多项式与极小多项式有相同的根集.

例 2 定义 $T \in \mathcal{L}(\mathbf{C}^3)$ 为

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

由于 T 的特征多项式为 $(z-6)^2(z-7)$, 故 T 的极小多项式只可能是 $(z-6)(z-7)$ 或 $(z-6)^2(z-7)$. 直接计算得

$$(T - 6I)(T - 7I) \neq 0,$$

故 T 的极小多项式为 $(z-6)^2(z-7)$.

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

$$T(z_1, z_2, z_3) = (6z_1, 6z_2, 7z_3).$$

由于 T 的特征多项式为 $(z-6)^2(z-7)$, 故 T 的极小多项式只可能是 $(z-6)(z-7)$ 或 $(z-6)^2(z-7)$. 直接计算得

$$(T - 6I)(T - 7I) = 0,$$

故 T 的极小多项式为 $(z-6)(z-7)$.

假定已知算子 $T \in \mathcal{L}(V)$ 关于某个基的矩阵 $\mathcal{M}(T)$, 我们可以来求极小多项式: 对 $m = 1, 2, \cdots$, 相继地考虑方程

$$a_0 \mathcal{M}(I) + a_1 \mathcal{M}(T) + \cdots + a_{m-1} \mathcal{M}(T^{m-1}) = -\mathcal{M}(T^m), \tag{1}$$

直到这个方程有一个解 $a_0, a_1, \cdots, a_{m-1}$. 从而标量 $a_0, a_1, \cdots, a_{m-1}$ 就是 T 的极小多项式的系数.

例 4 设 $T \in \mathcal{L}(\mathbf{C}^5)$, 它关于标准基的矩阵为

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

直接计算 $\mathcal{M}(T)$ 的各次幂, 可得方程 (1) 直到 $m = 5$ 才有解, 且 T 的极小多项式为 $z^5 - 6z + 3$.

9.10 若尔当形 (Jordan Form)

我们知道, 如果 V 是复线性空间, 那么每一个 $T \in \mathcal{L}(V)$ 都有一个具有较好形式的上三角矩阵 (参见 §9.6) . 本节将会得到更好的结论: V 有一个基, 使得 T 关于这个基的矩阵, 除了对角线以及紧位于对角线上方的元素之外, 其余元素均为 0.

首先来看幂零算子的两个例子.

例 1 设 $N \in \mathcal{L}(V)$ 是定义为

$$N(z_1, z_2, z_3, z_4) = (0, z_1, z_2, z_3)$$

的幂零算子. 取 $v = (1, 0, 0, 0)$, 则 $\underline{N^3v, N^2v, Nv, v}$ 是 $\mathbf{F}^4$ 的基, N 关于这个基的矩阵为

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

下面介绍一个稍微复杂一点的幂零算子的例子.

例 2 设 $N \in \mathcal{L}(V)$ 是定义为

$$N(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0)$$

的幂零算子. 它不像上例中的幂零算子那样好, 没有向量 $v \in \mathbf{F}^6$ 使得 $N^5v, N^4v, N^3v, N^2v, Nv, v$ 构成 $\mathbf{F}^6$ 的基. 但是如果取 $v_1 = (1, 0, 0, 0, 0, 0), v_2 = (0, 0, 0, 1, 0, 0), v_3 = (0, 0, 0, 0, 0, 1)$, 那么

$\underline{N^2v_1, Nv_1, v_1}, \underline{Nv_2, v_2}, \underline{v_3}$ 构成 $\mathbf{F}^6$ 的基. N 关于这个基的矩阵为

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

以下命题表明, 每个幂零算子 $N \in \mathcal{L}(V)$ 都与上例性质相似. 具体来说, 存在有限个向量 $v_1, \cdots, v_s \in V$, 使得 V 有一个由形如 $N^k v_j$ 的向量构成的基, 其中 j 从 1 取到 s, 而 k 则从 m_j 降序取到 0, 其中 m_j 是使得 $N^{m_j} v_j \neq 0$ 的最大非负整数. 以下命题的矩阵解释, 参见后面幂零算子若尔当基存在性的证明.

命题 (幂零算子具有降幂排序基): 设 $N \in \mathcal{L}(V)$ 是幂零的, 则存在 $v_1, \cdots, v_s \in V$ 与非负整数 $m_1, \cdots, m_s$, 使得

(1) $\underline{N^{m_1}v_1, \cdots, Nv_1, v_1}, \cdots, \underline{N^{m_s}v_s, \cdots, Nv_s, v_s}$ 构成 V 的基.

(2) $N^{m_1+1}v_1 = \cdots = N^{m_s+1}v_s = 0$.

在以下的定义中, 每个 A_j 的对角线上都是 T 的特征值 λ_j(因为上三角矩阵对角线上的元素就是算子的特征值), 而紧位于对角线上方的元素都为 1, A_j 的其余元素均为 0. 这些 λ_j 不需要是不同的, A_j 也可以只是一个 1×1 矩阵 $\begin{pmatrix} \lambda_j \end{pmatrix}$.

定义 (若尔当基): 设 $T \in \mathcal{L}(V)$, $v_1, \cdots, v_n$ 是 V 的基. 我们称 $v_1, \cdots, v_n$ 是 V 关于 T 的若尔当基 (Jordan basis), 如果 T 关于这个基有分块对角矩阵

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_s \end{pmatrix},$$

其中每个 A_j 都是形如

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

的上三角矩阵.

下面的命题告诉我们, 幂零算子的降幂排序基就是它的若尔当基.

命题 (幂零算子必有若尔当基)：设 $N \in \mathcal{L}(V)$ 是幂零算子, 则 V 有一个基是 N 的若尔当基.

证明 考虑 N 的降幂排序基 $\underline{N^{m_1}v_1, \cdots, Nv_1, v_1}, \cdots, \underline{N^{m_s}v_s, \cdots, Nv_s, v_s}$. 首先, 由于每一个 $\operatorname{span}\{N^{m_j}v_j, \cdots, Nv_j, v_j\}$ 都是 N 的不变子空间. 因此, N 关于降幂排序基具有分块对角矩阵. 又因为对于每个 j, N 都将组 $N^{m_j}v_j, \cdots, Nv_j, v_j$ 中的第一个向量映成 0, 并将这个组中的其余向量映成它的前一个向量, 因此对角线上的每块矩阵都形如

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

结论得证. □

定理 (复算子必有若尔当基)：设 V 是复线性空间, $T \in \mathcal{L}(V)$, 则 V 有一个基是 T 的若尔当基.

证明 设 $\lambda_1, \cdots, \lambda_m$ 是 T 的全部互异特征值, 则

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

因为 $(T - \lambda_j I)|_{G(\lambda_j, T)}$ 是幂零的, 由前一命题知 $G(\lambda_j, T)$ 具有 $(T - \lambda_j I)|_{G(\lambda_j, T)}$ 的若尔当基. 由

$$T|_{G(\lambda_j, T)} = (T - \lambda_j I)|_{G(\lambda_j, T)} + \lambda_j I|_{G(\lambda_j, T)}$$

知这个若尔当基也是 $T|_{G(\lambda_j, T)}$ 的若尔当基 (相应矩阵对角线上的元素从 0 变为了 λ_j). 将这些基组合起来就得到了 V 的一个基, 它是 T 的若尔当基. □

注：1870 年法国数学家卡米耶 · 若尔当首先发表了上述结论的证明.

10 实线性空间上的算子

上一章我们学习了有限维复线性空间上算子的结构. 本章将利用关于复线性空间上算子的结果来学习实线性空间上的算子.

本章用 $\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$, 用 V 表示 $\mathbf{F}$ 上的有限维非零线性空间.

10.1 实线性空间的复化

我们马上就会看到, 一个实线性空间可以自然地嵌入到一个复线性空间中, 后者称为 V 的复化. V 上的每个算子都可以扩张为 V 的复化上的算子. 因此关于复线性空间上算子的结果可以转化为实线性空间上算子的信息.

我们先定义实线性空间的复化.

定义 (实线性空间的复化): 设 V 是实线性空间. V 的复化 (记为 $V_{\mathbf{C}}$) 等于 $V\times V$, 其元素是有序对 (u,v), 但我们把它记作 $u+\mathrm{i}v$, 其中 $u,v\in V$.

定义 ($V_{\mathbf{C}}$ 上的加法与数乘): 设 V 是实线性空间, $V_{\mathbf{C}}$ 是 V 的复化. 定义 $V_{\mathbf{C}}$ 上的加法为

$$(u_1+\mathrm{i}v_1)+(u_2+\mathrm{i}v_2)=(u_1+u_2)+\mathrm{i}(v_1+v_2).$$

其中 $u_1,u_2,v_1,v_2\in V$. 定义 $V_{\mathbf{C}}$ 上的复标量乘法为

$$(a+\mathrm{i}b)(u+\mathrm{i}v)=(au-bv)+\mathrm{i}(av+bu),$$

其中 $a,b\in\mathbf{R}$, $u,v\in V$.

上述复标量乘法定义的动机来自于普通的复数乘法. 记住这一点, 就不必去背上面的定义了.

通过将 $u\in V$ 与 $u+\mathrm{i}0\in V_{\mathbf{C}}$ 等同起来, 就可以把 V 看做 $V_{\mathbf{C}}$ 的子集. 因此, 从 V 构造 $V_{\mathbf{C}}$ 可以看做是从 $\mathbf{R}^n$ 构造 $\mathbf{C}^n$ 的推广.

命题 ($V_{\mathbf{C}}$ 是复线性空间): 设 V 是实线性空间, 则 $V_{\mathbf{C}}$ 按照上面定义的加法与标量乘法作成一个复线性空间.

证明 请读者自行验证. □

注: $V_{\mathbf{C}}$ 的零元是 $0+\mathrm{i}0$, 通常写作 0.

关于复化, 你认为成立的可能都成立, 通常都只需要验证一下, 就像以下命题阐述的那样.

命题 (V 的基是 $V_{\mathbf{C}}$ 的基): 设 V 是实线性空间, $v_1,\cdots,v_n$ 是 V 的基, 那么 $v_1,\cdots,v_n$ 也是 $V_{\mathbf{C}}$ 的基.

证明 $V=\operatorname{span}\{v_1,\cdots,v_n\}$, 因此 $V_{\mathbf{C}}$ 中 $\operatorname{span}\{v_1,\cdots,v_n\}$ 包含了 $v_1,\cdots,v_n,\mathrm{i}v_1,\cdots,\mathrm{i}v_n$, 故

$$V_{\mathbf{C}}=\operatorname{span}\{v_1,\cdots,v_n\}.$$

又设 $c_1, \cdots, c_n \in \mathbf{C}$ 满足

$$c_1 v_1 + \cdots + c_n v_n = 0 = 0 + \mathrm{i}0.$$

那么

$$(\operatorname{Re} c_1) v_1 + \cdots + (\operatorname{Re} c_n) v_n = 0,$$
$$(\operatorname{Im} c_1) v_1 + \cdots + (\operatorname{Im} c_n) v_n = 0.$$

因为 $v_1, \cdots, v_n$ 在 V 中线性无关, 因此 $c_1 = \cdots = c_n = 0$, 故 $v_1, \cdots, v_n$ 在 $V_{\mathbf{C}}$ 中也线性无关. □

推论 (V **与** $V_{\mathbf{C}}$ **具有相同的维数**) : 实线性空间 V 与复线性空间 $V_{\mathbf{C}}$ 具有相同的维数.

证明 由前述命题即得. □

10.2 算子的复化

定义 (算子的复化) : 设 V 是实线性空间, $T \in \mathcal{L}(V)$, 定义 T 的复化 $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$ 为对任意的 $u, v \in V$,

$$T_{\mathbf{C}}(u + \mathrm{i}v) = Tu + \mathrm{i}\,Tv.$$

请自行验证, 若 V 是实线性空间, $T \in \mathcal{L}(V)$, 那么 $T_{\mathbf{C}}$ 确实属于 $\mathcal{L}(V_{\mathbf{C}})$. 这里的关键是根据复标量乘法的定义, 我们可以证明对任意的 $u, v \in V$ 与 $\lambda \in \mathbf{C}$, 均有

$$T_{\mathbf{C}}(\lambda(u + \mathrm{i}v)) = \lambda T_{\mathbf{C}}(u + \mathrm{i}v).$$

注 : 由前述定义直接可得对任意的 $v \in V$, 都有

$$T_{\mathbf{C}} v = Tv.$$

因此算子的复化可以理解为扩大原算子的定义域 (从 V 扩大为 $V_{\mathbf{C}}$) .

下面的例子是理解典型算子复化的好方法.

例 1 设 A 为 $n \times n$ 实矩阵. 定义 $T \in \mathcal{L}(\mathbf{R}^n)$ 为

$$Tx = Ax,$$

其中将 $\mathbf{R}^n$ 中的元素看做 $n \times 1$ 列向量. 将 $\mathbf{C}^n$ 等同于 $\mathbf{R}^n$ 的复化, 对每个 $z \in \mathbf{C}^n$ 则有

$$T_{\mathbf{C}} z = Az,$$

其中将 $\mathbf{C}^n$ 中的元素看做 $n\times 1$ 列向量. 即是说, 如果 T 是由 A 确定的 $\mathbf{R}^n$ 上的矩阵乘算子, 那么复化 $T_{\mathbf{C}}$ 也是由 A 确定的矩阵乘算子, 只是定义域扩大为 $\mathbf{C}^n$.

下面的命题是有意义的, 因为实线性空间的基也是它的复化的基.

命题 (T **与** $T_{\mathbf{C}}$ **有相同的矩阵**) : 设 V 为实线性空间, $T\in\mathcal{L}(V)$, $v_1,\cdots,v_n$ 是 V 的基, 则

$$\mathcal{M}(T)=\mathcal{M}(T_{\mathbf{C}}).$$

证明 注意到 $T_{\mathbf{C}}v_j=Tv_j$, 由此易得 $\mathcal{M}(T)=\mathcal{M}(T_{\mathbf{C}})$. □

我们知道, 有限维复线性空间上的算子必有特征值, 因而有一维不变子空间. 但有限维实线性空间上的算子可能没有特征值, 因此可能没有一维不变子空间. 幸运的是, 实线性空间如果没有一维不变子空间, 那么它一定有二维不变子空间. 注意到复化是怎样给出以下定理的一个简单证明的.

定理 (**每个算子必有一维或二维不变子空间**) : 非零有限维线性空间上的算子必有一维或二维不变子空间.

证明 因为复线性空间总有一维不变子空间, 因此只需对实线性空间证明. 设 V 是实线性空间, $T\in\mathcal{L}(V)$, 则 $T_{\mathbf{C}}$ 有特征值. 设 $a+\mathrm{i}b$ 为 $T_{\mathbf{C}}$ 的特征值, 则存在不全为 0 的向量 $u,v\in V$, 使得

$$T_{\mathbf{C}}(u+\mathrm{i}v)=(a+\mathrm{i}b)(u+\mathrm{i}v).$$

也即

$$Tu+\mathrm{i}Tv=(au-bv)+\mathrm{i}(av+bu).$$

因此

$$\begin{aligned}Tu&=au-bv,\\ Tv&=av+bu.\end{aligned}$$

令 $U=\operatorname{span}\{u,v\}$. 由上式知 U 在 T 下不变. 因为 u,v 不全为 0, 因此 $\dim U=1$ 或 2, 即 U 是 T 的一维或二维不变子空间. □

10.3 复化的极小多项式

设 V 是实线性空间, 反复利用 $T_{\mathbf{C}}$ 的定义知对任意的正整数 n 与 $u,v\in V$, 均有

$$(T_{\mathbf{C}})^n(u+\mathrm{i}v)=T^nu+\mathrm{i}\,T^nv. \tag{1}$$

以下的命题说明 $T_{\mathbf{C}}$ 的极小多项式的系数都是实数.

命题 (T **的极小多项式等于** $T_{\mathbf{C}}$ **的极小多项式**)：设 V 是实线性空间, $T \in \mathcal{L}(V)$, 则 $T_{\mathbf{C}}$ 的极小多项式等于 T 的极小多项式.

证明 设 $p \in \mathcal{P}(\mathbf{R})$ 是 T 的极小多项式, 则由 (1) 式知 $p(T_{\mathbf{C}}) = (p(T))_{\mathbf{C}} = 0$, 故 p 为 $T_{\mathbf{C}}$ 的零化多项式.

设 $q \in \mathcal{P}(\mathbf{C})$ 是 $T_{\mathbf{C}}$ 的首一零化多项式, 则 $q(T_{\mathbf{C}}) = 0$. 下证 $\deg q \geqslant \deg p$. 注意到对任意的 $u \in V$, 有 $q(T_{\mathbf{C}})u = 0$. 令 h 表示第 j 个系数是 q 的第 j 个系数的实部的多项式, 则 $h \in \mathcal{P}(\mathbf{R})$ 是首一的, 且 $h(T) = 0$. 因此 $\deg q = \deg h \geqslant \deg p$. 综上, p 为 $T_{\mathbf{C}}$ 的极小多项式. □

10.4 复化的特征值

现在转向算子的复化特征值问题. 和前面一样, 我们期望的性质都成立.

先来证明 $T_{\mathbf{C}}$ 的所有实特征值恰为 T 的所有特征值. 这里给出两种证明方法. 第一种较为初等. 第二种更短, 并且给出一些有用的见解.

命题 ($T_{\mathbf{C}}$ **的所有实特征值恰为** T **的所有特征值**)：设 V 是实线性空间, $T \in \mathcal{L}(V)$, 则 λ 为 $T_{\mathbf{C}}$ 的实特征值当且仅当 λ 为 T 的特征值.

证明 **方法一**: ($\Longleftarrow$) 设 v 为 T 的相应于 λ 的特征向量, 则 $Tv = \lambda v$. 由于 $T_{\mathbf{C}}v = Tv$, 故 $T_{\mathbf{C}}v = \lambda v$. 因此 λ 为 $T_{\mathbf{C}}$ 的特征值.

($\Longrightarrow$) 存在不全为 0 的向量 $u, v \in V$ 使得 $T_{\mathbf{C}}(u + \mathrm{i}v) = \lambda(u + \mathrm{i}v)$, 也即 $Tu + \mathrm{i}Tv = \lambda(u + \mathrm{i}v)$, 从而 $Tu = \lambda u$, $Tv = \lambda v$. 因为 u, v 不全为 0, 故 λ 为 T 的特征值.

方法二:

$$\begin{aligned}
\lambda \text{ 为 } T_{\mathbf{C}} \text{ 的实特征值} &\Longleftrightarrow \lambda \text{ 为 } T_{\mathbf{C}} \text{ 的极小多项式的实根} \\
&\Longleftrightarrow \lambda \text{ 为 } T \text{ 的极小多项式的根} \\
&\Longleftrightarrow \lambda \text{ 为 } T \text{ 的特征值.}
\end{aligned}$$

结论得证. □

以下命题表明, $T_{\mathbf{C}}$ 对于特征值 λ 与其复共轭 $\bar{\lambda}$ 的表现是对称的.

命题 ($T_{\mathbf{C}} - \lambda I$ **与** $T_{\mathbf{C}} - \bar{\lambda} I$ **的关系**)：设 V 为实线性空间, $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{C}$, $u, v \in V$, j 为非负整数, 则

$$(T_{\mathbf{C}} - \lambda I)^j (u + \mathrm{i}v) = 0 \quad \Longleftrightarrow \quad (T_{\mathbf{C}} - \bar{\lambda} I)^j (u - \mathrm{i}v) = 0.$$

证明 对 j 使用归纳法. □

上述结论有一个重要的推论如下, 它说的是若一个复数是 $T_{\mathbf{C}}$ 的特征值, 则它的复共轭也是 $T_{\mathbf{C}}$ 的特征值.

命题 ($T_{\mathbf{C}}$ **的非实特征值成对出现**)：设 V 是实线性空间, $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{C}$, 则 λ 是 $T_{\mathbf{C}}$ 的特征值当且仅当 $\bar{\lambda}$ 是 $T_{\mathbf{C}}$ 的特征值.

证明 前述命题中 j 取 1 即得. □

按照定义, 实线性空间上算子的特征值都是实数. 有时候数学家们会非正式地提到实线性空间上算子的复特征值, 他们指的是这个算子的复化的特征值.

回忆一下, 特征值的重数是指相应的广义特征子空间的重数.

命题 (λ **的重数等于** $\bar{\lambda}$ **的重数**)：设 V 是实线性空间, $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{C}$ 是 $T_{\mathbf{C}}$ 的特征值, 则 λ 作为 $T_{\mathbf{C}}$ 的特征值的重数等于 $\bar{\lambda}$ 作为 $T_{\mathbf{C}}$ 的特征值的重数.

证明 设 $u_1+\mathrm{i}v_1, \cdots, u_m+\mathrm{i}v_m$ 是 $G(\lambda, T_{\mathbf{C}})$ 的基, 请自行验证 $u_1-\mathrm{i}v_1, \cdots, u_m-\mathrm{i}v_m$ 是 $G(\bar{\lambda}, T_{\mathbf{C}})$ 的基. 因此 λ 与 $\bar{\lambda}$ 均有重数 m. □

我们已经见过 $\mathbf{R}^2$ 上没有特征值的算子的例子了. 以下命题说明 $\mathbf{R}^3$ 上不可能出现这样的情况.

命题 (**奇数维实线性空间上特征值的存在性**)：奇数维实线性空间上的算子必有特征值.

证明 设 V 为奇数维实线性空间, $T \in \mathcal{L}(V)$. 由前两个命题知 $T_{\mathbf{C}}$ 的非实特征值的重数之和为偶数. 因为 $T_{\mathbf{C}}$ 的所有特征值的重数之和等于 $\dim V_{\mathbf{C}}$ 为奇数 (因为 $\dim V_{\mathbf{C}} = \dim V$), 因此 $T_{\mathbf{C}}$ 必有实特征值. 由本节第一个命题知 T 有特征值. □

10.5 复化的特征多项式

上一章我们定义了有限维复线性空间上算子的特征多项式. 要定义有限维实线性空间上算子的特征多项式, 以下命题是关键的一步.

命题 ($T_{\mathbf{C}}$ **的特征多项式的系数均为实数**)：设 V 为实线性空间, $T \in \mathcal{L}(V)$, 则 $T_{\mathbf{C}}$ 的特征多项式的系数均为实数.

证明 设 λ 是 $T_{\mathbf{C}}$ 的 m 重特征值, 则由上节命题知 $\bar{\lambda}$ 也为 $T_{\mathbf{C}}$ 的 m 重特征值. 由特征多项式的定义, $T_{\mathbf{C}}$ 的特征多项式包含因子 $(z-\lambda)^m$ 与 $(z-\bar{\lambda})^m$. 将这两个因子相乘, 得

$$(z-\lambda)^m(z-\bar{\lambda})^m = (z^2 - 2(\operatorname{Re}\lambda)z + |\lambda|^2)^m.$$

上式右端多项式的系数都是实数.

$T_{\mathbf{C}}$ 的特征多项式是上面形式的项与 $(z-t)^d$ 形式的项的乘积, 其中 t 是 $T_{\mathbf{C}}$ 的实特征值, 且重数为 d. 因此 $T_{\mathbf{C}}$ 的特征多项式的系数都是实数. □

现在我们可以定义实线性空间上算子的特征多项式了.

定义 (**特征多项式**)：设 V 为实线性空间, $T \in \mathcal{L}(V)$, 我们称 $T_{\mathbf{C}}$ 的特征多项式为 T 的特征多项式.

性质 (**实线性空间上算子的特征多项式**)：设 V 为实线性空间, $T \in \mathcal{L}(V)$, 则

(1) T 的特征多项式的系数都是实的.

(2) T 的特征多项式的次数等于 $\dim V$.

(3) T 的所有特征值恰为 T 的特征多项式的所有实零点.

上一章我们对复线性空间证明了凯莱-哈密顿定理, 现在我们也能对实线性空间证明它了.

定理 (**凯莱-哈密顿定理**)：设 V 是实线性空间, $T \in \mathcal{L}(V)$, q 为 T 的特征多项式, 则 $q(T) = 0$.

证明 由复的凯莱-哈密顿定理知 $q(T_{\mathbf{C}}) = 0$, 从而 $q(T) = 0$. □

推论：设 V 是实线性空间, $T \in \mathcal{L}(V)$, 则 T 的极小多项式的次数至多为 $\dim V$.

证明 由凯莱-哈密顿定理即得. □

现在我们证明另一个我们之前只对复的情况才知道的结果.

性质 (**特征多项式是极小多项式的多项式倍**)：设 V 是实线性空间, $T \in \mathcal{L}(V)$, 则 T 的特征多项式是 T 的极小多项式的多项式倍.

证明 由凯莱-哈密顿定理与零化多项式的性质即得. □

10.6 实内积空间上的正规算子

复谱定理完整地描述了复内积空间上的正规算子. 本小节将完整地描述了实内积空间上的正规算子.

我们先描述二维实内积空间上非自伴的正规算子.

命题 (**二维实内积空间上非自伴正规算子的刻画**)：设 V 是二维实内积空间, $T \in \mathcal{L}(V)$, 则下面各条等价:

(1) T 是非自伴的正规算子.

(2) T 关于 V 的每个规范正交基有形如 $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ 的形式, 其中 $b \neq 0$.

(3) T 关于 V 的某个规范正交基有形如 $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ 的形式, 其中 $b > 0$.

证明 我们将证明 (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (1) .

(1) $\Longrightarrow$ (2) 设 e_1, e_2 为 V 的任一规范正交基. 又设

$$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

由正规算子的特点知 $\|Te_1\| = \|T^*e_1\|$, 从而 $b^2 = c^2$. 由于 T 非自伴, 故 $b \neq 0$, $c \neq b$. 从而 $c = -b$.

(2) $\Longrightarrow$ (3) 由 (2) 可设 e_1, e_2 为 V 的规范正交基, 且

$$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

其中 $b \neq 0$. 若 $b > 0$, 结论成立. 若 $b < 0$, 则

$$\mathcal{M}(T, (e_1, -e_2)) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

其中 $-b > 0$, 结论成立.

(3) $\Longrightarrow$ (1) 设 T 关于规范正交基 e_1, e_2 下有矩阵

$$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

其中 $b > 0$. 直接计算得 $\mathcal{M}(TT^*) = \mathcal{M}(T^T)$, 故 T 正规. 又 $b > 0$, 故 $\mathcal{M}(T) \neq \mathcal{M}(T^*)$, 从而 T 非自伴. $\square$

下面的结论告诉我们, 正规算子限制到不变子空间上仍然是正规算子. 这将允许我们对 $\dim V$ 使用归纳法来证明正规算子的刻画.

引理 (正规算子与不变子空间): 设 V 是内积空间, $T \in \mathcal{L}(V)$ 是正规的, U 是 V 的在 T 下不变的子空间, 则

(1) $U^\perp$ 在 T 下不变.

(2) U 在 T^* 下不变.

(3) $(T|_U)^* = (T^*)_U$.

(4) $T|_U \in \mathcal{L}(U)$ 与 $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ 均为正规算子.

注：上述引理中 T 正规这一条件是必不可少的. 例如定义 $T \in \mathcal{L}(\mathbf{R}^2)$ 为

$$T(x, y) = (x + y, y).$$

令 $U = \operatorname{span}\{(1,0)\}$, 则 U 在 T 下不变. 但是 $U^{\perp} = \operatorname{span}\{(1,0)\}$ 却并不是在 T 下不变的, 因为对 $(1,0) \in U^{\perp}$, 有

$$T(1,0) = (1,1) \notin U^{\perp}.$$

以下定理表明实内积空间上的正规算子接近于有对角矩阵. 具体来说, 我们得到每个块最大为 2×2 的分块对角矩阵.

定理 (实内积空间上正规算子的刻画)：设 V 为实内积空间, $T \in \mathcal{L}(V)$, 则下面各条等价:

(1) T 是正规的.

(2) T 关于 V 的某个基有分块对角矩阵, 且这个矩阵对角线上的块或是 1×1 矩阵, 或是形如

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

的 2×2 矩阵, 其中 $b > 0$.

注：我们不能期望得到比上述定理更好的结果, 因为在实内积空间上存在关于任意基都没有对角矩阵的正规算子. 例如由 $T(x, y) = (-y, x)$ 定义的 $T \in \mathcal{L}(\mathbf{R}^2)$ 是正规的 (请自行验证), 但它没有特征值, 因此它关于任何基都没有上三角矩阵, 更不用说对角矩阵了.

10.7 实内积空间上的等距同构

正如我们将要看到的, 以下例子是实内积空间上等距同构的关键组成部分. 此外, 它表明 $\mathbf{R}^2$ 上的等距同构可以没有特征值.

例 1 设 $\theta \in \mathbf{R}$, 则 $\mathbf{R}^2$ 上 (以原点为中心) 的逆时针旋转 θ 角度的算子是等距同构, 这在几何上是显然的. 这个算子关于标准基的矩阵为

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

若 θ 不是 π 的整数倍, 则 $\mathbf{R}^2$ 上没有可以映为其自身标量倍的非零向量, 因此这个算子没有特征值.

以下定理表明, 实内积空间上的每个等距同构都是由以下三部分组成: 一部分是二维子空间上的旋转, 一部分是恒等算子, 一部分是乘以 -1.

定理 (实内积空间上等距同构的刻画)：设 V 为实内积空间, $T \in \mathcal{L}(V)$, 则下面各条等价:

(1) T 是等距同构.

(2) T 关于 V 的某个基有分块对角矩阵, 且这个矩阵对角线上的块是由 1 或 -1 构成的 1×1 矩阵, 或是形如

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

的 2×2 矩阵, 其中 $\theta \in (0, \pi)$.

11 迹与行列式

本书的重点是线性映射与算子, 而不是矩阵. 但本章对矩阵的关注要多一些, 因为我们将要定义并讨论算子的行列式与迹, 然后把这些概念与矩阵的相应概念联系起来. 最后我们将要解释在体积和积分理论中所起到的重要作用.

本章我们用 $\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$, 用 V 表示 $\mathbf{F}$ 上的有限维非零线性空间.

11.1 基的变更

为了研究迹和行列式, 我们需要了解在基变更时算子的矩阵是如何变化的, 因此我们先给出关于基变更的必要材料.

恒等算子 $I \in \mathcal{L}(V)$ 关于任何基的矩阵都是对角线上为 1, 其余位置为 0 的对角矩阵. 在如下的定义中, 我们仍然用 I 表示这个矩阵.

定义 (单位矩阵)：设 n 为正整数, 我们称 $n \times n$ 矩阵

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

为单位矩阵 (identity matrix), 记作 I.

我们用 I 表示恒等算子与单位矩阵, 你应该能从上下文中确定 I 是指的什么. 例如在等式 $\mathcal{M}(I) = I$ 中, 左边的 I 是恒等算子, 右边的 I 是单位矩阵.

设 A 是与 I 大小相同的 ($\mathbf{F}$ 上的) 矩阵, 那么请自行验证

$$AI = IA = A.$$

定义（可逆）：设 A 为方阵. 若存在一个同样大小的方阵 B 使得

$$AB = BA = I,$$

则称 A 是可逆的 (invertible), 且称 B 是 A 的逆 (inverse).

与线性映射逆映射的唯一性证明类似, 我们可以证明若 A 的逆矩阵存在, 则一定是唯一的. 因此我们可以使用记号 A^{-1} 来表示 A 的逆.

注：有些数学家使用术语非奇异的和奇异的, 它们的意思分别与可逆的和不可逆的相同.

下面我们要对 $T \in \mathcal{L}(V)$ 用到 V 的两个基, 而不像之前大多数时候那样使用同一个基. 以下结论的证明参见线性映射乘积的矩阵.

命题（线性映射乘积的矩阵）：设 $S, T \in \mathcal{L}(V)$, $u_1, \cdots, u_n$ 和 $v_1, \cdots, v_n$ 以及 $w_1, \cdots, w_n$ 都是 V 的基, 那么

$$\begin{aligned}&\mathcal{M}(ST, (u_1, \cdots, u_n), (w_1, \cdots, w_n))\\ =&\mathcal{M}(S, (v_1, \cdots, v_n), (w_1, \cdots, w_n))\mathcal{M}(T, (u_1, \cdots, u_n), (v_1, \cdots, v_n)).\end{aligned}$$

以下命题讨论的是恒等算子 I 关于两个不同的基的矩阵. 注意把 u_j 写成 $v_1, \cdots, v_n$ 的线性组合, 对应的系数就是 $\mathcal{M}(I, (u_1, \cdots, u_n), (v_1, \cdots, v_n))$ 的第 j 列.

命题（恒等算子关于两个基的矩阵）：设 $u_1, \cdots, u_n$ 与 $v_1, \cdots, v_n$ 都是 V 的基, 那么 $\mathcal{M}(I, (u_1, \cdots, u_n), (v_1, \cdots, v_n))$ 与 $\mathcal{M}(I, (v_1, \cdots, v_n), (u_1, \cdots, u_n))$ 都是可逆的, 且它们互为逆.

证明 用 I 代替前述命题的 S 与 T, 用 u_j 代替 w_j 即得

$$\begin{aligned}I &= \mathcal{M}(I, (u_1, \cdots, u_n), (v_1, \cdots, v_n))\mathcal{M}(I, (v_1, \cdots, v_n), (u_1, \cdots, u_n)),\\ I &= \mathcal{M}(I, (v_1, \cdots, v_n), (u_1, \cdots, u_n))\mathcal{M}(I, (u_1, \cdots, u_n), (v_1, \cdots, v_n)).\end{aligned}$$

由此即得结论. □

例 1 考虑 $\mathbf{F}^2$ 的基 $(4, 2), (5, 3)$ 和 $(1, 0), (0, 1)$. 显然

$$\mathcal{M}(I, ((4, 2), (5, 3)), ((1, 0), (0, 1))) = \begin{pmatrix} 4 & 5 \\ 2 & 3 \end{pmatrix}.$$

请自行验证上面矩阵的逆为

$$\begin{pmatrix} \frac{3}{2} & -\frac{5}{2} \\ -1 & 2 \end{pmatrix}.$$

从而由上述命题知

$$\mathcal{M}(I,((1,0),(0,1)),((4,2),(5,3))) = \begin{pmatrix} \frac{3}{2} & -\frac{5}{2} \\ -1 & 2 \end{pmatrix}.$$

现在我们可以看到在基变更时 T 的矩阵是怎样变化的了. 回忆一下, 记号 $\mathcal{M}(T,(u_1,\cdots,u_n))$ 是 $\mathcal{M}(T,(u_1,\cdots,u_n),(u_1,\cdots,u_n))$ 的简写.

命题 (基变更公式): 设 $T \in \mathcal{L}(V)$, $u_1,\cdots,u_n$ 与 $v_1,\cdots,v_n$ 是 V 的两个基. 又记 $\mathcal{M}(I,(u_1,\cdots,u_n),(v_1,\cdots,v_n))$ 为 A, 则

$$\mathcal{M}(T,(u_1,\cdots,u_n)) = A^{-1}\mathcal{M}(T,(v_1,\cdots,v_n))A.$$

证明 由线性映射乘积的矩阵知

$$\mathcal{M}(T,(u_1,\cdots,u_n)) = A^{-1}\mathcal{M}(T,(u_1,\cdots,u_n),(v_1,\cdots,v_n)),$$
$$\mathcal{M}(T,(u_1,\cdots,u_n),(v_1,\cdots,v_n)) = \mathcal{M}(T,(v_1,\cdots,v_n))A.$$

将第二式带入第一式即得结论. □

11.2 迹

定义 (算子的迹): 设 $T \in \mathcal{L}(V)$, 它的迹 (trace) 按如下方式定义:

(1) 若 $\mathbf{F} = \mathbf{C}$, 则定义 T 的迹为 T 的全体特征值之和 (按重数重复).

(2) 若 $\mathbf{F} = \mathbf{R}$, 则定义 T 的迹为 T_{C} 的全体特征值之和 (按重数重复).

算子 T 的迹记为 $\operatorname{tr} T$. 迹和特征多项式联系紧密. 设 $\lambda_1,\cdots,\lambda_n$ 为 T 的特征值 (若为实线性空间, 则为 T_{C} 的特征值), 其中每个特征值按重数重复. 则由定义, T 的特征多项式等于

$$(z-\lambda_1)\cdots(z-\lambda_n).$$

将上式展开, 则 T 的特征多项式可写为

$$z^n - (\lambda_1+\cdots+\lambda_n)z^{n-1} + \cdots + (-1)^n\lambda_1\cdots\lambda_n.$$

由上式立即得到以下定理.

定理 (**迹与特征多项式**)：设 $T \in \mathcal{L}(V)$, $\dim V = n$, 则 $\operatorname{tr} T$ 等于 T 的特征多项式中 z^{n-1} 的系数的相反数.

本节的其余部分主要来讨论如何利用 T 关于任意一个基的矩阵计算 $\operatorname{tr} T$.

定义 (**矩阵的迹**)：定义方阵 A 的迹为 A 的对角线元素之和, 记作 $\operatorname{tr} A$.

现在我们已经定义了算子的迹与矩阵的迹. 你或许已经发现, 我们在两个不同的环境中使用了两个相同的词 "迹". 只有证明两个概念本质上是一样的, 这个术语才是合适的. 我们将会看到

$$\operatorname{tr} T = \operatorname{tr} \mathcal{M}(T, (v_1, \cdots, v_n))$$

确实是对的, 其中 $v_1, \cdots, v_n$ 是 V 的任意一个基.

引理 (**AB 的迹等于 BA 的迹**)：设 A 与 B 为同阶方阵, 则 $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

现在我们可以证明算子关于某个基的矩阵的对角线元素之和并不依赖于这个基.

定理 (**迹与基的选取无关**)：设 $T \in \mathcal{L}(V)$, $u_1, \cdots, u_n$ 与 $v_1, \cdots, v_n$ 是 V 的两个基, 则

$$\operatorname{tr} \mathcal{M}(T, (u_1, \cdots, u_n)) = \operatorname{tr} \mathcal{M}(T, (v_1, \cdots, v_n)).$$

证明 设 $A = \mathcal{M}(I, (u_1, \cdots, u_n), (v_1, \cdots, v_n))$, 则

$$\begin{aligned}
\operatorname{tr} \mathcal{M}(T, (u_1, \cdots, u_n)) &= \operatorname{tr} \left(A^{-1} \left(\mathcal{M}(T, (v_1, \cdots, v_n)) A\right)\right) \\
&= \operatorname{tr} \left(\left(\mathcal{M}(T, (v_1, \cdots, v_n)) A\right) A^{-1}\right) \\
&= \operatorname{tr} \mathcal{M}(T, (v_1, \cdots, v_n)).
\end{aligned}$$

其中第一个等号利用了基变更公式, 第二个等号利用了前述引理. □

以下定理是本节最重要的结果. 它说的是算子的迹等于算子的矩阵的迹. 这个定理没有指明用到的基, 因为根据上面的结果, 对每个基来说, 算子的矩阵的迹都是一样的.

定理 (**算子的迹等于算子的矩阵的迹**)：若 $T \in \mathcal{L}(V)$, 则 $\operatorname{tr} T = \operatorname{tr} \mathcal{M}(T)$.

证明 根据前述定理, 我们只需证明对某个基有 $\operatorname{tr} T = \operatorname{tr} \mathcal{M}(T)$. 若 V 是复线性空间, 则取 T 的若尔当基即得结论. 若 V 是实线性空间, 则把复的情况用于 $T_{\mathbf{C}}$ 即可. □

利用上述定理, 我们不必求出 $T \in \mathcal{L}(V)$ 的任何特征值就能算出它的迹.

例 1 考虑 $\mathbf{C}^5$ 上的一个算子, 它关于某个基的矩阵为

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

我们不知道这个算子的任何特征值的精确公式, 但我们知道它的特征值之和为 0, 因为它的矩阵的对角线元素之和为 0.

通过转换成矩阵的迹的语言, 我们可以给出算子的迹的一些有用性质的简单证明.

性质 (算子的迹是可加的)：设 $S, T \in \mathcal{L}(V)$, 则 $\operatorname{tr}(S+T) = \operatorname{tr} S + \operatorname{tr} T$.

证明 取 V 的一个基, 则

$$\begin{aligned} \operatorname{tr}(S+T) &= \operatorname{tr} \mathcal{M}(S+T) \\ &= \operatorname{tr}(\mathcal{M}(S) + \operatorname{tr} \mathcal{M}(T)) \\ &= \operatorname{tr} \mathcal{M}(S) + \operatorname{tr} \mathcal{M}(T) \\ &= \operatorname{tr} S + \operatorname{tr} T. \end{aligned}$$

其中第三个等号由矩阵的迹的定义直接可得. □

利用前面的结论, 我们可以得到下面这个奇妙的结果. 这个结果在无限维线性空间上的推广可以导出现代物理 (特别是量子理论) 的一些重要结果.

命题 (恒等算子不能表示为 ST 与 TS 之差)：不存在 $S, T \in \mathcal{L}(V)$ 使得

$$ST - TS = I.$$

证明 取 V 的一个基, 则

$$\begin{aligned} \operatorname{tr}(ST - TS) &= \operatorname{tr}(ST) - \operatorname{tr}(TS) \\ &= \operatorname{tr} \mathcal{M}(ST) - \operatorname{tr} \mathcal{M}(TS) \\ &= \operatorname{tr}(\mathcal{M}(S)\mathcal{M}(T)) - \operatorname{tr}(\mathcal{M}(T)\mathcal{M}(S)) \\ &= 0. \end{aligned}$$

其中第一个等号利用了算子迹的可加性, 最后一个等号利用了 $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. 由于 I 的迹等于 $\dim V$, 不等于 0, 因此 $ST - TS$ 与 I 有不同的迹, 它们不相等. □

11.3 行列式

现在可以定义算子的行列式了. 注意下面的定义仿照算子迹的定义方式, 用特征值之积代替了特征值之和.

定义 (算子的行列式): 设 $T \in \mathcal{L}(V)$, 它的行列式 (determinant) $\det T$ 按如下方式定义:
(1) 若 $\mathbf{F} = \mathbf{C}$, 则定义 T 的行列式为 T 的全体特征值之积 (按重数重复).
(2) 若 $\mathbf{F} = \mathbf{R}$, 则定义 T 的行列式为 T_{C} 的全体特征值之积 (按重数重复).

行列式和特征多项式联系紧密. 设 $\lambda_1, \cdots, \lambda_n$ 为 T 的特征值 (若为实线性空间, 则为 T_{C} 的特征值), 其中每个特征值按重数重复. 由定义, T 的特征多项式等于

$$(z-\lambda_1)\cdots(z-\lambda_n).$$

将上式展开, 则 T 的特征多项式可写为

$$z^n-(\lambda_1+\cdots+\lambda_n)z^{n-1}+\cdots+(-1)^n\lambda_1\cdots\lambda_n.$$

由此即得下面的两个结论.

命题 (行列式与特征多项式): 设 $T \in \mathcal{L}(V)$, $\dim V = n$, 则 $\det T$ 等于 $(-1)^n$ 乘以 T 的特征多项式的常数项.

命题 (特征多项式与迹、行列式): 设 $T \in \mathcal{L}(V)$, 则 T 的特征多项式可写为

$$z^n-(\operatorname{tr} T)z^{n-1}+\cdots+(-1)^n(\det T).$$

根据我们的定义, 下面的结论就有了一个简单的证明.

命题 (可逆当且仅当行列式非零): 设 $T \in \mathcal{L}(V)$, 则 T 可逆当且仅当 $\det T \neq 0$.

证明 首先设 V 是复线性空间, 则 T 可逆当且仅当 T 没有特征值 0, 当且仅当特征值的乘积不为 0.

再设 V 为实线性空间, 则 T 可逆当且仅当 T 没有特征值 0, 当且仅当 T_{C} 没有特征值 0(因为 T 与 T_{C} 有相同的实特征值). 因此我们再次得到 $T \in \mathcal{L}(V)$, 则 T 可逆当且仅当 $\det T \neq 0$. □

有些教材把以下定理作为特征多项式的定义, 而把我们的特征多项式定义作为结果.

定理 (特征多项式的另一种刻画): 设 $T \in \mathcal{L}(V)$, 则 T 的特征多项式等于 $\det(zI-T)$.

在上一节处理迹时, 我们发现公式 "迹 = 对角线元素之和" 对任意基都成立, 那么行列式也是如此吗? 也就是说, 算子的行列式等于算子关于任意基的矩阵的对角线元素之积吗?

遗憾的是, 行列式要比迹复杂得多, T 的行列式未必等于 T 关于任意基的矩阵 $\mathcal{M}(T)$ 的对角线元素之积.

下面的结论是行列式的一个重要性质, 它使得我们能够把算子的行列式与它的矩阵的行列式联系起来. 注意, 这个证明比关于迹的相应结果的证明复杂得多.

命题 (行列式是可乘的): 设 A 与 B 是大小相同的方阵, 则

$$\det(AB) = \det A \cdot \det B = \det(BA).$$

现在我们可以证明, 算子的矩阵的行列式与计算这个矩阵所使用的基无关.

定理 (行列式与基的选取无关): 设 $T \in \mathcal{L}(V)$, $u_1, \cdots, u_n$ 与 $v_1, \cdots, v_n$ 是 V 的两个基, 则

$$\det \mathcal{M}(T, (u_1, \cdots, u_n)) = \det \mathcal{M}(T, (v_1, \cdots, v_n)).$$

证明 设 $A = \mathcal{M}(I, (u_1, \cdots, u_n), (v_1, \cdots, v_n))$, 则

$$\begin{aligned}
\det \mathcal{M}(T, (u_1, \cdots, u_n)) &= \det\left(A^{-1}\left(\mathcal{M}(T, (v_1, \cdots, v_n))A\right)\right) \\
&= \det\left(\left(\mathcal{M}(T, (v_1, \cdots, v_n))A\right)A^{-1}\right) \\
&= \det \mathcal{M}(T, (v_1, \cdots, v_n)).
\end{aligned}$$

其中第一个等号利用了基变更公式, 第二个等号利用了 $\det(CD) = \det(DC)$. □

以下定理表明算子的行列式等于算子的矩阵的行列式. 这个定理没有指明用到的基, 因为根据上面的结果, 对每个基来说, 算子的矩阵的行列式都是一样的.

定理 (算子的行列式等于它的矩阵的行列式): 若 $T \in \mathcal{L}(V)$, 则 $\det T = \det \mathcal{M}(T)$.

证明 根据前述定理, 我们只需证明对某个基有 $\det T = \det \mathcal{M}(T)$. 若 V 是复线性空间, 则取 T 的若尔当基即得结论. 若 V 是实线性空间, 则把复的情况用于 $T_{\mathbf{C}}$ 即可. □

通过转换成矩阵行列式的语言, 我们可以给出算子行列式的一些有用性质的简单证明.

性质 (算子的行列式是可乘的): 设 $S, T \in \mathcal{L}(V)$, 则 $\det(ST) = (\det S)(\det T) = \det(TS)$.

证明 取 V 的一个基, 则

$$\begin{aligned}\det(ST) &= \det \mathcal{M}(ST)\\ &= \det(\mathcal{M}(S)\mathcal{M}(T))\\ &= \det \mathcal{M}(S) \cdot \det \mathcal{M}(T)\\ &= (\det S)(\det T).\end{aligned}$$

同理可得 $\det(TS) = (\det T)(\det S)$. 由 $\mathbf{F}$ 的交换性可知结论成立. □

11.4 行列式符号的解释

我们在最后一章之前就已经证明了线性代数的基本结果. 虽然行列式对于更高等的课题是有价值的研究工具, 但它在基础线性代数中并未发挥多少作用 (当该课题得到适当处理时) .

行列式在大学数学中确实有一个重要的应用, 即用于计算某些体积和积分. 这一小节我们来解释实线性空间上行列式的符号的含义. 之后在最后一小节, 我们将利用所学习的线性代数知识来弄清楚行列式和这些应用之间的联系 (因此我们将会利用线性代数处理分析中的一些内容).

回忆一下, 内积空间中的等距同构是保持范数的算子. 以下命题表明, 每个等距同构的行列式的绝对值都等于 1.

命题 (等距同构行列式的模为 1): 设 V 是内积空间, $S \in \mathcal{L}(V)$ 为等距同构, 则 $|\det S| = 1$.

回忆一下, T^*T 是半正定的, 它有唯一的半正定平方根 $\sqrt{T^*T}$. 因为 $\sqrt{T^*T}$ 是半正定的, 故它的特征值都是非负实数, 从而 $\det \sqrt{T^*T} \geqslant 0$. 这在下面的例子中起了重要作用.

例 1 设 V 是实内积空间, $T \in \mathcal{L}(V)$ 是可逆的 (因此 $\det T$ 是正的或者负的, 不为 0). 找出 $\det T$ 的符号的一个几何解释.

解 首先考虑等距同构 $S \in \mathcal{L}(V)$. 由上述命题, $\det S$ 等于 1 或 -1. 注意到

$$\{v \in V : Sv = -v\}$$

是特征子空间 $E(-1, S)$. 从几何的角度考虑, 我们可以称这个子空间为 S 的反向子空间. 由实内积空间上等距同构的刻画我们知道, 当这个子空间的维数是偶数的时候 $\det S = 1$, 当这个子空间的维数是奇数的时候 $\det S = -1$.

下面回到任意可逆算子 $T \in \mathcal{L}(V)$. 由极分解可知存在 $S \in \mathcal{L}(V)$ 使得 $T = S\sqrt{T^*T}$. 从而

$$\det T = (\det S)(\det \sqrt{T^*T}).$$

注意到 $\det \sqrt{T^*T}$ 是非负的, 因此 $\det T$ 是正的还是负的取决于 $\det S$ 是正的还是负的. 由前一段的讨论可知这取决于 S 的反向子空间 $E(-1, S)$ 是偶数维还是奇数维的.

注：我们没有给出“反向”的正式定义, 因为它只是作为一种直观来帮助我们理解而已.

11.5 体积

以下命题是研究体积的一个关键工具. 回忆一下, 根据上一节的讨论, 我们知道 $\det\sqrt{T^*T} \geqslant 0$.

命题 (T **与** $\sqrt{T^*T}$ **的行列式**)：设 V 是内积空间, $T \in \mathcal{L}(V)$, 则 $|\det T| = \det\sqrt{T^*T}$.

证明 由极分解定理与等距同构行列式的绝对值为 1 即得. □

现在转向 $\mathbf{R}^n$ 中的体积问题. 在这一小节的剩余部分, 我们取定一个正整数 n, 并且只考虑带有标准内积的实内积空间 $\mathbf{R}^n$.

我们将仅使用直观的体积概念, 因为我们的目的是理解线性代数, 而体积的概念属于分析学 (不过很快就会看到体积与行列式的紧密联系). 因此, 本节其余部分将依赖直观的体积概念, 而不依赖其严格的发展, 但在接下来的线性代数部分我们还是要保持一贯的严密. 如果适当解释, 这里关于体积所说的一切都是正确的: 这里使用的直观方法都可以运用分析学的手段转化成恰当的正确定义、正确陈述和正确证明.

对于 $\mathbf{R}^n$ 的子集 Ω, 我们用 $\operatorname{vol}\Omega$ 表示它的体积. 对于在 Ω 上有定义的函数 T, 我们用记号 $T(\Omega)$ 表示集合 $\{Tx : x \in \Omega\}$. 即是说

$$T(\Omega) = \{Tx : x \in \Omega\}.$$

对于 $T \in \mathcal{L}(\mathbf{R}^n)$ 与 $\Omega \subseteq \mathbf{R}^n$, 我们要利用 T 与 $\operatorname{vol}\Omega$ 给出 $\operatorname{vol}T(\Omega)$ 的公式. 先来看看半正定算子. 注意到半正定算子的行列式一定是非负的, 即下面的 $\det T \geqslant 0$.

命题 (半正定算子 T **使得体积缩胀** $\det T$ **倍)**：设 $T \in \mathcal{L}(\mathbf{R}^n)$ 是半正定的, $\Omega \subseteq \mathbf{R}^n$, 则

$$\operatorname{vol}T(\Omega) = (\det T)(\operatorname{vol}\Omega).$$

下一个工具是以下命题, 它说的是等距同构不改变体积.

命题 (等距同构不改变体积)：设 $S \in \mathcal{L}(\mathbf{R}^n)$ 为等距同构, $\Omega \subseteq \mathbf{R}^n$, 则

$$\operatorname{vol}S(\Omega) = \operatorname{vol}\Omega.$$

现在我们可以证明算子 $T \in \mathcal{L}(\mathbf{R}^n)$ 使得体积改变了 $|\det T|$ 倍了. 注意极分解定理对证明极为重要.

定理（T 使得体积缩胀 $|\det T|$ 倍）：设 $T \in \mathcal{L}(\mathbf{R}^n)$, $\Omega \subseteq \mathbf{R}^n$, 则

$$\operatorname{vol} T(\Omega) = |\det T|(\operatorname{vol} \Omega).$$

证明 由极分解定理知存在等距同构 $S \in \mathcal{L}(\mathbf{R}^n)$, 使得 $T = S\sqrt{T^*T}$, 从而

$$\begin{aligned}
\operatorname{vol} T(\Omega) &= \operatorname{vol} S\sqrt{T^*T}(\Omega) \\
&= \operatorname{vol} \sqrt{T^*T}(\Omega) \\
&= (\det \sqrt{T^*T})(\operatorname{vol} \Omega) \\
&= |\det T|(\operatorname{vol} \Omega).
\end{aligned}$$

其中第二个等号利用了等距同构不改变体积, 第三个等号利用了半正定算子使体积缩胀行列式倍, 最后一个等号利用了本节第一个命题. □

上述定理导致了行列式出现在重积分的变量替换中. 相关的讨论我们在下一节展开.

例 1 设 a, b, c 为正数. 找出一个已知体积的集合 $\Omega \subseteq \mathbf{R}^3$ 和一个算子 $T \in \mathcal{L}(\mathbf{R}^3)$, 使得 $T(\Omega)$ 等于椭球

$$\left\{(x, y, z) \in \mathbf{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} < 1\right\}.$$

并求该椭球的体积.

解 记上述椭球为 E. 令 Ω 为单位球

$$\Omega = \{(x, y, z) \in \mathbf{R}^3 : x^2 + y^2 + z^2 < 1\}.$$

定义算子 $T \in \mathcal{L}(\mathbf{R}^3)$ 为

$$T(x, y, z) = (ax, by, cz).$$

请自行验证 $E = T(\Omega)$. 易知 $\det T = abc$ (取标准基即可)，故

$$\operatorname{vol} E = \operatorname{vol} T(\Omega) = |\det T|(\operatorname{vol} \Omega) = \frac{4}{3}\pi abc.$$

11.6 重积分的变量替换公式

现在我们定义可微和导数的概念. 注意, 在这个语境下, 导数是一个算子, 而不像一元微积分中那样是一个数. 以下定义中 T 的唯一性证明留作习题.

定义 (可微与导数)：设 Ω 是 $\mathbf{R}^n$ 的开子集，函数 $f:\Omega\to\mathbf{R}^n$, $x\in\Omega$. 若存在算子 $T\in\mathcal{L}(\mathbf{R}^n)$ 使得

$$\lim_{h\to 0}\frac{\|f(x+h)-f(x)-Th\|}{\|h\|}=0,$$

则称 f 在 x 处是可微的 (differentiable), 且称满足上式的唯一算子 $T\in\mathcal{L}(\mathbf{R}^n)$ 为 f 在 x 处的导数 (derivative), 记作 $f'(x)$.

导数的思想是: 对固定的 x 和很小的 $\|h\|$, 有

$$f(x+h)\approx f(x)+f'(x)h.$$

设 Ω 是 $\mathbf{R}^n$ 的开子集, 函数 $f:\Omega\to\mathbf{R}^n$. 我们可以记

$$f(x)=(f_1(x),\cdots,f_n(x)).$$

其中每个 f_i 都是从 Ω 到 $\mathbf{R}$ 的函数. f_i 对第 j 个坐标的偏导数记作 D_jf_i, 它在 $x\in\Omega$ 处的值为 $D_jf_i(x)$. 若 f 在 x 处可微, 那么 $f'(x)$ 关于 $\mathbf{R}^n$ 的标准基的矩阵的第 i 行第 j 列的元素为 $D_jf_i(x)$, 也就是说

$$\mathcal{M}(f'(x))=\begin{pmatrix}D_1f_1(x) & \cdots & D_nf_1(x)\\ \vdots & & \vdots\\ D_1f_n(x) & \cdots & D_nf_n(x)\end{pmatrix}.$$

现在可以给出变量替换公式了. 之所以这么称呼它, 是因为我们可以把 $y=f(x)$ 看成一个变量替换, 就像后面的两个例子所阐述的那样. 需要注意的是, 下面给出的证明并不是一个严格的数学证明, 它只是为了传递结果正确的原因.

定理 (重积分的变量替换)：设 Ω 是 $\mathbf{R}^n$ 的开子集，函数 $\sigma:\Omega\to\mathbf{R}^n$ 在 Ω 上可微, f 是定义在 $\sigma(\Omega)$ 上的实值函数，则

$$\int_{\sigma(\Omega)}f(y)\,\mathrm{d}y=\int_{\Omega}f(\sigma(x))|\det\sigma'(x)|\,\mathrm{d}x.$$

证明 设 $x\in\Omega$, Γ 是包含 x 的小子集, 使得在集合 $\sigma(\Gamma)$ 上 f 约等于常数 $f(\sigma(x))$.

把一个集合中的每个向量都加上一个固定的向量, 体积不变. 利用导数可以给出 σ 在 x 附近的近似, 由此可得

$$\begin{aligned}\operatorname{vol}\sigma(\Gamma)&\approx\operatorname{vol}[(\sigma'(x)(\Gamma)]\\&\approx|\det\sigma'(x)|(\operatorname{vol}\Gamma).\end{aligned}$$

这里第二个约等于利用了 $\sigma'(x)$ 是线性变换 (注意我们并没有要求 f 与 σ 是线性的). 设 $y=\sigma(x)$, 上式左边乘以 $f(y)$, 右边乘以 $f(\sigma(x))$(因为 $y=\sigma(x)$, 故两个量相等), 得到

$$f(y)\mathrm{vol}\,\sigma(\Gamma)\approx f(\sigma(x))|\det\sigma'(x)|(\mathrm{vol}\,\Gamma).$$

把 Ω 分成许多小块, 并把上式对应于个小块的等式相加, 即得所求. □

作变量替换需要注意的是, 在作替换 $y=\sigma(x)$ 时, 一定会包含因子 $|\det\sigma'(x)|$. 最后通过两个重要的例子来说明这一点.

例 1 (**极坐标**) 定义 $\sigma:\mathbf{R}^2\to\mathbf{R}^2$ 为

$$\sigma(r,\theta)=(r\cos\theta,r\sin\theta).$$

这里使用 r,θ 而不是 x_1,x_2 作为坐标, 对熟悉极坐标的每个人来说都很显然. 请自行验证 σ' 在标准基下的矩阵为

$$\begin{pmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{pmatrix}.$$

上面这个矩阵的行列式等于 r, 这解释了在利用极坐标计算积分时为什么会有一个因子 r. 例如下式是 f 在 $\mathbf{R}^2$ 的一个圆盘上的积分, 注意那个额外的因子 r:

$$\int_{-1}^{1}\left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}f(x,y)\,\mathrm{d}y\right)\mathrm{d}x=\int_0^{2\pi}\left(\int_0^1 f(r\cos\theta,r\sin\theta)r\,\mathrm{d}r\right)\mathrm{d}\theta.$$

例 2 (**球坐标**) 定义 $\sigma:\mathbf{R}^3\to\mathbf{R}^3$ 为

$$\sigma(\rho,\varphi,\theta)=(\rho\sin\varphi\cos\theta,\rho\sin\varphi\sin\theta,\rho\cos\varphi),$$

这里使用 ρ,θ,φ 而不是 x_1,x_2,x_3 作为坐标, 对熟悉球坐标的每个人来说都很显然. 请自行验证, 对于 σ 的导数 σ', 相应于标准基的矩阵为

$$\begin{pmatrix}\sin\varphi\cos\theta & \rho\cos\varphi\cos\theta & -\rho\sin\varphi\sin\theta\\ \sin\varphi\sin\theta & \rho\cos\varphi\sin\theta & \rho\sin\varphi\cos\theta\\ \cos\varphi & -\rho\sin\varphi & 0\end{pmatrix}$$

上述矩阵的行列式等于 $\rho^2\sin\varphi$, 这解释了在利用球坐标计算积分时为什么会有一个因子 $\rho^2\sin\varphi$.

例如, 下式是函数 f 在 $\mathbf{R}^3$ 的一个球上的积分, 注意那个额外的因子 $\rho^2\sin\varphi$:

$$\begin{aligned}&\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} f(x,y,z)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x\\ =&\int_0^{2\pi}\int_0^{\pi}\int_0^1 f(\rho\sin\varphi\cos\theta,\rho\sin\varphi\sin\theta,\rho\cos\varphi)\rho^2\sin\varphi\,\mathrm{d}\rho\,\mathrm{d}\varphi\,\mathrm{d}\theta.\end{aligned}$$

11.7 习题

1. 设 A 是分块上三角矩阵

$$A=\begin{pmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

对角线上的每个 A_j 都是方阵. 证明 $\det A=(\det A_1)\cdots(\det A_m)$.