线性变换

黄利兵

数学科学学院

2023 年 3 月 13 日

主要内容

导言

- 线性变换理论是线性代数的重要基础理论之一.
- 线性变换虽然是抽象概念, 但它是用几何语言描述的, 又可以具体化为矩阵, 从而在某种意义上给矩阵提供了几何直观.
- 研究一个具体的线性变换时, 可以考虑整个空间是否可以分解为若干个不变子空间的直和, 从而将高维问题变成低维问题来考虑.

回顾: 线性映射

设 V, W 是数域 P 上的两个线性空间. 如果映射 $f\colon V \to W$ 满足

$$\begin{aligned} f(\alpha+\beta) &= f(\alpha)+f(\beta), \\ f(k\alpha) &= kf(\alpha), \quad \forall \alpha,\beta \in V, \quad k \in P, \end{aligned}$$

则称 f 是线性空间 V, W 之间的线性映射.

- 与线性映射有关的重要子空间
  - 核: $0 \in W$ 的原像, 记作 $f^{-1}(0)$ 或 $\ker f$;
  - 值域或像: f 的值域, 记作 $f(V)$ 或 $\operatorname{im} f$.
- 维数公式

$$\dim V = \dim \ker f + \dim \operatorname{im} f.$$

线性变换的定义

线性变换就是从一个线性空间 V 到其自身的线性映射.

定义

设 V 是数域 P 上的线性空间. 若映射 $\mathcal{A}: V \to V$ 满足

$$\mathcal{A}(\alpha+\beta) = \mathcal{A}(\alpha) + \mathcal{A}(\beta), \quad \mathcal{A}(k\alpha) = k\mathcal{A}(\alpha),$$

对任意 $\alpha, \beta \in V$ 和 $k \in P$ 成立, 则称 $\mathcal{A}$ 为 V 上的 **线性变换**.

注

- 定义中的两个条件也可等价地写为一个条件

$$\begin{aligned} \mathcal{A}(k\alpha + l\beta) =& k\mathcal{A}(\alpha) + l\mathcal{A}(\beta), \\ & \forall \alpha, \beta \in V, \quad k, l \in P. \end{aligned}$$

- 由于线性变换是线性映射的特殊情形, 所以每个线性变换也都有自己的核与像. 特别地, 核空间 $\ker \mathcal{A}$ 的维数称为 $\mathcal{A}$ 的零度, 记作 $\mathrm{N}(\mathcal{A})$ 或 $\mathrm{null}(\mathcal{A})$; 像空间的维数 $\dim \mathcal{A}(V)$ 称为 $\mathcal{A}$ 的秩, 记作 $R(\mathcal{A})$ 或 $\mathrm{rank}(\mathcal{A})$. 零度与秩的和等于空间的维数.

线性变换的性质

命题

设 $\mathcal{A}$ 是线性空间 V 上的线性变换, 则有

- $\mathcal{A}(0) = 0$, $\mathcal{A}(-\alpha) = -\mathcal{A}(\alpha)$.
- 若 $\alpha = k_1\alpha_1 + \cdots + k_s\alpha_s$, 则 $\mathcal{A}(\alpha) = k_1\mathcal{A}(\alpha_1) + \cdots + k_s\mathcal{A}(\alpha_s)$.
- 若 $\alpha_1, \cdots, \alpha_s$ 线性相关, 则 $\mathcal{A}(\alpha_1),\cdots, \mathcal{A}(\alpha_s)$ 也线性相关.

证明.

只有中间这条需要证明. 事实上, 反复利用线性变换的定义就可得到

$$
\begin{aligned}
&\mathcal{A}(k_1\alpha_1 + \cdots + k_s\alpha_s) \\
=&\mathcal{A}(k_1\alpha_1 + \cdots + k_{s-1}\alpha_{s-1}) + k_s\mathcal{A}(\alpha_s) \\
=&\mathcal{A}(k_1\alpha_1 + \cdots + k_{s-2}\alpha_{s-2}) \\
&\quad + k_{s-1}\mathcal{A}(\alpha_{s-1}) + k_s\mathcal{A}(\alpha_s) \\
=&\cdots \\
=&k_1\mathcal{A}(\alpha_1) + \cdots + k_s\mathcal{A}(\alpha_s).
\end{aligned}
$$

□

线性变换的例子 (一)

例

设 $V = P^{n\times 1}$. 固定一个矩阵 $A \in P^{n\times n}$, 则可定义映射 $\mathcal{A}: V \to V$ 如下

$$\mathcal{A}(\alpha) = A\alpha, \quad \forall \alpha \in P^{n\times 1} = V.$$

由于 $\mathcal{A}(k\alpha + l\beta) = A(k\alpha + l\beta) = kA\alpha + lA\beta = k\mathcal{A}(\alpha) + l\mathcal{A}(\beta)$, 易知 $\mathcal{A}$ 是线性变换.

例

设 $V = C^{\infty}(0,1)$, 则求导是线性变换, 因为

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\big(u(x) + v(x)\big) &= \frac{\mathrm{d}}{\mathrm{d}x}u(x) + \frac{\mathrm{d}}{\mathrm{d}x}v(x), \\ \frac{\mathrm{d}}{\mathrm{d}x}\big(cu(x)\big) &= c \cdot \frac{\mathrm{d}}{\mathrm{d}x}u(x). \end{aligned}$$

类似地, 在线性空间 $C^{\infty}[0,1]$ 上, 由 $u(x) \mapsto \int_0^x u(t)\,\mathrm{d}t$ 所定义的映射 $\mathcal{S}$ 也是线性变换.

线性变换的例子 (二)

例

设 $V = P[x]$, $a \in P$, 定义映射 $L_a : P[x] \to P[x]$ 如下

$$L_a(f(x)) = f(x+a),$$

则 L_a 是线性变换.

例 (旋转)

考虑平面上的所有几何向量所构成的线性空间 $V = \mathbb{R}^2$. 将每个向量旋转 θ 角, 这个运动过程描述了平面上的一个变换, 记作 I_θ. 向量 $(r\cos t, r\sin t)$ 在这个变换下被变为 $(r\cos(t+\theta), r\sin(t+\theta))$. 把坐标写为列向量, 则有

$$\begin{bmatrix} r\cos(t+\theta) \\ r\sin(t+\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r\cos t \\ r\sin t \end{bmatrix}.$$

因此, 向量 $\begin{bmatrix} x \\ y \end{bmatrix}$ 被变为 $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$. 由第一个例子可知它是线性变换.

线性变换的例子 (三)

例 (切变 (shear))

给定实数 k. 在平面直角坐标系中, 将每一点 (x, y) 沿着与 x 轴平行的方向平移 ky 个单位, 到达点 $(x + ky, y)$. 这个变换也可写为矩阵形式 $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$. 由第一个例子可知这个变换是线性变换.

例 (投影)

在空间直角坐标系中, 考虑这样的变换, 将点 (x, y, z) 投影到 xOy 平面, 得到 $(x, y, 0)$. 这个变换可写为矩阵形式 $(x, y, z)^{\mathsf{T}} \mapsto \mathrm{diag}(1, 1, 0)(x, y, z)^{\mathsf{T}}$. 由第一个例子可知这个变换是线性变换.

定义

将任一向量变为 0 的变换, 称为零变换, 记作 0. 将任一向量变为它自身的变换, 称为恒等变换, 记作 id. 一般地, 给定 $k \in P$, 变换 $\alpha \mapsto k\alpha$ 称为数乘变换, 记作 k 或 $k \cdot \mathrm{id}$.

线性变换的相等

两个线性变换相等, 是指它们的作用效果相同.

定义

设 $\mathcal{A}$, $\mathcal{B}$ 是线性空间 V 上的线性变换. 如果 $\mathcal{A}(\alpha) = \mathcal{B}(\alpha)$ 对任意向量 $\alpha \in V$ 成立, 则称 $\mathcal{A}$ 与 $\mathcal{B}$ 相等, 记作 $\mathcal{A} = \mathcal{B}$.

例

在平面上, 旋转 180° 与数乘 -1(即位似比为 -1 的位似变换) 这两个线性变换是相等的.

例

在 $\mathbb{R}[x]$ 中, 定义如下两个线性变换 $\mathcal{A}$ 与 $\mathcal{B}$

$$\begin{aligned} \mathcal{A}(u(x)) &= u(x) - u(0), \\ \mathcal{B}(u(x)) &= \int_0^x u'(t)\,\mathrm{d}t. \end{aligned}$$

由微积分基本定理可知 $\mathcal{A} = \mathcal{B}$.

线性变换的加法

定义

设 $\mathcal{A}, \mathcal{B}$ 是线性空间 V 上的线性变换, 它们的和 $\mathcal{A}+\mathcal{B}: V \to V$ 定义为

$$(\mathcal{A}+\mathcal{B})(\alpha) = \mathcal{A}(\alpha) + \mathcal{B}(\alpha), \quad \forall \alpha \in V.$$

注意

$$\begin{aligned} & (\mathcal{A}+\mathcal{B})(k\alpha + l\beta) \\ =& \mathcal{A}(k\alpha + l\beta) + \mathcal{B}(k\alpha + l\beta) \\ =& k\mathcal{A}(\alpha) + l\mathcal{A}(\beta) + k\mathcal{B}(\alpha) + l\mathcal{B}(\beta) \\ =& k(\mathcal{A}+\mathcal{B})(\alpha) + l(\mathcal{A}+\mathcal{B})(\beta), \end{aligned}$$

可见 $\mathcal{A}+\mathcal{B}$ 仍是 V 上的线性变换.

例

在线性空间 $C^{\infty}(0,1)$ 中, 求导与恒等变换的和, 是这样的变换

$$u(x) \mapsto u'(x) + u(x).$$

线性变换加法的性质

由定义容易验证, 线性变换的加法满足交换律和结合律:

$$
\begin{aligned}
(\mathcal{A}+\mathcal{B})(\alpha) &= \mathcal{A}(\alpha)+\mathcal{B}(\alpha) \\
&= \mathcal{B}(\alpha)+\mathcal{A}(\alpha) = (\mathcal{B}+\mathcal{A})(\alpha) \\
\Longrightarrow &\quad \mathcal{A}+\mathcal{B}=\mathcal{B}+\mathcal{A}, \\
(\mathcal{A}+(\mathcal{B}+\mathcal{C}))(\alpha) &= \mathcal{A}(\alpha)+(\mathcal{B}+\mathcal{C})(\alpha) \\
&= \mathcal{A}(\alpha)+\mathcal{B}(\alpha)+\mathcal{C}(\alpha) \\
&= (\mathcal{A}+\mathcal{B})(\alpha)+\mathcal{C}(\alpha) \\
&= ((\mathcal{A}+\mathcal{B})+\mathcal{C})(\alpha) \\
\Longrightarrow &\quad \mathcal{A}+(\mathcal{B}+\mathcal{C})=(\mathcal{A}+\mathcal{B})+\mathcal{C}.
\end{aligned}
$$

并且, 任意线性变换与零变换的和仍等于自身:

$$
\begin{aligned}
(0+\mathcal{A})(\alpha) &= 0(\alpha)+\mathcal{A}(\alpha)=\mathcal{A}(\alpha) \\
\Longrightarrow &\quad 0+\mathcal{A}=\mathcal{A}.
\end{aligned}
$$

线性变换的数量乘法

定义

设 $\mathcal{A}$ 是线性空间 V 上的线性变换, $k \in P$, 则 k 与 $\mathcal{A}$ 的数量乘法 $k\mathcal{A}: V \to V$ 定义为

$$(k\mathcal{A})(\alpha) = k \cdot \mathcal{A}(\alpha), \quad \forall \alpha \in V.$$

注意

$$\begin{aligned}
(k\mathcal{A})(\alpha+\beta) &= k \cdot \mathcal{A}(\alpha+\beta) \\
&= k\mathcal{A}(\alpha) + k\mathcal{A}(\beta) \\
&= (k\mathcal{A})(\alpha) + (k\mathcal{A})(\beta), \\
(k\mathcal{A})(l\alpha) &= k\mathcal{A}(l\alpha) \\
&= kl\mathcal{A}(\alpha) \\
&= l \cdot (k\mathcal{A})(\alpha),
\end{aligned}$$

可见 $k\mathcal{A}$ 仍是 V 上的线性变换.

线性变换的数乘的性质

由定义容易验证, 数量乘法满足结合律和两条分配律:

$$
\begin{aligned}
&((kl)\mathcal{A})(\alpha) = (kl)\cdot\mathcal{A}(\alpha)\\
&\quad = k\cdot l\cdot\mathcal{A}(\alpha) = (k(l\mathcal{A}))(\alpha)\\
&\Longrightarrow \quad (kl)\mathcal{A} = k(l\mathcal{A}),\\
&(k(\mathcal{A}+\mathcal{B}))(\alpha) = k\cdot((\mathcal{A}+\mathcal{B})(\alpha))\\
&\quad = k\cdot(\mathcal{A}(\alpha)+\mathcal{B}(\alpha))\\
&\quad = (k\mathcal{A})(\alpha)+(k\mathcal{B})(\alpha)\\
&\Longrightarrow \quad k(\mathcal{A}+\mathcal{B}) = k\mathcal{A}+k\mathcal{B},\\
&((k+l)\mathcal{A})(\alpha) = (k+l)(\mathcal{A}(\alpha))\\
&\quad = k\mathcal{A}(\alpha)+l\mathcal{A}(\alpha)\\
&\quad = (k\mathcal{A}+l\mathcal{A})(\alpha)\\
&\Longrightarrow \quad (k+l)\mathcal{A} = k\mathcal{A}+l\mathcal{A}.
\end{aligned}
$$

线性变换的全体

命题

将线性空间 V 上的所有线性变换构成的集合记作 $\mathrm{End}\,V$. 那么, $\mathrm{End}\,V$ 关于上述加法和数量乘法构成数域 P 上的线性空间.

证明.

前面已经验证了线性变换的加法满足封闭性、交换律、结合律, 数量乘法满足封闭性、结合律和分配律, 为证明 $\mathrm{End}\,V$ 构成线性空间, 只需验证 0 元素、负元素的存在性, 并验证 $1\cdot\mathcal{A}=\mathcal{A}$.

- $\mathrm{End}\,V$ 中的 0 元素就是零变换, 因为 $0+\mathcal{A}=\mathcal{A}$;
- $\mathcal{A}$ 的负元素就是 $(-1)\cdot\mathcal{A}$, 因为 $(\mathcal{A}+(-1)\cdot\mathcal{A})(\alpha)=\mathcal{A}(\alpha)+(-1)\cdot\mathcal{A}(\alpha)=0$, 表明 $\mathcal{A}+(-1)\cdot\mathcal{A}=0$;
- $1\cdot\mathcal{A}=\mathcal{A}$, 是因为 $(1\cdot\mathcal{A})(\alpha)=\mathcal{A}(\alpha)$, $\forall\alpha\in V$.

$\square$

线性变换的乘法

线性变换的乘法就是映射的复合.

定义

设 $\mathcal{A}, \mathcal{B}$ 是线性空间 V 上的线性变换, 它们的乘积 $\mathcal{AB}: V \to V$ 定义为

$$(\mathcal{AB})(\alpha) = \mathcal{A}(\mathcal{B}(\alpha)), \quad \forall \alpha \in V.$$

特别地, 数量乘法 $k\mathcal{A}$ 也可看成数乘变换 k 与线性变换 $\mathcal{A}$ 的乘法.
注意

$$\begin{aligned}(\mathcal{AB})(k\alpha + l\beta) &= \mathcal{A}(\mathcal{B}(k\alpha + l\beta))\\ &= \mathcal{A}(k\mathcal{B}(\alpha) + l\mathcal{B}(\beta))\\ &= k\mathcal{A}(\mathcal{B}(\alpha)) + l\mathcal{A}(\mathcal{B}(\beta))\\ &= k(\mathcal{AB})(\alpha) + l(\mathcal{AB})(\beta),\end{aligned}$$

可见 $\mathcal{AB}$ 仍是 V 上的线性变换.

线性变换乘法的性质

命题

线性变换的乘法具有以下性质

- 结合律: $\mathcal{A}(\mathcal{BC}) = (\mathcal{AB})\mathcal{C}$;
- 对加法的分配律: $\mathcal{A}(\mathcal{B}+\mathcal{C}) = \mathcal{AB}+\mathcal{AC}$;
- 与数乘的交换律: $\mathcal{A}(k\mathcal{B}) = k(\mathcal{AB}) = (k\mathcal{A})\mathcal{B}$.

注意, 乘法交换律一般不成立.

例

在 $R[x]$ 上, 求导 $\mathcal{D}$ 与积分 $\mathcal{S}$ 是两个线性变换, 容易验证, $\mathcal{DS} = \mathrm{id}$, 但 $\mathcal{SD} \neq \mathrm{id}$.

线性变换的乘方

定义

设 $\mathcal{A}$ 是线性空间 V 上的线性变换. 约定 $\mathcal{A}^0 = \mathrm{id}$, 并递推地定义 $\mathcal{A}^{n+1} = \mathcal{A}\mathcal{A}^n$, $n \in \mathbb{N}$. 称 $\mathcal{A}^n$ 为 $\mathcal{A}$ 的 n 次幂.

例

在平面上, 若 $\mathcal{A}$ 是 30° 旋转, 则 $\mathcal{A}^4$ 是 120° 旋转.

例

在 $P[x]_n$ 中, 若 $\mathcal{D}$ 是求导, 则 $\mathcal{D}^n$ 是零变换.

由乘法结合律可知

$$\mathcal{A}^m\mathcal{A}^n = \mathcal{A}^{m+n}, \quad (\mathcal{A}^m)^n = \mathcal{A}^{mn}, \quad m, n \in \mathbb{N}.$$

线性变换的多项式 (一)

定义

设 V 是数域 P 上的线性空间, $\mathcal{A}$ 是 V 上的线性变换. 对于

$$f(x) = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0 \in P[x],$$

称如下的线性变换

$$a_m \mathcal{A}^m + a_{m-1} \mathcal{A}^{m-1} + \cdots + a_1 \mathcal{A} + a_0 \cdot \mathrm{id}$$

为 $\mathcal{A}$ 的一个多项式, 记作 $f(\mathcal{A})$.

例

如果 I_{90° 是平面上的 90° 旋转, $f(x) = x^2 + 1$, 那么

$$f(\mathrm{I}_{90^\circ}) = 0.$$

例

设 $a \in P$ 且 $a \neq 0$. 考虑 $P[x]_n$ 上的线性变换 L_a 如下

$$L_a(u(x)) = u(x+a), \quad \forall u(x) \in P[x]_n.$$

又设 $\mathcal{D}$ 为求导, 即 $\mathcal{D}(u(x)) = u'(x)$. 那么, 由 Taylor 公式可知

$$u(x+a) = \sum_{k=0}^{n-1} \frac{a^k}{k!} \frac{\mathrm{d}^k u(x)}{\mathrm{d}x^k},$$

因而

$$L_a = \sum_{k=0}^{n-1} \frac{a^k}{k!} \mathcal{D}^k.$$

线性变换的多项式 (二)

现在固定一个线性变换 $\mathcal{A}$, 考虑它的不同多项式. 若 $f(x), g(x) \in P[x]$, 则容易验证, $f(\mathcal{A})g(\mathcal{A}) = g(\mathcal{A})f(\mathcal{A})$.

命题

给定线性空间 V 上的线性变换 $\mathcal{A}$, 则从 $P[x]$ 到 $\mathrm{End}\,V$ 的映射 $f(x) \mapsto f(\mathcal{A})$ 是线性映射.

证明.

显然 $f(x) + g(x)$ 被映为 $f(\mathcal{A}) + g(\mathcal{A})$, 且 $kf(x)$ 被映为 $kf(\mathcal{A})$. □

这个线性映射的核是理解线性变换的一个重要工具.

定义

对于线性变换 $\mathcal{A}$, 如果非零多项式 $u(x)$ 满足 $u(\mathcal{A}) = 0$, 则称 $u(x)$ 为 $\mathcal{A}$ 的一个零化多项式. 如果 $\mathcal{A}$ 有零化多项式, 则其中次数最低的首一多项式称为 $\mathcal{A}$ 的最小多项式或最低多项式.

线性变换的多项式 (三)

命题

如果线性变换 $\mathcal{A}$ 有最小多项式 $d(x)$, 则 $\mathcal{A}$ 的每个零化多项式都是 $d(x)$ 的倍式.

证明.

任取零化多项式 $f(x)$, 则 $(f(x), d(x))$ 也是零化多项式, 其次数 $\leq \deg d(x)$, 从而一定有 $(f(x), d(x)) = d(x)$, 即 $d(x) \mid f(x)$. $\square$

例

- 数乘变换 k 的最小多项式是 $x - k$. 特别地, 零变换的最小多项式是 x, 恒等变换的最小多项式是 $x - 1$.
- 考虑空间直角坐标系中将 (x, y, z) 变为 $(x, y, 0)$ 的投影变换 π_z. 容易看到 $\pi_z^2 = \pi_z$, 所以 $x^2 - x$ 是它的零化多项式. 同时 $\pi_z \neq 0$, $\pi_z \neq \mathrm{id}$, 因此它的最小多项式是 $x^2 - x$.
- 在 $P[x]$ 中, 求导 $\mathcal{D}$ 没有零化多项式, 从而没有最小多项式. 事实上, 任取 m 次多项式 $f(x)$, 则有 $f(\mathcal{D}) \neq 0$, 这是因为 $(f(\mathcal{D}))(x^{m+1}) \neq 0$.

可逆线性变换

定义

如果线性空间 V 上的线性变换 $\mathcal{A}$ 是一一映射, 则称 $\mathcal{A}$ 为可逆线性变换.

命题

- 线性变换 $\mathcal{A}$ 是可逆的, 当且仅当它是单射, 也当且仅当它是满射;
- id 是可逆线性变换;
- 若 $\mathcal{A}$ 是可逆线性变换, 则 $\mathcal{A}^{-1}$ 也是可逆线性变换, 且 $(\mathcal{A}^{-1})^{-1} = \mathcal{A}$;
- 若 $\mathcal{A}, \mathcal{B}$ 都是可逆线性变换, 则 $\mathcal{A}\mathcal{B}$ 也是可逆线性变换, 且 $(\mathcal{A}\mathcal{B})^{-1} = \mathcal{B}^{-1}\mathcal{A}^{-1}$.

对于可逆线性变换 $\mathcal{A}$, 还可定义

$$\mathcal{A}^{-n} = (\mathcal{A}^{-1})^n.$$

这样, 可逆线性变换的乘方运算可以扩充到负的指数. 容易验证

$$\begin{aligned} \mathcal{A}^m\mathcal{A}^n &= \mathcal{A}^{m+n}, \\ (\mathcal{A}^m)^n &= \mathcal{A}^{mn}, \quad \forall m, n \in \mathbb{Z}. \end{aligned}$$

线性变换的矩阵 (一)

在有限维线性空间中, 通过取一组基, 可以将抽象的向量具体化为列向量. 我们现在说明, 这也可以将抽象的线性变换具体化为矩阵, 线性变换的运算就对应着矩阵的运算.

定义

设 $\mathcal{A}$ 是有限维线性空间 V 上的线性变换. 取定 V 的一组基 $\alpha_1, \cdots, \alpha_n$, 将向量 $\mathcal{A}(\alpha_i)$ 在这组基下的坐标写为列向量 a_i, $1 \leq i \leq n$, 则这些列向量构成的矩阵 $A = (a_1, \cdots, a_n)$ 称为 $\mathcal{A}$ 在这组基下的矩阵.

注

采用形式的乘法, 我们可将上述定义写为

$$\mathcal{A}(\alpha_1, \cdots, \alpha_n) = (\mathcal{A}\alpha_1, \cdots, \mathcal{A}\alpha_n) = (\alpha_1, \cdots, \alpha_n) \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix},$$

其中 $a_i = (a_{1i}, \cdots, a_{ni})^{\mathsf{T}}$ 为 $\mathcal{A}(\alpha_i)$ 的坐标.

线性变换的矩阵 (二)

显然零变换的矩阵为零矩阵, 数乘变换 k 的矩阵为数量矩阵 kE_n.

例

给定 $A \in P^{n\times n}$, 考虑 $P^{n\times 1}$ 上的线性变换 $v \mapsto Av$. 在 $P^{n\times 1}$ 的基 $\varepsilon_1, \cdots, \varepsilon_n$ 下, $A\varepsilon_i$ 的坐标恰好是 A 的第 i 列, $1 \le i \le n$. 因此, 这个线性变换在该组基下的矩阵就是 A.

例

取 $P[x]_n$ 的一组基 $1, x, \frac{x^2}{2!}, \cdots, \frac{x^{n-1}}{(n-1)!}$. 我们考虑求导变换 $\mathcal{D}$ 在这组基下的矩阵. 由于 $\mathcal{D}(1) = 0$, 且当 $1 \le k \le n-1$ 时 $\mathcal{D}\left(\frac{x^k}{k!}\right) = \frac{x^{k-1}}{(k-1)!}$, 因此 $\mathcal{D}$ 在这组基下的矩阵为

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

线性变换的矩阵 (三)

取定一组基以后, 每个线性变换在这组基下对应着一个矩阵. 那么反过来, 是否每个矩阵也对应着一个线性变换呢?

命题

取定数域 P 上线性空间 V 的一组基 $\alpha_1, \cdots, \alpha_n$, 并取定 $A \in P^{n\times n}$. 那么, 存在唯一的线性变换 $\mathcal{A}$, 使得它在这组基下的矩阵为 A.

证明.

设 A 的列向量分别为 $a_1, \cdots, a_n$. 以 a_i 为坐标的向量记作 γ_i, $1 \le i \le n$.
存在性: 我们定义变换 $\mathcal{A}$ 如下

$$\mathcal{A}(x_1\alpha_1 + \cdots + x_n\alpha_n) = x_1\gamma_1 + \cdots + x_n\gamma_n.$$

容易验证 $\mathcal{A}$ 是线性变换.
唯一性: 如果 $\mathcal{A}$ 在这组基下的矩阵为 $A = (a_1, \cdots, a_n)$, 则 $\mathcal{A}(\alpha_i)$ 是坐标为 a_i 的向量, 从而上面的等式必定成立. $\square$

线性变换与矩阵的对应 (一)

命题

设 $\alpha_1, \cdots, \alpha_n$ 是线性空间 V 的一组基. 线性变换 $\mathcal{A}$ 在这组基下的矩阵为 A, 向量 α 在这组基下的坐标为 x, 那么, $\mathcal{A}(\alpha)$ 在这组基下的坐标为 Ax.

证明.

由于 α 在这组基下的坐标为 $x = (x_1, \cdots, x_n)^{\mathrm{T}}$, 所以

$$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n.$$

我们有 $\mathcal{A}(\alpha) = x_1\mathcal{A}(\alpha_1) + \cdots + x_n\mathcal{A}(\alpha_n)$. 由于 $\mathcal{A}(\alpha_i)$ 的坐标 a_i 为 A 的第 i 列, 所以 $\mathcal{A}(\alpha)$ 的坐标为

$$\begin{aligned} & x_1 a_1 + \cdots + x_n a_n \\ =&(a_1, \cdots, a_n)(x_1, \cdots, x_n)^{\mathrm{T}} = Ax. \end{aligned}$$

□

这个结论告诉我们, 在有限维线性空间中取定一组基以后, 采用坐标方式来描述的话, 每个线性变换都可写为本章第一个例子的形式.

线性变换与矩阵的对应 (二)

命题

设 $\mathcal{A}$, $\mathcal{B}$ 是有限维线性空间 V 上的线性变换, 它们在同一组基下的矩阵分别为 A, B. 那么, 线性变换 $\mathcal{A}+\mathcal{B}$, $k\mathcal{A}$, $\mathcal{A}\mathcal{B}$ 在这组基下的矩阵分别是 $A+B$, kA, AB.

证明.

仅以 $\mathcal{A}\mathcal{B}$ 为例进行证明. 设这组基为 $\alpha_1, \cdots, \alpha_n$, 那么

$$\begin{aligned}\mathcal{A}(\alpha_1, \cdots, \alpha_n) &= (\alpha_1, \cdots, \alpha_n)A,\\ \mathcal{B}(\alpha_1, \cdots, \alpha_n) &= (\alpha_1, \cdots, \alpha_n)B.\end{aligned}$$

这样一来

$$\begin{aligned}\mathcal{A}\mathcal{B}(\alpha_1, \cdots, \alpha_n) =& \mathcal{A}(\alpha_1, \cdots, \alpha_n)B\\ =& (\alpha_1, \cdots, \alpha_n)AB.\end{aligned}$$

可见 $\mathcal{A}\mathcal{B}$ 在这组基下的矩阵为 AB. □

推论

线性变换与矩阵的对应 (三)

证明.

取定 V 的一组基, 考虑如下映射 $\varphi : \mathrm{End}\, V \to P^{n\times n}$, 对每个线性变换 $\mathcal{A}$, $\varphi(\mathcal{A})$ 是 $\mathcal{A}$ 在这组基下的矩阵. 前面的命题表明 φ 是一一映射, 且保持加法与数乘, 因此是线性同构. □

推论

若 V 是有限维线性空间, $\mathcal{A} \in \mathrm{End}\, V$, 则 $\mathcal{A}$ 的最小多项式一定存在.

证明.

设 $\dim V = n$. 由上面的命题可知 $\mathrm{End}\, V$ 的维数为 n^2, 从而 $\mathcal{A}^0, \mathcal{A}^1, \cdots, \mathcal{A}^{n^2}$ 一定线性相关. □

由于线性变换与矩阵之间的这种对应关系, 我们也可讨论矩阵的零化多项式, 最小多项式等.

线性变换与矩阵的对应 (四)

命题

线性空间 V 上的线性变换 $\mathcal{A}$ 在某组基下的矩阵为 A, 则有

- $\mathcal{A}$ 可逆当且仅当 A 可逆, 且这时 $\mathcal{A}^{-1}$ 在同一组基下的矩阵是 A^{-1};
- 对于 $f(x) \in P[x]$, 线性变换 $f(\mathcal{A})$ 在这组基下的矩阵为 $f(A)$.

证明.

- $\mathcal{A}$ 可逆当且仅当 $\mathcal{A}$ 是满射, 即 $\mathcal{A}(V) = V$. 注意 $\mathcal{A}(V)$ 是由 $\mathcal{A}(\alpha_1), \cdots, \mathcal{A}(\alpha_n)$ 张成的, 所以 $\mathcal{A}(V) = V$ 当且仅当这组向量线性无关, 从而当且仅当它们的坐标线性无关, 也即 A 的列向量组线性无关, 当且仅当 A 可逆. 由 $\mathcal{A}(\alpha_1, \cdots, \alpha_n) = (\alpha_1, \cdots, \alpha_n)A$ 即可得到
$$\mathcal{A}^{-1}(\alpha_1, \cdots, \alpha_n) = (\alpha_1, \cdots, \alpha_n)A^{-1}.$$
- 利用线性变换的加法、数乘以及乘法与相应的矩阵加法、数乘以及乘法之间的对应关系, 这个结论是明显的.

□

线性变换与矩阵的对应 (五)

命题

设 $\mathcal{A}$ 是线性空间 V 上的线性变换, 它在两组基 $\alpha_1, \cdots, \alpha_n$ 和 $\beta_1, \cdots, \beta_n$ 下的矩阵分别为 A, B. 又设从基 $\alpha_1, \cdots, \alpha_n$ 到基 $\beta_1, \cdots, \beta_n$ 的过渡矩阵为 T, 则有

$$B = T^{-1}AT.$$

证明.

由于 $\mathcal{A}(\alpha_1, \cdots, \alpha_n) = (\alpha_1, \cdots, \alpha_n)A$, 且 $(\beta_1, \cdots, \beta_n) = (\alpha_1, \cdots, \alpha_n)T$, 我们有

$$\begin{aligned} \mathcal{A}(\beta_1, \cdots, \beta_n) &= \mathcal{A}(\alpha_1, \cdots, \alpha_n)T \\ &=(\alpha_1, \cdots, \alpha_n)AT = (\beta_1, \cdots, \beta_n)T^{-1}AT. \end{aligned}$$

可见 $\mathcal{A}$ 在基 $\beta_1, \cdots, \beta_n$ 下的矩阵为

$$B = T^{-1}AT.$$

□

相似矩阵 (一)

定义

对于矩阵 $A, B \in P^{n\times n}$, 如果存在可逆矩阵 T, 使得 $B = T^{-1}AT$, 则称 A 与 B 相似, 记为 $A \sim B$.

利用这一定义, 我们可以把前面的命题重新叙述为

命题

同一线性变换在不同基下的矩阵是相似的. 反之, 相似的两个矩阵可看作同一线性变换在两组不同基下的矩阵.

容易验证, 相似是矩阵之间的等价关系, 即

命题

矩阵的相似具有

- 反身性: 任意 n 阶矩阵 A 与自身相似;
- 对称性: 若 A 与 B 相似, 则 B 与 A 相似;
- 传递性: 若 A 与 B 相似, B 与 C 相似, 则 A 与 C 相似.

相似矩阵 (二)

证明.

- 由 $A = E_n^{-1} A E_n$ 可知 A 与自身相似;
- 若 A 与 B 相似, $B = T^{-1}AT$, 则也有 $A = TBT^{-1}$, 所以 B 与 A 相似;
- 若 $B = T^{-1}AT$, $C = S^{-1}BS$, 则有 $C = S^{-1}T^{-1}ATS = (TS)^{-1}A(TS)$, 所以 A 与 C 相似.

$\square$

命题

设 $A, B \in P^{n\times n}$. 如果 A 与 B 相似, 则有

- A^n 与 B^n 相似;
- 对任意多项式 $f(x)$, 矩阵 $f(A)$ 与 $f(B)$ 相似;
- $\mathrm{tr}(A) = \mathrm{tr}(B)$;
- $\det(A) = \det(B)$;
- $R(A) = R(B)$.

这里第 (5) 条性质是明显的. 因此我们只证明 (1), (2), (3), (4).

相似矩阵 (三)

证明.

设 $B = T^{-1}AT$, 其中 T 为可逆矩阵.

- 对 n 用数学归纳法证明 $B^n = T^{-1}A^nT$. 要从 n 过渡到 $n+1$, 只需注意到

$$B^{n+1} = BB^n = (T^{-1}AT)(T^{-1}A^nT) = T^{-1}AA^nT = T^{-1}A^{n+1}T.$$

- 设 $f(x) = \sum a_k x^k$, 则

$$f(B) = \sum a_k B^k = \sum a_k T^{-1}A^kT = T^{-1}\left(\sum a_k A^k\right)T = T^{-1}f(A)T.$$

- 利用迹的性质 $\mathrm{tr}(MN) = \mathrm{tr}(NM)$, 我们有

$$\mathrm{tr}(B) = \mathrm{tr}(T^{-1}AT) = \mathrm{tr}(ATT^{-1}) = \mathrm{tr}(A).$$

- $\det(B) = \det(T^{-1}AT) = \det(T^{-1})\det(A)\det(T) = \det(A).$

□

相似矩阵 (四)

由于同一线性变换 $\mathcal{A}$ 在不同基下的矩阵是相似的, 它们有相同的迹和行列式. 这种量不依赖于基的选取, 所以是线性变换本身的不变量, 可称之为 $\mathcal{A}$ 的迹和行列式, 分别记作 $\mathrm{tr}(\mathcal{A})$ 和 $\det(\mathcal{A})$.

例

设 $\varepsilon_1, \varepsilon_2$ 是 V 的一组基, 线性变换 $\mathcal{A}$ 在这组基下的矩阵为 $A = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$. 现在另取一组基 $\eta_1 = \varepsilon_1 - \varepsilon_2$, $\eta_2 = -\varepsilon_1 + 2\varepsilon_2$, 试求 $\mathcal{A}$ 在基 η_1, η_2 下的矩阵 B, 并计算 A^k.

解答

从基 $\varepsilon_1, \varepsilon_2$ 到基 η_1, η_2 的过渡矩阵为 $T = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$. 因此, $\mathcal{A}$ 在基 η_1, η_2 下的矩阵为 $B = T^{-1}AT = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

易知 $B^k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$, 所以 $A^k = TB^kT^{-1} = \begin{bmatrix} 1+k & k \\ -k & 1-k \end{bmatrix}$.

相似矩阵 (五)

这个例子告诉我们, 可以用矩阵相似的性质来简化矩阵的计算.
计算 A^k 的另一种方法是利用零化多项式. 注意 A 有一个零化多项式是 $(x-1)^2$, 而 $x^k = 1 + k(x-1) + (x-1)^2 \cdot u(x)$, 所以

$$A^k = E_2 + k \cdot (A - E_2) = \begin{bmatrix} 1+k & k \\ -k & 1-k \end{bmatrix}.$$

思考题

- (**) 设 $\mathcal{A}$ 是有限维线性空间 V 上的线性变换. 我们知道, $\mathcal{A}$ 的秩定义为 $\mathcal{A}(V)$ 的维数. 同时, $\mathcal{A}$ 在不同基下的矩阵有相同的秩, 它也可作为 $\mathcal{A}$ 的秩的另一种定义. 那么, 这两个定义是否一致?
- (****) 是否存在 $A \in \mathbb{Q}^{2\times 2}$, 使得 $A^6 = E_2$, $A^3 \neq E_2$, $A^2 \neq E_2$, 且 A 中所有元素之和为 2023?

不变子空间

线性变换在不同基下的矩阵是不同的. 自然的问题是, 能否找到一组基, 使得这组基下的矩阵最简单? 这是我们接下来要探索的主要问题. 一个直观的想法是先对线性空间进行分解，使这个线性变换分解为维数较低的子空间上的线性变换.

定义

设 $\mathcal{A}$ 是线性空间 V 上的线性变换. 如果子空间 $W \subset V$ 满足 $\mathcal{A}W \subset W$, 即任意 $\alpha \in W$, 有 $\mathcal{A}\alpha \in W$, 则称 W 是 $\mathcal{A}$ 的不变子空间, 简称 $\mathcal{A}$-子空间.

如果 W 是 $\mathcal{A}$ 的不变子空间, 则 $\mathcal{A}$ 也可看作 W 上的线性变换, 称为 $\mathcal{A}$ 在 W 上的限制, 记作 $\mathcal{A}|_W$.

例

零子空间和 V 都是 $\mathcal{A}$ 的不变子空间, 称为平凡不变子空间.

例

$\mathcal{A}$ 的核和像都是 $\mathcal{A}$ 的不变子空间.

例

考虑 $P[x]$ 上的求导变换 $\mathcal{D}$. 因为次数 $< n$ 的多项式求导后次数仍 $< n$, 所以 $P[x]_n$ 是 $\mathcal{D}$ 的不变子空间.

下面探索不变子空间的一些简单性质.

命题

不变子空间的交仍是不变子空间, 不变子空间的和仍是不变子空间.

证明.

设 W_1, W_2 是 $\mathcal{A}$-子空间, 那么, $\mathcal{A}W_1 \subset W_1$, $\mathcal{A}W_2 \subset W_2$.
由 $\mathcal{A}(W_1 \cap W_2) = (\mathcal{A}W_1) \cap (\mathcal{A}W_2)$ 可知 $W_1 \cap W_2$ 是 $\mathcal{A}$-子空间.
对任意 $w_1 + w_2 \in W_1 + W_2$(其中 $w_1 \in W_1$, $w_2 \in W_2$), 有

$$\mathcal{A}(w_1 + w_2) = \mathcal{A}w_1 + \mathcal{A}w_2 \in W_1 + W_2.$$

可见 $W_1 + W_2$ 是 $\mathcal{A}$-子空间. □

商空间上诱导的线性变换

回忆一下, 对于线性空间 V 的子空间 W, 可定义商空间 V/W 为所有同余类 $[\alpha]$ 的集合, 其中

$$[\alpha] = \alpha + W = \{\alpha + \gamma \mid \gamma \in W\}, \quad \alpha \in V.$$

商空间 V/W 的加法和数乘分别定义为

$$[\alpha] + [\beta] = [\alpha + \beta], \quad k[\alpha] = [k\alpha].$$

命题

如果 W 是 $\mathcal{A}$-子空间, 则在商空间 V/W 上存在唯一的线性变换 $\overline{\mathcal{A}}$, 使得 $\overline{\mathcal{A}}([\alpha]) = [\mathcal{A}(\alpha)]$ 对任意 $\alpha \in V$ 成立.

称这里的 $\overline{\mathcal{A}}$ 为 $\mathcal{A}$ 在 V/W 上诱导的线性变换.

证明.

存在性: 直接把 $\overline{\mathcal{A}}([\alpha]) = [\mathcal{A}(\alpha)]$ 作为 $\overline{\mathcal{A}}$ 的定义. 我们首先验证这个定义是合理的. 事实上, 因为 W 是 $\mathcal{A}$-子空间, 所以当 $\alpha - \beta \in W$ 时, $\mathcal{A}(\alpha) - \mathcal{A}(\beta) \in W$. 这就证明了当 $[\alpha] = [\beta]$ 时, $[\mathcal{A}(\alpha)] = [\mathcal{A}(\beta)]$.

证明 (续).

其次, 我们验证上述映射 $\overline{\mathcal{A}}$ 是线性的:

$$\begin{aligned}&\overline{\mathcal{A}}(k[\alpha]+l[\beta])=\overline{\mathcal{A}}([k\alpha+l\beta])\\=&[\mathcal{A}(k\alpha+l\beta)]=[k\mathcal{A}(\alpha)+l\mathcal{A}(\beta)]\\=&k[\mathcal{A}(\alpha)]+l[\mathcal{A}(\beta)]=k\overline{\mathcal{A}}([\alpha])+l\overline{\mathcal{A}}([\beta]).\end{aligned}$$

唯一性是显然的. □

例

在 $P[x]$ 中, 考虑线性变换 $\mathcal{A}$ 如下

$$\mathcal{A}(g(x))=xg(x).$$

任取非零多项式 $f(x)$, 则 $f(x)$ 的所有倍式构成的子空间 W 是 $\mathcal{A}$ 的不变子空间. 在商空间 $P[x]/W$ 上, 诱导的线性变换为

$$\overline{\mathcal{A}}([g(x)])=[xg(x)].$$

命题

设 W 是 $\mathcal{A}$-子空间. 将 W 的一组基 $\alpha_1, \cdots, \alpha_r$ 扩充为 V 的一组基 $\alpha_1, \cdots, \alpha_r, \alpha_{r+1}, \cdots, \alpha_n$. 那么, $\mathcal{A}$ 在这组基下的矩阵为分块上三角矩阵 $\begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix}$, 其中 A_1 为 $\mathcal{A}|_W$ 在基 $\alpha_1, \cdots, \alpha_r$ 下的矩阵, A_3 为 $\overline{\mathcal{A}}$ 在基 $[\alpha_{r+1}], \cdots, [\alpha_n]$ 下的矩阵.

证明.

设 $\mathcal{A}$ 在基 $\alpha_1, \cdots, \alpha_n$ 下的矩阵为 $A = (a_{ij})$, 即 $\mathcal{A}\alpha_i = a_{1i}\alpha_1 + \cdots + a_{ni}\alpha_n$.
当 $1 \leq i \leq r$ 时, $\alpha_i \in W$, 所以 $\mathcal{A}\alpha_i \in W$, 即 $\mathcal{A}\alpha_i$ 可被 $\alpha_1, \cdots, \alpha_r$ 线性表出. 这就证明了 A 的前 r 列中后 $n-r$ 行全为零, 而前 r 行前 r 列构成的矩阵 A_1 恰好是 $\mathcal{A}|_W$ 的矩阵.
当 $r+1 \leq i \leq n$ 时,

$$\begin{aligned} [\mathcal{A}\alpha_i] &= [a_{1i}\alpha_1 + \cdots + a_{ni}\alpha_n] \\ &= [a_{r+1,i}\alpha_{r+1} + \cdots + a_{ni}\alpha_n] \\ &= a_{r+1,i}[\alpha_{r+1}] + \cdots + a_{ni}[\alpha_n]. \end{aligned}$$

可见 A 的后 $n-r$ 行后 $n-r$ 列构成的矩阵 A_3 恰好是 $\overline{\mathcal{A}}$ 的矩阵. □

利用上述证明中的思路, 不难得到

引理

设 V 能写为 $\mathcal{A}$-子空间 W_1 和 W_2 的直和, 并设 $\alpha_1, \cdots, \alpha_r$ 和 $\alpha_{r+1}, \cdots, \alpha_n$ 分别是 W_1 和 W_2 的基. 那么, $\mathcal{A}$ 在基 $\alpha_1, \cdots, \alpha_n$ 下的矩阵为准对角矩阵 $\mathrm{diag}(A_1, A_2)$, 其中 A_i 是 $\mathcal{A}|_{W_i}$ 的矩阵, $i = 1, 2$.

由此利用数学归纳法就有

命题

如果 V 能写为 $\mathcal{A}$-子空间 $W_1, \cdots, W_s$ 的直和, 则存在一组基, 使得 $\mathcal{A}$ 在这组基下的矩阵为准对角矩阵 $\mathrm{diag}(A_1, \cdots, A_s)$, 其中 A_i 是 $\mathcal{A}|_{W_i}$ 的矩阵, $1 \leq i \leq s$.

只要找到合适的不变子空间分解, 就可使线性变换的矩阵为准对角矩阵. 那么, 如何寻找不变子空间呢?

引理

如果有线性变换 $\mathcal{B}$ 满足 $\mathcal{AB} = \mathcal{BA}$, 则 $\mathcal{B}$ 的核与像都是 $\mathcal{A}$ 的不变子空间.

证明.

对于 $\ker \mathcal{B}$ 中任一元素 ξ, 有

$$\mathcal{BA}\xi = \mathcal{AB}\xi = 0,$$

所以 $\mathcal{A}\xi \in \ker \mathcal{B}$. 这就证明了 $\ker \mathcal{B}$ 是 $\mathcal{A}$-子空间.
类似地, 对于 $\mathcal{B}V$ 中任一元素 $\mathcal{B}\eta$, 有

$$\mathcal{A}(\mathcal{B}\eta) = \mathcal{BA}\eta \in \mathcal{B}V,$$

所以 $\mathcal{B}V$ 是 $\mathcal{A}$-子空间. □

作为上述引理的推论, 任取 $f(x) \in P[x]$, 则 $f(\mathcal{A})$ 与 $\mathcal{A}$ 可交换, 所以 $\ker f(\mathcal{A})$ 和 $\mathrm{im} f(\mathcal{A})$ 都是 $\mathcal{A}$ 的不变子空间.

引理 (分解引理)

设 $f(x), g(x) \in P[x]$ 是互素的多项式, 那么, $\ker f(\mathcal{A})g(\mathcal{A}) = \ker f(\mathcal{A}) \oplus \ker g(\mathcal{A})$.

证明.

由于 $f(x)$ 与 $g(x)$ 互素, 存在多项式 $u(x), v(x)$ 使得

$$f(x)u(x) + g(x)v(x) = 1,$$

所以 $f(\mathcal{A})u(\mathcal{A}) + g(\mathcal{A})v(\mathcal{A}) = \mathrm{id}$.
现在我们先证明

$$\ker f(\mathcal{A})g(\mathcal{A}) = \ker f(\mathcal{A}) + \ker g(\mathcal{A}).$$

一方面, $\ker f(\mathcal{A})$ 和 $\ker g(\mathcal{A})$ 都是 $W := \ker f(\mathcal{A})g(\mathcal{A})$ 的子空间, 因此它们的和也是 W 的子空间.
另一方面, 对 W 中任一元素 ξ, 令 $\alpha = g(\mathcal{A})v(\mathcal{A})\xi$, $\beta = f(\mathcal{A})u(\mathcal{A})\xi$, 则由前面的等式可得 $\xi = \alpha + \beta$, 并容易验证 $\alpha \in \ker f(\mathcal{A})$, $\beta \in \ker g(\mathcal{A})$.
两方面合起来, 就证明了 $W = \ker f(\mathcal{A}) + \ker g(\mathcal{A})$.
如果 $\eta \in \ker f(\mathcal{A}) \cap \ker g(\mathcal{A})$, 则 $\eta = u(\mathcal{A})f(\mathcal{A})\eta + v(\mathcal{A})g(\mathcal{A})\eta = 0$. 因此这个和是直和. □

特征值和特征向量

上面的引理提示我们, 可以适当取多项式 $f(x)$ 来制造不变子空间. 但对随意选取的多项式 $f(x)$, 核空间 $\ker f(\mathcal{A})$ 通常是平凡的. 怎样的多项式才能得到非平凡的不变子空间呢? 我们可以从 1 次多项式或 1 维的不变子空间开始考虑.

定义

设 V 是数域 P 上的线性空间, $\mathcal{A} \in \operatorname{End} V$. 对于 $\lambda_0 \in P$, 记 $E_{\lambda_0} = \ker(\mathcal{A} - \lambda_0 \mathrm{id})$. 如果 $E_{\lambda_0} \neq \{0\}$, 则称 λ_0 为 $\mathcal{A}$ 的一个 特征值, E_{λ_0} 中的非零向量称为属于特征值 λ_0 的特征向量.

简单地说, 如果有非零向量 ξ 使得

$$\mathcal{A}\xi = \lambda_0 \xi,$$

则 λ_0 是一个特征值, ξ 是属于特征值 λ_0 的一个特征向量, 而 E_{λ_0} 就是所有属于 λ_0 的特征向量以及零向量构成的线性空间, 也称为特征子空间. 特别地, E_0 就是 $\mathcal{A}$ 的核.
注意如果 ξ 是一个特征向量, 则 $\mathcal{A}\xi$ 与 ξ 平行, 所以 ξ 可张成 $\mathcal{A}$ 的一个 1 维不变子空间.

例

在线性空间 $V = C^\infty(0,1)$ 上, 考虑变换 $\mathcal{D} = \frac{\mathrm{d}}{\mathrm{d}x}$, 则任意实数 k 都是 $\mathcal{D}$ 的特征值, 因为 $\mathcal{D}(\mathrm{e}^{kx}) = k \cdot \mathrm{e}^{kx}$, 表明 e^{kx} 是属于特征值 k 的特征向量.

例

考虑 $\mathbb{R}^3$ 中的投影变换 π, 使得 $\pi(x,y,z) = (x,y,0)$. 由 $\pi(0,0,1) = 0$ 可知 $(0,0,1)$ 是属于特征值 0 的特征向量, 由 $\pi(1,0,0) = (1,0,0)$ 可知 $(1,0,0)$ 是属于特征值 1 的特征向量.

下面主要考虑有限维线性空间.

引理

设 V 是有限维线性空间, $\mathcal{A} \in \mathrm{End}\, V$. 那么, $\lambda_0 \in P$ 是 $\mathcal{A}$ 的特征值, 当且仅当 $\det(\lambda_0 \mathrm{id} - \mathcal{A}) = 0$.

证明.

取定 V 的一组基, 设 $\mathcal{A}$ 在这组基下的矩阵为 A.
由定义, λ_0 是特征值, 当且仅当存在非零向量 ξ 使得 $\mathcal{A}\xi = \lambda_0\xi$. 用坐标来描述, 也就是存在非零列向量 x 使得 $Ax = \lambda_0 x$, 也即 $(\lambda_0 E_n - A)x = 0$. 这个齐次线性方程组有非零解当且仅当 $\det(\lambda_0 E_n - A) = 0$. 注意 $\lambda_0 E_n - A$ 恰好是线性变换 $\lambda_0\mathrm{id} - \mathcal{A}$ 的矩阵, 从而引理得证. □

定义

设 $\mathcal{A}$ 是数域 P 上有限维线性空间 V 上的线性变换. 称 $f(\lambda) = \det(\lambda\mathrm{id} - \mathcal{A})$ 为 $\mathcal{A}$ 的特征多项式.

由这个定义, 前述引理也可叙述为: λ_0 是 $\mathcal{A}$ 的特征值, 当且仅当它是特征多项式 (在数域 P 上) 的根.

定义

设 $A \in P^{n\times n}$. 称 $f(\lambda) = \det(\lambda E_n - A)$ 为 A 的特征多项式, 它的根称为 A 的特征根或特征值. 若 $\lambda_0 \in P$ 是 A 的特征值, 则齐次线性方程组 $(\lambda_0 E_n - A)x = 0$ 的非零解称为 A 的属于 λ_0 的特征向量.

在 n 维线性空间 V 上, 要求一个线性变换 $\mathcal{A}$ 的特征值和特征向量, 可以先取一组基 $\alpha_1, \cdots, \alpha_n$, 求出 $\mathcal{A}$ 在这组基下的矩阵 A, 再计算 A 的特征值和特征向量; A 的每个特征值也就是 $\mathcal{A}$ 的一个特征值, 而 A 的每个特征向量是 $\mathcal{A}$ 的某个特征向量在前述基下的坐标.
求矩阵 A 的特征值及特征向量的步骤如下:

- 计算特征多项式 $f(\lambda) = \det(\lambda E_n - A)$;
- 求 $f(\lambda)$ 的根, 它们就是 A 的特征值;
- 对每个特征值 λ_0, 解齐次线性方程组 $(\lambda_0 E_n - A)x = 0$, 其基础解系就可表出所有属于 λ_0 的特征向量.

例

求矩阵 $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$ 的特征值和特征向量.

解答

为计算 A 的特征多项式, 我们对矩阵 $\lambda E_3 - A$ 作初等行变换

$$\begin{bmatrix} \lambda-1 & -2 & -2 \\ -2 & \lambda-1 & -2 \\ -2 & -2 & \lambda-1 \end{bmatrix} \to \begin{bmatrix} 2 & 2 & 1-\lambda \\ -2 & \lambda-1 & -2 \\ \lambda-1 & -2 & -2 \end{bmatrix}$$

$$\to \begin{bmatrix} 2 & 2 & 1-\lambda \\ 0 & \lambda+1 & -\lambda-1 \\ 0 & -\lambda-1 & -2+(1-\lambda)^2/2 \end{bmatrix} \to \begin{bmatrix} 2 & 2 & 1-\lambda \\ 0 & \lambda+1 & -\lambda-1 \\ 0 & 0 & (\lambda+1)(\lambda-5)/2 \end{bmatrix}.$$

可见 A 的特征多项式为 $(\lambda+1)^2(\lambda-5)$, 从而特征值为 -1(二重) 和 5.

解答 (续)

将 $\lambda = 5$ 代入上述初等行变换过程, 可知线性方程组 $(5E_3 - A)x = 0$ 的基础解系为 $\eta_3 = (1,1,1)^{\mathrm{T}}$. 即属于 5 的特征向量为 $k(1,1,1)'$, $k \neq 0$.
同理可得属于 -1 的特征向量为 $k(1,-1,0)^{\mathrm{T}} + l(1,0,-1)^{\mathrm{T}}$, 其中 k, l 不全为零.

例

求 $P[x]_n$ 的线性变换 $\mathcal{D} = \frac{\mathrm{d}}{\mathrm{d}x}$ 的特征值及特征向量.

解答

取 $P[x]_n$ 的一组基 1, x, $\frac{x^2}{2!}$, $\cdots$, $\frac{x^{n-1}}{(n-1)!}$, 则 $\mathcal{D}$ 在这组基下的矩阵为

$$D = \begin{bmatrix} 0 & E_{n-1} \\ 0 & 0 \end{bmatrix},$$

从而可求得 D 的特征多项式为 λ^n, 特征值为 0. 由于线性方程组 $Dx = 0$ 的基础解系为 $(1,0,\cdots,0)^{\mathrm{T}}$, 所以 $\mathcal{D}$ 的特征向量为非零常数 k.

例

求平面旋转变换 I_θ 的特征值与特征向量.

解答

在平面直角坐标系中, I_θ 的矩阵为

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

从而特征多项式为

$$\lambda^2 - 2\lambda\cos\theta + 1.$$

当 $\cos\theta \neq \pm 1$ 时, 特征多项式无实根, 这时 I_θ 没有特征值.
当 $\cos\theta = 1$, 即 $\theta = 2k\pi$ 时, $I_\theta = \mathrm{id}$, 这时特征值为 1, 任意非零向量都是特征向量.
当 $\cos\theta = -1$, 即 $\theta = (2k-1)\pi$ 时, $I_\theta = -\mathrm{id}$, 这时特征值为 -1, 任意非零向量都是特征向量.

定理 (Hamilton-Cayley)

设 $A \in P^{n\times n}$, 则 A 的特征多项式是 A 的一个零化多项式.

证明.

令 $B = \lambda E_n - A$, 则 A 的特征多项式 $f(\lambda) = \det(B)$, 并有 $BB^* = f(\lambda)E_n$.
注意矩阵 B^* 中每个元素都是 λ 的多项式, 且次数 $< n$, 可将 B^* 写为

$$\lambda^{n-1}B_{n-1} + \lambda^{n-2}B_{n-2} + \cdots + \lambda B_1 + B_0,$$

其中 $B_{n-1}, B_{n-2}, \cdots, B_0 \in P^{n\times n}$. 又设

$$f(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0,$$

则等式 $BB^* = f(\lambda)E_n$ 可展开写为

$$\begin{aligned}&(\lambda E_n - A)(\lambda^{n-1}B_{n-1} + \lambda^{n-2}B_{n-2}\\ &\qquad + \cdots + \lambda B_1 + B_0)\\ =&\lambda^n E_n + \lambda^{n-1}a_{n-1}E_n + \cdots + \lambda a_1 E_n + a_0 E_n.\end{aligned}$$

比较两端各项的系数, 可得

证明 (续).

$$\begin{aligned} B_{n-1} &= E_n, \\ B_{n-2} - AB_{n-1} &= a_{n-1}E_n, \\ \cdots \quad & \quad \cdots \\ B_0 - AB_1 &= a_1 E_n, \\ -AB_0 &= a_0 E_n. \end{aligned}$$

依次用 $A^n, A^{n-1}, \cdots, A, E_n$ 左乘上述各式, 再相加, 则左端为零, 右端为 $f(A)$, 定理得证. $\square$

思考题

(***) 有同学说, 直接在等式

$$(\lambda E_n - A)B^* = f(\lambda)E_n$$

中将 λ 替换为 A, 不就证明了 $f(A) = 0$ 吗? 这个说法有什么问题?

利用矩阵与线性变换之间的对应关系, Hamilton-Cayley 定理也可叙述为

推论

设 $\mathcal{A}$ 是 n 维线性空间 V 上的线性变换, 则 $\mathcal{A}$ 的特征多项式是 $\mathcal{A}$ 的一个零化多项式.

由于最小多项式是所有零化多项式的公因式, 我们有

推论

设 $\mathcal{A}$ 是 n 维线性空间 V 上的线性变换, 则 $\mathcal{A}$ 的最小多项式是特征多项式的因式, 次数 $\leq n$.

特征多项式与不变子空间

特征多项式的一个直接应用是帮助我们寻找不变子空间.

引理

设 $\mathcal{A}$ 是 n 维线性空间 V 上的线性变换, W 是 $\mathcal{A}$ 的不变子空间. 设 $\mathcal{A}|_W$ 和 $\mathcal{A}$ 的特征多项式分别为 $f_1(\lambda)$, $f(\lambda)$; 它们的最小多项式分别为 $d_1(\lambda)$, $d(\lambda)$. 那么, $f_1(\lambda) \mid f(\lambda)$, $d_1(\lambda) \mid d(\lambda)$.

证明.

将 W 的一组基扩充为 V 的一组基, 则 $\mathcal{A}$ 在这组基下为准上三角形矩阵 $A = \begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix}$, 其中 A_1 为 $\mathcal{A}|_W$ 的矩阵.

注意 A 的特征多项式 $f(\lambda) = \det(\lambda I - A) = \det(\lambda I - A_1)\det(\lambda I - A_3)$, 而 A_1 的特征多项式 $f_1(\lambda) = \det(\lambda I - A_1)$. 可见 $f_1(\lambda)$ 是 $f(\lambda)$ 的因式.

对任意多项式 $g(\lambda)$, 有 $g(A) = \begin{bmatrix} g(A_1) & * \\ 0 & g(A_3) \end{bmatrix}$. 取 $g(\lambda)$ 为 A 的最小多项式 $d(\lambda)$, 则有 $d(A_1) = 0$, 即 $d(\lambda)$ 是 A_1 的一个零化多项式, 所以 $d_1(\lambda) | d(\lambda)$. □

命题

设 $\mathcal{A}$ 是 n 维线性空间 V 上的线性变换, 其特征多项式和最小多项式分别为 $f(\lambda)$, $d(\lambda)$. 如果 V 能分解为 $\mathcal{A}$-子空间 W_1, W_2, $\cdots$, W_s 的直和, 那么, $\mathcal{A}|_{W_i}$ 的特征多项式 $f_i(\lambda)$, 最小多项式 $d_i(\lambda)$ 满足

$$f(\lambda) = f_1(\lambda)f_2(\lambda)\cdots f_s(\lambda),$$
$$d(\lambda) = [d_1(\lambda), d_2(\lambda), \cdots, d_s(\lambda)].$$

证明.

分别取各个不变子空间的一组基, 合在一起, 构成 V 的一组基. $\mathcal{A}$ 在这组基下的矩阵为准对角矩阵 $A = \mathrm{diag}(A_1, A_2, \cdots, A_s)$, 其中 A_i 为 $\mathcal{A}|_{W_i}$ 的矩阵.
于是

$$f(\lambda) = \det(\lambda I - A) = \det(\lambda I - A_1)\cdots\det(\lambda I - A_s) = f_1(\lambda)\cdots f_s(\lambda).$$

注意 $d(A) = \mathrm{diag}(d(A_1), \cdots, d(A_s))$. 因此, 当 $d(A) = 0$ 时一定有 $d(A_i) = 0$, 从而 $d_i(\lambda)|d(\lambda)$. 反之, 当 $d_i(\lambda)|d(\lambda)$ 时, 一定有 $d(A_i) = 0 (1 \le i \le s)$, 从而 $d(A) = 0$. 由次数的最小性可知 $d(\lambda) = [d_1(\lambda), \cdots, d_s(\lambda)]$. □

根子空间分解

定理

设 V 是复数域上的 n 维线性空间, $\mathcal{A} \in \mathrm{End}\, V$ 的特征多项式 $f(\lambda)$ 有如下分解

$$f(\lambda) = (\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_s)^{n_s},$$

其中 $\lambda_1, \cdots, \lambda_s$ 是互不相同的复数. 则 V 可分解为不变子空间 $R_j = \ker(\mathcal{A} - \lambda_j \mathrm{id})^{n_j}$, $j = 1, \cdots, s$ 的直和.

证明.

反复利用分解引理的结论, 即可得到:

$$\ker f(\mathcal{A}) = \ker(\mathcal{A} - \lambda_1 \mathrm{id})^{n_1} \oplus \cdots \oplus \ker(\mathcal{A} - \lambda_s \mathrm{id})^{n_s}.$$

由 Hamilton-Cayley 定理, $f(\mathcal{A}) = 0$, 所以上式左端为 V, 定理得证. □

称 $R_j = \ker(\mathcal{A} - \lambda_j \mathrm{id})^{n_j}$ 为属于 λ_j 的根子空间. 由于 λ_j 是特征多项式的 n_j 重根, 所以我们称 n_j 为 λ_j 的代数重数.

引理

若 λ_0 是 $\mathcal{A}$ 的一个特征值, $g(x)$ 是 $\mathcal{A}$ 的一个零化多项式, 则 $g(\lambda_0) = 0$.

证明.

取属于 λ_0 的特征向量 ξ. 易知对任意多项式 $g(x)$, 都有 $g(\mathcal{A})\xi = g(\lambda_0)\xi$. 特别地, 当 $g(x)$ 为零化多项式时, 左端为零, 从而必有 $g(\lambda_0) = 0$. $\square$

命题

属于 λ_j 的根子空间 R_j 的维数, 恰好等于 λ_j 的代数重数.

证明.

注意 $\mathcal{A}|_{R_j}$ 有一个零化多项式是 $(\lambda - \lambda_j)^{n_j}$, 所以由上述引理可知 $\mathcal{A}|_{R_j}$ 只有一个特征值 λ_j, 从而它的特征多项式形如 $(\lambda - \lambda_j)^{m_j}$(其中 m_j 是 R_j 的维数); 这个特征多项式又是 $f(\lambda)$ 的因式, 所以 $m_j \leq n_j$.

又由 $V = R_1 \oplus \cdots \oplus R_s$ 可知 $n = m_1 + \cdots + m_s$. 因此一定有 $m_j = n_j$. $\square$

可对角化的线性变换

对于怎样的线性变换，可以适当取一组基，使得它在这组基下的矩阵为对角矩阵？等价地，怎样的矩阵可以相似于对角矩阵？这样的线性变换或矩阵称为可对角化的.

引理

若 $\lambda_1, \cdots, \lambda_s$ 是 $\mathcal{A}$ 的互不相同的特征值，则特征子空间 $E_{\lambda_1}, \cdots, E_{\lambda_s}$ 的和是直和. 因此，属于不同特征值的特征向量是线性无关的.

证明.

注意一次多项式 $\lambda - \lambda_1, \cdots, \lambda - \lambda_s$ 两两互素，因此由分解引理可知 $\ker(\mathcal{A} - \lambda_1 \mathrm{id}), \cdots, \ker(\mathcal{A} - \lambda_s \mathrm{id})$ 的和是直和. □

我们将特征子空间 E_{λ_j} 的维数称为特征值 λ_j 的几何重数.

定理

设 V 是数域 P 上的 n 维线性空间, $\mathcal{A} \in \mathrm{End}\, V$, 则以下条件等价:

(1) $\mathcal{A}$ 可对角化;

(2) $\mathcal{A}$ 有 n 个线性无关的特征向量;

(3) V 可写为不同特征子空间的直和;

(4) $\mathcal{A}$ 的最小多项式是不同一次因式的乘积.

证明.

(1)⇒(2). 取这样一组基, 使得 $\mathcal{A}$ 在这组基下的矩阵为对角矩阵. 容易发现, 这组基中每个向量都是特征向量.

(2)⇒(3). 这 n 个线性无关的特征向量分别属于相应的特征子空间. 因此, 这些特征子空间的维数之和 $\geq n$. 又由引理, 不同的特征子空间的和是直和, 所以它们的直和恰等于 V.

(3)⇒(4). 设不同特征值为 $\lambda_1, \cdots, \lambda_s$, 相应的特征子空间分别为 $E_1, \cdots, E_s$. 那么, $\mathcal{A}|_{E_j}$ 的最小多项式为 $\lambda - \lambda_j$. 因此, $\mathcal{A}$ 的最小多项式是 $\lambda - \lambda_1, \cdots, \lambda - \lambda_s$ 的最小公倍式, 即它们的乘积.

证明 (续).

(4)⇒(1). 如果 $\mathcal{A}$ 的最小多项式 $d(\lambda) = (\lambda - \lambda_1)\cdots(\lambda - \lambda_s)$, 那么, 由分解引理可知

$$V = \ker(\mathcal{A} - \lambda_1 \mathrm{id}) \oplus \cdots \oplus \ker(\mathcal{A} - \lambda_s \mathrm{id}).$$

分别取各个特征子空间 $\ker(\mathcal{A} - \lambda_i \mathrm{id})$ 的一组基, 合在一起构成 V 的基, 则 $\mathcal{A}$ 在这组基下的矩阵为对角矩阵. □

推论

设 V 是 n 维线性空间, 且 $\mathcal{A} \in \mathrm{End}\, V$ 可对角化, 则它的每个特征值的几何重数与代数重数相等.

证明.

注意每个特征子空间 $E_j = \ker(\mathcal{A} - \lambda_j \mathrm{id})$ 包含在相应的根子空间 $R_j = \ker(\mathcal{A} - \lambda_j \mathrm{id})^{n_j}$ 中, 所以每个特征值的几何重数 $\leq$ 代数重数.
如果 $\mathcal{A}$ 可对角化, 则 V 可写为各特征子空间的直和, 意味着所有特征值的几何重数之和为 n, 恰与代数重数之和相等. 因此, 每个特征值的几何重数都与代数重数相等. □

前面的定理有一种重要且常见的特殊情形.

推论

如果 $\mathcal{A}$ 有 $\dim V$ 个互不相同的特征值, 则 $\mathcal{A}$ 可对角化.

例

$\mathbb{C}^{n\times n}$ 中矩阵 $A = \begin{bmatrix} 0 & E_{n-1} \\ 1 & 0 \end{bmatrix}$ 是否相似于对角矩阵?

解答

容易算得 A 的特征多项式为 $\lambda^n - 1$. 它有 n 个互不相同的特征值

$$\lambda_k = \cos\frac{2k\pi}{n} + \sqrt{-1}\sin\frac{2k\pi}{n}, \quad k = 0, 1, \cdots, n-1.$$

所以 A 相似于对角矩阵 $\operatorname{diag}(\lambda_0, \lambda_1, \cdots, \lambda_{n-1})$.

思考题

(**) 上述矩阵 A 的特征向量是什么?

复线性空间的线性变换

下面我们将利用前面的知识继续讨论复线性空间的线性变换, 目标是找到一组基, 使得该线性变换在这组基下的矩阵最简单. 我们先从二维和三维的情形看起.

定理

设 V 是 2 维复线性空间, $\mathcal{A} \in \mathrm{End}\, V$ 的特征多项式和最小多项式分别为 $f(\lambda)$, $d(\lambda)$. 那么

- 如果 $f(\lambda)$ 有两个不同的根 λ_1, λ_2, 则存在 V 的一组基, 使得 $\mathcal{A}$ 的矩阵为 $\mathrm{diag}(\lambda_1, \lambda_2)$;
- 如果 $f(\lambda)$ 有两个相等的根 λ_0, 且 $d(\lambda) = \lambda - \lambda_0$, 则 $\mathcal{A} = \lambda_0 \mathrm{id}$, 这时 $\mathcal{A}$ 在任意一组基下的矩阵为 $\lambda_0 E_2$;
- 如果 $f(\lambda)$ 有两个相等的根 λ_0, 且 $d(\lambda) \neq \lambda - \lambda_0$, 则存在 V 的一组基, 使得 $\mathcal{A}$ 的矩阵为 $\begin{bmatrix} \lambda_0 & 1 \\ 0 & \lambda_0 \end{bmatrix}$.

容易看到, 在前两种情形, 最小多项式能写为不同一次因式的乘积, 所以 $\mathcal{A}$ 可对角化. 因此下面我们只证明最后一种情形, 即 $f(\lambda) = d(\lambda) = (\lambda - \lambda_0)^2$ 的情形.

证明.

只证明最后一种情形. 由于 $\mathcal{A}-\lambda_0\mathrm{id}\neq 0$, 一定存在某个向量 ξ 使得 $(\mathcal{A}-\lambda_0\mathrm{id})\xi=\eta\neq 0$. 注意 $(\mathcal{A}-\lambda_0\mathrm{id})^2=0$, 所以 $(\mathcal{A}-\lambda_0\mathrm{id})\eta=0$, 即 η 是属于 λ_0 的特征向量. 而 ξ 不是特征向量, 所以 η 与 ξ 线性无关, 构成 V 的一组基.
由 $\mathcal{A}\eta=\lambda_0\eta$, $\mathcal{A}\xi=\eta+\lambda_0\xi$ 可知 $\mathcal{A}$ 在这组基下的矩阵为 $\begin{bmatrix}\lambda_0 & 1\\ 0 & \lambda_0\end{bmatrix}$. □

定理

设 V 是 3 维复线性空间, $\mathcal{A}\in\mathrm{End}\,V$. 那么, $\mathcal{A}$ 在适当基下的矩阵是以下 6 种情形之一:

$$\begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3\end{bmatrix},\quad \begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_2\end{bmatrix},\quad \begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & 1\\ & & \lambda_2\end{bmatrix},$$
$$\begin{bmatrix}\lambda_1 & & \\ & \lambda_1 & \\ & & \lambda_1\end{bmatrix},\quad \begin{bmatrix}\lambda_1 & & \\ & \lambda_1 & 1\\ & & \lambda_1\end{bmatrix},\quad \begin{bmatrix}\lambda_1 & 1 & \\ & \lambda_1 & 1\\ & & \lambda_1\end{bmatrix}.$$

证明.

如果 $\mathcal{A}$ 有三个不相等的特征值, 则 $\mathcal{A}$ 可对角化, 这就是第一种情形.
如果 $\mathcal{A}$ 有两个不等的特征值, 则其中一个重数为 1(记作 λ_1), 另一个重数为 2(记作 λ_2). 这时根子空间 R_1, R_2 的维数分别为 $1, 2$. 利用上一个定理的结论, 不难发现 $\mathcal{A}$ 的矩阵为情形二或情形三.
如果 $\mathcal{A}$ 仅有一个特征值 (记作 λ_1), 重数为 3, 则需按最小多项式 $d(\lambda)$ 的次数来进行讨论:

- 若 $d(\lambda) = \lambda - \lambda_1$, 则 $\mathcal{A} = \lambda_1 \mathrm{id}$, 这就是情形四;
- 若 $d(\lambda) = (\lambda - \lambda_1)^2$, 记 $\mathcal{B} = \mathcal{A} - \lambda_1 \mathrm{id}$, 则 $\mathcal{B} \neq 0$, $\mathcal{B}^2 = 0$. 考虑不变子空间 $W = \mathcal{B}V$ 和 $U = \ker \mathcal{B}$, 易知 $W \subset U$, 且 $\dim W + \dim U = 3$, 所以 $\dim W = 1$, $\dim U = 2$. 取 W 的基 $\mathcal{B}\xi = \eta \neq 0$, 则易知 ξ, η 线性无关, 且 $\mathcal{B}\eta = 0$. 将 η 扩充为 U 的一组基 η, α, 于是 α, η, ξ 构成 V 的一组基, $\mathcal{A}$ 在这组基下的矩阵为情形五;
- 若 $d(\lambda) = (\lambda - \lambda_1)^3$, 则 $\mathcal{B} = \mathcal{A} - \lambda_1 \mathrm{id}$ 满足 $\mathcal{B}^2 \neq 0$, $\mathcal{B}^3 = 0$. 取 $\alpha \in V$ 使得 $\mathcal{B}^2\alpha \neq 0$, 则 $\mathcal{B}^2\alpha, \mathcal{B}\alpha, \alpha$ 构成 V 的一组基. $\mathcal{A}$ 在这组基下的矩阵为情形六.

□

二、三维的结果可以统一起来.

定义

对正整数 k, 令 $N_k = \begin{bmatrix} & E_{k-1} \\ 0 & \end{bmatrix}$, 并记 $J_k(\lambda_0) = \lambda_0 E_k + N_k$, 即

$$J_k(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & 0 & \cdots & 0 \\ 0 & \lambda_0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_0 \end{bmatrix},$$

称 $J_k(\lambda_0)$ 为 k 阶 Jordan 块. 如果一个准对角矩阵的每个对角块都是 Jordan 块, 则称它为 *Jordan* 矩阵.

定理

对二维或三维复线性空间的线性变换, 总可找到一组基, 使得它在这组基下的矩阵为 *Jordan* 矩阵.

幂零变换

现在继续讨论更高维数的复线性空间的线性变换. 利用根子空间分解定理, 只需考虑该线性变换在每个根子空间中的限制就可以了. 在每个根子空间中, 该线性变换只有一个特征值, 可不妨设这个特征值为零.

定义

设 $\mathcal{N}$ 是线性空间 W 上的线性变换. 若存在正整数 k 使得 $\mathcal{N}^k = 0$, 则称 $\mathcal{N}$ 为幂零变换.

引理

设 $\mathcal{N}$ 是 W 上的幂零变换, α 是 W 中非零向量, 则存在正整数 r 满足 $\mathcal{N}^{r-1}\alpha \neq 0$, $\mathcal{N}^r\alpha = 0$. 这时 $\alpha, \mathcal{N}\alpha, \cdots, \mathcal{N}^{r-1}\alpha$ 线性无关, 它们所张成的子空间称为一个循环子空间, 记作 $\langle\alpha\rangle$.

通常称 $\mathcal{N}^{r-1}\alpha, \cdots, \mathcal{N}\alpha, \alpha$ 为循环子空间 $\langle\alpha\rangle$ 的自然基底. 在这组基中, 恰有一个向量属于 $\ker\mathcal{N}$, 就是 $\mathcal{N}^{r-1}\alpha$. $\mathcal{N}|_{\langle\alpha\rangle}$ 在这组基下的矩阵是 r 阶 Jordan 块 $J_r(0) = N_r$.

命题

设 $\mathcal{N}$ 是 n 维线性空间 W 上的幂零变换, 则 W 可写为若干个循环子空间的直和.

证明.

对维数 n 用数学归纳法. 当 $n=1$ 时结论是显然的. 假设当 $n<m$ 时结论成立, 那么, 对于 m 维线性空间 W, 我们考虑不变子空间 $\mathcal{N}W$. 易知 $\dim \mathcal{N}W < m$(否则 $\mathcal{N}$ 可逆, 不可能有 $\mathcal{N}^k = 0$). 所以由归纳假设可知, $\mathcal{N}W$ 可写为若干个循环子空间的直和, 设

$$\mathcal{N}W = \langle \eta_1 \rangle \oplus \cdots \oplus \langle \eta_t \rangle.$$

其中, 由于 $\eta_j \in \mathcal{N}W$, 可设 $\eta_j = \mathcal{N}\alpha_j$. 于是我们得到了 W 中的循环子空间 $\langle \alpha_1 \rangle$, $\cdots$, $\langle \alpha_t \rangle$.

在每个循环子空间的基底中, 恰有一个向量属于 $\ker \mathcal{N}$. 在这些向量的基础上添加 $\alpha_{t+1}, \cdots, \alpha_s$, 就得到 $\ker \mathcal{N}$ 的一组基.

于是我们最终得到 $W = \langle \alpha_1 \rangle \oplus \cdots \oplus \langle \alpha_s \rangle$. $\square$

α_1				
$\mathcal{N}\alpha_1$	α_2			
$\mathcal{N}^2\alpha_1$	$\mathcal{N}\alpha_2$			
$\dots$	$\dots$	α_t		
$\dots$	$\dots$	$\mathcal{N}\alpha_t$		
$\dots$	$\dots$	$\dots$		
$\mathcal{N}^{r_1-1}\alpha_1$	$\mathcal{N}^{r_2-1}\alpha_2$	$\mathcal{N}^{r_t-1}\alpha_t$	α_{t+1}	α_s

上图中, 每一列是相应的循环子空间的一组基. 而按行看, 最下方一层恰好是 $\ker\mathcal{N}$ 的一组基, 最下方 2 层恰好是 $\ker\mathcal{N}^2$ 的一组基, $\cdots$, 最下方 j 层恰好是 $\ker\mathcal{N}^j$ 的一组基.

因此, 在 W 的循环子空间分解中, 维数 $\geq j$ 的循环子空间的个数等于 $\ker\mathcal{N}^j$ 与 $\ker\mathcal{N}^{j-1}$ 的维数之差. 因而, 维数 $= j$ 的循环子空间的个数等于

$$2\dim\ker\mathcal{N}^j - \dim\ker\mathcal{N}^{j+1} - \dim\ker\mathcal{N}^{j-1}.$$

复线性空间线性变换的标准形

定理

设 V 是复数域上的 n 维线性空间, $\mathcal{A} \in \mathrm{End}\, V$, 则存在 V 的一组基, 使得 $\mathcal{A}$ 在这组基下的矩阵为 *Jordan* 矩阵.

证明.

由根子空间分解定理, V 可写为根子空间 $R_1, \cdots, R_s$ 的直和, 其中

$$R_j = \ker(\mathcal{A} - \lambda_j \mathrm{id})^{n_j}.$$

令 $\mathcal{N}_j = (\mathcal{A} - \lambda_j \mathrm{id})|_{R_j}$, 则 $\mathcal{N}_j$ 是 R_j 上的幂零变换, 从而 R_j 可写为若干个循环子空间的直和. 由于 $\mathcal{N}_j$ 在每个循环子空间的自然基底下的矩阵为 Jordan 块, 所以 $\mathcal{N}_j$ 在 R_j 的一组基下的矩阵为 Jordan 矩阵. 进而 $\mathcal{A}|_{R_j}$ 在相同的基下的矩阵为 Jordan 矩阵. □

这个定理也可用矩阵来叙述: 设 $A \in \mathbb{C}^{n\times n}$, 则 A 一定相似于某个 Jordan 矩阵. 该 Jordan 矩阵称为 A 的 Jordan 标准形.

例

求 $A=\begin{bmatrix}4 & 5 & -2\\ -2 & -2 & 1\\ -1 & -1 & 1\end{bmatrix}$ 的 Jordan 标准形.

解答

对于 2 阶或 3 阶矩阵, 只需计算出特征多项式和最小多项式, 就可利用前面的分类定理得到 *Jordan* 标准形.
易知 A 的特征多项式为 $(\lambda-1)^3$. 又依次计算可知 $A-E_3\neq 0$, $(A-E_3)^2\neq 0$, 因此最小多项式为 $(\lambda-1)^3$. 故 A 的 *Jordan* 标准形为 $J_3(1)$.

例

设 $A=\begin{bmatrix}1 & 2 & 1\\ 1 & -1 & 1\\ 2 & 0 & 1\end{bmatrix}$, 求可逆矩阵 T, 使得 $T^{-1}AT$ 为 Jordan 矩阵.

解答

容易算得 A 的特征多项式为 $(\lambda+1)^2(\lambda-3)$, 从而最小多项式可能为 $(\lambda+1)(\lambda-3)$ 或 $(\lambda+1)^2(\lambda-3)$.

计算可知 $A-3E_3=\begin{bmatrix}-2 & 2 & 1\\ 1 & -4 & 1\\ 2 & 0 & -2\end{bmatrix}$, $(A+E_3)(A-3E_3)=\begin{bmatrix}0 & -4 & 2\\ 0 & 2 & -1\\ 0 & 4 & -2\end{bmatrix}\neq 0$.

因此, A 的最小多项式为 $(\lambda+1)^2(\lambda-3)$, 故 A 的 *Jordan* 标准形为 $J=\operatorname{diag}(J_2(-1),3)$.

取 $\beta=(1,1,-2)^{\mathsf{T}}$, 则 $\alpha=(A+E_3)\beta=(2,-1,-2)^{\mathsf{T}}$, 且 $(A+E_3)\alpha=0$. 又易算得属于 3 的特征向量 $\gamma=(2,1,2)^{\mathsf{T}}$. 因此, 取 $T=(\alpha,\beta,\gamma)$, 就有 $AT=TJ$, 即 $T^{-1}AT=J$.

思考题

(***) 上面的向量 β 是怎样选取的? 这种做法对一般的复矩阵都有效吗?

Jordan 标准形的应用

例

设 V 是 n 维复线性空间, $\mathcal{A}$ 是 V 上的线性变换. 证明存在可对角化的线性变换 $\mathcal{D}$ 和幂零线性变换 $\mathcal{N}$, 使得 $\mathcal{A}=\mathcal{D}+\mathcal{N}$, 且 $\mathcal{D}\mathcal{N}=\mathcal{N}\mathcal{D}$.

证明.

取这样一组基, 使得 $\mathcal{A}$ 在这组基下的矩阵为 Jordan 标准形 J. 取 J 的所有对角元按顺序排成对角矩阵 D, 并令 $N=J-D$, 则 N 是幂零矩阵.

为验证 D 与 N 交换, 只需注意到每个 Jordan 块中, 对角元构成的是数量矩阵, 它总与剩下的部分可交换.

现在, 分别取 $\mathcal{D}$ 和 $\mathcal{N}$ 为矩阵 D, N 所对应的线性变换, 即证. $\square$

思考题

(***) 如果 n 维复线性空间 V 上的线性变换 $\mathcal{A}$ 的 n 个特征值是 $\lambda_1,\cdots,\lambda_n$(可能有相等的), 证明: 对任意复多项式 $g(x)$, 线性变换 $g(\mathcal{A})$ 的 n 个特征值是 $g(\lambda_1),\cdots,g(\lambda_n)$.

例

如果 n 阶复矩阵 A 恰有一个特征值为 1, 而所有其他特征值的模长都 <1, 证明: $\lim_{k\to\infty} A^k = uv^{\mathsf{T}}$, 其中 u 是 A 的属于特征值 1 的特征向量, v 是 A^{T} 的属于特征值 1 的特征向量, 且 $v^{\mathsf{T}}u = 1$.

证明.

先对 Jordan 矩阵来证明. 注意 Jordan 块 $J_m(\lambda_0) = \lambda_0 E_m + N_m$ 满足

$$\begin{aligned} J_m(\lambda_0)^k = \lambda_0^k E_m + C_k^1 \lambda_0^{k-1} N_m + C_k^2 \lambda_0^{k-2} N_m^2 \\ + \cdots + C_k^m \lambda_0^{k-m} N_m^m, \end{aligned}$$

可见, 若 $|\lambda_0| < 1$, 则 $\lim_{k\to\infty} J_m(\lambda_0)^k \to 0$.

若 Jordan 矩阵 J 恰有一个特征值为 1, 而所有其他特征值的模长都 <1, 则可设 $J = \operatorname{diag}(1, J_{m_1}(\lambda_1), \cdots, J_{m_s}(\lambda_s))$, 其中 $|\lambda_1| < 1, \cdots, |\lambda_s| < 1$. 由前面的结果可知 $\lim_{k\to\infty} J^k = \operatorname{diag}(1, 0, \cdots, 0) = xx^{\mathsf{T}}$, 其中列向量 $x = (1, 0, \cdots, 0)^{\mathsf{T}}$.

证明 (续).

再考虑一般情形. 设 A 的 Jordan 标准形为 J, 则有可逆矩阵 P 使得 $P^{-1}AP = J$, 其中 P 的第一列就是属于特征值 1 的特征向量 u.

又在等式 $P^{-1}AP = J$ 两端取转置可得 $P^{\mathsf{T}}A^{\mathsf{T}}(P^{-1})^{\mathsf{T}} = J^{\mathsf{T}}$, 可见 $(P^{-1})^{\mathsf{T}}$ 的第一列是 A^{T} 的属于特征值 1 的特征向量 v. 由 $P^{-1}P$ 的第一行第一列为 1 可知 $v^{\mathsf{T}}u = 1$.

于是最终我们得到

$$\lim_{k\to\infty} A^k = \lim_{k\to\infty} PJ^kP^{-1} = Pxx^{\mathsf{T}}P^{-1} = uv^{\mathsf{T}},$$

证毕.

思考题

(***) 设 $n \geq 2$ 且 $A \in \mathbb{C}^{n\times n}$ 满足

$$\operatorname{rank}(A) \neq \operatorname{rank}(A^2).$$

证明存在非零矩阵 B, 使得 $AB = BA = B^2 = 0$.