(对称性) 若 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似, 则反过来 $\boldsymbol{B}$ 也与 $\boldsymbol{A}$ 相似;

(传递性) 若 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似, $\boldsymbol{B}$ 与 $\boldsymbol{C}$ 相似, 则 $\boldsymbol{A}$ 也与 $\boldsymbol{C}$ 相似.

证明. 用定义可以直接验证, 略. □

因此, 当 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 是相似的时候, 我们也称它们为**相似等价**的.

推论 6.2.13. 有限维线性空间 V 上的一个线性变换在 V 的不同基下的矩阵是相似的.

命题 6.2.14 (相似矩阵的基本性质)**.** 设 $\boldsymbol{A}, \boldsymbol{A}_i, \boldsymbol{B}, \boldsymbol{B}_i$ 都是 F 上的方阵.

(1) 若 $\boldsymbol{A} \sim \boldsymbol{B}$, 则 $\boldsymbol{A}^{\mathsf{T}} \sim \boldsymbol{B}^{\mathsf{T}}$.

(2) 若 $\boldsymbol{A} \sim \boldsymbol{B}$, 则 $\boldsymbol{A}^m \sim \boldsymbol{B}^m$, 其中 m 为正整数.

(3) 若 $\boldsymbol{A} \sim \boldsymbol{B}$, 设 $f(x)$ 是一个一元多项式, 则矩阵多项式 $f(\boldsymbol{A}) \sim f(\boldsymbol{B})$.

(4) 若 $\boldsymbol{A}_i \sim \boldsymbol{B}_i$, $i = 1, 2, \ldots, s$, 则 $\operatorname{diag}(\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_s) \sim \operatorname{diag}(\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_s)$.

(5) 对于任意的排列 $(i_1, i_2, \ldots, i_s) \in \mathcal{S}_s$, 我们有 $\operatorname{diag}(\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_s) \sim \operatorname{diag}(\boldsymbol{A}_{i_1}, \boldsymbol{A}_{i_2}, \ldots, \boldsymbol{A}_{i_s})$

(6) 若 $\boldsymbol{A} \sim \boldsymbol{B}$, 且 $\boldsymbol{A}$ 可逆, 则 $\boldsymbol{B}$ 也可逆, 且 $\boldsymbol{A}^{-1} \sim \boldsymbol{B}^{-1}$.

(7) 相似的矩阵有相同的行列式、迹和秩.

证明. (1) 若 $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 则 $\boldsymbol{A}^{\mathsf{T}} = \boldsymbol{T}^{\mathsf{T}}\boldsymbol{B}^{\mathsf{T}}(\boldsymbol{T}^{\mathsf{T}})^{-1}$, 从而 $\boldsymbol{A}^{\mathsf{T}} \sim \boldsymbol{B}^{\mathsf{T}}$.

(2) 若 $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 则对任意的正整数 m, 有

$$\boldsymbol{A}^m = \underbrace{(\boldsymbol{T}^{-1}\boldsymbol{B}\underbrace{\boldsymbol{T})(\boldsymbol{T}^{-1}}_{\text{相乘抵消}}\boldsymbol{B}\boldsymbol{T})\cdots(\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T})}_{m\text{组}} = \boldsymbol{T}^{-1}\boldsymbol{B}^m\boldsymbol{T},$$

(3) 若 $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 而 $f(x) = \sum_{i=0}^{m} a_i x^i$, 那么

$$\begin{aligned} f(\boldsymbol{A}) &= \sum_{i=0}^{m} a_i \boldsymbol{A}^i = \sum_{i=0}^{m} a_i (\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T})^m = \sum_{i=0}^{m} a_i \boldsymbol{T}^{-1}\boldsymbol{B}^m\boldsymbol{T} \\ &= \boldsymbol{T}^{-1}\left(\sum_{i=0}^{m} a_i \boldsymbol{B}^i\right)\boldsymbol{T} = \boldsymbol{T}^{-1} f(\boldsymbol{B}) \boldsymbol{T}. \end{aligned}$$

(4) 若 $\boldsymbol{A}_i=\boldsymbol{T}_i^{-1}\boldsymbol{B}_i\boldsymbol{T}_i$, $i=1,2,\ldots,s$, 那么

$$\begin{aligned}\boldsymbol{A}&=\operatorname{diag}(\boldsymbol{A}_1,\ldots,\boldsymbol{A}_s)=\operatorname{diag}(\boldsymbol{T}_1^{-1}\boldsymbol{B}_1\boldsymbol{T}_1,\ldots,\boldsymbol{T}_s^{-1}\boldsymbol{B}_s\boldsymbol{T}_s)\\&=\operatorname{diag}(\boldsymbol{T}_1^{-1},\ldots,\boldsymbol{T}_s^{-1})\operatorname{diag}(\boldsymbol{B}_1,\ldots,\boldsymbol{B}_s)\operatorname{diag}(\boldsymbol{T}_1,\ldots,\boldsymbol{T}_s)\\&=\underbrace{\operatorname{diag}(\boldsymbol{T}_1,\ldots,\boldsymbol{T}_s)^{-1}}_{\boldsymbol{T}^{-1}}\operatorname{diag}(\boldsymbol{B}_1,\ldots,\boldsymbol{B}_s)\underbrace{\operatorname{diag}(\boldsymbol{T}_1,\ldots,\boldsymbol{T}_s)}_{\text{记作 }\boldsymbol{T}}.\end{aligned}$$

(5) 由于任何一个排列都可以通过交换相邻位置, 化为顺序排列, 通过化归的方法, 这儿我们只需考虑 $s=2$ 且 $(i_1,i_2)=(2,1)$ 的情形. 此时, 设 $\boldsymbol{A}_1$ 与 $\boldsymbol{A}_2$ 分别为 n_1 阶方阵和 n_2 阶方阵, 并令 $\boldsymbol{T}=\begin{pmatrix}\boldsymbol{O}&\boldsymbol{I}_{n_1}\\\boldsymbol{I}_{n_2}&\boldsymbol{O}\end{pmatrix}$. 不难验证, $\boldsymbol{T}$ 可逆, 满足 $\boldsymbol{T}^{-1}=\begin{pmatrix}\boldsymbol{O}&\boldsymbol{I}_{n_2}\\\boldsymbol{I}_{n_1}&\boldsymbol{O}\end{pmatrix}$, 并且有 $\boldsymbol{T}\operatorname{diag}(\boldsymbol{A}_1,\boldsymbol{A}_2)\boldsymbol{T}^{-1}=\operatorname{diag}(\boldsymbol{A}_2,\boldsymbol{A}_1)$.

这儿的证明与思路请用分块矩阵的初等变换的方法来理解

(6) 若 $\boldsymbol{A}=\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 且 $\boldsymbol{A}$ 可逆, 那么 $\boldsymbol{B}=\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1}$ 是可逆矩阵的乘积, 从而也可逆, 并有 $\boldsymbol{A}^{-1}=\boldsymbol{T}^{-1}\boldsymbol{B}^{-1}\boldsymbol{T}$.

(7) 若 $\boldsymbol{A}=\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 则

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

以及

$$\operatorname{tr}(\boldsymbol{B})=\operatorname{tr}((\boldsymbol{T}^{-1}\boldsymbol{A})\boldsymbol{T})=\operatorname{tr}(\boldsymbol{T}(\boldsymbol{T}^{-1}\boldsymbol{A}))=\operatorname{tr}(\boldsymbol{A}),$$

即 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 有相同的行列式与迹. 又由于 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相抵等价, 从而有相同的秩. □

例 6.2.15. *如果方阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^2=\boldsymbol{A}$, 则称 $\boldsymbol{A}$ 为一个***幂等矩阵***. 证明: 与幂等矩阵相似的矩阵仍然是幂等矩阵.*

证明. 只需验证: 对于任意的可逆矩阵 $\boldsymbol{T}$ 有 $(\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T})^2=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 当且仅当 $\boldsymbol{A}^2=\boldsymbol{A}$. 而这是显然的. □

例 6.2.16. *如果方阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^2=\boldsymbol{I}$, 则称 $\boldsymbol{A}$ 为一个***对合矩阵***. 证明: 与对合矩阵相似的矩阵仍然是对合矩阵.*

证明. 只需验证: 对于任意的可逆矩阵 $\boldsymbol{T}$ 有 $(\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T})^2=\boldsymbol{I}$ 当且仅当 $\boldsymbol{A}^2=\boldsymbol{I}$. 而这是显然的. □

例 6.2.17. 如果方阵 $\boldsymbol{A}$ 的某个正整数次幂对于零矩阵, 则称 $\boldsymbol{A}$ 为一个**幂零矩阵**. 使得 $\boldsymbol{A}^s = \boldsymbol{O}$ 成立的最小正整数 s 称为 $\boldsymbol{A}$ 的**幂零指数**. 证明: 与幂零矩阵相似的矩阵仍是幂零矩阵, 并且它们的幂零指数相等.

证明. 只需验证: 对于任意的可逆矩阵 $\boldsymbol{T}$ 和任意的正整数 s, 有 $(\boldsymbol{T}^{-1}\boldsymbol{A}^s\boldsymbol{T})^s = \boldsymbol{O}$ 当且仅当 $\boldsymbol{A}^s = \boldsymbol{O}$. 而这是显然的. □

例 6.2.18. 证明: 如果 n 阶实方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 满足 $\boldsymbol{AB} - \boldsymbol{BA} = \boldsymbol{A}$, 则 $\boldsymbol{A}$ 不可逆.

证明. 用反证法, 反设 $\boldsymbol{A}$ 可逆, 对等式左乘 $\boldsymbol{A}^{-1}$ 后, 我们得到 $\boldsymbol{B} - \boldsymbol{A}^{-1}\boldsymbol{BA} = \boldsymbol{I}$. 对该等式取矩阵的迹, 我们有 $\operatorname{tr}(\boldsymbol{B} - \boldsymbol{A}^{-1}\boldsymbol{BA}) = \operatorname{tr}(\boldsymbol{I})$, 即 $0 = n$. 这是一个矛盾. □

注 6.2.19. (1) 若同阶方阵满足 $\boldsymbol{A}_1 \sim \boldsymbol{A}_2$, $\boldsymbol{B}_1 \sim \boldsymbol{B}_2$, 我们一般情况下是无法推出 $(\boldsymbol{A}_1 + \boldsymbol{B}_1) \sim (\boldsymbol{A}_2 + \boldsymbol{B}_2)$ 的, 也无法推出 $(\boldsymbol{A}_1\boldsymbol{B}_1) \sim (\boldsymbol{A}_2\boldsymbol{B}_2)$. 为了说明这一点, 可以利用接下来要提到的相似不变量. 留作习题

(2) 在第四章中我们见到, 一般矩阵的相抵关系也是等价关系. 相抵的同型矩阵构成一个等价类, 我们称之为相抵等价类. 同型的矩阵相抵当且仅当它们有相同的相抵标准形, 当且仅当它们有相等的秩. 秩被称为一个**相抵不变量**, 即, 相抵的矩阵通过计算都会得到的相等的量.

(3) 相似的同阶方阵也构成一个等价类, 称为**相似等价类**. 对于相似等价类, 我们也可以讨论相应的**相似标准形** (请参考教材 *§6.5 若尔当标准形这一小节的内容*) 和**相似不变量** (即相似的矩阵通过计算都会得到的相等的量).

(4) 谈相抵矩阵时, 不要求矩阵为方阵. 谈相似矩阵时, 必须要求为方阵.

(5) 在前面的讨论中, 我们已经见到了行列式、迹和秩这三个常见的相似不变量. 由此很容易看出, 虽然相似的方阵必然相抵, 但是反过来一般不对.

(6) 用定义容易验证, 与单位矩阵 $\boldsymbol{I}_n$ 相似的矩阵仅有 $\boldsymbol{I}_n$ 自身. 这说明 $\boldsymbol{I}_n$ 所在的相似等价类仅有 $\boldsymbol{I}_n$ 一个元素. 更一般地, 不难验证, 若一个相似等价类仅有一个元素, 则它必为某个数量阵 (形如 $a\boldsymbol{I}_n$) 的相似等价类 (其证明与教材 $P114$ 作业题 $\#10$ 是类似的*). 除此之外, 一般而言, 相似等价类很大, 包含了很多不同的矩阵. 留作习题

*思路如下: 首先注意到 $\boldsymbol{A} = (a_{ij})_{n\times n}$ 所在的相似等价类仅有 $\boldsymbol{A}$ 的充要条件是 $\boldsymbol{A}$ 与任何可逆同阶方阵 $\boldsymbol{T}$ 乘法可交换. 接下来, 对于 $1 \le i \ne j \le n$, 考虑 $\boldsymbol{T} = \boldsymbol{I} + \boldsymbol{E}_{ij}$, 利用乘法可交换这一条件看出 $a_{ii} = a_{jj}$, $a_{ki} = 0\ (k \ne i)$, $a_{jk} = 0\ (k \ne j)$.

基于上面的讨论, 当我们谈方阵 $\boldsymbol{A}$ 可能的相似标准形, 一般不能指望其为数量阵, 那么退而求其次, 我们希望它能为对角阵.

定义 6.2.20. 给定一个方阵 $\boldsymbol{A}$, 若存在可逆阵 $\boldsymbol{T}$, 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 是对角阵, 则称 $\boldsymbol{A}$ 为**可相似对角化的**, 或简称为**可对角化的** (*diagonalizable*), 而称 $\boldsymbol{T}$ 将 $\boldsymbol{A}$ **对角化** (*diagonalize*).

注 6.2.21. 并不是每个相似等价类都有对角阵, 即, 不是每个方阵都可以对角化. 例如, 我们之后将要在例 *6.4.1* 中证明 $\begin{pmatrix}2&1&0\\0&2&1\\0&0&2\end{pmatrix}$ 不可对角化.

注 6.2.22. 若 $f(x)$ 为一个一元多项式, 且 $\boldsymbol{A}=\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 在前面我们已经看到了 $f(\boldsymbol{A})=\boldsymbol{T}^{-1}f(\boldsymbol{B})\boldsymbol{T}$. 而若 $\boldsymbol{B}=\mathrm{diag}(\mu_1,\ldots,\mu_n)$ 为对角阵, 则 $f(\boldsymbol{B})=\mathrm{diag}(f(\mu_1),\ldots,f(\mu_n))$, 其计算非常简单, 从而 $f(\boldsymbol{A})$ 的计算也会相对比较简单. 这给出了我们关心可对角化矩阵的一个理由.

例 6.2.23. 在教材例 *4.2.6* 中, 我们考虑了 *Fibonacci* 数列, $F_1=F_2=1$, $F_n=F_{n-1}+F_{n-2}$. 我们推出了递推公式

$$F_n=(1,0)\begin{pmatrix}1&1\\1&0\end{pmatrix}^{n-2}\begin{pmatrix}1\\1\end{pmatrix}.$$

这儿我们需要计算矩阵 $\boldsymbol{A}=\begin{pmatrix}1&1\\1&0\end{pmatrix}$ 的高阶幂. 显然, 直接计算是比较复杂的. 但是我们可以找到可逆矩阵 $\boldsymbol{T}=\begin{pmatrix}\frac{1+\sqrt5}{2}&\frac{1-\sqrt5}{2}\\1&1\end{pmatrix}$, 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\boldsymbol{D}=\begin{pmatrix}\frac{1+\sqrt5}{2}&\\&\frac{1-\sqrt5}{2}\end{pmatrix}$. 从而

$$\begin{aligned}F_n=&(1,0)\left(\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^{-1}\right)^{n-2}\begin{pmatrix}1\\1\end{pmatrix}=(1,0)\boldsymbol{T}\boldsymbol{D}^{n-2}\boldsymbol{T}^{-1}\begin{pmatrix}1\\1\end{pmatrix}\\=&(1,0)\begin{pmatrix}\frac{1+\sqrt5}{2}&\frac{1-\sqrt5}{2}\\1&1\end{pmatrix}\begin{pmatrix}\left(\frac{1+\sqrt5}{2}\right)^{n-2}&\\&\left(\frac{1-\sqrt5}{2}\right)^{n-2}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt5}&\frac{\sqrt5-1}{2\sqrt5}\\-\frac{1}{\sqrt5}&\frac{\sqrt5+1}{2\sqrt5}\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}\\=&\frac{\left(\frac{1+\sqrt5}{2}\right)^n-\left(\frac{1-\sqrt5}{2}\right)^n}{\sqrt5}.\end{aligned}$$

在之后我们会系统地讨论这样的 $\boldsymbol{T}$ 与 $\boldsymbol{D}$ 的找法.

例 6.2.24. 证明*:* 如果 $\boldsymbol{A}$ 可以对角化, 则 $\boldsymbol{A}\sim\boldsymbol{A}^{\mathsf{T}}$. (事实上, 在命题 *6.5.2* 中我们将利用若尔当标准形的理论来证明该结论对于一般的复方阵也成立, 从而不需要加“可对角化”这一条件.)

证明. 若 $\boldsymbol{A}$ 可对角化, 则存在对角阵 $\boldsymbol{D}$ 使得 $\boldsymbol{A}\sim\boldsymbol{D}$, 于是 $\boldsymbol{A}^{\mathsf{T}}\sim\boldsymbol{D}^{\mathsf{T}}$. 但是, $\boldsymbol{D}^{\mathsf{T}}=\boldsymbol{D}$. 由相似关系的传递性可知, $\boldsymbol{A}\sim\boldsymbol{A}^{\mathsf{T}}$. □

6.3 特征值与特征向量

一个很自然的问题是, 例 6.2.23 中的 $\boldsymbol{T}$ 与 $\boldsymbol{D}$ 是如何找到的? 我们可以换一个角度来看这个问题. 若记 $\boldsymbol{T}=(\boldsymbol{x}_1,\boldsymbol{x}_2)$, $\boldsymbol{D}=\operatorname{diag}(\lambda_1,\lambda_2)$, 则 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\boldsymbol{D}$ 当且仅当 $\boldsymbol{A}\boldsymbol{T}=\boldsymbol{T}\boldsymbol{D}$, 当且仅当 $\boldsymbol{A}(\boldsymbol{x}_1,\boldsymbol{x}_2)=(\lambda_1\boldsymbol{x}_1,\lambda_2\boldsymbol{x}_2)$. 从而, 这样的 $\boldsymbol{T}$ 与 $\boldsymbol{D}$ 若存在, 必满足如下的两条.

(1) 由于 $\boldsymbol{T}$ 可逆, 列向量 $\boldsymbol{x}_1,\boldsymbol{x}_2$ 线性无关. 特别地, 它们都是非零向量.

(2) 解矩阵方程组, 我们有 $\boldsymbol{A}\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i$ $(i=1,2)$.

由上面的讨论, 我们引入特征值与特征向量的概念.

特征值与特征向量的定义

定义 6.3.1. *设 $\boldsymbol{A}$ 为数域 F 上的 n 阶方阵. 如果存在 $\lambda\in F$ 和非零列向量 $\boldsymbol{x}\in F^n$ 使得 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$, 则称 λ 为 $\boldsymbol{A}$ 的*一个**特征值** *(eigenvalue, characteristic value), 并称 $\boldsymbol{x}$ 为***属于 λ 的一个特征向量** *(eigenvector, characteristic vector).*

注 6.3.2. *(1) 若 $\boldsymbol{x}$ 是矩阵 $\boldsymbol{A}$ 关于 λ 的特征向量, 则 $\boldsymbol{A}\boldsymbol{x}$ 是与 $\boldsymbol{x}$ 平行的向量, λ 是拉伸的系数.*

(2) 由前面例子的引入, 我们知道特征值是相似对角阵 (如果存在) 的主对角线上的元素.

(3) 特征值 λ 作为一个标量并不一定非零.

习题 6.3.3. *(1) 设 $\boldsymbol{u}=(3,1)^{\mathsf{T}}$ 和 $\boldsymbol{v}=(1,2)^{\mathsf{T}}$ 是 $\mathbb{R}^2$ 中的向量, 并设 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 是一个 2 阶方阵 $\boldsymbol{A}$ 分别相应于特征值 2 和 3 的特征向量. 设 $\mathscr{A}:\mathbb{R}^2\to\mathbb{R}^2$ 是线性变换, 由 $\boldsymbol{x}\mapsto\boldsymbol{A}\boldsymbol{x}$ 给出. 记 $\boldsymbol{w}=\boldsymbol{u}+\boldsymbol{v}$. 画出 $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$, $\mathscr{A}(\boldsymbol{u})$, $\mathscr{A}(\boldsymbol{v})$ 和 $\mathscr{A}(\boldsymbol{w})$.*

(2) 设 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 是 $\boldsymbol{A}$ 分别对应于特征值 -1 和 3 的特征向量, 重新做上面的小题.

例 6.3.4. *在例 6.2.23 中, 由*

$$\begin{pmatrix}1&1\\1&0\end{pmatrix}\begin{pmatrix}\frac{1+\sqrt5}{2}&\frac{1-\sqrt5}{2}\\1&1\end{pmatrix}=\begin{pmatrix}\frac{1+\sqrt5}{2}&\frac{1-\sqrt5}{2}\\1&1\end{pmatrix}\begin{pmatrix}\frac{1+\sqrt5}{2}&\\&\frac{1-\sqrt5}{2}\end{pmatrix}$$

可知，$\boldsymbol{x}_1 = \left(\frac{1+\sqrt{5}}{2}, 1\right)^{\mathsf{T}}$ 是方阵 $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ 相对于特征值 $\lambda_1 = \frac{1+\sqrt{5}}{2}$ 的一个特征向量，$\boldsymbol{x}_2 = \left(\frac{1-\sqrt{5}}{2}, 1\right)^{\mathsf{T}}$ 是方阵 $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ 相对于特征值 $\lambda_2 = \frac{1-\sqrt{5}}{2}$ 的一个特征向量.

考察集合 $V_{\boldsymbol{A}}(\lambda) := \{\, \boldsymbol{x} \in F^n \mid \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \,\}$. 它是齐次方程组 $(\lambda \boldsymbol{I}_n - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ 的解空间, 从而构成 F^n 的一个子空间. 由定义 6.3.1 知, λ 是 $\boldsymbol{A}$ 的一个特征值当且仅当 $V_{\boldsymbol{A}}(\lambda)$ 存在非零向量. 若该条件满足, 我们称 $V_{\boldsymbol{A}}(\lambda)$ 为**属于 λ 的特征子空间** (eigenspace, characteristic space). 一个向量是属于 λ 的特征向量当且仅当它是 $V_{\boldsymbol{A}}(\lambda)$ 中的一个非零向量.

类似地, 我们可以讨论线性变换的特征值与特征向量的问题.

定义 6.3.5. 设 F 为数域, V 为 F 上的线性空间, $\mathscr{A}$ 为 V 上的线性变换. 若存在 $\lambda \in F$ 以及非零向量 $\boldsymbol{x} \in V$, 满足 $\mathscr{A}(\boldsymbol{x}) = \lambda \boldsymbol{x}$, 则称 λ 为 $\mathscr{A}$ 的**一个特征值**, 并称 $\boldsymbol{x}$ 为**属于 λ 的一个特征向量**.

类似地, 我们可以考察集合 $V_{\mathscr{A}}(\lambda) := \{\, \boldsymbol{x} \in V \mid \mathscr{A}(\boldsymbol{x}) = \lambda \boldsymbol{x} \,\}$. 由于 $\mathscr{A}$ 为线性变换, $\boldsymbol{0} \in V_\lambda$, 从而该集合非空. 容易用定义验证, $V_{\mathscr{A}}(\lambda)$ 对于向量的加法与数乘运算封闭, 从而构成 V 的一个子空间. 由定义 6.3.5 知, λ 是 $\mathscr{A}$ 的一个特征值当且仅当 $V_{\mathscr{A}}(\lambda)$ 存在非零向量. 若该条件满足, 我们称 $V_{\mathscr{A}}(\lambda)$ 为**属于 λ 的特征子空间**. 一个向量是属于 λ 的特征向量当且仅当它是 $V_{\mathscr{A}}(\lambda)$ 中的一个非零向量.

例 6.3.6. 若 $\mathscr{A}$ 为恒等变换, 则 $\mathscr{A}$ 仅有特征值 1, 特征子空间 $V_{\mathscr{A}}(1) = V$ 为全空间, 即, V 中任意的非零向量都是属于 1 的特征向量.

习题 6.3.7. 求矩阵 $\boldsymbol{A} = \begin{pmatrix} 4 & 2 & 3 & 3 \\ 0 & 2 & h & 3 \\ 0 & 0 & 4 & 14 \\ 0 & 0 & 0 & 2 \end{pmatrix}$ 中的 h, 使得 $\lambda = 5$ 的特征子空间的维数为 2.

若 $\mathscr{A}$ 是数组空间 F^n 上的一个线性变换, 在自然基 $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ 下的矩阵为 $\boldsymbol{A}$, 则容易验证, $\boldsymbol{A}$ 与 $\mathscr{A}$ 的特征值和特征向量完全吻合. 关于这一点的更进一步的讨论, 我们有如下的命题.

命题 6.3.8. 若 $\mathscr{A}$ 是线性空间 V 上的一个线性变换, 在基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{A}$. 那么, 我们有以下几条. 教材命题 6.3.1

(1) $\boldsymbol{A}$ 与 $\mathscr{A}$ 具有相同的特征值.

(2) 设 λ 是 $\mathscr{A}$ 的一个特征值, 则 $\boldsymbol{x}$ 是 $\mathscr{A}$ 的关于 λ 的特征向量的充要条件是 $\boldsymbol{x}$ 在基 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 下的坐标列向量 $\boldsymbol{X}$ 是 $\boldsymbol{A}$ 关于 λ 的特征向量. 从而,

$$V_{\mathscr{A}}(\lambda)=\left\{\,\boldsymbol{x}\in V \mid \boldsymbol{x} \text{ 在基 } \boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n \text{ 下的坐标 } \boldsymbol{X} \text{ 属于 } V_{\boldsymbol{A}}(\lambda)\,\right\}.$$

证明. 设 $\lambda\in F$, $\boldsymbol{x}\in V$ 在 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 下的坐标为 $\boldsymbol{X}=(x_1,\ldots,x_n)^{\mathsf{T}}$, 即, $\boldsymbol{x}=(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n)\boldsymbol{X}=\sum_i x_i\boldsymbol{\alpha}_i$. 此时, 我们有

$$\mathscr{A}(\boldsymbol{x})=\mathscr{A}\left(\sum_i x_i\boldsymbol{\alpha}_i\right)=\sum_i x_i\mathscr{A}(\boldsymbol{\alpha}_i)=(\mathscr{A}(\boldsymbol{\alpha}_1),\ldots,\mathscr{A}(\boldsymbol{\alpha}_n))\boldsymbol{X}=(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n)\boldsymbol{A}\boldsymbol{X},$$

以及

$$\lambda\boldsymbol{x}=\lambda\left((\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n)\boldsymbol{X}\right)=(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n)\lambda\boldsymbol{X}.$$

由坐标表示的唯一性可知 $\boldsymbol{A}\boldsymbol{X}=\lambda\boldsymbol{X}$ 的充要条件是 $\mathscr{A}(\boldsymbol{x})=\lambda\boldsymbol{x}$. 另一方面, $\boldsymbol{x}$ 是零向量当且仅当 $\boldsymbol{X}$ 是零向量. 从而, 命题的各条容易验证. □

注 6.3.9. *线性变换的特征值与特征向量的求解问题, 一般都是利用命题 6.3.8, 化成在某组基下的矩阵的相应问题的求解. 关于这方面具体的例子, 见教材 P175 的例 6.3.6. 学生课后自习.*

习题 6.3.10. *在以下题目中, 设 $\boldsymbol{A}$ 是线性变换 $\mathscr{A}$ 在自然基下的矩阵. 不写出 $\boldsymbol{A}$, 直接描述 (求出) $\boldsymbol{A}$ 的特征值与特征子空间.*

(1) $\mathscr{A}$ 是 $\mathbb{R}^2$ 的相对于某条过原点的直线的反射变换 (镜像映射).

(2) $\mathscr{A}$ 是 $\mathbb{R}^3$ 的相对于某条过原点的直线的旋转变换.

特征值与特征向量的计算 设 $\boldsymbol{A}=(a_{ij})$ 为数域 F 上的一个 n 阶方阵, 则

$$\begin{aligned}\lambda\in F \text{ 为 } \boldsymbol{A} \text{ 的特征值} &\Leftrightarrow V_A(\lambda) \text{ 有非零向量} \Leftrightarrow (\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0} \text{ 有非零解}\\ &\Leftrightarrow \lambda\boldsymbol{I}-\boldsymbol{A} \text{ 不可逆} \Leftrightarrow \text{行列式 } |\lambda\boldsymbol{I}-\boldsymbol{A}|=0.\end{aligned}$$

我们可以观察到

$$|\lambda\boldsymbol{I}-\boldsymbol{A}|=\begin{vmatrix}\lambda-a_{11} & -a_{12} & \cdots & -a_{1n}\\ -a_{21} & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & -a_{n-1,n}\\ -a_{n1} & \cdots & -a_{n,n-1} & \lambda-a_{nn}\end{vmatrix}$$

$$= \lambda^n + \text{中间次数的复杂项} + (-1)^n \det(\boldsymbol{A}).$$

最后的这个多项式是关于未知元 λ 的 F 上的 n 次多项式, 其中间次数的系数较为复杂, 我们稍后会作进一步解释. 该多项式将被记作 $p_{\boldsymbol{A}}(\lambda)$, 称为 $\boldsymbol{A}$ 的**特征多项式**[*] (characteristic polynomial). 相应地, 方程 $p_{\boldsymbol{A}}(\lambda) = 0$ 称为矩阵 $\boldsymbol{A}$ 的**特征方程** (characteristic equation). 具体取值 $\lambda_0 \in F$ 为 $\boldsymbol{A}$ 的特征值, 当且仅当 λ_0 是 $p_{\boldsymbol{A}}(\lambda) = 0$ 的根. 由域上的多项式的理论, 我们知 $\boldsymbol{A}$ 的特征值最多有 n 个. 特别地, $\boldsymbol{A}$ 的特征值的个数有限.

注 6.3.11. 类似的讨论可以说明, $|\lambda \boldsymbol{I} + \boldsymbol{A}|$ 也是关于未知元 λ 的 F 上的 n 次多项式, 从而仅有有限多个 $\lambda \in F$ 使得 $\lambda \boldsymbol{I} + \boldsymbol{A}$ 不可逆. 这是 "微小摄动法" 的理论前提.

例 6.3.12. (1) 对于任意的 2 阶实对称阵 $\boldsymbol{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, 其特征多项式为 $p_{\boldsymbol{A}}(\lambda) = \lambda^2 - (a+c)\lambda + (ac - b^2)$, 从而特征值为 $\lambda_{1,2} = \dfrac{a + c \pm \sqrt{(a-c)^2 + 4b^2}}{2}$, 都是实数.

(2) 对于任意的 2 阶反对称实矩阵 $\boldsymbol{A} = \begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix}$, 其特征多项式为 $p_{\boldsymbol{A}}(\lambda) = \lambda^2 + b^2$, 从而特征值为 $\lambda_{1,2} = \pm b\mathrm{i}$, 都是纯虚数或 0.

例 6.3.13. 设 $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1s} \\ \boldsymbol{O} & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \boldsymbol{O} & \cdots \boldsymbol{O} & \boldsymbol{A}_{ss} \end{pmatrix}$ 为准上三角的分块矩阵, 其中主对角线上的子块 $\boldsymbol{A}_{11}, \boldsymbol{A}_{22}, \ldots, \boldsymbol{A}_{ss}$ 都是方阵, 那么由定义不难验证, 特征多项式

$$p_{\boldsymbol{A}}(\lambda) = p_{\boldsymbol{A}_{11}}(\lambda) p_{\boldsymbol{A}_{22}}(\lambda) \cdots p_{\boldsymbol{A}_{ss}}(\lambda).$$

特别地, 若 $\boldsymbol{A} = (a_{ij})$ 是上三角的方阵, 则 $p_{\boldsymbol{A}}(\lambda) = (\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn})$. 当然, 若 $\boldsymbol{A}$ 是 (准) 下三角的方阵时, 讨论也是类似的.

注 6.3.14. 我们观察到多项式 $\lambda^2 + 1$ 在 $\mathbb{R}$ 上无根, 在 $\mathbb{C}$ 上有一对不同的根. 故, 考虑特征值时, 我们更加倾向于在 $\mathbb{C}$ 上考察.

问题: 如何在 $\mathbb{C}$ 上求一个 n 阶方阵 $\boldsymbol{A}$ 的特征值和相应的特征向量?

[*] 有的教材和数学软件将 $|\boldsymbol{A} - \lambda \boldsymbol{I}|$ 称为特征多项式. 这个变动不是本质的.