

§1.2 线性变换及其矩阵表示

定义: T 称为由 V^n 到 V^m 的**变换 (映射)**, 如果 T 将 V^n 中的向量映射到 V^m 中的向量, 写作

$$T: \alpha \in V^n \rightarrow \beta = T\alpha \in V^m$$

其中 β 为 α 在 T 下的**像**, α 称为 β 的**原像**.

变换的例子

- $T: \alpha = (x_1, x_2)^T \in \mathbb{R}^2 \rightarrow \beta = (0, x_2)^T \in \mathbb{R}^2$
- $T: \alpha = (x_1, x_2)^T \in \mathbb{R}^2 \rightarrow \beta = (2x_1, 2x_2)^T \in \mathbb{R}^2$
- $T: \alpha = p(t) \in \mathbf{P}(t) \rightarrow \beta = (p(t))^2 \in \mathbf{P}(t)$

线性变换

定义: T 为由 V^n 到 V^m 的变换, 如果对于任意的 $k \in F, \alpha, \beta \in V^n$, 都有

$$T(\alpha + \beta) = T\alpha + T\beta, \quad T(k\alpha) = kT\alpha,$$

即变换 T 与线性运算可交换, 则 T 是**线性变换**.

特别的, 当 T 是 V^n 到自身的一个线性变换, 则称 T 是 V^n 的**线性变换**.

例 1. 给定 $A \in F^{m \times n}$, 定义由 V^n 到 V^m 的变换 T 为

$$T: x \in F^n \rightarrow y = Ax \in F^m.$$

由矩阵运算的性质得知, 易证 T 是一个线性变换.

例 2. 给定 $P \in F^{m \times m}$ 和 $Q \in F^{n \times n}$, 定义由 V^n 到 V^m 的变换 T 为

$$T: X \in F^{m \times n} \rightarrow Y = PXQ \in F^{m \times n}.$$

由矩阵运算的性质得知, 易证 T 是一个线性变换.

例 3. 对于 $\mathbf{P}_n(t)$ 中的多项式求导运算 $\frac{d}{dt}$, 记为 D , 即

$$Dp(t) = \frac{d}{dt}p(t), \quad p(t) \in \mathbf{P}_n(t)$$

因求导运算和线性运算可以交换顺序, 可知 D 是 $\mathbf{P}_n(t)$ 的一个线性变换.

例 4. V 的

- 恒等变换 $I: I\alpha = \alpha \forall \alpha \in V$
- 零变换 $O: O\alpha = 0, \forall \alpha \in V$

都是 V 的线性变换.

线性变换的一些简单性质:

- (1). $O\alpha = 0, \forall \alpha \in V$.
 $T0 = 0, T(-\alpha) = -T\alpha, \forall \alpha \in V$.
- (2). $T(\sum_{i=1}^r k_i \alpha_i) = \sum_{i=1}^r k_i T\alpha_i$, 即任意一组向量的线性组合的像, 等于它们的像的线性组合.
- (3). 一组线性相关的向量 $\alpha_1, \alpha_2, \dots, \alpha_r$, 它们在 T 下的像 $T\alpha_1, T\alpha_2, \dots, T\alpha_r$ 也线性相关.
注意: 线性无关的一组向量在 T 下的像可能是线性相关的, 例如零变换把线性无关的向量都映射为零向量.

线性变换的矩阵表示

设 T 是 V^n 到 V^m 的线性变换, 在 V^n 和 V^m 中分别取基 $\mathcal{B}_\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$

和 $\mathcal{B}_\beta = \{\beta_1, \beta_2, \dots, \beta_m\}$, 则 α_j 的像 $T\alpha_j$ ($1 \leq j \leq n$) 可由基 \mathcal{B}_β 唯一线性表出:

$$T\alpha_j = \sum_{i=1}^m \beta_i a_{ij} = (\beta_1, \beta_2, \dots, \beta_m) \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

将 $T\alpha_j$ 按 $j = 1, 2, \dots, n$ 的顺序排列, 则有

$$(T\alpha_1, \dots, T\alpha_n) = (\beta_1, \dots, \beta_m) \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

令 $T\mathcal{B}_\alpha := (T\alpha_1, T\alpha_2, \dots, T\alpha_n)$, 则上式可简写为

$$T\mathcal{B}_\alpha = \mathcal{B}_\beta A$$

则这称为**线性变换 T 的矩阵表示**, 其中 A 称为 T 在基偶 $\{\mathcal{B}_\alpha, \mathcal{B}_\beta\}$ 下的矩阵.

特别的, 若 $V^n = V^m$ 且 $\mathcal{B}_\alpha = \mathcal{B}_\beta$, 则 $T\mathcal{B}_\alpha = \mathcal{B}_\alpha A$, 此时称 n 阶方阵 A 为 T 在基 \mathcal{B}_α 下的矩阵.

例 5. 求 $\mathbf{P}_n(t)$ 的线性变换 $D = \frac{d}{dt}$ 在基 $\mathcal{B} = \{1, t, \dots, t^n\}$ 下的矩阵.

解: 由 $D\mathcal{B} = \mathcal{B}A$, 其中 $D\mathcal{B} = (0, 1, 2t, \dots, nt^{n-1})$,
即 $D\mathcal{B} = (0, 1, 2t, \dots, nt^{n-1}) = (1, t, t^2, \dots, t^n)A$,
由 $jt^{j-1} = \sum_{i=1}^n a_{ij}t^i$, 知 $a_{j-1,j} = j$ 而其他 a_{ij} 为零.

可得 $A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & n \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$

例 6 求 $\mathbf{P}_2(t)$ 到 $\mathbf{P}_3(t)$ 的线性变换 J

$$J(p(t)) = \int_0^t p(t) dt$$

基偶 $\{\mathcal{B}_1 = \{1, t, t^2\}, \mathcal{B}_2 = \{1, t, \frac{t^2}{2}, \frac{t^3}{3}\}\}$ 下的矩阵.

解: 由 $J\mathcal{B}_1 = \mathcal{B}_2 A$, 其中 $J\mathcal{B}_1 = (t, \frac{t^2}{2}, \frac{t^3}{3})$, 即

$$J\mathcal{B}_1 = (t, \frac{t^2}{2}, \frac{t^3}{3}) = (1, t, \frac{t^2}{2}, \frac{t^3}{3}) A,$$

可得 $A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$

由 T 的矩阵表示, 确定 T 的像的坐标:

设 $\alpha \in V^n$, 则可由 B_α 线性表出,

$$\alpha = \sum_{i=1}^n x_i \alpha_i = B_\alpha x$$

即 α 在 B_α 下的坐标为 x . 对于线性变换 T ,

$$T\alpha = \sum_{i=1}^n x_i T\alpha_i = TB_\alpha x$$

将 TB_α 用 V^n 的基 B_β 表出, 若 $TB_\alpha = B_\beta A$, 则

$$T\alpha = B_\beta Ax,$$

即 $T\alpha$ 在 B_β 下的坐标为 Ax .

定理: 设 $B_\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 和 $B_\beta = \{\beta_1, \beta_2, \dots, \beta_n\}$ 分别是 V^n 和 V^m 的基, 对于给定的 $m \times n$ 矩阵 $A = [a_{ij}]$, 则存在唯一的从 V^n 到 V^m 的线性变换 T , 使得它在 $\{B_\alpha, B_\beta\}$ 下的矩阵是 A , 即 $TB_\alpha = B_\beta A$.

证明: 先证明线性变换 T 的存在性.

对于任意 $\alpha \in V^n$, 设其在基 B_α 下的坐标为 x , 即 $\alpha = B_\alpha x$. 现定义变换 T 将 α 映为 $B_\beta Ax$, 即

$$T: \alpha = B_\alpha x \rightarrow \beta = B_\beta Ax$$

我们仍需验证 T 是线性变换.

证明 (续): 验证 T 是线性变换.

设 V^n 中的 $\alpha = \mathcal{B}_\alpha x$ 和 $\gamma = \mathcal{B}_\alpha z$, 而 $k, l \in F$, 则

$$\begin{aligned} T(k\alpha + l\gamma) &= T(k\mathcal{B}_\alpha x + l\mathcal{B}_\alpha z) \\ &= T(\mathcal{B}_\alpha(kx + lz)) \\ &= \mathcal{B}_\beta A(kx + lz) \\ &= k\mathcal{B}_\beta Ax + l\mathcal{B}_\beta Az \\ &= kT(\mathcal{B}_\alpha x) + lT(\mathcal{B}_\alpha z) \\ &= kT\alpha + lT\gamma \end{aligned}$$

则 T 是线性变换, 且由 T 定义易得 $T\mathcal{B}_\alpha = \mathcal{B}_\beta A$.

证明 (续): 再证明线性变换 T 的唯一性.

如果存在两个线性变换 T_1 和 T_2 , 满足

$$T_1 \mathcal{B}_\alpha = T_2 \mathcal{B}_\alpha = \mathcal{B}_\beta A.$$

那么线性变换 $\tilde{T} = T_1 - T_2$ 满足对于某个 \tilde{A} 有

$$\tilde{T} \mathcal{B}_\alpha = \{0, 0, \dots, 0\} = \mathcal{B}_\beta \tilde{A}$$

这里 \tilde{A} 只能为零, 即 \tilde{T} 将任何向量映为 0. 从而

$$T_1 \alpha = (T_2 + \tilde{T}) \alpha = T_2 \alpha + \tilde{T} \alpha = T_2 \alpha$$

对于任意 α 成立, 因此映射唯一.

V^n 到 V^m 的线性变换 T , 在给定的基偶 $\{B_\alpha, B_\beta\}$ 下, 对应一个矩阵 A . 反之, 对于同样的空间和基偶, 如果给定矩阵 A , 则对应一个线性变换 T .

接下来的问题:

- 如果取 V^n 和 V^m 的另一组基 $\{B_{\alpha'}, B_{\beta'}\}$, 那么 T 对应另外一个矩阵 B , 那么矩阵 A 和 B 之间有什么关系?
- 怎样选取 V^n 和 V^m 的基, 使得 T 的矩阵表示尽可能简单?

设 n 阶方阵 P 是基 \mathcal{B}_α 到 $\mathcal{B}_{\alpha'}$ 的变换矩阵, 而 n 阶方阵 Q 是基 \mathcal{B}_β 到 $\mathcal{B}_{\beta'}$ 的变换矩阵, $m \times n$ 矩阵 A, B 分别是 T 在基偶 $\{\mathcal{B}_\alpha, \mathcal{B}_\beta\}$ 和 $\{\mathcal{B}_{\alpha'}, \mathcal{B}_{\beta'}\}$ 下的矩阵, 那么由关系式

$$\begin{aligned}\mathcal{B}_{\alpha'} &= \mathcal{B}_\alpha P, & \mathcal{B}_{\beta'} &= \mathcal{B}_\beta Q, \\ T\mathcal{B}_\alpha &= \mathcal{B}_\beta A, & T\mathcal{B}_{\alpha'} &= \mathcal{B}_{\beta'} B\end{aligned}$$

可以推出 $\mathcal{B}_\beta AP = T\mathcal{B}_\alpha P = T\mathcal{B}_{\alpha'} = \mathcal{B}_{\beta'} B = \mathcal{B}_\beta QB$

$$\mathcal{B}_\beta (AP - QB) = 0$$

因 \mathcal{B}_β 是基, 则 $AP = QB, A = QBP^{-1}, B = Q^{-1}AP$

如果 $A = QBP^{-1}$ 或 $B = Q^{-1}AP$, 其中 P, Q 为可逆方阵, 那么称 A 和 B 是相抵 (或等价) 的. 如上证明了, 一个 V^n 到 V^m 的线性变换在不同基偶下的矩阵是相抵的.

假如 $V^m = V^n$, $B_\alpha = B_\beta$, $B_{\alpha'} = B_{\beta'}$, 那么 $Q = P$, 则有 $A = PBP^{-1}$. 此时方阵 A 与 B 是相似的. 即一个 V^n 到自身的线性变换在不同基偶下的矩阵是相似的.

之前的第二个问题等价于: 与 A 相抵 (或相似) 的最简单的矩阵是什么?

关于线性代数中矩阵知识的回忆

线性代数主要解决两类问题:

第一类: 给定 A 与 b , 求 x 使得 $Ax = b$.

- 常用思路: 求 A 的逆, 将 A 进行变换等等.
- 可能涉及知识: 矩阵的可逆性、初等变换和秩等等.

第二类: 给定 A , 求 λ 和 x 使得 $Ax = \lambda x$.

- 常用思路: 将 A 进行相似化简等.
- 可能涉及知识: 矩阵的特征值和特征向量, 特征多项式, 相似变换等.

以前线性代数中主要考虑:

- n 维欧氏向量空间, \mathbb{R}^n 和 \mathbb{C}^n 等.
- 矩阵对于向量的作用, 以及矩阵的性质 (如秩、特征值等等)

在这门课程中考虑:

- 一般的线性空间, 除欧氏空间外, 也包括如多项式空间 $\mathbf{P}(t)$ 和 $\mathbf{P}_n(t)$ 等其他众多例子.
- 考虑线性变换的作用, 以及线性变换的性质

一般来说, 关于线性空间和线性变换的讨论, 都可以转化为关于欧式空间 (坐标空间) 和矩阵的讨论.

线性变换的核与值域

定义: 设 T 是从 V^n 到 V^m 的线性变换, 则

$$N(T) := \{\alpha \in V^n \mid T\alpha = 0\}$$

$$R(T) := \{\beta \in V^m \mid \beta = T\alpha, \alpha \in V^n\}$$

分别称为 T 的核和 T 的值域.

- $N(T)$ 是 V^n 的一个子空间, 也被称为 T 的零空间, 其维数称为 T 的零度, 记作 $\text{null } T$;
- $R(T)$ 是 V^m 的一个子空间, 也被称为 T 的值空间, 其维数称为 T 的秩, 记作 $\text{rank } T$.

一个线性变换 T 将 V^n 中的任一向量 α 映射为 V^m 的向量 $T\alpha$. V^n 的每一个向量都有在 T 下的像, 但未必 V^m 的每一个向量都有原像. 这发生在 $\text{rank } T < m$ 的时候, 例如:

- $T: \alpha = (x_1, x_2)^T \in \mathbb{R}^2 \rightarrow \beta = (0, x_2)^T \in \mathbb{R}^2$
- $T: \alpha = p(t) \in \mathbf{P}_n(t) \rightarrow \beta = \frac{d}{dt}p(t) \in \mathbf{P}_n(t)$

但可以通过重新定义合适的空间 $T: V^n \rightarrow R(T)$, 使 $R(T)$ 中每一个向量都有原像

- $T: \alpha = (x_1, x_2)^T \in \mathbb{R}^2 \rightarrow \beta = (0, x_2)^T \in \{0\} \times \mathbb{R}$
- $T: \alpha = p(t) \in \mathbf{P}_n(t) \rightarrow \beta = \frac{d}{dt}p(t) \in \mathbf{P}_{n-1}(t)$

定理: 设 T 是从 V^n 到 V^m 的线性变换, 则

$$\text{null } T + \text{rank } T = n$$

证明: 若 $\text{null } T = k$, 并设 $B_0 = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 是 $N(T)$ 的一组基. 因为 $N(T) \subset V^n$, 所以可以把这 k 个向量扩充为 V^n 的基

$$B_1 = \{\alpha_1, \alpha_2, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n\}.$$

我们将证明 $\{T\alpha_{k+1}, T\alpha_{k+2}, \dots, T\alpha_n\}$ 是 $R(T)$ 的一组基, 从而得到 $\text{rank } T = n - k$.

证明 (续): 任取 $\alpha \in V^n$, 则 α 可由 B_1 线性表出:
 $\alpha = \sum_{i=1}^n b_i \alpha_i$. 又因为 $\alpha_i \in N(T)$ ($1 \leq i \leq k$), 则

$$T\alpha = \sum_{i=1}^n b_i T\alpha_i = \sum_{i=k+1}^n b_i T\alpha_i.$$

即 $R(T)$ 中的向量 $T\alpha$ 可由 $\{T\alpha_{k+1}, T\alpha_{k_2}, \dots, T\alpha_n\}$ 线性表出.

仍需证明 $\{T\alpha_{k+1}, T\alpha_{k_2}, \dots, T\alpha_n\}$ 线性无关.

证明 (续): 假如存在 $c_{k+1}, c_{k_2}, \dots, c_n \in F$ 使得

$$\sum_{i=k+1}^n c_i T\alpha_i = 0.$$

则有 $T(\sum_{i=k+1}^n c_i \alpha_i) = 0$, 从而 $\sum_{i=k+1}^n c_i \alpha_i \in N(T)$.

于是 $\sum_{i=k+1}^n c_i \alpha_i$ 可由 B_0 线性表出, 同时也由 $\alpha_{k+1}, \dots, \alpha_n$ 线性表出. 但 B_0 与 $\alpha_{k+1}, \dots, \alpha_n$ 线性无关, 故 $\sum_{i=k+1}^n c_i \alpha_i = 0$, 即 $c_{k+1} = \dots = c_n = 0$. 于是 $\{T\alpha_{k+1}, T\alpha_{k_2}, \dots, T\alpha_n\}$ 线性无关. 综上, 这是 $R(T)$ 的一组基, 则定理得证.

注:

若 $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 是 $N(T)$ 的一组基, 并被扩充为 V^n 的基 $\mathcal{B}_\alpha = \{\alpha_{k+1}, \alpha_{k+2}, \dots, \alpha_n, \alpha_1, \alpha_2, \dots, \alpha_k\}$.

由定理知 $\{T\alpha_{k+1}, T\alpha_{k+2}, \dots, T\alpha_n\}$ 是 $R(T)$ 的一组基, 也可以被扩充为 V^m 中的一组

基 $\mathcal{B}_\beta = \{T\alpha_{k+1}, T\alpha_{k+2}, \dots, T\alpha_n, \beta_1, \beta_2, \dots, \beta_{k+m-n}\}$.

则 T 在基偶 $\{\mathcal{B}_\alpha, \mathcal{B}_\beta\}$ 下的矩阵是

$$\begin{pmatrix} I_{(n-k)} & O_{(n-k) \times k} \\ O_{(m+k-n) \times (n-k)} & O_{(m+k-n) \times k} \end{pmatrix}.$$

其中 I_{n-k} 为 $(n-k)$ 阶单位阵, O 为零矩阵.

再次回顾例子

- $T: \alpha = (x_1, x_2)^T \in \mathbb{R}^2 \rightarrow \beta = (0, x_2)^T \in \mathbb{R}^2$
- $T: \alpha = p(t) \in \mathbf{P}_n(t) \rightarrow \beta = \frac{d}{dt}p(t) \in \mathbf{P}_n(t)$

第一个例子中 $N(T) = \text{span}\{(1, 0)^T\}$, $\text{null } T = 1$.
对应的 $\text{rank } T = 2 - 1 = 1 < 2$.

第二个例子中 $N(T) = \text{span}\{1\}$, $\text{null } T = 1$. 对应的 $\text{rank } T = (n + 1) - 1 = n < n + 1$.

另, 虽 $\dim N(T) + \dim R(T) = \text{null } T + \text{rank } T = n$
但 $N(T) + R(T)$ 不一定等于 V^n . 比如上面第二个例子.

不变子空间

定义: 设 T 是 V 的线性变换, W 是 V 的子空间, 如果对于任意的 $\alpha \in W$, 都有 $T\alpha \in W$. 则称 W 是 T 的**不变子空间**.

简单例子 (HW)

- T 的核 $N(T)$ 和 T 的值域 $R(T)$ 都是 T 的不变子空间
- T 的不变子空间的交空间与和空间也是 T 的不变子空间.

利用不变子空间简化 T 的矩阵表示

如果 W_i ($1 \leq i \leq s$) 都是 T 的不变子空间, 且有

$$V^n = W_1 \oplus W_2 \oplus \dots \oplus W_s,$$

其中每个 W_i 的维数有 $\dim W_i = n_i$, 则有
 $n_1 + n_2 + \dots + n_s = n$. 现在每个 W_i 中取一个基
 $\{\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in_i}\}$ ($i = 1, 2, \dots, s$), 并把它们顺序排
列为 V^n 的基 $B =$

$$\{\alpha_{11}, \alpha_{12}, \dots, \alpha_{1n_1}, \alpha_{21}, \alpha_{22}, \dots, \alpha_{2n_2}, \dots, \alpha_{s1}, \alpha_{s2}, \dots, \alpha_{sn_s}\},$$

那么 T 在 \mathcal{B} 下的矩阵是**对角块矩阵**:

$$\begin{pmatrix} A_{11} & & & \\ & A_{22} & & \\ & & \dots & \\ & & & A_{ss} \end{pmatrix} := \text{diag}\{A_{11}, A_{22}, \dots, A_{ss}\},$$

其中 A_{ii} ($1 \leq i \leq s$) 是 n_i 阶方阵. 反之, 如果 T 在 \mathcal{B} 下的矩阵是上述的对角块矩阵, 则

$$W_i = \text{span}\{\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in_i}\}, \quad (i = 1, 2, \dots, s)$$

是 T 的不变子空间, 且 V^n 是这些不变子空间的直和.

对角矩阵

一般来说, 对角矩阵是最简单的方阵, 也有许多特殊的性质. 我们想要知道, 如果线性变换 T 在基 $\mathcal{B} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下的矩阵是对角矩阵:

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \dots & \\ & & & \lambda_s \end{pmatrix} := \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_s\},$$

那么 T 应该要满足什么要求呢?

对角矩阵对应的线性变换

由 $T\mathcal{B} = \mathcal{B}A$, 有

$$(T\alpha_1, \dots, T\alpha_n) = (\alpha_1, \dots, \alpha_n) \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \dots & \\ & & & \lambda_s \end{pmatrix},$$

其中每个分量满足

$$T\alpha_i = \lambda_i \alpha_i, \quad 1 \leq i \leq n$$

这与特征值、特征向量的定义相关.

特征值与特征向量

定义: 设 T 是 $V^n(F)$ 的一个线性变换, 如果存在 $\lambda_0 \in F, \xi \in V^n(F)$ 且 $\xi \neq 0$, 使

$$T\xi = \lambda_0\xi$$

则称 λ_0 是 T 的一个**特征值**, ξ 称为 T 关于 λ_0 的**特征向量**.

求 T 的特征值与特征向量

在 V^n 中取一个基 $B = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, 且设 T 在 B 下的矩阵是 A . 如果 ξ 是 T 的一个特征向量, λ_0 是相应的特征值, 即 $T\xi = \lambda_0\xi, \xi \neq 0$, 那么 ξ 可由 B 的线性表出:

$$\xi = \sum_{i=1}^n x_i \alpha_i = B \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

因而有 $T\xi = TBx = BAx, \quad \lambda_0\xi = B(\lambda_0x)$.
代入得, $BAx = B(\lambda_0x)$ 或 $B(Ax - \lambda_0x) = 0$.

于是 ξ 在 B 下的坐标向量 x 满足:

$$Ax = \lambda_0 x.$$

则线性变换 T 的特征值问题与对应矩阵 A 的特征值问题是一一对应的. 对于给定的基 B ,

$$T\xi = \lambda_0 \xi \Leftrightarrow Ax = \lambda_0 x$$

由于相似矩阵有相同的特征多项式, 所以我们可以把 A 的特征多项式

$$f(\lambda) := \det(\lambda I_n - A) = \lambda^n + b_{n-1}\lambda^{n-1} + \dots + b_1\lambda + b_0$$

称为 T 的特征多项式, 于是 T 的特征值就是 T 的特征多项式的根.

例 8. $\mathbf{P}_2(t)$ 的线性变换 T 定义为

$$Tp(t) = p(t) + (t+1)\frac{d}{dt}p(t)$$

求 T 的特征值和特征向量.

解: 取 $\mathbf{P}_2(t)$ 的一组基 $\mathcal{B} = \{1, t, t^2\}$, 则 T 在 \mathcal{B}

下的矩阵是 $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$, 这里 A 的特征值

为 $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3$, 对应的特征向量分别为 $(1, 0, 0)^T, (1, 1, 0)^T, (1, 2, 1)^T$. 因此, T 的特征值为 $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3$, 对应的 T 特征向量分别为是多项式 $t, t+1, 1+2t+t^2$.

定理: T 关于不同特征值的特征向量线性无关.
(即如果 $T\alpha_i = \lambda_i\alpha_i$ 且 $\lambda_i \neq \lambda_j$, 其中 $i, j = 1, 2, \dots, r$ 且 $i \neq j$, 那么 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性无关.)

证明: 用数学归纳法证明.

- 由于特征向量是非零向量, 所以单个的特征向量线性无关.
- 假设 T 关于 k 个不同特征值的特征向量是线性无关的, 要证 T 关于 $k+1$ 个不同特征值 $\lambda_1, \lambda_2, \dots, \lambda_{k+1}$ 的特征向量 $\alpha_1, \alpha_2, \dots, \alpha_{k+1}$ 也线性无关.

证明 (续): 现证明 T 关于不同特征值 $\lambda_1, \lambda_2, \dots, \lambda_{k+1}$ 的特征向量 $\alpha_1, \alpha_2, \dots, \alpha_{k+1}$ 线性无关.

设有等式成立

$$b_1\alpha_1 + b_2\alpha_2 + \dots + b_{k+1}\alpha_{k+1} = 0 \quad (1)$$

等式 (1) 两端同乘以 λ_{k+1} , 得

$$b_1\lambda_{k+1}\alpha_1 + b_2\lambda_{k+1}\alpha_2 + \dots + b_{k+1}\lambda_{k+1}\alpha_{k+1} = 0$$

等式 (1) 两端同时作用线性变换 T , 得

$$b_1\lambda_1\alpha_1 + b_2\lambda_2\alpha_2 + \dots + b_{k+1}\lambda_{k+1}\alpha_{k+1} = 0$$

以上两式相减, 得

$$b_1(\lambda_{k+1}-\lambda_1)\alpha_1 + b_2(\lambda_{k+1}-\lambda_2)\alpha_2 + \dots + b_k(\lambda_{k+1}-\lambda_k)\alpha_k = 0$$

证明 (续): 根据归纳法假设 $\alpha_1, \alpha_2, \dots, \alpha_k$ 是线性无关的, 故

$$b_i(\lambda_{k+1} - \lambda_i) = 0, \quad 1 \leq i \leq k,$$

但 $\lambda_{k+1} - \lambda_i \neq 0$, 所以有 $b_i = 0$ ($i = 1, 2, \dots, k$). 从而等式 (1) 变为

$$b_{k+1}\alpha_{k+1} = 0.$$

又因为 $\alpha_{k+1} \neq 0$, 则 $b_{k+1} = 0$, 于是 $\alpha_1, \alpha_2, \dots, \alpha_{k+1}$ 是线性无关的.

定理: 设 $\lambda_1, \lambda_2, \dots, \lambda_k$ 是 T 的不同特征值. 而 $\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ir_i}$ ($1 \leq i \leq k$) 是 T 关于 λ_i 的 r_i 个线性无关特征向量, 则向量组

$$\{\alpha_{11}, \alpha_{12}, \dots, \alpha_{1r_1}, \alpha_{21}, \alpha_{22}, \dots, \alpha_{2r_2}, \dots, \alpha_{k1}, \alpha_{k2}, \dots, \alpha_{kr_k}\}$$

线性无关.

证明思路: 同样用数学归纳法证明.

- 只考虑关于一个特征值 λ_1 时, 由条件, $\alpha_{11}, \alpha_{12}, \dots, \alpha_{1r_1}$ 是线性无关的特征向量.
- 假设 T 关于 k 个不同特征值的特征向量组是线性无关的, 要证 T 关于 $k+1$ 个不同特征值 $\lambda_1, \lambda_2, \dots, \lambda_{k+1}$ 的特征向量组也线性无关.

线性变换的特征子空间

对于 T 的任一特征值 λ_0 , T 关于 λ_0 的所有特征向量, 再添上零向量组成一个集合

$$V_{\lambda_0} := \{\alpha \in V^n \mid T\alpha = \lambda_0\alpha\}$$

容易验证, 对于任意 $\alpha, \beta \in V_{\lambda_0}$ 和 $k, l \in F$, 有

$$T(k\alpha + l\beta) = kT\alpha + lT\beta = k\lambda_0\alpha + l\lambda_0\beta = \lambda_0(k\alpha + l\beta)$$

即 V_{λ_0} 是 $V^n(F)$ 的一个子空间, 称为 T 关于 λ_0 的**特征子空间**. $\dim V_{\lambda_0}$ 称为 λ_0 的**几何重数**.

如果 $\alpha \in V_{\lambda_0}$, 则 $T(T\alpha) = \lambda_0(T\alpha)$, 易知特征子空间 V_{λ_0} 是 T 的一个不变子空间.

如果 $\lambda_1, \lambda_2, \dots, \lambda_s$ 是 T 的所有不同的特征值, 则 T 的特征多项式 $f(\lambda)$ 可以表示为

$$f(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \dots (\lambda - \lambda_s)^{n_s},$$

且其中 $n_1 + n_2 + \dots + n_s = n$, 而 n_i ($1 \leq i \leq s$) 称为特征值 λ_i 的代数重数.

问题: 几何重数 v.s. 代数重数?

定理: 设 $\lambda_1, \lambda_2, \dots, \lambda_s$ 是 T 的所有不同特征值. 对任一 λ_i ($1 \leq i \leq s$), 都有

$$\dim V_{\lambda_i} \leq n_i$$

即任何特征值的几何重数不大于其代数重数.

证明: 不失一般性, 对 λ_1 进行证明.

设 $\dim V_{\lambda_1} = k$, 则 T 关于 λ_1 有 k 个线性无关的特征向量 $\alpha_1, \alpha_2, \dots, \alpha_k$. 从而 $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 是 T 的不变子空间 V_{λ_1} 的一个基, 把它扩充为 V^n 的基

$$\mathcal{B} = \{\alpha_1, \alpha_2, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n\},$$

那么 T 在 \mathcal{B} 下的矩阵为 $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$,

证明 (续):

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

其中 A_{11} 为 k 阶对角矩阵 $\text{diag}\{\lambda_1, \lambda_1, \dots, \lambda_1\}$. 且 A_{21} 为零矩阵, 所以 A 的特征多项式

$$\det(\lambda I_n - A) = (\lambda - \lambda_1)^k \cdot \det(\lambda I_{n-k} - A_{22}).$$

其中 $(\lambda - \lambda_1)$ 的次数不小于 $k = \dim V_{\lambda_1}$, 即 λ_1 的几何重数 $k \leq n_1$.

注:

若 $\lambda_1, \lambda_2, \dots, \lambda_s$ 是 T 的所有不同的特征值, 则有

$$V_{\lambda_1} + V_{\lambda_2} + \dots + V_{\lambda_s} = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_s},$$

根据定理, 有

$$\begin{aligned} \dim V_{\lambda_1} + \dim V_{\lambda_2} + \dots + \dim V_{\lambda_s} &\leq n_1 + \dots + n_s = n, \\ V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_s} &\subset V^n \end{aligned}$$

即特征子空间的维数之和小于等于 n .

可对角化

定义: T 称为是**可对角化的**, 如果存在 V^n 的基 B , 使 T 在 B 下的矩阵是对角矩阵.

定理: T 是可对角化的充分必要条件是下列等价条件之一成立:

- (1) T 有 n 个线性无关的特征向量;
- (2) $\dim V_{\lambda_i} = n_i, 1 \leq i \leq s$.
- (3) $V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_s} = V^n$.

证明: 三者等价性易证, 我们这里只证**(1) 是充要条件**.

证明: 先证(1) 是必要条件.

设 T 在基 $\mathcal{B} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下的矩阵是对角矩阵: $\text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 即

$$(T\alpha_1, T\alpha_2, \dots, T\alpha_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)\text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$$

则 $T\alpha_i = \lambda_i\alpha_i$, $1 \leq i \leq n$. 即 α_i 是 T 的特征向量, λ_i 是相应的特征值, 从而 T 有 n 个线性无关的特征向量.

证明: 再证(1) 是充分条件.

若 T 有 n 个线性无关的特征向量, 则有

$$T\alpha_i = \lambda_i\alpha_i, \quad 1 \leq i \leq n.$$

且这 n 个向量是一组基, 从而 T 在这组基下的矩阵是对角矩阵. 对角线上的数 $\lambda_1, \lambda_2, \dots, \lambda_n$ 都是 T 的特征值.

推论: 若 T 有 n 个不同的特征值, 则 T 必可对角化. 于是, 对于 $A \in \mathbb{C}^{n \times n}$, 若 A 的特征多项式没有重根, 则 A 必可对角化. 换句话说, A 相似于一个对角矩阵, 即存在可逆矩阵 P , 使得 $P^{-1}AP$ 为对角矩阵, 且 P 的每个列向量都是 A 的特征向量.

例 7. 证明 $\mathbf{P}_2(t)$ 的线性变换 $D = \frac{d}{dt}$ 是不可对角化的.

证: 取 $\mathbf{P}_2(t)$ 的一个基 $\mathcal{B} = \{1, t, t^2\}$, D 在 \mathcal{B} 下的矩阵是

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

A 的特征多项式是 $\det(\lambda I_3 - A) = \lambda^3$, 则 $\lambda = 0$ 是它的三重根, 0 的代数重数为 3.

例 8. 证明 $\mathbf{P}_2(t)$ 的线性变换 $D = \frac{d}{dt}$ 是不可对角化的.

证 (续): 同时齐次线性方程组

$$(A - 0I_3)x = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

的解由向量 $(1, 0, 0)^T$ 张成. 则 A 关于 0 的特征子空间由 $(1, 0, 0)^T$ 张成, 即 0 的几何重数为 1 .

因为特征值 0 的几何重数小于代数重数, 故 D (和 A) 不可对角化.

例 9. \mathbb{R}_3 的线性变换 T 定义为

$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

问 T 是否可对角化?

解: 取 \mathbb{R}_3 的标准基 $B = \{e_1, e_2, e_3\}$, T 在 B 下的矩阵是

$$A = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix},$$

A 的特征多项式是 $\det(\lambda I_3 - A) = (\lambda - 5)(\lambda - 3)^2$,
则 $\lambda_1 = 5$ 代数重数为 1, $\lambda_2 = 3$ 的代数重数为 2.

解 (续): 关于 $\lambda_1 = 5$, 齐次线性方程组

$$(A - 5I_3)x = \begin{pmatrix} -1 & 0 & 1 \\ 2 & -2 & 2 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

解由向量 $(1, 2, 1)^T$ 张成, 即 5 的几何重数为 1.
关于 $\lambda_1 = 3$, 齐次线性方程组

$$(A - 3I_3)x = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

解由线性无关的向量 $(1, 0, -1)^T$ 和 $(0, 1, 0)^T$ 张成, 即 3 的几何重数为 2.

解 (续): 则每个特征值的代数重数等于几何重数, 由定理, T 可对角化.

具体来说, 在 \mathbb{R}^3 取基

$$\mathcal{B} = \{(1, 2, 1)^T, (1, 0, -1)^T, (0, 1, 0)^T\}$$

则 T 在 \mathcal{B} 下的矩阵是对角矩阵 $\text{diag}\{5, 3, 3\} =$

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

例 10. 证明矩阵

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

在实数域上是不可对角化的, 但在复数域上是可对角化的.

解: A 的特征多项式是

$\det(\lambda I_2 - A) = \lambda^2 - 2\lambda + 2$, 这在实数域上没有根, 所以也不存在对应的特征向量, 即 A 在实数域上不可对角化. 但在复数域上, 特征多项式有两个根 $\lambda_1 = 1 + i$ 和 $\lambda_2 = 1 - i$, 从而 A 可对角化

解 (续): 由 $(\lambda_1 I_2 - A)x_1 = (\lambda_2 I_2 - A)x_2 = 0$ 可求出

$$x_1 = (1, i)^T, \quad x_2 = (1, -i)^T.$$

则

$$\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} = \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}.$$