第五章 特征值和特征向量

主要内容

5.1 基本概念

5.2 对角化和谱分解

5.3 矩阵的相似

本章只考虑方阵

5.1 基本概念

(a) 问题的引入以及一些准备

(b) 特征值和特征向量的定义 (重要概念)

(c) 特征多项式—如何求特征值和特征向量

(d) 矩阵的特征值、特征向量与特征多项式的一些性质

(a) 问题的引入: 方阵的幂

π

考虑某地的城镇化。假定每年有97%的人口选择留在城市，3%的人口选择由城镇迁移至乡村；每年有95%的人口选择留在农村，5%的人口选择由乡村迁移至城镇。设总人口为100万。

如果初始有$u_{country,0}$=60万人口在农村，$u_{city,0}$=40万人在城市。一年后，

$$\begin{bmatrix} u_{city,1} \\ u_{country,1} \end{bmatrix} = \begin{bmatrix} 0.97u_{city,0} + 0.05u_{country,0} \\ 0.03u_{city,0} + 0.95u_{country,0} \end{bmatrix} = \begin{bmatrix} 0.97 & 0.05 \\ 0.03 & 0.95 \end{bmatrix} \begin{bmatrix} u_{city,0} \\ u_{country,0} \end{bmatrix} = \begin{bmatrix} 418{,}000 \\ 582{,}000 \end{bmatrix}$$

于是矩阵$A = \begin{bmatrix} 0.97 & 0.05 \\ 0.03 & 0.95 \end{bmatrix}$记住了人口迁移的情况.

问题：预测多年以后的人口分布情况。

两个假设：总人口不变；每年的迁移比例也不变。

第k年的人口分布$\boldsymbol{u}_k = \begin{bmatrix} u_{city} \\ u_{country} \end{bmatrix}_k = A \begin{bmatrix} u_{city} \\ u_{country} \end{bmatrix}_{k-1} = \cdots = A^k \begin{bmatrix} u_{city} \\ u_{country} \end{bmatrix}_0$.

问题转化为计算矩阵的方幂。通过**矩阵的对角化**计算方幂是一个常用方法。

可对角化矩阵计算方幂：

如果方阵A可对角化，即有可逆矩阵X使得$A = X\Lambda X^{-1}$, Λ为对角阵

那么, $A^2 = X\Lambda X^{-1}X\Lambda X^{-1} = X\Lambda^2 X^{-1}$; $A^k = X\Lambda^k X^{-1}$.

线性映射视角看待矩阵可对角化：

$A = X\Lambda X^{-1} \Rightarrow AX = X\Lambda$

$$A[\boldsymbol{x}_1, \dots, \boldsymbol{x}_n] = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$\boxed{A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i}$$

因此，$\mathbb{R}^n$的一组基$\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$满足$A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$

例:

$\mathbb{R}^2$上的反射变换$H_{\boldsymbol{v}} = I_2 - 2\boldsymbol{v}\boldsymbol{v}^T$，$\boldsymbol{v}$是单位向量。

令$\boldsymbol{w}$是与$\boldsymbol{v}$正交的单位向量，则$H_{\boldsymbol{v}}\boldsymbol{w} = \boldsymbol{w}, H_{\boldsymbol{v}}\boldsymbol{v} = -\boldsymbol{v}$,

因此，$H_{\boldsymbol{v}}[\boldsymbol{v}\ \ \boldsymbol{w}] = [-\boldsymbol{v}\ \ \boldsymbol{w}] = [\boldsymbol{v}\ \ \boldsymbol{w}]\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

令$X = [\boldsymbol{v}\ \ \boldsymbol{w}] \in M_2(\mathbb{R}), H_{\boldsymbol{v}} = X\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}X^{-1}$.

$$H_{\boldsymbol{v}}^k = X\begin{bmatrix} (-1)^k & 0 \\ 0 & 1^k \end{bmatrix}X^{-1}.$$

$\boldsymbol{v}, \boldsymbol{w}$满足方程$H_{\boldsymbol{v}}\boldsymbol{x} = \lambda\boldsymbol{x}$,对某个$\lambda \in \mathbb{R}$.

例:

正交矩阵$Q=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$，在$\mathbb{R}^2$上给出的线性变换是逆时针旋转90°.

几何上：不存在$\lambda \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^2$使得$Q\boldsymbol{x}=\lambda\boldsymbol{x}$.

代数上：如果有λ使得方程$Q\boldsymbol{x}=\lambda\boldsymbol{x}$有解，则

$\lambda I_2 - Q$不可逆，即$\det(\lambda I_2 - Q)=0$。

我们得到方程$\lambda^2+1=0$.

$\lambda^2+1=0$.在实数中无解。

但是这个方程在复数中有解，$\lambda=i,-i$

且有$\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1\\ i\end{bmatrix}=-i\begin{bmatrix}1\\ i\end{bmatrix}$，$\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\begin{bmatrix}i\\ 1\end{bmatrix}=i\begin{bmatrix}i\\ 1\end{bmatrix}$. 令$X=\begin{bmatrix}1 & i\\ i & 1\end{bmatrix}$,

$Q^k=X\begin{bmatrix}(-i)^k & \\ & i^k\end{bmatrix}X^{-1}$.

代数基本定理

一个复系数一元n次的一元多项式$p(x) \in \mathbb{C}[x]$在$\mathbb{C}$上恰好有n个根（计算重数）

即存在因式分解$p(x) = a_0(x - x_1)^{n_1} \dots (x - x_k)^{n_k}$，$n_1 + \cdots + n_k = n, x_i \in \mathbb{C}$

本章出现的矩阵为**复矩阵**，数为**复数**和向量为**复向量**。

给定复向量 $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$，定义 $\overline{\boldsymbol{x}} = \begin{bmatrix} \overline{x_1} \\ \vdots \\ \overline{x_n} \end{bmatrix}$

给定复矩阵 $A = [a_{ij}]$，定义 $\bar{A} = [\overline{a_{ij}}]$

改为复矩阵后的变化:

复矩阵与实矩阵的**相同之处**：

矩阵的运算，子空间与基，高斯消元解方程的过程，行列式的计算和定义等和实矩阵一样。

复矩阵与实矩阵的**不同之处**：

$\mathbb{C}^n$的内积应该定义为$\overline{\boldsymbol{v}}^T\boldsymbol{w}$，正交性改为$\boldsymbol{v}\perp\boldsymbol{w}$，如果$\overline{\boldsymbol{v}}^T\boldsymbol{w}=0$.

于是$\overline{\boldsymbol{v}}^T\boldsymbol{v}=0$当且仅当$|v_1|^2+\cdots+|v_n|^2=0$，当且仅当$v_1=\cdots=v_n=0$.

(b) 特征值和特征向量的定义

给定$A \in M_n(\mathbb{C})$，如果对$\lambda \in \mathbb{C}$，存在非零向量$\boldsymbol{x} \in \mathbb{C}^n$，使得

$$\boxed{A\boldsymbol{x} = \lambda\boldsymbol{x},}$$

则称λ为A（在$\mathbb{C}$上）的一个**特征值**, 称非零向量$\boldsymbol{x}$为A的一个属于特征值为λ的**特征向量**。（$\lambda, \boldsymbol{x}$）称为**特征对**。

如果A为实方阵，如果特征对（$\lambda, \boldsymbol{x}$）满足$\lambda \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^n$，则分别称二者为$A$在$\mathbb{R}$上的特征值和特征向量，称该二元组为$A$在$\mathbb{R}$上的特征对。

注意:

(1) 只有方阵才有特征值和特征向量

(2) 零向量**不是**特征向量

(3) 如果$\boldsymbol{x}$是属于特征值λ的特征向量，则$c\boldsymbol{x}, c \neq 0$也是属于特征值$\lambda$的特征向量

(4) 如果$\boldsymbol{x}$是A的属于特征值λ的特征向量，则$\boldsymbol{x}$是A^k属于特征值λ^k的特征向量

$$A^k(\boldsymbol{x}) = A^{k-1}(\lambda\,\boldsymbol{x}) = \cdots = \lambda^k \boldsymbol{x}$$

例:

如果A可逆，$(\lambda, \boldsymbol{x})$是$A$的特征对，则$\lambda \neq 0$且$(\lambda^{-1}, \boldsymbol{x})$是$A^{-1}$特征对。

证明:

首先$\lambda \neq 0$，否则与A可逆矛盾

$\boldsymbol{x} = I\boldsymbol{x} = A^{-1}A\boldsymbol{x} = A^{-1}\lambda\boldsymbol{x}.$

因此 $A^{-1}\boldsymbol{x} = \lambda^{-1}\boldsymbol{x}.$

问题：如何求特征值、特征向量?

命题:

A为n阶方阵，$\lambda_0 \in \mathbb{C}$. 以下叙述等价：

(1) λ_0为A的特征值

(2) 方程$(\lambda_0 I_n - A)\boldsymbol{x} = \mathbf{0}$有非零解

(3) 方阵$\lambda_0 I_n - A$不可逆

(4) $\det(\lambda_0 I_n - A) = 0$

特别地，0是A的特征值当且仅当A不可逆

证明：

$(\lambda_0, \boldsymbol{x}_0)$是一个特征对，则$\boldsymbol{x}_0 \neq \mathbf{0}$且$A\boldsymbol{x}_0 = \lambda_0 \boldsymbol{x}_0$

于是$\boldsymbol{x}_0$满足方程$(\lambda_0 I_n - A)\boldsymbol{x} = \mathbf{0}$

(c) 基本概念：特征值多项式:

以λ为不定变元的多项式

$$p_A(\lambda) = \det(\lambda I_n - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & & \vdots \\ \vdots & & \ddots & -a_{n-1,n} \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

称为A的**特征多项式。**

$p_A(\lambda)$次数为n，最高次数项为λ^n.

例:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$p_A(\lambda) = |\lambda I_3 - A| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} \\ -a_{21} & \lambda - a_{22} & -a_{23} \\ -a_{31} & -a_{32} & \lambda - a_{33} \end{vmatrix} = (\lambda - a_{11})(\lambda - a_{22})(\lambda - a_{33}) + \text{次数} \le 1\text{的项}.$$

定理:

设n 阶方阵A的特征多项式为$p_A(\lambda)$，那么

(1) λ_0是A的特征值当且仅当$p_A(\lambda_0) = |\lambda_0 I - A| = 0$

(2) 向量$\boldsymbol{x}_0 \in \mathbb{C}^n$是$A$的属于$\lambda_0$的特征向量当且仅当$\boldsymbol{x}_0 \neq \mathbf{0}$且$\boldsymbol{x}_0 \in \mathcal{N}(\lambda_0 I_n - A)$.

定义:

解空间$\mathcal{N}(\lambda_0 I_n - A)$称为$A$的属于$\lambda_0$的**特征子空间**.

$\{A$的属于λ_0的特征向量$\} = \mathcal{N}(\lambda_0 I_n - A) \backslash \{\mathbf{0}\}$

例:

二阶方阵$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$的特征多项式

$$p_A(\lambda) = \det(\lambda I_2 - A) = \begin{vmatrix} \lambda - a & -b \\ -c & \lambda - d \end{vmatrix}$$

$$= \lambda^2 - (a + d)\lambda + (ad - bc).$$

$a + d$为A的迹$(trace)$, $ad - bc$为A的行列式。

例:

上三角或下三角矩阵的特征值为对角线上的元素。

证明:

以下三角矩阵为例。设 $L = \begin{bmatrix} a_{11} & & \\ \vdots & \ddots & \\ a_{1n} & \cdots & a_{nn} \end{bmatrix}$.

$p_L(\lambda) = |\lambda I_n - L| = (\lambda - a_{11}) \dots (\lambda - a_{nn})$,

$p_L(\lambda) = 0$当且仅当λ为$a_{11}, \dots, a_{nn}$中一个数

命题:

$p_A(\lambda) = p_{A^T}(\lambda)$. 因此，$A$与$A^T$有相同的特征值。

证明:

$$p_A(\lambda) = \det(\lambda I - A) = \det(\lambda I - A)^T = \det\left(\lambda I - A^T\right) = p_{A^T}(\lambda).$$

例:

$Q=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$. 求特征值、特征向量

$p_Q(\lambda)=\lambda^2+1=(\lambda-i)(\lambda+i)$. 因此, Q的特征值为$i,-i$.

$\det Q=1=(-i)(i)$, $\quad trace(Q)=-i+i=0=a_{11}+a_{22}$.

求解方程$(iI_2-Q)\boldsymbol{x}=\mathbf{0}$得到特征子空间$\mathcal{N}(iI_2-Q)=\{k\begin{bmatrix}i\\ 1\end{bmatrix}\}$

求解方程$(-iI_2-Q)\boldsymbol{x}=\mathbf{0}$得到特征子空间$\mathcal{N}(-iI_2-Q)=\{k\begin{bmatrix}1\\ i\end{bmatrix}\}$

例：特征值的分布—Gershgorin圆盘定理

对n阶方阵$A=[a_{ij}]_{i,j}$，定义如下n个圆盘

$G_i(A)=\left\{z \middle| \mid z-a_{ii} \mid \le \sum_{j\neq i} |a_{ij}| \right\}$，则

A的全部特征值一定落在这n个圆盘之一.

证明:

这是对角占优矩阵可逆的应用：以$n=3$为例，

如果$|\lambda_0-a_{11}|>|a_{12}|+|a_{13}|, |\lambda_0-a_{22}|>|a_{21}|+|a_{23}|, |\lambda_0-a_{33}|>|a_{31}|+|a_{32}|$,

则$\lambda_0 I-A=\begin{bmatrix}\lambda_0-a_{11} & -a_{12} & -a_{13}\\ -a_{21} & \lambda_0-a_{22} & -a_{23}\\ -a_{31} & -a_{32} & \lambda_0-a_{33}\end{bmatrix}$是对角占优矩阵，可逆。

如果λ_0是A的特征值，则矩阵$\lambda_0 I - A$不可逆。于是，λ_0不会同时满足上面三个条件。

那么，λ_0一定满足下列三个条件中的至少一个

$|\lambda - a_{11}| \le |a_{12}| + |a_{13}|, |\lambda - a_{22}| \le |a_{21}| + |a_{23}|, |\lambda - a_{33}| \le |a_{31}| + |a_{32}|.$

即λ_0一定落在上述三个圆盘中的一个。

小结:

求A的特征值λ_0 $\Longleftrightarrow$ 求$p_A(\lambda) = 0$的解

求λ_0对应的特征子空间 $\Longleftrightarrow$ 求$\mathcal{N}(\lambda_0 I - A)$的一组基

(d) 矩阵的特征值、特征向量与特征多项式的一些性质:

命题：设A为n阶矩阵，特征值为$\lambda_1, \dots, \lambda_n \in \mathbb{C}$。则

(1) 行列式=特征值之积，即$|A| = \lambda_1\lambda_2 \dots \lambda_n$.

(2) 迹=特征值之和，即$\mathrm{trace}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

(3) $p_A(\lambda) = \lambda^n - \mathrm{trace}(A)\lambda^{n-1} + \cdots + (-1)^n \det A$.

证明：设$\lambda_1, \dots, \lambda_n \in \mathbb{C}$为$A$的特征值。

$p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0 = (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)$.

$p_A(0) = (-1)^n \lambda_1\lambda_2 \dots \lambda_n = (-1)^n |A|$.

因此，$|A| = \lambda_1\lambda_2 \dots \lambda_n$.

π

将$p_A(\lambda) = \begin{vmatrix} \lambda - a_{11} & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{n1} & \cdots & \lambda - a_{nn} \end{vmatrix}$沿第一行展开，得到

$$p_A(\lambda) = (\lambda - a_{11})(\lambda - a_{22}) \dots (\lambda - a_{nn}) + (\text{次数} \leq n - 2\text{的项})$$

$$= \lambda^n - (a_{11} + \cdots + a_{nn})\lambda^{n-1} + (\text{次数} \leq n - 2\text{的项})$$

另一方面，

$$p_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)$$

$$= \lambda^n - (\lambda_1 + \cdots + \lambda_n)\lambda^{n-1} + (\text{次数} \leq n - 2\text{的项})$$

对比得知，$a_{11} + \cdots + a_{nn} = \lambda_1 + \cdots + \lambda_n$.

代数重数与几何重数:

$A \in M_n(\mathbb{C})$, 由代数学基本定理,

$p_A(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \dots (\lambda - \lambda_k)^{n_k}$, 其中$\lambda_i \in \mathbb{C}$互不相同,

$n_1, \dots, n_k \geq 1, n_1 + \cdots + n_k = n.$

$AM(\lambda_i) = n_i$ 称为特征值λ_i的**代数重数** (Algebraic Multiplicity)

$GM(\lambda_i) = \dim \mathcal{N}(\lambda_i I_n - A)$称为$\lambda_i$的**几何重数**(Geometric Multiplicity)

我们下小节会看到，对任意特征值λ，总有 $GM(\lambda) \leq AM(\lambda)$

例:

$Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. p_Q(\lambda) = (\lambda - i)(\lambda + i).$

$AM(i) = GM(i) = 1, AM(-i) = GM(-i) = 1$

实矩阵的复特征值与复特征向量:

命题：$A \in M_n(\mathbb{R})$. $\lambda_0 \in \mathbb{C} \backslash \mathbb{R}$. $(\lambda_0, \boldsymbol{x}_0)$为$A$的特征对。那么$(\overline{\lambda_0}, \overline{\boldsymbol{x}_0})$也为$A$的特征对且$AM(\lambda_0) = AM\left(\overline{\lambda_0}\right), GM(\lambda_0) = GM\left(\overline{\lambda_0}\right)$.

证明：由于$A\boldsymbol{x}_0 = \lambda_0 \boldsymbol{x}_0$，取共轭得到$\bar{A}\overline{\boldsymbol{x}_0} = \overline{\lambda_0}\overline{\boldsymbol{x}_0}$.

由于A为实矩阵，$\bar{A} = A$, 于是$A\overline{\boldsymbol{x}_0} = \overline{\lambda_0}\overline{\boldsymbol{x}_0}$. 因此$\left(\overline{\lambda_0}, \overline{\boldsymbol{x}_0}\right)$也为$A$的特征对。

设$n_0 = AM(\lambda_0)$，$n_0' = AM\left(\overline{\lambda_0}\right)$。由于$A$为实矩阵，$p_A(\lambda) \in \mathbb{R}[\lambda]$为实多项式，

由$p_A(\lambda) = (\lambda - \lambda_0)^{n_0}\left(\lambda - \overline{\lambda_0}\right)^{n_0'} \cdots$为实多项式，得到$n_0 = n_0'$

由于$\mathcal{N}(\lambda_0 I - A)$的线性无关的向量$\boldsymbol{x}_1, \dots, \boldsymbol{x}_d$取共轭后得到线性无关的向量$\overline{\boldsymbol{x}_1}, \dots, \overline{\boldsymbol{x}_d}$。因此 $GM(\lambda_0) = GM\left(\overline{\lambda_0}\right)$.

小结: 对于实矩阵，复特征值成对出现。“成对”意思是:

(1) λ是特征值则 $\bar{\lambda}$也是特征值

(2) λ与$\bar{\lambda}$的重数（代数重数、几何重数）相等

本节小结：

(1) A可对角化：有可逆矩阵X使得$A = X\Lambda X^{-1}$, Λ为对角阵

$X = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]$, $A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$

(2) 特征值和特征向量：

特征多项式$\det(\lambda I_n - A) = 0$求特征值；

解方程$(\lambda I_n - A)\boldsymbol{x} = \mathbf{0}$求特征向量

(3) $|A| = \lambda_1 \lambda_2 \cdots \lambda_n$； $\operatorname{trace}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

(4) 对于实矩阵，复特征值成对出现