第五章　特征值与特征向量

2021 年 12 月 15 日

动机

- 把线性空间 V 上的一个线性变换 $\vec{x} \mapsto A\vec{x}$ 分解成容易理解的元素.

  具体来说, 就是问是否存在一个 (线性无关) 向量组 $\{\vec{\xi_1}, \cdots\}$, 其中 $A\vec{\xi_i} = \lambda_i \vec{\xi_i}$, $i = 1, 2, \cdots$;

  且对于 $\forall \vec{x} \in V$ 可以用该向量组的向量来展开, 即 $\vec{x} = a_1\vec{\xi_1} + a_2\vec{\xi_2} + \cdots + a_m\vec{\xi_m}$?

- 这时我们称 $\vec{\xi_i}$ 是矩阵 A (或者说是对应的线性变换) 的一个特征向量. 上面的式子 $\vec{x} = \sum_i a_i\vec{\xi_i}$ 就是把向量 $\vec{x}$ 按照 A 的特征向量展开.

内容

定义

A 为 $n \times n$ 矩阵, $\vec{x}$ 为非零向量, 若存在数 λ 使得 $A\vec{x} = \lambda\vec{x}$, 则称 λ 为 A 的一个特征值, $\vec{x}$ 称为对应于 λ 的特征向量.

- λ 是 A 的特征值当且仅当 $(A - \lambda I)\vec{x} = \vec{0}$ 有非零解 $\Leftrightarrow \det(A - \lambda I) = 0$.
- $A - \lambda I$ 的零空间, 即 $\mathrm{Nul}\,(A - \lambda I)$, 称为矩阵 A 关于 λ 的特征子空间, 记为 V_λ, 也就是说 $V_\lambda := \{\vec{x} \,|\, A\vec{x} = \lambda\vec{x}\}$.
- 只有 λ 是 A 的特征值时, 才有 $\dim V_\lambda > 0$.

示例

例

- 上 (下) 三角矩阵的特征值
- λ 是 A 的特征值 $\Leftrightarrow$ λ 是 A^T 的特征值.
- A 有特征值 $0 \Leftrightarrow \det A = 0$, 即 A 不可逆.
- A 有特征值 λ, 则 A^n 有特征值 λ^n; 若 A 可逆, 则 A^{-1} 有特征值 λ^{-1}.

  讨论: A 的伴随矩阵 $\operatorname{adj} A$ 的特征值.

关于方阵 A 的多项式

设有形如 $P(x) = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_0$ 的关于 x 的 m 阶多项式，定义 P 关于 n 阶方阵 A 的矩阵多项式为

$$P(A) \equiv a_m A^m + a_{m-1} A^{m-1} + \cdots + a_0 I_n,$$

其中 I_n 是 n 阶单位矩阵.

定义

若有 $P(A) = 0$, 则称多项式 P 为矩阵 A 的一个零化多项式.

定理

若 $P(A) = 0$, λ 是矩阵 A 的一个特征值, 则有 $P(\lambda) = 0$.

例

- $A^n = 0$, 则 A 只有零特征值.
- $A^2 = 1$, 则 A 有特征值 ± 1; 其对应的特征向量可为 $\vec{v} \pm A\vec{v}$ (若其 $\neq \vec{0}$).
- $A^2 = A$, 则 A 有特征值 0 和 1; 其对应的特征向量可分别为 $\vec{v} - A\vec{v}$ 和 $A\vec{v}$.

内容

特征多项式

定理 (特征值与特征方程的根)

λ 是 n 阶矩阵 A 的特征值的充要条件: λ 满足特征方程 $\det(A-\lambda I)=0$.

$p_A(\lambda) := \det(A-\lambda I)$ 称为矩阵 A 的特征多项式.

- $p_A(\lambda)$ 是关于 λ 的 n 次多项式, 最高次项为 $(-\lambda)^n$.
- 特征方程有 n 个复根 $\lambda_1, \cdots, \lambda_n$ (包括重根, 参考代数学基本定理).
- λ_i 作为特征方程根的重数称为它的代数重数.
- $p_A(\lambda)$ 是矩阵 A 的零化多项式, 即有 $p_A(A)=0$ (Cayley-Hamilton 定理).

- $p_A(\lambda)$ 可以展开为

$$\begin{aligned} p_A(\lambda) &= (\lambda_1-\lambda)\cdots(\lambda_n-\lambda) \\ &= (-\lambda)^n + \left[\sum_{i=1}^n \lambda_i\right](-\lambda)^{n-1} + \left[\sum_{i<j}\lambda_i\lambda_j\right](-\lambda)^{n-2} + \cdots + \prod_{i=1}^n \lambda_i. \end{aligned}$$

- 对比 $\det(A-\lambda I)$ 的展开系数, 我们有

$\mathrm{Tr}\, A = \sum_i \lambda_i$,
A 的所有二阶主子式之和 $= \sum_{i<j}\lambda_i\lambda_j = \frac{1}{2}\left[(\mathrm{Tr}\, A)^2 - \mathrm{Tr}\left(A^2\right)\right]$,
$\cdots$
$\det A$ (即n阶主子式) $= \prod_i \lambda_i$.

$p_A(\lambda)$ 的系数其实是关于 $\{\lambda_1,\ldots,\lambda_n\}$ 的各个基本对称多项式, 亦可以写成关于 $\{\mathrm{Tr}\, A,\ldots,\mathrm{Tr}\,(A^n)\}$ 的多项式. 它们构成了关于 A 的线性变换的一些不变量.

内容

定义

如果方阵 A 相似于对角矩阵, 即 $\exists$ 可逆矩阵 P, s.t. $A = PDP^{-1}$, 其中 D 为对角矩阵, 则称 A 可对角化.

$n \times n$ 矩阵 A 可对角化的充要条件是 A 有 n 个线性无关的特征向量.

这些向量组成了矩阵 P 的列向量, 其相应的特征值组成了对角矩阵 D 的主对角线上的元素.

即由 $A\vec{v}_i = \lambda_i \vec{v}_i,\ i = 1, \ldots, n$, $\{\vec{v}_1, \ldots, \vec{v}_n\}$ 线性无关, 有

$$A\begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_n\end{bmatrix} = \begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix} \Rightarrow A = PDP^{-1}.$$

对角化的步骤

- 利用特征多项式求出特征值.
- 把特征值逐个代入 $(A-\lambda I)\vec{x}=\vec{0}$, 求出特征向量.
- 利用特征向量构造矩阵 P, 用对应的特征值构造矩阵 D.
- $A=PDP^{-1}$.

矩阵可以对角化的一些判断

设 A 是 $n \times n$ 的矩阵, 若 A 可以对角化,

- 一个充分条件是 A 有 n 个互异的特征值.

定理 (互异特征值的特征向量线性无关)

设 $\lambda_1, \cdots, \lambda_r$ 是 $n \times n$ 矩阵 A 相异的特征值, 对应的特征向量为 $\vec{v}_1, \cdots, \vec{v}_r$, 则向量组 $\{\vec{v}_1, \cdots, \vec{v}_r\}$ 线性无关.

- 对于代数重数 > 1 的特征值 λ, 要求其特征子空间的维数 $\dim V_\lambda =$ 代数重数.

  $\dim V_\lambda$ 又称之为特征值 λ 的几何重数. 一般的, 对于每个特征值我们有几何重数 $\leq$ 代数重数.

- 充要条件: 对于每个特征值, 要有几何重数 $=$ 代数重数.

总是可以对角化的一些特殊矩阵

例 (实对称矩阵: $A^T = A$)

- 特征值总是实的;
- 属于不同特征值的特征向量是正交的;
- 可以用正交矩阵对角化, 即 $A = ODO^{-1}$; 其中 O 是正交矩阵, 即满足 $O^T O = I$.

内容

回顾: 坐标向量 – 向量在不同基下的表示

设 $\{\vec{a}_1, \cdots, \vec{a}_n\}$ 及 $\{\vec{b}_1, \cdots, \vec{b}_n\}$ 是 n 维线性空间 V 的两个基, 过渡矩阵为 P ($n \times n$ 的可逆矩阵), 即 $[\vec{b}_1 \ \cdots] = [\vec{a}_1 \ \cdots]P$; $\vec{v}$ 是 V 中的一个向量, 有

$$\vec{v} = [\vec{a}_1 \ \cdots][\vec{v}]_\alpha = [\vec{b}_1 \ \cdots][\vec{v}]_\beta.$$

$[\vec{v}]_\alpha$ ($[\vec{v}]_\beta$) 称为向量 $\vec{v}$ 在 α-基 (β-基) 下的坐标向量.

向量的表示依赖于所选取的基.

线性变换的矩阵表示

设有 V 上的线性变换 $T: V \to V$. 对于 $\vec{v} \in V$, 有

$$T(\vec{v}) = T\left([a_1 \ \cdots][\vec{v}]_\alpha\right) = [\,T(a_1) \ \cdots][\vec{v}]_\alpha \equiv [a_1 \ \cdots]\, A\, [\vec{v}]_\alpha.$$

A 称为线性变换在 α-基下的矩阵表示. 类似的有 β-基下的矩阵表示 B:

$$T(\vec{v}) = T\left([b_1 \ \cdots][\vec{v}]_\beta\right) = [\,T(b_1) \ \cdots][\vec{v}]_\beta \equiv [b_1 \ \cdots]\, B\, [\vec{v}]_\beta.$$

线性变换的矩阵表示

设有 V 上的线性变换 $T: V \to V$. 对于 $\vec{v} \in V$, 有

$$T(\vec{v}) = T\left([a_1 \ \cdots][\vec{v}]_\alpha\right) = [T(a_1) \ \cdots][\vec{v}]_\alpha \equiv [a_1 \ \cdots]\, A\, [\vec{v}]_\alpha.$$

A 称为线性变换在 α-基下的矩阵表示. 类似的有 β-基下的矩阵表示 B:

$$T(\vec{v}) = T\left([b_1 \ \cdots][\vec{v}]_\beta\right) = [T(b_1) \ \cdots][\vec{v}]_\beta \equiv [b_1 \ \cdots]\, B\, [\vec{v}]_\beta.$$

Q: A 与 B 的联系?

线性变换的矩阵表示

设有 V 上的线性变换 $T: V \to V$. 对于 $\vec{v} \in V$, 有

$$T(\vec{v}) = T([a_1 \ \cdots][\vec{v}]_\alpha) = [T(a_1) \ \cdots][\vec{v}]_\alpha \equiv [a_1 \ \cdots] A [\vec{v}]_\alpha.$$

A 称为线性变换在 α-基下的矩阵表示. 类似的有 β-基下的矩阵表示 B:

$$T(\vec{v}) = T([b_1 \ \cdots][\vec{v}]_\beta) = [T(b_1) \ \cdots][\vec{v}]_\beta \equiv [b_1 \ \cdots] B [\vec{v}]_\beta.$$

Q: A 与 B 的联系?

A: $A = PBP^{-1}$.

线性变换的矩阵表示

设有 V 上的线性变换 $T: V \to V$. 对于 $\vec{v} \in V$, 有

$$T(\vec{v}) = T([a_1 \ \cdots][\vec{v}]_\alpha) = [T(a_1) \ \cdots][\vec{v}]_\alpha \equiv [a_1 \ \cdots]\,A\,[\vec{v}]_\alpha.$$

A 称为线性变换在 α-基下的矩阵表示. 类似的有 β-基下的矩阵表示 B:

$$T(\vec{v}) = T([b_1 \ \cdots][\vec{v}]_\beta) = [T(b_1) \ \cdots][\vec{v}]_\beta \equiv [b_1 \ \cdots]\,B\,[\vec{v}]_\beta.$$

Q: A 与 B 的联系?

A: $A = PBP^{-1}$.

我们称 A 与 B 是相似的, 它们通过过渡矩阵 P 联系起来. 把 A 对角化相当于寻找相应变换的对角矩阵表示. 与 A 相似的所有矩阵可以看成是同一线性变换在不同基下的表示.

相似矩阵

定义

两个 $n \times n$ 的矩阵 A 和 B, 如果存在可逆矩阵 P 使得 $P^{-1}AP = B$, 我们就说 A 和 B 是相似的, 可记为 $A \sim B$.

- 相似关系是一个等价关系.
- 同一个线性变换在不同基下的矩阵表示是相似的, 它们通过两个基之间的过渡矩阵联系起来.
- 相似矩阵具有相同的特征多项式.
- 线性变换的不变量不依赖于所选取的表示, 如特征值、特征多项式等.

例

Q: 设有 $\mathbb{R}^3$ 的两个基 $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ 及 $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$, 有线性变换 T, 满足

$$T(\vec{v}_1) = \vec{w}_1 + \vec{w}_2 + \vec{w}_3, \quad T(\vec{v}_2) = \vec{w}_2 + \vec{w}_3, \quad T(\vec{v}_3) = \vec{w}_3.$$

已知 $T(\vec{v}) = \vec{w}_1$, 求 $\vec{v}$.

A: 引入矩阵 A 形如

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

注意到我们有 $A[\vec{v}]_v = [T(\vec{v})]_w$. 由 $T(\vec{v}) = \vec{w}_1$, 有

$$A\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

其中 $\vec{x} = [\vec{v}]_v$. 解得 $\vec{x} = (1, -1, 0)$, 即 $\vec{v} = \vec{v}_1 - \vec{v}_2$.

内容

示例: $\mathbb{R}^2$ 中的旋转变换

旋转变换

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

特征值 $\lambda_\pm = e^{\pm i\theta}$, 特征向量 $\vec{v}_+ = (1, -i)$, $\vec{v}_- = (-i, 1)$,

$$A = \frac{1}{2} \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix} \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$$

事实上, 对于形如 $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ 的实矩阵可以看成是旋转 + 倍乘的复合变换.

有复特征值的实矩阵

对于有复特征值的 2×2 的实矩阵 A, 我们有

$$A\vec{v} = \lambda\vec{v}, \qquad A\vec{v}^{*} = \lambda^{*}\vec{v}^{*}.$$

由于 $\lambda \neq \lambda^{*}$, 则 $\vec{v}$, $\vec{v}^{*}$ 线性无关, 或者说 $\mathrm{Re}\,\vec{v}$ 和 $\mathrm{Im}\,\vec{v}$ 线性无关.

$$A\left(\mathrm{Re}\,\vec{v} + i\,\mathrm{Im}\,\vec{v}\right) = \left(\lambda_1 + i\lambda_2\right)\left(\mathrm{Re}\,\vec{v} + i\,\mathrm{Im}\,\vec{v}\right) \Rightarrow$$

$$\begin{cases} A\,\mathrm{Re}\,\vec{v} = \lambda_1 \mathrm{Re}\,\vec{v} - \lambda_2 \mathrm{Im}\,\vec{v}, \\ A\,\mathrm{Im}\,\vec{v} = \lambda_2 \mathrm{Re}\,\vec{v} + \lambda_1 \mathrm{Im}\,\vec{v} \end{cases} \quad \text{或} \quad A\left[\mathrm{Re}\,\vec{v}\ \mathrm{Im}\,\vec{v}\right] = \left[\mathrm{Re}\,\vec{v}\ \mathrm{Im}\,\vec{v}\right]\begin{bmatrix} \lambda_1 & \lambda_2 \\ -\lambda_2 & \lambda_1 \end{bmatrix}$$

A 对应的线性变换在 $\{\mathrm{Re}\,\vec{v}, \mathrm{Im}\,\vec{v}\}$ 基下的是一个简单的旋转 + 倍乘的复合变换.

内容

差分方程 $\vec{x}_{k+1} = A\vec{x}_k$

- 若 $\vec{x}_0$ 是 A 的特征向量, 则 $\vec{x}_k = \lambda^k \vec{x}_0$ 满足方程.
- 设 A 可以对角化, 特征值 $|\lambda_1| > |\lambda_2| \cdots \geq |\lambda_n|$, 对应的特征向量 $\{\vec{v}_1, \cdots, \vec{v}_n\}$ 构成 $\mathbb{R}^n$ 的一个基.

  对于给定初态 $\vec{x}_0$, 有 $\vec{x}_0 = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n$, 则

$$\begin{aligned} \vec{x}_k &= A^k\vec{x}_0 = c_1\lambda_1^k\vec{v}_1 + \cdots + c_n\lambda_n^k\vec{v}_n \\ &= \lambda_1^k\left[c_1\vec{v}_1 + \cdots + c_n\left(\lambda_n/\lambda_1\right)^k\vec{v}_n\right] \\ &\xrightarrow{k\to\infty} \vec{v}_1 \end{aligned}$$

本章作业

- 5.1 节: 5, 13, 23, 25, 29, 31, 33
- 5.2 节: 11, 13, 17, 27
- 5.3 节: 3, 5, 15, 17, 21, 27
- 5.4 节: 5, 11, 13, 15, 19
- 5.5 节: 5, 11, 17

课堂练习

- 5.1 节: 3, 9, 21, 27
- 5.2 节: 9, 19
- 5.3 节: 1, 13, 23
- 5.4 节: 3, 9, 25
- 5.5 节: 23, 24