第七章　对称矩阵与二次型

2022 年 1 月 5 日

动机

- 对称矩阵 $A^T = A$ 理论形式优美, 经常在各种应用中出现, 如二次型, 二次曲线, 对称变换, 简正坐标与简正模...
- 其在复数域的推广: 厄密矩阵 (厄密矩阵 A 满足 $A = A^\dagger := \left(A^T\right)^*$) 在量子力学中有广泛应用.

在下面的讨论中, 所有向量和矩阵元素均为实数.

内容

谱定理

A 是一个 $n \times n$ 的对称矩阵, 我们有

- A 的所有特征值都是实的.
- 不同特征值的特征子空间相互正交.
- 对于每个特征值, 其几何重数 $=$ 代数重数.
- 可以正交对角化, 即有 $A = ODO^{-1}$, 其中 D 为实对角矩阵, O 为正交矩阵.

舒尔 (Schur) 分解

定理

若 $n \times n$ 矩阵 A 有 n 个实特征值 (包含重数), 则有舒尔分解 $A = URU^T$, 其中 U 是正交矩阵, R 是上三角矩阵, 其对角元为 A 的特征值.

参看教材第六章补充习题 15, 16

从舒尔分解定理出发, 如果进一步假设 A 是对称矩阵, 由 $A^T = A$ 可知 R 应该是对角的. 关于对称矩阵正交对角化的结论便可以得到证明.

谱分解

A 是对称矩阵, 有 $A = ODO^{-1} = ODO^T$, 正交矩阵 $O = [\vec{u}_1 \ \cdots \ \vec{u}_n]$, D 是对角矩阵. 展开,

$$
\begin{aligned}
A &= ODO^T = \begin{bmatrix} \lambda_1 \vec{u}_1 & \cdots & \lambda_n \vec{u}_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix} \\
&= \lambda_1 \vec{u}_1 \vec{u}_1^T + \cdots + \lambda_n \vec{u}_n \vec{u}_n^T & \text{(谱分解)} \\
&= \lambda_1 |1\rangle\langle 1| + \cdots + \lambda_n |n\rangle\langle n|. & \text{(狄拉克记号)}
\end{aligned}
$$

$\vec{u}_i \vec{u}_i^T$ 是投影矩阵, $(\vec{u}_i \vec{u}_i^T)\vec{x}$ 给出了 $\vec{x}$ 在 $\vec{u}_i$ 方向上的正交投影.

投影矩阵

定义

一个 $n \times n$ 的矩阵 B 称为投影矩阵, 如果它满足以下两个条件:

- $B^2 = B$;
- $B^T = B$.

$\forall \vec{y} \in \mathbb{R}^n$, $B\vec{y}$ 就是 $\vec{y}$ 在 $\mathrm{Col}\, B$ 上的正交投影.

证明: 只需要证明 $B^T(\vec{y} - B\vec{y}) = \vec{0}$ 即可.

- 对于一个单位向量 $\vec{u}$, $\vec{u}\vec{u}^T$ 是投影矩阵.
- 一般的, 对于一个具有单位正交列的矩阵 Q, QQ^T 是一个投影矩阵: $QQ^T = \mathrm{proj}_{\mathrm{Col}\, Q}$.

对称矩阵的一些性质

A 是 $n \times n$ 的对称矩阵

- $(\mathrm{Col}\,A)^{\perp} = \mathrm{Nul}\,A^T = \mathrm{Nul}\,A \;\Rightarrow\; \mathbb{R}^n = \mathrm{Col}\,A \oplus \mathrm{Nul}\,A.$
- 非零特征值对应的特征向量生成 A 的列空间 $\mathrm{Col}\,A$, 零特征值的特征向量生成 $\mathrm{Nul}\,A$.
- A 的非零特征值的数目 $= \mathrm{rank}\,A$.
- A 是正定矩阵的充要条件是 $A = R^T R$, 其中 R 是可逆上三角矩阵, 且主对角元是正的.

  参看教材第七章补充习题 7

内容

A 是 $n \times n$ 的对称矩阵, 形如 $\vec{x}^T A \vec{x}$ 的表达式称为一个二次型.

- 二次型定义了一个 $\mathbb{R}^n \to \mathbb{R}$ 的映射.
- 二次型与对称矩阵的对应关系: $\vec{x}^T A \vec{x} \longleftrightarrow A$.
- (主轴定理) 存在变量的正交变换 $\vec{x} = O\vec{y}$, 它将二次型 $\vec{x}^T A \vec{x}$ 变换为不含交叉项的二次型 $\vec{y}^T D \vec{y}$.
- 上述矩阵 O 的列向量构成了二次型 $\vec{x}^T A \vec{x}$ 的主轴, 向量 $\vec{y}$ 是向量 $\vec{x}$ 在主轴正交坐标系下的坐标向量.

二次型的分类

定义

一个二次型 Q 是:

- 正定的: $\forall \vec{x} \neq \vec{0},\ Q(\vec{x}) > 0.$
- 负定的: $\forall \vec{x} \neq \vec{0},\ Q(\vec{x}) < 0.$
- 不定的: $Q(\vec{x})$ 既有正值又有负值.

二次型分类与对应矩阵 A 特征值之间的关系:

- $Q(\vec{x})$ 正定, 当且仅当 A 的所有特征值是正数 $\Rightarrow$ 正定矩阵.
- $Q(\vec{x})$ 负定, 当且仅当 A 的所有特征值是负数 $\Rightarrow$ 负定矩阵.
- $Q(\vec{x})$ 不定, 当且仅当 A 的特征值既有正的, 又有负的.

惯性指数

二次型 $Q(\vec{x}) = \vec{x}^T A\vec{x}$ 在坐标 (表象) 变换 $\vec{x} = P\vec{y}$ 下的变换:
$Q = \vec{y}^T \left(P^T AP\right) \vec{y}$, 即 $A \to P^T AP$. 我们称 A 作了一个合同变换.

在合同变换下, 矩阵的特征值会变. 但另一方面, 可以证明, 在合同变换下大于零, 小于零及等于零的特征值数目不会发生变化. 我们把大于 (小于) 零的特征值数目称为其正 (负) 惯性指数, 它们是关于这个二次型的不变量.

内容

条件优化

Q: 在 $\|\vec{x}\| = 1$ 的条件下，二次型 $Q(\vec{x}) = \vec{x}^T A\vec{x}$ 的极值.

A: $Q(\vec{x})$ 的极值在沿主轴方向上达到.

注：二次型 $Q = Q(\vec{x}) = \vec{x}^T A\vec{x}$ 可以通过正交变换 $\vec{x} = O\vec{y}$ 变成对角的形式:
$Q = Q(\vec{y}) = \vec{y}^T D\vec{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$. 由约束 $\|\vec{x}\| = 1 = \|\vec{y}\|$, 我们可以看到对于固定长度的向量，二次型 Q 在沿主轴方向上取极值.

在这里我们可以看到二次型的极值问题和相应的对称矩阵的特征值问题是有密切联系的.

关于极值的定理

定理

A 是一个 $n \times n$ 的对称矩阵, 其正交对角化的形式为 $A = ODO^{-1}$, 设对角矩阵 D 的元素按照降序排列 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, 在 O 中对应的列向量为 $\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_n$. 那么对 $k = 2, \cdots, n$ 时, 在以下限制条件下:

$$\vec{x}^T\vec{x} = 1, \quad \vec{x}^T\vec{u}_1 = 0, \ \cdots \ \vec{x}^T\vec{u}_{k-1} = 0$$

二次型 $Q(\vec{x}) = \vec{x}^TA\vec{x}$ 的最大值是特征值 λ_k, 且这个最大值在 $\vec{x} = \vec{u}_k$ 处可以达到.

注: 这个定理可以推广到更一般的情况. 例如我们可以证明在 $\vec{x}^T\vec{x} = 1$ 及其它任意$k-1$ 个齐次线性约束下, 二次型 $Q(\vec{x})$ 的最大值不会小于 λ_k.

内容

动机及思想

- 对于一般的 $m \times n$ 的矩阵 $A: \mathbb{R}^n \to \mathbb{R}^m$, 变换 $\vec{x} \mapsto A\vec{x}$ 的全部几何特性都可用二次型 A^TA 来说明.

  比如说 $A\vec{x}$ 的长度: 由 $\|A\vec{x}\|^2 = \vec{x}^TA^TA\vec{x}$, 可知使得 $A\vec{x}$ 最长的单位向量 $\vec{x}$ 应沿着 A^TA 决定的第一主轴方向, 即其最大特征值 λ_1 对应的特征向量方向. 这时 $\|A\vec{x}\| = \sigma_1\|\vec{x}\|$. $\sigma_i := \sqrt{\lambda_i}$ 称为 A 的一个奇异值.

- 根据 A 的特性我们可以同时把两个空间 $\mathbb{R}^n$ 和 $\mathbb{R}^m$ 分解成一些空间的直和.

- 奇异值分解: $A = U\Sigma V^T$, 其中 U 和 V 分别是 $m \times m$ 和 $n \times n$ 的正交矩阵, $\Sigma_{m\times n}$ 是类对角矩阵, 非零对角元为 A 的奇异值 σ_i, $i = 1, \cdots$.

A^TA 对空间的分解

设 $\{\vec{v}_1, \cdots, \vec{v}_n\}$ 是由 A^TA 的特征向量构成的 $\mathbb{R}^n$ 的一个标准正交基, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 是特征值, 其中 r 个是非零的, 即 $\lambda_1 \neq 0, \ \cdots \ , \lambda_r \neq 0$, 而 $\lambda_{r+1} = \cdots = \lambda_n = 0$. 我们有

- $\{A\vec{v}_1, \cdots, A\vec{v}_r\}$ 生成 $\mathrm{Col}\, A$, 且 $\{A\vec{v}_1, \cdots, A\vec{v}_r\}$ 是其正交基.
- $\mathrm{rank}\, A = r$, r 为非零奇异值的数目.
- $\mathrm{Col}\, A$ 基的单位化: $\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i$, $i = 1, \cdots, r$.
- $\{\vec{u}_1, \cdots, \vec{u}_r, \vec{u}_{r+1}, \cdots, \vec{u}_m\}$ 是 $\mathbb{R}^m$ 的一个标准正交基. 有 $\mathrm{Span}\, \{\vec{u}_1, \cdots, \vec{u}_r\} = \mathrm{Col}\, A$, 及 $\mathrm{Span}\, \{\vec{u}_{r+1}, \cdots, \vec{u}_m\} = (\mathrm{Col}\, A)^{\perp} = \mathrm{Nul}\, A^T$.

至此, 对 $\mathbb{R}^m$ 空间的分解已经完成: $\mathbb{R}^m = \mathrm{Col}\, A \oplus \mathrm{Nul}\, A^T$.

A^TA 对空间的分解 (续)

- 注意到 $A\vec{v}_{r+1} = \cdots = A\vec{v}_n = \vec{0}$, 即有 $\mathrm{Span}\,\{\vec{v}_{r+1}, \cdots \vec{v}_n\} = \mathrm{Nul}\,A$.
- $\mathrm{Span}\,\{\vec{v}_1, \cdots, \vec{v}_r\} = (\mathrm{Nul}\,A)^{\perp} = \mathrm{Row}\,A$.

至此, 对 $\mathbb{R}^n$ 空间的分解已经完成: $\mathbb{R}^n = \mathrm{Row}\,A \oplus \mathrm{Nul}\,A$. $\{\vec{v}_1, \cdots, \vec{v}_r\}$ 和 $\{\vec{v}_{r+1}, \cdots, \vec{v}_n\}$ 分别是其标准正交基.

奇异值分解

定理

设 A 是秩为 r 的 $m \times n$ 的矩阵, 那么存在一个类对角的 $m \times n$ 矩阵 Σ, 其对角元素是 A 的前 r 个奇异值, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, 及一个 $m \times m$ 的正交矩阵 U 和一个 $n \times n$ 的正交矩阵 V, 使得 $A = U\Sigma V^T$.

实际上, 矩阵 U 即为前面的 $[\vec{u}_1 \ \cdots \ \vec{u}_m]$, 矩阵 V 即为前面的 $[\vec{v}_1 \ \cdots \ \vec{v}_n]$, 类对角矩阵 Σ 形如 $\Sigma = \begin{bmatrix} D_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}$, 其中 D 是 $r \times r$ 的对角矩阵, 对角元为 $\sigma_1, \cdots, \sigma_r$.

本章作业

课堂练习