第四章 矩阵分解

第 12 讲 谱分解与 Cholesky 分解

黄定江

DaSE @ ECNU

djhuang@dase.ecnu.edu.cn

引入：本讲主要内容

本讲主要讨论两类特殊的矩阵：对称矩阵和正半定矩阵的分解。

- 对称矩阵的谱分解（特征分解）：可以把任意对称矩阵分解成三个矩阵的积，包括一个正交矩阵和一个实的对角矩阵。
- 正定矩阵的 Cholesky 分解：可以把任意对称正定矩阵分解成一个具有正的对角元的下三角矩阵和其转置的乘积。

特征值与物理或力学中振动的频谱相联系，所以特征分解也称为谱分解！

引入：特征分解和 Cholosky 分解的应用

机器学习、人工智能和信号处理中的应用

- 机器学习：PCA 降维
- 数据压缩：图像压缩

机器学习、人工智能中矩阵函数模型的性质和计算应用

- 确定二次函数的凸性
- 分解二次函数为带有正交向量的线性函数的平方和
- 根据某些二次优化问题的最优值刻画对称矩阵特征值的特性
- 降低求逆求行列式等的数值计算复杂性

12.1.1 对称矩阵的谱分解：矩阵的特征分解和对称矩阵性质回顾

矩阵的特征分解定理

一个矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 可以分解为 $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$, 其中 $\boldsymbol{P}$ 是由特征向量构成的可逆矩阵，$\boldsymbol{D}$ 是对角矩阵且对角元是 $\boldsymbol{A}$ 的特征值，当且仅当 $\boldsymbol{A}$ 有 n 个线性无关的特征向量。

关于对称矩阵特征值特征向量的有用性质

- 对称矩阵总是具有实特征值
- 对称矩阵的不同特征值对应的特征向量是相互正交的

通过以下定理的证明，我们可以给这两个性质保障。

定理 1

设 $\boldsymbol{A}$ 是实对称矩阵，则 $\boldsymbol{A}$ 的特征值皆为实数。

证明.

设 λ_0 是 $\boldsymbol{A}$ 的特征值，于是有非零向量 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 满足

$$\boldsymbol{A}\boldsymbol{x} = \lambda_0 \boldsymbol{x}$$

令

$$\overline{\boldsymbol{x}} = (\overline{x}_1, \overline{x}_2, \cdots, \overline{x}_n)$$

其中 $\overline{x}_i$ 是 x_i 的共轭复数，则 $\overline{\boldsymbol{A}\boldsymbol{x}} = \overline{\lambda}_0 \overline{\boldsymbol{x}}$。考察等式

$$\overline{\boldsymbol{x}}^T(\boldsymbol{A}\boldsymbol{x}) = \overline{\boldsymbol{x}}^T \boldsymbol{A}^T \boldsymbol{x} = (\boldsymbol{A}\overline{\boldsymbol{x}})^T \boldsymbol{x} = (\overline{\boldsymbol{A}\boldsymbol{x}})^T \boldsymbol{x}$$

其左边为 $\lambda_0 \overline{\boldsymbol{x}}^T \boldsymbol{x}$，右边为 $\overline{\lambda}_0 \overline{\boldsymbol{x}}^T \boldsymbol{x}$。故

$$\lambda_0 \overline{\boldsymbol{x}}^T \boldsymbol{x} = \overline{\lambda}_0 \overline{\boldsymbol{x}}^T \boldsymbol{x}$$

又因 $\boldsymbol{x}$ 是非零向量

$$\overline{\boldsymbol{x}}^T \boldsymbol{x} = \overline{x}_1 x_1 + \overline{x}_2 x_2 + \cdots + \overline{x}_n x_n \neq \boldsymbol{0}$$

故 $\lambda_0 = \overline{\lambda}_0$，即 λ_0 是一个实数。证毕。 □

谱分解定理

定理 2

设实矩阵 $\boldsymbol{A}$ 是 n 阶方阵，则下面 3 个命题等价：

1. $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$。
2. 存在一个正交矩阵 $\boldsymbol{Q}$ 使得 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{\Lambda}$，其中 $\boldsymbol{\Lambda}$ 是对角阵。
3. 存在 n 个 $\boldsymbol{A}$ 的特征向量构成 $\mathbb{R}^n$ 的一个标准正交基。

谱分解定理证明

证明：1 $\rightarrow$ 2：

利用数学归纳法，当 $n=1$ 时，易知 1 推 2 成立。

假设当 $k=n-1$ 时结论成立。

设 λ_1 是 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ 的一个特征值，对应的特征向量为 $\boldsymbol{q}_1$，即 $\boldsymbol{A}\boldsymbol{q}_1=\lambda_1\boldsymbol{q}_1$，且我们令 $||\boldsymbol{q}_1||=1$。

我们可以将 $(\boldsymbol{q}_1)$ 扩充成一个标准正交基 $(\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots,\boldsymbol{q}_n)$ 从而得到了一个正交矩阵 $\boldsymbol{Q}$，又因为 $\boldsymbol{A}\boldsymbol{q}_i\in\mathbb{R}^n$ 可以由 $\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots,\boldsymbol{q}_n$ 线性表出，并且 $\boldsymbol{A}\boldsymbol{q}_1=\lambda_1\boldsymbol{q}_1$，那么

$$\boldsymbol{A}\boldsymbol{Q}=\boldsymbol{A}[\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots,\boldsymbol{q}_n]=[\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots,\boldsymbol{q}_n]\begin{pmatrix}\lambda_1 & \boldsymbol{a}^{\mathrm{T}}\\ \boldsymbol{0} & \boldsymbol{C}\end{pmatrix}=\boldsymbol{Q}\boldsymbol{A}_1$$

谱分解定理证明

$$\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{A}[\boldsymbol{q}_1, \boldsymbol{q}_2, ..., \boldsymbol{q}_n] = [\boldsymbol{q}_1, \boldsymbol{q}_2, ..., \boldsymbol{q}_n]\begin{pmatrix} \lambda_1 & \boldsymbol{a}^{\mathrm{T}} \\ \boldsymbol{0} & \boldsymbol{C} \end{pmatrix} = \boldsymbol{Q}\boldsymbol{A}_1$$

故 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{A}_1\boldsymbol{Q}^{\mathrm{T}}$，$\boldsymbol{A}_1 = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}$。$\boldsymbol{A}$ 对称，所以 $\boldsymbol{A}_1$ 对称，从而 $\boldsymbol{a}^{\mathrm{T}} = \boldsymbol{0}^{\mathrm{T}}, \boldsymbol{C}$ 对称。
根据假设，存在正交矩阵 $\boldsymbol{Q}_1$ 使得 $\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{C}\boldsymbol{Q}_1 = \boldsymbol{\Lambda}_1$，$\boldsymbol{C} = \boldsymbol{Q}_1\boldsymbol{\Lambda}_1\boldsymbol{Q}_1^{\mathrm{T}}$，

$$\boldsymbol{A}_1 = \begin{pmatrix} \lambda_1 & \\ & \boldsymbol{C} \end{pmatrix} = \begin{pmatrix} 1 & \\ & \boldsymbol{Q}_1 \end{pmatrix}\begin{pmatrix} \lambda_1 & \\ & \boldsymbol{\Lambda}_1 \end{pmatrix}\begin{pmatrix} 1 & \\ & \boldsymbol{Q}_1^{\mathrm{T}} \end{pmatrix}$$

即令 $\boldsymbol{Q}_2 = \boldsymbol{Q}\begin{pmatrix} 1 & \\ & \boldsymbol{Q}_1 \end{pmatrix}$，$\begin{pmatrix} \lambda_1 & \\ & \boldsymbol{\Lambda}_1 \end{pmatrix} = \boldsymbol{\Lambda}$ 即有 $\boldsymbol{A} = \boldsymbol{Q}_2\boldsymbol{\Lambda}\boldsymbol{Q}_2^{\mathrm{T}}$，$\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_2 = \boldsymbol{\Lambda}$。

谱分解定理证明

2→3: 将 $\boldsymbol{Q}$ 按列分块 $\boldsymbol{Q}=[\boldsymbol{q}_1,\boldsymbol{q}_2,...,\boldsymbol{q}_n]$，记 $\boldsymbol{\Lambda}=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}$

$$\boldsymbol{A}\boldsymbol{Q}=[\boldsymbol{A}\boldsymbol{q}_1,\boldsymbol{A}\boldsymbol{q}_2,...,\boldsymbol{A}\boldsymbol{q}_n]=\boldsymbol{Q}\boldsymbol{\Lambda}=[\lambda_1\boldsymbol{q}_1,\lambda_2\boldsymbol{q}_2,...,\lambda_n\boldsymbol{q}_n]$$

故，$\boldsymbol{q}_1,\boldsymbol{q}_2,...,\boldsymbol{q}_n$ 是矩阵 $\boldsymbol{A}$ 的特征向量。

谱分解定理证明

3→1: 令 $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ 是矩阵 $\boldsymbol{A}$ 的 n 个两两正交且模为 1 的特征向量。
则 $\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n]$ 是一正交矩阵。
记 $\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$
其中 λ_i 是 $\boldsymbol{q}_i$ 对应的特征值。

$$\boldsymbol{A}\boldsymbol{Q} = [\boldsymbol{A}\boldsymbol{q}_1, \boldsymbol{A}\boldsymbol{q}_2, \ldots, \boldsymbol{A}\boldsymbol{q}_n] = [\lambda_1\boldsymbol{q}_1, \lambda_2\boldsymbol{q}_2, \ldots, \lambda_n\boldsymbol{q}_n] = \boldsymbol{Q}\boldsymbol{\Lambda}$$

故 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}$
$\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{Q}\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{A}$

谱分解定义

定义 1

设对称矩阵 $\boldsymbol{A}$ 为 n 阶方阵，如果 $\boldsymbol{A}$ 可以被分解为 $\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}$，其中 $\boldsymbol{Q}=[\boldsymbol{q}_1,\boldsymbol{q}_2,\dots,\boldsymbol{q}_n]$ 是由特征向量 $\boldsymbol{q}_1,\boldsymbol{q}_2,\dots,\boldsymbol{q}_n$ 组成的 n 阶方阵，$\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,...,\lambda_n)$ 是由特征值 $\lambda_1,\lambda_2,...,\lambda_n$ 组成的 n 阶对角矩阵，则这种分解叫做对称矩阵的**谱分解**或者**特征分解**。

秩 -1 矩阵和

我们也可以将 $\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}$ 改写成秩 -1 矩阵和的形式：

$$\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}=\sum_{i=1}^{n}\lambda_i\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}}$$

。

求解对称方阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 的特征分解步骤

- 计算矩阵 $\boldsymbol{A}$ 的特征值 $\lambda_1, \ldots, \lambda_n$，即求特征方程
$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0$$
的 n 个根。
- 求特征值对应的 n 个相互正交的特征向量 $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$，即求解方程组并单位化
$$\boldsymbol{A}\boldsymbol{q}_i = \lambda_i \boldsymbol{q}_i, i = 1, \ldots, n$$
- 记矩阵 $\boldsymbol{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)$。
- 最终得到矩阵 $\boldsymbol{A}$ 的特征分解为
$$\boldsymbol{A} = \boldsymbol{Q} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \boldsymbol{Q}^T$$

例 1

求实对称矩阵 $\boldsymbol{A}=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}$ 的特征分解.

(1) 计算特征值和正交单位特征向量.

$|\lambda\boldsymbol{I}-\boldsymbol{A}|=\begin{vmatrix}\lambda-2 & -1\\ -1 & \lambda-2\end{vmatrix}=0$

特征值

$$\lambda_1=3,\lambda_2=1$$

对应的特征向量通过求解

$$\boldsymbol{A}\boldsymbol{q}_1=3\boldsymbol{q}_1,\boldsymbol{A}\boldsymbol{q}_2=\boldsymbol{q}_2,$$

并单位化得到，所以有

$$\boldsymbol{q}_1=(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}})^T,\boldsymbol{q}_2=(-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}})^T$$

(2) 写出左特征向量方阵 $\boldsymbol{Q}$ 和特征值方阵 $\boldsymbol{\Lambda}$.

$$\boldsymbol{Q}=[\boldsymbol{q}_1,\boldsymbol{q}_2]=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix},\qquad \boldsymbol{\Lambda}=\begin{bmatrix}\lambda_1 & \\ & \lambda_2\end{bmatrix}=\begin{bmatrix}3 & \\ & 1\end{bmatrix}$$

又因为 $\boldsymbol{A}$ 是实对称矩阵，所以 $\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}3 & \\ & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$.

(3) 我们也可以将其改写成秩 -1 矩阵和的形式：

$$\boldsymbol{A}=3\cdot\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}+1\cdot\begin{bmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}=3\cdot\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}+\begin{pmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{pmatrix}$$

几何理解

从线性空间的角度看，在一个定义了内积的线性空间里，对一个 N 阶对称方阵进行特征分解，就产生了该空间的 N 个标准正交基，矩阵对应的变换将空间中的向量投影到这 N 个基上。N 个特征向量就是 N 个标准正交基，而特征值的模则代表向量在每个基上的投影长度的伸缩倍数。

12.1.2 谱分解与优化

实矩阵的特征分解可以用于优化二次方程 $Q(\boldsymbol{x}) = \boldsymbol{x}^\mathsf{T}\boldsymbol{A}\boldsymbol{x}$，其中限制 $\|\boldsymbol{x}\|_2 = 1$。当 $\boldsymbol{x}$ 等于 $\boldsymbol{A}$ 的某个特征向量时，$Q(\boldsymbol{x})$ 为对应的特征值。

在限制条件下，函数 $Q(\boldsymbol{x})$ 的最大值是最大特征值，最小值是最小特征值：

- 如果 $\boldsymbol{A}$ 是正半定矩阵则 $\min\limits_{||\boldsymbol{x}||_2=1} Q(\boldsymbol{x}) \geq 0$
- 如果 $\boldsymbol{A}$ 是负半定矩阵则 $\max\limits_{||\boldsymbol{x}||_2=1} Q(\boldsymbol{x}) \leq 0$

例如，我们知道无向图的 Laplacian 矩阵 $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ 是正半定的，它有一个特征值是 0，从而

$$\min_{\boldsymbol{x}:\|\boldsymbol{x}\|_2=1} \boldsymbol{x}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{x} = 0.$$

瑞利商

定义 2

设矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 为一对称阵，比率式 $R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}, (0 \neq \boldsymbol{x} \in \mathbb{R}^n)$ 被称为瑞利商。

定理 3

给定一个对称矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$，瑞利商 $R(\boldsymbol{x})$ 有如下性质：

$$\lambda_{min}(\boldsymbol{A}) \leq R(\boldsymbol{x}) \leq \lambda_{max}(\boldsymbol{A})$$

并且有

$$\lambda_{max}(\boldsymbol{A}) = \max_{\boldsymbol{x}:\|\boldsymbol{x}\|_2=1} \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}, \qquad \lambda_{min}(\boldsymbol{A}) = \min_{\boldsymbol{x}:\|\boldsymbol{x}\|_2=1} \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x},$$

当 $\boldsymbol{x} = \boldsymbol{u}_1$ 或 $\boldsymbol{x} = \boldsymbol{u}_n$ 时，瑞利商取到最大值或最小值，其中 $\boldsymbol{u}_1, \boldsymbol{u}_n$ 分别是最大特征值和最小特征值对应的特征向量。

证明.

矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 为一对称阵，那么设它的 n 个特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 对应的标准正交的特征向量为 $\boldsymbol{q}_1, \boldsymbol{q}_2, \dots \boldsymbol{q}_n$。

可以将 $\boldsymbol{x}$ 表示为

$$\boldsymbol{x} = a_1\boldsymbol{q}_1 + a_2\boldsymbol{q}_2 + \cdots + a_n\boldsymbol{q}_n$$

则有瑞利商分母

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = (\sum_{i=0}^{n} a_i\boldsymbol{q}_i)^{\mathrm{T}}(\sum_{j=0}^{n} a_j\boldsymbol{q}_j) = \sum_{i=0}^{n} a_i^2\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i = \sum_{i=0}^{n} a_i^2$$

而瑞利商分子

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = (\sum_{i=0}^{n} a_i\boldsymbol{q}_i)^{\mathrm{T}}\boldsymbol{A}(\sum_{j=0}^{n} a_j\boldsymbol{q}_j) = \sum_{i=0}^{n}\sum_{j=0}^{n} a_i\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{A}a_j\boldsymbol{q}_j$$
$$= \sum_{i=0}^{n}\sum_{j=0}^{n} a_ia_j\lambda_j\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = \sum_{i=0}^{n} a_i^2\lambda_i\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i = \sum_{i=0}^{n} a_i^2\lambda_i$$

证明. 续.

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \sum_{i=0}^{n} a_i^2, \qquad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \sum_{i=0}^{n} a_i^2\lambda_i$$

又 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$，所以 $\lambda_n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \leq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \leq \lambda_1\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$,

$$\lambda_{min}(\boldsymbol{A}) \leq R(\boldsymbol{x}) \leq \lambda_{max}(\boldsymbol{A})$$

当 $\boldsymbol{x} = \boldsymbol{u}_1 = a_1\boldsymbol{q}_1, (a_1 \neq 0)$ 时：

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = a_1^2, \qquad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = a_1^2\lambda_1, \qquad R(\boldsymbol{x}) = \lambda_{max}(\boldsymbol{A})$$

当 $\boldsymbol{x} = \boldsymbol{u}_n = a_n\boldsymbol{q}_n, (a_n \neq 0)$ 时：

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = a_n^2, \qquad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = a_n^2\lambda_1, \qquad R(\boldsymbol{x}) = \lambda_{min}(\boldsymbol{A})$$

如果限制 $\|\boldsymbol{x}\| = 1$，此时 $R(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$。当 $\boldsymbol{x} = \boldsymbol{q}_1$ 时，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 取到极大值 λ_{max}；当 $\boldsymbol{x} = \boldsymbol{q}_n$ 时，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 取到极小值 λ_{min}。

正 (半) 定矩阵的特征值

通过对瑞利商的讨论，我们也得到了如下结论：

$$\boldsymbol{A} \succeq 0 \iff \lambda_i(\boldsymbol{A}) \geq 0, i = 1, \cdots, n$$

$$\boldsymbol{A} \succ 0 \iff \lambda_i(\boldsymbol{A}) > 0, i = 1, \cdots, n$$

证明.

$$\boldsymbol{A} \succeq 0 \iff R(x) \geq 0 \iff \lambda_{min}(\boldsymbol{A}) \geq 0$$

也就是说，如果 $\boldsymbol{A}$ 是正半定矩阵，它的任何一个特征值都非负。

$$\boldsymbol{A} \succ 0 \iff R(x) > 0 \iff \lambda_{min}(\boldsymbol{A}) > 0$$

也就是说，如果 $\boldsymbol{A}$ 是正定矩阵，它的任何一个特征值都为正。 □

注记

矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 为一对称阵，那么 $f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 可以改写为一系列线性函数 $y_i(\boldsymbol{x}) = \boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{x}$ 的平方的加权和的形式，

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n} \lambda_i y_i(\boldsymbol{x})^2.$$

$y_i(\boldsymbol{x})$ 的权重 $\boldsymbol{q}_i$ 是**相互正交**的，就是 $\boldsymbol{A}$ 的第 i 个特征值 λ_i 对应的单位特征向量。

- n 阶对称矩阵可以记为 S^n，
- n 阶正半定矩阵可以记为 S^n_+，
- n 阶正定矩阵可以记为 S^n_{++}。

Poincare 不等式

定理3可以推广到更一般的形式，称为对称矩阵特征值的极小极大准则。

我们先证明下面这个定理：

定理 4

[Poincare 不等式] 令 $\boldsymbol{A} \in S^n$，并令 $\mathbb{V}$ 是 $\mathbb{R}^n$ 中的任意一个 k 维子空间，这里 $1 \leq k \leq n$。那么，存在单位向量 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$，$\|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2 = 1$，使得

$$\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} \leq \lambda_k(\boldsymbol{A}), \qquad \boldsymbol{y}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{y} \geq \lambda_{n-k+1}(\boldsymbol{A})$$

证明.

令 $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^{\mathrm{T}}$ 是 $\boldsymbol{A}$ 的谱分解，记 $\mathbb{Q} = \mathrm{Col}\,(\boldsymbol{U}_k)$ 是 $\boldsymbol{U}_k = [\boldsymbol{u}_k, \cdots, \boldsymbol{u}_n]$ 张成的子空间。由于 $\mathbb{Q}$ 是 $n-k+1$ 维的，$\mathbb{V}$ 维度为 k，$\mathbb{V} \cap \mathbb{Q}$ 一定是非空的。

证明. 续.

选取一个单位向量 $\boldsymbol{x} \in \mathbb{V} \cap \mathbb{Q}$。则存在 $\boldsymbol{\eta}, \|\boldsymbol{\eta}\| = 1$ 使得 $\boldsymbol{x} = \boldsymbol{U}_k\boldsymbol{\eta}$，那么

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{U}_k^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}_k\boldsymbol{\eta} = \sum_{i=k}^{n} \lambda_i(\boldsymbol{A})\eta_i^2$$

$$\leq \lambda_k(\boldsymbol{A})\sum_{i=k}^{n}\eta_i^2 = \lambda_k(\boldsymbol{A})$$

这就证明了命题中的第一个不等式。

对于第二个不等式，我们可以对 $-\boldsymbol{A}$ 用同样的处理方式即可证明。

极小极大准则 (Minimax principle)

推论 1

[极小极大准则] 令 $\boldsymbol{A} \in S^n$，并令 $\mathbb{V}$ 是 $\mathbb{R}^n$ 中的任意一个 k 维子空间，这里 $1 \le k \le n$。那么，对于 $k \in \{1, \cdots, n\}$，有

$$\begin{aligned}\lambda_k(\boldsymbol{A}) &= \max_{\dim \mathbb{V}=k} \min_{\boldsymbol{x}\in\mathbb{V}, \|\boldsymbol{x}\|_2=1} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} \\ &= \min_{\dim \mathbb{V}=n-k+1} \max_{\boldsymbol{x}\in\mathbb{V}, \|\boldsymbol{x}\|_2=1} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}\end{aligned}$$

证明.

根据 Poincare 不等式，如果 $\mathbb{V}$ 是 $\mathbb{R}^n$ 的 k 维子空间，那么 $\min_{\boldsymbol{x}\in\mathbb{V}, \|\boldsymbol{x}\|_2=1} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} \le \lambda_k(\boldsymbol{A})$。

如果我们令 $\mathbb{V} = \mathbf{span}\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k\}$，那么我们就得到了第一个等式。

对 $-\boldsymbol{A}$ 用同样的处理方式，我们会得到第二个等式。 $\square$

极小极大准则可以用于比较两个对称矩阵和的特征值和原矩阵特征值的大小关系。

推论 2

令 $\boldsymbol{A}, \boldsymbol{B} \in S^n$，对每个 $k = 1, \cdots, n$，有

$$\lambda_k(\boldsymbol{A}) + \lambda_{min}(\boldsymbol{B}) \leq \lambda_k(\boldsymbol{A} + \boldsymbol{B}) \leq \lambda_k(\boldsymbol{A}) + \lambda_{max}(\boldsymbol{B})$$

证明.

根据推论 1，我们有

$$\begin{aligned}
\lambda_k(\boldsymbol{A} + \boldsymbol{B}) &= \min_{\dim \mathbb{V} = n-k+1} \max_{\boldsymbol{x}: \|\boldsymbol{x}\|_2 = 1} (\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} + \boldsymbol{x}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{x}) \\
&\geq \min_{\dim \mathbb{V} = n-k+1} \max_{\boldsymbol{x}: \|\boldsymbol{x}\|_2 = 1} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} + \lambda_{min}(\boldsymbol{B}) \\
&= \lambda_k(\boldsymbol{A}) + \lambda_k(\boldsymbol{B})
\end{aligned}$$

这就证明了命题中不等式的左半部分，对于右半部分，可以用类似的方法证明。 □

12.2.1 正半定矩阵：定义回顾

对于一个对称矩阵，我们可以对其进行谱分解。但是如果我们对矩阵进一步约束，令其为一个正定矩阵，那么我们则可以对其进行另外一种特殊的三角分解：Cholesky 分解。

我们首先回顾一下关于正半定和正定矩阵的定义和性质。

定义 3

对于对称矩阵 $\boldsymbol{A}$，如果对任意的 $\boldsymbol{x} \in \mathbb{R}^n$ 均有 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \geq 0$ 或 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$，则称 $\boldsymbol{A}$ 是**正半定矩阵**或**正定矩阵**，记做 $\boldsymbol{A} \succeq 0$ 或 $\boldsymbol{A} \succ 0$。

正半定矩阵性质回顾

正半定矩阵的性质

- 正半定矩阵的所有顺序主子式均为非负的。
- 正半定矩阵的行列式是非负的。
- 两个正半定矩阵的和是正半定的。
- 非负实数与正半定矩阵的数乘矩阵是正半定的。

对于分块对角矩阵：

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{O}_{n\times m} \\ \boldsymbol{O}_{m\times n} & \boldsymbol{B} \end{pmatrix}$$

我们很容易验证，如果 $\boldsymbol{A}, \boldsymbol{B}$ 都是正 (半) 定的，那么 $\boldsymbol{M}$ 就是正 (半) 定的。

Schur 补

下面介绍一个定理，可以让我们判断一个非分块对角矩阵是否是正定的。

定理 5

(Schur 补) 令 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$，$\boldsymbol{B} \in \mathbb{R}^{m\times m}$，$\boldsymbol{X} \in \mathbb{R}^{n\times m}$，$\boldsymbol{A}, \boldsymbol{B}$ 都是对称矩阵，并且 $\boldsymbol{B}$ 是正定矩阵。对称矩阵 $\boldsymbol{M}$:

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{T}} & \boldsymbol{B} \end{pmatrix},$$

定义 $\boldsymbol{A}$ 在 $\boldsymbol{M}$ 中的 *Schur* 补为：

$$\boldsymbol{S} := \boldsymbol{A} - \boldsymbol{X}\boldsymbol{B}^{-1}\boldsymbol{X}^{\mathrm{T}}.$$

那么，

$$\boldsymbol{M} \succeq 0 (\boldsymbol{M} \succ 0) \iff \boldsymbol{S} \succeq 0 (\boldsymbol{S} \succ 0)$$

证明.

定义分块矩阵

$$\boldsymbol{C}=\begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{O}_{n\times m} \\ -\boldsymbol{B}^{-1}\boldsymbol{X}^{\mathrm{T}} & \boldsymbol{I}_m \end{bmatrix}$$

这是一个分块下三角矩阵，对角元非零，因此可逆。容易验证：

$$\boldsymbol{C}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{C}=\begin{pmatrix} \boldsymbol{S} & \boldsymbol{O}_{n\times m} \\ \boldsymbol{O}_{m\times n} & \boldsymbol{B} \end{pmatrix}$$

$\boldsymbol{M}$ 正 (半) 定当且仅当 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{C}$ 正 (半) 定。又 $\boldsymbol{B}$ 是正定的，因此 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{C}$ 正 (半) 定当且仅当 $\boldsymbol{S}$ 正 (半) 定。 □

我们判断一个矩阵是正 (半) 定矩阵后，就可以对其进行 Cholesky 分解。

12.2.2 Cholesky 分解

定义 4

设 $\boldsymbol{A}=[a_{ij}]\in\mathbb{R}^{n\times n}$ 是正定矩阵，

$$\boldsymbol{A}=\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$$

称为矩阵 $\boldsymbol{A}$ 的 *Cholesky* 分解，其中，$\boldsymbol{G}\in\mathbb{R}^{n\times n}$ 是一个具有正的对角线元素的下三角矩阵，即

$$\boldsymbol{G}=\begin{pmatrix} g_{11} & & & \\ g_{21} & g_{22} & & \\ \vdots & \vdots & \ddots & \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{pmatrix} \tag{1}$$

Cholesky 分解

比较 $\boldsymbol{A} = \boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$ 两边，易得

$$a_{ij} = \sum_{k=1}^{j} g_{jk}g_{ik} = \sum_{k=1}^{j-1} g_{jk}g_{ik} + g_{jj}g_{ij}$$

从而有

$$g_{jj}g_{ij} = a_{ij} - \sum_{k=1}^{j-1} g_{jk}g_{ik} \tag{2}$$

把 $g_{jj}g_{ij}$ 记为 $v(i)$。如果知道了 $\boldsymbol{G}$ 的前 $j-1$ 列，那么 $v(i)$ 就是可计算的。在式(2)中令 $i=j$，立即有 $g_{jj}^2 = v(j) = a_{jj} - \sum_{k=1}^{j-1} g_{jk}g_{jk}$。然后，由式(2)得

$$g_{ij} = v(i)/g_{jj} = v(i)/\sqrt{v(j)} \tag{3}$$

Cholesky 分解算法

总结以上结论，可得到计算 Cholesky 分解的算法：

Algorithm 2 Cholesky分解

```
for j = 1 : n do
    for i = j : n do
        v(i) = a_ij;
        for k = 1 : j − 1 do
            v(i) = v(i) − g_jk g_ik;
        end for
        g_ij = v(i)/√v(j);
    end for
end for
```

Cholesky 分解定理

以上分析结果可以归纳为下面的定理。

定理 6

[Cholesky 分解] 如果 $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ 是正定矩阵，则 *Cholesky* 分解 $\boldsymbol{A} = \boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$ 存在且是唯一的，其中，下三角矩阵 $\boldsymbol{G} \in \mathbb{R}^{n \times n}$ 的非零元素由式(3)决定。

下三角矩阵 $\boldsymbol{G}$ 称为 Cholesky 三角。另外，Cholesky 分解也叫做平方根方法，因为下三角矩阵 $\boldsymbol{G}$ 可以视为矩阵 $\boldsymbol{A}$ 的“平方根”。

例 2

求矩阵 $\boldsymbol{A}$ 的 *Cholesky* 分解

$$\boldsymbol{A}=\begin{bmatrix} 4 & -1 & -1 \\ -1 & 4.25 & 2.75 \\ 1 & 2.75 & 3.5 \end{bmatrix}$$

解

显然 $\boldsymbol{A}^T=\boldsymbol{A}$，特征值 $\lambda_1=1.15>0, \lambda_2=3.9>0, \lambda_3=6.7>0$，因此，$\boldsymbol{A}$ 为对称正定矩阵。故存在 $\boldsymbol{A}=\boldsymbol{G}\boldsymbol{G}^T$，则有：

解

$$g_{11} = \sqrt{a_{11}} = 2,\ g_{21} = \frac{a_{21}}{g_{11}} = -0.5,\ g_{31} = \frac{a_{31}}{g_{11}} = 0.5,$$

$$g_{22} = \sqrt{a_{22} - g_{22}^2} = 2,$$

$$g_{32} = \frac{a_{32} - g_{31}g_{21}}{g_{22}} = 1.5$$

$$g_{33} = \sqrt{a_{33} - g_{31}^2 - g_{32}^2} = 1$$

可得:

$$\boldsymbol{G} = \begin{bmatrix} 2 & 0 & 0 \\ -0.5 & 2 & 0 \\ 0.5 & 1.5 & 1 \end{bmatrix}$$

由非方阵得到对称正半定矩阵

定理 7

设矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，我们可以由 $\boldsymbol{A}$ 得到一个对称正半定矩阵

$$\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A}$$

理解为什么这个定理成立将对我们如何使用对称矩阵很有启发。

因为对称性要求 $\boldsymbol{S} = \boldsymbol{S}^{\mathrm{T}}$，而我们得到的

$$\boldsymbol{S} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{\mathrm{T}} = \boldsymbol{S}^{\mathrm{T}}$$

而正半定性要求 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x} \geq 0$。因为标量积计算的是平方和 (平方和本身总是正的或零的），我们则有

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}) \geq 0$$

本讲小结

对称矩阵的特征分解

- 存在性和特征分解求解
- 瑞利商
- 极小极大准则

正定矩阵的 Cholosky 分解

- 存在性和 Cholosky 分解求解
- 正（半）定矩阵的特征值
- 舒尔补

机器学习中各种对称、正半定矩阵模型的数值计算的重要工具，利用对角矩阵和三角矩阵的优异性质，可大大降低求逆求行列式等数值计算的复杂性，并将用于奇异值分解和线性方程组等的计算求解等。

缺点：限制在方阵！