

MIT 18.06 线性代数

第二十一讲：特征值和特征向量

整理自 MIT 18.06 课程及相关公开笔记

引言

本讲标志着线性代数课程进入了新的核心模块——**特征值** (Eigenvalues) 与**特征向量** (Eigenvectors)。特征值不仅是描述矩阵固有属性的重要标量，也是理解矩阵对角化、矩阵幂运算以及微分方程组求解的基础。本讲将在方阵的背景下，系统介绍特征值与特征向量的定义、求解算法以及一些特殊情形。

1 特征值与特征向量

1.1 核心方程与几何意义

对于 n 阶方阵 A ，如果存在一个非零向量 $\mathbf{x}$ 和一个标量 λ ，使得如下线性变换成立：

$$A\mathbf{x} = \lambda\mathbf{x}$$

则称 $\mathbf{x}$ 为矩阵 A 的**特征向量**， λ 为 A 的**特征值**。

几何解释：矩阵 A 可以视为对向量空间 $\mathbb{R}^n$ 的一个线性变换。在一般向量经过 A 作用后，方向会发生改变，而特征向量 $\mathbf{x}$ 是那些在变换前后方向保持不变（或者方向完全相反，或者变为零向量）的向量；标量 λ 表示变换在该方向上的拉伸或压缩倍数。

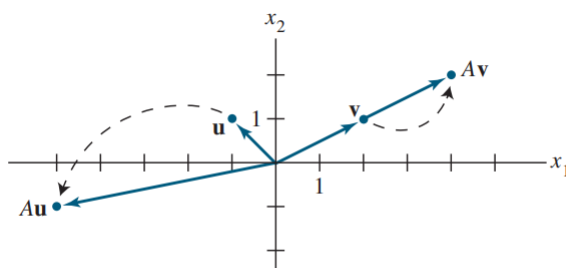


FIGURE 1 Effects of multiplication by A .

1.2 特征值为零的情形

如果矩阵 A 存在特征值 $\lambda = 0$ ，根据定义有 $A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$ ，这意味着存在非零向量 $\mathbf{x}$ 使得齐次方程 $A\mathbf{x} = \mathbf{0}$ 有解。因此，矩阵 A 奇异（不可逆）的充要条件是它至少有一个特征值为 0。

1.3 投影矩阵的特征值

考虑一个投影矩阵 P ，其作用是将向量投影到某个平面（子空间）上。

- 对于位于该投影平面内的任意向量 $\mathbf{x}$ ，投影后保持不变，因此 $P\mathbf{x} = \mathbf{x}$ ，即特征值为 1。
- 对于垂直于该平面的向量 $\mathbf{x}$ ，投影后为零向量，因此 $P\mathbf{x} = \mathbf{0}$ ，即特征值为 0。

投影矩阵的特征向量张成了整个空间。

2 特征值的求解

2.1 特征方程

要直接求解方程 $A\mathbf{x} = \lambda\mathbf{x}$ 中的 λ 和 $\mathbf{x}$ 并不容易。将方程移项：

$$A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \implies (A - \lambda I)\mathbf{x} = \mathbf{0}$$

要使得非零向量 $\mathbf{x}$ 是此齐次方程的解，系数矩阵 $(A - \lambda I)$ 必须是奇异矩阵，因此，其行列式必须等于 0。这给出了**特征方程** (Characteristic Equation)：

$$|A - \lambda I| = 0$$

计算步骤：

1. 求解特征方程 $|A - \lambda I| = 0$ ，解出的 n 个根即为矩阵的 n 个特征值（在复数域内有 n 个解，可能有重根）。
2. 对于每一个特征值 λ_i ，将其代入 $(A - \lambda_i I)\mathbf{x} = \mathbf{0}$ ，利用高斯消元法求此方程组的零空间，零空间中的非零向量即为对应于 λ_i 的特征向量。

2.2 迹与行列式

根据特征方程的根与系数的关系，可以得到两个的重要性质：

- **特征值之和等于矩阵的迹** (trace)：

$$\sum_{i=1}^n \lambda_i = \text{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

- **特征值之积等于矩阵的行列式**：

$$\prod_{i=1}^n \lambda_i = |A|$$

对于二阶矩阵，若已知其中一个特征值，可以直接通过这两个性质快速求出另一个特征值。

3 示例

3.1 二阶对称矩阵

求解矩阵 $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ 的特征值与特征向量。

解：构造特征方程：

$$|A - \lambda I| = \begin{vmatrix} 3 - \lambda & 1 \\ 1 & 3 - \lambda \end{vmatrix} = (3 - \lambda)^2 - 1 = 0$$

展开后得到 $\lambda^2 - 6\lambda + 8 = 0$ （其中 6 是迹，8 是行列式的值），解得特征值 $\lambda_1 = 4, \lambda_2 = 2$ 。

分别回代求零空间：

- 当 $\lambda_1 = 4$ 时，代入 $(A - 4I)\mathbf{x} = \mathbf{0}$ ，得 $\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{x} = \mathbf{0}$ ，特征向量可取 $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 。
- 当 $\lambda_2 = 2$ 时，代入 $(A - 2I)\mathbf{x} = \mathbf{0}$ ，得 $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \mathbf{x} = \mathbf{0}$ ，特征向量可取 $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ 。

3.2 矩阵加法对特征值与特征向量的影响

在 A 的基础上，考虑矩阵 $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ，其特征值为 1 和 -1，特征向量分别为 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 。观察 $A = B + 3I$ ，由 $B\mathbf{x} = \lambda'\mathbf{x}$ 可推出：

$$A\mathbf{x} = (B + 3I)\mathbf{x} = B\mathbf{x} + 3\mathbf{x} = \lambda'\mathbf{x} + 3\mathbf{x} = (\lambda' + 3)\mathbf{x}$$

注意事项：一般矩阵的求和的特征值不等于特征值的和（即 $|\lambda(A) + \lambda(B)| \neq |\lambda(A + B)|$ ），除非这两个矩阵具有相同的特征向量。

4 特殊情况

4.1 虚数特征值

考虑旋转矩阵 $Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ ，该矩阵将平面内所有向量逆时针旋转 90° 。由于任意向量在经过 90° 旋转后方向都会发生改变，因此该矩阵在实数域内**不存在实数特征向量**。

其特征方程为：

$$|Q - \lambda I| = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0$$

解得复数特征值 $\lambda_1 = i$ 和 $\lambda_2 = -i$ 。由此可得对称矩阵与反对称矩阵的性质：

- 对于对称矩阵 $A^T = A$ ，特征值始终为**实数**。
- 对于反对称矩阵 $A^T = -A$ ，特征值为**纯虚数**。

4.2 重根与特征向量的短缺

对于三角矩阵，其特征值恰好是主对角线上的元素。例如，对于上三角矩阵 $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ ：

$$|A - \lambda I| = \begin{vmatrix} 3 - \lambda & 1 \\ 0 & 3 - \lambda \end{vmatrix} = (3 - \lambda)^2 = 0$$

解得 **二重根** $\lambda_1 = \lambda_2 = 3$ ，将 $\lambda = 3$ 代入求特征向量：

$$(A - 3I)\mathbf{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0}$$

该方程只能解出一个线性无关的特征向量 $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ，这意味着尽管是二阶矩阵，但缺少了一个独立的特征向量。这种特征向量不足的现象，通常意味着矩阵**无法相似对角化**（即退化）。

5 对称矩阵特征向量的正交性

对于对称矩阵，其不同特征值所对应的特征向量必然是互相正交的。证明如下：假设对称矩阵 $A = A^T$ 有两个不同的特征值 $\lambda_1 \neq \lambda_2$ ，对应的特征向量分别为 $\mathbf{x}_1, \mathbf{x}_2$ ，由定义可知：

$$A\mathbf{x}_1 = \lambda_1\mathbf{x}_1, \quad A\mathbf{x}_2 = \lambda_2\mathbf{x}_2$$

用 $\mathbf{x}_2^T$ 左乘第一个方程：

$$\mathbf{x}_2^T A\mathbf{x}_1 = \lambda_1 \mathbf{x}_2^T \mathbf{x}_1$$

利用对称性 $A = A^T$ ，等式左边可改写为：

$$\mathbf{x}_2^T A\mathbf{x}_1 = (\mathbf{x}_2^T A^T)\mathbf{x}_1 = (A\mathbf{x}_2)^T \mathbf{x}_1 = \lambda_2 \mathbf{x}_2^T \mathbf{x}_1$$

因此， $\lambda_1 \mathbf{x}_2^T \mathbf{x}_1 = \lambda_2 \mathbf{x}_2^T \mathbf{x}_1 \implies (\lambda_1 - \lambda_2) \mathbf{x}_2^T \mathbf{x}_1 = 0$ ，由于 $\lambda_1 \neq \lambda_2$ ，故必然有 $\mathbf{x}_2^T \mathbf{x}_1 = 0$ ，即 $\mathbf{x}_1 \perp \mathbf{x}_2$ 。

总结

本讲介绍了线性代数中的特征值与特征向量理论：

- **定义**：满足方程 $A\mathbf{x} = \lambda\mathbf{x}$ 的标量 λ 与向量 $\mathbf{x}$ 。
- **求解核心**：特征方程 $|A - \lambda I| = 0$ 用于求出特征值， $(A - \lambda I)\mathbf{x} = \mathbf{0}$ 用于求出特征向量。
- **重要性质**： $\sum \lambda_i = \text{tr}(A)$ 、 $\prod \lambda_i = |A|$ 、对称矩阵不同特征值对应的特征向量正交。
- **特殊情况**：实数矩阵可能有复数特征值（如旋转矩阵）；三角矩阵若有重根，可能导致特征向量短缺，矩阵不可相似对角化。

补充：特征值和特征向量的应用

特征值与特征向量不仅是抽象的线性代数概念，更是现代数据科学、机器学习和复杂网络分析中不可或缺的核心工具。

- **特征向量**：是在线性变换（矩阵乘法）作用下，仅发生长度伸缩而保持方向不变的向量，它代表了系统或数据集内在的“稳定结构”。
- **特征值**：是特征向量在变换中被拉伸或压缩的倍数，特征值的大小直接反映了该特征向量在系统中的“重要性”或“影响力度”。

5.1 应用一：人脸识别

问题：如何高效且精确地表示和识别一张人脸图像？

1. **降维与数据转化**：将高维的像素矩阵（人脸图像）展开为一维向量，构建全体样本的数据集。
2. **寻找核心特征**：计算人脸图像数据集的**协方差矩阵**，该矩阵包含了像素之间变化的统计信息。
3. **求解特征脸**：协方差矩阵的特征向量被提取出来，在计算机视觉中被称为“**特征脸**”(Eigenfaces)，这些特征向量构成了一组最佳的基底，能够最高效地描述人脸的变化。
4. **特征值衡量重要性**：特征值的大小表示了该特征脸在整体人脸变化中的贡献度或重要性。

重构与应用：任何一张新的人脸图像，都可以表示为“平均人脸”与一组特征脸的线性组合（加权求和）。通过比较这些权重系数，即可实现快速的人脸识别与分类。



(a)



(b)

Figure 1: (a) Face images used as the training set. (b) The average face Ψ .



Figure 2: Seven of the eigenfaces calculated from the images of Figure 1, without the background removed.

图 1: Face Recognition Using Eigenfaces (Turk & Pentland, 1991)

5.2 应用二：数据降维

问题：在处理高维数据（例如具有数千个特征的基因表达数据或图像数据）时，如何剔除冗余噪声，保留最重要的信息？

1. **寻找主成分：**计算数据的**协方差矩阵**，并求解其特征值与特征向量。
2. **定义方向的重要性：**特征值的大小直接代表了数据在对应特征向量方向上的**方差大小**，特征值越大，说明数据在这个方向上的波动越大（即信息量越丰富）。
3. **降维操作：**将特征值从大到小排列，选取前 k 个最大的特征值，将数据投射到对应的 k 个特征向量所张成的子空间上，即可实现数据降维。

结论：通过丢弃特征值很小（即方差很小、可视为噪声）的特征向量方向，PCA 可以在保留数据集主要结构的基础上，极大地降低计算复杂度。

5.3 应用三：网页排名

问题：在拥有数十亿网页的万维网中，如何客观评估每一个网页的重要性或权威性？

1. **构建转移矩阵：**将整个互联网的网页视为**节点**，超链接视为**有向边**，构造一个巨大的**转移矩阵**。矩阵的每个元素代表用户从网页 i 跳转到网页 j 的概率。
2. **求解主特征向量：**PageRank 算法的核心在于寻找该转移矩阵的**主特征向量**（即对应最大特征值 $\lambda = 1$ 的特征向量）。
3. **特征值的理论保证：**根据 Perron-Frobenius 定理，该转移矩阵的最大特征值必为 1，保证了稳态分布的数学确定性。

PageRank 值：特征向量中的每一个元素值，即为对应网页的 PageRank 得分，得分为 1 的特征向量代表了互联网经过无数次随机漫游后的**稳态概率分布**，这正是网页相对重要性的量化度量。

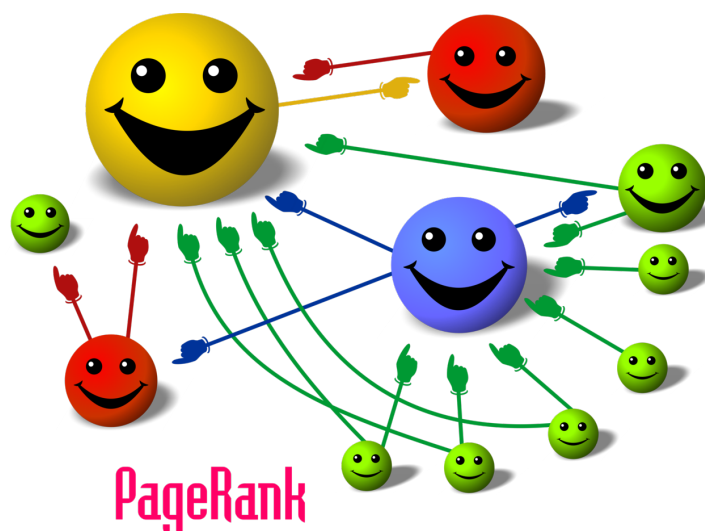


图 2: PageRank 的卡通概念图