

MIT 18.06 线性代数

第二十二讲：相似对角化

整理自 MIT 18.06 课程及相关公开笔记

引言

上一讲引入了特征值与特征向量的概念。本讲将在此基础上，介绍一种全新的矩阵分解形式——**对角化** (Diagonalization)，并展示如何利用对角化极大地简化矩阵的幂运算。同时，本讲将通过差分方程 (Difference Equations) 和斐波那契数列的具体实例，展示特征值与特征向量在分析动态系统中的核心作用。

1 矩阵的对角化

1.1 对角化分解的推导

设 n 阶方阵 A 拥有 n 个线性无关的特征向量 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ ，将它们作为列向量拼成一个 $n \times n$ 的可逆矩阵 S ：

$$S = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$$

根据特征值和特征向量的定义，有 $A\mathbf{x}_i = \lambda_i\mathbf{x}_i$ 。计算矩阵 A 与 S 的乘积：

$$AS = A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{x}_1 & \lambda_2\mathbf{x}_2 & \cdots & \lambda_n\mathbf{x}_n \end{bmatrix}$$

将右边的向量组合提取为矩阵形式：

$$AS = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = SA$$

由于 S 的列向量线性无关， S 是可逆矩阵。在等式两边同时左乘 S^{-1} ，即可得到矩阵的对角化公式：

$$S^{-1}AS = \Lambda \quad \text{或写成} \quad A = SAS^{-1}$$

这里 Λ 是对角矩阵，其对角线上的元素正是矩阵 A 的 n 个特征值 λ_i 。这一分解被称为**相似对角化**，矩阵 A 与对角阵 Λ 称为**相似矩阵**，因为它们具有相同的特征值。

1.2 对角化的前提条件

矩阵 A 可相似对角化，当且仅当它拥有 n 个线性无关的特征向量。如果矩阵拥有 n 个**互不相同**的特征值，那么它一定拥有 n 个线性无关的特征向量，从而一定能被对角化。大部分矩阵都能满足这一条件。

2 利用对角化计算矩阵的幂

2.1 特征值与特征向量的幂性质

根据特征值的定义 $A\mathbf{x} = \lambda\mathbf{x}$ ，考察 A^2 对同一个特征向量 $\mathbf{x}$ 的作用：

$$A^2\mathbf{x} = A(A\mathbf{x}) = A(\lambda\mathbf{x}) = \lambda A\mathbf{x} = \lambda^2\mathbf{x}$$

由此可见，矩阵 A^2 的特征向量与 A 相同，其特征值为 λ^2 。同理可推广至任意 k 次幂：

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

2.2 幂运算的矩阵分解

利用对角化公式 $A = S\Lambda S^{-1}$ ，可以直接导出矩阵幂的一般公式：

$$A^2 = (S\Lambda S^{-1})(S\Lambda S^{-1}) = S\Lambda(S^{-1}S)\Lambda S^{-1} = S\Lambda^2 S^{-1}$$

$$A^k = S\Lambda^k S^{-1}$$

这说明计算矩阵的 k 次幂只需计算对角矩阵的 k 次幂（即对角线上各特征值的 k 次幂），然后夹乘前后矩阵即可，极大简化了高次幂的计算。

2.3 矩阵幂收敛于零的条件

基于 $A^k = S\Lambda^k S^{-1}$ ，可以得出一个判断动态系统极限行为的重要定理：当 $k \rightarrow \infty$ 时，矩阵 A^k 趋于零矩阵，当且仅当矩阵 A 的所有特征值的绝对值均严格小于 1，即：

$$|\lambda_i| < 1 \quad (\text{对所有 } i = 1, 2, \dots, n)$$

这个条件在差分方程的稳定性分析中起着决定性作用。

3 重特征值与可对角化

矩阵是否可对角化，与是否拥有重特征值紧密相关：

- **互异特征值**：若矩阵 A 有 n 个互不相同的特征值，则这些特征值对应的特征向量必然线性无关，因此矩阵必然可对角化。
- **重特征值**：若矩阵 A 有重特征值，它仍然可能拥有 n 个线性无关的特征向量。例如， n 阶单位矩阵 I 的所有特征值均为 1，但它拥有 n 个标准基向量作为线性无关的特征向量，因此可以对角化。
- **特征向量短缺**：对于某些带有重特征值的矩阵，可能会出现特征向量短缺的情况。例如，对于三角矩阵 $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ ，其特征方程为 $(\lambda - 2)^2 = 0$ ，仅有重根 2，求解零空间 $(A - 2I)\mathbf{x} = \mathbf{0}$ ：

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0} \implies \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

该矩阵仅有 1 个线性无关的特征向量，低于 $n = 2$ ，因此无法对角化。

4 差分方程与动态系统

4.1 一阶差分方程的通解

在宏观经济学的离散动态建模中，经常会遇到一阶线性差分方程：

$$\mathbf{u}_{k+1} = A\mathbf{u}_k$$

给定初始向量 $\mathbf{u}_0$ ，其解一般形式为 $\mathbf{u}_k = A^k \mathbf{u}_0$ 。结合对角化的幂运算公式，可以将初始向量 $\mathbf{u}_0$ 分解为特征向量的线性组合：

$$\mathbf{u}_0 = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \cdots + c_n \mathbf{x}_n = S\mathbf{c}$$

将该形式代入差分方程中，可以得到系统在第 k 步的显式解：

$$\begin{aligned}\mathbf{u}_1 &= A\mathbf{u}_0 = c_1 \lambda_1 \mathbf{x}_1 + c_2 \lambda_2 \mathbf{x}_2 + \cdots + c_n \lambda_n \mathbf{x}_n \\ \mathbf{u}_k &= A^k \mathbf{u}_0 = c_1 \lambda_1^k \mathbf{x}_1 + c_2 \lambda_2^k \mathbf{x}_2 + \cdots + c_n \lambda_n^k \mathbf{x}_n = S\Lambda^k \mathbf{c}\end{aligned}$$

这个公式清晰地揭示了系统的长期行为：每一个特征值 λ_i 决定了对应特征向量方向的增长或衰减速率。

4.2 斐波那契数列的应用

斐波那契数列定义为： $F_0 = 0, F_1 = 1$ ，并且满足递推公式 $F_{k+2} = F_{k+1} + F_k$ 。这是一个二阶差分方程，为了将其转化为一阶矩阵形式，令：

$$\mathbf{u}_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$$

根据递推公式，可以构造出如下等价的一阶系统：

$$\mathbf{u}_{k+1} = \begin{bmatrix} F_{k+2} \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = A\mathbf{u}_k$$

求该矩阵 A 的特征方程：

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

解得两个特征值：

$$\lambda_1 = \frac{1 + \sqrt{5}}{2} \approx 1.618, \quad \lambda_2 = \frac{1 - \sqrt{5}}{2} \approx -0.618$$

其中 λ_1 为黄金分割比，由于其绝对值大于 1，它主导了序列的增长速度。当 $k \rightarrow \infty$ 时， $|\lambda_2| < 1$ 使得 λ_2^k 迅速衰减，因此：

$$F_{100} \approx c_1 \lambda_1^{100} = c_1 \left(\frac{1 + \sqrt{5}}{2} \right)^{100}$$

对应的特征向量为 $\mathbf{x}_i = \begin{bmatrix} \lambda_i \\ 1 \end{bmatrix}$ 。由初始向量 $\mathbf{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 可求出组合系数 c_1 和 c_2 ，最终得到斐波那契数列的精确通项公式：

$$F_k = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^k - \frac{1}{\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2} \right)^k$$

总结

- **对角化**将矩阵乘法分解为 $A = S\Lambda S^{-1}$ ，该分解大大简化了矩阵幂运算，并且特征值的收敛条件决定了差分方程解的长远趋势。
- 矩阵拥有 n 个线性无关的特征向量是对角化的必要条件，重特征值可能导致特征向量短缺（即矩阵退化为不可对角化矩阵）。
- **差分方程**的解可以通过特征向量展开求得。斐波那契数列的经典例子诠释了如何通过求解特征值与特征向量，直接从解析层面预测离散数列的长期极限行为。