

MIT 18.06 线性代数

第二十五讲：对称矩阵与正定矩阵

整理自 MIT 18.06 课程及相关公开笔记

引言

本讲将围绕一类性质优良的方阵——**对称矩阵** (symmetric matrix) 展开，并引出与之紧密相关的**正定矩阵** (positive definite matrix)。对称矩阵的特征值和特征向量具有明确的实数性质和正交性，这使得它们在数学理论与工程实践中都被频繁使用；而正定矩阵则通过将矩阵的**主元**、**特征值**和**行列式**这三个核心概念整合在一起，从而为许多优化问题与稳定性分析提供了统一的分析框架。

1 对称矩阵

1.1 谱定理与对称矩阵的对角化

由于对称矩阵存在 n 个线性无关且相互正交的特征向量，因此它必然可以被对角化；对于通常的可对角化矩阵，我们有 $A = S\Lambda S^{-1}$ 。而将特征向量矩阵 S 取为标准正交矩阵 Q ，并利用 $Q^{-1} = Q^T$ ，则对称矩阵可以表示为：

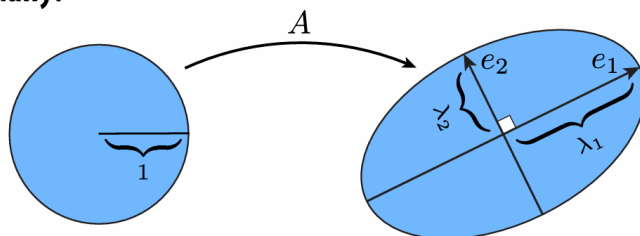
$$A = Q\Lambda Q^{-1} = Q\Lambda Q^T$$

其中 Q 的列向量是标准正交的特征向量， Λ 为对角元为特征值的对角矩阵。

这种分解形式在数学上被称为**谱定理** (spectral theorem)，它将特征值视为矩阵的“谱”；在力学中则被称为**主轴定理** (principal axis theorem)，其含义是：在合适的坐标轴（即特征向量方向）下，矩阵所对应的二次型将不再含有交叉项，可化为对角形式。此外，任何形如 $Q\Lambda Q^T$ 的矩阵都是对称矩阵，因为 $(Q\Lambda Q^T)^T = Q\Lambda Q^T$ 。

■ Algebraically, matrix A is *symmetric* if $A^T = A$

■ Visually?



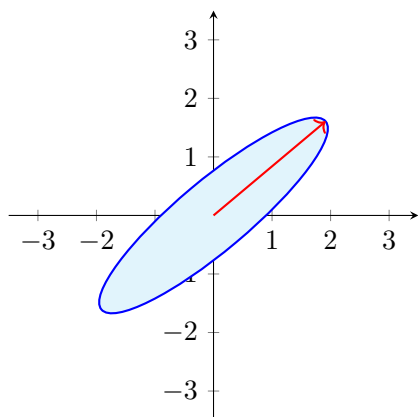
Key idea: unequal stretch along orthogonal axes

Spectral theorem: symmetric matrix has real eigenvalues (stretch amounts), orthogonal eigenvectors (stretch directions)

1.2 几何直观与二次型的矩阵表示

二次型 (Quadratic Form) 是含有多个变量的二次齐次多项式，在几何上通常对应着椭圆、椭圆等二次曲面。通过旋转坐标轴消除交叉项，可以将复杂的曲线方程简化为只含平方项的**标准形**。

任意线性变换



正交旋转变换

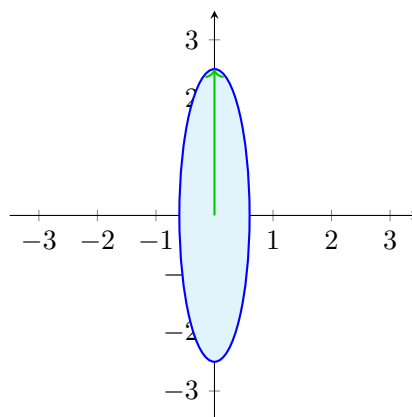


图 1: 正交旋转变换不改变形状

1.2.1 二次型的矩阵形式

对于含交叉项的二次型方程，例如： $5x^2 + 8xy + 5y^2 = 1$ ，可以将该多项式写成向量与矩阵乘积的**二次型矩阵形式**：

$$f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$$

其中 $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ ，而 A 是对应于该二次型的**对称矩阵**。矩阵构造法则：

- x^2 和 y^2 的系数直接放在矩阵的**主对角线**上；
- 交叉项 ($8xy$) 的系数需要**对半拆开**，分别放在矩阵对称位置的非对角线上（即 4 和 4）。

于是，上面的方程可以转化为：

$$\mathbf{x}^T A \mathbf{x} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1$$

1.2.2 正交变换

为了消除交叉项，必须进行**正交变换** (Orthogonal Transformation)。正交变换对应一个**正交矩阵** P (满足 $P^T P = I$)，它代表坐标系的一种**旋转**。这种变换的特点是：只会改变图形的坐标朝向，而不会改变图形自身的形状、长宽比例和体积。

设新坐标为 $\mathbf{y}$ ，则变量代换关系为：

$$\mathbf{x} = P \mathbf{y}$$

代入二次型得：

$$\mathbf{x}^T A \mathbf{x} = (P \mathbf{y})^T A (P \mathbf{y}) = \mathbf{y}^T (P^T A P) \mathbf{y}$$

1.2.3 二次型的标准型

我们寻找正交矩阵 P 的目标是让变换后的矩阵 $P^T A P$ 成为一个**对角矩阵** Λ ，即让非对角线上的交叉项系数全部变为 0。根据谱定理，对于任意实对称矩阵 A ，总存在一个正交矩阵 Q ，使得：

$$Q^T A Q = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

此时，原二次型方程被化为了**标准形** (Standard Form)：

$$\mathbf{y}^T \Lambda \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 = 1$$

在这组新坐标下，椭圆、椭球的对称轴就与新坐标轴完全对齐了。

1.3 寻找方向就是寻找特征向量

那么，如何找到那个让图形“变正”的特殊旋转角度呢？这个特殊方向是图形本身就具备的，即椭圆的**长轴和短轴**（在三维空间中对应椭球的三根**主轴**）。

在主轴的方向上取一个向量 $\mathbf{x}$ ，施加矩阵变换 A 后，该向量的方向不发生变化，仅长度发生伸缩： $A\mathbf{x} = \lambda\mathbf{x}$ ，这就是矩阵 A 的特征向量，对应的伸缩倍数 λ 为特征值。显然，这些特征向量（两两垂直）对应的方向，就是我们需要的“旋转对齐方向”。在计算中，将二次型化为标准形的步骤：

1. **求特征值**：求解矩阵 A 的特征方程 $|A - \lambda I| = 0$ ，解出的 λ_i 就是标准形中平方项的系数。对于本例中的矩阵 A ，解特征方程可得 $\lambda_1 = 8, \lambda_2 = 2$ 。
2. **求特征向量**：将每一个特征值 λ_i 分别代入特征方程 $(A - \lambda_i I)\mathbf{x}_i = \mathbf{0}$ ，求得对应的线性无关的特征向量 $\mathbf{x}_i$ （即椭圆长轴和短轴的方向）。
3. **单位化并构造正交矩阵**：将求得的所有特征向量进行**单位化**处理（使其模长为 1），并将单位化后的向量按列排成一个矩阵 P ，此时矩阵 P 即为所需的**正交变换矩阵**。
4. **写出标准形**： $\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$ ；对于二维情形，本例的标准形即为： $8y_1^2 + 2y_2^2 = 1$ 。

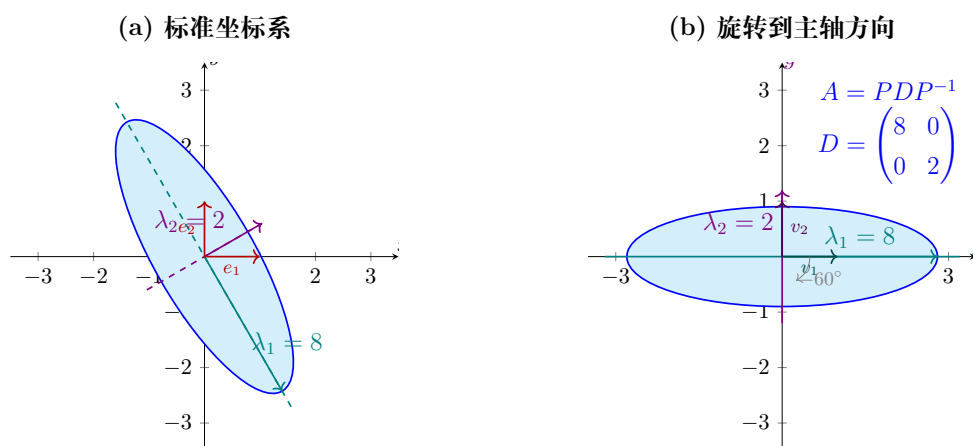


图 2: 坐标轴旋转到主轴方向后，椭圆的主轴与坐标轴对齐，矩阵变为对角形

1.4 对称矩阵特征值为实数的证明

对称矩阵的特征值均为实数，这一结论的证明过程如下：

设 A 为实对称矩阵，且有特征值 λ 和对应的特征向量 x ，满足：

$$Ax = \lambda x$$

对该式取共轭，由于 A 为实矩阵，得到：

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

对上式两边取转置，可得：

$$\bar{x}^T A^T = \bar{x}^T \bar{\lambda}$$

由于 A 对称，即 $A^T = A$ ，因此：

$$\bar{x}^T A = \bar{x}^T \bar{\lambda}$$

在等式两边右乘 x ，得到：

$$\bar{x}^T Ax = \bar{x}^T \bar{\lambda}x$$

另一方面，由原特征方程 $Ax = \lambda x$ ，在等式左边左乘 $\bar{x}^T$ ，得到：

$$\bar{x}^T Ax = \bar{x}^T \lambda x$$

比较以上两个结果，有：

$$\bar{x}^T \lambda x = \bar{x}^T \bar{\lambda}x$$

即：

$$(\lambda - \bar{\lambda}) \bar{x}^T x = 0$$

而 $\bar{x}^T x = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 > 0$ （因为特征向量 x 非零），因此必须有 $\lambda = \bar{\lambda}$ ，即 λ 为实数。

对于复矩阵而言，若满足 $A = \bar{A}^T$ （即共轭转置等于自身，称为 Hermite 矩阵），则上述推理同样成立，其特征值也为实数，特征向量同样具有正交性。

1.5 对称矩阵是投影矩阵的组合

根据谱分解 $A = Q\Lambda Q^T$ ，将矩阵乘法按列乘以行的方式展开：

$$A = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots + \lambda_n q_n q_n^T$$

其中 $q_k q_k^T$ 正是沿方向 q_k 进行投影的投影矩阵。因此，任意实对称矩阵都可以表示为若干个相互正交的投影矩阵的线性组合，其组合系数即为对应的特征值，这是谱定理的另一种理解方式。

1.6 主元的符号与特征值的符号

1.6.1 惯性定理

在实际计算中，对于大规模矩阵，直接通过求解 $|A - \lambda I| = 0$ 来获得特征值往往不够方便。而对于对称矩阵，一个重要的结论是：

正主元的个数 = 正特征值的个数

负主元的个数 = 负特征值的个数

换言之，对称矩阵主元的正负号与特征值的正负号完全一致。该结论可通过合同变换的**惯性定理** (Sylvester's Law of Inertia) 加以证明：消元过程给出 $A = LDL^T$ ，而对角化给出 $A = Q\Lambda Q^T$ ，两者比较可知 D 与 Λ 合同，合同变换不改变正负**惯性指数**，因此主元和特征值的符号相匹配。这提供了一种比直接求特征值更方便的判断方式，通过消元获得主元即可了解特征值的符号分布。

此外，利用矩阵 $A + bI$ 的特征值比 A 的特征值大 b 这一性质，还可以通过检查 $A + bI$ 的主元符号来判断 A 的特征值与 b 的大小关系，从而对特征值的具体位置进行估计。

1.6.2 合同变换

设 A 和 B 是两个 n 阶实对称矩阵。如果存在一个可逆矩阵 P ，使得：

$$B = P^T A P$$

1. 消元过程 (LDL^T 分解)，由于 L 是单位下三角矩阵，必然可逆。

将 A 两边同时左乘 L^{-1} 、右乘 $(L^{-1})^T$ ，可得：

$$D = L^{-1} A (L^{-1})^T$$

令 $P = (L^{-1})^T$ ，则：

$$D = P^T A P$$

所以 D 与 A 合同。

2. 正交对角化 ($Q\Lambda Q^T$ 分解)，由于 Q 是正交矩阵，满足 $Q^T = Q^{-1}$ ，必然可逆。同理可得：

$$\Lambda = Q^T A Q$$

所以 Λ 也与 A 合同。

合同关系具有传递性：如果 D 与 A 合同， A 与 Λ 合同，则 D 与 Λ 合同。

相似 (Similarity) 刻画的是“同一个线性变换在不同基下的表现”，关注的是特征值。

合同 (Congruence) 刻画的是“同一个二次型在不同坐标系下的表达式”，关注的是惯性指数。

2 正定矩阵

2.1 定义与基本判定

正定矩阵是指所有特征值均为正数的实对称矩阵。由于对称矩阵的特征值与主元符号一致，因此等价于：当对称矩阵的所有主元均为正数时，该矩阵即为正定矩阵。

例如，对于矩阵

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$$

消元后得到主元为 5 和 2.2，均大于零，因此 A 是正定矩阵。计算特征值：

$$|A - \lambda I| = \lambda^2 - 8\lambda + 11 = 0$$

解得 $\lambda = 4 \pm \sqrt{5}$ ，二者均为正数，这与主元符号一致的结论相符。

2.2 正定矩阵的充要条件

对于实对称矩阵 A ，以下条件彼此等价，均可作为正定矩阵的判定依据：

- (1) 对任意非零向量 x ，均有 $x^T A x > 0$ ；
- (2) 所有特征值 $\lambda_i > 0$ ；
- (3) 所有主元均大于 0；
- (4) 全部顺序主子式均大于 0；
- (5) 存在可逆矩阵 D ，使得 $A = D^T D$ ；
- (6) A 合同于单位矩阵 I 。

其中，顺序主子式是指取矩阵前 k 行、前 k 列所构成的子矩阵的行列式， $1 \leq k \leq n$ 。这一条件给出了一个便于手算检验的途径。

2.3 正定矩阵的必要条件

若 A 是正定矩阵，则可推出以下必要条件：

- (1) 主对角元素 $a_{ii} > 0$ ；
- (2) 行列式 $|A| > 0$ 。

需要注意的是，行列式大于零只是正定的必要条件，而非充分条件。例如矩阵

$$\begin{bmatrix} -1 & 0 \\ 0 & -3 \end{bmatrix}$$

的行列式为 $3 > 0$ ，但其特征值均为负数，因此并不是正定矩阵。这也说明了“全部顺序主子式均大于 0”这一条件的必要性——仅靠整体行列式无法排除对角线元素为负的情况。

总结

- **对称矩阵的基本性质**：实对称矩阵的特征值均为实数，并且可以选出一组相互正交的特征向量。利用标准正交特征向量矩阵 Q ，对称矩阵可分解为 $A = Q\Lambda Q^T$ ，这一分解形式称为谱定理。
- **对称矩阵的展开与投影解释**：将 $A = Q\Lambda Q^T$ 按列乘以行展开，可得 $A = \sum_{k=1}^n \lambda_k q_k q_k^T$ ，表明每个对称矩阵都是一组相互正交的投影矩阵的线性组合。
- **主元与特征值的符号对应**：对于对称矩阵，正主元的个数等于正特征值的个数，负主元的个数等于负特征值的个数。这一性质为判断特征值符号提供了便捷途径。
- **正定矩阵的定义与判定**：正定矩阵是特征值全为正的对称矩阵，等价地，也可通过主元全为正、顺序主子式全为正、存在可逆矩阵 D 使 $A = D^T D$ 、或 A 合同于单位矩阵等方式加以判定。
- **正定矩阵的必要条件**：正定矩阵的主对角元素和行列式均大于零，但这些条件单独使用时并不充分，需要结合顺序主子式等其它判据。