14 Dive into Eigen Decomposition 深入特征值分解

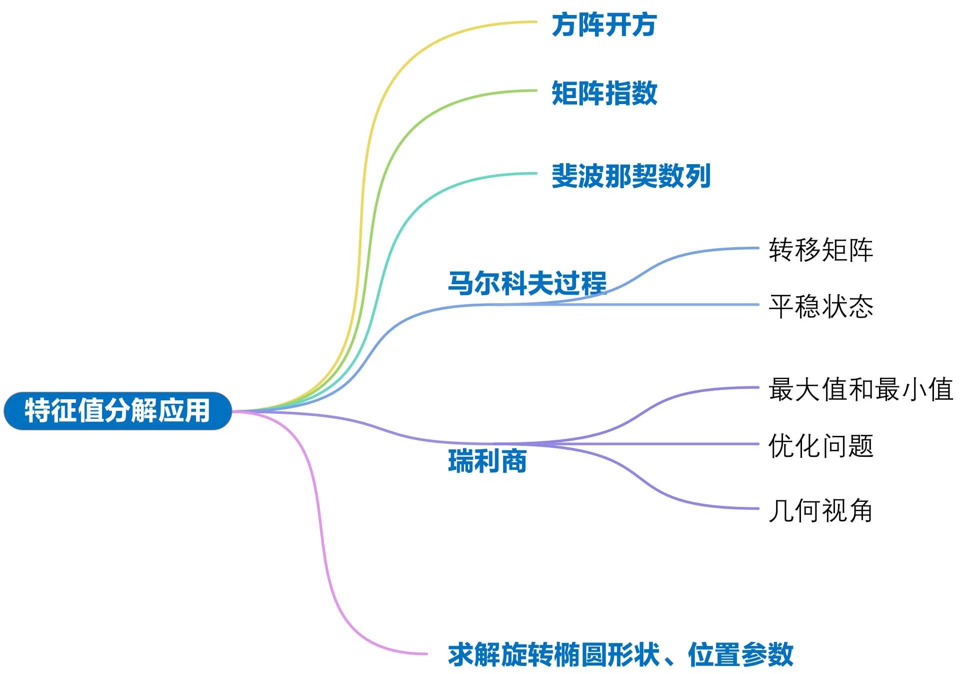
无处不在的特征值分解

生命之殇，并非求其上，却得其中；而是求其下，必得其下。

The greater danger for most of us lies not in setting our aim too high and falling short; but in setting our aim too low, and achieving our mark.

—— 米开朗琪罗 (Michelangelo) | 文艺复兴三杰之一 | 1475 ~ 1564

- ◀ `numpy.meshgrid()` 产生网格化数据
- ◀ `numpy.prod()` 指定轴的元素乘积
- ◀ `numpy.linalg.inv()` 矩阵求逆
- ◀ `numpy.linalg.eig()` 特征值分解
- ◀ `numpy.diag()` 以一维数组的形式返回方阵的对角线元素，或将一维数组转换成对角阵
- ◀ `seaborn.heatmap()` 绘制热图


特征值分解应用
方阵开方
矩阵指数
斐波那契数列
马尔科夫过程
转移矩阵
平稳状态
瑞利商
最大值和最小值
优化问题
几何视角
求解旋转椭圆形状、位置参数

14.1 方阵开方

本章是上一章的延续，本章继续探讨特征值分解及其应用。这一节介绍利用特征值分解完成方阵开方。

如果方阵 $\boldsymbol{A}$ 可以写作：

$$\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B} \tag{1}$$

$\boldsymbol{B}$ 是 $\boldsymbol{A}$ 的平方根。利用特征值分解，可以求得 $\boldsymbol{A}$ 的平方根。

首先对矩阵 $\boldsymbol{A}$ 特征值分解：

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} \tag{2}$$

令：

$$\boldsymbol{B} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{V}^{-1} \tag{3}$$

$\boldsymbol{B}^2$ 可以写成：

$$\boldsymbol{B}^2 = \left(\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{V}^{-1}\right)^2 = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\underbrace{\boldsymbol{V}^{-1}\boldsymbol{V}}_{\boldsymbol{I}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{V}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} = \boldsymbol{A} \tag{4}$$

即：

$$\boldsymbol{A}^{\frac{1}{2}} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{V}^{-1} \tag{5}$$

类似地，方阵 $\boldsymbol{A}$ 的立方根可以写成：

$$\boldsymbol{A}^{\frac{1}{3}} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{3}}\boldsymbol{V}^{-1} \tag{6}$$

继续推广，可以得到：

$$\boldsymbol{A}^{p} = \boldsymbol{V}\boldsymbol{\Lambda}^{p}\boldsymbol{V}^{-1} \tag{7}$$

其中，p 为任意实数。

举个例子

给定如下方阵 $\boldsymbol{A}$，求解如下矩阵的平方根：

$$\boldsymbol{A} = \begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix} \tag{8}$$

对 $\boldsymbol{A}$ 进行特征值分解得到：

$$A=\begin{bmatrix}1.25 & -0.75\\ -0.75 & 1.25\end{bmatrix}=V\Lambda V^{-1}=\begin{bmatrix}\sqrt{3}/2 & \sqrt{3}/2\\ -\sqrt{3}/2 & \sqrt{3}/2\end{bmatrix}\begin{bmatrix}2 & 0\\ 0 & 1/2\end{bmatrix}\begin{bmatrix}\sqrt{3}/2 & -\sqrt{3}/2\\ \sqrt{3}/2 & \sqrt{3}/2\end{bmatrix} \tag{9}$$

矩阵 B 为：

$$\begin{aligned}B=V\Lambda^{\frac{1}{2}}V^{-1}&=\begin{bmatrix}\sqrt{2}/2 & \sqrt{2}/2\\ -\sqrt{2}/2 & \sqrt{2}/2\end{bmatrix}\begin{bmatrix}\sqrt{2} & 0\\ 0 & \sqrt{2}/2\end{bmatrix}\begin{bmatrix}\sqrt{2}/2 & -\sqrt{2}/2\\ \sqrt{2}/2 & \sqrt{2}/2\end{bmatrix}\\ &=\begin{bmatrix}1 & 1/2\\ -1 & 1/2\end{bmatrix}\begin{bmatrix}\sqrt{2}/2 & -\sqrt{2}/2\\ \sqrt{2}/2 & \sqrt{2}/2\end{bmatrix}=\begin{bmatrix}3\sqrt{2}/4 & -\sqrt{2}/4\\ -\sqrt{2}/4 & 3\sqrt{2}/4\end{bmatrix}\end{aligned} \tag{10}$$

Bk4_Ch14_01.py 求解上述例子中 A 的平方根。

14.2 矩阵指数：幂级数的推广

给定一个标量 a，指数 e^a 可以用幂级数展开表达：

$$e^a=\exp(a)=1+a+\frac{1}{2!}a^2+\frac{1}{3!}a^3+\cdots \tag{11}$$

对于 (11) 这个式子感到生疏的读者，可以回顾《数学要素》第 17 章有关泰勒展开内容。

类似地，对于方阵 A，可以定义**矩阵指数** (matrix exponential) e^A 为一个收敛幂级数：

$$e^A=\exp(A)=I+A+\frac{1}{2!}A^2+\frac{1}{3!}A^3+\cdots \tag{12}$$

如果 A 可以特征值分解得到如下等式，计算 (12) 则容易很多：

$$A=V\Lambda V^{-1} \tag{13}$$

其中，

$$\Lambda=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D\end{bmatrix} \tag{14}$$

利用特征值分解，A^k 可以写作：

$$A^k=V\Lambda^kV^{-1} \tag{15}$$

其中，k 为非负整数。

将 (15) 代入 (12)，得到：

$$
\begin{aligned}
\mathrm{e}^{\boldsymbol{A}} &= \exp(\boldsymbol{A}) = \boldsymbol{V}\boldsymbol{V}^{-1} + \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} + \frac{1}{2!}\boldsymbol{V}\boldsymbol{\Lambda}^{2}\boldsymbol{V}^{-1} + \frac{1}{3!}\boldsymbol{V}\boldsymbol{\Lambda}^{3}\boldsymbol{V}^{-1} + \cdots \\
&= \boldsymbol{V}\left(\boldsymbol{I} + \boldsymbol{\Lambda} + \boldsymbol{\Lambda}^{2} + \boldsymbol{\Lambda}^{3} + \cdots\right)\boldsymbol{V}^{-1}
\end{aligned}
\tag{16}
$$

特别地，对角方阵 $\boldsymbol{\Lambda}$ 矩阵指数为：

$$
\mathrm{e}^{\boldsymbol{\Lambda}} = \exp(\boldsymbol{\Lambda}) = \boldsymbol{I} + \boldsymbol{\Lambda} + \frac{1}{2!}\boldsymbol{\Lambda}^{2} + \frac{1}{3!}\boldsymbol{\Lambda}^{3} + \cdots
\tag{17}
$$

容易计算对角阵 $\boldsymbol{\Lambda}$ 矩阵指数 $\mathrm{e}^{\boldsymbol{\Lambda}}$：

$$
\begin{aligned}
\mathrm{e}^{\boldsymbol{\Lambda}} &= \exp(\boldsymbol{\Lambda}) = \boldsymbol{I} + \boldsymbol{\Lambda} + \frac{1}{2!}\boldsymbol{\Lambda}^{2} + \frac{1}{3!}\boldsymbol{\Lambda}^{3} + \cdots \\
&= \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} + \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix} + \frac{1}{2!}\begin{bmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_D^2 \end{bmatrix} + \cdots \\
&= \lim_{n\to\infty}\begin{bmatrix} \sum_{k=1}^{n}\frac{1}{k!}\lambda_1^k & & & \\ & \sum_{k=1}^{n}\frac{1}{k!}\lambda_2^k & & \\ & & \ddots & \\ & & & \sum_{k=1}^{n}\frac{1}{k!}\lambda_D^k \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{e}^{\lambda_1} & & & \\ & \mathrm{e}^{\lambda_2} & & \\ & & \ddots & \\ & & & \mathrm{e}^{\lambda_D} \end{bmatrix}
\end{aligned}
\tag{18}
$$

将 (17) 代入 (16)，得到：

$$
\mathrm{e}^{\boldsymbol{A}} = \boldsymbol{V}\,\mathrm{e}^{\boldsymbol{\Lambda}}\,\boldsymbol{V}^{-1}
\tag{19}
$$

将 (18) 代入 (19)，得到：

$$
\mathrm{e}^{\boldsymbol{A}} = \boldsymbol{V}\begin{bmatrix} \mathrm{e}^{\lambda_1} & & & \\ & \mathrm{e}^{\lambda_2} & & \\ & & \ddots & \\ & & & \mathrm{e}^{\lambda_D} \end{bmatrix}\boldsymbol{V}^{-1}
\tag{20}
$$

可以用 `scipy.linalg.expm()` 计算矩阵指数。

14.3 斐波那契数列：求通项式

本系列丛书《数学要素》第 14 章介绍过斐波那契数列 (Fibonacci number)，本节介绍如何使用特征值分解推导得到斐波那契数列通项式。

斐波那契数列可以通过如下**递归** (recursion) 方法获得：

$$\begin{cases} F_0 = 0 \\ F_1 = F_2 = 1 \\ F_n = F_{n-1} + F_{n-2}, \quad n > 2 \end{cases} \tag{21}$$

包括第 0 项，斐波那契数列的前 11 项为：

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55 \tag{22}$$

构造列向量

将斐波那契数列连续每两项写成列向量形式：

$$\boldsymbol{x}_0 = \begin{bmatrix} F_0 \\ F_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{x}_1 = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{x}_3 = \begin{bmatrix} F_3 \\ F_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \boldsymbol{x}_4 = \begin{bmatrix} F_4 \\ F_5 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \cdots \tag{23}$$

图 1 所示为列向量连续变化过程，能够看到它们逐渐收敛到一条直线上。这条直线通过原点，斜率就是**黄金分割** (golden ratio)：

$$\varphi = \frac{\sqrt{5}+1}{2} \approx 1.61803 \tag{24}$$

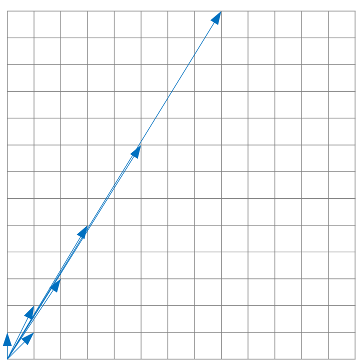


图 1. 斐波那契数列列向量连续变化过程

连续列向量间关系

数列的第 $k+1$ 项 $\boldsymbol{x}_{k+1}$ 和第 k 项 $\boldsymbol{x}_k$ 之间的关系可以写成如下矩阵运算：

$$\boldsymbol{x}_{k+1} = \begin{bmatrix} F_{k+1} \\ F_{k+2} \end{bmatrix} = \boldsymbol{A}\boldsymbol{x}_k = \boldsymbol{A}\begin{bmatrix} F_k \\ F_{k+1} \end{bmatrix} \tag{25}$$

其中

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \tag{26}$$

观察上式 $\boldsymbol{A}$，发现 $\boldsymbol{A}$ 对应的几何操作是“剪切 + 置换”的合成。

有了 (25)，$\boldsymbol{x}_k$ 可以写成：

$$\begin{aligned} \boldsymbol{x}_k &= \boldsymbol{A}\boldsymbol{x}_{k-1} \\ &= \boldsymbol{A}^2\boldsymbol{x}_{k-2} \\ &= \boldsymbol{A}^3\boldsymbol{x}_{k-3} \\ &\cdots \\ &= \boldsymbol{A}^k\boldsymbol{x}_0 \end{aligned} \tag{27}$$

特征值分解

对 $\boldsymbol{A}$ 进行特征值分解：

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} \tag{28}$$

其中，

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix}, \quad \boldsymbol{V}^{-1} = \frac{1}{\lambda_2 - \lambda_1}\begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix} \tag{29}$$

$\boldsymbol{A}$ 的特征方程为：

$$\lambda^2 - \lambda - 1 = 0 \tag{30}$$

求解 (30)，可以得到两个特征值：

$$\lambda_1 = \frac{1-\sqrt{5}}{2}, \quad \lambda_2 = \frac{1+\sqrt{5}}{2} \tag{31}$$

$\boldsymbol{x}_k$ 可以写成：

$$\boldsymbol{x}_k = \boldsymbol{V}\boldsymbol{\Lambda}^k\boldsymbol{V}^{-1}\boldsymbol{x}_0 \tag{32}$$

将 (29) 代入 (32)，得到：

$$
\begin{aligned}
\boldsymbol{x}_k &= \frac{1}{\lambda_2 - \lambda_1}\begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix}\begin{bmatrix} \lambda_1^k & \\ & \lambda_2^k \end{bmatrix}\begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \frac{1}{\lambda_2 - \lambda_1}\begin{bmatrix} \lambda_2^k - \lambda_1^k \\ \lambda_2^{k+1} - \lambda_1^{k+1} \end{bmatrix}
\end{aligned} \tag{33}
$$

即，

$$
\begin{bmatrix} F_k \\ F_{k+1} \end{bmatrix} = \frac{1}{\lambda_2 - \lambda_1}\begin{bmatrix} \lambda_2^k - \lambda_1^k \\ \lambda_2^{k+1} - \lambda_1^{k+1} \end{bmatrix} \tag{34}
$$

确定通项式

因此 F_k 可以写成：

$$
F_k = \frac{\lambda_2^k - \lambda_1^k}{\lambda_2 - \lambda_1} \tag{35}
$$

将 (31) 代入 (35) 得到 F_k 解析式：

$$
F_k = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k}{\sqrt{5}} \tag{36}
$$

至此，我们通过特征值分解得到斐波那契数列通项式解析式。

14.4 马尔科夫过程的平稳状态

本系列丛书在《数学要素》鸡兔同笼三部曲中虚构了"“鸡兔互变”的故事。本节回顾这个故事，并介绍如何用特征值分解求解其平稳状态。

图 2 描述鸡兔互变的比例，每晚有 30%的小鸡变成小兔，其他小鸡不变；同时，每晚有 20%小兔变成小鸡，其余小兔不变。这个转化的过程叫做**马尔科夫过程** (Markov process)。

马尔科夫过程满足以下三个性质：(1) 可能输出状态有限；(2) 下一步输出的概率仅仅依赖上一步的输出状态；(3) 概率值相对于时间为常数。

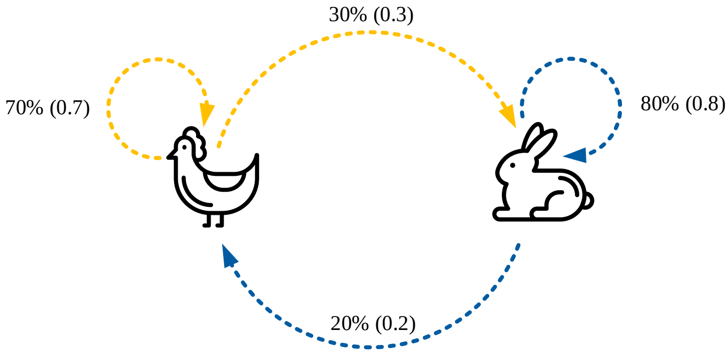


图 2. 鸡兔互变的比例

“鸡兔互变”这个例子中，第 k 天，鸡兔的比例用列向量 $\boldsymbol{\pi}(k)$ 表示；其中，$\boldsymbol{\pi}(k)$ 第一行元素代表小鸡的比例，第二行元素代表小兔的比例。第 $k+1$ 天，鸡兔的比例用列向量 $\boldsymbol{\pi}(k+1)$ 表示。

变化的比例写成方阵 $\boldsymbol{T}$，$\boldsymbol{T}$ 通常叫做**转移矩阵** (transition matrix)。

这样 $k \to k+1$ 变化过程可以写成：

$$k \to k+1: \quad \boldsymbol{T}\boldsymbol{\pi}(k) = \boldsymbol{\pi}(k+1) \tag{37}$$

对于鸡兔互变，$\boldsymbol{T}$ 为：

$$\boldsymbol{T} = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix} \tag{38}$$

求平稳状态

观察图 3，我们初步得出结论不管初始状态向量 ($k = 0$) 如何，鸡兔比例最后都达到了一定的平衡，也就是：

$$\boldsymbol{T}\boldsymbol{\pi} = \boldsymbol{\pi} \tag{39}$$

有了本书特征值分解相关的知识，相信大家一眼就看出来 (39) 代表的关系就是特征值分解。

对 $\boldsymbol{T}$ 进行特征值分解得到两个特征向量：

$$\boldsymbol{v}_1 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 0.5547 \\ 0.8321 \end{bmatrix} \tag{40}$$

鸡兔总比例之和为 1，且非负。因此选择 $\boldsymbol{v}_2$ 来计算 $\boldsymbol{\pi}$：

$$\boldsymbol{\pi} = \frac{1}{0.5547 + 0.8321}\boldsymbol{v}_2 = \frac{1}{0.5547 + 0.8321}\begin{bmatrix} 0.5547 \\ 0.8321 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} \tag{41}$$

这个 $\boldsymbol{\pi}$ 叫做**平稳状态** (steady state)。

Bk4_Ch14_02.py 绘制图 3。

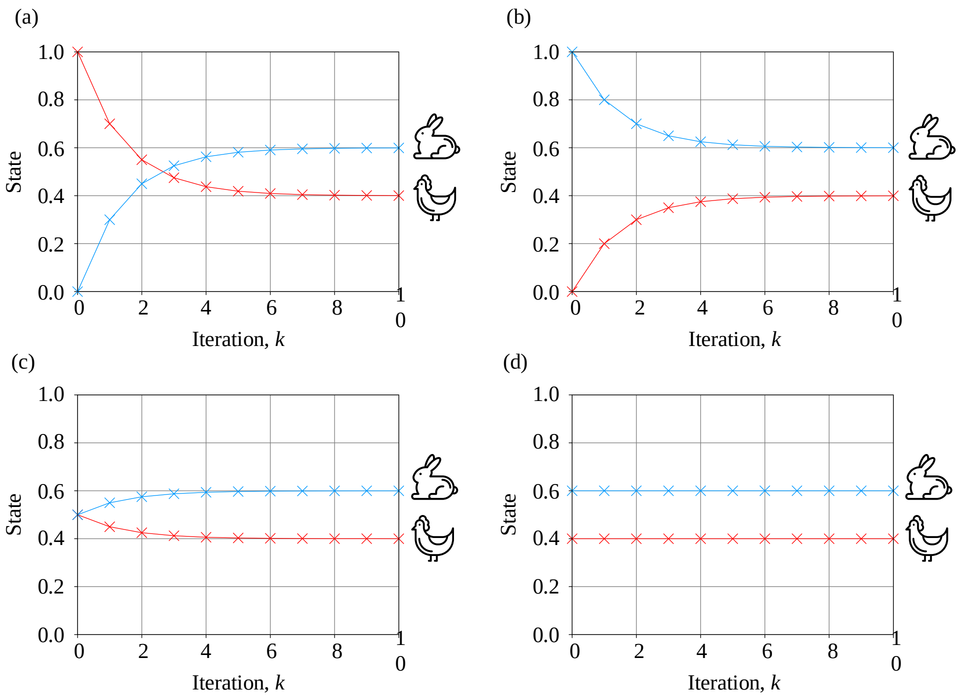


图 3. 不同初始状态条件下平稳状态

看过本系列丛书《数学要素》一册的读者应该还记得图 4 这幅图，它从几何视角描述了不同初始状态向量条件下，经过连续 12 次变化，向量都收敛于同一方向。

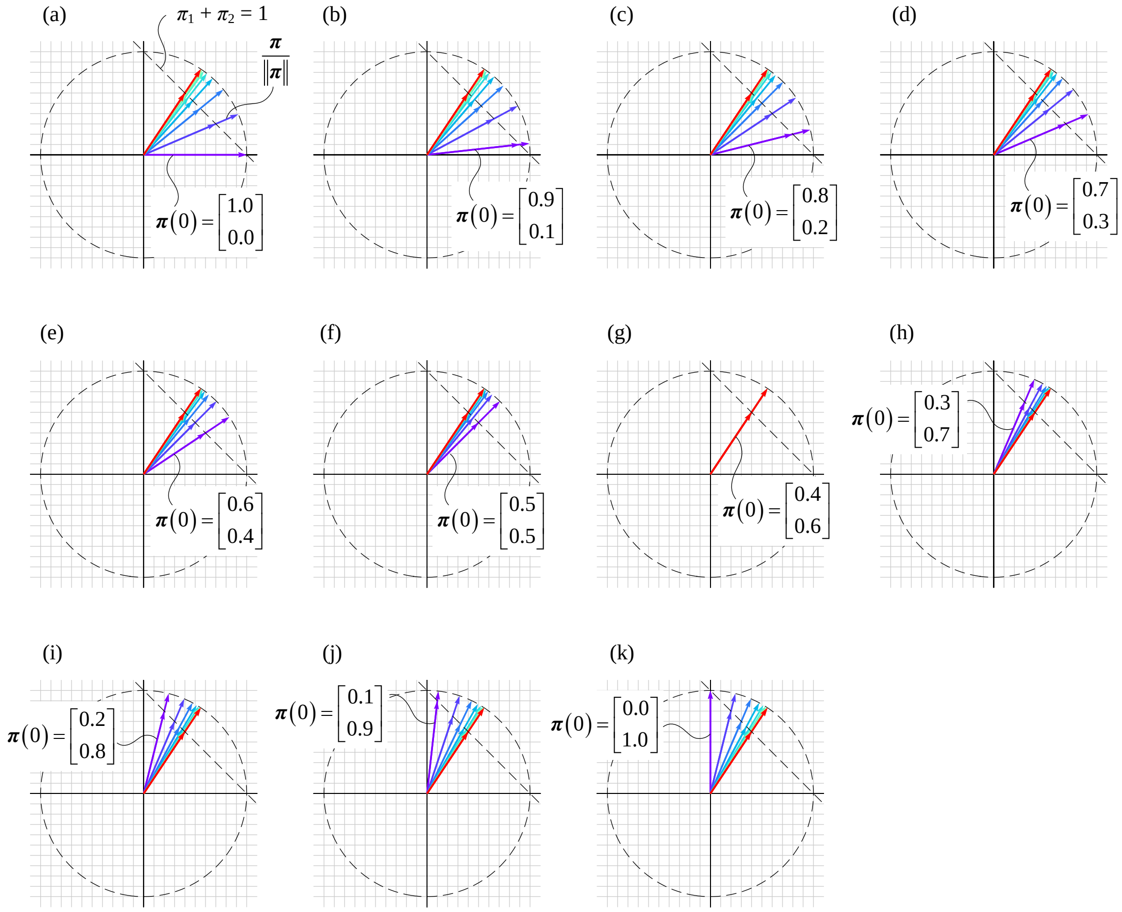


图 4. 连续 12 夜鸡兔互变比例，几何视角，图片来自《数学要素》

14.5 瑞利商

瑞利商 (Rayleigh quotient) 在很多机器学习算法中扮演重要角色，瑞利商和特征值分解有着密切关系。本节利用几何视角可视化瑞利商，让大家深入理解瑞利商这个概念。

定义

给定实数对称矩阵 $\boldsymbol{A}$，它的瑞利商定义为：

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}} \tag{42}$$

其中，$\boldsymbol{x} = [x_1, x_2, \ldots, x_D]^{\mathrm{T}}$。

⚠ 注意，(42) 中 $\boldsymbol{x}$ 不能为零向量 $\boldsymbol{0}$，也就是说，$x_1, x_2, \ldots, x_D$ 不能同时为 0。

先给出结论，瑞利商 $R(\boldsymbol{x})$ 的取值范围：

$$\lambda_{\min} \le R(\boldsymbol{x}) \le \lambda_{\max} \tag{43}$$

其中，$\lambda_{\min}$ 和 $\lambda_{\max}$ 分别为矩阵 $\boldsymbol{A}$ 的最小和最大特征值。

最大值和最小值

求解 (42) 中 $R(\boldsymbol{x})$ 的最大、最小值，等价于 $R(\boldsymbol{x})$ 分母为定值条件下，求解分子的最大值和最小值。

令 $\boldsymbol{x}$ 为单位向量，即：

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \|\boldsymbol{x}\|_2^2 = 1 \quad \Leftrightarrow \quad \|\boldsymbol{x}\|_2 = 1 \tag{44}$$

$\boldsymbol{A}$ 为对称矩阵，对其特征值分解得到：

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} \tag{45}$$

$R(\boldsymbol{x})$ 的分子可以写成：

$$\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\right)^{\mathrm{T}}\boldsymbol{\Lambda}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\right) = \left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\right)^{\mathrm{T}}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D\end{bmatrix}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\right) \tag{46}$$

令

$$\boldsymbol{y} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x} \tag{47}$$

这样，(47) 可以写成：

$$\boldsymbol{y}^{\mathrm{T}}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D\end{bmatrix}\boldsymbol{y} = \begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_D\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D\end{bmatrix}\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_D\end{bmatrix} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_D y_D^2 \tag{48}$$

类似地，$R(\boldsymbol{x})$ 的分母可以写成：

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\right)^{\mathrm{T}}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\right) = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} = y_1^2 + y_2^2 + \cdots + y_D^2 = 1 \tag{49}$$

这样，瑞利商就可以简洁地写成以 $\boldsymbol{y}$ 为自变量的函数 $R(\boldsymbol{y})$：

$$R(\boldsymbol{y}) = \frac{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_D y_D^2}{y_1^2 + y_2^2 + \cdots + y_D^2} \tag{50}$$

举个例子

下面，我们以 2×2 矩阵为例，讲解如何求解瑞利商。给定 $\boldsymbol{A}$ 为：

$$\boldsymbol{A} = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix} \tag{51}$$

$R(\boldsymbol{x})$ 为：

$$R(\boldsymbol{x}) = \frac{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} = \frac{1.5x_1^2 + x_1x_2 + 1.5x_2^2}{x_1^2 + x_2^2} \tag{52}$$

$\boldsymbol{A}$ 的两个特征值分别为 $\lambda_1 = 2$，$\lambda_2 = 1$。$R(\boldsymbol{x})$ 等价于 $R(\boldsymbol{y})$，根据 (50)，$R(\boldsymbol{y})$ 写成：

$$R(\boldsymbol{y}) = \frac{y_1^2 + 2y_2^2}{y_1^2 + y_2^2} \tag{53}$$

推导最值

求解 $R(\boldsymbol{y})$ 的最大、最小值，等价于 $R(\boldsymbol{y})$ 分母为 1 条件下，分子的最大值和最小值。

简单推导 $R(\boldsymbol{y})$ 最大值：

$$R(\boldsymbol{y}) = y_1^2 + 2y_2^2 \le 2\underbrace{\left(y_1^2 + y_2^2\right)}_{1} = 2 \tag{54}$$

推导 $R(\boldsymbol{y})$ 最小值：

$$R(\boldsymbol{y}) = y_1^2 + 2y_2^2 \ge \underbrace{\left(y_1^2 + y_2^2\right)}_{1} = 1 \tag{55}$$

几何视角

下面我们用几何方法来解释瑞利商。

(52) 的分母为 1，意味着分母代表的几何图形是个单位圆，即，

$$x_1^2 + x_2^2 = 1 \tag{56}$$

(52) 分子对应二次函数：

$$f(x_1, x_2) = 1.5x_1^2 + x_1x_2 + 1.5x_2^2 \tag{57}$$

这个二次函数对应的等高线图如所示图 5 (a) 所示。$f(x_1, x_2)$ 等高线和单位圆相交的交点中找到 $f(x_1, x_2)$ 取得最大值和最小值点。最大特征值 λ_1 对应的特征向量 $\boldsymbol{v}_1$，$\boldsymbol{v}_1$ 这个方向上做一条直线，直线和单位圆交点 (x_1, x_2) 对应的就是瑞利商的最大值点；此时，瑞利商的最大值为 λ_1。

图 6 (a) 所示为 $f(x_1, x_2)$ 曲面，以及单位圆在曲面上的映射得到的曲线。

从视角来看，上述问题实际上是个含约束优化问题，本书第 18 章将介绍如何利用拉格朗日乘子法将含约束优化问题转化为无约束优化问题。

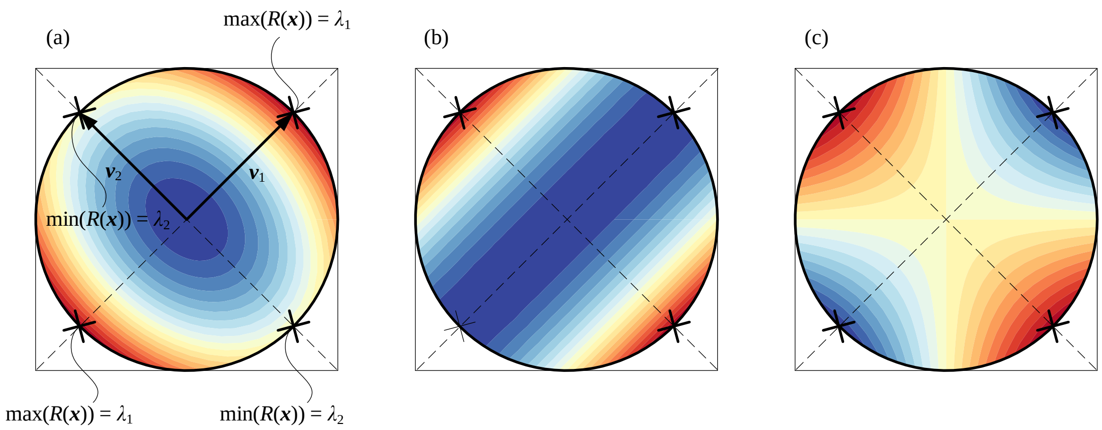


图 5. 平面上可视化 $f(x_1, x_2)$ 和单位圆

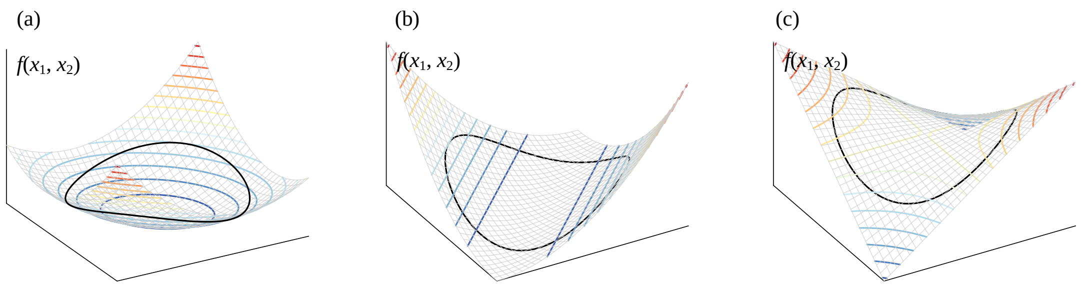


图 6. 三维空间中可视化 $f(x_1, x_2)$ 和单位圆

Bk4_Ch14_03.py 绘制图 5 和图 6。

采用单位圆作为限制条件是为了简化瑞利商对应的优化问题，而且单位圆正好是单位向量终点的落点。实际上满足瑞利商最大值的点 (x_1, x_2) 有无数个，它们都位于特征向量 $\boldsymbol{v}_1$ 所在直线上。我们能从图 7 中一睹瑞利商 $R(x_1, x_2)$ 曲面形状真容，以及瑞利商最大值和最小值对应的 (x_1, x_2) 坐标值。

注意，瑞利商 $R(x_1, x_2)$ 在 $(0, 0)$ 没有定义。

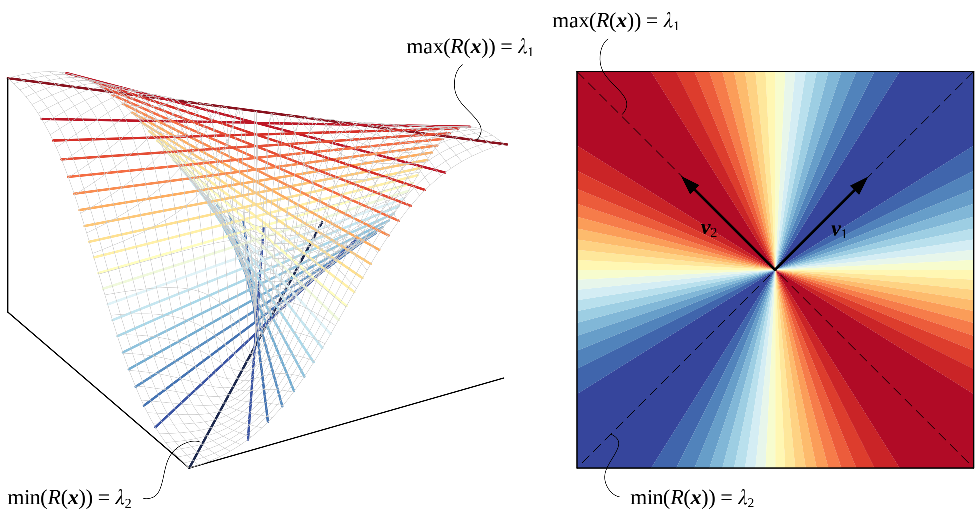


图 7. 三维空间中可视化瑞利商

再举两个例子

给定矩阵 $\boldsymbol{A}$:

$$\boldsymbol{A}=\begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix} \tag{58}$$

它的特征值分别为 $\lambda_1 = 1$，$\lambda_2 = 0$。$f(x_1, x_2)$ 等高线和曲面如图 5 (b) 和图 6 (b)所示。

图 5 (c) 等高线对应的矩阵 $\boldsymbol{A}$ 为:

$$\boldsymbol{A}=\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \tag{59}$$

它的特征值分别为 $\lambda_1 = 1$，$\lambda_2 = -1$。图 6 (c) 所示为 $f(x_1, x_2)$ 曲面的形状。

三维空间

以上探讨的三种情况都是以 2 × 2 矩阵为例。在三维空间中，$D = 3$ 这种情况，(44) 对应的是一个单位圆球体，将 $f(x_1, x_2, x_3)$ 三元函数的数值以等高线的形式映射到单位圆球体，得到图 8。

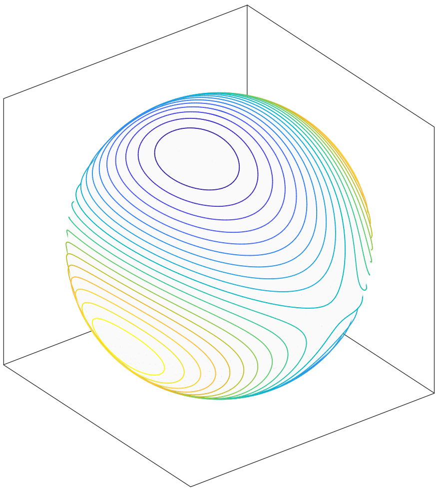


图 8. 三维单位球体表面瑞利商值等高线

14.6 再谈椭圆：特征值分解

从《数学要素》一册开始，本系列丛书几次三番谈及椭圆。这是因为圆锥曲线，特别是椭圆，在机器学习中扮演重要角色。本章最后将结合线性变换和特征值分解再聊聊椭圆。

平面上，圆心位于原点半径为 1 的正圆叫做**单位圆** (unit circle)，解析式可以写成如下形式：

$$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}-1=0 \tag{60}$$

其中 $\boldsymbol{z}$ 为，

$$\boldsymbol{z}=\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{61}$$

利用 L^2 范数，(60) 可以写成：

$$\|\boldsymbol{z}\|=1 \tag{62}$$

经过 $\boldsymbol{A}$ 映射向量 $\boldsymbol{z}$ 变成 $\boldsymbol{x}$：

$$\boldsymbol{x}=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\boldsymbol{A}\boldsymbol{z} \tag{63}$$

假设 $\boldsymbol{A}$ 可逆，也就是说 $\boldsymbol{A}$ 对应的几何操作可逆，$\boldsymbol{z}$ 可以写成：

$$\boldsymbol{z}=\boldsymbol{A}^{-1}\boldsymbol{x} \tag{64}$$

将 (64) 代入 (60) 得到：

$$\left(\boldsymbol{A}^{-1}\boldsymbol{x}\right)^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{x}-1=0 \tag{65}$$

利用 L^2 范数，(65) 还可以写成：

$$\left\| A^{-1} x \right\| = 1 \tag{66}$$

整理上式得到如下二次型：

$$x^{\mathrm{T}} \underbrace{\left(A A^{\mathrm{T}} \right)^{-1}}_{Q} x - 1 = 0 \tag{67}$$

举个例子

以本章开头 (8) 给出的矩阵 A 为例，在 A 的映射下 $z \to x = Az$：

$$x = \underbrace{\begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix}}_{A} z \tag{68}$$

如图 9 所示，满足 (60) 的向量 z 终点落在单位圆上。经过 $x = Az$ 映射后，向量 x 终点落在旋转椭圆上。

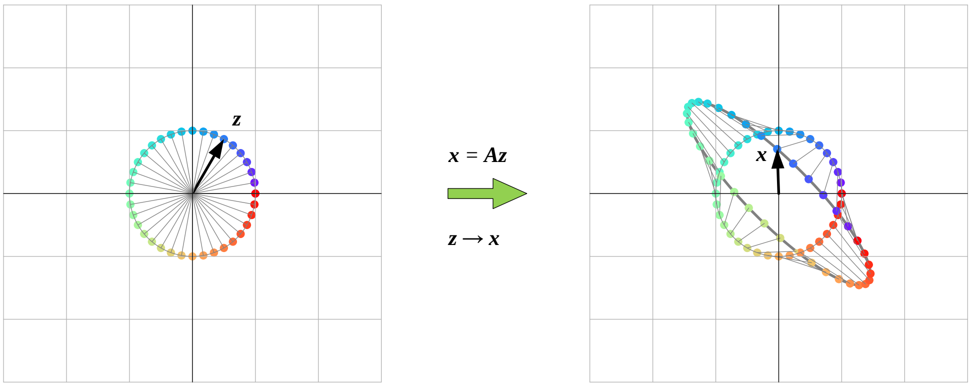


图 9. 单位圆到旋转椭圆

将 (8) 给定 A 代入 (67)，得到图 9 右侧旋转椭圆解析式如下：

$$2.125 x_1^2 + 3.75 x_1 x_2 + 2.125 x_2^2 - 1 = 0 \tag{69}$$

如果有人问我们，图 9 右侧旋转椭圆的半长轴、半短轴多长？椭圆长轴旋转角度多大？

为了解决这些问题，我们需要借助特征值分解。

特征值分解

令 Q 为：

$$Q = \left(A A^{\mathrm{T}} \right)^{-1} = \begin{bmatrix} 2.125 & 1.875 \\ 1.875 & 2.125 \end{bmatrix} \tag{70}$$

$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 显然是个对称矩阵，对称矩阵的逆还是对称矩阵，因此 $\boldsymbol{Q}$ 是对称矩阵。对 $\boldsymbol{Q}$ 进行特征值分解得到：

$$\boldsymbol{Q} = \left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} \tag{71}$$

强调一下，本节特征值分解的对象为 $\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}$，而不是 $\boldsymbol{A}$。

利用 (8) 给定 $\boldsymbol{A}$ 计算 $\boldsymbol{Q}$ 具体值，并特征值分解得到：

$$\underbrace{\begin{bmatrix} 2.125 & 1.875 \\ 1.875 & 2.125 \end{bmatrix}}_{\boldsymbol{Q}} = \underbrace{\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}}_{\boldsymbol{V}} \underbrace{\begin{bmatrix} 4 & 0 \\ 0 & 0.25 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}} \tag{72}$$

大家已经清楚上式中的 $\boldsymbol{V}$、$\boldsymbol{\Lambda}$ 对应的几何操作分别是"旋转"、"缩放"。请大家注意，$\boldsymbol{\Lambda}$ 并不是单位圆到椭圆的缩放比例。我们还需对 $\boldsymbol{\Lambda}$ 再多一步处理。

几何视角：缩放 → 旋转

整理 (71) 得到 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 对应的特征值分解：

$$\begin{aligned} \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} &= \left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\right)^{-1} = \left(\boldsymbol{V}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{\mathrm{T}} \\ &= \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}}\left(\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}}\right)^{\mathrm{T}} \end{aligned} \tag{73}$$

由于 $\boldsymbol{Q}$ 为对称矩阵，特征值分解得到的 $\boldsymbol{V}$ 为正交矩阵，因此存在 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}$。如上推导用到了这个关系。

上式告诉我们 $\boldsymbol{A}$ 相当等价于：

$$\boldsymbol{A} \sim \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}} \tag{74}$$

注意，$\boldsymbol{A} \neq \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}}$。这是因为，$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}$，不能推导得到 $\boldsymbol{A} = \boldsymbol{B}$。本书第 5 章强调过这一点。

从几何角度来看，$\boldsymbol{A}$ 这个映射相当于被分解成"旋转 ($\boldsymbol{V}$) + 缩放 ($\boldsymbol{\Lambda}^{\frac{-1}{2}}$)"。将 (74) 代入 (63)，得到：

$$\boldsymbol{x} = \boldsymbol{V}\underbrace{\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{z}}_{\boldsymbol{y}} \tag{75}$$

$\boldsymbol{z}$ 先经过缩放 ($\boldsymbol{\Lambda}^{\frac{-1}{2}}$) 得到 $\boldsymbol{y}$，$\boldsymbol{y}$ 经过旋转 ($\boldsymbol{V}$) 得到 $\boldsymbol{x}$：

$$\begin{gathered} \boldsymbol{y} = \boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{z} \\ \boldsymbol{x} = \boldsymbol{V}\boldsymbol{y} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{z} \end{gathered} \tag{76}$$

图 10 所示为上述“单位圆 → 正椭圆 → 旋转椭圆”几何变换过程。比较图 9 和图 10，容易发现形状上旋转椭圆完全相同。但是大家如果仔细比较图 9 和图 10 上，可以发现“彩灯”位置并不相同。这个差异来自于 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}$ 不能推导得到 $\boldsymbol{A} = \boldsymbol{B}$。

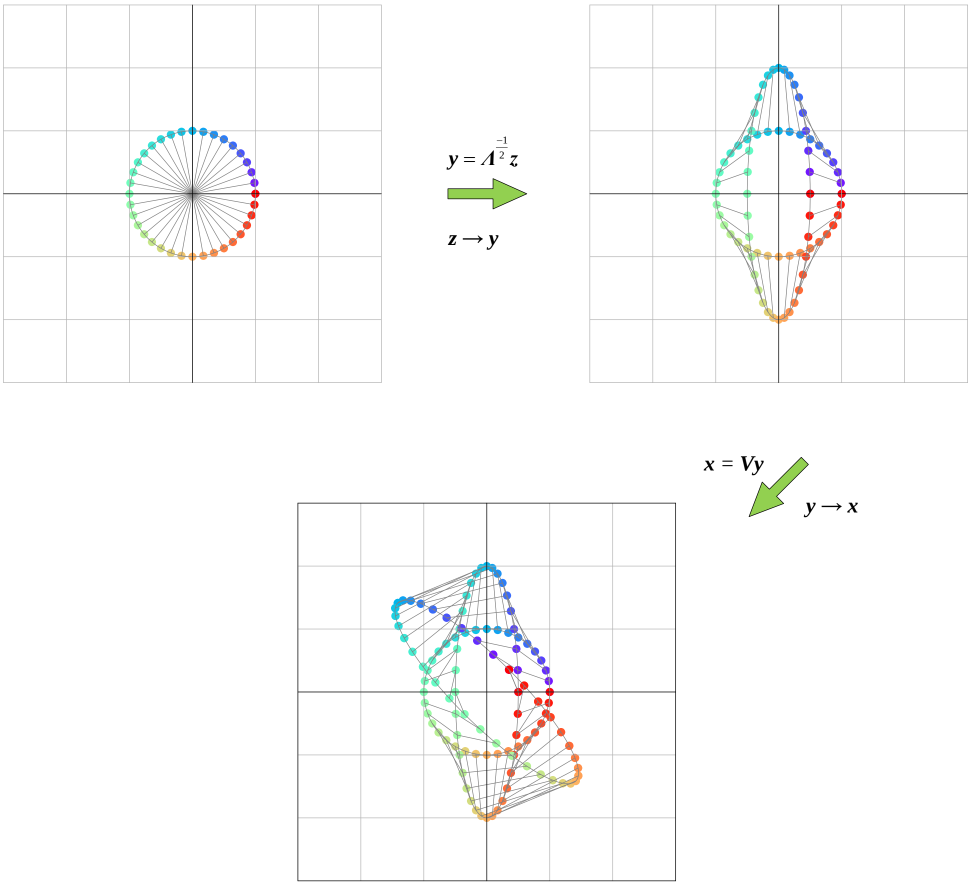


图 10. 单位圆 → 正椭圆 → 旋转椭圆

椭圆长、短轴

利用 $\boldsymbol{y}$ 和 $\boldsymbol{z}$ 的关系，(60) 可以写成：

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{y} - 1 = 0 \tag{77}$$

即：

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} - 1 = 0 \quad \Rightarrow \quad \lambda_1 y_1^2 + \lambda_2 y_2^2 = 1 \tag{78}$$

写成大家熟悉的椭圆形式：

$$\frac{y_1^2}{\left(1\big/\sqrt{\lambda_1}\right)^2} + \frac{y_2^2}{\left(1\big/\sqrt{\lambda_2}\right)^2} = 1 \tag{79}$$

如果 $\lambda_1 > \lambda_2 > 0$，上式中这个正椭圆的半长轴长度为 $\sqrt{1/\lambda_2}$，半短轴长度为 $\sqrt{1/\lambda_1}$。实际上，我们在本书第 5 章接触过这个结论。

代入具体值，得到：

$$\frac{y_1^2}{0.5^2} + \frac{y_2^2}{2^2} - 1 = 0 \tag{80}$$

图 11 所示为旋转椭圆的长轴、短轴位置，以及半长轴、半短轴长度。

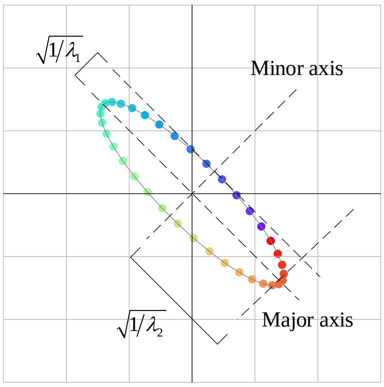
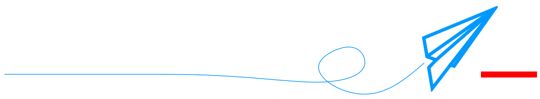


图 11. 旋转椭圆长轴、短轴



本章主要着墨在特征值分解的应用，比如方阵开方、矩阵指数、斐波那契数列、马尔科夫过程平衡状态等等。

本章特别值得注意的一个知识点是瑞利商，数据科学和机器学习很多算法中都离不开瑞利商。希望大家能从几何视角理解瑞利商的最值。本书还将在拉格朗日乘子法中继续探讨瑞利商。

本章最后讨论了如何用特征值分解获得旋转椭圆的半长轴、半短轴长度，以及旋转角度等位置信息。这部分内容和《概率统计》一册中多元高斯分布关系密切。

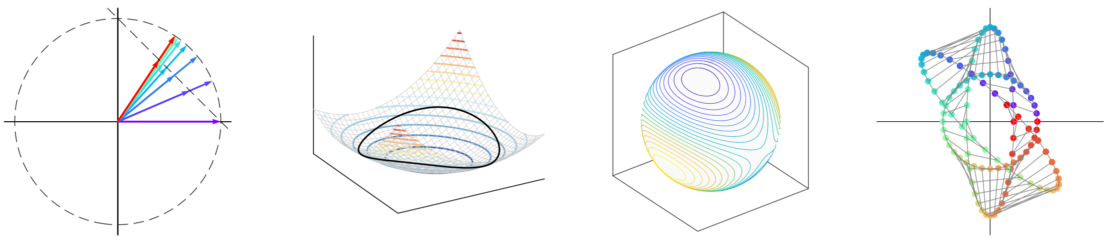


图 12. 总结本章重要内容的四副图