13 Eigen Decomposition 特征值分解

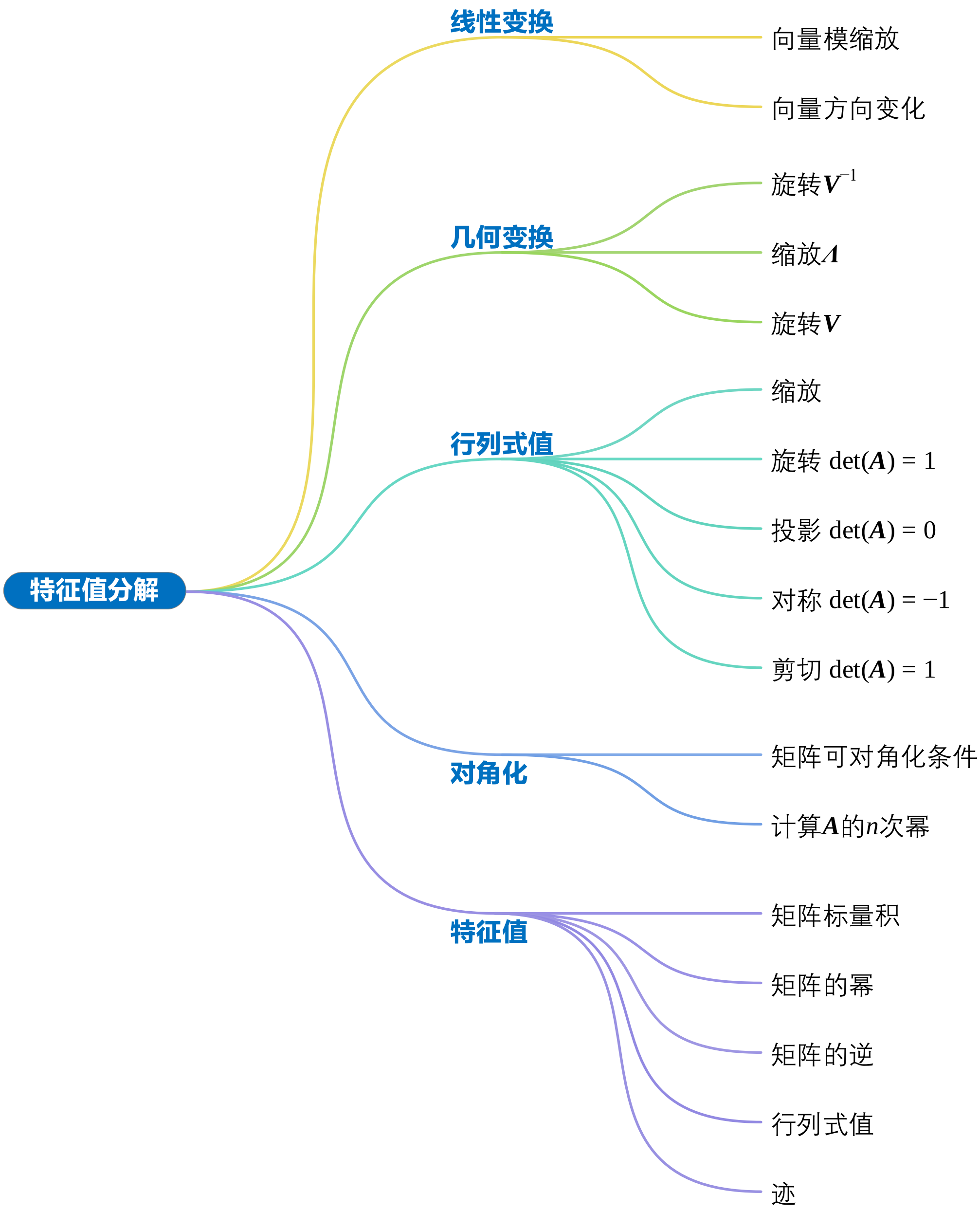
旋转 → 缩放 → 旋转

如果不能用数学表达，人类任何探索都不能被称之为真正的科学。

No human investigation can be called real science if it cannot be demonstrated mathematically.

—— 列奥纳多·达·芬奇 (Leonardo da Vinci) | 文艺复兴三杰之一 | 1452 ~ 1519

- `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.cos()` 计算余弦值
- `numpy.sin()` 计算正弦值
- `numpy.tan()` 计算正切值
- `numpy.flip()` 指定轴翻转数组
- `numpy.fliplr()` 左右翻转数组
- `numpy.flipud()` 上下翻转数组


特征值分解
线性变换
向量模缩放
向量方向变化
几何变换
旋转$\boldsymbol{V}^{-1}$
缩放$\boldsymbol{\Lambda}$
旋转$\boldsymbol{V}$
行列式值
缩放
旋转 $\det(\boldsymbol{A}) = 1$
投影 $\det(\boldsymbol{A}) = 0$
对称 $\det(\boldsymbol{A}) = -1$
剪切 $\det(\boldsymbol{A}) = 1$
对角化
矩阵可对角化条件
计算$\boldsymbol{A}$的n次幂
特征值
矩阵标量积
矩阵的幂
矩阵的逆
行列式值
迹

13.1 几何角度看特征值分解

本书第 8 章讲解线性变换时提到，几何视角下，方阵对应缩放、旋转、投影、剪切等几何变换中一种甚至多种的组合，而矩阵分解可以帮我们找到这些几何变换的具体成分。本章专门要讲的特征值分解能帮我们找到某些特定方阵中“缩放”和“旋转”这两个成分。

举个例子

给定如下一个矩阵 $\boldsymbol{A}$，具体如下：

$$\boldsymbol{A}=\begin{bmatrix}1.25 & -0.75\\ -0.75 & 1.25\end{bmatrix} \tag{1}$$

矩阵 $\boldsymbol{A}$ 乘向量 $\boldsymbol{w}_1$ 得到一个新向量 $\boldsymbol{A}\boldsymbol{w}_1$，比如：

$$\boldsymbol{w}_1=\begin{bmatrix}1\\ 0\end{bmatrix},\quad \boldsymbol{A}\boldsymbol{w}_1=\begin{bmatrix}1.25 & -0.75\\ -0.75 & 1.25\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}=\begin{bmatrix}1.25\\ -0.75\end{bmatrix} \tag{2}$$

如图 1 所示，从几何角度，对比原向量 $\boldsymbol{w}_1$，经过 $\boldsymbol{A}$ 的映射，$\boldsymbol{A}\boldsymbol{w}_1$ 的方向和模都发生了变化。也就是说，$\boldsymbol{A}$ 起到了缩放、旋转两方面作用。

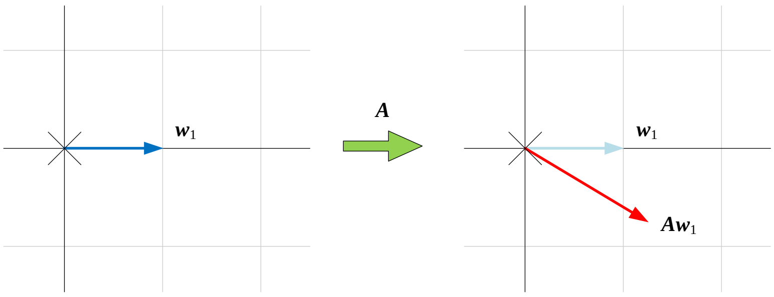


图 1. 我们发现相比原向量 $\boldsymbol{w}_1$，新向量 $\boldsymbol{A}\boldsymbol{w}_1$ 的方向和模都发生变化

图 2 给出 81 个不同朝向向量 $\boldsymbol{w}$，它们都是单位向量，即向量模均为 1。经过 $\boldsymbol{A}$ 的映射得到图 3 所示 81 个不同 $\boldsymbol{A}\boldsymbol{w}$ 结果。图 3 中，多数情况，$\boldsymbol{w}$ (蓝色箭头) 到 $\boldsymbol{A}\boldsymbol{w}$ (红色箭头) 同时发生旋转、缩放。请大家特别注意图 3 中如下四个向量 (背景为浅蓝色)：

$$\boldsymbol{w}_{11}=\begin{bmatrix}\sqrt{2}/2\\ \sqrt{2}/2\end{bmatrix},\quad \boldsymbol{w}_{31}=\begin{bmatrix}-\sqrt{2}/2\\ \sqrt{2}/2\end{bmatrix},\quad \boldsymbol{w}_{51}=\begin{bmatrix}-\sqrt{2}/2\\ -\sqrt{2}/2\end{bmatrix},\quad \boldsymbol{w}_{71}=\begin{bmatrix}\sqrt{2}/2\\ -\sqrt{2}/2\end{bmatrix} \tag{3}$$

矩阵 $\boldsymbol{A}$ 和这四个向量相乘得到的结果和原向量相比，仅仅发生缩放，也就是向量模变化，但是方向没有变化。$\boldsymbol{A}$ 对这些向量只产生缩放变换，不产生旋转效果，那么这些向量就称为 $\boldsymbol{A}$ 特征向量，伸缩的比例就是特征值。

⚠ 注意，准确来说，如果 $\boldsymbol{w}$ 是 $\boldsymbol{A}$ 的特征向量，$\boldsymbol{A}$ 和 $\boldsymbol{A}\boldsymbol{w}$ 方向平行，同向或反向。

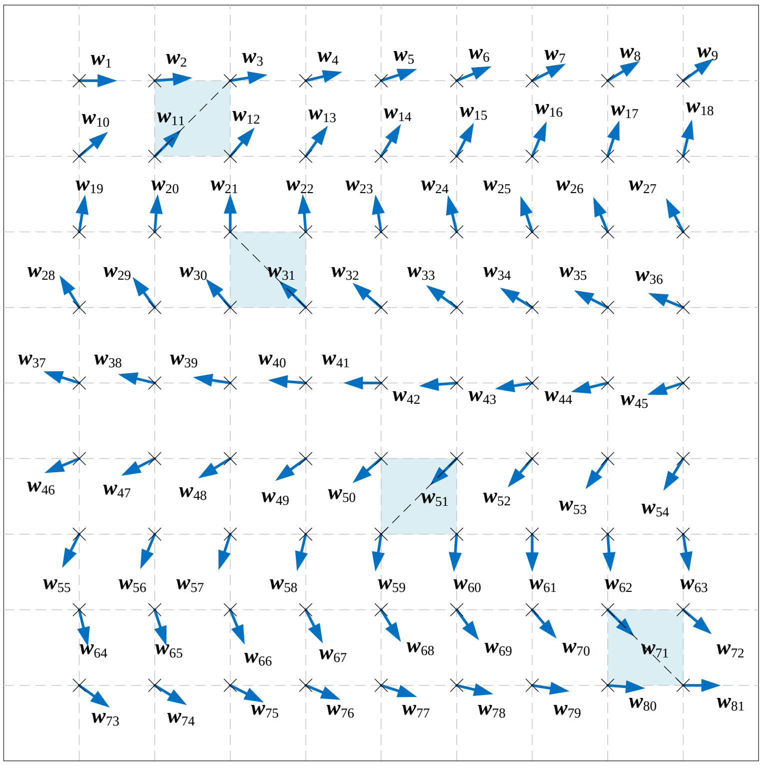


图 2. 81 个朝向不同方向的单位向量

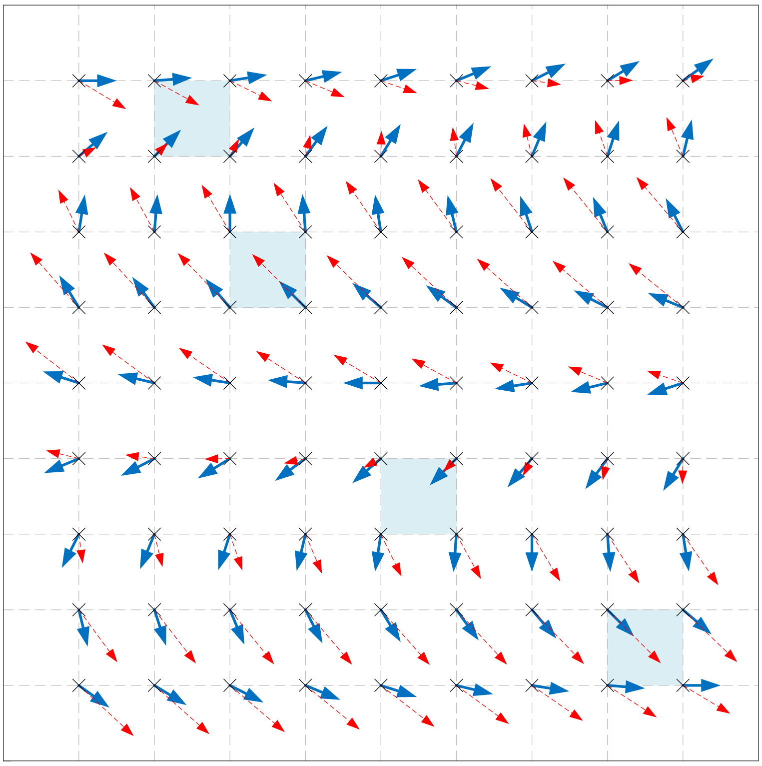


图 3. 矩阵 $\boldsymbol{A}$ 乘 $\boldsymbol{w}$ 得到的 81 个不同结果

单位圆

为了更好看清矩阵 $\boldsymbol{A}$ 的作用，我们将不同朝向的向量都放在一个单位圆中，如图 4 左图。

图 4 左图中，向量的终点落在单位圆上。为了方便可视化，图 4 左图只展示四个蓝色箭头的线段，它们都是特征向量。图 4 右图为经过 $\boldsymbol{A}$ 映射后得到向量，终点落在旋转椭圆上。对比图 4 椭圆和正圆的缩放比例，大家可以试着估算特征值大小。不禁感叹，椭圆真是无处不在。本书后文椭圆还将出现在不同场合，特别是和协方差矩阵相关的内容中。

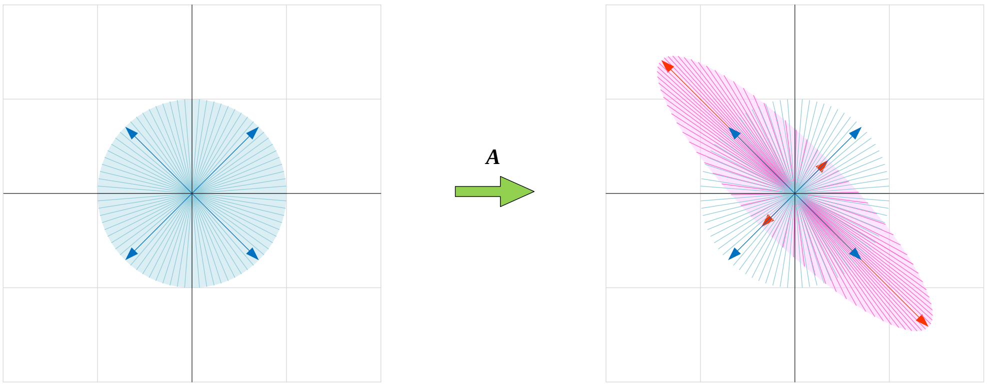


图 4. 矩阵 $\boldsymbol{A}$ 对一系列向量的映射结果

Bk4_Ch13_01.py 绘制图 2、图 3、图 4。需要说明的是，为了方便大家理解以及保证图形的矢量化，丛书不会直接使用 Python 出图。所有图片后期都经过多道美化工序。因此，大家使用代码获得的图片和书中图片存在一定差异，但是图片美化中绝不会篡改数据。

13.2 旋转 → 缩放 → 旋转

根据本书第 11 章所述，矩阵 $\boldsymbol{A}$ 的特征值分解可以写成：

$$\boldsymbol{A} = \overset{\text{Rotate}}{\boldsymbol{V}} \overset{\text{Scale}}{\boldsymbol{\Lambda}} \overset{\text{Rotate}}{\boldsymbol{V}^{-1}} \tag{4}$$

几何视角，$\boldsymbol{A}$ 乘任意向量 $\boldsymbol{w}$ 代表“旋转 → 缩放 →旋转”，即，

$$\boldsymbol{A}\boldsymbol{w} = \overset{\text{Rotate}}{\boldsymbol{V}} \overset{\text{Scale}}{\boldsymbol{\Lambda}} \overset{\text{Rotate}}{\boldsymbol{V}^{-1}} \boldsymbol{w} \tag{5}$$

⚠ 注意几何变换顺序是从右向左，即旋转 ($\boldsymbol{V}^{-1}$) → 缩放 ($\boldsymbol{\Lambda}$) →旋转 ($\boldsymbol{V}$)。

举个 2 × 2 矩阵的例子

(4) 等式右乘 $\boldsymbol{V}$ 得到：

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda} \tag{6}$$

将 $\boldsymbol{V}$ 展开写成 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 并代入上式得到：

$$\boldsymbol{A}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2\end{bmatrix}\begin{bmatrix}\lambda_1 & \\ & \lambda_2\end{bmatrix} \tag{7}$$

展开 (7) 得到：

$$\begin{bmatrix}\boldsymbol{A}\boldsymbol{v}_1 & \boldsymbol{A}\boldsymbol{v}_2\end{bmatrix} = \begin{bmatrix}\lambda_1\boldsymbol{v}_1 & \lambda_2\boldsymbol{v}_2\end{bmatrix} \tag{8}$$

对于上一节给出的例子，将具体数值代入 (4)，得到：

$$\underbrace{\begin{bmatrix}1.25 & -0.75\\ -0.75 & 1.25\end{bmatrix}}_{\boldsymbol{A}} = \underbrace{\begin{bmatrix}\sqrt{2}/2 & -\sqrt{2}/2\\ \sqrt{2}/2 & \sqrt{2}/2\end{bmatrix}}_{\boldsymbol{V}}\underbrace{\begin{bmatrix}0.5 & 0\\ 0 & 2\end{bmatrix}}_{\boldsymbol{\Lambda}}\underbrace{\begin{bmatrix}\sqrt{2}/2 & \sqrt{2}/2\\ -\sqrt{2}/2 & \sqrt{2}/2\end{bmatrix}}_{\boldsymbol{V}^{-1}} \tag{9}$$

下面，我们分别讨论 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 的几何特征。

第一特征向量

$\boldsymbol{v}_1$ 为：

$$\boldsymbol{v}_1 = \begin{bmatrix}\sqrt{2}/2\\ \sqrt{2}/2\end{bmatrix} \tag{10}$$

$\boldsymbol{A}$ 乘 $\boldsymbol{v}_1$ 得到 $\boldsymbol{A}\boldsymbol{v}_1$：

$$\boldsymbol{A}\boldsymbol{v}_1 = \begin{bmatrix}1.25 & -0.75\\ -0.75 & 1.25\end{bmatrix}\begin{bmatrix}\sqrt{2}/2\\ \sqrt{2}/2\end{bmatrix} = \begin{bmatrix}\sqrt{2}/4\\ \sqrt{2}/4\end{bmatrix} = \underbrace{\frac{1}{2}}_{\lambda_1}\times\begin{bmatrix}\sqrt{2}/2\\ \sqrt{2}/2\end{bmatrix} \tag{11}$$

可以发现，相比 $\boldsymbol{v}_1$，$\boldsymbol{A}\boldsymbol{v}_1$ 方向没有发生变化，$\boldsymbol{A}$ 仅仅产生缩放作用，缩放的比例为 $\lambda_1 = 1/2$。

图 5 中蓝色箭头代表 $\boldsymbol{v}_1$，将 (4) 代入 (11)，将 $\boldsymbol{A}$ 拆解为“旋转→缩放→旋转”三步几何操作：

$$\boldsymbol{A}\boldsymbol{v}_1 = \overset{\text{Rotate}}{\boldsymbol{V}}\ \overset{\text{Scale}}{\boldsymbol{\Lambda}}\ \overset{\text{Rotate}}{\boldsymbol{V}^{-1}}\boldsymbol{v}_1 \tag{12}$$

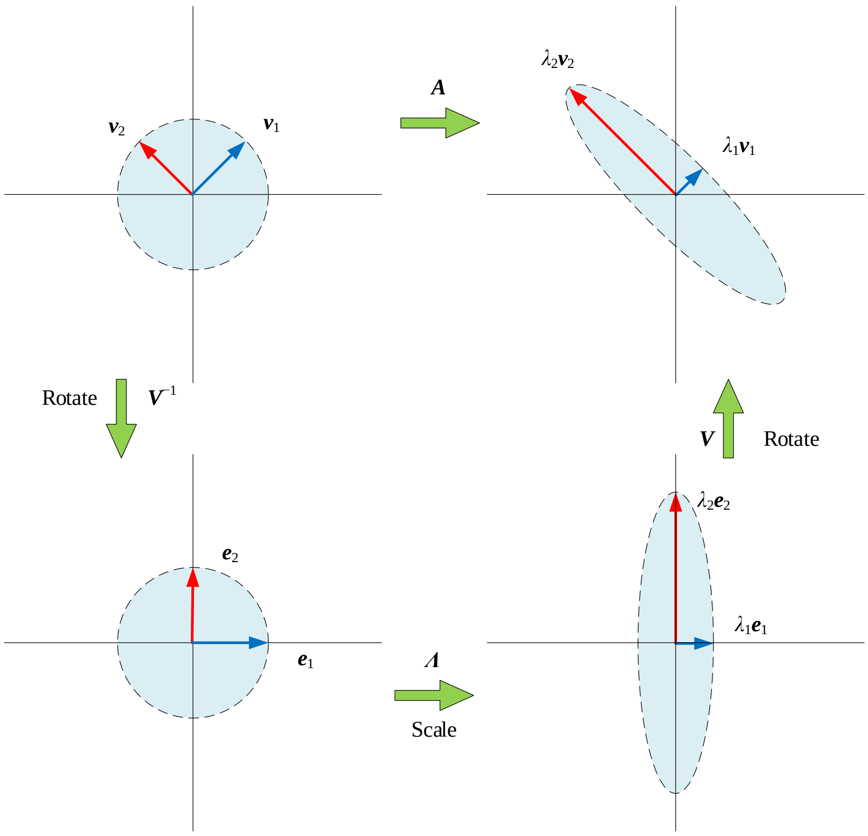


图 5. "旋转→缩放→旋转"操作

$\boldsymbol{V}^{-1}\boldsymbol{v}_1$ 相对 $\boldsymbol{v}_1$ 顺时针旋转 45°:

$$\boldsymbol{V}^{-1}\boldsymbol{v}_1=\begin{bmatrix}\sqrt{2}/2 & \sqrt{2}/2\\ -\sqrt{2}/2 & \sqrt{2}/2\end{bmatrix}\begin{bmatrix}\sqrt{2}/2\\ \sqrt{2}/2\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}=\boldsymbol{e}_1 \tag{13}$$

然后再利用 $\boldsymbol{\Lambda}$ 完成缩放操作，得到 $\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{v}_1$:

$$\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{v}_1=\begin{bmatrix}0.5 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}0.5\\0\end{bmatrix}=\underset{\lambda_1}{0.5}\boldsymbol{e}_1 \tag{14}$$

最后利用 $\boldsymbol{V}$ 完成逆时针旋转 45°，得到 $\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{v}_1$:

$$\begin{aligned}\underbrace{\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}}_{\boldsymbol{A}}\boldsymbol{v}_1&=\begin{bmatrix}\sqrt{2}/2 & -\sqrt{2}/2\\ \sqrt{2}/2 & \sqrt{2}/2\end{bmatrix}\underset{\lambda_1}{0.5}\boldsymbol{e}_1\\ &=\begin{bmatrix}\sqrt{2}/2 & -\sqrt{2}/2\\ \sqrt{2}/2 & \sqrt{2}/2\end{bmatrix}\begin{bmatrix}0.5\\0\end{bmatrix}=\begin{bmatrix}\sqrt{2}/4\\ \sqrt{2}/4\end{bmatrix}=\underset{\lambda_1}{0.5}\begin{bmatrix}\sqrt{2}/2\\ \sqrt{2}/2\end{bmatrix}\\ &=\lambda_1\boldsymbol{v}_1\end{aligned} \tag{15}$$

第二特征向量

类似地，下面讨论 $\boldsymbol{A}$ 乘 $\boldsymbol{v}_2$ 对应的“旋转→缩放→旋转”操作。

$\boldsymbol{v}_2$ 为：

$$\boldsymbol{v}_2 = \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \tag{16}$$

$\boldsymbol{A}$ 乘 $\boldsymbol{v}_2$ 得到 $\boldsymbol{A}\boldsymbol{v}_2$：

$$\boldsymbol{A}\boldsymbol{v}_2 = \begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix} \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \end{bmatrix} = \underset{\lambda_2}{2} \times \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \tag{17}$$

$\boldsymbol{V}^{-1}\boldsymbol{v}_2$ 将 $\boldsymbol{v}_2$ 顺时针旋转 45°：

$$\underset{\text{Rotate}}{\boldsymbol{V}^{-1}} \boldsymbol{v}_2 = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \boldsymbol{e}_2 \tag{18}$$

再缩放得到 $\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{v}_2$：

$$\underset{\text{Scale}}{\boldsymbol{\Lambda}} \boldsymbol{V}^{-1}\boldsymbol{v}_2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \underset{\lambda_2}{2} \boldsymbol{e}_2 \tag{19}$$

最后旋转得到 $\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{v}_2$：

$$\begin{aligned} \underset{\text{Rotate}}{\boldsymbol{V}} \boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{v}_2 &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \underset{\lambda_2}{2} \boldsymbol{e}_2 \\ &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \end{bmatrix} = \underset{\lambda_2}{2} \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \\ &= \lambda_2 \boldsymbol{v}_2 \end{aligned} \tag{20}$$

整个几何变换过程如图 5 中红色箭头所示。

Bk4_Ch13_02.py 绘制图 5。

13.3 再谈行列式值和线性变换

计算本章第一节给出矩阵 $\boldsymbol{A}$ 的行列式值 $\det(\boldsymbol{A})$：

$$\det(\boldsymbol{A}) = \det\left(\begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix}\right) = 1 \tag{21}$$

本书前文提到过 2 × 2 矩阵行列式值相当于几何变换前后“面积缩放系数”。上式中 $\boldsymbol{A}$ 的行列式值为 1，因此几何变换前后面积没有任何缩放。

这一点也可以通过 $\boldsymbol{\Lambda}$ 的行列式值加以验证：

$$\begin{aligned} \det(\boldsymbol{A}) &= \det\left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\right) = \det(\boldsymbol{V})\det(\boldsymbol{\Lambda})\det\left(\boldsymbol{V}^{-1}\right) \\ &= \det(\boldsymbol{\Lambda})\det\left(\boldsymbol{V}\boldsymbol{V}^{-1}\right) = \det(\boldsymbol{\Lambda}) \\ &= \lambda_1\lambda_2 = \frac{1}{2}\times 2 = 1 \end{aligned} \tag{22}$$

上式说明，如果 $\boldsymbol{A}$ 可以进行特征值分解，矩阵 $\boldsymbol{A}$ 的行列式值等于 $\boldsymbol{A}$ 的所有特征值之积。

图 6 给出一个正方形，内部和边缘整齐排列散点。在 $\boldsymbol{A}$ 的作用下，正方形完成“旋转→缩放→旋转”三步几何操作。不难发现，得到的菱形和原始正方形的面积一致，这一点印证了 $|\boldsymbol{A}| = 1$。

回过头来看图 4 右图旋转椭圆，它的半长轴长度为 2，而半短轴长度为 1/2。但是，得到的椭圆面积和原来单位圆面积一样。

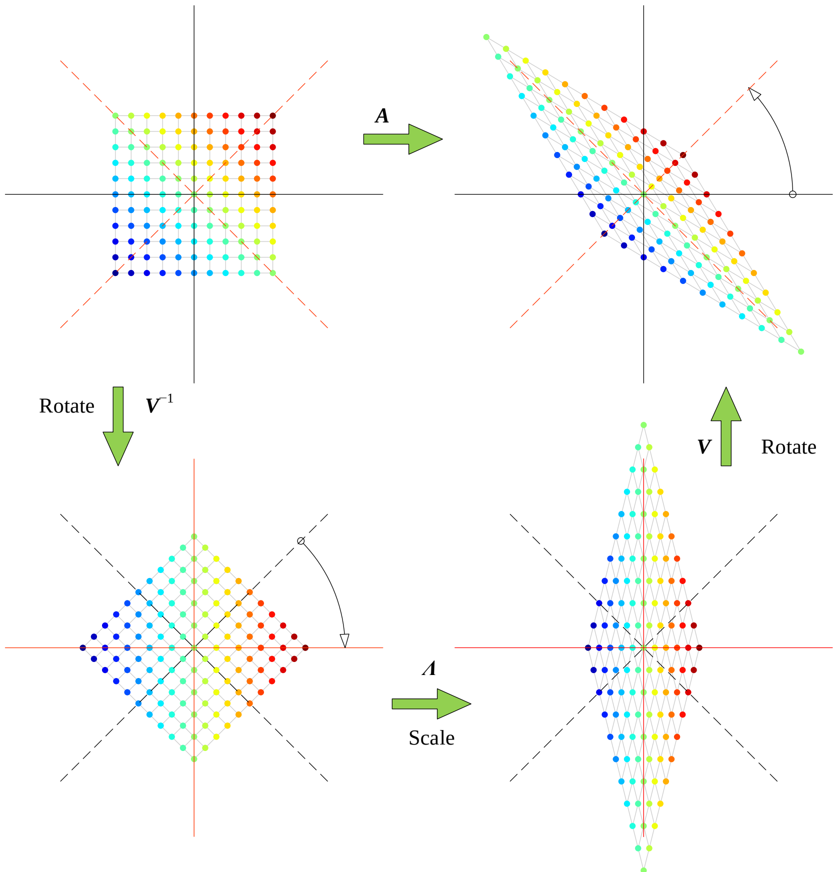


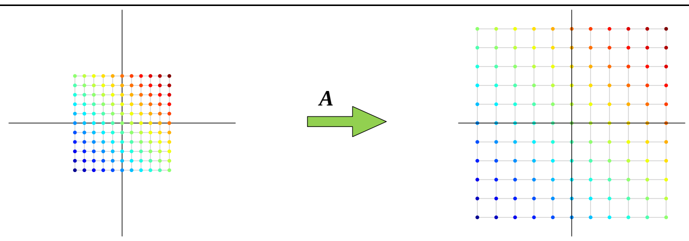
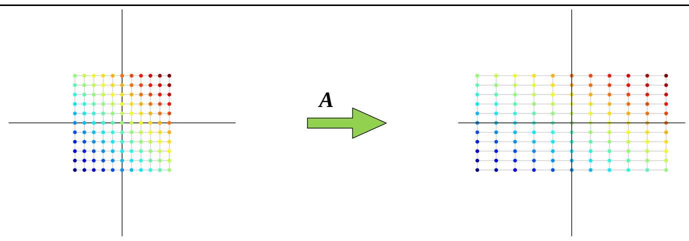
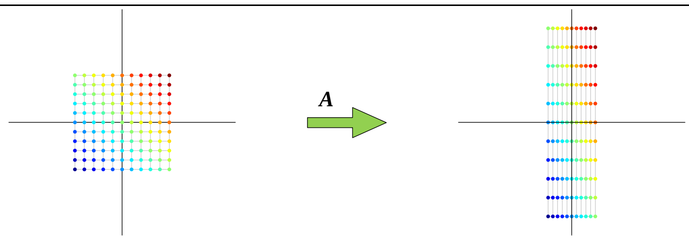
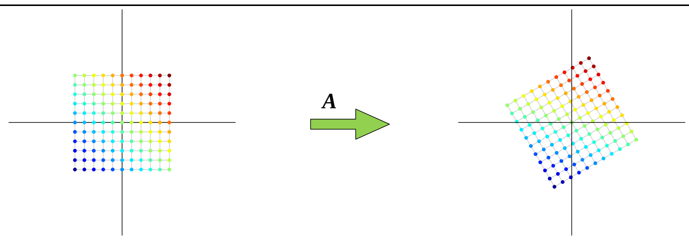
图 6. 正方形经过矩阵 $\boldsymbol{A}$ 线性变换

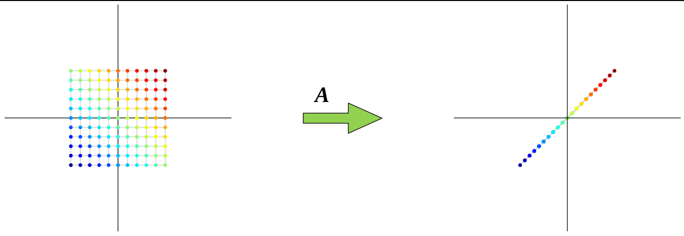
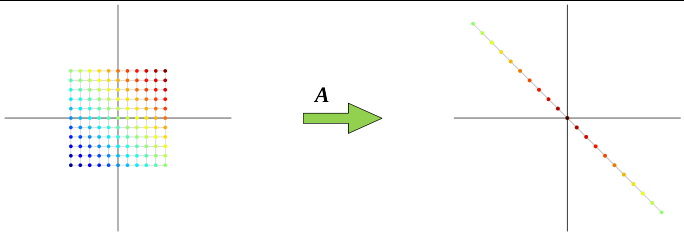
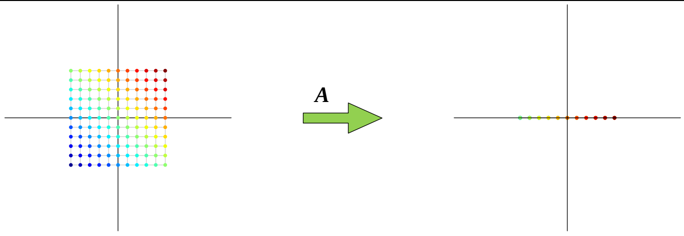
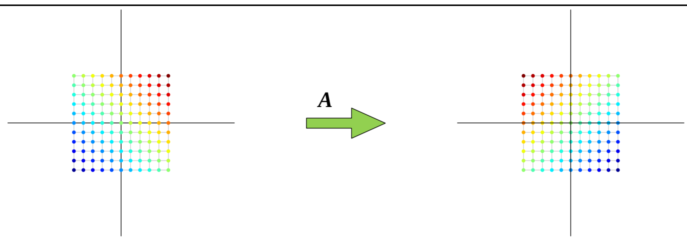
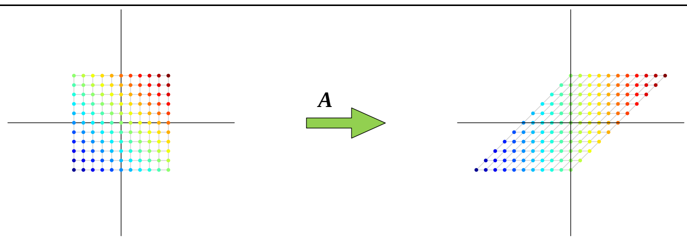
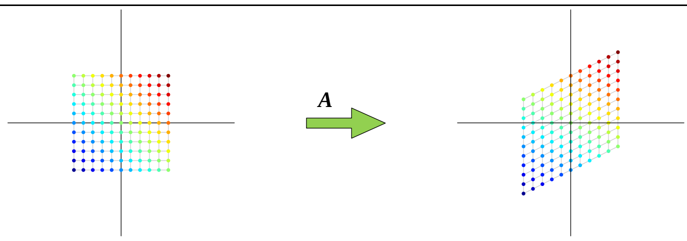
常见线性变换矩阵的特征值和行列式值

表 1 总结常见 2×2 矩阵对应的线性变换、特征值、行列式值。表 1 告诉我们特征值可以为正数、负数、0，甚至是复数。复数特征值都是成对出现，且共轭。下一章会专门讲解特征值分解中出现复数现象。

此外，请大家自行判断表中哪些矩阵可逆，也就是几何变换可逆。

表 1. 常见 2×2 矩阵对应的线性变换、特征值、行列式值

矩阵 $\boldsymbol{A}$	几何特征
等比例缩放 $\boldsymbol{A}=\begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix}$ $\begin{cases}\lambda_1 = 2\\ \lambda_2 = 2\end{cases}$ $\det(\boldsymbol{A}) = 4$	
不等比例缩放 $\boldsymbol{A}=\begin{bmatrix}2 & 0\\ 0 & 1\end{bmatrix}$ $\begin{cases}\lambda_1 = 2\\ \lambda_2 = 1\end{cases}$ $\det(\boldsymbol{A}) = 2$	
不等比例缩放 $\boldsymbol{A}=\begin{bmatrix}0.5 & 0\\ 0 & 2\end{bmatrix}$ $\begin{cases}\lambda_1 = 2\\ \lambda_2 = 0.5\end{cases}$ $\det(\boldsymbol{A}) = 1$	
旋转 $\boldsymbol{A}=\begin{bmatrix}\sqrt{3}/2 & -0.5\\ 0.5 & \sqrt{3}/2\end{bmatrix}$ $\begin{cases}\lambda_1 = \sqrt{3}/2 + 0.5i\\ \lambda_2 = \sqrt{3}/2 - 0.5i\end{cases}$ $\det(\boldsymbol{A}) = 1$	

投影 $\boldsymbol{A}=\begin{bmatrix}0.5 & 0.5\\ 0.5 & 0.5\end{bmatrix}$ $\begin{cases}\lambda_1=1\\ \lambda_2=0\end{cases}$ $\det(\boldsymbol{A})=0$	
非正交映射 $\boldsymbol{A}=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}$ $\begin{cases}\lambda_1=2\\ \lambda_2=0\end{cases}$ $\det(\boldsymbol{A})=0$	
横轴投影 $\boldsymbol{A}=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$ $\begin{cases}\lambda_1=1\\ \lambda_2=0\end{cases}$ $\det(\boldsymbol{A})=0$	
纵轴对称 $\boldsymbol{A}=\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}$ $\begin{cases}\lambda_1=1\\ \lambda_2=-1\end{cases}$ $\det(\boldsymbol{A})=-1$	
剪切 $\boldsymbol{A}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$ $\begin{cases}\lambda_1=1\\ \lambda_2=1\end{cases}$ $\det(\boldsymbol{A})=1$	
剪切 $\boldsymbol{A}=\begin{bmatrix}1 & 0\\ 0.5 & 1\end{bmatrix}$ $\begin{cases}\lambda_1=1\\ \lambda_2=1\end{cases}$ $\det(\boldsymbol{A})=1$	

13.4 对角化、谱分解

如果存在一个非奇异矩阵 $\boldsymbol{V}$ 和一个对角矩阵 $\boldsymbol{D}$，使得方阵 $\boldsymbol{A}$ 满足：

$$\boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V} = \boldsymbol{D} \tag{23}$$

则称 $\boldsymbol{A}$ **可对角化** (diagonalizable)。

只有可对角化的矩阵才能特征值分解：

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1} \tag{24}$$

其中，矩阵 $\boldsymbol{D}$ 就是特征值矩阵。

如果 $\boldsymbol{A}$ 可以对角化，矩阵 $\boldsymbol{A}$ 的平方可以写成：

$$\boldsymbol{A}^2 = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1} = \boldsymbol{V}\boldsymbol{D}^2\boldsymbol{V}^{-1} = \boldsymbol{V}\begin{bmatrix} (\lambda_1)^2 & & & \\ & (\lambda_2)^2 & & \\ & & \ddots & \\ & & & (\lambda_D)^2 \end{bmatrix}\boldsymbol{V}^{-1} \tag{25}$$

类似地，$\boldsymbol{A}$ 的 n 次幂可以写成：

$$\boldsymbol{A}^n = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1} = \boldsymbol{V}\boldsymbol{D}^n\boldsymbol{V}^{-1} = \boldsymbol{V}\begin{bmatrix} (\lambda_1)^n & & & \\ & (\lambda_2)^n & & \\ & & \ddots & \\ & & & (\lambda_D)^n \end{bmatrix}\boldsymbol{V}^{-1} \tag{26}$$

特别地，如果 $\boldsymbol{A}$ 为对称矩阵，$\boldsymbol{A}$ 的特征值分解可以写成：

$$\begin{aligned}\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} &= \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D\end{bmatrix}\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\end{bmatrix} \\ &= \lambda_1\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}} + \lambda_2\boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots\lambda_D\boldsymbol{v}_D\boldsymbol{v}_D^{\mathrm{T}} = \sum_{j=1}^{D}\lambda_j\boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{T}} \\ &= \lambda_1\boldsymbol{v}_1\otimes\boldsymbol{v}_1 + \lambda_2\boldsymbol{v}_2\otimes\boldsymbol{v}_2 + \cdots\lambda_D\boldsymbol{v}_D\otimes\boldsymbol{v}_D = \sum_{j=1}^{D}\lambda_j\boldsymbol{v}_j\otimes\boldsymbol{v}_j\end{aligned} \tag{27}$$

上式告诉我们为什么特征分解又叫**谱分解** (spectral decomposition)，因为特征值分解将矩阵拆解成一系列特征值和特征向量张量积乘积，就好比将白光分解成光谱中各色光一样。

以鸢尾花为例

本书第 10 章计算了鸢尾花数据矩阵 $\boldsymbol{X}$ 的格拉姆矩阵 $\boldsymbol{G}$，如图 7 所示。图 7 中 $\boldsymbol{G}$ 中元素没有保留任何小数位。

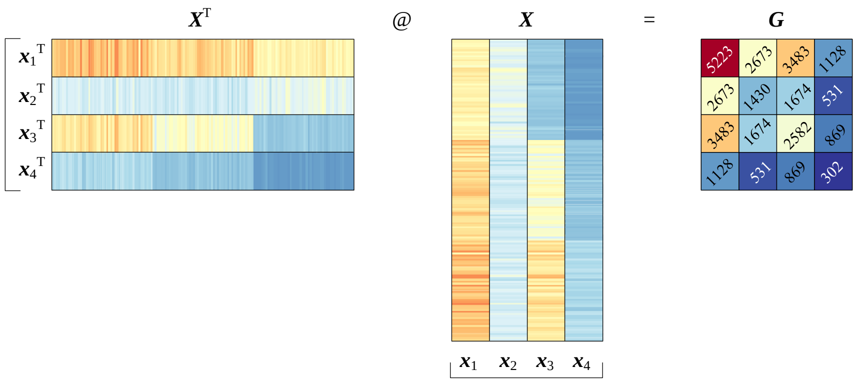


图 7. 矩阵 $\boldsymbol{X}$ 的格拉姆矩阵，图片来自本书第 10 章

格拉姆矩阵 $\boldsymbol{G}$ 为对称矩阵，对 $\boldsymbol{G}$ 特征值分解得到：

$$
\boldsymbol{G}=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}=\begin{bmatrix} 0.75 & 0.28 & 0.50 & 0.32 \\ 0.38 & 0.54 & -0.67 & -0.31 \\ 0.51 & -0.70 & -0.05 & -0.48 \\ 0.16 & -0.34 & -0.53 & 0.75 \end{bmatrix}\begin{bmatrix} 9208.3 & & & \\ & 315.4 & & \\ & & 11.9 & \\ & & & 3.5 \end{bmatrix}\begin{bmatrix} 0.75 & 0.28 & 0.50 & 0.32 \\ 0.38 & 0.54 & -0.67 & -0.31 \\ 0.51 & -0.70 & -0.05 & -0.48 \\ 0.16 & -0.34 & -0.53 & 0.75 \end{bmatrix}^{\mathrm{T}} \quad (28)
$$

上式中，$\boldsymbol{V}$ 仅保留两位小数位，特征值仅保留一位小数位。

(28) 也回答了本书第 10 章矩阵 $\boldsymbol{V}$ 从哪里来的问题。除了特征值分解，本书第 15、16 章介绍的奇异值分解也可以帮助我们获得矩阵 $\boldsymbol{V}$。

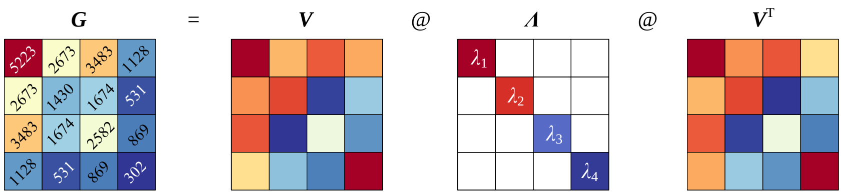


图 8. 矩阵 $\boldsymbol{X}$ 的格拉姆矩阵的特征值分解

利用谱分解方式展开 (28) 得到：

$$
\begin{aligned} \boldsymbol{G} &= \lambda_1 \boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \lambda_3 \boldsymbol{v}_3 \otimes \boldsymbol{v}_3 + \lambda_4 \boldsymbol{v}_4 \otimes \boldsymbol{v}_4 \\ &= 9208.3 \boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + 315.4 \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + 11.9 \boldsymbol{v}_3 \otimes \boldsymbol{v}_3 + 3.5 \boldsymbol{v}_4 \otimes \boldsymbol{v}_4 \end{aligned} \quad (29)
$$

由于 $\boldsymbol{V}$ 是规范正交基，因此在 $\mathbb{R}^4$ 空间中，$\boldsymbol{V}$ 的作用仅仅是旋转。而真正决定具体哪个 $\boldsymbol{v}_j$“更重要”的是特征值 λ_j 大小。观察上式容易发现，随着特征值 λ_j 不断减小，对应 $\boldsymbol{v}_j \otimes \boldsymbol{v}_j$ 的影响力也在衰减。图 9 中五幅热图采用相同色谱，$\lambda_1 \boldsymbol{v}_1 \otimes \boldsymbol{v}_1$ 影响力最大，剩下三个成分影响几乎可以忽略不计。请大家根据本书第 10 章代码，自行编写代码绘制本节热图。

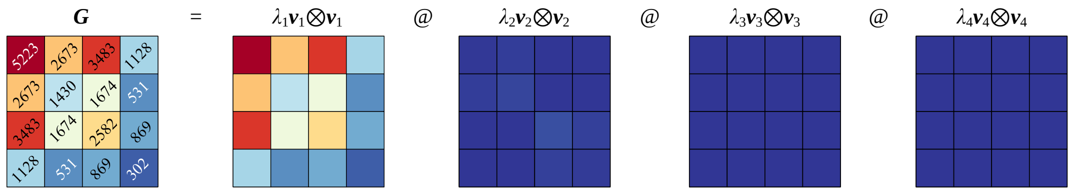


图 9. 矩阵 $\boldsymbol{X}$ 的格拉姆矩阵的谱分解

13.5 聊聊特征值

本节介绍特征值重要性质。

前文几次提到，给定矩阵 $\boldsymbol{A}$，其特征值 λ 和特征向量 $\boldsymbol{v}$ 关系为：

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v} \tag{30}$$

$\boldsymbol{A}$ 标量积 $k\boldsymbol{A}$ 对应的特征值为 λk，即，

$$(k\boldsymbol{A})\boldsymbol{v} = (k\lambda)\boldsymbol{v} \tag{31}$$

矩阵 $\boldsymbol{A}^2$ 的特征向量仍然为 $\boldsymbol{v}$，特征值为 λ^2：

$$\boldsymbol{A}^2\boldsymbol{v} = \boldsymbol{A}(\boldsymbol{A}\boldsymbol{v}) = \boldsymbol{A}(\lambda\boldsymbol{v}) = \lambda(\boldsymbol{A}\boldsymbol{v}) = \lambda^2\boldsymbol{v} \tag{32}$$

推广上式，n 为任意整数，$\boldsymbol{A}^n$ 的特征值为 λ^n：

$$\boldsymbol{A}^n\boldsymbol{v} = \lambda^n\boldsymbol{v} \tag{33}$$

(33) 也可以推广得到：

$$\boldsymbol{A}^n\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda}^n \tag{34}$$

如果逆矩阵 $\boldsymbol{A}^{-1}$ 存在，$\boldsymbol{A}^{-1}$ 的特征向量仍为 $\boldsymbol{v}$，特征值为 $1/\lambda$：

$$\boldsymbol{A}^{-1}\boldsymbol{v} = \frac{1}{\lambda}\boldsymbol{v} \tag{35}$$

前文提到，矩阵 $\boldsymbol{A}$ 的行列式值为其特征值乘积：

$$\det(\boldsymbol{A}) = \prod_{j=1}^{D} \lambda_j \tag{36}$$

如果任一特征值为 0，则 det($\boldsymbol{A}$) 为零，这种情况 $\boldsymbol{A}$ 起到降维作用。这正是本书第 4 章讲过的多维空间“平行体”和“正立方体”的体积关系。

特征向量 $\boldsymbol{v}_j$ 和特征值 λ_j 矩阵一一对应。对于多维方阵，多个不同的特征向量可以有相同的特征值，也就是在这些特征向量 $\boldsymbol{v}_j$ 方向上伸缩比例相同。

$\boldsymbol{A}$ 标量积 $k\boldsymbol{A}$ 的行列式值为：

$$\det\left(k\boldsymbol{A}\right)=k^{D}\prod_{j=1}^{D}\lambda_j \tag{37}$$

这相当于“平行体”和“正立方体”每个维度上边长都等比例缩放，缩放系数为 k。而体积的缩放比例为 k^D。

如果方阵 $\boldsymbol{A}$ 的形状为 $D \times D$，且 $\boldsymbol{A}$ 的秩 (rank) 为 r，则 $\boldsymbol{A}$ 有 $D - r$ 个特征值为 0。

矩阵 $\boldsymbol{A}$ 的迹等于其特征值之和：

$$\operatorname{tr}\left(\boldsymbol{A}\right)=\sum_{i=1}^{D}\lambda_i \tag{38}$$

我们将会在主成分分析 (Principal Component Analysis, PCA) 中用到 (38) 结论。

下图四副子图其实是一张图，它代表着特征值分解的几何视角——旋转 → 缩放 → 旋转。这一点对于理解特征值分解尤其重要。

此外，请大家特别注意对称矩阵的特征值分解，结果中 $\boldsymbol{V}$ 为正交矩阵，即规范正交基。

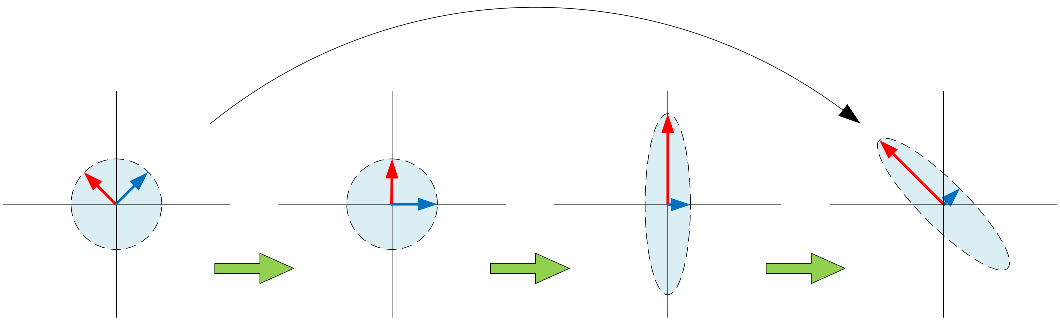


图 10. 总结本章重要内容的四副图