

# 13

## Eigen Decomposition

# 特征值分解

旋转 → 缩放 → 旋转



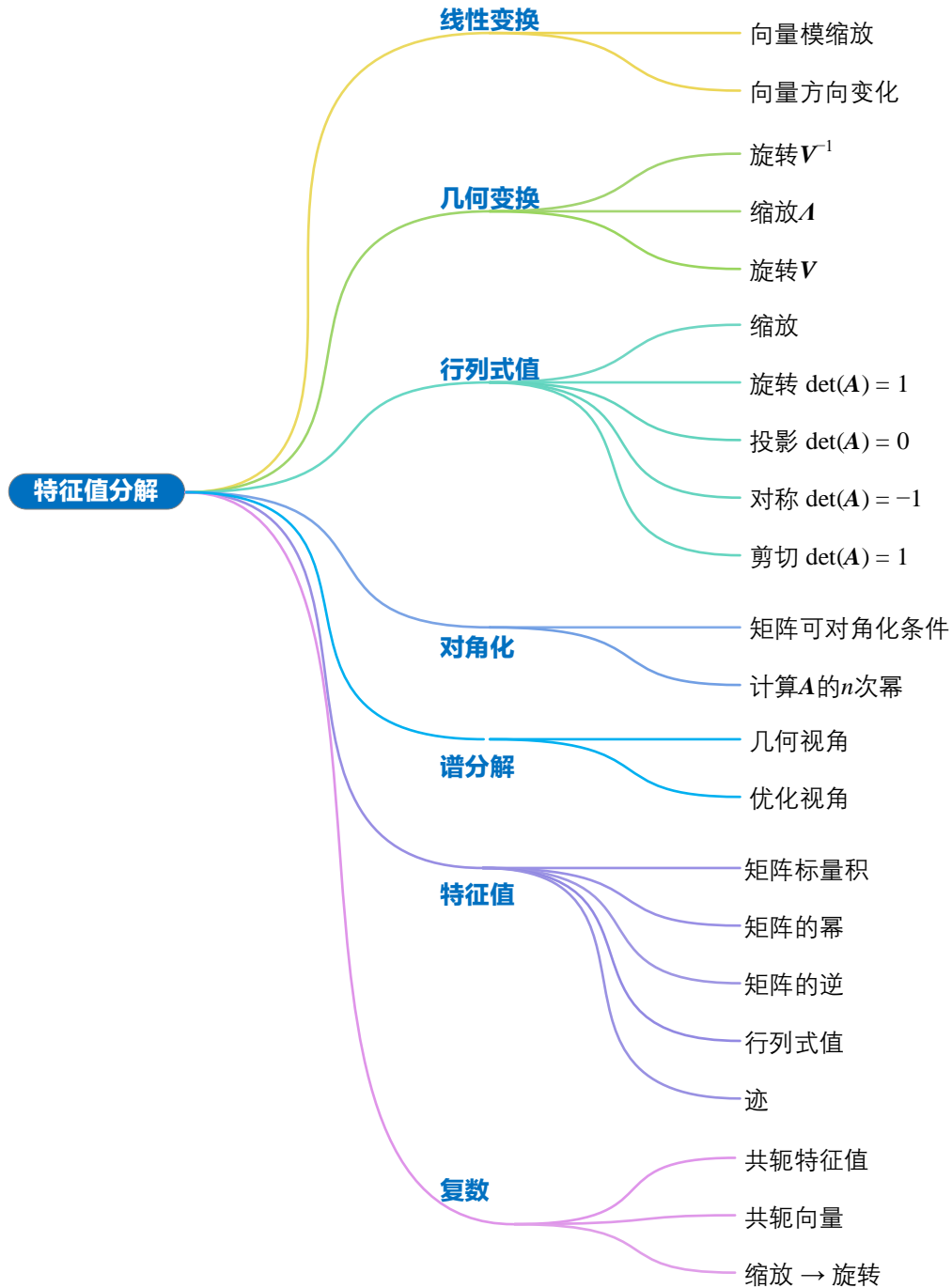
如果不能用数学表达，人类任何探索都不能被称之为真正的科学。

*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



## 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  $A$ ，具体如下：

$$A = \begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix} \quad (1)$$

矩阵  $A$  乘向量  $w_1$  得到一个新向量  $Aw_1$ ，比如：

$$w_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad Aw_1 = \begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.25 \\ -0.75 \end{bmatrix} \quad (2)$$

如图 1 所示，从几何角度，对比原向量  $w_1$ ，经过  $A$  的映射， $Aw_1$  的方向和模都发生了变化。也就是说， $A$  起到了缩放、旋转两方面作用。

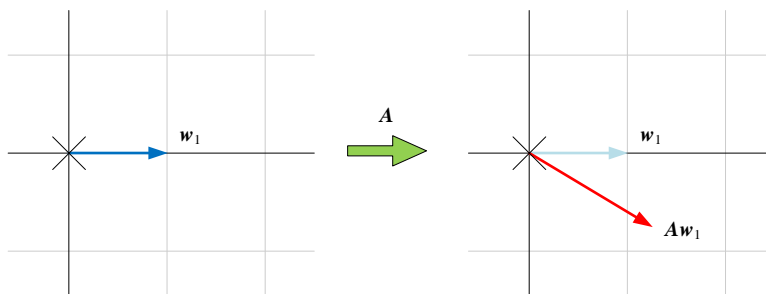


图 1. 我们发现相比原向量  $w_1$ ，新向量  $Aw_1$  的方向和模都发生变化

图 2 给出 81 个不同朝向向量  $w$ ，它们都是单位向量，即向量模均为 1。经过  $A$  的映射得到图 3 所示 81 个不同  $Aw$  结果。图 3 中，多数情况， $w$  (蓝色箭头) 到  $Aw$  (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)：

$$w_{11} = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}, \quad w_{31} = \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}, \quad w_{51} = \begin{bmatrix} -\sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix}, \quad w_{71} = \begin{bmatrix} \sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix} \quad (3)$$

矩阵  $A$  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。 $A$  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  $A$  特征向量，伸缩的比例就是特征值。

**▲ 注意**，准确来说，如果  $w$  是  $A$  的特征向量， $A$  和  $Aw$  方向平行，同向或反向。

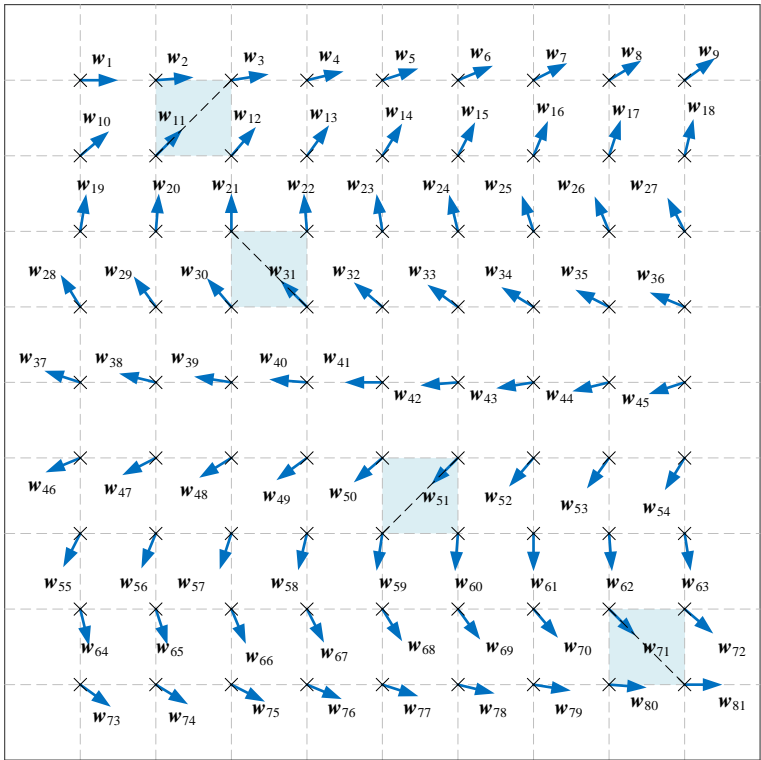


图 2. 81 个朝向不同方向的单位向量

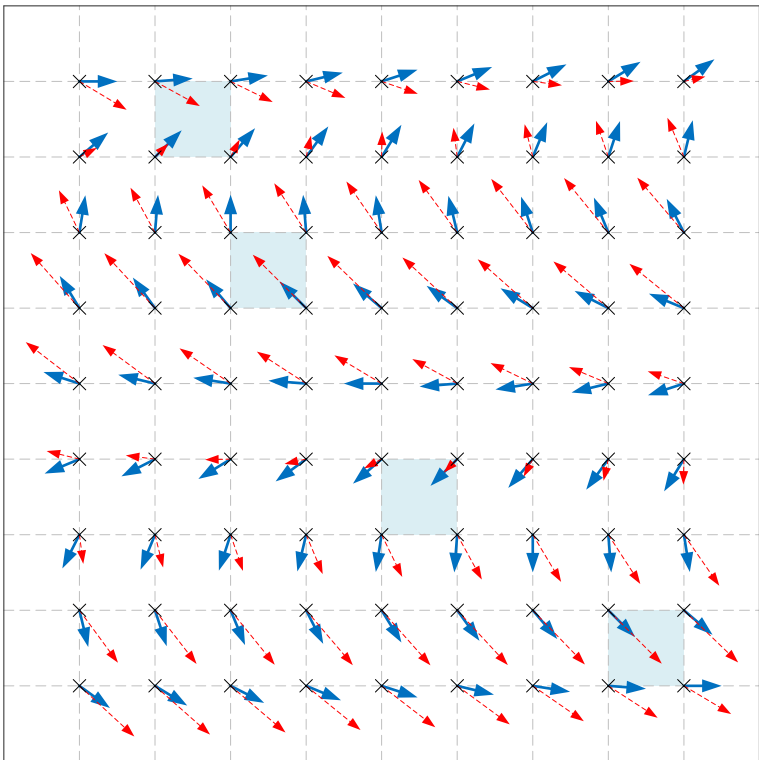


图 3. 矩阵  $A$  乘  $w$  得到的 81 个不同结果

## 单位圆

为了更好看清矩阵  $A$  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  $A$  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。

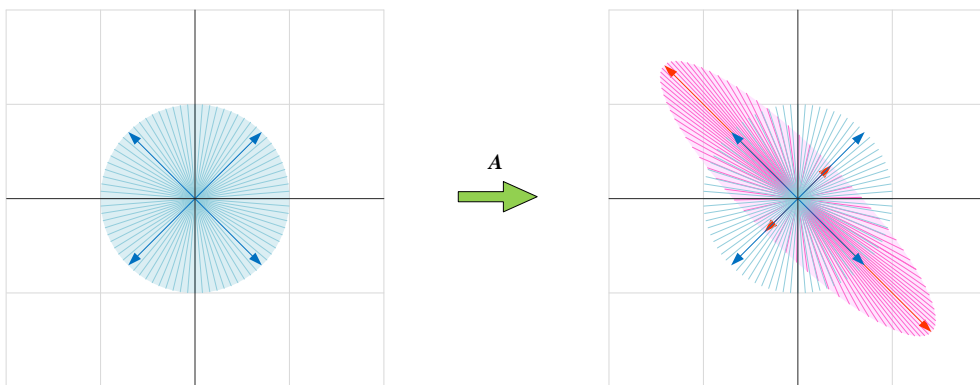


图 4. 矩阵  $A$  对一系列向量的映射结果



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  $A$  的特征值分解可以写成：

$$A = \overset{\text{Rotate}}{V} \overset{\text{Scale}}{A} \overset{\text{Rotate}}{V^{-1}} \quad (4)$$

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

$$Aw = \overset{\text{Rotate}}{V} \overset{\text{Scale}}{A} \overset{\text{Rotate}}{V^{-1}} w \quad (5)$$

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

## 举个 $2 \times 2$ 矩阵的例子

(4) 等式右乘  $V$  得到：

$$AV = VA \quad (6)$$

将  $V$  展开写成  $[v_1, v_2]$  并代入上式得到：

$$A[v_1 \ v_2] = [v_1 \ v_2] \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \quad (7)$$

展开 (7) 得到：

$$[Av_1 \ Av_2] = [\lambda_1 v_1 \ \lambda_2 v_2] \quad (8)$$

对于上一节给出的例子，将具体数值代入 (4)，得到：

$$\underbrace{\begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix}}_A = \underbrace{\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}}_V \underbrace{\begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix}}_A \underbrace{\begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}}_{V^{-1}} \quad (9)$$

下面，我们分别讨论  $v_1$  和  $v_2$  的几何特征。

## 第一特征向量

$v_1$  为：

$$v_1 = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \quad (10)$$

$A$  乘  $v_1$  得到  $Av_1$ ：

$$Av_1 = \begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/4 \\ \sqrt{2}/4 \end{bmatrix} = \frac{1}{2} \times \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \quad (11)$$

可以发现，相比  $v_1$ ， $Av_1$  方向没有发生变化， $A$  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  $v_1$ ，将 (4) 代入 (11)，将  $A$  拆解为“旋转→缩放→旋转”三步几何操作：

$$Av_1 = \overset{\text{Rotate}}{V} \overset{\text{Scale}}{A} \overset{\text{Rotate}}{V^{-1}} v_1 \quad (12)$$

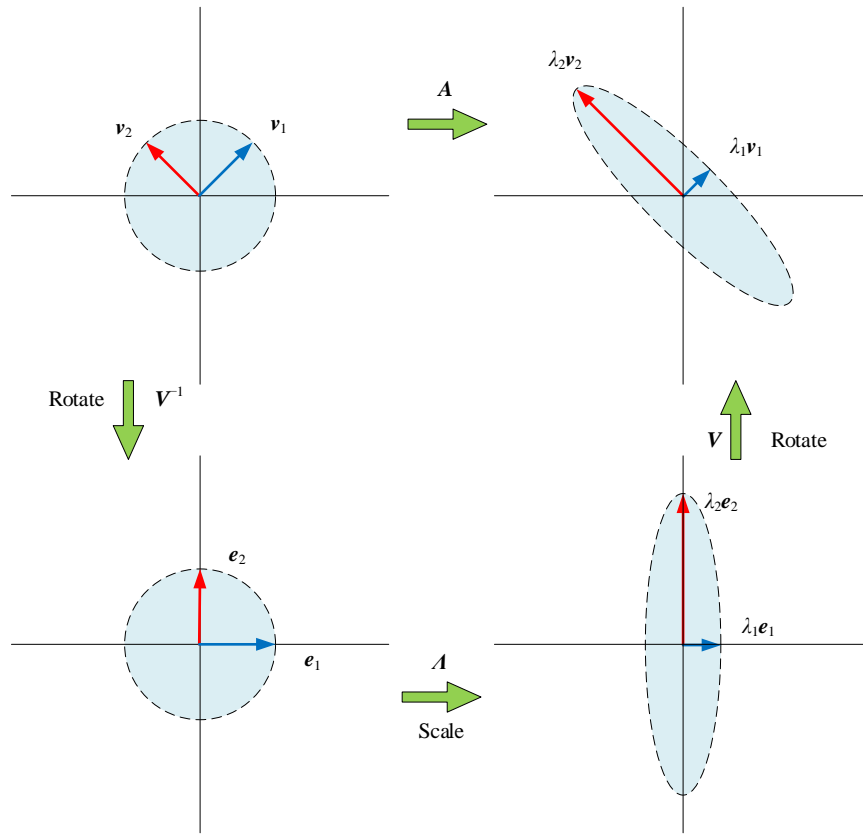


图 5. “旋转→缩放→旋转”操作

$V^{-1}v_1$  相对  $v_1$  顺时针旋转  $45^\circ$ :

$$V^{-1}v_1 = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = e_1 \quad (13)$$

然后再利用  $A$  完成缩放操作，得到  $AV^{-1}v_1$ :

$$AV^{-1}v_1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} = \begin{matrix} 0.5 \\ \lambda_1 \end{matrix} e_1 \quad (14)$$

最后利用  $V$  完成逆时针旋转  $45^\circ$ ，得到  $VAV^{-1}v_1$ :

$$\begin{aligned} \underbrace{VAV^{-1}}_A v_1 &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{matrix} 0.5 \\ \lambda_1 \end{matrix} e_1 \\ &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/4 \\ \sqrt{2}/4 \end{bmatrix} = \begin{matrix} 0.5 \\ \lambda_1 \end{matrix} \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \\ &= \lambda_1 v_1 \end{aligned} \quad (15)$$

## 第二特征向量

类似地，下面讨论  $A$  乘  $v_2$  对应的“旋转→缩放→旋转”操作。

$v_2$  为：

$$v_2 = \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \quad (16)$$

$A$  乘  $v_2$  得到  $Av_2$ ：

$$Av_2 = \begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix} \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \end{bmatrix} = 2 \times_{\lambda_2} \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \quad (17)$$

$V^{-1}v_2$  将  $v_2$  顺时针旋转  $45^\circ$ ：

$$\underset{\text{Rotate}}{V^{-1}} v_2 = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = e_2 \quad (18)$$

再缩放得到  $AV^{-1}v_2$ ：

$$\underset{\text{Scale}}{A} V^{-1} v_2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} = 2 \times_{\lambda_2} e_2 \quad (19)$$

最后旋转得到  $VAV^{-1}v_2$ ：

$$\begin{aligned} \underset{\text{Rotate}}{V} AV^{-1} v_2 &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} 2 \times_{\lambda_2} e_2 \\ &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \end{bmatrix} = 2 \times_{\lambda_2} \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \\ &= \lambda_2 v_2 \end{aligned} \quad (20)$$

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  $A$  的行列式值  $\det(A)$ ：



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

本书前文提到过  $2 \times 2$  矩阵行列式值相当于几何变换前后“面积缩放系数”。上式中  $\mathbf{A}$  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  $\mathbf{A}$  的行列式值加以验证：

$$\begin{aligned} \det(\mathbf{A}) &= \det(\mathbf{V}\mathbf{A}\mathbf{V}^{-1}) = \det(\mathbf{V})\det(\mathbf{A})\det(\mathbf{V}^{-1}) \\ &= \det(\mathbf{A})\det(\mathbf{V}\mathbf{V}^{-1}) = \det(\mathbf{A}) \\ &= \lambda_1\lambda_2 = \frac{1}{2} \times 2 = 1 \end{aligned} \quad (22)$$

上式说明，如果  $\mathbf{A}$  可以进行特征值分解，矩阵  $\mathbf{A}$  的行列式值等于  $\mathbf{A}$  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  $\mathbf{A}$  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  $|\mathbf{A}| = 1$ 。

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。

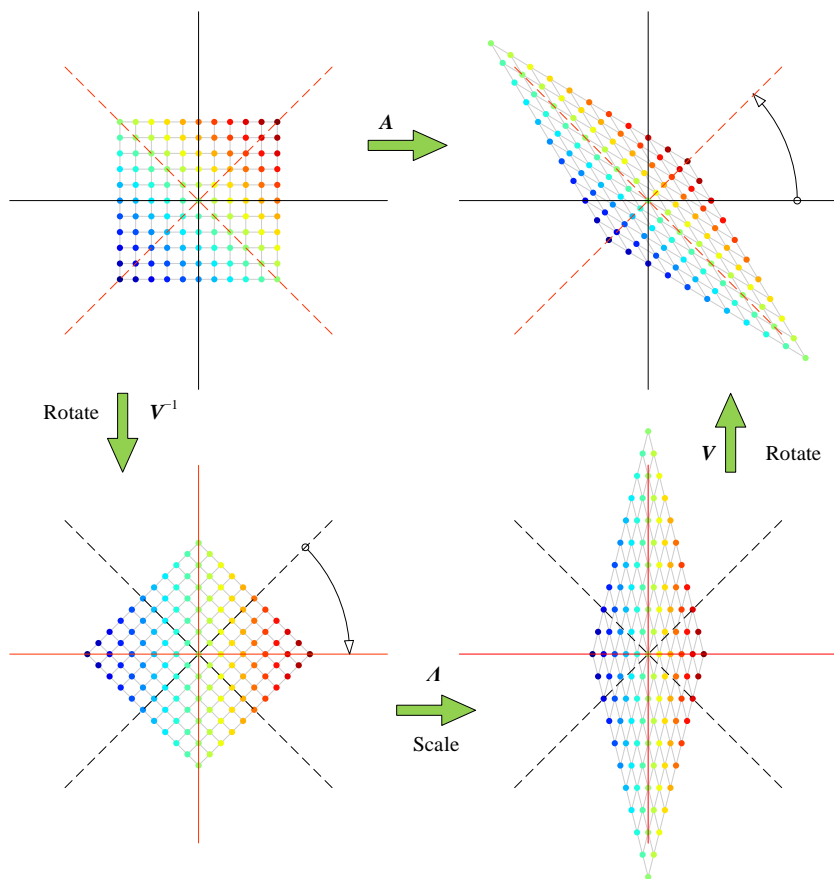


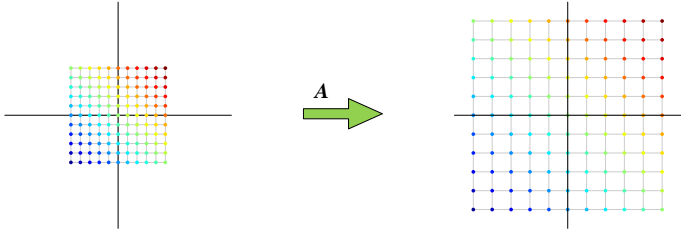
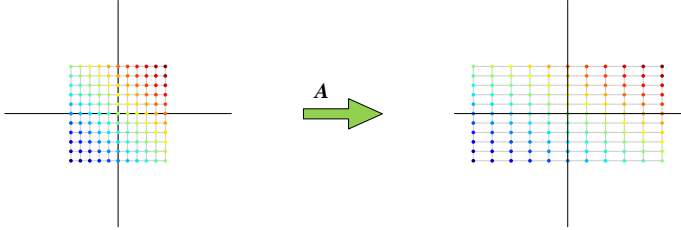
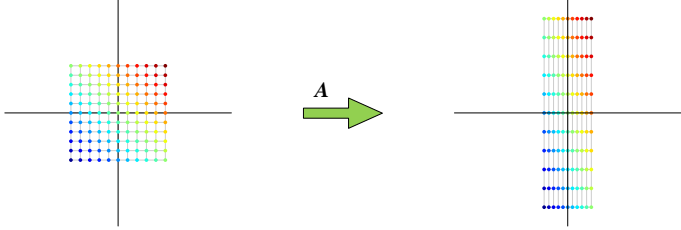
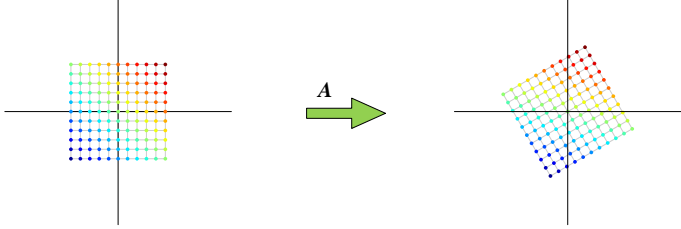
图 6. 正方形经过矩阵  $\mathbf{A}$  线性变换

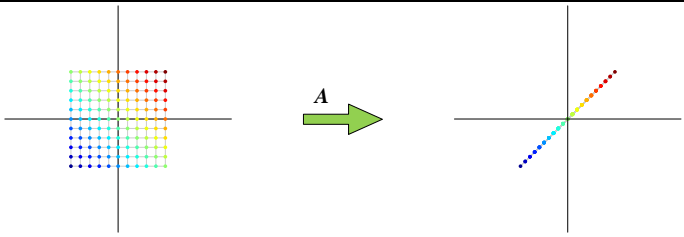
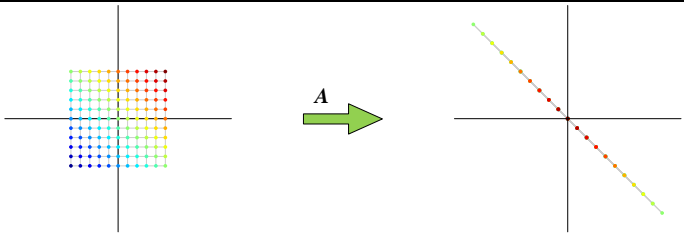
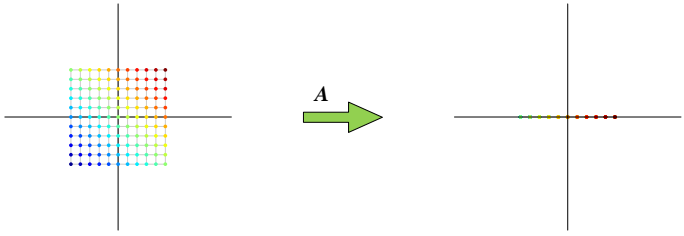
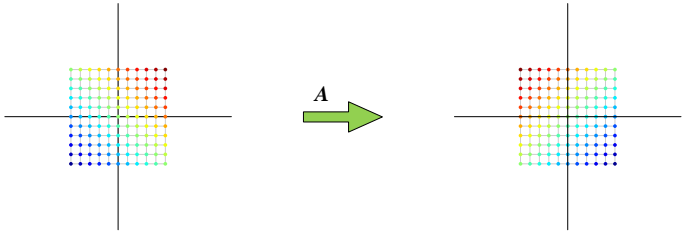
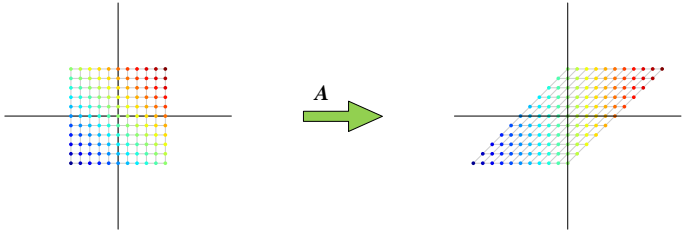
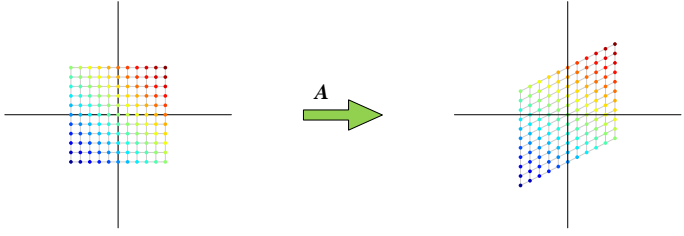
## 常见线性变换矩阵的特征值和行列式值

表 1 总结常见  $2 \times 2$  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。本章最后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  $2 \times 2$  矩阵对应的线性变换、特征值、行列式值

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

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

## 13.4 对角化、谱分解

### 可对角化

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

$$V^{-1}AV = D \quad (23)$$

则称  $A$  可对角化 (diagonalizable)。

只有可对角化的矩阵才能特征值分解：

$$A = VDV^{-1} \quad (24)$$

其中，矩阵  $D$  就是特征值矩阵。

如果  $A$  可以对角化，矩阵  $A$  的平方可以写成：

$$A^2 = VDV^{-1}VDV^{-1} = VD^2V^{-1} = V \begin{bmatrix} (\lambda_1)^2 & & & \\ & (\lambda_2)^2 & & \\ & & \ddots & \\ & & & (\lambda_D)^2 \end{bmatrix} V^{-1} \quad (25)$$

类似地， $A$  的  $n$  次幂可以写成：

$$A^n = VDV^{-1}VDV^{-1} = VD^nV^{-1} = V \begin{bmatrix} (\lambda_1)^n & & & \\ & (\lambda_2)^n & & \\ & & \ddots & \\ & & & (\lambda_D)^n \end{bmatrix} V^{-1} \quad (26)$$

### 谱分解

特别地，如果  $A$  为对称矩阵， $A$  的特征值分解可以写成：

$$\begin{aligned} A &= VAV^T = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_D \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_D^T \end{bmatrix} \\ &= \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T + \cdots + \lambda_D \mathbf{v}_D \mathbf{v}_D^T = \sum_{j=1}^D \lambda_j \mathbf{v}_j \mathbf{v}_j^T \\ &= \lambda_1 \mathbf{v}_1 \otimes \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 \otimes \mathbf{v}_2 + \cdots + \lambda_D \mathbf{v}_D \otimes \mathbf{v}_D = \sum_{j=1}^D \lambda_j \mathbf{v}_j \otimes \mathbf{v}_j \end{aligned} \quad (27)$$

其中,  $V$  为正交矩阵, 满足  $V^T V = V V^T = I$ 。

上式告诉我们为什么对称矩阵的特征分解又叫**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。

再进一步, 将  $V$  整理到 (27) 等式的左边:

$$V^T A V = A \quad (28)$$

同样将  $V$  写成其列向量并展开上式,

$$\begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_D^T \end{bmatrix} A \underbrace{\begin{bmatrix} v_1 & v_2 & \cdots & v_D \end{bmatrix}}_V = \underbrace{\begin{bmatrix} v_1^T A v_1 & v_1^T A v_2 & \cdots & v_1^T A v_D \\ v_2^T A v_1 & v_2^T A v_2 & \cdots & v_2^T A v_D \\ \vdots & \vdots & \ddots & \vdots \\ v_D^T A v_1 & v_D^T A v_2 & \cdots & v_D^T A v_D \end{bmatrix}}_{V^T A V} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}}_A \quad (29)$$

观察上式, 我们发现, 当  $i=j$  时, 方阵对角线元素满足:

$$v_j^T A v_j = \lambda_j \quad (30)$$

当  $i \neq j$  时, 方阵非对角线元素满足:

$$v_i^T A v_j = 0 \quad (31)$$

### 谱分解格拉姆矩阵

本书中见到的对称矩阵多数是格拉姆矩阵。对于数据矩阵  $X$ , 它的格拉姆矩阵  $G$  为  $G = X^T X$ 。  $G$  就是 (29) 中的矩阵  $A$ , 代入得到:

$$\begin{bmatrix} v_1^T X^T X v_1 & v_1^T X^T X v_2 & \cdots & v_1^T X^T X v_D \\ v_2^T X^T X v_1 & v_2^T X^T X v_2 & \cdots & v_2^T X^T X v_D \\ \vdots & \vdots & \ddots & \vdots \\ v_D^T X^T X v_1 & v_D^T X^T X v_2 & \cdots & v_D^T X^T X v_D \end{bmatrix} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}}_A \quad (32)$$

令  $y_j = X v_j$ 。如图 7 所示, 由于  $y_j$  是单位矩阵, 矩阵乘积  $X v_j$  相当于数据矩阵  $X$  向  $\text{span}(v_j)$  投影结果为  $y_j$ 。

(32) 可以写成:

$$\begin{bmatrix} y_1^T y_1 & y_1^T y_2 & \cdots & y_1^T y_D \\ y_2^T y_1 & y_2^T y_2 & \cdots & y_2^T y_D \\ \vdots & \vdots & \ddots & \vdots \\ y_D^T y_1 & y_D^T y_2 & \cdots & y_D^T y_D \end{bmatrix} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}}_A \quad (33)$$

观察上式，我们发现当  $i \neq j$  时， $y_i$  和  $y_j$  正交。我们在本书第 10 章介绍过这一结论，上述推导让我们“知其所以然”。

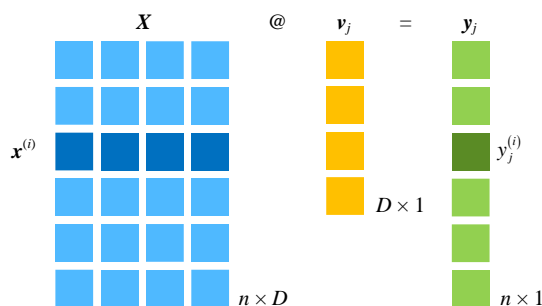


图 7. 数据矩阵  $X$  向  $\text{span}(v_j)$  投影结果为  $y_j$

注意，(32) 上式中矩阵每个元素显然都是标量。本书之前一直强调，看到矩阵乘积结果为标量时，一定要想一想矩阵乘积能否写成  $L^2$  范数。

(33) 对角线元素显然可以写成  $L^2$  范数：

$$\|y_j\|_2^2 = \|Xv_j\|_2^2 = \lambda_j \quad (34)$$

## 几何视角

该怎么理解 (34)？我们还是要拿出看家本领——几何视角。如图 8 所示，用散点  $\bullet$  代表数据矩阵  $X$ ，散点  $\bullet$  向  $\text{span}(v_j)$  投影结果为  $y_j$ ，即图中  $\times$ 。 $y_j$  中的每个值就是  $\times$  到原点的距离。

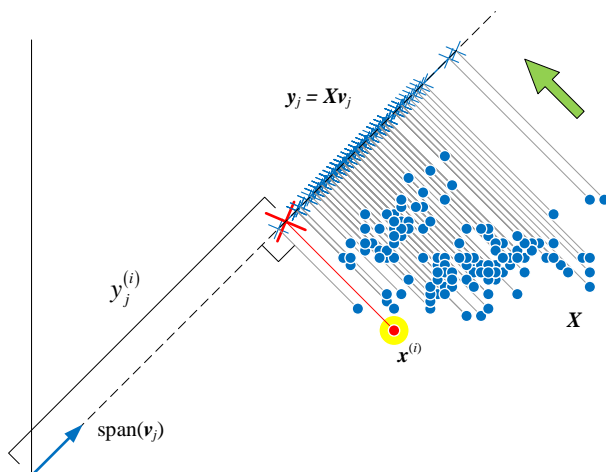


图 8. 数据矩阵  $X$  向  $\text{span}(v_j)$  投影结果为  $y_j$ ，几何视角

矩阵  $X$  的第  $i$  行行向量为  $x^{(i)}$ ，即图 8 中红点  $\bullet$ 。 $x^{(i)}$  向  $v_j$  投影结果  $y_j^{(i)}$  就是  $x^{(i)}$  在  $\text{span}(v_j)$  的坐标：

$$\mathbf{y}_j^{(i)} = \mathbf{x}^{(i)} \mathbf{v}_j \quad (35)$$

有了这个视角，我们知道 (34) 中  $\|\mathbf{y}_j\|_2^2$  代表  $\mathbf{y}_j^{(i)}$  到原点距离 (有正负) 的平方和，即：

$$\|\mathbf{y}_j\|_2^2 = \left(y_j^{(1)}\right)^2 + \left(y_j^{(2)}\right)^2 + \cdots + \left(y_j^{(n)}\right)^2 = \lambda_j \quad (36)$$

而距离的平方和恰好等于特征值  $\lambda_j$ 。

若 (33) 中特征值  $\lambda_j$  按大小排列，即  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D$ ，这说明特征向量  $\mathbf{v}_j$  也有主次之分。数据矩阵  $\mathbf{X}$  朝不同特征向量  $\mathbf{v}$  投影，得到的  $\|\mathbf{y}\|_2^2 = \|\mathbf{X}\mathbf{v}\|_2^2$  有大、有小。有大小之分，就意味有优化存在。

我们先给结论，在  $\mathbb{R}^D$  有无数个  $\mathbf{v}$  中， $\mathbf{X}$  朝特征向量  $\mathbf{v}_1$  投影对应的  $\|\mathbf{y}\|_2^2 = \|\mathbf{X}\mathbf{v}\|_2^2$  最大，这个最大值为  $\lambda_1$ 。本书第 18 章将提供优化视角告诉我们“为什么”。

### 以鸢尾花为例

本书第 10 章计算了鸢尾花数据矩阵  $\mathbf{X}$  的格拉姆矩阵  $\mathbf{G}$ ，如图 9 所示。图 9 中  $\mathbf{G}$  中元素没有保留任何小数位。

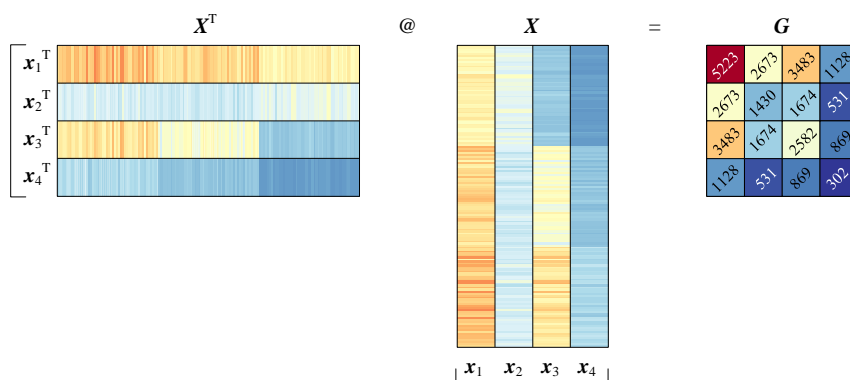


图 9. 矩阵  $\mathbf{X}$  的格拉姆矩阵，图片来自本书第 10 章

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

$$\mathbf{G} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \begin{bmatrix} 0.75 & 0.28 & 0.50 & 0.32 \\ 0.38 & 0.54 & -0.67 & -0.31 \\ 0.51 & -0.70 & -0.05 & -0.48 \\ 0.16 & -0.34 & -0.53 & 0.75 \end{bmatrix} \begin{bmatrix} 9208.3 & & & \\ & 315.4 & & \\ & & 11.9 & \\ & & & 3.5 \end{bmatrix} \begin{bmatrix} 0.75 & 0.28 & 0.50 & 0.32 \\ 0.38 & 0.54 & -0.67 & -0.31 \\ 0.51 & -0.70 & -0.05 & -0.48 \\ 0.16 & -0.34 & -0.53 & 0.75 \end{bmatrix}^T \quad (37)$$

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



(37) 也回答了本书第 10 章矩阵  $V$  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  $V$ 。

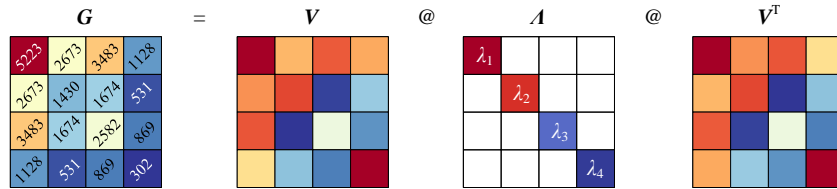


图 10. 矩阵  $X$  的格拉姆矩阵的特征值分解

利用谱分解方式展开 (37) 得到：

$$\begin{aligned} G &= \lambda_1 \mathbf{v}_1 \otimes \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 \otimes \mathbf{v}_2 + \lambda_3 \mathbf{v}_3 \otimes \mathbf{v}_3 + \lambda_4 \mathbf{v}_4 \otimes \mathbf{v}_4 \\ &= 9208.3 \mathbf{v}_1 \otimes \mathbf{v}_1 + 315.4 \mathbf{v}_2 \otimes \mathbf{v}_2 + 11.9 \mathbf{v}_3 \otimes \mathbf{v}_3 + 3.5 \mathbf{v}_4 \otimes \mathbf{v}_4 \end{aligned} \quad (38)$$

由于  $V$  是规范正交基，因此在  $\mathbb{R}^4$  空间中， $V$  的作用仅仅是旋转。而真正决定具体哪个  $\mathbf{v}_j$  “更重要”的是特征值  $\lambda_j$  大小。观察上式容易发现，随着特征值  $\lambda_j$  不断减小，对应  $\mathbf{v}_j \otimes \mathbf{v}_j$  的影响力也在衰减。图 11 中五幅热图采用相同色谱， $\lambda_1 \mathbf{v}_1 \otimes \mathbf{v}_1$  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。

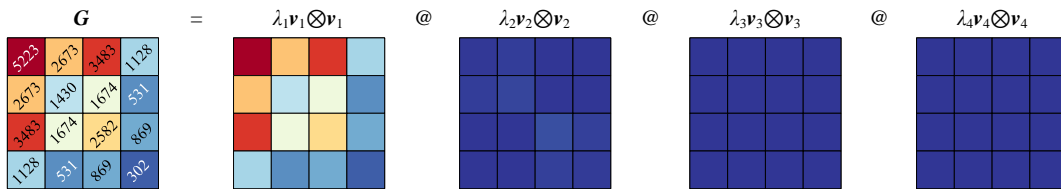


图 11. 矩阵  $X$  的格拉姆矩阵的谱分解

## 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  $A$ ，其特征值  $\lambda$  和特征向量  $\mathbf{v}$  关系为：

$$A\mathbf{v} = \lambda\mathbf{v} \quad (39)$$

$A$  标量积  $kA$  对应的特征值为  $\lambda k$ ，即，

$$(kA)\mathbf{v} = (k\lambda)\mathbf{v} \quad (40)$$



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

$$A^2\mathbf{v} = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda(A\mathbf{v}) = \lambda^2\mathbf{v} \quad (41)$$

推广上式， $n$  为任意整数， $A^n$  的特征值为  $\lambda^n$ ：

$$A^n\mathbf{v} = \lambda^n\mathbf{v} \quad (42)$$

(42) 也可以推广得到：

$$A^n\mathbf{V} = \mathbf{V}A^n \quad (43)$$

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

$$A^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v} \quad (44)$$

前文提到，矩阵  $A$  的行列式值为其特征值乘积：

$$\det(A) = \prod_{j=1}^D \lambda_j \quad (45)$$

如果任一特征值为 0，则  $\det(A)$  为零，这种情况  $A$  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  $\mathbf{v}_j$  和特征值  $\lambda_j$  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  $\mathbf{v}_j$  方向上伸缩比例相同。

$A$  标量积  $kA$  的行列式值为：

$$\det(kA) = k^D \prod_{j=1}^D \lambda_j \quad (46)$$

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

矩阵  $A$  的迹等于其特征值之和：

$$\text{tr}(A) = \sum_{i=1}^D \lambda_i \quad (47)$$

我们将会在主成分分析 (Principal Component Analysis, PCA) 中用到 (47) 结论。

## 13.6 特征值分解中的复数现象

本本章前文在对实数矩阵进行特征值分解时，我们偶尔发现特征值、特征向量存在虚数。这一节讨论这个现象。

### 举个例子

给定如下  $2 \times 2$  实数矩阵  $A$ :

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad (48)$$

对  $A$  进行特征值分解，得到两个特征值分别为：

$$\lambda_1 = 1 + i, \quad \lambda_2 = 1 - i \quad (49)$$

### 共轭特征值、共轭特征向量

这对共轭特征值出现的原因是，方阵  $A$  特征方程有一对复数解：

$$|A - \lambda I| = 0 \quad (50)$$

求解出的非实数的特征值会以共轭复数形式成对出现，因此它们也常被称作**共轭特征值** (conjugate eigenvalues)。所谓**共轭复数** (complex conjugate)，是指两个实部相等，虚部互为相反数的复数。

(48) 中  $A$  的特征值  $\lambda_1$  和  $\lambda_2$  对应的特征向量分别是：

$$\mathbf{v}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -i \\ 1 \end{bmatrix} \quad (51)$$

这样的特征向量，常被称作**共轭特征向量** (conjugate eigenvector)。展开来说，本书前文讲述的向量矩阵等概念都是建立在实数  $\mathbb{R}^n$  基础之上，我们可以把同样的数学工具推广到复数空间  $\mathbb{C}^n$  上。

$\mathbb{C}^n$  中的任意复向量  $\mathbf{x}$  的共轭向量  $\bar{\mathbf{x}}$ ，也是  $\mathbb{C}^n$  中的向量。 $\bar{\mathbf{x}}$  中每个元素是  $\mathbf{x}$  对应元素的共轭复数。比如，给定复数向量  $\mathbf{x}$  和对应的共轭向量  $\bar{\mathbf{x}}$  如下：

$$\mathbf{x} = \begin{bmatrix} 1+i \\ 3-2i \end{bmatrix}, \quad \bar{\mathbf{x}} = \begin{bmatrix} 1-i \\ 3+2i \end{bmatrix} \quad (52)$$

### 一个特殊的 $2 \times 2$ 矩阵

给定矩阵  $A$  如下：

$$A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \quad (53)$$

其中， $a$  和  $b$  均为实数，且不同时等于 0。

容易求得  $A$  的复数特征值为一对共轭复数：

$$\lambda = a \pm bi \quad (54)$$

两者的关系如图 12 所示。图 12 横轴为实部，纵轴为虚部。

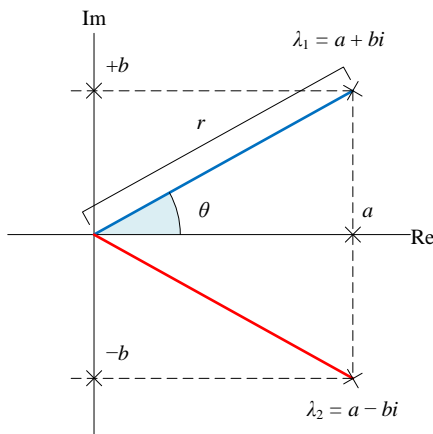


图 12. 一对共轭特征值

两个共轭特征值的模相等，令  $r$  复数特征值的模，容易发现， $r$  是矩阵  $A$  行列式值的平方根：

$$r = |\lambda| = \sqrt{a^2 + b^2} = \sqrt{|A|} \quad (55)$$

因此， $A$  可以写成：

$$A = \sqrt{a^2 + b^2} \begin{bmatrix} \frac{a}{\sqrt{a^2 + b^2}} & \frac{-b}{\sqrt{a^2 + b^2}} \\ \frac{b}{\sqrt{a^2 + b^2}} & \frac{a}{\sqrt{a^2 + b^2}} \end{bmatrix} = r \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = \underbrace{\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}}_R \underbrace{\begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}}_S \quad (56)$$

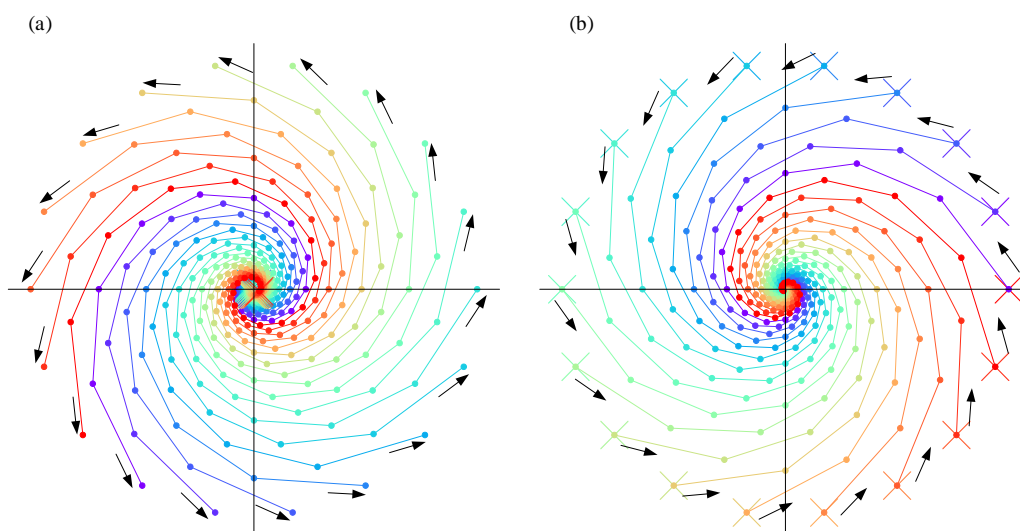
图 12 所示复平面上， $\theta$  为  $(0, 0)$  到  $(a, b)$  线段和水平轴正方向夹角， $\theta$  也称作复数  $\lambda_1 = a + bi$  的辐角。

## 几何视角

有了上述分析，矩阵  $A$  的几何变换就变得很清楚， $A$  是缩放 ( $S$ ) 和旋转 ( $R$ ) 的复合。给平面上某个位置的  $x_0$ ，用矩阵  $A$  不断作用在  $x_0$  上：

$$x_n = A^n x_0 \quad (57)$$

如图 13 (a) 所示，当缩放系数  $r = 1.2 > 1$ ，我们可以看到，随着  $n$  增大，向量  $x_n$  不断旋转向外。如图 13 (b) 所示，当缩放系数  $r = 0.8 < 1$ ，随着  $n$  增大，向量  $x_n$  不断旋转向内。注意，图 12 中平面是复平面，横轴是实数轴，纵轴是虚数轴。而图 13 则是实数  $x_1 x_2$  平面。

图 13. 在矩阵  $A$  几何变换重复下，向量的  $x$  位置变化

Bk4\_Ch13\_03.py 绘制图 13。



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。

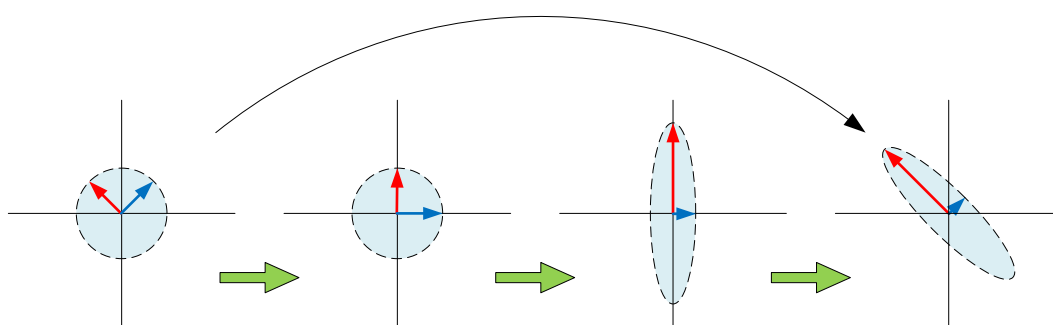


图 14. 总结本章重要内容的四副图

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

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：[jiang.visualize.ml@gmail.com](mailto:jiang.visualize.ml@gmail.com)

本章最后以我们在对实数矩阵分解中遇到的复数现象为例，介绍了共轭特征值和共轭特征向量。注意，复数矩阵自有一套体系，比如复数矩阵的转置叫做**埃尔米特转置** (Hermitian transpose)，记号一般用上标  $H$ 。复数矩阵相关内容不在本书范围内，感兴趣的读者可以自行学习。