13 Eigen Decomposition 特征值分解

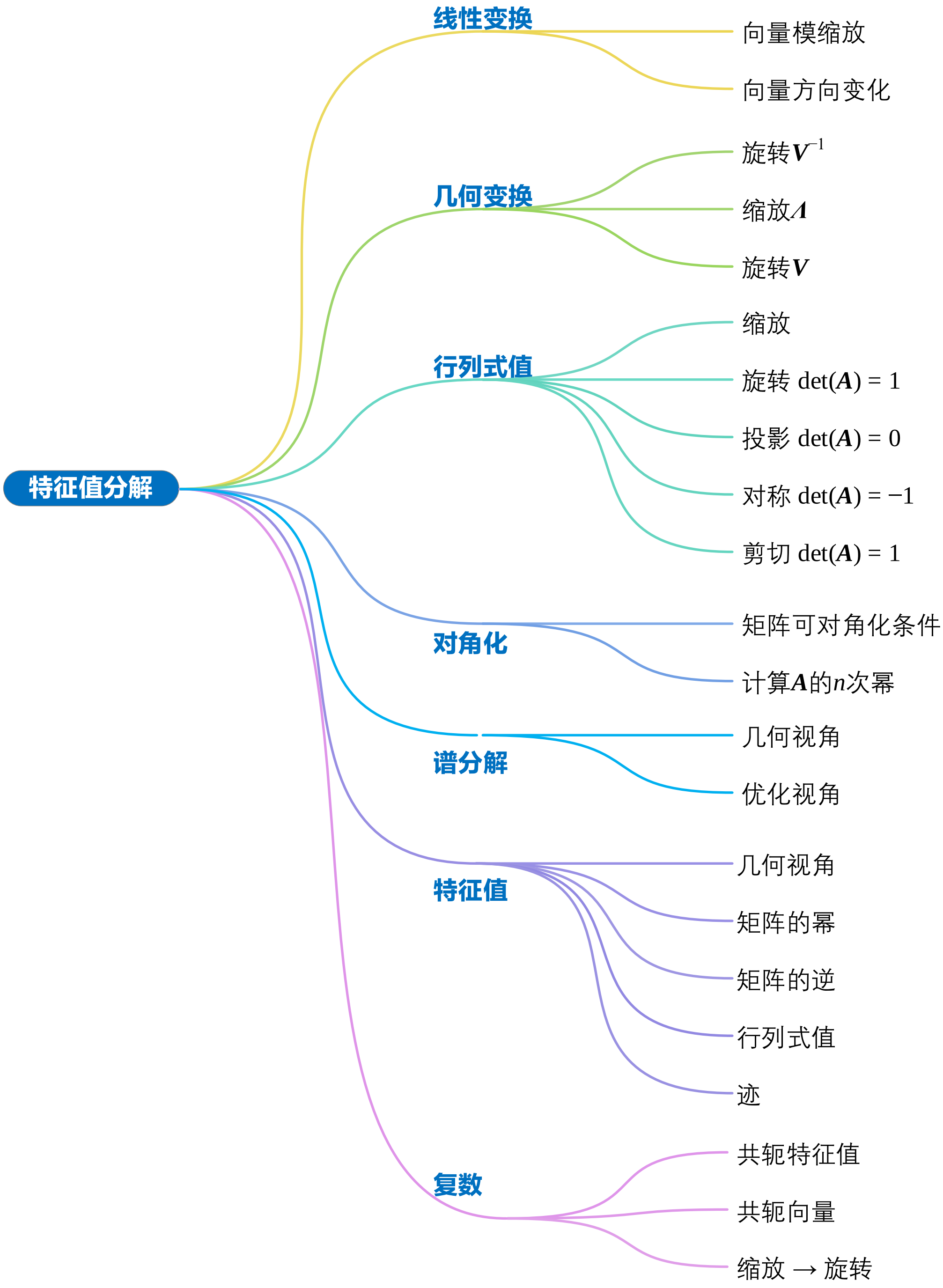
旋转 → 缩放 → 旋转

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V^{-1}
缩放Λ
旋转V
行列式值
缩放
旋转 $\det(A) = 1$
投影 $\det(A) = 0$
对称 $\det(A) = -1$
剪切 $\det(A) = 1$
对角化
矩阵可对角化条件
计算A的n次幂
谱分解
几何视角
优化视角
特征值
几何视角
矩阵的幂
矩阵的逆
行列式值
迹
复数
共轭特征值
共轭向量
缩放 → 旋转

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。

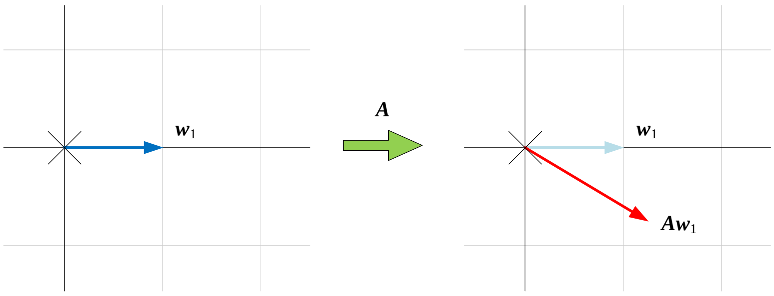


图 1. 我们发现相比原向量 $\boldsymbol{w}_1$，新向量 $\boldsymbol{A}\boldsymbol{w}_1$ 的方向和模都发生变化

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，$\boldsymbol{w}$ 和 $\boldsymbol{A}\boldsymbol{w}$ 方向平行，同向或反向。

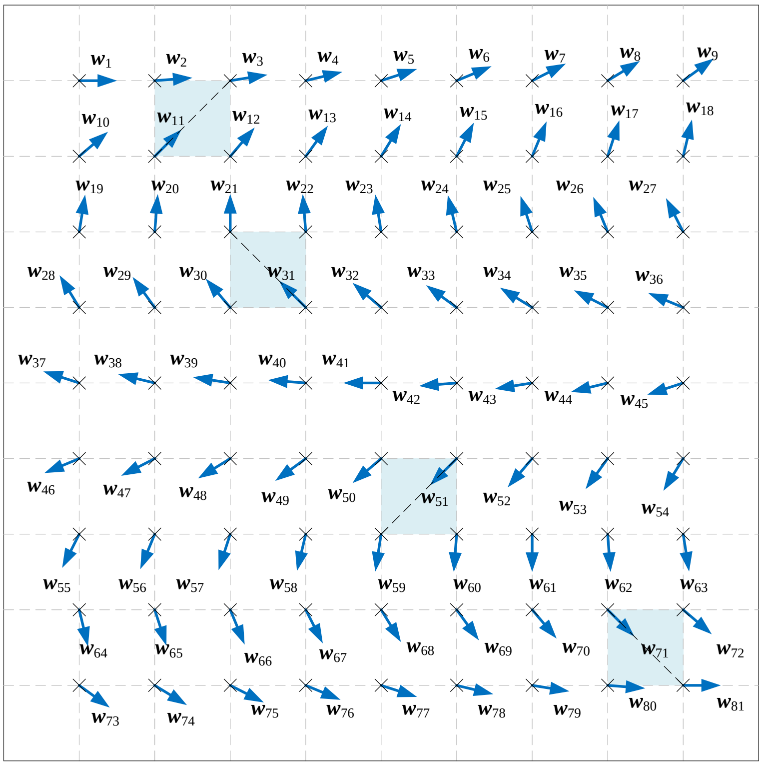


图 2. 81 个朝向不同方向的单位向量

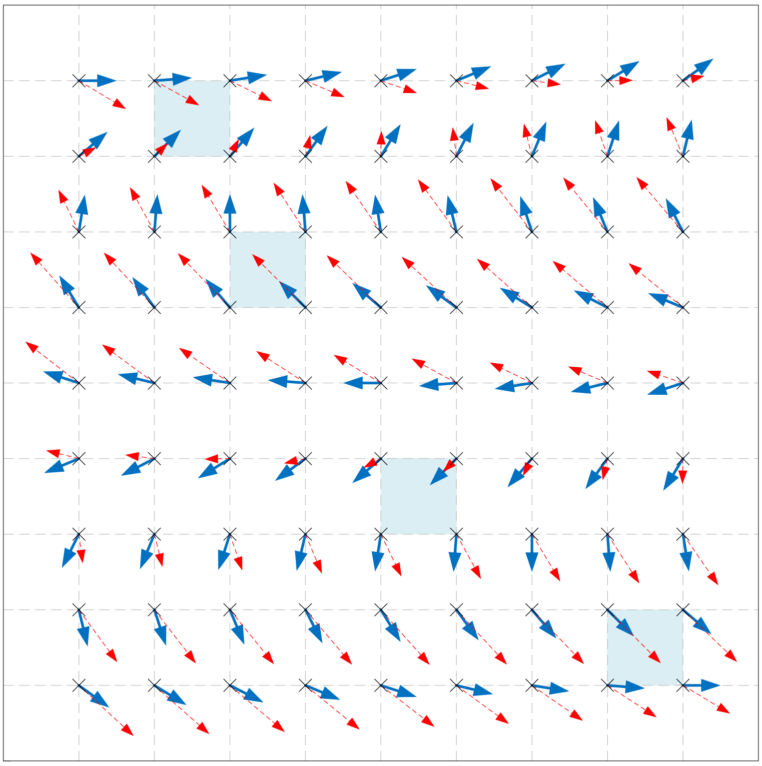


图 3. 矩阵 $\boldsymbol{A}$ 乘 $\boldsymbol{w}$ 得到的 81 个不同结果

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。

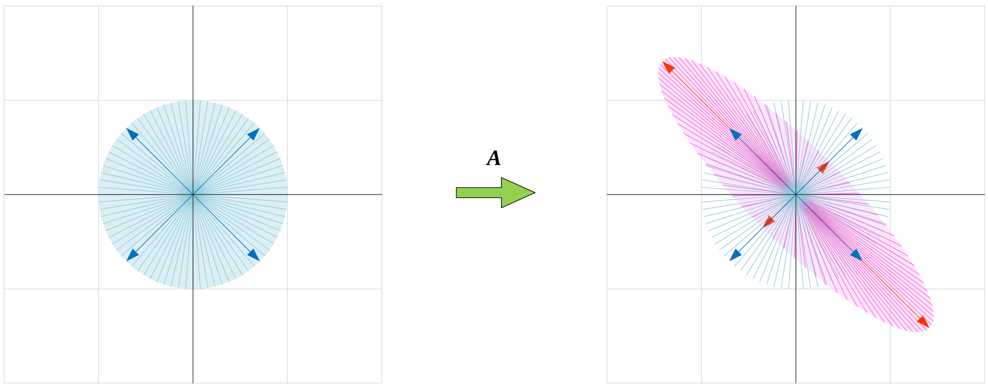


图 4. 矩阵 $\boldsymbol{A}$ 对一系列向量的映射结果

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 (V^{-1}) → 缩放 (Λ) →旋转 (V)。此外，准确来说，只有 V 是正交矩阵且 $\det(V) = 1$，V 才是旋转矩阵 (rotation matrix)，对应的几何操作才是纯粹的旋转。

举个 2 × 2 矩阵的例子

(4) 等式右乘 V 得到：

$$AV = V\Lambda \tag{6}$$

将 V 展开写成 $[v_1, v_2]$ 并代入上式得到：

$$A\begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 \end{bmatrix}\begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \tag{7}$$

展开 (7) 得到：

$$\begin{bmatrix} Av_1 & Av_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 \end{bmatrix} \tag{8}$$

对于上一节给出的例子，将具体数值代入 (4)，得到：

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

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：

$$v_1 = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \tag{10}$$

A 乘 v_1 得到 Av_1：

$$Av_1 = \begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix}\begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/4 \\ \sqrt{2}/4 \end{bmatrix} = \underbrace{\frac{1}{2}}_{\lambda_1} \times \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \tag{11}$$

可以发现，相比 v_1，Av_1 方向没有发生变化，A 仅仅产生缩放作用，缩放比例为 $\lambda_1 = 1/2$。

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：

$$Av_1 = \overset{\text{Rotate}}{V}\ \overset{\text{Scale}}{\Lambda}\ \overset{\text{Rotate}}{V^{-1}}\ v_1 \tag{12}$$

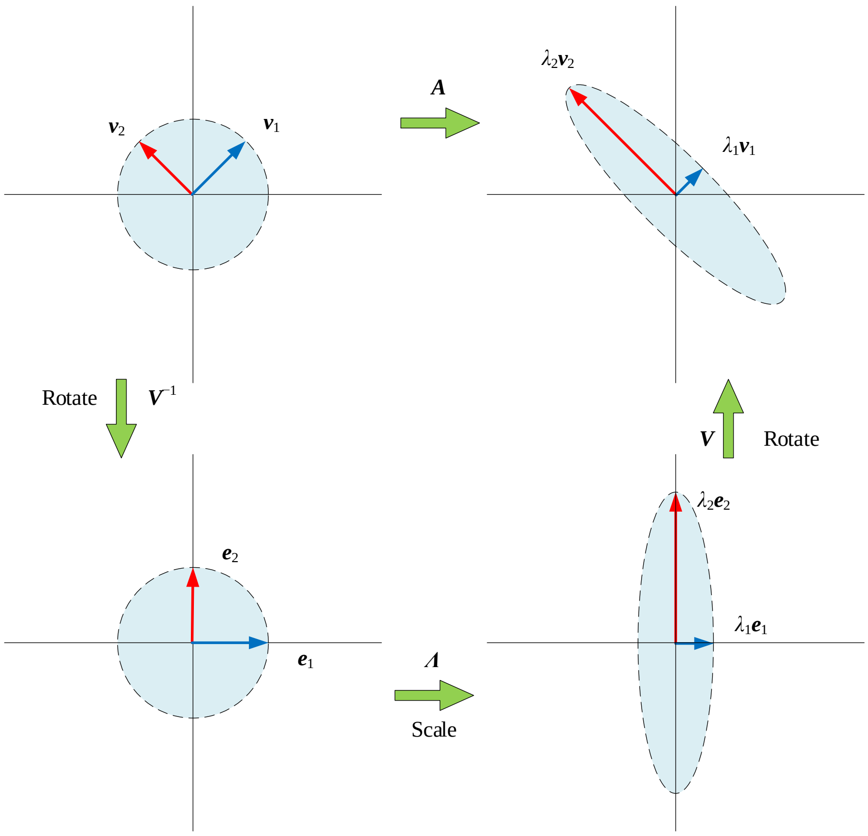


图 5. "旋转→缩放→旋转"操作

$\boldsymbol{V}^{-1}\boldsymbol{v}_1$ 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 $\boldsymbol{\Lambda V}^{-1}\boldsymbol{v}_1$：

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

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

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

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 $\boldsymbol{A}\boldsymbol{v}_2$：

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

相比 $\boldsymbol{v}_2$，$\boldsymbol{A}\boldsymbol{v}_2$ 方向没有发生变化，$\boldsymbol{A}$ 产生缩放作用，缩放比例为 $\lambda_2=2$。

$\boldsymbol{V}^{-1}\boldsymbol{v}_2$ 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 $\det(\boldsymbol{A})$：

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

本书第 4 章提到过，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。

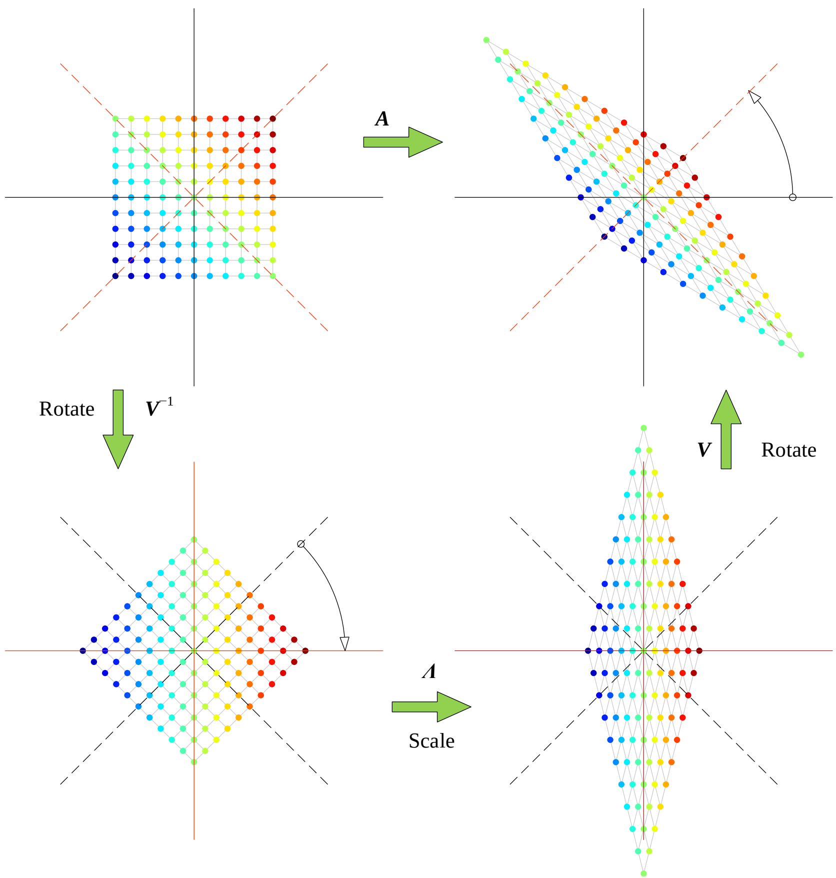


图 6. 正方形经过矩阵 $\boldsymbol{A}$ 线性变换

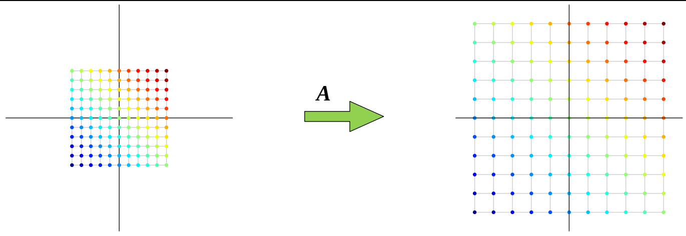
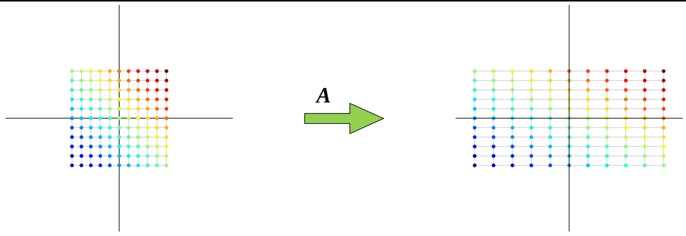
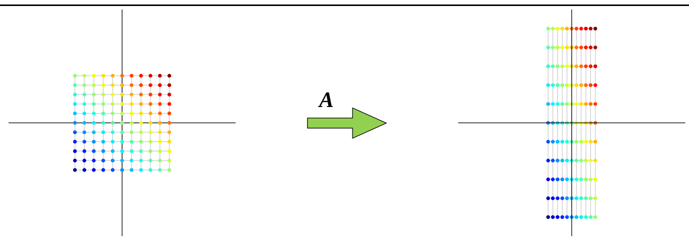
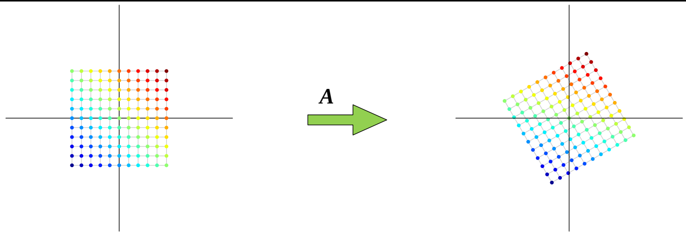
线性变换、特征值、行列式值

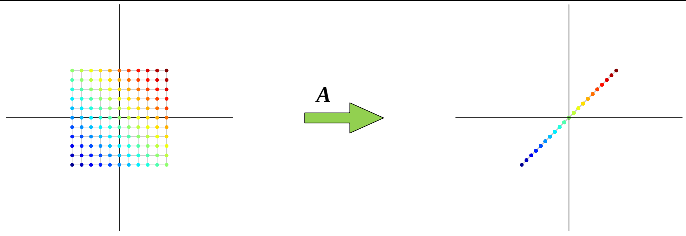
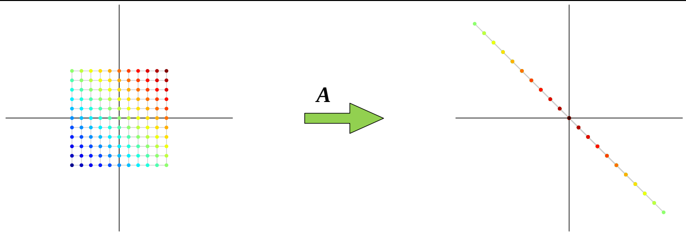
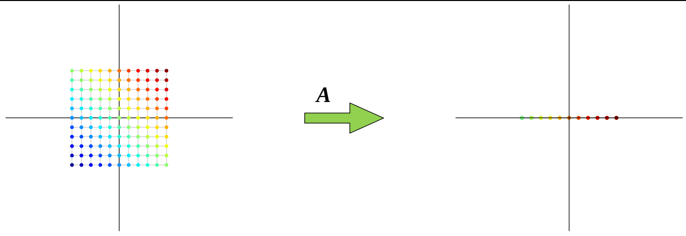
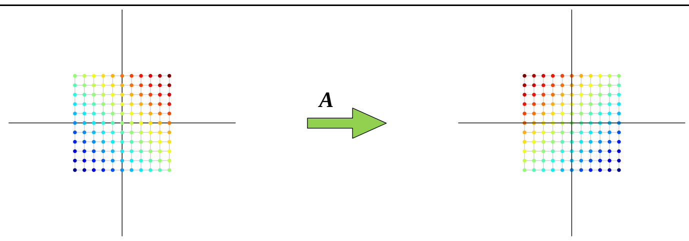
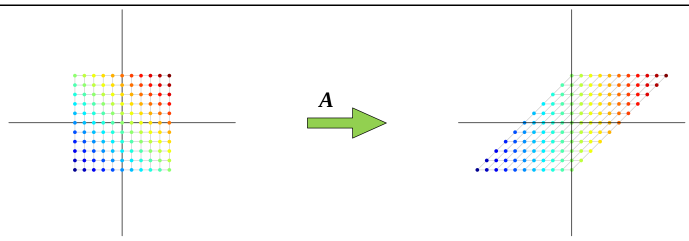
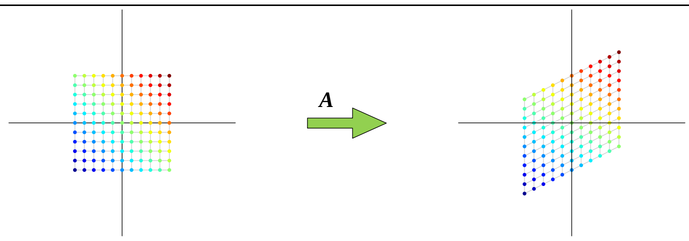
表 1 总结常见 2 × 2 矩阵对应的线性变换、特征值、行列式值。表 1 告诉我们特征值可以为正数、负数、0，甚至是复数。复数特征值都是成对出现，且共轭。本章最后专门讲解特征值分解中出现复数现象。

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。

本章用 Streamlit 制作了一个 App，大家可以自行输入矩阵 $\boldsymbol{A}$ 的值，然后绘制表 1 不同散点图。请参考 Streamlit_Bk4_Ch13_04.py。

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

13.4 对角化、谱分解

可对角化

如果存在一个非奇异矩阵 $\boldsymbol{V}$ 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：

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

谱分解

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

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

其中，$\boldsymbol{V}$ 为正交矩阵，满足 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}$。

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之和，就好比将白光分解成光谱中各色光一样。

再进一步，将 $\boldsymbol{V}$ 整理到 (27) 等式的左边：

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V} = \boldsymbol{\Lambda} \tag{28}$$

同样将 $\boldsymbol{V}$ 写成其列向量并展开上式，

$$\underbrace{\begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}} \end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}} \boldsymbol{A} \underbrace{\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D \end{bmatrix}}_{\boldsymbol{V}} = \underbrace{\begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_1 & \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_D \\ \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_1 & \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_D \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_1 & \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_D \end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}}_{\boldsymbol{\Lambda}} \tag{29}$$

观察上式，我们发现，当 $i = j$ 时，方阵对角线元素满足：

$$\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j = \lambda_j \tag{30}$$

当 $i \neq j$ 时，方阵非对角线元素满足：

$$\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j = 0 \tag{31}$$

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 $\boldsymbol{X}$，它的格拉姆矩阵 $\boldsymbol{G}$ 为 $\boldsymbol{G} = \boldsymbol{X}^{\mathrm{T}}\mathbf{X}$。$\boldsymbol{G}$ 就是 (29) 中的矩阵 $\boldsymbol{A}$，代入得到：

$$\underbrace{\begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_1 & \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_D \\ \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_1 & \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_D \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_1 & \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_D \end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{V}} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}}_{\boldsymbol{\Lambda}} \tag{32}$$

特别地，如果 $\boldsymbol{X}$ 列满秩，$\boldsymbol{G}$ 可逆，$\boldsymbol{G}$ 逆矩阵的特征值分解为：

$$\boldsymbol{G}^{-1} = \boldsymbol{V} \underbrace{\begin{bmatrix} 1/\lambda_1 & & & \\ & 1/\lambda_2 & & \\ & & \ddots & \\ & & & 1/\lambda_D \end{bmatrix}}_{\boldsymbol{\Lambda}^{-1}} \boldsymbol{V}^{\mathrm{T}} \tag{33}$$

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

(32) 可以写成：

$$\underbrace{\begin{bmatrix} \boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{y}_1 & \boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{y}_2 & \cdots & \boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{y}_D \\ \boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{y}_1 & \boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{y}_2 & \cdots & \boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{y}_D \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{y}_D^{\mathrm{T}}\boldsymbol{y}_1 & \boldsymbol{y}_D^{\mathrm{T}}\boldsymbol{y}_2 & \cdots & \boldsymbol{y}_D^{\mathrm{T}}\boldsymbol{y}_D \end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{V}} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}}_{\boldsymbol{\Lambda}} \tag{34}$$

观察上式，我们发现当 $i \neq j$ 时，$\boldsymbol{y}_i$ 和 $\boldsymbol{y}_j$ 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”。

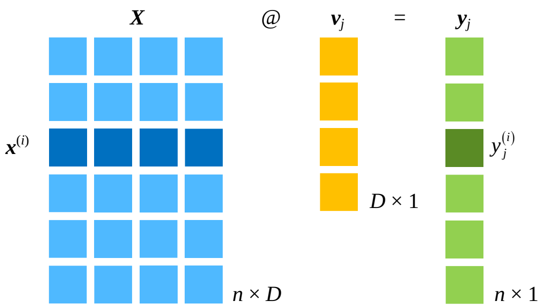


图 7. 数据矩阵 $\boldsymbol{X}$ 向 span($\boldsymbol{v}_j$) 投影结果为 $\boldsymbol{y}_j$

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。

(34) 对角线元素显然可以写成 L^2 范数：

$$\left\|\boldsymbol{y}_j\right\|_2^2 = \left\|\boldsymbol{X}\boldsymbol{v}_j\right\|_2^2 = \lambda_j \tag{35}$$

几何视角

该怎么理解 (35)?

我们还是要拿出看家本领——几何视角。

如图 8 所示，用散点 ● 代表数据矩阵 $\boldsymbol{X}$，散点 ● 向 span($\boldsymbol{v}_j$) 投影结果为 $\boldsymbol{y}_j$，即图中 ×。$\boldsymbol{y}_j$ 中的每个值就是 × 到原点的距离。

图 8 中红点 ● 代表矩阵 $\boldsymbol{X}$ 的第 i 行行向量为 $\boldsymbol{x}^{(i)}$。$\boldsymbol{x}^{(i)}$ 向 $\boldsymbol{v}_j$ 投影结果 $y_j^{(i)}$ 就是 $\boldsymbol{x}^{(i)}$ 在 span($\boldsymbol{v}_j$) 的坐标：

$$y_j^{(i)} = \boldsymbol{x}^{(i)}\boldsymbol{v}_j \tag{36}$$

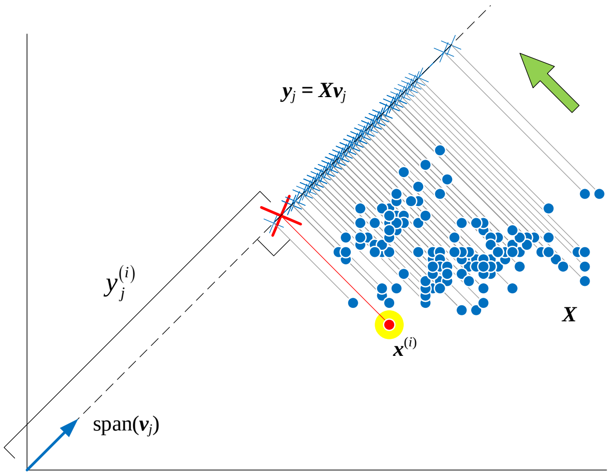


图 8. 数据矩阵 X 向 span(v_j) 投影结果为 y_j，几何视角

有了这个视角，我们知道 (35) 中 $\|y_j\|_2^2$ 代表 $y_j^{(i)}$ 到原点距离 (有正负) 的平方和，即：

$$\|y_j\|_2^2 = \left(y_j^{(1)}\right)^2 + \left(y_j^{(2)}\right)^2 + \cdots \left(y_j^{(n)}\right)^2 = \lambda_j \tag{37}$$

注意，这些距离的平方和恰好等于特征值 λ_j。

若 (34) 中特征值 λ_j 按大小排列，即 $\lambda_1 \geqslant \lambda_2 \geqslant \dots \lambda_D$。这说明特征向量 v_j 也有主次之分。数据矩阵 X 朝不同特征向量 v 投影，得到的 $\|y\|_2^2 = \|Xv\|_2^2$ 有大有小。

如果某个特征值为 0，这说明在它之前的特征向量已经“解释了”矩阵 X 的所有成分。轮到之后的特征向量，投影分量必然为 0。

有大小之分，就意味存在优化问题。我们先给结论，在 $\mathbb{R}^D$ 有无数个 v 中，X 朝第一特征向量 v_1 投影对应的 $\|y_1\|_2^2 = \|Xv_1\|_2^2$ 最大，最大值为 λ_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 X 的格拉姆矩阵 G，如图 9 所示。图 9 中 G 中元素没有保留任何小数位。

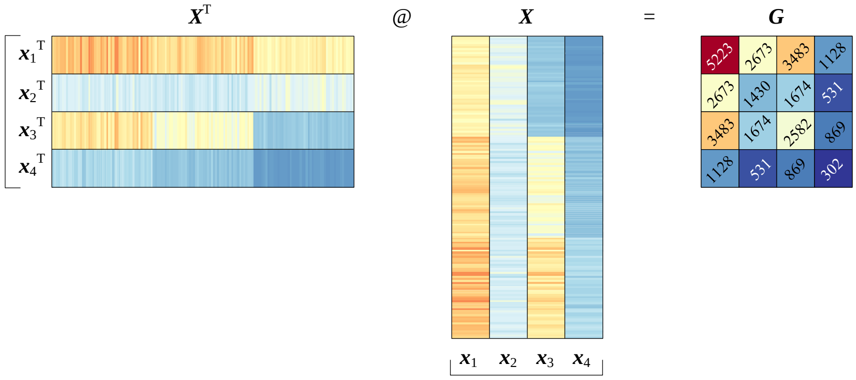


图 9. 矩阵 X 的格拉姆矩阵，图片来自本书第 10 章

格拉姆矩阵 G 为对称矩阵，对 G 特征值分解得到：

$$G = V\Lambda V^{T} = \begin{bmatrix} 0.75 & 0.28 & 0.50 & 0.32 \\ 0.38 & 0.54 & -0.67 & -0.31 \\ 0.51 & -0.70 & -0.05 & -0.48 \\ 0.16 & -0.34 & -0.53 & 0.75 \end{bmatrix} \begin{bmatrix} 9208.3 & & & \\ & 315.4 & & \\ & & 11.9 & \\ & & & 3.5 \end{bmatrix} \begin{bmatrix} 0.75 & 0.28 & 0.50 & 0.32 \\ 0.38 & 0.54 & -0.67 & -0.31 \\ 0.51 & -0.70 & -0.05 & -0.48 \\ 0.16 & -0.34 & -0.53 & 0.75 \end{bmatrix}^{T} \tag{38}$$

上式中，V 仅保留两位小数位，特征值仅保留一位小数位。

(38) 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。

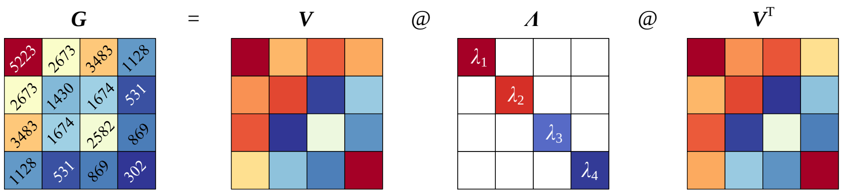


图 10. 矩阵 X 的格拉姆矩阵的特征值分解

利用谱分解方式展开 (38) 得到：

$$\begin{aligned} G &= \lambda_1 v_1 \otimes v_1 + \lambda_2 v_2 \otimes v_2 + \lambda_3 v_3 \otimes v_3 + \lambda_4 v_4 \otimes v_4 \\ &= 9208.3 v_1 \otimes v_1 + 315.4 v_2 \otimes v_2 + 11.9 v_3 \otimes v_3 + 3.5 v_4 \otimes v_4 \end{aligned} \tag{39}$$

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 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 11 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。根据本书第 10 章代码，请大家自行编写代码绘制本节热图。

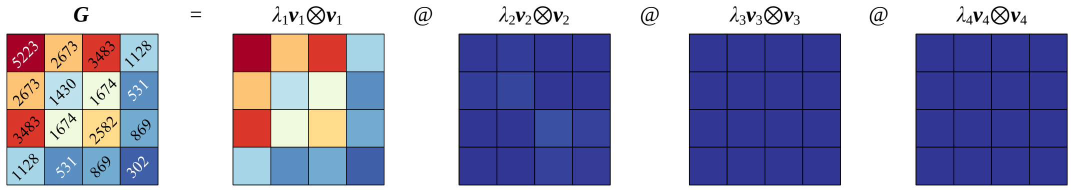


图 11. 矩阵 $\boldsymbol{X}$ 的格拉姆矩阵的谱分解

13.5 聊聊特征值

几何视角

本书第 4 章在讲解行列式值时，简单介绍过特征值。从几何角度来看，如图 12 (a) 所示，当矩阵 $\boldsymbol{A}$ 的形状为 2 × 2 时，以它的两个列向量 $\boldsymbol{a}_1$ 和 $\boldsymbol{a}_2$ 为边的平行四边形面积就是 $\boldsymbol{A}$ 的行列式值。如图 12 (b) 所示，当 $\boldsymbol{A}$ 的形状为 3 × 3 时，$\boldsymbol{a}_1$、$\boldsymbol{a}_2$、$\boldsymbol{a}_3$ 为边的平行六面体体积便是 $\boldsymbol{A}$ 的行列式值。

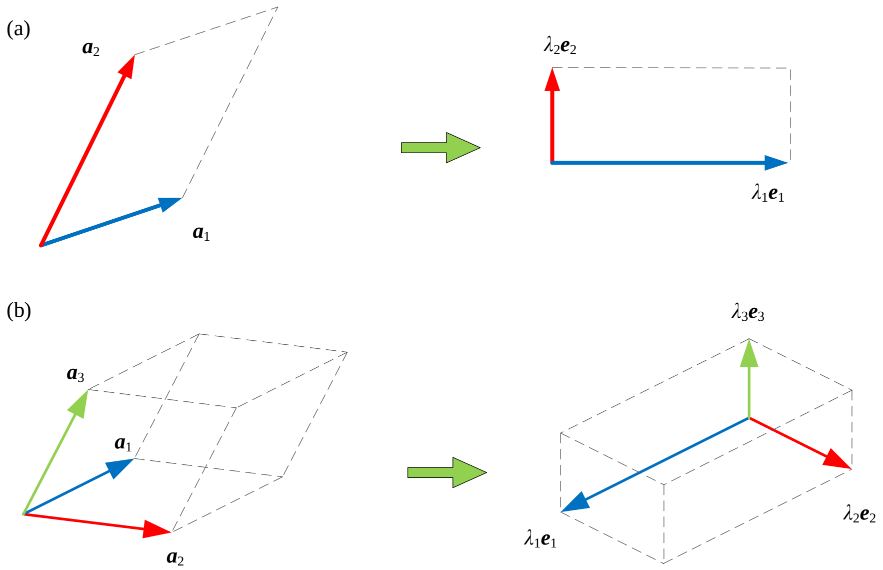


图 12. 特征值的几何性质

比如，给定矩阵 $\boldsymbol{A}$ 为：

$$\boldsymbol{A}=\begin{bmatrix}3 & 2\\1 & 4\end{bmatrix},\quad \boldsymbol{a}_1=\begin{bmatrix}3\\1\end{bmatrix},\quad \boldsymbol{a}_2=\begin{bmatrix}2\\4\end{bmatrix} \tag{40}$$

$\boldsymbol{a}_1$ 和 $\boldsymbol{a}_2$ 为边的平行四边形面积为 10，即 $|\boldsymbol{A}| = 10$。

对矩阵 $\boldsymbol{A}$ 特征值分解后得到的特征值写成矩阵形式，并分别作用于 $\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$：

$$\boldsymbol{\Lambda}=\begin{bmatrix}5 & 0\\0 & 2\end{bmatrix} \quad\Rightarrow\quad \begin{cases}\boldsymbol{\lambda}_1=\lambda_1\boldsymbol{e}_1=5\boldsymbol{e}_1\\ \boldsymbol{\lambda}_2=\lambda_2\boldsymbol{e}_2=2\boldsymbol{e}_2\end{cases} \tag{41}$$

如图 12 (a) 所示，$\lambda_1\boldsymbol{e}_1$ 和 $\lambda_2\boldsymbol{e}_2$ 为边的平行四边形为矩形。容易计算矩形的面积为 $\lambda_1\lambda_2 = 10$，即 $|\boldsymbol{\Lambda}| = \lambda_1\lambda_2$。图 12 (a) 左右两个图形的面积相同，即 $|\boldsymbol{A}| = |\boldsymbol{\Lambda}| = 10$。

从几何角度来看，对角化实际上就是，平行四边形转化为矩形，或者，平行六面体转化为立方体的过程，如图 12 (b)。

如图 13 所示，当矩阵 $\boldsymbol{A}$ 非满秩时，也就是说 $\boldsymbol{A}$ 的列向量线性相关。如果 $\boldsymbol{A}$ 可以对角化，特征值分解后至少一个特征值为 0。这样的话，得到的立方体的体积为 0。也就是说，原来的平行六面体体积也为 0，即 $|\boldsymbol{A}| = 0$。从线性映射角度来看，$\boldsymbol{A}$ 起到降维作用。

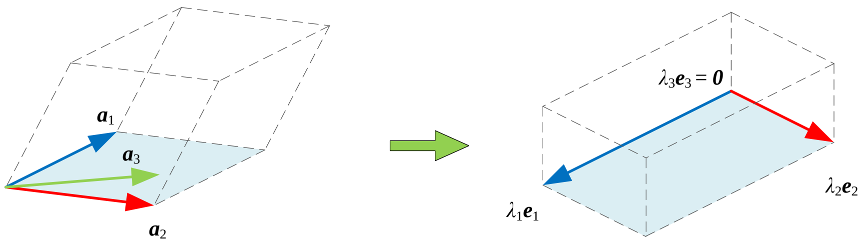


图 13. 特征值的几何性质，线性相关

重要性质

下面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：

$$\boldsymbol{A}\boldsymbol{v}=\lambda\boldsymbol{v} \tag{42}$$

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

$$(k\boldsymbol{A})\boldsymbol{v}=(k\lambda)\boldsymbol{v} \tag{43}$$

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

$$\boldsymbol{A}^2\boldsymbol{v}=\boldsymbol{A}(\boldsymbol{A}\boldsymbol{v})=\boldsymbol{A}(\lambda\boldsymbol{v})=\lambda(\boldsymbol{A}\boldsymbol{v})=\lambda^2\boldsymbol{v} \tag{44}$$

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

$$A^n v = \lambda^n v \tag{45}$$

(45) 也可以推广得到：

$$A^n V = V\Lambda^n \tag{46}$$

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

$$A^{-1} v = \frac{1}{\lambda} v \tag{47}$$

前文提到，矩阵 A 的行列式值为其特征值乘积：

$$\det(A) = \prod_{j=1}^{D} \lambda_j \tag{48}$$

A 标量积 kA 的行列式值为：

$$\det(kA) = k^D \prod_{j=1}^{D} \lambda_j \tag{49}$$

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：

$$\operatorname{tr}(A) = \sum_{i=1}^{D} \lambda_i \tag{50}$$

我们将会在**主成分分析** (Principal Component Analysis, PCA) 中用到 (50) 结论。

13.6 特征值分解中的复数现象

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 × 2 实数矩阵 A：

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \tag{51}$$

对 A 进行特征值分解，得到两个特征值分别为：

$$\lambda_1 = 1+i,\quad \lambda_2 = 1-i \tag{52}$$

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 $\boldsymbol{A}$ 特征方程有一对复数解：

$$\left|\boldsymbol{A} - \lambda \boldsymbol{I}\right| = 0 \tag{53}$$

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。

(51) 中 $\boldsymbol{A}$ 的特征值 λ_1 和 λ_2 对应的特征向量分别是：

$$\boldsymbol{v}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix},\quad \boldsymbol{v}_2 = \begin{bmatrix} -i \\ 1 \end{bmatrix} \tag{54}$$

这样的特征向量被称作**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 $\mathbb{R}^n$ 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 $\boldsymbol{x}$ 的共轭向量 $\bar{\boldsymbol{x}}$，也是 $\mathbb{C}^n$ 中的向量。$\bar{\boldsymbol{x}}$ 中每个元素是 $\boldsymbol{x}$ 对应元素的共轭复数。比如，给定复数向量 $\boldsymbol{x}$ 和对应的共轭向量 $\bar{\boldsymbol{x}}$ 如下：

$$\boldsymbol{x} = \begin{bmatrix} 1+i \\ 3-2i \end{bmatrix},\quad \bar{\boldsymbol{x}} = \begin{bmatrix} 1-i \\ 3+2i \end{bmatrix} \tag{55}$$

一个特殊的 2 × 2 矩阵

给定矩阵 $\boldsymbol{A}$ 如下：

$$\boldsymbol{A} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \tag{56}$$

其中，a 和 b 均为实数，且不同时等于 0。

容易求得 $\boldsymbol{A}$ 的复数特征值为一对共轭复数：

$$\lambda = a \pm bi \tag{57}$$

两者的关系如图 14 所示。图 14 横轴为实部，纵轴为虚部。

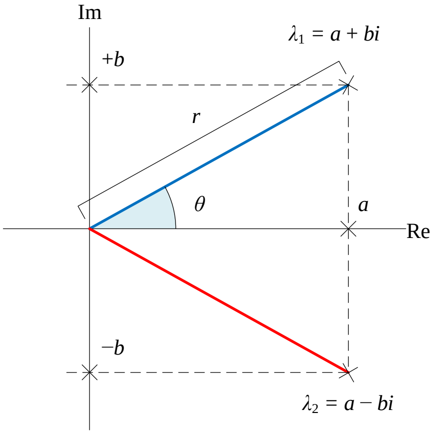


图 14. 一对共轭特征值

图 14 中，两个共轭特征值的模相等，令 r 为复数特征值的模，容易发现，r 是矩阵 $\boldsymbol{A}$ 行列式值的平方根：

$$r = |\lambda| = \sqrt{a^2 + b^2} = \sqrt{|\boldsymbol{A}|} \tag{58}$$

因此，$\boldsymbol{A}$ 可以写成：

$$\boldsymbol{A} = \sqrt{a^2+b^2}\begin{bmatrix} \dfrac{a}{\sqrt{a^2+b^2}} & \dfrac{-b}{\sqrt{a^2+b^2}} \\ \dfrac{b}{\sqrt{a^2+b^2}} & \dfrac{a}{\sqrt{a^2+b^2}} \end{bmatrix} = r\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \underbrace{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}_{\boldsymbol{R}}\underbrace{\begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}}_{\boldsymbol{S}} \tag{59}$$

图 14 所示复平面上，θ 为 (0, 0) 到 (a, b) 线段和水平轴正方向夹角，θ 也称作为复数 $\lambda_1 = a + bi$ 的辐角。

几何视角

有了上述分析，矩阵 $\boldsymbol{A}$ 的几何变换就变得很清楚，$\boldsymbol{A}$ 是缩放 ($\boldsymbol{S}$) 和旋转 ($\boldsymbol{R}$) 的复合。给平面上某个 $\boldsymbol{x}_0$，将矩阵 $\boldsymbol{A}$ 不断作用在 $\boldsymbol{x}_0$ 上：

$$\boldsymbol{x}_n = \boldsymbol{A}^n \boldsymbol{x}_0 \tag{60}$$

如图 15 (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 $\boldsymbol{x}_n$ 不断旋转向外发散。

如图 15 (b) 所示，当缩放系数 $r = 0.8 < 1$，随着 n 增大，向量 $\boldsymbol{x}_n$ 不断旋转向内收缩。

注意，图 14 中平面是复平面，横轴是实数轴，纵轴是虚数轴。而图 15 则是实数 x_1x_2 平面。

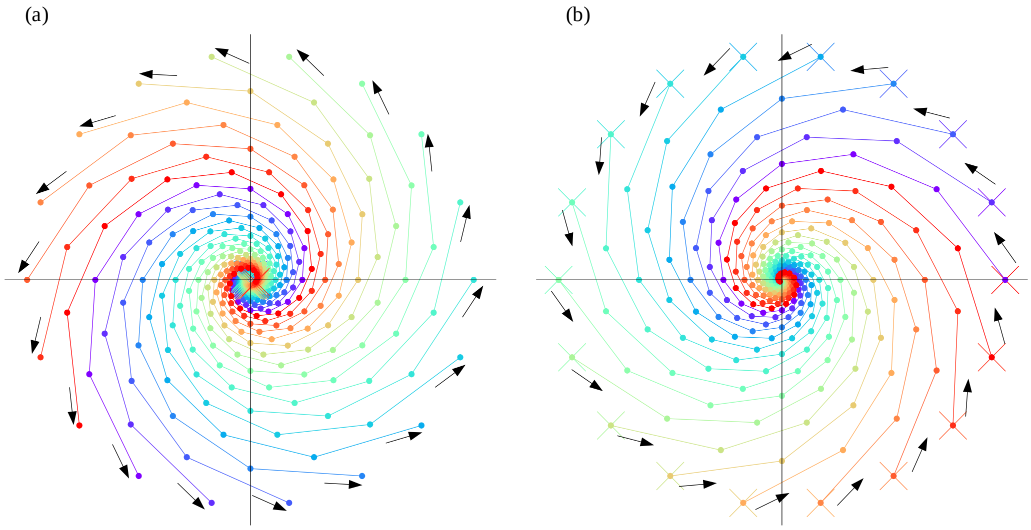


图 15. 在矩阵 $\boldsymbol{A}$ 几何变换重复下，向量的 $\boldsymbol{x}$ 位置变化

Bk4_Ch13_03.py 绘制图 15。

下图四幅子图其实是一张图，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。

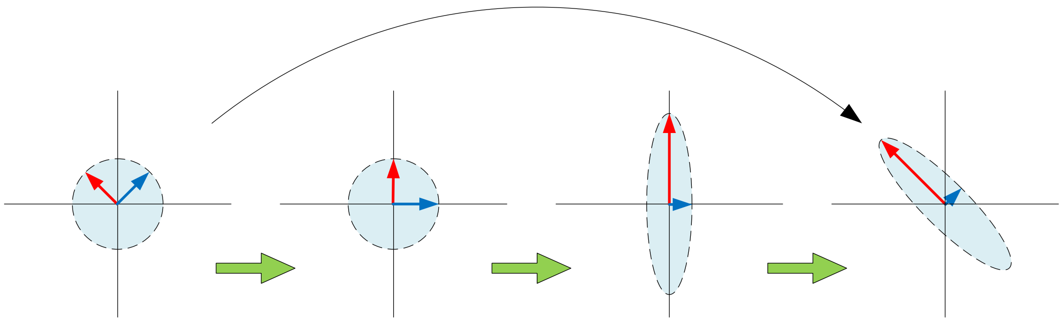


图 16. 总结本章重要内容的四幅图

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

注意，这幅图中展示的几何变换对应的是谱分解，即对称矩阵的特征值分解。更准确地说，谱分解得到的正交矩阵 $\boldsymbol{V}$ 的行列式值为 1 时，$\boldsymbol{V}$ 才叫旋转矩阵，应的几何操作才是纯粹的旋转。当正交矩阵 $\boldsymbol{V}$ 的行列式值为−1 时，$\boldsymbol{V}$ 对应的几何操作是“旋转 + 镜像”。但是为了方便理解，本书很多场合将特征值分解对应的几何操作“简单粗暴”地写成“旋转 → 缩放 → 旋转”。

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，复数矩阵自有一套运算体系，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 $^{\mathrm{H}}$。复数矩阵中的“正交矩阵”叫做**酉矩阵**、**幺正矩阵** (unitary matrix)。再比如，复数矩阵中的“对称矩阵”叫做**正规矩阵** (normal matrix)。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。