13 Eigen Decomposition 特征值分解

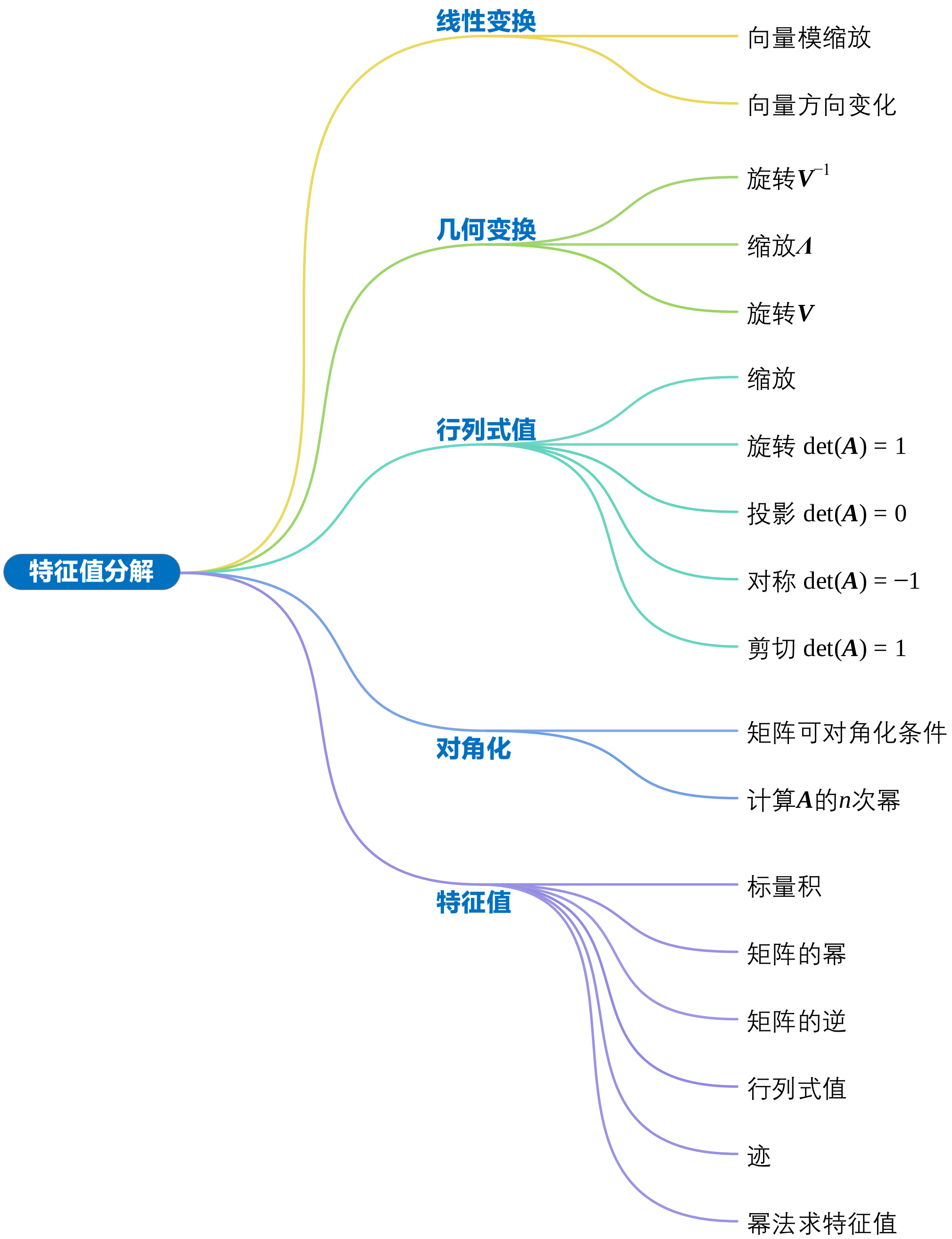
旋转 → 缩放 → 旋转

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
- `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
标量积
矩阵的幂
矩阵的逆
行列式值
迹
幂法求特征值

13.1 几何角度看特征值分解

本书前文讲解线性变换时提到，几何视角下，方阵对应缩放、旋转、投影、剪切等各种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，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。

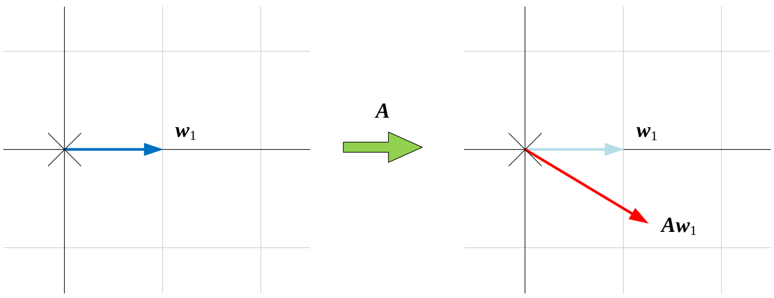


图 1. 我们发现相比原向量 $\boldsymbol{w}_1$，新向量 $\boldsymbol{A}\boldsymbol{w}_1$ 的方向和模都发生变化

下面，给出如图 2 所示的 81 个不同朝向向量 $\boldsymbol{w}$，它们都是单位向量，也就是模均为 1。用矩阵 $\boldsymbol{A}$ 乘这些向量，我们来分析一下 $\boldsymbol{A}\boldsymbol{w}$ 结果特点。

图 3 所示为 $\boldsymbol{A}\boldsymbol{w}$ 对比 $\boldsymbol{w}$ 的结果。$\boldsymbol{w}$ (蓝色箭头) 到 $\boldsymbol{A}\boldsymbol{w}$ (红色箭头) 这个变化过程中，原向量 $\boldsymbol{w}$ 主要发生旋转、缩放。

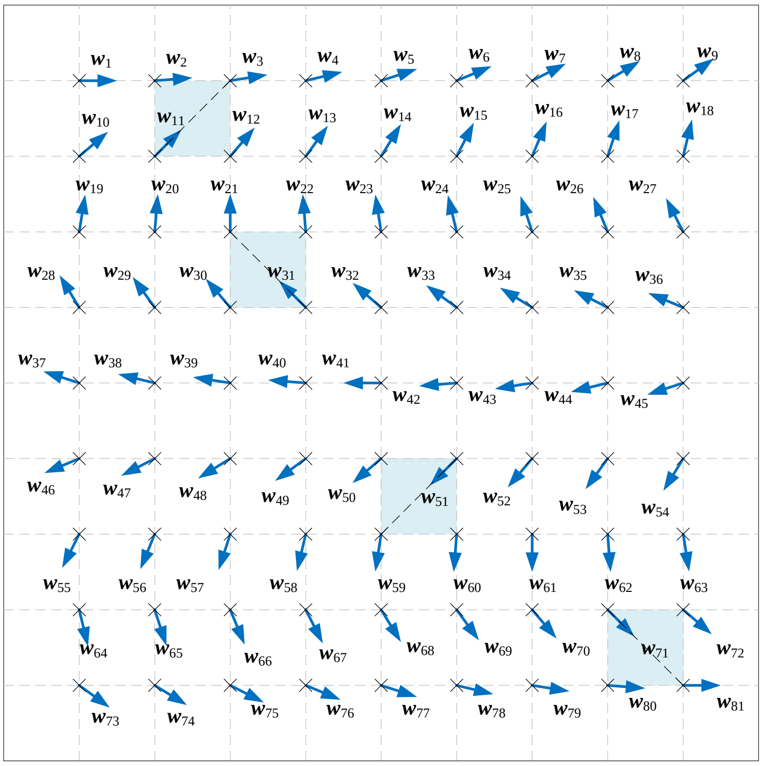


图 2. 81 个朝向不同方向的单位向量

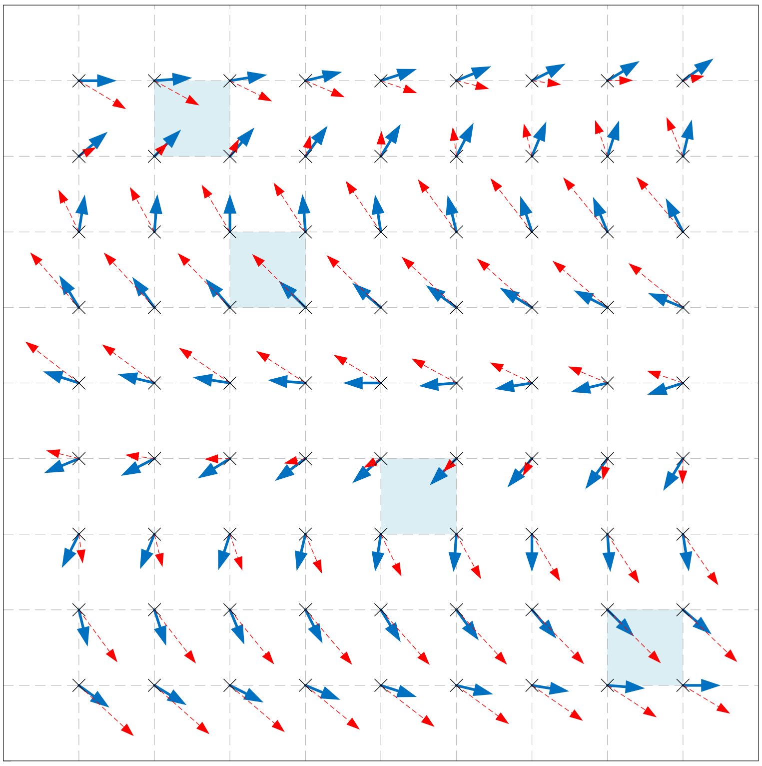


图 3. 矩阵 $\boldsymbol{A}$ 乘 $\boldsymbol{w}$ 得到的结果

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，仅仅发生缩放变化，也就是向量长度变化，但是方向没有变化。

但是如果矩阵 $\boldsymbol{A}$ 对某些向量只发生缩放变换，不产生旋转效果，那么这些向量就称为这个矩阵的特征向量，伸缩的比例就是特征值。

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；注意，为了方便可视化，图 4 左图只展示四个有箭头的线段，它们都是特征向量。

图 4 左图中，向量的终点落在单位圆上。图 4 右图为经过 $\boldsymbol{A}$ 转换后得到向量终点落椭圆上。图 4 中箭头对应的向量就是特征向量。通过图 4 中椭圆相对单位圆的缩放比例，大家也可以大概估算特征值大小。

此外，我们不禁感叹，椭圆真是无处不在；本书后文椭圆还将出现在不同场合。

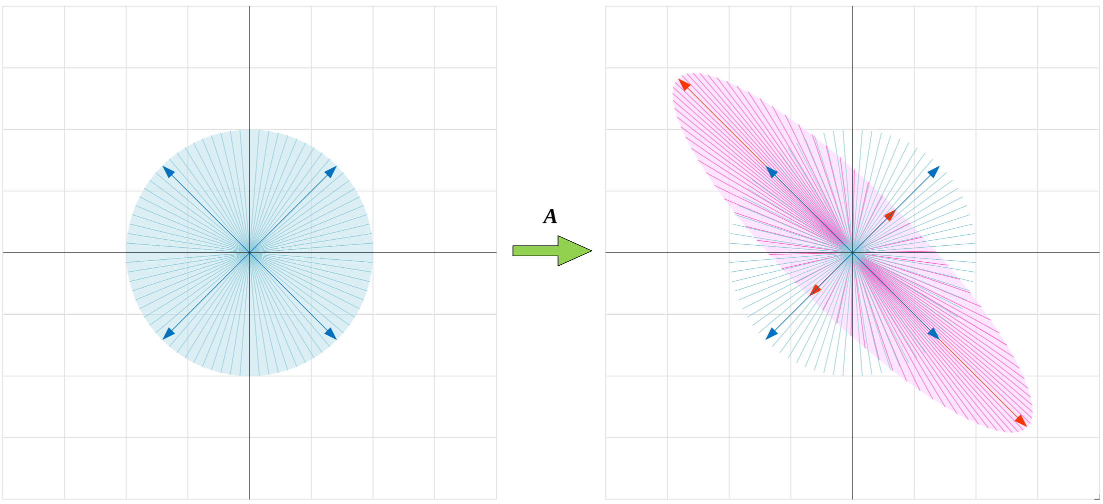


图 4. 矩阵 $\boldsymbol{A}$ 对一系列向量的作用结果

`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，所有图片都经过后期多道工序美化。因此，大家使用代码获得的图片可能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。

```python
# Bk4_Ch13_01.py

import numpy as np
import matplotlib.pyplot as plt

A = np.array([[1.25,  -0.75],
              [-0.75, 1.25]])

xx1, xx2 = np.meshgrid(np.linspace(-8, 8, 9), np.linspace(-8, 8, 9))
num_vecs = np.prod(xx1.shape);

thetas = np.linspace(0, 2*np.pi, num_vecs)

thetas = np.reshape(thetas, (-1, 9))
thetas = np.flipud(thetas);

uu = np.cos(thetas);
vv = np.sin(thetas);

fig, ax = plt.subplots()

ax.quiver(xx1,xx2,uu,vv,
          angles='xy', scale_units='xy',scale=1,
          edgecolor='none', facecolor= 'b')

plt.ylabel('$x_2$')
plt.xlabel('$x_1$')
plt.axis('scaled')
ax.set_xlim([-10, 10])
ax.set_ylim([-10, 10])
ax.grid(linestyle='--', linewidth=0.25, color=[0.5,0.5,0.5])
ax.set_xticks(np.linspace(-10,10,11));
ax.set_yticks(np.linspace(-10,10,11));
plt.show()

# Matrix multiplication
V = np.array([uu.flatten(),vv.flatten()]).T;
W = V@A;

uu_new = np.reshape(W[:,0],(-1, 9));
vv_new = np.reshape(W[:,1],(-1, 9));

fig, ax = plt.subplots()

ax.quiver(xx1,xx2,uu,vv,
          angles='xy', scale_units='xy',scale=1,
          edgecolor='none', facecolor= 'b')

ax.quiver(xx1,xx2,uu_new,vv_new,
          angles='xy', scale_units='xy',scale=1,
          edgecolor='none', facecolor= 'r')

plt.ylabel('$x_2$')
plt.xlabel('$x_1$')
plt.axis('scaled')
ax.set_xlim([-10, 10])
ax.set_ylim([-10, 10])
ax.grid(linestyle='--', linewidth=0.25, color=[0.5,0.5,0.5])
ax.set_xticks(np.linspace(-10,10,11));
ax.set_yticks(np.linspace(-10,10,11));
plt.show()


fig, ax = plt.subplots()
ax.quiver(xx1*0,xx2*0,uu,vv,
          angles='xy', scale_units='xy',scale=1,
          edgecolor='none', facecolor= 'b')

ax.quiver(xx1*0,xx2*0,uu_new,vv_new,
          angles='xy', scale_units='xy',scale=1,
          edgecolor='none', facecolor= 'r')
```

```
plt.ylabel('$x_2$')
plt.xlabel('$x_1$')
plt.axis('scaled')
ax.set_xlim([-2, 2])
ax.set_ylim([-2, 2])
ax.grid(linestyle='--', linewidth=0.25, color=[0.5,0.5,0.5])
ax.set_xticks(np.linspace(-2,2,5));
ax.set_yticks(np.linspace(-2,2,5));
plt.show()
```

13.2 旋转 → 缩放 → 旋转

根据上文所述，矩阵 $\boldsymbol{A}$ 的特征值分解可以写成：

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

注意，如果没有特殊说明，本书中的特征向量都是单位向量。

$\boldsymbol{A}$ 乘任意向量是一个"旋转 → 缩放 →旋转" 顺序操作，即，

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

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 $\boldsymbol{V}$ 展开写作 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$：

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

(7) 展开得到：

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

对于上一节给出的例子，将具体数值代入 (4)，可以得到：

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

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

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 $\boldsymbol{v}_1$，得到的结果 $\boldsymbol{A}\boldsymbol{v}_1$ 方向没有发生变化，仅仅施加缩放效果，缩放的比例为 $\lambda_1 = 1/2$。

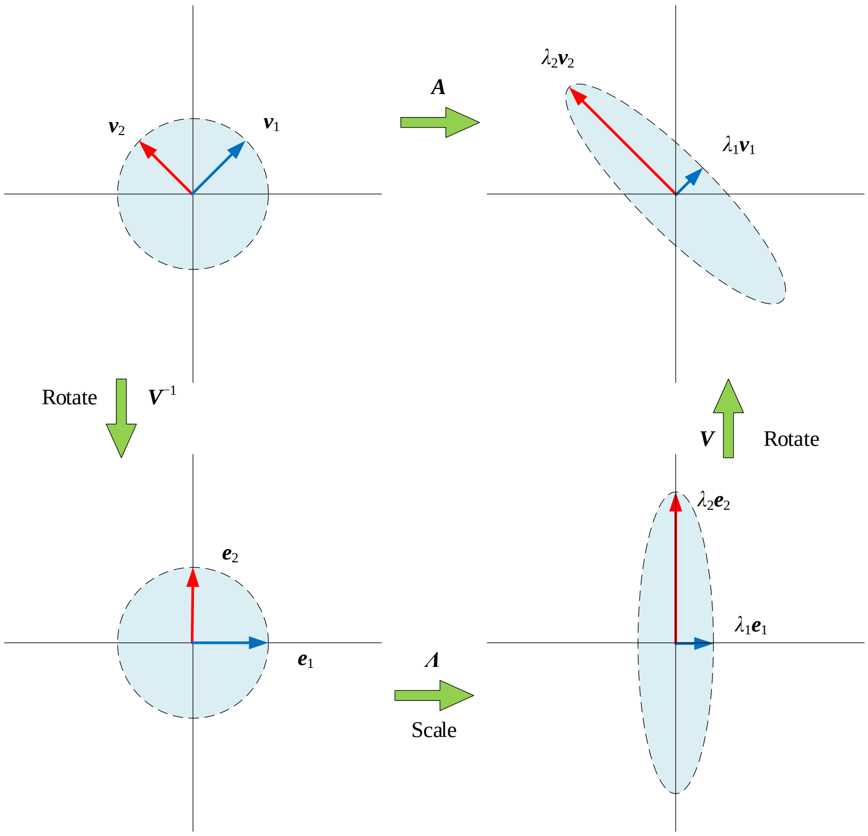


图 5. "旋转→缩放→旋转"操作

如图 5 中蓝色箭头所示，将 (4) 代入 (11)，将 $\boldsymbol{A}$ 拆解为"旋转→缩放→旋转"：

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

$V^{-1}v_1$ 相对 v_1 顺时针旋转 45°：

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

然后再利用 Λ 完成缩放操作，得到 $\Lambda V^{-1}v_1$：

$$\Lambda V^{-1}v_1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} = \underset{\lambda_1}{0.5} e_1 \tag{14}$$

最后逆时针旋转 45°，得到 $V\Lambda V^{-1}v_1$：

$$\begin{aligned} \underbrace{V\Lambda V^{-1}}_{A} v_1 &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \underset{\lambda_1}{0.5} e_1 \\ &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/4 \\ \sqrt{2}/4 \end{bmatrix} = \underset{\lambda_1}{0.5} \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \\ &= \lambda_1 v_1 \end{aligned} \tag{15}$$

第二特征向量

同理，讨论 A 乘 v_2 对应的“旋转→缩放→旋转”操作。

v_2 为：

$$v_2 = \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \tag{16}$$

A 乘 v_2 得到 Av_2：

$$Av_2 = \begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix} \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \end{bmatrix} = \underset{\lambda_2}{2} \times \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \tag{17}$$

$V^{-1}v_2$ 将 v_2 顺时针旋转 45°：

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

再缩放得到 $\Lambda V^{-1}v_2$：

$$\underset{\text{Scale}}{\Lambda} V^{-1} v_2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \underset{\lambda_2}{2} e_2 \tag{19}$$

最后旋转得到 $V\Lambda V^{-1}v_2$：

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

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。

```python
# Bk4_Ch13_02.py

import numpy as np
import matplotlib.pyplot as plt

def visualize(X_circle,X_vec,title_txt):

    fig, ax = plt.subplots()

    plt.plot(X_circle[0,:], X_circle[1,:],'k',
             linestyle = '--',
             linewidth = 0.5)

    plt.quiver(0,0,X_vec[0,0],X_vec[1,0],
              angles='xy', scale_units='xy',scale=1,
              color = [0, 0.4392, 0.7529])

    plt.quiver(0,0,X_vec[0,1],X_vec[1,1],
              angles='xy', scale_units='xy',scale=1,
              color = [1,0,0])

    plt.axvline(x=0, color= 'k', zorder=0)
    plt.axhline(y=0, color= 'k', zorder=0)

    plt.ylabel('$x_2$')
    plt.xlabel('$x_1$')

    ax.set_aspect(1)
    ax.set_xlim([-2.5, 2.5])
    ax.set_ylim([-2.5, 2.5])
    ax.grid(linestyle='--', linewidth=0.25, color=[0.5,0.5,0.5])
    ax.set_xticks(np.linspace(-2,2,5));
    ax.set_yticks(np.linspace(-2,2,5));
    plt.title(title_txt)
    plt.show()


theta = np.linspace(0, 2*np.pi, 100)

circle_x1 = np.cos(theta)
circle_x2 = np.sin(theta)

V_vec = np.array([[np.sqrt(2)/2, -np.sqrt(2)/2],
                  [np.sqrt(2)/2,  np.sqrt(2)/2]])

X_circle = np.array([circle_x1, circle_x2])

# plot original circle and two vectors
visualize(X_circle,V_vec,'Original')
```

```
A = np.array([[1.25, -0.75],
              [-0.75, 1.25]])

# plot the transformation of A

visualize(A@X_circle, A@V_vec,'$A$')


#%% Eigen deomposition

# A = V @ D @ V.T

lambdas, V = np.linalg.eig(A)

D = np.diag(np.flip(lambdas))
V = V.T # reverse the order

print('=== LAMBDA ===')
print(D)
print('=== V ===')
print(V)

# plot the transformation of V.T

visualize(V.T@X_circle, V.T@V_vec,'$V^T$')

# plot the transformation of D @ V.T

visualize(D@V.T@X_circle, D@V.T@V_vec,'$\u039BV^T$')

# plot the transformation of V @ D @ V.T

visualize(V@D@V.T@X_circle, V@D@V.T@V_vec,'$V\u039BV^T$')

# plot the transformation of A

visualize(A@X_circle, A@V_vec,'$A$')
```

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。

这一点也可以通过特征值加以验证：

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

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，正方形以及完成“旋转→缩放→旋转”等几何变化。

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，它的长轴放大 2 倍，而短轴缩小至原来一半。但是，得到的椭圆面积和原来的单位圆一样。

此外，观察上式可以发现，矩阵 $\boldsymbol{A}$ 的行列式值等于所有特征值之积。

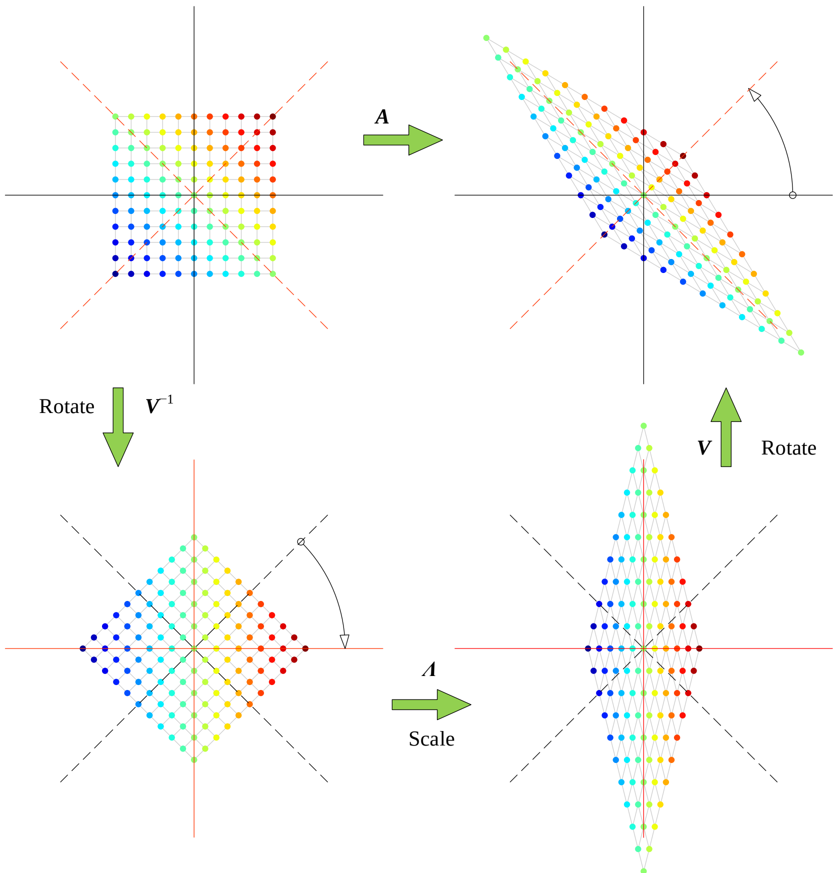


图 6. 正方形经过矩阵 $\boldsymbol{A}$ 线性变换

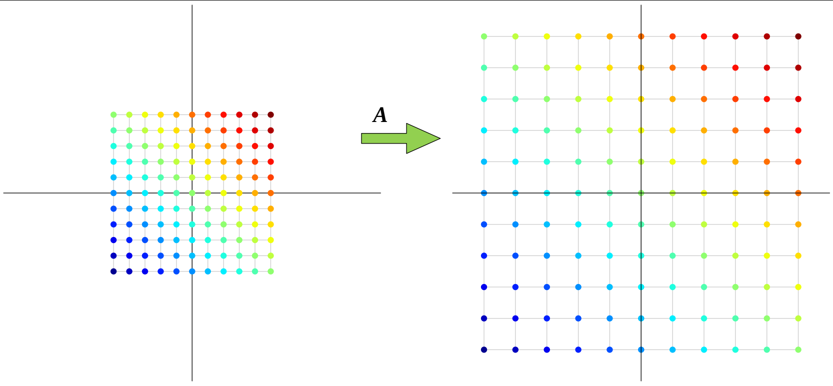
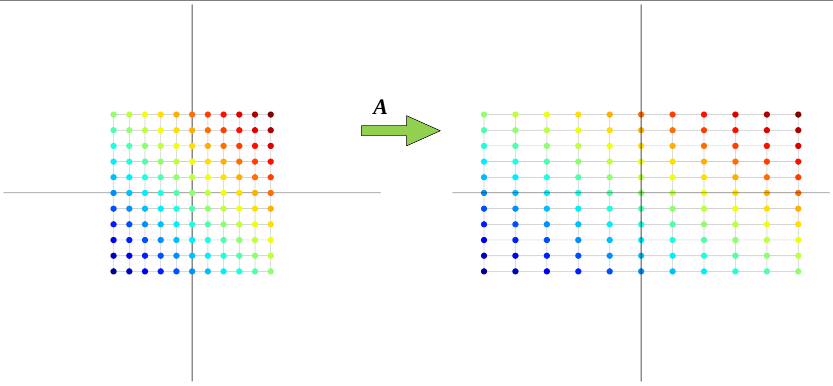
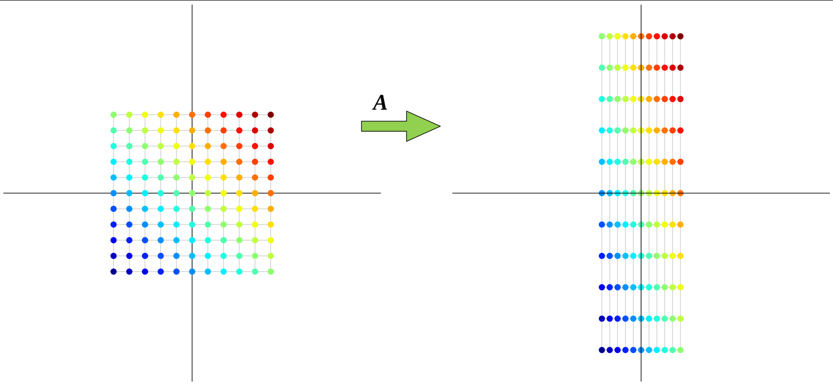
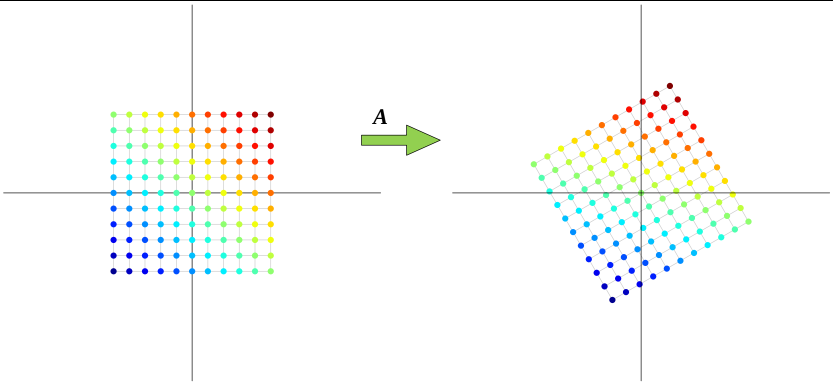
常见线性变换矩阵的特征值和行列式值

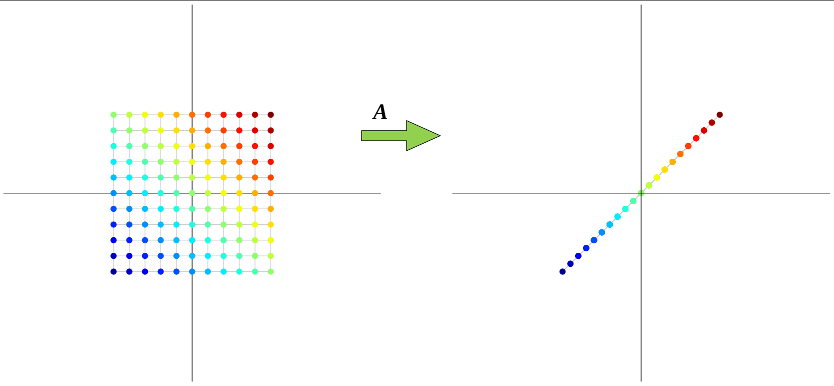
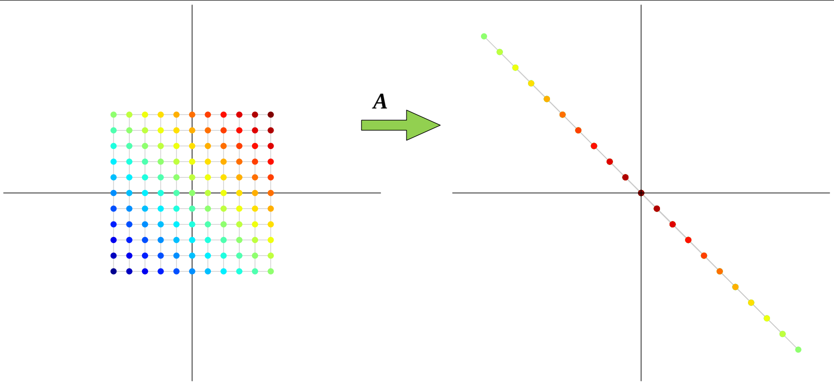
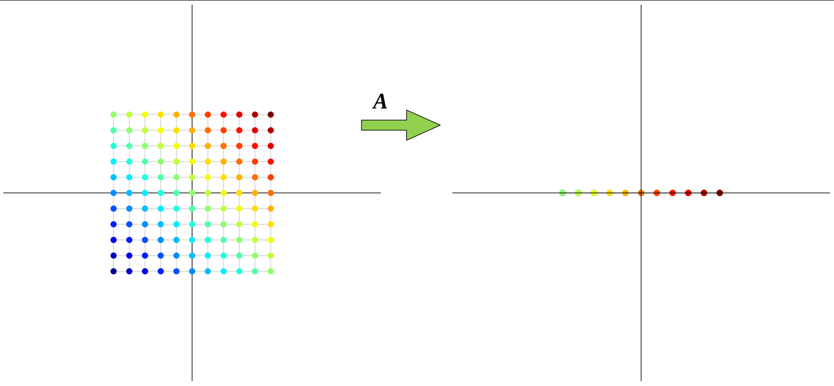
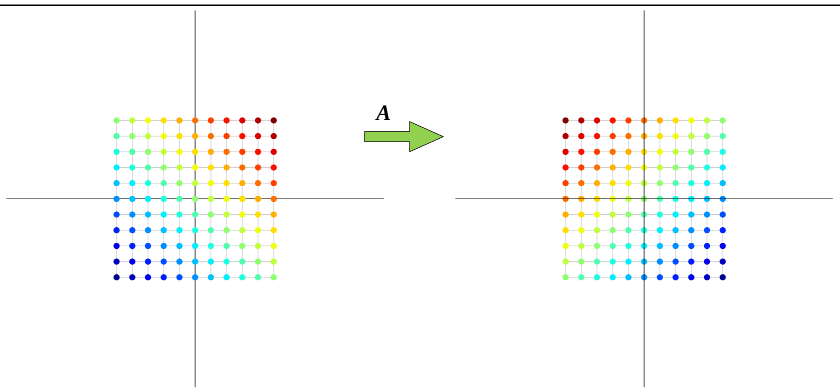
表 1 总结常见线性变换中特征值、行列式值关系。

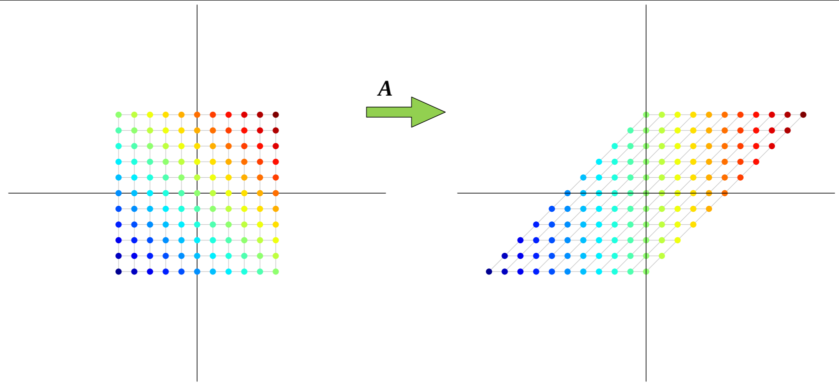
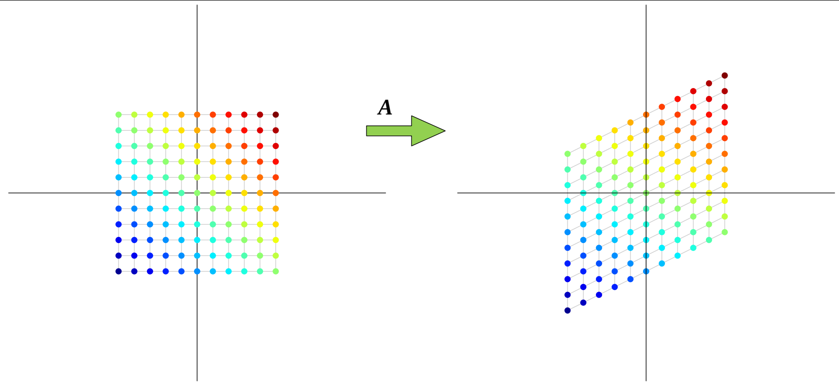
观察表 1，可以发现特征值可以为正数、负数、0，甚至是复数。复数特征值都是成对出现，且共轭。下一章会专门讲解特征值分解中出现复数的情况。

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. 常见线性变换中特征值、行列式值关系

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 $\boldsymbol{A}=\begin{bmatrix}0.5 & 0.5\\0.5 & 0.5\end{bmatrix}$ $\begin{cases}\lambda_1=1\\\lambda_2=0\end{cases}$ $\det(\boldsymbol{A})=0$	
非正交投影 $\boldsymbol{A}=\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix}$ $\begin{cases}\lambda_1=2\\\lambda_2=0\end{cases}$ $\det(\boldsymbol{A})=0$	
投影 $\boldsymbol{A}=\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$ $\begin{cases}\lambda_1=1\\\lambda_2=0\end{cases}$ $\det(\boldsymbol{A})=0$	
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：将方阵转化为对角阵

如果选在一个非奇异矩阵 $\boldsymbol{V}$ 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 $\boldsymbol{A}$ 为**可对角化** (diagonalizable)，$\boldsymbol{V}$ 将 $\boldsymbol{A}$ 对角化。

只有可对角化的矩阵才可以进行特征值分解：

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

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：

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

特别地，如果 A 为对称矩阵，A 的特征值分解可以写成：

$$\begin{aligned} A = V\Lambda V^{\mathrm{T}} &= \begin{bmatrix} v_1 & v_2 & \cdots & v_D \end{bmatrix}\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}\begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ \vdots \\ v_D^{\mathrm{T}} \end{bmatrix} \\ &= \lambda_1 v_1 v_1^{\mathrm{T}} + \lambda_2 v_2 v_2^{\mathrm{T}} + \cdots \lambda_D v_D v_D^{\mathrm{T}} = \sum_{j=1}^{D} \lambda_j v_j v_j^{\mathrm{T}} \\ &= \lambda_1 v_1 \otimes v_1 + \lambda_2 v_2 \otimes v_2 + \cdots \lambda_D v_D \otimes v_D = \sum_{j=1}^{D} \lambda_j v_j \otimes v_j \end{aligned} \tag{27}$$

上式告诉我们为什么特征分解又称谱分解 (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。

13.5 深入聊聊特征值

本节介绍特征值重要性质。

给定矩阵 A，特征值 λ 和特征向量 v 关系为：

$$Av = \lambda v \tag{28}$$

A 标量积 kA，kA 对应的特征值为 λk，即，

$$(kA)v = (k\lambda)v \tag{29}$$

矩阵 A^2 的特征向量仍然为 v，特征值为 λ^2：

$$A^2 v = A(Av) = A(\lambda v) = \lambda(Av) = \lambda^2 v \tag{30}$$

推广上式，n 为任意整数，A^n 的特征值为 λ^n：

$$A^n v = \lambda^n v \tag{31}$$

(31) 也可以推广得到：

$$A^n V = V\Lambda^n \tag{32}$$

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

$$A^{-1}v = \frac{1}{\lambda}v \tag{33}$$

矩阵 A 的行列式值为其特征值乘积:

$$\det(A) = \prod_{i=1}^{D} \lambda_i \tag{34}$$

上式中，λ_i 可以相等。如果任一特征值为 0，则 $\det(A)$ 为零，这种情况相当于降维。这正是前文讲过的多维空间“平行体”和“正立方体”的体积关系。

特征向量 v_i 和特征值 λ_i 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 v_i 方向上伸缩比例相同。

A 标量积 kA 的行列式值:

$$\det(kA) = k^D \prod_{i=1}^{D} \lambda_i \tag{35}$$

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

矩阵 A 的迹为其特征值之和:

$$\operatorname{tr}(A) = \sum_{i=1}^{D} \lambda_i \tag{36}$$

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

13.6 幂法求特征值

幂法 (power method) 可以用来求某个矩阵特征向量中模最大的特征值和特征向量。

幂法指的是，对于一个 $D \times D$ 方阵 A，先取任意 D 行一列向量 x_0，然后进行如下迭代:

$$\begin{aligned} x_1 &= Ax_0 \\ x_2 &= Ax_1 = A^2 x_0 \\ x_3 &= Ax_2 = A^3 x_0 \\ &\cdots \\ x_n &= Ax_{n-1} = A^n x_0 \end{aligned} \tag{37}$$

假设矩阵 A 有 D 个线性无关的特征向量，按照模的大小排列这些特征值:

$$|\lambda_1| \ge |\lambda_2| \ge \cdots |\lambda_D| \tag{38}$$

它们对应的特征向量为:

$$\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D \tag{39}$$

它们作为基底可以张成 D 维向量空间，初始向量 $\boldsymbol{x}_0$ 可以由它们线性组合得到：

$$\boldsymbol{x}_0 = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \ldots + a_D\boldsymbol{v}_D \tag{40}$$

将 (40) 代入 (37) 最后一个等式，得到：

$$\boldsymbol{x}_n = a_1\boldsymbol{A}^n\boldsymbol{v}_1 + a_2\boldsymbol{A}^n\boldsymbol{v}_2 + \ldots + a_D\boldsymbol{A}^n\boldsymbol{v}_D \tag{41}$$

上一节介绍，给定矩阵 $\boldsymbol{A}$，特征值 λ 对应的特征向量为 $\boldsymbol{v}$；而 $\boldsymbol{A}^n$ 的特征值为 λ^n，即，

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

将 (31) 代入(41)，得到：

$$\boldsymbol{x}_n = a_1\lambda_1^n\boldsymbol{v}_1 + a_2\lambda_2^n\boldsymbol{v}_2 + \ldots + a_D\lambda_D^n\boldsymbol{v}_D \tag{43}$$

整理上式，得到：

$$\boldsymbol{x}_n = \lambda_1^n\left(a_1\boldsymbol{v}_1 + a_2\left(\frac{\lambda_2}{\lambda_1}\right)^n\boldsymbol{v}_2 + \ldots + a_D\left(\frac{\lambda_D}{\lambda_1}\right)^n\boldsymbol{v}_D\right) \tag{44}$$

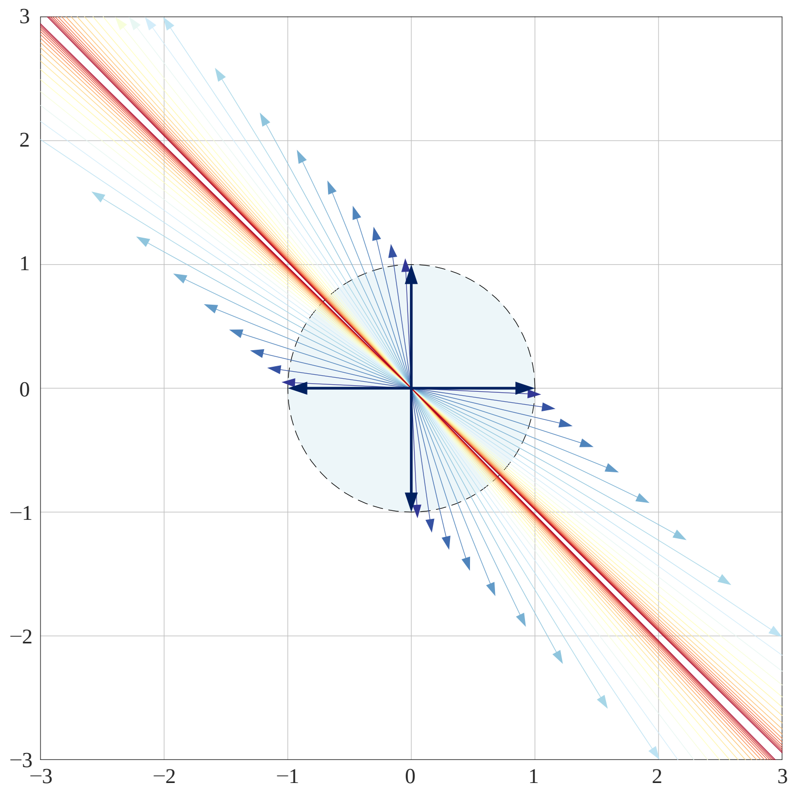


图 7. 四个向量经过幂乘收敛于最大特征值对应方向

因为 λ_1 为最大实数特征值，因此当 n 足够大时，(44) 收敛到：

$$\boldsymbol{x}_n \to \lambda_1^n a_1 \boldsymbol{v}_1 \tag{45}$$

图 7 所示为四个单位向量经过幂乘收敛于最大特征值对应方向。

实际计算时，为了避免计算过程中出现绝对值过大或过小的数参加运算，通常在每步迭代时，将用向量的 L^∞范数“归一化”处理，即：

$$\boldsymbol{z}_n = \frac{\boldsymbol{x}_n}{\left\|\boldsymbol{x}_n\right\|_\infty}, \quad \boldsymbol{x}_{n+1} = \boldsymbol{A}\boldsymbol{z}_n \tag{46}$$

请大家自行编程绘制图 7。

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 $\boldsymbol{V}$ 为正交矩阵，即规范正交基。

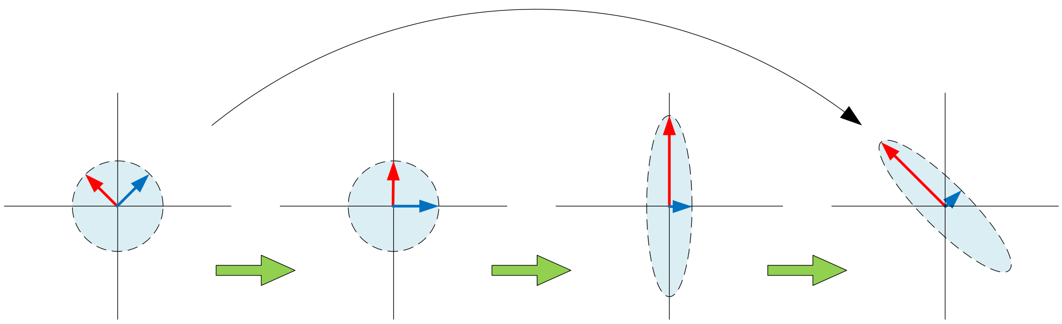


图 8. 总结本章重要内容的四副图