

# 20

## Revisit Conic Sections

# 再谈圆锥曲线

从矩阵运算和几何变换视角



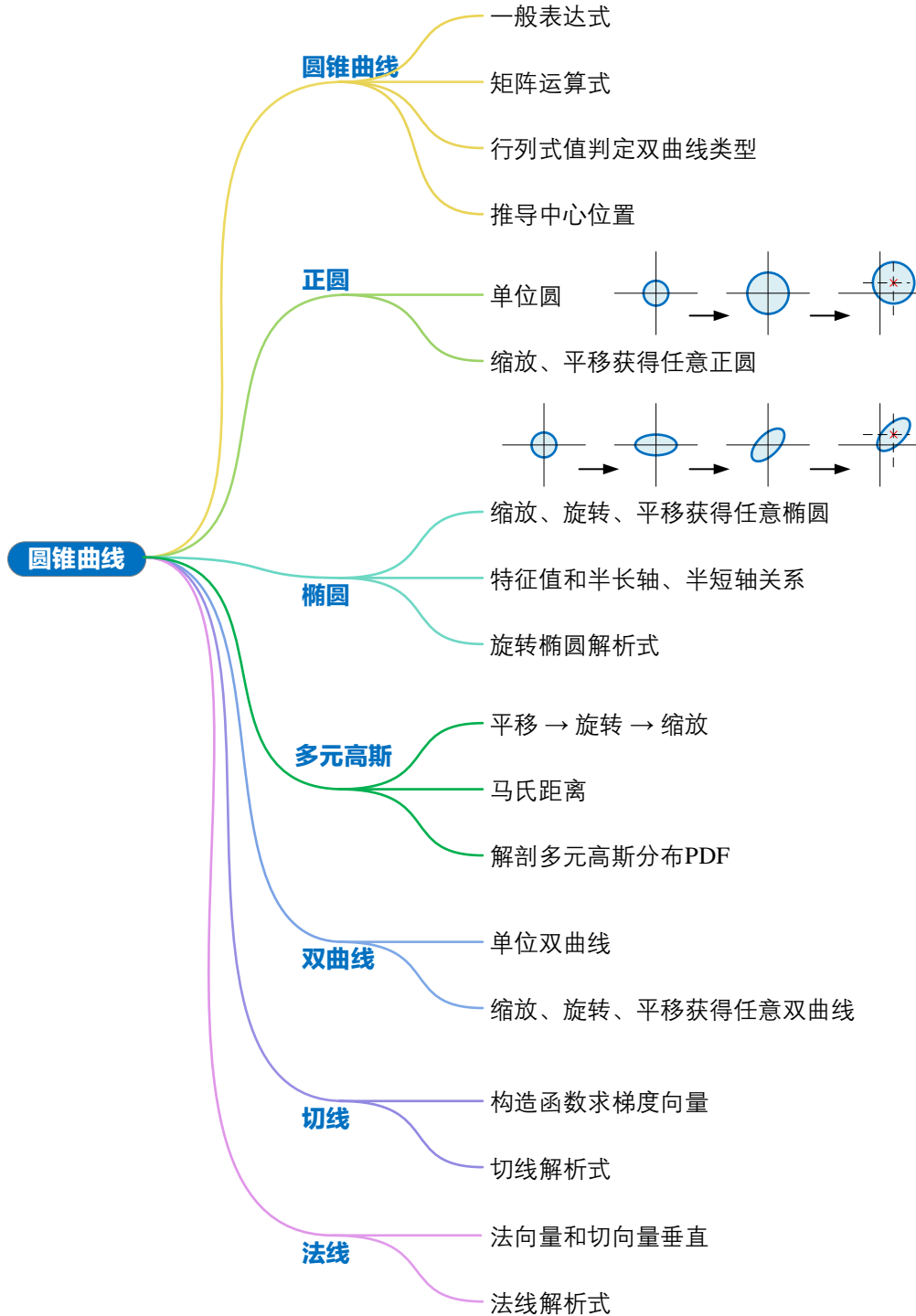
滴水穿石，靠的不是力量，而是持之以恒。

*Dripping water hollows out stone, not through force but through persistence.*

—— 奥维德 (Ovid) | 古罗马诗人 | 43 BC ~ 17/18 AD



- ◀ matplotlib.pyplot.contour() 绘制等高线图
- ◀ numpy.cos() 计算余弦值
- ◀ numpy.linalg.eig() 特征值分解
- ◀ numpy.linalg.inv() 矩阵求逆
- ◀ numpy.sin() 计算正弦值
- ◀ numpy.tan() 计算正切值



## 20.1 无处不在的圆锥曲线

本套丛书每一册几乎都离不开圆锥曲线这个话题。

《数学要素》第 8、9 章详细介绍过圆锥曲线相关性质，《概率统计》一册会讨论圆锥曲线和高斯分布千丝万缕的联系。同时我们也看到条件概率、回归分析和主成分分析中，圆锥曲线扮演重要角色。《机器学习》一册介绍的很多算法中，决策边界就是圆锥曲线。

利用本书前文讲解的线性代数工具，本章将从矩阵运算和几何变换角度探讨圆锥曲线。这个视角将帮助大家更加深刻理解圆锥曲线在概率统计、数据学科和机器学习的重要作用。

### 一般表达式

圆锥曲线一般表达式如下：

$$Ax_1^2 + Bx_1x_2 + Cx_2^2 + Dx_1 + Ex_2 + F = 0 \quad (1)$$

把 (1) 写成矩阵运算式：

$$\frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} D \\ E \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + F = 0 \quad (2)$$

(2) 进一步写成：

$$\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{w}^T \mathbf{x} + F = 0 \quad (3)$$

其中，

$$\mathbf{Q} = \begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} D \\ E \end{bmatrix} \quad (4)$$

矩阵  $\mathbf{Q}$  的行列式值为：

$$\det \mathbf{Q} = \det \begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix} = 4AC - B^2 \quad (5)$$

矩阵  $\mathbf{Q}$  的行列式值决定圆锥曲线的形状：

- ◀ 当  $4AC - B^2 > 0$ ，上式为**椭圆** (ellipse)；特别地，当  $A = C$  且  $B = 0$ ，解析式为**正圆** (circle)；
- ◀  $4AC - B^2 = 0$  时，解析式为**抛物线** (parabola)；
- ◀  $4AC - B^2 < 0$  时，解析式为**双曲线** (hyperbola)。

这实际上回答了本系列丛书《数学要素》中提出的一个问题——为什么用  $4AC - B^2$  判断圆锥曲线形状。

## 中心

当  $4AC - B^2$  不等于 0，圆锥曲线中椭圆、正圆和双曲线这三类曲线存在中心。依照 (2) 构造如下二元函数  $f(x_1, x_2)$ ：

$$f(x_1, x_2) = \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} D \\ E \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + F \quad (6)$$

下面介绍如何求解圆锥曲线中心。

$f(x_1, x_2)$  对  $[x_1, x_2]^T$  一阶导数为  $[0, 0]^T$  时，也就是梯度向量为  $\mathbf{0}$  时， $(x_1, x_2)$  为  $f(x_1, x_2)$  驻点：

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (7)$$

这个驻点就是圆锥曲线的中心。推导圆锥曲线中心位置：

$$\begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} D \\ E \end{bmatrix} = \mathbf{0} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = - \begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix}^{-1} \begin{bmatrix} D \\ E \end{bmatrix} \quad (8)$$

回忆  $2 \times 2$  方阵的逆：

$$\begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix}^{-1} = \frac{1}{\underbrace{4AC - B^2}_{\text{Determinant}}} \begin{bmatrix} 2C & -B \\ -B & 2A \end{bmatrix} \quad (9)$$

将 (9) 代入 (8)，得到圆锥曲线中心  $\mathbf{c}$  坐标：

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \frac{1}{B^2 - 4AC} \begin{bmatrix} 2CD - BE \\ 2AE - BD \end{bmatrix} \quad (10)$$

大家通过上式也知道了，为什么对于椭圆、正圆和双曲线会要求  $4AC - B^2$  不等于 0。

## 20.2 正圆：从单位圆到任意正圆

### 单位圆

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

$$\mathbf{x}^T \mathbf{x} - 1 = 0 \quad (11)$$

其中  $\mathbf{x}$  为,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (12)$$

展开 (11) 得到:

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 1 = x_1^2 + x_2^2 - 1 = 0 \quad (13)$$

当然, (11) 可以用  $L^2$  范数、向量内积等方式表达单位圆, 比如:

$$\begin{aligned} \|\mathbf{x}\|_2 - 1 &= 0 \\ \|\mathbf{x}\|_2^2 - 1 &= 0 \\ \mathbf{x} \cdot \mathbf{x} - 1 &= 0 \\ \langle \mathbf{x}, \mathbf{x} \rangle - 1 &= 0 \end{aligned} \quad (14)$$

其中,  $\|\mathbf{x}\|_2 - 1 = 0$  可以写成  $\|\mathbf{x} - \mathbf{0}\|_2 - 1 = 0$ , 代表  $\mathbf{x}$  距离原点  $\mathbf{0}$  的  $L^2$  范数 (欧几里得距离) 为 1。

## 缩放

圆心位于原点半径为  $r$  的正圆解析式为:

$$\mathbf{x}^T \mathbf{x} - r^2 = 0 \quad (15)$$

上式相当于,

$$\mathbf{x}^T \begin{bmatrix} 1/r^2 & 0 \\ 0 & 1/r^2 \end{bmatrix} \mathbf{x} - 1 = 0 \quad (16)$$

将 (16) 写成:

$$\mathbf{x}^T \underbrace{\begin{bmatrix} 1/r & 0 \\ 0 & 1/r \end{bmatrix}}_{S^{-1}} \underbrace{\begin{bmatrix} 1/r & 0 \\ 0 & 1/r \end{bmatrix}}_{S^{-1}} \mathbf{x} - 1 = 0 \quad (17)$$

令矩阵  $S$  为:

$$S = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \quad (18)$$

由于  $S$  为对角方阵, 因此 (17) 可以进一步整理为:

$$\mathbf{x}^T S^{-1} S^{-1} \mathbf{x} - 1 = (\mathbf{S}^{-1} \mathbf{x})^T \mathbf{S}^{-1} \mathbf{x} - 1 = 0 \quad (19)$$

从几何变换视角来观察, 相信大家已经在 (19) 中看到  $S$  起到缩放作用。

## 缩放 + 平移

圆心位于  $\mathbf{c} = [c_1, c_2]^T$  半径为  $r$  的正圆解析式为：

$$(\mathbf{x} - \mathbf{c})^T \begin{bmatrix} 1/r^2 & 0 \\ 0 & 1/r^2 \end{bmatrix} (\mathbf{x} - \mathbf{c}) - 1 = 0 \quad (20)$$

(20) 也可以写成：

$$\begin{aligned} (\mathbf{x} - \mathbf{c})^T (\mathbf{x} - \mathbf{c}) - r^2 &= 0 \\ \|\mathbf{x} - \mathbf{c}\|_2 - r &= 0 \\ \|\mathbf{x} - \mathbf{c}\|_2^2 - r^2 &= 0 \\ (\mathbf{x} - \mathbf{c}) \cdot (\mathbf{x} - \mathbf{c}) - r^2 &= 0 \\ \langle (\mathbf{x} - \mathbf{c}), (\mathbf{x} - \mathbf{c}) \rangle - r^2 &= 0 \end{aligned} \quad (21)$$

不同参考资料中圆锥曲线的表达各有不同。本节不厌其烦地罗列圆锥曲线的各种形式解析式目的只有一个，让大家知道这些表达的等价关系，从而对它们不再感到陌生、畏惧。

此外，本书前文一直强调，看到矩阵乘法结果为标量时，要考虑是否能将其写成范数，并从距离角度理解。

为了让大家看到我们熟悉的正圆解析式，进一步展开整理 (20) 得到：

$$\begin{aligned} (\mathbf{x} - \mathbf{c})^T \begin{bmatrix} 1/r^2 & 0 \\ 0 & 1/r^2 \end{bmatrix} (\mathbf{x} - \mathbf{c}) &= \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right)^T \begin{bmatrix} 1/r^2 & 0 \\ 0 & 1/r^2 \end{bmatrix} \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right) \\ &= \begin{bmatrix} x_1 - c_1 & x_2 - c_2 \end{bmatrix} \begin{bmatrix} 1/r^2 & 0 \\ 0 & 1/r^2 \end{bmatrix} \begin{bmatrix} x_1 - c_1 \\ x_2 - c_2 \end{bmatrix} \\ &= \frac{(x_1 - c_1)^2}{r^2} + \frac{(x_2 - c_2)^2}{r^2} = 1 \end{aligned} \quad (22)$$

## 从单位圆到一般正圆

在 (20) 中，大家应该看到了平移。

下面探讨圆心位于原点的单位圆如何一步步经过几何变换得到 (22) 中对应的圆心位于  $\mathbf{c} = [c_1, c_2]^T$  半径为  $r$  的正圆。

平面内，单位圆解析式写成：

$$\mathbf{z}^T \mathbf{z} - 1 = 0 \quad (23)$$

$\mathbf{z}$  通过先等比例缩放，再平移得到  $\mathbf{x}$ ：

$$\mathbf{x} = \underbrace{\begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}}_{\text{Scale}} \mathbf{z} + \underbrace{\mathbf{c}}_{\text{Translate}} \quad (24)$$

整理上式， $\mathbf{z}$  可以写作：

$$\mathbf{z} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}^{-1} (\mathbf{x} - \mathbf{c}) \quad (25)$$

将 (25) 代入 (23)，得到：

$$\left( \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}^{-1} (\mathbf{x} - \mathbf{c}) \right)^T \left( \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}^{-1} (\mathbf{x} - \mathbf{c}) \right) - 1 = 0 \quad (26)$$

整理上式，

$$(\mathbf{x} - \mathbf{c})^T \begin{bmatrix} 1/r & 0 \\ 0 & 1/r \end{bmatrix} \begin{bmatrix} 1/r & 0 \\ 0 & 1/r \end{bmatrix} (\mathbf{x} - \mathbf{c}) - 1 = 0 \quad (27)$$

即，

$$(\mathbf{x} - \mathbf{c})^T \begin{bmatrix} 1/r^2 & 0 \\ 0 & 1/r^2 \end{bmatrix} (\mathbf{x} - \mathbf{c}) - 1 = 0 \quad (28)$$

可以发现 (28) 和 (20) 完全一致。也就是说，如图 1 所示，单位圆可以通过“缩放 + 平移”，得到圆心位于  $\mathbf{c}$  半径为  $r$  的圆。

沿着这一思路，下一节我们讨论如何通过几何变换一步步将单位圆转换成任意旋转椭圆。

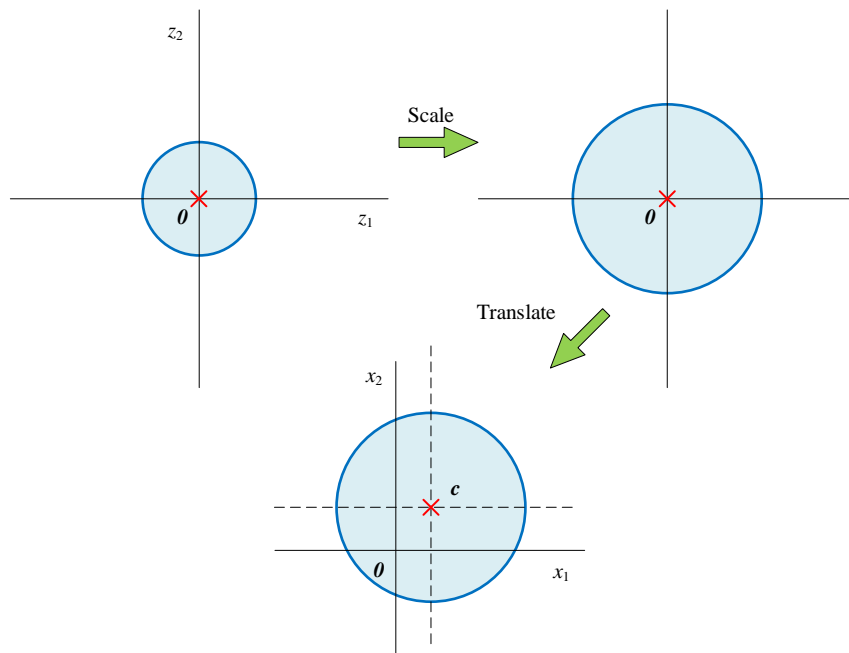


图 1. 单位圆变换得到圆心位于  $\mathbf{c}$  半径为  $r$  的正圆

## 20.3 单位圆到旋转椭圆：缩放 → 旋转 → 平移

这一节介绍如何利用“缩放 → 旋转 → 平移”几何变换，将单位圆变成中心位于任何位置的旋转椭圆。

利用上一节 (23) 给出的单位圆解析式中  $z$ ，对  $z$  先用  $S$  缩放，再通过  $R$  逆时针旋转  $\theta$ ，最后平移  $c$ ，得到  $x$ ：

$$\underbrace{R}_{\text{Rotate}} \underbrace{S}_{\text{Scale}} z + \underbrace{c}_{\text{Translate}} = x \quad (29)$$

其中，

$$R = \underbrace{\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}}_{\text{Rotate}}, \quad S = \underbrace{\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}}_{\text{Scale}}, \quad c = \underbrace{\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}}_{\text{Translate}} \quad (30)$$

如果  $a > b > 0$ ， $a$ 、 $b$  分别为椭圆半长轴、半短轴长度。

从向量空间角度来看，(29) 代表仿射变换；当  $c = \mathbf{0}$  时，不存在平移，(29) 代表线性变换。

将 (30) 代入 (29)，得到：

$$\underbrace{\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}}_{\text{Rotate}} \underbrace{\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}}_{\text{Scale}} \underbrace{\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}}_{\text{Translate}} + \underbrace{\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}}_{\text{Translate}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (31)$$

图 2 所示为从单位圆经过“缩放 → 旋转 → 平移”几何变换得到中心位于  $c$  的旋转椭圆过程。



对这些几何变换感到陌生的读者，请回顾本书第 8 章。



再次强调，单位圆默认圆心位于原点，半径为 1。此外，请大家注意从欧氏距离、等距线、 $L^2$  范数等视角理解正圆。



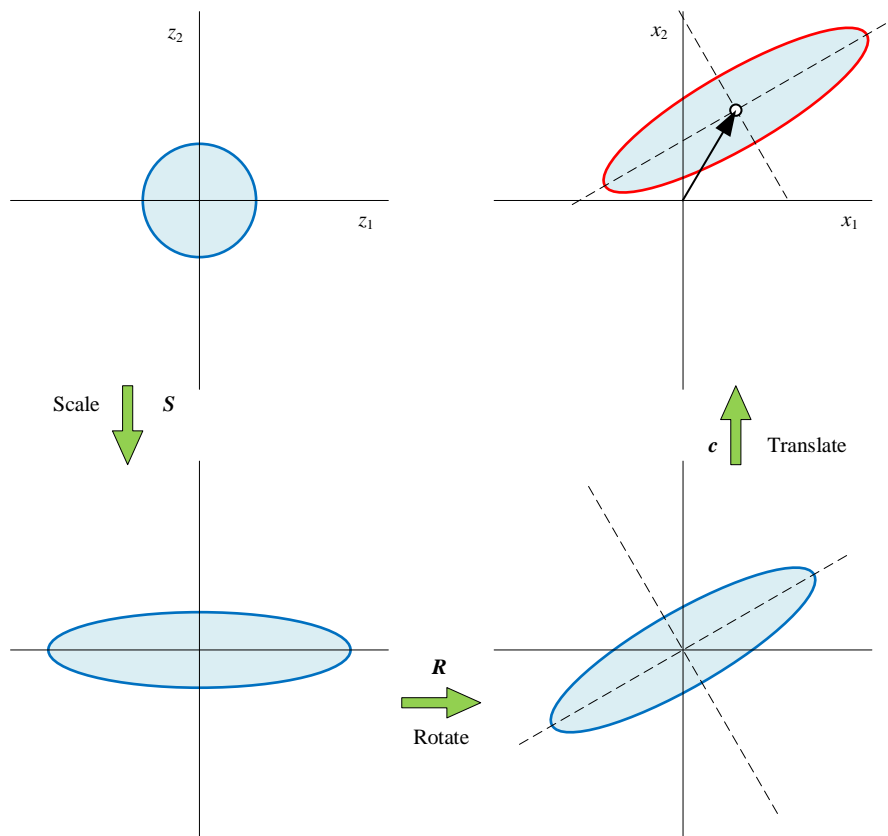


图 2. 从单位圆得到旋转椭圆

整理 (29) 得到  $z$  的解析式为：

$$z = S^{-1} R^{-1} (x - c) \quad (32)$$

$R$  为正交矩阵，所以：

$$R^{-1} = R^T \quad (33)$$

(32) 写成：

$$z = S^{-1} R^T (x - c) \quad (34)$$

反方向来看图 2，中心位于任何位置的旋转椭圆可以通过“平移 → 旋转 → 缩放”变成单位圆。

将 (34) 代入 (23) 正圆解析式，得到：

$$(S^{-1} R^T (x - c))^T (S^{-1} R^T (x - c)) - 1 = 0 \quad (35)$$

进一步整理，得到旋转椭圆解析式：

$$(x - c)^T R S^{-2} R^T (x - c) - 1 = 0 \quad (36)$$

令

$$\boldsymbol{Q} = \boldsymbol{R} \boldsymbol{S}^{-2} \boldsymbol{R}^T = \boldsymbol{R} \begin{bmatrix} a^{-2} & \\ & b^{-2} \end{bmatrix} \boldsymbol{R}^T \quad (37)$$

### 特征值分解

对  $\boldsymbol{Q}$  特征值分解，得到：

$$\boldsymbol{Q} = \boldsymbol{R} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \boldsymbol{R}^T \quad (38)$$

比较 (37) 和 (38)，得出特征值矩阵和缩放矩阵关系：

$$\begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} = \begin{bmatrix} a^{-2} & \\ & b^{-2} \end{bmatrix} \quad (39)$$

即，

$$a = \frac{1}{\sqrt{\lambda_1}}, \quad b = \frac{1}{\sqrt{\lambda_2}} \quad (40)$$

其中， $\lambda_2 > \lambda_1 > 0$ 。

这样，我们便得到椭圆半长轴和半短轴长度和矩阵  $\boldsymbol{Q}$  特征值之间的关系。前文说过，如果  $a > b > 0$ ，椭圆的半长轴长度为  $a$ ，半短轴长度为  $b$ 。而  $a/b$  的比值为：

$$\frac{a}{b} = \frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}} \quad (41)$$

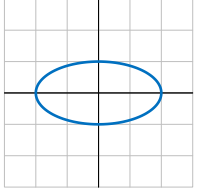
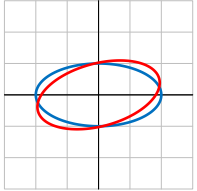
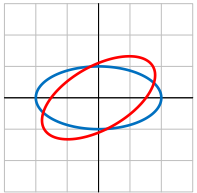
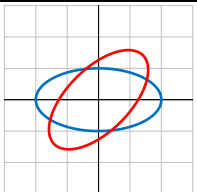
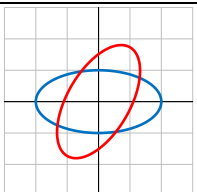
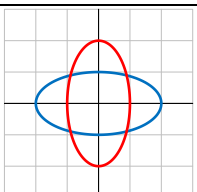
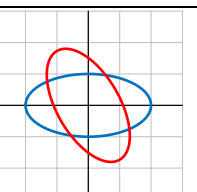
我们在本书第 14 章也讨论过如何用解特征值分解获得椭圆半长轴和半短轴。此外第 14 章还比较“缩放 + 旋转”和“缩放 + 剪切”这两种几何变换路线。

### 只考虑旋转

表 1 中给出一系列不同旋转角度椭圆解析式和对应  $\boldsymbol{Q}$  的特征值分解。表 1 中不同椭圆半长轴和半短轴长度保持一致，唯一变化的就是旋转角度。大家如果对几个不同  $\boldsymbol{Q}$  特征值分解，容易发现它们特征值完全相同，也就是椭圆的半长轴、半短轴长度一致。

表 1. 旋转椭圆解析式、 $\boldsymbol{Q}$  的特征值分解

旋转角度	椭圆解析式 (最多保留小数点后 4 位)	对 $\boldsymbol{Q}$ 特征值分解 (最多保留小数点后 4 位)	图像

$\theta = 0^\circ$ (0)	$\mathbf{x}^T \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x} - 1 = 0$	$\begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	
$\theta = 15^\circ$ ( $\pi/12$ )	$\mathbf{x}^T \begin{bmatrix} 0.3002 & -0.1875 \\ -0.1875 & 0.9498 \end{bmatrix} \mathbf{x} - 1 = 0$	$\begin{bmatrix} 0.3002 & -0.1875 \\ -0.1875 & 0.9498 \end{bmatrix} = \begin{bmatrix} 0.9659 & -0.2588 \\ 0.2588 & 0.9659 \end{bmatrix} \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.9659 & 0.2588 \\ -0.2588 & 0.9659 \end{bmatrix}$	
$\theta = 30^\circ$ ( $\pi/6$ )	$\mathbf{x}^T \begin{bmatrix} 0.4375 & -0.3248 \\ -0.3248 & 0.8125 \end{bmatrix} \mathbf{x} - 1 = 0$	$\begin{bmatrix} 0.4375 & -0.3248 \\ -0.3248 & 0.8125 \end{bmatrix} = \begin{bmatrix} 0.8660 & -0.5000 \\ 0.5000 & 0.8660 \end{bmatrix} \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.8660 & 0.5000 \\ -0.5000 & 0.8660 \end{bmatrix}$	
$\theta = 45^\circ$ ( $\pi/4$ )	$\mathbf{x}^T \begin{bmatrix} 0.6250 & -0.3750 \\ -0.3750 & 0.6250 \end{bmatrix} \mathbf{x} - 1 = 0$	$\begin{bmatrix} 0.6250 & -0.3750 \\ -0.3750 & 0.6250 \end{bmatrix} = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix}$	
$\theta = 60^\circ$ ( $\pi/3$ )	$\mathbf{x}^T \begin{bmatrix} 0.8125 & -0.3248 \\ -0.3248 & 0.4375 \end{bmatrix} \mathbf{x} - 1 = 0$	$\begin{bmatrix} 0.8125 & -0.3248 \\ -0.3248 & 0.4375 \end{bmatrix} = \begin{bmatrix} 0.5000 & -0.8660 \\ 0.8660 & 0.5000 \end{bmatrix} \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5000 & 0.8660 \\ -0.8660 & 0.5000 \end{bmatrix}$	
$\theta = 90^\circ$ ( $\pi/2$ )	$\mathbf{x}^T \begin{bmatrix} 1 & 0 \\ 0 & 0.25 \end{bmatrix} \mathbf{x} - 1 = 0$	$\begin{bmatrix} 1 & 0 \\ 0 & 0.25 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	
$\theta = 120^\circ$ ( $2\pi/3$ )	$\mathbf{x}^T \begin{bmatrix} 0.8125 & 0.3248 \\ 0.3248 & 0.4375 \end{bmatrix} \mathbf{x} - 1 = 0$	$\begin{bmatrix} 0.8125 & 0.3248 \\ 0.3248 & 0.4375 \end{bmatrix} = \begin{bmatrix} -0.5000 & -0.8660 \\ 0.8660 & -0.5000 \end{bmatrix} \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -0.5000 & 0.8660 \\ -0.8660 & -0.5000 \end{bmatrix}$	

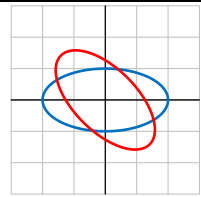
$\theta = 145^\circ$ $(3\pi/4)$	$\mathbf{x}^T \begin{bmatrix} 0.4967 & 0.3524 \\ 0.3524 & 0.7533 \end{bmatrix} \mathbf{x} - 1 = 0$	$\begin{bmatrix} 0.4967 & 0.3524 \\ 0.3524 & 0.7533 \end{bmatrix}$ $= \begin{bmatrix} -0.8192 & -0.5736 \\ 0.5736 & -0.8192 \end{bmatrix} \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -0.8192 & 0.5736 \\ -0.5736 & -0.8192 \end{bmatrix}$	
------------------------------------	--	---	---

图 3 所示为单位圆经过几何变换获得中心位于 (2, 1) 的几个不同旋转角度椭圆的示例。

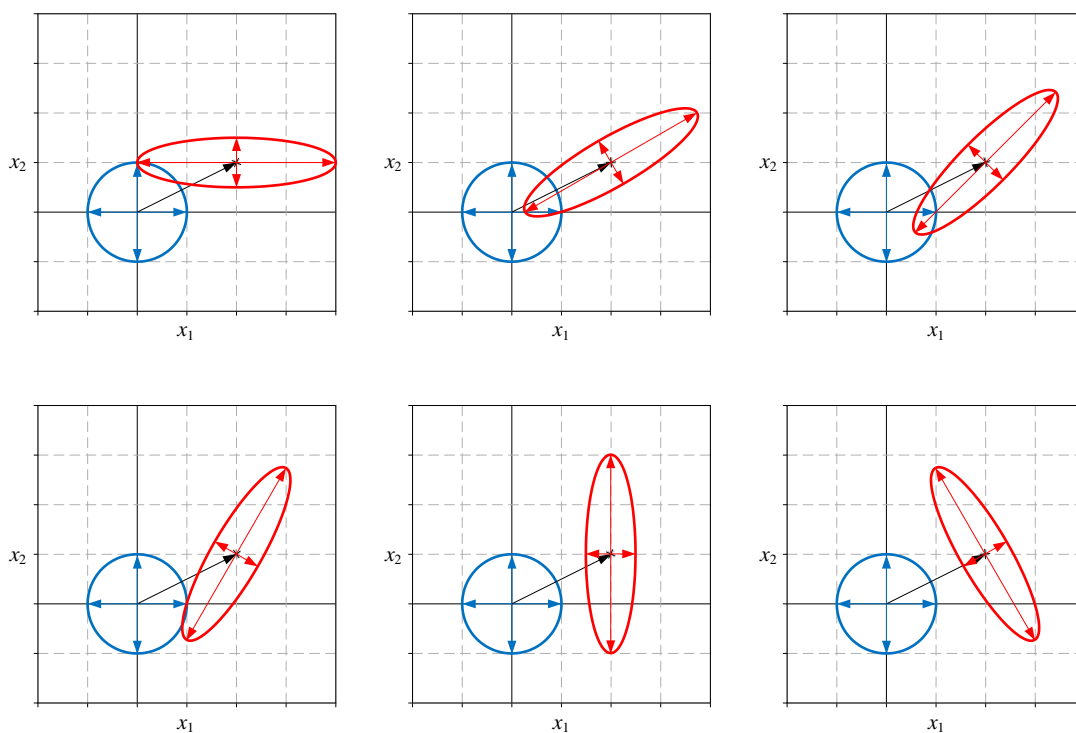


图 3. 通过单位圆获得几个不同的旋转椭圆



Bk4\_Ch20\_01.py 绘制图 3。

### 一般解析式

为了方便整理旋转椭圆解析式，省略 (34) 平移项  $\mathbf{c}$ ，将 (34) 展开得到：

$$\begin{aligned}
 \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} &= \mathbf{S}^{-1} \mathbf{R}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}^{-1} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
 &= \begin{bmatrix} 1/a & 0 \\ 0 & 1/b \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
 &= \begin{bmatrix} \frac{\cos \theta}{a} x_1 + \frac{\sin \theta}{a} x_2 \\ -\frac{\sin \theta}{b} x_1 + \frac{\cos \theta}{b} x_2 \end{bmatrix}
 \end{aligned} \tag{42}$$

将 (42) 代入 (23)，整理得到旋转椭圆解析式：

$$\frac{[x_1 \cos(\theta) + x_2 \sin(\theta)]^2}{a^2} + \frac{[x_1 \sin(\theta) - x_2 \cos(\theta)]^2}{b^2} = 1 \tag{43}$$



(43) 和《数学要素》第 8 章给出的旋转椭圆解析式完全一致。

对比 (36) 和 (43)，相信大家已经体会到用矩阵运算表达椭圆解析式极为简洁。(43) 还仅仅是在二维平面上中心位于原点的椭圆解析式，当中心不在原点，或者维度升高，(43) 这种解析式显然不能胜任描述复杂的椭圆或椭球。更重要的是，借助特征值分解等线性代数工具，(36) 让我们能够分析椭圆或者椭球的几何特点，比如中心位置、长短轴长度、旋转等等。

## 20.4 多元高斯分布：矩阵分解、几何变换、距离

本节介绍如何用上一节介绍的“平移 → 旋转 → 缩放”解剖多元高斯分布。

### 多元高斯分布

多元高斯分布的**概率密度函数** (Probability Density Function, PDF) 解析式如下：

$$f_{\boldsymbol{x}}(\boldsymbol{x}) = \frac{\exp\left(-\frac{1}{2} \overbrace{(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}^{\text{Ellipse}}\right)}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \tag{44}$$

注意，上式中希腊字母  $\boldsymbol{x}$  代表  $D$  维随机变量构成的列向量， $\boldsymbol{x} = [X_1, X_2, \dots, X_D]^T$ 。

相信大家已经在它的分子中看到了旋转椭圆解析式  $(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$ 。这就是为什么很多机器学习算法能够和以椭圆为代表的圆锥曲线扯上关系，因为这些算法中都含有多元高斯分布成分。

本系列丛书会用椭圆等高线描述 (44)。(44) 中  $\boldsymbol{\Sigma}$  的不同形态还会影响到椭圆的形状，如图 4 所示。实际上多元高斯分布 PDF 等高线是多维空间层层包裹的多维椭球面，为了方便展示，我们选择了椭圆这个可视化方案。

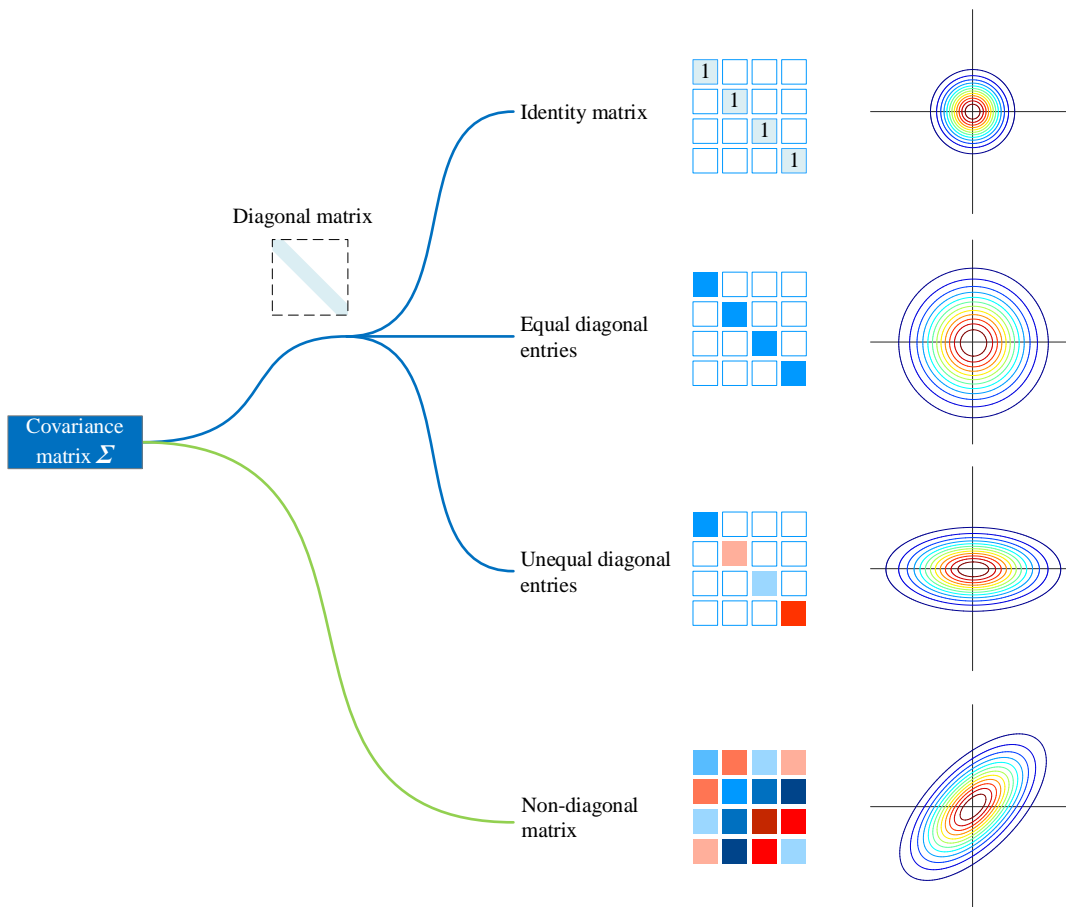


图 4. 协方差矩阵的形态影响高斯密度函数形状

### 特征值分解协方差矩阵

协方差矩阵  $\Sigma$  为对称矩阵，对  $\Sigma$  特征值分解得到：

$$\Sigma = V\Lambda V^T \quad (45)$$

其中， $V$  为正交矩阵。通过上式可以得到对协方差矩阵的逆  $\Sigma^{-1}$  的特征值分解：

$$\Sigma^{-1} = (V\Lambda V^T)^{-1} = (V^T)^{-1} \Lambda^{-1} V^{-1} = V\Lambda^{-1} V^T \quad (46)$$

进一步，将 (46) 代入 (44) 中的椭圆解析式，并整理得到：

$$\begin{aligned} (x - \mu)^T V\Lambda^{-1} V^T (x - \mu) &= (x - \mu)^T V\Lambda^{-\frac{1}{2}} \Lambda^{-\frac{1}{2}} V^T (x - \mu) \\ &= \left[ \Lambda^{-\frac{1}{2}} V^T (x - \mu) \right]^T \Lambda^{-\frac{1}{2}} V^T (x - \mu) \end{aligned} \quad (47)$$

也就是说， $(x - \mu)^T \Sigma^{-1} (x - \mu)$  可以拆成  $\Lambda^{-\frac{1}{2}} V^T (x - \mu)$  的“平方”。

### 平移 → 旋转 → 缩放

大家应该对图 5 这四幅子图并不陌生，我们在本书第 8 章用它们解释过常见几何变换。下面，我们再聊聊它们和多元高斯分布的联系。

几何视角来看，(47) 中， $\Lambda^{-\frac{1}{2}}V^T(x-\mu)$  代表中心在  $\mu$  的旋转椭圆，通过“平移 → 旋转 → 缩放”转换成单位圆的过程。

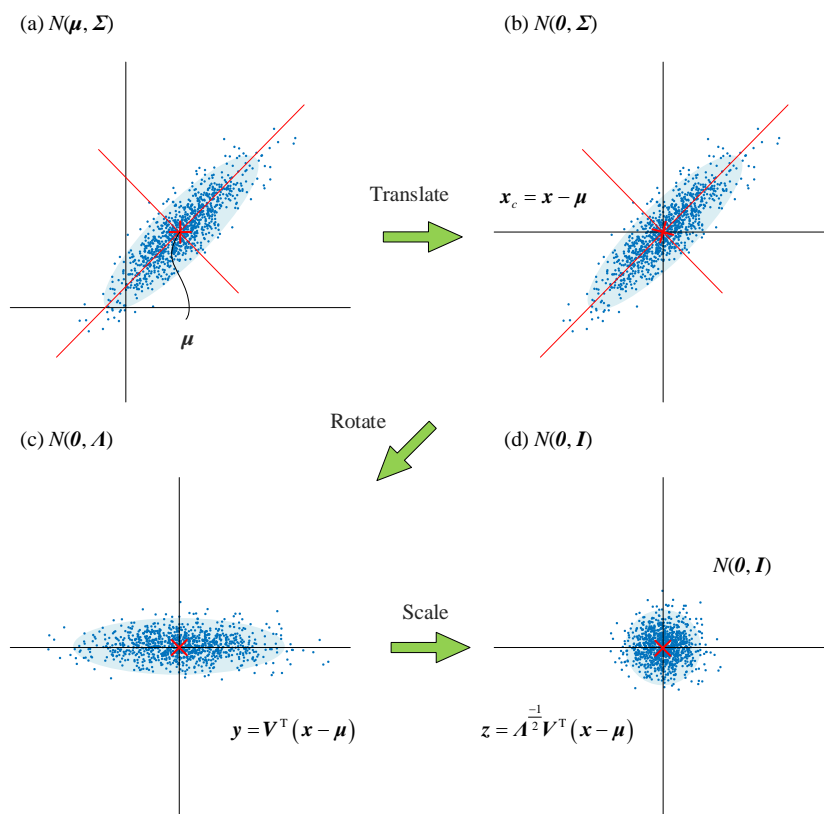


图 5. 平移 → 旋转 → 缩放

从统计视角来思考，图 5 (a) 中旋转椭圆代表多元高斯分布  $N(\mu, \Sigma)$ ，随机数质心位于  $\mu$ ，椭圆形状描述了协方差矩阵  $\Sigma$ 。图 5 (a) 中散点是服从  $N(\mu, \Sigma)$  的随机数。

图 5 (a) 中散点经过平移得到  $x_c = x - \mu$ ，这是一个去均值（中心化过程）。图 5 (b) 中旋转椭圆代表多元高斯分布  $N(0, \Sigma)$ 。随机数质心平移到原点。

图 5 (b) 中椭圆旋转之后得到图 5 (c) 中正椭圆，对应：

$$y = V^T x_c = V^T (x - \mu) \quad (48)$$

正椭圆的半长轴、半短轴长度蕴含在特征值矩阵  $\Lambda$  中。图 5 (c) 中随机数服从  $N(0, \Lambda)$ 。

最后一步是缩放，从图 5 (c) 到图 5 (d)，对应：

$$\mathbf{z} = \mathbf{A}^{-\frac{1}{2}} \mathbf{y} = \mathbf{A}^{-\frac{1}{2}} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \quad (49)$$

图 5 (d) 中单位圆则代表多元高斯分布  $N(\mathbf{0}, \mathbf{I})$ 。

利用向量  $\mathbf{z}$ ，多元高斯分布 PDF 可以写成：

$$f_{\mathbf{z}}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2} \mathbf{z}^T \mathbf{z}\right)}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} = \frac{\exp\left(-\frac{1}{2} \|\mathbf{z}\|_2^2\right)}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \quad (50)$$

$\mathbf{z}$  的模  $\|\mathbf{z}\|$  实际上代表“整体”z 分数。

类比的话，一元高斯分布的概率密度函数可以写成：

$$f(x) = \frac{\exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right)}{(2\pi)^{\frac{1}{2}} \sigma} = \frac{\exp\left(-\frac{1}{2} z^2\right)}{(2\pi)^{\frac{1}{2}} (\sigma^2)^{\frac{1}{2}}} \quad (51)$$

大家应该更容易在上式分子中看到 z 分数的平方。

反向来看， $\mathbf{x} = \mathbf{V} \mathbf{A}^{\frac{1}{2}} \mathbf{z} + \boldsymbol{\mu}$  代表通过“缩放 → 旋转 → 平移”把单位圆转换成中心在  $\boldsymbol{\mu}$  的旋转椭圆。也就是把  $N(\mathbf{0}, \mathbf{I})$  转换成  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 。从数据角度来看，我们可以通过“缩放 → 旋转 → 平移”，把服从  $N(\mathbf{0}, \mathbf{I})$  的随机数转化为服从  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  的随机数。

### 欧氏距离 → 马氏距离

本书前文反复提到，看到  $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$  这种二次型，就要考虑它是否代表某种距离！将  $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$  写成  $L^2$  范数平方形式：

$$\begin{aligned} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) &= \left[ \mathbf{A}^{-\frac{1}{2}} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \right]^T \mathbf{A}^{-\frac{1}{2}} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \\ &= \left\| \mathbf{A}^{-\frac{1}{2}} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \right\|_2^2 \\ &= \|\mathbf{z}\|_2^2 \end{aligned} \quad (52)$$

也就是说， $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$  开方得到：

$$d = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})} = \left\| \mathbf{A}^{-\frac{1}{2}} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \right\| = \|\mathbf{z}\| \quad (53)$$

上式就是大名鼎鼎的马氏距离。马氏距离，也叫**马哈距离** (Mahal distance)，全称**马哈拉诺比斯距离** (Mahalanobis distance)。



马氏距离是机器学习中重要的距离度量。马氏距离的独特之处在于，它通过引入协方差矩阵在计算距离时考虑了数据的分布。此外，马氏距离无量纲量 (unitless 或 dimensionless)，它将各个特征数据标准化。也就是说，马氏距离可以看做是多元数据的  $z$  分数。比如，马氏距离为 2，意味着某点距离数据质心的距离为 2 倍标准差。

比对来看， $(x - \mu)^T (x - \mu)$  代表  $x$  和  $\mu$  两点之间欧氏距离平方。 $\sqrt{(x - \mu)^T (x - \mu)} = \|x - \mu\|$  代表欧氏距离。

在本系列丛书《数学要素》一册第 7 章中，我们知道，地理上的相近，不代表关系的紧密。比如，相隔万里的好友，近在咫尺的路人。马氏距离就是考虑了样本数据“亲疏关系”的距离度量。

### 马氏距离：以鸢尾花为例

为了让大家更好地理解马氏距离，下面我们以鸢尾花数据为例展开讲解。

用鸢尾花花萼长度、花瓣长度两个特征。为了方便，令花萼长度为  $x_1$ ，花瓣长度为  $x_2$ 。二元数据质心所在位置为：

$$\mu = \begin{bmatrix} 5.843 \\ 3.758 \end{bmatrix} \quad (54)$$

图 6 中散点代表鸢尾花样本数据。图 6 对比欧氏距离和马氏距离。

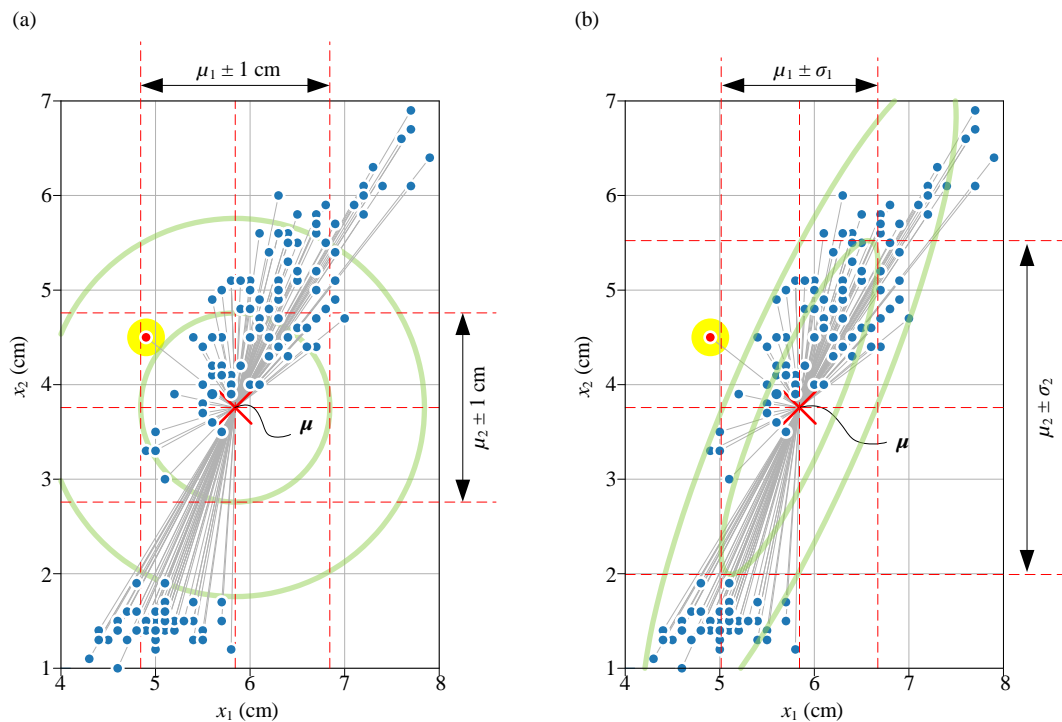


图 6. 欧氏距离和马氏距离

平面上任意一个鸢尾花样本点  $\mathbf{x}$  到质心  $\boldsymbol{\mu}$  的欧氏距离为：

$$\begin{aligned} d &= \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{x} - \boldsymbol{\mu})} = \sqrt{\left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5.843 \\ 3.758 \end{bmatrix} \right)^T \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5.843 \\ 3.758 \end{bmatrix} \right)} \\ &= \sqrt{(x_1 - 5.843)^2 + (x_2 - 3.758)^2} \end{aligned} \quad (55)$$

图 6 (a) 所示的两个同心圆距离质心  $\boldsymbol{\mu}$  为 1 cm 和 2 cm。欧氏距离显然没有考虑数据之间的亲疏关系。举个例子，图 6 (a) 中红色点  $\bullet$  距离质心的欧氏距离略大于 1 cm。但是对于整体样本数据， $\bullet$  显得鹤立鸡群，格格不入。

图 6 中鸢尾花数据协方差矩阵  $\boldsymbol{\Sigma}$  为：

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.685 & 1.274 \\ 1.274 & 3.116 \end{bmatrix} \quad (56)$$

协方差的逆  $\boldsymbol{\Sigma}^{-1}$  为：

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 6.075 & -2.484 \\ -2.484 & 1.336 \end{bmatrix} \quad (57)$$

代入具体值，图 6 (b) 的马氏距离解析式为：

$$\begin{aligned} d &= \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \begin{bmatrix} 6.075 & -2.484 \\ -2.484 & 1.336 \end{bmatrix} (\mathbf{x} - \boldsymbol{\mu})} \\ &= \sqrt{\left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5.843 \\ 3.758 \end{bmatrix} \right)^T \begin{bmatrix} 6.075 & -2.484 \\ -2.484 & 1.336 \end{bmatrix} \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5.843 \\ 3.758 \end{bmatrix} \right)} \\ &= \sqrt{6.08x_1^2 - 4.97x_1x_2 + 1.34x_2^2 - 52.32x_1 + 18.99x_2 + 117.21} \end{aligned} \quad (58)$$

图 6 (b) 中两个椭圆就是马氏距离  $d = 1$  和  $d = 2$  时对应的等高线。再次强调，马氏距离没有单位，它相当于 z 分数。准确地说，马氏距离的单位是“标准差”。

再看图 6 (b) 中红色点  $\bullet$ ，它的马氏距离远大于 2。也就是说，考虑整体数据分布亲疏情况，红色点  $\bullet$  离样本数据“远的多”。

显然，相比欧氏距离，在度量数据之间亲疏关系上，马氏距离更胜任。

我们可以用 `scipy.spatial.distance.mahalanobis()` 函数计算马氏距离，Scikit-Learn 库中也有计算马氏距离的函数。



本系列丛书会在《概率统计》一册有一章专门讲解马氏距离，我们会继续鸢尾花这个例子。在《数据科学》一册，我们还会用马氏距离判断离群点。

## 高斯函数

将 (53) 中马氏距离  $d$  代入多元高斯分布概率密度函数，得到：

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}d^2\right)}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \quad (59)$$

上式，我们看到高斯函数  $\exp(-1/2 \bullet)$  把“距离度量”转化成“亲近度”。如图 7 所示，从几何角度来看，这是一个二次曲面到高斯函数曲面的转换。

从统计角度来看，距离中心  $\boldsymbol{\mu}$  越远，对应的概率越小。概率密度值可以无限接近 0，但是不为 0，这说明虽然是小概率事件，但是“万事皆可能”。强调一下，概率密度不同于概率值，概率密度函数积分或多重积分之后可以得到概率值。

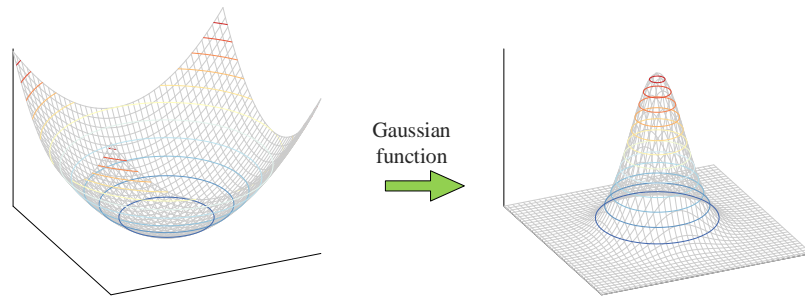


图 7. “距离度量”转化成“亲近度”

## 分母：行列式值

从体积角度来看，“平移 → 旋转 → 缩放”这一系列几何变换带来的面积/体积缩放系数和 (44) 分母中的  $|\boldsymbol{\Sigma}|^{\frac{1}{2}}$  有直接关系。

把系数  $|\boldsymbol{\Sigma}|^{\frac{1}{2}}$  从 (44) 分母移到分子可以写成  $|\boldsymbol{\Sigma}|^{-\frac{1}{2}}$ 。而  $\boldsymbol{\Sigma}^{-\frac{1}{2}}$  相当于：

$$\boldsymbol{\Sigma}^{-\frac{1}{2}} \sim \boldsymbol{A}^{-\frac{1}{2}} \boldsymbol{V}^T (\mathbf{x} - \boldsymbol{\mu}) \quad (60)$$

本书第 5 章和第 14 章都强调过， $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{B}\boldsymbol{B}^T$ ，不能推导得到  $\boldsymbol{A} = \boldsymbol{B}$ 。

$|\boldsymbol{\Sigma}|^{-\frac{1}{2}}$  开根号的原因很容易理解， $\boldsymbol{\Sigma}^{-1}$  中有“两份”上述“平移 → 旋转 → 缩放”几何变换，因此  $|\boldsymbol{\Sigma}^{-1}|$  代表缩放比例的平方。

注意，协方差矩阵  $\boldsymbol{\Sigma}$  真正起到缩放的成分是特征值矩阵  $\boldsymbol{A}$ ，即  $|\boldsymbol{\Sigma}| = |\boldsymbol{A}|$ 。强调一下， $|\bullet|$  这个运算符是求行列式值，不是绝对值。行列式值完成“矩阵 → 标量”运算，这个标量结果对应面积/体积缩放系数。

从统计角度来看，对比 (44) 和 (51)， $|\boldsymbol{\Sigma}|$  相当于方差。

## 分母：体积归一化

如图 8 所示，从几何角度来看，(44) 分母中  $(2\pi)^{\frac{D}{2}}$  一项起到归一化作用，为了保证概率密度函数曲面和整个水平面包裹的体积为 1，即概率为 1。

同理，(51) 分母中  $(2\pi)^{\frac{1}{2}}$  用来保证  $f(x)$  和整条横轴围成图像面积为 1。

再次强调，概率密度经过积分或多重积分可以得到概率。本系列丛书《数学要素》第 18 章介绍过一元高斯函数积分、二元高斯函数“偏积分”和二重积分，建议大家回顾。

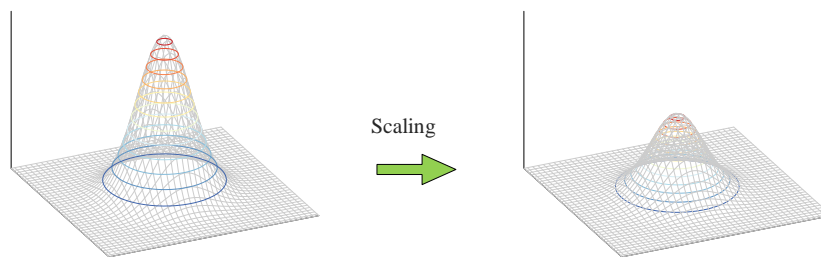


图 8. 体积归一化

## 解剖多元高斯分布 PDF

有了以上分析，理解、记忆多元高斯分布的概率密度函数解析式，就变得格外容易了：

$$\begin{aligned}
 d &= \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} && \text{Mahal distance} \\
 \|z\| &&& \text{z-score} \\
 z &= A^{\frac{-1}{2}} V^T (x - \mu) && \text{Translate} \rightarrow \text{rotate} \rightarrow \text{scale} \\
 \left[ A^{\frac{-1}{2}} V^T (x - \mu) \right]^T A^{\frac{-1}{2}} V^T (x - \mu) &&& \text{Eigen decomposition} \\
 (x - \mu)^T \Sigma^{-1} (x - \mu) &&& \text{Ellipse/ellipsoid} \\
 f_x(x) &= \frac{\exp\left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right)}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \\
 &\downarrow \text{Distance} \rightarrow \text{similarity} && \downarrow \text{Normalization} \\
 &&& \text{Multivariable calculus} \\
 &&& \downarrow \text{Scaling} \\
 &&& \text{Eigenvalues}
 \end{aligned}
 \tag{61}$$

有关多元高斯分布的故事才刚刚开始，本系列丛书《概率统计》中多元高斯分布占据大半江山。

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)



我们用 Streamlit 和 Plotly 绘制二元高斯分布概率密度函数曲面、平面等高线，大家可以调节均方差、相关性系数来观察图像变化。请参考 Streamlit\_Bk4\_Ch20\_04.py。

## 20.5 从单位双曲线到旋转双曲线

本节讲解如何把单位双曲线变换得到任意双曲线。平面上，**单位双曲线** (unit hyperbola) 定义如下：

$$\mathbf{z}^T \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{z} - 1 = 0 \quad (62)$$

➡ (44) 对单位双曲线感到陌生的读者请看到本系列丛书《数学要素》第 9 章。同时也建议大家回顾这章中讲解的双曲函数。

展开 (62) 得到：

$$z_1^2 - z_2^2 = 1 \quad (63)$$

和前文思路完全一致，首先对  $\mathbf{z}$  通过  $\mathbf{S}$  缩放，再通过  $\mathbf{R}$  逆时针旋转  $\theta$ ，最后平移  $\mathbf{c}$ ：

$$\underbrace{\mathbf{R}}_{\text{Rotate}} \underbrace{\mathbf{S}}_{\text{Scale}} \mathbf{z} + \underbrace{\mathbf{c}}_{\text{Translate}} = \mathbf{x} \quad (64)$$

同样展开得到：

$$\underbrace{\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}}_{\text{Rotate}} \underbrace{\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}}_{\text{Scale}} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \underbrace{\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}}_{\text{Translate}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (65)$$

后续推导和上一节完全一致，我们可以得到任意双曲线的解析式。鉴于我们已经放弃代数解析式表达复杂圆锥曲线，因此不建议大家展开推导。图 9 所示为通过单位双曲线旋转得到的一系列双曲线。图 9 中蓝色和红色双曲线仅仅存在旋转关系，没有经过缩放和平移操作。

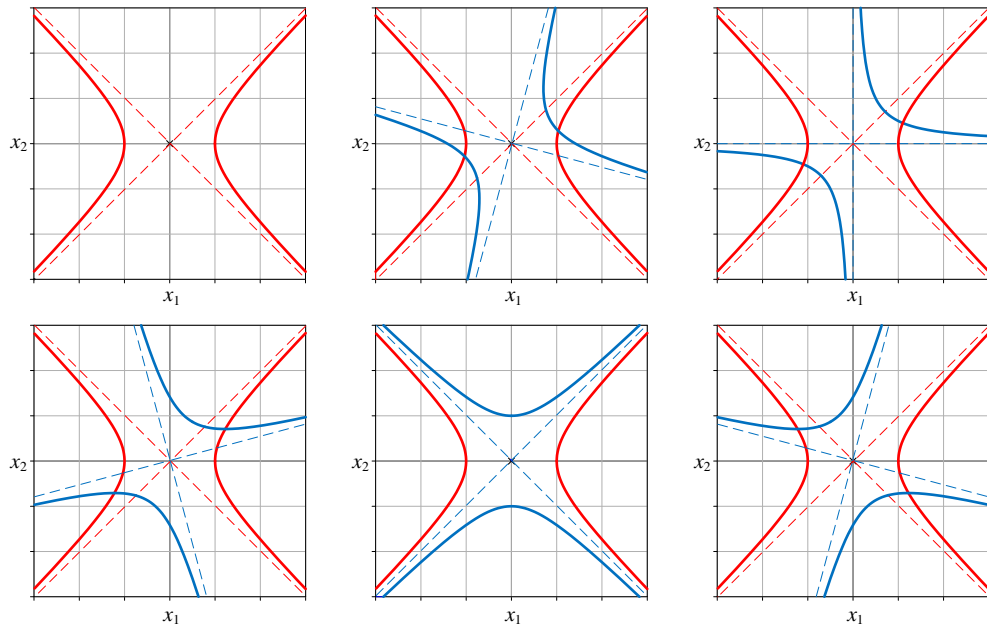
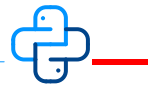


图 9. 通过单位双曲线旋转得到的一系列双曲线



请大家自行修改 `Bk4_Ch20_02.py` 参数绘制不同几何变换条件下双曲线。

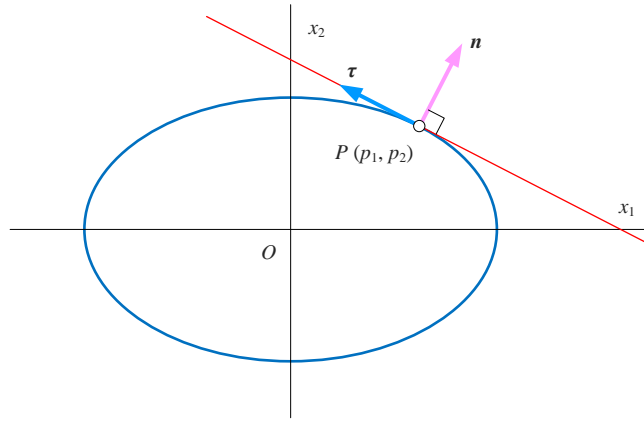
## 20.6 切线：构造函数，求梯度向量

本节探讨如何求解圆锥曲线切线解析式。

### 椭圆

首先以椭圆为例求解其切线解析式。标准椭圆解析式为：

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1 \quad (66)$$

图 10. 椭圆上点  $P$  处切向量和法向量

先构造一个二元函数  $f(x_1, x_2)$ ，如下：

$$f(x_1, x_2) = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \quad (67)$$

如图 10 所示，椭圆上  $P(p_1, p_2)$  一点处  $f(x_1, x_2)$  梯度，即法向量  $\mathbf{n}$  为：

$$\mathbf{n} = \nabla f(\mathbf{x})|_{(p_1, p_2)} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} \bigg|_{(p_1, p_2)} = \begin{bmatrix} \frac{2p_1}{a^2} \\ \frac{2p_2}{b^2} \end{bmatrix} \quad (68)$$

如图 10 所示，切线上任意一点和点  $P$  构成向量，垂直于法向量  $\mathbf{n}$ ，因此两者内积为 0，即，

$$\begin{bmatrix} \frac{2p_1}{a^2} \\ \frac{2p_2}{b^2} \end{bmatrix} \cdot \begin{bmatrix} x_1 - p_1 \\ x_2 - p_2 \end{bmatrix} = \frac{2p_1}{a^2}(x_1 - p_1) + \frac{2p_2}{b^2}(x_2 - p_2) = 0 \quad (69)$$

整理上式，得到  $P(p_1, p_2)$  点处椭圆切线解析式：

$$\frac{p_1}{a^2}x_1 + \frac{p_2}{b^2}x_2 = \frac{p_1^2}{a^2} + \frac{p_2^2}{b^2} \quad (70)$$

图 11 所示为某个给定椭圆上不同点切线。



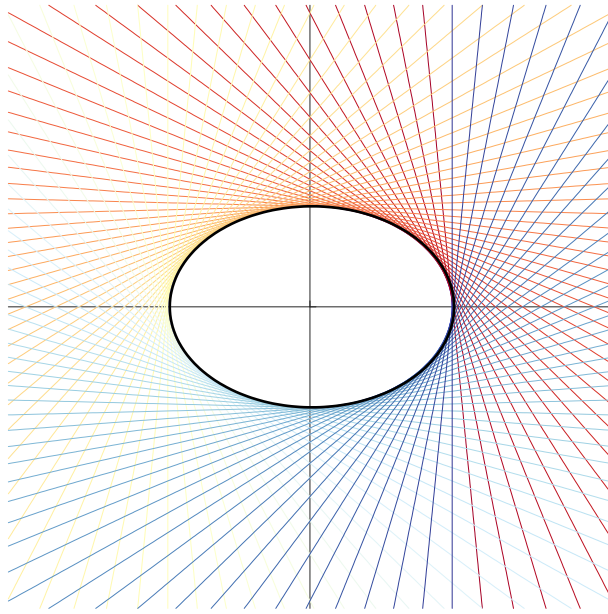


图 11. 椭圆切线分布

### 正圆

正圆是椭圆特殊形式，将  $a = b = r$  带入上式，可获得圆心位于原点的正圆上  $P(p_1, p_2)$  点切线解析式：

$$p_1x_1 + p_2x_2 = p_1^2 + p_2^2 = r^2 \quad (71)$$

图 12 所示为中心位于原点单位圆不同点上切线。

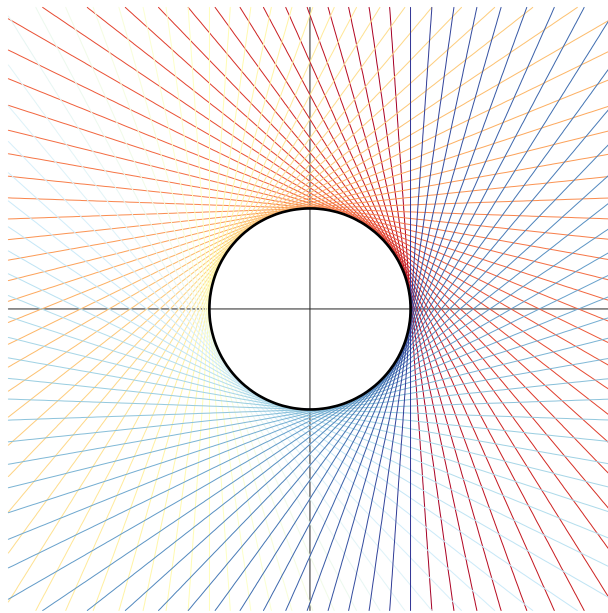


图 12. 单位圆切线分布



## 双曲线

同样的方法可以求解标准双曲线切线。焦点位于横轴标准椭圆解析式写作：

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1 \quad (72)$$

类似地，先构造一个二元函数  $f(x_1, x_2)$ ，如下：

$$f(x_1, x_2) = \frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} \quad (73)$$

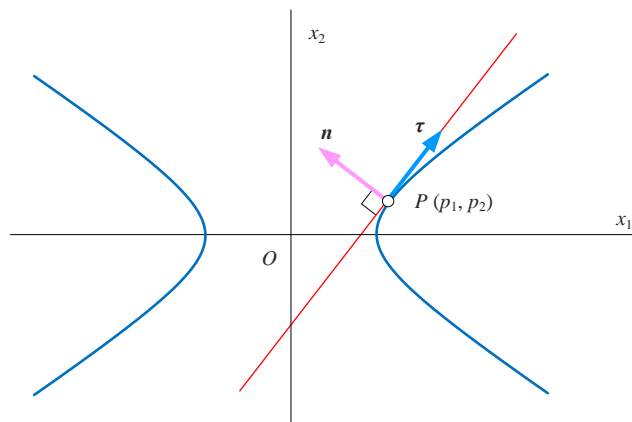


图 13. 双曲线上点  $P$  处切向量和法向量

如图 13 所示，双曲线上  $P(p_1, p_2)$  点处函数  $f(x_1, x_2)$  梯度，即法向量  $\mathbf{n}$  为：

$$\mathbf{n} = \nabla f(\mathbf{x}) \Big|_{(p_1, p_2)} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} \Big|_{(p_1, p_2)} = \begin{bmatrix} \frac{2p_1}{a^2} \\ -\frac{2p_2}{b^2} \end{bmatrix} \quad (74)$$

如图 13 所示，切线上任意一点和点  $P$  构成向量，垂直于法向量  $\mathbf{n}$ ，通过内积为 0 得到以下等式：

$$\begin{bmatrix} \frac{2p_1}{a^2} \\ -\frac{2p_2}{b^2} \end{bmatrix} \cdot \begin{bmatrix} x_1 - p_1 \\ x_2 - p_2 \end{bmatrix} = \frac{2p_1}{a^2}(x_1 - p_1) - \frac{2p_2}{b^2}(x_2 - p_2) = 0 \quad (75)$$

整理上式，得到双曲线上  $P(p_1, p_2)$  点处切线解析式：

$$\frac{p_1}{a^2}x_1 - \frac{p_2}{b^2}x_2 = \frac{p_1^2}{a^2} - \frac{p_2^2}{b^2} \quad (76)$$

图 14 展示单位双曲线不同点切线。

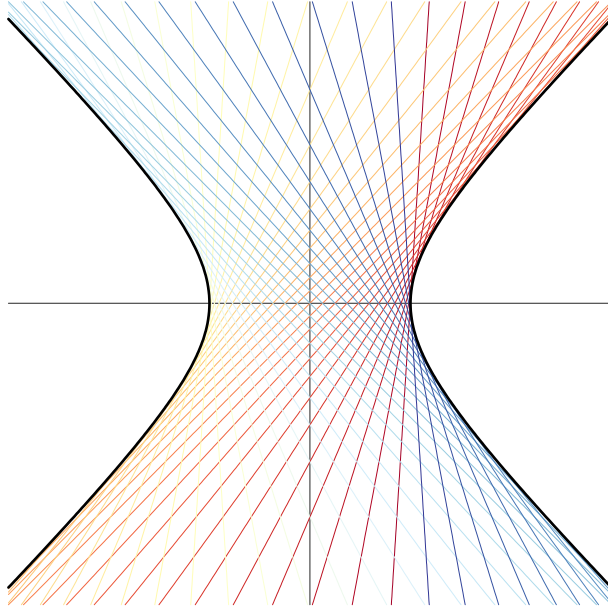


图 14. 双曲线左右两侧切线分布

### 圆锥曲线一般式

本章前文给出圆锥曲线常见一般表达式，同样据此构造一个二元函数  $f(x_1, x_2)$ ：

$$f(x_1, x_2) = Ax_1^2 + Bx_1x_2 + Cx_2^2 + Dx_1 + Ex_2 + F \quad (77)$$

圆锥曲线任意一点  $P(p_1, p_2)$  处二元函数  $f(x_1, x_2)$  梯度，即法向量  $\mathbf{n}$  为：

$$\mathbf{n} = \nabla f(\mathbf{x})|_{(p_1, p_2)} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}|_{(p_1, p_2)} = \begin{bmatrix} 2Ap_1 + Bp_2 + D \\ Bp_1 + 2Cp_2 + E \end{bmatrix} \quad (78)$$

切线上任意一点和点  $P$  构成向量，垂直于法向量  $\mathbf{n}$ ，因此两者向量内积为 0：

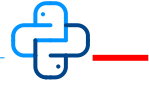
$$\begin{bmatrix} 2Ap_1 + Bp_2 + D \\ Bp_1 + 2Cp_2 + E \end{bmatrix} \cdot \begin{bmatrix} x_1 - p_1 \\ x_2 - p_2 \end{bmatrix} = 0 \quad (79)$$

即

$$(2Ap_1 + Bp_2 + D)(x_1 - p_1) + (Bp_1 + 2Cp_2 + E)(x_2 - p_2) = 0 \quad (80)$$

整理得到圆锥曲线任意一点  $P(p_1, p_2)$  处切线解析式：

$$(2Ap_1 + Bp_2 + D)x_1 + (Bp_1 + 2Cp_2 + E)x_2 = 2Ap_1^2 + 2Bp_1p_2 + 2Cp_2^2 + Dp_1 + Ep_2 \quad (81)$$



Bk4\_Ch20\_03.py 绘制图 11，请大家修改代码自行绘制本节和下一节其他图像。

## 20.7 法线：法向量垂直于切向量

### 椭圆

(67) 所示标准椭圆上  $P(p_1, p_2)$  一点处切向量  $\tau$  为：

$$\tau = \begin{bmatrix} \frac{\partial f}{\partial x_2} \\ -\frac{\partial f}{\partial x_1} \end{bmatrix}_{(p_1, p_2)} = \begin{bmatrix} \frac{2p_2}{b^2} \\ -\frac{2p_1}{a^2} \end{bmatrix} \quad (82)$$

$P(p_1, p_2)$  处切向量  $\tau$  显然垂直其法向量  $n$ 。

椭圆上  $P(p_1, p_2)$  点处法线解析式：

$$\begin{bmatrix} \frac{2p_2}{b^2} \\ -\frac{2p_1}{a^2} \end{bmatrix} \cdot \begin{bmatrix} x_1 - p_1 \\ x_2 - p_2 \end{bmatrix} = \frac{2p_2}{b^2}(x_1 - p_1) - \frac{2p_1}{a^2}(x_2 - p_2) = 0 \quad (83)$$

整理得到：

$$\frac{p_2}{b^2}x_1 - \frac{p_1}{a^2}x_2 = \frac{p_1p_2}{b^2} - \frac{p_1p_2}{a^2} \quad (84)$$

图 15 所示为椭圆法线分布情况。

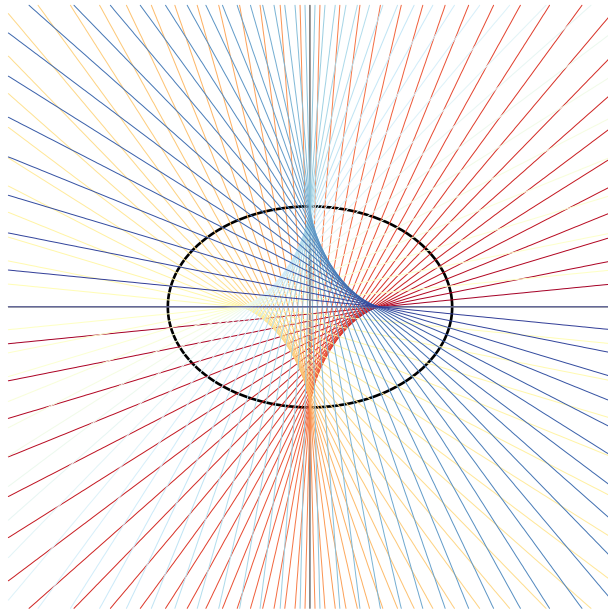


图 15. 椭圆法线分布

### 圆锥曲线一般式

下面推导一般圆锥曲线的法线。(77) 圆锥曲线解析式上  $P$  点处切向量  $\tau$  为：

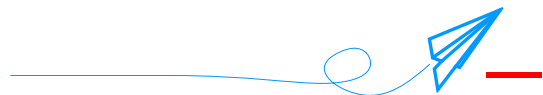
$$\tau = \left[ \begin{array}{c} \frac{\partial f}{\partial x_2} \\ -\frac{\partial f}{\partial x_1} \end{array} \right]_{(p_1, p_2)} = \left[ \begin{array}{c} Bp_1 + 2Cp_2 + E \\ -(2Ap_1 + Bp_2 + D) \end{array} \right] \quad (85)$$

得到过  $P$  点圆锥曲线法线直线方程，如下：

$$\left[ \begin{array}{c} Bp_1 + 2Cp_2 + E \\ -(2Ap_1 + Bp_2 + D) \end{array} \right] \cdot \left[ \begin{array}{c} x_1 - p_1 \\ x_2 - p_2 \end{array} \right] = 0 \quad (86)$$

整理得到如下法线解析式：

$$(Bp_1 + 2Cp_2 + E)x_1 - (2Ap_1 + Bp_2 + D)x_2 = B(p_1^2 - p_2^2) + (2C - 2A)p_1p_2 + Ep_1 - Dp_2 \quad (87)$$



下面这幅图最能总结本章的核心内容。它虽然是四幅子图，却代表着一个连贯的几何变换操作。不管是从旋转椭圆到单位圆，还是从单位圆到旋转椭圆，请大家务必记住每步几何变换对应的线性代数运算。

理解这些几何变换对于理解协方差矩阵、多元高斯分布、主成分分析和很多机器学习算法有至关重要的作用。

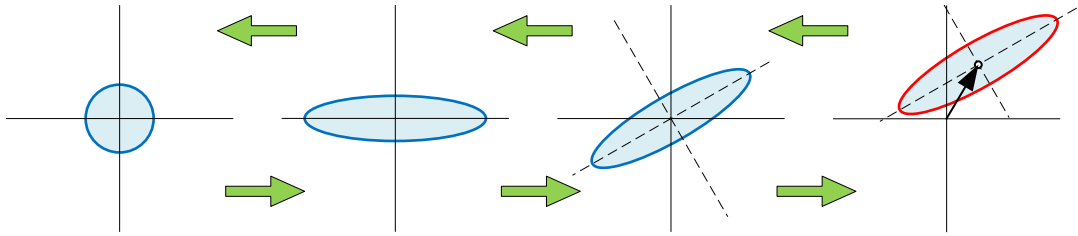


图 16. 总结本章重要内容的四幅图

本章利用矩阵分解、几何视角解剖了多元高斯分布的概率密度函数。请大家特别注意理解“平移 → 旋转 → 缩放”这三步几何操作，以及马氏距离的意义。