

# 11

## Multivariate Gaussian Distribution

# 多元高斯分布

几何、代数、概率统计的完美结合



在我看来，数学科学是一个不可分割的有机体，其生命力取决于各部分的联系。

***Mathematical science is in my opinion an indivisible whole, an organism whose vitality is conditioned upon the connection of its parts.***

—— 大卫·希尔伯特 (David Hilbert) | 德国数学家 | 1862 ~ 1943



- ▶ `numpy.cov()` 计算协方差矩阵
- ▶ `numpy.diag()` 如果 A 为方阵，`numpy.diag(A)` 函数提取对角线元素，以向量形式输入结果；如果 a 为向量，`numpy.diag(a)` 函数将向量展开成方阵，方阵对角线元素为 a 向量元素
- ▶ `numpy.linalg.eig()` 特征值分解
- ▶ `numpy.linalg.inv()` 计算逆矩阵
- ▶ `numpy.linalg.norm()` 计算范数
- ▶ `numpy.linalg.svd()` 奇异值分解
- ▶ `scipy.spatial.distance.euclidean()` 计算欧氏距离
- ▶ `scipy.spatial.distance.mahalanobis()` 计算马氏距离
- ▶ `seaborn.heatmap()` 绘制热图
- ▶ `seaborn.kdeplot()` 绘制 KDE 核概率密度估计曲线
- ▶ `seaborn.pairplot()` 绘制成对分析图
- ▶ `sklearn.decomposition.PCA()` 主成分分析函数

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)



## 11.1 矩阵角度：一元、二元、三元到多元

### 一元

本书第 9 章讲解了一元高斯分布的 PDF 解析式，具体如下：

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

图 1 (a) 所示为一元高斯分布 PDF 的图像。

### 二元

第 10 章中，我们看到二元高斯分布的 PDF 解析式：

$$f_{x,y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{x,y}^2}} \times \exp\left(\underbrace{-\frac{1}{2}\frac{1}{(1-\rho_{x,y}^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{x,y}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)}_{\text{Ellipse}}\right) \quad (2)$$

图 1 (b) 所示为二元高斯分布 PDF 的图像。

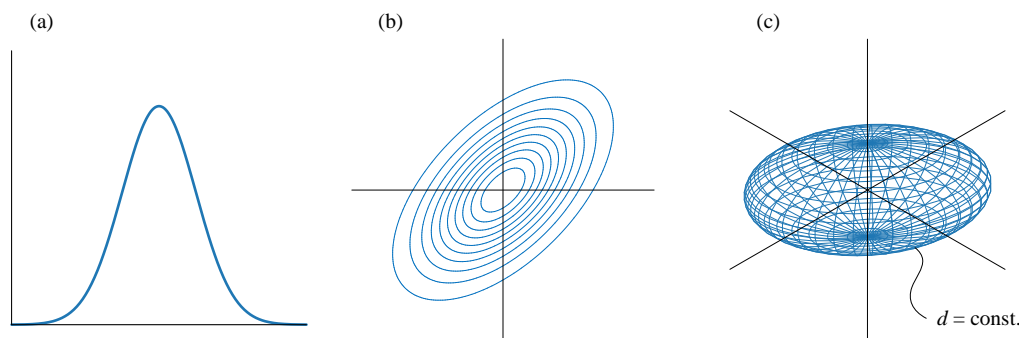


图 1. 一元、二元、三元高斯分布的几何形态

### 三元

(2) 已经很复杂，我们再看看三元高斯分布 PDF 解析式。在  $\sigma_1 = \sigma_2 = \sigma_3 = 1$ ,  $\mu_1 = \mu_2 = \mu_3 = 0$  条件下，三元高斯分布 PDF 解析式如下：

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \frac{\exp\left(\frac{-1}{2}d^2\right)}{(2\pi)^{\frac{3}{2}}\sqrt{1+2\rho_{1,2}\rho_{1,3}\rho_{2,3}-(\rho_{1,2}^2+\rho_{1,3}^2+\rho_{2,3}^2)}} \quad (3)$$

其中

$$d^2 = \frac{x_1^2(\rho_{2,3}^2-1)+x_2^2(\rho_{1,3}^2-1)+x_3^2(\rho_{1,2}^2-1)+2[x_1x_2(\rho_{1,2}-\rho_{1,3}\rho_{2,3})+x_1x_3(\rho_{1,3}-\rho_{1,2}\rho_{2,3})+x_2x_3(\rho_{2,3}-\rho_{1,3}\rho_{2,3})]}{(\rho_{1,2}^2+\rho_{1,3}^2+\rho_{2,3}^2-2\rho_{1,2}\rho_{1,3}\rho_{2,3}-1)} \quad (4)$$

当  $d$  为确定值时，上式代表一个**椭球** (ellipsoid)，如图 1 (c) 所示。也就是说三元高斯分布 PDF 的几何图形是嵌套的椭球。

相信大家已经看到了三元高斯分布 PDF 解析式的复杂程度。更不用说，(3) 的解析式是在  $\sigma_1 = \sigma_2 = \sigma_3 = 1$ ， $\mu_1 = \mu_2 = \mu_3 = 0$  这个极特殊条件下获得的。

到了四元、五元、更高元高斯分布 PDF 解析式时，代数展开式已经完全不够用了。因此，对于多元高斯分布，我们需要矩阵算式。

## 多元

“鸢尾花书”读者应该已经很熟悉多元正态分布 PDF，具体如下：

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \quad (5)$$

其中， $\boldsymbol{\chi}$ 、 $\mathbf{x}$ 、 $\boldsymbol{\mu}$  均为列向量：

$$\boldsymbol{\chi} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_D \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_D \end{bmatrix} \quad (6)$$

向量  $\boldsymbol{\mu}$  常常被称作**质心** (centroid)， $D$  为高斯分布的特征数，比如二元高斯分布  $D=2$ 。

协方差矩阵  $\boldsymbol{\Sigma}$  为：

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,D} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{D,1} & \sigma_{D,2} & \cdots & \sigma_{D,D} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,D}\sigma_1\sigma_D \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2,D}\sigma_2\sigma_D \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,D}\sigma_1\sigma_D & \rho_{2,D}\sigma_2\sigma_D & \cdots & \sigma_D^2 \end{bmatrix} \quad (7)$$

⚠ 特别需要大家注意的是，如果 (5) 成立，协方差矩阵  $\boldsymbol{\Sigma}$  必须为正定矩阵。如果为  $\boldsymbol{\Sigma}$  半正定， $\boldsymbol{\Sigma}$  的行列式值为 0，而 (5) 分母不能为 0。 $\boldsymbol{\Sigma}$  半正定说明  $\boldsymbol{\chi}$  存在线性相关。

一组随机变量构成的列向量  $\boldsymbol{\chi}$  服从如 (5) 多元高斯分布，记做：

$$\boldsymbol{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_D \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_D \end{bmatrix}, \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,D} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{D,1} & \sigma_{D,2} & \cdots & \sigma_{D,D} \end{bmatrix} \right) \quad (8)$$

或更简便地记做：

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (9)$$

再次强调，这个语境下， $\boldsymbol{x}$  为随机变量构成的列向量，每一行代表一个随机变量；而  $\boldsymbol{X}$  代表数据矩阵，每一列对应一个随机变量所有样本。

### 多元 → 一元

$D = 1$  时，质心为：

$$\boldsymbol{\mu} = [\mu] \quad (10)$$

协方差矩阵为：

$$\boldsymbol{\Sigma} = [\sigma^2] \quad (11)$$

(5) 分子中的二次型 (quadratic form) 可以展开为：

$$(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = (x - \mu) \sigma^{-2} (x - \mu) = \left( \frac{x - \mu}{\sigma} \right)^2 \quad (12)$$

我们看到的是 Z 分数的平方。这和 (1) 解析式完全一致。

### 多元 → 二元

再以二元 ( $D = 2$ ) 高斯分布为例，它的质心：

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad (13)$$

二元高斯分布的协方差矩阵  $\boldsymbol{\Sigma}$  具体为：

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{1,2} \sigma_1 \sigma_2 \\ \rho_{1,2} \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix} \quad (14)$$

协方差矩阵的行列式值  $|\boldsymbol{\Sigma}|$ ：

$$|\boldsymbol{\Sigma}| = \sigma_1^2 \sigma_2^2 - \rho_{1,2}^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 (1 - \rho_{1,2}^2) \quad (15)$$

再次强调，如果相关性系数为 $\pm 1$ ，行列式值 $|\Sigma|$ 为 0。相关性系数取值范围为 $(-1, 1)$ 时，协方差矩阵的逆 $\Sigma^{-1}$ 为：

$$\Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho_{1,2}^2)} \begin{bmatrix} \sigma_2^2 & -\rho_{1,2} \sigma_1 \sigma_2 \\ -\rho_{1,2} \sigma_1 \sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{1 - \rho_{1,2}^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho_{1,2}}{\sigma_1 \sigma_2} \\ \frac{-\rho_{1,2}}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix} \quad (16)$$

对于二元高斯分布，(5) 分子中的二次型展开：

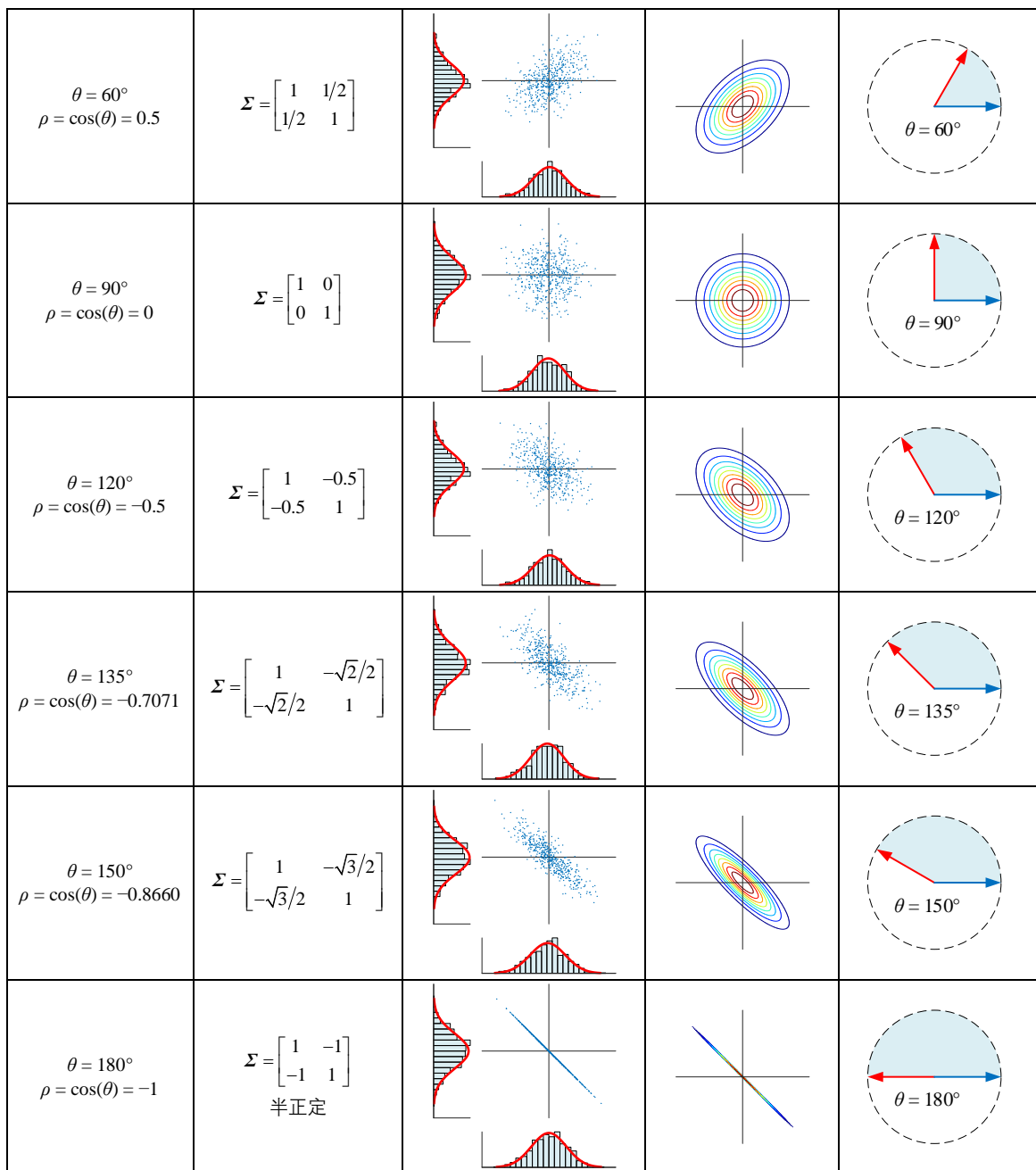
$$\begin{aligned} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) &= \begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{bmatrix} \frac{1}{1 - \rho_{1,2}^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho_{1,2}}{\sigma_1 \sigma_2} \\ \frac{-\rho_{1,2}}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \\ &= \frac{1}{1 - \rho_{1,2}^2} \left[ \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho_{1,2} \left( \frac{x_1 - \mu_1}{\sigma_1} \right) \left( \frac{x_2 - \mu_2}{\sigma_2} \right) + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \end{aligned} \quad (17)$$

分别将 (17) 和 (15) 代入 (5) 可以得到二元高斯分布 PDF 解析式。

表 1 所示为不同线性相关系数的可视化方案。

表 1. 不同相关性系数的可视化方案

相关性系数	协方差矩阵	散点图	PDF 等高线	向量
$\theta = 0^\circ$ $\rho = \cos(\theta) = 1$	$\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ 半正定			
$\theta = 30^\circ$ $\rho = \cos(\theta) = 0.8660$	$\Sigma = \begin{bmatrix} 1 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1 \end{bmatrix}$			
$\theta = 45^\circ$ $\rho = \cos(\theta) = 0.7071$	$\Sigma = \begin{bmatrix} 1 & \sqrt{2}/2 \\ \sqrt{2}/2 & 1 \end{bmatrix}$			



### 随机变量独立

特别地，如果  $(X_1, X_2)$  服从二元高斯分布，并且随机变量  $X_1$  和  $X_2$  独立，这样  $(X_1, X_2)$  的协方差矩阵为：

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \quad (18)$$

注意，这个协方差矩阵为对角阵。

根据上一章所学，我们知道  $X_1$  和  $X_2$  各自的边缘概率密度函数分别为：

$$\begin{aligned} f_{X_1}(x_1) &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2\right) \\ f_{X_2}(x_2) &= \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right) \end{aligned} \quad (19)$$

如果  $(X_1, X_2)$  服从二元高斯函数，且  $X_1$  和  $X_2$  独立， $(X_1, X_2)$  概率密度函数可以写成两个边缘概率密度函数的乘积：

$$\begin{aligned} \underbrace{f_{X_1, X_2}(x_1, x_2)}_{\text{Joint}} &= \frac{1}{2\pi\sigma_1\sigma_2} \times \exp\left(-\frac{1}{2}\left(\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right)\right) \\ &= \underbrace{\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2\right)}_{\text{Marginal, } f_{X_1}(x_1)} \times \underbrace{\frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right)}_{\text{Marginal, } f_{X_2}(x_2)} \end{aligned} \quad (20)$$

这种情况，二元高斯分布 PDF 等高线为正椭圆。

## 11.2 高斯分布：椭圆、椭球、超椭球

### 椭圆分布

上一章提过高斯分布是**椭圆分布** (elliptical distribution) 的一种特殊形式。而椭圆分布的 PDF 一般形式为：

$$f(\mathbf{x}) = k \cdot g\left(\underbrace{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}_{\text{Ellipse}}\right) \quad (21)$$

本书第 7 章介绍的学生  $t$ -分布、逻辑分布、拉普拉斯分布也都是椭圆分布家族成员。

### 二元高斯分布：椭圆结构

回顾上一章介绍的二元高斯分布的椭圆结构。如图 2 所示，椭圆中心对应质心  $\boldsymbol{\mu}$ ，椭圆和  $\pm\sigma$  标准差构成的长方形相切，四个切点分别为  $A$ 、 $B$ 、 $C$  和  $D$ ，对角切点两两相连得到两条直线  $AC$ 、 $BD$ 。

$AC$  相当于在给定  $X_2$  条件下  $X_1$  的条件概率期望值； $BD$  相当于在给定  $X_1$  条件下  $X_2$  的条件概率期望值，这是本书第 12 章要讨论的话题。



在椭圆的学习中，我们很关注椭圆的长轴、短轴，对应图 2 中两条红线  $EG$ 、 $FH$ 。 $EG$  通过椭圆圆心  $O$  最长的线段，为椭圆长轴； $FH$  通过椭圆中心  $O$  最短的线段，为椭圆短轴。获得长轴、短轴的长度、角度需要用到特征值分解，这是本章后续要讨论的内容。



而长轴就是主成分分析的第一主元方向，这是本书第 14、25 章要讨论的话题。

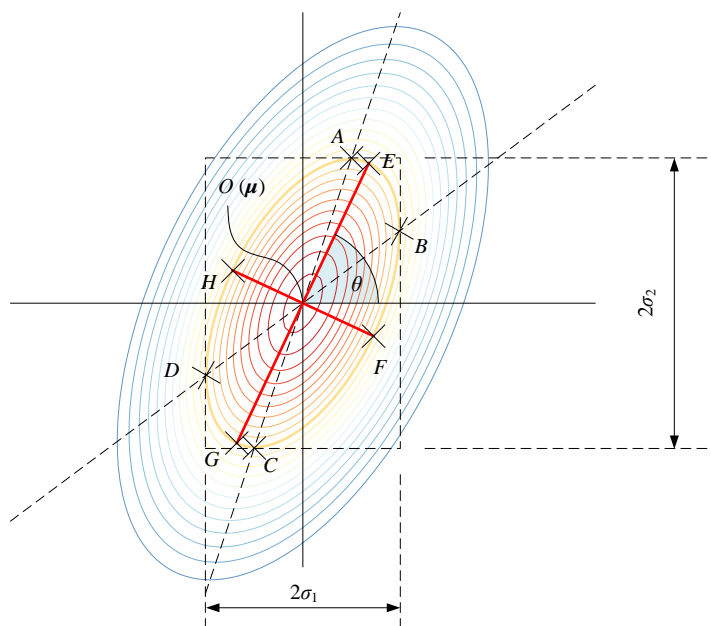


图 2. 椭圆和  $\pm\sigma$  标准差长方形的关系



Bk5\_Ch11\_01.py 绘制图 2。

### 三元高斯分布

前文提过，三元高斯分布 PDF 的几何图形是一层层“嵌套”的椭球。为了看见三元高斯分布 PDF 的椭球，我们采用“切片”这种可视化方案。



《可视之美》一册介绍过这种可视化方案。

图 3 可视化三元高斯分布的 PDF，这个高斯分布的质心位于原点，协方差矩阵为单位矩阵。图 3 的子图是在  $X_3$  在 5 个不同值上的“切片”，代表  $f_{X_1, X_2, X_3}(x_1, x_2, x_3 = c)$ 。容易看出来， $f_{X_1, X_2, X_3}(x_1, x_2, x_3 = c)$  的等高线是正圆。

图 4 所示为这个三元高斯分布的边缘分布。图中我们看到了协方差矩阵分块。



本书第 12、13 章还会进一步介绍协方差矩阵分块的应用场景。

图 5 和图 6 可视化协方差矩阵为对角阵的三元高斯分布，子图中我们看到的多是正椭圆。图 7 和图 8 可视化协方差矩阵为一般正定矩阵的三元高斯分布，我们看到的多是旋转椭圆。

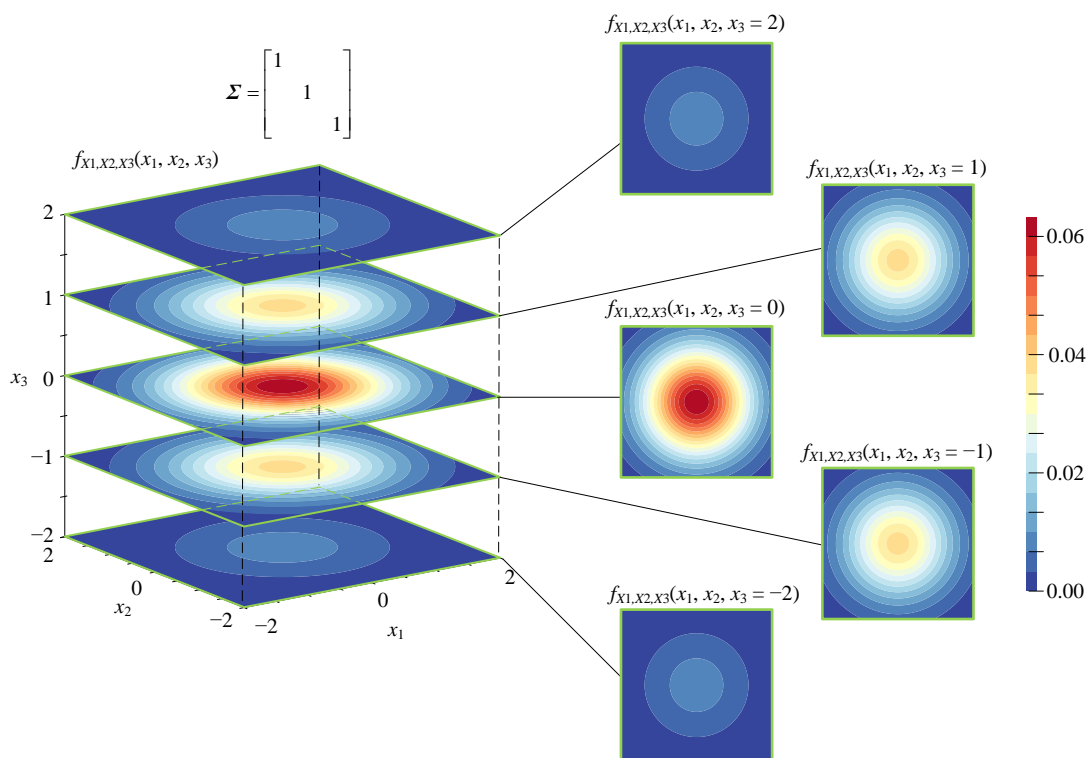


图 3. 三元高斯分布切片，协方差为单位矩阵

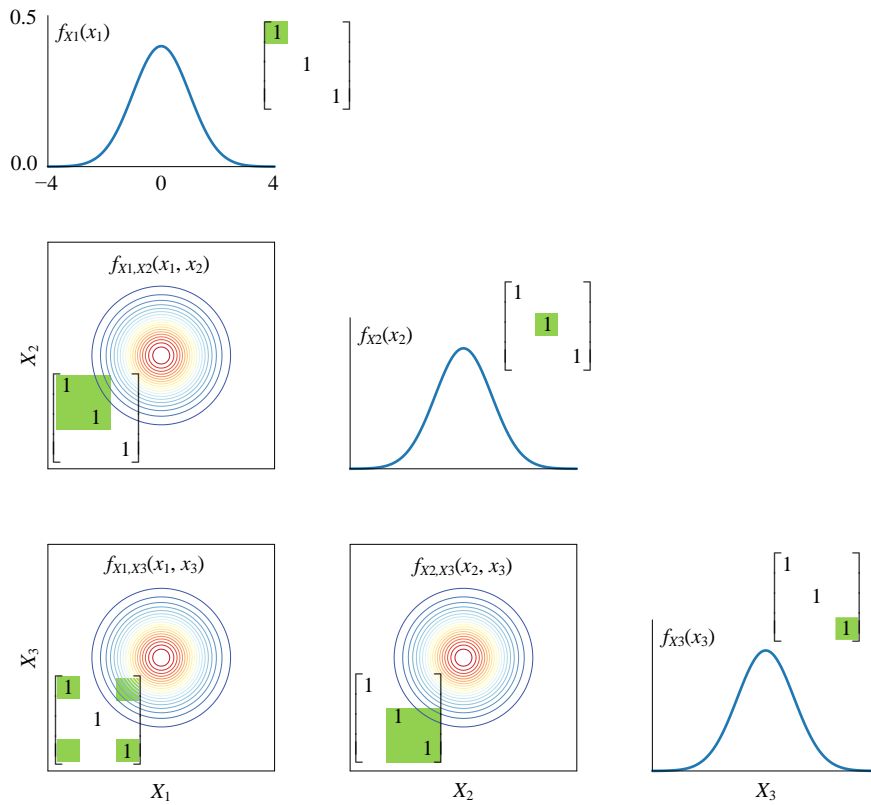


图 4. 三元高斯分布的边缘分布，协方差为单位矩阵

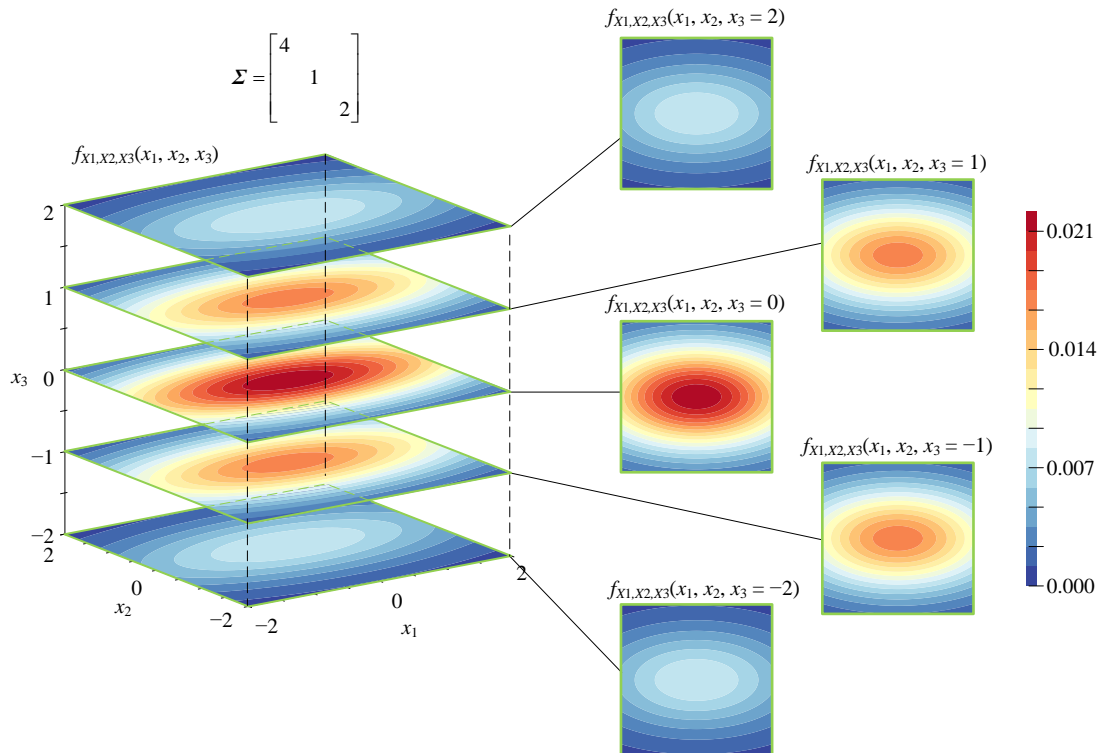


图 5. 三元高斯分布切片，协方差矩阵为对角矩阵

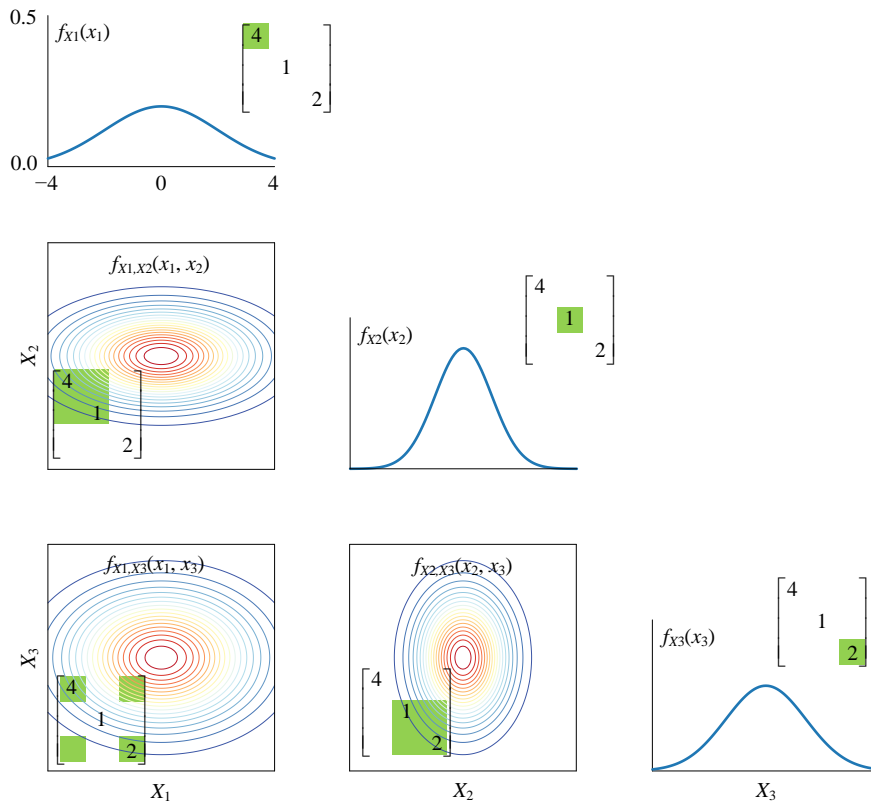


图 6. 三元高斯分布的边缘分布，协方差矩阵为对角矩阵

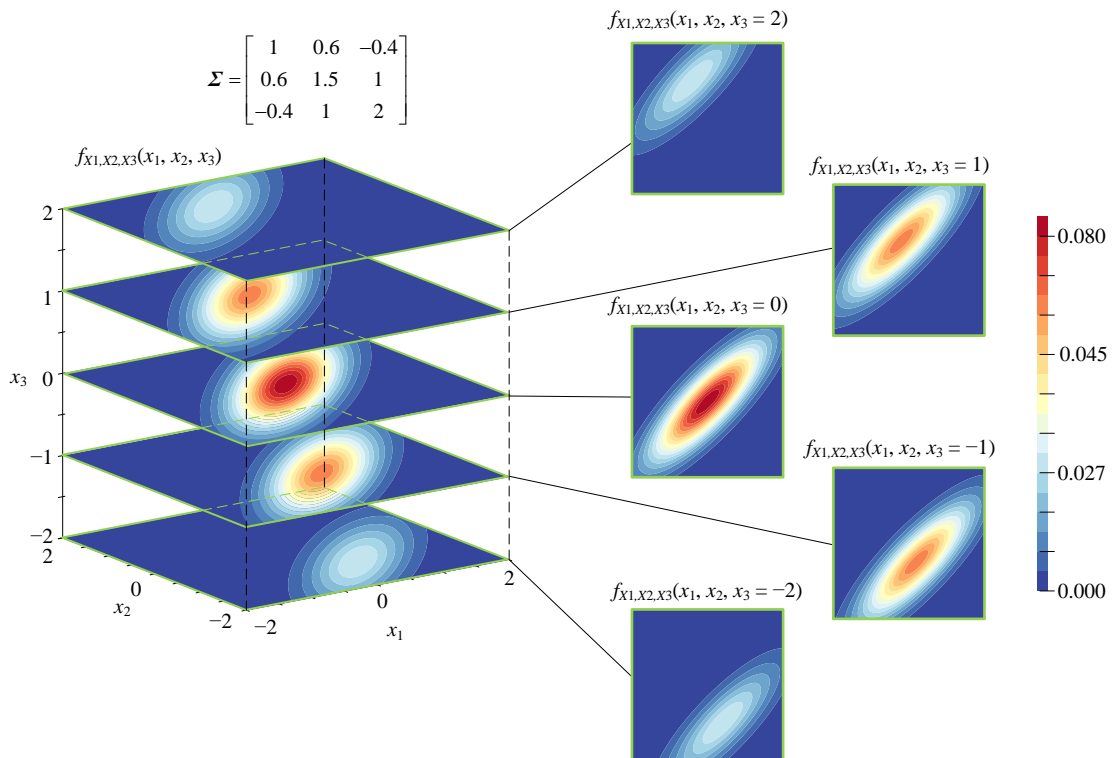


图 7. 三元高斯分布切片，协方差矩阵正定

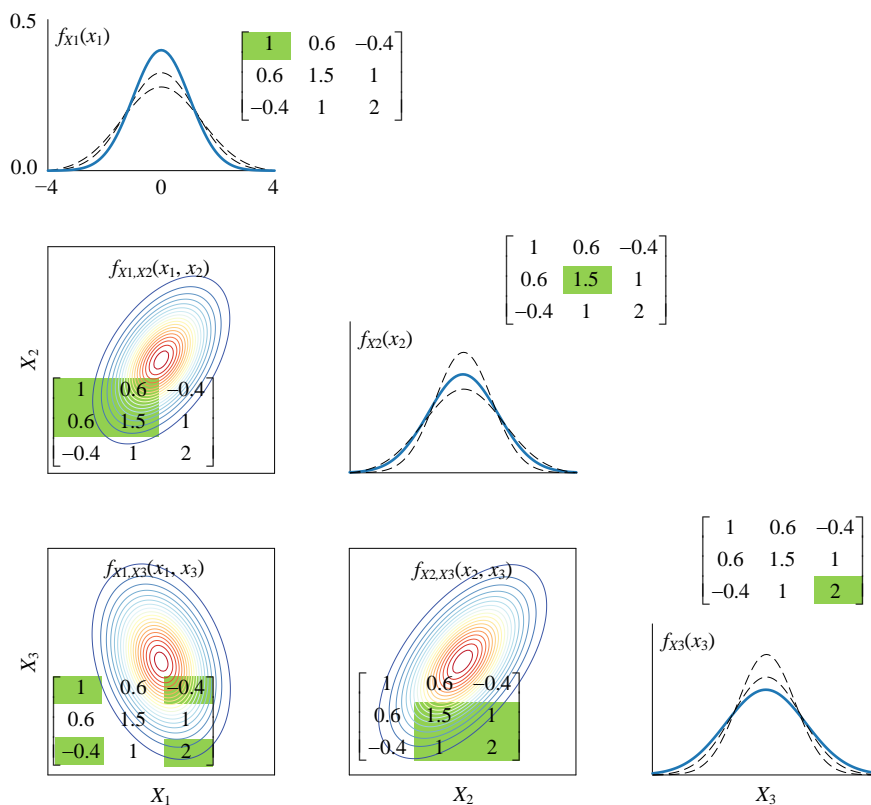


图 8. 三元高斯分布的边缘分布，协方差矩阵正定

这一章，我们用 `plotly.graph_objects.Volume()` 可视化三维高斯分布，这个 App 中大家可以调整分布参数。请大家参考 `Streamlit_Bk5_Ch11_03.py`。

## 11.3 解剖多元高斯分布 PDF

《矩阵力量》第 20 章介绍过如何用“平移 → 旋转 → 缩放”解剖多元高斯分布，本节把其中重要的内容“抄”了过来。

### 特征值分解协方差矩阵

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

$$\Sigma = V\Lambda V^T \quad (22)$$

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

如果  $\Sigma$  正定，利用 (22) 获得  $\Sigma^{-1}$  的特征值分解：

$$\Sigma^{-1} = V \Lambda^{-1} V^T \quad (23)$$

由此，将  $(x - \mu)^T \Sigma^{-1} (x - \mu)$  拆成  $\Lambda^{-\frac{1}{2}} V^T (x - \mu)$  的“平方”：

$$(x - \mu)^T V \Lambda^{-1} V^T (x - \mu) = \left[ \Lambda^{-\frac{1}{2}} V^T (x - \mu) \right]^T \Lambda^{-\frac{1}{2}} V^T (x - \mu) = \left\| \Lambda^{-\frac{1}{2}} V^T (x - \mu) \right\|_2^2 \quad (24)$$

### 平移 → 旋转 → 缩放

(24) 的几何解释是，旋转椭圆通过“平移  $(x - \mu) \rightarrow$  旋转  $(V^T) \rightarrow$  缩放  $(\Lambda^{-\frac{1}{2}})$ ”转换成单位圆，具体过程如图 9 所示。

图 9 (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 9 (a) 中散点是服从  $N(\mu, \Sigma)$  的随机数。

图 9 (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 9 (b) 中旋转椭圆代表多元高斯分布  $N(0, \Sigma)$ 。随机数质心也随之平移到原点。

图 9 (b) 中椭圆旋转之后得到图 9 (c) 中正椭圆，对应：

$$y = V^T x_c = V^T (x - \mu) \quad (25)$$

协方差矩阵  $\Sigma$  通过特征值分解得到特征值矩阵  $\Lambda$ 。而正椭圆的半长轴、半短轴长度蕴含在特征值矩阵  $\Lambda$  中，这算是拨开云雾的过程。图 9 (c) 中随机数服从  $N(0, \Lambda)$ 。

最后一步是缩放，从图 9 (c) 到图 9 (d)，对应：

$$z = \Lambda^{-\frac{1}{2}} y = \Lambda^{-\frac{1}{2}} V^T (x - \mu) \quad (26)$$

图 9 (d) 中单位圆则代表多元标准分布  $N(0, I)$ 。这意味着满足  $N(0, I)$  的随机变量为独立同分布。**独立同分布** (Independent and identically distributed, IID) 是指一组随机变量中每个变量的概率分布都相同，且这些随机变量互相独立。

利用向量  $z$ ，多元高斯分布 PDF 可以写成：

$$f_z(x) = \frac{\exp\left(-\frac{1}{2} z^T z\right)}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} = \frac{\exp\left(-\frac{1}{2} \|z\|_2^2\right)}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \quad (27)$$

$z$  的模  $\|z\|$  实际上代表“整体”Z 分数。

### 缩放 → 旋转 → 平移

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)

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。

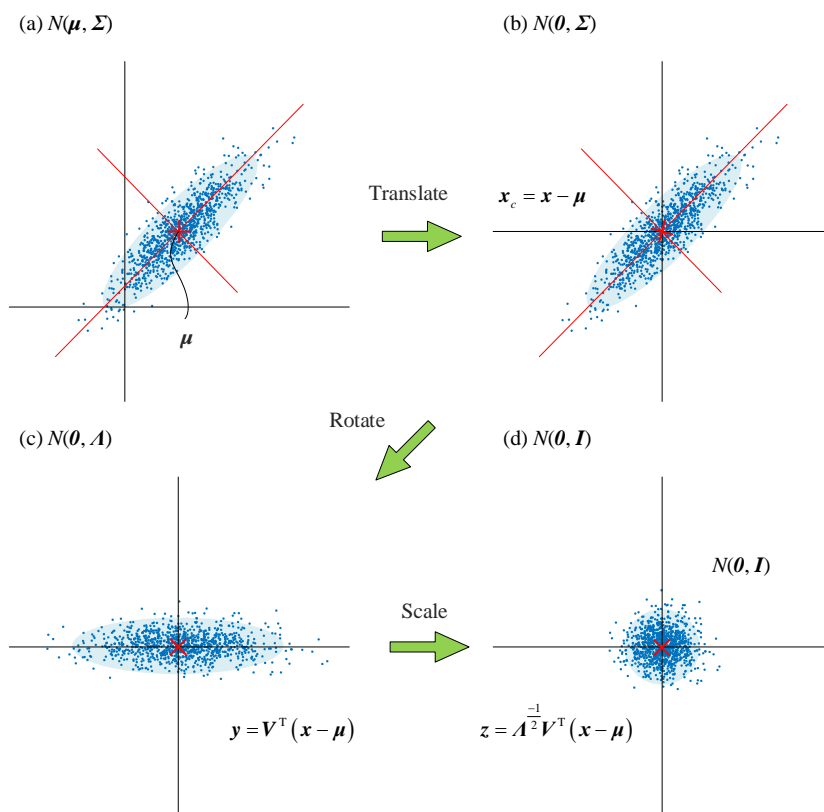


图 9. 平移 → 旋转 → 缩放，图片来自《矩阵力量》

## 马氏距离

马氏距离可以写成：

$$d = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})} = \left\| \mathbf{A}^{\frac{-1}{2}} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \right\| = \|\mathbf{z}\| \quad (28)$$

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 (unitless 或 dimensionless)，它将各个特征数据标准化。本书第 23 章将专门讲解马氏距离及其应用。

## 高斯函数

将 (28) 中马氏距离  $d$  代入多元高斯分布概率密度函数，得到：

$$f_{\mathbf{z}}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}d^2\right)}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \quad (29)$$

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”。图 10 所示为马氏距离图像。大家可以发现这个曲面为开口朝上的锥面，等高线为旋转椭圆。

图 10 (b) 中白色虚线正圆代表距离质心  $\boldsymbol{\mu}$  欧氏距离为 1 的等高线。欧氏距离是最自然的距离度量。而马氏距离则引入协方差矩阵  $\boldsymbol{\Sigma}$ ，计算距离时考虑数据的分布情况。



本书第 23 章将区分欧氏距离和马氏距离。

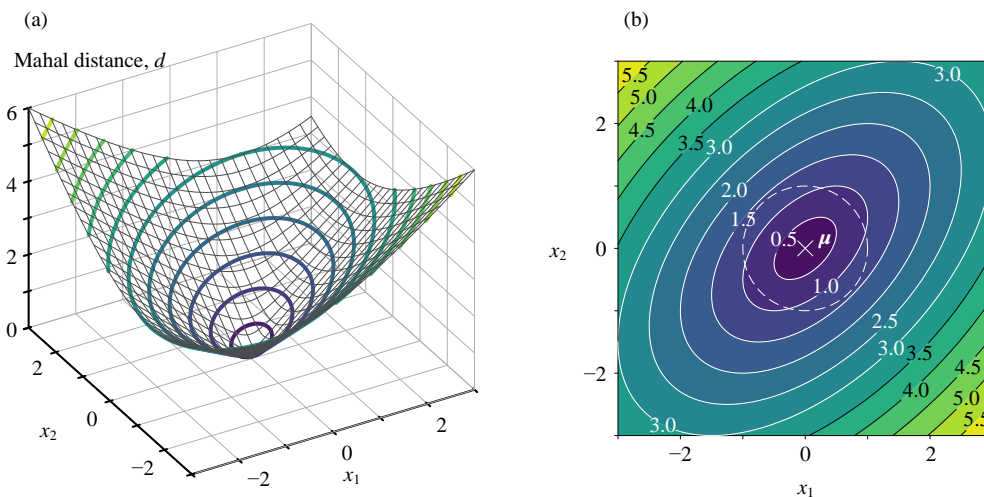


图 10. 马氏距离椭圆等高线

将具体马氏距离  $d$  值代入 (29)，可以得到高斯概率密度值。也就是说，图 10 每一个椭圆都对应一个概率密度值。这就是图 11 中等高线的含义。



请大家注意区分，椭圆等高线到底是代表马氏距离，还是概率密度值。



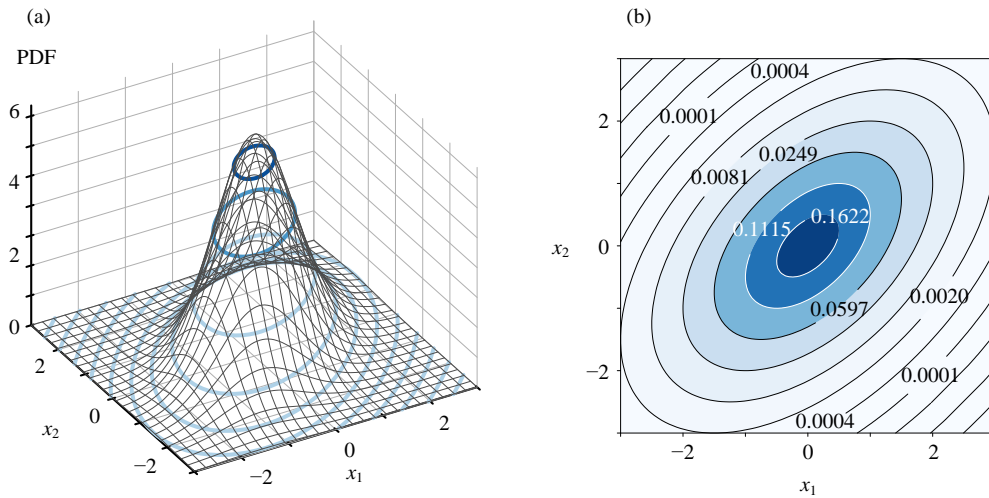


图 11. 高斯分布 PDF 椭圆等高线

### 分母：行列式值

把  $|\Sigma|^{-\frac{1}{2}}$  从 (5) 分母移到分子可以写成  $|\Sigma|^{-\frac{1}{2}}$ 。而  $\Sigma^{-\frac{1}{2}}$  相当于：

$$\Sigma^{-\frac{1}{2}} \sim A^{-\frac{1}{2}} V^T (\mathbf{x} - \boldsymbol{\mu}) \quad (30)$$

从体积角度来看，“平移 → 旋转 → 缩放”几何变换带来的面积/体积缩放系数便是  $|\Sigma|^{-\frac{1}{2}}$ 。准确来说，只有“缩放”才影响面积/体积，因此  $|\Sigma|^{-\frac{1}{2}} = |A|^{-\frac{1}{2}}$ 。

### 分母：体积归一化

从几何角度来看，(5) 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。

## 11.4 平移 → 旋转

本节以二元高斯分布 PDF 为例，利用特征值分解这个工具进一步深入理解多元高斯分布。

### 特征值分解

形状为  $2 \times 2$  协方差矩阵  $\Sigma$ ，它的特征值和特征向量关系为：

$$\begin{cases} \Sigma \mathbf{v}_1 = \lambda_1 \mathbf{v}_1 \\ \Sigma \mathbf{v}_2 = \lambda_2 \mathbf{v}_2 \end{cases} \quad (31)$$

(31) 可以写成：

$$\Sigma \underbrace{\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}}_{\mathbf{V}} = \underbrace{\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}}_{\mathbf{V}} \underbrace{\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}}_{\mathbf{A}} \quad (32)$$

即，

$$\Sigma \mathbf{V} = \mathbf{V} \mathbf{A} \quad (33)$$

将  $\Sigma$  具体值代入 (31) 得到：两个特征值对应的特征向量如下：

$$\begin{cases} \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \mathbf{v}_1 = \lambda_1 \mathbf{v}_1 \\ \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \mathbf{v}_2 = \lambda_2 \mathbf{v}_2 \end{cases} \quad (34)$$

两个特征值可以通过下式求得：

$$\begin{aligned} \lambda_1 &= \frac{\sigma_1^2 + \sigma_2^2}{2} + \sqrt{(\rho_{1,2}\sigma_1\sigma_2)^2 + \left(\frac{\sigma_1^2 - \sigma_2^2}{2}\right)^2} \\ \lambda_2 &= \frac{\sigma_1^2 + \sigma_2^2}{2} - \sqrt{(\rho_{1,2}\sigma_1\sigma_2)^2 + \left(\frac{\sigma_1^2 - \sigma_2^2}{2}\right)^2} \end{aligned} \quad (35)$$

当  $\rho_{1,2} = 0$  且  $\sigma_1 = \sigma_2$  时，(35) 中两个特征值相等。这种条件下，概率密度等高线为正圆。

### 长轴、短轴

大家已经清楚，二元高斯分布的 PDF 平面等高线是椭圆。如图 12 所示， $\sqrt{\lambda_1}$  就是椭圆半长轴长度， $\sqrt{\lambda_2}$  就是半短轴长度：

$$\begin{aligned} EO = GO = \sqrt{\lambda_1} &= \sqrt{\frac{\sigma_x^2 + \sigma_y^2}{2} + \sqrt{(\rho_{x,y}\sigma_x\sigma_y)^2 + \left(\frac{\sigma_x^2 - \sigma_y^2}{2}\right)^2}} \\ FO = HO = \sqrt{\lambda_2} &= \sqrt{\frac{\sigma_x^2 + \sigma_y^2}{2} - \sqrt{(\rho_{x,y}\sigma_x\sigma_y)^2 + \left(\frac{\sigma_x^2 - \sigma_y^2}{2}\right)^2}} \end{aligned} \quad (36)$$

$\mathbf{v}_1$  和  $\mathbf{v}_2$  具体值为：

$$\begin{aligned} \mathbf{v}_1 &= \begin{bmatrix} \frac{\sigma_1^2 - \sigma_2^2}{2} + \sqrt{\left(\rho_{1,2}\sigma_1\sigma_2\right)^2 + \left(\frac{\sigma_1^2 - \sigma_2^2}{2}\right)^2} \\ \rho_{1,2}\sigma_1\sigma_2 \\ 1 \end{bmatrix} \\ \mathbf{v}_2 &= \begin{bmatrix} \frac{\sigma_1^2 - \sigma_2^2}{2} - \sqrt{\left(\rho_{1,2}\sigma_1\sigma_2\right)^2 + \left(\frac{\sigma_1^2 - \sigma_2^2}{2}\right)^2} \\ \rho_{1,2}\sigma_1\sigma_2 \\ 1 \end{bmatrix} \end{aligned} \quad (37)$$

图 12 中， $\mathbf{v}_1$  对应的就是椭圆半长轴方向， $\mathbf{v}_2$  对应半短轴方向。在主成分分析中， $\mathbf{v}_1$  就是第一主元方向。 $\mathbf{v}_2$  便是第二主元方向。

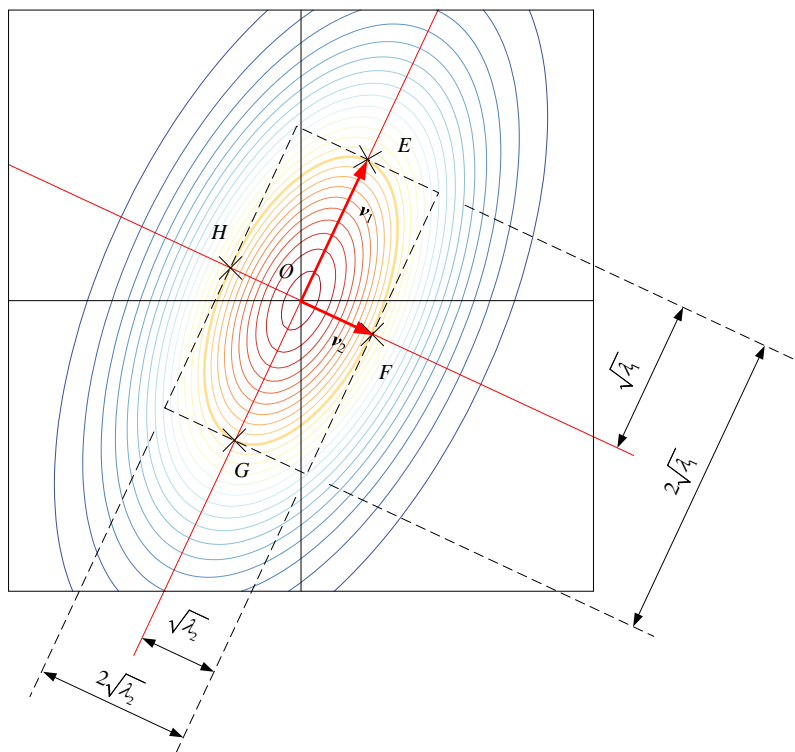


图 12. 椭圆的长轴、短轴

实际上，将  $(X_1, X_2)$  投影到  $\mathbf{v}_1$  得到的随机变量的方差就是  $\lambda_1$ ，对应的标准差为  $\sqrt{\lambda_1}$ 。将  $(X_1, X_2)$  投影到  $\mathbf{v}_2$  得到的随机变量的方差为  $\lambda_2$ ，其标准差为  $\sqrt{\lambda_2}$ 。

## 随机变量的线性变换

从另外一个角度来看，如图 13 所示，某个满足二元高斯分布随机变量  $(X_1, X_2)$  朝若干方向投影。我们先给出结论，这些方向中，向  $\mathbf{v}_1$  投影得到的随机变量方差最大，向  $\mathbf{v}_2$  投影得到的随机变量方差最小。

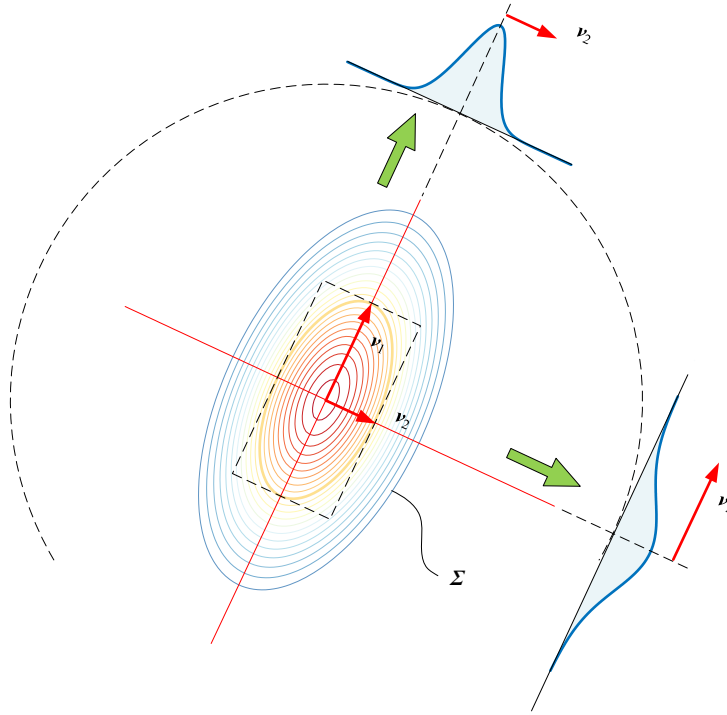


图 13. 二元高斯分布朝不同方向投影

假设二元随机变量列向量  $\mathbf{x} = [X_1, X_2]^T$  满足图 13 所示二元高斯分布。而  $\mathbf{x}$  先中心化，再向  $\mathbf{v}_1$  投影得到  $Y_1$ ：

$$Y_1 = (\mathbf{x} - \boldsymbol{\mu}_x)^T \mathbf{v}_1 = \left( \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \right)^T \begin{bmatrix} v_{1,1} \\ v_{2,1} \end{bmatrix} = (X_1 - \mu_1)v_{1,1} + (X_2 - \mu_2)v_{2,1} \quad (38)$$

从数据角度，上述过程如图 14 所示。

对  $Y_1$  求方差：

$$\begin{aligned} \text{var}(Y_1) &= E[(Y_1 - \mu_{Y_1})^2] = E\left[\left((\mathbf{x} - \boldsymbol{\mu}_x)^T \mathbf{v}_1\right)^T (\mathbf{x} - \boldsymbol{\mu}_x)^T \mathbf{v}_1\right] \\ &= \mathbf{v}_1^T E\left[\overbrace{\left((\mathbf{x} - \boldsymbol{\mu}_x)^T\right) (\mathbf{x} - \boldsymbol{\mu}_x)^T}^{\boldsymbol{\Sigma}_x}\right] \mathbf{v}_1 \\ &= \mathbf{v}_1^T \boldsymbol{\Sigma}_x \mathbf{v}_1 \end{aligned} \quad (39)$$

因为  $Y_1$  已经中心化，所以上式中  $\mu_{Y_1} = 0$ 。

将  $\Sigma_X$  的特征值分解代入 (39) 得到：

$$\begin{aligned}\text{var}(Y_1) &= \mathbf{v}_1^T \Sigma_X \mathbf{v}_1 = \mathbf{v}_1^T \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{bmatrix} \mathbf{v}_1 \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda_1\end{aligned}\quad (40)$$

实际上就是随机变量的线性变换，我们将会在本书第 14 章继续这一话题。

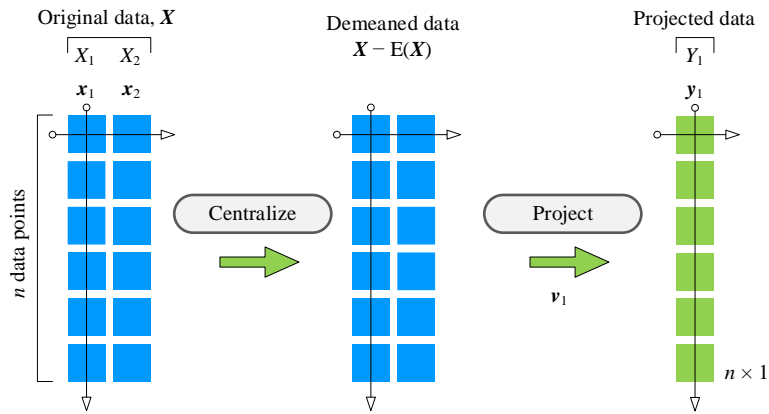


图 14.  $X$  先中心化，再向  $\mathbf{v}_1$  投影得到  $\mathbf{y}_1$

## 椭圆旋转

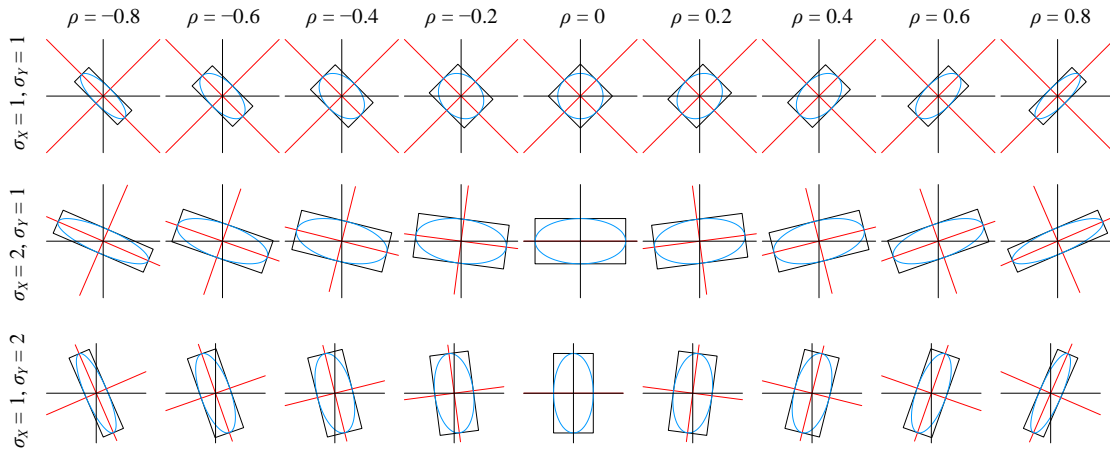
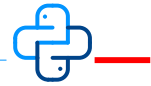
椭圆旋转角度  $\theta$ ：

$$\theta = \frac{1}{2} \arctan \left( \frac{2\rho_{1,2}\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2} \right) \quad (41)$$

图 15 所示为在  $\sigma_1$ 、 $\sigma_2$  大小不同， $\rho_{1,2}$  取值不同对椭圆旋转的影响。

观察 (41)，发现椭圆的旋转角度和  $\sigma_1$ 、 $\sigma_2$ 、 $\rho_{1,2}$  有关。

特别地，当  $\sigma_1 = \sigma_2$  时，如果  $\rho_{1,2}$  为小于 1 的正数，椭圆的旋转角度为  $45^\circ$ ；如果  $\rho_{1,2}$  为大于 -1 的负数，椭圆的旋转角度为  $-45^\circ$ 。

图 15. 在  $\sigma_1$ 、 $\sigma_2$  大小不同， $\rho_{1,2}$  取值不同对椭圆旋转的影响

Bk5\_Ch11\_02.py 绘制图 15。

### 特征值之和

可以发现 (35) 中两个特征值之和，等于协方差矩阵  $\Sigma$  的两个方差之和：

$$\lambda_1 + \lambda_2 = \sigma_1^2 + \sigma_2^2 \quad (42)$$



这正是《矩阵力量》讲到的特征值分解中，原矩阵迹等于特征值矩阵的迹。建议大家回顾特征值分解的优化视角。

### 特征值之积

两个特征值乘积为：

$$\begin{aligned} \lambda_1 \lambda_2 &= \left( \frac{\sigma_1^2 + \sigma_2^2}{2} \right)^2 - \left( (\rho_{1,2} \sigma_1 \sigma_2)^2 + \left( \frac{\sigma_1^2 - \sigma_2^2}{2} \right)^2 \right) \\ &= \sigma_1^2 \sigma_2^2 - \rho_{1,2}^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 (1 - \rho_{1,2}^2) \end{aligned} \quad (43)$$

这和协方差矩阵  $\Sigma$  行列式值相等：

$$|\Sigma| = \sigma_1^2 \sigma_2^2 - \rho_{1,2}^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 (1 - \rho_{1,2}^2) \quad (44)$$

## 谱分解

$\Sigma$  的谱分解可以进一步写成：

$$\Sigma = V\Lambda V^T = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{bmatrix} = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T \quad (45)$$

本书下一章还会继续这一话题。

## 平移 → 旋转

令

$$\mathbf{y} = V^T(\mathbf{x} - \boldsymbol{\mu}) \quad (46)$$

发现上式  $V^T(\mathbf{x} - \boldsymbol{\mu})$  相当于  $\mathbf{x}$  经过平移  $(\mathbf{x} - \boldsymbol{\mu})$ 、旋转  $(V^T)$  两步操作得到  $\mathbf{y}$ 。整个过程如图 16 所示。

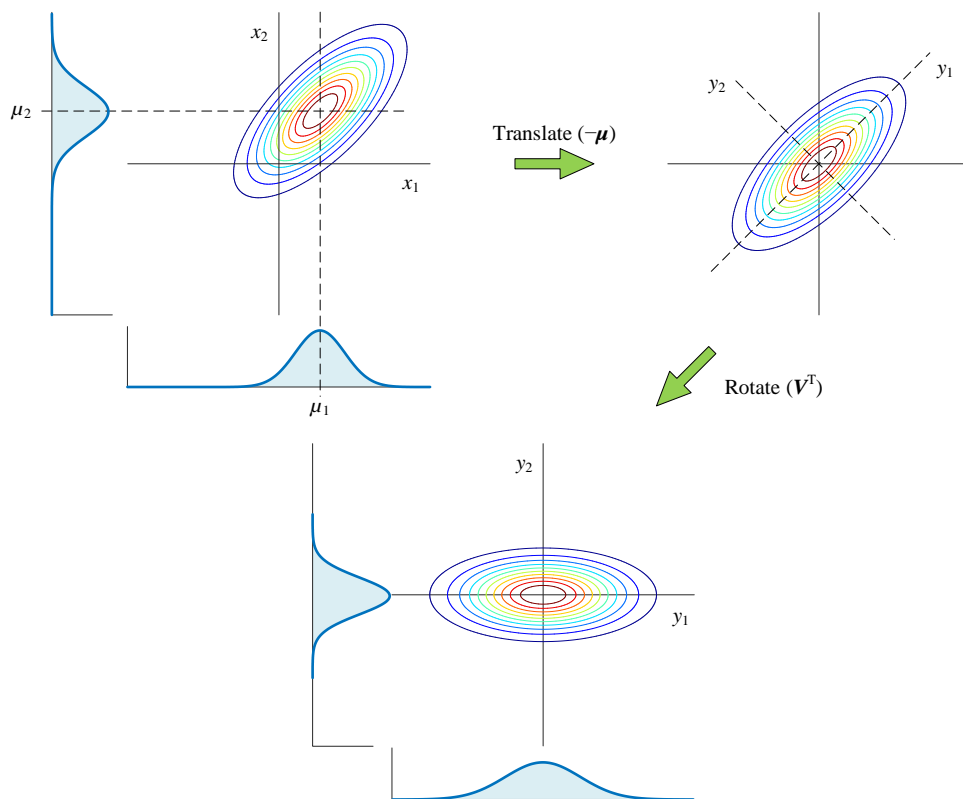


图 16. 椭圆先平移再旋转

这样  $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$  可以写成：

$$\mathbf{y}^T \mathbf{A}^{-1} \mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_q \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_q \end{bmatrix}^{-1} \begin{bmatrix} y_1 & y_2 & \cdots & y_q \end{bmatrix}^T = \sum_{j=1}^D \frac{y_j^2}{\lambda_j} \quad (47)$$

其中， $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D$ 。上式代表着一个多维空间正椭球体。

平移  $(\mathbf{x} - \boldsymbol{\mu})$ 、旋转  $(\mathbf{V}^T)$  两步几何变换只改变椭球的空间位置和旋转角度，不改变椭球本身的几何尺寸。也就是说， $|\boldsymbol{\Sigma}| = |\mathbf{A}|$ 。

特别地，当  $D = 2$  时，令  $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$  为 1，(47) 可以写成平面正椭圆：

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} = 1 \quad (48)$$

显然这个椭圆中心位于原点，同样这就解释了为什么图 12 中椭圆的半长轴为  $\sqrt{\lambda_1}$ ，半短轴为  $\sqrt{\lambda_2}$ 。

反过来， $\mathbf{y}$  先经过旋转、再平移得到  $\mathbf{x}$ ：

$$\mathbf{x} = \mathbf{V}\mathbf{y} + \boldsymbol{\mu} \quad (49)$$

## 独立

二元随机变量  $(Y_1, Y_2)$  对应的二元高斯分布 PDF 为：

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= \frac{1}{2\pi\sqrt{\lambda_1\lambda_2}} \times \exp\left(-\frac{1}{2}\left(\frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2}\right)\right) \\ &= \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{\lambda_1}} \exp\left(-\frac{1}{2}\frac{y_1^2}{\lambda_1}\right)}_{f_{Y_1}(y_1)} \times \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{\lambda_2}} \exp\left(-\frac{1}{2}\frac{y_2^2}{\lambda_2}\right)}_{f_{Y_2}(y_2)} \end{aligned} \quad (50)$$

可以发现随机变量  $Y_1$  和  $Y_2$  独立。如图 16 所示，随机变量  $Y_1$  对应的方差为  $\lambda_1$ ，标准差为  $\sqrt{\lambda_1}$ ；随机变量  $Y_2$  对应的方差为  $\lambda_2$ ，标准差为  $\sqrt{\lambda_2}$ 。

## 11.5 平移 → 旋转 → 缩放

$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$  可以整理为：



$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = [\mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu})]^T \mathbf{A}^{-\frac{1}{2}} \mathbf{A}^{-\frac{1}{2}} [\mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu})] = \left( \mathbf{A}^{-\frac{1}{2}} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \right)^2 \quad (51)$$

这就是前文讲到的“开方”。

令：

$$\mathbf{z} = \mathbf{A}^{-\frac{1}{2}} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \quad (52)$$

上式相当于  $\mathbf{x}$  经过平移、旋转和缩放，最后得到  $\mathbf{z}$ ，整个过程如图 17 所示。

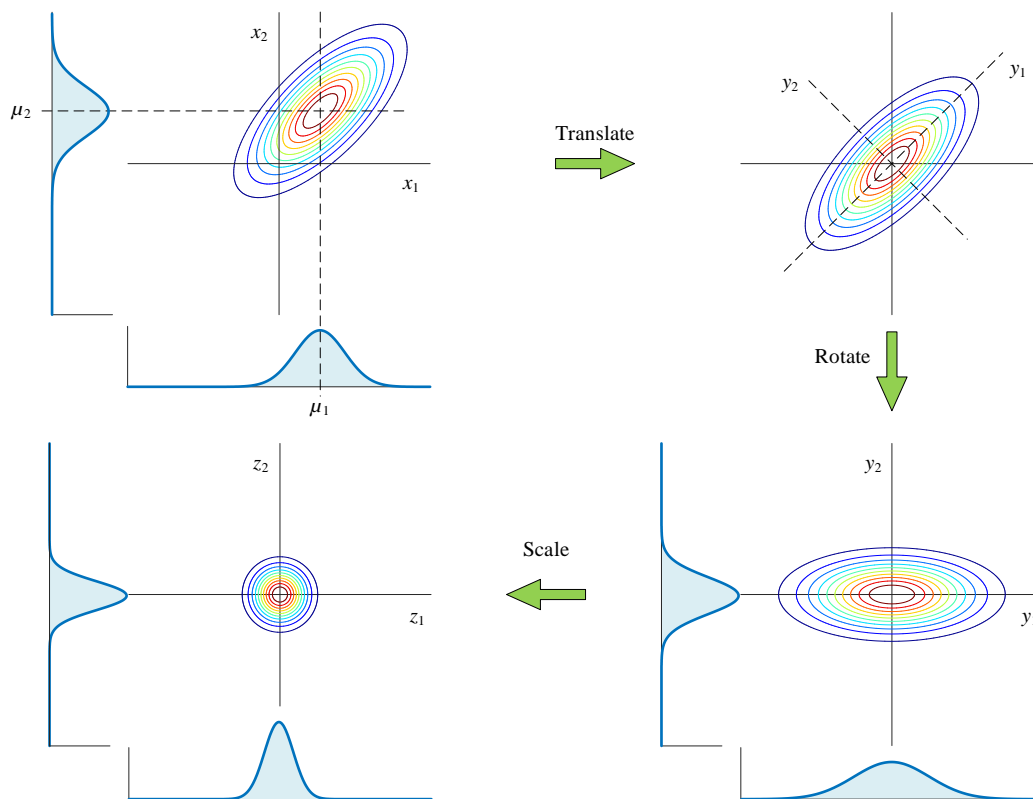


图 17. 椭圆先平移、再旋转，最后缩放，得到单位圆

## 单位球体

将 (52) 代入 (51)，得到的解析式：

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{z}^T \mathbf{z} = z_1^2 + z_2^2 + \cdots + z_D^2 = \sum_{j=1}^D z_j^2 \quad (53)$$

当上式为 1 时，它代表多维空间的单位球体。

反过来，也可以利用  $\mathbf{z}$  通过缩放、旋转、平移，反求  $\mathbf{x}$ ：

$$\mathbf{x} = \underset{\text{Rotate}}{\mathbf{V}} \underset{\text{Scale}}{\mathbf{D}} \mathbf{z} + \underset{\text{Translate}}{\boldsymbol{\mu}} \quad (54)$$

图 18 展示 (54) 对应的几何变换。

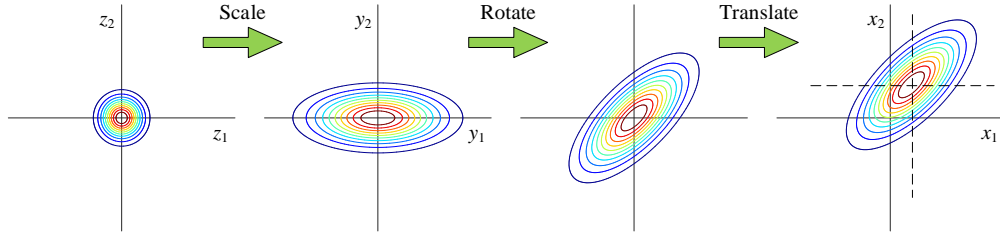


图 18. 单位圆先缩放，再旋转，最后平移

### 数据视角

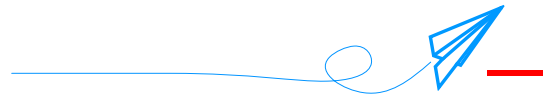
类似 (38)，从数据角度来看，如果数据矩阵  $\mathbf{X}$  服从  $N(\mathbf{E}(\mathbf{X}), \boldsymbol{\Sigma}_x)$ 。对  $\mathbf{X}$  先中心化，再向  $\mathbf{V}$  投影，最后缩放得到  $\mathbf{Z}$ ：

$$\mathbf{Z} = (\mathbf{X} - \mathbf{E}(\mathbf{X})) \mathbf{V} \mathbf{A}^{\frac{-1}{2}} \quad (55)$$

$\mathbf{Z}$  的协方差矩阵为单位矩阵  $\mathbf{I}$ ：

$$\begin{aligned} \boldsymbol{\Sigma}_z &= \frac{\mathbf{Z}^T \mathbf{Z}}{n-1} = \frac{\left( (\mathbf{X} - \mathbf{E}(\mathbf{X})) \mathbf{V} \mathbf{A}^{\frac{-1}{2}} \right)^T \left( (\mathbf{X} - \mathbf{E}(\mathbf{X})) \mathbf{V} \mathbf{A}^{\frac{-1}{2}} \right)}{n-1} \\ &= \mathbf{A}^{\frac{-1}{2}} \mathbf{V}^T \overbrace{\frac{(\mathbf{X} - \mathbf{E}(\mathbf{X}))^T (\mathbf{X} - \mathbf{E}(\mathbf{X}))}{n-1}}^{\boldsymbol{\Sigma}_x} \mathbf{V} \mathbf{A}^{\frac{-1}{2}} \\ &= \mathbf{A}^{\frac{-1}{2}} \mathbf{V}^T \boldsymbol{\Sigma}_x \mathbf{V} \mathbf{A}^{\frac{-1}{2}} = \mathbf{I} \end{aligned} \quad (56)$$

也就是说，如果  $\mathbf{X}$  服从多维高斯分布的话， $\mathbf{Z}$  服从 IID 标准正态分布。



本章将一元、二元、三元高斯分布提高到了多元。而多元高斯分布离不开矩阵运算。

利用特征值分解，我们从几何角度理解多元高斯分布 PDF 中隐含的“平移 → 旋转 → 缩放”。这对理解协方差矩阵、马氏距离、主成分分析等概念至关重要。

以后希望大家每次见到多元高斯分布 PDF 式子，对它的每个组成部分的作用都能如数家珍、滔滔不绝。