

## 12

## Cholesky Decomposition

## Cholesky 分解

获取列向量空间位置坐标



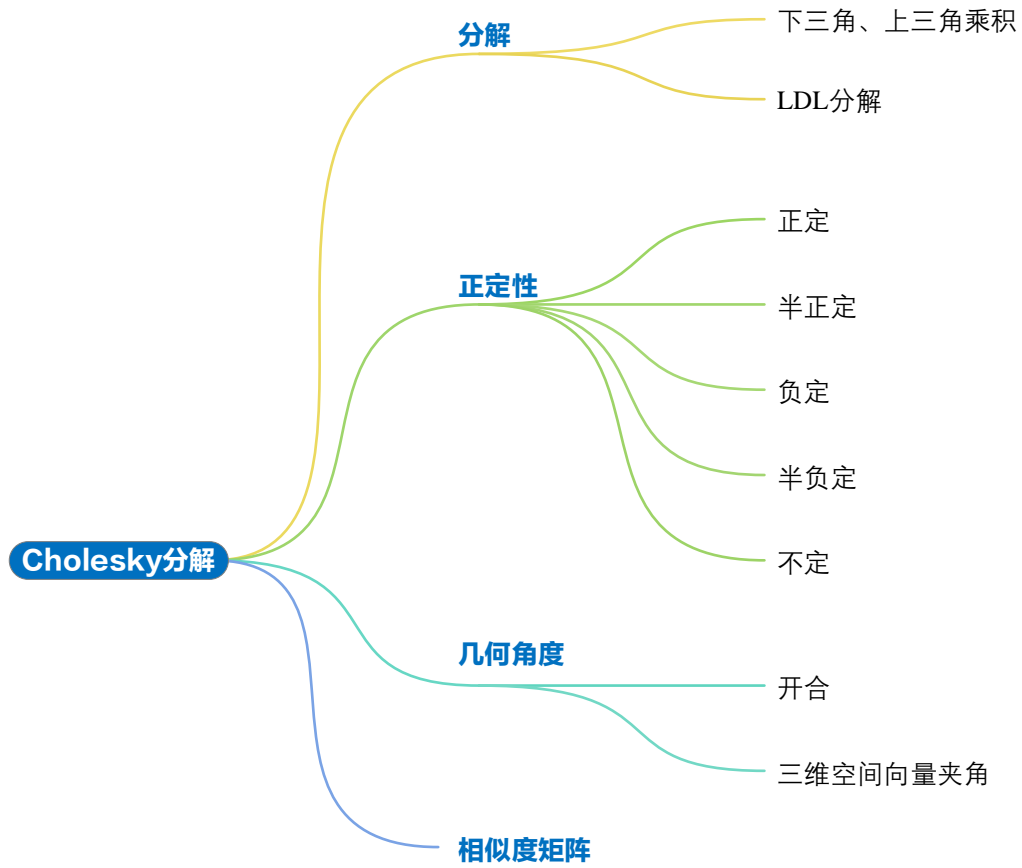
每个人都是天才。但是，如果你以爬树的能力来判断一条鱼，那么那条鱼终其一生都会认为自己愚蠢无能。

***Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.***

—— 阿尔伯特·爱因斯坦 (Albert Einstein) | 理论物理学家 | 1879 ~ 1955



- ◀ `ax.contour3D()` 绘制三维曲面等高线
- ◀ `ax.plot_wireframe()` 绘制线框图
- ◀ `math.radians()` 将角度转换成弧度
- ◀ `matplotlib.pyplot.contour()` 绘制平面等高线
- ◀ `matplotlib.pyplot.contourf()` 绘制平面填充等高线
- ◀ `matplotlib.pyplot.plot()` 绘制线图
- ◀ `matplotlib.pyplot.quiver()` 绘制箭头图
- ◀ `matplotlib.pyplot.scatter()` 绘制散点图
- ◀ `numpy.arccos()` 反余弦
- ◀ `numpy.cos()` 计算余弦值
- ◀ `numpy.deg2rad()` 将角度转化为弧度
- ◀ `numpy.linalg.cholesky()` Cholesky 分解
- ◀ `numpy.linalg.eig()` 特征值分解



## 12.1 Cholesky 分解

实数矩阵的 Cholesky 分解由法国军官、数学家安德烈·路易·科列斯基 (André-Louis Cholesky) 最先发明。科列斯基本人在一战结束前夕战死沙场，Cholesky 分解是由科列斯基的同事在他死后发表的，并以科列斯基的名字命名。

通过上一章学习，大家知道 Cholesky 分解将方阵  $A$  分解为一个下三角矩阵  $L$  以及它的转置  $L^T$  的乘积：

$$A = LL^T \quad (1)$$

定义上三角矩阵  $R$  为  $L^T$  的话，(1) 可以写成：

$$A = R^T R \quad (2)$$

### LDL 分解

在 Cholesky 分解基础上，上一章又介绍了 LDL 分解。LDL 分解，将上述矩阵  $A$  分解成下三角矩阵  $L$ 、对角阵方阵  $D$ 、 $L^T$  三者乘积，即，

$$A = LDL^T \quad (3)$$

注意，(3) 中下三角矩阵  $L$  为对角线元素均为 1。

假设对角方阵  $D$  对角线元素非负，LDL 分解可以进一步写成：

$$A = LD^{1/2} (D^{1/2})^T L^T = LD^{1/2} (LD^{1/2})^T \quad (4)$$

$D^{1/2}$  也是个对角方阵， $D^{1/2}$  对角线上元素是  $D$  的对角线元素的非负平方根。

令，

$$B = D^{1/2} \quad (5)$$

(4) 可写成：

$$A = LB(LB)^T \quad (6)$$

用上三角矩阵  $R$  来替换  $L^T$ ，(6) 可以写成：

$$A = R^T BBR = (BR)^T BR \quad (7)$$

## 12.2 正定矩阵才可以进行 Cholesky 分解

本书前文提到，并非所有矩阵都可以做 Cholesky 分解，只有**正定矩阵** (positive-definite matrix) 才可以。

方阵  $A$  为**正定矩阵** (positive definite matrix) 时，需要满足在  $\mathbf{x} \neq \mathbf{0}$  ( $\mathbf{x}$  为非零列向量) 条件下：

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$$

(8)

注意，正定矩阵的特征值均为正。这一点，本书后续将深入讨论。

### 几何视角

从几何角度更容易理解正定矩阵，以如下  $2 \times 2$  矩阵为例：

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

(9)

定义二元函数  $y = f(x_1, x_2)$ ：

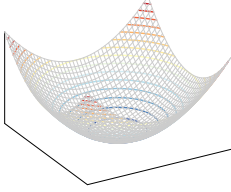
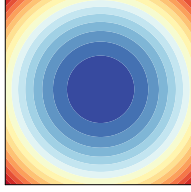
$$\begin{aligned} y &= f(x_1, x_2) \\ &= \mathbf{x}^T \mathbf{A} \mathbf{x} \\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= ax_1^2 + 2bx_1x_2 + cx_2^2 \end{aligned}$$

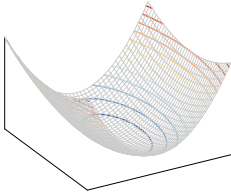
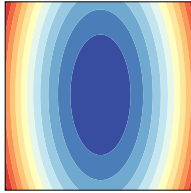
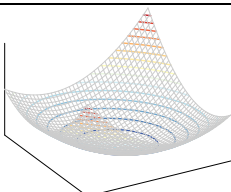
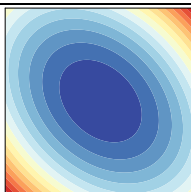
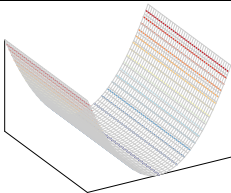
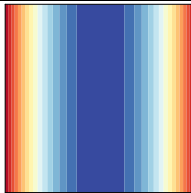
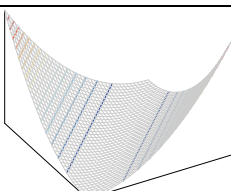
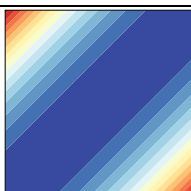
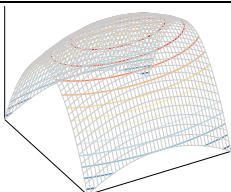
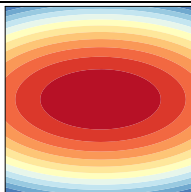
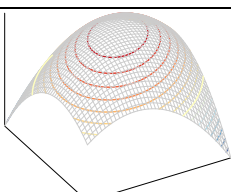
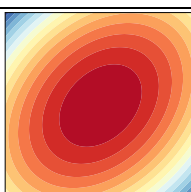
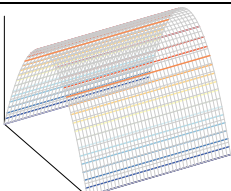
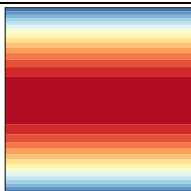
(10)

可以发现函数  $y = f(x_1, x_2)$  为二次型，这就把正定性和丛书《数学要素》讲过的二次曲面联系起来了。

除了正定矩阵，还有半正定、负定、半负定、不定这几种正定性。表 1 总结几种正定性例子、曲面、等高线特征。希望读者能够通过表中几何图形建立正定性的直观印象，本书后续将专门深入讲解正定性这一概念。

表 1. 几种正定性

正定性	例子	三维曲面	平面等高线
正定 (positive definite)	开口向上正圆抛物面 $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$		

	开口向上正椭圆 抛物面 $A = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$		
	开口向上旋转椭圆 抛物面 $A = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}$		
半正定 (positive semi-definite)	山谷面 $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$		
	旋转山谷面 $A = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}$		
负定 (negative definite)	开口向下正椭圆 抛物面 $A = \begin{bmatrix} -0.5 & 0 \\ 0 & -2 \end{bmatrix}$		
	开口向下旋转椭圆 抛物面 $A = \begin{bmatrix} -1.5 & 0.5 \\ 0.5 & -1.5 \end{bmatrix}$		
半负定 (negative semi-definite)	山脊面 $A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$		

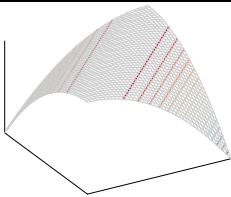
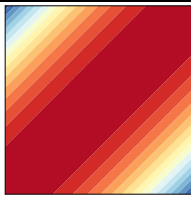
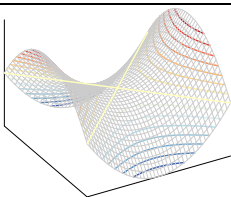
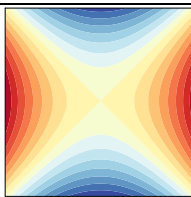
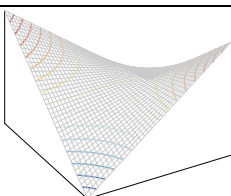
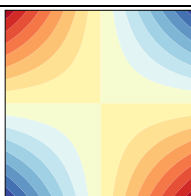
本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

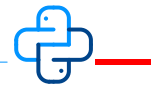
版权归清华大学出版社所有，请勿商用，引用请注明出处。

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)

	旋转山脊面 $A = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}$		
不定 (indefinite)	马鞍面 $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$		
	旋转马鞍面 $A = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$		



Bk4\_Ch12\_01.py 绘制表 1 三维曲面和等高线。请注意改变 a、b、c 三个系数取值。

```
# Bk4_Ch12_01.py

import numpy as np
from matplotlib import pyplot as plt

x1 = np.arange(-2,2,0.05)
x2 = np.arange(-2,2,0.05)

xx1_fine, xx2_fine = np.meshgrid(x1,x2)

a = 0; b = -1; c = 0;

yy_fine = a*xx1_fine**2 + 2*b*xx1_fine*xx2_fine + c*xx2_fine**2

# 3D visualization

fig, ax = plt.subplots()
ax = plt.axes(projection='3d')

ax.plot_wireframe(xx1_fine,xx2_fine,yy_fine,
                  color = [0.8,0.8,0.8],
                  linewidth = 0.25)

ax.contour3D(xx1_fine,xx2_fine,yy_fine,15,
             cmap = 'RdYlBu_r')

ax.view_init(elev=30, azim=60)
ax.xaxis.set_ticks([])
ax.yaxis.set_ticks([])
ax.zaxis.set_ticks([])
plt.tight_layout()
ax.set_proj_type('ortho')
plt.show()
```

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)

```
# 2D visualization
fig, ax = plt.subplots()
ax.contourf(xx1_fine, xx2_fine, yy_fine, 15,
            cmap = 'RdYlBu_r')

ax.xaxis.set_ticks([])
ax.yaxis.set_ticks([])
ax.set_aspect('equal')

plt.show()
```

## 12.3 几何角度：开合

本节，我们从一个有趣的几何视角来分析一种特殊矩阵的 Cholesky 分解。

### 以 $2 \times 2$ 矩阵为例

给定如  $2 \times 2$  矩阵  $P$ ，它的主对角元素为 1，非主对角线元素为余弦值  $\cos\theta_{1,2}$ ：

$$P = \begin{bmatrix} 1 & \cos\theta_{1,2} \\ \cos\theta_{1,2} & 1 \end{bmatrix} \quad (11)$$

对矩阵  $P$  进行 Cholesky 分解可以得到：

$$P = LL^T = \underbrace{\begin{bmatrix} 1 & 0 \\ \cos\theta_{1,2} & \sin\theta_{1,2} \end{bmatrix}}_L \underbrace{\begin{bmatrix} 1 & \cos\theta_{1,2} \\ 0 & \sin\theta_{1,2} \end{bmatrix}}_{L^T} = \begin{bmatrix} 1 & \cos\theta_{1,2} \\ \cos\theta_{1,2} & 1 \end{bmatrix} \quad (12)$$

矩阵  $P$  的 Cholesky 分解还可以写成：

$$P = R^T R = \underbrace{\begin{bmatrix} 1 & 0 \\ \cos\theta_{1,2} & \sin\theta_{1,2} \end{bmatrix}}_{R^T} \underbrace{\begin{bmatrix} 1 & \cos\theta_{1,2} \\ 0 & \sin\theta_{1,2} \end{bmatrix}}_R \quad (13)$$

将  $R$  写成：

$$R = \begin{bmatrix} 1 & \cos\theta_{1,2} \\ 0 & \sin\theta_{1,2} \end{bmatrix} = [\mathbf{r}_1 \quad \mathbf{r}_2] \quad (14)$$

在平面直角坐标系中， $\mathbf{e}_1$  和  $\mathbf{e}_2$  分别代表水平和竖直正方向的单位向量。 $R$  分别乘  $\mathbf{e}_1$  和  $\mathbf{e}_2$ ，得到  $\mathbf{r}_1$  和  $\mathbf{r}_2$ ：

$$\begin{aligned} \mathbf{r}_1 &= R\mathbf{e}_1 = \begin{bmatrix} 1 & \cos\theta_{1,2} \\ 0 & \sin\theta_{1,2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \mathbf{r}_2 &= R\mathbf{e}_2 = \begin{bmatrix} 1 & \cos\theta_{1,2} \\ 0 & \sin\theta_{1,2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta_{1,2} \\ \sin\theta_{1,2} \end{bmatrix} \end{aligned} \quad (15)$$

很容易判断  $\mathbf{r}_1$  和  $\mathbf{r}_2$  均为单位向量。

而向量  $\mathbf{r}_1$  和  $\mathbf{r}_2$  夹角余弦值为：

$$\cos \theta = \frac{\mathbf{r}_1 \cdot \mathbf{r}_2}{\|\mathbf{r}_1\| \|\mathbf{r}_2\|} = \cos \theta_{1,2} \quad (16)$$

## 几何视角

如图 1 所示，从几何角度来讲， $\mathbf{R} (\mathbf{L}^T)$  的相当于把原本正交的  $[\mathbf{e}_1, \mathbf{e}_2]$  标准正交基转化成具有一定夹角的  $[\mathbf{r}_1, \mathbf{r}_2]$  非正交基，且  $\mathbf{e}_1 = \mathbf{r}_1$ ，相当于“锚定”。这类似 QR 分解中， $\mathbf{q}_1$  平行于  $\mathbf{x}_1$ 。再次强调，虽然  $[\mathbf{r}_1, \mathbf{r}_2]$  中每个向量为单位向量，但是  $[\mathbf{r}_1, \mathbf{r}_2]$  为非正交基。

图 1 所示这种几何变换像是“门合页”的开合。

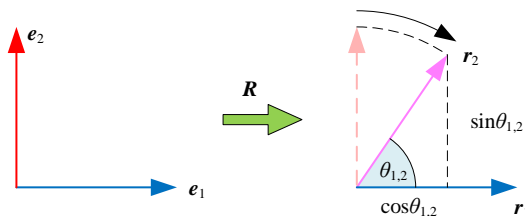


图 1. 开合

$0 < \cos \theta_{1,2} < 1$  时，即  $0^\circ < \theta_{1,2} < 90^\circ$ ，“合页”从直角  $90^\circ$  关闭角度至  $\theta_{1,2}$ 。

$-1 < \cos \theta_{1,2} < 0$  时，即  $90^\circ < \theta_{1,2} < 180^\circ$ ，“合页”从直角  $90^\circ$  打开至  $\theta_{1,2}$ 。我们给这种几何变换取个名字，叫做“开合”。注意，这种所谓“开合”的几何变换只适用于标准正交基。

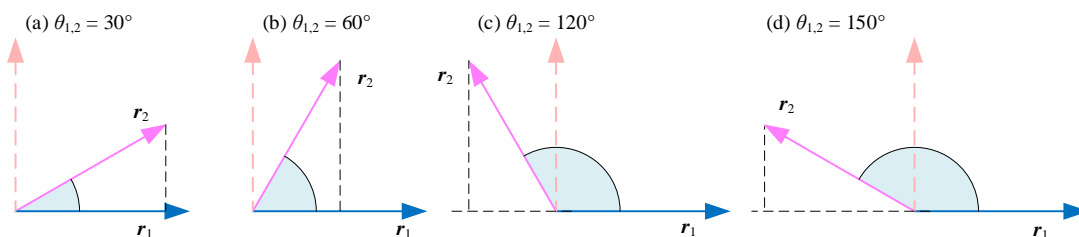


图 2. 不同的开合角度  $\cos \theta_{1,2}$

计算 (14) 中  $\mathbf{R}$  的行列式值：

$$|\mathbf{R}| = \begin{vmatrix} 1 & \cos \theta_{1,2} \\ 0 & \sin \theta_{1,2} \end{vmatrix} = \sin \theta_{1,2} \quad (17)$$



这个行列式值结果表明“开合”前后，图形的面积缩比为  $\sin\theta_{1,2}$ 。这和我们在图 3 中看到一致。 $[e_1, e_2]$  构造正方形面积为 1，而  $[r_1, r_2]$  构造的平行四边形面积为  $\sin\theta_{1,2}$ 。

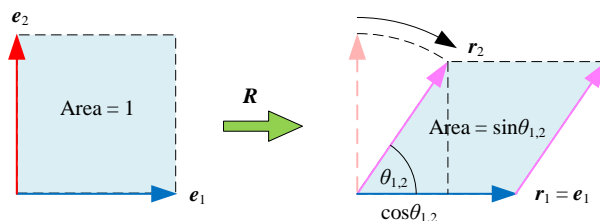


图 3. 开合对应的面积变化

### 举个例子

给定  $P$  为：

$$P = \begin{bmatrix} 1 & \cos 60^\circ \\ \cos 60^\circ & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad (18)$$

对  $P$  进行 Cholesky 分解得到：

$$P = LL^T = \underbrace{\begin{bmatrix} 1 & 0 \\ 0.5 & \sqrt{3}/2 \end{bmatrix}}_{R^T} \underbrace{\begin{bmatrix} 1 & 0.5 \\ 0 & \sqrt{3}/2 \end{bmatrix}}_R \quad (19)$$

图 4 所示为  $e_1$  和  $e_2$  经过 (19) 中  $R$  转换得到向量  $r_1$  和  $r_2$ 。

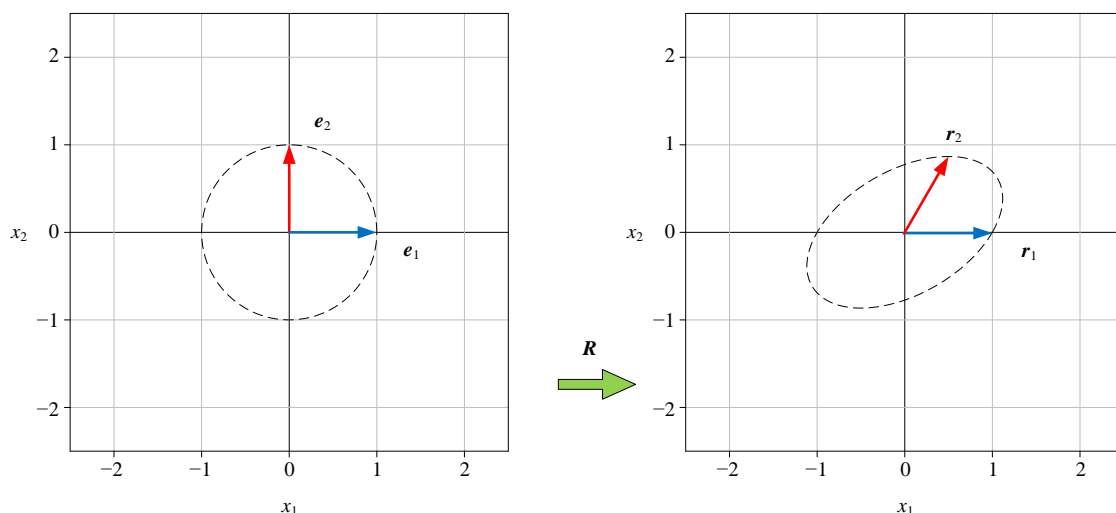


图 4.  $e_1$  和  $e_2$  经过  $R$  转换得到向量  $r_1$  和  $r_2$

## 12.4 几何变换：缩放 → 开合

给定  $\Sigma$  具体形式如下：

$$\Sigma = \begin{bmatrix} a^2 & a \cdot b \cdot \cos \theta_{1,2} \\ a \cdot b \cdot \cos \theta_{1,2} & b^2 \end{bmatrix} \quad (20)$$

其中， $a$  和  $b$  都是正数。

先把  $\Sigma$  写成：

$$\Sigma = \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_S \underbrace{\begin{bmatrix} 1 & \cos \theta_{1,2} \\ \cos \theta_{1,2} & 1 \end{bmatrix}}_P \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_S \quad (21)$$

将 (11) 代入 (21)，得到：

$$\Sigma = (RS)^T (RS) = \underbrace{\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}}_S \underbrace{\begin{bmatrix} 1 & 0 \\ \cos \theta_{1,2} & \sin \theta_{1,2} \end{bmatrix}}_{R^T} \underbrace{\begin{bmatrix} 1 & \cos \theta_{1,2} \\ 0 & \sin \theta_{1,2} \end{bmatrix}}_R \underbrace{\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}}_S \quad (22)$$

上式相当于对  $\Sigma$  直接进行 Cholesky 分解的结果：

$$\Sigma = L_{\Sigma} (L_{\Sigma})^T \quad (23)$$

将  $RS$  先后作用在  $e_1$  和  $e_2$  上，得到  $x_1$  和  $x_2$ ：

$$\begin{aligned} x_1 &= RS e_1 = \begin{bmatrix} 1 & \cos \theta_{1,2} \\ 0 & \sin \theta_{1,2} \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ x_2 &= RS e_2 = \begin{bmatrix} 1 & \cos \theta_{1,2} \\ 0 & \sin \theta_{1,2} \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = b \begin{bmatrix} \cos \theta_{1,2} \\ \sin \theta_{1,2} \end{bmatrix} \end{aligned} \quad (24)$$

这相当于，先对  $e_1$  和  $e_2$  先缩放 ( $S$ )，再开合 ( $R$ )。请读者格外注意几何变换的先后顺序，这个顺序就是矩阵乘法的从右向左的顺序。

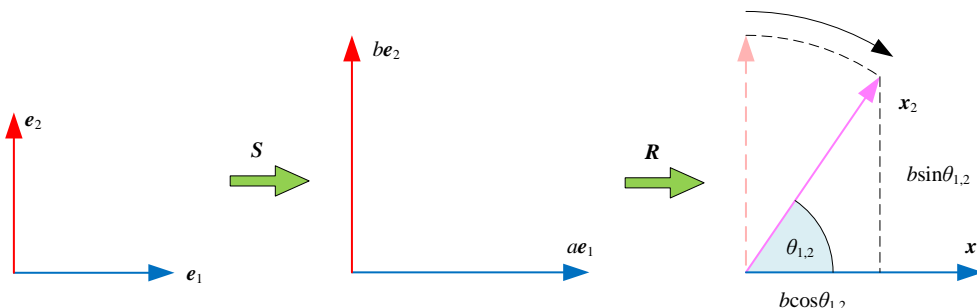


图 5. 先缩放再开合

计算 (24) 中，向量  $x_1$  和  $x_2$  夹角余弦值为：

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)

$$\cos \theta = \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{\|\mathbf{x}_1\| \|\mathbf{x}_2\|} = \frac{a \cdot b \cdot \cos \theta_{1,2}}{a \cdot b} = \cos \theta_{1,2} \quad (25)$$

发现向量  $\mathbf{x}_1$  和  $\mathbf{x}_2$  夹角等同于向量  $\mathbf{r}_1$  和  $\mathbf{r}_2$  夹角。

### 举个例子

给定  $\Sigma$  具体值为,

$$\Sigma = \begin{bmatrix} 1.5^2 & 1.5 \times 2 \times \cos 60^\circ \\ 1.5 \times 2 \times \cos 60^\circ & 2^2 \end{bmatrix} = \begin{bmatrix} 2.25 & 1.5 \\ 1.5 & 4 \end{bmatrix} \quad (26)$$

对  $\Sigma$  进行 Cholesky 分解得到:

$$\Sigma = (\mathbf{R}_\Sigma)^\top (\mathbf{R}_\Sigma) = \begin{bmatrix} 1.5 & 0 \\ 1 & 1.732 \end{bmatrix} \begin{bmatrix} 1.5 & 1 \\ 0 & 1.732 \end{bmatrix} \quad (27)$$

图 6 所示为  $\mathbf{e}_1$  和  $\mathbf{e}_2$  经过  $\mathbf{R}_\Sigma$  转换得到向量  $\mathbf{x}_1$  和  $\mathbf{x}_2$ :

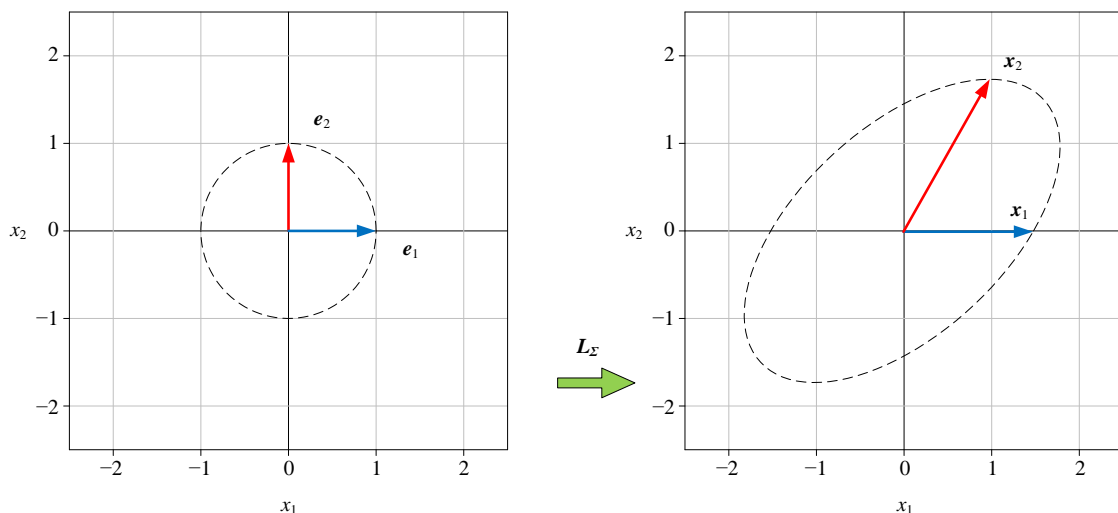


图 6.  $\mathbf{e}_1$  和  $\mathbf{e}_2$  经过  $\mathbf{R}_\Sigma$  转换得到向量  $\mathbf{x}_1$  和  $\mathbf{x}_2$

按照 (22),  $\Sigma$  可以分解成:

$$\Sigma = \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & 2 \end{bmatrix}}_S \underbrace{\begin{bmatrix} 1 & 0 \\ 0.5 & \sqrt{3}/2 \end{bmatrix}}_{R^\top} \underbrace{\begin{bmatrix} 1 & 0.5 \\ 0 & \sqrt{3}/2 \end{bmatrix}}_R \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & 2 \end{bmatrix}}_S \quad (28)$$

图 7 所示为  $\mathbf{e}_1$  和  $\mathbf{e}_2$  分别经过  $S$  和  $R$  转换, 得到向量  $\mathbf{x}_1$  和  $\mathbf{x}_2$ 。

本系列丛书一般用  $\Sigma$  来代表协方差矩阵。本节之所以用矩阵  $\Sigma$ ，这是因为大家很快会发现 Cholesky 分解和协方差矩阵之间的关系。而本章前文中提到的矩阵  $P$ ，就是本书之后要讲的相关性系数矩阵。类比的话，矩阵  $P$  中的余弦值就是相关性系数。

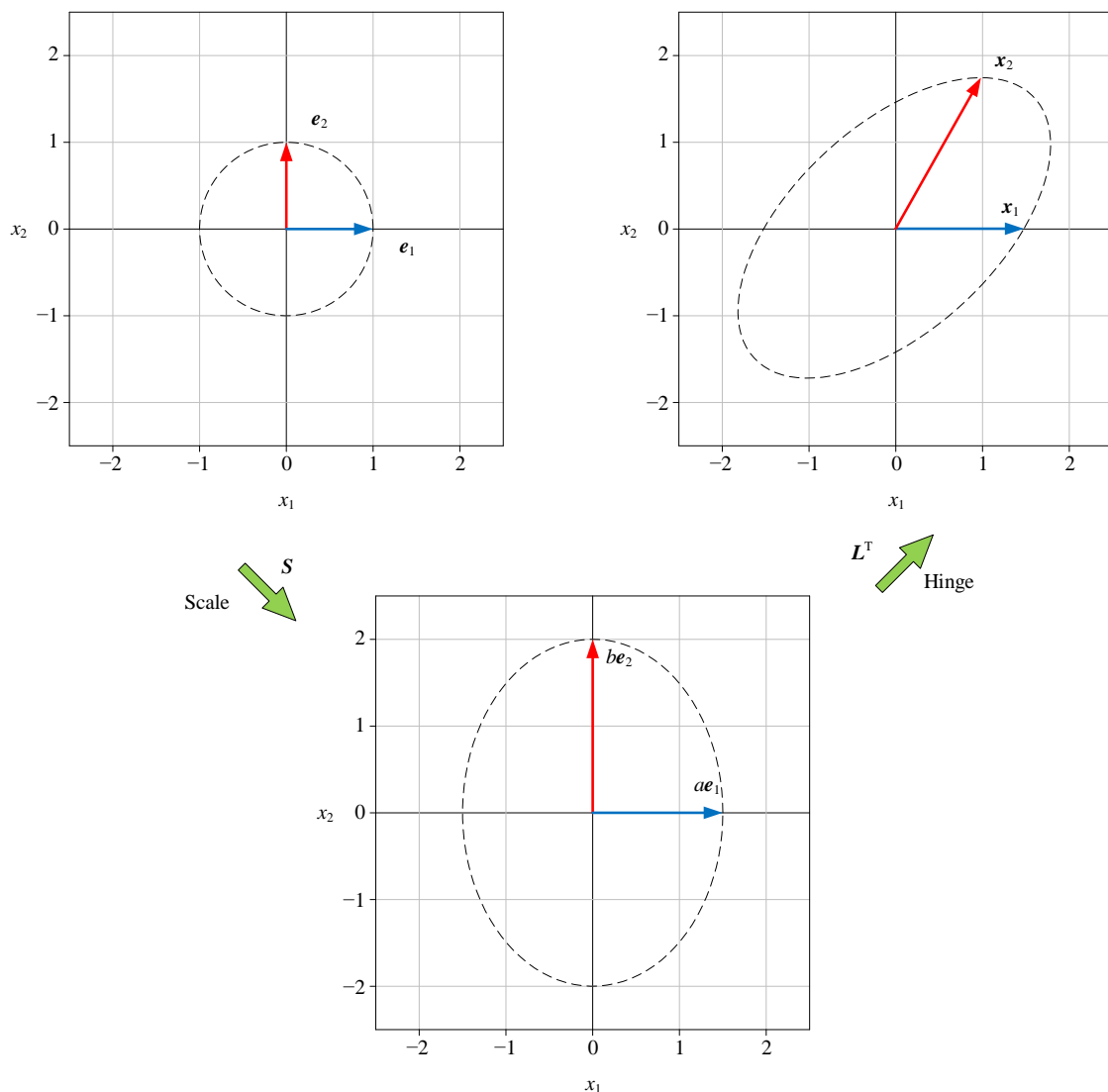


图 7.  $e_1$  和  $e_2$  分别经过  $S$  和  $R$  转换

## 12.5 推广到三维空间

本节利用立体几何视角探讨 Cholesky 分解。

给定如下  $3 \times 3$  矩阵  $P$ ,

$$\mathbf{P} = \begin{bmatrix} 1 & \cos \theta_{1,2} & \cos \theta_{1,3} \\ \cos \theta_{1,2} & 1 & \cos \theta_{2,3} \\ \cos \theta_{1,3} & \cos \theta_{2,3} & 1 \end{bmatrix} \quad (29)$$

其中,  $\theta_{1,2}$ 、 $\theta_{1,3}$ 、 $\theta_{2,3}$ 三个角度均大于等于  $0^\circ$ 。

对  $\mathbf{P}$  进行 Cholesky 分解:

$$\mathbf{P} = \mathbf{R}^T \mathbf{R} \quad (30)$$

其中,

$$\mathbf{R} = \begin{bmatrix} 1 & \cos \theta_{1,2} & \cos \theta_{1,3} \\ 0 & \sqrt{1 - \cos^2 \theta_{1,2}} & \frac{\cos \theta_{2,3} - \cos \theta_{1,3} \cos \theta_{1,2}}{\sqrt{1 - \cos^2 \theta_{1,2}}} \\ 0 & 0 & \sqrt{1 - \cos^2 \theta_{1,3} - \frac{(\cos \theta_{2,3} - \cos \theta_{1,3} \cos \theta_{1,2})^2}{1 - \cos^2 \theta_{1,2}}} \end{bmatrix} \quad (31)$$

相当于:

$$\mathbf{r}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{r}_2 = \begin{bmatrix} \cos \theta_{1,2} \\ \sqrt{1 - \cos^2 \theta_{1,2}} \\ 0 \end{bmatrix}, \quad \mathbf{r}_3 = \begin{bmatrix} \cos \theta_{1,3} \\ \frac{\cos \theta_{2,3} - \cos \theta_{1,3} \cos \theta_{1,2}}{\sqrt{1 - \cos^2 \theta_{1,2}}} \\ \sqrt{1 - \cos^2 \theta_{1,3} - \frac{(\cos \theta_{2,3} - \cos \theta_{1,3} \cos \theta_{1,2})^2}{1 - \cos^2 \theta_{1,2}}} \end{bmatrix} \quad (32)$$

将  $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$  代入 (30) 得到:

$$\mathbf{P} = \begin{bmatrix} 1 & \cos \theta_{1,2} & \cos \theta_{1,3} \\ \cos \theta_{1,2} & 1 & \cos \theta_{2,3} \\ \cos \theta_{1,3} & \cos \theta_{2,3} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^T \mathbf{r}_1 & \mathbf{r}_1^T \mathbf{r}_2 & \mathbf{r}_1^T \mathbf{r}_3 \\ \mathbf{r}_2^T \mathbf{r}_1 & \mathbf{r}_2^T \mathbf{r}_2 & \mathbf{r}_2^T \mathbf{r}_3 \\ \mathbf{r}_3^T \mathbf{r}_1 & \mathbf{r}_3^T \mathbf{r}_2 & \mathbf{r}_3^T \mathbf{r}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{r}_1 & \mathbf{r}_1 \cdot \mathbf{r}_2 & \mathbf{r}_1 \cdot \mathbf{r}_3 \\ \mathbf{r}_2 \cdot \mathbf{r}_1 & \mathbf{r}_2 \cdot \mathbf{r}_2 & \mathbf{r}_2 \cdot \mathbf{r}_3 \\ \mathbf{r}_3 \cdot \mathbf{r}_1 & \mathbf{r}_3 \cdot \mathbf{r}_2 & \mathbf{r}_3 \cdot \mathbf{r}_3 \end{bmatrix} \quad (33)$$

观察 (33) 对角线, 可以容易判断  $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$  均为单位向量, 但是  $[\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$  并非正交基。

而  $\mathbf{P}$  中非对角线元素  $\cos \theta_{i,j}$  就是  $\mathbf{r}_i$  和  $\mathbf{r}_j$  向量夹角的余弦值。下面验证一下。

计算向量  $\mathbf{r}_1$  和  $\mathbf{r}_2$  角度的余弦值:

$$\frac{\mathbf{r}_1 \cdot \mathbf{r}_2}{\|\mathbf{r}_1\| \|\mathbf{r}_2\|} = \cos \theta_{1,2} \quad (34)$$

计算向量  $\mathbf{r}_1$  和  $\mathbf{r}_3$  角度的余弦值:

$$\frac{\mathbf{r}_1 \cdot \mathbf{r}_3}{\|\mathbf{r}_1\| \|\mathbf{r}_3\|} = \cos \theta_{1,3} \quad (35)$$

计算向量  $\mathbf{r}_2$  和  $\mathbf{r}_3$  角度的余弦值：

$$\frac{\mathbf{r}_2 \cdot \mathbf{r}_3}{\|\mathbf{r}_2\| \|\mathbf{r}_3\|} = \cos \theta_{2,3} \quad (36)$$

## 几何视角

如图 8 所示，利用  $\mathbf{R}$ ，我们完成了标准正交基  $[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$  向非正交基  $[\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$  的转换。

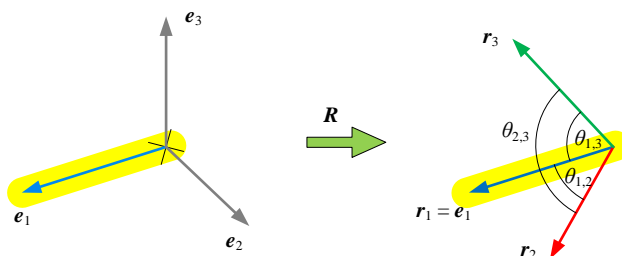


图 8. 三维系转化成满足指定两两夹角的坐标系

换个角度，(29) 中矩阵  $\mathbf{P}$  指定了目标向量两两夹角  $\cos \theta_{1,2}$ 、 $\cos \theta_{1,3}$ 、 $\cos \theta_{2,3}$ 。即  $\mathbf{r}_1$  和  $\mathbf{r}_2$  的夹角为  $\cos \theta_{1,2}$ ， $\mathbf{r}_1$  和  $\mathbf{r}_3$  的夹角为  $\cos \theta_{1,3}$ ， $\mathbf{r}_2$  和  $\mathbf{r}_3$  的夹角为  $\cos \theta_{2,3}$ 。我们想要找到空间中满足这个条件的三个单位向量。

对  $\mathbf{P}$  进行 Cholesky 分解得到矩阵  $\mathbf{R}$ ，它的列向量  $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$  就是我们想要找的三个向量的空间坐标点。特别地， $\mathbf{r}_1$  和  $\mathbf{e}_1$  相同。好就好比，在构造  $[\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$  这个非正交基时， $\mathbf{r}_1$  锚定在  $\mathbf{e}_1$ 。

## 两个例子

图 9 给出两个例子，在给定  $\cos \theta_{1,2}$ 、 $\cos \theta_{1,3}$ 、 $\cos \theta_{2,3}$  三个角度条件下，我们可以利用 Cholesky 分解矩阵  $\mathbf{P}$  计算得到满足夹角条件的三个单位向量  $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$ 。

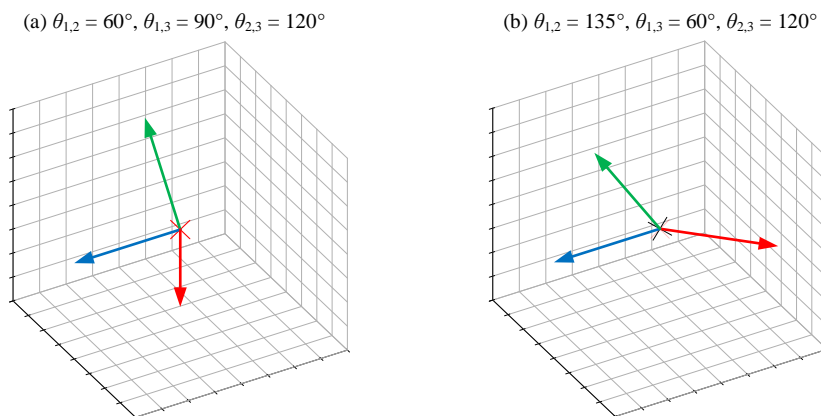


图 9. 给定三个夹角，确定向量三维空间位置

## 条件

在图 8 中，任意两个夹角必须大于等于第三个夹角，也就是必须满足如下三个不等式：

$$\begin{aligned}\theta_{1,2} + \theta_{1,3} &\geq \theta_{2,3} \\ \theta_{1,2} + \theta_{2,3} &\geq \theta_{1,3} \\ \theta_{1,3} + \theta_{2,3} &\geq \theta_{1,2}\end{aligned}\quad (37)$$

另外，三个角度夹角必须小于等于  $360^\circ$ ：

$$\theta_{1,2} + \theta_{1,3} + \theta_{2,3} \leq 360^\circ \quad (38)$$

试想一个有趣的现象，在图 8 中，如果  $\cos\theta_{1,2} = \cos\theta_{1,3} + \cos\theta_{2,3}$ ，这意味着  $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$  三个向量在一个平面上，也就是  $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$  线性相关。

这种情况，矩阵  $\mathbf{R}$  不满秩，也就是说  $\mathbf{P}$  也不满秩。正定矩阵满秩，也就是说这种情形  $\mathbf{P}$  不可以 Cholesky 分解。

而三个夹角之和等于  $360^\circ$  时，即  $\cos\theta_{1,2} + \cos\theta_{1,3} + \cos\theta_{2,3} = 360^\circ$ ， $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$  三个向量也在一个平面上， $\mathbf{P}$  也不可以 Cholesky 分解。

最后，如果  $\theta_{1,2}$ 、 $\theta_{1,3}$ 、 $\theta_{2,3}$  任一角度为  $0^\circ$ ，这意味着存在两个向量共线，这种情况  $\mathbf{P}$  也不可以 Cholesky 分解。

也就是为了保证  $\mathbf{P}$  可以 Cholesky 分解，即正定，需要满足以下条件：

$$\begin{aligned}\theta_{1,2} &> 0^\circ, \quad \theta_{1,3} > 0^\circ, \quad \theta_{2,3} > 0^\circ \\ \theta_{1,2} + \theta_{1,3} &> \theta_{2,3}, \quad \theta_{1,2} + \theta_{2,3} > \theta_{1,3}, \quad \theta_{1,3} + \theta_{2,3} > \theta_{1,2} \\ \theta_{1,2} + \theta_{1,3} + \theta_{2,3} &< 360^\circ\end{aligned}\quad (39)$$

## 夹角相同

再看一组特殊情况，向量两两夹角相同，即，

$$\theta_{1,2} = \theta_{1,3} = \theta_{2,3} \quad (40)$$

打个比方，这个例子像是一把雨伞的开合。假设雨伞只有三个伞骨，雨伞开合时，伞骨之间的两两夹角相等。

雨伞合起来时，三个伞骨并拢，相当于三个向量之间夹角为  $0^\circ$ ，即共线。

如果雨伞最大开度可以让伞面为平面，这时三个伞骨之间夹角为  $120^\circ$ ，也就是三个向量在一个平面上。

有了这两个极限情况，我们知道向量之间夹角取值范围为  $[0^\circ, 120^\circ]$ 。

图 10 给出四个不同开合角度。实际上，图 10 (d) 对应的矩阵  $\mathbf{P}$  是不能进行 Cholesky 分解，这是因为  $\theta_{1,2}$ 、 $\theta_{1,3}$ 、 $\theta_{2,3}$  三个角度都是  $120^\circ$ ，因此  $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$  在一个平面上，线性相关。

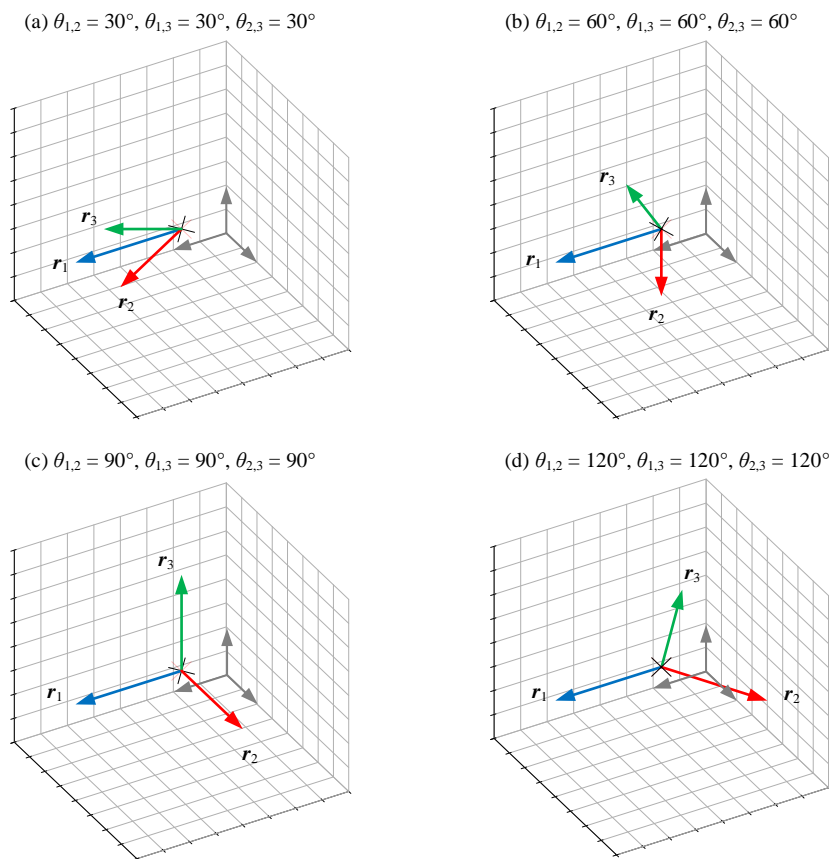


图 10. 相等角度条件下，确定向量三维空间位置

至此，我们利用空间几何视角，探讨了 Cholesky 分解以及满足 Cholesky 分解条件。





Bk4\_Ch12\_02.py 绘制图 9 和图 10。请读者自行设定夹角条件，看看哪些角度组合能够进行 Cholesky 分解，哪些不能。

```
# Bk4_Ch12_02.py

import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np
import math as m

cos_theta_12 = np.cos(m.radians(135))
cos_theta_13 = np.cos(m.radians(60))
cos_theta_23 = np.cos(m.radians(120))

P = np.array([[1, cos_theta_12, cos_theta_13],
              [cos_theta_12, 1, cos_theta_23],
              [cos_theta_13, cos_theta_23, 1]])

L = np.linalg.cholesky(P)
R = L.T

fig = plt.figure()
ax = fig.add_subplot(projection='3d')

plt.plot(0,0,0,color = 'r', marker = 'x',
         markersize = 12)

colors = ['b', 'r', 'g']
for i in np.arange(0,3):

    vector = R[:,i]
    v = np.array([vector[0],vector[1],vector[2]])
    vlength=np.linalg.norm(v)
    ax.quiver(0,0,0,vector[0],vector[1],vector[2],
             length=vlength, color = colors[i])

ax.set_xlim([-1,1])
ax.set_ylim([-1,1])
ax.set_zlim([-1,1])

ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('z')

ax.xaxis.set_ticklabels([])
ax.yaxis.set_ticklabels([])
ax.zaxis.set_ticklabels([])
ax.view_init(35, 60)
ax.set_proj_type('ortho')
ax.set_box_aspect([1,1,1])
```

## 12.6 从格拉姆矩阵到相似度矩阵

有了本章前文内容铺垫，下面我们回头来看一下本书前文介绍的一个重要的概念——格拉姆矩阵。

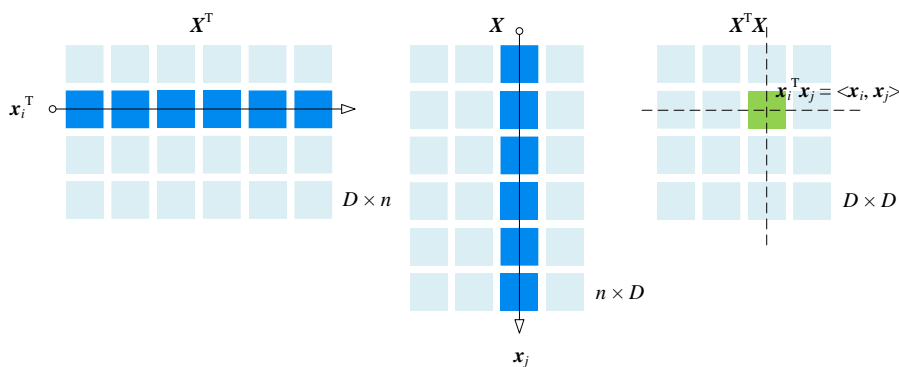


图 11. 格拉姆矩阵

如图 11 所示，数据矩阵  $X$  的格拉姆矩阵  $G$  可以写成标量积形式：

$$G = \begin{bmatrix} \mathbf{x}_1 \cdot \mathbf{x}_1 & \mathbf{x}_1 \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_1 \cdot \mathbf{x}_D \\ \mathbf{x}_2 \cdot \mathbf{x}_1 & \mathbf{x}_2 \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_2 \cdot \mathbf{x}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_D \cdot \mathbf{x}_1 & \mathbf{x}_D \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_D \cdot \mathbf{x}_D \end{bmatrix} = \begin{bmatrix} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle & \langle \mathbf{x}_1, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_1, \mathbf{x}_D \rangle \\ \langle \mathbf{x}_2, \mathbf{x}_1 \rangle & \langle \mathbf{x}_2, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_2, \mathbf{x}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{x}_D, \mathbf{x}_1 \rangle & \langle \mathbf{x}_D, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_D, \mathbf{x}_D \rangle \end{bmatrix} \quad (41)$$

### 确定列向量坐标

对  $G$  进行 Cholesky 分解得到：

$$G = R_G^T R_G \quad (42)$$

将  $R_G$  写成列向量：

$$R_G = [\mathbf{r}_{G,1} \quad \mathbf{r}_{G,2} \quad \cdots \quad \mathbf{r}_{G,D}] \quad (43)$$

将 (43) 代入 (42) 得到：

$$G = \begin{bmatrix} \mathbf{r}_{G,1}^T \\ \mathbf{r}_{G,2}^T \\ \vdots \\ \mathbf{r}_{G,D}^T \end{bmatrix} [\mathbf{r}_{G,1} \quad \mathbf{r}_{G,2} \quad \cdots \quad \mathbf{r}_{G,D}] = \begin{bmatrix} \langle \mathbf{r}_{G,1}, \mathbf{r}_{G,1} \rangle & \langle \mathbf{r}_{G,1}, \mathbf{r}_{G,2} \rangle & \cdots & \langle \mathbf{r}_{G,1}, \mathbf{r}_{G,D} \rangle \\ \langle \mathbf{r}_{G,2}, \mathbf{r}_{G,1} \rangle & \langle \mathbf{r}_{G,2}, \mathbf{r}_{G,2} \rangle & \cdots & \langle \mathbf{r}_{G,2}, \mathbf{r}_{G,D} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{r}_{G,D}, \mathbf{r}_{G,1} \rangle & \langle \mathbf{r}_{G,D}, \mathbf{r}_{G,2} \rangle & \cdots & \langle \mathbf{r}_{G,D}, \mathbf{r}_{G,D} \rangle \end{bmatrix} \quad (44)$$

(41) 等价于 (44)，向量模和向量夹角之间完全等价。这相当于我们在  $\mathbb{R}^D$  中找到了  $X$  每个列向量的具体坐标！

以鸢尾花数据矩阵  $X$  为例， $X$  可以写成四个列向量左右排列，即  $X = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4]$ 。这些列向量都有 150 个元素，显然不能直接在  $\mathbb{R}^4$  空间中展示。

图 12 所示为计算  $X$  的 Gram 矩阵  $G$  过程热图。如前文所述，矩阵  $G$  中包含了  $[x_1, x_2, x_3, x_4]$  各个列向量的模，以及它们之间两两夹角。一个向量就两个元素——大小和方向， $G$  这相当于集成了  $[x_1, x_2, x_3, x_4]$  每个向量关键信息。

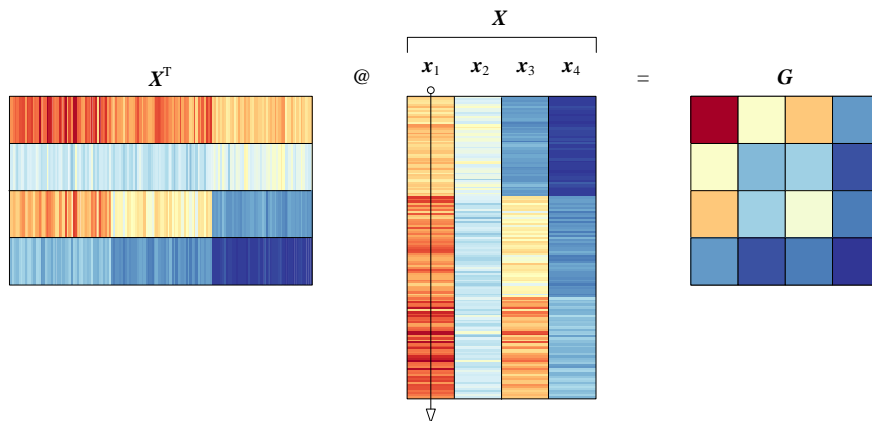


图 12. 鸢尾花数据格拉姆矩阵

如图 13 所示，对 Gram 矩阵  $G$  进行 Cholesky 分解得到上三角矩阵  $R_G$ ， $R_G$  的列向量长度为 4，它们在  $\mathbb{R}^4$  空间中，等价于  $[x_1, x_2, x_3, x_4]$ 。

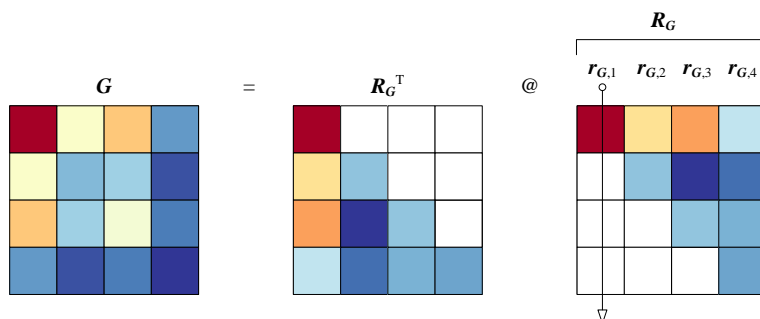


图 13. 对格拉姆矩阵  $G$  进行 Cholesky 分解

## 向量夹角

以向量夹角余弦形式展开  $G$  中向量积：

$$G = \begin{bmatrix} \|x_1\| \|x_1\| \cos \theta_{1,1} & \|x_1\| \|x_2\| \cos \theta_{2,1} & \cdots & \|x_1\| \|x_D\| \cos \theta_{1,D} \\ \|x_2\| \|x_1\| \cos \theta_{1,2} & \|x_2\| \|x_2\| \cos \theta_{2,2} & \cdots & \|x_2\| \|x_D\| \cos \theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|x_D\| \|x_1\| \cos \theta_{1,D} & \|x_D\| \|x_2\| \cos \theta_{2,D} & \cdots & \|x_D\| \|x_D\| \cos \theta_{D,D} \end{bmatrix} \quad (45)$$

观察矩阵  $G$ ，它包含了数据矩阵  $X$  中列向量的两个重要信息——模  $\|\mathbf{x}_i\|$ 、方向（向量两两夹角  $\cos\theta_{ij}$ ）。

定义缩放矩阵  $S$ ，具体形式如下：

$$S = \begin{bmatrix} \|\mathbf{x}_1\| & & & \\ & \|\mathbf{x}_2\| & & \\ & & \ddots & \\ & & & \|\mathbf{x}_D\| \end{bmatrix} \quad (46)$$

对  $G$  左右乘上  $S$  矩阵的逆，得到矩阵  $C$ ：

$$C = S^{-1}GS^{-1} = \begin{bmatrix} \frac{\mathbf{x}_1 \cdot \mathbf{x}_1}{\|\mathbf{x}_1\|\|\mathbf{x}_1\|} & \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|} & \cdots & \frac{\mathbf{x}_1 \cdot \mathbf{x}_D}{\|\mathbf{x}_1\|\|\mathbf{x}_D\|} \\ \frac{\mathbf{x}_2 \cdot \mathbf{x}_1}{\|\mathbf{x}_2\|\|\mathbf{x}_1\|} & \frac{\mathbf{x}_2 \cdot \mathbf{x}_2}{\|\mathbf{x}_2\|\|\mathbf{x}_2\|} & \cdots & \frac{\mathbf{x}_2 \cdot \mathbf{x}_D}{\|\mathbf{x}_2\|\|\mathbf{x}_D\|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbf{x}_D \cdot \mathbf{x}_1}{\|\mathbf{x}_D\|\|\mathbf{x}_1\|} & \frac{\mathbf{x}_D \cdot \mathbf{x}_2}{\|\mathbf{x}_D\|\|\mathbf{x}_2\|} & \cdots & \frac{\mathbf{x}_D \cdot \mathbf{x}_D}{\|\mathbf{x}_D\|\|\mathbf{x}_D\|} \end{bmatrix} \quad (47)$$

### 余弦相似度矩阵

矩阵  $C$  有自己的名字——余弦相似度矩阵 (cosine similarity matrix)。这是因为  $C$  的每个元素实际上计算的是  $\mathbf{x}_i$  和  $\mathbf{x}_j$  向量的夹角  $\theta_{ij}$  余弦值  $\cos\theta_{ij}$ ，即，

$$C = \begin{bmatrix} 1 & \cos\theta_{2,1} & \cdots & \cos\theta_{1,D} \\ \cos\theta_{1,2} & 1 & \cdots & \cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \cos\theta_{1,D} & \cos\theta_{2,D} & \cdots & 1 \end{bmatrix} \quad (48)$$

相比格拉姆矩阵  $G$ ，余弦相似度矩阵  $C$  中只包含了  $X$  列向量两两夹角  $\cos\theta_{ij}$  这个单一信息。

对  $C$  进行 Cholesky 分解得到：

$$C = LL^T = R^T R \quad (49)$$

将  $R$  写成  $[\mathbf{r}_1, \mathbf{r}_1, \dots, \mathbf{r}_D]$ ， $C$  可以写成：

$$C = R^T R = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \vdots \\ \mathbf{r}_D^T \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_D \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^T \mathbf{r}_1 & \mathbf{r}_1^T \mathbf{r}_2 & \cdots & \mathbf{r}_1^T \mathbf{r}_D \\ \mathbf{r}_2^T \mathbf{r}_1 & \mathbf{r}_2^T \mathbf{r}_2 & \cdots & \mathbf{r}_2^T \mathbf{r}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{r}_D^T \mathbf{r}_1 & \mathbf{r}_D^T \mathbf{r}_2 & \cdots & \mathbf{r}_D^T \mathbf{r}_D \end{bmatrix} = \begin{bmatrix} 1 & \cos\theta_{2,1} & \cdots & \cos\theta_{1,D} \\ \cos\theta_{1,2} & 1 & \cdots & \cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \cos\theta_{1,D} & \cos\theta_{2,D} & \cdots & 1 \end{bmatrix} \quad (50)$$

根据本章前文分析，我们知道  $\mathbf{r}_1, \mathbf{r}_1, \dots, \mathbf{r}_D$  都是单位向量。 $\mathbf{r}_1, \mathbf{r}_1, \dots, \mathbf{r}_D$  两两夹角余弦值满足相似度矩阵  $C$  的要求。

这一重要性质在蒙特卡洛模拟 (Monte Carlo simulation) 中有重要应用。如图 14，本章介绍的内容可以用来产生满足指定相关性系数要求的随机数。

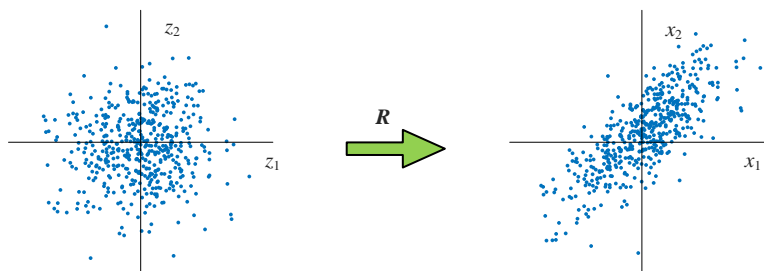
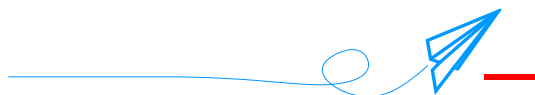


图 14. 产生满足指定相关性矩阵要求的随机数



本章从几何视角讲解了 Cholesky 分解。只有正定矩阵才可以进行 Cholesky 分解，这一点可以用来判断矩阵是否为正定。我们创造了“开合”这个词用来描述 Cholesky 分解得到的上三角矩阵对应的几何变换。

对 Gram 矩阵进行 Cholesky 分解可以帮我们确定原数据矩阵的列向量空间坐标。此外，我们将在后续协方差矩阵和蒙特卡罗模拟中再聊到 Cholesky 分解。

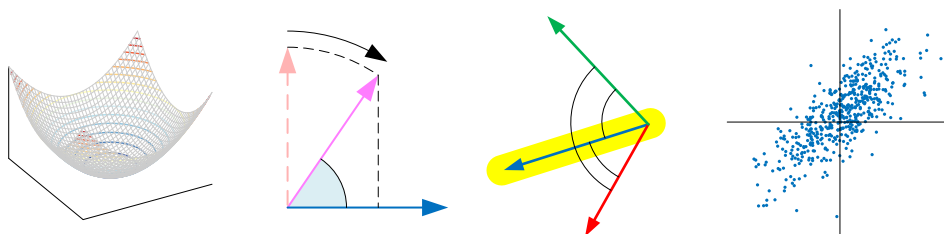


图 15. 总结本章重要内容的四副图