

# 12

## Cholesky Decomposition

# Cholesky 分解

适用于正定矩阵



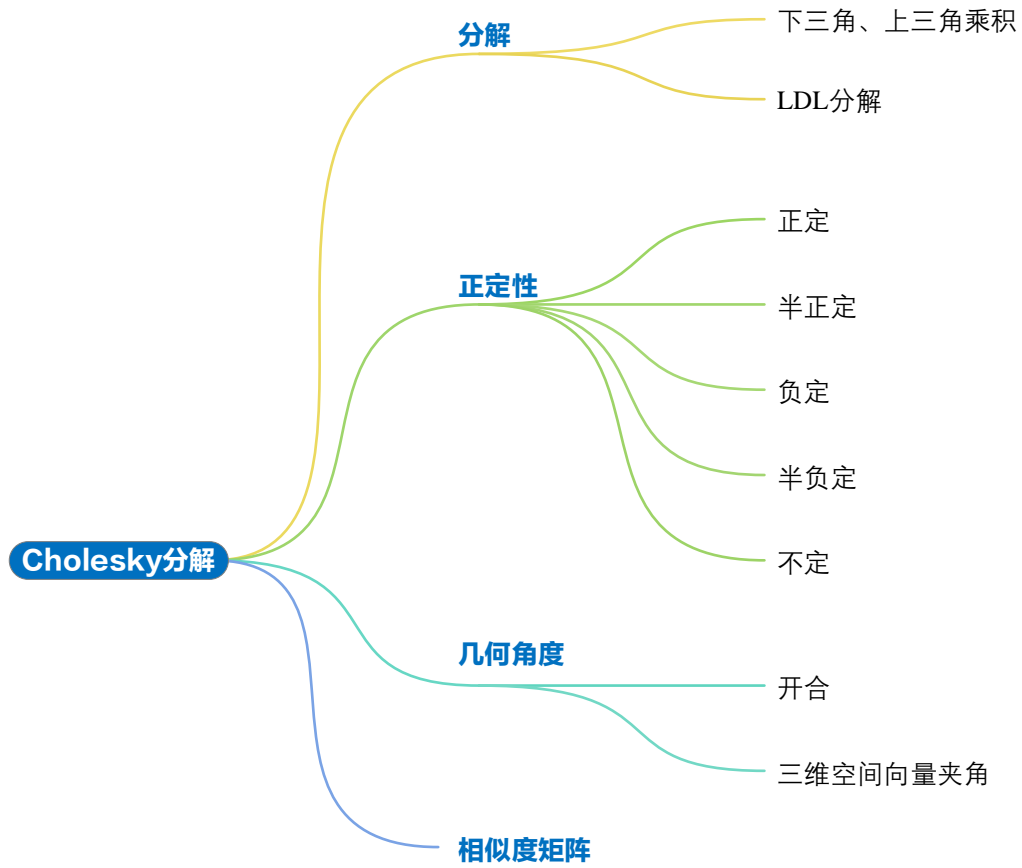
每个人都是天才。但是，如果你以爬树的能力来评判一条鱼，那么那条鱼终其一生都会认为自己愚蠢无能。

***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()` 计算反余弦
- ◀ `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



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)

## 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：

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

利用上三角矩阵  $R (=L^T)$ ，(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，

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

(3) 中下三角矩阵  $L$  为对角线元素均为 1。从几何视角来看， $L$  相当于我们在本书第 8 章中提到剪切。

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：

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

$LB$  相当于  $A$  的平方根。

用上三角矩阵  $R$  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

上一章提到，并非所有矩阵都可以做 Cholesky 分解，只有**正定矩阵** (positive-definite matrix) 才能 Cholesky 分解。

在  $\mathbf{x}$  为非零列向量 ( $\mathbf{x} \neq \mathbf{0}$ ) 条件下，如果方阵  $\mathbf{A}$  满足：

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \quad (8)$$

则称方阵  $\mathbf{A}$  为**正定矩阵** (positive definite matrix)。(8) 中列向量  $\mathbf{x}$  的行数和矩阵  $\mathbf{A}$  行数一致。二次型  $\mathbf{x}^T \mathbf{A} \mathbf{x}$  的结果是标量。此外，正定矩阵的特征值均为正。

### 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：

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad (9)$$

注意，正定矩阵都是对称方阵。

定义二元函数  $y = f(x_1, x_2)$ ：

$$y = f(x_1, x_2) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = ax_1^2 + 2bx_1x_2 + cx_2^2 \quad (10)$$

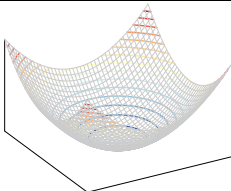
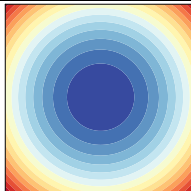
函数  $y = f(x_1, x_2)$  就是本书第 5 章提到的二次型。更重要的是，上式把正定性和丛书《数学要素》讲过的二次曲面联系起来。

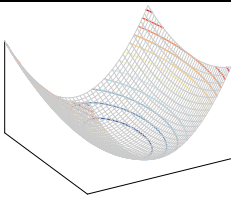
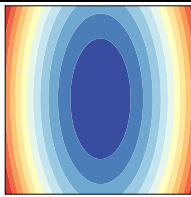
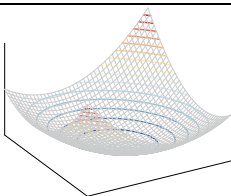
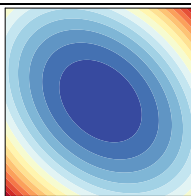
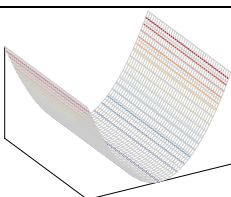
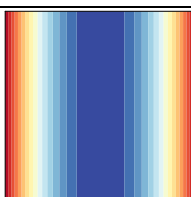
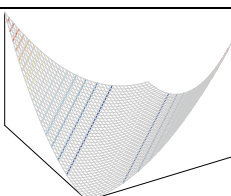
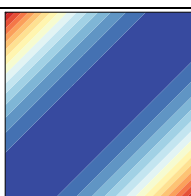
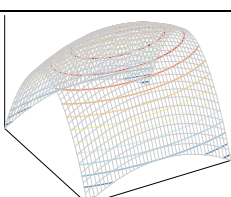
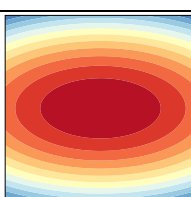
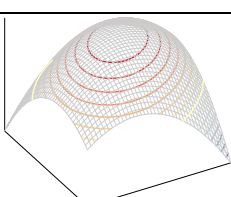
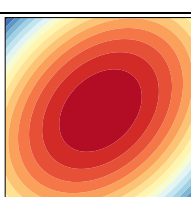
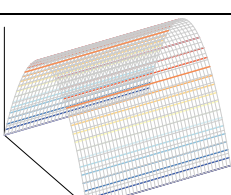
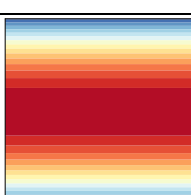
除了正定矩阵，还有半正定、负定、半负定、不定这几种正定性。表 1 总结几种正定性、曲面、等高线特征。希望读者能够通过表中几何图形建立正定性的直观印象。此外，请大家自行分析表中曲面的极值特征。

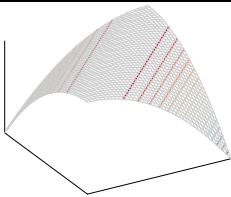
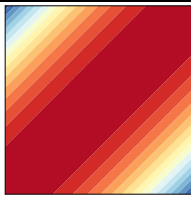
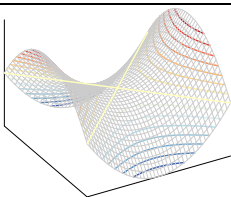
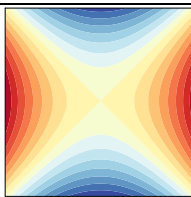
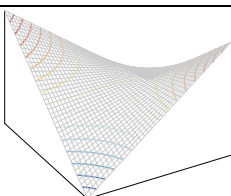
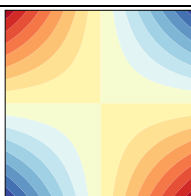


本书第 21 章将专门讨论矩阵的正定性。

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 $A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$		

	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。

## 12.3 几何角度：开合

本节，我们从一个有趣的几何视角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：

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

利用上三角矩阵  $R$ ，矩阵  $P$  的 Cholesky 分解还可以写成：

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

将  $\mathbf{R}$  写成：

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

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， $[\mathbf{e}_1, \mathbf{e}_2]$  是  $\mathbb{R}^2$  空间的标准正交基。 $\mathbf{R}$  分别乘  $\mathbf{e}_1$  和  $\mathbf{e}_2$ ，得到  $\mathbf{r}_1$  和  $\mathbf{r}_2$ ：

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

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  $\mathbf{r}_2$  夹角余弦值正是  $\cos \theta_{1,2}$ ：

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

## 几何视角

如图 1 所示，从几何角度来讲， $\mathbf{R}$  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”。

**⚠** 再次强调，虽然  $[\mathbf{r}_1, \mathbf{r}_2]$  中每个列向量为单位向量，但是并不正交，因此  $[\mathbf{r}_1, \mathbf{r}_2]$  为非正交基。

如图 1 所示， $[\mathbf{e}_1, \mathbf{e}_2]$  的夹角为 90 度，经过  $\mathbf{R}$  变换后， $[\mathbf{r}_1, \mathbf{r}_2]$  的夹角变成  $\theta_{1,2}$ 。这种几何变换像是“门合页”的开合。我们给这种几何变换取个名字，就叫做“开合”。

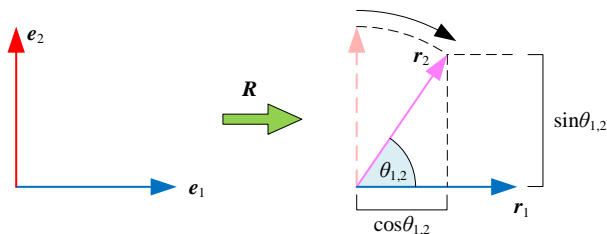


图 1. 开合

图 2 所示为四种不同开合角度。 $0 < \cos\theta_{1,2} < 1$  时，即  $0^\circ < \theta_{1,2} < 90^\circ$ ，“门合页”从直角  $90^\circ$  关闭至  $\theta_{1,2}$ ，具体如图 2 (a) (b) 所示两例。

$-1 < \cos\theta_{1,2} < 0$  时，即  $90^\circ < \theta_{1,2} < 180^\circ$ ，“合页”从直角  $90^\circ$  打开至  $\theta_{1,2}$ ，具体如图 2 (c) (d) 所示两例。

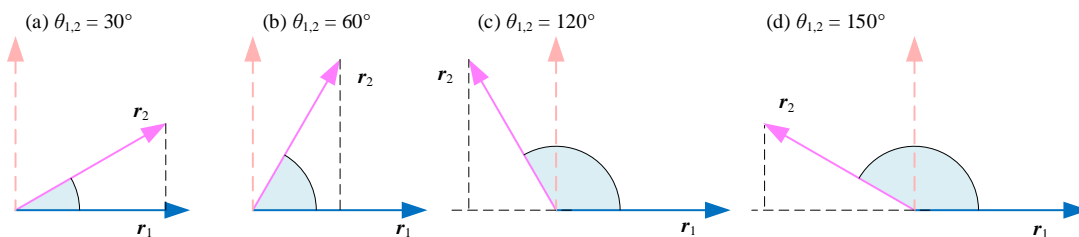


图 2. 不同的开合角度  $\cos\theta_{1,2}$

## 行列式值

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，图形的面积缩放比例为  $\sin\theta_{1,2}$ 。这和我们图 3 中看到一致。 $[\mathbf{e}_1, \mathbf{e}_2]$  构造正方形面积为 1，而  $[\mathbf{r}_1, \mathbf{r}_2]$  构造的平行四边形面积为  $\sin\theta_{1,2}$ 。

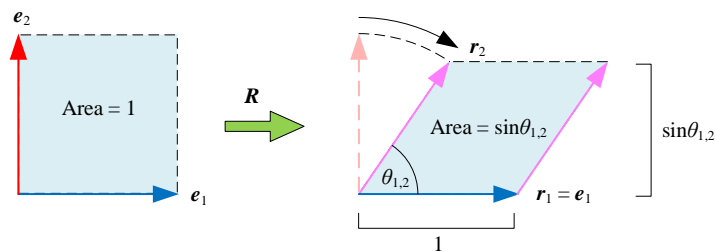


图 3. 开合对应的面积变化

## 举个例子

给定  $\mathbf{P}$  为：

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

对  $\mathbf{P}$  进行 Cholesky 分解得到：



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

图 4 所示为  $e_1$  和  $e_2$  经过 (19) 中  $R$  转换得到向量  $r_1$  和  $r_2$ ，而正圆经过  $R$  转换变成旋转椭圆。大家可能会问这个旋转椭圆的半长轴和半短轴长度分别为多少，这就需要借助特征值分解来计算。

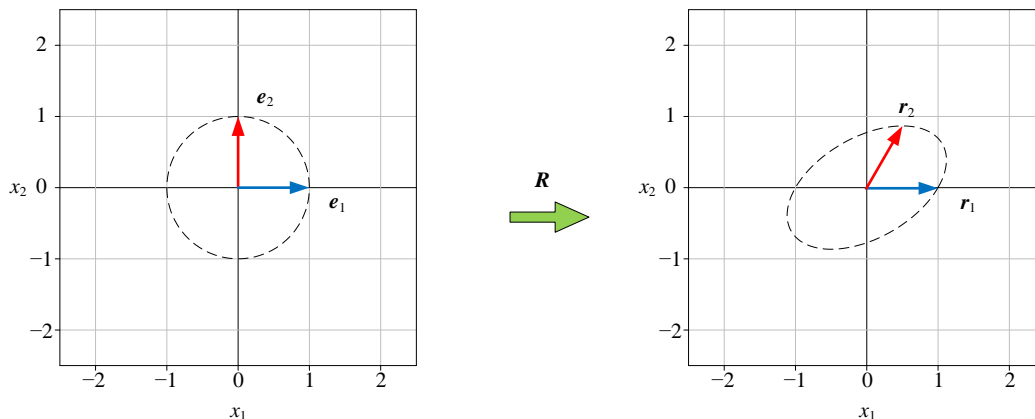


图 4.  $e_1$  和  $e_2$  经过  $R$  转换得到向量  $r_1$  和  $r_2$

## 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。

将  $RS$  ( $S$  先、 $R$  后) 作用在在  $e_1$  和  $e_2$  上, 得到  $x_1$  和  $x_2$ :

$$\begin{aligned} x_1 &= RS e_1 = \begin{bmatrix} 1 & \cos \theta_{1,2} \\ 0 & \sin \theta_{1,2} \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ x_2 &= RS e_2 = \begin{bmatrix} 1 & \cos \theta_{1,2} \\ 0 & \sin \theta_{1,2} \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = b \begin{bmatrix} \cos \theta_{1,2} \\ \sin \theta_{1,2} \end{bmatrix} \end{aligned} \quad (23)$$

这相当于, 对  $e_1$  和  $e_2$  先缩放 ( $S$ ), 再开合 ( $R$ )。

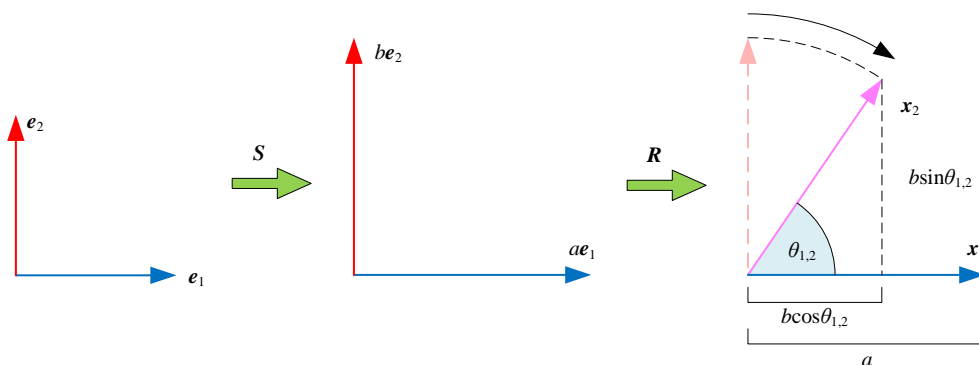


图 5. 先缩放再开合

计算 (23) 中, 向量  $x_1$  和  $x_2$  夹角余弦值为:

$$\cos \theta = \frac{x_1 \cdot x_2}{\|x_1\| \|x_2\|} = \frac{a \cdot b \cdot \cos \theta_{1,2}}{a \cdot b} = \cos \theta_{1,2} \quad (24)$$

发现向量  $x_1$  和  $x_2$  夹角等同于向量  $r_1$  和  $r_2$  夹角。

### 举个例子

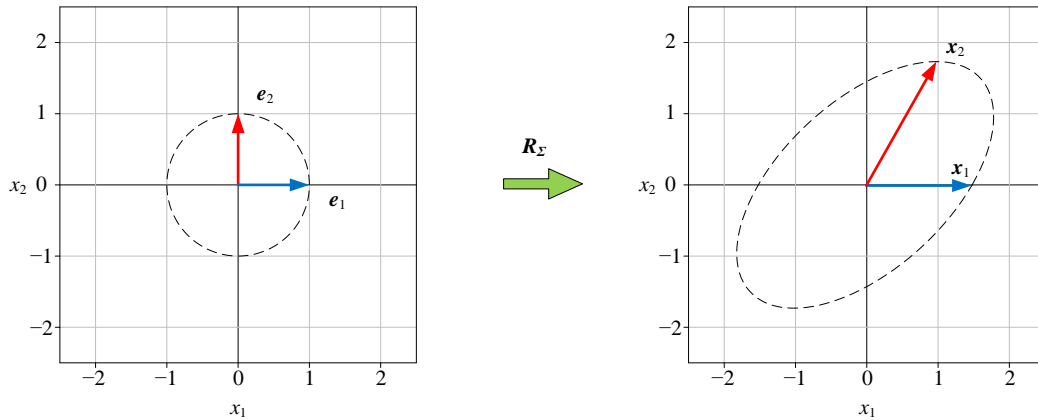
给定  $\Sigma$  具体值为:

$$\Sigma = \begin{bmatrix} 1.5^2 & 1.5 \times 2 \times \cos 60^\circ \\ 1.5 \times 2 \times \cos 60^\circ & 2^2 \end{bmatrix} = \begin{bmatrix} 2.25 & 1.5 \\ 1.5 & 4 \end{bmatrix} \quad (25)$$

对  $\Sigma$  进行 Cholesky 分解得到:

$$\Sigma = (R_\Sigma)^T (R_\Sigma) = \begin{bmatrix} 1.5 & 0 \\ 1 & 1.732 \end{bmatrix} \begin{bmatrix} 1.5 & 1 \\ 0 & 1.732 \end{bmatrix} \quad (26)$$

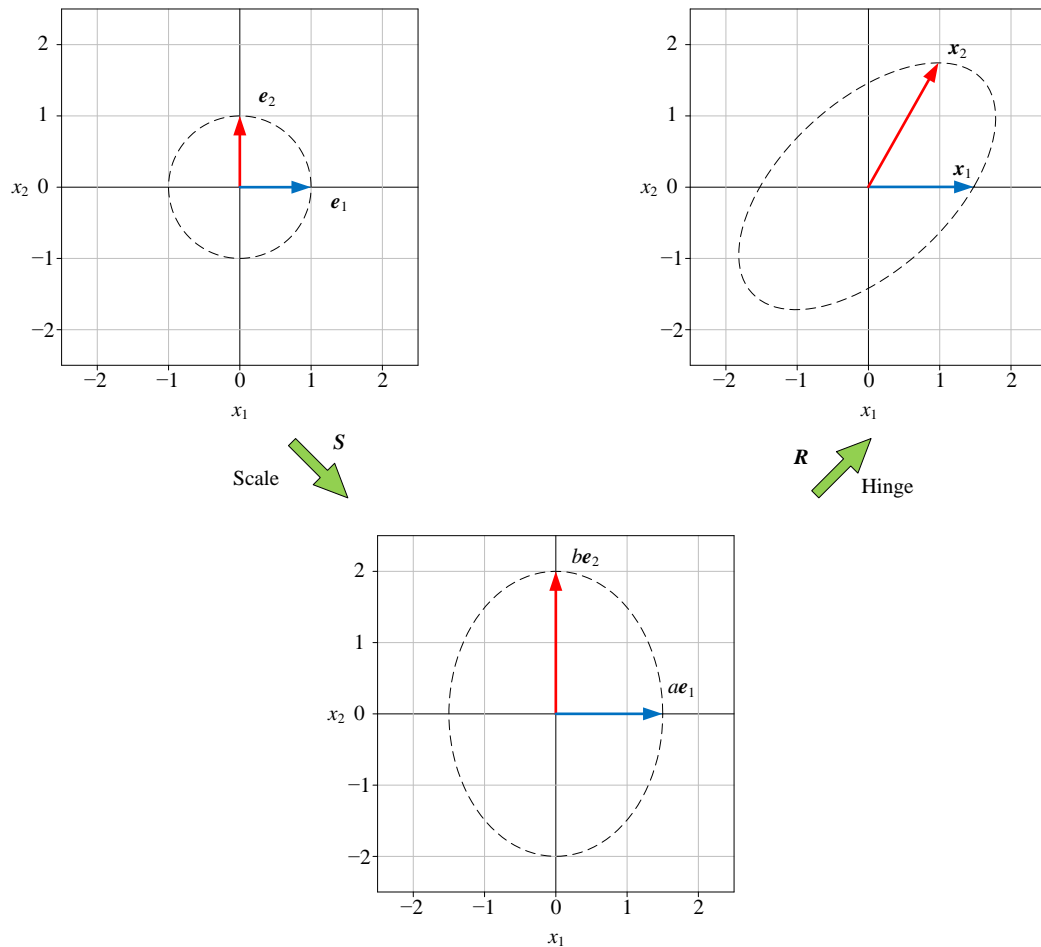
图 6 所示为  $e_1$  和  $e_2$  经过  $R_\Sigma$  转换得到向量  $x_1$  和  $x_2$ 。

图 6.  $e_1$  和  $e_2$  经过  $R_T$  转换得到向量  $x_1$  和  $x_2$ 

按照 (22),  $\Sigma$  可以分解成:

$$\Sigma = \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & 2 \end{bmatrix}}_S \underbrace{\begin{bmatrix} 1 & 0 \\ 0.5 & \sqrt{3}/2 \end{bmatrix}}_{R^T} \underbrace{\begin{bmatrix} 1 & 0.5 \\ 0 & \sqrt{3}/2 \end{bmatrix}}_R \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & 2 \end{bmatrix}}_S = (RS)^T RS \quad (27)$$

图 7 所示为  $e_1$  和  $e_2$  分别经过  $S$  和  $R$  转换, 得到向量  $x_1$  和  $x_2$ 。



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)

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

对 (25) 中  $\Sigma$  进行 LDL 分解：

$$\Sigma = \begin{bmatrix} 2.25 & 1.5 \\ 1.5 & 4 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 2/3 & 1 \end{bmatrix}}_L \underbrace{\begin{bmatrix} 2.25 & 0 \\ 0 & 3 \end{bmatrix}}_D \underbrace{\begin{bmatrix} 1 & 2/3 \\ 0 & 1 \end{bmatrix}}_{L^T} \quad (28)$$

将对角矩阵  $D$  写成  $BB$ ,

$$\Sigma = \begin{bmatrix} 2.25 & 1.5 \\ 1.5 & 4 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 2/3 & 1 \end{bmatrix}}_L \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & \sqrt{3} \end{bmatrix}}_B \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & \sqrt{3} \end{bmatrix}}_B \underbrace{\begin{bmatrix} 1 & 2/3 \\ 0 & 1 \end{bmatrix}}_{L^T} = (BL^T)^T BL^T \quad (29)$$

上式中的  $L$  对应的几何变换为剪切,  $B$  对应缩放。也就是说, 先剪切 ( $L^T$ ) 再缩放 ( $B$ ), 我们也可以获得图 6 右图。请大家自行绘制分步几何变换图像。

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紧密联系。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。

## 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  $P$ ,

$$P = \begin{bmatrix} 1 & \cos \theta_{1,2} & \cos \theta_{1,3} \\ \cos \theta_{1,2} & 1 & \cos \theta_{2,3} \\ \cos \theta_{1,3} & \cos \theta_{2,3} & 1 \end{bmatrix} \quad (30)$$

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

对  $P$  进行 Cholesky 分解：

$$P = R^T R \quad (31)$$

其中,

$$R = \begin{bmatrix} 1 & \cos \theta_{1,2} & \cos \theta_{1,3} \\ 0 & \sqrt{1 - \cos^2 \theta_{1,2}} & \frac{\cos \theta_{2,3} - \cos \theta_{1,3} \cos \theta_{1,2}}{\sqrt{1 - \cos^2 \theta_{1,2}}} \\ 0 & 0 & \sqrt{1 - \cos^2 \theta_{1,3} - \frac{(\cos \theta_{2,3} - \cos \theta_{1,3} \cos \theta_{1,2})^2}{1 - \cos^2 \theta_{1,2}}} \end{bmatrix} \quad (32)$$

相当于：

$$\mathbf{r}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{r}_2 = \begin{bmatrix} \cos \theta_{1,2} \\ \sqrt{1 - \cos^2 \theta_{1,2}} \\ 0 \end{bmatrix}, \quad \mathbf{r}_3 = \begin{bmatrix} \cos \theta_{1,3} \\ \frac{\cos \theta_{2,3} - \cos \theta_{1,3} \cos \theta_{1,2}}{\sqrt{1 - \cos^2 \theta_{1,2}}} \\ \sqrt{1 - \cos^2 \theta_{1,3} - \frac{(\cos \theta_{2,3} - \cos \theta_{1,3} \cos \theta_{1,2})^2}{1 - \cos^2 \theta_{1,2}}} \end{bmatrix} \quad (33)$$

将  $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$  代入 (31) 得到：

$$\mathbf{P} = \begin{bmatrix} 1 & \cos \theta_{1,2} & \cos \theta_{1,3} \\ \cos \theta_{1,2} & 1 & \cos \theta_{2,3} \\ \cos \theta_{1,3} & \cos \theta_{2,3} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^T \mathbf{r}_1 & \mathbf{r}_1^T \mathbf{r}_2 & \mathbf{r}_1^T \mathbf{r}_3 \\ \mathbf{r}_2^T \mathbf{r}_1 & \mathbf{r}_2^T \mathbf{r}_2 & \mathbf{r}_2^T \mathbf{r}_3 \\ \mathbf{r}_3^T \mathbf{r}_1 & \mathbf{r}_3^T \mathbf{r}_2 & \mathbf{r}_3^T \mathbf{r}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{r}_1 & \mathbf{r}_1 \cdot \mathbf{r}_2 & \mathbf{r}_1 \cdot \mathbf{r}_3 \\ \mathbf{r}_2 \cdot \mathbf{r}_1 & \mathbf{r}_2 \cdot \mathbf{r}_2 & \mathbf{r}_2 \cdot \mathbf{r}_3 \\ \mathbf{r}_3 \cdot \mathbf{r}_1 & \mathbf{r}_3 \cdot \mathbf{r}_2 & \mathbf{r}_3 \cdot \mathbf{r}_3 \end{bmatrix} \quad (34)$$

观察 (34) 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  $[\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$  为非正交基。

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  $\mathbf{r}_2$  夹角的余弦值：

$$\frac{\mathbf{r}_1 \cdot \mathbf{r}_2}{\|\mathbf{r}_1\| \|\mathbf{r}_2\|} = \cos \theta_{1,2} \quad (35)$$

$\mathbf{r}_1$  和  $\mathbf{r}_3$  夹角的余弦值为：

$$\frac{\mathbf{r}_1 \cdot \mathbf{r}_3}{\|\mathbf{r}_1\| \|\mathbf{r}_3\|} = \cos \theta_{1,3} \quad (36)$$

$\mathbf{r}_2$  和  $\mathbf{r}_3$  夹角的余弦值为：

$$\frac{\mathbf{r}_2 \cdot \mathbf{r}_3}{\|\mathbf{r}_2\| \|\mathbf{r}_3\|} = \cos \theta_{2,3} \quad (37)$$

## 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。

换个角度，(30) 中矩阵  $\mathbf{P}$  指定了目标向量两两“相对夹角”余弦值  $\cos \theta_{1,2}$ 、 $\cos \theta_{1,3}$ 、 $\cos \theta_{2,3}$ 。即  $\mathbf{r}_1$  和  $\mathbf{r}_2$  的相对夹角余弦值为  $\cos \theta_{1,2}$ ， $\mathbf{r}_1$  和  $\mathbf{r}_3$  的相对夹角余弦值为  $\cos \theta_{1,3}$ ， $\mathbf{r}_2$  和  $\mathbf{r}_3$  的相对夹角余弦值为  $\cos \theta_{2,3}$ 。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  $\mathbf{e}_1$ 。

再次强调一下， $\cos\theta_{1,2}$ 、 $\cos\theta_{1,3}$ 、 $\cos\theta_{2,3}$  确定的角度是向量之间的“相对夹角”。而  $[r_1, r_2, r_3]$  两两列向量确定的角度则是参考标准正交基的“绝对夹角”，这是因为  $r_1 = e_1$ 。

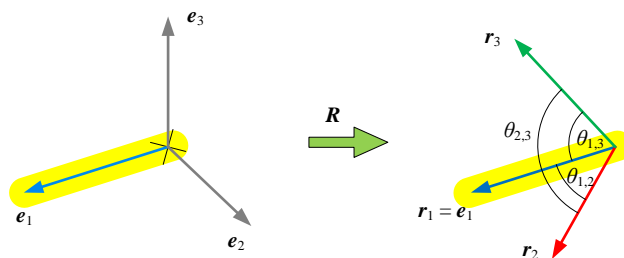


图 8. 三维系转化成满足指定两两夹角的坐标系

## 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  $P$  计算得到满足夹角条件的三个单位向量  $r_1$ 、 $r_2$ 、 $r_3$ 。

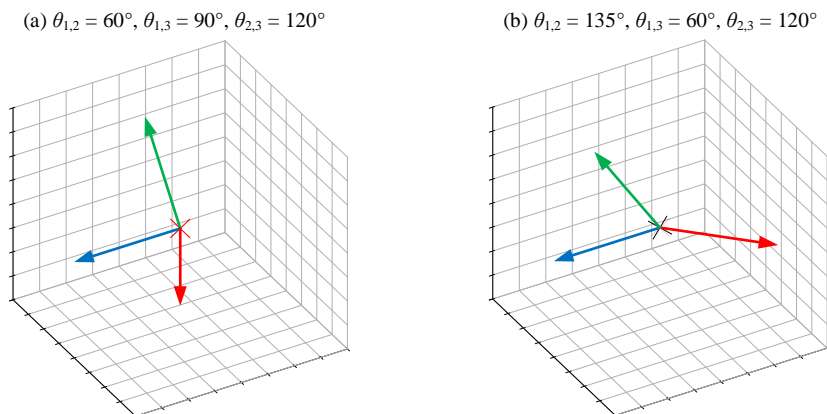


图 9. 给定三个夹角，确定向量三维空间位置

## 前提条件

在图 8 中，任意两个夹角之和必须大于等于第三个夹角，且任意角度不能为  $0^\circ$ ，也就是必须满足如下三个不等式：

$$\begin{aligned}\theta_{1,2} + \theta_{1,3} &\geq \theta_{2,3} > 0^\circ \\ \theta_{1,2} + \theta_{2,3} &\geq \theta_{1,3} > 0^\circ \\ \theta_{1,3} + \theta_{2,3} &\geq \theta_{1,2} > 0^\circ\end{aligned}\tag{38}$$

另外，三个角度夹角必须小于等于  $360^\circ$ ：

$$\theta_{1,2} + \theta_{1,3} + \theta_{2,3} \leq 360^\circ \quad (39)$$

试想一个有趣的现象，在图 8 中，如果  $\theta_{1,2} = \theta_{1,3} + \theta_{2,3}$ ，这意味着  $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$  三个向量在一个平面上， $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$  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，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  $\theta_{1,2} + \theta_{1,3} + \theta_{2,3} = 360^\circ$ ， $\mathbf{r}_1$ 、 $\mathbf{r}_2$ 、 $\mathbf{r}_3$  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 (30) 中  $\mathbf{P}$  可以 Cholesky 分解，即正定，需要满足以下条件：

$$\begin{aligned} \theta_{1,2} > 0^\circ, \quad \theta_{1,3} > 0^\circ, \quad \theta_{2,3} > 0^\circ \\ \theta_{1,2} + \theta_{1,3} > \theta_{2,3}, \quad \theta_{1,2} + \theta_{2,3} > \theta_{1,3}, \quad \theta_{1,3} + \theta_{2,3} > \theta_{1,2} \\ \theta_{1,2} + \theta_{1,3} + \theta_{2,3} < 360^\circ \end{aligned} \quad (40)$$

### 夹角相同

再看一组特殊情况，(30) 中  $\mathbf{P}$  两两夹角相同，即，

$$\theta_{1,2} = \theta_{1,3} = \theta_{2,3} = \theta \quad (41)$$

此时， $\mathbf{P}$  可以写成：

$$\mathbf{P} = \begin{bmatrix} 1 & \cos \theta & \cos \theta \\ \cos \theta & 1 & \cos \theta \\ \cos \theta & \cos \theta & 1 \end{bmatrix} \quad (42)$$

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。三个向量必然线性相关。

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  $120^\circ$ ，三个向量在一个平面上，也线性相关。

有了这两个极限情况，我们知道向量之间夹角  $\theta$  取值范围为  $[0^\circ, 120^\circ]$ ，而  $\cos \theta$  的取值范围为  $[-0.5, 1]$  ( $\cos(120^\circ) = -0.5$ ,  $\cos(0^\circ) = 1$ )。这也就是说，这种情况下， $\mathbf{P}$  的两个极端取值为：

$$\mathbf{P} = \begin{bmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (43)$$

上式中两个  $\mathbf{P}$  都不能进行 Cholesky 分解，因为  $\mathbf{P}$  都不满秩。

图 10 给出四个不同开合角度。图 10 (d) 对应的 (43) 第一个矩阵  $\mathbf{P}$ ,  $\theta_{1,2}$ 、 $\theta_{1,3}$ 、 $\theta_{2,3}$  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。

从统计角度来看,  $\mathbf{P}$  代表相关性系数矩阵。如果其中任意两个随机变量的相关性系数相等, 满足 (42) 相关性系数的取值范围为  $[-0.5, 1]$ 。

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。

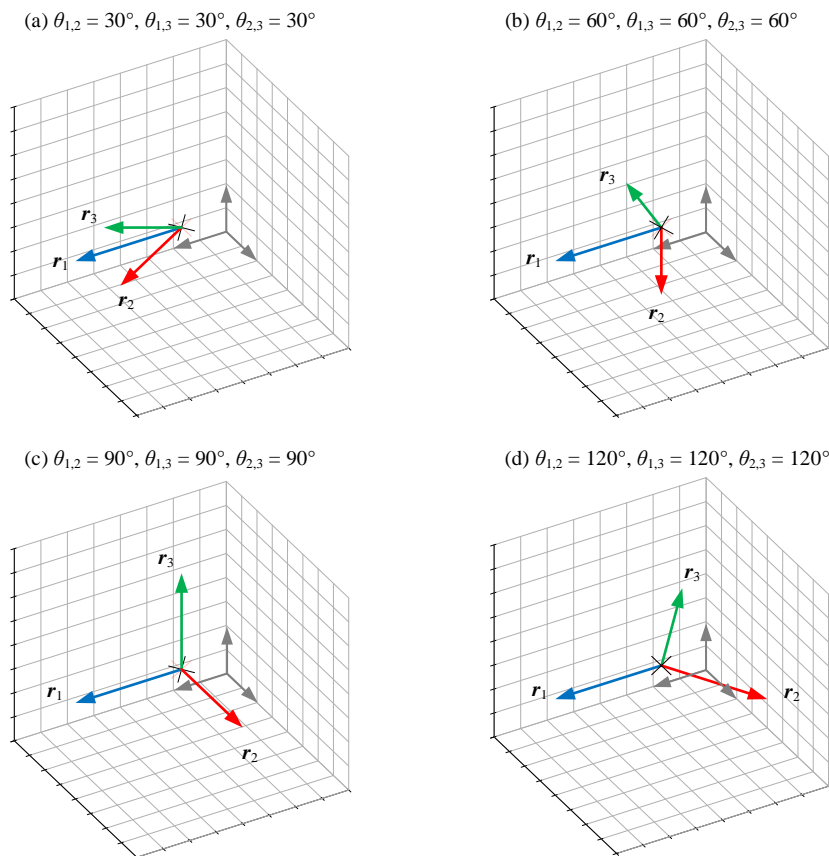
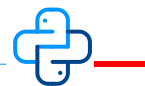


图 10. 相等角度条件下, 确定向量三维空间位置



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。



## 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。

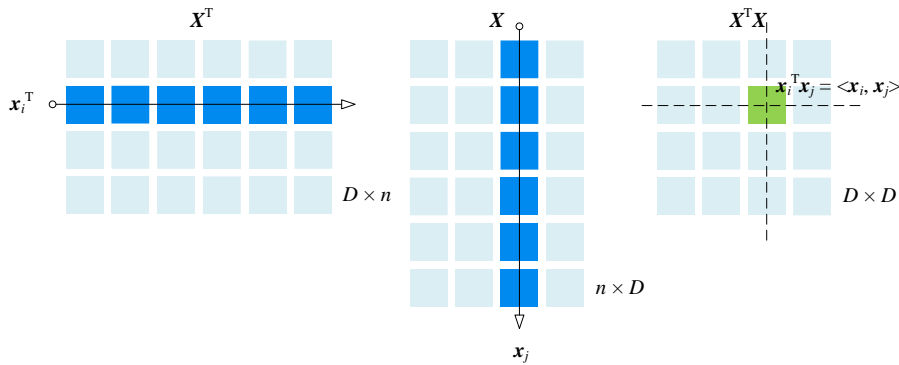


图 11. 格拉姆矩阵

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：

$$G = \begin{bmatrix} \mathbf{x}_1 \cdot \mathbf{x}_1 & \mathbf{x}_1 \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_1 \cdot \mathbf{x}_D \\ \mathbf{x}_2 \cdot \mathbf{x}_1 & \mathbf{x}_2 \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_2 \cdot \mathbf{x}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_D \cdot \mathbf{x}_1 & \mathbf{x}_D \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_D \cdot \mathbf{x}_D \end{bmatrix} = \begin{bmatrix} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle & \langle \mathbf{x}_1, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_1, \mathbf{x}_D \rangle \\ \langle \mathbf{x}_2, \mathbf{x}_1 \rangle & \langle \mathbf{x}_2, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_2, \mathbf{x}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{x}_D, \mathbf{x}_1 \rangle & \langle \mathbf{x}_D, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_D, \mathbf{x}_D \rangle \end{bmatrix} \quad (44)$$

### 确定列向量坐标

对  $G$  进行 Cholesky 分解得到：

$$G = R_G^T R_G \quad (45)$$

将  $R_G$  写成一排列向量：

$$R_G = [\mathbf{r}_{G,1} \quad \mathbf{r}_{G,2} \quad \cdots \quad \mathbf{r}_{G,D}] \quad (46)$$

将 (46) 代入 (45) 得到：

$$G = \begin{bmatrix} \mathbf{r}_{G,1}^T \\ \mathbf{r}_{G,2}^T \\ \vdots \\ \mathbf{r}_{G,D}^T \end{bmatrix} \begin{bmatrix} \mathbf{r}_{G,1} & \mathbf{r}_{G,2} & \cdots & \mathbf{r}_{G,D} \end{bmatrix} = \begin{bmatrix} \langle \mathbf{r}_{G,1}, \mathbf{r}_{G,1} \rangle & \langle \mathbf{r}_{G,1}, \mathbf{r}_{G,2} \rangle & \cdots & \langle \mathbf{r}_{G,1}, \mathbf{r}_{G,D} \rangle \\ \langle \mathbf{r}_{G,2}, \mathbf{r}_{G,1} \rangle & \langle \mathbf{r}_{G,2}, \mathbf{r}_{G,2} \rangle & \cdots & \langle \mathbf{r}_{G,2}, \mathbf{r}_{G,D} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{r}_{G,D}, \mathbf{r}_{G,1} \rangle & \langle \mathbf{r}_{G,D}, \mathbf{r}_{G,2} \rangle & \cdots & \langle \mathbf{r}_{G,D}, \mathbf{r}_{G,D} \rangle \end{bmatrix} \quad (47)$$

(44) 等价于 (47)，向量模和向量夹角之间完全等价。这“相当于”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  $X = [x_1, x_2, x_3, x_4]$ 。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余弦值。

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。

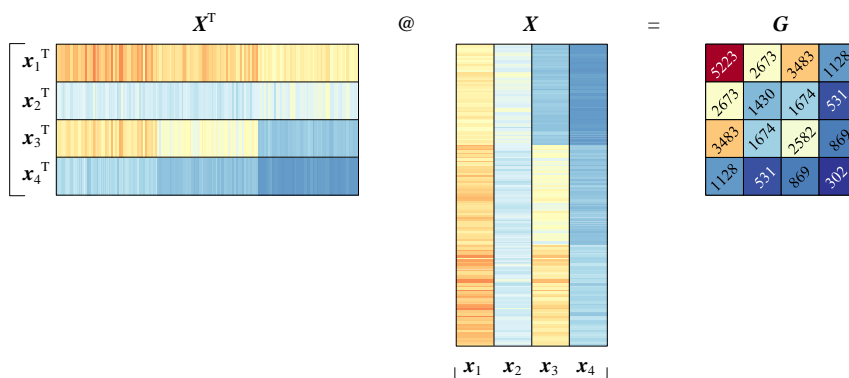


图 12. 鸢尾花数据矩阵  $X$  格拉姆矩阵，图片来自本书第 10 章

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”  $[x_1, x_2, x_3, x_4]$ 。

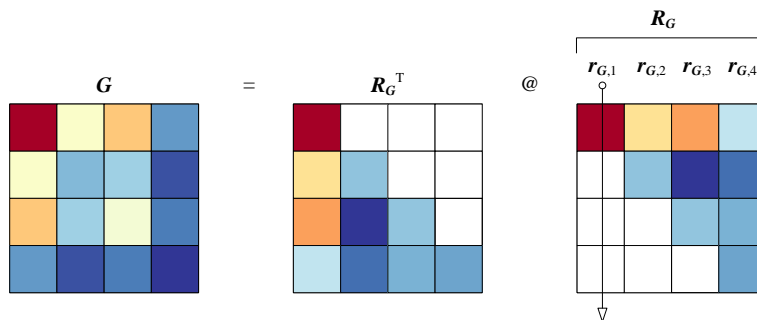


图 13. 对格拉姆矩阵  $G$  进行 Cholesky 分解

## 向量夹角

以向量夹角余弦形式展开  $G$  中向量积：

$$\mathbf{G} = \begin{bmatrix} \|\mathbf{x}_1\| \|\mathbf{x}_1\| \cos \theta_{1,1} & \|\mathbf{x}_1\| \|\mathbf{x}_2\| \cos \theta_{2,1} & \cdots & \|\mathbf{x}_1\| \|\mathbf{x}_D\| \cos \theta_{1,D} \\ \|\mathbf{x}_2\| \|\mathbf{x}_1\| \cos \theta_{1,2} & \|\mathbf{x}_2\| \|\mathbf{x}_2\| \cos \theta_{2,2} & \cdots & \|\mathbf{x}_2\| \|\mathbf{x}_D\| \cos \theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|\mathbf{x}_D\| \|\mathbf{x}_1\| \cos \theta_{1,D} & \|\mathbf{x}_D\| \|\mathbf{x}_2\| \cos \theta_{2,D} & \cdots & \|\mathbf{x}_D\| \|\mathbf{x}_D\| \cos \theta_{D,D} \end{bmatrix} \quad (48)$$

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

定义缩放矩阵  $\mathbf{S}$ ，具体形式如下：

$$\mathbf{S} = \begin{bmatrix} \|\mathbf{x}_1\| & & & \\ & \|\mathbf{x}_2\| & & \\ & & \ddots & \\ & & & \|\mathbf{x}_D\| \end{bmatrix} \quad (49)$$

对  $\mathbf{G}$  左右分别乘上  $\mathbf{S}$  的逆，得到  $\mathbf{C}$ ：

$$\mathbf{C} = \mathbf{S}^{-1} \mathbf{G} \mathbf{S}^{-1} = \begin{bmatrix} \frac{\mathbf{x}_1 \cdot \mathbf{x}_1}{\|\mathbf{x}_1\| \|\mathbf{x}_1\|} & \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{\|\mathbf{x}_1\| \|\mathbf{x}_2\|} & \cdots & \frac{\mathbf{x}_1 \cdot \mathbf{x}_D}{\|\mathbf{x}_1\| \|\mathbf{x}_D\|} \\ \frac{\mathbf{x}_2 \cdot \mathbf{x}_1}{\|\mathbf{x}_2\| \|\mathbf{x}_1\|} & \frac{\mathbf{x}_2 \cdot \mathbf{x}_2}{\|\mathbf{x}_2\| \|\mathbf{x}_2\|} & \cdots & \frac{\mathbf{x}_2 \cdot \mathbf{x}_D}{\|\mathbf{x}_2\| \|\mathbf{x}_D\|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbf{x}_D \cdot \mathbf{x}_1}{\|\mathbf{x}_D\| \|\mathbf{x}_1\|} & \frac{\mathbf{x}_D \cdot \mathbf{x}_2}{\|\mathbf{x}_D\| \|\mathbf{x}_2\|} & \cdots & \frac{\mathbf{x}_D \cdot \mathbf{x}_D}{\|\mathbf{x}_D\| \|\mathbf{x}_D\|} \end{bmatrix} \quad (50)$$

矩阵  $\mathbf{C}$  中元素就是向量两两夹角余弦值。

### 余弦相似度矩阵

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

$$\mathbf{C} = \begin{bmatrix} 1 & \cos \theta_{2,1} & \cdots & \cos \theta_{1,D} \\ \cos \theta_{1,2} & 1 & \cdots & \cos \theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \cos \theta_{1,D} & \cos \theta_{2,D} & \cdots & 1 \end{bmatrix} \quad (51)$$

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

对  $\mathbf{C}$  进行 Cholesky 分解得到：

$$\mathbf{C} = \mathbf{L} \mathbf{L}^T = \mathbf{R}^T \mathbf{R} \quad (52)$$

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

$$C = R^T R = \begin{bmatrix} r_1^T \\ r_2^T \\ \vdots \\ r_D^T \end{bmatrix} \begin{bmatrix} r_1 & r_2 & \cdots & r_D \end{bmatrix} = \begin{bmatrix} r_1^T r_1 & r_1^T r_2 & \cdots & r_1^T r_D \\ r_2^T r_1 & r_2^T r_2 & \cdots & r_2^T r_D \\ \vdots & \vdots & \ddots & \vdots \\ r_D^T r_1 & r_D^T r_2 & \cdots & r_D^T r_D \end{bmatrix} = \begin{bmatrix} 1 & \cos \theta_{1,2} & \cdots & \cos \theta_{1,D} \\ \cos \theta_{2,1} & 1 & \cdots & \cos \theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \cos \theta_{D,1} & \cos \theta_{D,2} & \cdots & 1 \end{bmatrix} \quad (53)$$

根据本章前文分析，我们知道  $r_1, r_2, \dots, r_D$  都是单位向量。

图 14 所示为鸢尾花数据矩阵的格拉姆矩阵  $G$ ，先转化成相似度矩阵  $C$ ，再转化成角度矩阵。

当然，我们也可以对鸢尾花数据先中心化，得到矩阵  $X_c$ 。再  $X_c$  计算的格拉姆矩阵，然后再计算其相似度矩阵，最后计算角度矩阵。请大家自行完成上述运算，并和图 14 结果比较。

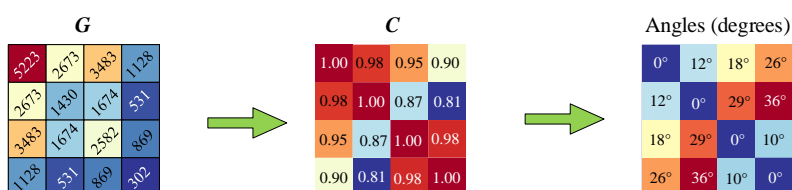


图 14. 格拉姆矩阵  $G$  转化成相似度矩阵  $C$ ，再转化成角度

本节介绍的内容在蒙特卡洛模拟 (Monte Carlo simulation) 中有重要应用。如图 15 所示，本章介绍的 Cholesky 分解结果可以用来产生满足指定相关性系数的随机数。

➡ 本系列丛书《概率统计》和《数据科学》两本会从理论、应用两个角度讲解蒙特卡洛模拟。

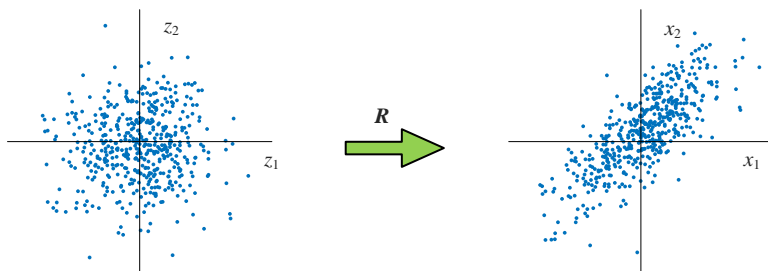


图 15. 产生满足指定相关性矩阵要求的随机数

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 分解可以帮助我们确定原数据矩阵的列向量空间等价坐标。此外，我们将在本系列丛书《概率统计》中有关协方差矩阵和蒙特卡罗模拟中再聊到 Cholesky 分解。

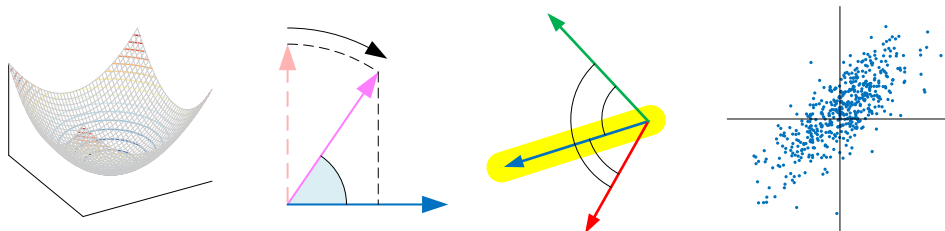


图 16. 总结本章重要内容的四幅图