

# 理解和分析协方差矩阵

Shuyong Chen

2022 年 10 月 29 日

## 1 前言

分析协方差矩阵的文章已经很多, 本文主要从几何的角度出发, 直观地显示协方差矩阵的特点, 为今后应用和分析协方差矩阵建立直觉。

## 2 定义

在概率论和统计学中, **协方差矩阵 (covariance matrix)**, 也称为**自动协方差矩阵 (auto-covariance matrix)**、**色散矩阵 (dispersion matrix)**、**方差矩阵 (variance matrix)** 或**方差-协方差矩阵 (variance-covariance matrix)**, 是一个平方矩阵, 给出给定随机向量的每对元素之间的协方差。任意协方差矩阵都是对称的和半正定的矩阵, 且其主对角线包含方差(即每个元素自身的协方差)。

直观地说, 协方差矩阵将方差的概念推广到多维。例如, 二维空间中随机点集合的变化不能完全用一个数字表示, 在  $x$  和  $y$  方向上的变化也不能包含所有必要的信息; 需要一个  $2 \times 2$  矩阵来充分描述二维变化。

随机向量  $\mathbf{X}$  的协方差矩阵通常由  $\mathbf{K}_{\mathbf{XX}}$  或  $\Sigma$  表示。

在本文中, 粗体无下标的  $\mathbf{X}$  和  $\mathbf{Y}$  用于引用随机向量, 非粗体有下标  $X_i$  和  $Y_i$  用于引用标量随机变量。

如果列向量的条目

$$\mathbf{X} = (X_1, X_2, \dots, X_n)^T$$

是随机向量, 每个变量都有有限的方差和期望值, 那么协方差矩阵  $\mathbf{K}_{\mathbf{XX}}$  是其  $(i, j)$  项为协方差的矩阵

$$K_{X_i X_j} = \text{cov}[X_i, X_j] = E[(X_i - E[X_i])(X_j - E[X_j])]$$

其中算子  $E$  表示其参数的期望值 (均值)。

换句话说,

$$\mathbf{K}_{\mathbf{XX}} = \begin{bmatrix} E[(X_1 - E[X_1])(X_1 - E[X_1])] & E[(X_1 - E[X_1])(X_2 - E[X_2])] & \cdots & E[(X_1 - E[X_1])(X_n - E[X_n])] \\ E[(X_2 - E[X_2])(X_1 - E[X_1])] & E[(X_2 - E[X_2])(X_2 - E[X_2])] & \cdots & E[(X_2 - E[X_2])(X_n - E[X_n])] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_n - E[X_n])(X_1 - E[X_1])] & E[(X_n - E[X_n])(X_2 - E[X_2])] & \cdots & E[(X_n - E[X_n])(X_n - E[X_n])] \end{bmatrix}$$

上述定义等同于矩阵等式

$$\boxed{K_{XX} = \text{cov}[\mathbf{X}, \mathbf{X}] = E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^T] = E[\mathbf{X}\mathbf{X}^T] - \boldsymbol{\mu}_X\boldsymbol{\mu}_X^T} \quad (1)$$

其中  $\boldsymbol{\mu}_X = E[\mathbf{X}]$ 。

### 3 性质

在线性代数教材中，有相当多的章节分析了实对称矩阵的特性，因此有不少特性可以直接拿来描述协方差矩阵。再加上半正定的特性，协方差矩阵的特性有：

1. 对称性
2. 半正定性 (可为零)
3. 实特征值和正特征值
4. 矩阵迹 (trace) 为正 (矩阵迹为特征值之和)
5. 行列式是正的 (行列式是特征值的乘积)
6. 对角线条目都是正数
7. 正交特征向量
8. 可对角化为  $Q\Lambda Q^T$
9. 可以得到 Cholesky 分解。
10.  $A^T A$  的秩与  $A$  的秩相同。
11.  $\ker(A^T A) = \ker(A)$

从几何的角度看，最重要的特性就是协方差矩阵是可正交对角化的

$$K_{XX} = R\Sigma^2 R^{-1}$$

其中， $R$  是以特征向量为列向量构成的正交矩阵，并且有特性  $R^{-1} = R^T$ ,  $RR^T = I$ ,  $\det(R) = 1$ ，因此  $R$  实际上是  $SO(n)$  的旋转矩阵； $\Sigma^2$  是由特征值，在这里是方差  $\sigma_1^2, \dots, \sigma_n^2$  构成的对角矩阵。对上式稍做变形

$$\begin{aligned} K_{XX} &= R\Sigma\Sigma R^T \\ &= (\Sigma R^T)^T (\Sigma R^T) \\ &= A^T A \end{aligned}$$

其中， $A = \Sigma R^T$ 。矩阵  $A^T A$  又被称为 Gramian 矩阵。矩阵  $A^T A$  (Gramian 矩阵) 具有以下性质：

- $\mathbf{A}^T \mathbf{A}$  是一个关键的矩阵结构，因为它在正交投影中起着重要的作用。协方差矩阵只是特例。
- $\mathbf{A}^T \mathbf{A}$  是协方差矩阵—你可以定义多元正态分布，其中  $\mathbf{A}^T \mathbf{A}$  是协方差矩阵，参见这里。
- 这相当于讨论对称半正定矩阵 (symmetric positive semidefinite matrices, s.p.s.d.) —对于某些矩阵  $\mathbf{A}$ ，每个对称半正定矩阵都可以写成  $\mathbf{A}^T \mathbf{A}$ 。

## 4 协方差矩阵几何解释

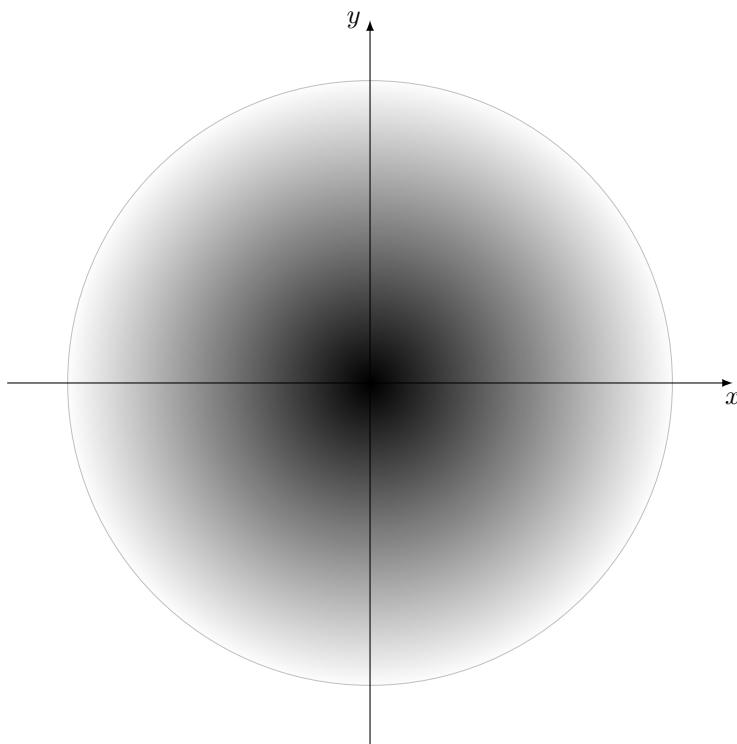
我们回到协方差矩阵是正交对角化形式

$$\mathbf{K}_{\mathbf{X}\mathbf{X}} = \mathbf{R}\mathbf{\Sigma}^2\mathbf{R}^T$$

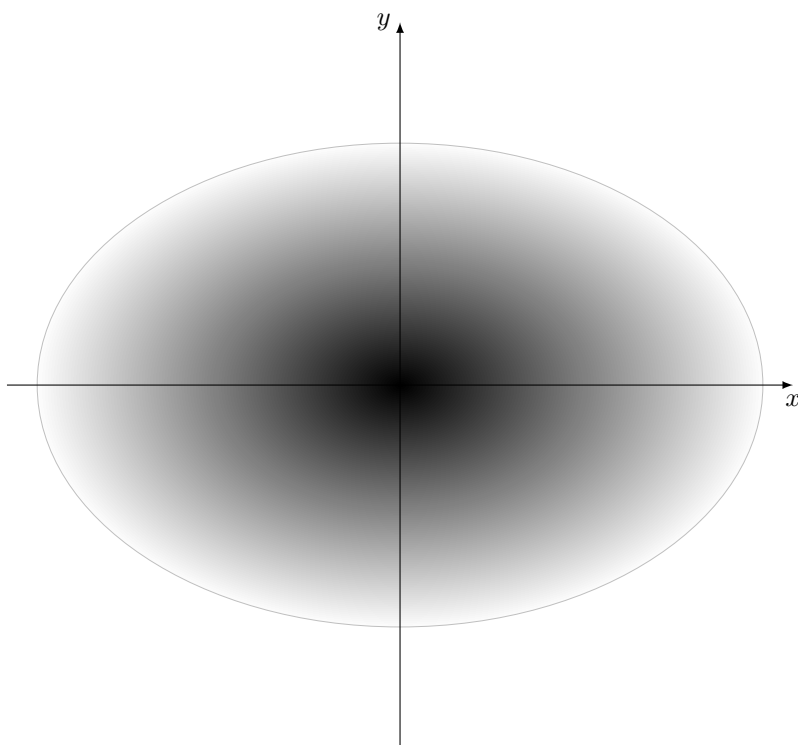
当协方差矩阵只有变量自身的方差而无变量之间的协方差时，矩阵可表示为

$$\mathbf{K}_{\mathbf{X}\mathbf{X}} = \mathbf{I}\mathbf{\Sigma}^2\mathbf{I}^T$$

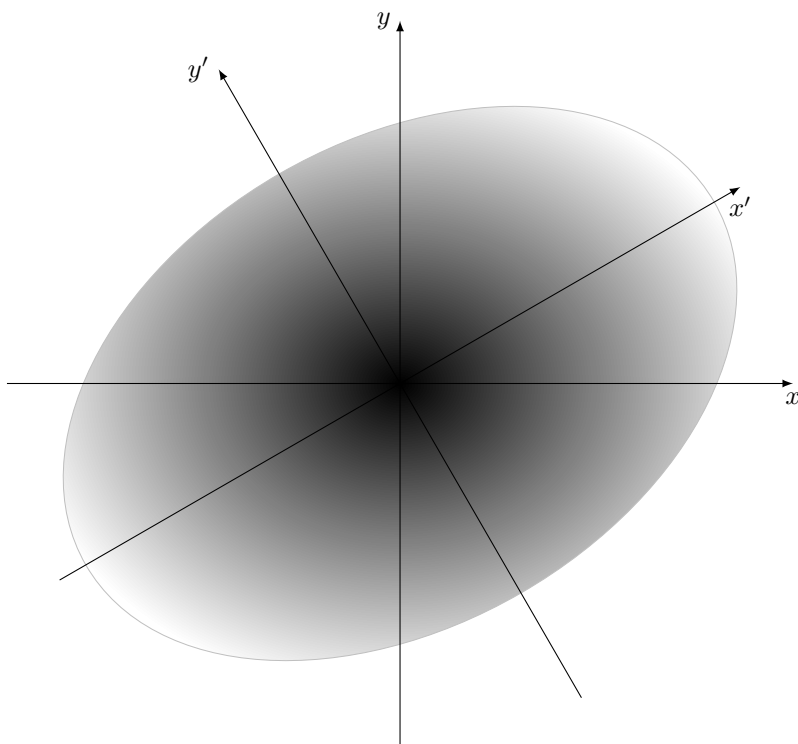
如果我们有高斯白色数据集合，表示为一个圆或一个球，



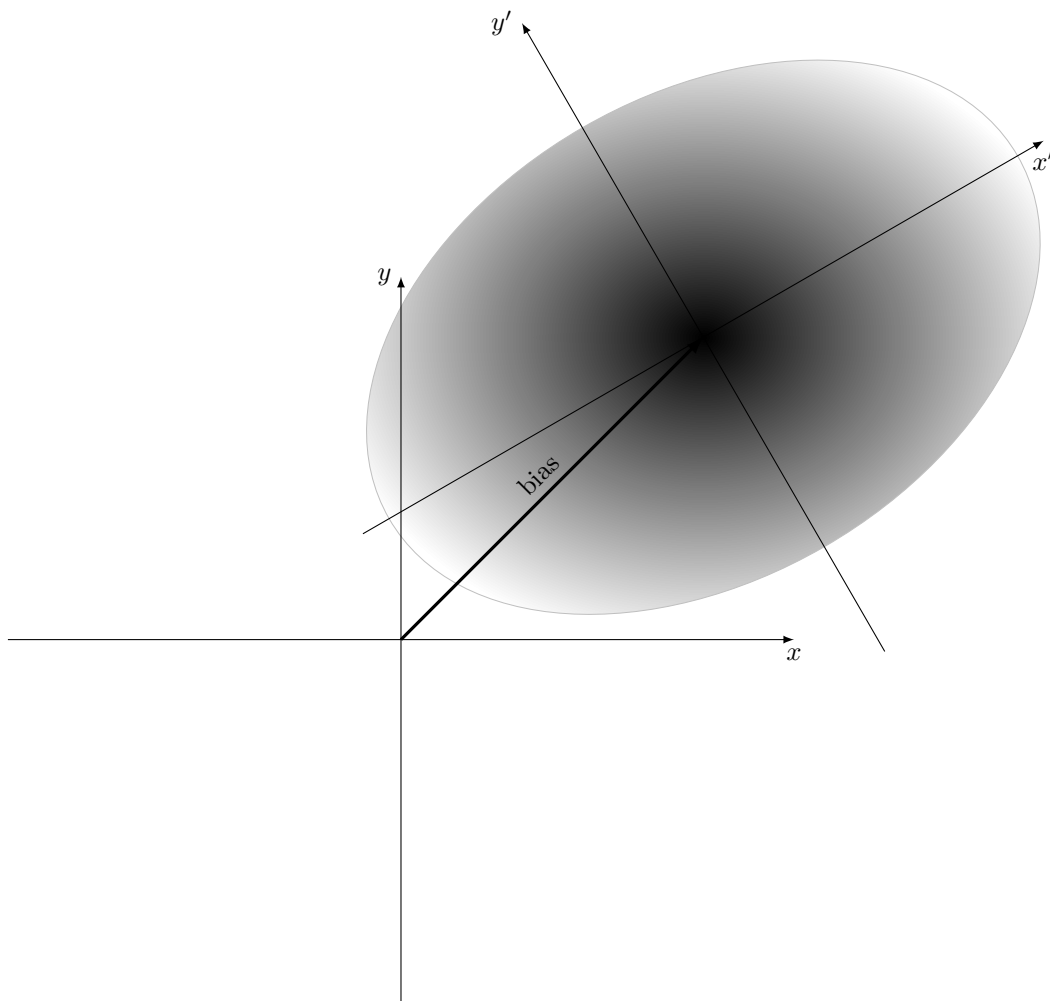
经过上式的线性变换，圆或球被拉伸为椭圆或椭球，数据在各轴的缩放程度由各轴的方差  $\sigma_i^2$  决定。



如果协方差矩阵变量之间的协方差不为零，则椭圆或椭球发生了旋转，因为  $\det(\mathbf{R}) = 1$ ，所以各个数据点到椭圆或椭球的中心的距离不会发生改变，也就是椭圆或椭球的形状不会发生改变。



**小结：** 对于高斯白色数据，协方差矩阵中的变量方差将圆拉伸为椭圆，变量协方差将椭圆旋转到新的方向， $\mathbf{R}$  中的列向量为新坐标系的正交基，但是新/旧坐标系的原点重合。最终的椭圆的形状，由  $\mathbf{R}$ ，也就是特征向量决定方向，由  $\sigma_i^2$ ，也就是特征值决定各个方向的大小。是数据中的非零均值，或称偏差 (bias)，将新坐标的原点移动到新的位置。



我们经常在三维空间中估计姿态，姿态的不确定性，也就是其  $3 \times 3$  协方差矩阵，方向为该矩阵的 3 个特征向量，一个圆球由 3 个特征值拉伸成的椭球，椭球的中心就是概率中心，就是姿态真实值最有可能存在的地方。在每一个穿过概率中心的直线上取概率密度，得到的就是真实值在各个方向上的高斯分布。

其实， $\mathbf{K}_{\mathbf{xx}} = \mathbf{R}\mathbf{\Sigma}^2\mathbf{R}^T$  是椭圆方程的二次型矩阵形式

$$\left(\frac{x_1}{\sigma_{x_1}}\right)^2 + \left(\frac{x_2}{\sigma_{x_2}}\right)^2 + \cdots + \left(\frac{x_n}{\sigma_{x_n}}\right)^2 = s \quad (2)$$

其中  $s$  定义椭圆的比例，表示所选的置信水平。等式 (2) 的左侧实际上表示独立正态分布数据样本的平方和。已知高斯数据点的平方和根据所谓的卡方分布进行分布。卡方分布定义为“自由度”，表

示未知量的数量。因此，通过计算卡方似然，我们可以很容易地获得上述总和，因此  $s$  等于特定值的概率。事实上，由于我们对置信区间感兴趣，我们正在寻找  $s$  小于或等于特定值的概率，该值可以使用累积卡方分布轻松获得。如果想偷懒，我们通常不需要计算这种概率，而只用在概率表中查找：  
<https://people.richland.edu/james/lecture/m170/tbl-chi.html>。

## 5 协方差矩阵最小化

协方差矩阵代表数据的不确定性，也就是数据在各维的扩散程度。在进行最优估计时，我们希望估计出来的数据最精确，首先需要对经过协方差矩阵变换后数据的总体扩散程度做一个度量，方法是计算矩阵的范数。我们知道，度量向量  $\mathbf{x}$  的大小用的是向量的欧几里德范数：

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$$

如果用类似的思路将矩阵看成是一个可以展开的折叠向量，则矩阵  $A$  的范数可以表示为

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$$

这种范数称为 Frobenius 范数 (Frobenius norm)，也称为 Hilbert-Schmidt 范数，后者在 (可能无限维的) Hilbert 空间上的算子上下文中使用得更频繁。经过一系列的变换和证明，Frobenius 范数可表示为

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\text{trace}(A^* A)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(A)}$$

其中， $\sigma_i(A)$  是矩阵  $A$  的奇异值。

对于  $n \times n$  的矩阵  $A$ ，它的特征值和矩阵的行列式和矩阵的迹有下面的关系：

$$\prod_{i=1}^n \lambda_i = \det(A)$$

$$\sum_{i=1}^n \lambda_i = \text{trace}(A)$$

由上可知，矩阵的迹表示的是该矩阵所有特征值之和。因为特征值的几何意义是在特征向量方向上的缩放的程度，所以矩阵的迹 (特征值之和) 的几何意义就是所有线性无关的特征向量的缩放程度之和。因为协方差矩阵的特点， $K_{\mathbf{XX}} = \mathbf{R}\Sigma^2\mathbf{R}^T$ ，并且  $\det(\mathbf{R}) = 1$ ，根据上一节的几何解释，是方差  $\sigma_i^2$  决定了椭圆的拉伸程度，也就是数据的扩散程度，所以，如果我们要度量协方差矩阵的总体扩散程度，只需要简单地计算  $\text{trace}(\Sigma^T \Sigma)$ 。

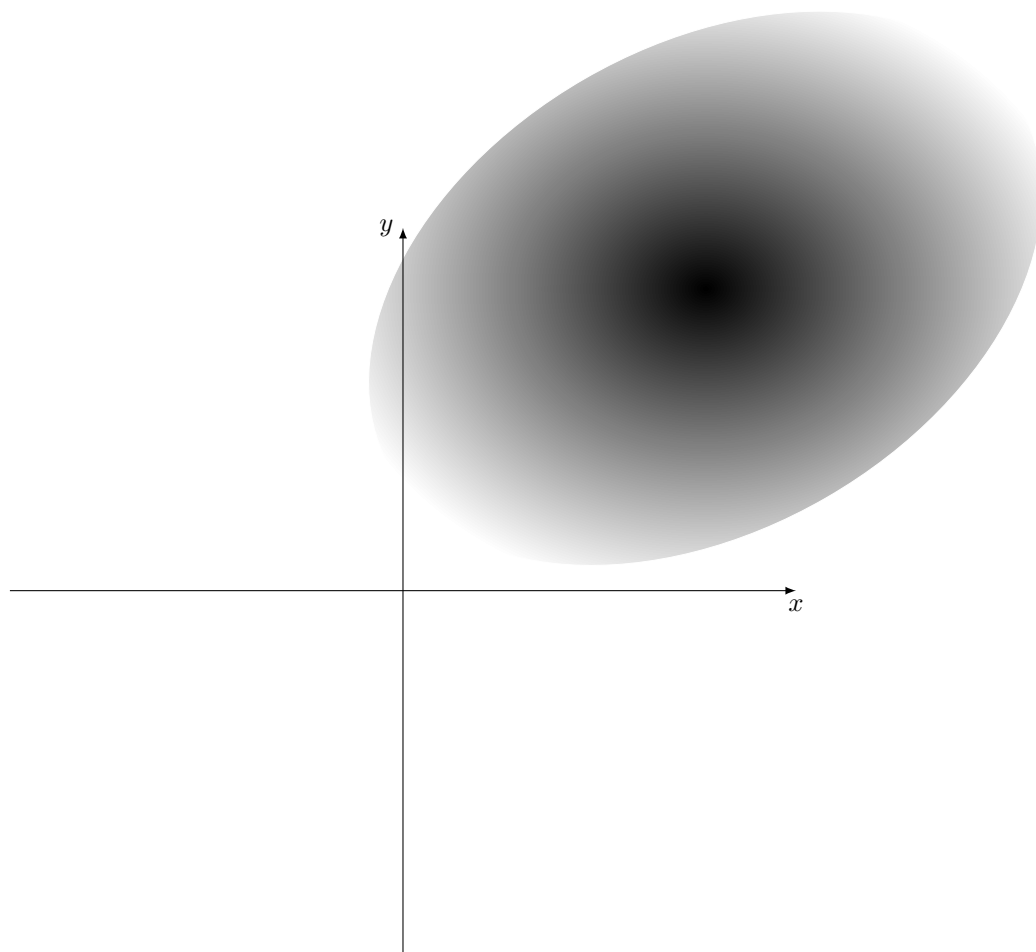
从另外一个方面理解，把  $\Sigma$  的对角线元素看成是标准差向量，则  $\text{trace}(\Sigma^T \Sigma)$  计算的是方差向量的范数。最小化矩阵的迹，就是从总体上最小化方差向量的范数。

在证明卡尔曼滤波器过程中要使得协方差矩阵的总体扩散程度最小，就要最小化协方差矩阵，使用的算法是最小化该矩阵的迹，所以，卡尔曼滤波器也被称为计算状态向量在 Hilbert 空间上的正交投影。

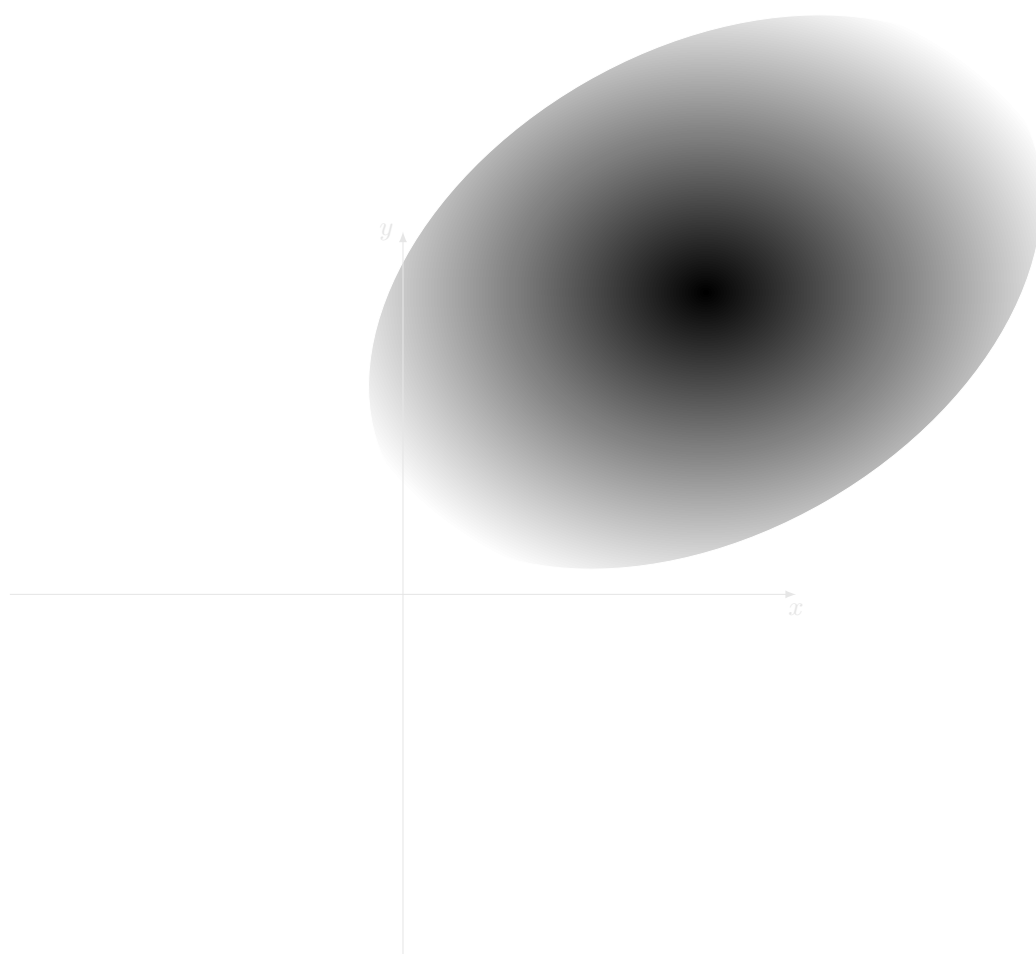
## 6 数据分析的直觉

如果我们相信多元数据的分布大多数是高斯分布，则协方差矩阵代表其不确定性，其形状为椭圆或椭球。由此可以更好地理解数据本身的意义。

假如我们测量得到这样的数据集，

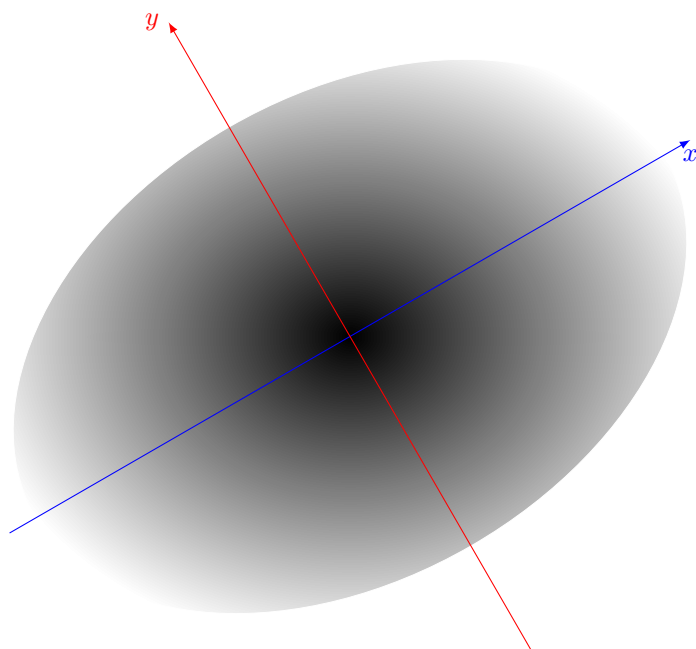


当坐标轴被忽略时，我们能做什么？

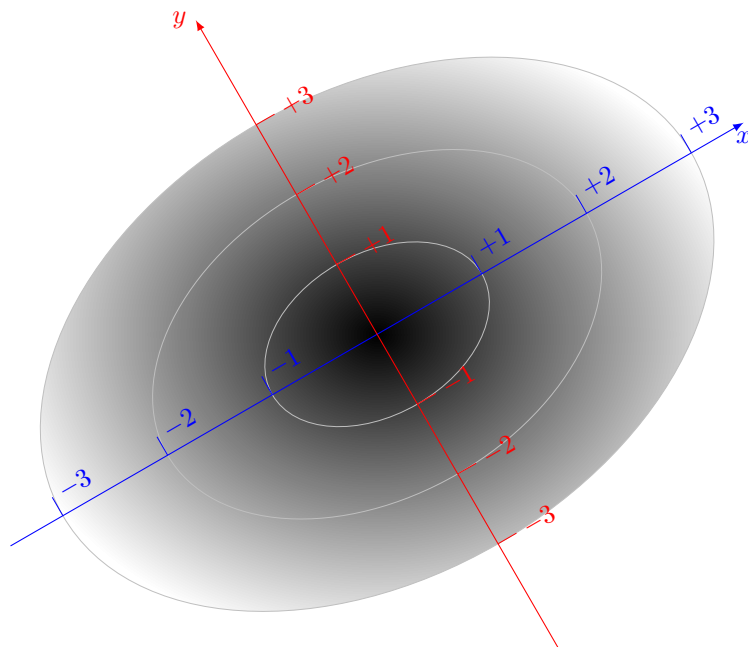


引用数据本身建议的坐标。**原点**将位于各点的中心点 (各点均值的点)。**第一个坐标轴** (下图中的蓝色) 将沿着点的“尖峰”延伸，这 (根据定义) 是方差最大的任意方向。**第二个坐标轴** (图中的红色) 将垂直于第一个坐标轴延伸。(在两个以上的维度中，将选择方差尽可能大的垂直方向，以此类推。)

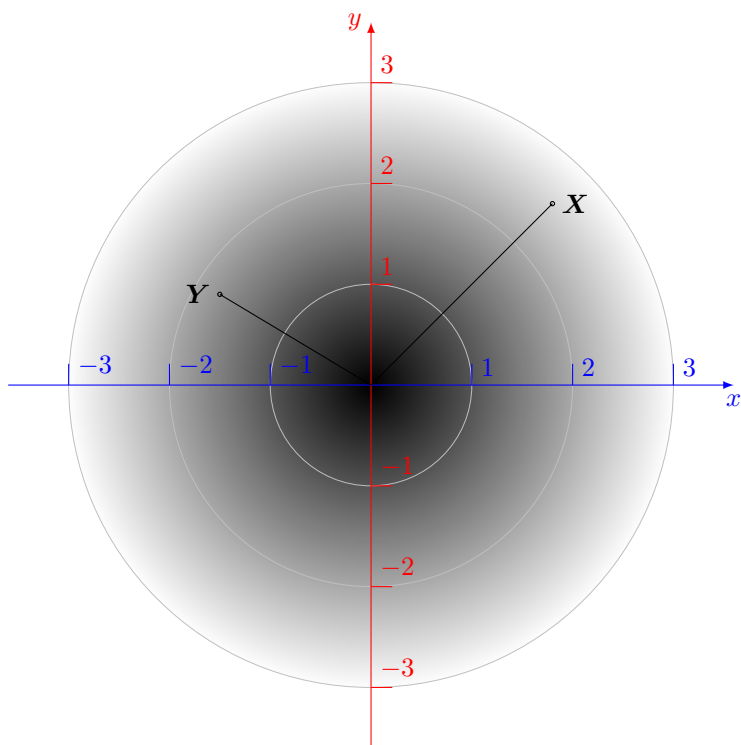




我们需要一个**刻度**。沿每个轴的标准偏差将很好地确定沿轴的单位。记住 68 – 95 – 99.7 规则：大约三分之二 (68%) 的点应该在原点的一个单位内 (沿轴)；大约 95% 应该在两个单位内。这样就很容易找到正确的单位。为便于参考，下图包括以下单位中的单位圆：



那看起来不像一个圆，是吗？这是因为这张图像被扭曲了 (两个轴上的数字之间的间距不同就证明了这一点)。让我们用轴的正确方向 (从左到右，从下到上) 和单位纵横比重新绘制它，这样水平方向上的一个单位实际上等于垂直方向上的一个单位：



我们是在这张图像中测量  $x$  和  $y$  的 Mahalanobis 距离并对其进行比较，而不是在原始图像中。

这里发生了什么？我们让数据告诉我们如何构建一个坐标系，以便在散点图中进行测量。仅此而已。尽管我们在这一过程中有一些选择（我们总是可以反转其中一个或两个轴；在极少数情况下，沿着“尖峰”的方向 — 主方向 — 不是唯一的），但它们不会改变最终绘图中的距离。

### 小结：

- 沿着新的坐标轴的单位向量是**特征向量**（协方差矩阵或其逆矩阵）。
- 我们注意到，将沿着每个特征向量的距离除以标准偏差（协方差的平方根），将不失真地把椭圆变成一个圆。让  $S$  代表协方差函数，两点  $x$  和  $y$  之间的新的距离（Mahalanobis 距离）是  $x$  到  $y$  的距离除以  $S(x - y, x - y)$  的平方根。相应的代数运算，现在认为  $S$  表示为矩阵， $x$  和  $y$  表示为向量，写为  $\sqrt{(x - y)^T S^{-1} (x - y)}$ 。无论用什么基来表示向量和矩阵，这都是有效的。特别是，这是在原始坐标中 Mahalanobis 距离的正确公式。
- 在最后一步中，轴被扩展的量是逆协方差矩阵的（平方根）**特征值**。等价地，轴被协方差矩阵的（根）特征值**收缩**。因此，散射越多，将椭圆转换为圆所需的收缩就越多。
- 尽管此过程始终适用于任意数据集，但对于近似多元正态分布的数据，它看起来非常漂亮（经典足球形状的云）。在其他情况下，平均点可能无法很好地表示数据中心，或者无法使用方差作为传播的度量以准确识别“尖峰”（数据的一般趋势）。

- 坐标原点的移动、轴的旋转和扩展共同形成**仿射变换**。除了初始偏移外，这是一个基的改变，从原来的基 (使用指向正坐标方向的单位向量) 到新的基 (使用单位特征向量的选择)。注意这里所说的仿射变换，指的是：先是平移，然后是基的改变。因为我们必须对“基的改变”做一些自由的理解，以包括不可逆转的线性变换：这是一个对 PCA 很重要的问题，它有效地放弃了一些基础元素。
- 这与主成分分析 (PCA) 密切相关。由此可见，让数据决定你用来描述它们和测量它们的差异的坐标，是及其优雅和实用性的。
- 对于多元正态分布 (我们可以使用概率密度的特性而不是点云的类似特性进行相同的构造)，Mahalanobis 距离 (到新原点) 出现在描述标准正态分布的概率密度的表达式  $\exp(-\frac{1}{2}x^2)$  中，以取代“ $x$ ”。因此，在新坐标系中，当投影到通过原点的任意直线上时，多元正态分布看起来像是**标准正态分布**。特别是，它在每个新坐标中都是标准正态分布。从这个角度来看，多元正态分布之间唯一本质上的区别在于它们使用了多少维度。(注意，该维度数量可能是，而且有时是，小于维度尺寸数量。)

## 7 Mahalanobis 距离

**Mahalanobis 距离 (Mahalanobis distance)** 是的一个衡量点  $P$  和概率分布  $D$  之间距离的方法，由 P. C. Mahalanobis 在 1936 年提出。它是对衡量  $P$  距离  $D$  的均值有多少标准差的概念的多维概括。 $P$  在  $D$  的均值处的距离为零，并随着  $P$  沿每个主分量轴远离均值而增长。如果这些轴中的每一个都被重新标定为具有单位方差，那么 Mahalanobis 距离就对应于变换空间中的标准欧几里德距离。因此，Mahalanobis 距离是无单位的、尺度不变的，并考虑到了数据集的相关性。

### 7.1 定义和性质

一组观测值为  $\mathbf{x} = (x_1, x_2, x_3, \dots, x_N)^\top$  与一组具有均值  $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \dots, \mu_N)^\top$  及其协方差矩阵  $\mathbf{S}$  的观测值的 Mahalanobis 距离定义为

$$D_M(\mathbf{x}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{S}^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

Mahalanobis 距离 (或其平方值即为“广义的点间距离的平方”) 也可以定义为两个具有相同的概率分布的随机向量  $\mathbf{x}$  和  $\mathbf{y}$  之间的不相似度测量，它们的概率分布用协方差矩阵  $\mathbf{S}$  表示：

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^\top \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})}$$

由于  $\mathbf{S}$  是协方差矩阵，它是半正定的，并且半正定矩阵的逆也是半正定的，我们得到的  $\mathbf{S}^{-1}$  也是半正定的。这就解释了为什么可以取平方根，因为所有的值都是正数。

如果协方差矩阵是单位矩阵，则 Mahalanobis 距离就会退化为欧几里德距离。如果协方差矩阵是对角的，则产生的距离度量称为**标准化欧几里德距离 (standardized Euclidean distance)**：

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^N \frac{(x_i - y_i)^2}{\sigma_i^2}}$$

其中  $\sigma_i$  是  $x_i$  和  $y_i$  在样本集合上的标准偏差。

Mahalanobis 距离在数据所跨越的空间的满秩线性变换下保持不变。这意味着，如果数据具有非平凡的零空间，则可以在将数据 (非退化地) 投影到数据的适当维度的任意空间后计算 Mahalanobis 距离。

我们可以找到有用的 Mahalanobis 距离的平方分解，这有助于解释多变量观测值离群的一些原因，也为识别离群值提供了一个图形工具。

## 7.2 直观的解释

考虑估计  $N$  维欧几里德空间中的一个测试点属于集合的概率的问题，其中我们得到的样本点肯定属于该集合。我们的第一步是找到样本点的中心点或质心。直观地说，测试点离这个质心越近，它就越有可能属于这个集合。

但是，我们还需要知道集合是分布在大范围内还是小范围内，这样我们就可以确定距中心的给定距离是否值得注意。最简单的方法是估计样本点到质心距离的标准偏差。如果测试点与质心之间的距离小于一个标准偏差，则我们可以得出结论，测试点极有可能属于集合。距离越远，测试点越有可能不属于集合。

通过将测试点和集合之间的标准化距离定义为  $\frac{\|\mathbf{x} - \boldsymbol{\mu}\|_2}{\sigma}$ ，可以量化这种直观的方法，其可解释为： $\frac{\text{testpoint} - \text{sample mean}}{\text{standard deviation}}$ 。通过将其插入正态分布，我们可以得出测试点属于集合的概率。

上述方法的缺点是，我们假设样本点以球形方式围绕质心分布。如果分布是明显的非球形的，例如椭球形的，那么我们期望测试点属于集合的概率不仅取决于与质心的距离，还取决于方向。在椭球体具有短轴的方向上，测试点必须更近，而在那些长轴的方向上，测试点可以离中心更远。

将其置于数学基础上，可以通过构建样本的协方差矩阵来估计最能代表集合概率分布的椭球体。Mahalanobis 距离是测试点到质心的距离除以测试点方向上椭球的宽度。

## 7.3 小结

回到上一章“分析数据直觉”的解释，Mahalanobis 距离定义为  $d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^\top \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})}$ ，其中  $\mathbf{S}$  是某些数据的协方差矩阵的估计；这意味着它是对称的。如果用于估计  $\mathbf{S}$  的列不是线性相关的，则  $\mathbf{S}$  是正定的。对称矩阵是可对角化的，其特征值和特征向量是实数。协方差矩阵的特征值均为正。特征向量可以选择为具有单位长度，并且是正交的，因此我们可以写为  $\mathbf{S} = \mathbf{R} \boldsymbol{\Sigma}^2 \mathbf{R}^\top$  和  $\mathbf{S}^{-1} = \mathbf{R}^\top \boldsymbol{\Sigma} \mathbf{R}$ 。将其插入到距离定义中，

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{[(\mathbf{x} - \mathbf{y})^\top \mathbf{R}^\top] \boldsymbol{\Sigma} \boldsymbol{\Sigma} [\mathbf{R}(\mathbf{x} - \mathbf{y})]} = \sqrt{\mathbf{z}^\top \mathbf{z}}$$

显然，方括号中的乘积是转置，乘以  $\mathbf{R}$  的效果是将向量  $(\mathbf{x} - \mathbf{y})$  旋转到一个正交基。最后， $\boldsymbol{\Sigma}$  是对角矩阵，由对角线上的每个元素倒置，然后取平方根来生成，是对每个向量的每个元素进行重新缩放。事实上， $\boldsymbol{\Sigma}$  正是正交空间中每个特征的逆标准偏差 (即  $(\boldsymbol{\Sigma}^2)^{-1}$  是一个精度矩阵，由于数据是在正交基

上, 所以矩阵是对角线的)。其效果是通过“扁平化”其轴线, 将旋转椭圆的物体变换成一个圆形。很明显,  $z^T z$  是以平方单位测量的, 所以取平方根会将距离返回到原始单位。

也就是, 对于经过协方差矩阵变换的数据, 由数据的各维方差将圆形白化数据拉伸成了椭圆, 沿各正交轴上的数据刻度不再相同, 因此在此空间中的数据不好相互比较。这时需要逆向操作, 将椭圆变换回圆形, 最后数据是在欧几里德空间中比较各自到概率中心的距离。

## 8 应用案例

卡尔曼滤波属于动态递归的最优化算法, 协方差矩阵在算法里面起着核心作用。所以对于卡尔曼滤波器当前的质量状态, 可以通过对当前的协方差矩阵状态进行判断。具体就是使用残差向量和残差协方差计算 Mahalanobis 距离进行判断。我们可以设置一个阈值, 比如 5, 代表着  $\pm 5$  倍的标准差, 根据常识,  $\pm 3$  倍的标准差已经包含 99.7% 的数据, 因此如果残差的 Mahalanobis 距离超过阈值则说明该估计值已经不可信。

因为在卡尔曼滤波器中, 协方差矩阵是由残差驱动而动态变化的。因为统计学上杠杆效应问题, 距离概率中心越远的离群值, 对协方差矩阵的影响越大。如果频繁发生 Mahalanobis 距离超过阈值的事件, 说明协方差矩阵质量出现了问题。

在卡尔曼滤波器的具体应用中, 即位姿估计, 也有一个有意思的比较。用卡尔曼滤波器进行位姿估计, 有两大类算法, 一种是直接法, 即对位姿直接用卡尔曼滤波器进行估计, 另一种是间接法, 即对位姿的误差用卡尔曼滤波器进行估计, 再将其作用到标称位姿, 则标称位姿就是最优估计。

抛开数学上的争议 (即位姿是群而不是向量, 因此不能使用减法, 而协方差的核心算子就是减法), 我们从数据变化的可能性考虑, 还是间接法比较好一些。

因为直接法是直接对位姿进行最优估计, 而位姿是有可能剧烈变化的。突变的位姿残差, 就会在循环迭代中使得残差的 Mahalanobis 距离增大, 影响协方差矩阵质量。这时候我们怎么判断这是外部的位姿剧烈变化还是我们的卡尔曼滤波器出现了问题?

但是对于间接法, 误差依赖于传感器的特性, 即测量偏差和随机游走噪声。我们可以对系统内部的传感器进行充分测试, 它的特性会比较稳定, 变化缓慢, 很少有突变, 因此协方差矩阵的方向、形状和偏差位置就相对稳定, 估计出来的误差也相对稳定, 因此理论上间接法会比较好。

## 9 总结

协方差矩阵是许多不同领域的有用工具。从中可以导出一个变换矩阵, 称为白化变换, 它允许从一个完全消除数据的相关性的视角, 找到以紧凑方式表示数据的最佳基 (有关协方差矩阵的形式证明和附加属性, 请参见瑞利商 (Rayleigh Quotient))。这被称为主成分分析 (principal component analysis, PCA) 和 Karhunen-Loève 变换 (Karhunen-Loève transform, KL-transform)。

协方差矩阵代表着多元数据的不确定性, 它各维数据自身的方差驱动白化数据由圆形变成椭圆, 数据之间的协方差驱动椭圆旋转。而为了度量数据点之间的差异, 又需要将椭圆变成圆形, 所用算法为 Mahalanobis 距离。所有一切的直觉, 就是让数据决定你用来描述它们和测量它们的差异的坐标。

## 10 参考资料

1. Covariance Matrix
2. A geometric interpretation of the covariance matrix
3. The matrix  $A^T A$  (Gramian matrix) has the following properties
4. Properties of the Covariance Matrix
5. Mahalanobis Distance
6. Bottom to top explanation of the Mahalanobis distance?