矩阵之美 —— 算法篇

2025 年 3 月 11 日

目　录

第 1 章 最小二乘法

最小二乘法（Least Squares, LS）自诞生以来，在诸多领域已经得到了广泛的应用. 本章将首先分别从行空间和列空间的角度给出线性方程组的两个图像，然后从代数、几何、概率等角度对最小二乘法给出全方位的诠释.

1.1 问题背景

1766 年，德国有一位名叫约翰 • 提丢斯（Johann Daniel Titius, 1729-1796）的大学教授，写了下面的数列

$$\frac{3\times 2^n+4}{10}, n=-\infty, 0, 1, 2, 3, \cdots .$$

令人惊奇的是，他发现这个数列的每一项与当时已知的六大行星（水星、金星、地球、火星、木星、土星）到太阳的距离比例（地球到太阳的距离定义为 1）有一定的联系. 提丢斯的朋友，天文学家波得（Johann Elert Bode, 1747-1826）深知这一发现的重要意义，就于 1772 年公布了提丢斯的这一发现. 这串数从此引起了众多科学家的极大重视，并被称为提丢斯 —— 波得定律（即太阳系行星与太阳平均距离的经验规则）. 当时，人们还没有发现天王星和海王星，认为土星就是距太阳最远的行星. 1781 年，英籍德国人赫歇尔（William Herschel, 1738-1822）在接近 19.6 的位置上（即数列中的第 8 项）发现了天王星，从此，人们就对这一定则深信不疑了. 根据这一定则，在数列的第 5 项即 2.8 的位置上也应该对应一颗行星，只是还没有被发现. 于是，许多天文学家和天文爱好者便以极大的热情，踏上了寻找这颗新行星的征程.

图 1.1　六大行星到太阳的距离（近似）比例，其中假定地球到太阳的距离为 1

1801 年，意大利天文学家皮亚齐（Giuseppe Piazzi, 1746-1826）终于在相应位置发现了一颗小行星（即谷神星）. 遗憾的是，经过 40 天的跟踪观测后，由于谷神星运行至太阳背后，皮亚齐失去了谷神星的位置. 随后全世界的科学家利用皮亚齐的观测数据开始寻找谷神星，但是根据大多数人计算的结果来寻找谷神星都没有收获. 时年 24 岁的高斯（Johann Carl Friedrich Gauss, 1777-1855）也采用了一种新方法（即为最小二乘法）计算了谷神星的轨道. 奥地利天文学家奥尔伯斯（Heinrich Wilhelm Matthias Olbers, 1758-1840）根据高斯计算出来的轨道重新发现了谷神星. 高斯使用的最小二乘

法于 1809 年发表在他的著作《天体运动论》中，而法国科学家勒让德（Adrien-Marie Legendre, 1752-1833）于 1806 年独立发现了最小二乘法，但因不为时人所知而默默无闻. 两人曾为谁最早创立最小二乘法原理发生争执. 最小二乘法自创立以来，在自然科学乃至社会科学的各个领域得到了广泛的应用.

由于最小二乘法大多用于线性方程组的求解，因此，接下来我们首先给出线性方程组的两个图像.

1.2 线性方程组的两个图像

对于下面包含 m 个方程、n 个未知量的线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}, \tag{1.1}$$

可以将其写成矩阵（向量）形式

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b},$$

其中

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

为线性方程组的系数矩阵，

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

分别为待求解的变量组成的向量和方程组右边的常数项组成的向量.

对于线性方程组 (1.1)，一般可以从两个角度来理解：行空间角度和列空间角度. 从行空间角度，该方程组可以认为是 n 维空间中 m 个超平面的交集. 从列空间角度，该方程组的常向量 $\boldsymbol{b}$ 可以认为是系数矩阵 $\mathbf{A}$ 的列向量的线性组合，其中 $\boldsymbol{x}$ 的分量为组合系数.

1.2.1 线性方程组的行空间图像

(1.1) 中的线性方程组由 m 个线性方程构成，其中每一个方程均代表 n 维空间中的一个超平面. 以其中的第 i 个方程为例，即

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i.$$

显然，该方程是 n 维空间中以系数矩阵 $\mathbf{A}$ 的第 i 个行向量 $\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}$ 为法方向的 $(n-1)$ 维超平面. 而线性方程组 (1.1) 的解则可以认为是这 m 个 $(n-1)$ 维超平面的交集，此即为线性方程组的行空间图像. 其中，由系数矩阵 $\mathbf{A}$ 的所有行向量张成的线性空间称为 $\mathbf{A}$ 的行空间.

下面以一个简单的线性方程组为例给出线性方程组行空间图像的直观理解，即

$$\begin{cases} 2x_1 + x_2 = 3 \\ x_1 + 2x_2 = 3 \end{cases}. \tag{1.2}$$

显然，该方程组由二维平面上的两条直线方程构成（如图 1.2 ），即直线方程 $2x_1+x_2 = 3$ 和直线方程 $x_1 + 2x_2 = 3$. 两条直线的法线分别为系数矩阵的两个行向量. 这两条直线的交点 (1,1) 即为该线性方程组的唯一解.

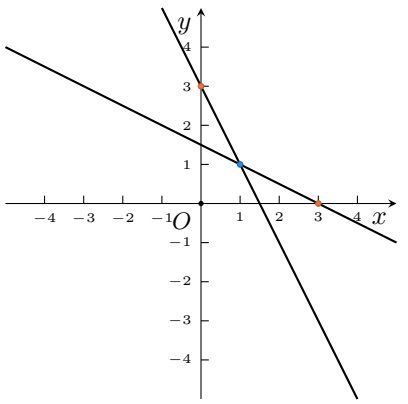


图 1.2　线性方程组的行空间图像

在方程组中的未知数和方程的个数都非常少的时候，利用方程组的行空间图像来理解线性方程组或许是一种不错的选择（如图 1.2 ）. 但当方程组的未知数和方程组的个数比较多的时候，行空间的方式将很难得到方程组的清晰图像. 比如当 $m = n = 3$ 时，线性方程组 (1.1) 的解对应于三维空间中三个二维平面的交集. 尽管我们仍然可以通过类似于图 1.2 的方式分别画出三个平面，然后再探讨它们的交集情况. 但是，与平面上两条直线的相交问题相比，三维空间中三个平面的相交情况无论从绘图，还是从交集结构角度，都明显要复杂一些. 进一步地，当方程组的变量的个数和方程的个数更多时，利用行空间图像的方式来理解线性方程组，将更为繁琐，甚至不可接受.

从行空间图像的角度来理解线性方程组，可以得到一些有悖于我们常识的有意思的结论. 比如，当 $m = n = 4$ 时，(1.1) 变为

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 = b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 = b_4 \end{cases}, \tag{1.3}$$

该方程组包含 4 个变量和 4 个方程. 其中每个方程对应着四维空间的一个三维超平面. 不妨假设该方程组的系数矩阵 $\mathbf{A}$ 非奇异. 从行空间图像角度，(1.3) 的解对应 4 个三维超平面的交集. 如果我们联立前两个方程，则可以得到四维空间中的一个二维平面 S_1. 相应地，后两个方程联立也同样可以得到四维空间中的另一个二维平面 S_2. 因此，从行空间图像角度，(1.3) 的解也可以认为对应着这两个二维平面的交集. 由系数矩阵的非奇异性，我们知道该方程组只有唯一解 $\boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{b}$. 这意味着，S_1 与 S_2 相交且只交于一点. 因此，在四维空间中，两个平面可以只相交于一个点.

从上面两个平面交于一个点的例子可以看出，在三维空间中不可能出现的事情，在四维空间中却真实地发生了. 因此，在大多数情况下，尽管行空间图像并不是一个处理线性方程组的好的方式，但它却有助于增进我们对高维世界的理解.

1.2.2 线性方程组的列空间图像

记 $\mathbf{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix}$，则线性方程组 (1.1) 可以表示为

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}, \tag{1.4}$$

即方程组 (1.1) 右边的常数向量 $\boldsymbol{b}$ 可以表示为系数矩阵 $\mathbf{A}$ 的列向量的线性组合，且组合系数为待求的未知变量. 此即为线性方程组的列空间图像，其中由系数矩阵 $\mathbf{A}$ 的所有列向量张成的线性空间称为 $\mathbf{A}$ 的列空间.

从列空间图像的角度，(1.2) 可以重新表示为

$$x_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}. \tag{1.5}$$

在 (1.5) 中，方程组的常数项向量表示为系数矩阵的两个列向量的线性组合，且组合系数分别为两个待求的变量 x_1, x_2. 显然 $x_1 = x_2 = 1$ 即为满足 (1.5) 的解. 从图 1.3 可以看出，我们也可以将 $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathrm{T}}$ 看作 $\boldsymbol{b}$ 在 $\eta = \{(2,1),(1,2)\}$ 坐标系下的坐标.

接下来，我们分别从行空间图像角度和列空间图像角度探讨一下线性方程组的解的存在性. 从行空间图像角度，当线性方程组的各个方程所对应的超平面的交集非空

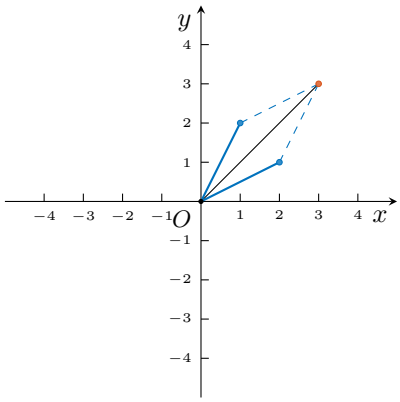


图 1.3　线性方程组的列空间图像

的时候，线性方程组有解；反之，则无解. 从列空间图像角度，当线性方程组右边的常数项向量位于系数矩阵的列空间时，方程有解；否则，则无解. 从中可以看出，相对于线性方程组的行空间图像，线性方程组的列空间图像更为简洁、清晰.

对于一个线性方程组，当常数项向量位于系数矩阵的列空间时，可以很容易得到方程组的精确解. 而当常数项向量不能由系数矩阵的列向量线性表出时，我们可以引入最小二乘法来得到方程组的近似解.

1.3 线性方程组的最小二乘解

由线性方程组的求解理论可知，在求解线性方程组时，当方程的个数多于未知数（变量）的个数时，方程往往无解，此类方程组称为矛盾方程组或超定方程组. 最小二乘法是求解矛盾方程组的经典方法. 下面分别从行空间图像角度和列空间角度给出最小二乘法的两种解法.

1.3.1 最小二乘法的行空间方法

不妨假设线性方程组 (1.1) 是矛盾方程组. 从行空间图像的角度，即不存在一组系数 $x_1, x_2, \cdots, x_n$，使得方程组 (1.1) 中的每一个方程都成立. 那么，退而求其次，我们希望存在一组系数 $x_1, x_2, \cdots, x_n$，使得方程组的每一个方程都尽量成立，即对于每一个方程，我们希望 $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i, i = 1, 2, \cdots, m$ 尽量成立. 因此，自然而然，我们可以用 $\left(\sum_{j=1}^{n} a_{ij}x_j - b_i\right)^2$ 来衡量第 i 个方程在这组系数下的误差. 此时，针对线性方程组，这组系数的求解问题就可以转化为如下优化模型

$$\min_{x_1,x_2,\cdots,x_n} f(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{m}\left(\sum_{j=1}^{n}a_{ij}x_j-b_i\right)^2. \tag{1.6}$$

为了得到上述优化模型的最优解, 我们可以首先求解目标函数的稳定点, 即满足目标函数相对于每个自变量的偏导数都为 0 的点. 由于

$$\frac{\partial f(x_1,x_2,\cdots,x_n)}{\partial x_k}=2\sum_{i=1}^{m}a_{ik}\left(\sum_{j=1}^{n}a_{ij}x_j-b_i\right)\overset{令}{=}0,k=1,2,\cdots,n,$$

因此，目标函数稳定点的求解可以转化为如下线性方程组的求解

$$\begin{cases}\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}a_{i1}a_{ij}x_j=\sum\limits_{i=1}^{m}a_{i1}b_i\\ \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}a_{i2}a_{ij}x_j=\sum\limits_{i=1}^{m}a_{i2}b_i\\ \qquad\vdots\\ \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}a_{in}a_{ij}x_j=\sum\limits_{i=1}^{m}a_{in}b_i\end{cases}. \tag{1.7}$$

显然 (1.7) 是一个有 n 个未知量 n 个方程的线性方程组，令

$$\tilde{\mathbf{A}}=\begin{bmatrix}\sum_{i=1}^{m}a_{i1}a_{i1} & \sum_{i=1}^{m}a_{i1}a_{i2} & \cdots & \sum_{i=1}^{m}a_{i1}a_{in}\\ \sum_{i=1}^{m}a_{i2}a_{i1} & \sum_{i=1}^{m}a_{i2}a_{i2} & \cdots & \sum_{i=1}^{m}a_{i2}a_{in}\\ \vdots & \vdots & \ddots & \vdots\\ \sum_{i=1}^{m}a_{in}a_{i1} & \sum_{i=1}^{m}a_{in}a_{i2} & \cdots & \sum_{i=1}^{m}a_{in}a_{in}\end{bmatrix},\quad \tilde{\boldsymbol{b}}=\begin{bmatrix}\sum_{i=1}^{m}a_{i1}b_i\\ \sum_{i=1}^{m}a_{i2}b_i\\ \vdots\\ \sum_{i=1}^{m}a_{in}b_i\end{bmatrix},$$

则 (1.7) 的解为

$$\boldsymbol{x}=\tilde{\mathbf{A}}^{-1}\tilde{\boldsymbol{b}}. \tag{1.8}$$

尽管 (1.8) 的表达式比较简洁，但由于 (1.7) 的系数矩阵 $\tilde{\mathbf{A}}$ 过于繁杂，因此实际应用中我们大多从线性方程组的列空间图像角度给出方程组的最小二乘解.

1.3.2 最小二乘法的列空间方法

当线性方程组 (1.1) 为矛盾方程组时，从列空间图像的角度，这意味着常数项向量 $\boldsymbol{b}$ 不在方程组 (1.1) 的系数矩阵 $\mathbf{A}$ 的列空间中，或者说 $\boldsymbol{b}$ 不能由 $\mathbf{A}$ 的列向量线性表出. 同样地，退而求其次，我们可以寻求找到一组系数 $x_1,x_2,\cdots,x_n$，使得 $x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n=\mathbf{A}\boldsymbol{x}$ 与向量 $\boldsymbol{b}$ 尽量接近. 不妨用这两项之差的 2-范数的平

方来衡量它们之间的误差，这样，这组系数的求解问题可以转化为如下优化模型

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) = \|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2. \tag{1.9}$$

为了得到 (1.9) 的最优解，我们同样需要先找到目标函数的稳定点. 首先，将目标函数展开可以得到

$$f(\boldsymbol{x}) = \|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = (\mathbf{A}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}(\mathbf{A}\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}.$$

利用矩阵微积分公式（相关知识见附录 B ）可以得到

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{x} - 2\mathbf{A}^{\mathrm{T}}\boldsymbol{b}.$$

令 $\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \mathbf{0}$, 有

$$\boldsymbol{x} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{b} = \mathbf{A}^{\dagger}\boldsymbol{b}, \tag{1.10}$$

此即为线性方程组的最小二乘解，其中 $\mathbf{A}^{\dagger} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}$ 称为矩阵 $\mathbf{A}$ 的广义逆.

事实上, 模型 (1.9) 与 (1.6) 是等价的, 所以 (1.9) 可以认为是 (1.6) 的矩阵（向量）表达. 注意到

$$\tilde{\mathbf{A}} = \mathbf{A}^{\mathrm{T}}\mathbf{A}, \quad \tilde{\boldsymbol{b}} = \mathbf{A}^{\mathrm{T}}\boldsymbol{b}.$$

因此，两种方式得到的线性方程组的最小二乘解 (1.10) 与 (1.8) 也是等价的. 但毫无疑问，与基于行空间图像的方法相比，基于列空间图像的最小二乘解更为简洁、直观、优美.

1.3.3 直线拟合

最小二乘法在诸多领域和方向得到了广泛的应用，而本节着重介绍最小二乘法在直线拟合中的应用. 已知一组观测点 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ 分布在直角坐标系中（如图 1.4 ），如何用一条直线拟合这些散点呢？

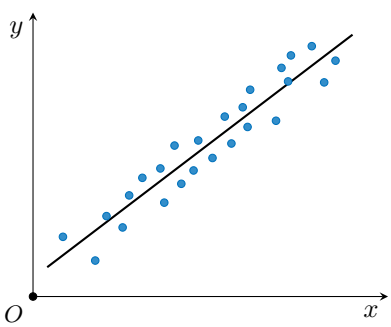


图 1.4　直线拟合示意图

不妨假设待求的直线方程为

$$y = ax + b, \tag{1.11}$$

其中 a 代表直线的斜率，b 为直线的截距，它们均为待求的未知变量. 显然，我们希望已知的 n 个观测都满足该直线方程，即

$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \\ \qquad\vdots \\ ax_n + b = y_n \end{cases}. \tag{1.12}$$

实际应用中，(1.12) 往往为矛盾方程，即不存在 a, b 使得上述方程组的各个方程都成立. 因而，退而求其次，我们接下来转而寻求 a, b，使得 (1.12) 尽量成立. 借鉴 (1.6)，我们可以建立直线拟合的优化模型为

$$\min_{a,b} f(a,b) = \sum_{i=1}^{n} (ax_i + b - y_i)^2. \tag{1.13}$$

方便起见，我们可以类比 (1.9) 将 (1.13) 中的目标函数重新表示为自变量为向量的情形，这样，(1.13) 可以转化为如下形式

$$\min_{\boldsymbol{c}} f(\boldsymbol{c}) = \|\mathbf{X}\boldsymbol{c} - \boldsymbol{y}\|^2, \tag{1.14}$$

其中,

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{x} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

根据 (1.10)，可得 (1.14) 的最小二乘解为

$$\boldsymbol{c} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}. \tag{1.15}$$

下面用一个非常简单的例子来直观展示最小二乘法的求解过程.

例 1.1 用直线 $y = ax + b$ 来拟合平面上的三个点：$(1,1), (2,1)$ 和 $(3,3)$.

解　根据给定的三个点的坐标，显然有 $x_1 = 1, x_2 = 2, x_3 = 3$；$y_1 = 1, y_2 = 1, y_3 = 3$. 记

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix},$$

根据公式 (1.15) 可得

$$\begin{aligned}\boldsymbol{c} &= \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y} \\ &= \left(\begin{bmatrix}1 & 2 & 3\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ 2 & 1\\ 3 & 1\end{bmatrix}\right)^{-1}\begin{bmatrix}1 & 2 & 3\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix}1\\ 1\\ 3\end{bmatrix} \\ &= \begin{bmatrix}1\\ -1/3\end{bmatrix},\end{aligned}$$

即，待求直线方程为 $y = x - \frac{1}{3}$.

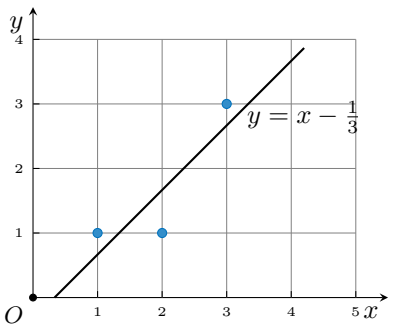


图 1.5　三个点及其对应的拟合直线

1.4 最小二乘法的几何解释

对于方程组 (1.1)，(1.4) 给出了其列空间表达，即 $x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}$. 当此方程组为矛盾方程组时，即当 $\boldsymbol{b}$ 不能由系数矩阵 $\mathbf{A}$ 的列向量线性表出时，(1.10) 给出的方程组的最小二乘解 $\boldsymbol{x} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{b}$ 即为我们想要的一组系数. 利用这组系数对 $\mathbf{A}$ 的各个列向量进行线性组合，将得到 $\mathbf{A}$ 的列空间的一个新元素 $\mathbf{A}\boldsymbol{x}$. 显然，这个新元素 $\mathbf{A}\boldsymbol{x}$ 在 $\mathbf{A}$ 的列空间，而向量 $\boldsymbol{b}$ 不在 $\mathbf{A}$ 的列空间，那么它们之间存在什么关系呢？

事实上，由于 $\mathbf{A}\boldsymbol{x} = \mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{b}$，记

$$\mathbf{P}_{\mathbf{A}} = \mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}, \tag{1.16}$$

显然 $\mathbf{A}\boldsymbol{x}$ 与 $\boldsymbol{b}$ 之间的关系为

$$\mathbf{A}\boldsymbol{x} = \mathbf{P}_{\mathbf{A}}\boldsymbol{b}. \tag{1.17}$$

即，将 $\mathbf{P}_{\mathbf{A}}$ 作用于 $\boldsymbol{b}$，就能得到 $\mathbf{A}\boldsymbol{x}$，那么矩阵 $\mathbf{P}_{\mathbf{A}}$ 又是一个什么矩阵呢？它具有什么性质呢？事实上，$\mathbf{P}_{\mathbf{A}}$ 就是由矩阵 $\mathbf{A}$ 构建的投影矩阵，将其作用于任何一个向

量，都会将该向量投影到矩阵 $\mathbf{A}$ 的列空间. 因此，当 $\boldsymbol{x}=\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{b}$ 时，$\mathbf{A}\boldsymbol{x}$ 就是向量 $\boldsymbol{b}$ 在 $\mathbf{A}$ 的列空间的投影，或者说，$\mathbf{A}\boldsymbol{x}$ 是 $\mathbf{A}$ 的列空间中距离向量 $\boldsymbol{b}$ 最近的向量. 也即，当 (1.1) 为矛盾方程组时，虽然不存在一组系数 $x_1,x_2,\cdots,x_n$, 使得 $x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n=\boldsymbol{b}$ 成立，但是方程组的最小二乘解 $\boldsymbol{x}=\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{b}$, 会使得 $x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n$（即 $\mathbf{A}\boldsymbol{x}$）最大程度靠近 $\boldsymbol{b}$.

总结而言，从几何角度，最小二乘法相当于将常数项向量 $\boldsymbol{b}$ 往系数矩阵 $\mathbf{A}$ 的列空间上进行正交投影.

接下来，我们给出投影矩阵的严格定义.

定义 1.1（**投影矩阵**） 具有对称性的幂等矩阵，即为投影矩阵.

从定义可以看出，当一个矩阵 $\mathbf{P}$ 为投影矩阵时，根据对称性，必然有 $\mathbf{P}^{\mathrm{T}}=\mathbf{P}$. 根据幂等性，必然有 $\mathbf{P}^2=\mathbf{P}$. 投影矩阵为幂等矩阵，意味着，在一个空间上投影一次和投影两次甚至多次，结果是一样的.

在实际应用中，我们还经常用到投影矩阵 $\mathbf{P}$ 的正交补投影矩阵 $\mathbf{P}^{\perp}$，其中

$$\mathbf{P}^{\perp}=\mathbf{I}-\mathbf{P}, \tag{1.18}$$

即正交补投影矩阵等于单位阵与投影矩阵之差. 对于 (1.16) 中的矩阵 $\mathbf{P}_{\mathbf{A}}$, 容易验证其满足投影矩阵的定义，即 $\mathbf{P}_{\mathbf{A}}^{\mathrm{T}}=\mathbf{P}_{\mathbf{A}}$，$\mathbf{P}_{\mathbf{A}}^2=\mathbf{P}_{\mathbf{A}}$. 相应地 $\mathbf{P}_{\mathbf{A}}$ 的正交补投影矩阵为 $\mathbf{P}_{\mathbf{A}}^{\perp}=\mathbf{I}-\mathbf{P}_{\mathbf{A}}$，将 $\mathbf{P}_{\mathbf{A}}^{\perp}$ 作用于任意列向量，则将该向量投影到矩阵 $\mathbf{A}$ 的列空间的正交补空间. 值得注意的是，上述投影矩阵 $\mathbf{P}_{\mathbf{A}}$ 的定义一般要求矩阵 $\mathbf{A}$ 为列满秩矩阵. 当 $\mathbf{A}$ 为行满秩矩阵的时候，可先将其转置为列满秩矩阵 $\mathbf{A}^{\mathrm{T}}$，相应的投影矩阵为 $\mathbf{P}_{\mathbf{A}^{\mathrm{T}}}=\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)^{-1}\mathbf{A}$，将 $\mathbf{P}_{\mathbf{A}^{\mathrm{T}}}$ 作用于任意向量，则可以将该向量投影到矩阵 $\mathbf{A}$ 的行空间.

仍以上述的直线拟合为例. 显然, 这 n 个观测对 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 可以认为是二维空间的 n 个散点. 同时, 也可以将这些观测看作 n 维样本空间的两个向量 $\boldsymbol{x}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}^{\mathrm{T}}$, $\boldsymbol{y}=\begin{bmatrix}y_1 & y_2 & \cdots & y_n\end{bmatrix}^{\mathrm{T}}$. 利用 $y=ax+b$ 对这组散点进行直线拟合, 相当于寻找参数 a,b 使得向量方程 $\boldsymbol{y}=a\boldsymbol{x}+b\mathbf{1}$ 成立或者尽量成立. 从样本空间来看, 其实就是找到向量 $\boldsymbol{x}$ 与 $\mathbf{1}$ 的最佳线性组合, 使得该组合与向量 $\boldsymbol{y}$ 尽量接近. 这相当于将向量 $\boldsymbol{y}$ 投影到向量 $\boldsymbol{x}$ 与 $\mathbf{1}$ 所张成的平面上，而投影点 $\boldsymbol{y}'$ 即为该平面上距离点 $\boldsymbol{y}$ 最近的点（图 1.6 ）.

1.5 最小二乘法的概率解释

本节将从概率角度，对最小二乘法给出进一步解释. 仍以平面上一组观测点 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 的直线拟合为例. 在模型 (1.14) 中，目标函数 $f(a,b)=\sum_{i=1}^{n}(ax_i+b-y_i)^2$ 代表因变量 y 的总体观测误差. 注意到，这个总体观测误差由 n

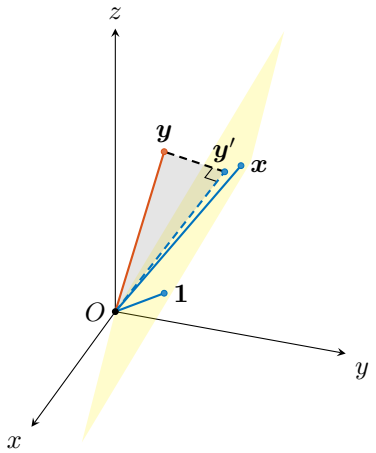


图 1.6　最小二乘法的几何解释，其中 $\boldsymbol{x}=\begin{bmatrix}1 & 2 & 3\end{bmatrix}^{\mathrm{T}}$，$\boldsymbol{y}=\begin{bmatrix}1 & 1 & 3\end{bmatrix}^{\mathrm{T}}$，$\mathbf{1}=\begin{bmatrix}1 & 1 & 1\end{bmatrix}^{\mathrm{T}}$, 最小二乘法对应着三个向量之间的投影关系. 当以 $y=ax+b$ 来拟合散点时，对应到样本空间，相当于寻求向量 $\boldsymbol{y}$ 到向量 $\boldsymbol{x}$ 和 $\mathbf{1}$ 所张成平面（$\operatorname{span}(\boldsymbol{x},\mathbf{1})$）上的投影 $\boldsymbol{y}'$

个部分构成. 其中的每一项 $(ax_i+b-y_i)^2$ 均代表相应观测点的因变量的观测误差. 即对于每个观测点的因变量的观测误差，都用模型解 (ax_i+b) 与观测值 y_i 的差的平方来衡量. 这正是相应的方法命名为最小二乘法而不是最小一乘或者最小三乘的原因所在. 接下来，从概率的角度对这一问题进行进一步阐述.

在直线拟合问题中，由于 (1.12) 一般为矛盾方程，因此我们可以把 (1.12) 写为如下的形式

$$\begin{cases} y_1 = ax_1 + b + \varepsilon_1 \\ y_2 = ax_2 + b + \varepsilon_2 \\ \quad\vdots \\ y_n = ax_n + b + \varepsilon_n \end{cases}, \tag{1.19}$$

其中，ε_i 为第 i 个因变量 y_i 的观测误差，它对应着不能被线性模型刻画的因素. 由于误差的不确定性，其中的每一个观测误差 ε_i 都可以看作一个随机变量. 假设 ε_i 服从均值为 0，标准差为 σ 的正态分布，即 $\varepsilon_i \sim N\left(0,\sigma^2\right)$，则其概率密度函数为[1]

$$f\left(\varepsilon_i\right)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right). \tag{1.20}$$

[1]在本书，模型误差函数和概率密度函数都用 $f(\cdot)$ 来表示，请读者注意区分.

这意味着，在给定 x_i 和参数 a,b 的情况下，因变量 y_i 也服从同样的正态分布，即

$$f\left(y_i \mid x_i; a, b\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(ax_i + b - y_i\right)^2}{2\sigma^2}\right). \tag{1.21}$$

我们可以定义所有观测数据关于参数 a,b 的似然（Likelihood）函数如下

$$L(a,b) = f\left(y_1, y_2, \cdots, y_n \mid x_1, x_2, \cdots, x_n; a, b\right). \tag{1.22}$$

显然，似然函数 $L(a,b)$ 为给定 $x_1, x_2, \cdots, x_n$ 和参数 a,b 情况下，$y_1, y_2, \cdots, y_n$ 的联合概率密度函数. 不妨假设所有的误差项 ε_i 独立同分布，那么联合概率密度函数等于所有因变量的概率密度函数的乘积，即

$$f\left(y_1, y_2, \cdots, y_n \mid x_1, x_2, \cdots, x_n; a, b\right) = \prod_{i=1}^{n} f\left(y_i \mid x_i; a, b\right), \tag{1.23}$$

因此，

$$L(a,b) = \prod_{i=1}^{n} f\left(y_i \mid x_i; a, b\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(ax_i + b - y_i\right)^2}{2\sigma^2}\right). \tag{1.24}$$

选择合适的参数 a,b，使得观测数据出现的可能性最大（即 $L(a,b)$ 最大），即为关于参数 a,b 的最大似然问题. 为了便于求解，定义 $l(a,b) = \ln(L(a,b))$，显然最大化 $L(a,b)$ 与最大化 $l(a,b)$ 是等价的. 由于

$$\begin{aligned} l(a,b) =& \ln\left(\prod_{i=1}^{n} f\left(y_i \mid x_i; a, b\right)\right) = \ln\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(ax_i + b - y_i\right)^2}{2\sigma^2}\right)\right) \\ =& \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(ax_i + b - y_i\right)^2}{2\sigma^2}\right)\right) \\ =& n \ln \frac{1}{\sqrt{2\pi}\sigma} - \sum_{i=1}^{n} \frac{\left(ax_i + b - y_i\right)^2}{2\sigma^2}, \end{aligned} \tag{1.25}$$

并且 $\frac{1}{2\sigma^2}$ 为常数. 因此，最大化 $l(a,b)$，就相当于最小化 $\sum_{i=1}^{n}(ax_i + b - y_i)^2$. 而 $\sum_{i=1}^{n}(ax_i + b - y_i)^2$ 正好为 (1.13) 的目标函数，即 $f(a,b) = \sum_{i=1}^{n}(ax_i + b - y_i)^2$. 因此，在观测误差满足独立同高斯分布的前提下，直线拟合问题的目标函数中的平方项是一个必然的结果.

从上面的推导可以看出，各个观测点的模型误差满足独立同分布的高斯分布是最小二乘法能够行之有效的前提. 如果此条件不能得到满足，用最小二乘法拟合数据很多情况下将不能得到合理的结果.

由于各个观测互不影响，因此不同观测点的观测误差之间的独立性一般都是成立的，那么接下来我们重点关注每一个观测点的模型误差是否都满足高斯分布. 事实上，概率论中的中心极限定理为这个前提条件的成立提供了一定的理论保证.

定理 1.1（中心极限定理） 假设随机变量 $\delta_1, \delta_2, \cdots, \delta_n$ 相互独立，在一定条件下[2]，它们的平均 $\frac{1}{n}\sum_{i=1}^{n}\delta_i$ 随着 n 的增大趋于高斯分布.

对于上述直线拟合问题中的观测数据的每一个误差项 ε_i，都可以认为它由多种不相干的因素构成，而每一个因素也可以认为是一个随机变量，不妨将其记为 $\varepsilon_{ij}, j = 1, 2, \cdots k$，其中 k 为随机因素的个数. 因此，ε_i 可以表示为多个随机变量的平均，即 $\varepsilon_i = \frac{1}{k}\sum_{j=1}^{k}\varepsilon_{ij}$. 而当这些随机因素的个数比较大时，根据中心极限定理，他们的平均 ε_i 必然趋于高斯分布. 因此，一般情况下，我们假设直线拟合中不能被模型刻画的观测误差满足高斯分布是合理的.

1.6 最小二乘法在应用中的问题

前面的几节给出最小二乘法的相关理论结果，但在实际应用中，针对不同的应用场景，最小二乘法还存在着各种问题. 接下来，我们将对其中的变量问题、约束问题、病态问题、异常问题、目标函数问题等几个常见的问题展开讨论.

1.6.1 变量问题

在本小节，我们仍以直线拟合为例，探讨最小二乘法中的变量问题. 在 1.3.3 节中，对于给定的一组散点 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，通过最小二乘法，我们给出了拟合这组散点的直线方程 $y = ax + b$. 这样做，其实隐含了一个前提条件：x 为自变量，y 为因变量. 那么我们不免要问，对于同样的这组散点，如果把 y 当做自变量，x 当做因变量，即用 $x = a'y + b'$ 来拟合这组散点是否可以得到同样的结果呢？或者说，用 $x = a'y + b'$ 来拟合这组散点，和用 $y = ax + b$ 来拟合这组散点是否会得到同一条直线呢？容易验证，如果它们为同一条直线，必有下列等式成立

$$\begin{cases} a' = \dfrac{1}{a} \\ b' = -\dfrac{b}{a} \end{cases}. \tag{1.26}$$

接下来我们把 y 当做自变量，x 当做因变量，用 $x = a'y + b'$ 来拟合这组散点. 同样记

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^{\mathrm{T}}, \quad \boldsymbol{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^{\mathrm{T}},$$

这组散点的直线拟合方程 $x = a'y + b'$ 的确定，相当于寻求系数 a', b'，使得下面的方

[2]在 Lindeberg-Levy 定理中这个条件为随机变量具有相同的分布，在 Lindeberg-Feller 定理中这个条件为 Lindeberg 条件，该条件不要求随机变量具有相同的分布.

程组尽量成立

$$\begin{cases} a'y_1 + b' = x_1 \\ a'y_2 + b' = x_2 \\ \qquad\quad \vdots \\ a'y_n + b' = x_n \end{cases}. \tag{1.27}$$

其对应的向量形式为

$$a'\boldsymbol{y} + b'\mathbf{1} = \boldsymbol{x},$$

令

$$\mathbf{Y} = \begin{bmatrix} \boldsymbol{y} & \mathbf{1} \end{bmatrix}, \quad \boldsymbol{c}' = \begin{bmatrix} a' & b' \end{bmatrix}^{\mathrm{T}},$$

则最佳系数 a', b' 的确定问题转化为如下优化模型

$$\min_{\boldsymbol{c}'} g\left(\boldsymbol{c}'\right) = \left\|\mathbf{Y}\boldsymbol{c}' - \boldsymbol{x}\right\|^2. \tag{1.28}$$

借鉴 (1.14) 可以得到 (1.28) 的最小二乘解为

$$\boldsymbol{c}' = \left(\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathrm{T}}\boldsymbol{x}. \tag{1.29}$$

至此，我们通过自变量和因变量位置的互换给出了两种直线拟合的方式. 那么，这两种直线拟合的方式是否等价呢？下面我们仍然用 1.3.3 节的例子1.1中的三个散点 $(1,1),(2,1),(3,3)$ 来对此问题进行验证. 记

$$\mathbf{Y} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},$$

根据公式 (1.29)，可以得到

$$\boldsymbol{c}' = \left(\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right)^{-1}\mathbf{Y}^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix} 0.75 \\ 0.75 \end{bmatrix},$$

对应的直线方程为（如图 1.7 ）

$$x = 0.75y + 0.75,$$

即 $a' = b' = 0.75$，显然 $a' \neq \frac{1}{a}, b' \neq -\frac{b}{a}$，即等式 (1.26) 不成立. 这说明，这两种直线拟合得到的结果是不等价的. 从图 1.7 也可以看出，利用 $x = a'y + b'$ 得到的拟合直线与利用 $y = ax + b$ 拟合的直线并不是同一条直线.

接下来，我们从模型的目标函数本身来对上述现象进行进一步解释. 当用 $y = ax + b$ 来拟合散点时，模型的目标函数（见 (1.13) 或者 (1.14)）反映的是各个散点在 Y 轴方向到拟合直线的距离的平方和（图 1.8a ）. 而当用 $x = a'y + b'$ 来拟合散点时，

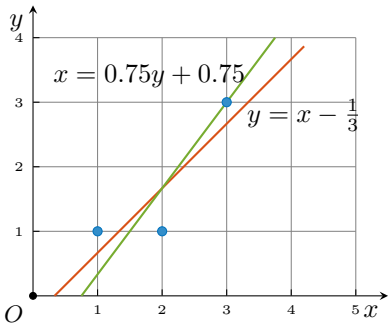
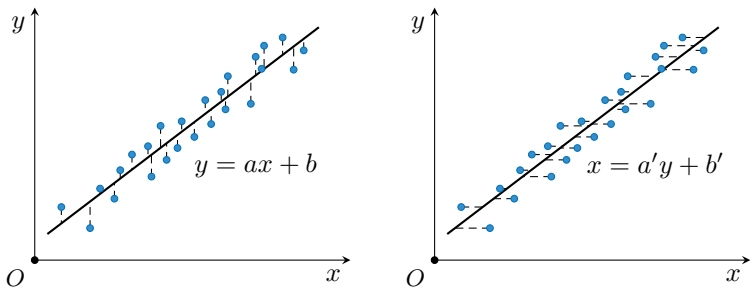


图 1.7　分别用 $y = ax + b$ 和 $x = a'y + b'$ 拟合散点，得到不同的直线方程

模型的目标函数（见 (1.28)）则反映的是各个散点在 X 轴方向到拟合直线的距离的平方和（图 1.8b ）. 即在 $y = ax + b$ 中，x 为自变量，y 为因变量，此时的最小二乘解把因变量 y 的观测误差降到最小. 而在 $x = a'y + b'$，y 为自变量，x 为因变量，此时的最小二乘解则使得因变量 x 的观测误差降至最小. 总的来说，最小二乘法总是使得因变量的观测误差达到最小.



(a)　　(b)

图 1.8　当 x 和 y 互为自变量和因变量时直线拟合的直观解释 (a) x 为自变量 (b) y 为自变量

从上面的讨论可以看出，对于给定的观测数据，选哪个变量作为自变量，哪个变量作为因变量将会对最终的最小二乘结果产生较大的影响. 那么，在实际的应用中，该如何界定自变量与因变量呢？一个基本的原则就是，应当选择观测误差小甚至没有观测误差的变量作为自变量；相应地，观测误差大的变量应当被选择为因变量.

真实的数据可能存在所有的观测变量误差都比较大的情况，此时用前面任何一种最小二乘法都可能导致较大的误差. 这种情况下，各个变量地位相当，没有明显的因果关系，解决此类问题一般需要引入总体最小二乘法. 图 1.9 给出两个变量时总体最小二乘法的直观解释. 它最终得到的直线将会使各个散点到该直线距离（即垂直距

离或垂线距离）的平方和最小. 由于总体最小二乘可以归结为主成分分析，我们将在第 2 章对其展开进一步讨论，这里就不再赘述.

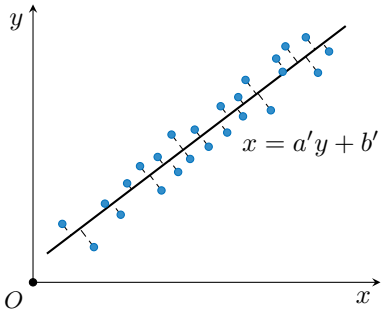


图 1.9　总体最小二乘法的直观解释，此时变量 x 和 y 都有较大的误差，且没有明显的因果关系

此外，从几何角度，当以 $x = a'y + b'$ 来拟合这组散点时，在样本空间中相当于寻找向量 $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$ 到向量 $\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^{\mathrm{T}}$ 与 $\mathbf{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^{\mathrm{T}}$ 所张成的平面的投影点 $\boldsymbol{x}'$（图 1.10 ）. 显然，这与图 1.6 所表示的是完全不同的几何关系. 这再次说明了用 $y = ax + b$ 进行直线拟合与用 $x = a'y + b'$ 进行直线拟合一般情况下将得到不同的结果.

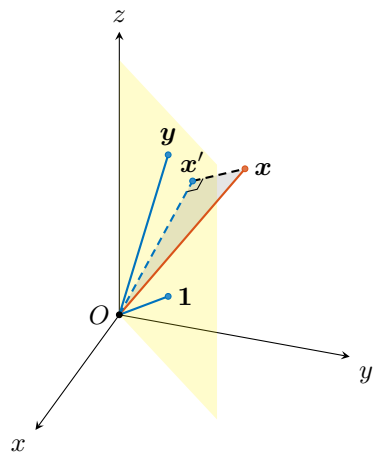


图 1.10　最小二乘法的几何解释：其中 $\boldsymbol{x} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^{\mathrm{T}}$，$\boldsymbol{y} = \begin{bmatrix} 1 & 1 & 3 \end{bmatrix}^{\mathrm{T}}$，$\mathbf{1} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^{\mathrm{T}}$, 最小二乘法对应着三个向量之间的投影关系. 当以 $x = a'y + b'$ 来拟合散点时，对应到样本空间，相当于寻求向量 $\boldsymbol{x}$ 到向量 $\boldsymbol{y}$ 和 $\mathbf{1}$ 所张成平面（$\mathrm{span}(\boldsymbol{y}, \mathbf{1})$）上的投影 $\boldsymbol{x}'$

1.6.2 约束问题

上述所有最小二乘法的理论和实例，均对未知量没有任何约束. 实际应用中，这些变量可能会有实际的物理意义，因此需要对它们添加相应的约束以满足相应的物理性质.

仍然假设 (1.1) 为矛盾方程组，同时假设未知变量 $\boldsymbol{x}$ 需要满足 m 个等式约束 $h_i(\boldsymbol{x})=0$，$i=1,2,\cdots,m$ 和 p 个不等式约束 $\mathrm{g}_j(\boldsymbol{x})\leq 0$，$j=1,2,\cdots,p$，则相应的最优化模型为

$$\begin{cases}\min\limits_{\boldsymbol{x}} & f(\boldsymbol{x})=\|\mathbf{A}\boldsymbol{x}-\boldsymbol{b}\|^2\\ \text{s.t.} & h_i(\boldsymbol{x})=0,\quad i=1,2,\cdots,m\cdot\\ & \mathrm{g}_j(\boldsymbol{x})\leq 0,\quad j=1,2,\cdots,p\end{cases}\tag{1.30}$$

一般情况下，此模型不存在类似于 (1.10) 的解析解. 在经典的最优化理论里面提供了解决这类问题的诸多方法，并且很多方法都被收录在诸多应用软件中，因此这里就不再对此问题展开讨论. 下面用一个简单的例子，从几何上给约束最小二乘一个直观的解读.

假设系数矩阵 $\mathbf{A}$ 由三维空间中的两个向量组成，即 $\mathbf{A}=\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2\end{bmatrix}$. 为了直观起见，不妨假设 $\boldsymbol{a}_1,\boldsymbol{a}_2$ 都在平面 Oxy 内（如图 1.11 所示），并且分别用点 A_1,A_2 表示. 同时假设常数项向量 $\boldsymbol{b}$ 在 Oyz 平面，并用点 B 表示.

当没有任何约束项时，线性方程组的求解问题等价于求解下面的最优化模型，

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})=\|\mathbf{A}\boldsymbol{x}-\boldsymbol{b}\|^2.\tag{1.31}$$

正如 (1.10) 所给出的，此模型存在解析解，并且此模型的解等价于点 B 往 OA_1,OA_2 所张成的平面的投影，对应投影点 B_1.

如果只考虑等式约束，并假设只有一个等式约束 $\mathbf{1}^{\mathrm{T}}\boldsymbol{x}=1$，其中 $\mathbf{1}=\begin{bmatrix}1 & 1\end{bmatrix}^{\mathrm{T}}$，即 $\boldsymbol{x}$ 的两个分量之和为 1. 此时，线性方程组的求解问题可以转化为如下等式约束的最优化问题

$$\begin{cases}\min\limits_{\boldsymbol{x}} & f(\boldsymbol{x})=\|\mathbf{A}\boldsymbol{x}-\boldsymbol{b}\|^2\\ \text{s.t.} & \mathbf{1}^{\mathrm{T}}\boldsymbol{x}=1\end{cases}.\tag{1.32}$$

此模型也存在解析解（具体请参考第 6.3 节），并且此模型的解等价于点 B 往 A_1,A_2 两个点所在直线的投影，对应投影点 B_2.

如果只考虑不等式约束，并假设只有一个不等式约束 $\boldsymbol{x}\geq\mathbf{0}$，即 $\boldsymbol{x}$ 的所有分量都是非负的. 此时，线性方程组的求解问题可以转化为如下非负约束的最优化问题

$$\begin{cases}\min\limits_{\boldsymbol{x}} & f(\boldsymbol{x})=\|\mathbf{A}\boldsymbol{x}-\boldsymbol{b}\|^2\\ \text{s.t.} & \boldsymbol{x}\geq\mathbf{0}\end{cases}.\tag{1.33}$$

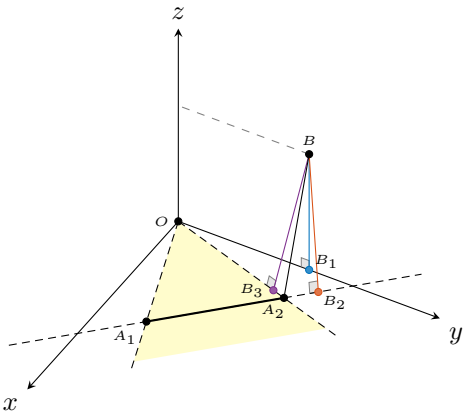


图 1.11　约束最小二乘法：点 B 的 4 种不同投影方式对应着 4 种不同约束的最小二乘法，其中 B_1 对应着模型的无约束最小二乘解，B_2 对应着模型的等式约束最小二乘解，B_3 对应着模型的非负约束最小二乘解，$B_4(A_2)$ 对应着模型非负、等式约束最小二乘解

此模型不存在解析解，但可以通过最优化理论中的罚函数、有效集等方法得到该模型的最优解. 此模型的解等价于点 B 往两个点 A_1, A_2 所对应的向量所夹的黄色区域的投影，对应投影点 B_3（此时的投影点可以理解为点 B 到黄色区域的最短距离点）.

如果同时考虑上述的等式约束和非负约束，此时，线性方程组的求解问题可以转化为如下两个约束的最优化问题

$$\begin{cases} \min\limits_{\boldsymbol{x}} & f(\boldsymbol{x}) = \|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \\ \text{s.t.} & \boldsymbol{x}^{\mathrm{T}}\mathbf{1} = 1, \boldsymbol{x} \geq \mathbf{0} \end{cases}. \tag{1.34}$$

此模型也不存在解析解，但是也可以用罚函数、有效集等方法得到该模型的最优解. 此模型的解等价于点 B 往 A_1, A_2 两个点所连接的线段的投影，对应投影点 B_4（此时的投影点可以理解为点 B 到该线段的最短距离点）. 值得注意的是，此时的投影点 B_4 与点 A_2 是重合的，因此在图 1.11 中我们并没有标出 B_4.

1.6.3 病态问题

对于一个线性方程组 $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$，当其系数矩阵的条件数很大时，此方程组为病态方程组. 此时，$\mathbf{A}$ 或者 $\boldsymbol{b}$ 的微小扰动都可能导致方程组的解 $\boldsymbol{x}$ 发生较大的变动. 比如线性方程组

$$\begin{cases} 5x + 7y = 0.5 \\ 7x + 10y = 0.7 \end{cases}, \tag{1.35}$$

的解为 $x=0.1, y=0$. 如果我们对 (1.35) 中右边的常数项 $\boldsymbol{b}$ 进行微调，使得线性方程组变为

$$\begin{cases} 5x+7y=0.51 \\ 7x+10y=0.69 \end{cases},$$

则方程的解变为 $x=0.27, y=-0.12$. 可见 $\boldsymbol{b}$ 的微小变动引起了方程的解的较大变动. 之所以会产生这个现象，正是因为该方程组的系数矩阵具有较大的条件数. 接下来我们首先给出矩阵的条件数的定义.

定义 1.2（矩阵的条件数） 矩阵 $\mathbf{A}$ 的条件数等于 $\mathbf{A}$ 的范数与 $\mathbf{A}^{-1}$ 的范数的乘积（矩阵范数相关定义见附录 A ），即 $\operatorname{cond}(\mathbf{A})=\|\mathbf{A}\|\|\mathbf{A}^{-1}\|$.

可以验证，当取矩阵的 2-范数时，矩阵 $\mathbf{A}$ 的条件数等于该矩阵的最大奇异值与最小奇异值之比，即

$$\operatorname{cond}(\mathbf{A})=\frac{\sigma_{\max}}{\sigma_{\min}},$$

其中 $\sigma_{\max}$ 为矩阵 $\mathbf{A}$ 的最大奇异值，而 $\sigma_{\min}$ 为 $\mathbf{A}$ 的最小奇异值.

在 (1.35) 中，计算可得，系数矩阵 $\mathbf{A}$ 有两个奇异值，分别为 $\sigma_1=14.933, \sigma_2=0.067$. 显然，矩阵 $\mathbf{A}$ 的两个奇异值相差悬殊，这必然导致 $\mathbf{A}$ 具有较大的条件数

$$\operatorname{cond}(\mathbf{A})=\frac{\sigma_1}{\sigma_2}=222.9955.$$

相应地，线性方程组 (1.35) 为病态方程组.

为了缓解线性方程组中系数矩阵的病态问题，常用的方法是在优化模型 (1.9) 中加入一个正则项（这也是岭回归[1]的思想），得到如下最优化模型

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})=\|\mathbf{A}\boldsymbol{x}-\boldsymbol{b}\|^2+\lambda\|\boldsymbol{x}\|^2. \tag{1.36}$$

这里 λ 为一个需要人为设定的正数. 为了得到 (1.36) 的解，我们可以先计算目标函数对自变量的导数，并令其等于零向量

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=2\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{x}-2\mathbf{A}^{\mathrm{T}}\boldsymbol{b}+2\lambda\boldsymbol{x}\overset{\text{令}}{=}\mathbf{0},$$

可得

$$\boldsymbol{x}=\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}+\lambda\mathbf{I}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{b}. \tag{1.37}$$

其中 $\mathbf{I}$ 为相应阶数的单位矩阵.

当矩阵 $\mathbf{A}$ 的条件数很大时，$\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 必然包含一个相对特别小的特征值，而对应的逆矩阵 $\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}$ 必然对某些方向（比如属于该特征值的特征向量方向）的向量具有极大的放大作用，这正是相应的线性方程组为病态方程组的原因所在. 而 (1.37) 中扰动项 $\lambda\mathbf{I}$ 的加入，使得 $\mathbf{A}^{\mathrm{T}}\mathbf{A}+\lambda\mathbf{I}$ 的条件数一般情况下都要远小于 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 的条件数，从而可以一定程度克服线性方程组的病态问题.

1.6.4 异常问题

在直线拟合问题中，当因变量和自变量的观测点的分布呈现一定的线性关系的时候，可以用最小二乘法得到很好的拟合结果. 但在实际应用中，当观测点受到噪声污染的时候（如图 1.12 ），直接利用最小二乘法对包括噪声点在内的所有观测点进行直线拟合在很多情况下难以得到理想的结果.

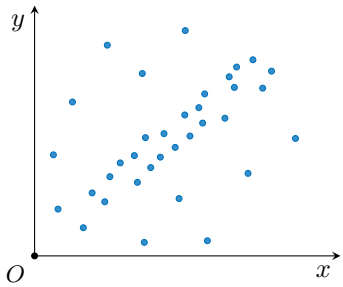


图 1.12　受噪声污染的观测数据

为了解决噪声污染下的直线拟合问题，研究人员发展了各种策略，其中常用的包括加权最小二乘法、Huber 回归、随机采样一致性（Random Sample Consensus, RANSAC）[2]等方法. 接下来，简要介绍一下 RANSAC 的大体思路.

利用 RANSAC 进行直线拟合可以分为初筛和拟合两个阶段. 在初筛阶段，随机给出包含 s 个散点的原始观测点的一个子集，然后对这 s 个点利用最小二乘法进行直线拟合. 设定一个阈值 d，记录下所有的观测点中与该直线的距离小于 d 的点（称为内点），并将所有的内点存储在一个集合 S_i 中. 重复以上步骤，当出现内点个数增加的情形，则更新 S_i. 对最终的内点集合 S_i 进行直线拟合即为我们想要的最终结果. 图 1.13 ∼ 图 1.15 给出了初筛过程的简单示意图，从中可以看出，图 1.15 包含了更多的内点，利用这些内点进行直线拟合显然会得到我们想要的结果.

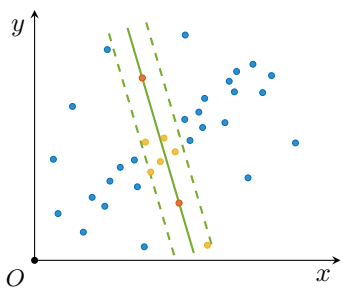


图 1.13　随机采样一致性. 随机子集的散点标为红色，距离随机子集的拟合直线的距离小于给定阈值的点标为黄色（本图中有 6 个点）

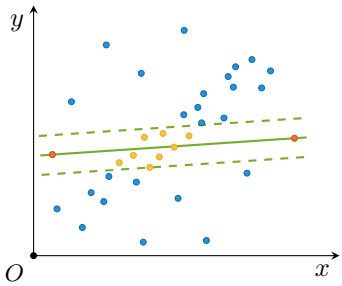


图 1.14　随机采样一致性. 随机子集的散点标为红色，距离随机子集的拟合直线的距离小于给定阈值的点标为黄色（本图中有 8 个点）

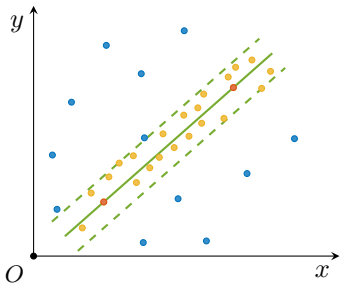


图 1.15　随机采样一致性. 随机子集的散点标为红色，距离随机子集的拟合直线的距离小于给定阈值的点标为黄色（本图中有 21 个点）

1.6.5 目标函数问题

最小二乘法不但可以用于直线拟合，而且可以用于曲线拟合；不仅可以处理单变量的观测数据，同时也可以处理多变量的观测数据（对应曲面拟合）.

给定一组观测数据 $(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$，下面给出用 m 次多项式来拟合这些散点的过程

$$y = f(x) = a_0 x^m + a_1 x^{m-1} + \cdots + a_{m-1} x + a_m.$$

注意到，参数 $\boldsymbol{a} = \begin{bmatrix} a_0 & a_1 & \cdots & a_m \end{bmatrix}^{\mathrm{T}}$ 的确定等价于下面的最优化问题

$$\min_{\boldsymbol{a}} g(\boldsymbol{a}) = \|\mathbf{X}\boldsymbol{a} - \boldsymbol{y}\|^2, \tag{1.38}$$

其中

$$\begin{aligned}
\boldsymbol{x}_i &= \begin{bmatrix} x_1^i & x_2^i & \cdots & x_n^i \end{bmatrix}^{\mathrm{T}}, i = 1, 2, \cdots, m, \\
\boldsymbol{y} &= \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^{\mathrm{T}}, \\
\mathbf{1} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^{\mathrm{T}}, \\
\mathbf{X} &= \begin{bmatrix} \boldsymbol{x}_m & \cdots & \boldsymbol{x}_1 & \mathbf{1} \end{bmatrix}^{\mathrm{T}},
\end{aligned} \quad .$$

则很容易得到 (1.38) 的最小二乘解为

$$\boldsymbol{a} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}. \tag{1.39}$$

比较 (1.15) 可以看出，直线拟合和曲线拟合并没有本质的区别.

在多数情况下，对于一组给定的观测点，我们首先需要选取适当的目标函数，然后利用最小二乘法对其进行直线或者曲线拟合. 但有些情况下，我们很难用一个简单的函数来描述给定的散点分布，如图 1.16 . 对于这种情况的曲线拟合，该如何进行处理呢？

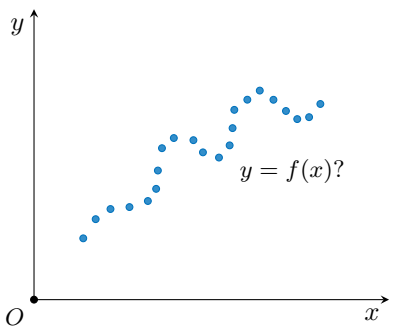


图 1.16　未知目标函数表达式的曲线拟合问题

第一种常用的方法是采用多项式对上述散点进行拟合. 如果多项式的次数设定合理，在多数情况下，多项式拟合都会得到不错的结果. 但在多项式拟合中，多项式的次数是一个需要人为设定的量，不合理的次数会导致过拟合或者欠拟合现象.

第二种常用的方法则是利用局部最小二乘法对任意形状的散点分布进行曲线拟合. 假设最终的拟合函数为 $y = f(x)$，自变量 x 的每一个位置处，相应的 y 值只需要利用该位置附近的若干观测点进行直线拟合确定. 该方法隐含了一个前提，即数据的分布在局部是线性或者接近线性的. 在观测点足够多，分布足够密的时候，利用局部最小二乘法一般都会得到不错的结果.

1.7 小　　结

至此，本章的内容总结为以下 6 条：

1. 从行空间角度，线性方程组的解为各个方程所对应的超平面的交集.
2. 从列空间角度，线性方程组的常数项向量为系数矩阵的各个列向量的线性表出，且表出系数即为待求的解.
3. 最小二乘法是求解矛盾方程组的常用方法.
4. 代数上，线性方程组的最小二乘解对应着矩阵的广义逆操作.
5. 几何上，线性方程组的最小二乘解等价于线性方程组的常数项向量在系数矩阵的列空间的正交投影.
6. 概率上，最小二乘法基于线性方程组模型误差服从高斯分布.

第 2 章 主成分分析

主成分分析（Principal Component Analysis, PCA）是最常用的数据降维手段. 本章首先介绍一些基本统计概念，然后给出主成分分析的模型，并推导其求解方法，接着从几何、子空间逼近、概率、信息论等角度对主成分分析给出不同的诠释.

2.1 问题背景

主成分分析首先是由卡尔. 皮尔森（Karl Pearson）在 1901 年引入的，但当时只针对非随机变量进行讨论. 之后哈罗德 • 霍特林（Harold Hotelling）将此方法推广到随机向量的情形. 主成分分析自其诞生以来，已经在诸多领域得到了极为广泛的应用.

下面我们首先从一个简单的例子入手，讨论一下主成分分析的应用背景. 对于图 2.1 中线段上的 5 个点 A, B, C, D, E，可以采用多种方式对其进行定量描述.

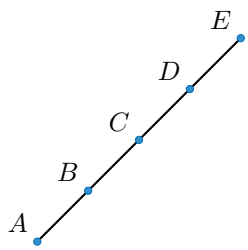


图 2.1　线段上的 5 个点

首先，我们可以把该线段放在实轴上（图 2.2 ），则可得到这 5 个点在实轴上的坐标分别为 a_1, a_2, a_3, a_4, a_5. 可以将其表示为向量的形式

$$\boldsymbol{x} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \end{bmatrix}. \tag{2.1}$$

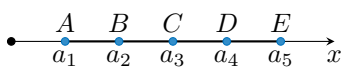


图 2.2　在实轴上，A, B, C, D, E 这 5 个点可以用它们的坐标定量描述，分别为 a_1, a_2, a_3, a_4, a_5

也可以把线段放在二维平面的直角坐标系中（图 2.3 ），这样就可以得到这 5 个点的坐标分别为 (x_1, y_1)，(x_2, y_2)，(x_3, y_3)，(x_4, y_4)，(x_5, y_5). 将其表示为矩阵形式，

有

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ y_1 & y_2 & y_3 & y_4 & y_5 \end{bmatrix}. \tag{2.2}$$

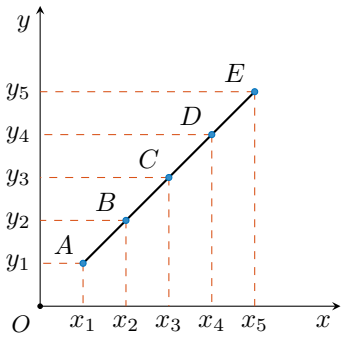


图 2.3　在平面直角坐标系中，A, B, C, D, E 这 5 个点可以用如下 5 个坐标定量描述: (x_1, y_1)，(x_2, y_2)，(x_3, y_3)，(x_4, y_4)，(x_5, y_5)

如果把线段放在三维空间的直角坐标系中（图 2.4 ），就可以得到这 5 个点的坐标分别为 $(\alpha_1, \beta_1, \gamma_1)$，$(\alpha_2, \beta_2, \gamma_2)$，$(\alpha_3, \beta_3, \gamma_3)$，$(\alpha_4, \beta_4, \gamma_4)$，$(\alpha_5, \beta_5, \gamma_5)$. 将其表示为矩阵形式，有

$$\mathbf{X} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \\ \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 \end{bmatrix}. \tag{2.3}$$

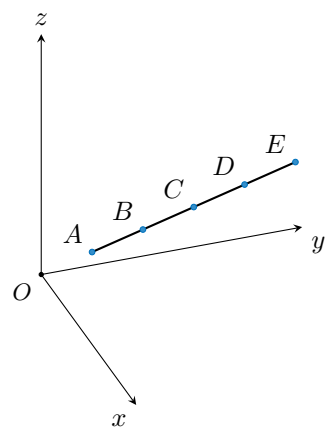


图 2.4　在三维空间直角坐标系中，A, B, C, D, E 这 5 个点可以用如下 5 个坐标定量描述：$(\alpha_1, \beta_1, \gamma_1)$，$(\alpha_2, \beta_2, \gamma_2)$，$(\alpha_3, \beta_3, \gamma_3)$，$(\alpha_4, \beta_4, \gamma_4)$，$(\alpha_5, \beta_5, \gamma_5)$

从 (2.1) 到 (2.3) 可以看出，尽管它们的表达形式各不相同，但是它们所表达的是同一个对象，即线段上的 5 个点. 很显然，这 5 个点的本征维度为 1，用一个维度或一个特征足以表达出这 5 个点包含的所有信息. 公式 (2.2) 和 (2.3) 虽然也可以完整地表达出这 5 个点的全部信息，但相对于公式 (2.1)，它们分别用了两个特征和三个特征，对应的表达则显得冗余且不直观，因此通常采用 (2.1) 对这 5 个点进行定量表达.

上面的例子表明，在对象本征维度已知的情况下，可以选用合适的坐标系简洁、直观地定量表达对象. 遗憾的是，在实际应用中，给定任意一个 p 个特征，n 个观测的数据对象

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}.$$

我们往往很难直观地判断出该矩阵的本征维度. 因此，给定一个数据对象，如何有效判断出该数据的本征维度，将是一个理论和应用上都非常有意义的问题. 主成分分析正是处理此类问题的最为常用的手段.

2.2 基本统计概念

在讨论主成分分析的基本原理之前，本节首先给出相关的基本数学概念.

2.2.1 随机变量的数字特征

定义 2.1 *若 X 为离散型随机变量，其概率质量函数为 $P(X = x_i) = p_i, i = 1, 2, \cdots$，如果级数 $\sum_{i=1}^{\infty} x_i p_i$ 绝对收敛，则称其为 X 的*数学期望*，简称 X 的*期望*，记作 $\mathrm{E}\, X$ 或 $\mathrm{mean}(X)$，即*

$$\mathrm{E}\, X = \mathrm{mean}(X) = \sum_{i=1}^{\infty} x_i p_i.$$

*若 X 为连续型随机变量，其概率密度函数为 $f(x)$，如果积分 $\int_{-\infty}^{+\infty} x f(x) dx$ 绝对收敛，则称其为 X 的*期望*，即*

$$\mathrm{E}\, X = \mathrm{mean}(X) = \int_{-\infty}^{+\infty} x f(x) dx.$$

定义 2.2 *若 X 为随机变量，且 $\mathrm{E}(X - \mathrm{E}\, X)^2$ 存在，则称其为 X 的*方差*，记作 $\mathrm{D}\, X$ 或 $\mathrm{var}(X)$，即*

$$\mathrm{D}\, X = \mathrm{var}(X) = \mathrm{E}(X - \mathrm{E}\, X)^2.$$

定义 2.3 设 X 和 Y 是两个随机变量，若 $\mathrm{E}((X-\mathrm{E}\,X)(Y-\mathrm{E}\,Y))$ 存在，则称其为随机变量 X 与 Y 的**协方差**，记为 $\mathrm{cov}(X,Y)$，即

$$\mathrm{cov}(X,Y)=\mathrm{E}((X-\mathrm{E}\,X)(Y-\mathrm{E}\,Y)),$$

相应地，称

$$\rho_{XY}=\frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{D}\,X\,\mathrm{D}\,Y}},$$

为 X 与 Y 的**相关系数**.

定义 2.4 设 X 与 Y 是两个随机变量，若 $\mathrm{E}\,X^k(k=1,2,\cdots)$ 存在，则称它为 X 的 k 阶**原点矩**；若 $\mathrm{E}(X-\mathrm{E}\,X)^k(k=1,2,\cdots)$ 存在，则称它为 X 的 k 阶**中心矩**；若 $\mathrm{E}\,X^kY^l(k,l=1,2,\cdots)$ 存在，则称它为 X 和 Y 的 $k+l$ 阶**混合原点矩**；若 $\mathrm{E}(X-\mathrm{E}\,X)^k(Y-\mathrm{E}\,Y)^l$ 存在，则称它为 X 和 Y 的 $k+l$ 阶**混合中心矩**.

显然，数学期望 $\mathrm{E}\,X$ 是随机变量 X 的一阶原点矩；方差 $\mathrm{D}\,X$ 是 X 的二阶中心矩；协方差 $\mathrm{cov}(X,Y)$ 是 X 与 Y 的二阶混合中心矩.

定义 2.5 设 n 维随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$，则称

$$\boldsymbol{\mu}=\mathrm{E}\,\boldsymbol{X}=\mathrm{mean}(X_1,X_2,\cdots,X_n)=(\mathrm{E}\,X_1,\mathrm{E}\,X_2,\cdots,\mathrm{E}\,X_n),$$

为 $\boldsymbol{X}$ 的**数学期望**，或简称 $\boldsymbol{X}$ 的**期望**.

定义 2.6 设 n 维随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$，记

$$\sigma_{ij}=\mathrm{cov}(X_i,X_j)=\mathrm{E}(X_i-\mathrm{E}\,X_i)(X_j-\mathrm{E}\,X_j),i,j=1,2,\cdots,n,$$

则称矩阵

$$\boldsymbol{\Sigma}=\mathrm{cov}(\boldsymbol{X})=\begin{bmatrix}\sigma_{11}&\sigma_{12}&\cdots&\sigma_{1n}\\ \sigma_{21}&\sigma_{22}&\cdots&\sigma_{2n}\\ \vdots&\vdots&\ddots&\vdots\\ \sigma_{n1}&\sigma_{n2}&\cdots&\sigma_{nn}\end{bmatrix},$$

为随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 的**协方差矩阵**. 如果记

$$\rho_{ij}=\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}},i,j=1,2,\cdots,n,$$

则称矩阵

$$\mathrm{corr}(\boldsymbol{X})=\begin{bmatrix}\rho_{11}&\rho_{12}&\cdots&\rho_{1n}\\ \rho_{21}&\rho_{22}&\cdots&\rho_{2n}\\ \vdots&\vdots&\ddots&\vdots\\ \rho_{n1}&\rho_{n2}&\cdots&\rho_{nn}\end{bmatrix},$$

为随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 的相关矩阵或相关系数矩阵.

在实际应用中，需要处理的往往是随机变量的观测数据，因此，接下来我们将给出几个常用的样本统计量[1].

2.2.2 样本统计量

定义 2.7 给定随机变量 X 的 n 个观测值 $x_1, x_2, \cdots, x_n$，分别称[2]

$$\mu = \mathrm{mean}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{2.4}$$

$$\sigma^2 = \mathrm{var}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2, \tag{2.5}$$

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k, k = 1, 2, \cdots, \tag{2.6}$$

$$m'_k = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^k, k = 1, 2, \cdots, \tag{2.7}$$

为随机变量 X 的样本 $x_1, x_2, \cdots, x_n$ 的样本均值、样本方差[3]、样本 k 阶原点矩和样本 k 阶中心矩. 其中 $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$ 为由 $x_1, x_2, \cdots, x_n$ 这 n 个观测构成的样本列向量.

定义 2.8 给定随机变量 X 的 n 个观测值 $x_1, x_2, \cdots, x_n$，以及随机变量 Y 的观测值 $y_1, y_2, \cdots, y_n$，称

$$\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu_{\boldsymbol{x}})(y_i - \mu_{\boldsymbol{y}}), \tag{2.8}$$

为随机变量 X 和 Y 的样本 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的样本协方差，其中 $\boldsymbol{x}$ 定义如上，而 $\boldsymbol{y}$ 是由 $y_1, y_2, \cdots, y_n$ 这 n 个观测构成的样本列向量 $\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^{\mathrm{T}}$，此外，$\mu_{\boldsymbol{x}}$ 和 $\mu_{\boldsymbol{y}}$ 分别为 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的均值.

---

[1] 样本统计量和总体参数（总体统计量）是统计学中两个基本的概念，一般用数据的样本统计量估计它的总体参数. 为了方便起见，本书对样本统计量和总体参数的符号表达不做区分.

[2] 对于 $\mathrm{mean}(\cdot)$，$\mathrm{var}(\cdot)$ 等符号，当括号中是随机变量时，它们表示的是随机变量的数字特征，而当括号中是观测向量时，它们表示的是观测数据的样本统计量.

[3] 样本方差的计算也可以采用公式 $\mathrm{var}(\boldsymbol{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$. 其中系数的选取取决于随机变量的均值情况. 当均值已知时，系数使用 $\frac{1}{n}$，而当均值为估计值的时候，系数使用 $\frac{1}{n-1}$. 在本书中，为了简单起见，如果不做特别说明，系数均取 $\frac{1}{n}$.

给定 p 维随机向量 $\boldsymbol{X}$ 的 n 个观测，我们可以得到一个 $p \times n$ 的观测矩阵 $\mathbf{X}$，即

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}. \tag{2.9}$$

此时，可以认为观测矩阵由随机向量的 n 个观测构成，即

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix},$$

其中 $\boldsymbol{x}_i = \begin{bmatrix} x_{1i} & x_{2i} & \cdots & x_{pi} \end{bmatrix}^{\mathrm{T}}, i = 1, 2, \cdots, n$ 为随机向量 $\boldsymbol{X}$ 的第 i 个观测. 也可以认为观测矩阵由 p 个随机变量的样本观测行向量构成，即

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix},$$

其中样本 $X_i = \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{in} \end{bmatrix}, i = 1, 2, \cdots, p$ 由随机向量 $\boldsymbol{X}$ 的第 i 个分量的 n 个观测构成[4].

定义 2.9 给定 p 维随机向量 $\boldsymbol{X}$ 的 n 个观测，即

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix},$$

称

$$\boldsymbol{\mu} = \mathrm{mean}(\mathbf{X}) = \begin{bmatrix} \mathrm{mean}(X_1) \\ \mathrm{mean}(X_2) \\ \vdots \\ \mathrm{mean}(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}, \tag{2.10}$$

为随机向量 $\boldsymbol{X}$ 的样本 $\mathbf{X}$ 的**样本均值向量**. 其中

$$\mu_i = \mathrm{mean}(X_i) = \frac{1}{n} \sum_{k=1}^{n} x_{ik}, i = 1, 2, \cdots, p,$$

[4]这里的 X_i 表示的是行向量，而在前文中大写斜体字母也被用来表示随机变量. 对于一个给定的大写斜体符号，读者可以结合上下文判断其含义.

为 X_i 的样本均值. 记

$$\sigma_{ij} = \mathrm{cov}(X_i, X_j) = \frac{1}{n}\sum_{k=1}^{n}(x_{ik} - \mu_i)(x_{jk} - \mu_j), i, j = 1, 2, \cdots, p,$$

为 X_i, X_j 的样本协方差. 则

$$\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}, \tag{2.11}$$

为随机向量 $\boldsymbol{X}$ 的样本 $\mathbf{X}$ 的**样本协方差矩阵**，简称为 $\mathbf{X}$ 的协方差矩阵.

此外，随机向量的其它样本统计量也可以给出类似的定义，这里就不再赘述. 可以认为，各个样本统计量是随机变量相应数字特征的估计值.

2.2.3 样本统计量的向量表示

1. 样本均值

给定随机变量 X 的 n 个观测值 $x_1, x_2, \cdots, x_n$，样本均值的计算可以表示为

$$\mu = \mathrm{mean}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\mathbf{1}^{\mathrm{T}}\boldsymbol{x}, \tag{2.12}$$

其中 $\mathbf{1}$ 是所有元素均为 1 的 n 维列向量.

2. 样本方差

给定随机变量 X 的 n 个观测值 $x_1, x_2, \cdots, x_n$，样本方差的计算可以表示为

$$\sigma^2 = \mathrm{var}(\boldsymbol{x}) = \frac{1}{n}(\boldsymbol{x} - \mu\mathbf{1})^{\mathrm{T}}(\boldsymbol{x} - \mu\mathbf{1}). \tag{2.13}$$

当样本均值 $\mu = 0$ 时，样本方差可以简单地表示为

$$\sigma^2 = \mathrm{var}(\boldsymbol{x}) = \frac{1}{n}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}. \tag{2.14}$$

需要注意的是，如果样本向量 $\boldsymbol{x}$ 为行向量，则需要将公式 (2.14) 调整为

$$\sigma^2 = \mathrm{var}(\boldsymbol{x}) = \frac{1}{n}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}. \tag{2.15}$$

3. 样本协方差

给定随机变量 X 的 n 个观测值 $x_1, x_2, \cdots, x_n$，以及随机变量 Y 的 n 个观测值 $y_1, y_2, \cdots, y_n$，它们的样本协方差的计算可以表示为

$$\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n}(\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})^{\mathrm{T}}(\boldsymbol{y} - \mu_{\boldsymbol{y}}\mathbf{1}). \tag{2.16}$$

当样本均值 $\mu_{\boldsymbol{x}} = 0, \mu_{\boldsymbol{y}} = 0$ 时，样本协方差可以简单地表示为

$$\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}. \tag{2.17}$$

4. 样本均值向量

给定 p 维随机向量 $\boldsymbol{X}$ 的 n 个观测值 $\mathbf{X}$（见 (2.9)），其样本均值向量的计算可以表示为

$$\boldsymbol{\mu} = \frac{1}{n}\mathbf{X}\mathbf{1}. \tag{2.18}$$

5. 样本协方差矩阵

给定 p 维随机向量 $\boldsymbol{X}$ 的 n 个观测值 $\mathbf{X}$，$\mathbf{X}$ 的样本协方差矩阵可以表示为

$$\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{X}) = \frac{1}{n}(\mathbf{X} - \boldsymbol{\mu}\mathbf{1}^{\mathrm{T}})(\mathbf{X} - \boldsymbol{\mu}\mathbf{1}^{\mathrm{T}})^{\mathrm{T}} = \frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}} - \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}. \tag{2.19}$$

当样本均值向量为零向量时，即 $\boldsymbol{\mu} = \mathbf{0}$，样本协方差矩阵可以简单地表示为

$$\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{X}) = \frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}. \tag{2.20}$$

值得注意的是，(2.19) 中 $(\mathbf{X} - \boldsymbol{\mu}\mathbf{1}^{\mathrm{T}})$ 这一项可以表示为

$$\mathbf{X} - \boldsymbol{\mu}\mathbf{1}^{\mathrm{T}} = \mathbf{X} - \frac{1}{n}\mathbf{X}\mathbf{1}\mathbf{1}^{\mathrm{T}} = \mathbf{X}\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right) = \mathbf{X}\mathbf{P}_{\mathbf{1}}^{\perp}, \tag{2.21}$$

其中，$\mathbf{P}_{\mathbf{1}}^{\perp} = \mathbf{I} - \mathbf{1}\mathbf{1}^{\dagger} = \mathbf{I} - \mathbf{1}\left(\mathbf{1}^{\mathrm{T}}\mathbf{1}\right)^{-1}\mathbf{1}^{\mathrm{T}} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}$ 为 n 维样本空间中元素全为 1 的向量 $\mathbf{1}$ 的正交补投影算子. 从 (2.21) 可以看出，数据在 p 维特征空间的中心化操作对应着其在 n 维样本空间的正交投影.

此外，关于 $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 的计算，可以根据 $\mathbf{X}$ 的不同分块方式，给出两种不同的理解方式. 当 $\mathbf{X}$ 按照如下方式分块时

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix},$$

矩阵 $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 中的每一个元素都可以认为是矩阵 $\mathbf{X}$ 中相应的两个行向量的内积，即

$$\mathbf{X}\mathbf{X}^{\mathrm{T}} = \begin{bmatrix} X_1X_1^{\mathrm{T}} & X_1X_2^{\mathrm{T}} & \cdots & X_1X_p^{\mathrm{T}} \\ X_2X_1^{\mathrm{T}} & X_2X_2^{\mathrm{T}} & \cdots & X_2X_p^{\mathrm{T}} \\ \vdots & \vdots & \ddots & \vdots \\ X_pX_1^{\mathrm{T}} & X_pX_2^{\mathrm{T}} & \cdots & X_pX_p^{\mathrm{T}} \end{bmatrix}. \tag{2.22}$$

当 $\mathbf{X}$ 按照如下方式分块时

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix},$$

矩阵 $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 可以认为是 n 个秩 1 矩阵的和，即

$$\mathbf{X}\mathbf{X}^{\mathrm{T}} = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}. \tag{2.23}$$

接下来以一个简单的 2×3 大小的矩阵

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix},$$

为例 $(p = 2, n = 3)$ 直观比较公式 (2.22) 和 (2.23) 的区别.

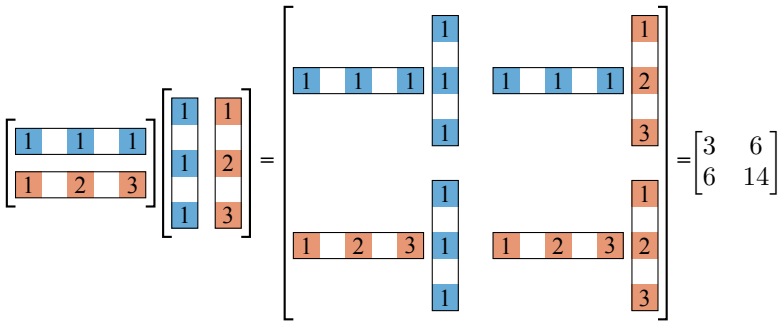


图 2.5　公式 (2.22) 矩阵乘法示意图

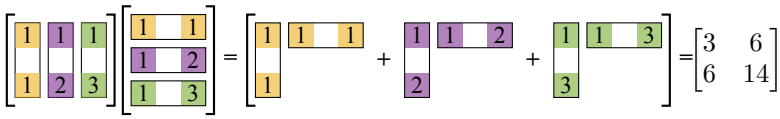


图 2.6　公式 (2.23) 矩阵乘法示意图

2.3　主成分分析的基本原理

给定一个 p 个特征，n 个观测的数据对象

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix},$$

可以将其看成 p 维空间的 n 个向量或者 n 个点. 尽管这 n 个向量在 p 维空间中，但是正如图 2.4 所展示的那样，它们的本征维度很可能不是 p. 也就是说，尽管我们用 p 维空间来承载这些观测数据，但这 n 个观测的整体信息很可能用一个更低维度的空间来表达就足够了. 那么，该如何确定数据的本征维度呢？

2.3.1 任意方向的方差

为了寻求足以承载数据所有信息的低维子空间，首先需要明确的就是用什么指标来表征数据的信息量. 主成分分析以方差为指标，通过求取数据的方差极值方向从而得到足以表达数据信息的低维子空间[3, 4]. 由于观测数据 $\mathbf{X}$ 包含 p 个特征，为了得到数据的方差极值方向 $\boldsymbol{u}=\begin{bmatrix}u_1 & u_2 & \cdots & u_p\end{bmatrix}^{\mathrm{T}}$，我们首先需要将数据都投影到 $\boldsymbol{u}$ 方向，得到

$$Y=\boldsymbol{u}^{\mathrm{T}}\mathbf{X}.$$

显然 Y 是一个包含 n 个元素的行向量，可以认为它是一个新的随机变量的 n 个观测. 为了方便起见，不妨假设 $\mathbf{X}$ 的均值向量为零向量，即 $\boldsymbol{\mu}=\mathrm{mean}(\mathbf{X})=\mathbf{0}$. 根据 (2.15) 可以得到 Y 的方差为

$$\mathrm{var}(Y)=\mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})=\frac{1}{n}\boldsymbol{u}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{u}=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u}, \tag{2.24}$$

其中 $\boldsymbol{\Sigma}=\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 是观测数据的协方差矩阵.

公式 (2.24) 清晰地表明，多元观测数据在任意方向的方差可以由数据的协方差矩阵 $\boldsymbol{\Sigma}$ 和投影方向 $\boldsymbol{u}$ 解析表达. 这个公式简洁、美观，令人震撼！首先，它进一步加深了我们对协方差矩阵的理解，即数据的协方差矩阵包含了数据的所有二阶统计信息. 其次，它极大便利了我们对方差极值方向的求解.

2.3.2 模型与求解

尽管 (2.24) 给出了数据在任意方向方差的解析表达，但是注意到该方差的大小会受到 $\boldsymbol{u}$ 的范数 $\|\boldsymbol{u}\|$ 的大小的影响. 当 $\|\boldsymbol{u}\|$ 趋于无穷大时，$\mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})$ 可以趋于无穷大；而当 $\|\boldsymbol{u}\|$ 趋于零时，$\mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})$ 也会趋于零. 因此，有必要对 $\boldsymbol{u}$ 加以限制，使得 $\mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})$ 不受 $\boldsymbol{u}$ 的长度的影响. 于是，就得到如下的主成分分析优化模型

$$\begin{cases}\max\limits_{\boldsymbol{u}} & \mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u}\\ \text{s.t.} & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1\end{cases}. \tag{2.25}$$

上述优化模型为等式约束的目标函数的极值问题，可以用拉格朗日乘子法建立如下目标函数

$$\mathcal{L}(\boldsymbol{u},\lambda)=\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u}+\frac{\lambda}{2}(1-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}). \tag{2.26}$$

该函数对 $\boldsymbol{u}$ 的偏导数如下

$$\frac{\partial\mathcal{L}(\boldsymbol{u},\lambda)}{\partial\boldsymbol{u}}=\boldsymbol{\Sigma}\boldsymbol{u}-\lambda\boldsymbol{u}.$$

函数 $\mathcal{L}(\boldsymbol{u},\lambda)$ 极值处的偏导数必然为零向量，因此，令 $\frac{\partial \mathcal{L}(\boldsymbol{u},\lambda)}{\partial \boldsymbol{u}} = \mathbf{0}$ 可得到

$$\boldsymbol{\Sigma u} = \lambda \boldsymbol{u}. \tag{2.27}$$

从 (2.27) 可以看出，观测数据方差极值方向的求取可以归结为其协方差矩阵的特征值与特征向量问题. 方程 (2.27) 两边同时左乘 $\boldsymbol{u}^{\mathrm{T}}$，结合 $\boldsymbol{u}$ 的单位向量约束，可得

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma u} = \lambda \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = \lambda. \tag{2.28}$$

公式 (2.28) 明确地告诉我们，协方差矩阵的特征值正好就是数据在对应特征向量方向的方差.

假设协方差矩阵 $\boldsymbol{\Sigma}$ 的 p 个特征值满足 $\lambda_1 \geq \cdots \geq \lambda_p$，相应的特征向量记为 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p$. 那么 $\boldsymbol{\Sigma}$ 的属于最大特征值 λ_1 的特征向量 $\boldsymbol{u}_1$ 即为数据的最大方差方向. 将数据 $\mathbf{X}$ 投影到该方向将得到数据的第一主成分

$$Y_1 = \boldsymbol{u}_1^{\mathrm{T}}\mathbf{X}.$$

接下来，我们致力于寻找数据的第二个方差极值方向，希望所求的新方向仍使得 $\mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma u}$ 达到极值，同时不受第一个方差极值方向的影响. 因此，第二个方差极值方向的求取可以归结为如下优化模型

$$\begin{cases} \max\limits_{\boldsymbol{u}} \quad \mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma u} \\ \text{s.t.} \quad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1, \quad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}_1 = 0 \end{cases}. \tag{2.29}$$

与 (2.25) 不同的是，(2.29) 通过引入正交约束来规避第一个方差方向的影响. 这事实上等价于在 $\boldsymbol{u}_1$ 的正交补空间寻找数据的方差极值方向. 同样地，可以利用拉格朗日乘子法对 (2.29) 进行求解. 首先构造 (2.29) 的拉格朗日函数为

$$\mathcal{L}(\boldsymbol{u},\lambda,\mu) = \frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma u} + \frac{1}{2}\lambda(1 - \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}) + \mu \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}_1, \tag{2.30}$$

该函数对 $\boldsymbol{u}$ 求偏导数得

$$\frac{\partial \mathcal{L}(\boldsymbol{u},\lambda,\mu)}{\partial \boldsymbol{u}} = \boldsymbol{\Sigma u} - \lambda \boldsymbol{u} + \mu \boldsymbol{u}_1,$$

令偏导数为零向量得

$$\boldsymbol{\Sigma u} = \lambda \boldsymbol{u} - \mu \boldsymbol{u}_1, \tag{2.31}$$

公式 (2.31) 两边同时左乘 $\boldsymbol{u}_1^{\mathrm{T}}$，得

$$\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{\Sigma u} - \lambda \boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{u} + \mu \boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{u}_1 = 0. \tag{2.32}$$

由于 $\boldsymbol{\Sigma u}_1 = \lambda \boldsymbol{u}_1$ 以及 $\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{u} = 0$，整理 (2.32) 可得 $\mu = 0$. 因此由 (2.31) 得 $\boldsymbol{\Sigma u} = \lambda \boldsymbol{u}$. 这意味着 (2.29) 的极值方向也可以归结为协方差矩阵 $\boldsymbol{\Sigma}$ 的特征向量问题. 又由于 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}_1 = 0$ 的约束，可以推断使得 (2.29) 达到最大值的方向必然是 $\boldsymbol{\Sigma}$ 的第二大特征值 λ_2 所对应的特征向量 $\boldsymbol{u}_2$，即 $\boldsymbol{u} = \boldsymbol{u}_2$. 以此类推，我们仅需要求解出 (2.27) 的所有非

零特征值所对应的特征向量，即可得到原始数据的所有方差极值方向. 记

$$\boldsymbol{\Lambda}=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_p\end{bmatrix},\quad \mathbf{U}=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_p\end{bmatrix},$$

它们分别为协方差矩阵 $\boldsymbol{\Sigma}$ 的特征值矩阵和特征向量矩阵. 则最终的主成分变换公式为

$$\mathbf{Y}=\mathbf{U}^{\mathrm{T}}\mathbf{X}, \tag{2.33}$$

其中前 k 个主成分的贡献率定义为

$$\eta(k)=\frac{\lambda_1+\lambda_2+\cdots+\lambda_k}{\lambda_1+\lambda_2+\cdots+\lambda_p}. \tag{2.34}$$

在应用中，可以根据给定的贡献率来确定需要保留的主成分的个数.

总结而言，对于给定的一个 p 个特征，n 个观测的数据 $\mathbf{X}$，对其进行主成分分析的主要步骤如下.

算法 2.1　主成分分析的主要步骤

1. 计算数据的均值向量且将数据中心化：$\boldsymbol{\mu}=\frac{1}{n}\mathbf{X}\mathbf{1}, \mathbf{X}=\mathbf{X}-\boldsymbol{\mu}\mathbf{1}^{\mathrm{T}}$
2. 计算协方差矩阵：$\boldsymbol{\Sigma}=\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}$
3. 对 $\boldsymbol{\Sigma}$ 进行特征分解得到其特征值与特征向量：$\boldsymbol{\Sigma}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}$
4. 主成分变换：$\mathbf{Y}=\mathbf{U}^{\mathrm{T}}\mathbf{X}$

以上的主成分分析过程采用的是数据的协方差矩阵，而在实际的应用中，为了消除量纲的影响，可以将协方差矩阵替换为相关系数矩阵. 下面给出一个主成分分析的简单例子.

例 2.1 假设观测数据为

$$\mathbf{X}=\begin{bmatrix}1 & 2 & 3\\ 1 & 1 & 3\end{bmatrix},$$

试对该数据进行主成分分析.

解　显然该数据可以视为二维平面上的三个点（即 $p=2, n=3$），它们的坐标分别为 $(1,1)$，$(2,1)$ 和 $(3,3)$.

1. 首先得到数据的均值向量为

$$\boldsymbol{\mu}=\frac{1}{3}\mathbf{X}\mathbf{1}=\frac{1}{3}\begin{bmatrix}1 & 2 & 3\\ 1 & 1 & 3\end{bmatrix}\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}=\begin{bmatrix}2\\ \frac{5}{3}\end{bmatrix},$$

然后将数据中心化，即

$$\mathbf{X} = \mathbf{X} - \boldsymbol{\mu}\mathbf{1}^{\mathrm{T}} = \begin{bmatrix} -1 & 0 & 1 \\ -\frac{2}{3} & -\frac{2}{3} & \frac{4}{3} \end{bmatrix},$$

这里 $\mathbf{1} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^{\mathrm{T}}$ 是一个分量全为 1 的列向量.

2. 计算协方差矩阵

$$\boldsymbol{\Sigma} = \frac{1}{3}\mathbf{X}\mathbf{X}^{\mathrm{T}} = \frac{1}{3}\begin{bmatrix} -1 & 0 & 1 \\ -\frac{2}{3} & -\frac{2}{3} & \frac{4}{3} \end{bmatrix}\begin{bmatrix} -1 & -\frac{2}{3} \\ 0 & -\frac{2}{3} \\ 1 & \frac{4}{3} \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 6 & 6 \\ 6 & 8 \end{bmatrix}$$

3. 对协方差矩阵进行特征分解得 $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}$，其中特征值和特征向量矩阵分别为

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 1.4536 & 0 \\ 0 & 0.1019 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{bmatrix} = \begin{bmatrix} 0.6464 & -0.7630 \\ 0.7630 & 0.6464 \end{bmatrix}.$$

4. 主成分变换

$$\mathbf{Y} = \mathbf{U}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} -1.1551 & -0.5087 & 1.6637 \\ 0.3321 & -0.4309 & 0.0988 \end{bmatrix},$$

其中 $Y_1 = \boldsymbol{u}_1^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} -1.1551 & -0.5087 & 1.6637 \end{bmatrix}$ 为第一主成分，它对应着数据 $\mathbf{X}$ 在 $\boldsymbol{u}_1$ 方向的投影. $Y_2 = \boldsymbol{u}_2^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} 0.3321 & -0.4309 & 0.0988 \end{bmatrix}$ 为第二主成分，它对应着数据 $\mathbf{X}$ 在 $\boldsymbol{u}_2$ 方向的投影.

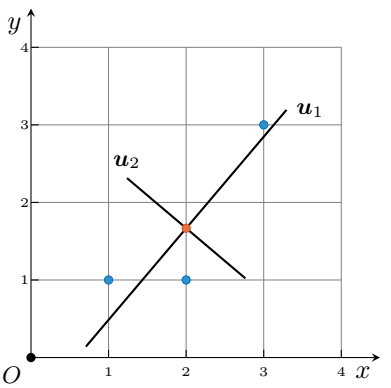


图 2.7 三个散点 $(1,1)$，$(2,1)$ 和 $(3,3)$ 的主成分变换. 第一主成分和第二主成分方向分别为 $\begin{bmatrix} 0.6464 & 0.7630 \end{bmatrix}^{\mathrm{T}}$，$\begin{bmatrix} -0.7630 & 0.6464 \end{bmatrix}^{\mathrm{T}}$. 数据在这两个方向的方差分别为 $\lambda_1 = 1.4536$，$\lambda_2 = 0.1019$

从上面的计算可知，数据在第一主成分方向上的方差为 $\lambda_1 = \boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u}_1 = 1.4536$，而在第二主成分方向的方差为 $\lambda_2 = \boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u}_2 = 0.1019$. 第一个成分的贡献率达到了

$\eta(1)=\frac{\lambda_1}{\lambda_1+\lambda_2}=\frac{1.4536}{1.4536+0.1019}=0.9345$，即数据的主要信息量分布在第一主成分.

2.4 主成分分析的几何解释

根据前面讲的主成分分析的基本原理和步骤，我们可以对任意给定的数据进行主成分分析. 以图 2.8 中二维平面上的散点为例，一旦给定一组散点，总是可以找到这组数据的最大主成分方向和最小主成分方向. 并且，这两个主成分方向可能会随着散点数量和散点分布形状的变化而变化. 然而，有趣的是，无论这两个主成分方向如何变化，它们之间始终存在一个不变的关系，即两个主成分方向必然互相垂直！这个现象固然可以通过协方差矩阵的基本性质给出代数解释（即，任意实对称矩阵的特征向量必然正交），但在其背后，是否隐藏着更为深刻的几何机制呢？

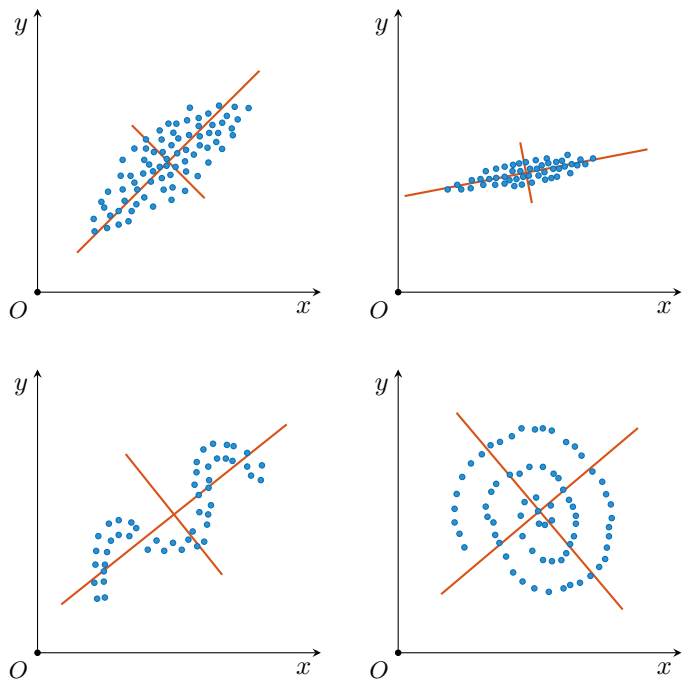


图 2.8　不同散点分布下最大主成分方向和最小主成分方向直观示意图

对于具有 p 个特征 n 个观测的 $p\times n$ 大小的数据矩阵 $\mathbf{X}$，我们一般把 $\mathbf{X}$ 的列向量所在的线性空间叫特征空间，$\mathbf{X}$ 的行向量所在的线性空间叫样本空间（第一章已经使用了样本空间的概念）. 这样，矩阵 $\mathbf{X}$ 就可以认为由 p 维特征空间的 n 个（列）向量构成，即

$$\mathbf{X}=\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n\end{bmatrix}. \tag{2.35}$$

也可以认为 $\mathbf{X}$ 由 n 维样本空间的 p 个（行）向量构成，即

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}. \tag{2.36}$$

由于主成分分析模型的目标函数为 $\mathrm{var}(\mathbf{u}^{\mathrm{T}}\mathbf{X})$，为了探寻主成分分析背后的几何机制，我们有必要分析一下该目标函数中的 $\mathbf{u}^{\mathrm{T}}\mathbf{X}$ 这一项. 对于任意一个 p 维单位向量 $\boldsymbol{u} = \begin{bmatrix} u_1 & u_2 & \cdots & u_p \end{bmatrix}^{\mathrm{T}}$，根据 (2.35) 和 (2.36) 这两种 $\mathbf{X}$ 的不同的分块，$Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X}$ 也相应地有两种解释. 首先，可以把 Y 看作 $\mathbf{X}$ 的 n 个列向量在 $\boldsymbol{u}$ 上的投影所得到的 n 个数值的集合，即

$$Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}_2 & \cdots & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}_n \end{bmatrix}. \tag{2.37}$$

其次，也可以把 Y 看作以 $\boldsymbol{u}$ 的 p 个分量作为权重的 $\mathbf{X}$ 的 p 个行向量的线性组合，即

$$Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X} = u_1X_1 + u_2X_2 + \cdots + u_pX_p. \tag{2.38}$$

基于 (2.37)，我们更倾向于把 Y 当作一维空间的 n 个点的集合；而基于 (2.38)，由于 $X_1, X_2, \cdots, X_p$ 均为 n 维空间的点或向量，因此他们的线性组合 Y 也更倾向于被认为是 n 维空间的一个点或向量.

仍以上节简单例子中的数据为例（直接使用中心化后的数据），即

$$\mathbf{X} = \begin{bmatrix} -1 & 0 & 1 \\ -\frac{2}{3} & -\frac{2}{3} & \frac{4}{3} \end{bmatrix}.$$

显然，如图 2.9 所示，它对应二维平面上的三个蓝色的点

$$\boldsymbol{x}_1 = \begin{bmatrix} -1 \\ -\frac{2}{3} \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 0 \\ -\frac{2}{3} \end{bmatrix}, \quad \boldsymbol{x}_3 = \begin{bmatrix} 1 \\ \frac{4}{3} \end{bmatrix}.$$

同时，如图 2.10 所示，它也对应三维空间上的两个蓝色的点

$$X_1 = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}, \quad X_2 = \begin{bmatrix} -\frac{2}{3} & -\frac{2}{3} & \frac{4}{3} \end{bmatrix}.$$

当选择 $\boldsymbol{u}$ 为 45° 角方向单位向量时，即 $\boldsymbol{u} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}^{\mathrm{T}}$. 此时

$$Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} -\frac{5\sqrt{2}}{6} & -\frac{\sqrt{2}}{3} & \frac{7\sqrt{2}}{6} \end{bmatrix}.$$

基于 (2.37)，对应图 2.9 中三个红色的点，它们相当于 $\mathbf{X}$ 的三个列向量在 $\boldsymbol{u}$ 方向的投影. 基于 (2.38)，$Y = \begin{bmatrix} -\frac{5\sqrt{2}}{6} & -\frac{\sqrt{2}}{3} & \frac{7\sqrt{2}}{6} \end{bmatrix}$ 对应图 2.10 中的一个红色的点，它相当于 $\mathbf{X}$ 的两个行向量以 $\boldsymbol{u}$ 的两个分量为权重的线性组合.

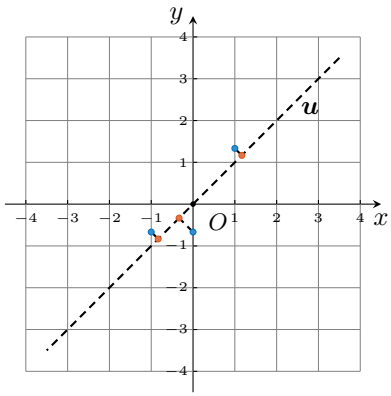


图 2.9　$Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X}$ 的特征空间解释. 它相当于 $\mathbf{X}$ 的各个列向量在 $\boldsymbol{u}$ 方向上的投影

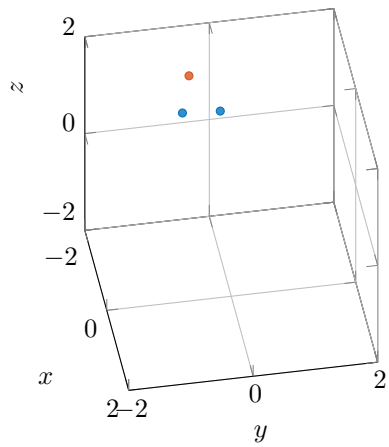


图 2.10　$Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X}$ 的样本空间解释. 它相当于 $\mathbf{X}$ 的各个行向量以 $\boldsymbol{u}$ 的各个分量为权重的线性组合

接下来我们再来分析一下主成分分析模型的目标函数:

$$\operatorname{var}(Y) = \operatorname{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u} = \frac{1}{n}\boldsymbol{u}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{u} = \frac{1}{n}\left\|\boldsymbol{u}^{\mathrm{T}}\mathbf{X}\right\|^2 = \frac{1}{n}\|Y\|^2. \tag{2.39}$$

注意到，(2.39) 中左边这一项 $\operatorname{var}(Y)$ 代表数据 $\mathbf{X}$ 在 $\boldsymbol{u}$ 方向的方差，它是 p 维空间中数据在 $\boldsymbol{u}$ 方向的统计量；而 (2.39) 中右边这一项中的 $\|Y\|^2$ 则代表 n 维样本空间中向量 Y 的几何量（长度的平方）. 也就是说，主成分分析的目标函数既可以当作 p 维特征空间的统计量，也可以看作 n 维样本空间的几何量. 那么读者不免要问，从这两个角度看主成分分析有什么区别么？或者说从样本空间看主成分分析有什么优势么？

我们知道，几何量是一个相对于统计量更为直观的量. 对于给定的散点，如果不经过计算的话，在特征空间中我们很难直接判断出数据的各个主成分方向. 如果不利

用协方差矩阵的代数性质的话，也很难直接判断出这些主成分之间的相互关系. 而在样本空间，特征空间中数据 $\mathbf{X}$ 在任意方向 $\boldsymbol{u}$ 的投影 $Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X}$ 的方差 $\mathrm{var}(Y)$ 对应样本空间的几何量 $\|Y\|^2$，这似乎提供了一条从样本空间探寻数据主成分的途径. 问题的关键在于对于 $\mathbf{X}$ 的行向量的所有可能的线性组合 $Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X}$（其中 $\boldsymbol{u}$ 为单位向量），它们在样本空间中到底具有什么结构，或者它们到底构成什么图形？

带着上面这个问题，我们遍历所有的二维单位向量 $\boldsymbol{u}$，并用它们的分量对上面例子中的 $\mathbf{X}$ 的两个行向量 $X_1 = \begin{bmatrix}-1 & 0 & 1\end{bmatrix}$ 和 $X_2 = \begin{bmatrix}-\frac{2}{3} & -\frac{2}{3} & \frac{4}{3}\end{bmatrix}$ 进行线性组合，这样可以得到一个三维空间中的曲线，如图 2.11 所示. 从图中可以看出，这个曲线似乎是一个椭圆. 而椭圆上有两个特殊的点，即椭圆长轴端点和椭圆短轴端点. 显然，椭圆长轴端点是椭圆上所有点中距离原点（椭圆中心）最远的点，而椭圆短轴端点是椭圆上所有点中距离原点最近的点.

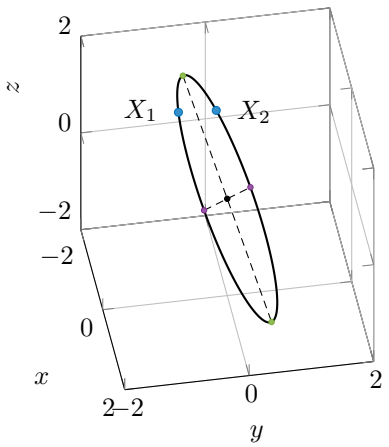


图 2.11　遍历二维单位向量对 $X_1 = \begin{bmatrix}-1 & 0 & 1\end{bmatrix}$ 和 $X_2 = \begin{bmatrix}-\frac{2}{3} & -\frac{2}{3} & \frac{4}{3}\end{bmatrix}$ 线性组合得到的点的集合构成样本空间中的一个椭圆，其长短轴端点（绿色点和紫色点）分别对应了数据的最大主成分和最小主成分

根据 (2.39) 中特征空间的方差极值和样本空间的距离极值的等价关系，长轴端点必然对应特征空间的最大方差方向，也就是说椭圆长轴端点就是数据的第一主成分！同理，椭圆短轴端点也必然是数据的最小主成分或第二主成分. 此外，由于椭圆的长短轴必然垂直，因此数据的各个主成分方向也必然垂直（请读者思考）. 需要强调的是，这个椭圆完全由给定的数据确定. 一旦数据给定了，这个椭圆在样本空间中的位置和性质也就确定了.

上面的例子展示了数据在具有两个特征三个观测的情况下，其在二维特征空间的方差极值方向等价于该数据在三维样本空间中由其确定的椭圆的长短轴端点. 一般情况下，对于 p 个特征 n 个观测的数据 $\mathbf{X}$，该数据在样本空间中也唯一确定一个 p 维超

椭球面，其上的点都由 $\mathbf{X}$ 的行向量线性组合而成，且组合系数向量为任意单位向量. 一旦数据给定了，这个超椭球面在 n 维样本空间中的位置和形状也就完全确定了. 此外，与上面的简单例子类似，数据 $\mathbf{X}$ 在 p 维特征空间的方差极值方向等价于该数据在 n 维样本空间中由其确定的超椭球面的各个轴的端点.

上面的结论都建立在一个基本的假设或者猜想之上，即用任意单位向量对原始数据 $\mathbf{X}$ 的行向量进行线性组合，得到的散点 $\boldsymbol{u}^{\mathrm{T}}\mathbf{X}$ 都位于 n 维样本空间的 p 维超椭球面上. 这个猜想是否成立呢？我们下面将给出严格的证明.

定理 2.1 对于具有 p 个特征 n 个观测的行满秩数据矩阵，用任意 p 维单位向量 $\boldsymbol{u}$ 对 $\mathbf{X}$ 的各个行向量进行线性组合得到的散点 $\boldsymbol{u}^{\mathrm{T}}\mathbf{X}$ 总位于 n 维样本空间中的一个 p 维超椭球面上.

证明 用任意一个单位向量 $\boldsymbol{u}$ 对 $\mathbf{X}$ 的各个行向量进行线性组合得到 $Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X}$, 显然 Y 是一个 n 维行向量. 记 $\boldsymbol{y} = Y^{\mathrm{T}}$，则 $\boldsymbol{y}$ 为 n 维列向量，且有

$$\mathbf{X}^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{y}. \tag{2.40}$$

由于 $\boldsymbol{y}$ 是 $\mathbf{X}^{\mathrm{T}}$ 的列向量的线性组合，即 $\boldsymbol{y}$ 位于 $\mathbf{X}^{\mathrm{T}}$ 的列空间. 因此，利用最小二乘法，可以得到 $\boldsymbol{u}$ 的精确解为

$$\boldsymbol{u} = (\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X}\boldsymbol{y}. \tag{2.41}$$

因为 $\boldsymbol{u}$ 为单位向量，即满足 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1$，因此 (2.41) 必然满足

$$\boldsymbol{y}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-2}\mathbf{X}\boldsymbol{y} = 1, \tag{2.42}$$

即，$\boldsymbol{y}$ 是满足下述方程的解

$$f(\boldsymbol{y}) = \boldsymbol{y}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-2}\mathbf{X}\boldsymbol{y} - 1 = 0. \tag{2.43}$$

显然 $f(\boldsymbol{y}) = 0$ 为 n 维空间中的二次超曲面. 鉴于 $\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-2}\mathbf{X}$ 是非负定实对称矩阵，根据解析几何中曲面类型的判断准则，可知，$f(\boldsymbol{y}) = 0$ 为 n 维空间中的一个 p 维超椭球面. ■

总结而言，类似于最小二乘法的几何解释，样本空间为主成分分析的理解也提供了一个更加清晰、直观的视角.

例 2.2 平面上在正方形内均匀分布的二维散点（图 2.12 ），请给出该数据的方差极值方向.

解 首先，在不经过计算的情况下，似乎很难直观判断出该数据在各个方向方差的大小，因此，似乎也很难直接判断出该数据的方差极值方向. 但是，根据定理 2.1，我们知道该数据可以确定一个样本空间的椭圆，由于椭圆的长短轴必然垂直，因此该数据如果存在最大和最小方差方向时，这两个方向也必然垂直. 但是由于该数据的特殊性，二维特征平面上任意两个垂直方向的方差必然相等. 因此该数据在各个方向的方差也

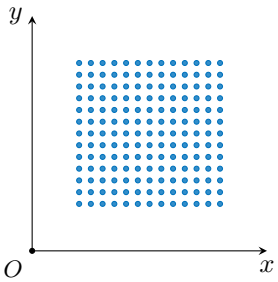


图 2.12　平面上以正方形形状均匀分布的散点

必然都相等. 也就是说，该数据所确定的样本空间的椭圆其实是圆.

2.5 主成分分析的子空间逼近解释

本节从子空间逼近的角度对 PCA 进行解释. 对于具有 p 个特征 n 个观测的数据，对其进行子空间逼近的含义就是我们希望找到一个能够尽量表征观测数据 $\mathbf{X}$ 的所有信息的低维子空间. 为了方便起见，仍然假设 $\mathbf{X}$ 的均值向量为零向量.

不妨假设需要寻找的低维子空间的维度为 $s(s<p)$，且 $\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_s$ 为该子空间的一组标准正交基. 记

$$\mathbf{Q}=\begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_s\end{bmatrix},$$

显然矩阵 $\mathbf{Q}$ 为列正交矩阵. 可以定义矩阵 $\mathbf{Q}$ 的正交投影矩阵为

$$\mathbf{P}_{\mathbf{Q}}=\mathbf{Q}\mathbf{Q}^{\dagger}=\mathbf{Q}(\mathbf{Q}^{\mathrm{T}}\mathbf{Q})^{-1}\mathbf{Q}^{\mathrm{T}}=\mathbf{Q}\mathbf{Q}^{\mathrm{T}}, \tag{2.44}$$

将 $\mathbf{X}$ 投影到 $\mathbf{Q}$ 的列空间得

$$\mathbf{Y}=\mathbf{P}_{\mathbf{Q}}\mathbf{X}=\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{X}. \tag{2.45}$$

如果矩阵 $\mathbf{Q}$ 的列空间是 $\mathbf{X}$ 的最佳逼近子空间，则必然满足 $\mathbf{Y}=\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{X}$ 与 $\mathbf{X}$ 尽量接近. 我们不妨用它们之差的 F-范数的平方来衡量它们之间的误差，于是可以建立如下关于 $\mathbf{Q}$ 的优化模型

$$\begin{cases}\min\limits_{\mathbf{Q}}\|(\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}})\mathbf{X}\|_{\mathrm{F}}^2\\ \text{s.t. }\mathbf{Q}^{\mathrm{T}}\mathbf{Q}=\mathbf{I}_s\end{cases}, \tag{2.46}$$

其中 $\mathbf{I}_p,\mathbf{I}_s$ 分别是 p 阶和 s 阶单位矩阵. 事实上，$\mathbf{P}_{\mathbf{Q}}^{\perp}=\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}}$ 是矩阵 $\mathbf{Q}$ 的正交补投影矩阵，因此模型 (2.46) 也可以解读为，寻求矩阵 $\mathbf{Q}$，使得 $\mathbf{X}$ 在 $\mathbf{Q}$ 的列空间的正

交补空间的投影分量尽量小. 又由于

$$\begin{aligned}\left\|(\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}})\mathbf{X}\right\|_{\mathrm{F}}^2&=\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\left(\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right)\left(\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right)\mathbf{X}\right)\\&=\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\left(\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right)\mathbf{X}\right)\\&=\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)-\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{X}\right),\end{aligned}\tag{2.47}$$

因此模型 (2.46) 中目标函数的最小化，等价于 $\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{X}\right)$ 的最大化. 进一步地，因为

$$\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{X}\right)=\mathrm{tr}\left(\mathbf{Q}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{Q}\right),$$

所以 (2.46) 可以转化为如下优化问题

$$\begin{cases}\max\limits_{\mathbf{Q}}\ \mathrm{tr}\left(\mathbf{Q}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{Q}\right)\\ \text{s.t. }\mathbf{Q}^{\mathrm{T}}\mathbf{Q}=\mathbf{I}_s\end{cases}.\tag{2.48}$$

而 (2.48) 的求解可以归结为矩阵 $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 的特征值与特征向量问题（请读者思考并验证）. 假设 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_s$ 为 $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 的前 s 个最大的特征值，$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_s$ 为相应的特征向量，则 $\mathbf{Q}=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_s\end{bmatrix}$ 为 (2.48) 的解，而由 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_s$ 这 s 个向量张成的 s 维子空间为待求的能够最大程度表征 $\mathbf{X}$ 所包含信息的子空间.

需要注意的是，尽管待求的子空间是唯一确定的，但是该空间的基可以有无穷多选择. 因此，理论上 (2.48) 有无穷多解. 比如对于任意一个 $s\times s$ 的正交矩阵 $\mathbf{Q}_1$，可以验证 $\mathbf{Q}\mathbf{Q}_1$ 也是 (2.48) 的解.

记 $\mathbf{X}=\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n\end{bmatrix}$，$\mathbf{Y}=\begin{bmatrix}\boldsymbol{y}_1 & \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_n\end{bmatrix}$，根据 (2.45)，$\boldsymbol{y}_i(i=1,2,\cdots,n)$ 可以认为是 $\boldsymbol{x}_i(i=1,2,\cdots,n)$ 在 $\mathbf{Q}$ 的列空间的投影. 而 (2.46) 的目标函数则相当于所有这 n 个点的距离的平方和，即

$$\left\|(\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}})\mathbf{X}\right\|_{\mathrm{F}}^2=\sum_{i=1}^{n}\left\|\boldsymbol{x}_i-\boldsymbol{y}_i\right\|^2=\sum_{i=1}^{n}d_i^2.\tag{2.49}$$

因此，模型 (2.46) 的子空间逼近模型也可以解读为：寻找一个特征空间的超平面[5]，使得观测数据的各个点到超平面的距离平方和最小. 通俗而言，这相当于寻找一个距离所有观测点都尽可能近的低维超平面，也可以理解为用一个超平面去拟合给定的观测点.

当观测数据的特征数为 $p=2$ 且待逼近的子空间维度为 $s=1$ 时，(2.46) 就相当于寻找一个距离二维平面上各个观测点最近的直线，而这正对应着直线拟合的总体最小二乘（图 2.13 ）.

例 2.3 试用主成分分析对平面上的三个点 $(1,1)$，$(2,1)$ 和 $(3,3)$ 进行直线拟合.

解 根据例 2.1 可知，这三个点的均值向量为 $\boldsymbol{\mu}=\begin{bmatrix}2 & \frac{5}{3}\end{bmatrix}^{\mathrm{T}}$，第一主成分方向为

[5]在本书中，超平面泛指线性空间的子空间.

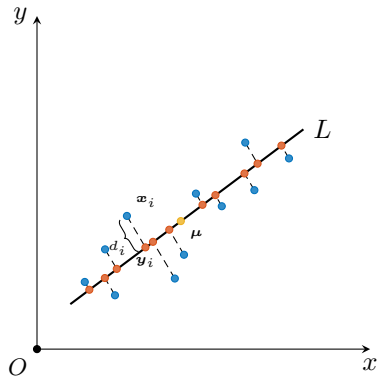


图 2.13　主成分分析的子空间逼近示意图. 当观测样本为二维平面上的数据，待逼近的子空间的维数为 1 时，主成分分析等价于直线拟合的总体最小二乘

$\boldsymbol{u}_1 = \begin{bmatrix} 0.6464 & 0.7630 \end{bmatrix}^{\mathrm{T}}$. 因此，这三个散点的一维最佳逼近子空间是一个经过点 $\boldsymbol{\mu} = \begin{bmatrix} 2 & \frac{5}{3} \end{bmatrix}^{\mathrm{T}}$，方向为 $\boldsymbol{u}_1 = \begin{bmatrix} 0.6464 & 0.7630 \end{bmatrix}^{\mathrm{T}}$ 的直线. 对应的直线方程经过整理为

$$0.7630x - 0.6464y - 0.4487 = 0. \tag{2.50}$$

为了便于比较，我们将 (2.50) 所对应的直线与基于 $y = ax + b$，$x = a'y + b'$ 所拟合的直线绘于同一平面内（如图 2.14 ），可以看出它们为三条不同的直线. 在实际应用中要根据实际情况选取合适的直线方程对给定的散点进行拟合.

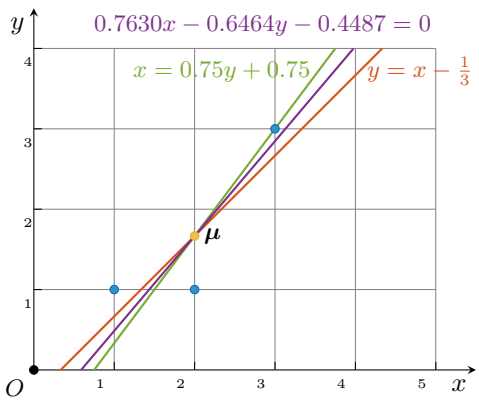


图 2.14　直线拟合的三种情形：$y = ax + b$，$x = a'y + b'$ 和 $ax + by + c = 0$

2.6 主成分分析的概率解释

对于给定的 p 个特征 n 个观测的数据 $\mathbf{X}$，仍假设需要寻找的低维子空间的维度为 $s(s<p)$，且仍然记 $\mathbf{Q}=\begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_s\end{bmatrix}$，其中 $\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_s$ 为该子空间的任意一组标准正交基. 假设 $\mathbf{X}$ 服从多元正态分布，那么，它在 $\mathbf{Q}$ 的列空间的正交补空间的投影 $(\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}})\mathbf{X}$ 也服从多元正态分布. 记

$$\mathbf{E}=(\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}})\mathbf{X}, \tag{2.51}$$

可以认为 $\mathbf{E}$ 是 $\mathbf{X}$ 的子空间逼近的误差项. 在 (2.46) 的目标函数中，用 $\mathbf{E}$ 的 F-范数的平方来衡量子空间逼近的误差的大小. 下面就为什么选择这一指标给出概率上的解读.

不妨假设 $\mathbf{E}$ 中的各个元素 $e_{ij}(1\leqslant i\leqslant p,1\leqslant j\leqslant n)$ 独立且服从均值为 0、方差为 σ^2 的高斯分布，即 $e_{ij}\sim N(0,\sigma^2)$. 对于所有的 e_{ij}，其概率密度函数都具有如下形式

$$f(e_{ij})=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{e_{ij}^2}{2\sigma^2}}. \tag{2.52}$$

因此，在给定矩阵参数 $\mathbf{Q}$ 的情况下，子空间逼近的误差的联合概率密度函数为

$$f(\mathbf{E}|\mathbf{Q})=\prod_{i,j}f(e_{ij}), \tag{2.53}$$

继而，可以构建关于矩阵参数 $\mathbf{Q}$ 的对数似然函数为

$$l(\mathbf{Q})=\ln f(\mathbf{E}|\mathbf{Q})=\ln\prod_{i,j}f(e_{ij})=\sum_{i,j}\ln\frac{1}{\sqrt{2\pi}\sigma}-\frac{1}{2\sigma^2}\sum_{i,j}e_{ij}^2. \tag{2.54}$$

显然，$l(\mathbf{Q})$ 的最大化等价于 $\sum_{i,j}e_{ij}^2$ 的最小化，而

$$\sum_{i,j}e_{ij}^2=\|\mathbf{E}\|_{\mathrm{F}}^2=\left\|(\mathbf{I}_p-\mathbf{Q}\mathbf{Q}^{\mathrm{T}})\mathbf{X}\right\|_{\mathrm{F}}^2, \tag{2.55}$$

因此，这就从概率上解释了 (2.46) 中为什么用 F-范数而不是别的指标来衡量误差的大小. 同时，这也一定程度说明了，只有当观测数据服从高斯分布时，主成分分析才是最佳的降维手段. 当数据分布不满足高斯分布时，(2.51) 中的模型误差项一般也不再服从高斯分布，后面的推导也就无从谈起. 此时，(2.46) 中的 F-范数失去了概率基础，它通常也不再是衡量模型误差的最佳选择.

2.7 主成分分析的信息论解释

信息是个很抽象的概念. 人们常常说信息很多，或者信息较少，但却很难具体地说出信息到底有多少. 信息论之父香农首先给出了信息熵的定义并用它来衡量信息量的大小.

定义 2.10 设 X 是一个离散型随机变量，其可能的取值为 $x_1,x_2,\cdots,x_n$，相应的概率

分别为 $P(x_1), P(x_2), \cdots, P(x_n)$，则 X 的**熵**为

$$H(X) = -\sum_{i=1}^{n} P(x_i) \ln P(x_i). \tag{2.56}$$

定义 2.11 设 X 是一个概率密度函数为 $f(x)$ 的随机变量，则 X 的熵（称为**微分熵**）定义为

$$H(X) = -\int_{-\infty}^{\infty} f(x) \ln f(x) \mathrm{d}x. \tag{2.57}$$

例 2.4 试计算高斯分布的信源的微分熵.

解 假设 X 是一个服从高斯分布的随机变量，且其概率密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

其中 μ, σ^2 分别为随机变量的均值和方差. 根据 (2.57)，X 的微分熵为

$$\begin{aligned} H(X) &= -\int_{-\infty}^{\infty} f(x) \ln f(x) \mathrm{d}x = -\int_{-\infty}^{\infty} f(x) \ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) \mathrm{d}x \\ &= -\int_{-\infty}^{\infty} f(x) \ln \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{d}x - \int_{-\infty}^{\infty} f(x) \ln e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x. \end{aligned}$$

因为

$$\int_{-\infty}^{\infty} f(x) \mathrm{d}x = 1,$$

所以

$$\begin{aligned} H(X) &= \frac{1}{2} \ln \left(2\pi\sigma^2\right) - \int_{-\infty}^{\infty} f(x) \ln e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \\ &= \frac{1}{2} \ln \left(2\pi\sigma^2\right) + \int_{-\infty}^{\infty} f(x) \frac{(x-\mu)^2}{2\sigma^2} \mathrm{d}x. \end{aligned}$$

又因为

$$\int_{-\infty}^{\infty} (x-\mu)^2 f(x) \mathrm{d}x = \sigma^2,$$

所以，高斯随机变量的微分熵为

$$H(X) = \frac{1}{2} \ln \left(2\pi\sigma^2\right) + \frac{1}{2} = \frac{1}{2} \ln \left(2\pi e\sigma^2\right). \tag{2.58}$$

从 (2.58) 可以看出，高斯分布的信源的微分熵只与方差正相关. 这也从另一个角度解释了主成分分析中用方差为指标来衡量数据信息量大小的合理性.

2.8 主成分分析在应用中的问题

尽管主成分分析是最常用的数据降维方法，但在应用中也有诸多问题需要注意. 接下来，将针对非高斯、量纲、维数、噪声等几个在主成分分析应用中经常遇到的问

题展开探讨.

2.8.1 非高斯问题

当观测数据服从高斯分布时，主成分分析是最佳的降维或特征提取手段. 在实际应用中，由于各种因素的影响，观测数据的高斯性假设往往很难得到保证. 此外，一些应用寻求的是数据的非高斯性比较强的方向，比如经典的鸡尾酒会问题. 在图 2.15 的场景中，分别在多个位置放置了多个麦克风，这些麦克风可以接收到场景中所有人的混合声音. 当麦克风数和人数都为三时，对应的混合模型为

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t), \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t), \\ x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t). \end{aligned}$$

其中 $s_1(t), s_2(t), s_3(t)$ 分别为三个人单独的语音时间序列，$x_1(t), x_2(t), x_3(t)$ 为三个麦克风接收到的语音时间序列，系数矩阵 $\mathbf{A}$ 则为相应的混合系数矩阵. 鸡尾酒会问题指的就是如何从这些麦克风收到的混合声音 $x_1(t), x_2(t), x_3(t)$ 中还原出场景中各个人的声音 $s_1(t), s_2(t), s_3(t)$. 独立成分分析是处理这类问题的常用手段，常用的方法有 FastICA [5]、JADE [6]等. 第 3 章将要介绍的主偏度分析也可以用于处理此类问题.

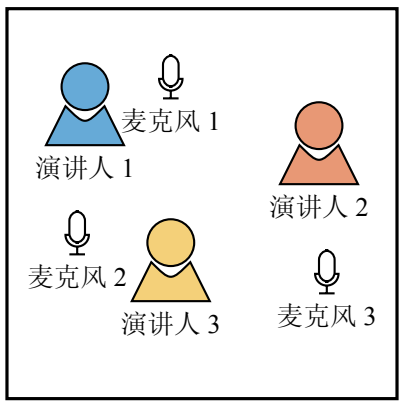


图 2.15　鸡尾酒会问题

2.8.2 量纲问题

在一些应用中，观测数据的不同特征往往具有不同的量纲. 比如，当分别用一把英寸刻度的尺子和一把厘米刻度的尺子同时测量某一物体的高度时，就会得到如图 2.16 所示的散点分布. 由于 1 英寸 =2.54 厘米，对同一个测量对象，以厘米为量纲

的测量结果在数值上就会明显要更大一些. 相应地，对于一组测量对象，以厘米为量纲的测量结果的方差在数值上也会更大一些. 因此，这就会导致厘米刻度的观测对于主成分的贡献要远大于英寸刻度的观测. 为了克服这种由量纲的不同引起的各个特征对主成分的贡献不均衡的现象，可以在对数据进行主成分分析之前首先将各个特征的观测标准化，或者使用数据的相关系数矩阵对这种类型的数据进行主成分分析.

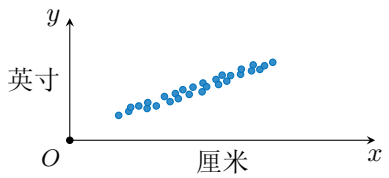


图 2.16　主成分分析中的量纲问题

当然，由于厘米刻度的测量精度一般要高于英寸刻度的测量精度，因此，以厘米为量纲的观测对主成分的贡献更大也是合理的. 从这个角度来说，对于图 2.16 中数据的主成分分析直接使用协方差矩阵也是无可厚非的.

2.8.3 维数问题

对于 p 个特征、n 个观测的数据 $\mathbf{X}$，数据的主成分分析可以归结为它的大小为 $p \times p$ 的协方差矩阵 $\boldsymbol{\Sigma}$ 的特征值与特征向量问题. 一般情况下，数据的特征数都比较小，且远小于观测数. 而在某些应用中，也会出现特征数很大且远大于观测数的情形，即 $p \gg n$. 此时，直接通过协方差矩阵的特征分析求取数据的各个主成分将具有相对较高的计算复杂度. 不妨设 $\mathbf{X}$ 的均值向量为零向量，则其协方差矩阵 $\boldsymbol{\Sigma} = \frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 的特征值与特征向量问题为

$$\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{u} = \lambda\boldsymbol{u}. \tag{2.59}$$

当 p 非常大时，(2.59) 的求解具有很高的计算复杂度. 不妨在公式两边同时左乘 $\mathbf{X}^{\mathrm{T}}$

$$\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{u} = \lambda\mathbf{X}^{\mathrm{T}}\boldsymbol{u}. \tag{2.60}$$

从 (2.60) 可以看出，$\mathbf{X}^{\mathrm{T}}\boldsymbol{u}$ 是 $\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 的特征向量. 由于 $p \gg n$，因此，相对于 $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}$，$\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 的特征值与特征向量的计算具有更低的计算复杂度. 但是，我们需要求解的并不是 $\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 的特征向量，而是 $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 的特征向量. 为此，我们只需要在 (2.60) 两边同时再左乘 $\mathbf{X}$，则有

$$\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{u} = \lambda\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{u}. \tag{2.61}$$

从 (2.61) 可以看出 $\mathbf{X}\mathbf{X}^{\mathrm{T}}\boldsymbol{u}$ 正是 $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 的特征向量.

因此，当 $p \gg n$ 时，为了求解观测数据 $\mathbf{X}$ 的协方差矩阵 $\boldsymbol{\Sigma}$ 的特征值与特征向量，

我们可以先得到 $\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 的特征向量 $\boldsymbol{v}$，再用其对 $\mathbf{X}$ 的各个列向量线性组合得到 $\mathbf{X}\boldsymbol{v}$ 即为 $\boldsymbol{\Sigma}$ 的特征向量.

2.8.4 噪声问题

在实际应用中，观测数据或多或少都会包含一定数量的噪声. 假设噪声为加性高斯噪声，则观测数据 $\mathbf{X}$ 可以分解为真实信号 $\mathbf{S}$ 与高斯噪声 $\mathbf{N}$ 两部分之和，即

$$\mathbf{X} = \mathbf{S} + \mathbf{N}.$$

当对此类含有噪声的数据进行主成分分析时，各个主成分的确定不但取决于真实信号各个方向方差的大小，而且也会受到噪声在特征空间方差分布的影响. 也就是说，主成分分析本身并没有辨别信号和噪声的能力，它只根据数据在某方向方差的大小进行主成分的判定. 当噪声在特征空间各个方向分布严重不均衡时，主成分分析的某些比较靠前的主成分很可能会包含大量的噪声. 针对这类问题，一般可以把主成分分析目标函数中的方差替换为信噪比，相应的优化模型如下

$$\begin{cases} \max\limits_{\boldsymbol{u}} \dfrac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{N}}\boldsymbol{u}} \\ \text{s.t. } \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1 \end{cases},$$

其中 $\boldsymbol{\Sigma}$ 为观测数据的协方差矩阵，$\boldsymbol{\Sigma}_{\mathbf{N}}$ 为观测数据中的噪声分量的协方差矩阵. 该模型属于典型的广义瑞利商的极值问题，这种问题的求解可以参考第 9.2 节中的相应内容.

2.9 小　　结

至此，本章的内容总结为以下 6 条：

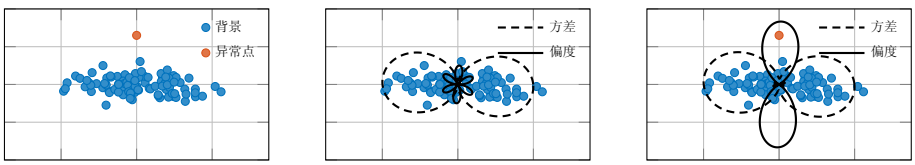
1. 当观测数据服从高斯分布时，主成分分析是最佳的降维或者特征提取手段.
2. 数据的协方差矩阵包含了数据的所有二阶统计信息，数据在任意方向的方差都可以由协方差矩阵和表征相应方向的单位向量解析表达.
3. 数据的主成分分析可以转化为数据协方差矩阵的特征值与特征向量分析.
4. 在几何上，任意的观测数据都对应一个样本空间的超椭球面，该超椭球面的各个长短轴端点对应数据的各个主成分.
5. 从子空间逼近角度，主成分分析等价于总体最小二乘法，它可以认为是用一个低维超平面拟合给定散点.
6. 从概率角度，主成分分析要求数据服从高斯分布. 且当数据服从高斯分布时，数据方差的大小等价于信息熵的大小.

第 3 章 主偏度分析

主偏度分析（Principal Skewness Analysis，PSA）是主成分分析从二阶统计到三阶统计的自然拓展. 本章首先介绍一些必备的数学概念，然后给出主偏度分析的模型及基于正交和非正交约束的求解方法. 接着探讨了主偏度分析与独立成分分析的关系，并借助于单形体首次揭示了主偏度分析的几何内涵.

3.1 问题背景

第 2 章已经表明，在数据满足高斯分布的情况下，主成分分析是最佳的降维或者特征提取手段. 然而，在实际应用中，需要处理的数据一般并不服从高斯分布，此时利用主成分分析对数据进行处理往往很难得到理想的效果. 如图 3.1a 所示的数据，红色目标点的存在会破坏数据的整体高斯性. 此时，若对这块数据进行主成分分析，红色目标点和蓝色背景点是否可以在数据的第一主成分中得到显著的区分呢？或者说，红色目标点的存在是否会显著增大数据在垂直方向的方差呢？



(a) (b) (c)

图 3.1 异常点对数据统计量的影响（为了方便展示，统计量映射图均经过了适当的缩放）(a) 包含异常点的数据 (b) 无异常点时数据的方差与偏度映射图 (c) 有异常点时数据的方差与偏度映射图

从图 3.1b 和图 3.1c 中的统计量映射图[1]可以看出，当蓝色背景点足够多时，红色异常点对数据的整体方差分布的影响并不大. 此时，对图 3.1a 中的数据进行主成分分析，在所得到的第一主成分（数据在水平方向的投影）中，红色目标点会淹没在蓝色背景之中. 因此，对于此类数据，主成分分析显然并不是理想的特征提取手段. 为了凸显红色目标点和蓝色背景点的差异，当务之急是寻找一个对类似这种游离在背景之外的孤立目标比较敏感的指标. 从图 3.1b 和 3.1c 可以看出，二阶统计量 —— 方差显然并不是一个合适的选择. 有趣的是，当我们选择目标函数为三阶统计量 —— 偏度时，情况发生了根本性的变化. 从图 3.1c 可以发现，红色目标点的存在使得数据在垂直方向

[1]统计量映射图的概念将在第 3.2.4 小节中详细介绍.

的偏度显著增加. 那么，一个更有趣的问题来了，我们是否可以发展一个类似于主成分分析的寻求数据偏度极值方向的特征提取手段呢？

3.2 基本概念

为了得到数据的偏度极值方向，我们首先给出几个基本的数学概念.

3.2.1 偏度的定义

均值和方差是两个最常用的随机变量的数字特征，接下来我们介绍随机变量的另外一个重要的数字特征 —— 偏度.

定义 3.1 给定一个随机变量 X，若 μ 为 X 的均值，σ 为 X 的标准差，则称

$$\operatorname{skew}(X)=\frac{\mathrm{E}(X-\mu)^3}{\sigma^3} \tag{3.1}$$

为随机变量 X 的偏度.

在实际应用中，往往只能得到随机变量的若干观测. 此时，我们只能得到随机变量偏度的估计值，也就是数据的三阶统计量 —— 样本偏度. 接下来，本书仍采用 $\operatorname{skew}(\boldsymbol{x})$ 来表示观测数据 $\boldsymbol{x}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}^{\mathrm{T}}$ 的样本偏度.

定义 3.2 给定随机变量 X 的 n 个观测值 $x_1,x_2,\cdots,x_n$，令 μ 为这些观测值的均值，则称

$$\operatorname{skew}(\boldsymbol{x})=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2)^{\frac{3}{2}}}, \tag{3.2}$$

为随机变量 X 的样本 $x_1,x_2,\cdots,x_n$ 的样本偏度（简称为偏度），其中 $\boldsymbol{x}$ 是由 n 个观测构成的样本列向量.

对于一个矩阵 $\mathbf{X}\in\mathbb{R}^{p\times n}$，可以将其视作 p 维随机向量的 n 个观测构成的数据. 这个数据在任意方向 $\boldsymbol{u}\in\mathbb{R}^{p\times 1}$ 上的偏度也能够根据 (3.2) 进行计算：首先得到投影数据 $Y=\boldsymbol{u}^{\mathrm{T}}\mathbf{X}$，然后根据 (3.2) 计算 Y 或 Y^{T} 的偏度即可.

相较于方差，偏度是数据分布偏斜方向和程度的一个度量，其值可以为任意实数. 如图 3.2 所示，对称分布的数据，其偏度值为零，正偏态通常意味着数据的分布向左偏斜，负偏态则反之. 因此，偏度经常被用来描述具有非对称分布的数据. 此外，鉴于

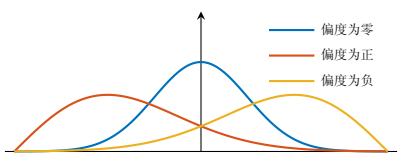


图 3.2　概率密度函数的分布与偏度的正负

远离中心的值对偏度的贡献较大，偏度还经常被用来捕捉数据中的异常点. 图 3.3 给出了直观的示例，分别展示了偏度为零、偏度为正和偏度为负这三种不同分布的简单模拟数据.

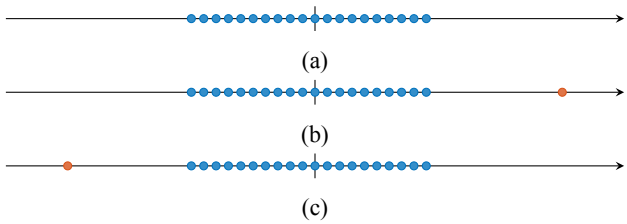


图 3.3　偏度与异常点的关系 (a) 偏度为零 (b) 偏度为正 (c) 偏度为负

3.2.2 数据白化

从公式 (3.2) 可以看出，对于给定的数据，其偏度的求取同时涉及分子和分母两项的计算. 为了简化这一过程，接下来我们引入数据的白化算子，其定义如下.

定义 3.3 给定数据 $\mathbf{X}=\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n\end{bmatrix}\in\mathbb{R}^{p\times n}$，若其均值向量和协方差矩阵为

$$\boldsymbol{\mu}=\frac{1}{n}\mathbf{X}\mathbf{1},\quad \boldsymbol{\Sigma}=\frac{1}{n}(\mathbf{X}-\boldsymbol{\mu}\mathbf{1}^{\mathrm{T}})(\mathbf{X}-\boldsymbol{\mu}\mathbf{1}^{\mathrm{T}})^{\mathrm{T}}.$$

那么，任意一个满足如下条件的矩阵 $\mathbf{W}$ 都可以称作数据 $\mathbf{X}$ 的白化算子

$$\mathbf{W}^{\mathrm{T}}\mathbf{W}=\boldsymbol{\Sigma}^{-1}. \tag{3.3}$$

而白化后的数据有如下表达式

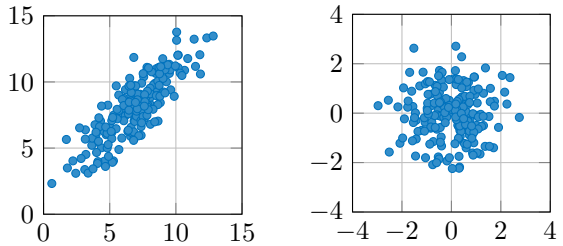
$$\hat{\mathbf{X}}=\mathbf{W}(\mathbf{X}-\boldsymbol{\mu}\mathbf{1}^{\mathrm{T}}). \tag{3.4}$$

在本书中，我们通常选择 $\mathbf{W}=\boldsymbol{\Sigma}^{-\frac{1}{2}}$ 作为白化算子. 可以验证，白化后数据 $\hat{\mathbf{X}}$ 的均值向量为零向量，协方差矩阵为单位矩阵，即

$$\hat{\boldsymbol{\mu}}=\frac{1}{n}\hat{\mathbf{X}}\mathbf{1}=\mathbf{0},\quad \hat{\boldsymbol{\Sigma}}=\frac{1}{n}\hat{\mathbf{X}}\hat{\mathbf{X}}^{\mathrm{T}}=\mathbf{I}_p, \tag{3.5}$$

这意味着白化后的数据在任意方向上的方差均为 1.

图 3.4a 中给出了一组二维数据，它的均值并不位于坐标系原点 $(0,0)$ 处，沿着不同方向的方差也明显不一样. 对这个数据进行白化，可以得到图 3.4b 中展示的结果. 特别地，对于二维平面上任意三个不共线的点，白化之后将变为正三角形的三个顶点；而三维空间上任意 4 个不共面的点，白化之后将变为正四面体的 4 个顶点. 对此，读者可自行验证.



(a)　　(b)

图 3.4　数据白化示例 (a) 原数据 (b) 白化后的数据

对于白化后的数据 $\hat{\mathbf{X}}$，其在任意方向 $\boldsymbol{u}$ 上的偏度计算公式为

$$\operatorname{skew}(\boldsymbol{u}^{\mathrm{T}}\hat{\mathbf{X}}) = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{u}^{\mathrm{T}}\hat{\mathbf{X}}\right)_i^3, \tag{3.6}$$

其中 $\left(\boldsymbol{u}^{\mathrm{T}}\hat{\mathbf{X}}\right)_i$ 表示向量 $\boldsymbol{u}^{\mathrm{T}}\hat{\mathbf{X}}$ 的第 i 个元素.

3.2.3 张量基本运算

在后文的推导中涉及到张量相关的一些运算，因此有必要对其进行简单的介绍. 其中，k 模积是最基本的张量运算之一，它可以看作是矩阵乘法的推广，其定义如下.

定义 3.4　给定张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ 与矩阵 $\mathbf{U} \in \mathbb{R}^{J \times I_k}$，两者的 k 模积操作可以表示为

$$\mathcal{A} \times_k \mathbf{U} \in \mathbb{R}^{I_1 \times \cdots \times I_{k-1} \times J \times I_{k+1} \times \cdots \times I_N}, \tag{3.7}$$

k 模积结果中的元素具有如下表达式（$a_{i_1 i_2 \cdots i_n}, u_{j i_k}$ 分别为张量 $\mathcal{A}$ 和矩阵 $\mathbf{U}$ 对应位置的元素）

$$(\mathcal{A} \times_k \mathbf{U})_{i_1 \cdots i_{k-1} j i_{k+1} \cdots i_N} = \sum_{i_k=1}^{I_k} a_{i_1 \cdots i_{k-1} i_k i_{k+1} \cdots i_N} u_{j i_k}. \tag{3.8}$$

我们知道，矩阵乘以一个向量可以看作是用该向量中的元素对矩阵的各个列向量进行线性组合. 同样地，张量与向量的 k 模积也可以看作用向量中的元素对张量沿着第 k 个维度的切片进行线性组合. 例如，对于一个三阶张量 $\mathcal{A} \in \mathbb{R}^{3\times3\times4}$，该张量 3 模积一个 1×4 大小的向量将会得到一个 3×3 的矩阵，具体操作如图 3.5 所示. 在这个例子中，线性组合的对象不再是向量，而是张量沿着第三个维度的切片，也就是矩阵.

类似地，张量与一个矩阵的 k 模积则可以看作是：使用该矩阵的每一行对张量的第 k 个维度的切片进行线性组合得到一个新的切片，然后将这些切片组合成一个新的张量. 例如，对于一个三阶张量 $\mathcal{A} \in \mathbb{R}^{3\times3\times4}$，其 3 模积一个 2×4 大小的矩阵会得到

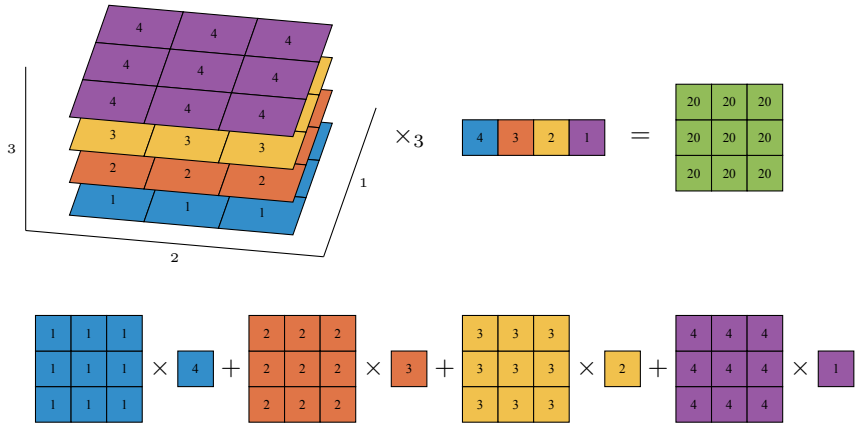


图 3.5　张量与向量 k 模积示意图

一个 $3 \times 3 \times 2$ 的张量，具体操作如图 3.6 所示.

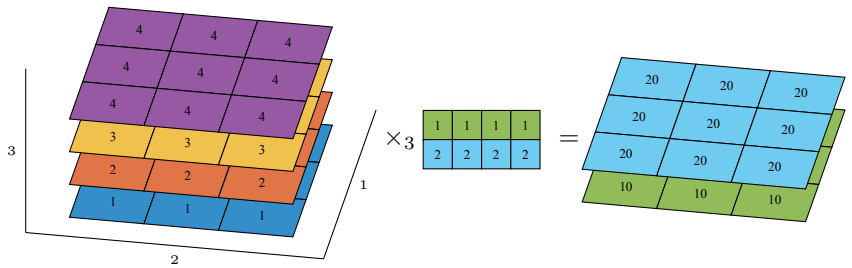


图 3.6　张量与矩阵 k 模积示意图

另一个常用的张量运算称作外积，其定义如下.

定义 3.5　给定 N 个向量，$\boldsymbol{a}_i \in \mathbb{R}^{I_i \times 1}, i = 1, 2, \cdots, N$，它们的外积为一个秩为 1 的张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$，记作

$$\mathcal{A} = \boldsymbol{a}_1 \circ \boldsymbol{a}_2 \circ \cdots \circ \boldsymbol{a}_N. \tag{3.9}$$

对于外积得到的张量，它的元素有如下表达式

$$a_{i_1 i_2 \cdots i_N} = a_1^{(i_1)} a_2^{(i_2)} \cdots a_N^{(i_N)}, \tag{3.10}$$

其中，$a_p^{(q)}$ 表示向量 $\boldsymbol{a}_p$ 的第 q 个元素.

可以发现，外积是一种特殊的 k 模积运算，即 (3.9) 可以表示为

$$\mathcal{A} = 1 \times_1 \boldsymbol{a}_1 \times_2 \boldsymbol{a}_2 \cdots \times_N \boldsymbol{a}_N. \tag{3.11}$$

特别地，当 $\boldsymbol{a}_1 = \boldsymbol{a}_2 = \cdots = \boldsymbol{a}_N = \boldsymbol{a}$（此时，$I_1 = I_2 = \cdots = I_N = I$）时，(3.9) 可以

简化为

$$\mathcal{A} = \boldsymbol{a}^{\circ N}. \tag{3.12}$$

此时，$\mathcal{A}$ 为一个 N 阶 I 维张量，其元素有如下表达式

$$a_{i_1 i_2 \cdots i_N} = a^{(i_1)} a^{(i_2)} \cdots a^{(i_N)}. \tag{3.13}$$

显然，任意交换 $i_1, i_2, i_3, \cdots, i_N$ 的顺序，都不会改变 $a_{i_1 i_2 i_3 \cdots i_N}$ 的取值，比如

$$a_{i_1 i_2 i_3 \cdots i_N} = a_{i_2 i_1 i_3 \cdots i_N},$$

因此 (3.12) 中的 $\mathcal{A}$ 为一个对称张量.

3.2.4 统计量映射图

对于一个高维数据 $\mathbf{X} \in \mathbb{R}^{p \times n}$，不妨将其在任意投影方向 $\boldsymbol{u}$ 上的 k 阶统计量记作 $s^{(k)}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})$. 如果将每个单位向量 $\boldsymbol{u}$ 都赋予一个长度，且该长度等于统计量 $s^{(k)}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})$ 的模（即绝对值），则在 p 维空间可以得到一个新的向量 $\left|s^{(k)}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\right| \boldsymbol{u}$. 所有这些新的向量（或新的点）构成的几何结构我们称之为数据 $\mathbf{X}$ 的统计量映射图. 显然，统计量映射图能够清晰地反映数据在各个方向的高阶统计分布情况. 具体来说，统计量映射图有如下的定义.

定义 3.6 *对于一个 p 维 n 个观测的数据 $\mathbf{X} \in \mathbb{R}^{p \times n}$，令 $s^{(k)}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})$ 为观测向量 $\boldsymbol{u}^{\mathrm{T}}\mathbf{X}$ 的 k 阶统计量，则如下点的集合被称作数据的 k 阶统计量映射图.*

$$\left\{ \boldsymbol{r}^{(k)} = |s^{(k)}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})| \boldsymbol{u} \,\middle|\, \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1, \boldsymbol{u} \in \mathbb{R}^{p \times 1} \right\} \tag{3.14}$$

对于一个二维标准高斯分布的数据，它在任意方向的方差均为常数 $\mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = s^{(2)}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = 1$，这意味着该数据的二阶统计量（方差）映射图为一个半径为 1 的圆

$$\left\{ \boldsymbol{r}^{(2)} = \boldsymbol{u} \,\middle|\, \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1, \boldsymbol{u} \in \mathbb{R}^{2 \times 1} \right\} \tag{3.15}$$

统计量映射图的维度与数据的维度有关，当数据分布在二维平面上时，对应的统计量映射图是二维空间中的曲线. 图 3.7 通过一个简单的示例展示了统计量映射图的构造过程. 其中所用的数据为

$$\mathbf{X} = \begin{bmatrix} -0.1722 & 0.2090 & 0.1925 & -0.2293 \\ -0.0003 & -0.4738 & 0.2318 & 0.2423 \end{bmatrix}.$$

观察可知，$\mathbf{X}$ 的均值向量为零向量. 如果选择投影方向 $\boldsymbol{u} = \begin{bmatrix} 0 & 1 \end{bmatrix}^{\mathrm{T}}$，则数据在该方向的方差为

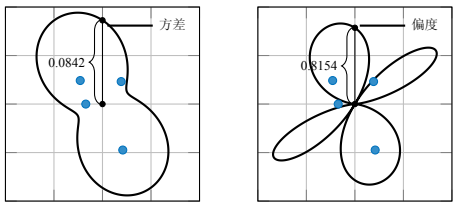
$$\mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = s^{(2)}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \frac{1}{4}(0.0003^2 + 0.4738^2 + 0.2318^2 + 0.2423^2) \approx 0.0842,$$

这意味着 $\boldsymbol{u} = \begin{bmatrix} 0 & 1 \end{bmatrix}^{\mathrm{T}}$ 对应该数据的二阶统计量（方差）映射图上的点 $(0, 0.0842)$. 同

样地，根据 (3.2)，可以计算得到数据在 $\boldsymbol{u}$ 方向的偏度为

$$\operatorname{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = s^{(3)}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \frac{\frac{1}{4}(-0.0003^3 - 0.4738^3 + 0.2318^3 + 0.2424^3)}{0.0842^{\frac{3}{2}}} \approx -0.8154,$$

这意味着 $\boldsymbol{u} = \begin{bmatrix} 0 & 1 \end{bmatrix}^{\mathrm{T}}$ 对应该数据的三阶统计量（偏度）映射图上的点 $(0, |-0.8154|) = (0, 0.8154)$.



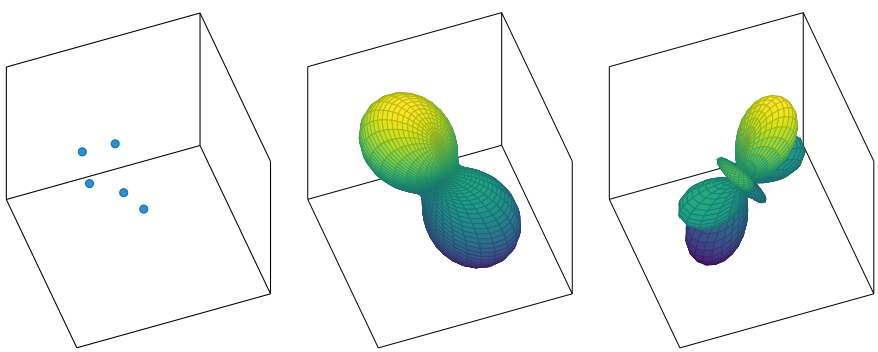
(a) (b)

图 3.7 二维数据统计量映射图（为了方便展示，统计量映射图均经过了适当的缩放）(a) 二阶统计量（方差）(b) 三阶统计量（偏度）

当数据分布在三维空间内时，比如

$$\mathbf{X} = \begin{bmatrix} 2.3459 & 0.0893 & 2.2103 & 0.7440 & 0.6762 \\ -0.4959 & 1.0007 & -1.8874 & -1.2499 & -0.2327 \\ 0.1599 & -1.0078 & 0.9440 & 1.6672 & -0.9105 \end{bmatrix},$$

其所对应的统计量映射图则是三维空间中的曲面，如图 3.8 所示. 可以发现，从统计量映射图中，我们能够直观地观察到数据在不同方向上投影结果的统计量大小.



(a) (b) (c)

图 3.8 三维数据统计量映射图 (a) 数据 (b) 二阶统计量（方差）(c) 三阶统计量（偏度）

3.3 主偏度分析

主偏度分析[7]是主成分分析从二阶统计到三阶统计的自然拓展，本节我们将首先介绍数据在任意方向偏度的解析表达，然后给出主偏度分析的优化模型与基本求解算法.

3.3.1 任意方向的偏度

在主成分分析中，对于任意 p 个特征 n 个观测的数据 $\mathbf{X}$，其在任意方向 $\boldsymbol{u}$ 的方差可以通过数据的协方差矩阵解析表达，即 $\mathrm{var}(\mathbf{X}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u}$. 那么，既然偏度是方差的三阶推广，我们不禁要问，数据在任意方向的偏度是否也有类似的简洁表达呢？接下来，为了方便起见，不妨假设数据 $\mathbf{X}$ 为白化后的数据，即其均值向量为零向量，协方差矩阵为单位矩阵. 于是，根据公式 (3.6)，数据 $\mathbf{X}$ 在任意方向 $\boldsymbol{u}$ 的偏度有

$$\begin{aligned}\mathrm{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) &= \mathrm{skew}\left(\begin{bmatrix}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}_2 & \cdots & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}_n\end{bmatrix}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}_i)^3 = \frac{1}{n}\sum_{i=1}^{n}\left((\boldsymbol{x}_i\circ\boldsymbol{x}_i\circ\boldsymbol{x}_i)\times_1\boldsymbol{u}^{\mathrm{T}}\times_2\boldsymbol{u}^{\mathrm{T}}\times_3\boldsymbol{u}^{\mathrm{T}}\right)\\ &= \left(\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\circ\boldsymbol{x}_i\circ\boldsymbol{x}_i\right)\times_1\boldsymbol{u}^{\mathrm{T}}\times_2\boldsymbol{u}^{\mathrm{T}}\times_3\boldsymbol{u}^{\mathrm{T}}\right).\end{aligned} \tag{3.16}$$

不难发现，$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\circ\boldsymbol{x}_i\circ\boldsymbol{x}_i$ 不再是一个矩阵，而是一个大小为 $p\times p\times p$ 的三阶对称张量（如图 3.9 所示），将其记作 $\mathcal{S}$，则 (3.16) 可以简化为

$$\mathrm{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \mathcal{S}\times_1\boldsymbol{u}^{\mathrm{T}}\times_2\boldsymbol{u}^{\mathrm{T}}\times_3\boldsymbol{u}^{\mathrm{T}}. \tag{3.17}$$

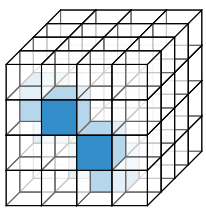


图 3.9　三阶对称张量 $\mathcal{S}$ 的示意图 ($p=4$)，蓝色方块具有相同的数值

出于传统习惯以及表达上的简洁，我们通常将 (3.17) 中 $\boldsymbol{u}^{\mathrm{T}}$ 的转置符号省略，即用下式表示数据在 $\boldsymbol{u}$ 方向的偏度

$$\mathrm{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \mathcal{S}\times_1\boldsymbol{u}\times_2\boldsymbol{u}\times_3\boldsymbol{u}. \tag{3.18}$$

事实上，数据在任意方向的方差公式 (2.24) 也可以类似于 (3.17) 重新表示为

$$\mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u} = \boldsymbol{\Sigma}\times_1\boldsymbol{u}^{\mathrm{T}}\times_2\boldsymbol{u}^{\mathrm{T}}. \tag{3.19}$$

类似于 (3.18)，我们也可以将其简化为

$$\mathrm{var}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \boldsymbol{\Sigma} \times_1 \boldsymbol{u} \times_2 \boldsymbol{u}. \tag{3.20}$$

类比主成分分析中协方差矩阵的概念，我们将 (3.17) 或 (3.18) 中的三阶张量命名为数据的协偏度张量 (Coskewness Tensor). 与协方差矩阵包含数据所有的二阶统计信息类似，协偏度张量包含了数据所有的三阶统计信息，这也正是数据在任意方向的偏度能有 (3.17) 或 (3.18) 这样解析表达式的原因所在. 对于一个 N 阶张量 $\mathcal{S} \in \mathbb{R}^{I\times I\times\cdots\times I}$，记

$$\mathcal{S}\boldsymbol{u}^m = \mathcal{S} \times_{N-m+1} \boldsymbol{u} \times_{N-m+2} \boldsymbol{u} \cdots \times_N \boldsymbol{u}, \tag{3.21}$$

其中 $m \leq N$. 容易验证，对于协偏度张量 $\mathcal{S} \in \mathbb{R}^{p\times p\times p}$，有

$$\mathcal{S}\boldsymbol{u}^3 = \mathcal{S} \times_1 \boldsymbol{u} \times_2 \boldsymbol{u} \times_3 \boldsymbol{u}, \quad \mathcal{S}\boldsymbol{u}^2 = \mathcal{S} \times_2 \boldsymbol{u} \times_3 \boldsymbol{u},$$

那么，数据在 $\boldsymbol{u}$ 方向的偏度可以表示为

$$\mathrm{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \mathcal{S}\boldsymbol{u}^3. \tag{3.22}$$

与任意方向方差的解析表达一样，公式 (3.17) 和 (3.22) 简洁优雅，为后续数据偏度极值方向的求解提供了极大便利.

3.3.2 协偏度张量的计算

在第 2 章，根据数据观测矩阵分块方式的不同，我们给出了两种协方差矩阵的计算方式. 在本章，我们仍按照观测矩阵的不同分块方式，给出两种不同的协偏度张量的计算方法. 仍假设 $\mathbf{X}$ 为白化后的数据，且其两种分块方式分别为

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \in \mathbb{R}^{p\times n}.$$

基于列向量的外积，可以给出协偏度张量的第一种计算公式为

$$\mathcal{S} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \circ \boldsymbol{x}_i \circ \boldsymbol{x}_i, \tag{3.23}$$

即，协偏度张量 $\mathcal{S}$ 为 n 个秩为 1 的张量 $\boldsymbol{x}_i \circ \boldsymbol{x}_i \circ \boldsymbol{x}_i$ 的平均，如图 3.10b 所示.

基于行向量的“内积”，可以给出协偏度张量中每一个元素的计算公式为

$$s_{ijk} = \frac{1}{n}\sum_{l=1}^{n} X_i(l)X_j(l)X_k(l). \tag{3.24}$$

其中 s_{ijk} 表示行向量 X_i, X_j, X_k 之间的协偏度，如图 3.10c 所示.

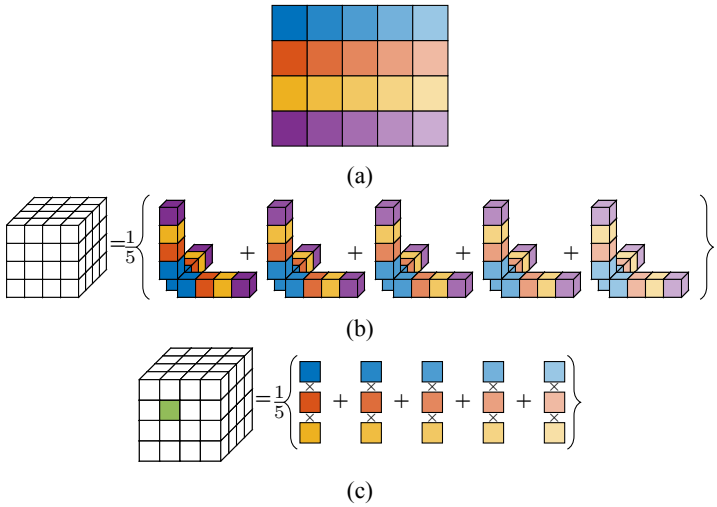


图 3.10　协偏度张量计算示意图 (a) 4×5 大小的白化数据 (b) 基于列向量外积的计算方式 (c) 基于行向量 “内积” 的计算方式来计算 s_{123}

尽管第一种分块对应的计算方式看上去更加简洁，但是在实际应用中，n 往往远远大于 p，因此在一些循环较慢的语言中（比如 MATLAB），这种计算方式的效率会非常低下. 而第二种分块对应的计算方式只需要进行 p^3 次计算，且每次计算都可以利用软件内置的向量点乘以及求平均的函数来获得 s_{ijk}，因此效率会更高. 此外，如果能充分利用统计张量的对称性，计算次数还可以进一步减少.

利用列克罗内克积算符（一种特殊的 Khatri－Rao 积[8]），我们还可以将张量的计算写成矩阵乘法的形式. 对于两个拥有相同列数的矩阵 $\mathbf{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix}$ 和 $\mathbf{B} = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_n \end{bmatrix}$，它们的列克罗内克积有如下表达式

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} \boldsymbol{a}_1 \otimes \boldsymbol{b}_1 & \boldsymbol{a}_1 \otimes \boldsymbol{b}_2 & \cdots & \boldsymbol{a}_n \otimes \boldsymbol{b}_n \end{bmatrix},$$

其中 $\otimes$ 代表克罗内克积，其定义和相关性质请参考第 3.4.1 小节. 对于数据矩阵 $\mathbf{X}$，其自身与自身的列克罗内克积为

$$\mathbf{X} * \mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1 \otimes \boldsymbol{x}_1 & \boldsymbol{x}_2 \otimes \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \otimes \boldsymbol{x}_n \end{bmatrix} \in \mathbb{R}^{p^2 \times n}.$$

可以验证，$\mathbf{X} * \mathbf{X}$ 与 $\mathbf{X}$ 的互协方差矩阵

$$\mathbf{S} = \frac{1}{n} (\mathbf{X} * \mathbf{X}) \mathbf{X}^{\mathrm{T}},$$

为一个大小为 $p^2 \times p$ 的矩阵，将其转化为一个 $p \times p \times p$ 的三阶对称张量后，其中的元素将与协偏度张量 $\mathcal{S}$ 中的元素一一对应.

事实上，$\mathbf{X} * \mathbf{X}$ 的每一列都可以看作是原数据的二次非线性项. 不妨将原数据与

其二次非线性项组合起来，得到如下的新的数据矩阵

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X} * \mathbf{X} \end{bmatrix} \in \mathbb{R}^{(p^2+p)\times n}.$$

该数据矩阵的协方差矩阵可以表示为如下的分块矩阵

$$\tilde{\mathbf{\Sigma}} = \frac{1}{n}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^{\mathrm{T}} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{S}^{\mathrm{T}} \\ \mathbf{S} & \mathbf{K} \end{bmatrix},$$

其中 $\mathbf{\Sigma}$ 为原数据 $\mathbf{X}$ 的协方差矩阵，包含了数据的全部二阶统计信息. 需要注意的是，在本小节中由于 $\mathbf{X}$ 为白化数据，因此 $\mathbf{\Sigma}$ 为单位矩阵. 而矩阵 $\mathbf{S} = \frac{1}{n}(\mathbf{X}*\mathbf{X})\mathbf{X}^{\mathrm{T}} \in \mathbb{R}^{p^2\times p}$ 对应了数据的协偏度张量 $\mathcal{S}$，因此其包含了数据的全部三阶统计信息. 至于矩阵 $\mathbf{K} = \frac{1}{n}(\mathbf{X} * \mathbf{X})(\mathbf{X} * \mathbf{X})^{\mathrm{T}} \in \mathbb{R}^{p^2\times p^2}$ 则包含了数据的全部四阶统计信息，将其转换为一个 $p\times p\times p\times p$ 的四阶对称张量后，其中的元素将与数据的四阶统计张量 $\mathcal{K} \in \mathbb{R}^{p\times p\times p\times p}$ 中的元素一一对应（读者可自行查阅四阶统计量 —— 峭度的相关定义，并进行验证）.

3.3.3 模型与求解

在得到数据在任意方向偏度的解析表达之后，我们可以给出如下的主偏度分析优化模型

$$\begin{cases} \max\limits_{\boldsymbol{u}} & \mathcal{S}\boldsymbol{u}^3 \\ \text{s.t.} & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1 \end{cases}. \tag{3.25}$$

为了得到模型的最优解，类似于主成分分析，我们仍然采用拉格朗日乘子法. 首先构建模型的拉格朗日函数为

$$\mathcal{L}(\boldsymbol{u}, \lambda) = \frac{1}{3}\mathcal{S}\boldsymbol{u}^3 + \frac{\lambda}{2}(1 - \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}). \tag{3.26}$$

(3.26) 两边同时对自变量求偏导有

$$\frac{\partial \mathcal{L}(\boldsymbol{u}, \lambda)}{\partial \boldsymbol{u}} = \mathcal{S}\boldsymbol{u}^2 - \lambda\boldsymbol{u},$$

令其等于零向量，可得

$$\mathcal{S}\boldsymbol{u}^2 = \lambda\boldsymbol{u}. \tag{3.27}$$

这意味着，数据的偏度极值方向必然满足 (3.27). 进一步地，可以发现 (3.27) 与矩阵的特征值与特征向量问题非常相似，因此我们称 (3.27) 为协偏度张量的特征值与特征向量问题. Lim 与 Qi 两位学者在 2005 年也分别从纯数学角度关注了类似的问题，并分别独立给出了张量 Z-特征对的概念[9, 10]，其定义如下.

定义 3.7（对称张量的 Z-特征对） 给定一个 m 阶 n 维的对称张量 $\mathcal{S}$，若如下式子满

足

$$\mathcal{S}\boldsymbol{u}^{m-1} = \lambda\boldsymbol{u}, \tag{3.28}$$

则称 $(\lambda, \boldsymbol{u})$ 为张量 $\mathcal{S}$ 的一个 Z-特征对. 其中，$\lambda \in \mathbb{R}$ 为 $\mathcal{S}$ 的特征值，$\boldsymbol{u} \in \mathbb{R}^{n\times 1}$ 为 λ 对应的特征向量，其满足 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1$.

显然，当 $m = 2$ 时，对称张量的 Z-特征对将会退化为矩阵特征对. 而公式 (3.27) 则可认为是公式 (3.28) 在 $m = 3$，且 $\mathcal{S}$ 为数据的协偏度张量时的特例. 可以验证，当向量 $\boldsymbol{u}$ 满足 (3.27) 时，其相应的特征值 λ 则正好为数据在这个方向的偏度值，即

$$\mathrm{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \mathcal{S}\boldsymbol{u}^3 = (\mathcal{S}\boldsymbol{u}^2)^{\mathrm{T}}\boldsymbol{u} = (\lambda\boldsymbol{u})^{\mathrm{T}}\boldsymbol{u} = \lambda. \tag{3.29}$$

我们知道，矩阵的特征值与特征向量问题已经得到了极为广泛和极其详尽的研究，相关的参考资料浩如烟海. 而遗憾的是，张量的特征对问题的研究目前尚处于初级阶段，可供参考的文献极为匮乏. 为此，我们采用比较常用的优化策略 —— 固定点迭代法来得到 (3.27) 的一个特征对. 具体步骤如下.

算法 3.1 固定点迭代法求解第一个特征对

1. 随机初始化向量 $\boldsymbol{u}$
2. 令 $\boldsymbol{u} \leftarrow \mathcal{S}\boldsymbol{u}^2$
3. 向量归一化：$\boldsymbol{u} \leftarrow \boldsymbol{u}/\|\boldsymbol{u}\|$
4. 重复步骤 2 和 3，直至收敛

接下来的问题是如何求解 (3.27) 的第二个特征对以及余下的所有特征对. 为此，我们这里采取正交约束的策略. 假设已获得 l 个特征向量 $\mathbf{U}_l = \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_l\end{bmatrix}$，对于第 $l+1$ 个特征向量 $\boldsymbol{u}_{l+1}$，为了防止其收敛到前 l 个特征向量，我们可以将 $\boldsymbol{u}_{l+1}$ 的搜索范围限制在前 l 个特征向量张成的空间的正交补空间. 相应地，求解第 $l+1$ 个特征向量的算法为：

算法 3.2 固定点迭代法求解第 $l+1$ 个特征对

1. 随机初始化向量 $\boldsymbol{u}_{l+1}$
2. 令 $\boldsymbol{u}_{l+1} \leftarrow \mathcal{S}\boldsymbol{u}_{l+1}^2$
3. 将 $\boldsymbol{u}_{l+1}$ 投影到 $\mathbf{U}_l$ 的列空间的正交补空间：$\boldsymbol{u}_{l+1} \leftarrow \mathbf{P}_{\mathbf{U}_l}^{\perp}\boldsymbol{u}_{l+1}$
4. 向量归一化：$\boldsymbol{u}_{l+1} \leftarrow \boldsymbol{u}_{l+1}/\|\boldsymbol{u}_{l+1}\|$
5. 重复步骤 2、3 和 4，直至收敛

其中，$\mathbf{P}_{\mathbf{U}_l}^{\perp} = \mathbf{I}_p - \mathbf{U}_l(\mathbf{U}_l^{\mathrm{T}}\mathbf{U}_l)^{-1}\mathbf{U}_l^{\mathrm{T}}$ 为矩阵 $\mathbf{U}_l$ 的正交补投影算子. 显然，$\mathbf{U}_l$ 为列正交矩阵，它的正交补投影算子可以简化为 $\mathbf{P}_{\mathbf{U}_l}^{\perp} = \mathbf{I}_p - \mathbf{U}_l\mathbf{U}_l^{\mathrm{T}}$.

从上面的步骤可以看出，在算法 3.2 中，每次迭代都需要额外将向量投影到已有的特征向量的正交补空间. 事实上，这个步骤是没必要的，即我们可以把上述步骤的

2,3 合并为

$$\begin{aligned}\boldsymbol{u}_{l+1} =& \mathbf{P}_{\mathbf{U}_l}^{\perp}\left(\mathcal{S}(\mathbf{P}_{\mathbf{U}_l}^{\perp}\boldsymbol{u}_{l+1})^2\right)\\ =& \mathcal{S}\times_1\mathbf{P}_{\mathbf{U}_l}^{\perp}\times_2(\mathbf{P}_{\mathbf{U}_l}^{\perp}\boldsymbol{u}_{l+1})\times_3(\mathbf{P}_{\mathbf{U}_l}^{\perp}\boldsymbol{u}_{l+1})\\ =& (\mathcal{S}\times_1\mathbf{P}_{\mathbf{U}_l}^{\perp}\times_2\mathbf{P}_{\mathbf{U}_l}^{\perp}\times_3\mathbf{P}_{\mathbf{U}_l}^{\perp})\boldsymbol{u}_{l+1}^2.\end{aligned} \tag{3.30}$$

记 $\mathcal{S}_l=\mathcal{S}\times_1\mathbf{P}_{\mathbf{U}_l}^{\perp}\times_2\mathbf{P}_{\mathbf{U}_l}^{\perp}\times_3\mathbf{P}_{\mathbf{U}_l}^{\perp}$，则调整后的固定点迭代算法的具体步骤如下.

算法 3.3 简化后的固定点迭代法求解第 $l+1$ 个特征对

1. 计算投影后的张量 $\mathcal{S}_l=\mathcal{S}\times_1\mathbf{P}_{\mathbf{U}_l}^{\perp}\times_2\mathbf{P}_{\mathbf{U}_l}^{\perp}\times_3\mathbf{P}_{\mathbf{U}_l}^{\perp}$
2. 随机初始化向量 $\boldsymbol{u}_{l+1}$
3. 令 $\boldsymbol{u}_{l+1}\leftarrow\mathcal{S}_l\boldsymbol{u}_{l+1}^2$
4. 向量归一化：$\boldsymbol{u}_{l+1}\leftarrow\boldsymbol{u}_{l+1}/\|\boldsymbol{u}_{l+1}\|$
5. 重复步骤 3 和 4，直至收敛

有趣的是，对于任意一个位于前 l 个特征向量所在的子空间中的向量 $\boldsymbol{u}$，都有 $\mathcal{S}_l\boldsymbol{u}^3=0$. 这一定程度揭示了正交约束的工作机制：对原本的协偏度张量进行正交补投影得到新协偏度张量使得数据在前 l 个特征向量张成子空间中任意方向上偏度均为零，从而阻止第 $l+1$ 个特征向量收敛到前 l 个特征向量.

例 3.1 试利用基于正交约束的主偏度分析算法计算如下 $2\times2\times2$ 对称张量的特征对

$$\mathcal{S}_{:,:,1}=\begin{bmatrix}0.9031 & -1.331\\ -1.331 & 0.5509\end{bmatrix},\quad \mathcal{S}_{:,:,2}=\begin{bmatrix}-1.331 & 0.5509\\ 0.5509 & 1.5886\end{bmatrix}.$$

解 我们首先给出该对称张量的三个精确特征向量分别为

$$\boldsymbol{u}_1=\begin{bmatrix}0.8716\\ -0.4903\end{bmatrix},\quad \boldsymbol{u}_2=\begin{bmatrix}0.1284\\ 0.9917\end{bmatrix},\quad \boldsymbol{u}_3=\begin{bmatrix}0.8739\\ 0.4861\end{bmatrix}.$$

利用上述主偏度分析求解算法中基于正交约束的固定点迭代算法，可以得到该对称张量的两个特征向量分别为

$$\boldsymbol{u}_1=\begin{bmatrix}0.8716\\ -0.4903\end{bmatrix},\quad \boldsymbol{u}_2=\begin{bmatrix}0.4093\\ 0.8716\end{bmatrix}.$$

从两组特征向量以及图 3.11 可以看出，使用正交约束的主偏度分析算法得到的第一个特征向量与精确解相同. 这是因为在求解第一个特征向量时不存在任何额外的约束. 但是，正交约束强制要求第二个特征向量与第一个特征向量正交，因此它与精确解的差距较大. 并且，由于正交约束的限制，算法只能得到两组解，而实际上该对称张量的特征对有三个.

我们知道，实对称矩阵的各个特征向量必然相互正交，但从上面的例子可以看出，对称张量的各个特征向量却未必正交. 因此，在主偏度分析中，正交约束的引入虽然一定程度简化了张量特征对求解的难度，但同时也带来了两个显著的问题，即特征对

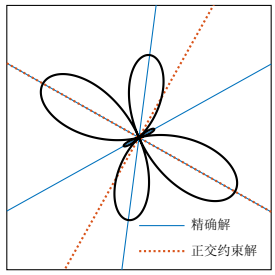


图 3.11　正交约束的主偏度分析算法获得的特征向量

精度的降低以及特征对数量的失准. 为了缓解这一问题，接下来我们将介绍一种非正交约束的主偏度分析算法.

3.4 非正交约束主偏度分析

上述主偏度分析中正交约束不仅仅使得投影后的统计张量在已知特征向量方向上的偏度为零，还迫使所有位于已知特征向量所在的子空间中的投影方向上的偏度都为零. 从这个角度而言，正交约束是一个过于“强硬”的约束，它不仅仅改变了已知特征向量方向的偏度，还将整个子空间方向上所有的偏度都修改了. 而接下来将要介绍的非正交主偏度分析[11]则在一定程度缓解了这一问题. 在此之前，我们首先给出克罗内克积（Kronecker Product）的概念和相关性质.

3.4.1 克罗内克积

克罗内克积是一种在矩阵理论中常见的运算，也被称为直积或张量积. 与常规的矩阵乘法不同的是，它可以将两个任意大小的矩阵组合成一个更大的矩阵. 克罗内克积有许多重要的性质，在诸多研究方向中都有着广泛的应用. 本小节将介绍克罗内克积的定义以及一些基本的性质.

定义 3.8　对于两个矩阵 $\mathbf{A} \in \mathbb{R}^{m\times n}$ 和 $\mathbf{B} \in \mathbb{R}^{p\times q}$，它们的克罗内克积定义为

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \in \mathbb{R}^{mp\times nq}. \tag{3.31}$$

克罗内克积具有如下几个基本的运算性质.

性质 3.1

1. **结合律**：对于任何矩阵 $\mathbf{A}, \mathbf{B}$ 和 $\mathbf{C}$，有 $(\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C})$.
2. **分配律**：如果 $\mathbf{A}, \mathbf{B},$ 和 $\mathbf{C}$ 是三个矩阵，并且它们的大小使得以下的运算都有定

义，那么有 $\mathbf{A}\otimes(\mathbf{B}+\mathbf{C})=\mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C}$.

3. **数乘**：如果 $\mathbf{A}$ 是一个矩阵，c 是一个常数，则有 $c(\mathbf{A}\otimes\mathbf{B})=(c\mathbf{A})\otimes\mathbf{B}=\mathbf{A}\otimes(c\mathbf{B})$.
4. **转置**：对于任何矩阵 $\mathbf{A}$ 和 $\mathbf{B}$，都有 $(\mathbf{A}\otimes\mathbf{B})^{\mathrm{T}}=\mathbf{A}^{\mathrm{T}}\otimes\mathbf{B}^{\mathrm{T}}$.
5. **逆矩阵**：如果矩阵 $\mathbf{A}$ 和 $\mathbf{B}$ 都是可逆的，那么 $\mathbf{A}\otimes\mathbf{B}$ 也可逆，且有 $(\mathbf{A}\otimes\mathbf{B})^{-1}=\mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$.
6. **混合积**：对于任何矩阵 $\mathbf{A},\mathbf{B},\mathbf{C}$ 和 $\mathbf{D}$，如果矩阵乘法 $\mathbf{AC}$ 和 $\mathbf{BD}$ 成立，有 $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=(\mathbf{AC})\otimes(\mathbf{BD})$.

此外，克罗内克积还满足如下的性质.

性质 3.2 给定两个方阵 $\mathbf{A}\in\mathbb{R}^{m\times m}$ 和 $\mathbf{B}\in\mathbb{R}^{n\times n}$，那么有

1. **行列式**：$\det(\mathbf{A}\otimes\mathbf{B})=(\det(\mathbf{A}))^m(\det(\mathbf{B}))^n$.
2. **迹**：$\mathrm{tr}(\mathbf{A}\otimes\mathbf{B})=\mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B})$.
3. **秩**：$\mathrm{rank}(\mathbf{A}\otimes\mathbf{B})=\mathrm{rank}(\mathbf{A})\,\mathrm{rank}(\mathbf{B})$.

关于克罗内克积，还有如下重要公式

$$\mathrm{vec}(\mathbf{AXB}^{\mathrm{T}})=(\mathbf{B}\otimes\mathbf{A})\,\mathrm{vec}(\mathbf{X}). \tag{3.32}$$

如果我们使用张量乘法的记号，那么上式可以写作

$$\mathrm{vec}(\mathbf{X}\times_1\mathbf{A}\times_2\mathbf{B})=(\mathbf{B}\otimes\mathbf{A})\,\mathrm{vec}(\mathbf{X}), \tag{3.33}$$

其中 $\mathrm{vec}(\cdot)$ 代表了向量化操作，可以用来将一个 $I_1\times I_2\times\cdots\times I_N$ 大小的张量转换为一个 $I_1I_2\cdots I_N\times 1$ 大小的向量. 进一步地，对于一个 N 阶张量 $\mathcal{S}$，我们有性质 3.3 .

性质 3.3 给定一个张量 $\mathcal{S}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$，以及 N 个矩阵 $\mathbf{U}_1\in\mathbb{R}^{J_1\times I_1}$，$\mathbf{U}_2\in\mathbb{R}^{J_2\times I_2}$，$\cdots$，$\mathbf{U}_N\in\mathbb{R}^{J_N\times I_N}$，那么有

$$\mathrm{vec}(\mathcal{S}\times_1\mathbf{U}_1\times_2\mathbf{U}_2\cdots\times_N\mathbf{U}_N)=(\mathbf{U}_N\otimes\mathbf{U}_{N-1}\cdots\otimes\mathbf{U}_1)\,\mathrm{vec}(\mathcal{S}). \tag{3.34}$$

此外，为了方便起见，对于一个向量 $\boldsymbol{x}$ 或一个矩阵 $\mathbf{X}$，我们称

$$\underbrace{\boldsymbol{x}\otimes\boldsymbol{x}\cdots\otimes\boldsymbol{x}}_{N\text{个}\boldsymbol{x}},\quad\underbrace{\mathbf{X}\otimes\mathbf{X}\cdots\otimes\mathbf{X}}_{N\text{个}\mathbf{X}},$$

为 $\boldsymbol{x}$ 或 $\mathbf{X}$ 的 N 次克罗内克积.

3.4.2 非正交约束

基于性质 3.3 ，可以将 (3.22) 中的偏度计算公式写成向量内积的形式

$$\mathcal{S}\boldsymbol{u}^3=(\boldsymbol{u}\otimes\boldsymbol{u}\otimes\boldsymbol{u})^{\mathrm{T}}\,\mathrm{vec}(\mathcal{S}). \tag{3.35}$$

类似地，正交约束中投影后的张量 $\mathcal{S}_l$ 可以表示为如下形式

$$\mathrm{vec}(\mathcal{S}_l)=(\mathbf{P}^{\perp}_{\mathbf{U}_l}\otimes\mathbf{P}^{\perp}_{\mathbf{U}_l}\otimes\mathbf{P}^{\perp}_{\mathbf{U}_l})\,\mathrm{vec}(\mathcal{S}). \tag{3.36}$$

因此，投影后的数据在 $\boldsymbol{u}$ 方向的偏度可以表示为

$$\mathcal{S}_l \boldsymbol{u}^3 = (\boldsymbol{u} \otimes \boldsymbol{u} \otimes \boldsymbol{u})^{\mathrm{T}} (\mathbf{P}_{\mathbf{U}_l}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_l}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_l}^{\perp}) \operatorname{vec}(\mathcal{S}). \tag{3.37}$$

进一步地，$\operatorname{vec}(\mathcal{S}_l)$ 和 $\operatorname{vec}(\mathcal{S}_{l-1})$ 之间存在如下关系.

引理 3.1 对于列正交矩阵 $\mathbf{U}_l = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_l \end{bmatrix}$，下面两个向量

$$\operatorname{vec}(\mathcal{S}_{l-1}) = (\mathbf{P}_{\mathbf{U}_{l-1}}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_{l-1}}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_{l-1}}^{\perp}) \operatorname{vec}(\mathcal{S}),$$
$$\operatorname{vec}(\mathcal{S}_l) = (\mathbf{P}_{\mathbf{U}_l}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_l}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_l}^{\perp}) \operatorname{vec}(\mathcal{S}),$$

之间满足如下关系

$$\operatorname{vec}(\mathcal{S}_l) = \mathbf{P}_l \operatorname{vec}(\mathcal{S}_{l-1}), \tag{3.38}$$

其中 $\mathbf{P}_l = \mathbf{P}_{\boldsymbol{u}_l}^{\perp} \otimes \mathbf{P}_{\boldsymbol{u}_l}^{\perp} \otimes \mathbf{P}_{\boldsymbol{u}_l}^{\perp}$ 为向量 $\boldsymbol{u}_l$ 的正交补投影矩阵的三次克罗内克积.

证明 由于正交约束得到的特征向量之间必然正交，因此 $\mathbf{P}_{\mathbf{U}_l}^{\perp}$ 可以拆分为多个投影矩阵的乘积

$$\mathbf{P}_{\mathbf{U}_l}^{\perp} = \mathbf{I} - \mathbf{U}_l \mathbf{U}_l^{\mathrm{T}} = \prod_{i=1}^{l} (\mathbf{I} - \boldsymbol{u}_i \boldsymbol{u}_i^{\mathrm{T}}) = \prod_{i=1}^{l} \mathbf{P}_{\boldsymbol{u}_i}^{\perp}. \tag{3.39}$$

根据性质 3.1 中的混合积性质，不难证明

$$\begin{aligned} \mathbf{P}_{\mathbf{U}_l}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_l}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_l}^{\perp} &= \left(\prod_{i=1}^{l} \mathbf{P}_{\boldsymbol{u}_i}^{\perp} \right) \otimes \left(\prod_{i=1}^{l} \mathbf{P}_{\boldsymbol{u}_i}^{\perp} \right) \otimes \left(\prod_{i=1}^{l} \mathbf{P}_{\boldsymbol{u}_i}^{\perp} \right) \\ &= \prod_{i=1}^{l} \mathbf{P}_{\boldsymbol{u}_i}^{\perp} \otimes \mathbf{P}_{\boldsymbol{u}_i}^{\perp} \otimes \mathbf{P}_{\boldsymbol{u}_i}^{\perp} \\ &= \prod_{i=1}^{l} \mathbf{P}_i, \end{aligned} \tag{3.40}$$

其中 $\mathbf{P}_i = \mathbf{P}_{\boldsymbol{u}_i}^{\perp} \otimes \mathbf{P}_{\boldsymbol{u}_i}^{\perp} \otimes \mathbf{P}_{\boldsymbol{u}_i}^{\perp}$ 为向量 $\boldsymbol{u}_i$ 的正交补投影矩阵的三次克罗内克积.

此外，根据性质 3.1 可以证明，$\mathbf{P}_i$ 和 $\prod_{i=1}^{l} \mathbf{P}_i$ 均为投影矩阵（请读者自己验证）. 并且任意交换乘积的顺序，都不影响 $\prod_{i=1}^{l} \mathbf{P}_i$ 的最终结果.

根据 (3.40) 不难发现，投影后的张量 $\mathcal{S}_l$ 可以写作

$$\begin{aligned} \operatorname{vec}(\mathcal{S}_l) &= \mathbf{P}_{\mathbf{U}_l}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_l}^{\perp} \otimes \mathbf{P}_{\mathbf{U}_l}^{\perp} \operatorname{vec}(\mathcal{S}) = \prod_{i=1}^{l} \mathbf{P}_i \operatorname{vec}(\mathcal{S}) \\ &= \mathbf{P}_l \prod_{i=1}^{l-1} \mathbf{P}_i \operatorname{vec}(\mathcal{S}) = \mathbf{P}_l \operatorname{vec}(\mathcal{S}_{l-1}). \end{aligned} \tag{3.41}$$

■

根据 (3.35) 和 引理 3.1 ，对于已经得到的特征向量 $\boldsymbol{u}_l$，有

$$\begin{aligned}\mathcal{S}_l\boldsymbol{u}_l^3 &= (\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)^{\mathrm{T}}\operatorname{vec}(\mathcal{S})\\ &= (\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)^{\mathrm{T}}\mathbf{P}_l\operatorname{vec}(\mathcal{S}_{l-1})\\ &= (\mathbf{P}_{\boldsymbol{u}_l}^{\perp}\boldsymbol{u}_l\otimes\mathbf{P}_{\boldsymbol{u}_l}^{\perp}\boldsymbol{u}_l\otimes\mathbf{P}_{\boldsymbol{u}_l}^{\perp}\boldsymbol{u}_l)^{\mathrm{T}}\operatorname{vec}(\mathcal{S}_{l-1})\\ &= \mathbf{0}_{p^3}^{\mathrm{T}}\operatorname{vec}(\mathcal{S}_{l-1})\\ &= 0.\end{aligned}\tag{3.42}$$

也就是说，经过投影矩阵 $\mathbf{P}_l$ 的作用，张量 $\mathcal{S}_l$ 中在 $\mathcal{S}_{l-1}$ 的基础上去除了特征向量 $\boldsymbol{u}_l$ 的信息，而张量 $\mathcal{S}_{l-1}$ 已不包含前 $l-1$ 个特征向量的任何信息，因此 S_l 的特征向量的求解也当然不会收敛到特征向量 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_l$ 中的任何一个.

注意到，经过投影矩阵 $\mathbf{P}_i$ 的投影后，不仅使得 $\boldsymbol{u}_i\otimes\boldsymbol{u}_i\otimes\boldsymbol{u}_i$ 为零向量，所有形如 $\boldsymbol{u}_i\otimes\boldsymbol{v}\otimes\boldsymbol{w},\boldsymbol{v}\otimes\boldsymbol{u}_i\otimes\boldsymbol{w},\boldsymbol{v}\otimes\boldsymbol{w}\otimes\boldsymbol{u}_i$ 的向量也都变为了零向量，比如

$$\begin{aligned}&(\mathbf{P}_{\boldsymbol{u}_i}^{\perp}\otimes\mathbf{P}_{\boldsymbol{u}_i}^{\perp}\otimes\mathbf{P}_{\boldsymbol{u}_i}^{\perp})(\boldsymbol{u}_i\otimes\boldsymbol{v}\otimes\boldsymbol{w})\\ =&(\mathbf{P}_{\boldsymbol{u}_i}^{\perp}\boldsymbol{u}_i)\otimes(\mathbf{P}_{\boldsymbol{u}_i}^{\perp}\boldsymbol{v})\otimes(\mathbf{P}_{\boldsymbol{u}_i}^{\perp}\boldsymbol{w})\\ =&\mathbf{0}_p\otimes(\mathbf{P}_{\boldsymbol{u}_i}^{\perp}\boldsymbol{v})\otimes(\mathbf{P}_{\boldsymbol{u}_i}^{\perp}\boldsymbol{w})\\ =&\mathbf{0}_{p^3}.\end{aligned}$$

因此，一定程度上可以说，正交约束产生的投影矩阵 $\mathbf{P}_i$ 是 “强硬的” 或者 “过剩的”. 事实上，为了消除上一次迭代时使用的张量 $\mathcal{S}_{l-1}$ 中特征向量 $\boldsymbol{u}_l$ 的信息，$\mathbf{P}_l$ 并不是唯一的选择. 对于已获得的特征向量 $\boldsymbol{u}_l$，我们只需要找到一个矩阵 $\tilde{\mathbf{P}}_l$，对张量进行投影 $\operatorname{vec}(\mathcal{S}_l)=\tilde{\mathbf{P}}_l\operatorname{vec}(\mathcal{S}_{l-1})$，使得投影后的张量满足如下等式即可

$$\mathcal{S}_l\boldsymbol{u}_l^3=(\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)^{\mathrm{T}}\tilde{\mathbf{P}}_l\operatorname{vec}(\mathcal{S}_{l-1})=0.\tag{3.43}$$

也就是说，我们只需找到一个矩阵 $\tilde{\mathbf{P}}_l$，使得 $\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l$ 位于它的核空间即可. 那么这样的矩阵 $\tilde{\mathbf{P}}_l$ 存在么？

利用克罗内克积的性质 3.1 ，我们可以发现

$$(\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)^{\mathrm{T}}(\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)=(\boldsymbol{u}_l^{\mathrm{T}}\boldsymbol{u}_l)\otimes(\boldsymbol{u}_l^{\mathrm{T}}\boldsymbol{u}_l)\otimes(\boldsymbol{u}_l^{\mathrm{T}}\boldsymbol{u}_l)=1\otimes1\otimes1=1,$$

因此向量 $\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l$ 的正交补投影矩阵为

$$\tilde{\mathbf{P}}_l=\mathbf{I}_{p^3}-(\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)(\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)^{\mathrm{T}}.$$

显然这个正交补投影矩阵 $\tilde{\mathbf{P}}_l$ 就是满足 (3.43) 要求的投影矩阵. 基于新的投影矩阵 $\tilde{\mathbf{P}}_l$，数据在 $\boldsymbol{u}$ 方向上的偏度可以表示为

$$\begin{aligned}\mathcal{S}_l\boldsymbol{u}^3&=(\boldsymbol{u}\otimes\boldsymbol{u}\otimes\boldsymbol{u})^{\mathrm{T}}\operatorname{vec}(\mathcal{S}_l)=(\boldsymbol{u}\otimes\boldsymbol{u}\otimes\boldsymbol{u})^{\mathrm{T}}\tilde{\mathbf{P}}_l\operatorname{vec}(\mathcal{S}_{l-1})\\ &=\Big(\tilde{\mathbf{P}}_l(\boldsymbol{u}\otimes\boldsymbol{u}\otimes\boldsymbol{u})\Big)^{\mathrm{T}}\operatorname{vec}(\mathcal{S}_{l-1}).\end{aligned}\tag{3.44}$$

从 (3.44) 可以看到，$\tilde{\mathbf{P}}_l$ 并没有直接作用到向量 $\boldsymbol{u}$ 上，而是作用到向量 $\boldsymbol{u}\otimes\boldsymbol{u}\otimes\boldsymbol{u}$ 上，因此得到的特征向量有可能相互之间不正交，这正是本算法被称作非正交约束的主偏度分析的原因所在.

相较而言，正交约束主偏度分析中 $\boldsymbol{u}$ 方向上的偏度可以表示为

$$\begin{aligned}\mathcal{S}_l\boldsymbol{u}^3 &= (\boldsymbol{u}\otimes\boldsymbol{u}\otimes\boldsymbol{u})^{\mathrm{T}}(\mathbf{P}_{\mathbf{U}_l}^{\perp}\otimes\mathbf{P}_{\mathbf{U}_l}^{\perp}\otimes\mathbf{P}_{\mathbf{U}_l}^{\perp})\operatorname{vec}(\mathcal{S})\\ &= (\mathbf{P}_{\mathbf{U}_l}^{\perp}\boldsymbol{u}\otimes\mathbf{P}_{\mathbf{U}_l}^{\perp}\boldsymbol{u}\otimes\mathbf{P}_{\mathbf{U}_l}^{\perp}\boldsymbol{u})^{\mathrm{T}}\operatorname{vec}(\mathcal{S}).\end{aligned} \tag{3.45}$$

这意味着投影矩阵 $\mathbf{P}_l$ 是直接对向量 $\boldsymbol{u}$ 进行了正交化处理，因此得到的特征向量之间必然正交.

此外，在新的投影矩阵 $\tilde{\mathbf{P}}_l$ 下，投影后的张量 $\mathcal{S}_l$ 和 $\mathcal{S}_{l-1}$ 之间存在引理 3.2 中给出的关系.

引理 3.2 *在非正交约束下，对于已获得的特征向量 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_l$，基于投影矩阵 $\tilde{\mathbf{P}}_l$ 投影后的张量 $\mathcal{S}_l$ 和 $\mathcal{S}_{l-1}$ 之间满足如下关系*

$$\mathcal{S}_l = \mathcal{S}_{l-1} - \mathcal{S}_{l-1}\boldsymbol{u}_l^3(\boldsymbol{u}_l\circ\boldsymbol{u}_l\circ\boldsymbol{u}_l). \tag{3.46}$$

证明 利用克罗内克积的性质 3.3 ，可以将 $\mathbf{S}_l$ 写成如下形式

$$\begin{aligned}\operatorname{vec}(\mathcal{S}_l) &= \prod_{i=1}^{l}\tilde{\mathbf{P}}_i\operatorname{vec}(\mathcal{S}) = \tilde{\mathbf{P}}_l\prod_{i=1}^{l-1}\tilde{\mathbf{P}}_i\operatorname{vec}(\mathcal{S})\\ &= \left(\mathbf{I}_{p^3} - (\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)(\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)^{\mathrm{T}}\right)\operatorname{vec}(\mathcal{S}_{l-1})\\ &= \operatorname{vec}(\mathcal{S}_{l-1}) - (\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)(\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)^{\mathrm{T}}\operatorname{vec}(\mathcal{S}_{l-1})\\ &= \operatorname{vec}(\mathcal{S}_{l-1}) - \mathcal{S}_{l-1}\boldsymbol{u}_l^3(\boldsymbol{u}_l\otimes\boldsymbol{u}_l\otimes\boldsymbol{u}_l)\\ &= \operatorname{vec}(\mathcal{S}_{l-1}) - \mathcal{S}_{l-1}\boldsymbol{u}_l^3\operatorname{vec}(\boldsymbol{u}_l\circ\boldsymbol{u}_l\circ\boldsymbol{u}_l)\\ &= \operatorname{vec}(\mathcal{S}_{l-1} - \mathcal{S}_{l-1}\boldsymbol{u}_l^3(\boldsymbol{u}_l\circ\boldsymbol{u}_l\circ\boldsymbol{u}_l)),\end{aligned}$$

即

$$\mathcal{S}_l = \mathcal{S}_{l-1} - \mathcal{S}_{l-1}\boldsymbol{u}_l^3(\boldsymbol{u}_l\circ\boldsymbol{u}_l\circ\boldsymbol{u}_l).$$

■

注意到 $\tilde{\mathbf{P}}_l$ 的大小为 $p^3\times p^3$，在数据维度较高时，计算 $\operatorname{vec}(\mathcal{S}_l) = \prod_{i=1}^{l}\tilde{\mathbf{P}}_l\operatorname{vec}(\mathcal{S})$ 的代价会很大. 而利用引理 3.2 中给出的迭代公式则可以大大降低计算量. 综上所述，可以得到利用非正交约束计算第 $l+1$ 个特征向量的流程，具体见算法 3.4 .

对比算法 3.3 和算法 3.4 的流程，可以发现两者几乎是相同的. 唯一的区别在于，正交约束的主偏度分析中投影后张量的计算公式如下

$$\mathcal{S}_l = \mathcal{S}\times_1\mathbf{P}_{\mathbf{U}_l}^{\perp}\times_2\mathbf{P}_{\mathbf{U}_l}^{\perp}\times_3\mathbf{P}_{\mathbf{U}_l}^{\perp},$$

算法 3.4 非正交约束主偏度分析求解第 $l+1$ 个特征向量的算法流程

1. 计算投影后的张量 $\mathcal{S}_l = \mathcal{S}_{l-1} - \mathcal{S}_{l-1}\boldsymbol{u}_l^3(\boldsymbol{u}_l \circ \boldsymbol{u}_l \circ \boldsymbol{u}_l)$
2. 随机初始化向量 $\boldsymbol{u}_{l+1}$
3. 令 $\boldsymbol{u}_{l+1} = \mathcal{S}_l\boldsymbol{u}_{l+1}^2$
4. 向量归一化：$\boldsymbol{u}_{l+1} = \boldsymbol{u}_{l+1}/\|\boldsymbol{u}_{l+1}\|$
5. 重复步骤 3 和 4，直至收敛

而非正交约束的主偏度分析中投影后张量的计算公式则为

$$\mathcal{S}_l = \mathcal{S}_{l-1} - \mathcal{S}_{l-1}\boldsymbol{u}_l^3(\boldsymbol{u}_l \circ \boldsymbol{u}_l \circ \boldsymbol{u}_l).$$

尽管两者的形式并不相同，但它们本质都是对向量化后的张量进行投影，只是两者使用了不同的投影矩阵，分别为

$$\begin{aligned}\operatorname{vec}(\mathcal{S}_l) &= \mathbf{P}_l \operatorname{vec}(\mathcal{S}_{l-1}) = \left((\mathbf{I} - \boldsymbol{u}_l\boldsymbol{u}_l^{\mathrm{T}}) \otimes (\mathbf{I} - \boldsymbol{u}_l\boldsymbol{u}_l^{\mathrm{T}}) \otimes (\mathbf{I} - \boldsymbol{u}_l\boldsymbol{u}_l^{\mathrm{T}})\right) \operatorname{vec}(\mathcal{S}_{l-1}),\\ \operatorname{vec}(\mathcal{S}_l) &= \tilde{\mathbf{P}}_l \operatorname{vec}(\mathcal{S}_{l-1}) = \left(\mathbf{I}_{p^3} - (\boldsymbol{u}_l \otimes \boldsymbol{u}_l \otimes \boldsymbol{u}_l)(\boldsymbol{u}_l \otimes \boldsymbol{u}_l \otimes \boldsymbol{u}_l)^{\mathrm{T}}\right) \operatorname{vec}(\mathcal{S}_{l-1}).\end{aligned}$$

注意到 $\tilde{\mathbf{P}}_l$ 的秩为 p^3-1，而根据性质 3.2 ，正交约束中的投影矩阵 $\mathbf{P}_l$ 的秩为 $(p-1)^3$. 由于 $p>1$ 时，$p^3-1>(p-1)^3$，因此，在非正交约束的主偏度分析中，经过投影后的张量 $\operatorname{vec}(\mathcal{S}_l) = \tilde{\mathbf{P}}_l \operatorname{vec}(\mathcal{S}_{l-1})$ 似乎有更多的信息得到保留. 对于这两个投影矩阵，我们有如下定理.

定理 3.1 *设已获得的第 l 个特征向量为 $\boldsymbol{u}_l$，正交约束和非正交约束在求解第 $l+1$ 个特征向量时对应的两个投影矩阵 $\mathbf{P}_l$ 和 $\tilde{\mathbf{P}}_l$ 的核空间满足如下关系*

$$\operatorname{null}\left(\tilde{\mathbf{P}}_l\right) \subseteq \operatorname{null}\left(\mathbf{P}_l\right),$$

其中 $\operatorname{null}(\cdot)$ 表示矩阵的核空间.

证明 非正交约束使用的投影矩阵 $\tilde{\mathbf{P}}_l$ 的核空间为第 l 个特征向量的三次克罗内克积构成的子空间，即

$$\operatorname{null}\left(\tilde{\mathbf{P}}_l\right) = \operatorname{span}(\boldsymbol{u}_l \otimes \boldsymbol{u}_l \otimes \boldsymbol{u}_l).$$

而正交约束使用的投影矩阵 $\mathbf{P}_l$ 的核空间为

$$\begin{aligned}\operatorname{null}\left(\mathbf{P}_l\right) &= \operatorname{span}(\boldsymbol{u}_l \otimes \boldsymbol{v} \otimes \boldsymbol{w}, \boldsymbol{v} \otimes \boldsymbol{u}_l \otimes \boldsymbol{w}, \boldsymbol{v} \otimes \boldsymbol{w} \otimes \boldsymbol{u}_l | \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^{p\times 1})\\ &= \operatorname{span}\left(\begin{bmatrix}\boldsymbol{u}_l \otimes \mathbf{I}_p \otimes \mathbf{I}_p & \mathbf{I}_p \otimes \boldsymbol{u}_l \otimes \mathbf{I}_p & \mathbf{I}_p \otimes \mathbf{I}_p \otimes \boldsymbol{u}_l\end{bmatrix}\right),\end{aligned}$$

其中 $\operatorname{span}(\cdot)$ 表示由给定的向量或者矩阵的列向量张成的子空间. 由于 $\boldsymbol{u}_l \otimes \mathbf{I}_p \otimes \mathbf{I}_p$ 为一个矩阵，并且可以通过线性组合得到 $\boldsymbol{u}_l \otimes \boldsymbol{u}_l \otimes \boldsymbol{u}_l$

$$(\boldsymbol{u}_l \otimes \mathbf{I}_p \otimes \mathbf{I}_p)(1 \otimes \boldsymbol{u}_l \otimes \boldsymbol{u}_l) = \boldsymbol{u}_l \otimes \boldsymbol{u}_l \otimes \boldsymbol{u}_l,$$

因此

$$\operatorname{span}(\boldsymbol{u}_l \otimes \boldsymbol{u}_l \otimes \boldsymbol{u}_l | 1 \le i \le l) \subseteq \operatorname{span}\left(\boldsymbol{u}_l \otimes \mathbf{I}_p \otimes \mathbf{I}_p\right).$$

所以，$\text{null}\left(\tilde{\mathbf{P}}_l\right)$ 是 $\text{null}\left(\mathbf{P}_l\right)$ 的子空间. ■

以 $l=1$ 为例，此时 $\mathbf{U}_1=\boldsymbol{u}_1$，在正交约束主偏度分析中

$$\mathbf{P}_1=(\mathbf{I}_p-\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}})\otimes(\mathbf{I}_p-\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}})\otimes(\mathbf{I}_p-\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}})$$

为 $\boldsymbol{u}_1$ 的正交补的三次克罗内克积，因此 $\mathbf{P}_1\,\text{vec}(\mathcal{S})$ 是将 $\text{vec}(\mathcal{S})$ 投影到 $\boldsymbol{u}_1\otimes\mathbf{I}_p\otimes\mathbf{I}_p,\mathbf{I}_p\otimes\boldsymbol{u}_1\otimes\mathbf{I}_p,\mathbf{I}_p\otimes\mathbf{I}_p\otimes\boldsymbol{u}_1$ 这三个矩阵的列向量张成的空间的正交补空间. 而在非正交约束主偏度分析中

$$\tilde{\mathbf{P}}_1=\mathbf{I}_{p^3}-(\boldsymbol{u}_1\otimes\boldsymbol{u}_1\otimes\boldsymbol{u}_1)(\boldsymbol{u}_1\otimes\boldsymbol{u}_1\otimes\boldsymbol{u}_1)^{\mathrm{T}}$$

为 $\boldsymbol{u}_1$ 的三次克罗内克积的正交补，所以 $\tilde{\mathbf{P}}_1\,\text{vec}(\mathcal{S})$ 是将 $\text{vec}(\mathcal{S})$ 投影到向量 $\boldsymbol{u}_1\otimes\boldsymbol{u}_1\otimes\boldsymbol{u}_1$ 的正交补空间. 显然，后一个正交补空间包含了前一个正交补空间. 因此，在求解 $\mathcal{S}$ 的第二个特征向量的时候，基于非正交约束的主偏度的分析可以在一个更大的空间进行搜索，理应得到更为精确的解.

利用非正交约束，理论上我们最多可以得到 p^3 个特征向量，并且得到的特征向量之间并不一定正交. 这是否意味着张量特征对求解问题被解决了呢？答案是否定的. 张量的特征向量要求 $\mathcal{S}\boldsymbol{u}_{l+1}^2=\lambda_{l+1}\boldsymbol{u}_{l+1}$，而非正交约束只能保证 $\mathcal{S}_l\boldsymbol{u}_{l+1}^2=\lambda_{l+1}\boldsymbol{u}_{l+1}$. 假设 $\boldsymbol{u}_{l+1}$ 严格为张量 $\mathcal{S}$ 的特征向量，由于 $\mathcal{S}_l\boldsymbol{u}_{l+1}^2$ 并不一定平行于 $\mathcal{S}\boldsymbol{u}_{l+1}^2$，所以非正交约束也不能保证得到的特征向量是协偏度张量的精确特征解.

例 3.2 试利用基于非正交约束的主偏度分析算法计算如下 $2\times2\times2$ 对称张量的特征对

$$\mathcal{S}_{:,:,1}=\begin{bmatrix}0.9031 & -1.331\\ -1.331 & 0.5509\end{bmatrix},\quad \mathcal{S}_{:,:,2}=\begin{bmatrix}-1.331 & 0.5509\\ 0.5509 & 1.5886\end{bmatrix}.$$

解

我们首先给出该对称张量的三个精确特征向量分别为

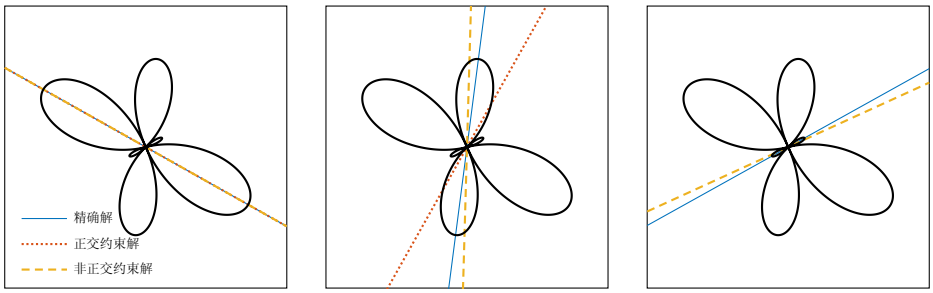
$$\boldsymbol{u}_1=\begin{bmatrix}0.8716\\ -0.4903\end{bmatrix},\quad \boldsymbol{u}_2=\begin{bmatrix}0.1284\\ 0.9917\end{bmatrix},\quad \boldsymbol{u}_3=\begin{bmatrix}0.8739\\ 0.4861\end{bmatrix}.$$

利用上述主偏度分析求解算法中基于非正交约束的固定点迭代算法，可以得到该对称张量的三个特征向量（人为设定的数量）分别为

$$\boldsymbol{u}_1=\begin{bmatrix}0.8716\\ -0.4903\end{bmatrix},\quad \boldsymbol{u}_2=\begin{bmatrix}0.0282\\ 0.9996\end{bmatrix},\quad \boldsymbol{u}_3=\begin{bmatrix}0.9096\\ 0.4156\end{bmatrix}.$$

我们将算法获得的特征向量绘制在图 3.12 中，并与精确解以及正交约束解进行对比. 从图 3.12a 中可以看出，和正交约束主偏度分析一样，算法得到的第一个特征向量与精确解相同. 从图 3.12b 中则能发现，由于非正交约束对张量 $\mathcal{S}$ 的修改相对较小，因此对应的解要更加接近精确解. 不过需要注意的是，张量 $\mathcal{S}$ 一共有三个特征向量，但正交约束的主偏度分析算法只能得到两个. 而非正交约束的主偏度分析算法能够通过人为设定特征对数量获得三个，但除了第一个特征向量外，其余两个特征向量并不精

确.



(a) (b) (c)

图 3.12 非正交约束的主偏度分析算法获得的特征向量 (a) 第一个特征向量 (b) 第二个特征向量 (c) 第三个特征向量

3.5 主偏度分析与独立成分分析

在信号处理中，独立成分分析（Independent Component Analysis, ICA）指的是一类将多元信号分解为多个独立子成分的盲源分离算法. 我们在第 2 章中提到的 “鸡尾酒会问题” 就是独立成分分析的一个重要应用场景.

独立成分分析相关的算法有很多，其中最著名的便是快速独立成分分析算法（FastICA）[12]. 本节将会介绍快速独立成分分析算法的基本原理，并将其与主偏度分析算法进行比较.

3.5.1 快速独立成分分析

对于一个大小为 $p \times n$ 的白化后的数据 $\mathbf{X}$，FastICA 算法的基本思想是找到一个投影方向 $\boldsymbol{u} \in \mathbb{R}^{p\times 1}$，使得投影后的数据 $\boldsymbol{u}^{\mathrm{T}}\mathbf{X}$ 的某种非高斯性指标尽量大，相应的优化模型为

$$\begin{cases} \max\limits_{\boldsymbol{u}} & G(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\mathbf{1} \\ \text{s.t.} & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1 \end{cases}, \tag{3.47}$$

其中 $G(\cdot)$ 是用于衡量数据非高斯性的函数，常用的非高斯性衡量函数包括 x^3，x^4，$\log\cosh(x)$，$-e^{-x^2/2}$ 等. $G(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})$ 代表对向量 $\boldsymbol{u}^{\mathrm{T}}\mathbf{X}$ 中的每一个元素都用函数 $G(\cdot)$ 进行非线性映射从而得到一个新的向量. 可以验证，当 $G(x) = x^3$ 时，优化模型 (3.47) 中的目标函数 $G(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\mathbf{1}$ 正好对应数据在 $\boldsymbol{u}$ 方向的偏度（详见第 3.5.2 小节公式 (3.53)）；而当 $G(x) = x^4$ 时，该目标函数则对应了数据在 $\boldsymbol{u}$ 方向的峭度.

FastICA 采用牛顿迭代法求解模型 (3.47)，首先构建如下的拉格朗日函数

$$\mathcal{L}(\boldsymbol{u},\lambda)=G(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\mathbf{1}+\frac{1}{2}\lambda(1-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}), \tag{3.48}$$

$\mathcal{L}(\boldsymbol{u},\lambda)$ 关于自变量的导数为

$$\frac{\partial\mathcal{L}(\boldsymbol{u},\lambda)}{\partial\boldsymbol{u}}=\mathbf{X}G'(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}-\lambda\boldsymbol{u}, \tag{3.49}$$

$\mathcal{L}(\boldsymbol{u},\lambda)$ 关于自变量的黑塞（Hessian）矩阵[12]为

$$\begin{aligned}\frac{\partial^2\mathcal{L}(\boldsymbol{u},\lambda)}{\partial\boldsymbol{u}\partial\boldsymbol{u}^{\mathrm{T}}}&=\mathbf{X}\operatorname{diag}\left(G''(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\right)\mathbf{X}^{\mathrm{T}}-\lambda\mathbf{I}_p\\&\approx\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}}G''(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\mathbf{1}-\lambda\mathbf{I}_p\\&=\left(\frac{1}{n}G''(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\mathbf{1}-\lambda\right)\mathbf{I}_p.\end{aligned} \tag{3.50}$$

其中 $G'(\cdot)$ 和 $G''(\cdot)$ 分别表示 $G(\cdot)$ 的一阶和二阶导数，它们的大小均与自变量的大小相同[2].

根据 (3.49) 和 (3.50)，利用牛顿迭代法，我们可以得到如下的迭代公式

$$\begin{aligned}\boldsymbol{u}\leftarrow&\boldsymbol{u}-\left(\frac{\partial^2\mathcal{L}(\boldsymbol{u},\lambda)}{\partial\boldsymbol{u}\partial\boldsymbol{u}^{\mathrm{T}}}\right)^{-1}\frac{\partial\mathcal{L}(\boldsymbol{u},\lambda)}{\partial\boldsymbol{u}}\\=&\boldsymbol{u}-\frac{\mathbf{X}G'(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}-\lambda\boldsymbol{u}}{\frac{1}{n}G''(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\mathbf{1}-\lambda}.\end{aligned} \tag{3.51}$$

由于接下来我们还需要对 $\boldsymbol{u}$ 进行归一化，因此在 (3.51) 右边乘以 $\left(\frac{1}{n}G''(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\mathbf{1}-\lambda\right)$ 并不会影响最终结果，所以可以将 (3.51) 改写为

$$\boldsymbol{u}\leftarrow\frac{1}{n}G''(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\mathbf{1}\boldsymbol{u}-\mathbf{X}G'(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}. \tag{3.52}$$

根据 (3.52) 给出的迭代公式，我们可以得到 FastICA 求解第一个独立成分的具体步骤（算法 3.5 ）.

算法 3.5 FastICA 求解第一个独立成分的流程

1. 随机初始化向量 $\boldsymbol{u}$
2. 令 $\boldsymbol{u}\leftarrow\frac{1}{n}G''(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})\mathbf{1}\boldsymbol{u}-\mathbf{X}G'(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}$
3. 向量归一化：$\boldsymbol{u}\leftarrow\boldsymbol{u}/\|\boldsymbol{u}\|$
4. 重复步骤 2 和 3，直至收敛

由于 FastICA 假设所有的独立成分之间相互正交，因此在求解第 $l+1$ 个独立成分时，迭代过程中的每一步都需要将 $\boldsymbol{u}_{l+1}$ 投影到前 l 投影方向 $\mathbf{U}_l=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_l\end{bmatrix}$ 的正交补空间中. 对应的算法流程见算法 3.6 .

[2]比如 $G(x)=x^3$ 时，其一阶导数为 $G'(x)=3x^2$，二阶导数为 $G''(x)=6x$. 当自变量为 n 维行向量 $\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}$ 时，我们有 $G\left(\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}\right)=\begin{bmatrix}x_1^3 & \cdots & x_n^3\end{bmatrix}$，$G'\left(\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}\right)=\begin{bmatrix}3x_1^2 & \cdots & 3x_n^2\end{bmatrix}$，$G''\left(\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}\right)=\begin{bmatrix}6x_1 & \cdots & 6x_n\end{bmatrix}$.

算法 3.6 FastICA 求解第 $l+1$ 个独立成分的流程

1. 随机初始化向量 $\boldsymbol{u}_{l+1}$
2. 令 $\boldsymbol{u}_{l+1} \leftarrow \frac{1}{n} G''(\boldsymbol{u}_{l+1}^{\mathrm{T}}\mathbf{X})\mathbf{1}\boldsymbol{u}_{l+1} - \mathbf{X}G'(\boldsymbol{u}_{l+1}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}$
3. 将 $\boldsymbol{u}_{l+1}$ 投影到 $\mathbf{U}_l$ 的正交补空间中：$\boldsymbol{u}_{l+1} \leftarrow \mathbf{P}_{\mathbf{U}_l}^{\perp}\boldsymbol{u}_{l+1}$
4. 向量归一化：$\boldsymbol{u} \leftarrow \boldsymbol{u}/\|\boldsymbol{u}\|$
5. 重复步骤 3 和 4，直至收敛

3.5.2 FastICA 与主偏度分析

对比算法 3.1 和 3.5 可以发现，FastICA 求解第一个独立成分与主偏度分析求解第一个特征对的流程非常接近，两者之间的主要区别在于投影方向的更新方式. FastICA 利用如下公式更新投影方向

$$\boldsymbol{u} \leftarrow \frac{1}{n} G''(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}\mathbf{1}\boldsymbol{u} - \mathbf{X}G'(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}},$$

而主偏度分析更新投影方向的公式为

$$\boldsymbol{u} \leftarrow \mathcal{S}\boldsymbol{u}^2.$$

注意到，FastICA 可以选择与主偏度分析相同的非高斯性指标，即偏度. 在这种情况下，非高斯性衡量函数为 $G(x) = x^3$，并且我们有

$$\mathrm{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \frac{1}{n} G(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}\mathbf{1} = \mathcal{S}\boldsymbol{u}^3. \tag{3.53}$$

这表明，当 FastICA 选择偏度作为非高斯性指标时，其目标函数与主偏度分析的目标函数是一致的. 进一步地，由于 $G'(x) = 3x^2$，我们可以证明

$$\begin{aligned}
\left(\mathbf{X}G'(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}\right)_i &= \sum_{l=1}^{n} x_{il} G'(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})_l = 3\sum_{l=1}^{n} x_{il}\left(\sum_{j=1}^{p} x_{jl}u_j\right)^2 \\
&= 3\sum_{l=1}^{n} x_{il}\left(\sum_{j=1}^{p}\sum_{k=1}^{p} x_{jl}x_{kl}u_j u_k\right) \\
&= 3\sum_{j=1}^{p}\sum_{k=1}^{p}\left(\sum_{l=1}^{n} x_{il}x_{jl}x_{kl}\right)u_j u_k \\
&= 3n(\mathcal{S}\boldsymbol{u}^2)_i,
\end{aligned} \tag{3.54}$$

其中 x_{ij} 表示矩阵 $\mathbf{X}$ 的第 i 行第 j 列元素. 又因为 $G''(x) = 6x$，所以

$$G''(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}\mathbf{1} = 6\boldsymbol{u}^{\mathrm{T}}\mathbf{X}\mathbf{1} = 0. \tag{3.55}$$

因此，当使用偏度作为非高斯性指标时，FastICA 的迭代公式 (3.52) 等价于

$$\boldsymbol{u} \leftarrow -3n\mathcal{S}\boldsymbol{u}^2. \tag{3.56}$$

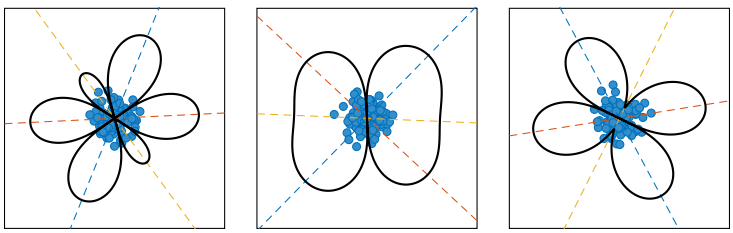
由于 $\mathrm{skew}(-\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = -\mathrm{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})$，并且在后续的步骤中我们需要对投影方向进行归一化，因此前面的系数 $-3n$ 并不影响最终的结果. 这意味着当 FastICA 选择偏度作为非高斯性指标时，其迭代公式与主偏度分析的迭代公式是等价的，两者的结果仅有可能存在正负号的区别. 此外，在 FastICA 中也使用了正交约束来求解剩余的独立成分，因此就结果而言，在初值相同的条件下，FastICA 与正交约束的主偏度分析算法是完全相同的.

注意到，在 FastICA 中每次迭代都需要观测矩阵 $\mathbf{X} \in \mathbb{R}^{p\times n}$ 的参与. 对于一些图像数据而言，p 往往较小只有个位数，而 n 为图像像素个数，通常达到几十万甚至上百万级别. 因此，FastICA 在处理这类样本数极大的数据时，计算量往往会非常大. 而主偏度分析算法的迭代只涉及到对三阶统计张量 $\mathcal{S} \in \mathbb{R}^{p\times p\times p}$ 的相关运算，因此计算量通常要远远小于 FastICA.

3.6 主偏度分析的几何解释

在第 2 章中，我们已经知道，对于任意的观测数据 $\mathbf{X} \in \mathbb{R}^{p\times n}$，其在样本空间都存在一个与方差指标相对应的超椭球面，而这个超椭球面正是数据的各个主成分方向相互垂直的几何本源. 那么，在样本空间中，是否也存在一个与偏度指标对应的几何结构呢？如果存在的话，这个几何结构是否蕴含了偏度极值方向分布的规律呢？

接下来，我们首先展示几个简单的例子，看看是否能从这些例子中窥探出数据偏度极值分布的规律.



(a)　　(b)　　(c)

图 3.13　偏度的几何图像 (a) 随机数据一 (b) 随机数据二 (c) 随机数据三

从图 3.13 可以看出，数据的偏度极值方向分布与方差极值方向分布展现出截然不同的现象. 虽然同是平面上的散点，它们的偏度极值方向不但不垂直，而且似乎没有任何规律可言. 我们从几何角度来解释主偏度分析的努力似乎要无功而返. 接下来，我们将要从最简单的情况出发，给出一些特殊数据的偏度分布情况，并基于此进一步探讨一般数据偏度极值分布的几何意义.

3.6.1 单形体的偏度映射图

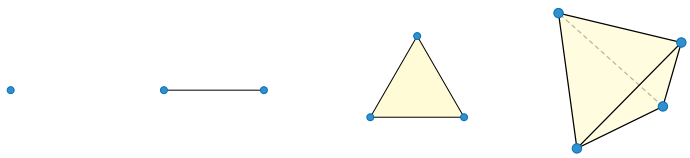
单形体是欧几里得空间中最简单、最基本的几何结构，其定义如下

定义 3.9 设有 $n+1$ 个 n 维向量 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n+1}$，并且 $\boldsymbol{x}_2-\boldsymbol{x}_1, \boldsymbol{x}_3-\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n+1}-\boldsymbol{x}_1$ 线性无关，那么由这 $n+1$ 个向量决定的点集

$$\left\{ \sum_{i=1}^{n+1} \theta_i \boldsymbol{x}_i \middle| \sum_{i=1}^{n+1} \theta_i = 1, \theta_i \geq 0 \right\},$$

被称作 n 维单形体.

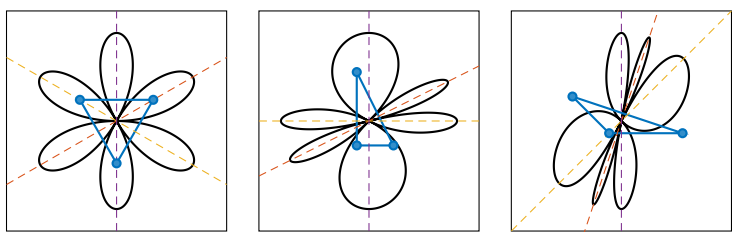
最常见的单形体有零维空间中的点，一维空间中的线段，二维空间中的三角形和三维空间中的四面体，如图 3.14 所示.



(a) (b) (c) (d)

图 3.14 常见的单形体 (a) 点 (b) 线段 (c) 三角形 (d) 四面体

对于一个 n 维空间中的单形体，我们可以用一个矩阵 $\mathbf{X} \in \mathbb{R}^{n\times(n+1)}$ 来表示它的 $n+1$ 个顶点，并根据定义 3.6 可以绘制出该数据的偏度映射图. 在本节（第 3.6 节），为了表达方便起见，我们也称 $\mathbf{X}$ 的偏度映射图为以 $\mathbf{X}$ 的 $n+1$ 个列向量为顶点的单形体的偏度映射图. 图 3.15 给出了几个常见的三角形的偏度映射图. 从中可以发现，对于任意类型的三角形，其偏度极值方向总与三角形的高的方向平行（分别垂直于三角形的三条边）.



(a) (b) (c)

图 3.15 不同类型的三角形的偏度映射图及局部偏度极值方向示意图 (a) 正三角形 (b) 直角三角形 (c) 钝角三角形

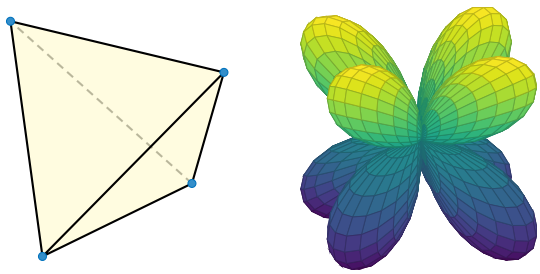
进一步将以上结论推广到更高维的单形体，我们有如下重要结论[13].

定理 3.2 对于一个与单形体顶点对应的数据 $\mathbf{X} \in \mathbb{R}^{n\times(n+1)}$，其局部偏度极值方向与单形体任意顶点到剩余 n 个顶点构成的子单形体的高的方向平行.

我们以三维空间的四面体为例对定理 3.2 进行进一步阐释，其中所用的数据为

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \tag{3.57}$$

显然，$\mathbf{X}$ 对应三维空间的正四面体. 该单形体以及对应的偏度映射图如图 3.16 所示.



(a)　　(b)

图 3.16　正四面体和它的偏度映射图 (a) 正四面体 (b) 偏度映射图

显然，这个四面体包含 4 个从顶点到底面的高，如图 3.17a 到 3.17d 所示. 此外，还有三个从边到边的公垂线，如图 3.17e 到 3.17g 所示. 通过计算，我们可以知道数据 $\mathbf{X}$ 对应的协偏度张量包含了 7 个特征对，并且在这 7 个特征对中，有 4 个对应了 $\mathbf{X}$ 的偏度极值方向，而剩下的三个则对应 $\mathbf{X}$ 的偏度鞍点方向. 经过验证，可以发现这 4 个偏度极值方向正好与图 3.17a 到 3.17d 中的 4 条高一一对应（平行），这与定理 3.2 中的结论完全一致. 有趣的是，其他三个鞍点方向正好分别与四面体的三条边到边的公垂线对应，如图 3.17e 到 3.17g 所示.

进一步地，除了单形体，对于由对称狄利克雷分布（Dirichlet Distribution）[14]生成的数据（如图 3.18 所示），我们也有如下重要结论[13].

定理 3.3 令 $\mathbf{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_p \end{bmatrix} \in \mathbb{R}^{p\times(p+1)}$ 为 p 维单形体的 $p+1$ 个顶点构成的矩阵，且混合系数矩阵 $\mathbf{D} \in \mathbb{R}^{(p+1)\times n}$ 满足对称狄利克雷分布，那么生成的数据 $\mathbf{X} = \mathbf{V}\mathbf{D} \in \mathbb{R}^{p\times n}$ 偏度极值方向的期望平行于单形体的高.

从图 3.19 可以看出，由服从对称狄利克雷分布的混合系数生成的数据的偏度映射图与单形体的偏度映射图的形状确实是相同的. 定理 3.3 在遥感图像混合像元分析中有重要的理论与实际意义，相关内容将在《矩阵之美 —— 应用篇》中详细介绍.

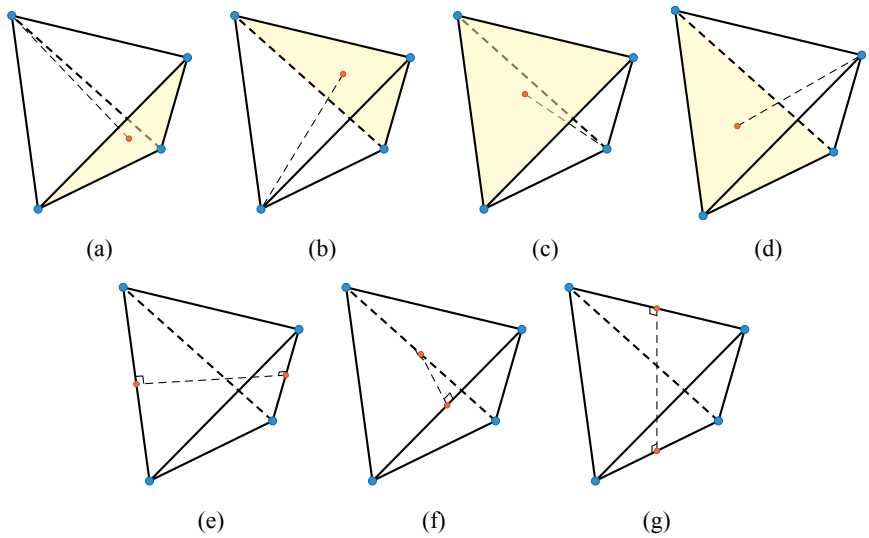


图 3.17　单形体的高与其构造的协偏度张量特征向量的关系示意图. (a) - (d) 4 个局部极大值偏度方向一一对应单形体的 4 条高线；(e) - (g) 三个鞍点方向对应连接不同顶点的两边的公垂线

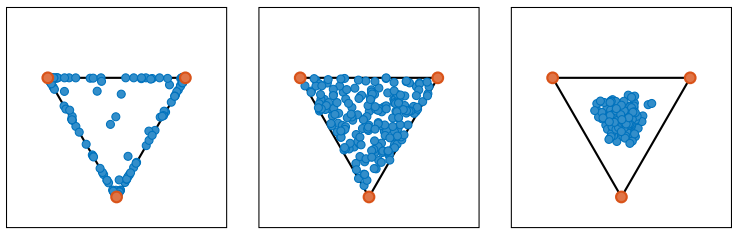


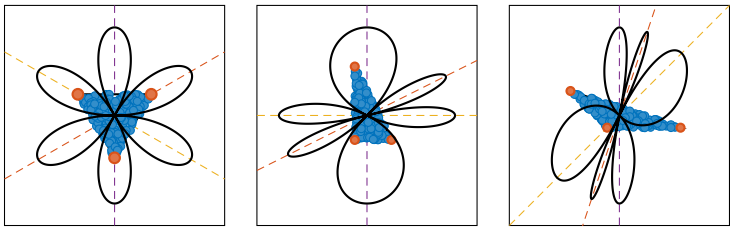
图 3.18　服从不同参数的对称狄利克雷分布的混合系数生成的数据（蓝色为数据点，红色为三角形顶点）

3.6.2 几何解释

借助于单形体的偏度映射图，我们可以尝试给出一般数据偏度映射图的几何解释. 设有白化后的数据 $\mathbf{X}$，对于 p 维空间中的任意单位向量 $\boldsymbol{u}$，$\mathbf{X}$ 在该方向的偏度为 $\mathrm{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X})$. 不妨令 $Y = \boldsymbol{u}^{\mathrm{T}}\mathbf{X}$. 容易验证，$Y$ 是 n 维样本空间的一个单位行向量. 因此有

$$\mathrm{skew}(\boldsymbol{u}^{\mathrm{T}}\mathbf{X}) = \mathrm{skew}(Y) = \mathrm{skew}(Y\mathbf{I}_n) \tag{3.58}$$

其中 $\mathbf{I}_n$ 为 $n \times n$ 大小的单位矩阵. 公式 (3.58) 表明，p 维空间中数据 $\mathbf{X}$ 在 $\boldsymbol{u}$ 方向的偏度等于 n 维空间中数据 $\mathbf{I}_n$ 在 Y 方向的偏度. 显而易见，以单位矩阵 $\mathbf{I}_n$ 的各个列向量



(a)　　(b)　　(c)

图 3.19　服从对称狄利克雷分布的混合系数生成的数据的偏度映射图（蓝色为数据点，红色为三角形顶点）

为顶点可以构成 n 维空间的一个 $n-1$ 维正单形体. 又由于 Y 的取值范围为整个 $\mathbf{X}$ 的行空间，这样一来，我们可以得到如下结论.

定理 3.4　*对于一个白化后的数据 $\mathbf{X} \in \mathbb{R}^{p\times n}$，其偏度映射图为 $\mathbf{X}$ 的行向量所构成的子空间与单位矩阵 $\mathbf{I}_n$ 对应的单形体的偏度映射图的交集.*

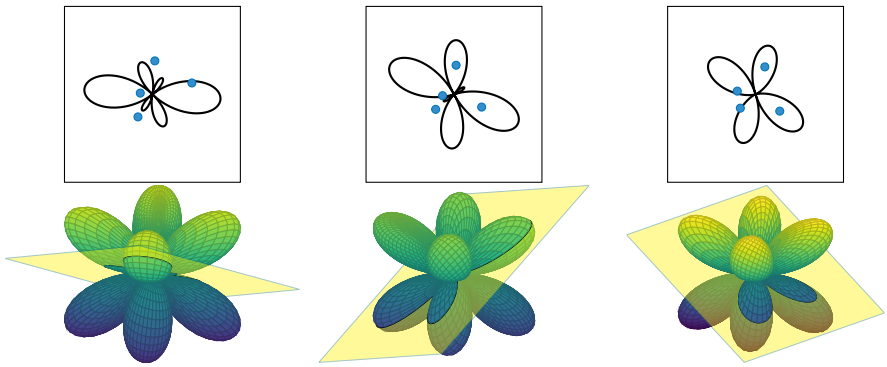
接下来，我们用一个简单的例子进一步来加强对定理 3.4 的理解，当 $n=4$ 时，$\mathbf{I}_n$ 对应四维单位矩阵，即

$$\mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.59}$$

以 $\mathbf{I}_4$ 的各个列向量为顶点可以构成四维空间中的一个正四面体，它的偏度映射图为一个四维空间的三维结构（类似于图 3.16b ）. 对于任意一个 2×4 大小的白化后的数据 $\mathbf{X}$，其偏度映射图为上述正四面体偏度映射图与 $\mathbf{X}$ 的行空间的交集，如图 3.20 所示. 类似地，对于任意一个 2×5 大小的白化后的数据 $\mathbf{X}$，其偏度映射图为四维单形体的偏度映射图与 $\mathbf{X}$ 的行空间的交集，如图 3.21 所示（由于四维单形体的偏度映射图是一个四维结构，因此我们无法展示其全貌，而只能给出它在某个三维空间的投影）.

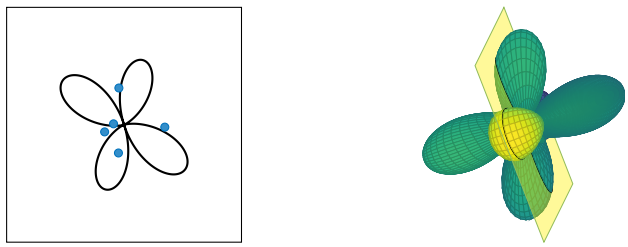
3.7　主偏度分析在应用中的问题

作为主成分分析的高阶版本，主偏度分析展现了 “矩阵” 在处理非对称乃至非高斯数据方面的能力，但在应用中，也存在一些需要注意的问题. 接下来，我们将对其中的收敛问题、噪声问题和精确解问题展开进一步探讨.



(a) (b) (c)

图 3.20 2×4 大小数据的偏度映射图与相应单形体偏度映射图的关系



(a) (b)

图 3.21 2×5 大小数据的偏度映射图与相应单形体偏度映射图的关系 (a) 数据的偏度映射图 (b) 数据所在切平面与相应单形体偏度映射图的交集

3.7.1 收敛问题

在正交约束和非正交约束的主偏度分析中，我们使用了不动点迭代法来获得协偏度张量的特征对. 但是在部分数据上，该方法可能会出现震荡不收敛的现象（图 3.22 中蓝色曲线）. 为了解决这一问题，我们可以借鉴动量梯度下降（Gradient Descent with Momentum）法的思想.

动量梯度下降法是对梯度下降法的改良版本，广泛地使用在各种优化问题中. 从物理的角度来看，梯度下降法是将每次迭代的梯度看作是当前的运动速度，从而控制迭代的方向，如图 3.23 中蓝色曲线所示. 而动量梯度下降法则是将每次迭代的梯度看作是加速度，基于上一次的速度和当前的加速度来计算当前的速度，然后再根据速度来迭代，从而减少了震荡的可能性，加快了收敛速度，如图 3.23 中红色曲线所示.

类似地，在主偏度分析中，我们也可以基于前两次的迭代结果来获得当前的迭代结果，即

$$\boldsymbol{u}^{(k+1)} = \mathcal{S} \times_2 \boldsymbol{u}^{(k)} \times_3 \boldsymbol{u}^{(k-1)}, \tag{3.60}$$

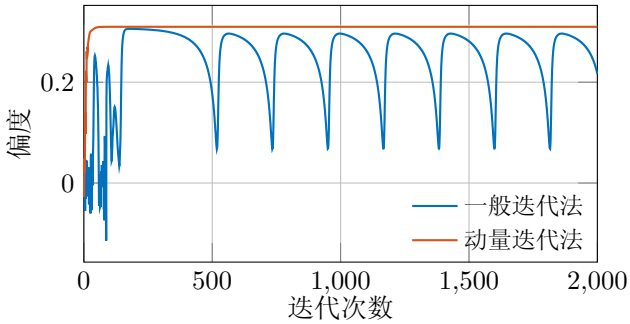


图 3.22　主偏度分析的收敛问题

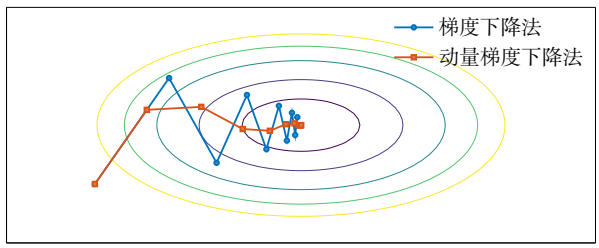


图 3.23　梯度下降法与动量梯度下降法的对比

其中，$\boldsymbol{u}^{(k)}$ 表示第 k 次迭代的结果. 这种迭代方式被称作主偏度分析的动量迭代[15]. 大量实验表明，(3.60) 中的动量梯度迭代法可以显著降低甚至消除算法的不收敛概率，提高主偏度分析的鲁棒性和适用性.

特别地，使用动量迭代法必然可以避免一般迭代法中的如下交替震荡情形

$$\boldsymbol{u}^{(k+1)} = \mathcal{S} \times_2 \boldsymbol{u}^{(k)} \times_3 \boldsymbol{u}^{(k)},$$

$$\boldsymbol{u}^{(k)} = \mathcal{S} \times_2 \boldsymbol{u}^{(k+1)} \times_3 \boldsymbol{u}^{(k+1)},$$

不妨假设使用动量迭代时出现了交替震荡，即

$$\boldsymbol{u}^{(k)} = \mathcal{S} \times_2 \boldsymbol{u}^{(k+1)} \times_3 \boldsymbol{u}^{(k)},$$

$$\boldsymbol{u}^{(k+1)} = \mathcal{S} \times_2 \boldsymbol{u}^{(k)} \times_3 \boldsymbol{u}^{(k+1)}.$$

与此同时，根据 $\mathcal{S}$ 的对称性，我们有

$$\mathcal{S} \times_2 \boldsymbol{u}^{(k+1)} \times_3 \boldsymbol{u}^{(k)} = \mathcal{S} \times_2 \boldsymbol{u}^{(k)} \times_3 \boldsymbol{u}^{(k+1)},$$

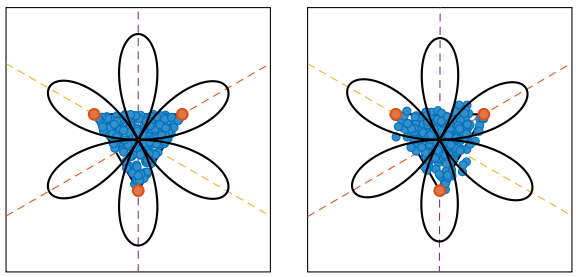
所以 $\boldsymbol{u}^{(k)} = \boldsymbol{u}^{(k+1)}$，即动量迭代必然可以克服交替震荡.

3.7.2 噪声问题

设有白化后的数据 $\mathbf{X}$，对其添加白噪声后得到数据 $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{N}$. 由于白噪声在各个方向上的偏度都为零，因此，添加噪声前后数据的协偏度张量成正比（请读者验证），即

$$\mathcal{S}_{\tilde{\mathbf{X}}} \propto \mathcal{S}_{\mathbf{X}}.$$

换句话说，白化数据的偏度对白噪声是不敏感的. 如图 3.24 所示的符合对称狄利克雷分布的数据，添加噪声前后的偏度极值方向几乎没有变化.



(a) (b)

图 3.24 白噪声对偏度的影响 (a) 添加白噪声前的偏度映射图 (b) 添加白噪声后的偏度映射图

但是需要注意的是，由于计算偏度前需要对数据进行白化，而只有在数据白化后噪声是白的，才能保持噪声对偏度没有影响. 这就要求原始数据中噪声的分布与数据的分布是相同或接近的，而这在实际应用中很难满足. 不过，大量的实验表明，数据中的白噪声对数据偏度分布的影响较小，因此可以认为偏度指标对于白噪声具有一定的鲁棒性.

3.7.3 精确解问题

本章所提及的算法都只能获得张量的部分特征对，且不能保证解的精确性. 而对于张量精确解的个数有定理 3.5 中给出的上界. 例如 $M(3,3) = 7$，这表示三阶三维对称张量最多可以有 7 个实特征对. 不难发现，随着维度 n 的增大，实特征对的上界值也呈指数增加. 此外，即使是两个阶数和维数完全相同的对称张量，它们的精确解的个数也完全可能不同.

定理 3.5 *阶数为 m，维度为 p 的对称张量实特征对的个数至多不超过*[16]

$$M(m,p) = \frac{(m-1)^p - 1}{m-2}.$$

即使定理 3.5 给出了对称张量实特征对个数的上界，但是如何求取这些特征对仍

然是一个难题. 迄今为止，公开发表的文献里面只有两种方法可以得到对称张量的所有精确特征对. 国际上首个可以获取张量全部精确特征对的算法[17]于 2014 年由中国科学院数学所的学者提出，主要基于半正定规划（Semi-Definite Programming，SDP）. 该方法利用拉格朗日函数的一阶梯度信息，通过进一步约束特征值的大小，并利用一些成熟的优化软件包，依次从大到小得到张量所有精确的实特征对. 另外一种算法[18]则主要基于同伦（Homotopic）这一概念，它将张量特征对求解问题视作一个具有 p 个变量, $p+1$ 个等式的非线性方程组（目标系统），接着设计一个相同规模但更加容易求解的非线性方程组（简单系统）. 然后结合目标系统和简单系统构建一个同伦系统，并且该系统可以通过控制参数来在简单系统和目标系统之间连续变换. 由于简单系统的解较容易得到，因此当控制同伦系统逐步从简单系统演变到目标系统的同时，简单系统的解也可以随之迭代为目标系统的解，从而获得张量的所有特征对.

在实际应用中，比如鸡尾酒会问题，并不是所有的特征对都是我们需要的，那么在张量的所有特征对中，我们应该如何选择需要的特征对呢？正如我们在图 3.17 中展示的，特征对可以分为局部极值和鞍点两类. 而在大量的实验中，我们发现通常独立成分对应了局部极值特征对，而鞍点往往是两个或多个成分的混合. 因此，在获得所有精确解后，往往需要利用其他信息（比如二阶导数信息）来筛选出我们需要的特征对.

3.8 小　　结

至此，本章的内容总结为以下 6 条：

1. 主偏度分析是三阶统计分析的标准工具，当数据的分布呈现明显的非对称结构时，可以考虑用主偏度分析对其进行特征提取.
2. 数据的协偏度张量包含了数据的所有三阶统计信息，数据在任意方向的偏度都可以由协偏度张量和表征相应方向的单位向量解析表达.
3. 数据的主偏度分析可以转化为数据协偏度张量的特征值与特征向量分析.
4. 单形体的偏度极值方向与单形体的各个高的方向一一对应，鞍点偏度方向与各个公垂线方向一一对应.
5. 主偏度分析与取偏度指标的 FastICA 效果等价，不过 FastICA 的每次迭代都需要所有数据的参与，而主偏度分析仅需要协偏度张量参与运算.
6. 在几何上，观测数据的偏度映射图对应于相应单形体偏度映射图的特定截面.

参考文献

[1] Shalabh. Theory of Ridge Regression Estimation with Applications[J]. Journal of the Royal Statistical Society Series A: Statistics in Society, 2022, 185(2): 742-743.

[2] FISCHLER Martin-A., BOLLES Robert-C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography [G]//FISCHLER Martin-A., FIRSCHEIN Oscar. Readings in Computer Vision. San Francisco (CA): Morgan Kaufmann, 1987: 726-740.

[3] PEARSON Karl. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space [J]. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901, 2(11): 559-572.

[4] HOTELLING Harold. Analysis of a Complex of Statistical Variables into Principal Components[J]. Journal of Educational Psychology, 1933, 24: 498-520.

[5] HYVÄRINEN Aapo, OJA Erkki. A Fast Fixed-Point Algorithm for Independent Component Analysis[J]. Neural computation, 1997, 9(7): 1483-1492.

[6] CARDOSO Jean-François, SOULOUMIAC Antoine. Blind Beamforming for Non-Gaussian Signals[C]//IEE proceedings F (radar and signal processing): vol. 140: 6. 1993: 362-370.

[7] GENG Xiurui, JI Luyan, SUN Kang. Principal Skewness Analysis: Algorithm and Its Application for Multispectral/Hyperspectral Images Indexing[J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(10): 1821-1825.

[8] LIU Shuangzhe. Matrix Results on the Khatri-Rao and Tracy-Singh Products[J]. Linear Algebra and its Applications, 1999, 289(1): 267-277.

[9] LIM Lek-Heng. Singular Values and Eigenvalues of Tensors: A Variational Approach [C]//1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005. 2005: 129-132.

[10] QI Liqun. Eigenvalues of a Real Supersymmetric Tensor[J]. Journal of Symbolic Computation, 2005, 40(6): 1302-1324.

[11] GENG Xiurui, WANG Lei. NPSA: Nonorthogonal Principal Skewness Analysis[J]. IEEE Transactions on Image Processing, 2020, 29: 6396-6408.

[12] HYVÄRINEN A., OJA E. Independent Component Analysis: Algorithms and Applications[J]. Neural Networks, 2000, 13(4): 411-430.

[13] GENG Xiurui, WANG Lei, ZHU Liangliang, et al. A New Property of the Triangle and Its Application[J]. Unpublished manuscript.

[14] OLKIN Ingram, RUBIN Herman. Multivariate Beta Distributions and Independence Properties of the Wishart Distribution[J]. The Annals of Mathematical Statistics, 1964, 35(1): 261-269.

[15] GENG Xiurui, MENG Lingbo, LI Lin, et al. Momentum Principal Skewness Analysis [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(11): 2262-2266.

[16] CARTWRIGHT Dustin, STURMFELS Bernd. The Number of Eigenvalues of a Tensor [J]. Linear Algebra and its Applications, 2013, 438(2): 942-952.

[17] CUI Chun-Feng, DAI Yu-Hong, NIE Jiawang. All Real Eigenvalues of Symmetric Tensors[J]. SIAM Journal on Matrix Analysis and Applications, 2014, 35(4): 1582-1601.

[18] CHEN Liping, HAN Lixing, ZHOU Liangmin. Computing Tensor Eigenvalues via Homotopy Methods[J]. SIAM Journal on Matrix Analysis and Applications, 2016, 37(1): 290-319.

附录 A 向量范数与矩阵范数

首先给出范数的概念：

定义 A.1（**范数**） 若 V 是数域上的线性空间，在其上定义了一个实值函数 $\|\cdot\|: V \to \mathbb{R}$，满足：

1. 非负性：$\|\boldsymbol{x}\| \geq 0, \|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = 0$
2. 齐次性：$\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$
3. 三角不等式：$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

其中 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 是 V 中的任意矢量，α 为数域上的任意标量. 则称 $\|\cdot\|$ 为线性空间 V 上的**范数**. 特别地，当线性空间的元素为矩阵时，相应的范数也称为**矩阵范数**.

接下来，我们给出常用的向量范数与矩阵范数.

A.1 向量范数

我们以欧氏空间为例，给出 n 维欧氏空间中元素的各种常用范数. 其中，下面用到的向量 $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$, $\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^{\mathrm{T}}$ 均为 n 维欧氏空间中的向量.

1. ℓ_1-范数：

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|.$$

由向量的 ℓ_1-范数（简称 1-范数）可以定义向量之间的**曼哈顿距离**（**Manhattan Distance**）

$$d_{\boldsymbol{xy}} = \|\boldsymbol{x} - \boldsymbol{y}\|_1 = \sum_{i=1}^{n} |x_i - y_i|.$$

图 A.1 给出了二维平面上到原点的曼哈顿距离为 1 的所有的点的几何结构.

2. ℓ_2-范数：

$$\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}}.$$

由向量的 ℓ_2-范数（简称 2-范数）定义可以定义向量之间的**欧氏距离**（**Euclidean Distance**）：

$$d_{\boldsymbol{xy}} = \|\boldsymbol{x} - \boldsymbol{y}\| = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}.$$

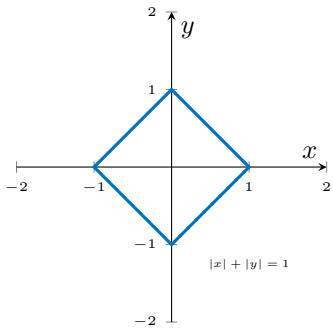


图 A.1　在二维平面上到原点的曼哈顿距离为 1 的点集的几何结构

图 A.2 给出了二维平面上到原点的欧氏距离为 1 的所有的点的几何结构.

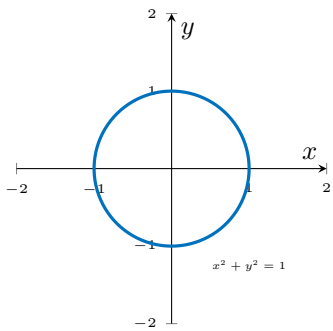


图 A.2　在二维平面上到原点的欧氏距离为 1 的点集的几何结构

3. ℓ_p-范数：

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}.$$

由向量的 ℓ_p-范数（简称 p 范数）可以定义向量之间的**明斯基距离（Minkowski Distance）**：

$$\boldsymbol{d}_{xy} = \|\boldsymbol{x} - \boldsymbol{y}\|_p = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}}.$$

图 A.3 分别给出了 $p = 0.5$ 和 $p = 4$ 时二维平面上到原点的明斯基距离为 1 的所有的点的几何结构.

4. ∞-范数：

$$\|\boldsymbol{x}\|_\infty = \max_{1 \le i \le n} |x_i|.$$

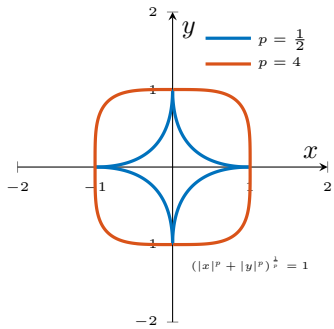


图 A.3　在二维平面上到原点明斯基距离为 1 的点集的几何结构

由向量的 ∞-范数可以定义向量之间的**切比雪夫距离**（**Chebyshev Distance**）：

$$d_{\boldsymbol{xy}} = \|\boldsymbol{x} - \boldsymbol{y}\|_\infty = \max_{1 \leq i \leq n} |x_i - y_i|.$$

∞-范数可以看作是 ℓ_p-范数在 p 趋于无穷时的极限情形，即有

$$\|\boldsymbol{x}\|_\infty = \lim_{p \to \infty} \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

相应地，到原点的切比雪夫距离为 1 的图形也延续上面几个图形的规律，为正方形结构（图 A.4 ）.

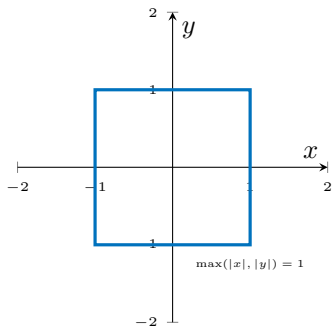


图 A.4　在二维平面上到原点的切比雪夫距离为 1 的点集的几何结构

5. ℓ_0-范数：

$$\|\boldsymbol{x}\|_0 = \sum_{i=1}^{n} \mathbb{I}(x_i),$$

其中，$\mathbb{I}(\cdot)$ 为指示函数，有如下表达式

$$\mathbb{I}(x_i)=\begin{cases}1, & x_i\neq 0\\ 0, & x_i=0\end{cases}.$$

简而言之，向量的 ℓ_0-范数（简称 0-范数）等于向量中非零元素的个数. 图 A.5 给出了二维平面上两个分量都不大于 1 且 ℓ_0-范数为 1 的所有的点的几何结构. 需要注意的是，向量的 ℓ_0-范数并不是普通意义的范数，因为它并不满足范数齐次性的性质.

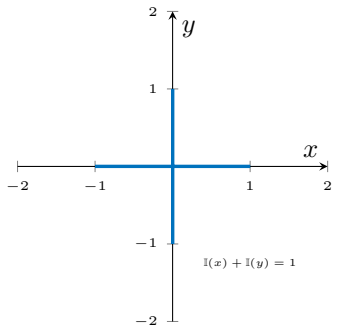


图 A.5　二维平面上两个分量都不大于 1 且 ℓ_0-范数为 1 的点集的几何结构

A.2 矩阵范数

常用的矩阵范数（其中 $\mathbf{A}$ 为 $m\times n$ 实矩阵）如下.

1. 弗罗贝尼乌斯范数（Frobenius Norm）

矩阵 $\mathbf{A}$ 的弗罗贝尼乌斯范数（简称 F-范数）是最常用的矩阵范数，公式如下

$$\|\mathbf{A}\|_{\mathrm{F}}=\sqrt{\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{A})}=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}^2\right)^{\frac{1}{2}}.$$

事实上，它相当于将矩阵 $\mathbf{A}$ 展成向量后的 ℓ_2-范数，即 $\|\mathbf{A}\|_{\mathrm{F}}=\|\,\mathrm{vec}(\mathbf{A})\|_2$.

2. p-范数：

矩阵 $\mathbf{A}$ 的 p-范数定义为

$$\|\mathbf{A}\|_p=\max_{\boldsymbol{x}\neq 0}\frac{\|\mathbf{A}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p},$$

其中，$\|\boldsymbol{x}\|_p$ 是向量 $\boldsymbol{x}$ 的 ℓ_p-范数. 需要注意的是，矩阵 $\mathbf{A}$ 的 2-范数和它的 F-范数一般情况下并不等价.

3. 行和范数（Row-sum Norm）

$$\|\mathbf{A}\|_{row} = \max_{1\leq i\leq m}\left\{\sum_{j=1}^{n}|a_{ij}|\right\}.$$

4. 列和范数（Column-sum Norm）

$$\|\mathbf{A}\|_{col} = \max_{1\leq j\leq n}\left\{\sum_{i=1}^{m}|a_{ij}|\right\}.$$

5. 谱范数（Spectrum Norm）

$$\|\mathbf{A}\|_{spec} = \sigma_{max} = \sqrt{\lambda_{max}},$$

其中 σ_{max} 是矩阵 $\mathbf{A}$ 的最大奇异值，即 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 的最大特征值 λ_{max} 的平方根.

6. 马哈拉诺比斯范数（Mahalanobis Norm）

$$\|\mathbf{A}\|_{\mathbf{\Omega}} = \sqrt{\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{\Omega}\mathbf{A})},$$

其中 $\mathbf{\Omega}$ 为正定矩阵.

7. 核范数（Nuclear Norm）

$$\|\mathbf{A}\|_{*} = \mathrm{tr}(\sqrt{\mathbf{A}^{\mathrm{T}}\mathbf{A}}).$$

经过简单的验证可知，矩阵的核范数等于矩阵的所有奇异值之和.

附录 B 矩阵微积分

B.1 实值标量函数相对于实向量的梯度

设 $f(\boldsymbol{x})$ 是一个以向量 $\boldsymbol{x}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}^{\mathrm{T}}$ 为自变量的实标量函数，则该函数相对于自变量的梯度定义为

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=\begin{bmatrix}\frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n}\end{bmatrix}^{\mathrm{T}}.$$

$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}$ 也可以记作 $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$，其中 $\nabla_{\boldsymbol{x}}=\begin{bmatrix}\frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \cdots & \frac{\partial}{\partial x_n}\end{bmatrix}^{\mathrm{T}}$ 为梯度算子.

类似地，也可以给出函数 $f(\boldsymbol{x})$ 相对于行向量 $\boldsymbol{x}^{\mathrm{T}}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}$ 的梯度为

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}=\begin{bmatrix}\frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n}\end{bmatrix}=\nabla_{\boldsymbol{x}^{\mathrm{T}}} f(\boldsymbol{x}).$$

从上面的式子可以看出：

1. 以列（行）向量为自变量的实标量函数，其对于自变量的梯度仍然为一个大小相同的列（行）向量.
2. 梯度的每个分量代表着函数在该分量所在坐标方向上的变化率.

实值标量函数相对于实向量的梯度，满足如下几个规则（请读者自行推导）：

1. 线性法则：若 $f(\boldsymbol{x})$ 和 $g(\boldsymbol{x})$ 分别是向量 $\boldsymbol{x}$ 的实值标量函数，c_1 和 c_2 为实常数，则
$$\frac{\partial(c_1 f(\boldsymbol{x})+c_2 g(\boldsymbol{x}))}{\partial \boldsymbol{x}}=c_1\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}+c_2\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}.$$
2. 乘积法则（标量版）：若 $f(\boldsymbol{x})$ 和 $g(\boldsymbol{x})$ 分别是向量 $\boldsymbol{x}$ 的实值标量函数，则
$$\frac{\partial(f(\boldsymbol{x})g(\boldsymbol{x}))}{\partial \boldsymbol{x}}=g(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}+f(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}.$$
3. 乘积法则（向量版）：若 $\boldsymbol{f}(\boldsymbol{x}), \boldsymbol{g}(\boldsymbol{x})$ 为实列向量函数，则
$$\frac{\partial(\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{g}(\boldsymbol{x}))}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\boldsymbol{g}(\boldsymbol{x})+\frac{\partial \boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}).$$
4. 商法则：若 $f(\boldsymbol{x})$ 和 $g(\boldsymbol{x})$ 分别是向量 $\boldsymbol{x}$ 的实值标量函数，且 $g(\boldsymbol{x})\neq 0$，则
$$\frac{\partial(f(\boldsymbol{x})/g(\boldsymbol{x}))}{\partial \boldsymbol{x}}=\frac{1}{g^2(\boldsymbol{x})}\left(g(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}-f(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}\right).$$
5. 链式法则：若 $\boldsymbol{g}(\boldsymbol{x})$ 是实列向量函数，则
$$\frac{\partial(f(\boldsymbol{g}(\boldsymbol{x})))}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\frac{\partial f(\boldsymbol{g})}{\partial \boldsymbol{g}}.$$

例 B.1 试求实值标量函数 $f(\boldsymbol{x})=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$ 相对于自变量 $\boldsymbol{x}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}^{\mathrm{T}}$ 的梯

度.

解 由于 $f(\boldsymbol{x})=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}=a_1x_1+\cdots+a_nx_n$，所以

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}}=\begin{bmatrix}\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial x_1} & \frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial x_2} & \cdots & \frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial x_n}\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}a_1 & a_2 & \cdots & a_n\end{bmatrix}^{\mathrm{T}}=\boldsymbol{a}.$$

同理，可得

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}=\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}^{\mathrm{T}}}=\begin{bmatrix}\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial x_1} & \frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial x_2} & \cdots & \frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial x_n}\end{bmatrix}=\begin{bmatrix}a_1 & a_2 & \cdots & a_n\end{bmatrix}=\boldsymbol{a}^{\mathrm{T}}.$$

例 B.2 试求实值标量函数 $f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{x}$ 相对于自变量 $\boldsymbol{x}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}^{\mathrm{T}}$ 的梯度.

解 首先有

$$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{x}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j,$$

其中，$f(\boldsymbol{x})$ 中含有 x_k 的项为

$$f_{x_k}(\boldsymbol{x})=\sum_{j=1,j\neq k}^{n}a_{kj}x_kx_j+\sum_{i=1,i\neq k}^{n}a_{ik}x_ix_k+a_{kk}x_kx_k.$$

$f_{x_k}(\boldsymbol{x})$ 相对于 x_k 的偏导数为

$$\begin{aligned}\frac{\partial f_{x_k}(\boldsymbol{x})}{\partial x_k}&=\sum_{j=1,j\neq k}^{n}a_{kj}x_j+\sum_{i=1,i\neq k}^{n}a_{ik}x_i+2a_{kk}x_k=\sum_{j=1}^{n}a_{kj}x_j+\sum_{i=1}^{n}a_{ik}x_i\\&=\mathbf{A}(k,:)\boldsymbol{x}+\mathbf{A}^{\mathrm{T}}(k,:)\boldsymbol{x}.\end{aligned}$$

因此，$f(\boldsymbol{x})$ 相对于 $\boldsymbol{x}$ 的梯度为

$$\begin{aligned}\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}&=\begin{bmatrix}\frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n}\end{bmatrix}^{\mathrm{T}}\\&=\begin{bmatrix}\frac{\partial f_{x_1}(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_{x_2}(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_{x_n}(\boldsymbol{x})}{\partial x_n}\end{bmatrix}^{\mathrm{T}}\\&=\mathbf{A}\boldsymbol{x}+\mathbf{A}^{\mathrm{T}}\boldsymbol{x}.\end{aligned}$$

B.2 实值向量函数相对于实向量的梯度

m 维列向量函数 $\boldsymbol{f}(\boldsymbol{x})=\begin{bmatrix}f_1(\boldsymbol{x}) & f_2(\boldsymbol{x}) & \cdots & f_m(\boldsymbol{x})\end{bmatrix}^{\mathrm{T}}$ 相对于 n 维行向量 $\boldsymbol{x}^{\mathrm{T}}$ 的梯度为一个 $m\times n$ 的矩阵，即

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}=\begin{bmatrix}\frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_1(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n}\\ \frac{\partial f_2(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_2(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\boldsymbol{x})}{\partial x_n}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_m(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n}\end{bmatrix}.$$

m 维列向量函数相对于列向量 $\boldsymbol{x}$ 的梯度，将是一个更高的列向量，即

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_n} \end{bmatrix} = \mathrm{vec}\left(\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}\right).$$

m 维行向量函数 $\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})$ 相对于 n 维列向量 $\boldsymbol{x}$ 的梯度为一个 $n \times m$ 的矩阵，即

$$\frac{\partial \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_2(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f_1(\boldsymbol{x})}{\partial x_2} & \frac{\partial f_2(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} & \frac{\partial f_2(\boldsymbol{x})}{\partial x_n} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix}.$$

m 维行向量函数 $\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})$ 相对于行向量 $\boldsymbol{x}^{\mathrm{T}}$ 的梯度，将是一个更长的行向量，即

$$\frac{\partial \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} = \left(\mathrm{vec}\left(\frac{\partial \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)\right)^{\mathrm{T}}.$$

例 B.3 试求实向量函数 $\boldsymbol{f}(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x}$ 相对于向量 $\boldsymbol{x}^{\mathrm{T}}$ 的梯度.

解 方法 1：由于

$$\boldsymbol{f}(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x} = \begin{bmatrix} \mathbf{A}(1,:)\boldsymbol{x} \\ \mathbf{A}(2,:)\boldsymbol{x} \\ \vdots \\ \mathbf{A}(n,:)\boldsymbol{x} \end{bmatrix},$$

所以

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial (\mathbf{A}(1,:)\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} \\ \frac{\partial (\mathbf{A}(2,:)\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} \\ \vdots \\ \frac{\partial (\mathbf{A}(n,:)\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}(1,:) \\ \mathbf{A}(2,:) \\ \vdots \\ \mathbf{A}(n,:) \end{bmatrix} = \mathbf{A}.$$

方法 2：由于

$$\boldsymbol{f}(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x} = x_1\mathbf{A}(:,1) + x_2\mathbf{A}(:,2) + \cdots + x_n\mathbf{A}(:,n),$$

所以

$$\begin{aligned} \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} &= \begin{bmatrix} \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_1} & \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}(:,1) & \mathbf{A}(:,2) & \cdots & \mathbf{A}(:,n) \end{bmatrix} = \mathbf{A}. \end{aligned}$$

同理，可以得到

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \mathrm{vec}(\mathbf{A}),$$

$$\frac{\partial \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left(\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}\right)^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}},$$

$$\frac{\partial \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} = \left(\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} = \mathrm{vec}(\mathbf{A})^{\mathrm{T}}.$$

B.3 实值函数相对于实矩阵的梯度

实值函数 $f(\mathbf{A})$ 相对于其自变量 $m \times n$ 矩阵 $\mathbf{A} = (a_{ij})$ 的梯度仍然是一个 $m \times n$ 的矩阵，即

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial f(\mathbf{A})}{\partial a_{11}} & \frac{\partial f(\mathbf{A})}{\partial a_{12}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{1n}} \\ \frac{\partial f(\mathbf{A})}{\partial a_{21}} & \frac{\partial f(\mathbf{A})}{\partial a_{22}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{A})}{\partial a_{m1}} & \frac{\partial f(\mathbf{A})}{\partial a_{m2}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{mn}} \end{bmatrix} = \nabla_{\mathbf{A}} f(\mathbf{A}).$$

事实上，实值函数相对于矩阵的梯度与实值函数相对于向量的梯度并没有本质的区别，它们之间可以通过下面的公式相互转化

$$\nabla_{\mathbf{A}} f(\mathbf{A}) = \mathrm{unvec}\left(\nabla_{\mathrm{vec}(\mathbf{A})} f(\mathrm{vec}(\mathbf{A}))\right),$$

其中 unvec(·) 是 vec(·) 的逆操作，它将一个向量转化为一个矩阵.

例 B.4 试求函数 $f(\mathbf{A}) = \boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y}$ 相对于矩阵 $\mathbf{A}$ 的梯度.

解 方法 1：逐元素求导，由于

$$f(\mathbf{A}) = \boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_i y_j.$$

因此

$$\frac{\partial f(\mathbf{A})}{\partial a_{ij}} = x_i y_j,$$

所以

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial f(\mathbf{A})}{\partial a_{11}} & \frac{\partial f(\mathbf{A})}{\partial a_{12}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{1n}} \\ \frac{\partial f(\mathbf{A})}{\partial a_{21}} & \frac{\partial f(\mathbf{A})}{\partial a_{22}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{A})}{\partial a_{m1}} & \frac{\partial f(\mathbf{A})}{\partial a_{m2}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{mn}} \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}.$$

方法 2：向量求导法，由于

$$f(\mathbf{A}) = \boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y} = \mathrm{vec}(\mathbf{A})^{\mathrm{T}}(\boldsymbol{y} \otimes \boldsymbol{x}).$$

因此

$$\frac{\partial f(\mathrm{vec}(\mathbf{A}))}{\partial \,\mathrm{vec}(\mathbf{A})} = \boldsymbol{y} \otimes \boldsymbol{x},$$

又因为

$$\boldsymbol{y} \otimes \boldsymbol{x} = \mathrm{vec}(\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}),$$

所以

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = \mathrm{unvec}(\nabla_{\mathrm{vec}(\mathbf{A})} f(\mathrm{vec}(\mathbf{A}))) = \mathrm{unvec}(\boldsymbol{y} \otimes \boldsymbol{x}) = \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}.$$

B.4 矩阵微分

对于一个以向量 $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$ 为自变量的实值函数 $f(\boldsymbol{x})$，其微分公式如下

$$\mathrm{d}f(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x})}{\partial x_i} \mathrm{d}x_i.$$

对于一个 $m \times n$ 矩阵 $\mathbf{X} = (x_{ij})$ 为自变量的实值函数 $f(\mathbf{X})$，其微分公式为

$$\mathrm{d}f(\mathbf{X}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\partial f(\mathbf{X})}{\partial x_{ij}} \mathrm{d}x_{ij}. \tag{B.1}$$

通过简单的计算，可以验证 (B.1) 可以重新表示为如下形式：

$$\mathrm{d}f(\mathbf{X}) = \mathrm{tr}\left(\left(\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \right)^{\mathrm{T}} \mathrm{d}\mathbf{X} \right), \tag{B.2}$$

其中

$$\mathrm{d}\mathbf{X}=\begin{bmatrix}\mathrm{d}x_{11} & \mathrm{d}x_{12} & \cdots & \mathrm{d}x_{1n}\\ \mathrm{d}x_{21} & \mathrm{d}x_{22} & \cdots & \mathrm{d}x_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \mathrm{d}x_{m1} & \mathrm{d}x_{m2} & \cdots & \mathrm{d}x_{mn}\end{bmatrix}.$$

需要指出的是，由于向量是矩阵的特例，所以当自变量 $\mathbf{X}$ 为向量时，(B.2) 仍然成立.

矩阵微分有如下常用的性质（请读者尝试证明）：

1. 常数矩阵的微分矩阵为零矩阵，即
$$\mathrm{d}(\mathbf{A})=\mathbf{0}.$$
2. 矩阵转置的微分矩阵等于原矩阵的微分矩阵的转置，即
$$\mathrm{d}(\mathbf{X}^{\mathrm{T}})=(\mathrm{d}\mathbf{X})^{\mathrm{T}}.$$
3. 矩阵微分算子为线性算子，即对于任意常数 a,b 和相同大小的矩阵函数 $\mathbf{X},\mathbf{Y}$，都有
$$\mathrm{d}(a\mathbf{X}+b\mathbf{Y})=a\mathrm{d}\mathbf{X}+b\mathrm{d}\mathbf{Y}.$$
4. 矩阵函数的迹的微分等于其微分的迹，即
$$\mathrm{d}(\mathrm{tr}(\mathbf{X}))=\mathrm{tr}(\mathrm{d}\mathbf{X}).$$
5. 两个矩阵函数的乘积的微分满足如下规则
$$\mathrm{d}(\mathbf{X}\mathbf{Y})=(\mathrm{d}\mathbf{X})\mathbf{Y}+\mathbf{X}\mathrm{d}\mathbf{Y}.$$
6. 矩阵函数的 Kronecker 积的微分满足如下规则
$$\mathrm{d}(\mathbf{X}\otimes\mathbf{Y})=(\mathrm{d}\mathbf{X})\otimes\mathbf{Y}+\mathbf{X}\otimes\mathrm{d}\mathbf{Y}.$$
7. 矩阵函数的逆的微分满足如下规则
$$\mathrm{d}(\mathbf{X}^{-1})=-\mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1}.$$

例 B.5 试用矩阵微分法求函数 $f(\mathbf{A})=\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y}$ 相对于矩阵 $\mathbf{A}$ 的梯度.

解 由于

$$\mathrm{d}f(\mathbf{A})=\mathrm{d}(\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y})=\mathrm{d}(\mathrm{tr}(\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y}))=\mathrm{tr}(\mathrm{d}(\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y}))=\mathrm{tr}(\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}}\mathrm{d}\mathbf{A}),$$

基于 (B.2)，有

$$\left(\frac{\partial f(\mathbf{A})}{\partial\mathbf{A}}\right)^{\mathrm{T}}=\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}},$$

因此

$$\frac{\partial f(\mathbf{A})}{\partial\mathbf{A}}=(\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}.$$

例 B.6 试用矩阵微分法求实值标量函数 $f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{x}$ 相对于向量 $\boldsymbol{x}$ 的梯度.

解 由于

$$\begin{aligned}\mathrm{d}f(\boldsymbol{x}) &= \mathrm{d}(\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{x}) = \mathrm{d}(\mathrm{tr}(\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{x})) = \mathrm{tr}(\mathrm{d}(\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{x})) = \mathrm{tr}((\mathrm{d}\boldsymbol{x}^{\mathrm{T}})\mathbf{A}\boldsymbol{x} + \boldsymbol{x}^{\mathrm{T}}\mathbf{A}\mathrm{d}\boldsymbol{x}) \\ &= \mathrm{tr}(\boldsymbol{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathrm{d}\boldsymbol{x}) + \mathrm{tr}(\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\mathrm{d}\boldsymbol{x}) = \mathrm{tr}((\boldsymbol{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} + \boldsymbol{x}^{\mathrm{T}}\mathbf{A})\mathrm{d}\boldsymbol{x}),\end{aligned}$$

基于 (B.2)，有

$$\left(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} + \boldsymbol{x}^{\mathrm{T}}\mathbf{A},$$

因此

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = (\boldsymbol{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} + \boldsymbol{x}^{\mathrm{T}}\mathbf{A})^{\mathrm{T}} = \mathbf{A}\boldsymbol{x} + \mathbf{A}^{\mathrm{T}}\boldsymbol{x}.$$

B.5 迹函数的梯度矩阵

对于一个 $n \times n$ 的实矩阵 $\mathbf{X} = (x_{ij})$，它的迹定义为它的对角元素之和，即

$$\mathrm{tr}(\mathbf{X}) = \sum_{i=1}^{n} x_{ii}.$$

显然，矩阵的迹可以认为是以矩阵为自变量的实值函数. 由于

$$\frac{\partial \,\mathrm{tr}(\mathbf{X})}{\partial x_{ij}} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}.$$

因此，矩阵的迹函数相对于该矩阵的梯度为单位矩阵，即

$$\frac{\partial \,\mathrm{tr}(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{I}.$$

关于矩阵的迹，上面的例 B.5 和 B.6 的推导过程中都用到了一个重要的性质：

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}),$$

即，矩阵乘积的迹与乘积顺序无关.

例 B.7 试求 $f(\mathbf{X}) = \mathrm{tr}(\mathbf{AXB})$ 相对于矩阵 $\mathbf{X}$ 的梯度.

解 由于

$$\mathrm{d}f(\mathbf{X}) = \mathrm{d}(\mathrm{tr}(\mathbf{AXB})) = \mathrm{tr}(\mathrm{d}(\mathbf{AXB})) = \mathrm{tr}(\mathbf{A}(\mathrm{d}\mathbf{X})\mathbf{B}) = \mathrm{tr}(\mathbf{BA}(\mathrm{d}\mathbf{X})),$$

基于 (B.2)，我们有

$$\left(\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}\right)^{\mathrm{T}} = \mathbf{BA},$$

因此

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = (\mathbf{BA})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}.$$

例 B.8 试求 $f(\mathbf{X}) = \mathrm{tr}(\mathbf{AX}^{-1}\mathbf{B})$ 相对于矩阵 $\mathbf{X}$ 的梯度.

解 由于

$$\begin{aligned}\mathrm{d}f(\mathbf{X}) &= \mathrm{d}(\mathrm{tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})) = \mathrm{tr}(\mathrm{d}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})) \\ &= \mathrm{tr}(\mathbf{A}(\mathrm{d}\mathbf{X}^{-1})\mathbf{B}) = \mathrm{tr}(-\mathbf{A}\mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1}\mathbf{B}) \\ &= \mathrm{tr}(-\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1}\mathrm{d}\mathbf{X}),\end{aligned}$$

基于 (B.2)，我们有

$$\left(\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}\right)^{\mathrm{T}} = -\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1},$$

因此

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = (-\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^{\mathrm{T}} = -\mathbf{X}^{-\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{X}^{-\mathrm{T}}.$$

例 B.9 试求 $f(\mathbf{X}) = \mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ 相对于矩阵 $\mathbf{X}$ 的梯度.

解 *方法 1：矩阵微分法，由于*

$$\begin{aligned}\mathrm{d}f(\mathbf{X}) &= \mathrm{d}(\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{X})) = \mathrm{tr}(\mathrm{d}(\mathbf{X}^{\mathrm{T}}\mathbf{X})) = \mathrm{tr}((\mathrm{d}\mathbf{X}^{\mathrm{T}})\mathbf{X} + \mathbf{X}^{\mathrm{T}}\mathrm{d}\mathbf{X}) \\ &= \mathrm{tr}((\mathrm{d}\mathbf{X}^{\mathrm{T}})\mathbf{X}) + \mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathrm{d}\mathbf{X}) = \mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathrm{d}\mathbf{X}) + \mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathrm{d}\mathbf{X}) = \mathrm{tr}(2\mathbf{X}^{\mathrm{T}}\mathrm{d}\mathbf{X}),\end{aligned}$$

基于 (B.2)，我们有

$$\left(\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}\right)^{\mathrm{T}} = 2\mathbf{X}^{\mathrm{T}},$$

因此

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}.$$

方法 2：向量求导法，由于

$$\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{X}) = \|\mathbf{X}\|_{\mathrm{F}}^2 = \mathrm{vec}(\mathbf{X})^{\mathrm{T}}\,\mathrm{vec}(\mathbf{X}),$$

因此

$$\frac{\partial\,\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}} = \mathrm{unvec}\left(\frac{\partial(\mathrm{vec}(\mathbf{X})^{\mathrm{T}}\,\mathrm{vec}(\mathbf{X}))}{\partial\,\mathrm{vec}(\mathbf{X})}\right) = \mathrm{unvec}(2\,\mathrm{vec}(\mathbf{X})) = 2\mathbf{X}.$$

从上面的例子可以看出，(B.2) 是矩阵迹函数自变量求导的基本工具. 我们常见的矩阵迹函数的梯度基本都可以基于 (B.2) 按照上述例子中的套路进行求解.

B.6 行列式的梯度矩阵

对于一个 $n \times n$ 的实满秩矩阵 $\mathbf{X} = (x_{ij})$，其行列式 $|\mathbf{X}|$ 也是一个以矩阵为自变量的实值函数. 为了计算该函数相对于矩阵的梯度，我们可以将行列式按第 i 行展开或者按第 j 列展开，对应的公式分别为

$$|\mathbf{X}| = \sum_{j=1}^{n} x_{ij}c_{ij}, \quad |\mathbf{X}| = \sum_{i=1}^{n} x_{ij}c_{ij},$$

其中 $\mathbf{C}=(c_{ij})$ 为矩阵 $\mathbf{X}$ 的代数余子式矩阵.

无论采用哪种行列式展开方式，都可以得到 $|\mathbf{X}|$ 相对于变量 x_{ij} 的梯度为

$$\frac{\partial|\mathbf{X}|}{\partial x_{ij}}=c_{ij}.$$

因此，$|\mathbf{X}|$ 相对于 $\mathbf{X}$ 的梯度为

$$\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}}=\mathbf{C}.$$

又因为矩阵的代数余子式矩阵是其伴随矩阵的转置，即 $\mathbf{C}=(\mathbf{X}^*)^{\mathrm{T}}$，且矩阵的伴随矩阵又可表示为矩阵的行列式与矩阵的逆的乘积，即 $\mathbf{X}^*=|\mathbf{X}|\mathbf{X}^{-1}$，所以 $|\mathbf{X}|$ 相对于 $\mathbf{X}$ 的梯度最终可以表示为

$$\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}}=|\mathbf{X}|\mathbf{X}^{-\mathrm{T}}. \tag{B.3}$$

基于 (B.2) 和 (B.3)，我们可以得到矩阵行列式的微分公式如下：

$$\mathrm{d}|\mathbf{X}|=\mathrm{tr}(|\mathbf{X}|\mathbf{X}^{-1}\mathrm{d}\mathbf{X}). \tag{B.4}$$

例 B.10 试求 $f(\mathbf{X})=|\mathbf{AXB}|$ 相对于矩阵 $\mathbf{X}$ 的梯度.

解 基于 (B.4)，我们有

$$\begin{aligned}\mathrm{d}f(\mathbf{X})&=\mathrm{d}|\mathbf{AXB}|=\mathrm{tr}(|\mathbf{AXB}|(\mathbf{AXB})^{-1}\mathrm{d}(\mathbf{AXB}))\\&=\mathrm{tr}(|\mathbf{AXB}|(\mathbf{AXB})^{-1}\mathbf{A}(\mathrm{d}\mathbf{X})\mathbf{B})\\&=\mathrm{tr}(|\mathbf{AXB}|\mathbf{B}(\mathbf{AXB})^{-1}\mathbf{A}\mathrm{d}\mathbf{X}).\end{aligned}$$

因此，$f(\mathbf{X})=|\mathbf{AXB}|$ 相对于矩阵 $\mathbf{X}$ 的梯度为

$$\frac{\partial|\mathbf{AXB}|}{\partial \mathbf{X}}=(|\mathbf{AXB}|\mathbf{B}(\mathbf{AXB})^{-1}\mathbf{A})^{\mathrm{T}}=|\mathbf{AXB}|\mathbf{A}^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}.$$

例 B.11 试求 $f(\mathbf{X})=|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|$ 相对于矩阵 $\mathbf{X}$ 的梯度.

解 基于 (B.4)，我们有

$$\begin{aligned}\mathrm{d}f(\mathbf{X})&=\mathrm{d}|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|=\mathrm{tr}(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|(\mathbf{X}^{\mathrm{T}}\mathbf{AX})^{-1}\mathrm{d}(\mathbf{X}^{\mathrm{T}}\mathbf{AX}))\\&=\mathrm{tr}(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|(\mathbf{X}^{\mathrm{T}}\mathbf{AX})^{-1}((\mathrm{d}\mathbf{X}^{\mathrm{T}})\mathbf{AX}+\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathrm{d}\mathbf{X}))\\&=\mathrm{tr}(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|(\mathbf{X}^{\mathrm{T}}\mathbf{AX})^{-1}(\mathrm{d}\mathbf{X}^{\mathrm{T}})\mathbf{AX})+\mathrm{tr}(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|(\mathbf{X}^{\mathrm{T}}\mathbf{AX})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathrm{d}\mathbf{X})\\&=\mathrm{tr}(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}(\mathrm{d}\mathbf{X})(\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{X})^{-1})+\mathrm{tr}(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|(\mathbf{X}^{\mathrm{T}}\mathbf{AX})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathrm{d}\mathbf{X})\\&=\mathrm{tr}(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|(\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathrm{d}\mathbf{X})+\mathrm{tr}(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|(\mathbf{X}^{\mathrm{T}}\mathbf{AX})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathrm{d}\mathbf{X})\\&=\mathrm{tr}(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|((\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}+(\mathbf{X}^{\mathrm{T}}\mathbf{AX})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A})\mathrm{d}\mathbf{X}).\end{aligned}$$

因此，$f(\mathbf{X})=|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|$ 相对于矩阵 $\mathbf{X}$ 的梯度为

$$\begin{aligned}\frac{\partial|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|}{\partial \mathbf{X}}&=(|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|((\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}+(\mathbf{X}^{\mathrm{T}}\mathbf{AX})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A}))^{\mathrm{T}}\\&=|\mathbf{X}^{\mathrm{T}}\mathbf{AX}|(\mathbf{AX}(\mathbf{X}^{\mathrm{T}}\mathbf{AX})^{-1}+\mathbf{A}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{X})^{-1}).\end{aligned}$$

当矩阵 $\mathbf{A}$ 为单位矩阵且 $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 可逆时，上式退化为

$$\frac{\partial|\mathbf{X}^{\mathrm{T}}\mathbf{X}|}{\partial\mathbf{X}} = 2|\mathbf{X}^{\mathrm{T}}\mathbf{X}|\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}.$$

同理可得，当 $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ 可逆时，我们有

$$\frac{\partial|\mathbf{X}\mathbf{X}^{\mathrm{T}}|}{\partial\mathbf{X}} = 2|\mathbf{X}\mathbf{X}^{\mathrm{T}}|(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X}.$$

B.7 黑塞矩阵

实值函数 $f(\boldsymbol{x})$ 相对于向量 $\boldsymbol{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^{\mathrm{T}}$ 的二阶偏导数称为该函数的黑塞矩阵，定义为

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial\boldsymbol{x}\partial\boldsymbol{x}^{\mathrm{T}}} = \frac{\partial}{\partial\boldsymbol{x}^{\mathrm{T}}}\left(\frac{\partial f(\boldsymbol{x})}{\partial\boldsymbol{x}}\right) = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2\partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2\partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n\partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n\partial x_n} \end{bmatrix} = \nabla^2_{\boldsymbol{x}} f(\boldsymbol{x}).$$

函数的黑塞矩阵一般记为 $\mathbf{H}$. 容易验证，函数的黑塞矩阵 $\mathbf{H}$ 是一个实对称矩阵.

类似地，我们可以给出实值函数 $f(\boldsymbol{x})$ 相对于向量 $\boldsymbol{x}$ 的三阶偏导数为

$$\nabla^3_{\boldsymbol{x}} f(\boldsymbol{x}) = \left(\frac{\partial^3 f(\boldsymbol{x})}{\partial x_i\partial x_j\partial x_k}\right).$$

此时，$\nabla^3_{\boldsymbol{x}} f(\boldsymbol{x})$ 将不再是一个矩阵，而是一个三阶张量，$\frac{\partial^3 f(\boldsymbol{x})}{\partial x_i\partial x_j\partial x_k}$ 是该张量的一个元素. 一般地，我们可以给出实值函数 $f(\boldsymbol{x})$ 相对于向量 $\boldsymbol{x}$ 的任意 k 阶偏导数为

$$\nabla^k_{\boldsymbol{x}} f(\boldsymbol{x}) = \left(\frac{\partial^k f(\boldsymbol{x})}{\partial x_{i_1}\partial x_{i_2}\cdots\partial x_{i_k}}\right).$$

相应地，$\nabla^k_{\boldsymbol{x}} f(\boldsymbol{x})$ 为一个 k 阶对称张量，且 $\frac{\partial^k f(\boldsymbol{x})}{\partial x_{i_1}\partial x_{i_2}\cdots\partial x_{i_k}}$ 是该张量的一个元素.

基于实值函数相对于向量的各阶偏导数，我们可以给出以向量为自变量的实值函数泰勒公式的一般表达式

$$f(\boldsymbol{x}+\Delta\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{k=1}^{\infty}\frac{1}{k!}(\nabla^k_{\boldsymbol{x}} f(\boldsymbol{x}))\times_1\Delta\boldsymbol{x}\times_2\Delta\boldsymbol{x}\cdots\times_k\Delta\boldsymbol{x}, \tag{B.5}$$

其中 $\nabla^1_{\boldsymbol{x}} f(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} f(\boldsymbol{x})$. 值得说明的是，(B.5) 中求和公式的每一项都有明确的物理意义. 在 $\Delta\boldsymbol{x}$ 为单位向量的情况下，

$$(\nabla^k_{\boldsymbol{x}} f(\boldsymbol{x}))\times_1\Delta\boldsymbol{x}\times_2\Delta\boldsymbol{x}\cdots\times_k\Delta\boldsymbol{x},$$

正好为函数 $f(\boldsymbol{x})$ 在 $\Delta\boldsymbol{x}$ 方向的 k 阶方向导数. 比如，$k=1$ 时，

$$(\nabla^1_{\boldsymbol{x}} f(\boldsymbol{x}))\times_1\Delta\boldsymbol{x} = (\Delta\boldsymbol{x})^{\mathrm{T}}\nabla_{\boldsymbol{x}} f(\boldsymbol{x}),$$

为函数 $f(\boldsymbol{x})$ 在 $\Delta\boldsymbol{x}$ 方向的方向导数. 当 $k=2$ 时，

$$(\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})) \times_1 \Delta\boldsymbol{x} \times_2 \Delta\boldsymbol{x} = (\Delta\boldsymbol{x})^{\mathrm{T}} \nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}) \Delta\boldsymbol{x} = (\Delta\boldsymbol{x})^{\mathrm{T}} \mathbf{H} \Delta\boldsymbol{x},$$

为函数 $f(\boldsymbol{x})$ 在 $\Delta\boldsymbol{x}$ 方向的 2 阶方向导数.

利用黑塞矩阵，可以给出实值函数的局部极小值条件，具体而言，我们有

定理 B.1（**局部极小值条件**） 如果 $\boldsymbol{x}^*$ 是函数 $f(\boldsymbol{x})$ 的局部极小值，并且 $\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 附近连续，则 $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^*) = \mathbf{0}, \nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}^*) \geq \mathbf{0}$. 其中 $\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}^*) \geq \mathbf{0}$ 代表 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 处的黑塞矩阵是半正定的.

同理，我们也可以给出实值函数的局部极大值条件，即

定理 B.2（**局部极大值条件**） 如果 $\boldsymbol{x}^*$ 是函数 $f(\boldsymbol{x})$ 的局部极大值，并且 $\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 附近连续，则 $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^*) = \mathbf{0}, \nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}^*) \leq \mathbf{0}$. 其中 $\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}^*) \leq \mathbf{0}$ 代表 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}^*$ 处的黑塞矩阵是半负定的.

例 B.12 试分析函数 $f(x, y) = x^2 + y^2 + 1$ 的局部极值情况.

解 首先计算函数对于自变量的一阶偏导数为

$$\frac{\partial f(x,y)}{\partial x} = 2x, \quad \frac{\partial f(x,y)}{\partial y} = 2y.$$

令一阶偏导数为零，可以得到该函数唯一的驻点（静止点）为 $x=0, y=0$. 然后计算该函数的黑塞矩阵为

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial^2 f(x,y)}{\partial x \partial y} \\ \frac{\partial^2 f(x,y)}{\partial y \partial x} & \frac{\partial^2 f(x,y)}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

显然，$\mathbf{H}$ 为正定矩阵. 根据定理 B.1，$x=0$, $y=0$ 是函数 $f(x, y) = x^2 + y^2 + 1$ 的局部极小值点. 事实上，$x=0$, $y=0$ 也是该函数的全局最小值点（图 B.1）.

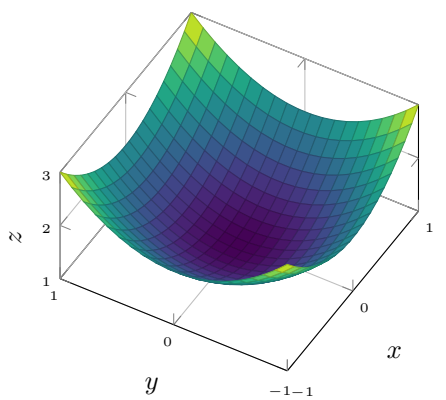


图 B.1 函数 $f(x, y) = x^2 + y^2 + 1$

例 B.13 试分析函数 $f(x,y)=x^3+y^3-3x-3y+1$ 的局部极值情况.

解　首先计算函数对于自变量的一阶偏导数为

$$\frac{\partial f(x,y)}{\partial x}=3x^2-3,\quad \frac{\partial f(x,y)}{\partial y}=3y^2-3.$$

令一阶偏导数等零可得到该函数的 4 个驻点分别为 $(1,1)$, $(1,-1)$, $(-1,1)$, $(-1,-1)$. 而该函数的黑塞矩阵为

$$\mathbf{H}=\begin{bmatrix}\frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial^2 f(x,y)}{\partial x\partial y}\\ \frac{\partial^2 f(x,y)}{\partial y\partial x} & \frac{\partial^2 f(x,y)}{\partial y^2}\end{bmatrix}=\begin{bmatrix}6x & 0\\ 0 & 6y\end{bmatrix}.$$

当 $x=1$, $y=1$ 时, $\mathbf{H}$ 为正定矩阵, 因此点 $(1,1)$ 是该函数的极小值点; 当 $x=-1$, $y=-1$ 时, $\mathbf{H}$ 为负定矩阵, 因此点 $(-1,-1)$ 是该函数的极大值点; 当 $x=1$, $y=-1$ 时, $\mathbf{H}$ 为不定矩阵, 因此点 $(1,-1)$ 是该函数的鞍点; 当 $x=-1$, $y=1$ 时, $\mathbf{H}$ 仍为不定矩阵, 因此点 $(-1,1)$ 也是该函数的鞍点（图 B.2 ）.

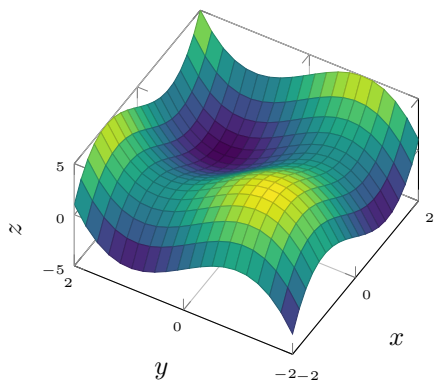


图 B.2　函数 $f(x,y)=x^3+y^3-3x-3y+1$