

第3章

矩阵与内积空间

在第2章中,我们从代数学的角度探讨了向量空间和线性变换,强调了矩阵在其中的作用。本章将从几何学的角度探讨这些概念,以便直观地理解它们。几何学需要在向量空间中建立长度和角度的概念。为了能在向量空间中引入这两个概念,数学家们巧妙地引入了内积。内积将向量空间中的代数学和几何学紧密地联系了起来。

3.1 范数

在解析几何中,向量是始于原点的有向线段,它的长度是原点到有向线段终点的距离。在向量空间中,向量的长度或大小用一个称为范数的函数来度量。

定义 3.1(范数) 设 V 是向量空间,如果 V 上定义的函数

$$\|\cdot\|:V\rightarrow\mathbb{R}, \quad x\mapsto\|x\| \quad (3.1)$$

对于 $\forall\lambda\in\mathbb{R}, \forall x,y\in V$,满足以下3个条件。

(1) 正定性: $\|x\|\geq 0$, 且 $\|x\|=0\iff x=0$ 。

(2) 齐次性: $\|\lambda x\|=|\lambda|\cdot\|x\|$ 。

(3) 三角不等式: $\|x+y\|\leq\|x\|+\|y\|$ 。

那么,称函数 $\|\cdot\|$ 为向量空间 V 上的**范数(norm)**。

定义3.1是在一般的向量空间上定义范数的,但机器学习的应用中只考虑有限维向量空间 \mathbb{R}^n 上的范数。对于 $x\in\mathbb{R}^n$,用下标表示向量的元素(分量),比方说 x_i 是向量 x 的第 i 个元素。

下面,我们来看 \mathbb{R}^n 上3种常见的范数。

(1) **L_1 范数**: 对于 $x\in\mathbb{R}^n$, x 的 L_1 范数被定义为

$$\|x\|_1=\sum_{i=1}^n|x_i|, \quad (3.2)$$

L_1 范数也称为曼哈顿范数。图3.1(a)显示了 \mathbb{R}^2 中所有 $\|x\|_1=1$ 的向量 x 的终点,形成了两条对角线分别在两条坐标轴上的菱形,其上任一点的横、纵坐标绝对值之和均为1。

(2) **L_2 范数**: 对于 $x\in\mathbb{R}^n$, x 的 L_2 范数被定义为

$$\|x\|_2=\sqrt{\sum_{i=1}^nx_i^2}=\sqrt{x^Tx}, \quad (3.3)$$

L_2 范数也称为欧几里得(Euclidean)范数。图3.1(b)显示了 \mathbb{R}^2 中所有 $\|x\|_2=1$ 的向量 x

的终点,形成了以原点为圆心的单位圆。

【注 3.1】 本书中如果没有特别说明,使用的范数默认为 L_2 范数。

(3) L_∞ 范数: 对于 $\mathbf{x} \in \mathbb{R}^n$, \mathbf{x} 的 L_∞ 范数被定义为

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|. \quad (3.4)$$

L_∞ 范数也称为无穷范数。图 3.1(c)显示了 \mathbb{R}^2 中所有 $\|\mathbf{x}\|_\infty = 1$ 的向量 \mathbf{x} 的终点,形成了以原点为中心、两组对边分别平行于两条坐标轴的正方形,其上任一点的横、纵坐标绝对值的最大值均为 1。

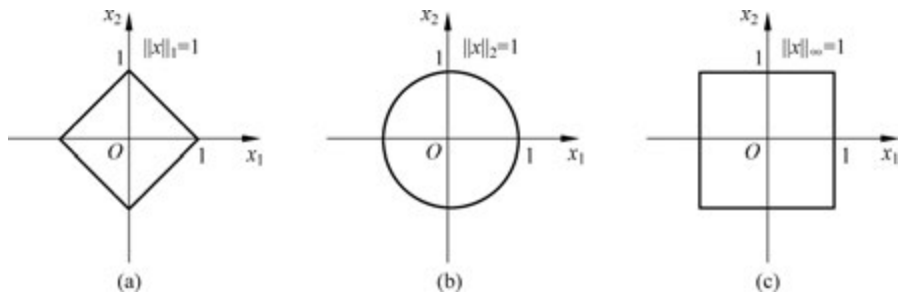


图 3.1 范数为 1 的向量的终点集

事实上,以上 3 种范数可以统一为如下形式的 L_p 范数

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad p=1,2,\dots. \quad (3.5)$$

当 $p=1$ 时,式(3.5)表示 L_1 范数;当 $p=2$ 时,式(3.5)表示 L_2 范数;当 $p \rightarrow \infty$ 时,式(3.5)表示 L_∞ 范数。

3.2 内积

内积可以用来计算向量的长度和两个向量之间的距离,也可以计算两个向量之间的角度,进而确定向量之间是否正交。我们对内积的认识总是从点积开始的。

3.2.1 点积

对于 $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$,

$$\mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i \quad (3.6)$$

称为这两个向量的点积(dot product)。点积的结果是一个数,所以又称为标量积(scalar product)。

点积是一种特殊的内积,而内积是更加一般化的概念。

3.2.2 一般内积

回忆线性映射的定义(参见 2.4 节),它是保持向量的加法和数乘运算的一元映射。把线性映射 $\Phi(\mathbf{x})$ 推广到二元,就是双线性映射。

定义 3.2(双线性映射) 设 V 是向量空间,双线性映射 $\Omega: V \times V \rightarrow \mathbb{R}$ 将 V 中的两个向

量映射为一个实数,它对于每个参数位置都是线性的,即对于 $\forall x, y, z \in V, \forall \lambda, \mu \in \mathbb{R}$, 有

$$\Omega(\lambda x + \mu y, z) = \lambda \Omega(x, z) + \mu \Omega(y, z), \quad (3.7)$$

$$\Omega(x, \lambda y + \mu z) = \lambda \Omega(x, y) + \mu \Omega(x, z). \quad (3.8)$$

式(3.7)表明 Ω 关于第一个参数位置是线性的; 式(3.8)表明 Ω 关于第二个参数位置是线性的。

定义 3.3(对称和正定) 设 V 是向量空间, $\Omega: V \times V \rightarrow \mathbb{R}$ 是双线性映射, 则

(1) 若 $\forall x, y \in V, \Omega(x, y) = \Omega(y, x)$, 则称 Ω 为对称的。

(2) 若 $\forall x \in V \setminus \{0\}, \Omega(x, x) > 0$, 且 $\Omega(0, 0) = 0$, 则称 Ω 为正定的。

定义 3.4(内积与内积空间) 设 V 是向量空间, $\Omega: V \times V \rightarrow \mathbb{R}$ 是双线性映射, 则

(1) 对称正定的双线性映射 $\Omega: V \times V \rightarrow \mathbb{R}$ 称为 V 上的内积(inner product), 内积通常记为 $\langle x, y \rangle$ 。

(2) $(V, \langle \cdot, \cdot \rangle)$ 称为内积空间或带有内积的(实)向量空间。若内积就是点积, 则称 $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ 为欧几里得空间, 简称欧氏空间。

【例 3.1】(非点积的内积)

设 $V = \mathbb{R}^2$, 定义

$$\langle x, y \rangle = x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2x_2 y_2,$$

则 $\langle \cdot, \cdot \rangle$ 是内积但不是点积。其证明作为练习留给读者。

3.2.3 对称正定矩阵

对称正定矩阵在机器学习中起着重要的作用, 而且可以用来定义内积。第 4 章的矩阵分解将涉及对称正定矩阵。

设 V 是 n 维向量空间, 带有内积 $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$, 一个有序基为 $B = (b_1, b_2, \dots, b_n)$, 则 $\forall x, y \in V$ 都可以写成基向量的线性组合, 即 $\exists \lambda_i, \mu_j \in \mathbb{R}, i, j = 1, 2, \dots, n$, 使 $x =$

$\sum_{i=1}^n \lambda_i b_i \in V, y = \sum_{j=1}^n \mu_j b_j \in V$ 。由于内积的双线性性质, 对 $\forall x, y \in V$ 有

$$\langle x, y \rangle = \left\langle \sum_{i=1}^n \lambda_i b_i, \sum_{j=1}^n \mu_j b_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \langle b_i, b_j \rangle \mu_j = \hat{x}^T A \hat{y},$$

其中 $A = [\langle b_i, b_j \rangle]_{n \times n}$, \hat{x}, \hat{y} 分别是 x, y 关于基 B 的坐标向量。这意味着 $\langle \cdot, \cdot \rangle$ 可由 A 来确定。内积 $\langle \cdot, \cdot \rangle$ 的对称性意味着 A 是对称矩阵。此外, 由内积的正定性知,

$$\forall x \in V \setminus \{0\}, \text{ 有 } x^T A x > 0. \quad (3.9)$$

定义 3.5(对称正定矩阵) 满足式(3.9)的对称矩阵 A 称为对称正定矩阵, 或者就称为正定矩阵(positive definite matrix)。若式(3.9)中的“ $>$ ”替换为“ \geq ”, 则称矩阵 A 为对称半正定矩阵。

【例 3.2】 判断下列矩阵是否正定。

$$A_1 = \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 9 & 6 \\ 6 & 3 \end{bmatrix}.$$

【解】 A_1 是对称的, 且 $\forall x \in \mathbb{R}^2 \setminus \{0\}$, 有

$$x^T A_1 x = [x_1, x_2] \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 9x_1^2 + 12x_1 x_2 + 5x_2^2 = (3x_1 + 2x_2)^2 + x_2^2 > 0,$$

因此 A_1 是正定的。

A_2 是对称的,但不是正定的,因为对于 $x \in \mathbb{R}^2 \setminus \{0\}$,有

$$x^T A_2 x = [x_1, x_2] \begin{bmatrix} 9 & 6 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 9x_1^2 + 12x_1x_2 + 3x_2^2 = (3x_1 + 2x_2)^2 - x_2^2,$$

所以 $x^T A_2 x$ 可能小于 0, 比如当 $x = [2, -3]^T$ 时。

向量空间的基选定后,任何对称正定矩阵都以如下方式确定一个内积。

定理 3.1 设 V 是实值有限维向量空间,有序基 $B = (b_1, b_2, \dots, b_n)$, $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ 是一个内积当且仅当存在一个对称正定矩阵 $A \in \mathbb{R}^{n \times n}$, 对于 $\forall x, y \in V$, 有

$$\langle x, y \rangle = \hat{x}^T A \hat{y}. \quad (3.10)$$

其中 \hat{x}, \hat{y} 分别是 x, y 关于基 B 的坐标向量。

由式(3.6)和式(3.10)易看出,点积是内积的特殊情况(当且仅当 $A = E_n$)。

【注 3.2】 对称正定矩阵 $A \in \mathbb{R}^{n \times n}$ 具有以下性质。

(1) A 的零空间(核)只由 0 组成,这是因为 $\forall x \neq 0 \Rightarrow x^T A x > 0$, 于是 $x \neq 0 \Rightarrow A x \neq 0$ 。

(2) A 的对角线元素 a_{ii} 是正的,这是因为 $a_{ii} = e_i^T A e_i > 0$, 其中 e_i 为 \mathbb{R}^n 标准基的第 i 个基向量, $i = 1, 2, \dots, n$ 。

3.3 长度和距离

3.1 节介绍了范数,它用来计算向量的长度。3.2 节引入了内积。内积可以与范数产生联系,因为任何内积 $\langle \cdot, \cdot \rangle$ 都可以自然地诱导出范数

$$\|x\| = \sqrt{\langle x, x \rangle}, \quad (3.11)$$

于是就可以用内积来计算向量的长度(length)。然而,并不是每个范数都是由内积诱导出来的,例如,曼哈顿范数[式(3.2)]并没有相应的内积。接下来我们只探讨由内积诱导的范数。

【注 3.3】(柯西-施瓦茨不等式) 内积 $\langle \cdot, \cdot \rangle$ 诱导的范数 $\|\cdot\|$ 满足柯西-施瓦茨不等式

$$|\langle x, y \rangle| \leq \|x\| \|y\|. \quad (3.12)$$

【例 3.3】(使用内积定义的向量长度)

现在,内积诱导的范数就可以用来计算向量长度。取向量 $x = [1, 1]^T \in \mathbb{R}^2$, 若使用点积作为内积,则向量 x 的长度为

$$\|x\| = \sqrt{x^T x} = \sqrt{1^2 + 1^2} = \sqrt{2};$$

若选用另一个内积

$$\langle x, y \rangle = x^T \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} y = x_1 y_1 - \frac{1}{2}(x_1 y_2 + x_2 y_1) + x_2 y_2,$$

则当 $x = [1, 1]^T$ 时,有

$$\langle x, x \rangle = x_1^2 - x_1 x_2 + x_2^2 = 1 - 1 + 1 = 1,$$

$$\|x\| = \sqrt{\langle x, x \rangle} = 1.$$

这个长度比用点积计算的小,可以看出,这是因为此时 $x_1 x_2 > 0$ 。

对于 $y = [1, -1]^T$, $y_1 y_2 < 0$, $\|y\| = \sqrt{\langle y, y \rangle} = \sqrt{y_1^2 - y_1 y_2 + y_2^2} = \sqrt{3}$, 此时比用点积计算的长度大。

定义 3.6(距离和度量) 给定内积空间为 $(V, \langle \cdot, \cdot \rangle)$, $\forall x, y \in V$,

$$d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle} \quad (3.13)$$

称为向量 x 与 y 之间的距离(distance)。若用点积作为内积,则此距离称为欧几里得距离。

映射

$$\begin{aligned} d: V \times V &\rightarrow \mathbb{R}, \\ (x, y) &\mapsto d(x, y), \end{aligned}$$

称为度量(metric)。

【注 3.4】 与向量的长度类似,向量之间的距离未必需要内积,这是因为范数未必由内积诱导。如果使用由内积诱导的范数,那么距离可能会随着内积的选择而变化。

度量 d 有以下性质。

- (1) d 是正定的,即 $\forall x, y \in V$, 有 $d(x, y) \geq 0$, 且 $d(x, y) = 0 \Leftrightarrow x = y$ 。
- (2) d 是对称的,即 $\forall x, y \in V$, 有 $d(x, y) = d(y, x)$ 。
- (3) d 满足三角不等式,即 $\forall x, y, z \in V$, 有 $d(x, z) \leq d(x, y) + d(y, z)$ 。

【注 3.5】 乍一看,内积和度量的性质非常相似,其实 $\langle x, y \rangle$ 和 $d(x, y)$ 的大小是相反的:两个向量越接近,度量就越小;但当长度给定时,两个向量越接近,内积越大。

3.4 角度和正交

内积还可以用来计算向量之间的夹角。

根据柯西-施瓦茨不等式[式(3.12)],设 $x \neq 0, y \neq 0$, 有

$$-1 \leq \frac{\langle x, y \rangle}{\|x\| \|y\|} \leq 1,$$

因此,存在唯一的 $\omega \in [0, \pi]$, 使

$$\cos \omega = \frac{\langle x, y \rangle}{\|x\| \|y\|}, \quad (3.14)$$

ω 就是向量 x 与 y 之间的夹角(included angle)。

【例 3.4】 用点积作为内积来计算向量 $x = [1, 1]^T$ 和 $y = [1, 2]^T$ 之间的夹角,根据式(3.14)得

$$\cos \omega = \frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle \langle y, y \rangle}} = \frac{x^T y}{\sqrt{x^T x y^T y}} = \frac{3}{\sqrt{10}},$$

因此,向量 x 与 y 之间夹角为 $\omega = \arccos \frac{3}{\sqrt{10}}$, 约为 18° 。

内积的一个重要作用是用来刻画正交性。

定义 3.7(正交性) 给定向量 x 与 y , x 与 y 正交(orthogonality)当且仅当 $\langle x, y \rangle = 0$, 记作 $x \perp y$ 。若 x 与 y 还是单位向量,即 $\|x\| = \|y\| = 1$, 则称 x 与 y 是规范正交的(orthonormal)。

【注 3.6】 由定义 3.7 知, 向量空间中的零向量(0)与每个向量正交。

正交是几何中垂直概念的推广。然而, 正如选取不同的内积会得到不同的向量长度一样, 向量之间的垂直关系也依赖于所选取的内积。下面看一个例子。

【例 3.5】 对 \mathbb{R}^2 中向量 $x = [1, 1]^T, y = [-1, 1]^T$, 用两个不同的内积来计算它们的夹角(见图 3.2)。

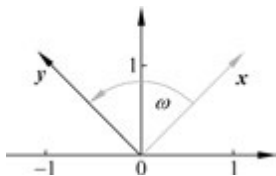


图 3.2 向量的正交性依赖于内积的选取

一是用点积作为内积, 显然 $x^T y = 0$, 因此 x 与 y 之间的夹角 $\omega = 90^\circ$, 即 $x \perp y$ 。

二是选取内积 $\langle x, y \rangle = x^T \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} y$, 则

$$\cos \omega = \frac{\langle x, y \rangle}{\|x\| \|y\|} = -\frac{1}{3} \Rightarrow \omega = \arccos\left(-\frac{1}{3}\right) \approx 109.5^\circ,$$

此时 x 与 y 不正交。因此, 关于一种内积正交的向量不一定关于另一种内积正交。

定义 3.8(正交矩阵) 方阵 $A \in \mathbb{R}^{n \times n}$ 为正交矩阵(orthogonal matrix)当且仅当 A 的各列向量是规范正交的, 这等价于

$$A^T A = A A^T = E, \quad (3.15)$$

即

$$A^{-1} = A^T, \quad (3.16)$$

也就是说, 正交矩阵的转置就是其逆矩阵。

【注 3.7】 由定义 3.8 易证, 正交矩阵的乘积仍为正交矩阵, 证明留给读者。

2.4.2 小节已经讲过, 矩阵对向量的线性变换, 直观上就是旋转和伸缩, 而且一般来说, 同一矩阵对不同的向量旋转的角度不同、伸缩的倍数也不同。但当使用点积作为内积时, 正交矩阵对向量的变换[称为正交变换(orthogonal transformation)]很特殊, 具体通过下面的注 3.8 和注 3.9 来解释。

【注 3.8】(保长性和保角性)

正交变换具有保长性和保角性。以下设 $A \in \mathbb{R}^{n \times n}$ 为正交矩阵, $x, y \in \mathbb{R}^n$ 为任意向量。

(1) 保长性:

$$\|Ax\|^2 = (Ax)^T (Ax) = x^T A^T A x = x^T E x = x^T x = \|x\|^2, \quad (3.17)$$

即 x 经过正交矩阵 A 的变换后, 长度没有变化。

(2) 保角性: 设 Ax 与 Ay 之间的夹角为 ω , 则

$$\cos \omega = \frac{(Ax)^T (Ay)}{\|Ax\| \|Ay\|} = \frac{x^T A^T A y}{\sqrt{x^T A^T A x y^T A^T A y}} = \frac{x^T y}{\|x\| \|y\|}, \quad (3.18)$$

即 x, y 经过正交矩阵 A 的变换后, 夹角没有变化。

综上, 正交矩阵对向量的变换, 不进行伸缩, 只改变方向, 且向量之间的夹角保持不变。

进一步还可以推知, 在正交变换下, 几何体的运动只可能是旋转、反射及它们的组合, 其

形状和大小都不发生改变,即正交变换是刚体运动,详见注 3.9。

【注 3.9】(旋转矩阵和反射矩阵)

由保长性和保角性可推知, \mathbb{R}^n 中的正交变换对刚体(形状和大小都不变的几何体)来说,实现的是①围绕原点的旋转;②关于过原点直线的反射;③前两种变换的复合。换句话说, $\mathbb{R}^{n \times n}$ 中的正交矩阵可以是旋转矩阵、反射矩阵或二者的乘积。

以 \mathbb{R}^2 中的正交变换为例,旋转矩阵形如

$$\text{Rot}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad (3.19)$$

对于 $\forall x \in \mathbb{R}^2$, 向量 $\text{Rot}(\theta)x$ 由 x 围绕原点逆时针旋转 θ 角度得到,如图 3.3(a)所示;反射矩阵形如

$$\text{Ref}(\theta) = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}, \quad (3.20)$$

对于 $\forall x \in \mathbb{R}^2$, 向量 $\text{Ref}(\theta)x$ 由 x 关于倾斜角为 θ 的过原点直线反射而得到,如图 3.3(b)所示,其中 $\alpha = \beta$ 但与 θ 无关。

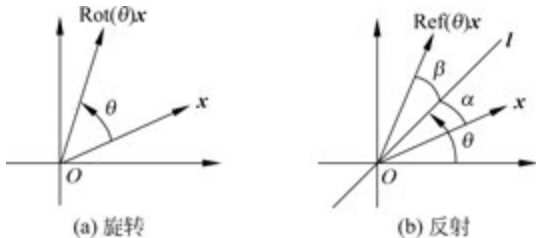


图 3.3 向量的旋转和反射(其中 $\alpha = \beta$)

【例 3.6】 令 $\theta = \frac{\pi}{2}$, 则 \mathbb{R}^2 中的旋转矩阵和反射矩阵分别为

$$\text{Rot}\left(\frac{\pi}{2}\right) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \text{Ref}\left(\frac{\pi}{2}\right) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix},$$

对于 $\forall \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, 显然, $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$, 它由 $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 逆时针旋转 $\frac{\pi}{2}$ 得到;

$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_1 \\ x_2 \end{bmatrix}$, 它与 $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 关于 x_2 坐标轴对称,如图 3.4 所示。

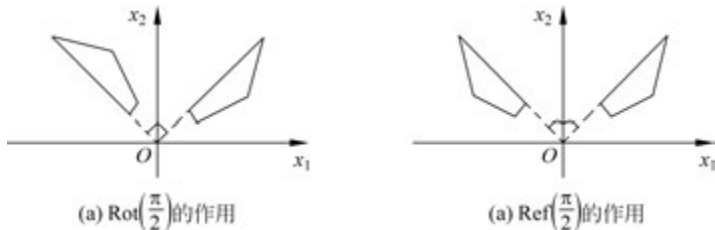


图 3.4 $\theta = \frac{\pi}{2}$ 时几何体的旋转和反射

3.5 规范正交基

我们已经知道, n 维向量空间的一个基由 n 个线性无关的向量组成。我们刚刚又学习了如何用内积计算向量的长度和向量之间的夹角。在此基础上, 可以定义一种特殊的基, 基向量长度都为 1 且两两正交。

定义 3.9 (规范正交基) 给定 n 维内积空间 \mathbb{V} , 以及 \mathbb{V} 的一个基 $\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$, 如果

$$\langle \epsilon_i, \epsilon_j \rangle = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}, \quad (3.21)$$

对任意的 $i, j=1, 2, \dots, n$ 都成立, 那么称基 $\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$ 为规范正交基。

“规范”即基向量的范数(长度)都等于 1, “正交”即基向量两两正交。

在 2.3.4 小节中, 使用高斯消元法可以为一组向量张成的向量空间找到其基向量。本章我们会介绍: 得到基向量(一般是非正交的)之后, 可以通过施密特正交化方法来构建规范正交基(详见 3.8.4 小节)。

【例 3.7】 欧氏空间 \mathbb{R}^n 的标准基 $\{e_1, e_2, \dots, e_n\}$ 是规范正交基, 其中内积是向量的点积。例如, 在 \mathbb{R}^2 中,

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

显然 $\|e_1\| = \sqrt{1^2 + 0^2} = 1$, $\|e_2\| = 1$, $e_1^T e_2 = 1 \times 0 + 0 \times 1 = 0$, $\{e_1, e_2\}$ 是规范正交基;

$$\epsilon_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \epsilon_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

因为 $\|\epsilon_1\| = \|\epsilon_2\| = 1$ 且 $\epsilon_1^T \epsilon_2 = 0$, 所以是 ϵ_1, ϵ_2 也组成 \mathbb{R}^2 的一个规范正交基。

3.6 正交补

前面我们定义了向量之间的正交, 现在进一步来探讨向量空间之间的正交。

定义 3.10 (正交补) 给定 D 维内积空间 \mathbb{V} 及其 d 维子空间 \mathbb{U} , \mathbb{U} 的正交补空间 \mathbb{U}^\perp 是 \mathbb{V} 的一个 $D-d$ 维子空间, 这个子空间包含 \mathbb{V} 中所有与 \mathbb{U} 中向量正交的向量, 即

$$\mathbb{V}^\perp = \{x \in \mathbb{V} : \forall u \in \mathbb{U}, x \perp u\}. \quad (3.22)$$

显然 $\mathbb{U} \cap \mathbb{V}^\perp = \{\mathbf{0}\}$ 。因而 $\forall x \in \mathbb{V}$ 都能被唯一分解为

$$x = \sum_{i=1}^d \lambda_i b_i + \sum_{j=1}^{D-d} \mu_j b_j^\perp, \quad \lambda_i, \mu_j \in \mathbb{R}, \quad (3.23)$$

其中 (b_1, b_2, \dots, b_d) 为 \mathbb{U} 的基, $(b_1^\perp, b_2^\perp, \dots, b_{D-d}^\perp)$ 为 \mathbb{U}^\perp 的基。

举个直观的正交补例子。如图 3.5 所示, 设 \mathbb{U} 是 \mathbb{R}^3 中的 2 维子空间, 直观上它是一个过原点的平面。若 \mathbb{R}^3 中的向量 w 与平面 \mathbb{U} 正交, 则它是正交补 \mathbb{U}^\perp 的基向量。所有与 w 正交的向量必在 \mathbb{U} 中。 w 称为 \mathbb{U} 的**法向量**(normal vector)。

在向量空间和仿射空间中, 通常用法向量描述超平面。

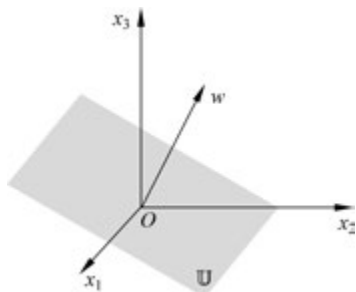


图 3.5 3 维空间中平面的正交补由其法向量张成

3.7 函数的内积

2.3.1 小节介绍过,广义的向量空间中的向量不局限于 \mathbb{R}^n 中的元素。这有两层意思:其一,向量可以不是只有一列或一行的矩阵;其二,向量空间可以是无限维的。记 $C[a, b]$ 为定义在 $[a, b]$ 上的连续函数的全体,易见 $C[a, b]$ 关于函数的加法和数乘是封闭的,并且满足 2.3.1 小节中注 2.13 列出的 8 条性质,因此 $C[a, b]$ 是一个向量空间,其中的向量就是任意定义在 $[a, b]$ 上的连续函数。那么,如何定义其中两个向量的内积呢?

设函数 $u(x), v(x) \in C[a, b]$, 则 $u(x), v(x)$ 都在 $[a, b]$ 上可积,二者的内积可以定义为

$$\langle u, v \rangle = \int_a^b u(x)v(x)dx. \quad (3.24)$$

于是可以通过函数的内积来定义范数 $\|u\| = \sqrt{\int_a^b [u(x)]^2 dx}$ 。此外可以定义正交:若式(3.24)的值为 0, 则函数 $u(x)$ 与 $v(x)$ 是正交的。

【例 3.8】(函数的正交) 设函数 $u = \sin x, v = \cos x$, 则按式(3.24)定义的内积中被积函数为 $f(x) = \sin x \cos x$, 如图 3.6 所示。由于 $f(-x) = -f(x)$, 所以该函数是奇函数, 且该函数在区间 $[-\pi, \pi]$ 上的积分值为 0, 故两个函数 $u = \sin x$ 与 $v = \cos x$ ($x \in [-\pi, \pi]$) 正交。

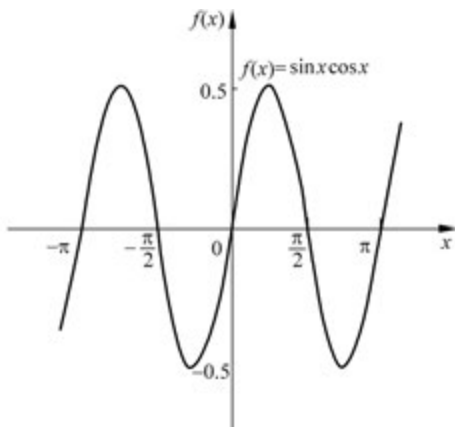


图 3.6 内积中被积函数的图像

【注 3.10】(余弦函数系) 若积分区间为 $[-\pi, \pi]$, 则函数集

$$\{1, \cos x, \cos 2x, \cos 3x, \dots\} \quad (3.25)$$

是正交的, 即其中任意两个函数正交, 这是因为 $\forall m, n \in \mathbb{N}$ 且 $m \neq n$, 有 $m+n, m-n \in \mathbb{N}_+$, 于是

$$\int_{-\pi}^{\pi} \cos mx \cos nx dx = \frac{1}{2} \left[\int_{-\pi}^{\pi} \cos(m+n)x dx + \int_{-\pi}^{\pi} \cos(m-n)x dx \right] = 0.$$

此函数集张成了一个函数子空间, 子空间中的函数形如

$$\sum_{k=0}^{\infty} \lambda_k \cos kx, \quad (3.26)$$

在 $[-\pi, \pi]$ 上都是偶函数。将一个偶函数 $g(x)$ 投影到这个子空间, 用式 (3.26) 近似 $g(x)$, 这正是傅里叶级数的思想。

3.8 正交投影

投影是一类重要的线性变换。机器学习中处理的数据通常是高维的, 而高维数据难以分析及可视化。然而, 高维数据通常只有少数维度包含大部分的信息, 大多数其他维度对描述数据的关键性质不是必需的。当然, 对高维数据进行压缩或可视化时, 会不可避免地丢失数据的信息。为了使压缩损失最小化, 我们需要保留数据信息量的最大维度。

第 1 章中已经介绍了, 数据可以表示为向量, 并且可以将原始高维数据投影到一个低维特征空间中, 在低维空间中分析数据。本节要探讨的正交投影, 是将高维数据投影到给定的低维子空间中, 保留原始数据尽可能多的信息并最小化原始数据和相应投影之间的误差。图 3.7 显示了正交投影的一个示例。

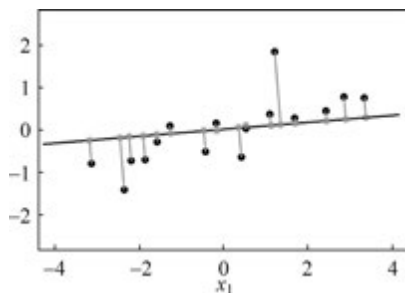


图 3.7 2 维数据集到 1 维子空间的正交投影

首先要明确投影是一种怎样的线性变换。

定义 3.11(投影) 设 V 是向量空间, U 是 V 的一个子空间, 若线性变换 $\pi: V \rightarrow U$ 满足

$$\pi^2 = \pi \circ \pi = \pi, \quad (3.27)$$

则 π 称为一个**投影**(projection)。

因为线性变换可以用变换矩阵来表示, 所以投影对应于一种特殊的变换矩阵 P_π , 其称为**投影矩阵**(projection matrix), 具有性质

$$P_\pi^2 = P_\pi. \quad (3.28)$$

接下来, 我们推导内积空间 $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ 到其子空间的**正交投影**(orthogonal

projection), 先从 1 维子空间开始, 最终推导出一般子空间上的正交投影。以下默认以点积 $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ 为内积。

3.8.1 1 维子空间上的正交投影

给定一条经过原点的直线(1 维子空间), 设其基向量为 $\mathbf{b} \in \mathbb{R}^n$ 。也就是说, 这条直线是由 \mathbf{b} 张成的 1 维子空间 $\mathbb{U} \subseteq \mathbb{R}^n$ 。要将向量 $\mathbf{x} \in \mathbb{R}^n$ 正交地投影到 \mathbb{U} 上, 就是要找到 \mathbb{U} 中最接近 \mathbf{x} 的向量 $\pi_{\mathbb{U}}(\mathbf{x})$ 。投影向量 $\pi_{\mathbb{U}}(\mathbf{x})$ 的本质可以从以下两个方面来刻画(如图 3.8 所示)。

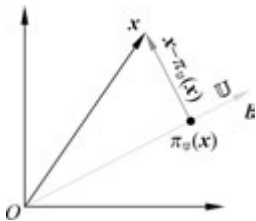


图 3.8 \mathbb{R}^2 到一维子空间的正交投影

(1) $\pi_{\mathbb{U}}(\mathbf{x})$ 最接近向量 \mathbf{x} , 即它们之间的距离 $\|\mathbf{x} - \pi_{\mathbb{U}}(\mathbf{x})\|$ 最小, 这意味着差向量 $\mathbf{x} - \pi_{\mathbb{U}}(\mathbf{x})$ 正交于直线 \mathbb{U} , 也就是 $\mathbf{x} - \pi_{\mathbb{U}}(\mathbf{x})$ 与 \mathbb{U} 的基向量 \mathbf{b} 正交, 所以 $\langle \mathbf{x} - \pi_{\mathbb{U}}(\mathbf{x}), \mathbf{b} \rangle = 0$ 。

(2) 向量 \mathbf{x} 在 \mathbb{U} 上的投影 $\pi_{\mathbb{U}}(\mathbf{x})$ 必是 \mathbb{U} 中的向量, 因此 $\pi_{\mathbb{U}}(\mathbf{x})$ 可由 \mathbb{U} 的基向量 \mathbf{b} 线性表示, 即 $\exists \lambda \in \mathbb{R}$, 有 $\pi_{\mathbb{U}}(\mathbf{x}) = \lambda \mathbf{b}$, 其中 λ 是 $\pi_{\mathbb{U}}(\mathbf{x})$ 关于 \mathbf{b} 的坐标。

下面按照求投影的坐标 λ 、投影向量 $\pi_{\mathbb{U}}(\mathbf{x})$ 、投影矩阵 \mathbf{P}_{π} 的顺序, 详细解析 1 维子空间上的正交投影。

第 1 步, 求投影的坐标 λ 。在正交条件下有

$$\langle \mathbf{x} - \pi_{\mathbb{U}}(\mathbf{x}), \mathbf{b} \rangle = 0 \quad \stackrel{\pi_{\mathbb{U}}(\mathbf{x}) = \lambda \mathbf{b}}{\Leftrightarrow} \quad \langle \mathbf{x} - \lambda \mathbf{b}, \mathbf{b} \rangle = 0,$$

利用内积的双线性及对称性, 得

$$\langle \mathbf{x}, \mathbf{b} \rangle - \lambda \langle \mathbf{b}, \mathbf{b} \rangle = 0 \Leftrightarrow \lambda = \frac{\langle \mathbf{x}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} = \frac{\langle \mathbf{b}, \mathbf{x} \rangle}{\|\mathbf{b}\|^2},$$

此时内积为点积, 于是可得

$$\lambda = \frac{\mathbf{b}^T \mathbf{x}}{\mathbf{b}^T \mathbf{b}} = \frac{\mathbf{b}^T \mathbf{x}}{\|\mathbf{b}\|^2}. \quad (3.29)$$

若 $\|\mathbf{b}\| = 1$, 则投影的坐标 λ 等于 $\mathbf{b}^T \mathbf{x}$ 。

第 2 步, 求投影向量 $\pi_{\mathbb{U}}(\mathbf{x}) \in \mathbb{U}$ 。因为 $\pi_{\mathbb{U}}(\mathbf{x}) = \lambda \mathbf{b}$, 由式(3.29)得

$$\pi_{\mathbb{U}}(\mathbf{x}) = \lambda \mathbf{b} = \frac{\mathbf{b}^T \mathbf{x}}{\|\mathbf{b}\|^2} \mathbf{b}. \quad (3.30)$$

第 3 步, 求投影矩阵 \mathbf{P}_{π} 。根据定义 3.11, 投影是线性变换, 因此正交投影对应一个投影矩阵 \mathbf{P}_{π} , 使

$$\pi_{\mathbb{U}}(\mathbf{x}) = \mathbf{P}_{\pi} \mathbf{x}. \quad (3.31)$$

由式(3.30), 有

$$\pi_{\mathbb{U}}(\mathbf{x}) = \lambda \mathbf{b} = \mathbf{b} \lambda = \mathbf{b} \frac{\mathbf{b}^T \mathbf{x}}{\|\mathbf{b}\|^2} = \frac{\mathbf{b} \mathbf{b}^T}{\|\mathbf{b}\|^2} \mathbf{x}, \quad (3.32)$$

比较式(3.31)和式(3.32),立即可知

$$\mathbf{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\mathrm{T}}{\|\mathbf{b}\|^2}. \quad (3.33)$$

其中 $\|\mathbf{b}\|^2 = \mathbf{b}^\mathrm{T}\mathbf{b}$ 一个数, $\mathbf{b}\mathbf{b}^\mathrm{T}$ 是一个秩为1的对称矩阵,故 \mathbf{P}_π 是一个秩为1的对称矩阵。

至此,由式(3.33)计算的矩阵 \mathbf{P}_π ,可将任意的 $\mathbf{x} \in \mathbb{R}^n$ 正交地投影到 \mathbf{b} 所张成的子空间 \mathbb{U} 上,投影向量即 $\mathbf{P}_\pi \mathbf{x} \in \mathbb{U}$ 。

【例 3.9】 设一条过原点的直线的方向向量为 $\mathbf{b} = [1, 2, 2]^\mathrm{T}$, 求出该直线上的正交投影矩阵 \mathbf{P}_π 。

【解】 由式(3.33),得

$$\mathbf{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\mathrm{T}}{\mathbf{b}^\mathrm{T}\mathbf{b}} = \frac{1}{9} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} [1 \ 2 \ 2] = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}.$$

可以举例验证 \mathbf{P}_π 为正交投影: ①随机选择一个 $\mathbf{x} \in \mathbb{R}^3$, 检验 $\mathbf{x} - \mathbf{P}_\pi \mathbf{x}$ 是否与 \mathbf{b} 正交;
②检验 \mathbf{P}_π 是否为投影矩阵。例如,对 $\mathbf{x} = [1, 1, 1]^\mathrm{T}$, 可求得

$$\mathbf{P}_\pi \mathbf{x} = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{5}{9} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \in \text{span} \left(\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right).$$

验证正交性:

$$\mathbf{b}^\mathrm{T}(\mathbf{x} - \mathbf{P}_\pi \mathbf{x}) = [1 \ 2 \ 2] \left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{5}{9} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right) = 0.$$

验证投影:

$$\mathbf{P}_\pi^2 = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} \cdot \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} = \mathbf{P}_\pi.$$

【注 3.11】 $\mathbf{P}_\pi \pi_\mathbb{U}(\mathbf{x}) = \pi_\mathbb{U}(\mathbf{x})$ 意味着 \mathbf{P}_π 的作用不会改变 $\pi_\mathbb{U}(\mathbf{x})$, 使用 4.2 节的知识, 可以说明投影 $\pi_\mathbb{U}(\mathbf{x})$ 是投影矩阵 \mathbf{P}_π 的特征向量, 其对应的特征值为 1。

3.8.2 一般子空间上的正交投影

接下来,研究 \mathbb{R}^n 中的向量到 m 维子空间 $\mathbb{U} \subseteq \mathbb{R}^n$ 的正交投影,其中 $m \leq n, m \in \mathbb{N}_+$ 。设向量组 $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m)$ 是 \mathbb{U} 的有序基。要将向量 $\mathbf{x} \in \mathbb{R}^n$ 正交地投影到 \mathbb{U} 上,就是要找到 \mathbb{U} 中最接近 \mathbf{x} 的向量 $\pi_\mathbb{U}(\mathbf{x})$ 。

与 1 维子空间的情形相仿,投影向量 $\pi_\mathbb{U}(\mathbf{x})$ 的本质可以从两个方面来刻画(图 3.9 是 \mathbb{R}^3 中的向量投影到 2 维子空间的示例)。

(1) 距离 $\|\mathbf{x} - \pi_\mathbb{U}(\mathbf{x})\|$ 最小,意味着差向量 $\mathbf{x} - \pi_\mathbb{U}(\mathbf{x})$ 正交于子空间 \mathbb{U} ,也就是 $\mathbf{x} - \pi_\mathbb{U}(\mathbf{x})$ 与 \mathbb{U} 的基向量 $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m$ 正交,因此

$$\langle \mathbf{x} - \pi_\mathbb{U}(\mathbf{x}), \mathbf{b}_i \rangle = 0, \quad i = 1, 2, \dots, m. \quad (3.34)$$

(2) 投影 $\pi_\mathbb{U}(\mathbf{x})$ 必是 \mathbb{U} 中的向量,可由 \mathbb{U} 的基 $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m)$ 来线性表示,即 $\exists \lambda_1,$

$\lambda_2, \dots, \lambda_m \in \mathbb{R}$, 有

$$\pi_U(x) = \sum_{i=1}^m \lambda_i b_i = B\lambda, \quad (3.35)$$

其中 $B = [b_1, b_2, \dots, b_m] \in \mathbb{R}^{n \times m}$, $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_m]^T \in \mathbb{R}^m$ 是 $\pi_U(x)$ 关于基 (b_1, b_2, \dots, b_m) 的坐标向量。

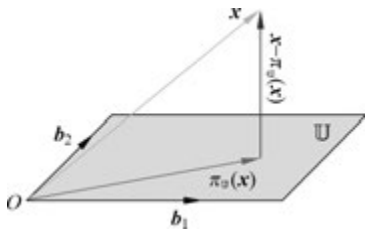


图 3.9 \mathbb{R}^3 到其二维子空间的正交投影

于是,我们可以分步骤来求投影的坐标 $\lambda_1, \lambda_2, \dots, \lambda_m$, 投影向量 $\pi_U(x)$ 和投影矩阵 P_π 。

第 1 步, 求投影的坐标 $\lambda_1, \lambda_2, \dots, \lambda_m$ 。在正交条件式 (3.34) 下, 并结合式 (3.35) 的表示, 有

$$\begin{aligned} b_1^T(x - B\lambda) &= 0, \\ b_2^T(x - B\lambda) &= 0, \\ &\vdots \\ b_m^T(x - B\lambda) &= 0, \end{aligned}$$

从而得到一个齐次线性方程组

$$\begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_m^T \end{bmatrix} [x - B\lambda] = 0 \Leftrightarrow B^T(x - B\lambda) = 0, \quad (3.36)$$

进而有

$$B^T B \lambda = B^T x. \quad (3.37)$$

因为 b_1, b_2, \dots, b_m 是 U 的基向量, 所以矩阵 B 列满秩, 并且 $B^T B$ 是可逆的 (证明留给读者)。于是可由式 (3.37) 解得

$$\lambda = (B^T B)^{-1} B^T x, \quad (3.38)$$

其中矩阵 $(B^T B)^{-1} B^T$ 是 B 的伪逆 (伪逆的概念详见 3.8.3 小节)。

第 2 步, 求投影向量 $\pi_U(x) \in U$ 。由于 $\pi_U(x) = B\lambda$, 结合 (3.38) 式可得

$$\pi_U(x) = B(B^T B)^{-1} B^T x. \quad (3.39)$$

第 3 步, 求投影矩阵 P_π 。将 $\pi_U(x) = P_\pi x$ 与式 (3.39) 相比较, 可得

$$P_\pi = B(B^T B)^{-1} B^T. \quad (3.40)$$

至此, 由式 (3.40) 计算的矩阵 P_π , 可将任意的 $x \in \mathbb{R}^n$ 正交地投影到 m 维子空间 U 上, 投影向量为 $P_\pi x \in U$ 。

【注 3.12】 本小节中一般子空间的正交投影矩阵公式适用于 3.8.1 小节的 1 维子空

间。若 $\dim(\mathbb{U})=1$, 则 $\mathbf{B}^T\mathbf{B}$ 是一个实数, 此时式(3.40)就可以写成 $\mathbf{P}_\pi = \frac{\mathbf{B}\mathbf{B}^T}{\mathbf{B}^T\mathbf{B}}$, 这正是式(3.33)中的投影矩阵。

【例 3.10】(到 2 维子空间上的正交投影)

对于 \mathbb{R}^3 的子空间 $\mathbb{U} = \text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}\right)$ 及向量 $\mathbf{x} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^3$, 将 \mathbf{x} 正交投影到子空间 \mathbb{U}

中, 求投影坐标 λ , 投影向量 $\pi_{\mathbb{U}}(\mathbf{x})$ 和投影矩阵 \mathbf{P}_π 。

【解】 容易判断 $[1 \ 1 \ 1]^T$ 和 $[0 \ 1 \ 2]^T$ 线性无关, 于是 \mathbb{U} 的生成集就是 \mathbb{U} 的一个基, 把 \mathbb{U} 的基向量写成矩阵

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}。$$

计算矩阵 $\mathbf{B}^T\mathbf{B}$ 和向量 $\mathbf{B}^T\mathbf{x}$, 有

$$\mathbf{B}^T\mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}, \quad \mathbf{B}^T\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}。$$

根据式(3.38), 得投影坐标

$$\lambda = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix};$$

根据式(3.35), 得投影向量

$$\pi_{\mathbb{U}}(\mathbf{x}) = \mathbf{B}\lambda = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix},$$

相应的投影误差用原始向量与其在 \mathbb{U} 上的投影的差向量的范数来度量, 即

$$\|\mathbf{x} - \pi_{\mathbb{U}}(\mathbf{x})\| = \|[1 \ 2 \ -1]^T\| = \sqrt{6};$$

根据式(3.40), 得投影矩阵

$$\mathbf{P}_\pi = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}。$$

可以分两步验证 \mathbf{P}_π 为正交投影: ①验证 $\mathbf{x} - \mathbf{P}_\pi\mathbf{x}$ 与 \mathbb{U} 的所有基向量正交; ②验证 $\mathbf{P}_\pi^2 = \mathbf{P}_\pi$ 。

【注 3.13】 投影向量 $\pi_{\mathbb{U}}(\mathbf{x})$ 虽然位于 \mathbb{R}^n 的 m 维子空间 \mathbb{U} 中, 但 $\pi_{\mathbb{U}}(\mathbf{x})$ 仍是 \mathbb{R}^n 中的向量。表示投影向量只需要 m 个坐标 $\lambda_1, \lambda_2, \dots, \lambda_m$ 。

【注 3.14】(PCA 的正交投影的特殊性) 回顾第 1 章介绍的 PCA 降维, 事实上就是找到一个正交投影矩阵将高维数据投影到低维子空间中, 在式(1.39)

$$\mathbf{z}'_n = \mathbf{W}'\mathbf{W}'^T\mathbf{z}_n$$

中, $\mathbf{z}_n \in \mathbb{R}^D$, $\mathbf{z}'_n \in \mathbb{R}^d$, 找到的投影矩阵就是 $\mathbf{P}_\pi = \mathbf{W}'\mathbf{W}'^T$ 。因为 $(\mathbf{W}'^T, \mathbf{W}')^{-1} = \mathbf{E}$, 所以这个

P_π 符合式(3.40)。对同一组原始数据而言,PCA 这个正交投影的特殊之处在于子空间中的投影方差是最大的(同时重构误差是最小的),而一般的正交投影不对子空间做特殊要求,如图 3.10 所示。

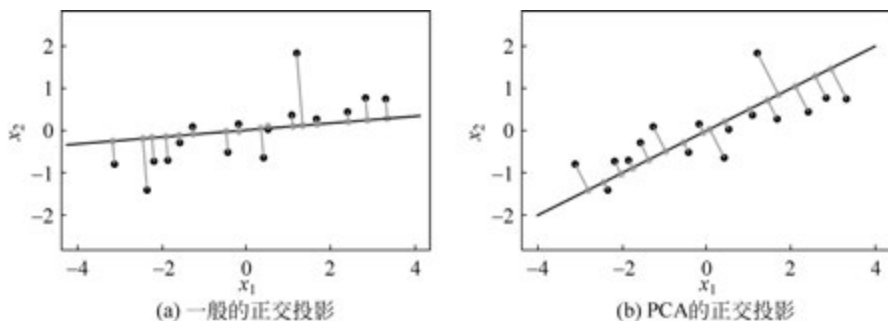


图 3.10 一般的正交投影与 PCA 的正交投影

3.8.3 线性方程组的最小二乘解

2.2.6 小节对线性方程组 $Ax=b$ 的解做了进一步的讨论,将其写成如下形式

$$x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n = b,$$

其中 $A=[\alpha_1, \alpha_2, \cdots, \alpha_n]$, $x=[x_1, x_2, \cdots, x_n]^T$, 并得出了结论: 线性方程组 $Ax=b$ 有解等价于常数向量 b 可由系数矩阵 A 的各列向量线性表示,换句话说, b 存在于 A 的各列向量张成的子空间中; 而 $Ax=b$ 无解等价于 b 不属于 A 的各列向量张成的子空间。

现在,使用本章关于正交投影的知识,就可以近似求解无解的线性方程组 $Ax=b$ 。虽然向量 b 不属于 A 的各列向量张成的子空间,但可以在该子空间中找到最接近 b 的向量,即 b 在该子空间中的正交投影,记为 b' 。这样通过求解线性方程组 $Ax=b'$ 得到原线性方程组的近似解(称为最小二乘解)。具体来说,由原线性方程组 $Ax=b$ 可得

$$A^T Ax = A^T b, \quad (3.41)$$

式(3.41)称为法方程。若 $A \in \mathbb{R}^{m \times n}$ 列满秩,则 $A^T A \in \mathbb{R}^{n \times n}$ 且 $r(A^T A) = r(A) = n$ 。于是 $A^T A$ 可逆,法方程式(3.41)的解为

$$x = (A^T A)^{-1} A^T b. \quad (3.42)$$

式(3.42)即原线性方程组 $Ax=b$ 的最小二乘解,其中矩阵 $(A^T A)^{-1} A^T$ 称为 A 的伪逆。一般来说, A 不是方阵,无法保证有逆矩阵,但只要 A 列满秩(此时 $A^T A$ 可逆),便可求其伪逆。在实际应用中,我们常给 $A^T A$ 加一个偏差项 ϵE 来保证可逆性在数值上的稳定。

3.8.4 施密特正交化

向量空间的基一般不是正交的。然而,为了方便计算内积,我们常常需要得到规范正交基。本小节就是来介绍,从非正交的基向量出发,可以使用施密特(Schmidt)正交化方法来构建规范正交基。

施密特正交化方法的核心是正交投影,可将 n 维向量空间 V 的任意基 (b_1, b_2, \cdots, b_n) 经过如下迭代构造为 V 的一个正交基 (u_1, u_2, \cdots, u_n) 。

$$\begin{aligned} u_1 &= b_1, \\ u_k &= b_k - \pi_{\text{span}(u_1, \cdots, u_{k-1})}(b_k), \quad k = 2, 3, \cdots, n, \end{aligned} \quad (3.43)$$

其中 $\pi_{\text{span}(u_1, \dots, u_{k-1})}(b_k)$ 表示第 k 个基向量 b_k 到由前 $k-1$ 个已构造出的正交向量 u_1, \dots, u_{k-1} 张成的子空间的投影, 用 b_k 减去该投影得到向量 u_k , u_k 与 u_1, \dots, u_{k-1} 张成的 $k-1$ 维子空间正交, 最终得到的 (u_1, u_2, \dots, u_n) 就是 V 的一个正交基。进一步将 u_1, u_2, \dots, u_n 单位化, 得到

$$\epsilon_k = \frac{u_k}{\|u_k\|}, \quad k=1, 2, \dots, n, \quad (3.44)$$

显然 $\|\epsilon_k\|=1, k=1, 2, \dots, n$ 。 $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ 就是 V 的一个规范正交基。

【例 3.11】 如图 3.11(a) 所示, \mathbb{R}^2 中的一个基为

$$b_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

使用施密特正交化方法, 可构造出如下的 \mathbb{R}^2 中的一个正交基 (u_1, u_2) (用点积作为内积)。

$$u_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix},$$

$$u_2 = b_2 - \pi_{\text{span}(u_1)}(b_2) = b_2 - \frac{u_1 u_1^T}{\|u_1\|^2} b_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

这些步骤如图 3.11(b) 和 (c) 所示。显然 $u_1^T u_2 = 0$, 即 u_1, u_2 是正交的。

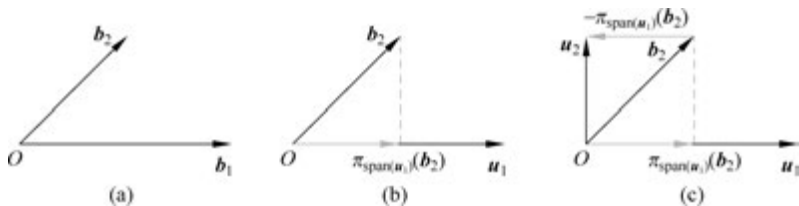


图 3.11 施密特正交化过程

3.8.5 仿射空间上的正交投影

前面我们学习了向量空间向低维子空间的投影, 接下来, 将讨论如何将向量投影到仿射空间中。

如图 3.12(a) 所示, 给定仿射空间 $L = x_0 + U$, 其中 b_1, b_2 是向量空间 U 的基向量。为确定向量 x 在 L 中的正交投影 $\pi_L(x)$, 需要先把问题转化为确定在向量空间中的投影。

不难看出, 任意向量 x 相对于 L 的位置就等同于 $x - x_0$ 相对于 $L - x_0$ 的位置, 而 $L - x_0 = U$ 为向量空间, 故先求 $x - x_0$ 在向量空间 U 中的正交投影, 即 $\pi_U(x - x_0)$, 如图 3.12(b) 所示。

对向量空间 U 做整体平移, 使其零元从原点移动到 x_0 的终点, 就得到了仿射空间 L , 如图 3.12(c) 所示, U 上的投影 $\pi_U(x - x_0)$ 跟随 U 的这个平移, 就得到了 x 在仿射空间 L 上的正交投影的定义

$$\pi_L(x) = x_0 + \pi_U(x - x_0). \quad (3.45)$$

由式 (3.45) 知, $\pi_L(x)$ 是 x 在仿射空间 $L = x_0 + U$ 中的正交投影, 也就是说 $\pi_L(x) - x_0$ 是 $x - x_0$ 在向量空间 U 中的正交投影。显然, 仿射空间上的“正交投影”这个术语, 是从它的方向空间借来的。

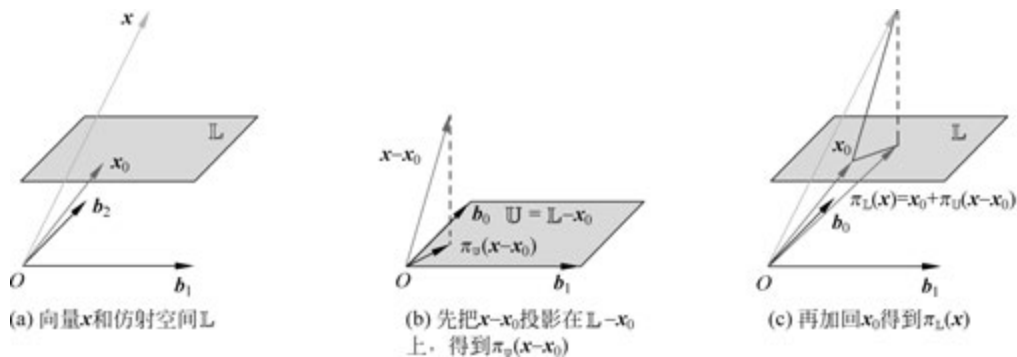


图 3.12 仿射空间上的投影

3.9 旋转

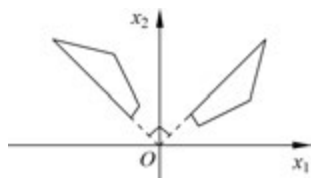
旋转在计算机图形学和机器人学等领域有着重要的应用。旋转是一种线性变换,也是欧氏空间 \mathbb{R}^n 中的自同构变换(一一映射的线性变换)。3.4节的注3.9中已介绍了, \mathbb{R}^n 中刚体(形状和大小都不变的几何体)围绕原点的旋转是一种正交变换, \mathbb{R}^2 中的旋转矩阵形如

$$\text{Rot}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

对于 $\forall x \in \mathbb{R}^2$, x 围绕原点逆时针旋转 θ 角度得到的向量就是 $\text{Rot}(\theta)x$ 。图3.13展示了几何体围绕原点旋转 $\frac{\pi}{2}$, 其所对应的旋转矩阵是

$$\text{Rot}(\theta) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

接下来,将由浅入深地介绍 $\mathbb{R}^2, \mathbb{R}^3, \mathbb{R}^n$ 中的旋转矩阵。

图 3.13 $\theta = \frac{\pi}{2}$ 时几何体的旋转

3.9.1 2维欧氏空间中的旋转

欧氏空间 \mathbb{R}^2 的标准基是 $\left\{e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$,它确定了 \mathbb{R}^2 中的标准坐标系。把这个坐标系逆时针旋转 θ 角,如图3.14所示,旋转后的向量依旧是线性无关的,因而仍然是 \mathbb{R}^2 的基。这意味着旋转是一个基变换。

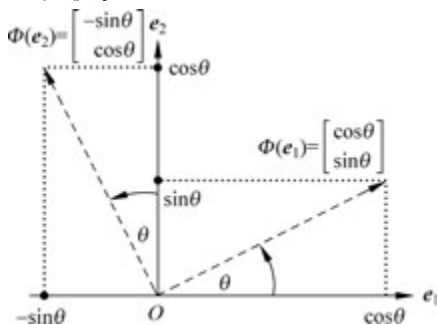


图 3.14 旋转是一个基变换

因为旋转 Φ 是线性变换,所以可用它的变换矩阵[称为旋转矩阵,记为 $\text{Rot}(\theta)$]来表示,观察图 3.14 中的三角函数关系,旋转后的基向量为

$$\Phi(e_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad \Phi(e_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix};$$

又因为 $[e_1 \ e_2]$ 就是单位矩阵 E_2 , 所以

$$[\Phi(e_1) \ \Phi(e_2)] = [e_1 \ e_2] \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad (3.46)$$

比较式(3.46)和式(2.48),由变换矩阵的定义,可知旋转矩阵为

$$R(\theta) = [\Phi(e_1) \ \Phi(e_2)] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \quad (3.47)$$

3.9.2 3 维欧氏空间中的旋转

与 \mathbb{R}^2 中的旋转不同, \mathbb{R}^3 中的旋转可以围绕任意一个过原点的轴,并且任一点的旋转过程都在垂直于旋转轴的平面(2 维子空间)内。如图 3.15(a)所示,旋转轴为过原点的任意向量 l , U 为与 l 垂直的过原点的平面, θ 为旋转角。尽管 \mathbb{R}^3 中的旋转轴方向是任意的,我们也可以使用类似于 3.9.1 小节的方法来确定一般旋转矩阵 R : 先确定标准基 e_1, e_2, e_3 旋转后的向量,即正交的 Re_1, Re_2, Re_3 , 然后组合 Re_1, Re_2, Re_3 就可以得到旋转矩阵 R 。当然,第一步确定标准基旋转后的向量并非易事。

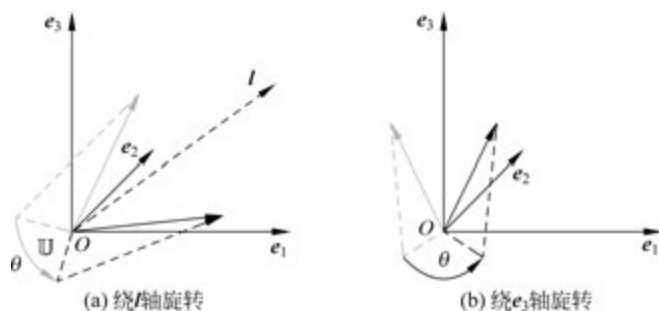


图 3.15 一般旋转和绕标准基的基向量旋转

在 3 维或更高维空间中,我们用右手螺旋法则来确定绕某一轴旋转的正方向,即让右手大拇指指向旋转轴的方向,其余四指握拳指向的方向为旋转的正方向。图 3.15(a)中,右手大拇指指向旋转轴 l 的方向,其余四指握拳指向的方向为旋转角 θ 的方向;图 3.15(b)中,旋转轴为 e_3 轴, θ 的方向就是旋转的正方向。

如果以 \mathbb{R}^3 中的标准基的基向量 e_i 为旋转轴,如图 3.15(b)所示,那么任意向量在旋转时其坐标的第 i 个分量就是固定的, $i=1, 2, 3$ 。这个特殊性凸显了标准基的基向量作为旋转轴的价值,有必要明确其旋转矩阵。

(1) 绕 e_1 轴旋转 θ 角: $\forall x \in \mathbb{R}^3$, 变换为 $\Phi_1(x) = R_1(\theta)x$, 旋转矩阵

$$R_1(\theta) = [\Phi_1(e_1) \ \Phi_1(e_2) \ \Phi_1(e_3)] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}, \quad (3.48)$$

显然 $\Phi_1(\mathbf{e}_1) = \mathbf{e}_1$, 向量旋转后的 \mathbf{e}_1 坐标不变, 旋转正方向就是 $\mathbf{e}_2\mathbf{e}_3$ 坐标平面上的逆时针方向。

(2) 绕 \mathbf{e}_2 轴旋转 θ 角: $\forall \mathbf{x} \in \mathbb{R}^3$, 变换为 $\Phi_2(\mathbf{x}) = \mathbf{R}_2(\theta)\mathbf{x}$, 旋转矩阵

$$\mathbf{R}_2(\theta) = [\Phi_2(\mathbf{e}_1) \quad \Phi_2(\mathbf{e}_2) \quad \Phi_2(\mathbf{e}_3)] = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}, \quad (3.49)$$

显然 $\Phi_2(\mathbf{e}_2) = \mathbf{e}_2$, 向量旋转后的 \mathbf{e}_2 坐标不变, 旋转正方向就是 $\mathbf{e}_3\mathbf{e}_1$ 坐标平面上的逆时针方向。

(3) 绕 \mathbf{e}_3 轴旋转 θ 角: $\forall \mathbf{x} \in \mathbb{R}^3$, 变换为 $\Phi_3(\mathbf{x}) = \mathbf{R}_3(\theta)\mathbf{x}$, 旋转矩阵

$$\mathbf{R}_3(\theta) = [\Phi_3(\mathbf{e}_1) \quad \Phi_3(\mathbf{e}_2) \quad \Phi_3(\mathbf{e}_3)] = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (3.50)$$

显然 $\Phi_3(\mathbf{e}_3) = \mathbf{e}_3$, 向量旋转后的 \mathbf{e}_3 坐标不变, 旋转正方向就是 $\mathbf{e}_1\mathbf{e}_2$ 坐标平面上的逆时针方向。

3.9.3 n 维欧氏空间中的旋转

作为 \mathbb{R}^3 的推广, \mathbb{R}^n 中的旋转也可以围绕任意一个过原点的轴, 在正交于旋转轴的 2 维子空间内进行。然而, n 维欧氏空间中确定一般旋转矩阵比较困难, 我们更感兴趣于标准基的基向量为旋转轴的情况。此时旋转可以理解为固定坐标的 $n-2$ 维, 只旋转剩余的 2 个基向量确定的坐标平面。3.9.2 小节已经在 \mathbb{R}^3 中做过类似的事情。以下给出这种特殊旋转的定义。

定义 3.12 (Givens 旋转) 设 \mathbb{V} 是 n 维欧氏空间, 若对于 $1 \leq i < j \leq n, i, j = 1, 2, \dots, n$, $\theta \in \mathbb{R}$, 自同构 $\Phi_{ij}: \mathbb{V} \rightarrow \mathbb{V}$ 具有变换矩阵

$$\mathbf{R}_{ij}(\theta) = \begin{bmatrix} \mathbf{E}_{i-1} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \cos\theta & \mathbf{0} & -\sin\theta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}_{j-i-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sin\theta & \mathbf{0} & \cos\theta & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{E}_{n-j} \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad (3.51)$$

则 $\mathbf{R}_{ij}(\theta)$ 称为 **Givens 旋转** 或 **Givens 矩阵**。本质上 $\mathbf{R}_{ij}(\theta)$ 是由单位矩阵 \mathbf{E}_n 做如下改变得来的:

$$r_{ii} = \cos\theta, \quad r_{ij} = -\sin\theta, \quad r_{ji} = \sin\theta, \quad r_{jj} = \cos\theta.$$

【注 3.15】 (3.51) 式中的 $\mathbf{0}$ 表示全零块, 在不同的位置上可能是零矩阵、零向量或仅一个 0。

Givens 矩阵当 $n=2$ 时就是式 (3.47) 中的旋转矩阵, 当 $n=3$ 时就是式 (3.48)~式 (3.50) 中的矩阵。

【注 3.16】 \mathbb{R}^n 中的 Givens 旋转以标准基的一个基向量为旋转轴, 向量在旋转时其坐标中的 $n-2$ 维是固定的, 比一般旋转大大简化。应用中可以把围绕任意过原点的轴的一般旋转用若干 Givens 旋转的复合来实现, 也就是把对应的一般旋转矩阵分解成若干 Givens

矩阵的乘积。如图 3.16 所示,其中图 3.16(b)是图 3.16(a)左下角的局部放大,可以看到,图 3.16(a)中围绕 l 轴旋转 θ 等价于图 3.16(b)中 3 个 Givens 旋转的复合: e_1e_2 坐标面内旋转 φ_1 , e_2e_3 坐标面内旋转 φ_2 , e_3e_1 坐标面内旋转 φ_3 。

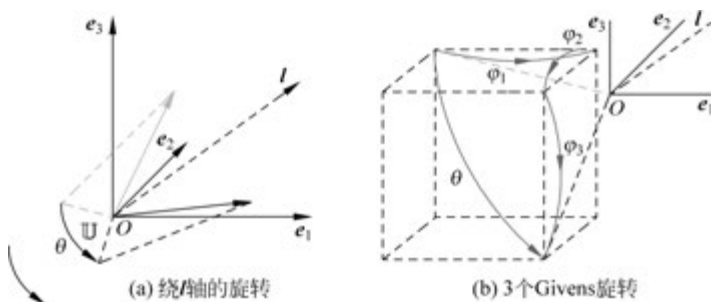


图 3.16 \mathbb{R}^3 中的一般旋转可由 3 个 Givens 旋转的复合来实现

3.9.4 旋转的性质

由于旋转矩阵是一种正交矩阵,旋转具有如下性质。

(1) 旋转保持距离不变,即 $\forall x, y$ 有 $\|x - y\| = \|R_\theta x - R_\theta y\|$, 任意两点之间的距离旋转之后保持不变。

(2) 旋转保持夹角不变,即 $R_\theta x$ 和 $R_\theta y$ 之间的夹角与 x 和 y 之间的夹角相等。

(3) 3 维或更高维向量的旋转通常不能交换顺序。因此,实际应用中旋转的顺序很重要,即使围绕同一旋转轴也不能交换顺序。只有在 2 维空间中向量旋转是可交换的,即 $\forall \phi, \theta \in [0, 2\pi], R(\phi)R(\theta) = R(\theta)R(\phi)$ 。

习题

1. 已知 $\langle \cdot, \cdot \rangle$: 对任意 $x = [x_1 \ x_2]^T \in \mathbb{R}^2, y = [y_1 \ y_2]^T \in \mathbb{R}^2$, 有

$$\langle x, y \rangle = x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2x_2 y_2,$$

证明 $\langle \cdot, \cdot \rangle$ 是一个内积。

2. 设 \mathbb{R}^2 中 $\langle \cdot, \cdot \rangle$: 对任意 $x, y \in \mathbb{R}^2$, 有

$$\langle x, y \rangle = x^T \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} y,$$

问 $\langle \cdot, \cdot \rangle$ 是点积吗?

3. 设

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad y = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix},$$

使用以下内积计算 x 与 y 之间的距离。

(1) $\langle x, y \rangle = x^T y$ 。

(2) $\langle x, y \rangle = x^T A y$, 其中 $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ 。

4. 设

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -1 \\ -1 \end{bmatrix},$$

使用以下内积计算 \mathbf{x} 与 \mathbf{y} 之间角度。

(1) $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ 。

(2) $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{A} \mathbf{y}$, 其中 $\mathbf{B} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$ 。

5. 证明正交矩阵的乘积仍为正交矩阵。

6. 欧氏空间 \mathbb{R}^5 默认以点积为内积, 其子空间 $\mathbf{U} \subseteq \mathbb{R}^5$ 和向量 $\mathbf{x} \in \mathbb{R}^5$ 分别为

$$\mathbf{U} = \text{span} \left(\begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -3 \\ 5 \\ 0 \\ 7 \end{bmatrix} \right), \quad \mathbf{x} = \begin{bmatrix} -1 \\ -9 \\ -1 \\ 4 \\ 1 \end{bmatrix}.$$

(1) 求 \mathbf{x} 在子空间 \mathbf{U} 中的正交投影 $\pi_{\mathbf{U}}(\mathbf{x})$ 。

(2) 求距离 $d(\mathbf{x}, \mathbf{U})$ 。

7. 设欧氏空间 \mathbb{R}^3 的内积为

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \mathbf{y},$$

并设 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ 为 \mathbb{R}^3 的标准基。

(1) 求 \mathbf{e}_2 在子空间 $\mathbf{U} = \text{span}(\mathbf{e}_1, \mathbf{e}_3)$ 中的投影 $\pi_{\mathbf{U}}(\mathbf{e}_2)$ 。

(2) 计算距离 $d(\mathbf{e}_2, \mathbf{U})$ 。

(3) 画出标准基向量和 $\pi_{\mathbf{U}}(\mathbf{e}_2)$ 。

8. 使用施密特正交化方法, 把 2 维子空间 $\mathbf{U} \subseteq \mathbb{R}^3$ 的基 $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2)$ 变换为规范正交基 $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2)$, 其中

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}.$$

9. 设 $x_1, x_2, \dots, x_n > 0$, 满足 $x_1 + x_2 + \dots + x_n = 1$, 利用柯西-施瓦茨不等式证明

(1) $\sum_{i=1}^n x_i^2 \geq \frac{1}{n}$ 。

(2) $\sum_{i=1}^n \frac{1}{x_i} \geq n^2$ 。

提示: 使用 \mathbb{R}^n 中的点积, 选择特定的向量 \mathbf{x}, \mathbf{y} , 运用柯西-施瓦茨不等式。

10. 设

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix},$$

将 $\mathbf{x}_1, \mathbf{x}_2$ 旋转 30° , 求出旋转矩阵及旋转后的向量。