9 Orthogonal Projection
正交投影

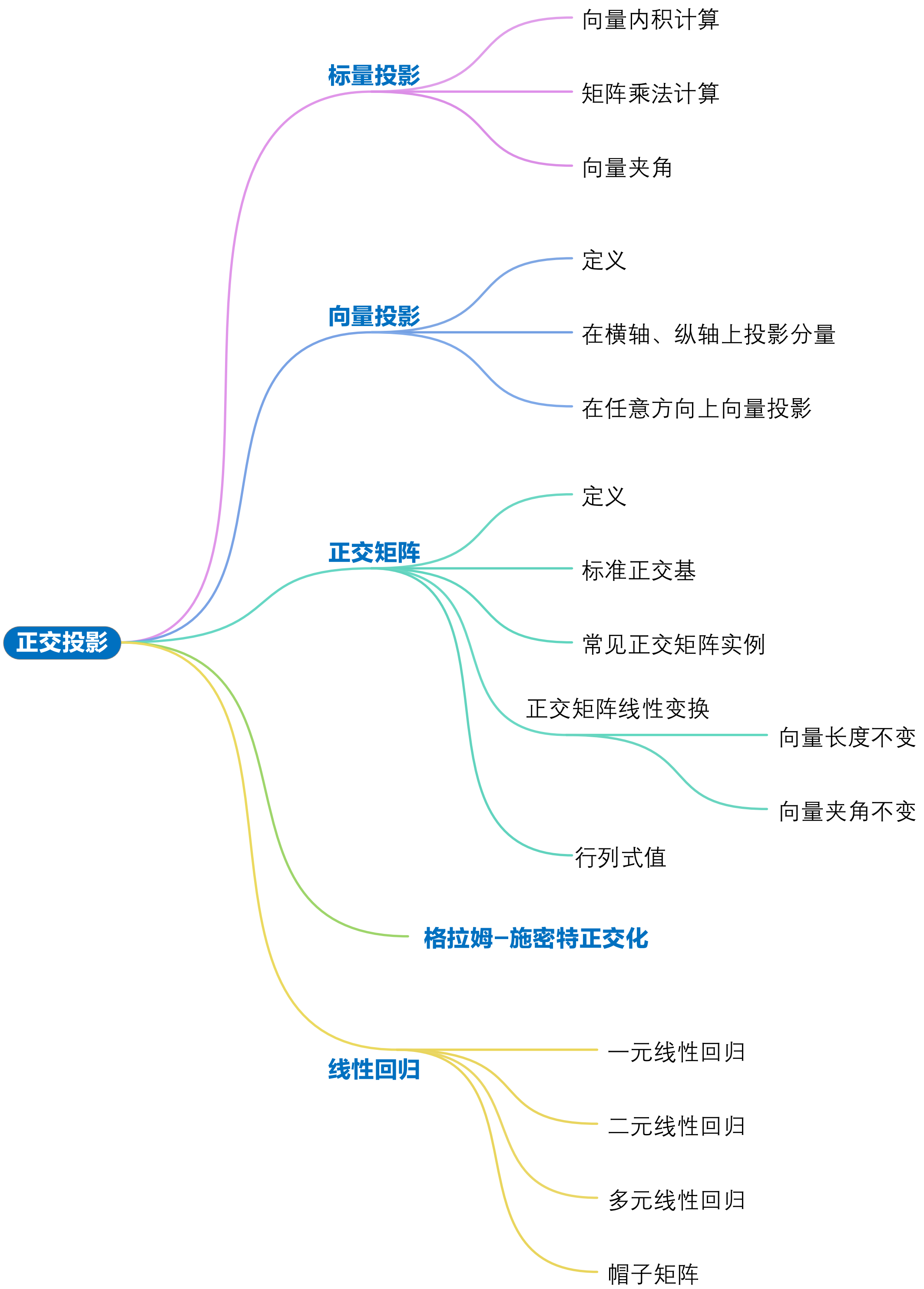
应用几乎无处不在

数学好比给了人类第六感。

Mathematics seems to endow one with something like a new sense.

—— 查尔斯·达尔文 (Charles Darwin) | 进化论之父 | 1809 ~ 1882

◀ `numpy.random.randn()` 生成满足标准正态分布的随机数
◀ `numpy.linalg.qr()` QR 分解
◀ `seaborn.heatmap()` 绘制热图


正交投影
标量投影
向量内积计算
矩阵乘法计算
向量夹角
向量投影
定义
在横轴、纵轴上投影分量
在任意方向上向量投影
正交矩阵
定义
标准正交基
常见正交矩阵实例
正交矩阵线性变换
向量长度不变
向量夹角不变
行列式值
格拉姆–施密特正交化
线性回归
一元线性回归
二元线性回归
多元线性回归
帽子矩阵

9.1 标量投影：结果为标量

正交

打个比方，正交投影类似正午头顶阳光将物体投影到地面上，如图 1 所示。此时，假设光线之间相互平行，和地面垂直。

把列向量 $\boldsymbol{x}$ 看成是一根木杆，而列向量 $\boldsymbol{v}$ 方向代表地面水平方向。$\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上的投影结果为 $\boldsymbol{z}$。很明显 $\boldsymbol{x}-\boldsymbol{z}$ 垂直于 $\boldsymbol{v}$，因此两者向量内积为 0：

$$(\boldsymbol{x}-\boldsymbol{z})\cdot\boldsymbol{v}=0 \tag{1}$$

用矩阵乘法，(1) 可以写成，

$$(\boldsymbol{x}-\boldsymbol{z})^{\mathrm{T}}\boldsymbol{v}=0 \tag{2}$$

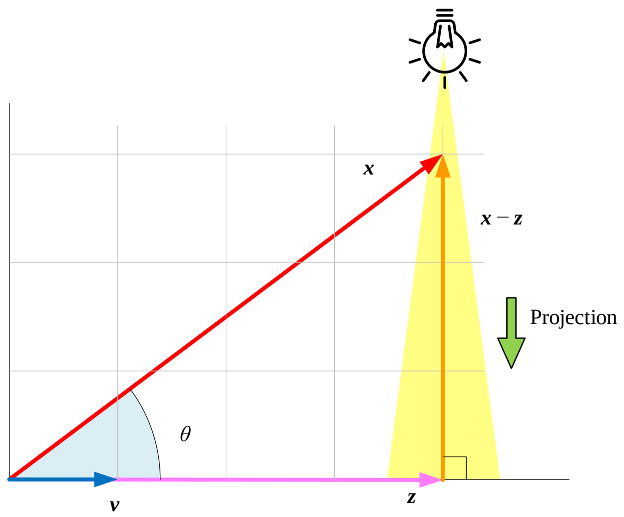


图 1. 正交投影的意义

由于 $\boldsymbol{z}$ 和非零向量 $\boldsymbol{v}$ 共线，因此 $\boldsymbol{z}$ 与 $\boldsymbol{v}$ 的单位向量共线，它们之间的关系为：

$$\boldsymbol{z}=s\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \tag{3}$$

其中，s 为标量，s 常被称作 $\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上的**标量投影** (scalar projection)。

将 (3) 代入 (2) 得到：

$$\left(\boldsymbol{x} - s\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}\right)^{\mathrm{T}}\boldsymbol{v} = 0 \tag{4}$$

(4) 经过整理得到 s 的解析式，也就是 $\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上的标量投影为：

$$s = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}}{\|\boldsymbol{v}\|} \tag{5}$$

上式可以写成如下四种形式：

$$s = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{x}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{v}\cdot\boldsymbol{x}}{\|\boldsymbol{v}\|} \tag{6}$$

⚠ 注意，$\boldsymbol{x}$ 和 $\boldsymbol{v}$ 为等长列向量。

特别地，如果 $\boldsymbol{v}$ 本身就是单位向量，(6) 可以写作：

$$s = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}\cdot\boldsymbol{v} = \boldsymbol{v}\cdot\boldsymbol{x} \tag{7}$$

本系列丛书，一般会用 $\boldsymbol{e}$、$\boldsymbol{v}$、$\boldsymbol{u}$ 等代表单位向量。

向量夹角

下面介绍从向量夹角入手推导标量投影。

如图 1 所示，向量 $\boldsymbol{x}$ 和 $\boldsymbol{v}$ 的夹角为 θ，这个夹角的余弦值 $\cos\theta$ 可以通过下式求解：

$$\cos\theta = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}}{\|\boldsymbol{x}\|\|\boldsymbol{v}\|} = \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}}{\|\boldsymbol{x}\|\|\boldsymbol{v}\|} = \frac{\boldsymbol{x}\cdot\boldsymbol{v}}{\|\boldsymbol{x}\|\|\boldsymbol{v}\|} = \frac{\boldsymbol{v}\cdot\boldsymbol{x}}{\|\boldsymbol{x}\|\|\boldsymbol{v}\|} \tag{8}$$

而 $\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上的标量投影 s 便是向量 $\boldsymbol{x}$ 的模乘以 $\cos\theta$：

$$s = \|\boldsymbol{x}\|\cos\theta = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{x}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{v}\cdot\boldsymbol{x}}{\|\boldsymbol{v}\|} \tag{9}$$

这样，我们便得到和 (6) 一致的结果。

⚠ 注意，两两向量夹角是“相对角度”；类比的话，极坐标系中的相对极轴的角度可以称为“绝对角度”。

9.2 向量投影：结果为向量

相对标量投影，我们更经常使用**向量投影** (vector projection)。

顾名思义，向量投影就是在标量投影 s 基础上再加 $\boldsymbol{v}$ 的方向，即标量投影 s 乘以 $\boldsymbol{v}$ 的单位向量。因此，$\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上的向量投影为：

$$\text{proj}_{\boldsymbol{v}}(\boldsymbol{x}) = s\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{x}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v} = \frac{\boldsymbol{v}\cdot\boldsymbol{x}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v} = \frac{\boldsymbol{x}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v} = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}\boldsymbol{v} = \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}\boldsymbol{v} \tag{10}$$

用尖括号<>表达标量积，$\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上的向量投影可以记做：

$$\text{proj}_{\boldsymbol{v}}(\boldsymbol{x}) = \frac{\langle\boldsymbol{x},\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\boldsymbol{v} \tag{11}$$

特别地，如果 $\boldsymbol{v}$ 为单位向量，$\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上的向量投影则可以写成：

$$\text{proj}_{\boldsymbol{v}}(\boldsymbol{x}) = \langle\boldsymbol{x},\boldsymbol{v}\rangle\boldsymbol{v} = (\boldsymbol{x}\cdot\boldsymbol{v})\boldsymbol{v} = (\boldsymbol{v}\cdot\boldsymbol{x})\boldsymbol{v} = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v})\boldsymbol{v} = (\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x})\boldsymbol{v} \tag{12}$$

举个例子

实际上，获得某一个向量的横、纵轴坐标，或者计算横纵轴的向量分量，也是一个投影过程。下面看一个实例。给定如下列向量 $\boldsymbol{x}$，

$$\boldsymbol{x} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \tag{13}$$

如图 2 所示，列向量 $\boldsymbol{x}$ 既可以代表平面直角坐标系上的一点，也可以代表一个起点为原点 (0, 0)、终点为 (4, 3) 的向量。

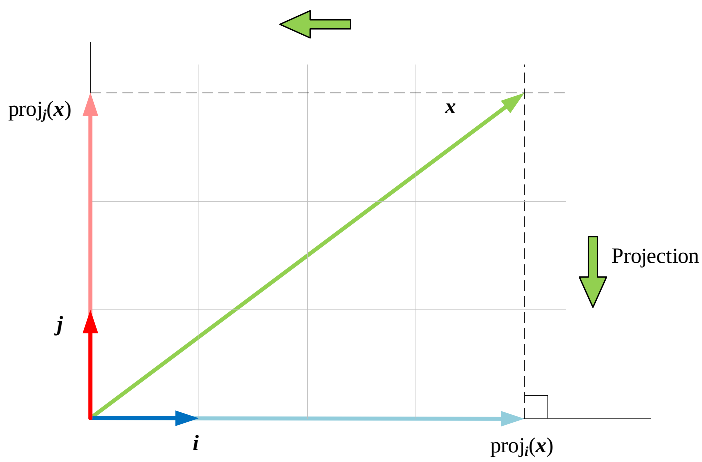


图 2. $\boldsymbol{x}$ 向 $\boldsymbol{i}$ 和 $\boldsymbol{j}$ 投影

$\boldsymbol{x}$ 向单位向量 $\boldsymbol{i} = [1, 0]^{\mathrm{T}}$ 方向上投影得到的标量投影就是 $\boldsymbol{x}$ 横轴坐标：

$$\boldsymbol{i}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{i} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^{\mathrm{T}}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 4 \tag{14}$$

$\boldsymbol{x}$ 向单位向量 $\boldsymbol{j} = [0, 1]^{\mathrm{T}}$ 方向上投影得到的标量投影就是 $\boldsymbol{x}$ 纵轴坐标：

$$j^{\mathrm{T}}x = x^{\mathrm{T}}j = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 3 \tag{15}$$

x 在单位向量 $i = [1, 0]^{\mathrm{T}}$ 方向上向量投影就是 x 在横轴上的分量：

$$\mathrm{proj}_i(x) = (x^{\mathrm{T}}i)i = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} i = 4i \tag{16}$$

x 在单位向量 $j = [0, 1]^{\mathrm{T}}$ 方向上向量投影就是 x 在纵轴上的分量：

$$\mathrm{proj}_j(x) = (x^{\mathrm{T}}j)j = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 \\ 1 \end{bmatrix} j = 3j \tag{17}$$

如果单位向量 v 为，

$$v = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} \tag{18}$$

x 在 v 方向上投影得到的标量投影为：

$$x^{\mathrm{T}}v = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} = 5 = \|x\| \tag{19}$$

如图 3 所示，可以发现，x 和 v 实际上共线，也就是夹角为 0°。这显然是个特例。

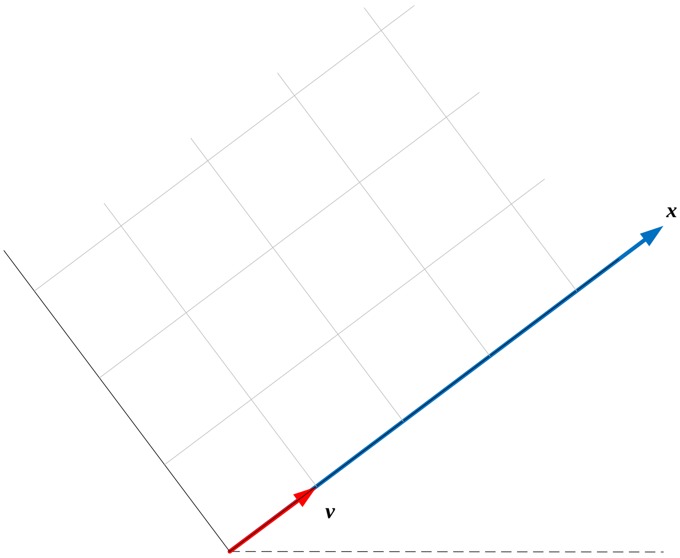


图 3. x 向 v 的投影

推导投影坐标

前文在讲解线性变换时介绍过，点 (x_1, x_2) 在通过原点、切向量为 $\boldsymbol{\tau}$ $[\tau_1, \tau_2]^{\mathrm{T}}$ 直线方向上投影得到的坐标 (z_1, z_2)，计算式如下：

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \frac{1}{\|\boldsymbol{\tau}\|^2} \begin{bmatrix} \tau_1^2 & \tau_1\tau_2 \\ \tau_1\tau_2 & \tau_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{20}$$

下面利用本节知识简单推导 (20)。

$\boldsymbol{x}$ 在 $\boldsymbol{\tau}$ 方向上的向量投影为：

$$\begin{aligned} \boldsymbol{z} &= \frac{\boldsymbol{x}\cdot\boldsymbol{\tau}}{\|\boldsymbol{\tau}\|^2}\boldsymbol{\tau} = \frac{x_1\tau_1 + x_2\tau_2}{\|\boldsymbol{\tau}\|^2}\begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \\ &= \frac{1}{\|\boldsymbol{\tau}\|^2}\begin{bmatrix} (x_1\tau_1 + x_2\tau_2)\tau_1 \\ (x_1\tau_1 + x_2\tau_2)\tau_2 \end{bmatrix} = \frac{1}{\|\boldsymbol{\tau}\|^2}\begin{bmatrix} \tau_1^2 x_1 + \tau_1\tau_2 x_2 \\ \tau_1\tau_2 x_1 + \tau_2^2 x_2 \end{bmatrix} = \frac{1}{\|\boldsymbol{\tau}\|^2}\begin{bmatrix} \tau_1^2 & \tau_1\tau_2 \\ \tau_1\tau_2 & \tau_2^2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{aligned} \tag{21}$$

⚠ 注意，不做特殊说明的话，本书中“投影”都是正交投影。

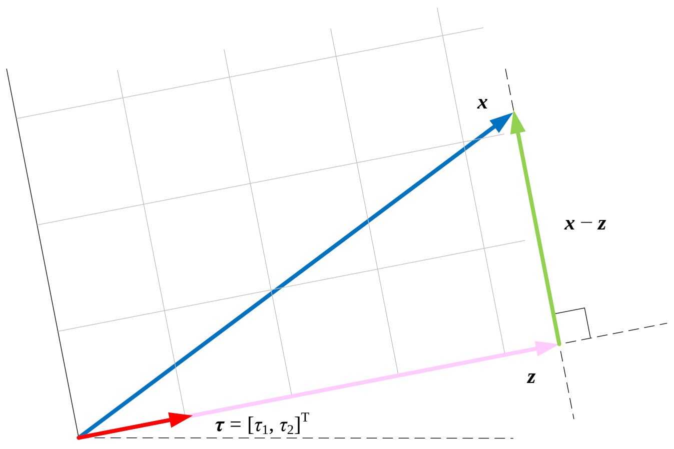


图 4. $\boldsymbol{x}$ 在 $\boldsymbol{\tau}$ 方向上投影

图 5 所示为点 A 向一系列通过原点方向不同直线的投影。

有读者可能会问，空间某点朝任意直线或超平面投影时，如果它们不通过原点，该如何计算投影点的坐标？这个问题将在本书第 19 章揭晓答案。

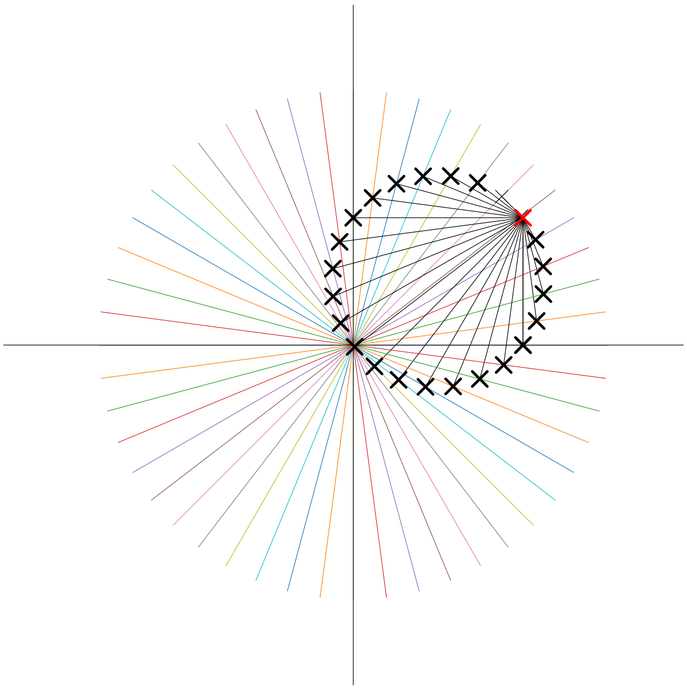


图 5. 点 A 到一系列通过原点的直线投影

向量张量积：无处不在

回过头再看 (12)，令 $\boldsymbol{v}$ 为单位向量，(12) 可以写成如下含有向量张量积的形式：

$$\operatorname{proj}_{\boldsymbol{v}}(\boldsymbol{x})=\underbrace{(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x})}_{\text{Scaler}}\boldsymbol{v}=\boldsymbol{v}\underbrace{(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x})}_{\text{Scaler}}=\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}=(\boldsymbol{v}\otimes\boldsymbol{v})_{2\times 2}\boldsymbol{x} \tag{22}$$

我们称 $\boldsymbol{v}\otimes\boldsymbol{v}$ 为投影矩阵 (projection matrix)。

利用向量张量积，(21) 可以写成：

$$\boldsymbol{z}=\frac{1}{\|\boldsymbol{\tau}\|^{2}}(\boldsymbol{\tau}\otimes\boldsymbol{\tau})_{2\times 2}\boldsymbol{x}=\left(\frac{\boldsymbol{\tau}}{\|\boldsymbol{\tau}\|}\otimes\frac{\boldsymbol{\tau}}{\|\boldsymbol{\tau}\|}\right)_{2\times 2}\boldsymbol{x}=(\hat{\boldsymbol{\tau}}\otimes\hat{\boldsymbol{\tau}})_{2\times 2}\boldsymbol{x} \tag{23}$$

其中，$\hat{\boldsymbol{\tau}}$ 代表 $\boldsymbol{\tau}$ 的单位向量。

一般情况，数据矩阵 $\boldsymbol{X}$ 中样本点的坐标值以行向量表达，$\boldsymbol{X}$ 向单位向量 $\boldsymbol{v}$ 方向投影坐标为：

$$\boldsymbol{Z}=\boldsymbol{X}\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}=\boldsymbol{X}(\boldsymbol{v}\otimes\boldsymbol{v})_{2\times 2} \tag{24}$$

请大家格外注意 (24)，我们下一节还要继续聊。此外，它也是下一章要讨论的核心内容。

`Bk4_Ch9_01.py` 绘制图 5。

9.3 正交矩阵：一个规范正交基

回顾"猪引发的投影问题"

本章前文介绍的是朝一个向量方向投影，比如向量 $\boldsymbol{x}$ 向 $\boldsymbol{v}$ 方向投影，这可以视作 $\boldsymbol{x}$ 向 $\boldsymbol{v}$ 张起的空间 span($\boldsymbol{v}$) 投影。

同理，向量也可以向一个有序基构造的平面投影。这个有序基可以是正交基，可以是非正交基。而数据科学和机器学习实践中，最常用的基底是规范正交基。正交矩阵的列向量就是范正交基向量。

本系列丛书《数学要素》聊过向量向一个平面投影。鸡兔同笼三步曲中，我们聊了农夫和需求 $\boldsymbol{y}$ (10 只兔、10 只鸡、5 只猪) 和"A-B 套餐"平面的关系，具体如图 6 所示。

$\boldsymbol{w}_1$ 和 $\boldsymbol{w}_2$ 张起 A-B 套餐"平面为 H = span($\boldsymbol{w}_1$, $\boldsymbol{w}_2$)，图 6 中 $\boldsymbol{y}$ 向 span($\boldsymbol{w}_1$, $\boldsymbol{w}_2$) 投影。而 span($\boldsymbol{w}_1$, $\boldsymbol{w}_2$) 的有序基为 [$\boldsymbol{w}_1$, $\boldsymbol{w}_2$]。请大家自行验证有序基 [$\boldsymbol{w}_1$, $\boldsymbol{w}_2$] 为非正交基。

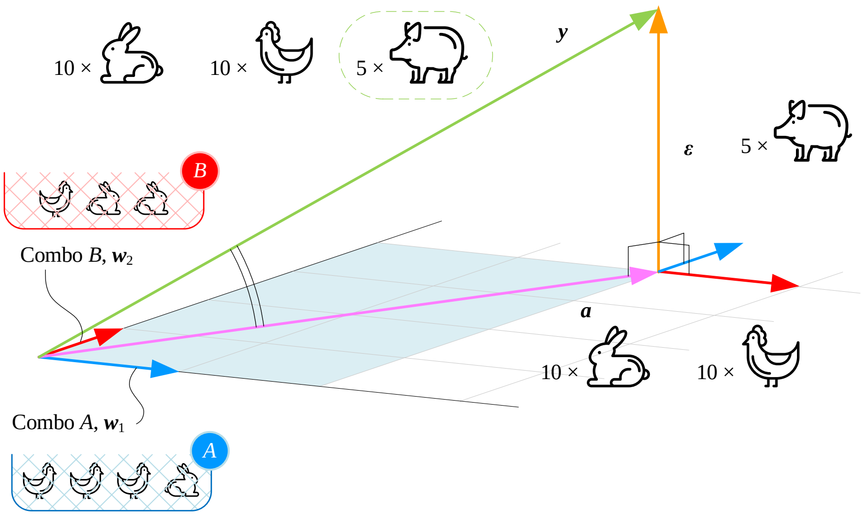


图 6. 农夫的需求和小贩提供的"A-B 套餐"平面存在 5 只猪的距离，来自本系列丛书《数学要素》

正交矩阵

满足下式的方阵 $\boldsymbol{V}$ 为**正交矩阵** (orthogonal matrix)：

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{I} \tag{25}$$

正交矩阵基本性质:

$$\begin{aligned}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} &= \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{I} \\ \boldsymbol{V}^{\mathrm{T}} &= \boldsymbol{V}^{-1}\end{aligned} \tag{26}$$

(26)中两式经常使用，必须烂熟于心。

举个实例，图 7 所示热图为一个 4 × 4 正交矩阵 $\boldsymbol{V}$ 和自己转置 $\boldsymbol{V}^{\mathrm{T}}$ 乘积为单位阵 $\boldsymbol{I}$。

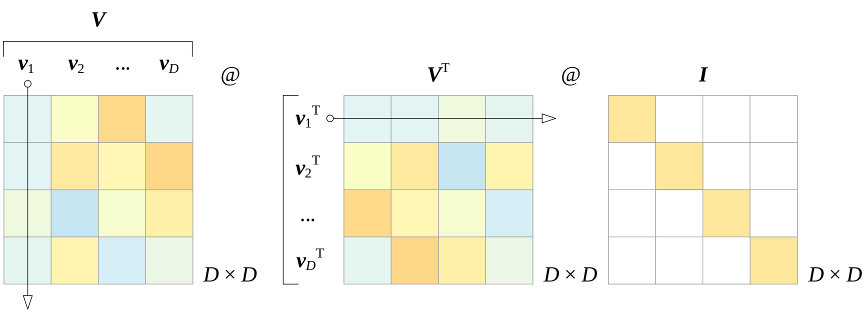


图 7. 正交阵 $\boldsymbol{V}$ 和自己转置 $\boldsymbol{V}^{\mathrm{T}}$ 乘积为单位阵 $\boldsymbol{I}$

前文的例子

其实我们已经接触过几种正交矩阵。本书前文讲过的矩阵 $\boldsymbol{I}$、$\boldsymbol{R}$、$\boldsymbol{T}$ 和 $\boldsymbol{P}$ 都是正交矩阵:

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \boldsymbol{T} = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}, \quad \boldsymbol{P} = \begin{bmatrix} & & & 1 \\ 1 & & & \\ & & 1 & \\ & 1 & & \end{bmatrix} \tag{27}$$

其中，$\boldsymbol{I}$ 为单位矩阵，$\boldsymbol{R}$ 作用是旋转，$\boldsymbol{T}$ 作用是镜像，$\boldsymbol{P}$ 是置换矩阵。

本书前文提到的如下两个矩阵也是正交矩阵:

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \tag{28}$$

它们都满足，方阵和自身转置乘积为单位阵，即:

$$\begin{aligned}
V^{\mathrm{T}}V &= \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}\begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
W^{\mathrm{T}}W &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\end{aligned} \tag{29}$$

矩阵乘法第一视角展开

将 (25) 中矩阵 $\boldsymbol{V}$ 写成一排列向量：

$$V_{D\times D} = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,D} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n,1} & v_{n,2} & \cdots & v_{n,D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D \end{bmatrix} \tag{30}$$

(25) 左侧可以写成：

$$V^{\mathrm{T}}V = \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}} \end{bmatrix}\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D \end{bmatrix} \tag{31}$$

(31) 展开得到：

$$V^{\mathrm{T}}V = \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_1 & \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_D \\ \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{v}_1 & \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{v}_D \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{v}_1 & \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{v}_D \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{32}$$

主对角线结果为 1，即，

$$\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j = \boldsymbol{v}_j \cdot \boldsymbol{v}_j = \left\|\boldsymbol{v}_j\right\|^2 = 1 \tag{33}$$

也就是说，矩阵 $\boldsymbol{V}$ 的每个列向量 $\boldsymbol{v}_j$ 为单位向量。

(32) 主对角线以外元素均为 0：

$$\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j = 0, \quad i \neq j \tag{34}$$

即 $\boldsymbol{V}$ 中任意两个列向量两两正交，即垂直。

至此，可以判定 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D\}$ 为规范正交基。写成有序基形式，就是矩阵 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$。$\boldsymbol{V}$ 张起一个 D 维超平面 $\mathrm{span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D)$。

大家应该已经意识到，(32) 就是 V^TV 矩阵乘法的第一视角。

批量化计算向量模和夹角

此外，(32) 告诉我们"批量"计算一系列向量模和两两夹角的方式——Gram 矩阵！

举个例子，给定矩阵 X，将其写成一系列列向量 $X = [x_1, x_2, \ldots, x_D]$。$X$ 的 Gram 矩阵为：

$$G = \begin{bmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & \cdots & x_1 \cdot x_D \\ x_2 \cdot x_1 & x_2 \cdot x_2 & \cdots & x_2 \cdot x_D \\ \vdots & \vdots & \ddots & \vdots \\ x_D \cdot x_1 & x_D \cdot x_2 & \cdots & x_D \cdot x_D \end{bmatrix} = \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \cdots & \langle x_1, x_D \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \cdots & \langle x_2, x_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_D, x_1 \rangle & \langle x_D, x_2 \rangle & \cdots & \langle x_D, x_D \rangle \end{bmatrix} \tag{35}$$

借助向量夹角余弦展开 G 中向量积：

$$G = \begin{bmatrix} \|x_1\|\|x_1\|\cos\theta_{1,1} & \|x_1\|\|x_2\|\cos\theta_{2,1} & \cdots & \|x_1\|\|x_D\|\cos\theta_{1,D} \\ \|x_2\|\|x_1\|\cos\theta_{1,2} & \|x_2\|\|x_2\|\cos\theta_{2,2} & \cdots & \|x_2\|\|x_D\|\cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|x_D\|\|x_1\|\cos\theta_{1,D} & \|x_D\|\|x_2\|\cos\theta_{2,D} & \cdots & \|x_D\|\|x_D\|\cos\theta_{D,D} \end{bmatrix} \tag{36}$$

观察矩阵 G，它包含了数据矩阵 X 中列向量的两个重要信息——模 $\|x_i\|$、方向 (向量两两夹角 $\cos\theta_{i,j}$)。再次强调，$\theta_{i,j}$ 为相对角度。

我们将会在本书第 12 章讲解 Cholesky 分解时继续深入讲解这一话题。

矩阵乘法第二视角展开

有了第一视角，大家自然会想到矩阵乘法的第二视角。

还是将 V 写成 $[v_1, v_2, \ldots, v_D]$，VV^T 则可以按如下方式展开：

$$VV^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_D \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_D^T \end{bmatrix} = v_1 v_1^T + v_2 v_2^T + \cdots v_D v_D^T = I_{D \times D} \tag{37}$$

(37) 可以写成张量积之和形式：

$$VV^T = v_1 \otimes v_1 + v_2 \otimes v_2 + \cdots v_D \otimes v_D = I_{D \times D} \tag{38}$$

上一节 (24) 对应数据矩阵 X 向单位向量 v 方向投影。如果 X 向规范正交基 $V = [v_1, v_2, \ldots, v_D]$ 投影，对应的运算则为：

$$\begin{aligned} X_{n\times D}VV^{\mathrm{T}} &= X\left(v_1\otimes v_1 + v_2\otimes v_2 + \cdots + v_D\otimes v_D\right)_{2\times 2} \\ &= \underbrace{Xv_1}_{z_1}\otimes v_1 + \underbrace{Xv_2}_{z_2}\otimes v_2 + \cdots + \underbrace{Xv_D}_{z_D}\otimes v_D \\ &= X_{n\times D}I_{D\times D} \\ &= X_{n\times D} \end{aligned} \tag{39}$$

大家可能已经糊涂了，上式折腾了半天，最后得到的还是原数据矩阵 X 本身！这实际上是原矩阵 X 在 $\mathbb{R}^D$ 中 $[v_1, v_2, \ldots, v_D]$ 这个特定规范正交基进行分解。

(39) 已经非常接近本书第 15、16 章要讲解的奇异值分解的思路。希望大家能搞清楚 (39) 背后的数学思想。

再进一步，如图 8 所示，下式代表一个规范正交基对单位矩阵的分解：

$$I_{D\times D} = v_1\otimes v_1 + v_2\otimes v_2 + \cdots + v_D\otimes v_D = \sum_{j=1}^{D} v_j\otimes v_j \tag{40}$$

其中，每个 $v_j\otimes v_j$ 都是一个特定方向的投影矩阵 (projection matrix)。这个视角同样重要，本章和下一章还将继续深入讨论。

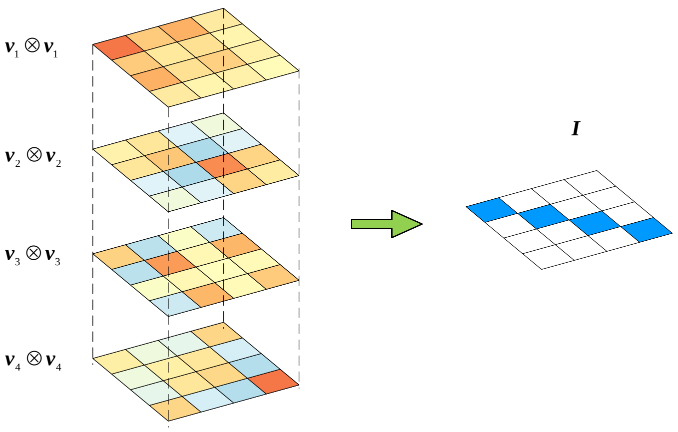


图 8. 对单位矩阵的分解

9.4 规范正交基性质

本节以 (28) 矩阵 V 为例介绍更多规范正交基的性质。

坐标

将 V 分解成两个列向量，

$$v_1 = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}, \quad v_2 = \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \tag{41}$$

这两个向量长度为 1，都是单位向量。

显然，V 这个矩阵的转置和 V 本身乘积是一个 2 × 2 单位矩阵。用矩阵乘法第一视角展开 $V^{\mathrm{T}}V$ 得到：

$$V^{\mathrm{T}}V = \begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} v_1^{\mathrm{T}}v_1 & v_1^{\mathrm{T}}v_2 \\ v_2^{\mathrm{T}}v_1 & v_2^{\mathrm{T}}v_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I \tag{42}$$

给定列向量 $x = [4, 3]^{\mathrm{T}}$。如图 9 (a) 所示，x 在标准正交基 $[e_1, e_2]$ 中的坐标为 (4, 3)。

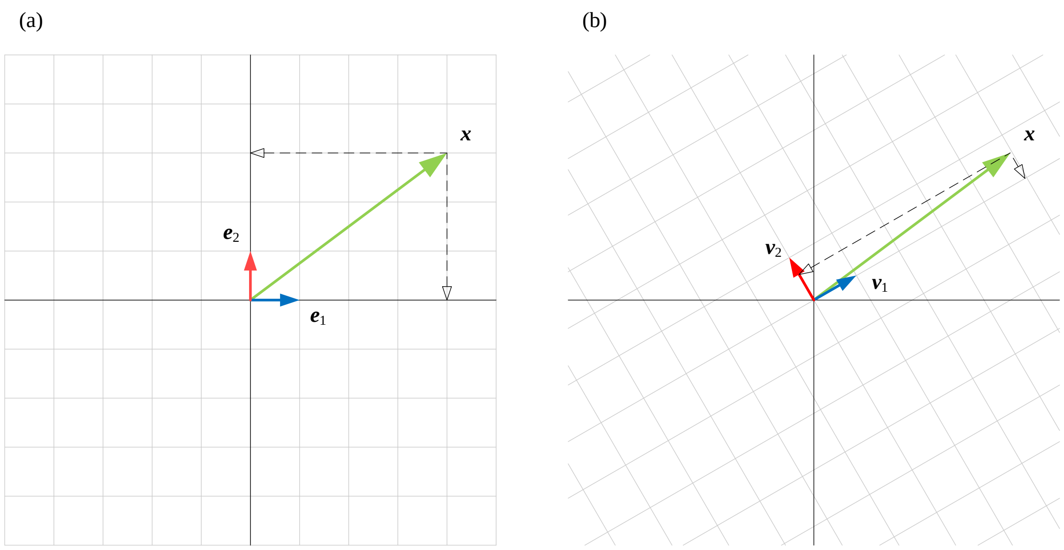


图 9. x 在不同规范正交系中的坐标

如图 9 (b) 所示，将 x 投影到这个 V 规范正交系中，得到的结果就是在 $[v_1, v_2]$ 这个规范正交系的坐标：

$$V^{\mathrm{T}}x = \begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \end{bmatrix} x = \begin{bmatrix} v_1^{\mathrm{T}}x \\ v_2^{\mathrm{T}}x \end{bmatrix} = \begin{bmatrix} \mathrm{proj}_{v_1}(x) \\ \mathrm{proj}_{v_2}(x) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 4.964 \\ 0.598 \end{bmatrix} \tag{43}$$

这说明，向量 x 在 $[v_1, v_2]$ 规范正交系中的坐标为 (4.964, 0.598)。

向量长度不变

经过正交矩阵 $\boldsymbol{V}$ 线性变换后，向量 $\boldsymbol{x}$ 的 L^2 范数，即向量长度，没有变化：

$$\begin{aligned}\left\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\right\|_2^2 &= \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\cdot\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x} = \left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\right)^{\mathrm{T}}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\right) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{x} \\ &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}\cdot\boldsymbol{x} = \left\|\boldsymbol{x}\right\|_2^2\end{aligned} \tag{44}$$

比较图 9 (a) 和 (b) 可以发现，不同规范正交系中 $\boldsymbol{x}$ 的长度确实没有变化。向量 $\boldsymbol{x}$ 在 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 规范正交系中的坐标为 (4.964, 0.598)，计算向量模：

$$\sqrt{4.964^2 + 0.598^2} = \sqrt{4^2 + 3^2} = 5 \tag{45}$$

图 10 所示为给定一点在不同二维规范正交系中的投影结果。

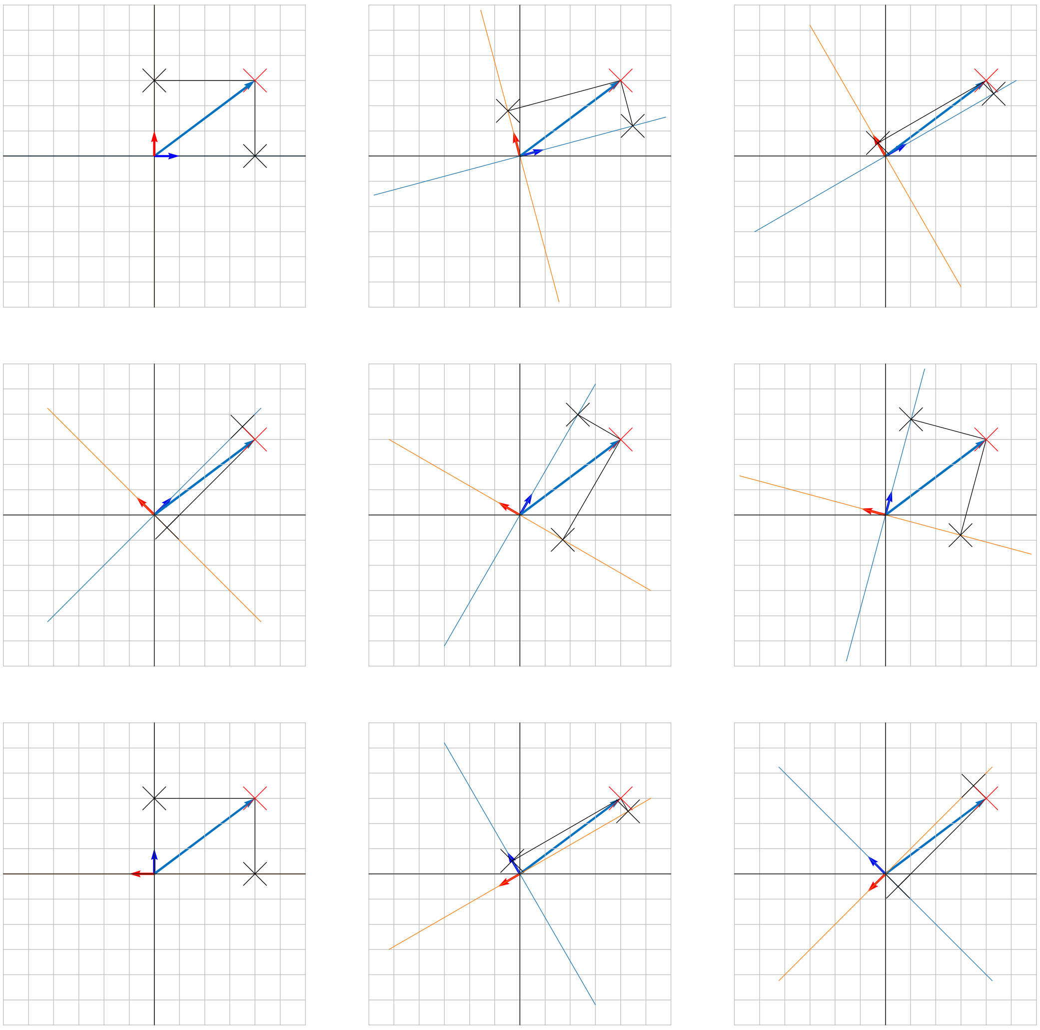


图 10. 平面中向量在不同坐标系的投影

Bk4_Ch9_02.py 绘制图 10。

夹角不变

$\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 经过正交矩阵 $\boldsymbol{V}$ 线性转化得到 $\boldsymbol{z}_i$ 和 $\boldsymbol{z}_j$。$\boldsymbol{z}_i$ 和 $\boldsymbol{z}_j$ 两者夹角等同于 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 夹角:

$$\frac{\boldsymbol{z}_i \cdot \boldsymbol{z}_j}{\|\boldsymbol{z}_i\|\|\boldsymbol{z}_j\|} = \frac{\boldsymbol{z}_i \cdot \boldsymbol{z}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|} = \frac{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i \cdot \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|} = \frac{\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i\right)^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|} = \frac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|} = \frac{\boldsymbol{x}_i \cdot \boldsymbol{x}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|} \tag{46}$$

如图 11 所示，发现正交矩阵 $\boldsymbol{A}$ 线性变换后，$\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 两者角度没有变化。这也不难理解，变化前后，向量都还在 $\mathbb{R}^2$ 中，只不过是坐标参考系发生了旋转，而 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 之间的“相对角度”完全没有发生改变。

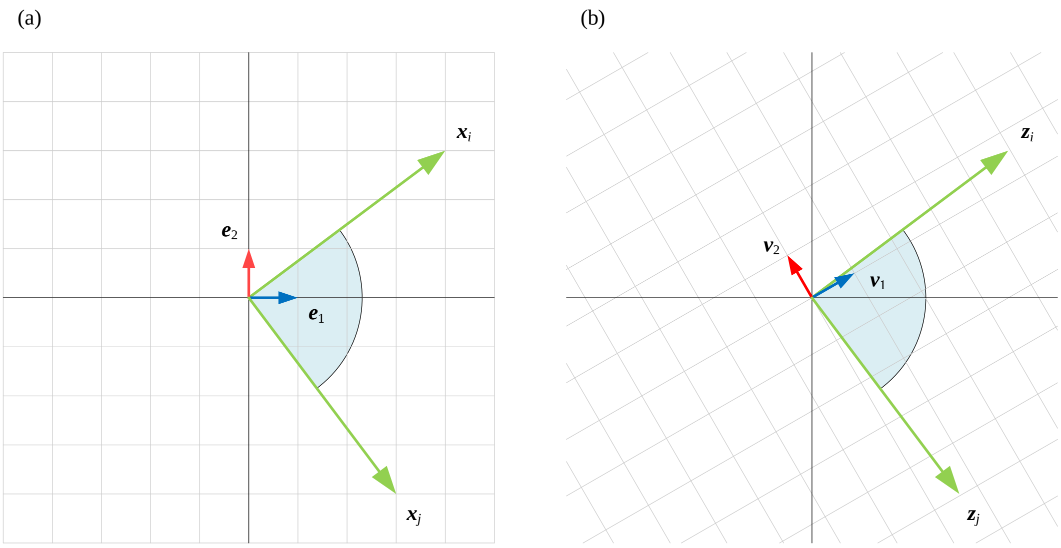


图 11. 不同规范正交系中，$\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 的夹角不变

行列式值

正交矩阵 $\boldsymbol{V}$ 还有一个有趣性质，$\boldsymbol{V}$ 行列式值为 1 或−1:

$$\left(\det\left(\boldsymbol{V}\right)\right)^2 = \det\left(\boldsymbol{V}^{\mathrm{T}}\right)\det\left(\boldsymbol{V}\right) = \det\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\right) = \det\left(\boldsymbol{I}\right) = 1 \tag{47}$$

也就是说，对于二维正交矩阵 $\boldsymbol{V}$，经过 $\boldsymbol{V}$ 线性变换后，面积不变。当 $\det(\boldsymbol{V})$ 为−1 时，图形会发生翻转。

9.5 再谈镜像：从投影视角

上一章聊几何变换时，我们介绍了镜像，并且直接给出完成镜像操作转换矩阵 $\boldsymbol{T}$ 的两种形式，具体如下。

$$\begin{aligned}\boldsymbol{T} &= \frac{1}{\|\boldsymbol{\tau}\|^2}\begin{bmatrix}\tau_1^2-\tau_2^2 & 2\tau_1\tau_2\\ 2\tau_1\tau_2 & \tau_2^2-\tau_1^2\end{bmatrix}\\ \boldsymbol{T} &= \begin{bmatrix}\cos 2\theta & \sin 2\theta\\ \sin 2\theta & -\cos 2\theta\end{bmatrix}\end{aligned} \tag{48}$$

本节用正交投影推导 (48) 中第二个转换矩阵形式。

如图 12 所示，镜像对称轴 l 这条直线通过原点，直线切向量 $\boldsymbol{\tau}$ 为：

$$\boldsymbol{\tau} = \begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} \tag{49}$$

向量 $\boldsymbol{x}$ 关于对称轴 l 得到镜像为 $\boldsymbol{z}$。

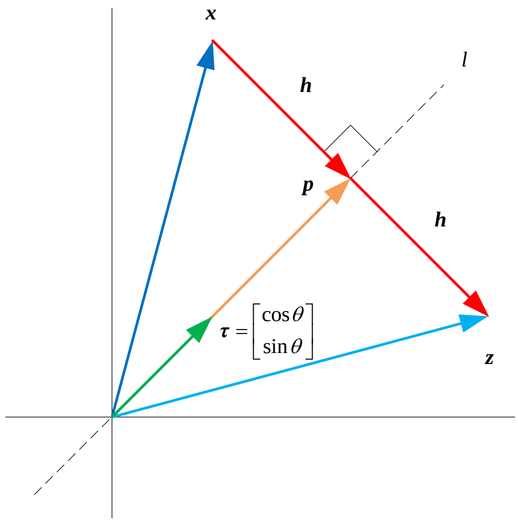


图 12. 投影视角看镜像

从投影角度来看，向量 $\boldsymbol{x}$ 在 $\boldsymbol{\tau}$ 方向投影为向量 $\boldsymbol{p}$。利用张量积 (投影矩阵) 形式，向量 $\boldsymbol{p}$ 可以写成：

$$\boldsymbol{p} = (\boldsymbol{\tau}\otimes\boldsymbol{\tau})\boldsymbol{x} \tag{50}$$

将 (49) 代入 (50)，整理得到：

$$p = (\tau \otimes \tau) x = \begin{bmatrix} \cos\theta\cos\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin\theta\sin\theta \end{bmatrix} x \tag{51}$$

利用三角恒等式，上式可以整理为：

$$p = \begin{bmatrix} (\cos 2\theta + 1)/2 & \sin 2\theta/2 \\ \sin 2\theta/2 & (1 - \cos 2\theta)/2 \end{bmatrix} x \tag{52}$$

令向量 $\boldsymbol{h}$ 为 $\boldsymbol{p}$ 和 $\boldsymbol{x}$ 之差，即：

$$h = p - x \tag{53}$$

根据正交投影，$\boldsymbol{h}$ 显然垂直 $\boldsymbol{p}$。观察图 12，由于 $\boldsymbol{z}$ 和 $\boldsymbol{x}$ 为镜像关系，因此两者之差为 $2\boldsymbol{h}$，也就是下式成立：

$$z = x + 2h \tag{54}$$

将 (53) 代入 (54) 整理得到：

$$z = 2p - x \tag{55}$$

将 (52) 代入 (55) 得到：

$$\begin{aligned} z &= 2\begin{bmatrix} (\cos 2\theta + 1)/2 & \sin 2\theta/2 \\ \sin 2\theta/2 & (1 - \cos 2\theta)/2 \end{bmatrix} x - Ix \\ &= \begin{bmatrix} 2\times(\cos 2\theta + 1)/2 - 1 & 2\times\sin 2\theta/2 \\ 2\times\sin 2\theta/2 & 2\times(1 - \cos 2\theta)/2 - 1 \end{bmatrix} x \\ &= \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} x \end{aligned} \tag{56}$$

这样，我们便用投影视角推导得到 (48) 中第二个镜像转换矩阵。请大家自行推导 (48) 中第一个镜像转换矩阵。

9.6 格拉姆–施密特正交化

格拉姆-施密特正交化 (Gram-Schmidt orthogonalization) 是求解欧氏空间规范正交基的一种方法。整个过程用到核心数学工具就是正交投影。

给定非正交 D 个线性不相关的向量 $[\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_D]$，通过格拉姆-施密特正交化，可以得到 D 个单位正交向量 $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \ldots, \boldsymbol{q}_D\}$，它们可以构造一个规范正交基 $[\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \ldots, \boldsymbol{q}_D]$。

正交化过程

格拉姆-施密特正交化过程如下所示：

$$
\begin{aligned}
&\boldsymbol{\eta}_1 = \boldsymbol{x}_1 \\
&\boldsymbol{\eta}_2 = \boldsymbol{x}_2 - \operatorname{proj}_{\boldsymbol{\eta}_1}(\boldsymbol{x}_2) \\
&\boldsymbol{\eta}_3 = \boldsymbol{x}_3 - \operatorname{proj}_{\boldsymbol{\eta}_1}(\boldsymbol{x}_3) - \operatorname{proj}_{\boldsymbol{\eta}_2}(\boldsymbol{x}_3) \\
&\cdots \\
&\boldsymbol{\eta}_D = \boldsymbol{x}_D - \sum_{j=1}^{D-1} \operatorname{proj}_{\boldsymbol{\eta}_j}(\boldsymbol{x}_D)
\end{aligned}
\tag{57}
$$

前两步

图 13 所示为格拉姆-施密特正交化前两步。

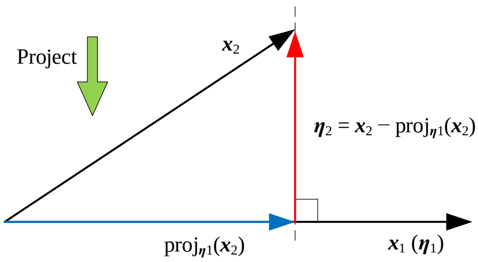


图 13. 格拉姆-施密特正交化前两步

获得 $\boldsymbol{\eta}_1$ 很容易，只需要 $\boldsymbol{\eta}_1 = \boldsymbol{x}_1$。

求解 $\boldsymbol{\eta}_2$ 需要利用 $\boldsymbol{\eta}_2$ 垂直于 $\boldsymbol{\eta}_1$ 这一条件，即：

$$
(\boldsymbol{\eta}_1)^{\mathrm{T}} \boldsymbol{\eta}_2 = 0 \tag{58}
$$

如图 13 所示，$\boldsymbol{x}_2$ 在 $\boldsymbol{\eta}_1$ 方向上投影为 $\operatorname{proj}_{\boldsymbol{\eta}1}(\boldsymbol{x}_2)$，剩余的向量分量垂直于 $\boldsymbol{x}_1$ ($\boldsymbol{\eta}_1$)，这个分量就是 $\boldsymbol{\eta}_2$：

$$
\boldsymbol{\eta}_2 = \boldsymbol{x}_2 - \operatorname{proj}_{\boldsymbol{\eta}_1}(\boldsymbol{x}_2) = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{\eta}_1}{\boldsymbol{\eta}_1^{\mathrm{T}}\boldsymbol{\eta}_1}\boldsymbol{\eta}_1 \tag{59}
$$

$\boldsymbol{\eta}_2$ 也有自己的名字，叫正交补 (orthogonal complement)。

下面验证 $\boldsymbol{\eta}_1$ 和 $\boldsymbol{\eta}_2$ 相互垂直：

$$
\begin{aligned}
(\boldsymbol{\eta}_1)^{\mathrm{T}} \boldsymbol{\eta}_2 &= (\boldsymbol{x}_1)^{\mathrm{T}} \left(\boldsymbol{x}_2 - \frac{\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{\eta}_1}{\boldsymbol{\eta}_1^{\mathrm{T}}\boldsymbol{\eta}_1}\boldsymbol{\eta}_1 \right) \\
&= \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 - \frac{\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_1}{\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1} \\
&= \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 - \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_1 = 0
\end{aligned}
\tag{60}
$$

第三步

如图 14 所示，第三步是 $\boldsymbol{x}_3$ 向 $[\boldsymbol{\eta}_1, \boldsymbol{\eta}_2]$ 张成的平面投影。令 $\boldsymbol{\eta}_3$ 为 $\boldsymbol{x}_3$ 中不在 $[\boldsymbol{\eta}_1, \boldsymbol{\eta}_2]$ 平面上的成分，即：

$$\boldsymbol{\eta}_3 = \boldsymbol{x}_3 - \text{proj}_{\boldsymbol{\eta}_1}(\boldsymbol{x}_3) - \text{proj}_{\boldsymbol{\eta}_2}(\boldsymbol{x}_3) \tag{61}$$

显然，$\boldsymbol{\eta}_3$ 垂直 $\text{span}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$，也就是说 $\boldsymbol{\eta}_3$ 分别垂直 $\boldsymbol{\eta}_1$ 和 $\boldsymbol{\eta}_2$。按此思路，不断反复投影直至得到所有正交向量 $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \ldots, \boldsymbol{\eta}_D\}$。

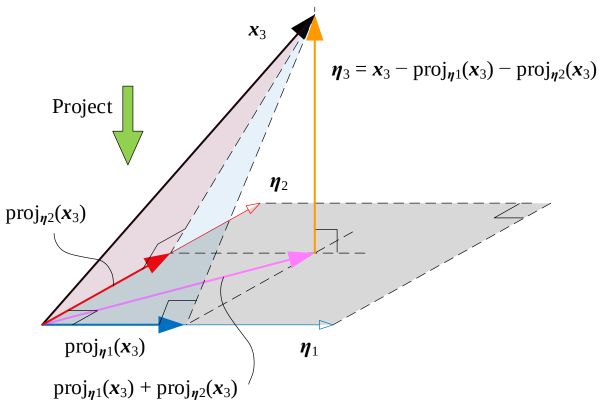


图 14. 格拉姆-施密特正交化第三步

单位化

最后单位化 (归一化) 获得单位正交向量 $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \ldots, \boldsymbol{q}_D\}$：

$$\boldsymbol{q}_1 = \frac{\boldsymbol{\eta}_1}{\|\boldsymbol{\eta}_1\|}, \quad \boldsymbol{q}_2 = \frac{\boldsymbol{\eta}_2}{\|\boldsymbol{\eta}_2\|}, \quad \boldsymbol{q}_3 = \frac{\boldsymbol{\eta}_3}{\|\boldsymbol{\eta}_3\|}, \quad \cdots, \quad \boldsymbol{q}_D = \frac{\boldsymbol{\eta}_D}{\|\boldsymbol{\eta}_D\|} \tag{62}$$

举个实例

给定 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 两个向量，利用格拉姆-施密特正交化获得两个正交向量：

$$\boldsymbol{x}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{63}$$

$\boldsymbol{\eta}_1$ 就是 $\boldsymbol{x}_1$，即，

$$\boldsymbol{\eta}_1 = \boldsymbol{x}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \tag{64}$$

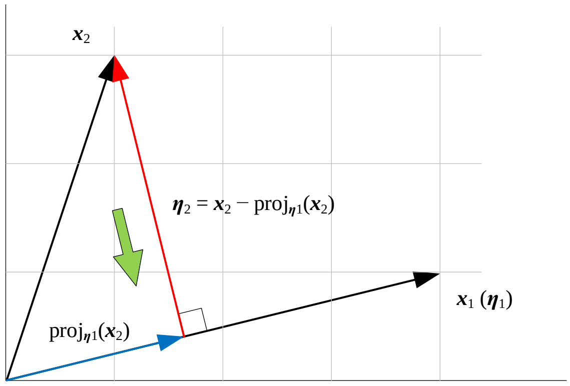


图 15. 格拉姆-施密特正交化第三步

x_2 在 η_1 (x_1) 方向上投影，得到向量投影：

$$\text{proj}_{\eta_1}(x_2) = \frac{x_2 \cdot \eta_1}{\eta_1 \cdot \eta_1}\eta_1 = \frac{4\times1+1\times3}{4\times4+1\times1}\times\begin{bmatrix}4\\1\end{bmatrix} = \frac{7}{17}\times\begin{bmatrix}4\\1\end{bmatrix} \tag{65}$$

计算 η_2

$$\begin{aligned}\eta_2 &= x_2 - \text{proj}_{\eta_1}(x_2)\\ &= \begin{bmatrix}1\\3\end{bmatrix} - \frac{7}{17}\times\begin{bmatrix}4\\1\end{bmatrix} = \frac{1}{17}\times\begin{bmatrix}-11\\44\end{bmatrix}\end{aligned} \tag{66}$$

最后对 η_1 和 η_2 单位化，得到 q_1 和 q_2：

$$\begin{cases} q_1 = \dfrac{\eta_1}{\|\eta_1\|} = \dfrac{1}{\sqrt{17}}\begin{bmatrix}4\\1\end{bmatrix} \\ q_2 = \dfrac{\eta_2}{\|\eta_2\|} = \dfrac{1}{\sqrt{17}}\begin{bmatrix}-1\\4\end{bmatrix} \end{cases} \tag{67}$$

格拉姆-施密特正交化可以通过 QR 分解完成，这是第 11 章矩阵分解要讲解的内容之一。

9.7 投影视角看线性回归

本系列丛书《数学要素》中，我们在鸡兔同笼三部曲中简单介绍过如何通过投影视角理解线性回归。本节在此基础上展开讲解。

一元线性回归

列向量 $\boldsymbol{y}$ 在 $\boldsymbol{x}$ 方向上正交投影得到向量 $\hat{\boldsymbol{y}}$。向量差 $\boldsymbol{y}-\hat{\boldsymbol{y}}$ 垂直于 $\boldsymbol{x}$。据此构造如下等式:

$$\boldsymbol{x}^{\mathrm{T}}\left(\boldsymbol{y}-\hat{\boldsymbol{y}}\right)=0 \tag{68}$$

显然 $\hat{\boldsymbol{y}}$ 和 $\boldsymbol{x}$ 共线，因此下式成立:

$$\hat{\boldsymbol{y}}=b\boldsymbol{x} \tag{69}$$

其中，b 为实数系数。大家在上式中是否已经看到线性回归的影子?

从数据角度思考，$\boldsymbol{x}$ 为因变量，$\boldsymbol{y}$ 为自变量。数据 $\boldsymbol{x}$ 方向能够解释 $\boldsymbol{y}$ 的一部分，即 $\hat{\boldsymbol{y}}$。不能解释的部分就是**残差** (residuals)，即 $\boldsymbol{y}-\hat{\boldsymbol{y}}$。

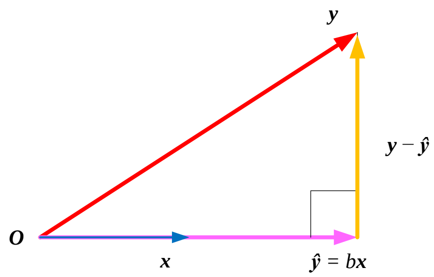


图 16. 向量 $\boldsymbol{y}$ 向 $\boldsymbol{x}$ 正交投影得到向量投影 $\hat{\boldsymbol{y}}$

将 (69) 代入 (68)，得到:

$$\boldsymbol{x}^{\mathrm{T}}\left(\boldsymbol{y}-b\boldsymbol{x}\right)=0 \tag{70}$$

容易求得系数 b 为:

$$b=\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \tag{71}$$

从而，$\hat{\boldsymbol{y}}$ 为:

$$\hat{\boldsymbol{y}}=\boldsymbol{x}\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \tag{72}$$

这样利用向量投影这个数学工具，我们解释了一元线性回归。

⚠ 注意，在上述分析中，我们没有考虑常数项。也就是说，线性回归模型为比例函数，截距为 0。

二元线性回归

下面我们聊一下二元线性回归。

如图 17 所示，两个线性无关向量 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 张成一个平面 span($\boldsymbol{x}_1$, $\boldsymbol{x}_2$)。向量 $\boldsymbol{y}$ 向该平面投影得到向量 $\hat{\boldsymbol{y}}$。向量 $\hat{\boldsymbol{y}}$ 是 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 线性组合:

$$\hat{\boldsymbol{y}} = b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 = \underbrace{\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}}_{\boldsymbol{X}} \underbrace{\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}_{\boldsymbol{b}} = \boldsymbol{X}\boldsymbol{b} \tag{73}$$

其中，b_1 和 b_2 为系数。

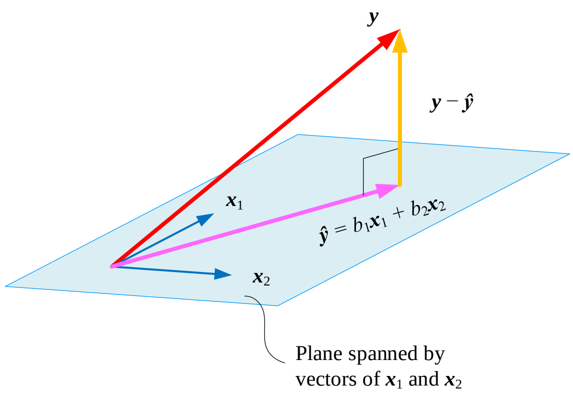


图 17. 向量 $\boldsymbol{y}$ 向平面 span($\boldsymbol{x}_1$, $\boldsymbol{x}_2$) 投影

$\boldsymbol{y} - \hat{\boldsymbol{y}}$ 垂直于 $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2]$，据此构造如下两个等式：

$$\begin{cases} \boldsymbol{x}_1^{\mathrm{T}}\left(\boldsymbol{y} - \hat{\boldsymbol{y}}\right) = 0 \\ \boldsymbol{x}_2^{\mathrm{T}}\left(\boldsymbol{y} - \hat{\boldsymbol{y}}\right) = 0 \end{cases} \Rightarrow \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \end{bmatrix}\left(\boldsymbol{y} - \hat{\boldsymbol{y}}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{74}$$

即，

$$\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{y} - \hat{\boldsymbol{y}}\right) = \boldsymbol{0} \tag{75}$$

将 (73) 代入 (75) 得到：

$$\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}\right) = \boldsymbol{0} \tag{76}$$

从而推导得到 $\boldsymbol{b}$ 的解析式：

$$\boldsymbol{b} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{77}$$

(77) 代入 (73)，可以得到：

$$\boxed{\hat{\boldsymbol{y}} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}} \tag{78}$$

上式中，$\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$ 常被称作**帽子矩阵** (hat matrix)。$\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$ 是我们在本书第 5 章提到的**幂等矩阵** (idempotent matrix)，也就是说下式成立：

$$\left(\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\right)^2 = \boldsymbol{X}\underbrace{\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}}_{\boldsymbol{I}}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}} \tag{79}$$

多元线性回归

以上结论也可以推广到如图 18 所示多元线性回归情形。D 个向量 $\boldsymbol{x}_1$、$\boldsymbol{x}_2$、...、$\boldsymbol{x}_D$ 张成超平面 $H = \text{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$，向量 $\boldsymbol{y}$ 在超平面 H 上投影结果为 $\hat{\boldsymbol{y}}$，即，

$$\hat{\boldsymbol{y}} = b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_D\boldsymbol{x}_D \tag{80}$$

误差 $\boldsymbol{y} - \hat{\boldsymbol{y}}$ 垂直垂直于 $H = \text{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$，也就是说 $\boldsymbol{y} - \hat{\boldsymbol{y}}$ 分别垂直于 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D$，即：

$$\begin{cases} \boldsymbol{x}_1^{\mathrm{T}}(\boldsymbol{y} - \hat{\boldsymbol{y}}) = 0 \\ \boldsymbol{x}_2^{\mathrm{T}}(\boldsymbol{y} - \hat{\boldsymbol{y}}) = 0 \\ \qquad \vdots \\ \boldsymbol{x}_D^{\mathrm{T}}(\boldsymbol{y} - \hat{\boldsymbol{y}}) = 0 \end{cases} \Rightarrow \underbrace{\begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_D^{\mathrm{T}} \end{bmatrix}}_{\boldsymbol{X}^{\mathrm{T}}} (\boldsymbol{y} - \hat{\boldsymbol{y}}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{81}$$

用之前的推导思路，我们也可以得到 (78)。

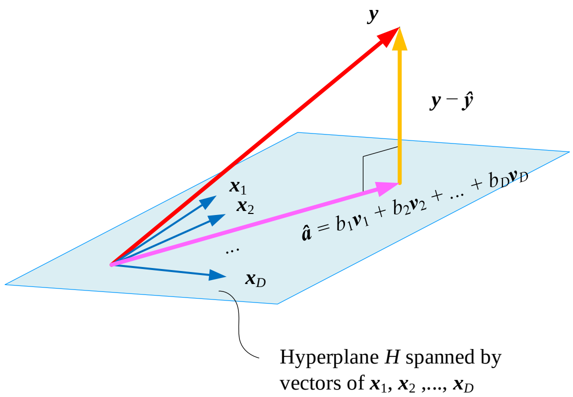


图 18. 向量 $\boldsymbol{y}$ 向超平面 $\text{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$ 投影

考虑常数项

而考虑常数项 b_0，无非就是在 (80) 中加入一个全 1 列向量 $\boldsymbol{1}$，即，

$$\hat{\boldsymbol{y}} = b_0\boldsymbol{1} + b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_D\boldsymbol{x}_D \tag{82}$$

而 $D + 1$ 个向量 $\boldsymbol{1}$、$\boldsymbol{x}_1$、$\boldsymbol{x}_2$、...、$\boldsymbol{x}_D$ 张成一个全新超平面 $\text{span}(\boldsymbol{1}, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$。而 $\boldsymbol{1}$ 经常写成 $\boldsymbol{x}_0$，新的 $\boldsymbol{X}$ 则为 $[\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$。按照本节前文思路，我们同样可以得到 (78)。

数据角度，$\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D$ 是一列列数值，但是几何视角下它们又是什么？本书第 12 章就试图回答这个问题。

更具一般性的正交投影

最后再回过头来看 (78)，我们可以发现这个式子实际上代表了更具一般性的正交投影。

数据矩阵 $\boldsymbol{X}_{n\times D}$ 的列向量 $[\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_D]$ 张成超平面 $H = \text{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$。

即便 $[\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_D]$ 之间并非两两正交，向量 $\boldsymbol{y}$ 依然可以在超平面 H 上正交投影，得到 $\hat{\boldsymbol{y}}$。

特殊地，如果假设 $\boldsymbol{X}$ 的列向量 $[\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_D]$ 两两正交，且列向量本身都是单位向量，可以得到：

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_D^{\mathrm{T}} \end{bmatrix}}_{X^{\mathrm{T}}} \underbrace{\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D \end{bmatrix}}_{X} = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_D \\ \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_D \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_D^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{x}_D^{\mathrm{T}}\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D^{\mathrm{T}}\boldsymbol{x}_D \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{83}$$

即：

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{I} \tag{84}$$

显然，$\boldsymbol{X}_{n\times D}$ 不能叫做正交矩阵，这是因为 $\boldsymbol{X}_{n\times D}$ 的形状为 $n \times D$，不是方阵。

将 (84) 代入 (78) 得到：

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{85}$$

将上式 $\boldsymbol{X}$ 写成 $[\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_D]$，并展开得到：

$$\hat{\boldsymbol{y}} = \underbrace{\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_D^{\mathrm{T}} \end{bmatrix}}_{X^{\mathrm{T}}} \boldsymbol{y} = \left(\boldsymbol{x}_1\boldsymbol{x}_1^{\mathrm{T}} + \boldsymbol{x}_2\boldsymbol{x}_2^{\mathrm{T}} + \cdots \boldsymbol{x}_D\boldsymbol{x}_D^{\mathrm{T}} \right)\boldsymbol{y} \tag{86}$$

进一步，使用向量张量积将上式写成：

$$\hat{\boldsymbol{y}} = \left(\boldsymbol{x}_1 \otimes \boldsymbol{x}_1 + \boldsymbol{x}_2 \otimes \boldsymbol{x}_2 + \cdots \boldsymbol{x}_D \otimes \boldsymbol{x}_D \right)\boldsymbol{y} \tag{87}$$

⚠ 再次强调，上式成立的前提是——$\boldsymbol{X}$ 的列向量 $[\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_D]$ 两两正交，且列向量本身都是单位向量。

这从另外一个侧面解释了我们为什么需要格拉姆-施密特正交化！也就是说，通过格拉姆-施密特正交化，$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_D]$ 变成 $\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, ..., \boldsymbol{q}_D]$。而 $[\boldsymbol{q}_1, \boldsymbol{q}_2, ..., \boldsymbol{q}_D]$ 两两正交，且列向量都是单位向量，即：

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\underbrace{\begin{bmatrix}\boldsymbol{q}_1^{\mathrm{T}}\\ \boldsymbol{q}_2^{\mathrm{T}}\\ \vdots\\ \boldsymbol{q}_D^{\mathrm{T}}\end{bmatrix}}_{\boldsymbol{Q}^{\mathrm{T}}}\underbrace{\begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_D\end{bmatrix}}_{\boldsymbol{Q}}=\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix} \tag{88}$$

从 $\boldsymbol{X}$ 到 $\boldsymbol{Q}$，本章利用的是格拉姆-施密特正交化，而本书第 11 章将用 QR 分解。此外，本书最后一章将介绍如何用矩阵分解结果计算线性回归系数。

到目前为止，相信大家已经领略到了矩阵乘法的伟力所在！本章前前后后用的无非就是矩阵乘法的各种变形、各种乘法视角。强烈建议大家回过头来再读一遍本书第 5 章，相信你会有一番新的收获。

本章从几何角度讲解正交投影及其应用。以下四幅图总结本书重要内容。

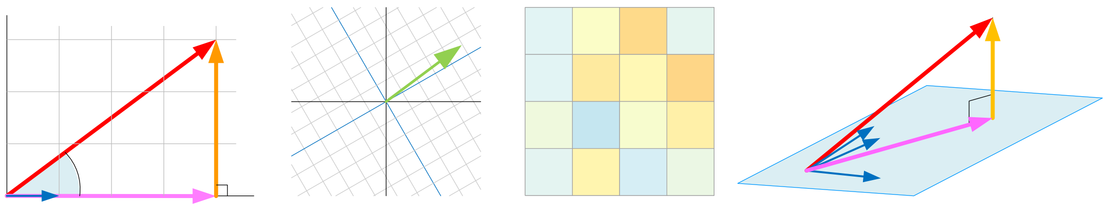


图 19. 总结本章重要内容的四副图

本书后续内容离不开投影这个线性代数工具！大家务必熟练掌握标量/向量投影，不管是用向量内积、矩阵乘法，还是张量积。

正交矩阵本身就是规范正交基。我们将会在数据投影、矩阵分解、数据空间等一系列话题中，反复用到正交矩阵。请大家务必注意正交矩阵的性质，以及两个展开视角。

手算格拉姆-施密特正交化没有意义，大家理解这个正交化思想就好。本书后续还会介绍其他正交化方法，重要的是大家能从几何、空间、数据视角区分不同正交化方法得到的规范正交系。

重要的事情，强调多少遍都不为过。有向量的地方，就有几何！几何视角是理解线性回归的最佳途径，本系列丛书《数据科学》还会展开讲解线性回归。

下一章以数据为视角，和大家聊聊正交投影如何帮助我们解密数据。