9 Orthogonal Projection 正交投影

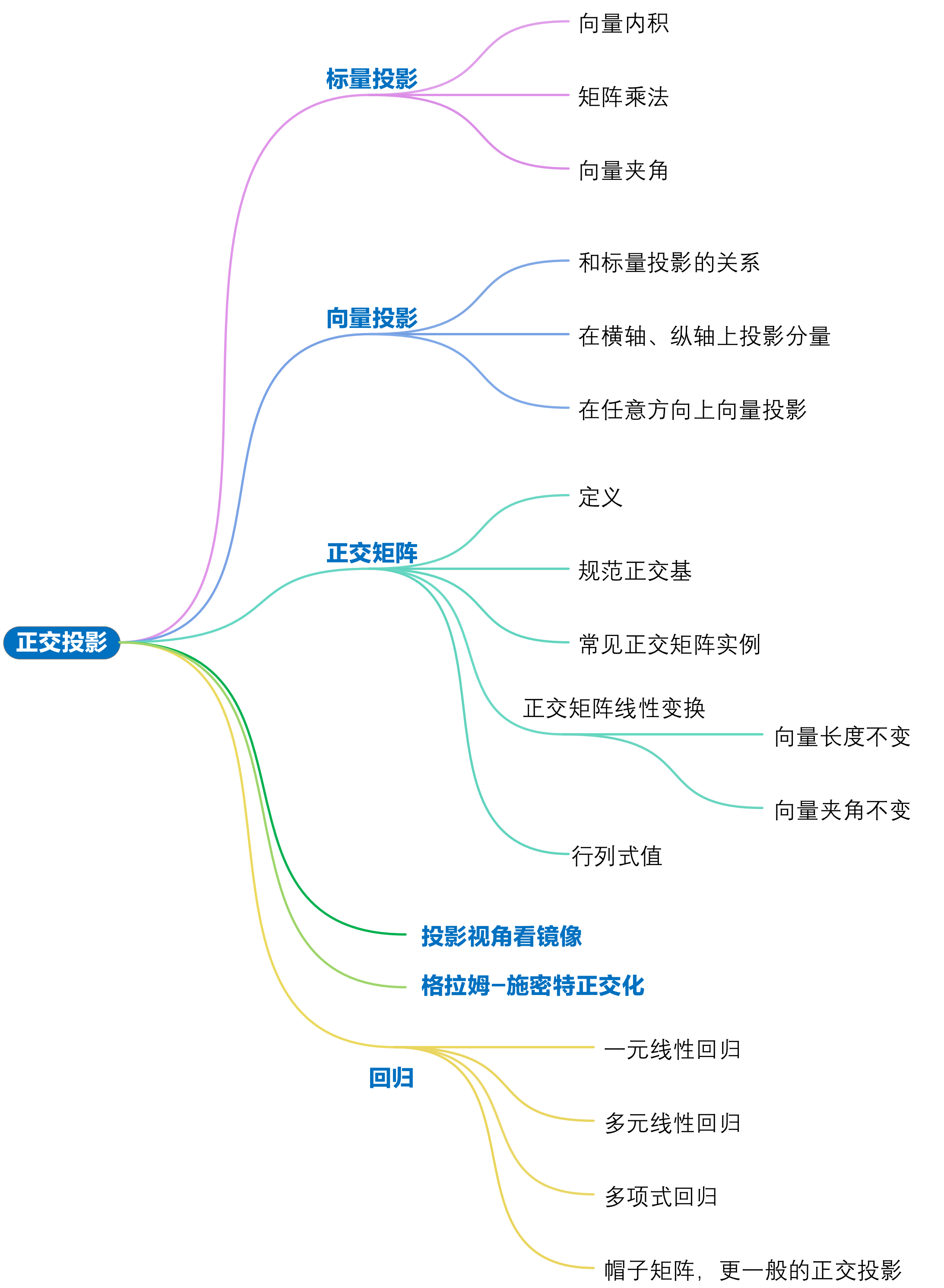
应用几乎无处不在

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
矩阵乘法
向量夹角
向量投影
和标量投影的关系
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
规范正交基
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
投影视角看镜像
格拉姆–施密特正交化
回归
一元线性回归
多元线性回归
多项式回归
帽子矩阵，更一般的正交投影

9.1 标量投影：结果为标量

正交

打个比方，**正交投影** (orthogonal projection) 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。

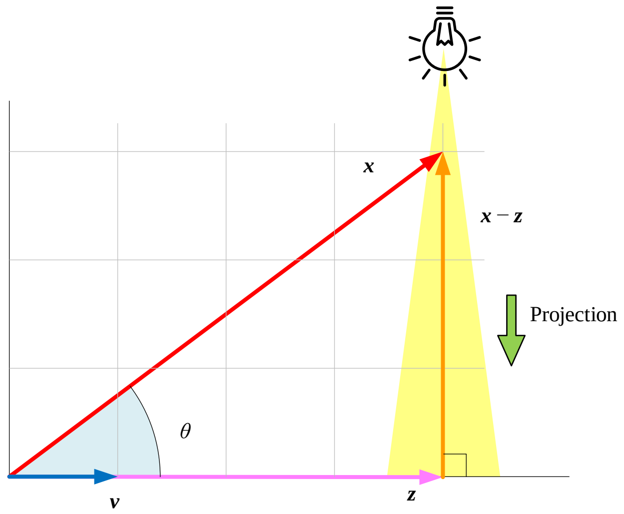


图 1. 正交投影的意义

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 $\boldsymbol{z}$。向量 $\boldsymbol{z}$ 的长度 (向量模) 就是 $\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上的**标量投影** (scalar projection)。

令，标量 s 为向量 $\boldsymbol{z}$ 的模。

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：

$$\boldsymbol{z} = s\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \tag{1}$$

很明显，如图 1 所示，$\boldsymbol{x}-\boldsymbol{z}$ 垂直于 $\boldsymbol{v}$，因此两者向量内积为 0：

$$(\boldsymbol{x}-\boldsymbol{z})\cdot\boldsymbol{v} = 0 \tag{2}$$

用矩阵乘法，(2) 可以写成，

$$(\boldsymbol{x}-\boldsymbol{z})^{\mathrm{T}}\boldsymbol{v} = 0 \tag{3}$$

将 (1) 代入 (3) 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：

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

上式可以写成如下几种形式：

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

⚠ 注意，$\boldsymbol{x}$ 和 $\boldsymbol{v}$ 为等行数列向量。

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

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

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

下面介绍如何从向量夹角入手推导标量投影。

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

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

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

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

这样，我们便得到和 (6) 一致的结果。

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，向量投影就是标量投影结果再乘上 $\boldsymbol{v}$ 的方向，即 s 乘以 $\boldsymbol{v}$ 的单位向量。因此，$\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上的向量投影实际上就是 (1)，即：

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

用尖括号<>表达标量积，x 在 v 方向上的向量投影可以记做：

$$\operatorname{proj}_v(x) = \frac{\langle x, v\rangle}{\langle v, v\rangle}v \tag{11}$$

特别地，如果 v 为单位向量，x 在 v 方向上的向量投影则可以写成：

$$\operatorname{proj}_v(x) = \langle x, v\rangle v = (x \cdot v)v = (v \cdot x)v = (x^{\mathrm{T}}v)v = (v^{\mathrm{T}}x)v \tag{12}$$

举个例子

实际上，获得平面上某一个向量的横、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 x，

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

如图 2 所示，列向量 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。

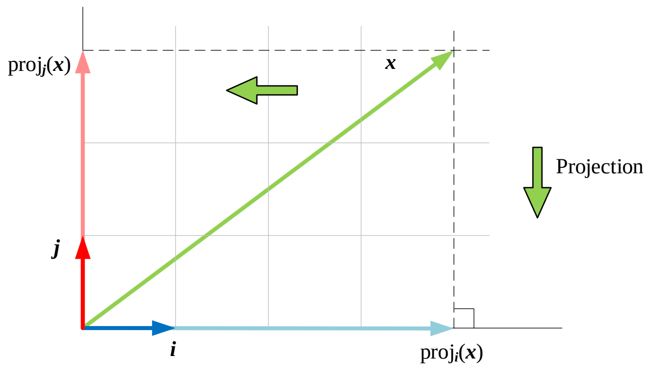


图 2. x 向 i 和 j 投影

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

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

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

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

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

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

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

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

如果单位向量 $\boldsymbol{v}$ 为，

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

$\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向上投影得到的标量投影为：

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

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

从向量空间角度来看，向量 $\boldsymbol{v}$ 张起的空间为 span($\boldsymbol{v}$)，这个向量空间维度为 1。由于 $\boldsymbol{x} = 5\boldsymbol{v}$，$\boldsymbol{x}$ 在 span($\boldsymbol{v}$) 坐标为 5。

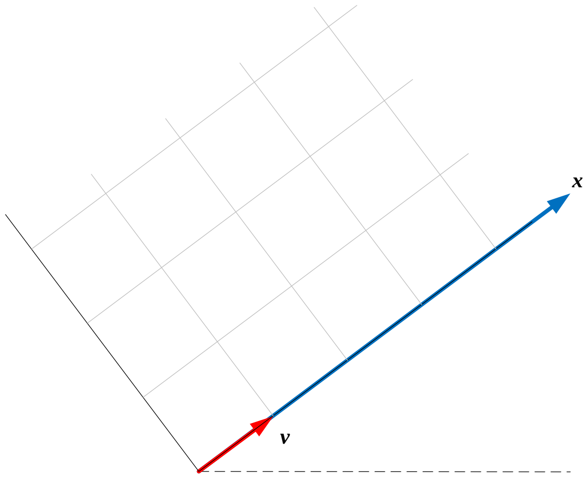


图 3. $\boldsymbol{x}$ 向 $\boldsymbol{v}$ 的投影

推导投影坐标

上一章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 $\boldsymbol{\tau}$ $[\tau_1, \tau_2]^{\mathrm{T}}$ 直线方向上正交投影得到点的坐标 (z_1, z_2) 为：

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

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 所示为点 A 向一系列通过原点、方向不同直线的投影坐标。

本书第 7 章强调过，向量空间一定都通过原点。大家可能会问，空间某点朝任意直线或超平面投影时，如果直线或超平面不通过原点，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。

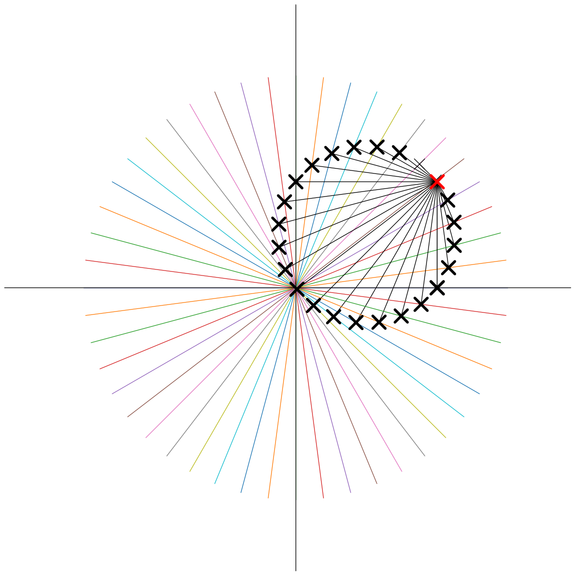


图 4. 点 A 向一系列通过原点的直线投影

Bk4_Ch9_01.py 绘制图 4。

向量张量积：无处不在

回过头再看 (12)，假设 $\boldsymbol{v}$ 为单位列向量，(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 $\boldsymbol{v}$ 方向投影得到的标量投影，即 $\boldsymbol{X}$ 在 $\operatorname{span}(\boldsymbol{v})$ 的坐标：

$$\boldsymbol{Z}=\boldsymbol{X}\boldsymbol{v} \tag{24}$$

$\boldsymbol{X}$ 向单位向量 $\boldsymbol{v}$ 方向投影得到的向量投影坐标则为：

$$\boldsymbol{Z}=\boldsymbol{X}\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}=\boldsymbol{X}(\boldsymbol{v}\otimes\boldsymbol{v})_{2\times 2} \tag{25}$$

请大家格外注意 (25)，我们下一章还要继续这个话题。此外，(25) 也是下一章要讨论的核心运算。

9.3 正交矩阵：一个规范正交基

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 $\boldsymbol{v}$ 张起的向量空间 $\operatorname{span}(\boldsymbol{v})$ 投影。同理，向量 $\boldsymbol{x}$ 也可以向一个有序基构造的平面/超平面投影。这个有序基可以是正交基，可以是非正交基。

数据科学和机器学习实践中，最常用的基底是规范正交基。正交矩阵的本身就是规范正交基。本节主要介绍正交矩阵的性质。

正交矩阵

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

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{I} \tag{26}$$

强调一下，$\boldsymbol{V}$ 为方阵是前提；否则即便满足上式也不能称之为正交矩阵。比如，如下长方形矩阵 $\boldsymbol{A}$ 也满足上式，但 $\boldsymbol{A}$ 不是正交矩阵：

$$\underbrace{\begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \end{bmatrix}}_{A^{\mathrm{T}}} \underbrace{\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \\ 0 & 0 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & \\ & 1 \end{bmatrix}}_{I} \tag{27}$$

但是，$\boldsymbol{A}$ 的列向量为单位向量、且两两正交，所以 $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2]$ 是规范正交基。

正交矩阵基本性质：

$$\begin{aligned} \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} &= \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{I} \\ \boldsymbol{V}^{\mathrm{T}} &= \boldsymbol{V}^{-1} \end{aligned} \tag{28}$$

(28) 中两式经常使用，必须烂熟于心。

举个实例，图 5 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 $\boldsymbol{I}$。

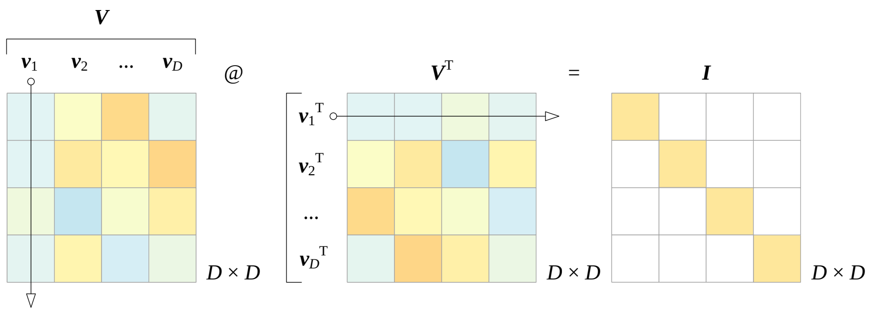


图 5. 正交阵 $\boldsymbol{V}$ 和自己转置 $\boldsymbol{V}^{\mathrm{T}}$ 乘积为单位阵 $\boldsymbol{I}$

前文的例子

其实我们已经接触过几种正交矩阵。本书前文提到的如下两个矩阵都是正交矩阵：

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \tag{29}$$

(29) 中 $\boldsymbol{V}$ 和 $\boldsymbol{W}$ 都满足方阵和自身转置乘积为单位阵，即：

$$\begin{aligned} \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} &= \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{30}$$

本书上一章讲过的矩阵 $\boldsymbol{R}$、$\boldsymbol{T}$ 和 $\boldsymbol{P}$ 都是正交矩阵：

$$\boldsymbol{R}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix},\quad \boldsymbol{T}=\begin{bmatrix}\cos 2\theta & \sin 2\theta\\ \sin 2\theta & -\cos 2\theta\end{bmatrix},\quad \boldsymbol{P}=\begin{bmatrix} & & & 1\\ 1 & & & \\ & & 1 & \\ & 1 & & \end{bmatrix} \tag{31}$$

其中，$\boldsymbol{R}$ 代表旋转，$\boldsymbol{T}$ 代表镜像，$\boldsymbol{P}$ 是置换矩阵。

矩阵乘法第一视角展开

将 (26) 中矩阵 $\boldsymbol{V}$ 写成一排列向量：

$$\boldsymbol{V}_{D\times D}=\begin{bmatrix}v_{1,1} & v_{1,2} & \cdots & v_{1,D}\\ v_{2,1} & v_{2,2} & \cdots & v_{2,D}\\ \vdots & \vdots & \ddots & \vdots\\ v_{n,1} & v_{n,2} & \cdots & v_{n,D}\end{bmatrix}=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D\end{bmatrix} \tag{32}$$

(26) 左侧可以写成：

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}}\\ \boldsymbol{v}_2^{\mathrm{T}}\\ \vdots\\ \boldsymbol{v}_D^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D\end{bmatrix} \tag{33}$$

(33) 展开得到：

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_1 & \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_D\\ \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{v}_1 & \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{v}_D\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{v}_1 & \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D^{\mathrm{T}}\boldsymbol{v}_D\end{bmatrix}=\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix} \tag{34}$$

大家应该已经意识到，(34) 就是 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}$ 矩阵乘法的第一视角。

$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}$ 主对角线结果为 1，即，

$$\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j=\boldsymbol{v}_j\cdot\boldsymbol{v}_j=\left\|\boldsymbol{v}_j\right\|^2=1\quad j=1,2,\ldots,D \tag{35}$$

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

(34) 主对角线以外元素均为 0：

$$\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j=0,\quad i\neq j \tag{36}$$

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 维向量空间 $\mathrm{span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D)$，$\mathbb{R}^D = \mathrm{span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D)$。也就是说，$[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$ 是张起 $\mathbb{R}^D$ 无数规范正交基的一组。

顺便提一嘴，由于 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}$，$\boldsymbol{V}^{\mathrm{T}}$ 本身也是一个规范正交基。$\boldsymbol{V}^{\mathrm{T}}$ 可以展开写成 $\boldsymbol{V}^{\mathrm{T}} = [\boldsymbol{v}^{(1)\mathrm{T}}, \boldsymbol{v}^{(2)\mathrm{T}}, \ldots, \boldsymbol{v}^{(j)\mathrm{T}}]$。

批量化计算向量模和夹角

此外，(34) 告诉我们“批量”计算一系列向量模和两两夹角的方式——**格拉姆矩阵** (Gram matrix)！

$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}$ 相当于 $\boldsymbol{V}$ 的格拉姆矩阵，通过对 (34) 的分析，我们知道格拉姆矩阵包含原矩阵的所有向量模、向量两两夹角这两类信息。

再举个例子，给定矩阵 $\boldsymbol{X}$，将其写成一组列向量 $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$。$\boldsymbol{X}$ 的格拉姆矩阵为：

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{x}_1 \cdot \boldsymbol{x}_1 & \boldsymbol{x}_1 \cdot \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_1 \cdot \boldsymbol{x}_D \\ \boldsymbol{x}_2 \cdot \boldsymbol{x}_1 & \boldsymbol{x}_2 \cdot \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_2 \cdot \boldsymbol{x}_D \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_D \cdot \boldsymbol{x}_1 & \boldsymbol{x}_D \cdot \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D \cdot \boldsymbol{x}_D \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{x}_1, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_1, \boldsymbol{x}_D \rangle \\ \langle \boldsymbol{x}_2, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_2, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_2, \boldsymbol{x}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{x}_D, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_D, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_D, \boldsymbol{x}_D \rangle \end{bmatrix} \tag{37}$$

借助向量夹角余弦展开 $\boldsymbol{G}$ 中向量积：

$$\boldsymbol{G} = \begin{bmatrix} \|\boldsymbol{x}_1\|\|\boldsymbol{x}_1\|\cos\theta_{1,1} & \|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\|\cos\theta_{1,2} & \cdots & \|\boldsymbol{x}_1\|\|\boldsymbol{x}_D\|\cos\theta_{1,D} \\ \|\boldsymbol{x}_2\|\|\boldsymbol{x}_1\|\cos\theta_{2,1} & \|\boldsymbol{x}_2\|\|\boldsymbol{x}_2\|\cos\theta_{2,2} & \cdots & \|\boldsymbol{x}_2\|\|\boldsymbol{x}_D\|\cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|\boldsymbol{x}_D\|\|\boldsymbol{x}_1\|\cos\theta_{D,1} & \|\boldsymbol{x}_D\|\|\boldsymbol{x}_2\|\cos\theta_{D,2} & \cdots & \|\boldsymbol{x}_D\|\|\boldsymbol{x}_D\|\cos\theta_{D,D} \end{bmatrix} \tag{38}$$

观察矩阵 $\boldsymbol{G}$，它包含了数据矩阵 $\boldsymbol{X}$ 中列向量的两个重要信息——模 $\|\boldsymbol{x}_i\|$、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 分解时继续深入探讨这一话题。

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 $\boldsymbol{V}$ 写成 $[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$，$\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}$ 则可以按如下方式展开：

$$\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}} \end{bmatrix} = \boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots \boldsymbol{v}_D\boldsymbol{v}_D^{\mathrm{T}} = \boldsymbol{I}_{D\times D} \tag{39}$$

(39) 可以写成一系列张量积之和：

$$\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \cdots \boldsymbol{v}_D \otimes \boldsymbol{v}_D = \boldsymbol{I}_{D\times D} \tag{40}$$

上一节 (25) 对应数据矩阵 $\boldsymbol{X}$ 向单位向量 $\boldsymbol{v}$ 向量投影。如果 $\boldsymbol{X}$ 向规范正交基 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_D]$ 张起的 D 维空间投影，得到的标量投影就是 $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{V}$，而向量投影结果为：

$$\begin{aligned} \boldsymbol{X}_{n\times D}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} &= \boldsymbol{X}\left(\boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \cdots + \boldsymbol{v}_D \otimes \boldsymbol{v}_D\right)_{2\times 2} \\ &= \underbrace{\boldsymbol{X}\boldsymbol{v}_1}_{z_1} \otimes \boldsymbol{v}_1 + \underbrace{\boldsymbol{X}\boldsymbol{v}_2}_{z_2} \otimes \boldsymbol{v}_2 + \cdots + \underbrace{\boldsymbol{X}\boldsymbol{v}_D}_{z_D} \otimes \boldsymbol{v}_D \\ &= \boldsymbol{X}_{n\times D}\boldsymbol{I}_{D\times D} \\ &= \boldsymbol{X}_{n\times D} \end{aligned} \tag{41}$$

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 $\boldsymbol{X}$ 本身！

(41) 已经非常接近本书第 15、16 章要讲解的奇异值分解的思路。下一章我们一起搞清楚 (41) 背后的数学思想。

再进一步，如图 6 所示，下式代表一个规范正交基对单位矩阵的分解：

$$\boldsymbol{I}_{D\times D} = \boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \cdots + \boldsymbol{v}_D \otimes \boldsymbol{v}_D = \sum_{j=1}^{D} \boldsymbol{v}_j \otimes \boldsymbol{v}_j \tag{42}$$

其中，每个 $\boldsymbol{v}_j \otimes \boldsymbol{v}_j$ 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。

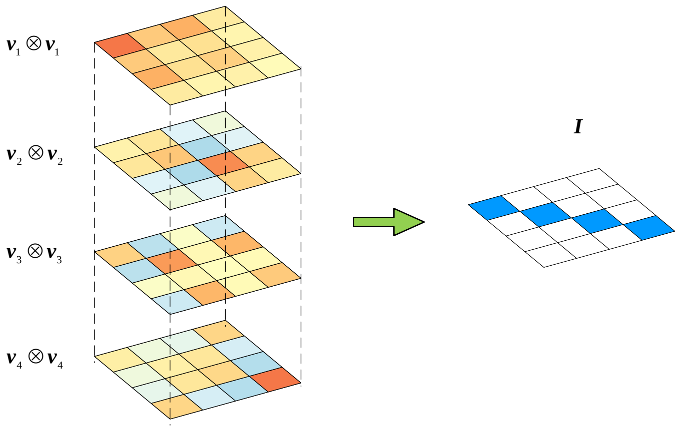


图 6. 对单位矩阵的分解

9.4 规范正交基性质

本节以 (29) 中矩阵 $\boldsymbol{V}$ 为例介绍更多规范正交基的性质。

坐标

将 $\boldsymbol{V}$ 分解成两个列向量，

$$\boldsymbol{v}_1 = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \tag{43}$$

这两个向量长度为 1，都是单位向量。

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

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_1 & \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_2 \\ \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{v}_1 & \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{I} \tag{44}$$

给定列向量 $\boldsymbol{x} = [4, 3]^{\mathrm{T}}$。如图 7 (a) 所示，$\boldsymbol{x}$ 在标准正交基 $[\boldsymbol{e}_1, \boldsymbol{e}_2]$ 中的坐标为 (4, 3)。

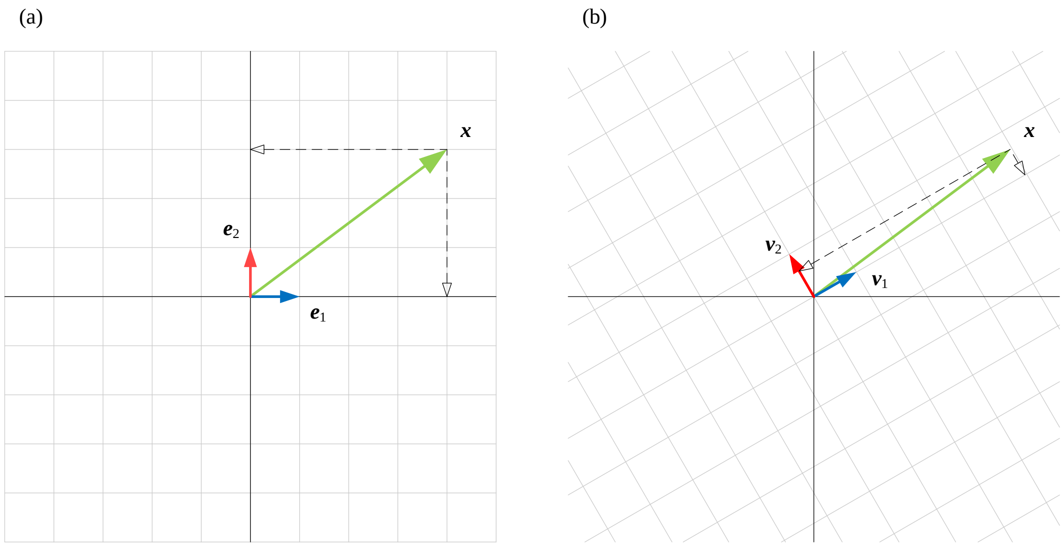


图 7. $\boldsymbol{x}$ 在不同规范正交系中的坐标

如图 7 (b) 所示，将 $\boldsymbol{x}$ 投影到 $\boldsymbol{V}$ 这个规范正交系中，得到的结果就是在 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 这个规范正交系的坐标：

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{x} \\ \boldsymbol{v}_2^{\mathrm{T}}\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \operatorname{proj}_{\boldsymbol{v}_1}(\boldsymbol{x}) \\ \operatorname{proj}_{\boldsymbol{v}_2}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 4.964 \\ 0.598 \end{bmatrix} \tag{45}$$

这说明，向量 $\boldsymbol{x}$ 在规范正交系 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 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，即向量模，没有变化：

$$\begin{aligned}\|V^{\mathrm{T}}x\|_2^2 &= V^{\mathrm{T}}x \cdot V^{\mathrm{T}}x = \left(V^{\mathrm{T}}x\right)^{\mathrm{T}}\left(V^{\mathrm{T}}x\right) = x^{\mathrm{T}}V^{\mathrm{T}}Vx \\ &= x^{\mathrm{T}}Ix = x^{\mathrm{T}}x = x \cdot x = \|x\|_2^2\end{aligned} \tag{46}$$

比较图 7 (a) 和 (b) 可以发现，不同规范正交系中 x 的长度确实没有变化。向量 x 在 $[v_1, v_2]$ 中坐标为 (4.964, 0.598)，计算其向量模：

$$\sqrt{4.964^2 + 0.598^2} = \sqrt{4^2 + 3^2} = 5 \tag{47}$$

图 8 所示为平面上给定向量在不同规范正交基中的投影结果。

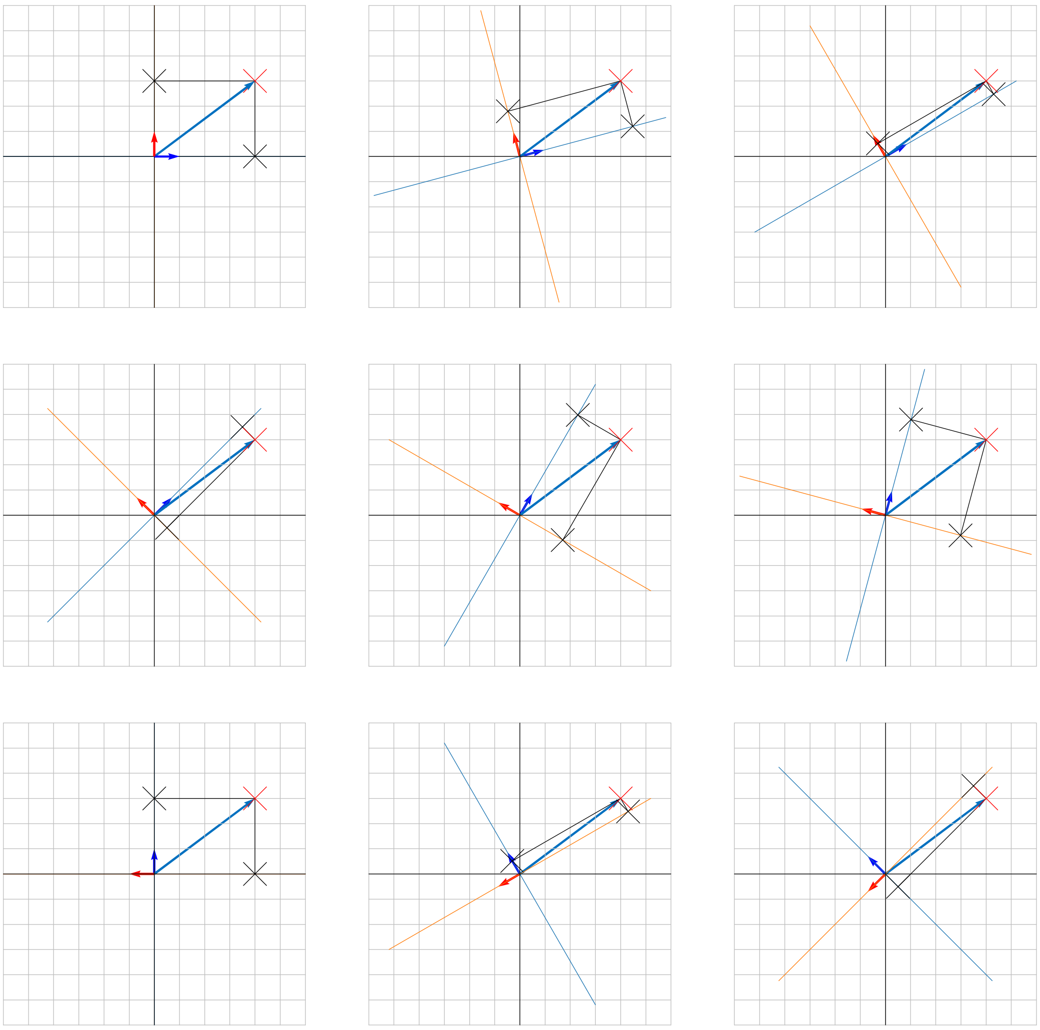


图 8. 平面中向量在不同坐标系的投影

Bk4_Ch9_02.py 绘制图 8。

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 $\boldsymbol{z}_j$ 夹角等同于 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 夹角：

$$\frac{\boldsymbol{z}_i \cdot \boldsymbol{z}_j}{\|\boldsymbol{z}_i\|\|\boldsymbol{z}_j\|}=\frac{\boldsymbol{z}_i \cdot \boldsymbol{z}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|}=\frac{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i \cdot \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|}=\frac{\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i\right)^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|}=\frac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|}=\frac{\boldsymbol{x}_i \cdot \boldsymbol{x}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|} \tag{48}$$

如图 9 所示，经过正交矩阵 $\boldsymbol{V}$ 线性变换后，$\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 两者相对角度等同于 $\boldsymbol{z}_i$ 和 $\boldsymbol{z}_j$ 相对角度。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。

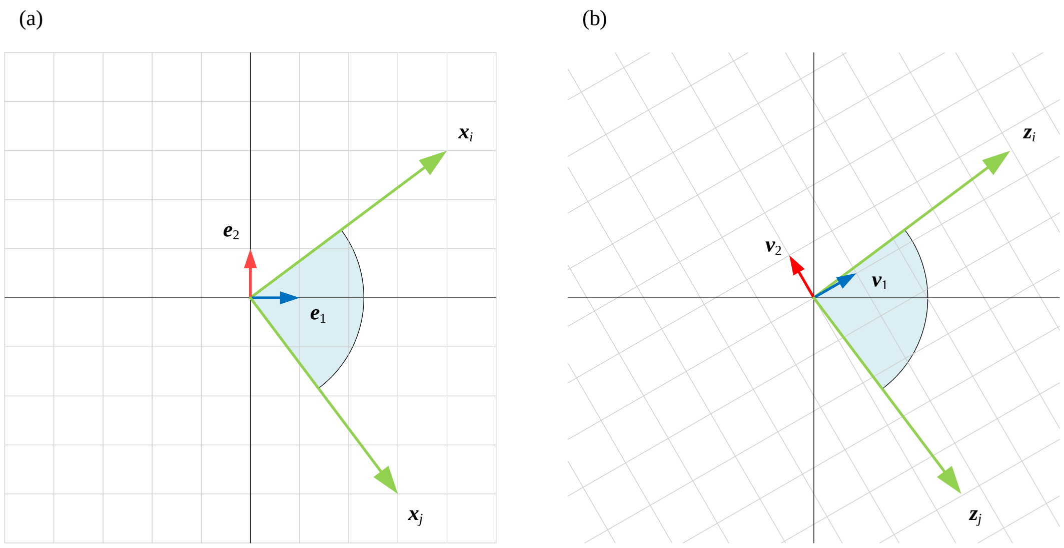


图 9. 不同规范正交系中，$\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 的夹角不变

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：

$$\left(\det(\boldsymbol{V})\right)^2=\det\left(\boldsymbol{V}^{\mathrm{T}}\right)\det(\boldsymbol{V})=\det\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\right)=\det(\boldsymbol{I})=1 \tag{49}$$

也就是说，经过 2 × 2 方阵 $\boldsymbol{V}$ 线性变换后，图形面积不变。当 $\det(\boldsymbol{V}) = -1$ 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 $\boldsymbol{T}$ 的一种形式，具体如下：

$$\boldsymbol{T} = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \tag{50}$$

本节用正交投影推导 (50)。

如图 10 所示，镜像对称轴 l 这条直线通过原点，直线切向量 $\boldsymbol{\tau}$ 为：

$$\boldsymbol{\tau} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{51}$$

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

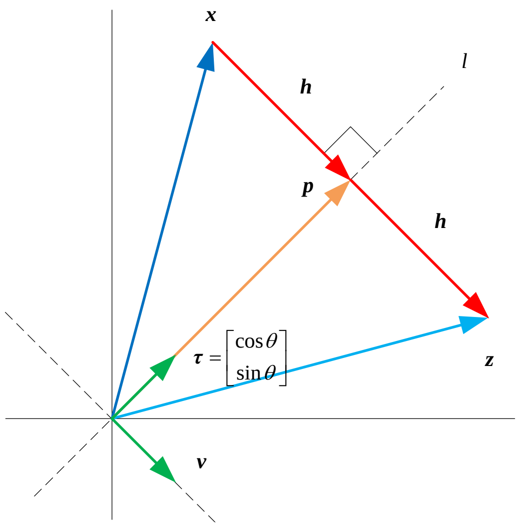


图 10. 投影视角看镜像

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：

$$\boldsymbol{p} = (\boldsymbol{\tau} \otimes \boldsymbol{\tau})\boldsymbol{x} \tag{52}$$

将 (51) 代入 (52)，整理得到：

$$\boldsymbol{p} = (\boldsymbol{\tau} \otimes \boldsymbol{\tau})\boldsymbol{x} = \begin{bmatrix} \cos\theta\cos\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin\theta\sin\theta \end{bmatrix}\boldsymbol{x} \tag{53}$$

利用三角恒等式，上式可以整理为：

$$\boldsymbol{p}=\begin{bmatrix}(\cos 2\theta+1)/2 & \sin 2\theta/2\\ \sin 2\theta/2 & (1-\cos 2\theta)/2\end{bmatrix}\boldsymbol{x} \tag{54}$$

令，向量 $\boldsymbol{h}$ 为 $\boldsymbol{p}$、$\boldsymbol{x}$ 之差，即：

$$\boldsymbol{h}=\boldsymbol{p}-\boldsymbol{x} \tag{55}$$

根据正交投影，$\boldsymbol{h}$ 显然垂直 $\boldsymbol{p}$。观察图 10，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：

$$\boldsymbol{z}=\boldsymbol{x}+2\boldsymbol{h} \tag{56}$$

将 (55) 代入 (56) 整理得到：

$$\boldsymbol{z}=2\boldsymbol{p}-\boldsymbol{x} \tag{57}$$

从另外一个角度来看，$\boldsymbol{x}+\boldsymbol{z}=2\boldsymbol{p}$。

将 (54) 代入 (57) 得到：

$$\begin{aligned}\boldsymbol{z}&=2\begin{bmatrix}(\cos 2\theta+1)/2 & \sin 2\theta/2\\ \sin 2\theta/2 & (1-\cos 2\theta)/2\end{bmatrix}\boldsymbol{x}-\boldsymbol{I}\boldsymbol{x}\\&=\begin{bmatrix}2\times(\cos 2\theta+1)/2-1 & 2\times\sin 2\theta/2\\ 2\times\sin 2\theta/2 & 2\times(1-\cos 2\theta)/2-1\end{bmatrix}\boldsymbol{x}\\&=\begin{bmatrix}\cos 2\theta & \sin 2\theta\\ \sin 2\theta & -\cos 2\theta\end{bmatrix}\boldsymbol{x}\end{aligned} \tag{58}$$

这样，我们便用投影视角推导得到 (50) 结果。

豪斯霍尔德矩阵

此外，将 (52) 代入 (57)，整理得到：

$$\boldsymbol{z}=2(\boldsymbol{\tau}\otimes\boldsymbol{\tau})\boldsymbol{x}-\boldsymbol{x}=(2\boldsymbol{\tau}\otimes\boldsymbol{\tau}-\boldsymbol{I})\boldsymbol{x} \tag{59}$$

在图 10 中，定义单位向量 $\boldsymbol{v}$ 垂直于切向量 $\boldsymbol{\tau}$，$[\boldsymbol{\tau}, \boldsymbol{v}]$ 为规范正交基，满足：

$$\boldsymbol{\tau}\otimes\boldsymbol{\tau}+\boldsymbol{v}\otimes\boldsymbol{v}=\boldsymbol{I} \tag{60}$$

$\boldsymbol{\tau}\otimes\boldsymbol{\tau}$ 可以写成：

$$\boldsymbol{\tau}\otimes\boldsymbol{\tau}=\boldsymbol{I}-\boldsymbol{v}\otimes\boldsymbol{v} \tag{61}$$

将 (61) 代入 (59) 得到：

$$\boldsymbol{z}=\underbrace{(\boldsymbol{I}-2\boldsymbol{v}\otimes\boldsymbol{v})}_{\boldsymbol{H}}\boldsymbol{x} \tag{62}$$

令 $\boldsymbol{H}$ 为：

$$\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{v} \otimes \boldsymbol{v} \tag{63}$$

矩阵 $\boldsymbol{H}$ 有自己的名字——**豪斯霍尔德矩阵** (Householder matrix)。矩阵 $\boldsymbol{H}$ 完成的转换叫做**豪斯霍尔德反射** (Householder reflection)，也叫初等反射。图 10 中向量 $\boldsymbol{v}$ 的方向就是反射面所在方向。

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：

$$\begin{aligned}
&\boldsymbol{\eta}_1 = \boldsymbol{x}_1 \\
&\boldsymbol{\eta}_2 = \boldsymbol{x}_2 - \operatorname{proj}_{\boldsymbol{\eta}_1}(\boldsymbol{x}_2) \\
&\boldsymbol{\eta}_3 = \boldsymbol{x}_3 - \operatorname{proj}_{\boldsymbol{\eta}_1}(\boldsymbol{x}_3) - \operatorname{proj}_{\boldsymbol{\eta}_2}(\boldsymbol{x}_3) \\
&\cdots \\
&\boldsymbol{\eta}_D = \boldsymbol{x}_D - \sum_{j=1}^{D-1} \operatorname{proj}_{\boldsymbol{\eta}_j}(\boldsymbol{x}_D)
\end{aligned} \tag{64}$$

前两步

图 11 所示为格拉姆-施密特正交化前两步。

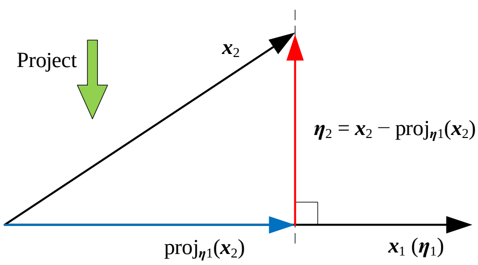


图 11. 格拉姆-施密特正交化前两步

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：

$$(\boldsymbol{\eta}_1)^{\mathrm{T}}\boldsymbol{\eta}_2 = 0 \tag{65}$$

如图 11 所示，$\boldsymbol{x}_2$ 在 $\boldsymbol{\eta}_1$ 方向上向量投影为 $\mathrm{proj}_{\boldsymbol{\eta}1}(\boldsymbol{x}_2)$，剩余的向量分量垂直于 $\boldsymbol{x}_1$ ($\boldsymbol{\eta}_1$)，这个分量就是 $\boldsymbol{\eta}_2$：

$$\boldsymbol{\eta}_2 = \boldsymbol{x}_2 - \mathrm{proj}_{\boldsymbol{\eta}_1}(\boldsymbol{x}_2) = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{\eta}_1}{\boldsymbol{\eta}_1^{\mathrm{T}}\boldsymbol{\eta}_1}\boldsymbol{\eta}_1 \tag{66}$$

$\boldsymbol{\eta}_2$ 也有自己的名字，叫 $\boldsymbol{\eta}_1$ 的**正交补** (orthogonal complement)。也可以说，$\boldsymbol{\eta}_1$ 和 $\boldsymbol{\eta}_2$ 互为正交补。下面验证 $\boldsymbol{\eta}_1$ 和 $\boldsymbol{\eta}_2$ 相互垂直：

$$\begin{aligned}(\boldsymbol{\eta}_1)^{\mathrm{T}}\boldsymbol{\eta}_2 &= (\boldsymbol{x}_1)^{\mathrm{T}}\left(\boldsymbol{x}_2 - \underbrace{\frac{\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{\eta}_1}{\boldsymbol{\eta}_1^{\mathrm{T}}\boldsymbol{\eta}_1}}_{\text{Scalar}}\boldsymbol{\eta}_1\right) \\ &= \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 - \frac{\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_1}{\underbrace{\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1}_{\text{Scalar}}} = \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 - \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_1 = 0\end{aligned} \tag{67}$$

第三步

如图 12 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 $[\boldsymbol{\eta}_1, \boldsymbol{\eta}_2]$ 平面上的向量分量，即：

$$\boldsymbol{\eta}_3 = \boldsymbol{x}_3 - \mathrm{proj}_{\boldsymbol{\eta}_1}(\boldsymbol{x}_3) - \mathrm{proj}_{\boldsymbol{\eta}_2}(\boldsymbol{x}_3) \tag{68}$$

显然，$\boldsymbol{\eta}_3$ 垂直 $\mathrm{span}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$，也就是说 $\boldsymbol{\eta}_3$ 分别垂直 $\boldsymbol{\eta}_1$ 和 $\boldsymbol{\eta}_2$。$\boldsymbol{\eta}_3$ 和 $\mathrm{span}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ 互为正交补。

按此思路，不断反复投影直至得到所有正交向量 $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \ldots, \boldsymbol{\eta}_D\}$。

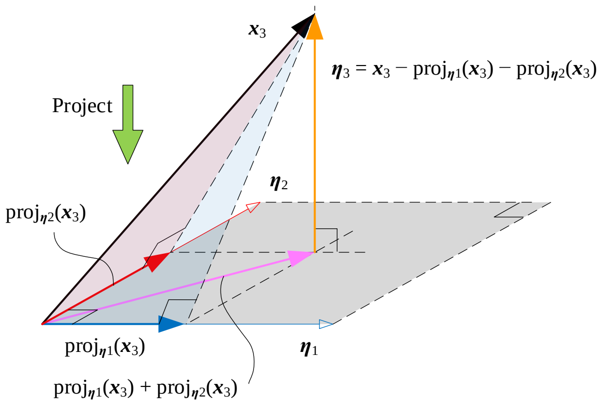


图 12. 格拉姆-施密特正交化第三步

单位化

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

$$\boldsymbol{q}_1 = \frac{\boldsymbol{\eta}_1}{\|\boldsymbol{\eta}_1\|}, \quad \boldsymbol{q}_2 = \frac{\boldsymbol{\eta}_2}{\|\boldsymbol{\eta}_2\|}, \quad \boldsymbol{q}_3 = \frac{\boldsymbol{\eta}_3}{\|\boldsymbol{\eta}_3\|}, \quad \cdots, \quad \boldsymbol{q}_D = \frac{\boldsymbol{\eta}_D}{\|\boldsymbol{\eta}_D\|} \tag{69}$$

值得强调的是，规范正交基 $[\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \ldots, \boldsymbol{q}_D]$ 的特别之处在于 $\boldsymbol{q}_1$ 平行 $\boldsymbol{x}_1$。本书后续还会介绍其他获得规范正交基的算法，请大家注意比对。

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：

$$\boldsymbol{x}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{70}$$

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

$$\boldsymbol{\eta}_1 = \boldsymbol{x}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \tag{71}$$

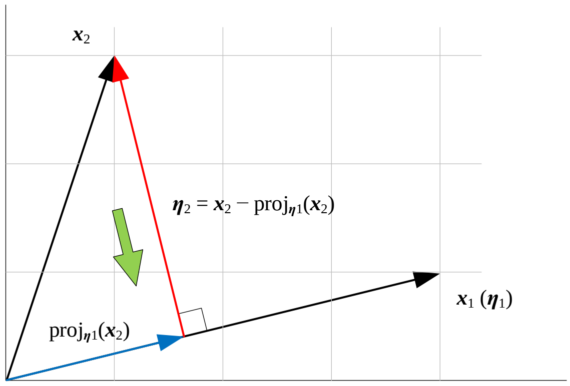


图 13. 格拉姆-施密特正交化第三步

$\boldsymbol{x}_2$ 在 $\boldsymbol{\eta}_1$ $(\boldsymbol{x}_1)$ 方向上投影，得到向量投影：

$$\operatorname{proj}_{\boldsymbol{\eta}_1}(\boldsymbol{x}_2) = \frac{\boldsymbol{x}_2 \cdot \boldsymbol{\eta}_1}{\boldsymbol{\eta}_1 \cdot \boldsymbol{\eta}_1}\boldsymbol{\eta}_1 = \frac{4\times1+1\times3}{4\times4+1\times1}\times\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \frac{7}{17}\times\begin{bmatrix} 4 \\ 1 \end{bmatrix} \tag{72}$$

计算 $\boldsymbol{\eta}_2$：

$$\begin{aligned}\boldsymbol{\eta}_2 &= \boldsymbol{x}_2 - \operatorname{proj}_{\boldsymbol{\eta}_1}\left(\boldsymbol{x}_2\right) \\ &= \begin{bmatrix}1\\3\end{bmatrix} - \frac{7}{17}\times\begin{bmatrix}4\\1\end{bmatrix} = \frac{1}{17}\times\begin{bmatrix}-11\\44\end{bmatrix}\end{aligned} \tag{73}$$

最后对 $\boldsymbol{\eta}_1$ 和 $\boldsymbol{\eta}_2$ 单位化，得到 $\boldsymbol{q}_1$ 和 $\boldsymbol{q}_2$：

$$\begin{cases}\boldsymbol{q}_1 = \dfrac{\boldsymbol{\eta}_1}{\|\boldsymbol{\eta}_1\|} = \dfrac{1}{\sqrt{17}}\begin{bmatrix}4\\1\end{bmatrix} \\ \boldsymbol{q}_2 = \dfrac{\boldsymbol{\eta}_2}{\|\boldsymbol{\eta}_2\|} = \dfrac{1}{\sqrt{17}}\begin{bmatrix}-1\\4\end{bmatrix}\end{cases} \tag{74}$$

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 投影视角看回归

本系列丛书《数学要素》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：

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

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

$$\hat{\boldsymbol{y}} = b\boldsymbol{x} \tag{76}$$

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？

从向量空间角度来看，$\operatorname{span}(\boldsymbol{x})$ 张起的向量空间维度为 1。$\hat{\boldsymbol{y}}$ 在 $\operatorname{span}(\boldsymbol{x})$ 中，$\hat{\boldsymbol{y}}$ 和 $\boldsymbol{x}$ 线性相关。

从数据角度思考，$\boldsymbol{x}$ 为自变量，$\boldsymbol{y}$ 为因变量。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 $\boldsymbol{\varepsilon} = \boldsymbol{y}-\hat{\boldsymbol{y}}$。$\boldsymbol{\varepsilon}$ 和 $\boldsymbol{x}$ 互为正交补。

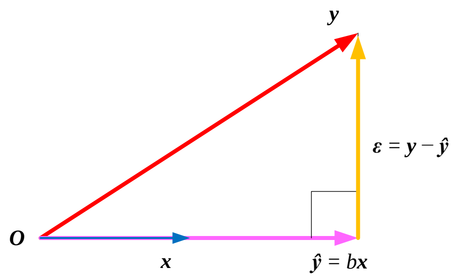


图 14. 向量 $\boldsymbol{y}$ 向 $\boldsymbol{x}$ 正交投影得到向量投影 $\hat{\boldsymbol{y}}$

将 (76) 代入 (75)，得到：

$$\boldsymbol{x}^{\mathrm{T}}\left(\boldsymbol{y}-b\boldsymbol{x}\right)=0 \tag{77}$$

容易求得系数 b 为：

$$b=\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \tag{78}$$

从而，$\hat{\boldsymbol{y}}$ 为：

$$\hat{\boldsymbol{y}}=\boldsymbol{x}\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \tag{79}$$

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，上述线性回归模型为比例函数，截距为 0。从图像上来看，比例函数过原点。

二元线性回归

下面我们聊一下二元线性回归。

如图 15 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：

$$\hat{\boldsymbol{y}}=b_1\boldsymbol{x}_1+b_2\boldsymbol{x}_2=\underbrace{\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2\end{bmatrix}}_{\boldsymbol{X}}\underbrace{\begin{bmatrix}b_1\\ b_2\end{bmatrix}}_{\boldsymbol{b}}=\boldsymbol{X}\boldsymbol{b} \tag{80}$$

其中，b_1 和 b_2 为系数。span($\boldsymbol{x}_1$, $\boldsymbol{x}_2$) 和 $\boldsymbol{\varepsilon}=\boldsymbol{y}-\hat{\boldsymbol{y}}$ 互为正交补。

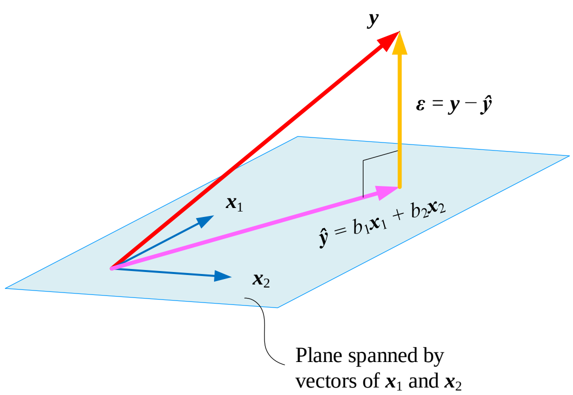


图 15. 向量 $\boldsymbol{y}$ 向平面 span($\boldsymbol{x}_1$, $\boldsymbol{x}_2$) 投影

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

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

注意，并不要求 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 相互正交。

整理 (81) 得到：

$$\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{y}-\hat{\boldsymbol{y}})=\boldsymbol{0} \tag{82}$$

将 (80) 代入 (82) 得到：

$$\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b})=\boldsymbol{0} \tag{83}$$

从而推导得到 $\boldsymbol{b}$ 的解析式：

$$\boldsymbol{b}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{84}$$

(84) 代入 (80)，可以得到：

$$\boxed{\hat{\boldsymbol{y}}=\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}} \tag{85}$$

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)。必须强调一点，只有 $\boldsymbol{X}$ 为列满秩时，$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 才存在逆。

$\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$ 是我们在本书第 5 章提到的**幂等矩阵** (idempotent matrix)，即下式成立：

$$\left(\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\right)^2=\boldsymbol{X}\underbrace{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}}_{\boldsymbol{I}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}=\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}} \tag{86}$$

多元线性回归

以上结论也可以推广到如图 16 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 $H=\mathrm{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$，向量 $\boldsymbol{y}$ 在超平面 H 上投影结果为 $\hat{\boldsymbol{y}}$，即，

$$\hat{\boldsymbol{y}}=b_1\boldsymbol{x}_1+b_2\boldsymbol{x}_2+\cdots+b_D\boldsymbol{x}_D \tag{87}$$

误差 $\boldsymbol{y}-\hat{\boldsymbol{y}}$ 垂直垂直于 $H=\mathrm{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$，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：

$$\begin{cases} \boldsymbol{x}_1^{\mathrm{T}}(\boldsymbol{y}-\hat{\boldsymbol{y}})=0 \\ \boldsymbol{x}_2^{\mathrm{T}}(\boldsymbol{y}-\hat{\boldsymbol{y}})=0 \\ \quad\vdots \\ \boldsymbol{x}_D^{\mathrm{T}}(\boldsymbol{y}-\hat{\boldsymbol{y}})=0 \end{cases} \Rightarrow \underbrace{\begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_D^{\mathrm{T}} \end{bmatrix}}_{\boldsymbol{X}^{\mathrm{T}}}(\boldsymbol{y}-\hat{\boldsymbol{y}})=\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{88}$$

$\mathrm{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$ 和 $\boldsymbol{\varepsilon}=\boldsymbol{y}-\hat{\boldsymbol{y}}$ 互为正交补。

用之前的推导思路，我们也可以得到 (85)。

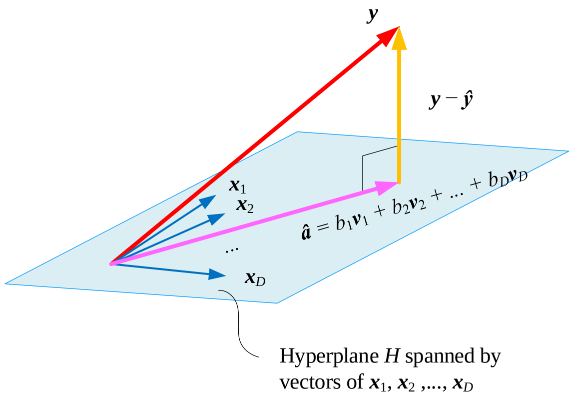


图 16. 向量 $\boldsymbol{y}$ 向超平面 span($\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D$) 投影

考虑常数项

而考虑常数项 b_0，无非就是在 (87) 中加入一个全 1 列向量 $\boldsymbol{1}$，即，

$$\hat{\boldsymbol{y}} = b_0\boldsymbol{1} + b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_D\boldsymbol{x}_D \tag{89}$$

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 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 (85)。

在多元线性回归中，$\boldsymbol{X}$ 也叫**设计矩阵** (design matrix)。

数据角度来看，$\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D$ 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。

多项式回归

有些应用场合，自变量和因变量之间存在明显的非线性关系，线性回归不足以描述这种关系。这种情况，我们需要借助非线性回归模型，比如**多项式回归** (polynomial regression)。

举个例子，一元三次多项式回归模型可以写成：

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + b_3 x^3 \tag{90}$$

这时，设计矩阵 $\boldsymbol{X}$ 为：

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}_{n\times 4} \tag{91}$$

举个例子，$\boldsymbol{x}$ 和 $\boldsymbol{y}$ 取值图 17 所示。

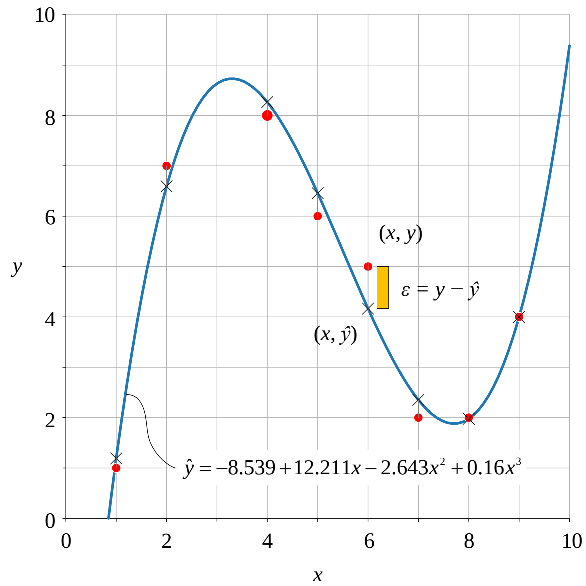


图 17. 一元三次多项式回归

一元三次多项式回归模型的自变量 $\boldsymbol{x}$、因变量 $\boldsymbol{y}$、设计矩阵 $\boldsymbol{X}$ 分别为：

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{bmatrix}_{8\times 1}, \quad \boldsymbol{y} = \begin{bmatrix} 1 \\ 7 \\ 8 \\ 6 \\ 5 \\ 2 \\ 2 \\ 4 \end{bmatrix}_{8\times 1}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 4 & 16 & 64 \\ 1 & 5 & 25 & 125 \\ 1 & 6 & 36 & 216 \\ 1 & 7 & 49 & 343 \\ 1 & 8 & 64 & 512 \\ 1 & 9 & 81 & 729 \end{bmatrix}_{8\times 4} \tag{92}$$

利用 (84) 计算得到系数向量 $\boldsymbol{b}$：

$$\boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \left(\underbrace{\begin{bmatrix} 8 & 42 & 276 & 1998 \\ 42 & 276 & 1998 & 15252 \\ 276 & 1998 & 15252 & 120582 \\ 1998 & 15252 & 120582 & 977676 \end{bmatrix}}_{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}} \right)^{-1} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \approx \begin{bmatrix} -8.539 \\ 12.211 \\ -2.643 \\ 0.160 \end{bmatrix} \tag{93}$$

三次一元多项式回归模型可以写成：

$$\hat{y} = -8.539 + 12.211x - 2.643x^2 + 0.16x^3 \tag{94}$$

对于给定的因变量 $\boldsymbol{y}$，因变量预测值为 $\hat{\boldsymbol{y}}$，误差为 $\boldsymbol{\varepsilon}$，它们的具体值如下：

$$
\boldsymbol{y}=\begin{bmatrix}1\\7\\8\\6\\5\\2\\2\\4\end{bmatrix}_{8\times 1},\quad \hat{\boldsymbol{y}}=\begin{bmatrix}1.189\\6.592\\8.266\\6.457\\4.165\\2.351\\1.976\\4.001\end{bmatrix}_{8\times 1},\quad \boldsymbol{\varepsilon}=\boldsymbol{y}-\hat{\boldsymbol{y}}=\begin{bmatrix}-0.189\\0.408\\-0.266\\-0.457\\0.835\\-0.351\\0.024\\-0.001\end{bmatrix}_{8\times 1} \tag{95}
$$

更具一般性的正交投影

最后再回过头来看 (85)，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 = 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：

$$
\underbrace{\begin{bmatrix}\boldsymbol{x}_1^{\mathrm{T}}\\\boldsymbol{x}_2^{\mathrm{T}}\\\vdots\\\boldsymbol{x}_D^{\mathrm{T}}\end{bmatrix}}_{\boldsymbol{X}^{\mathrm{T}}}\underbrace{\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D\end{bmatrix}}_{\boldsymbol{X}}=\begin{bmatrix}\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_D\\\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_D\\\vdots & \vdots & \ddots & \vdots\\\boldsymbol{x}_D^{\mathrm{T}}\boldsymbol{x}_1 & \boldsymbol{x}_D^{\mathrm{T}}\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D^{\mathrm{T}}\boldsymbol{x}_D\end{bmatrix}=\begin{bmatrix}1 & 0 & \cdots & 0\\0 & 1 & \cdots & 0\\\vdots & \vdots & \ddots & \vdots\\0 & 0 & \cdots & 1\end{bmatrix} \tag{96}
$$

即：

$$
\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}=\boldsymbol{I} \tag{97}
$$

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 (97) 代入 (85) 得到：

$$
\hat{\boldsymbol{y}}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{98}
$$

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

$$
\hat{\boldsymbol{y}}=\underbrace{\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D\end{bmatrix}}_{\boldsymbol{X}}\underbrace{\begin{bmatrix}\boldsymbol{x}_1^{\mathrm{T}}\\\boldsymbol{x}_2^{\mathrm{T}}\\\vdots\\\boldsymbol{x}_D^{\mathrm{T}}\end{bmatrix}}_{\boldsymbol{X}^{\mathrm{T}}}\boldsymbol{y}=\left(\boldsymbol{x}_1\boldsymbol{x}_1^{\mathrm{T}}+\boldsymbol{x}_2\boldsymbol{x}_2^{\mathrm{T}}+\cdots\boldsymbol{x}_D\boldsymbol{x}_D^{\mathrm{T}}\right)\boldsymbol{y} \tag{99}
$$

进一步，使用向量张量积将上式写成：

$$
\hat{\boldsymbol{y}}=\left(\boldsymbol{x}_1\otimes\boldsymbol{x}_1+\boldsymbol{x}_2\otimes\boldsymbol{x}_2+\cdots\boldsymbol{x}_D\otimes\boldsymbol{x}_D\right)\boldsymbol{y} \tag{100}
$$

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，即满足下式：

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \underbrace{\begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}} \\ \boldsymbol{q}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{q}_D^{\mathrm{T}} \end{bmatrix}}_{\boldsymbol{Q}^{\mathrm{T}}} \underbrace{\begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_D \end{bmatrix}}_{\boldsymbol{Q}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{101}$$

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。

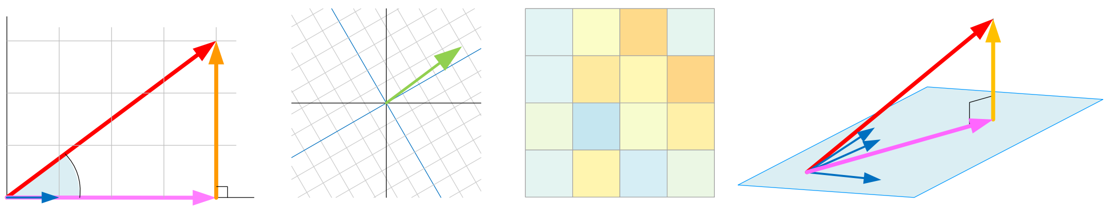


图 18. 总结本章重要内容的四幅图

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结果差异。

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《概率统计》、《数据科学》还会从不同角度展开讲解线性回归。

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。