

# 高级计量经济学

## Lecture 3: The Geometry of Linear Regression

黄嘉平

工学博士 经济学博士  
深圳大学中国经济特区研究中心 讲师

**办公室** 粤海校区汇文楼1510  
**E-mail** [huangjp@szu.edu.cn](mailto:huangjp@szu.edu.cn)  
**Website** <https://huangjp.com>

欧氏空间

# 欧氏空间

## Euclidean Space

$n$  维欧氏空间  $E^n$  是在  $n$  维实向量的集合  $\mathbb{R}^n$  中加入内积

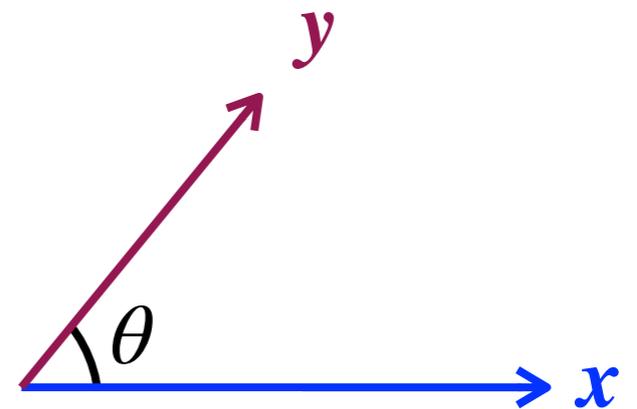
$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} \equiv \mathbf{x}^\top \mathbf{y} \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

$E^n$  中向量的长度:  $\|\mathbf{x}\| = (\mathbf{x}^\top \mathbf{x})^{1/2}$

内积与内角:  $\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta$

$\mathbf{x}, \mathbf{y}$  平行时, 则  $\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \|\mathbf{y}\|$

$\mathbf{x}, \mathbf{y}$  互相垂直时 (即正交  $\mathbf{x} \perp \mathbf{y}$ ), 则  $\langle \mathbf{x}, \mathbf{y} \rangle = 0$



Cauchy-Schwartz inequality:  $|\mathbf{x}^\top \mathbf{y}| \leq \|\mathbf{x}\| \|\mathbf{y}\|$  因为  $-1 \leq \cos \theta \leq 1$

# 子空间

## Subspace

$E^n$  的子空间也是一个欧氏空间  $E^k, k \leq n$ 。

基底 (basis vectors) 与张成 (span)

若  $E^n$  中的  $n$  个向量  $\mathbf{x}_1, \dots, \mathbf{x}_n$  线性独立, 且任意  $\mathbf{y} \in E^n$  可以写成  $\mathbf{x}_1, \dots, \mathbf{x}_n$  的线性结合, 则  $\mathbf{x}_1, \dots, \mathbf{x}_n$  被称为  $E^n$  的基底,  $E^n$  可以由  $\mathbf{x}_1, \dots, \mathbf{x}_n$  张成, 记作  $E^n = \mathcal{S}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ 。

$\mathcal{S}(\mathbf{x}_1, \dots, \mathbf{x}_k), \mathbf{x}_1, \dots, \mathbf{x}_k \in E^n, k \leq n$ , 是由  $\mathbf{x}_1, \dots, \mathbf{x}_k$  张成的  $E^n$  的子空间。

当矩阵  $X = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_k]$  时,  $\mathcal{S}(\mathbf{x}_1, \dots, \mathbf{x}_k)$  被称作  $X$  的列空间, 记作  $\mathcal{S}(X)$ 。

# 正交子空间

## Orthogonal Subspaces

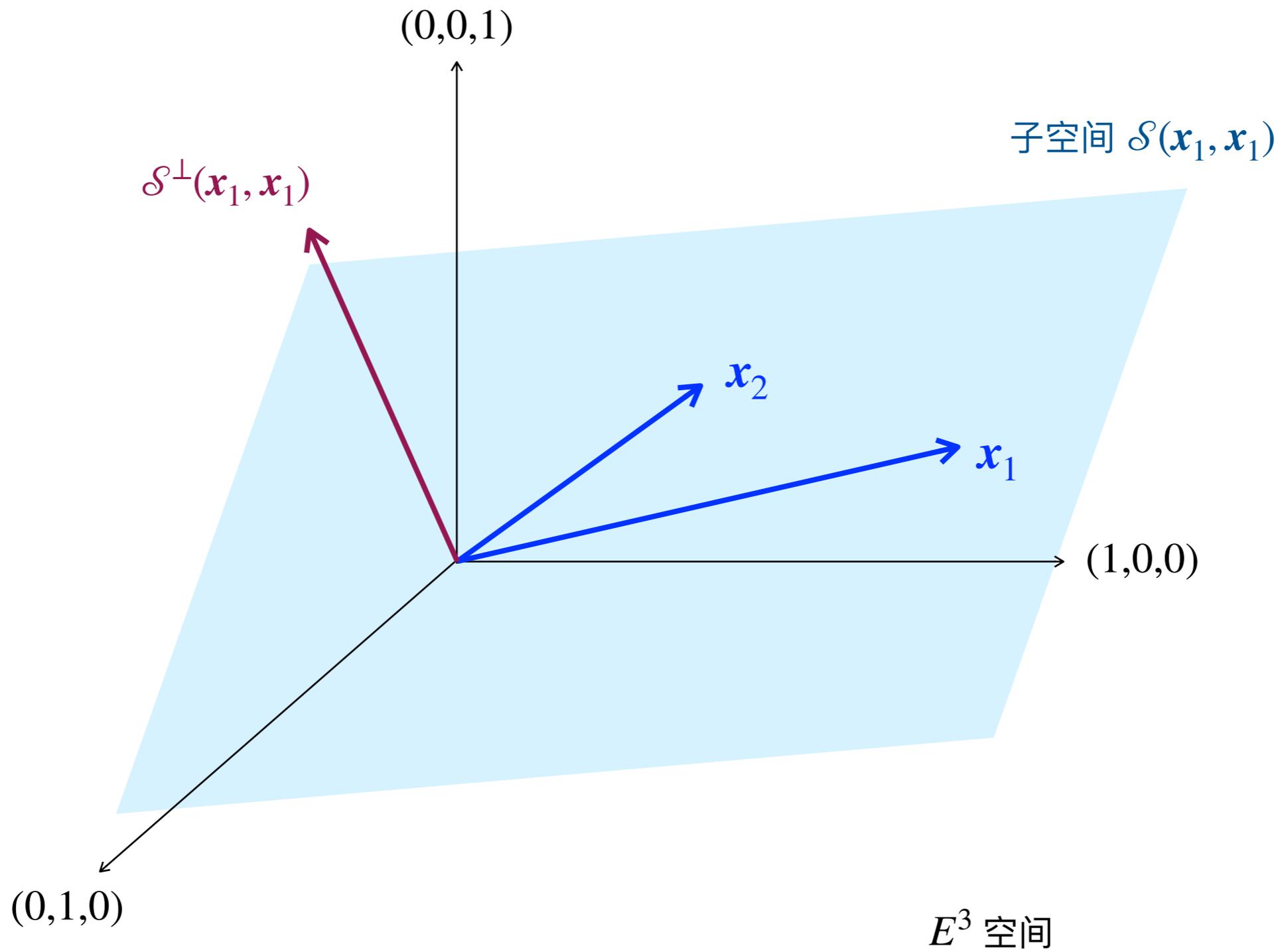
对于  $E^n$  的两个子空间  $S_1$  和  $S_2$ ，若任意  $v_1 \in S_1, v_2 \in S_2$  的内积为零，则称  $S_1$  和  $S_2$  正交。

$E^n$  中所有和  $\mathcal{S}(X)$  正交的向量的集合  $\mathcal{S}^\perp(X)$  被称作  $\mathcal{S}(X)$  在  $E^n$  中的正交补空间，即

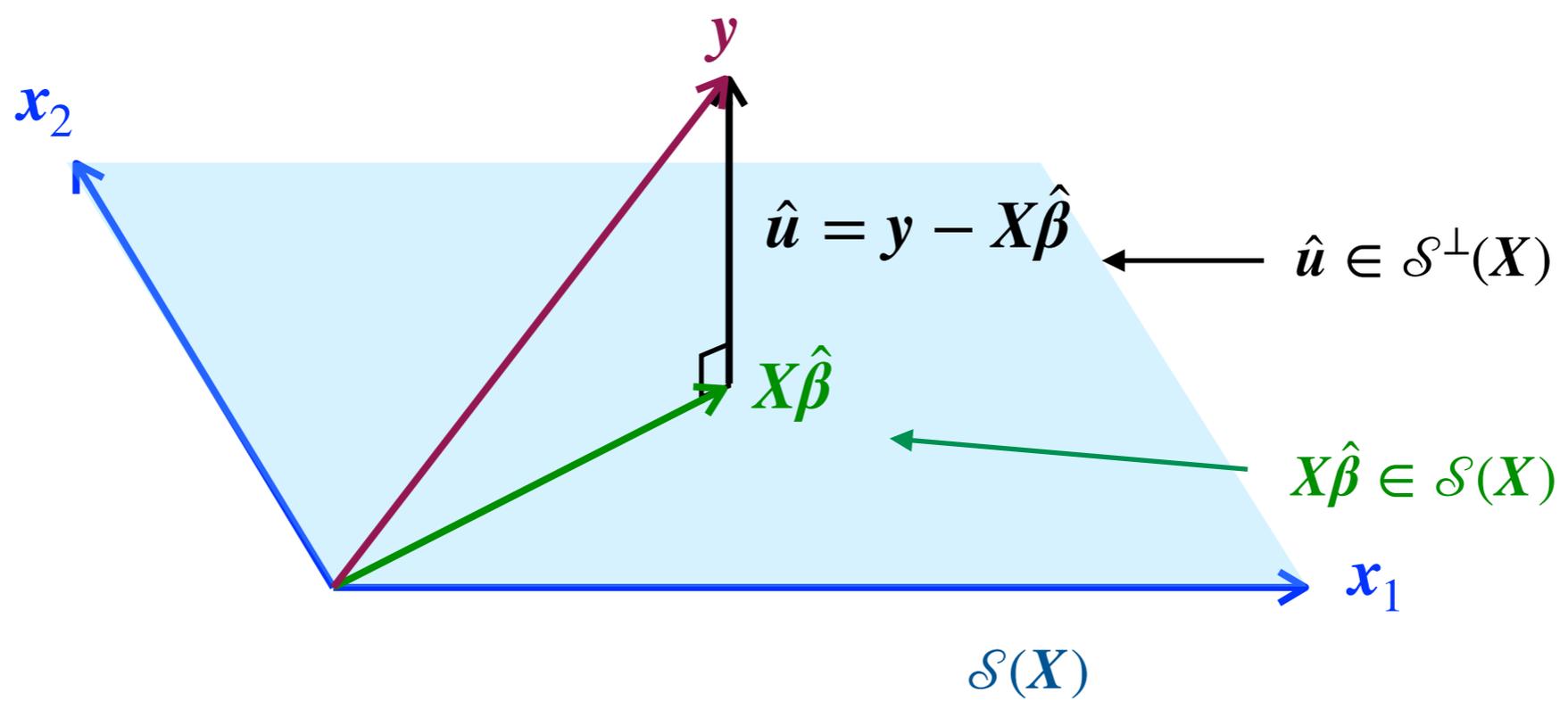
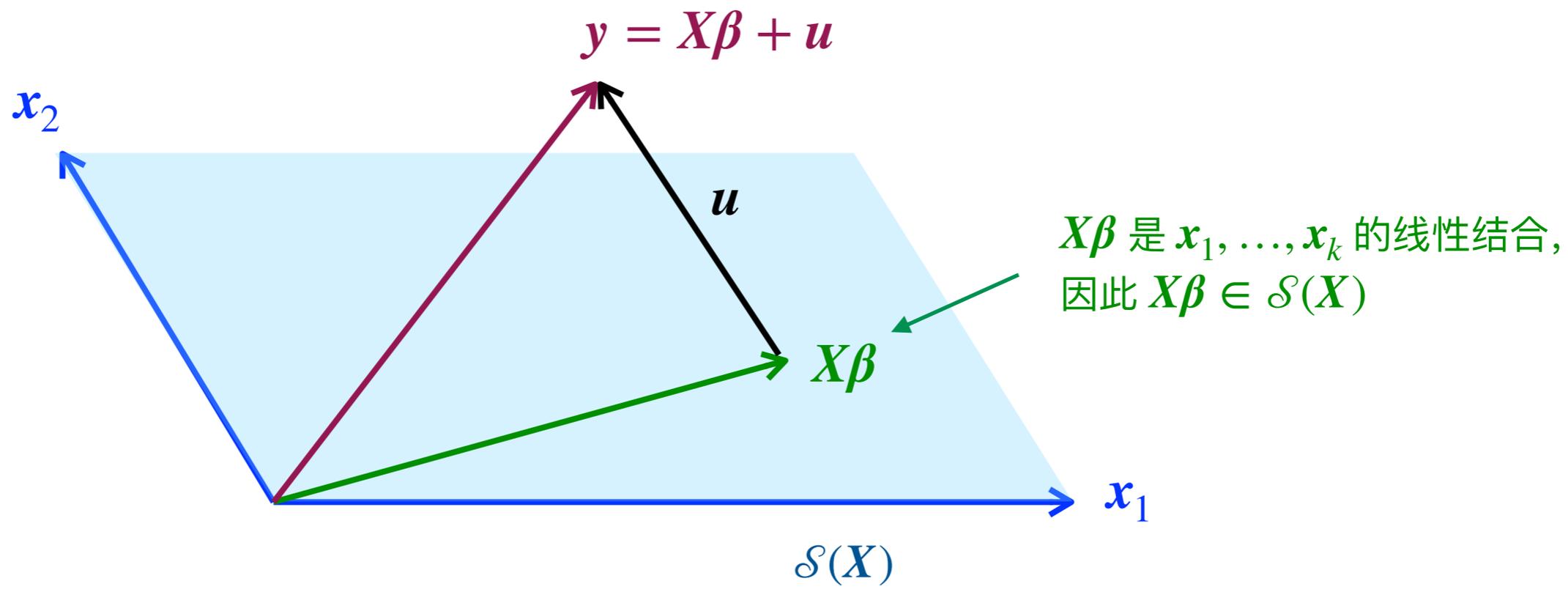
$$\mathcal{S}^\perp(X) = \{y \in E^n \mid y \perp x \text{ for all } x \in \mathcal{S}(X)\}$$

维度：空间的维度等于其基底的数量

$$\dim(E^n) = n, \dim(\mathcal{S}(X)) = k \Rightarrow \dim(\mathcal{S}^\perp(X)) = n - k$$



根据定义，任意的  $a \in \mathcal{S}(X)$  和任意的  $b \in \mathcal{S}^\perp(X)$  都满足  $a \perp b$



$\hat{u} \in \mathcal{S}^\perp(X)$ , 因此  $\hat{u} \perp x_1, \hat{u} \perp x_2$

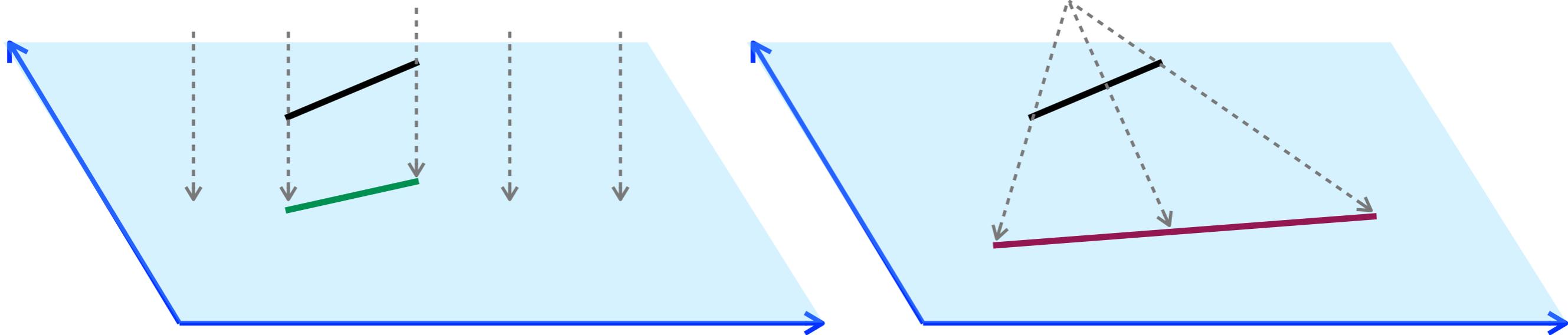
# 正交投影

# 正交投影

## Orthogonal Projection

投影：将  $E^n$  中的每一点和其子空间中的一点关联起来的函数，或映射（mapping）。这时要保持原本在子空间中的点不变。

正交投影：将  $E^n$  的每一点映射到子空间中距离其最近的点。



绿色直线为正交投影

紫色直线为透视投影

# 投影矩阵

## Projection Matrix

$$\begin{aligned}\hat{u} &= y - X\hat{\beta} \\ &= y - X(X^\top X)^{-1}X^\top y \\ &= (I - X(X^\top X)^{-1}X^\top)y\end{aligned}$$

令

$$\begin{aligned}P_X &= X(X^\top X)^{-1}X^\top, \\ M_X &= I - X(X^\top X)^{-1}X^\top = I - P_X\end{aligned}$$

则  $P_X y = X\hat{\beta}$  为  $y$  的 OLS 预测值  $\hat{y}$ ,  $M_X y = \hat{u}$  为 OLS 残差。

我们可以称  $P_X$  为投影矩阵, 因其将  $y$  投影到子空间  $\mathcal{S}(X)$ ; 称  $M_X$  为残差生成矩阵 (residual maker), 因其将  $y$  投影到  $\mathcal{S}^\perp(X)$ 。

# 投影矩阵的性质

- $P_X X = X(X^\top X)^{-1} X^\top X = X$
- $M_X X = (I - P_X)X = X - X = \mathbf{O}$  (零矩阵)
- $P_X$  和  $M_X$  是对称矩阵 注意  $(A^\top)^{-1} = (A^{-1})^\top$
- $P_X$  和  $M_X$  是幂等矩阵 (idempotent matrix)

$$P_X^2 = X(X^\top X)^{-1} X^\top X(X^\top X)^{-1} X^\top = X(X^\top X)^{-1} X^\top = P_X$$
$$M_X^2 = (I - P_X)(I - P_X) = I - 2P_X + P_X^2 = I - P_X = M_X$$

- $P_X M_X = M_X P_X = \mathbf{O}$

# 投影的值域与子空间

$P_X$  的值域是  $\mathcal{S}(X)$  整体。 $M_X$  的值域是  $\mathcal{S}^\perp(X)$  整体。

1. **值域包含在子空间内**：对于任意的  $y \in \mathbb{R}^n$ ，  
 $P_X y = X\hat{\beta}$  是  $X$  的列的线性结合，因此  $P_X y \in \mathcal{S}(X)$ ；  
 $X^\top M_X y = X^\top M_X^\top y = (M_X X)^\top y = \mathbf{0}$ ，因此  $M_X y \in \mathcal{S}^\perp(X)$ 。
2. **子空间包含在值域内**： $\mathcal{S}(X)$  中的任意一点  $x$  都可以写成  
 $x = a_1 x_1 + \cdots + a_k x_k = Xa$ ，因为  $P_X x = P_X Xa = Xa = x$ ，所以  
 $\mathcal{S}(X)$  中所有的点都是其自身通过  $P_X$  得到的像。  
 $\mathcal{S}^\perp(X)$  中的任意一点  $z$  都满足  $X^\top z = \mathbf{0}$ ，此时  
 $M_X z = (I - P_X)z = z - P_X z$ ，因为  $P_X z = X(X^\top X)^{-1}X^\top z = \mathbf{0}$ ，所以  
 $M_X z = z$ ，即  $\mathcal{S}^\perp(X)$  中所有的点都是其自身通过  $M_X$  得到的像。

# 其他重要公式

- 标准方程 (normal equation)

$$X^T \hat{u} = \mathbf{0} \Leftrightarrow X^T (y - X\hat{\beta}) = \mathbf{0} \Leftrightarrow X^T X\hat{\beta} = X^T y$$

正交条件

与最小二乘法相同

- 线性模型的正交分解

由定义可知  $P_X + M_X = I$ , 因此

$$y = P_X y + M_X y = \hat{y} + \hat{u}$$

通过勾股定理可得  $\|y\|^2 = \|\hat{y}\|^2 + \|\hat{u}\|^2$ , 即 TSS = ESS + SSR

# 解释变量的线性变换

# 子空间 $\mathcal{S}(X)$ 的基底

$\mathcal{S}(X)$  可以由不同的基底张成。

令  $A$  为  $k \times k$  的非奇异矩阵, 即  $\det(A) \neq 0$ ,  $A$  可逆, 则

$$XA = X[a_1 \ a_2 \ \dots \ a_k] = [Xa_1 \ Xa_2 \ \dots \ Xa_k]$$

这里每个  $Xa_i$  都是  $\mathcal{S}(X)$  中的一点, 因此由  $(Xa_1, Xa_2, \dots, Xa_k)$  张成的子空间  $\mathcal{S}(XA) \subseteq \mathcal{S}(X)$ 。

任意的  $x \in \mathcal{S}(X)$  都可以表达为  $x = Xb = XAA^{-1}b = (XA)(A^{-1}b)$ , 因此  $x$  是  $XA$  的列的线性结合, 即  $x \in \mathcal{S}(XA)$ 。因此  $\mathcal{S}(X) \subseteq \mathcal{S}(XA)$ 。

最终可得  $\mathcal{S}(X) = \mathcal{S}(XA)$ 。

# 解释变量的线性变换

$XA$  可以看做线性回归中解释变量的线性变换。变换前后的投影矩阵  $P_X$  和  $P_{XA}$  代表同一正交投影。

$$\begin{aligned} P_{XA} &= XA(A^\top X^\top XA)^{-1}A^\top X^\top \\ &= XAA^{-1}(X^\top X)^{-1}(A^\top)^{-1}A^\top X^\top \quad \Rightarrow \quad P_{XA}y = P_Xy \\ &= X(X^\top X)^{-1}X^\top \\ &= P_X \\ \Rightarrow M_{XA} &= M_X, M_{XA}y = M_Xy \\ \Rightarrow \hat{\beta}_{XA} &= (A^\top X^\top XA)^{-1}A^\top X^\top y = A^{-1}(X^\top X)^{-1}(A^\top)^{-1}A^\top X^\top y = A^{-1}\hat{\beta}_X \end{aligned}$$

针对线性回归  $y = X\beta + u$ ，若将解释变量进行线性变换，则变换后的模型  $y = XA\beta + u$  和原模型的预测值相同，残差相同，但 OLS 估计值  $\hat{\beta}$  会发生变化。

# 解释变量的单位转换

由上一页的结论可导出：解释变量的单位转换不影响预测值和残差，但影响系数的估计值。

假设模型包含温度这一解释变量和常数项。温度的单位可以是摄氏 ( $C$ ) 或华氏 ( $F$ )，二者的关系是  $F = 32 + \frac{9}{5}C$ 。令  $\mathbf{1}$  代表要素为 1 的向量，则

$$[\mathbf{1} \quad F] = [\mathbf{1} \quad C] \begin{bmatrix} 1 & 32 \\ 0 & \frac{9}{5} \end{bmatrix}$$

即以摄氏记录的数据可以通过  $\begin{bmatrix} 1 & 32 \\ 0 & \frac{9}{5} \end{bmatrix}$  转换为华氏。若摄氏和华氏下的回归系数分别为  $[\beta_1 \quad \beta_2]$  和  $[\alpha_1 \quad \alpha_2]$ ，则可得  $\beta_1 = \alpha_1 + 32\alpha_2$ ,  $\beta_2 = \frac{9}{5}\alpha_2$ 。