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
Website https://huangjp.com

欧氏空间

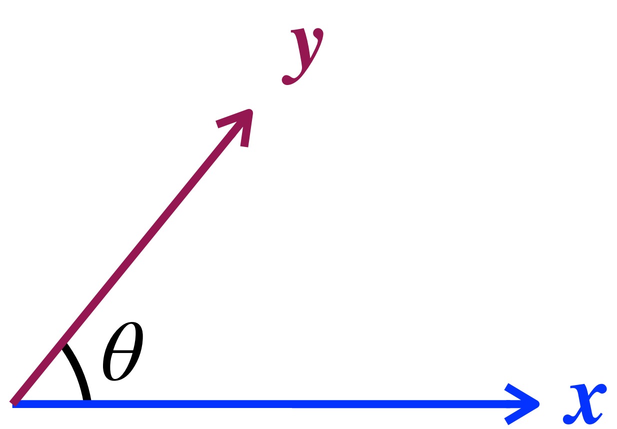
欧氏空间

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

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

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



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

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

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

Cauchy-Schwartz inequality：$|\boldsymbol{x}^\top \boldsymbol{y}| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$ 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 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 线性不相关，且任意 $\boldsymbol{y} \in E^n$ 可以写成 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 的线性结合，则 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 被称为 E^n 的基底，E^n 可以由 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 张成，记作 $E^n = \mathcal{S}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ 。

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

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

正交子空间

Orthogonal Subspaces

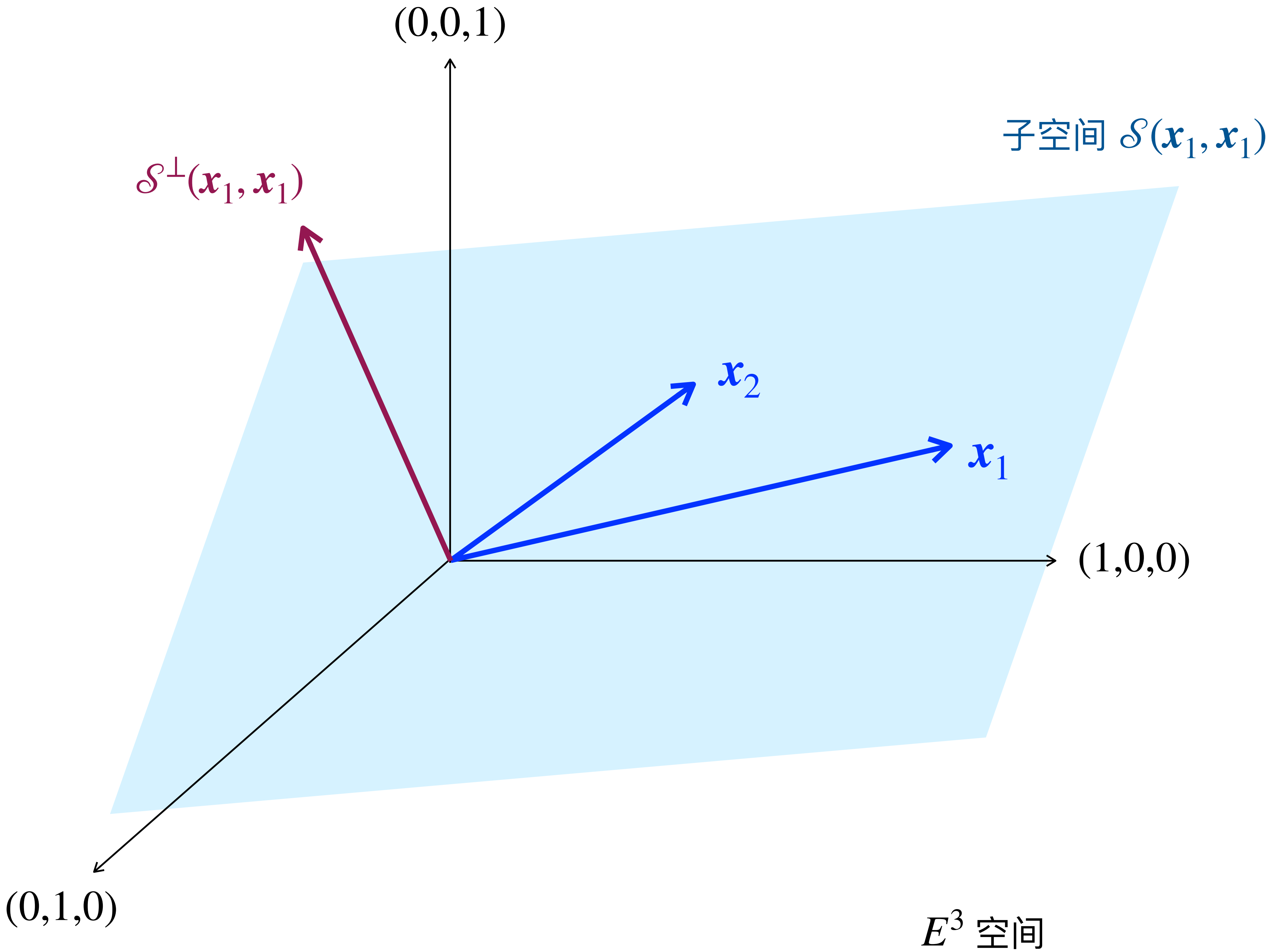
对于 E^n 的两个子空间 S_1 和 S_2，若任意 $\boldsymbol{v}_1 \in S_1, \boldsymbol{v}_2 \in S_2$ 的内积为零，则称 S_1 和 S_2 正交。

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

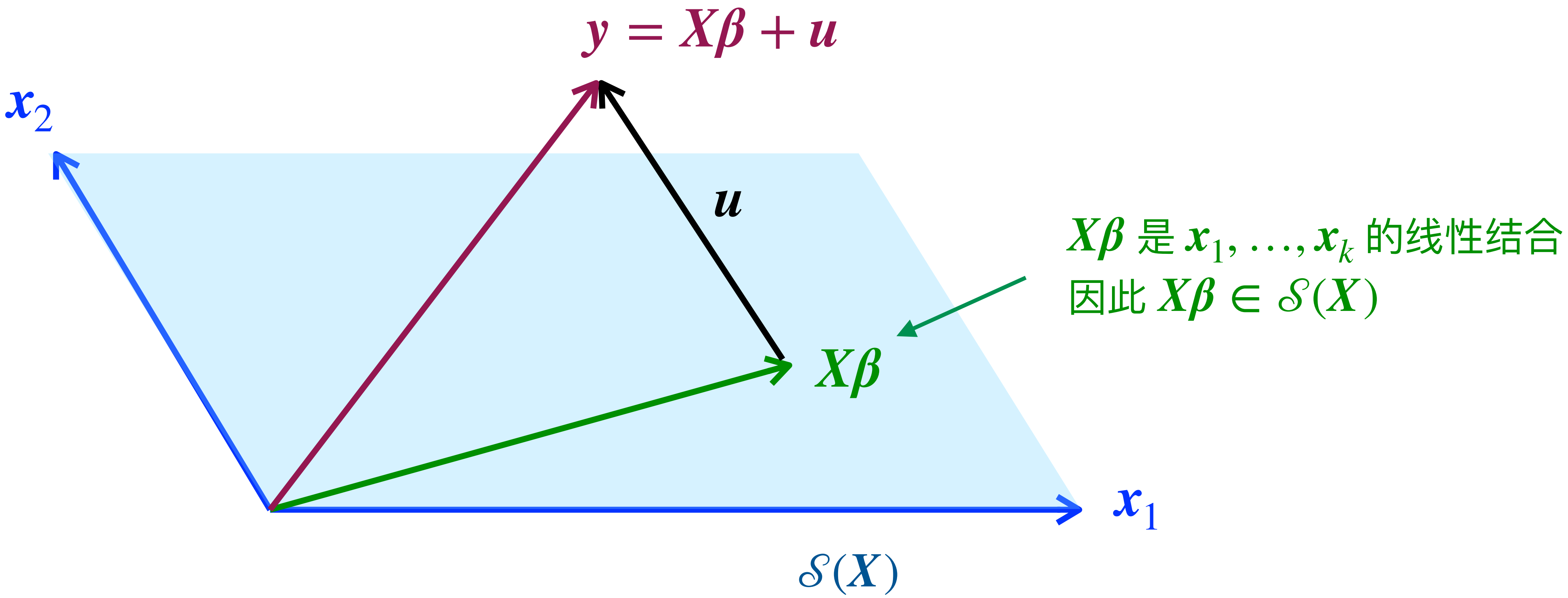
$$\mathcal{S}^\perp(\boldsymbol{X}) = \{\boldsymbol{y} \in E^n \mid \boldsymbol{y} \perp \boldsymbol{x} \text{ for all } \boldsymbol{x} \in \mathcal{S}(\boldsymbol{X})\}$$

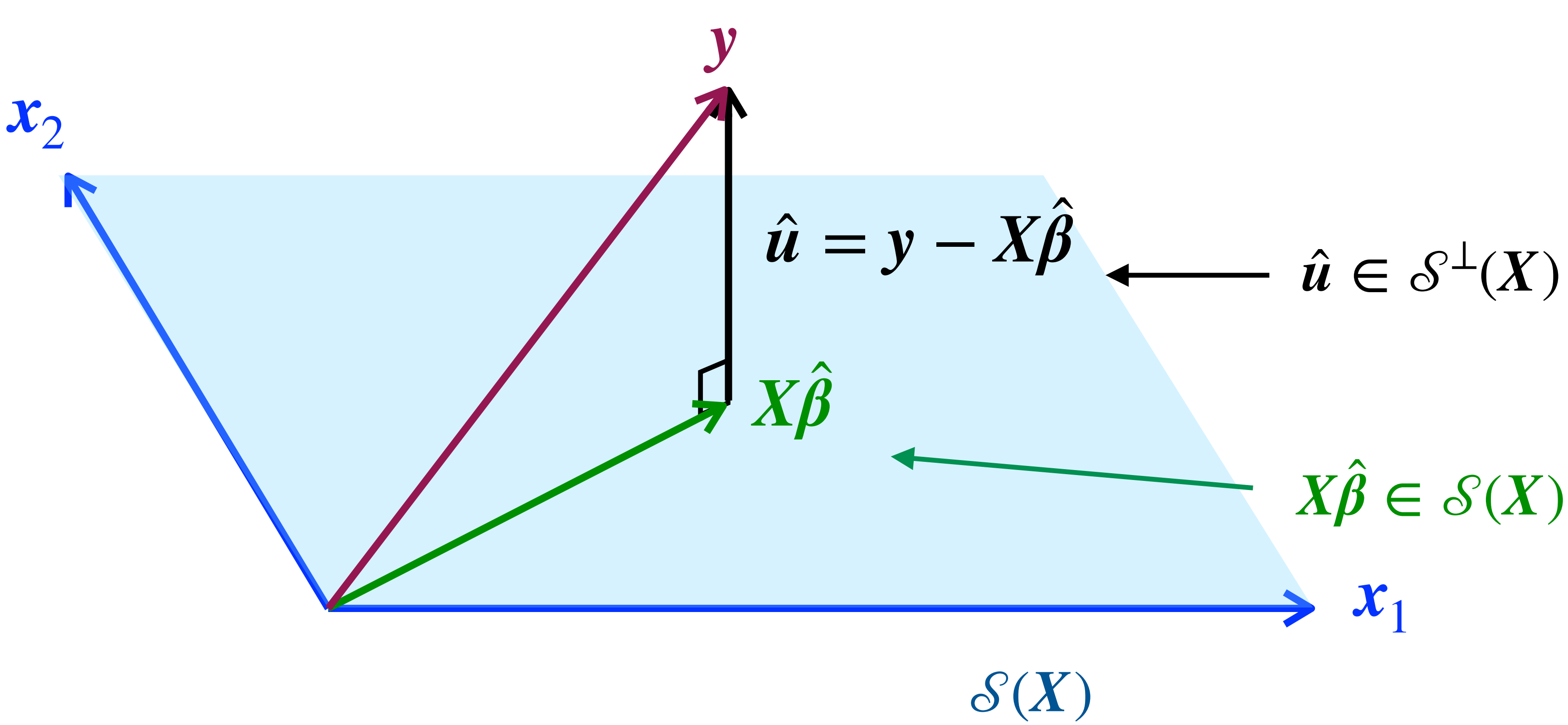
维度：空间的维度等于其基地的数量

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



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





$\hat{\boldsymbol{u}} \in \mathcal{S}^{\perp}(\boldsymbol{X})$，因此 $\hat{\boldsymbol{u}} \perp \boldsymbol{x}_1, \hat{\boldsymbol{u}} \perp \boldsymbol{x}_2$

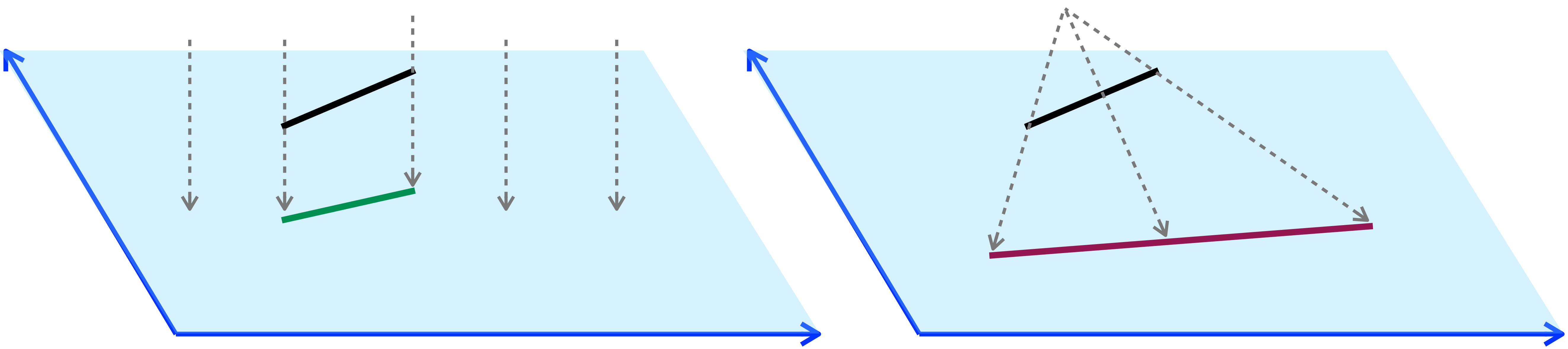
正交投影

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

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

令

$$
\begin{aligned}
&\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top, \\
&\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}}
\end{aligned}
$$

则 $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ 为 $\boldsymbol{y}$ 的 OLS 预测值 $\hat{\boldsymbol{y}}$，$\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} = \hat{\boldsymbol{u}}$ 为 OLS 残差。

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

投影矩阵的性质

- $\boldsymbol{P_X X} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{X}$
- $\boldsymbol{M_X X} = (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = \boldsymbol{O}$（零矩阵）
- $\boldsymbol{P_X}$ 和 $\boldsymbol{M_X}$ 是对称矩阵

注意 $(\boldsymbol{A}^\top)^{-1} = (\boldsymbol{A}^{-1})^\top$

- $\boldsymbol{P_X}$ 和 $\boldsymbol{M_X}$ 是幂等矩阵（idempotent matrix）

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

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

投影的值域与子空间

$\boldsymbol{P}_X$ 的值域是 $\mathcal{S}(\boldsymbol{X})$ 整体。$\boldsymbol{M}_X$ 的值域是 $\mathcal{S}^{\perp}(\boldsymbol{X})$ 整体。

1. 值域包含在子空间内：对于任意的 $\boldsymbol{y} \in \mathbb{R}^n$，
   $\boldsymbol{P}_X\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{X}$ 的列的线性结合，因此 $\boldsymbol{P}_X\boldsymbol{y} \in \mathcal{S}(\boldsymbol{X})$；
   $\boldsymbol{X}^\top\boldsymbol{M}_X\boldsymbol{y} = \boldsymbol{X}^\top\boldsymbol{M}_X{}^\top\boldsymbol{y} = (\boldsymbol{M}_X\boldsymbol{X})^\top\boldsymbol{y} = \boldsymbol{0}$，因此 $\boldsymbol{M}_X\boldsymbol{y} \in \mathcal{S}^{-1}(\boldsymbol{X})$。

2. 子空间包含在值域内：$\mathcal{S}(\boldsymbol{X})$ 中的任意一点 $\boldsymbol{x}$ 都可以写成
   $\boldsymbol{x} = a_1\boldsymbol{x}_1 + \cdots + a_k\boldsymbol{x}_k = \boldsymbol{X}\boldsymbol{a}$，因为 $\boldsymbol{P}_X\boldsymbol{x} = \boldsymbol{P}_X\boldsymbol{X}\boldsymbol{a} = \boldsymbol{X}\boldsymbol{a} = \boldsymbol{x}$，所以 $\mathcal{S}(\boldsymbol{X})$ 中所有的点都是其自身通过 $\boldsymbol{P}_X$ 得到的像。

   $\mathcal{S}^{-1}(\boldsymbol{X})$ 中的任意一点 $\boldsymbol{z}$ 都满足 $\boldsymbol{X}^\top\boldsymbol{z} = \boldsymbol{0}$，此时
   $\boldsymbol{M}_X\boldsymbol{z} = (\boldsymbol{I} - \boldsymbol{P}_X)\boldsymbol{z} = \boldsymbol{z} - \boldsymbol{P}_X\boldsymbol{z}$，因为 $\boldsymbol{P}_X\boldsymbol{z} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{z} = \boldsymbol{0}$，所以 $\boldsymbol{M}_X\boldsymbol{z} = \boldsymbol{z}$，即 $\mathcal{S}^{-1}(\boldsymbol{X})$ 中所有的点都是其自身通过 $\boldsymbol{M}_X$ 得到的像。

其他重要公式

- 标准等式（normal equation）

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

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 $\boldsymbol{P}_X + \boldsymbol{M}_X = \boldsymbol{I}$，因此

$$\boldsymbol{y} = \boldsymbol{P}_X \boldsymbol{y} + \boldsymbol{M}_X \boldsymbol{y} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{u}}$$

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

独立变量的线性变换

子空间 $\mathscr{S}(\boldsymbol{X})$ 的基底

$\mathscr{S}(\boldsymbol{X})$ 可以由不同的基底张成。

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

$$\boldsymbol{X}\boldsymbol{A} = \boldsymbol{X}\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_k\end{bmatrix} = \begin{bmatrix}\boldsymbol{X}\boldsymbol{a}_1 & \boldsymbol{X}\boldsymbol{a}_2 & \dots & \boldsymbol{X}\boldsymbol{a}_k\end{bmatrix}$$

这里每个 $\boldsymbol{X}\boldsymbol{a}_i$ 都是 $\mathscr{S}(\boldsymbol{X})$ 中的一点，因此由 $(\boldsymbol{X}\boldsymbol{a}_1, \boldsymbol{X}\boldsymbol{a}_2, \dots, \boldsymbol{X}\boldsymbol{a}_k)$ 张成的子空间 $\mathscr{S}(\boldsymbol{X}\boldsymbol{A}) \subseteq \mathscr{S}(\boldsymbol{X})$。

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

最终可得 $\mathscr{S}(\boldsymbol{X}) = \mathscr{S}(\boldsymbol{X}\boldsymbol{A})$。

独立变量的线性变换

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

$$
\begin{aligned}
P_{XA} &= XA(A^\top X^\top XA)^{-1}A^\top X^\top \\
&= XAA^{-1}(X^\top X)^{-1}(A^\top)^{-1}A^\top X^\top \\
&= X(X^\top X)^{-1}X^\top \\
&= P_X
\end{aligned}
\quad \Rightarrow \quad P_{XA}y = P_X y
$$

$$
\Rightarrow \quad W_{XA} = W_X,\ W_{XA}y = W_X y
$$

$$
\Rightarrow \quad \hat{\beta}_{XA} = (A^\top X^\top XA)^{-1}A^\top X^\top y = A^{-1}(X^\top X)^{-1}(A^\top)^{-1}A^\top X^\top y = A^{-1}\hat{\beta}_X
$$

针对线性回归 $y = X\beta + u$，若将独立变量进行线性变换，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。

独立变量的单位转换

由上一页的结论可导出：独立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独立变量和常数项。温度的单位可以是摄氏（$\boldsymbol{C}$）或华氏（$\boldsymbol{F}$），二者的关系是 $\boldsymbol{F} = 32 + \frac{9}{5}\boldsymbol{C}$。令 $\boldsymbol{\iota}$ 代表要素为 1 的向量，则

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

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 $[\beta_1 \ \beta_2]$ 和 $[\alpha_1 \ \alpha_2]$，则可得 $\beta_1 = \alpha_1 + 32\alpha_2, \beta_2 = \frac{9}{5}\alpha_2$。