高级计量经济学

Lecture 2: The Linear Model

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

线性模型的矩阵表达

数据矩阵

		y	x_2	x_3	……	x_i	……	x_k
	country	**WAGES**	**PRICES**	**GDP**	**EMPLOY**	**MONEY1**	**MONEY2**	**UNEMPLOY**
观测值 1	**Ausstralia**	4.41	3.75	3.04	1.68	9.1	10.98	8.68
	Austria	4.15	2.71	2.55	0.65	5.37	7.37	5.48
	Belgium	3.99	2.37	2.16	0.34	5.87	-999999	8.49
	Canada	3.76	2.83	2.03	1.17	6.13	8.51	9.51
	Denmark	3.78	2.61	2.02	0.02	3.21	4.51	7.68
	Finland	5.65	3.11	1.78	-1.06	5.97	-999999	10.44
	France	3.55	2.4	2.08	0.28	5.19	4.05	10.84
观测值 t	**Germany**	4.08	2.78	2.71	0.08	9.08	7.59	6.86
	Greece	14.18	13.09	2.08	0.87	14.46	14.14	8.84
	Iceland	-999999	8.42	1.54	-0.13	10.73	13.64	3.07
	Ireland	4.5	2.51	6.4	2.16	-999999	-999999	13.81
	Italy	6.16	4.86	1.68	-0.3	4.75	4.17	10.46
	Japan	2.28	1.47	2.81	1.06	5.56	4.49	2.65
	Korea	12.83	6.13	7.73	2.57	-999999	17.59	2.4
	Luxembourg	-999999	2.53	5.64	3.02	-999999	7.22	2.39
	Mexico	-999999	30.59	2.67	4.56	-999999	32.53	3.71
	Netherlands	2.59	2.22	2.86	1.88	6.31	6.61	6.41
观测值 n	**Norway**	4.31	3.12	2.98	0.36	5.98	4.66	5.07

线性模型的矩阵表达

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

线性回归模型 $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$

我们规定 $\boldsymbol{X}_t$ 代表矩阵 $\boldsymbol{X}$ 的第 t 行，
$\boldsymbol{x}_i$ 代表矩阵 $\boldsymbol{X}$ 的第 i 列。

$$\Leftrightarrow \quad \begin{aligned} y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + u_1 \\ y_2 &= \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + u_2 \\ &\vdots \\ y_n &= \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + u_n \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} y_1 &= \boldsymbol{X}_1\boldsymbol{\beta} + u_1 \\ y_2 &= \boldsymbol{X}_2\boldsymbol{\beta} + u_2 \\ &\vdots \\ y_n &= \boldsymbol{X}_n\boldsymbol{\beta} + u_n \end{aligned}$$

一般情况下我们设 $\boldsymbol{X}$ 的第一列中的要素都为 1，则 β_1 为截距。

系数的估计

矩估计

Method-of-moments Estimation

k 阶理论矩：$E[X^k]$　　　k 阶样本矩：$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

矩估计就是假设样本矩等于理论矩，从而建立方程，以此求出参数的估计值。例如，总体均值为一阶矩 $\mu = E[X]$，则其估计值为 $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$。

针对 $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$，我们假设 $E[\boldsymbol{X}_t^\top u_t] = \mathbf{0}$ for all t，则

$$\mathbf{0} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_t^\top u_t = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_t^\top (y_t - \boldsymbol{X}_t\boldsymbol{\beta})$$

$$= \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_t^\top y_t - \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_t^\top \boldsymbol{X}_t\boldsymbol{\beta}$$

$$\Rightarrow \quad \boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{X}^\top \boldsymbol{X\beta}$$

$$\Rightarrow \quad \boxed{\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}}$$

MM估计量

最小二乘估计

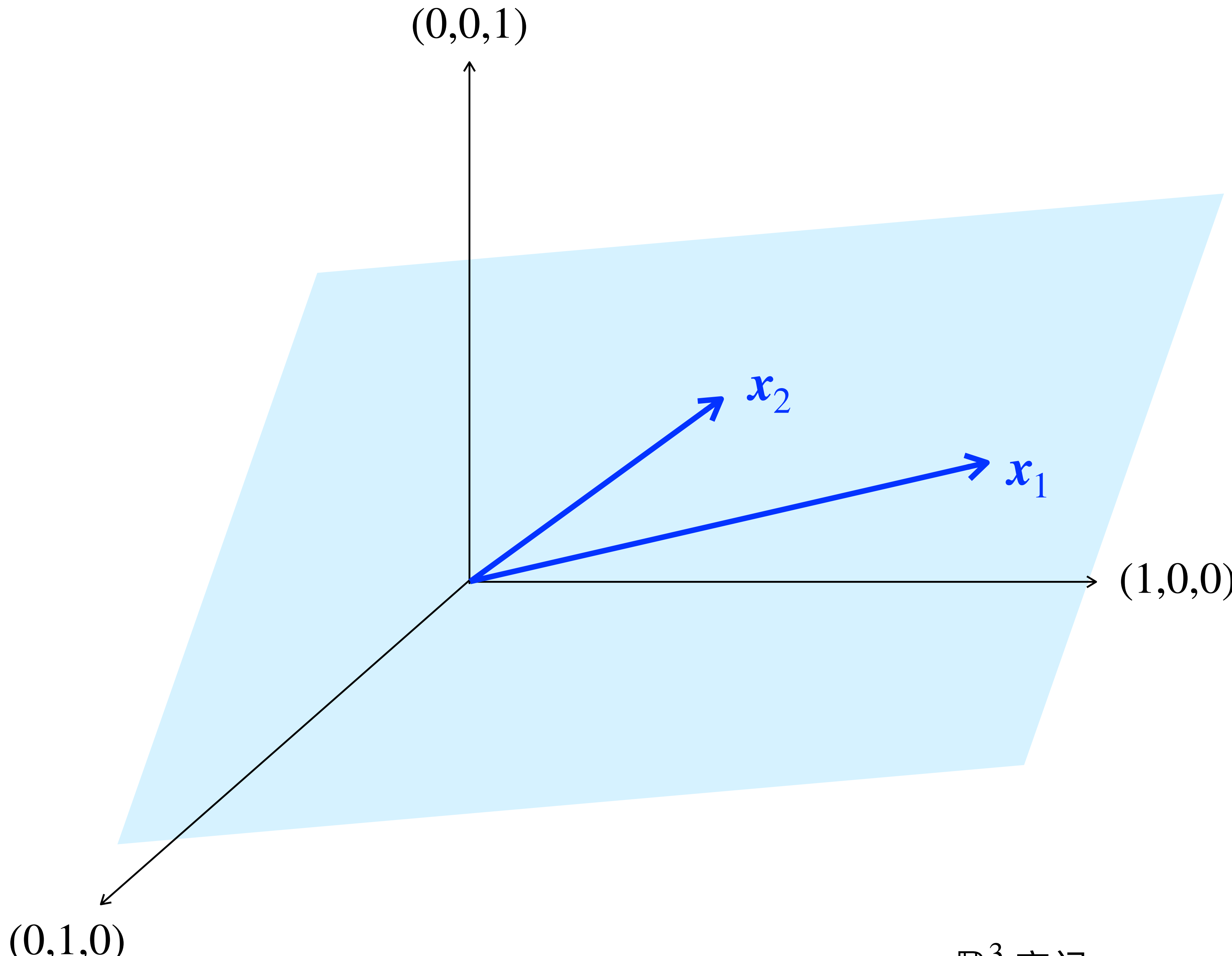
Ordinary Least Squares Estimation

针对任意 $\boldsymbol{\beta}$，我们称 $\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}$ 为残差（residuals）。残差平方和（sum of squared residuals）为

$$\mathrm{SSR}(\boldsymbol{\beta})=\sum_{t=1}^{n}(y_t-\boldsymbol{X}_t\boldsymbol{\beta})^2=(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^{\top}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$$

OLS 的目的是找到使 $\mathrm{SSR}(\boldsymbol{\beta})$ 取值最小的 $\hat{\boldsymbol{\beta}}$，即

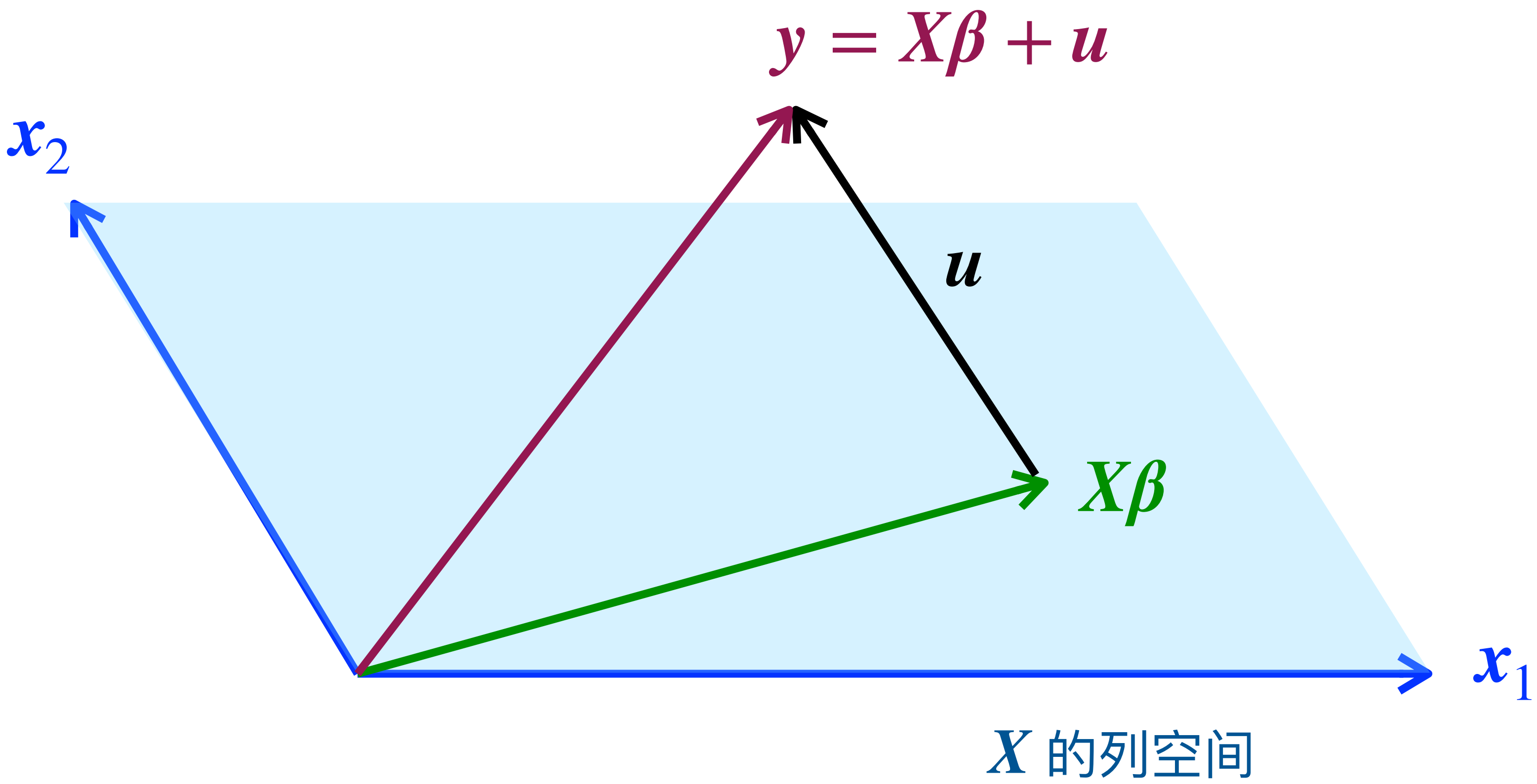
$$\hat{\boldsymbol{\beta}}=\underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\arg\min}\ \mathrm{SSR}(\boldsymbol{\beta})$$

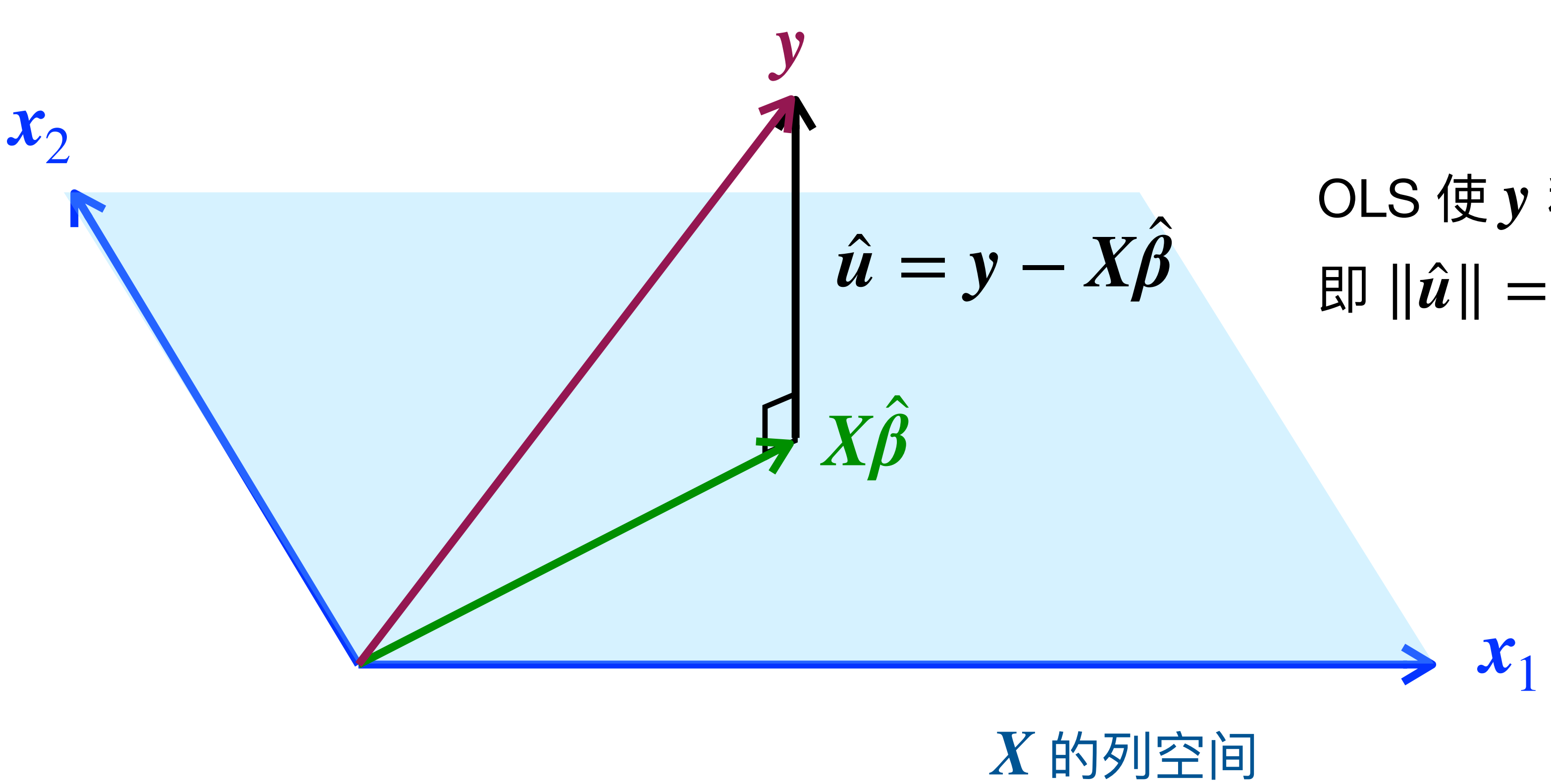


由 $a\boldsymbol{x}_1 + b\boldsymbol{x}_2$ 定义的平面

平面上的任意一点都可以写成 $\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix} = \boldsymbol{X}\boldsymbol{a}$ 的形式，因此也称作 $\boldsymbol{X}$ 的列空间（column space）。

$\mathbb{R}^3$ 空间





OLS 使 $\boldsymbol{y}$ 和 $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ 间的距离最小,
即 $\|\hat{\boldsymbol{u}}\| = \sqrt{\hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}}} \leq \sqrt{\boldsymbol{u}^\top \boldsymbol{u}} = \|\boldsymbol{u}\|$

OLS 估计量

$$
\begin{aligned}
\mathrm{SSR}(\boldsymbol{\beta}) &= (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \\
&= \boldsymbol{y}^\top\boldsymbol{y} - \boldsymbol{y}^\top\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{y} + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{y}^\top\boldsymbol{y} - 2\boldsymbol{y}^\top\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}
\end{aligned}
$$

若 $\boldsymbol{A}$ 为对称矩阵，则 $\frac{\partial \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$。$\boldsymbol{X}^\top\boldsymbol{X}$ 为对称矩阵。

若 $\boldsymbol{a}^\top\boldsymbol{x} = \sum_{i=1}^n a_i x_i$，则 $\frac{\partial \boldsymbol{a}\boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial \boldsymbol{a}\boldsymbol{x}}{\partial x_1} \\ \vdots \\ \frac{\partial \boldsymbol{a}\boldsymbol{x}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \boldsymbol{a}$。

一阶条件为

$$
\frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \beta_1} \\ \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \beta_2} \\ \vdots \\ \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \beta_k} \end{bmatrix} = -2\boldsymbol{X}^\top\boldsymbol{y} + 2\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}
$$

$$
\Rightarrow \quad \boldsymbol{X}^\top\boldsymbol{y} = \boldsymbol{X}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}}
$$

$$
\Rightarrow \quad \boxed{\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}}
$$

OLS估计量 = MM估计量

其他必要条件

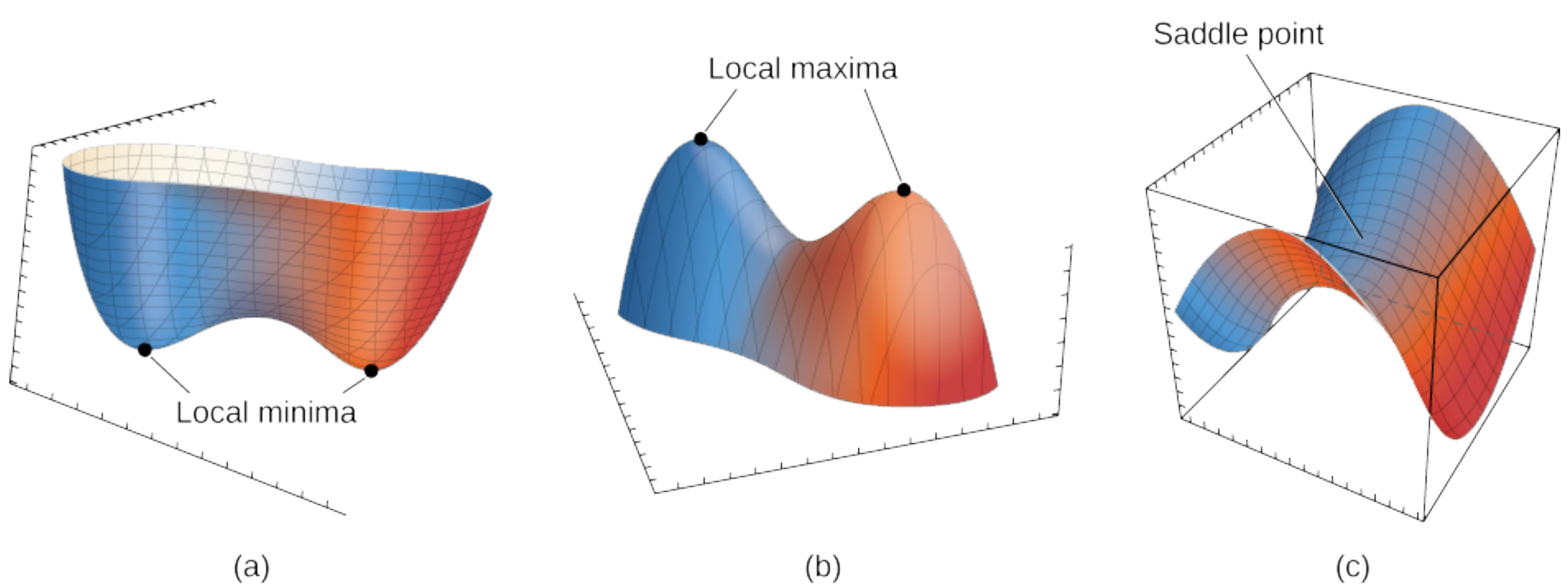
- OLS 解的存在条件

为保证 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ 存在，我们需要保证 $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ 存在，即 $\mathrm{rank}(\boldsymbol{X}^\top \boldsymbol{X}) = k$（full rank 满秩）。该条件可以通过假设 $\mathrm{rank}(\boldsymbol{X}) = k$ 得到满足。

OLS假设之一，若在 $n > k$ 时 $\mathrm{rank}(\boldsymbol{X}) < k$，就会出现共线性问题。

- OLS 解真的使 SSR 最小化

一般情况下一阶条件不能保证最优解：



图片地址 https://math.libretexts.org/Courses/University_of_California_Davis/UCD_Mat_21C:_Multivariate_Calculus/13:_Partial_Derivatives/13.7:_Extreme_Values_and_Saddle_Points

我们需要讨论二阶条件的 Hessian 矩阵。

二阶条件

Hessian Matrix

$$\frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\;\partial \boldsymbol{\beta}^{\top}} = \begin{bmatrix} \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial {\beta_1}^2} & \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \beta_1\;\partial \beta_2} & \cdots & \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \beta_1\;\partial \beta_k} \\ \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \beta_2\;\partial \beta_1} & \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial {\beta_2}^2} & \cdots & \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \beta_2\;\partial \beta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \beta_k\;\partial \beta_1} & \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial \beta_k\;\partial \beta_2} & \cdots & \frac{\partial\,\mathrm{SSR}(\boldsymbol{\beta})}{\partial {\beta_k}^2} \end{bmatrix} = 2X^{\top}X$$ 是正定矩阵（positive definite）。

$\boxed{\boldsymbol{A} \text{ 是正定矩阵} \Leftrightarrow \boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x} > 0 \text{ for all } \boldsymbol{x} \neq 0}$

如果 $n \times k$ 矩阵 $\boldsymbol{A}$ 满足 $n > k$, $\mathrm{rank}(\boldsymbol{A}) = k$，则 $\boldsymbol{A}^{\top}\boldsymbol{A}$ 是正定矩阵，$\boldsymbol{A}\boldsymbol{A}^{\top}$ 是半正定矩阵（positive semidefinite）。

证明：

因为 $\mathrm{rank}(\boldsymbol{A}) = k$，$\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{0}$，因此

$$\boldsymbol{x}^{\top}(\boldsymbol{A}^{\top}\boldsymbol{A})\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^{\top}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{z}^{\top}\boldsymbol{z} > 0$$

因为 $n > k$，且 $\mathrm{rank}(\boldsymbol{A}^{\top}) = \mathrm{rank}(\boldsymbol{A}) = k$，所以 $\boldsymbol{s} = \boldsymbol{A}^{\top}\boldsymbol{x}$ 可以为 $\boldsymbol{0}$。因此

$$\boldsymbol{x}^{\top}(\boldsymbol{A}\boldsymbol{A}^{\top})\boldsymbol{x} = (\boldsymbol{A}^{\top}\boldsymbol{x})^{\top}(\boldsymbol{A}^{\top}\boldsymbol{x}) = \boldsymbol{s}^{\top}\boldsymbol{s} \geq 0 \qquad \square$$

$\Rightarrow\; 2X^{\top}X$ 是正定矩阵，满足二阶条件。

二阶条件的含义：$\mathrm{SSR}(\boldsymbol{\beta})$ 是 $\boldsymbol{\beta}$ 的凸函数（convex function）。

一个特例的推导

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$$

求解 $\min_{\beta_1, \beta_2} \sum \left[y_i - (\beta_1 x_{t1} - \beta_2 x_{t2}) \right]^2$

一阶条件为：

$$\frac{\partial S}{\partial \beta_1} = -2 \sum x_{t1} y_i + 2\beta_1 \sum x_{t1}^2 + 2\beta_2 \sum x_{t1} x_{t2} = 0$$

$$\frac{\partial S}{\partial \beta_2} = -2 \sum x_{t2} y_i + 2\beta_1 \sum x_{t1} x_{t2} + 2\beta_2 \sum x_{t2}^2 = 0$$

解一阶条件，得

$$\hat{\beta}_1 = \frac{(\sum x_{t1} y_i)(\sum x_{t2}^2) - (\sum x_{t2} y_i)(\sum x_{t1} x_{t2})}{(\sum x_{t1}^2)(\sum x_{t2}^2) - (\sum x_{t1} x_{t2})^2}$$

$$\hat{\beta}_2 = \frac{(\sum x_{t2} y_i)(\sum x_{t1}^2) - (\sum x_{t1} y_i)(\sum x_{t1} x_{t2})}{(\sum x_{t1}^2)(\sum x_{t2}^2) - (\sum x_{t1} x_{t2})^2}$$

课后练习：

1. 确认左面的结果和公式

   $$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

   得到的结果一致。
2. 确认二阶条件成立。

y 的分解：TSS = ESS + SSR

将 OLS 估计量 $\hat{\boldsymbol{\beta}}$ 代入回归模型，可得

$$\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}, \qquad \hat{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \text{ 是 OLS 残差}$$

此时，

$$\begin{aligned}
\boldsymbol{y}^\top \boldsymbol{y} &= (\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}})^\top (\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}) \\
&= (\boldsymbol{X}\hat{\boldsymbol{\beta}})^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} + (\boldsymbol{X}\hat{\boldsymbol{\beta}})^\top \hat{\boldsymbol{u}} + \hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}} \\
&= \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} + 2\hat{\boldsymbol{u}}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}}
\end{aligned}$$

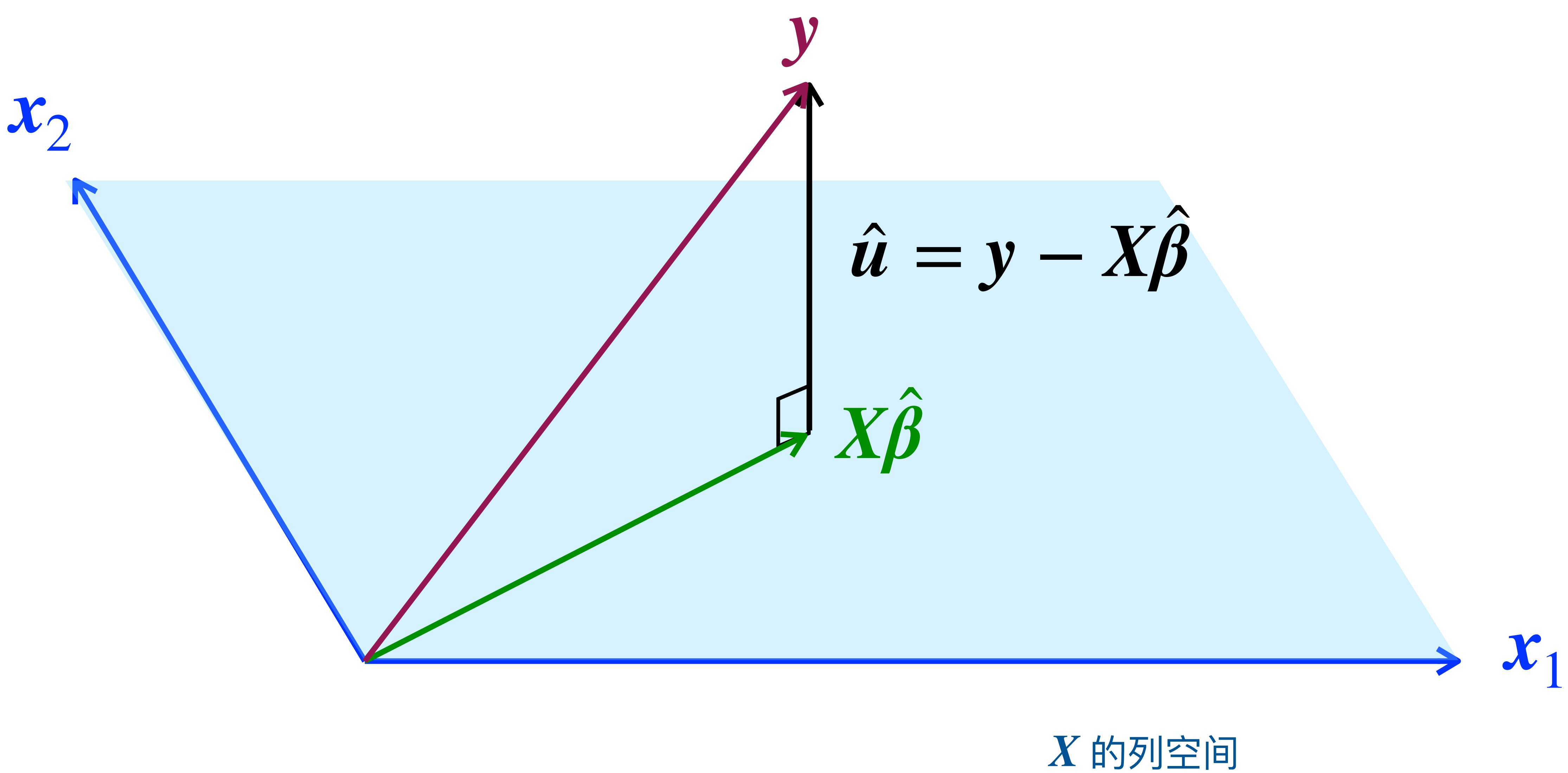
已知 $\hat{\boldsymbol{u}}^\top \boldsymbol{X} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top \boldsymbol{X} = \boldsymbol{y}^\top \boldsymbol{X} - \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{0}^\top$ 由 $\hat{\boldsymbol{\beta}}$ 表达式可得

因此可得出

$$\underbrace{\boldsymbol{y}^\top \boldsymbol{y}}_{\text{Total Sum of Squares}} = \underbrace{\hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}}}_{\text{Explained Sum of Squares}} + \underbrace{\hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}}}_{\text{Sum of Squared Residuals}}$$

$$\boldsymbol{y}^\top \boldsymbol{y} = \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}}$$

$$\Leftrightarrow \quad \|\boldsymbol{y}\|^2 = \|\boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2 + \|\hat{\boldsymbol{u}}\|^2$$ （勾股定理）



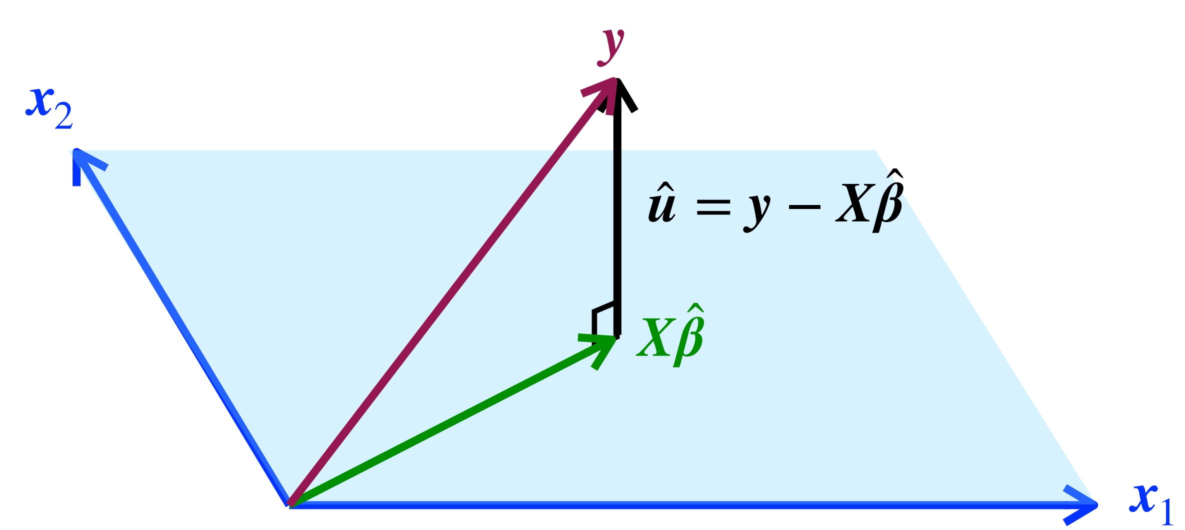
正交向量与 OLS

Orthogonal Vectors and OLS

若向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 满足 $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^\top \boldsymbol{b} = \boldsymbol{b}^\top \boldsymbol{a} = 0$ （即内积为零），则称 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 正交，写为 $\boldsymbol{a} \perp \boldsymbol{b}$。

OLS 估计中，残差 $\hat{\boldsymbol{u}}$ 与 $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ 正交，
因此可得

$$
\begin{aligned}
& (\boldsymbol{X}\hat{\boldsymbol{\beta}})^\top \hat{\boldsymbol{u}} = 0 \\
\Leftrightarrow \quad & (\boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = 0 \\
\Leftrightarrow \quad & \hat{\boldsymbol{\beta}}^\top (\boldsymbol{X}^\top \boldsymbol{y} - \boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}}) = 0
\end{aligned}
$$



假设 $\hat{\boldsymbol{\beta}} \neq \mathbf{0}$，则 $\boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$