高级计量经济学

Appendix 3: Review of Matrix and Linear Algebra

Hansen, B. (2022). Econometrics. Princeton University Press.

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

矩阵和向量

矩阵的表达方式

Matrix notation

- **标量（scalar）**是一个单独的数，一般用标准体小写字母表达，例如 a
- **向量（vector）**是由 k 个数组成的清单，用粗体小写字母表达，例如

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} \quad \text{或} \quad \boldsymbol{\alpha} = (a_1 \quad a_2 \quad \cdots \quad a_k)$$

也可以写成 $\boldsymbol{\alpha} = (a_1, a_2, \cdots, a_k)$

- **矩阵（matrix）**是由 $k \times r$ 个数字组成二维排列，一般用粗体大写字母表达，例如

$$\boldsymbol{A} = \begin{array}{cl} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2r} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & \boxed{a_{ij}} & \cdots & a_{ir} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kj} & \cdots & a_{kr} \end{pmatrix} & \begin{array}{l} \\ \\ \\ \longleftarrow \text{第 } i \text{ 行} \\ \\ \\ \end{array} \\ \begin{array}{c} \qquad\qquad\qquad \uparrow \\ \qquad\qquad\qquad \text{第 } j \text{ 列} \end{array} & \end{array}$$

a_{ij} 称为 $\boldsymbol{A}$ 的 (i, j) 要素

$\boldsymbol{A}$ 也可以简写成 $\boldsymbol{A} = (a_{ij})_{k \times r}$ 或 $\boldsymbol{A} = (a_{ij})$

矩阵的表达方式

Matrix notation

- 矩阵的**转置**（**transpose**）$\boldsymbol{A}^\top$ 是将矩阵 $\boldsymbol{A}$ 以主对角线为轴翻转后获得的矩阵，即

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1r} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kr} \end{pmatrix} \quad \Leftrightarrow \quad \boldsymbol{A}^\top = \begin{pmatrix} a_{11} & \cdots & a_{k1} \\ \vdots & & \vdots \\ a_{1r} & \cdots & a_{kr} \end{pmatrix}$$

例：$\begin{pmatrix} 2 & 3 & 0 \\ -1 & 1 & 5 \end{pmatrix}^\top = \begin{pmatrix} 2 & -1 \\ 3 & 1 \\ 0 & 5 \end{pmatrix}$

也可以写作 $\boldsymbol{A}'$ 或 $\boldsymbol{A}^t$

- 如果 $\boldsymbol{A}$ 的行数等于列数，则称其为**方阵**（**square matrix**）
- 方阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A} = \boldsymbol{A}^\top$，即 $a_{ij} = a_{ji}$ 时，称其为**对称**（**symmetric**）矩阵
- $k \times k$ **单位矩阵**（**identity matrix**）$\boldsymbol{I}_k$ 是主对角要素为 1，其他要素为 0 的方阵，即

$$\boldsymbol{I}_k = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

- 主对角要素下面/上面的要素都为零的方阵称为上三角/下三角（upper triangular/lower triangular）矩阵，两者统称三角矩阵
- **分块矩阵**（**partitioned matrix**）是将一个大矩阵表达成几个小矩阵的矩阵，例如 $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}$

矩阵计算

Matrix operations

- 矩阵的加法：$\boldsymbol{A}+\boldsymbol{B}=(a_{ij}+b_{ij})$
- 矩阵乘以标量：$c\boldsymbol{A}=\boldsymbol{A}c=(c\times a_{ij})$
- 向量的内积（inner product）： 两个列向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的内积定义为

$$\boldsymbol{a}\cdot\boldsymbol{b}=\boldsymbol{a}^\top\boldsymbol{b}=a_1b_1+a_2b_2+\cdots+a_kb_k=\sum_{i=1}^{k}a_kb_k$$

- 矩阵的乘法：$k\times r$ 矩阵 $\boldsymbol{A}$ 和 $r\times s$ 矩阵 $\boldsymbol{B}$ 的积定义为

$$\boldsymbol{AB}=\begin{pmatrix}\boldsymbol{a}_1^\top\\\boldsymbol{a}_2^\top\\\vdots\\\boldsymbol{a}_k^\top\end{pmatrix}\begin{pmatrix}\boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_s\end{pmatrix}=\begin{pmatrix}\boldsymbol{a}_1^\top\boldsymbol{b}_1 & \boldsymbol{a}_1^\top\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1^\top\boldsymbol{b}_s\\\boldsymbol{a}_2^\top\boldsymbol{b}_1 & \boldsymbol{a}_2^\top\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2^\top\boldsymbol{b}_s\\\vdots & \vdots & & \vdots\\\boldsymbol{a}_k^\top\boldsymbol{b}_1 & \boldsymbol{a}_k^\top\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_k^\top\boldsymbol{b}_s\end{pmatrix}$$

也可以表达为 $\boldsymbol{C}=\boldsymbol{AB}\ \Leftrightarrow\ c_{ij}=\boldsymbol{a}_i^\top\boldsymbol{b}_j$，$\boldsymbol{C}$ 为 $k\times s$ 矩阵

矩阵计算的性质

Some properties of matrix operations

- 矩阵的加法满足<u>交换律</u>和<u>结合律</u>
  - 交换律：$\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A}$
  - 结合律：$\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})=(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}$
- 矩阵的乘法满足<u>结合律</u>和<u>分配律</u>
  - 结合律：$\boldsymbol{A}(\boldsymbol{B}\boldsymbol{C})=(\boldsymbol{A}\boldsymbol{B})\boldsymbol{C}$
  - 分配律：$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C})=\boldsymbol{A}\boldsymbol{B}+\boldsymbol{A}\boldsymbol{C}$
- 矩阵的乘法<u>不满足交换律</u>：$\boldsymbol{A}\boldsymbol{B}\neq\boldsymbol{B}\boldsymbol{A}$ 举例证明这条性质
- 与单位矩阵的积：$k\times r$ 矩阵 $\boldsymbol{A}$ 满足 $\boldsymbol{I}_k\boldsymbol{A}=\boldsymbol{A}\boldsymbol{I}_r=\boldsymbol{A}$
- **正交（orthogonal）**向量与矩阵
  - 正交向量：如果 $\boldsymbol{a}^\top\boldsymbol{b}=0$，则称项量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 正交
  - 正交矩阵：如果 $\boldsymbol{A}^\top\boldsymbol{B}=\boldsymbol{O}$，则称矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 正交。这里 $\boldsymbol{O}$ 为零矩阵。$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 正交意味着 $\boldsymbol{A}$ 的任意列向量 与 $\boldsymbol{B}$ 的任意列向量正交

方阵的迹

Trace

- $k \times k$ 方阵 $\boldsymbol{A}$ 的**迹（trace）**定义为其主对角要素之和，即

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{k} a_{ii}$$

- 一些显而易见的性质
  - $\operatorname{tr}(c\boldsymbol{A}) = c \operatorname{tr}(\boldsymbol{A})$
  - $\operatorname{tr}(\boldsymbol{A}^\top) = \operatorname{tr}(\boldsymbol{A})$
  - $\operatorname{tr}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B})$
  - $\operatorname{tr}(\boldsymbol{I}_k) = k$
  - 对于 $k \times r$ 矩阵 $\boldsymbol{A}$ 和 $r \times k$ 矩阵 $\boldsymbol{B}$，$\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$ 尝试证明这条性质

矩阵的秩与逆

Rank and inverse

- $k \times r$ 矩阵 $\boldsymbol{A} = (\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_r)$ 的**秩（rank）**是 $\boldsymbol{A}$ 的线性独立的列的最大个数，记作 $\mathrm{rank}(\boldsymbol{A})$
- $\mathrm{rank}(\boldsymbol{A}) \leq \min(k, r)$
- 如果 $\mathrm{rank}(\boldsymbol{A}) = \min(k, r)$，则称 $\boldsymbol{A}$ **满秩（has full rank）**
- 当 $r \leq k$ 时，$\mathrm{rank}(\boldsymbol{A}) = r$ 称为列满秩。列满秩代表所有的列都是线性独立的
- $k \times k$ 方阵 $\boldsymbol{A}$ 如果满足 $\mathrm{rank}(\boldsymbol{A}) = k$，则称 $\boldsymbol{A}$ 为**非奇异矩阵（non-singular matrix）**。这意味着不存在能使 $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{0}$ 的非零向量 $\boldsymbol{c}$
- 如果 $k \times k$ 方阵 $\boldsymbol{A}$ 是非奇异矩阵，则存在唯一的 $k \times k$ 方阵 $\boldsymbol{B}$，使

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}_k$$

  此时，称 $\boldsymbol{B}$ 为 $\boldsymbol{A}$ 的**逆矩阵（inverse matrix）**，并写作 $\boldsymbol{A}^{-1} = \boldsymbol{B}$
- $\boldsymbol{A}$ 与 $\boldsymbol{C}$ 是非奇异矩阵时，
  - $(\boldsymbol{A}^\top)^{-1} = (\boldsymbol{A}^{-1})^\top$
  - $(\boldsymbol{AC})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{A}^{-1}$

行列式

Determinant

- 行列式（determinant）是每个方阵对应的一个标量，定义为

$$\det(\boldsymbol{A}) = |\boldsymbol{A}| = \sum_{\pi} \varepsilon_{\pi}\, a_{1j_1} a_{2j_2} \cdots a_{kj_k}$$

此式中 $\pi = (j_1, j_2, \ldots, j_k)$ 是 $(1,2,\ldots,k)$ 的一个排列，而每个 π 中较大数字排在较小数字之前的组合数为偶数时 $\varepsilon_{\pi} = 1$，为奇数时 $\varepsilon_{\pi} = -1$

例如：$\varepsilon_{(3,2,1)} = -1$

- 行列式的定义比较复杂，通常我们会直接记住常用的 2×2 和 3×3 方阵的行列式公式

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \qquad \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - afh - bdi, \qquad \begin{vmatrix} a & b & c & a & b \\ d & e & f & d & e \\ g & h & i & g & h \end{vmatrix}$$

- 行列式的一些性质：
  $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\top})$, $\det(c\boldsymbol{A}) = c^k \det(\boldsymbol{A})$, $\det(\boldsymbol{AB}) = \det(\boldsymbol{BA}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$,
  $\det(\boldsymbol{A}^{-1}) = \det(\boldsymbol{A})^{-1}$, $\det(\boldsymbol{A}) \neq 0 \Leftrightarrow \boldsymbol{A}$ 是非奇异矩阵，$\boldsymbol{A}$ 是三角矩阵 $\Leftrightarrow \det(\boldsymbol{A}) = \prod_{i=1}^{k} a_{ii}$

特征值

Eigenvalues

- 关于 $k \times k$ 矩阵 $\boldsymbol{A}$ 的方程 $\boldsymbol{A}\boldsymbol{h} = \lambda \boldsymbol{h}$ 存在非零解 $\boldsymbol{h}$ 时，称 $\boldsymbol{h}$ 为 $\boldsymbol{A}$ 的**特征向量（eigenvector）**，λ 为对应的**特征值（eigenvalue）**

- 方程 $\boldsymbol{A}\boldsymbol{h} = \lambda \boldsymbol{h}$ 可以改写为 $(\boldsymbol{A} - \lambda \boldsymbol{I}_k)\boldsymbol{h} = \boldsymbol{0}$，此时存在非零解的充分必要条件是

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}_k) = 0$$

  此方程称为**特征方程（characteristic equation）**，其左边是关于 λ 的 k 次多项式，因此称之为**特征多项式（characteristic polynomial）**，λ 为特征方程的解（共有 k 个，可重复，可为实数或复数）

- 令 λ_i 和 $\boldsymbol{h}_i$, $i = 1,\ldots,k$ 为 $\boldsymbol{A}$ 的特征值和特征向量，则
  - $\det(\boldsymbol{A}) = \prod_{i=1}^{k} \lambda_i$
  - $\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{k} \lambda_i$
  - $\boldsymbol{A}$ 为非奇异矩阵 $\Leftrightarrow$ 所有的特征值 λ_i 都不为零
  - $\boldsymbol{A}\boldsymbol{B}$ 和 $\boldsymbol{B}\boldsymbol{A}$ 具有相同的非零特征值
  - 如果 $\boldsymbol{B}$ 是非奇异矩阵，则 $\boldsymbol{A}$ 与 $\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}$ 具有相同的特征值

谱分解

Spectral decomposition

- 令 λ_i 和 $\boldsymbol{h}_i$, $i = 1,\ldots,k$ 为 $\boldsymbol{A}$ 的特征值和特征向量，则方程 $\boldsymbol{A}\boldsymbol{h} = \lambda\boldsymbol{h}$ 可以写为下面的矩阵形式

$$\boldsymbol{A}\boldsymbol{H} = \boldsymbol{H}\boldsymbol{\Lambda}$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix}$$

其中，$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{h}_1 & \cdots & \boldsymbol{h}_k \end{pmatrix}, \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$

谱分解（spectral decomposition）：如果 $\boldsymbol{A}$ 是 $k \times k$ 的实对称矩阵，则 $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top$。$\boldsymbol{A}$ 的所有特征值都为实数，且 $\boldsymbol{H}$ 满足 $\boldsymbol{H}^\top\boldsymbol{H} = \boldsymbol{I}_k$

- 如果 $\boldsymbol{A}$ 是可逆实对称矩阵，则 $\boldsymbol{A}^{-1} = (\boldsymbol{H}^\top)^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^{-1} = \boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^\top$，因此 $\boldsymbol{A}^{-1}$ 与 $\boldsymbol{A}$ 拥有相同的特征向量，其特征值是 $\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_k^{-1}$

（半）正定矩阵

Positive (semi)definite matrices

如果 $k \times k$ 的实对称矩阵 $\boldsymbol{A}$，对于任意非零向量 $\boldsymbol{c} \neq \mathbf{0}$ 都有

$$\boldsymbol{c}^\top \boldsymbol{A} \boldsymbol{c} > 0 \quad (\boldsymbol{c}^\top \boldsymbol{A} \boldsymbol{c} \geq 0)$$

则称 $\boldsymbol{A}$ 为（半）**正定矩阵（positive (semi)definite matrix）**，并可记作 $\boldsymbol{A} > 0$ （$\boldsymbol{A} \geq 0$）

$(c_1 \quad c_2)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = c_1^2 - 2c_1c_2 + c_2^2 = (c_1 - c_2)^2 \geq 0$，因此 $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ 是半正定矩阵

- $\boldsymbol{c}^\top \boldsymbol{A} \boldsymbol{c} = \sum_{i=1}^{k} \sum_{j=1}^{k} c_i a_{ij} c_j = \sum_{i=1}^{k} a_{ii} c_i^2 + \sum_{i \neq j} a_{ij} c_i c_j$ 是关于 c_i 的二次函数，因此称为**二次型（quadratic form）**
- 如果 $\boldsymbol{A}$ 是正定矩阵，则 $\boldsymbol{A}$ 为非奇异矩阵（$\det(\boldsymbol{A}) > 0$），且 $\boldsymbol{A}^{-1}$ 也是正定矩阵
- $\boldsymbol{A}$ 是（半）正定矩阵 $\Leftrightarrow$ $\boldsymbol{A}$ 的所有特征值都为正（非负）实数
- 如果 $\boldsymbol{A}$ 是半正定矩阵，则 $\operatorname{rank}(\boldsymbol{A})$ 等于正特征值的个数

行列式与正定矩阵

Determinants and positive definite matrices

- 从 $k \times k$ 矩阵 $\boldsymbol{A}$ 中删除任意 $k-r$ 行和任意 $k-r$ 列后，剩下的 $r \times r$ 矩阵的行列式称为 $\boldsymbol{A}$ 的 r 阶**余子式（minor）**

- 如果第 i 行被删除时第 i 列也被删除，则剩下的 $r \times r$ 矩阵的行列式称为 $\boldsymbol{A}$ 的 r 阶**主子式（principle minor）**

- 由 $\boldsymbol{A}$ 的前 r 行和前 r 列组成的 $r \times r$ 矩阵的行列式称为 $\boldsymbol{A}$ 的 r 阶**顺序主子式（leading principle minor）**

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kk} \end{pmatrix}$$

- $\boldsymbol{A}$ 是正定矩阵
  $\Leftrightarrow$ $\boldsymbol{A}$ 的所有<u>顺序主子式</u>都为正

- $\boldsymbol{A}$ 是半正定矩阵
  $\Leftrightarrow$ $\boldsymbol{A}$ 的所有<u>主子式</u>都为非负

例：

3×3 矩阵 $\boldsymbol{A} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$

有 7 个主子式：

a_{11}，a_{22}，a_{33}，$|\boldsymbol{A}|$，以及

$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, $\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$, $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$

其中，a_{11}，$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$，$|\boldsymbol{A}|$ 是顺序主子式

Cholesky 分解

Cholesky decomposition

- 如果 $k \times k$ 实矩阵 $\boldsymbol{A}$ 是正定矩阵，则存在矩阵 $\boldsymbol{B}$ 使 $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}^\top$。我们称 $\boldsymbol{B}$ 为 $\boldsymbol{A}$ 的矩阵平方根（matrix square root）并记作 $\boldsymbol{B} = \boldsymbol{A}^{\frac{1}{2}}$。矩阵平方根并不是唯一的

Cholesky 分解：如果 $k \times k$ 实矩阵 $\boldsymbol{A}$ 是正定矩阵，则存在唯一满秩的下三角矩阵 $\boldsymbol{L}$ 使 $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$

- 例如对于 3×3 矩阵，

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} = \begin{pmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{pmatrix} \begin{pmatrix} \ell_{11} & \ell_{21} & \ell_{31} \\ 0 & \ell_{22} & \ell_{32} \\ 0 & 0 & \ell_{33} \end{pmatrix} = \begin{pmatrix} \ell_{11}^2 & \ell_{11}\ell_{21} & \ell_{11}\ell_{31} \\ \ell_{21}\ell_{11} & \ell_{21}^2 + \ell_{22}^2 & \ell_{21}\ell_{31} + \ell_{22}\ell_{32} \\ \ell_{31}\ell_{11} & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} & \ell_{31}^2\ell_{32}^2\ell_{33}^2 \end{pmatrix}$$

注意：$\boldsymbol{A}$ 是正定矩阵 $\Rightarrow$ $\boldsymbol{A}$ 是对称矩阵

这里共有 6 个线性方程与 6 个未知量（$\boldsymbol{L}$ 的要素），$\boldsymbol{A}$ 是正定矩阵（所有顺序主子式为正）可以保证解的存在性和唯一性，并保证 $\boldsymbol{L}$ 的主对角要素为正

- $\boldsymbol{L}$ 是 $\boldsymbol{A}$ 的一个矩阵平方根

多变量函数求导

Matrix calculus

- $\boldsymbol{x} = (x_1, x_2, \cdots, x)^\top$ 为 k 维实数向量，令 $g(\boldsymbol{x}) = g(x_1, x_2, \cdots, x_k) : \mathbb{R}^k \to \mathbb{R}$ 为 $\boldsymbol{x}$ 的实数值函数
- 称

$$\frac{\partial}{\partial \boldsymbol{x}} g(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} g(\boldsymbol{x}) \\ \frac{\partial}{\partial x_2} g(\boldsymbol{x}) \\ \vdots \\ \frac{\partial}{\partial x_k} g(\boldsymbol{x}) \end{pmatrix} \quad 和 \quad \frac{\partial}{\partial \boldsymbol{x}^\top} g(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} g(\boldsymbol{x}) & \frac{\partial}{\partial x_2} g(\boldsymbol{x}) & \cdots & \frac{\partial}{\partial x_k} g(\boldsymbol{x}) \end{pmatrix}$$

为 g 的**梯度（gradient）**向量

- 称

$$\frac{\partial^2}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} g(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} g(\boldsymbol{x}) & \frac{\partial^2}{\partial x_1 \partial x_2} g(\boldsymbol{x}) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_k} g(\boldsymbol{x}) \\ \frac{\partial^2}{\partial x_2 \partial x_1} g(\boldsymbol{x}) & \frac{\partial^2}{\partial x_2^2} g(\boldsymbol{x}) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_k} g(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_k \partial x_1} g(\boldsymbol{x}) & \frac{\partial^2}{\partial x_k \partial x_2} g(\boldsymbol{x}) & \cdots & \frac{\partial^2}{\partial x_k^2} g(\boldsymbol{x}) \end{pmatrix}$$

为 g 的**海赛（Hessian）**矩阵

多变量函数求导

Matrix calculus

- $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{a}^\top \boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^\top \boldsymbol{a}) = \boldsymbol{a}$

$$\boldsymbol{a}^\top \boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{a} = \textstyle\sum_{i=1}^{k} a_i x_i \quad \Rightarrow \quad \frac{\partial}{\partial x_i}(\boldsymbol{a}^\top \boldsymbol{x}) = \frac{\partial}{\partial x_i}(\boldsymbol{x}^\top \boldsymbol{a}) = a_i$$

- $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^\top A) = A, \ \frac{\partial}{\partial \boldsymbol{x}^\top}(A\boldsymbol{x}) = A$

$$\text{令 } A = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k),\ \text{则 } \boldsymbol{x}^\top A = (\boldsymbol{x}^\top \boldsymbol{a}_1, \boldsymbol{x}^\top \boldsymbol{a}_2, \cdots, \boldsymbol{x}^\top \boldsymbol{a}_k)$$

$$\Rightarrow \quad \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^\top A) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^\top \boldsymbol{a}_1, \boldsymbol{x}^\top \boldsymbol{a}_2, \cdots, \boldsymbol{x}^\top \boldsymbol{a}_k) = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k) = A$$

- $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^\top A\boldsymbol{x}) = (A + A^\top)\boldsymbol{x}, \ \frac{\partial^2}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top}(\boldsymbol{x}^\top A\boldsymbol{x}) = A + A^\top$

如果 A 是对称矩阵，则 $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^\top A\boldsymbol{x}) = 2A\boldsymbol{x}, \ \frac{\partial^2}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top}(\boldsymbol{x}^\top A\boldsymbol{x}) = 2A$

$$\boldsymbol{x}^\top A\boldsymbol{x} = \textstyle\sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} x_i x_j$$

$$\Rightarrow \quad \frac{\partial}{\partial x_i}(\boldsymbol{x}^\top A\boldsymbol{x}) = a_{ij} x_j + a_{ji} x_j, \quad \frac{\partial^2}{\partial x_i \partial x_j}(\boldsymbol{x}^\top A\boldsymbol{x}) = a_{ij} + a_{ji}$$

多变量函数的极值和最值

Extrema of functions of several variables

$\boldsymbol{x}^*$ 是函数 g 的**驻点**（**stationary point**，包括极大值点、极小值点和鞍点）$\Leftrightarrow$ $\frac{\partial}{\partial \boldsymbol{x}} g(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{x}^*} = \boldsymbol{0}$

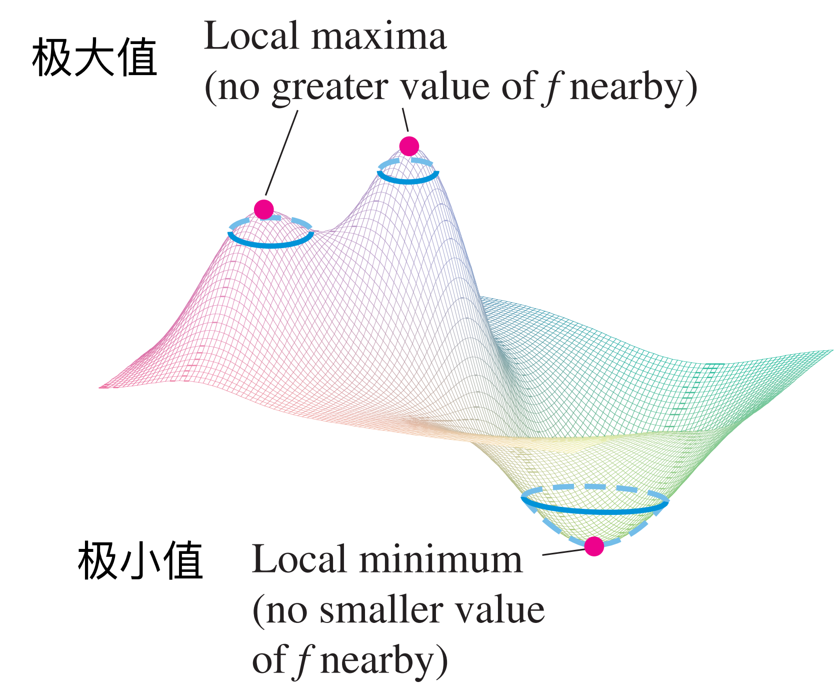
- 必要条件

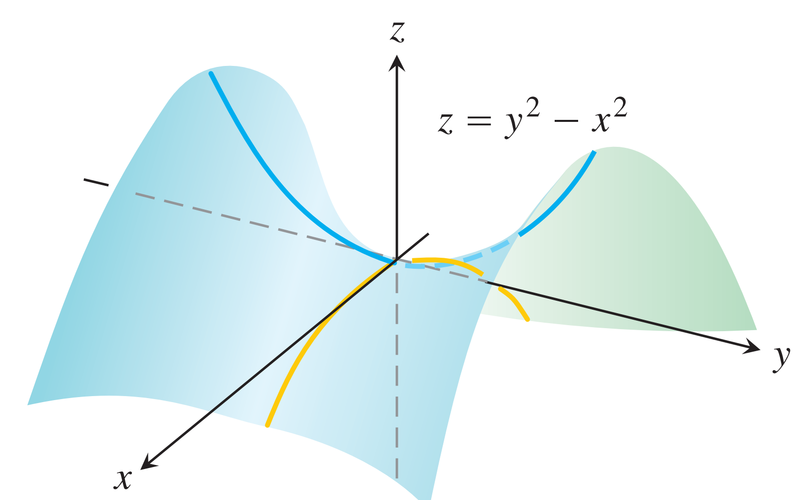
  $\boldsymbol{x}^*$ 是函数 g 的<u>最值点</u> $\Rightarrow$ $\frac{\partial}{\partial \boldsymbol{x}} g(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{x}^*} = \boldsymbol{0}$

- 充分条件

  令 $S \subseteq \mathbb{R}^k$ 为凸集合，并假设 $\boldsymbol{x}^* \in S$ 是函数 g 的一个驻点，则

  - 海赛矩阵 $\frac{\partial^2}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} g(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{x}^*}$ 是<u>正定矩阵</u> $\Rightarrow$ $\boldsymbol{x}^*$ 是 g 的<u>极小值点</u>
  - g 在 S 中是凸函数（$\Leftrightarrow$ 对于任意 $\boldsymbol{x} \in S$，海赛矩阵 $\frac{\partial^2}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} g(\boldsymbol{x})$ 都是<u>半正定矩阵</u>） $\Rightarrow$ $\boldsymbol{x}^*$ 是 g 在 S 中的<u>最小值点</u>





鞍点（saddle point）

图出自 Thomas, Weir, & Hass. *Thomas' Calculus*, 12e. Addison-Wesley