

第七讲：凸函数

优化问题的约束和目标函数

杨 林

大 纲

1. 凸函数的定义

2. 一阶条件

3. 二阶条件

4. 单变量条件

大 纲

1. 凸函数的定义

2. 一阶条件

3. 二阶条件

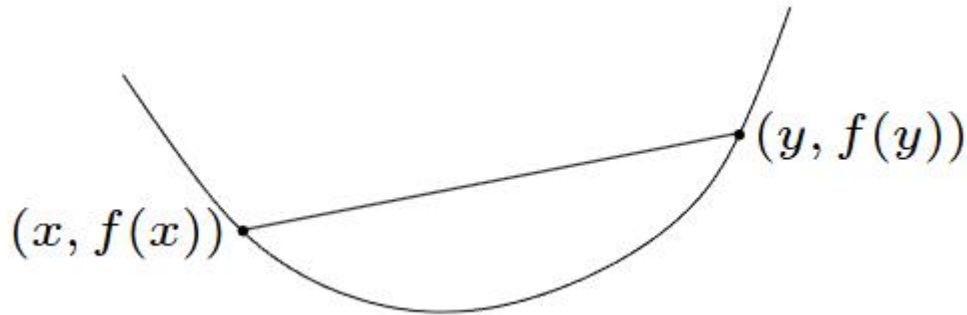
4. 单变量条件

1 凸函数的定义

■ **定义1 (凸函数)**: 对于 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 若 $\mathbf{dom} f$ 是一个凸集, 并且对于 $\forall x, y \in \mathbf{dom} f, 0 \leq \theta \leq 1$ 有

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

则称 f 是凸函数.



□ f 是严格凸的, 若 $\mathbf{dom} f$ 是凸集且

$$f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)$$

对于任意的 $x, y \in \mathbf{dom} f, x \neq y$ 且 $0 \leq \theta \leq 1$

□ f 是凹函数, 当且仅当 $-f$ 是凸函数

1 凸函数的定义

仿射函数是凸函数和凹函数；所有范数都是凸的

■ \mathbb{R}^n 中的例子：

仿射函数： $f(x) = a^T x + b$

范数： $\|x\|_p = (\sum_{i=1}^p |x_i|^p)^{\frac{1}{p}}, p \geq 1; \|x\|_\infty = \max_k |x_k|$ (满足三角不等式)

■ $\mathbb{R}^{m \times n}$ 中的例子：

仿射函数： $f(x) = \text{tr}(A^T X) + b = \sum_{i=1}^m \sum_{j=1}^n A_{ij} X_{ij} + b$

注： X 为矩阵， $\text{tr}(X)$ 为迹函数

1 凸函数的定义

□ 例 1：使用定义检查凸性

1. 范数(使用三角不等式和齐次性)

$$f(\theta x + (1 - \theta)y) \leq f(\theta x) + f((1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

2. “最大值函数”：对任意 $0 \leq \theta \leq 1$ ，函数 $f(x)$ 满足

$$\begin{aligned} f(\theta x + (1 - \theta)y) &= \max_i (\theta x_i + (1 - \theta)y_i) \\ &\leq \theta \max_i x_i + (1 - \theta) \max_i y_i \\ &= \theta f(x) + (1 - \theta)f(y) \end{aligned}$$

1 凸函数的定义

■ **定义2(扩展值延伸)**: f 的延伸函数 \tilde{f} 满足

$$\tilde{f}(x) = f(x), x \in \mathbf{dom} f, \tilde{f}(x) = \infty, x \notin \mathbf{dom} f$$

经常用来简化表达; 在 $\mathbb{R} \cup \{\infty\}$ 上, 也满足

$$0 \leq \theta \leq 1 \Rightarrow \tilde{f}(\theta x + (1 - \theta)y) \leq \theta \tilde{f}(x) + (1 - \theta)\tilde{f}(y)$$

■ **证明:**

显然当 $\theta = 0$ 或 1 时成立.

若 x 或 $y \notin \mathbf{dom} f$ 且 $\theta \neq 0, 1$, 则右侧为 $+\infty$, 因此不等式仍然成立.

1 凸函数的定义

定义3: 函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 的 α -下水平集

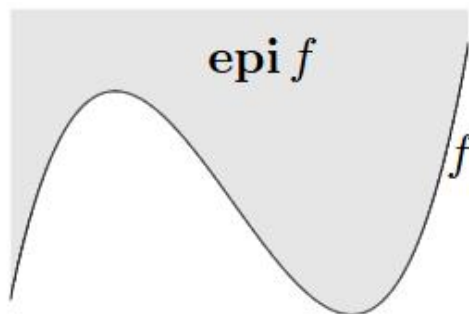
$$C_\alpha = \{x \in \mathbf{dom} f \mid f(x) \leq \alpha\}$$

凸函数的下水平集是凸集(逆命题不成立)

定义4: 函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 的上镜图

$$\mathbf{epi} f = \{(x, t) \in \mathbb{R}^{n+1} \mid x \in \mathbf{dom} f, f(x) \leq t\}$$

当且仅当 $\mathbf{epi} f$ 是凸集时, f 是凸函数(结合下图可直观理解)



1 凸函数的定义

□ 下水平集是凸集的证明:

如果 $x, y \in C_\alpha$, 则有 $f(x) \leq \alpha, f(y) \leq \alpha$

因此对于任意 $0 \leq \theta \leq 1$, 有

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y) \leq \alpha$$

即 $\theta x + (1 - \theta)y \in C_\alpha$, 得证

1 凸函数的定义

□ Jensen不等式:

如果 f 是凸函数, 那么对于任意 $0 \leq \theta \leq 1$ 有

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

□ 拓展: 如果 f 是凸函数, 那么

$$f(Ez) \leq Ef(z)$$

基本不等式是具有离散分布 $p(x) = \theta, p(y) = 1 - \theta$ 的特殊情况.

大 纲

1. 凸函数的定义

2. 一阶条件

3. 二阶条件

4. 单变量条件

2 一阶条件

■ 可微函数：

f 可微当且仅当 $\text{dom } f$ 是开集且梯度

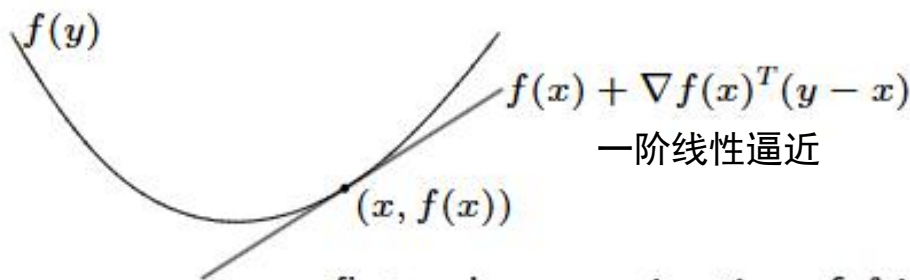
$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$$

存在于每个 $x \in \text{dom } f$.

一阶条件： 可微函数 f （定义域为凸）是凸函数当且仅当

$f(y) \geq f(x) + \nabla f(x)^T(y - x)$, 对所有 $x, y \in \text{dom } f$;

f 是严格凸函数当且仅当 $f(y) > f(x) + \nabla f(x)^T(y - x)$



first-order approximation of f is global underestimator

2 一阶条件

■ 证明:

\Leftarrow : 对所有 $0 < t \leq 1$, 若 $f(x + t(y - x)) \leq (1 - t)f(x) + tf(y)$, 则存在

$$\begin{aligned} f(y) &\geq \frac{1}{t}f(x + t(y - x)) - \frac{1 - t}{t}f(x) \\ &= f(x) + \frac{f(x + t(y - x)) - f(x)}{t} \end{aligned}$$

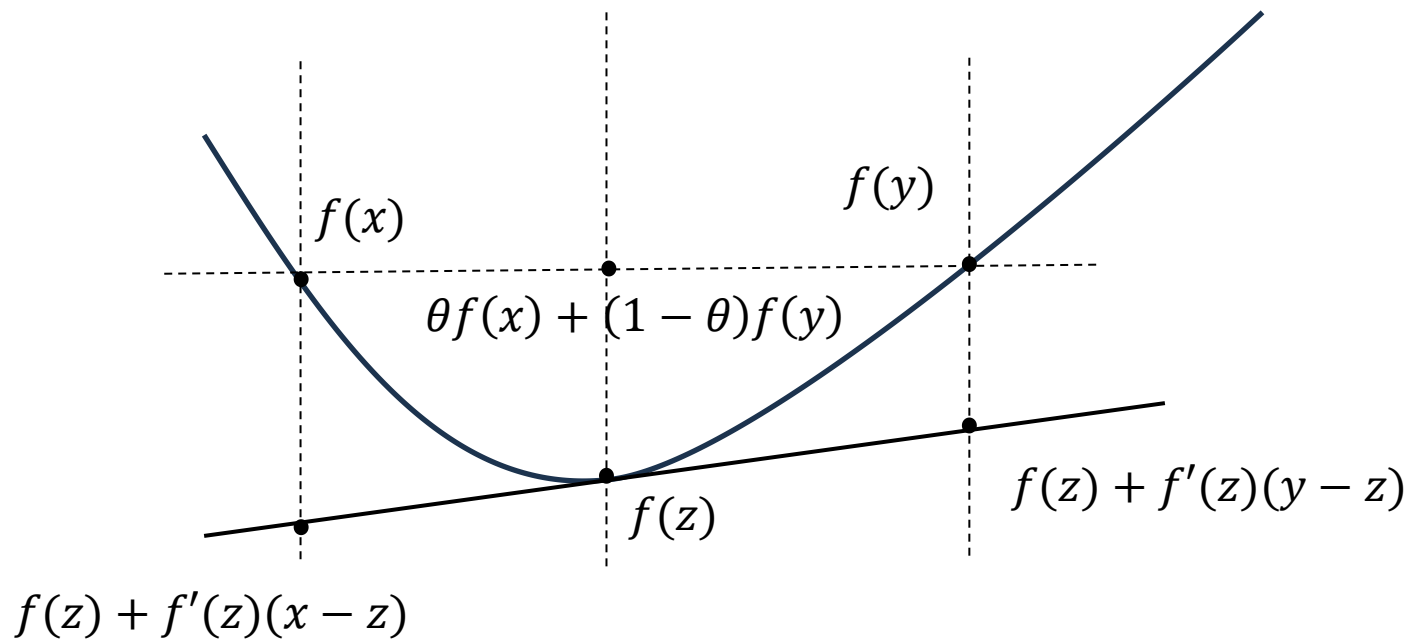
$$f(y) \geq f(x) + f'(x)(y - x) \quad (t \text{ 趋向 } 0)$$

\Rightarrow : 设 $z = \theta x + (1 - \theta)y$. 那么有 $f(x) \geq f(z) + f'(z)(x - z)$ 以及 $f(y) \geq f(z) + f'(z)(y - z)$. 因此

$$\begin{aligned} &\theta f(x) + (1 - \theta)f(y) \\ &\geq \theta f(z) + \theta f'(z)(x - z) + (1 - \theta)f(z) + (1 - \theta)f'(z)(y - z) \\ &= f(z) + f'(z)[\theta x + (1 - \theta)y - z] \\ &= f(z) \end{aligned}$$

2 一阶条件

$$z = \theta x + (1 - \theta)y$$



2 一阶条件

□ 例 2:

设函数 $f : \mathbb{R} \rightarrow \mathbb{R}$ 是凸函数。试证明: 利用凸函数的定义证明, 对于三变量 $x_1 < x_2 < x_3$, 函数 f 满足如下公式 (单调性质):

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2}$$

2 一阶条件

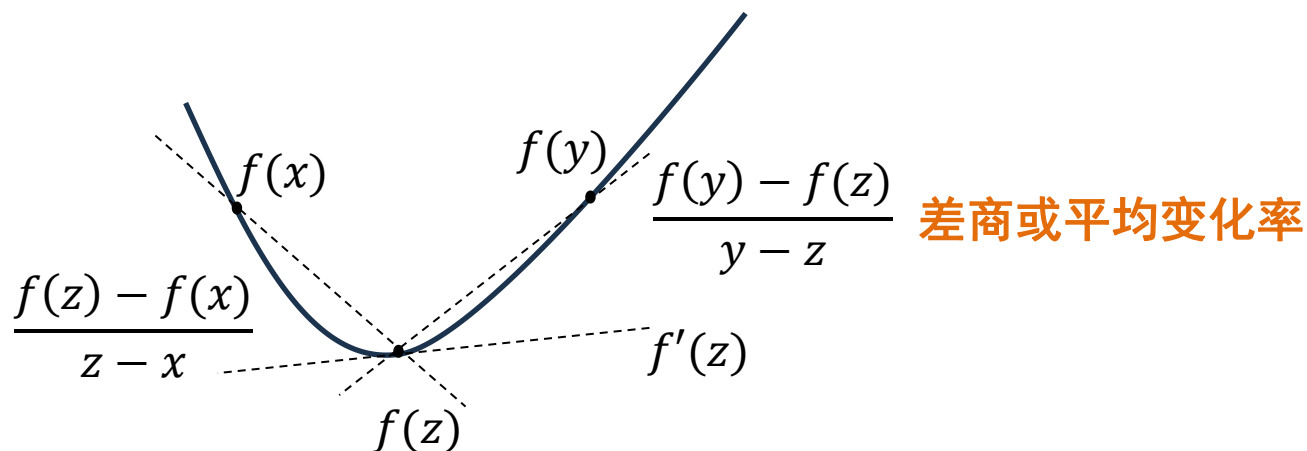
■ 例 2 的证明:

$$f(x_3) \geq f(x_2) + f'(x_2)(x_3 - x_2)$$

$$f(x_1) \geq f(x_2) + f'(x_2)(x_1 - x_2)$$

$$\frac{f(x_3) - f(x_2)}{x_3 - x_2} \geq f'(x_2), \quad \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq f'(x_2)$$

$$\frac{f(x_3) - f(x_2)}{x_3 - x_2} \geq f'(x_2) \geq \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$



大 纲

1. 凸函数的定义

2. 一阶条件

3. 二阶条件

4. 单变量条件

3 二阶条件

■ 二阶条件:

f 是二次可微的若 $\mathbf{dom} f$ 是开集且Hessian $\nabla^2 f(x) \in S^n$

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \dots, n$$

存在于每个 $x \in \mathbf{dom} f$.

二阶条件: 对于具有凸域的二次可微函数 f

□ f 是凸函数当且仅当

$$\nabla^2 f(x) \succeq 0, \text{ 对于所有 } x \in \mathbf{dom} f$$

□ 如果 $\nabla^2 f(x) \succ 0$, 对于所有 $x \in \mathbf{dom} f$, 那么 f 是严格凸的

3 二阶条件

■ 证明:

\Leftarrow : 观察 $f(y) \geq f(x) + f'(x)(y - x)$, 即 $f'(x) \leq \frac{f(y)-f(x)}{y-x}$

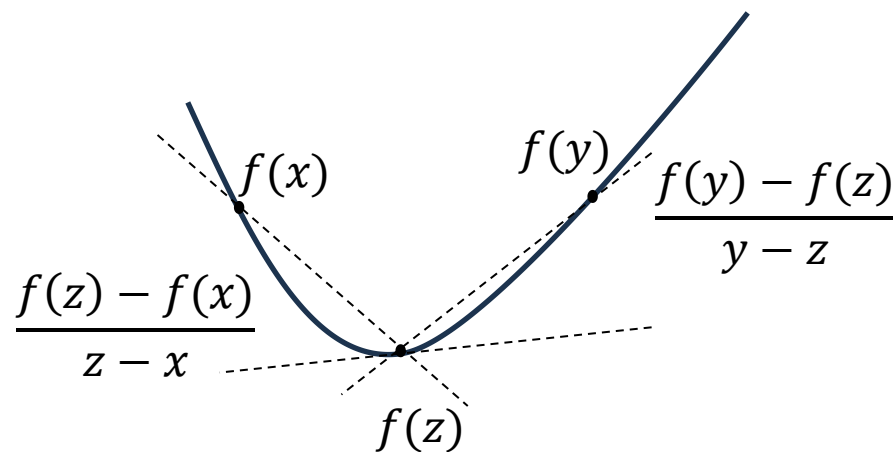
以及 $f(x) \geq f(y) + f'(y)(x - y)$, 即 $f'(y) \geq \frac{f(y)-f(x)}{y-x}$

因此 $f'(y) \geq \frac{f(y)-f(x)}{y-x} \geq f'(x)$

\Rightarrow : 存在 $z \in [x, y]$, 使得 $\frac{f(y)-f(x)}{y-x} = f'(z)$ (拉格朗日中值定理)

因此 $f'(x) \leq f'(z) = \frac{f(y)-f(x)}{y-x}$

所以 $f(y) \geq f(x) + (y - x)f'(x)$



3 二阶条件

□ 例 3:

设函数 $f: \mathbb{R} \rightarrow \mathbb{R}$ 是凸函数, 其定义域为 \mathbb{R} . 函数在 \mathbb{R} 上有上界. 证明该函数是常数.

■ 证明:

任取一点 x . 如果 $f'(x) \neq 0$, 不妨设 $f'(x) > 0$.

根据凸函数的性质, 我们有 $f'(z) \geq f'(x)$ 对所有的 $z \geq x$ 都成立. 假设 $f'(x) = \delta > 0$. 那么我们有

$$f(z) - f(x) \geq \delta(z - x)$$

因此当 $z \rightarrow \infty$, $f(z)$ 趋向于正无穷。此时, $f(x)$ 不存在上界。因此 $f'(x) > 0$ 不成立。同样, $f'(x) < 0$ 也不成立。因此, 只能有 $f'(x) = 0$ 。得证

• 思考: 如果 $x \in \mathbb{R}^n$ 呢?

3 二阶条件

■ 例 4（判断凹凸）：

□ 仿射函数： \mathbb{R} 上的 $f(x) = ax + b$, 对于任意 $a, b \in \mathbb{R}$,

$$f''(x) = 0$$

□ 指数函数： $f(x) = e^{ax}$, 对于任意 $a \in \mathbb{R}$

$$f''(x) = a^2 e^{ax}$$

□ 幂函数： \mathbb{R}_{++} 上的 $f(x) = x^\alpha$, 对于 $\alpha \geq 1$ 或者 $\alpha \leq 0$

$$f''(x) = \alpha(\alpha - 1)x^{\alpha-2} \quad [\alpha \text{ 与 } \alpha - 1 \text{ 同号}]$$

□ 幂函数： \mathbb{R}_{++} 上的 $f(x) = x^\alpha$, 对于 $0 \leq \alpha \leq 1$

$$f''(x) = \alpha(\alpha - 1)x^{\alpha-2} \quad [\alpha \text{ 与 } \alpha - 1 \text{ 异号}]$$

□ 负熵函数： \mathbb{R}_{++} 上的 $f(x) = -x \log x$

$$f''(x) = -\frac{1}{x}, \quad x > 0$$

3 二阶条件

□ 例 4 (续) :

□ **二次函数**: $f(x) = (1/2)x^T Px + q^T x + r, (P \in S^n, \text{若 } P \geq 0 \text{ 则是凸函数})$

$$\nabla f(x) = Px + q, \quad \nabla^2 f(x) = P$$

□ **最小二乘函数**: $f(x) = \|Ax - b\|_2^2, (\text{对任意 } A \text{ 都是凸函数})$

$$\nabla f(x) = 2A^T(Ax - b), \quad \nabla^2 f(x) = 2A^T A$$

R. K. 矩阵的转置点乘矩阵自身是一个半正定矩阵

□ **Quadratic-over-linear**: $f(x, y) = x^2/y, (\text{当 } y > 0 \text{ 时为凸函数})$

$$\nabla^2 f(x, y) = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \geq 0$$

□ **一个非凸的例子**: $f(x) = 1/x^2$ (除非限制在 \mathbb{R}_{++} 上)

预备知识：矩阵的导数

$$\frac{\partial(\mathbf{a}^T \mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}^T \mathbf{a})}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial(\mathbf{a}^T \mathbf{X} \mathbf{b})}{\partial \mathbf{X}} = \mathbf{a} \mathbf{b}^T = \mathbf{a} \otimes \mathbf{b} \in \mathbb{R}^{n \times n}$$

$$\frac{\partial(\mathbf{a}^T \mathbf{X}^T \mathbf{b})}{\partial \mathbf{X}} = \mathbf{b} \mathbf{a}^T = \mathbf{b} \otimes \mathbf{a} \in \mathbb{R}^{n \times n}$$

$$\frac{\partial(\mathbf{x}^T \mathbf{A} \mathbf{x})}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}$$

$$\frac{\partial[(\mathbf{X} \mathbf{b} + \mathbf{c})^T \mathbf{A} (\mathbf{X} \mathbf{b} + \mathbf{c})]}{\partial \mathbf{X}} = (\mathbf{A} + \mathbf{A}^T)(\mathbf{X} \mathbf{b} + \mathbf{c}) \mathbf{b}^T$$

$$\frac{\partial(\mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{b})}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{a} \otimes \mathbf{b} + \mathbf{b} \otimes \mathbf{a})$$

注： \otimes ：克罗内克积

3 二阶条件

□ 例 4 (续) :

□ **Log-sum-exp**: $f(x) = \log \sum_{k=1}^n \exp(x_k)$ 是凸函数

$$\nabla^2 f(x) = \frac{1}{1^T z} \mathbf{diag}(z) - \frac{1}{(1^T z)^2} z^T z, (z_k = \exp(x_k))$$

要证明 $\nabla^2 f(x) \succeq 0$, 我们必须验证对于所有的 v , $v^T \nabla^2 f(x) v \geq 0$:

$$v^T \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k z_k v_k)^2}{(\sum_k z_k)^2} \geq 0$$

因为 $(\sum_k z_k v_k^2)(\sum_k z_k) \geq (\sum_k z_k v_k)^2$ (由柯西不等式得), $a_k = \sqrt{z_k} v_k$, $b_k = \sqrt{z_k}$

□ **几何平均数函数**: $f(x) = (\prod_{k=1}^n x_k)^{1/n}$ 在 \mathbb{R}_{++}^n 上是凹函数 (证明过程与 Log-sum-exp 类似)

3 二阶条件

□ 例 4 (续) :

验证以下函数是否为凸函数 (给出分析过程)

$$(1) f(x, y) = x^2 y^2 + \frac{x}{y}, \quad x > 0, y > 0$$

$$(2) f(x, y) = \log(e^x + e^y) - \log x, \quad x > 0$$

$$(3) f(x, y) = \exp\{x^2 + e^{-y}\}, \quad x > 0, y > 0$$

$$(4) f(x, y) = xy \log xy, \quad x > 0, y > 0$$

$$(5) f(x, y) = -\log(cx + dy), \quad c, d \in \mathbb{R}$$

3 二阶条件

■ 例 4 的证明:

$$(1) f(x, y) = x^2 y^2 + \frac{x}{y}, \quad x > 0, y > 0$$

计算Hessian 矩阵:

$$H = \begin{bmatrix} 2y^2 & 4xy - 1/y^2 \\ 4xy - 1/y^2 & 2x^2 + 2x/y^3 \end{bmatrix}$$

取点 (1, 1), Hessian 行列式为 $\det(H) = (2)(4) - (4 - 1)^2 = 8 - 9 = -1 < 0$, 故矩阵不定

所以函数非凸

注: $\det(H)$: 矩阵 H 的行列式

3 二阶条件

■ 例 4的证明:

$$(2) f(x, y) = \log(e^x + e^y) - \log x, \quad x > 0$$

计算Hessian 矩阵:

$$H = \begin{bmatrix} e^x e^y / (e^x + e^y)^2 + 1/x^2 & -e^x e^y / (e^x + e^y)^2 \\ -e^x e^y / (e^x + e^y)^2 & e^x e^y / (e^x + e^y)^2 \end{bmatrix}$$

$$\text{对任意向量 } v = (v_1, v_2), \quad v^T H v = \frac{1}{x^2} v_1^2 + \frac{e^x e^y}{(e^x + e^y)^2} (v_1 - v_2)^2 \geq 0$$

所以函数是凸函数

3 二阶条件

■ 例 4的证明:

$$(3)f(x, y) = \exp \{x^2 + e^{-y}\}, x > 0, y > 0$$

令 $g(x, y) = x^2 + e^{-y}$, 则 $f = e^g$. 计算Hessian 矩阵:

$$H = e^g \begin{bmatrix} 4x^2 + 2 & -2xe^y \\ -2xe^y & e^{-2y} + e^{-y} \end{bmatrix}$$

主对角线元素 $4x^2 + 2, e^{-2y} + e^{-y} > 0$, 且 $\det(H) > 0$, 故矩阵正定

所以函数是凸函数

3 二阶条件

■ 例 4的证明:

$$(4) f(x, y) = xy \log xy, \quad x > 0, y > 0$$

计算Hessian 矩阵:

$$H = \begin{bmatrix} y/x & \log(xy) + 2 \\ \log(xy) + 2 & x/y \end{bmatrix}$$

Hessian 行列式为 $\det(H) = 1 - (\log(xy) + 2)^2$

当 $xy = e^{-1}$ 时, $\det(H) = 1 - (1)^2 = 0$

当 $xy = e^1$ 时, $\det(H) = 1 - (1 + 2)^2 = -8$

所以函数非凸

3 二阶条件

■ 例 4的证明:

$$(5) f(x, y) = -\log(cx + dy), \quad c, d \in \mathbb{R}$$

函数 $\log(cx + dy)$ 在定义域 $cx + dy > 0$ 上为凹函数
(Hessian 矩阵为半负定)

$$H = \begin{bmatrix} -\frac{d^2}{(cx + dy)^2} & -\frac{cd}{(cx + dy)^2} \\ -\frac{cd}{(cx + dy)^2} & -\frac{c^2}{(cx + dy)^2} \end{bmatrix}$$

其负数 $-\log(cx + dy)$ 为凸函数

大 纲

1. 凸函数的定义

2. 一阶条件

3. 二阶条件

4. 单变量条件

4 单变量条件

■ 判断凸性的其他方法：将函数限制在一条直线上

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是凸函数当且仅当对于 $\mathbf{dom} f$ 中的任意 x 和 \mathbb{R}^n 中的任意 v , 函数 $g: \mathbb{R} \rightarrow \mathbb{R}$

$$g(t) = f(x + tv), \mathbf{dom} g = \{t : x + tv \in \mathbf{dom} f\}$$

(关于 t) 是凸函数.

可以通过检查单变量函数的凸性来验证 f 的凸性.

■ 证明:

\Rightarrow : 取任意两点 x_1 和 x_2 , 它们必须位于一条直线上, 该直线的形式为 $x + tv$, \dots

\Leftarrow : $x + tv$ 与 $\mathbf{dom} f$ 相交, 形成一个新的凸集. 很明显, 新凸集中的点满足凸性条件

4 单变量条件

□ 例 7: $f: S^n \rightarrow \mathbb{R}$, 其中 $f(X) = \log \det X$, $\mathbf{dom} f = S_{++}^n$

$$g(t) = \log \det (X + tV) = \log \det \left(X^{\frac{1}{2}} \left(I + t(X^{-\frac{1}{2}})^T V X^{-\frac{1}{2}} \right) X^{\frac{1}{2}} \right)$$

$$= \log \det X + \log \det (I + tX^{-\frac{1}{2}} V X^{-\frac{1}{2}})$$

$$= \log \det X + \sum_{i=1}^n \log (1 + t\lambda_i)$$

其中 λ_i 是 $(X^{-\frac{1}{2}})^T V X^{-\frac{1}{2}}$ 的特征值. 因此, 下式成立

$$g'(t) = \sum_{i=1}^n \frac{\lambda_i}{1 + t\lambda_i}, g''(t) = - \sum_{i=1}^n \frac{\lambda_i^2}{(1 + t\lambda_i)^2}$$

g 在 t 上是凹函数 (对于任意选择的 $X \succ 0$ 和 V); 因此 f 是凹函数.

谢谢！