第二章 基础知识

修贤超

https://xianchaoxiu.github.io

目录

凸函数的定义

- **定义 2.16** 设 $f: \mathbb{R}^n \to \mathbb{R}$ 为适当函数，如果 $\mathrm{dom}\ f$ 是凸集，且

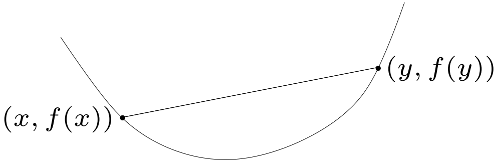
$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$$

  对所有 $x, y \in \mathrm{dom}\ f$, $0 \le \theta \le 1$ 都成立，则称 f 是**凸函数**

- 若对所有 $x, y \in \mathrm{dom}\ f$，$x \ne y$，$0 < \theta < 1$，有

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

  则称 f 是**严格凸函数**



一元凸函数的例子

- 仿射函数　对任意 $a, b \in \mathbb{R}$，$ax + b$ 是 $\mathbb{R}$ 上的凸 (凹)函数
- 指数函数　对任意 $a \in \mathbb{R}$，e^{ax} 是 $\mathbb{R}$ 上的凸函数
- 绝对值的幂　对 $p \geq 1$，$|x|^p$ 是 $\mathbb{R}$ 上的凸函数
- 幂函数　对 $\alpha \geq 1$ 或 $\alpha \leq 0$，x^α 是 $\mathbb{R}_{++}$ 上的凸函数
- 幂函数　对 $0 \leq \alpha \leq 1$，x^α 是 $\mathbb{R}_{++}$ 上的凹函数
- 对数函数　$\log x$ 是 $\mathbb{R}_{++}$ 上的凹函数
- Sigmoid 函数、Heaviside 函数、ReLU 函数 . . .

多元凸函数的例子

- 所有的仿射函数既是凸函数，又是凹函数

$$f(x) = a^\top x + b$$

$$f(X) = \mathrm{Tr}(A^\top X) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b$$

- 所有的范数都是凸函数

$$f(x) = \|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p} \ (p \geq 1)$$

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^\top X))^{1/2}$$

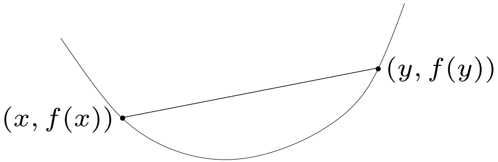
强凸函数

- 定义 2.17 若存在常数 $m > 0$，使得

$$g(x) = f(x) - \frac{m}{2}\|x\|^2$$

为凸函数，则称 $f(x)$ 为强凸函数

- 为了方便也称 $f(x)$ 为 m-强凸函数
- 命题 2.3 设 f 为强凸函数且存在最小值，则 f 的最小值点唯一



凸函数判定定理

- **定理 2.6** $f(x)$ **是凸函数当且仅当对每个** $x \in \mathrm{dom}\ f$，$v \in \mathbb{R}^n$，**函数** $g: \mathbb{R} \to \mathbb{R}$ **是关于** t **的凸函数**

$$g(t) = f(x+tv), \quad \mathrm{dom}\ g = \{t \mid x+tv \in \mathrm{dom}\ f\}$$

- **例 2.4** $f(X) = -\log\det X$ **是凸函数，其中** $\mathrm{dom}\ f = \mathcal{S}^n_{++}$

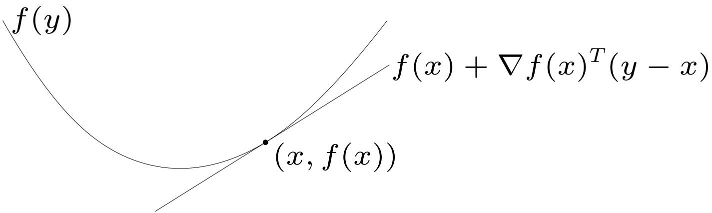
证明 **任取** $X \succ 0$ **以及方向** $V \in \mathcal{S}^n$，**将** f **限制在直线** $X+tV$ **上，则**

$$\begin{aligned} g(t) &= -\log\det(X+tV) \\ &= -\log\det X - \log\det(I + tX^{-1/2}VX^{-1/2}) \\ &= -\log\det X - \sum_{i=1}^{n} \log(1+t\lambda_i) \end{aligned}$$

一阶条件

- **定理 2.7** 对于定义在凸集上的可微函数 f，则 f 是凸函数当且仅当

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \quad \forall x, y \in \mathrm{dom}\ f$$



- **定理 2.8** 设 f 为可微函数，则 f 为凸函数当且仅当 $\mathrm{dom}\ f$ 为凸集且 ∇f 为单调映射

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq 0, \quad \forall\ x, y \in \mathrm{dom}\ f$$

二阶条件

- **定理 2.9** 设 f 为定义在凸集上的二阶连续可微函数，则 f 是凸函数当且仅当

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in \mathrm{dom}\ f$$

- 如果 $\nabla^2 f(x) \succ 0\ \forall x \in \mathrm{dom}\ f$，则 f 是**严格凸函数**
- **例 2.5** 最小二乘函数 $f(x) = \|Ax - b\|_2^2$

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

对任意 A，函数 f 都是凸函数

上方图

- **定理 2.10** 函数 $f(x)$ 为凸函数当且仅当其上方图 epi 是凸集

  证明 (必要性) 若 f 为凸函数，则对任意 $(x_1, y_1), (x_2, y_2) \in \text{epi } f, t \in [0,1]$ 有

$$ty_1 + (1-t)y_2 \geq tf(x_1) + (1-t)f(x_2) \geq f(tx_1 + (1-t)x_2)$$

  故 $(tx_1 + (1-t)x_2, ty_1 + (1-t)y_2) \in \text{epi } f, t \in [0,1]$

  (充分性) 若 $\text{epi } f$ 是凸集，则对任意 $x_1, x_2 \in \text{dom } f, t \in [0,1]$ 有

$$(tx_1 + (1-t)x_2, tf(x_1) + (1-t)f(x_2)) \in \text{epi } f$$

$$\Downarrow$$

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

凸函数的判断方法

- 用定义验证（通常将函数限制在一条直线上）
- 利用一阶条件、二阶条件
- 直接研究 f 的上方图 $\mathrm{epi}\ f$
- 说明 f 可由简单的凸函数通过保凸运算得到
  - 非负加权和
  - 与仿射函数复合
  - 逐点取最大值
  - 与标量向量函数复合

非负加权和与仿射函数的复合

- **定理 2.11** (1) 若 f 是凸函数，则 αf 是凸函数，其中 $\alpha \geq 0$
- **定理 2.11** (2) 若 f_1, f_2 是凸函数，则 $f_1 + f_2$ 是凸函数
- **定理 2.11** (3) 若 f 是凸函数，则 $f(Ax+b)$ 是凸函数

例子

- 线性不等式的对数障碍函数

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^\top x), \quad \operatorname{dom} f = \{x \mid a_i^\top x < b_i, i = 1, ..., m\}$$

- 仿射函数的（任意）范数 $f(x) = \|Ax + b\|$

逐点取最大值

- 定理 2.11 (4) 若 $f_1, \cdots, f_m$ 是凸函数，则

$$f(x) = \max\{f_1(x), \cdots, f_m(x)\}$$

是凸函数

例子

- 分段线性函数

$$f(x) = \max_{i=1,\cdots,m} (a_i^\top x + b_i)$$

- $x \in \mathbb{R}^n$ 的前 r 个最大分量之和

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

$$\Updownarrow$$

$$f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \le i_1 < i_2 < \cdots < i_r \le n\}$$

逐点取上界

- 定理 2.11 (5) 若对每个 $y \in \mathcal{A}$，$f(x,y)$ 是关于 x 的凸函数，则

$$g(x) = \sup_{y \in \mathcal{A}} f(x,y)$$

是凸函数

例子

- 集合 C 点到给定点 x 的最远距离

$$f(x) = \sup_{y \in C} \|x - y\|$$

- 对称矩阵 $X \in \mathcal{S}^n$ 的最大特征值

$$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^\top X y$$

与函数的复合

- **定理 2.11** (6) 给定函数 $g: \mathbb{R}^n \to \mathbb{R}$ 和 $h: \mathbb{R} \to \mathbb{R}$,

$$f(x) = h(g(x))$$

若 g 是凸函数, h 是凸函数且单调不减；g 是凹函数, h 是凸函数且单调不增，那么 f 是凸函数

例子

- 如果 g 是凸函数，则 $\exp g(x)$ 是凸函数
- 如果 g 是正值凹函数，则 $1/g(x)$ 是凸函数

取下确界

- **定理 2.11** (7) 若 $f(x,y)$ 关于 (x,y) 整体是凸函数，$\mathcal{C}$ 是凸集，则

$$g(x) = \inf_{y \in \mathcal{C}} f(x,y)$$

是凸函数

例子

- 考虑函数 $f(x,y) = x^\top Ax + 2x^\top By + y^\top Cy$，海瑟矩阵满足

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \succeq 0, \quad C \succ 0$$

则 $f(x,y)$ 为凸函数. 对 y 求最小值得

$$g(x) = \inf_y f(x,y) = x^\top (A - BC^{-1}B^\top)x$$

- 点 x 到凸集 $\mathcal{S}$ 的距离 $\mathrm{dist}(x,\mathcal{S}) = \inf_{y \in \mathcal{S}} \|x - y\|$ 是凸函数

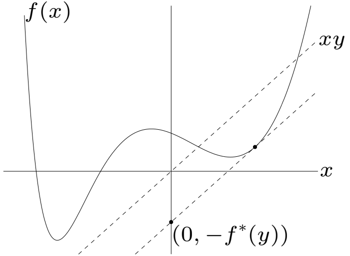
目录

共轭函数

- **定义 2.19** 适当函数 f 的**共轭函数**定义为

$$f^*(y) = \sup_{x \in \mathrm{dom}\ f} (y^\top x - f(x))$$

- 无论 f 是否是凸函数，f^* 恒为凸函数



- **命题 2.5** Fenchel 不等式 $f(x) + f^*(y) \geq x^\top y \quad \forall x, y$

例 2.6

- 考虑二次函数

$$f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$$

  - 强凸情形（$A \succ 0$）知

$$f^*(y) = \frac{1}{2}(y-b)^\top A^{-1}(y-b) - c$$

  - 一般凸情形（$A \succeq 0$）知

$$f^*(y) = \frac{1}{2}(y-b)^\top A^\dagger(y-b) - c, \quad \mathrm{dom}\ f^* = \mathcal{R}(A) + b$$

    这里 $\mathcal{R}(A)$ 为 A 的像空间

例 2.7

- 给定凸集 $\mathcal{C}$, 示性函数为

$$I_C(x) = \begin{cases} 0, & x \in \mathcal{C} \\ +\infty, & x \notin \mathcal{C} \end{cases}$$

共轭函数

$$\begin{aligned} I^*(y) &= \sup_x \ \{y^\top x - I_{\mathcal{C}}(x)\} \\ &= \sup_{x \in \mathcal{C}} \ y^\top x \end{aligned}$$

- $I^*(y)$ 称为凸集 $\mathcal{C}$ 的**支撑函数**

二次共轭函数

- **定义 2.20** 任一函数 f 的二次共轭函数定义为

$$f^{**}(x) = \sup_{y \in \mathrm{dom}\ f^*} (x^\top y - f^*(y))$$

- **定理 2.12** 若 f 为闭凸函数，则

$$f^{**}(x) = f(x)$$

- **性质** 若 f 为闭凸函数，则

$$y \in \partial f(x) \quad \Leftrightarrow \quad x \in \partial f^*(y) \quad \Leftrightarrow \quad x^\top y = f(x) + f^*(y)$$

目录

次梯度

- 回顾可微凸函数 f 的一阶条件

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$

- 定义 2.21 设 f 为适当凸函数，x 为 $\mathrm{dom}\ f$ 中的一点. 若向量 $g \in \mathbb{R}^n$ 满足

$$f(y) \geq f(x) + g^\top (y - x), \quad \forall y \in \mathrm{dom}\ f$$

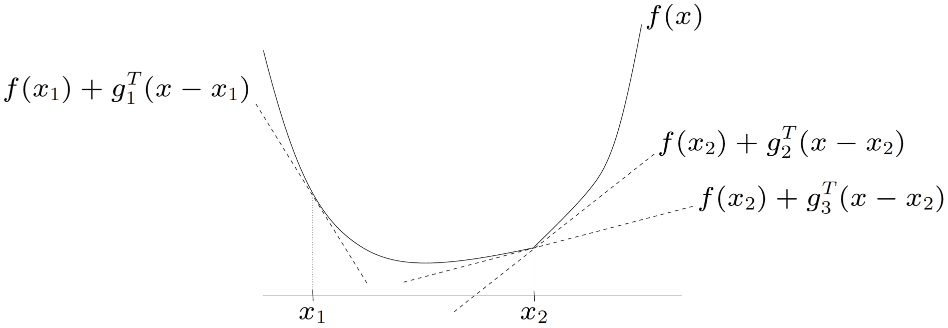
则称 g 为函数 f 在点 x 处的一个次梯度. 进一步，称集合

$$\partial f(x) = \{g \mid g \in \mathbb{R}^n, f(y) \geq f(x) + g^\top (y - x), \forall y \in \mathrm{dom}\ f\}$$

为 f 在点 x 处的次微分

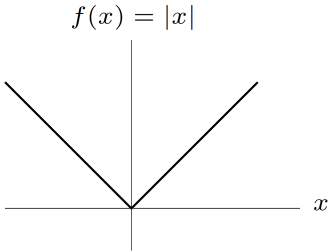
次梯度

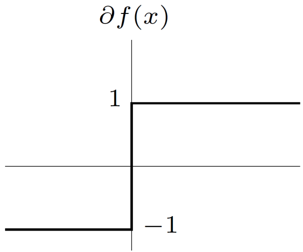
- g_1 是点 x_1 处的次梯度
- g_2, g_3 是点 x_2 处的次梯度



例子

- 绝对值函数 $f(x) = |x|$





- 欧几里得范数 $f(x) = \|x\|_2$

$$\text{若 } x \neq 0,\ \partial f(x) = \frac{1}{\|x\|_2}x, \quad \text{若 } x = 0,\ \partial f(x) = \{g \mid \|g\|_2 \leq 1\}$$

次梯度的性质

- **定理 2.13** 设 f 是凸函数，则 $\partial f(x)$ 有如下性质
  - 对任何 $x \in \mathrm{dom}\ f$，$\partial f(x)$ 是一个闭凸集（可能为空集）
  - 如果 $x \in \mathrm{int\ dom}\ f$，则 $\partial f(x)$ 非空有界集
- **命题 2.6** 设 凸函数 $f(x)$ 在 $x_0 \in \mathrm{int\ dom}\ f$ 处可微，则

$$\partial f(x_0) = \{\nabla f(x_0)\}$$

- **定理 2.14** 设 $f: \mathbb{R}^n \to \mathbb{R}$ 为凸函数，$x, y \in \mathrm{dom}\ f$，则

$$(u - v)^\top (x - y) \geq 0$$

其中 $u \in \partial f(x)$，$v \in \partial f(y)$

次梯度的计算规则

- **非负线性组合** 若凸函数 f_1, f_2 满足 $\text{int dom } f_1 \cap \text{dom } f_2 \neq \emptyset$ 且 $x \in \text{dom } f_1 \cap \text{dom } f_2$。若

$$f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x), \quad \alpha_1, \alpha_2 \geq 0$$

则 $f(x)$ 的次微分

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$$

- **线性变量替换** 设 h 为适当凸函数，f 满足

$$f(x) = h(Ax + b)$$

若存在 $x^\sharp \in \mathbb{R}^m$，使得 $Ax^\sharp + b \in \text{int dom } h$，则

$$\partial f(x) = A^\top \partial h(Ax + b), \quad \forall\, x \in \text{int dom } f$$

两个函数之和的次梯度

- **定理 2.15** 设 $f_1, f_2 : \mathbb{R}^n \to (-\infty, +\infty]$ 是凸函数，则对任意的 $x_0 \in \mathbb{R}^n$ 有

$$\partial f_1(x_0) + \partial f_2(x_0) \subseteq \partial(f_1 + f_2)(x_0)$$

进一步，若 $\text{int dom } f_1 \cap \text{dom } f_2 \neq \emptyset$，则对任意的 $x_0 \in \mathbb{R}^n$ 有

$$\partial(f_1 + f_2)(x_0) = \partial f_1(x_0) + \partial f_2(x_0).$$

- Moreau-Rockafellar 定理

函数族的上确界

- 定理 2.16 **设** $f_1, f_2, \cdots, f_m: \mathbb{R}^n \to (-\infty, +\infty]$ **均为凸函数，令**

$$f(x) = \max\{f_1(x), f_2(x), \cdots, f_m(x)\}, \quad \forall x \in \mathbb{R}^n$$

对 $x_0 \in \bigcap_{i=1}^{m} \text{int dom } f_i$**，定义** $I(x_0) = \{i \mid f_i(x_0) = f(x_0)\}$**，则**

$$\partial f(x_0) = \text{conv} \bigcup_{i \in I(x_0)} \partial f_i(x_0)$$

  - $I(x_0)$ **表示点** x_0 **处"有效"函数的指标**
  - $\partial f(x_0)$ **是点** x_0 **处"有效"函数的次微分并集的凸包**

- **如果** f_i **可微**, $\partial f(x_0) = \text{conv}\{\nabla f_i(x_0) \mid i \in I(x_0)\}$

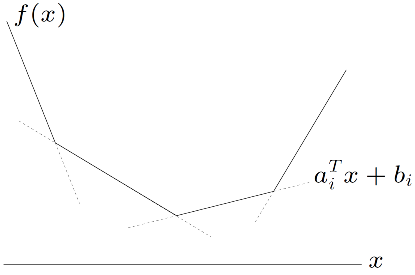
例 2.11

- **分段线性函数**

$$f(x) = \max_{i=1,2,\cdots,m} \{a_i^\top x + b_i\}$$

点 x 处的次微分是一个多面体

$$\partial f(x) = \mathrm{conv}\{a_i \mid i \in I(x)\}, \quad I(x) = \{i \mid a_i^\top x + b_i = f(x)\}$$



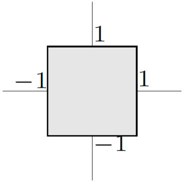
例 2.12

- ℓ_1-范数

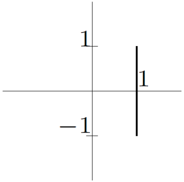
$$f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^n} s^\top x$$

点 x 处的次微分是

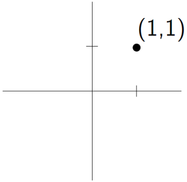
$$\partial f(x) = J_1 \times \cdots \times J_n, \quad J_k = \begin{cases} [-1,1], & x_k = 0 \\ \{1\}, & x_k > 0 \\ \{-1\}, & x_k < 0 \end{cases}$$



$\partial f(0,0) = [-1,1] \times [-1,1]$



$\partial f(1,0) = \{1\} \times [-1,1]$



$\partial f(1,1) = \{(1,1)\}$

Q&A

Thank you!

感谢您的聆听和反馈