

BCD & ADMM 算法收敛理论

Lecture 2: 凸分析基础

罗自炎

北京交通大学数统学院

E-mail: zyluo@bjtu.edu.cn

参考资料

- 教材与参考文献：
 - 最优化：建模、算法与理论
 - Bolte et al., MP 2014
 - Fazel et al., SIMAX 2013
 - Rockafellar & Wets, Variational Analysis
- 致谢： 北京大学文再文教授; 清华大学张立平教授

目 录

- 向量范数和矩阵范数
- 凸集
- 凸函数
- 共轭函数
- 次梯度

向量范数

- 最常用的向量范数 ℓ_p 范数($p \geq 1$): 令 $v \in \mathbb{R}^n$,

$$\|v\|_p = \left(\sum_{i=1}^n |v_i|^p \right)^{\frac{1}{p}}.$$

- ℓ_∞ 范数: $\|v\|_\infty = \max_{1 \leq j \leq n} |v_{(j)}|$.
- 由正定矩阵 A 诱导的向量范数: $\|v\|_A = \sqrt{v^T A v}$.
- Cauchy-Schwarz不等式: 设 $a, b \in \mathbb{R}^n$, 则 $|a^T b| \leq \|a\|_2 \|b\|_2$, 且等号成立的条件是 a 与 b 线性相关.

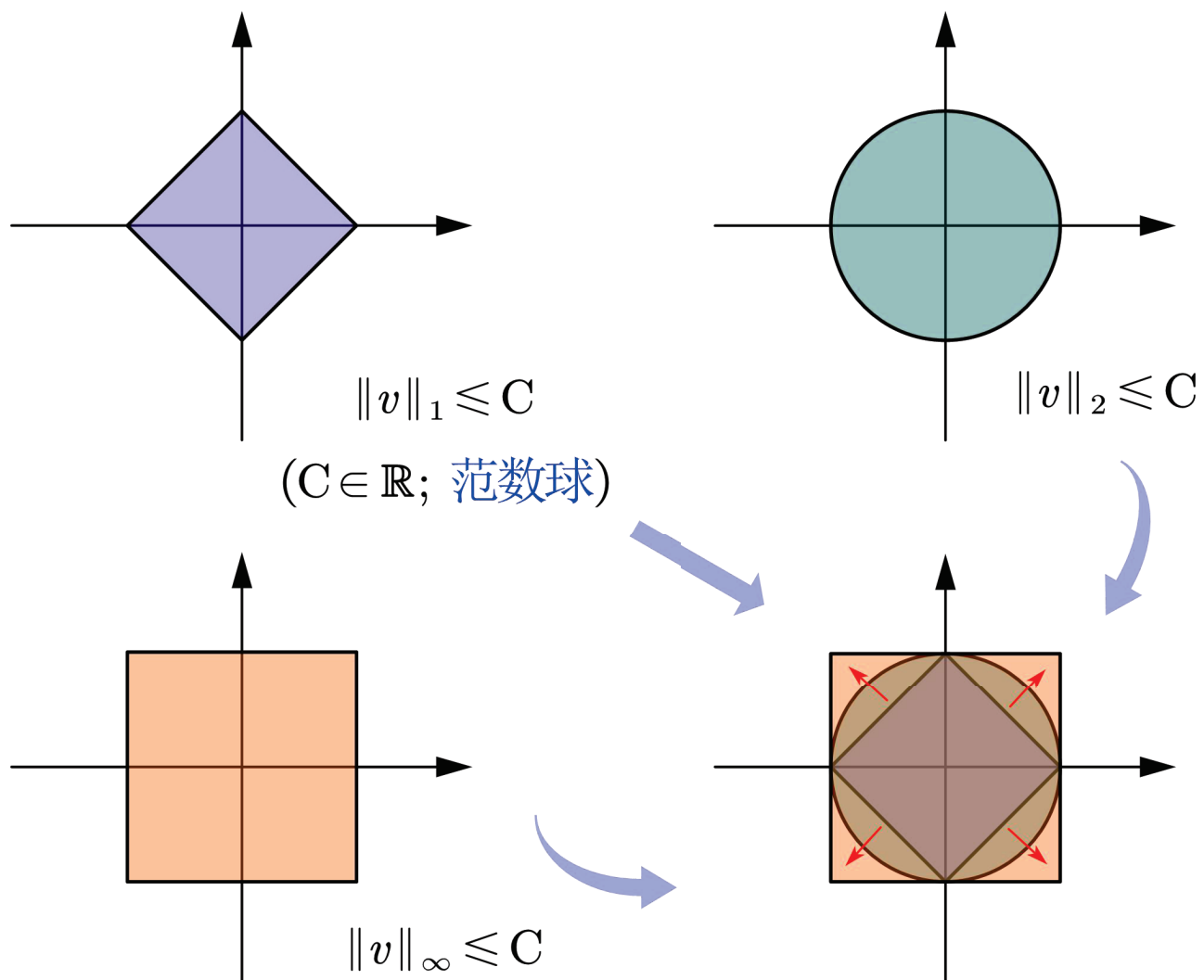


Figure 1: l_1 范数, l_2 范数和 l_∞ 范数

矩阵范数、核范数

Definition 1. 如果函数 $\|\cdot\| : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^+$ 满足:

- 正定性: 对 $\forall A \in \mathbb{R}^{m \times n}$, 有 $\|A\| \geq 0$ 且 $\|A\| = 0 \Leftrightarrow A = 0_{m \times n}$;
- 正齐次性: 对任意 $A \in \mathbb{R}^{m \times n}$ 和 $\alpha \in \mathbb{R}$, 有 $\|\alpha A\| = |\alpha| \|A\|$;
- 三角不等式: 对于任意 $A, B \in \mathbb{R}^{m \times n}$, 有 $\|A + B\| \leq \|A\| + \|B\|$.

则称 $\|\cdot\|$ 是定义在向量空间 $\mathbb{R}^{m \times n}$ 上的矩阵范数.

矩阵 ℓ_p 范数

类似于向量 ℓ_p 范数, 矩阵的 ℓ_p 范数: ℓ_1 -范数, F-范数. 令 $A \in \mathbb{R}^{m \times n}$,

- ℓ_1 -范数: $\|A\|_1 = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|.$

- Frobenius范数(F-范数): $\|A\|_F = \sqrt{\text{Tr}(AA^T)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}.$

- F-范数具有正交不变性: 对于任意的正交矩阵 $U \in \mathbb{R}^{m \times m}$ 和 $V \in \mathbb{R}^{n \times n}$, 有

$$\|UAV\|_F^2 = \|A\|_F^2.$$

- 矩阵迹的性质: $\text{Tr}(AB) = \text{Tr}(BA).$

- 矩阵的内积: 设 $A, B \in \mathbb{R}^{m \times n}$, 矩阵 A, B 的内积定义为

$$\langle A, B \rangle = \text{Tr} (AB^T) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ij}.$$

- F-范数的Cauchy不等式: 设 $A, B \in \mathbb{R}^{m \times n}$, 则

$$|\langle A, B \rangle| \leq \|A\|_F \|B\|_F,$$

等号成立当且仅当 A 和 B 线性相关.

Definition 2. 给定矩阵 $A \in \mathbb{R}^{m \times n}$, 其核范数定义为

$$\|A\|_* = \sum_{i=1}^r \sigma_i,$$

其中 $r = \text{rank} (A)$, $\sigma_i (i = 1, \dots, r)$ 为 A 的所有非零奇异值.

Question: 核范数是矩阵范数吗?

矩阵的算子范数

Definition 3. 给定矩阵 $A \in \mathbb{R}^{m \times n}$, \mathbb{R}^m 中的向量范数 $\|\cdot\|_{(m)}$ 和 \mathbb{R}^n 中的向量范数 $\|\cdot\|_{(n)}$, 其诱导的矩阵范数为

$$\|A\|_{(m,n)} = \max_{x \in \mathbb{R}^n, \|x\|_{(n)}=1} \|Ax\|_{(m)}.$$

将 $\|\cdot\|_{(m)}$ 和 $\|\cdot\|_{(n)}$ 取向量的 ℓ_p 范数, 可诱导下面的矩阵的 p 范数.

- A 的 1-范数: $\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|.$
- A 的谱范数: $\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_{\max}(A^T A)}.$
- A 的 ∞ 范数: $\|A\|_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|.$

矩阵算子范数的性质

矩阵算子范数的相容性: $\|Ax\|_{(m)} \leq \|A\|_{(m,n)} \|x\|_{(n)}.$

具体地说,

$$\|Ax\|_2 \leq \|A\|_2 \|x\|_2.$$

Theorem 1. 设 $A \in \mathbb{R}^{n \times n}$, 则对 A 的谱范数 $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)},$

- $\|A\|_2^2 = \|A^T\|_2^2 = \|A^T A\|_2 = \|AA^T\|_2.$
- 对于任意 n 阶正交矩阵 C, D 有

$$\|CA\|_2 = \|AD\|_2 = \|CAD\|_2 = \|A\|_2.$$

凸集

- 仿射集: $x_1, x_2 \in \mathcal{C} \Rightarrow \theta x_1 + (1 - \theta)x_2 \in \mathcal{C}, \forall \theta \in \mathbb{R}$.
- 凸集: $x_1, x_2 \in \mathcal{C} \Rightarrow \theta x_1 + (1 - \theta)x_2 \in \mathcal{C}, \forall 0 \leq \theta \leq 1$.
- 仿射集当然都是凸集.
- 若 \mathcal{S} 是凸集, 则 $k\mathcal{S} = \{ks | k \in \mathbb{R}, s \in \mathcal{S}\}$ 是凸集.
- 若 \mathcal{S} 和 \mathcal{T} 均是凸集, 则 $\mathcal{S} + \mathcal{T} = \{s + t | s \in \mathcal{S}, t \in \mathcal{T}\}$ 是凸集.
- 若 \mathcal{S} 和 \mathcal{T} 均是凸集, 则 $\mathcal{S} \cap \mathcal{T}$ 是凸集.
- 设 \mathcal{S} 是凸集, 则 $\text{int}\mathcal{S}, \text{cl}(\mathcal{S})$ 均是凸集.

光滑函数

Definition 4 (梯度). 给定函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 且 f 在点 x 的一个邻域内有意义, 若存在向量 $g \in \mathbb{R}^n$ 满足

$$\lim_{p \rightarrow 0} \frac{f(x+p) - f(x) - g^T p}{\|p\|} = 0,$$

其中 $\|\cdot\|$ 是任意的向量范数, 就称 f 在点 x 处可微 (或Fréchet 可微). 此时 g 称为 f 在点 x 处的梯度, 记作 $\nabla f(x)$. 如果对区域 D 上的每一个点 x 都有 $\nabla f(x)$ 存在, 则称 f 在 D 上可微. 若 $\nabla f(x)$ 均存在且连续, 则称 f 为连续可微函数. (记号: $f \in C^1$)

- 令 $p = \varepsilon e_i$, e_i 是第 i 个分量为 1 的单位向量,

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right]^T.$$

光滑函数

习题1: 计算二元函数 $f(x) = (\max\{0, x_1\})^2 + (\max\{0, x_2\})^2$ 的梯度.

光滑函数

Definition 5 (海瑟矩阵(Hessian)). 如果函数 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 在点 x 处的二阶偏导数 $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ $i, j = 1, 2, \dots, n$ 都存在, 则

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

称为 f 在点 x 处的海瑟矩阵.

- 当 $\nabla^2 f(x)$ 在区域 D 上的每个点 x 处都存在时, 称 f 在 D 上二阶可微. 若 $\nabla^2 f(x)$ 在 D 上还连续, 则称 f 在 D 上二阶连续可微, 可以证明此时海瑟矩阵是一个对称矩阵. (记号: $f \in C^2$)

光滑函数

Definition 6 (雅克比(Jacobi)矩阵). 设向量值函数 $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 为 $F(x) = (f_1(x), f_2(x), \dots, f_m(x))^{\top}$. 若 F 在点 x 处的偏导数 $\frac{\partial f_i(x)}{\partial x_j}$, $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ 都存在, 则

$$JF(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \dots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \dots & \frac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & \dots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

称为 F 在点 x 处的雅克比矩阵. (引入记号: $\nabla F(x) = (JF(x))^{\top}$)

光滑函数

- Hessian & Jacobian: $\nabla^2 f(x) = J(\nabla f(x))$
- Chain Rule: 若 $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $G : \mathbb{R}^m \rightarrow \mathbb{R}^q$ 均为连续可微函数, 定义 $H : \mathbb{R}^n \rightarrow \mathbb{R}^q$ 为 $H(x) = G(F(x))$. 则

$$JH(x) = JG(F(x))JF(x), \quad \nabla H(x) = \nabla F(x)\nabla G(F(x))$$

光滑函数

习题2: 计算最小二乘损失函数 $f(x) = \frac{1}{2} \|Ax - b\|^2$ 的梯度与海瑟矩阵, 其中 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

光滑函数

Theorem 2 (泰勒展开). 设 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微, 则 f 在 $\bar{x} \in \mathbb{R}^n$ 处的一阶泰勒展开式 (Taylor's Expansion) 为:

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^\top (x - \bar{x}) + o(\|x - \bar{x}\|)$$

若 f 二阶连续可微, 则 f 在 $\bar{x} \in \mathbb{R}^n$ 处的二阶泰勒展开式为:

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^\top (x - \bar{x}) + \frac{1}{2} (x - \bar{x})^\top \nabla^2 f(\bar{x}) (x - \bar{x}) + o(\|x - \bar{x}\|^2)$$

其中 $o(\|x - \bar{x}\|)$ 表示 $\|x - \bar{x}\|$ 的高阶无穷小量.

▷ 中值定理的形式:

- $\exists t \in (0, 1), f(x + p) = f(x) + \nabla f(x + tp)^\top p.$
- $\exists t \in (0, 1), \nabla f(x + p) = \nabla f(x) + \int_0^1 \nabla^2 f(x + tp) p \, dt,$

$$f(x + p) = f(x) + \nabla f(x)^\top p + \frac{1}{2} p^\top \nabla^2 f(x + tp) p$$

L -光滑函数

Definition 7 (梯度Lipschitz连续). 设 f 连续可微, 若存在常数 $L > 0$, 对 $\forall x, y \in \text{dom } f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$ 有

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|,$$

则称 f 是 L -光滑函数. (记号: $f \in C^{1,1}$)

习题3: 判断如下函数是否为 L -光滑函数:

(1) $f(x) = \|x_+\|^2$, 其中 $x_+ := (\max\{0, x_1\}, \dots, \max\{0, x_n\})^\top$;

(2) $f(x) = \frac{1}{2}\|Ax - b\|^2$, 其中 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

L -光滑函数

▷ 重要性质:

- **二次上界**: 设 f 是 L -光滑函数且 $\text{dom } f$ 为凸集, 则 f 有二次上界:

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|^2 \quad \forall x, y \in \text{dom } f.$$

Hint: 构造辅助函数 $g(t) = f(x + t(y - x))$, $t \in [0, 1]$

- 设 f 可微, $\text{dom } f = \mathbb{R}^n$, 且存在一个全局极小点 x^* . 若 f 是 L -光滑的, 则对 $\forall x \in \mathbb{R}^n$ 有

$$\frac{1}{2L} \|\nabla f(x)\|^2 \leq f(x) - f(x^*).$$

Hint: $f(x^*) \leq \inf_{y \in \mathbb{R}^n} f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2} \|y - x\|^2$ & 无约束凸优化一阶最优性条件

广义实值函数与适当函数

Definition 8 (广义实值函数(extended real-valued function)).

令 $\overline{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{\pm\infty\}$ 为广义实数空间, 称映射 $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ 为广义实值函数.

Definition 9 (适当函数). 给定广义实值函数 $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ 和非空集合 \mathcal{X} . 如果存在 $x \in \mathcal{X}$ 使得 $f(x) < +\infty$, 并且对任意的 $x \in \mathcal{X}$, 都有 $f(x) > -\infty$, 那么称函数 f 关于集合 \mathcal{X} 是适当的.

习题4: 给定非空集合 $C \subseteq \mathbb{R}^n$, 定义 C 的示性函数(indicator function)

$$\text{为: } I_C(x) := \begin{cases} 0, & x \in C; \\ +\infty, & x \notin C. \end{cases} \quad \text{判断 } I_C \text{ 是否有适当函数.}$$

凸函数

Definition 10. 设 f 为适当函数, 若 $\text{dom } f$ 是凸集,

- f 是凸函数: 若对所有 $x, y \in \text{dom } f$ 和 $0 \leq \theta \leq 1$ 有

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

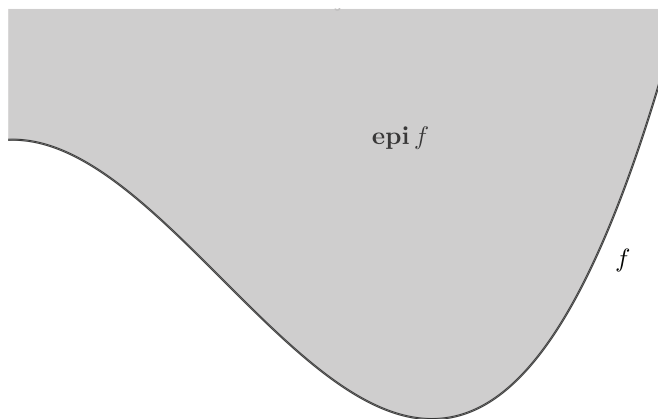
- f 是严格凸函数: 若对所有 $x, y \in \text{dom } f$ 和 $x \neq y, 0 < \theta < 1$, 有

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

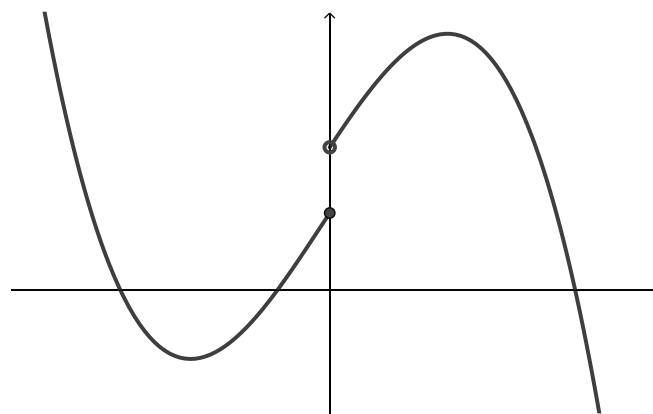
下水平集和上方图

Definition 11. 设广义实值函数 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$.

- f 的 α -下水平集: $C_\alpha = \{x \mid f(x) \leq \alpha\}$.
- f 的上方图: $\text{epi } f = \{(x, t) \in \mathbb{R}^{n+1} \mid f(x) \leq t\}$.
- f 为闭函数: $\text{epi } f$ 为闭集.
- f 为下半连续函数: 对任意的 $x \in \mathbb{R}^n$, 有 $\liminf_{y \rightarrow x} f(y) \geq f(x)$.



(a) 上方图 $\text{epi } f$



(b) 下半连续函数 $f(x)$

闭函数与下半连续函数

Theorem 3. 设广义实值函数 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, 则下列命题等价:

- ① $f(x)$ 的任意 α -下水平集都是闭集;
- ② $f(x)$ 是下半连续的;
- ③ $f(x)$ 是闭函数.

下半连续函数的性质:

- **加法:** 若 f 与 g 均为适当的下半连续函数, 且 $\text{dom} f \cap \text{dom} g \neq \emptyset$, 则 $f + g$ 也是下半连续函数.
- **仿射函数的复合:** 若 f 为下半连续函数, 则 $f(Ax + b)$ 也为下半连续函数;
- **上确界:** 若 $f_\lambda, \lambda \in \Lambda$ 均为下半连续函数, 则 $\sup_{\lambda \in \Lambda} f_\lambda(x)$ 也为下半连续函数.

凸函数的性质

- 设 f 为凸函数, 则 f 的所有 α -下水平集都是凸集.
- **Jensen不等式**: 设 f 是凸函数, 则对于 $1 \leq i \leq m, x_i \in \text{dom} f$, $0 \leq \theta_i \leq 1$ 且 $\sum_{i=1}^m \theta_i = 1$, 有

$$f\left(\sum_{i=1}^m \theta_i x_i\right) \leq \sum_{i=1}^m \theta_i f(x_i).$$

- **概率Jensen不等式**: 设 f 是凸函数, 则对任意随机变量 z ,

$$f(\mathbf{E}z) \leq \mathbf{E}f(z).$$

- 设 $f: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 为凸函数, 则 f 在 $\text{int dom} f$ 连续.

凸函数的判定

- 函数 $f(x)$ 为凸函数当且仅当其上方图 $\text{epi} f$ 是凸集.
- 设 f 为可微函数且 $\text{dom} f$ 是凸集, 则 f 是凸函数当且仅当

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \quad \forall x, y \in \text{dom} f. \quad (1)$$

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq 0, \quad \forall x, y \in \text{dom} f. \quad (2)$$

f 是严格凸函数当且仅当(1)或(2)对所有 $x, y \in \text{dom} f$ 且 $x \neq y$ 严格成立.

- 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

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

如果 $\nabla^2 f(x) \succ 0 \quad \forall x \in \text{dom} f$, 则 f 是严格凸函数.

强凸函数

- **强凸函数**: 若存在常数 $m > 0$, 使得 $g(x) = f(x) - \frac{m}{2}\|x\|^2$ 为凸函数.
- **强凸函数**: 若存在常数 $m > 0$, 使得对 $\forall x, y \in \text{dom} f, \theta \in (0, 1)$ 有

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y) - \frac{m}{2}\theta(1 - \theta)\|x - y\|^2.$$

- 设 f 为可微函数且 $\text{dom} f$ 是凸集, 则 f 是 m -强凸函数当且仅当

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq m\|x - y\|^2, \quad \forall x, y \in \text{dom} f.$$

- **二次下界**: 设 f 为可微 m -强凸函数, 则 f 的所有 α -下水平集有界,

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}\|x - y\|^2 \quad \forall x, y \in \text{dom} f.$$

Hint: $g(x) = f(x) - \frac{m}{2}\|x\|_2^2$ 凸 & 一阶判别条件

保凸函数运算

- **非负加权和:** 若 f_1, f_2 是凸函数且 $\alpha_1, \alpha_2 \geq 0$, 则 $\alpha_1 f_1 + \alpha_2 f_2$ 是凸函数.
- **与仿射函数的复合:** 若 f 是凸函数, 则 $f(Ax + b)$ 是凸函数.
【例】 $f(x) = \|Ax + b\|$.
- **逐点取最大值:** 若 f_1, \dots, f_m 是凸函数, 则 $\max\{f_1(x), \dots, f_m(x)\}$ 是凸函数. 【例】 $f(x) = \max_{i=1, \dots, m} (a_i^T x + b_i)$.
- **取下确界:** 若 $f(x, y)$ 关于 (x, y) 整体是凸函数, C 是凸集, 则

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

是凸函数.

【例】 点 x 到凸集 S 的距离 $\text{dist}(x, S) = \inf_{y \in S} \|x - y\|$ 是凸函数.

- 取上确界: 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】

1. 集合 C 的支撑函数: $S_C(x) = \sup_{y \in C} y^T x$ 是凸函数.
2. 点 x 到集合 C 的最远距离: $f(x) = \sup_{y \in C} \|x - y\|$ 是凸函数.
3. 对称矩阵 $X \in \mathbb{S}^n$ 的最大特征值

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

是凸函数.

- 与标量函数的复合: 给定函数 $g : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R} \rightarrow \mathbb{R}$, 令 $f(x) = h(g(x))$. 若 g 是凸函数且 h 是单调增凸函数, 则 f 是凸函数; 若 g 是凹函数且 h 是单调减凸函数, 则 f 是凸函数.

【例】若 g 是凸函数, 则 $\exp g(x)$ 是凸函数; 若 g 是正值凹函数, 则 $1/g(x)$ 是凸函数.

- 与向量函数的复合: 给定函数 $g : \mathbb{R}^n \rightarrow \mathbb{R}^k$, $h : \mathbb{R}^k \rightarrow \mathbb{R}$, 令

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), \dots, g_k(x)).$$

若 g_i 是凸函数, h 是凸函数且关于每个分量单调增, 则 f 是凸函数;
若 g_i 是凹函数, h 是凸函数且关于每个分量单调减, 则 f 是凸函数.

【例】若 g_i 是正值凹函数, 则 $\sum_{i=1}^m \ln g_i(x)$ 是凹函数; 若 g_i 是凸函数, 则 $\ln \sum_{i=1}^m \exp g_i(x)$ 是凸函数.

- 透视函数: 定义 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 的透视函数 $g: \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$ 如下:

$$g(x, t) = tf\left(\frac{x}{t}\right), \quad \text{dom } g = \{(x, t) \mid \frac{x}{t} \in \text{dom } f, t > 0\}.$$

若 f 是凸函数, 则 g 是凸函数.

【例】

1. 相对熵函数 $g(x, t) = t \log t - t \log x$ 是 \mathbb{R}_{++}^2 上的凸函数.
2. 若 f 是凸函数, 则

$$g(x) = (c^T x + d)f\left(\frac{Ax + b}{c^T x + d}\right)$$

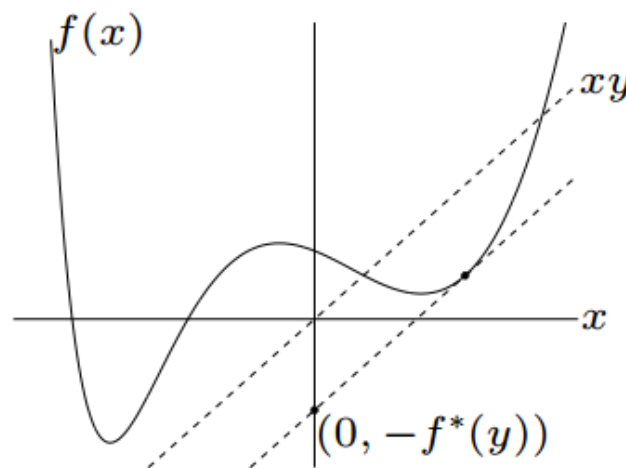
是区域 $\left\{x \mid c^T x + d > 0, \frac{Ax + b}{c^T x + d} \in \text{dom } f\right\}$ 上的凸函数.

共轭函数

Definition 12. 适当函数 f 的共轭函数为:

$$f^*(y) = \sup_{x \in \text{dom } f} \{y^T x - f(x)\}.$$

- f^* 恒为凸函数, 无论 f 是否是凸函数.
- Fenchel不等式: $f(x) + f^*(y) \geq x^T y$.



常见函数的共轭函数

- 强凸二次函数 $f(x) = \frac{1}{2}x^T Qx$, $Q \in \mathbb{S}_{++}^n$ 的共轭函数:

$$f^*(y) = \sup_x \{y^T x - \frac{1}{2}x^T Qx\} = \frac{1}{2}y^T Q^{-1}y.$$

- 给定非空凸集 C , 示性函数 $I_C(x)$ 的共轭函数为支撑函数 $I_C^*(y)$

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

$$I_C^*(y) = \sup_x \{y^T x - I_C(x)\} = \sup_{x \in C} y^T x.$$

- 范数 $f(x) = \|x\|$ 的共轭函数:

$$f^*(y) = \sup_x \{y^T x - \|x\|\} = \begin{cases} 0 & \|y\|_* \leq 1, \\ +\infty & \|y\|_* > 1. \end{cases}$$

对偶范数: $\|y\|_* = \sup_{\|x\| \leq 1} y^T x.$

范数与对偶范数的关系: $x^T y \leq \|x\| \|y\|_*, \forall x, y \in \mathbb{R}^n.$

Proof. 若 $\|y\|_* \leq 1$, 则对 $\forall x \in \mathbb{R}^n$ 有 $x^T y \leq \|x\|$, 且当 $x = 0$ 时等号成立, 从而 $f^*(y) = \sup_x \{y^T x - \|x\|\} = 0.$

若 $\|y\|_* > 1$, 则至少存在一个 x 使得 $\|x\| \leq 1, x^T y > 1$. 从而对 $\forall t > 0$,

$$f^*(y) = \sup_x \{y^T x - \|x\|\} \geq y^T(tx) - \|tx\| = t(y^T x - \|x\|), \quad (3)$$

当 $t \rightarrow +\infty$ 时, (3) 的右端趋于 $+\infty$. □

双共轭函数

Definition 13. 函数 f 的**双共轭函数**为:

$$f^{**}(x) = \sup_{y \in \text{dom } f^*} \{x^T y - f^*(y)\}.$$

注:

(i) f^{**} 恒为闭凸函数, 且由Fenchel不等式知

$$f^{**}(x) \leq f(x) \quad \forall x; \quad \text{或等价地, } \text{epi } f \subseteq \text{epi } f^{**}.$$

(ii) 若 f 为闭凸函数, 则

$$f^{**}(x) = f(x) \quad \forall x; \quad \text{或等价地, } \text{epi } f = \text{epi } f^{**}.$$

Proof. 用反证法. 若 $(x, f^{**}(x)) \notin \text{epi} f$, 则 $\exists a \in \mathbb{R}^n, b, c \in \mathbb{R}, (a, b) \neq 0$ 且 $b \leq 0$ (若 $b > 0$, 令 $s \rightarrow +\infty$, 可推出矛盾), 使得

$$\begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} z - x \\ s - f^{**}(x) \end{bmatrix} \leq c < 0 \quad \forall (z, s) \in \text{epi} f. \quad (4)$$

若 $b < 0$, 在(4)中取 $s = f(z)$, 则

$$a^T z + b f(z) - a^T x - b f^{**}(x) \leq c. \quad (5)$$

令 $y = -\frac{a}{b}$, 由(4)得 $f^*(y) - x^T y + f^{**}(x) \leq -\frac{c}{b} < 0$, 与Fenchel不等式矛盾.

若 $b = 0$, 取 $\hat{y} \in \text{dom } f^*$, 则对 $\forall \varepsilon > 0$, $\forall (z, s) \in \text{epi} f$, 由Fenchel不等式知, $\hat{y}^T z - s \leq \hat{y}^T z - f(z) \leq f^*(\hat{y})$, 从而有

$$\begin{bmatrix} a + \varepsilon \hat{y} \\ -\varepsilon \end{bmatrix}^T \begin{bmatrix} z - x \\ s - f^{**}(x) \end{bmatrix} \leq c + \varepsilon (f^*(\hat{y}) - x^T \hat{y} + f^{**}(x)) < 0,$$

□

次梯度

Definition 14. 设 f 为适当凸函数, $x \in \text{dom } f$.

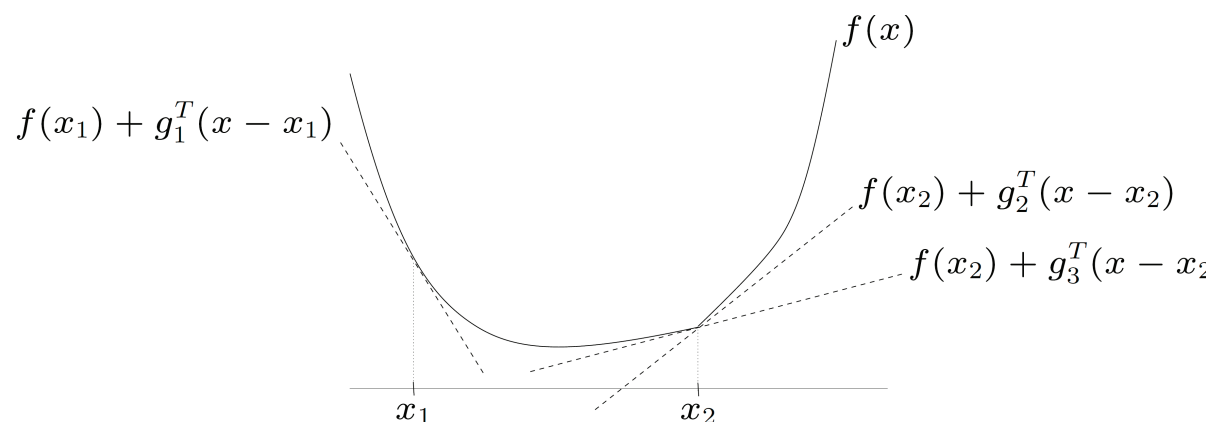
- 若向量 $g \in \mathbb{R}^n$ 满足

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

则称 g 为函数 f 在点 x 处的一个次梯度.

- f 在点 x 处的次微分:

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



- 若 f 是可微凸函数, 则 $\nabla f(x)$ 是 f 在点 x 处的一个次梯度.
- 次梯度 g 可提供 $f(y)$ 的一个全局下界: $f(x) + g^T(y - x)$.
- 次梯度 g 可诱导出上方图 $\mathbf{epi} f$ 在点 $(x, f(x))$ 处的一个支撑超平面:

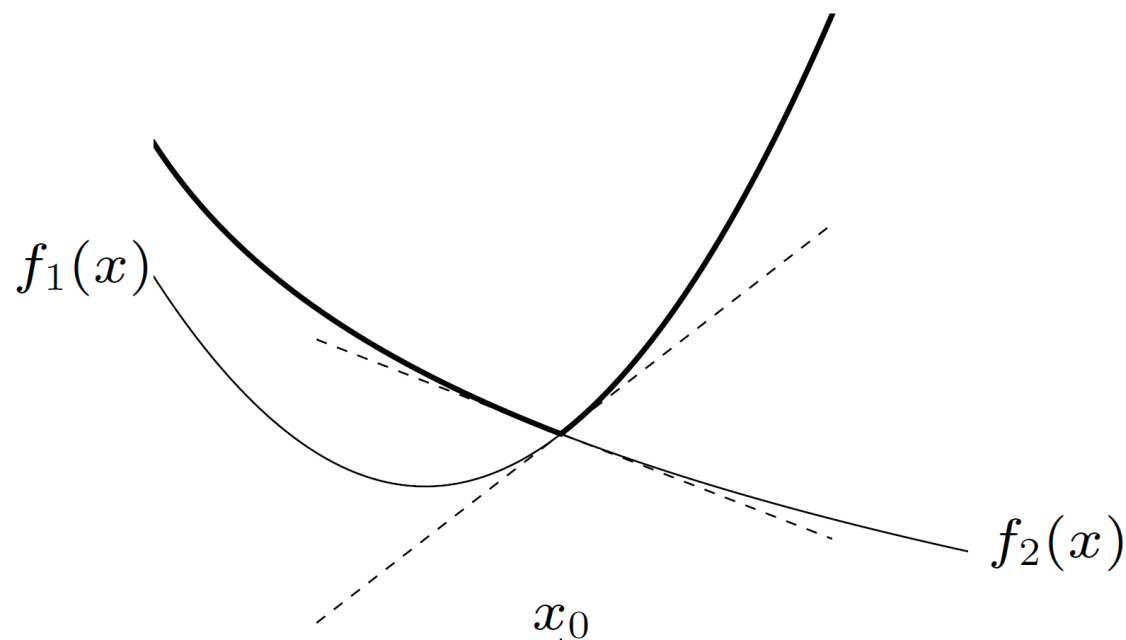
$$\begin{bmatrix} g \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \quad \forall (y, t) \in \mathbf{epi} f.$$

- **次梯度存在性**: 设 f 是凸函数, 若 $x \in \mathbf{intdom} f$, 则 $\partial f(x) \neq \emptyset$.
- ℓ_2 范数 $f(x) = \|x\|_2$ 的次微分:

$$\partial \|x\|_2 = \begin{cases} \left\{ \frac{x}{\|x\|_2} \right\} & \text{if } x \neq 0, \\ \{g : \|g\|_2 \leq 1\} & \text{if } x = 0. \end{cases}$$

- 绝对值函数 $f(x) = |x|$ 在点 $x = 0$ 处的次微分: $\partial f(0) = [-1, 1]$.

【例】 $f(x) = \max\{f_1(x), f_2(x)\}$ f_1, f_2 是可微凸函数.



- $f(x)$ 在点 x_0 处的次梯度可取范围 $[\nabla f_1(x_0), \nabla f_2(x_0)]$.
- 若 $f_1(\hat{x}) > f_2(\hat{x})$, 则 f 在点 \hat{x} 处的次梯度等于 $\nabla f_1(\hat{x})$.
- 若 $f_1(\hat{x}) < f_2(\hat{x})$, 则 f 在点 \hat{x} 处的次梯度等于 $\nabla f_2(\hat{x})$.

次梯度的性质

- 设 f 是凸函数, 则对 $\forall x \in \mathbf{dom} f$, $\partial f(x)$ 是闭凸集(可能为空集).
- 设 f 是凸函数, 则对 $\forall x \in \mathbf{intdom} f$, $\partial f(x)$ 是非空有界集.
- 设凸函数 $f(x)$ 在 $x_0 \in \mathbf{intdom} f$ 处可微, 则 $\partial f(x_0) = \{\nabla f(x_0)\}$.
- 次梯度的单调性 设 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 是凸函数, $x, y \in \mathbf{dom} f$, 则

$$(u - v)^T (x - y) \geq 0, \quad \forall u \in \partial f(x), \forall v \in \partial f(y).$$

- 次梯度的连续性 设 $f(x)$ 是闭凸函数且 $\partial f(\bar{x}) \neq \emptyset$. 若

$$\lim_{k \rightarrow \infty} x^k = \bar{x}, \quad g^k \in \partial f(x^k) \text{ 且 } \lim_{k \rightarrow \infty} g^k = \bar{g},$$

则 $\bar{g} \in \partial f(\bar{x})$.

凸函数的方向导数

Definition 15. 对于凸函数 f , 给定点 $x_0 \in \text{dom } f$ 以及方向 $d \in \mathbb{R}^n$, 其方向导数定义为

$$\partial f(x_0; d) = \inf_{t>0} \frac{f(x_0 + td) - f(x_0)}{t}.$$

注: 对于可微函数, $\partial f(x_0; d) = \nabla f(x_0)^T d$.

- **方向导数有限:** 设 $f(x)$ 为凸函数, $x_0 \in \text{int dom } f$, 则对 $\forall d \in \mathbb{R}^n$, $\partial f(x_0; d)$ 有限.
- **方向导数和次梯度:** 设 $f: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 为凸函数, $x_0 \in \text{int dom } f$, 则对 $\forall d \in \mathbb{R}^n$, 有

$$\partial f(x_0; d) = \max_{g \in \partial f(x_0)} g^T d.$$

次梯度的计算

- **凸函数的非负线性组合：** 设 $\alpha_1, \alpha_2 \geq 0$, 凸函数 f_1, f_2 满足 $\text{int dom } f_1 \cap \text{dom } f_2 \neq \emptyset$, 若

$$f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x), \quad x \in \text{dom } f_1 \cap \text{dom } f_2$$

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, $f(x) = h(Ax + b)$. 若存在 $x^\# \in \mathbb{R}^m$ 使得 $Ax^\# + b \in \text{int dom } h$, 则

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

► 两个函数之和的次梯度

Theorem 4 (Moreau-Rockafellar定理). 设 $f_1, f_2 : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 是凸函数, 则对任意的 $x_0 \in \mathbb{R}^n$,

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

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

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

► 凸函数族最大值的次梯度

Theorem 5. 设 $f_1, \dots, f_m : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 为凸函数, 令

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

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

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

【例】 $f(x) = \max_{i=1,2,\dots,m} \{a_i^T x + b_i\}$ 的次梯度.

► 逐点上确界函数的次梯度

Theorem 6. 设 $\{f_\alpha \mid \mathbb{R}^n \rightarrow (-\infty, +\infty]\}_{\alpha \in \mathcal{A}}$ 是一族凸函数, 令

$$f(x) = \sup_{\alpha \in \mathcal{A}} f_\alpha(x).$$

对 $x_0 \in \bigcap_{\alpha \in \mathcal{A}} \text{int dom } f_\alpha$, 定义 $I(x_0) = \{\alpha \in \mathcal{A} \mid f_\alpha(x_0) = f(x_0)\}$, 则

$$\text{conv} \bigcup_{\alpha \in I(x_0)} \partial f_\alpha(x_0) \subseteq \partial f(x_0).$$

若 \mathcal{A} 是紧集且 f_α 关于 α 连续, 则

$$\text{conv} \bigcup_{\alpha \in I(x_0)} \partial f_\alpha(x_0) = \partial f(x_0).$$

► 固定分量的函数极小值的次梯度

Theorem 7. 设 $h : \mathbb{R}^n \times \mathbb{R}^m \rightarrow (-\infty, +\infty]$ 是关于 (x, y) 的凸函数, $f(x) = \inf_y h(x, y)$. 对 $\hat{x} \in \mathbb{R}^n$, 设 $\hat{y} \in \mathbb{R}^m$ 满足 $h(\hat{x}, \hat{y}) = f(\hat{x})$, 且存在 $g \in \mathbb{R}^n$ 使得 $(g, 0) \in \partial h(\hat{x}, \hat{y})$, 则 $g \in \partial f(\hat{x})$.

► 复合函数的次梯度

Theorem 8. 设 $f_1, f_2, \dots, f_m : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 为 m 个凸函数, $h : \mathbb{R}^m \rightarrow (-\infty, +\infty]$ 为关于各分量单调递增的凸函数, 令

$$f(x) = h(f_1(x), f_2(x), \dots, f_m(x)).$$

设 $z = (z_1, \dots, z_m) \in \partial h(f_1(\hat{x}), \dots, f_m(\hat{x}))$, $g_i \in \partial f_i(\hat{x})$, 则

$$g \stackrel{\text{def}}{=} z_1 g_1 + z_2 g_2 + \dots + z_m g_m \in \partial f(\hat{x}).$$

强凸函数共轭函数的性质

Theorem 9. 设 $f(x)$ 是适当且闭的强凸函数, 强凸参数为 $\mu > 0$, 则 $f^*(y)$ 在全空间 \mathbb{R}^n 上有定义, 且是 $\frac{1}{\mu}$ -光滑函数.

Proof. 对任意的 $y \in \mathbb{R}^n$, $f(x) - x^T y$ 是强凸函数, 因此对任意的 $y \in \mathbb{R}^n$, 存在唯一的 $x \in \text{dom } f$, 使得 $f^*(y) = x^T y - f(x)$. 根据最优性条件

$$y \in \partial f(x) \Leftrightarrow f^*(y) = x^T y - f(x).$$

由于 $f(x)$ 是闭凸函数, 二次共轭为其本身, 于是对同一组 x, y 有

$$x^T y - f^*(y) = f(x) = f^{**}(x) = \sup_y \{x^T y - f^*(y)\}.$$

这说明 y 也使得 $x^T y - f^*(y)$ 取到最大值. 根据一阶最优性条件,

$$x \in \partial f^*(y).$$

再根据 x 的唯一性容易推出 $\partial f^*(y)$ 中只含一个元素, 故 $f^*(y)$ 可微且 $\nabla f^*(y) = x$.

下证 $f^*(y)$ 为梯度 $\frac{1}{\mu}$ -Lipschitz连续的. 对任意的 y_1, y_2 , 存在唯一的 $x_1, x_2 \in \mathbf{dom} f$ 使得

$$y_1 \in \partial f(x_1), \quad y_2 \in \partial f(x_2).$$

根据次梯度性质以及 $f(x) - \frac{\mu}{2}\|x\|^2$ 是凸函数,

$$f(x_2) \geq f(x_1) + (y_1 - \mu x_1)^T(x_2 - x_1),$$

$$f(x_1) \geq f(x_2) + (y_2 - \mu x_2)^T(x_1 - x_2).$$

两式相加得

$$(y_1 - y_2)^T(x_1 - x_2) \geq \mu\|x_1 - x_2\|^2.$$

因 $x_1 = \nabla f^*(y_1)$, $x_2 = \nabla f^*(y_2)$, 故有

$$(y_1 - y_2)^T(\nabla f^*(y_1) - \nabla f^*(y_2)) \geq \mu\|\nabla f^*(y_1) - \nabla f^*(y_2)\|^2.$$

因此, $\nabla f^*(y)$ 是 $\frac{1}{\mu}$ -利普希茨连续的.

□

非凸函数的次微分

Definition 16. 设 $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 为适当、下半连续函数.

- 对给定的 $x \in \text{dom } f$, 满足如下条件的所有向量 $u \in \mathbb{R}^n$ 的集合定义为 f 在点 x 处的 **Fréchet 次微分**:

$$\liminf_{y \rightarrow x, y \neq x} \frac{f(y) - f(x) - u^\top (y - x)}{\|y - x\|},$$

记为 $\hat{\partial}f(x)$; 若 $x \notin \text{dom } f$, 定义 $\hat{\partial}f(x) = \emptyset$.

- f 在点 x 处的 **极限次微分(或Mordukhovich次微分, 或简称为次微分)**:

$$\partial f(x) = \left\{ u \in \mathbb{R}^n \mid \exists x^k \rightarrow x, f(x^k) \rightarrow f(x), u^k \in \hat{\partial}f(x^k), u^k \rightarrow u \right\}.$$

$$\text{i.e., } \partial f(x) = \limsup_{y \rightarrow_f x} \hat{\partial}f(y).$$

Thank you!