提纲

回顾可微凸函数f 的一阶条件:

$$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}}(y - x)$$

- f 在点x 处的一阶近似是f 的一个全局下界
- $\nabla f(x)$ 可以诱导出上方图$\mathbf{epi}\, f$ 在点$(x, f(x))$ 处的支撑超平面

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

若f 不可微，可否类似地定义一种梯度，使之具有梯度的一些性质?

次梯度

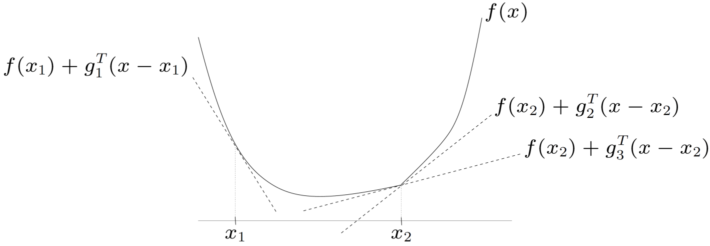
- 设f 为适当凸函数，x为定义域$\mathbf{dom}\, f$中的一点．若向量$g \in \mathbb{R}^n$ 满足
$$f(y) \geq f(x) + g^{\mathrm{T}}(y-x), \quad \forall y \in \mathbf{dom}\, f,$$
则称g 为函数f 在点x 处的一个**次梯度**．
- 进一步地，称集合
$$\partial f(x) = \{g \mid g \in \mathbb{R}^n, f(y) \geq f(x) + g^{\mathrm{T}}(y-x), \forall y \in \mathbf{dom}\, f\}$$
为f 在点x 处的**次微分**.

次梯度

- $f(x)+g^{\mathrm{T}}(y-x)$ 是$f(y)$ 的一个全局下界
- g 可以诱导出上方图$\mathbf{epi}\, f$ 在点$(x,f(x))$ 处的一个支撑超平面

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

- 如果f 是可微凸函数，那么$\nabla f(x)$ 是f 在点x 处的一个次梯度
- 例：g_2,g_3 是点x_2 处的次梯度；g_1 是点x_1 处的次梯度



次梯度存在性

设f为凸函数，$\mathbf{dom}\, f$为其定义域．如果$x \in \mathbf{int\ dom}\, f$，则$\partial f(x)$是非空的，其中$\mathbf{int\ dom}\, f$的含义是集合$\mathbf{dom}\, f$的所有内点.

证明：

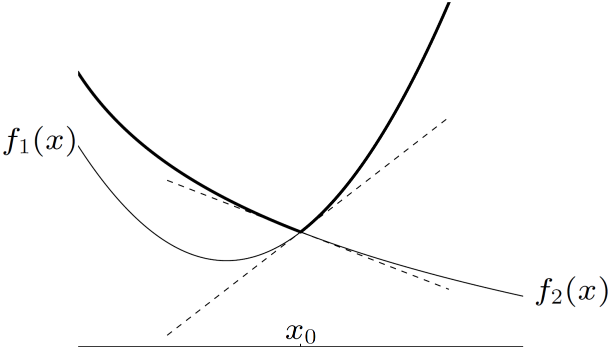
- $(x, f(x))$ 是$\mathbf{epi}\, f$ 边界上的点
- 因此存在$\mathbf{epi}\, f$ 在点$(x, f(x))$ 处的支撑超平面:

$$\exists\ (a, b) \neq 0, \quad \begin{bmatrix} a \\ b \end{bmatrix}^{\mathrm{T}} \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \quad \forall\ (y, t) \in \mathbf{epi}\, f$$

- 令$t \to +\infty$, 可知$b \leq 0$
- 取$y = x + \epsilon a \in \mathbf{dom}\, f$, $\epsilon > 0$, 可知$b \neq 0$
- 因此$b < 0$ 并且$g = a/|b|$ 是f 在点x 处的次梯度

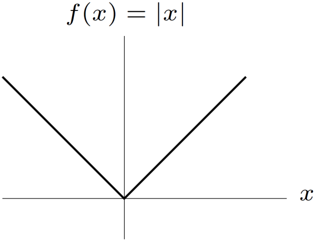
例

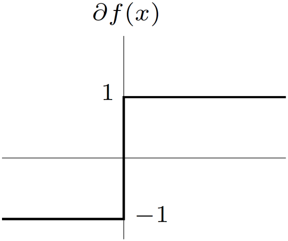
$$f(x) = \max\{f_1(x), f_2(x)\} \quad f_1, f_2 \text{ 是可微凸函数}$$



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

- 绝对值函数 $f(x) = |x|$





- 欧几里得范数 $f(x) = \|x\|_2$

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

提纲

次微分是闭凸集

对任何$x \in \mathbf{dom}\, f$，$\partial f(x)$ 是一个闭凸集（可能为空集）.

证明：

- 设$g_1, g_2 \in \partial f(x)$，并设$\lambda \in (0,1)$，由次梯度的定义

$$f(y) \geq f(x) + g_1^{\mathrm{T}}(y-x), \quad \forall y \in \mathrm{dom} f,$$
$$f(y) \geq f(x) + g_2^{\mathrm{T}}(y-x), \quad \forall y \in \mathrm{dom} f.$$

  由上面第一式的λ 倍加上第二式的$(1-\lambda)$ 倍，我们可以得到$\lambda g_1 + (1-\lambda) g_2 \in \partial f(x)$，从而$\partial f(x)$ 是凸集.

- 令$g_k \in \partial f(x)$为次梯度且$g_k \to g$，则

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

  在上述不等式中取极限，并注意到极限的保号性，最终我们有

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

  这说明$\partial f(x)$为闭集.

内点的次微分非空有界

如果$x \in \textbf{int dom} f$，则$\partial f(x)$ 非空有界集.

证明：

- 非空可由次梯度存在性直接得出
- 取充分小的$r > 0$, 使得

$$B = \{x \pm re_i | i = 1, \cdots, n\} \subset \textbf{dom} f$$

- 对任意非零的$g \in \partial f(x)$, 存在$y \in B$ 满足

$$f(y) \geq f(x) + g^{\mathrm{T}}(y - x) = f(x) + r\|g\|_\infty$$

- 由此得到$\partial f(x)$ 有界:

$$\|g\|_\infty \leq \frac{\max_{y \in B} f(y) - f(x)}{r} < +\infty$$

可微函数的次微分

设凸函数$f(x)$在$x_0 \in \textbf{int dom}\, f$处可微，则$\partial f(x_0) = \{\nabla f(x_0)\}$.

证明：

- 根据可微凸函数的一阶条件可知梯度$\nabla f(x_0)$为次梯度.
- 下证$f(x)$在点x_0处不可能有其他次梯度. 设$g \in \partial f(x_0)$，根据次梯度的定义，对任意的非零$v \in \mathbb{R}^n$且$x_0 + tv \in \textbf{dom}\, f, t > 0$有

$$f(x_0 + tv) \geq f(x_0) + tg^{\mathrm{T}} v.$$

若$g \neq \nabla f(x_0)$，取$v = g - \nabla f(x_0) \neq 0$，上式变形为

$$\frac{f(x_0 + tv) - f(x_0) - t\nabla f(x_0)^{\mathrm{T}} v}{t\|v\|} \geq \frac{(g - \nabla f(x_0))^{\mathrm{T}} v}{\|v\|} = \|v\|.$$

- 不等式两边令$t \to 0$，根据Fréchet 可微的定义，左边趋于0，而右边是非零正数，可得到矛盾.

次梯度的单调性

设$f : \mathbb{R}^n \to \mathbb{R}$为凸函数，$x, y \in \mathbf{dom}\, f$，则$(u-v)^\mathrm{T}(x-y) \geq 0$, 其中$u \in \partial f(x)$，$v \in \partial f(y)$．

证明：

- 由次梯度的定义，

$$f(y) \geq f(x) + u^\mathrm{T}(y-x),$$
$$f(x) \geq f(y) + v^\mathrm{T}(x-y).$$

- 将以上两个不等式相加即得结论．

次梯度的连续性

设$f(x)$ 是闭凸函数且∂f 在点$\bar{x}$ 附近存在且非空．若序列$x^k \to \bar{x}$，$g^k \in \partial f(x^k)$ 为$f(x)$ 在点x^k 处的次梯度，且$g^k \to \bar{g}$，则$\bar{g} \in \partial f(\bar{x})$．

证明：

- 对任意$y \in \mathrm{dom} f$，根据次梯度的定义，

$$f(y) \geq f(x^k) + \langle g^k, y - x^k \rangle.$$

- 对上述不等式两边取下极限，我们有

$$\begin{aligned} f(y) &\geq \liminf_{k\to\infty} [f(x^k) + \langle g^k, y - x^k \rangle] \\ &\geq f(\bar{x}) + \langle \bar{g}, y - \bar{x} \rangle, \end{aligned}$$

其中第二个不等式利用了$f(x)$的下半连续性以及$g^k \to \bar{g}$，由此可推出$\bar{g} \in \partial f(\bar{x})$．

次梯度的计算规则

弱次梯度计算: 得到一个次梯度

- 足以满足大多数不可微凸函数优化算法
- 如果可以获得任意一点处$f(x)$ 的值，那么总可以计算一个次梯度

强次梯度计算: 得到$\partial f(x)$，即所有次梯度

- 一些算法、最优性条件等，需要完整的次微分
- 计算可能相当复杂

下面我们假设$x \in \textbf{int dom} f$

基本规则

- **可微凸函数：** 若凸函数f 在点x 处可微，则$\partial f(x) = \{\nabla f(x)\}$.
- **凸函数的非负线性组合：** 设凸函数f_1, f_2 满足$\textbf{int dom}\, f_1 \cap \textbf{dom}\, f_2 \neq \emptyset$，而$x \in \textbf{dom}\, f_1 \cap \textbf{dom}\, f_2$. 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$Ax^\sharp + b \in \textbf{int dom}\, h$，则

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

两个函数之和的次梯度

设$f_1, f_2 : \mathbb{R}^n \to (-\infty, +\infty]$ 是两个凸函数，则对任意的$x_0 \in \mathbb{R}^n$，

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

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

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

证明:

- 第一个结论由次梯度的定义是显然的．以下我们证第二个结论．
- 对于任意给定的x_0，设$g \in \partial(f_1 + f_2)(x_0)$．如果$f_1(x_0) = +\infty$，则$(f_1 + f_2)(x_0) = +\infty$．由次梯度的定义，我们有

$$(f_1 + f_2)(x) \geq (f_1 + f_2)(x_0) + g^{\mathrm{T}}(x - x_0)$$

对任意$x \in \mathbb{R}^n$ 成立，故$f_1 + f_2 \equiv +\infty$．这与$\textbf{int dom}\, f_1 \cap \textbf{dom}\, f_2 \neq \emptyset$ 矛盾，因此以下我们假设$f_1(x_0), f_2(x_0) < +\infty$．

- 定义如下两个集合，容易验证S_1, S_2 均为非空凸集

$$\begin{aligned} S_1 &= \{(x-x_0, y) \in \mathbb{R}^n \times \mathbb{R} \mid y > f_1(x) - f_1(x_0) - g^{\mathrm{T}}(x-x_0)\}, \\ S_2 &= \{(x-x_0, y) \in \mathbb{R}^n \times \mathbb{R} \mid y \le f_2(x_0) - f_2(x)\}, \end{aligned}$$

- 设$(x-x_0, y) \in S_1 \cap S_2$，则

$$\begin{aligned} y &> f_1(x) - f_1(x_0) - g^{\mathrm{T}}(x-x_0), \\ y &\le f_2(x_0) - f_2(x). \end{aligned}$$

  上两式相减即得

$$(f_1+f_2)(x) < (f_1+f_2)(x_0) + g^{\mathrm{T}}(x-x_0),$$

  这与$g \in \partial(f_1+f_2)(x_0)$ 矛盾．因此$S_1 \cap S_2 = \emptyset$．
- 根据凸集分离定理，存在非零的(a,b) 和另一个实数c ，使得

$$\begin{aligned} a^{\mathrm{T}}(x-x_0) + by &\le c, \quad \forall (x-x_0, y) \in S_1, && (1) \\ a^{\mathrm{T}}(x-x_0) + by &\ge c, \quad \forall (x-x_0, y) \in S_2. && (2) \end{aligned}$$

  注意到$(0,0) \in S_2$，故$c \le 0$．此外，$(0,\varepsilon) \in S_1$ 对任何$\varepsilon > 0$ 成立，由此可得$c = 0$ 以及$b \le 0$．

- 如果$b=0$，则由上两式即得$a^{\mathrm{T}}(x-x_0)=0$ 对任何$x\in\textbf{dom}\, f_1\cap\textbf{dom}\, f_2$ 成立．现在取$\hat{x}\in\textbf{int dom}\, f_1\cap\textbf{dom}\, f_2$，并设$\delta>0$ 使得点$\hat{x}$ 处的邻域$N_\delta(\hat{x})\subset\textbf{int dom}\, f_1\cap\mathrm{dom}f_2$，则

$$a^{\mathrm{T}}u=a^{\mathrm{T}}(\hat{x}+u-x_0)$$

对任何$u\in\mathbb{R}^n$ 成立．此时再令$u=\frac{\delta a}{2\|a\|_2}$ 即得$a=0$．但这与(a,b) 非零矛盾，故b 不可能为0．

- 现将(1) 式除以$-b$，并令$\hat{a}=-\frac{a}{b}$，就得到

$$\begin{aligned}\hat{a}^{\mathrm{T}}(x-x_0)&\le y,\quad\forall(x-x_0,y)\in S_1,\\\hat{a}^{\mathrm{T}}(x-x_0)&\ge y,\quad\forall(x-x_0,y)\in S_2.\end{aligned}$$

利用上面两个式子和S_1 和S_2 的定义可以分别得到$g+\hat{a}\in\partial f_1(x_0)$ 和$-\hat{a}\in\partial f_2(x_0)$．因此$g=(g+\hat{a})+(-\hat{a})\in\partial f_1(x_0)+\partial f_2(x_0)$．

函数族的上确界

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

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

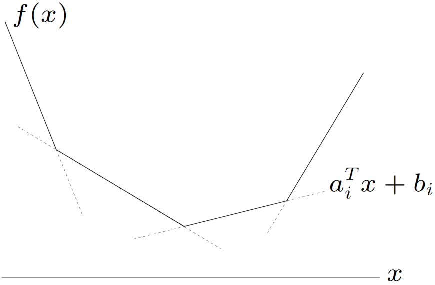
对$x_0\in\bigcap_{i=1}^{m}\textbf{int dom}\, f_i$，定义$I(x_0)=\{i\mid f_i(x_0)=f(x_0)\}$，则

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

- $I(x_0)$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, $\partial f(x_0)=\textbf{conv}\{\nabla f_i(x_0)\mid i\in I(x_0)\}$

例：分段线性函数

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



- 点x处的次微分是一个多面体

$$\partial f(x) = \mathbf{conv}\{a_i \mid i \in I(x)\}$$

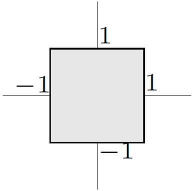
其中$I(x) = \{i \mid a_i^{\mathrm{T}} x + b_i = f(x)\}$

例: ℓ_1-范数

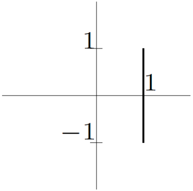
$$f(x) = \|x\|_1 = \max_{s \in \{-1,1\}^n} s^{\mathrm{T}} x$$

- 次微分是区间的乘积

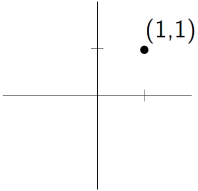
$$\partial f(x) = J_1 \times \cdots \times J_n, \quad J_k = \begin{cases} [-1, 1], & x_k = 0 \\ \{1\}, & x_k > 0 \\ \{-1\}, & x_k < 0 \end{cases}$$



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



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



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