

次梯度

文再文

北京大学北京国际数学研究中心

教材《最优化：建模、算法与理论》配套电子教案

<http://bicmr.pku.edu.cn/~wenzw/optbook.html>

致谢：本教案由朱楨源协助准备

- 1 次梯度的定义
- 2 次梯度的性质
- 3 凸函数的方向导数
- 4 次梯度的计算规则
- 5 对偶和最优性条件

一阶条件

回顾可微凸函数 f 的一阶条件:

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

- 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

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

则称 g 为函数 f 在点 x 处的一个次梯度。

- 进一步地，称集合

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

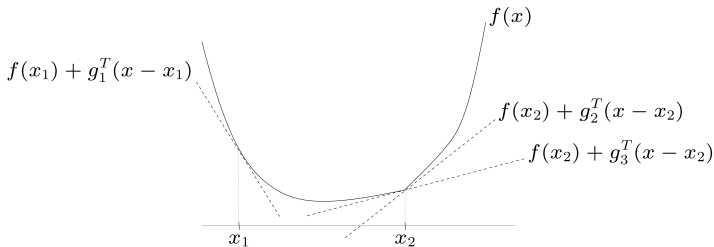
为 f 在点 x 处的次微分。

次梯度

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

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

- 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 $x \in \mathbf{int} \mathbf{dom} f$, 则 $\partial f(x)$ 是非空的, 其中 $\mathbf{int} \mathbf{dom} f$ 的含义是集合 $\mathbf{dom} f$ 的所有内点.

证明:

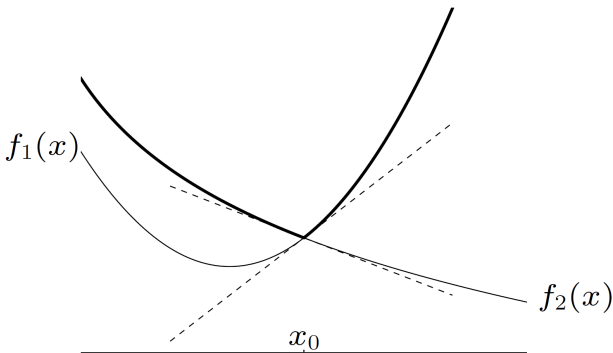
- $(x, f(x))$ 是 $\mathbf{epi} f$ 边界上的点
- 因此存在 $\mathbf{epi} f$ 在点 $(x, f(x))$ 处的支撑超平面:

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

- 令 $t \rightarrow +\infty$, 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

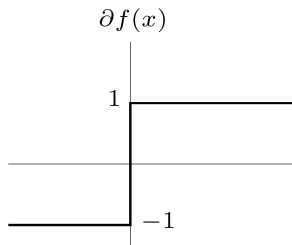
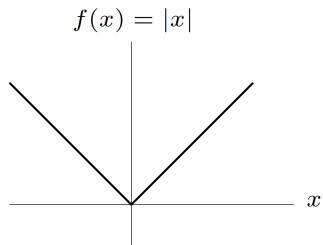
例

$f(x) = \max\{f_1(x), f_2(x)\}$ f_1, f_2 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 $f(x) = \|x\|_2$

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

1 次梯度的定义

2 次梯度的性质

3 凸函数的方向导数

4 次梯度的计算规则

5 对偶和最优性条件

次微分是闭凸集

对任何 $x \in \text{dom } f$, $\partial f(x)$ 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

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

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

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 $g_k \rightarrow g$, 则

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

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

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

这说明 $\partial f(x)$ 为闭集.

内点的次微分非空有界

如果 $x \in \text{int dom } f$, 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

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

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

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

- 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 $x_0 \in \text{int dom } f$ 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 $x_0 + tv \in \text{dom } f, t > 0$ 有

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

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

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

- 不等式两边令 $t \rightarrow 0$, 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 为凸函数, $x, y \in \mathbf{dom} f$, 则 $(u - v)^T(x - y) \geq 0$, 其中 $u \in \partial f(x)$, $v \in \partial f(y)$.

证明:

- 由次梯度的定义,

$$f(y) \geq f(x) + u^T(y - x),$$

$$f(x) \geq f(y) + v^T(x - y).$$

- 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 $x^k \rightarrow \bar{x}$, $g^k \in \partial f(x^k)$ 为 $f(x)$ 在点 x^k 处的次梯度, 且 $g^k \rightarrow \bar{g}$, 则 $\bar{g} \in \partial f(\bar{x})$.

证明:

- 对任意 $y \in \text{dom}f$, 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

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

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

例：非次可微函数

如下函数在点 $x = 0$ 处不是次可微的：

- $f : \mathbf{R} \rightarrow \mathbf{R}, \text{dom } f = \mathbf{R}_+$

$$x = 0 \text{ 时, } f(x) = 1, x > 0 \text{ 时, } f(x) = 0$$

- $f : \mathbf{R} \rightarrow \mathbf{R}, \text{dom } f = \mathbf{R}_+$

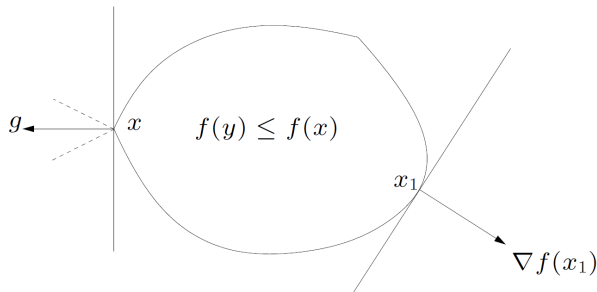
$$f(x) = -\sqrt{x}$$

$\text{epi } f$ 在点 $(0, f(0))$ 处的唯一支撑超平面是垂直的

次梯度和水平子集

如果 g 是 f 在点 x 处的一个次梯度, 那么

$$f(y) \leq f(x) \implies g^T(y - x) \leq 0$$



点 x 处的非零次梯度可定义水平子集 $\{y \mid f(y) \leq f(x)\}$ 的支撑超平面

- 1 次梯度的定义
- 2 次梯度的性质
- 3 凸函数的方向导数
- 4 次梯度的计算规则
- 5 对偶和最优性条件

方向导数

- 一般函数：设 f 为适当函数，给定点 x_0 以及方向 $d \in \mathbb{R}^n$ ，方向导数（若存在）定义为

$$\lim_{t \downarrow 0} \phi(t) = \lim_{t \downarrow 0} \frac{f(x_0 + td) - f(x_0)}{t},$$

其中 $t \downarrow 0$ 表示 t 单调下降趋于0.

- 凸函数：易知 $\phi(t)$ 在 $(0, +\infty)$ 上是单调不减的，上式中的极限号 \lim 可以替换为下确界 \inf . 上述此时极限总是存在（可以为无穷），进而凸函数总是可以定义方向导数.
- 方向导数的定义：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

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

方向导数有限

设 $f(x)$ 为凸函数, $x_0 \in \text{int dom } f$, 则对任意 $d \in \mathbb{R}^n$, $\partial f(x_0; d)$ 有限.

证明:

- 首先 $\partial f(x_0; d)$ 不为正无穷是显然的.
- 由于 $x_0 \in \text{int dom } f$, 根据次梯度的存在性定理可知 $f(x)$ 在点 x_0 处存在次梯度 g .
- 根据方向导数的定义, 我们有

$$\begin{aligned}\partial f(x_0; d) &= \inf_{t>0} \frac{f(x_0 + td) - f(x_0)}{t} \\ &\geq \inf_{t>0} \frac{tg^T d}{t} = g^T d.\end{aligned}$$

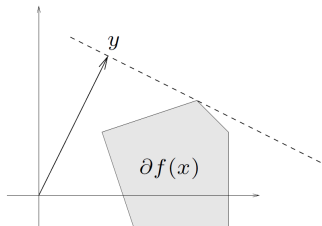
其中的不等式利用了次梯度的定义.

- 这说明 $\partial f(x_0; d)$ 不为负无穷.

方向导数和次梯度

设 $f: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 为凸函数, 点 $x_0 \in \text{int dom } f$, d 为 \mathbb{R}^n 中任一方向, 则

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



$\partial f(x; y)$ 是 $\partial f(x)$ 的支撑函数

- 对于可微函数, $\partial f(x_0; d) = \nabla f(x_0)^T d$
- 这也说明 $\partial f(x_0; d)$ 对所有的 $x_0 \in \text{int dom } f$, 以及所有的 d 都存在

证明：

- 记 $q(v) = \partial f(x_0; v)$. 根据方向导数有限性的证明过程可得,

$$q(d) = \partial f(x_0; d) \geq g^T d, \quad \forall g \in \partial f(x_0).$$

这说明 $\partial f(x_0; d)$ 是 $g^T d$ 的一个上界, 接下来说明该上界为上确界.

- 构造函数

$$h(v, t) = t \left(f \left(x_0 + \frac{v}{t} \right) - f(x_0) \right),$$

可知 $h(v, t)$ 为 $\tilde{f}(v) = f(x_0 + v) - f(x_0)$ 的透视函数, 并且

$$q(v) = \inf_{t' > 0} \frac{f(x_0 + t'v) - f(x_0)}{t'} \stackrel{t=1/t'}{=} \inf_{t > 0} h(v, t).$$

由于透视函数 $h(v, t)$ 为凸函数, 取下确界仍为凸函数, 因此 $q(v)$ 关于 v 是凸函数.

- 由方向导数的有限性直接可以得出 $\text{dom } q = \mathbb{R}^n$, 因此 $q(v)$ 在全空间任意一点次梯度存在.

- 对方向 d , 设 $\hat{g} \in \partial q(d)$, 则对任意 $v \in \mathbb{R}^n$ 以及 $\lambda \geq 0$, 我们有

$$\lambda q(v) = q(\lambda v) \geq q(d) + \hat{g}^T(\lambda v - d).$$

- 令 $\lambda = 0$, 我们有 $q(d) \leq \hat{g}^T d$; 令 $\lambda \rightarrow +\infty$, 我们有

$$q(v) \geq \hat{g}^T v,$$

- 进而推出

$$f(x + v) \geq f(x) + q(v) \geq f(x) + \hat{g}^T v.$$

这说明 $\hat{g} \in \partial f(x)$ 且 $\hat{g}^T d \geq q(d)$. 即 $q(d)$ 为 $g^T d$ 的上确界, 且当 $g = \hat{g}$ 时上确界达到.

更一般的推广

- 设 f 为适当凸函数，且在 x_0 处次微分不为空集，则对任意 $d \in \mathbb{R}^n$ 有

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

且当 $\partial f(x_0; d)$ 不为无穷时，上确界可以取到。

提纲

- 1 次梯度的定义
- 2 次梯度的性质
- 3 凸函数的方向导数
- 4 次梯度的计算规则
- 5 对偶和最优性条件

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 $x \in \text{int dom } f$

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 $\text{int dom } f_1 \cap \text{dom } f_2 \neq \emptyset$ ，而 $x \in \text{dom } f_1 \cap \text{dom } f_2$ 。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

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

两个函数之和的次梯度

设 $f_1, f_2 : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 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 $\text{int dom } f_1 \cap \text{dom } f_2 \neq \emptyset$, 则对任意的 $x_0 \in \mathbb{R}^n$,

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

证明:

- 第一个结论由次梯度的定义是显然的. 下面我们证第二个结论.
- 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

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

对任意 $x \in \mathbb{R}^n$ 成立, 故 $f_1 + f_2 \equiv +\infty$. 这与 $\text{int dom } f_1 \cap \text{dom } f_2 \neq \emptyset$ 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

$$S_1 = \{(x - x_0, y) \in \mathbb{R}^n \times \mathbb{R} \mid y > f_1(x) - f_1(x_0) - g^T(x - x_0)\},$$

$$S_2 = \{(x - x_0, y) \in \mathbb{R}^n \times \mathbb{R} \mid y \leq f_2(x_0) - f_2(x)\},$$

- 设 $(x - x_0, y) \in S_1 \cap S_2$ ，则

$$y > f_1(x) - f_1(x_0) - g^T(x - x_0),$$

$$y \leq f_2(x_0) - f_2(x).$$

上两式相减即得

$$(f_1 + f_2)(x) < (f_1 + f_2)(x_0) + g^T(x - x_0),$$

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

$$a^T(x - x_0) + by \leq c, \quad \forall (x - x_0, y) \in S_1, \quad (1)$$

$$a^T(x - x_0) + by \geq c, \quad \forall (x - x_0, y) \in S_2. \quad (2)$$

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

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

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

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

$$\hat{a}^T(x - x_0) \leq y, \quad \forall (x - x_0, y) \in S_1,$$

$$\hat{a}^T(x - x_0) \geq y, \quad \forall (x - x_0, y) \in S_2.$$

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

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

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

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

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

证明：

- 若 $f(x_0) = +\infty$ ，则 $f_i(x_0) = +\infty, i \in I(x_0)$ ，于是等式两端均为 \emptyset 。
- 下设 $f(x_0) < +\infty$ 。 $\forall i \in I(x_0)$ ，容易验证 $\partial f_i(x_0) \subseteq \partial f(x_0)$ 。再由次微分是闭凸集可知

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

- 另一方面，设 $g \in \partial f(x_0)$ 。假设 $g \notin \mathbf{conv} \bigcup_{i \in I(x_0)} \partial f_i(x_0)$ ，由严格分离定理（注意到 $\mathbf{conv} \bigcup_{i \in I(x_0)} \partial f_i(x_0)$ 和 $\{g\}$ 均为闭凸集）和方向导数与次梯度的关系，存在 $a \in \mathbb{R}^n$ 和 $b \in \mathbb{R}$ ，使得

$$a^T g > b \geq \max_{i \in I(x_0)} \sup_{\xi \in \partial f_i(x_0)} a^T \xi = \max_{i \in I(x_0)} \partial f_i(x_0; a).$$

- 因为

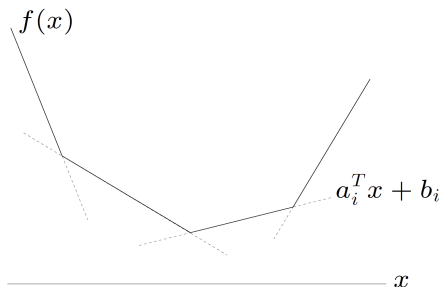
$$\begin{aligned}\partial f(x_0; a) &= \lim_{t \rightarrow 0^+} \frac{f(x_0 + ta) - f(x_0)}{t} \\ &= \max_{i \in I(x_0)} \lim_{t \rightarrow 0^+} \frac{f_i(x_0 + ta) - f_i(x_0)}{t} \\ &= \max_{i \in I(x_0)} \partial f_i(x_0; a).\end{aligned}$$

故 $a^T g > \partial f(x_0; a)$.

- 但由于 $g \in \partial f(x_0)$, 我们有 $f(x_0 + ta) \geq f(x_0) + tg^T a$, 因而 $\partial f(x_0; a) \geq a^T g$, 这就导致矛盾. 故 $g \in \mathbf{conv} \bigcup_{i \in I(x_0)} \partial f_i(x_0)$.

例：分段线性函数

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



- 点 x 处的次微分是一个多面体

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

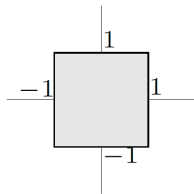
其中 $I(x) = \{i \mid a_i^T x + b_i = f(x)\}$

例: ℓ_1 -范数

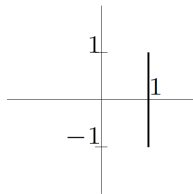
$$f(x) = \|x\|_1 = \max_{s \in \{-1, 1\}^n} s^T x$$

- 次微分是区间的乘积

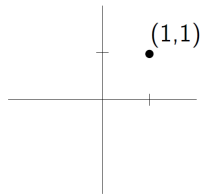
$$\partial f(x) = J_1 \times \cdots \times J_n, \quad J_k = \begin{cases} [-1, 1], & x_k = 0 \\ \{1\}, & x_k > 0 \\ \{-1\}, & x_k < 0 \end{cases}$$



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



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



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

逐点上确界函数

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

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

- 如果还有 \mathcal{A} 是紧集且 f_α 关于 α 连续，则

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

例：最大特征值函数

$A(x) = A_0 + x_1A_1 + \cdots + x_nA_n$ 并且系数 A_i 对称，令

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

计算点 \hat{x} 处的一个次梯度：

- 选择特征值 $\lambda_{\max}(A(\hat{x}))$ 对应的任一单位特征向量 y
- $y^T A(x) y$ 在点 \hat{x} 处的梯度是 f 的一个次梯度：

$$(y^T A_1 y, \cdots, y^T A_n y) \in \partial f(\hat{x})$$

固定分量的函数极小值

$$f(x) = \inf_y h(x, y), \quad h \text{ 关于 } (x, y) \text{ 联合凸}$$

计算点 \hat{x} 处的一个次梯度：

- 设 $\hat{y} \in \mathbb{R}^m$ 满足 $h(\hat{x}, \hat{y}) = f(\hat{x})$
- 存在 $g \in \mathbb{R}^n$ 使得 $(g, 0) \in \partial h(\hat{x}, \hat{y})$, 则 $g \in \partial f(\hat{x})$

证明：对任意 $x \in \mathbb{R}^n, y \in \mathbb{R}^m$

$$\begin{aligned} h(x, y) &\geq h(\hat{x}, \hat{y}) + g^T(x - \hat{x}) + 0^T(y - \hat{y}) \\ &= f(\hat{x}) + g^T(x - \hat{x}) \end{aligned}$$

于是

$$f(x) = \inf_y h(x, y) \geq f(\hat{x}) + g^T(x - \hat{x})$$

例: 距离函数

设 C 是 \mathbb{R}^n 中一闭凸集, 令

$$f(x) = \inf_{y \in C} \|x - y\|_2$$

计算点 \hat{x} 处的一个次梯度:

- 若 $f(\hat{x}) = 0$, 则容易验证 $g = 0 \in \partial f(\hat{x})$;
- 若 $f(\hat{x}) > 0$, 取 \hat{y} 为 \hat{x} 在 C 上的投影, 即 $\hat{y} = \mathcal{P}_C(\hat{x})$, 计算

$$g = \frac{1}{\|\hat{x} - \hat{y}\|_2}(\hat{x} - \hat{y}) = \frac{1}{\|\hat{x} - \mathcal{P}_C(\hat{x})\|_2}(\hat{x} - \mathcal{P}_C(\hat{x}))$$

复合函数

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

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

计算点 \hat{x} 处的一个次梯度:

- $z = (z_1, z_2, \dots, z_m) \in \partial h(f_1(\hat{x}), f_2(\hat{x}), \dots, f_m(\hat{x}))$ 以及 $g_i \in \partial f_i(\hat{x})$
- $g \stackrel{\text{def}}{=} z_1 g_1 + z_2 g_2 + \dots + z_m g_m \in \partial f(\hat{x})$

证明:

$$\begin{aligned} f(x) &\geq h(f_1(\hat{x}) + g_1^T(x - \hat{x}), f_2(\hat{x}) + g_2^T(x - \hat{x}), \dots, f_m(\hat{x}) + g_m^T(x - \hat{x})) \\ &\geq h(f_1(\hat{x}), f_2(\hat{x}), \dots, f_m(\hat{x})) + \sum_{i=1}^m z_i g_i^T(x - \hat{x}) \\ &= f(\hat{x}) + g^T(x - \hat{x}), \end{aligned}$$

最优值函数

设函数 f_i 是凸函数, 定义 $h(u, v)$ 为如下凸问题的最优值

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq u_i, \quad i = 1, \dots, m \\ & Ax = b + v \end{aligned}$$

计算点 (\hat{u}, \hat{v}) 处的一个次梯度:

- 假设 $h(\hat{u}, \hat{v})$ 有限, 强对偶成立

$$\begin{aligned} \max \quad & \inf_x \left(f_0(x) + \sum_i \lambda_i (f_i(x) - \hat{u}_i) + \nu^T (Ax - b - \hat{v}) \right) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

- 如果 $\hat{\lambda}, \hat{\nu}$ 是最优对偶变量, 那么 $(-\hat{\lambda}, -\hat{\nu}) \in \partial h(\hat{u}, \hat{v})$

证明:由弱对偶原理可得

$$\begin{aligned} h(u, v) &\geq \inf_x \left(f_0(x) + \sum_i \hat{\lambda}_i (f_i(x) - u_i) + \hat{v}^T (Ax - b - v) \right) \\ &= \inf_x \left(f_0(x) + \sum_i \hat{\lambda}_i (f_i(x) - \hat{u}_i) + \hat{v}^T (Ax - b - \hat{v}) \right) \\ &\quad - \hat{\lambda}^T (u - \hat{u}) - \hat{v}^T (v - \hat{v}) \\ &= h(\hat{u}, \hat{v}) - \hat{\lambda}^T (u - \hat{u}) - \hat{v}^T (v - \hat{v}) \end{aligned}$$

期望函数

u 是一个随机变量, h 是关于 x 的凸函数, 令

$$f(x) = \mathbb{E}h(x, u)$$

计算点 \hat{x} 处的一个次梯度:

- 选择一个函数 g 满足 $g(u) \in \partial_x h(\hat{x}, u)$
- $g = \mathbb{E}_u g(u) \in \partial f(\hat{x})$

证明: 由 h 的凸性和 $g(u)$ 的定义,

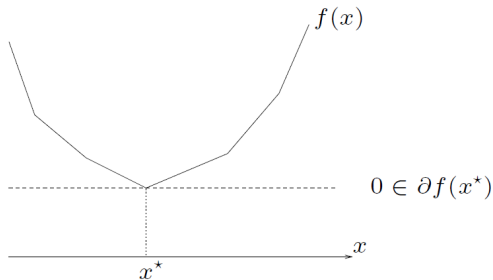
$$\begin{aligned} f(x) &= \mathbb{E}h(x, u) \\ &\geq \mathbb{E} (h(\hat{x}, u) + g(u)^T(x - \hat{x})) \\ &= f(\hat{x}) + g^T(x - \hat{x}) \end{aligned}$$

- 1 次梯度的定义
- 2 次梯度的性质
- 3 凸函数的方向导数
- 4 次梯度的计算规则
- 5 对偶和最优性条件

最优性条件：无约束问题

x^* 是 $f(x)$ 的极小点当且仅当

$$0 \in \partial f(x^*)$$



证明：根据定义

$$f(y) \geq f(x^*) + 0^T(y - x^*), \quad \forall y \quad \Leftrightarrow \quad 0 \in \partial f(x^*)$$

例: 分片线性极小

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

- 最优性条件

$$0 \in \mathbf{conv}\{a_i \mid i \in I(x^*)\}, \quad \text{其中 } I(x) = \{i \mid a_i^T x + b_i = f(x)\}$$

- 也就是说, x^* 是最优解当且仅当存在 λ 使得

$$\lambda \geq 0, \quad \mathbf{1}^T \lambda = 1, \quad \sum_{i=1}^m \lambda_i a_i = 0, \quad \lambda_i = 0 \text{ for } i \notin I(x^*)$$

- 这是等价线性规划问题的最优性条件

$$\begin{array}{ll} \min & t \\ \text{s.t.} & Ax + b \leq t\mathbf{1} \end{array} \qquad \begin{array}{ll} \max & b^T \lambda \\ \text{s.t.} & A^T \lambda = 0 \\ & \lambda \geq 0, \quad \mathbf{1}^T \lambda = 1 \end{array}$$

最优性条件：约束问题

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m \end{aligned}$$

如果强对偶成立, 那么 x^*, λ^* 是最优原始、对偶变量当且仅当

- 1 x^* 是可行的
- 2 $\lambda^* \geq 0$
- 3 $\lambda_i^* f_i(x^*) = 0, i = 1, \dots, m$
- 4 x^* 是下式的一个极小值点

$$L(x, \lambda^*) = f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x)$$

- 当 $\text{dom } f_i = \mathbf{R}^n$ 时, Karush-Kuhn-Tucker 条件为条件 1, 2, 3 以及

$$0 \in \partial L_x(x^*, \lambda^*) = \partial f_0(x^*) + \sum_{i=1}^m \lambda_i^* \partial f_i(x^*).$$

- 对于可微函数 f_i , 上式成为

$$0 = \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*)$$