

最优化方法 第八讲

Weiwen Wang(王伟文)

暨南大学

2025 年 5 月 5 日

目录

- 1 次梯度
 - 次梯度定义
 - 凸函数次梯度存在性
 - 凸函数次梯度与方向导数
 - 次梯度计算

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定义 1

称向量 $\mathbf{g}$ 为函数 f 在点 $\mathbf{x}_0 \in \text{dom } f$ 的次梯度, 若对任意 $\mathbf{y} \in \text{dom } f$ 有

$$f(\mathbf{y}) \geq f(\mathbf{x}_0) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x}_0 \rangle.$$

函数 f 在点 $\mathbf{x}_0 \in \text{dom } f$ 的次梯度集合记为 $\partial f(\mathbf{x}_0)$, 称为 f 在点 $\mathbf{x}_0$ 处的次微分.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

考虑函数 $f(x) = \max\{0, x\}, x \in \mathbb{R}$. 对任意 $y \in \mathbb{R}$ 和 $g \in [0, 1]$,

$$f(y) = \max\{0, y\} \geq f(0) + g \cdot y = g \cdot y.$$

函数 f 在 0 处次梯度不唯一, 有 $\partial f(0) = [0, 1]$.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定义 2 (对偶范数)

给定 $\mathbb{R}^d$ 上的范数 $\|\cdot\|$, 其对偶范数定义为

$$\|\mathbf{y}\|_* = \max_{\mathbf{x}} \{\langle \mathbf{x}, \mathbf{y} \rangle : \|\mathbf{x}\| \leq 1\} \quad \forall \mathbf{y} \in \mathbb{R}^d.$$

- 广义 Cauchy-Schwartz 不等式

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{y}\|_* \|\mathbf{x}\|$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

设 $f(x) = \|x\|, x \in \mathbb{R}^d$, 则

$$\partial f(\mathbf{0}) = \{\mathbf{g} \in \mathbb{R}^d : \|\mathbf{g}\|_* \leq 1\}$$

$\mathbf{g} \in \partial f(\mathbf{0})$ 当且仅当

$$\|\mathbf{y}\| \geq \langle \mathbf{g}, \mathbf{y} \rangle \quad \forall \mathbf{y} \in \mathbb{R}^d.$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

$\mathbf{g} \in \partial f(\mathbf{0})$ 当且仅当

$$\|\mathbf{y}\| \geq \langle \mathbf{g}, \mathbf{y} \rangle \quad \forall \mathbf{y} \in \mathbb{R}^d.$$

若 $\mathbf{y} = \mathbf{0}$, 上式对任意 $\mathbf{g} \in \mathbb{R}^d$ 成立.

考虑 $\mathbf{y} \neq \mathbf{0}$, 上述不等式等价于

$$\langle \mathbf{g}, \frac{\mathbf{y}}{\|\mathbf{y}\|} \rangle \leq 1 \quad \forall \mathbf{y} \in \mathbb{R}^d \setminus \{\mathbf{0}\}.$$

不等式两端同时关于 $\mathbf{y}$ 取最大值, 由对偶范数定义可得

$$\|\mathbf{g}\|_* = \max_{\mathbf{y} \in \mathbb{R}^d \setminus \{\mathbf{0}\}} \langle \mathbf{g}, \frac{\mathbf{y}}{\|\mathbf{y}\|} \rangle \leq 1.$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

设 $f(\mathbf{x}) = \|\mathbf{x}\|_1, \forall \mathbf{x} \in \mathbb{R}^d$,

$$\partial f(\mathbf{0}) = \{\mathbf{g} : \|\mathbf{g}\|_\infty \leq 1\} = [-1, 1]^d.$$

特别地, 若 $f(x) = |x|$, 则

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

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定理 3

对任意 $\mathbf{x} \in \text{dom } f$, $\partial f(\mathbf{x})$ 为闭凸集,

$$\partial f(\mathbf{x}) = \bigcap_{\mathbf{y} \in \text{dom } f} \{\mathbf{g} : f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle\}$$

即半空间 $\{\mathbf{g} : f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle\}$ 的交集.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定理 4 (支撑超平面定理)

设 $\mathcal{C} \subset \mathbb{R}^d$ 为非空凸集, 若 $\mathbf{y} \notin \mathcal{C}$, 则存在非零向量 $\mathbf{p}$ 使得

$$\langle \mathbf{x}, \mathbf{p} \rangle \leq \langle \mathbf{y}, \mathbf{p} \rangle \quad \forall \mathbf{x} \in \mathcal{C}.$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定理 5

设 f 为凸函数, 若 $\mathbf{x}_0 \in \text{int}(\text{dom } f)$, 则 $\partial f(\mathbf{x}_0)$ 为非空有界集.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

证明.

因为 f 为凸函数, 故 $\text{epi } f$ 为凸集, 且 $(\mathbf{x}_0, f(\mathbf{x}_0)) \in \text{bd}(\text{epi } f)$. 由支撑超平面定理知存在非零向量 $(\mathbf{p}, -\alpha) \in \mathbb{R}^d \times \mathbb{R}$ 使得

$$\langle \mathbf{x}_0, \mathbf{p} \rangle - \alpha f(\mathbf{x}_0) \geq \langle \mathbf{y}, \mathbf{p} \rangle - \alpha \tau \quad \forall (\mathbf{y}, \tau) \in \text{epi } f.$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

证明.

因为 f 为凸函数, 故 $\text{epi } f$ 为凸集, 且 $(\mathbf{x}_0, f(\mathbf{x}_0)) \in \text{bd}(\text{epi } f)$. 由支撑超平面定理知存在非零向量 $(\mathbf{p}, -\alpha) \in \mathbb{R}^d \times \mathbb{R}$ 使得

$$\langle \mathbf{x}_0, \mathbf{p} \rangle - \alpha f(\mathbf{x}_0) \geq \langle \mathbf{y}, \mathbf{p} \rangle - \alpha \tau \quad \forall (\mathbf{y}, \tau) \in \text{epi } f.$$

若取 $\tau = f(\mathbf{y})$, 且有 $\alpha > 0$, 上式可以整理得到

$$f(\mathbf{y}) \geq f(\mathbf{x}_0) + \langle \frac{\mathbf{p}}{\alpha}, \mathbf{y} - \mathbf{x}_0 \rangle.$$

此时有 $\frac{\mathbf{p}}{\alpha} \in \partial f(\mathbf{x}_0)$. 故下证 $\alpha > 0$. □

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

取 $\mathbf{y} = \mathbf{x}_0$, $\tau = f(\mathbf{x}_0) + 1$,

$$\langle \mathbf{x}_0, \mathbf{p} \rangle - \alpha f(\mathbf{x}_0) \geq \langle \mathbf{x}_0, \mathbf{p} \rangle - \alpha(f(\mathbf{x}_0) + 1)$$

解得 $\alpha \geq 0$.若 $\mathbf{p} = 0$, 则 $\alpha > 0$, 由上述讨论知 $\frac{\mathbf{p}}{\alpha} = 0 \in \partial f(\mathbf{x}_0)$, 故考虑 $\mathbf{p} \neq 0$ 的情形.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

取 $\mathbf{y} = \mathbf{x}_0$, $\tau = f(\mathbf{x}_0) + 1$,

$$\langle \mathbf{x}_0, \mathbf{p} \rangle - \alpha f(\mathbf{x}_0) \geq \langle \mathbf{x}_0, \mathbf{p} \rangle - \alpha(f(\mathbf{x}_0) + 1)$$

解得 $\alpha \geq 0$.若 $\mathbf{p} = 0$, 则 $\alpha > 0$, 由上述讨论知 $\frac{\mathbf{p}}{\alpha} = 0 \in \partial f(\mathbf{x}_0)$, 故考虑 $\mathbf{p} \neq 0$ 的情形.由凸函数的局部 Lipschitz 连续性, 存在 $\epsilon > 0$ 及 $M_\epsilon > 0$ 使得

$$f(\mathbf{z}) - f(\mathbf{x}_0) \leq M_\epsilon \|\mathbf{x} - \mathbf{z}\| \quad \forall \mathbf{z} \in \mathcal{B}_{\|\cdot\|}(\mathbf{x}_0, \epsilon)$$

又因为 $(\mathbf{z}, f(\mathbf{z})) \in \text{epi } f, \forall \mathbf{z} \in \mathcal{B}_{\|\cdot\|}(\mathbf{x}_0, \epsilon)$, 故

$$\langle \mathbf{p}, \mathbf{z} - \mathbf{x}_0 \rangle \leq \alpha(f(\mathbf{z}) - f(\mathbf{x}_0)) \leq \alpha M_\epsilon \|\mathbf{x}_0 - \mathbf{z}\| \quad \forall \mathbf{z} \in \mathcal{B}_{\|\cdot\|}(\mathbf{x}_0, \epsilon)$$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

$$\langle \mathbf{p}, \mathbf{z} - \mathbf{x}_0 \rangle \leq \alpha(f(\mathbf{z}) - f(\mathbf{x}_0)) \leq \alpha M_\epsilon \|\mathbf{x}_0 - \mathbf{z}\| \quad \forall \mathbf{z} \in \mathcal{B}_{\|\cdot\|}(\mathbf{x}_0, \epsilon)$$

取 $\mathbf{z} = \mathbf{x}_0 + \epsilon \frac{\mathbf{p}}{\|\mathbf{p}\|}$ 代入上式得

$$0 < \epsilon \|\mathbf{p}\| \leq \alpha M_\epsilon \epsilon$$

故有 $\alpha > 0$, 由前述讨论知 $\frac{\mathbf{p}}{\alpha} \in \partial f(\mathbf{x}_0)$.

综上所述, $\frac{\mathbf{p}}{\alpha} \in \partial f(\mathbf{x}_0)$, 故 $\partial f(\mathbf{x}_0)$ 非空.



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

下证有界性, 对任意 $\mathbf{g} \in \partial f(\mathbf{x}_0)$, 结合次梯度定义及凸函数局部 Lipschitz 连续性,

$$\langle \mathbf{g}, \mathbf{z} - \mathbf{x}_0 \rangle \leq f(\mathbf{z}) - f(\mathbf{x}_0) \leq M_\epsilon \|\mathbf{z} - \mathbf{x}_0\| \quad \forall \mathbf{z} \in \mathcal{B}_{\|\cdot\|}(\mathbf{x}_0, \epsilon) \setminus \{\mathbf{x}_0\}.$$

故

$$\left\langle \mathbf{g}, \frac{\mathbf{z} - \mathbf{x}_0}{\|\mathbf{z} - \mathbf{x}_0\|} \right\rangle \leq M_\epsilon \quad \forall \mathbf{z} \in \mathcal{B}_{\|\cdot\|}(\mathbf{x}_0, \epsilon) \setminus \{\mathbf{x}_0\}.$$

即 $\langle \mathbf{g}, \mathbf{h} \rangle \leq 1, \forall \|\mathbf{h}\| = 1$, 由对偶范数定义知

$$\|\mathbf{g}\|_* \leq M_\epsilon.$$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定理 6

对于函数 f , 定义 *Fenchel* 对偶

$$f_*(\mathbf{s}) = \sup_{\mathbf{y} \in \text{dom } f} \{\langle \mathbf{s}, \mathbf{y} \rangle - f(\mathbf{y})\}$$

及 *Fenchel* 对偶的对偶

$$f_{**}(\mathbf{x}) = \sup_{\mathbf{s} \in \text{dom } f_*} \{\langle \mathbf{x}, \mathbf{s} \rangle - f_*(\mathbf{s})\}$$

则对于任意 $\mathbf{x} \in \text{dom } f$, 有 $f(\mathbf{x}) \geq f_{**}(\mathbf{x})$; 且若存在 $\mathbf{x} \in \text{dom } f$ 满足 $\partial f(\mathbf{x}) \neq \emptyset$, 则 $\partial f(\mathbf{x}) \subseteq \text{dom } f_*$ 且 $f(\mathbf{x}) = f_{**}(\mathbf{x})$.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

证明.

由 Fenchel 对偶定义, $\forall \mathbf{x} \in \text{dom } f$ 及 $\forall \mathbf{s} \in \text{dom } f_*$,

$$f_*(\mathbf{s}) \geq \langle \mathbf{s}, \mathbf{x} \rangle - f(\mathbf{x})$$

故

$$\langle \mathbf{s}, \mathbf{x} \rangle - f_*(\mathbf{s}) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \text{dom } f, \forall \mathbf{s} \in \text{dom } f_*$$

不等式左右两端同时关于 $\mathbf{s} \in \text{dom } f_*$ 取最大值即可得到

$$f_{**}(\mathbf{x}) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \text{dom } f.$$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

对 $\mathbf{x} \in \text{dom } f$, 若 $\mathbf{g} \in \partial f(\mathbf{x})$, 由次梯度定义

$$f(\mathbf{y}) \geq f(\mathbf{x}) - \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \quad \forall \mathbf{y} \in \text{dom } f.$$

整理得到

$$\langle \mathbf{g}, \mathbf{y} \rangle - f(\mathbf{y}) \leq \langle \mathbf{g}, \mathbf{x} \rangle - f(\mathbf{x}) \quad \forall \mathbf{y} \in \text{dom } f.$$

故 $f_*(\mathbf{g}) \leq \langle \mathbf{g}, \mathbf{x} \rangle - f(\mathbf{x}) < \infty$, 即 $\mathbf{g} \in \text{dom } f_*$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

已知 $f(\mathbf{x}) \geq f_{**}(\mathbf{x})$, 要证 $f(\mathbf{x}) = f_{**}(\mathbf{x})$ 只需再证 $f_{**}(\mathbf{x}) \geq f(\mathbf{x})$.

有定义知

$$f_{**}(\mathbf{x}) \geq \langle \mathbf{x}, \mathbf{s} \rangle - f_*(\mathbf{s}) \quad \forall \mathbf{s} \in \text{dom } f_*.$$

若 $\mathbf{g} \in \partial f(\mathbf{x})$, 则 $\mathbf{g} \in \text{dom } f_*$, 由前述讨论

$$f_*(\mathbf{g}) \leq \langle \mathbf{g}, \mathbf{x} \rangle - f(\mathbf{x}).$$

取 $\mathbf{s} = \mathbf{g}$, 结合上述不等式即可得到

$$f_{**}(\mathbf{x}) \geq f(\mathbf{x}).$$

综上所述 $f(\mathbf{x}) = f_{**}(\mathbf{x})$.



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

引理 7

设 f 是凸函数, $\mathbf{x}_0 \in \text{int}(\text{dom } f)$, 则 f 在点 $\mathbf{x}$ 处的方向导数 $f'(\mathbf{x}, \cdot)$ 是凸函数, 且有

$$f(\mathbf{y}) \geq f(\mathbf{x}) + f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) \quad \forall \mathbf{y} \in \text{dom } f.$$

$$f'(\mathbf{x}_0, \tau \mathbf{p}) = \tau f'(\mathbf{x}_0, \mathbf{p}) \quad \forall \tau > 0$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定理 8

设 f 是凸函数, $\mathbf{x}_0 \in \text{int}(\text{dom } f)$, 则

$$\partial_2 f'(\mathbf{x}_0; \mathbf{0}) = \partial f(\mathbf{x}_0)$$

其中 ∂_2 为关于第二个参量的次微分, 且对任意 $\mathbf{p} \in \mathbb{R}^d$,

$$f'(\mathbf{x}_0; \mathbf{p}) = \max\{\langle \mathbf{g}, \mathbf{p} \rangle : \mathbf{g} \in \partial f(\mathbf{x}_0)\}$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

证明.

1. 由方向导数定义及 $\mathbf{g} \in \partial f(\mathbf{x}_0)$

$$f'(\mathbf{x}_0; \mathbf{p}) = \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x}_0 + \alpha \mathbf{p}) - f(\mathbf{x}_0)}{\alpha} \geq \langle \mathbf{g}, \mathbf{p} \rangle.$$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

因为 $f'(\mathbf{x}_0; \mathbf{0}) = 0$, 故 $\mathbf{g} \in \partial_2 f'(\mathbf{x}_0; \mathbf{0})$, 即 $\partial f(\mathbf{x}_0) \subseteq \partial_2 f'(\mathbf{x}_0; \mathbf{0})$.

由引理 6, $\forall \mathbf{y} \in \text{dom } f$

$$f(\mathbf{y}) \geq f(\mathbf{x}_0) + f'(\mathbf{x}_0, \mathbf{y} - \mathbf{x}_0) \geq f(\mathbf{x}_0) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x}_0 \rangle \quad \forall \mathbf{g} \in \partial_2 f'(\mathbf{x}_0; \mathbf{0}).$$

故 $\mathbf{g} \in \partial f(\mathbf{x}_0)$.

综上所述 $\partial f(\mathbf{x}_0) = \partial_2 f'(\mathbf{x}_0; \mathbf{0})$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

2. 因为 $f'(\mathbf{x}_0, \cdot)$ 为凸函数且定义域为 $\mathbb{R}^d$, 故对任意 $\mathbf{p} \in \mathbb{R}^d$, $\partial_2 f'(\mathbf{x}_0; \mathbf{p})$ 非空. 设 $\mathbf{h} \in \partial_2 f'(\mathbf{x}_0; \mathbf{p})$, 取任意 $\tau > 0$ 及 $\mathbf{v} \in \mathbb{R}^d$, 由引理 6 及次梯度定义

$$\tau f'(\mathbf{x}_0; \mathbf{v}) = f'(\mathbf{x}_0; \tau \mathbf{v}) \geq f'(\mathbf{x}_0; \mathbf{p}) + \langle \mathbf{h}, \tau \mathbf{v} - \mathbf{p} \rangle$$

不等式两端同时除以 τ , 取 $\tau \rightarrow \infty$

$$f'(\mathbf{x}_0; \mathbf{v}) \geq \langle \mathbf{h}, \mathbf{v} \rangle \quad \forall \mathbf{v} \in \mathbb{R}^d.$$

这表明 $\mathbf{h} \in \partial_2 f'(\mathbf{x}_0; \mathbf{0}) = \partial f(\mathbf{x}_0)$.

取 $\tau = 0$, 可整理得到

$$f'(\mathbf{x}_0; \mathbf{p}) \leq \langle \mathbf{h}, \mathbf{p} \rangle \leq f'(\mathbf{x}_0; \mathbf{p}).$$

因此 $f'(\mathbf{x}_0; \mathbf{p}) = \max\{\langle \mathbf{g}, \mathbf{p} \rangle : \mathbf{g} \in \partial f(\mathbf{x}_0)\}$.



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

引理 9

设 f 为凸函数, 若 f 在 $\mathbf{x}_0 \in \text{int}(\text{dom } f)$ 处可微, 则 $\partial f(\mathbf{x}_0) = \{\nabla f(\mathbf{x}_0)\}$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

证明.

对任意 $\mathbf{p} \in \mathbb{R}^d$, 由梯度与方向导数的关系及定理 7

$$f'(\mathbf{x}_0; \mathbf{p}) = \langle \nabla f(\mathbf{x}_0), \mathbf{p} \rangle \geq \langle \mathbf{g}, \mathbf{p} \rangle \quad \forall \mathbf{g} \in \partial f(\mathbf{x}_0).$$

取 $\mathbf{p} = \mathbf{g} - \nabla f(\mathbf{x}_0)$ 代入上式整理得到

$$\|\nabla f(\mathbf{x}_0) - \mathbf{g}\|^2 \leq 0 \quad \forall \mathbf{g} \in \partial f(\mathbf{x}_0).$$

故 $\mathbf{g} = \nabla f(\mathbf{x}_0)$.



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

引理 10

设 f 为凸函数, $\mathbf{x}_0 \in \text{int}(\text{dom } f)$, 若存在向量函数 $g(\mathbf{x}) \in \partial f(\mathbf{x}_0)$ 在 $\mathbf{x}_0$ 处连续, 则 f 在 $\mathbf{x}_0$ 处可微且 $\nabla f(\mathbf{x}_0) = g(\mathbf{x}_0)$.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

证明.

要证 $\nabla f(\mathbf{x}_0) = g(\mathbf{x}_0)$, 则证

$$f'(\mathbf{x}_0; \mathbf{h}) = \langle g(\mathbf{x}_0), \mathbf{h} \rangle \quad \forall \mathbf{h} \in \mathbb{R}^d.$$

由定理 7, 对任意 $\mathbf{h} \in \mathbb{R}^d$, 存在充分小的 $\alpha_0 > 0$, 使得

$$\langle g(\mathbf{x}_0), \mathbf{h} \rangle \leq \frac{1}{\alpha} [f(\mathbf{x}_0 + \alpha \mathbf{h}) - f(\mathbf{x}_0)] \quad \forall \alpha \in (0, \alpha_0)$$

由因为, $g(\mathbf{x}_0 + \alpha \mathbf{h}) \in \partial f(\mathbf{x}_0 + \alpha \mathbf{h})$, 故

$$\frac{1}{\alpha} [f(\mathbf{x}_0 + \alpha \mathbf{h}) - f(\mathbf{x}_0)] \leq \langle g(\mathbf{x}_0 + \alpha \mathbf{h}), \mathbf{h} \rangle$$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

故 $\forall \mathbf{h} \in \mathbb{R}^d$,

$$\langle g(\mathbf{x}_0), \mathbf{h} \rangle \leq \frac{1}{\alpha} [f(\mathbf{x}_0 + \alpha \mathbf{h}) - f(\mathbf{x}_0)] \leq \langle g(\mathbf{x}_0 + \alpha \mathbf{h}), \mathbf{h} \rangle \quad \forall \alpha \in (0, \alpha_0)$$

取 $\alpha \rightarrow 0$,

$$f'(\mathbf{x}_0; \mathbf{h}) = \langle g(\mathbf{x}_0), \mathbf{h} \rangle \quad \forall \mathbf{h} \in \mathbb{R}^d.$$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

命题 11

设 f 为凸函数, $\mathbf{x}_0 \in \text{int}(\text{dom } f)$, 若 f 在 $\mathbf{x}_0$ 处有唯一次梯度 $g(\mathbf{x}_0)$, 则 f 在 $\mathbf{x}_0$ 处可微且 $\nabla f(\mathbf{x}_0) = g(\mathbf{x}_0)$.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

设 $f(\mathbf{x}) = \|\mathbf{x}\|_2$, 若 $\mathbf{x} = \mathbf{0}$,

$$\partial f(\mathbf{0}) = \{\mathbf{g} \in \mathbb{R}^d : \|\mathbf{g}\|_2 \leq 1\}.$$

若 $\mathbf{x} \neq \mathbf{0}$, f 可微,

$$\nabla f(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}$$

综上

$$\partial f(\mathbf{x}) = \begin{cases} \left\{ \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \right\}, & \mathbf{x} \neq \mathbf{0} \\ \{\mathbf{g} \in \mathbb{R}^d : \|\mathbf{g}\|_2 \leq 1\}, & \mathbf{x} = \mathbf{0} \end{cases}$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定义 12 (支撑函数)

设 $C \subseteq \mathbb{R}^d$ 为非空集合, 定义 C 上的支撑函数为

$$\sigma_C(\mathbf{x}) = \max_{\mathbf{y} \in C} \langle \mathbf{x}, \mathbf{y} \rangle.$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

引理 13

若 $A, B \subseteq \mathbb{R}^d$ 为非空闭凸集, 则 $A = B$ 当且仅当 $\sigma_A = \sigma_B$.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定理 14

设 f 为凸函数, 定义线性算子

$$\mathcal{A}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}: \quad \mathbb{R}^d \rightarrow \mathbb{R}^p$$

则 $\varphi(\mathbf{x}) = f(\mathcal{A}(\mathbf{x}))$ 是凸函数. 如果 $\text{dom } \varphi$ 非空, 对 $\mathbf{x} \in \text{int}(\text{dom } \varphi)$ 满足 $\mathcal{A}(\mathbf{x}) \in \text{int}(\text{dom } f)$, 则

$$\partial\varphi(\mathbf{x}) = \mathbf{A}^T \partial f(\mathcal{A}(\mathbf{x})).$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

证明.

2. $\partial f(\mathcal{A}(\mathbf{x}))$ 非空有界集,

$$\begin{aligned}
 \varphi'(\mathbf{x}; \mathbf{h}) &= \lim_{\alpha \rightarrow 0^+} \frac{\varphi(\mathbf{x} + \alpha \mathbf{h}) - \varphi(\mathbf{x})}{\alpha} \\
 &= \lim_{\alpha \rightarrow 0^+} \frac{f[\mathcal{A}(\mathbf{x} + \alpha \mathbf{h})] - f(\mathcal{A}(\mathbf{x}))}{\alpha} \\
 &= \lim_{\alpha \rightarrow 0^+} \frac{f[\mathcal{A}(\mathbf{x}) + \alpha \mathbf{A}\mathbf{h}] - f(\mathcal{A}(\mathbf{x}))}{\alpha} \\
 &= f'(\mathcal{A}(\mathbf{x}); \mathbf{A}\mathbf{h}) \\
 &= \max_{\mathbf{g}} \{ \langle \mathbf{g}, \mathbf{A}\mathbf{h} \rangle : \mathbf{g} \in \partial f(\mathcal{A}(\mathbf{x})) \} \\
 &= \max_{\mathbf{g}} \{ \langle \mathbf{g}, \mathbf{h} \rangle : \mathbf{g} \in \mathbf{A}^T \partial f(\mathcal{A}(\mathbf{x})) \}
 \end{aligned}$$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

续.

 $\partial\varphi(\mathbf{x})$ 非空有界集, 且

$$\varphi'(\mathbf{x}; \mathbf{h}) = \max_{\mathbf{g}} \{\langle \mathbf{g}, \mathbf{h} \rangle : \mathbf{g} \in \partial\varphi(\mathbf{x})\}.$$

又因为 $\partial\varphi(\mathbf{x}), \partial f(\mathcal{A}(\mathbf{x}))$ 均为闭凸集, 故

$$\partial\varphi(\mathbf{x}) = \mathbf{A}^T \partial f(\mathcal{A}(\mathbf{x})).$$



次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

引理 15

f_1, f_2 均为闭凸函数, 若 $\alpha_1, \alpha_2 \geq 0$, 则 $\alpha_1 f_1 + \alpha_2 f_2$ 亦为闭凸函数.
对任意 $\mathbf{x} \in \text{int}(\text{dom } f) = \text{int}(\text{dom } f_1) \cap \text{int}(\text{dom } f_2)$,

$$\partial f(\mathbf{x}) = \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x}).$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

定理 16

设 $f_1, f_2, \dots, f_m$ 为凸函数, $\text{dom } f_i$ 非空. 定义

$$f(\mathbf{x}) = \max_{i \in [m]} f_i(\mathbf{x}),$$

若 $\mathbf{x} \in \cap \text{int}(\text{dom } f)$, 则

$$\partial f(\mathbf{x}) = \text{conv} \left(\bigcup_{i \in I(\mathbf{x})} \partial f_i(\mathbf{x}) \right),$$

其中 $I(\mathbf{x}) = \{i \in [m] : f_i(\mathbf{x}) = f(\mathbf{x})\}$.

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

$$f(\mathbf{x}) = \|\mathbf{x}\|_{\infty} = \max_{i \in [d]} |x_i| = \max_{i \in [d]} f_i(\mathbf{x}), \quad f_i(\mathbf{x}) = |x_i|.$$

$$\partial f(\mathbf{x}) \begin{cases} \{\mathbf{g} : \|\mathbf{g}\|_1 \leq 1\}, \mathbf{x} = \mathbf{0} \\ \left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \operatorname{sgn}(x_i) \mathbf{e}_i : \sum_{i \in I(\mathbf{x})} \lambda_i = 1, \lambda_i \geq 0 \quad \forall i \in I(\mathbf{x}) \right\}, \mathbf{x} \neq \mathbf{0} \end{cases}$$

次梯度

次梯度定义

凸函数次梯度存在性

凸函数次梯度与方向导数

次梯度计算

$$f(\mathbf{x}) = \max_{i \in [m]} \mathbf{a}_i^T \mathbf{x} + \mathbf{b}_i$$

$$\partial f(\mathbf{x}) = \left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \mathbf{a}_i : \sum_{i \in I(\mathbf{x})} \lambda_i = 1, \lambda_i \geq 0 \ \forall i \in I(\mathbf{x}) \right\}$$

$$I(\mathbf{x}) = \{i : f(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} + \mathbf{b}_i\}.$$