次梯度

1 次梯度与次微分

定义 1. $\mathbf{g}$ 是凸函数 f 在 $\mathbf{x} \in \mathrm{dom}\ f$ 处的次梯度，如果满足

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

定义 2. $f(\mathbf{x})$ 的次微分 $\partial f(\mathbf{x})$ 是所有次梯度的集合，即

$$\partial f(\mathbf{x}) = \{\mathbf{g} | \mathbf{g}^T(\mathbf{y}-\mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x}), \forall \mathbf{y} \in \mathrm{dom} f\}$$

性质 1. $\partial f(\mathbf{x})$ 是一个闭凸集，并且可以是空集

注 1. 根据定义，次微分是一些半空间的交，因此是一个凸集，当交集为空时，该集合为空

性质 2. 如果 $\mathbf{x} \in \mathrm{int}\ \mathrm{dom} f$ ，则 $\partial f(\mathbf{x})$ 非空有界

证明. 当 $\mathbf{x} \in \mathrm{int}\ \mathrm{dom}\ f$ 时，$(\mathbf{x}, f(\mathbf{x}))$ 是凸集 epi f 的边界，因此在 $(\mathbf{x}, f(\mathbf{x}))$ 存在一个支撑超平面，

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

即

$$\mathbf{a}^T\mathbf{y} + bt \leq \mathbf{a}^T\mathbf{x} + bf(\mathbf{x}) \quad \forall\ (\mathbf{y}, t) \in \mathrm{epi}\ f$$

若 $b > 0$，当 $t \to \infty$ 时不等式不成立
若 $b = 0$，不等式变为 $\mathbf{a}^T\mathbf{y} \leq \mathbf{a}^T\mathbf{x}$，令 $\mathbf{y} = \mathbf{x} + \epsilon\mathbf{a}$，其中 $\epsilon > 0$ 且充分小，此时不等式将不成立
因此 $b < 0$，对不等式做一个变形得到

$$-\frac{\mathbf{a}^T}{b}(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$$

由次梯度的定义可以得到，$-\frac{\mathbf{a}^T}{b} \in \partial f(\mathbf{x})$，因此当 $\mathbf{x} \in \mathrm{int}\ \mathrm{dom}\ f$ 时，集合 $\partial f(\mathbf{x})$ 非空。下证 $\partial f(\mathbf{x})$ 有界
当 $\mathbf{x} \in \mathrm{int}\ \mathrm{dom}\ f$ 时，取 $r > 0$ 且充分小，定义如下包含 2n 个点的集合

$$B = \{\mathbf{x} \pm r\mathbf{e}_k | k = 1, ..., n\} \subset \mathrm{dom}\ f$$

并定义 $M=\max_{\mathbf{y}\in B} f(\mathbf{y})<\infty$，对任意 $\mathbf{g}\in\partial f(\mathbf{x})$，总存在 $\mathbf{y}\in B$ 满足

$$r\|\mathbf{g}\|_\infty=\mathbf{g}^T(\mathbf{y}-\mathbf{x})$$

若 $\|\mathbf{g}\|_\infty=|g_k|$，则取 $\mathbf{y}=\mathbf{x}+r\mathrm{sign}(g_k)\mathbf{e}_k$。因为 $\mathbf{g}$ 是次梯度，则满足

$$f(\mathbf{x})+r\|\mathbf{g}\|_\infty=f(\mathbf{x})+\mathbf{g}^T(\mathbf{y}-\mathbf{x})\le f(\mathbf{y})\le M$$

可以得到

$$\|\mathbf{g}\|_\infty\le\frac{M-f(\mathbf{x})}{r}\quad\forall\,\mathbf{g}\in\partial f(\mathbf{x})$$

因此 $\partial f(\mathbf{x})$ 是有界的 □

1.1 可次微分计算例子

例 1. 绝对值函数 $f(x)=|x|$

因为 $f(x)$ 只有在 $x=0$ 处不可导，因此只需单独考虑这点的次微分，当 $x>0$ 时，$f(x)=x$，导数即为次微分，此时 $\partial f(x)=1$，同理可得当 $x<1$ 时，$\partial f(x)=-1$。当 $x=0$ 时利用定义求次微分

$$f(y)\ge f(x)+g(y-x)$$

即

$$|y|\ge gy$$

当 $y=0$ 时，不等式恒成立，当 $y>0$ 时，$g\le 1$，当 $y<0$ 时，$g\ge -1$，因此当 $x=0$ 是，$\partial f(x)\in[-1,1]$。综上

$$\partial f(x)=\begin{cases}1 & x>0\\ [-1,1] & x=0\\ -1 & x<-1\end{cases}$$

例 2. 欧几里得范数 $f(\mathbf{x})=\|\mathbf{x}\|_2$

1.2 不可次微分例子

例 3. $f:\mathbb{R}\to\mathbb{R},\mathrm{dom}\ f=\mathbb{R}_+$，其定义如下

$$f(x)=\begin{cases}1x=0\\ 0x=1\end{cases}$$

在 $x=0$ 处不可次微分

例 4. $f:\mathbb{R}\to\mathbb{R}, \mathrm{dom}\ f=\mathbb{R}_+$，其定义如下

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

在 $x=0$ 处不可次微分

由定义 $g\in\partial f(x)$

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

代入 $x=0$，即

$$f(y)\geq gy \quad \Longleftrightarrow \quad g\leq -\sqrt{\frac{1}{y}} \quad \forall\ y\in\mathrm{dom}\ f$$

得到 $g<-\infty$，因此在 $x=0$ 处不可次微分

1.3 次微分的单调性

凸函数的次微分是单调运算，即

$$(\mathbf{u}-\mathbf{v})^T(\mathbf{x}-\mathbf{y})\geq 0 \forall \mathbf{x},\mathbf{y},\mathbf{u}\in\partial f(\mathbf{x}),\mathbf{v}\in\partial f(\mathbf{y})$$

证明. 由次微分的定义可得

$$f(\mathbf{y})\geq f(\mathbf{x})+\mathbf{u}^T(\mathbf{y}-\mathbf{x}) \qquad (1)$$
$$f(\mathbf{x})\geq f(\mathbf{y})+\mathbf{v}^T(\mathbf{x}-\mathbf{y}) \qquad (2)$$

$(1)+(2)$ 即得

$$(\mathbf{u}-\mathbf{v})^T(\mathbf{x}-\mathbf{y})$$

□

1.4 次梯度和水平集

1.5 次梯度计算运算规则

1.5.1 可导函数

如果函数 f 是可导的，则

$$\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$$

1.5.2 非负线性组合

如果 $f(\mathbf{x})=\alpha_1 f_1(x)+\alpha_2 f_2(x), \alpha_1, \alpha_2 \geq 0$，则

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

注 2. 等式右边为集合的加法

1.5.3 对变量的仿射变换

如果 $f(\mathbf{x})=h(A\mathbf{x}+\mathbf{b})$，则

$$\partial f(\mathbf{x})=A^T \partial h(A\mathbf{x}+\mathbf{b})$$

1.5.4 逐点最大

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

定义 $I(\mathbf{x})=\{i|f_i(\mathbf{x})=f(x)\}$
弱梯度计算规则：选取任意 $k \in I(\mathbf{x})$, 计算 f_k 在点 $\mathbf{x}$ 的其中一个梯度
强梯度计算规则：$\partial f(\mathbf{x})=\operatorname{conv}\{\bigcup_{i\in I(\mathbf{x})} \partial f_i(\mathbf{x})\}$。即 active 函数在 $\mathbf{x}$ 点的次微分的并集的凸包，如果 f_i 是可导的，则 $\partial f(\mathbf{x})=\operatorname{conv}\{\nabla f_i(\mathbf{x})|i\in I(\mathbf{x})\}$

例 5. $f(\mathbf{x})=\max_{i=1,...,m}(\mathbf{a}_i^T\mathbf{x}+b_i)$，则在 $\mathbf{x}$ 点的次微分为

$$\partial f(\mathbf{x})=\operatorname{conv}\{\mathbf{a}_i|i\in I(\mathbf{x})\}$$

是一个多面体，其中 $I(\mathbf{x})=\{i|\mathbf{a}_i^T\mathbf{x}+b_i=f(\mathbf{x})\}$

1.5.5 最小化

已知函数 $f(\mathbf{x})=\inf_{\mathbf{y}} h(\mathbf{x},\mathbf{y})$，并且 h 关于 $(\mathbf{x},\mathbf{y})$ 联合凸。在 $\hat{\mathbf{x}}$ 处的次梯度的计算方式为
1. 假设最小值存在，计算 $h(\hat{\mathbf{x}},\mathbf{y})$ 的最小值 $\hat{\mathbf{y}}$
2. 计算次梯度 $(\mathbf{g},\mathbf{0})\in\partial h(\hat{\mathbf{x}},\hat{\mathbf{y}})$
则 $\mathbf{g}$ 即为 $f(\mathbf{x})$ 的一个次梯度

证明. 因为 $(\mathbf{g},\mathbf{0})\in\partial h(\hat{\mathbf{x}},\hat{\mathbf{y}})$，则由次梯度的定义可得，对 $\forall\ \mathbf{x},\mathbf{y}$ 成立

$$h(\mathbf{x},\mathbf{y})\geq h(\hat{\mathbf{x}},\hat{\mathbf{y}})+\mathbf{g}^T(\mathbf{x}-\hat{\mathbf{x}})+\mathbf{0}^T(\mathbf{y}-\hat{\mathbf{y}})$$

又因为 $h(\hat{\mathbf{x}},\hat{\mathbf{y}}) = \inf_{\mathbf{y}} h(\hat{\mathbf{x}},\mathbf{y}) = f(\hat{\mathbf{x}})$，上述不等式变为

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

对 $\forall\ \mathbf{x},\mathbf{y}$ 成立，因此

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

即 $\mathbf{g}$ 是 $f(\mathbf{x})$ 的次梯度 □

例 6. 计算 $f(\mathbf{x}) = \inf_{\mathbf{y}\in C} \|\mathbf{x}-\mathbf{y}\|_2$，其中 C 是闭凸集

$h(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2$，当 $\hat{\mathbf{x}} \in C$ 时，$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}} \|\hat{\mathbf{x}}-\mathbf{y}\| = \hat{\mathbf{x}}$，此时 $h(\hat{\mathbf{x}},\hat{\mathbf{y}}) = \|\hat{\mathbf{x}}-\hat{\mathbf{x}}\|_2 = 0$，根据次梯度的定义

$$h(\mathbf{x},\mathbf{y}) \geq h(\hat{\mathbf{x}},\hat{\mathbf{y}}) + \mathbf{g}^T(\mathbf{x}-\hat{\mathbf{x}}) + \mathbf{0}^T(\mathbf{y}-\hat{\mathbf{y}}) \quad \forall\ \mathbf{x},\mathbf{y}$$

即

$$\|\mathbf{x}-\mathbf{y}\|_2 \geq \mathbf{g}^T(\mathbf{x}-\hat{\mathbf{x}}) \quad \forall\ \mathbf{x},\mathbf{y}$$

取 $\mathbf{g} = 0$，此时 $\|\mathbf{x}-\mathbf{y}\|_2 \geq 0$ 对所有 $\mathbf{x},\mathbf{y}$ 恒成立

1.5.6 组合

已知函数 $f(\mathbf{x}) = h(f_1(\mathbf{x}),...,f_k(\mathbf{x}))$，其中 h 是单调不减凸函数，f_i 也为凸函数。在 $\hat{\mathbf{x}}$ 处的次梯度的计算方式为
1. 计算 $\mathbf{z} \in \partial h(f_1(\hat{\mathbf{x}}),...,f_k(\hat{\mathbf{x}}))$ 和 $g_i \in \partial f_i(\hat{\mathbf{x}})$
2. 计算次梯度 $\mathbf{g} = z_1 g_1 + ... + z_k g_k \in \partial f(\hat{\mathbf{x}})$

证明. 因为 $g_i \in \partial f_i(\hat{\mathbf{x}})$，则

$$f_i(\mathbf{x}) \geq f_i(\hat{\mathbf{x}}) + \mathbf{g}_i^T(\mathbf{x}-\hat{\mathbf{x}})$$

又因为 f 单调不减，因此

$$\begin{aligned} f(\mathbf{x}) &= h(f_1(\mathbf{x}),...,f_k(\mathbf{x})) \\ &\geq h(f_1(\mathbf{x}) + \mathbf{g}_1^T(\mathbf{x}-\hat{\mathbf{x}}),...,f_k(\mathbf{x}) + \mathbf{g}_k^T(\mathbf{x}-\hat{\mathbf{x}})) \end{aligned}$$

因为 $\mathbf{z} \in \partial h(f_1(\hat{\mathbf{x}}), ..., f_k(\hat{\mathbf{x}}))$，可以得到

$$\begin{aligned}
h(f_1(\mathbf{x}) + \mathbf{g}_1^T(\mathbf{x} - \hat{\mathbf{x}}), ..., f_k(\mathbf{x}) + \mathbf{g}_k^T(\mathbf{x} - \hat{\mathbf{x}})) &\geq h(f_1(\hat{\mathbf{x}}), ..., f_k(\hat{\mathbf{x}})) + \mathbf{z}^T \begin{bmatrix} \mathbf{g}_1^T(\mathbf{x} - \hat{\mathbf{x}}) \\ \vdots \\ \mathbf{g}_k^T(\mathbf{x} - \hat{\mathbf{x}}) \end{bmatrix} \\
&= h(f_1(\hat{\mathbf{x}}), ..., f_k(\hat{\mathbf{x}})) + (z_1\mathbf{g}_1 + ... + z_k\mathbf{g}_k)^T(\mathbf{x} - \hat{\mathbf{x}}) \\
&= f(\mathbf{x}) + \mathbf{g}^T(\mathbf{x} - \hat{\mathbf{x}})
\end{aligned}$$

因此 $\mathbf{g} = z_1 g_1 + ... + z_k g_k \in \partial f(\hat{\mathbf{x}})$ □

1.5.7 最优值函数

凸优化问题

$$\begin{aligned}
\min \;\; & f_0(\mathbf{x}) \\
s.t \;\; & f_i(\mathbf{x}) \leq u_i, i = 1, ..., m \\
& A\mathbf{x} = \mathbf{b} + \mathbf{v}
\end{aligned}$$

最优值记为 $f(\mathbf{u}, \mathbf{v})$

假设 $f(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ 是有限的，并且强对偶成立，原问题的对偶问题为

$$\begin{aligned}
\max \inf_x \; & \left(f_0(\mathbf{x}) + \sum_i \lambda_i (f_i(\mathbf{x}) - u_i) + \nu^T(A\mathbf{x} - \mathbf{b} - \mathbf{v}) \right) \\
& s.t. \; \lambda \succeq 0
\end{aligned}$$

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

证明. 由弱对偶定理可以得到

$$\begin{aligned}
f(\mathbf{u}, \mathbf{v}) &\geq \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_i \hat{\lambda}_i (f_i(x) - u_i) + \hat{\nu}^T(A\mathbf{x} - \mathbf{b} - \mathbf{v}) \right) \\
&\geq \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_i \hat{\lambda}_i (f_i(\mathbf{x}) - \hat{u}_i) + \nu^T(A\mathbf{x} - \mathbf{b} - \hat{\mathbf{v}}) \right) - \hat{\lambda}^T(\mathbf{u} - \hat{\mathbf{u}}) - \hat{\nu}^T(\mathbf{v} - \hat{\mathbf{v}}) \\
&= f(\hat{\mathbf{u}}, \hat{\mathbf{v}}) - \hat{\lambda}^T(\mathbf{u} - \hat{\mathbf{u}}) - \hat{\nu}^T(\mathbf{v} - \hat{\mathbf{v}})
\end{aligned}$$

由次梯度的定义可以得到 $(-\hat{\lambda}, -\hat{\nu}) \in \partial f(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ □

1.5.8 方差

已知函数

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

其中 $\mathbf{u}$ 是随机变量，且对任意 $\mathbf{u}$，h 关于 $\mathbf{x}$ 是凸的，在 $\hat{\mathbf{x}}$ 处的次梯度的计算方式为

1. 选择一个函数 $g(\mathbf{u}) \in \partial_{\mathbf{x}} h(\hat{\mathbf{x}}, \mathbf{u})$

2.$\mathbf{g} = \mathrm{E}_u g(\mathbf{u}) \in \partial f(\hat{\mathbf{x}})$

证明. 因为 $g(\mathbf{u}) \in \partial_{\mathbf{x}} h(\hat{\mathbf{x}}, \hat{\mathbf{u}})$，则

$$h(\mathbf{x}, \mathbf{u}) \geq h(\hat{\mathbf{x}}, \mathbf{u}) + g(\mathbf{u})^T(\mathbf{x} - \hat{\mathbf{x}})$$

因此

$$f(\mathbf{x}) = \mathrm{E}\ h(\mathbf{x}, \mathbf{u}) \geq \mathrm{E}\ \left(h(\hat{\mathbf{x}}, \mathbf{u}) + g^T(u)(\mathbf{x} - \hat{\mathbf{x}})\right)$$

由期望的性质得

$$\mathrm{E}\ \left(h(\hat{\mathbf{x}}, \mathbf{u}) + g^T(u)(\mathbf{x} - \hat{\mathbf{x}})\right) = f(\hat{\mathbf{x}}) + \mathbf{g}^T(\mathbf{x} - \hat{\mathbf{x}})$$

其中 $\mathbf{g} = \mathrm{E}_{\mathbf{u}} g(\mathbf{u})$，由梯度的定义可得 $\mathbf{g} \in \partial f(\hat{\mathbf{x}})$ □

2 对偶和最优性条件

2.1 无约束优化问题

定理 1. $\mathbf{x}^*$ 最小化 $f(\mathbf{x})$ 当且仅当

$$\mathbf{0} \in \partial f(\mathbf{x}^*)$$

证明. $\mathbf{x}^*$ 为最小值，则可以得到

$$\begin{aligned} f(\mathbf{y}) &\geq f(\mathbf{x}^*) + \mathbf{0}^T(\mathbf{y} - \mathbf{x}^*) \\ &= f(\mathbf{x}^*) + \mathbf{0}^T(\mathbf{y} - \mathbf{x}^*) \quad \forall \mathbf{y} \in \mathrm{dom} f \end{aligned}$$

由次梯度的定义可以得到

$$\mathbf{0} \in \partial f(\mathbf{x}^*)$$

□

例 7.

2.2 约束优化问题

约束优化问题的最优解利用 KKT 条件进行刻画，有如下约束优化问题

$$\min f_0(\mathbf{x})$$
$$s.t. f_i(\mathbf{x}) \le 0 \quad i = 1, ..., m$$

其中 $\mathrm{dom}\ f_i = \mathbb{R}^n$，即 f_i 处处可次微分。如果强对偶条件成立，则 $\mathbf{x}^*, \lambda^*$ 分别是原问题和对偶问题的最优解当且仅当

1.$\mathbf{x}^*$ 是可行解

2.$\lambda^* \succeq 0$

3.$\lambda_i f_i(\mathbf{x}^*) = 0, i = 1, ..., m$

4.$\mathbf{x}^*$ 是 $L(\mathbf{x}, \lambda) = f_0(\mathbf{x}) + \sum_{i=1}^m {\lambda_i}^* f_i(\mathbf{x})$ 的最小值点，即

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

3 方向导数

3.1 一般函数的方向导数

定义 3. *对于一般的函数 f，$f(x)$ 在 y 方向上的导数为*

$$\begin{aligned} f^{'}(\mathbf{x}; \mathbf{y}) &= \lim_{\alpha \searrow 0} \frac{f(\mathbf{x} + \alpha \mathbf{y}) - f(\mathbf{x})}{\alpha} \\ &= \lim_{t \to \infty} \left(t f(\mathbf{x} + \frac{1}{t}\mathbf{y}) - t f(\mathbf{x}) \right) \end{aligned}$$

性质 3. *$f^{'}(\mathbf{x}; \mathbf{y})$ 关于 $\mathbf{y}$ 是齐次的，即*

$$f^{'}(\mathbf{x}; \lambda \mathbf{y}) = \lambda f^{'}(\mathbf{x}; \mathbf{y}) \quad \forall\ \lambda \ge 0$$

3.2 凸函数的方向导数

对于凸函数的方向导数，将一般函数方向导数定义中的 lim 替换成 inf，即

$$\begin{aligned} f^{'}(\mathbf{x}; \mathbf{y}) &= \inf_{\alpha \searrow 0} \frac{f(\mathbf{x} + \alpha \mathbf{y}) - f(\mathbf{x})}{\alpha} \\ &= \inf_{t \to \infty} \left(t f(\mathbf{x} + \frac{1}{t}\mathbf{y}) - t f(\mathbf{x}) \right) \end{aligned}$$

证明. 容易得到，$h(\mathbf{y})=f(\mathbf{x}+\mathbf{y})-f(\mathbf{x})$ 是关于 y 的凸函数，且 $h(0)=0$。下证 $th(\mathbf{y}/t)$ 关于 t 单调不增，对任意的 $t>s>0$ 总有

$$\begin{aligned} th(\mathbf{y}/t)-sh(\mathbf{y}/s) &= t[f(\mathbf{x}+\mathbf{y}/t)-f(\mathbf{x})]-s[f(\mathbf{x}+\mathbf{y}/s)-f(\mathbf{x})] \\ &= t\left[f(\mathbf{x}+\mathbf{y}/t)-\left(\frac{t-s}{t}f(x)+\frac{s}{t}f(\mathbf{x}+\mathbf{y}/s)\right)\right] \end{aligned}$$

因为 $0<\frac{t-s}{t}<1, 0<\frac{s}{t}<1$，且 $\frac{t-s}{t}\mathbf{x}+\frac{s}{t}\mathbf{x}+\mathbf{y}/s=\mathbf{x}+\mathbf{y}/t$，由凸函数的定义可得

$$\frac{t-s}{t}f(x)+\frac{s}{t}f(\mathbf{x}+\mathbf{y}/s)\geq f(\mathbf{x}+\mathbf{y}/t)$$

因此

$$\left[f(\mathbf{x}+\mathbf{y}/t)-\left(\frac{t-s}{t}f(x)+\frac{s}{t}f(\mathbf{x}+\mathbf{y}/s)\right)\right]\leq 0$$

即 $th(\mathbf{y}/t)$ 关于 t 单调不增，因此可以得到

$$\begin{aligned} f^{'}(\mathbf{x};\mathbf{y}) &= \lim_{t\to\infty} th(\mathbf{y}/t)=\inf_{t>0} th(\mathbf{y}/t) \\ &= \inf_{t\to\infty}\left(tf(\mathbf{x}+\frac{1}{t}\mathbf{y})-tf(\mathbf{x})\right) \end{aligned}$$

□

性质 4. $f^{'}(\mathbf{x};\mathbf{y})$ 关于 $\mathbf{y}$ 是凸的

性质 5. 由方向导数的定义，可以得到

$$f(\mathbf{x}+\alpha\mathbf{y})\geq f(\mathbf{x})+\alpha f^{'}(\mathbf{x};\mathbf{y}) \quad \forall\ \alpha\geq 0$$

即 $f^{'}(\mathbf{x};\mathbf{y})$ 定义了 f 在 $\mathbf{y}$ 方向上的下界

3.3 方向导数与次梯度

对于凸函数 $f,\mathbf{x}\in \text{int dom}f$，方向导数与次微分具有如下关系

$$f^{'}(\mathbf{x};\mathbf{y})=\sup_{\mathbf{g}\in\partial f(\mathbf{x})}\mathbf{g}^T\mathbf{y}$$

即 $f^{'}(x;y)$ 是 $\partial f(\mathbf{x})$ 的支撑函数
对于可导函数，则 $f^{'}(\mathbf{x};\mathbf{y})=\nabla f(\mathbf{x})^T\mathbf{y}$

证明. 如果 $\mathbf{g} \in \partial f(\mathbf{x})$，则由方向的定义可以得到

$$f^{'}(\mathbf{x};\mathbf{y}) \geq \inf_{\alpha>0} \frac{f(\mathbf{x})+\alpha\mathbf{g}^T\mathbf{y}-f(\mathbf{x})}{\alpha} = \mathbf{g}^T\mathbf{y}$$

下证存在 $\hat{\mathbf{g}} \in \partial f(\mathbf{x})$ 满足 $f^{'}(\mathbf{x};\mathbf{y}) = \hat{\mathbf{g}}^T\mathbf{y}$
设 $\hat{\mathbf{g}}$ 是 $f^{'}(\mathbf{x};\mathbf{y})$ 在 $\mathbf{y}$ 方向的一个次梯度，则由次梯度的定义和方向导数的性质可得

$$\lambda f^{'}(\mathbf{x};\mathbf{v}) = f^{'}(\mathbf{x};\lambda\mathbf{v}) \geq f^{'}(\mathbf{x};\mathbf{y}) + \hat{\mathbf{g}}^T(\lambda\mathbf{v}-\mathbf{y}), \quad \forall \mathbf{v}, \lambda \geq 0$$

令 $\lambda \to \infty$ 可以得到

$$f^{'}(\mathbf{x};\mathbf{v}) \geq \hat{\mathbf{g}}^T\mathbf{v}$$

有前面 $f^{'}(\mathbf{x};\mathbf{y})$ 定义了 f 在 $\mathbf{y}$ 方向上的下界，可以得到

$$f(\mathbf{x}+\mathbf{v}) \geq f(\mathbf{x}) + f^{'}(\mathbf{x};\mathbf{v}) \geq f(\mathbf{x}) + \hat{\mathbf{g}}^T\mathbf{v} \quad \forall\ \mathbf{v}$$

则由次梯度的定义可以得到

$$\hat{\mathbf{g}} \in \partial f(\mathbf{x})$$

令 $\lambda = 0$ 可以得到

$$f^{'}(\mathbf{x};\mathbf{y}) \leq \hat{\mathbf{g}}^T\mathbf{y}$$

□

3.4 下降方向和次梯度

定义 4. 如果 $f^{'}(\mathbf{x};\mathbf{y}) < 0$，则 $\mathbf{y}$ 是 f 在 $\mathbf{x}$ 处的下降方向

注 3. 如果 $f(\mathbf{x}) \neq 0$，则可微函数 f 的负梯度是下降方向，但负的次梯度并不总是下降方向。比如 $f(\mathbf{x}_1,\mathbf{x}_2) = |\mathbf{x}_1| + 2|\mathbf{x}_2|$，$\mathbf{g} = (1,2) \in \partial f(1,0)$，但是 $\mathbf{y} = -\mathbf{g} = (-1,-2)$ 不是 f 在 $(1,0)$ 的下降方向