无约束不可微问题的最优性理论

仍考虑问题:

$$\min_{x \in \mathbb{R}^n} \quad f(x),$$

但其中$f(x)$ 为不可微函数.

- 很多实际问题的目标函数不是光滑的, 例如$f(x) = \|x\|_1$.
- 对于此类问题, 由于目标函数可能不存在梯度和海瑟矩阵, 因此无约束可微问题的一阶和二阶条件不适用, 此时我们必须使用其他最优性条件来判断不可微问题的最优点.

凸优化问题一阶充要条件

对于目标函数是凸函数的情形，我们已经引入了次梯度的概念并给出了其计算法则.一个自然的问题是:可以利用次梯度代替梯度来构造最优性条件吗?

定理 (凸优化问题一阶充要条件)

假设f 是适当且凸的函数，则x^* 为全局极小点当且仅当$0 \in \partial f(x^*)$.

- **必要性**:因为x^* 为全局极小点，所以

$$f(y) \geq f(x^*) = f(x^*) + 0^{\mathrm{T}}(y - x^*), \quad \forall y \in \mathbb{R}^n \Rightarrow 0 \in \partial f(x^*).$$

- **充分性**:如果$0 \in \partial f(x^*)$, 那么根据次梯度的定义

$$f(y) \geq f(x^*) + 0^{\mathrm{T}}(y - x^*) = f(x^*), \quad \forall y \in \mathbb{R}^n.$$

因而x^* 为一个全局极小点.

凸优化问题的一阶充要条件:注记

- $0 \in \partial f(x^*)$ 是x^* 为全局最优解的充要条件, 这个结论比一般的无约束可微问题一阶必要条件要强, 其原因是凸问题有非常好的性质, 它的稳定点中不存在鞍点.
- 因此, 可通过计算凸函数的次梯度来求解其对应的全局极小点.
- 相较于非凸函数, 凸函数的最优性分析简单, 计算以及验证起来比较方便, 在实际建模中受到广泛的关注.

复合优化问题

- 在实际问题中，目标函数不一定是凸函数，但它可以写成一个光滑函数与一个非光滑凸函数的和.
- 其中目标函数的光滑项可能是凸的，比如LASSO 问题、图像去噪问题和盲反卷积问题.
- 也可能是非凸的，例如字典学习问题和神经网络的损失函数.
- 因此研究此类问题的最优性条件十分必要.
- 考虑一般复合优化问题

$$\min_{x\in\mathbb{R}^n} \quad \psi(x) \stackrel{\text{def}}{=\!=} f(x) + h(x),$$

其中f 为光滑函数(可能非凸)，h 为凸函数(可能非光滑).

复合优化问题的一阶必要条件

定理 (复合优化问题的一阶必要条件)

令x^* 为复合优化问题的一个局部极小点, 那么

$$-\nabla f(x^*) \in \partial h(x^*),$$

其中$\partial h(x^*)$ 为凸函数h 在点x^* 处的次梯度集合.

- 由$\psi(x) = f(x) + h(x)$ 及方向导数的定义

$$\begin{aligned}
\psi'(x^*; d) &= \lim_{t \to 0+} \frac{\psi(x^* + td) - \psi(x^*)}{t} \\
&= \nabla f(x^*)^{\mathrm{T}} d + \partial h(x^*; d) \\
&= \nabla f(x^*)^{\mathrm{T}} d + \sup_{\theta \in \partial h(x^*)} \theta^{\mathrm{T}} d.
\end{aligned}$$

复合优化问题的一阶必要条件:证明

- 若$-\nabla f(x^*) \notin \partial h(x^*)$, 由$h$ 是凸函数知$\partial h(x^*)$ 是有界闭凸集, 又根据严格分离定理知, 存在$d \in \mathbb{R}^n$ 以及常数b, 使得

$$\theta^{\mathrm{T}} d < b < -\nabla f(x^*)^{\mathrm{T}} d, \quad \forall \theta \in \partial h(x^*) \Rightarrow \psi'(x^*; d) < 0.$$

- 这说明对充分小的非负实数t,

$$\psi(x^* + td) < \psi(x^*).$$

这与x^* 的局部极小性矛盾.因此$-\nabla f(x^*) \in \partial h(x^*)$.

- 该定理给出了当目标函数含有非光滑凸函数时的一阶必要条件.
- 由于目标函数可能是整体非凸的, 因此一般没有一阶充分条件.

实例: ℓ_1 范数优化问题

$$\min_{x\in\mathbb{R}^n} \quad \psi(x) \stackrel{\text{def}}{=\!=} f(x) + \mu\|x\|_1,$$

其中$f(x):\mathbb{R}^n \to \mathbb{R}$ 为光滑函数, 系数$\mu > 0$.

- $\|x\|_1$ 不是可微的, 但我们可以计算其次微分

$$\partial_i\|x\|_1 = \begin{cases} \{1\}, & x_i > 0, \\ [-1,1], & x_i = 0, \\ \{-1\}, & x_i < 0. \end{cases}$$

- 若x^* 是局部最优解, 则$-\nabla f(x^*) \in \mu\partial\|x^*\|_1$, 从而有必要条件:

$$\nabla_i f(x^*) = \begin{cases} -\mu, & x_i^* > 0, \\ a \in [-\mu,\mu], & x_i^* = 0, \\ \mu, & x_i^* < 0. \end{cases}$$

- 如果$f(x)$ 是凸的(比如在LASSO 问题中$f(x) = \frac{1}{2}\|Ax-b\|^2$), 那么满足上式的x^* 就是问题的全局最优解.