BCD&ADMM 收敛理论

罗自炎

zyluo@bjtu.edu.cn

AI4M优化方向暑期班 北京大学
2024.7

Lecture 3：无约束优化最优性条件

无约束优化最优性条件

- 解存在性与唯一性
- 无约束光滑优化最优性条件
- 无约束非光滑优化最优性条件

无约束优化最优性条件

最优解存在性(Weierstrass定理)

数学模型： $\min\limits_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in \mathcal{X}$

其中$\mathcal{X} \subseteq \mathbb{R}^n$为可行域. 记最优解集为$\operatorname{argmin}_{\mathcal{X}} f$

定理1 设$f\colon \mathcal{X} \to (-\infty, +\infty]$是恰当且闭的，则$\operatorname{argmin}_{\mathcal{X}} f$为非空紧集，若如下任一条件成立：

(1) $dom\ f = \{x \in \mathcal{X}: f(x) < +\infty\}$是有界的;

(2) $\exists\alpha$使得下水平集$C_\alpha = \{x \in \mathcal{X}: f(x) < \alpha\}$非空有界;

(3) f 是强制的, i.e.,$\{x^k\} \subseteq \mathcal{X}, x^k \to +\infty \implies \lim\limits_{k\to\infty} f(x^k) = +\infty$

无约束优化最优性条件

最优解唯一性

强拟凸函数： 给定非空凸集$\mathcal{X}$ 与函数$f: \mathcal{X} \rightarrow (-\infty, +\infty]$,若对任意的$x \neq y$以及$\lambda \in (0,1)$, 均有： $f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$, 则称$f$为$\mathcal{X}$上的**强拟凸函数**. （**严格凸 $\Rightarrow$ 强拟凸 $\nRightarrow$ 凸**）

定理2 给定**非空凸紧集**$\mathcal{X} \subseteq \mathbb{R}^n$,设$f: \mathcal{X} \rightarrow (-\infty, +\infty]$是**适当、闭、强拟凸**函数，则$\operatorname{argmin}_{\mathcal{X}} f$为**单点集**，i.e., 存在唯一的$x^*$ 满足：

$$f(x^*) < f(x), \qquad \forall x \in \mathcal{X} \setminus \{x^*\}.$$

无约束优化最优性条件

无约束光滑优化

数学模型：

$$\min_{x\in\mathbb{R}^n} f(x) \qquad (1)$$

其中$f:\mathbb{R}^n \to \mathbb{R}$为连续可微函数.

函数的下降方向： 设$\bar{x} \in \mathbb{R}^n$是任一给定向量，$d \in \mathbb{R}^n$且$d \neq 0$. 若存在$\delta > 0$，对任意$\lambda \in (0, \delta)$，有$f(\bar{x} + \lambda d) < f(\bar{x})$，则称$d$为函数$f$在点 $\bar{x}$ 处的一个**下降方向**.

$\boldsymbol{d^T \nabla f(x) < 0 \implies d}$**是**$\boldsymbol{f}$**在**$\boldsymbol{x}$**处的下降方向**

无约束优化最优性条件

一阶必要条件

定理3 若$x^* \in \mathbb{R}^n$是无约束优化问题(1)的一个局部最优解，则有：

$$\nabla f(x^*) = 0.$$

证明 反证法. 假设$\nabla f(x^*) \neq 0$,取$d = -\nabla f(x^*) \neq 0$. 则对充分小的数$\alpha > 0$,

由一阶泰勒展开式可得：

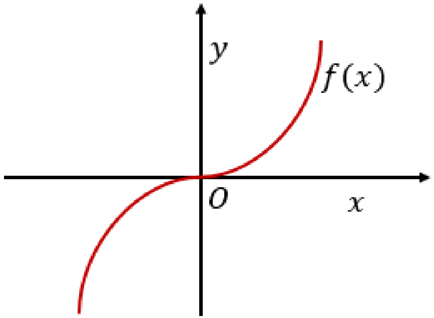
$$\begin{aligned} f(x^* + \alpha d) &= f(x^*) + \alpha \nabla f(x^*)^T d + o(\alpha\|d\|) \\ &= f(x^*) - \alpha\|\nabla f(x^*)\|^2 + o(\alpha\|d\|) \\ &< f(x^*) \end{aligned}$$

与x^*是局部最优解矛盾，从而定理得证.

无约束优化最优性条件

一阶条件的非充分性

例题1 一元函数$f(x) = x^3$在$x^* = 0$处有:
$\nabla f(x^*) = f'(x^*) = 0$, 然而$x^* = 0$ 不是函数 f的局部极小值点,也不是局部极大值点. 因此定理1中的一阶最优性条件**不是充分条件**.



定义1 对于无约束优化问题(1)，称满足$\nabla f(x^*) = 0$ 的点$x^* \in \mathbb{R}^n$为问题(1)的**稳定点**或**驻点**（stationary point）. 若 x^* 既不是极大值点，也不是极小值点，则称之为**鞍点**（saddle point）.

无约束优化最优性条件

二阶必要条件

定理4 若$x^* \in \mathbb{R}^n$是无约束优化问题(1)的一个局部最优解，则有：

$$\nabla f(x^*) = 0 \text{ 且} \nabla^2 f(x^*)\text{为半正定矩阵.}$$

证明 由定理1可知$\nabla f(x^*) = 0$，因而只需证明$\nabla^2 f(x^*)$为半正定矩阵.

设$d \in \mathbb{R}^n$为任一非零向量，由f在x^*二阶泰勒展开式可得：

$$\begin{aligned} f(x^* + \alpha d) &= f(x^*) + \alpha \nabla f(x^*)^T d + \frac{1}{2}\alpha^2 d^T \nabla^2 f(x^*) d + o(\alpha^2 \|d\|^2) \\ &= f(x^*) + \frac{1}{2}\alpha^2 d^T \nabla^2 f(x^*) d + o(\alpha^2 \|d\|^2) \end{aligned}$$

$$\Rightarrow \frac{f(x^*+\alpha d) - f(x^*)}{\alpha^2} = \frac{1}{2} d^T \nabla^2 f(x^*) d + \boxed{\frac{o(\alpha^2 \|d\|^2)}{\alpha^2}} \rightarrow 0, \text{ 当}\alpha \rightarrow 0$$

由x^*为局部最优解可知,当$|\alpha|$充分小时,$f(x^* + \alpha d) - f(x^*) \geq 0$,

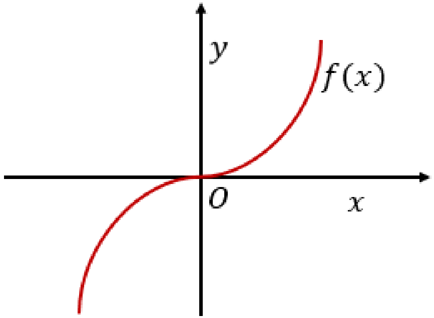
从而得到$d^T \nabla^2 f(x^*) d \geq 0, \forall d \neq 0$. 半正定性得证.

无约束优化最优性条件

二阶条件的非充分性

例题2 一元函数$f(x) = x^3$在$x^* = 0$处有:

$\nabla f(x^*) = f'(x^*) = 0, \nabla^2 f(x^*) = f''(x^*) = 0$ (半正定). 然而$x^* = 0$ 不是函数f的局部极小值点,也不是局部极大值点. 因此定理2中的二阶最优性条件**不是充分条件**.



无约束优化最优性条件

二阶充分条件

定理5 若$x^* \in \mathbb{R}^n$满足$\nabla f(x^*) = 0$且$\nabla^2 f(x^*)$为**正定矩阵**.则x^*是无约束优化问题(1)的一个**严格**局部最优解.

证明 设$d \in \mathbb{R}^n$为任一非零向量，由f在x^*二阶泰勒展开式可得：

$$f(x^* + \alpha d) = f(x^*) + \alpha \nabla f(x^*)^T d + \frac{1}{2}\alpha^2 d^T \nabla^2 f(x^*) d + o(\alpha^2 \|d\|^2)$$

$$= f(x^*) + \frac{1}{2}\alpha^2 d^T \nabla^2 f(x^*) d + \boxed{o(\alpha^2 \|d\|^2)} \rightarrow 0,\ \text{当}\alpha \rightarrow 0$$

由$\nabla^2 f(x^*)$为正定矩阵以及d非零可知, $d^T \nabla^2 f(x^*) d > 0$.

因而存在$\delta > 0$, 使得当$\alpha \in (0, \delta)$时有$\frac{1}{2}\alpha^2 d^T \nabla^2 f(x^*) d + o\left(\alpha^2 \|d\|^2\right) > 0$. 从而：

$f(x^* + \alpha d) - f(x^*) > 0$. 结合非零$d$的任意性, x^*的严格局部最优性得证.

无约束优化最优性条件

例题3：求解无约束优化问题 $\min_{x\in\mathbb{R}^2} f(x) = \frac{1}{3}x_1^3 + \frac{1}{3}x_2^3 - x_2^2 - x_1$

解 写出目标函数的梯度与海塞矩阵如下：

$$\nabla f(x) = \begin{pmatrix} x_1^2 - 1 \\ x_2^2 - 2x_2 \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} 2x_1 & 0 \\ 0 & 2x_2 - 2 \end{pmatrix}$$

由$\nabla f(x) = 0$ 可得该优化问题的所有稳定点如下:

$$x^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, x^{(2)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, x^{(3)} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, x^{(4)} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

计算可知：只有在$x^{(1)}$处相应的海塞矩阵$\nabla^2 f(x^{(1)})$为正定矩阵，其它三个稳定点处的海塞矩阵都不是半正定的.

因而$x^{(1)}$是该优化问题的一个严格局部最优解. **是全局最优解吗？**

无约束优化最优性条件

例题4：求解无约束优化问题 $\min\limits_{x\in\mathbb{R}^2} f(x) = 5x_1^2 - 6x_1x_2 + 5x_2^2$

解 写出目标函数的梯度与海塞矩阵如下：

$$\nabla f(x) = \begin{pmatrix} 10x_1 - 6x_2 \\ 10x_2 - 6x_1 \end{pmatrix}, \qquad \nabla^2 f(x) = \begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}$$

由$\nabla f(x) = 0$ 可得 $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$为该优化问题的稳定点.

进一步，由于$\boxed{\forall x, \nabla^2 f(x) = \begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}\text{为正定矩阵,}}$ ⇨ f为凸函数

可知二阶充分条件在x^*处满足，从而x^*为严格局部最优解.

结合f为凸函数可知，全局最优解与局部最优解等价.

因而x^* 为该凸规划问题的**唯一全局最优解**.

无约束优化最优性条件

无约束光滑凸优化

定理6 若无约束优化问题(1) 的目标函数f是连续可微的凸函数, 则 $x^* \in \mathbb{R}^n$是凸优化(1)的全局最优解当且仅当$\nabla f(x^*) = 0$.

证明

充分性 由凸函数f的一阶判别条件可知：$\forall x \in \mathbb{R}^n$,有

$$f(x) - f(x^*) \geq \nabla f(x^*)^T(x - x^*) = 0$$

因而x^*是凸规划(1)的全局最优解.

必要性 由一阶必要最优性条件（定理3）得证.

无约束优化最优性条件

无约束非光滑优化—凸优化

定理7 若无约束优化问题(1) 的目标函数f是**适当凸**函数, 则$x^* \in \mathrm{dom} f$是凸规划(1)的全局最优解当且仅当$0 \in \partial f(x^*)$.

一阶充要条件

证明 **充分性** 由凸函数f的次梯度定义可知：$\forall x \in \mathrm{dom}\, f$,有

$$f(x) \geq f(x^*) + 0^T(x - x^*) = f(x^*), \quad (*)$$

因而x^*是凸规划(1)的全局最优解.

必要性 由（*）可知$0 \in \partial f(x^*)$.

无约束优化最优性条件

无约束非光滑优化—非凸优化

定理8 设优化问题(1) 的目标函数f是**适当下半连续**的, 若$x^* \in \mathrm{dom} f$ 是问题(1)的一个局部最优解，则$0 \in \hat{\partial} f(x^*) \subseteq \partial f(x^*)$.

一阶必要条件

证明 由函数f的Fréchet 次微分定义，只需证明：$\forall x \in \mathrm{dom}\, f$,有

$$\liminf_{x \to x^*, x \neq x^*} \frac{f(x) - f(x^*) + 0^T(x - x^*)}{\|x - x^*\|} \geq 0$$

$$\Uparrow$$

x^*是局部最优解

无约束优化最优性条件

无约束非光滑优化—复合优化

数学模型： $$\min_{x\in\mathbb{R}^n} \psi(x) = f(x) + h(x) \tag{2}$$

其中$f \in C^1$(可能非凸), h为**适当闭**函数（可能非光滑）.

定理9 若$x^* \in \operatorname{dom}\psi$是复合优化问题(2) 的一个局部最优解, 则：

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

一阶必要条件

证明 定理8 $\Longrightarrow 0 \in \partial\psi(x^*) = \nabla f(x^*) + \partial h(x^*)$ 【VA,Ex8.8(c) or Thm10.1】

无约束优化最优性条件

习题

设函数$f:\mathbb{R}^n \to \mathbb{R}$连续可微, 常数$\mu > 0$，试写出如下$\ell_1$范数优化问题的一阶最优性条件：

$$\min_{x\in\mathbb{R}^n} \psi(x) := f(x) + \mu\|x\|_1$$

Thank You !