Subject: Section2, Foundations of Smooth Optimization

Date: from 9 July, 2025 to July 12

Contents

Foundations of Smooth Optimization

Outline

本节内容作为优化理论学习的基础部分，介绍了极小点、L-Smooth 光滑性、凸性三个关键概念，引入了 Taylor Theorem 这一重要工具，用于刻画和分析无约束与约束优化问题的极小点性质。在上述内容基础上，讨论了连续可微与二阶可微情况下极小点的必要条件与充分条件，并给出了详细证明.

Preliminary / Basic Tools

Definition

Definition 1. 考虑函数 $f : \mathcal{D} \to \mathbb{R}$，其中 $\mathcal{D} = dom(f) \subset \mathbb{R}^n$，在优化问题中有如下定义：

1. ***Local minimizer:*** 称 $x^* \in \mathcal{D}$ 为局部极小点*(local minimizer)*若存在邻域 $\mathcal{N}(x^*)$ 使得 $\forall\, x \in \mathcal{N} \cap \mathcal{D}$，$f(x) \geqslant f(x^*)$.
2. ***Global minimizer:*** 称 $x^* \in \mathcal{D}$ 为全局极小点*(global minimizer)*若 $\forall\, x \in \mathcal{D}$，$f(x) \geqslant f(x^*)$.
3. ***Strict local minimizer:*** 称邻域 $\mathcal{N}(x^*)$ 中的局部极小点 $x^* \in \mathcal{D}$ 为严格局部极小点*(strict local minimizer)*若 $\forall\, x \in \mathcal{N}(x^*)$，$x \neq x^*$，$f(x) > f(x^*)$
4. ***Isolated local minimizer:*** 称邻域 $\mathcal{N}(x^*)$ 中的局部极小点 $x^* \in \mathcal{D}$ 为孤立局部极小点若 $\mathcal{N}(x^*)$ 中除 x^* 外无其他极小点.
5. ***Unique minimizer:*** 称 $x^* \in \mathcal{D}$ 为唯一极小点*(unique minimizer)*若其是唯一的全局极小点.

Definition 2. 对于约束优化问题

$$\min_{x\in\Omega} \; f(x) \tag{1}$$

其中 $\Omega \subset \mathcal{D} \subset \mathbb{R}^n$ 为闭集. 对问题*(1)*的解有如下定义：

1. ***Local solution:*** 称 x^* 为局部解*(local solution)*若存在邻域 $\mathcal{N}(x^*)$ 使得 $\forall\, x \in \mathcal{N} \cap \Omega$，$f(x) \geqslant f(x^*)$.
2. ***Global solution:*** 称 x^* 为全局解*(global solution)*若 $\forall\, x \in \Omega$，$f(x) \geqslant f(x^*)$.

Taylor's Theorem

优化问题的分析中一个重要的工具是 Taylor's Theorem，它说明了光滑函数 f 如何被依赖于 f 的低阶导数的多项式局部逼近.

Theorem 1 (Taylor's Theorem)**.** 给定连续可微函数*(continuously differentiable)* $f : \mathbb{R}^n \to \mathbb{R}$，给定 $x, p \in \mathbb{R}^n$，有如下两种形式的 *Taylor's theorem*：

$$\textit{integral form: } f(x+p) = f(x) + \int_0^1 \nabla f(x+\gamma p)^T p \, d\gamma, \tag{2a}$$

$$\textit{mean-value form: } f(x+p) = f(x) + \nabla f(x+\gamma p)^T p, \textit{ for some } \gamma \in (0,1). \tag{2b}$$

进一步，若 f 二次可微(*twice continuously differentiable*)，则有：

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+\gamma p) p \, d\gamma \tag{3a}$$

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+\gamma p) p, \textit{ for some } \gamma \in (0,1) \tag{3b}$$

Corollary 1. 若 f 连续可微，则

$$f(x+p) = f(x) + \nabla f(x)^T p + o(\|p\|). \tag{4}$$

Proof. 由 mean-value form 的 Taylor's Theorem 式(2b) 可得

$$\begin{aligned} f(x+p) &= f(x) + \nabla f(x+\gamma p)^T p = f(x) + \nabla f(x)^T p + \left(\nabla f(x+\gamma p) - \nabla f(x)\right)^T p \\ &= f(x) + \nabla f(x)^T p + O(\|\nabla f(x+\gamma p) - \nabla f(x)\| \cdot \|p\|) = f(x) + \nabla f(x)^T + o(\|p\|). \end{aligned} \tag{5}$$

其中最后一个等式是由于 $\|\nabla f(x+\gamma p) - \nabla f(x)\| \xrightarrow{p \to 0} 0$，因此 $\|\nabla f(x+\gamma p) - \nabla f(x)\| = O(\|p\|)$，故 $O(\|\nabla f(x+\gamma p) - \nabla f(x)\| \cdot \|p\|) = o(p)$. □

L-Smoothness

在研究优化问题时，导数的 Lipschitz 系数 L 是函数的一个重要性质，其描述了函数变化的剧烈程度.

Definition 3 (L-smoothness)**.** 对于连续可微函数 f，称其是 *L-smooth* 的或有 *L-Lipschitz* 导数，若其导数是 *L-Lipschitz*的，即满足：

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x-y\|, \; \forall \, x, y \in dom(f) \tag{6}$$

Note 1. 对于一个函数 f 而言，*L-Lipschitz* 是指其自身满足 $\|f(x) - f(y)\| \leqslant \|x-y\|$，*L-smooth* 是指其梯度 ∇f 是 *L-Lipschitz* 的.

Lemma 1. 给定一 *L-smooth* 函数 f，则其在 y 处的函数值上界能够被一在 y 点处函数值为 $f(y)$ 的二次函数(*quadratic function*)控制：

$$f(y) \leqslant f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|^2. \tag{7}$$

Proof. 根据 integral form 的 Taylor's Theorem 式(2a) 可知

$$f(y) = f(x+y-x) = f(x) + \int_0^1 \nabla f(x+\gamma(y-x))^T (y-x) \, \mathrm{d}\gamma \tag{8}$$

由于 f 的导数具有 L-Lipschitz 性质，结合式(8)可得

$$\begin{aligned} & f(y) - f(x) - \nabla f(x)^T (y-x) = \int_0^1 \left(\nabla f(x+\gamma(y-x)) - \nabla f(x)\right)^T (y-x) \, \mathrm{d}\gamma \\ & \leqslant \int_0^1 \left(\|\nabla f(x+\gamma(y-x)) - \nabla f(x)\| \cdot \|y-x\|\right) \, \mathrm{d}x \leqslant \int_0^1 L\gamma \|y-x\|^2 \, \mathrm{d}\gamma = \frac{L}{2}\|y-x\|^2. \end{aligned} \tag{9}$$

□

对于一二次可微的 L-smooth 函数 f，则 L 可以通过其 Hessian 矩阵 $\nabla^2 f(x)$ 刻画.

Lemma 2. 给定一二次可微函数 f，若其 *L-smooth*，则

$$\nabla^2 f(x) \preceq LI,\ \forall x \in dom(f). \tag{10}$$

其中 I 为单位矩阵. 反之，若 $-LI \preceq \nabla^2 f(x) \preceq LI$，则 f 是 *L-smooth* 的.

证明详见附录B.

Characterizing Minima of Smooth Functions

Necessary / Sufficient Condition for Smooth Unconstrained Optimization

本节将使用在 Preliminary 中介绍的工具分析如下无约束优化问题(unconstrained optimization problem)的解:

$$\min_{x\in\mathbb{R}^n} f(x), \tag{11}$$

其中 f 为光滑函数.

首先给出必要条件(necessary condition)，即当 x^* 为局部解时满足的条件.

Theorem 2 (Necessary Conditions for Smooth Unconstrained Optimization)**.**

1. 对于连续可微函数 f，若 x^* 为式*(11)*的局部解，那么 $\nabla f(x^*) = 0$.
2. 对于二次连续可微函数 f，若 x^* 为式*(11)*的局部解，那么 $\nabla f(x^*) = 0$ 且 $\nabla^2 f(x^*)$ 是半正定的*(positive semidefinite)*.

Note 2. *Theorem 2* 中第一条称为一阶必要条件***(first-order necessary condition)***，第二条称为二阶必要条件***(second-order necessary condition)***. 证明详见附录*A*.

下面给出充分条件(sufficient condition)，即当 x^* 满足要求时一定为(严格)局部解.

Theorem 3 (Sufficient Conditions for Smooth Unconstrained Optimization)**.**
对于二次连续可微函数 f 和点 x^*，若 $\nabla f(x^*) = 0$，且 $\nabla^2 f(x^*)$ 是正定*(positive definite)*的，那么 x^* 为为式*(11)*的严格局部解.

Proof. 由于 $\nabla^2 f(x^*)$ 是正定的，根据其连续性，存在 $\rho > 0, \epsilon > 0$，使得 $\forall\, p \in \mathbb{R}^n$，$\|p\| \leqslant \rho, \forall\, \gamma \in (0,1)$，有

$$\nabla^2 f(x^* + \gamma p) \geqslant \epsilon I \tag{12}$$

由二次 mean-value form 的 Taylor's Theorem 式(3b)，存在 $\gamma \in (0,1)$ 使得

$$f(x^*+p) = f(x^*) + \nabla f(x^*)^T p + \frac{1}{2} p^T \nabla^2 f(x^* + \gamma p) p \geqslant f(x^*) + \frac{1}{2}\epsilon \|p\|^2,\ \forall\, p \in \mathbb{R}^n,\ \|p\| \leqslant \rho. \tag{13}$$

取 $\mathcal{N} = \{x \in \mathbb{R}^n \mid \|x - x^*\| < \rho\}$，则对于任意 $x \in \mathcal{N}$，有

$$f(x) \geqslant f(x^*) + \frac{1}{2}\epsilon \|x - x^*\|^2 > f(x^*),\ \forall x \in \mathcal{N} \setminus \{x^*\}. \tag{14}$$

因此 x^* 是严格局部解. □

Convex Sets and Functions

Convex Sets and Functions

凸函数在优化中起着核心作用，因为这些凸函数实例很容易验证最优性，并且保证在合理的计算量内可以发现这种最优性。

Definition 4 (Convex Set)**.** 称集合 $\Omega \subset \mathbb{R}^n$ 为凸集(*convex set*)，若

$$x, y \in \Omega \Rightarrow (1-\alpha)x + \alpha y \in \Omega,\ \forall\, \alpha \in [0,1]. \tag{15}$$

Note 3. 凸集的几何意义是：任意两点之间的连线都在集合内.

Definition 5 ((Weekly) Convex Function)**.** 称函数 f 为(弱)凸函数(*weekly convex function*)，若

$$f((1-\alpha)x + \alpha y) \leqslant (1-\alpha)f(x) + \alpha f(y),\ \forall\, x, y \in dom(f),\ \forall\, \alpha \in [0,1]. \tag{16}$$

Note 4. 凸函数的几何意义是：任意两点之间的连线在函数图像上方. 或 f 的*epigraph*，定义为 $epi\ f = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid t \geqslant f(x)\}$，是一个凸集.

Definition 6 (Strongly Convex Function)**.** 称连续可微的凸函数 f 为强凸函数(*strongly convex function*)，若

$$f((1-\alpha)x+\alpha y) \leqslant (1-\alpha)f(x)+\alpha f(y) - \frac{m}{2}\alpha(1-\alpha)\|x-y\|^2,\ \forall\, x, y \in dom(f),\ \forall\, \alpha \in [0,1]. \tag{17}$$

其中 $m > 0$ 是强凸性参数(*strong convexity parameter*)，并称 f 是强凸的且凸模(*modulus of convexity*)为 m.

Theorems on Convex Functions

对于凸优化问题(目标函数、约束集均为凸的优化问题)，一个重要特征是其局部解也是全局解.

Theorem 4 (Global Optimality of Convex Functions)**.** 对于约束优化问题*(1)*，若 f 是凸函数且 Ω 是闭凸集，则

1. 约束优化问题*(1)*的局部解也是全局解.
2. 由约束优化问题*(1)*全局解构成的集合是凸集.

Proof. **Part 1.** 反证法. 假设 $x^* \in \Omega$ 为局部最优解但非全局最优解，那么存在 $\bar{x} \in \Omega$，使得 $f(\bar{x}) < f(x^*)$. 根据凸性可知 $\forall\, \alpha \in [0,1]$，

$$f(x^* + \alpha(\bar{x} - x^*)) \leqslant \alpha f(\bar{x}) + (1-\alpha)f(x^*) < f(x^*) \tag{18}$$

因此对于 x^* 的邻域 $\mathcal{N}$，都存在足够小的 $\alpha \in [0,1]$，使得 $x^* + \alpha(\bar{x} - x^*) \in \mathcal{N} \cap \Omega$，且 $f(x^* + \alpha(\bar{x} - x^*)) < f(x^*)$，这与 x^* 为局部解矛盾！

Part 2. 设 $\hat{x}, x^*$ 均为全局解，故满足 $f(\hat{x}) = f(x^*)$，那么根据函数 f 的凸性，有

$$f(x^*) \leqslant f(\alpha x^* + (1-\alpha)\hat{x}) \leqslant \alpha f(x^*) + (1-\alpha)f(\hat{x}) = f(x^*) \tag{19}$$

因此得到 $f(\alpha x^* + (1-\alpha)\hat{x}) = f(x^*)$ 也是全局解，故根据定义全局解集合为凸集. □

Lemma 3 (Fundamental Characterization of Convexity of a Smooth Function)**.**
若 f 为连续可微的凸函数，那么

$$f(y) \geqslant f(x) + \nabla f(x)^T(y-x), \forall\, x, y \in dom(f) \tag{20}$$

Proof. 由 Corollary 1 可知 $\forall\, x, y \in \mathrm{dom}(f)$

$$f(x+\alpha(y-x)) = f(x) + \alpha\nabla f(x)^T(y-x) + o(\alpha\|y-x\|) \leqslant (1-\alpha)f(x) + \alpha f(y) \tag{21}$$

将 $o(\alpha\|y-x\|)$ 写为 $o(\alpha)$ 并化简，可以得到

$$f(y) \geqslant f(x) + \nabla f(x)^T(y-x) + o(1) \tag{22}$$

令 $\alpha \downarrow 0$，即可得到

$$f(y) \geqslant f(x) + \nabla f(x)^T(y-x), \forall\, x, y \in \mathrm{dom}(f) \tag{23}$$

□

Theorem 5. 若 f 连续可微且为凸函数，那么若 $\nabla f(x^*) = 0$，则 x^* 是无约束优化问题*(11)*的全局解.

Proof. 设 $\nabla f(x^*) = 0$，根据 Lemma 3，可得

$$f(y) \geqslant f(x^*) + \nabla f(x^*)^T(y-x^*) = f(x^*), \forall\, y \in \mathrm{dom}(f) \tag{24}$$

因此 x^* 为全局解. □

Lemma 4. 若二阶连续可微函数 f 是凸函数，那么 $\nabla^2 f(x) \succeq 0, \forall\, x \in dom(f)$.

Proof. 根据 mean-value form 的 Taylor Theorem 式(3b)可知存在 $\gamma \in (0,1)$，使得

$$f(y) - f(x) = \nabla f(x)^T(y-x) + \frac{1}{2}(y-x)^T\nabla^2 f(x+\gamma(y-x))(y-x),\ \forall\, x, y \tag{25}$$

由 Lemma 3可知

$$f(y) \geqslant f(x) + \nabla f(x)^T(y-x), \forall\, x, y \in \mathrm{dom}(f) \tag{26}$$

结合上式，得

$$(y-x)^T\nabla^2 f(x+\gamma(y-x))(y-x) \geqslant 0, \forall\, x, y \in \mathrm{dom}(f) \tag{27}$$

取 $y = x + \alpha p$，其中 $\alpha > 0$，p 为任意单位向量，得

$$p^T\nabla^2 f(x+\alpha\gamma p)p \geqslant 0 \tag{28}$$

此时取 $\alpha \downarrow 0$，p 为单位特征向量，即得 $\nabla^2 f(x) \succeq 0, \forall\, x \in \mathrm{dom}(f)$. □

Corollary 2. 给定一二阶连续可微的凸函数 f，那么其满足 *L-smooth* 当且仅当 $\forall\, x \in dom(f)$，$0 \preceq \nabla^2 f(x) \preceq LI$.

Proof. (充分性) 已知 $\forall\, x \in \mathrm{dom}(f)$，$0 \preceq \nabla^2 f(x) \preceq LI$，根据 Lemma 2即可知 f 满足 L-smooth.

(必要性) 已知二阶连续可微的凸函数 f 满足 L-smooth. 根据 Lemma 2可知 $\nabla^2 f(x) \preceq LI$. 由于 f 凸，因此 $\nabla^2 f(x) \succeq 0$. □

Theorems on Strongly Convex Functions

Lemma 5 (Equivalent Definition for Strongly Convex Functions)**.** 对于连续可微函数 f，以下条件等价：

1. f 是强凸的，且强凸性参数为 $m > 0$.
2. $f(y) \geqslant f(x) + \nabla f(x)^T(y-x) + \frac{m}{2}\|y-x\|^2,\ \forall\, x, y \in dom(f)$.

Proof. 根据定义可知

$$\begin{aligned} & f((1-\alpha)x+\alpha y) = f(x) + \alpha\nabla f(x)^T(y-x) + o(\alpha\|y-x\|) \\ & \leqslant (1-\alpha)f(x) + \alpha f(y) - \frac{m}{2}\alpha(1-\alpha)\|x-y\|^2,\ \forall\, x, y \in \mathrm{dom}(f),\ \forall\, \alpha \in [0,1] \end{aligned} \tag{29}$$

化简可得

$$f(y) \geqslant f(x) + \nabla f(x)^T(y-x) + \frac{m}{2}\|y-x\|^2 + o(1) \tag{30}$$

令 $\alpha \downarrow 0$，得到

$$f(y) \geqslant f(x) + \nabla f(x)^T(y-x) + \frac{m}{2}\|y-x\|^2 \tag{31}$$

□

Note 5. 该命题说明了连续可微强凸函数 f 在 y 处的下界可以被一个在 y 处函数值为 $f(y)$ 的二次函数控制. 并且与 *Lemma 3* 形式是类似的.

Theorem 6. 对于连续可微且为强凸函数的函数 f，若 $\nabla f(x^*) = 0$，则 x^* 是无约束优化问题*(11)*的唯一全局解.

Proof. 根据 Lemma 5，若 $\nabla f(x^*) = 0$，则

$$f(y) \geqslant f(x) + \frac{m}{2}\|y-x\|^2, \forall\, y \in \mathrm{dom}(f) \tag{32}$$

因此若 $y \neq x^*$，$f(y) > f(x^*)$，即 x^* 为唯一全局解. □

Lemma 6. 对于 $\mathbb{R}^n$ 上二阶连续可微函数 f，那么 f 强凸且凸模为 m 当且仅当 $\forall\, x \in dom(f)$，$\nabla^2 f(x) \succeq mI$.

证明见附录C.

A Proof of Theorem 2

Proof of Theorem 2

Proposition 1.

1. 对于连续可微函数 f，若 x^* 为式*(11)*的局部解，那么 $\nabla f(x^*) = 0$.
2. 对于二次连续可微函数 f，若 x^* 为式*(11)*的局部解，那么 $\nabla f(x^*) = 0$ 且 $\nabla^2 f(x^*)$ 是半正定的*(positive semidefinite)*.

Proof Method 1. **Part 1.** 由 Corollary 1

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + o(\|x - x^*\|). \tag{33}$$

由于 x^* 是局部解，存在邻域 $\mathcal{N}(x^*)$ 使得 $\forall x \in \mathcal{N}(x^*)$，$f(x) \geqslant f(x^*)$，因此

$$f(x) - f(x^*) = \nabla f(x^*)^T (x - x^*) + o(\|x - x^*\|) \geqslant 0,\ \forall x \in \mathcal{N}(x^*). \tag{34}$$

取 $x = x^* + \epsilon p \in \mathcal{N}(x^*)$，其中 p 为任意单位向量，$\epsilon > 0$，则有

$$\epsilon \cdot \nabla f(x^*)^T p + o(\epsilon) \geqslant 0,\ \forall p \in \mathbb{R}^n. \tag{35}$$

由于 $o(\epsilon) \to 0$ 当 $\epsilon \to 0$，因此有

$$\nabla f(x^*)^T p \geqslant 0,\ \forall p \in \mathbb{R}^n. \tag{36}$$

分别取 $p = \nabla f(x^*)/\|\nabla f(x^*)\|$ 和 $p = -\nabla f(x^*)/\|\nabla f(x^*)\|$，则有

$$\|\nabla f(x^*)\|^2 \geqslant 0,\ \text{且}\ -\|\nabla f(x^*)\|^2 \geqslant 0. \tag{37}$$

因此有 $\nabla f(x^*) = 0$.

Part 2. 根据 Part 1，立刻得到 $\nabla f(x^*) = 0$. 只需证明 $\nabla^2 f(x^*)$ 是半正定的.
根据二次 mean-value form 的 Taylor's Theorem 式(3b)，存在 $\gamma \in (0, 1)$ 使得

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x + \gamma(x - x^*))(x - x^*). \tag{38}$$

由于 x^* 是局部解，存在邻域 $\mathcal{N}(x^*)$ 使得 $\forall x \in \mathcal{N}(x^*)$，$f(x) \geqslant f(x^*)$，因此

$$\begin{aligned} f(x) - f(x^*) &= \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x + \gamma(x - x^*))(x - x^*) \\ &\xlongequal{\nabla f(x^*)=0} \frac{1}{2}(x - x^*)^T \nabla^2 f(x + \gamma(x - x^*))(x - x^*) \geqslant 0,\ \forall x \in \mathcal{N}(x^*) \end{aligned} \tag{39}$$

取 $x = x^* + \epsilon p$，其中 p 为任意单位向量，$\epsilon > 0$，则有

$$\frac{1}{2}\epsilon^2 p^T \nabla^2 f(x^* + \gamma \epsilon p) p \geqslant 0,\ \forall p \in \mathbb{R}^n. \tag{40}$$

令 $\epsilon \to 0$，则有

$$p^T \nabla^2 f(x^*) p \geqslant 0, \, \forall p \in \mathbb{R}^n. \tag{41}$$

取 p 为任意单位特征向量，记 λ_p 为相应特征值，则有

$$\lambda_p p^T p = \lambda_p \|p\|^2 \geqslant 0 \Rightarrow \lambda_p \geqslant 0. \tag{42}$$

因此有 $\nabla^2 f(x^*)$ 是半正定的(positive semidefinite)，即 $\nabla^2 f(x^*) \succeq 0$. □

Proof Method 2. **Part 1.** 反证法. 假设 $\nabla f(x^*) \neq 0$，考虑 $f(x^* - \alpha \nabla f(x^*)), \alpha > 0$，使用 mean-value form Taylor Theorem 式(2b)可得

$$f(x^* - \alpha \nabla f(x^*)) = f(x^*) - \alpha \nabla f(x^* - \gamma \alpha \nabla f(x^*))^T \nabla f(x^*) \text{ for some } \gamma \in (0,1) \tag{43}$$

根据 ∇f 的连续性，对于足够小的 α 有

$$\nabla f(x^* - \gamma \alpha \nabla f(x^*))^T \nabla f(x^*) \geqslant \frac{1}{2} \|\nabla f(x^*)\|^2, \, \forall \, \gamma \in (0,1) \tag{44}$$

代入式(43)中，得对于足够小的 α，有

$$\nabla f(x^* - \gamma \alpha \nabla f(x^*)) \leqslant f(x^*) - \frac{\alpha}{2} \|\nabla f(x^*)\|^2 < f(x^*) \tag{45}$$

因此无论如何选取邻域 $\mathcal{N}(x^*)$，都可以选取足够小的 α 使得 $x^* - \gamma \alpha \nabla f(x^*) \in \mathcal{N}(x^*)$ 且 $\nabla f(x^* - \gamma \alpha \nabla f(x^*)) < f(x^*)$，故 $f(x^*)$ 并非局部解，矛盾！

Part 2. 根据 Part 1，立刻得到 $\nabla f(x^*) = 0$. 只需证明 $\nabla^2 f(x^*)$ 是半正定的.

反证法. 假设 $\nabla^2 f(x^*)$ 存在负特征值，因此 $\exists \, v \in \mathbb{R}^n, \lambda > 0$，使得 $v^T \nabla^2 f(x^*) v \leqslant -\lambda$. 根据二次 mean-value form 的 Taylor′s Theorem 式(3b)，存在 $\gamma \in (0,1)$ 使得

$$f(x^* + \alpha v) = f(x^*) + \alpha \nabla f(x^*)^T v + \frac{\alpha^2}{2} v^T \nabla^2 f(x + \gamma (x^* + \gamma \alpha v)) v. \tag{46}$$

根据二次可微性可知对于足够小的 α 有

$$v^T \lambda^2 f(x^* + \gamma \alpha v) v \leqslant -\frac{\lambda}{2}, \, \forall \, \gamma \in (0,1) \tag{47}$$

代入上式，结合 $\nabla f(x^*) = 0$ 可得

$$f(x^* + \alpha v) \leqslant f(x^*) - \frac{1}{4} \alpha^2 \lambda < f(x^*) \tag{48}$$

同理，x^* 非局部解，矛盾！ □

B Proof of Lemma 2

Proof of Lemma 2

Proposition 2. 给定一二次可微函数 f，若其 *L-smooth*，则

$$\nabla^2 f(x) \preceq LI, \, \forall x \in dom(f). \tag{49}$$

其中 I 为单位矩阵. 反之，若 $-LI \preceq \nabla^2 f(x) \preceq LI$，则 f 是 *L-smooth* 的.

Proof. **Part 1:** 根据 Lemma 1，令 $y = x + \alpha p, \alpha > 0$，则

$$f(x+\alpha p) \leqslant f(x) + \nabla f(x)^T(\alpha p) + \frac{L}{2}\|\alpha p\|^2 = f(x) + \alpha \nabla f(x)^T p + \frac{L\alpha^2}{2}\|p\|^2. \tag{50}$$

由二次 mean-value form 的 Taylor's Theorem 式(3b)，存在 $\gamma \in (0,1)$ 使得

$$f(x+\alpha p) - f(x) - \alpha \nabla f(x)^T p = \frac{\alpha^2}{2} p^T \nabla^2 f(x+\gamma\alpha p) p. \tag{51}$$

结合上述两式可得

$$p^T \nabla^2 f(x+\gamma\alpha p) p \leqslant L\|p\|^2,\ \forall p \in \mathbb{R}^n,\ \forall \alpha > 0. \tag{52}$$

取 $\alpha \downarrow 0$，p 为任意特征向量，记 λ_p 为相应特征值，则

$$\lambda_p p^T p = \lambda_p \|p\|^2 \leqslant L\|p\|^2 \Rightarrow \lambda_p \leqslant L \Rightarrow \nabla^2 f(x) \preceq LI \tag{53}$$

Part 2: 已知 $-LI \preceq \nabla^2 f(x) \preceq LI$，那么 $\forall\, x$，$\|\nabla^2 f(x)\| \leqslant L$. 根据二次 integral form 的 Taylor's Theorem 式(3a)，

$$\nabla f(y) = \nabla f(x+y-x) = \nabla f(x) + \int_0^1 \nabla^2 f(x+\gamma(y-x))(y-x)\,\mathrm{d}\gamma \tag{54}$$

因此有

$$\begin{aligned} &\|\nabla f(y) - \nabla f(x)\| = \|\int_0^1 \nabla^2 f(x+\gamma(y-x))(y-x)\,\mathrm{d}\gamma\| \\ &\leqslant \int_0^1 \|\nabla^2 f(x+\gamma(y-x))\|\|(y-x)\|\,\mathrm{d}\gamma \leqslant \int_0^1 L\|y-x\|\,\mathrm{d}\gamma = L\|y-x\| \end{aligned} \tag{55}$$

□

C Proof of Lemma 6

Proof of Lemma 6

Proposition 3. 对于 $\mathbb{R}^n$ 上二阶连续可微函数 f，那么 f 强凸且凸模为 m 当且仅当 $\forall\, x \in$ *dom*(f)，$\nabla^2 f(x) \succeq mI$.

Proof. (必要性) 对于二阶连续可微强凸函数 f，由 mean value form 的 Taylor's Theorem 式(3b)可知 $\forall x, u \in \mathbb{R}^n$，$\forall\, \alpha > 0$，有

$$f(x+\alpha u) = f(x) + \alpha \nabla f(x)^T u + \frac{1}{2}\alpha^2 u^T \nabla^2 f(x+\gamma\alpha u) u,\ \text{for some } \gamma \in (0,1) \tag{56}$$

由强凸性可得

$$f(x+\alpha u) \geqslant f(x) + \alpha \nabla f(x)^T u + \frac{m}{2}\alpha^2 \|u\|^2 \tag{57}$$

结合上述两式，得

$$u^T \nabla^2 f(x+\gamma\alpha u) \geqslant m\|u\|^2 \tag{58}$$

令 $\alpha \downarrow 0$，得 $u^T \nabla^2 f(x) \geqslant m\|u\|^2$，即得 $\nabla^2 f(x) \succeq mI$.

(充分性) 已知 $\nabla^2 f(x) \succeq mI, \forall\, x$，同样由 mean value form 的 Taylor's Theorem 式(3b)可得

$$f(z) = f(x) + \nabla f(x)^T(z-x) + \frac{1}{2}(z-x)^T \nabla^2 f(z+\gamma(z-x))(z-x), \text{ for some } \gamma \in (0,1) \tag{59}$$

由 $\nabla^2 f(x) \succeq mI, \forall\, x$，则

$$(z-x)^T \nabla^2 f(z+\gamma(z-x))(z-x) \geqslant m\|z-x\|^2 \tag{60}$$

因此得到

$$f(z) \geqslant f(x) + \nabla f(x)^T(z-x) + \frac{m}{2}\|z-x\|^2, \forall\, x, z \in \mathrm{dom}(f) \tag{61}$$

此即强凸性的等价定义. □

Last updated: 12 July, 2025

References