APM462: Nonlinear Optimization

Tianyu Du

September 16, 2019

Contents

1 Preliminaries

1.1 Mean Value Theorems and Taylor Approximations.

Definition 1.1. Let $f : S \subset \mathbb{R}^n \to \mathbb{R}$, the **gradient** of f at $x \in S$, if exists, is a vector $\nabla f(x) \in \mathbb{R}^n$ characterized by the property

$$\lim_{v \to 0} \frac{f(x+v) - f(x) - \nabla f(x) \cdot v}{||v||} = 0 \tag{1.1}$$

Theorem 1.1 (The First Order of Mean Value Theorem)**.** Let f be a C^1 real-valued function defined on $\mathbb{R}^n$, then for any $x, v \in \mathbb{R}^n$, there exists some $\theta \in (0,1)$ such that

$$f(x+v) = f(x) + \nabla f(x + \theta v) \cdot v \tag{1.2}$$

Proof. Let $x, v \in \mathbb{R}^n$, define $g(t) : \mathbb{R} \to \mathbb{R} := f(x+tv)$, which is C^1. By the mean value theorem on $\mathbb{R}^{\mathbb{R}}$, there exists $\theta \in (0,1)$ such that $g(0+1) = g(0) + g'(\theta)(1-0)$, that is, $f(x+v) = f(x) + g'(\theta)$. Note that $g'(\theta) = \nabla(x + \theta v) \cdot v$, what desired is immediate. ∎

Proposition 1.1 (The First Order Taylor Approximation)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function, then

$$f(x+v) = f(x) + \nabla f(x) \cdot v + o(||v||) \tag{1.3}$$

that is

$$\lim_{||v|| \to 0} \frac{f(x+v) - f(x) - \nabla f(x) \cdot v}{||v||} = 0 \tag{1.4}$$

Proof. By the mean value theorem, $\exists \theta \in (0,1)$ such that $f(x+v) - f(x) = \nabla f(x + \theta v) \cdot v$. The limit becomes $\lim_{||v|| \to 0} \frac{[\nabla f(x+\theta v) - \nabla f(x)] \cdot v}{||v||} = \lim_{||v|| \to 0; x+\theta v \to x} \frac{[\nabla f(x+\theta v) - \nabla f(x)] \cdot v}{||v||}$. Since $f \in C^1$, $\lim_{x+\theta v \to x} \nabla f(x+\theta v) = \nabla f(x)$. And $\frac{v}{||v||}$ is a unit vector, and every component of it is bounded, as the result, the limit of inner product vanishes instead of explodes. ∎

Theorem 1.2 (The Second Order Mean Value Theorem)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2 function, then for any $x, v \in \mathbb{R}^n$, there exists $\theta \in (0,1)$ satisfying

$$f(x+v) = f(x) + \nabla f(x) \cdot v + \frac{1}{2} v' H_f(x + \theta v)\ v \tag{1.5}$$

where H_f is the Hessian matrix of f, may also be written as $\nabla^2 f$.

Proposition 1.2 (The Second Order Taylor Approximation)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2 function, and $x, v \in \mathbb{R}^n$, then

$$f(x+v) = f(x) + \nabla f(x) \cdot v + \frac{1}{2} v' H_f(x)\ v + o(||v||^2) \tag{1.6}$$

that is

$$\lim_{||v|| \to 0} \frac{f(x+v) - f(x) - \nabla f(x) \cdot v - \frac{1}{2} v' H_f(x)\ v}{||v||^2} = 0 \tag{1.7}$$

Proof. By the second mean value theorem, there exists $\theta \in (0,1)$ such that the limit is equivalent to

$$\lim_{||v|| \to 0} \frac{1}{2} \left(\frac{v}{||v||} \right)' [H_f(x + \theta v) - H_f(x)] \frac{v}{||v||} \tag{1.8}$$

Since $f \in C^2$, the limit of $[H_f(x+\theta v) - H_f(x)]$ is in fact $\mathbf{0}_{n \times n}$. And every component of unit vector $\frac{v}{||v||}$ is bounded, the quadratic form converges to zero as an immediate result. ∎

It is often noted that the gradient at a particular $x_0 \in dom(f) \subset \mathbb{R}^n$ gives the direction f increases most rapidly. Let $x_0 \in dom(f)$, and v be a unit vector representing a *feasible direction* of change. That is, there exists $\delta > 0$ such that $x_0 + tv \in dom(f)\ \forall t \in [0, \delta)$. Then the rate of change of f along feasible direction v can be written as

$$\frac{d}{dt}\bigg|_{t=0} f(x_0 + tv) = \nabla f(x_0) \cdot v = ||\nabla f(x_0)||\ ||v|| \cos(\theta) \tag{1.9}$$

where $\theta = \angle(v, \nabla f(x_0)$. And the derivative is maximized when $\theta = 0$, that is, when v and ∇f point the same direction.

1.2 Implicit Function Theorem

Theorem 1.3 (Implicit Function Theorem)**.** Let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be a C^1 function, let $(a, b) \in \mathbb{R}^n \times \mathbb{R}$ such that $f(a, b) = 0$. If $\nabla f(a, b) \neq 0$, then $\{(x, y) \in \mathbb{R}^n \times \mathbb{R} : f(x, y) = 0\}$ is locally a graph of a function $g : \mathbb{R}^n \to \mathbb{R}$.

Remark 1.1. $\nabla f(x_0) \perp$ level set of f near x_0.

2 Convexity

2.1 Terminologies

Definition 2.1. Set $\Omega \subset \mathbb{R}^n$ is **convex** if and only if

$$\forall x_1, x_2 \in \Omega,\ \lambda \in [0, 1],\ \lambda x_1 + (1 - \lambda)x_2 \in \Omega \tag{2.1}$$

Definition 2.2. A function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is **convex** if and only if Ω is convex, and

$$\forall x_1, x_2 \in \Omega,\ \lambda \in [0, 1],\ f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{2.2}$$

Definition 2.3. A function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is **strictly convex** if and only if Ω is convex and

$$\forall x_1, x_2 \in \Omega,\ \lambda \in (0, 1),\ f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{2.3}$$

2.2 Basic Properties of Convex Functions

Definition 2.4. A function $f : \Omega \to \mathbb{R}$ is **concave** if and only if $-f$ is **convex**.

Proposition 2.1. (i) If f_1, f_2 are convex on Ω, so is $f_1 + f_2$;

(ii) If f is convex on Ω, then for any $a > 0$, af is also convex on Ω;

(iii) Any **sub-level/lower contour set** of a convex function f

$$SL(c) := \{x \in \mathbb{R}^n : f(x) \leq c\} \tag{2.4}$$

is convex.

Proof of (iii). Let $c \in \mathbb{R}$, and $x_1, x_2 \in SL(c)$. Let $s \in [0, 1]$. Since $x_1, x_2 \in SL(c)$, and $f(\cdot)$ is convex, $f(sx_1 + (1 - s)x_2) \leq sf(x_1) + (1 - s)f(x_2) \leq sc + (1 - s)c = c$. Which implies $sx_1 + (1 - s)x_2 \in SL(c)$. ■

Example 2.1. $f(x) : \mathbb{R}^n \to \mathbb{R} := ||x||$ is convex.

Proof. Note that for any $u, v \in \mathbb{R}^n$, by triangle inequality, $||u-(-v)|| \leq ||u-0|| + ||0-(-v)|| = ||u|| + ||v||$. Consequently, let $u, v \in \mathbb{R}^n$ and $s \in [0,1]$, then $||su+(1-s)v|| \leq ||su|| + ||(1-s)v|| = s||u|| + (1-s)||v||$. Therefore, $||\cdot||$ is convex. ■

2.3 Characteristics of C^1 Convex Functions

Theorem 2.1 (C^1 criterions for convexity)**.** Let $f \in C^1$, then f is convex on a convex set Ω if and only if

$$\forall x, y \in \Omega,\ \ f(y) \geq f(x) + \nabla f(x) \cdot (y-x) \tag{2.5}$$

that is, *the linear approximation is never an overestimation of value of f.*

Proof. ($\Longrightarrow$) Suppose f is convex on a convex set Ω. Then $f(sy+(1-s)x) \leq sf(y)+(1-s)f(x)$ for every $x, y \in \Omega$ and $s \in [0,1]$, which implies, for every $s \in (0,1]$:

$$\frac{f(sy+(1-s)x)-f(x)}{s} \leq f(y) - f(x) \tag{2.6}$$

By taking the limit of $s \to 0$,

$$\lim_{s\to 0} \frac{f(x+s(y-x))-f(x)}{s} \leq f(y)-f(x) \tag{2.7}$$

$$\Longrightarrow \ \frac{d}{ds}\bigg|_{s=0} f(x+s(y-x)) \leq f(y)-f(x) \tag{2.8}$$

$$\Longrightarrow \ \nabla f(x)\cdot(y-x) \leq f(y)-f(x) \tag{2.9}$$

($\Longleftarrow$) Let $x_0, x_1 \in \Omega$, let $s \in [0,1]$. Define $x^* := sx_0 + (1-s)x_1$, then

$$f(x_0) \geq f(x^*) + \nabla f(x^*)\cdot(x_0 - x^*) \tag{2.10}$$

$$\Longrightarrow \ f(x_0) \geq f(x^*) + \nabla f(x^*)\cdot[(1-s)(x_0-x_1)] \tag{2.11}$$

Similarly,

$$f(x_1) \geq f(x^*) + \nabla f(x^*)\cdot(x_1 - x^*) \tag{2.12}$$

$$\Longrightarrow \ f(x_1) \geq f(x^*) + \nabla f(x^*)\cdot[s(x_1-x_0)] \tag{2.13}$$

Therefore, $sf(x_0)+(1-s)f(x_1) \geq f(x^*)$. ■

Theorem 2.2 (C^2 criterion for convexity)**.** $f \in C^2$ is a convex function on a convex set $\Omega \subset \mathbb{R}^n$ if and only if $\nabla^2 f(x) \succcurlyeq 0$ for all $x \in \Omega$.

Remark 2.1. When f is defined on $\mathbb{R}$, the C^2 criterion becomes $f''(x) \geq 0$.

Proof. ($\Longleftarrow$) Suppose $\nabla^2 f(x) \succcurlyeq 0$ for every $x \in \Omega$, let $x, y \in \Omega$. By the second order MVT,

$$f(y) = f(x) + \nabla f(x)\cdot(y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(x+s(y-x))(y-x) \text{ for some } s \in [0,1] \tag{2.14}$$

$$\Longrightarrow \ f(y) \geq f(x) + \nabla f(x)\cdot(y-x) \tag{2.15}$$

So f is convex by the C^1 criterion of convexity.
($\Longrightarrow$) Let $v \in \mathbb{R}^n$. Suppose, for contradiction, that for some $x \in \Omega$, $\nabla^2 f(x) \not\succcurlyeq 0$. If such $x \in \partial\Omega$, note that $v^T\nabla^2 f(\cdot)v$ is continuous because $f \in C^2$, then there exists $\varepsilon > 0$ such that $\forall x' \in V_\varepsilon(x) \cap \Omega^{int}$, $v^T \nabla^2 f(x')v <$

0. Hence, one may assume with loss of generality that such $x \in \Omega^{int}$. Because $x \in \Omega^{int}$, exists $\varepsilon' > 0$, such that $V_{\varepsilon'}(x) \subseteq \Omega^{int}$. Define $\hat{v} := \frac{v}{\sqrt{\varepsilon'}}$, then for every $s \in [0,1]$, $\hat{v}^T \nabla^2 f(x + s\hat{v})\hat{v} < 0$. Let $y = x + \hat{v}$, by the mean value theorem, $f(y) = f(x) + \nabla f(x) \cdot (y - x) + \frac{1}{2}(y-x)^T \nabla^2 f(x + s(y-x))(y-x)$ for some $s \in [0,1]$. This implies $f(y) < f(x) + \nabla f(x) \cdot (y - x)$, which contradicts the C^1 criterion for convexity. ■

2.4 Minimum and Maximum of Convex Functions

Theorem 2.3. Let $\Omega \subset \mathbb{R}^n$ be a convex set, and $f : \Omega \to \mathbb{R}$ is a convex function. Let

$$\Gamma := \left\{ x \in \Omega : f(x) = \min_{x \in \Omega} f(x) \right\} \equiv \operatorname*{argmin}_{x \in \Omega} f(x) \tag{2.16}$$

If $\Gamma \neq \varnothing$, then

(i) Γ is convex;

(ii) any local minimum of f is the global minimum.

Proof (i). Let $x, y \in \Gamma$, $s \in [0,1]$, then $sx + (1-s)y \in \Omega$ because Ω is convex. Since f is convex, $f(sx + (1-s)y) \leq sf(x) + (1-s)f(y) = \min_{x \in \Omega} f(x)$. The inequality must be equality since it would contradicts the fact that $x, y \in \Gamma$. Therefore, $sx + (1-s)y \in \Gamma$. ■

Proof (ii). Let $x \in \Omega$ be a local minimizer for f, but assume, for contradiction, it is not a global minimizer. That is, there exists some other y such that $f(y) < f(x)$. Since f is convex,

$$f(x + t(y-x)) = f((1-t)x + ty) \leq (1-t)f(x) + tf(y) < f(x) \tag{2.17}$$

for every $t \in (0,1]$. Therefore, for every $\varepsilon > 0$, there exists $t^* \in (0,1]$ such that $x + t^*(y-x) \in V_\varepsilon(x)$ and $f(x + t^*(y-x)) < f(x)$, this contradicts the fact that x is a local minimum. ■

Theorem 2.4. Let $\Omega \subset \mathbb{R}^n$ be a convex set, and $f : \Omega \to \mathbb{R}$ is a convex function. Then

$$\max_{x \in \Omega} f(x) = \max_{x \in \partial\Omega} f(x) \tag{2.18}$$

Proof. As we assumed, Ω is closed, therefore $\partial\Omega \subseteq \Omega$. Hence, $\max_{x \in \Omega} f \geq \max_{x \in \partial\Omega} f$. Suppose $\max_{x \in \Omega} f > \max_{x \in \partial\Omega} f$, let $x^* := \operatorname{argmax}_{x \in \Omega} f \in \Omega^{int}$. Then we can construct a string line through x^* and intersects $\partial\Omega$ at two points, $y_1, y_2 \in \partial\Omega$, such that $x^* = sy_1 + (1-s)y_2$ for some $s \in (0,1)$. Further, since f is convex, $f(x^*) \leq sf(y_1) + (1-s)f(y_2) \leq s \max_{\partial\Omega} f + (1-s)\max_{\partial\Omega} f = \max_{\partial\Omega} f$, which leads to a contradiction. ■

Proposition 2.2. Let f be continuous function on $\mathbb{R}^n$, then the level set and sub-level set of f are closed. Further, since the intersection of arbitrary closed sets is closed, so the set defined by collection of continuous functions $\{h_i; g_j\}$

$$\{x \in \mathbb{R}^n : h_i(x) = 0 \wedge g_j(x) \leq 0 \ \forall i, j\} \tag{2.19}$$

is closed.

Theorem 2.5 (Extreme Value Theorem)**.** Let $\Omega \subset \mathbb{R}^n$ be a compact set, and $f : \Omega \to \mathbb{R}$ is continuous, then f attains its maximum and minimum on Ω.

Corollary 2.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function, if there exists $a \in \mathbb{R}^n$ such that $f(x) \geq f(a)$ for every $x \notin \mathcal{B}(r,a)$, then f attains its minimum in $\mathcal{B}(r,a)$.

Proposition 2.3. When $\Omega = \mathbb{R}^n$, the unconstrained minimization has the following properties

(i) $\operatorname{argmax} f = \operatorname{argmin}(-f)$;

(ii) $\max f = -\min(-f)$