Optimization

Wenxiao Yang[*]

[*]Department of Mathematics, University of Illinois at Urbana-Champaign

2022

目录

1 Unconstrained Optimization

1.1 Conditions for Optimality

Function: $f : \mathbb{R}^n \to \mathbb{R}^n$, $x \in \&,\ \& \subseteq \mathbb{R}^n$.
Terminology: x^* will always be the optimal input at some function.

1.2 Global minimizer, Local minimizer

Definition 1.
Say x^* is a global minimizer(minimum) of f if $f(x^*) \leq f(x), \forall x \in \&$.
Say x^* is a unique global minimizer(minimum) of f if $f(x^*) < f(x), \forall x \neq x^*$.
Say x^* is a local minimizer(minimum) of f if $\exists r > 0$ so that $f(x^*) \leq f(x)$ when $\|x - x^*\| < r$.

A minimizer is strict if $f(x^*) < f(x)$ for all relevant x.

1.3 Optimization in $\mathbb{R}$

1.3.1 Theorem: local minimizer $\Rightarrow f'(x^*) = 0$

Theorem 1. *If $f(x)$ is differentiable function and x^* is a local minimizer, then $f'(x^*) = 0$.*

证明.
Def of $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$
Def of local minimizer: $f(x^*) - f(x) \geq 0, |x^* - x| < r$
when $0 < h < r$, $\frac{f(x+h)-f(x)}{h} \geq 0$; when $-r < h < 0$, $\frac{f(x+h)-f(x)}{h} \leq 0$. Then $f'(x) = 0$. □

1.3.2 Theorem: $f'(x^*) = 0, f''(x^*) \geq 0 \Rightarrow$ local minimizer

Theorem 2. *If $f : \mathbb{R} \to \mathbb{R}$ is a function with a continuous second derivative and x^* is a critical point of f (i.e. $f'(x) = 0$), then:*
(1): If $f''(x) \geq 0,\ \forall x \in \mathbb{R}$, then x^* is a global minimizer on $\mathbb{R}$.
(2): If $f''(x) \geq 0,\ \forall x \in [a, b]$, then x^* is a global minimizer on $[a, b]$.
(3): If we only know $f''(x^*) \geq 0$, x^* is a local minimizer.

证明.
(1)$f(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2} f''(\xi)(x - x^*)^2 = f(x^*) + 0 + \textit{something non negative} \geq f(x^*)\ \forall x$
(2) Similar to (1)
(3)$f''(x^*) \geq 0,\ f''$ continuous $\Rightarrow \exists r$ s.t. $f''(x) \geq 0\ \forall x \in [x^* - \frac{r}{2}, x^* + \frac{r}{2}]$, then x is a local minimizer. □

1.4 Optimization in $\mathbb{R}^n$

1.4.1 Necessary Conditions for Optimality: Local Extremum $\Rightarrow \nabla f(x^*) = 0$

A base point x, we consider an arbitrary direction u. $\{x + tu | t \in \mathbb{R}\}$

For $\alpha > 0$ sufficiently small:

1. $f(x^*) \leq f(x^* + \alpha u)$

2. $g(\alpha) = f(x^* + tu) - f(x^*) \geq 0$

3. $g(\beta)$ is continuously differentiable for $\beta \in [0, \alpha]$

By chain rule,

$$g'(\beta) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^* + \beta u)u_i$$

By Mean Value Theorem,

$$g(\alpha) = g(0) + g'(\beta)\alpha \text{ for some } \beta \in [0, \alpha]$$

Thus

$$g(\alpha) = \alpha \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^* + \beta u)u_i \geq 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^* + \beta u)u_i \geq 0$$

Letting $\alpha \to 0$ and hence $\beta \to 0$, we get

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^*)u_i \geq 0 \text{ for all } u \in \mathbb{R}^n$$

By choosing $u = [1, 0, ..., 0]^T$, $u = [-1, 0, ..., 0]^T$, we get

$$\frac{\partial f(x^*)}{\partial x_1} \geq 0, \ \frac{\partial f(x^*)}{\partial x_1} \leq 0 \Rightarrow \frac{\partial f(x^*)}{\partial x_1} = 0$$

Similarly, we can get

$$\nabla f(x^*) = [\frac{\partial f(x^*)}{\partial x_1}, \frac{\partial f(x^*)}{\partial x_2}, ..., \frac{\partial f(x^*)}{\partial x_n}]^T = 0$$

Theorem 3. *If f is continuously differentiable and x^* is a local extremum. Then $\nabla f(x^*) = 0$.*

1.4.2 Stationary Point, Saddle Point

All points x^* s.t. $\nabla f(x^*) = 0$ are called stationary points.
Thus, all extrema are stationary points.
But not all stationary points have to be extrema.
Saddle points are the stationary points neither local minimum nor local maximum.

Example 1. *$f(x) = x^3$, $x = 0$ is a stationary point but not extrema. (saddle point)*

1.4.3 Second Order Necessary Condition

Definition 2. *The Hessian of f at point x is an $n \times n$ symmetric matrix denoted by $\nabla^2 f(x)$ with* $[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$

Theorem 4. *Suppose f is twice continuously differentiable and x^* in local <u>minimum</u>. Then*

$$\nabla f(x^*) = 0 \text{ and } \nabla^2 f(x^*) \geq 0$$

证明.

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

Let α be small enough so that $g(\alpha) = f(x^* + \alpha u) - f(x^*) \geq 0$.

By Taylor series expansion,

$$g(\alpha) = g(0) + \alpha g'(0) + \frac{\alpha^2}{2} g''(0) + O(\alpha^2)$$

$$g'(\alpha) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^* + \beta u) u_i = \nabla f(x^* + \alpha u)^T u$$

$$g''(\alpha) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^* + \beta u) u_i u_j = u^T \nabla^2 f(x^* + \alpha u) u$$

$$\begin{aligned} g'(0) = \nabla f(x^*)^T u = 0;\ g''(0) &= u^T \nabla^2 f(x^*) u \\ g(\alpha) &= \frac{\alpha^2}{2} u^T \nabla^2 f(x^*) u + O(\alpha^2) \geq 0 \\ \text{When } \alpha \to 0, \text{ we get } u^T \nabla^2 f(x^*) u &\geq 0,\ \forall u \in \mathbb{R}^n \\ &\Rightarrow \nabla^2 f(x^*) \geq 0 \end{aligned}$$

□

1.4.4 Sufficient Conditions for Optimality

Theorem 5. *Suppose f is twice continuously differentiable in a neighborhood of x^* and (1) $\nabla f(x^*) = 0$; (2) $\nabla^2 f(x^*)$ is PD ($u^T \nabla^2 f(x^*) u > 0$, $\forall u \in \mathbb{R}^n$). Then x^* is local minimum.*

证明.

Consider $u \in \mathbb{R}^n$, $\alpha > 0$ and let

$$\begin{aligned} g(\alpha) &= f(x^* + \alpha u) - f(x^*) \\ &= \frac{\alpha^2}{2} u^T \nabla^2 f(x^*) u + O(\alpha^2) \geq 0 \\ &= \frac{\alpha^2}{2} [u^T \nabla^2 f(x^*) u + 2\frac{O(\alpha^2)}{\alpha^2}] \\ & u^T \nabla^2 f(x^*) u > 0;\ \frac{O(\alpha^2)}{\alpha^2} \to 0 \\ &\Rightarrow g(\alpha) > 0 \text{ for } \alpha \text{ sufficiently small for all } u \neq 0 \\ &\Rightarrow x^* \text{ is local minimum.} \end{aligned}$$

(specially if $\|u\| = 1$, $u^T \nabla^2 f(x^*) u \geq \lambda_{\min}(\nabla^2 f(x^*))$, $\lambda_{\min}(\nabla^2 f(x^*))$ is the minimal eigenvalues of $\nabla^2 f(x^*)$.) □

1.4.5 Using Optimality Conditions to Find Minimum

1. Find all points satisfying necessary condition $\nabla f(x) = 0$ (all stationary points)
2. Filter out points that don't satisfy $\nabla^2 f(x) \geq 0$
3. Points with $\nabla^2 f(x) > 0$ are strict local minimum.
4. Among all points with $\nabla^2 f(x) \geq 0$, declare a global minimum, one with the smallest value of f, assuming that global minimum exists.

Example 2. $f(x) = 2x^2 - x^4$

$$
\begin{aligned}
f'(x) &= 4x - 4x^3 = 0 \\
&\Rightarrow x = 0, x = 1, x = -1 \text{ are stationary points} \\
f''(x) &= 4 - 12x^2 = \begin{cases} 4 & \text{if } x = 0 \\ -8 & \text{if } x = 1, -1 \end{cases} \\
&\Rightarrow x = 0 \text{ is the only local min, and it is strict}
\end{aligned}
$$

But $-f(x) \to \infty$ as $|x| \to \infty \Rightarrow$ no global min, but global max exists. $f(1), f(-1)$ are strict local max and both global max.

2 MATH 484

A base point x, we consider an arbitrary direction u. $\{x + tu | t \in \mathbb{R}\}$
We define the restriction of f to the line through x in the direction of u to be the function:

$$\phi_u(t) = f(x + tu)$$

Lemma 1. *x^* is a global minimizer of f iff for all u, $t = 0$ is the global minimizer of $\phi_u(t)$*

证明.
$(\Rightarrow)$ $\phi_u(0) = f(x^*) \leq f(x^* + tu) = \phi_u(t)$
$(\Leftarrow)$ Let $X \in \mathbb{R}^n$, $u = X - x^*$. $\phi_u(0) \leq \phi_u(1) \Rightarrow f(x^*) \leq f(x^* + u) = f(x)$ □

2.0.1 The first-derivative test in $\mathbb{R}^n$: $\phi'_u(t) = \nabla f(x + tu) \cdot u$

First derivative of $f : \mathbb{R}^n \to \mathbb{R}$, Easier: $\phi'_u(t)$?
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathsf{g} : \mathbb{R} \to \mathbb{R}^n$:

$$\frac{\partial f(\mathsf{g}(t))}{\partial t} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathsf{g}(t)) \frac{d}{dt} g_i(t)$$

$$\frac{\partial \phi_u(t)}{\partial t} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x+tu)u_i$$

The gradient of f: $\nabla f(x) = (\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_d})^T \Rightarrow \phi_u'(t) = \nabla f(x+tu) \cdot u$

Fine print: Chain rule only works when all $\frac{\partial f}{\partial x_k}$ exists and are continuous.

Example 3. $f(x,y) = x^2 + 3xy - 1,\ x^* = (0,0),\ u = (3,2)$

$\phi_u(t) = f(x^* + tu) = f(3t, 2t) = 27t^2 - 1$

$\phi_u'(t) = 54t$

$\nabla f(x,y) = (2x+3y, 3x)$

$\phi_u'(t) = \nabla f(x+tu) \cdot u = 54t$

2.0.2 Theorem 4: ∇f is continuous, x^* is a global minimizer of f $\Rightarrow \nabla f(x^*) = 0$

Theorem 6 (Theorem 2.1)**.** *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, if ∇f is continuous and x^* is a global minimizer of f, then $\nabla f(x^*) = 0$. (When $\nabla f(x^*) = 0$, we call x^* a critical point of f.)*

x^* is a global minimizer $\Rightarrow$ x^* is a critical point, inverse may not true.

2.0.3 The second-derivative test in $\mathbb{R}^n$

$$\phi_u'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x+tu)u_i$$

$$\phi_u''(t) = \sum_{i=1}^{n}\sum_{j=1}^{n} u_i u_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x+tu)$$

2.0.4 Hessian matrix

Define Hessian matrix of f and write Hf. That is,

$$\phi_u''(t) = u^T Hf(x+tu)u$$

Fine print: Chain rule only works when all $\frac{\partial^2 f}{\partial x_i \partial x_j}$ exists and are continuous. ($\Rightarrow Hf$ is continuous)

2.0.5 Theorem 5: Hf is continuous, $\nabla f(x^*) = 0$, $u^T Hf(x^*)u \geq 0, \forall u$ $\Rightarrow$ x^* is a global minimizer of f

Theorem 7. *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, if Hf is continuous and x^* is a critical point of f. If for any u, that $u^T Hf(x^*)u \geq 0$. Then x^* is a global minimizer of f.*

proved by Taylor

Theorem 8 (Taylor)**.** *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, if Hf is continuous and x^* is a critical point of f, then*

$$f(x) = f(x^*) = \nabla f(x^*)(x - x^*) + \frac{1}{2}(x - x^*)^T Hf(z)(x - x^*)$$

for some z on the line between x and x^*

2.1 Minimizing over other sets

What if the domain of f: $D \subset \mathbb{R}^n$

(1): want x^* to be in the interior of D, not on the boundary (want to be able to "look" from x^* in any direction.)

(2): want x^* to "see" all other points in D using straight line u.

Convexity

good domain e.g. Ball: $B(x^*, r) = \{x|||x - x^*|| < r\}$

2.1.1 Theorem 7: ∇f is continuous, x^*(interior of D) is a local minimizer of $f \Rightarrow \nabla f(x^*) = 0$

Theorem 9 (Theorem 4.1, 类似 Theorem 2.1)**.** *Suppose $f : D \to \mathbb{R}$ has continuous ∇f and x^* is not on the boundary of D. If x^* is a local minimizer of f, then x^* is a critical point of f: $\nabla f(x^*) = 0$*

2.1.2 Theorem 8: Hf is continuous, x^*(interior of D) $\nabla f(x^*) = 0$, $\exists r$ s.t. $u^T Hf(x^*)u \geq 0, \forall x \in B(x^*, r), \forall u \Rightarrow x^*$ is a local minimizer of f

Theorem 10. *Given a function $f : D \to \mathbb{R}$, if Hf is continuous and x^* is a critical point of f in the interior of D. Suppose $\exists r$ s.t. for any u, that $u^T Hf(x^*)u \geq 0$ whenever $x \in B(x^*, r) \subset D$. Then x^* is a local minimizer of f.*