Numerical Optimization

Lecture 11: Optimality Conditions (Unconstrained Optimization)

王浩

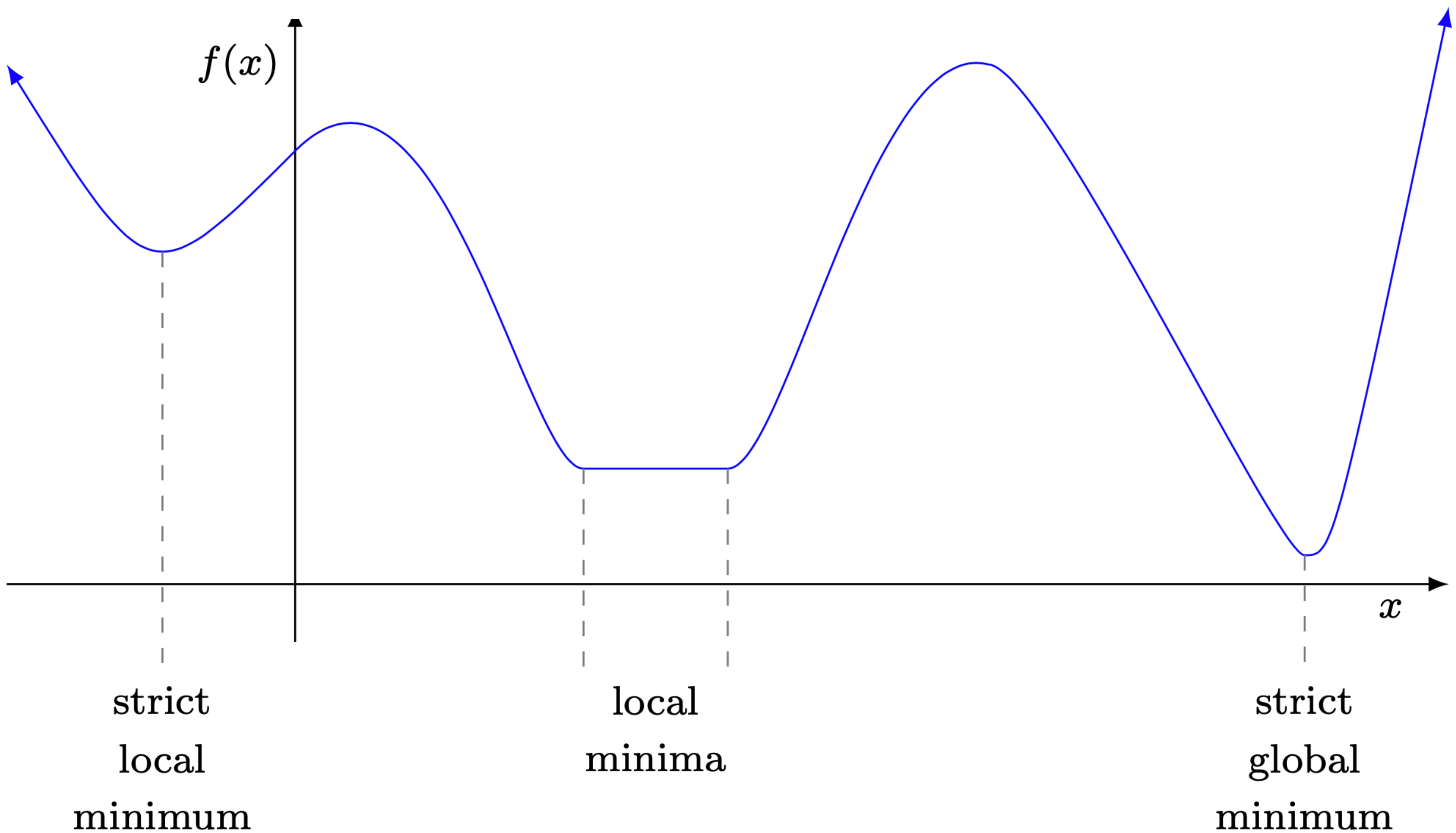
信息科学与技术学院

Email: wanghao1@shanghaitech.edu.cn

Unconstrained optimization

Consider the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} \; f(x).$$



Local $\Rightarrow$ global minimum in convex optimization

A special fact in convex optimization is that all local minima are global minima.

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then a local minimum of f is a global minimum of f.
If f is strictly convex, then there exists at most one global minimum of f.

Proof.

To derive a contradiction, suppose that x_* is a local minimum of f that is not a global minimum. Then, there exists $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) < f(x_*)$. By convexity of f, we have for all $\alpha \in (0, 1)$ that

$$f(\alpha x_* + (1-\alpha)\bar{x}) \leq \alpha f(x_*) + (1-\alpha) f(\bar{x}) < f(x_*).$$

This means that f has a value strictly lower than $f(x_*)$ at every point on the line segment $(x_*, \bar{x}]$, which violates the local minimality of x_*. (The statement about strictly convex f can be proved in a similar manner.)

Global vs. local minimization

- Unfortunately, for nonconvex optimization, the conditions in the definitions of global and local minima are not entirely useful.
- Unless we can verify strict quasiconvexity, we rarely have global information about f, and so have no way to verify if a point is a global minimizer.
- Thus, in nonconvex optimization, we often focus on finding a local minimizer.
- Using calculus, we can derive local optimality conditions that aid in determining if a point is a local minimizer.
- In this manner, we rarely (if ever) use the aforementioned definitions directly.

First-order necessary condition

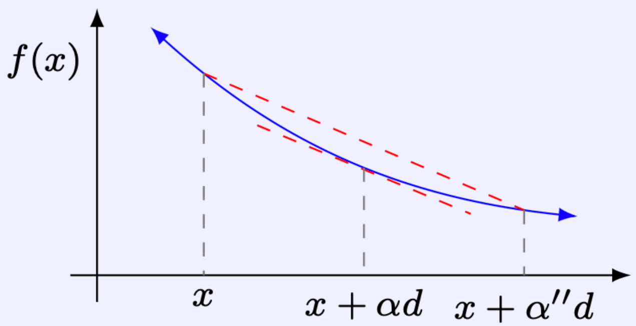
Theorem (First-order necessary condition)

If $f \in \mathcal{C}$ and x_* is a local minimizer of f, then $\nabla f(x_*) = 0$.

Proof.

For $x \in \mathbb{R}^n$ with $\nabla f(x) \neq 0$, let $d = -\nabla f(x)$ (with $\nabla f(x)^T d = -\|\nabla f(x)\|_2^2 < 0$). Since ∇f is continuous, there exists $\alpha' > 0$ such that $d^T \nabla f(x + \alpha d) < 0$ for all $\alpha \in [0, \alpha']$, i.e., the directional derivative remains negative some way along d. By the Mean Value Theorem (3.1.4), for any $\alpha'' \in (0, \alpha']$ we have

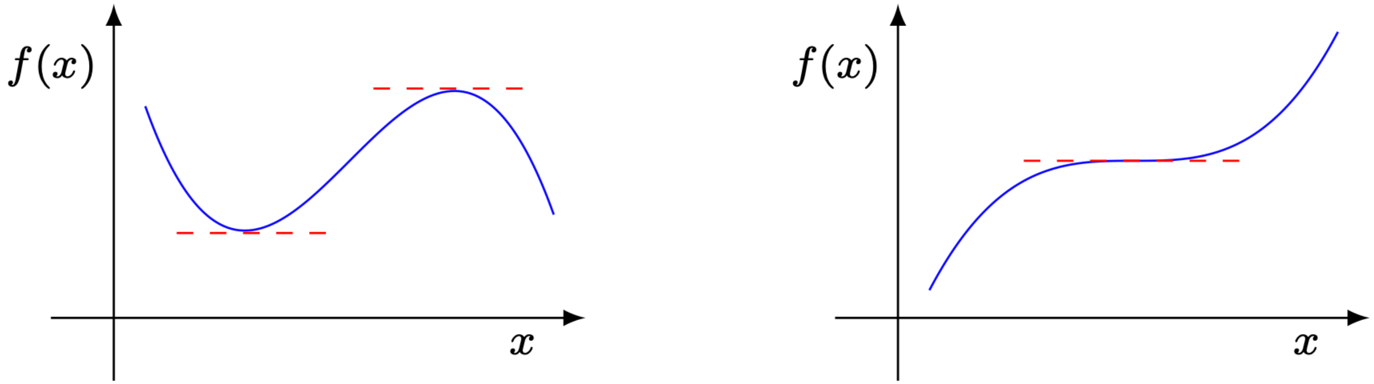
$$f(x + \alpha'' d) = f(x) + \alpha'' d^T \nabla f(x + \alpha d) \ \text{ for some } \ \alpha \in (0, \alpha'').$$



Thus, $f(x + \alpha'' d) < f(x)$ for all $\alpha'' \in (0, \alpha']$.

Stationary points

- We can limit our search to points where $\nabla f(x_*) = 0$.
- However, $\nabla f(x_*) = 0$ does not imply that we have a local minimizer!



- At least we know that if $\nabla f(x) \neq 0$, then x is not a local minimizer.

Definition (Stationary point)

A point $x \in \mathbb{R}^n$ is a stationary point for $f \in \mathcal{C}$ if $\nabla f(x) = 0$.

Convex optimization

If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex (but not necessarily real-valued or differentiable), then we have the following stronger result.

Theorem (First-order necessary and sufficient condition)

If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex and $0 \in \partial f(x_*)$, then x_* is a global minimizer of f.

In fact, we can say more to characterize the solution set of a convex problem. . .

Second-order necessary condition

Theorem (Second-order necessary condition)

If $f \in \mathcal{C}^2$ and x_* is a local minimizer of f, then $\nabla^2 f(x_*) \succeq 0$.

Proof.

For $x \in \mathbb{R}^n$ with $\nabla f(x) = 0$ but $\nabla f^2(x) \not\succeq 0$, let $d \in \mathbb{R}^n$ satisfy $d^T \nabla f(x) d < 0$. (We call such a d a direction of negative curvature.) Since $\nabla^2 f$ is continuous, there exists $\alpha' > 0$ such that

$$d^T \nabla^2 f(x + \alpha d) d < 0 \quad \text{for all} \quad \alpha \in [0, \alpha'],$$

i.e., the curvature remains negative some way along d. By Taylor's Theorem (3.1.5), for all $\alpha'' \in (0, \alpha']$ and some $\alpha \in (0, \alpha'')$ we have

$$\begin{aligned} f(x + \alpha'' d) &= f(x) + \alpha'' \nabla f(x)^T d + \tfrac{1}{2} \alpha''^2 d^T \nabla^2 f(x + \alpha d) d \\ &= f(x) + \tfrac{1}{2} \alpha''^2 d^T \nabla^2 f(x + \alpha d) d \\ &< f(x). \end{aligned}$$

Thus, x cannot be a minimizer.

Discussion

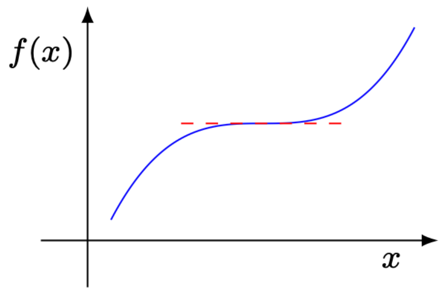
- Thus, at a local minimizer x_*, the Hessian of f is positive semidefinite.
- We already know that at a minimizer x_*, we have $\nabla f(x_*) = 0$, so together

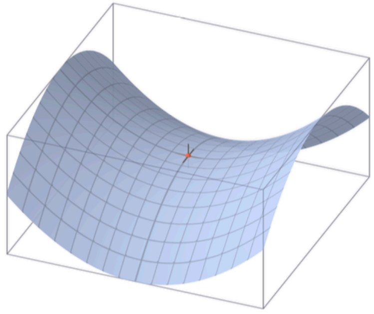
$$\nabla f(x_*) = 0 \quad \text{and} \quad \nabla^2 f(x_*) \succeq 0$$

  must be true at any local minimizer x_* of f.
- We can limit our search to points with zero gradient, then throw out any points where the Hessian is not positive semidefinite.

Necessary, but not sufficient

The fact that we may have $\nabla^2 f(x_*)d = 0$ for some d makes these insufficient.





- $f(x) = 1 + (x-4)^3$ has

$$\nabla f(x)|_{x=4} = 3(x-4)^2\big|_{x=4} = 0 \quad \text{and} \quad \nabla^2 f(x)\big|_{x=4} = 6(x-4)|_{x=4} = 0,$$

so the second order necessary conditions are satisfied at $x = 4$!

- $f(x) = x_1^4 - x_2^4$ has

$$\nabla f(x)|_{x=0} = \begin{bmatrix} 4x_1^3 \\ -4x_2^3 \end{bmatrix}\Bigg|_{x=0} = 0 \quad \text{and} \quad \nabla^2 f(x)\big|_{x=0} = \begin{bmatrix} 12x_1^2 & 0 \\ 0 & -12x_2^2 \end{bmatrix}\Bigg|_{x=0} = 0$$

so the second order necessary conditions are satisfied at $x = 0$.

- Note: the second order necessary conditions can be satisfied at a maximizer!

Second-order sufficient conditions

Theorem (Second-order sufficient conditions)

If $f \in \mathcal{C}^2$, $\nabla f(x_*) = 0$, and $\nabla^2 f(x_*) \succ 0$, then x_* is a strict local minimizer.

Proof sketch.

Since $\nabla^2 f$ is continuous, it remains positive definite near x_*. Taylor's Theorem (3.1.5) and $\nabla f(x_*) = 0$ then imply that, for some $\alpha \in (0, 1)$,

$$f(x_* + d) = f(x_*) + \tfrac{1}{2} d^T \nabla^2 f(x_* + \alpha d) d.$$

Hence, f must take larger values at other points near x_*. (See textbook.)

- A nice fact, when we can actually use it!
- By designing algorithms that find a sequence of points with decreasing function values, one hopes that maximizers and saddle points are avoided, i.e., one often focuses on finding a point with zero gradient. That being said, one can search over negative curvature directions to find a point satisfying the second-order necessary conditions, but, in general, a point satisfying the second-order sufficient conditions may not exist.

定义 5.2 (下降方向) 对于可微函数 f 和点 $x \in \mathbb{R}^n$，如果存在向量 d 满足

$$\nabla f(x)^{\mathrm{T}} d < 0,$$

那么称 d 为 f 在点 x 处的一个**下降方向**.

由下降方向的定义，容易验证: 如果 f 在点 x 处存在一个下降方向 d，那么对于任意的 $T > 0$，存在 $t \in (0, T]$，使得

$$f(x + td) < f(x).$$

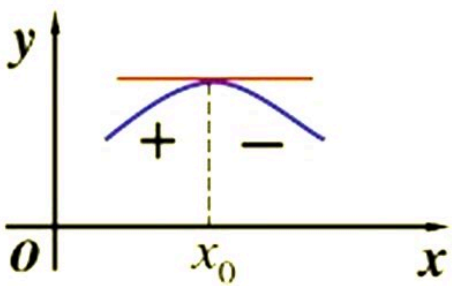
回忆：单变量函数的极值条件

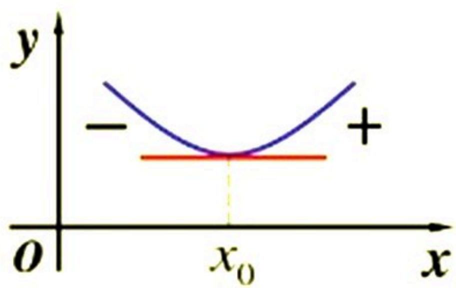
定理（第一充分条件）

(1) 如果$x \in (x_0-\delta, x_0)$，有$f'(x)>0$；而$x \in (x_0, x_0+\delta)$，有$f'(x)<0$，则$f(x)$在x_0处取得极大值.

(2) 如果$x \in (x_0-\delta, x_0)$，有$f'(x)<0$；而$x \in (x_0, x_0+\delta)$有$f'(x)>0$，则$f(x)$在x_0处取得极小值.

(3) 如果当$x \in (x_0-\delta, x_0)$及$x \in (x_0, x_0+\delta)$时，$f'(x)$符号相同，则$f(x)$在x_0处无极值.





单变量函数的极值条件

定理 （第二充分条件）设 $f(x)$ 在 x_0 处具有二阶导数，且 $f'(x_0)=0$，$f''(x_0)\neq 0$，**则**

(1) 当 $f''(x_0)<0$ 时，函数 $f(x)$ 在 x_0 处取得极大值；

(2) 当 $f''(x_0)>0$ 时，函数 $f(x)$ 在 x_0 处取得**极小值**。

第三充分条件

定理 假定 $f(x)$ 在 $x=x_0$ 处具有直到 n 阶的连续导数，且

$$f'(x_0)=f''(x_0)=\Lambda=f^{(n-1)}(x_0)=0,\text{但 } f^{(n)}(x_0)\neq 0$$

证明当 n 为偶数时，$f(x_0)$ 是 $f(x)$ 的极值

当 n 为奇数时，$f(x_0)$ 不是 $f(x)$ 的**极值**。

Optimization Problem and System of Equations

$$\min_x \quad f(x)$$

$$\Updownarrow$$

$$F(x) = 0$$