$\boxed{\text{Starting } 9/2}$

TA Office Hours (Will be added to syllabus with zoom links)

Salma	:	Mon	9-10am
		Fri	4-5 pm
Ning	:	Thurs	10:30-11:30 am
Thabo	:	Thurs	3-4pm
Jinke	:	Wed	3-4 pm

The gradient gives a <u>first-order model</u> of f

$$f(x+s) \approx f(x) + \nabla f(x)^T s$$

<u>Theorem (First-Order Taylor Approximation)</u>

Let f have L-Lipschitz continuous gradient.

(i.e. $\forall_{x,y \in \mathbb{R}^d}$, $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$)

Then $|f(x+s) - (f(x) + \nabla f(x)^T s)| \leq \frac{L}{2}\|s\|_2^2$.

Proof. HW1 Q2(a)

The Hessian gives a second-order model

$$f(x+s) \approx f(x) + \nabla f(x)^T s + \frac{1}{2} s^T \nabla^2 f(x) s$$

Theorem (Second-Order Taylor Approximation)

Let f have Q-Lipschitz continuous $\nabla^2 f(x)$

Then

$$\left| f(x+s) - \left(f(x) + \nabla f(x)^T s + \frac{1}{2} s^T \nabla^2 f(x) s \right) \right| \leq \frac{Q}{6} \|s\|_2^3 .$$

Proof. HW 1 Q2(b).

Optimality Conditions

1. Global, Local, Strict Minimizers
2. First-Order Necessary Condition
3. First-Order Sufficient Cond Under Convexity
4. Second-Order Necessary Condition
5. Second-Order Sufficient Condition

Basic Problem: Unconstrained Minimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

A vector $x^* \in \mathbb{R}^d$ is a <u>global minimizer</u> if

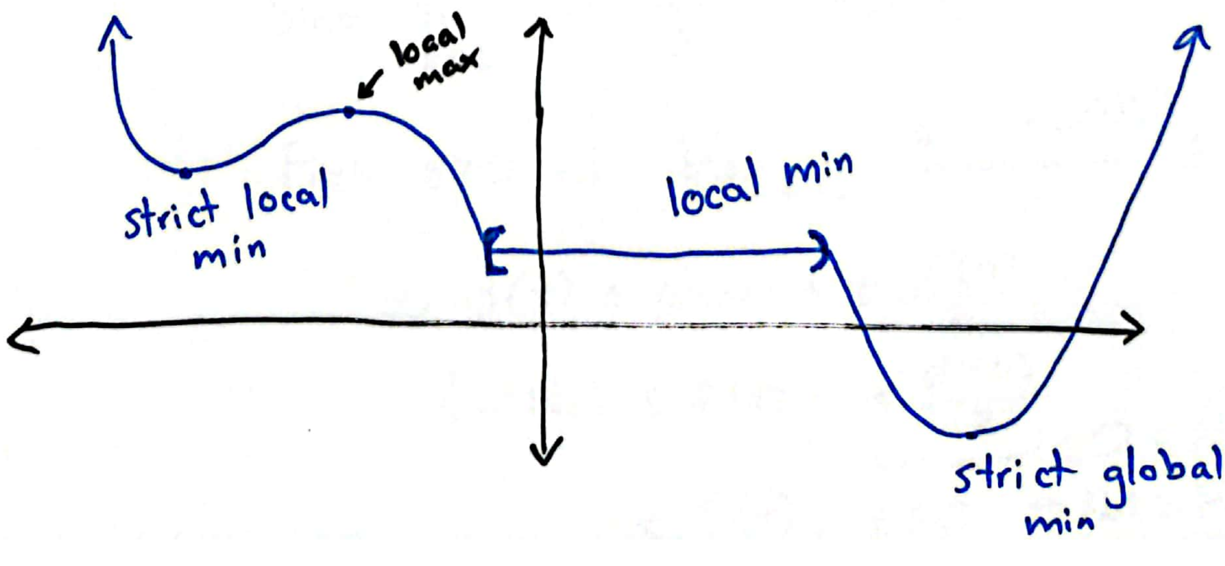
$$\forall x \in \mathbb{R}^d \quad f(x^*) \le f(x).$$

A vector x^* is a <u>local minimizer</u> if

$$\exists \varepsilon > 0 \; \forall x \in \underbrace{B(x^*, \varepsilon)}_{\{x \mid \|x - x^*\|_2 \le \varepsilon\}} \quad f(x^*) \le f(x)$$

A vector x^* is a <u>strict local minimizer</u> if

$$\exists \varepsilon > 0 \; \forall x \in B(x^*, \varepsilon) \setminus \{x^*\} \quad f(x^*) < f(x).$$



2. Theorem (First-Order Necessary Condition)

Suppose f is cont. diff.

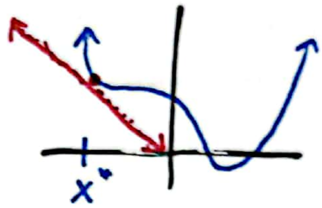
If x^* is a local min, then

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

Proof Idea. $\nabla f(x^*) \neq 0$

$\Rightarrow$ Some direction downhill

$\Rightarrow$ Better nearby point.



Proof. Consider any $x^* \in \mathbb{R}^d$ with $\nabla f(x^*) \neq 0$

Let $s = -\dfrac{\nabla f(x^*)}{\|\nabla f(x^*)\|}$ and define

$$\phi(t) = f(x + ts)$$

We know $\phi'(0) = \nabla f(x^*)^T s = -\|\nabla f(x)\|_2$
$< 0.$

Since $\lim_{t \to 0} \dfrac{\phi(t) - \phi(0)}{t} = \phi'(0).$

Pick $\varepsilon > 0$ s.t. $\forall t \leq \varepsilon$, $\dfrac{\phi(t) - \phi(0)}{t} \leq \dfrac{\phi'(0)}{2}$

$$\Rightarrow \phi(t) \leq \phi(0) + t\,\frac{\phi'(0)}{2}$$

$$f(x^* + ts) \leq f(x^*) + t\,\underbrace{\frac{\phi'(0)}{2}}_{<0}$$

$< f(x^*)$, $t \neq 0$. $\Rightarrow$ Not a local min $\square$

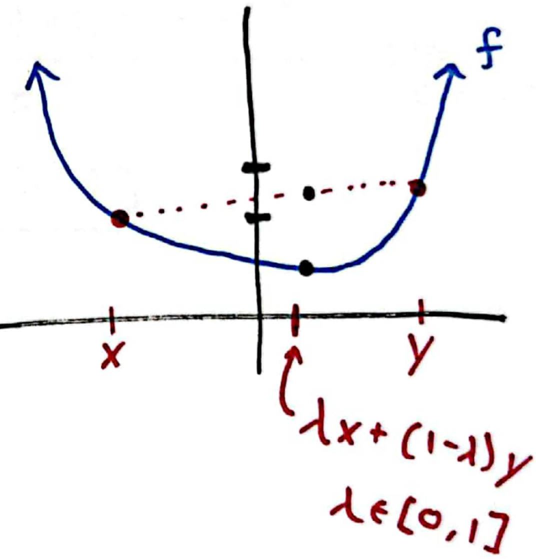
Is this sufficient?

(Does every $\nabla f(x^*)=0$ have x^* as a local min)

No! Local max also $\nabla f(x)=0$
(saddle point)

3. First-Order Sufficient Condition Under Convexity

> We say f is <u>convex</u> if
>
> $\forall x,y \in \mathbb{R}^d,\ \lambda \in [0,1]$
>
> $f(\lambda x+(1-\lambda)y) \le \lambda f(x)+(1-\lambda)f(y)$.



Theorem (First-Order Conditions Under Convexity)

Suppose f is cont. diff. and convex.

Then x^* has $\nabla f(x^*)=0$ if and only if x^* is a global min.

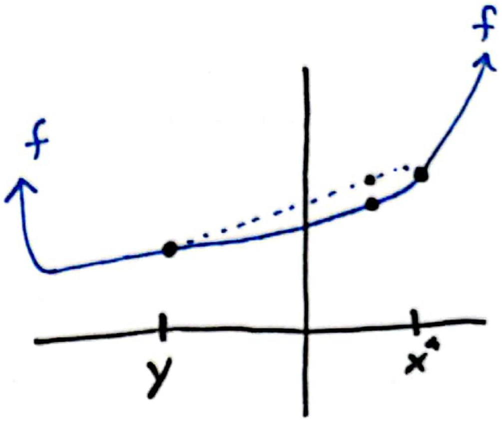
Proof. $\Leftarrow$ (as global min $\Rightarrow$ local min $\Rightarrow \nabla f(x^*)=0$)

$\boxed{\Rightarrow}$ Lets show the contrapositive.

Suppose x^* is not a global min

Let y s.t. $f(y) < f(x^*)$

and $\phi(t) = f(x^* + t(y - x^*))$



Then $\phi'(0) = \nabla f(x^*)^T (y - x^*)$

$$= \lim_{t \to 0} \frac{f(x^* + t(y - x^*)) - f(x^*)}{t}$$

$$\leq \lim_{t \to 0} \frac{(1-t) f(x^*) + t f(y) - f(x^*)}{t} \quad \text{(by convexity)}$$

$$= \lim_{t \to 0} \frac{t(f(y) - f(x^*))}{t}$$

$$= f(y) - f(x^*)$$

$$< 0.$$

$\Rightarrow \nabla f(x^*) \neq 0.$ $\square$

4. Theorem (Second-Order Necessary Condition)

Suppose f is twice diff.

If x^* is local min, then

$\nabla f(x^*) = 0$ and $s^T \nabla^2 f(x^*) s \geq 0 \quad \forall s \in \mathbb{R}^d$

$\Updownarrow$

$\nabla^2 f(x)$ is positive semidefinite

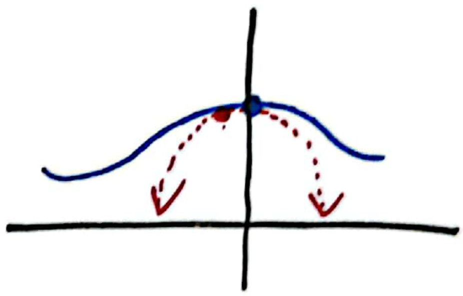
$\Leftrightarrow \nabla^2 f(x) \succeq 0$

Proof Idea. If $\nabla f(x^*) = 0$ but $s^T \nabla^2 f(x) s < 0$ then

negative curvature

$\Rightarrow$ Nearby better point.

$\Rightarrow$ Not local min.



Proof. Suppose $\nabla f(x^*) = 0$ but $s^T \nabla^2 f(x^*) s < 0$

for some $s \in \mathbb{R}^d$

(for example, s as the most negative eigenvector of $\nabla^2 f(x)$)

Then $\phi(t) = f(x^* + ts)$, has $\phi'(0) = 0$

$\phi''(0) = s^T \nabla^2 f(x^*) s$

< 0.

$$\frac{1}{2}\phi''(0) = \lim_{t\to 0} \frac{\phi(t) - \phi(0)}{t^2} < 0$$

Pick t small enough $\frac{\phi(t) - \phi(0)}{t^2} \leq \frac{\phi''(0)}{2 \cdot 2}$.

$$\Rightarrow \phi(t) \leq \phi(0) + \underbrace{t^2 \frac{\phi''(0)}{2 \cdot 2}}_{<0}$$

$$\Rightarrow f(x^* + ts) < f(x^*)$$

$\Rightarrow$ Not local min. $\square$