

Lecture 2

Scribe?

Agenda

Aug 31/23

- Calculus review
- Optimality Conditions
- Basics of convexity ← Most likely next time!

Calculus Review

Consider a smooth function

$f: \mathbb{R}^d \rightarrow \mathbb{R}$, i.e., there exists
 $\nabla f(\bar{x})$ s.t. $\forall x$

$$\lim_{h \rightarrow 0} \frac{f(\bar{x} + h) - (f(\bar{x}) + \nabla f(\bar{x})^T h)}{\|h\|} = 0.$$

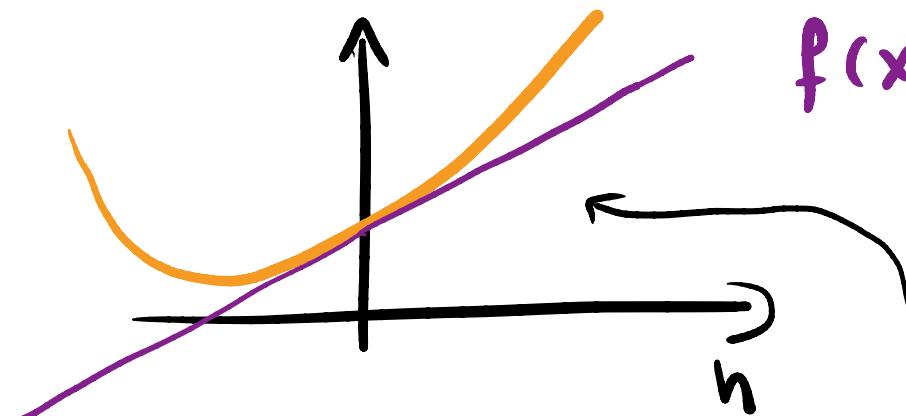
Equivently,

$$\nabla f(\bar{x}) = \begin{pmatrix} \frac{\partial f(\bar{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\bar{x})}{\partial x_n} \end{pmatrix}.$$

Intuition

$$f(x+h)$$

$$f(x) + \nabla f(x)^T h$$



Zoom

In the limit the slopes are equal.

We say that a function is

twice differentiable at x if

\exists a linear mapping $\nabla^2 f(x) : \mathbb{R}^d \rightarrow \mathbb{R}^d$

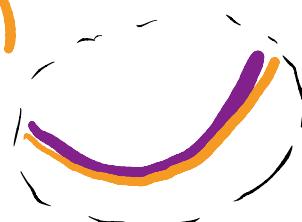
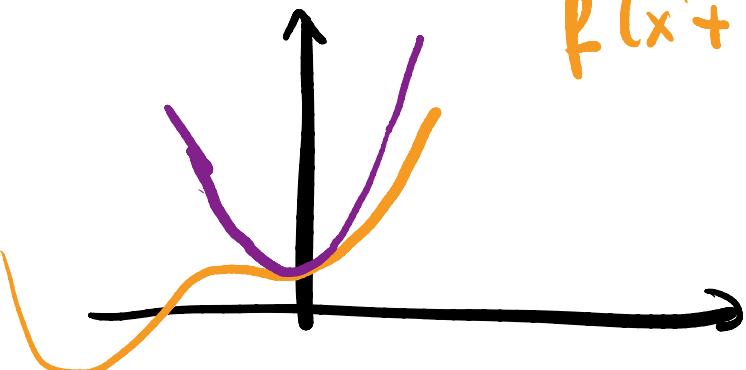
s.t.

$$\lim_{h \rightarrow 0} \frac{f(\bar{x} + h) - (f(\bar{x}) + \nabla f(\bar{x})^T h + \frac{1}{2} h^T \nabla^2 f(\bar{x}) h)}{\|h\|^2} = 0.$$

Intuition

$$f(x) + \nabla f(x)^T h$$

$$f(x+h)$$



Same second order behavior in the limit.

Theorem: Let $f: \mathbb{R}^d \rightarrow \mathbb{R}$. Fix $\bar{x}, \bar{s} \in \mathbb{R}^d$

Define $\varphi(t) = f(\bar{x} + t\bar{s})$.

- If f is differentiable

so is φ and

$$\varphi'(t) = \bar{s}^T \nabla f(\bar{x} + t\bar{s})$$

Chain rule

- If f is twice differentiable

so is φ and

$$\varphi''(t) = \bar{s}^T \nabla^2 f(\bar{x} + t\bar{s}) \bar{s}.$$

-1

Why do we care?

This gives a nice way to compute derivatives!

Theorem (First-order Taylor approximation)

Let f have an L -Lipschitz continuous gradient ($\forall x, y \quad \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$).

Then

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

Proof Exercise in HW L.

+

□

Theorem (Second order approximation)

Let f have a α -Lipschitz continuous Hessian $\nabla^2 f^2(\bar{x})$.

Then

$$|f(\bar{x} + h) - (f(\bar{x}) + \nabla f(\bar{x})^T h + \frac{1}{2} h^T \nabla^2 f(\bar{x}) h)| \leq \frac{\alpha}{6} \|h\|^3$$

Proof Exercise in HW L.

+

□

Why do we care?

If you end up doing research, approximations will simplify things for you!

Both in calculations (physicists are famous for doing this) and for algorithms

Optimality conditions

Types of minimizers

Assume $f: \mathbb{R}^d \rightarrow \mathbb{R}$ is a function.

Consider

$$\min_{\bar{x}} f(\bar{x})$$

► Global optimizers (The holy grail)

A point \bar{x}^* is a global minimizer if $\forall \bar{x} \in \mathbb{R}^d$

$$f(\bar{x}^*) \leq f(\bar{x})$$

► Local minimizer

A point \bar{x}^* is a local minimizer if

$$\exists \varepsilon \text{ s.t. } \forall \bar{x} \in B_\varepsilon(\bar{x}^*) \quad \{x \mid \|x - \bar{x}^*\|_2 \leq \varepsilon\}$$

$$f(\bar{x}^*) \leq f(\bar{x})$$

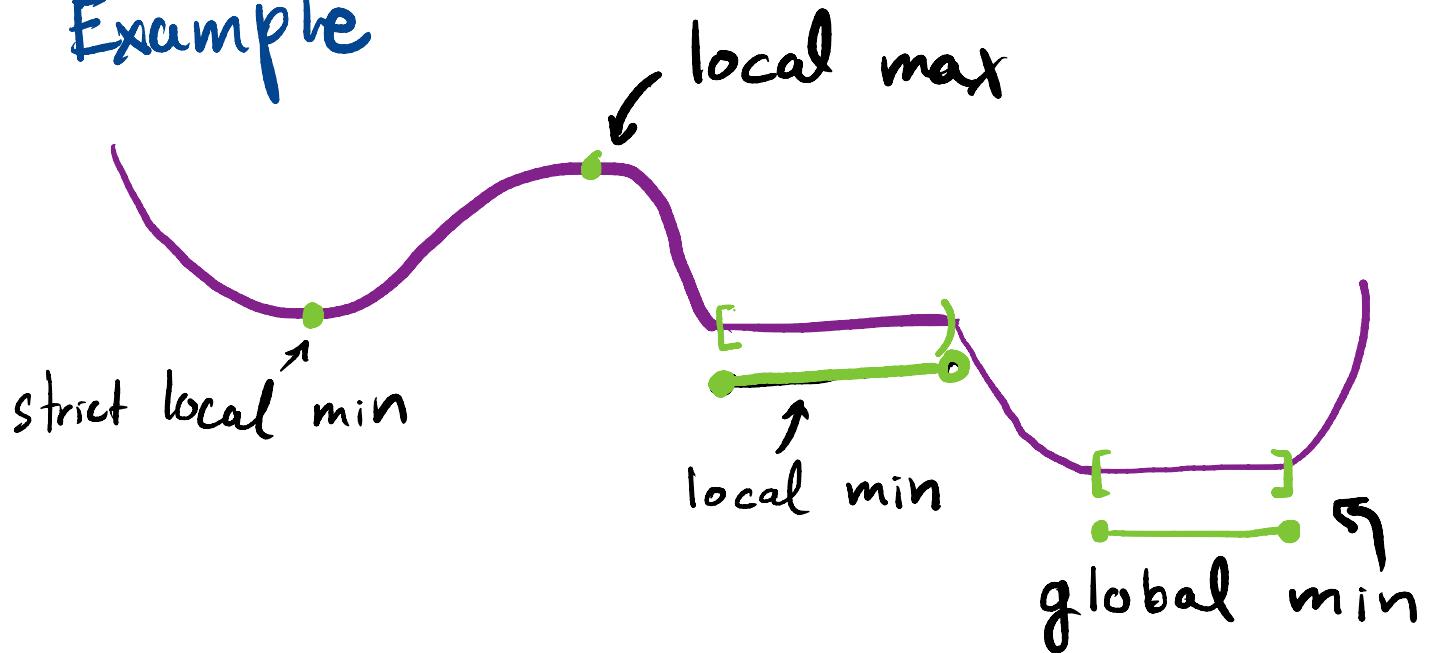
A point \bar{x}^* is a strict (local) minimizer

$$\forall \bar{x} \in \mathbb{R}^d \quad (\exists \varepsilon \quad \forall \bar{x} \in B_\varepsilon(\bar{x}^*))$$

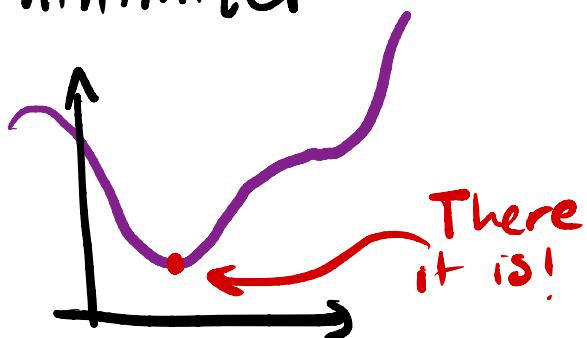
$$f(\bar{x}^*) < f(\bar{x})$$

Notice the
STRICT equation

Example



When we draw a function in 2d, it's easy to find a minimizer



But when the domain is high-dim, e.g., $d = 200$, it's impossible

so we need to be able to find conditions that ensure (or hint) that a point is an optimum.

That's what optimality conditions are all about!

We will cover 4 types of optimality conditions

- First-order necessary condition
- First-order sufficient condition
- Second-order necessary condition
- Second-order sufficient condition

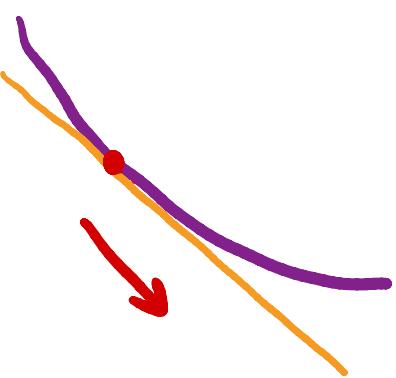
Theorem First-order necessary condition

Suppose f is cont diff (C^1)

If \bar{x}^* is a local min,

$$\Rightarrow \nabla f(\bar{x}^*) = 0.$$

Intuition



If $\nabla f(x^*) \neq 0$
 \Rightarrow Direction downhill
 \Rightarrow Better nearby point.

Proof: Assume, looking for a contradiction, $\nabla f(\bar{x}^*) \neq 0$.

Define

$$\phi(t) = f(\bar{x}^* + t \bar{s})$$

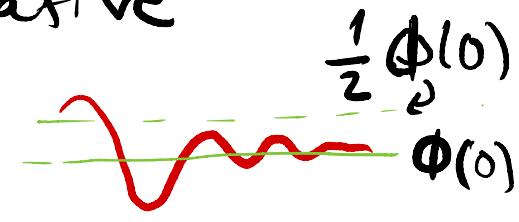
$$\bar{s} = -\frac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|}$$

Then, by the chain rule

$$\phi'(0) = s^\top \nabla f(\bar{x}^*) = -\|\nabla f(\bar{x}^*)\| < 0.$$

By the def. of the derivative

$$\phi'(0) = \lim_{t \rightarrow 0} \frac{\phi(t) - \phi(0)}{t}.$$



Then, all for small enough $t > 0$

$$\begin{aligned}\phi(t) - \phi(0) &\leq \frac{1}{2} \phi'(0) t \\ &\leq -\frac{1}{2} \|\nabla f(\bar{x})\| t < 0\end{aligned}$$

$$\Rightarrow f(\bar{x}^* + t \bar{s}) < f(x^*)$$

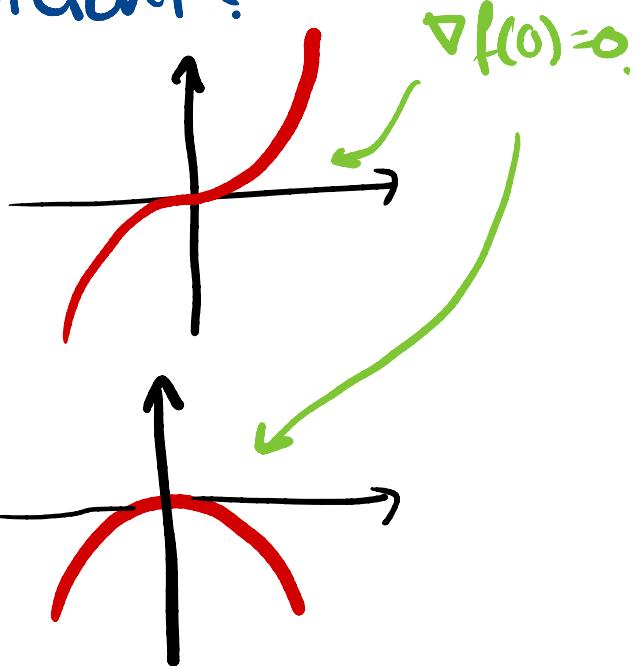
↓

□

Critical

Is $\nabla f(x^*) = 0$ sufficient?

No! $f(x) = x^3$



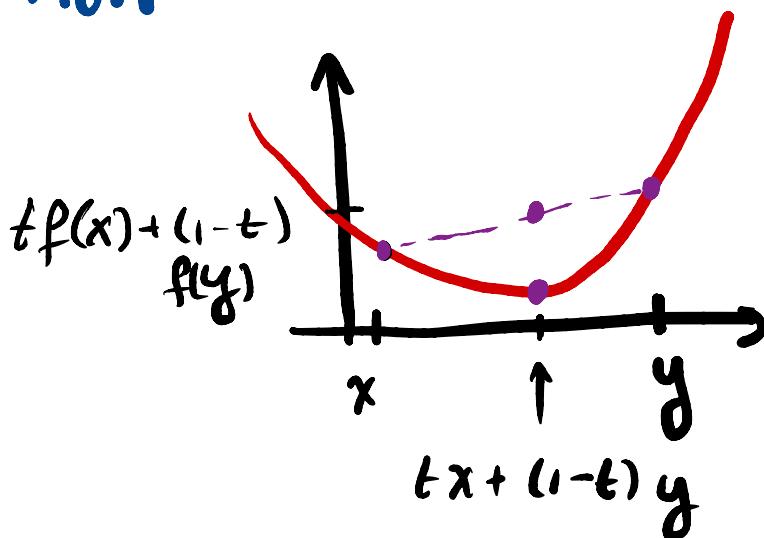
$$f(x) = -x^2$$

In order to define a sufficient condition we need to define a special family of functions

Def: A function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ is convex if $\forall \bar{x}, \bar{y} \in \mathbb{R}^d$ and $\forall t \in [0,1]$

$$f(t\bar{x} + (1-t)\bar{y}) \leq tf(\bar{x}) + (1-t)f(\bar{y})$$

Intuition



Theorem (1st-order sufficient condition)

Assume that $f: \mathbb{R}^d \rightarrow \mathbb{R}$ is a smooth convex function.

Then, x^* satisfies $\nabla f(x^*) = 0$
iff x^* is a global minimizer.