Lecture 2 Aug/29/24

Agenda

- Calculus review
- Optimality conditions

Calculus Review

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

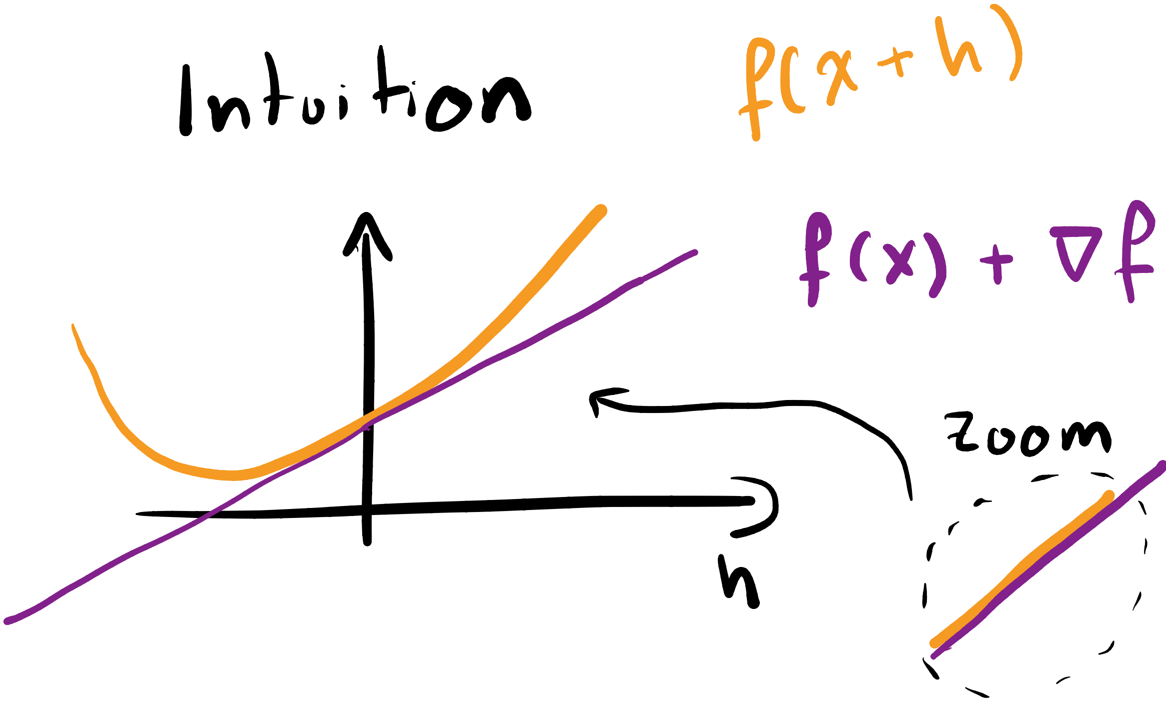
$$\lim_{h \to 0} \frac{f(\bar{x}+\bar{h}) - (f(\bar{x}) + \nabla f(\bar{x})^T \bar{h})}{\|h\|} = 0.$$

If f is differentiable at x, then

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

This doesn't necessarily hold even if the partial derivatives exist.

Intuition



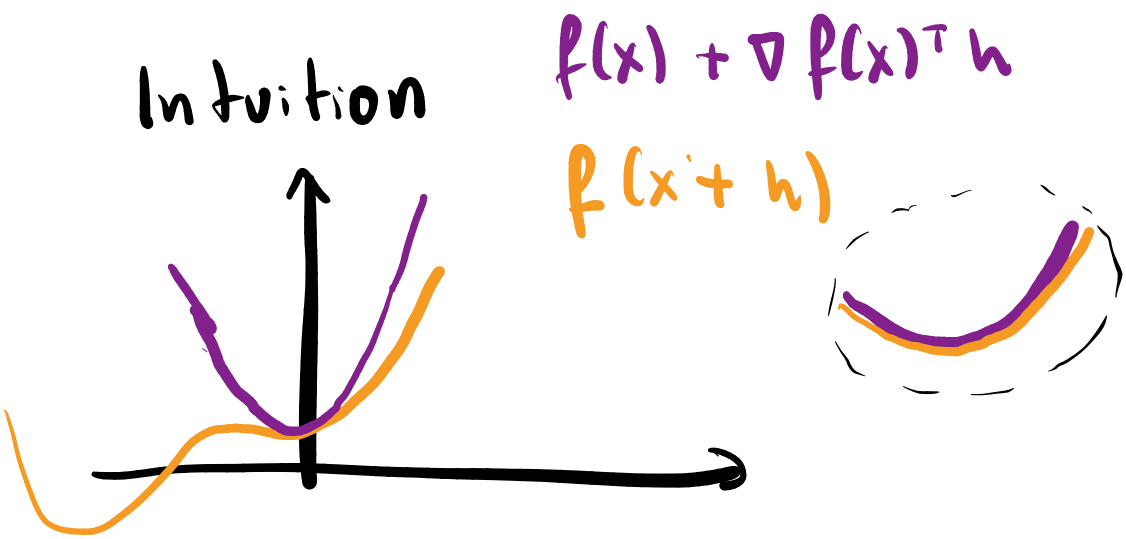
In the limit the slopes are equal.

We say that a function is twice differentiable at x if $\exists$ a linear mapping $\nabla^2 f(\bar{x}) : \mathbb{R}^d \to \mathbb{R}^d$ s.t.

$$\lim_{h \to 0} \frac{f(\bar{x} + \bar{h}) - \left(f(\bar{x}) + \nabla f(\bar{x})^T \bar{h} + \frac{1}{2} \bar{h} \nabla^2 f(\bar{x}) \bar{h}\right)}{\|\bar{h}\|^2}$$

$$= 0.$$

Intuition



Same second order behavior in the limit.

Theorem: Let $f:\mathbb{R}^d \to \mathbb{R}$. 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

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

$\dashv$

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 $(\forall x, y \ \|\nabla f(\bar{x}) - \nabla f(\bar{y})\| \leq L\|\bar{x} - \bar{y}\|)$.

Then

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

Proof Exercise in HW 1. $\dashv$

$\square$

Theorem (Second order approximation)

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

w.r.t the operator norm.

Then

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

Proof

Exercise in HW 1. $\dashv$

$\square$

Why do we care?

If you end up doing research, approximations will simplify things for you! Both in calculations (physists are famous for doing this) and for algorithms.

Many optimazation algorithms iteratively minimize approximations!

More about this later.

Optimality conditions

Types of minimizers

Assume $f: \mathbb{R}^d \to \mathbb{R}$ is a function.
Consider
$$\min_{\bar{x}} f(\bar{x})$$

- **Global optimizers** (The holy grail)

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

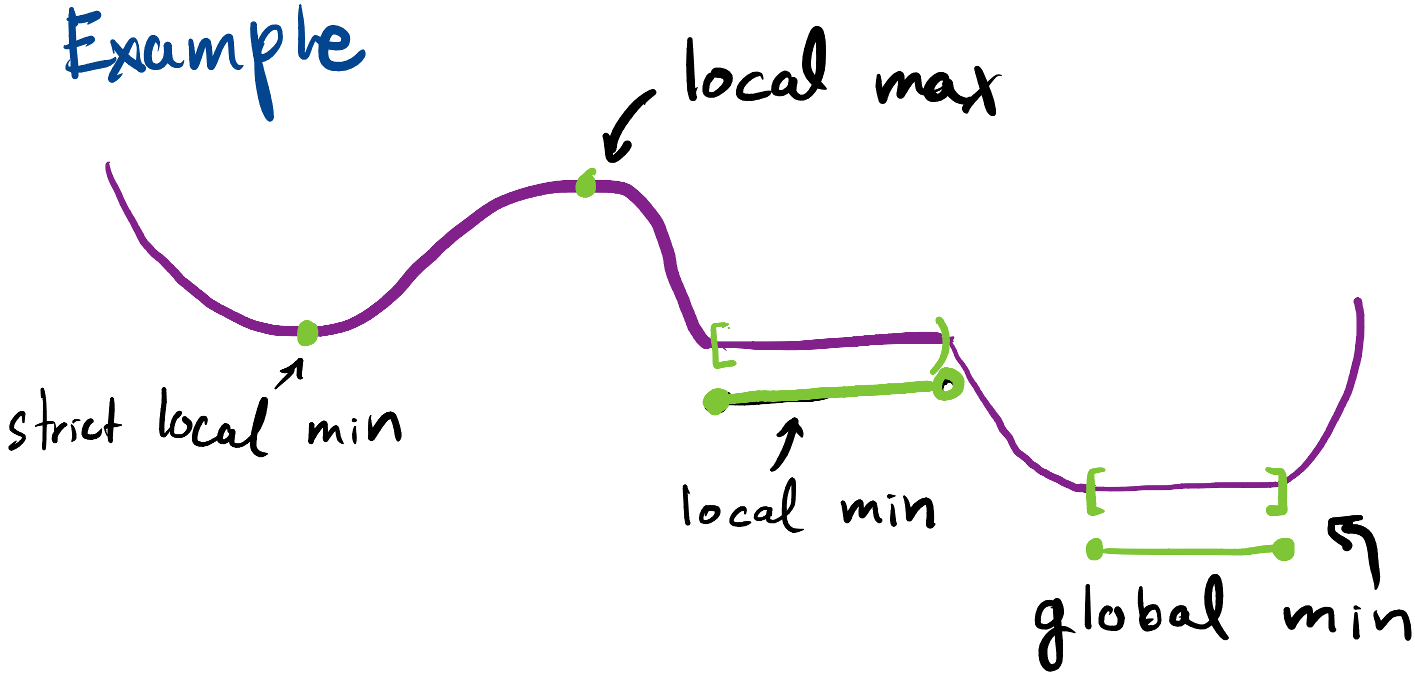
- **Local minimizer**

A point x^* is a local minimizer if $\exists \varepsilon$ s.t $\forall \bar{x} \in B_\varepsilon(x^*)$ $\{x \mid \|x - \bar{x}\|_2 \leq \varepsilon\}$
$$f(\bar{x}^*) \leq f(\bar{x})$$

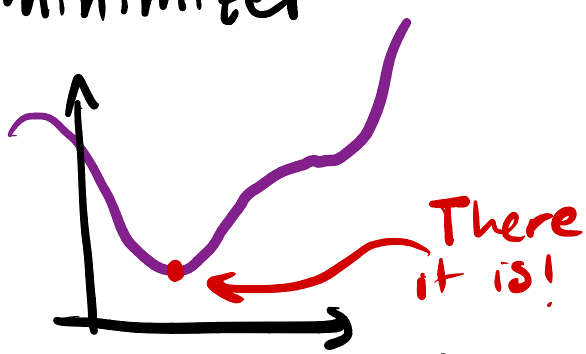
A point x^* is a strict (local) minimizer
$\forall \bar{x} \in \mathbb{R}^d$ $(\exists \varepsilon \ \forall \bar{x} \in B_\varepsilon(\bar{x}^*))$
$$f(\bar{x}^*) < f(\bar{x})$$

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 for convex functions
- Second-order nessary condition
- Second-order sufficient condition

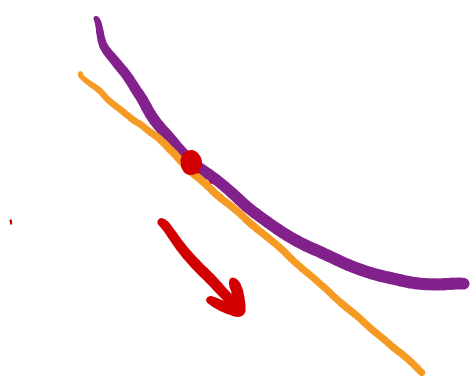
Theorem First-order nessary condition

Suppose f is cont diff (C^1)

If x^* is a local min,

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

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

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

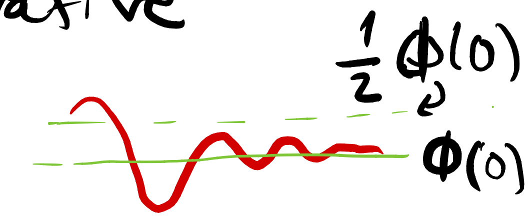
$$\phi(t) = f(\bar{x}^* + t\,\bar{s})$$

Then, by the chain rule

$$\phi'(0) = s^T \nabla f(\bar{x}^*) = -\|\nabla f(\bar{x}^*)\| < 0$$

By the def. of the derivative

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



Then, all for small enough $t > 0$

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

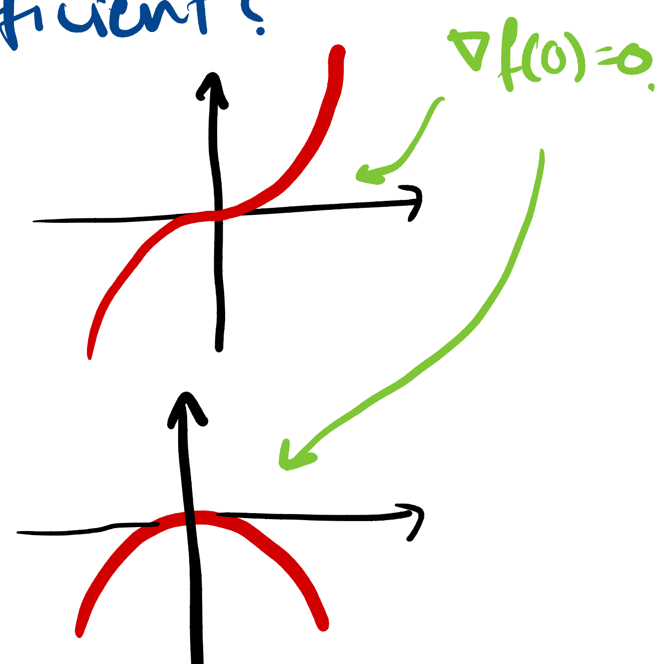
$$\Rightarrow \quad f(\bar{x}^* + t\bar{s}) < f(x^*) \qquad \lightning \qquad \square$$

Critical

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

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

$f(x) = -x^2$

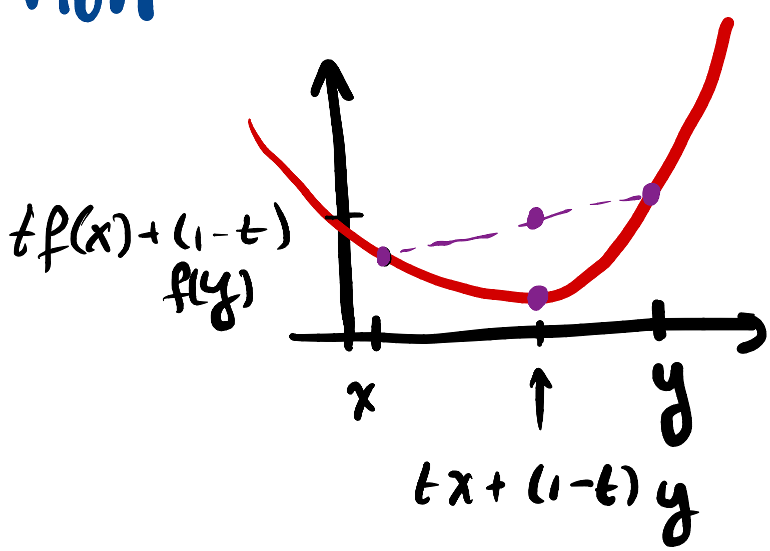


In order to define a sufficient condition we need to define a special family of functions

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

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

Intuition



Theorem (1st-order sufficient condition)

Assume that $f: \mathbb{R}^d \to \mathbb{R}$ is a smooth convex function.
Then, x^* satisfies $\nabla f(x^*) = 0$ iff x^* is a global minimizer.