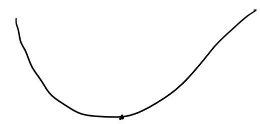
* Minima of convex functions:

(P) $\min_{x \in \mathcal{X}} f(x)$: f is convex, $\mathcal{X}$ is convex, closed, and nonempty

* <u>Thm 2.6</u> Consider (P). We have the following:

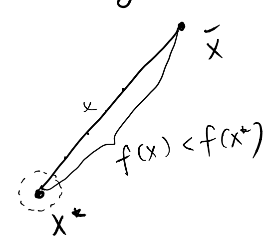
(a) Any local solution to (P) is also a global solution.



(b) The set of global solutions to (P) is convex.

<u>Proof:</u>

(a) Suppose f.p.o.c. that x^* is a local but not a global solution. Then $\exists\, \bar{x} \in \mathcal{X}$ s.t. $f(\bar{x}) < f(x^*)$.



As $\mathcal{X}$ is convex, $\forall\, \alpha \in (0,1)$:

$(1-\alpha)\, x^* + \alpha \bar{x} \in \mathcal{X}$.

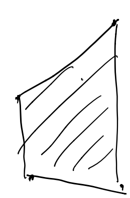
As f is convex, $\forall\, \alpha \in (0,1)$:

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

$\Rightarrow$ Every neighborhood of x^* must include a point $(1-\alpha) x^* + \alpha \bar{x}$ for some $\alpha > 0$ that will have a <u>strictly</u> lower function value. $\Rightarrow$ x^* cannot be a local solution.

(b) Let $x^*, \bar{x} \in \mathcal{X}$ be any two global solutions.

$\mathcal{X}$ is convex $\Rightarrow \forall\, \alpha \in (0,1)$: $(1-\alpha)\, x^* + \alpha \bar{x} \in \mathcal{X}$.



f is convex $\Rightarrow \forall\, \alpha \in (0,1)$:

$$f((1-\alpha) x^* + \alpha \bar{x}) \leq (1-\alpha) f(x^*) + \alpha f(\bar{x})$$
$$= f(x^*) = f(\bar{x})$$

$\Rightarrow f((1-\alpha) x^* + \alpha \bar{x}) = f(x^*)$

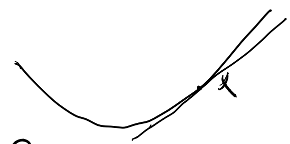
$\Rightarrow (1-\alpha) x^* + \alpha \bar{x}$ is a global solution.

$\Rightarrow$ the set of global solutions must be convex. $\blacksquare$

* Thm.

(a) Let f be cont.'ly diff.'able. f is convex if and only if

$$\forall x, y: \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle .$$

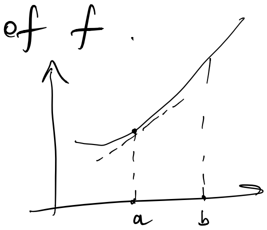


(b) Let f be twice cont.'ly diff.'able. f is convex if and only if $\forall x: \nabla^2 f(x) \succcurlyeq 0$.

* Thm 2.7 Let f be cont.'ly diff.'able and convex. If $\nabla f(x^*) = 0$, then x^* is a global min of f.

Pf: Use Part (a) of the Thm above:

$$\forall x: \underbrace{f(x) \geq f(x^*)} + \langle \cancelto{0}{\nabla f(x^*)}, x - x^* \rangle .$$



(for constrained setups, we would use $\langle \nabla f(x^*), x - x^* \rangle \geq 0, \forall x$)

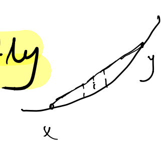
* Strongly convex functions:

* Def. Given $m > 0$, we say that $f: \mathbb{R}^d \to \bar{\mathbb{R}}$ is m-strongly convex (or strongly convex w/ modulus m), if $\forall x, y \in \mathbb{R}^d$:

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha) f(x) + \alpha f(y) - \frac{m}{2}(1-\alpha)\alpha \|y - x\|^2 .$$

* Ex: 1) When f is cont.'ly diff.'able, equivalently

$$\forall x, y: f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{m}{2}\|y - x\|^2 .$$



2) When f is twice cont.'ly diff.'able, equivalently:

$$\forall x: \nabla^2 f(x) \succcurlyeq m I .$$

* Thm 2.8. Suppose that $f: \mathbb{R}^d \to \bar{\mathbb{R}}$ is cont.'ly diff.'able and m-strongly convex for some $m > 0$. If $\nabla f(x^*) = 0$, then x^* is the unique global min of f.

Proof: From Ex 1):

$$\forall x: f(x) \geq f(x^*) + \langle \cancelto{0}{\nabla f(x^*)}, x - x^* \rangle + \underbrace{\frac{m}{2}\|x - x^*\|^2}_{>0 \text{ unless } x = x^*}$$

∎

* <u>Growth of sequences</u>:

$\{a_k\}_{k\geq 1}$, $\{b_k\}_{k\geq 1}$, $\forall k : a_k, b_k \geq 0$.

$a_k \leq 10 b_k$
$a_k = \mathcal{O}(b_k)$

* "Big-Oh" notation:

$a_k = \mathcal{O}(b_k) \iff (\exists M > 0)(\exists K < \infty)(\forall k \geq K) : a_k \leq M b_k$.

(E.g., $k = \mathcal{O}(\frac{1}{10} k^2)$, $k = \mathcal{O}(\frac{1}{10!} k)$)

* If $a_k = \mathcal{O}(b_k)$ and $b_k = \mathcal{O}(a_k)$, we write $a_k = \Theta(b_k)$.

* "Little-Oh" notation:

$a_k = o(b_k) \iff \lim_{k\to\infty} \frac{a_k}{b_k} = 0$.

---

* <u>Algorithmic setup</u>:

1) first-order oracle model:

$$x \longrightarrow \boxed{\text{oracle}} \longrightarrow \underline{f(x)}, \underline{\nabla f(x)}$$

2) second-order oracle model:

$$x \longrightarrow \boxed{\text{oracle}} \longrightarrow f(x), \nabla f(x), \nabla^2 f(x)$$

* All algorithms we consider in this class are iterative:
  → start w/ some x_0, get oracle answers for x_0, choose x_1
  → at iteration k, get oracle answers for x_k, choose x_{k+1}

* <u>Basic Descent Methods</u>:

* Assumptions for this part:

(A1) f is L-smooth for some $L < \infty$ (thus also cont.'ly diff.'able)

(A2) $\mathcal{X} = \mathbb{R}^d$, i.e., the problem is unconstrained

<u>Note</u>: for now, and until explicitly stated otherwise, we are <u>not</u> assuming that f is convex.

* Def. $p \in \mathbb{R}^d$ is a descent direction for f at x if $f(x+tp) < f(x)$ for all suff. small $t > 0$.

* Prop 3.2. If f is cont.'ly diff.'able (in a neighborhood of x), then any p s.t. $\langle \nabla f(x), p \rangle < 0$ is a descent direction.

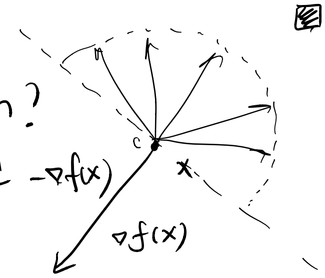
Proof: TT + continuity of ∇f: $(y = x + tp)$

$$f(x+tp) = f(x) + t \langle \nabla f(x + \mu tp), p \rangle \text{ for some } \mu \in (0,1).$$

We know that $\langle \nabla f(x), p \rangle < 0$. As ∇f is continuous, for all suff. small $t > 0$:

$$t \langle \nabla f(x + \mu tp), p \rangle < 0.$$

$$\Rightarrow f(x+tp) < f(x).$$



* What would be a good descent direction?
  - could try to move in the direction of $-\nabla f(x)$
  - justification:

    Look at all p w/ $\|p\|_2 = 1$. Then:

$$\inf_{\|p\|_2 = 1} \langle \nabla f(x), p \rangle = -\|\nabla f(x)\|_2 \text{ attained for } p = -\frac{\nabla f(x)}{\|\nabla f(x)\|_2}.$$