Amath 515 Jan 13

* More concepts: closed, proper, level bounded

* strong convexity

* Measuring rates.

Last time: Characterized behavior

of continuously differentiable (C^1)

convex fcns.

Cor 2.19 For $f \in C^1$ and convex, TFAE:

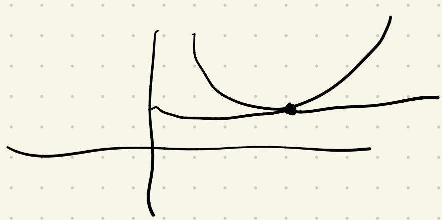
1. x is a global minimizer for f
2. x is a local minimizer for f
3. x is a stationary pt ($\nabla f(x) = 0$).

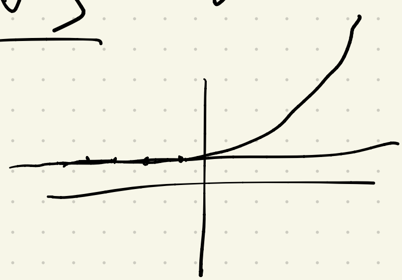
Pf: $1 \Rightarrow 2$ trivial, $2 \Rightarrow 3$ we did earlier

$3 \Rightarrow 1$: $f(y) \geq f(x) + \langle \cancel{\nabla f(x)}, y - x \rangle$ $\forall x, y$

At x, $\nabla f(x) = 0$

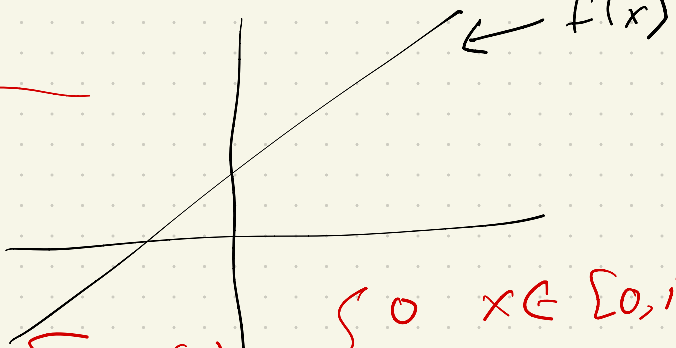
$f(y) \geq f(x)$ $\forall y$ ✓





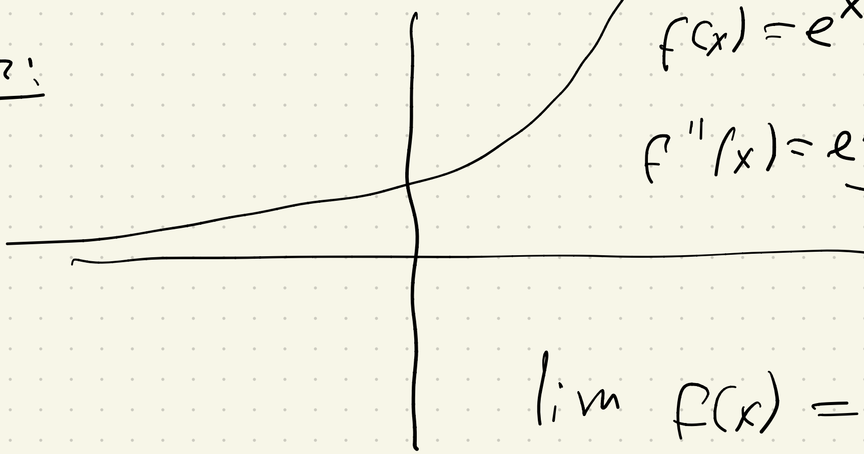
Even for convex fcns, a minimizer need not exist.

Ex 1:



$$\delta_{[0,1]}(x) = \begin{cases} 0 & x \in [0,1] \\ \infty & x \notin [0,1] \end{cases}$$

Ex 2:



$f(x) = e^x$

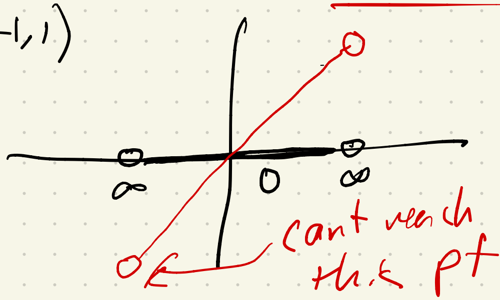
$f''(x) = e^x > 0$

$$\lim_{x \to \infty} f(x) = 0$$

Ex 3: $f(x) = \delta_{(-1,1)}(x)$

To exclude

$g(x) = x + \delta_{(-1,1)}$



Read p. 5–11 of 'variational analysis'

Def: Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$

Def: For a $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, we call

$$\underline{\text{dom } f} = \{x \in \mathbb{R}^n : f(x) < \infty\}.$$

↪ 'effective domain'

$f(x) = \delta_{[0,1]}(x)$, then $\text{dom } f = [0,1]$

$$\inf f = \inf_{x \in \mathbb{R}^n} f(x) = \inf_{x \in \text{dom} f} f(x)$$

= smallest value f can take.

$$\text{argmin } f = \{x : f(x) = \inf f\}$$

could be empty.

Ex: $f(x) = e^x$, $\text{dom } f = \mathbb{R}$

$\inf f = 0$

$\text{argmin } f = \{\}$

* argmin = set of global minimizers.

* Lets exclude some fcns.

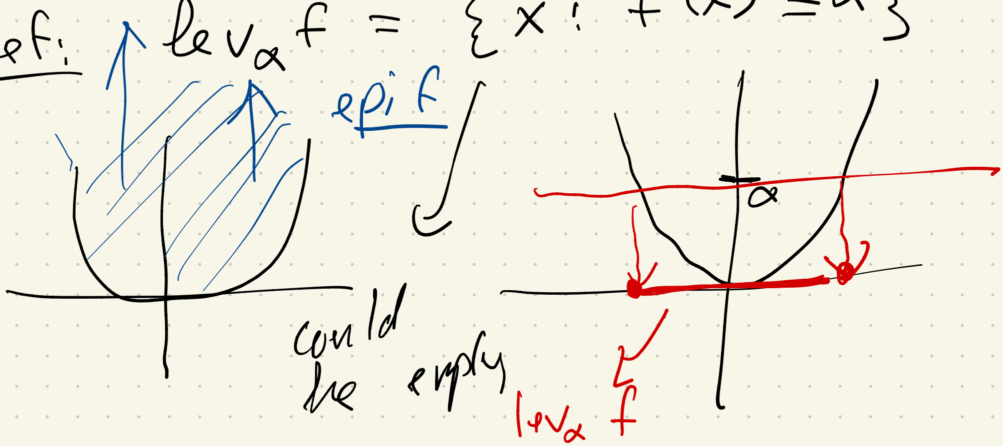
- $f \equiv \infty$, not useful
- $f(x) = -\infty$, not useful

Def: f is a proper function if $f(x) < \infty$ for at least one x, and $f(x)$ never equals $-\infty$ for any x.

* All functions we work with are proper.

Def: $\text{epi}\, f = \left\{ \begin{bmatrix} \alpha \\ x \end{bmatrix} : f(x) \le \alpha \right\}$

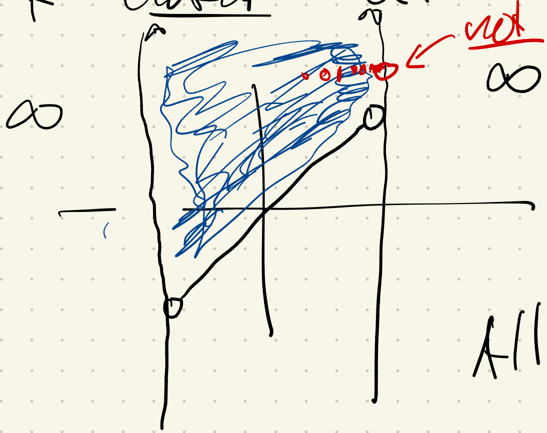
Def: $\text{lev}_\alpha f = \{ x : f(x) \le \alpha \}$



Now, we require <u>epi f</u> (equivalently, all $\text{lev}_\alpha f$) to be <u>closed</u>.

<u>Def:</u> f is a <u>closed</u> function if epi f (equivalently, all $\text{lev}_\alpha f$) are <u>closed sets</u>.

f <u>closed</u> eliminates $f(x) + \delta_{(0,1)}$

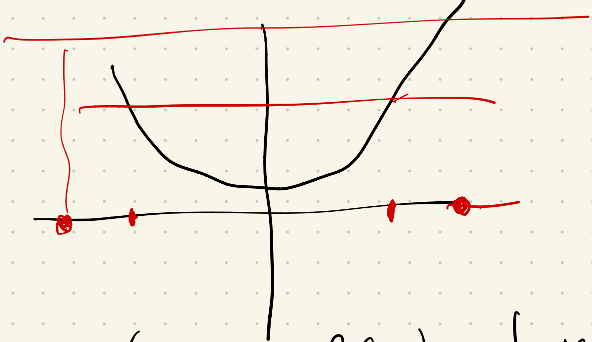


~~not~~ in epigraph

$f(1) = \infty$

All fcns we consider are <u>proper</u> and <u>closed</u>.

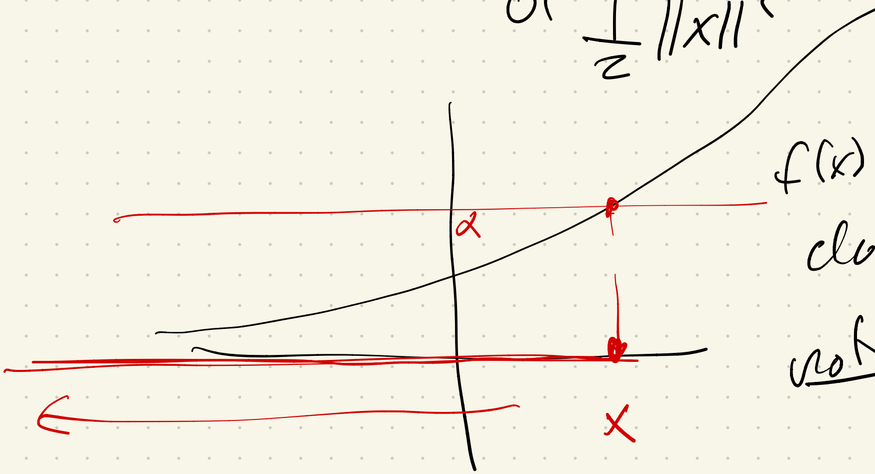
<u>Def:</u> $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is <u>level bounded</u> if all $\text{lev}_\alpha f$ are bounded (possibly empty)



is level bdd

ex: show $f(x) = \frac{1}{2}x^2$ has bdd level sets

or $\frac{1}{2}\|x\|^2$



$f(x) = e^x$ closed, but not level bdd

Thm: Rock-Wets, 1.9

Suppose $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ is closed, proper, and level bounded. then inf f is finite, and argmin f is nonempty and compact (closed & bounded).

$f(x) = e^x$, what are $\text{lev}_\alpha f(x)$?

$\text{lev}_\alpha f = \{x : f(x) \leq \alpha\}$

Given α, easily find x, $f(x) = \alpha$:

$\alpha = e^x$, so $\underline{x = \ln(\alpha)}$

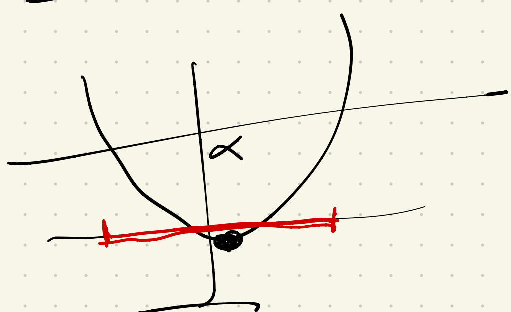
So $\text{lev}_\alpha e^x = \underline{(-\infty, \ln \alpha]}$

$\longrightarrow$ closed set.

All closed, $\therefore$ e^x closed fcn.

But $\text{lev}_\alpha e^x$ not bounded.

$f(x) = \frac{1}{2} x^2$



$\alpha = \frac{1}{2} x^2$, $x = \pm \sqrt{2\alpha}$

If $\alpha \geq 0$, then $\text{lev}_\alpha\left(\frac{1}{2}x^2\right) = [-\sqrt{2\alpha}, \sqrt{2\alpha}]$

$\alpha < 0$, $\text{lev}_\alpha(\,) = \{\}$

<u>Upshot:</u> the more assumptions we make, the nicer the behavior we can expect.

<u>Consider</u> closed, proper, <u>level bdd</u>, then they have minimizers.

+ C^1, minimizers identified by <u>$\nabla f(x) = 0$</u>

+ <u>convex</u>: $\nabla f(x) = 0 \Rightarrow$ global minimizer

+ get global minimizer is unique

Convex, but not level bdd.

* Very useful concept: convex fcns that are bounded below by quadratics.

<u>Def:</u> a function $f: \mathbb{R}^n \to \mathbb{R}$, C^1 is called <u>α-strongly convex</u> if

$$\underbrace{f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle}_{\text{convex grad ineq.}} + \underbrace{\frac{\alpha}{2}\|y-x\|^2}_{\text{quadratic}}$$

<u>Ex:</u> $f(x) = \frac{1}{2}\|x\|^2$ is 1-strongly cvx

$f(x) = g(x) + \frac{\alpha}{2}\|x\|^2$, g convex, is α-strongly convex

<u>Ex:</u> Show $f \in C^1$ is α-strongly cvx $\iff f(x) - \frac{\alpha}{2}\|x\|^2$ is cvx

Thm 2.23: For $f \in C^1$, TFAE:

1. f is α-strongly cvx
2. $\langle \nabla f(y) - \nabla f(x), y-x \rangle \geq \alpha \|y-x\|^2$
3. If f twice cont. diff'ble, then

$$\nabla^2 f \succeq \alpha I$$

$$\hookrightarrow \nabla^2 f - \alpha I \text{ psd}$$

$$\hookrightarrow \lambda(\nabla^2 f) \geq \alpha$$

Ex: Show that every α-strongly cvx function is level bounded.

$\therefore$ every α-strongly cvx fcn <u>has</u> a minimizer.

Ex: Show that an α-strongly cvx function must have a <u>unique</u> minimizer.

Next time: consider $\min_x f(x)$,

f is C^1.

Algo: $x^{k+1} = x^k - \gamma \nabla f(x^k)$

If all we know about f is that it is C^1, all we can look at is $\|\nabla f(x^k)\|$ → "are we close to a stationary pt?"

If in addition, f is convex and level bounded, minimizers exist and $f^* = \inf f$ is finite.

Can now look at $f(x^k) - f^*$

how big is this?

Finally, if f is α-strongly convex, then a unique minimizer exists, call it x^*, and can directly look at $\|x^k - x^*\|$

Three measures: $\|\nabla f(x^k)\|$

for three cases.

$f(x^k) - f^* \leftarrow$

$\|x^k - x^*\|$

Any of these measures will give us some sequence indexed by k, the iteration counter

$(x^{k+1} = x^k - \gamma \nabla f(x^k))$

$a_k = \|\nabla f(x^k)\|$

$a_k = f(x^k) - f^*$

$a_k = \|x^k - x^*\|$

a_k hopefully $\downarrow 0$

How fast?

How do we measure speed at which some sequence $a_k \downarrow 0$?

① $a_k \downarrow 0$ at a sublinear rate

if can find constants $c > 0, \theta > 0$

where $a_k \leq \frac{c}{k^\theta}$

Usually, we see $\theta = \frac{1}{2}, 1,$ or 2

$$a_k \leq \frac{c}{\sqrt{k}} \qquad a_k \leq \frac{c}{k^2}$$

(slowest) (fastest)

Usually, you terminate algorithm at a target ε for a_k

Q: After what k will $a_k \leq \varepsilon$?

$a_k \leq \varepsilon$, want min k so $\varepsilon = \frac{c}{k^\theta}$

$$k^\theta = \frac{c}{\varepsilon}$$

$$k \geq \left(\frac{c}{\varepsilon}\right)^{\frac{1}{\theta}}$$

c, θ both matter

$c = 1, \ \varepsilon = 1e^{-6} \quad \frac{c}{\varepsilon} = 10^6$

$q = \frac{1}{2}, \ k \gtrsim (10^6)^2 = 10^{12}$

$q = 2, \ k \geq (10^6)^{1/2} = 10^3$

(2) $a_k \downarrow 0$ at a linear (or geom) rate if $\underbrace{a_k \leq c(1-q)^k}_{*}$ $q \in (0,1)$

Ex: follow book with index p 28 to check why $*$ implies that given ε, $a_k \leq \varepsilon$ after $k \geq \frac{1}{q} \ln\left(\frac{c}{\varepsilon}\right)$

(3) $a_k \downarrow 0$ at a quadratic rate if

$a_{k+1} \leq c \cdot a_k^2$

$\leq \frac{1}{c}(c\, a_0)^{2^{k+1}}$ [Magic]