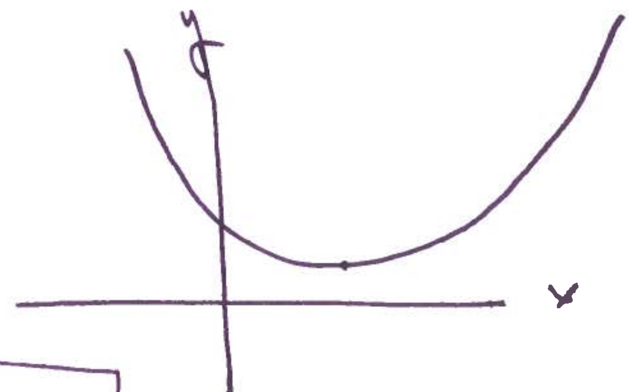
OH: 3-4 pm ~~in~~ ~~AHG 11~~ as in canvas announcement.

Also by appointment (skype)

Convex Opt'n
Lecture 2
Fall, 2016
– Constantine Caramanis

Last time: This class – algorithms for large scale

Example:



$$\boxed{y = ax^2 + bx + c}$$

curvature controlled by a

~~1/2 step size cannot be~~

For gradient descent:

$$x^+ = x - \gamma \cdot \underline{\nabla f(x)}$$

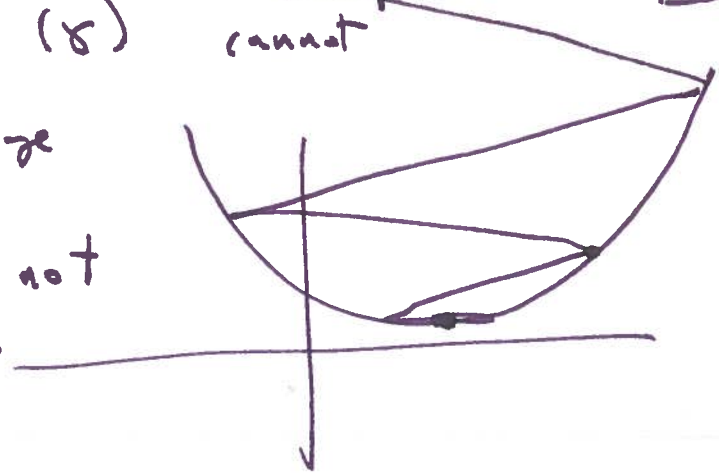
$\min f(x)$

Side note:

Opt: $f(\underline{x})$

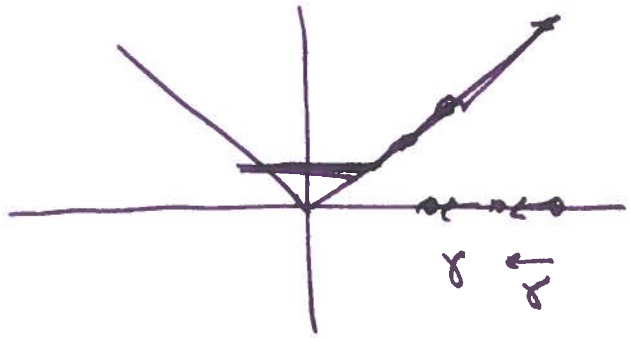
Stat: $f(\theta), f(p)$

1. step size (γ) cannot be too large



2. Step size does not need to be too small!

point 2: 'step size need not be too small'



$$f(x) = |x|$$

$$\nabla f(x) = \begin{cases} +1 & x > 0 \\ -1 & x < 0 \end{cases}$$

$x = 0$: later...

What happens w/ constant step size.

$$x^+ = x - \gamma \cdot \operatorname{sgn}(x)$$

No self-tuning.

Consequence: if you want ε error

then $\gamma \sim \varepsilon$.

$\#$ steps required $\sim 1/\varepsilon$.

Exercise: If you don't take too big step size for quadratic, then $\#$ steps required $\sim \log(1/\varepsilon)$

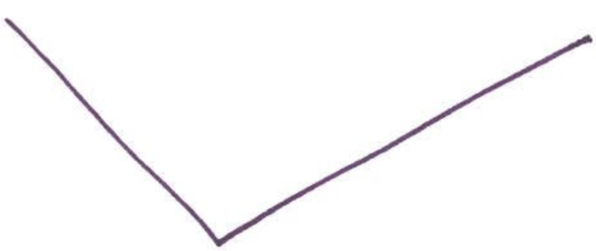
Fast convergence when curvature

is not too big

and not too small – namely,

not zero.

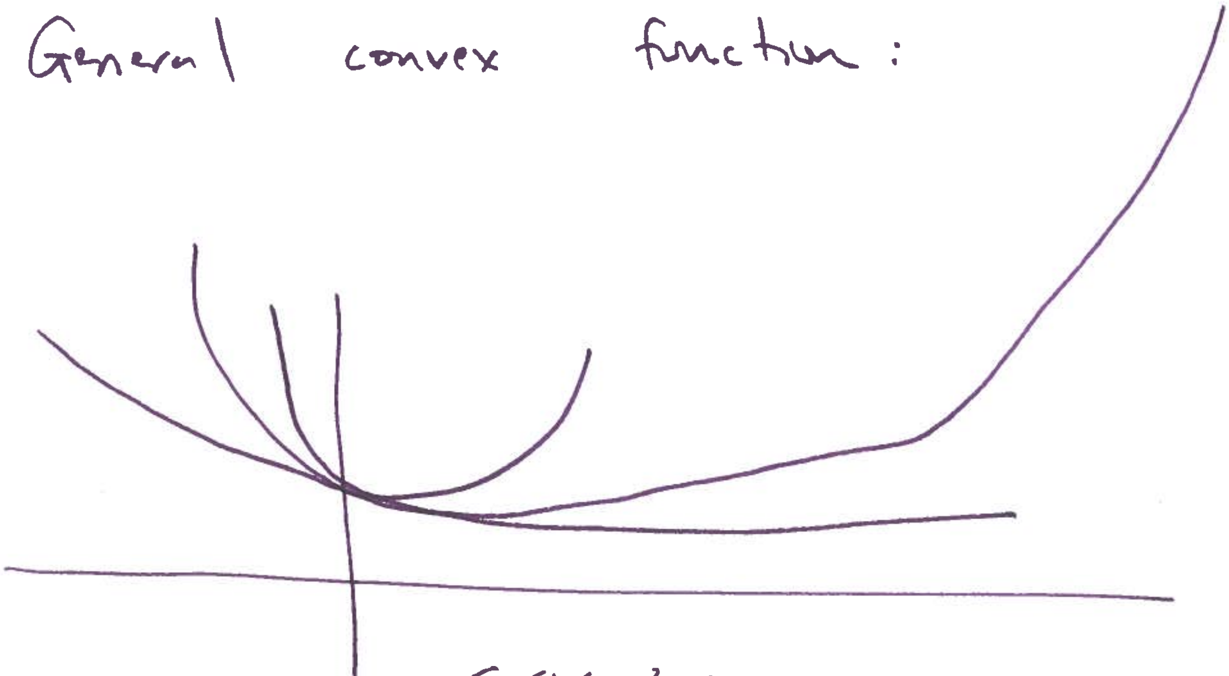
$\Rightarrow$ small step size



– no curvature

How will we use these simple ideas?

General convex function:



Simple but important idea: upper & lower bds on curvature.

This class: some basic definitions

$$\min: f(x)$$
$$\text{st}: x \in \mathcal{X}$$

Def'n: A set $\mathcal{X} \subseteq \mathbb{R}^n$ is convex, if

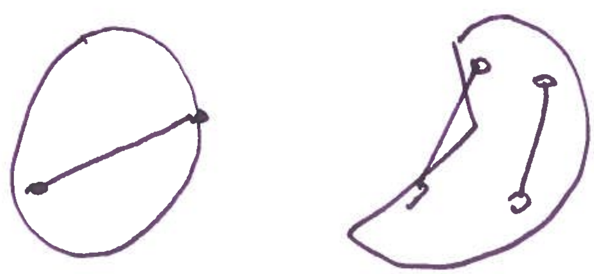
(much of convex analysis & opt'n extended beyond $\mathbb{R}^n$)

$\forall x_1, x_2 \in \mathcal{X}$, $\forall \lambda \in [0,1]$

$$x = \lambda x_1 + (1-\lambda) x_2 \in \mathcal{X}.$$

Geometric: If $x_1, x_2 \in \mathcal{X}$ then line segment $[x_1, x_2] \subseteq \mathcal{X}$

Ex:



Suppose M_1, M_2 are $n \times n$, symmetric matrices

Moreover: M_1, M_2 have all e-values ≥ 0.

$\mathcal{M}$ = set of all symm $n \times n$ matrices w/ e-values ≥ 0.

Q: Is $\mathcal{M}$ convex?

Fact: M symm $n \times n$ matrix

then $M = \sum_i \lambda_i v_i v_i^T$

where $\lambda_i \in \mathbb{R}$, $v_i \perp v_j$, $\|v_i\|_2 = 1$.

~~Check~~ ~~Pf~~

Check: λ_i are the e-values.

$M v_j = \lambda_j v_j$:

$M v_j = \sum_i (\lambda_i v_i v_i^T) v_j = \cancel{\lambda_j v_j v_j}$

$= \lambda_j v_j \underbrace{v_j^T v_j}_{1} = \lambda_j v_j$.

~~Check~~ **Exercise**:

M symm, $n \times n$, all e-vals ≥ 0

then $\forall x \in \mathbb{R}^n$, $x^T M x \geq 0$.

Exercise: If M symm $n \times n$, and $x^T M x \geq 0$ $\forall x \in \mathbb{R}^n \Rightarrow \lambda_i \geq 0$.

$\mathcal{M}$ set of all symm $n \times n$ matrices with e-vals ≥ 0 —

~~[illegible]~~ Claim: $\mathcal{M}$ is convex

pf: Need to show: $M_1, M_2 \in \mathcal{M}$

Then $\forall \lambda \in [0,1]$,

$$\left(\lambda M_1 + (1-\lambda) M_2\right) \in \mathcal{M}.$$

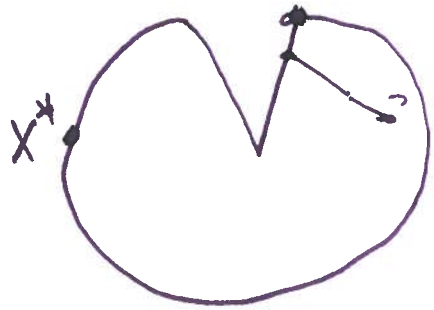
By fact on previous page, equivalently we can check:

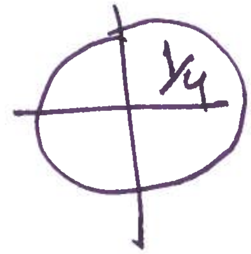
$$x^T \left(\lambda M_1 + (1-\lambda) M_2\right) x \geq 0 \quad \forall x \in \mathbb{R}^n.$$

$$= \lambda \underbrace{x^T M_1 x}_{\geq 0} + (1-\lambda) \underbrace{x^T M_2 x}_{\geq 0} \geq 0$$

$\Rightarrow$ $\mathcal{M}$ is convex.

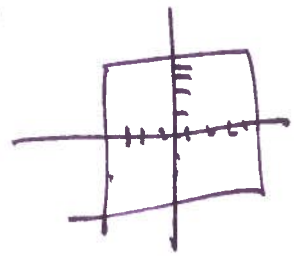
Why?



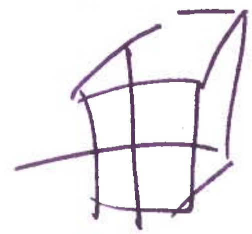


$\subseteq \mathbb{R}^n$

2^{-n}.

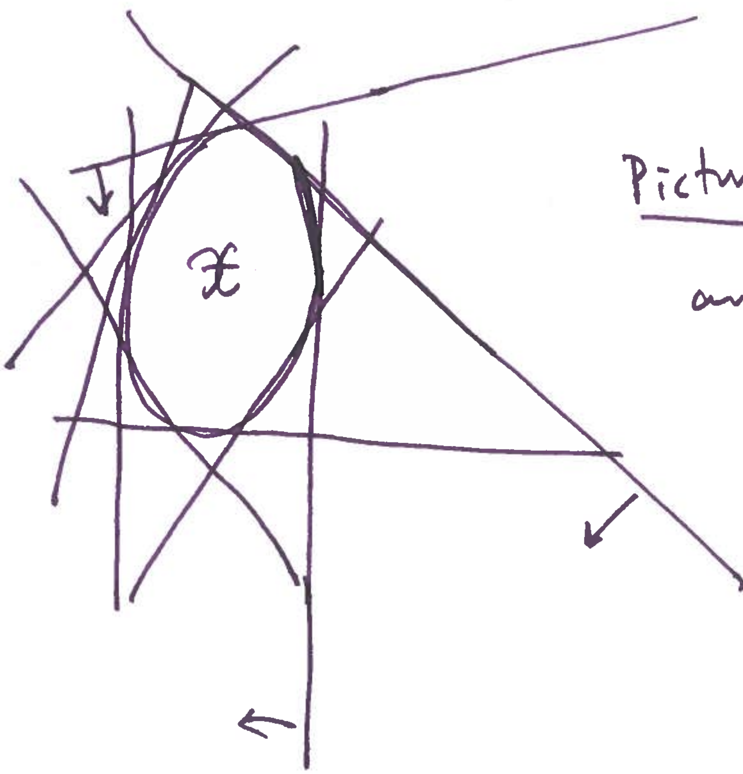


$(1/\varepsilon)^2$



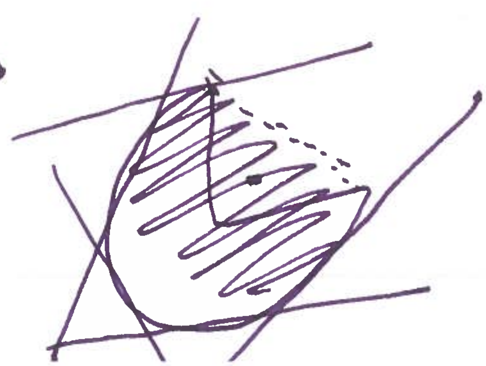
$\left(\frac{1}{\varepsilon}\right)^n$.

Another way to represent convex sets; another def'n.



Picture illustrates: $\mathfrak{X}$ is closed and convex iff

$$\mathfrak{X} = \bigcap_{H \supseteq \mathfrak{X}} H$$

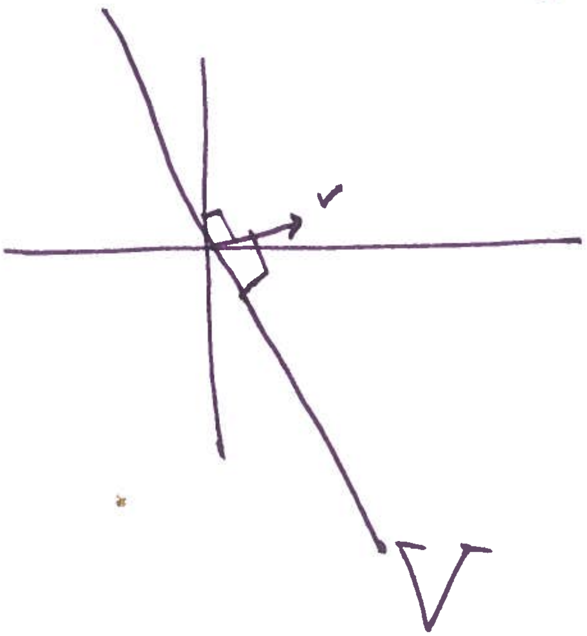


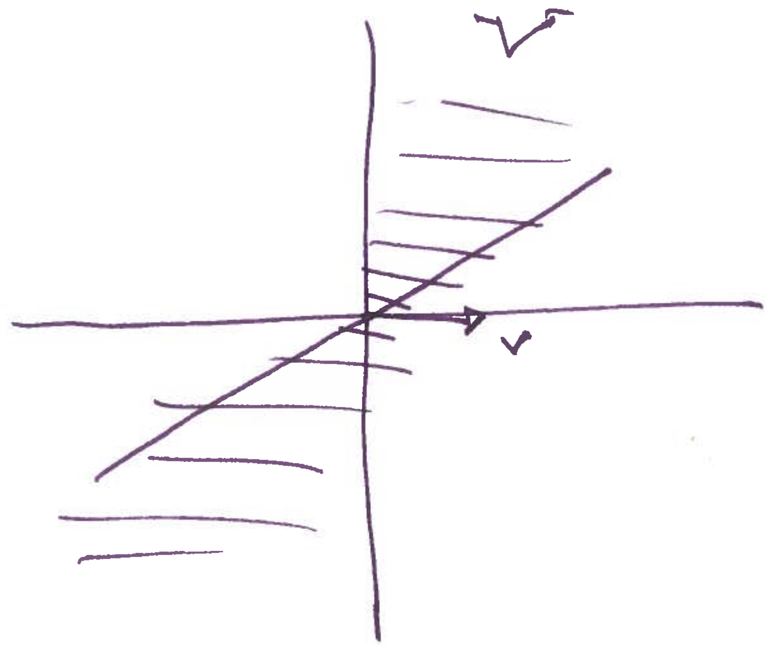
Import aside: Half-space

$\mathbb{R}^n$, pick $v \in \mathbb{R}^n$

$$V = \{ x : \langle x, v \rangle = 0 \}$$

Hyperplane ~~passing through origin~~ a





$x_1, \dots, x_n$ s.t. $x_1 v_1 + \dots + x_n v_n = 0$

"Hyperplane" dim'n $n-1$ linear space.

"co-dim 1"

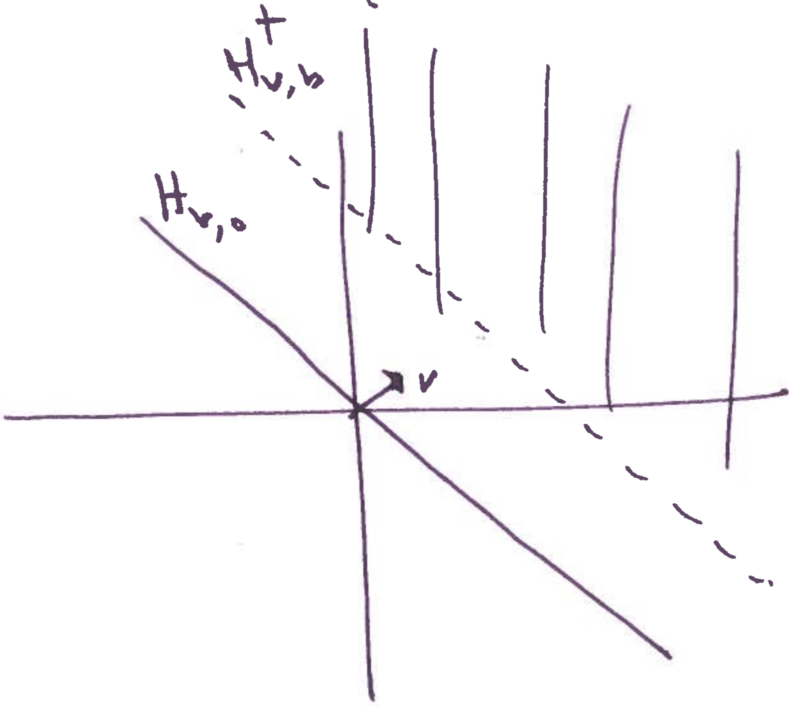
Half-space:

~~Scrap Page 2~~

$\cancel{H_+ = \{x :}$

$$H_{v,b} = \{x : \langle x, v \rangle = b\}$$

$$H^+_{v,b} = \{x : \langle x, v \rangle \geq b\}$$



<u>Fact</u>: If $\mathcal{X}$ is a closed convex set and $x \notin \mathcal{X}$ then $\exists$ H st

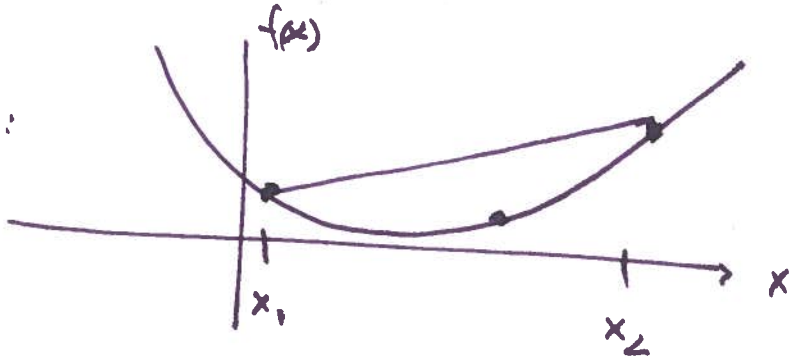
$\mathcal{X} \subseteq (H^+)^\circ$ $\quad x \in (H^-)^\circ$

Convex functions

$$\min: f(x) \quad \text{st}: x \in \mathcal{X}$$

Scrap Page1

Picture :



* concave if same holds w/ ineq. reversed.

$$f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

1 **Def'n**: f is called convex iff $\forall x_1, x_2$

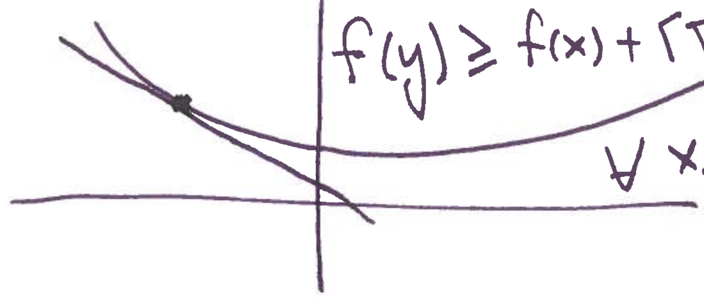
$\forall \lambda \in [0,1]$: $f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$.

Note: No regularity assumptions about f.

Def'n 2: A fn f ~~on~~ that is diff'ble is convex iff

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle$$

$\forall x, y$.



$$f(x+v) = f(x) + \langle \nabla f(x), v \rangle$$ 1st order T. approx'n.

Def'n 3: A fn f that is twice diff'ble, is convex iff $\nabla^2 f(x) \succeq 0$, i.e., has non-neg e-vals at all x.

Note:

Scrap Page

$$f(x) = \frac{1}{2} x^T Q x + q^T x + c$$

$$\nabla f(x) = Qx + q$$

$$\nabla^2 f(x) = Q \longleftrightarrow \text{indep. of } x!$$

Not true in general.

$$f(x) = ax^3 + bx^2 + cx + d$$

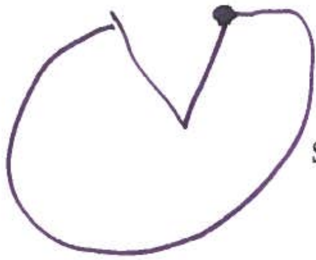
$$\frac{d^2 f}{dx^2} = 6ax + 2b$$

say $a = b = 1$. $f''(x) = 6x + 2$

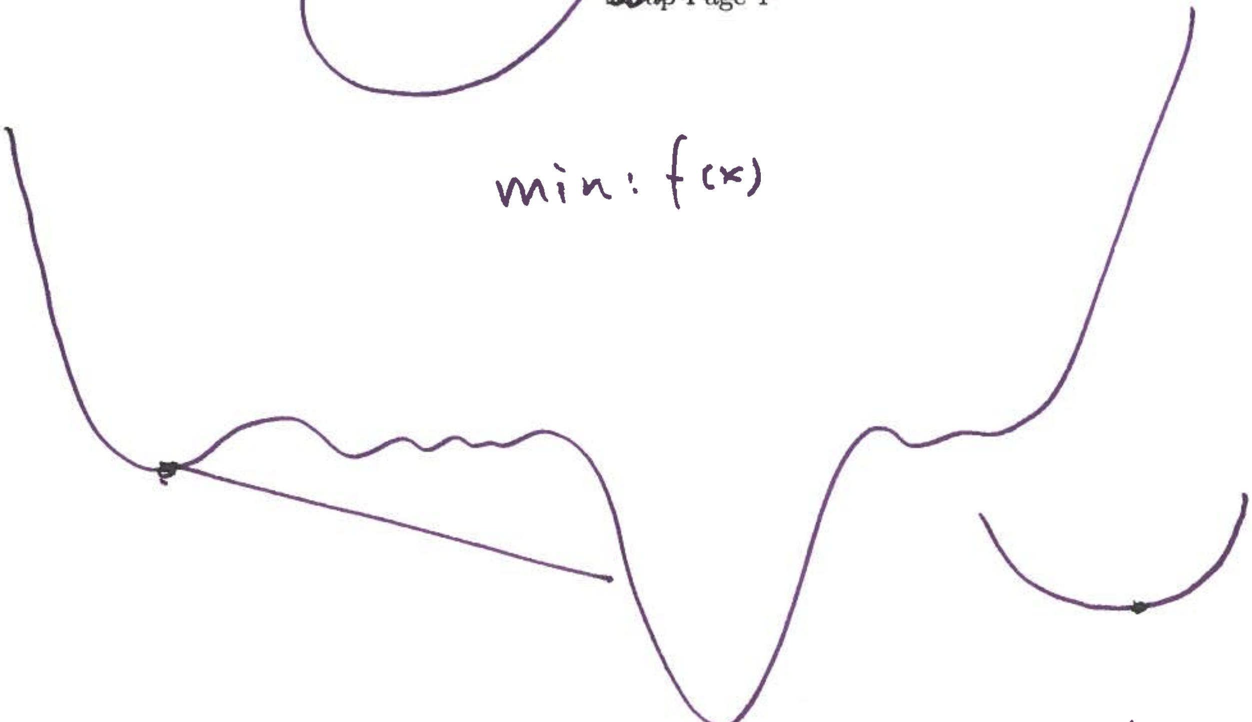
Convex for $x \geq -1/3$

non convex $x < -1/3$.

Why?



~~Scrap Page 1~~



Claim: Suppose f is diff'ble and is convex. ~~Then~~ ~~f~~ ~~f~~ ~~then~~, If ~~pf~~. we are at $\hat{x}$ s.t. $f'(\hat{x}) = 0$, then $\hat{x}$ is optimal.

pf: by def'n 2 of convex fn:

$$f(y) \geq f(\hat{x}) + \cancel{\langle \nabla f(\hat{x}), y - \hat{x} \rangle} \quad \forall y$$

$\blacksquare$

Regression. Quadratic fns & beyond...

→ convex fns.

Convex opt'n can be used for non-convex problems.

<u>Regression</u>: $y = X\beta$

$\beta \in \mathbb{R}^p \qquad X \quad n \times p$

n = # data points $\qquad p$ = # dec vars
deg. of f.
dim'n.

Generally: $p > n$.

$(p - n)$

ex: $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$p \gg n$. But: β is sparse! only has a few non-zero values.

eq. $\qquad$ # vars. $\qquad$ β 4 non-zero

$$[y] = \begin{bmatrix} & X & \end{bmatrix} [\beta]$$

Natural guess :

$$\min: \|X\beta - y\|_2^2$$
$$\text{st}: \underbrace{\beta \text{ has } 4 \text{ non-zeros}}_{\text{non-convex constraint.}}$$

$p = 10^6$

$$\binom{10^6}{4} \sim (10^6)^4.$$

Convex opt idea :

$$\min: \|X\beta - y\|_2^2 + \lambda \|\beta\|_1$$

$\Sigma |\beta_i|$