Introduction to
Optimisation Course

École Centrale de Lyon
January 2023

- Convex optimisation appears everywhere in science and engineering, mostly on the form of a convex program:

$$\underset{x \in K \subset \mathbb{R}^n}{\operatorname{argmin}} f(x)$$

where
- K is a convex constraint set
- $f(x)$ is a convex function

Example:

Linear inverse problems:
(medical imaging, photography, astronomy, etc.)

...

$$y = Ax$$

A: linear degradation process

x: image we want to recover

We can recover a sparse image by solving:

$$\underset{x \in K}{\operatorname{argmin}} \|x\|_1$$

where $K = \{x \in \mathbb{R}^n / y = Ax\}$

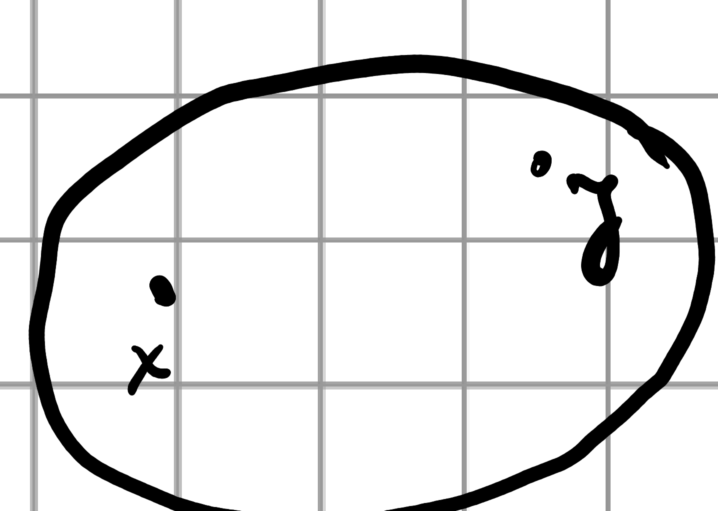
REFERENCES:

- "Mathematical Introduction to Compressed Sensing" by Foucart and Rauhut (Appendix B)
- "Convex Optimization" by Boyd and Vandemberghe (Chapters 2, 3, 4, 5, 9)
- "Proximal Algorithms" by Parikh et al. (Chapters 2, 3, 4)

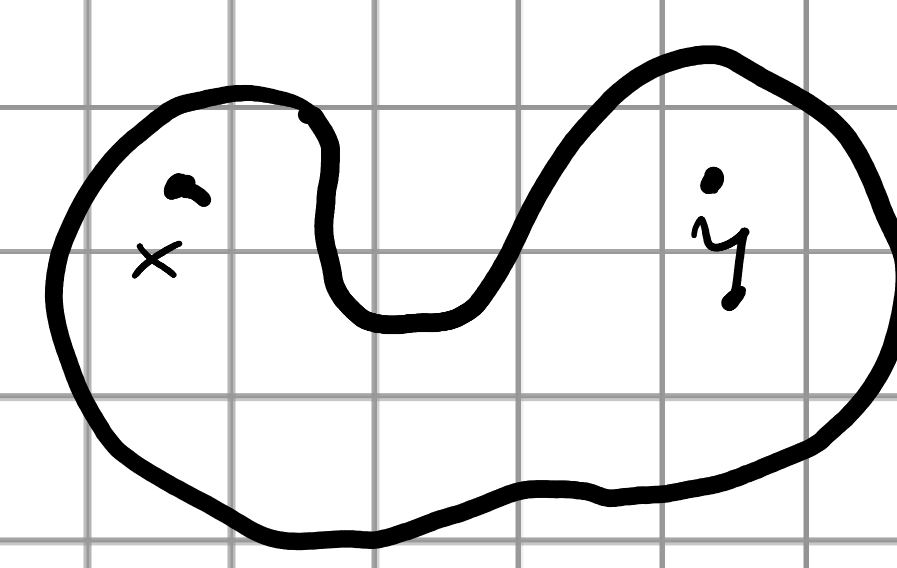
CONVEX SETS

Def convex set is a subset $K \subseteq \mathbb{R}^n$

$$\text{iif } \forall t \in [0,1],\ tx + (1-t)y \in K \quad \forall x, y \in K$$



CONVEX

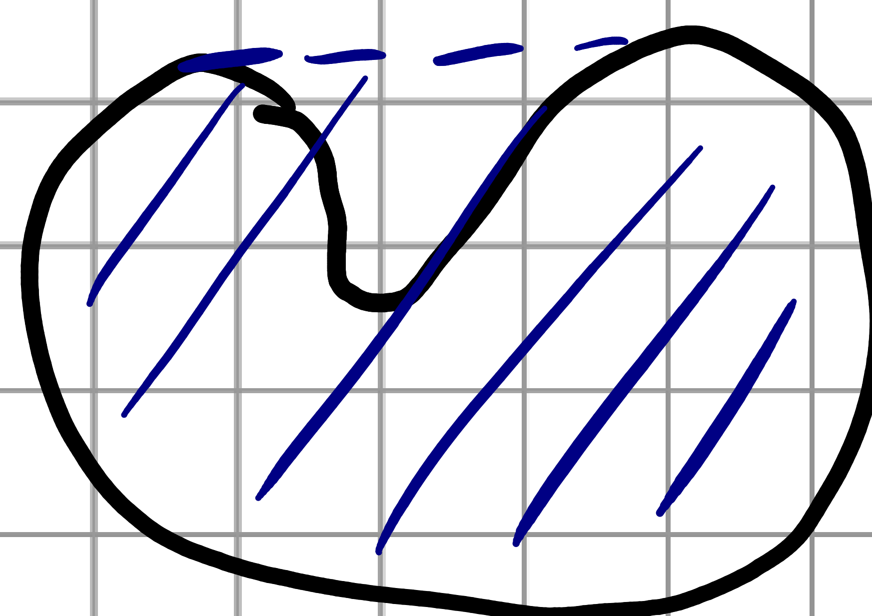


NON-CONVEX

- The intersection of convex sets is convex!
- The union of convex sets is NOT necessarily convex.

Def Convex hull of a subset $K \subset \mathbb{R}^n$

$$\mathrm{conv}(K) = \left\{ \sum_{j=1}^{N} t_j x_j \; : \; N \geq 1, \; t_j \geq 0, \; \sum_{j=1}^{N} t_j = 1, \; x_1, \dots, x \right\}$$



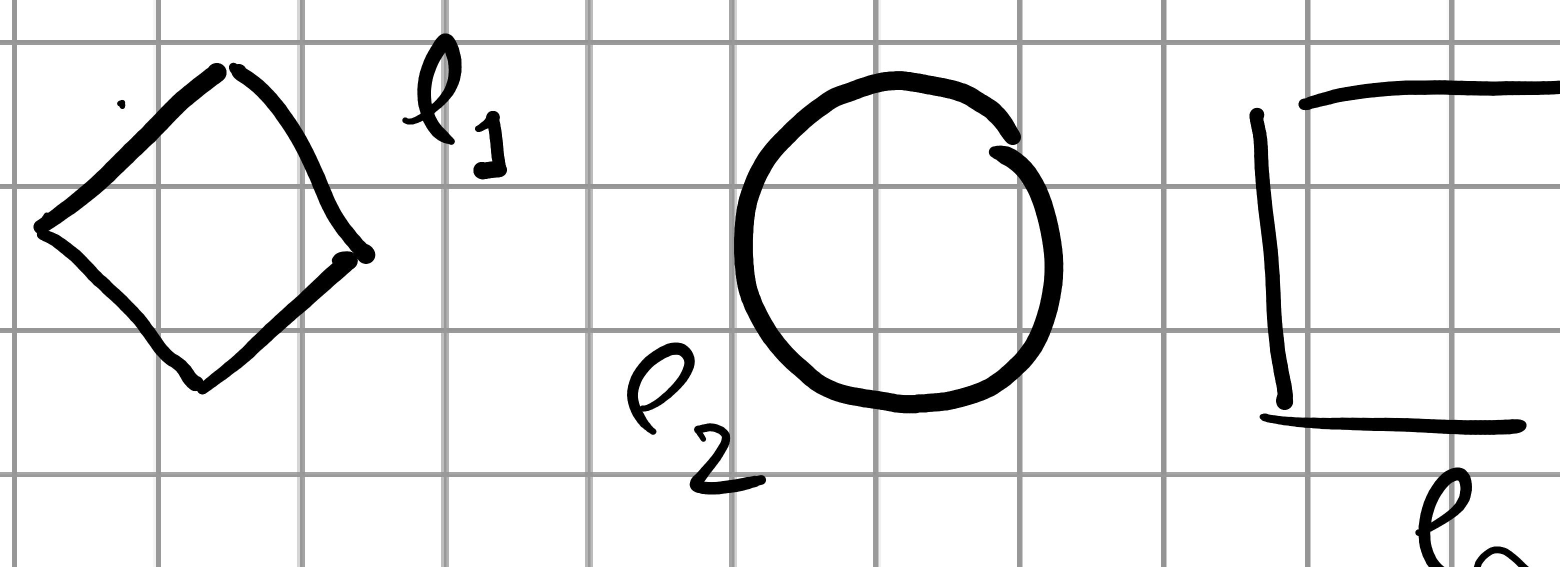
K

conv(K)

$K \subseteq \mathrm{conv}(K)$

- it is the smallest set containing K
- K is convex $\Leftrightarrow$ $\mathrm{conv}(K) = K$

- Examples of convex sets.
  - p-balls $B_r^p = \{ x : \|x\|_p \leq r \}$

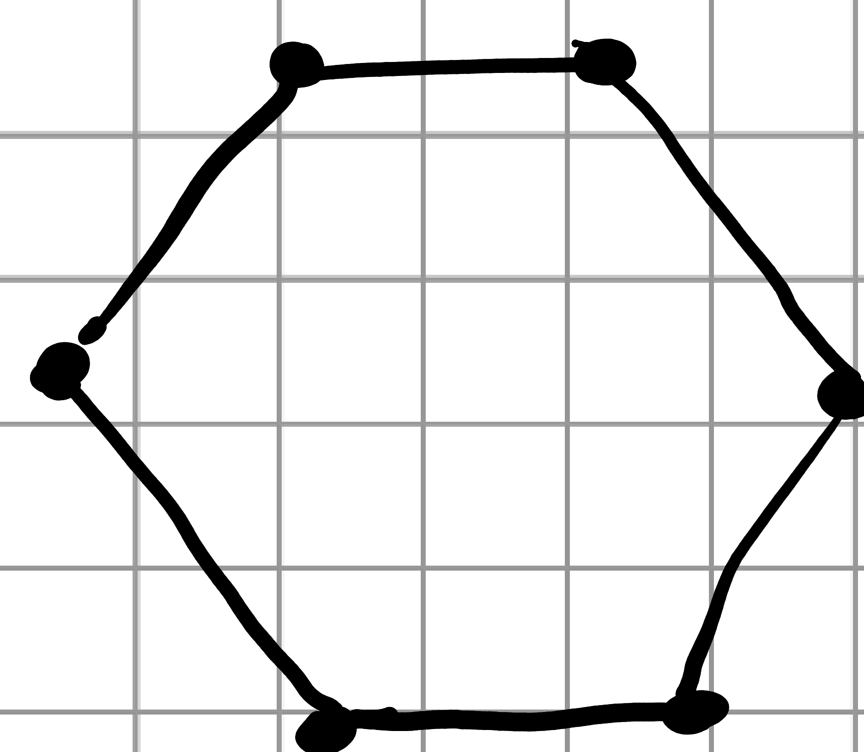


  - affine subspaces $\{ x : Ax = y \}$
  - halfspaces (ω, λ) $\{ x : x^T \omega \leq \lambda \}$

Def: Extreme points

Let $K \subset \mathbb{R}^n$ be a convex set. A point $x \in K$ is an extreme point iif $x = ty + (1-t)z$ and $y, z \in K$, $t \in [0,1]$ $\implies y = z = x$

FRANCE IS A CONVEX SET!



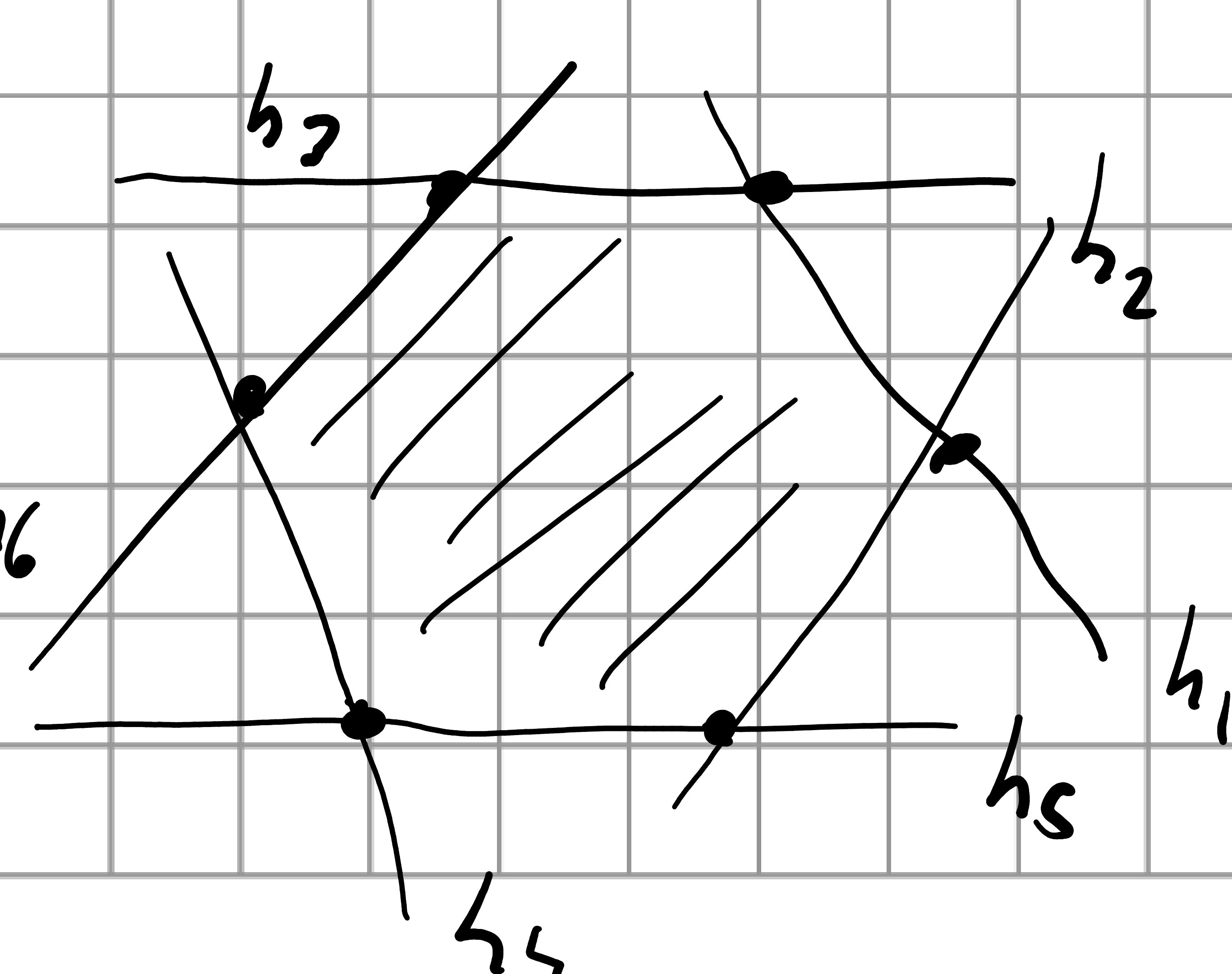
EXTREME POINT

Theorem: A convex set is the convex hull of its extreme points.

Duality:

Any convex set can be represented as an intersection of a collection H of halfspaces

$$K = \bigcap_{h \in H} h$$



CONVEX FUNCTIONS

$$f : \mathbb{R}^n \longrightarrow \,]-\infty, \infty] = \mathbb{R} \cup \{\infty\}$$

WITH CONVENTION $x + \infty = \infty, \; x \in \mathbb{R}$

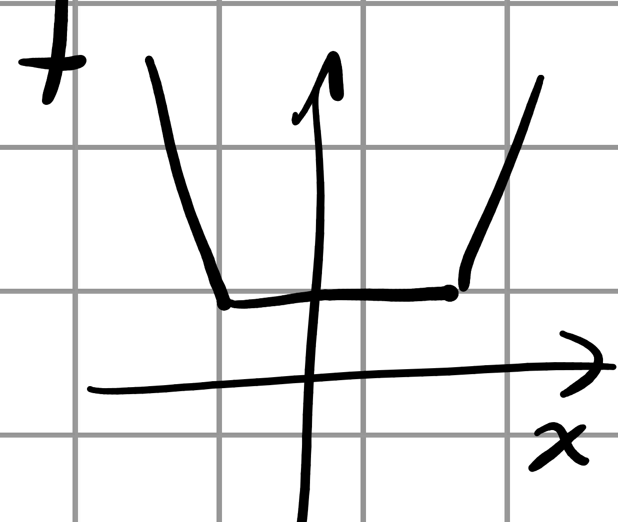
$\lambda \infty = \infty, \; \lambda > 0$

DOMAIN:

$$\text{DOM}(f) = \{ x \in \mathbb{R}^n, \; f(x) \neq \infty \}$$

– A FUNCTION IS <u>proper</u> IF $\text{dom}(f)$ IS NOT EMPTY.

DEFINITIONS: $f : \mathbb{R}^n \to \,]-\infty, \infty)$ IS CALLED



- CONVEX if $f(tx + (1-t)y) \leq t f(x) + (1-t) f(y)$

$\forall x, y \in \mathbb{R}^n, \; \forall t \in [0, 1]$

- STRICTLY CONVEX if

$$f(tx + (1-t)y) < t f(x) + (1-t) f(x)$$

$\forall x, y \in \mathbb{R}^n, \; \forall t \in \,]0, 1[$

- STRONGLY CONVEX WITH param $\gamma > 0$ if

$$f(tx + (1-t)y) \leq t f(x) + (1-t) f(y) - \frac{\gamma}{2} t(1-t) \|x - y\|^2$$

Useful properties:

$f : \mathbb{R} \to \mathbb{R}$ convex non decreasing

$g : \mathbb{R}^n \to \mathbb{R}$ convex

$\Rightarrow$ $f \circ g$ is convex

Example $g(x) = x^2$ $f(x) = ax$ $a \geq 0$

$f \circ g(x) = ax^2$ is convex for any $a \geq 0$

Examples of convex functions:

Quadratic function • $f(x) = x^T A x$

is convex if $A \in \mathbb{R}^{n \times n}$ is positive semi definite

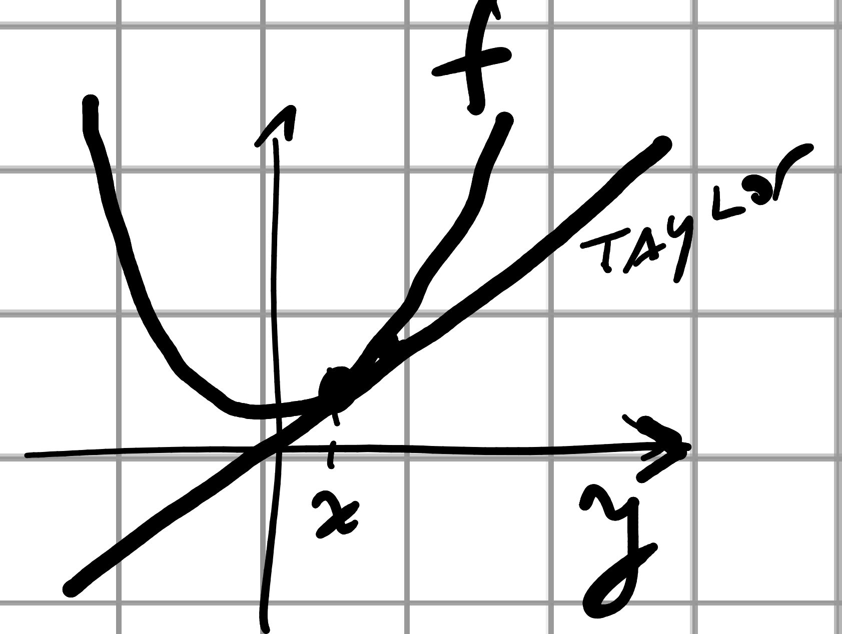
and is strongly convex if A is positive definite

Indicator • $f(x) = \mathbb{1}_{x \in K}$ where K is a convex set

$$= \begin{cases} +\infty & \text{if } x \notin K \\ 0 & \text{if } x \in K \end{cases}$$

Proposition: Let $f: \mathbb{R}^n \to \mathbb{R}$ be a DIFFERENTIABLE FUNCTION

- f is convex iif $\forall x, y \in \mathbb{R}^n$



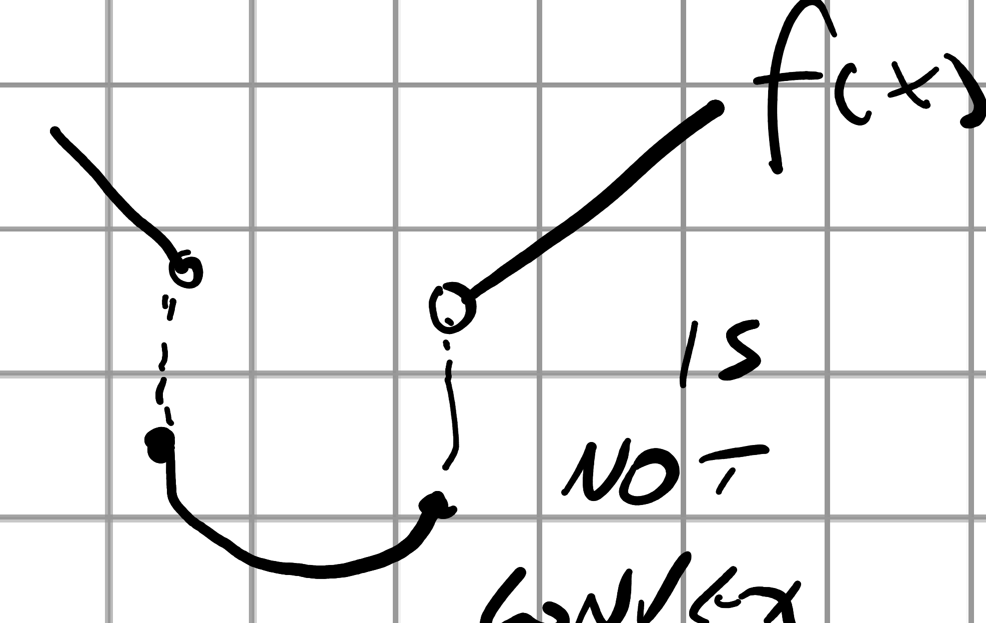
$$f(y) \geq f(x) + \nabla f^T (y-x)$$

- f is strongly convex iif $\forall x, y \in \mathbb{R}^n$

$$f(x) \geq f(y) + \nabla f^T (x-y) + \frac{\gamma}{2} \|x-y\|^2$$

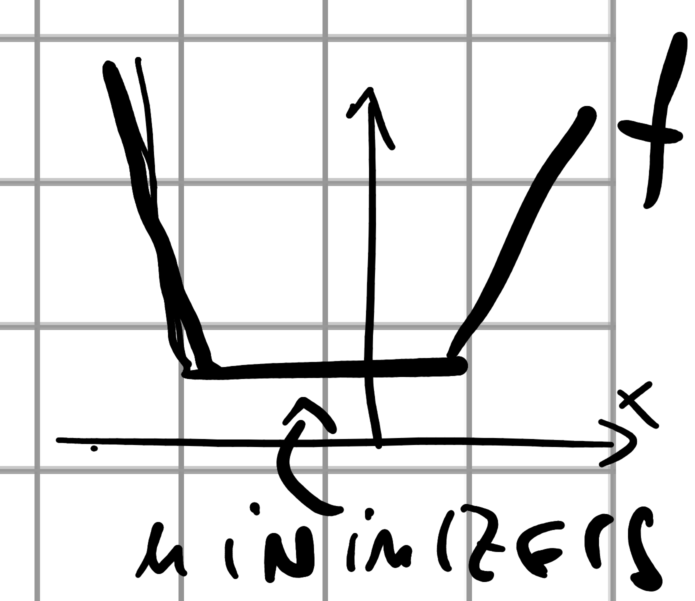
- If f is TWICE DIFFERENTIABLE then it is convex iif $\forall x \in \mathbb{R}^n$ $\nabla^2 f(x) \succeq 0$ (MATRIX IS PSD)

- A CONVEX FUNCTION IS CONTINUOUS ON THE INTERIOR OF ITS DOMAIN

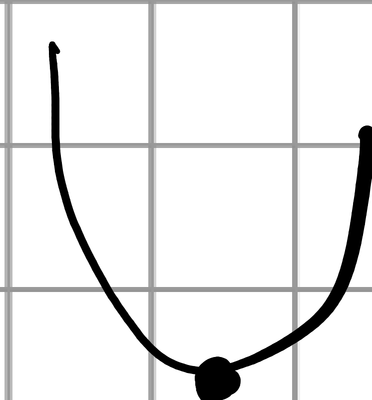


!! **Prop.** $f : \mathbb{R}^n \to]-\infty, +\infty]$ be a convex function

- A local minimizer of f is a <u>global</u> minimizer
- The set of minimizers is convex



- If f is stricty convex there is a unique minimizer



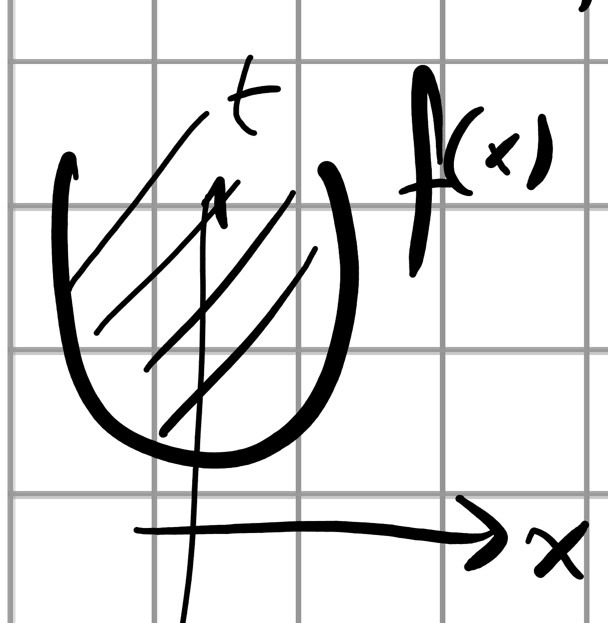
Theorem

Let $f : \mathbb{R}^n \times \mathbb{R}^m \to]-\infty, +\infty]$ be a convex function.

Then $g(x) = \inf_{y \in \mathbb{R}^m} f(x,y)$ is a convex function

$g : \mathbb{R}^n \to]-\infty, +\infty]$

Epigraph:

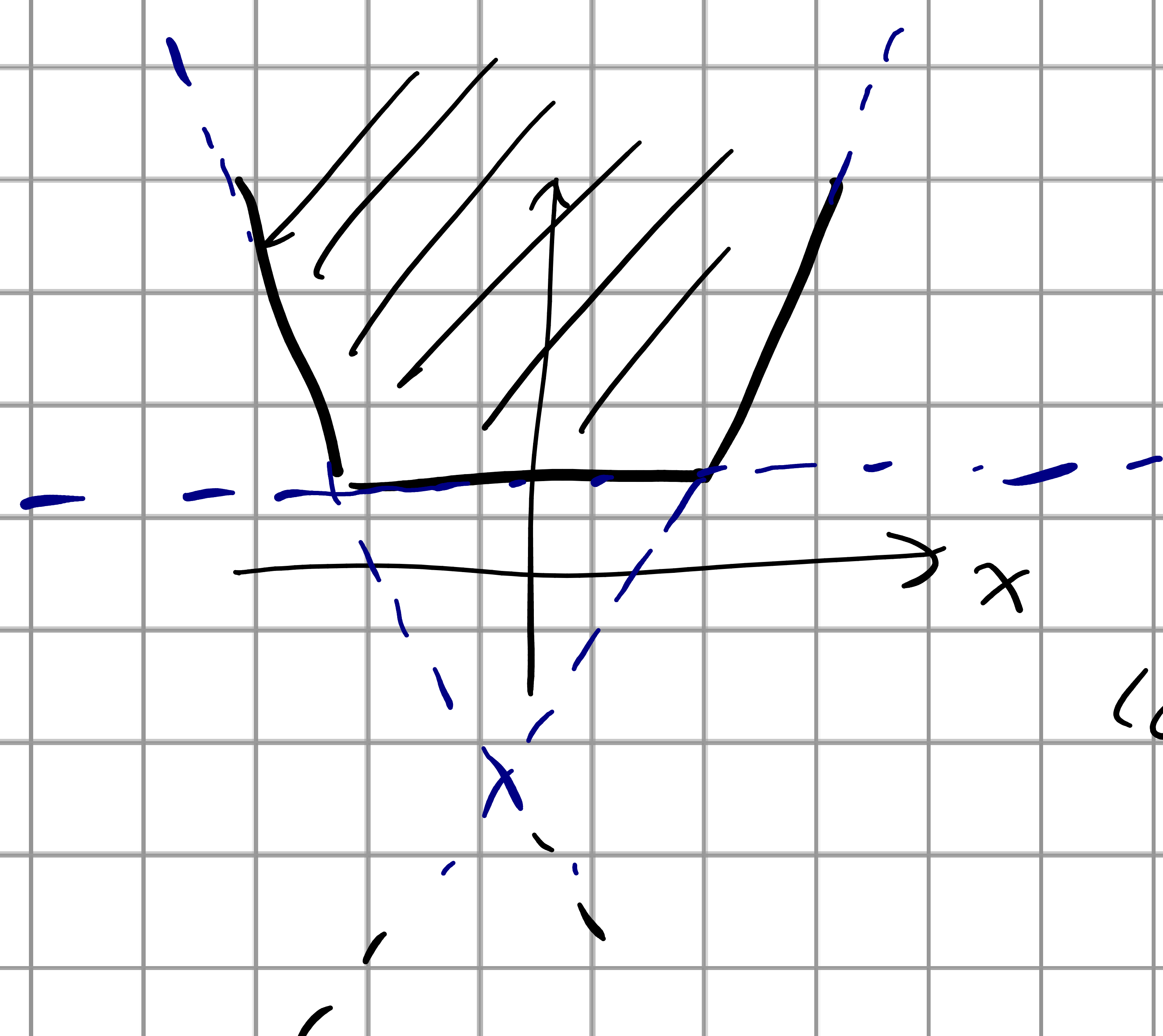


$$epi(f) = \{ (t, x) \in \mathbb{R}^{n+1}, \; f(x) \leq t \}$$

$epi(f)$ is a convex set $\iff$ f is a convex function

CONVEX CONJUGATE

DUALITY FOR FUNCTIONS:

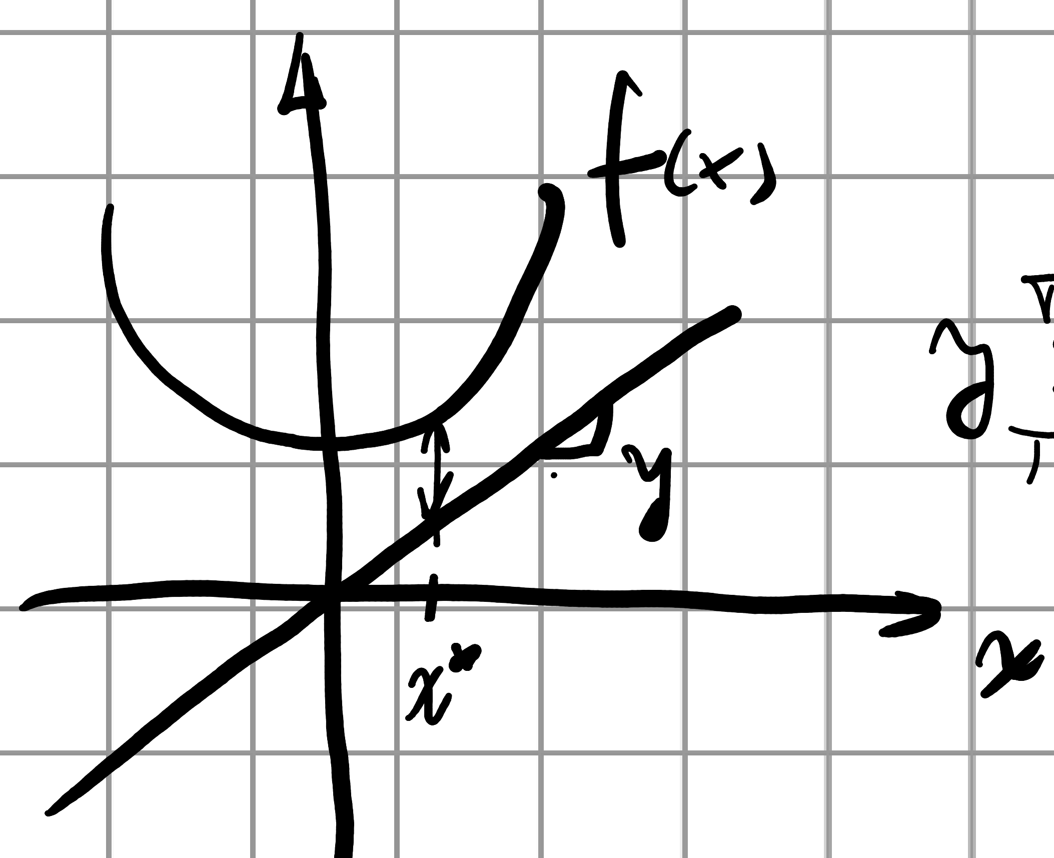


IDEA: epi(f) CAN BE EXPRESSED AS INTERSECTION OF HALFSPACES

$\Updownarrow$

CONVEX f CAN BE EXPRESSED AS SUPREMUM OVER LINEAR FUNCTIONS

$$f(x) = \sup_y \; y^T x - \underbrace{b_y}_{\text{HOW SHOULD WE CHOOSE IT?}}$$



$y^T x$ HAS TO BE BELOW $f(x)$!

$(y^T x + b_y \le f(x))$

DEFINITION: FENCHEL-DUAL (CONVEX CONJUGATE)

$$f^*(y) = \sup_x \; y^T x - f(x)$$

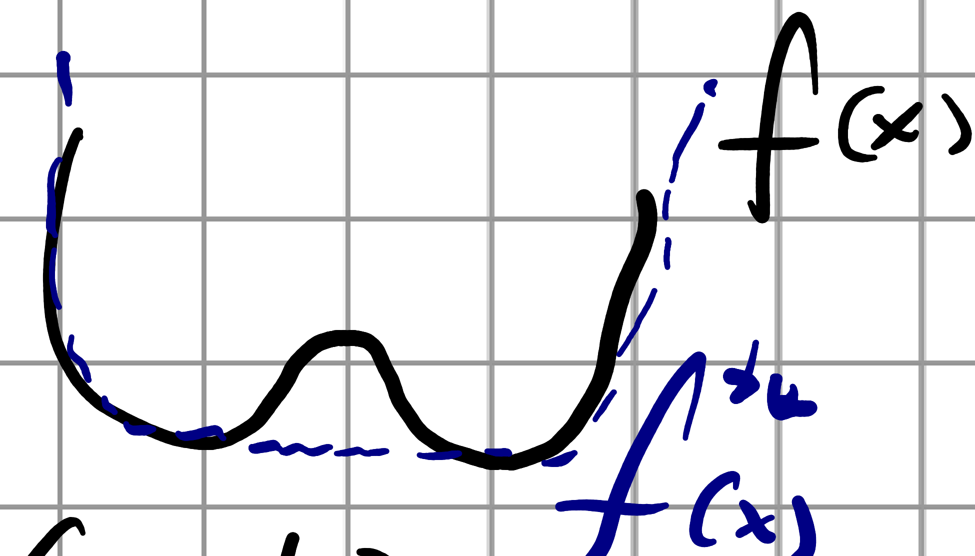
$f^*(y)$ IS CONVEX INDEPENDENTLY OF WHETHER f IS CONVEX OR NOT

FENCHEL-YOUNG $\forall x, y \in \mathbb{R}^n$

$$x^T y \leq f(x) + f^*(y)$$

PROPERTIES OF CONVEX CONJUGATE: $f: \mathbb{R}^n \to]-\infty, +\infty]$

1) f^* is LOWER SEMI CONTINUOUS (LSC)
2) f^{**} is THE LARGEST LSC CONVEX FUNCTION SATISFYING $f^{**}(x) \leq f(x)$ $\forall x \in \mathbb{R}$



3) $\mathrm{epi}(f^{**}) = \mathrm{conv}(\mathrm{epi}(f))$
4) If f is CONVEX $f^{**} = f$!
5) For $\lambda \neq 0$ $[f(\lambda x)]^* = f^*\left(\frac{y}{\lambda}\right)$
6) For $\lambda > 0$, $(\lambda f(x))^* = \lambda f^*\left(\frac{y}{\lambda}\right)$
7) For $z \in \mathbb{R}^n$ $(f(x-z))^* = f^*(y) + z^T y$

Example:

$$\cdot\ f(x) = \|x\|_p \Rightarrow f^*(y) = \mathbb{1}_{\underbrace{\|y\|_q \leq 1}_{\text{"DUAL BALL"}}}$$

WHERE $1 = \frac{1}{p} + \frac{1}{q}$

$l_2 \quad p=2 \Rightarrow q=2 \quad l_2$

$l_1 \quad p=1 \Rightarrow q=\infty \quad l_\infty$

DEF. THE SUBDIFFERENTIAL OF A CONVEX FUNCTION $f: \mathbb{R}^n \to]-\infty, \infty]$ AT POINT $x \in \mathbb{R}^n$ IS

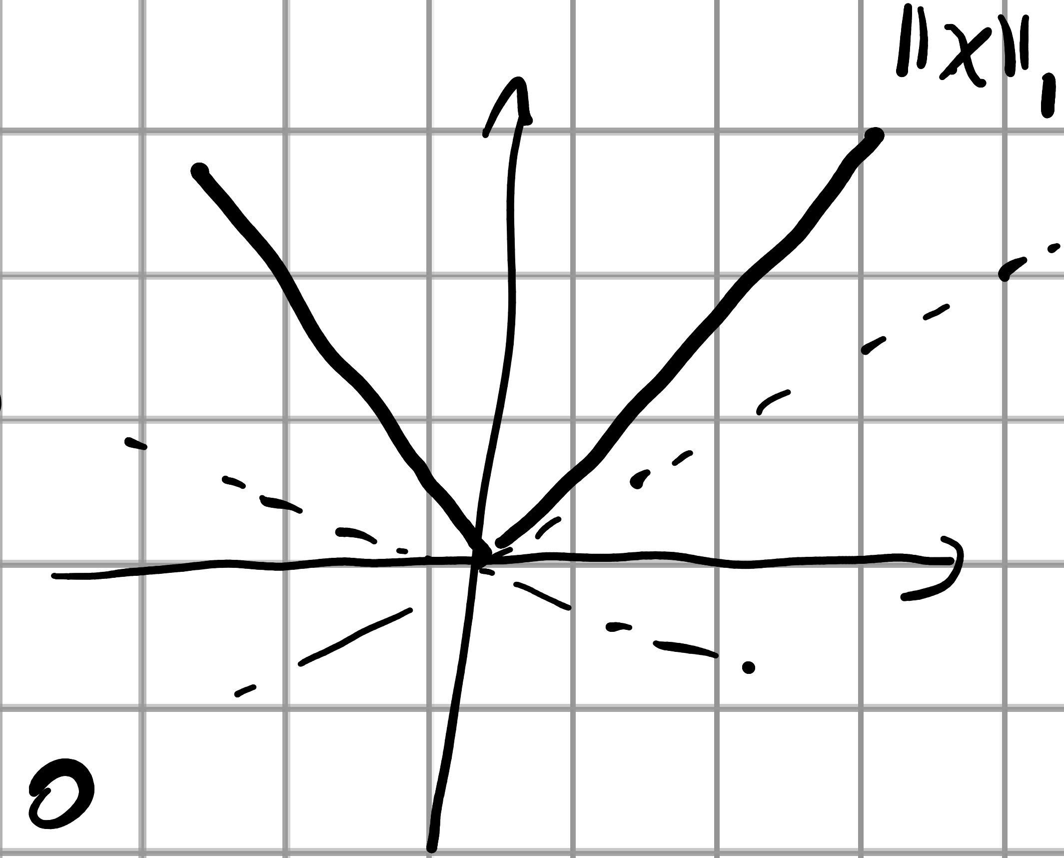
THE SET $$\partial f(x) = \{ v \in \mathbb{R}^n,\ f(y) \geq f(x) + v^T(y-x)\ \forall y \in \mathbb{R}^n \}$$

- THE SUBDIFFERENTIAL OF A CONVEX FUNCTION IS ALWAYS NON-EMPTY
- IF f IS DIFFERENTIABLE AT x

$$\Rightarrow \partial f(x) = \{ \nabla f(x) \}$$ SINGLE POINT.

Example: $f(x) = \|x\|_1$

$$\partial f(x) = \begin{cases} \{\text{sign}(x)\} & \text{if } x \neq 0 \\ [-1, 1] & \text{if } x = 0 \end{cases}$$



Theorem: $x \in \mathbb{R}^n$ is a minimum of convex f iff $0 \in \partial f(x)$

Theorem $f : \mathbb{R}^n \to]-\infty, +\infty]$ convex, l.s.c., $x, y \in \mathbb{R}^n$

(a) $y \in \partial f(x)$

(b) $x \in \partial f^*(y)$

(c) $y^T x = f(x) + f^*(y)$ ← equality!

(a) $\iff$ (b) $\iff$ (c)

e.g. $y = 0$ $\quad 0 \in \partial f(x) \iff \underbrace{x \in \partial f^*(0)}_{\text{DUAL WAY TO OPTIMIZE } f(x)}$

PROXIMAL MAPPING : $f : \mathbb{R}^n \longrightarrow]-\infty, \infty]$
CONVEX

$$\text{prox}_f(x) = \underset{x \in \mathbb{R}^n}{\text{argmin}}\ f(x) + \frac{1}{2}\|x - z\|^2$$

↳ IT HAS A UNIQUE MINIMIZER SINCE $f + \|\cdot - z\|^2$ IS STRONGLY CONVEX!

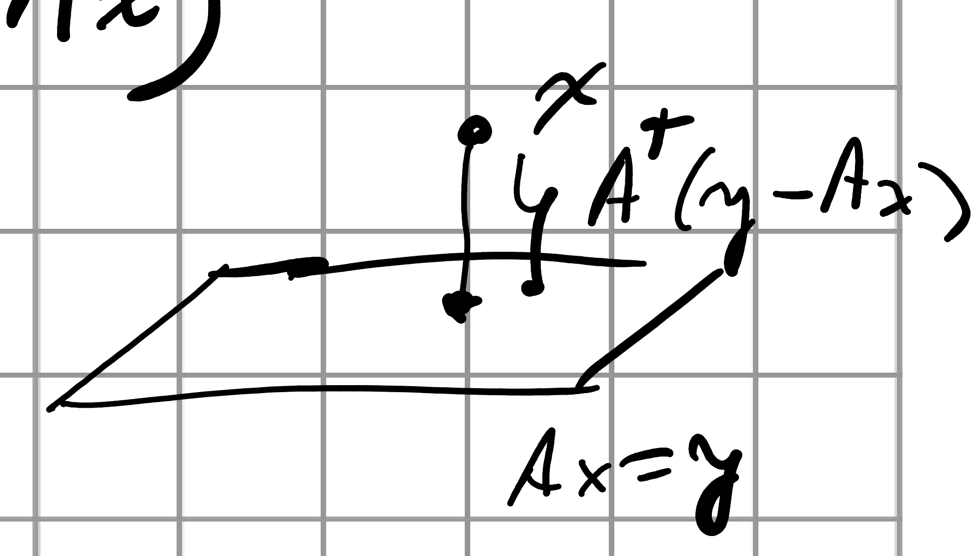
EXAMPLES:

① $f(x) = \mathbb{1}_{x \in K}$ $\quad$ K CONVEX

$\Rightarrow \text{prox}_f(x)$ IS THE ORTHOGONAL PROJECTION ONTO K

$$K = \{ x \in \mathbb{R}^n \,/\, Ax = y \}$$

$$\text{prox}_f(x) = x + A^+(y - Ax)$$



② $f(x) = \frac{\lambda}{2}\|y - Ax\|^2$

$$\underset{x}{\text{argmin}}\ \frac{1}{2}\|x - z\|^2 + \frac{\lambda}{2}\|y - Ax\|^2$$

$$\frac{d}{dx} \rightarrow (x - z) + \lambda A^T(Ax - y) = 0$$

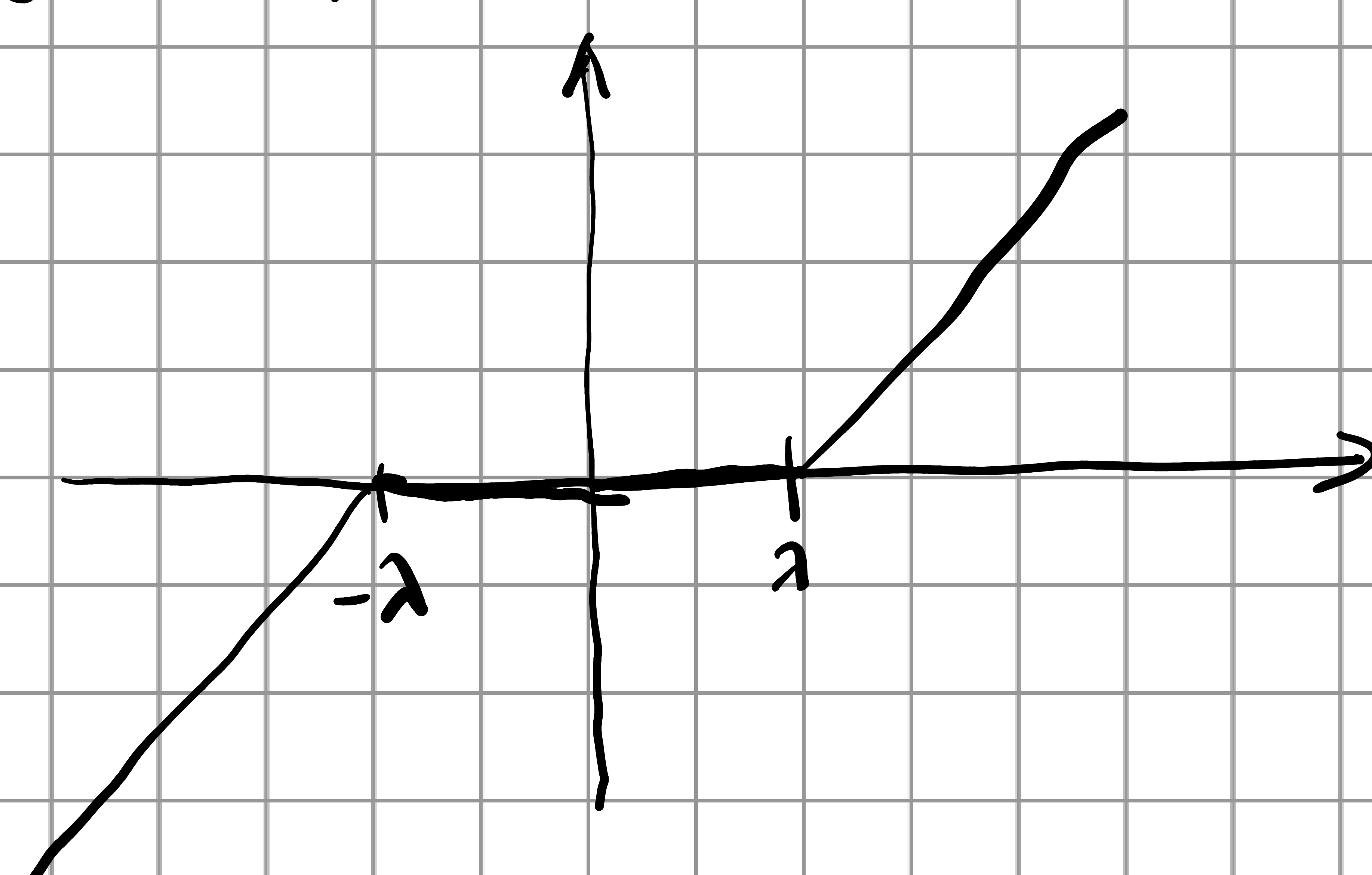
$$x = (I + \lambda A^T A)^{-1}(z + \lambda y)$$

(3) $f(x) = \lambda \|x\|_1$

$$\arg\min_x \lambda \|x\|_1 + \frac{1}{2}\|x - z\|^2$$

$$= \begin{cases} y - \lambda & \text{if } y \geq \lambda \\ y + \lambda & \text{if } y \leq -\lambda \\ 0 & \text{if } |y| \leq \lambda \end{cases}$$

SOFT THRESHOLDING



Proposition $f: \mathbb{R}^n \to]-\infty, \infty]$ CONVEX

$$x = \mathrm{prox}_f(z) \iff z \in x + \partial f(x)$$

$\Downarrow$ ALSO WRITTEN AS

$$x = (\partial f + I)^{-1} z$$

THEOREM: MOREAU'S IDENTITY

$f: \mathbb{R}^n \to]-\infty, \infty]$ CONVEX L.S.C.

THEN $\forall z \in \mathbb{R}^n$

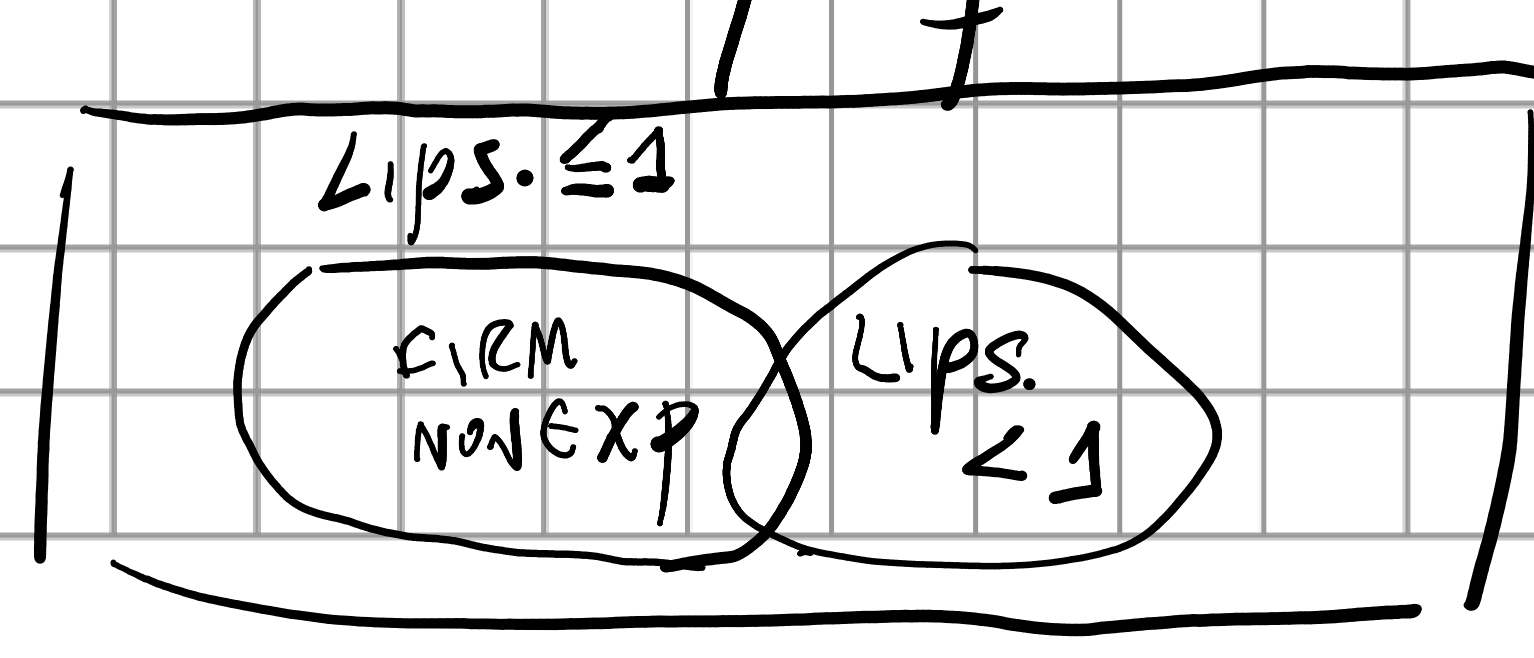
$$\boxed{\mathrm{prox}_f(z) + \mathrm{prox}_{f^*}(z) = z}$$

THEOREM $\forall z, z' \in \mathbb{R}^n$

$$\|\mathrm{prox}_f(z) - \mathrm{prox}_f(z')\|^2 \le (z - z')^T (\mathrm{prox}_f(z) - \mathrm{prox}_f(z'))$$

THAT IS $\mathrm{prox}_f(\cdot)$ IS FIRMLY NON-EXPANSIVE

(LIPS. CONS. ≤ 1 BUT MAYBE 1)



Convex Optimisation problems:

$$\min_{x \in \mathbb{R}^n} \overbrace{f(x)}^{\text{objective}} \quad \text{subject to} \quad \underbrace{\begin{aligned} g_i(x) &= 0 \\ h_j(x) &\le 0 \end{aligned}}_{\substack{\text{CONSTRAINTS} \\ x \in K}}$$

f, g, h CONVEX

EQUIVALENT TO $\min_{x \in K} f_0(x)$

or

$$\min_{x \in \mathbb{R}^n} f_0(x) + \mathbb{1}_{x \in K}$$

CONVEX OPTIMIZATION problems $\rightarrow$ DIFF./NON DIFF.

CONVEX OPTIMIZATION problems $\rightarrow$ CONSTRAINED/UNCONSTRAINED

UNCONSTRAINED MINIMIZATION

$$\arg\min_x f(x)$$

① GRADIENT DESCENT.

- ASSUMES THAT f IS DIFFERENTIABLE

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

η: STEP SIZE

IF f IS LIPSCHITZ

$$\|f(x) - f(y)\| \leq L \|x - y\|$$

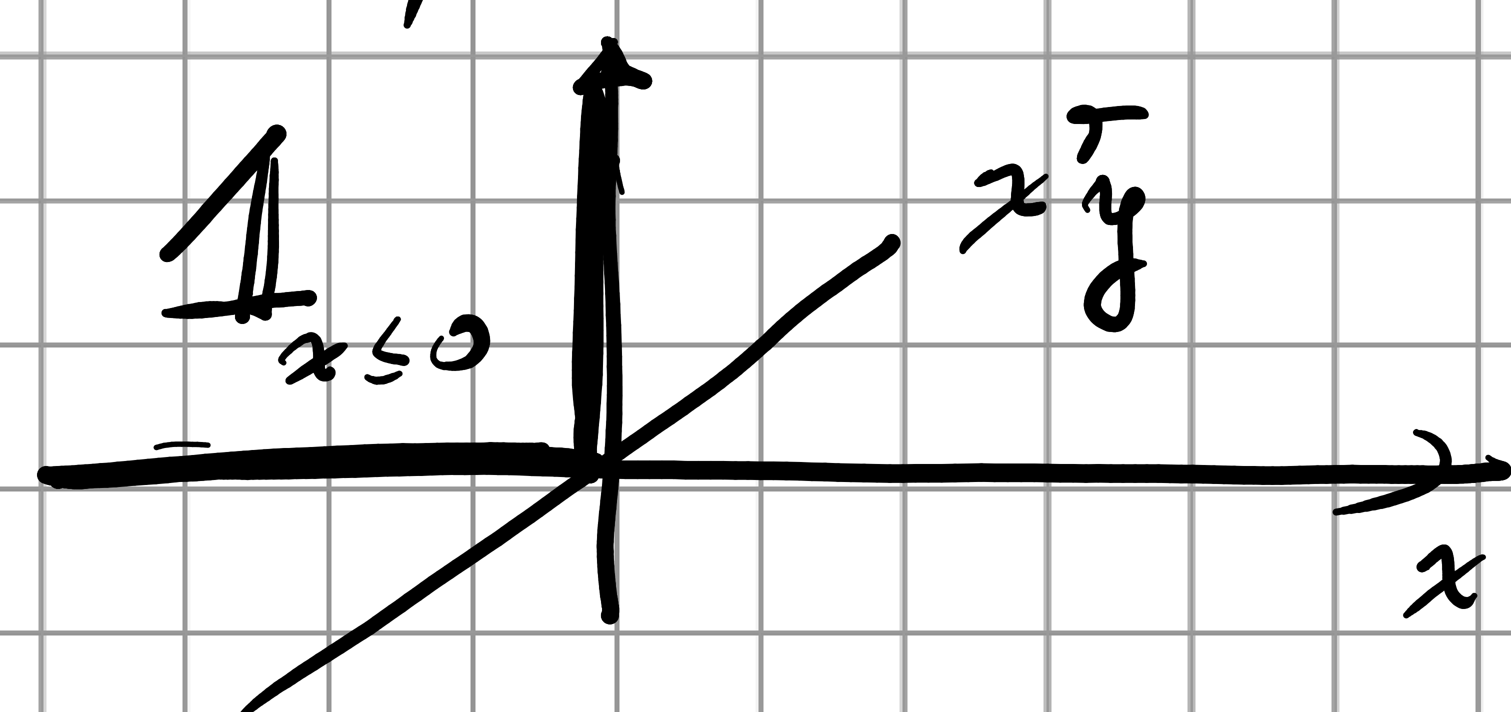
THEN $\eta < \frac{2}{L}$ GUARANTEES CONVERGENCE TO A GLOBAL MINIMUM.

WHAT HAPPENS IF $f(x)$ IS NOT DIFFERENTIABLE?

CONSTRAINED MINIMIZATION WITH LINEAR CONSTRAINTS

$$\operatorname{argmin}_x \underbrace{f_0(x) + \mathbb{1}_{Ax-b\leq 0}}_{f(x)}$$

IDEA RELAX INDICATOR FUNCTIONS
by LINEAR FUNCTIONS



$$y^T x \leq \mathbb{1}_{x\leq 0} \quad \text{for ALL } x$$
$$\forall\, y \geq 0$$

LAGRANGIAN

$$\mathcal{L}(x,y) = f_0(x) + y^T(Ax-b) \leq f(x)$$

HOLDS $\forall x \in \mathbb{R}^n$

1) LOWER BOUND

$$H(y) = \inf_x f_0(x) + y^T(Ax-b)$$

$$= -\sup_x y^T(b-Ax) - f_0(x)$$

$$= -f^*(A^T y) - y^T b$$

CONVEX CONJUGATE

WE HAVE

$$H(y) \leq f(x) \qquad \forall x \in \mathbb{R}^n, \; y \geq 0$$

$\Rightarrow$ pick BEST LOWER BOUND

$$p^* = \sup_y H(y)$$

WEAK DUALITY: $p^* \leq \min_x f(x)$

STRONG DUALITY: $p^* = \min_x f(x)$

<u>DUAL</u> opt<u>imization</u> pro<u>blem</u>

$$\hat{y} = \arg\max_{y \in \mathbb{R}^m} -f^*(A^T y) - b^T y$$

THEN $\hat{x} = \arg\min_x L(x, \hat{y})$

In particular we consider:

$$\arg\min_{x \in \mathbb{R}^n} \underbrace{f(x)}_{\substack{\text{DIFFERENTIABLE} \\ \text{CONVEX} \\ \text{LIPSCHITZ CONSTANT } L}} + \overbrace{g(x)}^{\substack{\text{NON-DIFFERENTIABLE} \\ \text{CONVEX}}}$$

② PROXIMAL GRADIENT / FORWARD BACKWARD

$$\boxed{x^{k+1} = \operatorname{prox}_{\gamma g}\left(x^k - \gamma \nabla f(x^k)\right)}$$

- FIRST WE REDUCE f WITH GRADIENT DESCENT

$$u^k = x^k - \gamma \nabla f(x^k)$$

- THEN WE REDUCE $g(x)$ WITH PROXIMAL GRAD. DESC.

$$x^{k+1} = \operatorname{prox}_{\lambda g}(u^k)$$

IF $\gamma < \frac{2}{L}$, THE ALGORITHM CONVERGES TO A GLOBAL MINIMIZER of $f(x) + g(x)$

③ DOUGLAS - RACHFORD SPLITTING

(ALTERNATING DIRECTIONS METHOD OF MULTIPLIERS)

ADMM

$$\underset{x,s}{\arg\min}\ f(x) + g(s)$$

$$\text{s.t. } x = s$$

(CONSTRAINED MINIMIZATION)

AUGMENTED LAGRANGIAN

$$L_\rho(x,s,y) = f(x) + g(s) + \underbrace{y^T(x-s)}_{\text{DUAL}} + \underbrace{\frac{\rho}{2}\|x-s\|^2}_{\text{AUGMENTATION}}$$

$$\begin{cases} x^{k+1} = \underset{x}{\arg\min}\ f(x) + y^T(x-s) + \frac{\rho}{2}\|x-s\|^2 \\ s^{k+1} = \underset{s}{\arg\min}\ g(s) + y^T(x-s) + \frac{\rho}{2}\|x-s\|^2 \\ y^{k+1} = y^k + \rho(x^{k+1} - s^{k+1}) \end{cases}$$

$\rightarrow$ DUAL ASCENT

USING THE FACT THAT

$$y^T(x-s) + \frac{\rho}{2}\|x-s\|^2 = x^T y - s^T y + \frac{\rho}{2} x^T x + \frac{\rho}{2} s^T s - \rho x^T s$$

$$\underset{x}{\propto} \frac{\rho}{2}\left\| x - \left(s - \frac{y}{\rho}\right)\right\|^2$$

$$\underset{s}{\propto} \frac{\rho}{2}\left\| s - \left(x - \frac{y}{\rho}\right)\right\|^2$$

$$\begin{cases} x^{k+1} = \operatorname{prox}_{\rho f}\left(s^k - \frac{y^k}{\rho}\right) \\ s^{k+1} = \operatorname{prox}_{\rho g}\left(x^{k+1} - \frac{y^k}{\rho}\right) \\ y^{k+1} = y^k + \rho\left(x^{k+1} - s^{k+1}\right) \end{cases}$$