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.)

$$y = \underset{\text{linear degradation process}}{A} \underset{\text{image we want to recover}}{x}$$

...

We can recover a sparse image by solving:

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

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

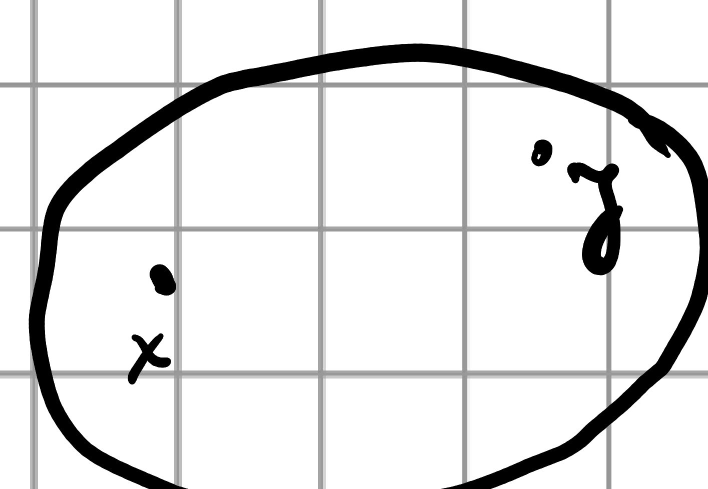
REFERENCES:

- "MATHEMATICAL INTRODUCTION TO COMPRESSED SENSING" by FOUCART AND RAUHUT (APPENDIX B)
- "CONVEX OPTIMIZATION" by Boyd and Vandenberghe (CHAPTERS 2, 3, 4, 5, 9)
- "PROXIMAL ALGORITHMS" by Parikh et al. (CHAPTERS 2, 3, 4)

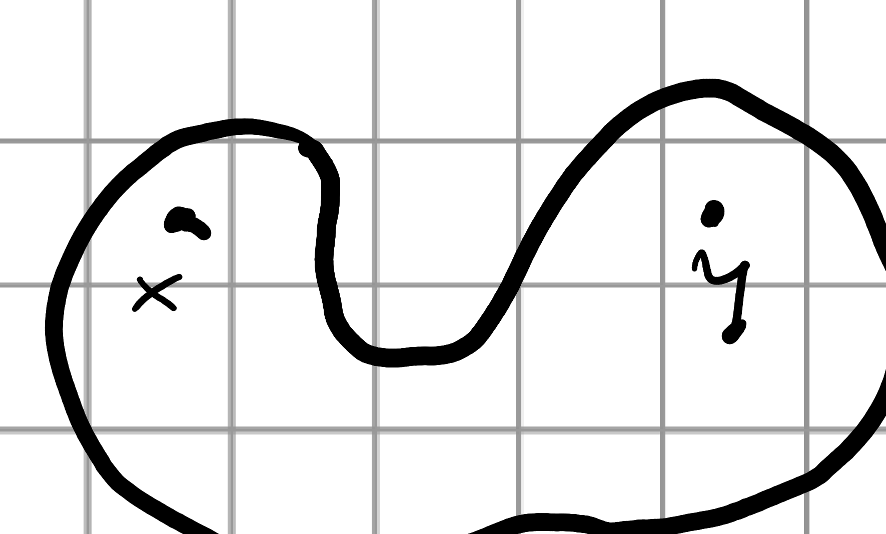
CONVEX SETS

Def CONVEX SET IS A SUBSET $K \subseteq \mathbb{R}^n$

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



CONVEX

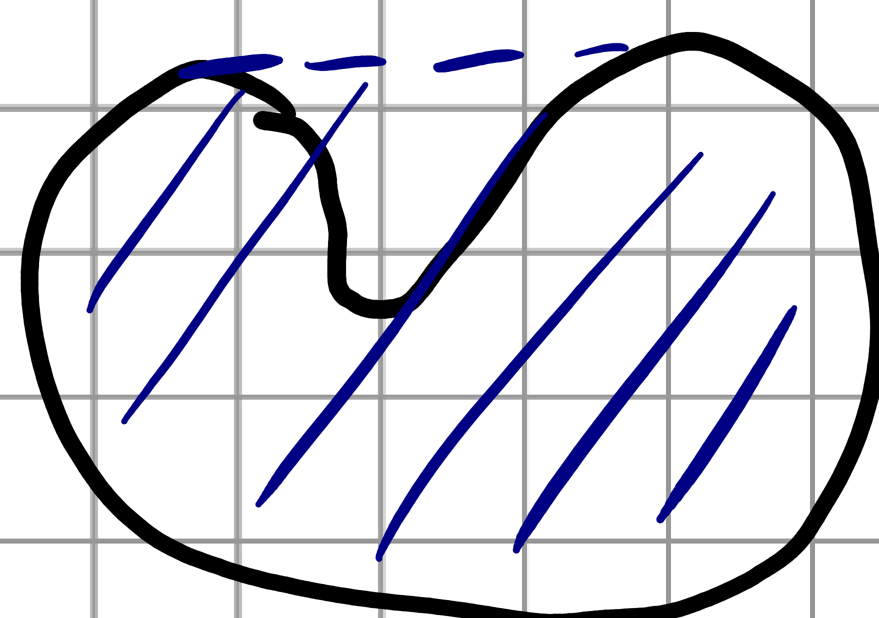


NON-CONVEX

- THE INTERSECTION OF CONVEX SETS IS CONVEX!
- THE UNION OF CONVEX SETS IS NOT NECESSARILY CONVEX.

Def CONVEX HULL OF A SUBSET $K \subset \mathbb{R}^n$

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



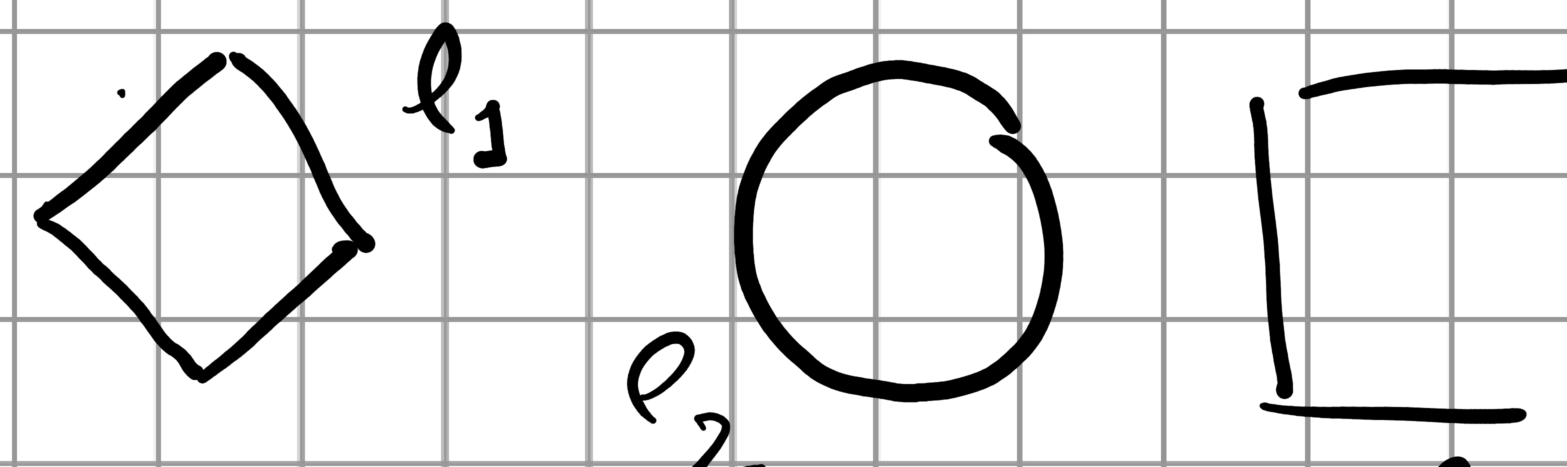
K

conv(K)

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

- IT IS THE SMALLEST SET CONTAINING K
- K IS CONVEX $\Leftrightarrow$ $\operatorname{conv}(K) = K$

- EXAMPLES OF CONVEX SETS.
  - p-BALLS $B_r^p = \{ x : \|x\|_p \leq r \}$



ℓ_1 ℓ_2 ℓ_∞

  - AFFINE SUBSPACES $\{ x : Ax = y \}$
  - HALF-SPACES (ω, λ) $\{ x : x^T \omega \leq \lambda \}$

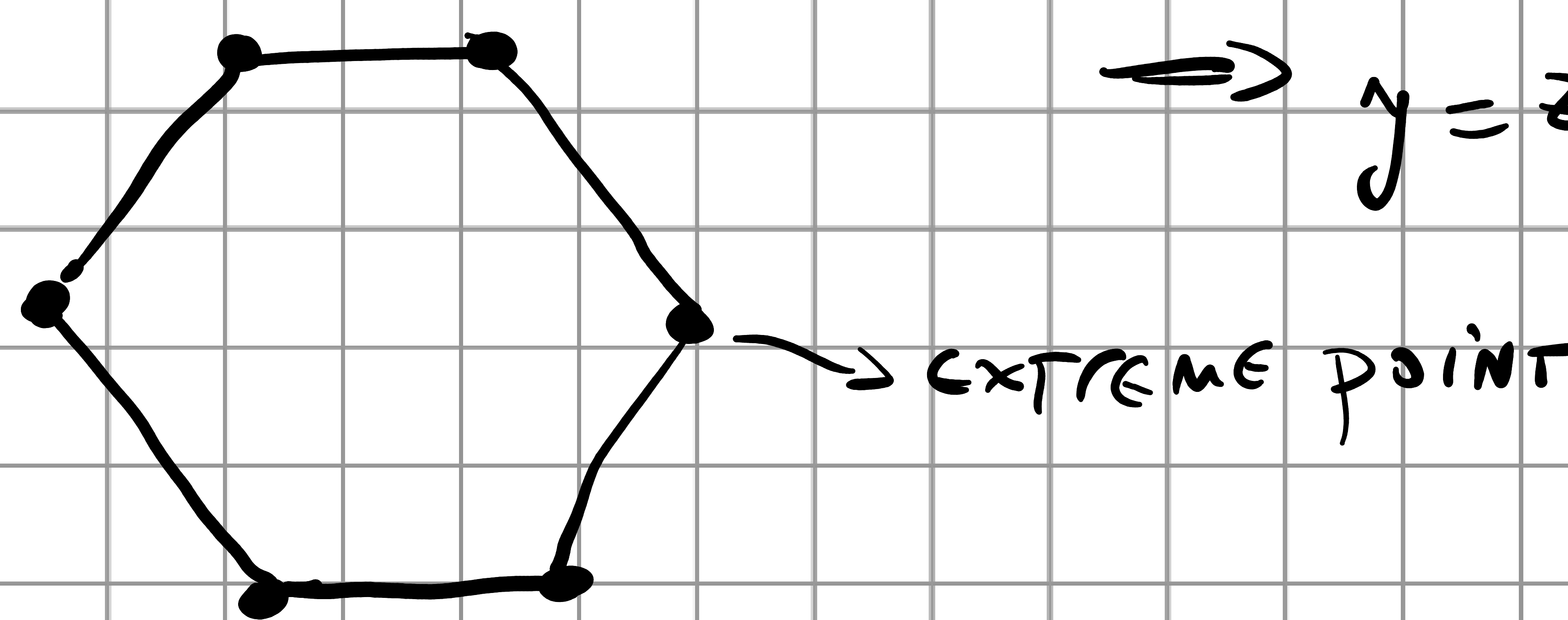
Def: Extreme points

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

$y, z \in K$
$t \in [0,1]$

$\Longrightarrow y = z = x$

France is a convex set!

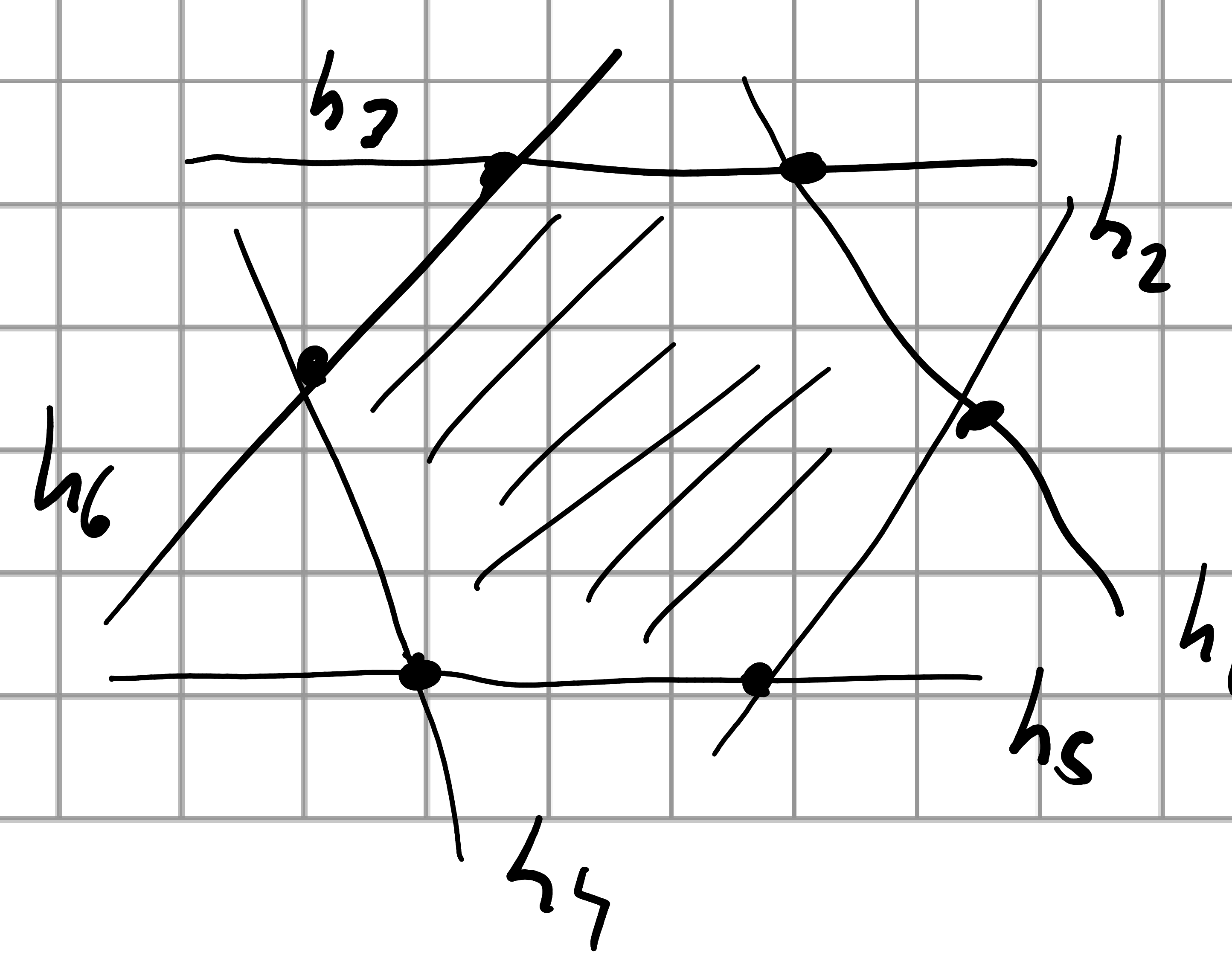


Theorem: A convex set is the convex hull of its extreme points.

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

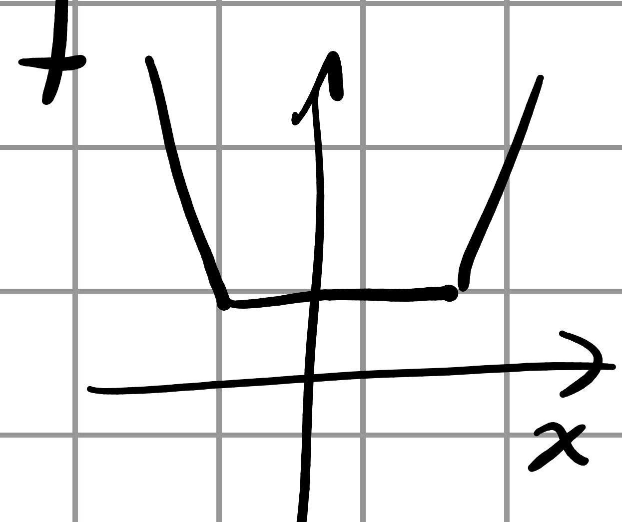
$\lambda \infty = \infty, \; \lambda > 0$

DOMAIN :

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

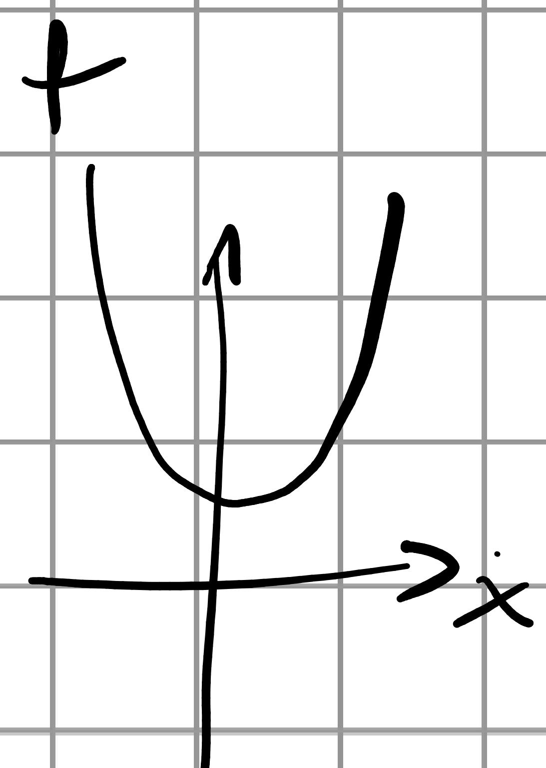
– A FUNCTION IS **proper** IF $\mathrm{dom}(f)$ IS NOT EMPTY.

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

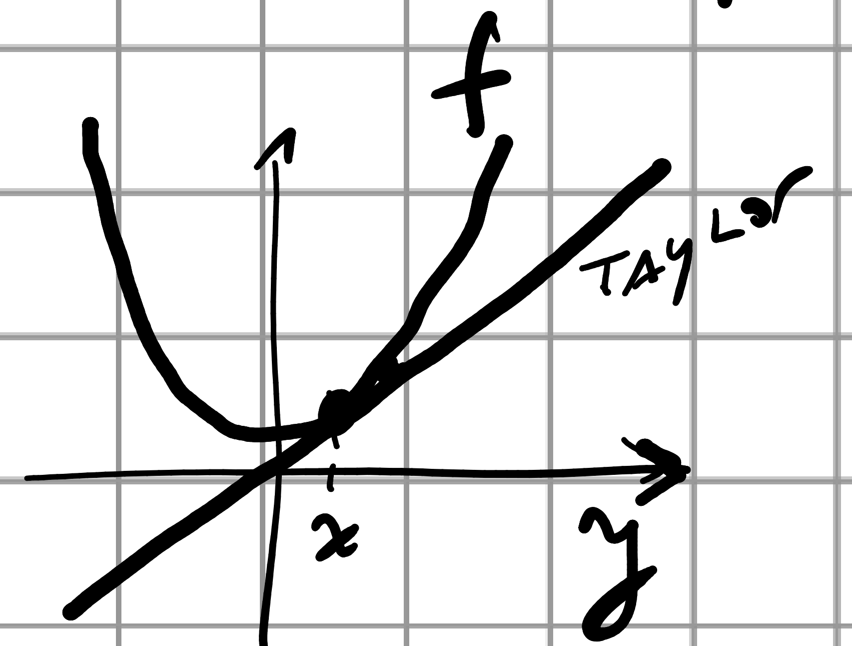
and is strongly convex if A is positive definite with $\gamma = \lambda_{min}(A)$

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



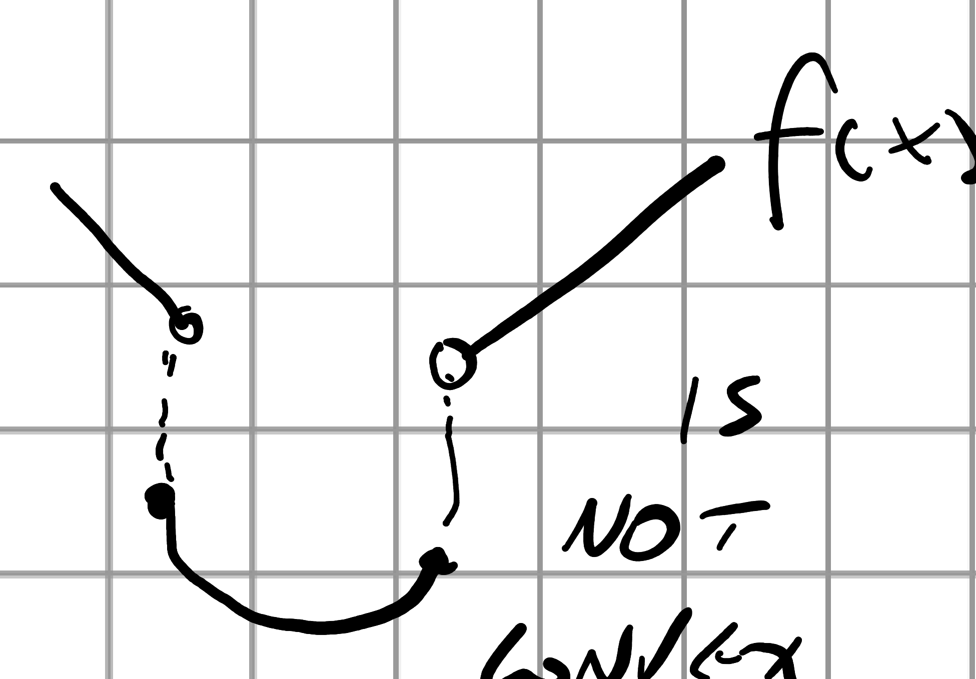
$$f(y) \geq f(x) + \nabla f^T (y - x)$$

- f IS STRONGLY CONVEX IIF $\forall x, y \in \mathbb{R}^n$

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

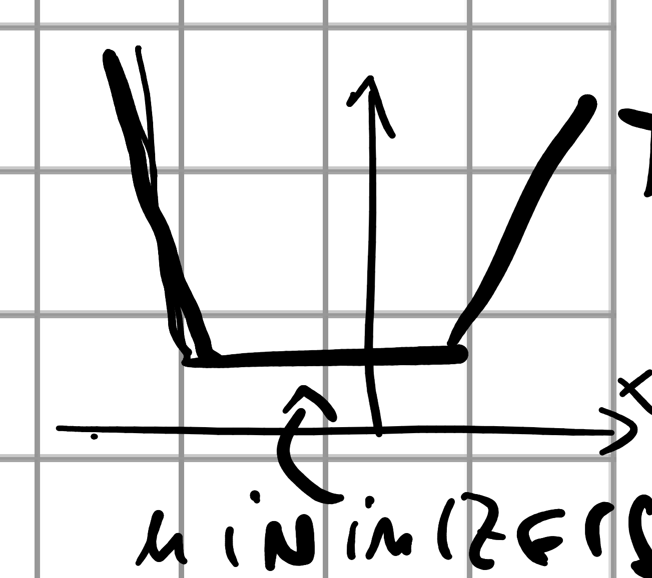
- If f is TWICE DIFFERENTIABLE THEN IT IS CONVEX IIF $\forall x \in \mathbb{R}^n$ $\nabla^2 f(x) \succcurlyeq 0$ (MATRIX IS PSD)
  - IT IS STRICTLY CONVEX IF $\nabla^2 f(x) > 0$

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

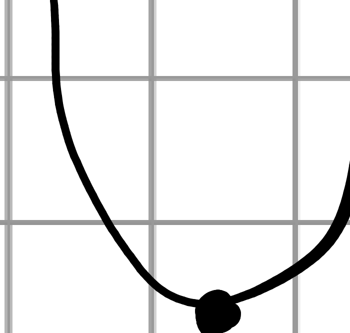


!! **Prop.** $f : \mathbb{R}^n \to]-\infty, +\infty]$ be a convex function

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



- If f is strictly convex there is a unique minimizer

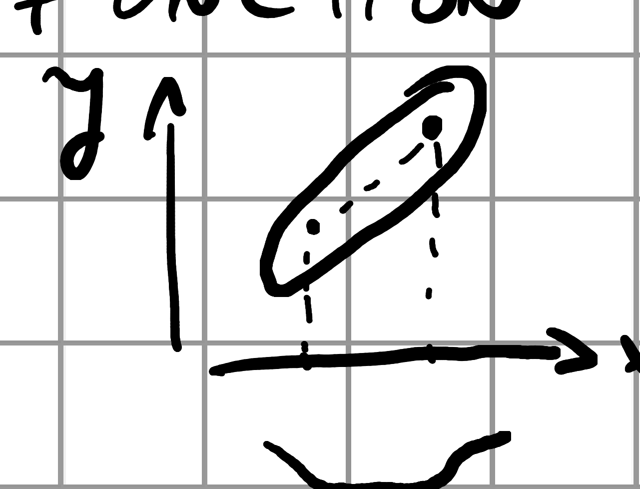


Theorem

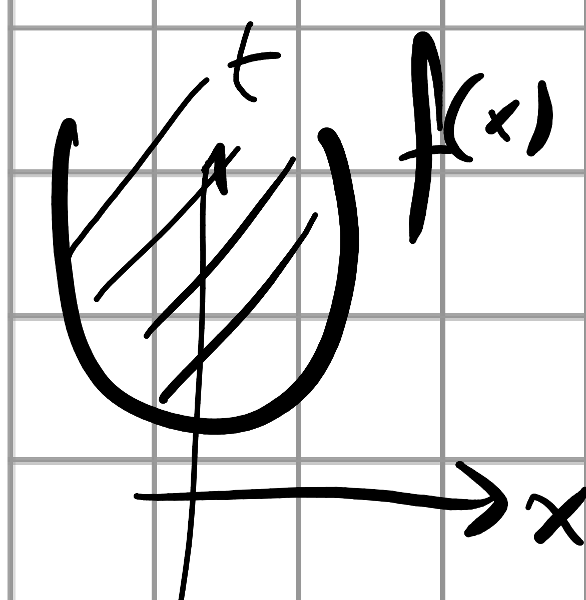
Let $f : \mathbb{R}^n \times \mathbb{R}^m \to]-\infty, +\infty]$ be a convex function.

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

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



Epigraph:

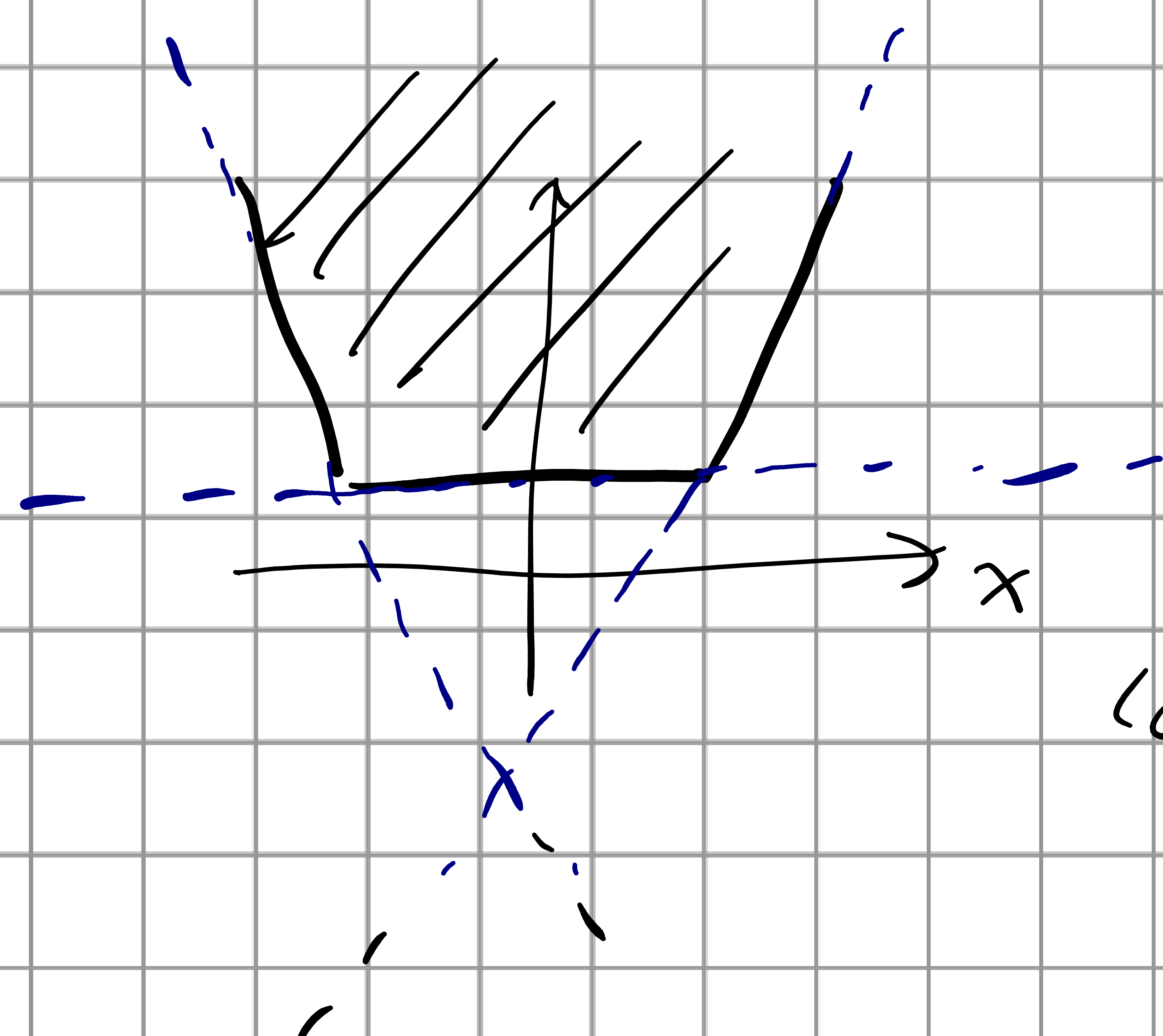


$$\mathrm{epi}(f) = \{ (t, x) \in \mathbb{R}^{n+1}, \ f(x) \leq t \}$$

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

CONVEX CONJUGATE

DUALITY FOR FUNCTIONS:



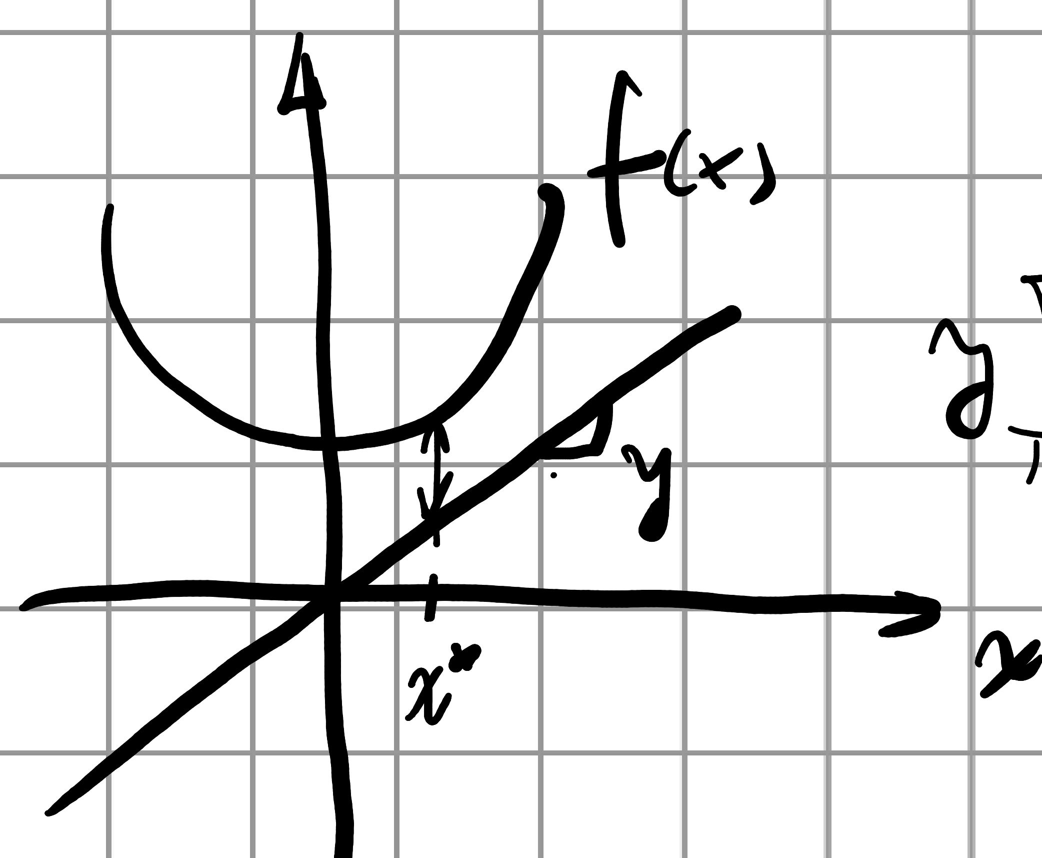
IDEA: epi(f) CAN BE EXPRESSED AS INTERSECTION OF HALFSPACES

$\Updownarrow$

CONVEX f CAN BE EXPRESSED AS SUPREMUM OVER LINEAR FUNCTIONS

$$f(x) = \sup_y \; y^T x - \underbrace{f^*(y)}_{}$$

HOW SHOULD WE CHOOSE IT?



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

$$\left(y^T x - b_y \leq f(x) \right)$$

DEFINITION: FENCHEL-DUAL (CONVEX CONJUGATE)

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

CAN BE APPLIED TO ANY $f(x)$ – CONVEX OR NOT!

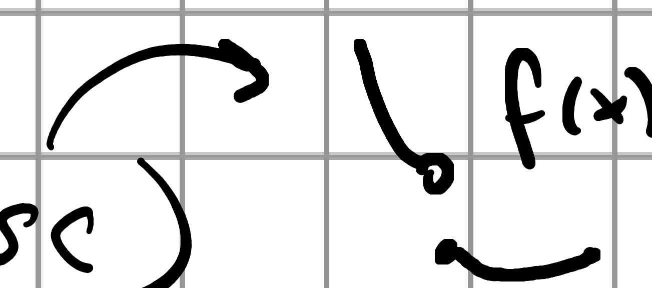
$f^*(y)$ IS CONVEX INDEPENDENTLY OF WHETHER f IS CONVEX OR NOT

Fenchel - Young $\quad \forall x, y \in \mathbb{R}^n$

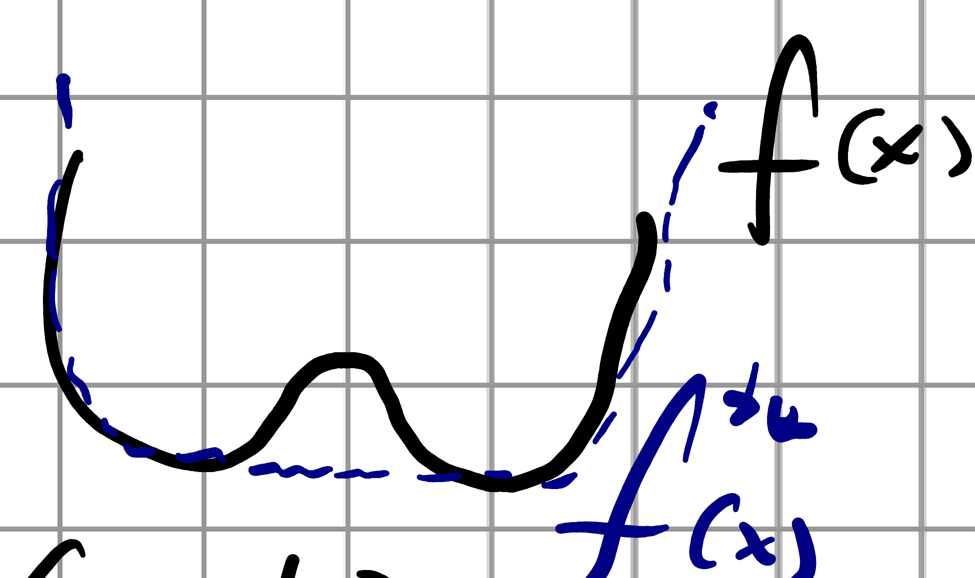
$$x^T y \leq f(x) + f^*(y)$$

Properties of convex conjugate: $\quad f : \mathbb{R}^n \to]-\infty, +\infty]$

1) f^* is lower semi continuous (LSC)



2) f^{**} is the largest LSC convex function satisfying $f^{**}(x) \leq f(x) \quad \forall x \in \mathbb{R}$



3) $\mathrm{epi}(f^{**}) = \mathrm{conv}(\mathrm{epi}(f))$

4) If f is convex $\quad f^{**} = f$!

5) For $\lambda \neq 0 \quad [f(\lambda x)]^* = f^*\left(\frac{y}{\lambda}\right)$

6) For $\lambda > 0, \quad (\lambda f(x))^* = \lambda f^*\left(\frac{y}{\lambda}\right)$

7) For $z \in \mathbb{R}^n \quad (f(x-z))^* = f^*(y) + z^T y$

Example:

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

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

$x \in \mathbb{R}$

e.g. $f(x) = |x|$

$\ell_2 \quad p=2 \Rightarrow q=2 \quad \ell_2$

$\ell_1 \quad p=1 \Rightarrow q=\infty \quad \ell_\infty$

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

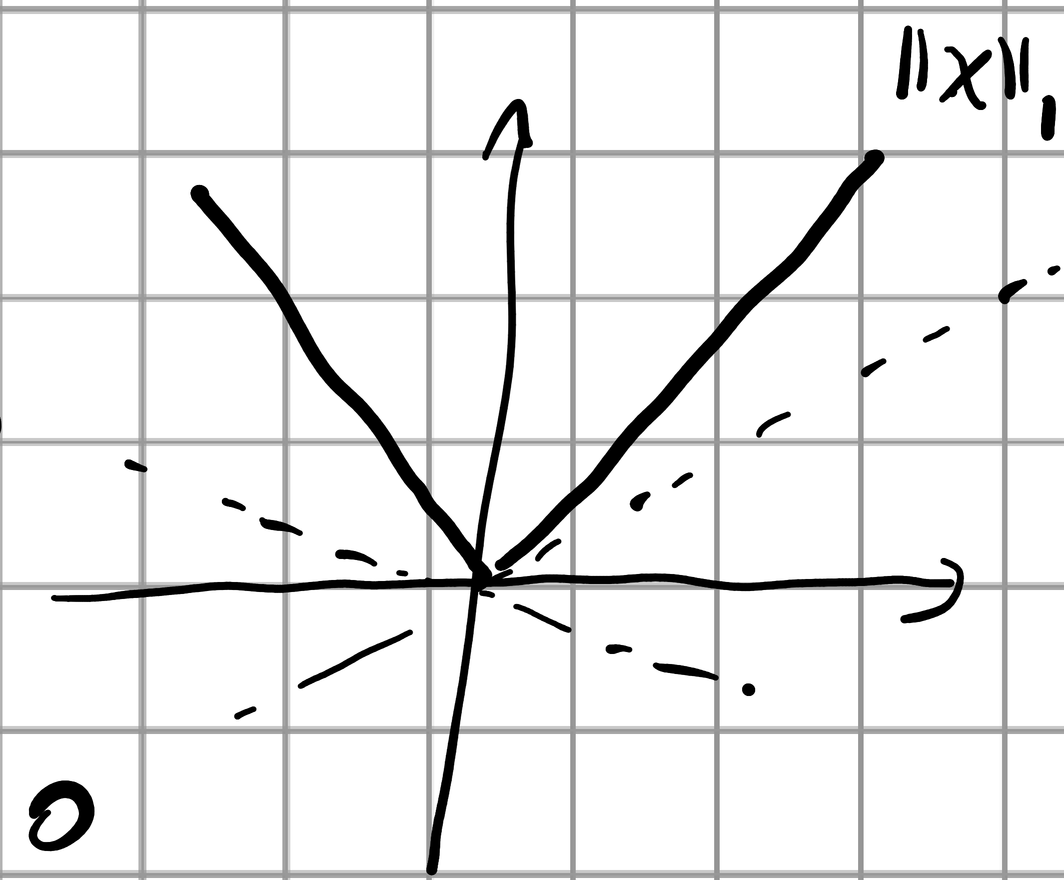
$$\partial f(x) = \{ v \in \mathbb{R}^n, \ f(y) \geq f(x) + v^T(y-x) \ \forall y \in \mathbb{R}^n \}$$

- 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 iif $0 \in \partial f(x)$

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!

$$(a) \Longleftrightarrow (b) \Longleftrightarrow (c)$$

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

Proximal Mapping: $f : \mathbb{R}^n \to]-\infty, \infty]$

CONVEX

$$\operatorname{prox}_f(x) = \arg\min_{x \in \mathbb{R}^n} f(x) + \frac{1}{2}\|x - z\|^2$$

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

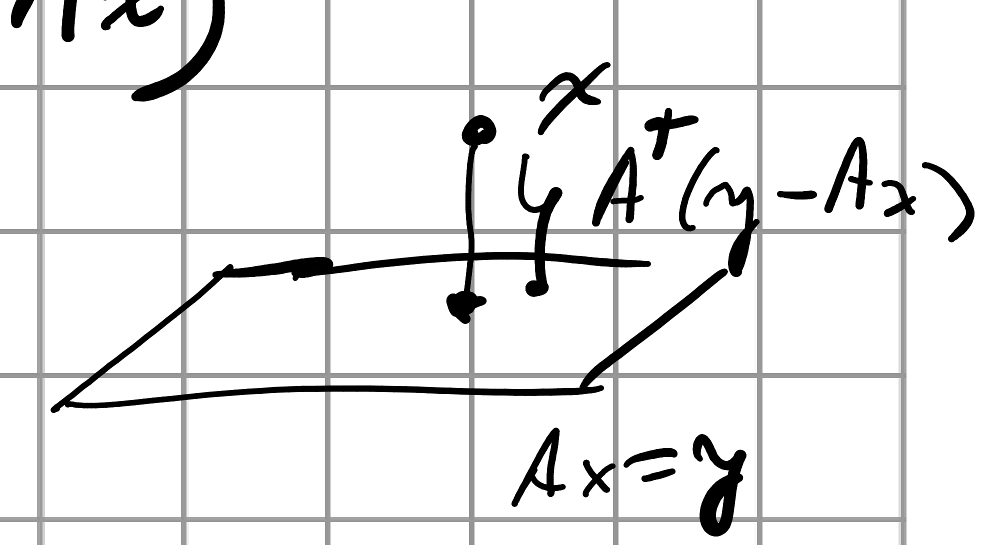
Examples:

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

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

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

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



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

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

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

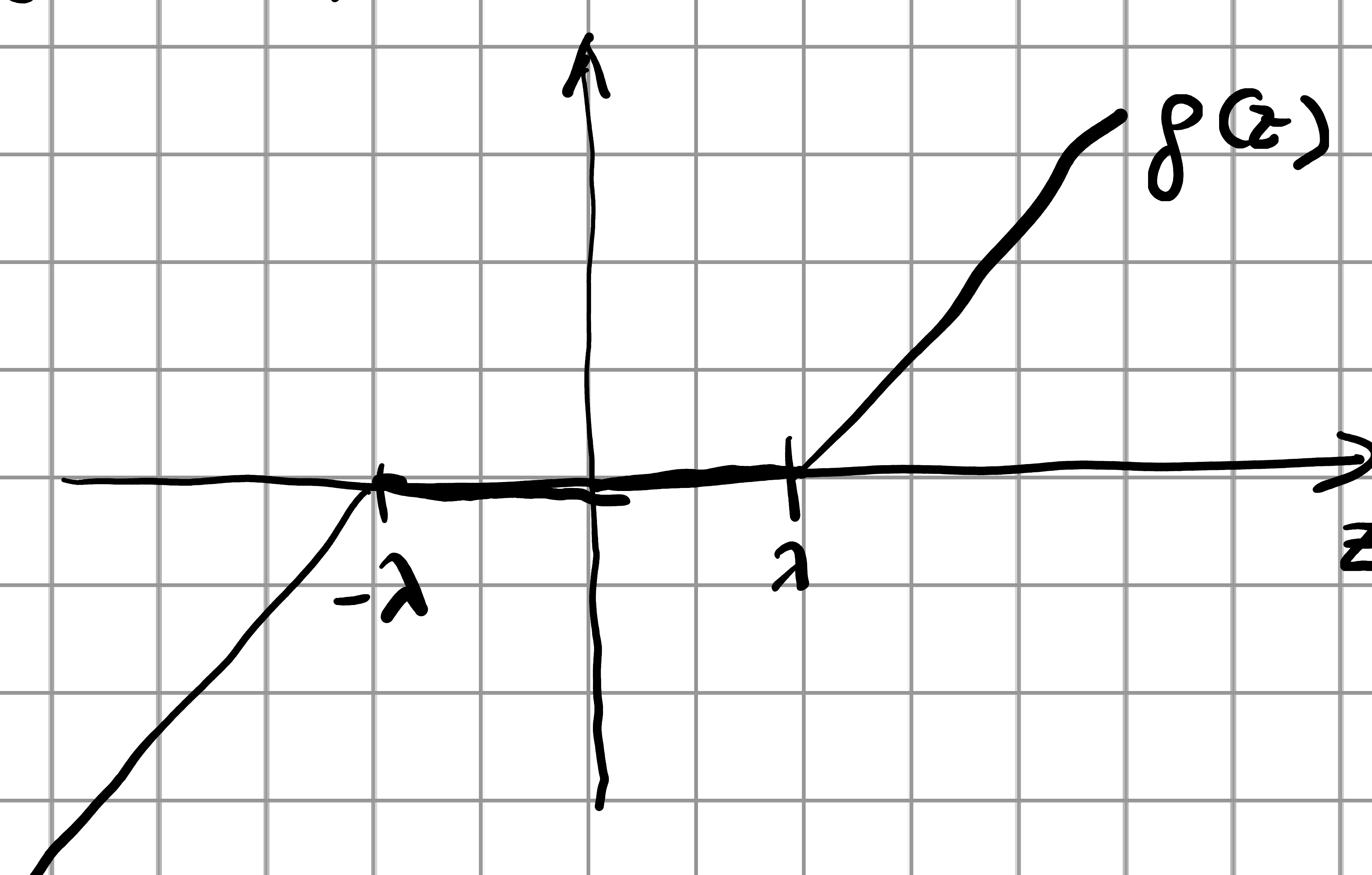
$$x = (I + \lambda A^T A)^{-1}(z + \lambda y)$$

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

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

$$g_i(z) = \begin{cases} z_i - \lambda & \text{if } z_i \geq \lambda \\ z_i + \lambda & \text{if } z_i \leq -\lambda \\ 0 & \text{if } |z_i| \leq \lambda \end{cases}$$

SOFT THRESHOLDING



SEPARABLE FUNCTION:

$$\lambda \|x\|_1 + \frac{1}{2}\|x - z\|^2 = \sum_{i=1}^{m} \lambda |x_i| - \frac{1}{2}(x_i - z_i)^2$$

$$= \sum_{i=1}^{m} f_i(x_i)$$

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

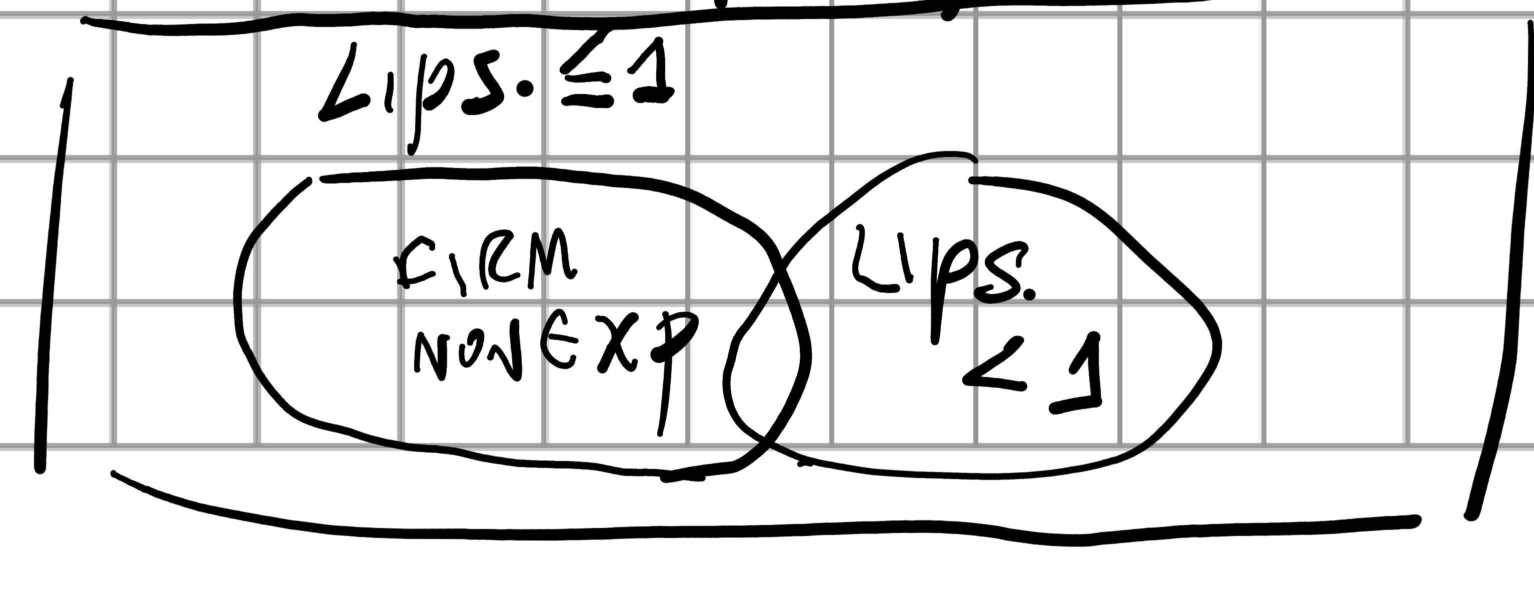
$$\boxed{\mathrm{prox}_f(z) + \mathrm{prox}_{f^*}(z) = z}$$

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

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

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

(LIPS. CONS. ≤ 1 BUT MAYBE 1)



Convex Optimisation problems:

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

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.

$\rightarrow$ 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?

PROXIMAL GRADIENT DESCENT

$$x^{k+1} = \text{prox}_{\eta f}(x^k)$$

PROPOSITION: If f is convex, this algorithm converges to a minimum $\forall \eta > 0$.

EXAMPLES: ① $f(x) = \lambda \|x\|_1$

$$x^{k+1} = \text{softthres}(x^k, \lambda\eta)$$

converges in $\dfrac{\|x^0\|_\infty}{\lambda\eta}$ iterations

② $f(x) = \frac{1}{2}\|Ax - b\|^2$

PGD

$$x^{k+1} = (\eta A^T A + I)^{-1} x^k + c$$

where $c = \left(A^T A + \frac{I}{\eta}\right)^{-1} \frac{x^k}{\eta}$

converges $\forall \eta > 0$!

GD

$$x^{k+1} = (I - A^T A \eta) x^k + c$$

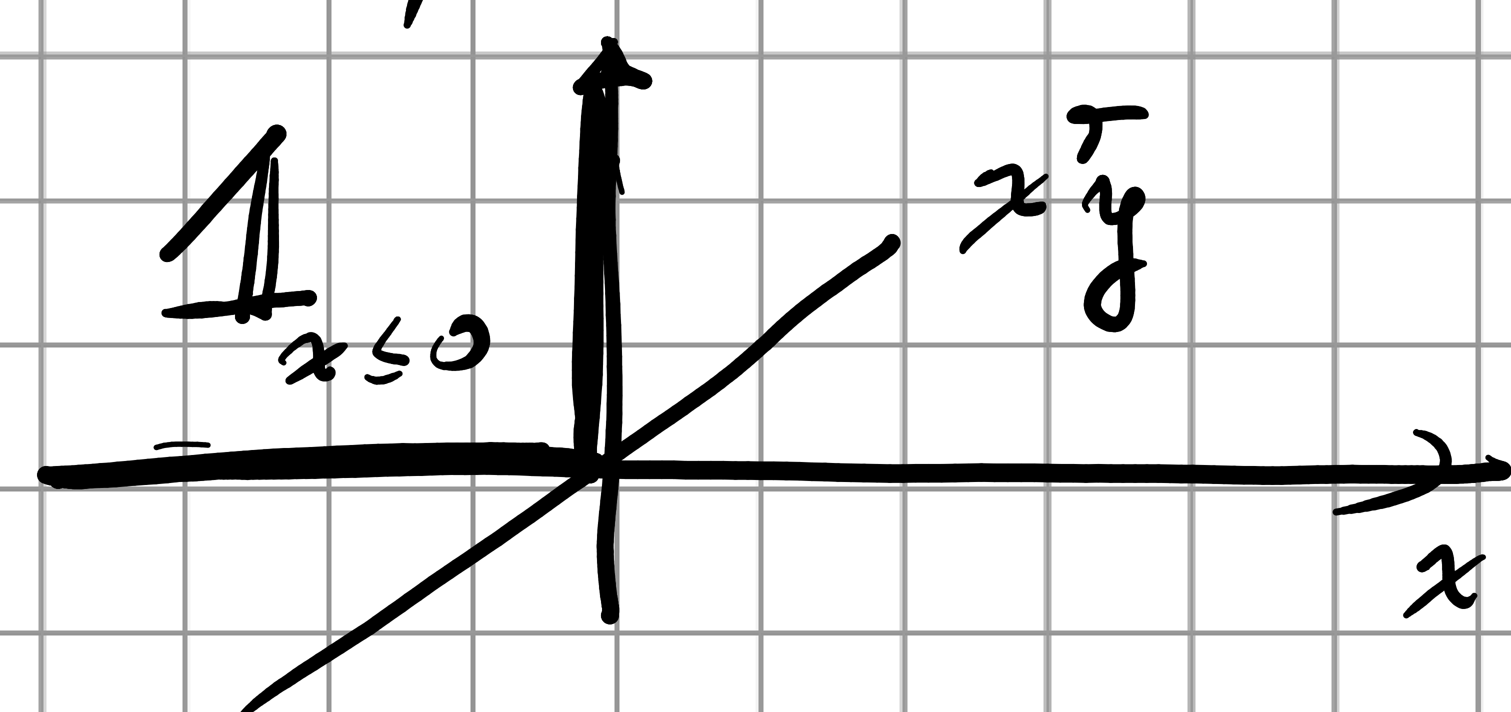
where $c = \eta A^T b$

converges for $\eta < \dfrac{2}{\lambda_{max}(A^T A)}$

CONSTRAINED MINIMIZATION WITH LINEAR CONSTRAINTS

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

IDEA RELAX INDICATOR FUNCTIONS
by LINEAR FUNCTIONS



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

$\boxed{\text{LAGRANGIAN}}$

$$L(x,y) = f_0(x) + y^T(Ax-b) \le f(x)$$

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

1) LOWER BOUND

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

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

$$= -\underbrace{f^*}_{\text{CONVEX CONJUGATE}}(A^T y) - y^T b$$

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

if there exist feasible point x

<u>DUAL</u> opt<u>imization</u> prob<u>lem</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

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

- 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.

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

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

$$\operatorname*{argmin}_{x,s} \; f(x) + g(s)$$

s.t. $x = s$

AUGMENTED LAGRANGIAN (CONSTRAINED MINIMIZATION)

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

$$\begin{cases} x^{k+1} = \operatorname*{argmin}_{x} \; f(x) + y^T(x-s) + \frac{\rho}{2}\|x-s\|^2 \\ s^{k+1} = \operatorname*{argmin}_{s} \; g(s) + y^T(x-s) + \frac{\rho}{2}\|x-s\|^2 \\ y^{k+1} = y^k + \rho\,(x^{k+1} - s^{k+1}) \rightarrow \text{DUAL ASCENT} \end{cases}$$

USING THE FACT THAT

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

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

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

ADMM STANDARD FORM

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

$\Updownarrow$ EQUIVALENT ALGORITHMS

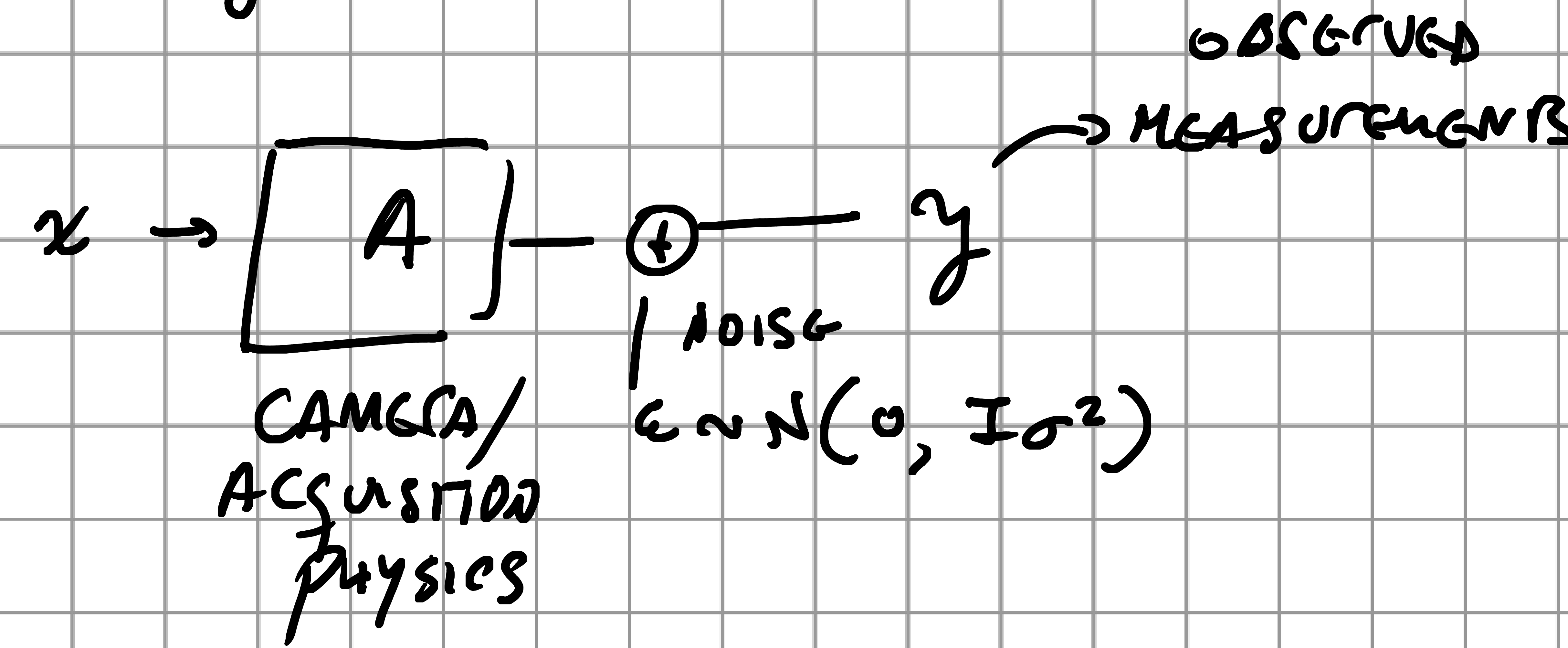
DOUGLAS RACHFORD ($\rho = 1$)

$$\begin{cases} x^{k+1} = \text{prox}_f(z^k) \\ z^{k+1} = \text{prox}_g(2x^{k+1} - z^k) + (z^k - x^k) \end{cases}$$

Inverse problems

$x \in \mathbb{R}^n$ $\quad y \in \mathbb{R}^m$

$$y = Ax + \varepsilon \qquad A \in \mathbb{R}^{m \times n}$$



- THERE ARE INFINITELY MANY SOLUTIONS
  if A is LOW RANK, $\operatorname{rank} A < n$ (EVEN IF $\varepsilon = 0$)

  ↳ WE REQUIRE REGULARIZATION

$$\underset{x}{\operatorname{argmin}} \; \underbrace{\overset{\text{DATA FIDELITY}}{\|Ax - y\|^2}}_{\substack{\text{CONVEX} \\ \text{BUT NOT} \\ \text{STRICTLY} \\ \text{CONVEX.}}} + \underbrace{\overset{\text{REGULARIZATION}}{\lambda \|x\|_1}}_{\substack{\text{PROMOTES} \\ \text{SPARSE} \\ \text{SOLUTIONS}}}$$