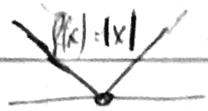
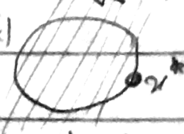
- **Motivation :** $\min_x f(x)$ $\quad f(x) \begin{cases} \text{cvx.} \\ \text{non-smooth} \end{cases}$

<u>ex</u> : $f(x) = |x|$



; constraint optim $\min_{x \in C} f(x) = f + \underbrace{\iota_C(x)}_{\sim}$



ML-Type Pbm

<u>SVM</u> : $\min_x \frac{1}{m} \sum_i \ell(\langle a_i, x \rangle, y_i)$ $\quad \ell =$

ℓ = non-smooth $\quad \ell(u_i \cdot y_i)$.

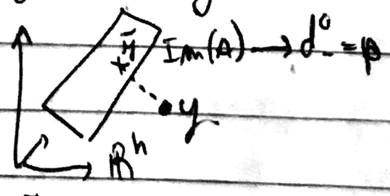
<u>Lasso</u> : $\min_x \frac{1}{2n} \underbrace{\sum_i \|\underbrace{\langle a_i, x\rangle}_{L(Ax)} - y_i\|^2}_{\|Ax - y\|^2} + \lambda \underbrace{\|x\|_1}_{\text{regul non-smooth}} = f(x)$.

- **<u>A detour</u> : least square & ridge**

$$\min_x \frac{1}{2n} \sum_{i=1}^{n} |\langle a_i, x \rangle - y_i|^2 = \frac{1}{2n} \|Ax - y\|^2$$

- if $A = \square \updownarrow n$, $p = n$: invertible ie $\ker(A) = 0$ $\iff \mathrm{Im}(A) = \mathbb{R}^n$ $\to \exists!$ sol° $x^* = A^{-1}y$.

- if $A = \updownarrow n \geq p$ (width p): injective ~~surjective~~ $\ker(A) = 0$ in general $y = Ax$ no sol°



$\mathrm{Im}(A) \to d° = p$

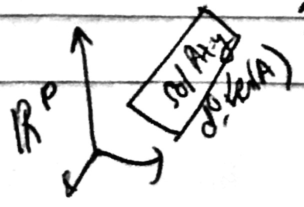
sol° $\min f(x) = \frac{1}{2} |y - Ax|^2$ $\quad \nabla f(x) = A^T(Ax - y) = 0$

$\iff x^* = \underbrace{(A^T A)}_{\in \mathbb{R}^{p \times p}}{}^{-1} A^T y$ invertible.

indeed $A^T A z = 0 \Rightarrow \langle A^T A z, z \rangle = 0 \Rightarrow \|Az\| = 0 \Rightarrow Az = 0 \Rightarrow z = 0$

- if $A = \square \updownarrow n \geq p$ (width p) surjective ie $\mathrm{Im}(A) = \mathbb{R}^n$
  not injective ie $\ker(A) \neq 0$ $\Rightarrow$ infinite de sol°

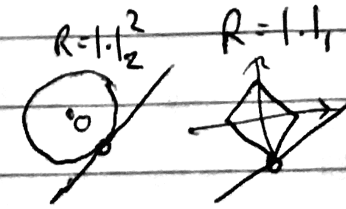
$\{x : Ax = y\}$ espace // $\ker(A)$



Need Regularizat°: $\min_x \frac{1}{2}|y-Ax|^2 + \lambda R(x)$

$\downarrow \lambda \to 0.$

$\min_{y=Ax} R(x).$



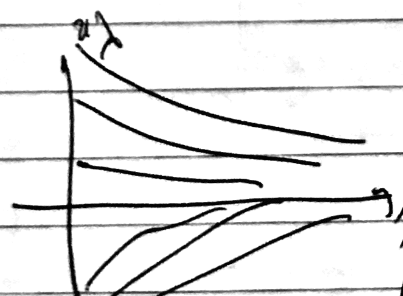
- **Ridge regression:** $\min_x \frac{1}{2}|y-Ax|^2 + \frac{\lambda}{2}|x|^2 = f(x).$

sol° $\nabla f(x) = A^*(Ax-y) + \lambda x = 0$

i.e. $x = \underbrace{(\underbrace{A^*A}_{p\times p} + \lambda Id_p)}_{\text{always invertible}}{}^{-1}\underbrace{A^*y}_{\mathbb{R}^p}$

Lemma: [woodbury formula] $(\underbrace{A^*A}_{p\times p} + \lambda Id_p)^{-1}A^* = A^*(\underbrace{AA^*}_{n\times n} + \lambda Id_n)^{-1}$

good idea if feature space very large (eg RKHS).

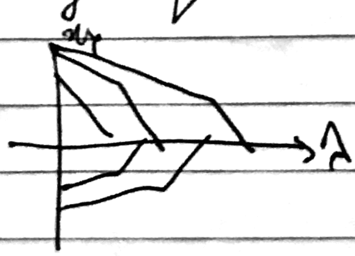


path non sparse

→ Cross validate λ!

- **Lasso for feature select°:** $\min_x \frac{1}{2}|y-Ax|^2 + \lambda|x|_1$

non smooth!!



Prox & FB

3

$\min_x f(x)$

<u>Grad Desc</u>: $x_{k+1} = x_k - \tau \nabla f(x_k) = (Id - \tau \nabla f)(x_k)$

<u>Thm</u> if $\partial^2 f(x) \leq L\, Id$, $x_k \to x$ if $\tau < \frac{2}{L}$

<u>Pbm</u>: if f non smooth $Id - \tau \nabla f$ undefined

<u>Implicit</u>: $x_{k+1} = \mathrm{prox}_{\tau f}(x_k)$ (Euler implicit aka proximal point Algo) (Rockafellar)

$\boxed{Ex}$ Projecteur, ℓ_2^2, ℓ_1 (cf feuille 4)

<u>Def</u> $\mathrm{prox}_{\tau f}(x) \triangleq \arg\min_z \frac{1}{2}\|x - z\|^2 + \tau f(z)$

at optimal z^*, $z - x + \tau \nabla f(z) = 0$

Rmq: If f smooth, $\mathrm{prox}_{\tau f}(x) = (Id + \tau \nabla f)^{-1}(x)$

[examples feuilles suivante]

<u>Thm</u>: $\forall \tau > 0$, $x_k \to x^*$ sol^c

→ useless!

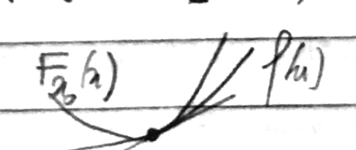
<u>Splitting</u>: $\min_x \underbrace{f(x)}_{\text{smooth}} + \underbrace{g(x)}_{\text{proximable}} = \mathcal{E}(x)$ $\boxed{ex}$

→ Constr: $\min_{x \in C} f(x)$

→ Lasso $\min \frac{1}{2}\|y - Ax\|^2 + \lambda \|x\|_1$

→ Dual SVM

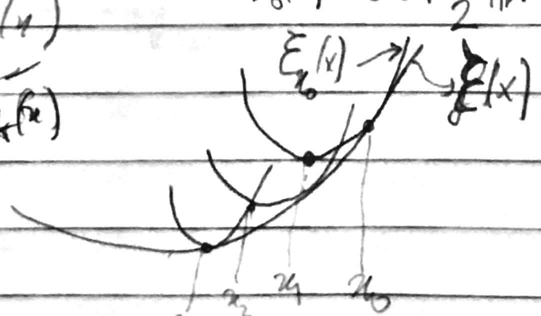
<u>Lem</u>: If $\partial^2 f(x) \leq \frac{L}{2} Id$, then $f(x) \leq \underbrace{f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{k}{2}\|x - x_0\|^2}_{F_{x_0}(x) = cste + \frac{k}{2}\|x - (x_0 - \underbrace{\frac{1}{k}}_{= \tau \leq \frac{1}{L}} \nabla f(x_0))\|^2}$ $k \geq L$



$\Rightarrow \mathcal{E}(x) \leq \underbrace{F_{x_k}(x) + g(x)}_{\triangleq \mathcal{E}_{x_k}(x)}$

Majorize-Minimize

$$\begin{cases} \mathcal{E}_{x_k}(x) \geq \mathcal{E}(x) \\ \mathcal{E}_{x_k}(x_k) = \mathcal{E}(x_k) \\ \mathcal{E}_{x_k}(x) - \mathcal{E}(x) \text{ smooth} \end{cases}$$



<u>Algo MM</u>: $x_{k+1} = \arg\min \mathcal{E}_{x_k}(x)$

Ici: $x_{k+1} \triangleq \arg\min_x \frac{k}{2}\|x - [x_k - \overset{\tau}{\frac{1}{k}} \nabla f(x_k)]\|^2 + g(x) = \mathrm{Prox}_{\tau g}(x_k - \tau \nabla f(x_k))$ [ex: grad projeté]

<u>Thm</u>: if $\tau < 2/L$, $x_k \to x^*$, $\mathcal{E}(x_k) - \mathcal{E}(x^*) \leq 1/k$ FISTA ↗

exemple ISTA $g(x) = \lambda\|x\|_1$

NESTEROV / FISTA

$$x_{k+1} = \mathrm{Prox}_{\tau g}\left[y_k - \tau_k \nabla f(y_k)\right]$$

$$y_{k+1} = x_{k+1} + \theta_k (x_{k+1} - x_k)$$

FISTA-like $\tau_k \triangleq \frac{1}{L}$ $\theta_k \to 1$ $\theta_k \triangleq \frac{k}{k+3}$

Thm: $\mathcal{E}(x_k) - \min \mathcal{E} = O(1/k^2)$ [Beck Teboulle]

$x_k \to x^*$ [Dossal / Chambolle]

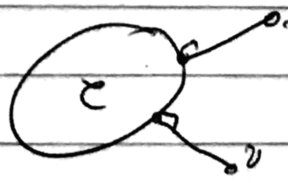
limit° : if $\partial^2 \mathcal{E}(x) \geq \mu \,\mathrm{Id}$ (strong convexity)

$\mathcal{E}(x_k) - \mathcal{E}(x^*) = O(\rho^k)$ [FB].

$\mathcal{E}(x_k) - \mathcal{E}(x^*) = O(1/k^\alpha)$ $\alpha = 3$ (?).

$\text{Prox}_{\tau f}(x) = \min_z \frac{1}{2}\|x-z\|^2 + \tau f(x)$

ex: $f(x) = \iota_C(x)$



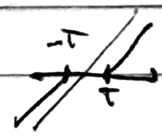
$\text{prox}_{\tau f} = \text{proj}_C$

ex: $f(x) = \|x\|^2 \rightsquigarrow z - x + \tau z = 0 \longrightarrow \text{prox}_{\tau f}(x) = \frac{x}{1+\tau}$

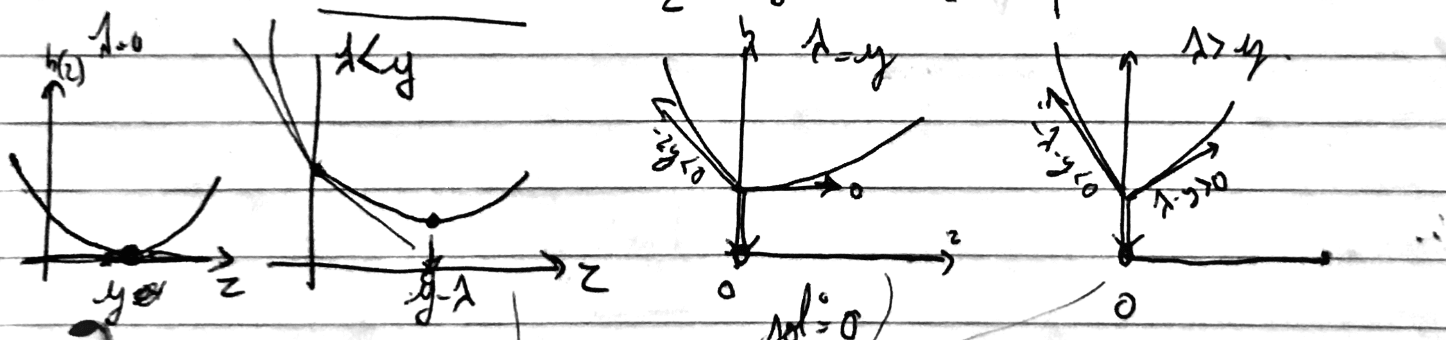
Rmq: $\text{prox}_{\tau f}$ contractant $\|\text{prox}_{\tau f}(x) - \text{prox}_{\tau f}(x')\| \le \|x - x'\|$

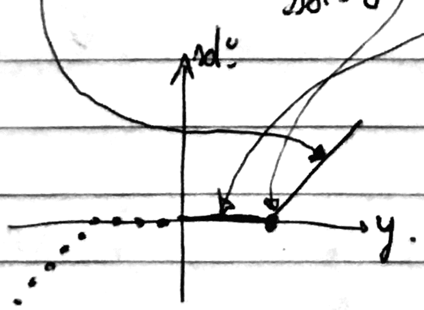
ex: $f(x) = \sum_i f_i(x_i) \Rightarrow \text{prox}_{\tau f}(x) = (\text{prox}_{\tau f_i}(x_i))_i$

ex: $f(x) = |x|$, $x \in \mathbb{R}$ $\quad \text{prox}_{\tau f}(x) = \text{soft}_\tau(x) = (|x|-\tau)_+ \text{sign}(x)$



Preuve dessin: $h(z) = \frac{1}{2}|z-y|^2 + \lambda|z|_1$ pour $x>0$.





$\text{prox}_{\tau|\cdot|}(-y) = -\text{prox}_{\tau|\cdot|}(y)$

Duality ⑤

Lagrange duality : special case of equality

PRIMAL $\min_{Ax=y} f(x)$

$$\min_x \max_h f(x) + \langle y - Ax, h \rangle$$

$$= \max_h \min_x f(x) + \langle y - Ax, h \rangle .$$

$$= \max_h \langle y, h \rangle + \left[\min_x f(x) - \langle x, A^* h \rangle \right]$$

$$= \max_h \langle y, h \rangle - \underbrace{\left[\max_x \langle x, A^* h \rangle - f(x) \right]}_{\triangleq f^*(A^* h)}$$

Thm minmax = maxmin if $\exists x_0, Ax_0 = y$

if f smooth $\partial_x = 0 \iff \nabla f(x^*) = A^* h$

DUAL $= \boxed{\max \; \langle y, h \rangle - f^*(A^* h)}$

Primal – Dual rel°
$$\begin{cases} Ax^* = y \\ \nabla f(x^*) = A^* h \end{cases}$$
constraints.

Advantage : Sometimes Dual is "simpler"
+ sometimes ∇f can be inverted to recover the primal from the dual

Application

$\boxed{\text{Clamp \& Duality}}$

Duality 6

$\boxed{\text{Goal: leverage strong convexity to deal with non smooth loss}}$

$$\min_{z} L(\underbrace{Xw}_{z}, y) + \frac{\lambda}{2}\|w\|^2$$

$$\min_{w,z}\ \max_h$$

$$\max_h \min_{w,z} L(z,y) + \frac{\lambda}{2}\|w\|^2 + \langle h, Xw - z\rangle$$

$$\max_h \left[\min_z -\langle z,h\rangle + L(z,y)\right] + \left[\min_w \langle w, X^*h\rangle + \frac{\lambda}{2}\|w\|^2\right]$$

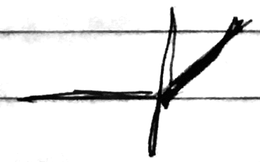
$$-L^*(+h,y) \quad -\left(\frac{\lambda}{2}\|\cdot\|^2\right)^*(-X^*h).$$

$$\boxed{\max_h -\frac{\lambda}{2}\left\|-\frac{X^*h}{\lambda}\right\|^2 - L^*(h,y)}$$

$$X^*h + \lambda w = 0 \Rightarrow w = -\frac{X^*h}{\lambda}.$$

sparse (support vect)

$\boxed{\text{ex}}$ $L(z,y) = \max(0, z\cdot y).$

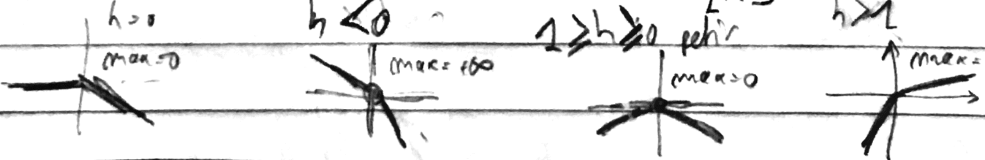


$y \to 0$

$y = \pm 1.$

$$L^*(h,y) = \max_z zh - (zy)^+ = \iota_{[??]}(h)$$

$y=1$: $((\cdot)^+)^* = \iota_{[0,1]}$



$y=-1$

$$\boxed{y\in\{-1,+1\}: L^*(h,y) = \iota_{[0,1]}(y \circ h)}$$

$\boxed{\text{Dual SVM}}$ $\max_h \|X^*h\|^2$ st $\forall i,\ h_i \cdot y_i \in [0,1] \rightarrow \boxed{w = -\frac{X^*h}{\lambda}}$