Optimisation convexe non-lisse

I / Motivations et rappels

1) Un problème d'optimisation "fréquent"

$$\min_{x \in \mathbb{R}^n} h(x) = f(x) + g(x)$$

avec $f : \mathbb{R}^n \to \mathbb{R}$ lisse à savoir un gradient lipschitzien :

$\exists\, L > 0$ tq $\forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^n \quad \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$

$g : \mathbb{R}^n \to \mathbb{R}$ convexe potentiellement non-lisse.

ex : i) $g = 0$ optimisation lisse non convexe

ii) $g(x) = \lambda \|x\|_1$ avec $\lambda > 0$: régularisation parcimonieuse

iii) Reformulation d'un problème d'optimisation avec contraintes

$\min_{x \in \mathcal{C}} f(x)$ avec $\mathcal{C}$ convexe $\neq \emptyset$ de $\mathbb{R}^n$

$\Rightarrow$ "pénalisation" $g(x) = \begin{cases} 0 \text{ si } x \in \mathcal{C} \\ +\infty \text{ sinon} \end{cases}$

$$\min_{x \in \mathbb{R}^n} h(x) = f(x) + g(x)$$

2) Exemples d'applications

exemple 1 : Moindres carrés régularisés.

On dispose d'un modèle linéaire :

$\forall x \in \mathbb{R}^n \quad f(x, \beta) = x^T \beta$ avec $\beta \in \mathbb{R}^n$ paramètres inconnus

On dispose d'observations $(x_i, y_i) \in (\mathbb{R}^n \times \mathbb{R}^p)$ permettant d'estimer β. D'où le problème d'optimisation :

$$\min_{\beta \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^{p} (f(x_i, \beta) - y_i)^2 \Leftrightarrow \min_{\beta \in \mathbb{R}^n} \frac{1}{2} \|X\beta - y\|_2^2$$

avec $X = \begin{bmatrix} x_1^T \\ \vdots \\ x_p^T \end{bmatrix} \in \mathcal{M}_{p,n}(\mathbb{R}) \quad y = (y_i) \in \mathbb{R}^p$

$\Rightarrow$ terme de régularisation : $\min_{\beta \in \mathbb{R}^n} \frac{1}{2} \|X\beta - y\|_2^2 + g(\beta)$

i) $g(\beta) = \frac{\lambda}{2} \|\beta\|_2^2$ régularisation de Tikhanov

ii) $g(\beta) = \lambda \|\beta\|_1$ régularisation parcimonieuse (LASSO)

iii) $g(\beta) = \frac{\lambda}{2} \|\beta\|_B^2$ avec $\|\beta\|_B^2 = \beta^T B \beta$ et $B \in S_n(\mathbb{R})$ définie positive

régularisation de Tikhanov généralisée

iv) $g(\beta) = \frac{\lambda}{p} \|\beta\|_p^p$ avec $p > 1$ et $\|\beta\|_p = \left[\sum_{i=1}^{n} |\beta_i|^p\right]^{1/p}$

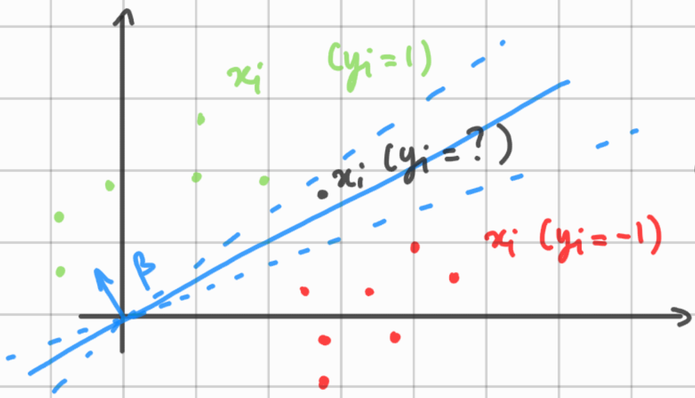
Rem : $S_n(\mathbb{R})$: matrice symétrique

exemple 2 : SVM séparateur à vaste marge

On dispose de données $(x_i)_{i \in [\![1,p]\!]} \in (\mathbb{R}^n)^p$ labélisées $(y_i)_{i \in [\![1,p]\!]} \in [-1,1]^p$

On cherche à construire un hyperplan séparant les données (x_i) selon leurs labels (y_i)

. Dans un premier temps on suppose qu'il existe un tel hyperplan de vecteur normal $\beta \in \mathbb{R}^n$ passant par l'origine.

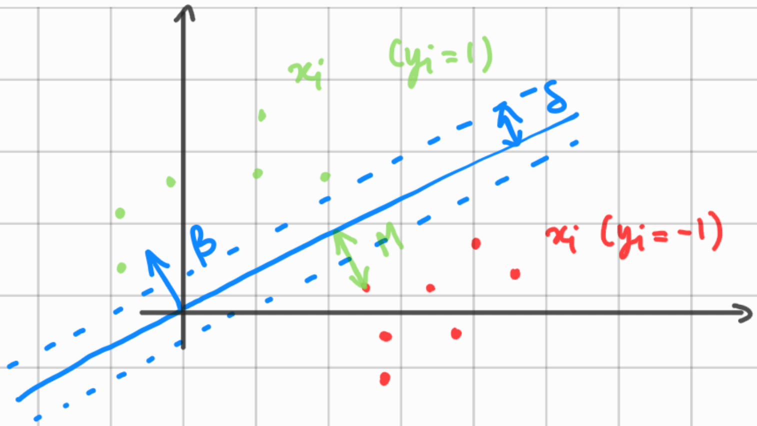


la classe du nouvel x_i dépend de l'hyperplan choisi

Q ? quel hyperplan choisir ?

$\Rightarrow$ incertitude : nombre de données, répartition dans $\mathbb{R}^n$, etc...

$\Rightarrow$ maximiser la distance minimale entre les données et l'hyperplan.



Condition de séparabilité : $\forall i \in [\![1,p]\!]$ $y_i x_i^T\beta \geq 0$
d'où le problème d'optimisation :

$$\max_{(\beta,M)\in\mathbb{R}^n\times\mathbb{R}^+} M$$
$$\text{s.c } \forall i\in[\![1,p]\!] \quad \frac{y_i x_i^T\beta}{\|\beta\|_2} \geq M$$

<u>Rem:</u> $d\left(z, \{\beta^T x=0\}\right) = \dfrac{|\beta^T z|}{\|\beta\|_2}$

- En pratique, la condition de séparabilité n'est pas vérifiée pour tout $i \in [\![1,p]\!]$

$\Rightarrow$ pénalisation des contraintes non satisfaites

On pose $\forall t\in\mathbb{R}$ $t^+ = \max(t,0)$

ceci conduit à formuler un autre problème d'optimisation :

$$\max_{(\beta,M)\in\mathbb{R}^n\times\mathbb{R}_+} M - \lambda\sum_{i=1}^{p}\left(1-\frac{y_i x_i^T\beta}{\|\beta\|_2 M}\right)_+ \quad \text{avec } \lambda>0$$

Avec $\|\beta\|_2 = \frac{1}{M}$ et en reformulant pour obtenir un problème de minimisation :

$$\min_{\beta\in\mathbb{R}^n} \|\beta\|_2^2 + \underbrace{\lambda\sum_{i=1}^{p}\left(1-y_i x_i^T\beta\right)_+}_{\text{non lisse}}$$

3) Rappels de convexité

def : Soit $\mathcal{C}\subset\mathbb{R}^n$ $\mathcal{C}$ est une partie convexe de $\mathbb{R}^n$ si

$\forall (x,y) \in \mathcal{C}^2 \quad \forall \lambda \in [0,1] \quad \lambda x + (1-\lambda) y \in \mathcal{C}$

ex i) partie affine $x_0 + S$ avec S sev de $\mathbb{R}^n$ et $x_0 \in \mathbb{R}^n$

$:= \{x_0 + s \text{, avec } s \in S\}$

ii) hyperplan $\{x \in \mathbb{R}^n \text{ tq } a^T x = b\}$

iii) Demi-espace $\{x \in \mathbb{R}^n \text{ tq } a^T x \leq b\}$

iv) polyhèdre : $\{Ax \leq b\}$ avec A matrice

v) ellipsoïde : $\{x \in \mathbb{R}^n \text{ tq } x^T \mathcal{C} x \leq 1\}$ avec $\mathcal{C} \in S_n(\mathbb{R})$ semie définie positive

propriété : opérations préservant la convexité

i) intersection

ii) somme

iii) multiplication par un scalaire.

iv) produit cartésien

v) image réciproque d'une application linéaire $U^{-1}(\mathcal{C})$ $\mathcal{C}$ convexe

vi) Image directe par une application linéaire

vii) projection : $\{x_1 \text{ tq } (x_1 | x_2) \in \mathcal{C}\}$ avec $\mathcal{C}$ convexe

def : fonction convexe sur $\mathcal{C}$ convexe

$f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C}$ convexe de $\mathbb{R}^n$

- f est convexe sur $\mathcal{C}$ (convexe) si $\forall (x,y) \in \mathcal{C}^2 \quad \forall \lambda \in [0,1]$

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

- f est strictement convexe sur $\mathcal{C}$ (convexe) si $\forall (x,y) \in \mathcal{C}^2 \quad \forall \lambda \in]0,1[$

$x \neq y \quad f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$

propriété : CNS convexité dans le cas dérivable

Soit $f : \Omega \to \mathbb{R}$ avec Ω ouvert et $\mathcal{C} \subset \Omega$ convexe

1) Si f est dérivable sur Ω, alors

f convexe sur $\mathcal{C}$ convexe $\Leftrightarrow \forall (x,y) \in \mathcal{C}^2 \quad f(y) \geq f'(x)(y-x) + f(x)$

$\Leftrightarrow \forall (x,y) \in \mathcal{C}^2 \quad f(y) \geq \nabla f(x)^T (y-x) + f(x)$

2) Si f est deux fois dérivable sur Ω, f convexe sur $\mathcal{C}$ convexe

$\Leftrightarrow \forall (x,y) \in \mathcal{C}^2 \quad f''(x)(y-x, y-x) \geq 0$

$\Leftrightarrow \forall (x,y) \in \mathcal{C}^2 \quad (y-x)^T \nabla^2 f(x)(y-x) \geq 0$

__propriété__ Soit $\mathcal{C}$ convexe de $\mathbb{R}^n$ Soit $f : \mathcal{C} \to \mathbb{R}$

f est convexe sur $\mathcal{C}$ convexe $\Leftrightarrow \forall x \in \mathcal{C} \ \forall v \in \mathbb{R}^n$

$t \to f(x+tv)$ est convexe sur

$\{ t \in \mathbb{R} \text{ tq } x + tv \in \mathcal{C} \}$ convexe

__propriété__

i) f convexe sur $\mathcal{C}$ convexe $\Rightarrow \alpha f$ convexe sur $\mathcal{C}$ convexe avec $\alpha > 0$

ii) combinaisons linéaires à coefficients positifs de fonctions convexes sont convexes.

iii) f convexe sur $\mathcal{C}$ convexe

Soient $A \in \mathcal{M}_{m,n}(\mathbb{R})$ et $b \in \mathbb{R}^m$ alors $\mathcal{C}^1 = \{ x \in \mathbb{R}^n \text{ tq } Ax+b \in \mathcal{C} \}$

est convexe et $x \mapsto f(Ax+b)$ convexe sur $\mathcal{C}$

iv) $(f_i)_{i \in [\![1,m]\!]}$ convexes sur $(\mathcal{C}_i)_{i \in [\![1,m]\!]}$ alors

$\max\limits_{i \in [\![1,m]\!]} f_i$ est convexe sur $\bigcap\limits_{i=1}^{m} \mathcal{C}_i$

v) $g : \mathbb{R}^n \to \mathbb{R}$ convexe sur $\mathcal{C} \subset \mathbb{R}^n$

$h : \mathbb{R}^n \to \mathbb{R}$ convexe sur $\mathcal{C}'$ tq $g(\mathcal{C}) \subset \mathcal{C}'$ et croissante

alors $h \circ g$ est convexe sur $\mathcal{C}$

vi) $g : \mathbb{R}^n \to \mathbb{R}^p$ avec $\forall i \in [\![1,p]\!]$ g_i convexe sur $\mathbb{R}^n$

$h : \mathbb{R}^p \to \mathbb{R}$ convexe et croissante vis-à-vis de chacun de ses arguments alors $f : \mathbb{R}^n \to \mathbb{R}$

$x \mapsto h \circ g(x) = h(g_1(x), \dots, g_p(x))$

est convexe sur $\mathbb{R}^n$

4) Régularité des fonctions convexes

<u>def</u> : épigraphe de f

Soit $f : \mathcal{C} \to \mathbb{R}$ on appelle épigraphe de f noté $\mathcal{E}(f)$,

$$\mathcal{E}(f) = \{(x, \omega) \in \mathcal{C} \times \mathbb{R} \text{ tq } f(x) \leq \omega\}$$

<u>propriété</u> :

$f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C}$ convexe de $\mathbb{R}^n$

f convexe sur $\mathcal{C}$ convexe $\Leftrightarrow$ $\mathcal{E}(f)$ convexe

<u>preuve</u> ($\Rightarrow$) Supposons f convexe sur $\mathcal{C}$ convexe

$\forall (x,y) \in \mathcal{C}^2 \quad \forall \lambda \in [0,1] \quad f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$

Soient $(x_1, \omega_1) \in \mathcal{E}(f)$, $(x_2, \omega_2) \in \mathcal{E}(f)$ et $\lambda \in [0,1]$

$(\lambda x_1 + (1-\lambda) x_2 , \lambda \omega_1 + (1-\lambda) \omega_2) \in \mathcal{E}(f)$?

$f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda \underbrace{f(x_1)}_{\leq \omega_1} + (1-\lambda) \underbrace{f(x_2)}_{\leq \omega_2}$ par convexité de f sur $\mathcal{C}$ / par définition de $\mathcal{E}(f)$

$\leq \lambda \omega_1 + (1-\lambda) \omega_2$

$\Rightarrow (\lambda x_1 + (1-\lambda) x_2 , \lambda \omega_1 + (1-\lambda) \omega_2) \in \mathcal{E}(f)$

$\Rightarrow \lambda (x_1, \omega_1) + (1-\lambda)(x_2, \omega_2) \in \mathcal{E}(f)$ $\Rightarrow \mathcal{E}(f)$ convexe

($\Leftarrow$) supposons $\mathcal{E}(f)$ convexe

Soient $(x,y) \in \mathcal{C}^2$, $\lambda \in [0,1]$

on $(x, f(x)) \in \mathcal{E}(f)$ et $(y, f(y)) \leq \mathcal{E}(f)$

d'où par convexité de $\mathcal{E}(f)$

$$\lambda (x, f(x)) + (1-\lambda)(y, f(y)) \in \mathcal{E}(f)$$

$\Rightarrow f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$

par def de $\mathcal{E}(f)$

<u>propriété</u> : Inégalité de Jensen

Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C}$ convexe de $\mathbb{R}^n$

$\forall (x_i)_{i\in[\![1,p]\!]} \in \mathcal{C}^p \quad \forall(\lambda_i)_{i\in[\![1,p]\!]} \in \mathbb{R}^p$ tq $\forall i \in [\![1,p]\!]$ $\lambda_i \geq 0$
et $\sum_{i=1}^{p} \lambda_i = 1$

alors $f\left(\sum_{i=1}^{p} \lambda_i x_i\right) \leq \sum_{i=1}^{p} \lambda_i f(x_i)$

<u>propriété</u>: Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C}$ convexe de $\mathbb{R}^n$
Soit $x_0 \in \mathring{\mathcal{C}}$ alors f est continue en x_0.

<u>preuve</u>: soit $x_0 \in \mathring{\mathcal{C}}$
Soit Δ un simplexe inclus dans $\mathring{\mathcal{C}}$ et contenant x_0 :
Notons $(\Delta_i)_{i\in[\![1,n+1]\!]}$ les sommets de Δ
$\forall x \in \Delta$, $\exists! (\lambda_i)_{i\in[\![1,n+1]\!]} \in \mathbb{R}^{n+1}$ tq $\lambda_i \geq 0$ $\forall i \in [\![1,n+1]\!]$,

$\sum_{i=1}^{n+1} \lambda_i = 1$ et $x = \sum_{i=1}^{n+1} \lambda_i \Delta_i$

(coordonnées barycentriques de x vis-à-vis de Δ)
d'où $f(x) \leq \sum_{i=1}^{n+1} \lambda_i f(\Delta_i)$ par convexité de f et inégalité de Jensen

$$\leq \max_{j\in[\![1,n+1]\!]} f(\Delta_j) \underbrace{\sum_{i=1}^{n+1} \lambda_i}_{=1 \text{ par def de } (\lambda_i)_{i\in[\![1,n+1]\!]}}$$

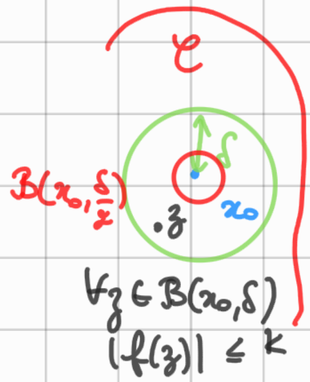
$$\leq \max_{j\in[\![1,n+1]\!]} f(\Delta_j)$$

$\Rightarrow$ f est majorée sur Δ
En particulier $\forall \delta > 0$ tq $B(x_0, \delta) \subset \Delta$,
f majorée sur $B(x_0, \delta)$
Fixons un tel δ et notons M un majorant de f sur $B(x_0, \delta)$
Soit $z \in B(x_0, \delta)$, on pose $z' = 2x_0 - z$
alors $z' \in B(x_0, \delta)$: $\|z' - x_0\| = \underbrace{\|x_0 - z\|}_{< \delta}$

Alors $x_0 \in [z, z']$

Par convexité de f sur $\mathcal{C}$: $f(x_0) \leq \frac{1}{2} f(z) + \frac{1}{2} \underbrace{f(z')}_{\leq M \text{ car } z' \in B(x_0,\delta)}$

$\Rightarrow f(z) \geq 2 f(x_0) - M$



<u>Bilan</u> : $2 f(x_0) - M \leq f(z) \leq M \quad \forall z \in B(x_0, \delta)$

$\Rightarrow f$ bornée sur $B(x_0,\delta)$

Soit $K > 0$ tq $\forall z \in B(x_0,\delta)$ $|f(z)| \leq K$

. Mq f est lipschitzienne sur $B(x_0, \frac{\delta}{2})$

Soient $(x,y) \in B(x_0, \frac{\delta}{2})^2$

On pose $\begin{cases} x' = x - \frac{\delta}{2} \frac{y-x}{\|y-x\|} \\ y' = y + \frac{\delta}{2} \frac{y-x}{\|y-x\|} \end{cases}$

alors $x' \in B(x_0,\delta)$: $\| x' - x_0 \| = \| x - x_0 - \frac{\delta}{2} \frac{y-x}{\|y-x\|} \|$

$$\leq \underbrace{\|x - x_0\|}_{< \frac{\delta}{2} \text{ car } x \in B(x_0,\frac{\delta}{2})} + \frac{\delta}{2}$$

$$< \delta$$

Idem pour $y' \in B(x_0,\delta)$

d'où $|f(x')| \leq K$ et $|f(y')| \leq K$

De plus, $x = \underbrace{\frac{2\|y-x\|}{2\|y-x\|+\delta}}_{= \lambda \in [0,1]} x' + \underbrace{\frac{\delta}{2\|y-x\|+\delta}}_{= 1-\lambda} y = \lambda x' + (1-\lambda) y$

Par convexité de f sur $\mathcal{C}$: $f(x) \leq \lambda f(x') + (1-\lambda) f(y)$

$\Rightarrow f(x) - f(y) \leq \lambda (f(x') - f(y))$

$\leq 2K\lambda$

$$\leq \frac{4K}{\frac{2}{\delta}\|y-x\|+\delta}\|y-x\|$$

$\Rightarrow f(x)-f(y) \leq \frac{4K}{\delta}\|y-x\|$

De même $y=\lambda y'+(1-\lambda)x$ et de la même manière
$f(y)-f(x) \leq \frac{4K}{\delta}\|y-x\|$

$\Rightarrow |f(y)-f(x)| \leq \frac{4K}{\delta}\|y-x\| \ \forall (x,y)\in B(x_0,\frac{\delta}{2})$

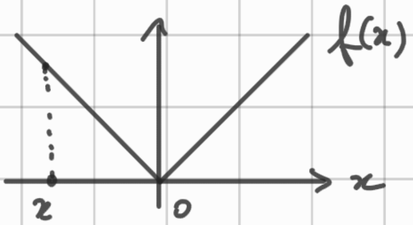
Bilan : f $\frac{4K}{\delta}$- lipschitzienne sur $B(x_0,\frac{\delta}{2})$
En particulier f continue en x_0.

II. Sous différentiel d'une fonction

1) Sous-gradient, sous-différentiel

def : $f:\mathcal{C}\to\mathbb{R}$ avec $\mathcal{C}\subset\mathbb{R}^n$ Soient $x\in\mathcal{C}$ et $g\in\mathbb{R}^n$
g est appelé sous-gradient de f en x si
$\forall y\in\mathcal{C} \quad f(y)\geq g^T(y-x)+f(x)$
On appelle sous différentiel de f en x noté $\partial f(x)$, l'ensemble des sous gradients de f en x

ex : $f:\mathbb{R}\to\mathbb{R}$
$x\mapsto|x|$



Soit $x\in\mathbb{R}$ Q? $\partial f(x)$?
1er cas $x<0$
alors $f(x)=-x$
Soit $g\in\mathbb{R}$ tq $\forall y\in\mathbb{R}$ $f(y)\geq g(y-x)+f(x)$

Soit $y \geq 0$ $f(y) = -y = -y + x - x = -(y-x) + f(x)$

$\Leftrightarrow -(y-x) + f(x) \geq g(y-x) + f(x)$
$\Leftrightarrow -(1+g)(y-x) \geq 0 \quad \forall y \geq 0$
$\Leftrightarrow g = -1$

Soit $y > 0$ $f(y) = y \geq -y \geq -(y-x) + f(x) \geq g(y-x) + f(x)$
ok avec $g = -1$
d'où $\partial f(x) = \{-1\}$

$2^{\text{ème}}$ cas $\partial f(x) = \{1\}$

$3^{\text{ème}}$ cas $x = 0$ $f(x=0)$
Soit $g \in \mathbb{R}$ tq $\forall y \in \mathbb{R}$ $f(y) \geq g(y-x) + f(x) \geq gy$
Soit $y > 0$ $f(y) = y$
$f(y) \geq gy \Leftrightarrow (1-g)\, y \geq 0$
$\Leftrightarrow g \leq 1$

. Soit $y \leq 0$ $f(y) = -y$ $f(y) \geq g(y) \Leftrightarrow -(1+g) y \geq 0$
$\Leftrightarrow g \geq -1$

$\Rightarrow g \in [-1, 1]$
On vérifie $\partial f(0) = [-1, 1] \setminus 0$

<u>propriété</u>: Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C} \subset \mathbb{R}^n$
Soit $x \in \mathcal{C}$ Alors $\partial f(x) = \emptyset$ ou $\partial f(x)$ est une partie convexe fermée de $\mathbb{R}^n$

preuve : Supposons $\partial f(x) \neq \emptyset$ alors $\partial f(x) = \bigcap_{y \in \mathcal{C}} \{ g \in \mathbb{R}^n \text{ tq } f(y) \geq g^T(y-x) + f(x) \}$
or $\forall y \in \mathcal{C}$ $\{ g \in \mathbb{R}^n \text{ tq } f(y) \geq g^T(y-x) + f(x) \}$
est convexe (demi-espace) et fermée comme image

réciproque d'un fermé par une application continue avec $\varphi : \mathbb{R}^n \to \mathbb{R}$
$g \mapsto f(y) - f(x) - g^T(y-x)$

$\Rightarrow \partial f(x)$ convexe fermé comme intersection de parties convexes et fermées

<u>propriété</u> : Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C} \subset \mathbb{R}^n$
Soit $x \in \mathring{\mathcal{C}}$ tq f continue en x alors $\partial f(x)$ est borné

<u>preuve</u> : Soit $x \in \mathring{\mathcal{C}}$ tq f continue en x
$x \in \mathring{\mathcal{C}}$: $\exists\, y_1 > 0$ tq $B(x_1, y_1) \subset \mathcal{C}$ f continue en x
$\forall \varepsilon > 0 \quad \exists y_2 > 0$ tq $\forall y \in B(x, y_2) \quad |f(y) - f(x)| < \varepsilon$

Posons $y = \min(y_1, y_2)$
Mq $\partial f(x)$ borné
Supposons le contraire
$\forall M > 0 \quad \exists g \in \partial f(x)$ tq $\|g\| > M$
Soit $M > 0$ Fixons un tel $g \in \partial f(x) \Rightarrow g \neq 0$
Soit $y = x + \frac{\eta}{2} \frac{g}{\|g\|}$

D'où $\|y - x\| = \frac{\eta}{2} \Rightarrow y \in B(x, y) \subset \mathcal{C}$

Par définition de $g \in \partial f(x)$
$f(y) \geq g^T(y-x) + f(x)$

$f(y) - f(x) \geq \frac{\eta}{2}\|g\| > \frac{\eta}{2} M$ par hypothèse.

Avec $M = \frac{2}{\eta}\varepsilon$ $\qquad f(y) - f(x) > \varepsilon$

Or $y \in B(x, y) \Rightarrow y \in B(x, y_2) \Rightarrow |f(y) - f(x)| < \varepsilon$
d'où $\varepsilon < f(y) - f(x) < \varepsilon$ contradiction

d'où $\partial f(x)$ borné

<u>propriété</u> : $f : \mathcal{C} \to \mathbb{R}$ convexe sur $\mathcal{C}$ convexe de $\mathbb{R}^n$

i) $\forall x \in \mathring{\mathcal{C}}$ $\partial f(x) \neq \emptyset$

Et $\partial f(x)$ compact convexe $\neq \emptyset$ de $\mathbb{R}^n$

ii) Si f dérivable en $x \in \mathring{\mathcal{C}}$ alors $\partial f(x) = \{\nabla f(x)\}$

preuve :

i) lemme : Soit $\mathcal{C}$ convexe $\neq \emptyset$ de $\mathbb{R}^n$

Soit $x_0 \in C_{\mathcal{C}} \cup (\overline{\mathcal{C}} \setminus \mathring{\mathcal{C}})$

alors $\exists a \in \mathbb{R}^n \setminus \{0\}$ tq $\sup\limits_{z \in \mathcal{C}} a^T z \leq a^T x_0$

$C_{\mathcal{C}}$ complémentaire de $\mathcal{C}$ dans $\mathbb{R}^n$

$\overline{\mathcal{C}} \setminus \mathring{\mathcal{C}}$ frontière de $\mathcal{C}$.

Soit $x \in \mathring{\mathcal{C}}$ En particulier $(x, f(x)) \in \overline{\mathcal{E}(f)} \setminus \mathring{\mathcal{E}}(f)$

De plus f convexe $\Rightarrow \mathcal{E}(f)$ convexe

d'où par lemme

$\exists (a, \alpha) \in (\mathbb{R}^n \times \mathbb{R}) \setminus \{0\}$ tq $\forall (y, \omega) \in \mathcal{C} \times \mathbb{R}$

$$[a^T \ \alpha] \begin{bmatrix} y \\ \omega \end{bmatrix} \leq [a^T \ \alpha] \begin{bmatrix} x \\ f(x) \end{bmatrix}$$

$$\Leftrightarrow a^T (y - x) + \alpha (\omega - f(x)) \leq 0$$

Si $\alpha > 0$ alors $a^T (y - x) \leq \alpha (f(x) - \omega) \xrightarrow[\omega \to +\infty]{} -\infty$ impossible

Si $\alpha = 0$ alors $a^T (y - x) \leq 0$

or $x \in \mathcal{C}^\circ$: $\exists N \in \mathbb{N}$ tq $\forall m \geq N$ $y_n = x + \frac{a}{m} \in \mathcal{C}$

d'où $\forall m \geq N$ $\frac{\|a\|^2}{m} \leq 0 \Rightarrow a = 0$ impossible car $\begin{bmatrix} a \\ \alpha \end{bmatrix} \neq 0$

d'où $\alpha < 0$

Il vient $a^T (y - x) \leq \alpha (f(x) - \omega) \Rightarrow f(x) - \omega \leq \left[\frac{a}{\alpha}\right]^T (y - x)$

$$\Rightarrow w \geq f(x) + \left[-\frac{a}{\alpha}\right]^T (y-x)$$

En particulier avec $w = f(y)$

$$\Rightarrow f(y) \geq f(x) + \left[-\frac{a}{\alpha}\right]^T (y-x) \ \forall y \in \mathcal{C} \Rightarrow \frac{-a}{\alpha} \in \partial f(x)$$

$$\Rightarrow \partial f(x) \neq \emptyset$$

ii) On suppose de plus f dérivable en $x \in \mathring{\mathcal{C}}$

+ f convexe sur $\mathcal{C}$ convexe et dérivable en $x \in \mathring{\mathcal{C}}$

$\Rightarrow \forall y \in \mathcal{C} \quad f(y) \geq \nabla f(x)^T (y-x) + f(x)$

$\Rightarrow \nabla f(x) \in \partial f(x)$

$\Rightarrow \{\nabla f(x)\} \subset \partial f(x)$

Soit $g \in \partial f(x)$

$\forall y \in \mathcal{C} \quad f(y) \geq g^T (y-x) + f(x)$

Or $x \in \mathring{\mathcal{C}}$ $\exists N \in \mathbb{N}$ tq $\forall n \in \mathbb{N}$ $y_n = x + \frac{u}{n} \in \mathcal{C}$ avec $U \in \mathbb{R}^n$

Fixons un tel N

$\forall n \in N \quad f(y_n) \geq \frac{1}{n} g^T U + f(x)$

Or f dérivable en x :

$f(y_n) = f(x) + \frac{1}{n} \nabla f(x)^T U + \frac{1}{n} \|u\| \varepsilon(\frac{1}{n} U)$ avec $\varepsilon(h) \xrightarrow[h \to 0]{} 0$

$$\Leftrightarrow \cancel{f(x)} + \frac{1}{n} \nabla f(x)^T U + \frac{1}{n} \|u\| \varepsilon\left(\frac{U}{n}\right) \geq \frac{1}{n} g^T U + \cancel{f(x)}$$

$$\Leftrightarrow (\nabla f(x) - g)^T U + \underbrace{\|u\| \varepsilon\left(\frac{u}{n}\right)}_{\xrightarrow[n \to +\infty]{} 0} \geq 0$$

À la limite $\quad (\nabla f(x) - g)^T U \geq 0 \quad \forall U \in \mathbb{R}^n$

$\Rightarrow g = \nabla f(u)$ et $\partial f(x) \subset \{\nabla f(x)\}$

Bilan : $\partial f(x) = \{\nabla f(x)\}$

<u>propriété</u> $f : \mathcal{C} \to \mathbb{R}$ convexe sur $\mathcal{C}$ convexe

Soit $x^* \in \mathcal{C}$

x^* minimum global de f sur $\mathcal{C}$ $\Leftrightarrow$ $0 \in \partial f(x^*)$

preuve : $0 \in \partial f(x^*) \Leftrightarrow \forall y \in \mathcal{C}\ f(y) \geq 0^T(y - x^*) + f(x^*)$

$\Leftrightarrow \forall y \in \mathcal{C}\ f(y) \geq f(x^*)$

$\Leftrightarrow$ x^* minimum global de f sur $\mathcal{C}$

2) Calculs de sous-gradients

Pour simplifier, on suppose avoir (f_i) convexes sur $\mathbb{R}^n$

<u>propriété</u>

i) soit $(\alpha_1, \alpha_2) \in {\mathbb{R}_+^*}^2$

On pose $f = \alpha_1 f_1 + \alpha_2 f_2$

alors $\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$

ii) $h : \mathbb{R}^n \to \mathbb{R}$

$x \mapsto f(Ax+b)$ avec $A \in \mathcal{M}_{p,n}(\mathbb{R})$ $b \in \mathbb{R}^p$

f convexe sur $\mathbb{R}^p$

$\partial h(x) = A^T \partial f(Ax+b)$

iii) $f : \mathbb{R}^n \to \mathbb{R}$

$x \mapsto \max\limits_{i \in [\![1,p]\!]} f_i(x)$

Soit $I_0 = \{ i \in [\![1,p]\!] \text{ tq } f_i(x) = \max\limits_{j \in [\![1,p]\!]} f(x) \}$

$\forall g\ \partial f_{i_0}(x)$ avec $i_0 \in I_0$, $g \in \partial f(x)$

iv) $f : \mathbb{R}^n \to \mathbb{R}$
$x \mapsto \sup_{\alpha \in A} f_\alpha(x)$

On note $I_0 = \{\alpha \in A \text{ tq } f_\alpha(x) = \sup_{\beta \in A} f_\beta(x)\}$

$\forall g \in \partial f_\alpha(x)$ avec $\alpha \in I_0$ $\quad g \in \partial f(x)$

III. Algorithmes de sous-gradient

1) algorithme du sous-gradient

(P) $\min_{x \in \mathbb{R}^n} f(x)$ avec f convexe sur $\mathbb{R}^n$ (convexe)

On suppose que f admet au moins un minimum $x^* \in \mathbb{R}^n$

Algo: $\quad x_0 \in \mathbb{R}^n$
Tant que
  calculer $g_k \in \partial f(x_k)$
  $x_{k+1} = x_k - \alpha_k g_k$ avec $\alpha_k > 0$
Fin Tant que

rem: f convexe sur $\mathbb{R}^n \Rightarrow \forall x \in \mathbb{R}^n \; \partial f(x) \neq \emptyset$ et $\partial f(x)$ est un compact convexe de $\mathbb{R}^n$

Soit $k \in \mathbb{N}$

$$\|x_{k+1} - x^*\|^2 = \|x_k - \alpha_k g_k - x^*\|^2$$
$$= \|x_k - x^*\|^2 - 2\alpha_k g_k^T (x_k - x^*) + \alpha_k^2 \|g_k\|^2$$

or $g_k \in \partial f(x_k)$

d'où $\forall y \in \mathbb{R}^n \quad f(y) \geq g_k^T (y - x_k) + f(x_k)$

En particulier $f(x^*) \geq g_k^T (x^* - x_k) + f(x_k)$

$$\Rightarrow -g_k^T (x_k - x^*) \leq f(x^*) - f(x_k)$$

d'où $\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\alpha_k (f(x_k) - f(x^*)) + \alpha_k^2 \|g_k\|^2$

Par récurrence

$$\|x_{k+1} - x^*\|^2 \leq \|x_1 - x^*\| - 2\sum_{l=1}^{k} \alpha_l (f(x_l) - f(x^*)) + \sum_{l=1}^{k} \alpha_l^2 \|g_l\|^2$$

On pose $f_{best}^k = \min_{j \in [\![1,k]\!]} f(x_j)$

On a $2\sum_{l=1}^{k} \alpha_l \underbrace{(f(x_l) - f(x^*))}_{\geq f_{best}^k - f(x^*)} \leq \|x_1 - x^*\|^2 + \sum_{l=1}^{k} \alpha_l^2 \|g_l\|^2$

$$\Rightarrow 0 \leq f_{best}^k - f(x^*) \leq \frac{\|x_1 - x^*\|^2 + \sum_{l=1}^{k} \alpha_l^2 \|g_l\|^2}{2\sum_{l=1}^{k} \alpha_l}$$

Soit $R > 0$ tq $\|x_1 - x^*\| \leq R$

$$\boxed{0 \leq f_{best}^k - f(x^*) \leq \frac{R^2 + \sum_{l=1}^{k} \alpha_l^2 \|g_l\|^2}{2\sum_{l=1}^{k} \alpha_l}}$$

On suppose de plus $\exists a > 0$ tq $\forall l \in \mathbb{N} \quad \|g_l\| \leq a$

$$\Rightarrow 0 \leq f_{best}^{k} - f(x^*) \leq \frac{R^2 + a^2 \sum_{l=1}^{k} \alpha_l^2}{2 \sum_{l=1}^{k} \alpha_l}$$

Q? Quelle est la stratégie de pas α_l ?

① Pas constant : $\forall l \in [\![1, k]\!]$ $\alpha_l = \alpha > 0$

$$0 \leq f_{best}^{k} - f(x^*) \leq \frac{R^2 - a^2\alpha^2 k}{2\alpha k} \xrightarrow[k \to +\infty]{} \frac{a^2\alpha}{2}$$

$$\Rightarrow f_{best}^{k} \in \mathcal{B}\left(f(x^*), \frac{a^2\alpha}{2}\right)$$

② $\begin{cases} \forall l \in \mathbb{N} \quad \alpha_l > 0 \\ \sum_{l=1}^{k} \alpha_l \xrightarrow[n \to +\infty]{} +\infty \\ \sum_{l=1}^{+\infty} \alpha_l^2 < +\infty \end{cases}$ Alors $0 \leq f_{best}^{k} - f(x^*) \leq \dfrac{R^2 + a^2 \sum_{l=1}^{k} \alpha_l^2}{\sum_{l=1}^{k} \alpha_l} \xrightarrow[k \to +\infty]{} 0$

donc $f_{best}^{k} \xrightarrow[k \to +\infty]{} f(x^*)$

③ $\forall l \in \mathbb{N}$ $\alpha_l = \dfrac{\gamma_l}{\|g_l\|}$ avec $\begin{cases} \gamma_l > 0 \\ \gamma_l \xrightarrow[l \to +\infty]{} 0 \\ \sum_{l=1}^{k} \gamma_l \xrightarrow[k \to +\infty]{} +\infty \end{cases}$

Et plus généralement :

$\begin{cases} \alpha_l > 0 \ \forall l \in \mathbb{N} \\ \alpha_l \xrightarrow[l \to +\infty]{} 0 \\ \sum_{l=1}^{k} \alpha_l \xrightarrow[k \to +\infty]{} +\infty \end{cases}$

$\alpha_l \xrightarrow[l \to +\infty]{} 0$

$\forall \varepsilon > 0 \ \exists N_1 \in \mathbb{N}$ tq $\forall l \in N_1$ $\underbrace{|\alpha_l|}_{= \alpha_l \text{ car } \alpha_l > 0} < \dfrac{\varepsilon}{a^2}$

En particulier $\alpha_l^2 \leq \alpha_l \dfrac{\varepsilon}{a^2}$

$\forall k \geq N_1 + 1$

$$0 \leq f_{best}^k - f(x^*) \leq \frac{R^2 + a^2 \sum_{l=1}^{N_1} \alpha_l^2}{2 \sum_{l=1}^{k} \alpha_l} + \frac{a^2 \sum_{l=N+1}^{k} \overbrace{\alpha_l^2}^{< \alpha_l \frac{\varepsilon}{a^2}}}{2 \sum_{l=1}^{k} \alpha_l}$$

$$< \frac{R^2 + a^2 \sum_{l=1}^{N_1} \alpha_l^2}{2 \sum_{l=1}^{k} \alpha_l} + \frac{\varepsilon}{2} \underbrace{\frac{\sum_{l=N_1+1}^{k} \alpha_l}{\sum_{l=1}^{k} \alpha_l}}_{\leq 1 \quad \text{car } \alpha_l > 0}$$

De plus $\sum_{l=1}^{k} \alpha_l \to +\infty$

$$\exists N_2 \in \mathbb{N} \text{ tq } \forall k \geq N_2 \sum_{l=1}^{k} \alpha_l \geq \frac{R^2 + a^2 \sum_{l=1}^{N_1} \alpha_l}{\varepsilon}$$

d'où $\forall k \geq \max(1+N_1, N_2)$

$$0 \leq f_{best}^k - f(x^*) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \text{et ce } \forall \varepsilon > 0$$

$$\Rightarrow f_{best}^k \xrightarrow[k \to +\infty]{} f(x^*)$$

Rem:

On s'intéresse à la stratégie de pas qui minimise

$$\varphi(\alpha) = \frac{R^2 + a^2 \|\alpha\|_2^2}{2 \|\alpha\|_1} \quad \text{avec } \alpha \in (\mathbb{R}_+^*)^{k+1}$$

Avec une telle stratégie de pas,
il pourra être nécessaire de faire $o(\frac{1}{\varepsilon^2})$ itérations pour atteindre $|f_{best}^k - f(x^*)| < \varepsilon$ (cf TD1)

$\varepsilon = 10^{-3} \Rightarrow 10^6$ itérations

$\Rightarrow$ l'algorithme est très lent dans le pire cas

2) algorithme du sous-gradient projeté

$\min_{x \in \mathcal{C}} f(x)$ avec $\mathcal{C}$ convexe fermé non vide de $\mathbb{R}^n$

(P) $\min_{x \in \mathcal{C}} f(x)$ avec $\mathcal{C}$ partie convexe, fermée, non vide de $\mathbb{R}^n$

On suppose que (P) admette au moins une solution $x^* \in \mathcal{C}$

<u>Rem:</u> $\mathcal{C}$ partie convexe fermée $\neq \emptyset$ de $\mathbb{R}^n$:
la projection orthogonale sur $\mathcal{C}$, notée $\Pi_{\mathcal{C}}$, est 1-lipschitzienne
(cf cours analyse Hilbertienne)

$\forall (x,y) \in (\mathbb{R}^n)^2 \quad \|\Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(y)\| \leq \|x-y\|$

Soit $x \in \mathbb{R}^n \quad \|x - \Pi_{\mathcal{C}}(x)\| = \min_{y \in \mathcal{C}} \|x-y\|$

<u>Idée:</u> projeté orthogonale sur $\mathcal{C}$ des itérations de l'algorithme de sous gradient

<u>Algo:</u> $x_0 \in \mathbb{R}^n$
Tant que
  Calculer $g_k \in \partial f(x_k)$
  $x_{k+1} = \Pi_{\mathcal{C}}(x_k - \alpha_k g_k)$
Fin Tant que

<u>Q?</u> Convergence?

Soit $z_{k+1} = x_k - \alpha_k g_k \in \mathbb{R}^n$ avec $g_k \in \partial f(x_k)$ $\alpha_k > 0$

$$\|z_{k+1} - x^*\| = \|x_k - \alpha_k g_k - x^*\|^2$$
$$\leq \|x_k - x^*\| - 2\alpha_k (f(x_k) - f(x^*)) + \alpha_k^2 \|g_k\|^2$$
car $g_k \in \partial f(x_k)$

Or $\|x_{k+1} - x^*\|^2 = \|\Pi_{\mathcal{C}}(z_{k+1}) - \Pi_{\mathcal{C}}(x^*)\|^2$

car $x^* \in \mathcal{C} \Rightarrow \Pi_{\mathcal{C}}(x^*) = x^*$

par définition de x_{k+1}

$\leq \|z_{k+1} - x^*\|^2$ car $\Pi_{\mathcal{C}}$ 1-lipschitzienne

$\leq \|x_k - x^*\| - 2\alpha_k (f(x_k) - f(x^*)) + \alpha_k^2 \|g_k\|^2$

On retrouve les mêmes résultats de convergence que l'algorithme de sous-gradient.

Notamment $o\left(\frac{1}{\varepsilon^2}\right)$ itérations pour atteindre $|f_{best}^k - f(x^*)| < \varepsilon$

Rem: $\Pi_{\mathcal{C}}(x)$ peut être difficile à calculer en pratique selon ce qu'est $\mathcal{C}$.

3) Cas particulier : contraintes convexes d'inégalité

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} f(x) \\ \text{tq } \forall i \in [\![1,p]\!] \ f_i(x) \leq 0 \end{cases}$$
avec f et $(f_i)_{i \in [\![1,p]\!]}$ convexe sur $\mathbb{R}^n$

En posant $\mathcal{C} = \{x \in \mathbb{R}^n \text{ tq } \forall i \in [\![1,p]\!] \ f_i(x) \leq 0\}$

$\Rightarrow \min\limits_{x \in \mathcal{C}} f(x)$ avec $\mathcal{C}$ convexe : $(x,y) \in \mathcal{C}^2 \quad \lambda \in [0,1]$

$\forall i \in [\![1,p]\!] \ f_i(\lambda x + (1-\lambda) y) \leq \lambda f_i(x) + (1-\lambda) f_i(y)$

par convexité de f_i

≤ 0 car $\lambda \in [0,1]$ et $(x,y) \in \mathcal{C}^2$

Il est possible d'utiliser l'algorithme de sous-gradient projeté, mais $\Pi_{\mathcal{C}}$ peut être difficile à calculer

Q? Autre stratégie?
Respect des contraintes à chaque itération ?

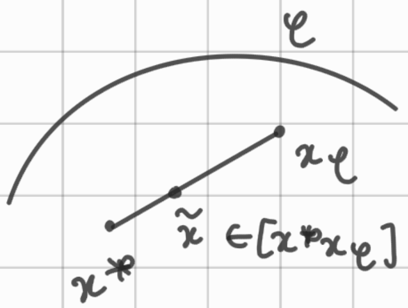
Algo: $x_0 \in \mathbb{R}^n$
Tant que
calculer $g_k \in \mathbb{R}^n$ tq
$$\begin{cases} g_k \in \partial f(x_k) \text{ si } x_k \in \mathcal{C} \\ g_k \in \partial f_i(x_k) \text{ avec } i \in [\![1,p]\!] \\ \qquad \text{tq } f_i(x_k) > 0 \end{cases}$$
$$x_{k+1} = x_k - \alpha_k g_k$$
Fin Tant que

On pose $f_{best}^k = \min_{j \in [\![1,k]\!]} \{ f(x_j) \text{ avec } x_j \in \mathcal{C} \}$

On suppose de plus que
$\exists x_{\mathcal{C}} \in \mathbb{R}^n$ tq $\forall i \in [\![1,p]\!]$ $f_i(x) < 0$
et $f(x_{\mathcal{C}}) \neq f(x^*)$ avec x^* une solution de (P)

avec (α_ℓ) tq $\begin{cases} \alpha_\ell > 0 \ \forall \ell \in \mathbb{N} \\ \sum \alpha_\ell \to +\infty \\ \sum \alpha_\ell^2 < +\infty \end{cases}$ alors $f_{best}^k \underset{k \to +\infty}{\to} f(x^*)$

Supposons le contraire
$\exists \varepsilon > 0$ tq $\forall N \in \mathbb{N}$ $\exists k \geq N$ tq $f_{best}^k \geq f(x^*) + \varepsilon$
En particulier $f_{best}^N \geq f_{best}^k \geq f(x^*) + \varepsilon$ et ce $\forall N \in \mathbb{N}$



Soit $\lambda \in [0,1]$, on pose $\tilde{x} = (1-\lambda) x^* + \lambda x_{\mathcal{C}}$

Par convexité de f

$$f(\tilde{x}) \leq (1-\lambda) f(x^*) + \lambda f(x_{\mathcal{C}})$$
$$\leq f(x^*) + \lambda (f(x_{\mathcal{C}}) - f(x^*))$$

par hypothèse $f(x_{\mathcal{C}}) > f(x^*)$

Avec $\lambda = \min\left(1, \dfrac{\varepsilon}{2(f(x_{\mathcal{C}}) - f(x^*))}\right)$

il vient $f(\tilde{x}) \leq f(x^*) + \dfrac{\varepsilon}{2}$

$\forall i \in [\![1,p]\!]$ $\quad f_i(\tilde{x}) \leq (1-\lambda) f_i(x^*) + \lambda f_i(x_{\mathcal{C}})$ par convexité de f_i

Or $f_i(x^*) \leq 0$ car $x^* \in \mathcal{C}$

$f_i(x_{\mathcal{C}}) > 0$ par def de $x_{\mathcal{C}}$

$$f_i(x) \leq \lambda f_i(x_{\mathcal{C}})$$
$$\leq \lambda \max_{j \in [\![1,p]\!]} f_j(x_{\mathcal{C}})$$
$$\leq -\mu$$

avec $\mu = \lambda \max_{j \in [\![1,p]\!]} f_j(x_{\mathcal{C}}) > 0$

d'où $\exists (\tilde{x}, \mu) \in \mathcal{C} \times \mathbb{R}_+^*$ tq

$$\begin{cases} f(\tilde{x}) \leq f(x^*) + \dfrac{\varepsilon}{2} \\ f_i(\tilde{x}) \leq -\mu \quad \forall i \in [\![1,p]\!] \end{cases}$$

$\forall k \geqslant \mathbb{N}$

$$\|x_{k+1} - \tilde{x}\|^2 = \|x_k - \alpha_k g_k - \tilde{x}\|^2$$
$$= \|x_k - \tilde{x}\|^2 - 2\alpha_k g_k^T (x_k - \tilde{x}) + \alpha_k^2 \|g_k\|^2$$

<u>1^{er} cas :</u> $x_k \in \mathcal{C}$ alors $g_k \in \partial f(x_k)$

et $\|x_{k+1} - \tilde{x}\| \leq \|x_k - \tilde{x}\|^2 - 2\alpha_k (f(x_k) - f(\tilde{x})) + \alpha_k^2 \|g_k\|^2$

or $x_k \in \mathcal{C} \Rightarrow f(x_k) \geqslant f_{best}^k$

$\geqslant f(x^*) + \varepsilon$ par hypothèse

de plus $f(\tilde{x}) \leq f(x^*) + \frac{\varepsilon}{2}$

$\Rightarrow \|x_{k+1} - \tilde{x}\|^2 \leq \|x_k - \tilde{x}\|^2 - \varepsilon \alpha_k + \alpha_k^2 \|g_k\|^2$

$2^{ème}$ cas : $\exists i \in [\![1, p]\!]$ tq $f_i(x_k) > 0$
alors $g_k \in \partial f_i(x_k)$

et $\|x_{k+1} - \tilde{x}\|^2 \leq \|x_k - \tilde{x}\|^2 - 2\alpha_k (f_i(x_k) - f_i(\tilde{x})) + \alpha_k^2 \|g_k\|^2$

Or $f_i(x_k) > 0$ et $f_i(\tilde{x}) \leq -\mu$

$\Rightarrow \|x_{k+1} - \tilde{x}\|^2 \leq \|x_k - \tilde{x}\|^2 - 2\mu\alpha_k + \alpha_k^2 \|g_k\|^2$

En posant $\delta = \min(2\mu, \varepsilon) > 0$
alors $\forall x_k \in \mathbb{R}^n$

$$\|x_{k+1} - \tilde{x}\|^2 \leq \|x_k - \tilde{x}\|^2 - \delta \alpha_k + \alpha_k^2 \|g_k\|^2$$

Par récurrence $\|x_{k+1} - \tilde{x}\|^2 \leq \|x_1 - \tilde{x}\|^2 - \delta \sum_{l=1}^{k} \alpha_l + \sum_{l=1}^{k} \alpha_l^2 \|g_l\|^2$

Avec la même hypothèse que pour l'algorithme du sous-gradient
$\exists a > 0$ tq $\forall \ell \in \mathbb{N}$ $\|g_\ell\| \leq a$

$$\sum_{\ell=1}^{k} \alpha_\ell \leq \frac{\|x_1 - \tilde{x}\|^2 + a^2 \sum_{\ell=1}^{k} \alpha_\ell^2}{\delta}$$

$$\leq \frac{\|x_1 - \tilde{x}\|^2 + a^2 \sum_{\ell=1}^{+\infty} \alpha_\ell^2}{\delta} \quad \text{car } \alpha_\ell > 0 \ \forall \ell \in \mathbb{N} \ \sum \alpha_\ell^2 < +\infty$$

à la limite $k \to +\infty$

$$+\infty \leq \frac{\|x_1 - \tilde{x}\|^2 + a^2 \sum_{\ell=1}^{+\infty} \alpha_\ell^2}{\delta}$$ d'où contradiction

d'où $f_{best}^k \to f(x^*)$

IV. Méthodes proximales

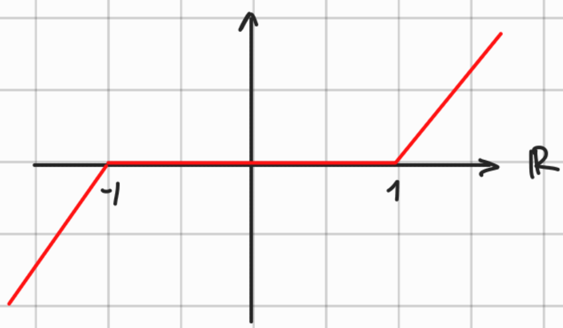
1) Fonction proximale

Soit $h : \mathbb{R}^n \to \mathbb{R}$ convexe. On définir la fonction proximale de h, notée $\text{prox}h$, par
$\forall x \in \mathbb{R}^n$ $\text{prox}h(x) = \underset{u \in \mathbb{R}^n}{\text{argmin}} \left(h(u) + \frac{1}{2} \|u - x\|_2^2 \right)$

ex: i) $h = 0$ alors $\text{prox}h(x) = x$ $\forall x \in \mathbb{R}^n$

ii) Soit $\mathcal{C}$ convexe de $\mathbb{R}^n$ Soit $h : \begin{array}{l} \mathbb{R}^n \to \mathbb{R} \\ x \mapsto \begin{cases} +\infty \text{ si } x \notin \mathcal{C} \\ 1 \text{ sinon} \end{cases} \end{array}$ alors $\text{prox}h(x) = \Pi_{\mathcal{C}}(x)$
↳ fermé non vide

iii) $\forall x \in \mathbb{R}^n$ $h(x) = \lambda \|x\|_1$ avec $\lambda > 0$ $\text{prox}h$ est appelée "seuillage doux", et est définie par : $\forall i \in [\![1, n]\!]$ $[\text{prox}h(x)]_i = \begin{cases} x_i + \lambda \text{ si } x_i < -\lambda \\ 0 \text{ si } |x_i| \leq \lambda \\ x_i - \lambda \text{ si } x_i > \lambda \end{cases}$



(Cf TD2)

2) Méthode du gradient proximal

On s'intéresse au problème d'optimisation (P) $\min_{x \in \mathbb{R}^n} f(x) = g(x) + h(x)$

avec g convexe dériv sur $\mathbb{R}^n$ et h convexe, potentiellement non-lisse (mais telle que $\operatorname{prox} h(x)$ est facilement calculable)

__Algo:__ $x_0 \in \mathbb{R}^n$ Tant que

$$x_{p+1} = \operatorname{prox}_{\alpha_p h}\big(\overbrace{x_p - \alpha_p \nabla g(x_p)}^{\text{itération de descente de gradient}}\big)$$

Fin Tant que

et (α_p) obtenue depuis : i) $\forall p \in \mathbb{N}\ \alpha_p = \alpha > 0$ (pas constant)

ii) Recherche linéaire

__ex:__ i) $h = 0$ $\quad x_{p+1} = x_p - \alpha_p \nabla g(x_p) \Rightarrow$ méthode de descente de gradient.

ii) $h(x) = \begin{cases} +\infty & \text{si } x \notin \mathcal{C} \\ 1 & \text{sinon} \end{cases}$ avec $\mathcal{C}$ convexe, fermé $\neq \emptyset$ de $\mathbb{R}^n$

$$x_{p+1} = \Pi_{\mathcal{C}}(x_p - \alpha_p \nabla g(x_p)) \Rightarrow \text{méthode du gradient projeté}$$

iii) $h(x) = \lambda \|x_1\| \Rightarrow$ "seuillage doux" pour minimiser $f(x) = g(x) + \lambda \|x\|_1$

__Rem:__ i) $x_{p+1} = \operatorname*{argmin}_{u \in \mathbb{R}^n} \left(\alpha_p h(u) + \frac{1}{2}\|u - x_p + \alpha_p \nabla g(x_p)\|_2^2\right)$

$$= \operatorname*{argmin}_{u \in \mathbb{R}^n} \left(h(u) + \frac{1}{2\alpha_p}\left[\|u - x_p\|^2 + 2\alpha_p \nabla g(x_p)(u - x_p) + \alpha_p^2 \|\nabla g(x_p)\|_2^2\right]\right)$$

← ne change pas le min mais seulement sa valeur

$$= \operatorname*{argmin}_{u \in \mathbb{R}^n} \left(h(u) + g(x_p) + \nabla g(x_p)^T(u - x_p) + \frac{1}{2\alpha_p}\|u - x_p\|_2^2 + \cancel{\frac{\alpha_p}{2}\|\nabla g(x_p)\|_2^2}\right)$$

$$= \operatorname*{argmin}_{u \in \mathbb{R}^n} \Big(h(u) + \underbrace{g(x_p) + \nabla g(x_p)^T(u - x_p) + \frac{1}{2\alpha_p}\|u - x_p\|_2^2}_{\text{modèle quadratique "dégénéré" de } g \text{ en } x_p}\Big)$$

ii) $u = \operatorname{prox} h(x) \Leftrightarrow 0 \in \partial \varphi(u)$ avec $\varphi(u) = h(u) + \frac{1}{2}\|u - x\|_2^2$

$\Leftrightarrow 0 \in \partial h(u) + \{u - x\}$

$\Leftrightarrow x - u \in \partial h(u)$

iii) la méthode du gradient proximal peut se réécrire

$$x_{p+1} = x_p - \alpha_p G_{\alpha_p}(x_p) \quad \text{avec} \quad G_\alpha(x) = \frac{1}{\alpha}\left[x - \text{prox}_{\alpha h}(x - \alpha \nabla g(x))\right]$$

En effet, $x_{p+1} - \alpha_p G_{\alpha_p}(x_p) = \text{prox}_{\alpha_p h}(x_p - \alpha_p \nabla g(x_p))$

D'après ii) $x_p - \alpha_p \nabla g(x_p) - x_p + \alpha_p G_{\alpha_p}(x_p) \in \partial \alpha_p h(x_p - \alpha_p G_{\alpha_p}(x_p))$

$$\Leftrightarrow \alpha_p\left[G_{\alpha_p}(x_p) - \nabla g(x_p)\right] \in \partial \alpha_p h(x_p - \alpha_p G_{\alpha_p}(x_p))$$

$$\Leftrightarrow \left[G_{\alpha_p}(x_p) - \nabla g(x_p)\right] \in \partial h(x_p - \alpha_p G_{\alpha_p}(x_p))$$

$$\Leftrightarrow G_{\alpha_p}(x_p) \in \partial h(x_p - \alpha_p h_{\alpha_p}(x_p)) + \{\nabla g(x_p)\}$$

De plus, $G_{\alpha_p}(x_p) = 0 \Leftrightarrow 0 \in \underbrace{\partial h(x_p - \alpha_p h_{\alpha_p}(x_p)) + \{\nabla g(x_p)\}}_{\partial f(x_p)}$ où $f(x) = g(x) + h(x)$

avec g dérivable convexe sur $\mathbb{R}^n$

$$\Leftrightarrow x_p \text{ minimum de } f.$$

D'où tout point fixe de la suite des itérés du gradient proximal est un minimum de f.

3) Recherche linéaire (choix des (α_p)) et convergence de l'algorithme

<u>Algo:</u> Recherche linéaire : backtracking + condition d'arrêt modifiée

Soient $\alpha_0 > 0$, $\beta \in]0;1[$

Tant que

$\alpha_{\ell+1} = \beta \alpha_\ell$

critère d'arrêt STOP si

$$g(x_p - \alpha_\ell G_{\alpha_\ell}(x_p)) \leq g(x_p) - \alpha_\ell G_{\alpha_\ell}(x_p)^T \nabla g(x_p) + \frac{\alpha_\ell}{2} \|G_{\alpha_\ell}(x_p)\|_2^2$$

Fin tant que

<u>Rem</u> : $G_{\alpha_\ell}(x_p)$ à calculer pour toute nouvelle valeur de α_ℓ

<u>prop</u> On suppose que ∇g est L-lipschitzienne avec $L > 0$:

$$\forall (x,y) \in (\mathbb{R}^n)^2 \quad \|\nabla g(x) - \nabla g(y)\|_2^2 \leq L\|x-y\|_2$$

Alors l'algorithme de recherche linéaire s'arrête sur $\alpha_\ell \geq \min\left(\alpha_0, \frac{\beta}{L}\right) := \alpha_{min}$

<u>preuve:</u> ∇g L- lipschitzien $\Rightarrow \forall (x,y) \in (\mathbb{R}^n)^2$

$$g(y) \leq g(x) + \nabla g(x)^T(y-x) + \frac{L}{2}\|y-x\|_2^2 \quad \text{(admis)}$$

D'où $g(x_p - \alpha_\ell G_{\alpha_\ell}(x_p)) \leq g(x_p) - \alpha_\ell \nabla g(x_p)^T G_{\alpha_\ell}(x_p) + \frac{L}{2}\alpha_\ell^2 \|G_{\alpha_\ell}(x_p)\|_2^2$

la condition est valide $\forall \alpha_\ell \in]0; \frac{1}{L}]$

$\Rightarrow$ à l'arrêt de la recherche linéaire $\alpha_\ell \geq \alpha_{min}$

(suite strictement décroissante qui tend vers 0)

<u>prop:</u> On suppose ∇g L-lipschitzien

À l'arrêt de la recherche linéaire ($\to (\alpha_p, G_{\alpha_p}(x_p))$)

$$\forall z \in \mathbb{R}^n \quad f(x_p - \alpha_p G_{\alpha_p}(x_p)) \leq f(z) + G_{\alpha_p}(x_p)^T(x_p - z) - \frac{\alpha_p}{2}\|G_{\alpha_p}(x_p)\|_2^2$$

<u>preuve:</u> Par définition du critère d'arrêt de la recherche linéaire :

$$g(x_p - \alpha_p G_{\alpha_p}(x_p)) \leq g(x_p) - \alpha_p G_{\alpha_p}(x_p)^T \nabla g(x_p) + \frac{\alpha_p}{2}\|G_{\alpha_p}(x_p)\|_2^2$$

$$\leq g(x_p) - \alpha_p G_{\alpha_p}(x_p)^T(G_{\alpha_p}(x_p) - \nabla g(x_p)) - \frac{\alpha_p}{2}\|G_{\alpha_p}(x_p)\|_2^2$$

D'après Rem iii) $G_{\alpha_p}(x_p) - \nabla g(x_p) \in \partial h(x_p - \alpha_p G_{\alpha_p}(x_p))$

d'où $\forall z \in \mathbb{R}^n \quad h(z) \geq [G_{\alpha_p}(x_p) - \nabla g(x_p)]^T(z - x_p + \alpha_p G_{\alpha_p}(x_p))$

$$+ h(x_p - \alpha_p G_{\alpha_p}(x_p))$$

De plus $g(x_p - \alpha_p G_{\alpha_p}(x_p)) + h(x_p - \alpha_p G_{\alpha_p}(x_p))$

$$\leq g(x_p) + h(x_p - \alpha_p G_{\alpha_p}(x_p)) + \alpha_p G_{\alpha_p}(x_p)^T(G_{\alpha_p}(x_p) - \nabla g(x_p))$$

$$- \frac{\alpha_p}{2}\|G_{\alpha_p}(x_p)\|_2^2$$

Or $h(x_p - \alpha_p G_{\alpha_p}(x_p)) \leq h(z) + [G_{\alpha_p}(x_p) - \nabla g(x_p)]^T(x_p - \alpha_p G_{\alpha_p}(x_p) - z)$

$$\Rightarrow f(x_p - \alpha_p G_{\alpha_p}(x_p)) \leq g(x_p) + h(z) + \alpha_p G_{\alpha_p}(x_p)^T(h_{\alpha_p}(x_p) - \nabla g(x_p))$$

$$+ [G_{\alpha_p}(x_p) - \nabla g(x_p)]^T(x_p - \alpha_p G_{\alpha_p}(x_p) - z) - \frac{\alpha_p}{2}\|G_{\alpha_p}(x_p)\|_2^2$$

$$f(x_p - \alpha_p G_{\alpha_p}(x_p)) \leq g(x_p) + h(z) + [G_{\alpha_p}(x_p) - \nabla g(x_p)]^T (x_p - z) - \frac{\alpha_p}{2} \|G_{\alpha_p}(x_p)\|_2^2$$

Or g convexe et dérivable sur $\mathbb{R}^n$

D'où $\forall z \in \mathbb{R}^n \quad g(z) \geq \nabla g(x_p)^T (z - x_p) + g(x_p)$

$$\Rightarrow g(x_p) \leq g(z) + \nabla g(x_p)^T (x_p - z)$$

Finalement $f(x_p - \alpha_p G_{\alpha_p}(x_p)) \leq g(z) + h(z) + \nabla g(x_p)^T (x_p - z)$
$+ [G_{\alpha_p}(x_p) - \nabla g(x_p)]^T [x_p - z] - \frac{\alpha_p}{2} \|G_{\alpha_p}(x_p)\|_2^2$

$$\Rightarrow f(x_p - \alpha_p G_{\alpha_p}(x_p)) \leq f(z) + G_{\alpha_p}(x_p)^T (x_p - z) - \frac{\alpha_p}{2} \|G_{\alpha_p}(x_p)\|_2^2 \quad \forall z \in \mathbb{R}^n$$

$$\operatorname{prox} h(x) = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left(h(u) + \frac{1}{2} \|u - x\|_2^2 \right)$$

De plus, $G_{\alpha_p}(x_p) = 0 \Leftrightarrow x_p$ minimum de f pour $\mathbb{R}^n$

<u>convergence</u> : On suppose ∇g est L-lipschitzien

On suppose également que $\forall p \in \mathbb{N}$ $\alpha_p = \frac{1}{L}$ ou α_p est obtenu depuis l'algorithme de recherche linéaire.

$\forall z \in \mathbb{R}^n \quad f(x_{p+1}) \leq f(z) + G_{\alpha_p}(x_p)^T (x_p - z) - \frac{\alpha_p}{2} \|G_{\alpha_p}(x_p)\|_2^2$

Avec $z = x_p \quad f(x_{p+1}) \leq f(x_p) - \frac{\alpha_p}{2} \|G_{\alpha_p}(x_p)\|_2^2$

Si x_p n'est pas un minimum de f $G_{\alpha_p}(x_p) \neq 0 \Rightarrow f(x_{p+1}) < f(x_p)$

$\Rightarrow$ méthode de descente

Avec $z = x^*$

$$f(x_{p+1}) \leq f(x^*) + G_{\alpha_p}(x_p)^T (x_p - x^*) - \frac{\alpha_p}{2} \|G_{\alpha_p}(x_p)\|_2^2$$

$$\Rightarrow 0 \leq f(x_{p+1}) - f(x^*) \leq G_{\alpha_p}(x_p)^T (x_p - x^*) - \frac{\alpha_p}{2} \|G_{\alpha_p}(x)\|_2^2$$

or $\|x_p - x^*\|_2^2 - \|x_{p+1} - x^*\|_2^2 = \|x_p - x^*\|_2^2 - \|x_p - \alpha_p G_{\alpha_p}(x_p) - x^*\|_2^2$

$$= 2\alpha_p G_{\alpha_p}(x_p)^T (x_p - x^*) + \alpha_p^2 \|G_{\alpha_p}(x_p)\|_2^2$$

$$= 2\alpha_p \left[G_{\alpha_p}(x_p)^T (x_p - x^*) - \frac{\alpha_p}{2} \|G_{\alpha_p}(x_p)\|_2^2 \right]$$

$$\Rightarrow G_{\alpha_p}(x_p)^T(x_p - x^*) - \frac{\alpha_p}{2}\|G_{\alpha_p}(x_p)\|_2^2 = \frac{1}{2\alpha_p}\left(\|x_p - x^*\|_2^2 - \|x_{p+1} - x^*\|_2^2\right)$$

Et $0 \leq f(x_{p+1}) - f(x^*) \leq \frac{1}{2\alpha_p}\left(\|x_p - x^*\|_2^2 - \|x_{p+1} - x^*\|_2^2\right)$ (*)

D'où $0 \leq \sum_{l=0}^{p}(f(x_{p+1}) - f(x^*)) \leq \frac{1}{2}\sum_{l=0}^{p}\frac{1}{\alpha_l}\left(\|x_l - x^*\|^2 - \|x_{l+1} - x^*\|_2^2\right)$

(x^* minimum de f sur $\mathbb{R}^n \Leftrightarrow \forall z \in \mathbb{R}^n$ $f(z) \geq f(x^*)$)

En posant $f_{best}^{p} = \min_{l \in [\![0,p]\!]} f(x_l)$

$$\Rightarrow 0 \leq p\left(f_{best}^{p} - f(x^*)\right) \leq \frac{1}{2}\sum_{l=0}^{p-1}\frac{1}{\alpha_l}\left(\|x_l - x^*\|_2^2 - \|x_{l+1} - x^*\|_2^2\right)$$

En posant $\alpha^* = \begin{cases} \frac{1}{L} & \text{si stratégie de pas constant} \\ \alpha_{min} & \text{si recherche linéaire} \end{cases}$

$$\Rightarrow 0 \leq p\left(f_{best}^{p} - f(x^*)\right) \leq \frac{1}{2\alpha^*}\underbrace{\sum_{l=0}^{p-1}(\underbrace{\|x_l - x^*\|^2 - \|x_{l+1} - x^*\|^2}_{\geq 0 \text{ par } (*)})}_{= \|x_0 - x^*\|_2^2 - \underbrace{\|x_p - x^*\|_2^2}_{\geq 0} \;\leq\; \|x_0 - x^*\|_2^2}$$

Soit $R \geq \|x_0 - x^*\|_2$

Alors $0 \leq f_{best}^{p} - f(x^*) \leq \frac{R^2}{2\alpha^*}\frac{1}{p} \xrightarrow[p \to +\infty]{} 0$ d'où $f_{best}^{p} \to f(x^*)$

On atteint $0 \leq f_{best}^{p} - f(x^*) \leq \varepsilon$ avec $\varepsilon > 0$

au plus tard si $p \geq \left(\frac{2\alpha^*}{R^2}\right)\frac{1}{\varepsilon}$

$\Rightarrow O(\frac{1}{\varepsilon})$ itérations pour atteindre la précision ε sur la valeur de $f(x^*)$

4) Méthode accélérée du gradient proximal

Sous les mêmes hypothèses que la méthode du gradient proximal, algorithme FISTA pour Fast Iterative Schrinkage-Thresholding algorithm.

$x_0 \in \mathbb{R}^n$

$y_0 = x_0$

Tant que

$$x_p = \text{prox}_{\alpha_p h}\left(y_{p-1} - \alpha_p \nabla g(y_{p-1})\right)$$

$$y_p = x_p + \frac{p-1}{p+2}(x_p - x_{p-1})$$

Fin tant que

Algo de recherche linéaire:

$t_0 = \alpha_{p-1}$, $\beta \in]0;1[$

Tant que

$t_\ell = \beta t_{\ell-1}$

Arrêt si $g(y_{p-1} - t_\ell \nabla g(y_{p-1})) \leq g(y_{p-1}) - t_\ell G_{\alpha_p}(y_{p-1})^T \nabla g(y_{p-1}) + \frac{t_\ell}{2} \|G_{t_\ell}(y_{p-1})\|_2^2$

Fin tant que

$\alpha_p = t_\ell$

Pour $\alpha_p = \frac{1}{L}$ ou issu de la recherche linéaire

$$0 \leq f^p_{best} - f(x^*) \leq \varepsilon \quad \text{avec } \varepsilon > 0$$

atteint en au plus $O\left(\frac{1}{\sqrt{\varepsilon}}\right)$ itérations.