CM
ON

hamza.ennaji
@univ-grenoble-alpes.fr

Cours ON : 1

Motivation et exemples :

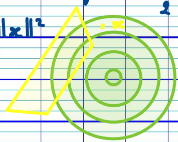
- Soit $f : \begin{cases} \mathbb{R}^n \longrightarrow \mathbb{R} \\ (P) \longrightarrow \min\limits_{x \in C} f(x) \end{cases}$ et $C \in \mathbb{R}^n$ on cherche $x^* \in C$

tq $\underset{\underset{f^*}{\shortparallel}}{f(x^*)} \leq f(x)$, $\forall x \in C$

ex: $\begin{cases} f(x) = \frac{1}{2} x^T A x + b^T x \\ C = \{x, Mx \leq d\} \end{cases}$

- Quand $A = I_d$ et $b = 0$ alors $f(x) = \frac{1}{2} \|x\|^2$

On cherche alors : $\min\limits_{x \in \{Mx \leq d\}} 1/2 \|x\|^2$



- Optimisation continue $\longrightarrow$ cadre lisse ou non lisse
- Optimisation discrète $\longrightarrow$ $C = \{0,1\}^n$ ou un graphe / réseau
- Contrôle optimal $\longrightarrow$ dimension infini
- Optimisation stochastique $\longrightarrow$ $f(x) = \underset{sup}{E}\left[\ell(x,y)\right]$

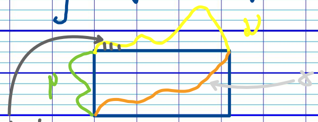
ex: $\Delta f = \sum\limits_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}$

$\begin{cases} -\Delta u = f, & \Omega \subset \mathbb{R}^n \\ u = 0, & \partial\Omega \end{cases}$ u résoud : $\min \frac{1}{2} \int_\Omega |\nabla u|^2 - \int_\Omega f u = E[u]$

$\begin{cases} u \in L^2 \text{ et } \nabla u \in L^2 \\ u_{|\partial\Omega} = 0 \end{cases}$

ex : Transport optimal de masse
Gaspart Monge
L. Kantorovich

$$\min \int |x-y|^2 \, d\gamma(x,y) \quad \text{où } \gamma \in \Gamma(\rho,\nu)$$



discrétisation par somme de Dirac → PL

$$\Gamma(\rho,\nu) = \{ P \in \mathbb{R}_+^{m\times m},\ P\mathbb{1}_m = \rho \text{ et } P^T\mathbb{1}_m = \nu \}$$

$$\min_{P\in\Gamma(\rho,\nu)} \langle C,P\rangle = \sum_{ij} c_{ij}P_{ij}$$

ex : Traitement d'images
Étant donné une image bruitée u_b, on cherche une image débruitée u

$$\min_u \frac{1}{2}\|u-u_b\|^2 + \lambda \underbrace{T_V(u)}_{=\int_\Omega |\nabla u|} = F(u) + G(\nabla u)$$

ex : Apprentissage : données (a_i, y_i) — a_i : description (étudiants / cellules) ; y_i : attribut (note, réponse) (normal, anormal)

– Objectif : trouver une fonction Φ prédiction qui associe approximativement les a_i aux y_i pour pouvoir prendre un nouvel ensemble de données $d = (a_i)$ pour prédire $z = (z_i)$.

Rappels :

1 - Topologie :

Définition :

- La boule fermée | ouverte centrée en $c \in \mathbb{R}^n$ de rayon $r > 0$ est définie par :

$$\bar{B}(c,r) = \{ x \in \mathbb{R}^n ,\ \|x - c\| \leq r \}$$

Définition : intérieur

- Soit $\mathcal{U} \subseteq \mathbb{R}^n$. On dit que $c \in \mathcal{U}$ est un point intérieur de $\mathcal{U}$ s'il existe $r > 0$ tq $B(c,r) \subseteq \mathcal{U}$

- On définit $\text{int}(\mathcal{U}) := \mathring{\mathcal{U}}$ par $\{ x \in \mathcal{U},\ B(x,r) \subseteq \mathcal{U},\ \text{pour } r > 0 \}$

ex : $\text{int}(\bar{B}(0,1)) = B(0,1)$
$\text{int}([0,1]) = (0,1)$
$\text{int}(\mathbb{R}_+) = \mathbb{R}_+^*$

Définition : ouverts

- $\mathcal{U} \subseteq \mathbb{R}^n$ est un ouvert si : $\forall x \in \mathcal{U},\ \exists r > 0,\ B(x,r) \subseteq \mathcal{U}$

Définition : fermés $\mathcal{U}$ fermé $\Leftrightarrow$ $\mathcal{U}^c$ ouvert

- $\mathcal{U}$ est fermé si : $\forall (x_n)_n \in \mathcal{U}$ tq $x_n \to x^*$, alors $x^* \in \mathcal{U}$

Proposition:

- Supposons que S est un fermé, alors $\forall \alpha \in \mathbb{R}$:

$$\left.\begin{array}{l} Lev(f,\alpha) = \{x \in S \,,\; f(x) \le \alpha\} \\ Con(f,\alpha) = \{x \in S \,,\; f(x) = \alpha\} \end{array}\right\} \text{ sont des fermés}$$

Définition:

- $D \subseteq \mathbb{R}^n$ est bornée s'il existe $r > 0$ tq $B(0,r) \supseteq D$
  D est compact s'il est fermé borné

Différentiabilité:

- Soit $f : S \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $x \in \dot{S}$ et $d \in \mathbb{R}^n \setminus \{0\}$.
  Si la limite $\lim\limits_{t \to 0^+} \dfrac{f(x+td) - f(x)}{t}$ existe, on l'appelle dérivée directionnelle.

dérivée de f dans la direction d

  On la note : $f'(x,d)$

- Pour $i = 1, \dots, n$ la dérivée de f dans la direction e_i est notée : $\dfrac{\partial f}{\partial x_i}(x)$

- Si $\partial_{x_i} f(x)$ existe pour tout x_i le gradient de f est :

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right)^T$$

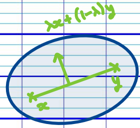
- On a, $f'(x;d) = \nabla f(x)^T . d$

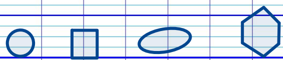
II - Convexité :

Définition :

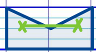
- Soit $C \subseteq \mathbb{R}^n$. On dit que C est convexe si :

$$\forall \lambda \in [0,1], \ \lambda x + (1-\lambda) y \in C, \ \forall x, y \in C$$



ex : 

Les boules : B, les hyperplans : $H = \{ x \in \mathbb{R}^n, \ a^T x = b \}$

contre-exemple :  est non convexe

Proposition : Intersection de convexe

- Soient $C_i \subseteq \mathbb{R}^n$ des ensembles convexes avec $i \in I$, alors :

$$C = \bigcap_{i \in I} C_i \text{ est convexe}$$

Théorème : Préservation de la convexité

1) Soient $C_1, \dots, C_m \subseteq \mathbb{R}^n$ des convexes de $\mathbb{R}^n$ et $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ alors :

$$C = \sum_{i=1}^{m} \lambda_i C_i \text{ est convexe}$$

2) Le produit : $\prod_{i=1}^{m} C_i$ est convexe

3) Si $C \subseteq \mathbb{R}^n$ est convexe et $A \in \mathbb{R}^{m \times n}$, alors :

$$A(C) = \{ Ax, \ x \in C \} \text{ est convexe}$$

Définition : Enveloppe convexe

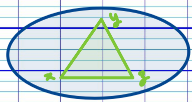
- Soient $(x_1, \dots, x_m) \in \mathbb{R}^n$. On appelle combinaison convexe des x_i :

$$\sum_{i=1}^{m} \lambda_i x_i \quad \text{avec} \quad \lambda_i \geq 0, \ \sum_{i=1}^{m} \lambda_i = 1$$

Théorème :

- Soit $C \subseteq \mathbb{R}^n$ un convexe et $x_1, \dots, x_m \in C$ alors :

$$\forall \lambda \in \Delta_m, \ \sum_{i=1}^{m} \lambda_i x_i \in C$$



avec $\Delta_m = \{ \lambda = (\lambda_1, \dots, \lambda_m), \ \lambda_i \geq 0 \text{ et } \sum_{i=1}^{m} \lambda_i = 1 \}$

Définition : Enveloppe

- Soit $S \in \mathbb{R}^n$. On appelle enveloppe convexe de $S := \mathrm{Conv}(S)$ l'ensemble de toutes les combinaisons convexes de S :

$$\mathrm{conv}(S) = \{ \sum_{i=1}^{m} \lambda_i x_i \ \text{ avec } x_i \in S_i, \ \lambda_i \geq 0 \text{ et } \sum_{i=1}^{m} \lambda_i = 1 \}$$

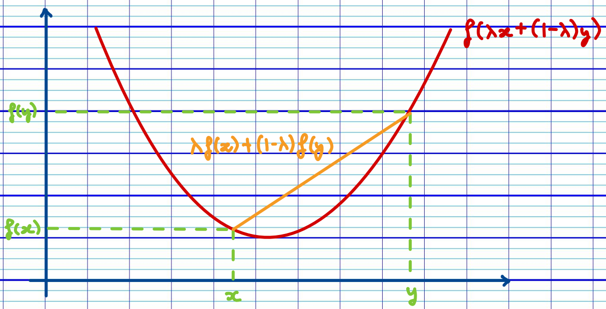
Définition : Fonctions convexes

- Soit $f : C \longrightarrow \mathbb{R}$. On dit que f est convexe si :

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y), \ \forall \lambda \in [0,1] \text{ et } \forall x, y \in C$$

– On dit que f est strictement convexe si :

$$f(\lambda x + (1-\lambda y)) < \lambda f(x) + (1-\lambda) f(y) \;,\; \forall \lambda \in (0,1) \text{ et } \forall x,y \in C$$



Définition : Concavité

– Si $-f$ est convexe, on dit que f est concave.

ex : les normes $\|.\|$ sont convexes.
les fonctions affines : $f(x) = a^T x + b$ sont convexes.

Théorème : Inégalité de Jensen

– Soit $f : C \longrightarrow \mathbb{R}$ convexe alors :

$$\forall\, x_1, \dots, x_m \in C \text{ et } \lambda \in \Delta_m$$

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

Théorème:

1. Si f est convexe et $\alpha \in \mathbb{R}^+$ alors αf est convexe
2. Si $f_1, \dots, f_m$ sont convexes sur $C \subseteq \mathbb{R}^n$ alors $\sum_{i=1}^{m} f_i$ est convexe sur C
3. Si $f_1, \dots, f_m$ sont convexes sur $C \in \mathbb{R}^n$ alors $f(x) = \sup_{i \in \{1, \dots, m\}} f_i(x)$ est convexe sur C.

Théorème : Caractérisation des fonctions convexes différentiables

Soit $f : C \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ différentiable alors :

$$f \text{ convexe} \Leftrightarrow \forall x, y \in C, \ f(y) \geq f(x) + \nabla^T f(x)(y-x)$$
$$\Leftrightarrow (\nabla f(x) - \nabla f(y))^T . (y-x) \geq 0$$

De plus, si f est C^2 alors f convexe $\Leftrightarrow \nabla^2 f(x) \geq 0$, $\forall x \in C$.

ex: $y^T \nabla^2 f(x) . y \geq 0 \Leftrightarrow \lambda_{min} (\nabla^2 f(x)) \geq 0$

plus petite valeur propre

Opérateur monotone: $\langle A(x) - A(y), x - y \rangle \geq 0$

ex: 1. $n = 1$, $f(x) = -\log(x)$ est convexe sur $\mathbb{R}_+^*$

2. $f(x) = \frac{1}{2} x^T A x$ et $A \in S^n$ (mat sym $A^T = A$)

Mq f convexe $\Leftrightarrow A \in S_+^n$

Remarque: Fonction à valeurs dans $\mathbb{R} \cup \{+\infty\}$

ex: $S \subseteq \mathbb{R}^n$, l'indicatrice de S : $\delta_S(x) = \begin{cases} 0 & \text{si } x \in S \\ +\infty & \text{sinon} \end{cases}$

– Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$, on note :
$\mathrm{dom}(f) = \{x \in \mathbb{R}^n,\ f(x) < +\infty\}$

– On dit que f est propre si $f \neq +\infty$, si $\exists x_n \in \mathbb{R}^n$ tq $f(x_n) < +\infty$

ex : 1. Fonction support (ou d'appui). Soit $S \subseteq \mathbb{R}^n$:

$$\sigma_S(x) = \sup_{y \in S} x^T y$$

2. $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$, $f^*(x) = \sup_{y \in \mathbb{R}^n} \{x^T y - f(y)\}$

Définition : Sous ensemble de niveau de fonctions (sub) level sets

– Soit $f : S' \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$. Le sous-ensemble de f de niveau $\alpha \in \mathbb{R}$ est :

$$\mathrm{Lev}(f, \alpha) = \{x \in S,\ f(x) \leq \alpha\}$$

Théorème :

– Soit $f : C \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction convexe, avec C convexe, alors :

$\forall \alpha \in \mathbb{R}$, $\mathrm{Lev}(f, \alpha)$ est convexe

⚠ Réciproque fausse

Définition :

– Une fonction $f : C \longrightarrow \mathbb{R}$ avec C convexe est dite quasi-convexe si :
$\forall \alpha \in \mathbb{R}$, $\mathrm{Lev}(f, \alpha)$ sont convexes.

ex: $f(x) = \sqrt{|x|}$

pour $\alpha < 0$: $\text{lev}(f,\alpha) = \emptyset$

pour $\alpha \geq 0$: $x \in \text{lev}(f,\alpha)$ si $\sqrt{|x|} \leq \alpha$ soit $|x| \leq \alpha^2$ ie $[-\alpha^2, \alpha^2]$

Optimisation convexe: $\min\limits_{x \in C} f(x)$ où $\begin{cases} f \text{ convexe} \\ C \text{ convexe} \end{cases}$

Rappel: Soit $x^* \in C$

On dit que x^* est un niveau local de f sur C s'il existe $r > 0$ tq:

$f(x^*) \leq f(x)$, $\forall x \in C \cap B(x^*, r)$

On dit que x^* est un minimum global de f sur C si :

$f(x^*) \leq f(x)$, $\forall x \in C$

Théorème:

- Soit $f : C \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ convexe avec C convexe. Soit $x^* \in C$, un minimum local de f sur C. Alors x^* est un minimum global de f.

Preuve:

- Soit $x^* \in C$ un minimum local de f sur C. Donc $\exists r > 0$ tq:

$f(x^*) \leq f(x)$, $\forall x \in C \cap B(x^*, r^*)$

Soit $y \in C \setminus \{x^*\}$, mq $f(x^*) \leq f(y)$

Soit $t \in [0,1]$, on considère : $\bar{x} = x^* + t(y - x^*)$ tq $\bar{x} \in C \cap B(x^*, r)$

Par convexité de C on a: $\bar{x} \in C \cap B(x^*, r)$ et par convexité de f:

$f(x^*) \leq f(\bar{x}) = f(x^* + t(y - x^*))$

$\leq tf(y) + (1-t) f(x^*)$

donc $t f(x^*) \leq t f(y)$ ie x^* est un minimum global de f.

- De même, on démontre :

Théorème:

- Soit $f : C \longrightarrow \mathbb{R}$ strictement convexe.
  Si $x^* \in C$ est un minimum strictement local alors x^* est un minimum strictement global.

Convexité de l'ensemble des minimiseurs:

Notation: On note $S := \underset{x \in C}{\operatorname{argmin}} f(x)$ l'ensemble des minimiseurs de f sur C.

Théorème:

- Soit $f : C \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction convexe, alors S est convexe.
  Si f est strictement convexe alors S contient au plus un élément.

Preuve:

- Si $S = \emptyset$, rien à démontrer.
- Sinon, on peut écrire $S = C \cap \operatorname{Lev}(f, f^*)$ qui est convexe par intersection de convexes.
  Autrement, soient $x, y \in S$ et $\lambda \in [0,1]$ alors:
  $$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$
  $$= \lambda f^* + (1-\lambda) f^* = f^*$$
  donc $\lambda x + (1-\lambda)y \in S$
  Si f est strictement convexe.
  Supposons qu'il existe $x, y \in S$ où $x \neq y$, alors $\frac{1}{2}x + \frac{1}{2}y \in C$
  et on a : $f(\frac{1}{2}x + \frac{1}{2}y) < \frac{1}{2}f(x) + \frac{1}{2}f(y) = f^*$ ce qui est absurde.

Projection orthogonal :

- Soit $C \subseteq \mathbb{R}^*$ un convexe fermé.

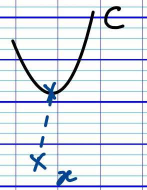
Définition :

- L'opérateur de projection orthogonale sur C est l'application (notée P_C ou Π_C)

$$P_C : \begin{cases} \mathbb{R}^n \longrightarrow C \\ x \longrightarrow P_C(x) = \underset{y \in C}{\arg\min}\{\|y-x\|^2\} \end{cases}$$

Théorème :

- Soit C un convexe fermé non vide dans $\mathbb{R}^n$, alors :



$\min_{y \in C} \|x-y\|^2$ admet une unique solution ($x^* = P_C(x)$)

ex : Pour $C = B(0,r)$, $r > 0$: $P_C(x) = \begin{cases} x & \text{si } \|x\| \leq r \\ r . \dfrac{x}{\|x\|} & \text{sinon} \end{cases}$

$\left\| r . \dfrac{x}{\|x\|} \right\| = r$

Théorème :

- Soit C un convexe fermé non vide de $\mathbb{R}^n$, alors pour $x \in \mathbb{R}^n$:

$$z = P_C(x) \text{ ssi } (x-z)^T(y-z) \leq 0 \, , \; \forall y \in C$$

Théorème :

- Soit C un convexe fermé non vide de $\mathbb{R}^n$.

1. $\forall v, w \in \mathbb{R}^n$, $(P_C(v) - P_C(w))^T (v-w) \geq \| P_C(v) - P_C(w)\|^2$
2. $\forall v, w \in \mathbb{R}^n$, $\|P_C(v) - P_C(w)\| \leq \|v-w\|$

CS

Inégalités :

- Young : $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$
- CS : $ab \leq |a||b|$

III - Descente : méthode du gradient

Motivation :

$\min\limits_{x \in \mathbb{R}^n} f(x)$

construire $x_0, x_1, \dots, x_n$

tq $f(x_n) \longrightarrow f^*$ valeur optimale

et si possible $x_n \xrightarrow[n \to \infty]{} x^*$ (dans un sens à préciser)

Définition : direction de descente

- Soit $f : \mathbb{R}^n \xrightarrow{C^1} \mathbb{R}$. On dit que $0 \neq d \in \mathbb{R}^n$ est une direction de descente de f en $x \in \mathbb{R}^n$ si $f'(x, d) = \nabla f(x)^T . d < 0$

Lemme :

- Soit $f \in C^1(\mathbb{R}^n)$ et $x \in \mathbb{R}^n$.

Si d est une direction de descente de f en x alors :

$\exists \varepsilon > 0$ tq $\forall t \in]0, \varepsilon]$: $f(x+td) < f(x)$

Schématiquement :

- Initialisation : $x_0 \in \mathbb{R}^n$
- Pour $k = 0, 1, 2 \dots$
  1. choisir d_k
  2. choisir t_k tq : $f(x_k + t_k d_k) < f(x_k)$
  3. prendre $x_{k+1} = x_k + t_k d_k$
  4. si un critère d'arrêt vérifié : $\|\nabla f(x_k)\| \leq \varepsilon$

     on renvoie x_{k+1}

- La recherche d'un pas t_k s'appelle la méthode de recherche linéaire, qui consiste à :

  $t_k \in \underset{t \geq 0}{\operatorname{argmin}}\, g(t) = f(x_k + t_k d_k)$
- Pas constant : $t_k = \bar{t}, \forall k$
- Recherche linéaire exacte : résoudre $(*)$
- Back-tracking : pour $s > 0$, $\alpha, \beta \in]0,1[$,

  tant que $f(x_k) - f(x_k + t_k d_k) < -\alpha_k t_k \nabla f(x_k)^T d_k$

  $t_k \leftarrow \beta t_k$

  donc $t_k = s\beta^{i_k}$ avec i_k le plus petit entier tq :

  $f(x_k) - f(x_k + s\beta^{i_k} d_k) \geq -\alpha s \beta^{i_k} \nabla f(x_k)^T d_k$

Exemple : recherche linéaire exacte

- $f(x) = x^T A x + 2b^T x + c$ avec $\begin{cases} A \in S_{++}^n \\ b \in \mathbb{R}^n \\ c \in \mathbb{R} \end{cases}$
- Soit d une direction de descente de f en $x \in \mathbb{R}^n$.

  On considère : $\min_{t \geq 0} g(t) = f(x + td)$

  On a : $g'(t) = 2(d^T A d)t + 2d^T \nabla f(x)$

  donc $g'(t) = 0 \Leftrightarrow t = \dfrac{-d^T \nabla f(x)}{d^T A d}$ car $A > 0$

Descente de gradient :

- Pour cette méthode, on prend $d_k = -\nabla f(x_k)$

  On a : $f'(x_k, -\nabla f(x_k)) = -\nabla f(x_k)^T \nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0$

  (pour x_k tq $\nabla f(x_k) \neq 0$)

Méthode de gradient :

- Donnée : tolérance $\varepsilon > 0$ (pour le critère d'arrêt)
- Initialisation : $x_0 \in \mathbb{R}^n$
- Étapes : pour $k = 0, 1, 2, \dots$

  1. choisir t_k avec une recherche linéaire sur :

     $$g(t) = f(t_k - t\nabla f(x_k))$$

  2. $x_{k+1} = x_k - t_k \nabla f(x_k)$
  3. arrêt si : $\|\nabla f(x_k)\| \leq \varepsilon$

Analyse de convergence :

- Soit $f \in C_L^{1,1}(\mathbb{R}^n)$ ie f est C^1 et $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$, $\forall x, y \in \mathbb{R}^n$

ex : 1. $f(x) = a^T x + b \in C_0^{1,1}$

2. $f(x) = x^T A x + 2b^T x + c$, $\nabla f(x) = 2(Ax+b)$

$$\|\nabla f(x) - \nabla f(y)\| = 2\|(Ax+b) - (Ay-b)\| \leq \underbrace{2\|A\|}_{L}\|x-y\|$$

Théorème :

- Soit $f \in C^2(\mathbb{R}^n)$, alors $f \in C_L^{1,1}(\mathbb{R}^n) \leftrightarrow \|\nabla^2 f(x)\| \leq L$, $\forall x \in \mathbb{R}^n$

Lemme :

- Soit $f \in C_L^{1,1}(\mathbb{R}^n)$, alors $\forall x \in \mathbb{R}^n$ et $\forall t > 0$:

$$f(x) - f(x - t\nabla f(x)) \geq t\left(1 - \frac{Lt}{2}\right) \|\nabla f(x)\|^2$$

- Pas constant tq $t_k = \bar{t} \in]0, \frac{2}{L}[$ on a :

$$f(x_k) - f(\underbrace{x_k - \bar{t}\nabla f(x_k)}_{x_{k+1}}) \geq \bar{t}\left(1 - \frac{L\bar{t}}{2}\right) \|\nabla f(x_k)\|^2$$

Pour avoir une plus grande borne, on regarde la fonction $t \longrightarrow t(1 - \frac{tL}{2})$ pour $t \in]0, 2/L[$

Donc on trouve $\bar{t} = \frac{1}{L}$.

On a donc :

$$\begin{aligned} f(x_k) - f(x_{k+1}) &= f(x_k) - f(x_k - \frac{1}{L}\nabla f(x_k)) \\ &\geq \frac{1}{L}\left(1 - \frac{L}{2}\frac{1}{L}\right)\|\nabla f(x_k)\|^2 \\ &= \frac{1}{2L}\|\nabla f(x_k)\|^2 \end{aligned}$$

- Recherche linéaire exacte : $x_{k+1} = x_k - t_k \nabla f(x_k)$

avec $t_k \in \underset{t \geq 0}{\operatorname{argmin}}\, g(t) = f(x_k - t\nabla f(x_k))$

Dans ce cas, par définition de t_k : $f(x_k - t_k\nabla f(x_k)) \leq f(x_k - \frac{1}{L}\nabla f(x_k))$

donc :

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\geq f(x_k) - f(x_k - \frac{1}{L}\nabla f(x_k)) \\ &\geq \frac{1}{2L}\|\nabla f(x_k)\|^2 \end{aligned}$$

- Pour le backtracking (avec $s > 0$ et $\alpha, \beta \in]0,1[$)

On choisit $t_k > \min\{s, \frac{2(1-\alpha)\beta}{L}\}$

Résumé :

- Soit $f \in C_L^{1,1}(\mathbb{R}^n)$ et $(x_n)_n$ la suite générée par la méthode de descente du gradient, alors, $f(x_k) - f(x_{k+1}) \geq M\|\nabla f(x_k)\|^2$

avec
$$M = \begin{cases} \bar{t}\left(1 - \frac{\bar{t}L}{2}\right) \\ \frac{1}{2L} \\ \alpha \min\{s, \frac{2(1-\alpha)\beta}{L}\} \end{cases}$$

Théorème : Convergence de gradients

- Soit $f \in C^{1,1}_L(\mathbb{R}^n)$ et $(x_n)_n$ suite générée par la méthode du gradient.
Supposons que f est bornée inférieurement, ie $\exists c \in \mathbb{R}$ tq $f(x) \geq c$ $\forall x \in \mathbb{R}^n$, alors :
  1. $(f(x_n))_n$ est décroissante et $f(x_{n+1}) < f(x_n)$ sauf si $\nabla f(x_n) = 0$
  2. $\nabla f(x_n) \xrightarrow[n \to +\infty]{} 0$

Démonstration :

1) On sait que $f(x_n) - f(x_{n+1}) \geq M \| \nabla f(x_n) \|^2 \geq 0$ donc quand $\nabla f(x_n) \neq 0$, $f(x_{n+1}) < f(x_n)$
2) Comme $(f(x_n))$ est décroissante et f bornée inf, alors $(f(x_n))$ est convergente, pour la suite $f(x_n) - f(x_{n+1}) \xrightarrow[n \to +\infty]{} 0$, donc d'après le calcul précédent au point 1) $\| \nabla f(x_n) \| \xrightarrow[n \to \infty]{} 0$

Théorème :

- Sous les mêmes hypothèses, on a $\forall n \geq 0$:

$$\min_{k=0,\dots,n} \| \nabla f(x_k) \| \leq \sqrt{\frac{f(x_0) - f^*}{M(n+1)}}$$

Rappels :

- Méthode de Newton pour chercher les zéros d'une fonction.
Soit $f : \mathbb{R} \to \mathbb{R}$, on cherche à résoudre $f(t) = 0$.
On regarde l'approximation d'ordre 1 de f autour d'un point $t_0 \in \mathbb{R}$ et
$\tilde{f}(t) = f(t_0) + f'(t_0)(t - t_0)$
donc $\tilde{f}(t) = 0 \Rightarrow t = t_0 - \dfrac{f(t_0)}{f'(t_0)}$

Itérativement, $t_{m+1} = t_m - \dfrac{f(t_m)}{f'(t_m)}$

De même, si $g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, pour résoudre $g(x) = 0_{\mathbb{R}^n}$ on remplace $g(x)$ par :
$\tilde{g}(x) = g(x_0) + \nabla g(x_0)(x - x_0)$ autour de $x_0 \in \mathbb{R}^n$
Donc, en supposant que $\det(\nabla g(x_0)) \neq 0$, on obtient que :
$x_{m+1} = x_m - (\nabla g(x_m))^{-1} g(x_m)$

- En optimisation, pour résoudre $\min\limits_{x \in \mathbb{R}^n} f(x)$, on considère le système linéaire $\nabla f(x) = 0$.
Pour rappel :
$\nabla f(x^*) = 0$ et f convexe revient à dire que x^* est optimal, on trouve donc :
$x_{m+1} = x_m - (\nabla^2 f(x_m))^{-1} \nabla f(x_m)$

Méthode de Newton :

- Soit $f : \mathbb{R}^n \to \mathbb{R}$ C^2 et $x_0 \in \mathbb{R}^n$.
On considère $\tilde{f}(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2} \langle \nabla^2 f(x_0)(x - x_0), x - x_0 \rangle$
La méthode de Newton, revient à générer la suite :
$x_{k+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}}\ \tilde{f}(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle \nabla^2 f(x_k)(x - x_k), x - x_k \rangle$
dans ce cas :
$\nabla \tilde{f}(x_{k+1}) = 0$ donne : $\nabla f(x_k) + \nabla^2 f(x_k)(x_{k+1} - x_k) = 0$
soit $x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$ où $\nabla^2 f(x_k) \in S^n_{++}$
donc la méthode de Newton est une méthode de descente avec comme direction :
$d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

Théorème : Convergence quadratique

- Soit $f \in C^2$ et supposons que :
  1. $\exists m > 0$ tq $\nabla^2 f(x) \succeq m I_n$, $\forall x \in \mathbb{R}^n$
  2. $\exists L > 0$ tq $\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L \|x - y\|$, $\forall x, y \in \mathbb{R}^n$

  Soit $(x_k)_k$ la suite générée par la méthode de Newton, alors :

  $\forall k \geq 0$, $\|x_{k+1} - x_k\|^2 \leq \frac{L}{2m} \|x_k - x^*\|^2$

Méthode du gradient projeté :

- On cherche $\min_{x \in C} f(x)$ et $C \subseteq \mathbb{R}^n$ convexe fermé.

Théorème :

- Soit $f : C \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ C^1 et $s > 0$, alors x^* est un point stationnaire de (P) ssi

  $x^* = P_C(x^* - s \nabla f(x^*))$

Remarque : on peut minimiser sur $\mathbb{R}^n$ avec : $\min_{x \in \mathbb{R}^n} \underbrace{f(x) + \mathbb{I}_C(x)}_{\tilde{f}}$

indicatrice matricielle

Gradien projeté :

- Soit $\epsilon > 0$.

  Initialisation : prendre $x_0 \in C$

  Étapes : 1. choisir t_k

  2. $x_{k+1} = P_C(x_k - t_k \nabla f(x_k))$

  3. si critère d'arrêt vérifié, on s'arrête et on renvoie x_{k+1}

Lemme :

- Soit $f \in C_L^{1,1}(C)$, avec C convexe fermé
$\forall x \in C$ et $\forall t \in \left(0, \frac{2}{L}\right)$, on a :

$$f(x) - f(P_c(x - t\nabla f(x))) \geq t\left(1 - \frac{Lt}{2}\right)\left\|\frac{1}{t}(x - P_c(x - t\nabla f(x)))\right\|^2$$

Démonstration:

- Utiliser le lemme de descente avec :
$y = P_c(x - t\nabla f(x))$ + propriété de projection

- Pour $M > 0$, on définit :
$G_M(x) = M\left(x - P_c\left(x - \frac{1}{M}\nabla f(x)\right)\right)$
Quand $C = \mathbb{R}^n$, $G_M(x) = \nabla f(x)$
Le lemme précédent devient : $f(x) - f(P_c(x - t\nabla f(x))) \geq t\left(1 - \frac{Lt}{2}\right)\overbrace{\|G_{1/t}(x)\|^2}^{\text{mesure d'optimalité}}$

Cas convexe:

- Soit $f \in C_L^{1,1}(C)$ convexe

Théorème:

- Sous ces conditions, la suite générée par la méthode du gradient projeté avec $t_k = t^* \in \left]0, \frac{1}{L}\right]$ vérifie :

1. $x_k \xrightarrow[k \to +\infty]{} x^* \in S = \underset{x \in C}{\operatorname{argmin}} f(x)$
2. $f(x_k) - f^* \leq \frac{1}{2t^* k}\|x_0 - x^*\|^2$

Conditions d'optimalité : contraintes affines

- Soit $(P) : \begin{cases} \min f(x) \\ a_i^T x \leq b_i,\ i=1,\dots,m \end{cases}$ avec $f \in C^1$ et $a_1,\dots,a_m \in \mathbb{R}^m$ et $b_1,\dots,b_m \in \mathbb{R}^m$.

Soit x^* un minimum local de (P).

Alors, $\exists \lambda_1,\dots,\lambda_m \geq 0$ tq : $\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i a_i = 0 \\ \lambda_i (a_i^T x^* - b_i) = 0,\ \forall i = 1,\dots,m \end{cases}$

- Il s'agit d'une condition nécessaire
- On a équivalence lorsque f est convexe :

x^* minimum de $f \leftrightarrow \exists \lambda_1,\dots,\lambda_m \geq 0$ tq on appelle λ_i le multiplicateur de Lagrange associé à la contrainte $a_i^T x \leq b_i$: $\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i a_i = 0 \\ \lambda_i (a_i^T x^* - b_i) = 0,\ \forall i = 1,\dots,m \end{cases}$

- On considère le problème : (P) $\begin{cases} \min f(x) \\ a_i^T x \leq b_i,\ i=1,\dots,m \\ c_j^T x = d_j,\ j=1,\dots,p \end{cases}$

Théorème : condition nécessaire

1) Si x^* est un minimum de (P) : $\exists \lambda_1,\dots,\lambda_m \geq 0$ et $\mu_1,\dots,\mu_p \in \mathbb{R}$ tq:

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i a_i + \sum_{j=1}^{p} \mu_j c_j = 0 \\ \lambda_i (a_i^T x^* - b_i) = 0,\ \forall i = 1,\dots,m \end{cases}$$

2) Condition suffisante lorsque f est convexe :

si f est convexe et x^* admissible (ie vérifie $\begin{cases} a_i^T x \leq b_i,\ i=1,\dots,m \\ c_j^T x = d_j,\ j=1,\dots,p \end{cases}$ tq $\exists \lambda_1,\dots,\lambda_m \geq 0$ et $\mu_1,\dots,\mu_p$

vérifiant $\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i a_i + \sum_{j=1}^{p} \mu_j c_j = 0 \\ \lambda_i (a_i^T x^* - b_i) = 0,\ \forall i = 1,\dots,m \end{cases}$

alors x^* est optimal pour (P)

Remarque: i. Les multiplicateurs de Lagrange associé à $a_i^T x \leq b$ sont toujours dans $\mathbb{R}_+$.

ii. Un point x^* vérifiant (C°) est dit point KKT (Karush-Kuhn-Tucker).

- Plus généralement, on considère (P): $\begin{cases} \min f(x) \\ g_i(x) \leq 0, \ i=1,\dots,m \\ h_j(x) = 0, \ j=1,\dots,p \end{cases}$ avec $f, g, h \in C^1$.

Définition:

- On appelle Lagrangien associé à (P) la fonction:

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- L'équivalent de l'équation $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i a_i + \sum_{j=1}^{p} \mu_j c_j = 0$, par le problème (P):

$$\begin{cases} \nabla_x L(x^*,\lambda,\mu) = 0 \\ \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) = 0 \end{cases}$$

KKT $\begin{cases} \nabla L = 0 \\ 2^{nd} \text{ pole } L = 0 \\ \text{contraintes} \end{cases}$ + convexe = mini

Remarque: le problème $\begin{cases} \min f(x) \\ a_i^T x \leq b_i \\ c_j^T x = d_j \end{cases}$ peut s'écrire de la forme: $\begin{cases} \min f(x) \\ Ax \leq b \\ Cx = d \end{cases}$ avec $A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix}$, $C = \begin{pmatrix} c_1^T \\ \vdots \\ c_p^T \end{pmatrix}$, $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ et $d = \begin{pmatrix} d_1 \\ \vdots \\ d_p \end{pmatrix}$

et son Lagrangien s'écrit: $L(x,\lambda,\mu) = f(x) + \lambda^T(Ax-b) + \mu^T(Cx-d)$

- **ex:** $\begin{cases} \min f(x,y,z) \quad \text{où } f(x,y,z) = \frac{1}{2}(x^2+y^2+z^2) \\ x+y+z = 3 \end{cases}$

- Le Lagrangien du problème s'écrit:

$$L(x,y,z,\lambda) = \frac{1}{2}(x^2+y^2+z^2) + \lambda(x+y+z-3)$$

- En écrivant les conditions KKT, on trouve:

$$\begin{cases} \nabla L(x,y,z,\lambda) = 0 \\ \lambda(x+y+z-3) = 0 \end{cases}$$, on a $\frac{\partial L}{\partial x} = x + \lambda$, $\frac{\partial L}{\partial y} = y + \lambda$, $\frac{\partial L}{\partial z} = z + \lambda$

donc $\nabla L = 0$ donne $x = y = z = -\lambda$

comme $x + y + z = 3$, on trouve $\lambda = -1$

le système : $\begin{cases} \frac{\partial L}{\partial x} = 0 \\ \frac{\partial L}{\partial y} = 0 \\ \frac{\partial L}{\partial z} = 0 \end{cases}$ donne $x = y = z = 1$

⚠ nous avons utiliser le fait que f est convexe, donc les conditions KKT sont suffisantes pour affirmer que $(x, y, z) = (1, 1, 1)$ est solution minimale.

Application : calcul de projection orthogonale

1) Projection sur un sous-espace affine, $y \in \mathbb{R}^n$:

$\min_{x \in C} f(x)$ avec $\begin{cases} f(x) = \|x - y\|^2 & , A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \text{ et } rg(A) = m \\ C = \{x \in \mathbb{R}^n, Ax = b\} \end{cases}$

Les conditions KKT sont nécessaires et suffisantes pour trouver $P_C(y)$.

Le Lagrangien du problème s'écrit :

$L(x, \lambda) = \|x - y\|^2 + 2\lambda^T(Ax - b)$

On a $\nabla_x L(x, \lambda) = 2(x - y) + 2A^T\lambda = 2x - 2(y - A^T\lambda)$

KKT : $\begin{cases} 2x - 2(y - A^T\lambda) = 0 \\ Ax = b \end{cases} \Rightarrow \begin{cases} x = y - A^T\lambda \\ Ax = b \end{cases}$

On remplace $x = y - A^T\lambda$ dans $Ax = b$ pour trouver $A(y - A^T\lambda) = b$ soit $\lambda = (AA^T)^{-1}(Ay - b)$

donc $P_C(y) = y - A^T\lambda = y - A^T(AA^T)^{-1}(Ay - b)$

- Cas particulier : $C = H = \{x, a^Tx = b\}$ $0 \neq a \in \mathbb{R}^n$ et $b \in \mathbb{R}$

On trouve : $P_H(y) = y - \frac{(a^Ty - b)}{\|a\|^2} a$.

2) Projection sur un demi-espace : $H^- = \{x \in \mathbb{R}^n, a^Tx \leq b\}$, $0 \neq a \in \mathbb{R}^n$ et $b \in \mathbb{R}$

On cherche à résoudre : $\min_{a^Tx \leq b} \|x - y\|^2$

Le Lagrangien du système s'écrit :

$L(x, \lambda) = \|x - y\|^2 + 2\lambda(a^Tx - b)$ avec $\lambda \geq 0$

Avec KKT : $\begin{cases} \nabla_x L(x,\lambda) = 0 & \rightarrow \ 2(x-y) + 2\lambda x = 0 \\ \lambda(a^T x - b) = 0 \\ a^T x \leq b \\ \lambda \geq 0 \end{cases}$

si $\lambda = 0$: $x = y$ donc $P_{H^-}(y) = y$ (car $a^T x \leq y$)

si $\lambda > 0$: on trouve $x = y - \lambda a$

on a : $a^T(y - \lambda a) = b \rightarrow a^T y - \lambda \|a\|^2 = b$ soit $\lambda = \frac{a^T y - b}{\|a\|^2} > 0$

Finalement : $P_{H^-}(y) = \begin{cases} y & \text{si } a^T y \leq b \\ y - \frac{a^T y - b}{\|a\|^2} & \text{sinon} \end{cases}$

3) Projection sur la boule : $\min\limits_{x \in \overline{B(0,r)} = \{x,\ \|x\|^2 \leq r^2\}} \|x - y\|^2$

$L(x,\lambda) = \|x - y\|^2 + 2\lambda(\|x\|^2 - r^2)$ avec $\lambda \geq 0$

On a $P_B(y) = \begin{cases} y & \text{si } \|y\|^2 \leq r^2 \\ r \cdot \frac{y}{\|y\|} & \text{sinon} \end{cases}$

IV - Sous-différentiel :

image, signal

Motivation : $\min\limits_{u} \frac{1}{2}\|u - d\|^2 + \lambda Tv(u)$ avec $Tv(u) = \|Du\|_1$

Théorème : différentiabilité des convexes

- Soit $f : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ convexe avec C convexe.

Soit $x_0 \in \mathring{C}$, alors $\exists \varepsilon > 0$ et $L > 0$ tq $B(x_0, \varepsilon) \subseteq C$ et $|f(x) - f(x_0)| \leq L\|x - x_0\|$, $\forall x \in B(x_0, \varepsilon)$

- Conséquence : si f est convexe sur C, alors f est différentiable pp sur C.

Définition :

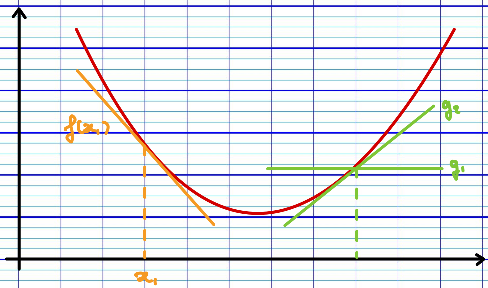
- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ propre (ie $-\infty < f$ et $dom(f) \neq \emptyset$)
- Soit $x \in dom(f)$

On dit que $g \in \mathbb{R}^n$ est un sous-gradient de f si :

$$f(y) \geq f(x) + \langle g, y-x \rangle, \quad \forall y \in \mathbb{R}^n$$

- On appelle sous-différentiel de f en x, l'ensemble des sous-gradients de f en x :

$$\partial f(x) = \{ g \in \mathbb{R}^n, \; f(y) \geq f(x) + \langle g, y-x \rangle, \; \forall y \in dom(f) \}$$



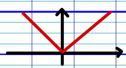
quand f est différentiable convexe et $f(y) \geq f(x) + \langle g, y-x \rangle$

- $y \longmapsto f(y) \geq f(x) + \langle g, y-x \rangle$ est une minorante affine de f en x.

Théorème :

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ propre. Alors, $\partial f(x)$ est un convexe fermé $\forall x \in \mathbb{R}^n$.

$$\partial f(x) = \bigcap_{y \in \mathbb{R}^n} \underbrace{\{ g : f(y) \geq f(x) + \langle g, y-x \rangle \}}_{H(y)}$$

- ex: $f(x) = |x|$ 

Pour $x > 0$: $f(x) = x$ et $\partial f(x) = \{+1\}$

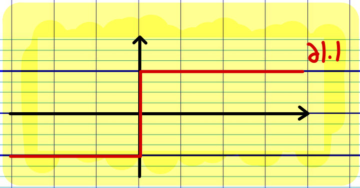
Pour $x < 0$: $f(x) = -x$ et $\partial f(x) = \{-1\}$

Soit $g \in \partial f(0)$, alors $\forall y \in \mathbb{R}$, $|y| \geq 0 + \langle g, y-0 \rangle$ ie $|y| \geq gy$

Pour $|g| \leq 1$ alors d'après l'inégalité de CS :

$\langle g, y \rangle \leq |g||y| \leq |y| \quad \forall y \in \mathbb{R}$

Inversement, $|g| = \max\limits_{|y| \leq 1} \langle g, y \rangle \leq \max\limits_{|y| \leq 1} |y| = 1$



- **Remarque:** $\partial f(x)$ est un ensemble qui peut-être $= \emptyset$

  On définit $\mathrm{dom}(\partial f) = \{x, \partial f(x) \neq \emptyset\}$

- ex: $f(x) = \begin{cases} -\sqrt{x} & x \geq 0 \\ +\infty & \text{sinon} \end{cases}$ avec $\mathrm{dom}(f) = \mathbb{R}^2$

  S'il existe $g \in \partial f(0)$ alors :

  $f(y) \geq f(0) + \langle g, y - 0 \rangle$, soit $-\sqrt{y} \geq yg \quad \forall y \in \mathbb{R}^+$

  Pour $y = 1$: $g \leq -1 < 0$

  Pour $y = \frac{1}{2g^2}$, on a : $-\sqrt{\frac{1}{2g^2}} \geq \frac{1}{2g^2} \cdot g = \frac{1}{2g}$

  donc $\frac{1}{2g^2} \leq \frac{1}{4g^2}$, ce qui est impossible

  donc $\partial f(0) = \emptyset$

Théorème:

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ propre, convexe et $\overset{\circ}{\mathrm{dom}(f)}$ alors $\partial f(x) \neq \emptyset$ et borné.

Proposition: $\partial f(x) \neq \emptyset \rightarrow$ convexité

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ propre tq $\mathrm{dom}(f)$ est convexe.

  Supposons que $\forall x \in \mathrm{dom}(f)$, $\partial f(x) \neq \emptyset$, alors f est convexe.

Théorème :

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ convexe propre et $x \in \overset{\circ}{\widehat{dom(f)}} = int(dom(f))$.
  Si f est différentiable en x, alors $\partial f(x) = \{\nabla f(x)\}$
  Inversement, si f admet un unique sous-gradient en x, alors f est différentiable en x et $\partial f(x) = \{\nabla f(x)\}$.

- ex : cône normal

- Soit $S \subset \mathbb{R}^n$ et $x \in \mathbb{R}^n$
  Soit $\delta_S(x) = \begin{cases} 0 & \text{si } x \in S \\ +\infty & \text{sinon} \end{cases}$
  Soit $x \in S$ alors $y \in \partial S(x)$:
  $$\delta_S(z) \geq \delta_S(x) + \langle y, z - x \rangle, \ \forall z \in S$$
  donc $\langle y, z - x \rangle \leq 0, \ \forall z \in S$
  donc $y \in \{ g \in \mathbb{R}^n, \langle g, z - x \rangle \leq 0, \forall z \in S \} = N_S(x)$

Sous-différentielle et optimisation :

Théorème : Fermat

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ convexe et propre.
  Alors, $x^* \in \underset{x \in \mathbb{R}^n}{\mathrm{Argmin}}\, f(x) \longleftrightarrow 0_{\mathbb{R}^n} \in \partial f(x^*)$

  $f(x) \geq f(x^*) + \langle 0, x - x^* \rangle$ et $\forall x \in dom(f)$

Théorème : CNS pour $\min_{x \in C} f(x)$ $\begin{cases} f \text{ convexe} \\ C \text{ convexe} \end{cases}$

- $x^* \in \underset{x \in C}{\mathrm{Argmin}}\, f(x) \Leftrightarrow 0 \in \partial f(x^*) + N_C(x^*)$

- Si f est différentiable dans x^*, $\partial f(x^*) = \{\nabla f(x^*)\}$

V - Algorithmes proximaux :

Définition : opérateur proximal

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$. On définit l'opérateur proximal de f par :

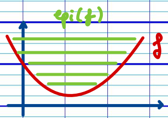
$$\mathrm{prox}_f(x) = \underset{y \in \mathbb{R}^n}{\mathrm{argmin}} \left\{ \frac{1}{2}\|x-y\|^2 + f(y) \right\}$$

Définition :

- On dit qu'une fonction convexe f est fermée si $\forall \alpha \in \mathbb{R}$, $\mathrm{Lev}(f, \alpha)$ est fermé.

De façon équivalente, son épigraphe :

$$\mathrm{epi}(f) = \{ (x,t) \in \mathbb{R}^{n+1},\ f(x) \leq t \text{ et } x \in \mathrm{dom}(f) \} \text{ est fermé.}$$



Théorème : 1er théorème du prox

- Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexe, propre et fermée.

Alors, $\mathrm{prox}_f(x)$ est un singleton $\forall x \in \mathbb{R}^n$.

Théorème :

tend vers l'infini à l'infini

- Si $y \longmapsto f(y) + \frac{1}{2}\|x-y\|^2$ est coércive alors :

$\forall x \in \mathbb{R}^n$, $\mathrm{prox}_f(x) \neq \emptyset$ et est un singleton.

- ex :

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

2) f affine, $f(x) = \langle a, x\rangle + b$ avec $a \in \mathbb{R}^n$ et $b \in \mathbb{R}$

$$\text{prox}_f(x) = \underset{y\in\mathbb{R}^n}{\text{argmin}} \{ \frac{1}{2}\|x-y\|^2 + \langle a, x\rangle + b \}$$
$$= \underset{y\in\mathbb{R}^n}{\text{argmin}} \{ \frac{1}{2}\|y-(x-a)\|^2 + \underbrace{\langle a, x\rangle + b - \frac{1}{2}\|a\|^2}_{\text{"}c\text{"}} \}$$
$$= x - a$$

3) f quadratique, $f = \frac{1}{2}x^TAx + b^Tx + c$ avec $A \in S_+^n$, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$

$$\text{prox}_f(x) = \underset{y\in\mathbb{R}^n}{\text{argmin}} \{ \underbrace{\frac{1}{2}\|x-y\|^2 + \frac{1}{2}y^TAy + b^Ty + c}_{g(y)} \}$$

On a : $\nabla g(y) = y - x + Ay + b$

donc $\nabla g(y) = 0 \Leftrightarrow y = (I+A)^{-1}(x-b)$

ie $\text{prox}_f(x) = (I+A)^{-1}(x-b)$

4) $f(x) = \lambda|x|$ avec $x \in \mathbb{R}$ et $\lambda > 0$

$$\text{prox}_f(x) = \underset{y\in\mathbb{R}}{\text{argmin}} \{ \underbrace{\frac{1}{2}|y-x|^2 + \lambda|y|}_{g(y)} \}$$

$$g(y) = \begin{cases} g_1(y) = \frac{1}{2}(y-x)^2 + \lambda y \\ g_2(y) = \frac{1}{2}(y-x)^2 - \lambda y \end{cases}$$

- Si le minimum est atteint en $y \gtrless 0$:

$g_1'(y) = 0 \Leftrightarrow y = x \mp \lambda$

donc $\text{prox}_f(x) = (|x| - \lambda)_+ . \text{sg}(x) = \mathcal{T}_\lambda(x)$

Règles de calcul :

- Soit $f : \prod_{i=1}^{m} \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$

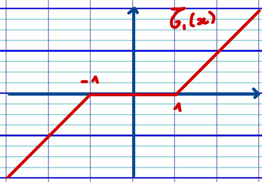
$f(x_1, \dots, x_m) = \sum_{i=1}^{m} f_i(x_i)$, $\forall x_i \in \mathbb{R}^n$ et $i = 1, \dots, m$

$\text{prox}_f(x) = \text{prox}_{f_1}(x_1) \times \dots \times \text{prox}_{f_m}(x_m)$

- **ex :** $f : \mathbb{R}^m \to \mathbb{R}$ convexe fermée, $f(x) = \sum_{i=1}^{m} f_i(x_i)$

alors $\text{prox}_f(x) = \left(\text{prox}_{f_i}(x_i)\right)_{i=1}^{m}$ vecteur

- **ex :** $f : \begin{cases} \mathbb{R}^m \longrightarrow \mathbb{R} \\ x \longrightarrow \lambda \|x\|_1 \end{cases}$ et $g(x) = \sum_{i=1}^{m} f(x_i)$ et $f(s) = s.|\lambda|$



Comme $\text{prox}_f(x) = \mathcal{T}_\lambda(x)$ alors $\text{prox}_g(x) = \left(\mathcal{T}_\lambda(x_i)\right)_{i=1}^{m}$

Théorème : composition avec une fonction affine

- Soit $g : \mathbb{R}^m \longrightarrow \mathbb{R} \cup \{+\infty\}$ convexe, propre et fermée où $f(x) = g(A(x) + b)$, $A : E \longmapsto \mathbb{R}^m$ tq $AA^T = \alpha I_d$ avec $\alpha > 0$. Alors :

$$\text{prox}_f(x) = x + \frac{1}{\alpha} A^T \, \text{prox}_{\alpha g}\left((A(x) + b) - (A(x) + b)\right)$$

- **ex :** $f : \begin{cases} \mathbb{R}^m \longrightarrow \mathbb{R} \\ x \longrightarrow |a^T x| \end{cases}$ avec $a \in \mathbb{R}^m \setminus \{0\}$

On a : $f(x) = g(a^T x)$ avec $g(t) = |t|$

ie $\text{prox}_g = \mathcal{T}_1$ et $A(x) = a^T x$

et $AA^T(x) = \|a\|^2 x$ ie $b = 0$ et $\alpha = \|a\|^2$

donc $\text{prox}_f(x) = x + \frac{1}{\|a\|^2}\left(\mathcal{T}_{\|a\|^2}(a^T x) - a^T x\right) a$

- Retour sur la projection :

Soit $g : \mathbb{R}^m \longrightarrow \mathbb{R} \cup \{+\infty\}$ avec $g(x) = \delta_C(x)$ et $C \subseteq \mathbb{R}^m$ non vide

Théorème : $\text{prox}_g(x) = \underset{y \in \mathbb{R}^m}{\text{argmin}} \left\{ \frac{1}{2}\|x - y\|^2 + \delta_C(y) \right\}$ où $\delta_S(x) = \begin{cases} 0 & \text{si } x \in S \\ +\infty & \text{sinon} \end{cases}$

$= \underset{y \in C}{\text{argmin}} \left\{ \frac{1}{2}\|x - y\|^2 \right\} = \text{Proj}_C(x)$

Théorème:

- Soit f une fonction convexe, propre et fermée, alors:

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

$$y = \mathrm{prox}_f(x) \Leftrightarrow x - y \in \partial f(y) \Leftrightarrow \langle x - y, z - y \rangle \leq f(z) - f(y), \ \forall z \in \mathbb{R}^n$$

- **ex:** $y = \mathrm{prox}_f(x) = \mathrm{argmin} \underbrace{\frac{1}{2}\|y - x\|^2 + f(y)}_{g(y)}$ alors $0 \in \partial g(y) = \nabla_y \left(\frac{1}{2}\|y - x\|^2\right) + \partial f(y)$

Descente de gradient proximal:

- $\min_{x \in \mathbb{R}^n} F(x) = f(x) + g(x)$ $\begin{cases} g : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\} \text{ convexe, propre et fermée} \\ f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\} \end{cases}$

On a $\mathrm{dom}(f)$ convexe et $\mathrm{dom}(g) \subseteq \mathrm{int}(\mathrm{dom}(f))$

$f \in C^{1,1}_{L_f}$ et on suppose que $\mathrm{argmin}\, F \neq \emptyset$

- **ex:** i. si $g \equiv 0$, $\mathrm{dom}(f) = \mathbb{R}^n$ $\quad \min_{x \in \mathbb{R}^n} f(x)$

ii. si $g = \delta_C$, $C \in \mathbb{R}^n$ convexe, fermé et non vide $\quad \min_{x \in C} f(x)$

iii. si $g = \lambda \|.\|_1$ $\quad \min_{x \in \mathbb{R}^n} f(x) + \lambda \|x\|_1$

- **Rappel:** on avait généré une suite $(x_k)_k$ avec:

$$x_{k+1} = \mathrm{Proj}_C(x_k - t_k \nabla f(x_k))$$

$\Leftrightarrow$

$$x_{k+1} = \underset{x}{\mathrm{argmin}} \left\{ \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle}_{\text{appl. linéaire}} + \underbrace{\frac{1}{2t_k} \|x - x_k\|^2}_{\text{prox}} \right\}$$

Rappel : $\min_{x \in C} f(x)$ $\quad (f(x) + \delta_C(x)$ où $\text{prox}_{\cdot} = \delta_C)$

- Itération de l'algorithme de gradient proximal (PGM)

$$x_{k+1} = \text{prox}\,(x_k - t_k \nabla f(x_k))$$

On cherche : $\min_{x \in \mathbb{R}^n} F = f + g$

i. g convexe, propre et fermée ($g \in \Gamma_0$)

ii. $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, propre, fermée, donc f convexe et $\text{dom}(g) \subseteq \text{int}(\text{dom}(f))$ et $f \in C^{1,1}_{L_f}$

iii. $\text{argmin}\, F \neq \emptyset$

PGM : Initialisation : $x_0 \in \text{int}(\text{dom}(f))$

Étapes : i. $\forall k \geq 0$, choisir $t_k > 0$ $\qquad O\left(\frac{1}{k}\right)$

ii. $x_{k+1} = \text{prox}_{t_k g}(x_k - t_k \nabla f(x_k)) = T_{t_k}^{f,g}(x_k)$

$\longrightarrow$ critères d'arrêt : $\frac{\|x_{k+1} - x_k\|}{\|x_k\|}$, $\|\nabla f(x_k)\|$, $\|G(x_k)\|$, ... $< \varepsilon$

- **ex :** i. $\min_{x \in \mathbb{R}^n} f(x) \longrightarrow x_{k+1} = x_k - t_k \nabla f(x_k)$ GM

ii. $\min_{x \in C} f(x) \longrightarrow x_{k+1} = \text{Proj}_C(x_k - t_k \nabla f(x_k))$ PGM

iii. $\min_{x \in \mathbb{R}^n} f(x) + \lambda \|x\|_1 \longrightarrow x_{k+1} = \mathcal{T}_{\lambda t_k}(x_k - t_k \nabla f(x_k))$ ISTA : iterative shrinkage thresholding algo

Lemme de descente :

- Sous les mêmes hypothèses.

$\forall x \in \text{int}(\text{dom}(f))$ et $L \in \left]\frac{L_f}{2}, +\infty\right[$, on a :

$$F(x) - F(T_L(x)) \geq \left(\frac{L - \frac{L_f}{2}}{L^2}\right) \|G_L^{f,g}(x)\|^2$$

où $G_L^{f,g} : \begin{cases} \text{int}(\text{dom}(f)) \longrightarrow \mathbb{R}^n \\ x \longrightarrow L(x - T_L(x)) \end{cases}$

- On appelle $G_L^{f,g}$: "gradient mapping"

On observe que :

$$x_{k+1} = T_L(x_k) = x_k - \frac{1}{L} G_L(x_k)$$

- **Remarque:** quand $f = 0$: $x_{k+1} = \text{prox}_{t_k g}(x_k)$ (algo du point proximal)

Théorème:

- Sous les mêmes hypothèses et en supposant en plus que f est convexe. Soit (x_k) la suite générée par l'algorithme de gradient proximal. Alors:

$$F(x_k) - F^* \leq \frac{L\|x_0 - x^*\|^2}{k}, \quad \forall x^* \in \text{argmin}(F)$$

Descente de gradient proximal accélérée:
i. $g \in I_0$
ii. $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$, $\overbrace{L_f\text{-smooth}}^{f \in C^{1,1}_{L}}$ et convexe
iii. $\text{argmin}(F = f + g) \neq \emptyset$

fast proximal gradient method

FISTA: Initialisation: prendre x_0, $y_0 = x_0$ et $t_0 = 1$

Étapes: i. pour $k \geq 0$, choisir $L_k > 0$

ii. $x_{k+1} = \text{prox}_{\frac{1}{L_k} g}\left(y_k - \frac{1}{L_k} \nabla f(y_k)\right)$

iii. $t_{k+1} = \dfrac{1 + \sqrt{1 + 4t_k^2}}{2}$

iv. $y_{k+1} = x_{k+1} + \left(\dfrac{t_k - 1}{t_{k+1}}\right)(x_{k+1} - x_k)$

$O\left(\frac{1}{k^2}\right)$

- Convergence de FISTA: $F(x_k) - \underset{= F(x^*)}{F^*} \leq \dfrac{L\|x_0 - x^*\|^2}{(k+1)^2}, \quad \forall x^* \in \text{argmin}\{F\}$

- **ex:** $\min \overbrace{\frac{1}{2}\|Ax - b\|^2}^{f(x)} + \overbrace{\lambda\|x\|_1}^{g(x)}$ avec $\begin{cases} A \in \mathbb{R}^{m \times n}, \quad \forall \lambda > 0 \\ b \in \mathbb{R}^m \end{cases}$

- Mq $L_f = \lambda_{max}(A^T A)$

$$\begin{cases} x_{k+1} = \mathcal{T}_{\frac{\lambda}{L_f}}\left(y_k - \frac{1}{L_f} A^T (Ay_k - b)\right) \\ t_{k+1} = \dfrac{1 + \sqrt{4t_k + 1}}{2} \\ y_{k+1} = x_{k+1} + \dfrac{t_k - 1}{t_k}(x_{k+1} - x_k) \end{cases}$$

problème primal

Dualité : $\min_x f(x) \qquad \max_y \phi(y)$

problème dual

Définition : Conjugué de Fenchel - Legendre

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$.

On définit son conjugué par :

$$f^* : \begin{cases} \mathbb{R}^n \longrightarrow \overline{\mathbb{R}} \\ y \longrightarrow f^*(y) = \sup_x \left(\langle x, y \rangle - f(x) \right) \end{cases} \quad \text{et } \delta_C(x) = \begin{cases} 0 \text{ si } x \in C \\ +\infty \text{ si } x \notin C \end{cases}$$

- **ex :** $f^*(y) = \sup_{x \in \mathbb{R}^n} \langle x, y \rangle - \delta_C(x)$

$= \sup_{x \in C} \langle x, y \rangle = \sigma_C(y)$ fonction support de C

Propriété de f^* : i. $f^* \in M$.

ii. si $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ propre $\{x, f(x) \neq 0\} \neq \emptyset$

$\forall x, y \in \mathbb{R}^n$, $\langle x, y \rangle \leq f^*(y) + f(x)$ Inégalité de Fenchel

Théorème :

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ convexe et propre

$\langle x, y \rangle = f^*(y) + f(x) \iff y \in \partial f(x)$ ♥ preuve exo

Si en plus f est fermée, alors :

$y \in \partial f(x) \iff x \in \partial f^*(y)$ essence de la dualité

Biconjugué :

- $f^{**}(x) = \sup_{y \in \mathbb{R}^n} \langle x, y \rangle - f^*(y)$ avec $f : E \longrightarrow \mathbb{R} \cup \{+\infty\}$ et $\langle \cdot, \cdot \rangle_{E^*, E}$

Théorème :

- Soit $f \in \Gamma_0(\mathbb{R}^n)$ alors $f^{**} = f$.

Règles de calcul :

Théorème : fonctions séparables

- Soit $g : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p} \longrightarrow \mathbb{R} \cup \{+\infty\}$ définie par $g(x_1, \dots, x_p) = \sum_{i=1}^{p} f_i(x_i)$ où $f_i : \mathbb{R}^{n_i} \longrightarrow \mathbb{R} \cup \{+\infty\}$ propres.

Alors :

$$g^*(y_1, \dots, y_p) = \sum_{i=1}^{p} f_i^*(y_i), \quad \forall y_i \in \mathbb{R}^{n_i} \text{ pour } i = 1, \dots, p$$

Démonstration :

- Soient $(y_1, \dots, y_p) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$, on a :

$$\begin{aligned} g^*(y_1, \dots, y_p) &= \sup_{\substack{x_i \in \mathbb{R}^{n_i} \\ i = 1, \dots, p}} \langle (y_1, \dots, y_p), (x_1, \dots, x_p) \rangle - g(x_1, \dots, x_p) \\ &= \sup_{\substack{x_i \in \mathbb{R}^{n_i} \\ i = 1, \dots, p}} \sum_{i=1}^{p} \langle y_i, x_i \rangle - \sum_{i=1}^{p} f_i(x_i) \\ &= \sum_{i=1}^{p} \sup_{\substack{x_i \in \mathbb{R}^{n_i} \\ i = 1, \dots, p}} \langle y_i, x_i \rangle - f(x_i) \\ &= \sum_{i=1}^{p} f^*(y_i) \end{aligned}$$

Théorème: changement de variable affine

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ et $A \in GL_n(\mathbb{R})$, où $a, b \in \mathbb{R}^n$ et $c \in \mathbb{R}$.

  Alors le conjugué de Fenchel de $g(x) = f(A(x-a)) + \langle b, x\rangle + c$ est donné par:

  $$g^*(y) = f^*((A^T)^{-1}(y-b)) + \langle a, y-b\rangle - c, \ \forall y \in \mathbb{R}^n$$

Démonstration:

- Soit $y \in \mathbb{R}^n$:

$$\begin{aligned}
g^*(y) &= \sup_x \langle x, y\rangle - g(x) \\
&= \sup_x \langle x, y\rangle - f(A(x-a)) - \langle b, x\rangle - c \\
&= \sup_z \langle y, A^{-1}(z) + a\rangle - f(z) - \langle b, A^{-1}(z) + a\rangle - c \\
&= \langle a, y-b\rangle - c + \sup_z \langle y-b, A^{-1}(z)\rangle - f(z) \\
&= \langle a, y-b\rangle - c + \sup_z \langle (A^{-1})^T(y-b), z\rangle - f(z) \\
&= f^*((A^{-1})^T(y-b)) + \langle a, y-b\rangle - c
\end{aligned}$$

$z = A(x-a)$

$\langle x, Ay\rangle = \langle A^T x, y\rangle$

Proposition:

- Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ et $\lambda > 0$:

  i. pour $g(x) = \lambda f(x)$, on a: $g^*(y) = \lambda f^*\left(\frac{y}{\lambda}\right)$

  ii. pour $g(x) = \lambda f\left(\frac{x}{\lambda}\right)$, on a: $g^*(y) = \lambda f^*(y)$

Démonstration:

- Soit $y \in \mathbb{R}^n$:

$$\begin{aligned}
g^*(y) &= \sup_x \langle x, y\rangle - g(x) \\
&= \sup_x \langle x, y\rangle - \lambda f(x) \\
&= \sup_x \lambda\left(\langle x, \frac{y}{x}\rangle - f(x)\right) \\
&= \lambda \sup_x \left(\langle \frac{x}{\lambda}, y\rangle - f(x)\right) \\
&= \lambda f^*\left(\frac{y}{\lambda}\right)
\end{aligned}$$

- ex:

1) $f(x) = \frac{1}{p}|x|^p$ pour $p>1$

alors $f^*(y) = \sup_x \langle x,y\rangle - \frac{1}{p}|x|^p$ et $(|x|)' = \frac{x}{|x|}$

$= \sup_x xy - f(x)$

$= \sup_x h(x)$

avec $h(x) = xy - \frac{1}{p}|x|^p$, qui est concave

on a : $h'(\bar{x}) = y - sg(\bar{x})|\bar{x}|^{p-1} = 0 \rightarrow y = sg(\bar{x})|\bar{x}|^{p-1}$

donc $sg(y) = sg(\bar{x})$ ie $|y| = |\bar{x}|^{p-1}$

soit $\bar{x} = sg(y)|y|^{\frac{1}{p-1}}$

donc $h(\bar{x}) = sg(y)y|y|^{\frac{1}{p-1}} - \frac{1}{p}|y|^{\frac{p}{p-1}}$

$= |y|.|y|^{\frac{1}{p-1}} - \frac{1}{p}|y|^{\frac{p}{p-1}}$

ainsi $f^*(y) = h(\bar{x}) = \left(1-\frac{1}{p}\right)|y|^{\frac{p}{p-1}}$

Soit q tq $\frac{1}{p} + \frac{1}{q} = 1$

On trouve finalement : $f^*(y) = \frac{1}{q}|y|^q$

2) Soit $f : \begin{cases} \mathbb{R}^n \longrightarrow \mathbb{R} \\ x \longrightarrow \frac{1}{2}x^TAx + b^Tx + c \end{cases}$

$f^*(y) = \sup_x \langle x,y\rangle - f(x)$

$= \sup_x h(x)$ avec $h(x) = -\frac{1}{2}x^TAx - b^Tx - c + \langle x,y\rangle$

On trouve :

$f^*(y) = \frac{1}{2}(y-b)^T A^{-1}(y-b) - c$

Dualité de Fenchel - Rockafellar :

- (P) : $\min_x h(Ax) + g(x)$ primal

avec $\begin{cases} h : \mathbb{R}^k \longrightarrow \mathbb{R} \cup \{+\infty\} \\ g : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\} \\ A : \mathbb{R}^n \longrightarrow \mathbb{R}^m \end{cases}$ propres

optimisation linéaire

- Le problème dual associé à (P) s'écrit :

$$(D) : \max_{y \in \mathbb{R}^m} -h^*(y) - g^*(-A^*y)$$

- On a : $\min_{x \in \mathbb{R}^n} h(Ax) + g(x) = \min_{x,z} \{ h(z) + g(x),\ z = Ax \}$

- Le Lagrangien du problème s'écrit :

$$L(x,z;y) = h(z) + g(x) + \langle y, Ax - z \rangle$$
$$= (-\langle y, Ax \rangle - g(x)) - (\langle y, z \rangle - h(z))$$

donc $\sup_{x,z} L(x,z;y) = -h^*(y) - g^*(-A^*y) = q(y)$

- **ex :** $f : \begin{cases} \mathbb{R}^n \longrightarrow \mathbb{R} \\ x \longrightarrow \frac{1}{2} x^T A x + b^T x + c \end{cases}$, avec $\begin{cases} A \in S_{++}^n \\ b \in \mathbb{R}^n \\ c \in \mathbb{R} \end{cases}$

Mq $f^*(y) = \frac{1}{2}(y-b)^T A^{-1} (y-b) - c$

On a $f^*(y) = \sup_x \langle x, y \rangle - f(x) = \sup_x \underbrace{\langle x, y \rangle - \frac{1}{2} x^T A x - \langle b, x \rangle - c}_{h(x) \text{ concave}}$

On peut écrire : $h(x) = \langle x, y - b \rangle - \frac{1}{2} x^T A x - c$

$\nabla_x h(\bar{x}) = (y-b) - Ax = 0 \rightarrow x = A^{-1}(y-b)$ est donc le maximum de h

donc $f^*(y) = h(\bar{x})$

$$= \langle A^{-1}(y-b), y-b \rangle - \frac{1}{2} (A^{-1}(y-b))^T \overbrace{\cancel{A\, A^{-1}}}^{Id} (y-b) - c$$
$$= (y-b)^T A^{-1} (y-b) - \frac{1}{2} (y-b)^T A^{-1} (y-b) - c$$
$$= \frac{1}{2} (y-b)^T A^{-1} (y-b) - c$$

- **ex :** $f : \begin{cases} \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\} \\ x \longrightarrow \begin{cases} -\sum_{i=1}^n \log(x_i) & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \end{cases}$

On a $f^*(y) = \sup_x \langle x, y \rangle - f(x)$

$$= \sup_x \underbrace{\sum_{i=1}^n x_i y_i + \log(x_i)}_{g(x) = -(\langle -x, y \rangle + f(x))}$$

Si $y > 0 \leftrightarrow \forall i, y_i > 0$ alors $f^*(y) = +\infty$

Si $y < 0$, $\nabla \tilde{g}(\bar{x}_i) = -\bar{x} - \frac{1}{y_i} = 0 \rightarrow \bar{x}_i = \frac{1}{y_i}$

alors $f^*(y) = \tilde{g}(\bar{x}_i) = -n - \sum_{i=1}^n \log(-y_i)$

Sinon on écrit que : $f(x) = \sum_{i=1}^{m} g(x_i)$ avec $g(t) = \begin{cases} -\log(t) & \text{si } t > 0 \\ +\infty & \text{sinon} \end{cases}$

On a : $g^*(s) = \sup_t \underbrace{st + \log(t)}_{h(t)}$

$h'(t) = s + \frac{1}{t}$ alors $\bar{t} = -\frac{1}{s}$

donc $f^*(y) = \begin{cases} \sum_{i=1}^{m} g^*(y_i) \\ +\infty \end{cases}$

- ex : $f : \begin{cases} \mathbb{R}^m \longrightarrow (-\infty, +\infty) \\ x \longrightarrow \begin{cases} \sum_{i=1}^{m} x_i \log(x_i) & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \end{cases}$

On peut écrire : $f(x) = \sum_{i=1}^{m} g(x_i)$

avec $g(t) = \begin{cases} t\log(t) & \text{si } t > 0 \\ +\infty & \text{sinon} \end{cases}$

Retour sur la dualité de Fenchel - Rockfeller :

$(P) : \min_{x \in \mathbb{R}^n} h(Ax) + g(x)$

$(D) : \max_{y \in \mathbb{R}^m} -h^*(y) - g^*(-A^T y)$

$Val(P) = \min_x h(Ax) + g(x)$, $\forall y \in dom(h^*)$ et $f(x) + f(y) \geq \langle x, y \rangle$

$\geq \min_x \langle y, Ax \rangle - h^*(y) + g(x)$

$= -h^*(y) - \sup_x \langle -A^T y, x \rangle - g(x)$

$= -h^*(y) - g^*(-A^T y)$

donc $Val(P) \geq Val(D) = \max_y -h^*(y) - g^*(-A^T y)$

Théorème de Fenchel - Rockafellar :

- $h, g \in \Gamma_0$ avec A expression linéaire si le " $\overbrace{\text{problème n'est pas dégénéré}}^{*}$ " alors :

$Val(P) = Val(D)$

$(*)$: $0 \in ri(\mathrm{dom}(h)) \setminus A(ri(\mathrm{dom} f))$

$ri(C) = \{ x \in \mathrm{Aff}(C),\ B_f(x,\varepsilon) \cap \mathrm{Aff}(C) \underset{\varepsilon>0}{\subseteq} C \}$

(P)	(D)
$\min_x \frac{1}{2}\|Ax-b\|_2^2 + \|x\|_1$	$\max_y -\frac{1}{2}\|y\|^2 - \langle b,y\rangle$ avec $\|A^T y\|_\infty \leq 1$
$\min_{\|x\|_q \leq 1} \|Ax-b\|_p$	$\max_{\|x\|_p \leq 1} -\|A^T y\|_q - \langle b,y\rangle$
$\min_x \langle c,x\rangle$, $Ax=b$, $x \in K$	$\max_y \langle b,y\rangle$, $A^T y - c \in K^\circ$

Pour l'exam : 1) ∇

2) sous-diff

3) prox

4) conjugué Fenchel

5) Méthode grad

6) Conditions KKT

7) min/max locaux

8) min/max globaux