Optimisation stochastique

1 Motivation et quelques rappels

1.1 Un problème d'optimisation "fréquent"

$$\min_{x \in \mathbb{R}^n} h(x) = \min_{x \in \mathbb{R}^n} f(x) + g(x) \tag{1}$$

avec :

- $f : \mathbb{R}^n \to \mathbb{R}$ lisse, à savoir à gradient lipschitzien :
  $\exists L > 0$ tel que $\forall x, y \in \mathbb{R}^n, ||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$
- $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexe, potentiellement non-lisse.

Exemple :

i) $g = 0$: problème d'optimisation lisse non-convexe.

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

iii) Reformulation d'un problème d'optimisation avec contraintes :
$\min\limits_{x \in \mathcal{C}} f(x)$ avec $\mathcal{C} \subset \mathbb{R}^n$ convexe non-vide.

1.2 Exemples d'applications

Exemple 1 : Moindres carrés régularisés
On dispose d'un modèle linéaire :

$$\forall x \in \mathbb{R}^n, f(x, \beta) = x^T\beta \qquad \beta \in \mathbb{R}^n \text{ paramètre du modèle}$$

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 suivant :

$$\min_{\beta \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^{p} (f(x_i, \beta) - y_i)^2 \quad \Leftrightarrow \quad \min_{\beta \in \mathbb{R}^n} \frac{1}{2}||X\beta - y||_2^2 \quad \text{avec } X = \begin{pmatrix} x_1^T \\ \vdots \\ x_p^T \end{pmatrix} \in \mathbb{M}_{p,n}(\mathbb{R}) \tag{2}$$

Texte manquant

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

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

Figure 1: SVM

iii) $g(\beta) = \frac{\lambda}{2}||\beta||_\beta^2 =$

Texte manquant

Exemple 1 : SVM (Séparateurs à Vaste Marge)
On dispose de données $(x_i)_{i\in\{1,\dots,p\}} \in \mathbb{R}^p$ labelisés $(y_i)_{i\in\{1,\dots,p\}} \in \{-1,1\}^p$
On cherche à construire un hyperplan séparant les données (x_i) selons leurs labels (y_i)

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

Quel hyperplan choisir ?
$\Rightarrow$ Incertitude : nombre de données, répartition dans $\mathbb{R}^n$, etc...
$\Rightarrow$ Maximiser la distance maximale entre les données et l'hyperplan.

Condition de séparabilité : $\quad \forall i \in \{1,\dots,p\}, y_i(x_i^T\beta) \geq 0$

D'où le problème d'optimisation suivant :

$$\max_{(\beta,M)\in(\mathbb{R}^n\times\mathbb{R}^+)} M \quad \text{s.c} \quad \forall i \in \{1,\dots,p\}, \quad y_i\frac{(x_i^T\beta)}{||\beta||_2} \geq M \tag{3}$$

Remarque : $\qquad d(z,\{\beta^Tx=0\}) = \frac{|\beta^Tz|}{||\beta||_2}$
En pratique, la condition de séparabilité n'est pas vérifiée pour tout $i \in \{1,\dots,p\}$
$\Rightarrow$ 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 :

$$\max_{(\beta,M)\in(\mathbb{R}^n\times\mathbb{R}^+)} M - \lambda\sum_{i=1}^{p}(1 - y_i\frac{(x_i^T\beta)}{||\beta||_2M})^+ \tag{4}$$

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

$$\min_{\beta\in\mathbb{R}^n} ||\beta||_2^2 + \lambda\sum_{i=1}^{p}(1 - y_i(x_i^T\beta))^+ \tag{5}$$

avec : $\sum_{i=1}^p(1-y_i(x_i^T\beta))^+$ non-lisse (non différentiable)

1.3 Rappels de convexité

Définition

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

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

Exemples :

i. partie affine : $\{x_0 + s \text{ avec } s \in S\}$ avec S un sous-espace vectoriel de $\mathbb{R}^n$ et $x_0 \in \mathbb{R}^n$.

ii. hyperplan : $\{x \in \mathbb{R}^n \text{ tel que } \alpha^T x = \beta\}$

iii. demi-espace : $\{x \in \mathbb{R}^n \text{ tel que } \alpha^T x \leq \beta\}$

iv. polyèdre : $\{x \in \mathbb{R}^n \text{ tel que } Ax \leq b\}$ avec $A \in \mathbb{M}_{m,n}(\mathbb{R})$ et $b \in \mathbb{R}^m$.

v. ellipoïde : $\{x \in \mathbb{R}^n \text{ tel que } x^T C x \leq 1\}$ avec $C \in \mathbb{S}_n(\mathbb{R})$ (matrice symétrique semi-définie positive).

Propriété

Opérations préservant la convexité :

- intersection
- somme
- multiplication par un scalaire
- produit cartésien
- image réciproque par une application linéaire
- image directe par une application linéaire
- projection : $\{x_1 \text{ tel que } (x_1, x_2) \in \mathcal{C}\}$ avec $\mathcal{C}$ convexe.

Définition - Fonctions convexes

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

- f est convexe sur $\mathcal{C}$ si :

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

- f est strictement convexe sur $\mathcal{C}$ si :

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

Propriété - CNS

Condition Nécessaire de convexité dans le cas dérivable.
Soit $f : \Omega \to \mathbb{R}$ avec Ω ouvert de $\mathbb{R}^n$ et $\mathcal{C} \subset \Omega$ convexe.

- Si f est dérivable sur Ω, alors f est convexe sur $\mathcal{C}$ convexe si et seulement si

$$\begin{aligned} &\forall x, y \in \mathcal{C}, f(y) \geq f(x) + f'(x)(y-x) \\ \Leftrightarrow &\forall x, y \in \mathcal{C}, f(y) \geq f(x) + \nabla f(x)^T (y-x) \end{aligned}$$

- Si f est deux fois dérivable sur Ω, alors f est convexe sur $\mathcal{C}$ convexe si et seulement si

$$\begin{aligned} &\forall x \in \mathcal{C}, f''(x)(y-x, y-x) \geq 0 \\ \Leftrightarrow &\forall x \in \mathcal{C}, (y-x)^T \nabla^2 f(x)(y-x) \geq 0 \end{aligned}$$

Propriété

- f convexe sur $\mathcal{C}$ convexe $\Rightarrow \alpha f$ convexe sur $\mathcal{C}$ convexe pour $\alpha > 0$
- Combinaisons linéaires à coefficients positifs de fonctions convexes sont convexes
- Soit f convexe sur $\mathcal{C}$ convexe. Soit $A \in \mathbb{M}_{m,n}(\mathbb{R})$ et $b \in \mathbb{R}^m$.
  Alors $\mathcal{C}' = \{x \in \mathbb{R}^n \text{ tel que } Ax + b \in \mathcal{C}\}$ est convexe et $x \mapsto f(Ax+b)$ est convexe sur $\mathcal{C}'$.
- Soit $(f_i)_{i \in \{1,\ldots,m\}}$ convexes sur $(\mathcal{C}_i)_{\{i \in \{1,\ldots,m\}\}}$
  Alors $\max\limits_{i \in \{1,\ldots,m\}} f_i$ convexe sur $\bigcap_{i=1}^m \mathcal{C}_i$.
- Soit $g : \mathbb{R}^n \to \mathbb{R}$ convexe sur $\mathcal{C} \subset \mathbb{R}^n$
  Soit $h : \mathbb{R} \to \mathbb{R}$ croissante et convexe sur $\mathcal{C}'$ tel que $g(\mathcal{C}) \subset \mathcal{C}'$
  Alors $h \circ g$ est convexe sur $\mathcal{C}$.
- Soit $g : \mathbb{R}^n \to \mathbb{R}^p$ avec $\forall i \in \{1, \ldots, p\}$, g_i convexe sur $\mathbb{R}^n$
  Soit $h : \mathbb{R}^p \to \mathbb{R}$ croissante et convexe vis-à-vis de chacun de ses arguments.
  Alors $\begin{aligned} f : \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto h \circ g(x) = h(g_1(x), \ldots, g_p(x)) \end{aligned}$ est convexe sur $\mathbb{R}^n$

1.4 Régularité des fonctions convexes

Définition - Epigraphe

Soit $f : \mathcal{C} \to \mathbb{R}$
On appelle *épigraphe* de f, noté $\varepsilon(f)$, l'ensemble suivant :

$$\varepsilon(f) = \{(x, w) \in \mathcal{C} \times \mathbb{R} \text{ tel que } f(x) \leq w\}$$

Propriété

Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C} \subset \mathbb{R}^n$ convexe.
f est convexe sur $\mathcal{C}$ si et seulement si $\varepsilon(f)$ est convexe.

Preuve :

i) Supposons f convexe sur $\mathcal{C}$ convexe
$\forall x, y \in \mathcal{C}$, $\forall \lambda \in [0, 1]$, $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
Soient $(x_1, w_1) \in \varepsilon(f)$ et $(x_2, w_2) \in \varepsilon(f)$ et $\lambda \in [0, 1]$
$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \leq \lambda w_1 + (1-\lambda)w_2$
$\Rightarrow \lambda(x_1, w_1) + (1-\lambda)(x_2, w_2) \in \varepsilon(f)$

ii) Supposons $\varepsilon(f)$ convexe
Soit $x, y \in \mathcal{C}$ et $\lambda \in [0, 1]$
$(x, f(x)), (y, f(y)) \in \varepsilon(f)$
$\Rightarrow \lambda(x, f(x)) + (1-\lambda)(y, f(y)) \in \varepsilon(f)$
$\Rightarrow (\lambda x + (1-\lambda)y, \lambda f(x) + (1-\lambda)f(y)) \in \varepsilon(f)$
$\Rightarrow f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$

Propriété - Inégalité de Jensen

Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C} \subset \mathbb{R}^n$ convexe.
Soit $x_1, \ldots, x_p \in \mathcal{C}$ et $\lambda_1, \ldots, \lambda_p \in \mathbb{R}^+$ tel que $\sum_{i=1}^p \lambda_i = 1$.

Alors $\quad f(\sum_{i=1}^p \lambda_i x_i) \leq \sum_{i=1}^p \lambda_i f(x_i)$

Propriété

Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C} \subset \mathbb{R}^n$ convexe. Soit $x_0 \in \mathring{\mathcal{C}}$
f est continue en x_0.

Preuve :
Soit $x_0 \in \mathring{\mathcal{C}}$. Soit Δ un simplexe inclus dans $\mathring{\mathcal{C}}$ et contenant x_0.
Notons $(s_i)_{i \in \{1,\ldots,n+1\}}$ les sommets de Δ.
$\forall x \in \Delta$, $\exists! (\lambda_i)_{i \in \{1,\ldots,n+1\}} \in \mathbb{R}^{n+1}$ tel que $\lambda_i \geq 0$ et $\sum_{i=1}^{n+1} \lambda_i = 1$ et $x = \sum_{i=1}^{n+1} \lambda_i s_i$
(coordonnées barycentriques de x vis-à-vis de Δ).

d'où

$$f(x) \leq \sum_{i=1}^{n+1} \lambda_i f(s_i) \text{ par convexité de } f \text{ et inégalité de Jensen.}$$
$$\leq \max_{i\in\{1,\ldots,n+1\}} f(s_i) \text{ car } \sum_{i=1}^{n+1} \lambda_i = 1$$

Donc f est majorée sur Δ.
En particulier, $\forall \delta > 0$ tel que $B(x_0, \delta) \subset \Delta$, f est majorée sur $B(x_0, \delta)$.
Fixons un tel δ et notons M un majorant de f sur $B(x_0, \delta)$.

Texte manquant

$\Rightarrow f(\delta) \leq 2f(x_0) - M$

Bilan : $2f(x_0) - M \leq f(z) \leq M$, $\forall z \in B(x_0, \delta)$
$\Rightarrow f$ est bornée sur $B(x_0, \delta)$

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

On montre que f est lipschitzienne sur $B(x_0, \frac{\delta}{2})$
Soit $x, y \in B(x_0, \frac{\delta}{2})$, $x \neq y$

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 $x' \in B(x_0, \delta)$: $||x' - x_0|| = ||x - \frac{\delta}{2}\frac{y-x}{||y-x||} - x_0|| \leq ||x - x_0|| + \frac{\delta}{2} < \delta$
De même, $y' \in B(x_0, \delta)$

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

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

Par convexité de f sur $\mathcal{C}$: $f(x) \leq \lambda f(x') + (1-\lambda) f(y)$
$\Rightarrow f(x) - f(y) \leq \lambda(f(x') - f(y)) \leq 2K\lambda \leq \frac{4K}{2||y-x||+\delta}||y - x||$
$\Rightarrow |f(x) - f(y)| \leq \frac{4K}{\delta}||y - x||$

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

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

2 Sous-différentiell d'une fonction

2.1 Sous-gradient et sous différentiel

Définition - Sous-differentiel et sous-gradient

Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C} \subset \mathbb{R}^n$ convexe. Soit $x \in \mathcal{C}$ et $g \in \mathbb{R}^n$
g est appelé *sous-gradient* de f en x si :

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

On appelle *sous-différentiel* de f en x, noté $\partial f(x)$, l'ensemble des sous-gradients de f en x :

$$\partial f(x) = \{g \in \mathbb{R}^n \text{ tel que } \forall y \in \mathcal{C}, f(y) \geq g^T(y-x) + f(x)\}$$

Exemple : $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$
Soit $x \in \mathbb{R}$

- si $x < 0$, alors $f(x) = -x$
  Soit $g \in \mathbb{R}$ tel que $\forall y \in \mathbb{R}$, $f(y) \geq g(y-x) + f(x)$
  - Soit $y \leq 0$, $f(y) = -y = -y + x - x = -(y-x) + f(x) \geq g(y-x) + f(x)$ avec $g = -1$
  - Soit $y > 0$, $f(y) = y \geq -y + x - x \geq -(y-x) + f(x) \geq g(y-x) + f(x)$ avec $g = -1$

  Donc $\partial f(x) = \{-1\}$
- si $x > 0$, $\partial f(x) = \{1\}$ (même raisonnement)
- si $x = 0$, $\partial f(x) = [-1, 1]$

Propriété

Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C} \subset \mathbb{R}^n$ convexe. Soit $x \in \mathcal{C}$
Alors $\partial f(x)$ est un convexe non-vide.

<u>Preuve :</u>
Supposons $\partial f(x) = \emptyset$
alors $\partial f(x) = \bigcap_{y \in \mathcal{C}} \{g \in \mathbb{R}^n \text{ tel que } f(y) \geq g^T(y-x) + f(x)\}$

or, $\forall y \in \mathcal{C}$, $\{g \in \mathbb{R}^n \text{ tel que } f(y) \geq g^T(y-x) + f(x)\}$ est convexe (demi-espace) et fermé (comme image réciproque d'un fermé par une application continue ψ)

$$\begin{aligned} \psi : \mathbb{R}^n &\to \mathbb{R} \\ g &\mapsto f(y) - f(x) - g^T(y-x) \end{aligned}$$

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

Propriété

Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C} \subset \mathbb{R}^n$ convexe. Soit $x \in \overset{\circ}{\mathcal{C}}$ tel que f est continue en x. Alors $\partial f(x)$ est borné.

<u>Preuve :</u>
Soit $x \in \overset{\circ}{\mathcal{C}}$ tel que f est continue en x.
$x \in \overset{\circ}{\mathcal{C}}$: $\exists y_1 > 0$ tel que $B(x, y_1) \subset \mathcal{C}$
f continue en x : $\forall \varepsilon > 0$, $\exists y_2 > 0$ tel que $\forall y \in B(x, y_2)$, $|f(y) - f(x)| < \varepsilon$

Posons $\eta = min(y_1, y_2)$

Montrons que $\partial f(x)$ est borné.
Supposons le contraire.
$\forall M > 0$, $\exists g \in \partial f(x)$ tel que $||g||_2 > M$
Soit $M > 0$. Fixons un tel g tel que $||g||_2 > M \Rightarrow g \neq 0$

Soit $y = x + \frac{\eta}{2}\frac{g}{||g||_2}$
d'où $||y - x||_2 = \frac{\eta}{2} < \eta \Rightarrow y \in B(x, \eta) \subset \mathcal{C}$

Par définition de g : $f(y) \geq g^T(y - x) + f(x)$
$\Rightarrow f(y) - f(x) \geq \frac{\eta}{2}||g||_2 > \frac{\eta}{2}M$ avec $M = \frac{2}{\eta}$
$\Rightarrow |f(y) - f(x)| > \varepsilon$

Or, $y \in B(x, \eta) \Rightarrow y \in B(x, y_2) \Rightarrow |f(y) - f(x)| < \varepsilon$

D'où $\varepsilon < |f(y) - f(x)| < \varepsilon$: contradiction.

Donc $\partial f(x)$ est borné.

Propriété

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

- $\forall x \in \overset{\circ}{\mathcal{C}}$, $\partial f(x) \neq \emptyset$ (Et $\partial f(x)$ compact convexe non-vide)
- Si f dérivable en $x \in \overset{\circ}{\mathcal{C}}$, alors $\partial f(x) = \{\nabla f(x)\}$

<u>Preuve :</u>

i) Soit $\mathcal{C} \subset \mathbb{R}^n$ convexe non-vide.
Soit $x_0 \in \mathcal{C}^c \cup (\overline{\mathcal{C}} \backslash \overset{\circ}{\mathcal{C}})$
alors $\exists \alpha \in \mathbb{R}^n \backslash \{0\}$ tel que $\sup\limits_{z \in \mathcal{C}} \alpha^T z \leq \alpha^T x_0$

Texte manquant

Soit $x \in \mathring{\mathcal{C}}$

Texte manquant

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

- f convexe sur $\mathcal{C}$ convexe et dérivable en $x \in \mathring{\mathcal{C}}$
$\Rightarrow \forall y \in \mathcal{C},\ f(y) \geq \nabla f(x)^T(y-x) + f(x)$
$\Rightarrow \nabla f(x) \in \partial f(x)$
$\Rightarrow \{\nabla f(x)\} \subset \partial f(x)$

- Soit $g \in \partial f(x)$
$\forall y \in \mathcal{C},\ f(y) \geq g^T(y-x) + f(x)$
Or, $x \in \mathring{\mathcal{C}}$: $\exists N \in \mathbb{N}$ tel que $y_N = x + \frac{u}{N} \in \mathcal{C}$ avec $u \in \mathbb{R}^n$

  Fixons un tel N.
  $\forall n \geq N,\ f(y_n) \geq \frac{1}{n} g^T u + f(x)$

  Or, f dérivable en x :
  $f(y_n) = f(x) + \frac{1}{n}\nabla f(x)^T u + \frac{1}{n}||u||_2 \varepsilon(\frac{1}{n}u)$ avec $\varepsilon(h) \xrightarrow[h \to 0]{} 0$
  $\Rightarrow f(x) + \frac{1}{n}\nabla f(x)^T u + \frac{1}{n}||u||_2 \varepsilon(\frac{1}{n}u) \geq \frac{1}{n} g^T u + f(x)$
  $\Rightarrow (\nabla f(x) - g)^T u + ||u||_2 \varepsilon(\frac{1}{n}u) \geq 0$

  A la limite : $(\nabla f(x) - g)^T u \geq 0, \forall u \in \mathbb{R}^n$
  $\Rightarrow g = \nabla f(x)$

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

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

Propriété

Soit $f : \mathcal{C} \to \mathbb{R}$ avec $\mathcal{C} \subset \mathbb{R}^n$ convexe. Soit $x^* \in \mathcal{C}$
Alors x^* est un minimum global de f sur $\mathcal{C}$ si et seulement si $0 \in \partial f(x^*)$

Preuve :
$0 \in \partial f(x^*) \Leftrightarrow \forall y \in \mathcal{C},\ f(y) \geq 0^T(y-x^*) + f(x^*) = f(x^*) \Leftrightarrow x^*$ est un minimum global de f sur $\mathcal{C}$

2.2 Calculs de sous-gradients

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

Propriété

- Soit $(\alpha_1, \alpha_2) \in (\mathbb{R}_+^*)^2$. On pose $f = \alpha_1 f_1 + \alpha_2 f_2$.
  Alors $\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$
- Soit $h : x \mapsto f(Ax + b)$ avec $A \in \mathbb{M}_{m,n}(\mathbb{R})$ et $b \in \mathbb{R}^m$
  Alors $\partial h(x) = A^T \partial f(Ax + b)$
- Soit $f : x \mapsto \max\limits_{i \in \{1,\ldots,m\}} f_i(x)$
  Soit $I_0 = \{i \in \{1, \ldots, m\}$ tel que $f_i(x) = f(x)\}$
  Alors $\forall g \in \partial f_{I_0}(x),\ g \in \partial f(x)$
- Soit $f : x \mapsto \sup\limits_{a \in A} f_a(x)$
  Soit $I_0 = \{a \in A$ tel que $f_a(x) = f(x)\}$
  Alors $\forall g \in \partial f_{I_0}(x),\ g \in \partial f(x)$

  Texte manquant

3 Algorithmes de sous-gradient

3.1 Algorithme du sous-gradient

On considèle le problème suivant :

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{avec } f \text{ convexe sur } \mathbb{R}^n. \tag{6}$$

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

On a l'Algorithme suivant :

Entrées : $x_0 \in \mathbb{R}^n$
Tant que
Texte manquant

Remarque : f est 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 $k \in \mathbb{N}$

$$\begin{aligned}||x_{k+1} - x^*||_2^2 &= ||x_k - \alpha_k g_k - x^*||_2^2 \\ &= ||x_k - x^*||_2^2 - 2\alpha_k g_k^T (x_k - x^*) + \alpha_k^2 ||g_k||_2^2\end{aligned}$$

Texte manquant

On pose : $f_{best}^k = \min\limits_{j \in \{1,\ldots,k\}} f(x_j)$

On a :

$2\sum\limits_{l=1}^{k}\alpha_l(f(x_l)-f^*) \leq ||x_1-x^*||_2^2 + \sum\limits_{l=1}^{k}||gl||_2^2$

$\Rightarrow 0 \leq f_{best}^k - f(x^*) \leq \frac{||x_1-x^*||_2^2+\sum\limits_{l=1}^{k}||gl||_2^2}{2\sum\limits_{l=1}^{k}\alpha_l}$

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

On a donc : $0 \leq f_{best}^k - f(x^*) \leq \frac{R^2+\sum\limits_{l=1}^{k}||gl||_2^2}{2\sum\limits_{l=1}^{k}\alpha_l}$

On suppose de plus : $\exists a>0$ tel que $\forall l \in \mathbb{N}^*$, $||gl|| \leq a$

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

<u>Quelle stratégie de pas α_l choisir ?</u>

- Pas constant : $\forall l \in \{1,\dots,k\}$, $\alpha_l = \alpha > 0$

  On a donc : $0 \leq f_{best}^k - f(x^*) \leq \frac{R^2+ka^2\alpha^2}{2k\alpha} \xrightarrow[k\to\infty]{} \frac{a^2\alpha}{2}$
  $\Rightarrow f_{best}^k \in B_f(f(x^*), \frac{a^2\alpha}{2})$

- $\forall l \in \{1,\dots,k\}, \alpha_l > 0$

-

$$\forall l \in \mathbb{N}, \alpha_l = \frac{\gamma_l}{gl} \qquad \text{avec} \qquad \begin{array}{l} \gamma_l > 0 \\ \lim\limits_{l\to\infty}\gamma_l = 0 \\ \sum\limits_{l=1}^{\infty}\gamma_l = \infty \end{array}$$

Texte manquant

$$\forall k \geq N_1+1, \qquad 0 \leq f_{best}^k - f(x^*) \leq \frac{R^2+a^2\sum\limits_{l=1}^{N_1}\alpha_l^2}{2\sum\limits_{l=1}^{k}\alpha_l} + \frac{a^2\sum\limits_{l=N_1+1}^{k}\alpha_l^2}{2\sum\limits_{l=1}^{k}\alpha_l}$$

$$= \frac{R^2+a^2\sum\limits_{l=1}^{N_1}\alpha_l^2}{2\sum\limits_{l=1}^{k}\alpha_l} + \frac{\varepsilon}{2}\frac{\sum\limits_{l=N_1+1}^{k}\alpha_l^2}{\sum\limits_{l=1}^{k}\alpha_l}$$

De plus, $\sum\limits_{l=1}^{k} \alpha_l \to \infty$

$\exists N_2 \in \mathbb{N}$ tel que $\forall k \geq N_2, \sum\limits_{l=1}^{k} \alpha_l^2 \geq \frac{R^2+a^2\sum\limits_{l=1}^{N_1}\alpha_l^2}{\varepsilon}$

D'où $\forall k \geq max(N_1+1, N_2)$, $0 \leq f_{best}^k - f(x^*) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$

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

Remarque : On s'intéresse à la stratégie de pas qui minimise $\psi(\alpha) = \frac{R^2+a^2||\alpha||_2^2}{2||\alpha||_1}$ avec $\alpha \in (\mathbb{R}_+^*)^k$
Avec une telle stratégie de pas, il pourra être nécessaire de faire $\mathcal{O}(\frac{1}{\varepsilon})$ itérations pour atteindre $|f_{best}^k - f(x^*)| \leq \varepsilon$
Pour $\varepsilon = 10^{-3}$, on a 10^6 itérations : l'algorithme est très long dans le pire des cas.

3.2 Algorithme du sous-gradient projeté

Soit $f : \mathbb{R}^n \to \mathbb{R}$ convexe. On considère le problème suivant :

$$\min_{x\in\mathcal{C}} f(x) \quad \text{avec } \mathcal{C} \text{ partie convexe, fermée =, non-vide de } \mathbb{R}^n. \tag{7}$$

On suppose que (7) admette au moins une solution $x^* \in \mathcal{C}$

Remarque : $\mathcal{C}$ partie convexe fermée non-vide de $\mathbb{R}^n$:
La projection orthogonale sur $\mathcal{C}$, notée $\Pi_{\mathcal{C}}$, est 1-lipschitzienne :

$$\forall (x,y) \in (\mathbb{R}^n)^2, ||\Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(y)||_2 \leq ||x-y||_2$$

Idée Projection orthogonale sur $\mathcal{C}$ des itérations de l'algorithme du sous-gradient.

On a l'algorithme suivant :

Entrées : $x_0 \in \mathbb{R}^n$
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

Y'a-t-il convergence ?
Soit $z_{k+1} = x_k - \alpha_k g_k \in \mathbb{R}^n$ avec $g_k \in \partial f(x_k)$ et $\alpha_k > 0$

$$\begin{aligned} ||z_{k+1} - x^*||_2^2 &= ||x_k - \alpha_k g_k - x^*||_2^2 \\ &\leq ||x_k - x^*||_2^2 - 2\alpha_k(f(x_k) - f(x^*)) + \alpha_k^2||g_k||_2^2 \end{aligned}$$

$$\begin{aligned}\text{Or,}\quad ||x_k - x^*||_2^2 &= ||\Pi_{\mathcal{C}}(z_{k+1}) - \Pi_{\mathcal{C}}(x^*)||_2^2 \qquad \text{car } x^* \in \mathcal{C} \Rightarrow \Pi_{\mathcal{C}}(x^*) = x^* \\ &\leq ||z_{k+1} - x^*||_2^2 \qquad \text{car } \Pi_{\mathcal{C}} \text{ est 1-lipschitzienne} \\ &\leq ||x_k - x^*||_2^2 - 2\alpha_k(f(x_k) - f(x^*)) + \alpha_k^2||g_k||_2^2\end{aligned}$$

On retrouve les mêmes contraintes de convergence que pour l'algorithme du sous-gradient.
Notamment, on a besoin de $\mathcal{O}(\frac{1}{\varepsilon})$ itérations pour atteindre $|f_{best}^k - f(x^*)| \leq \varepsilon$

Remarque : $\Pi_{\mathcal{C}}$ peut être difficile à calculer en pratique selon ce qu'est $\mathcal{C}$

3.3 Cas particulier : contraintes convexes d'inégalité

On considère le problème suivant :

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} f(x) \quad \text{avec } f \text{ et } (f_i)_{i \in \{1,\dots,p\}} \text{ convexe sur } \mathbb{R}^n \\ \forall i \in \{1,\dots,p\}, f_i(x) \leq 0 \end{cases} \tag{8}$$

En posant $\mathcal{C} = \{x \in \mathbb{R}^n \text{ tel que } \forall i \in \{1,\dots,p\}, f_i(x) \leq 0\}$

Le problème (8) devient :

$$\min_{x \in \mathcal{C}} f(x) \qquad \text{avec } \mathcal{C} \text{ convexe} \tag{9}$$

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

Quelles autres stratégies ?
On cherche un algorithme qui respecte les contraintes à chaque itérations.

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

- Calculer $g_k \in \mathbb{R}^n$ tel que :
  - $g_k \in \partial f(x_k)$ si $x_k \in \mathcal{C}$
  - $g_k \in \partial f_i(x_k)$ tel que $f_i(x_k) > 0$.
- $x_{k+1} = x_k - \alpha_k g_k$

Fin Tant que

On pose $f_{best}^k = \min\limits_{j \in \{1,\dots,k\}} f(x_j)$ avec $x_j \in \mathcal{C}$

On suppose de plus que : $\exists x_l \in \mathbb{R}^n$ tel que $\forall i \in \{1,\dots,p\}$, $f_i(x_l) < 0$ et $f(x_l) \neq f(x^*)$ (avec x^* solution de (9))

Avec (α_l) tel que : $\begin{cases} \forall l \in \mathbb{N}^*, \alpha_l > 0 \\ \sum \alpha_l = \infty \\ \sum \alpha_l^2 < \infty \end{cases}$

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

<u>Preuve :</u>
Supposons le contraire : $\exists \varepsilon > 0$ tel que $\forall N \in \mathbb{N}$, $\exists k \geq N$ tel que $f_{best}^k \geq f(x^*) + \varepsilon$
En particulier, $f_{best}^N \geq f_{best}^k \geq f(x^*) + \varepsilon, \forall N \in \mathbb{N}$

Soit $\lambda \in [0,1]$. On pose $\tilde{x} = (1-\lambda)x^* + \lambda x_l \in \mathcal{C}$ (par convexité de $\mathcal{C}$)

Par convexité de f, $\quad \begin{aligned} f(\tilde{x}) &\leq (1-\lambda)f(x^*) + \lambda f(x_l) \\ &\leq f(x^*) + \lambda(f(x_l) - f(x^*)) \end{aligned}$

Texte manquant

$\forall i \in \{1, \ldots, p\}$, $f_i(\tilde{x}) \leq (1-\lambda)f_i(x^*) + \lambda f_i(x_e)$ par convexité de f_i
Or, $f_i(x^*) \leq 0$ car $x^* \in \mathcal{C}$ et $f_i(x_l) < 0$ par définition de x_l

$$\begin{aligned} f_i(\tilde{x}) &\leq \lambda f_i(x_l) \\ &\leq \lambda \max_{i \in \{1,\ldots,p\}} f_i(x_l) \\ &\leq -\mu \qquad \text{avec } \mu = \lambda \max_{i \in \{1,\ldots,p\}} f_i(x_l) > 0 \end{aligned}$$

D'où $\exists (\tilde{x}, \mu) \in \mathcal{C} \times \mathbb{R}_+^*$ tel que $\begin{cases} f(\tilde{x}) \leq f(x^*) + \frac{\varepsilon}{2} \\ \forall i \in \{1, \ldots, p\}, f_i(\tilde{x}) \leq -\mu \end{cases}$

$\forall k \in \mathbb{N}$, $\quad \begin{aligned} ||x_{k+1} - \tilde{x}||_2^2 &= ||x_k - \alpha_k g_k - \tilde{x}||_2^2 \\ &= ||x_k - \tilde{x}||_2^2 - 2\alpha_k g_k^T(x_k - \tilde{x}) + \alpha_k^2 ||g_k||_2^2 \end{aligned}$

- 1er cas : $x_k \in \mathcal{C}$
  Alors $g_k \in \partial f(x_k)$ et $||x_{k+1} - \tilde{x}||_2^2 \leq ||x_k - \tilde{x}||_2^2 - 2\alpha_k(f(x_k) - f(\tilde{x})) + \alpha_k^2 ||g_k||_2^2$
  Or, $x_k \in \mathcal{C}$

Texte manquant

4 Méthodes proximales

4.1 Fonction proximale

Soit $h : \mathbb{R}^n \to \mathbb{R}$ convexe. On définit la fonction proximale de h, noté $prox_h$, par :

$$\forall x \in \mathbb{R}^n, \quad \text{prox}_h(x) = \underset{u \in \mathbb{R}^n}{\text{argmin}} \left\{ h(u) + \frac{1}{2}||u - x||_2^2 \right\} \tag{10}$$

<u>Exemple :</u>

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

ii) *Texte manquant*

iii) $\forall x \in \mathbb{R}^n, h(x) = \lambda ||x||_1$ avec $\lambda > 0$
Alors prox_h est appelé *Seuillage doux*, et est défini par :

$$\forall i \in \{1, \dots, n\}, \quad [\text{prox}_h(x)]_i = \begin{cases} x_i - \lambda & \text{si } x_i > \lambda \\ 0 & \text{si } |x_i| \leq \lambda \\ x_i + \lambda & \text{si } x_i < -\lambda \end{cases} \tag{11}$$

4.2 Méthode du gradient proximal

On s'intéresse au problème d'optimisation suivant :

$$\min_{x \in \mathbb{R}^n} f(x) = g(x) + h(x) \tag{12}$$

Avec :

- g convexe et dérivable sur $\mathbb{R}^n$
- h convexe, potentiellement non-lisse (mais telle que prox_h soit facile à calculer)

Entrées : $x_0 \in \mathbb{R}^n$
Tant que
- $x_{k+1} = \text{prox}_{\alpha_k} h(x_k - \alpha_k \nabla g(x_k))$

et α_k obtenue depuis :

i) $\forall k \in \mathbb{N}, \alpha_k = \alpha > 0$ (pas constant)

ii) Recherche linéaire

<u>Exemple :</u>

i) $h = 0$: $x_{k+1} = x_k - \alpha_k \nabla g(x_k)$ (méthode de descente de gradient)

ii) $h(x) = \begin{cases} 1 \text{ si } x \in \mathcal{C} & \text{avec } \mathcal{C} \text{ convexe fermée non-vide de } \mathbb{R}^n \\ +\infty \text{ sinon} \end{cases}$ $\quad x_{k+1} = \Pi_{\mathcal{C}}(x_k - \alpha_k \nabla g(x_k))$ (méthode du gradient projeté)

iii) $h(x) = \lambda ||x||_1$ avec $\lambda > 0$
Seuillage doux pour minimiser $f(x) = g(x) + \lambda ||x||_1$

Remarque :

i)

$$
\begin{aligned}
x_{k+1} &= \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \alpha_k h(u) + \frac{1}{2} ||u - x_k + \alpha_k \nabla g(x_k)||_2^2 \right\} \\
&= \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ h(u) + \frac{1}{2\alpha_k} ||u - x_k||_2^2 + 2\alpha_k \nabla g(x_k)^T (u - x_k) + \alpha^2 ||\nabla g(x_k)||_2^2 \right\} \\
&= \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ h(u) + g(x_k) + \nabla g(x_k)^T (u - x_k) + \frac{1}{2\alpha_k} ||u - x_k||_2^2 + \frac{\alpha_k}{2} ||\nabla g(x_k)||_2^2 \right\} \\
&= \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ h(u) + g(x_k) + \nabla g(x_k)^T (u - x_k) + \frac{1}{2\alpha_k} ||u - x_k||_2^2 \right\}
\end{aligned}
$$

Avec : $g(x_k) + \nabla g(x_k)^T (u - x_k) + \frac{1}{2\alpha_k} ||u - x_k||_2^2$ LE modèle quadratique *dégénéré* de g en x_k

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

iii) La méthode du gradient proximal peut se réécrire :
$x_{k+1} = x_k + \alpha_k G_{\alpha_k}(x_k)$ avec $G_{\alpha_k}(x) = \frac{1}{\alpha}(x - \text{prox } \alpha h(x - \alpha \nabla g(x)))$

En effet,
$$
\begin{aligned}
& x_k - \alpha_k G_{\alpha_k}(x_k) = \text{prox } \alpha_k h(x_k - \alpha_k \nabla g(x_k)) \\
& \Leftrightarrow x_k \alpha_k \nabla g(x_k) - x_k + \alpha_k G_{\alpha_k}(x_k) \in \partial \alpha_k h(x_k - \alpha_k G_{\alpha_k}(x_k)) \qquad \text{par ii)} \\
& \Leftrightarrow \alpha_k \left[G_{\alpha_k}(x_k) - \nabla g(x_k) \right] \in \partial \alpha_k h(x_k - \alpha_k G_{\alpha_k}(x_k)) \\
& \Leftrightarrow G_{\alpha_k}(x_k) - \nabla g(x_k) \in \partial h(x_k - \alpha_k G_{\alpha_k}(x_k)) \\
& \Leftrightarrow G_{\alpha_k}(x_k) \in \partial h(x_k - \alpha_k G_{\alpha_k}(x_k)) + \{\nabla g(x_k)\}
\end{aligned}
$$

De plus, $G_{\alpha_k}(x_k) = 0 \Leftrightarrow 0 \in \partial h(x_k) + \{\nabla g(x_k)\} \Leftrightarrow x_k$ est un minimum de f

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

Recherche linéaire : backtracking et condition d'arrêt modifiée
Entrées : $\alpha_0 > 0$, $\beta \in]0, 1[$, $x_k \in \mathbb{R}^n$, $\nabla g(x_k) \in \mathbb{R}^n$
Tant que

- $\alpha_{l+1} = \beta \alpha_l$
- Critère d'arrêt :
$g(x_k - \alpha_l G_{\alpha_l}(x_k)) \leq g(x_k) - \alpha_l G_{\alpha_l}(x_k)^T \nabla g(x_k) + \frac{\alpha_l}{2} ||G_{\alpha_l}(x_k)||_2^2$

Fin Tant que

4.3 Recherche linéaire (choix des pas) et convergence de l'algorithme

Propriété

On suppose que ∇g est L-lipschitzienne avec $L > 0$:
$\forall (x, y) \in (\mathbb{R}^n)^2, ||\nabla g(x) - \nabla g(y)||_2 \leq L||x - y||_2$
Alors l'algorithme de recherche linéaire s'arrête sur $\alpha_l \geq \min\,(\alpha_0, \frac{\beta}{L}) := \alpha_{min}$

Preuve :
∇g est L-lipschitzienne $\Rightarrow \forall (x, y) \in (\mathbb{R}^n)^2, g(y) \leq g(x) + \nabla g(x)^T (y - x) + \frac{L}{2} ||y - x||_2^2$
$\Rightarrow g(x_k - \alpha_l G_{\alpha_l}(x_k)) \leq g(x_k) - \alpha_l G_{\alpha_l}(x_k)^T \nabla g(x_k) + \frac{L}{2} \alpha_l^2 ||G_{\alpha_l}(x_k)||_2^2$

La condition est valide pour tout $\alpha_l \in]0, \frac{1}{L}[$
Donc à l'arrêt de la recherche linéaire, $\alpha_l \geq \min\,(\alpha_0, \frac{\beta}{L})$

Propriété

On suppose que ∇g est L-lipschitzienne. A l'arrêt de la recherche linéaire $(\alpha_k, G_{\alpha_k}(x_k))$, on a :
$\forall z \in \mathbb{R}^n, \quad f(x_k - \alpha_k G_{\alpha_k}(x_k)) \leq f(z) + G_{\alpha_k}(x_k)^T (x_k - z) - \frac{\alpha_k}{2} ||G_{\alpha_k}(x_k)||_2^2$

Preuve : Par définition du critère d'arrêt de la recherche linéaire :

$$\begin{aligned}
g(x_k - \alpha_k G_{\alpha_k}(x_k)) &\leq g(x_k) - \alpha_k G_{\alpha_k}(x_k)^T \nabla g(x_k) + \frac{\alpha_k}{2} ||G_{\alpha_k}(x_k)||_2^2 \\
&\leq g(x_k) + \alpha_k G_{\alpha_k}(x_k)^T (G_{\alpha_k}(x_k) - \nabla g(x_k)) - \frac{\alpha_k}{2} ||G_{\alpha_k}(x_k)||_2^2
\end{aligned}$$

On a : $G_{\alpha_k}(x_k) - \nabla g(x_k) \in \partial h(x_k - \alpha_k G_{\alpha_k}(x_k))$
Donc : $\forall z \in \mathbb{R}^n, h(z) \geq [G_{\alpha_k}(x_k) - \nabla g(x_k)]^T (z - x_k + \alpha_k G_{\alpha_k}(x_k)) + h(x_k - \alpha_k G_{\alpha_k}(x_k))$

De plus,

$$\begin{aligned}
g(x_k - \alpha_k G_{\alpha_k}(x_k)) + h(x_k - \alpha_k G_{\alpha_k}(x_k)) &\leq g(x_k) + h(x_k - \alpha_k G_{\alpha_k}(x_k)) \\
&+ \alpha_k G_{\alpha_k}(x_k)^T (G_{\alpha_k}(x_k) - \nabla g(x_k)) \\
&- \frac{\alpha_k}{2} ||G_{\alpha_k}(x_k)||_2^2 \\
\Leftrightarrow f(x_k - \alpha_k G_{\alpha_k}(x_k)) &\leq g(x_k) + h(x_k - \alpha_k G_{\alpha_k}(x_k)) \\
&+ \alpha_k G_{\alpha_k}(x_k)^T (G_{\alpha_k}(x_k) - \nabla g(x_k)) \\
&- \frac{\alpha_k}{2} ||G_{\alpha_k}(x_k)||_2^2
\end{aligned}$$

Or, $h(x_k - \alpha_k G_{\alpha_k}(x_k)) \leq h(z) + [G_{\alpha_k}(x_k) - \nabla g(x_k)]^T (x_k - z - \alpha_k G_{\alpha_k}(x_k))$

$$\begin{aligned}
\Rightarrow f(x_k - \alpha_k G_{\alpha_k}(x_k)) \leq g(x_k) + h(z) &+ \alpha_k G_{\alpha_k}(x_k)^T (G_{\alpha_k}(x_k) - \nabla g(x_k)) \\
&+ [G_{\alpha_k}(x_k) - \nabla g(x_k)]^T (x_k - z - \alpha_k G_{\alpha_k}(x_k)) \\
&- \frac{\alpha_k}{2} ||G_{\alpha_k}(x_k)||_2^2
\end{aligned}$$

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

Or, g est convexe et dérivable sur $\mathbb{R}^n$ d'où : $\forall z \in \mathbb{R}^n, g(z) \geq \nabla g(x_k)^T (z - x_k) + g(x_k)$
$\Rightarrow g(x_k) \leq g(z) + \nabla g(x_k)^T (x_k - z)$

Texte manquant