Gradient à pas optimal

Référence : [HU09] p.66-69.

Théorème 0.1 *Soit $A \in S_n^{++}(\mathbb{R})$. Soit $b \in \mathbb{R}^n$. Soit $c \in \mathbb{R}$. Soit :*

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto < Ax, x > + < b, x > + c \end{aligned}$$

On considère le problème d'optimisation suivant : minimiser $f(x)$, avec $x \in \mathbb{R}^n$ et on souhaite déterminer si ce problème admet une unique solution et si tel est le cas, la vitesse de convergence de la suite (x_n) donnée par l'algorithme du gradient à pas optimal vers la solution.

Démonstration

Étape 1 : Montrons que le problème admet une unique solution déterminée par une équation.

$\rightsquigarrow$ **Existence**

- $\mathbb{R}^n$ est un fermé.
- f est continue car polynomiale en les x_i.
- Montrons que f est coercive. Pour tout $x \in \mathbb{R}^n$, on a :

$$|f(x)| = \left| \frac{1}{2} < Ax, x > + < b, x > + c \right| \geq \frac{1}{2} | < Ax, x > | - | < b, x > + c | \geq \frac{1}{2} | < Ax, x > | - \parallel b \parallel \parallel x \parallel - |c|$$

d'après l'inégalité de Minkowski, l'inégalité triangulaire et celle de Cauchy-Schwarz.
De plus, comme $A \in S_n^{++}(\mathbb{R})$, alors A est diagonalisable en base orthonormée, notons alors $\lambda_1 \geq \ldots \geq \lambda_n$ ses valeurs propres dans l'ordre décroissant. Alors :

$$| < Ax, x > | \geq \lambda_n \parallel x \parallel^2$$

(car $\lambda_{min} = \min_{x \neq 0} \dfrac{< Ax, x >}{\parallel x \parallel^2}$). D'où :

$$|f(x)| \geq \frac{1}{2} \lambda_n \parallel x \parallel^2 - \parallel b \parallel \parallel x \parallel - |c|$$

Or f est strictement positive lorsque $x \longrightarrow \infty$ (car $A \in S_n^{++}(\mathbb{R})$) donc lorsque x est grand :

$$f(x) \geq \frac{1}{2} \lambda_n \parallel x \parallel^2 - \parallel b \parallel \parallel x \parallel - |c|$$

D'où f coercive, ie $\lim f(x) = +\infty$ lorsque $\parallel x \parallel \longrightarrow +\infty$.

Ainsi en appliquant le lemme 2, on obtient l'existence d'une solution pour le problème.

$\rightsquigarrow$ **Unicité**

- $\mathbb{R}^n$ est convexe.
- Montrons que f est strictement convexe sur $\mathbb{R}^n$.
  Déterminons pour cela la matrice hessienne de f. Soit $x = (x_1, \ldots, x_n)$, alors :

$$f(x_1, \ldots, x_n) = \frac{1}{2} \left(x_1 \sum_{i=1}^n a_{1i} x_i + \ldots + x_n \sum_{i=1}^n a_{ni} x_i \right) + (b_1 x_1 + \ldots + b_n x_n) + c$$

D'où :

$$\frac{\partial f}{\partial x_1} = \frac{1}{2} \left(\sum_{i=1}^n a_{1i} x_i + x_1 a_{11} + \ldots + x_n a_{n1} \right) + b_1$$

Ainsi :

$$\frac{\partial f}{\partial x_i \partial x_1} = \frac{1}{2}\left(a_{1i} + a_{i1}\right) = a_{1i}$$

car A est symétrique.

Ainsi, on montre que la matrice hessienne de f est A et comme $A \in S_n^{++}(\mathbb{R})$, alors f est strictement convexe (on aurait pu déterminer la matrice hessienne de f en utilisant la formule de la différentielle, car $f(x+h) = f(x) + < \nabla f(x), h > + \frac{1}{2} < Hessf(x)h, h >$).

D'où l'unicité du minimum d'après le lemme 4.

$\rightsquigarrow$ **Équation du minimum** Comme f est différentiable, en particulier en $\overline{x}$ et que $\overline{x}$ est le minimum de f sur $\mathbb{R}^n$, alors $\nabla f(\overline{x}) = 0$. En effet :

– Montrons que f est différentiable. Soient $x, h \in \mathbb{R}^n$:

$$\begin{aligned}
f(x+h) &= \frac{1}{2} < A(x+h), x+h > + < b, x+h > + c \\
&= \frac{1}{2} < Ax, x > + \frac{1}{2} < Ax, h > + \frac{1}{2} < Ah, x > + \frac{1}{2} < Ah, h > + < b, x > + < b, h > + c \\
&= f(x) + \frac{1}{2} < Ax, h > + \frac{1}{2} < h, {}^tAx > + < b, h > + \frac{1}{2} < Ah, h > \\
&= f(x) + \frac{1}{2} < Ax, h > + \frac{1}{2} < h, Ax > + < b, h > + o(\| h \|) \quad \text{car } A \text{ est symétrique ;} \\
&= f(x) + < Ax + b, h > + o(\| h \|)
\end{aligned}$$

D'où f différentiable et le gradient de f vaut $Ax + b$.

– Ainsi $\overline{x} = -A^{-1}b$ et si l'on remplace l'expression de $\overline{x}$ dans f, on obtient :

$$\begin{aligned}
\overline{f} = f(\overline{x}) &= \frac{1}{2} < A(-A^{-1}b), -A^{-1}b > + < b, -A^{-1}b > + c \\
&= \frac{1}{2} < b, A^{-1}b > - < b, A^{-1}b > + c \\
&= -\frac{1}{2} < A^{-1}b, b > + c
\end{aligned}$$

Étape 2 : Déterminons des relations sur les itérées dans la méthode du gradient à pas optimal.

L'algorithme du gradient à pas optimal consiste à construire une suite x_k de la manière suivante :

– initialisation : $x_0 \in \mathbb{R}^n$;
– définition de l'itérée x_{k+1} à partir de x_k (lorsque $\nabla f(x_k) \neq 0$) : $x_{k+1} = x_k + t_k d_k$; où $d_k = -\nabla f(x_k)$ et t_k est l'unique réel positif minimisant $t \longmapsto f(x_k + t d_k)$ sur $\mathbb{R}$.

Vérifions maintenant les relations suivantes pour tout $k \in \mathbb{N}$:

$$\begin{aligned}
t_k &= \frac{\| d_k \|^2}{< Ad_k, d_k >} \\
d_{k+1} &= d_k - t_k A d_k \\
< d_{k+1}, d_k > &= 0 \\
f(x_{k+1}) - \overline{f} &= (f(x_k) - \overline{f}) \left(1 - \frac{\| d_k \|^4}{< Ad_k, d_k >< A^{-1}d_k, d_k >}\right)
\end{aligned}$$

où $\overline{f}$ désigne la valeur optimale dans le problème de minimisation.

$\rightsquigarrow$ **Déterminons t_k.**

$$\begin{aligned}
f(x_k + td_k) &= \frac{1}{2} < A(x_k + td_k), x_k + td_k > + < b, x_k + td_k > + c \\
&= \frac{1}{2} < Ax_k, x_k > + \frac{1}{2} < Ax_k, td_k > + \frac{1}{2} < Atd_k, x_k > + \frac{1}{2} < Atd_k, td_k > + < b, x_k > + < b, td_k > + c \\
&= f(x_k) + \frac{1}{2} t^2 < Ad_k, d_k > + t < Ax_k, d_k > + t < b, d_k > \quad \text{car } A \text{ est symétrique ;} \\
&= f(x_k) + \frac{1}{2} t^2 < Ad_k, d_k > + t < Ax_k + b, d_k >
\end{aligned}$$

La fonction $t \longmapsto f(x_k+td_k)$ est minimisée sur $\mathbb{R}$ (on le montre en effectuant une étude de fonction), lorsque $d_k = -\nabla f(x_k) \neq 0$, en un seul point qui est déterminé par :

$$\frac{df}{dt}(x_k + td_k) = t < Ad_k, d_k > + < Ax_k + b, d_k >$$
$$t = -\frac{< Ax_k + b, d_k >}{< Ad_k, d_k >}$$

Or $d_k = -\nabla f(x_k) = -(Ax_k + b)$, d'où :

$$t_k = \frac{\| d_k \|^2}{< Ad_k, d_k >} > 0$$

$\rightsquigarrow$ **Déterminons une formule liant d_{k+1} et d_k et déduisons-en celle sur le produit scalaire de ces deux éléments.**

$$\begin{aligned} d_{k+1} = -(Ax_{k+1} + b) &= -(A(x_k + t_k d_k) + b) \\ &= -Ax_k - b - At_k d_k \\ &= d_k - t_k Ad_k \end{aligned}$$

D'où :

$$\begin{aligned} < d_{k+1}, d_k > &=< d_k, d_k > -t_k < Ad_k, d_k > \\ &=\| d_k \|^2 - \frac{\| d_k \|^2}{< Ad_k, d_k >} < Ad_k, d_k > \\ &= 0 \end{aligned}$$

d'après la formule de t_k que nous venons de démontrer.
$\rightsquigarrow$ **Déterminons une formule sur $f(x_{k+1}) - \overline{f}$.**

$$\begin{aligned} f(x_{k+1}) &= f(x_k + t_k d_k) \\ &= f(x_k) + \frac{1}{2} t_k^2 < Ad_k, d_k > +t_k < Ax_k + b, d_k > \\ &= f(x_k) + \frac{1}{2} \frac{\| d_k \|^4}{(< Ad_k, d_k >)^2} < Ad_k, d_k > + \frac{\| d_k \|^2}{< Ad_k, d_k >} < Ax_k + b, d_k > \\ &= f(x_k) + \frac{1}{2} \frac{\| d_k \|^4}{< Ad_k, d_k >} + \frac{\| d_k \|^2}{< Ad_k, d_k >} \times - \| d_k \|^2 \\ &= f(x_k) - \frac{1}{2} \frac{\| d_k \|^4}{< Ad_k, d_k >} \end{aligned}$$

D'où en soustrayant par $\overline{f}$:

$$f(x_{k+1}) - \overline{f} = f(x_k) - \overline{f} - \frac{1}{2} \frac{\| d_k \|^4}{< Ad_k, d_k >} = (f(x_k) - \overline{f}) \left(1 - \frac{\| d_k \|^4}{2 < Ad_k, d_k > (f(x_k) - \overline{f})}\right)$$

Déterminons pour finir $< A^{-1} d_k, d_k >$:

$$\begin{aligned} < A^{-1} d_k, d_k > &=< A^{-1}(Ax_k + b), Ax_k + b > \\ &=< x_k + A^{-1} b, Ax_k + b > \\ &=< x_k, Ax_k > + < x_k, b > + < A^{-1} b, Ax_k > + < A^{-1} b, b > \\ &=< x_k, Ax_k > + < x_k, b > + < {}^t A A^{-1} b, x_k > + < A^{-1} b, b > \\ &=< x_k, Ax_k > +2 < x_k, b > + < A^{-1} b, b > \quad \text{car } A \text{ symétrique, donc } {}^t A = A\,; \\ &= 2 \left(\frac{1}{2} < Ax_k, x_k > + < b, x_k > + \frac{1}{2} < A^{-1} b, b > \right) \end{aligned}$$

Or $\overline{f} = -\frac{1}{2} < A^{-1} b, b > +c$, d'où $-\overline{f} + c = \frac{1}{2} < A^{-1} b, b >$, ainsi :

$$< A^{-1} d_k, d_k >= 2 \left(\frac{1}{2} < Ax_k, x_k > + < b, x_k > +c - \overline{f} \right) = 2(f(x_k) - \overline{f})$$

D'où :

$$f(x_{k+1} - \overline{f} = (f(x_k) - \overline{f}) \left(1 - \frac{\| d_k \|^4}{< Ad_k, d_k >< A^{-1} d_k, d_k >}\right)$$

Étape 3 : Déterminons la vitesse de convergence de la suite (x_k) vers $\overline{x}$.
Soient $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ les valeurs propres de A.
On pose $c_2(A) = \dfrac{\lambda_1}{\lambda_n}$ le conditionnement de A.
Montrons que $\forall k \in \mathbb{N}$:

$$f(x_k) - \overline{f} \leq (f(x_0) - \overline{f}) \left(\frac{c_2(A) - 1}{c_2(A) + 1} \right)^{2k}$$

$$\| x_k - \overline{x} \| \leq \left(\frac{2(f(x_0) - \overline{f})}{\lambda_n} \right)^{1/2} \left(\frac{c_2(A) - 1}{c_2(A) + 1} \right)^{k}$$

Et déduisons-en la vitesse de convergence de l'algorithme du gradient à pas optimal.
$\rightsquigarrow$ **Déterminons la première inégalité.** D'après l'inégalité de Kantorovich, $\forall x \in \mathbb{R}^n$:

$$< Ax, x >< A^{-1}x, x > \leq \frac{1}{4} \left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}} \right)^2 \| x \|^4$$

On l'applique à d_k (toujours lorsque $d_k \neq 0$) :

$$< Ad_k, d_k >< A^{-1}d_k, d_k > \leq \frac{1}{4} \left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}} \right)^2 \| d_k \|^4$$

D'où :

$$\frac{\| d_k \|^4}{< Ad_k, d_k >< A^{-1}d_k, d_k >} \geq \frac{4}{\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}} \right)^2} = \frac{4}{\frac{\lambda_n}{\lambda_1} \left(1 + \frac{\lambda_1}{\lambda_n} \right)^2} = \frac{4c_2(A)}{(1 + c_2(A))^2}$$

Donc $\forall k \in \mathbb{N}$:

$$\begin{aligned}
f(x_{k+1}) - \overline{f} &= (f(x_k) - \overline{f}) \left(1 - \frac{\| d_k \|^4}{< Ad_k, d_k >< A^{-1}d_k, d_k >} \right) \\
&\leq (f(x_k) - \overline{f}) \left(1 - \frac{4c_2(A)}{(1 + c_2(A))^2} \right) \\
&\leq (f(x_k) - \overline{f}) \left(\frac{(c_2(A) + 1)^2}{(c_2(A) + 1)^2} - \frac{4c_2(A)}{(1 + c_2(A))^2} \right) \\
&\leq (f(x_k) - \overline{f}) \left(\frac{c_2(A) - 1}{c_2(A) + 1} \right)^2
\end{aligned}$$

Ainsi par récurrence sur k, on obtient :

$$f(x_k) - \overline{f} \leq (f(x_0) - \overline{f}) \left(\frac{c_2(A) - 1}{c_2(A) + 1} \right)^{2k}$$

$\rightsquigarrow$ **Déterminons la dernière inégalité.** Par ailleurs, on a :

$$f(x_k) - \overline{f} = \frac{1}{2} < Ax_k, x_k > + < b, x_k > + c - \overline{f}$$

Or $c - \overline{f} = \frac{1}{2} < A^{-1}b, b >$ et $\overline{x} = -A^{-1}b$, ie $-b = A\overline{x}$, d'où :

$$\begin{aligned}
f(x_k) - \overline{f} &= \frac{1}{2} < Ax_k, x_k > + < b, x_k > + \frac{1}{2} < A^{-1}b, b > \\
&= \frac{1}{2} < Ax_k, x_k > - < A\overline{x}, x_k > + \frac{1}{2} < \overline{x}, A\overline{x} > \\
&= \frac{1}{2} < Ax_k, x_k > - \frac{1}{2} < A\overline{x}, x_k > - \frac{1}{2} < A\overline{x}, x_k > + \frac{1}{2} < \overline{x}, A\overline{x} > \\
&= \frac{1}{2} < A(x_k - \overline{x}), x_k > + \frac{1}{2} < A\overline{x}, \overline{x} - x_k > \\
&= \frac{1}{2} < A(x_k - \overline{x}), x_k > - \frac{1}{2} < \overline{x}, A(x_k - \overline{x}) > \\
&= \frac{1}{2} < A(x_k - \overline{x}), x_k - \overline{x} > \\
&\geq \frac{1}{2} \lambda_n \parallel x_k - \overline{x} \parallel^2
\end{aligned}$$

Ainsi :

$$\parallel x_k - \overline{x} \parallel^2 \leq \frac{2}{\lambda_n}(f(x_k) - \overline{f}) \leq \frac{2}{\lambda_n}(f(x_0) - \overline{f}) \left(\frac{c_2(A) - 1}{c_2(A) + 1} \right)^{2k}$$

D'où le résultat attendu, ie :

$$\parallel x_k - \overline{x} \parallel \leq \left(\frac{2(f(x_0) - \overline{f})}{\lambda_n} \right)^{1/2} \left(\frac{c_2(A) - 1}{c_2(A) + 1} \right)^{k} \simeq C \left(\frac{c_2(A) - 1}{c_2(A) + 1} \right)^{k}$$

ce qui signifie que la vitesse de convergence est exponentielle de raison $\frac{c_2(A) - 1}{c_2(A) + 1}$.

Étape4 : Interprétation D'après les deux dernières formules que nous venons de démontrer, on remarque que plus $c_2(A)$ est proche de 1 et plus la méthode du gradient à pas optimal converge rapidement.
Le cas limite serait celui où $c_2(A) = 1$, ce qui suppose que $f(x) = \lambda \parallel x - \overline{x} \parallel^2$ pour un certain $\lambda > 0$, auquel cas on atteint $\overline{x}$ dès la première itération x_1.
Par contre lorsque $c_2(A)$ est grand, ie lorsque les valeurs propres extrêmes sont très différentes, la méthode est (dans le pire des cas) très lente.
Pour être sûr d'avoir $\frac{f(x_k) - \overline{f}}{f(x_0) - \overline{f}} \leq \epsilon$, il faudrait $k \geq \frac{\log(\epsilon)}{2 \log \left(\frac{c_2(A) - 1}{c_2(A) + 1} \right)} \simeq \frac{c_2(A)}{4} \log \left(\frac{1}{\epsilon} \right)$ quand $c_2(A) \longrightarrow +\infty$.

Lemmes utilisés

Lemme 0.1 (Inégalité de Kantorovich) *Soit* $A \in S_n^{++}(\mathbb{R})$. *Alors* $\forall x \in \mathbb{R}^n$,

$$\parallel x \parallel^4 \leq < Ax, x >< A^{-1}x, x > \leq \frac{1}{4} \left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}} \right)^2 \parallel x \parallel^4$$

où λ_1 *et* λ_n *désignent respectivement la plus grande et la plus petite valeur propre de* A.

Démonstration Par homogénéité de la norme, il suffit de montrer l'inégalité pour $\parallel x \parallel = 1$.
Soient $\lambda_1 \geq \ldots \geq \lambda_n$ les valeurs propres de A et soit $P \in \mathcal{O}_n(\mathbb{R})$ tel que : $A = {}^tPDP$ avec $D = diag(\lambda_1, \ldots, \lambda_n)$ (cette décomposition existe car A est symétrique).
Alors $A^{-1} = ({}^tPDP)^{-1} = P^{-1}D^{-1}\,{}^tP^{-1} = {}^tPD^{-1}P$ car $P \in \mathcal{O}_n(\mathbb{R})$, donc $P^{-1} = {}^tP$ et avec $D^{-1} = diag\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_n} \right)$.
Ainsi $< Ax, x > = < {}^tPDPx, x > = < DPx, Px >$ et $< A^{-1}x, x > = < D^{-1}Px, Px >$.
La transformation :

$$\begin{aligned}
P : S = \{x \in \mathbb{R}^n, \parallel x \parallel = 1\} &\longrightarrow S \\
x &\longmapsto Px
\end{aligned}$$

est une bijection (car de réciproque $x \longmapsto {}^tPx$) de la sphère unité sur elle-même (car $\| Px \|^2 = < Px, Px > = < x, x >$). Effectuons alors le changement de variable $y = Px$, il faut alors montrer que $\forall y \in \mathbb{R}^n$ unitaire :

$$1 \leq < Dy, y >< D^{-1}y, y > \leq \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2$$

Considérons pour cela le graphe de la fonction $\theta : x > 0 \longmapsto \frac{1}{x}$ (on suppose que $\lambda_n < \lambda_1$ sinon il n'y a rien à démontrer car $D = Id$ et $1 \leq \| y \|^4 \leq 1$).
Étant donné $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ de norme 1, on pose $\alpha_i = y_i^2$.
Ainsi $< Dy, y > = \sum_{i=1}^n \lambda_i y_i^2$ (respectivement $< D^{-1}y, y > = \sum_{i=1}^n \frac{1}{\lambda_i} y_i^2$) est une combinaison convexe des λ_i (respectivement des $\frac{1}{\lambda_i}$).
Pour $k = 1, \ldots, n$, soit M_k le point du graphe de θ de coordonnées $(\lambda_k, \frac{1}{\lambda_k})$.
Soit M le barycentre des points M_k affectés des coefficients α_k, M est donc dans l'enveloppe convexe D délimitée par le graphe de θ et la droite (M_1M_n). Deux points de même abscisse que M jouent alors des rôles important :

- le premier M' est le point du graphe de θ ;
- le deuxième M'' est celui de (M_1M_n).

On sait que l'ordonnée de M est $\sum_{i=1}^n \frac{1}{\lambda_i}\alpha_i$.
Déterminons l'ordonnée de M'. M' est le point d'intersection de θ et de la droite $x = \sum_{i=1}^n \lambda_i \alpha_i := \overline{\lambda}$, donc son ordonnée vaut $y_{M'} = \frac{1}{\overline{\lambda}}$ et écrire que l'ordonnée de M' est inférieure à celle de M donne :

$$\frac{1}{\overline{\lambda}} \leq \sum_{i=1}^n \frac{1}{\lambda_i}\alpha_i$$

ie

$$< D^{-1}y, y >< Dy, y > \geq 1$$

Déterminons l'ordonnée de M''. M'' est le point d'intersection de la droite (M_1M_n) et de $x = \overline{\lambda}$, déterminons à présent une équation de la droite passant par (M_1M_n) :

$$a = \frac{\frac{1}{\lambda_1} - \frac{1}{\lambda_n}}{\lambda_1 - \lambda_n} = \frac{\frac{\lambda_n - \lambda_1}{\lambda_n\lambda_1}}{\lambda_1 - \lambda_n} = -\frac{1}{\lambda_n\lambda_1}$$

et

$$b = \frac{1}{\lambda_1} + \frac{1}{\lambda_n}$$

D'où :

$$y_{M''} = \frac{1}{\lambda_1} + \frac{1}{\lambda_n} - \frac{\overline{\lambda}}{\lambda_1\lambda_n}$$

Et écrire que l'ordonnée de M'' est supérieure à celle de M donne :

$$\sum_{i=1}^n \frac{1}{\lambda_i}\alpha_i \leq \frac{1}{\lambda_1} + \frac{1}{\lambda_n} - \frac{\overline{\lambda}}{\lambda_1\lambda_n}$$

ie

$$\left(\sum_{i=1}^n \lambda_i\alpha_i\right)\left(\sum_{i=1}^n \frac{1}{\lambda_i}\alpha_i\right) \leq \frac{\overline{\lambda}(\lambda_1 + \lambda_n - \overline{\lambda})}{\lambda_1\lambda_n}$$

Et comme la fonction $u \longmapsto \frac{u(\lambda_1 + \lambda_n - u)}{\lambda_1\lambda_n}$ atteint son maximum en $u = \frac{\lambda_1 + \lambda_n}{2}$ (on le montre via une étude de cette fonction), alors :

$$< Dy, y >< D^{-1}y, y > \leq \left(\frac{\lambda_1 + \lambda_n}{2}\right)^2 \frac{1}{\lambda_1\lambda_n} = \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2$$

D'où le résultat !

Lemme 0.2 *Soit f une fonction continue et coercive. On considère le problème de minimisation sur un fermé X. Alors on a existence d'au moins une solution au problème de minimisation.*

Démonstration Il suffit de se ramener au théorème qu'on connaît qui nous dit qu'une application continue sur un compact est bornée et atteint ses bornes.
La coercivité de f nous donne :

$$\forall A > 0, \exists M > 0, \| x \| > M \Rightarrow f(x) > A$$

Choisissons $y \in X$ et $A = f(y)$, alors $\exists M > 0$ tel que $\| x \| > M \Rightarrow f(x) > f(y)$.
On en déduit que $\inf_{x \in X} f(x) = \inf\{f(x), \| x \| \leq M\}$.
Or on sait que $\{x \in X, \| x \| \leq M\}$ est fermé et borné (car $= \{x \in \mathbb{R}^n, \| x \| \leq M\} \cap X$ et le premier ensemble de cette intersection est l'image réciproque d'un fermé par l'application continue norme), donc compact et on se ramène bien au théorème voulu.

Lemme 0.3 *Si f est deux fois différentiable, alors on a équivalence entre :*

1. *f est strictement convexe sur $\mathbb{R}^n$;*
2. *la matrice hessienne de f est définie positive.*

Lemme 0.4 *Soit X convexe. Si f est strictement convexe, alors f a au plus un minimiseur.*

Démonstration Par l'absurde, supposons que f a deux minimiseurs x_1 et x_2, ie $f(x_1) = f(x_2) = \min_{x \in X} f(x)$.
Mais $\frac{1}{2}(x_1 + x_2) \in X$ (car X est convexe) et :

$$f\left(\frac{1}{2}(x_1 + x_2)\right) \leq \frac{1}{2}(f(x_1) + f(x_2)) = f(x_1) = min_{x \in X} f(x)$$

absurde !

Références

[HU09] Jean-Baptiste Hiriart-Urruty. *Optimisation et analyse convexe.* EDP Sciences, 2009.