Méthode du gradient à pas optimal

Julie Parreaux

2018-2019

Référence du développement : Mathématiques pures et appliquées [3, p.412].

Leçons où on présente le développement : 162 (Systèmes d'équations linéaires) ; 219 (Extremum), 223 (Suites numériques), 229 (Fonctions monotones et convexes), 233 (Analyse numérique matricielle).

1 Introduction

La méthode du gradient est une méthode d'analyse numérique permettant de trouver un extremum (qui peut être local) d'une fonction à plusieurs variables sans contraintes. Même sans contraintes, l'optimisation multivariée (comme ici) est complexe numériquement au sens ou sa complexité peut très facilement exploser.

Il existe plusieurs variantes de cette méthode que l'on nomme descente de gradient. L'idée "magique" des méthodes de descente de gradient est de se ramené à un problème de minimisation sans contrainte à une variable (que l'on sait relativement bien traiter) à l'aide d'une suite minimisante unidirectionnelles. L'idée intuitive est que l'on va minimiser la fonction dans une unique direction. Le choix de cette direction (qui est le gradient de la fonction au point où on est) et combien de temps on l'exploite sont des arguments cruciaux de ces méthodes. En effet, selon leurs choix la convergence va pouvoir être plus ou moins puissante et rapide. Cependant, il faut prendre se rappeler de conserver des complexités raisonnables pour ces calculs.

2 Méthode du gradient à pas optimal

Schéma du développement

1. Donner l'algorithme + illustration par un dessin.
2. Montrer que $\forall k \in \mathbb{N}$, $\langle R_{k+1}, R_k \rangle = 0$.
3. Montrer la convergence de la suite X_n vers $\overline{X}$.
   (a) Montrer que $\|X_{n+1} - \overline{X}\|_A^2 = -\alpha_{n+1}\|R_n\|^2 + \|X_n - \overline{X}\|_A^2$.
   (b) Montrer que $\|X_n - \overline{X}\|_A \leq \left(\frac{\mu_n - \mu_1}{\mu_n + \mu_1}\right)^n \|X_0 - \overline{X}\|_A$.
4. Estimation de l'approximation.

2.1 Application aux cas de la résolution d'un système linéaire

La méthode du gradient à pas optimal est une méthode de recherche d'extremum basé sur une descente de gradient. Cependant, en changeant le point de vue sur le système, on peut utiliser cette méthode pour trouver la solution.

Théorème ([2, p.409 ex.4])**.** *Soit $A \in S_n^{++}(\mathbb{R})$. Alors $AX_0 = B$ si et seulement si X_0 minimise l'application $f : \mathbb{R}^n \to \mathbb{R}$ qui à X associe $\frac{1}{2}\langle AX, X\rangle - \langle B, X\rangle$.*

Démonstration. $\Leftarrow$ Supposons que $AX_0 = B$ avec $X_0 \in \mathbb{R}^n$. Alors, $\forall H \in \mathbb{R}^n$, comme $A \in S_n^{++}(\mathbb{R})$,

$$\begin{aligned}
f(X_0 + H) &= \tfrac{1}{2}\langle A(X_0+H),(X_0+H)\rangle - \langle B,(X_0+H)\rangle && \text{(Par définition de } f\text{)}\\
&= \tfrac{1}{2}\langle AX_0,X_0\rangle + \tfrac{1}{2}\langle HX_0,X_0\rangle + \tfrac{1}{2}\langle AX_0,H\rangle + \tfrac{1}{2}\langle HX_0,H\rangle - \langle B,X_0\rangle - \langle H,X_0\rangle \\
& && \text{(Manipulation du produit scalaire)}\\
&= f(X_0) + \langle AX_0,H\rangle + \tfrac{1}{2}\langle AH,H\rangle - \langle B,H\rangle && \text{(Par définition de } f\text{)}\\
&= f(X_0) + \langle B,H\rangle + \tfrac{1}{2}\langle AH,H\rangle - \langle B,H\rangle && (B = AX_0)\\
&= f(X_0) + \tfrac{1}{2}\langle AH,H\rangle \\
&\geq f(X_0) && (\tfrac{1}{2}\langle AH,H\rangle > 0 \text{ ssi } H \neq 0)
\end{aligned}$$

$\Rightarrow$ Supposons que $X_0 \in \mathbb{R}^n$ minimise f.

— $\nabla f(X_0) = 0$ (caractérisation des extremums)

— $f(X+H) \to f(X) + \langle AX - B, H\rangle$ quand $\|H\| \to 0$.

Ainsi, $\nabla f(X_0) = AX_0 - B = 0$ et X_0 est bien solution du système.

□

Lemme 1. *Toute matrice $A \in \Sigma_n^{++}(\mathbb{R})$ de valeurs propres minimale et maximale μ_1 et μ_n, vérifie pour tout $X \in \mathbb{R}^n \setminus \{0\}$:*

$$\frac{\|X\|^4}{\|X\|^2_{A^{-1}}\|X\|^2_A} \geq 4\frac{\mu_1\mu_n}{(\mu_1+\mu_n)^2}$$

Démonstration. On se place dans une base orthonormée de $\mathbb{R}^n$ constitués des vecteurs propres de A (existe par $A \in \Sigma_n^{++}(\mathbb{R})$). On étudie ensuite l'application $\phi(x) = \frac{\mu_n}{x} + \frac{x}{\mu_1}$. □

2.2 Algorithme du gradient à pas optimal

On va donc étudier la convergence de l'algorithme du gradient à pas optimal (algorithme 1) dans le cadre de fonction de ces fonctions définies à partir d'un système linéaire. Cet algorithme est un algorithme d'approximation : on se donne ϵ un critère d'arrêt qui détermine la précision de notre solution. Ce paramètre influe sur la complexité de la méthode puisque ni la convergence ni la vitesse de convergence n'est remise en cause par ce paramètre. De plus, en fonction du critère d'arrêt équivalent que l'on peut choisir, il nous faut garder en mémoire les deux derniers termes de la suite. La figure 1 est un exemple d'application de cette méthode.

Algorithm 1 Méthode du gradient à pas optimal

function (Gradient-optimal)($X_0 \in \mathbb{R}^n$, $R_0 = AX_0 - B$, $\epsilon \geq 0$)
  while $\frac{\|R_n\|}{\|R_0\|} \leq \epsilon$ **do**
    $\alpha_{n+1} \leftarrow \frac{\|R_n\|^2}{\langle AR_n, R_n\rangle}$ ▷ Pas conjugué
    $X_{n+1} \leftarrow X_n - \alpha_{n+1}R_n$ ▷ Nouvelle valeur de X
    $R_{k+1} \leftarrow AX_n - B$ ▷ Estimation de l'erreur
  end while
end function

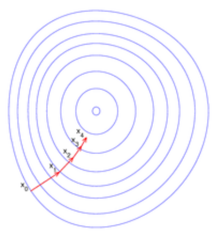

FIGURE 1 – Illustration de la méthode du gradient

2.3 Correction de cet algorithme et estimation de l'erreur

Théorème. *Soit $A \in S_n^{++}(\mathbb{R})$ et $X_0 \in \mathbb{R}^N$. L'algorithme du gradient (algorithme 1) converge vers la solution $\overline{X}$ du système $AX = B$ et on a pour tout $n \in \mathbb{N}$,*

$$\|X_n - \overline{X}\| \leq \left(\frac{\mathrm{Cond}(A)-1}{\mathrm{Cond}(A)+1}\right)^n \sqrt{\mathrm{Cond}(A)}\|X_0 - \overline{X}\|$$

Démonstration. Soit $A \in S_n^{++}(\mathbb{R})$ et $x_0 \in \mathbb{R}^N$. On note $\|X\|^2$ la norme définie par $\langle X, X\rangle$ et $\|X\|^2_A$ la norme définie par $\langle AX, X\rangle$ (les propriétés sur le produit scalaire nous assure les propriétés de norme).

Étape 1 : montrons que $\forall k \in \mathbb{N}$, $\langle R_{k+1}, R_k \rangle = 0$ Soit $k \in \mathbb{N}$.

$$\begin{aligned}
\langle R_{k+1}, R_k \rangle &= \langle AX_{k+1} - B, R_k \rangle && \text{(Définition de } R_{k+1}\text{)}\\
&= \langle AX_k - B - \alpha_{k+1} AR_k, R_k \rangle && \text{(Définition de } X_{k+1}\text{)}\\
&= \langle AX_k - B, R_k \rangle - \alpha_{k+1} \langle AR_k, R_k \rangle && \text{(Manipulation du produit scalaire)}\\
&= \langle R_k, R_k \rangle - \alpha_{k+1} \langle AR_k, R_k \rangle && \text{(Définition de } R_k\text{)}\\
&= \|R_k\|^2 - \frac{\|R_k\|^2}{\|R_k\|_A^2} \|R_k\|_A^2 && \text{(Définition des normes et de } \alpha_{k+1}\text{)}\\
&= \|R_k\|^2 - \|R_k\|^2\\
&= 0
\end{aligned}$$

Étape 2 : montrons la convergence de la suite X_n vers $\overline{X}$

Étape a : montrons que $\|X_{n+1} - \overline{X}\|_A^2 = -\alpha_{n+1}\|R_n\|^2 + \|X_n - \overline{X}\|_A^2$

$$\begin{aligned}
\|X_{n+1} - \overline{X}\|_A^2 &= \left\langle A(X_{n+1} - \overline{X}), X_{n+1} - \overline{X} \right\rangle && \text{(Définition de la norme)}\\
&= \left\langle A(X_{n+1} - \overline{X}), X_{n+1} - X_n + X_n - \overline{X} \right\rangle && \text{(Introduction de } X_n\text{)}\\
&= \left\langle A(X_{n+1} - \overline{X}), X_{n+1} - X_n \right\rangle + \left\langle A(X_{n+1} - \overline{X}), X_n - \overline{X} \right\rangle && \text{(Manipulation du produit scalaire)}\\
&= \left\langle A(X_{n+1} - \overline{X}), -\alpha_{n+1} R_n \right\rangle + \left\langle A(X_{n+1} - \overline{X}), X_n - \overline{X} \right\rangle && \text{(Définition de } \alpha_{n+1} R_n\text{)}\\
&= \langle AX_{n+1} - B, -\alpha_{n+1} R_n \rangle + \left\langle A(X_{n+1} - \overline{X}), X_n - \overline{X} \right\rangle && (\overline{X} \text{ est solution du système})\\
&= \langle R_{n+1}, -\alpha_{n+1} R_n \rangle + \left\langle A(X_{n+1} - \overline{X}), X_n - \overline{X} \right\rangle && \text{(Définition de } R_n\text{)}\\
&= -\alpha_{n+1} \langle R_{n+1}, R_n \rangle + \left\langle A(X_{n+1} - \overline{X}), X_n - \overline{X} \right\rangle && \text{(Manipulation du produit scalaire)}\\
&= \left\langle A(X_{n+1} - \overline{X}), X_n - \overline{X} \right\rangle && \text{(Étape 1)}\\
&= \left\langle A(X_n - \alpha_{n+1} R_n - \overline{X}), X_n - \overline{X} \right\rangle && \text{(Définition de } X_{n+1}\text{)}\\
&= \left\langle A(X_n - \overline{X}), X_n - \overline{X} \right\rangle - \alpha_{n+1} \left\langle AR_n, X_n - \overline{X} \right\rangle && \text{(Manipulation du produit scalaire)}\\
&= \|X_n - \overline{X}\|_A^2 - \alpha_{n+1} \left\langle R_n, A(X_n - \overline{X}) \right\rangle && \text{(Manipulation du produit scalaire)}\\
&= \|X_n - \overline{X}\|_A^2 - \alpha_{n+1} \langle R_n, AX_n - B \rangle && (\overline{X} \text{ est solution du système})\\
&= \|X_n - \overline{X}\|_A^2 - \alpha_{n+1} \langle R_n, R_n \rangle && \text{(Définition de } R_n\text{)}\\
&= \|X_n - \overline{X}\|_A^2 - \alpha_{n+1} \|R_n\|^2 && \text{(Définition de la norme)}
\end{aligned}$$

Étape b : montrons que $\|X_n - \overline{X}\|_A \le \left(\frac{\mu_n - \mu_1}{\mu_n + \mu_1}\right)^n \|X_0 - \overline{X}\|_A$ On a [2, p.413, ex.7],

$$\begin{aligned}
\|X_n - \overline{X}\|_A^2 &= \left\langle A(X_n - \overline{X}), X_n - \overline{X} \right\rangle && \text{(Définition de la norme)}\\
&= \left\langle A(X_n - \overline{X}), A^{-1}(A(X_n - \overline{X})) \right\rangle && \text{(Introduction de } A\text{)}\\
&= \left\langle AX_n - B, A^{-1}(AX_n - B) \right\rangle && (\overline{X} \text{ est solution du système})\\
&= \left\langle R_n, A^{-1}(R_n) \right\rangle && \text{(Définition de } R_n\text{)}\\
&= \|R_n\|_{A^{-1}}^2 && \text{(Définition de la norme)}
\end{aligned}$$

Ainsi,

$$\begin{aligned}
\|X_{n+1} - \overline{X}\|_A^2 &= \|R_n\|_{A^{-1}}^2 - \alpha_{n+1} \|R_n\|^2 && \text{(Par ce qui précède)}\\
&= \|R_n\|_{A^{-1}}^2 - \frac{\|R_n\|^4}{\|R_n\|_A^2} && \text{(Définition de } \alpha_{n+1}\text{)}\\
&= \|R_n\|_{A^{-1}}^2 \left(1 - \frac{\|R_n\|^4}{\|R_n\|_A^2 \|R_n\|_{A^{-1}}^2}\right) && \text{(En factorisant)}\\
&= \|X_n - \overline{X}\|_A^2 \left(1 - \frac{\|R_n\|^4}{\|R_n\|_A^2 \|R_n\|_{A^{-1}}^2}\right) && \text{(Par ce qui précède)}\\
&\le \|X_n - \overline{X}\|_A^2 \left(1 - \frac{4\mu_1 \mu_n}{(\mu_1 \mu_n)^2}\right) && (A \in S_n^{++}(\mathbb{R}))\\
&\le \frac{(\mu_n - \mu_1)^2}{(\mu_1 + \mu_n)^2} \|X_n - \overline{X}\|_A^2 && \text{(Par le lemme 1)}
\end{aligned}$$

En passant à la racine carré et en itérant, on obtient $\|X_n - \overline{X}\|_A \le \left(\frac{\mu_n - \mu_1}{\mu_1 + \mu_n}\right)^n \|X_0 - \overline{X}\|_A$. En remarquant que $0 < \frac{\mu_n - \mu_1}{\mu_1 + \mu_n} < 1$, on a bien la convergence de la suite.

Étape 3 : estimons l'approximation De plus, pour tout $X = \sum x_i e_i \in \mathbb{R}^n$ exprimé dans une base de vecteurs propres de A, on a :

$$\mu_1 \|X\|^2 = \sum \mu_1 x_i^2 \leq \|X\|_A^2 = \sum \mu_i x_i^2 \leq \sum \mu_n x_i^2 = \mu_n \|X\|^2$$

Alors, en appliquant les inégalités précédentes, on obtient :

$$\|X_n - \overline{X}\| \leq \frac{1}{\sqrt{\mu_1}} \|X_n - \overline{X}\|_A \leq \frac{1}{\sqrt{\mu_1}} \left(\frac{\mu_n - \mu_1}{\mu_1 + \mu_n} \right)^n \|X_0 - \overline{X}\|_A \leq \sqrt{\frac{\mu_n}{\mu_1}} \left(\frac{\mu_n - \mu_1}{\mu_1 + \mu_n} \right)^n \|X_0 - \overline{X}\|$$

Or, $\mathrm{Cond}(A) = \frac{\mu_n}{\mu_1}$, d'où $\sqrt{\frac{\mu_n}{\mu_1}} \left(\frac{\mu_n - \mu_1}{\mu_1 + \mu_n} \right) = \sqrt{\mathrm{Cond}(A)} \left(\frac{\mathrm{Cond}(A) - 1}{\mathrm{Cond}(A) + 1} \right)$. D'où le résultat. $\square$

3 Complément : Méthode du gradient à pas optimal via les suites stationnarisantes

Nous voulons démontrer la convergence de cette méthode. On va alors commencer par étudier quelques propriétés sur les suites stationnarisante et minimisante d'une fonction f [1, p.494].

Définition. Une suite $(x_k)_k$ de $\mathbb{R}^n$ est dite :

- stationnarisante pour f si $\lim_{k \to \infty} \nabla f(x_k) = 0$
- minimisante pour f si $\lim_{k \to \infty} f(x_k) = \alpha = \inf_{x \in \mathbb{R}^n} f(x)$.

Attention : il ne faut pas confondre suite minimisante et suite convergeant vers une fonction optimale : prendre par exemple $f(x) = e^x$. Dans le cas de la méthode du gradient à pas optimal, si la fonction que l'on souhaite minimiser n'est pas convexe, alors la suite $(x_k)_k$ définie par la méthode reste stationnarisante et ne sera pas minimisante.

Proposition 1. *Si f est convexe et différentiable sur $\mathbb{R}^n$, alors toute suite bornée stationnarisante pour f est minimisante pour f.*

Démonstration. Soit $(x_k)_k$ une suite stationnarisante pour f. Comme f est supposé convexe, on a $\forall x \in \mathbb{R}^n, f(x) \geq f(x_k) \langle \nabla f(x_k), x - x_k \rangle$. On obtient :

$$\begin{array}{rcll} \inf_{\mathbb{R}^n} f & \leq & \sup_{K \in \mathbb{N}} \inf_{k \geq K} f(x_k) = \liminf_{k \to \infty} f(x_k) & \text{(définitions de inf et sup)} \\ & \leq & \limsup_{k \to \infty} f(x_k) & (\liminf \leq \limsup) \\ & \leq & \limsup_{k \to \infty} f(x_k) + \lim_{k \in \infty} \langle \nabla f(x_k), x - x_k \rangle & (\lim_{k \in \infty \langle \nabla f(x_k), x - x_k \rangle} = 0) \\ & \leq & \limsup_{k \to \infty} [f(x_k) + \langle \nabla f(x_k), x - x_k \rangle] & (\lim < \limsup \text{ et linéarité de } \limsup) \\ & \leq & f(x) & (f(x) \geq f(x_k) \langle \nabla f(x_k), x - x_k \rangle) \end{array}$$

Comme l'inégalité précédente reste vraie pour tout $x \in \mathbb{R}^n$, on a

$$\inf_{\mathbb{R}^n} f \underbrace{\leq}_{\text{sup}} \liminf_{k \to \infty} f(x_k) \underbrace{\leq}_{\text{limites inf et sup}} \limsup_{k \to \infty} f(x_k) \underbrace{\leq}_{\text{inégalité précédente}} \inf_{\mathbb{R}^n} f$$

On en déduit que $\liminf_{k \to \infty} f(x_k) = \limsup_{k \to \infty} f(x_k)$, ce qui implique que $(f(x_k))_k$ admet une limite. De plus, les inégalités précédentes donnent $\inf_{\mathbb{R}^n} f = \limsup_{k \to \infty} f(x_k)$, ce qui implique, par unicité de la limite, que $\lim_{k \to \infty} f(x_k) = \inf_{\mathbb{R}^n} f$. $\square$

On va maintenant s'intéresser plus précisément à la suite $(x_k)_k$ définie par la méthode du gradient à pas optimal.

Lemme 2. *A chaque étape de la méthode, la direction $d_k = \nabla f(x_k)$ est une direction descendante, c'est-à-dire $\min_{[0,b]} g_k(t) < f(x_k)$.*

Démonstration. Comme on calcul $d_k = \nabla f(x_k)$, on sait que la méthode ne satisfait pas le critère de l'arrêt $||\nabla f(x_k)|| \leq \epsilon$ n'a pas stoppé la méthode. On en déduit :

$$\begin{array}{rcll} \langle d_k, \nabla f(x_k) \rangle & = & \langle -\nabla f(x_k), \nabla f(x_k) \rangle & \text{(définition de } d_k) \\ & = & -\langle \nabla f(x_k), \nabla f(x_k) \rangle & \text{(linéarité du produit scalaire)} \\ & = & -||\nabla f(x_k)||^2 & \text{(relation norme - produit scalaire)} \\ & < & 0 & (\epsilon > 0 \text{ et } ||\nabla f(x_k)||^2 < \epsilon) \end{array}$$

On raisonne par l'absurde : on suppose que $\min_{t\in[0,b]} g_k(t) \geq f(x_k)$. Alors,

$$\begin{aligned}
\forall t \in [0,b], g_k(t) \geq f(x_k) &\Leftrightarrow \forall t \in [0,b], f(x_k + td_k) \geq f(x_k) && \text{(définition de } g_k\text{)} \\
&\Rightarrow \langle d_k, \nabla f(x_k)\rangle = \lim_{t\to 0, t>0} \frac{f(x_k+td_k)-(x_k)}{t} && \text{(définition du gradient)} \\
&\Rightarrow \langle d_k, \nabla f(x_k)\rangle \geq \lim_{t\to 0, t>0} \frac{\overbrace{f(x_k)-(x_k)}^{=0}}{\underbrace{t}_{>0}} && \text{(première équivalence)} \\
&\Rightarrow \langle d_k, \nabla f(x_k)\rangle \geq 0
\end{aligned}$$

On obtient donc une contradiction avec le premier item. □

Proposition 2. *Si f est continuement différentiable sur $\mathbb{R}$ et coersive alors la suite $(x_k)_k$ est bornée, admet au moins une sous-suite convergente (Attention : il y a une erreur dans l'énoncé dans la référence) et toute sous-suite convergente est stationnaire.*

Démonstration. On suppose $\nabla f(x_k) \neq 0$, $\forall k \in \mathbb{N}$. En effet, dans le cas où $\nabla f(x_k) = 0$, la méthode s'arrête puisqu'on a trouver un extrema (qui peut être local). De plus, par la définition par récurrence, si $\nabla f(x_k) = 0$, alors la suite est constante donc convergente et stationnarisante : elle vérifie les propriétés que l'on veut.

Étape 1 : existence d'une sous-suite convergente Le lemme 2 nous assure que $\forall k \in \mathbb{N}$, $g_k(t) < f(x_k)$ ce qui est équivalent à dire que $f(x_{k+1}) = f(x_k + td_k) < f(x_k)$. Donc la suite $(f(x_k))_k$ est décroissante. De plus, la décroissance de la suite $(f(x_k))_k$, implique que $f(x_k) \leq f(x_0)$ donc $(x_k)_k$ est contenu dans le sous-niveau $S_{f(x_0)} = \{x \in \mathbb{R}^n, f(x) \leq f(x_0)\}$.

De plus, la coerisivité de f, nous assure que $S_{f(x_0)}$ est compact. Or, comme la suite $(x_k)_k$ est contenu dans $S_{f(x_0)}$ qui est compact, on en déduit que la suite $(x_k)_k$ est borné. Donc l'ensemble des points adhérents de $(x_k)_k$, $\mathrm{Adh}((x_k)_k)$, est non nul. D'où l'existence d'une sous-suite convergente.

Étape 2 : Montrer que tous les points de $\mathrm{Adh}((x_k)_k)$ sont des points stationnaires de f. Soit $\bar{x} = \lim_{n\to\infty} vx_{k_n}$.

Le lemme 2 nous assure que $\forall t \in]0,b]$ et $\forall n \in \mathbb{N}$, on a :

$$f\left(x_{k_{n+1}}\right) \leq f(x_{k_n+1}) \leq f(x_{k_n} - t\nabla f(x_{k_n}))$$

Comme f est continuement dérivable sur $\mathbb{R}^n$, $\forall t \in]0,b]$:

$$\frac{f(\bar{x} - t\nabla f(\bar{x})) - f(\bar{x})}{t} \underbrace{=}_{\bar{x}=\lim_{n\to\infty} x_{k_n}} \frac{f(x_{k_n} - t\nabla f(x_{k_n})) - f(x_{k_n})}{t} \leq 0$$

On en déduit que :

$$\langle \nabla f(\bar{x}), -\nabla f(\bar{x})\rangle \underbrace{=}_{\text{def du gradient}} \lim_{t\to 0} \frac{f(\bar{x} - t\nabla f(\bar{x})) - f(\bar{x})}{t} \underbrace{\geq}_{\text{précédent}} 0$$

Soit,

$$0 \underbrace{\leq}_{\text{précedent}} \langle \nabla f(\bar{x}), -\nabla f(\bar{x})\rangle \underbrace{=}_{\text{linéarité}} -\langle \nabla f(\bar{x}), \nabla f(\bar{x})\rangle \underbrace{=}_{\text{norme}} -||\nabla f(\bar{x})|| \underbrace{\leq}_{\text{positivité norme}} 0$$

Donc, $||\nabla f(\bar{x})|| = 0$, $\bar{x}$ est stationnaire. La suite $(x_k)_k$ est stationnarisante. □

Théorème. *Supposons que la fonction f est convexe et continuement différentiable sur $\mathbb{R}^n$.*

1. *Si f est coersive, alors la suite $(x_k)_k$ admet au moins une sous-suite convergente et toutes les sous-suites convergentes sont minimisante.*
2. *Si f est fortement convexe, alors la suite $(x_k)_k$ est minimisante et converge vers une solution optimale d'un problème sans contraintes.*

Démonstration. Supposons que la fonction f est convexe et continuement différentiable sur $\mathbb{R}^n$.

1. C'est une conséquence des propositions 1 et 2. La proposition 2 nous assure que la suite $(x_k)_k$ est bornée et admet une sous-suite convergente et toute sous-suite convergente est stationnarisante. La proposition 1 nous assure que la suite $(x_k)_k$ admet une sous-suite convergente et toute sous-suite convergente est minimisante. D'où le résultat.

2. Supposons maintenant que f est fortement convexe et différentiable, donc f est coersive.

   Le point 1 de notre théorème, il existe une sous-suite convergente et toute sous-suite convergente est minimisante. Soit $(x_n)_n$ une de ces sous-suites et $\overline{x}$ sa limite. Comme f est continue et que la suite $(x_n)_n$ est minimisante : $f(\overline{x}) = \inf_{\mathbb{R}^n} f$. On en déduit que $\lim f(x_k) = f(\overline{x}) = \inf f(x)$.

   Comme f est fortement convexe, f est strictement convexe. Donc $\mathrm{Argmin}_{\mathbb{R}^n} f = \{\overline{x}\}$. Donc, pour toute sous-suite convergente de $(x_k)_k$ converge vers $\overline{x}$ (par le lemme 2 et est incluse $S_{f(x_0)}(f)$). Par la compacité de $S_{f(x_0)}(f)$, $(x_k)_k$ converge vers $\overline{x}$.

□

Références

[1] J.-A. Weil A. Yger. *Mathématiques appliqués, L3*. Pearson Education, 2009.

[2] J.-J. Risler ; P. Boyer. *Algèbre pour la licence 3*. Dunod, 2006.

[3] Ramis ; Warusfel. *Cours de mathématiques pures et appliquées*. De Boecl, 2010.