Chapitre II :

Optimisation non- linéaire sans contraintes

1- l'optimisation sans contraintes

On aborde le problème d'optimisation sans contrainte qui est non-linéaire. Le problème est posé dans R^n, donc On veut résoudre min $x \in R^n$ f(x) ou max $x \in R^n$ f(x) i.e. on cherche v, valeur optimale et x_0 tel que $f(x_0)=v$.

2 Minimum local et minimum globale

Soit $x_0 \in S$

1. On dit que x_0 est un minimum local si
$$\exists V \in \mho\,(x_0) \quad tel\,que \quad \forall\, x \,\epsilon\, \mho\,, f(x) \,\geq f(x_0)$$
2. On dit que x_0 est un minimum local strict si
$$\exists V \in \mho\,(x_0) \quad tel\,que \quad \forall\, x \,\epsilon\, \mho\,, f(x) > f(x_0)$$
3. On dit que x_0 est un minimum global si
$$\forall\, x \,\epsilon\, S\,, f(x) \,\geq f(x_0)$$
4. On dit que x_0 est un minimum global strict si
$$\forall\, x \,\epsilon\, S\,, f(x) > f(x_0)$$

On definit de la meme facon un maximum local, un maximum local strict, maximum global et un maximum global strict en renversant les inegalités

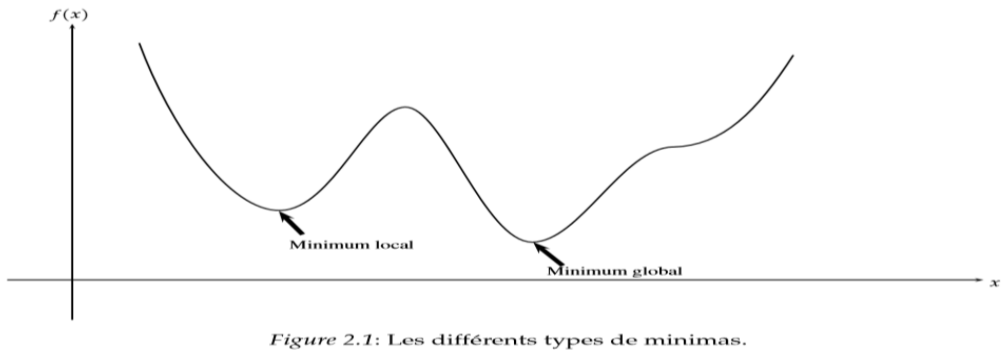


Figure 2.1: Les différents types de minimas.

Remarque : la qualité du minimum local dépend de la ''taille'' de V. plus V est ''Grant'', mieux c'est. Et si S = V, alors le minimum local est un minimum global

- ✓ **Propriétés**

1. Si x0 est un minimum global, alors c'est un minimum local.
2. Si on connait tous les minima locaux alors le plus petit est le minimum global

2.1 En pratique :

Nous notons Ps le problème $min_{x \in S} f(x)$ et P le problème $min_{x \in R^n} f(x)$

Lemme

Soit $S \subset R^n$ un ouvert . alors

x_0 est solution de Ps $\Leftrightarrow \begin{cases} x_0 \; solution \; de \; P \\ x_0 \; \in S \end{cases}$

En pratique, pour résoudre Ps on résout d'abord P , c'est plus simple puis

1. Si P a une solution $x_0 \; \in S$ alors x_0 est solution de Ps. Sinon, si $x_0 \; \notin S$ alors Ps n'a pas de solution
2. Si P n'a pas de solution, on ne sait pas pour Ps

2.2.condition du premier ordre

2.2.1. Dérivabilité et notion de gradient

Définition

Si f : $R^n \to R$. Si f est différentiable, alors on définit le gradient de f en x par

$$[\frac{\partial f}{\partial x_1}(x),..,\frac{\partial f}{\partial x_n}(x)]. \; On \; le \; note \; \nabla f(x)$$

Proposition on a le développement de Taylor suivant

$$f(x+v) = f(x) + \nabla f(x).v + \varepsilon(v)$$

Avec $\frac{\|\varepsilon(v)\|}{\|v\|} \to 0 \;\; quand \;\; \|v\| \to 0$

Exemple

Soit f : $R^2 \to R$ définit par $f(x_1 + x_2) = x_1 + x_2^2$

Alors

$$\nabla f(\bar{x}_1,\bar{x}_2) = [\frac{\partial f}{\partial x_1}(\bar{x}),\frac{\partial f}{\partial x_2}(\bar{x})] = [1 \quad 2\bar{x}_2]$$

Ou $\bar{x} = (\bar{x}_1 + \bar{x}_2)$

2.2.2. Recherche des points stationnaires

Définition

X_0 est un point stationnaire ssi $\nabla f(x_0) = 0$

Proposition : Condition nécessaire d'optimalité

Supposons que S est Ouvert et que f est différentiable sur S

Si x_0 est un minimum ou maximum local alors $\nabla f(x_0) = 0$

Corollaire les minima et les maxima locaux sont des point stationnaires et par conséquent, le minimum global et le maximum global aussi

Remarque

1. S doit être ouvert pour pouvoir appliquer la proposition

   Soit s = [0,1] et f(x) = x. alors f'(x) = 1 $\neq 0 \ \forall x$ et portant 1 est un maximum
2. La condition nécessaire exige seulement que f soit différentiable en $\bar{x}$ mais comme on ne sait pas a priori si les points candidat $\bar{x}$ se trouvent dans la zone de différentiable, on exige une condition plus forte : f différentiable sur tout S. toutefois il n'en pas toujours ainsi :

   ✓ **Propriétés**

- Soit $\boldsymbol{f}: S \rightarrow R$ telle que $f(x) = \min[\, f1\ (x), \dots, fm\ (x)]$ avec les fi différentiables sur S . alors f est différentiable en tout point $\bar{x}$ minimum local de f et donc $\nabla f(\bar{x}\ ) = 0$
- Dans certains cas , on peut conclure même si f n'est pas différentiable en $\bar{x}$

  Soit $\bar{x} \in$ S point en lequel f est continue

  Si $\exists V \in \mho(\bar{x})$ $ouvert\ tel\ que$

  1) $f\ est\ differentiable\ sur\ \mho\{\bar{x}\}$

  2) $< \nabla f(x), x - \bar{x} > \geq O(resp > 0) \forall x \in V\{\bar{x}\}$

  Alors $\bar{x}$ est minimum local (resp strict) de f.

Exemple : Soit f définie par f(x,y) = $x^{\frac{2}{3}} + x^{\frac{4}{5}} + 1$

Alors f n'est pas différentiable en (0,0) mais c'est un minimum strict de f sur R^2.

3. Attention, la réciproque est fausse : on peut avoir $\nabla f(\bar{x}) = 0$ sans que $\bar{x}$ ne soit ni un minimum ni un maximum

Conclusion : Dans les points stationnaires, il y a les minima et les maxima (locaux et globaux) et il y a aussi d'autres points.

Définition Un point stationnaire qui n'est ni un minimum ni un maximum est un point singulier.

Remarque Si le nombre de variables est :

- 1 : alors c'est un point d'inflexion. Ex : $f(x) = x^3$
- -2 : ça se complique. Ex : $f(x_1,x_2) = x_1^2 - x_2^2$. Alors
  $\rightarrow x_0 = (0,0)$ est un maximum pour $C_1 = \{(x_1,x_2)/x_1 = 0,\ y = x_1^2 - x_2^2\}$
  $\rightarrow x_0 = (0,0)$ est un minimum pour $C_2 = \{(x_1,x_2)/x_2 = 0,\ y = x_1^2 - x_2^2\}$

Définition: Un point selle ou col pour $f : R^2 \rightarrow R$ est un point $x_0 = (x_0^1, x_0^2)$ tel que $\exists \Delta_1, \Delta_2$ droites passant par x_0 telles que $f(\Delta_1)$ présente un maximum en x_0 et $f(\Delta_2)$ présente un minimum en x_0

Remarque Tous les points singuliers ne sont pas des points selle

2.3 Conditions du second ordre

2.3.1 La matrice Hessienne

Définition

Si $f : R^n \rightarrow R$ est de classe C^2 alors on définit la matrice Hessienne de f en x par

$$H_f(x) = \frac{\partial^2 f}{\partial x^2}(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

Proposition

1. H_f (x) est une matrice symétrique car : $\frac{\partial^2 f}{\partial_{x_i}\partial_{x_j}} = \frac{\partial^2 f}{\partial_{x_j}\partial_{x_i}}$ =d`es que f est deux fois dérivable (Schwarz)
2. 2. On a le développement de Taylor suivant :

$$f(x+v) = f(x) + \nabla f(x).v + \frac{1}{2}v^t H_f(x)v + \epsilon(v)$$

$avec\ \frac{||\epsilon(v)||}{||v||} \longrightarrow 0\ quand\ ||v||^2 \longrightarrow 0.$

Exemple : Soit f : $R^n \longrightarrow R$ telle que $f(x_1,x_2) = x_1 + x_2^2$.

$$\text{Alors} \begin{cases} \nabla f(x) = [1 \quad 2x_2] \\ H_f(x) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \end{cases}$$

2.3.2 Rappels sur les matrices

Définition Soit M ∈ $R^{n\times n}$ symétrique.

1. On dit que M est semi-définie positive ssi x^t M x ≥ 0 ∀x ∈ R^n

2. On dit que M est définie positive ssi x^t M x ≥ 0 ∀x ∈ R^n{0}

Remarque

On définit de la mème facon les matrices semi-définie , négative et définie négative en renversant les inégalités.

Rappel

Les mineurs principaux d'une matrice M sont les déterminants des sous-matrices k × k de M obtenues en barrant un sous-ensemble de (n − k) colonnes et le sous-ensemble correspondant des (n − k) lignes

2.3.3 Conditions du second ordre

Rappel

Les points stationnaires sont de deux types :

optimaux locaux/globaux et points singuliers (points selle et autres). Or ce que l'on recherche c'est un optimum global et on sait que s'il existe il fait partie des extrema locaux. Donc une fois que l'on a déterminé les points stationnaires ($\nabla f(x) = 0$) on doit faire le tri. D'ou la nécessité de faire une étude du second ordre.

En effet si x_0 est tel que $\nabla f(x_0) = 0$ alors

$f(x_0 + v) = f(x_0) + 1/2\, v^t H f(x)\, v + \epsilon(||v||^2)$

il faut donc déterminer le signe de $v^t H f(x) v$

Si H f est définie positive, $f(x_0 + v) > f(x_0)$ et x_0 est un minimum local strict.

Si H f est semi définie positive, $f(x_0 + v) \geq f(x_0)$ et x_0 est un minimum local.

Si le signe de $v^t H f(x) v$ varie, x_0 est un point singulier.

2.3.3.1.Conditions nécessaires

Si $x_0 \in S$ ouvert est un minimum local de f et si f est C^2, alors

$$\begin{cases} \nabla f(x_0) = 0 \\ y^t H f(x) y \geq 0 \forall y \in \mathbb{R}^n \end{cases}$$

La deuxième condition signifie que H f est semi-définie positive.

2.3.3.2.Condition suffisante

Soit $f : S \subset R^n \rightarrow R$ est C^2 en $x^0 \in S$.

Si $\nabla f(x_0)=0$ et $y^t Hf(x)y \geq 0\ \forall y \in R^n$, alors x_0 est un minimum local strict de f.

Remarque

1. Si Hf est seulement positive et non définie positive, on n'est pas assuré d'avoir un minimum local

exemple : $f : R \rightarrow R\ x \rightarrow x^3$, alors $x^{-} = 0$, $Hf(^{-}x) = 0 \geq 0$ et pourtant on n'a pas d'extremum en 0

2. En dimension 2, soit $A = \begin{pmatrix} r & s \\ s & t \end{pmatrix} \in M_2(R)$
   a. Si $\det(A) = rt - s^2 > 0$ alors les *vap* λ_1 et λ_2 de A sont de même signe car $\det(A) = \lambda_1\lambda_2$. De plus, $tr(A) = \lambda_1 + \lambda_2 = r + t$,et comme *r* et *t* sont de même signe $\det(A) = rt - s^2 > 0$, on en déduit que $\lambda_1 + \lambda_2$ a le signe de r. Les *vap* ont donc le signe de r.
   Donc

– det(A) > 0, A est définie positive ⇔ r > 0, A est définie négative ⇔ r < 0
– det(A) < 0, λ_1 et λ_2 sont de signes opposés donc A n'est ni définie positive ni définie négative

b. Considérons f : U ⊂ $R^2 \to R$ C^2 telle que ∇f$(\bar{x}, \bar{y})$=0 avec $(\bar{x}, \bar{y}) \in$ U.
Posons $\frac{\partial^2 f}{\partial x^2}(\bar{x}, \bar{y}), r = \frac{\partial^2 f}{\partial y^2}(\bar{x}, \bar{y})\ et\ s = \frac{\partial^2 f}{\partial x \partial y}(\bar{x}, \bar{y})$
Alors Hf$(\bar{x}, \bar{y})$ = A et donc , on a la proposition suivante :

Proposition
– Si $rt - s^2 > 0$ et r > 0, f a un minimum local en $(\bar{x}, \bar{y})$.
– Si $rt - s^2 > 0$ et r < 0, f a un maximum local en $(\bar{x}, \bar{y})$.
– Si $rt - s^2 = 0$, on ne peut pas conclure.

Exemple
Soit $f(x_1,x_2) = m\, x_1^2 + x_2^2$ définie de R^2 dans R, avec m ≠ 0.
Alors $\nabla f(x_1,x_2) = (2mx_1,2x_2)$.
On cherche les $x = (x_1,x_2)$ tels que ∇f(x)=0 qui équivaut a $\begin{cases} 2mx_1 = 0 \\ 2x_2 = 0 \end{cases}$

$$Hf(x_0) = \begin{pmatrix} 2m & 0 \\ 0 & 2 \end{pmatrix} \in M_2(R)$$

– si m > 0, alors Hf(x_0) est définie positive (vap > 0) donc par la condition suffisante x_0 = (0,0) est un minimum local strict.
– si m < 0, alors les vap n'ont pas le même signe donc par la condition nécessaire, x_0 = (0,0) ne peut pas être un minimum local.

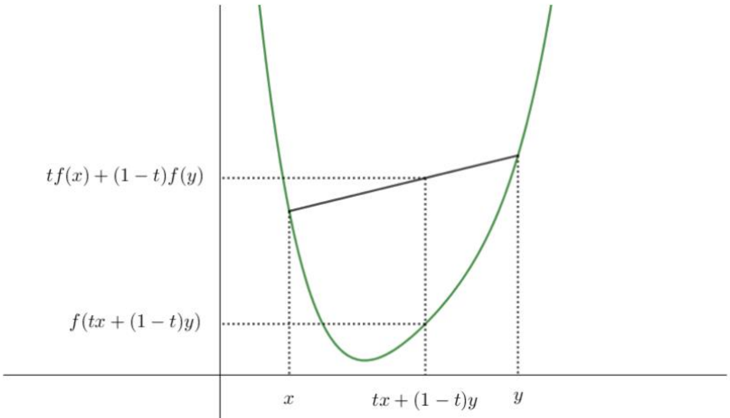
2.3.4 Fonctions convexes

Une fonction f est convexe sur un intervalle I si, pour tous x et y de I , pour tout de [0,1] :
f(tx + (1−t) y) ≤ tf(x) + (1−t) f(y).

Elle est strictement convexe si

$\forall$(x,y) ∈ I^2, x ≠ y, $\forall$t∈]0,1[, f(tx+(1−t)y)<tf(x)+(1−t)f(y).

Interprétation graphique :

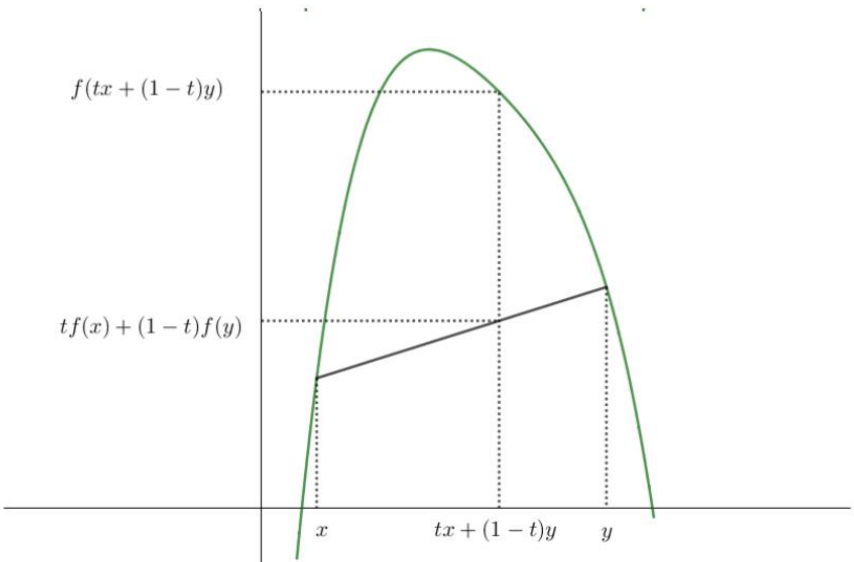


Une fonction est convexe si et seulement si sa courbe représentative est au-dessous de ses cordes

Une fonction f est concave si – f est convexe, c'est-à-dire si, pour tous x et y de I, pour tout t de [0 , 1] :

$$f(tx + (1-t)\,y) \geq tf(x) + (1-t)\,f(y)$$

Interprétation graphique :

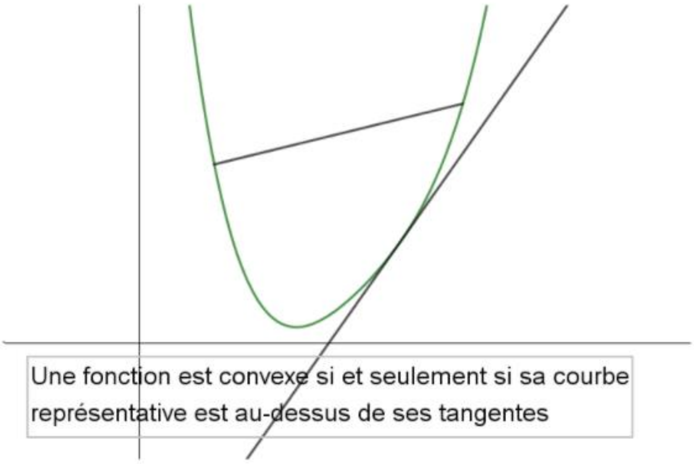


Une fonction est concave si et seulement si sa courbe représentative est au-dessus de ses cordes

La convexité se lit sur les dérivées de f :

Proposition:

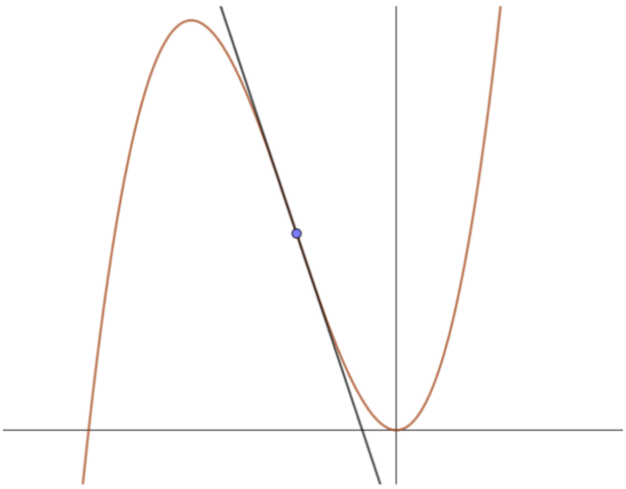
- Si f est dérivable sur II, f est convexe si et seulement si f' est roissante.
- Si f est deux fois dérivable sur II, f est convexe si et seulement si $f'' \geq 0$.



Définition : Si f est dérivable en x_0, x_0 est un point d'inflexion pour f si la tangente au point $(x_0, f(x_0))$ traverse la courbe représentative de f en ce point.

En particulier, si f est deux fois dérivable en x_0, et si x_0 est un point d'inflexion de f, $f''(x0) = 0$.

Exemple :



En un point d'inflexion, la tangente traverse
la courbe représentative.

3. Méthode de décente de gradient

Principe

On se place dans un espace de Hilbert V, et on cherche à calculer numériquement un x (qui n'est pas forcément unique) tel que

$\forall y \in V, f(x) \leq f(y)$

Le principe est de construire un algorithme itératif de la forme

$x^{k+1} = x^k - \rho_k d^k$

d^k est la direction de descente, ρ_k est le pas. Il est, soit fixé, éventuellement le même pour toutes les étapes (on parle alors de méthode à pas fixe), soit calculé à chaque étape de façon à minimiser f dans la direction d^k (on parle alors de méthode à pas optimal).

3.1.Méthode du gradient

Ici on choisit à chaque étape $d^k = \nabla J(x^k)$.

- **Méthode à pas variable**

On se donne le pas ρ_k, il peut être différent d'une étape à l'autre.

Théorème. Si f est convexe dérivable sur V , l'algorithme de gradient à pas variable converge vers la solution optimale pour $0 < a \leq \rho_k \leq b$

- **Méthode à pas optimal**

Ici on choisit à chaque étape ρ_k de façon que

$$f(x^k - \rho_k \nabla J(x^k)) = \inf_{\rho \in R} f(x^k - \rho \nabla J(x^k))$$

Théorème.

Si f est convexe dérivable sur V, l'algorithme de gradient à pas optimal est bien défini et converge vers la solution optimale.

3.1.1. Estimations et convergence dans le cas quadratique

- **Méthode à pas optimal**

On prend ici une direction de descente d^k quelconque dans R^n, non orthogonale à r^k. A chaque étape, la valeur du paramètre optimal ρ_k est donnée par

$$\rho_k = \frac{(r^k, d^k)}{(Ad^k, d^k)}$$

et l'on a $(r^{k+1}, d^k) = 0$.

Notons $E(v) = 1/2\,(A(v-u), v-u)$,

on a alors

$$E(x^{k+1}) = (1 - \gamma k)E(x^k)$$

avec $\gamma k = \frac{1}{2} \frac{(r^k, d^k)^2}{(Ad^k, d^k)(A^{-1}r^k, r^k)}$.

Puisque la quantité γk est par construction telle que $0 \leq \gamma k \leq 1$, on a l'estimation suivante :

si la direction de descente est telle que

$$\left(\frac{r^k}{||r^k||}, \frac{d^k}{||d^k||}\right)^2 \geqslant \mu > 0$$

alors $\gamma_k > \gamma = \mu / K(A)$ (où K(A) est le conditionnement de A, c'est-à-dire le rapport de la plus grande à la plus petite valeur propre), et donc

$$E(x^{k+1}) \leq (1 - \gamma)\, E(x^k)$$

On dit que la méthode **converge linéairement**.

- **Méthode de gradient à pas constant**

On choisit à chaque étape $\rho_k = \rho$. On a alors l'estimation

$$||x^k - x||^2 \leq \max_{1 \leq i \leq n} |1 - \rho\lambda_i|^k ||x^0 - x||^2$$

On en déduit que la méthode converge si et seulement si $\rho < 2/\ \lambda_n$ où λ_n est la plus grande valeur propre de A. Ici encore, la convergence est linéaire

3.2. Méthode du gradient conjugué

L'idée de la méthode est de construire itérativement des directions $d_0, \ldots, d_k$ mutuellement conjuguées. A chaque étape k la direction d_k est obtenue comme combinaison linéaire du gradient en x^k et de la direction précédente d^{k-1}, les coefficients étant choisis de telle manière que d^k soit conjuguée avec toutes les directions précédentes. Si l'on note $g^k = \nabla f(x^k)$, l'algorithme prend la forme suivante

On se donne x^0 et on pose $d^0 = -g^0$

$$x^{k+1} = x^k + \rho^k d^k,$$

avec $\rho^k = -\frac{g_k^T d^k}{d_k^T B d^k}$, cette formule définit bien le pas optimal

$$d^{k+1} = -g^{k+1} + \beta^k d^k,$$

avec $\beta^k = \frac{g_{k+1}^T B d^k}{d_k^T B d^k}$

En effet on a bien

$$\nabla f(x^{k+1})^T\, d^k = g^{kT}\, d^k + \rho^k d^{k\,T} B\, d^k = 0$$

- ✓ **Convergence de la méthode du gradient conjugué**

Le résultat suivant va nous permettre de retrouver d'une autre manière la propriété de convergence finie de l'algorithme du GC :

Proposition Soit A une matrice définie positive et x^k le vecteur obtenu à l'étape k de l'algorithme du GC. Alors on a $E(x^k) \leq E(x^0)$ min p∈Pk,p(0)=1 max z∈σ(A) p(z) ^{2} .