PSI* 2020/21 Rev 3 correction

Coercivité, méthode du gradient

Epreuve X-ENS 2012 PSI.

Préambule

1. On écrit la définition de la limite infinie en posant $A = |f(0)| + 1$. Il existe $M > 0$ tel que pour tout $x \in \mathbb{R}^n$, $\|x\| > M \Rightarrow f(x) \geq |f(0)| + 1$.

2. Soit K la boule fermée de $\mathbb{R}^n$ de centre 0 et de rayon M. Il s'agit d'une partie fermée de $\mathbb{R}^n$ qui est de dimension finie. C'est donc un fermé borné, et f est continue sur cette partie, donc admet une borne inférieure, atteinte en au moins un point. On note x^* un tel point.

On a déjà pour tout $x \in K$, $f(x^*) \leq f(x)$. Notamment $f(x^*) \leq f(0) < |f(0)| + 1$.

Alors, pour tout $x \notin K$, on a $\|x\| > M$ et par suite $f(x) \geq |f(0)| + 1 > f - x^*)$. La propriété $f(x^*) \leq f(x)$ est prouvée pour tout x.

$\boxed{\text{Une fonction coercive de } \mathbb{R}^n \text{ dans } \mathbb{R} \text{ présente un minimum global.}}$

3. Comme f est une fonction de classe $\mathcal{C}^1$ sur $\mathbb{R}^n$ qui est un ouvert, tout extremum (même local) de f est un point critique : $\boxed{\nabla f(x^*) = 0.}$

J'ai considéré - ça se discute - qu'on ne demandait pas de preuve mais l'énoncé précis du théorème, avec le mot "ouvert". Sinon on peut dire : chaque $h \mapsto f(x_1^*, \ldots, x_i^* + h, \ldots, x_n^*)$ *admet un extremum en* $h = 0$*, donc une dérivée nulle :* $\frac{\partial f}{\partial x_i}(x^*) = 0$.

Partie 1

4. On utilise Cauchy-Schwarz pour encadrer le terme (b, x) : $|(b, x)| \leq \|b\|.\|x\|$.

Par suite, $g(x) \geq \frac{1}{2} C \|x\|^2 - \|b\|.\|x\| = \|x\| \big(C\|x\| - \|b\| \big)$.

Pour $\|x\| > \frac{2\|b\|}{C}$, on voit que $\|g(x)\| \geq \|x\|.\|b\|$, et cela tend vers l'infini lorsque $\|x\|$ tend vers l'infini.

$\boxed{\text{La fonction } g \text{ est coercive.}}$

5. On peut tout exprimer en coordonnées et dériver par rapport à chaque variable (il y aurait un calcul plus adapté par DL global, mais c'est moins l'optique du programme actuel). En regroupant les termes identiques par symétrie de A,

$$g(x_1, x_2, x_3) = \frac{1}{2}(a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2) + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 - b_1x_1 - b_2x_2 - b_3x_3$$

Par sommes et produits de fonctions coordonnées, g est une fonction $\mathcal{C}^1$ sur $\mathbb{R}^3$, avec

$$\nabla g(x_1, x_2, x_3) = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 - b_3 \end{pmatrix}$$

On reconnaît ce vecteur gradient : c'est $\boxed{\nabla g(x) = Ax - b.}$

Comme g est coercive, le préliminaire montre qu'elle possède un minimum global x^*. De plus, en un tel minimum, le gradient de g s'annule. Le ou les points où g atteint son minimum sont à chercher parmi les solutions de l'équation $Az = b$.

Or la relation $Az = b$ est un système linéaire. Examinons les propriétés de la matrice A. Si x est un vecteur de $\ker A$, $(Ax, x) = 0$ et par suite $C\|x\|^2 \leq 0$. Puisque $C > 0$ cela montre que $x = 0$. Ainsi A est injective.

$\boxed{\text{Le système } Az = b \text{ admet une unique solution } x^* = A^{-1}b, \text{ unique extremum local de } g, \text{ qui est en fait minimum global.}}$

6. Calculons comme indiqué

$$u_{k+1} - x^* = u_k - \alpha A u_k + \alpha b - x^* = (u_k - x^*) - \alpha A(u_k - x^*) = (\mathrm{I}_3 - \alpha A)(u_k - x^*).$$

7. Une récurrence évidente permet d'établir que $u_k - x^* = (\mathrm{I}_3 - \alpha A)^k (u_0 - x^*)$.

On sait que comme A est symétrique réelle, elle est diagonalisable dans une base orthonormale $B = (e_1, \ldots, e_2, e_3)$. On note P la matrice de passage de la base canonique à la base B. Soient $\lambda_1, \lambda_2, \lambda_3$ les valeurs propres de A. Observons d'abord que l'hypothèse faite sur A donne une contrainte sur ces valeurs propres : $(Ae_i, e_i) = \lambda_i(e_i, e_i) = \lambda_i \geq C > 0$. Par ailleurs,

$$\mathrm{I} - \alpha A = \mathrm{I} - \alpha P D P^{-1} = P D' P^{-1}, \qquad \text{où } D = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix}, \qquad D' = \begin{pmatrix} 1 - \alpha\lambda_1 & & \\ & 1 - \alpha\lambda_2 & \\ & & 1 - \alpha\lambda_3 \end{pmatrix}.$$

Pour chaque i, on a $0 < \alpha\lambda_i \leq \alpha L < 2$. Ainsi $-1 < 1 - \alpha\lambda_i < 1$.

La matrice D a tous ses coefficients diagonaux dans $]-1, 1[$. On peut affirmer que les puissances $(D')^k$ ont une limite nulle et par suite $(\mathrm{I} - \alpha A)^k$ tend vers 0.

On trouve enfin que $\boxed{\text{la suite } (u_k) \text{ converge vers } x^*.}$

Partie 2

8. On envisage les trois cas

- si $x_k = 0$, alors $x_{k+1} = 0$ et ainsi $h(x_k) = h(x_{k+1})$.
- si $x_k > 0$, alors $x_{k+1} = x_k - t_k$ et ainsi $-x_k < x_{k+1} < x_k$. On en tire $0 < |x_{k+1}| < |x_k|$, puis par croissance de $h : x \mapsto x^2$ sur $\mathbb{R}_+$, $h(x_{k+1}) < h(x_k)$.
- si $x_k < 0$, alors $x_{k+1} = x_k + t_k$ et ainsi $x_k < x_{k+1} < -x_k$. On en tire $0 < |x_{k+1}| < |x_k|$, et de nouveau $h(x_{k+1}) < h(x_k)$.

Dans tous les cas, $\boxed{\forall k \in \mathbb{N}, h(x_{k+1}) \leq h(x_k).}$

9. On peut calculer le terme général de cette suite

$$v_k = \sum_{p=1}^{k} (v_p - v_{p-1}) + v_0 = 2 - \sum_{p=1}^{k} \frac{1}{2^p} = 2 - \left(1 - \frac{1}{2^k}\right) = 1 + \frac{1}{2^k}$$

Les termes de la suite sont tous positifs. On pose donc pour tout k, $\varepsilon_k = -1$ et $t_k = \frac{1}{2^{k+1}} \leq \frac{1}{2} < 2|v_k|$. On a bien $v_{k+1} = v_k + t_k\varepsilon_k$ avec les conditions désirées sur t_k et ε_k. La suite forme bien une descente par gradient pour h. Cependant sa limite est 1 qui n'est pas un point où h atteint son minimum global.

$\boxed{\text{On a trouvé un exemple de descente par gradient qui converge vers un point autre que le minimum global.}}$

10. De nouveau le calcul complet est possible, et utile pour trouver le signe de w_k.

$$w_k = \sum_{p=1}^{k} (w_p - w_{p-1}) + w_0 = 2 + 2\sum_{p=1}^{k} (-1)^p + 3\sum_{p=1}^{k} \left(\frac{-1}{2}\right)^p = 2\sum_{p=1}^{k} (-1)^p + 1 + \frac{(-1)^k}{2^k} = \begin{cases} 1 + \frac{1}{2^k} & \text{si k est pair} \\ -1 - \frac{1}{2^k} & \text{si k est impair} \end{cases}$$

La sous-suite des termes d'indices pairs (w_{2k}) converge vers 1, celle des termes d'indice impair vers -1. Donc la suite (w_k) diverge.

On pose cette fois $t_k = 2 + \frac{3}{2^{k+1}} < 2\left(1 + \frac{1}{2^k}\right)$ et $\varepsilon_k = (-1)^{k+1}$ qui est de signe opposé à x_k. On a bien les hypothèses pour affirmer que (w_k) forme une descente par gradient.

$\boxed{\text{On a trouvé un exemple de descente par gradient qui ne converge pas.}}$

Partie 3

11. Si le vecteur $u = \nabla f(x)$ est non nul, on forme $d = -\frac{u}{\|u\|}$ et on a bien les relations $\|d\| = 1$ et $(d, \nabla f(x)) = -\frac{\|u\|^2}{\|u\|} < 0$.

$\boxed{\text{Si } x \text{ n'est pas un point critique, } D_x \text{ est non vide (et réciproquement !).}}$

Prenons $d \in D_x$ et formons le DL de $f(x+td)$: on étudie la fonction le long d'une droite d'origine x et dirigée par d (expression $\phi(t) = f(\gamma(t))$ avec $\gamma(t) = x + td$). Comme f est $\mathcal{C}^1$,

$$\phi'(t) = \frac{\mathrm{d}}{\mathrm{d}t} f(x+td) = \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t)) = (\gamma'(t)|\nabla f(\gamma(t)) = (d, \nabla f(x+td))$$

Rappel : si on n'est pas à l'aise avec cette formule, on peut le considérer comme $f(x_1 + td_1, x_n + td_n)$.

Notamment en $t = 0$, cette dérivée vaut $\phi'(0) = (d, \nabla f(x))$, et on note cette valeur α, avec $\alpha < 0$ par définition de d.

$$f(x+td) = f(x) + t(d, \nabla f(x)) + o(t) = f(x) + \alpha t + o(t).$$

Reprenons la définition de $o(t)$: il existe $\eta > 0$ tel que si $t \in]0, \eta[$, alors le terme en $o(t)$ est majoré par $\frac{1}{2}|\alpha|t = -\frac{1}{2}\alpha t$. Ainsi pour $t \in]0, \eta[$, $f(x+td) - f(x) < \alpha t - \frac{1}{2}\alpha t < 0$.

L'intervalle $]0, \eta[$ appartient à $T_{d,x}$.

$\boxed{\text{Pour tout } d \in D_x, T_{d,x} \text{ est non vide. Précisément, il existe } \eta_d > 0 \text{ tel que }]0, \eta_d[\subset T_{d,x}.}$

NB : il est important de réaliser que c'est pour chaque d qu'on trouve un η_d (raisonnement uniquement le long d'une direction) ; rien ne dit qu'il existe un η valable pour tous les $d \in D$.

12. Attention quand on dit "coïncide" ça va dans les deux sens : dans le cas de la fonction h on a deux définitions (celle de la partie 2 et celle de la partie 3) ; il faut montrer que (x_k) vérifie l'une si et seulement si elle vérifie l'autre.

Commençons par poser un peu le cadre : ici $n = 1$, $\mathbb{R}^n = \mathbb{R}$, le produit scalaire est le produit ordinaire, le gradient de h est $h'(x) = 2x$, les seuls vecteurs de norme 1 sont 1 et -1.

Le cas $x_k = 0$ donne lieu à la même convention pour les deux définitions : $x_{k+1} = x_k$. Supposons maintenant $x_k \neq 0$.

La relation $x_k > 0$ revient au même que $(\nabla h(x_k)|1) > 0$, donc $\varepsilon_k = -1$ revient au même que $d_k = -1$. De même $\varepsilon_k = 1$ revient au même que $d_k = 1$.

Comparons maintenant ε_k et t_k. Les scalaires t_k sont choisis dans $\mathbb{R}_+^*$ pour respecter la condition

$$(x_k + t_k\varepsilon_k)^2 < x_k^2 \Leftrightarrow 2t_k x_k \varepsilon_k + t_k^2 < 0 \Leftrightarrow t_k(2x_k\varepsilon_k + t_k) < 0 \Leftrightarrow t_k < -2x_k\varepsilon_k = 2|x_k|,$$

puisque x_k et ε_k sont de signe opposé. On a prouvé, pour des $t_k > 0$, l'équivalence des conditions $t_k \in]0, 2|x_k|[$ et $t_k \in T_{x_k, d_k}$.

$\boxed{\text{Les deux définitions de la descente par gradient proposées pour la fonction } h \text{ coïncident.}}$

13. La suite des valeurs $f_k = f(x_k)$ est décroissante puisqu'on a soit $x_{k+1} = x_k$, soit $f(x_{k+1}) < f(x_k)$. En outre (f_k) est minoré par le minimum global $f(x^*)$ dont on a prouvé l'existence en préambule. La suite (f_k), décroissante et minorée, est donc convergente.

La suite (f_k) est donc bornée par une constante $K > 0$. En reprenant l'argument de la question 1, il existe une valeur $N > 0$ telle que $\|x\| > N \Rightarrow f(x) > K$. Ainsi les vecteurs x_k vérifient tous $\|x_k\| \leq N$.

$\boxed{\text{La suite } (x_k) \text{ est bornée ; la suite } (f(x_k)) \text{ est convergente.}}$

14. Calculons comme demandé, en faisant systématiquement apparaître r_k : $u_{k+1} = u_k - \alpha r_k$ et $Au_k = r_k + b$:

$$\begin{aligned} g(u_{k+1}) - g(u_k) &= \frac{1}{2}(A(u_k - \alpha r_k), u_k - \alpha r_k) - (b, u_k - \alpha r_k) - \frac{1}{2}(Au_k, u_k) + (b, u_k) \\ &= -\alpha(Au_k, r_k) + \frac{1}{2}\alpha^2(Ar_k, r_k) + \alpha(b, r_k) = -\alpha\|r_k\|^2 + \frac{1}{2}\alpha^2(Ar_k, r_k). \end{aligned}$$

vérifions qu'on retrouve bien la définition de la descente par gradient :

- si $\nabla g(u_k) = 0$, alors $u_{k+1} = u_k$

- sinon, on pose $d_k = \frac{-\nabla g(u_k)}{\|\nabla g(u_k)\|}$. On trouve bien que $d_k \in D_{u_k}$. On pose encore $t_k = \alpha\|\nabla g(u_k)\| > 0$. La propriété restant à vérifier pour la descente par gradient est que

$$\forall k \in \mathbb{N}, g(u_{k+1}) < g(u_k), \qquad \text{c'est-à-dire } \alpha < \frac{2\|r_k\|^2}{(Ar_k, r_k)}$$

Or d'après les propriétés de A, ce quotient vérifie $\frac{2\|r_k\|^2}{(Ar_k, r_k)} \leq \frac{2\|r_k\|^2}{C\|r_k\|^2} = \frac{2}{C}$.

Ainsi, $\boxed{\text{si on demande } 0 < \alpha < \frac{2}{C}\text{, alors } (u_k) \text{ est une suite de descente par gradient pour la fonction } g.}$

Partie 4

15. On peut estimer la quantité étudiée $\rho_k = t_k(d_k, \nabla f(x_k))$ qui est négative. Elle vérifie $\rho_k \geq \frac{f(x_{k+1})-f(x_k)}{m_1}$ puisque $m_1 > 0$.

Comme la suite $(f(x_k))$ est convergente, elle est bornée, et c'est également le cas des différences $f(x_{k+1}) - f(x_k)$. Ainsi (ρ_k) est minorée.

Mieux : si on forme la série $\sum\limits_{k=0}^{K} \rho_k \geq \frac{f(x_{K+1})-f(x_0)}{m_1}$: la somme partielle est également minorée. Or on sait que soit cette série à termes négatifs converge, soit la somme partielle tend vers $-\infty$. On peut donc affirmer que $\sum \rho_k$ converge et par suite que ρ_k tend vers 0.

$\boxed{\text{Sous l'hypothèse (2), } \rho_k = t_k(d_k, \nabla f(x_k)) \text{ a une limite nulle.}}$

16. S'il existe un entier k tel que $(d_k, \nabla f(x_k)) = 0$ on a vu que cela signifiait que $\nabla f(x_k)$ et $x_{k+1} = x_k$. . Dans ce cas, la suite est constante à partir d'un certain rang, et $(d_k, \nabla f(x_k)) = 0$ à partir d'un certain rang.

Prenons maintenant le cas le plus général : pour tout k, $(d_k, \nabla f(x_k)) \neq 0$. On va manipuler des valeurs absolues pour que les termes soient positifs.

Si $t_k \geq C_1$, alors $|(d_k, \nabla f(x_k))| \leq \frac{1}{C_1} t_k |(d_k, \nabla f(x_k))|$.

Sinon, on sait que $t_k \geq C_2 |(d_k, \nabla f(x_k))|$ et ainsi $t_k |(d_k, \nabla f(x_k))| \geq C_2 |(d_k, \nabla f(x_k))|^2$.

Réunissons tout cela en une seule majoration :

$$|(d_k, \nabla f(x_k))| \leq \max\left(\frac{1}{C_1} t_k |(d_k, \nabla f(x_k))|, \sqrt{\frac{1}{C_2} t_k |(d_k, \nabla f(x_k))|}\right)$$

Le majorant a une limite nulle.

$\boxed{\text{Sous les hypothèses (2) et (3), } (d_k, \nabla f(x_k)) \text{ tend vers 0.}}$

17. On étudie le terme $(BX|X)$ pour le vecteur $X = \nabla f(x_k) \neq 0$. Comme B est symétrique réelle, elle est diagonalisable en base orthonormale B. On note $\mu_1 \leq \cdots \leq \mu_n$ les valeurs propres rangées par ordre croissant, et on sait que $\mu_1 > 0$.

En développant X dans la base $B = (e_1, \ldots, e_n)$, il vient $X = \sum\limits_{i=1}^{n} x_i e_i$, $BX = \sum\limits_{i=1}^{n} x_i \mu_i e_i$ et comme la base est orthonormale

$$(BX, X) = \sum_{i=1}^{n} x_i^2 \mu_i \geq \sum_{i=1}^{n} x_i^2 \mu_1 = \|X\|^2 \mu_1$$

De même on trouve $\|BX\|^2 \leq \mu_n^2 \|X\|^2$.

Dès lors, $|(d_k, \nabla f(x_k))| = \frac{1}{\|BX\|} |(BX, X)| \geq \frac{\mu_1}{\mu_n} \|X\| = \frac{\mu_1}{\mu_n} \|\nabla f(x_k)\|$.

$\boxed{\text{Sous ces hypothèses, } \nabla f(x_k) \text{ tend vers 0.}}$

18. L'existence a déjà été prouvée ; montrons l'unicité par l'absurde, en supposant que $\tilde{x}$ est un minimum global, distinct de x^*. On se place alors au milieu en posant $\lambda = \frac{1}{2}$:

$$f(\frac{1}{2}x^* + \frac{1}{2}\tilde{x}) < \frac{1}{2}f(x^*) + \frac{1}{2}f(\tilde{x}) = f(x^*).$$

On a trouvé un point pour lequel la valeur de f est strictement inférieure au minimum global ; c'est absurde. On n'a pas utilisé l'hypothèse $\mathcal{C}^2$.

$\boxed{\text{Dans le cas } f \text{ strictement convexe et coercive, il y a unicité du minimum global.}}$

On prouve, en vue de la question suivante, les propriétés issues de la stricte convexité : on compare d'abord des pentes

$$\text{si } z = \lambda x_1 + (1-\lambda)x_2, \qquad \lambda(f(z) - f(x_1)) < (1-\lambda)(f(x_2) - f(z))$$

et encore

$$\lambda(z-x_1)\frac{f(z)-f(x_1)}{z-x_1} < (1-\lambda)(x_2-z)\frac{f(x_2)-f(z)}{x_2-z}$$

Mais comme $\lambda(z-x_1)=(1-\lambda)(x_2-z)>0$, il vient $\frac{f(z)-f(x_1)}{z-x_1} < \frac{f(x_2)-f(z)}{x_2-z}$.

On a prouvé cela pour tout z de la forme $z=\lambda x_1+(1-\lambda)x_2$, c'est-à-dire tout point de $]x_1,x_2[$. L'implication sur les pentes est donc

$$x_1 < z < x_2 \Rightarrow \frac{f(x_1)-f(z)}{x_1-z} < \frac{f(x_2)-f(z)}{x_2-z}$$

Si on passe à la limite quand x_2 tend vers z, il vient $x_1 < z \Rightarrow \frac{f(x_1)-f(z)}{x_1-z} \leq f'(z)$ et ensuite si x_1 tend vers z, $f'(x_1)\leq f'(z)$. C'est la croissance de f'.

Mais on a aussi établi $x_1 < z \Rightarrow f(x_1) \leq (x_1-z)f'(z)+f(z)$, c'est-à-dire que le graphe est au-dessus de la tangente (d'un côté, mais l'autre se traite en faisant jouer un rôle symétrique à x_2).

Avec la (stricte) convexité donnée, il y a (stricte) croissance de f' et le graphe est au-dessus des tangentes.

19. On est dans le cas $n=1$, avec ce qui a été dit au-dessus sur ce que deviennent les notations dans ce cadre. Commençons par le second point.

Par la question 16, on a $d_k f'(x_k)$ qui tend vers 0 avec $d_k=\pm 1$, donc $f'(x_k)$ qui tend vers 0. La suite (x_k) est bornée (question 13) donc elle admet au moins une sous-suite convergente (Bolzano-Weierstrass, admis ici). Soit une telle sous-suite $x_{\varphi(k)}$ et soit x_∞ sa limite. Alors $f(x_k)$ converge vers $f(x_\infty)$ par continuité et de même $f'(x_k)$ converge vers $f'(x_\infty)$. Ainsi $f'(x_\infty)=0$. Par convexité, la fonction f est au-dessus de la tangente en x_∞, donc de la droite horizontale $y=f(x_\infty)$. Ceci montre que x_∞ est l'unique minimum global de f.

Toute suite extraite de (x_k) qui converge converge vers x_*. Comme il est demandé de l'admettre, on peut énoncer que cela prouve que x_k converge vers x^*.

Avec les hypothèses de la question 18, ainsi que (2) et (3), la suite (x_k) converge vers x^*.

Voyons l'algorithme... ça ne paraît pas simple ! Voici une proposition incomplète :

1. Je délimite un domaine borné contenant x_0 et x^*. Pour cela on peut exploiter la monotonie de f' : je cherche a et b tels que $f'(a)<0, f'(b)>0$. Quitte à l'agrandir, le segment $[a,b]$ contient x_0.

2. J'appelle L un majorant de $|f''|$ sur $[a,b]$. J'ai donc une formule de Taylor :

$$f(x_{k+1}) = f(x_k)+t_k d_k f'(x_k)+R_k = f(x_k)-t_k|f'(x_k)|+R_k, \qquad \text{où } |R_k| \leq \frac{1}{2}t_k^2 L$$

On souhaite avoir t_k vérifiant une condition du type $-t_k|f'(x_k)|+R_k \leq -m_1 t_k|f'(x_k)|$.

Il suffit pour cela d'avoir $m_1 t_k|f'(x_k)|+\frac{1}{2}t_k^2 L \leq t_k|f'(x_k)|$.

Je choisis donc $m_1=\frac{1}{2}$ et j'impose $t_k=\frac{1}{L}|f'(x_k)|$. Cela me donne (2) (et notamment qu'il s'agit d'une descente par gradient), et (3).

3. Il y a un point qui n'est pas évident : est-ce que je reste dans $[a,b]$? Pour le garantir, il me semble que le plus accessible est d'utiliser la monotonie de f. Par exemple si on arrive à se débrouiller pour partir avec a,b tels que $f(a)=f(b)$, il est sûr que $f(x_k)\leq f(a)$ et donc que x_k reste dans $[a,b]$. Mais je ne sais pas si on attendait un questionnement aussi précis...