253 UTILISATION DE LA NOTION DE CONVEXITÉ EN ANALYSE.

Soit E un $\mathbb{R}$-espace vectoriel normé.

I. Généralités autour de la convexité

[Gou08, §1.5/2.3, p51/94] [BMP05, §1.5.1, p26]

I. A. Ensembles et fonctions convexes

Définition 1. $A \subset E$ est convexe si $\forall a, b \in A, \forall t \in [0,1], (1-t)a + tb \in A$.

Exemple 2. Un intervalle de $\mathbb{R}$ est convexe, une boule de E est convexe.

Définition 3. $f : E \longrightarrow \mathbb{R}$ est dite
- convexe si $\forall a, b \in E, \forall t \in]0,1[, f((1-t)a+tb) \leq (1-t)f(a) + tf(b)$,
- strictement convexe si $\forall a, b \in E, \forall t \in]0,1[, f((1-t)a+tb) < (1-t)f(a) + tf(b)$,
- concave si $-f$ est convexe.

Exemple 4. Une application affine est convexe et concave.

Exemple 5. Toute norme $\|.\|$ de E est convexe, non strictement convexe dès que $E \neq \{0\}$.

Proposition 6. *$f : E \longrightarrow \mathbb{R}$ est convexe si et seulement si*
- *$E_f = \{(x,y) \in \mathbb{R}^2 \mid f(x) \leq y\}$ est convexe,*
- *$\forall n \in \mathbb{N}^*, \forall (a_1, \dots, a_n) \in E^n, \forall (t_1, \dots, t_n) \in (\mathbb{R}_+)^n \setminus \{0_{\mathbb{R}^n}\}, f(\mathrm{Bar}((a_i, t_i)_i)) \leq \mathrm{Bar}((f(a_i), t_i)_i))$.*

Proposition 7. *Soit $f : I \longrightarrow \mathbb{R}$ où I est un intervalle de $\mathbb{R}$. f est convexe si et seulement si :*
- *pour tout $a \in I$, $x \longmapsto \frac{f(x)-f(a)}{x-a}$ est croissante sur $I \setminus \{a\}$,*
- *pour tout $a, x \in I$, $f'(a)(x-a) + f(a) \leq f(x)$,*
- *f' est croissante,*
- *$f'' \geq 0$.*

les trois dernières équivalences ayant lieu lorsque f est assez régulière ($\mathcal{C}^1$ ou $\mathcal{C}^2$).

On déduit de la première équivalence qu'une fonction convexe sur I est continue sur $\mathring{I}$ et admet une dérivée à droite et à gauche en tout point de $\mathring{I}$.

Exemple 8. $\exp$ est convexe, $\log$ est concave.

Exemple 9. Fonction convexe non continue sur le bord : voir annexe.

I. B. Inégalités de convexité

Exemple 10. On a des inégalités provenant de la convexité/concavité de certaines fonctions :

$$\forall x \in \mathbb{R}, \mathrm{e}^x \geq 1 + x \qquad \forall y \in]-1, +\infty[, \ln(1+y) \leq y \qquad \forall \theta \in \left[0, \frac{\pi}{2}\right], \frac{2}{\pi}\theta \leq \sin(\theta) \leq \theta$$

Proposition 11. [Inégalité arithmético-géométrique]
Soient $n \in \mathbb{N}^*$ et $a_1, \dots, a_n \in \mathbb{R}_+$. Alors $\sqrt[n]{a_1 \dots a_n} \leq \frac{1}{n}\sum_{i=1}^n a_i$.

Exemple 12. On peut généraliser pour toute famille $a_1, \dots, a_n \in \mathbb{R}_+$ et $b_1, \dots, b_n \in \mathbb{R}_+^*$, on a

$$(a_1^{b_1} \dots a_n^{b_n})^{1/b} \leq \frac{1}{b}\sum_{i=1}^n b_i a_i \qquad \text{où} \quad b = \sum_{i=1}^n b_i$$

Proposition 13. [Inégalité de Young]
Soient $p, q > 0$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Alors pour tout $a, b \in \mathbb{R}_+$, on a $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$ avec égalité si et seulement si $a^p = b^q$.

II. Exemples d'applications dans certains espaces

II. A. Espaces de Hilbert

[BMP05, §3.1.2, p95] [Bre99, ChV/VIII, p79–82/136]

Soit $(H, \langle . \mid . \rangle)$ un espace de Hilbert.

Théorème 14. [Projection sur un convexe fermé]
Soit C un convexe fermé de Hilbert H. Alors :

$$\forall x \in H, \exists! p \in C \mid \|x - p\| = d(x, C)$$

De plus, p est l'unique élément de C satisfaisant $\forall c \in C, \Re(\langle x - p \mid c - p \rangle) \leq 0$.

Remarque 15. On a en fait juste besoin de C complet et H préhilbertien.

Exemple 16. Si C n'est pas convexe : soit $H = \mathbb{R}$ et $C = \mathbb{R} \setminus]-1, 1[$. Alors 1 et -1 minimisent la distance de 0 à C.

Corollaire 17. *Soit F un sous-espace vectoriel fermé de H. Alors $H = F \oplus F^\perp$.*

Application 18. Existence de l'espérance conditionnelle d'une variable aléatoire L^2.

Théorème 19. [Théorème de Riesz-Fréchet]
Pour toute application $\phi \in H'$, il existe un unique $f \in H$ tel que $\forall v \in H, \phi(v) = \langle f \mid v \rangle$.
De plus $\phi \longmapsto f$ est une isométrie ($\|f\|_H = \|\phi\|_{H'}$).

Application 20. Existence et unicité de l'adjoint d'un opérateur $T \in \mathcal{L}(H)$.

Théorème 21. [Théorème de Lax-Milgram]
Soit $a : H^2 \longrightarrow \mathbb{R}$ une forme bilinéaire continue et coercive. Alors pour tout $\ell \in H'$, il existe un unique $u \in H$ tel que $\forall v \in H, a(u,v) = \ell(v)$.
Si de plus a est symétrique alors u est l'unique élément de H minimisant $v \longmapsto \frac{1}{2}a(v,v) - \ell(v)$.

Application 22. [Problème de Dirichlet faible]
Soit $I =]0,1[$ et $f \in L^2(I)$. Alors

$$\exists! u \in H_0^1(I) \mid \forall v \in H_0^1(I), \int_I u'.v' + \int_I uv = \int_I fv$$

De plus u minimise la fonction $v \longmapsto \frac{1}{2}\int_I |v'|^2 + v^2 - \int_I fv$.

II. B. Espaces L^p

[BP15, Ch9, p157–194] [Bre99, ChIV, p54]

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. Soit $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.
Soit $p \in [1, +\infty]$ et q son exposant conjugué ($\frac{1}{p} + \frac{1}{q} = 1$).

Définition 23. Pour $f : E \longrightarrow \mathbb{K}$ mesurable, on définit $\|f\|_p = \left(\int_E |f|^p \, d\mu\right)^{\frac{1}{p}}$ lorsque p est fini, et $\|f\|_\infty = \inf\{M > 0 \mid |f| \leq M \ \mu\text{-p.p.}\}$. On définit l'espace vectoriel :

$$\mathcal{L}^p(E, \mathcal{A}, \mu) = \{f : E \longrightarrow \mathbb{K} \text{ mesurable} \mid \|f\|_p < +\infty\}$$

Théorème 24. [Inégalité de Hölder]
Pour toutes fonctions $f, g : E \longrightarrow \mathbb{K}$ mesurables, on a $\|fg\|_1 \leq \|f\|_p \|g\|_q$.

Corollaire 25. [Inégalité de Minkowski]
Pour toutes fonctions $f, g : E \longrightarrow \mathbb{K}$ mesurables, on a $\|f+g\|_p \leq \|f\|_p + \|g\|_p$.

Définition 26. On note $L^p(E, \mathcal{A}, \mu) = \mathcal{L}^p(E, \mathcal{A}, \mu)/\{\|.\|_p = 0\}$ (ou $L^p(E)$, ou L^p) l'ensemble des fonctions de $\mathcal{L}^p(E, \mathcal{A}, \mu)$ quotienté par la relation d'égalité μ-p.p.. L'application $\|.\|_p$ passe au quotient et définit ainsi une application sur $L^p(E, \mathcal{A}, \mu)$ notée également $\|.\|_p$.

Corollaire 27. *$\|.\|_p$ est une norme et l'espace $(L^p(E), \|.\|_p)$ est un espace vectoriel normé.*

Application 28. [Inclusion entre espaces L^p]
Soient $p, p' \in [1, +\infty]$.
- lorsque $\mu(E) < +\infty$, on a $p < p' \Longrightarrow L^{p'} \subset L^p$,
- contre-exemple dans le cas où $\mu(E) = +\infty$: $f : x \longrightarrow (1+x)^{-1}\mathbb{1}_{\mathbb{R}_+^*}(x) \in L^2 \setminus L^1$.
- on a $p' < p \Longrightarrow \ell^{p'} \subset \ell^p$.

Proposition 29. [Inégalité de Jensen]
Si μ est une mesure de probabilité, alors pour toute fonction $f \in L^1$ à valeurs réelles et $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction convexe, on a $\varphi(\int_E f d\mu) \leq \int_E \varphi(f) d\mu$.

Application 30. On retrouve que $f(\sum_i \lambda_i x_i) \leq \sum_i \lambda_i f(x_i)$ dans le cas d'une fonction convexe avec des $(\lambda_i)_{1 \leq i \leq n}$ positifs de somme 1.

Application 31. [Théorème de Young]
Soient $p, q, r \in [1, +\infty]$ tels que $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$.
Alors si $f \in L^p$ et $g \in L^q$, on a $f * g \in L^r$ et $\|f * g\|_r \leq \|f\|_p \|g\|_q$.
Lorsque $q = 1$ et $p = r$, c'est une application de l'inégalité de Jensen.
Plus généralement, hormis le cas $p = q = 1$, c'est une application de l'inégalité de Hölder.

II. C. Inégalités en probabilités

[Ouv09, p162/508]

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une variable aléatoire.

Proposition 32. [Inégalité de Jensen]
Si $X \in L^1$ et ϕ est convexe, alors $\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$. De plus, si ϕ est strictement converxe, on a égalité si et seulement si X est constante presque sûrement.

Remarque 33. En particulier on retrouve $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$ et $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$.

Application 34. Soit $X_1, \ldots, X_n$ des v.a. i.i.d. de loi $\mathcal{E}(\theta)$ pour un paramètre $\theta > 0$ inconnu.
Alors $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur sans biais fortement convergeant de $\mathbb{E}[\mathcal{E}(\theta)] = \frac{1}{\theta}$.
Alors $\frac{1}{\overline{X}_n}$ est un estimateur fortement convergeant avec biais de θ.

Application 35. L'inégalité de Jensen reste valable si l'on remplace l'espérance par une espérance conditionnelle.
Ainsi pour φ convexe, si $(X_n)_{n \in \mathbb{N}}$ est une martingale, alors $(\varphi(X_n))_{n \in \mathbb{N}}$ est une sur-martingale.
On a aussi $\|\mathbb{E}[Y \mid \mathcal{B}]\|_p \leq \|Y\|_p$ pour toute v.a. $Y \in L^p$ et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$.

Proposition 36. [Inégalité de Hoeffding]
Soient $(X_i)_{1\leq i\leq n}$ des variables aléatoires indépendantes telles que pour tout i, $X_i \in [a_i, b_i]$ p.s. pour des réels $a_i \leq b_i$. Alors pour tout $t \geq 0$:

$$\mathbb{P}(|\sum_{i=1}^n X_i - \mathbb{E}[X_i]| \geq t) \leq 2\exp(-2t^2/\sum_{i=1}^n (b_i - a_i)^2)$$

Remarque 37. Par un argument de convexité on montre que chaque X_i est $(b_i - a_i)/2$-sous-gaussienne (vérifie $\mathbb{E}[e^{\lambda X_i}] \leq \exp(\lambda^2(b_i - a_i)^2/8)$ pour tout $\lambda \in \mathbb{R}$). L'inégalité de Hoeffding est alors une application d'une inégalité similaire sur les variables sous-gaussiennes.

Application 38. Soient $(X_i)_{1\leq i\leq n}$ des variables aléatoires indépendantes et identiquement distribuéesde Rademacher centrées, c'est-à-dire $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. Alors :

$$\forall t > 0, \mathbb{P}(|\sum_{i=1}^n X_i| \geq t) \leq 2\exp(-\tfrac{t^2}{2n})\mathbb{1}_{t\leq n}$$

Définition 39. [Fonction génératrice, fonction caractéristique]
Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ presque sûrement. On définit g_X la fonction génératrice de X par $g_X(s) = \mathbb{E}[s^X] = \sum_{k\geq 0} \mathbb{P}(X = k)s^k$ (lorsque la série converge).

Proposition 40. *g_X est bien définie, croissante et convexe sur $[0, 1]$. Si de plus $P(X = k) > 0$ pour au moins un $k \geq 2$, alors g_X est strictement convexe sur $[0, 1]$.*

Proposition 41. [Processus de branchement de Galton-Watson]
Soient $(X_i^j)_{i,j\geq 1}$ des v.a. i.i.d. à valeurs dans $\mathbb{N}$. On note μ leur loi, m leur espérance et on définit le processus $(Z_n)_{n\in\mathbb{N}}$ par $Z_0 = 1$ et $Z_{n+1} = \sum_{i=1}^{Z_n} X_i^{n+1}$ pour $n \in \mathbb{N}$. et enfin $\rho = \mathbb{P}(\exists n \geq 0 \mid Z_n = 0)$. Alors si $\mu \neq \delta_1$, on a :
- *si $m \leq 1$, $\rho = 1$ et on a extinction du processus presque surement,*
- *si $m > 1$, $\rho < 1$ et il y a une probabilité strictement positive de survie.*

III. Recherche d'extrema et de points fixes

[Rou99, Ch4, p140] [FGN07, §4.51, p287] [FGN12, §4.51, p39–41] [BMP05, §1.3.2, p19]

Proposition 42. *Soit $f : \mathbb{R}^p \longrightarrow \mathbb{R}$.*
- *Si f est différentiable, f est convexe (resp. strictement convexe) si et seulement si pour tout $x \neq y \in \mathbb{R}^p$, $f(y) - f(x) \geq \langle \nabla f(x) \mid y - x\rangle$ (resp. $f(y) - f(x) > \langle \nabla f(x) \mid y - x\rangle$,*
- *Si f est $\mathcal{C}^2$, f est convexe (resp. strictement convexe) si et seulement si $\nabla^2 f(x)$ est positive (resp. définie positive) pour tout $x \in \mathbb{R}^p$.*

Exemple 43. Soit $A \in \mathcal{S}_n(\mathbb{R})$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$ et $f : x \longmapsto \frac{1}{2}\langle Ax \mid x\rangle - \langle b \mid x\rangle + c$.
Si A est positive (resp. définie positive), f est convexe (resp. strictement convexe). Ainsi par exemple si $(E, \langle . \mid . \rangle)$ est euclidien, l'application $x \longmapsto \|x\|^2$ est strictement convexe.

Soit $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ une fonction convexe.

Proposition 44. *Soit C un convexe de $\mathbb{R}^p$ et $M = \{x \in C \mid f(x) = \inf_C f\}$ l'ensemble des minimums globaux de f sur C.*
On suppose qu'il existe un minimum local $x_\star$ de f sur C. Alors :
- *M est un ensemble convexe et $x_\star \in M$ est un minimum global,*
- *si de plus f est strictement convexe, alors $M = \{x_\star\}$.*

Proposition 45. [Inégalité d'Euler]
Soit C un convexe de $\mathbb{R}^p$. Si f admet un minimum local sur C en $x_\star$ et si f est différentiable en $x_\star$ alors $df(x_\star)(y - x_\star) \geq 0$ pour tout $y \in C$.

Théorème 46. [Méthode de Newton]
Soit $f : [c, d] \longrightarrow \mathbb{R}$ une fonction $\mathcal{C}^2$, où $c < d$, et telle que $f(c) < 0 < f(d)$ et $f' > 0$ sur $[c, d]$.
On considère la suite récurrente définie par $x_0 \in [c, d]$ et $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ pour $n \in \mathbb{N}$.
Alors en notant a l'unique 0 de f, on a :
- *(i) il existe $\alpha > 0$ tel que pour tout $x_0 \in [a - \alpha, a + \alpha]$, $(x_n)_{n\in\mathbb{N}}$ converge vers a de manière quadratique et il existe $C > 0$ telle que $\forall n \in \mathbb{N}, |x_{n+1} - a| \leq C\,|x_n - a|^2$.*
- *(ii) si de plus $f'' > 0$ (f est strictement convexe) sur $[a, d]$, alors pour tout $x_0 \in]a, d]$, $(x_n)_{n\in\mathbb{N}}$ est strictement décroissante et $x_{n+1} - a \sim_{n\to+\infty} \frac{f''(a)}{2f'(a)}(x_n - a)^2$.*

Application 47. Approximation des zéros d'un polynôme.

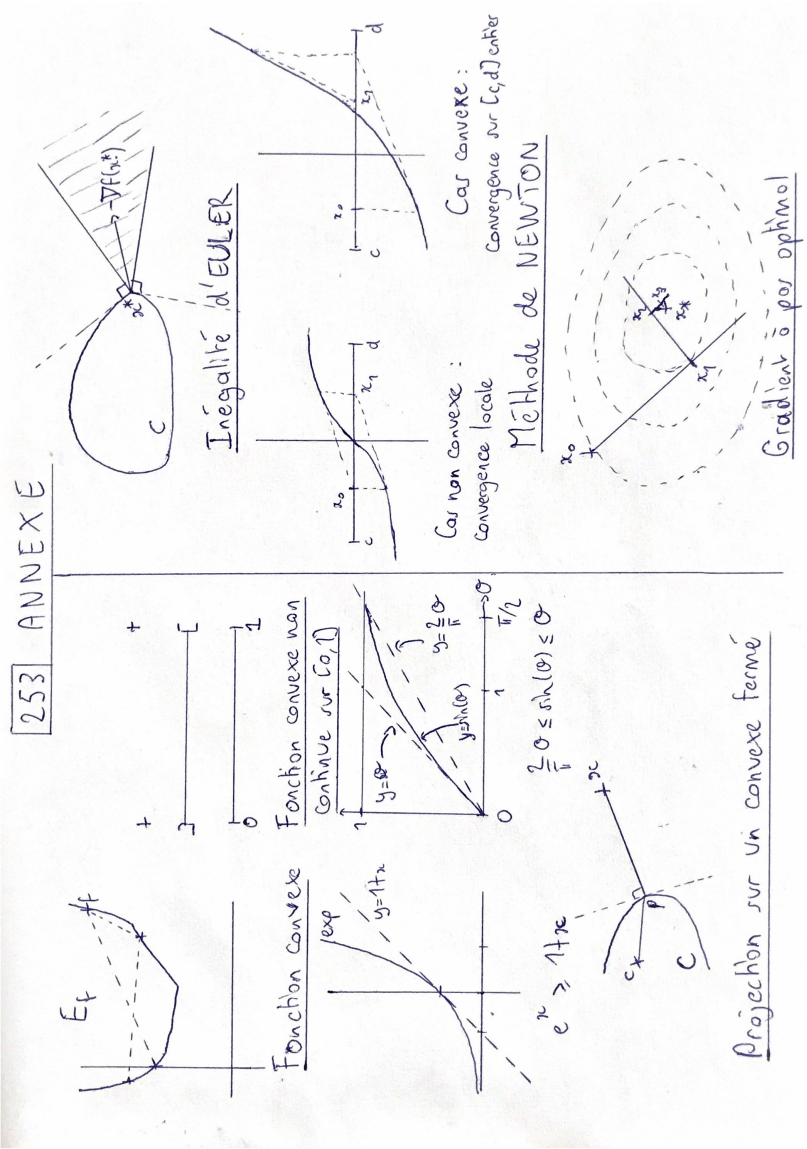
Théorème 48. [Méthode de gradient à pas optimal]
Soit $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ elliptique, c'est-à-dire f est $\mathcal{C}^1$ et telle qu'il existe $\alpha > 0$ satisfaisant

$$\forall x, y \in \mathbb{R}^p, \langle \nabla f(x) - \nabla f(y) \mid x - y\rangle \geq \alpha\,\|x - y\|^2$$

On définit $(x_n)_{n\in\mathbb{N}}$ par $x_0 \in \mathbb{R}^p$ et $x_{n+1} = x_n - \rho_n \nabla f(x_n)$ où $\rho_n = \operatorname{argmin}_{\rho>0} f(x_n - \rho\nabla f(x_n))$. Alors $(x_n)_{n\in\mathbb{N}}$ converge vers l'unique minimum global de f.

Application 49. Soit $A \in \mathcal{S}_n^{++}(\mathbb{R})$, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$. L'application $f : x \longmapsto \frac{1}{2}\langle Ax \mid x\rangle - \langle b \mid x\rangle$ admet un unique minimum $\overline{x}$ caractérisé par $\nabla f(\overline{x}) = A\overline{x} - b = 0$. On converge vers cette solution par l'algorithme de gradient à pas optimal.

ANNEXE



SPEECH

La convexité est une propriété forte sur des ensembles ou des fonctions, qui va se traduire par tout un tas de propriétés, caractérisations, rappelées au début du plan, et qui va offrir un cadre très adéquat pour les problèmes d'optimisation. La définition de la convexité est une propriété de dimension finie (nombre fini de points), donc on va pouvoir avoir des raisonnements géométriques, notamment dans les espaces de HILBERT.
La convexité fournit de nombreuses inégalités pratiques provenant de certaines fonctions, fréquemment utilisés notamment lorsque l'on travaille au voisinage du point d'égalité.
L'inégalité arithmético-géométrique va avoir pour application l'inégalité de YOUNG, qui elle même sera très utile dans la construction des espaces L^p, avec l'inégalité de HÖLDER.
La convexité est un véritable outil dans de nombreux cadres, c'est ce que l'on voit dans la partie 2. Déjà dans les HILBERT, un théorème fondamental est le théorème de projection sur un convexe fermé qui va avoir de nombreuses conséquences : avec notamment le théorème de RIESZ-FRÉCHET puis de LAX-MILGRAM et ses applications aux EDP, mais aussi l'existence de l'espérance conditionnelle en probabilités.
Ensuite la construction des espaces L^p repose sur les inégalités de HÖLDER et de MINKOWSKI, qui vont permettre d'assurer que l'on a une norme sur cet espace, et qui reposent sur l'inégalité de YOUNG vue précédemment. La convexité va aussi servir lorsque l'on s'intéresse aux convolutions de fonctions, par exemple avec le théorème de YOUNG.
Enfin en probabilités, la convexité intervient avec l'inégalité de JENSEN, mais elle permet aussi d'avoir des inégalités de concentration, comme par exemple l'inégalité de HOEFFDING : le fait d'additionner des variables indépendantes va augmenter leur régularité. Enfin la fonction génératrice de variables aléatoires à valeurs dans $\mathbb{N}$ est convexe sur $[0,1]$, et on pourra appliquer cela dans le premier développement sur le processus de branchement de GALTON-WATSON, où l'on fait une distinction de cas dont chacun utilise différentes propriétés de convexité.
Enfin dans une dernière partie, je parle d'optimisation puisque la convexité fournit un cadre intéressant : tout minimum local va être global, et unique si on a stricte convexité. La convexité d'une fonction va permettre à la méthode de NEWTON de passer d'un comportement local à un comportement global. Enfin en deuxième développement on montrera la convergence de l'algorithme de gradient à pas optimal dans le cas d'une fonction elliptique, c'est-à-dire fortement convexe dans un certain sens (convexe + coercive etc).

QUESTIONS

Q Soit $f : U \longrightarrow \mathbb{R}$ une fonction convexe définie sur un convexe U de $\mathbb{R}^n$, supposé borné. On suppose qu'il existe C convexe de U et $\varepsilon > 0$ tel que $\forall x \in C, \overline{B}(x, 2\varepsilon) \subset U$. Montrer que f est lispchitzienne sur C.

R On prend deux points x, y distincts et on regarde sur la droite (xy) les points x_ε et y_ε à distance ε de x et y en dehors de $[x, y]$. Ces points sont dans U par hypothèse, et on peut écrire x comme barycentre de x_ε et y, puis y comme barycentre de x et y_ε. En utilisant alors la convexité de f, on peut majorer $|f(x) - f(y)|$ par $\frac{d}{d+\varepsilon} \max(|f(x) - f(y_\varepsilon)|, |f(y) - f(x_\varepsilon)|)$

où $d = d(x, y)$.
Reste alors à utiliser que f est continue sur U (propriété un peu compliqué qui est admise pour l'exercice), en on obtient que $|f(x) - f(y)| \leq \frac{d}{\varepsilon} 2 \|f\|_{\infty, \overline{C_\varepsilon}}$ où $C_\varepsilon = C + B(0, \varepsilon)$ est un ensemble borné de U, et f est continue sur le compact $\overline{C_\varepsilon}$.

R Soit Z une variable aléatoire réelle et $r \geq s > 0$. Montrer que $\mathbb{E}[|Z|^r]^{1/r} \geq \mathbb{E}[|Z|^s]^{1/s}$.

Q Il s'agit d'appliquer l'inégalité de HÖLDER :

$$\mathbb{E}[|Z|^s] \leq \mathbb{E}[|Z|^r]^{s/r}$$

BIBLIOGRAPHIE

[BMP05] V. BECK, J. MALICK et G. PEYRÉ : *Objectif Agrégation*. H&K, 2ème édition, 2005.

[BP15] M. BRIANE et G. PAGÈS : *Théorie de l'intégration*. Vuibert, 6ème édition, 2015.

[Bre99] H. BREZIS : *Analyse fonctionnelle : théorie et applications*. Dunod, 1999.

[FGN07] S. FRANCINOU, H. GIANELLA et S. NICOLAS : *Oraux X-ENS - Analyse 1*. Cassini, 2007.

[FGN12] S. FRANCINOU, H. GIANELLA et S. NICOLAS : *Oraux X-ENS - Analyse 4*. Cassini, 2012.

[Gou08] X. GOURDON : *Les maths en tête - Analyse*. Ellipses, 2ème édition, 2008.

[Ouv09] J.-Y. OUVRARD : *Probabilités : Tome 2*. Cassini, 3ème édition, 2009.

[Rou99] F. ROUVIÈRE : *Petit guide de calcul différentiel*. Cassini, 1999.