Leçon 219 - Extremums : existence, caractérisation, recherche. Exemples et applications.

Extrait du rapport de jury

Cette leçon offre aux candidates et candidats une multitude d'approches possibles : utilisation de la topologie, du calcul différentiel, de la convexité (fonctions convexes, projection sur un convexe fermé et leurs multiples applications), de l'holomorphie.
Les candidates et candidats peuvent proposer des problèmes d'optimisation sous contraintes, si possible autres que la preuve de l'inégalité arithmético-géométrique. À ce sujet, une bonne compréhension de la méthode des multiplicateurs de Lagrange requiert celle de la notion d'espace tangent, qui en donne une justification beaucoup plus claire que certains raisonnements purement matriciels.
Les algorithmes de recherche d'extremums ont également leur place dans cette leçon (méthode de Newton, du gradient à pas optimal, problème des moindres carrés, etc).
Les candidates et candidats solides peuvent s'intéresser aux diverses versions du principe du maximum (fonctions holomorphes ou harmoniques, équations aux dérivées partielles), au calcul des variations, ou réfléchir à l'unicité de la meilleure approximation dans divers espaces fonctionnels, à commencer par celle des fonctions continues sur un segment par des polynômes de degré au plus égal à d.

Présentation de la leçon

Je vais vous présenter la leçon 219 intitulée : " Extremums : existence, caractérisation, recherche. Exemples et applications.". Le recherche d'extremum est omniprésente tant pour les sciences théoriques que pratiques, car il s'agit de cerner le plus finement possible les limites d'une fonction donnée, que ce soit pour en contrôler les écarts ou pour en optimiser l'utilisation (notamment avec des calculs de distance ou de marge financière par exemple). Ce n'est qu'avec l'avènement du calcul différentiel de Newton que la physique et les mathématiques tirent le plein parti de cette notion. Dès lors, les méthodes n'ont cessé de proliférer mettant au jour des propriétés toujours plus fines des objets et des espaces étudiés, et c'est dans cette variété des méthodes, unifiées autour de la recherche d'extremums que s'insère cette leçon.
Dans une première partie on s'intéresse d'abord aux conditions d'existence et d'unicité. On commence par donner les définitions générales dans une première partie (maximum/-minimum local, extremum local, point critique, etc.) ainsi qu'un premier résultat. Dans une deuxième partie on se place dans le cadre de fonctions continues sur un compact et il est alors possible de montrer que ces fonctions sont bornées et atteignent leurs bornes. On s'intéresse ensuite aux fonctions coercives avec quelques résultats dans le cadre d'espaces vectoriels puis aux fonctions convexes en montrant que les minimums locaux sont en fait globaux ce qui peut déboucher sur des résultats d'optimisation.
Dans une deuxième partie on fait le lien avec le calcul différentiel en regardant comment caractériser un extremum ainsi que l'analyse complexe. On commence par des conditions du premier ordre via le théorème de Rolle et l'égalité des accroissements finis. On s'intéresse ensuite à des conditions du deuxième ordre qui donnent des caractérisations des extrema via l'étude de la hessienne. Dans le cas d'une fonction définie sur $\mathbb{R}^2$, la proposition 23 nous donne ces caractérisations que l'on met en pratique avec un exemple. On continue cette deuxième partie avec de l'optimisation sous contraintes où le résultat majeur est le théorème des extrema liés qui nous donne une foule d'applications possibles et on termine avec une dernière partie sur l'holomorphie en donnant quelques grands résultats d'analyse complexe tels que les théorèmes de Liouville et de d'Alembert et le principe du maximum.
Enfin dans une dernière partie on donne des applications en optimisation numérique en commençant par la méthode de Newton qui permet d'approximer un zéro d'une fonction. On s'intéresse ensuite au problème des moindres carrés qui consiste à trouver une régression linéaire d'un nuage de points (technique très utilisée en statistiques et en physique-chimie). On étudie ensuite la cas des espaces de Hilbert qui sont des cas propices à la géométrie et donc notamment des projections sur des convexes et des sous-espaces vectoriels. On termine enfin cette leçon avec deux derniers points : le premier consacré aux matrices de Gram qui permettent de calculer des distances à des sous-espaces vectoriels et le second dédié à la méthode de descente du gradient où l'on donne d'abord quelques résultats sur les fonctions α-convexes avant d'en venir à la méthode du gradient ainsi que son fonctionnement et sa convergence.
On trouvera en annexe une illustration de minima ainsi que de la méthode du gradient.

Plan général

I - Existence et unicité d'extremum
1 - Premières définitions
2 - Fonctions continues sur un compact et coercivité
3 - Convexité

II - Extrema d'applications différentiables ou holomorphes
1 - Condition du premier ordre et applications
2 - Condition du deuxième ordre
3 - Optimisation sous contraintes
4 - Holomorphie

III - Applications en optimisation numérique
1 - Méthode de Newton
2 - Le problème des moindres carrés
3 - Le cas des espaces de Hilbert
4 - Matrices de Gram
5 - Méthode de descente du gradient

IV - Annexe
1 - Illustration de minima
2 - Illustration de la méthode du gradient

Cours détaillé

Dans toute cette leçon, on considère des fonctions à valeurs réelles.

I Existence et unicité d'extremum

I.1 Premières définitions

Dans toute cette sous-partie, on considère I un intervalle de $\mathbb{R}$ d'intérieur non vide, $a \in I$ et $f : I \longrightarrow \mathbb{R}$.

Définition 1 : Max./min. local [Deschamps (1), p.560] :
On dit que f admet un :
$*$ **maximum local en** a lorsqu'il existe un réel $\eta > 0$ tel que la fonction $f|_{I\cap[a-\eta;a+\eta]}$ admette un maximum en a.
$*$ **minimum local en** a lorsqu'il existe un réel $\eta > 0$ tel que la fonction $f|_{I\cap[a-\eta;a+\eta]}$ admette un minimum en a.

Définition 2 : Extremum local [Deschamps (1), p.560] :
On dit que f admet un **extremum local en** a lorsque f admet un maximum ou minimum local en a.

Remarque 3 : [Deschamps (1), p.560 + 561]
$*$ La fonction f admet un maximum local en a si, et seulement si, la fonction $-f$ admet un minimum local en a.
$*$ un extremum global est un extremum local mais la réciproque est fausse.

Théorème 4 : [Deschamps (1), p.561]
Si a est un point intérieur de I, que f est dérivable en a et que f admet un extremum local en a, alors $f'(a) = 0$.

Remarque 5 : [Deschamps (1), p.561]
Le théorème précédent donne une condition nécessaire d'existence d'un extremum mais pas suffisante. En effet, la fonction $x \longmapsto x^3$ a une dérivée qui s'annule en 0 mais il ne s'agit pas d'un extremum local.

Définition 6 : Point critique [Deschamps (1), p.561] :
On appelle **point critique de** f tout point de I en lequel f est dérivable et de dérivée nulle.

I.2 Fonctions continues sur un compact et coercivité

Théorème 7 : [Théorème des bornes atteintes [Deschamps (2), p.294]
Tout fonction continue définie sur une partie compacte non vide et à valeurs réelles est bornée et atteint ses bornes.

Remarque 8 :
Ce résultat d'optimisation est utilisé par exemple pour démontrer qu'en dimension finie toutes les normes sont équivalentes.

Définition 9 : Fonction coercive [Beck, p.30] :
On dit qu'une fonction f est coercive lorsque $\lim_{\|x\|\to+\infty} f(x) = +\infty$.

Exemple 10 :
Une norme est coercive alors qu'une application linéaire ne l'est pas.

Proposition 11 : [Beck, p.30]
Soient E un espace vectoriel de dimension finie et $f : E \longrightarrow \mathbb{R}$ une fonction continue.
Si C est un fermé de E et si f est coercive, alors f est minorée sur C et atteint son minimum.

Proposition 12 : [Beck, p.30]
Soient $(F, \|\cdot\|)$ un espace vectoriel normé et E un sous-espace vectoriel de F de dimension finie.
Pour tout $x \in F$, il existe $x^* \in E$ tel que $\|x - x^*\| = \inf_{y\in E} \|x - y\|$.

Théorème 13 : [Demailly, p.40]
Soit $f \in \mathcal{C}^0([a, b], \mathbb{R})$.
Pour tout $n \in \mathbb{N}$, il existe un unique polynôme q_n (appelé **polynôme de meilleure approximation uniforme de f à l'ordre n**) tel que $d(f, \mathbb{R}_n[X]) = \|f - q_n\|_\infty$.

I.3 Convexité

Dans toute cette sous-partie, on considère $(E, \|\cdot\|)$ un espace vectoriel de dimension finie, C une partie convexe de E et $f : C \longrightarrow \mathbb{R}$ une application.

Définition 14 : Fonction (strictement) convexe [Deschamps (2), p.158] :
f est dite **(strictement) convexe** sur C lorsque pour tout $(x, y) \in C^2$ et tout $\lambda \in [0; 1]$ on a $f(\lambda x + (1 - \lambda)y) \leq (<) \lambda f(x) + (1 - \lambda)f(y)$.

Théorème 15 : [Rombaldi, p.347]
Si f est convexe sur I et dérivable en $\alpha \in \mathring{I}$ avec $f'(\alpha) = 0$, alors elle admet un minimum global en α.

Théorème 16 : [Rombaldi, p.347]
Si f est convexe sur I et admet un minimum local en $\alpha \in I$, alors ce minimum est global en α.

II Extrema d'applications différentiables ou holomorphes

Dans toute cette sous-partie, on considère $\mathcal{U}$ un ouvert de $\mathbb{R}^n$.

II.1 Condition du premier ordre et applications

Proposition 17 : [Gourdon (1), p.335]
Si f admet un extremum relatif en a et est différentiable en a, alors $df_a = 0$.

Définition 18 : Point critique [Gourdon (1), p.336] :
On dit que a est un **point critique de** f lorsque f est différentiable en a et que $df_a = 0$.

Théorème 19 : Théorème de Rolle [Deschamps (1), p.562] :
Soit $f : [a; b] \longrightarrow \mathbb{R}$ une fonction continue sur $[a; b]$, dérivable sur $]a; b[$ telle que $f(a) = f(b)$.
Il existe $c \in]a; b[$ tel que $f'(c) = 0$.

Théorème 20 : Égalité des accroissements finis [Deschamps (1), p.563] :
Soit $f : [a; b] \longrightarrow \mathbb{R}$ une fonction continue sur $[a; b]$ et dérivable sur $]a; b[$. Il existe $c \in]a; b[$ tel que $f(b) - f(a) = f'(c)(b - a)$.

Corollaire 21 : [Deschamps (1), p.564]
Soient I un intervalle de $\mathbb{R}$ et $f : I \longrightarrow \mathbb{R}$.
Si f est continue sur I et dérivable sur l'intérieur de I, alors f est croissante si, et seulement si, $f'(x) \geq 0$ pour tout x dans l'intérieur de I.

II.2 Condition du deuxième ordre

Théorème 22 : [Gourdon (1), p.336]
Si f est une fonction de classe $\mathcal{C}^2$ et que $df_a = 0$, alors :
$*$ Si f admet un minimum (resp. un maximum) relatif en a, alors $H_f(a)$ est positive (resp. négative).
$*$ Si $H_f(a)$ est définie positive (resp. définie négative), alors f admet un minimum (resp. un maximum) relatif en a.

Proposition 23 : [Gourdon (1), p.337]

Soit x_0 un point critique de f tel que $d^2 f_{x_0} = \begin{pmatrix} r & s \\ s & r \end{pmatrix}$.

On a alors :
$*$ Si $rt - s^2 > 0$ et $r + t > 0$, alors x_0 est un minimum local.
$*$ Si $rt - s^2 > 0$ et $r + t < 0$, alors x_0 est un maximum local.
$*$ Si $rt - s^2 < 0$, alors x_0 n'est pas un extremum.
$*$ Si $rt - s^2 = 0$, alors on ne peut pas conclure.

Exemple 24 : [Gourdon (1), p.337]

La fonction f définie de $\mathbb{R}^2$ dans $\mathbb{R}$ par $f(x,y) = x^4 + y^4 - 2(x-y)^2$ possède 3 points critiques : $(0,0)$, $(\sqrt{2}, -\sqrt{2})$ et $(-\sqrt{2}, \sqrt{2})$.
De plus, le point $(0,0)$ est un point col et les points $(\sqrt{2}, -\sqrt{2})$ et $(-\sqrt{2}, \sqrt{2})$ sont des minimums relatifs (et même absolus).

II.3 Optimisation sous contraintes

Théorème 25 : Théorème des extrema liés [Gourdon (1), p.337] :

Soient $f, g_1, ..., g_p$ des fonctions réelles de classe $\mathcal{C}^1$ définies sur l'ouvert $\mathcal{U}$ de E et $\Gamma = \{x \in U \text{ tq } g_1(x) = ... = g_p(x) = 0\}$.
Si $f|_\Gamma$ admet un extremum relatif en $c \in \Gamma$ et si $dg_{1,c}, ..., dg_{r,c}$ sont linéairement indépendants, alors il existe $\lambda_1, ..., \lambda_r \in \mathbb{R}$ (uniques et appelés **multiplicateurs de Lagrange**) tels que $df_c = \sum_{i=1}^r \lambda_i dg_{i,c}$.

Exemple 26 : [Gourdon (1), p.339]

Il est possible de retrouver l'inégalité arithmético-géométrique à partir du théorème des extrema liés.

Développement 1 : [cf. GOURDON (1)]

Proposition 27 : [Gourdon, p.332]

Le déterminant est de classe $\mathcal{C}^\infty$ sur $\mathcal{M}_n(\mathbb{R})$ et :

$$\forall M \in \mathcal{M}_n(\mathbb{R}), \forall H \in \mathcal{M}_n(\mathbb{R}), \; d(\det)_M(H) = \operatorname{Tr}(\operatorname{Com}(M)^\intercal H)$$

Proposition 28 : [Gourdon (1), p.341]

Si l'on munit $\mathcal{M}_n(\mathbb{R})$ de la norme $\|\cdot\|_2 : M \longrightarrow \left(\sum_{i,j=1}^n m_{i,j}^2\right)^{\frac{1}{2}}$, alors le groupe des matrices orthogonales directes de $\mathrm{SO}_n(\mathbb{R})$ est l'ensemble des éléments de $\mathrm{SL}_n(\mathbb{R})$ de norme minimale.

II.4 Holomorphie

Dans toute cette sous-partie, on considère Ω un ouvert non vide de $\mathbb{C}$.

Théorème 29 : Inégalités de Cauchy [Tauvel, p.84] :

Soient $R > 0$ et $f \in \mathcal{H}(\mathcal{D}(0,R))$.
On a alors pour tout $z \in \mathcal{D}(0,R)$, $f(z) = \sum_{n=0}^{+\infty} a_n z^n$ et pour tout $r \in]0; R[$ et $n \in \mathbb{N}$, on a :

$$|a_n| = \frac{1}{n!}\left|f^{(n)}(a)\right| \leq \frac{\sup_{|z|=r} |f(z)|}{r^n}$$

Théorème 30 : Théorème de Liouville [Tauvel, p.84] :

Toute fonction entière et bornée est constante.

Théorème 31 : Théorème de D'Alembert [Tauvel, p.84] :

Tout polynôme d'une variable à coefficient complexe et non constant a au moins une racine dans $\mathbb{C}$.

Proposition 32 : [Tauvel, p.85]

Soit $f \in \mathcal{H}(\Omega)$.
La fonction f possède la propriété de la moyenne sur Ω

Proposition 33 : [Tauvel, p.86]

Soit f une fonction continue sur Ω.
Si f vérifie la propriété de la moyenne sur Ω, alors f vérifie le principe du maximum sur Ω.

Lemme 34 : Lemme de Schwartz [Tauvel, p.87] :

Soit $f \in \mathcal{H}(\mathcal{D}(0,1))$ telle que $f(0) = 0$ et $|f(z)| < 1$ pour tout $z \in \mathcal{D}(0,1)$.
$*$ Pour tout $z \in \mathcal{D}(0,1)$, $|f(z)| \leq |z|$. $\quad *$ $|f'(0)| \leq 1$.
En outre, si $|f'(0)| = 1$ ou s'il existe $z \in \mathcal{D}(0,1) \backslash \{0\}$ tel que $|f(z)| = |z|$, alors il existe $\lambda \in \mathbb{C}$ tel que $|\lambda| = 1$ et $f(z) = \lambda z$ pour tout $z \in \mathcal{D}(0,1)$.

III Applications en optimisation numérique

III.1 Méthode de Newton

Théorème 35 : Méthode de Newton [Rouvière, p.152] :
Soit $f : [c;d] \longrightarrow \mathbb{R}$ une fonction de classe $\mathcal{C}^2$.
On suppose que $f(c) < 0 < f(d)$ et $f' > 0$ sur $[c;d]$ et on considère la suite $(x_n)_{n\in\mathbb{N}}$ définie par $x_0 \in [c;d]$ et $x_{n+1} = F(x_n)$ où $F(x) = x - \frac{f(x)}{f'(x)}$.
$*$ f possède un unique zéro noté a et pour tout $x \in [c;d]$, il existe $z \in [a;x]$ tel que $F(x) - a = \frac{1}{2}\frac{f''(z)}{f'(x)}(x-a)^2$.
$*$ Il existe $C > 0$ tel que pour tout $x \in [c;d]$, $|F(x) - a| \leq C|x-a|^2$ et il existe $\alpha > 0$ tel que $I = [a-\alpha; a+\alpha]$ soit stable par F et que pour tout $x_0 \in I$, la suite $(x_n)_{n\in\mathbb{N}}$ a une convergence d'ordre 2 vers a dans I.
$*$ Si on suppose de plus que $f'' > 0$ sur $[c;d]$, alors l'intervalle $I = [a;d]$ est stable par F et pour tout $x_0 \in I$, la suite $(x_n)_{n\in\mathbb{N}}$ est strictement décroissante (ou constante) avec :

$$0 \leq x_{n-1} - a \leq C(x_n - a)^2 \text{ et } x_{n+1} - a \underset{n\to+\infty}{\sim} \frac{1}{2}\frac{f''(a)}{f'(a)}(x_n - a)^2$$

Remarque 36 :
Le premier résultat est un résultat local, tandis que dans le second résultat, on suppose f convexe pour avoir la convergence d'ordre 2.

III.2 Le problème des moindres carrés

Étant donnés n points (x_i, y_i) du plan usuel $\mathbb{R}^2$ avec les x_i non tous égaux entre eux, le but de la méthode des moindres carrés est de montrer (et trouver !) des nombres λ et μ uniques qui rendent minimum la somme $\sum_{i=1}^n (\lambda_i x_i + \mu - y_i)^2$.

Si l'on considère $n, p \in \mathbb{N}^*$, $A \in \mathcal{M}_{n,p}(\mathbb{R})$ et $b \in \mathcal{M}_{n,1}(\mathbb{R})$, cela revient à chercher un vecteur $x \in \mathbb{R}^p$ tel que $\|b - Ax\|_2 = \inf_{y\in\mathbb{R}^p} \|b - Ay\|_2$.

Proposition 37 :
$*$ Le problème des moindres carrés admet toujours une solution.
$*$ x est solution du problème des moindres carrés si, et seulement si, $A^\intercal Ax = A^\intercal b$.
$*$ Si $n \geq p$ et $\mathrm{rg}(A) = p$, alors $A^\intercal A$ est inversible et il existe une unique matrice réelle B triangulaire inférieure, à diagonale positive telle que $A^\intercal A = BB^\intercal$ et alors $x = (BB^\intercal)^{-1}A^\intercal b$ est une solution du problème des moindres carrés.

Remarque 38 : [Rouvière, p.386]
On peut étendre la méthode en cherchant un polynôme de degré m plutôt qu'une droite.

III.3 Le cas des espaces de Hilbert

Dans toute cette partie, on considère $(H, <\cdot;\cdot>_H)$ un espace de Hilbert.

Développement 2 : [cf. HASSAN]

Théorème 39 : [Hassan, p.489]
Soit C un convexe fermé non vide de $(H, <\cdot;\cdot>_H)$ un espace hilbertien.
Pour tout $x \in H$, il existe un unique $c \in C$ tel que $d(x, C) = \|x - c\|_H$ avec pour tout $z \in C$, $\mathrm{Re}(<z - c; x - c>_H) \leq 0$.
De plus, en notant P_C la projection sur C, on a P_C 1-lipschitzienne (donc continue).

Théorème 40 : [Hassan, p.490]
Soit F un sous-espace vectoriel fermé de $(H, <\cdot;\cdot>_H)$.
L'application $P_F : H \longrightarrow H$ est linéaire telle que :
$*$ P_F est une projection telle que $P_F \circ P_F = P_F$.
$*$ $\mathrm{Ker}(P_F) = F^\perp$ et $\mathrm{Im}(P_F) = F$.
$*$ Pour $x \in H$, $P_F(x)$ est l'unique élément y de F tel que $x - y \in F^\perp$ (P_F est donc la projection orthogonale sur F).

Corollaire 41 : [Hassan, p.491]
Soit F un sous-espace vectoriel de H.
$*$ On a $\overline{F}^\perp = F^\perp$, $\left(F^\perp\right)^\perp = F$ et $H = \overline{F} \oplus F^\perp$.
$*$ F est dense dans H si, et seulement si, $F^\perp = \{0_H\}$.

Remarque 42 :
L'hypothèse de fermeture dans le corollaire précédent n'est pas superflue. En effet, pour $H = \ell^2(\mathbb{N})$ et F l'ensemble des suites de H nulles à partir d'un certain rang, on a $F^\perp = \{0_H\}$ mais $E \neq F$.

Proposition 43 : [Hassan, p.491]
L'application :

$$J : \left|\begin{array}{ccc} H & \longrightarrow & H^* \\ y & \longmapsto & \varphi_y : \left|\begin{array}{ccc} H & \longrightarrow & \mathbb{K} \\ x & \longmapsto & <x;y>_H \end{array}\right. \end{array}\right.$$

est une isométrie semi-linéaire.

Théorème 44 : Théorème de représentation de Riesz [Hassan, p.492] :
Pour toute forme linéaire continue φ, il existe un unique $y \in H$ tel que $\varphi = \varphi_y$.

III.4 Matrices de Gram

Dans toute cette sous-partie, on considère n un entier naturel non nul.

Définition 45 : Matrice de Gram [Gourdon (2), p.274] :
On considère $x_1, ..., x_n$ n vecteurs de E.
On appelle **matrice de Gram** la matrice $(< x_i; x_j >)^2_{(i,j)\in[\![1;n]\!]}$ et on note $G(x_1, ..., x_n)$ son déterminant.

Théorème 46 : [Gourdon (2), p.275]
Si F est un sous-espace vectoriel de E de dimension n muni d'une base $(e_1, ..., e_n)$, alors pour tout $x \in E$, on a : $d(x, F)^2 = \frac{G(e_1,...,e_n,x)}{G(e_1,...,e_n)}$.

Corollaire 47 : Inégalité d'Hadamard [Gourdon (2), p.273] :
$*$ Si $x_1, ..., x_n \in E$, alors $G(x_1, ..., x_n) \leq \prod_{i=1}^n \|x_i\|^2$.
$*$ Si $v_1, ..., v_n \in \mathbb{C}$, alors $|\det(v_1, ..., v_n)| \leq \prod_{i=1}^n \|v_i\|$.

Remarque 48 :
De plus, on a égalité si, et seulement si, la famille de vecteurs contient un vecteur nul ou est orthogonale. Ainsi, dans le cas d'un espace préhilbertien réel E, le volume d'un parallélotope P de E est égal au produit des longueurs de ses côtés si, et seulement si, P est un parallélotope droit.

III.5 Méthode de descente du gradient

Dans toute cette sous-partie, on munit $\mathbb{R}^n$ du produit scalaire $< \cdot; \cdot >$ usuel et de la norme euclidienne associée et on considère C un ouvert convexe de $\mathbb{R}^n$.

Définition 49 : Application α-convexe [Gourdon (1), p.365] :
On considère $J : C \longrightarrow \mathbb{R}$.
On dit que J est une **application α-convexe** $(\alpha \geq 0)$ lorsque :

$$\forall x, y \in C,\ \forall \delta \in [0; 1],\ J((1-\delta)x + \delta y) \leq (1-\delta)J(x) + \delta J(y) - \frac{\delta(1-\delta)}{2}\alpha \|x - y\|^2$$

Proposition 50 : [Gourdon (1), p.365]
Soit $J : C \longrightarrow \mathbb{R}$.
Si J est α-convexe, alors J est convexe (et en particulier J est continue).

Proposition 51 : [Gourdon (1), p.365]
Soit $J : C \longrightarrow \mathbb{R}$.
Si J est α-convexe et de classe $\mathcal{C}^1$, alors :

$$\forall (x, y) \in C^2,\ < \operatorname{grad}_y(J) - \operatorname{grad}_x(J); y - x > \geq \alpha \|y - x\|^2$$

Proposition 52 : [Gourdon (1), p.366]
Soit $J : \mathbb{R}^n \longrightarrow \mathbb{R}$.
Si J est α-convexe, alors il existe un unique $x_0 \in \mathbb{R}^n$ tel que $J(x_0) = \inf_{x\in\mathbb{R}^n} J(x)$.

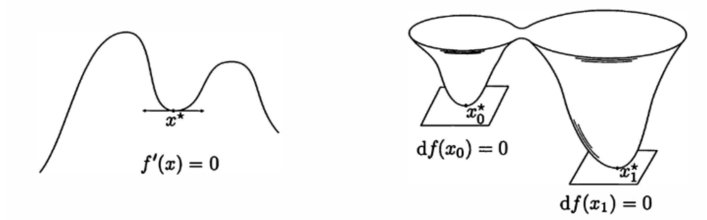
Remarque 53 :
La proposition précédente permet de justifier l'algorithme du gradient qui sert à minimiser $J : \mathbb{R}^n \longrightarrow \mathbb{R}$:
- Initialisation :
On choisir $v_0 \in \mathbb{R}^d$, un paramètre τ (le pas) et $\varepsilon > 0$.

- Boucle en $k \geq 0$:
Calculer $\operatorname{grad}_{v_k}(J)$ et poser $d_k = -\operatorname{grad}_{v_k}(J)$.
Si $d_k = 0$, l'algorithme a terminé.
Sinon, on pose $v_{k+1} = v_k + \tau d_k$ et tant que $|v_{k+1} - v_k| > \varepsilon$, on incrémente k de 1 et on recommence continue la boucle.

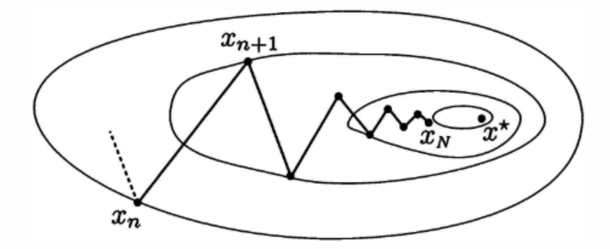
Théorème 54 :
Soit $J : \mathbb{R}^n \longrightarrow \mathbb{R}$ une application α-convexe telle que $\operatorname{grad}_J : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ est ℓ-lipschitzienne $(\ell > \alpha)$.
Si $\tau \in \left]0; \frac{2\alpha}{\ell^2}\right[$, alors la suite $(v_k)_{k\in\mathbb{N}}$ donnée par l'algorithme de la méthode du gradient converge vers l'unique solution du problème $J(x_0) = \inf_{x\in\mathbb{R}^n} J(x)$.

IV Annexe

IV.1 Illustration de minima



IV.2 Illustration de la méthode du gradient



Remarques sur le plan

— Il faut bien être au point sur l'optimisation et les fonctions convexes.

— On peut également parler de norme d'opérateurs ou aller dans d'autres problèmes géométriques.

Liste des développements possibles

— Matrices minimisant la norme sur $\mathrm{SL}_n(\mathbb{R})$.

— Théorème de projection sur un convexe fermé.

Bibliographie

— Claude Deschamps, *Tout-en-un MPSI*.

— Claude Deschamps, *Tout-en-un MP/MP**.

— Vincent Beck, *Objectif agrégation*.

— Jean-Pierre Demailly, *Analyse numérique et équations différentielles*.

— Jean-Étienne Rombaldi, *Mathématiques pour l'agrégation, Analyse et Probabilités*.

— Xavier Gourdon, *Les maths en tête, Analyse*.

— Patrice Tauvel, *Analyse complexe pour la Licence 3*.

— François Rouvière, *Petit guide du calcul différentiel*.

— Nawfal El Hage Hassan, *Topologie générale et espaces normés*.

— Xavier Gourdon, *Les maths en tête, Algèbre et Probabilités*.