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

Refs : [Gou2] Gourdon, Analyse. [All] Allaire, Analyse numérique et optimisation. [HL] Hirsch Lacombe, Éléments d'Analyse fonctionnelle. [Tau] C Tauvel, Analyse complexe. [Ci] Ciarlet, Analyse numérique matricielle et optimisation.

Devt : 11, 12 John Loewner [FGN3] ; 13/14/15 Ref Riesz [HL] ; 32 33 Extrema liés [Aver] ; 40 Gradient optimal [Ci].

Cadre : On se donne $J : E \to \mathbb{R}$ avec E un espace vectoriel normé. On cherche les extrema locaux ou globaux de J, quitte à remplacer J par $-J$, on peut se contenter de regarder les minima.

I. Existence et unicité du minimum.

1) Compacité. Coercivité.

[Gou2] 32 33

Prop 1 : Si K est compact dans E, et J est continue, alors J est bornée sur K et atteint ses bornes.

Appli 2 : Si K, K' sont deux compacts de E, il existe $x, x' \in K \times K'$ tel que $d(x_1, x_2) = d(K, K')$

[All] 290 292

Théo 3 : Si E est de dimension finie, $K \subseteq E$ un fermé non vide $J : K \to \mathbb{R}$ continue et coercive, c'est à dire
$$\lim_{\|x\| \to \infty} J(x) = +\infty.$$
Alors J admet un minimum sur K. De plus, on peut extraire de toute suite minimisante de J sur K une sous-suite convergeant vers le minimum.

Appli 4 : Soit $[a, b]$ un segment réel. Pour $f : [a, b] \to \mathbb{R}$ continue et $n \in \mathbb{N}$, il existe $P \in \mathbb{R}_n[X]$ tel que $\|f - P\|_\infty$ soit minimal.

Ex 5 : Sur l'espace de Hilbert $l^2(\mathbb{N})$, la fonctionnelle $J x = (\|x\|^2 - 1)^2 + \sum \frac{x_i^2}{i}$ n'admet pas de minimum même si J est coercive.

Appl 6 (D'Alembert Gauss) Pour $P \in \mathbb{C}[X]$ non constant, il existe dans $\mathbb{C}$ au moins une racine de P.

2) Convexité.

[All] 294

Pour récupérer des résultats comparables à la dimension finie. On doit rajouter des hypothèses de convexité.

Def 7 : Pour $K \subseteq E$ un convexe non vide, on dit que $J : K \to \mathbb{R}$ est convexe si : $\forall x, y \in K\ \theta \in [0, 1]$, $J(\theta x + (1-\theta) y) \leq \theta J(x) + (1-\theta) J(y)$. On dit que J est strictement convexe si l'inégalité précédente est stricte pour $x \neq y$ et $\theta \notin \{0, 1\}$. Enfin, on dit que J est α-convexe pour $\alpha > 0$ si
$$J(\theta x + (1-\theta) y) \leq \theta J(x) + (1-\theta) J(y) - \frac{\alpha}{2} \theta (1-\theta) \|x - y\|^2.$$
pour les conditions précédentes.

[All]

Prop 8 : Si J est une fonction convexe sur K un ensemble convexe, tout minimum local de J est en fait un minimum global et l'ensemble de points minimums de K forme un ensemble convexe (éventuellement vide). Si de plus J est strictement convexe, il existe au plus un minimum

[Gou2]

Appl 9 : Pour $x_1, \dots, x_n \in \mathbb{R}_+$ on a $\left(\prod_{i=1}^n x_i\right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n x_i$ (inégalité arithmético-géométrique).

Appl 10 (Hölder) Pour $p, q \geq 1$, $\frac{1}{p} + \frac{1}{q} = 1$, $f \in L^p$, $g \in L^q$, alors $fg \in L^1$ et on a $\|fg\|_1 \leq \|f\|_p \|g\|_q$.

[FGN3] 229

Lemme 11 (Convexité logarithmique du déterminant). Pour $A, B \in S_n^{++}(\mathbb{R})$ $\alpha, \beta \in \mathbb{R}_+$, $\alpha + \beta = 1$. Alors $\det(\alpha A + \beta B) \geq \det A^\alpha \det B^\beta$ et l'inégalité est stricte si $0 < \alpha < 1$ et $A \neq B$.

Théo 12 : (John Loewner) Soit $K \neq \emptyset$ un compact de $\mathbb{R}^n$ d'intérieur non vide. Alors il existe un unique ellipsoïde centré en 0 de volume minimal qui contient K. DVP

3) Cas des espaces hilbertiens.

[HL] 91 101

On suppose ici que $E = (E, \langle \cdot, \cdot \rangle)$ est un espace de Hilbert, on va retrouver certains résultats analogues à la dimension finie.

Théo 13 : Soit $\emptyset \neq C \subseteq E$ une partie convexe fermée. Alors
$$\forall x \in E, \exists! y \in C \mid d(x, y) = d(x, C).$$
DVP

On l'appelle projection (orthogonale) de x sur C, il est caractérisé par la propriété $\forall g \in C$, $\mathrm{Re}(y - z, y - x) \leq 0$.

Cor 14 : Si $F \subseteq E$ est un sous-espace vectoriel fermé, alors $E = F \oplus F^\perp$.

Appli 15 (Théorème de Riesz) L'application de E dans E' définie par $y \mapsto \langle \cdot, y \rangle$ est une isométrie surjective.

Théo 16 (Lax Milgram) Si E est un $\mathbb{R}$-espace de Hilbert $a : E \times E \to \mathbb{R}$ une forme bilinéaire continue et coercive : $a(x, x) \geq \alpha \|x\|^2$ pour un $\alpha > 0$. Alors pour tout $v \in E$, il existe $u \in E$ tel que $a(u, v) = L(v)$ où L est une forme linéaire continue.

De plus si a est symétrique, u se caractérise comme minimum de la fonctionnelle $\frac{1}{2} a(x, x) - L(x)$ (fonctionnelle quadratique).

[All]

Appli 17 : On considère l'équation $\{-\Delta u = f \text{ sur } \Omega,\ u=0 \text{ sur } \partial\Omega\}$, la résolution de ce problème dans $H_0^1(\Omega)$ est équivalente à celle de la formulation variationnelle $\int_\Omega \nabla u \nabla v \, d\lambda = \int_\Omega f v \quad \forall v \in H_0^1(\Omega)$.

Théo 18 : Soit $C \subseteq E$, $J : C \to \mathbb{R}$ continue coercive convexe, alors J admet un minimum.

4) Cas des fonctions holomorphes.

[Rou C] 85 86

Propo 19 : Soit $U \subseteq \mathbb{C}$ un ouvert et $f : U \to \mathbb{C}$ une fonction continue, on a équivalence :
(i) Pour tout disque $D(a,r) \subseteq U$, on a $f(a) = \frac{1}{2\pi}\int_0^{2\pi} f(a+re^{it})\,dt$
(ii) Pour tout disque $D(a,r) \subseteq U$, on a $f(a) = \frac{1}{\pi r^2}\iint_{D(a,r)} f(x+iy)\,dx\,dy$.
On dit alors que f respecte la propriété de la moyenne sur U.

Prop 20 : $f : U \to \mathbb{C}$ holomorphe respecte la propriété de la moyenne sur U.

Théo 21 : Soit $U \subseteq \mathbb{C}$ un ouvert et f continue sur U, si f vérifie la propriété de la moyenne sur U, alors f respecte le principe du maximum :
Si $|f|$ admet un extremum relatif en $a \in U$, f est constante au voisinage de a.

Cor 22 : Si $f : U \to \mathbb{C}$ est continue, vérifie la propriété de la moyenne sur U (U ouvert connexe borné de $\mathbb{C}$). Alors $|f|$ atteint son maximum sur ∂U, et est constante si ce maximum est atteint en un point intérieur.

II. Caractérisation des extremums locaux.

1) Différentiabilité et points critiques.

[Gou2] 315 316

Prop 23 : Si $J : U \subseteq \mathbb{R}^n \to \mathbb{R}$ admet un extremum relatif en $a \in U$ et si J est différentiable en a, alors $dJ_a = 0$ (condition que a vérifie l'équation d'Euler). $U \subseteq \mathbb{R}^n$ est un ouvert.

Rq 24 : Dans le cas général, cette condition n'est pas suffisante, elle donne seulement une liste de candidats possibles : $x \mapsto x^3$ admet un point critique en 0 sans s'annuler.

Prop 24 : Si $J : U \to \mathbb{R}$ est de classe C^2 et soit $a \in U$ telque $dJ_a = 0$. Par la formule de Taylor Young, on a au voisinage de a :
$$J(a+h) = J(a) + \frac{Q(h)}{2} + o(\|h\|^2)$$
où $Q = d^2J_a$ est une forme quadratique.
- Si J admet en a un minimum local, alors Q est positive
- Si Q est définie positive, alors J admet un minimum local isolé en a.

Ex 25 : À nouveau ces conditions laissent un cas douteux, où $x \mapsto x^3$ en 0 donne encore un contre exemple.

[Gou2] 317 318

Ex 26 : Si $n=2$. On pose $A = \begin{pmatrix} r & s \\ s & t \end{pmatrix} \in \mathcal{S}_2(\mathbb{R})$ la hessienne de J. On a par le théorème précédent
- $rt - s^2 > 0$ et $r > 0 \Rightarrow J$ admet un minimum relatif en a
- $rt - s^2 < 0 \Rightarrow J$ n'admet pas d'extremum en a
- $rt - s^2 = 0 \Rightarrow$ cas douteux.

Ex 27 : Si $J(x,y) = x^4 + y^4 - 2(x-y)^2$. Alors J a trois points critiques $(0,0)$ $(\sqrt{2}, -\sqrt{2})$ et $(-\sqrt{2}, \sqrt{2})$. Il y a un minimum local en $\pm(\sqrt{2}, -\sqrt{2})$ mais en $(0,0)$, on ne peut pas conclure (en fait il n'y en a pas).

2) Fonctions convexes.

[All] 305 309

Prop 28 : Soit $J : E \to \mathbb{R}$ différentiable. On a équivalence entre
- J est convexe sur E
- $\forall x, y \in E$, $J(y) \geq J(x) + (\nabla J(x), y-x)$
- $(\nabla J(x) - \nabla J(y), x-y) \geq 0 \quad \forall x, y \in E$.

(et les mêmes équivalence avec le cas strictement convexe).

Prop 29 : Avec les notations précédentes, et $\alpha > 0$. Les conditions suivantes sont équivalentes
- J est α-convexe sur E
- $\forall x, y \in E$, $J(y) \geq J(x) + (\nabla J(x), y-x) + \frac{\alpha}{2}\|x-y\|^2$
- $(\nabla J(x) - \nabla J(y), x-y) \geq \alpha \|x-y\|^2 \quad \forall x, y \in E$.

Théo 30 : (Inéquation d'Euler, cas convexe) Soit $u \in K$ convexe. On suppose J différentiable en u. Si u est un minimum local de J sur K, alors
$$(\nabla J(u), x-u) \quad \forall x \in K$$
On a la réciproque si J est convexe (u est alors même un minimum global)

Prop 31 : Si J est convexe sur E, alors $u \in E$ est extremum si et seulement si $\nabla J(u) = 0$.

3) Optimisation sous contrainte.

On peut chercher à minimiser J sur un ensemble $K \subseteq E$, on considère deux cas principaux : $K = \{x \in E \mid f(x) = 0\}$ pour une f (contraintes égalités)
$K = \{x \in E \mid f(x) \leq 0\}$ pour une f (contraintes inégalité).

[Aves]

Théo 32 : (Extrema liés) Soient $f, g_1 \dots g_r : U \to \mathbb{C}$ des fonctions C^1. On pose
- $(g_1 \dots g_r) = g : U \to \mathbb{R}^r$
- $\Gamma = \{x \in U \mid g(x) = 0\}$.

DVP

Si $a \in \Gamma$ est un extremum local de $f|_\Gamma$, alors : $dg_i a$ est une famille libre de $(\mathbb{R}^n)^*$ alors $\exists \lambda_1 \dots \lambda_r \in \mathbb{R}$ telque $df_a = \sum \lambda_i dg_i a$ on appelle les λ_i les multiplicateurs de Lagrange.

Appl 33 : Si $(E, \langle , \rangle)$ est euclidien, tout endomorphisme symétrique est diagonalisable sur une base orthonormée de E.

[All] 316. 317

Pour le cas des contraintes inégalités. On doit rajouter des conditions sur les contraintes.
On pose $f = (f_1 \dots f_m) : E \to \mathbb{R}^m$

Def 34 : Pour $x \in K$, on pose $I(x) = \{ i \in [\![1,m]\!] \mid f_i(x) = 0 \}$ les contraintes actives en x

Def 35 : On dit que les contraintes sont qualifiées en $x \in K$ s'il existe $\bar{w} \in E$ telle que pour tout $i \in I(x)$, on a $(f_i'(x), \bar{w}) < 0$ ou $(f_i'(x), \bar{w}) = 0$ et f_i est affine.

Théo 36 : Si J et les f_i sont différentiables en u et les contraintes sont qualifiées en u. Alors si u est un minimum local de J sur K, il existe $\lambda_1, \dots, \lambda_m \geq 0$, les multiplicateurs de Lagrange, tels que $\nabla J(u) + \sum_{i=1}^{p} \lambda_i f_i'(u) = 0$ $\lambda_i \geq 0$, $\lambda_i = 0$ si $f_i(u) < 0$

III. Optimisation numérique.

1) Méthodes de relaxation et de gradient. (dim $< \infty$)

[Lia] 184 189

On cherche, étant donnée J et u_0 de construire une suite convergeant vers un minimum de J (ça suppose qu'un tel minimum existe).

1ère méthode : Méthode de relaxation. On obtient X^{n+1} à partir de X^n en minimisant successivement J le long des directions canoniques : Si X^n est fixé, on pose pour $i \in [\![1,n]\!]$ X_i^{n+1} tel que $J(X_1^{n+1}, \dots, X_i^{n+1}, X_{i+1}^n, \dots, X_n^n) = \inf_{t \in \mathbb{R}} J(X_1^{n+1}, \dots, X_{i-1}^{n+1}, t, \dots, X_n^n)$

On pose enfin $X^{n+1} = (X_1^{n+1}, \dots, X_n^{n+1}) \in \mathbb{R}^n$.

Théo 37 : Si la fonctionnelle J est elliptique (α-convexe), alors la méthode de relaxation est bien posée et converge vers la solution globale.

Rq 38 : Un exemple fameux de fonctionnelle α-convexe est donné par les fonctionnelles quadratiques : Pour $A \in M_n(\mathbb{R})$, $b \in \mathbb{R}^n$, on pose $J(X) = \frac{1}{2}(AX, X) - (b, X)$. On a alors $\nabla J(X) = AX - b$. Si $A \in S_n^{++}(\mathbb{R})$ alors J est α-convexe et admet comme unique minimum la solution de $AX = b$. La minimisation de J est alors une méthode de résolution d'un système linéaire. Avec cette lecture, la méthode de relaxation correspond à la méthode de Gauss Seidel.

On peut chercher à être plus subtil que les vecteurs de base canonique comme choix de direction et prendre $\nabla J(X^n)$, en effet, pour $x \in E$, on a

$$J(x - \rho \nabla J(x)) \simeq J(x) - \rho \|\nabla J(x)\|^2 \quad \text{pour } \rho \text{ assez petit.}$$

Les méthodes de la forme $X^{n+1} = X^n - \rho^n \nabla J(X^n)$ sont dites méthodes de gradient.
On a essentiellement deux options raisonnables quant au choix de ρ^n

- Prendre ρ constant (ne dépendant pas de n). C'est le gradient à pas fixe
- Prendre ρ^n minimisant la fonction à une variable $\rho \mapsto J(X^n - \rho \nabla J(X^n))$ c'est le gradient à pas optimal.

[All] 339 340

Théo 39 : Si J est une fonctionnelle α-convexe différentiable, et ∇J est lipschitzien sur E. Alors pour $\rho \in]0, 2\alpha / C^2[$, l'algorithme du gradient à pas fixe converge quelque soit le point de départ, vers le minimum de J.

[Lia]

Théo 40 : Si J est α-convexe différentiable et ∇J est lipschitzien sur tout borné de E. Alors l'algorithme de gradient à pas optimal converge. DVP

Rq 41 : La condition du théorème 39 n'est pas optimale, dans le cas d'une fonctionnelle quadratique, on peut aller jusqu'à $\rho = 2/\lambda_n$ où $\lambda_n = \rho(A)$. (Le pas fixe optimal est alors donné par $2/(\lambda_1 + \lambda_n)$).

Rq 42 : A priori, le calcul du pas optimal dans l'algorithme associé peut beaucoup rallonger les calculs. Dans le cas des fonctionnelles quadratiques à nouveau. Ce pas optimal est explicite, et donné par $\frac{-\|\nabla J(X^n)\|^2}{(AX^n, X^n)}$.

2) Méthode de Newton.

[All] 347 348

Si E est de dimension finie. $F : \mathbb{R}^n \to \mathbb{R}^n$ de classe C^2, si u est un zéro de F tel que $dF(u)$ est inversible. La méthode de Newton consiste à résoudre $F(X) = 0$ en posant

$$X^{n+1} = X^n - (dF(X^n))^{-1} F(X^n) \quad \text{pour } n \geq 0.$$

Pour un problème d'optimisation, on va chercher à résoudre $\nabla J(X) = 0$ par la méthode de Newton (si J est assez régulière)

Rq 43 : Notons que cette méthode n'a aucun intérêt pour une fonctionnelle quadratique les itérations consistent déjà à inverser des systèmes linéaires.

Théo 44 : Si $F : \mathbb{R}^n \to \mathbb{R}^n$ est de classe C^2, u un zéro régulier de F. Il existe un $\varepsilon > 0$ tel que si X^0 est assez proche de la solution u (à ε près). La méthode de Newton définie ci-dessus converge vers u, et il existe $C > 0$ telle que

$$\|X^{n+1} - u\| \leq C \|X^n - u\|^2$$

(la convergence est excellente).

Rq 45 : En pratique, on se rapproche de la solution par un algorithme grossier, avant de lancer la méthode de Newton.

Rq 46 : En dimension 1, on retrouve la méthode de Newton "classique".