

Chapitre 2

CONDITIONS NÉCESSAIRES D'OPTIMALITÉ APPROCHÉE

"Good modern science implies good variational problems."

M. S. BERGER (1983)

"Nous devons nous contenter d'améliorer indéfiniment nos approximations." K. POPPER (1984)

Une condition nécessaire d'optimalité standard affirme que si $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ est minimisée en \bar{x} et que f est différentiable en \bar{x} (de différentielle $Df(\bar{x})$), alors $Df(\bar{x}) = 0$. La situation que l'on va examiner dans ce chapitre est celle où il n'y a pas (nécessairement) de minimiseurs de f sur E mais seulement des *minimiseurs approchés*, disons à ε près,

$$f(u) \leq \inf_E f + \varepsilon.$$

Que peut-on dire en de tels u ? Une première tentative – mauvaise – est de penser que $Df(u)$ y est "petit", disons $\|Df(u)\|_* \leq \varepsilon \dots$ Il n'en est rien, mais nous verrons que nous pouvons dire des choses en u , des *conditions nécessaires d'optimalité approchée*.

Points d'appui / Prérequis :

- Bases du calcul différentiel (dans les espaces de Banach)
- Rudiments d'analyse dans les espaces de Banach, de Hilbert.

1 Condition nécessaire d'optimalité approchée ou principe variationnel d'EKELAND

1.1 Le théorème principal : énoncé, illustrations, variantes

Contexte :

$(E, \|\cdot\|)$ est un espace de Banach

$f : E \rightarrow \mathbb{R} \cup \{+\infty\}$, non identiquement égale à $+\infty$, bornée inférieurement sur E

f est semicontinue inférieurement sur E .

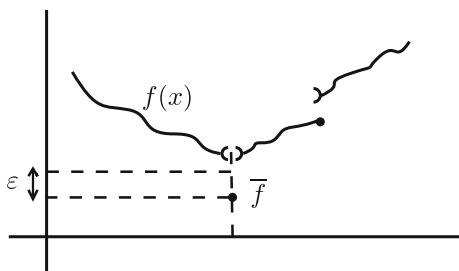
Quelques commentaires sur ces hypothèses :

- On l'aura noté, le contexte est très général... on est loin de l'hypothèse de différentiabilité sur f par exemple.
- On aurait pu prendre (E, d) espace métrique complet (et, de fait, certaines applications de ce qu'on va exposer se font dans un tel contexte), mais on a choisi $(E, \|\cdot\|)$ Banach car cela allège l'écriture et nous replace dans un contexte déjà étudié au Chapitre 1.
- f a été supposée bornée inférieurement, $\bar{f} := \inf_E f > -\infty$, c'est le minimum pour pouvoir parler de u , solution (ou minimiseur de f) à ε près (pour $\varepsilon > 0$) :

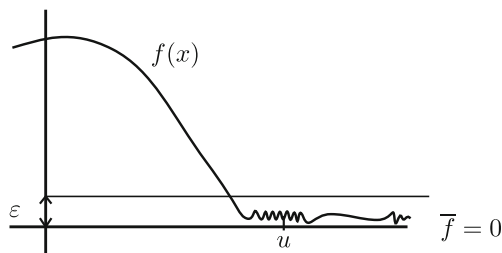
$$\left(\inf_E f \leq \right) f(u) \leq \inf_E f + \varepsilon. \quad (2.1)$$

Notons que, contrairement à la minimisation exacte, l'existence de minimiseurs à ε près (pour $\varepsilon > 0$) ne pose aucun problème : il y a toujours des minimiseurs à ε près ! Cela résulte de la définition même de $\inf A$ lorsque $A \subset \mathbb{R}$. L'unicité des minimiseurs à ε près n'est pas un problème non plus, il y a, généralement, une multitude de minimiseurs à ε près.

Une situation très particulière où ça n'est pas le cas est comme suit :



Un exemple introduction de mise en garde :



Ici f est dérivable sur \mathbb{R} . Même si u est un minimiseur à ε près de f , avec ε très petit, disons $\varepsilon = 10^{-6}$, la dérivée de f en u peut être très grande, disons $|f'(u)| = 10^{12}$!

Théorème 2.1 (I. Ekeland, 1974)

Pour $\varepsilon > 0$ une tolérance donnée, soit u un minimiseur à ε près de f sur E , c'est-à-dire vérifiant $f(u) \leq \bar{f} + \varepsilon$.

Alors, pour tout $\lambda > 0$, il existe $v \in E$ tel que :

- (i) $f(v) \leq f(u)$;
- (ii) $\|v - u\| \leq \lambda$;
- (iii) $\forall x \in E, x \neq v, f(v) < f(x) + \frac{\varepsilon}{\lambda} \|v - x\|$.

Commentaires

- Il s'agit bien d'un théorème d'existence : "il existe v tel que...". Le v exhibé dépend des choix précédents, on aurait pu noter $v_{\varepsilon, u, \lambda}$.
- (i) implique que le v exhibé fait aussi bien que u puisque

$$f(v) \leq f(u) \leq \bar{f} + \varepsilon,$$

v est aussi un minimiseur à ε près de f sur E .

- (ii) exprime que l'on contrôle la distance de v (exhibé) à u (donné au départ), et cette distance, c'est nous qui la contrôlons puisque $\lambda > 0$ est un choix libre de départ !
- Mais il faut compenser quelque part... plus λ est petit, plus grande est la perturbation $x \mapsto \frac{\varepsilon}{\lambda} \|x - v\|$ qui apparaît dans la formulation (iii).
- (iii) exprime un résultat de minimisation (globale). En effet, soit

$$\varphi : E \rightarrow \mathbb{R} \cup \{+\infty\}$$

$$x \mapsto \varphi(x) := f(x) + \frac{\varepsilon}{\lambda} \|v - x\|.$$

(la forme perturbée de f).

Comme $\varphi(v) = f(v)$, ce que dit (iii) n'est ni plus ni moins que

$$\forall x \in E, x \neq v, \varphi(v) < \varphi(x),$$

c'est-à-dire que v est un minimiseur global (strict) de φ sur E .

On notera que le u (de départ) a complètement disparu dans cette formulation...

Un premier raccourci (d'utilisation du théorème) consiste à prendre $\lambda = 1$, de sorte qu'on peut énoncer : Sous les hypothèses du théorème,

$$\begin{aligned} \forall \varepsilon > 0, \exists v_\varepsilon \text{ tel que} \\ f(v_\varepsilon) < f(x) + \varepsilon \|x - v_\varepsilon\| \text{ pour tout } x \in E, x \neq v_\varepsilon. \end{aligned} \quad (2.2)$$

C'est un résultat (raccourci) auquel nous ferons appel de temps en temps.

Une deuxième variante consiste à faire un compromis entre ε et λ : on choisit délibérément $\lambda = \sqrt{\varepsilon}$, ce qui fait que $\frac{\varepsilon}{\lambda} = \sqrt{\varepsilon}$ aussi. Cela donne donc :

Corollaire 2.2

$\varepsilon > 0$ étant donné, soit u un minimiseur à ε près de f sur E . Il existe alors $v_\varepsilon \in E$ tel que :

- (i) $f(v_\varepsilon) \leq f(u)$ (et donc $\leq \overline{f} + \varepsilon$);
- (ii) $\|v_\varepsilon - u\| \leq \sqrt{\varepsilon}$;
- (iii) $\forall x \in E, x \neq v_\varepsilon, f(v_\varepsilon) < f(x) + \sqrt{\varepsilon} \|v - x\|$.

Avant de faire la démonstration (complète) du théorème d'Ekeland, exposons deux illustrations.

1^{ère} illustration : Problème de minimisation avec contraintes.

Considérons

$$(\mathcal{P}) \begin{cases} \text{Minimiser } f(x), \\ x \in S, \end{cases}$$

où $f : E \rightarrow \mathbb{R}$ est continue, S un fermé non vide de E (lequel est toujours un Banach), et f est bornée inférieurement sur S ($\inf_S f > -\infty$).

À $\varepsilon > 0$ fixé, on dit que $u \in S$ est une solution à ε près de (\mathcal{P}) , ou bien est un minimiseur à ε près de f sur S , lorsque $f(u) \leq \inf_S f + \varepsilon$. La condition nécessaire d'optimalité approchée, adaptée au présent contexte, donne ceci :

Soit u un ε minimiseur de f sur S . Alors, pour tout $\lambda > 0$, il existe $v \in S$ tel que :

- (i) $f(v) \leq f(u)$;
- (ii) $\|v - u\| \leq \lambda$;
- (iii) $\forall x \in S, x \neq v, f(v) < f(x) + \frac{\varepsilon}{\lambda} \|x - v\|$.

La démonstration en est simple. Considérons $\tilde{f} : E \rightarrow \mathbb{R} \cup \{+\infty\}$ définie par $\tilde{f} := f + i_S$ (d'où $\tilde{f}(x) = f(x)$ si $x \in S, +\infty$ sinon). Il est clair que minimiser f sur S (exactement ou à ε près) équivaut à minimiser \tilde{f} sur E (exactement ou à ε près), car $\inf_S f = \inf_E \tilde{f}$.

La fonction \tilde{f} , somme de la fonction continue f et de la fonction s.c.i. i_S (n'oublions pas que S a été supposé fermé), est s.c.i. sur E . D'après le théorème principal, il existe $v \in E$ tel que :

- (i) $\tilde{f}(v) \leq \tilde{f}(u) = f(u)$, donc $\tilde{f}(v) < +\infty$, et $v \in S, \tilde{f}(v) = f(v)$;
- (ii) $\|v - u\| \leq \lambda$ (rien ne change ici);
- (iii) $f(v) = \tilde{f}(v) < \tilde{f}(x) + \frac{\varepsilon}{\lambda} \|x - v\|$ pour tout $x \in E, x \neq v$,
soit encore

$$f(v) < f(x) + \frac{\varepsilon}{\lambda} \|x - v\| \text{ pour tout } x \in S, x \neq v.$$

2^{ème} illustration : Quand la différentiabilité entre en jeu.

Commençons par un exercice sous forme de challenge...

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable et bornée inférieurement sur \mathbb{R}^n . Alors, pour tout $\varepsilon > 0$, il existe x_ε tel que $\|\nabla f(x_\varepsilon)\| \leq \varepsilon$.

Comment démontreriez-vous ce résultat ? Pas facile, hein ?

Faisons donc entrer en jeu la différentiabilité dans la condition nécessaire d'optimalité approchée d'Ekeland. Pour un aparté de révision sur les différentes notions de différentiabilité utiles, se reporter à l'Annexe.

Corollaire 2.3 Soit $f : E \rightarrow \mathbb{R}$ continue et Gâteaux-différentiable sur E ; on suppose de plus que f est bornée inférieurement sur E .

Pour un $\varepsilon > 0$ donné, soit u un minimiseur à ε près de f sur E . Alors il existe $v_\varepsilon \in E$ tel que :

- (i) $f(v_\varepsilon) \leq f(u)$;
- (ii) $\|v_\varepsilon - u\| \leq \sqrt{\varepsilon}$;
- (iii) $\|D_G f(v_\varepsilon)\|_* \leq \sqrt{\varepsilon}$.

En raccourci cela donne : $\forall \varepsilon > 0, \exists v_\varepsilon$ tel que $\|D_G f(v_\varepsilon)\|_* \leq \sqrt{\varepsilon}$; ce qui permet de résoudre l'exercice proposé au-dessus.

Démonstration du corollaire :

Seul le point (iii) est à démontrer. Nous savons que

$$\forall x \in E, \varphi(v_\varepsilon) \leq \varphi(x), \quad (2.3)$$

où $\varphi(x) := f(x) + \frac{\varepsilon}{\lambda} \|v_\varepsilon - x\|$. Ce qu'exprime (2.3) est que v_ε est un minimiseur (global, d'ailleurs) de φ sur E . Mais on ne peut affirmer que $D\varphi(v_\varepsilon) = 0$ car φ n'est pas différentiable en v_ε . Rappelons-nous (et revoyons sous forme d'exercice si nécessaire) qu'une norme sur E (quelle qu'elle soit) n'est jamais différentiable en 0. Exploitions néanmoins (2.3) avec divers choix de x . Soit $d \neq 0$ dans E et $\alpha > 0$; avec les choix successifs de $x = v_\varepsilon + \alpha d$ et de $x = v_\varepsilon - \alpha d$, on obtient à partir de (2.3) :

$$\begin{aligned} f(v_\varepsilon + \alpha d) - f(v_\varepsilon) &\geq -\sqrt{\varepsilon} \alpha \|d\|, \\ f(v_\varepsilon - \alpha d) - f(v_\varepsilon) &\geq -\sqrt{\varepsilon} \alpha \|d\|, \end{aligned}$$

soit encore

$$\begin{aligned} \frac{f(v_\varepsilon + \alpha d) - f(v_\varepsilon)}{\alpha} &\geq -\sqrt{\varepsilon} \|d\|, \\ \frac{f(v_\varepsilon - \alpha d) - f(v_\varepsilon)}{(-\alpha)} &\leq \sqrt{\varepsilon} \|d\|. \end{aligned}$$

Comme f est Gâteaux-différentiable en v_ε , un passage à la limite, $\alpha \rightarrow 0$, dans les deux inégalités au-dessus conduit à :

$$\begin{aligned} \langle D_G f(v_\varepsilon), d \rangle &\geq -\sqrt{\varepsilon} \|d\|, \\ \langle D_G f(v_\varepsilon), d \rangle &\leq \sqrt{\varepsilon} \|d\|, \end{aligned}$$

d'où

$$|\langle D_G f(v_\varepsilon), d \rangle| \leq \sqrt{\varepsilon} \|d\|.$$

Par conséquent,

$$\|D_G f(v_\varepsilon)\|_* := \sup_{\substack{d \in E \\ \|d\| \leq 1}} |\langle D_G f(v_\varepsilon), d \rangle| \leq \sqrt{\varepsilon}. \quad \text{CQFD} \quad \square$$

Si on revient à l'exemple de mise en garde du début du paragraphe (cf. page 27) : "en u minimiseur à $\varepsilon = 10^{-6}$ près de f , la dérivée n'est pas petite... mais il y a un v pas trop loin de u , $|v - u| \leq 10^{-3}$, lui-même minimiseur à 10^{-6} près de f , en lequel la dérivée est petite, $|f'(v)| \leq 10^{-3}$ précisément...". Avouez que ça ne se devine pas !

1.2 La démonstration du théorème principal

Le résultat central qui va servir est le suivant ; on a tous fait cet exercice quand on était petit...

Lemme 2.4 Soit (S_k) une suite *décroissante* (au sens de l'inclusion) de *fermés* de E (espace de Banach, donc *complet*). On suppose que

$$\text{diam}(S_k) := \sup_{x, y \in S_k} \|x - y\| \rightarrow 0 \text{ quand } k \rightarrow +\infty.$$

Alors, $\bigcap_{k=0}^{+\infty} S_k$ n'est pas vide et est réduit à un seul point (c'est ce qu'on appelle un singleton).

On va construire de manière récursive une suite de points x_k de E et une suite de fermés (non vides) S_k de E :

$$x_0 \searrow S_0 \nearrow x_1 \searrow S_1 \nearrow \dots x_k \searrow S_k \nearrow x_{k+1} \searrow \dots (x_k) \searrow \dots (S_k)$$

Initialisation du processus :

$x_0 := u$, le minimiseur à ε près de f sur E figurant comme donnée première du théorème.

$$S_0 := \left\{ x \in E \mid f(x) + \frac{\varepsilon}{\lambda} \|x - x_0\| \leq f(u) \right\}.$$

S_0 est un ensemble de sous-niveau de la fonction $x \mapsto f(x) + \frac{\varepsilon}{\lambda} \|x - x_0\|$, laquelle est s.c.i. (somme d'une fonction s.c.i. et d'une fonction continue), donc S_0 est fermé. De plus, S_0 n'est pas vide puisque $x_0 \in S_0$.

Ayant x_k , comment on définit S_k

Ayant x_k , on définit S_k comme suit :

$$S_k := \left\{ x \in E \mid f(x) + \frac{\varepsilon}{\lambda} \|x - x_k\| \leq f(x_k) \right\}.$$

Pour les mêmes raisons que celles évoquées plus haut, pour $k = 0$, S_k est un fermé de E et il contient x_k .

Ayant S_k , comment on définit x_{k+1}

Soit $m_k := \inf_{S_k} f$. Comme

$$-\infty < \inf_E f \leq m_k \leq f(x_k) (< +\infty),$$

il est loisible de choisir $x_{k+1} \in S_k$ tel que

$$f(x_{k+1}) \leq \frac{1}{2} [f(x_k) + m_k].$$

(il n'est pas exclu que x_{k+1} puisse être pris égal à x_k si $f(x_k) = m_k$).

Puis on définit S_{k+1} comme plus haut, et ainsi de suite.

Analysons les propriétés des suites (de points) (x_k) et de fermés (S_k) que l'on vient de définir. Les choses ne sont pas difficiles, mais il faut y aller progressivement.

(\mathcal{P}_1) (S_k) est décroissante : $\forall k, S_{k+1} \subset S_k$.

Soit en effet $x \in S_{k+1}$. Cela signifie, par définition même de S_{k+1} ,

$$f(x) + \frac{\varepsilon}{\lambda} \|x - x_{k+1}\| \leq f(x_{k+1}). \quad (2.4)$$

De par l'inégalité triangulaire, on en déduit :

$$\begin{aligned} f(x) + \frac{\varepsilon}{\lambda} \|x - x_k\| &\leq f(x) + \frac{\varepsilon}{\lambda} \|x - x_{k+1}\| + \frac{\varepsilon}{\lambda} \|x_{k+1} - x_k\| \\ &\leq f(x_{k+1}) + \frac{\varepsilon}{\lambda} \|x_{k+1} - x_k\| \text{ (grâce à (2.4))}, \\ &\leq f(x_k) \text{ (puisque } x_{k+1} \in S_k \text{ par construction)}. \end{aligned}$$

D'où, finalement,

$$f(x) + \frac{\varepsilon}{\lambda} \|x - x_k\| \leq f(x_k),$$

qui traduit bien le fait que $x \in S_k$.

(\mathcal{P}_2) (m_k) est croissante.

Comme $S_{k+1} \subset S_k$,

$$m_{k+1} := \inf_{S_{k+1}} f \geq \inf_{S_k} f =: m_k.$$

(\mathcal{P}_3) Décroissance géométrique de $(f(x_k) - m_k)_k$:

$$f(x_{k+1}) - m_{k+1} \leq \frac{1}{2} [f(x_k) - m_k]. \quad (2.5)$$

En effet,

$$f(x_{k+1}) \leq \frac{1}{2} [f(x_k) + m_k] \text{ (par construction de } x_{k+1}),$$

$$m_k \leq m_{k+1} \text{ (démontré au point } (\mathcal{P}_2));$$

cela implique

$$f(x_{k+1}) - m_{k+1} \leq \frac{1}{2} [f(x_k) + m_k] - m_k = \frac{1}{2} [f(x_k) - m_k].$$

(\mathcal{P}_4) Le diamètre de S_k , $\delta_k := \text{diam}(S_k)$, tend vers 0 quand $k \rightarrow +\infty$.

Par définition, $\delta_k = \sup_{a, b \in S_k} \|a - b\|$.

Soit $a \in S_k$. Par définition même de S_k ,

$$f(a) + \frac{\varepsilon}{\lambda} \|a - x_k\| \leq f(x_k).$$

En conséquence,

$$\begin{aligned} m_k + \frac{\varepsilon}{\lambda} \|a - x_k\| &\leq f(x_k), \\ \|a - x_k\| &\leq \frac{\lambda}{\varepsilon} [f(x_k) - m_k]. \end{aligned}$$

En réitérant l'inégalité (2.5), il s'ensuit :

$$\|a - x_k\| \leq \frac{\lambda}{\varepsilon} \frac{1}{2^k} [f(x_0) - m_0].$$

Si b est un autre élément (quelconque) de S_k ,

$$\|a - b\| \leq \|a - x_k\| + \|x_k - b\| \leq \frac{\lambda}{\varepsilon} \frac{1}{2^{k-1}} [f(x_0) - m_0].$$

In fine,

$$\delta_k \leq \frac{\lambda}{\varepsilon} \frac{1}{2^{k-1}} [f(x_0) - m_0],$$

et $\delta_k \rightarrow 0$ quand $k \rightarrow +\infty$.

Avec toutes ces propriétés énoncées de (S_k) , on fait appel au lemme ^{+∞} rappelé en début de démonstration : $\bigcap_{k=0} S_k = \{v\}$. Montrons que ce v fait notre affaire, c'est-à-dire que les propriétés (i), (ii) et (iii) annoncées du théorème sont bel et bien vérifiées.

Propriété (i). Puisque $v \in S_0$ (forcément...),

$$f(v) + \frac{\varepsilon}{\lambda} \|v - u\| \leq f(u) \quad (\text{de par la définition même de } S_0), \quad (2.6)$$

d'où $f(v) \leq f(u)$.

Propriété (ii). De (2.6) il vient :

$$\bar{f} + \frac{\varepsilon}{\lambda} \|v - u\| \leq f(u) \leq \bar{f} + \varepsilon \left(\text{rappelons que } \bar{f} = \inf_E f \right);$$

d'où $\|v - u\| \leq \lambda$.

Propriété (iii). C'est le point le plus délicat... On va démontrer (iii) sous la forme contraposée suivante

$$\left(x \in E, f(x) + \frac{\varepsilon}{\lambda} \|x - v\| \leq f(v) \right) \Rightarrow (x = v). \quad (2.7)$$

On est d'accord que cela revient au même ?

Partons donc de $x \in E$ vérifiant $f(x) + \frac{\varepsilon}{\lambda} \|x - v\| \leq f(v)$.

Pour tout k , $\|x - v\| \geq \|x - x_k\| - \|x_k - v\|$ (toujours cette fichue inégalité triangulaire); donc

$$f(x) + \frac{\varepsilon}{\lambda} \|x - x_k\| - \frac{\varepsilon}{\lambda} \|x_k - v\| \leq f(v),$$

soit encore

$$\begin{aligned} f(x) + \frac{\varepsilon}{\lambda} \|x - x_k\| &\leq f(v) + \frac{\varepsilon}{\lambda} \|x_k - v\| \\ &\leq f(x_k) \text{ puisque } v \in S_k. \end{aligned}$$

En somme :

$$f(x) + \frac{\varepsilon}{\lambda} \|x - x_k\| \leq f(x_k) \text{ pour tout } k,$$

ce qui revient à dire :

$$x \in S_k \text{ pour tout } k,$$

soit $x \in \bigcap_{k=0}^{+\infty} S_k = \{v\}$, donc $x = v$.

On a donc démontré (2.7), c'est-à-dire que mis à part $x = v$,

$$f(x) + \frac{\varepsilon}{\lambda} \|x - v\| > f(v).$$

□

1.3 Compléments

- Le théorème d'Ekeland est un outil d'Analyse appliquée très puissant, aussi puissant sans doute que "la technologie des approximations successives pour

les points fixes d'applications contractantes" (voir plus loin pour un lien entre les deux). Deux points que nous soulignons toutefois :

- L'importance du caractère *complet* de E ... Il a même été démontré que, peu ou prou, le théorème d'Ekeland s'applique si et seulement si E est complet.
- Avec $\varepsilon = \frac{1}{k}$, on exhibe v_k tel que $f(v_k) \leq \bar{f} + \frac{1}{k}$. On est donc tenté – j'ai vu ça plusieurs fois chez les étudiants – de passer à la limite sur k , en extrayant une sous-suite convergente (v_{k_n}) de (v_k) ... sauf que (v_k) n'a pas forcément de sous-suite convergente. Si tel était le cas, si $v_{k_n} \rightarrow v$ quand $n \rightarrow +\infty$,

$$\begin{aligned} \liminf_{n \rightarrow +\infty} f(v_{k_n}) &\geq f(v) \text{ car } f \text{ est s.c.i.,} \\ \limsup_{n \rightarrow +\infty} f(v_{k_n}) &\leq \bar{f}, \end{aligned}$$

soit $f(v) = \bar{f}$... On est loin de telles situations, c'est plus volontiers que " v_k s'échappe à l'infini" (for whatever that means...).

- Le contexte classique de la méthode des approximations successives pour les points fixes des applications contractantes est le suivant :

(E, d) est un espace métrique complet; φ est une contraction sur E , c'est-à-dire il existe $0 < k < 1$ tel que :

$$\forall x, y \in E, d[\varphi(x), \varphi(y)] \leq k d(x, y). \quad (2.8)$$

Alors φ a un point fixe et un seul (un seul point $\bar{x} \in E$ pour lequel $\varphi(\bar{x}) = \bar{x}$). L'unicité du point fixe ne pose pas problème, c'est son *existence* qui en pose. Voyons comment le théorème d'Ekeland permet d'y accéder facilement.

Définissons $f : E \rightarrow \mathbb{R}$ par $f(x) := d[x, \varphi(x)]$. Bien sûr, f est continue et bornée inférieurement sur E ($\inf_E f \geq 0$ puisque $f \geq 0$). Choisissons $\varepsilon > 0$ de telle sorte que $\varepsilon < 1 - k$ (possible puisque $1 - k > 0$). Grâce au raccourci énoncé en page 28 (cf. Corollaire 2.2), il existe $v \in E$ tel que

$$f(v) \leq f(x) + \varepsilon d(x, v) \text{ pour tout } x \in E. \quad (2.9)$$

Nous proposons $\bar{x} := \varphi(v)$; démontrons que cet \bar{x} fait notre affaire, c'est-à-dire que $\varphi(\bar{x}) = \bar{x}$.

Premier point : exploitation de la propriété de contraction (2.8) avec \bar{x} et v , soit

$$d[\bar{x}, \varphi(\bar{x})] = d[\varphi(v), \varphi(\bar{x})] \leq k d(\bar{x}, v). \quad (2.10)$$

Deuxième point : exploitation de l'inégalité (2.9) avec \bar{x} et v , soit

$$f(v) := d[v, \underbrace{\varphi(v)}_{=\bar{x}}] \leq \underbrace{f(\bar{x})}_{=d[\bar{x}, \varphi(\bar{x})]} + \varepsilon d(\bar{x}, v). \quad (2.11)$$

En combinant (2.10) et (2.11), cela donne :

$$d(v, \bar{x}) \leq (k + \varepsilon) d(\bar{x}, v),$$

ce qui est impossible à tenir avec $d(v, \bar{x}) > 0$ puisque $k + \varepsilon < 1$.

Donc $d(v, \bar{x}) = 0$, c'est-à-dire $(\varphi(\bar{x}) =) v = \bar{x}$.

Dans cette manière de faire – élégante au demeurant – on a perdu une chose : la méthode ou technique des approximations successives, celle qui faisait qu'on approchait le point fixe \bar{x} de φ par la suite définie par : $x_{k+1} := \varphi(x_k)$.

- Lorsque E est de dimension finie, ce qui, reconnaissons-le, n'est pas le contexte habituel des problèmes variationnels, il est possible de démontrer des variantes du théorème d'Ekeland avec des perturbations modelées sur $\|\cdot\|^p$, $p \geq 1$, et donc éventuellement différentiables (comme c'est le cas pour la norme euclidienne $\|\cdot\|$ et $p = 2$).

Ceci nous rapproche de ce qui va être démontré au § 2.

Théorème 2.5 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ semicontinue inférieurement et bornée inférieurement sur \mathbb{R}^n . Soit $\lambda > 0$ et $p \geq 1$.

La tolérance $\varepsilon > 0$ étant donnée, soit u un minimiseur à ε près de f sur \mathbb{R}^n , i.e. vérifiant $f(u) \leq \bar{f} + \varepsilon$.

Alors il existe $v \in \mathbb{R}^n$ tel que :

- (i) $f(v) \leq f(u)$ [et même $f(v) + \frac{\varepsilon}{\lambda} \|v - u\|^p \leq f(u)$];
- (ii) $\|v - u\| \leq \lambda$;
- (iii) $\forall x \in \mathbb{R}^n$, $f(v) + \frac{\varepsilon}{\lambda^p} \|v - u\|^p \leq f(x) + \frac{\varepsilon}{\lambda^p} \|x - u\|^p$.

Démonstration : Considérons la fonction $\theta := \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ définie par

$$\theta(x) := f(x) + \frac{\varepsilon}{\lambda^p} \|x - u\|^p.$$

f est s.c.i. et bornée inférieurement sur \mathbb{R}^n ; $\|x - u\|^p \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$. Ces deux raisons font que f est s.c.i. et 0-coercive sur \mathbb{R}^n ($f(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$).

Par conséquent – et c'est là que la dimension finie de $E = \mathbb{R}^n$ joue un rôle – il existe $v \in \mathbb{R}^n$ minimisant θ sur \mathbb{R}^n . Vérifions que ce v fait notre affaire.

Point (i). $\theta(v) \leq \theta(u)$, soit $f(v) + \frac{\varepsilon}{\lambda^p} \|v - u\|^p \leq f(u)$.

Point (ii). On a

$$\bar{f} + \frac{\varepsilon}{\lambda^p} \|v - u\|^p \leq f(v) + \frac{\varepsilon}{\lambda^p} \|v - u\|^p \leq f(u) \leq \bar{f} + \varepsilon,$$

d'où

$$\frac{\varepsilon}{\lambda^p} \|v - u\|^p \leq \varepsilon,$$

et donc $\|v - u\| \leq \lambda$.

Point (iii). $\theta(v) \leq \theta(x)$ pour tout $x \in \mathbb{R}^n$ se traduit par :

$$f(v) + \frac{\varepsilon}{\lambda^p} \|v - u\|^p \leq f(x) + \frac{\varepsilon}{\lambda^p} \|x - u\|^p \text{ pour tout } x \in \mathbb{R}^n,$$

c'est-à-dire l'inégalité de (iii) annoncée. □

Remarque 2.6 Dans le cas particulier où $p = 1$, l'inégalité (iii) du théorème ci-dessus dit :

$$\forall x \in \mathbb{R}^n, f(v) + \frac{\varepsilon}{\lambda} \|v - u\| \leq f(x) + \frac{\varepsilon}{\lambda} \|x - u\|.$$

Il s'ensuit :

$$\forall x \in \mathbb{R}^n, f(v) \leq f(x) + \frac{\varepsilon}{\lambda} \left[\|x - u\| - \|v - u\| \right] \leq f(x) + \frac{\varepsilon}{\lambda} \|x - v\|,$$

ce qui est (l'essentiel de) l'inégalité (iii) du théorème d'Ekeland.

2 Condition nécessaire d'optimalité approchée ou principe variationnel de BORWEIN-PREISS

2.1 Le théorème principal : énoncé, quelques illustrations

Dans ce paragraphe, l'idée est de présenter une condition nécessaire d'optimalité approchée ou principe variationnel avec des perturbations "lisses" de f , de la forme $\|\cdot\|^p$ par exemple. Le résultat ne sera pas décliné dans toute sa généralité, mais dans un contexte simplifié : l'espace sous-jacent sera un Hilbert et la perturbation de type $\|\cdot\|^2$.

Contexte :

$(H, \langle \cdot, \cdot \rangle)$ est un espace de *Hilbert* ($\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$) est la norme associée à $\langle \cdot, \cdot \rangle$.

$f : H \rightarrow \mathbb{R} \cup \{+\infty\}$, non identiquement égale à $+\infty$, *bornée inférieurement* sur H .

f est *semicontinue inférieurement* sur H .

Un des avantages de la norme hilbertienne $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ est qu'elle est très manipulable pour les calculs (rappelons que $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle$) et que la fonction $x \mapsto \|x\|^2$ est \mathcal{C}^∞ sur H .

Théorème 2.7 (J. Borwein et D. Preiss, 1987)

La tolérance $\varepsilon > 0$ étant donnée, soit u tel que $f(u) < \bar{f} + \varepsilon$. Alors, pour tout $\lambda > 0$, *il existe* v et w dans H tels que :

- (i) $f(v) < \bar{f} + \varepsilon$;
- (ii) $\|v - u\| < \lambda$ et $\|w - v\| < \lambda$;
- (iii) v minimise la fonction $x \mapsto g(x) := f(x) + \frac{\varepsilon}{\lambda^2} \|x - w\|^2$ sur H .

Commentaires

- C'est encore un théorème *d'existence* : "il existe v et w ...", mais cette fois-ci ce sont *deux* points qui sont exhibés.
- (i) indique que le v exhibé fait aussi bien que u .
- (ii) contrôle les distances de v et w par rapport au u de départ : $\|v - u\| < \lambda$ mais aussi $\|w - u\| < 2\lambda$.
- La fonction perturbant f dans (iii) est \mathcal{C}^∞ cette fois. Voyons ce que signifie (iii) géométriquement. Introduisons pour cela $p(x) = -\frac{\varepsilon}{\lambda^2} \|x - w\|^2$; le graphe de p est parabolique, tourné vers le bas (car p est quadratique concave), son sommet est atteint en $x = w$.

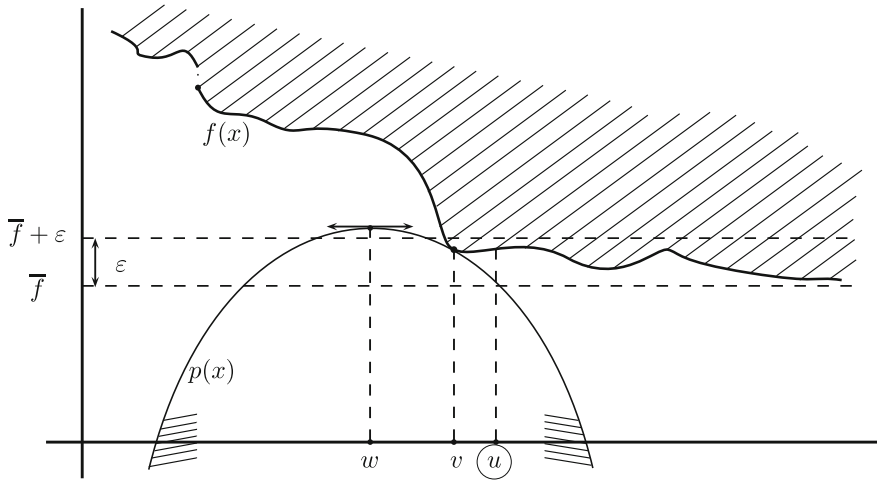
Réécrivons (iii) de manière différente mais équivalente :

$$\begin{aligned} & \forall x \in H, g(x) \geq g(v) \\ \Leftrightarrow f(x) + \frac{\varepsilon}{\lambda^2} \|x - w\|^2 & \geq f(v) + \frac{\varepsilon}{\lambda^2} \|v - w\|^2, \end{aligned}$$

soit encore :

$$\forall x \in H, f(x) - f(v) \geq p(x) - p(v). \quad (2.12)$$

Ainsi, le graphe de f est au-dessus du graphe parabolique de p , et les deux se touchent au point $(v, f(v))$.



(iii) peut d'ailleurs être raffiné en précisant que v est point de minimisation unique de la fonction perturbée g sur H , bref le point de contact $(v, f(v)) = (v, p(v))$ entre les deux graphes est le seul.

- En général, $v \neq w$, il n'y a aucune raison pour qu'ils coïncident.

La pente de p au point v (de contact) est $\nabla p(v) = -\frac{2\varepsilon}{\lambda^2}(v - w)$. Avec les estimations données en (ii), $\|\nabla p(v)\| < \frac{2\varepsilon}{\lambda}$.

Le vecteur $\nabla p(v)$ jouerait un rôle de "sous-gradient" ou de "gradient par dessous" de f en x ... Évidemment, si f se trouvait être différentiable en v (in whatever sense), $\nabla f(v) = \nabla p(v)$.

Précisons le rôle du point v par rapport à f , avec des substituts de conditions nécessaires d'optimalité, du 1^{er} comme du 2nd ordre.

Corollaire 2.8 Le vecteur $s := \nabla p(v)$ vérifie :

(C1) [sorte de condition de minimalité du 1^{er} ordre]

$$\liminf_{x \rightarrow v} \frac{f(x) - f(v) - \langle s, x - v \rangle}{\|x - v\|} \geq 0 ;$$

(C2) [sorte de condition de minimalité du 2nd ordre]

$$\liminf_{x \rightarrow v} \frac{f(x) - f(v) - \langle s, x - v \rangle}{\|x - v\|^2} \geq -\frac{\varepsilon}{\lambda^2} .$$

Démonstration : Comme

$$\frac{f(x) - f(v) - \langle s, x - v \rangle}{\|x - v\|} = \|x - v\| \frac{f(x) - f(v) - \langle s, x - v \rangle}{\|x - v\|^2};$$

il est facile de voir que (C1) est une conséquence de (C2).

(C2) est une condition de "minoration de courbure" de f en v par $-\frac{\varepsilon}{\lambda^2}$, laquelle est la courbure en tout point de la fonction quadratique p .

Soit $\theta(x) := f(x) - p(x)$, θ mesure l'écart entre les deux fonctions f et p . On a déjà observé (cf. (2.12)) que $\theta(x) \geq \theta(v)$ pour tout $x \in H$. En conséquence,

$$\liminf_{x \rightarrow v} \frac{\theta(x) - \theta(v)}{\|x - v\|^2} \geq 0. \quad (2.13)$$

Sachant que $p(x) = p(v) + \langle s, x - v \rangle - \frac{\varepsilon}{\lambda^2} \|x - v\|^2$ (c'est le développement de Taylor à l'ordre 2 de p en x , exact puisque p est quadratique), on a :

$$\begin{aligned} \theta(x) &= f(x) - p(x) = f(x) - p(v) - \langle s, x - v \rangle + \frac{\varepsilon}{\lambda^2} \|x - v\|^2, \\ -\theta(v) &= p(v) - f(v), \end{aligned}$$

d'où

$$\theta(x) - \theta(v) = f(x) - f(v) - \langle s, x - v \rangle + \frac{\varepsilon}{\lambda^2} \|x - v\|^2.$$

Le résultat (C2) annoncé résulte alors de (2.13). \square

On se souvient des conditions nécessaires d'optimalité suivantes :

Si $f : H \rightarrow \mathbb{R}$ est minimisée (même localement) en \bar{x} et que f est deux fois différentiable en \bar{x} , alors $\nabla f(\bar{x}) = 0$ et $D^2 f(\bar{x})$ est "positive", i.e. pour tout $d \in H$, $D^2 f(\bar{x})(d, d) \geq 0$.

En particulier,

$$\liminf_{x \rightarrow \bar{x}} \frac{f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle}{\|x - \bar{x}\|^2} \geq 0. \quad (2.14)$$

Mais que se passe-t-il quand il n'y a pas de minimiseur exact comme \bar{x} ?

On a alors une sorte de conditions d'optimalité du 1^{er} et 2nd ordre asymptotiques, avec des points qui "s'échappent à l'infini" ; elles sont bien sûr obtenues à partir de principes variationnels concernant des minimiseurs approchés de f .

Proposition 2.9 Outre les hypothèses sur f au début du paragraphe (p. 38), supposons que f soit Gâteaux-différentiable sur H . Soit (x_k) une suite minimisante pour f , c'est-à-dire telle $f(x_k) \rightarrow \bar{f}$ quand $k \rightarrow +\infty$.

Il existe alors une suite (v_k) de points de H vérifiant les trois propriétés suivantes :

- (i) $f(v_k) \rightarrow \bar{f}$ quand $k \rightarrow +\infty$ [(v_k) est aussi une suite minimisante pour f];
- (ii) $\|v_k - x_k\| \rightarrow 0$ quand $k \rightarrow +\infty$ [l'écart entre v_k et x_k se resserre au fur et à mesure que k augmente].
- (iii) $\|\nabla_G f(v_k)\| \rightarrow 0$ quand $k \rightarrow +\infty$ [condition nécessaire d'optimalité du 1^{er} ordre asymptotique].
- (iv) $\liminf_{k \rightarrow +\infty} \left[\liminf_{x \rightarrow v_k} \frac{f(x) - f(v_k) - \langle \nabla_G f(v_k), x - v_k \rangle}{\|x - v_k\|^2} \right] \geq 0$ (2.14_∞)
[condition nécessaire d'optimalité du 2nd ordre asymptotique; une sorte de version "asymptotisée" de (2.14)]

Démonstration : Pour k entier ≥ 1 , soit $\varepsilon_k := f(x_k) - \bar{f} + \frac{1}{k}$. Par construction, $\varepsilon_k > 0$, et par hypothèse $\varepsilon_k \rightarrow 0$. Évidemment – et cela a été fait pour :

$$f(x_k) < \bar{f} + \varepsilon_k.$$

Appliquons le théorème de BORWEIN- PREISS avec $u = x_k$, $\varepsilon = \varepsilon_k$ et $\lambda_k = (\varepsilon_k)^{1/3}$ par exemple. Il existe alors v_k et w_k tels que :

- $f(v_k) < \bar{f} + \varepsilon_k$, d'où $f(v_k) \rightarrow \bar{f}$ quand $k \rightarrow +\infty$;
- $\|v_k - x_k\| < \lambda_k = (\varepsilon_k)^{1/3}$, d'où $\|v_k - x_k\| \rightarrow 0$ quand $k \rightarrow +\infty$;
- $\|s_k = \nabla_G f(v_k)\| < \frac{2\varepsilon_k}{\lambda_k} = 2(\varepsilon_k)^{2/3}$, d'où $\|\nabla_G f(v_k)\| \rightarrow 0$ quand $k \rightarrow +\infty$.

Par ailleurs, appliquant la condition (C2) du corollaire 2.8 de la page 39, gardant à l'esprit que $s_k = \nabla p(v_k) = \nabla_G f(v_k)$,

$$\liminf_{x \rightarrow v_k} \frac{f(x) - f(v_k) - \langle \nabla_G f(v_k), x - v_k \rangle}{\|x - v_k\|^2} \geq -\frac{\varepsilon_k}{\lambda_k^2} = -(\varepsilon_k)^{1/3}.$$

L'inégalité (2.14_∞) s'ensuit. □

La démonstration du Théorème de BORWEIN- PREISS n'est pas facile, en tout cas pas aussi directe que celle d'EKELAND. Voici ce qu'on peut en dire :

- Si H est de dimension finie (H espace euclidien), il est possible d'en faire une démonstration dans l'esprit de celle du Théorème 2.5 de la page 36.
- Dans un contexte d'espace de Hilbert, outre la démonstration d'origine dans [BP], il y a celle de Clarke, Ledyaev, Stern et Wolenski dans leur livre ([CLSW], Chap. 1, § 4 et 5), mais il faut avoir traité d'autres choses avant

(l'inf-convolution avec des fonctions quadratiques)... c'est souvent comme cela en mathématiques.

Dans un contexte encore plus général, E est un espace de Banach, le théorème de BORWEIN- PREISS a fait des petits, il y a de nombreux articles qui ont été écrits sur le sujet, [FHV] en est un exemple choisi. Le Chapitre 8 de [Sc] est entièrement consacré à ces principes variationnels.

2.2 Applications en théorie de l'approximation hilbertienne

Le problème-modèle en approximation hilbertienne est le suivant :

Étant donné $x \in H$ (espace de Hilbert), S une partie fermée non vide de H , résoudre le problème de minimisation suivant

$$(\mathcal{P}_x) \begin{cases} \text{Minimiser } \|x - c\| \text{ (ou, ce qui revient au même, } \frac{1}{2} \|x - c\|^2) \\ c \in S. \end{cases}$$

Comme $\|\cdot\|$ est la norme hilbertienne, on a bien fait de "lisser" la fonction-objectif en prenant $f(x) := \frac{1}{2} \|x - c\|^2$. La fonction f se trouve être \mathcal{C}^∞ et convexe sur H (quadratique convexe, de fait).

Il y a deux objets mathématiques importants associés à la résolution de (\mathcal{P}_x) , à savoir :

– la *fonction-distance* d_S (ou ses associés)

$$\begin{aligned} d_S : H &\rightarrow \mathbb{R} \\ x &\mapsto d_S(x) := \inf_{c \in S} \|x - c\|. \end{aligned}$$

– la "*multiapplication*" solutions de (\mathcal{P}_x) , ou *multiapplication-projection* sur S

$$\begin{aligned} P_S : H &\rightrightarrows H \\ x &\mapsto P_S(x) := \{c \in S \mid \|x - c\| = d_S(x)\}. \end{aligned}$$

Au fond, P_S est une application de H dans $\mathcal{P}(S)$... et, bien entendu, $P_S(x)$ peut être vide. Quand $P_S(x)$ est réduit à un seul élément, un singleton donc, nous écrirons $P_S(x) = p_S(x)$ (grand P vs. petit p).

2.2.1 La fonction-distance et ses associés

* **Premières propriétés de la fonction-distance d_S**

- d_S est (toujours) 1-Lipschitz sur H , c'est-à-dire :

$$\forall u, v \in H, |d_S(u) - d_S(v)| \leq \|u - v\|. \quad (2.15)$$

Démonstration : la faire sous forme d'exercice.

C'est une propriété globale assez étonnante car S peut être extrêmement compliqué comme ensemble...

- Définition "duale" de $d_S(x)$, $x \notin S$:

$$d_S(x) = \sup \{r \geq 0 \mid \overline{B}(x, r) \cap S = \emptyset\} \quad (2.16)$$

où $\overline{B}(x, r)$ désigne la boule fermée de centre x et de rayon r . Un petit dessin aide à la compréhension géométrique de (2.16).

- d_S est convexe si et seulement si S est convexe (il en est de même de d_S^2).
Démonstration : la faire sous forme d'exercice.

- La fonction $\varphi_S : H \rightarrow \mathbb{R}$ définie par

$$\varphi_S(x) := \frac{1}{2} [\|x\|^2 - d_S^2(x)] \quad (2.17)$$

est toujours convexe.

En voilà une propriété étonnante !... car, ne l'oublions pas, S est un fermé quelconque ! La démonstration en est facile : il suffit d'exprimer φ_S comme le supremum d'une famille de fonctions (clairement) convexes.

Une conséquence est que

$$x \mapsto \frac{1}{2} d_S^2(x) = \frac{1}{2} \|x\|^2 - \varphi_S(x) \quad (2.18)$$

est (toujours) la différence de deux fonctions convexes, dont une $(\frac{1}{2} \|\cdot\|^2)$ est même convexe \mathcal{C}^∞ .

La classe $DC(H)$ de fonctions "différences-de-convexes" sur H est importante dans les problèmes variationnels non convexes ; on y reviendra abondamment au Chapitre 5.

Retenons de ce paragraphe qu'il y a trois fonctions importantes associées au problème (\mathcal{P}_x) :

la fonction-distance d_S ;

sa version "adoucie" $\frac{1}{2} d_S^2$ (car, élever au carré adoucit les mœurs...) ;

la fonction convexe φ_S .

La fonction distance d_S ne fait pas la différence entre la frontière $\text{Fr } S$ de S et son intérieur $\overset{\circ}{S}$:

$$\{x \in H \mid d_S(x) \leq 0\} = \{x \in H \mid d_S(x) = 0\} = S = (\text{Fr } S) \cup \overset{\circ}{S}.$$

Il y a une fonction qui fait ça, c'est une cousine de d_S , la *fonction-distance signée* Δ_S , définie comme suit :

$$\Delta_S(x) := \begin{cases} d_S(x) & \text{si } x \notin S, \\ -d_{S^c}(x) & \text{si } x \in S. \end{cases} \quad [S^c \text{ est le complémentaire de } S \text{ dans } H]$$

On a supposé implicitement que S , outre le fait de ne pas être vide, n'est pas tout l'espace H . Sous une forme d'écriture plus ramassée,

$$\Delta_S = d_S - d_{S^c}.$$

Voici quelques propriétés de la fonction Δ_S , qu'on pourra démontrer sous forme d'exercices :

$$\{x \in H \mid \Delta_S(x) > 0\} = S^c,$$

$$\{x \in H \mid \Delta_S(x) = 0\} = \text{Fr } S,$$

$$\{x \in H \mid \Delta_S(x) < 0\} = \overset{\circ}{S}, \quad (\text{un petit dessin peut aider à la compréhension de ces propriétés})$$

$$\Delta_{S^c} = -\Delta_S \quad (\text{il n'y a pas d'ambiguïté dans la définition puisque } d_{S^c} = d_{\overline{S^c}})$$

$$\Delta_S \text{ est 1-Lipschitz sur } H$$

$$\Delta_S \text{ est convexe si et seulement si } S \text{ est convexe.}$$

* Quid de la différentiabilité de d_S , de d_S^2 ?

- Si $x \in \overset{\circ}{S}$, la question ne se pose pas : d_S est nulle dans un voisinage de x , donc d_S est (Fréchet-) différentiable en x et $\nabla d_S(x) = 0$.
- Si $x \in \text{Fr } S$, la question se pose : d_S peut être différentiable en x (essayez avec un petit dessin dans le plan!), même s'il est plus probable que d_S ne soit pas différentiable en x . En tout cas, si d_S est différentiable en $x \in \text{Fr } S$, alors $\nabla d_S(x) = 0$ nécessairement (ayez un réflexe variationnel ! la fonction d_S est minimisée en x , et d_S a été supposée différentiable en x). Un autre point d'intérêt : la fonction d_S^2 est toujours différentiable en $x \in \text{Fr } S$ (c'est toujours l'effet adoucissant du passage au carré) avec, bien sûr, $\nabla d_S^2(x) = 0$.

En effet, si $x \in \text{Fr } S$, $d_S(x) = 0$, de sorte que

$$\begin{aligned} |d_S(x+h) + d_S(x)| &= |d_S(x+h) - d_S(x)| \leq \|h\|, \\ |d_S(x+h) - d_S(x)| &\leq \|h\|, \end{aligned}$$

grâce à la propriété de d_S d'être 1-Lipschitz sur H . Ainsi

$$|d_S^2(x+h) - d_S^2(x)| \leq \|h\|^2,$$

ce qui assure ce qui été annoncé.

- Si $x \notin S$, la fonction d_S (ou $\frac{1}{2} d_S^2$) peut être différentiable en x comme elle peut ne pas l'être. En tout cas, si d_S est différentiable en x ,

$$\nabla \left(\frac{1}{2} d_S^2 \right) (x) = d_S(x) \nabla d_S(x). \quad (2.19)$$

En clair :

$\{x \notin S \mid d_S \text{ est différentiable en } x\} = \{x \notin S \mid d_S^2 \text{ est différentiable en } x\}$,
 $(d_S \text{ différentiable sur } S^c) \Leftrightarrow (d_S^2 \text{ différentiable sur } H)$
 [grâce à ce qui a été signalé au point précédent]

2.2.2 La multiapplication-projection sur S

* Caractérisation des éléments de $P_S(x)$

Théorème 2.10 (caractérisation de " \bar{x} est un projeté de x sur S ")

Soit $x \notin S$. Les assertions suivantes sont équivalentes :

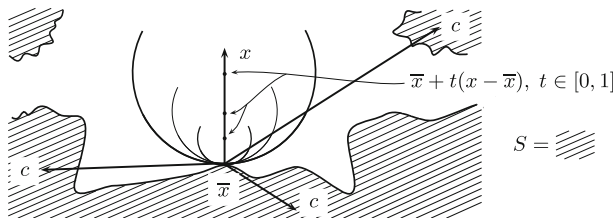
- (i) $\bar{x} \in P_S(x)$ (i.e., $\bar{x} \in S$ et $\|x - \bar{x}\| = d_S(x)$);
- (ii) $\bar{x} \in S$ et

$$\forall c \in S, \langle x - \bar{x}, c - \bar{x} \rangle \leq \frac{1}{2} \|c - \bar{x}\|^2; \quad (2.20)$$

- (iii) $\bar{x} \in S$ et

$$\forall t \in]0, 1], \bar{x} \in P_S[\bar{x} + t(x - \bar{x})]. \quad (2.21)$$

Il est assez étonnant qu'on obtienne une *caractérisation* des solutions de notre problème (\mathcal{P}_x)... Avec (2.20) on a une *condition nécessaire et suffisante d'optimalité globale* dans un problème qui n'est pas convexe ! La démonstration du théorème est facile, c'est du pur calcul hilbertien sur la norme (ou plutôt son carré).



On voit sur cette figure que $\langle x - \bar{x}, c - \bar{x} \rangle$ peut être positif, une chose qu'on n'a pas lorsque S est convexe.

Démonstration du théorème : On allégera l'écriture en ne répétant pas "pour tout $c \in S$ " dans les assertions.

(i) signifie :

$$\begin{aligned}
 & \bar{x} \in S \text{ et } \|x - \bar{x}\| \leq \|x - c\| \text{ pour tout } c \in S \\
 \Leftrightarrow & \bar{x} \in S \text{ et } \|x - \bar{x}\|^2 \leq \|x - c\|^2 \\
 \Leftrightarrow & \bar{x} \in S \text{ et } \|x - \bar{x}\|^2 \leq \|x - \bar{x}\|^2 + \|\bar{x} - c\|^2 + 2 \langle x - \bar{x}, \bar{x} - c \rangle \\
 & \quad [\text{utilisant le fait que } \|x - c\|^2 = \|x - \bar{x} + \bar{x} - c\|^2] \\
 \Leftrightarrow & \bar{x} \in S \text{ et } 2 \langle x - \bar{x}, c - \bar{x} \rangle \leq \|c - \bar{x}\|^2, \tag{2.22}
 \end{aligned}$$

qui n'est autre que (ii).

Par ailleurs, (2.22) est équivalent à :

$$\begin{aligned}
 & \bar{x} \in S \text{ et } 2 \langle x - \bar{x}, c - \bar{x} \rangle \leq \frac{1}{t} \|c - \bar{x}\|^2 \text{ pour tout } t \in]0, 1] \\
 \Leftrightarrow & \bar{x} \in S \text{ et } 2 \langle [\bar{x} + t(x - \bar{x})] - \bar{x}, c - \bar{x} \rangle \leq \|c - \bar{x}\|^2 \text{ pour tout } t \in]0, 1].
 \end{aligned}$$

Grâce à ce qui a été démontré plus haut, ceci est précisément la caractérisation du fait que $\bar{x} \in P_S[\bar{x} + t(x - \bar{x})]$. \square

Remarques :

- Évidemment, $P_S(x) = \{x\}$ lorsque $x \in S$.
Si $x \notin S$ et que $\bar{x} \in P_S(x)$, dès lors que $t \in]0, 1]$, \bar{x} se trouve être l'unique projeté sur S de $x_t := x + t(\bar{x} - x)$. Cela se "voit" sur la figure de cette même page, et se démontre facilement. Posons $\alpha := d_S(x)$. La boule $\bar{B}(x, \alpha)$ ne peut rencontrer S qu'à sa frontière ($\bar{y} \in S$ et $\|x - \bar{y}\| < \alpha$ contredit la définition de $\alpha = d_S(x)$). Donc $\bar{B}(x, \alpha) \cap S = [\text{Sphère}(x, \alpha)] \cap S$. Par suite, $\bar{B}(x_t, \|x_t - \bar{x}\|)$ ne rencontre S qu'en \bar{x} , c'est-à-dire

$$P_S(x_t) = \{\bar{x}\}.$$

- La caractérisation (2.20) est une sorte d'inéquation variationnelle qui rappelle celle caractérisant le projeté \bar{x} de x sur S lorsque S est convexe, à savoir :

$$\bar{x} \in S \text{ et } \langle x - \bar{x}, c - \bar{x} \rangle \leq 0 \text{ pour tout } c \in S. \quad (2.23)$$

Une question qui vient à l'esprit naturellement ici est : Comment se fait-il que le terme quadratique à droite de l'inégalité (2.20) ait disparu quand S est convexe ? Voici la réponse. Partons de l'inégalité dans (2.20). Pour un choix de $c \in S$ (convexe), considérons $c_\alpha := \bar{x} + \alpha(c - \bar{x})$ avec $\alpha \in]0, 1[$. Puisque S est convexe, c_α est encore dans S ; il vérifie donc l'inégalité de (2.20) :

$$\langle x - \bar{x}, [\bar{x} - \alpha(c - \bar{x})] - \bar{x} \rangle \leq \frac{1}{2} \|[\bar{x} + \alpha(c - \bar{x})] - \bar{x}\|^2,$$

soit

$$\langle x - \bar{x}, c - \bar{x} \rangle \leq \frac{\alpha}{2} \|c - \bar{x}\|^2.$$

Un passage à la limite, $\alpha \rightarrow 0$, conduit à l'inégalité espérée (2.23).

* Propriétés de la multiapplication P_S

Elles sont rassemblées dans la proposition suivante.

Proposition 2.11

- (i) $P_S(x)$ est une partie fermée bornée de S .
- (ii) Si $x \notin S$, $P_S(x) \subset \text{Fr } S$.
- (iii) Le graphe de P_S , à savoir $\{(x, y) \mid y \in P_S(x)\}$ est fermé dans $H \times H$.
- (iv) La multiapplication P_S est localement bornée, c'est-à-dire : si $B \subset H$ est borné,

$$P_S(B) := \{y \in P_S(x) \mid x \in B\} \text{ est borné.}$$

- (v) P_S est une *multiapplication monotone* (croissante), c'est-à-dire vérifiant :

$$\left(\begin{array}{c} x, x' \in H \\ y \in P_S(x), y' \in P_S(x') \end{array} \right) \Rightarrow (\langle y - y', x - x' \rangle \geq 0). \quad (2.24)$$

Démonstration : Les points (i) à (iv) sont faciles à démontrer à partir de la définition de $P_S(x)$ ou de la caractérisation de $\bar{x} \in P_S(x)$ (cf. sous-paragraphe précédent). Contentons-nous de vérifier (v).

À partir de la caractérisation de $y \in P_S(x)$, $y' \in P_S(x')$, on a :

$$\begin{aligned}\langle x - y, y' - y \rangle &\leq \frac{1}{2} \|y' - y\|^2 \quad (\text{choix particulier de } c = y'), \\ \langle x' - y', y - y' \rangle &\leq \frac{1}{2} \|y - y'\|^2 \quad (\text{choix particulier de } c = y).\end{aligned}$$

Par suite, en additionnant les deux inégalités au-dessus :

$$\langle x - x' + y' - y, y' - y \rangle \leq \|y - y'\|^2,$$

soit

$$\langle x - x', y' - y \rangle \leq 0.$$

□

2.2.3 Différentiabilité de d_S vs. unicité de la projection sur S

Il y a un lien étonnant entre la différentiabilité de d_S en $x \notin S$ et le fait que x admette une projection sur S au plus.

Proposition 2.12 Soit $x \notin S$.

- (i) Si d_S est différentiable en x (au sens de Gâteaux suffit), alors le problème d'approximation (\mathcal{P}_x) a *au plus* une solution. Si $\bar{x} = p_S(x)$, alors :

$$\nabla d_S(x) = \frac{x - \bar{x}}{\|x - \bar{x}\|}. \quad (2.25)$$

- (ii) Réciproque lorsque H est de dimension finie. Si $P_S(x)$ est réduit à un seul élément, alors d_S est différentiable en x (au sens de Fréchet même).

Démonstration.

- (i) Soit $\bar{x} \in P_S(x)$ (if any !). Considérons $t \in]0, 1]$ et formons le quotient différentiel

$$q_t := \frac{d_S[x + t(\bar{x} - x)] - d_S(x)}{t}.$$

La propriété de 1-Lipschitz sur H de d_S fait que :

$$d_S[x + t(\bar{x} - x)] = d_S[x + t(\bar{x} - x)] - d_S(\bar{x}) \leq (1 - t) \|x - \bar{x}\|.$$

Puisque $d_S(x) = \|x - \bar{x}\|$,

$$q_t \leq -\|x - \bar{x}\|.$$

Comme d_S a été supposée Gâteaux-différentiable en x , un passage à la limite ($t \rightarrow 0$) dans l'inégalité au-dessus conduit à

$$\langle \nabla d_S(x), \bar{x} - x \rangle \leq -\|x - \bar{x}\|. \quad (2.26)$$

Par la propriété de 1-Lipschitz sur H de d_S , on sait que $\|\nabla d_S(x)\| \leq 1$ nécessairement (on est d'accord?). Il résulte donc de l'inégalité de Cauchy-Schwarz et de (2.26) :

$$\langle \nabla d_S(x), \frac{\bar{x} - x}{\|\bar{x} - x\|} \rangle = -1.$$

Ceci impose que $\|\nabla d_S(x)\| = 1$. On est donc dans le cas d'égalité de l'inégalité de Cauchy-Schwarz, ce qui donne

$$\nabla d_S(x) = -\frac{\bar{x} - x}{\|\bar{x} - x\|} = \frac{x - \bar{x}}{d_S(x)}.$$

Le vecteur (unitaire) $\nabla d_S(x)$ ne peut pointer dans deux directions différentes, il n'y a donc qu'un \bar{x} dans $P_S(x)$ (lorsqu'il y en a).

(ii) La démonstration de la réciproque est laissée sous forme d'exercice. \square

Remarques.

- La Proposition 2.12 ne dit pas qu'il y a une solution au problème (\mathcal{P}_x) ... Le test d'existence est le suivant (en présence de différentiabilité de d_S en x , bien sûr) :
Si $\bar{x} := x - d_S(x)\nabla d_S(x) \in S$, (\mathcal{P}_x) a pour solution \bar{x} ; si $\bar{x} \notin S$, (\mathcal{P}_x) n'a pas de solution.
- La différentiabilité des fonctions cousines $\frac{1}{2}d_S^2$ et φ_S est, bien sûr, liée à celle de d_S . Si d_S est différentiable en $x \notin S$ et que $P_S(x) = \{\bar{x}\}$, il en est de même de $\frac{1}{2}d_S^2$ et φ_S avec

$$\begin{aligned} \nabla \left(\frac{1}{2} d_S^2 \right) (x) &= x - \bar{x}, \\ \nabla \varphi_S(x) &= \bar{x}. \end{aligned}$$

La fonction φ_S apparaît donc comme une "fonction primitive de la projection sur S " (for whatever it means).

2.2.4 Existence et unicité générique en approximation hilbertienne

Quand (\mathcal{P}_x) a-t-il une solution ? Quand (\mathcal{P}_x) a-t-il une et une seule solution ? Nous montrons ici que c'est "presque toujours" le cas. Évidemment, les questions posées concernent les points $x \notin S$.

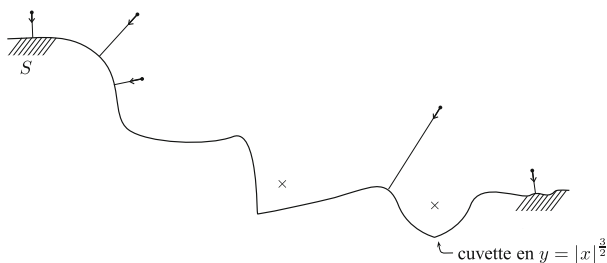
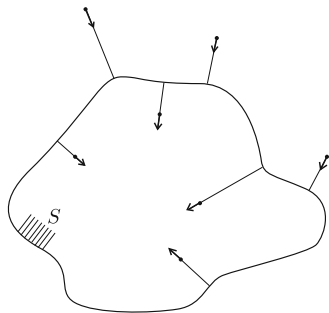


Fig. 2.1 Champ de gradients de d_S , pointant toujours vers S .

\times : points de non-différentiabilité de d_S

Fig. 2.2 Champ de gradients de la fonction-distance signée Δ_S



Théorème 2.13 On a :

- (i) $\{x \in H \mid P_S(x) \neq \emptyset\}$ est *dense* dans H
- (ii) $\{x \in H, P_S(x) \text{ est un singleton}\}$ est *dense* dans H .

Démonstration.

- (i) Soit $z \notin S$ et $\eta > 0$; il s'agit de trouver \tilde{z} tel que : $\|z - \tilde{z}\| \leq \eta$ et $P_S(\tilde{z}) \neq \emptyset$.

Fixons $\varepsilon > 0$ tel que $\varepsilon [d_S(z) + 3] < \eta$... choix bizarre, mais nous verrons pourquoi il a été fait.

Prenons $c_0 \in S$ tel que

$$\text{et } \left. \begin{aligned} \|c_0 - z\|^2 &< d_S^2(z) + \varepsilon \\ \|c_0 - z\| &< d_S(z) + 1. \end{aligned} \right\} \quad (2.27)$$

(c'est tout à fait possible, il suffit de penser à la définition de $d_S(z)$).

Nous allons appliquer le théorème de BORWEIN- PREISS à la fonction $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ que voici :

$$\forall x \in H, f(x) := \|x - z\|^2 + i_S(x).$$

La fonction f est s.c.i. sur H (somme d'une fonction continue et d'une fonction s.c.i.), minorée sur H par 0. En fait, $\inf_H f = d_S^2(z)$.

Par construction (cf. (2.27)), c_0 est un minimiseur à ε près de f sur H . D'après le théorème de Borwein-Preiss, appliqué avec le choix de $\lambda = 1$, il existe v et $w \in H$ tels que :

$$(\alpha) \quad \|v - c_0\| < 1, \|w - v\| < 1$$

$$(\beta) \quad v \in S \text{ et est un minimiseur de } x \mapsto f(x) + \varepsilon \|x - w\|^2 \text{ sur } H.$$

Explicitons ce que dit (β) :

$$\|v - z\|^2 + \varepsilon \|v - w\|^2 \leq \|c - z\|^2 + \varepsilon \|c - w\|^2 \text{ pour tout } c \in S,$$

ce qui est la même chose que

$$\|v - z\|^2 - \|c - z\|^2 \leq \varepsilon [\|c - w\|^2 - \|v - w\|^2] \text{ pour tout } c \in S. \quad (2.28)$$

Or

$$\begin{aligned} \|v - z\|^2 - \|c - z\|^2 &= \|v - z\|^2 - \|c - v + v - z\|^2 \\ &= -\|c - v\|^2 + 2 \langle v - z, c - v \rangle, \end{aligned}$$

$$\begin{aligned} \|c - w\|^2 - \|v - w\|^2 &= \|c - v + v - w\|^2 - \|v - w\|^2 \\ &= \|c - v\|^2 + 2 \langle c - v, v - w \rangle. \end{aligned}$$

Ainsi, (2.28) devient :

$$- [2 \langle v - z, c - v \rangle + \|c - v\|^2] \leq \varepsilon [2 \langle v - w, c - v \rangle + \|c - v\|^2],$$

d'où

$$2 \langle z - v + \varepsilon(w - v), c - v \rangle \leq (1 + \varepsilon) \|c - v\|^2;$$

$$\left\langle \frac{z - v + \varepsilon(w - v)}{1 + \varepsilon}, c - v \right\rangle \leq \frac{1}{2} \|c - v\|^2. \quad (2.29)$$

En définissant $\tilde{z} := v + \frac{z - v + \varepsilon(w - v)}{1 + \varepsilon}$, on s'assure (d'après (2.29)) que

$$\langle \tilde{z} - v, c - v \rangle \leq \frac{1}{2} \|c - v\|^2,$$

et ce pour tout $c \in S$. Or ceci est précisément la caractérisation du fait que v , dont on sait déjà qu'il est dans S , est un élément de $P_S(\tilde{z})$ (cf.

l'inégalité de caractérisation (2.20)).

Ensuite,

$$\begin{aligned}
 \|\tilde{z} - z\| &= \left\| v + \frac{z - v + \varepsilon(w - v)}{1 + \varepsilon} \right\| \\
 &= \left\| \left(1 - \frac{1}{1 + \varepsilon}\right)(v - z) + \frac{\varepsilon}{1 + \varepsilon}(w - v) \right\| \\
 &\leq \varepsilon \|v - z + w - v\| \\
 &\leq \varepsilon \left[\|v - c_0\| + \|c_0 - z\| + \|w - v\| \right] \\
 &\leq \varepsilon \left[1 + (d_S(z) + 1) + 1 \right] \text{ (cf. (2.27) et } (\alpha)) \\
 &\leq \varepsilon \left[3 + d_S(z) \right] \leq \eta.
 \end{aligned}$$

En somme, on a trouvé \tilde{z} tel que : $\|\tilde{z} - z\| \leq \eta$ et $v \in P_S(\tilde{z})$.

- (ii) À partir du moment où $P_S(z) \neq \emptyset$, $z \notin S$, on sait que pour $z_t = z + t(\bar{z} - z)$, $t \in]0, 1]$, $\bar{z} \in P_S(z)$, $P_S(z_t) = \{\bar{z}\}$ (cf. la 1^{ère} remarque dans la page 46). On peut donc prendre z_t aussi proche de z que voulu. Le résultat de densité annoncé s'ensuit. \square

Quand on projette $x \notin S$ sur S , quels points de $\text{Fr } S$ touche-t-on ? En fait, "presque tous" : "presque tout point de $\text{Fr } S$ est le projeté de quelqu'un". En termes mathématiques, cela donne le théorème suivant.

Théorème 2.14 On a :

$P_S(S^c) := \{\bar{x} \in P_S(x), x \notin S\}$ est une partie dense de $\text{Fr } S$.

Démonstration. Soit $x_f \in \text{Fr } S$ et $\eta > 0$. Le résultat du théorème précédent nous permet d'affirmer qu'il existe $x \notin S$ tel que : $\|x_f - x\| \leq \frac{\eta}{2}$ et $P_S(x) \neq \emptyset$. Ainsi, tout point \bar{x} de $P_S(x)$ est dans $\{\bar{x} \in P_S(x), x \notin S\}$ bien sûr, et

$$\|\bar{x} - x_f\| \leq \|\bar{x} - x\| + \|x - x_f\| \leq 2 \|x - x_f\| \leq \eta.$$

Le résultat de densité annoncé est ainsi démontré. \square

Retenons la portée générale des deux théorèmes de densité démontrés dans ce § 2.2.4 : H est un espace de Hilbert et S est un fermé quelconque de H !

3 Prolongements possibles

Les principes variationnels par perturbations de la fonction originelle à minimiser ne s'arrêtent pas à ceux exposés aux § 1 et 2. Un exemple additionnel est *le principe variationnel de C. Stegall (1978)*; son énoncé étant simple, donnons-le.

Soit $S \subset H$ *fermé borné* (non vide), soit $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$, finie en au moins un point de S , *semicontinue inférieurement* sur H , et *bornée inférieurement* sur S . Alors, pour un ensemble *dense* de points a de H , le problème de la minimisation de (la fonction perturbée) $x \mapsto f(x) - \langle a, x \rangle$ sur S a *une et une seule* solution.

Nous ne faisons que signaler l'existence d'un autre principe variationnel (du même acabit) dans des espaces de Banach (d'un certain type), c'est celui de Deville, Godefroy et Zizler [DGZ]. Traiter de tous ces principes variationnels occuperait presque tout le Cours... Ce n'est pas notre objectif : les principes variationnels de ce chapitre sont des *outils* dont chacun pourra se servir dans le contexte d'application qui est le sien.

Annexe

On rappelle dans cette annexe les trois types de différentiabilité utilisées en analyse et calcul variationnel, dans le contexte des fonctions numériques seulement.

Soit donc $(E, \|\cdot\|)$ espace de Banach et $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ finie dans un voisinage de x .

F-différentiabilité. C'est la différentielle usuelle, telle qu'étudiée en L3. On dit que f est différentiable au sens de M. Fréchet (F-différentiable en abrégé) en x s'il existe $l^* \in E^*$ telle que

$$\frac{f(x+u) - f(x) - \langle l^*, u \rangle}{\|u\|} \rightarrow 0 \text{ quand } u \neq 0 \rightarrow 0$$

(ou encore : $f(x+u) = f(x) + \langle l^*, u \rangle + o(\|u\|)$)

l^* , noté $D_F f(x)$ ou simplement $Df(x)$, est un élément de E^* .

Si l'espace source de f est un espace de Hilbert $(H, \langle \cdot, \cdot \rangle)$, la forme linéaire continue $D_F f(x)$ est *représentée* par un élément de H , noté $\nabla_F f(x)$

(ou $\nabla f(x)$ simplement) et appelé gradient de f en x :

$$\forall d \in H, D_F f(x) d = \langle \nabla_F f(x), d \rangle.$$

G-différentiabilité. On dit que f est différentiable au sens de R. Gâteaux (G-différentiable en abrégé) en x lorsque

$$\forall d \in E, \frac{f(x + \alpha d) - f(x)}{\alpha} \text{ a une limite lorsque } \alpha \rightarrow 0,$$

et que cette limite (qui dépend de d) est une forme linéaire continue de d :

$$\forall d \in E, \frac{f(x + \alpha d) - f(x)}{\alpha} \rightarrow \langle D_G f(x), d \rangle.$$

H-différentiabilité. Il y a une différentiabilité intermédiaire, au sens de J. Hadamard. Une manière de la présenter est comme ceci.

Soit \mathcal{B} la famille des *compacts* de E . On dit que f est différentiable au sens de J. Hadamard (H-différentiable en abrégé) en x lorsqu'il existe $l^* \in E^*$, noté $D_H f(x)$, telle que

$$\lim_{\alpha \rightarrow 0} \frac{f(x + \alpha d) - f(x)}{\alpha} = \langle D_H f(x), d \rangle \text{ uniformément pour } d \in S,$$

et ce pour tout $S \in \mathcal{B}$.

(2.30)

Cette manière d'exprimer les choses permet une comparaison directe avec la F-différentiabilité et la G-différentiabilité.

La F-différentiabilité de f en x s'écrit, de manière équivalente, comme dans la définition (2.30), en prenant pour \mathcal{B} la collection des *fermés bornés* de E .

La G-différentiabilité de f en x s'écrit, de manière équivalente, comme en (2.30), en prenant pour \mathcal{B} la collection des *ensembles finis de points* de E .

La comparaison entre les trois types de différentiabilité est maintenant claire :

$$(F\text{-différentiabilité}) \Rightarrow (H\text{-différentiabilité}) \Rightarrow (G\text{-différentiabilité}).$$

La H-différentiabilité (et donc la F-différentiabilité) de f en x implique la continuité de f en x ; ce n'est pas le cas pour la G-différentiabilité. La semi-continuité inférieure n'est pas acquise non plus avec la G-différentiabilité ;

ce qui fait qu'on a des énoncés de théorèmes avec des hypothèses comme "soit f s.c.i. et G-différentiable sur E ", laquelle est assurée avec "soit f F-différentiable sur E ".

Si E est de dimension finie

$$(H\text{-différentiabilité}) \Leftrightarrow (F\text{-différentiabilité}).$$

Si f vérifie une condition de Lipschitz dans un voisinage de x , alors

$$(G\text{-différentiabilité en } x) \Leftrightarrow (H\text{-différentiabilité en } x).$$

En pratique, dans un contexte de problèmes variationnels :

- la F-différentiabilité est une requête exigeante, souvent inaccessible... et pourtant beaucoup de résultats du Calcul différentiel reposent sur cette hypothèse.
- la G-différentiabilité est plus accessible, et souvent on commence par là, même pour accéder à la F-différentiabilité. Malheureusement, la G-différentiabilité ne permet pas les règles de calcul à la chaîne ("chain rules").

La dimension infinie pose des obstacles inattendus ; ainsi, même si $f : \mathcal{O} \subset E \rightarrow \mathbb{R}$ est Lipschitz et convexe dans un voisinage ouvert convexe \mathcal{O} de x , il peut y avoir un "gros trou" entre les différentiabilités G-H et F de f en x .

Fonctions continûment différentiables (de classe \mathcal{C}^1). Là, il n'y a pas de distinguo à faire (ouf !). Si \mathcal{O} est un ouvert de E , avoir f X-différentiable sur \mathcal{O} et $D_X f : \mathcal{O} \rightarrow E^*$ continue sur \mathcal{O} revient au même avec $X = G, H$ ou F .

Exercices

Exercice 1 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable, telle que $f(x)/\|x\| \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$ (c'est la 1-coercivité de f sur \mathbb{R}^n). Montrer qu'alors

$$\{\nabla f(x) \mid x \in \mathbb{R}^n\} = \mathbb{R}^n.$$

Hint : Pour $v \in \mathbb{R}^n$, considérer $g_v(x) := f(x) - \langle v, x \rangle$.

Exercice 2 Soit $f : E \rightarrow \mathbb{R}$ continue et Gâteaux-différentiable sur E (espace de Banach). On suppose qu'il existe $r > 0$ et c tels que :

$$\forall x \in E, f(x) \geq r \|x\| - c.$$

Montrer que $D_G f(E) := \{D_G f(x) \mid x \in E\}$ est dense dans rB^* (B^* est la boule unité de X^* pour la norme $\|\cdot\|_*$).

Hint : Étant donné $x^* \in rB^*$, considérer la fonction perturbée

$$g : x \in E \mapsto g(x) := f(x) - \langle x^*, x \rangle.$$

Appliquer à g le Corollaire 2.3 de la page 29.

Exercice 3 Soit $f : E \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 sur E (espace de Banach). On dit que f vérifie la condition (de compacité) de Palais-Smale lorsque :

$$\left. \begin{array}{l} (x_n) \subset E, (f(x_n))_n \text{ est bornée} \\ Df(x_n) \rightarrow 0 \text{ dans } X^* \end{array} \right\} \Rightarrow \left(\begin{array}{l} \text{il existe une sous-suite de } (x_n) \\ \text{qui converge (pour la topologie forte)} \end{array} \right).$$

Supposons donc que f vérifie la condition de Palais-Smale et qu'elle est bornée inférieurement sur E .

Montrer qu'il existe $\bar{x} \in E$ minimisant f sur E .

Hint : Appliquer le théorème d'Ekeland à f , avec $\varepsilon = \frac{1}{n}$ ($n \in \mathbb{N}^*$).

Exercice 4 (Minimisation approchée sur un sous-espace)

Soit $f : H \rightarrow \mathbb{R}$ semicontinue inférieurement et G-différentiable sur H (par exemple, f F-différentiable sur H couvre ces deux hypothèses). Soit V un sous-espace vectoriel fermé de H .

1) Montrer que si $\bar{x} \in V$ minimise f sur V , alors

$$\nabla f(\bar{x}) \in V^\perp.$$

2) Supposons f bornée inférieurement sur V . Montrer que pour tout $\varepsilon > 0$, il existe $\bar{x}_\varepsilon \in V$ vérifiant :

$$\left\{ \begin{array}{l} f(\bar{x}_\varepsilon) \leq \inf_V f + \varepsilon ; \\ |\langle \nabla f(\bar{x}_\varepsilon), d \rangle| \leq \varepsilon \text{ pour tout } d \in V \text{ tel que } \|d\| \leq 1. \end{array} \right.$$

Montrer que cette dernière condition équivaut à :

$$\nabla f(\bar{x}_\varepsilon) \in V^\perp + \overline{B}(0, \varepsilon).$$

Exercice 5 (Un théorème de point fixe inhabituel)

Soit $(X, \|\cdot\|)$ un espace de Banach, $\varphi : X \rightarrow \mathbb{R}$ une fonction semicontinue inférieurement et bornée inférieurement sur X . On considère $f : X \rightarrow X$ vérifiant

$$\|x - f(x)\| \leq \varphi(x) - \varphi[f(x)] \text{ pour tout } x \in X,$$

et on se propose de démontrer que f a un point fixe.

- 1) Montrer qu'il existe $\bar{x} \in X$ tel que

$$\varphi(\bar{x}) \leq \varphi(y) + \frac{1}{2} \|y - \bar{x}\| \text{ pour tout } y \in X.$$

En déduire que $\bar{x} = f(\bar{x})$.

- 2) Quelle différence essentielle voyez-vous entre ce résultat et les différents théorèmes de points fixes que vous avez rencontrés au cours de vos études ?

Exercice 6 (Un résultat inhabituel d'existence d'un minimiseur)

Soit $(X, \|\cdot\|)$ un espace de Banach. Soit $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction semicontinue inférieurement et bornée inférieurement sur X , non identiquement égale à $+\infty$. On fait l'hypothèse suivante : il existe $\alpha > 0$ tel que pour tout x vérifiant $f(x) > \inf_X f$, on peut trouver $\tilde{x} \neq x$ tel que

$$f(\tilde{x}) + \alpha \|x - \tilde{x}\| \leq f(x).$$

- 1) Montrer qu'il existe $\bar{x} \in X$ tel que $f(\bar{x}) = \inf_X f$.
- 2) Soit S l'ensemble des minimiseurs de f sur X . Montrer

$$d_S(x) \leq \frac{1}{\alpha} \left[f(x) - \inf_X f \right] \text{ pour tout } x \in X.$$

Exercice 7 (La règle de Fermat asymptotique)

Soit H un espace de Hilbert : $\langle \cdot, \cdot \rangle$ désigne le produit scalaire et $\|\cdot\|$ la norme associée. Si $f : H \rightarrow \mathbb{R}$ est minimisée en \bar{x} et qu'elle y est Gâteaux-différentiable, alors $\nabla f(\bar{x}) = 0$ (c'est la règle de Fermat). C'est la version "asymptotique" de cette règle que nous proposons d'établir dans cet exercice.

Considérons : $H \rightarrow \mathbb{R}$ semicontinue inférieurement, Gâteaux-différentiable sur H (par exemple, la Fréchet-différentiabilité de f sur H assure ces deux

conditions), et bornée inférieurement sur H . Montrer qu'il existe alors une suite (x_k) telle que :

$$f(x_k) \rightarrow \inf_H f \text{ et } \nabla f(x_k) \rightarrow 0 \quad \text{quand } k \rightarrow +\infty.$$

Références

- [E1] I. Ekeland. "On the variational principle". *J. Math. Anal. Appl.* 47 (1974), p. 324–353.
- [E2] I. Ekeland. "Nonconvex minimization problems". *Bull. Amer. Math. Soc.* 1 (1979), p. 443–474.
- [F] D.G. De Figueiredo. *Lectures on the Ekeland Variational Principle with Applications and Detours*. Tata Institute of Fundamental Research, Bombay, 1989.
- [BP] J.M. Borwein and D. Preiss. "A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions". *Trans. Amer. Math. Soc.* 303 (1987), p. 517–527.
- [L] P.D. Loewen. *Optimal Control Via Nonsmooth Analysis*. CRM Proceedings & Lecture notes, American Mathematical Society, 1993.
- [CLSW] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern and P.R. Wolenski. *Nonsmooth Analysis and Control Theory*. Graduate texts in mathematics, Springer Verlag, 1998.
- [FHV] M. Fabian, P. Hájek and J. Vanderwerff. "On smooth variational principles in Banach spaces". *J. Math. Anal. Appl.* 197 (1996), p. 153–173.
- [St] C. Stegall. "Optimization of functions on certain subsets of Banach spaces". *Math. Ann.* 236 (1978), p. 171–176.
- [DGZ] R. Deville, G. Godefroy and V.E. Zizler. "A smooth variational principle with applications to Hamilton-Jacobi equations in infinite dimensions". *J. Funct. Anal.* 111 (1993), p. 192–212.
- [Sc] W. Schirotzek. *Nonsmooth Analysis*. Universitext, Springer Verlag, 2007.
- [BZ] J.M. Borwein and Q.J. Zhu. *Techniques of Variational Analysis*. CMB books in mathematics, Springer Verlag, 2005.

Nous signalons les articles d'origine... il vaut mieux souvent revenir aux sources. L'article-revue [E2] reste, trente après sa publication, une très bonne référence pour l'énoncé et quelques-unes des premières applications du principe variationnel d'Ekeland. Notre § 2, sur le principe variationnel de Borwein-Preiss est tiré de ([L], Chap. 3).

<http://www.springer.com/978-3-642-30734-8>

Bases, outils et principes pour l'analyse variationnelle

Hiriart-Urruty, J.-B.

2013, XIII, 171 p. 36 ill., Softcover

ISBN: 978-3-642-30734-8