Chapitre 1

Sous-différentiabilité

On rappelle quelques résultats et propriétés concernant la sous-différentiabilité des fonctions convexes. Parmi les nombreux et excellents ouvrages traitant de ce sujet, on pourra consulter [Rockafellar (1970)], [Ekeland and Temam (1999)] et [Bauschke & Combettes (2011)]. Dans ce chapitre on fera souvent référence à [Gilbert (2018)], disponible sur demande auprès de l'auteur.

1.1 Définition et propriétés élémentaires

On note $\mathbb{U}$ l'espace de Hilbert de dimension finie $\mathbb{R}^n$, et on se donne une fonction F définie sur $\mathbb{U}$ à valeurs dans $]-\infty, +\infty]$. La fonction F sera dite propre si elle n'est pas identiquement égale à $+\infty$ (par définition, elle ne prend jamais la valeur $-\infty$). Le domaine (effectif) de la fonction F est le sous-ensemble de $\mathbb{R}^n$ sur lequel cette fonction est finie :

$$\mathrm{dom} F = \big\{ u \in \mathbb{U} \;,\;\; F(u) < +\infty \big\} \;.$$

Définition 1. Soit $u \in \mathbb{U}$ un élément du domaine de la fonction F. On dit que F est *sous-différentiable* en u si F possède une minorante affine exacte en u.[1] La pente r d'une telle minorante affine est appelée un *sous-gradient* de la fonction F en u ; l'ensemble de ces sous-gradients est appelé le *sous-différentiel* de F en u et est noté $\partial F(u)$. Dans le cas où la fonction F n'est pas sous-différentiable en u, on a $\partial F(u) = \emptyset$.

Cette définition fournit la caractérisation suivante du sous-différentiel de F en $u \in \mathrm{dom} F$:

$$r \in \partial F(u) \quad \Longleftrightarrow \quad F(v) \geq F(u) + \big\langle r \,, v - u \big\rangle \quad \forall v \in \mathbb{U} \;, \tag{1.1}$$

qui montre que $\partial F(u)$ est toujours un ensemble convexe fermé, éventuellement vide. Un premier résultat important pour l'optimisation découle directement de cette caractérisation :

$$u^\sharp \in \arg\min_{u \in \mathbb{U}} F(u) \quad \Longleftrightarrow \quad 0 \in \partial F(u^\sharp) \;. \tag{1.2}$$

Une autre conséquence de cette définition est le caractère monotone du sous-différentiel :

$$\big\langle r - s \,, u - v \big\rangle \geq 0 \quad \forall r \in \partial F(u) \;,\;\; \forall s \in \partial F(v) \;, \tag{1.3}$$

que l'on obtient en écrivant l'inégalité (1.1), d'une part au point v pour $r \in \partial F(u)$, d'autre part au point u pour $s \in \partial F(v)$, et en additionnant les deux résultats.

Dans le cas des fonctions convexes, on dispose du critère suivant de sous-différentiabilité.

1. Une minorante affine exacte de F en u est une fonction f de la forme $f(v) = \alpha + \langle r \,, v \rangle$, telle que $F(v) \geq f(v)$ pour tout $v \in \mathbb{U}$, avec $F(u) = f(u)$.

Théorème 1. *On suppose que F est une fonction convexe, bornée supérieurement par une constante finie au voisinage d'un point u_0. Alors, la fonction F est continue et sous-différentiable sur l'intérieur de son domaine (qui est non vide) :*

$$\partial F(u) \neq \emptyset \quad \forall u \in \mathrm{int}\big(\mathrm{dom} F\big) \ .$$

Pour la preuve, voir [Ekeland and Temam (1999), Chapter I, Proposition 2.5 & Proposition 5.2].

On dispose aussi du corollaire suivant.

Corollaire 1. *Toute fonction F convexe semi-continue inférieurement (s.c.i.) définie sur un espace de Hilbert $\mathbb{U}$ est continue et sous-différentiable sur l'intérieur de son domaine.*

Pour la preuve, voir [Ekeland and Temam (1999), Chapter I, Corollary 2.5].

Pour les fonctions convexes, les notions de différentiabilité et de sous-différentiabilité sont reliées par le théorème suivant.

Théorème 2. *On suppose que F est une fonction convexe. Si F est Gâteaux-différentiable*[2] *au point u, alors elle est sous-différentiable en u et on a : $\partial F(u) = \big\{\nabla F(u)\big\}$. Réciproquement, si F est sous-différentiable en u et si son sous-différentiel est réduit à un unique point r, alors F est Gâteaux-différentiable en u et on a : $\nabla F(u) = r$.*

Pour la preuve, voir [Ekeland and Temam (1999), Chapter I, Proposition 5.3].

Remarque 1. *(Notion facultative).* Soit G une fonction définie sur $\mathbb{U}$ à valeurs dans $[-\infty, +\infty[$, que l'on suppose *concave*. La fonction $F = -G$ est alors convexe, et on définit le *sur-différentiel*[3] de la fonction G en u, noté $\overline{\partial} G(u)$, de la manière suivante :

$$\overline{\partial} G(u) = -\partial(-G)(u) \ .$$

Avec cette définition, un élément $r \in \overline{\partial} G(u)$ vérifie l'inégalité :

$$G(u') \leq G(u) + \langle r \,, u' - u\rangle \quad \forall u' \in \mathbb{U} \ ,$$

et G possède alors une majorante affine exacte au point u. Les propriétés du sous-différentiel se traduisent sans difficulté en terme de sur-différentiel. $\diamond$

1.2 Lien avec la dérivée directionnelle $\ominus$

On rappelle la définition de la dérivée directionnelle d'une fonction.

Définition 2. Soit une fonction F définie sur $\mathbb{U}$ à valeurs dans $]-\infty, +\infty]$, propre. La dérivée directionnelle en u de la fonction F suivant la direction δ est définie par :

$$\mathrm{D}F(u)(\delta) = \lim_{h \to 0^+} \frac{F(u + h\delta) - F(u)}{h} \ .$$

On rappelle quelques propriétés classiques de la dérivée directionnelle.

2. Voir [Ekeland and Temam (1999), Chapter I, Definition 5.2] pour la différentielle au sens de Gâteaux.
3. aussi appelé sous-différentiel concave

Proposition 1. *Soit une fonction F définie sur $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$, convexe propre. Alors, la fonction $h \mapsto \big(F(u+h\delta)-F(u)\big)/h$ est croissante, la limite définissant $\mathrm{D}F(u)(\delta)$ existe dans $\overline{\mathbb{R}}$ et l'on a :*

$$\mathrm{D}F(u)(\delta) = \inf_{h>0} \frac{F(u+h\delta)-F(u)}{h} . \tag{1.4}$$

De plus, la fonction $\delta \mapsto \mathrm{D}F(u)(\delta)$ est convexe.

Démonstration. Soit $0 < h_1 \leq h_2$. La relation suivante est toujours vérifiée :

$$u + h_1\delta = \frac{h_1}{h_2}(u+h_2\delta) + \Big(1 - \frac{h_1}{h_2}\Big)u .$$

Par convexité de F, on obtient que :

$$\frac{F(u+h_1\delta)-F(u)}{h_1} \leq \frac{F(u+h_2\delta)-F(u)}{h_2} ,$$

et donc que l'application $h \mapsto \big(F(u+h\delta)-F(u)\big)/h$ est croissante. La relation (1.4) se déduit alors de la définition de la dérivée directionnelle. La convexité de la fonction $\delta \mapsto \mathrm{D}F(u)(\delta)$ est une conséquence directe de la relation (1.4) et de la convexité de F. □

Pour plus de détails sur l'application dérivée directionnelle et ses propriétés, on pourra consulter [Gilbert (2018), Chapitre 3, Proposition 3.14 & Proposition 3.15].

Remarque 2. Dans la proposition ci-dessus, si l'on suppose F concave plutôt que convexe, alors la fonction $h \mapsto \big(F(u+h\delta)-F(u)\big)/h$ est décroissante et on a :

$$\mathrm{D}F(u)(\delta) = \sup_{h>0} \frac{F(u+h\delta)-F(u)}{h} ,$$

et la limite existe dans $\overline{\mathbb{R}}$. ◇

Proposition 2. *Soit une fonction F définie sur $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$, convexe propre, continue en un point u. Alors, pour tout $\delta \in \mathbb{U}$,*

$$\mathrm{D}F(u)(\delta) = \sup_{r\in\partial F(u)} \langle r\,,\delta\rangle .$$

De plus, la fonction $\delta \mapsto \mathrm{D}F(u)(\delta)$ est continue.

Pour la preuve, voir [Gilbert (2018), Chapitre 3, Proposition 3.65].

On rappelle la notion de *fonction d'appui*. Soit U un sous-ensemble de $\mathbb{U}$. La fonction d'appui $\sigma_U : \mathbb{U} \to]-\infty,+\infty]$ est définie par :

$$\sigma_U(\delta) = \sup_{r\in U}\langle r\,,\delta\rangle .$$

Comme enveloppe supérieure de fonctions linéaires continues, la fonction d'appui σ_U est toujours convexe s.c.i.. On rappelle le résultat classique suivant (voir [Gilbert (2018), Chapitre 3, Proposition 3.10])

Proposition 3. *Soit U_1 et U_2 deux sous-ensembles de $\mathbb{U}$. Alors,*

$$\sigma_{U_1} = \sigma_{U_2} \quad\Longleftrightarrow\quad \mathrm{cl}\big(\mathrm{conv}(U_1)\big) = \mathrm{cl}\big(\mathrm{conv}(U_2)\big) ,$$

où $\mathrm{conv}(U)$ *(resp.* $\mathrm{cl}(U)$*) est l'enveloppe convexe (resp. la fermeture) de l'ensemble U.*

La proposition suivante fait le lien entre fonction d'appui et sous-différentiel.

Proposition 4. *Soit une fonction F définie sur $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$, convexe propre s.c.i. sous-différentiable. Si la dérivée directionnelle $\delta \mapsto \mathrm{D}F(u)(\delta)$ est s.c.i., alors on a pour tout $\delta \in \mathbb{U}$:*

$$\sigma_{\partial F(u)}(\delta) = \mathrm{D}F(u)(\delta) \ .$$

La preuve de cette proposition est une conséquence directe de [Aubin (1984), Proposition 3.8].

Exercice 1. Montrer que si F est est une fonction concave, semi-continue supérieurement (s.c.s.), sur-différentiable, et si sa dérivée directionnelle est s.c.s., le résultat précédent devient :

$$\sigma_{\overline{\partial} F(u)}(\delta) = -\mathrm{D}F(u)(-\delta) \ .$$

1.3 Lien avec la conjuguée de Fenchel

On rappelle la définition de la conjuguée de Fenchel d'une fonction.

Définition 3. Soit une fonction F définie sur $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$, propre. La conjuguée de Fenchel $F^\star$ de F est la fonction définie par :

$$F^\star(r) = \sup_{u\in\mathbb{U}} \langle r \,, u\rangle - F(u) \quad \forall r \in \mathbb{U} \ .$$

La conjuguée de Fenchel est donc une fonction définie sur $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$. La bi-conjuguée de Fenchel $F^{\star\star}$ de F est la conjuguée de Fenchel de $F^\star$.

On rappelle quelques propriétés classiques de la conjuguée de Fenchel (voir [Gilbert (2018), Chapitre 3, Proposition 3.43 & Proposition 3.45]).

Proposition 5.

1. *La conjuguée de Fenchel $F^\star$ de F est convexe s.c.i..*
2. *Si F est propre et possède une minorante affine, alors $F^\star$ est propre.*
3. *Si F est convexe s.c.i. propre, alors, $F^{\star\star} = F$.*

Le lien entre conjuguée de Fenchel et sous-différentiel est donné par le théorème suivant.

Théorème 3. *Soit F une fonction définie sur l'espace $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$ propre, et soit $u \in \mathrm{dom}F$. Alors,*

$$\partial F(u) = \big\{ r \in \mathbb{U},\ F(u) + F^\star(r) = \langle r \,, u\rangle \big\} \ .$$

Si on suppose de plus que F est convexe s.c.i., alors :

$$\partial F^\star(r) = \big\{ u \in \mathbb{U},\ F(u) + F^\star(r) = \langle r \,, u\rangle \big\} \ .$$

Démonstration. Par définition du sous-différentiel, on a :

$$\begin{aligned} r \in \partial F(u) \quad &\iff \quad F(v) \geq F(u) + \langle r \,, v-u\rangle \quad \forall v \in \mathbb{U} \ , \\ &\iff \quad \langle r \,, u\rangle - F(u) \geq \langle r \,, v\rangle - F(v) \quad \forall v \in \mathbb{U} \ , \\ &\iff \quad \langle r \,, u\rangle - F(u) \geq \sup_{v\in\mathbb{U}} \langle r \,, v\rangle - F(v) \ , \end{aligned}$$

ce qui montre que le sup en v est atteint au point u, d'où

$$r \in \partial F(u) \quad \iff \quad \langle r \,, u\rangle - F(u) = \sup_{v\in\mathbb{U}} \langle r \,, v\rangle - F(v) = F^\star(r) \ ,$$

d'où le premier résultat. Le second résultat est une conséquence directe du premier résultat appliqué à $F^\star$ et de l'égalité $F^{\star\star} = F$ dans le cas où F est convexe s.c.i. propre. ∎

1.4 Calcul sous-différentiel

Le résultat suivant est une conséquence directe de la définition de la sous-différentiabilité.

Proposition 6.

— *Soit F définie sur $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$, et soit $\lambda > 0$. Alors,*

$$\partial\big(\lambda F\big)(u) = \lambda \partial F(u) \qquad \forall u \in \mathbb{U} \; . \tag{1.5a}$$

— *Soit F et G définies sur $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$. Alors,*

$$\partial\big(F+G\big)(u) \supset \partial F(u) + \partial G(u) \qquad \forall u \in \mathbb{U} \; . \tag{1.5b}$$

Dans le cas convexe, le théorème suivant précise dans quel cas l'inclusion (1.5b) est en fait une égalité.

Théorème 4. *Soit F et G deux fonctions définies sur $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$, convexes et s.c.i.. On suppose qu'il existe $u_0 \in \mathrm{dom}F \cap \mathrm{dom}G$, tel que F ou G soit continue en u_0. Alors,*

$$\partial\big(F+G\big)(u) = \partial F(u) + \partial G(u) \qquad \forall u \in \mathbb{U} \; . \tag{1.6}$$

Pour la preuve, voir [Gilbert (2018), Chapitre 3, Proposition 3.70].

On s'intéresse enfin au cas d'une fonction composée. On considère l'espace $\mathbb{V} = \mathbb{R}^m$, une application linéaire continue Λ définie sur $\mathbb{V}$ à valeurs dans $\mathbb{U}$ et une fonction F définie sur $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$ convexe s.c.i..

Théorème 5. *Soit un point $v_0 \in \mathbb{V}$ tel que F est continue en $\Lambda \cdot v_0$. Alors*

$$\partial\big(F \circ \Lambda\big)(v) = \Lambda^\top \partial F(\Lambda \cdot v) \qquad \forall v \in \mathbb{V} \; . \tag{1.7}$$

Pour la preuve, voir [Gilbert (2018), Chapitre 3, Proposition 3.71].

1.5 Sous-différentiel et opérateur maximal monotone ⊖

Soit $\mathbb{U}$ l'espace de Hilbert de dimension finie $\mathbb{R}^n$, et soit M une multi-application $\mathbb{U} \rightrightarrows \mathbb{U}$, c'est-à-dire un opérateur qui à tout $u \in \mathbb{U}$ associe un sous-ensemble de $\mathbb{U}$. On rappelle les notions suivantes :

- *domaine* de M : $\mathrm{dom}M = \big\{u \in \mathbb{U} \; , \;\; M(u) \neq \emptyset\big\}$,
- *graphe* de M : $\mathrm{gr}M = \big\{(u,r) \in \mathbb{U} \times \mathbb{U} \; , \;\; r \in M(u)\big\}$,
- *inverse* de M : $M^{-1} : \mathbb{U} \rightrightarrows \mathbb{U}$, tel que $u \in M^{-1}(r) \iff r \in M(u)$.

Les deux définitions suivantes précisent les notions d'opérateur monotone et maximal.

Définition 4. On dit qu'un opérateur $M : \mathbb{U} \rightrightarrows \mathbb{U}$ est *monotone* s'il vérifie :

$$\big\langle r-s \, , u-v \big\rangle \geq 0 \quad \forall (u,r) \in \mathrm{gr}M \; , \;\; \forall (v,s) \in \mathrm{gr}M \; .$$

Définition 5. On dit qu'un opérateur $M : \mathbb{U} \rightrightarrows \mathbb{U}$ est *monotone maximal* s'il est monotone et s'il n'existe pas d'opérateur M' tel que le graphe de M soit strictement inclus dans celui de M' : si le couple $(v,s) \in \mathbb{U} \times \mathbb{U}$ est tel que $\big\langle r-s \, , u-v \big\rangle \geq 0$ pour tout $(u,r) \in \mathrm{gr}M$, alors $(v,s) \in \mathrm{gr}M$.

Les propriétés en tant qu'opérateur du sous-différentiel d'une fonction F sont données par le théorème suivant.

Théorème 6. *Le sous-différentiel d'une fonction F définie sur un espace de Hilbert de dimension finie $\mathbb{U}$ à valeurs dans $]-\infty,+\infty]$ convexe propre est un opérateur monotone. Si on suppose de plus que la fonction F est s.c.i., alors son sous-différentiel est un opérateur maximal monotone.*

Démonstration. La preuve du premier point a déjà été faite (voir l'équation (1.3)). Pour le second point, soit $(v,s) \in \mathbb{U} \times \mathbb{U}$, tel que

$$\big\langle r - s \,, u - v \big\rangle \geq 0 \quad \forall (u,r) \in \mathrm{gr}(\partial F) \ . \tag{1.8}$$

Considérons le problème :

$$\min_{u \in \mathbb{U}} F(u) + \frac{1}{2} \big\| u - (v+s) \big\|^2 \ .$$

Le critère de ce problème étant fortement convexe, propre, s.c.i. (puisque F l'est), il admet une solution unique $u^\sharp$ qui vérifie la condition d'optimalité :

$$v + s - u^\sharp \in \partial F(u^\sharp) \ .$$

Prenant alors $(u,r) = (u^\sharp, v+s-u^\sharp)$, l'inégalité (1.8) conduit à $\langle v - u^\sharp, u^\sharp - v \rangle \geq 0$, et donc $v = u^\sharp$. On déduit de la condition d'optimalité que $s \in \partial F(u^\sharp) = \partial F(v)$. On a donc $(v,s) \in \mathrm{gr}(\partial F)$ d'où le caractère maximal de l'opérateur. □

1.6 Sous-différentiel des fonctions marginales

On s'intéresse ici aux fonctions obtenues après une opération d'optimisation portant sur une partie des arguments d'une fonction. De telles fonctions sont appelées *fonctions marginales*, et on va étudier leurs propriétés selon les caractéristiques des fonctions dont elles proviennent.

1.6.1 Cas d'une fonction convexe-convexe

Soit J une fonction définie sur le produit $\mathbb{U} \times \mathbb{V}$ de deux espaces de Hilbert de dimension finie, à valeurs dans $]-\infty,+\infty]$, et soit U^{ad} un sous-ensemble de l'espace $\mathbb{U}$. On définit la *fonction marginale* de J :

$$\Phi(v) = \inf_{u \in U^{\mathrm{ad}}} J(u,v) \ , \tag{1.9}$$

ainsi que l'ensemble des solutions associées (éventuellement vide) :

$$\widehat{U}(v) = \operatorname*{arg\,min}_{u \in U^{\mathrm{ad}}} J(u,v) \ .$$

On rappelle quelques propriétés élémentaires de la fonction marginale Φ.

Proposition 7. *On suppose que*

- *la fonction J est propre et telle que* $\mathrm{dom} J \cap (U^{\mathrm{ad}} \times \mathbb{V}) \neq \emptyset$,
- *la fonction Φ ne prend jamais la valeur $-\infty$.*

Alors, la fonction Φ est propre.

Démonstration. Par la première hypothèse, il existe $(u_0, v_0) \in U^{\mathrm{ad}} \times \mathbb{V}$ tel que $J(u_0, v_0) < +\infty$. On en déduit que $\Phi(v_0) < +\infty$, et donc Φ est propre par la seconde hypothèse. $\square$

Remarque 3. La seconde hypothèse de cette proposition sera certainement vérifiée si, pour tout $v \in \mathrm{dom}\Phi$, l'inf de la fonction $u \mapsto J(u, v)$ sur U^{ad} est atteint en un point $\widehat{u}$. $\diamond$

Proposition 8. *On suppose que la fonction J est conjointement convexe en (u, v) et que l'ensemble U^{ad} est convexe. Alors, la fonction Φ est convexe.*

Démonstration. Pour tout $(u_1, u_2) \in U^{\mathrm{ad}} \times U^{\mathrm{ad}}$ et pour tout $\alpha \in [0, 1]$, on a par convexité de U^{ad} que $\alpha u_1 + (1 - \alpha)u_2 \in U^{\mathrm{ad}}$. Par définition de Φ, on a :

$$\begin{aligned}\Phi\big(\alpha v_1 + (1-\alpha)v_2\big) &\leq J\big(\alpha u_1 + (1-\alpha)u_2, \alpha v_1 + (1-\alpha)v_2\big) \ , \\ &\leq \alpha J(u_1, v_1) + (1-\alpha)J(u_2, v_2) \ ,\end{aligned}$$

par convexité de J. Comme cette inégalité est vraie pour tout $(u_1, u_2) \in U^{\mathrm{ad}} \times U^{\mathrm{ad}}$, on en déduit :

$$\Phi\big(\alpha v_1 + (1-\alpha)v_2\big) \leq \alpha\Phi(v_1) + (1-\alpha)\Phi(v_2) \ ,$$

d'où le résultat. $\square$

Proposition 9. *On suppose que la fonction J est convexe, que l'ensemble U^{ad} est convexe et qu'il existe $(u_0, v_0) \in U^{\mathrm{ad}} \times \mathbb{V}$ tel que l'application $v \mapsto J(u_0, v)$ soit bornée supérieurement par une constante M dans un voisinage $\mathcal{V}(v_0)$ de v_0. Alors, la fonction Φ est continue et sous-différentiable sur l'intérieur de son domaine.*

Démonstration. Par la proposition 8, on a que Φ est convexe. Pour tout $v \in \mathcal{V}(v_0)$, on a $\Phi(v) \leq J(u_0, v) \leq M$. Une application directe du théorème 1 fournit la conclusion. $\square$

Remarque 4. Cette proposition ne dit rien quant à la sous-différentiabilité de la fonction Φ au bord de son domaine. $\diamond$

Pour énoncer le résultat principal de cette section, on introduit la *fonction caractéristique* de l'ensemble $U^{\mathrm{ad}} \times \mathbb{V}$:

$$\chi_{U^{\mathrm{ad}}\times\mathbb{V}}(u, v) = \begin{cases} 0 & \text{si } u \in U^{\mathrm{ad}} \text{ et } v \in \mathbb{V} \\ +\infty & \text{sinon} \end{cases} \ ,$$

et on introduit la fonction $\widetilde{J} = J + \chi_{U^{\mathrm{ad}}\times\mathbb{V}}$. De manière évidente, le problème (1.9) se met sous la forme équivalente :

$$\Phi(v) = \inf_{u\in\mathbb{U}} \widetilde{J}(u, v) \ .$$

On fait les hypothèses suivantes.

Hypothèse 1.

1. U^{ad} est une partie convexe fermé non vide de $\mathbb{U}$, et $\mathrm{dom}J \cap (U^{\mathrm{ad}} \times \mathbb{V}) \neq \emptyset$.
2. J est convexe, s.c.i., propre, sous-différentiable sur un ouvert contenant $U^{\mathrm{ad}} \times \mathbb{V}$.
3. Pour tout $v \in \mathbb{V}$, l'application $u \mapsto J(u, v)$ est coercive sur U^{ad}.
4. Il existe $(u_0, v_0) \in U^{\mathrm{ad}} \times \mathbb{V}$ tel que $v \mapsto J(u_0, v)$ soit bornée sur un voisinage de v_0.
5. Il existe $(u_0', v_0') \in U^{\mathrm{ad}}\times\mathbb{V}$ tel que l'une des fonctions J ou $\chi_{U^{\mathrm{ad}}\times\mathbb{V}}$ soit continue en (u_0', v_0').

Théorème 7. *Sous les hypothèses 1.1—1.4, la fonction marginale Φ est convexe, propre, et elle est continue et sous-différentiable sur l'intérieur de son domaine. Soit un point $v \in \mathrm{dom}\Phi$, et soit $\widehat{u} \in \widehat{U}(v)$. Le sous-différentiel de Φ est donné par :*

$$\partial\Phi(v) = \left\{ s \ , \ \ (0, s) \in \partial\widetilde{J}(\widehat{u}, v) \right\} . \tag{1.10}$$

Démonstration. Par l'hypothèse 1.1, le domaine de Φ est non vide. Les hypothèses 1.2 et 1.3 impliquent que, pour tout $v \in \mathrm{dom}\Phi$, l'inf dans le problème (1.9) est atteint : l'ensemble $\widehat{U}(v)$ est non vide et $\Phi(v)$ est fini. La fonction Φ ne prend jamais la valeur $-\infty$ et est donc propre. L'hypothèse 1.2 permet d'assurer que la fonction Φ est convexe, et l'hypothèse 1.4 implique que Φ est continue et sous-différentiable sur l'intérieur de son domaine (proposition 9).

Soit alors $v \in \mathrm{dom}\Phi$. On a :

$$s \in \partial\Phi(v) \quad \Longleftrightarrow \quad \Phi(v') \geq \Phi(v) + \left\langle s \, , v' - v \right\rangle \quad \forall v' \in \mathbb{V}$$

par définition du sous-différentiel,

$$\Longleftrightarrow \quad \Phi(v') \geq \widetilde{J}(\widehat{u}, v) + \left\langle s \, , v' - v \right\rangle \quad \forall v' \in \mathbb{V}$$

pout n'importe quel point $\widehat{u} \in \widehat{U}(v)$ solution du problème (1.9),

$$\Longleftrightarrow \quad \widetilde{J}(u', v') \geq \widetilde{J}(\widehat{u}, v) + \left\langle s \, , v' - v \right\rangle \quad \forall (u', v') \in \mathbb{U} \times \mathbb{V}$$

par définition de Φ,

$$\Longleftrightarrow \quad (0, s) \in \partial\widetilde{J}(\widehat{u}, v)$$

par définition du sous-différentiel, d'où le résultat. □

Remarque 5. Une conséquence de la démonstration est que l'ensemble $\left\{ s \ , \ \ (0, s) \in \partial\widetilde{J}(\widehat{u}, v) \right\}$ *ne dépend pas* du minimiseur $\widehat{u}$ choisi dans l'ensemble des solutions $\widehat{U}(v)$ car tous les éléments de $\widehat{U}(v)$ conduisent au même ensemble. En particulier, le fait que $\partial\Phi(v)$ soit réduit à un singleton (et donc que Φ soit différentiable) n'est pas lié à l'unicité de la solution du problème (1.9). ◇

Remarque 6. L'hypothèse 1.4 impliquant que Φ est sous-différentiable sur l'intérieur de son domaine est très forte. En fait, il suffirait de montrer que la fonction Φ est s.c.i. et d'utiliser le corollaire 1 pour arriver aux conclusions du théorème 7. Remplaçant l'hypothèse 1.4 par :

1.4b. Pour tout $\overline{v} \in \mathbb{V}$, il existe un voisinage $\mathcal{V} \subset \mathbb{V}$ de $\overline{v}$ et il existe un sous-ensemble fermé borné $\mathcal{B} \subset \mathbb{U}$ tels que
$$\forall v \in \mathcal{V} \ , \ \ \forall u \in \widehat{U}(v) \ , \ \ u \in \mathcal{B} \ ,$$

la proposition suivante fournit le caractère s.c.i. de la fonction Φ.

Proposition 10. *On fait les hypothèses 1.1—1.3, et on suppose que l'hypothèse 1.4b est satisfaite. Alors, la fonction Φ est semi-continue inférieurement.*

Démonstration. On rappelle ([Gilbert (2018), Annexe A, Proposition A.1]) que la fonction Φ est s.c.i. si tous ses ensembles de sous-niveau $\mathrm{lev}_\alpha\Phi = \left\{ v \in \mathbb{V} \ , \ \ \Phi(v) \leq \alpha \right\}$ sont fermés. Soit $\{v_k\}_{k\in\mathbb{N}}$ une suite d'éléments de l'ensemble de sous-niveau $\mathrm{lev}_\alpha\Phi$, telle que $v_k \to \overline{v}$. On peut alors construire une suite $\{u_k\}_{k\in\mathbb{N}}$, avec $u_k \in \widehat{U}(v_k)$. Par l'hypothèse 1.4b, cette suite est contenue dans un sous-ensemble fermé borné de $\mathbb{U}$ et on peut donc en extraire une sous-suite $\{u_{\varphi(k)}\}_{k\in\mathbb{N}}$ qui converge vers un point $\overline{u} \in \mathbb{U}$. Comme la suite $\{(u_{\varphi(k)}, v_{\varphi(k)})\}_{k\in\mathbb{N}}$ converge vers $(\overline{u}, \overline{v})$, on obtient par semi-continuité inférieure de J que $(\overline{u}, \overline{v}) \in \mathrm{lev}_\alpha J$ puisque $J(u_{\varphi(k)}, v_{\varphi(k)}) = \Phi(v_{\varphi(k)}) \leq \alpha$ pour tout k. Par définition de Φ, on a que $\Phi(\overline{v}) \leq J(\overline{u}, \overline{v}) \leq \alpha$, d'où l'on conclut que $\overline{v} \in \mathrm{lev}_\alpha \Phi$. On a ainsi montré que tous les ensembles $\mathrm{lev}_\alpha \Phi$ sont fermés, et donc que la fonction Φ est s.c.i. □

En conclusion, on notera que l'hypothèse 1.4b est vérifiée si l'ensemble U^{ad} est borné. ◇

Le corollaire suivant permet de mieux appréhender le sous-différentiel de la fonction Φ par rapport au sous-différentiel de la fonction $v \mapsto J(u, v)$, c'est-à-dire le sous-différentiel partiel en v de la fonction J, que l'on note $\partial_v J(u, v)$.

Corollaire 2. *On se place sous les mêmes hypothèses que dans le théorème 7 et l'on suppose de plus que l'hypothèse 1.5 est satisfaite. Soit un point $v \in \mathrm{dom}\Phi$. Alors :*

$$\partial\Phi(v) \subset \bigcap_{\widehat{u}\in\widehat{U}(v)} \partial_v J(\widehat{u}, v) . \tag{1.11}$$

Démonstration. Sous l'hypothèse supplémentaire 1.5, le théorème 4 s'applique et on a donc l'égalité : $\partial\widetilde{J}(u, v) = \partial J(u, v) + \partial\chi_{U^{\mathrm{ad}}}(u) \times \{0\}$.[4] Soit alors $s \in \partial\Phi(v)$. Par le théorème 7, $(0, s)$ est un sous-gradient de $\widetilde{J}$ et est donc de la forme $(p, s) + (r, 0)$, avec $p + r = 0$ et avec $s \in \mathrm{proj}_{\mathbb{V}}\left(\partial J(\widehat{u}, v)\right) \subset \partial_v J(\widehat{u}, v)$ (voir la remarque 8). On a montré que :

$$s \in \partial\Phi(v) \quad \Longrightarrow \quad s \in \partial_v J(\widehat{u}, v) \quad \forall \widehat{u} \in \widehat{U}(v) ,$$

d'où le résultat. □

Dans le cas différentiable, on dispose d'un résultat plus précis.

Corollaire 3 (Cas différentiable)**.** *On suppose de plus que la fonction J est différentiable. Alors la fonction Φ est elle aussi différentiable et l'on a :*

$$\nabla\Phi(v) = \nabla_v J(\widehat{u}, v) \quad \forall \widehat{u} \in \widehat{U}(v) .$$

Démonstration. Dans le cas J différentiable, une solution $\widehat{u}$ du problème (1.9) est caractérisée par :

$$\nabla_u J(\widehat{u}, v) \in -\partial\chi_{U^{\mathrm{ad}}}(\widehat{u}) ,$$

ce qui implique que $\big(0, \nabla_v J(\widehat{u}, v)\big) \in \partial\widetilde{J}(\widehat{u}, v)$ et donc que Φ est sous-différentiable en $v \in \mathrm{dom}\Phi$. Par unicité de la seconde composante de $\partial\widetilde{J}(\widehat{u}, v)$ on en déduit que Φ est différentiable. □

Remarque 7. Le corollaire 3 reste vrai si l'on suppose que seule l'application partielle $v \mapsto J(u, v)$ est différentiable, car la différentiabilité de Φ repose sur l'unicité de la seconde composante dans le sous-différentiel J. ◇

Remarque 8. Soit F une fonction définie sur le produit d'espace $\mathbb{U} \times \mathbb{V}$. On suppose que F est propre, sous-différentiable en un point $(u, v) \in \mathrm{dom}F$. Il est instructif de comparer le sous-différentiel partiel $\partial_u F(u, v)$ avec la projection du sous-différentiel $\partial F(u, v)$ sur l'espace $\mathbb{U}$. Par définition du sous-différentiel et du sous-différentiel partiel, on a toujours :

$$\mathrm{proj}_{\mathbb{U}}\left(\partial F(u, v)\right) \subset \partial_u F(u, v) .$$

4. Il est aisé de montrer que, pour tout $v \in \mathbb{V}$, on a : $\partial\chi_{U^{\mathrm{ad}}\times\mathbb{V}}(u, v) = \partial\chi_{U^{\mathrm{ad}}}(u) \times \{0\}$.

L'inclusion réciproque est vraie sous certaines conditions. Soit en effet un point $(u, v) \in \mathbb{U} \times \mathbb{V}$ et soit $r \in \partial_u F(u, v)$. On pose :

$$\Psi(v) = \inf_{u' \in \mathbb{U}} F(u', v) - \langle r , u' \rangle \ .$$

On montre par application directe de la définition que la conjuguée de Fenchel de Ψ vérifie :

$$\Psi^\star(s) = F^\star(r, s) \ .$$

Par définition de Ψ et comme $r \in \partial_u F(u, v)$, on a :

$$\Psi(v) = F(u, v) - \langle r , u \rangle \ .$$

Supposons que la fonction marginale Ψ soit propre et qu'elle soit sous-différentiable en v. Soit alors $s \in \partial\Psi(v)$. Par le théorème 3, on a :

$$\Psi(v) + \Psi^\star(s) = \langle s , v \rangle \ .$$

On en déduit alors que :

$$F^\star(r, s) = \Psi^\star(s) = \langle s , v \rangle - \Psi(v) = \langle r , u \rangle + \langle s , v \rangle - F(u, v) \ ,$$

ce qui prouve que $(r, s) \in \partial F(u, v)$, et donc que :

$$\partial_u F(u, v) \subset \mathrm{proj}_{\mathbb{U}} \left(\partial F(u, v) \right) \ ,$$

d'où l'on déduit l'égalité des ensembles $\partial_u F(u, v)$ et $\mathrm{proj}_{\mathbb{U}} \left(\partial F(u, v) \right)$.

Cette égalité n'est pas toujours vraie. Dans [Bauschke & Combettes (2011), Remark 16.7], on trouve l'exemple suivant. Soit χ_B la fonction caractéristique de la boule unité de $\mathbb{R}^2$. Le sous-différentiel de cette fonction au point $(1, 0)$ est $\mathbb{R}_+ \times \{0\}$, d'où $\mathrm{proj}_{\mathbb{V}} \left(\partial\chi_B(1, 0) \right) = \{0\}$. Pourtant, l'application composante $v \mapsto \chi_B(1, v)$ est égale à la fonction caractéristique de l'ensemble réduit au point 0 dont le sous-différentiel en 0 est l'espace $\mathbb{R}$: $\partial_v \chi_B(1, 0) = \mathbb{R}$. $\diamond$

1.6.2 Cas d'une fonction convexe-concave

Soit L une fonction définie sur le produit $\mathbb{U} \times \mathbb{V}$ de deux espaces de Hilbert de dimension finie, à valeurs dans $\mathbb{R}$, et soit U^{ad} un sous-ensemble de l'espace $\mathbb{U}$. On définit alors la *fonction marginale* :

$$H(\lambda) = \inf_{u \in U^{\mathrm{ad}}} L(u, \lambda) \ , \tag{1.12}$$

et l'ensemble des solutions associées (éventuellement vide) :

$$\widehat{U}(\lambda) = \arg\min_{u \in U^{\mathrm{ad}}} L(u, \lambda) \ .$$

On fait les hypothèses suivantes.

Hypothèse 2.

1. Le sous-ensemble U^{ad} est convexe fermé.
2. Pour tout $\lambda \in \mathbb{V}$, l'application $u \mapsto L(u, \lambda)$ est propre, convexe, s.c.i., coercive sur U^{ad}.

3. Pour tout $u \in \mathbb{U}$, l'application $\lambda \mapsto L(u,\lambda)$ est concave, continue, sur-différentiable.[5]

Les hypothèses 2.1 et 2.2 assurent que l'infimum est atteint dans le problème (1.12) et donc que l'ensemble des solutions $\widehat{U}(\lambda)$ est non vide.

Théorème 8. *Soit L une application vérifiant l'hypothèse 2 et soit λ un point de $\mathbb{V}$ en lequel la fonction H est sur-différentiable. Alors :*

$$\overline{\partial} H(\lambda) = \mathrm{cl}\Big(\mathrm{conv}\Big(\bigcup_{\widehat{u}\in\widehat{U}(\lambda)} \overline{\partial}_\lambda L(\widehat{u},\lambda)\Big)\Big) . \tag{1.13}$$

Démonstration. Par l'hypothèse 2, on a que l'ensemble $\widehat{U}(\lambda)$ est non vide pour tout $\lambda \in \mathbb{V}$. La fonction H, en tant qu'enveloppe inférieure de fonctions concaves semi-continues supérieurement (s.c.s.), est concave s.c.s, sur-différentiable sur l'intérieur de son domaine.

On va prouver que les dérivées directionnelles en λ suivant une direction quelconque δ des fonctions L et H vérifient :

$$\mathrm{D}H(\lambda)(\delta) = \inf_{u\in\widehat{U}(\lambda)} \mathrm{D}_\lambda L(u,\lambda)(\delta) . \tag{1.14}$$

Soit $u \in \widehat{U}(\lambda)$. Par définition de H et comme u est solution de (1.12), on a :

$$\frac{H(\lambda+h\delta)-H(\lambda)}{h} \leq \frac{L(u,\lambda+h\delta)-L(u,\lambda)}{h} ,$$

et donc :

$$\mathrm{D}H(\lambda)(\delta) \leq \inf_{u\in\widehat{U}(\lambda)} \mathrm{D}_\lambda L(u,\lambda)(\delta) .$$

Montrons maintenant que $\exists\, \widehat{u} \in \widehat{U}(\lambda)$ tel que $\mathrm{D}H(\lambda)(\delta) \geq \mathrm{D}_\lambda L(\widehat{u},\lambda)(\delta)$, ce qui impliquera la relation (1.14). On définit l'ensemble :

$$B_h = \Big\{ u \in U^{\mathrm{ad}} \ , \ \frac{L(u,\lambda+h\delta)-H(\lambda)}{h} \leq \mathrm{D}H(\lambda)(\delta) \Big\} .$$

— L'ensemble B_h est fermé car L est s.c.i. en u.
— L'ensemble B_h est borné car L est coercive en u.
— L'ensemble B_h est non vide car il contient $\widehat{U}(\lambda+h\delta)$. En effet, pour $\widehat{u} \in \widehat{U}(\lambda+h\delta)$,

$$\frac{L(\widehat{u},\lambda+h\delta)-H(\lambda)}{h} = \frac{H(\lambda+h\delta)-H(\lambda)}{h} \leq \mathrm{D}H(\lambda)(\delta) ,$$

la dernière inégalité étant une conséquence de la remarque 2, d'où le résultat.
— Si $h_1 \leq h_2$, alors $B_{h_1} \subset B_{h_2}$. En effet, par concavité de la fonction L par rapport à λ,

$$L(u,\lambda+h_1\delta)-H(\lambda) \geq \frac{h_1}{h_2}\big(L(u,\lambda+h_2\delta)-H(\lambda)\big) + \Big(1-\frac{h_1}{h_2}\Big)\big(L(u,\lambda)-H(\lambda)\big) ,$$

et le résultat se déduit de la définition de B_{h_1} et de ce que $L(u,\lambda)-H(\lambda) \geq 0$.

5. Voir la remarque 1.

Soit $h_0 > 0$. D'après les propriétés précédentes, les ensembles B_h pour $h \leq h_0$ sont fermés bornés et donc compacts. Posons

$$\widehat{B} = \bigcap_{0<h\leq h_0} B_h \ .$$

Si cette intersection est vide, on peut trouver un nombre fini de compacts B_h d'intersection vide, ce qui contredit la propriété d'inclusion des B_h et le fait qu'ils soient tous non vides. On en déduit donc que $\widehat{B} \neq \emptyset$. Soit alors $\widehat{u} \in \widehat{B}$. Pour tout $h > 0$, on a par définition de B_h :

$$L(\widehat{u}, \lambda + h\delta) - H(\lambda) \leq h\mathrm{D}H(\lambda)(\delta) \ . \tag{1.15}$$

Passant à la limite en h et utilisant la continuité de L en λ, on obtient que $L(\widehat{u}, \lambda) \leq H(\lambda)$. On en déduit donc que $\widehat{u} \in \widehat{U}(\lambda)$. De l'inégalité (1.15), divisant par $h > 0$, utilisant le fait que $H(\lambda) = L(\widehat{u}, \lambda)$ et faisant tendre h vers 0, on obtient :

$$\mathrm{D}_\lambda L(\widehat{u}, \lambda)(\delta) \leq \mathrm{D}H(\lambda)(\delta) \ ,$$

ce qui est l'inégalité recherché. On a donc montré la relation (1.14).

Comme l'application $\lambda \mapsto L(u, \lambda)$ est continue, on déduit de la proposition 2 que l'application $\delta \mapsto \mathrm{D}_\lambda L(u, \lambda)(\delta)$ est continue en tout point u, et donc que l'application $\delta \mapsto \mathrm{D}H(\lambda)(\delta)$ est s.c.s. comme enveloppe inférieure de fonctions continues d'après (1.14). Soit λ un point en lequel H est sur-différentiable. Appliquant la proposition 4 (transposée au cas sur-différentiable) à la fonction H, utilisant la relation (1.14), et appliquant de nouveau la proposition 4 à la fonction $L(u, \cdot)$, on obtient :

$$\begin{aligned}
\sigma_{\overline{\partial} H(\lambda)}(\delta) &= -\mathrm{D}H(\lambda)(\delta) \\
&= \sup_{u \in \widehat{U}(\lambda)} -\mathrm{D}_\lambda L(u, \lambda)(\delta) \\
&= \sup_{\widehat{u} \in \widehat{U}(\lambda)} \sigma_{\overline{\partial}_\lambda L(\widehat{u}, \lambda)}(\delta) \ ,
\end{aligned}$$

et donc, par définition de la fonction d'appui,

$$\begin{aligned}
&= \sup_{\widehat{u} \in \widehat{U}(\lambda)} \ \sup_{r \in \overline{\partial}_\lambda L(\widehat{u}, \lambda)} \langle r \, , \delta \rangle \\
&= \sup_{r \in \bigcup_{\widehat{u} \in \widehat{U}(\lambda)} \overline{\partial}_\lambda L(\widehat{u}, \lambda)} \langle r \, , \delta \rangle \\
&= \sigma_{\bigcup_{\widehat{u} \in \widehat{U}(\lambda)} \overline{\partial}_\lambda L(\widehat{u}, \lambda)}(\delta) \ .
\end{aligned}$$

Les deux ensembles $\overline{\partial} H(\lambda)$ et $\bigcup_{\widehat{u} \in \widehat{U}(\lambda)} \overline{\partial}_\lambda L(\widehat{u}, \lambda)$ ont ainsi la même fonction d'appui, et le premier d'entre eux est convexe fermé. On en déduit par la proposition 3 :

$$\overline{\partial} H(\lambda) = \mathrm{cl}\Big(\mathrm{conv}\Big(\bigcup_{\widehat{u} \in \widehat{U}(\lambda)} \overline{\partial}_\lambda L(\widehat{u}, \lambda)\Big)\Big) \ ,$$

ce qui achève la démonstration. □

Application au cas différentiable. Si la fonction $\lambda \mapsto L(u, \lambda)$ est *différentiable*, et si le problème (1.12) a une *unique* solution, c'est-à dire si $\widehat{U}(\lambda) = \{\widehat{u}(\lambda)\}$, alors la fonction H est différentiable et l'on a :

$$\nabla H(\lambda) = \nabla_\lambda L(\widehat{u}(\lambda), \lambda) \ .$$

Remarque 9. On obtient des résultats analogues en considérant la fonction marginale obtenue en *maximisant* par rapport à λ. On définit :

$$\Psi(u) = \sup_{\lambda \in \Lambda^{\mathrm{ad}}} L(u,\lambda) \ ,$$
$$\widehat{\Lambda}(u) = \arg\max_{\lambda \in \Lambda^{\mathrm{ad}}} L(u,\lambda) \ .$$

En supposant que l'application $u \mapsto L(u,\lambda)$ est convexe, continue, sous-différentiable, et que l'application $\lambda \mapsto L(u,\lambda)$ est propre, concave, s.c.s., coercive sur Λ^{ad}, on obtient

$$\partial\Psi(u) = \mathrm{cl}\Bigg(\mathrm{conv}\Big(\bigcup_{\widehat{\lambda} \in \widehat{\Lambda}(u)} \partial_u L(u,\widehat{\lambda})\Big)\Bigg) \ ,$$

ce qui est une généralisation du résultat concernant l'enveloppe supérieure d'une famille de fonctions vu au §1.6.3. ◇

Remarque 10. Une étude très détaillée de la sous-différentiabilité des fonctions marginales dans les cas convexe-convexe, convexe-concave, convexe-concave-convexe, et même convexe-concave-convexe-concave pourra être trouvée dans [Mataoui (1990), Chapitre 2]. [6]

À titre d'illustration, considérons le cas de la fonction de perturbation :

$$\min_{u \in U^{\mathrm{ad}}} J(u) \quad \text{sous} \quad \Theta(u) - v \in -C \ , \tag{1.16}$$

On pose

$$L(u,v,\lambda) = J(u) + \big\langle \lambda \,, \Theta(u) - v \big\rangle \quad \text{avec } \lambda \in C^\star \ .$$

On définit d'abord :

$$G(v,\lambda) = \inf_{u \in U^{\mathrm{ad}}} L(u,v,\lambda) \quad , \quad \widehat{U}(v,\lambda) = \arg\min_{u \in U^{\mathrm{ad}}} L(u,v,\lambda) \ .$$

On suppose alors que $(u,v) \mapsto L(u,v,\lambda)$ vérifie les conditions du théorème 7 et du corollaire 3 pour tout $\lambda \in C^\star$. Alors, la fonction G est convexe s.c.i. propre en v, concave s.c.s en λ, et

$$\nabla_v G(v,\lambda) = \nabla_v L(\widehat{u},v,\lambda) \quad \forall \widehat{u} \in \widehat{U}(v,\lambda) \ ,$$

et l'on déduit de la forme spécifique de L que l'on a : $\nabla_v G(v,\lambda) = -\lambda$.

On définit ensuite :

$$\Phi(v) = \sup_{\lambda \in C^\star} G(v,\lambda) \quad , \quad \widehat{\Lambda}(v) = \arg\max_{\lambda \in C^\star} G(v,\lambda) \ .$$

On notera que les conditions d'existence d'un point-selle implique que $\Phi(v)$ est la valeur optimale du problème (1.16). Sous les hypothèses du théorème 8 et de la remarque 9, on a

$$\begin{aligned}\partial\Phi(v) &= \mathrm{cl}\Bigg(\mathrm{conv}\Big(\bigcup_{\widehat{\lambda} \in \widehat{\Lambda}(v)} \nabla_v G(v,\widehat{\lambda})\Big)\Bigg) \ ,\\ &= -\widehat{\Lambda}(v) \ .\end{aligned}$$

Appliquée en $v = 0$, cette dernière relation exprime la fameuse « interprétation marginaliste des multiplicateurs » d'un problème d'optimisation sous contraintes. ◇

6. Dans cette étude, certains détails doivent cependant être corrigés. . .

1.6.3 Cas d'une enveloppe supérieure

Soit I un *ensemble quelconque* d'indices et soit $\{J_i\}_{i\in I}$ une famille de fonctions toutes définies sur $\mathbb{U}$ à valeurs dans $]-\infty, +\infty]$. On définit la fonction J comme l'enveloppe supérieure de cette famille de fonctions :

$$J(u) = \sup_{i\in I} J_i(u) \ ,$$

et on note $I(u)$ le sous-ensemble (éventuellement vide) d'indices où le sup est atteint :

$$I(u) = \big\{ i \in I \ , \ \ J(u) = J_i(u) \big\} \ .$$

On montre aisément que, si les fonctions J_i sont convexes (resp. s.c.i.), alors la fonction J est elle aussi convexe (resp. s.c.i.). En effet, il est immédiat de voir que la relation suivante sur les épigraphes est vérifiée :

$$\mathrm{epi}\Big(\sup_{i\in I} J_i\Big) = \bigcap_{i\in I} \mathrm{epi} J_i \ ,$$

d'où ces conclusions. Par contre, le fait que les fonctions J_i soient propres n'implique pas que la fonction J le soit (considérer par exemple l'enveloppe supérieure des fonctions constantes). Le théorème suivant donne des informations sur le sous-différentiel de la fonction marginale J.

Théorème 9. *On suppose que les fonctions J_i sont propres convexes et que la fonction J est elle aussi propre. Alors,*

$$\partial J(u) \supset \mathrm{cl}\bigg(\mathrm{conv}\Big(\bigcup_{i\in I(u)} \partial J_i(u)\Big)\bigg) \ .$$

Démonstration. La convexité des fonctions J_i entraîne celle de la fonction J. Soit $u \in \mathrm{dom}J$, soit $i \in I(u)$ et soit $r_i \in \partial J_i(u)$. Pour tout $v \in \mathbb{U}$, on a :

$$\begin{aligned} J(v) &\geq J_i(v) \\ &\geq J_i(u) + \langle r_i \, , v-u \rangle \\ &= J(u) + \langle r_i \, , v-u \rangle \ , \end{aligned}$$

la dernière égalité venant de la définition de $I(u)$. On en déduit que $r_i \in \partial J(u)$. Le sous-différentiel étant un ensemble convexe fermé, on déduit le résultat. □

Remarque 11. Ce théorème ne donne pas d'information sur le sous-différentiel de la fonction J lorsque l'ensemble $I(u)$ est vide, ou encore lorsque $\partial J_i(u) = \emptyset$ pour tout $i \in I(u)$. ◇

Obtenir l'égalité plutôt que l'inclusion dans le théorème ci-dessus est plus difficile et nécessite des hypothèses additionnelles. On a par exemple le résultat suivant.

Théorème 10. *On suppose que les fonctions J_i sont propres convexes et que la fonction J est elle aussi propre. On suppose de plus que l'ensemble d'indices I est une partie compacte d'un espace métrique et que les applications $i \mapsto J_i(u)$ sont s.c.s. pour tout $u \in \mathrm{dom}J$. Alors, pour tout $u \in \mathrm{int}\big(\mathrm{dom}J\big)$, on a :*

$$\partial J(u) = \mathrm{cl}\bigg(\mathrm{conv}\Big(\bigcup_{i\in I(u)} \partial J_i(u)\Big)\bigg) \ .$$

Pour la preuve, voir [Hiriart-Urruty and Lemaréchal (2001), Theorem 4.4.2]. Ce théorème peut aussi être vu comme un cas particulier du théorème 8 et de la remarque 9, que l'on a vu au §1.6.2.

Le corollaire suivant est une conséquence directe de ce théorème dans le cas différentiable.

Corollaire 4. *On suppose en plus des hypothèses du théorème 10 que les fonctions J_i sont différentiables. Alors,*

$$\partial J(u) = \mathrm{cl}\Big(\mathrm{conv}\big\{\nabla J_i(u)\ ,\ \ i \in I(u)\big\}\Big)\ .$$

Ce résultat montre que la fonction J est différentiable en u pourvu que l'ensemble $I(u)$ soit réduit à un indice unique.

1.7 Conditions d'optimalité

On considère le problème d'optimisation suivant :

$$\min_{u \in U^{\mathrm{ad}}} F(u)\ , \tag{1.17}$$

où U^{ad} est une partie convexe fermée non vide d'un espace de Hilbert de dimension finie $\mathbb{U} = \mathbb{R}^n$, et où $F : \mathbb{U} \to\ ]-\infty, +\infty]$ est une fonction propre convexe s.c.i. coercive sur U^{ad} et sous-différentiable. On suppose que le problème n'est pas trivial, à savoir que $\mathrm{dom} F \cap U^{\mathrm{ad}} \neq \emptyset$. Ces hypothèses permettent de garantir l'existence d'une solution $u^\sharp$ du problème d'optimisation (1.17). Notant $\chi_{U^{\mathrm{ad}}}$ la fonction caractéristique de l'ensemble U^{ad}, le problème (1.17) se met sous la forme équivalente :

$$\min_{u \in \mathbb{U}} F(u) + \chi_{U^{\mathrm{ad}}}(u)\ .$$

Par la relation (1.2), la condition d'optimalité associée s'écrit :

$$0 \in \partial\big(F + \chi_{U^{\mathrm{ad}}}\big)(u^\sharp)\ .$$

Sous l'hypothèse qu'il existe un point de U^{ad} en lequel F est continue, et utilisant le théorème 4, cette condition d'optimalité prend la forme :

$$0 \in \partial F(u^\sharp) + \partial \chi_{U^{\mathrm{ad}}}(u^\sharp)\ ,$$

ce qui s'écrit de manière équivalente :

$$\exists r \in \partial F(u^\sharp)\ ,\ \ \exists s \in \partial \chi_{U^{\mathrm{ad}}}(u^\sharp)\ ,\ \ r + s = 0\ .$$

De plus,

$$\begin{aligned} s \in \partial \chi_{U^{\mathrm{ad}}}(u^\sharp) \quad &\Longleftrightarrow \quad \chi_{U^{\mathrm{ad}}}(u) \geq \chi_{U^{\mathrm{ad}}}(u^\sharp) + \big\langle s\, , u - u^\sharp \big\rangle \qquad \forall u \in \mathbb{U}\ , \\ &\Longleftrightarrow \quad \big\langle s\, , u - u^\sharp \big\rangle \leq 0 \qquad \forall u \in U^{\mathrm{ad}}\ . \end{aligned}$$

La dernière équivalence permet d'identifier le sous-différentiel de la fonction $\chi_{U^{\mathrm{ad}}}$ en $u \in U^{\mathrm{ad}}$ et le *cône normal* $N_u U^{\mathrm{ad}}$ de l'ensemble U^{ad} au point u :[7]

$$\partial \chi_{U^{\mathrm{ad}}}(u) = N_u U^{\mathrm{ad}}\ . \tag{1.18}$$

On en déduit le théorème suivant.

Théorème 11. *Sous les hypothèses mentionnées ci-dessus, une condition nécessaire et suffisante pour que $u^\sharp \in U^{\mathrm{ad}}$ soit solution du problème (1.17) est :*

$$\exists r \in \partial F(u^\sharp)\ ,\ \ \big\langle r\, , u - u^\sharp \big\rangle \geq 0 \quad \forall u \in U^{\mathrm{ad}}\ . \tag{1.19}$$

7. Par définition, le cône normal de U^{ad} au point u_0 est : $N_{u_0} U^{\mathrm{ad}} = \big\{ r \in \mathbb{U}\ ,\ \ \langle r\, , u - u_0 \rangle \leq 0 \quad \forall u \in U^{\mathrm{ad}} \big\}$.

1.8 Conditions de Karush-Kuhn-Tucker

On considère le problème d'optimisation suivant :

$$\min_{u \in \mathbb{U}} F(u) \ , \tag{1.20a}$$

sous la contrainte :

$$\Theta(u) \in -C \subset \mathbb{V} \ . \tag{1.20b}$$

On suppose que la fonction $F : \mathbb{U} \to]-\infty, +\infty]$ est propre, convexe, s.c.i., coercive et sous-différentiable. L'espace $\mathbb{V}$ est l'espace $\mathbb{R}^m$, et on suppose que C est un cône convexe fermé saillant de $\mathbb{V}$. On rappelle qu'un cône C est un sous-ensemble de $\mathbb{V}$ tel que si $v \in C$, alors αv appartient aussi à C pour tout $\alpha \geq 0$, et qu'il est dit saillant si $C \cap -C = \{0\}$. On note $C^\star$ le cône dual de C :

$$C^\star = \big\{ \lambda \in \mathbb{V}, \ \langle \lambda \, , v \rangle \geq 0 \ \ \forall v \in \mathbb{V} \big\} \ .$$

On suppose que la fonction Θ, définie sur l'espace $\mathbb{U}$ à valeurs dans l'espace $\mathbb{V}$, est :

- continue,
- C-convexe : $\Theta\big(\alpha u_1 + (1-\alpha) u_2\big) - \alpha \Theta\big(u_1\big) - (1-\alpha)\Theta\big(u_2\big) \in -C \quad \forall \alpha \in [0,1]$,
- C-sous-différentiable : $\partial\Theta(u) = \big\{ \theta \in \mathcal{L}(\mathbb{U}, \mathbb{V}) \ , \ \ \Theta(w) - \Theta(u) - \theta \cdot (w-u) \in C \quad \forall w \in \mathbb{U} \big\}$.

Remarque 12. On vérifie aisément que la propriété de C-convexité de Θ est équivalente à la propriété de convexité de l'application $\xi_\lambda : u \mapsto \big\langle \lambda \, , \Theta(u) \big\rangle$ pour tout $\lambda \in C^\star$. Pour ce qui est de la sous-différentiabilité, on montre facilement l'inclusion suivante pour tout $\lambda \in C^\star$:

$$\partial\xi_\lambda(u) \supset \big\{ r \in \mathbb{U}, \ r = \theta^\top \cdot \lambda, \ \theta \in \partial\Theta(u) \big\} \ .$$

Mais l'inclusion réciproque n'est pas garantie ! Lorsque les ensembles dans la relation précédente sont égaux, on dit que l'application Θ est *régulièrement* sous-différentiable. ◇

Les hypothèses sur Θ font que l'ensemble $U^\Theta = \big\{ u \in \mathbb{U} \ , \ \ \Theta(u) \in -C \big\}$ est un convexe fermé de $\mathbb{U}$, et on peut alors prouver, grâce aux hypothèses faites sur F, l'existence d'une solution $u^\sharp$ du problème d'optimisation (1.20).

Pour obtenir les conditions d'optimalité du problème (1.20), on va introduire une hypothèse de *qualification des contraintes*. Pour cela, on introduit le cône suivant défini en un point $u_0 \in \mathbb{U}$:

$$C^{\mathrm{L}}_{u_0} = \big\{ r \in \mathbb{U}, \ r = \theta^\top \cdot \lambda, \ \theta \in \partial\Theta(u_0), \ \lambda \in C^\star \ \text{ et } \ \big\langle \lambda \, , \Theta(u_0) \big\rangle = 0 \big\} \ .$$

Proposition 11. *En tout point $u_0 \in \mathbb{U}$, on a l'inclusion :*

$$C^{\mathrm{L}}_{u_0} \subset N_{u_0} U^\Theta \ .$$

Démonstration. Soit $r \in C^{\mathrm{L}}_{u_0}$: il existe $\theta \in \partial\Theta(u_0)$ et il existe $\lambda \in C^\star$ vérifiant $\big\langle \lambda \, , \Theta(u_0) \big\rangle = 0$, tels que $r = \theta^\top \cdot \lambda$. On a donc, pour tout $u \in U^\Theta$:

$$\begin{aligned} \langle r \, , u - u_0 \rangle &= \big\langle \lambda \, , \theta \cdot (u - u_0) \big\rangle \\ &\leq \big\langle \lambda \, , \Theta(u) - \Theta(u_0) \big\rangle \\ &\leq \big\langle \lambda \, , \Theta(u) \big\rangle \\ &\leq 0 \ . \end{aligned}$$

On en déduit donc que $r \in N_{u_0} U^\Theta$. □

On donne alors la définition suivante, dite *hypothèse de qualification des contraintes*.

Définition 6. On dit que les contraintes sont qualifiées en un point $u_0 \in \mathbb{U}$ si l'on a :

$$C^{\mathrm{L}}_{u_0} = N_{u_0} U^{\Theta} \ .$$

Soit $u^\sharp$ une solution du problème (1.20). On a vu au §1.7 qu'il existait alors un $r \in \partial F(u^\sharp)$ tel que

$$r \in -N_{u^\sharp} U^{\Theta} \ .$$

Supposant les contraintes qualifiées en $u^\sharp$, cette condition est équivalente à $r \in -C^{\mathrm{L}}_{u^\sharp}$. Par définition du cône $C^{\mathrm{L}}_{u^\sharp}$, on en déduit le théorème suivant, qui énonce les conditions de Karush, Kuhn et Tucker (KKT) dans le cas sous-différentiable.

Théorème 12. *Sous les hypothèses précédentes, le fait que $u^\sharp$ soit solution du problème* (1.20) *est caractérisé par les conditions suivantes :*

$$\exists\, \lambda \in \mathbb{V} \ , \ \exists\, r \in \partial F(u^\sharp) \ , \ \exists\, \theta \in \partial\Theta(u^\sharp) \ \ \textit{tels que :}$$

$$r + \theta^\top \cdot \lambda = 0 \ , \tag{1.21a}$$

$$\lambda \in C^\star \ , \ \ \Theta(u^\sharp) \in -C \ , \tag{1.21b}$$

$$\big\langle \lambda \, , \Theta(u^\sharp) \big\rangle = 0 \ . \tag{1.21c}$$

On notera que ces conditions sont l'exacte traduction au cas sous-différentiable des conditions classiques de Karush, Kuhn et Tucker.

Remarque 13. Il n'y a aucune difficulté supplémentaire à ajouter un ensemble admissible et donc à considérer le problème

$$\min_{u \in U^{\mathrm{ad}}} F(u) \ ,$$

sous la contrainte :

$$\Theta(u) \in -C \subset \mathbb{V} \ ,$$

avec U^{ad} sous-ensemble convexe fermé de l'espace $\mathbb{U}$. Il suffit en effet de considérer l'ensemble admissible $U^{\mathrm{ad}} \cap U^{\Theta}$, dont la fonction indicatrice est $\chi_{U^{\mathrm{ad}}} + \chi_{U^{\Theta}}$. Sous l'hypothèse supplémentaire $\mathrm{int}\big(U^{\mathrm{ad}}\big) \cap U^{\Theta} \neq \emptyset$, une application directe du théorème 4 à la relation (1.18) montre que :

$$N_{u^\sharp} U^{\mathrm{ad}} \cap U^{\Theta} = N_{u^\sharp} U^{\mathrm{ad}} + N_{u^\sharp} U^{\Theta} \ .$$

La première condition (1.21a) prend alors la forme :

$$r + \theta^\top \cdot \lambda \in -N_{u^\sharp} U^{\mathrm{ad}} \ ,$$

et les conditions de Karush, Kuhn et Tucker deviennent dans ce cas :

$$\exists\, \lambda \in \mathbb{V} \ , \ \exists\, r \in \partial F(u^\sharp) \ , \ \exists\, \theta \in \partial\Theta(u^\sharp) \ \ \text{tels que :}$$

$$\begin{aligned}
&\big\langle r + \theta^\top \cdot \lambda \, , u - u^\sharp \big\rangle \geq 0 \quad \forall u \in U^{\mathrm{ad}} \ , \\
&\lambda \in C^\star \ , \ \ \Theta(u^\sharp) \in -C \ , \\
&\big\langle \lambda \, , \Theta(u^\sharp) \big\rangle = 0 \ ,
\end{aligned}$$

qui généralisent donc celles énoncées au théorème 12. ◇