Module A2

Éléments de calcul sous-différentiel

L'objectif de ce module est de généraliser la notion de différentiabilité à des fonctions non différentiables. Suivant la manière de procéder, la généralisation obtenue permet de retrouver certaines des propriétés de la différentielle, mais – évidemment – jamais la totalité. C'est donc le contexte d'utilisation qui détermine le choix de cette généralisation. Dans ce cours, on considère une généralisation qui conserve des propriétés intéressantes pour l'optimisation.

Sauf mention contraire, $\mathcal{X}$ est un espace vectoriel réel de dimension finie, muni du produit scalaire noté $\langle \cdot, \cdot \rangle$ et on note $\|\cdot\|$ la norme euclidienne associée. Pour toutes les autres notations et définitions, le lecteur est invité à se reporter aux cours précédents.

1 Sous-différentiabilité : cas convexe

1.1 Définition et exemples

On commence par considérer le cas des fonctions convexes. Pour généraliser la notion de différentielle / gradient, on commence par rappeler que, si $J : \mathcal{X} \to \mathbb{R}$ est une fonction convexe différentiable, alors on a pour tout $x^0 \in \mathcal{X}$

$$\forall\, x \in \mathcal{X}, \qquad J(x) \geq J(x^0) + \langle \nabla J(x^0), x - x^0 \rangle$$

Cette propriété est intéressante, car elle implique en particulier que x^0 est un minimiseur de J si et seulement si $\nabla J(x^0) = 0$. Il s'agit de la règle de Fermat, qui sera présentée plus en détails dans la section prochaine. Ainsi, un critère pour définir une généralisation de la différentielle peut être le suivant : que cette règle reste vraie pour une fonction convexe non différentiable. Il s'ensuit qu'une manière naturelle de procéder est d'introduire la définition suivante :

Définition 1 (Sous-différentiel et sous-gradients d'une fonction convexe)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe** et propre. Soit $x \in \operatorname{dom} J$. On définit le *sous-différentiel* de J au point x^0 comme étant l'ensemble $\partial J(x^0)$ des vecteurs $p \in \mathcal{X}$, appelés *sous-gradients* de J au point x^0, tels que

$$\forall\, x \in \mathcal{X}, \qquad J(x) \geq J(x^0) + \langle p, x - x^0 \rangle$$

Lorsque $x^0 \notin \operatorname{dom} J$, on pose $\partial J(x^0) = \emptyset$.

Si J est convexe et différentiable, alors on a évidemment l'inclusion suivante :

$$\forall\, x^0 \in \mathcal{X}, \qquad \nabla J(x^0) \in \partial J(x_0)$$

La proposition **??** démontre qu'en réalité, $\nabla J(x^0)$ est l'unique sous-gradient de J en x^0. Avant d'énoncer ce résultat, démontrons un lemme qui sera utilisé de manière récurrente dans les preuves qui suivront dans cette section et la suivante.

Lemme 1

Soit $(p, q) \in \mathcal{X}^2$. On suppose que

$$\langle p - q, x - z \rangle \leq o(\|x - z\|)$$

Alors on a $p = q$.

Démonstration : Soit ε une fonction réelle telle que $\varepsilon(t) \to 0$ lorsque $t \to 0$. L'hypothèse du lemme s'écrit donc

$$\forall\, (x, z) \in \mathcal{X}^2, \qquad \langle p - q, x - z \rangle \leq \|x - z\|\, \varepsilon(\|x - z\|)$$

On choisit $x = z + t\, h$ avec $h \in \mathcal{X}$ et $t > 0$. En divisant par t puis en faisant tendre t vers 0, on obtient alors

$$\forall\, h \in \mathcal{X}, \qquad \langle p - q, h \rangle \leq 0$$

ce qui prouve que $p - q = 0$. ■

Proposition 1

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction **convexe** différentiable en $x_0 \in \mathcal{X}$. Alors on a

$$\partial J(x^0) = \big\{ \nabla J(x_0) \big\}$$

> Sous des conditions relativement faibles, on peut établir que la réciproque est vraie (c'est-à-dire que si le sous-différentiel est réduit à un point, alors la fonction est différentiable).

Démonstration : Soit $x^0 \in \mathcal{X}$. Écrivons le développement limité de J en x^0 :

$$\forall\, x \in \mathcal{X}, \qquad J(x) = J(x^0) + \langle \nabla J(x^0), x - x^0 \rangle + o(\|x^0 - x\|)$$

avec $\varepsilon(t) \to 0$ lorsque $t \to 0$. Soit $p \in \partial J(x^0)$. Par définition du sous-gradient, on a

$$\forall\, x \in \mathcal{X}, \qquad J(x) \geq J(x^0) + \langle \nabla J(x^0), x - x^0 \rangle + \langle \nabla J(x^0) - p, x^0 - x \rangle$$

de sorte que, combiné avec la relation précédente, il en résulte que

$$\langle \nabla J(x^0) - p, x^0 - x \rangle \leq o(\|x^0 - x\|)$$

Le lemme 1 prouve alors que $\nabla J(x^0) = p$. ■

Dans le cas général (non différentiable), les sous-gradients $p \in \partial J(x^0)$ d'une fonction convexe sont des vecteurs tels que les fonctions

$$x \mapsto J(x^0) + \langle p, x - x^0 \rangle$$

sont des *minorantes affines exactes* en x^0 de la fonction J, c'est-à-dire une fonction affine qui prend en tout point $x \in \mathcal{X}$ une valeur inférieure à celle de la fonction J au même point, et qui prend la même valeur de J au point x^0. Plus précisément, les sous-gradients de J en x^0 sont exactement les *gradients* des minorantes affines de J exactes en x^0.

Proposition 2 (Existence des sous-gradients I)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe**, s.c.i. et propre. Alors on a

$$\forall\, x^0 \in \operatorname{int}(\operatorname{dom} J), \qquad \partial J(x^0) \neq \emptyset$$

Démonstration : Admis. La preuve fait appel à la notion d'hyperplan séparateur.

Autrement dit, si J est convexe, alors il admet en tout point de l'intérieur de son domaine une minorante affine exacte. Réciproquement, si une fonction au domaine convexe admet une minorante affine exacte en tout point de son domaine, alors on peut montrer qu'elle est convexe :

Proposition 3

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. On suppose que $\operatorname{dom} J$ est **convexe** et que, pour tout $x \in \operatorname{dom} J$, il existe un vecteur $p \in \mathcal{X}$ tel que

$$\forall\, z \in \mathcal{X}, \qquad J(z) \geq J(x) + \langle p, z - x\rangle$$

Alors J est convexe et $p \in \partial J(x)$.

Dans la proposition **??**, les fonctions $z \mapsto J(x) + \langle p, z - x\rangle$ sont des minorantes affines de J exactes en J. Cette proposition se traduit donc de la manière suivante : si J admet en tout point de son domaine convexe une minorante affine exacte, alors J est convexe.

Démonstration : Soit $(x_1, x_2) \in (\operatorname{dom} J)^2$ et $\lambda \in [\,0\,;1\,]$. Posons

$$z = \lambda\, x_1 + (1-\lambda)\, x_2$$

Par hypothèse sur $\operatorname{dom} J$, on a $z \in \operatorname{dom} J$. Il existe donc $p \in \mathcal{X}$ tel que

$$J(x_1) \geq J(z) + \langle p, x_1 - z\rangle \qquad \text{et} \qquad J(x_2) \geq J(z) + \langle p, x_2 - z\rangle$$

Multiplions la première inégalité par λ et la seconde par $(1-\lambda)$, puis sommons les deux inégalités obtenues :

$$\lambda\, J(x_1) + (1-\lambda)\, J(x_2) \geq J(z)$$

On reconnaît l'inégalité de Jensen, ce qui prouve que J est convexe. ■

Donnons quelques exemples de sous-différentiels, qui illustrent une manière de le calculer. On pourra voir que la détermination du sous-différentiel est en général difficile, hormis dans le cas où la fonction est différentiable.

Exemple

Sous-différentiel de la norme euclidienne. On pose

$$J : \left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & \|x\| \end{array} \right.$$

la norme euclidienne. Cette fonction admet des sous-gradients en tout point de $\mathcal{X}$ car elle est continue. Soit $x^0 \neq 0$. Puisque $\|x^0\| = \sqrt{\|x^0\|^2}$ et que la racine carrée est dérivable sur $]\,0\,;+\infty\,[$, la fonction J est différentiable en x^0 et on a

$$\partial J(x^0) = \left\{ \frac{x^0}{\|x^0\|} \right\}$$

Les sous-gradients de J en 0 sont les vecteurs vérifiant

$$\forall\, x \in \mathcal{X}, \qquad \|x\| \geq \langle p, x\rangle$$

L'inégalité de Cauchy-Schwarz assure que les éléments de la boule unité fermée conviennent. Pour s'assurer que ce sont les seuls sous-gradients possibles, on suppose qu'il en existe un de norme strictement supérieure à 1. Dans ce cas, la relation précédente appliquée à $x = p$ assure que

$$\|p\| \geq \|p\|^2$$

ce qui est absurde. On a donc finalement démontré que

$$\forall\, x^0 \in \mathcal{X}, \qquad \partial J(x^0) = \begin{cases} \left\{ \dfrac{x^0}{\|x^0\|} \right\} & \text{si } x^0 \neq 0 \\ \left\{ p \in \mathcal{X} \mid \|p\| \leq 1 \right\} & \text{si } x^0 = 0 \end{cases}$$

Dans la figure 1, on illustre ce résultat dans le cas où $\mathcal{X} = \mathbb{R}$. On y trace en particulier quelques minorantes affines de J exactes en 0.

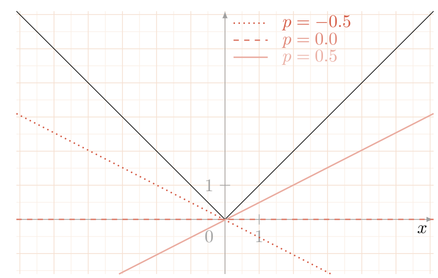

Figure 1 – Sous-gradients en 0 de la valeur absolue. En noir, la courbe représentative de la valeur absolue, en rouge, des minorantes affines exactes en 0.

L'exercice suivant montre qu'une fonction peut admettre des sous-gradients en tout point de son domaine, même si celui-ci est fermé.

Exercice

Sous-différentiel de l'indicatrice d'une boule. Soit $C \in \mathcal{X}$ et $R > 0$. On pose $\mathcal{B}(C, R)$ la boule fermée de centre C et de rayon R, et on considère

$$J = \chi_{\mathcal{B}(C,R)}$$

l'indicatrice de cet ensemble convexe, fermé et non vide. En utilisant la règle de bascule (module A9), montrer que

$$\forall\, x^0 \in \mathcal{X}, \qquad \partial J(x^0) = \begin{cases} \{0\} & \text{si } \|x^0 - C\| < R \\ \left\{ \lambda\,(x^0 - C) \mid \lambda \geq 0 \right\} & \text{si } \|x^0 - C\| = R \end{cases}$$

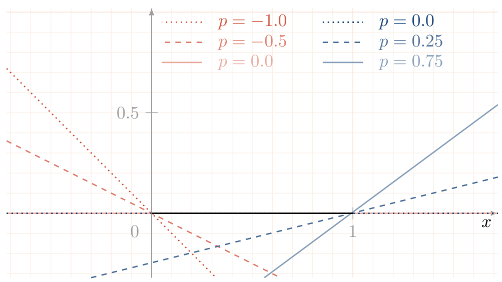

Figure 2 – Sous-gradients en 0 et en 1 de l'indicatrice du segment $[\,0\,;1\,]$. En noir, la courbe représentative de l'indicatrice, en rouge (resp. en bleu), des minorantes affines exactes en 0 (resp. en 1).

> De manière générale, les sous-gradients de l'indicatrice de l'ensemble convexe fermé non vide $\mathcal{C} \subset \mathcal{X}$ en un point x^0 de cet ensemble sont les vecteurs p vérifiant
> $$\forall\, x \in \mathcal{C}, \qquad 0 \geq \langle p, x - x^0 \rangle$$
> L'ensemble de ces points est aussi connu sous le nom *cône normal* à $\mathcal{C}$.

Exemple

Non-existence de sous-gradient au bord du domaine. On considère la fonction réelle définie par

$$\forall\, x \in \mathbb{R}, \qquad \partial J(x) = \begin{cases} -\sqrt{x} & \text{si } x \geq 0 \\ +\infty & \text{sinon} \end{cases}$$

On peut vérifier (laissé en exercice) que J est convexe, s.c.i. et propre. Elle est différentiable en tout point de l'intérieur de son domaine. En $0 \in \operatorname{dom} J$ cependant, on voit graphiquement (cf. figure 3) que toutes les droites passant par l'origine coupe la courbe représentative de J. Vérifions cette observation ; supposons qu'il existe $p \in \mathbb{R}$ tel que

$$\forall\, x \in \mathbb{R}, \qquad J(x) \geq 0 + p\,(x - 0)$$

ce qui revient à vérifier l'existence d'un $p \in \mathbb{R}$ tel que

$$\forall\, x > 0, \qquad -\sqrt{x} \geq p\,x \qquad \text{soit} \qquad -1 \geq p\,\sqrt{x}$$

la relation précédente étant automatiquement vérifiée si $x \leq 0$. Il en découle que p ne peut être positif ; or, p ne peut être strictement négatif, car il vérifierait

$$\forall\, x > 0, \qquad -\frac{1}{p} = \frac{1}{|p|} \leq \sqrt{x}$$

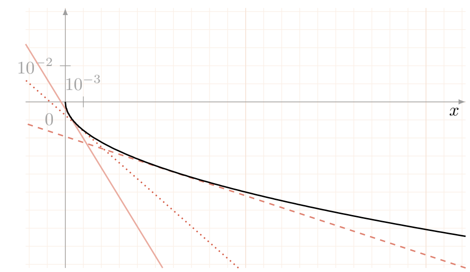

Figure 3 – Cas de non-existence de sous-gradient au bord du domaine. En noir, la courbe représentative de la fonction $J : x \mapsto -\sqrt{x} + \chi_{[\,0\,;+\infty\,[}$. En rouge, différentes minorantes affines de J.

1.2 Propriétés

Établissons quelques propriétés intéressantes du sous-différentiel et des sous-gradients d'une fonction convexe.

Proposition 4

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction **convexe** et propre. Soit $(x, z) \in (\operatorname{dom} J)^2$. Alors on a

$$(p, q) \in \partial J(x) \times \partial J(z) \qquad \Longrightarrow \qquad \langle p - q, x - z \rangle \geq 0$$

On dit que l'opérateur multi-valué ∂J est *monotone*. Lorsque J est différentiable, l'opérateur ∇J est monotone. Lorsque J est une fonction réelle, cette propriété n'est autre que la croissance de la dérivée J'.

Démonstration : Soit $(x, z) \in (\operatorname{dom} J)^2$ et soit $(p, q) \in \partial J(x) \times \partial J(z)$. En écrivant l'inégalité définissant le sous-gradient p au point z et celle définissant q au point x, puis en sommant les deux relations obtenues, on aboutit au résultat désiré. ■

Proposition 5

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ deux fonctions **convexes** propres telles que $\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$. Alors on a

$$\forall\, x \in \mathcal{X}, \qquad \partial f(x) + \partial g(x) \subset \partial (f + g)(x)$$

Démonstration : Commençons par remarquer que $\mathrm{dom}(f+g) = \mathrm{dom}\, f \cap \mathrm{dom}\, g$. Soit $x \in \mathrm{dom}(f+g)$ et $(p,q) \in \partial f(x) \times \partial g(x)$. En écrivant les inégalités définissant p et q, et en les sommant, on obtient

$$\forall\, z \in \mathcal{X}, \qquad f(z) + g(z) \geq f(x) + g(x) + \langle p+q, z-x \rangle$$

On reconnaît la définition de $(p+q) \in \partial(f+g)(x)$. ■

L'inclusion dans la proposition **??** peut être stricte. Cependant, à l'aide de notions qui dépassent le cadre de ce cours, il est possible de démontrer les deux résultats suivants :

Proposition 6

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ deux fonctions **convexes** propres telles que $\mathrm{dom}\, f \cap \mathrm{dom}\, g \neq \emptyset$. On suppose qu'il existe $x_0 \in \mathrm{dom}\, f \cap \mathrm{dom}\, g$ tel que f soit continue en x_0. Alors on a

$$\forall\, x \in \mathcal{X}, \qquad \partial(f+g)(x) = \partial f(x) + \partial g(x)$$

Démonstration : Admis.

Proposition 7

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe** propre et soit $K : \mathcal{Z} \to \mathcal{X}$ un opérateur linéaire. Soit $b \in \mathcal{X}$. Définissons l'application affine suivante :

$$\forall\, x \in \mathcal{Z}, \qquad A\, x = K\, x + b$$

On suppose $A^{-1}\, \mathrm{dom}\, f$ est non vide et qu'il existe $x_0 \in \mathrm{dom}\, f$ tel que f soit continue en $A\, x_0$. Alors on a

$$\forall\, z \in \mathcal{Z}, \qquad \partial(J \circ A)(z) = K^* \partial J(A\, z)$$

Démonstration : Admis.

Pour les autres propriétés du sous-différentiel non spécifiques au cas convexe, le lecteur est invité à consulter la section suivante.

2 Sous-différentiabilité : cas général

2.1 Définition

Définition 2 (Sous-différentiel de Fréchet)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. Soit $x \in \mathrm{dom}\, J$. On définit le *sous-différentiel de* Fréchet de J au point x^0 comme étant l'ensemble $\hat{\partial} J(x^0)$ des vecteurs $p \in \mathcal{X}$ tels que

$$J(x) \geq J(x^0) + \langle p, x - x^0 \rangle + o(\|x - x^0\|)$$

Lorsque $x^0 \notin \mathrm{dom}\, J$, on pose $\hat{\partial} J(x^0) = \emptyset$.

Le sous-différentiel de Fréchet ne possède pas des propriétés suffisamment intéressantes pour en faire une généralisation utile du sous-différentiel d'une fonction convexe. Aussi, introduit-on la définition suivante :

Définition 3 (Sous-différentiel limitant)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. Soit $x \in \mathrm{dom}\, J$. On définit le *sous-différentiel (limitant)* de J au point x^0 comme étant l'ensemble $\partial J(x^0)$ des vecteurs $p \in \mathcal{X}$ tels qu'il existe deux suites $(x_k)_{k \in \mathbb{N}} \in (\mathrm{dom}\, J)^{\mathbb{N}}$ et $(p_k)_{k \in \mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ telles que $p_k \in \hat{\partial} J(x_k)$ pour tout $k \in \mathbb{N}$ et vérifiant

$$\lim_{k \to +\infty} x_k = x^0, \qquad \lim_{k \to +\infty} J(x_k) = J(x^0) \qquad \text{et} \qquad \lim_{k \to +\infty} p_k = p$$

Lorsque $x^0 \notin \mathrm{dom}\, J$, on pose $\partial J(x^0) = \emptyset$.

Grâce au passage à la limite, on augmente le nombre de points dans $\mathrm{dom}\, J$ pour lesquels la fonction J admet des sous-gradients (le *domaine* du sous-différentiel), en même temps que on définit davantage de sous-gradients :

Proposition 8

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. Alors on a

$$\forall\, x \in \mathcal{X}, \qquad \hat{\partial} J(x) \subset \partial J(x)$$

Démonstration : Laissé au lecteur.

Ce faisant, comme on le verra dans un module A3, on conserve des propriétés qui nous seront utiles en optimisation.

Remarque : D'après la définition du sous-différentiel limitant, si deux fonctions J_1 et J_2 ont même domaine et que, pour tout x de leur domaine commun, ils partagent le même sous-différentiel de Fréchet, alors si de plus, pour toute suite $(x_k)_{k \in \mathbb{N}}$ dans $\mathrm{dom}\, J_1$ de limite $x \in \mathrm{dom}\, J_1$, on a l'équivalence

$$\lim_{k \to +\infty} J_1(x_k) = J_1(x) \qquad \Longleftrightarrow \qquad \lim_{k \to +\infty} J_2(x_k) = J_2(x)$$

alors ils partagent également le même sous-différentiel limitant en tout point de leur domaine. C'est le cas par ailleurs dès que les deux fonctions J_1 et J_2 sont continues sur leur domaine commun.

> Il existe d'autres manières de définir un sous-différentiel limitant, notamment en considérant d'autres modes de convergence pour le passage à la limite et/ou d'autres ensembles de sous-gradients que ceux de Fréchet. Parmi les constructions les plus connues, on peut citer le sous-différentiel de Clarke, qui est défini comme l'enveloppe convexe du passage à la limite (dans le sens introduit dans la définition **??**) des *gradients* (aux points où la fonction est différentiable).

Le lecteur aura remarqué l'ambiguïté introduite par l'utilisation de la même notation pour le sous-différentiel d'une fonction convexe (définition **??**) et le sous-différentiel limitant (définition **??**). La proposition suivante prouve que cette ambiguïté n'est pas gênante :

Proposition 9

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe** propre. Alors les deux ensembles définis dans les définitions 1 et 2 coïncident, et on a par ailleurs

$$\forall\, x \in \mathcal{X}, \qquad \hat{\partial} J(x) = \partial J(x)$$

C'est pourquoi l'on n'utilisera désormais plus l'adjectif *limitant*.

Démonstration : *Pour des raisons de lisibilité, on se contente de démontrer partiellement ce résultat.* Pour lever temporairement l'ambiguïté de notation, on distingue $\partial^{\text{convexe}} J(x)$ l'ensemble des sous-gradients tels que définis par la définition **??** et $\partial^{\text{limitant}} J(x)$ l'ensemble des sous-gradients tels que définis par la définition **??**. Il est immédiat que

$$\forall\, x \in \mathcal{X}, \qquad \partial^{\text{convexe}} J(x) \subset \hat{\partial} J(x) \subset \partial^{\text{limitant}} J(x)$$

On admet que l'inclusion inverse est vraie. Soit $p \in \partial^{\text{limitant}} J(x)$. Il existe par définition deux suites $(x_k)_{k\in\mathbb{N}}$ et $(p_k)_{k\in\mathbb{N}}$ telles que $p_k \in \hat{\partial} J(x_k)$ pour tout $k \in \mathbb{N}$ et

$$\lim_{k\to+\infty} x_k = x, \qquad \lim_{k\to+\infty} J(x_k) = J(x) \qquad \text{et} \qquad \lim_{k\to+\infty} p_k = p$$

Or les p_k vérifient $\qquad \forall\, z \in \mathcal{X}, \qquad J(z) \geq J(x_k) + \langle p_k, z - x_k \rangle$

Il suffit alors de passer à la limite dans cette relation pour tout $z \in \mathcal{X}$ fixé. ■

Dans le cas différentiable, les notions de sous-différentiel / sous-gradients et de gradient peuvent être confondues :

Proposition 10

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. Soit $x^0 \in \operatorname{dom} J$.

(i) Si J est différentiable en x^0, alors $\hat{\partial} J(x^0) = \big\{\nabla J(x^0)\big\}$.

(ii) Si, de plus, ∇J est continu au voisinage de x^0, alors $\partial J(x^0) = \big\{\nabla J(x^0)\big\}$.

Dans tous les cas, si J est différentiable en $x^0 \in \operatorname{dom} J$, alors $\nabla J(x^0) \in \partial J(x^0)$.

Démonstration : Démontrons successivement les deux affirmations.

- **Affirmation (i).** Puisque J est différentiable en x^0, on a par définition que

$$J(x) = J(x^0) + \langle \nabla J(x^0), x - x^0 \rangle + o(\|x - x^0\|)$$

ce qui assure que $\nabla J(x^0) \in \hat{\partial} J(x^0)$. Soit $p \in \hat{\partial} J(x^0)$. On a donc

$$J(x^0) + \langle \nabla J(x^0), x - x^0 \rangle + o(\|x - x^0\|) \geq J(x^0) + \langle p, x - x^0 \rangle + o(\|x - x^0\|)$$

ce qui donne après simplification

$$\langle p - \nabla J(x^0), x - x^0 \rangle \leq o(\|x - x^0\|)$$

Le lemme 1 démontre que $p = \nabla J(x^0)$.

- **Affirmation (ii).** Soit $p \in \partial J(x^0)$. Par définition du sous-différentiel, il existe deux suites $(x_k)_{k\in\mathbb{N}} \in (\operatorname{dom} J)^{\mathbb{N}}$ et $(p_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ vérifiant

$$\lim_{k\to+\infty} x_k = x^0, \quad \lim_{k\to+\infty} J(x_k) = J(x^0) \qquad \text{et} \qquad \lim_{k\to+\infty} p_k = p$$

Par hypothèse, il existe un voisinage $\mathcal{V}$ de x^0 sur lequel J est différentiable. À partir d'un certain indice $k_0 \in \mathbb{N}$, les x_k appartiennent à $\mathcal{V}$. D'après l'affirmation (i), on a

$$\forall\, k \geq k_0, \quad \hat{\partial} J(x_k) = \{\nabla J(x_k)\} \qquad \text{soit} \qquad \forall\, k \geq k_0, \quad p_k = \nabla J(x_k)$$

On en déduit que $\qquad p = \lim_{k\to+\infty} \nabla J(x_k) = \nabla J(x^0)$

par continuité de ∇J sur $\mathcal{V}$. ■

Exemple

Sous-différentiel d'une fonction concave non différentiable. On considère la fonction concave $J = -|\cdot|$. Puisque J est de classe $\mathcal{C}^1$ sur $]-\infty\,;0[\,\cup\,]\,0\,;+\infty[$, on en déduit que

$$\forall\, x \neq 0, \qquad \hat{\partial} J(x) = \partial J(x) = \left\{-\frac{x}{|x|}\right\} \in \{1, -1\}$$

Aussi, il s'ensuit immédiatement que $\partial J(0) = \{1, -1\}$. Par ailleurs, $\hat{\partial} J(0) = \emptyset$. En effet, dans le cas contraire, il existerait $p \in \{1, -1\}$ tel que

$$\forall\, x \in \mathbb{R}^*, \qquad -|x| \geq p\,x + |x|\,\varepsilon(|x|) \qquad \text{soit} \qquad -1 \geq p\,\frac{x}{|x|} + \varepsilon(|x|)$$

avec $\varepsilon(t) \to 0$ lorsque $t \to 0$. En faisant tendre x vers 0 par valeurs inférieures si p est négatif et par valeurs supérieures sur p est positif, on obtient une contradiction.

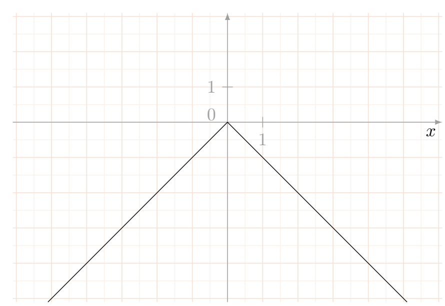

Figure 4 – Courbe représentative de l'opposé de la valeur absolue (en noir).

On termine ce paragraphe avec un résultat intéressant (que l'on admettra).

Proposition 11

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction s.c.i. Alors il existe une suite $(x_k)_{k\in\mathbb{N}} \in (\operatorname{dom} J)^{\mathbb{N}}$ vérifiant

$$\lim_{k\to+\infty} x_k = x^0, \qquad \lim_{k\to+\infty} J(x_k) = J(x^0) \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \quad \hat{\partial} J(x_k) \neq \emptyset$$

Autrement dit, l'ensemble des points qui admettent un sous-gradient de Fréchet est dense dans le domaine de J.

Démonstration : Admis.

2.2 Règles de calcul sous-différentiel

On donne ici quelques règles permettant de calculer le sous-différentiel.

Proposition 12

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction et soit $\alpha > 0$. Alors on a

$$\forall\, x \in \mathcal{X}, \qquad \partial(\alpha J)(x) = \alpha\, \partial J(x)$$

Démonstration : Il suffit de remarquer que

$$\alpha\, J(x) \geq \alpha\, J(x_k) + \langle p_k, x - x_k\rangle + o(\|x - x_k\|)$$

équivaut à $$J(x) \geq J(x_k) + \left\langle \frac{p_k}{\alpha}, x - x_k \right\rangle + o(\|x - x_k\|)$$

de sorte que $\hat{\partial}(\alpha\, J)(x) = \alpha\, \hat{\partial} J(x)$. On conclut alors par passage à la limite. ■

Proposition 13

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction et soit $\alpha < 0$. Alors on a

$$\forall\, x \in \mathcal{X}, \qquad \partial(\alpha J)(x) = -\alpha\, \partial(-J)(x)$$

Démonstration : Laissé au lecteur.

Contre-exemple

Sous-différentiel d'une somme. On considère les deux fonctions suivantes :

$$f : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \max\{0, x\} \end{array} \right. \qquad \text{et} \qquad g : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \min\{0, x\} \end{array} \right.$$

La fonction f est convexe ; en 0, son sous-différentiel vaut $\partial f(0) = [\,0\,;1\,]$. Par ailleurs, on peut vérifier que g étant continûment différentiable sur $]\,0\,;+\infty\,[$ et sur $]-\infty\,;0\,[$, on a

$$\partial g(x) = \{\nabla g(x)\} = \{0\} \text{ si } x > 0 \qquad \text{et} \qquad \partial g(x) = \{\nabla g(x)\} = \{1\} \text{ si } x < 0$$

Par passage à la limite, on vérifie que $\partial g(0) = \{0, 1\}$. Ainsi,

$$\partial f(0) + \partial g(0) = [\,0\,;2\,]$$

Or, en remarquant que $f(x) + g(x) = x$ quel que soit $x \in \mathbb{R}$, on en déduit que

$$\partial(f + g)(0) = \{1\} \subset [\,0\,;2\,]$$

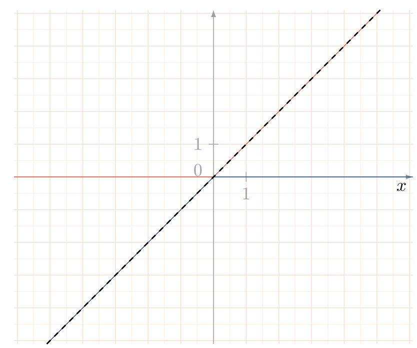

Figure 5 – En noir, la courbe représentative de la fonction $J = f + g$ avec $f : x \mapsto \max\{0, x\}$ (en rouge) et $g : x \mapsto \min\{0, x\}$ (en bleu).

En général, le sous-différentiel de la somme de deux fonctions n'est pas égal à la somme des sous-différentiels de chaque terme de la somme. Il existe toutefois trois cas intéressants pour lesquels on a bien cette égalité, à savoir le cas convexe (déjà vu plus haut), le cas différentiable et le cas séparable.

Proposition 14

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ et $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ deux fonctions telles que $\operatorname{dom}(f + g) \neq \emptyset$. Soit $x^0 \in \operatorname{dom}(f + g)$. On suppose que $f(x^0)$ est fini et que g est continûment différentiable au voisinage de x^0. Alors on a

$$\partial(f + g)(x^0) = \partial f(x^0) + \nabla g(x^0)$$

Démonstration : On démontre séparément les deux inclusions.

- **Inclusion** $\supset$. Soit $p \in \partial f(x^0)$. Par définition, il existe une suite $(x_k)_{k\in\mathbb{N}}$ dans $\operatorname{dom} f$ telle que

$$\lim_{k\to+\infty} x_k = x^0 \qquad \text{et} \qquad \lim_{k\to+\infty} f(x_k) = f(x^0)$$

Par continuité de g au voisinage de x^0, la suite $(x_k)_{k\in\mathbb{N}}$ vérifie

$$\lim_{k\to+\infty} x_k = x^0 \qquad \text{et} \qquad \lim_{k\to+\infty} \big(f(x_k)+g(x_k)\big) = f(x^0)+g(x^0)$$

Par ailleurs, il existe une suite $(p_k)_{k\in\mathbb{N}}$ dans $\mathcal{X}$ telle que

$$\forall\, k\in\mathbb{N}, \quad f(x) \geq f(x_k) + \langle p_k, x-x_k\rangle + o(\|x-x_k\|) \qquad \text{et} \qquad \lim_{k\to+\infty} p_k = p$$

La fonction g étant différentiable au voisinage x^0, on en déduit qu'il existe un rang $k_0\in\mathbb{N}$ tel que

$$\forall\, k\geq k_0, \quad g(x) \geq g(x_k) + \langle \nabla g(x_k), x-x_k\rangle + o(\|x-x_k\|)$$

En sommant les deux relations précédentes, il en découle que

$$\forall\, k\geq k_0, \quad f(x)+g(x) \geq f(x_k)+g(x_k) + \langle p_k+\nabla g(x_k), x-x_k\rangle + o(\|x-x_k\|)$$

La continuité de ∇g au voisinage de x^0 assure alors finalement que

$$\lim_{k\to+\infty} \big(p_k + \nabla g(x_k)\big) = p + \nabla g(x^0)$$

ce qui démontre le résultat annoncé.

- **Inclusion $\subset$.** Posons $F=f+g$ et $G=-g$. La fonction G étant continûment différentiable autour de x^0, on a démontré au point précédent que

$$\partial F(x^0)+\nabla G(x^0) \subset \partial(F+G)(x^0) \qquad \text{soit} \qquad \partial(f+g)(x^0)-\nabla g(x^0)\subset \partial f(x^0)$$

Il suffit alors d'ajouter le vecteur $\nabla g(x^0)$ aux deux membres de cette inclusion pour obtenir le résultat souhaité. ■

Cette proposition permet de démontrer une relation utile dans le cadre des fonctions fortement convexes :

Corollaire 1

Soit $J:\mathcal{X}\to\mathbb{R}$ une fonction **fortement convexe**, de module α, et propre. Soit $x\in\operatorname{dom} J$ et $p\in\partial J(x)$. Alors on a

$$\forall\, z\in\mathcal{X}, \qquad J(z) \geq J(x) + \langle p, z-x\rangle + \frac{\alpha}{2}\,\|z-x\|^2$$

Démonstration : La forte convexité de J équivalant à la convexité de la fonction

$$x\mapsto J(x) - \alpha\,\|x-x^0\|^2/2$$

(d'après la proposition 15 du module A1) on utilise la proposition précédente pour démontrer que

$$p\in\partial J(x) \qquad \Longleftrightarrow \qquad p-\alpha\,(x-x^0) \in \partial\left(x\mapsto J(x)-\frac{\alpha}{2}\,\|x-x^0\|^2\right)$$

Or, par définition du sous-gradient d'une fonction convexe, ce dernier vecteur vérifie

$$\forall\, z\in\mathcal{X}, \qquad J(z)-\frac{\alpha}{2}\,\|z-x^0\|^2 \geq J(x)-\frac{\alpha}{2}\,\|x-x^0\|^2 + \langle p-\alpha\,(x-x^0), z-x\rangle$$

On obtient après simplification le résultat souhaité. ■

Proposition 15

Soit $f:\mathcal{X}\to\mathbb{R}\cup\{+\infty,-\infty\}$ et $g:\mathcal{Z}\to\mathbb{R}\cup\{+\infty,-\infty\}$ deux fonctions telles que le domaine de la fonction

$$J:\left\{\begin{array}{ccc} \mathcal{X}\times\mathcal{Z} & \to & \mathbb{R}\cup\{+\infty\} \\ (x,z) & \mapsto & f(x)+g(z)\end{array}\right.$$

soit non vide. Alors on a

$$\forall\,(x,z)\in\mathcal{X}\times\mathcal{Z}, \qquad \partial J(x,z) = \partial f(x)\times\partial g(z)$$

On dit que J est *(additivement) séparable*. Notons que les espaces $\mathcal{X}$ et $\mathcal{Z}$ peuvent être de dimension différente.

Démonstration : Remarquons que $\operatorname{dom} J = \operatorname{dom} f\times\operatorname{dom} g$.

- Commençons par montrer que la propriété voulue est vraie pour le sous-différentiel de Fréchet. Soit $(x^0,z^0)\in\operatorname{dom} J$ et $p\in\hat\partial f(x^0)$ et $q\in\hat\partial g(z^0)$. On a alors

$$f(x) \geq f(x^0) + \langle p, x-x^0\rangle + o(\|x-x^0\|)$$

et

$$g(z) \geq g(z^0) + \langle q, z-z^0\rangle + o(\|z-z^0\|)$$

soit, en posant $X=(x,z)$, $X^0=(x^0,z^0)$ et $P=(p,q)$, et en sommant les deux inégalités précédentes

$$J(X) \geq J(X^0) + \langle P, X-X^0\rangle + o(\|X-X^0\|)$$

ce qui implique que $P\in\hat\partial J(X^0)$. Réciproquement, si $(p,q)\in\partial J(x^0,y^0)$, alors il suffit d'écrire la relation précédente pour $X=(x,0)$ et $X=(0,z)$, en remarquant que $X\to 0$ si et seulement si $x\to 0$ et $z\to 0$ respectivement.

- Démontrons le résultat principal. Soit $(x,z)\in\operatorname{dom} J$. Alors $p\in\partial f(x)$ et $q\in\partial g(z)$ si et seulement s'il existe quatre suites $(x_k)_{k\in\mathbb{N}}$, $(z_k)_{z\in\mathbb{N}}$, $(p_k)_{z\in\mathbb{N}}$ et $(q_k)_{z\in\mathbb{N}}$ dans $\operatorname{dom} f$, $\operatorname{dom} g$, $\mathcal{X}$ et $\mathcal{Z}$ respectivement, vérifiant d'une part

$$\lim_{k\to+\infty} x_k = x, \quad \lim_{k\to+\infty} f(x_k) = f(x), \quad \lim_{k\to+\infty} z_k = z, \quad \lim_{k\to+\infty} g(z_k) = g(z)$$

ce qui s'écrit de manière équivalente

$$\lim_{k\to+\infty} (x_k,z_k) = (x,z) \qquad \text{et} \qquad \lim_{k\to+\infty} \big(f(x_k),g(z_k)\big) = \big(f(x),g(z)\big)$$

et d'autre part

$$\forall\, k\in\mathbb{N}, \quad p_k\in\hat\partial f(x_k), \quad \lim_{k\to+\infty} p_k = p, \quad q_k\in\hat\partial g(z_k), \quad \lim_{k\to+\infty} q_k = q$$

ce qui s'écrit de manière équivalente

$$\forall\, k\in\mathbb{N}, \qquad (p_k,q_k)\in\hat\partial f(x_k)\times\hat\partial g(z_k) \qquad \text{et} \qquad \lim_{k\to+\infty} (p_k,q_k) = (p,q)$$

Puisqu'on a démontré plus haut l'identité $\hat\partial f(x_k)\times\hat\partial g(z_k) = \hat\partial J(z_k)$, on voit que ce qui précède est vrai si et seulement si $(p,q)\in\partial J(x,z)$. ■

2.3 Sous-différentiels partiels

Dans ce paragraphe, on considère le cas de fonctions à plusieurs variables

$$J : \left\{ \begin{array}{ccc} \mathcal{X}_1 \times \cdots \times \mathcal{X}_n & \to & \mathbb{R} \cup \{+\infty, -\infty\} \\ (x_1, \ldots, x_n) & \mapsto & J(x_1, \ldots, x_n) \end{array} \right.$$

où les X_i sont des espaces vectoriels de dimension finie. Pour plus de lisibilité, on se restreint au cas des fonctions à deux variables

$$J : \left\{ \begin{array}{ccc} \mathcal{X} \times \mathcal{Z} & \to & \mathbb{R} \cup \{+\infty, -\infty\} \\ (x, z) & \mapsto & J(x, z) \end{array} \right.$$

avec $\mathcal{Z}$ un espace vectoriel de dimension finie.

Définition 4 (Sous-différentiels partiels)

Soit $J : \mathcal{X} \times \mathcal{Z} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. Pour tout $(x^0, z^0) \in \operatorname{dom} J$, on appelle le *sous-différentiel partiel* par rapport à la variable x (resp. z) de la fonction J au point (x^0, z^0), noté $\partial_x J(x^0, z^0)$ (resp. $\partial_z J(x^0, z^0)$) le sous-différentiel de la fonction partielle $x \mapsto J(x, z^0)$ au point x^0 (resp. le sous-différentiel de la fonction partielle $z \mapsto J(x^0, z)$ au point z^0).

Autrement dit, $p \in \partial_x J(x^0, z^0)$ si et seulement s'il existe deux suites $(x_k)_{k\in\mathbb{N}}$ et $(p_k)_{z\in\mathbb{N}}$ dans $\operatorname{dom} f$ et $\mathcal{X}$ respectivement, vérifiant d'une part

$$\lim_{k\to+\infty} x_k = x, \qquad \lim_{k\to+\infty} J(x_k, z_0) = J(x, z_0)$$

et d'autre part

$$\forall\, k \in \mathbb{N}, \qquad J(x, z^0) \in J(x^0, z^0) + \langle p_k, x - x^0 \rangle + o(\|x - x^0\|)$$

et telles que

$$\lim_{k\to+\infty} p_k = p$$

(Le lecteur est invité à expliciter la définition de $\partial_z J(x^0, z^0)$ de la même manière).

Proposition 16

Soit $J : \mathcal{X} \times \mathcal{Z} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. Soit $(x^0, z^0) \in \operatorname{dom} J$. Si J est continûment différentiable dans un voisinage de (x^0, z^0), alors

$$\partial_x J(x^0, z^0) = \left\{ \frac{\partial J}{\partial x}(x^0, z^0) \right\} \qquad \text{et} \qquad \partial_z J(x^0, z^0) = \left\{ \frac{\partial J}{\partial z}(x^0, z^0) \right\}$$

Démonstration : Il suffit d'écrire la définition des sous-différentiels partiels et d'appliquer la proposition **??**. ■

Une conséquence immédiate de cette proposition est la structure en produit cartésien du sous-différentiel lorsque la fonction est de classe $\mathcal{C}^1$ autour du point considéré. En effet, puisque

$$\nabla J(x^0, z^0) = \begin{pmatrix} \dfrac{\partial J}{\partial x}(x^0, z^0) \\ \dfrac{\partial J}{\partial z}(x^0, z^0) \end{pmatrix}$$

il s'ensuit naturellement que

$$\partial J(x^0, z^0) = \left\{ \frac{\partial J}{\partial x}(x^0, z^0) \right\} \times \left\{ \frac{\partial J}{\partial z}(x^0, z^0) \right\} = \partial_x J(x^0, z^0) \times \partial_z J(x^0, z^0)$$

Dans le cas non différentiable, cette identité n'est plus vérifiée en général, comme le montre l'exemple suivant.

Contre-exemple

Lien entre le sous-différentiel et les sous-différentiels partiels dans le cas non différentiable. On considère la fonction suivante :

$$J : \left\{ \begin{array}{ccc} \mathbb{R}^2 & \to & \mathbb{R} \\ (x, z) & \mapsto & |x + z| + 2\,|x - z| \end{array} \right.$$

On peut vérifier que J et ses fonctions partielles sont convexes et continues, de sorte que leurs sous-différentiels sont définies comme dans la définition **??**. Par ailleurs, J non différentiable en tout point (a, a) avec $a \in \mathbb{R}$. Soit $a \neq 0$. Il est aisé de vérifier que

$$\partial_x J(a, a) = \nabla(x \mapsto |x + a|)(a) + 2\,\partial(x \mapsto |x - a|)(a) = \frac{a}{|a|} + 2\,[\,-1\,;1\,]$$

et que

$$\partial_z J(a, a) = \nabla(z \mapsto |a + z|)(a) + 2\,\partial(z \mapsto |a - z|)(a) = \frac{a}{|a|} + 2\,[\,-1\,;1\,]$$

de sorte que, puisque $a/|a| \in \{1, -1\}$, on a

$$(0, 0) \in \partial_x J(a, a) \times \partial_z J(a, a)$$

Montrons que $(0, 0) \neq \partial J(a, a)$. En effet, dans le cas contraire, on aurait par définition

$$\forall (x, z) \in \mathbb{R}^2, \qquad |x + z| + 2\,|x - z| \geq 2\,|a|$$

ce qui n'est pas le cas lorsque $x = z = 0$. Il s'ensuit que

$$\partial_x J(a, a) \times \partial_z J(a, a) \neq \partial J(a, a)$$

Il est possible en réalité de trouver des exemples où les deux ensembles

$$\partial_x J(x, z) \times \partial_z J(x, z) \qquad \text{et} \qquad \partial J(x, z)$$

sont égaux, le premier inclus dans le second ou l'inverse, ou ni l'un ni l'autre. On voit donc que, dans le cas non différentiable, il existe toute une variété de situations possibles, et ce, même dans le cas pourtant relativement maîtrisé des fonctions convexes. Ainsi, l'implication

$$(0, 0) \in \partial_x J(x, z) \times \partial_z J(x, z) \qquad \Longrightarrow \qquad (0, 0) \in \partial J(x, z)$$

est généralement une propriété cruciale pour une certaine classe de méthodes d'optimisation du premier ordre.

Proposition 17

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$, $g : \mathcal{Z} \to \mathbb{R} \cup \{+\infty, -\infty\}$ et $h : \mathcal{X} \times \mathcal{Z} \to \mathbb{R} \cup \{+\infty, -\infty\}$ trois fonctions telles que le domaine de la fonction

$$J : \left\{ \begin{array}{ccc} \mathcal{X} \times \mathcal{Z} & \to & \mathbb{R} \cup \{+\infty\} \\ (x, z) & \mapsto & f(x) + g(z) + h(x, z) \end{array} \right.$$

soit non vide. Soit $(x^0, z^0) \in \operatorname{dom} J$. On suppose que h est continûment différentiable au voisinage de (x^0, z^0). Alors on a

$$\partial J(x^0, z^0) = \partial_x J(x^0, z^0) \times \partial_z J(x^0, z^0)$$

avec

$$\partial_x J(x^0, z^0) = \frac{\partial h}{\partial x}(x^0, z^0) + \partial f(x^0) \qquad \text{et} \qquad \partial_z J(x^0, z^0) = \frac{\partial h}{\partial z}(x^0, z^0) + \partial g(z^0)$$

Démonstration : Il s'agit d'une conséquence des propositions **??**, **??** et **??**. ■

3 Fermeture du sous-différentiel

Une conséquence de la définition du sous-différentiel est la propriété suivante :

Proposition 18 (Fermeture du sous-différentiel)

Soit $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. Soit $x^0 \in \operatorname{dom} J$. On suppose qu'il existe une suite $(x_k)_{k\in\mathbb{N}}$ dans $\mathcal{X}$ telle que

$$\lim_{k\to+\infty} x_k = x^0 \qquad \text{et} \qquad \lim_{k\to+\infty} J(x_k) = J(x^0)$$

et une suite $(p_k)_{k\in\mathbb{N}}$ dans $\mathcal{X}$ telle que

$$\forall\, k \in \mathbb{N}, \qquad p_k \in \partial J(x_k) \qquad \text{et} \qquad \lim_{k\to+\infty} p_k = p$$

Alors on a $p \in \partial J(x^0)$.

Démonstration : On suppose que les suites $(x_k)_{k\in\mathbb{N}}$ et $(p_k)_{k\in\mathbb{N}}$ vérifient les hypothèses de la proposition. Soit $k \in \mathbb{N}$. Par hypothèse, il existe deux suites $(x_{k,j})_{j\in\mathbb{N}}$ et $(p_{k,j})_{j\in\mathbb{N}}$ dans $\operatorname{dom} J$ et $\mathcal{X}$ respectivement, telles que

$$\lim_{j\to+\infty} x_{k_j} = x_k \qquad \text{et} \qquad \lim_{j\to+\infty} J(x_{k_j}) = J(x_k)$$

$$\text{et} \qquad \forall\, j \in \mathbb{N}, \qquad J(x) \geq J(x_{k,j}) + \langle p_{k,j}, x - x_{k,j}\rangle + o(\|x - x_{k,j}\|)$$

Il est alors possible de construire une suite $(\hat{x}_k)_{k\in\mathbb{N}}$ définie par $\hat{x}_k = x_{k,j_k}$ pour un $j_k \in \mathbb{N}$ donné vérifiant $j_k > j_{k'}$ si $k > k'$, de sorte que

$$\lim_{k\to+\infty} \hat{x}_k = x, \qquad \lim_{k\to+\infty} J(\hat{x}_k) = J(x) \qquad \text{et} \qquad \lim_{k\to+\infty} \hat{p}_k = p$$

en définissant $\hat{p}_k = p_{k,j_k}$ pour tout $k \in \mathbb{N}$. Il s'ensuit alors que $p \in \partial J(x^0)$. ■

Lorsque J est de la forme

$$J : \left\{ \begin{array}{ccc} \mathcal{X} \times \mathcal{Z} & \to & \mathbb{R} \cup \{+\infty, -\infty\} \\ (x, z) & \mapsto & J(x, z) \end{array} \right.$$

la fermeture du sous-différentiel s'écrit : pour tout $(x^0, z^0) \in \operatorname{dom} J$ et toute suite $((x_k, z_k))_{k\in\mathbb{N}}$ dans $\mathcal{X} \times \mathcal{Z}$ telle que

$$\lim_{k\to+\infty} (x_k, z_k) = (x^0, z^0) \qquad \text{et} \qquad \lim_{k\to+\infty} J(x_k, z_k) = J(x^0, z^0)$$

et pour toute suite $(p_k, q_k)_{k\in\mathbb{N}}$ dans $\mathcal{X} \times \mathcal{Z}$ telles que

$$\forall\, k \in \mathbb{N}, \qquad (p_k, q_k) \in \partial J(x_k, z_k)$$

si $(p_k, q_k)_{k\in\mathbb{N}}$ est convergente, de limite respective (p, q), alors on a

$$(p, q) \in \partial J(x^0, z^0)$$

Cependant, comme on le verra dans un exemple plus bas, cette propriété n'est pas préservée lorsque l'on remplace le sous-différentiel ∂J par les sous-différentiels partiels $\partial_x J$ et $\partial_z J$. Or, on verra que, si la propriété de fermeture du sous-différentiel est une propriété utile pour prouver la convergence de certaines méthodes d'optimisation du premier ordre, elle n'est cependant pas toujours suffisante, et que l'on a besoin d'une propriété analogue pour les sous-différentiels partiels. Aussi introduit-on les définitions suivantes :

Définition 5 (Fermeture paramétrique des sous-différentiels partiels)

Soit $J : \mathcal{X} \times \mathcal{Z} \to \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. Soit $(x^0, z^0) \in \operatorname{dom} J$ et une suite $((x_k, z_k))_{k\in\mathbb{N}}$ dans $\mathcal{X} \times \mathcal{Z}$ telle que

$$\lim_{k\to+\infty} (x_k, z_k) = (x^0, z^0) \qquad \text{et} \qquad \lim_{k\to+\infty} J(x_k, z_k) = J(x^0, z^0) \qquad (\star)$$

- On dit que $\partial_x J$ (resp. $\partial_z J$) est *paramétriquement fermé en* (x^0, z^0) *relativement à* $((x_k, z_k))_{k\in\mathbb{N}}$ si pour toute suite $(p_k)_{k\in\mathbb{N}}$ dans $\mathcal{X}$ (resp. $(q_k)_{k\in\mathbb{N}}$ dans $\mathcal{Z}$) telle que

$$\forall\, k \in \mathbb{N}, \qquad p_k \in \partial_x J(x_k, z_k) \qquad \text{et} \qquad \lim_{k\to+\infty} p_k = p$$

$$(\text{resp.} \quad \forall\, k \in \mathbb{N}, \qquad q_k \in \partial_z J(x_k, z_k) \qquad \text{et} \qquad \lim_{k\to+\infty} q_k = q)$$

  on a $p \in \partial_x J(x^0, z^0)$ (resp. $q \in \partial_z J(x^0, z^0)$).
- On dit $\partial_x J$ (resp. $\partial_z J$) est *fermé en* (x^0, z^0) si $\partial_x J$ (resp. $\partial_z J$) est paramétriquement fermé en (x^0, z^0) relativement à n'importe quelle suite vérifiant $(\star)$.
- On dit que $\partial_x J$ (resp. $\partial_z J$) est *fermé* si $\partial_x J$ (resp. $\partial_z J$) est fermé en (x^0, z^0) en tout point $(x^0, z^0) \in \operatorname{dom} J$.

Contrairement à la fermeture du sous-différentiel, la fermeture paramétrique du sous-différentiel n'est pas toujours acquise, comme le montre l'exemple suivant.

Contre-exemple

On considère la fonction définie sur $\mathbb{R}^2$ par

$$\forall\,(x,z)\in\mathbb{R}^2,\qquad J(x,z)=\begin{cases}\sqrt{x\,z} & \text{si } x\geq 0 \text{ et } z\geq 0\\ +\infty & \text{sinon}\end{cases}$$

Soit $(x^0,z^0)=(0,0)\in\operatorname{dom}J$. Soit $(x_k)_{k\in\mathbb{N}}$ une suite de $]\,0\,;+\infty\,[$ tendant vers 0, et posons $z_k=x_k$ pour tout $k\in\mathbb{N}$. Pour tout $k\in\mathbb{N}$, la fonction partielle $x\mapsto J(x,z_k)$ est visiblement dérivable sur $]\,0\,;+\infty\,[$, de sorte que

$$\partial_x J(x_k,z_k)=\left\{\frac{\sqrt{z_k}}{2\sqrt{x_k}}\right\}=\left\{\frac{1}{2}\right\}\qquad\text{avec}\qquad\lim_{k\to+\infty}\frac{1}{2}=\frac{1}{2}$$

Par continuité de J sur son domaine, on a par ailleurs

$$\lim_{k\to+\infty}J(x_k,z_k)=0=J(0,0)$$

Calculons $\partial_x J(0,0)$. Puisque $J(x,0)=\chi_{[\,0\,;+\infty\,[}(x)$, on peut montrer que

$$\partial_x J(0,0)=\,]-\infty\,;0\,]\not\ni\frac{1}{2}$$

ce qui prouve que $\partial_x J$ n'est pas fermé en $(0,0)$.

Proposition 19

Soit $J:\mathcal{X}\times\mathcal{Z}\to\mathbb{R}$ une fonction. Soit $(x^0,z^0)\in\operatorname{dom}J$. On suppose que J continûment différentiable au voisinage de (x^0,z^0). Alors ses sous-différentiels partiels $\partial_x J$ et $\partial_z J$ sont fermés en (x^0,z^0).

Démonstration : Laissé au lecteur.

Proposition 20

Soit $f:\mathcal{X}\to\mathbb{R}\cup\{+\infty,-\infty\}$, $g:\mathcal{Z}\to\mathbb{R}\cup\{+\infty,-\infty\}$ et $h:\mathcal{X}\times\mathcal{Z}\to\mathbb{R}\cup\{+\infty,-\infty\}$ trois fonctions telles que le domaine de la fonction

$$J:\left\{\begin{array}{ccc}\mathcal{X}\times\mathcal{Z} & \to & \mathbb{R}\cup\{+\infty\}\\ (x,z) & \mapsto & f(x)+g(z)+h(x,z)\end{array}\right.$$

soit non vide. Soit $(x^0,z^0)\in\operatorname{dom}J$. On suppose que h est continûment différentiable au voisinage de (x^0,z^0). Alors ses sous-différentiels partiels $\partial_x J$ et $\partial_z J$ sont fermés en (x^0,z^0).

Démonstration : Soit $((x_k,z_k))_{k\in\mathbb{N}}$ une suite vérifiant

$$\lim_{k\to+\infty}(x_k,z_k)=(x^0,z^0)\qquad\text{et}\qquad\lim_{k\to+\infty}J(x_k,z_k)=J(x^0,z^0)$$

Soit $(p_k)_{k\in\mathbb{N}}$ une suite convergente de limite p, telle que $p_k\in\partial_x J(x_k,z_k)$ pour tout $k\in\mathbb{N}$. D'après la proposition **??**, on a

$$\forall\,k\in\mathbb{N},\qquad q_k=p_k-\frac{\partial h}{\partial x}(x_k,z_k)\in\partial f(x_k)$$

Par continuité de ∇h, la suite des q_k converge vers

$$p-\frac{\partial h}{\partial x}(x^0,z^0)$$

La fermeture du sous-différentiel de f assure alors que

$$p-\frac{\partial h}{\partial x}(x^0,z^0)\in\partial f(x^0)\qquad\text{soit}\qquad p\in\partial_x J(x^0,z^0)$$

On démontre de manière analogue que $\partial_z J$ est fermé en (x^0,z^0). ■

Proposition 21

Soit $J:\mathcal{X}\times\mathcal{Z}\to\mathbb{R}\cup\{+\infty\}$ une fonction biconvexe. Soit $(x^0,z^0)\in\operatorname{dom}J$.

(i) Si J est continue en (x^0,z^0), alors son sous-différentiel partiel $\partial_x J$ (resp. $\partial_z J$) est fermé en (x^0,z^0) relativement à toute suite $((x_k,z_k))_{k\in\mathbb{N}}$ telle que

$$\forall\,x\in\mathcal{X},\qquad\lim_{k\to+\infty}J(x,z_k)=J(x,z^0)$$

$$(\text{resp.}\qquad\forall\,z\in\mathcal{Z},\qquad\lim_{k\to+\infty}J(x_k,z)=J(x^0,z))$$

(ii) Si, de plus, J est continue, alors ses sous-différentiels partiels $\partial_x J$ et $\partial_z J$ sont fermés en (x^0,z^0).

On rappelle qu'une fonction $J:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}\cup\{+\infty,-\infty\}$ est dite *biconvexe* si ses deux fonctions partielles $x\mapsto J(x,y)$ et $y\mapsto J(x,y)$ sont convexes pour tout $x\in\mathcal{X}$ et tout $y\in\mathcal{Y}$ pour lesquels elles sont bien définies.

Démonstration : Soit $(x^0,z^0)\in\operatorname{dom}J$ et une suite $((x_k,z_k))_{k\in\mathbb{N}}$ dans $\mathcal{X}\times\mathcal{Z}$ vérifiant la relation $(\star)$ dans la définition **??**. Soit $(p_k)_{k\in\mathbb{N}}$ une suite convergente de limite p, telle que $p_k\in\partial_x J(x_k,z_k)$ pour tout $k\in\mathbb{N}$. Puisque les fonctions $x\mapsto J(x,z_k)$ sont convexes, on a

$$\forall\,x\in\mathcal{X},\qquad J(x,z_k)\geq J(x_k,z_k)+\langle p_k,x-x_k\rangle$$

Passons à la limite lorsque k tend vers $+\infty$. Le membre de gauche tend vers $J(x,z^0)$ soit par hypothèse sur la suite $(z_k)_{k\in\mathbb{N}}$, soit par continuité ; la continuité de J en (x_0,z_0) assure alors que

$$J(x,z^0)\geq J(x^0,z^0)+\langle p,x-x^0\rangle$$

ce qui prouve que $p\in\partial_x(J,x^0,z^0)$. ■

Pour aller plus loin

Sous-différentiel. Le sous-différentiel limitant défini dans ce module possède la propriété, outre de généraliser les notions de gradient et de sous-différentiel d'une fonction convexe, de préserver la règle de FERMAT, qui est une condition nécessaire d'optimalité (et suffisante dans le cas convexe), comme on le verra dans le module B1.
Les sous-gradients interviennent par ailleurs dans de nombreux contextes, qu'il s'agisse de l'opérateur proximal et les enveloppes de MOREAU (module A6), la caractérisation des points-selles d'une fonction convexe-concave non différentiable (module A8) ou encore la conjuguée convexe à travers la règle de bascule (module A9). De manière générale, les méthodes d'optimisation dites du premier ordre reposent fortement sur ce concept, puisqu'il permet de mesurer la distance des itérations aux points recherchés.

Sous-différentiels partiels. La propriété de fermeture paramétrique des sous-différentiels partiels et l'inclusion du produit cartésien des sous-différentiels partiels dans le sous-différentiel (au moins aux points critiques de la fonction) sont deux conditions cruciales pour assurer la convergence des schémas de type éclatement de variables tels que les méthodes de descentes par blocs (module B5). La seconde en particulier permet de lier les points critiques partiels et les points critiques globaux d'une fonction à plusieurs variables.