# Module A2
# Calcul sous-différentiel

Sauf mention contraire, $\mathcal{X}$ (resp. $\mathcal{Y}$) est un espace de Hilbert, muni du produit scalaire noté $\langle \cdot, \cdot \rangle$ et on note $\|\cdot\|$ la norme qui découle du produit scalaire.

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 numérique.

## 1 Sous-différentiabilité : cas convexe

### 1.1 Fonctions convexes différentiables

La convexité d'une fonction différentiable sur un espace de Hilbert peut être caractérisée par son gradient.

**Proposition 1**

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction **différentiable**. Alors on a équivalence entre les énoncés suivants :

(i) $f$ est convexe ;

(ii) $\forall\,(x_1, x_2) \in \mathcal{X}^2, \qquad f(x_2) \geq f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle$

(iii) l'opérateur $\nabla f : \mathcal{X} \to \mathcal{X}$ est *monotone*, c'est-à-dire que

$$\forall\,(x_1, x_2) \in \mathcal{X}^2, \qquad \langle \nabla f(x_2) - \nabla f(x_1), x_2 - x_1 \rangle \geq 0$$

> Pour plus de détails sur la différentiabilité sur les espaces de Hilbert, et en particulier sur la définition du gradient d'une application différentiable, le lecteur pourra consulter le module **A1 : Différentiabilité sur un espace de Hilbert. Différentiabilité de Gateaux.** (3MA261).

Démonstration : Admis. La preuve se trouve dans le module **A1 : Différentiabilité sur un espace de Hilbert. Différentiabilité de Gateaux.** (3MA261).

Dans la proposition précédente, l'énoncé (ii) peut se réécrire de manière équivalente : pour tout $x^0 \in \mathcal{X}$,

$$\forall\, x \in \mathcal{X}, \qquad f(x) \geq \langle \nabla f(x^0), x \rangle + \underbrace{f(x^0) + \langle \nabla f(x^0), x^0 \rangle}_{\text{constante}}$$

Autrement dit, la fonction $f$ est minorée par une forme affine. Ce genre de fonctions est appelé *minorante affine*.

**Définition 1** (Minorante affine)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction. On appelle *minorante affine* de $f$ toute fonction affine $g : \mathcal{X} \to \mathbb{R}$ telle que

$$\forall x \in \mathcal{X}, \qquad g(x) \leq f(x)$$

On dit par ailleurs que $g$ est *exacte* en $x^0 \in \operatorname{dom} f$ si $g(x^0) = f(x^0)$.

Géométriquement, une minorante affine de $f$ définit donc (par son graphe) un hyperplan affine situé en-dessous du graphe de la fonction de $f$. Si une minorante affine est exacte en un point, alors l'hyperplan affine touche le graphe en un point. Une minorante affine étant une fonction affine, elle est de la forme

$$g : \left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & \langle p, x \rangle + c \end{array} \right.$$

avec $p \in \mathcal{X}$ et $c \in \mathbb{R}$. Ainsi, si $g$ est exacte en $x^0$, on a nécessairement

$$f(x^0) = g(x^0) = \langle p, x^0 \rangle + c$$

de sorte que $c = f(x^0) - \langle p, x^0 \rangle$. Par conséquent,

**Proposition 2** (Forme d'une minorante affine exacte)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction et $x^0 \in \operatorname{dom} f$. Toute minorante affine de $f$ exacte en $x^0$ est de la forme

$$g : \left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & f(x^0) + \langle p, x - x^0 \rangle \end{array} \right. \qquad \text{avec } p \in \mathcal{X}.$$

Une conséquence de la proposition 1 est donc la suivante : si $f$ est une fonction convexe différentiable, alors elle admet une minorante affine. Plus précisément, elle admet une minorante affine exacte en tout point $x^0 \in \mathcal{X}$, de la forme

$$g : x \mapsto f(x^0) + \langle \nabla f(x^0), x - x^0 \rangle$$

En adaptant la preuve de la proposition 1, on peut également démontrer que si $f$ est différentiable en $a$ (et non sur tout l'espace), alors $f$ admet une minorante affine exacte en $a \in \mathcal{X}$, de la même forme (c'est-à-dire que $p = \nabla f(a)$). On va démontrer qu'il s'agit de l'unique minorante affine exacte en $a$. Pour cela, on commence par établir le résultat suivant :

**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 $\varepsilon$ 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 $p - q = 0$. ■

On peut maintenant démontrer le résultat annoncé :

**Proposition 3**

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction **convexe** différentiable en $x^0 \in \mathcal{X}$. Alors $f$ admet comme unique minorante affine exacte en $x^0$

$$\left\{ \begin{array}{ccl} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & f(x^0) + \langle \nabla f(x^0), x - x^0\rangle \end{array} \right.$$

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

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

Puisque $f$ est convexe, on a

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

Soit $g : x \mapsto f(x^0) + \langle p, x - x^0\rangle$ une minorante affine de $f$ exacte en $x^0$. On a donc

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

de sorte que, en combinant cette inégalité avec le développement limité écrit plus haut,

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

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

Ainsi, pour les fonctions **convexes différentiables** en $x^0$, le gradient $\nabla f(x^0)$ (qui définit de manière équivalente la différentielle de $f$ en $x^0$) est l'**unique** vecteur $p$ tel que

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

soit une minorante affine de $f$ exacte en $x^0$. La réciproque est vraie, comme l'affirme le résultat (admis) suivant :

**Proposition 4**

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction **convexe** et $x^0 \in \operatorname{dom} f$. On suppose que $f$ admet une unique minorante affine exacte en $x^0$

$$\left\{ \begin{array}{ccl} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & f(x^0) + \langle p, x - x^0\rangle \end{array} \right.$$

Alors $f$ est différentiable en $x^0$ et on a $\nabla f(x^0) = p$.

DÉMONSTRATION : Admis. La preuve se trouve dans le livre *Convex Analysis*, de R.T. ROCKAFELLAR (1960) et utilise l'identité entre la différentiabilité de FRÉCHET et la différentiabilité de GATEAUX dans le cas convexe (cf. **A1 : Différentiabilité sur un espace de HILBERT. Différentiabilité de GATEAUX.** (3MA261)).

## 1.2 Sous-gradients convexes, sous-différentiel convexe

On vient de voir dans le paragraphe précédent que, pour une fonction **convexe**, les minorantes affines exactes permettaient de caractériser la différentiabilité et, le cas échéant, le gradient (et donc la différentielle) de la fonction en tout point où elle est différentiable.

Qu'en est-il des points où la fonction n'est pas différentiable ? Considérons le cas de la fonction valeur absolue

$$f : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ t & \mapsto & |t| \end{array} \right.$$

qui est dérivable partout sauf en 0. Ainsi, pour tout $t^0 > 0$, elle admet une unique minorante affine exacte en $t^0$ de la forme (puisque $f'(t^0) = 1$)

$$g : t \mapsto t^0 + t - t^0 = t$$

et pour tout $t^0 < 0$, elle admet une unique minorante affine exacte en $t^0$ de la forme

$$g : t \mapsto -t^0 - (t - t^0) = -t$$

Intéressons-nous à présent au cas $t^0 = 0$. Toute minorante affine de $f$ vérifie

$$\forall\, t \in \mathbb{R}, \qquad |t| \geq \underbrace{p\, t + c}_{= g(t)}$$

Si $g$ est exacte en 0, alors $g(0) = f(0) = 0$, de sorte que $c = 0$. Ainsi, il s'agit de trouver les valeurs de $p$ telles que

$$\forall\, t \in \mathbb{R}, \qquad |t| \geq p\, t$$

Si cette inégalité est vraie, alors elle est vérifiée pour $t > 0$ ; dans ce cas, $|t| = t$ et on peut simplifier par $t$, et obtenir $1 \geq p$. Un raisonnement similaire avec $t < 0$ assure que $-1 \leq p$. Autrement dit, si $g$ est une minorante affine de $f$ exacte en 0, alors $p$ doit être compris entre $-1$ et 1. Réciproquement, si

$$-1 \leq p \leq 1$$

alors en multipliant par $t \in \mathbb{R}$ (et en changeant éventuellement le sens des inégalités), on obtient que

$$-|t| \leq p\, t \leq |t|$$

Il s'ensuit donc que $t \mapsto p\, t$ est bien une minorante affine de $f$. Finalement, on a montré que les droites $t \mapsto p\, t$ avec $p \in [\,-1\,;1\,]$ sont les uniques minorantes affines de $f$ exactes en 0 (cf. Figure 1).

Une autre propriété intéressante du gradient en optimisation **convexe différentiable** est la suivante : si $f$ est une fonction **convexe différentiable**, alors

$$\nabla f(x^*) = 0 \qquad \Longleftrightarrow \qquad \forall\, x \in \mathcal{X}, \quad f(x^*) \leq f(x)$$

(il s'agit d'une conséquence directe de la proposition 1). Autrement dit, $x^*$ est un minimiseur de $f$ (on y reviendra en détails dans le module **B1 : Méthodes d'optimisation du premier ordre**). Si l'on interprète ce résultat à l'aide des minorantes affines, on

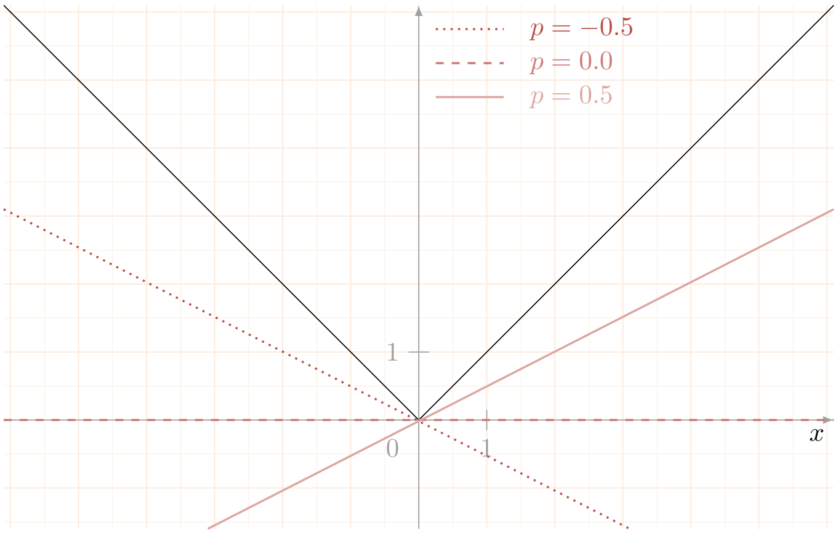


Figure 1 – Minorantes affines exactes en 0 de la valeur absolue. En noir, la courbe représentative de la valeur absolue, en rouge, des minorantes affines exactes en 0.

voit que $x^*$ est un minimiseur de $f$ si et seulement si l'unique minorante affine exacte en $x^*$ est “horizontale” (c'est-à-dire constante). Or, ce résultat reste partiellement vraie pour toute fonction, même non différentiable : $x^*$ est un minimiseur de $f$ si et seulement si $x^* \in \operatorname{dom} f$ et que

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

Autrement dit, $x^* \in \operatorname{dom} f$ est un minimiseur de $f$ si et seulement si $f$ admet en $x^*$ une minorante affine exacte constante. On peut vérifier que c'est bien le cas pour la valeur absolue (dont l'unique minimiseur est 0).

Ainsi, on voit que le vecteur $p$ dans la définition d'une minorante affine exacte caractérise à la fois la différentiabilité de la fonction au point considéré (lorsqu'il est unique, et donne le gradient le cas échéant) et l'optimalité de ce même point s'il peut s'annuler. Or, en optimisation convexe différentiable du premier ordre, la différentiabilité est principalement utilisée pour caractériser les minimiseurs. Aussi, si l'on souhaite généraliser le concept de différentiabilité à des fonctions non différentiables (mais convexes) tout en préservant l'intérêt principal de celui-ci dans le cas de l'optimisation, une possibilité intéressante est de décorréler la notion du gradient de la notion de différentiabilité, et à la place se focaliser sur son rôle dans la notion de minorante affine exacte.

**Définition 2** (Sous-différentiel et sous-gradients d'une fonction convexe)

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

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

Lorsque $x^0 \notin \operatorname{dom} f$, on pose $\partial f(x^0) = \emptyset$. Si $\partial f(x^0) \neq \emptyset$, on dit que $f$ est *sous-différentiable* en $x^0$.

Si $f$ est **convexe**, les propositions 3 et 4 assurent que $f$ est différentiable en $x^0$ si et seulement si

$$\partial f(x^0) = \{\nabla f(x^0)\}$$

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

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

sont les minorantes affines exactes en $x^0$ de la fonction $f$.

**Proposition 5** (Existence des sous-gradients)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe** et propre. On suppose qu'il existe $x^0 \in \operatorname{dom} f$ tel que $f$ soit continue en $x^0$. Alors on a

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

DÉMONSTRATION : Admis. La preuve fait appel à la notion d'hyperplan séparateur. On pourra trouver la preuve de ce résultat dans le polycopié [Szarek] par exemple.

> En dimension finie, on peut remplacer l'intérieur topologique par l'intérieur relatif. L'hypothèse de continuité y est en outre automatiquement satisfaite.

CONTRE-EXEMPLE

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

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

On peut vérifier (laissé en exercice) que $f$ est convexe, s.c.i. et propre. Elle est différentiable en tout point de l'intérieur de son domaine. En $0 \in \operatorname{dom} f$ cependant, on voit graphiquement (cf. figure 2) que toutes les droites passant par l'origine coupe la courbe représentative de $f$. Vérifions cette observation ; supposons qu'il existe $p \in \mathbb{R}$ tel que

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

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

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

Donnons quelques exemples de sous-différentiels, qui illustrent chacun 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.

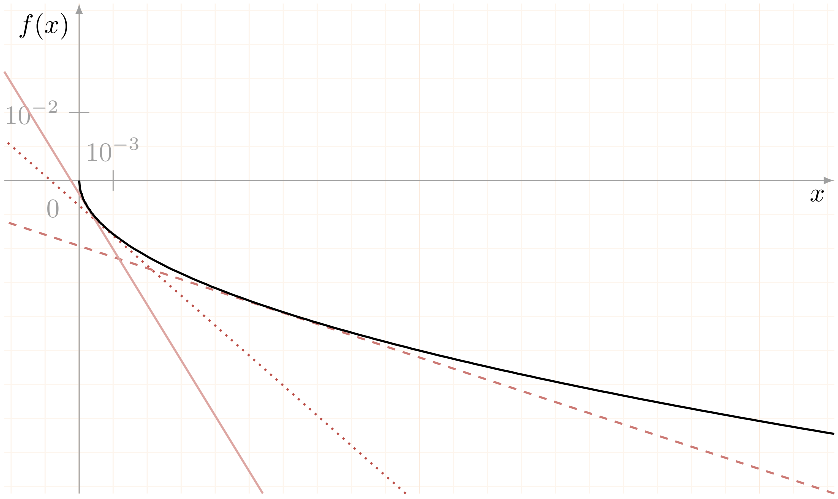


FIGURE 2 – Cas de non-existence de sous-gradient au bord du domaine. En noir, la courbe représentative de la fonction $x \mapsto -\sqrt{x}$. En rouge, différentes minorantes affines exactes de $f$.

EXEMPLE

**Sous-différentiel de la norme.** On pose

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

la norme issue du produit scalaire. 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 $f$ est différentiable en $x^0$ et on a

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

Les sous-gradients de $f$ 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

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

L'exercice suivant montre qu'une fonction peut admettre des sous-gradients en tout point de son domaine, même si celui-ci est fermé. (Attention, la résolution de cet exercice utilise des résultats qui ne seront établis que dans le module **A8 : Dualité de FENCHEL et conjuguée convexe**).

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

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

l'indicatrice de cet ensemble convexe, fermé et non vide. En utilisant la règle de bascule (module **A8 : Dualité de Fenchel et conjuguée convexe**), montrer que

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

Dans la figure 3, on illustre le cas réel, où on considère le segment $[\,0\,;1\,]$.

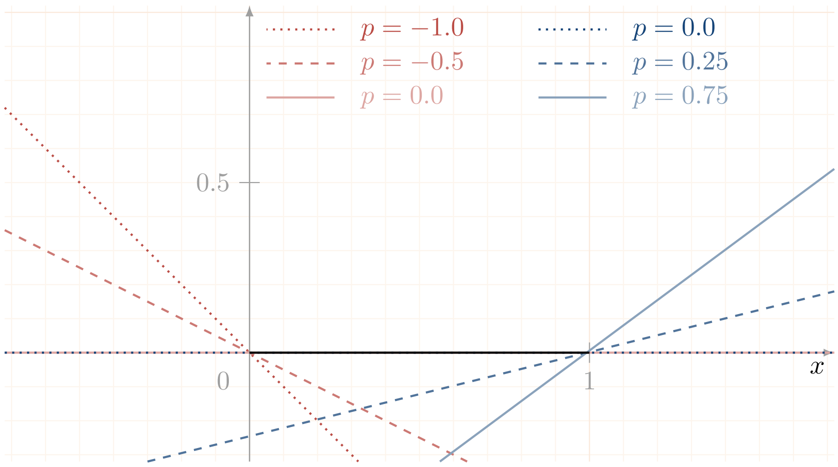


FIGURE 3 – Sous-gradients en 0 et en 1 de l'indicatrice du segment $[\,0\,;1\,]$. En noir, la courbe représentative de l'indicatrice, en rouge (resp. en bleu), des minorantes affines exactes en 0 (resp. en 1).

> 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 $\mathcal{C}$.

## 1.3 Propriétés et règles de calcul

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

**Proposition 6** (Monotonie du sous-différentiel)

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

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

On dit que l'opérateur multi-valué $\partial f$ est *monotone*. Cette notion généralise celle de la monotonie de l'opérateur $\nabla f$, qui elle-même généralise la croissance de la dérivée $f'$ dans le cas réel.

DÉMONSTRATION : Soit $(x,z) \in (\operatorname{dom} f)^2$ et soit $(p,q) \in \partial f(x) \times \partial f(z)$. 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. ■

Il est important de noter que, contrairement au cas différentiable, il n'y a pas nécessairement identité entre le sous-différentiel de la somme de deux fonctions convexes et la somme des sous-différentiels :

**Proposition 7**

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ deux fonctions **convexes** propres. Alors on a

$$\forall\, x \in \operatorname{dom} f \cap \operatorname{dom} g, \qquad \partial f(x) + \partial g(x) \subset \partial(f+g)(x)$$

DÉMONSTRATION : Commençons par remarquer que $\operatorname{dom}(f+g) = \operatorname{dom} f \cap \operatorname{dom} g$. Soit $x \in \operatorname{dom}(f+g)$ 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 7 peut être stricte, comme le montre l'exemple suivant :

EXEMPLE

**Sous-différentiel d'une somme** On considère la fonction réelle définie par

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

et la fonction $g = \chi_{]-\infty\,;0\,]}$. Ces deux fonctions sont convexes, propres et s.c.i. et on a (cf. exemple page 6) :

$$\partial f(0) = \emptyset \qquad \text{et} \qquad \partial g(0) = [\,0\,;+\infty\,[$$

Ainsi, $\partial f(0) + \partial g(0) = \emptyset$. Or, on a

$$\forall\, x \in \mathbb{R}, \qquad (f+g)(x) = \begin{cases} 0 & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases}$$

de sorte que $\partial(f+g)(0) = \,]-\infty\,;+\infty\,[$.

Le théorème d'ATTOUCH–BREZIS (**A8 : Dualité de FENCHEL**) montre que :

**Proposition 8**

*On suppose que $\mathcal{X}$ est de dimension finie.* Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ deux fonctions **convexes** s.c.i. propres telles que

$$0 \in \operatorname{intr}(\operatorname{dom} f - \operatorname{dom} g)$$

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

DÉMONSTRATION : *La preuve utilise des outils introduits dans le module* ***A8 : Dualité de Fenchel et conjuguée convexe***. Soit $x \in \mathcal{X}$ et $p \in \partial(f+g)(x)$. Puisque $f$ est convexe, s.c.i. et propre, la règle de bascule assure que $x \in \partial f^*(p)$, tandis que l'identité de LEGENDRE–FENCHEL permet d'écrire

$$f(x) + f^*(p) = \langle p, x \rangle$$

Les hypothèses sur $f$ et $g$ permettent par ailleurs d'appliquer le théorème d'ATTOUCH–BREZIS, qui implique que

$$\exists\, q \in \mathcal{X}, \qquad f^*(p) = f^*(p-q) + g^*(q)$$

de sorte que $\qquad f(x) + g(x) + f^*(p-q) + g^*(q) = \langle p-q, x \rangle + \langle q, x \rangle$

Or, d'après l'inégalité de FENCHEL–YOUNG, on a

$$g(x) + g^*(q) \geq \langle q, x \rangle$$

Par conséquent, on a nécessairement

$$f(x) + f^*(p-q) \leq \langle p-q, x \rangle$$

L'inégalité inverse est à nouveau donnée par l'inégalité de FENCHEL–YOUNG ; aussi, on a l'égalité. L'identité de LEGENDRE–FENCHEL assure donc que $p-q \in \partial f(x)$. D'une manière analogue, on démontre que $q \in \partial g(x)$. Autrement dit, puisque $p = (p-q)+q$, on en déduit que $\partial(f+g)(x) \subset \partial f(x) + \partial g(x)$. ■

REMARQUE : Dans la section suivante, on verra deux autres cas d'égalité pour le sous-différentiel de la somme de deux fonctions convexes.

**Proposition 9**

*On suppose que $\mathcal{X}$ et $\mathcal{Z}$ sont de dimension finie.* Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ 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 :

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

On suppose $\mathrm{intr}(\mathrm{dom}\, f) \cap \mathrm{Im} A$ est non vide. Alors on a

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

On rappelle que $K^*$ est l'*opérateur adjoint* de $K$, l'unique opérateur linéaire vérifiant

$$\forall\, (z, x) \in \mathcal{Z} \times \mathcal{X}, \qquad \langle x, K\, z \rangle = \langle K^* x, z \rangle$$

On note ainsi que cette règle de calcul est cohérente avec la différentiation de $f \circ A$ lorsque $f$ est différentiable.

DÉMONSTRATION : Admis.

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

## 2 Sous-différentiabilité : cas général

On va continuer notre entreprise de généralisation de la notion de différentiabilité (et de gradient) aux fonctions non différentiables, mais maintenant non convexes. L'idée centrale est de proposer une définition qui

1. coïncide avec la différentielle lorsque la fonction est différentiable ;
2. coïncide avec le sous-différentiel (convexe) lorsque la fonction est convexe ;
3. préserve la règle de FERMAT (**B1 : Méthodes d'optimisation du premier ordre**).

## 2.1 Définition

Commençons par proposer une première généralisation du sous-différentiel :

**Définition 3** (Sous-différentiel de FRÉCHET)

Soit $f : \mathcal{X} \to \overline{\mathbb{R}}$ une fonction. Soit $x \in \operatorname{dom} f$. On définit le *sous-différentiel de* FRÉCHET de $f$ au point $x^0$ comme étant l'ensemble $\hat{\partial} f(x^0)$ des vecteurs $p \in \mathcal{X}$ tels que

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

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

Comparons cette définition avec celle du sous-différentiel convexe. On rappelle que le sous-différentiel convexe n'est défini que pour une fonction convexe. Si $f$ est une fonction convexe sous-différentiable en $x^0$, alors par définition, pour tout $p \in \partial f(x^0)$, on a

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

La fonction $g$ est une minorante affine de $f$ exacte en $x^0$ : autrement dit, $g$ est une fonction affine, donc différentiable, de gradient $\nabla g(x^0) = p$ et $f(x^0) = g(x^0)$. Supposons à présent que $f$ n'est pas convexe et que $f$ est sous-différentiable (au sens de FRÉCHET) en $x^0$. Considérons la fonction suivante :

$$\tilde{g} : \left\{ \begin{array}{ccl} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & \min\big(f(x), f(x^0) + \langle p, x - x^0 \rangle\big) \end{array} \right.$$

La fonction $\tilde{g}$ vérifie $\tilde{g}(x^0) = f(x^0)$ et, par construction, est telle que

$$\forall\, x \in \mathcal{X}, \qquad f(x) \geq \tilde{g}(x)$$

Par ailleurs, on a pour tout $h \neq 0$

$$\begin{aligned} \tilde{g}(x^0 + h) &= \min\big(f(x^0 + h), f(x^0) + \langle p, h \rangle\big) \\ &= f(x^0) + \langle p, h \rangle + \|h\| \begin{cases} 0 & \text{si } f(x^0) + \langle p, h \rangle \leq f(x^0 + h) \\ \dfrac{f(x^0 + h) - f(x^0) - \langle p, h \rangle}{\|h\|} & \text{sinon} \end{cases} \end{aligned}$$

Or, si $f(x^0) + \langle p, h \rangle > f(x^0 + h)$, alors par définition du sous-différentiel de FRÉCHET, on a que

$$0 > \frac{f(x^0 + h) - f(x^0) - \langle p, h \rangle}{\|h\|} \geq \varepsilon(\|h\|)$$

avec $\varepsilon(t) \to 0$ lorsque $t \to 0$, de sorte que la fonction définie par l'accolade tend vers 0 lorsque $h$ tend vers 0. Autrement dit, on a démontré que $\tilde{g}$ est différentiable (au sens de FRÉCHET) en $x^0$, de gradient $\nabla \tilde{g}(x^0) = p$. Ainsi, on voit que la définition du sous-différentiel de FRÉCHET élargit celle du sous-différentiel convexe, en ce sens qu'elle considère des "minorantes" différentiables au lieu de minorantes affines.

**Proposition 10**

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe** propre. Alors on a l'identité
$$\forall\, x \in \mathcal{X}, \qquad \partial f(x) = \hat{\partial} f(x)$$

Démonstration : Admis. L'inclusion $\partial f(x) \subset \hat{\partial} f(x)$ découle immédiatement de la définition du sous-différentiel de Fréchet. On admet l'inclusion $\hat{\partial} f(x) \subset \partial f(x)$.

Ainsi, le sous-différentiel de Fréchet généralise le sous-différentiel convexe. Il présente l'avantage d'avoir une définition très proche de celle du sous-différentiel convexe. Cependant, elle ne possède pas des propriétés suffisamment intéressantes pour en faire une généralisation utile du sous-différentiel d'une fonction convexe, en particulier car, contrairement au sous-différentiel convexe, le sous-différentiel de Fréchet n'est pas stable par passage à la limite (or, nous verrons que l'analyse des algorithmes d'optimisation requiert souvent de tels passages à la limite). Aussi, introduit-on la définition suivante :

**Définition 4** (Sous-différentiel limitant)

Soit $f : \mathcal{X} \to \overline{\mathbb{R}}$ une fonction. Soit $x \in \operatorname{dom} f$. On définit le *sous-différentiel (limitant)* de $f$ au point $x^0$ comme étant l'ensemble $\partial f(x^0)$ des vecteurs $p \in \mathcal{X}$ tels qu'il existe $(x_k)_{k\in\mathbb{N}} \in (\operatorname{dom} f)^{\mathbb{N}}$ et $(p_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ telles que $p_k \in \hat{\partial} f(x_k)$ pour tout $k \in \mathbb{N}$ et vérifiant
$$\lim_{k\to+\infty} x_k = x^0, \qquad \lim_{k\to+\infty} f(x_k) = f(x^0) \qquad \text{et} \qquad \lim_{k\to+\infty} p_k = p$$
Lorsque $x^0 \notin \operatorname{dom} f$, on pose $\partial f(x^0) = \emptyset$.

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

**Proposition 11**

Soit $f : \mathcal{X} \to \overline{\mathbb{R}}$ une fonction. Alors on a
$$\forall\, x \in \mathcal{X}, \qquad \hat{\partial} f(x) \subset \partial f(x)$$

Démonstration : Laissé au lecteur.

Par ailleurs, la construction même du sous-différentiel limitant assure une stabilité par passage à la limite (cf. §3. Fermeture du sous-différentiel).

> 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 4) 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 2) et le sous-différentiel limitant (définition 4). La proposition suivante prouve que cette ambiguïté n'est pas gênante :

**Proposition 12**

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe** propre. Alors les deux ensembles définis dans les définitions 2 et 4 coïncident.

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

Démonstration : Pour lever temporairement l'ambiguïté de notation, on va noter $\partial^{\text{convexe}} f(x)$ l'ensemble des sous-gradients tels que définis par la définition 2 (sous-différentiel convexe) et $\partial^{\text{limitant}} f(x)$ l'ensemble des sous-gradients tels que définis par la définition 4 (sous-différentiel limitant). On vient de montrer que

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

Soit $p \in \partial^{\text{limitant}} f(x)$. Il existe par définition deux suites $(x_k)_{k\in\mathbb{N}}$ et $(p_k)_{k\in\mathbb{N}}$ telles que $p_k \in \hat{\partial} f(x_k)$ pour tout $k \in \mathbb{N}$ et

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

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

puisque $\partial^{\text{convexe}} f(x) = \hat{\partial} f(x)$. 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 13**

Soit $f : \mathcal{X} \to \overline{\mathbb{R}}$ une fonction. Soit $x^0 \in \operatorname{dom} f$.

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

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

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

Démonstration : Démontrons successivement les deux énoncés.

- **Énoncé (i).** Puisque $f$ est différentiable en $x^0$, on a par définition que
$$f(x) = f(x^0) + \langle \nabla f(x^0), x - x^0\rangle + o(\|x - x^0\|)$$
ce qui assure que $\nabla f(x^0) \in \hat{\partial} f(x^0)$. Soit $p \in \hat{\partial} f(x^0)$. On a donc
$$f(x^0) + \langle \nabla f(x^0), x - x^0\rangle + o(\|x - x^0\|) \geq f(x^0) + \langle p, x - x^0\rangle + o(\|x - x^0\|)$$
ce qui donne après simplification
$$\langle p - \nabla f(x^0), x - x^0\rangle \leq o(\|x - x^0\|)$$
Le lemme 1 démontre que $p = \nabla f(x^0)$.
- **Énoncé (ii).** Soit $p \in \partial f(x^0)$. Par définition du sous-différentiel, il existe deux suites $(x_k)_{k\in\mathbb{N}} \in (\operatorname{dom} f)^{\mathbb{N}}$ et $(p_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ vérifiant
$$\lim_{k\to+\infty} x_k = x^0, \qquad \lim_{k\to+\infty} f(x_k) = f(x^0) \qquad \text{et} \qquad \lim_{k\to+\infty} p_k = p$$

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

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

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

par continuité de $\nabla f$ sur $\mathcal{V}$. ■

EXEMPLE

**Sous-différentiel d'une fonction concave non différentiable.** On considère la fonction concave $f = -|\cdot|$. Puisque $f$ est de classe $\mathcal{C}^1$ sur $\mathbb{R} \setminus \{0\}$, on en déduit que

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

Aussi, il s'ensuit immédiatement que $\partial f(0) = \{1, -1\}$. Par ailleurs, $\hat{\partial} f(0) = \emptyset$. 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.

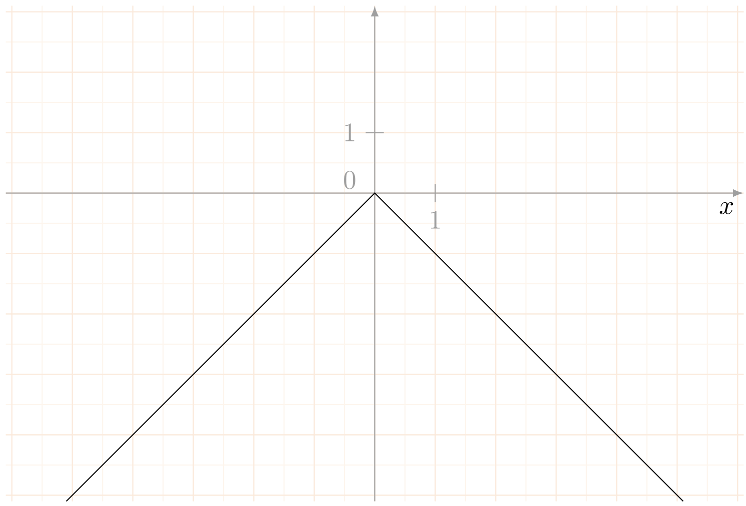


FIGURE 4 – Courbe représentative de l'opposé de la valeur absolue (en noir).

## 2.2 Règles de calcul sous-différentiel

**Proposition 14**

Soit $f : \mathcal{X} \to \overline{\mathbb{R}}$ une fonction et soit $\alpha > 0$. Alors on a

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

Démonstration : Il suffit de remarquer que

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

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

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

**Proposition 15**

Soit $f : \mathcal{X} \to \overline{\mathbb{R}}$ une fonction et soit $\alpha < 0$. Alors on a

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

Démonstration : Laissé au lecteur.

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 16**

Soit $f : \mathcal{X} \to \overline{\mathbb{R}}$ et $g : \mathcal{X} \to \overline{\mathbb{R}}$ deux fonctions telles que $\mathrm{dom}(f+g) \neq \emptyset$. Soit $x^0 \in \mathrm{dom}(f+g)$. 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 $\mathrm{dom}\, f$ 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 $\nabla g$ 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. ■

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

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

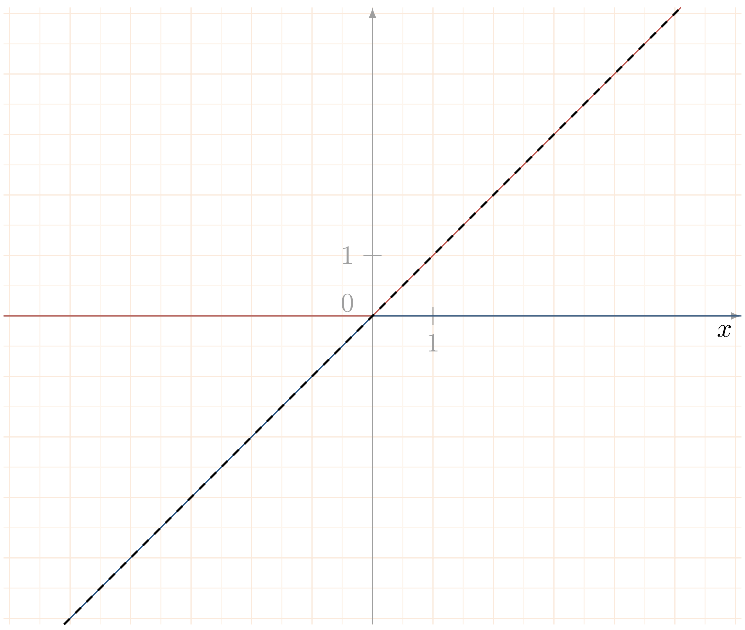


Figure 5 – En noir, la courbe représentative de la fonction $f = f + g$ avec $f : x \mapsto \max\{0, x\}$ (en rouge) et $g : x \mapsto \min\{0, x\}$ (en bleu).

La proposition 16 permet de démontrer une relation utile dans le cadre des fonctions fortement convexes :

**Corollaire 1**

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **fortement convexe**, de module $\alpha$, propre. On suppose que $f$ est sous-différentiable en $x \in \operatorname{dom} f$ avec $p \in \partial f(x)$. Alors on a

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

Démonstration : La forte convexité de $f$ équivalant à la convexité de la fonction

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

(d'après la proposition 13 du module **A1 : Éléments de topologie**), on utilise la proposition précédente pour démontrer que

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

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

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

On obtient après simplification le résultat souhaité. ■

**Proposition 17** (Somme séparable)

Soit $f : \mathcal{X} \to \overline{\mathbb{R}}$ et $g : \mathcal{Z} \to \overline{\mathbb{R}}$ deux fonctions s.c.i. telles que le domaine de la fonction

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

soit non vide. Alors on a

$$\forall\, (x, z) \in \operatorname{dom} J, \qquad \partial J(x, z) = \partial f(x) \times \partial g(z)$$

On dit que $f$ 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 $(p, q) \in \hat{\partial} J(x^0, y^0)$, 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.

- **Inclusion $\partial f(x) \times \partial g(z) \subset \partial J(x, z)$.** Soit $(x, z) \in \operatorname{dom} J$ et $p \in \partial f(x)$ et $q \in \partial g(z)$. Alors par définition, 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 implique

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

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 $\hat{\partial} f(x_k) \times \hat{\partial} g(z_k) = \hat{\partial} J(x_k, z_k)$, on en déduit que $(p, q) \in \partial J(x, z)$.

- **Inclusion $\partial f(x) \times \partial g(z) \supset \partial J(x, z)$.** Soit $(x, z) \in \operatorname{dom} J$ et $(p, q) \in \partial J(x, z)$. Alors par définition, 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

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

et d'autre part que, pour tout $k \in \mathbb{N}$,

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

Remarquons que, puisque $f$ et $g$ sont s.c.i.,

$$\begin{aligned} f(x) + g(z) = \lim_{k\to+\infty} f(x_k) + g(z_k) &= \liminf_{k\to+\infty} f(x_k) + g(z_k) \\ &\geq \liminf_{k\to+\infty} f(x_k) + \liminf_{k\to+\infty} g(z_k) \geq f(x) + g(z) \end{aligned}$$

Ainsi, les inégalités ci-dessus sont des égalités, et en particulier,

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

On en déduit alors que $p \in \partial f(x)$ et $q \in \partial g(z)$. ■

## 2.3 Sous-différentiels partiels

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

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

où les $\mathcal{X}_i$ sont des espaces de Hilbert. Pour plus de lisibilité, on se restreint au cas

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

avec $\mathcal{Z}$ un espace de Hilbert.

**Définition 5** (Sous-différentiels partiels)

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

$$\partial_x f(x^0, z^0) \qquad (\text{resp. } \partial_z f(x^0, z^0))$$

le sous-différentiel de la fonction partielle

$$x \mapsto f(x, z^0) \text{ au point } x^0 \qquad (\text{resp. } z \mapsto f(x^0, z) \text{ au point } z^0)$$

Autrement dit, $p \in \partial_x f(x^0, z^0)$ 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

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

et d'autre part

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

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

**Proposition 18**

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

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

Démonstration : Il suffit d'écrire la définition des sous-différentiels partiels et d'appliquer la proposition 13. ■

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

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

il s'ensuit naturellement que

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

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 :

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

On peut vérifier que $f$ et ses fonctions partielles sont convexes et continues, de sorte que leurs sous-différentiels sont définies comme dans la définition 2. Par ailleurs, $f$ non différentiable en tout point $(a, a)$ avec $a \in \mathbb{R}$. Soit $a \neq 0$. Il est aisé de vérifier que

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

et que

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

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

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

Montrons que $(0,0) \neq \partial f(a,a)$. 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

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

> On voit donc que, dans le cas non différentiable, l'implication
>
> $$(0,0) \in \partial_x f(x,z) \times \partial_z f(x,z) \qquad \Longrightarrow \qquad (0,0) \in \partial f(x,z)$$
>
> est parfois une **hypothèse** cruciale en optimisation du premier ordre.

**Proposition 19**

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

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

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 13, 16 et 17. ■

# 3 Fermeture du sous-différentiel

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

**Proposition 20** (Fermeture du sous-différentiel)

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

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

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

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

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

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 $\operatorname{dom} f$ et $\mathcal{X}$ respectivement, telles que

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

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

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

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

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

Lorsque $J$ est de la forme $$J : \left\{ \begin{array}{ccc} \mathcal{X}\times\mathcal{Z} & \to & \overline{\mathbb{R}} \\ (x,z) & \mapsto & f(x,z) \end{array} \right.$$

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 $\partial J$ par les sous-différentiels **partiels** $\partial_x J$ et $\partial_z J$. Or, pour l'analyse de certains schémas d'optimisation, la propriété de fermeture du sous-différentiel n'est pas toujours suffisante, et l'on a besoin d'une propriété analogue pour les sous-différentiels partiels. Aussi introduit-on les définitions suivantes :

**Définition 6** (Fermeture paramétrique des sous-différentiels partiels)

Soit $J : \mathcal{X}\times\mathcal{Z} \to \overline{\mathbb{R}}$ 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 continûment 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

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

Calculons $\partial_x f(0,0)$. 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 $\partial_x f$ n'est pas fermé en $(0,0)$.

**Proposition 21**

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 22**

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

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

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 vers $p$, telle que $p_k \in \partial_x J(x_k, z_k)$ pour tout $k \in \mathbb{N}$. D'après la proposition 16, 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 $\nabla h$, la suite des $q_k$ converge vers

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

La fermeture du sous-différentiel de $J$ 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 $(x^0, z^0)$. ■

On termine ce module en considérant le cas biconvexe :

**Proposition 23**

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

(i) Le sous-différentiel partiel $\partial_x J$ (resp. $\partial_z J$) de $J$ 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 $\partial_z f$ sont fermés en $(x^0, z^0)$.

On rappelle qu'une fonction $J : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ 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 6. 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 : Méthodes d'optimisation du premier ordre**.
Les sous-gradients interviennent par ailleurs dans de nombreux contextes, qu'il s'agisse de l'opérateur proximal et les enveloppes de Moreau (module **A5 : Opérateur proximal de Moreau**), la caractérisation des points-selles d'une fonction convexe-concave non différentiable (module **A6 : Dualité min-max**) ou encore la conjuguée convexe à travers la règle de bascule (module **A8 : Dualité de Fenchel et conjuguée convexe**). 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 **B6 : Éclatement de variables**). La seconde en particulier permet de lier les points critiques partiels et les points critiques globaux d'une fonction à plusieurs variables.