كلية العلوم و التقنيات فاس
Faculté des Sciences et Techniques de Fès

جامعة سيدي محمد بن عبد الله
Université Sidi Mohamed Ben Abdellah

Master Mathématique et Application au Calcul Scientifique

(MACS)

MEMOIRE DE FIN D'ETUDES

Pour l'obtention du Diplôme de Master Sciences et Techniques (MST)

Optimisation non différentiable

♦ ***Réalisé par :*** ***OUHABI Ayoub***

♦ ***Encadré par :*** ***Pr. Abdelmajid HILALI***
Pr. Ahmed EL HILALI ALAOUI

Soutenu le 18 Juin 2019

Devant le jury composé de:

Anisse OUADGHIRI	**Faculté des Sciences et Techniques Fès**	**Président**
Ghizlane CHAIBI	**Faculté des Sciences et Techniques Fès**	**Examinateur**
Mohammed EL KHOMSSI	**Faculté des Sciences et Techniques Fès**	**Examinateur**
Abdelmajid HILALI	**Faculté des Sciences et Techniques Fès**	**Encadrant**
Ahmed EL HILALI ALAOUI	**Faculté des Sciences et Techniques Fès**	**Encadrant**

Année Universitaire 2018 / 2019

Résumé

On s'intéresse dans ce document à l'analyse non différentiable et ses applications en optimisation à travers l'étude de sous-différentiel au sens d'analyse convexe et au sens de Clarke. Des règles de calcul ont été développé dans des contextes différents à l'aide des sous-différentiel introduit par Rockafellar et Clarke et plus précisément avec des données localement lipschitziennes ou convexes.

Mots clés :

Fonctions convexes ; Sous différentiel au sens d'analyse convexe ; Sous différentiel au sens de Clarke ; Condition d'optimalité ; Cône tangent ; Cône normale ; Calcule sous-différentiel ; Problème d'optimisation.

Dédicace

Au nom de Dieu Le Clément et Le Miséricordieux Louange à ALLAH le Tout-puissant

A mes très chers parents

A ma soeur Hajar et à mes frères Hafid et Mohamed

A toutes ma familles

A touts(es) mes meilleurs ami(e)s

A tous ceux qui me sont chers

Toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail.

Table des matières

Table des figures

Remerciement

En tout premier lieu, Je remercie le bon Dieu, tout puissant, de m'avoir donné la force pour survivre, ainsi que l'audace pour dépasser toutes les difficultés.

Ma plus grande gratitude va à mes encadrant Mr Ahmed ELHILALI ALAOUI et Mr Abdelmajid HILALI pour leurs disponibilité permanente, la confiance qu'ils m'ont accordée, et ses précieux conseils qui m'ont permis de mener à bien ce travail.

Je remercie également les membres de jury qui ont accepté d'évaluer mon travail.

Afin de n'oublier personne, mes vifs remerciements s'adressent à tous ceux qui m'ont aidé à la réalisation de ce travail.

Merci infiniment à toutes et à tous.

Notations principales

c.f	confère.
i.e	c'est à dire.
ssi	si et seulement si.
CNS	conditions nécessaires et suffisantes.
B **et** B^*	boules unités de $\mathbb{R}^n$ et $(\mathbb{R}^n)^*$ (respectivement en abrégé resp).
epi(f**)**	épigraphe de la fonction f.
s.c.i/s.c.s	semi continuité inférieure/supérieure.
$\overline{\mathbb{R}}$	$= \mathbb{R} \cup +\infty$.
$\langle , \rangle$	crochet de dualité.
$\|.\|$	norme sur l'espace $\mathbb{R}^n$.
$(\mathbb{R}^n)^*$	dual topologique de $\mathbb{R}^n$.
$x \rightharpoonup x_0$	convergence faible étoile.
$x_k \to x$	$(x)_{k \geqslant 1}$ tend vers x quand k tend vers $+\infty$.
$d_S(x)$	distance de x à S.
$\nabla f(x)$	gradient de f en x.
$f'(x)$	dérivée de f en x.
$f'_d(x)$	dérivée à droite de f en x.
$f'_g(x)$	dérivée à gauche de f en x.
$f'(x; d)$	dérivée directionnelle de f.
$f'_G(x)$	dérivée de Gâteaux de f.
$f'_F(x)$	dérivée de Fréchet de f.
A^*	adjoint d'un opérateur linéaire continue A.
i_Ω	fonction indicatrice de Ω.
S_α	ensemble de niveau α.

$int\Omega$ intérieur de Ω .

$cl\Omega$ adhérence de Ω.

$co\Omega$ enveloppe convexe d'un sous-ensemble Ω.

$\overline{Conv}(\mathbb{R}^n)$ ensemble des fonctions $f : \mathbb{R}^n \to R$ convexes fermées propres.

$\phi : \mathbb{R}^n \rightleftarrows \mathbb{R}^n$ multifonction de $\mathbb{R}^n$ dans $\mathbb{R}^n$.

Imf image de f.

f^{-1} image réciproque de f.

$Dom(\phi)$ domaine de multifonction ϕ.

$Ker(f)$ noyau de f.

$gph\phi$ graphe de multifonction ϕ.

$\limsup\limits_{x \to x_0} = \{y \in \mathbb{R}^n \mid \exists x_k \to x_0 \ et \ \exists y_k \to y \ tel que \ y_k \in f(x_k) \ pour \ tou \ k \in \mathbb{N}\}.$

$\liminf\limits_{x \to x_0} = \{y \in \mathbb{R}^n \mid \forall x_k \to x_0 \, et \, \exists y_k \in f(x_k) \, avec \, k \in \mathbb{N} \, tel \, que \, y_k \to y, k \to +\infty\}.$

Introduction

Le concept de la différentiabilité joue un rôle crucial dans l'étude des problèmes d'optimisation. Les origines de l'analyse non lisse remontent au début des années 70 quand les théoriciens du contrôle et de la programmation non-linéaire ont essayé d'établir des conditions nécessaires d'optimalité pour des problèmes avec des données non différentielles ou avec des fonctions non différentiable. Beaucoup d'efforts ont été consacrés à développer cette nouvelle direction de l'analyse. Parmi les plus importantes contributions, nous pouvons citer les travaux d'**Aubin**, **Borwein**, **Clarke**, **Ioffe**, **Rockafellar**, **Thibault**, etc. ainsi que les nombreuses références contenues dans les monographies récentes de Rockafellar-wets [18], Mordukhovich [15], pour se rendre compte du développement ainsi que des nombreuses applications de l'analyse non différentiable. En particulier, les nombreuses variantes de dérivées généralisées étudiées par ces auteurs permettent de traiter des applications dans des domaines très variés.

Dans le but de résoudre ce genre de problèmes, divers concepts de dérivées généralisées ont été proposés pour remplacer la dérivée. Le but était de définir la dérivée généralisée pour chaque point dans le domaine d'une fonction appartenant à une classe particulière telle les fonctions convexes ou bien les fonctions localement lipschitziennes ou bien les fonctions D.C (Différence de deux fonctions convexes). Le premier sous-différentiel était introduit par **Rockafellar** [17]. Après, **Clarke** a introduit la dérivé généralisé et le sous-différentiel au sens de Clarke [9], etc. Même si les dérivées généralisées sont très utiles dans l'étude des problèmes non différentiables, leurs définitions sont compliquées et elles sont souvent difficiles à calculer, ce qui a poussé à penser aux concepts géométriques.

Le concept géométrique des vecteurs perpendiculaires à un ensemble était utilisé par **Clarke** [9], tandis que **Hirriart-Urruty** [5] était le premier à montrer comment obtenir une formule explicite pour le cône tangent convexe correspondant. La construction de dérivées généralisées possibles à valeurs non convexes a été développée par Mordukhovich et cela en utilisant ces vecteurs normaux. Le sous-différentiel proximal était introduit par Mordukhovich [MN] mais il a été présenté explicitement par **Rockafellar** [17] où les caractérisations du sous-gradient généralisé de Clarke et le cône normal de Clarke ont été donnés.

Quand une fonction f atteint un minimum en un point x^* , on a alors $0 \in \partial f(x^*)$ et ceci pour toute notion de sous-différentiel. Toute fois on conçoit bien que cette notion n'est vraiment intéressante que si l'on peut exprimer ou estimer le sous-différentiel d'une somme ou d'une composition, c'est à dire si l'on peut établir de bonnes règles de calcul sous-différentiel pour cette notion.

Ce document de synthèse s'articule autour de l'analyse non différentiable et de ses applications en optimisation. En effet, dans un premier chapitre, on fait un certain nombre de rappels sur l'espace de Hilbert $\mathbb{R}^n$ et aussi une étude générale sur les fonctions convexes définies sur $\mathbb{R}^n$, en fin la notion de la différentiabilité au sens de Fréchet, Gâteau et Hadamard.

Dans le deuxième chapitre, on va étudié le sous-différentiel au sens d'analyse convexe. Ce sous-différentiel introduit par **Rockafellar**, comme la pente des minorantes affines de la fonction f en un point x. Ce chapitre apporte aussi des règles de calculs pour le sous-différentiel dans des contextes différents de calcul (pour la somme et la composition par exemple), en terminera par des applications à l'optimisation.

Dans le troisième chapitre, nous nous intéressons à l'étude de sous-différentiel généralisé introduit par **Clarke** [9], comme généralisation de sous différentiel au sens d'analyse convexe. Dans ce chapitre nous avons développé des règles de calcul de sous-différentiel avec des données localement lipschitziennes, aussi des applications à l'optimisation mais cette fois la fonction f est une fonction D.C (différence de deux fonctions convexes). A notre connaissance, ce qui est nouveau c'est le cas où les données du problème sont des fonctions D.C. En conclusion, un problème ouvert est donné.

CHAPITRE 1

ENSEMBLES ET FONCTIONS CONVEXES

Ce chapitre présente les définitions, exemples et propriétés de base des ensembles et fonctions convexes dans l'espace euclidien $\mathbb{R}^n$.

1.1 Préliminaire

Nous commençons par examiner les notions classiques et les propriétés de l'espace euclidien $\mathbb{R}^n$. Les preuves des résultats présentés dans cette section se trouvent dans des ouvrages standard sur le calcul avancé et algèbre linéaire.

Notons $\mathbb{R}^n$ l'ensemble des n-uplets des nombres réels $x = (x_1, ..., x_n)$. Alors $\mathbb{R}^n$ est un espace vectoriel avec les deux lois suivantes :

$$(x_1, ..., x_n) + (y_1, ..., y_n) = (x_1 + y_1, ..., x_n + y_n),$$

$$\lambda(x_1, ..., x_n) = (\lambda x_1, ..., \lambda x_n),$$

où $(x_1, ..., x_n), (y_1, ..., y_n) \in \mathbb{R}^n$et $\lambda \in \mathbb{R}$. L'élément zéro de $\mathbb{R}^n$ et le nombre zéro de $\mathbb{R}$ sont souvent désignés par la même notation 0 si aucune confusion ne survient.

Pour tout $x = (x_1, ..., x_n) \in \mathbb{R}^n$, on l'identifie avec le vecteur colonne $x = (x_1, ..., x_n)^T$ où le symbole "T" représente le transpose de vecteur. Étant donné $x = (x_1, ..., x_n) \in \mathbb{R}^n$ et $y = (y_1, ..., y_n) \in \mathbb{R}^n$, le produit scalaire de x et y est défini par :

$$< x, y >:= \sum_{i=1}^{n} x_i y_i.$$

La proposition suivante donne quelques propriétés importantes du produit scalaire dans $\mathbb{R}^n$.

Proposition 1.1.

Pour $x,y,z \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$,on a :

1. $< x, x >\geqslant 0$, et $< x, x >= 0$ si et seulement si $x = 0$.
2. $< x, y >=< y, x >$.
3. $< \lambda x, y >= \lambda < x, y >=< x, \lambda y >$.
4. $< x, y + z >=< x, y > + < x, z >$.

La norme euclidienne de $x = (x_1, ..., x_n) \in \mathbb{R}^n$ est définie par

$$\| x \| := \sqrt{x_1^2 + ... + x_n^2} = \sqrt{< x, x >}.$$

Proposition 1.2.

Pour $x,y,z \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$,on a :

1. $\| x \|\geqslant 0$, et $\| x \|= 0$ si et seulement si $x = 0$.
2. $\| \lambda x \|=| \lambda |\| x \|$.
3. $\| x + y \|\leqslant\| x \| + \| y \|$ (Inégalité triangulaire).
4. $|< x, y >|\leqslant\| x \|\| y \|$ (Inégalité de Cauchy Schwarz).

L'utilisation de la norme euclidienne nous permet d'introduire les boules dans $\mathbb{R}^n$, qui peuvent être utilisées pour définir d'autres notions topologiques dans $\mathbb{R}^n$.

Définition 1.3.

La boule fermée centré en $x_0 \in \mathbb{R}^n$ avec le rayon $r \geqslant 0$ et la boule fermée unité de $\mathbb{R}^n$ est définie respectivement par :

$$\overline{B}(x_0; r) := \{x \in \mathbb{R}^n \mid \|x - x_0\|\leqslant r\} \quad et \quad \overline{B} := \{x \in \mathbb{R}^n \mid \|x\|\leqslant 1\}.$$

La boule ouverte centré en $x_0 \in \mathbb{R}^n$ avec le rayon $r \geqslant 0$ et la boule ouverte unité de $\mathbb{R}^n$ est définie respectivement par :

$$B(x_0; r) := \{x \in \mathbb{R}^n \mid \|x - x_0\|< r\} \quad et \quad B := \{x \in \mathbb{R}^n \mid \|x\|< 1\}.$$

Pour Ω ,Ω_1 et Ω_2 des sous-ensembles de $\mathbb{R}^n$ et pour $\lambda \in \mathbb{R}$, on définis les opérations suivantes :

$$\Omega_1 + \Omega_2 := \{x + y \mid x \in \Omega_1, y \in \Omega_2\}, \quad \lambda\Omega := \{\lambda x \mid x \in \Omega\}.$$

La proposition suivante peut être prouvée facilement.

Proposition 1.4.

Soit Ω_1 et Ω_2 deux sous-ensemble de $\mathbb{R}^n$.
(i) Si Ω_1 est un ouvert ou Ω_2 est un ouvert, alors $\Omega_1 + \Omega_2$ est un ouvert.
(ii) Si Ω_1 est un fermé et Ω_2 est un compact, alors $\Omega_1 + \Omega_2$ est un fermé.

Il est facile de voir que $B = B(0; 1)$ et $B(x_0; r) = x_0 + rB$.

Définition 1.5.

Soit $\Omega \subset \mathbb{R}^n$. Alors x_0 est un point intérieur de Ω s'il existe $\delta > 0$ tel que

$$B(x_0; \delta) \subset \Omega.$$

l'ensemble de tous les points intérieurs de Ω est désigné par $int\Omega$.
De plus, on dit que Ω est ouvert si tout point de Ω est un point intérieur de Ω.

Nous obtenons que Ω est un ouvert si et seulement si pour tout $x_0 \in \Omega$ il existe $\delta > 0$ tel que $B(x_0; \delta) \subset \Omega$. Il est évident de voir que ensemble vide $\emptyset$ et l'espace $\mathbb{R}^n$ sont ouverts dans $\mathbb{R}^n$. Par ailleurs, toute boule ouverte est un ouvert.

Définition 1.6.

Un ensemble $\Omega \subset \mathbb{R}^n$ est un fermé si son complémentaire $\Omega^c = \mathbb{R}^n \setminus \Omega$ est un ouvert de $\mathbb{R}^n$.

Remarque 1.6.1. *Il en résulte que l' ensemble vide $\emptyset$ et l'espace $\mathbb{R}^n$ sont des fermés dans $\mathbb{R}^n$.*

Proposition 1.7.

(i) La réunion de toute familles d'ensembles ouverts dans $\mathbb{R}^n$ est ouvert.
(ii) L'intersection finie de toute familles d'ensembles ouverts dans $\mathbb{R}^n$ est ouvert .
(iii) La réunion finie de toute familles d'ensembles fermés dans $\mathbb{R}^n$ est fermé .
(iv) L'intersection de toute familles d'ensembles fermés dans $\mathbb{R}^n$ est fermé.

Définition 1.8.

Soit $\{x_k\}$ une suite dans $\mathbb{R}^n$.On dit que $\{x_k\}$ converge vers $\bar{x}$ si $\lim\limits_{k\to\infty} \|x_k - \bar{x}\| = 0$. Dans ce cas on écrit

$$\lim_{k\to\infty} x_k = \bar{x}.$$

Cette notion nous permet de définir les concepts topologiques importants suivants pour les ensembles.

Définition 1.9.

Soit Ω un sous-ensemble non vide de $\mathbb{R}^n$. Alors :
(i) La fermeture de Ω, noté par $\overline{\Omega}$ ou $cl\Omega$, est l'ensemble des limites de toutes les suites convergentes appartenant à Ω.
(ii) La frontière de Ω noté par $Fr\Omega$, est l'ensemble $\overline{\Omega} \setminus int\Omega$.

Nous pouvons voir que la fermeture de Ω est l'intersection de tous les ensembles fermés contenant Ω et que l'intérieur de Ω est l'union de tous les ensembles ouverts contenus dans Ω. Il découle de la définition que $x_0 \in \overline{\Omega}$ si et seulement si pour tout $\delta > 0$ on a $B(x_0;\delta) \cap \Omega \neq \emptyset$. Par ailleurs $x_0 \in Fr\Omega$ si et seulement si pour tout $\delta > 0$ la boule fermée rencontre les deux ensembles Ω et son complémentaire Ω^c.

Définition 1.10.

Soit $\{x_k\}$ une suite dans $\mathbb{R}^n$ et soit $\{k_l\}$ une suite strictement croissante d'entiers positifs. Alors la nouvelle suite $\{x_{k_l}\}$ est appelé sous suite de $\{x_k\}$.

On dit qu'un ensemble Ω est borné s'il est contenu dans une boule centrée à l'origine de rayon $r > 0$, c-à-d, $\Omega \in B(0; r)$. Ainsi, une suite $\{x_k\}$ est bornée s'il existe $r > 0$ avec

$$\|x_k\| \leqslant r \quad \forall k \in \mathbb{N}.$$

Le résultat important suivant est connu sous le nom de théorème de **Balzano-Weierstrass**.

Théoreme 1.11 ([5]).

Toute suite bornée dans $\mathbb{R}^n$ admet une sous-suite convergente.

Le résultat suivant joue un rôle très important dans l'analyse et l'optimisation.

Définition 1.12.

On dit qu'un ensemble est compact dans $\mathbb{R}^n$ si pour toute suite dans Ω on peut extraire une sous-suite convergente vers un point de Ω.

Le résultat suivant est une conséquence du théorème de Bolzano-Weierstrass.

Théoreme 1.13.

Un sous-ensemble Ω de $\mathbb{R}^n$ est compact si et seulement si il est fermé et borné.

Rappelons maintenant les notions de bornes pour les sous-ensembles de la droite réelle.

Définition 1.14.

Soit D un sous-ensemble de la droite réelle. le nombre $m \in \mathbb{R}$ (resp $M \in \mathbb{R}$) est un minorant (resp majorant) de D si on a

$$x \geqslant m \quad (resp \quad x \leqslant M) \quad \forall x \in D,$$

De plus, on dit que l'ensemble D est borné s'il est simultanément minoré et majoré.

Définition 1.15.

Si l'ensemble des majorants d'une partie D de $\mathbb{R}$ admet un plus petit élément M on dit que M est la borne supérieure de D et on note $M = sup(D)$. Cette borne est alors unique.

Si l'ensemble des minorants d'une partie D de $\mathbb{R}$ admet un plus grand élément m, on dit que m est la borne inférieure de D et on note $m = inf(D)$. Cette borne est alors unique

Définition 1.16.

Soient $f : \Omega \to \overline{\mathbb{R}}$ une fonction à valeurs réelles et $x_0 \in \Omega$ avec $f(x_0) < \infty$. Alors f est continue en x_0 si pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que

$$\forall x \in \Omega, \|x - x_0\| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon.$$

Théoreme 1.17.

On dit que f est continue sur Ω s'il est continue en tout point de Ω.

Autrement f est continue en x_0 (avec $f(x_0) < \infty$) si et seulement si pour toute suite $\{x_k\}$ dans Ω converge vers x_0 la suite $\{f(x_k)\}$ converge vers $f(x_0)$.

Définition 1.18.

f est dite semi-continue inférieurement $(s.c.i)$ sur $\mathbb{R}^n$ si :

$$\forall x \in \mathbb{R}^n, \quad f(x) \leqslant \lim_{y \longrightarrow x} inf(f(y)).$$

f est dite semi-continue supérieurement $(s.c.s)$ sur $\mathbb{R}^n$. Si $(-f)$ est semi-continue inférieurement $(s.c.i)$ sur $\mathbb{R}^n$.

Définition 1.19.

Soit $\alpha \in \mathbb{R}$, on appelle sous niveau de f, l'ensemble :

$$\mathbf{S}_\alpha = \{\mathbf{x} \in \mathbb{R}^\mathbf{p} : \mathbf{f}(\mathbf{x}) \leqslant \alpha\}$$

Proposition 1.20.

Les propriétés suivantes sont équivalentes :

1. f est semi-continue inférieurement en tout point de $\mathbb{R}^p$.
2. pour tout réel $\alpha \in \mathbb{R}$, S_α est fermé .

Définition 1.21.

Soit $f : \Omega \to \overline{\mathbb{R}}$ et soit $x_0 \in \Omega$ avec $f(x_0) < \infty$. On dit que f admet un minimum local en $x_0 \in \Omega$ s'il existe $\delta > 0$ tel que

$$f(x) \geqslant f(x_0) \quad pour\ tout\ x \in B(x_0;\delta) \cap \Omega.$$

On dit que f a un minimum global en $x_0 \in \Omega$ si

$$f(x) \geqslant f(x_0) \quad pour\ tout\ x \in \Omega.$$

Les notions de maximum local et global peuvent être définies de la même manière.

Enfin, dans cette section, nous formulons un résultat fondamental de l'analyse mathématique et optimisation connue sous le nom de théorème d'existence de Weierstrass.

Théoreme 1.22.

Soit $f : \Omega \to \overline{\mathbb{R}}$ une fonction continue, où Ω un sous-ensemble non vide, compact de $\mathbb{R}^n$. Alors il existe $\bar{x} \in \Omega$ et $\bar{u} \in \Omega$ tel que

$$f(\bar{x}) = \inf\{f(x) \mid x \in \Omega\} \quad et \quad f(\bar{u}) = \sup\{f(x) \mid x \in \Omega\}.$$

1.2 Ensembles convexes

Dans ce paragraphe et ce qui suivra, on se limite à quelques faits fondamentaux de l'analyse convexe qui nous sont indispensables pour la suite du mémoire. De bonnes références pour aller plus loin sur ce sujet sont les ouvrages [1] et [5], ainsi que l'un des ouvrages "fondateurs", celui de **Reckafellar** [17].

Nous commençons l'étude de la convexité des ensembles, puis passons aux fonctions.

Étant donné deux éléments a et b dans $\mathbb{R}^n$, définir le segment

$$[a,b] := \{\lambda a + (1-\lambda)b \mid \lambda \in [0,1]\}.$$

Notez que si $a = b$, alors ce segment se réduit à un singleton $[a,b] = \{a\}$.

Définition 1.23.

Un sous ensemble Ω de $\mathbb{R}^n$ est convexe si $[a,b] \subset \Omega$ pour tout $a,b \in \Omega$. D'une manière équivalente Ω est convexe si $\forall a, b \in \Omega,\ \forall \lambda \in [0,1] : \lambda a + (1-\lambda)b \in \Omega$.

Étant donnés $x_1, ..., x_n \in \mathbb{R}^n$, l'élément $x = \sum_{i=1}^{n} \lambda_i x_i$, où $\sum_{i=1}^{n} \lambda_i = 1$ et $\lambda_i \geqslant 0$ est appelé combinaison convexe de $x_1, ..., x_n$.

Proposition 1.24.

Un sous ensemble Ω de $\mathbb{R}^n$ est convexe si et seulement si il contient toutes les combinaisons convexes de ses éléments.

Démonstration. Condition suffisante est évidente. Pour justifier la nécessité, nous montrons par induction que toute combinaison convexe $\sum_{i=1}^{n} \lambda_i x_i$ d'éléments dans Ω est un élément de Ω. cette conclusion découle directement de la définition pour $n = 1, 2$.

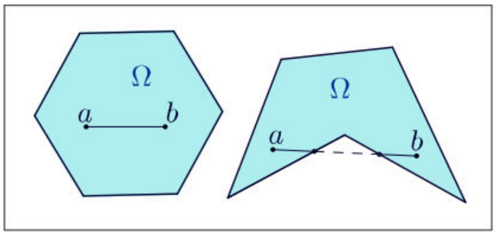


FIGURE 1.1 – Ensemble convexe et ensemble non convexe

Fixé maintenant un entier positif $n \geqslant 2$ et suppose que chaque combinaison convexe de $k \in \mathbb{N}$ élément de Ω, où $k \leqslant n$, appartient à Ω . De la combinaison convexe

$$y := \sum_{i=1}^{n+1} \lambda_i x_i, \; \sum_{i=1}^{n+1} \lambda_i = 1, \, \lambda_i \geqslant 0.$$

et observer que si $\lambda_{n+1} = 1$, alors $\lambda_1 = \lambda_2 = ... = \lambda_n = 0$, donc $y = x_{n+1} \in \Omega$. dans ce cas où $\lambda_{n+1} < 1$ nous obtenons les représentations

$$\sum_{i=1}^{n} \lambda_i = 1 - \lambda_{n+1} \; et \; \sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} = 1,$$

D'après hypothèse nous obtenons

$$z := \sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} x_i \in \Omega.$$

Il donne donc la relation

$$y = (1-\lambda_{n+1}) \sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} x_i + \lambda_{n+1} x_{n+1} = (1-\lambda_{n+1}) z + \lambda_{n+1} x_{n+1} \in \Omega$$

et ainsi compléter la preuve de la proposition. □

Étant données un sous ensemble Ω de $\mathbb{R}^n$, alors l'intersection de convexes contenant Ω est un convexe qui contient Ω.

Définition 1.25.

Soit Ω un sous ensemble de $\mathbb{R}^n$. On définit l'enveloppe convexe de Ω, noté $co\Omega$ par :

$$co\Omega := \bigcap \{C \subset \mathbb{R}^n \mid C \; convexe \, et \; \Omega \subset C\}.$$

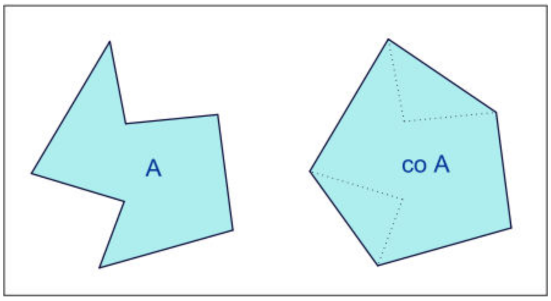


FIGURE 1.2 – Ensemble non convexe et son enveloppe convexe

Proposition 1.26.

Enveloppe convexe $co\Omega$ est le plus petit ensemble convexe contenant Ω.

Proposition 1.27.

Pour tout sous-ensemble Ω de $\mathbb{R}^n$, son enveloppe convexe admet la représentation

$$co\Omega = \left\{ \sum_{i=1}^{n} \lambda_i a_i \mid \sum_{i=1}^{n} \lambda_i = 1,\ \lambda_i \geqslant 0,\, a_i \in \Omega,\, n \in \mathbb{N} \right\}.$$

1.3 Fonctions convexes

cette section recueille des faits de base sur les fonctions convexes générales (à valeurs réelles étendue) y compris leurs caractérisations analytiques et géométriques, les propriétés importantes ainsi que leurs spécifications pour des sous-classes particulières.

Définition 1.28.

Le domaine et épigraphe de $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ est définie, respectivement, par :

$$domf := \{x \in \mathbb{R}^n : f(x) < +\infty\} \ et$$

$$Epi(f) = \{(x, \alpha) \in domf \times \mathbb{R} : \ f(x) \leqslant \alpha\}$$

La définition suivant donne une caractérisation géométrique de la convexité de fonction par la convexité de l'ensemble épigraphique associé.

Définition 1.29.

Une fonction $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ est convexe si et seulement si épigraphe $epif$ est un sous ensemble convexe de $\mathbb{R}^n \times \mathbb{R}$.

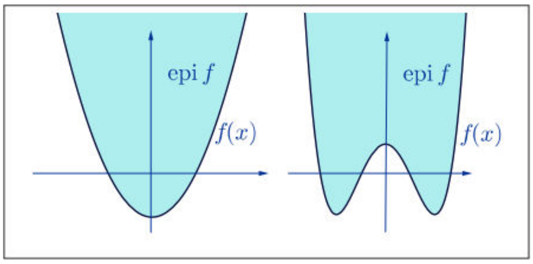


FIGURE 1.3 – Épigraphe

Proposition 1.30.

Soit $f : \Omega \to \overline{\mathbb{R}}$ une fonction définie sur $\Omega \subset \mathbb{R}^n$.la fonction f est convexe sur Ω si

$$\forall x, y \in \Omega,\ \forall \lambda \in [0,1] :\ f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y) \quad (1.1)$$

- Si l'inégalité dans 1.1 est strict pour $x \neq y$, alors f est strictement convexe sur Ω.
- f est dite concave si $(-f)$ est convexe.

Étant donnée une fonction $f : \Omega \to \overline{\mathbb{R}}$, le prolongement de f à $\mathbb{R}^n$ est définie par :

$$\tilde{f}(x) = \begin{cases} f(x) & \text{si } x \in \Omega \\ \infty & \text{si } non \end{cases} \quad (1.2)$$

Évidemment, si f est convexe sur le convexe Ω, alors $\tilde{f}(x)$ est convexe sur $\mathbb{R}^n$.
De plus, si f est une fonction convexe, alors elle est également convexe sur chaque sous-ensemble convexe de $\mathbb{R}^n$. Cela permet de considérer sans perte de généralité des fonctions convexes à valeurs réelles étendues sur tout l'espace $\mathbb{R}^n$.

Illustrons la convexité des fonctions par des exemples.

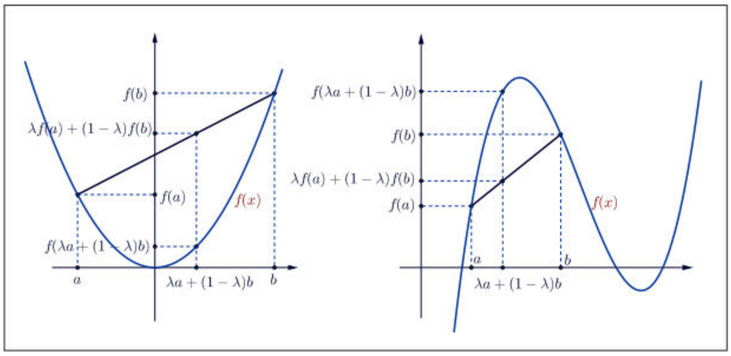


FIGURE 1.4 – fonction convexe et fonction non convexe

Exemple 1.30.1. *Les fonctions suivantes sont convexes :*

(i) $f(x) :=< a, x > +b$ *pour* $x \in \mathbb{R}^n$*, où* $a \in \mathbb{R}^n$ *et* $b \in \mathbb{R}$.

(ii) $g(x) :=\|x\|$ *pour* $x \in \mathbb{R}^n$.

(iii) $h(x) := x^2$ *pour* $x \in \mathbb{R}^n$.

Exemple 1.30.2. *Soit A une matrice* $n \times n$ *symétrique.On dit que A est semi définie positive si* $< Au, u >\geqslant 0$ *pour tout* $u \in \mathbb{R}^n$*. Vérifions que A est semi définie positive si et seulement si la fonction* $f : \mathbb{R}^n \to \mathbb{R}$ *définie par*

$$f(x) := \frac{1}{2} < Ax, x >, \qquad x \in \mathbb{R}^n.$$

est convexe. En effet, un calcul direct montre que pour tout $x, y \in \mathbb{R}^n$ *et* $\lambda \in [0, 1]$ *on a*

$$\lambda f(x)+(1-\lambda)f(y)-f(\lambda x+(1-\lambda)y) = \frac{1}{2}\lambda(1-\lambda) < A(x-y),(x-y) > . \quad (1.3)$$

Si la matrice A est semi définie positive, alors $< A(x-y),(x-y) >\geqslant 0$*, donc la fonction* f *est convexe par (1.3) . Inversement, en suppose la convexité de* f *et en utilisant l'égalité (1.3) pour* $x = u$ *et* $y = 0$ *vérifier que A est semi définie positive.*

La caractérisation suivante de la convexité est connue sous le nom **d'inégalité Jensen**.

Théoreme 1.31.

Une fonction $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ est convexe si et seulement si pour tous $\lambda_i \geqslant 0$ comme $i = 1, ..., m$ avec $\sum_{i=1}^{m} \lambda_i = 1$, et pour tous éléments $x_i \in \mathbb{R}^n$, $i = 1, ..., m$, alors

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leqslant \sum_{i=1}^{m} \lambda_i f(x_i). \tag{1.4}$$

Démonstration.
Depuis la relation (1.4) et pour $m = 2$ on a immédiatement la convexité de f , nous avons seulement besoin de prouver que toute fonction convexe f satisfie l'inégalité Jensen (1.4. On le monter par récurrence, pour m=1 l'inégalité (1.4) est trivial, pour m=2 on a l'inégalité (1.4) d'après la définition de convexité, supposons que l'inégalité (1.4) est vérifiée pour tout $m = k$ avec $k \geqslant 2$. Soit $\lambda_i \geqslant 0$, $i = 1, ...k + 1$, avec $\sum_{i=1}^{k} \lambda_i = 1 - \lambda_{k+1}$.

Si $\lambda_{k+1} = 1$, alors $\lambda_1 = \lambda_2 = ... = \lambda_k = 0$, donc (1.4) est vérifie. Supposons que $0 \leqslant \lambda_{n+1} < 1$ nous obtenons

$$\sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} = 1,$$

Par un calcul direct basés sur la convexité nous obtenons

$$\begin{aligned}
f\left(\sum_{i=1}^{k} \lambda_i x_i\right) &= f\left((1-\lambda_{k+1})\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x_i + \lambda_{k+1} x_{k+1}\right) \\
&\leqslant (1-\lambda_{k+1}) f\left(\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x_i\right) + \lambda_{k+1} f(x_{k+1}) \\
&\leqslant (1-\lambda_{k+1}) \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} f(x_i) + \lambda_{k+1} f(x_{k+1}) \\
&= \sum_{i=1}^{k+1} \lambda_i f(x_i)
\end{aligned}$$

ceci justifie l'inégalité (1.4) et complète la preuve du théorème. □

Maintenant, nous montrons que la convexité est préservée sous certaines opérations importantes.

Proposition 1.32.

Soit $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ une suite de fonctions convexes. Alors les fonctions suivantes sont convexes :
(i)— La multiplication par un scalaire λf, pour tout $\lambda > 0$.
(ii)— La somme de fonctions $\sum_{i=1}^{m} f_i$.
(iii)— Le maximum de fonctions $\max_{1 \leqslant i \leqslant m} f_i$.

Proposition 1.33.

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe et soit $\phi : \mathbb{R} \to \overline{\mathbb{R}}$ une fonction croissante et convexe . Alors la composition $\phi \circ f$ est convexe.

Proposition 1.34.

Soit $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ pour $i \in I$ une suite de fonctions convexes avec l'ensemble d'indice $I \neq \emptyset$. Alors la fonction $f(x) := \sup_{i \in I} f_i(x)$ est convexe.

Dans cette section, nous collectons quelques propriétés de fonctions convexes d'une variable réelle (c'est-à-dire une fonction convexe dont le domaine est un intervalle de $\mathbb{R}$) qui nous seront utiles dans la suite. La plupart des idées importantes de la théorie de la convexité peuvent être exprimées dans ce cadre dimensionnel.

Dans la suite de cette section, I désigne un intervalle arbitraire (fini ou infini) de $\mathbb{R}$

Lemme 1.35.

Soit $f : \mathbb{R} \to \overline{\mathbb{R}}$ une fonction convexe, supposons que son domaine est un intervalle ouvert I pour tout $a, b \in I$ et $a < x < b$, on a l'inégalité

$$\frac{f(x)-f(a)}{x-a} \leqslant \frac{f(b)-f(a)}{b-a} \leqslant \frac{f(b)-f(x)}{b-x}$$

Démonstration. Soit $a, b, x \in I$ avec $a < x < b$ on pose $t = \dfrac{x-a}{b-a} \in (0,1)$. Alors

$$f(x) = f(a + x - a) = f\Big(a + \frac{x-a}{b-a}(b-a)\Big) = f\big(a + t(b-a)\big) = f\big(tb + (1-t)a\big).$$

D'après la convexité de f on obtient

$$f(x) = f\big(tb + (1-t)a\big) \leqslant tf(b) + (1-t)f(a) \ \textit{donc}$$

$$f(x) - f(a) \leqslant tf(b) - tf(a) = \frac{x-a}{b-a}\Big(f(b) - f(a)\Big),$$

par suite

$$\frac{f(x)-f(a)}{x-a} \leqslant \frac{f(b)-f(a)}{b-a}.$$

De même on trouve

$$f(x)-f(b) \leqslant (t-1)f(b)+(1-t)f(a) = (t-1)\Big(f(b)-f(a)\Big) = \frac{x-b}{b-a}\Big(f(b)-f(a)\Big)$$

qui implique finalement que

$$\frac{f(b)-f(a)}{b-a} \leqslant \frac{f(b)-f(x)}{b-x}.$$

□

Théoreme 1.36.

Supposons que $f : \mathbb{R} \to \overline{\mathbb{R}}$ est dérivable sur son domaine, qui est un intervalle ouvert I. Alors f est convexe si et seulement si f' est croissante sur I.

Démonstration. Supposons que f est convexe et soit $a, b \in I$ tel que $a < b$. Par le lemme (1.35), on a

$$\frac{f(x)-f(a)}{x-a} \leqslant \frac{f(b)-f(a)}{b-a}$$

pour tout $x \in (a,b)$.Ceci implique par la définition de dérivé que

$$f'(a) \leqslant \frac{f(b)-f(a)}{b-a}$$

De même, on arrive à

$$\frac{f(b)-f(a)}{b-a} \leqslant f'(b)$$

et conclure que $f'(a) \leqslant f'(b)$, c.à.d , f' est une fonction croissante.

Pour prouver le contraire, supposons que f' est croissante et soient $x, y \in I$ avec $x < y$ et $t \in (0,1)$. Alors

$$x < x_t < y \ \ pour x_t := tx + (1-t)y.$$

Par le théorème des accroissement finis, on trouve c_1, c_2 tel que $x < c_1 < x_t < c_2 < y \quad et$

$$f(x_t) - f(x) = f'(c_1)(x_t - x) = f'(c_1)(1-t)(y-x),$$
$$f(x_t) - f(y) = f'(c_2)(x_t - y) = f'(c_2)t(x-y).$$

Par suite

$$tf(x_t) - tf(x) = f'(c_1)t(1-t)(y-x),$$
$$(1-t)f(x_t) - (1-t)f(y) = f'(c_2)t(1-t)(x-y).$$

Puisque $f'(c_1) \leqslant f'(c_2)$, en ajoutant ces dernières inégalité on trouve

$$f(x_t) \leqslant tf(x) + (1-t)f(y).$$

Ce qui prouve la convexité de f . □

Corollaire 1.37.

Soit $f : \mathbb{R} \to \overline{\mathbb{R}}$ une fonction 2 fois dérivable sur son domaine qui est un intervalle ouvert I. Alors f est convexe si et seulement si $f''(x) \geqslant 0$ pour tout $x \in I$.

Démonstration. $f''(x) \geqslant 0$ pour tout $x \in I$ si et seulement si $f'(x)$ est croissante sur I. Alors la conclusion découle directement du théorème (1.36). □

Proposition 1.38.

Soit $f : I \to \mathbb{R}$ une fonction convexe . Alors $\forall x \in intI, f$ admit la dérivé à droite $f'_d(x)$ et la dérivé à gauche $f'_g(x)$ qui sont finis.
De plus $\forall x, y \in intI$ tel que $x < y$, on a

$$f'_g(x) \leqslant f'_d(x) \leqslant f'_g(y) \leqslant f'_d(y).$$

Corollaire 1.39.

Soit $f : I \to \mathbb{R}$ une fonction convexe . Alors f est continue sur $intI$.

Définition 1.40.

Une fonction $f : I \to \mathbb{R}$ est localement convexe si pour tout $x \in I$ il existe un intervalle ouvert $I_x \subset I$ tel que f est convexe sur I_x.

Théoreme 1.41.

Soit $f : I \to \mathbb{R}$ une fonction à valeurs dans $\mathbb{R}$. Alors tout fonction localement convexe est globalement convexe.

Pour montrer ce théorème on a besoin du lemme suivant :

Lemme 1.42.

Soit I un intervalle de $\mathbb{R}$ et soit $a, b \in I$, avec $a < b$. Si $f : I \to \mathbb{R}$ est convexe sur $I \bigcap [a, +\infty)$ et $I \bigcap (-\infty, b]$, Alors f est convexe sur I.

Démonstration. Soit $a, b \in I$, si a ou b est un extremum de I alors f est convexe sur I.
Si non : Soit $x, y \in I$ avec $x < y$ et $t \in [0,1]$ posons $z = tx + (1-t)y$ Montrons que $f(z) \leqslant tf(x) + (1-t)f(y)$.

1ère **cas :** **x<a<z<b<y**

alors $\exists r, s \in [0,1]$ tel que $a = rx + (1-r)z$ et $z = sa + (1-s)y$ donc,

$$\begin{aligned} z &= s(rx + (1-r)z) + (1-s)y \\ &= \frac{sr}{1-s(1-r)}x + \frac{1-s}{1-s(1-r)}y, \end{aligned}$$

donc $t = \dfrac{sr}{1-s(1-r)}$
f continue sur $[a, b]$, et f convexe sur $I \cap [a, +\infty[$ et $I \cap]-\infty, b]$ donc

$$f(z) \leqslant tf(x) + (1-t)f(y).$$

D'où f convexe sur I.

2ème **cas :** **x<a<b<z<y**

donc, $\exists r, s \in [0,1]$ tel que $z = rb + (1-r)y$ et $b = sa + (1-s)y$ et $a = tx + (1-t)z$ d'après la convexité de f sur $I \cap [a, +\infty[$ et $I \cap]-\infty, b]$ on trouve que f est convexe sur I.

3ème **cas :** **x<z<a<b<y**

D'une manière similaire de 2ème **cas** donc f est convexe sur I.

□

Démonstration du théorème. Soit $z \in I$ fixé tel que z n'est pas un point extrême. Soit

$$d =: \sup\{y \in [z, b] \mid f \, convexe \, sur I \cap [z, y]\}$$

f localement convexe sur I et z n'est pas un point extrême de I on a : $z < d$ et f convexe sur $I \cap [z, d)$. Si $d \in I$ d'après la convexité local de f on a f convexe sur $I \cap (d-r, d+r)$, $r > 0$.

Puisque f convexe sur $I \cap (d-r, d+r)$ et f convexe sur $I \cap [z, d)$ donc d'après le lemme f convexe sur $I \cap [z, d+r)$

Donc il suffit de montrer que $I \cap [z, d+r) = I \cap [z, +\infty[, r > 0$.
Supposons que $I \cap [z, d+r) \neq I \cap [z, +\infty[, r > 0$
donc $\exists c \in I \cap [z, +\infty)$ tel que $c \notin I \cap [z, d+r)$
c'est-à-dire $c \in [z, +\infty)$ et $c \notin [z, d+r)$.

Soit $\mathbf{V}_x$ un intervalle ouvert de x (Voisinage de x), pour tout $x \in [z, c]$. On a :

$$[z, c] = \bigcup_{x \in [z,c]} \mathbf{V}_x,$$

$[z, c]$ est un compact donc, $\exists x_1, ..., x_n \in [z, c]$ tel que

$$[z, c] = \bigcup_{i=1}^{n} \mathbf{V}_{x_i},$$

on suppose que l'ensemble d'indexation est bien ordonné, i.e $x_1 < ... < x_n$. Donc forcement $\mathbf{V}_{x_i} \cap \mathbf{V}_{x_i+1} \neq \emptyset$ pour tout $i = 1, ..., n$, puisque $f/_{\mathbf{V}_{x_i}}$ est convexe pour tout $i = 1, ..., n$ et d'après le lemme alors,

$$f/_{\bigcup_{i=1}^{n} \mathbf{V}_{x_i}} \textit{ est convexe.}$$

Finalement f convexe sur $[z, c]$, ce qui est absurde car $\big(d =: \sup\{y \in [z, b] \mid f \textit{ convexe sur } I \cap [z, y]\}\big)$, alors $I \cap [z, d+r) = I \cap [z, +\infty[, r > 0$.

De plus f est convexe sur $I \cap]-\infty, z]$ et f convexe sur voisinage z donc f convexe sur $I \cap (z-s, z+s)$, $s > 0$ d'après le lemme f est convexe sur I.

Ce qui achève la preuve.

□

1.4 Intérieur relatif d'un ensemble

Nous commençons cette section avec le définition et les propriétés des ensembles affines. Soit a, $b \in R$, la ligne reliant les deux points est

$$\mathfrak{L}[a, b] := \{\lambda a + (1-\lambda) b : \lambda \in \mathbb{R}\}.$$

noté que si $a = b$, alors $\mathfrak{L}[a, b] = \{a\}$.

Définition 1.43.

Un sous ensemble Ω de $\mathbb{R}^n$ est affine si pour tout $a,b \in \Omega$ on a $\mathfrak{L}[a,b] \subset \Omega$

Par exemple, tout point, ligne et plans dans $\mathbb{R}^3$ sont des ensembles affines.
l'ensemble vide et l'espace entier sont toujours affines.
Il résulte de la définition que l'intersection d'ensembles affines est affine . Ce qui nous amène à la construction de l'enveloppe affine d'un ensemble.

Définition 1.44.

L'enveloppe affine d'un ensemble $\Omega \subset \mathbb{R}^n$ est

$$\textbf{aff}\,\Omega := \bigcap \{C \mid C \textit{ affine et } \Omega \subset C\}.$$

Maintenant, nous sommes prêts à définir une notion majeure d'intérieurs relatifs des ensembles convexes.

Définition 1.45.

Soit Ω un ensemble convexe. On dit que $x \in \Omega$ est un élément de l'intérieur relatif de Ω noté $ri\Omega$ s'il existe $\varepsilon > 0$ tel que $B(x;\varepsilon) \cap \textbf{aff}\Omega \subset \Omega$.

Le prochain est l'un des résultats les plus fondamentaux de la géométrie convexe.

Théoreme 1.46.

Soit Ω un ensemble convexe non vide de $\mathbb{R}^n$. Alors
(i) $ri(C) \neq \emptyset$.
(ii) $[a,b) \subset ri\Omega$ pour tout $a \in ri\Omega$ et $b \in \overline{\Omega}$.

Nous concluons cette section par les propriétés suivantes de l'intérieur relatif.

Propriété 1.47.

Soit Ω un ensemble convexe non vide de $\mathbb{R}^n$. Pour les ensembles convexes $ri\Omega$ et $\overline{\Omega}$,
on a
(i) $\overline{ri\Omega} = \overline{\Omega}$.
(ii) $ri\Omega = ri\overline{\Omega}$

1.5 Diverses notions de différentiabilité

On rappelle dans ce paragraphe les trois types de différentiabilité utilisées en analyse, et étendu dans le contexte des fonctions numériques seulement.
Soient E un espace de Banach et Ω un ouvert de E. Soit $f : \Omega \subset E \to \mathbb{R}$ une application dans voisinage de x élément de Ω.

1.5.1 différentiabilité de Fréchet

Définition 1.48.

On dit que f est différentiable au sens de M.Fréchet (F.différentiable en abrégé) en $x \in \Omega$, s'il existe une forme linéaire continue l sur $\mathbb{R}^n$ tel que :

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - l(h)}{||h||} = 0.$$

La forme linéaire l, notée $D_F f(x)$ ou simplement $Df(x)$, un élément de l'espace dual de E.

Remarque 1.48.1. *Si E est un espace de Hilbert, la forme linéaire, continue $D_F f(x)$ est représentée par un élément de E^*, noté $\nabla_F f(x)$ ou ($\nabla f(x)$ simplement) et appelé gradient de f en x :*

$$\forall d \in E,\ D_F f(x)d = \left\langle \nabla f(x), d \right\rangle.$$

Théoreme 1.49.

Soit f une fonction différentiable. Les propositions suivantes sont équivalentes :

1. f est convexe.
2. Ses hyperplans tangents sont des minorants :

$$\forall (x,y) \in E \times E,\ f(y) \geqslant f(x) + < \nabla f(x), y - x > .$$

3. Le gradient de f est un opérateur monotone :

$$\forall (x,y) \in E \times E,\ < \nabla f(y) - \nabla f(x), y - x > \geqslant 0.$$

1.5.2 Dérivée directionnelle

L'analyse du comportement de la fonction dans certaines directions sera également importante dans le cadre des méthodes d'optimisation. Nous introduisons le concept de dérivée directionnelle.

Définition 1.50.

Soit f une fonction à valeur réel et $x_0 \in domf$ .La dérivée directionnelle de f en x_0 dans la direction $d \in E$ est la limite suivante

$$f'(x_0; d) := \lim_{\lambda \to 0^+} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}. \tag{1.5}$$

Si cette limite existe,on dit que $f'(x_0; d)$ est la dérivée directionnelle de f en x.

Notez que la relation (2.1) est parfois appelée la dérivée directionnelle droite f en x dans la direction d. Sa contrepartie gauche est définie par

$$f'_{-}(x_0; d) := \lim_{\lambda \to 0^-} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

Il est facile de voir que

$$f'_{-}(x_0; d) = -f'(x_0; -d) \quad \forall d \in E.$$

Lemme 1.51.

Soit f une fonction convexe à valeurs réel et $x_0 \in dom f$ et soit $d \in E$, définie

$$\phi(\lambda) := \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}, \quad \lambda > 0.$$

Alors la fonction ϕ est croissante sur $]0, +\infty]$.

Démonstration. On fixe, $0 < \lambda_1 < \lambda_2$ on pose

$$x_0 + \lambda_1 d := \frac{\lambda_1}{\lambda_2}(x_0 + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2})x_0.$$

on applique f qui est convexe on trouve

$$\phi(\lambda_1) \leqslant \phi(\lambda_2),$$

donc ϕ est croissante sur $]0, +\infty]$. □

Proposition 1.52.

Soit f et $x_0 \in dom f$, la dérivée directionnelle $f'(x_0; d)$ existe dans tout direction $d \in E$ De plus

$$f'(x_0; d) = \inf_{\lambda > 0} \phi(\lambda), \quad d \in E$$

Démonstration. D'après le lemme (1.51) ϕ est croissante . Ainsi nous avons

$$f'(x_0; d) = \lim_{\lambda \to 0^+} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda} = \lim_{\lambda \to 0^+} \phi(\lambda) = \inf_{\lambda > 0} \phi(\lambda),$$

ce qui prouve la proposition. □

Corollaire 1.53.

Si f est une fonction convexe, alors $f'(x_0; d)$ est un nombre réel pour tout $x_0 \in int(dom f)$ et $d \in E$.

Rappelons que la dérivabilité selon tout les directions en x n'implique pas nécessairement la différentiabilité de f en x . Il suffit de prendre $f(x) = |x|$.

1.5.3 différentiabilité de Gâteau

Définition 1.54.

On dit que f est différentiable au sens de R. Gâteau (G-différentiable en abrégé) en x_0 lorsque :

$$\forall d \in E,\ \lim_{\lambda \to 0^+} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}$$

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

$$\lim_{\lambda \to 0^+} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda} = \left\langle D_G f(x_0), d. \right\rangle$$

Remarque 1.54.1. *On peut deéfinir aussi le Gradient à partir de la différentielle au sens de Gâteaux.*

Définition 1.55.

Soit f est Gateaux différentiable alors $f'_G(x) \in \mathbb{R}^n$ admet par le théorème de Riesz un représentant c unique définie par :

$$D_G f(x)(y) = \left\langle c, y \right\rangle,\ \forall y \in E,$$

est appelé gradient de f en x et noté : $c = \nabla f(x)$.

Il est clair que si f est Fréchet différentiable alors f est Gâteaux différentiable. La réciproque est fausse en général comme le montre l'exemple simple suivant :

Exemple 1.55.1. *Considérons la fonction* $f : \mathbb{R}^2 \to \mathbb{R}$ *définie par*

$$f(x) = \begin{cases} 1 & si\ x = y^2\ et\ y \neq 0 \\ 0 & si\ non \end{cases}$$

f *est G-différentiable en* $(0,0)$ *mais elle n'est pas F-différentiable en* $(0,0)$ *puisqu'elle n'est pas continue en ce point.*

Remarque 1.55.1. *Cependant sous certaines conditions, la réciproque est vrai.*

Proposition 1.56.

Si $D_G f(x)$ est continue en x alors $D_G f(x)$ existe.

1.5.4 différentiabilité de Hadamard

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

Définition 1.57.

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

$$\lim_{\lambda \to 0} \frac{f(x+\lambda d) - f(x)}{\lambda} = \left\langle D_H f(x), d \right\rangle \textit{ uniformément pour } d \in S,$$

et ce pour tout $S \subset \mathscr{B}$.

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

- La F-différentiabilité de f en x s'écrit, de manière équivalente, comme dans la définition (1.57), en prenant pour $\mathscr{B}$ la collection des fermés bornés de E.
- La G-différentiabilité de f en x s'écrit, de manière équivalente, comme en (1.57), en prenant pour $\mathscr{B}$ la collection des ensembles finis de points de E.

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

Proposition 1.58.

F-différentiabilité $\Rightarrow$ H-différentiabilité $\Rightarrow$ G-différentiabilité.

La H-différentiabilité (et donc la F-différentiabilité) de f en x implique la continuité de f en x ; ce n'est pas le cas pour la G-différentiabilité. La semi-continuité inférieure n'est pas acquise non plus avec la G-différentiabilité ;ce qui fait qu'on a des énoncés de théorèmes avec des hypothèses comme "soit f s.c.i. et G-différentiable sur E", laquelle est assurée avec "soit f F-différentiable sur E ".

Remarque 1.58.1. *Si f vérifie une condition de Lipschitz dans un voisinage de x, alors*
H-différentiabilité en x $\Longleftrightarrow$ G-différentiabilité en x.

CHAPITRE 2

SOUS DIFFÉRENTIEL AU SENS D'ANALYSE CONVEXE

Dans ce chapitre, nous nous introduisons les outils spécifiques à l'optimisation non différentiable tels que le sous-différentiel et l'opérateur proximal. Nous verrons également comment s'écrivent les conditions nécessaires d'optimalité dans le cadre non lisse.

2.1 Sous-gradient et sous différentiel

le sous différentiel est un concept permettant de décrire la variation locale d'une fonction convexe (à valeur réelles) non nécessairement différentiable dans un sens classique. Le sous différentie d'une fonction convexe est l'ensemble des pentes de toutes les minorants affines de la fonction, voir [17], [1], [6], [22].

Définition 2.59.

Soit f une fonction convexe . Un vecteur $\eta \in \mathbb{R}^n$ est appelé sous gradient de f au point $x_0 \in dom(f)$ si :

$$\forall x \in dom(f)\; f(x) - f(x_0) \geqslant < \eta, x - x_0 >$$

L'applcation $a(x) = f(x_0) + < \eta, x - x_0 >$ est une minorante affine de f vérifiant :

$$a(x) \leqslant f(x)\; pour\; tout\; x \in \mathbb{R}^n\; a(x_0) = f(x_0).$$

La définition (2.59) exprime que la fonction affine continue

$$x \to \langle \eta, y \rangle + f(x_0) - \langle \eta, x_0 \rangle$$

de pente η, minore f sur $\mathbb{R}^n$ et coïncide avec elle en x_0. Autre manière de dire les choses : η est un sous-gradient de f en x_0 si, et seulement si, x_0 est un minimiser de la fonction perturbée $x \to f(x) - \langle \eta, x \rangle$ sur $\mathbb{R}^n$.

Les appellations sous-gradient ou sous-différentiel doivent faire penser que ces concepts ont quelque chose à voir avec les objets du Calcul différentiel mais qu'ils interviennent "par dessous les fonctions". le sous-différentiel est défini pour n'importe quelle fonction, mais nous verrons qu'il fonctionne bien essentiellement dans le cas les fonctions convexes. Des généralisations du concept seront abordées au Chapitre suivant.

Donnons quelques exemples.

Exemple 2.59.1. *Soit $f : \mathbb{R} \to \mathbb{R}$, défini par $f(x) = |x|$. Alors,*

$$\partial f(x) = \{-1\} \text{ si } x < 0 \text{ , } \{1\} \text{ si } x > 0 \text{ , } [-1, 1] \text{ si } x = 0.$$

En parallèle de cette fonction, considérons $g = i_{[-1,1]}$ (la fonction indicatrice de $[-1, 1]$). Alors,

$$\partial g(x) = \{0\} \text{ si } -1 < x < 1 \text{ , } \mathbb{R}^- \text{ si } x = -1 \text{ , } \mathbb{R}^+ \text{ si } x = 1.$$

Les graphes de ces deux multiaplications sous-différentiels,∂f et ∂g, tracés ci-dessous, sont à garder à l'esprit car ils sont dans une relation particulière.

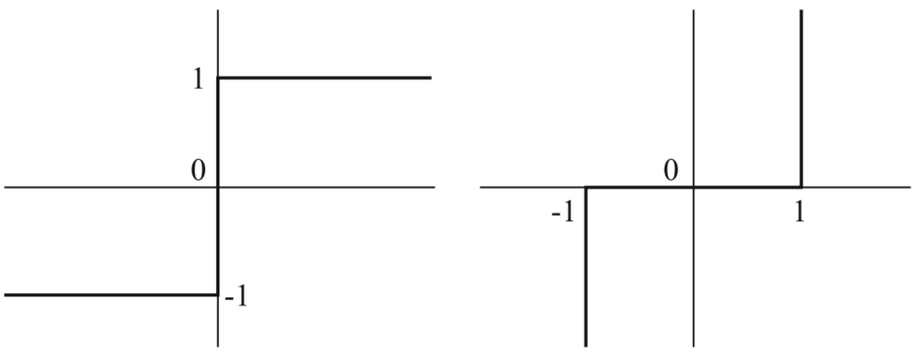


FIGURE 2.1 – graphes des multifonctions $\partial|.|$ et $\partial i_.$

Soit $\Omega \subset \mathbb{R}^n$ non vide, soit $x \in \Omega$. Alors d'après la Définition (2.59), on a

$$\eta \in \partial i_\Omega(x) \Leftrightarrow \langle \eta, y - x \rangle \leqslant 0 \ \forall y \in \Omega. \tag{2.1}$$

L'ensemble $\partial i_\Omega(x)$ est appelé **cone normal** à Ω en x, il est désormais noté $N(\Omega, x)$ ou $N_\Omega(x)$. La signification géométrique de l'inégalité présente dans (2.1) est clair : η fait un "angle obtus" avec tout vecteur $y - x$ s'appuyant sur $y \in \Omega$.

Remarque 2.59.1. *Une fonction peut admettre plusieurs sous-gradients en un point où elle n'est pas différentiable.*

Définition 2.60.

On appelle sous-différentiel de f en x_0 l'ensemble des sous-gradients de f en x_0, noté par :

$$\partial f(x_0) = \{\eta \in \mathbb{R}^n \mid \eta \ sous-gradient \ de \ f \ en \ x_0\}.$$

Propriété 2.61.

1. Lien géométrique avec l'épigraphe de f. On a :

$$\eta \in \partial f(x_0) \Leftrightarrow \Big((\eta, -1) \in \mathbb{R}^n \times \mathbb{R} \ \textit{est normal à} \ epi(f) \ en \ \big(x_0, f(x_0)\big)$$

$$\text{i.e } \ (\eta, -1) \in N_{epi(f)}\big(x_0, f(x_0)\big)\Big).$$

2. Lien avec la Transformation de Legendre-Fenchel voir section (2.4). On a :

$$\begin{aligned} \eta \in \partial f(x_0) &\Leftrightarrow \big(f^*(\eta) + f(x_0) \geqslant \langle \eta, x_0 \rangle\big) \\ &\Leftrightarrow \big(f^*(\eta) + f(x_0) = \langle \eta, x_0 \rangle\big) \end{aligned}$$

En clair, il y a égalité dans l'inégalité de Fenchel (section (2.4)) exactement lorsque $\eta \in \partial f(x_0)$.

Démonstration. 1. D'après (2.1) on a :

$$(\eta, -1) \in N_{epi(f)}\big(x_0, f(x_0) \Leftrightarrow \langle (\eta, -1); (x, f(x)) - (x_0, f(x_0)) \rangle \leqslant 0, \ \forall (x, f(x)) \in epi(f),$$

$$\Leftrightarrow \langle \eta ; x - x_0 \rangle + \langle -1 ; f(x) - f(x_0) \rangle \leqslant 0, \ \forall (x, f(x)) \in epi(f),$$

D'où

$$\langle \eta ; x - x_0 \rangle + f(x_0) \rangle \leqslant f(x), \ \forall x \in \mathbb{R}^n.$$

Ce qui achève la preuve.

2. La proposition (2.78) donne la première équivalence, pour la deuxième si $\eta \in \partial f(x_0)$ on a :

$$\langle \eta, x \rangle - \langle \eta, x_0 \rangle + f(x_0) \leqslant f(x), \ x \in \mathbb{R}^n,$$

par suite

$$\langle \eta, x_0 \rangle - f(x_0) \geqslant \sup_{x \in \mathbb{R}^n} \{ \langle \eta, x \rangle - f(x) \} = f^*(\eta), \ x \in \mathbb{R}^n,$$

donc,

$$\langle \eta, x_0 \rangle \geqslant f(x_0) + f^*(\eta), \ x \in \mathbb{R}^n.$$

D'où le résultat.

□

Pourvu qu'il y ait coïncidence des valeurs en x_0, la sous-différentiation ne sait pas discerner tout ce qui est entre f et $\overline{co}f$:

$$\left(\overline{co}f \leqslant g \leqslant f \ et \ f(x_0) = g(x_0) \right) \Rightarrow \left(\partial f(x_0) = \partial g(x_0) \right).$$

Propriété 2.62.

Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe.

(i)– Si $\partial f(x_0) \neq \emptyset$, alors $\overline{co}f$ et f coïncident en x_0.

(ii)– Si $\eta \in \partial f(x_0)$, alors $x_0 \in \partial f^*(\eta)$.

Proposition 2.63.

Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe, semi-continue inférieur et fini en un point, alors on a :

$$\eta \in \partial f(x_0) \Leftrightarrow x_0 \in \partial f^*(\eta).$$

Géométriquement, cela signifie que les graphes des multiapplications ∂f et ∂f^* sont "inverse" l'un de l'autre :

$$(x_0, \eta) \in \ graphe \ de \ \partial f \Leftrightarrow (x_0, \eta) \in \ graphe \ de \ \partial f^*.$$

C'est le moment de revoir l'exemple qui conduit à la figure (2.1) : la fonction g n'y autre que f^*.

> **Remarque 2.63.1.** ***Étude lorsque*** n=1 ***(dimention 1)*** *Signalons ici que bien entendu,* $\mathbb{R}^*$ *s'identifie à* $\mathbb{R}$ *et l'on considère que* $\partial f(x)$ *est un sous ensemble de* $\mathbb{R}$.

Théoreme 2.64.

Soit f une fonction convexe et x dans l'intérieur de $dom(f)$. Alors :

$$\partial f(x) = \left[f'_g(x), f'_d(x)\right].$$

Interprétation géométrique

L'interprétation géométrique du sous-différentiel est la suivante. Il est formée par toutes les directions des hyperplans qui passent par le point $(x; f(x))$ et restent "sous" le graphe de la fonction f. Ces hyperplans sont appelés hyperplans support ou hyperplans d'appui au graphe de f en x.

Le corollaire suivant explique pourquoi il faut voir $\partial f(x)$ comme une extension de $f'(x)$.

Corollaire 2.65.

Soit x dans l'intérieur de $dom(f)$.
f est dérivable en x si et seulement si $\partial f(x)$ est un singleton, et dans ce cas

$$\partial f(x) = \left\{f'(x)\right\}.$$

Étude de dimension quelconque pour un espace de Banach X arbitraire, nous n'avons pas une caractérisation aussi agréable qu'en dimension un, même lorsque x est intérieur à $dom(f)$.
Nous allons juste signaler quelques propriétés.

Théoreme 2.66.

Soit $x \in int(dom(f))$. Alors f est G-dérivable en x si et seulement si $\partial f(x)$ est un singleton, et dans ce cas

$$\partial f(x) = \{D_G f(x)\}.$$

Remarque 2.66.1. *Lorsque la dimension de X est finie, l'existence de $D_G f(x)$ est équivalente à celle de $D_F f(x)$. Ainsi, lorsque X est de dimension finie, on peut énoncer le théorème précédent en : f est Fréchet-dérivable en x si et seulement si $\partial f(x)$ est un singleton, et dans ce cas*

$$\partial f(x) = \{D_F f(x)\}.$$

2.2 Existence de sous-gradient

Théoreme 2.67.

Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe qui est finie et continue en au moins un point x. Alors, le sous-différentiel de f est non vide en tout point de l'intérieur de son domaine (ce qui lui-même non vide) et en particulier en x.

Une autre définition du sous-différentiel
On donne maintenant un autre point de vue sur le sous-différentiel d'une fonction convexe plus en rapport, avec la notion de dérivée directionnelle.

Théoreme 2.68.

Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe. En un point $x \in \mathbb{R}^n$, on a :

$$\partial f(x) = \left\{x^* \in \mathbb{R}^n \mid f'(x;d) \geqslant < x^*, d >, \forall d \in \mathbb{R}^n\right\}.$$

Lemme 2.69.

$\partial f(x_0)$ est un ensemble convexe fermé.
De plus, si $x \notin dom(f)$ on a $\partial f(x) = \emptyset$.

Démonstration. Montrons que $\partial f(x_0)$ est convexe.
Soient η_1 , $\eta_2 \in \partial f(x_0)$ on a $\forall x \in \mathbb{R}^n$

$$f(x) - f(x_0) \geqslant < \eta_1, x - x_0 >$$

$$f(x) - f(x_0) \geqslant < \eta_2, x - x_0 > .$$

Soit $\alpha \in [0,1]$, alors

$$\alpha f(x) - \alpha f(x_0) \geqslant \alpha < \eta_1, x - x_0 >$$

$$(1-\alpha) f(x) - (1-\alpha) f(x_0) \geqslant (1-\alpha) < \eta_2, x - x_0 >,$$

en sommant , on voit que $\alpha\eta_1 + (1-\alpha)\eta_2 \in \partial f(x_0)$.
Ce qui achève la preuve. □

Corollaire 2.70.

Si $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ est une fonction convexe, continue en x et $x \in dom(f)$, alors la fonction $d \mapsto f'(x;d)$ est la fonction support de $\partial f(x)$. Autrement dit,

$$f'(x;d) = \sup_{x^* \in \partial f(x)} \left\langle x^*, d \right\rangle \ \forall d \in \mathbb{R}^n.$$

Si $\partial f(x)$ est réduit à un singleton, la dérivée directionnelle $f'(x;d)$ sera une fonction linéaire en d, autrement dit f est G-différentiable au point x. La réciproque est également vraie.

Théoreme 2.71.

Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe, propre et dérivable en $x \in \mathbb{R}^n$. Elle admet une dérivée $f'(x)$ si et seulement si $\partial f(x)$ réduit à un seul point, qui est $f'(x)$, et donc :

$$f(y) - f(x) \geqslant \left\langle f'(x), y - x \right\rangle \ , \forall y \in \mathbb{R}^n.$$

2.3 Règles de calcul du sous-différentiel

La correspondance $x \rightrightarrows \partial f(x)$ est une application multivoque ou multi-fonction, c'est-à-dire qu'elle associe à un point $x \in \mathbb{R}^n$ le sous ensemble $\partial f(x)$ de $\mathbb{R}^n$. De plus elle est ici à "valeurs" convexes fermées.
Nous verrons ci-dessous l'avantage de sous-différentiel pour l'optimisation. Signalons

que l'on dispose d'un calcul sous-différentiel, malheureusement moins maniable que le calcul différentiel.

On considère ici un certain nombre d'opération préservant la convexité des fonctions et envisage le calcul de sous-différentiel des fonction obtenues à partir de ceux des fonctions de départ.

la multi-fonction $x \rightrightarrows \partial f(x)$ est monotone (croissante), c'est-à-dire*

$$\left(\eta_1 \in \partial f(x_1)\ et\ \eta_2 \in \partial f(x_2)\right) \Rightarrow \left(\left\langle \eta_1 - \eta_2, x_1 - x_2 \right\rangle \geqslant 0\right).$$

Ce résultat immédiatement de la définition de sous-différentiel écrite avec $x = x_1$ et $y = x_2$, puis $x = x_2$ et $y = x_1$.
Soit $x_1, ..., x_k$ k points, $\eta_1, ..., \eta_k$ k sous-gradients de f, avec $\eta_i \in \partial f(x_i)$ pour tout $i = 1, ..., k$. Alors,

$$\sum_{i=0}^{k} \left\langle s_i, x_{i+1} - x_i \right\rangle,$$

en convenant que $x_{k+1} = x_1$.
Par définition de sous-différentiel on a :

$$\partial(\alpha f) = \alpha \partial f(x),\ \forall \alpha > 0, \forall x \in \mathbb{R}^n.$$

Proposition 2.72.

Soit $f_1, f_2, ...f_n$ des fonctions convexes de $\mathbb{R}^n$ dans $\overline{\mathbb{R}}$ on a :

1. $\partial\left(\sum_{k=1}^{n} f_k\right)(x) \supset \sum_{k=1}^{n} (\partial f_k(x)),\ \ \forall x \in \mathbb{R}^n.$
   Avec l'égalité si $\bigcap_{i=1}^{n} ri(dom(f_i)) \neq \emptyset$ ou bien s'il existe un point dans $\mathbb{R}^n$ en lequel au moins $(n-1)$ fonctions sont continues et la dernière est finie.
2. Soit $I_x = \left\{i,\ f_i(x) = \max_j f_j(x)\right\}$. Alors :
   $\partial\left(\max_i f_i\right)(x) = \overline{co\left(\bigcup_{i \in I_x} \partial f_i(x)\right)}$

*. Lorsque l'inégalité est dans l'autre sens, $\left\langle \eta_1 - \eta_2, x_1 - x_2 \right\rangle \leqslant 0$, on parle de multi-fonction monotone décroissante ou, plutôt, dissipative.

Démonstration.

1- Montrons l'inclusion : Soit $x^* \in \sum_{k=1}^{n} (\partial f_k(x))$,
donc il existe $(x_1^*, ..., x_n^*) \in \partial f_1(x) \times, ..., \times \partial f_n(x)$ tel que $x^* = x_1^* + ... + x_n^*$ on a :

$$\begin{aligned}\forall z \in \mathbb{R}^n \; f(z) = f_1(z) + ... + f_n(z) &\geqslant f_1(x) + ... + f_n(x) + \langle x_1^*, z - x \rangle + ... + \langle x_n^*, z - x \rangle \\ &= f(x) + \langle x_1^* + ... + x_n^*, z - x \rangle \\ &= f(x) + \langle x^*, z - x \rangle.\end{aligned}$$

Donc, $x^* \in \partial f(x)$. Inversement, si $\bigcap_{i=1}^{n} ri(dom(f_i)) \neq \emptyset$, alors d'après la propriété (2.61) on a :

$$x^* \in \partial f(x) \Leftrightarrow \langle x^*, x \rangle = \sum_{k=1}^{n} f_k(x) + \left(\sum_{k=1}^{n} f_k(x^*) \right)^*.$$

Par le théorème 16.4 ([17])

$$x^* \in \partial f(x) \Leftrightarrow \langle x^*, x \rangle = \sum_{k=1}^{n} f_k(x) + \inf\{f_1^*(x_1^*) + ... + f_1^*(x_n^*) \mid x^* = x_1^* + ... + x_n^*\}$$

où pour x^* inf est atteint par certains $x_1^*, ..., x_n^*$. Ainsi $\partial f(x)$ est constitué des vecteurs de la forme $x_1^* + ... + x_n^*$ tel que

$$\langle x_1^*, x \rangle + ... + \langle x_n^*, x \rangle = \sum_{k=1}^{n} f_k(x) + \sum_{k=1}^{n} f_k^*(x_k^*).$$

Puisque on a toujours

$$\langle x_k^*, x \rangle \leqslant f_k(x) + f_k^*(x_k^*),$$

avec égalité si, et seulement, si

$$x_k^* \in \partial f_k^*(x).$$

Ce qui achève la preuve.

□

Proposition 2.73.

Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe et $A : \mathbb{R}^n \to \mathbb{R}^n$ un opérateur linéaire continue alors :

$$\partial(f \circ A)(x) \supset A^* \partial f(Ax), \; \forall x \in \mathbb{R}^n.$$

L'égalité aura lieu s'il existe un point dans $dom(f) \cap ImA$ ou f est continue.

Démonstration. Si $x^* \in A^*\partial f(Ax)$ alors, $x^* = A^* y^*$ avec $y^* \in \partial f(Ax)$

$\forall z \in \mathbb{R}^n$ on a :

$$\begin{aligned} f(Az) &\geqslant f(Ax) + \left\langle y^*, Az - Ax \right\rangle \\ &= f(Ax) + \left\langle A^* y^*, z - x \right\rangle \\ &= f(Ax) + \left\langle x^*, z - x \right\rangle \end{aligned}$$

. D'où $x^* \in \partial(f \circ A)(x)$.

Inversement, supposons qu'il existe un point dans $dom(f) \cap ImA$. Alors d'après théorème 16.3 ([17])

$$(f \circ A)^*(x^*) = \inf\{f^*(y^*) \mid A^* y^* = x^*\}$$

où inf est atteint pour chaque x^* tel que $(fA)^*(x^*) \neq +\infty$. donnons $x^* \in \partial(f \circ A)(x)$, d'après la propriété (2.61) on a :

$$(f \circ A)^*(x^*) + (f \circ A)(x) = \left\langle x^*, x \right\rangle,$$

il existe y^* tel que $A^* y^* = x^*$ et $(f \circ A)(x) + f^*(y^*) = \left\langle A^* y^*, x \right\rangle$,

$$(f \circ A)(x) + f^*(y^*) = \left\langle y^*, Ax \right\rangle,$$

donc, $y^* \in \partial f(Ax)$.

$$D'où \qquad x^* = A^* y^* \in A^* \partial f(Ax). \qquad C.Q.F.D.$$

□

Lemme 2.74.

Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe différentiable sur son domaine , alors

$$\forall x \in int(dom(f)) \quad \partial f(x) = \{\nabla f(x)\}.$$

Lemme 2.75.

Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction tel que $f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x)$ où f_1, f_2 deux fonctions convexes continues ou s.c.i et $(x, \alpha_1, \alpha_2) \in \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^+$, alors

$$\partial f(x) = \{\eta \in \mathbb{R}^n, \exists(\eta_1, \eta_2) \in \partial f_1(x) \times \partial f_2(x) \ et \ \eta = \eta_1 + \eta_2\}.$$

Lemme 2.76.

Soient $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe fermé, et $g : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ une fonction convexe croissante, alors pour tout $x \in int(dom(f))$ on a

$$\partial g \circ f(x) = \{\eta_1\eta_2 \ , \ \eta_1 \in g(f(x)) \ , \ \eta_2 \in f(x)\}$$

2.4 La transformation de Legendre-Fenchel.

une nouvelle transformée de fonction, portant le nom de W. Fenchel et A.M. Legendre (l'intervention de ce deuxième nom sera expliquée un peu plus loin). Comme cela a déjà été dit en début de ce chapitre, dès que nous parlerons d'une fonction $f : E \to \overline{\mathbb{R}}$, il s'agira d'une fonction non identiquement égale à $+\infty$ et minorée par une fonction affine continue :

$$f(x) \geqslant \langle s_0, x\rangle - r_0 \ \textit{pour tout} \ \ x \in \mathbb{R}^n, \tag{2.2}$$

pour un certain $s_0 \in \mathbb{R}^n$et un certain $r_0 \in \mathbb{R}$.

Définition 2.77.

La transformée de Legendre-Fenchel de f est la fonction f^* définie sur $(\mathbb{R}^n)^* = \mathbb{R}^n$ de la manière suivante :

$$\forall s \in (\mathbb{R}^n)^*, \ f^*(s) := \sup\left[\langle s, x\rangle - f(x)\right].$$

Autres appellations pour f^* : conjuguée de f, polaire de f.

Une première interprétation économique de $f^*(s)$: Supposons qu'un bien x soit vendu au prix s et qu'il ait coûté $f(x)$ à produire ; la meilleure marge en vendant au prix s, parmi toutes les quantités x de biens pouvant être produites, est $\sup\limits_{x\in\mathbb{R}^n}\left[\langle s, x\rangle - f(x)\right]$.
Autre lecture de la Définition (2.77) :

$$-f^*(s) = \inf_{x\in\mathbb{R}^n}\left[f(x) - \langle s, x\rangle\right].$$

Ainsi, $f^*(s)$ est, au signe près, le résultat de la minimisation de f perturbée par une forme linéaire continue $\langle s, .\rangle$.

Avec les hypothèses sur f faites dès le début, f^* n'est pas identiquement égale à $+\infty$ (en effet, $f^*(s_0) < +\infty$ pour la pente s_0 de (2.2), et ne prend jamais la valeur $+\infty$). De plus, par définition-construction, f^* est toujours une fonction convexe s.c.i. (pour la topologie $\sigma(\mathbb{R}^n, \mathbb{R}^n)$). Il suffit pour voir cela d'écrire f comme le supremum d'une famille de fonctions affines continues :

$$f^* := \sup\left[\left\langle ., x\right\rangle - f(x)\right]$$

On se rappelle qu'avec la transformée de Fourier $\mathscr{F}f$ de f, on a $\mathscr{F}f(0) = \int_{\mathbb{R}^n} f(x)dx$.
Avec la transformée de Legendre-Fenchel, on a quelque chose de similaire :

$$f^*(0) = -\inf_{x\in\mathbb{R}^n} f(x).$$

Exemple 2.77.1. ► *$f : \mathbb{R} \to \mathbb{R}$ définie par $f(x) = \frac{1}{p}|x|^p$, où $p > 1$.*
Alors, en désignant par q le "réel conjugué de p", i.e tel que $\frac{1}{p} + \frac{1}{q} = 1$,

$$f^*(s) = \frac{1}{q}|s|^q \textit{ pour tout } s \in \mathbb{R}.$$

► *$f : \mathbb{R} \to \mathbb{R}$ définie par $f(x) = -\ln(x)$ si $x > 0$, $+\infty$ si $x \leqslant 0$.Alors*

$$f^*(s) = -ln(-s) - 1 \textit{ si } s < 0\ , +\infty \textit{ ailleurs.}$$

► *$f : \mathbb{R}^n \to \mathbb{R}$ définie par $f(x) = \frac{1}{2}\left\langle Ax, x\right\rangle$, où A est supposée définie positive. Alors, f^* a la même allure que f :*

$$f^*(s) = \frac{1}{2}\left\langle A^{-1}s, s\right\rangle.$$

► *$f : \mathbb{R} \to \mathbb{R}$ définie par $f(x) = \exp(x)$. Alors :*

$$f^*(s) = s\ln(s) - s \textit{ si } s > 0,\ 0 \textit{ si } s = 0,\ +\infty \textit{ si } s < 0.$$

L'inégalité suivante vient immédiatement de la définition-construction de f^* :

Proposition 2.78.

Pour tout $x \in \mathbb{R}^n$ et $s \in \mathbb{R}^n$, $\left\langle s, x\right\rangle \leqslant f(x) + f^*(s)$.

Proposition 2.79.

pour tout fonction $f : \mathbb{R}^n \to \mathbb{R}$, f^* est convexe fermée.

2.5 Application à l'optimisation

2.5.1 Cas abstrait

Proposition 2.80 (CNS).

Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe. Alors x_0 est un minimum global de f sur $\mathbb{R}^n$ si et seulement si $0 \in \partial f(x_0)$.

Remarque 2.80.1. *Cette condition est donc la CNS pour la minimisation d'une fonction convexe. Elle présente de nombreux mérites.*

1. *Lorsque f est G-différentiable en x_0, on retrouve la CNS usuelle : $D_G f(x_0) = 0$.*
2. *Bien entendu, on peut calculer $\partial f(.)$ partout, ce qui n'est pas le cas de $D_G f(.)$.*
3. *Elle permet de traiter les contraintes convexes (c.f. ci dessous).*

maintenant à minimiser une fonction convexe f sur un ensemble convexe D. Cela revient à minimiser la fonction convexe $f_D = f + i_D$. Si l'on suppose que f est continue en au moins un point de D, on a pour tout x, donc la condition nécessaire et suffisante pour que x_0 soit un minimum sur D de f est :

$$0 \in \partial f(x_0) + N(D, x_0).$$

Au passage, si de plus f est G-différentiable en x_0. on retrouve la CNS usuelle :

$$-D_G f(x_0) \in N(D, x_0).$$

2.5.2 Cas explicite

Soient $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ et $g_i : \mathbb{R}^n \to \mathbb{R},\ i = \overline{1, ..., n}$, sont convexes.
Considérons le problème

$$\begin{cases} \min f(x) \\ g_i(x) \leqslant 0, \ i = \overline{1, ..., n}, \\ x \in \mathbb{R}^n \end{cases} \tag{2.3}$$

Théoreme 2.81 ([14]).

Si a est minimum local de problème (2.3), alors il existe des multiplicateurs $\lambda_i \in \mathbb{R}^+$ tel que :

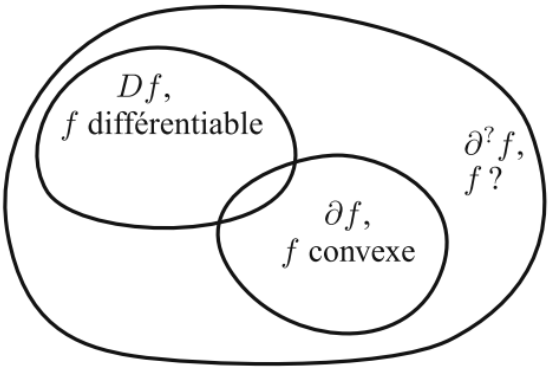
$(a)-$ $g_i(a) \leqslant 0$ et $\lambda_i g_i(a) = 0$ pour $i = \overline{1, ..., n}$.

$(a)-$ $0 \in \lambda_0 \partial f(a) + \sum_{i=1}^{n} \lambda_i \partial g_i(a)$.

Chapitre 3

SOUS DIFFÉRENTIEL AU SENS DE CLARKE

Les problèmes variationnels ou d'optimisation font intervenir, de manière naturelle, des fonctions qui ne sont pas différentiables. Certes ces fonctions sont différentiables en la plupart des points, mais ne le sont pas aux "points intéressants". Les objectifs d'un calcul différentiel généralisé sont, au moins : "que ça fonctionne" (eu égard aux opérations usuelles de l'analyse) ; "que ça s'utilise" (Algorithmique, problèmes applicatifs). En démarrant ce chapitre, il y a déjà deux contextes dans lesquels on sait évoluer et qu'il s'agit d'englober et de généraliser : celui des fonctions différentiables et celui des fonctions convexes. Ainsi, tout nouvel objet mathématique visant à "différentier des fonctions non différentiables" devra se réduire à la différentielle usuelle dans le cas des fonctions différentiables (ou du moins continûment différentiables) et à celui de sous-différentiel dans le cas de fonctions convexes.



Dans cette cohérence ascendante cherchant à toucher une classe de fonctions aussi vaste que possible, nous sommes conduits à faire des choix parmi tous les sous-différentiels généralisés proposés par les mathématiciens lors des trente-cinq dernières années. Ces choix dépendent de ce qu'on veut faire :

– S'il s'agit de traiter les problèmes variationnels ou d'optimisation dans leur formulation abstraite, dériver des conditions nécessaires d'optimalité par exemple, il y a alors plusieurs sous-différentiels généralisés possibles.

– S'il s'agit d'algorithmique pour traiter des problèmes non différentiables, il n'y a pas besoin de généralité maximale mais bien de disposer d'un outil avec des règles de calcul robustes. Dans ce but, nous consacrerons la ce chapitre au gradient généralisé ou sous-différentiel généralisé au sens de F. Clarke.

3.1 Fonction lipschitzienne

Les preuves des résultats donnés dans ce paragraphe peuvent être trouvées dans Soit $\Omega \subset \mathbb{R}^n$ un ouvert.

Définition 3.82.

Une fonction $f : \Omega \to \mathbb{R}$ est dite localement lipschitzienne de rang k, au voisinage de x si :

$$\forall \varepsilon > 0, \forall (y,z) \in B(x;\varepsilon)\ |f(y) - f(z)| \leqslant k\|y - z\|$$

Une fonction localement Lipschitzienne en x présente ainsi, un taux d'accroissement qui est borné dans un voisinage de x. D'autre part , une fonction localement Lipschitzienne en x n'est pas forcement différentiable en x.

La classe des fonctions localement Lipschitzienne sur Ω est remarquablement stable pour toutes les opérations usuelles de l'analyse.

Proposition 3.83.

Soit $f, g : \Omega \to \mathbb{R}$ deux fonctions localement lipschitzienne sur Ω, soit $\lambda, \mu \in \mathbb{R}$.Alors

1. $\lambda f + \mu g$ est localement lipschitzienne sur Ω.
2. fg est localement lipschitzienne sur Ω.
3. $|f|$ est localement lipschitzienne sur Ω
4. Si de plus $f(x) \neq 0$, $\forall x \in \Omega$, $\dfrac{1}{f}$ est localement lipschitzienne sur Ω.
5. $\max(f, g)$ et $\min(f, g)$ est localement lipschitzienne sur Ω

Parmi les classes de fonctions déjà rencontrées et qui sont localement Lipschitz,on a :

Proposition 3.84.

1. Soit Ω convexe et $f : \Omega \to \mathbb{R}$ convexe (ou concave), continue sur Ω, alors f localement lipschitzienne sur Ω.
2. Soit $f : \Omega \to \mathbb{R}$ continûment différentiable sur Ω, alors f localement lipschitzienne sur Ω.

Exemple 3.84.1.
1— *Une application est 0-lipschitzienne si et seulement si elle est constante [**?**].*
2— *Il existe néanmoins des fonctions différentiables sur Ω qui ne sont pas localement Lipschitz sur Ω par exemple la fonction convexe $f : x \in [0,1] \to -\sqrt{x}$ est dérivable mais pas localement Lipschitz au voisinage de 0. car $\lim\limits_{\substack{x \to 0 \\ x \neq 0}} f'(x) = -\infty$ la constante de Lipschitz est explosé au bord du domaine.*

Théoreme 3.85.

Une fonction $f : \Omega \to \mathbb{R}$ localement Lipschitz sur Ω est différentiable presque partout sur Ω

Une propriété est dite vrai presque partout sur Ω si la propriété vrai en tous points de Ω, à l'exception de ceux d'un ensemble de mesure de Lebesgue* nulle.

Rappelons qu'ici les différentiabilités au sens de Gâteaux ou Fréchet sont équivalentes (cf. Chapitre 1). De plus, le caractère localement Lipschitz de f fait que $\nabla f(x')$, là où il existe dans un voisinage de x, est "contrôlé" par la constante de Lipschitz, il "n'explose pas". En termes plus mathématiques, pour tout $x \in \Omega$, il existe $r > 0$ et $L \geqslant 0$ tels que

$$\left\{\nabla f(x') \mid x' \in \overline{B}(x,r) \; et \, f \;\; est \;\; différentiable \;\; en \;\; x'\right\} \subset \overline{B}(0,L).$$

Ceci est dû au fait que, pour x' voisin de x,

$$\left|\frac{f(x'+td)-f(x')}{t}\right| \leqslant L\|d\|. \tag{3.1}$$

Quand on pense différentiabilité de f en x, on pense inévitablement à des quotients différentiels

$$\frac{f(x'+td)-f(x')}{t}, \; où \; d \in \mathbb{R}^n \; et \; t > 0. \tag{3.2}$$

Qu'en faire lorsque f n'est pas différentiable en x ? On a beau essayer des limites supérieures ou inférieures quand $t \to 0^+$ à partir de (3.2), on récupère à l'arrivée une sorte de dérivée directionnelle généralisée $f^{(1)}(x;d)$ dont la seule propriété tangible est qu'elle est positivement homogène en la direction d : $f^{(1)}(x;\alpha d) = \alpha f^{(1)}(x;d)$ pour $\alpha > 0$.
Une approche différente, décisive quant à l'utilité du concept qui va suivre, consiste à considérer le quotient différentiel de (3.2) pas en x seul mais dans un voisinage de x. Elle est due à F. Clarke (1973) et a marqué le renouveau de ce qu'on appelle parfois l'Analyse non-lisse (Nonsmooth analysis en anglais).

3.2 La dérivée directionnelle généralisée

On considère toujours, une fonction $f : \Omega \in \mathbb{R}^n \to \mathbb{R}$ localement Lipschitz sur l'ouvert Ω de $\mathbb{R}^n$ et $x \in \Omega$.
Contrairement au cas convexe, l'hypothèse que la fonction $f : \mathbb{R}^n \to \mathbb{R}$ est localement Lipschitzienne en $x \in R^n$ n'est pas suffisante pour l'existence des dérivées directionnelles de f. C'est pourquoi on doit généraliser le concept de dérivée directionnelle.

$*$. Il existe une unique mesure λ sur la tribu Borélien $\mathfrak{B}(\mathbb{R})$ qui étend la notion de longueur des intervalles :
$\lambda([a,b]) = b-a \quad \lambda$ est σ-finie. On l'appelle la mesure de Lebesgue.

Définition 3.86.

La dérivée directionnelle généralisée de f en x, au sens de Clarke, est la limite supérieure suivante

$$f_c(x_0;d) := \limsup_{\substack{x\to x_0\\ \lambda\to 0^+}} \frac{f(x+\lambda d)-f(x)}{\lambda}. \tag{3.3}$$

Comme cela était attendu, $f_c(x_0;0) = 0$ et $f_c(x_0;\alpha d) = \alpha f_c(x_0;d)$ pour tout $\alpha > 0$. Plus surprenant, et essentiel pour la suite des évènements, est la propriété de convexité que voici :

Propriété 3.87.

La fonction $d \in \mathbb{R}^n \to f_c(x_0;d)$ est convexe continue sur $\mathbb{R}^n$. De plus :

$$\forall d \in \mathbb{R}^n,\ |f_c(x_0;d)| \leqslant L\|d\|,$$

Où L est une constante de Lipschitz pour f dans un voisinage de x.

Démonstration. Puisque $f_c(x_0;.)$ est positivement homogène $f_c(x_0;\alpha d) = \alpha f_c(x_0;d)$ pour tout $d \in \mathbb{R}^n$ et tout $\alpha > 0$), la convexité de $f_c(x_0;.)$ revient à sous-additivité. A-t-on

$$\forall u,v \in \mathbb{R}^n\ \ f_c(x_0;u+v) \leqslant f_c(x_0;u) + f_c(x_0;v)?$$

On a clairement

$$f_c(x_0;u+v) := \limsup_{\substack{x\to x_0\\ \lambda\to 0^+}} \frac{f(x+\lambda(u+v))-f(x)}{\lambda} \tag{3.4}$$

$$\leqslant \limsup_{\substack{x\to x_0\\ \lambda\to 0^+}} \frac{f(x+\lambda u+\lambda v)-f(x+\lambda u)}{\lambda} + \limsup_{\substack{x\to x_0\\ \lambda\to 0^+}} \frac{f(x+\lambda u)-f(x)}{\lambda} \tag{3.5}$$

$$\leqslant f_c(x_0;v) + f_c(x_0;u). \tag{3.6}$$

Comme le montre nettement la démonstration ci-dessus, c'est vraiment cette approche qui a consisté à aller voir "ce qui se passe autour de x" qui a permis d'accéder à la convexité de $f_c(x;.)$. Il suffit $u = \alpha d$ et $v = (1-\alpha)d'$ où $\alpha \in [0,1]$ et $d,d' \in \mathbb{R}^n$.

La majoration dans la (3.87) vient immédiatement du fait que

$$\left|\frac{f(x+\lambda d)-f(x)}{\lambda}\right| \leqslant L\|d\|,$$

pour $\lambda > 0$ assez petit et x voisin de x_0. □

On aurait pu être tenté de prendre une limite inférieure au lieu d'une limite supérieure dans (3.3) :

$$f_{\underline{c}}(x_0; d) := \liminf_{\substack{x \to x_0 \\ \lambda \to 0^+}} \frac{f(x+\lambda d)-f(x)}{\lambda}. \tag{3.7}$$

Cela n'aurait pas changé le font de l'affaire puisque

$$f_{\underline{c}}(x_0; d) = -f_c(x_0; -d).$$

Signalons avant d'aller plus loin que la limite supérieure

$$f_c(x_0; d) := \limsup_{\substack{x \to x_0 \\ \lambda \to 0^+}} \frac{f(x+\lambda d)-f(x)}{\lambda} = \inf_{\substack{\varepsilon>0 \\ r>0}} \sup_{\substack{\lambda \in]0,\varepsilon] \\ x \in \overline{B}(x_0, r)}} \frac{f(x+\lambda d)-f(x)}{\lambda},$$

est "atteinte" par une suite $\{x_k\} \to x_0$ et $\{\lambda_k\} \to 0^+$, c'est-à-dire : Il existe une suite $\{x_k\}$ converge vers x_0 et une suite $\{\lambda\} > 0$ converge vers 0 telle que

$$\limsup_{k \to +\infty} \frac{f(x_k+\lambda_k d)-f(x_k)}{\lambda_k}.$$

Cela peut aider dans certaines démonstrations.

Propriété 3.88.

1. $f_c(x; d) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ est semi-continue supérieurement, autrement :

$$\forall \{x_k\} \to x,\ \forall \{d_k\} \to d,\ \limsup_{k \to +\infty} \leqslant f_c(x; d).$$

2. "Symétrisation " :

$$\forall d \in \mathbb{R}^n, (-f)_c(x; d) = f_c(x; -d).$$

Démonstration. Montrons d'abord (2). Par définition,

$$f_c(x_0;-d) := \limsup_{\substack{x\to x_0\\ \lambda\to 0^+}} \frac{f(x-\lambda d)-f(x)}{\lambda}.$$

Avec le changement de variables $u := x-\lambda d$, le quotient différentiel ci-dessus n'est autre que

$$\frac{(-f)(u+\lambda d)-(-f)(u)}{\lambda}.$$

Donc

$$\limsup_{\substack{u\to x_0\\ \lambda\to 0^+}} \frac{(-f)(u+\lambda d)-(-f)(u)}{\lambda} = (-f)_c(x_0;d)$$

□

Puisque $f_c(x_0;.)$ est convexe et continue sur $\mathbb{R}^n$ (et même Lipschitz sur $\mathbb{R}^n$), positivement homogène, il est tentant de considérer les formes linéaires continues minorant de $f_c(x_0;.)$. ce qui donne naissance au sous-différentiel généralisé (au sens de Clarke) de f en x_0.

3.3 Sous-différentiel de Clarke

Définition 3.89.

Le sous-différentiel généralisé de f en x_0, au sens de Clarke, est

$$\partial_c f(x_0) := \{x^* \in \mathbb{R}^n \mid < x^*, d > \leqslant f_c(x_0;d) \text{ pour tout } d \in \mathbb{R}^n\}.$$

On aurait pu être tenté d'utiliser la fonction concave $f_{\underline{c}}(x_0;.)$ et les formes linéaires continues majorant $f_{\underline{c}}(x_0;d)$. Cela n'aurait rien changé puisque, grâce a la relation $f_{\underline{c}}(x_0;d) = -f_c(x_0;-d)\ \forall d \in \mathbb{R}^n$, il découle

$$\{x^* \in \mathbb{R}^n \mid < x^*, d > \leqslant f_c(x_0;d) \text{ pour tout } d \in \mathbb{R}^n\} = \left\{x^* \in \mathbb{R}^n \mid < x^*, d' > \geqslant f_{\underline{c}}(x_0;d') \text{ pour tout } d \in \mathbb{R}\right.$$

Désormais, c'est toute la machinerie de l'analyse convexe (chapitre 2) qui va être appliquée à $\partial_c f(x_0)$ via la fonction convexe $f_c(x_0;.)$.

Propriété 3.90.

(i)– $\partial_c f(x_0)$ est un convexe $\sigma(\mathbb{R}^n, \mathbb{R}^n)$-compact[a] non vide de $\mathbb{R}^n$, sa fonction d'appui $f_c(x_0; .)$, i.e.

$$\forall d \in \mathbb{R}^n,\ f_c(x_0; d) = \sup_{x^* \in \partial_c f(x_0)} < x^*, d > .$$

(ii)– Si f continûment différentiable sur Ω, alors :

$$\partial_c f(x_0) = \{Df(x)\} \textit{ pour tout } x \in \Omega$$

(iii)– Si f est convexe et continue sur Ω, alors

$$\partial_c f(x_0) = \partial f(x_0) \left[\textit{sous-différentiel de } f \textit{ en } x_0 \textit{ au sens d'analyse convexe} \right].$$

(iv)– Si $f = \max(f_1, ..., f_n)$,où chaque f_i continûment différentiable sur Ω, alors : $\partial_c f(x_0) = co\{Df_i(x) \mid f_i(x) = f(x)\}$.

a. Compact par la topologie faible-* qui est la moins fine rendant toutes les applications $\xi_x : \mathbb{R}^n \to \mathbb{R}$ continues.

Propriété 3.91.

Si $f : \Omega \to \mathbb{R}^n$ est localement Lipschitzienne sur Ω, alors pour tout $x \in \Omega$:

$$f_c(x_0; d) = \limsup_{x \to x_0} \left\{ \left\langle \nabla f(x), d \right\rangle \mid f \textit{ différentiable en } x_0 \right\}. \tag{3.8}$$

Une version un peu plus générale que (3.8) est comme suite. Supposons que f admette en tout point x au voisinage de x_0, une dérivée directionnelle usuelle :

$$f'(x; d) = \lim_{\lambda \to 0^+} \frac{f(x + \lambda d) - f(x)}{\lambda},\ d \in \mathbb{R}^n.\ \textit{Alors, pour tout } d \in \mathbb{R}^n,$$
$$f_c(x_0; d) = \limsup_{x \to x_0} f'(x; d).$$

La dérivée directionnelle généralisée $f_c(x_0; d)$ apparaît donc comme une "version régularisée (en allant regarder autour de x_0 de la dérivée directionnelle usuelle $f'(x_0; d)$. Donnons quelques exemple d'illustrations diverses.

Exemple 3.91.1. *Soit* $f : x \in \mathbb{R} \mapsto f(x) = -|x|$. *Alors,* $\partial_c f(0) = [-1, 1]$. *De manière plus générale, si* $f : \Omega \to \mathbb{R}$ *est concave et continue sur* Ω*, alors*

$$\partial f(x) = \{x^* \in \mathbb{R}^n \mid f(y) \leqslant f(x) + \left\langle x^*, y - x \right\rangle\},$$

c'est-à-dire le sur-différentiel de f *en* x.

Exemple 3.91.2. *Soit* $f : \mathbb{R} \to \mathbb{R}$ *définie par :*

$$f(x) = x^2 \sin\left(\frac{1}{x}\right) \text{ si } x \neq 0,\ f(0) = 0.$$

f *est dérivable sur* $\mathbb{R}$ *mais pas continûment dérivable sur* $\mathbb{R}$. *De fait,*

$$f'(x) = 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) \text{ si } x \neq 0,$$

laquelle dérivée n'a pas de limite quand $x \to 0$.
Or f *localement Lipschitz on a* $\partial_c f(0) = [-1, 1]$. *Ainsi, alors que* $f'(0) = 0$.

D'un manière plus générale, si la fonction localement Lipschitz $f : \Omega \to \mathbb{R}$ est Fréchet-différentiable en $x \in \Omega$, $Df(x) \in \partial_c f(x)$. Ceci n'est une faiblesse car, rappelons-nous que (Propriétés 3.72, (ii)) $\partial_c f(x) = \{Df(x)\}$ en tout $x \in \Omega$ lorsque f est continûment différentiable sur Ω

3.4 Règle de calcul de sous-différentiel

Les règles de calcul basiques sur les sous-différentiels généralisés sont directement dérivées des règles de calcul sur les sous-différentiels de fonctions convexes (du Chapitre 2). En effet, $\partial_c f(x)$ est le sous-différentiel en 0 de la fonction convexe positivement homogène $f_c(x; .)$:

$$x^* \in \partial_c f(x) \Longleftrightarrow f_c(x; 0) + \left\langle x^*, d - 0 \right\rangle \text{ pour tout } d \in \mathbb{R}^n.$$

Proposition 3.92.

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction localement Lipschitz.

1. $\partial(\alpha f)(x) = \alpha \partial f(x)$ pour tout $\alpha \in \mathbb{R}$. En particulier,
$$\partial(-f)(x) = -\partial f(x).$$
2. $$\partial(f+g)(x) \subset \partial f(x) + \partial g(x).$$
3. Si $\overline{x}$ est un minimiseur local ou un maximiseur local de f, alors :
$$0 \in \partial f(\overline{x}.$$
4. $$Si\ \{x_k\} \to x,\ x_k^* \in \partial f(x_k)\ et\ si\ x_k^* \to x^*$$
(pour la topologie faible-*, $\sigma(\mathbb{R}^n, \mathbb{R}^n)$, alors
$$x^* \in \partial f(x).$$
5. Théorème des accroissement finis (ou de la valeur moyenne) : Supposons $[x, y]\Omega$; il existe alors $\lambda \in]0, 1[$ tel que
$$f(y) - f(x) \in \left\langle \partial[x + \lambda(y-x)], y - x \right\rangle$$
$$\left(:= \left\{ \left\langle x^*, y - x \right\rangle \mid x^* \in \partial f[x + \lambda(y-x)] \right\} \right)$$
6. Si $f = \max(f_1, ..., f_k)$,
$$\partial f(x) \subset co\,\{\partial f_i(x) \mid \ tels\ que\ f_i(x) = f(x)\}.$$
7. Un exemple de règle de calcul sur les fonctions composées : Supposons que $f = g \circ F$, avec F continûment différentiable sur Ω_1 et g localement lipschitzienne sur Ω_2. alors :
$$\partial f(x) = [DF(x)]^* \partial g[F(x)], \tag{3.9}$$
où $[DF(x)]^* : \mathbb{R}^n \to \mathbb{R}^m$ désigne l'adjointe de la différentielle $DF(x) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$.
Il y a égalité en (3.9) lorsque $DF(x)$ est surjective.
$$\begin{tikzcd} \Omega_1 \subset \mathbb{R}^n \arrow[rr, "F"] \arrow[dr, "f = g\circ F"'] & & \Omega_2 \subset \mathbb{R}^m \arrow[dl, "g"] \\ & \mathbb{R} & \end{tikzcd}$$

- Démontrer en premier lieu des relations d'inégalité entre dérivées directionnelles généralisées ;
- Appliquer les règles de calcul sous-différentiel(de fonction convexes) à ces fonctions dérivées directionnelle généralisées ;
- En déduire des règles de comparaison, sous forme d'inclusion,entre sous différentiels généralisés.

Démonstration. Nous n'en esquisserons que quelques-unes pour illustrer le cheminement présenté plus haut.

1. Pour démontrer (1), on utilise le fait que $(\alpha f)_c(x;d) = f_c(x;\alpha d)$ pour tout $d \in \mathbb{R}^n$.

2. On commence par démontrer que

$$(f+g)_c(x;d) \leqslant f_c(x;d) + g_c(x;d) \ \textit{pour tout } d \in \mathbb{R}^n.$$

3. En un point $\overline{x}$ minimiseur local de f,

$$f_c(\overline{x};d) \geqslant 0 \textit{pour tout } \ d \in \mathbb{R}^n.$$

4. On commence par démontrer que

$$f_c(\overline{x};d) \leqslant \max\left\{f_c^i(\overline{x};d) \ | \ i \textit{ tels que} f^i(\overline{x}) = f(\overline{x})\right\} \ \textit{pour tout } \ d \in \mathbb{R}^n.$$

□

Quelques commentaires :

– L'inclusion (2) de (4) , et non l'égalité, peut surprendre. En fait, il n'en est rien, c'est l'égalité qui aurait été étonnante, vu la généralité des fonctions en jeu et la manière "tarabiscotée" dont le sous-différentiel généralisé est construit. Pour prendre un exemple simple, si $f(x) = -g(x) = |x|$,

$$\partial f(0) = \partial g(0) = [-1,+1], \ \textit{alors que } \partial(f+g)(0) = 0.$$

– La relation (5) est très simple, et pourtant elle est très utile, ne serait-ce qu'en algorithmique où on est fréquemment en situation de comparer $f(x_k + \lambda_k d_k)$ à $f(x_k)$. Or

$$f(x_k + \lambda_k d_k) = f(x_k) + \lambda_k \langle s_k, d_k \rangle,$$

où $s_k \in \partial f(\theta_k)$ et θ_k est un point intermédiaire entre x_k et $x_k + \lambda_k d_k$.

– Avoir des égalités dans les inclusions des règles de calcul 6.12 requiert, apriori, des hypothèses fortes sur le comportement des fonctions au voisinage de x. L'une d'entre elles est que, pour les fonctions f en jeu, la dérivée directionnelle usuelle $f'(x;.)$ existe et coïncide avec la dérivée directionnelle généralisée $f_c(x;.)$. Certes, ceci est vérifié pour les fonctions continûment différentiables ou les fonctions convexes, mais a peu de chances de l'être pour une fonction non convexe qui ne serait pas différentiable en x.

3.5 Application à l'optimisation

3.5.1 Sans contraintes

Théoreme 3.93 ([11]).

On suppose que $x_0 \in int(dom(f))$ est un minimum local de f sur $\mathbb{R}^n$. Alors pour tout $d \in \mathbb{R}^n$,

$$f_c'(x_0; d) \geqslant 0,$$

ou encore

$$0 \in \partial_c f(x_0).$$

3.5.2 Avec contraintes

Cône tangent de Clarke

Soit $S \cap \mathbb{R}^n$, nous désignons par $d_S(.)$ la fonction distance définie par :

$$d_S(x) = \inf_{y \in S} \|x - y\|, \ \forall x \in \mathbb{R}^n.$$

$d_S(.)$ est localement Lipschitzienne de rapport 1, par suite

$$\partial_c d_S(x) \neq \emptyset, \ \forall x \in \mathbb{R}^n.$$

Définition 3.94.

Soit $x \in \mathbb{R}^n$, on appelle Cône tangente de Clarke à S en x, l'ensemble noté :

$$T_c(S,x) = \left\{ d \in \mathbb{R}^n \mid f_c'(x;d) \leqslant 0 \right\} \; où \; f = d_S(.).$$

Et $N_c(S,x) = \overline{\cup_{\lambda > 0} \lambda \partial_c d_S(x)}$ est appelé Cône normal de clarke à S en x.

Proposition 3.95.

$T_c(S,x)$ est un Cône convexe fermé de $\mathbb{R}^n$, et donc

$$T_c(S,x) = [N_c(S,x)]^\circ$$

Maintenant on minimise une fonction f localement Lipschitzienne sur un ensemble S. Cela revient à minimiser globalement la fonction $f_S = f + kd_S(.), k > 0$. Donc la condition nécessaire pour que x_0 soit un minimum local de f sur S est :

$$0 \in \partial f(x_0) + N_c(S, x_0).$$

3.5.3 Cas explicite

Soit f, g_i $i = \overline{1,...,n}$ localement Lipschitzienne et considérons le problème suivant :

$$\begin{cases} \min f(x) \\ g_i(x) \leqslant 0, \; i = \overline{1,...,n}, \\ x \in \mathbb{R}^n \end{cases} \tag{3.10}$$

On appelle Lagrangien du problème (3.10) la fonction $L : \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+^n \to \mathbb{R}$, donnée par :

$$L(x,\lambda,\alpha) = \lambda f(x) + \sum_{i=1}^{n} \alpha_i g_i(x), \; \forall x \in \mathbb{R}^n$$

Théoreme 3.96 ([11]).

Si a est minimum local de (3.10) alors :

$$\begin{cases} \exists \overline{\lambda} \geqslant 0, \exists \overline{\alpha} \in \mathbb{R}^n_+ \textit{ non tous nuls} \\ \overline{\alpha_i} = 0 \textit{ si } i \notin I(a), \left(\sum_{i=1}^{n} \overline{\alpha_i} g_i(a) = 0 \right) \\ \textit{et } \; 0 \in \partial_c L(a, \overline{\lambda}, \overline{\alpha}) \subset \overline{\lambda} \partial_c f(a) + \sum_{i=1}^{n} \overline{\alpha_i} \partial_c g_i(a) \end{cases} \tag{3.11}$$

Remarque 3.96.1. *Ce théorème redonne le théorème (2.81) de* ***Rockafellar*** *dans le cas convexe ainsi que le théorème de* ***kuhn-tuker*** *dans le cas différentiable.*

3.6 Fonctions D.C

Dans cette section nous sommes attachés à étudier la calasse de fonctions importantes en optimisation non convexe, non différentiable : les fonctions qui peuvent s'écrire comme différence de fonctions convexes (fonctions D.C).

Définition 3.97.

Soient C un ensemble convexe non vide de $\mathbb{R}^n$ et $f : C \to \overline{\mathbb{R}}$ une fonction.

- ► La fonction f est dite D.C sur C si elle peut s'écrire comme différence de deux fonctions g, h convexes sur C,

$$f(x) = g(x) - h(x), \qquad \forall x \in C.$$

- ► On dit que $g - h$ est une décomposition D.C de f et g, h sont des composantes D.C de f.

$DC(\mathbb{R}^n) :=$ ensemble des fonctions qui s'écrivent comme des différences de fonctions convexes sur $\mathbb{R}^n$.

Propriété 3.98.

$DC(\mathbb{R}^n)$ est stable par les propriétés usuelles de l'Analyse telles que : addition, soustraction, multiplication, maximum d'un nombre fini de fonctions, etc.

Proposition 3.99.

Si f une fonction D.C sur un ensemble C et f admet une décomposition D.C comme $g-h$ alors pour tout fonction φ convexe finie sur C.

$$(g+\varphi)-(h+\varphi) \quad \textit{fournit aussi une décomposition D.C de } f$$

Proposition 3.100.

Tout fonction de D.C admet une infinité de décomposition D.C.

3.6.1 Application à l'optimisation

Considérons le problème suivant :

$$\begin{cases} minimiser f(x) := g(x)-h(x) \\ x \in \mathbb{R}^n \end{cases} \tag{3.12}$$

où g, h sont des fonctions convexes s.c.i propre sur $\mathbb{R}^n$. la deuxième fonction h est partout finie et continue sur $\mathbb{R}^n$. Si ça n'est pas le cas, comme nous minimisons dans (3.12), nous donnons la priorité à $+\infty$ c'est-à-dire que nous adoptons la règle de calcul $(+\infty)-(+\infty) = +\infty$ pour le cas ou cela se produirait.

Définition 3.101.

Soient g, h deux fonctions convexes s.c.i propres.
Un point $x^* \in dom(g) \cap dom(h)$ est un minimum local de $g-h$ s'il existe un voisinage (une boule) $\mathbf{V}_{x^*}$ de x^* tel que

$$(g-h)(x^*) \leqslant (g-h)(x), \;\; \forall x \in \mathbf{V}_{x^*}.$$

Proposition 3.102.

Soit $\overline{x}$ un minimum local de $f = g - h$ sur $\mathbb{R}^n$. Alors :

$$\partial h(\overline{x}) \subset \partial g(\overline{x}) \tag{3.13}$$

Démonstration. Pour $x \in \overline{B}(\overline{x}, r)$ on a :

$$f(x) = g(x) - h(x) \geqslant f(\overline{x}) = g(\overline{x}) - h(\overline{x}),$$

soit encore

$$g(x) - g(\overline{x}) \geqslant h(x) - h(\overline{x}).$$

Soit $\overline{x}^* \in \partial h(\overline{x})$. D'après la définition de sous-différentiel on trouve

$$h(x) - h(\overline{x}) \geqslant \left\langle \overline{x}^*, x - \overline{x} \right\rangle$$

donc,

$$g(x) - g(\overline{x}) \geqslant \left\langle \overline{x}^*, x - \overline{x} \right\rangle \quad \forall x \in \overline{B}(\overline{x}, r).$$

Grâce à la convexité de g qui "globalise" les inégalités, la relation au dessus s'étend à tout $\mathbb{R}^n$.

$$\overline{x}^* \in \partial g(\overline{x}).$$

□

La condition (3.13) est "orientée" vers la minimisation, et la condition nécessaire vérifiée par un maximum local $\overline{x}$ serait $\partial h(\overline{x}) \subset \partial g(\overline{x})$. Pour symétriser quelque peu les choses, **Toland** a eu l'idée d'introduire la notion de point critique (ou stationnaire) suivante.

Définition 3.103.

Un point $\overline{x} \in \mathbb{R}^n$ est appelé T-critique (ou T-stationnaire) de $f = g - h$ lorsque $\partial g(\overline{x}) \cap \partial h(\overline{x}) \neq \emptyset$.

Lorsque $\overline{x}$ est un point T-critique, la valeur $f(\overline{x}) = g(\overline{x}) - h(\overline{x})$ est appelée valeur T-critique de f. Évidemment, cette notion de T-criticité de f dépend de la décomposition D.C. $f = g - h$ de f. Comme conséquence de la Proposition précédente, nous avons :

• Si $\overline{x}$ est un minimum local de $f = g - h$ et si $\partial h(\overline{x}) \neq \emptyset$,alors $\overline{x}$ est un point T-critique de f.

• Si $\overline{x}$ est un maximum local de $f = g - h$ et si $\partial g(\overline{x}) \neq \emptyset$,alors $\overline{x}$ est un point T-critique de f.

3.7 Problème ouvert

Des questions importantes susceptibles de constituer des problèmes ouverts pour la problématique que nous avons abordée.

- tout fonction D.C admet plusieurs décompositions de D.C, est ce qu'il existe une décomposition meilleure (optimale) parmi ces décompositions.

Bibliographie

[1] D. Azé. Éléments d'analyse convexe. Éditions Ellipses, Paris, 1997.

[2] J.P. Aubin et I. Ekeland. Applied non linear analysis, Wiley, New York. 1984.

[3] S. Artstein-Avidan and V. Milman. "The concept of duality in convex analysis, and the characterization of the Legendre transform". Annals of Mathematics, 169, p. 661–674,2009.

[4] S. Balac et L. Chupin, Analyse et algèbre : cours de mathématiques de deuxième année avec exercices corrigés et illustrations avec Maple, PPUR, p. 558 2008 .

[5] J. Baptiste et Hiriart-Urruty, Bases, outils et principes pour l'analyse variationnelle, Institut de Mathématiques de Toulouse,vol.70, pp. 141-168 Springer Link 2013.

[6] D. Bertsekas, Convex analysis and Optimization, Massachusetts institute of Technology, 2003.

[7] H. Brezis, Analyse fonctionnelle, Théorie et application Edition Masson, 1983.

[8] J. M. Borwein and D. Preiss, A smooth variational principale with applications to subdifferentiability of convex Functions. Trans. Amer. math. Soc. 303 , pp 371-381, 1987.

[9] F. H. Clarke, Generalized gradients and applications. Transactions of the American Mathematical Society, 205, 247–247, 1975.

[10] F.H. Clarke, Necessary conditions for nonsmooth problems in optimal control and the calculus of variations, Ph.D Thesis, University of Washington 1973.

[11] F.H. Clarke, Nonsmooth analysis and Optimization, Proceedings of the International Congress of Mathematicians Helsinki, p. 848-853, 1978.

[12] A. EL HILALI ALAOUI, Fonctions s' écrivant comme différence de deux fonctions convexes : Fonctions primal lower-nile. 1996.

[13] I. Ekeland and R. Temam. Convex analysis and variational problems. Reprinted by SIAM Publications, Classics in, Applied Mathematics, 28, 1999.

[14] A. D. Ioffe, Necessary and sufficient conditions for a local minimum 1,2,3, SIAM J.Control Optim. Vol.17, N2, March 1979.

[15] B. S. Mordukhovich, and N. M. Nam, An Easy Path to Convex Analysis and Applications. Synthesis Lectures on Mathematics and Statistics, 6(2), pp 1–218, 2013.

[16] J. P. penot, Calcul sous-differentiel et optimisation, Journal of functional analysis, N27, pp. 248-276, 1978 .

[17] R.T. Rockafellar, Convex Analysis, Princeton University Press, 1970.

[18] R. T. Rockafellar, and R. J. B. Wets, Variational Analysis Grundlehren Der Mathematischen Wissenschaften, vol. 317 1998.

[19] R. T. Rockafellar, Generalized directional derivatives and subgradients of nonconvex functions, Canad.J. Math. 32, pp. 257-280, 1980.

[20] J. S. Treiman, Clarke's gradients and epsilon-subgradients in Banach spaces, Trans. Amer. Math. Soc. vol. 294 pp 65-78, 1986.

[21] J. S. Treiman, Shrinking generalized gradient, Nonlinear Anal. Theory Math. Appl. 12, pp. 1429-1449, 1988.

[22] M. Willem, Analyse Convexe et Optimisation, ciaco s.c, 1987.