UNIVERSITY OF ALBERTA

PROTO-DIFFERENTIATION AND EPI-DIFFERENTIATION

BY

SANATH KUMARA BORALUGODA ©

A thesis
submitted to the Faculty of Graduate Studies and Research
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

EDMONTON, ALBERTA

FALL 1994

National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-315-95007-2

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR	: **Sanath Kumara Boralugoda**
TITLE OF THESIS	: **Proto-differentiation and Epi-differentiation**
DEGREE	: **Master of Science**
YEAR THIS DEGREE GRANTED	: **Fall 1994**

S. K. Boralugoda

Department of Mathematics
University of Sri Jayewardenepura
Nugegoda
Sri Lanka.

Date: June 3rd, 1994

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled

PROTO-DIFFERENTIATION AND EPI-DIFFERENTIATION

submitted by **Sanath Kumara Boralugoda** in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE**.

Dr. R. A. Poliquin (Supervisor)

Vaclav Zizler

Dr. V. Zizler (Chairman)

Dr. D. Wiens (External Examiner)
Department of Statistics and Applied
Probability, University of Alberta.

Date : June 27th, 1994.

THIS THESIS IS DEDICATED TO MY PARENTS

ABSTRACT

The convergence of sets is the key ingredient in defining the generalized derivatives of functions and *multifunctions*. Two such notions of set convergences are studied and are used to define *proto-derivatives* and *epi-derivatives*. *Amenable* functions, a central class of functions in finite dimensional optimization theory and applications, are further studied. The definition and calculus results of *fully* amenable functions are extended to Banach spaces. A criteria is given to derive the relationship between second-order epi-derivatives of fully amenable functions and the proto-derivatives of their associated *subgradient* mappings. In particular, a formula of the proto-derivatives of subgradient mappings of a *max-function* is extended.

ACKNOWLEDGEMENTS

I would like to thank Dr. James Muldowney, Graduate Chairman of the Department of Mathematics , for organizing my graduate program at the University of Alberta.

It is with great pleasure that I would like to thank my supervisor Dr. René Poliquin, for his invaluable guidance throughout the preparation of this thesis.

I am also thankful to Dr. Václav Zizler, for his advice and interest in my research.

CONTENTS

CHAPTER 1

INTRODUCTION

Optimization theory has changed immensely, during the last four decades, and generated new kinds of mathematics with far reaching consequences. not only for practical applications but for the very foundations of analysis. The origins of analytic optimization lie in the classical calculus which deals with smooth (continuously differentiable) functions. Recently, attempts were taken to weaken these smoothness requirements and develop a general theory of nonsmooth (not necessarily smooth) analysis. As nonsmoothness occurs naturally and frequently in mathematics and optimization, we are led to study differential properties of non differentiable functions (*subdifferential* properties).

In this thesis, we discuss two such notions of generalized differentiation; namely the *proto-differentiation* of set-valued mappings (*multifunctions*) and the *epi-differentiation* of extended-real-valued functions (they may take $+\infty$). We define these generalized derivatives by replacing the traditional pointwise convergence of difference quotients by more general notions of set convergence. specifically, for multifunctions we replace the pointwise convergence of "difference quotient" multifunctions by the convergence of sets of their graphs and for functions by the sets of corresponding *epigraphs*; which consists of all the points lying on or above the graph.

It turns out that some of the most important multifunctions in optimization theory, such as multifunctions expressing feasibility or optimality, are actually differentiable in such a generalized sense. Moreover, a large class of functions used in optimization is epi-differentiable and the epi-derivatives can be used to obtain the optimality conditions of nonsmooth problems as in the classical case.

In this thesis, an attempt is made to present most of the results in a real

Banach space setting and the functions are assumed to be extended-real-valued unless otherwise specified.

1.1. Outline of the Thesis.

In **Chapter 2**, we review two notions of set convergence; namely *Painlevé-Kuratowski* (PK) and *Attouch-Wets* (AW) convergences. We use these notions to define the *proto-derivatives* of multifunctions.

A family of subsets $\{ S_t \}_{t>0}$ of a Banach space $\mathcal{X}$ (or more generally any topological space) parameterized by $t > 0$ is said to Painlevé-Kuratowski converge(PK) to a set $S \subset \mathcal{X}$ if

$$s\text{-}\limsup_{t \downarrow 0} S_t = S = s\text{-}\liminf_{t \downarrow 0} S_t.$$

Here $s\text{-}\limsup_{t \downarrow 0} S_t$ is the set of all accumulation points of sequences from the sets S_t and $s\text{-}\liminf_{t \downarrow 0} S_t$ is the set of limit points of such sequences. It follows that the limit set S is closed. In finite dimensions, this convergence of sets $\{ S_t \}_{t>0}$ is the pointwise convergence of the distance functions $\xi \to d(\xi, S_t)$ (Proposition 2.1.5). This convergence was first introduced by Painlevé in 1902 and later popularized by Kuratowski in his famous book TOPOLOGIE.

Of course, one obtains different kinds of PK-convergences with respect to various topologies (e.g. weak, weak-star, etc.). An important convergence was introduced by Mosco **[16]**, this convergence amounts to the sets $\{ S_t \}_{t>0}$ converging to S in PK-sense in both weak and strong topologies. It turns out however that the good properties of Mosco convergence are limited inherently to reflexive Banach spaces (Beer and Borwein **[7]**).

On the other hand, set convergence in the Attouch-Wets (AW) sense **[3]** refers to the uniform convergence of the distance functions $\xi \to d(\xi, S_t)$ on all bounded sets of $\mathcal{X}$ (Proposition 2.2.2). When $\mathcal{X}$ is finite dimensional, the three notions are equivalent. In general AW-convergence implies PK-convergence and Mosco convergence is really appropriate for convex sets (since the closed convex sets are weakly closed) where it lies between the other two types of convergences.

The sets to which we want to apply such convergence in order to define proto-differentiability are the graphs of difference quotient multifunctions :

Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. A multifunction $\Gamma : \mathcal{X} \rightrightarrows \mathcal{Y}$ is (PK) *proto-differentiable* at a point x and for a particular element $v \in \Gamma(x)$ if the difference quotient multifunctions

$$\xi \rightrightarrows \left[\Gamma(x + t\xi) - v\right]/t$$

regarded as a family index by $t > 0$, graph-converge (PK sense) as $t \downarrow 0$. If so, the limit mapping is denoted by $\Gamma'^{(pk)}_{x,v}$ and called the (PK) *proto-derivative* of Γ at x and v. It assigns to each $\xi \in \mathcal{X}$ a subset $\Gamma'^{(pk)}_{x,v}(\xi)$ of $\mathcal{Y}$, which could be empty for some choice of ξ.

This notion of proto-differentiation was introduced by Rockafellar in **[27]**, although the idea of differentiating a set-valued mapping by constructing an appropriate tangent cone to its graph was first developed in detail in the book of Aubin and Ekeland [**4**]. When the mapping happens to be single-valued and (Fréchet) differentiable, its proto-derivative reduces to the usual derivative. In general, the proto-derivative is a positively homogeneous multifunction having a closed graph (Proposition 2.3.9) i.e., a set-valued analog of a continuous linear operator.

It is often said that the best way to grasp a mathematical idea is through a geometric interpretation, so it is appropriate to describe the proto-derivative as a certain tangent cone to the graph of multifunction. It turns out that, the distinguishing feature of proto-differentiability is its requirement that the *contingent* cone and *derivable* cone coincide (Proposition 2.3.7). We discuss these analytic and geometric properties of proto-derivatives in the last section of the chapter.

Chapter 3 introduces a more appropriate notion for the convergence of sequence of extended-real-valued functions: the epi-convergence which corresponds geometrically to set convergence of the epigraphs of the functions. Specifically, one obtains (strong) epi-, Mosco epi-, AW epi- convergences with respect to the

PK, Mosco and AW convergences of sets. However, in finite dimensions all these notions of convergences coincide as all forms of set convergences agree.

Historically, the notion of epi-convergence arose as a property that guarantees the continuity of *Legendre-Fenchel* transform ($f \to f^*$). Epi-convergence was first considered by Wijsman **[34]** in 1966. He proved that for convex functions in a finite dimensional setting epi-convergence makes the Legendre-Fenchel transform continuous. Later, this property was extended to reflexive Bnach spaces by Mosco **[16]** (w.r.t. Mosco epi-convergence) and more recently to non reflexive spaces by Attouch and Wets **[3]** (w.r.t. AW epi-convergence).

For our purposes we study the Mosco epi-convergence (M-convergence) in reflexive Banach spaces. The following characterization of M-convergence is due to Mosco **[16]**, which makes verifying M-convergence easier.

A family of functions $\varphi_t : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is M-converges to $\varphi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ when for all $\xi \in \mathcal{X}$ the following hold,

(a) $(\forall \xi_t \overset{w}{\to} \xi) \qquad \liminf_{t \downarrow 0} \varphi_t(\xi_t) \geq \varphi(\xi)$.

(b) $(\exists t_n \downarrow 0)$ and $(\xi_n \overset{s}{\to} \xi)$ with $\limsup_{n \to \infty} \varphi_{t_n}(\xi_n) \leq \varphi(\xi)$.

Our notation w indicates the weak convergence and s for the strong one.

Now the natural question is to investigate the connections between (Mosco) epi-convergence and pointwise (p) convergence. Simple examples can be constructed even in the finite dimensional setting, to show that M-convergence and p-convergence do not imply each other (Example 3.1.5). However, for a subclass of closed convex functions (equi-lower semicontinuous) these convergences agree, in finite dimensions as shown by Salinetti and Wets **[31]**.

This notion of convergence has recently been successful in overcoming the failures of pointwise convergence in many problems of Calculus of Variations, Optimization, Stochastic Programming, etc. (see, for instance the book **[1]**).

Next, we define the Mosco epi-derivatives by replacing the usual pointwise convergence of difference quotients by Mosco epi-convergence. The strong feature of (Mosco) epi-derivative is that it corresponds to a geometric concept of approximation much like the one used in classical differential analysis. In **Chapter 4** we identify a central class of functions for which such derivatives do always exist.

Chapter 4 presents a general class of functions very useful in optimization theory yet enjoys a sharper form of subdifferential calculus. The class of functions which can be written as a composition of convex function (extended valued) and a smooth mapping has been recognized as a model of greater promise in optimization theory and applications. Recently, Rockafellar introduced amenable functions (in finite dimensional setting) as a well chosen class of convexly composite functions. Here we introduce amenable functions in a Banach space setting:
A function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is *fully amenable* at $\bar{x}$, a point of the *effective* domain of f (i.e. $\bar{x} \in \operatorname{dom} f := \{\, x \mid f(x) < +\infty \,\}$), if in a neighbourhood of $\bar{x}$, we have $f = g \circ F$ where $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ is a piecewise linear-quadratic convex function (Definition 4.1.2) and $F : \mathcal{X} \to \mathcal{Y}$ is a $\mathcal{C}^2$ (Fréchet) mapping. In addition the following constraint qualification is satisfied at $\bar{x}$:

$$\mathbb{R}_+\big(\operatorname{dom} g - F(\bar{x})\big) - DF(\bar{x})\mathcal{X} = \mathcal{Y}, \tag{1.1.1}$$

where $DF(\bar{x})$ is the Fréchet differential of F at $\bar{x}$.

Fully amenability is the refinement of amenability (Definition 4.1.1) that supports second-order as well as first-order subdifferential theory. It turns out that, in finite dimensional case, fully amenable functions are (twice) epi-differentiable and the basic calculus rules apply (see **[21]**, for an exposition). Problems that can be written using fully amenable functions include standard nonlinear programming problems, convex problems, nonsmooth (and nonconvex) problems in which the objective is the maximum of a finite collection of smooth functions, and much more (see Examples in Section 1).

The constraint qualification (1.1.1) is devised to handle the case where $F(x)$ is a boundary point of $\operatorname{dom} g$. In the finite dimensional case, our definition of amenability reduces to the one introduced by Rockafellar as the condition (1.1.1) is equivalent to the basic constraint qualification $N_{\mathrm{dom}\, g}\big(F(\bar{x})\big) \cap \ker\big(DF(x)^*\big) = \{\,0\,\}$. One sees that the amenability is a local property, this is because the constraint qualification (1.1.1) is a local condition, as shown in the Theorem 4.1.9.

The next task is to investigate the subdifferential properties of amenable functions. The situation is quite settled in the finite dimensional case. In fact, Rockafellar established in [**26**] that when f is fully amenable at x it is twice epi-differentiable at x for every $v \in \partial f(x)$ and the epi-derivatives can be given by usable formulas (as in 4.2.20). Here $\partial f(x)$ is the set of *subgradients* (replacement for the gradient, see Definition 4.2.2), which can be taken in any sense since all forms of subgradients agree as far as fully amenable functions are concerned. This is even true in the infinite dimensional case.

In Section 2, we prove several new results (extensions). The first one in our agenda is to show that all forms of subgradients agree in the case of fully amenable functions. This is addressed through Theorem 4.2.7 and Theorem 4.2.5. Indeed, Theorem 4.2.7 shows that fully amenable functions are Clarke regular while Theorem 4.2.5 shows that if a Clarke regular function is Mosco epi-differentiable then all forms of subgradients agree. Hence it follows that all forms of subgradients agree from the known result that fully amenable functions are Mosco epi-differentiable.

Next we turn to the second-order analysis. A relationship of fundamental importance in determining the proto-derivatives of the subgradient mapping $\partial f : \mathcal{X} \rightrightarrows \mathcal{X}^*$ of a convex function f is the following:

$$f \quad \text{is twice Mosco epi-differentiable at } x \text{ for a functional } v \in \partial f(x)$$

$$\Longleftrightarrow$$

$$\partial f \quad \text{is (PK) proto-defferentiable at } x \text{ for } v,$$

and then

$$\big(\partial f\big)'^{(pk)}_{x,v}(\xi) = \partial\big(\tfrac{1}{2} f''^{(m)}_{x,v}\big)(\xi) \text{ for all } \xi. \tag{1.1.2}$$

This result was obtained in finite dimensions in Rockafellar [**30**] and generalized to reflexive Banach spaces by Do [**12**]. However, Poliquin [**18**] proved that for any fully amenable function f on $\mathbb{R}^n$, the proto-derivatives of ∂f exists and can be determined through (1.1.2) from the formulas known for second-order epi-derivatives of f. Since the subgradients provide first-order information, this result can be viewed as giving second-order information on fully amenable functions.

Our aim here is to extend second-order results of fully amenable functions and establish the relation (1.1.2) in a reflexive Banach space setting. We must begin with the underline properties of functions g appearing in the definition of fully amenable functions. We prove in Theorem 4.2.12 that any piecewise linear-quadratic function g is twice Mosco epi-differentiable, and that first- and second-order Mosco epi-derivatives are given by simple formulas (as a directional derivative). This was first established by Rockafellar in finite dimensions ([**26**], Theorem 3.1). Extending these results to reflexive spaces one has to replace the epi-convergence by the slightly more complicated notion of Mosco epi-convergence.

Mosco epi-differentiability of f is addressed next. However, one cannot expect the relation (1.1.2) to hold in its full strength (as it does in finite dimensional case). This is because one has to impose conditions over F ($f = g \circ F$) to guarantee the Mosco epi-convergence of second-order difference quotients of f. In particular, from a recent result of Cominetti [**11**] it follows that if $F(x) \in \operatorname{int}(\operatorname{dom} g)$ and the mapping $\xi \to \langle D^2F(x)\xi, \xi\rangle$ from $\mathcal{X}$ to $\mathcal{Y}$ is weakly continuous then f is twice Mosco epi-differentiable at x.

Assuming some additional smoothness assumptions on the Banach space $\mathcal{X}$, we prove in Theorem 4.2.13 that the mapping ∂f is (PK) proto-differentiable at x and the relation (1.1.2) is satisfied as well. An explicit formula for the proto-derivative is also given in Theorem 4.2.13.

In **Chapter 5**, we consider a subclass of fully amenable functions which can be written as the pointwise maximum of finitely many $\mathcal{C}^2$ functions. This subclass is sufficient to cover standard problems of mathematical programming with finitely many constraints (Example 4.1.8). Subdifferential properties of these max-functions have been studied by several authors Auslender and Cominetti [**6**], Penot **[17]**, Poliquin and Rockafellar **[22]** to name a few.

Here we extend the work of Poliquin and Rockafellar **[22]** to Banach spaces. In fact, Theorem 5.1.1 establishes the first- and second-order subdifferential results of max-functions through the extended results of fully amenable functions in **Chapter 4**. Specifically, proto-derivatives of subgradient mappings of max-functions are determined through (1.1.2). In later half of the chapter we compare our proto-derivative formula of the subgradient mapping of max-function with another approach **[6]** that used the direct definition of proto-derivative. Even in the finite dimensional case, it is hard to see whether these formulas agree. However, in the special case of maximum of finitely many linear functions we show that these formulas agree (Theorem 5.2.2). We conclude our work deriving the subdifferential properties of the absolute value function from Theorem 5.2.2.

CHAPTER 2

SET CONVERGENCE AND PROTO-DIFFERENTIABILITY

The convergence of sets is the key ingredient in the definition of generalized differentiability of set-valued mappings (multifunctions). Various forms of convergence have been studied for this matter and among them Painlevé-Kuratowski (PK) and Attouch-Wets (AW) convergence are extensively used in this area. For PK and AW convergences see **[5]**, **[27]** and **[2]**, **[3]** respectively.

2.1. PK-Convergence

Definition 2.1.1. *A family of subsets $\{ S_t \}_{t>0}$ of a Banach space $\mathcal{X}$ parameterized by $t > 0$ is said to PK-converge to a set $S \subset \mathcal{X}$ in the strong (norm) topology as $t \downarrow 0$, written $S_t \xrightarrow{pk} S$, if*

$$s\text{-}\limsup_{t\downarrow 0} S_t = S = s\text{-}\liminf_{t\downarrow 0} S_t,$$

where

$$s\text{-}\limsup_{t\downarrow 0} S_t := \left\{ \xi \in \mathcal{X} \,\middle|\, \exists\ t_n \downarrow 0,\ and\ \xi_n \xrightarrow{s} \xi \quad with \quad \xi_n \in S_{t_n} \right\}$$

is the outer limit set of $\{ S_t \}_{t>0}$, and

$$s\text{-}\liminf_{t\downarrow 0} S_t := \left\{ \xi \in \mathcal{X} \,\middle|\, \forall\ t_n \downarrow 0,\ \exists \xi_n \xrightarrow{s} \xi \ \ with\ \xi_n \in S_{t_n}\ \ for\ n\ sufficiently\ large\ \right\}$$

is the inner limit set of $\{ S_t \}_{t>0}$.

Analogously, we can also define, with the obvious modifications, $w\text{-}\limsup$, $w\text{-}\liminf$, $w^*\text{-}\limsup$ etc..., and obtain different kinds of PK-convergences. In Chapter 3, we introduced one of such called *Mosco* convergence which guarantees the convergence in both strong and weak topologies. In finite dimensions, all these notions of convergences coincide.

Proposition 2.1.2. *Assume $S_t \overset{pk}{\to} S$. We have the following properties for set limits:*

(a) *The set S is unique and closed.*

(b) $s\text{-}\liminf_{t\downarrow 0} S_t \subset s\text{-}\limsup_{t\downarrow 0} S_t$ *and these sets are closed.*

(c) *The inner limit $s\text{-}\liminf_{t\downarrow 0} S_t$ consists of all limit points of sequences ξ_n selected with $\xi_n \in S_{t_n}$ while the outer limit $s\text{-}\limsup_{t\downarrow 0} S_t$ consists of all cluster points of such sequences.*

Proof. Follows easily from the definitions of set limit. □

Example 2.1.3. *Let $S_t := \{(x,y) \mid x^2 + y^2 - (2/t)y \le 0 \,; t > 0\} \subset \mathbb{R}^2$. Then S_t forms an increasing family of sets as $t\downarrow 0$ that PK-converge to $S := \{(x,y) \mid y \ge 0\}$.*

Example 2.1.4. *Consider the family of sets defined for $t > 0$ by*

$$S_t := \begin{cases} D_1 & \text{if } t \text{ rational,} \\ D_2 & \text{if } t \text{ irrational.} \end{cases}$$

where D_1 and D_2 are closed sets in $\mathcal{X}$.
Then $s\text{-}\limsup_{t\downarrow 0} S_t = D_1 \cup D_2$ and $s\text{-}\liminf_{t\downarrow 0} S_t = D_1 \cap D_2$. So $\{ S_t \}_{t>0}$ fails to converge as $t\downarrow 0$ unless $D_1 = D_2$.

According to the definitions, all types of limit sets are closed and consideration of sets that are not closed are really unnecessary. For instance, the constant sequence $S_t = \mathbb{Q}^n \subset \mathbb{R}^n$ (where $\mathbb{Q}$ is the set of rational numbers) converges to $\mathbb{R}^n$, not $\mathbb{Q}^n$, since $\lim_{t\downarrow 0} S_t = \operatorname{cl} \mathbb{Q}^n = \mathbb{R}^n$.

Let $\{ S_t \}_{t>0}$ be a family of sets and S be a closed set in a finite dimensional space $\mathcal{X}$. Then we have the following distance function characterization for the set limit.(cf. Salinetti & Wets **[32]**)

Proposition 2.1.5. *Let $\mathcal{X}$ be a finite dimensional space.*

$$S_t \overset{pk}{\to} S \iff S \text{ closed and } d(\xi, S_t) \to d(\xi, S) \text{ for all } \xi \text{ in } \mathcal{X}.$$

where $d(\xi, D)$ is the distance (generated by the norm) from ξ to the set D.

Proof. First suppose that for all $\xi \in \mathcal{X}$, $\lim_{t\downarrow 0} d(\xi, S_t) = d(\xi, S)$. If $\xi \in S$ then $d(\xi, S) = 0$ and hence by the hypothesis we have that $\lim_{t\downarrow 0} d(\xi, S_t) = 0$. This means that for all $t_n \downarrow 0$, there exists ξ_n in S_{t_n} with $\xi_n \to \xi$. Hence $\xi \in \liminf_{t\downarrow 0} S_t$. This is true for every ξ in S and consequently

$$S \subset \liminf_{t\downarrow 0} S_t. \tag{2.1.1}$$

To prove $S_t \overset{pk}{\to} S$ it remains to show that $\limsup_{t\downarrow 0} S_t \subset S$. Let $\xi \in \limsup_{t\downarrow 0} S_t$. Then there must exist $t_n \downarrow 0$, and ξ_n in S_{t_n} with $\xi_n \to \xi$ and hence by the assumption we have $d(\xi, S) = \lim_{t_n \downarrow 0} d(\xi, S_{t_n}) = 0$. Which implies $\xi \in S$, since S closed. This is true for every $\xi \in \limsup_{t\downarrow 0} S_t$ and thus

$$\limsup_{t\downarrow 0} S_t \subset S. \tag{2.1.2}$$

Inclusions (2.1.1) and (2.1.2) yield $S_t \overset{pk}{\to} S$.(Note that this part of the proof is valid in any space)

Now we suppose that we have $S_t \overset{pk}{\to} S$. Recall that by definition S is closed. For any $\xi \in \mathcal{X}$, there exists $t_n \downarrow 0$, $z_n \in \mathrm{cl}(S_{t_n})$ and $z \in S$ such that $d(\xi, S_{t_n}) = \|\xi - z_n\|$ and $d(\xi, S) = \|\xi - z\|$. By hypothesis $z \in S$ implies that $z \in \liminf_{t\downarrow 0} S_t$ and for all $t_n \downarrow 0$, there exists $y_n \in S_{t_n}$ with $y_n \to z$. Clearly for all $\xi \in \mathcal{X}$, we have that $\lim_{n\to\infty} d(\xi, y_n) = d(\xi, z) = d(\xi, S)$. On the other hand, we have also that for all $n \in \mathbb{N}$, $d(\xi, y_n) \geq d(\xi, z_n)$. Therefore, $\limsup_{n\to\infty} d(\xi, S_{t_n}) \leq d(\xi, S)$ and also note that the $\{z_n\}$ is bounded. So we have

$$\limsup_{t\downarrow 0} d(\xi, S_t) \leq d(\xi, S). \tag{2.1.3}$$

Since $\{z_n\}$ is bounded it has at least one cluster point, say y. Then $y = \lim_{k\to\infty} z_{n_k}$, where $z_{n_k} \in S_{t_{n_k}}$. Thus, $y \in \limsup_{t\downarrow 0} S_t = S$. We have that for all k, $d(\xi, z_{n_k}) \geq d(\xi, y) - d(y, z_{n_k})$ and consequently

$$\liminf_{k\to\infty} d(\xi, z_{n_k}) \geq d(\xi, y) \geq d(\xi, S)$$

This implies

$$\liminf_{t \downarrow 0} d(\xi, S_t) \geq d(\xi, S). \qquad (2.1.4)$$

Relations (2.1.3) and (2.1.4) yield the result. □

Remark 1. The pointwise convergence of distance functions implies the PK-convergence of $\{ S_t \}_{t>0}$ in any infinite dimensional Banach spaces.

Remark 2. The proof of only if part is based on the pointwise convergence of distance functions. Actually , these functions converge uniformly on all bounded sets in the finite dimensional space $\mathcal{X}$, by the fact that the distance functions $\xi \to d(\xi, S_t)$ are uniformly Lipschitz (with modulus 1).

The following characterization enable us to approximate the sets S_t using the set limit S. For instance, see Proposition 2.3.2.

Proposition 2.1.6. *Let $\mathcal{X}$ be a finite dimensional space. $S_t \overset{pk}{\to} S \iff S$ is a closed set and such that for arbitrary large $\rho > 0$ and arbitrary small $\varepsilon > 0$, there exists $\tau > 0$ for which*

$$S_t \cap \rho B \subset S + \varepsilon B \quad \textit{and} \quad S \cap \rho B \subset S_t + \varepsilon B \quad \textit{when} \quad t \in (0, \tau).$$

Proof. Suppose $S_t \overset{pk}{\to} S$. Then S is a closed set(by definition).

If the first inclusion does not hold, there must exists $\rho > 0, \varepsilon > 0$, and sequence $t_n \downarrow 0$ such that $\xi_n \in S_{t_n} \cap \rho B \not\subset S + \varepsilon B$. The sequence $\{\xi_n\}$ is bounded and therefore has at least one cluster point ξ. So there exists a subsequence $\{\xi_{n_k}\}$ of $\{\xi_n\}$ such that $\xi_{n_k} \in S_{t_{n_k}}$ with $\xi_{n_k} \to \xi$. By definition then $\xi \in \limsup_{t \downarrow 0} S_t$, but it cannot be in S, because $d(\xi_n, S) \geq \varepsilon$ for all n. This contradicts $\limsup_{t \downarrow 0} S_t = S$, thus $S_t \cap \rho B \subset S + \varepsilon B$ for all t in $(0, \tau)$.

If the second inclusion does not hold, there must exist $\rho > 0, \varepsilon > 0$ and a sequence $t_n \downarrow 0$ such that $\xi_n \in S \cap \rho B \not\subset S_{t_n} + \varepsilon B$. By compactness of the set $(S \cap \rho B)$, there exists a subsequence $\{\xi_{n_k}\}$ of $\{\xi_n\}$ such that $\xi_{n_k} \to \xi$ in S. Thus, by the hypothesis $\xi \in \limsup_{t \downarrow 0} S_t$ but $d(\xi_n, S_{t_n}) \geq \varepsilon$. So $\xi \notin \limsup_{t \downarrow 0} S_t$, a contradiction. Thus we have $S \cap \rho B \subset S_t + \varepsilon B$ for all $t \in (0, \tau)$.

Now suppose we have that $S_t \cap \rho B \subset S + \varepsilon B$ and $S \cap \rho B \subset S_t + \varepsilon B$ for all t in $(0, \tau)$. To verify that $S_t \overset{pk}{\to} S$ it is enough to show $\limsup_{t \downarrow 0} S_t \subset S$ and $S \subset \liminf_{t \downarrow 0} S_t$. If $\limsup_{t \downarrow 0} S_t \not\subset S$, there must exist $t_n \downarrow 0$. $\xi_n \in S_{t_n}$ with $\xi_n \to \xi \notin S$. Pick ρ big enough to have $\|\xi_n\| \leq \rho$ for all n. Thus, $\xi_n \in (S_{t_n} \cap \rho B)$ and by the first hypothesis $\xi_n \in (S + \varepsilon B)$ for all $\varepsilon > 0$. Since S is closed and $\varepsilon > 0$ was arbitrary, we conclude that $\xi \in S$. This is a contradiction, thus $\limsup_{t \downarrow 0} S_t \subset S$.

If $S \not\subset \liminf_{t \downarrow 0} S_t$, there is a ξ in S but not in $\liminf_{t \downarrow 0} S_t$. Take $\rho \geq \|\xi\|$ and by the hypothesis for all $\varepsilon > 0$, there exists $\tau > 0$ such that $\xi \in S \cap \rho B \subset S_t + \varepsilon B$ when $t \in (0, \tau)$. This implies $\xi \in S_t + \varepsilon B$ for all $\varepsilon > 0$ and $t \in (0, \tau)$. Hence, for all $t_n \downarrow 0$, there exists $\xi_n \in S_{t_n}$ such that $d(\xi, S_{t_n}) \leq \varepsilon$ for all $\varepsilon > 0$. Thus, $\xi \in \liminf_{t \downarrow 0} S_t$, a contradiction. We have that $S \subset \liminf_{t \downarrow 0} S_t$ and completes the proof. ∎

Note. We always have $\Longleftarrow$ in any space.

For a sequence of sets in a separable metric space we have the following version of Bolzano-Weierstrass property to the set-valued framework (cf. [**5**], Theorem 1.1.7).

Theorem 2.1.7. *(Zarankiewicz) Every sequence of subsets S_n of a separable metric space $\mathcal{X}$ contains a subsequence which has a (possibly empty) limit.*

Proof. Since $\mathcal{X}$ is separable, there exists a countable family of open subsets $\mathcal{U}_m$ satisfying the following property:

$$\forall \text{ open subset } \mathcal{U},\ \forall\, x \in \mathcal{U}, \exists\, \mathcal{U}_m \text{ such that } x \in \mathcal{U}_m \subset \mathcal{U}.$$

Let us consider a sequence of subsets S_n. We shall construct a sequence of subsequences $\{S_n^{(m)}\}_{n>0}$ by induction.

For $m = 0$, we set $S_n^{(0)} := S_n$. Assume that the $m-1$ first subsequences $\{S_n^{(p)}\}_{n>0}$, $0 \leq p \leq m-1$ have been constructed. Consider the m^{th} open subset $\mathcal{U}_m$. Then either for every subsequence n_j, $\mathcal{U}_m \cap (\limsup_{j\to\infty} S_{n_j}^{(m-1)}) \neq \emptyset$ in which case we set $S_j^{(m)} := S_j^{(m-1)}$, or there exists a subsequence n_j such that $\mathcal{U}_m \cap$

$(\limsup_{j\to\infty} S_{n_j}^{(m-1)}) = \emptyset$ in which case we set $S_j^{(m)} := S_{n_j}^{(m-1)}$.(The choice of such a subsequence does not matter.) These sequences $\{S_n^{(m)}\}_{n>0}$ being constructed, we extract the diagonal subsequence $D_n := S_n^{(n)}$. We claim that it has a set limit.

If not, there would exist $x_0 \in \limsup_{n\to\infty} D_n$ and $x_0 \notin \liminf_{n\to\infty} D_n$. The later condition means that there exists an open neighbourhood $\mathcal{U}$ of x_0 and a subsequence D_{n_j} such that $\mathcal{U} \cap D_{n_j} = \emptyset$ for any j. Let us fix an open subset $\mathcal{U}_m$ such that $x_0 \in \mathcal{U}_m \subset \mathcal{U}$. We thus deduce that $\mathcal{U}_m \cap (\limsup_{j\to\infty} D_{n_j}) = \emptyset$.

Since for $n_j \geq m$, $D_{n_j} := S_{n_j}^{n_j} = S_{p_j}^{(m-1)}$ for some p_j, we observe that D_{n_j} is a subsequence of the sequence$\{S_n^{(m-1)}\}_{n>0}$, the upper limit of which is disjoint from $\mathcal{U}_m$. By the very construction of $\{S_n^{(m)}\}_{n>0}$, we infer that $S_j^{(m)} = S_{p_j}^{(m-1)}$ and consequently, that

$$\mathcal{U}_m \cap (\limsup_{j\to\infty} S_j^{(m)}) = \mathcal{U}_m \cap (\limsup_{j\to\infty} S_{p_j}^{(m-1)}) = \emptyset.$$

Since $D_n := S_n^{(n)} = S_{p_n}^{(m)}$ for some p_n, we deduce that the sequence $\{D_n\}_{n\geq m}$ is a subsequence of the the sequence $\{S_j^{(m)}\}_{j>0}$. Thus

$$x_0 \in \limsup_{n\to\infty} D_n \subset \limsup_{j\to\infty} S_j^{(m)} \subset \mathcal{X} \setminus \mathcal{U}_m$$

which contradicts the fact that x_0 belongs to $\mathcal{U}_m$. □

2.2. Attouch-Wets Convergence

Let $e(C, D)$ denotes the one-sided Hausdorff distance from the set C to the set D in $\mathcal{X}$; it is defined as follows:

$$e(C, D) := \begin{cases} 0 & \text{if } C = \emptyset, \\ \infty & \text{if } C \neq D, D = \emptyset. \\ \sup_{x\in C} d(x, D) & \text{otherwise.} \end{cases}$$

For any $\rho \geq 0$, the ρ-Hausdorff distance between C and D is given by

$$\text{haus}_\rho(C, D) := \max\{e(C_\rho, D), e(D_\rho, C)\}$$

where for any set $H \subset \mathcal{X}$, $H_\rho := H \cap \rho B$, B denotes the closed unit ball in $\mathcal{X}$. The ρ-Hausdorff distance also given as :

$$\text{haus}_\rho(C, D) = \inf \left\{ \varepsilon > 0 \,\middle|\, C_\rho \subset (D + \varepsilon B) \text{ and } D_\rho \subset (C + \varepsilon B) \right\}.$$

Definition 2.2.1. *A sequence of sets* $\{ S_t \}_{t>0}$ *is said to Attouch-Wets(AW) converge to* S, *denoted* $S_t \overset{aw}{\to} S$, *if for all* ρ *big enough* $\lim_{t \downarrow 0} \text{haus}_\rho(S_t, S) = 0$.

We already seen (Proposition 2.1.5 and Remark 2) that in finite dimensional spaces PK-convergence of sets S_t refers to the (uniform) convergence of distance functions $x \to d(x, S_t)$ (on all bounded sets). AW-convergence is much stronger than PK-convergence in the sense that it is equivalent to the uniform convergence of the distance functions on all bounded sets in any space.

Proposition 2.2.2. *Let* $\mathcal{X}$ *be any normed space.*

$S_t \overset{aw}{\to} S \iff$ *for all* x *in any bounded set* E *of* $\mathcal{X}$ *and for all* $\varepsilon > 0$ *there exists* $\tau > 0$ *such that* $|d(x, S_t) - d(x, S)| < \varepsilon$, *whenever* t *in* $(0, \tau)$.

Lemma 2.2.3. *For* C, $D \subset \mathcal{X}$ *and* $\rho \geq 0$, *let*

$$\delta_\rho(C, D) := \sup_{\|y\| \leq \rho} |d(y, C) - d(y, D)|$$

where $\delta_\rho(C, D) := \infty$ *if at least one of the sets is empty. Then for any* $\rho \geq 0$,

$$\delta_\rho(C, D) \geq \text{haus}_\rho(C, D) \tag{2.2.1}$$

and for all $\rho \geq d(0, C)$,

$$\delta_\rho(C, D) \leq \text{haus}_{3\rho}(C, D) \tag{2.2.2}$$

Proof of Lemma 2.2.3. *Since* $\rho B \supset C_\rho$ *for any* $\rho \geq 0$,

$$\delta_\rho(C, D) \geq \sup_{y \in C_\rho} d(y, D) = e(C_\rho, D) \text{ and hence } \delta_\rho(C, D) \geq \text{haus}_\rho(C, D).$$

Now fix $\rho \geq d(0, C)$. Since the distance functions $y \to d(y, C)$ are Lipschitz in y we have $d(y, C) \leq d(0, C) + \|y\|$ for $y \in \mathcal{X}$. For all $y \in \mathcal{X}$ such that $\|y\| \leq \rho$ we have that $d(y, C) \leq 2\rho$ and thus $d(y, C) = d(y, C_{3\rho})$. It follows that

$$\sup_{\|y\| \leq \rho} \{d(y, D) - d(y, C)\} \leq \sup\{d(y, D) - d(y, C_{3\rho})\} \leq e(C_{3\rho}, D).$$

With the symmetric inequality, obtained when interchanging the roles C and D, this becomes $\delta_\rho(C, D) \leq \text{haus}_{3\rho}(C, D)$. □

Proof of Proposition 2.2.2. Let E be any bounded set in $\mathcal{X}$. We first suppose $S_t \overset{aw}{\to} S$, i.e. for any given $\varepsilon > 0$, there exists $\tau > 0$ such that $\text{haus}_\rho(S_t, S) < \varepsilon$, for all $t \in (0, \tau)$ and $\rho \geq \bar{\rho}$ for some $\bar{\rho}$. Now pick ρ' such that

$$\rho' := \sup \left\{ \bar{\rho}, \ \|x\|, \ d(0, S) \,\middle|\, x \in E \right\}.$$

Fix $\rho \geq \rho', \varepsilon > 0$, then by inequality (2.2.2), and the AW-convergence of sets $\{ S_t \}_{t>0}$, there exists $\tau' > 0$ such that

$$\delta_\rho(S, S_t) \leq \text{haus}_{3\rho}(S, S_t) = \text{haus}_{3\rho}(S_t, S) < \varepsilon \ \text{ for all } t \in (0, \tau')$$

This implies $\sup\limits_{x \in E} |d(x, S_t) - d(x, S)| < \varepsilon$ for all $t \in (0, \tau')$ and have the desired result.

Conversely, suppose for any bounded set E and given $\varepsilon > 0$, there exists $\tau > 0$ such that $|d(x, S_t) - d(x, S)| < \varepsilon$ for all $x \in E$, and $t \in (0, \tau)$. Take any $\rho > 0$ and put $E = \rho B$. Then by the hypothesis we have $\sup\limits_{y \in \rho B} |d(x, S_t) - d(x, S)| = \delta_\rho(S_t, S) < \varepsilon$, for all $t \in (0, \tau)$. Combining this with the inequality (2.2.1) of the lemma, we have $\text{haus}_\rho(S_t, S) \leq \delta_\rho(S_t, S) < \varepsilon$ for all $t \in (0, \tau)$, which completes the proof. □

In finite dimensional case, the notions of PK-convergence and the AW-convergence coincide. To be more precise we have:

Proposition 2.2.4. *Let $\{S_t, \ S; \ t > 0\}$ be a family of sets in a finite dimensional space $\mathcal{X}$.*

(a) If $S_t \overset{pk}{\to} S$ then $\lim\limits_{t \downarrow 0} \text{haus}_\rho(S_t, S) = 0$ for all $\rho > 0$

(b) For all ρ big enough $\lim\limits_{t \downarrow 0} \text{haus}_\rho(S_t, S) = 0$ and S be closed then we have $S_t \overset{pk}{\to} S$.

Proof.

(a) Suppose $S_t \overset{pk}{\to} S$. It suffices to show that, for all $\rho > 0$,

$$\lim_{t \downarrow 0} e((S_t)_\rho, \limsup_{t \downarrow 0} S_t) \;=\; 0, \tag{2.2.3}$$

and

$$\lim_{t \downarrow 0} e((\liminf_{t \downarrow 0} S_t)_\rho, S_t) \;=\; 0. \tag{2.2.4}$$

Let $LS = \limsup_{t \downarrow 0} S_t$, and $LI = \liminf_{t \downarrow 0} S_t$. There is nothing to prove if $LS = \emptyset$, since then, for any $\rho > 0$, there always exists a sequence $t_n \downarrow 0$ with $(S_{t_n})\rho = \emptyset$. Let us thus assume that $LS \neq \emptyset$. If (2.2.3) does not hold, there exist $\varepsilon > 0$, for all $t_n \downarrow 0$ with $e((S_{t_n})_\rho, LS) > \varepsilon$, or equivalently for all $t_n \downarrow 0$, there exists $\xi_n \in (S_{t_n})_\rho$ such that $d(\xi_n, LS) > \varepsilon$. The sequence $\{\xi_n\}$ is bounded by ρ and hence admits at least one cluster point, say $\xi \in \rho B$, which also belongs to LS. For this ξ, we have that

$$\lim_{\xi_n \to \xi} d(\xi_n, LS) = d(\xi, LS) \geq \varepsilon > 0,$$

which contradicts the fact that $\xi \in (LS)_\rho$.

Again if $(LI)_\rho = \emptyset$, there is nothing to prove because $e((LI)_\rho, S_t) = 0$ for whatever S_t. Otherwise, simply observe that $(LI)_\rho \subset LI$, that $e(C_\rho, D) \leq e(C, D)$, and $\lim_{t \downarrow 0} e((LI)_\rho, S_t) = 0$ as follows from the definition of the *liminf* of a family of sets.

(b) Suppose that $\lim_{t \downarrow 0} \text{haus}_\rho(S_t, S) = 0$ for all $\rho > \bar{\rho}$ for some $\bar{\rho} > 0$. This implies that

$$\lim_{t \downarrow 0} e((S_t)_\rho, S) = 0 \text{ and } \lim_{t \downarrow 0} e(S_\rho, S_t) = 0, \text{ i.e.,}$$

$$\lim_{t \downarrow 0} \sup_{x \in S_t \cap \rho B} d(x, S) = 0 = \lim_{t \downarrow 0} \sup_{x \in S \cap \rho B} d(x, S_t)$$

Therefore, any given $\varepsilon > 0$ one can find $\tau > 0$ satisfying

$$\sup_{x \in S_t \cap \rho B} d(x, S) < \varepsilon \text{ and } \sup_{x \in S \cap \rho B} d(x, S_t) < \varepsilon \text{ for all } t \in (0, \tau).$$

So we have

$$S_t \cap \rho B \subset S + \varepsilon B \text{ and } S \cap \rho B \subset S_t + \varepsilon B \text{ for all } \varepsilon > 0,\ \rho > \bar{\rho} \text{ and } t \in (0, \tau).$$

Since S is closed and above relations with Proposition 2.1.6 yield $S_t \stackrel{pk}{\to} S$. □

Note. Part(b) is valid in any space.

2.3. Proto-differentiability of a Set-valued mapping

A wide class of multifunctions important in optimization enjoys a differential property that we call *proto-differentiability*. We define the proto-differentiability of a set-valued mapping in terms of graphical convergence (i.e. set convergence of graphs) of associated difference quotient multifunctions. This notion was first introduced by Rockafellar in [**27**].

Definition 2.3.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces, $\Gamma : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a set-valued mapping, $x \in$ dom Γ (i.e. $\Gamma(x) \neq \emptyset$) and v a vector in $\Gamma(x)$. The first order difference quotients $\Gamma_t : \mathcal{X} \rightrightarrows \mathcal{Y}$ are defined by*

$$\Gamma_t(\xi) := \big[\Gamma(x + t\xi) - v\big]/t; \ t > 0.$$

We say that Γ is (PK) proto-differentiable at x relative to v with proto-derivative $\Gamma'^{(pk)}_{x,v}$ if gphΓ_t (PK)converge to gph$\Gamma'^{(pk)}_{x,v}$ (where gph stands for graph) as t tends to 0.

As the derivative of a function is used to approximate the function the same can be said of the proto-derivative in a finite dimensional space.

Proposition 2.3.2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be finite dimensional spaces and $\Gamma : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a set valued mapping with $x \in$ domΓ, $v \in \Gamma(x)$. In order that Γ is proto-differentiable at x relative to v, it is necessary and sufficient that there exists a closed graph multifunction $D : \mathcal{X} \rightrightarrows \mathcal{Y}$(i.e. $\{(\xi, D(\xi)); \ \xi \in \mathcal{X}\}$ is closed in $\mathcal{X} \times \mathcal{Y}$), which will be $\Gamma'^{(pk)}_{x,v}$, for which the following holds. For any $\rho > 0$(arbitrary large) and any $\varepsilon > 0$(arbitrary small), there exits $\tau > 0$ such that for all $t \in (0, \tau)$*

$$\text{gph}\Gamma_t \cap \rho(B_{\mathcal{X}} \times B_{\mathcal{Y}}) \subset \text{gph}D + \varepsilon(B_{\mathcal{X}} \times B_{\mathcal{Y}})$$

$$\text{gph}D \cap \rho(B_{\mathcal{X}} \times B_{\mathcal{Y}}) \subset \text{gph}\Gamma_t + \varepsilon(B_{\mathcal{X}} \times B_{\mathcal{Y}})$$

where $B_{\mathcal{X}}$ and $B_{\mathcal{Y}}$ are closed unit balls in $\mathcal{X}$ and $\mathcal{Y}$ respectively.

Proof. Set $S_t :=\text{gph}\Gamma_t$ and $S :=\text{gph}D$ and use Proposition 2.1.6. ◻

Note. We always have $\Longleftarrow$ in any space.

Even if a given multifunction happens to be a single-valued mapping its proto-derivative might not be single-valued as we seen by the following example.

Example 2.3.3. *Consider $\Gamma(x) = \sqrt{|x|} : \mathbb{R} \to \mathbb{R}$. Let Γ_t be the first order difference quotients of Γ at 0 relative to $0 = \Gamma(0)$. Then $\Gamma_t(\xi) = \sqrt{|\xi|/t}$ converges graphically to*

$$\Gamma'_{0,0}(\xi) := \begin{cases} [0,\infty) & \text{if } \xi = 0, \\ \emptyset & \text{if } \xi \neq 0, \end{cases}$$

and is multivalued at $\xi = 0$. Here $\Gamma'_{0,0}$ is the proto-derivative of Γ at 0 (relative to 0).

From a geometric point of view, proto-differentiation of a multifunction corresponds to looking at certain tangent cones to the graph of the multifunction.

Definition 2.3.4. *Let $C \subset \mathcal{X}$ be a subset of $\mathcal{X}$ and $x \in C$.*

(a) *The contingent cone to C at x, denoted by $K_C(x)$, is defined by*

$$K_C(x) := \left\{ \xi \,\middle|\, \exists t_n \downarrow 0, \text{ and } \exists \xi_n \xrightarrow{s} \xi \text{ with } x + t_n\xi_n \in C \text{ for all } n \right\}.$$

(b) *The intermediate or adjacent(derivable) cone to C at x, denoted by $A_C(x)$, is defined by*

$$A_C(x) := \left\{ \xi \,\middle|\, \forall t_n \downarrow 0, \exists \xi_n \xrightarrow{s} \xi \text{ with } x + t_n\xi_n \in C \text{ for } n \text{ sufficientely large } \right\}.$$

(c) *The Clarke tangent cone to C at x, denoted by $T_C(x)$, is defined by*

$$T_C(x) := \left\{ \xi \,\middle|\, \forall t_n \downarrow 0, \forall x_n \xrightarrow{s} x \text{ with } x_n \in C, \exists \xi_n \xrightarrow{s} \xi \right.$$
$$\left. \text{such that } x_n + t_n\xi_n \in C \text{ for } n \text{ sufficientely large } \right\}.$$

We see at once that

$$K_C(x) = s\text{-}\limsup_{t\downarrow 0}\,[C - x]/t, \quad A_C(x) = s\text{-}\liminf_{t\downarrow 0}\,[C - x]/t, \text{ and}$$

$$T_C(x) = s\text{-}\liminf_{\substack{C \ni x' \xrightarrow{s} x \\ t \downarrow 0}} \left[C - x'\right]/t$$

so that they are closed cones and we have the obvious inclusions

$$T_C(x) \subset A_C(x) \subset K_C(x).$$

Definition 2.3.5. *The set C is said to be derivable at x provided $A_C(x) = K_C(x)$. In this case $s\text{-}\lim_{t\downarrow 0}[C - x]/t$ exists and is known as "the approximation cone" to C at x. In addition, if $T_C(x) = K_C(x)$ we say that C is Clarke regular at x.*
In particular, if $x \in \operatorname{int} C$ we have the Clarke regularity since $T_C(x) = K_C(x) = \mathcal{X}$. A different link can be forged through the notion of epigraphs.

Definition 2.3.6. *The function f is Clarke (subdifferentially) regular at x if the set $epif$ is Clarke regular at $(x, f(x))$.*
This property is of strong interest in nonsmooth analysis because of its simplifying effect on various formulae for "generalized gradients"; see Clarke [8].

Setting $C :=$gphΓ in definition 2.3.5 since gph$\Gamma_t = \left[\text{gph}\Gamma - (x, v)\right]/t$ we have the following geometric characterization for proto-differentiability.

Proposition 2.3.7. *The multifunction $\Gamma : \mathcal{X} \rightrightarrows \mathcal{Y}$ is proto-differentiable at a point $v \in \Gamma(x)$ if and only if the set gphΓ is derivable at (x, v). The graph of the proto-derivative multifunction $\Gamma'^{(pk)}_{x,v}$ then equals the approximating cone to gphΓ at (x, v). In particular, if gphΓ is Clarke regular at (x, v) then Γ is proto-differentiable at x relative to v.*
The following formula enable us to calculate the proto-derivative.

Proposition 2.3.8. *Assume that the proto-derivative $\Gamma'_{x,v}$ exists. Then for every $\xi \in \mathcal{X}$ one has*

$$\Gamma'^{(pk)}_{x,v}(\xi) = s\text{-}\lim\sup_{\substack{\xi' \xrightarrow{s} \xi \\ t \downarrow 0}} \left[\Gamma(x + t\xi') - v\right]/t. \tag{2.3.1}$$

Proof. It can be easily verified that a point w belongs to the right side of (2.3.1) if and only if $(\xi, w) \in s\text{-}\limsup_{t\downarrow 0}[\text{gph}\,\Gamma - (x, v)]/t$. Then the result follows at once from the Proposition 2.3.7. □

The next proposition contains the basic features of proto-derivatives.

Proposition 2.3.9. *Let $\Gamma : \mathcal{X} \rightrightarrows \mathcal{Y}$ be proto-differentiable at x relative to $v \in \Gamma(x)$. Then the proto-derivative $\Gamma'^{(pk)}_{x,v} : \mathcal{X} \rightrightarrows \mathcal{Y}$ has closed graph and satisfies $0 \in \Gamma'^{(pk)}_{x,v}(0)$ and $\Gamma'^{(pk)}_{x,v}(\lambda\xi) = \lambda\Gamma'^{(pk)}_{x,v}(\xi)$ for all $\xi \in \mathcal{X}$ and $\lambda > 0$. Moreover $\Gamma'^{(pk)}_{x,v}(0)$ is a closed cone which includes the contingent cone to $\Gamma(x)$ at v.*

If $\mathcal{Y} = \mathcal{X}^*$ and Γ is a monotone operator (i.e. $\forall(x,p) \in gph\Gamma$, $\forall(y,q) \in gph\Gamma$, $\langle p-q, x-y\rangle \geq 0$), then for every $\xi \in dom\Gamma'^{(pk)}_{x,v}$ and for every $\eta_1, \eta_2 \in \Gamma'^{(pk)}_{x,v}(\xi)$, one has $\langle \eta_1, \xi\rangle = \langle \eta_2, \xi\rangle$.

Proof. The verification of first part follows from Propositions 2.3.7 and 2.3.8 and the proof of second part we refer to C.Do [**12**], Proposition 3.2. □

The following calculus of proto-derivative also follows from the Proposition 2.3.8.

Suppose $\Gamma = \bar{\Gamma} + g$ where $\bar{\Gamma} : \mathcal{X} \rightrightarrows \mathcal{Y}$ is proto-differentiable at x relative to $\bar{v} \in \bar{\Gamma}(x)$ and $g : \mathcal{X} \to \mathcal{Y}$ is a function (single-valued) that is Fréchet differentiable at x. Then Γ is proto-differentiable at x relative to $v = \bar{v} + g(x)$ with

$$\Gamma'^{(pk)}_{x,v}(\xi) = \bar{\Gamma}'^{(pk)}_{x,\bar{v}}(\xi) + Dg(x)\xi,$$

where $Dg(x)$ is the Fréchet differential of g at x.

The notion of proto-differentiability is appropriate for the derivation of some of the most important set-valued maps involved in optimization. Specifically, set-valued mappings expressing feasibility or optimality, are proto-differentiable, as seen next.

Example 2.3.10. *(Proto-differentiability of feasible set)*
Let $G : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ have the form

$$G(x) := \Big\{ u \in \mathbb{R}^n \;\Big|\; f_i(x,u) \leq 0, i = 1, \ldots, s \quad f_i(x,u) = 0, i = s+1, \ldots, m. \Big\} \tag{2.3.2}$$

where $f_i : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$ is C^1 for $i = 1, \ldots, m$. Suppose for a particular x and element $u \in G(x)$ that the constrained system in (2.3.2) satisfies the Mangasarian-Fromovitz constraint qualification. i.e.

$$\begin{cases} \text{the only multipliers } y_i \geq 0, \ i = 1, \ldots, s \text{ satisfying} \\ \sum_{i=1}^{m} y_i \nabla_u f_i(x, u) = 0 \text{ are } y_1 = 0, \ldots, y_m = 0. \end{cases}$$

It is verified that(see [**27**], Example 5.5) then G is proto-differentiable at x relative to u and the proto-derivative is given by

$$G'_{x,u}(\xi) := \Big\{ w \in \mathbb{R}^n \;\Big|\; \nabla_x f_i(x, u)\xi + \nabla_u f_i(x, u)w \leq 0, \text{ for all } i \in I(x, u); \\ \nabla_x f_i(x, u)\xi + \nabla_u f_i(x, u)w = 0, \text{ for } i = s + 1, \ldots, m \Big\},$$

where $I(x, u)$ denotes the indices of the inequality constraints in (2.3.2) that are active at u, i.e. the indices $i = \{1, \ldots, s\}$ such that $f_i(x, u) = 0$.

CHAPTER 3

EPI-CONVERGENCE AND EPI-DIFFERENTIATION

3.1. Epi-Convergence

The traditional concept of convergence of functions—the pointwise convergence—relates poorly to many important operations in optimization like, "max", "min", "argmin", "argmax".

For instance, a sequence of functions $f_n : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ converging pointwise to a function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ (i.e. $f_n(x) \to f(x)\ \forall x$, denote $f_n \xrightarrow{p} f$) does not generally imply that $\inf f_n \xrightarrow{p} \inf f$. Therefore, a different concept of convergence is required. The natural notion turns out to be the one corresponding geometrically to set convergence of the *epigraphs* of the functions.

The terminology and notation we used here is the standard one of *Convex Analysis*. For a function f from a Banach space $\mathcal{X}$ to $\mathbb{R} \cup \{+\infty\}$ we define the following. The *effective domain* of f is denoted by

$$\mathrm{dom} f := \left\{ x \in \mathcal{X} \,\middle|\, f(x) < +\infty \right\}$$

and its *epigraph*

$$\mathrm{epi} f := \left\{ (x, \alpha) \in \mathcal{X} \times \mathbb{R} \,\middle|\, f(x) \le \alpha \right\}.$$

The function f is said to be convex if $\mathrm{epi} f$ is convex in $\mathcal{X} \times \mathbb{R}$ and further it is said to be *closed* if f is *lower semicontinuous* (same as $\mathrm{epi} f$ closed) and *proper* (i.e. $\mathrm{dom}\, f$ is nonempty).

Definition 3.1.1. *Let $\mathcal{X}$ be a finite dimensional space. A family of functions $\{ \varphi_t \}_{t>0}$ from $\mathcal{X}$ to $\mathbb{R} \cup \{+\infty\}$ parameterized by $t > 0$ is said to epi-converge to a function $\varphi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ as t tends to 0, denoted by $\varphi_t \xrightarrow{e} \varphi$, if $\mathrm{epi}\, \varphi_t \xrightarrow{pk} \mathrm{epi}\, \varphi$, i.e.,*

$$\limsup_{t \downarrow 0} \mathrm{epi} \varphi_t = \mathrm{epi} \varphi = \liminf_{t \downarrow 0} \mathrm{epi} \varphi_t.$$

In infinite dimensional case, depending on the topology of the space, we obtain different kinds of e-convergences.

Definition 3.1.2.

(a) A family of functions $\{ \varphi_t \}_{t>0}$ from a Banach space $\mathcal{X}$ to $\mathbb{R} \cup \{+\infty\}$ is said to Mosco-epi(M) converge to a function $\varphi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ as $t \downarrow 0$, denoted by $\varphi_t \overset{m}{\to} \varphi$, if epi φ_t *converge to* epi φ *in both weak and strong topologies, i.e.,*

$$w\text{-}\limsup_{t\downarrow 0} \operatorname{epi} \varphi_t \subset \operatorname{epi} \varphi \subset s\text{-}\liminf_{t\downarrow 0} \operatorname{epi} \varphi_t.$$

Since we always have

$$s\text{-}\liminf_{t\downarrow 0} S_t \subset w\text{-}\liminf_{t\downarrow 0} S_t, \quad s\text{-}\limsup_{t\downarrow 0} S_t \subset w\text{-}\limsup_{t\downarrow 0} S_t$$

for any family of sets $\{ S_t \}_{t>0}$, in other words, $\varphi_t \overset{m}{\to} \varphi$ if all four sets above are equal:

$$s\text{-}\liminf_{t\downarrow 0} \operatorname{epi} \varphi_t = w\text{-}\liminf_{t\downarrow 0} \operatorname{epi} \varphi_t = s\text{-}\limsup_{t\downarrow 0} \operatorname{epi} \varphi_t = w\text{-}\limsup_{t\downarrow 0} \operatorname{epi} \varphi_t.$$

where

$$\begin{aligned}
w\text{-}\limsup_{t\downarrow 0} \operatorname{epi} \varphi_t &:= \Big\{ (\xi, \alpha) \in \mathcal{X} \times \mathbb{R} \,\Big|\, \exists\, t_n \downarrow 0, \quad \exists\, (\xi_n, \alpha_n) \overset{w}{\to} (\xi, \alpha) \\
&\qquad \textit{with } \varphi_{t_n}(\xi_n) \le \alpha_n \Big\}, \\
w\text{-}\liminf_{t\downarrow 0} \operatorname{epi} \varphi_t &:= \Big\{ (\xi, \alpha) \in \mathcal{X} \times \mathbb{R} \,\Big|\, \forall\, t_n \downarrow 0, \quad \exists\, (\xi_n, \alpha_n) \overset{w}{\to} (\xi, \alpha) \\
&\qquad \textit{with } \varphi_{t_n}(\xi_n) \le \alpha_n \textit{ for } n \textit{ sufficiently large } \Big\}, \\
s\text{-}\limsup_{t\downarrow 0} \operatorname{epi} \varphi_t &:= \Big\{ (\xi, \alpha) \in \mathcal{X} \times \mathbb{R} \,\Big|\, \exists\, t_n \downarrow 0, \quad \exists\, (\xi_n, \alpha_n) \overset{s}{\to} (\xi, \alpha) \\
&\qquad \textit{with } \varphi_{t_n}(\xi_n) \le \alpha_n \Big\}, \\
s\text{-}\liminf_{t\downarrow 0} \operatorname{epi} \varphi_t &:= \Big\{ (\xi, \alpha) \in \mathcal{X} \times \mathbb{R} \,\Big|\, \forall\, t_n \downarrow 0, \quad \exists\, (\xi_n, \alpha_n) \overset{s}{\to} (\xi, \alpha) \\
&\qquad \textit{with } \varphi_{t_n}(\xi_n) \le \alpha_n \textit{ for } n \textit{ sufficiently large } \Big\}.
\end{aligned}$$

(b) We shall say $\{ \varphi_t \}_{t>0}$ Attouch-Wets(AW) converge to φ as $t \downarrow 0$, denoted by $\varphi_t \overset{aw}{\to} \varphi$, if epi $\varphi_t \overset{aw}{\to}$ epi φ, *i.e.,*

$$\lim_{t\downarrow 0} \operatorname{haus}_\rho(\operatorname{epi} \varphi_t, \operatorname{epi} \varphi) = 0 \textit{ for all } \rho \textit{ sufficiently large.}$$

When $\mathcal{X}$ is finite dimensional these notions are equivalent to the e-convergence as defined in 3.1.1. (cf. Proposition 2.2.4. and the fact that weak and strong topologies coincide when the space is finite dimensional) and in general AW-convergence implies M-convergence; see Attouch-Wets [3], Proposition 4.5.

Mosco obtained the following important characterization of M-convergence.

Lemma 3.1.3. *(Mosco [16], Lemma 1.10)*
$\varphi_t \xrightarrow{m} \varphi$ *if and only if each point ξ belongs to the reflexive Banach space $\mathcal{X}$ one has*

$$(\forall \xi_t \xrightarrow{w} \xi) \quad \liminf_{t \downarrow 0} \varphi_t(\xi_t) \geq \varphi(\xi) \tag{3.1.1}$$

$$(\exists t_n \downarrow 0) \quad (\exists \xi_n \xrightarrow{s} \xi) \quad \limsup_{n \to \infty} \varphi_{t_n}(\xi_n) \leq \varphi(\xi) \tag{3.1.2}$$

Equality then holds in (3.1.2) and $\lim_{n\to\infty} \varphi_{t_n}(\xi_n) = \varphi(\xi)$. *Note that φ_t strongly epi-converges to φ if the inequality (3.1.1) holds for every ξ_t converging strongly to ξ.*

Proof. It is straightforward to verify that (3.1.1) is equivalent to $w\text{-}\limsup_{t\downarrow 0}(\text{epi}\varphi_t)$ $\subset \text{epi}\varphi$ and that (3.1.2) is equivalent to $s\text{-}\liminf_{t\downarrow 0}(\text{epi}\,\varphi_t) \supset \text{epi}\varphi$ which, in view of

$$w\text{-}\limsup_{t\downarrow 0}(\text{epi}\varphi_t) \subset \text{epi}\varphi \subset s\text{-}\liminf_{t\downarrow 0}(\text{epi}\varphi_t)$$

and the definition of M-convergence, yield the desired result. □

Note. See Mosco [16] for a detailed proof.

The following result is about Mosco convergence of monotone sequences of convex functions.

Proposition 3.1.4. *Let $\{\varphi_t\}_{t>0}$ be a family of closed proper convex functions.*

(a) if φ_t increases as $t\downarrow 0$, then φ_t M-converges to $\sup_{t>0}\varphi_t$.

(b) if φ_t decreases as $t\downarrow 0$, then φ_t M-converges to $\text{cl}[\inf_{t>0}\varphi_t]$, *where 'cl' denotes the lower semicontinuous closure (weak or strong since* $\inf_{t>0}\varphi_t$ *is convex).*

Proof. First recall that $\varphi_t \xrightarrow{m} \varphi$ if and only if

(i) epi $\varphi \subset s\text{-}\liminf_{t\downarrow 0} \operatorname{epi}\varphi_t$.

(ii) $w\text{-}\limsup_{t\downarrow 0} \operatorname{epi}\varphi_t \subset \operatorname{epi}\varphi$.

(a) Let $\varphi = \sup_{t>0}\varphi_t$. Then $\operatorname{epi}\varphi = \bigcap_{t>0}\operatorname{epi}\varphi_t$, hence (i) holds. If $w\text{-}\limsup_{t\downarrow 0}$ epi $\varphi_t = \emptyset$, the inclusion is trivial. Suppose $(x,\alpha) \in w\text{-}\limsup_{t\downarrow 0}\operatorname{epi}\varphi_t$. Then there exist $t_n\downarrow 0$ and $(x_n,\alpha_n) \xrightarrow{w} (x,\alpha)$ with $(x_n,\alpha_n) \in \operatorname{epi}\varphi_{t_n}$. Since $\operatorname{epi}\varphi_{t_n}$ is decreasing, then for $k > 0$ we have $(x_n,\alpha_n) \in \operatorname{epi}\varphi_{t_k}$ for all $n > k$, hence, since $\operatorname{epi}\varphi_{t_k}$ is weakly closed, $(x,\alpha) \in \operatorname{epi}\varphi_{t_k}$. Therefore $(x,\alpha) \in \bigcap_{k>0}\operatorname{epi}\varphi_{t_k}$ which implies $(x,\alpha) \in \operatorname{epi}\varphi$. Thus $w\text{-}\limsup_{t\downarrow 0}\operatorname{epi}\varphi_t \subset \operatorname{epi}\varphi$, that is (ii) holds.

(b) Now let $\varphi = \operatorname{cl}[\inf_{t>0}\varphi_t]$ then $\operatorname{epi}\varphi = \operatorname{cl}[\bigcup_{t>0}\operatorname{epi}\varphi_t]$ is a closed convex set, hence epi φ is weakly closed. Therefore (ii) holds. Moreover, (i) holds, for $d\big((x,\alpha),\operatorname{epi}\varphi_t\big) \to 0$ as $t\downarrow 0$ for each $(x,\alpha) \in \operatorname{epi}\varphi$, because $\{\operatorname{epi}\varphi_t\}_{t>0}$ is increasing. □

Recall that the family of (extended) real valued functions $\{\varphi_t\}_{t>0}$ is said to converge *pointwise* (p-converges) to the function φ as $t\downarrow 0$, written $\varphi_t \xrightarrow{p} \varphi$, if for all $\xi \in \mathcal{X}$, $\varphi(\xi) = \lim_{t\downarrow 0}\varphi_t(\xi)$, or in other words

$$\limsup_{t\downarrow 0}\varphi_t(\xi) \le \varphi(\xi) \le \liminf_{t\downarrow 0}\varphi_t(\xi).$$

Even in the finite dimensional setting neither type of convergence implies the other :

Example 3.1.5. *Consider the sequence of closed convex functions on* $\mathbb{R}$*:*
For n= *1, 2,...*

$$f_n(x) := \begin{cases} x^n & ; x \in [0,1], \\ +\infty & ; \textit{otherwise}. \end{cases}$$

It is easy to verify that

$$f_n \xrightarrow{p} f := \begin{cases} 0 & ; x \in [0,1), \\ 1 & ; x = 1, \\ +\infty & ; \textit{otherwise}. \end{cases}$$

and

$$f_n \xrightarrow{e} f' := \begin{cases} 0 & ; x \in [0,1], \\ +\infty & ; \textit{otherwise}. \end{cases}$$

It turns out that e and p convergences agree on a subclass of closed convex functions.

Theorem 3.1.6. *(Salinetti-Wets* [**31**]*)*

Let $\mathcal{X}$ be finite dimensional and suppose φ_t, φ are closed convex functions. Then

(a) *if* $\mathrm{int}(\mathrm{dom}\,\varphi) \neq \emptyset$ *and* $\varphi_t \xrightarrow{p} \varphi$ *then* $\varphi_t \xrightarrow{e} \varphi$.

(b) *if* $\varphi_t \xrightarrow{e} \varphi$ *then* $\varphi_t \xrightarrow{p} \varphi$ *on* $\mathrm{int}(\mathrm{dom}\,\varphi)$.

Proof. See Salinetti and Wets [**31**], Corollary $2C$ and $3B$.

Salinetti and Wets also showed that two types of convergences are equivalent for the class of convex functions which are "equi-lower semicontinuous" and is the maximal class of convex functions which the equivalence can be obtained.

3.2. Epi-differentiation

We define the *epi-derivatives* replacing the pointwise limit of difference quotients of the classical directional derivative by the epi-limit. These epi-derivatives enjoy a rich and exhaustive calculus for a large class of functions used in optimization and we can even obtain optimality conditions. These optimality conditions are quite simple in nature; see Theorem 3.2.6. These calculus result can be found in the recent paper by Poliquin and Rockafellar [**21**].

Definition 3.2.1. *Let f be a function defined on a reflexive Banach space $\mathcal{X}$, and $x \in \mathcal{X}$ a point at which f is finite.*

(a) *We say that f is (Mosco) epi-differentiable at x if the first order difference quotients $\varphi_{x,t} : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$\varphi_{x,t}(\xi) := \big[f(x + t\xi) - f(x)\big]/t \;;\; \xi \in \mathcal{X} \;\; (t > 0)$$

(Mosco) epi-converge(as $t \downarrow 0$) to some function φ having $\varphi(0) \neq -\infty$. We then write $f_x^{\prime(m)}$ instead of φ and is the first order (Mosco) epi-derivative of f at x.

(b) *Let $v \in \mathcal{X}^*$(where $\mathcal{X}^*$ is the dual to $\mathcal{X}$) and consider the second order difference quotient functions*

$$\psi_{x,v,t}(\xi) := \big[f(x + t\xi) - f(x) - t\langle v, \xi\rangle\big]/(1/2)t^2 \;\;;\; \xi \in \mathcal{X} \;\; (t > 0)$$

where $\langle v, \xi \rangle := v(\xi)$. If these functions (Mosco) epi-converge(as $t \downarrow 0$) to some function ψ having $\psi(0) \neq -\infty$, then we say that f is twice (Mosco) epi-differentiable at x relative to v, and ψ is called the second order (Mosco) epi-derivative of f at x relative to v. We then write $f''^{(m)}_{x,v}$ instead of ψ.

When $\mathcal{X}$ is finite dimensional, one may characterized the epigraph of f'_x as the approximating cone (see Section 2.3) to epigraph of f at $(x, f(x))$, that is to say,

$$\mathrm{epi} f'_x = \lim_{t \downarrow 0} \left[\mathrm{epi} f - (x, f(x))\right]/t.$$

This is immediate from the relation

$$\mathrm{epi} \varphi_{x,t} = \left[\mathrm{epi} f - (x, f(x))\right]/t$$

($\varphi_{x,t}$ as defined in 3.2.1) and the definition of epi-convergence. Therefore, f is epi-differentiable at x if and only if the epigraph epi$f$ of f is derivable at $(x, f(x))$. In particular, if f is Clarke regular at x in the sense that the contingent cone to epi$f$ at $(x, f(x))$ equals the Clarke tangent cone there, then f is epi-differentiable at x.

The epi-differentiation of f is related to the proto-differentiation of a multifunction in the following way: Given $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$, we associate to it the multifunction $\Gamma : \mathcal{X} \rightrightarrows \mathbb{R}$, where

$$\Gamma(x) := \begin{cases} [f(x), \infty) & \text{if } f(x) < \infty, \\ \emptyset & \text{if } f(x) = \infty. \end{cases}$$

Then $\mathrm{gph}\,\Gamma = \mathrm{epi}\, f$, and the above characterization with Proposition 2.3.7. give us:

Γ is proto-differentiable at $(x, f(x))$ if and only if f is epi-differentiable at x.

The following proposition contains the basic features of second-order epi-derivatives.

Proposition 3.2.2. *The second-order Mosco epi-derivative $f''^{(m)}_{x,v}$, if it exists, is sequentially weakly lower semicontinuous, proper, positively homogeneous of degree 2 and $f''^{(m)}_{x,v}(0) = 0$.*

Proof. Any function ψ expressible as a Mosco epi-limit is sequentially weakly lower semicontinuous, because the limit set $\operatorname{epi}\psi$ is necessarily (sequentially) weakly closed by the definition of the M-convergence. Next we prove that $f''^{(m)}_{x,v}$ is positively homogeneous of degree 2, i.e.,

$$f''^{(m)}_{x,v}(\lambda\xi) = \lambda^2 f''^{(m)}_{x,v}(\xi) \text{ for all } \xi \in \mathcal{X} \text{ and } \lambda > 0.$$

Take $\xi \in \mathcal{X}$ and choose $t_n \downarrow 0$ and $\xi_n \xrightarrow{s} \xi$ with

$$\left[f(x + t_n\xi_n) - f(x) - t_n\langle v, \xi_n\rangle\right]/(1/2)t_n^2 \to f''^{(m)}_{x,v}(\xi).$$

It follows that

$$\lim_{n\to\infty} \left[f(x + (t_n/\lambda)\lambda\xi_n) - f(x) - (t_n/\lambda)\langle v, \lambda\xi_n\rangle\right]/(1/2)(t_n/\lambda)^2 = \lambda^2 f''^{(m)}_{x,v}(\xi),$$

and using (3.1.2) we deduce that $f''^{(m)}_{x,v}(\lambda\xi) \leq \lambda^2 f''^{(m)}_{x,v}(\xi)$. This inequality applied to $\lambda\xi$ and $1/\lambda$ instead of ξ and λ give us the converse inequality $\lambda^2 f''^{(m)}_{x,v}(\xi) \leq f''^{(m)}_{x,v}(\lambda\xi)$.

Now we prove $f''^{(m)}_{x,v}$ is proper. Suppose not, i.e., there must exists $\xi \in \mathcal{X}$ such that $f''^{(m)}_{x,v}(\xi) = -\infty$. Then for all $n \in \mathbb{N}$ we have $f''^{(m)}_{x,v}(\xi/n) = -\infty$ by the homogeneity of the M-limit. We deduce from the already established semicontinuity that $f''^{(m)}_{x,v}(0) = -\infty$, which contradicts the epi-differentiability of f at x. Finally, we have $f''^{(m)}_{x,v}(0) = 0$ as a result of homogeneity and $f''^{(m)}_{x,v}(0) > -\infty$. □

Definition 3.2.3. *A vector $v \in \mathcal{X}^*$ is a (Mosco) epi-gradient for f at x if for every $\xi \in \mathcal{X}$ one has $\langle v, \xi\rangle \leq f'^{(m)}_x(\xi)$. We simply say that f is twice (Mosco) epi-differentiable at x (without reference to a particular v) if f is (Mosco) epi-differentiable at x, has at least one (Mosco) epi-gradient there, and with respect to every (Mosco) epi-gradient v, is twice (Mosco) epi-differentiable at x relative to v.*

Proposition 3.2.4. *If f happens to be of class C^2 (Fréchet), one has*

$$f'^{(m)}_x(\xi) = \langle Df(x), \xi\rangle \text{ for all } \xi$$

and the unique Mosco epi-gradient at x is $v = Df(x)$, and
$f''^{(m)}_{x,v}(\xi) = \langle D^2 f(x)\xi, \xi\rangle$ *for all* ξ *when* $v = Df(x)$, *and the mapping* $\xi \to \langle D^2 f(x)\xi, \xi\rangle$ *is weakly lower semicontinuous.*

Where $Df(x)$ and $D^2 f(x)$ are the first and second order Fréchet differentials of f respectively.

Proof. Since f is C^1, for any $\xi \in \mathcal{X}$ and $t_n \downarrow 0$ there must exists $\xi_n \overset{s}{\to} \xi$ such that

$$[f(x + t_n\xi_n) - f(x)]/t_n \to \langle Df(x), \xi\rangle.$$

This also holds for any $\xi_t \overset{w}{\to} \xi$. Indeed, for any $\xi_t \overset{w}{\to} \xi$, using the Mean Value Theorem, we may find for each $t > 0$ an $\alpha_t \in (0, 1)$ such that with $x_t = x + t\alpha_t\xi_t$ one has

$$\frac{f(x + t\xi_t) - f(x)}{t} = Df(x_t)\xi_t.$$

Now, since ξ_t are norm bounded implies $x_t \overset{s}{\to} x$ and f is C^1 we have

$$\begin{aligned}\lim_{t\downarrow 0} \frac{f(x + t\xi_t) - f(x)}{t} &= \lim_{t\downarrow 0} Df(x_t)\xi_t \\ &= Df(x)\xi,\end{aligned}$$

the last equality by the fact that continuous linear functionals are weakly continuous. Thus we conclude that $f'^{(m)}_x(\xi) = \langle Df(x), \xi\rangle$.

Now let v be any Mosco epi-gradient of f at x, i.e., $f'^{(m)}_x(\xi) \geq \langle v, \xi\rangle$ for all $\xi \in \mathcal{X}$. So we have $Df(x)(\xi) \geq v(\xi)$ for all ξ and replacing ξ by $-\xi$ we get the converse inequality since the Df and v are linear functionals in $\mathcal{X}^*$. Then for all ξ, $Df(x)(\xi) = v(\xi)$ so that the Mosco epi-gradient $Df(x)$ is unique. The proof of the second part follows from a similar argument to the first. □

Proposition 3.2.5. *Suppose that f is twice Mosco epi-differential at x. Let g be any C^2 function with the mapping $\xi \to \langle D^2 g(x)\xi, \xi\rangle$ is weakly lower semicontinuous. Then the function $h = f + g$ is twice Mosco epi-differentiable at x. The Mosco epi-gradient of h at x are the vectors of the form $u = v + Dg(x)$ such that v is a Mosco epi-gradient of f at x, and for any such u one has*

$$h''^{(m)}_{x,u}(\xi) = f''^{(m)}_{x,v}(\xi) + \langle D^2 g(x)\xi, \xi\rangle.$$

Proof. Clearly

$$\frac{h(x+t\xi)-h(x)-t\langle u,\xi\rangle}{(1/2)t^2}=\frac{f(x+t\xi)-f(x)-t\langle v,\xi\rangle}{(1/2)t^2}+\frac{g(x+t\xi)-g(x)-t\langle Dg(x),\xi\rangle}{(1/2)t^2}$$

We denote these quotients respectively by $\Phi_t(\xi)$, $\Gamma_t(\xi)$ and $\Delta_t(\xi)$.

Since g is C^2, for $\xi \in \mathcal{X}$ and $t_n \downarrow 0$ there must exists $\xi_n \xrightarrow{s} \xi$ such that

$$\begin{aligned}\limsup_{n\to\infty}\Phi_{t_n}(\xi_n) &\le \limsup_{n\to\infty}\Gamma_{t_n}(\xi_n)+\limsup_{n\to\infty}\Delta_{t_n}(\xi_n)\\ &= f''^{(m)}_{x,v}(\xi)+\langle D^2g(x)\xi,\xi\rangle. \end{aligned} \tag{3.2.1}$$

On the other hand, for any $t_n \downarrow 0$ and $\xi_n \xrightarrow{w} \xi$ we have

$$\begin{aligned}\liminf_{n\to\infty}\Phi_{t_n}(\xi_n) &\ge \liminf_{n\to\infty}\Gamma_{t_n}(\xi_n)+\liminf_{n\to\infty}\Delta_{t_n}(\xi_n)\\ &\ge f''^{(m)}_{x,v}(\xi)+\langle D^2g(x)\xi,\xi\rangle, \end{aligned} \tag{3.2.2}$$

by the second part of Proposition 3.2.4. Relations (3.2.1) and (3.2.2) implies h is twice Mosco epi-differentiable at x and have that

$$h''^{(m)}_{x,u}(\xi)=f''^{(m)}_{x,v}(\xi)+\langle D^2g(x)\xi,\xi\rangle.$$

□

We conclude this section, illustrating how epi-derivatives can be used to obtain the optimality conditions for a wide class of functions involved in optimization.

Theorem 3.2.6. *(Optimality Conditions)*
Let $f : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ be a strongly lower semicontinuous function and x be a point where f is finite and twice Mosco epi-differentiable.

(a) (Necessary Condition) If f has a local minimum at x, then 0 is a Mosco epi-gradient of f at x and ${f''_{x,0}}^{(m)}(\xi) \ge 0$ for all $\xi \in \mathcal{X}$.

(b) (Sufficient Condition) If $\mathcal{X}$ is finite dimensional and 0 is a epi-gradient of f at x and $f''_{x,0}(\xi) > 0$ for all $\xi \ne 0$, then f has a local minimum at x in the strong sense, i.e.,

$$\exists\alpha>0 \text{ with } f(x')\ge f(x)+\alpha\|x'-x\|^2 \quad \text{for all } x' \text{ near } x \tag{3.2.3}.$$

Proof.

(a) Because $f_x'^{(m)}(\xi)$ exists at x, then for any $\xi \in \mathcal{X}$ there exists $t_n \downarrow 0$ and $\xi_n \xrightarrow{s} \xi$ such that

$$\frac{f(x + t_n\xi_n) - f(x)}{t_n} \to f_x'^{(m)}(\xi).$$

If f has a local minimum at x then $\left[f(x + t_n\xi_n) - f(x)\right]/t_n \geq 0$ for n sufficiently large and going to the limit we deduce that $f_x'(\xi) \geq 0$. This implies that 0 is a Mosco epi-gradient of f at x. Since f is twice Mosco epi-differentiable at x, we also have

$$\frac{f(x + t_n\xi_n) - f(x)}{(1/2)t_n^2} \to f_{x,0}''^{(m)}(\xi).$$

Hence we deduce that $f_{x,0}''^{(m)}(\xi) \geq 0$.

(b) Suppose this is not the case. Then for all $\alpha > 0$ there exist $x_n^\alpha \to x$ such that $x_n^\alpha \neq x$ and $f(x_n^\alpha) < f(x) + \alpha\|x_n^\alpha - x\|^2$. Take $\xi_n := \frac{x_n^\alpha - x}{\|x_n^\alpha - x\|}$, which converges (w.l.o.g.) to some ξ with $\|\xi\| = 1$, so that setting $t_n := \|x_n^\alpha - x\|$ we get

$$f_{x,0}''(\xi) \leq \liminf_{n\to\infty} \frac{f(x + t_n\xi_n) - f(x)}{(1/2)t_n^2} \leq 2\alpha.$$

This is true for all α positive, and hence $f_{x,0}''(\xi) \leq 0$, a contradiction. □

CHAPTER 4

AMENABLE FUNCTIONS

4.1. Introduction

In the previous chapters, we introduced two concepts of generalized differentiation, under the designations of epi-differentiation and proto-differentiation for functions and multifunctions. However, the calculus we have developed has been for smooth functions; it is natural to ask to what extent it would be possible to treat more general functions. Recently, significant progress on this issue has been made by Rockafellar introducing a special class of functions, termed as " amenable". See [**26**], [**28**], [**29**].

This class is general enough to handle most applications in optimization yet well-behaved enough to carry out a sharper form of subdifferential calculus. Amenable functions enjoy above properties because their inherent nature ensure the wide applicability and capture the local aspects of convexity and smoothness as we see by the definitions and examples below. See also [**20**] – [**22**].

Definition 4.1.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. A function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is amenable at $\bar{x} \in \operatorname{dom} f$, if on some (strong)open neighbourhood of V of $\bar{x}$ there is a C^1(Fréchet) mapping $F : V \to \mathcal{Y}$ and a proper, lower semicontinuous(lsc), convex function $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ such that $f(x) = g(F(x))$ for $x \in V$ and the following constraint qualification is satisfied at $\bar{x}$:*

$$\mathbb{R}_+\big(\operatorname{dom} g - F(\bar{x})\big) - DF(\bar{x})\mathcal{X} = \mathcal{Y} \tag{4.1.1}$$

where $DF(\bar{x})$ is the Fréchet differential of F at $\bar{x}$.

To study the second-order differential properties the following refinement of amenability is useful. First we need to introduce the piecewise linear-quadratic functions:

Definition 4.1.2. *A function $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ with effective domain $D = \{u \mid g(u) < +\infty\}$ will be called piecewise linear-quadratic if D can be expressed* *as the union of finitely many sets D_j (for $j \in J$, a finite index set), such that D_j is a convex (polyhedral) set given by*

$$D_j = \cap_{i=1}^{k_j} E_i^j, \text{ where } E_i^j = \{\, y \in \mathcal{Y} \,|\, \langle (y_i^j)^*, y \rangle \le a_i^j,\ (y_i^j)^* \in \mathcal{Y}^*,\ a_i^j \in \mathbb{R} \,\},$$

and the restriction of g to D_j is a polynomial function Q_j :

$$Q_j(u) = \tfrac{1}{2} B_j(u, u) + \langle l_j^*, u \rangle + c_j \text{ where } c_j \in \mathbb{R},\ l_j^* \in \mathcal{Y}^* \text{ and } B_j : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$$

is a continuous symmetric bilinear function.

Definition 4.1.3. *A function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is fully amenable at $x \in \operatorname{dom} f$, if the conditions in definition 4.1.1 can be satisfied with extra stipulation that F is a C^2(Fréchet) mapping and g is piecewise linear-quadratic convex function.*

In the finite dimensional case, we see that the constraint qualification (4.1.1) is equivalent to

$$N_{\operatorname{dom} g}\big(F(\bar{x})\big) \cap \ker\big(DF(\bar{x})^*\big) = 0. \tag{4.1.2}$$

Where $N_{\operatorname{dom} g}\big(F(\bar{x})\big) := \{\xi \mid \langle \xi, z - F(\bar{x}) \rangle \le 0 \text{ for all } z \text{ in } \operatorname{dom} g\}$ is the normal cone to $\operatorname{dom} g$ at $F(\bar{x})$ and $DF(\bar{x})^*$ denote the transpose of the matrix $DF(x)$. In other words, there is no $\xi \neq 0$ in $N_{\operatorname{dom} g}\big(F(\bar{x})\big)$ with $DF(x)^*\xi = 0$. This is known as the *basic constraint qualification* used by Rockafellar in **[26]** defining amenable functions in finite dimensions.

To see that (4.1.1) is equivalent to (4.1.2), first notice that the condition (4.1.1) is equivalent to $\Big(\mathbb{R}^+\big(\operatorname{dom} g - F(\bar{x})\big)\Big)^\circ \cap \big(DF(\bar{x})\mathcal{X}\big)^\circ = 0$. Then one can easily identify that $\Big(\mathbb{R}^+\big(\operatorname{dom} g - F(\bar{x})\big)\Big)^\circ = N_{\operatorname{dom} g}\big(F(\bar{x})\big)$ and $\big(DF(x)\mathcal{X}\big)^\circ = \ker\big(DF(\bar{x})^*\big)$.

Here C° denotes the *polar* of a cone C of $\mathcal{X}$:

$$C^\circ := \{\, x^* \in \mathcal{X}^* \,|\, \langle x^*, y \rangle \le 0 \text{ for all } y \text{ in } C \,\}.$$

In general (4.1.1) implies (4.1.2) and (4.1.2) implies

$$\operatorname{cl}\big[\mathbb{R}_+\big(\operatorname{dom} g - F(\bar{x})\big) - DF(x)\mathcal{X}\big] = \mathcal{Y}.$$

The closure operation is not needed in finite dimensions (cf. **[23]**, Convex Analysis).

Example 4.1.4. *Any proper, lower semicontinuous, convex function f over a Banach space $\mathcal{X}$ is amenable at all points in* $\operatorname{dom} f$. *Any convex, piecewise linear-quadratic function f is fully amenable at all points in* $\operatorname{dom} f$.

Here the mapping F in definitions 4.1.1 and 4.1.3 can be taken to be the identity. Note that the condition (4.1.1) is satisfied at $\bar{x} \in \operatorname{dom} f$ since $DF(\bar{x}) =$ identity operator and $\mathcal{X} = \mathcal{Y}$ gives $DF(\bar{x})\mathcal{X} = \mathcal{Y}$.

Example 4.1.5. *Any C^1 function f is everywhere amenable, whereas any C^2 function f is everywhere fully amenable.*

Take $\mathcal{Y} = \mathbb{R}$ and $g(w) = w$ in the definitions 4.1.1 and 4.1.3. Then $\operatorname{dom} g = \mathbb{R}$, so that condition (4.1 .1) is trivially satisfied.

Example 4.1.6. *The class of functions which can be written as the maximum of finitely many smooth functions is amenable and we focus on subdifferential properties of these max-functions in Chapter 5.*
If $f = \max\{f_1, \ldots, f_k\}$ for a family of C^1 functions $f_i : \mathcal{X} \to \mathbb{R}$. Then f is everywhere amenable. If each f_i is C^2, f is everywhere fully amenable.
Set $\mathcal{Y} = \mathbb{R}^k$ and $F(x) = \big(f_1(x), \ldots, f_k(x)\big)$ along with $g(w_1, \ldots, w_k) = \max\{w_1, \ldots, w_k\}$. Then g is piecewise linear and $\operatorname{dom} g = \mathbb{R}^k$ so that the condition (4.1.1) is automatically satisfied.

Example 4.1.7. *Consider the mathematical programming problem*

$$(\mathcal{P}_0) \qquad \text{minimize } f_0(x) \text{ over all } x \in E \text{ satisfying } f_i(x) \in I_i \text{ for } i = 1, \ldots, k,$$

where the functions $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 0, 1, \ldots, k$, are of class C^2, the nonempty set E is polyhedral (implying convex), and each nonempty set I_i is a closed (but not necessarily bounded) interval in $\mathbb{R}$.
The problem $(\mathcal{P}_0)$ can be reformulated as

$$(\mathcal{P}_0') \qquad \text{minimize } f(x) \text{ over all } x \in \mathbb{R}^n,$$

where $f(x) = g\big(F(x)\big)$ for

$$F(x) = \big(f_0(x), f_1(x), \ldots, f_k(x), x\big), \quad g(u_0, u_1, \ldots, u_k, x) = u_0 + \delta_D(u_1, \ldots, u_k, x)$$

with $D = I_1 \times \ldots \times I_k \times E$.

Let $C := \{ x \in E \mid f_i(x) \in I_i;\ i = 1, \ldots, k \}$. *Then* f *is fully amenable at any point* $\bar{x} \in C$ *at which the following basic constraint qualification is satisfied:*

$$\begin{cases} \text{there is no } (y_1, \ldots, y_k) \neq (0, \ldots, 0) \text{ with }, \\ y_i \in N_{I_i}\big(f_i(\bar{x})\big) \text{ and } -\sum_{i=1}^k y_i \nabla f_i(\bar{x}) \in N_E(x). \end{cases}$$

To see that f *is fully amenable at* $\bar{x}$, *observe that* D *is a polyhedral set,* g *is piecewise linear and the condition (4.1.2) is reduced to the one above since* $N_D\big(F(x)\big) = N_{I_1}\big(f_1(\bar{x})\big) \times \ldots \times N_{I_k}\big(f_k(\bar{x})\big) \times N_E(\bar{x})$ *and* $\nabla F(\bar{x}) = \big(\nabla f_0(\bar{x}), \nabla f_1(x), \ldots, \nabla f_k(x), u\big)$, *where* $u = (1, \ldots, 1) \in \mathbb{R}^n$.

Example 4.1.8. *Consider the constrained nonlinear programming problem*

$$\begin{aligned} (\mathcal{P}) \qquad & \text{minimize } f_0(x) \\ & f_i(x) \leq 0,\ i = 1, \ldots, s, \\ & f_i(x) = 0,\ i = s+1, \ldots, k, \\ & x \in E. \end{aligned}$$

and a penalty representation of $(\mathcal{P})$:

$$(\mathcal{P}_{pen}) \qquad \text{minimize } f_0(x) + \sum_{i=1}^{s} r_i \big[f_i(x)\big]_+ + \sum_{i=s+1}^{k} r_i \big|f_i(x)\big|$$

where $r_i > 0$ *for all* i, $\big[f(x)\big]_+ = \max\{0, f(x)\}$, f_i *and* E *as in Example 4.1.7. The objective of the problem* $(\mathcal{P}_{pen})$ *can be easily reformulated as maximum of finitely many* C^2 *functions, and hence by Example 4.1.6. it is fully amenable.*

Although amenability may seem to be a condition focussed on a single point at a time, it is truly a local condition on a neighbourhood of a point. This is because the constraint qualification (4.1.1) holds not just for x, but for all x in some neighbourhood of $\bar{x}$ relative to dom f; this is shown in the following theorem.

Theorem 4.1.9. *If the constraint qualification (4.1.1) holds at a point* $x \in \operatorname{dom} f$ *then it holds at all points* $\tilde{x} \in \operatorname{dom} f$ *in some (strong)neighbourhood of* x.

This can be proved using a result of S.Kurcyusz and J.Zowe ([**15**]. Theorem 5.2). In that paper they studied a " regularity condition" for the following mathematical programming problem in Banach spaces:

$$(\mathcal{P}): \ \text{minimize } f(x) \text{ subject to } x \in C \text{ and } h(x) \in K$$

where $f : \mathcal{X} \to \mathbb{R}$ is differentiable and $h : \mathcal{X} \to \mathcal{Y}$ is C^1, C is closed convex subset of $\mathcal{X}$ and K is a closed convex cone in $\mathcal{Y}$ with vertex at 0. The set of all feasible solutions for $(\mathcal{P})$ is denoted by M, i.e., $M = C \cap h^{-1}(K)$.
A point $\bar{x} \in M$ is said to be *regular* if

$$Dh(\bar{x})C(\bar{x}) - K\big(h(\bar{x})\big) = \mathcal{Y} \tag{4.1.3}$$

Where $C(\bar{x}) := \{\lambda(c - \bar{x}) | c \in C, \lambda \geq 0\}$, $K(z) := \{k - \lambda z \mid k \in K, \lambda \geq 0\}$ $(z \in \mathcal{Y})$. They proved that the condition (4.1.3) is a local condition:

Theorem 4.1.10. *(Kurcyusz & Zowe)*
Suppose $\bar{x}$ is a regular point for problem $(\mathcal{P})$. If $\|Dh(\bar{x}) - D\tilde{h}(\tilde{x})\|$, $haus\big((C - \bar{x})_1, (\tilde{C} - \tilde{x})_1\big)$ and $haus\Big(\big(K - h(\bar{x})\big)_1, \big(\tilde{K} - \tilde{h}(\tilde{x})\big)_1\Big)$ are small enough, then $\tilde{x}$ is a regular point for $(\tilde{\mathcal{P}})$,
where $\tilde{x}$ is a feasible point for the perturbed problem:

$$(\tilde{\mathcal{P}}): \ \text{minimize } f(x) \text{ subject to } x \in \tilde{C} \text{ and } \tilde{h}(x) \in \tilde{K},$$

haus denote the Hausdorff distance as defined in Section 2.2, and

$$(C - \bar{x})_1 := (C - \bar{x}) \cap B_{\mathcal{X}}, \quad (K - \bar{x})_1 := (K - \bar{x}) \cap B_{\mathcal{Y}}$$

where $B_{\mathcal{X}}$ and $B_{\mathcal{Y}}$ are closed unit balls in $\mathcal{X}$ and $\mathcal{Y}$ respectively.

Proof of Theorem 4.1.9. Let $\bar{x} \in \operatorname{dom} f$ at which f is amenable. Thus f has a local representation $f = g \circ F$ for all x in a neighbourhood V of $\bar{x}$ and the constraint qualification (4.1.1) is satisfied at $\bar{x}$, i.e.,

$$\mathbb{R}_+\big(\operatorname{dom} g - F(\bar{x})\big) - DF(\bar{x})\mathcal{X} = \mathcal{Y}.$$

To see amenability is a local condition, it suffices to show that the constraint qualification (4.1.1) is satisfied in a neighbourhood of $\bar{x}$ relative to $\operatorname{dom} f \cap V$. This easily follows from Theorem 4.1.10, once we established:

(i) regularity condition (4.1.3) is equivalent to the constraint qualification (4.1.1) for some choice of h, C and K.

(ii) assumptions in Theorem 4.1.10. are satisfied.

To see (i), simply take $h := F$, $C := \mathcal{X}$, $K := \mathbb{R}_+ \operatorname{dom} g$ and $x \in \operatorname{dom} f \cap V$. With this also set $\tilde{h} := F$, $\tilde{C} := \mathcal{X}$, $\tilde{K} := \mathbb{R}_+ \operatorname{dom} g$ in Theorem 4.1.10. Then we have $\|DF(\bar{x}) - DF(\tilde{x})\| \leq K\|\bar{x} - \tilde{x}\|$ for some $K > 0$ and for all $\tilde{x}$ in $\operatorname{dom} f \cap V$, $\operatorname{haus}\big((\mathcal{X} - \bar{x})_1, (\mathcal{X} - \tilde{x})_1\big) = 0$ and the quantity $\operatorname{haus}\Big(\big(\mathbb{R}_+ \operatorname{dom} g - F(x)\big)_1, \big(\mathbb{R}_+ \operatorname{dom} g - F(\tilde{x})\big)_1\Big)$ is small. The latter is true because, by the definition of Hausdorff distance for any $y \in \mathbb{R}_+ \operatorname{dom} g - F(\bar{x})$, one can find $\tilde{y} \in \mathbb{R}_+ \operatorname{dom} g - F(\tilde{x})$ such that $d(y, \tilde{y}) = \|F(\bar{x}) - F(\tilde{x})\|$. Since F is continuous we can make $d(y, \tilde{y})$ arbitrary small by choosing a suitable neighbourhood of $\bar{x}$. The above arguments imply that the existence of a neighbourhood $\bar{V}$ of $\bar{x}$ relative to $\operatorname{dom} f$, satisfying (ii) and then by the theorem we have

$$\mathbb{R}_+\big(\operatorname{dom} g - F(\tilde{x})\big) - DF(\tilde{x})\mathcal{X} = \mathcal{Y} \text{ for all } \tilde{x} \in \bar{V}.$$

□

4.2. Epi-differentiability of Amenable functions

Amenable functions are not in general differentiable (not even convex) in the usual sense. It turns out that epi-differentiation is more appropriate for this class of functions. In fact, fully amenable functions, in finite dimensional space, are twice epi-differentiable; see Rockafellar [26]. Moreover, proto-differentiability of the subgradient mapping (see definition below) is equivalent to the second-order epi-differentiability; see Theorem 4.2.9.

First, we need to introduce the *subdifferential* of a function which is analogous to the differential in case of differentiable functions. There are several ways of defining the subdifferential of a nonconvex functions, but they all coincide as far as fully amenable functions are concerned. See Corollary 4.2.8.

Definition 4.2.1. *Let $C \subset \mathcal{X}$ and x in C. The Clarke normal cone to C at x, denoted $N_C(x)$, is the cone polar to $T_C(x)$:*

$$N_C(x) := \big(T_C(x)\big)^\circ = \big\{\, v \in \mathcal{X}^* \;\big|\; \langle v, \xi\rangle \le 0 \text{ for all } \xi \text{ in } T_C(x) \,\big\}$$

we recall that $T_C(x)$ is the Clarke tangent cone to C at x (Definition 2.3.4).

Definition 4.2.2. *Let $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be finite at a point x. Then the (generalized) subdifferential $\partial f(x)$ of f at x is given by*

$$\partial f(x) := \big\{\, v \in \mathcal{X}^* \;\big|\; (v, -1) \in N_{\operatorname{epi} f}\big(x, f(x)\big) \,\big\}.$$

A functional v of ∂f is termed a subgradient.

The above definition, which of course is analogous to the fact that $\big(f'(x), -1\big)$ is normal to the graph of the smooth function $f : \mathbb{R} \to \mathbb{R}$ and one can easily see that the subgradient set reduces to the gradient $f'(x)$ when f is smooth.
The following Lemma gives another useful characterization.

Lemma 4.2.3. *Let $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be a Clarke regular function (Definition 2.3.6) and x be a point in $\operatorname{dom} f$. Then*

$$\partial f(x) = \big\{\, v \in \mathcal{X}^* \;\big|\; f'_s(x;\xi) \ge \langle v, \xi\rangle \text{ for all } \xi \text{ in } \mathcal{X} \,\big\} \tag{4.2.1}$$

where $f'_s(x;\xi) := \liminf\limits_{\substack{\xi_n \xrightarrow{s} \xi \\ t_n \downarrow 0}} \big[f(x + t_n\xi_n) - f(x)\big]/t_n$.

Proof. Suppose first that $v \in \partial f(x)$. We wish to prove that $f'_s(x;\xi) \ge \langle v, \xi\rangle$ for all ξ. By the definition of ∂f and the regularity of f, we have

$$(v, -1) \in N_{\operatorname{epi} f}\big(x, f(x)\big) = \Big(T_{\operatorname{epi} f}\big(x, f(x)\big)\Big)^\circ = \Big(K_{\operatorname{epi} f}\big(x, f(x)\big)\Big)^\circ.$$

By the definition of $f'_s(x;\xi)$, for any ξ, there exists $t_n \downarrow 0$ and $\xi_n \xrightarrow{s} \xi$ with

$$\frac{\big(x + t_n\xi_n, f(x + t_n\xi_n)\big) - \big(x, f(x)\big)}{t_n} \xrightarrow{s} \big(\xi, f'_s(x;\xi)\big).$$

which implies $\big(\xi, f'_s(x;\xi)\big)$ in $T_{\mathrm{epi}\, f}\big(x, f(x)\big)$, by the regularity of f. Since $(v,-1)$ belongs to the polar of this set we have $\langle (v,-1), (\xi, f'_s(x;\xi))\rangle \le 0$ for all ξ and hence the result.

Conversely, assume that $v \in \mathcal{X}^*$ with $f'_s(x;\xi) \ge \langle v, \xi\rangle$ for all ξ. It is enough to show that $(v,-1) \in N_{\mathrm{epi}\, f}\big(x, f(x)\big)$, or equivalently $\langle (v,-1), (\xi,\alpha)\rangle \le 0$ (i.e., $\langle v,\xi\rangle \le \alpha$) for all (ξ,α) in $T_{\mathrm{epi}\, f}\big(x, f(x)\big)$. For each (ξ,α) there must exist (x_n,α_n) in epi f such that $\dfrac{(x_n,\alpha_n) - (x, f(x))}{t_n} \xrightarrow{s} (\xi,\alpha)$. Now, by our assumption we have

$$\frac{f(x_n) - f(x) - \langle v, x_n - x\rangle}{\|x_n - x\|} \ge 0 \ \text{ for } \ t_n = \|x_n - x\|$$

which also implies $\dfrac{\alpha_n - f(x) - \langle v, x_n - x\rangle}{\|x_n - x\|} \ge 0$, since (x_n,α_n) in epi f and going to the limit we deduce that $\alpha - \langle v,\xi\rangle \ge 0$, which completes the proof. □

When f happens to be Mosco epi-differentiable and Clarke regular, all above forms of subdifferentials coincide as we see next.

Lemma 4.2.4. *Assume f is Mosco epi-differentiable at x. Then*

$$f'^{(m)}_x(\xi) = f'_s(x;\xi) \ \text{ for all } \xi.$$

Proof. Let $f'_w(x;\xi) := \liminf\limits_{\substack{\xi_n \xrightarrow{w} \xi \\ t_n \downarrow 0}} \big[f(x + t_n\xi_n) - f(x)\big]/t_n$. Then by the characterization of Mosco epi-limit (cf. Lemma 3.1.3) we deduce that

$$f'^{(m)}_x(\xi) \le f'_w(x;\xi) \le f'_s(x;\xi) \le f'^{(m)}_x(\xi)$$

and hence the result. □

Theorem 4.2.5. *Let $\mathcal{X}$ be a reflexive Banach space and f be Clarke regular at a point x in* dom f. *If f is Mosco epi-differentiable at x, then*

$$\partial f(x) = \big\{\, v \in \mathcal{X}^* \;\big|\; f'^{(m)}_x(\xi) \ge \langle v,\xi\rangle \ \text{ for all } \xi \text{ in } \mathcal{X} \,\big\} \tag{4.2.2}$$

$$= \big\{\, v \in \mathcal{X}^* \;\big|\; f(y) \ge f(x) + \langle v, y - x\rangle + o(\|y - x\|) \,\big\} \tag{4.2.3}$$

Here $o(\|y-x\|)$ *refers to an expression having the property that* $\frac{o(y-x)}{\|y-x\|} \to 0$ *as* $(y-x) \to 0$.

Proof. The first equality is obvious from Lemmas 4.2.3 and 4.2.4 . To see the second equality it is enough to show that the right side of (4.2.3) agrees with the same in (4.2.2). Suppose v satisfies (4.2.3) and choose ξ in $\mathcal{X}$, $t_n \downarrow 0$, and $\xi_n \xrightarrow{s} \xi$ such that $[f(x+t_n\xi_n) - f(x)]/t_n \to f_x^{\prime(m)}(\xi)$. Then from (4.2.3) we have

$$\langle v, \xi_n \rangle + \frac{o(\|t_n\xi_n\|)}{t_n} \le \frac{f(x+t_n\xi_n) - f(x)}{t_n}$$

and going to the limit we deduce $\langle v, \xi \rangle \le f_x^{\prime(m)}(\xi)$. This is true for any ξ we conclude that v belongs to the right side of (4.2.2).

Conversely, suppose v in $\mathcal{X}^*$ with $f_x^{\prime(m)}(\xi) \ge \langle v, \xi \rangle$ for all ξ. If v does not belong to the right side of (4.2.3) one can find $\alpha < 0$ and $y_n \xrightarrow{s} x$ such that

$$\frac{f(y_n) - f(x) - \langle v, y_n - x \rangle}{\|y_n - x\|} \le \alpha.$$

Since $\mathcal{X}$ is reflexive we may extract a subsequence, still denote y_n, such that $\xi_n := \frac{(y_n - x)}{\|y_n - x\|} \xrightarrow{w} \xi$ for some ξ in $\mathcal{X}$, and setting $t_n = \|y_n - x\|$ we get

$$\frac{f(x+t_n\xi_n) - f(x)}{t_n} \le \langle v, \xi_n \rangle + \alpha.$$

Letting $n \to \infty$ and the characterization of Mosco epi-limit (cf. Lemma 3.1.3) we get

$$f_x^{\prime(m)}(\xi) \le \langle v, \xi \rangle + \alpha < \langle v, \xi \rangle,$$

a contradiction. □

We shall begin the further study of epi-differentiability of fully amenable functions by first stating the following known result of the first-order epi-derivatives.

Theorem 4.2.6. *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be reflexive Banach spaces,* f *be fully amenable as in definition 4.1.3, and* $f = g \circ F$ *is a local representation around* $\bar{x} \in \operatorname{dom} f$ *in the sense required in that definition. Then* f *is Mosco epi-differentiable at* x *with*

$$f_{\bar{x}}^{\prime(m)}(\xi) = g_{F(\bar{x})}^{\prime(m)}\big(DF(\bar{x})\xi\big) \text{ and } \partial f(\bar{x}) = DF(\bar{x})^* \partial g\big(F(\bar{x})\big)$$

where $DF(\bar{x})^*$ represents the adjoint operator of $DF(\bar{x})$.

Proof. This result was obtained in finite dimensions in Rockafellar [**26**] and generalized to infinite dimensions by Cominetti [**11**]. ◻

The following theorem shows that fully amenability is much stronger condition than Clarke regularity.

Theorem 4.2.7. *Let f be fully amenable at $\bar{x} \in \operatorname{dom} f$. Then f is Clarke regular at $\bar{x}$.*

Proof. We first show that the set $C := \operatorname{dom} f$ is Clarke regular. The set $C = \{\, x \mid F(x) \in \operatorname{dom} g \,\}$ can be represented in the form

$$C = \{\, x \in \mathcal{X} \mid \langle y_i^*, F(x)\rangle \le a_i \text{ for } i = 1, \dots, p\,; \quad \langle y_i^*, F(x)\rangle = a_i \text{ for } i = p+1, \dots, q \,\},$$

for some choice of $y_i^* \in Y^*$. This can be viewed as

$$C = \{\, x \in \mathcal{X} \mid h_i(x) \le 0 \text{ for } i = 1, \dots, p\,; \;\; h_i(x) = 0 \text{ for } i = p+1, \dots, q \,\},$$

where $h_i(x) = \langle y_i^*, F(x)\rangle - a_i$.
Now let $h := (h_1, \dots, h_q) : \mathcal{X} \to \mathbb{R}^q$ and

$$M = \{\, (u_1, \dots, u_q) \mid u_i \le 0 \text{ for } i = 1, \dots, p\,; \;\; u_i = 0 \text{ for } i = p+1, \dots, q \,\}.$$

Then $C = h^{-1}(M)$. Note that the set M is convex and hence $T_M\big(h(x)\big) = K_M\big(h(x)\big)$. Then the Clarke regularity of C follows by a result of Aubin and Frankowska ([**5**], Corollary 4.3.4) :
It says that if a continuously differentiable map h from a Banach space $\mathcal{X}$ to a finite dimensional space $\mathcal{Y}$ and a closed convex set $M \subset \mathcal{Y}$ satisfying the constraint condition

$$Dh(\bar{x})\mathcal{X} + T_M\big(h(\bar{x})\big) = \mathcal{Y}$$

then at any $\bar{x} \in h^{-1}(M)$, the set $h^{-1}(M)$ is Clarke regular.

In our case the above constraint condition is equivalent to

$$Dh(\bar{x})\mathcal{X} + T_M\big(h(\bar{x})\big) = \mathbb{R}^q, \text{ or eqivalently}$$
$$\ker\big(Dh^*(\bar{x})\big) \cap N_M\big(h(\bar{x})\big) = \{0\} \text{ for any } \bar{x} \in C.$$

Where

$$N_M\big(h(\bar{x})\big) = \big\{ (\lambda_1, \dots, \lambda_q) \,\big|\, \lambda_i \geq 0 \text{ for } i = 1, \dots, p \text{ active}$$
$$\lambda_i = 0 \text{ for } i = 1, \dots, p \text{ inactive } \big\}.$$

This means that the constraint condition is satisfied if the only multipliers $\lambda_i \geq 0,\ i = 1, \dots, p$ such that

$$\textstyle\sum_{i=1}^q \lambda_i Dh_i(\bar{x}) = 0 \text{ is } \quad \lambda_1 = \dots = \lambda_q = 0,$$

$$\text{i.e., } DF(\bar{x})^*\big(\textstyle\sum_{i=1}^q \lambda_i y_i^*\big) = 0 \text{ is } \lambda_1 = \dots = \lambda_q = 0.$$

Because $\sum_{i=1}^q \lambda_i y_i^* \in N_{\mathrm{dom}\, g}\big(F(\bar{x})\big)$ and our constraint qualification it follows that $\sum_{i=1}^q \lambda_i y_i^* = 0$. So the constraint condition is satisfied if the only $(\lambda_1, \dots, \lambda_q)$ such that $\sum_{i=1}^q \lambda_i y_i^* = 0$ is $\lambda = (\lambda_1, \dots, \lambda_q) = 0$. This can always be assumed without lose of generality because of the following claim. By [**5**], Corollary 4.3.4 C is Clarke regular.

Claim. *Without lose of generality we may assume that y_i^* are chosen such that if $\sum_{i=1}^q \lambda_i y_i^* = 0$ with $\lambda_i \geq 0$ for $i = 1, \dots, p$; $\lambda_i = 0$ for $i = 1, \dots, p$ inactive, implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_q = 0$.*

Proof of Claim. For $y \in \mathcal{Y}$, consider the system

$$(*) \begin{cases} \langle y_i^*, y \rangle \leq a_i \text{ for } i = 1, \dots, p. \\ \langle y_i^*, y \rangle = a_i \text{ for } i = p+1, \dots, q. \end{cases}$$

Assume $\sum_{i=1}^q \lambda_i y_i^* = 0$ with say $\lambda_1 > 0$ then

$$y_1^* = -\textstyle\sum_{i=2}^p (\lambda_i/\lambda_1) y_i^* - \sum_{i=p+1}^q (\lambda_i/\lambda_1) y_i^*.$$

Which gives

$$\langle y_1^*, F(\bar{x}) \rangle = -\textstyle\sum_{i=2}^p (\lambda_i/\lambda_1) \langle y_i^*, F(\bar{x}) \rangle - \sum_{i=p+1}^q (\lambda_i/\lambda_1) \langle y_i^*, F(\bar{x}) \rangle,$$
$$a_1 = -\textstyle\sum_{i=2}^p (\lambda_i/\lambda_1) a_i - \sum_{i=p+1}^q (\lambda_i/\lambda_1) a_i. \qquad (**)$$

Choose any y in $(*)$. Then we have $-(\lambda_i/\lambda_1)\langle y_i^*, y\rangle \geq -(\lambda_i/\lambda_1)a_i$ and hence

$$\begin{aligned}\langle y_1^*, y\rangle &= -\textstyle\sum_{i=2}^{p}(\lambda_i/\lambda_1)\langle y_i^*, y\rangle - \sum_{i=p+1}^{q}(\lambda_i/\lambda_1)\langle y_i^*, y\rangle,\\ &\geq -\textstyle\sum_{i=2}^{p}(\lambda_i/\lambda_1)a_i - \sum_{i=p+1}^{q}(\lambda_i/\lambda_1)a_i,\\ &= a_1. \qquad (\text{ by } (**))\end{aligned}$$

Thus, not only do we have $\langle y_i^*, y\rangle \leq a_1$ but also $\langle y_1^*, y\rangle \geq a_1$ which implies $\langle y_1^*, y\rangle = a_1$. By renumbering if necessary we may assume that $\lambda_i = 0$ for $i = 1, 2, \ldots, p$.

Assume $\sum_{i=p+1}^{q} \lambda_i y_i^* = 0$ with say $\lambda_{p+1} \neq 0$ then $y_{p+1}^* = -\sum_{i=p+2}^{q}(\lambda_i/\lambda_{p+1})y_i^*$. So that any y that satisfy $\langle y_i^*, y\rangle = a_i$ for $i = p+2, \ldots, q$ has

$$\langle y_{p+1}^*, y\rangle = -\textstyle\sum_{i=p+2}^{q}(\lambda_i/\lambda_{p+1})\langle y_i^*, y\rangle = -\sum_{i=p+2}^{q}(\lambda_i/\lambda_{p+1})a_i = a_{p+1}.$$

Thus, the equality $\langle y_{p+1}^*, y\rangle = a_{p+1}$ can be dropped from the system. $\square$

Now, we proceed to show that f is Clarke regular at $\bar{x} \in C$.
Since f is fully amenable at $\bar{x}$, it has the local representation $f(x) = g\big(F(x)\big)$. The function g, being piecewise linear-quadratic, is locally Lipschitzian relative to its effective domain D. Let $\mu > 0$ be a Lipschitz constant that works for a neighbourhood of $u = F(\bar{x})$, and define

$$\tilde{g}(u') = \inf_{w\in\mathcal{Y}} \big\{ g(w) + \tilde{\mu}\|u' - w\| \big\}, \quad \text{where } \tilde{\mu} > \mu.$$

i.e., $\tilde{g}$ is the infimal convolution of g and $\tilde{\mu}\|\cdot\|$, since g is convex so is $\tilde{g}$. By the choice of $\tilde{\mu}$ one will have

$$\tilde{g}(u') = g(u') \text{ for all } u' \text{ in some neighbourhood of } u.$$

(namely any neighbourhood where μ acts as a Lipschitz constant) In particular, $\tilde{g}$ is finite at certain points, but also

$$\tilde{g}(u') \leq g(u) + \tilde{\mu}\|u' - u\| < \infty \quad \text{for all } u',$$

and by convexity $\tilde{g}$ is finite everywhere on $\mathcal{Y}$. Hence $\tilde{g}$ is locally Lipschitz and everywhere Clarke regular. Thus, we have

$$f(x') = \tilde{g}\big(F(x')\big) + \delta_C(x') \text{ for all } x' \text{ near } x.$$

The function $\tilde{f}(x') = \tilde{g}(F(x'))$ is Clarke regular, because composition of Clarke regular locally Lipschitz function with a smooth mapping preserves Clarke regularity ([**9**], Theorem 2.3.10). The Clarke regularity of C at $\bar{x}$ implies indicator function δ_C is Clarke regular at $\bar{x}$. Applying Rockafellar ([**24**], Corollary 2 of Theorem 2) we are able to conclude that the sum function is Clarke regular at $\bar{x}$ and consequently f is Clarke regular at $\bar{x}$. □

Corollary 4.2.8. *Let f be fully amenable at x. Then the various forms of subgradient sets $\partial f(x)$ agree.*

Proof. Since f is fully amenable at x, it is Clarke regular (Theorem 4.2.7) and Mosco epi-differential at x (Theorem 4.2.6). Then the result follows from Theorem 4.2.5. □

The next theorem, which links the proto-derivative of the subgradient mapping to the subgradient of the second order epi-derivative, enables us to calculate proto-derivatives of the subgradient mapping of fully amenable functions easily.

Theorem 4.2.9. *Let $\mathcal{X}$ be a finite dimensional space. If f is fully amenable at x, it is in fact twice epi-differentiable there relative to every $v \in \partial f(\bar{x})$. Moreover, the subgradient mapping ∂f is then proto-differentiable at $\bar{x}$ relative to every $v \in \partial f(x)$ with*

$$(\partial f)'_{\bar{x},v}(\xi) = \partial\big(\tfrac{1}{2} f''_{\bar{x},v}\big)(\xi) \text{ for all } \xi \text{ in } \mathcal{X}. \tag{4.2.4}$$

Proof. See Poliquin and Rockafellar [**21**] Theorem 2.9. and references therein.□

This formula stems from a classical result of C^2 functions. When f happens to be a C^2 function on $\mathbb{R}^n$, the subgradient mapping ∂f reduces then to the usual gradient mapping $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$, and proto-differentiation of ∂f comes down to ordinary differentiation of ∇f:

$$D(\nabla f)(\bar{x})(\xi) = \lim_{\substack{\xi' \to \xi \\ t \downarrow 0}} \frac{\nabla f(\bar{x} + t\xi') - \nabla f(\bar{x})}{t} = \nabla^2 f(\bar{x})\xi \tag{4.2.5}$$

Where $\nabla^2 f(\bar{x})$ is the matrix of second derivatives of f at $\bar{x}$.

The matrix $\nabla^2 f(\bar{x})$ can also be obtained as a limit of second-order difference quotients:

$$\lim_{\substack{\xi' \to \xi \\ t \downarrow 0}} \frac{f(\bar{x} + t\xi') - f(\bar{x}) - t\langle \xi', \nabla f(\bar{x})\rangle}{(\frac{1}{2})t^2} = \langle \xi, \nabla^2 f(x)\xi \rangle \tag{4.2.6}$$

If we denote the function $\xi \to \langle \xi, \nabla^2 f(\bar{x})\xi \rangle$ in (4.2.6) by $D^2 f(x)$, then the gradient mapping associated with it is twice the mapping $\xi \to \nabla^2 f(x)\xi$ in (4.2.5), so that symbolically we have $\nabla\big(D^2 f(\bar{x})\big) = 2D\big(\nabla f\big)(x)$ for all x.

It is natural to wonder whether one can extend the results of Theorem 4.2.9 to the infinite dimensional case. In this regard, more recently, R. Cominetti [**11**], proved that for a general class of functions (on a reflexive space) which consist of composition of C^2 mapping with a locally Lipschitz, convex outer function with extra regularity assumptions (this class of functions include the fully amenable functions) is twice Mosco epi-differentiable. However, he neither established the proto-differentiability of the subgradient mapping nor the relation (4.2.4). Our aim here is to establish these properties to some extent in a reflexive space which admits a C^2 function with its second Fréchet derivative is bigger than a multiple of the norm square.

We will need the following well known results for convex functions. Here is a similar version of Theorem 4.2.9 for the convex case:

Theorem 4.2.10. *Let $\mathcal{X}$ be a reflexive Banach space, and $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous, proper and convex. Let $\bar{v} \in \partial f(x)$. Then f is twice Mosco epi-differentiable at $\bar{x}$ relative to $\bar{v}$ if and only if ∂f is (PK)proto-differentiable at $\bar{x}$ relative to $\bar{v}$. In which case the formula (4.2.4) holds.*

This result was obtained in finite dimensions in Rockafellar [30] and generalized to infinite dimensions by Do [**12**]. □

The following lemma is a classic result of convex functions which relates the directional derivative of a convex function to its subgradient set.

Lemma 4.2.11. *Let f be a convex function and x be a point in dom f. Then v*

is a subgradient of f at x if and only if

$$f'(x;\xi) \geq \langle v,\xi\rangle \ \text{ for all } \xi$$

where $f'(x;\xi) := \lim_{t\downarrow 0} \dfrac{f(x+t\xi)-f(x)}{t}$ *is the usual (one-sided) directional derivative of f at x in the direction ξ. In fact, the lower semicontinuous closure of $f'(x;\cdot)$ (weak or strong since it is convex) is the support function of the closed convex set $\partial f(x)$.*

Proof. It is well known from convex analysis, that a functional v belongs to $\partial f(x)$ if and only if $f(y) \geq f(x) + \langle v, y-x\rangle$ for all y. Setting $y = x + t\xi$ the above inequality becomes $\dfrac{f(x+t\xi)-f(x)}{t} \geq \langle v,\xi\rangle$ for every ξ and $t > 0$. Since the different quotients decrease (cf. Convex Analysis [**23**], Theorem 23.1) to $f'(x;\xi)$ as $t\downarrow 0$, this inequality is equivalent to the one in the theorem.

Conversely, take any v in $\mathcal{X}^*$ with $\lim_{t\downarrow 0} \dfrac{f(x+t\xi)-f(x)}{t} \geq \langle v,\xi\rangle$ for all ξ. Since f is convex, we get $\inf_{t>0} \dfrac{f(x+t\xi)-f(x)}{t} \geq \langle v,\xi\rangle$. Setting $y = x+\xi$ and $t = 1$ in the above inequality we deduce $f(y) \geq f(x) + \langle v, y-x\rangle$ for all y, which completes the first part of the proof.

One can also infer that

$$\partial f(x) = \left\{ v \,\middle|\, \operatorname{cl} f'(x;\xi) \geq \langle v,\xi\rangle \ \text{ for all } \xi \right\}$$

Since $\operatorname{cl} f'(x;\xi) \leq f'(x;\xi)$, we only need to verify that any v satisfies $f'(x,\xi) \geq \langle v,\xi\rangle$ for all ξ belongs to the above set. This follows as

$$\langle v,\xi\rangle = \liminf_{\xi'\to\xi} \langle v,\xi'\rangle \leq \liminf_{\xi'\to\xi} f'(x;\xi') = \operatorname{cl} f'(x;\xi).$$

Now, we proceed to prove the well known subgradient characterization.
First consider,

$$\left(\operatorname{cl} f'(x;v)\right)^* = \sup_{\xi} \left\{ \langle v,\xi\rangle - \operatorname{cl} f'(x;\xi) \right\},$$

where $*$ represent the *conjugate* of a function, which is given by

$$\forall x^* \in \mathcal{X}^*, \ \ \varphi^*(x^*) := \sup_{x\in\mathcal{X}} \left\{ \langle x^*, x\rangle - \varphi(x) \right\}, \ \ \text{for any function } \varphi : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}.$$

If v in $\partial f(x)$ then $\langle v, \xi\rangle - \operatorname{cl} f'(x, \xi) \le 0$ for all ξ and we have that $\left(\operatorname{cl} f'(x; v)\right)^* = 0$. But, by the positively homogeneity of $f'(x; \cdot)$ we get $\left(\operatorname{cl} f'(x; v)\right)^* = \infty$ when $\langle v, \xi\rangle - \operatorname{cl} f'(x; \xi) > 0$ for some ξ. Thus $\left(\operatorname{cl} f'(x; v)\right)^* = \delta_{\partial f(x)}(v)$, where δ is the indicator function.

On the other hand, $\delta^*_{\partial f(x)}(y) = \sup_v \left\{ \langle v, y\rangle - \delta_{\partial f(x)}(v) \right\} = \sup_{v \in \partial f(x)} \langle v, y\rangle = \sigma_{\partial f(x)}(y)$. Connecting above results through the fact that

$$\operatorname{cl} f'(x; v) = \left(\operatorname{cl} f'(x; v)\right)^{**},$$

for convex functions we conclude $\operatorname{cl} f'(x; v) = \sigma_{\partial f(x)}(v)$. Where $**$ represent the *biconjugate* of a function, which is given by

$$\forall x \in \mathcal{X},\ \ \varphi^{**}(x) := \sup_{x^* \in \mathcal{X}^*} \left\{ \langle x^*, x\rangle - \varphi^*(x^*) \right\}, \ \text{ for any function } \varphi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}.$$

□

A prerequisite to the study of epi-derivatives of fully amenable functions is an understanding of such derivatives in case of the proper convex functions $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ which are piecewise linear-quadratic in the sense of Definition 4.1.2.

Theorem 4.2.12. *Let g be piecewise linear-quadratic convex function. At any point $u \in D = \operatorname{dom} g$, the function g is twice Mosco epi-differentiable. Its first Mosco epi-derivative function $g_u^{\prime(m)}$ is expressed simply by taking limits along rays:*

$$g_u^{\prime(m)}(w) = \lim_{t \downarrow 0} \frac{g(u + tw) - g(u)}{t}. \tag{4.2.7}$$

The function $g_u^{\prime(m)}$ is convex and piecewise linear with effective domain

$$\operatorname{dom} g_u^{\prime(m)} = T_D(u).$$

It is the support function of $\partial g(u)$, which is nonempty convex polyhedron and coincides with the set of all Mosco epi-gradients y of g at u. For any $y \in \partial g(u)$ the second Mosco epi-derivative function $g_{u,y}^{\prime\prime(m)}$ is likewise expressed simply by taking limits along rays:

$$g_{u,y}^{\prime\prime(m)}(w) = \lim_{t \downarrow 0} \frac{g(u + tw) - g(u) - t\langle y, w\rangle}{\frac{1}{2}t^2}. \tag{4.2.8}$$

The function $g''^{(m)}_{u,y}$ *is convex and piecewise linear-quadratic with effective domain*

$$\operatorname{dom} g''^{(m)}_{u,y} = \{\, w \in \mathcal{Y} \,|\, g'^{(m)}_u(w) = \langle y, w\rangle \,\} = N_{\partial g(u)}(y). \tag{4.2.9}$$

Thus for $y \in \partial g(u)$ *one has*

$$g''^{(m)}_{u,y}(w) = \begin{cases} \gamma_u(w) & \text{if } \langle y, w\rangle = g'^{(m)}_u(w), \\ +\infty & \text{if } \langle y, w\rangle < g'^{(m)}_u(w), \end{cases} \tag{4.2.10}$$

where for $w \in \operatorname{dom} g'^{(m)}_u$ *one defines*

$$\gamma_u(w) = \lim_{t \downarrow 0} \frac{g(u+tw) - g(u) - t g'^{(m)}_u(w)}{\frac{1}{2}t^2} < +\infty \qquad [\, = 0 \text{ if } g \text{ is actually piecewise-linear }]. \tag{4.2.11}$$

Proof. Consider a representation of g as in Definition 4.1.2 in terms of (polyhedral) sets $D_j (j \in J)$. Fix $u \in D$ and let $J_u = \{\, j \in J \,|\, u \in D_j \,\}$. For each $j \in J_u$ write:

$$g(u') = g(u) + \tfrac{1}{2} B_j(u'-u, u'-u) + \langle l_j^*, (u'-u)\rangle \text{ for } u' \in D_j$$

for some $l_j^* \in \mathcal{Y}^*$ and $B_j : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ a continuous symmetric bilinear function.

Claim 1. *For each* $j \in J_u$ *there exists* $\varepsilon_j > 0$ *such that*

$$[D_j - u] \cap \varepsilon_j B = T_{D_j}(u) \cap \varepsilon_j B.$$

Proof of Claim 1. First notice that the tangent cone $T_{D_j}(u)$ has a particularly simple form because D_j is polyhedral, namely

$$T_{D_j}(u) = \{\, w \in \mathcal{Y} \,|\, \exists \tau > 0 \text{ with } u + tw \in D_j \text{ for all } t \in (0, \tau) \,\}.$$

For notational convenience, let $D_j = D_1$ and write

$$D_1 = \cap_{i=1}^k E_i \text{ and } E_i = \{\, u \,|\, \langle l_i^*, u\rangle \le a_i \,\}.$$

Suppose $w \in T_{D_1}(u) \cap \varepsilon B$ with $\varepsilon > 0$. Then there exists $\tau > 0$ such that $u + tw \in D_1$ for all $t \in (0, \tau)$, which implies that $\langle l_i^*, u + tw\rangle \le a_i$ for all i and $t \in (0, \tau)$. Thus, we get

$$\langle l_i^*, u\rangle + t\langle l_i^*, w\rangle \le a_i \text{ for all } i \text{ and } t \in (0, \tau) \tag{4.2.12}$$

If $\langle l_i^*, u\rangle = a_i$ for some i then (4.2.12) implies that $\langle l_i^*, w\rangle \leq 0$. Therefore we have $\langle l_i^*, u+w\rangle \leq a_i$ which implies that $u+w \in E_i$. Otherwise, i.e., $\langle l_i^*, u\rangle < a_i$, take $\varepsilon := \min_{i\in\{1,\dots,k\}} \varepsilon_i$, where $\varepsilon_i = \dfrac{a_i - l_i^*(u)}{\|l_i^*\|}$ if $\langle l_i^*, u\rangle < a_i$. Then

$$\begin{aligned}\langle l_i^*, u+w\rangle &= \langle l_i^*, u\rangle + \langle l_i^*, w\rangle,\\ &\leq \langle l_i^*, u\rangle + \|l_i^*\|\|w\|,\\ &\leq \langle l_i^*, u\rangle + \|l_i^*\|\varepsilon_i,\\ &= a_i,\end{aligned}$$

and hence $u+w \in E_i$. Therefore we have $u+w \in E_i$ for all i, which implies $u+w \in D_1$.

On the other hand, suppose $w \in (D_1 - u)\cap \varepsilon B$. Then $w = y - u$ for some $y \in D_1$. Now consider,

$$\begin{aligned}\langle l_i^*, u+tw\rangle &= \langle l_i^*, u\rangle + t\langle l_i^*, w\rangle, \qquad (t>0)\\ &= \langle l_i^*, u\rangle + tl_i^*(y-u),\\ &\begin{cases}\leq a_i & \text{if } \langle l_i^*, u\rangle = a_i, \text{ (for all } t)\\ \leq a_i & \text{if } \langle l_i^*, u\rangle < a_i, \text{ (for small enough } t)\end{cases}\end{aligned}$$

which implies that $u+tw \in E_i$ for all i and small enough t. Thus we have $u+tw \in D_1$ and hence $w \in T_{D_1}(u)\cap \varepsilon B$. □

We also have

$$T_D(u) = \cup_{j\in J_u} T_{D_j}(u)$$

The inclusion $\supset$ is obvious. Now take $w \in T_D(u)$, i.e., there is $\tau > 0$ such that $u + tw \in D$ for all $t \in (0,\tau)$. Since $u \in D_j$ for all $j \in J_u$ there exists $\tau'(<\tau)$ such that $u+tw \in D_j$ for all $t \in (0,\tau')$ and for some $j \in J_u$. Thus $w \in T_{D_j}(u)$ and hence the result.

By Claim 1, for each $j \in J_u$ there exists $\varepsilon_j > 0$ such that

$$[D_j - u]\cap \varepsilon_j B = T_{D_j}(u)\cap \varepsilon_j B.$$

Let $\varepsilon := \min\{\varepsilon_j \mid j \in J_u\}$. Then for arbitrary $\rho > 0$ one has for all $w \in \rho B$ and $t \in (0, \varepsilon/\rho)$:

$$\frac{g(u+tw)-g(u)}{t} = \begin{cases}\langle l_j^*, w\rangle + \frac{1}{2}tB_j(w,w) & \text{if } w \in T_{D_j}(u),\ j\in J_u,\\ +\infty & \text{if } w \notin T_D(u),\end{cases} \tag{4.2.13}$$

Let $\varphi_{u,t}(w)$ and $\sigma_{u,t}(w)$ denote the left and right sides of (4.2.13), respectively; (4.2.13) asserts that $\varphi_{u,t}$ and $\sigma_{u,t}$ agree on the ball ρB when $t \in (0, \varepsilon/\rho)$.

Claim 2.

$$B_j(w,w) \geq 0 \text{ } when \text{ } w \in T_{D_j}(u),\ j \in J_u.$$

Proof of Claim 2. By the representation of g, B_j is convex on D_j for j in J_u. This is equivalent to the convexity of the restriction of B_j to each line segment in D_j. This is the same as the convexity of the function $\varphi(t) = B_j(u+tw, u+tw)$ on the open interval $\{t \mid u + tw \in D_j\}$ for each $u \in D_j$ and $w \in \mathcal{Y}$. For any $w \in \mathcal{Y}$ we have

$$\varphi''(t) = D^2 B_j(u+tw, u+tw)(w,w) = B_j(w,w).$$

Since $\varphi(t)$ is convex we have $B_j(w,w) \geq 0$. In particular, this is true when $w \in T_{D_j}(u)$, $j \in J_u$. □

We see from (4.2.13) that $\sigma_{u,t}(w)$ is a closed, convex, proper function and by Claim 2 it decreases as $t \downarrow 0$. Then by Theorem 3.1.4 (part (b)) the Mosco epi-limit exists and equals to $\sigma = \mathrm{cl}\left[\inf_{t>0} \sigma_{u,t}\right]$, i.e.,

$$\sigma(w) = \begin{cases} \langle l_j^*, w\rangle & \text{if } w \in T_{D_j}(u),\ j \in J_u, \\ +\infty & \text{if } w \notin T_D(u). \end{cases} \tag{4.2.14}$$

This means that $g_u'^{(m)}$ exists and equals to the usual directional derivative at u:

$$g_u'^{(m)}(w) = \sigma(w) \text{ for all } w. \tag{4.2.15}$$

By Lemma 4.2.11 we have

$$g_u'^{(m)}(w) = \sup_{y \in \partial g(u)} \langle y, w\rangle.$$

Therefore $\partial g(u)$ consists of the vectors y satisfying $g_u'^{(m)}(w) \geq \langle y, w\rangle$ for all w, which are by definition the Mosco epi-gradients of g at u. One can also characterized the $\partial g(u)$ by way of the piecewise linear-quadratic nature of g:
For $j \in J_u$, we write

$$g(u) = Q_j(u) + \delta_D(u), \text{ where } Q_j(u) = \tfrac{1}{2}B_j(u,u) + \langle l_j^*, u\rangle + c_j.$$

Thus,

$$\partial g(u) = \mathrm{co}\left\{ DQ_j(u) \,\middle|\, j \in J_u \right\} + N_D(u), \text{ where } DQ_j(u) := B_j(u,\cdot) + l_j^*.$$

The convex sets in the above sum are finitely generated so that they are polyhedral (as in **[23]**). Thus, $\partial g(u)$ is a convex polyhedron as being the sum of polyhedral convex sets.

To see the second-order results, we transform (4.2.13) into the assertion that

$$\frac{g(u+tw) - g(u) - \langle y, w\rangle}{\frac{1}{2}t^2} = \begin{cases} B_j(w,w) + \frac{2}{t}\langle l_j^* - y, w\rangle & \text{if } w \in T_{D_j}(u),\ j \in J_u, \\ +\infty & \text{if } w \notin T_D(u), \end{cases} \tag{4.2.16}$$

This being true for all $w \in \rho B$ when $t \in (0, \varepsilon/\rho)$. Let

$$\psi(w) = \begin{cases} B_j(w,w) & \text{if } w \in T_{D_j}(u),\ j \in J_u, \\ +\infty & \text{if } w \notin T_D(u). \end{cases} \tag{4.2.17}$$

and observe that

$$\psi(w) = \lim_{t \downarrow 0} \frac{g(u+tw) - g(u) - tg_u^{\prime(m)}(w)}{\frac{1}{2}t^2} \text{ for all } w \in \mathcal{Y} \tag{4.2.18}$$

by virtue of (4.2.13)-(4.2.15). Denote the difference quotients in (4.2.16) by $\varphi_{u,y,t}(w)$. Recalling that (4.2.14) gives $g_u^{\prime(m)}(w)$, we can write (4.2.16) as

$$\varphi_{u,y,t}(w) = \psi(w) + \tfrac{2}{t}[g_u^{\prime(m)}(w) - \langle y, w\rangle], \tag{4.2.19}$$

an equation that holds for all $w \in \rho B$ and $t \in (0, \varepsilon/\rho)$. Under the assumption that y is a Mosco epi-gradient of g at u, we have $g_u^{\prime(m)}(w) - \langle y, w\rangle \geq 0$ for all w. The closed, convex, proper functions $\varphi_{u,y,t}$ increase as $t \downarrow 0$ and hence by Theorem 3.1.4 (part (a)) the Mosco epi-limit exists and equals to $\psi_0(w) = \sup_{t>0} \varphi_{u,y,t}(w)$, i.e.,

$$\psi_0(w) = \begin{cases} \psi(w) & \text{if } g_u^{\prime(m)}(w) - \langle y, w\rangle = 0, \\ +\infty & \text{if } g_u^{\prime(m)}(w) - \langle y, w\rangle > 0\ . \end{cases}$$

Thus $g_{u,y}^{\prime\prime(m)}$ exists and equals ψ_0. □

Remark. This result was first established by Rockafellar in finite dimensions; see **[26]** Theorem 3.1.

Next we state our main theorem. It extends known results of second-order epi-derivatives of fully amenable functions, to reflexive spaces.

Theorem 4.2.13. *Let $\mathcal{X}$ and $\mathcal{Y}$ be reflexive Banach spaces, f be fully amenable as in definition 4.1.3., and $f = g \circ F$ is a local representation around $\bar{x} \in \operatorname{dom} f$ in the sense required in that definition. Assume also that the mapping $\xi \to D^2F(\bar{x})(\xi,\xi)$ from $\mathcal{X}$ to $\mathcal{Y}$ is weakly continuous (in particular when $\mathcal{X}$ is finite dimensional). Assume $F(\bar{x}) \in \operatorname{int}(\operatorname{dom} g)$, then f is twice Mosco epi-differentiable at $\bar{x}$ relative to all $v \in \partial f(\bar{x})$ and the Mosco second epi-derivative is given by*

$$f''^{(m)}_{\bar{x},v}(\xi) = \begin{cases} \max\limits_{y \in Y(\bar{x},v)} \left\{ g''^{(m)}_{F(\bar{x}),y}\big(DF(\bar{x})\xi\big) + \langle y, D^2F(\bar{x})(\xi,\xi)\rangle \right\} & \text{if } \xi \in \Xi(\bar{x},v), \\ +\infty & \text{if } \xi \notin \Xi(\bar{x},v), \end{cases} \tag{4.2.20}$$

where $Y(\bar{x},v) = \{\, y \in \partial g\big(F(\bar{x})\big) \mid DF(\bar{x})^* y = v \,\}$ is a nonempty bounded polyhedral convex set and $\Xi(\bar{x},v) = N_{\partial f(\bar{x})}(v) = \{\, \xi \mid f'^{(m)}_{\bar{x}}(\xi) = \langle v, \xi\rangle \,\}$.

Moreover, if $\mathcal{X}$ admits a $\mathcal{C}^2$ function φ with $\langle D^2\varphi(\bar{x})\xi,\xi\rangle \geq K\|\xi\|^2$ for all ξ and some $K > 0$, and the mapping $\xi \to \langle D^2\varphi(\bar{x})\xi,\xi\rangle$ is weakly lower semicontinuous, then ∂f is (PK) proto-differentiable at $\bar{x}$ relative to every $v \in \partial f(\bar{x})$ and the proto-derivative $(\partial f)'^{(pk)}_{\bar{x},v}$ satisfies

$$(\partial f)'^{(pk)}_{\bar{x},v}(\xi) = \partial\big(\tfrac{1}{2} f''^{(m)}_{\bar{x},v}\big)(\xi) \text{ for all } \xi, \tag{4.2.21}$$

and is given by

$$(\partial f)'^{(pk)}_{\bar{x},v}(\xi) = \operatorname{co}\left\{ DF(\bar{x})^* (\partial g)'^{(pk)}_{F(\bar{x}),y}\Big(DF(\bar{x})\xi\Big) + \partial\big(\langle y, D^2F(\bar{x})(\cdot,\cdot)\rangle\big)(\xi) \;;\; y \in \operatorname{ext} Y(\bar{x},v) \right\}, \tag{4.2.22}$$

where $\operatorname{ext} Y(\bar{x},v)$ is the set of extreme points of $Y(\bar{x},v)$ and co denotes the closed convex hull.

The following lemma will be needed.

Lemma 4.2.14. *Let $f : C \to \mathbb{R}$ be a $\mathcal{C}^2$ function and C be a convex set in a Banach space $\mathcal{X}$. If for each x in C, $D^2f(x)(\xi,\xi) \geq 0$ for all ξ in $\mathcal{X}$, then f is convex over C.*

Proof of Lemma. The convexity of f on C is equivalent to the convexity of the restriction of f to each line segment in C. This is the same as the convexity of the function $\psi(t) := f(x + t\xi)$ on the open interval $\{t \mid x + t\xi \in C\}$ for each x in C and ξ in $\mathcal{X}$. Then $\psi''(t) = D^2 f(x + t\xi)(\xi, \xi) \geq 0$ by our assumption, which implies ψ is convex for each x in C and ξ in $\mathcal{X}$. □

Proof of Theorem 4.2.13. The first part of the proof is due to Cominetti; see for instance [11], Theorem 4.4. In it he proved for a class of functions on a reflexive space, which consist of composition of $\mathcal{C}^2$ mapping F with a locally Lipschitz (relative to $\operatorname{dom} g$), convex outer function g with extra regularity assumptions, is twice Mosco epi-differentiable, and the formula (4.2.20) holds.
Cominetti used the following regularity condition:

$$0 \in \operatorname{core}\big[\operatorname{dom} g - F(\bar{x}) - DF(\bar{x})\mathcal{X}\big],$$

where for a set $C \subset \mathcal{X}$, $\operatorname{core} C := \big\{ u \in C \,\big|\, \forall\, u' \in \mathcal{X},\ \exists\, \varepsilon > 0,\ \forall\, \lambda \in [-\varepsilon, \varepsilon],\ u + \lambda u' \in C \big\}$.

This condition is obviously satisfied in our case, since always we have, $\operatorname{dom} g - F(x) \subset \operatorname{dom} g - F(x) - DF(x)\mathcal{X}$ and $F(\bar{x})$ is an interior point of $\operatorname{dom} g$. He also assumed that the outer function g is twice Mosco epi-differentiable and the epi-derivative coincides with the usual directional derivative. This is of course true for the fully amenable case since g is piecewise linear-quadratic (cf. Theorem 4.2.12).

The proof of the second result is based on a recent paper by R. Poliquin; [18], Proposition 4.2. In [18] he proved that a fully amenable function on $\mathbb{R}^n$ is lower-$\mathcal{C}^2$ (i.e. locally the sum of the function and a nonnegative multiple of a norm square is convex) at $\bar{x}$ where $F(\bar{x}) \in \operatorname{int}(\operatorname{dom} g)$, ∂f is proto-differentiable at $\bar{x}$ relative to $v \in \partial f(\bar{x})$ and the relation (4.2.21) is satisfied as well. We proceed to establish the relation (4.2.21) extending the above results to reflexive Banach spaces.

Take $\bar{x} \in \mathcal{X}$ with $F(\bar{x}) \in \operatorname{int}(\operatorname{dom} g)$. Then there exists $r > 0$ such that $B\big(F(\bar{x}), r\big) \subset \operatorname{dom} g$. Now consider $U^* = \displaystyle\bigcup_{y \in B(F(\bar{x}), r)} \partial g(y)$.

Claim 1. *U^* is a norm bounded set in $\mathcal{Y}^*$.*

Proof of Claim 1. Take any $u^* \in U^*$. Then $u^* \in \partial g(\bar{y})$ for some $\bar{y}$ in $B(F(\bar{x}), r)$. Since $\bar{y}$ belongs to interior of dom g and g is piecewise linear-quadratic convex, it is locally Lipschitz at $\bar{y}$. Then there must exist $M > 0$ and a neighbourhood V of $\bar{y}$ such that

$$|g(y) - g(z)| \leq M\|y - z\| \text{ whenever } y, z \in V.$$

Since $u^* \in \partial g(\bar{y})$, then for all y in V we have

$$\langle u^*, y - \bar{y} \rangle \leq g(y) - g(\bar{y}) \leq M\|y - \bar{y}\|$$

which implies that $\|u^*\| \leq M$ and concludes the proof of Claim 1. □

Claim 2. *There exists $\lambda > 0$ and $\rho > 0$ such that the function $f(x) + \rho\varphi(x) + \delta_{\bar{x}+\lambda B}(x)$ is convex, where φ is C^2 function such that $\langle D^2\varphi(x)\xi, \xi\rangle \geq K\|\xi\|^2$ for all ξ and some K positive and the mapping $\xi \to \langle D^2\varphi(x)\xi, \xi\rangle$ is weakly lower semicontinuous.*

Proof of Claim 2. Let $W := F^{-1}\Big(B\big(F(\bar{x}), r\big)\Big)$. We wish to prove that for some $\rho > 0$, $f(x) + \rho\varphi(x) + \delta_{\bar{x}+\lambda B}(x)$ is convex, where $\lambda > 0$ such that $\bar{x} + \lambda B \subset W$. Since g is convex and continuous on dom g, for any x in W, we have (cf. Convex Analysis, [**23**])

$$\begin{aligned} f(x) &= g\big(F(x)\big) = g^{**}\big(F(x)\big) \\ &= \sup_{u^* \in U^*} \big\{ \langle u^*, F(x)\rangle - g^*\big(u^*\big) \big\}, \qquad (U^* \text{ is strongly bounded}) \end{aligned}$$

where g^* and g^{**} denotes the conjugate and biconjugate of g respectively. Then for any $\rho > 0$, we have

$$f(x) + \rho\varphi(x) = \sup_{u^* \in U^*} \Big\{ \langle u^*, F(x)\rangle - g^*\big(u^*\big) + \rho\varphi(x) \Big\}.$$

Now consider for a fixed u^*, and any x in W,

$$h_{u^*}(x) := \langle u^*, F(x)\rangle - g^*(u^*) + \rho\varphi(x),$$

and differentiating twice with respect to x we get

$$D^2 h_{u^*}(x)(\xi,\xi) = \langle u^*, D^2F(x)(\xi,\xi)\rangle + \rho D^2\varphi(x)(\xi,\xi), \tag{4.2.23}$$

where

$$\begin{aligned} |\langle u^*, D^2F(x)(\xi,\xi)\rangle| &\leq \|u^*\| \ \|D^2F(x)(\xi,\xi)\| \\ &\leq M\|D^2F(x)\| \ \|\xi\|^2 \qquad (\text{ by Claim 1 }) \end{aligned}$$

Since F is $\mathcal{C}^2$, there exists $L > 0$ such that $\|D^2F(x)\| \leq L$ for all x in W. Therefore, $|\langle u^*, D^2F(x)(\xi,\xi)\rangle| \leq K'\|\xi\|^2$ for all x in W, where $K' = ML > 0$. By our assumption and (4.2.23) we have, $D^2h_{u^*}(x)(\xi,\xi) \geq -K'\|\xi\|^2 + \rho K\|\xi\|^2$. Taking $\rho = K'/K$ we have that $D^2h_{u^*}(x)(\xi,\xi) \geq 0$ for all x in W and ξ in X. Then by Lemma 4.2.14, $h_{u^*}(x)$ is convex on $B(\bar{x},\lambda)$ and the fact that the supremum of convex functions is convex give us $f(x) + \rho\varphi(x) + \delta_{x+\lambda B}(x)$ is convex, which completes the proof of claim 2. □

Now let $h(x) := f(x) + \rho\varphi(x) + \delta_{\bar{x}+\lambda B}(x)$. We know from Claim 2, that h is convex and we deduce

$$\partial h(x) = \partial f(x) + \rho D\varphi(x) \ \text{ on } \ \text{int}(\bar{x} + \lambda B).$$

We denote by x^* the Fréchet differential of φ at x. Here ∂f can be taken in any sense of the characterization of Theorem 4.2.5 since f is fully amenable at x.

Now for any v in $\partial f(\bar{x})$, from the first part of the proof, f is twice Mosco epi-differentiable at $\bar{x}$ and by the Proposition 3.2.5 we have

$$h''^{(m)}_{\bar{x},v+\rho\bar{x}^*}(\xi) = f''^{(m)}_{\bar{x},v}(\xi) + \rho D^2\varphi(\bar{x})(\xi,\xi). \tag{4.2.24}$$

Since h is convex, lower semicontinuous, and twice Mosco epi-differentiable at x, then by Theorem 4.2.10, ∂h is proto-differentiable at x relative to $v + \rho x^*$ and for all ξ we have

$$(\partial h)'^{(pk)}_{\bar{x},v+\rho\bar{x}^*}(\xi) = \partial\Big(\tfrac{1}{2}h''^{(m)}_{\bar{x},v+\rho\bar{x}^*}\Big)(\xi). \tag{4.2.25}$$

We now show that (4.2.24) and (4.2.25) imply that ∂f is proto-differentiable at $\bar{x}$ relative to v and that

$$(\partial h)'^{(pk)}_{\bar{x},v+\rho\bar{x}^*}(\xi) = (\partial f)'^{(pk)}_{\bar{x},v}(\xi) + \rho D^2\varphi(\bar{x})\xi.$$

To see this, let

$$(\xi, u) \in s\text{-}\limsup_{t\downarrow 0} \frac{\operatorname{gph}\partial f - (\bar{x}, v)}{t},$$

i.e., there exists $t_n \downarrow 0$ and $v_n \in \partial f(x_n)$ with $(1/t_n)(v_n - v) \xrightarrow{s} u$ and $(1/t_n)(x_n - \bar{x}) \xrightarrow{s} \xi$. But,

$$\frac{x_n^* - \bar{x}^*}{t_n} = \frac{D\varphi(x_n) - D\varphi(\bar{x})}{t_n} \xrightarrow{s} D^2\varphi(\bar{x})\xi := \eta \text{ (say) }.$$

Thus $(1/t_n)[(v_n + \rho x_n^*) - (v + \rho\bar{x}^*)] \xrightarrow{s} u + \rho\eta$ and eventually, $(v_n + \rho x_n^*) \in \partial h(x_n)$. This implies

$$\begin{aligned}(\xi, u + \rho\eta) &\in s\text{-}\limsup_{t\downarrow 0} \frac{\operatorname{gph}\partial h - (\bar{x}, v + \rho\bar{x}^*)}{t} \\ &= s\text{-}\liminf_{t\downarrow 0} \frac{\operatorname{gph}\partial h - (\bar{x}, v + \rho\bar{x}^*)}{t}\end{aligned}$$

i.e., $(u + \rho\eta) \in (\partial h)'^{(pk)}_{\bar{x},v+\rho\bar{x}^*}(\xi)$.

Hence, for all $\mu_n \downarrow 0$ there exists $w_n' \in \partial h(x_n')$, such that

$$(1/\mu_n)(x_n' - \bar{x}) \xrightarrow{s} \xi \text{ and } (1/\mu_n)[w_n' - (v + \rho\bar{x}^*)] \xrightarrow{s} u + \rho\eta.$$

Eventually, $w_n' = v_n' + \rho x_n'^*$, where $v_n' \in \partial f(x_n')$. Since $\dfrac{x_n'^* - \bar{x}^*}{\mu_n} \xrightarrow{s} D^2\varphi(\bar{x})\xi = \eta$ and hence

$$(1/\mu_n)(x_n' - \bar{x}) \xrightarrow{s} \xi \text{ and } (1/\mu_n)(v_n' - v) \xrightarrow{s} u,$$

i.e.,

$$(\xi, u) \in s\text{-}\liminf_{t\downarrow 0} \frac{\operatorname{gph}\partial f - (\bar{x}, v)}{t}.$$

We conclude that ∂f is proto-differentiable at $\bar{x}$, relative to v.

To establish the relation (4.2.21), notice that

$$\begin{aligned}(\partial f)'^{(pk)}_{\bar{x},v}(\xi) = \left[(\partial h)'^{(pk)}_{\bar{x},v+\rho\bar{x}^*}(\xi) - \rho D^2\varphi(\bar{x})\xi\right] &= \partial\left(\tfrac{1}{2}h''^{(m)}_{\bar{x},v+\rho\bar{x}^*}\right)(\xi) - \rho D^2\varphi(\bar{x})\xi \\ &= \partial\left(\tfrac{1}{2}f''^{(m)}_{\bar{x},v}\right)(\xi).\end{aligned}$$

Finally, to obtain the proto-derivative formula (4.2.22) all we need to do, according to (4.2.21), is evaluate the subgradient of the second order Mosco epi-derivative. First, notice that the formula (4.2.20) can be written as, for any $\bar{y} \in Y(\bar{x}, v)$,

$$f''^{(m)}_{\bar{x},v}(\xi) = g''^{(m)}_{F(\bar{x}),\bar{y}}\big(DF(\bar{x})\xi\big) + \max_{y \in \operatorname{ext} Y(\bar{x},v)} \Big\{\langle y, D^2F(\bar{x})(\xi,\xi)\rangle\Big\} \qquad (4.2.20')$$

where $\operatorname{ext} Y(\bar{x}, v)$ is the set of extreme points of $Y(\bar{x}, v)$ in (4.2.20).

These formulas are of course equal because for any $y \in \partial g(u)$ one actually has (by Theorem 4.2.12)

$$g''^{(m)}_{u,y}(w) = \begin{cases} \lim\limits_{t \downarrow 0} \Big[g(u + tw) - g(u) - tg'_u(w)\Big] / \frac{1}{2}t^2 & \text{if } \langle y, w\rangle = g'_u(w), \\ +\infty & \text{if } \langle y, w\rangle < g'_u(w), \end{cases}$$

where $u = F(\bar{x})$, $w = DF(\bar{x})\xi$, and for any $y \in Y(\bar{x}, v)$

$$\langle y, w\rangle = \langle y, DF(\bar{x})\xi\rangle = \langle DF(\bar{x})^* y, \xi\rangle = \langle v, \xi\rangle.$$

Notice also that, the set $Y(\bar{x}, v)$ is (nonempty) polyhedron by its definition, inasmuch as the set $\partial g\big(F(\bar{x})\big) = \partial g(u)$ is a polyhedron (Theorem 4.2.12). It is norm bounded as from Claim 1. Thus, the maximum in the second order formula (4.2.20) can be taken over $\operatorname{ext} Y(\bar{x}, v)$ (which is a finite set).

According to formula (4.2.20′) and the subgradient of maximum of finitely many quadratic functions (see Clarke [9]) we only need to show that for any $y \in Y(\bar{x}, v)$ we have

$$\partial\big(g''^{(m)}_{F(\bar{x}),y} \circ DF(\bar{x})\big)(\xi) = DF(\bar{x})^* \partial g''^{(m)}_{F(\bar{x}),y}\big(DF(\bar{x})\xi\big). \qquad (4.2.26)$$

To show (4.2.26), first notice that

$$\begin{aligned} \big(g''^{(m)}_{F(\bar{x}),y} \circ DF(\bar{x})\big)'_{\xi}(\xi') &= \lim_{t \downarrow 0} \frac{\Big(g''^{(m)}_{F(\bar{x}),y} \circ DF(\bar{x})\Big)(\xi + t\xi') - \Big(g''^{(m)}_{F(\bar{x}),y} \circ DF(\bar{x})\Big)(\xi)}{t}, \\ &= \lim_{t \downarrow 0} \frac{g''^{(m)}_{F(\bar{x}),y}\big(DF(\bar{x})\xi + tDF(\bar{x})\xi'\big) - g''^{(m)}_{F(\bar{x}),y}\big(DF(\bar{x})\xi\big)}{t}, \\ &= \big(g''^{(m)}_{F(\bar{x}),y}\big)'_{DF(\bar{x})\xi}\big(DF(\bar{x})\xi'\big). \end{aligned} \qquad (4.2.27)$$

Since the sets in (4.2.26) are closed convex, to show that they are equal, it suffices to show that they have the same support functions.

To see this, we start with the support function of $DF(\bar{x})^*\partial g''^{(m)}_{F(\bar{x}),y}\big(DF(\bar{x})\xi\big)$:

$$\begin{aligned}
\sup_{w\in\partial g''^{(m)}_{F(\bar{x}),y}\left(DF(\bar{x})\xi\right)} \langle DF(\bar{x})^* w, \xi'\rangle &= \sup_{w\in\partial g''^{(m)}_{F(\bar{x}),y}\left(DF(\bar{x})\xi\right)} \langle w, DF(\bar{x})\xi'\rangle, \\
&= \mathrm{cl}\left[\big(g''^{(m)}_{F(\bar{x}),y}\big)'_{DF(\bar{x})\xi}\big(DF(\bar{x})\xi'\big)\right] \;\; \text{(Lemma 4.2.11)} \\
&= \big(g''^{(m)}_{F(\bar{x}),y}\big)'_{DF(\bar{x})\xi}\big(DF(\bar{x})\xi'\big) \;\; (\, g''^{(m)}_{F(\bar{x}),y} \;\; lsc \;) \\
&= \big(g''^{(m)}_{F(\bar{x}),y}\circ DF(\bar{x})\big)'_{\xi}(\xi') \qquad \text{(by 4.2.27)} \\
&= \sup_{\eta\in\partial\left(g''^{(m)}_{F(\bar{x}),y}\circ DF(\bar{x})\right)(\xi)} \langle \eta, \xi'\rangle
\end{aligned}$$

thus we have the equality in (4.2.26). Therefore, the proto-derivative formula for the subgradient mapping is

$$\big(\partial f\big)'^{(pk)}_{\bar{x},v}(\xi) = \mathrm{co}\,\Big\{\, DF(\bar{x})^*(\partial g)'^{(pk)}_{F(\bar{x}),y}\Big(DF(\bar{x})\xi\Big) + \partial\Big(\langle y, D^2F(\bar{x})(\cdot,\cdot)\rangle\Big)(\xi) \; ; \; y\in \mathrm{ext}\, Y(\bar{x},v)\,\Big\}.$$

□

CHAPTER 5

POINTWISE MAXIMA

5.1. Subdifferential Properties of Max-Functions

Now we turn to the study of the subdifferential properties of a useful class of functions in optimization; the maximum $f = \max_i f_i$ of finitely many C^2 functions. This class is of special interest because it is the simplest class of nonsmooth functions and also of its wide applicability in optimization. For instance, approximation and penalization procedures for mathematical programming problems can be expressed as the pointwise maximum of certain other functions which are themselves smooth (Example 4.1.8).

Even though, the max-functions are the simplest kind of nonsmooth functions, direct derivation of proto-derivative formulas of its subgradient mappings becomes highly complex (for instance, see [6] Proposition 12). Fortunately, we are able to take advantage of this class being a subclass of fully amenable functions (Example 4.1.6) and hence the derivation of subdifferential calculus becomes much easier. Here is a more detailed description.

Theorem 5.1.1. *Let $f(x) = \max\{f_1(x), \ldots, f_k(x)\}$ defined on a reflexive Banach space $\mathcal{X}$, where each function $f_i : \mathcal{X} \to \mathbb{R}$ is C^2 (Fréchet). Then f is everywhere fully amenable as in the Definition 4.1.3, and $f = g \circ F$ is a local representation around x in $\mathcal{X}$ in the sense required in that defintion. Assume also that the mapping $\xi \to \langle D^2F(x)\xi, \xi\rangle$ from $\mathcal{X}$ to $\mathbb{R}^k$ is weakly continuous. Then*

$$\partial f(x) = \mathrm{co}\left\{ Df_i(x) \,\middle|\, i \in I(x) \right\}, \qquad f_x'^{(m)}(\xi) = \max_{i \in I(x)} \langle Df_i(x), \xi\rangle \tag{5.1.1}$$

where $I(x)$ denotes the set of all indices i such that $f_i(x) = f(x)$ and co denotes the closed convex hull.

For any $v \in \partial f(x)$, the second-order Mosco epi-derivative of f at x for v exists *and given by*

$$f''^{(m)}_{x,v}(\xi) = \begin{cases} \max_{y \in Y(x,v)} \sum_{i=1}^{k} y_i \langle D^2 f_i(x)\xi, \xi\rangle & \text{if } \xi \in \Xi(x,v), \\ +\infty & \text{if } \xi \notin \Xi(x,v), \end{cases} \tag{5.1.2}$$

where $Y(x,v)$ is a polyhedral set and $\Xi(x,v)$ is a polyhedral cone, namely

$$\begin{aligned} Y(x,v) &= \Big\{ y \,\Big|\, y_i \geq 0 \text{ if } i \in I(x),\ y_i = 0 \text{ if } i \notin I(x), \\ &\qquad\qquad \textstyle\sum_{i=1}^{k} y_i = 1,\ \sum_{i=1}^{k} y_i Df_i(x) = v \Big\}, \\ \Xi(x,v) &= N_{\partial f(x)}(v) = \Big\{ \xi \,\Big|\, \langle Df_i(x) - v, \xi\rangle \leq 0 \text{ for all } i \in I(x) \Big\}. \end{aligned} \tag{5.1.3}$$

Furthermore, if $\mathcal{X}$ admits a $\mathcal{C}^2$ function φ with $\langle D^2\varphi(x)\xi, \xi\rangle \geq K\|\xi\|^2$ for all ξ and some $K > 0$, and the mapping $\xi \to \langle D^2\varphi(x)\xi, \xi\rangle$ is weakly lower semicontinuous , then ∂f is (PK) proto-differentiable at x for v with

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \Big\{ \sum_{i=1}^{k} y_i D^2 f_i(x)\xi \,\Big|\, y \in Y_{\max}(x,v,\xi) \Big\} + N_{\Xi(x,v)}(\xi) & \text{if } \xi \in \Xi(x,v), \\ \emptyset & \text{if } \xi \notin \Xi(x,v), \end{cases} \tag{5.1.4}$$

where $Y_{\max}(x,v,\xi)$ is the closed face of $Y(x,v)$ consisting of the multiplier vectors y that achieve the maximum in (5.1.2).

Proof. To see that f is everywhere fully amenable, simply observe that $f(x) = g\big(F(x)\big)$ for

$$F(x) = \big(f_1(x), \ldots, f_k(x)\big), \qquad g(w_1, \ldots, w_k) = \max\{w_1, \ldots, w_k\}, \tag{5.1.5}$$

and g is piecewise linear (polyhedral). Condition (4.1.1) is automatically satisfied since $\operatorname{dom} g$ is all of $\mathbb{R}^k$.

Applying Theorem 4.2.6 to the case of F and g in (5.1.5), we have

$$\partial f(x) = DF(x)^* \partial g\big(F(x)\big), \qquad f'^{(m)}_{x}(\xi) = g'^{(m)}_{F(x)}\big(DF(x)\xi\big),$$

where

$$\begin{aligned} DF(x) &= \big(Df_1(x), \ldots, Df_k(x)\big), \\ \partial g(w) &= \Big\{ y \,\Big|\, y_i \geq 0 \text{ if } w_i = g(w),\ y_i = 0 \text{ if } w_i < g(w),\ \textstyle\sum_{i=1}^{k} y_i = 1 \Big\}. \end{aligned}$$

Therefore

$$\partial f(x) = \left\{ y_1 Df(x) + \ldots + y_m Df_k(x) \,\middle|\, y_i \geq 0,\ \textstyle\sum_{i=1}^{k} y_i = 1,\ y_i = 0 \text{ for } i \notin I(x) \right\}.$$
$$= \text{co}\left\{ Df_i(x) \,\middle|\, i \in I(x) \right\}.$$

Since g is piecewise-linear, its Mosco epi-derivative $g_u'^{(m)}(w)$ agrees with the usual directional derivative (cf. Theorem 4.2.12) and hence

$$g_u'^{(m)}(w) = \lim_{t \downarrow 0} \frac{g(u + tw) - g(u)}{t}.$$
$$= \lim_{t \downarrow 0} \max_i \frac{\{u_i + tw_i\} - g(u)}{t}.$$

(Since for t small enough, any index i not satisfying $u_i = g(u)$ can be ignored in the maximum.)

$$= \lim_{t \downarrow 0} \max \left\{ \frac{(u_i + tw_i - u_i)}{t} \,\middle|\, i \text{ such that } u_i = g(u) \right\}.$$
$$= \max \left\{ w_i \,\middle|\, i \text{ such that } u_i = g(u) \right\}.$$

Thus, $f_x'^{(m)}(\xi) = g_{F(x)}'^{(m)}(DF(x)\xi)$,

$$= \max_{i \in I(x)} \langle Df_i(x), \xi \rangle.$$

Also note that $f_x'^{(m)}(\xi)$ is also equals to its directional derivative.

Indeed,

$$f'(x;\xi) = \lim_{t \downarrow 0} \frac{f(x + t\xi) - f(x)}{t}.$$
$$= \lim_{t \downarrow 0} \max_i \frac{f_i(x + t\xi) - f(x)}{t}.$$
$$= \lim_{t \downarrow 0} \max_{i \in I(x)} \frac{f_i(x + t\xi) - f_i(x)}{t}.$$
$$= \max_{i \in I(x)} \langle Df_i(x), \xi \rangle.$$
$$= f_x'^{(m)}(\xi).$$

The second-order epi-differentiability of f follows from the Theorem 4.2.13, since g is piecewise linear the term $g_{F(x),y}''^{(m)}(DF(x)\xi)$ in (4.2.20) vanishes and also the multiplier set $Y(x, v)$ has particularly simple form:

$$
\begin{aligned}
Y(x,v) &= \{\, y \in \partial g(F(x)) \mid DF(x)^* y = v \,\}, \\
&= \Big\{\, y \,\Big|\, y_i \geq 0 \text{ if } i \in I(x),\ y_i = 0 \text{ if } i \notin I(x), \\
&\qquad\qquad \textstyle\sum_{i=1}^k y_i = 1,\ \sum_{i=1}^k y_i Df_i(x) = v \,\Big\}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
\Xi(x,v) &= \{\, \xi \mid f_x'^{(m)}(\xi) \leq \langle v, \xi\rangle \,\}, \\
&= \{\, \xi \mid \max_{i \in I(x)} \langle Df_i(x), \xi\rangle \leq \langle v, \xi\rangle \,\}, \ \text{(by 5.1.1)} \\
&= \{\, \xi \mid \langle Df_i(x) - v, \xi\rangle \leq 0 \text{ for all } i \in I(x) \,\}.
\end{aligned}
$$

Finally, we obtain the proto-derivative formula using (5.1.2),

$$
f_{x,v}''^{(m)}(\xi) = \max_{y \in Y(x,v)} \sum_{i=1}^{k} y_i \langle D^2 f_i(x)\xi, \xi\rangle + \delta_{\Xi(x,v)}(\xi),
$$

and the relation $(\partial f)_{x,v}'^{(pk)}(\xi) = \partial\big(\frac{1}{2} f_{x,v}''^{(m)}\big)(\xi)$.

Thus,

$$
\begin{aligned}
(\partial f)'_{x,v}(\xi) &= \Big\{ \textstyle\sum_{i=1}^k y_i D^2 f_i(x)\xi \,\Big|\, y \in Y_{\max}(x,v,\xi) \Big\} + \partial\big(\delta_{\Xi(x,v)}\big)(\xi), \\
&= \begin{cases} \Big\{ \sum_{i=1}^k y_i D^2 f_i(x)\xi \,\Big|\, y \in Y_{\max}(x,v,\xi) \Big\} + N_{\Xi(x,v)}(\xi) & \text{if } \xi \in \Xi(x,v), \\ \emptyset & \text{if } \xi \notin \Xi(x,v). \end{cases}
\end{aligned}
$$

□

Remark. These results were first established in finite dimensions by Poliquin and Rockafellar [**22**]. The assumptions on Banach space $\mathcal{X}$ are redundant in finite dimensions.

5.2. Comparison of Proto-derivative Formulas

Several authors have worked on the proto-differentiation of the subgradient mapping ∂f of the max-function f. Among them Poliquin and Rockafellar [**22**], Auslender and Cominetti [**6**], and Penot [**17**] work is significant. They used two different approaches to calculate proto-derivatives. Poliquin and Rockafellar utilized their favourable amenable setting, on the other hand, Auslender and Cominetti used the direct definition to calculate proto-derivative. All these work were confined to finite dimensions.

As a result of the max-function is fully amenable, the proto-derivative formula given in [**22**], that we extend in Theorem 5.1.1., is much simpler than others and does not need extra assumptions in finite dimensions. However, the direct approach of Auslender and Cominetti gave only a formula for the outer graphical limit of the proto-derivative of ∂f and the usage of formula is limited to the special situations because of the complexity of the formula. Later, Penot [**17**] extended Auslender and Cominetti formula to infinite dimensions and established the proto-differentiability under sharper restrictions.

The major difficulty still remains is to see whether these formulas agree when the proto-derivative exists. Even in the finite dimensions, the reconciliation of these formulas is not a simple task.

First of all we need to introduce the Auslender and Cominetti formula:

Recall that $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is proto-differentiable at x for $v \in \partial f(x)$ (same as $\operatorname{gph} \partial f$ is derivable at (x, v)) if

$$\limsup_{t \downarrow 0} \left[\operatorname{gph} \partial f - (x, v)\right]/t = \liminf_{t \downarrow 0} \left[\operatorname{gph} \partial f - (x, v)\right]/t,$$

and note that

$$(\xi, u) \in \limsup_{t \downarrow 0} \left[\operatorname{gph} \partial f - (x, v)\right]/t \quad \Longleftrightarrow \quad u \in \limsup_{\substack{\xi' \to \xi \\ t \downarrow 0}} \left[\partial f(x + t\xi') - v\right]/t.$$

The following formula is given by Auslender and Cominetti for the outer graphical limit of the difference quotients, denote as $\partial^2 f_{x,v}(\xi)$: For any $v \in \partial f(x)$,

$$\partial^2 f_{x,v}(\xi) = \limsup_{\substack{\xi' \to \xi \\ t \downarrow 0}} \left[\partial f(x + t\xi') - v\right]/t,$$

$$= \bigcup_{I^* \in S(x,v,\xi)} \bigcup_{y \in Y(I^*,v)} \Big[\textstyle\sum_{i=1}^k y_i \nabla^2 f_i(x)\xi + E(I^*, y) \Big], \qquad (5.2.1)$$

where

$$Y(I^*, v) := \big\{ y \in Y(x, v) \,\big|\, y_i = 0 \text{ for } i \notin I^* \big\},$$

$$S(x, v, \xi) := \big\{ I^* \subset I(x) \,\big|\, Y(I^*, v) \neq \emptyset \text{ and } \exists t_n \downarrow 0,\ \xi_n \to \xi,$$

$$\text{with } I^* = I(x + t_n \xi_n) \text{ for all } n \big\},$$

$$E(I^*, y) := \big\{ \textstyle\sum_{i=1}^k \sigma_i \nabla f_i(x) \,\big|\, \sum_{i=1}^k \sigma_i = 0,\ \sigma_i = 0 \text{ if } i \notin I^*,\ \sigma_i \geq 0 \text{ if } y_i = 0 \big\}.$$

Notice that I^* of $S(x, v, \xi)$ is the set of "infinitly repeated index sets" at x in a given direction ξ with $I^* = I(x + t_n \xi_n)$. One of the main reasons of the complexity of the formula (5.2.1) is to identify which index sets I^* belong to the collection $S(x, v, \xi)$.

It was observed by Auslender and Cominetti that each set I^* of $S(x, v, \xi)$ is included in $I(x, \xi) = \big\{ i \in I(x) \,\big|\, \nabla f_i(x)\xi = f'(x; \xi) \big\}$, the set of indices achieving the maximum in (5.1.1).

Indeed, if $I^* \in S(x, v, \xi)$, i.e., $Y(I^*, v) \neq \emptyset$ and $\exists t_n \downarrow 0, \xi_n \to \xi$ with $I^* = I(x + t_n \xi_n)$ for all n, then for $i \in I^*$

$$\begin{aligned} f'(x; \xi) &= \lim_{n \to \infty} \frac{f(x + t_n \xi) - f(x)}{t_n}, \\ &= \lim_{n \to \infty} \Big[\frac{f(x + t_n \xi) - f(x + t_n \xi_n)}{t_n} + \frac{f(x + t_n \xi_n) - f(x)}{t_n} \Big], \\ &= 0 + \lim_{n \to \infty} \frac{f(x + t_n \xi_n) - f(x)}{t_n}, \quad (f \text{ locally Lipschitz}) \\ &= \lim_{n \to \infty} \frac{f_i(x + t_n \xi_n) - f_i(x)}{t_n}, \\ &= \nabla f_i(x)\xi. \end{aligned}$$

which implies $I^* \subset I(x, \xi)$.

With this observation we shall show that the effective domain of the formula (5.2.1) is always contained in (5.1.4).

Proposition 5.2.1.

$$Y\big(I(x, \xi), v\big) \neq \emptyset \iff \langle v, \xi \rangle = f'(x; \xi).$$

Proof. First assume that $y \in Y\big(I(x,\xi), v\big)$. Then

$$v = \sum_{i=1}^k y_i \nabla f_i(x),\ \sum_{i=1}^k y_i = 1 \text{ with } y_i = 0 \text{ for } i \notin I(x,\xi)$$

and hence

$$\begin{aligned}\langle v, \xi\rangle &= \sum_{i=1}^k y_i \langle \nabla f_i(x), \xi\rangle,\\ &= \sum_{i\in I(x,\xi)} y_i \langle \nabla f_i(x), \xi\rangle,\\ &= \sum_{i\in I(x,\xi)} y_i f'(x;\xi),\\ &= f'(x;\xi).\end{aligned}$$

Conversely, assume that $f'(x;\xi) = \langle v, \xi\rangle$. We wish to prove the set $Y\big(I(x,\xi), v\big)$ is nonempty, i.e., there exists y of $Y(x,v)$ with $y_i = 0$ for $i \notin I(x,\xi)$.
Since $v = \sum_{i\in I(x)} y_i \nabla f_i(x)$ with $\sum_{i\in I(x)} y_i = 1$, we have

$$\begin{aligned}f'(x;\xi) &= \sum_{i\in I(x)} y_i \langle \nabla f_i(x), \xi\rangle,\\ &= \sum_{i\in I(x,\xi)} y_i \langle \nabla f_i(x), \xi\rangle + \sum_{i\in I(x)\setminus I(x,\xi)} y_i \langle \nabla f_i(x), \xi\rangle,\\ &< \sum_{i\in I(x,\xi)} y_i f'(x;\xi) + \sum_{i\in I(x)\setminus I(x,\xi)} y_i f'(x;\xi),\\ &= f'(x;\xi).\end{aligned}$$

which is impossible. Thus, $y_i = 0$ for all i in $I(x) \setminus I(x,\xi)$, and hence $y_i = 0$ for $i \notin I(x;\xi)$. □

It follows from the above discussion and the proposition, that whenever $\partial^2 f_{x,v}(\xi)$ is nonempty, there exists $I^* \in S(x,v,\xi)$, hence $I^* \subset I(x,\xi)$, such that $Y(I^*, v)$ is non void. From this we have $Y\big(I(x,\xi), v\big) \neq \emptyset$ and by Proposition 5.2.1. $\langle v, \xi\rangle = f'(x;\xi)$, and hence $(\partial f)'_{x,v}(\xi) \neq \emptyset$.

The first step of comparison of formulas (5.1.4) and (5.2.1) was carried out by Poliquin and Rockafellar in [**22**]. They observed that, in the special case of $\{\, \nabla f_i(x) \,|\, i \in I(x;\xi) \,\}$ linearly independent the above formulas agree. They also showed that for a fixed $\xi \in \Xi(x,v)$ and for any $y \in Y(x,v)$,

$$N_{\Xi(x,v)}(\xi) = E\big(I(x,\xi), y\big),$$

and illustrated another difficulty by giving an example to the case that $E(I^*, y)$ can be a proper subset of $N_{\Xi(x,v)}(\xi)$

Here we consider a very special case, the maximum of finitely many linear functionals on $\mathbb{R}$, and we shall show that the various formulas agree.

Theorem 5.2.2. *Consider, for any $x \in \mathbb{R}$, $f(x) = \max_{i\in\{1,\ldots,k\}} f_i(x)$; where $f_i(x) = a_i x + b_i$. Then the set of subgradients of f at x is a closed interval in $\mathbb{R}$, namely*

$$\partial f(x) = [a, b], \quad \text{where } a = \min_{i\in I(x)} a_i, \quad b = \max_{i\in I(x)} a_i.$$

For each $v \in \partial f(x)$, ∂f is proto-differentiable at x, explicitly

(i) If $\partial f(x)$ is a singleton, i.e., $a = b = v = \partial f(x)$, then

$$(\partial f)'_{x,v}(\xi) = 0 \text{ for all } \xi.$$

(ii) If $v \in (a, b)$, where $a \neq b$, then

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \mathbb{R} & \text{if } \xi = 0, \\ \emptyset & \text{if } \xi \neq 0. \end{cases}$$

(iii) If $v = a$ and $a \neq b$, then

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \{0\} & \text{if } \xi < 0. \\ \mathbb{R}_+ \cup \{0\} & \text{if } \xi = 0. \\ \emptyset & \text{if } \xi > 0. \end{cases}$$

(iv) If $v = b$ and $a \neq b$, then

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \emptyset & \text{if } \xi < 0, \\ \mathbb{R}_- \cup \{0\} & \text{if } \xi = 0, \\ \{0\} & \text{if } \xi > 0. \end{cases}$$

Moreover, the proto-derivative formulas (5.1.4) and (5.2.1) agree and equals (i)-(iv) in each case.

Proof. It is plain that $\partial f(x) = [a, b]$, since $\nabla f_i(x) = a_i$ for each i. We first show that the proto-derivative formula (5.1.4) reduced to (i)-(iv) in each case.

(i) $\partial f(x) = v$

First note that $N_{\Xi(x,v)}(\xi) = 0$ for all ξ since $\Xi(x,v) = N_{\partial f(x)}(v) = \mathbb{R}$. Then by (5.1.4),

$$(\partial f)'_{x,v}(\xi) = 0 \text{ for all } \xi.$$

(ii) $v \in (a,b)$

In this case $\Xi(x,v) = N_{\partial f(x)}(v) = \{0\}$ and hence $N_{\Xi(x,v)}(\xi) = \mathbb{R}$ for $\xi = 0$. Thus by (5.1.4),

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \mathbb{R} & \text{if } \xi = 0, \\ \emptyset & \text{if } \xi \neq 0. \end{cases}$$

(iii) $v = a$

If $v = a$ then notice that $\Xi(x,v) = N_{\partial f(x)}(v) = \mathbb{R}_- \cup \{0\}$ and

$$N_{\Xi(x,v)}(\xi) = \begin{cases} \{0\} & \text{if } \xi < 0, \\ \mathbb{R}_+ \cup \{0\} & \text{if } \xi = 0. \end{cases}$$

Hence by (5.1.4),

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \{0\} & \text{if } \xi < 0, \\ \mathbb{R}_+ \cup \{0\} & \text{if } \xi = 0, \\ \emptyset & \text{if } \xi > 0. \end{cases}$$

(iv) $v = b$

Here, we have $\Xi(x,v) = N_{\partial f(x)}(v) = \mathbb{R}_+ \cup \{0\}$ and

$$N_{\Xi(x,v)}(\xi) = \begin{cases} \{0\} & \text{if } \xi < 0, \\ \mathbb{R}_- \cup \{0\} & \text{if } \xi = 0. \end{cases}$$

Thus by (5.1.4),

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \emptyset & \text{if } \xi < 0, \\ \mathbb{R}_- \cup \{0\} & \text{if } \xi = 0, \\ \{0\} & \text{if } \xi > 0. \end{cases}$$

Next we show that the formula (5.2.1) agrees with (i)-(iv) in each case. First note that the formula (5.2.1) has a particularly simple form, since the functions involved in the formula are linear, namely

$$(\partial f)'_{x,v}(\xi) = \bigcup_{I^* \in S(x,v,\xi)} \ \bigcup_{y \in Y(I^*,v)} E(I^*, y),$$

and the infinitely repeated index set I^* is always a singleton for $\xi \neq 0$. Therefore, any $x \in \mathbb{R}$, I^* can be written as

$$I^* = \begin{cases} \{\ell\} & \text{if } \xi > 0, \\ \{m\} & \text{if } \xi < 0, \\ \{\ell, m\} & \text{if } \xi = 0, \quad \text{where } \ell, m \in \{1, \dots, k\}. \end{cases}$$

(i) $\partial f(x) = v$

Since $\partial f(x)$ is a singleton, I^* is also a singleton for all ξ, say $I^* = \{p\}$, and the condition $\sum_{i=1}^{k} y_i \nabla f_i(x) = v$ in $Y(x, v)$ is trivially satisfied. Hence, $Y\big(\{p\}, v\big) = (0, \dots, 0, 1, 0, \dots, 0)$, here the p^{th} coordinate is 1.

$$E\big(\{p\}, v\big) = \Big\{ \textstyle\sum_{i=1}^{k} \sigma_i \nabla f_i(x) \,\Big|\, \sum_{i=1}^{k} \sigma_i = 0, \ \sigma_i = 0 \text{ for all } i \neq p \Big\} = \{0\}.$$

Therefore,

$$(\partial f)'_{x,v}(\xi) = 0 \text{ for all } \xi.$$

(ii) $v \in (a, b)$

When $\xi > 0$, $I^* = \{\ell\}$, and

$$Y\big(\{\ell\}, v\big) = \Big\{ y \,\Big|\, y_i = 0 \text{ for all } i \neq \ell, \ y_\ell b = v \text{ with } y_\ell = 1 \Big\}.$$

Since $v \neq b$ the set $Y(\{\ell\}, v)$ is empty. Similarly, for $\xi < 0$ we have $Y(\{m\}, v) = \emptyset$. When $\xi = 0$, $I^* = \{\ell, m\}$, and

$$Y\big(\{\ell, m\}, v\big) = \Big\{ y \,\Big|\, y_i = 0 \text{ for all } i \neq \ell, m, \ y_\ell b + y_m a = v, \\ \text{with } y_\ell + y_m = 1 \text{ and } y_\ell, y_m \geq 0 \Big\}.$$

Since $v \in (a, b)$, the quantities y_ℓ and y_m has to be (strictly) positive and hence

$$\begin{aligned} E\big(\{\ell, m\}, v\big) &= \Big\{ \sigma_\ell b + \sigma_m a \,\Big|\, \sigma_\ell + \sigma_m = 0 \text{ and } \sigma_\ell, \sigma_m \in \mathbb{R} \Big\}, \\ &= \Big\{ \sigma_\ell (b - a) \,\Big|\, \sigma_\ell \in \mathbb{R} \Big\}, \\ &= \mathbb{R}. \end{aligned}$$

Thus,

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \mathbb{R} & \text{if } \xi = 0, \\ \emptyset & \text{if } \xi \neq 0. \end{cases}$$

(iii) $v = a$

When $\xi > 0$. $I^* = \{\ell\}$, and

$$Y\big(\{\ell\}, v\big) = \big\{\, y \,\big|\, y_i = 0 \text{ for all } i \neq \ell,\ y_\ell b = a \text{ with } y_\ell = 1 \,\big\}.$$

Since $b \neq a$ the set $Y\big(\{\ell\}, v\big)$ is empty.

When $\xi < 0$, $I^* = \{m\}$, and

$$\begin{aligned} Y\big(\{m\}, v\big) &= \big\{\, y \,\big|\, y_i = 0 \text{ for all } i \neq m,\ y_m a = a \text{ with } y_m = 1 \,\big\}, \\ &= (0, \ldots, 0, 1, 0, \ldots, 0);\ \text{ here } m^{th} \text{ coordinate is } 1\ . \end{aligned}$$

Then

$$E\big(\{m\}, v\big) = \big\{\, \textstyle\sum_{i=1}^k \sigma_i \nabla f_i(x) \,\big|\, \sum_{i=1}^k \sigma_i = 0,\ \sigma_i = 0 \text{ for all } i \neq m \,\big\} = \{0\}.$$

When $\xi = 0$, $I^* = \{\ell, m\}$, and

$$\begin{aligned} Y\big(\{\ell, m\}, v\big) &= \big\{\, y \,\big|\, y_i = 0 \text{ for all } i \neq \ell, m,\ y_\ell b + y_m a = a, \\ &\qquad\qquad \text{with } y_\ell + y_m = 1 \text{ and } y_\ell, y_m \geq 0 \,\big\}, \\ &= (0, \ldots, 0, 1, 0, \ldots, 0);\ \text{ here } m^{th} \text{ coordinate is } 1\ . \end{aligned}$$

Hence,

$$\begin{aligned} E\big(\{\ell, m\}, v\big) &= \big\{\, \sigma_\ell b + \sigma_m a \,\big|\, \sigma_\ell + \sigma_m = 0,\ \sigma_\ell \geq 0 \text{ since } y_\ell = 0 \,\big\}, \\ &= \big\{\, \sigma_\ell (b - a) \,\big|\, \sigma_\ell \geq 0 \,\big\}, \\ &= \mathbb{R}_+ \cup \{0\}. \end{aligned}$$

Thus,

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \{0\} & \text{if } \xi < 0, \\ \mathbb{R}_+ \cup \{0\} & \text{if } \xi = 0, \\ \emptyset & \text{if } \xi > 0. \end{cases}$$

(iv) $v = b$

When $\xi > 0$, $I^* = \{\ell\}$ and

$$\begin{aligned} Y\big(\{\ell\}, v\big) &= \big\{\, y \,\big|\, y_i = 0 \text{ for all } i \neq \ell,\ y_\ell b = b \text{ with } y_\ell = 1 \,\big\}, \\ &= (0, \ldots, 0, 1, 0, \ldots, 0);\ \text{ here } \ell^{th} \text{ coordinate is } 1\ . \end{aligned}$$

$$E(\{\ell\},v) = \left\{ \textstyle\sum_{i=1}^{k}\sigma_i \nabla f_i(x) \,\middle|\, \sum_{i=1}^{k}\sigma_i = 0,\ \sigma_i = 0 \text{ for all } i \neq \ell \right\} = \{0\}.$$

When $\xi < 0$, $I^* = \{m\}$, and

$$Y(\{m\},v) = \left\{ y \,\middle|\, y_i = 0 \text{ for all } i \neq m,\ y_m a = b \text{ with } y_m = 1 \right\}.$$

Since $b \neq a$ the set $Y(\{m\},v)$ is empty.
When $\xi = 0$, $I^* = \{\ell, m\}$, and

$$\begin{aligned} Y(\{\ell,m\},v) = \Big\{ y \,\Big|\, & y_i = 0 \text{ for all } i \neq \ell, m,\ y_\ell b + y_m a = b, \\ & \text{with } y_\ell + y_m = 1 \text{ and } y_\ell, y_m \geq 0 \Big\}, \\ = & (0,\ldots,0,1,0,\ldots,0); \text{ here } \ell^{th} \text{ coordinate is } 1 . \end{aligned}$$

Hence

$$\begin{aligned} E(\{\ell,m\},v) &= \left\{ \sigma_\ell b + \sigma_m a \,\middle|\, \sigma_\ell + \sigma_m = 0,\ \sigma_m \geq 0 \text{ since } y_m = 0 \right\}, \\ &= \left\{ \sigma_m(a-b) \,\middle|\, \sigma_m \geq 0 \right\}, \\ &= \mathbb{R}_- \cup \{0\}. \quad (\text{ since } a - b < 0) \end{aligned}$$

Thus

$$(\partial f)'_{x,v}(\xi) = \begin{cases} \emptyset & \text{if } \xi < 0, \\ \mathbb{R}_- \cup \{0\} & \text{if } \xi = 0, \\ \{0\} & \text{if } \xi > 0. \end{cases}$$

□

One can easily derive the subdifferential properties of the absolute value function, from Theorem 5.2.2.

Corollary 5.2.3. *Let $f(x) = |x|$ for $x \in \mathbb{R}$. Then*

$$\partial f(x) = \begin{cases} 1 & \text{if } x > 0, \\ [-1,1] & \text{if } x = 0, \\ -1 & \text{if } x < 0, \end{cases}$$

and the subgradient mapping $\partial f : \mathbb{R} \rightrightarrows \mathbb{R}$ *is everywhere proto-differentiable with*

$$(\partial f)'_{x>0,1}(\xi) = 0 \ \textit{for all} \ \ \xi,$$

$$(\partial f)'_{x<0,-1}(\xi) = 0 \ \textit{for all} \ \ \xi,$$

$$(\partial f)'_{x=0,v\in(-1,1)}(\xi) = \begin{cases} \mathbb{R} & \textit{if } \xi = 0, \\ \emptyset & \textit{if } \xi \neq 0, \end{cases}$$

$$(\partial f)'_{x=0,v=-1}(\xi) = \begin{cases} \{0\} & \textit{if } \xi < 0, \\ \mathbb{R}_+ \cup \{0\} & \textit{if } \xi = 0, \\ \emptyset & \textit{if } \xi > 0, \end{cases}$$

$$(\partial f)'_{x=0,v=1}(\xi) = \begin{cases} \emptyset & \textit{if } \xi < 0, \\ \mathbb{R}_- \cup \{0\} & \textit{if } \xi = 0, \\ \{0\} & \textit{if } \xi > 0. \end{cases}$$

Proof. Apply Theorem 5.2.2 observing that $|x| = \max\{x, -x\}$. □

REFERENCES

1. H. Attouch, **Variational Convergence for Functions and Operators**, Pitman, New York, 1984.
2. H. Attouch and J. B. Wets, "Epigraphical Analysis," in *Analyse Non Linéaire* (Perpignan, 1987). **6** (1989), Suppl., 73–100.
3. H. Attouch and J. B. Wets, "Quantitative stability of variational systems. I. The epigraphical distance," Trans. Amer. Math. Soc. **328** (1991), 695–729.
4. J. P. Aubin and I. Ekeland, **Applied Nonlinear Analysis**, Wiley, 1984.
5. J. P. Aubin and H. Frankowska, **Set-Valued Analysis**, Birkhäuser, 1990.
6. A. Auslender and R. Cominetti, "A comparative study of multifunction differentiability with applications in mathematical programming," Math. Oper. Research **16** (1991), 240–258.
7. J. Beer and J. Borwein, "Mosco convergence and duality," Preprint.
8. F. H. Clarke, "Generalized gradients and applications," Trans. Amer. Math. Soc. **205** (1975), 247–262.
9. F. H. Clarke, **Optimization and Nonsmooth Analysis**, Wiley, 1983.
10. F. H. Clarke, **Methods of Dynamic and Nonsmooth Optimization**, CBMS-NSF Regional Conference Series in Applied Mathematics, **57** SIAM Publications, Philadelphia, PA, 1989.
11. R. Cominetti, "On Pseudo-differentiability," Trans. Amer. Math. Sco. **324** (1991), 843–865.
12. C. N. Do, "Generalized second derivatives of convex functions in reflexive Banach spaces," Trans. Amer. Math. Sco. **334** (1992), 281–301.
13. A. D. Ioffe, "Variational analysis of a composite function: a formula for the lower second-order epi-derivative," J. Math. Anal. Appl. **160**, No. 2, (1991), 379–405.

14. A. D. Ioffe and V. M. Tihomirov, **Theory of Extremal Problems**, North-Holland, 1979.

15. S. Kurcyusz and J. Zowe, "Regularity and stability for the mathematical programming problem in Banach spaces," Appl. Math. Optim. **5** (1979), 49–62.

16. U. Mosco, "Convergence of convex sets and of solutions of variational inequalities," Adv. in Math. **3** (1969), 510–585.

17. J. P. Penot, "On the differentiability of the subdifferential of a max-function," preprint, 1992.

18. R. A. Poliquin, "Proto-differentiation of subgradient set-valued mappings," Canadian J. Math. **42** (1990), 520–532.

19. R. A. Poliquin, "An extension of Attouch's Theorem and its applications to second-order epi-differentiation of convexly composite functions," Trans. Amer. Math. Soc. **332** (1992), 861–874.

20. R. A. Poliquin and R. T. Rockafellar, "Amenable functions in optimization," to appear in the Proceedings of the International School of Mathematics G. Stampacchia, 10th course: Nonsmooth Optimization: Methods and Applications, 1993.

21. R. A. Poliquin and R. T. Rockafellar, "A calculus of epi-derivatives applicable to optimization," Canadian J. Math., to appear.

22. R. A. Poliquin and R. T. Rockafellar, "Proto-derivative formulas for basic subgradient mappings in mathematical programming," Canadian J. Math., to appear.

23. R. T. Rockafellar, **Convex Analysis**, Princeton University Press, Princeton, NJ, 1970.

24. R. T. Rockafellar, "Directionally Lipschitzian functions and subdifferential calculus," Proc. London Math. Soc. **39** (1979), 331-355.

25. R. T. Rockafellar, "Favorable classes of Lipschitz-continuous functions in subgradient optimization, " Progress in Nondifferentiable Optimization, E. Nur-

minski (ed.). II ASA Collaborative Proceedings Series. International Institute of Applied Systems Analysis. Laxenburg. Austria 1982. 125–144.

26. R. T. Rockafellar. " First- and second-order epi-differentiability in nonlinear programming." Trans. Amer. Math. Sco. **307** (1988). 75–107.

27. R. T. Rockafellar. "Proto-differentiability of set-valued mappings and its applications in optimization." in Analyse Non Linéaire. H. Attouch et al.(eds.). Gauthier-Villars. Paris (1989). 449-482.

28. R. T. Rockafellar. "Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives." Math. of Oper. Research **14** (1989). 462–484.

29. R. T. Rockafellar. "Nonsmooth analysis and parametric optimization." in Methods of Nonconvex Analysis (A. Cellina. ed.) Springer-verlag Lecture Notes in Math. No. 1446 (1990). 137–151.

30. R. T. Rockafellar. "Generalized second derivatives of convex functions and saddle functions." Trans. Amer. Math. Sco. **320** (1990). 810–822.

31. G. Salinetti and J. B. Wets. "On the relations between two types of convergence for convex functions." J. Math. Anal. Appl. **60** (1977). 211–226.

32. G. Salinetti and J. B. Wets. " On the convergence of sequence of convex sets in finite dimensions." SIAM Rev. **21** (1979). 18–33.

33. J. B. Wets. **Convergence of Convex Functions. Variational Inequalities and Convex Optimization Problems**. Wiley. Chichester (1980). 375–419.

34. R. Wijsman. "Convergence of sequence of convex sets. cones and functions II." Trans. Amer. Math. Soc. **123** (1966). 32–45.