YASAR UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
MATHEMATICS

MASTER THESIS

SUBDIFFERENTIALS IN NON-SMOOTH ANALYSIS

Ece GÜRBÜZ

Supervisor
Assist. Prof. Dr. Shahlar MAHARRAMOV

Izmir, 2011

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “Subdifferentials in Nonsmooth Analysis” adlı çalışmanın, tarafımdan bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin “References” bölümünde gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

26.05.11

Ece GÜRBÜZ

ACKNOWLEDGEMENTS

I would like to thank my supervisor Assist. Prof. Dr. Shahlar Maharramov for his patience, willingness to help to his students and his guidance which enabled me to develop an understanding of the subject and to write this thesis.

I am also grateful to all my teachers and professors, especially to Assist. Prof. Dr. Gökşen Bacak Turan who motivated me to attend graduate school, for their help during my study at university.

I would also like to thank my family for their support and encourgement that provided me through my entire life.

Without one of these people, I would not have finished this thesis.

ÖZET

Yüksek Lisans Tezi

DÜZGÜN OLMAYAN ANALİZDE SUBDİFFERANSİYELLER

Ece GÜRBÜZ

Yaşar Üniversitesi

Fen Bilimleri Enstitüsü

Matematik

İlk bölümde, zayıf subdifferentialların bazı özellikleri ele alındı. Yayınlarda [2,12,13] tanımlanmış zayıf subdifferentialların tanım ve özellikleri kullanılarak, düzgün olmayan ve konveks olmayan analizdeki zayıf subdifferansiyeller ile ilgili bazı teoremlerin ispatları yapıldı.

İkinci bölümde, herhangibir X Banach uzayında, genelleştirilmiş gradyantların analizi incelendi.

Üçüncü bölümde, kesikli optimal teori alanında bir araştırma sunuldu. Bir parametreye bağlı basamak kontrol problemi incelendi. Basamak kontrol problemi için kesikli maksimum prensibinin yeni bir versiyonu türetildi.

Anahtar Sözcükler: *Zayıf Subdifferansiyel, Subdifferansiyel, Superdifferansiyel, Optimal Kontrol Problem.*

ABSTRACT

Master Thesis

SUBDIFFERENTIALS IN NON-SMOOTH ANALYSIS

Ece GÜRBÜZ

Yasar University

Institute of Natural and Applied Sciences

Master of Mathematics

In the first chapter, some properties of the weak subdifferential is considered. By using definition and properties of the weak subdifferential which described in the papers [2,12,13], we prove some theorem connecting weak subdifferential in non-smooth and non-convex analysis.

In the second chapter, we consider the calculus of generalized gradients in an arbitrary Banach space *X*.

In the third chapter, we discuss the discrete optimal control theory. The step control problem depending on a parameter is investigated. No smoothness of the cost function φ is assumed and new versions of the discrete maximum principle for the step control problem are derived.

Keywords: *Weak Subdifferential, Subdifferential, Superdifferential, Optimal Control Problem.*

CONTEXT

INTRODUCTION

Nonsmooth analysis had its origins in the early 1970s when control theorists and nonlinear programmers attempted to deal with necessary optimality conditions for problems with nonsmooth data or with nonsmooth functions (such as the pointwise maximum of several smooth functions) that arise even in many problems with smooth data. The first such canonical generalized gradient was the generalized gradient introduced by Clarke in his work [9]. He applied this generalized gradient systematically to nonsmooth problems in a variety of problems. Several of the other frequently used generalized derivative concepts are the co-derivatives introduced by Mordukhovich[5], approximate and geometric subdifferentials introduced by Ioffe [10], Michel and Penot's derivatives [11], Rockafellar and Wets [12] provide a comprehensive overview of the field. Since a nonconvex set has no supporting hyperline at each boundary point, the notion of subgradient have been generalized by most researches on optimality conditions for nonconvex problems. The notion of weak subdifferential which is a generalization of the classic subdifferential, is introduced by Azimov and Gasimov[2]. In first section of the thesis, we investigate relationships between Frechet lower subdifferential and weak subdifferential, prove some theorem connecting weak subdifferential. With the start, we give some definition which will be usefull for us some parts of the current paper. Let $(X, \|.\|_X)$ be a real normed space, and let X^* be a topological dual of X.

The second section of the thesis is devoted to the generalized gradient and its applications. The calculus of generalized gradients is the best-known and most frequently invoked part of nonsmooth analysis. Unlike proximal calculus, it can be developed in an arbitrary Banach space X. In the second chapter, we make a fresh start in such a setting, and we begin with functions and not sets. We present the basis results for the class of locally Lipschitz functions. Then the associated geometric concepts are introduced, including for the first time a look at tangency. In fact, we examine two notions of tangency; sets for which they coincide are termed regular and enjoy useful properties. We proceed to relate the generalized gradient to the constructs of the preceeding chapter when X is Hilbert space. Finally, we derive a useful limiting-gradient characterization when the underlying space is finite dimensional.

In the last section, third chapter, we investigate necessary optimality condition for switching system in the discrete case. Some applied problems in fields such as economy, military defense, and chemistry are inherently multistage problems in nonsmooth optimization. In such problems, there are several stages which are characterized by their own equations, controls, phase coordinates, constants, etc. Usually these stages can be connected to each other by additional conditions. Here problems will be considered where these relations are given by switching points which are controlled by a given parameter. These multistage processes will be called step control systems or discrete systems with varying structure.

CHAPTER 1

SOME PROPERTIES OF THE WEAK SUBDIFFERENTIAL

Definition 1.1:

F is called strictly differentiable at x with a strict derivative $\nabla F(x)$ if

$$\lim_{u \to x, \bar{u} \to x} \frac{F(\bar{u}) - F(u) - (\nabla F(x), \bar{u} - u)}{\|\bar{u} - u\|} = 0$$

Definition 1.2: Let F: $X \to R$ be a single-valued function, and let $\bar{x} \in X$ be a given point where $F(\bar{x})$ is finite. A pair $(x^*, c) \in X^* \times R_+$ is called the weak subgradient of F at $\bar{x}$ if $F(x) - F(\bar{x}) \geq (x^*, x - \bar{x}) - c\|x - \bar{x}\|$ for all $x \in X$, here R_+ is the set of nonnegative real numbers.

The set $\partial^W F(\bar{x}) = \{(x^*, c) \in X^* \times R_+ : F(x) - F(\bar{x}) \geq (x^*, x - \bar{x}) - c\|x - \bar{x}\|\}$

for all $x \in X$ is called weak subdifferential for the F at the point $\bar{x} \in X$.

Remark 1.3: It is noted in the references [2, remark 2.3] that if there is a continuous (superlinear) concave function

$$g(x) = (x^*, x - \bar{x}) + F(\bar{x}) - c\|x - \bar{x}\|$$

such that $g(x) \leq F(x)$ for all $x \in X$ and $g(\bar{x}) = F(\bar{x})$, then the pair

$(x^*, c) \in X^* \times R_+$ is a weak subgradient of F at $\bar{x} \in X$. This opinion is also necessary for the pair $(x^*, c) \in X^* \times R_+$ to be weak subgradient of F at $\bar{x}$.

We can add extra opinion such that the norm of the gradient ∇g is a bounded above. In fact, if we take gradient of the functional

$$g(x) = (x^*, x - \bar{x}) + F(\bar{x}) - c\|x - \bar{x}\|,$$

we get

$$\nabla g(x) = x^* - c\frac{(x - \bar{x})}{\|x - \bar{x}\|}.$$

Then, if we calculate norm of the gradient of the functional $\nabla g(x)$, we get

$$\|\nabla g(x)\| = \left\|x^* - c\frac{(x - \bar{x})}{\|x - \bar{x}\|}\right\| \leq \|x^*\| + \left\|c\frac{(x - \bar{x})}{\|x - \bar{x}\|}\right\| = \|x^*\| + c\frac{\|x - \bar{x}\|}{\|x - \bar{x}\|} = \|x^*\| + c$$

$\Rightarrow$ $\|\nabla g(x)\| \leq \|x^*\| + c$ for all $x \in X$ and $x \neq \bar{x}$.

It means that the gradient $\|\nabla g(x)\|$ is bounded above by the number $\|x^*\| + c$.

Definition 1.4: The set $\partial F(\bar{x}) = \left\{x^* \in X^* : \liminf_{x \to \bar{x}} \frac{F(x)-F(\bar{x})-(x^*,\ x-\bar{x})}{\|x-\bar{x}\|} \geq 0\right\}$ is called a Frechet subdifferential of the F at $\bar{x}$.

Let us note that the Frechet subdifferential may be empty for some functions.

Example 1.5: Take $F: R \to R: F(x) = -|x|, x \in R$. Easy calculation shows that Frechet subdifferential for above example at the point zero is empty, i.e. , $\partial F(0) = \emptyset$

Theorem 1.6: If x^* is a Frechet subgradient for the functional $F: X \to R$ at the point $\bar{x}$, then the couple (x^*, c) is a weak subdifferential for the functional F(x) at $\bar{x}$ for any nonnegative $c \in R_+$.

Proof: Let x^* is a Frechet subgradient for the functional $F: X \to R$ at the point $\bar{x}$. Then by using above maintained definition of the Frechet subdifferential, we can write this definition equavelentely as follows

$$F(x) - F(\bar{x}) - (x^*,\ x - \bar{x}) \geq o\|x - \bar{x}\|.$$

It is easy to show that right side of the last inequality no less than $-c\|x - \bar{x}\|$ for any nonnegative *c*. Then it follows that

$$F(x) - F(\bar{x}) - (x^*,\ x - \bar{x}) \geq o\|x - \bar{x}\| \geq -c\|x - \bar{x}\|.$$

Last equation says that (x^*, c) is a weak subdifferential for the functional *F(x)* at $\bar{x}$. Following theorem is a analogous of the proposition 4, p.52 in the article 12.

Theorem 1.7: Let *F(x)* is a finite at $\bar{x}$, $g \in C^1$ in a neighborhood of $\bar{x}$. Then if $(x^*, c) \in \partial^W(F + g)$, then $(x^* - g^{'}(\bar{x}), -2c) \in \partial^W F(\bar{x})$.

Proof: If we put –*g* in the definition of weak subdifferential, then

$-g(x) + g(\bar{x}) + c\|x - \bar{x}\| \geq (-g(\bar{x}), x - \bar{x})$ for all $x \in X$. Since $(x^*, c) \in \partial^W(F + g)$, we have $F(x) + g(x) - F(\bar{x}) - g(\bar{x}) + c\|x - \bar{x}\| \geq (x^*, x - \bar{x})$

for all $x \in X$ near $\bar{x}$. Upon adding these inequalities we arrive that ,

$$F(x) - F(\bar{x}) + 2c\|x - \bar{x}\| \geq (x^* - g^{'}(\bar{x}), x - \bar{x}) \Rightarrow$$

$$F(x) - F(\bar{x}) \geq (x^* - g^{'}(\bar{x}), x - \bar{x}) - 2c\|x - \bar{x}\|$$

Last inequality means that $(x^* - g^{'}(\bar{x}), -2c) \in \partial^W F(\bar{x})$.

Theorem 1.8: If F is positively homogenous, then $\partial^W F(\lambda\bar{x}) = \partial^W F(\bar{x})$ for the positive real number λ.

Proof: Let F is a weak subdifferential at the point $\bar{x}$ with the pair $(x^*, c) \in (X^*, R_+)$

Then by using definition of weak subdifferential, we can show weak subdifferential set for the function $F(\lambda x)$ at the point $\bar{x}$ as follow $\partial^W F(\lambda\bar{x}) = \{(x^*, c): F(x) - F(\bar{x}) - (x^*,\ x - \bar{x}) \geq c\|x - \bar{x}\|\}$, which means the pair $(x^*, c) \in (X^*, R_+)$ is also weak subdifferential for the functional F at the point $\bar{x}$.

Theorem 1.9: Let *F(x)* is finite at $\bar{x}$, and (x^*, c) is weak subdifferential for *F(x)* at $\bar{x}$, which $x^* = c(x - \bar{x})$ for some $x \in X$.

Then $F(x) \geq F(\bar{x})$ for all $x \in X$, which satisfying $\|x - \bar{x}\| \geq 1$.

Proof: If $x^* = c(x - \bar{x})$ then $F(x) - F(\bar{x}) \geq (x^*,\ x - \bar{x}) - c\|x - \bar{x}\|$

$$F(x) - F(\bar{x}) \geq c\|x - \bar{x}\|^2 - c\|x - \bar{x}\| = c\|x - \bar{x}\|(\|x - \bar{x}\| - 1)$$

Then for $\|x - \bar{x}\| \geq 1$ we get $F(x) \geq F(\bar{x})$

Theorem 1.10: If F is strictly differentiable at $\bar{x}$ with a derivative $\nabla F(\bar{x})$. Then for any $(x^*, c) \in \partial^W F(U)$ there exist $\delta > 0$ such that $x^* \in \nabla F(\bar{x}) + \frac{3c}{2} B^*$, where $u \in B_\delta(\bar{x})$-sphere with radius δ and B^* - unit sphere.

Proof: It follows from definition of strict differentiable that for any $\varepsilon > 0$ there exist $\delta > 0$ such that

$$|F(\bar{u}) - F(u) - (\Delta F(\bar{x}), \bar{u} - u)| \leq \frac{\varepsilon}{2}\|\bar{u} - u\|, \forall \bar{u}, u \in B_\delta(\bar{x}). \qquad (1)$$

Let take $u \in B_\delta(\bar{x})$ and assume there exist $(x^*, c) \in \partial^W F(U)$. Then if we take $\varepsilon = c$, it follows definition of weak subdifferential that

$$F(\bar{u}) - F(u) - (x^*, \bar{u} - u) \geq -c\|\bar{u} - u\| \qquad (2)$$

From (1) and (2) we get

$$(\Delta F(\bar{x}) - x^*, \bar{u} - u) \leq \frac{c}{2}\|\bar{u} - u\| + c\|\bar{u} - u\| = \frac{3c}{2}$$

From last relation we can write that

$\|\Delta F(\bar{x}) - x^*\| \leq \frac{3c}{2}$ which reduce to the $x^* \in \Delta F(\bar{x}) + \frac{3c}{2}$.

CHAPTER 2

CLARKE SUBDIFFERENTIAL(GENERALIZED)

2.1 Definition and Basic Properties

Throughout this chapter, *X* is a real Banach space. Let $f: X \to \mathbb{R}$ be Lipschitz of rank *K* near a given point $x \in X$; that is, for some $\in > 0$, we have

$$|f(y) - f(z)| \leq K\|y - z\| \forall y, z \in B(x; \varepsilon).$$

The generalized directional derivative of *f* at *x* in the direction *v*, denoted

$$f^\circ(x; v) \coloneqq \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t},$$

where of course *y* is a vector in *X* and *t* is a positive scalar. Note that this definition does not presuppose the existence of any limit (since it involves an upper limit only), that it involves only the behavior of *f* arbitrarily near *x*, and that it differs from the traditional definition of the directional derivative in that the base point (*y*) of the difference quotient varies. The utility of f° stems from the following basic properties. (A function g is positively homogenous if $g(\lambda v) = \lambda g(v)$ for $\lambda \geq 0$, and subadditive if $g(v + w) \leq g(v) + g(w)$.)

Proposition 2.1.1: Let *f* be Lipschitz of rank K near *x*. Then:

(a) The function $v \to f^\circ(x; v)$is finite, positively homogeneous, and subadditive on *X*, and satisfies

$$\left|f^\circ(x; v)\right| \leq K\|v\|.$$

(b) $f^\circ(x; v)$ is upper semicontinuous as a function of $(x; v)$ and as a function of *v* alone, is a Lipschitz of rank K on *X*.

(c) $f^\circ(x; -v) = (-f)^\circ(x; v)$.

Proof: In view of the Lipschitz condition, the absolute value of the difference quotient in the definition of $f^\circ(x; v)$ is bounded by $K\|v\|$ when *y* is sufficiently near *x* and *t* sufficiently near 0. It follows that $\left|f^\circ(x; v)\right|$ admits the same upper bound. The fact that $f^\circ(x; \lambda v) = \lambda f^\circ(x; v)$ for any $\lambda \geq 0$ is immidiate, so let us turn now to the subadditivity. With all the upper limits below understood to be taken as $y \to x$ and t$\downarrow$ 0, we calculate:

$$f^{\circ}(x; v+w) = limsup \frac{f(y+tv+tw) - f(y)}{t}$$

$$\leq limsup \frac{f(y+tv+tw) - f(y+tw)}{t}$$

$$+ limsup \frac{f(y+tw) - f(y)}{t}$$

(Since the upper limit of a sum is bounded above by the sum of the upper limits). The first upper limit in this last expression is $f^{\circ}(x; v)$, since the term $y + tw$ represents in essence just a dummy variable converging to x.

We conclude

$$f^{\circ}(x; v+w) \leq f^{\circ}(x; v) + f^{\circ}(x; w).$$

which establishes (a).

Now let $\{x_i\}$ and $\{v_i\}$ be arbitrary sequences converging to x and v, respectively. For each i, by definition of the upper limit, there exist y_i in X and $t_i > 0$ such that

$$\|y_i - x_i\| + t_i < \frac{1}{i},$$

$$f^{\circ}(x_i, v_i) - \frac{1}{i} \leq \frac{f(y_i + t_i v_i) - f(y_i)}{t_i}$$

$$= \frac{f(y_i + t_i v) - f(y_i)}{t_i} + \frac{f(y_i + t_i v_i) - f(y_i + t_i v)}{t_i}$$

Note that the last term is bounded in magnitude by $K\|v_i - v\|$(in view of the Lipschitz condition). Upon taking upper limits (as $i \to \infty$), we derive

$$\limsup_{i \to \infty} f^{\circ}(x_i; v_i) \leq f^{\circ}(x; v),$$

which establishes the upper semicontinuity.

Finally, let any v and w in X be given. We have

$$f(y+tv) - f(y) \leq f(y+tw) - f(y) + K\|v-w\|t$$

for y near x, t near 0. Dividing by t and taking upper limits as $y \to x$, $t \downarrow 0$, gives

$$f^{\circ}(x; v) \leq f^{\circ}(x; w) + K\|v - w\|.$$

Since this also holds with v and w switched, (b) follows. to prove (c), we calculate:

$$f^{\circ}(x;-v) := \limsup_{\substack{x'\to x \\ t\downarrow 0}} \frac{f(x' - tv) - f(x')}{t}$$

$$= \limsup_{\substack{u\to x \\ t\downarrow 0}} \frac{(-f)(u+tv) - (-f)(u)}{t}, where\ u := x' - tv$$

$$=(-f)^{\circ}(x;v),$$

as stated.

Exercise 2.1.2: Let *f* and *g* be Lipschitz near *x*. Prove that for any $v \in X$,

$$(f+g)^{\circ}(x;v) \le f^{\circ}(x;v) + g^{\circ}(x;v).$$

A function such as $v \mapsto f^{\circ}(x;v)$ which is positively homogeneous and subadditive on *X* is the *support function* of a uniquely determined closed convex set in X^* (The dual space of continuous linear functionals on *X*).

Some terminology is in order. Given a nonempty subset Σ of X^*, its support function is the function $H_\Sigma: X \to (-\infty, \infty]$ defined as follows:

$$H_\Sigma(v) := sup\{\langle \zeta, v\rangle : \zeta \in \Sigma\},$$

where we have used the familiar convention of denoting the value of the linear functional ζ at v by $\langle \zeta, v\rangle$. We gather some useful facts about support functions in the next result.

Proposition 2.1.3:

(a) Let Σ be a nonempty subset of X^*. Then H_Σ is positively homogeneus, subadditive, and lower semicontinuous.

(b) If Σ is convex and w^*-closed, then a point ζ in X^* belongs to Σ iff we have $H_\Sigma(v) \ge \langle \zeta, v\rangle$for all v in X.

(c) More generally, if Σ and Λ are two nonempty, convex, and w^*-closed subsets of X^*, then $\Sigma \supset \Lambda$ iff $H_\Sigma(v) \ge H_\Lambda(v)$ for all v in X.

(d) If $p: X \to \mathbb{R}$ is positively homogeneous and subadditive and bounded on the unit ball, then there is a uniquely defined nonempty, convex, and w^*-compact subset Σ of X^* such that $p = H_\Sigma$.

Proof: That H_Σ is positively homogeneous and subadditive follows immediately from its definition. As the upper envelope of continuous functions, H_Σ is automatically lower semicontinuous, whence (a). We turn now to (b), which is easily seen to amount to the following assertion: if $\zeta \notin \Sigma$, then for some $v \in X$ we have $H_\Sigma(v) < \langle \zeta, v \rangle$. This is proven by applying the Hahn-Banach Seperation Theorem (see e.g., Rudin (1973)) to the topological vector space consisting of X^* with its weak*-topology, bearing in mind that the dual of that space is identified with X. The proof of (c) is immediate in light of (b); there remains (d).

Given p, we set

$$\Sigma := \{\zeta \in X^* : p(v) \geq \langle \zeta, v \rangle \; \forall v \in X\}.$$

Then Σ is seen to be convex as a consequence of the properties of p, and w*-closed as the intersection of a family of w*-closed subsets. If K is a bound for p on $B(0;1)$, then we have $\langle \zeta, v \rangle \leq K$ for all $v \in B(0;1)$ for any element ζ of Σ. It follows that Σ is bounded, and hence w*-compact by Alaoğlu's Theorem. Clearly we have $p \geq H_\Sigma$; let us prove equality. Let $v \in X$ be given. Then, by a standart form of the Hahn-Banach Theorem (Rudin (1973, Theorem 3.2)), there exists $\zeta \in X^*$ such that $\langle \zeta, w \rangle \leq p(w) \; \forall w \in X$, with $\langle \zeta, v \rangle = p(v)$. Then $\zeta \in \Sigma$, so that $H_\Sigma(v) = p(v)$as required. Finally, the uniqueness of Σ follows from c.

Returning now to our function *f*, and taking for the function p of the proposition the function $f^\circ(x;\cdot)$, we define the generalized gradient of f at *x*, denoted $\partial f(x)$, to be the (nonempty) w*-compact subset of X^* whose support function is $f^\circ(x;\cdot)$. Thus $\zeta \in \partial f(x)$ iff $f^\circ(x;\cdot) \geq \langle \zeta, v \rangle$ for all v in X. Since $f^\circ(x;\cdot)$ does not depend on which one of the two equivalent norms on X is chosen, it follows that $\partial f(x)$ too is independent of the particular norm on X.

Some immediate intiution about ∂f is available from the following exercise, where we see that the relationship between f° and ∂f generalizes the classical formula $f'(x; v) = \langle f'(x), v \rangle$ for the directional derivative $f'(x; v)$.

We proceed now to derive some of the basic properties of the generalized gradient. A multivalued function F is said to be upper semicontinuous at x if for all $\varepsilon > 0$ there exist $\delta > 0$ such that

$$\|x - y\| < \delta \Longrightarrow F(y) \subset F(x) + \varepsilon B.$$

We denote by $\|\zeta\|_*$ the norm in X^*:

$$\|\zeta\|_* := sup\{\langle\zeta, v\rangle : v \in X, \|v\| = 1\},$$

and B_* denotes the open unit ball in X^*:

Proposition 2.1.4: Let f be Lipschitz of rank K near x. Then:

(a) $\partial f(x)$ is a nonempty, convex, weak*-compact subset of X^*, and $\|\zeta\|_* \le K$ for every $\zeta \in \partial f(x)$.

(b) For every v in X we have $f^\circ(x; v) = max\{\langle\zeta, v\rangle : \zeta \in \partial f(x)\}$.

(c) $\zeta \in \partial f(x)$ iff $f^\circ(x; v) \ge \langle\zeta, v\rangle\ \forall v \in X$.

(d) If $\{x_i\}$ and $\{\zeta_i\}$ are sequences in X and X^* such that $\zeta_i \in \partial f(x_i)$ for each i, and if x_i converges to x and ζ is a weak* cluster point of the sequence $\{\zeta_i\}$, then we have $\zeta \in \partial f(x)$.

(e) If X is finite dimensional, then ∂f is upper semicontinuous at x.

Proof: We have already noted that $\partial f(x)$ is nonempty and w^*-compact. Each $\zeta \in \partial f(x)$ satisfies $\langle\zeta, v\rangle \le f^\circ(x; v) \le K\|v\|$ for all v in X, whence $\|\zeta\|_* \le K$. The assertions (b), (c) merely reiterate that $f^\circ(x;\cdot)$ is the support function of $\partial f(x)$.

Let us prove the closure propety (d). Fix $v \in X$. For each i, we have $f^\circ(x_i; v) \ge \langle\zeta_i, v\rangle$ (in view of (c)). The sequence $\{\langle\zeta_i, v\rangle\}$ is bounded in $\mathbb{R}$, and contains terms that are arbitrarily near $\langle\zeta, v\rangle$. Let us extract a subsequence of $\{\zeta_i\}$ (without relabeling) such that $\langle\zeta_i, v\rangle \to \langle\zeta, v\rangle$. Then passing to the limit in the preceeding inequality gives $f^\circ(x; v) \ge \langle\zeta, v\rangle$, since f° is upper semicontinuous in x (Proposition 2.1.1). Since v is arbitrary, it follows (from (c) again) that $\zeta \in \partial f(x)$.

We turn now to (e). Let $\varepsilon > 0$ be given; then we wish to show that for all y sufficiently near x, we have

$$\partial f(y) \subset \partial f(x) + \varepsilon\bar{B}.$$

If this is not the case, then there is a sequence y_i converging to x and points $\zeta_i \in \partial f(y_i)$ such that $\zeta_i \notin \partial f(x) + \varepsilon\bar{B}$. We can therefore seperate ζ_i from the compact convex set in question: for some $v_i \neq 0$ we have

$$\langle \zeta_i, v_i \rangle \geq max\{\langle \zeta, v_i \rangle : \zeta \in \partial f(x) + \varepsilon \bar{B}\}$$
$$= f^{\circ}(x; v_i) + \varepsilon \|v_i\|.$$

Because of positive homogeneity, we can take $\|v_i\| = 1$. Note that the sequence $\{\zeta_i\}$ is bounded. Since we are in finite dimensions, we can extract convergent subsequences from $\{\zeta_i\}$ and $\{v_i\}$ (we do not relabel):

$\zeta_i \to \zeta,\ v_i \to v,$ where $\|v\| = 1$. The inequality above gives in the limit $\langle \zeta, v \rangle \geq f^{\circ}(x; v) + \varepsilon$, while invoking part (d) yields $\zeta \in \partial f(x)$. But then (c) is contradicted. This completes the proof.

2.2 Basic Calculus

We will derive an assortment of formulas that facilitate the calculation of ∂f when f is synthesized from simpler functionals through linear combinations, maximization, composition, and so on. We always assume that the given functions are Lipschitz near the point of interest; as we will see, this property has the useful feature of being preserved under the operations in question .

Proposition 2.2.1: For any scalar λ,we have $\partial(\lambda f)(x) = \lambda \partial f(x)$.

Proof: Note that λf is Lipschitz near x, of rank $|\lambda| K$. When λ is nonnegative, $(\lambda f)^{\circ} = \lambda f^{\circ}$, and the result follows immediately. To complete the proof, it sufficies to consider now the case $\lambda = -1$. An element ζ of X^* belongs to $\partial(-f)(x)$ iff $(-f)^{\circ}(x; v) \geq \langle \zeta, v \rangle$ for all v. By Proposition 2.1.1(c), this is equivalent to: $f^{\circ}(x; -v) \geq \langle \zeta, v \rangle$ for all v, which is equivalent to $-\zeta$ belonging to $\partial f(x)$ by Proposition 2.1.3(c).)

We now examine the generalized gradient of the sum of the two functions f and g, each of which is Lipschitz near x. It is easy to see that $f + g$ is also Lipschitz near x, and we would like to relate $\partial(f + g)$ to $\partial f(x) + \partial g(x)$. We will now do so, and introduce a technique that will be used many times: That of proving an inclusion between closed convex sets by proving an equivalent inequality between support functions.

The support function of $\partial(f + g)(x)$, evaluated at v, is $(f + g)^{\circ}(x; v)$ (by definition), while that of $\partial f(x) + \partial g(x)$ is $f^{\circ}(x; v) + g^{\circ}(x; v)$ (the support function of a sum of sets is the sum of the support functions). Since the sum of two w*- compact sets is w*-compact (addition is w*-continuous on $X^* \times X^*$), it follows that the general inequality

$$(f+g)^{\circ}(x;v) \leq f^{\circ}(x;v) + g^{\circ}(x;v)$$

noted in Exercises 2.1.2 is equivalent to the inclusion

$$\partial(f+g)(x) \subset \partial f(x) + \partial g(x),$$

as observed in Proposition 2.1.3 (c).

The extension of this inclusion (a sum rule) to finite linear combinations is immediate.

Proposition 2.2.2: Let f_i (i = 1, 2, …, n) be Lipschitz near x, and let λ_i (i =1,2,…,n) be scalars. Then $f := \sum_{i=1}^{n} \lambda_i f_i$ is Lipschitz near x, and we have

$$\partial\left(\sum_{i=1}^{n} \lambda_i f_i\right)(x) \subset \sum_{i=1}^{n} \lambda_i \partial f_i(x).$$

Exercise 2.2.3: Prove Proposition 2.2.2, and give an example with $X = \mathbb{R}$ and $n = 2$ for which the inclusion is strict.

Theorem 2.2.4: (Lebourg's Mean Value Theorem) Let x and y belong to X, and suppose that f is Lipschitz on an open set containing the line segment $[x,y]$. Then there exists a point u in (x, y) such that

$$f(y) - f(x) \in \langle \partial f(u), y - x \rangle.$$

Proof: We will need the following special chain rule for the proof. We denote by x_t the point $x + t(y-x)$.

Lemma 2.2.5: The function g:[0,1]→ R defined by $g(t) = f(x_t)$ is Lipschitz on (0,1), and we have

$$\partial g(t) \subset \langle \partial f(x_t), y - x \rangle.$$

Proof of the Lemma: The fact that g is Lipschitz is plain. The two closed convex sets appearing in the equation are in fact intervals in R, so it suffices to prove that for $v = \pm 1$, we have

$$\max\{\partial g(t)v\} \leq max\{\langle \partial f(x_t), y - x \rangle v\}.$$

Now the left-hand side is just $g^{\circ}(t;v)$; that is,

$$\limsup_{\substack{s \to t \\ \lambda \downarrow 0}} \frac{g(s+\lambda v) - g(s)}{\lambda}$$

$$= \limsup_{\substack{s \to t \\ \lambda \downarrow 0}} \frac{f\big(x + [s+\lambda v](y-x)\big) - f\big(x + s(y-x)\big)}{\lambda}$$

$$\leq \limsup_{\substack{y \to x_t \\ \lambda \downarrow 0}} \frac{f\left(y' + \lambda v(y-x)\right) - f(y')}{\lambda}$$

$$= f^{\circ}\left(x_t; v(y-x)\right)$$

$$= max\langle \partial f(x_t), v(y-x)\rangle,$$

which completes the proof of the lemma.

Now to the proof of the theorem. Consider the function θ on [0,1] defined by

$$\theta(t) = f(x_t) + t[f(x) - f(y)].$$

Note that $\theta(0) = \theta(1) = f(x)$, so that there is a point t in (0,1) at which θ attains a local minimum and local maximum (by continuity). We have $0 \in \partial\theta(t)$. We may calculate $\partial\theta(t)$ by appealing to Propositions 2.2.1 and 2.2.2, and the Lemma 2.2.5. We deduce

$$0 \in f(x) - f(y) + \langle \partial f(x_t), y - x\rangle,$$

which is the assertion of the theorem (take $u = x_t$).

Theorem 2.2.6: (The Chain Rule) Let $F: X \to \mathbb{R}^n$ be Lipschitz near *x*, and let $g: \mathbb{R}^n \to \mathbb{R}$ be Lipschitz near *F*(*x*). Then the function $f(x') \coloneqq g(F(x'))$ is Lipschitz near *x*, and we have

$$\partial f(x) \subset \overline{co}^{*}\{\partial\langle \gamma, F(\cdot)\rangle(x) : \gamma \epsilon \partial g\left(F(x)\right)\},$$

where $\overline{co}^{*}$ signifies the w*-closed convex hull.

Proof: It is an inclusion between two convex weak*-compact sets that is at issue; the corresponding support function inequality amounts to the statement that for given *v*, there exists γ in $\partial g\left(F(x)\right)$, and ζ in the generalized gradient at *x* of the function $x' \mapsto \langle \gamma, F(x')\rangle$, such that $f^{\circ}(x; v) \leq \langle \zeta, v\rangle$. We will prove the theorem by producing such a pair ζ and ζ.

To begin, we give ourselves sequences $y_i \longrightarrow x$ and $t_i \downarrow 0$ realizing the limsup in the definition of $f^{\circ}(x; v)$; i.e., such that

$$\lim_{i \longrightarrow \infty} \frac{f(y_i + t_i v) - f(y_i)}{t_i} = f^{\circ}(x; v).$$

Applying the Mean Value Theorem 2.2.4 gives, for each *i*, an element $\gamma_i \in \partial g(z_i)$ such that

$$\frac{f(y_i + t_i v) - f(y_i)}{t_i} = \frac{g(F(y_i + t_i v)) - g(F(y_i))}{t_i} = \langle \gamma_i, \frac{F(y_i + t_i v) - F(y_i)}{t_i}\rangle$$

where z_i lies on the line segment joining $F(y_i)$ and $F(y_i + t_i v)$. It follows that $z_i \to F(x)$, and that for a suitable subsequence we have $\gamma_i \to \gamma \in \partial g\big(F(x)\big)$ (we eschew rebaleling). This is the required γ; we turn now to exhibiting ζ.

By the Mean Value Theorem again, there exists $\zeta_i \in \partial\langle \gamma, F(\cdot)\rangle(w_i)$ such that

$$\langle \gamma, \frac{F(y_i + t_i v) - F(y_i)}{t_i}\rangle = \langle \zeta_i, v\rangle,$$

where w_i is on the line segment joining y_i and $y_i + t_i v$. It follows that $w_i \to x$, that the sequence $\{\zeta_i\}$ is bounded in X^*, and that $\{\langle \zeta_i, v\rangle\}$ is bounded in $\mathbb{R}$. We may pass again to a subsequence to arrange for $\langle \zeta_i, v\rangle$ to converge to some limit; having done so, let ζ be a weak*-cluster point of $\{\zeta_i\}$. Then $\langle \zeta_i, v\rangle \longrightarrow \langle \zeta, v\rangle$ necessarily, and $\zeta \in \partial\langle \gamma, F(\cdot)\rangle(x)$ (Proposition 2.1.4(d)).

Combining the above, we arrive at

$$\frac{f(y_i + t_i v) - f(y_i)}{t_i} = \langle (\gamma_i - \gamma) + \gamma, \frac{F(y_i + t_i v) - F(y_i)}{t_i}\rangle$$

$$= \langle \gamma_i - \gamma, \frac{F(y_i + t_i v) - F(y_i)}{t_i}\rangle + \langle \zeta_i, v\rangle.$$

Now the term

$$\frac{F(y_i + t_i v) - F(y_i)}{t_i}$$

is bounded because F is Lipschitz, and we know $\gamma_i \to \gamma$. Therefore passing to the limit yields

$$f^{\circ}(x; v) = \lim_{i \to \infty} \frac{f(y_i + t_i v) - f(y_i)}{t_i} = \langle \zeta, v\rangle,$$

which confirms that ζ has the required properties.

2.3 Relation to Derivatives

We remind the reader that some basic definitions and facts about classical differentiability. (These carry over to the present banach space setting when the $\langle \cdot, \cdot \rangle$ is given the duality pairing interpretation.)

Proposition 2.3.1: Let f be Lipschitz near x.

(a) If f admits a Gateaux derivative $f_G'(x)$ at x, then $f_G'(x) \in \partial f(x)$.

(b) If f is continuously differentiable at x, then $\partial f(x) = \{f'(x)\}$.

Proof: By definition we have the following relation between $f_G'(x)$ and the one-sided directional derivatives:

$$f'(x; v) = \langle f_G'(x), v \rangle \ \forall v \in \mathbb{R}^n.$$

But clearly, $f'(x; v) \leq f^{\circ}(x; v)$. That $f_G'(x)$ belongs to $\partial f(x)$ now follows from Proposition 2.1.4(c).

Now suppose that *f* is C^1 in a neighboorhood of *x*, and fix $v \in X$. For *y* near *x* and $t > 0$ near 0, we have

$$\frac{f(y + tv) - f(y)}{t} = \langle f'(z), v \rangle$$

for some $z \in (y, y + tv)$, by the classical Mean Value Theorem. As $y \to x$ and $t \downarrow 0$, the point *z* converges to *x*, and because $f'(\cdot)$ is continuous (as a map between the Banach Spaces *X* and X^*), we derive $f^{\circ}(x; v) \leq \langle f'(x) \rangle$. It follows now from Proposition 2.1.4(c) that$\langle \zeta, v \rangle \leq \langle f'(x), v \rangle$ whenever $\zeta \in \partial f(x)$. Since v is arbitary, we conclude that $\partial f(x)$ is the singleton $\{f'(x)\}$.

2.4 Convex and Regular Functions

A real valued function *f* defined on an open convex subset *U* of *X* is termed convex provided that for any two points $x, y \in U$ we have

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) \ \forall t \in [0,1].$$

Proposition 2.4.1: If *f* is a convex function on *U* that is bounded above on a neighboorhood of some point in *U*, then for any *x* in *U*, *f* is Lipschitz near *x*.

Proposition 2.4.2: Let *f* be convex on *U* and Lipschitz near $x \in U$. Then the directional derivatives $f'(x; v)$ exist, and we have $f'(x; v) = f^{\circ}(x; v)$. A vector ζ belongs to $\partial f(x)$ iff

$$f(y) - f(x) \geq \langle \zeta, y - x \rangle \ \ \forall y \in U.$$

Proof: It follows directly from the definition of convex function that for small $t > 0$, the function

$$t \mapsto \frac{f(x' + tv) - f(x')}{t}$$

is nondecreasing. This fact, together with the Lipschitz hypothesis, implies the existence and finiteness of the directional derivative for all x' near x, for all v:

Now fix $\delta > 0$, and observe that $f^{\circ}(x; v)$ can be written as

$$f^{\circ}(x; v) = \lim_{\varepsilon \downarrow 0} \sup_{\|x' - x\| \le \varepsilon\delta} \sup_{0 < t < \varepsilon} \frac{f(x' + tv) - f(x')}{t}.$$

The preceding remarks show that an alternative expression for $f^{\circ}(x; v)$ is

$$\lim_{\varepsilon \downarrow 0} \sup_{\|x' - x\| \le \varepsilon\delta} \frac{f(x' + \varepsilon v) - f(x')}{\varepsilon}.$$

If K is a Lipschitz constant for f near x, then for all x' in $B(x; \varepsilon\delta B)$, for all ε sufficiently small, we have

$$\left| \frac{f(x' + \varepsilon v) - f(x')}{\varepsilon} - \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} \right| \le 2\delta K,$$

so that

$$f^{\circ}(x; v) \le \lim_{\varepsilon \downarrow 0} \left\{ \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} + 2\delta K \right\} = f'(x; v) + 2\delta K.$$

Since δ is the arbitrary we deduce $f^{\circ}(x; v) \le f'(x; v)$, and hence equality, since $f^{\circ} \le f'$ inherently. Finally, we observe

$$\begin{aligned} \zeta \in \partial f(x) &\Leftrightarrow f^{\circ}(x; v) \ge \langle \zeta, v \rangle \ \forall v, \\ &\Leftrightarrow f'(x; v) \ge \langle \zeta, v \rangle \ \forall v, \\ &\Leftrightarrow \inf_{t > 0} \frac{f(x + tv) - f(x)}{t} \ge \langle \zeta, v \rangle \ \forall v, \\ &\Leftrightarrow f(y) - f(x) \ge \langle \zeta, y - x \rangle \ \forall y \in U. \end{aligned}$$

It turns out that the property of having directional derivatives $f'(x; v)$ that coincide with $f^{\circ}(x; v)$ is precisely what is required to make our calculus rules more exact. We give this property a name; the function f is regular at x provided that f is Lipschitz near x and admits directional derivatives $f'(x; v)$ at x for all v, with

$$f'(x; v) = f^{\circ}(x; v).$$

Evidently, functions which are continuously differentiable at x are regular at x, since then $f'(x; v) = \langle f'(x), v \rangle = f^{\circ}(x; v)$. Also, convex functions which are Lipschitz near x are regular there, by the preceding proposition.

Exercise 2.4.3: Give an example of a function which is neither C^1 nor convex near x, but which is regular at x.

Let us now illustrate how regularity sharpens certain calculus rules, such as that for the sum of two functions. If f and g are Lipschitz near x, we know (Proposition 2.2.2) that

$$\partial(f+g)(x) \subset \partial f(x) + \partial g(x).$$

Suppose now that f and g are regular at x. Then we can argue as follows to get the opposite inclusion: for any v,

$$\begin{aligned} &max\{\langle \zeta + \xi, v\rangle : \zeta \in \partial f(x), \xi \in \partial g(x)\} \\ &= f^{\circ}(x;v) + g^{\circ}(x;v) \\ &= f^{'}(x;v) + g^{'}(x;v) \\ &= (f+g)^{'}(x;v) \leq (f+g)^{\circ}(x;v) \\ &= max\{\langle \zeta, v\rangle : \zeta \in \partial(f+g)(x)\}. \end{aligned}$$

This inequality between support functions is equivalent to the inclusion

$$\partial f(x) + \partial g(x) \subset \partial(f+g)(x),$$

so that equailty actually holds. A bonus consequence of this argument is the fact that $(f+g)^{'}(x;\cdot)$ and $(f+g)^{\circ}(x;\cdot)$ coincide, so that $f+g$ inherits regularity from f and g. In fact, it is clear that any (finite) nonnegative linear combination of regular functions is regular.

The following theorem subsumes the case of a finite sum just discussed. The setting is that of the Chain Rule 2.2.6, of which this is a refinement.

Theorem 2.4.4: Let $F: X \to \mathbb{R}^n$ be such that each component function f_i of F is regular at x. Let $g: \mathbb{R}^n \to \mathbb{R}$ be regular at $F(x)$, and suppose that each $\gamma \in \partial g(F(x))$ has nonnegative components. Then the function $f(x^{'}) := g(F(x^{'}))$ is regular at x, and we have

$$\partial f(x) = \overline{co}^{*}\{\partial\langle \gamma, F(\cdot)\rangle(x) : \gamma \in \partial g(F(x))\}.$$

Proof: We ask the reader to check as a first step that f admits directional derivatives at x:

$$f^{'}(x;v) = g^{'}\Big(F(x); F^{'}(x;v)\Big),$$

where $F^{'}(x;v)$ signifies the vector in $\mathbb{R}^n$ whose ith component is $f_i(x;v)$. Now consider, for given $v \in X$, the maximum of the inner product of v taken with elements from the right side of the equality asserted by the theorem. That maximum equals

$$\max\{\langle\gamma, F(\cdot)\rangle^{\circ}(x; v): \gamma \in \partial g(F(x))\}$$
$$= \max\{\langle\gamma, F(\cdot)\rangle'(x; v): \gamma \in \partial g(F(x))\}$$

(for $\langle\gamma, F(\cdot)\rangle$ is regular at x, as a nonnegative linear combination of function regular at x, since γ is nonnegative).

$$= \max\{\langle\gamma, F(\cdot)\rangle^{\circ}(x; v): \gamma \in \partial g(F(x))\}$$
$$= g^{\circ}\left(F(x); F'(x; v)\right)$$
$$= g'\left(F(x); F'(x; v)\right) \qquad \text{(since } g \text{ is regular at } F(x)\text{)}$$
$$= f'(x; v) \qquad \text{(as noted above)}$$
$$\leq f^{\circ}(x; v).$$

But this last term is the support function of the left side, evaluated at v, implying the opposite inclusion to the one furnished by Theorem 2.2.6. It follows that the two sets coincide, and that $f^{\circ}(x; v)$ and $f'(x; v)$ agree.

2.5 Tangents and Normals

Let S be a nonempty closed subset of X. There is a globally Lipschitz function associated with S that completely characterizes it: its distance function $d_S(\cdot)$ is given by

$$d_S(x) := \inf\{\|x - s\|: s \in S\}.$$

We can apply our Lipschitz calculus to $d_S(\cdot)$ in order to define geometric constructs for S. In this light, a rather natural way to define a direction v tangent to S at $x \in S$ is as follows: we require $d_S^{\circ}(x; v) \leq 0$. (That is, d_S should not increase in the direction, as measured by the generalized directional derivative.) We remark that since $d_S^{\circ}(x; v) \geq 0$ for all v (show), it is equivalent to require $d_S^{\circ}(x; v) = 0$. We proceed to adopt this definition: the tangent cone to S at x, denoted $T_S(x)$, is the set of all those $v \in X$ satisfying $d^{\circ}(x; v) \leq 0$.

It is occasionally useful to have the following alternate, direct, characterization of $T_S(x)$ on hand, and reassuring to know that tangency does not depend on the choice of equivalent norms for X (as d_S does):

Proposition 2.5.1: An element v of X is tangent to S at x iff, for every sequence x_i in S converging to x and sequence t_i in $(0, \infty)$ decreasing to 0, there exists a sequence v_i in X converging to v such that $x_i + t_i v_i \in S$ for all i.

Proof: Suppose first that $v \in T_S(x)$, and that sequence $x_i \to x$ (with $x_i \in S, t_i \downarrow 0$ are given. We must produce the sequence v_i alluded to in the statement of the theorem. Since $d_S^{\circ}(x; v) = 0$ by assumption, we have

$$\lim_{i\to\infty} \frac{d_S(x_i + t_i v) - d_S(x_i)}{t_i} = \lim_{i\to\infty} \frac{d_S(x_i + t_i v)}{t_i} = 0.$$

Let s_i be a point in S which satisfies

$$\|x_i + t_i v - s_i\| \leq d_S(x_i + t_i v) + \frac{t_i}{i}$$

and let us set

$$v_i = \frac{s_i - x_i}{t_i}.$$

Then $\|v - v_i\| \to 0$; that is, v_i converges to v. Furthermore,

$$x_i + t_i v_i = s_i \in S,$$

as required.

Now, for the converse. Let v have the stated property concerning sequences, and choose a sequence y_i converging to x and t_i decreasing to 0 such that

$$\lim_{i\to\infty} \frac{d_S(y_i + t_i v) - d_S(y_i)}{t_i} = d_S^{\circ}(x; v).$$

Our purpose is to prove this quantity nonpositive, for then $v \in T_S(x)$ by definition. Let s_i in S satisfy

$$\|s_i - y_i\| \leq d_S(y_i) + \frac{t_i}{i}.$$

It follows that s_i converges to x. Thus there is a sequence v_i converging to v such that $s_i + t_i v_i \in S$. But then, since d_S is Lipschitz of rank 1,

$$\begin{aligned} d_S(y_i + t_i v) &\leq d_S(s_i + t_i v_i) + \|y_i - s_i\| + t_i\|v - v_i\| \\ &\leq d_S(y_i) + t_i\left(\|v - v_i\| + \frac{1}{i}\right). \end{aligned}$$

We deduce that the limit above is nonpositive, which completes the proof.

2.6 The Normal Cone

In the case of classical manifolds in $\mathbb{R}^n$, the tangent space and the normal space are orthogonal to one another. When convex cones are involved, it is polarity that serves to obtain one from the other. We define the normal cone to S at x, denoted $N_S(x)$, as follows:

$$N_S(x) \coloneqq T_S(x)^{\circ} \coloneqq \{\zeta \in X^*: \langle \zeta, v\rangle \leq 0 \; \forall v \in T_S(x)\}.$$

Proposition 2.6.1:

(a) $N_S(x)$ is a w^*-closed convex cone.

(b) $N_S(x) = cl^*\{\cup_{\lambda \geq 0} \lambda\, \partial d_S(x)\}$.

(c) $T_S(x)$ is in turn the polar of $N_S(x)$; that is,

$$T_S(x) = N_S(x)^{\circ} = \{v \in X: \langle \zeta, x \rangle \leq 0\ \forall \zeta \in N_S(x)\}.$$

Proof: Property (a) is immediate. Let $\zeta \in \partial d_S(x)$, and suppose that $v \in T_S(x)$. Since $d_S^{\circ}(x; v) \leq 0$ by definition of $T_S(x)$, and since $d_S^{\circ}(x; \cdot)$ is the support function of $\partial d_S(x)$, we deduce $\langle \zeta, v \rangle \leq 0$. This shows that $\partial d_S(x)$ lies in $N_S(x)$, which implies that the set Σ appearing on the right in (b) is contained in $N_S(x)$. To complete the proof of (b), let ζ be a point in the complement of Σ. By the Seperation Theorem, there exists $v \in X$ such that

$$H_\Sigma(v) < \langle \zeta, v \rangle.$$

It follows that $\langle \zeta, v \rangle > 0$ and $H_\Sigma(v) \leq 0$, since Σ is a cone. Therefore $\langle v, \theta \rangle \leq 0\ \forall \theta \in \partial d_S(x)$, whence $d_S^{\circ}(x; v) \leq 0$. We conclude that $v \in T_S(x)$. Since $\langle \zeta, v \rangle > 0$, it follows that $\zeta \notin N_S(x)$; (b) is proven.

We turn now to the proof of (c). Let $v \in T_S(x)$. Then $\langle \zeta, v \rangle \leq 0\ \forall \zeta \in \partial d_S(x)$, which implies $\langle \zeta, v \rangle \leq 0\ \forall \zeta \in N_S(x)$ in view of (b). Thus $v \in N_S(x)^{\circ}$. Conversely, let $v \in N_S(x)^{\circ}$. Then $\langle \zeta, v \rangle \leq 0\ \forall \zeta \in \partial d_S(x)$, because of (b). But then $d_S^{\circ}(x; v) \leq 0$ and $v \in T_S(x)$.

We postpone the proof of the fact that T_S and N_S coincide with the classical tangent and normal spaces when S is a smooth manifold. An other special case of interest is the convex one, which we now examine.

Proposition 2.6.2: Let S be convex. Then

$$T_S(x) = \mathrm{cl}\{\lambda\,(s - x): \lambda \geq 0, s \in S\}$$

and

$$N_S(x) = \{\zeta \in X^*: \langle \zeta, x' - x \rangle \leq 0\ \forall x' \in S\}.$$

Proof: The convexity of the set S readily implies that of the function $d_S(\cdot)$. It follows then from Proposition 2.4.2 that $d_S'(x; v)$ exists and coincides with $d_S^{\circ}(x; v)$. Consequently, $T_S(x)$ consists of those $v \in X$ for which

$$\lim_{t \downarrow 0} \frac{d_S(x + tv)}{t} = 0.$$

This in turn is equivalent to the existence of $s(t) \in S$ such that

$$\frac{\|x + tv - s(t)\|}{t} \to 0 \ \text{as}\ t \downarrow 0.$$

Setting $u(t) := x + tv - s(t)/t,$ this can be expressed in the form

$$v = \left(\frac{1}{t}\right)(s(t) - x) + u(t),$$

where $u(t) \to 0$ as $t \downarrow 0$. This is equivalent to the characterization of $T_S(x)$ given in the statement of the proposition. The expression for $N_S(x)$ then follows immediately from this characterization, together with the fact that $N_S(x)$ is the polar of $T_S(x)$.

When *S* is the epigraph of a function, we would expect some relationship to exist between its tangent and normal cones on the one hand, and the generalized gradient of the function on the other. In fact, a complete duality exists, as we now see.

Theorem 2.6.3: Let f be lipschitz near *x*. Then:

(a) $T_{epif}\big(x, f(x)\big) = \text{epi} f^{\circ}(x;\cdot)$; and

(b) $\zeta \in \partial f(x) \Leftrightarrow (\zeta, -1) \in N_{epif}\big(x, f(x)\big)$.

Proof: Suppose first that (v, r) lies in $T_{epif}\big(x, f(x)\big)$. Choose sequences $y_i \to x, t_i \downarrow 0$, such that

$$\lim_{i\to\infty} \frac{f(y_i + t_i v) - f(y_i)}{t_i} = f^{\circ}(x; v).$$

Note that $(y_i, f(y_i))$ is a sequence in epi$f$ converging to $(x, f(x))$. Accordingly, by Proposition 2.5.1, there exists a sequence (v_i, r_i) converging to (v, r) such that $(y_i, f(y_i)) + t_i(v_i, r_i) \in \text{epi}\, f$. Thus

$$f(y_i) + t_i r_i \geq f(y_i + t_i v_i).$$

We rewrite this as

$$\frac{f(y_i + t_i v_i) - f(y_i)}{t_i} \leq r_i.$$

Taking limits, we obtain $f^{\circ}(x; v) \leq r$ as desired.

We now show that for any v, for any $\delta \geq 0$, the point $(v, f^{\circ}(x; v) + \delta)$ lies in $T_{\text{epi}f}\big(x, f(x)\big)$; this will complete the proof of (a). Let (x_i, r_i) be any sequence in epi$f$ converging to $\big(x, f(x)\big)$, and let $t_i \downarrow 0$. We must produce a sequence (v_i, s_i) converging to $(v, f^{\circ}(x; v) + \delta)$ with the property that $(x_i, r_i) + t_i(v_i, s_i)$ lies in epi$f$ for each i; that is, such that $r_i + t_i s_i \geq f(x_i + t_i v_i)$.

Let us define $v_i = v$ and

$$s_i \coloneqq \max\left\{f^{\circ}(x;v) + \delta, \frac{f(x_i + t_i v) - f(x_i)}{t_i}\right\}.$$

Observe first that $s_i \to f^{\circ}(x;v) + \delta$, since

$$\limsup_{i\to\infty} \frac{f(x_i + t_i v) - f(x_i)}{t_i} \le f^{\circ}(x;v).$$

We have

$$r_i + t_i s_i \ge r_i + [f(x_i + t_i v) - f(x_i)]$$

and $r_i \ge f(x_i)$ (since $(x_i, r_i) \in \operatorname{epi} f$), which together give

$$r_i + t_i s_i \ge f(x_i + t_i v)$$

showing that $(x_i + t_i v, r_i + t_i s_i)$ belongs to epi f, as required.

We turn now to (b). We know that $\delta \in \partial f(x)$ iff $f^{\circ}(x;v) \ge \langle \zeta, v\rangle\ \forall v$; that is, precisely when for any v and any $r \ge f^{\circ}(x;v)$ we have

$$\langle (\zeta, -1), (v, r)\rangle \le 0.$$

By (a), this last inequality holds for all the (v, r) in question iff it holds for all $(v, r) \in T_{\operatorname{epi} f}(x, f(x))$; that is, precisely when $(\zeta, -1)$ lies in the polar of $T_{\operatorname{epi} f}(x, f(x))$, namely $N_{\operatorname{epi} f}(x, f(x))$.

CHAPTER 3

ONE CLASS NONSMOOTH DISCRETE STEP CONTROL PROBLEM

Example 3.1: (see reference [6])

A car moves according to the law $\dot{x} = y,\ \dot{y} = ug_1(y),\ \ u \in U$ at the time interval $\Delta_1 = [t_0, t_1]$, and according to $\dot{x} = y,\ \dot{y} = ug_2(y),\ u \in \sigma(y(t_1))$ at the time interval $\Delta_2 = [t_1, T]$. The initial and final time moments t_0 and T are fixed while instant t_1 is not fixed. The set $U = [0,1]$ and the functions $g_1,\ g_2, \sigma$ are positive and differentiable in R^1. The car starts from the origin $(x^0, y^0) = (0,0)$. The state variables x and y are assumed to be continuous on the whole interval $\Delta = [0, T]$. It is required to maximize $x(T)$. To find the necessary optimality conditions, we have to build Hamilton-Pontryagin functions for each step and derive the optimality condition at the switching moment t_1 using steps Δ_1 and Δ_2. In this example, switching moments are interesting for us because at the switching point we have to derive the optimality condition. By using increment formula and conjugate systems we can get the necessary condition for this step control system.

Example 3.2:

Consider a rocket with two types of engines that work consecutively. The work of the second engine depends on the first one. Moreover, the rocket moves from one controlling area to a second one that changes all the structure (controls, functions, conditions, etc.). For the smooth case, some articles were published previously[14,31,6,8,4,32]. In [14,16,19,17] the authors had gained the necessary optimality condition of first order and investigated singular control, time with delay and sufficient optimality condition as a Krotov type for discrete switching optimal control problem.

In [31], the author does not make any assumptions about the number of switches, nor about the mode sequence. They simply are determined by the solution of the problem. Sufficient and necessary optimality conditions for optimality are formulated for the second optimization problem. If they exist, bang-bang-type solutions of the embedded optimal control problem are solutions of the original problem. Otherwise, suboptimal solutions are obtained via the Chattering lemma by the author. In [4] the author develops a computational method for solving an optimal control problem which is governed by a switched dynamical time system with time delay. Then, we derive the required gradient of the cost function which is obtained via solving a number of delay differential equation forward in time. On this basis author solved this problem as a mathematical programming. All this results dedicated in the smooth case optimal switching control problem (in all these papers the cost functional is smooth). In the present paper, the author's main aim is to formulate necessary optimality conditions for nonsmooth case and the switching points which depend on certain parameters, by using the Frechet superdifferential. (see, e.g.,[26,27,28]). To start our discussion, first we have to describe certain points about nonsmooth analysis.

3.1 Tools of Nonsmooth Analysis

If φ_k is lower semicontinuous around *x*, then its basic subdifferential can be shown by:

$$\partial\varphi(x) = \limsup_{x \xrightarrow{\varphi} \bar{x}} \hat{\partial}\varphi(x).$$

Here,

$$\hat{\partial}\varphi\ (x) := \left\{ x^* \in R^n \mid \liminf_{u \to x} \frac{\varphi(u) - \varphi(x) - \langle x^*, u - x \rangle}{|u - x|} \geq 0 \right\}$$

is the Frechet subdifferential. By using plus-minus symmetric constructions, we can write

$$\partial^+\varphi(x) := -\partial(-\varphi)(x),\ \hat{\partial}^+\varphi(x) := -\hat{\partial}(-\varphi)(x)$$

where ∂^+ denotes a basic superdifferential and $\hat{\partial}^+$ denotes a Frechet superdifferential.

Here

$$\hat{\partial}\varphi^+(x) := \left\{ x^* \in R^n \mid \limsup_{u \to x} f \frac{\varphi(u) - \varphi(x) - \langle x^*, u - x \rangle}{|u - x|} \leq 0 \right\}$$

For a Locally Lipschitzian function, the subdifferential and superdifferential may be different. For example, if we take $\varphi(x) = |x|$ on R, then $\partial\varphi(0) = [-1,1]$, while $\partial^+\varphi(0) = \{-1,1\}$.

We can give upper regularity of the function at the point by using definitions of superdifferential and Frechet superdifferential. Also if the extended-real-valued function is Lipschitz continuous around the given point and upper regular at this point then the Frechet superdifferential is not empty.

Definition 3.1.1: φ is upper regular at $\bar{x}$ if $\partial^{+}\varphi(\bar{x}) = \hat{\partial}^{+}\varphi(\bar{x})$**,** If $\partial\varphi(x^{0}) = \hat{\partial}\varphi(x^{0})$ then, this function lower regular at x^{0}.

Proposition 3.1.2: Let $\varphi : R^{n} \to \bar{R}$ be Lipschitz continuous around $\bar{x}$ and upper regular at this point.Then

$$0 \neq \bar{\partial}^{+}\varphi(\bar{x}) = \bar{\partial}\varphi(\bar{x})$$

It is distance between the point x and set Ω, when $x \in \Omega$

$$dist(x;\Omega) = \inf_{u\in\Omega}\|x - w\|$$

and define the Euclidean projector of x to Ω by:

$$\Pi(x;\Omega) := \{w \in \Omega \mid \|x - w\| = dist(x;\Omega\}.$$

If the set Ω is closed, then the set $\Pi(x;\Omega)$ is nonempty for every $x \in R^{n}$.

This nonconvex cone to closed sets and corresponding subdifferential of lower semicontinuous extended–real –valued functions satisfying these requirements were introduced by Mordukhovich in the beginning of 1975. The initial motivation came from the intention to derive necessary optimality conditions for optimal control problems with endpoint geometric constraints by passing to the limit from free endpoint control problems, which are much easier to handle. This was published in [25] (first in Russian and then translated into English), where the original normal cone definition was given in finite dimensional spaces by:

$$N(x;\Omega) := \limsup_{x\to\bar{x}}\left[cone\left(x - \Pi\left(x;\Omega\right)\right)\right],$$

via the Euclidean projector, while the basic subdifferential $\partial\varphi(\bar{x})$ was defined geometrically via the normal cone to the epigraph of φ. Here it is assumed that φ is a real valued finite function and the basic subdifferential is defined as:

$$\partial\varphi(\bar{x}) := \{x^{*} \in R^{n} (x^{*},-1) \in N((\bar{x},\varphi(\bar{x})); epi\varphi)\}.$$

Here $epi\varphi := \{(x,\mu) \in R^{n+1} | \mu \geq \varphi(x)\}$ and is called the epigraph of a given extended real valued function. Note that this cone is nonconvex (see, ref[26,27,28]) and for the locally Lipschitzian functions, the convex hull of a subdifferential has a Clarke generalized gradient, $\overline{\varphi}_k(x^0) = co\partial\varphi(x^0)$.

Furthermore it is equal to Clarke generalized subdifferential at this point (for proof, see [3]). By using all these nonsmooth analysis tools, we will try to find the superdifferential form of the necessary optimality condition for the step discrete system.

3.2 Necessary optimality condition

Consider a controlling process, which is described by the following discrete system with varying structure:

minimize: $$S(u,v) = \sum_{i=1}^{3} \varphi_i(x_i(t_i)) \tag{3.1}$$

subject to:

$$x_i(t+1) = f_i(t, x_i(t), u_i(t)),\ t \in T_i = \{t_{i-1}, t_{i-1}+1, ..., t_i - 1\},\ i = 1,2,3, \tag{3.2}$$

$$\begin{cases} x_1(t_0) = g_1(v_1) \\ x_i(t_{i-1}) = g_i(x_{i-1}(t_{i-1}), v_i), \quad i = 2,3 \end{cases} \tag{3.3}$$

$$u_i(t) \in U_i \subset R^r,\ t \in T_i,\ i = 1,2,3. \tag{3.4}$$

$$v_i \in V_i,\ i = 1,2,3. \tag{3.5}$$

Here $v_i, i = 1,2,3$ are q-dimensional controlling parameters and $V_i \subseteq R^q$, $i = 1,2,3$, i.e. $v_i \in V_i$, $i = 1,2,3$.

For these equations it is clear that the system's conditions are described in 3 stages (for a rocket entering from space to the atmosphere and then into water). In any stage, the system is described by its equation, controls, *switching points*, and controlling parameters for switching points. In case, there is no switching point, we can apply the Pontryagin's maximum principle to any part of the system, but in this case it is difficult. For this we have to get new conditions of the switching points. In this problem, $g_1 : R^q \to R^n$ are assumed to be at least twice continuously differentiable vector-valued functions, $g_i : R^n \times R^q \to R^n$ are given at least twice continuously differentiable vector-valued functions, i = 2,3, $f_i : R \times R^n \times R^r \to R^n$ are given continuous, at least twice continuously partially differentiable vector-valued functions with respect to *x*, $\varphi_i : R^n \to R$ are given at least functions. We do not assume any smoothness on the cost functional $\varphi_i \quad i = 1,2,3$, u_i(t): *R* → $U_i \subset R^r$ are controls and $v_i \in V_i \subset R^q$ are controlling *parameters*. The sets U_i, V_i, are assumed to be nonempty and bounded. The pair $\left(u_i^{\,0}(t), v_i^{\,o}\right)$ which takes its volume from these sets is called an admissible control. A pair $\left(u_i^{\,0}(t), v_i^{\,o}\right)$ with the properties (3.4) and (3.5) is called admissible. The triple $\left(u_i^{\,0}(t), v_i^{\,o}, x_i^0(t)\right)$ is an admissible process. For the fixed admissible control $\left(u_i^{\,0}(t), v_i^{\,o}\right)$ we introduce the following notation:

$$H_i\left(t, x_i, u, \Psi_i^0\right) = \Psi_i^{0'}(t) \cdot f_i\left(t, x_i, u_i\right),$$

$$\Delta_{u_i} H_i[t] \equiv H_i\left(t, x_i^0(t), u_i(t), \Psi_i^0(t)\right) - H_i\left(t, x_i^0(t), u_i^0(t), \Psi_i^0(t)\right),$$

$$\frac{\partial H_i[t]}{\partial x_i} = \frac{\partial H_i\left(t, x_i^0(t), u_i^0(t), \psi_i^0(t)\right)}{\partial x_i}, \ \Delta_{v_1} g_1[v_1] \equiv g_1(v_1) - g_1\left(v_1^0\right)$$

$$\Delta_{v_i} g_i\left(x_{i-1}^0(t_{i-1}), v_i^0\right) \equiv g_i\left(x_{i-1}^0(t_{i-1}), v_i\right) - g_i\left(x_{i-1}^0(t_{i-1}), v_i^0\right), i = 2,3$$

$$L_1\left(v_1, \Psi_1^0(t_0 - 1)\right) = \Psi_1^{0'}(t_0 - 1) g_1(v_1),$$

$$L_2\left(x_1(t_1), v_2, \Psi_2^0(t_1 - 1)\right) = \Psi_2^{0'}(t_1 - 1) g_2\left(x_1(t_1), v_2\right)$$

$$L_3\left(x_2(t_2), v_3, \Psi_3^0(t_2 - 1)\right) = \Psi_3^{0'}(t_2 - 1) g_3\left(x_2(t_2), v_3\right).$$

Theorem 3.2.1: Assume that $\varphi_i : R^n \to R$ is finite at $x_i^{\,0}(t_i)$ and $\hat{\partial}\varphi\left(x^0(t_i) \neq 0\right)$.

If the sets

$$f_i\left(t, x_i^0(t), U_i\right) = \left\{\alpha_i : \alpha_i = f_i\left(t, x_i^0(t), u_i\right), u_i \in U\right\},\ i = 1,2,3$$

$$g_1(V_1) = \{\alpha_4 : \alpha_4 = g_1(v_1), v_1 \in V_1\}$$

$$g_i\left(x_{i-1}^0(t_{i-1}), V_i\right) = \left\{\alpha_i : \alpha_i = g_i\left(x_{i-1}^0(t_{i-1}), v_i\right), v_i \in V\right\},\quad i = 2,3$$

are convex, then for the optimality of an admissible control $\left(u^0(t), v^0\right)$ in the problem described in (3.1)-(3.5) it is necessary that for any $x_i^* \in \hat{\partial}\varphi\left(x^0(t_i)\right)$ the following conditions are true:

Discrete maximum principle for the control

$$\sum_{t=t_{i-1}}^{t_i - 1} \Delta_{u_i(t)} H_i[t] \leq 0,\ \text{for all}\ u_i(t) \in U_i,\quad i = 1,2,3,\ t \in T_i \tag{3.6}$$

Discrete maximum principle for the controlling parameter $v_i^0,\quad i = 1,2,3$

$$\max_{v_1 \in V_{1i}} L_1\left(v_1, \psi_1^0(t_0 - 1)\right) = L_1\left(v_1^0, \psi_1^0(t_0 - 1)\right) \tag{3.7}$$

$$\max_{v_i \in V_i} L_i\left(x_{i-1}^0(t_{i-1}), v_i, \psi_i^0(t_{i-1} - 1)\right) = L_i\left(x_{i-1}^0(t_{i-1}), v_i^0, \psi_i^0(t_{i-1} - 1)\right), i = 2,3 \tag{3.8}$$

where $\psi(.)$ is an adjoint trajectory which satisfies equation (3.11). If the set $f_i\left(t, x^0, U\right)$ is convex, then the necessary optimality condition is global over all $u_i \in U_i$.

Proof: In the control problem, one of the methods to get the necessary optimality conditions is to use the increment formula. For this, we have to calculate the increment formula, to find a conjugate system for the corresponding problems and use an analog of needle variations in the continuous case. The rest of the increment formula can be estimated using the step method.

For the optimal pair $\left(u^0(t), v^0\right)$ we can write increment of the functional following form

$$\Delta S\left(u^0, v^0\right) = \sum_{i=1}^{3}\left[\varphi_i\left(x_i(t_i)\right) - \varphi_i\left(x_i^0(t_i)\right)\right] \geq 0 .$$

Using nonsmooth analysis tools we can write that, for any $x_i^* \in \hat{\partial}^+\varphi\left(x_i^0(t_i)\right)$ we can write

$$\varphi_i\left(x_i(t_i)\right) - \varphi_i\left(x_i^0(t_i)\right) \leq \left\langle x_i^*, (\Delta x_i^0(t)\right\rangle + 0\left(\Delta x_i^0(t)\right)$$

Then, the increment of the functional takes the following form:

$$\Delta S\left(u^0, v^0\right) = \sum_{i=1}^{3}\left\langle x_i^*, \Delta x_i^0(t)\right\rangle + 0\left(\Delta x_i^0(t)\right).$$

Let us multiply both sides of equation (3.2) by $\psi_i(t)$ and sum it up from $i = 1$ to 3. By using this sum, the definition of nonsmooth analysis, and Taylor's increment formula, after some calculations we can write the increment of the functional at an arbitrary admissible pair $\left(u_i(t), v_i\right)$ as:

$$\Delta S\left(u^0, v^0\right) = \sum_{i=1}^{3}\left[\left\langle x_i^*, \Delta x_i(t)\right\rangle\right] + \sum_{i=1}^{3}\sum_{t=t_{i-1}}^{t_i-1}\psi_i^{0'}(t-1)\Delta x_i(t) - \sum_{i=1}^{3}\sum_{t=t_{i-1}}^{t_i-1}\left[H_i\left(t, \bar{x}_i(t), \bar{u}_i(t), \psi_i^0(t)\right) -\right.$$

$$\left. - H_i\left(t, x_i^0(t), u_i^0(t), \psi_i^0(t)\right)\right] + \sum_{i=1}^{3}\psi_i^{0'}(t_i - 1)\Delta x_i(t_i) - \psi_1^{0'}(t_0 - 1)\Delta_{\bar{v}_1} g_1\left(v_1^0\right) - \psi_2^{0'}(t_1 - 1) \times$$

$$\times\left[g_2\left(\bar{x}_1(t_1), \bar{v}_2\right) - g_2\left(x_1^0(t_1), v_2^0\right)\right] - \psi_3^{0'}(t_2 - 1)\left[g_3\left(\bar{x}_2(t_2), \bar{v}_3\right) - g_2\left(x_2^0(t_2), v_3^0\right)\right] =$$

$$= \sum_{i=1}^{3}\left[\varphi_i\left(\bar{x}_i(t_i)\right) - \varphi_i\left(x_i^0(t)\right)\right] + \sum_{i=1}^{3}\sum_{t=t_{i-1}}^{t_i-1}\psi_i^{0'}(t-1)\Delta x_i(t) - \sum_{i=1}^{3}\sum_{t=t_{i-1}}^{t_i-1}\Delta_{\bar{u}_i} H_i[t] -$$

$$- \sum_{i=1}^{3}\sum_{t=t_{i-1}}^{t_i-1}\left[H_i\left(t, \bar{x}_i(t), \bar{u}_i(t), \psi_i^0(t)\right) - H_i\left(t, x_i(t), \bar{u}_i(t), \psi_i^0(t)\right)\right] + \sum_{i=1}^{3}\psi_i^{0'}(t_i - 1)\Delta x_i(t_i) -$$

$$- \Delta_{\bar{v}_1} L_1\left(v_1^0, \psi_1^0(t-1)\right) - \Delta_{\bar{v}_2} L_2\left(x_1^0(t), v_2^0, \psi_2^0(t_1 - 1)\right) - \Delta_{\bar{v}_3} L_3\left(x_2^0(t_2), v_3^0, \psi_3^0(t_2 - 1)\right) -$$

$$- \left[L_2\left(\bar{x}_1(t_1), \bar{v}_2, \psi_2^0(t_1 - 1)\right) - L_2\left(x_1^0(t_1), \bar{v}_2, \psi_2^0(t_1 - 1)\right)\right] - \left[L_3\left(\bar{x}_2(t_2), \bar{v}_3, \psi_3^0(t_2 - 1)\right) -\right.$$

$$\left. - L_3\left(x_2^0(t_2), \bar{v}_3, \psi_3^0(t_2 - 1)\right)\right]. \tag{3.9}$$

where by definition:

$$\eta_1\left(u^0, v^0; \Delta u, \Delta v\right) = \sum_{i=1}^{3} \sum_{t=t_{i-1}}^{t_{i-1}} \frac{\partial \Delta_{\overline{u}_i} H'[t]}{\partial x} \Delta x_i(t) - o_3\left(\left\|\Delta x_1(t_1)\right\|\right) - o_4\left(\left\|\Delta x_2(t_2)\right\|\right) -$$

$$-\frac{\partial \Delta_{\overline{v}_2} L_2\left(x_1^0(t_1), v_2^0, \psi_2^0(t_1-1)\right)}{\partial x_1} \Delta x_1(t_1) - \frac{\partial \Delta_{\overline{v}_3} L_3\left(x_2^0(t_2), v_3, \psi_2^0(t_2-1)\right)}{\partial x_2} \Delta x_2(t_2) +$$

$$+\sum_{i=1}^{3} o_1^{(i)} \left\|\Delta x_i(t_i)\right\| - \sum_{i=1}^{3} \sum_{t=t_{i-1}}^{t_i-1} o_2^{(i)} \left\|\Delta x_i(t)\right\| . \tag{3.10}$$

Here $o_i(\cdot), \quad i = 1,..,8$ are defined by the expansions:

$$\varphi_i\left(\overline{x}_i(t_i)\right) - \varphi_i\left(x_i(t_i)\right) = \frac{\partial \varphi_i'\left(x_i^0(t_i)\right)}{\partial x_i} \Delta x_i(t_i) + o_1^{(i)}\left(\left\|\Delta x_i(t_i)\right\|\right), \ i = \overline{1,3}$$

$$H_i\left(t, \overline{x}_i(t), \overline{u}_i(t), \psi_i^0(t)\right) - H_i\left(t, x_i^0(t), \overline{u}_i\ (t), \psi_i^0(t)\right) = \frac{\partial H_i'\left(t, x_i^0(t), \overline{u}_i(t), \psi_i^0(t)\right)}{\partial x_i} \times$$

$$\times \Delta x_i(t) + o_2^{(i)}\left(\left\|\Delta_i(t)\right\|\right), \ i = \overline{1,3}$$

$$L_2\left(\overline{x}_1(t_1), \overline{v}_2, \psi_2^0(t_1-1)\right) - L_2\left(x_1^0(t_1), \overline{v}_2, \psi_2^0(t_1-1)\right) = \frac{\partial L_2'\left(x_1^0(t_1), \overline{v}_2, \psi_2^0(t_1-1)\right)}{\partial x_1} \Delta x_1(t_1) +$$

$$+ o_3\left(\left\|\Delta x_1(t_1)\right\|\right),$$

$$L_3\left(\overline{x}_2(t_2), \overline{v}_3, \psi_3^0(t_2-1)\right) - L_3\left(x_2^0(t_2), \overline{v}_3, \psi_3^0(t_2-1)\right) = \frac{\partial L_3'\left(x_2^0(t_2), \overline{v}_3, \psi_3^0(t_2-1)\right)}{\partial x_2} \Delta x_2(t_2) +$$

$$+ o_4\left(\left\|\Delta x_2(t_2)\right\|\right),$$

Now take $\Psi_i^0(t), \quad i=1,2,3,$ as solutions of the following linear difference equations:

$$\begin{cases} \Psi_i^0(t-1) = \dfrac{\partial H_i[t]}{\partial x_i}, \; i=1,2,3, t \in T_i \\ \Psi_1^0(t_1-1) = -x_1^* + \dfrac{\partial L_2\left(x_1^0(t_1), v_2^0, \psi_2^0(t_1-1)\right)}{\partial x_1} \\ \Psi_2^0(t_2-1) = -x_2^* + \dfrac{\partial L_3\left(x_2^0(t_2), v_3^0, \psi_3^0(t_2-1)\right)}{\partial x_1} \\ \Psi_3^0(t_3-1) = -x_3^* \end{cases} \tag{3.11}$$

The increment formula (3.9) reduces to a simpler one:

$$\Delta S(u^0, v^0) = -\sum_{i=1}^{3} \sum_{t=t_{i-1}}^{t_i-1} \Delta_{\bar{u}_i} H[t] - \Delta_{\bar{v}_1} L_1\left(v_1^0, \psi_1^0(t_0-1)\right) -$$

$$- \Delta_{\bar{v}_3} L_3\left(x_2^0(t_2), v_3^0, \psi_3^0(t_2-1)\right) + \eta_1\left(u^0, v^0; \Delta u, \Delta v\right) \tag{3.12}$$

Let $\left(u_i^o(t), v_i^o\right)$ be an optimal pair, and assume that the sets of admissible velocities are convex along the process $\left(u_i(t), v_i, x_i(t)\right)$, i. e., the sets

$$f_i\left(t, x_i^0(t), U_i\right) = \left\{\alpha_i : \alpha_i = f_i\left(t, x_i^0(t), u_i\right), u \in U_i\right\}, \qquad i = 1,2,3$$

$$g_1(V_1) = \left\{\alpha_4 : \alpha_4 = g_1(v_1), v_1 \in V_1\right\},$$

$$g_i\left(x_{i-1}^0(t_{i-1}), V_i\right) = \left\{\alpha_{i+3} : \alpha_{i+3} = g_i\left(x_{i-1}^0(t_{i-1}), v_i\right), v \in V_i\right\}, \qquad i = 2,3,$$

are convex. Let $\varepsilon \in [0,1]$ be an arbitrary number. Denote the increment of the optimal pair by:

$$(\Delta u_i(t;\varepsilon) = u_i(t;\varepsilon) - u_i^0(t), \;\; \Delta v_i(\varepsilon) = v_i(\varepsilon) - v_i^0, \; t \in T_i, \; i = 1, \; 2, \; 3, \tag{3.13}$$

Then, by convexity, for each $u_i(t)\in U_i$, $v_i\in V$, $t\in T_i$, $i=1,2,3$, there are $u_i(t,\varepsilon)\in U_i$, $v_i(\varepsilon)\in V_i$, $i=1,\ 2,\ 3$ such that

$$\Delta_{u_{i(t;\varepsilon)}} f_i[t]=\varepsilon\Delta_{u_{i(t)}} f_i[t],\ i=1,\ 2,\ 3,$$

$$\Delta_{v_{1(\varepsilon)}} g_1\left(v_1^0\right)=\varepsilon\Delta_{v_1} g_1\left(v_1^0\right),$$

$$\Delta_{v_{i(\varepsilon)}} g_i\left(x_{i-1}^0\left(t_{i-1}\right),v_i^0\right)=\varepsilon\Delta_{v_i} g_i\left(x_{i-1}^0\left(t_{i-1}\right),v_i^0\right),\ i=2,\ 3.$$

Equation (3.10) introduces an increment of the solution $x_i(t)$ which is denoted by $\{\Delta x_i(t;\varepsilon), i=1,2,3\}$. Using the step methods, we can prove that $\|\Delta x_i(t;\varepsilon)\|\le Z_{11}\varepsilon$, $t\in T_i\cup t_i$, $i=1,2,3$. Using these estimates in (3.12) it can easily be seen that the necessary optimality condition is $\Delta S\left(u^0,v^0\right)\ge 0$.

Corollary 3.2.2: Assume that φ_i is Lipschitz continuous around at x_i^0, upper regular at this point and the sets

$$f_i\left(t_i x_i^0(t),U_i\right)=\left\{\alpha_i:\alpha_i=f_i\left(t_i x_i^0(t),u_i\right),u_i\in U\right\},\ i=1,2,3$$
$$g_1\left(V_1\right)=\left\{\alpha_4:\alpha_4=g_1\left(v_1\right),v_1\in V_1\right\}$$
$$g_i\left(x_{i-1}^0\left(t_{i-1}\right),V_i\right)=\left\{\alpha_i:\alpha_i=g_i\left(x_{i-1}^0\left(t_{i-1}\right),v_i\right),v_i\in V\right\},\ i=2,3$$

are convex. Then for the optimality of an admissible control $\left(u^0(t),v^0\right)$ in the problem given through (3.1)-(3.5), it is necessary that for any $x_i^*\in\bar{\partial}\varphi\left(x^0\left(t_i\right)\right)$ the following conditions are true:

Discrete maximum principle for the control $u_i^0(t), i=1,2,3$

$$\sum_{t=t_{i-1}}^{t_i-1}\Delta_{u_{i(t)}} H_i[t]\le 0,\qquad \text{for all}\qquad u_i(t)\in U_i, i=1,2,3,\ t\in T_i \tag{3.14}$$

Discrete maximum principle for the controlling parameter $v_i^0, i=1,2,3$

$$\max_{v_i\in V_i} L_i\left(x_{i-1}^0\left(t_{i-1}\right),v_i,\psi_i^0\left(t_{i-1}-1\right)\right)=L_i\left(x_{i-1}^0\left(t_{i-1}\right),v_i^0,\psi_i^0\left(t_{i-1}-1\right)\right), i=2,3$$

$$\max_{v_1\in V_1} L_1\left(v_1,\psi_1^0\left(t_0-1\right)\right)=L_1\left(v_1^0,\psi_1^0\left(t_0-1\right)\right) \tag{3.15}$$

where $\psi(.)$ is adjoin trajectory and satisfies the system described under (3.11).

It is easy to prove this theorem by using the tools of nonsmooth analysis given above. It should be noted that the system of linear difference equations (3.11) is the conjugate system for the problem (3.1)-(3.5). If we take smoothness on the cost functional φ_i then we can get some following corollary and analogies of Pontryagin maximum principle.

3.3 Necessary Optimality Conditions Using the Linearizing Principle as An Analogue of Euler Equation in Nonsmooth Case.

If the cost functional is differentiable, the functions f_i, g_i have also partial derivatives with respect to u_i, v_i, respectively, and the sets U_i and V_i are convex, then another necessary optimality condition can be obtained using the linearizing maximum principle of Pontryagin. The proof of the next following corollaries to a large extent similar to the proof of Theorem 3.2.1 and is omitted. For the proof the interested reader is referred to the thesis [18].

Corollary 3.3.1: (The superdifferential form of linearizing maximum principle).

If the sets U_i, V_i are convex, then, for the optimality of the pair $\left(u^0(t), v^0\right)$, it is necessary that the following inequalities hold:

$$\sum_{t=t_{i-1}}^{t_i-1} \frac{\partial H_i'[t]}{\partial u_i}\left(u_i(t)-u_i^0(t)\right) \le 0 \qquad \text{for all} \quad u_i(t) \in U_i,\ t \in T_i, \quad i = 1,2,3 \tag{3.16}$$

$$\frac{\partial L_1'\left(v_1^0, \psi_1^0(t_0-1)\right)}{\partial v_1}\left(v_1 - v_1^0\right) \le 0, \quad \text{for all} \quad v_i \in V_i \tag{3.17}$$

$$\frac{\partial L_i'\left(x_{i-1}^0(t_{i-1}), v_i^0, \Psi_i^0(t_{i-1}-1)\right)}{\partial v_i}\left(v_i - v_i^0\right) \le 0, \quad \textit{for all } v_i \in V_i, i = 2,3. \tag{3.18}$$

In the case of openness of the sets $U_i, V_i,\ i = 1,2,3$ also using Euler's equation one can derive the necessary optimality conditions:

Corollary 3.3.2: (An analogue of Euler equation): If the sets $U_i\ , V_i$ are open, then for the optimality of the pair $\left(u^0(t), v^0\right)$, it is necessary that the following equations holds:

$$\frac{\partial H_i^{'}[t]}{\partial u_i} = 0, \quad t \in T_i, \quad i = 1,2,3 \tag{3.19}$$

$$\frac{\partial L_1^{'}\left(v_1^0, \psi_1^0(t_0 - 1)\right)}{\partial v_1} = 0 \text{ and} \tag{3.20}$$

$$\frac{\partial L_i^{'}\left(x_{i-1}^0(t_{i-1}), v_i^0, \psi_i^0(t_{i-1} - 1)\right)}{\partial v_i} = 0, \text{ for } i = 2,3 \tag{3.21}$$

CONCLUSION

In this thesis, weak subdifferential which introduced by Azimov and Gazimov [2] is studied. We proved some necessary theorem of the weak subdifferential and got useful properties on it. We also investigate the properties of the generalized gradient in the Clarke means.

One of the principial advantages of the theory of generalized gradients is the complete duality that it induces between tangency and normality, and between functions ans sets (via the epigraph, or via the distance function). Note that in developing the theory, we choose the generalized directional derivative as primitive notion, and used it to define generalized gradient, the tangent cone (via d_S) and then, by polarity, the normal cone.

A further research topic is the development of methods for searching optimality condition for the nonsmooth optimal control problem by using weak subdifferential. Open problems including the existence of the solution, the exploration of the necessary conditions in the nonsmooth case, the solution of the HJB(Hamilton-Jacobi-Belman) equation, the use of numerical methods, etc., still present considerable challenges.

REFERENCES

1. Aris, R. (1964). Discrete Dynamic Programming, Blaisdell, New York.

2. Azimov, A.Y and Gasimov, R.N. (1999). On weak canjugasy, weak subdifferential and duality with zero gap in nonconvex optimization. Int. j. appl. math., (171-192).

3. Bellman, R. (1957). Dynamic Programming, Princeton University Press, Princeton, New Jersey.

4. Changzhi, W., Teo, K., Rui, L. and Zhao, Y. (2006). Optimal control of switching system with time delay. Applied Mathematics Letters 19.1062-1067.

5. Clarke, F.H. (1983). Optimization and Nonsmooth Analysis, Wiley-Interscience, New York, USA, 9-30.

6. D'Apice, C., Garavello, M., Manzo, R., and Piccoli, B. (2003) Hybrid optimal control: case study of a car with gears, International Journal of Control 76, (1272-1284).

7. Dmitruk, A.V. and Kaganovich, A.M. (2008). The Hyprid Maximum Principle is a Consequence of PontryaginMaximum Principle. System and Control Letters., Volume 57, Issue 11, (964-970).

8. Egertedt M., Wardi, Y. and Delmotte, F. (2003). Optimal Control of switching times in switched dynamical systems. Proc.IEEE Conference. Decision Control 2138-2143.

9. Gabasov, R. and Tarasenko, N. (1971). Necessary high-order conditions of optimality for discrete systems. (Russian, English), Autom.Remote Control, (50-57), translation from Avtom. Telemekh., No.1, (58-65).

10. Ioffe, A. D. (1990). Proximal analysis and approximate subdifferentials. J. London Math. Soc. 41 (175-192).

11. Kasimbeyli, R. (2010). A nonlinear cone seperation theorem and scalaration in nonconvex vector optimization, SIAM J. Optim. Vol.20, No.3, (1591-1619).

12. Kasimbeyli, R. and Gonca, I. (2010). The properties of the weak subdifferentials, Gazi University Journal of Science 23(1):49-53.

13. Kasimbeyli, R. and Mamedov, M. (2009). On weak subdifferential, directional derivatives, and radial epiderivatives for nonconvex functions. SIAM J. Optim. Vol.20, No.2, pp.841-855.

14. Magerramov, S. F. and Mansimov, K. B. (2001). Optimization of a class of discrete step control systems. (Russian, English)(3) Comput. Math. Math. Phys. 41, (334-339). translation from Zh. Vychisl. Mat. Mat. Fiz. 41, No.3, (360-366)

15. Magerramov, S. F. and Mansimov, K.B. (2001). Optimization of a class of discrete step control systems, (Russian, English)-Comput. Math. Math Phys. 41, 3, (334-339), Translation from Zh. Vychisl Mat. Mat. Fiz. 41, No.3, (360-366).

16. Maharramov, S. and Dempe. S. (2005). Optimization a class discrete system with varying structure Preprint 2005-US34,TU Berkgakademie Freiberg, Faculty of Mathematics and Informatics, Germany, 20pages (link:http://www.mathe.tu-freiberg.de/~dempe/Artikel/maharramov.pdf)

17. Maharramov, S. F. (2001). Investigation of singular controls in one discrete system with variable structure and delay. Proceeding of the Institute of Mathematics and Mechanics. Volume 14, 169-175.

18. Maharramov, S. F. (2003). Uslovija optimal'nosti dlja odnogo klassa diskretnikh zadach optimal'nogo upravlenija. PhD thesis, Akademija Nauk Azerbajdshana, Inst. Kibernetika.

19. Maharramov, S. F. (2003). Analysis of the quasisingular control for the a step control problem. Azerbajdshan Republic "Tahsil" Society , Journal "Bilgi". Physical, Mathematics & Earth Sciences "Bilgi", no.1, (44-50)

20. Maharramov, S. F. (2010). Necessary optimality conditions for switching control problem. American Institute of mathematical sciencese, Journal of industrial and management optimization, Vol.6, No 1, (47-57).

21. Maharramov, S. F. with Sengul, M. and Gurbuz, E. (2010). The Proceedings of 24th Mini Conference. On Continuous Optimization and Information-Based Technologies in the Financial Sectors. Some Properties of Weak Subdifferential. 60-65.

22. Mansimov, K. B. and Diskretnie, K. (1985). Sufficient Conditions of Krotov Type In Discrete 2-Parameter System, Automation And Remote Control 46 (8): (932-937) Part 1.

23. Michel, P. and Penot, J.P. (1985). Calcul sous-differential pour des fonctions Lipschitziennes et non-Lipschiziennes, C. R. Acad. Sci. Paris, Ser. I Math. 298.

24. Mordukhovic, B.S. Variational Analysis and Generalized Differention, Vol (330-584), Springer-Verlag, Berl-2005.

25. Mordukhovich, B. S. (1976). Maximum principle in problems of time optimal control with nonsmooth constraints, J. Appl. Math. Mech. 40, 960-969

26. Mordukhovich, B. S. (1988). Approximation Methods in Problems of Optimization and Control, Nauka, Main Physical and Mathematical Editions, Moscow, 360 (Russian)

27. Mordukhovich, B. S. (2005). Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications, Vol. 330, 331 of Grundlehren Series (Fundamental Principles of Mathematical Sciences), Springer-Verlag, NewYork.

28. Mordukhovich, B. S. and Shvartsman, I. (2002). Discrete maximum principle for nonsmooth optimal control problems with delas, Cybernet.Systemss.Anal., 38. 255-264.

29. Propoy, A. (1969). Elementi teorii optimalnix diskretnix prosesov. M.Hauka, 384

30. Rockkafeller, R.T. and R.J.-B. Wets. Variational Analysis. Springer, Berlin, 1998.

31. Sorin, C.B. and DeCarlo, A.R. Optimal control of switching stystem. Automatica 41 (1) (11-27)

32. Xu, X. P. and Antsaklis, A. (2000). A dynamic programming approach for optimal control of switched system. Proc. IEEE Conference. Decision Control 1822-1827.