



ADDIS ABABA UNIVERSITY
GRADUATE PROGRAM
DEPARTMENT OF MATHEMATICS

GRADUATE SEMINAR REPORT

ON

**OPTIMALITY CONDITIONS FOR NONSMOOTH OPTIMIZATION
AND MORDUKHOVICH SUBDIFFERENTIALS**

***(SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR THE DEGREE OF MASTER OF SCIENCE IN MATHEMATICS)***

By Belay Bekele

Advisor: Dr. Semu Mitiku

ADDIS ABABA
JANUARY, 2012

Declaration

I declare that this project has been composed by me and that no part of the project has formed the basis for the award of any Degree, Diploma, Associate ship, Fellowship, or any other similar title to me.

Belay Bekele

Permission

*This is to certify that this project is compiled by **Mr. Belay Bekele W/Michael** in the Department of Mathematics, Addis Ababa University, under my supervision. I hereby also confirm that the project can be submitted for evaluation by examiners and eventual defense.*

Date: -----

Advisor:

Semu Mitiku (Ph.D)

Examiners:

1. -----

Berhanu Guta (Ph.D)

2. -----

K.Venkateswarlu (Ph.D)

To the memory of my Father

Abstract

The differentiability assumptions plays a vital role in nonlinear programming, because most of methods of finding the optimum point in non linear programming starts by finding the gradient of the function and then the stationary points. For unconstrained optimization problems, checking the Positive definiteness of the Hessian matrix at stationary points, one can conclude whether those stationary points are optimum points or not if the objective function is differentiable. Similarly, if the objective function and functions in the constraint set are differentiable, the well known optimality condition called Karush Kuhn Tucker (KKT) condition leads to find the optimum point(s) of the given optimization problem. But, since finding the gradient of the function for non-differentiable functions is not possible, we treat the problem by finding the subgradient, the directional derivative, finding the Mordukhovich normal cone depending on the convexity of the function. Consequently, the optimization procedures for the optimization problems on which functions in the problem are not differentiable is different from the optimization procedures for the optimization problems in which the objective function as well as functions in constraints are differentiable. This project focuses on finding the optimality conditions for optimizations problems without any differentiability assumptions. The subgradient and directional derivative approach are used to solve nonsmooth optimization problem of convex type; and the Mordukhovich extremal principle is applied to solve nonsmooth optimization problems of non convex type.

Acknowledgment

First of all, I would like to thank my advisor, Dr.Semu Mitiku, who was eager and cooperative to help in any difficulty I faced in this project as well as who seriously supervises the project and who shaped the outline of the project to appear in this manner.

I would also like to thank my classmates who gave their constructive comment and who participated on correcting connotations involved in this project.

Finally, I would like to thank my wife Tesfanesh Temesgen and my daughter Betselot, who provide suitable situations for me to proceed on the project continuously and who comment on my work.

Table of Contents

1. Introduction	7
1.1 Method for unconstrained smooth optimization	7
1.2 Karush-Kuhn- Tucker (KKT) necessary optimality condition for constrained smooth optimization	9
1.3 Statement of the problem.....	11
1.4 Objectives	11
2. Non smooth and convex optimization	11
2.1 Directional Derivatives and Subdifferentials for Convex Functions	12
2.1.1 Unconstrained convex and non smooth optimization	15
2.1.2 constrained non smooth and convex optimization	16
2.2 optimality conditions	17
3. Mordukhovich optimality conditions	21
3.1 Frechet derivative and subdifferentials	21
3.2 Multifunction and Sequential Pain-lieveKuratowski upper limit	23
3.3 Mordukhovich Normals and sub differentials.....	24
3.4 Coderivatives	28
3.5 Exremal principles.....	30
3.6 Sequentially normally compact sets.....	30
3.7 Calculus of Mordukovich subdifferentials	32
3.8 Calculus of Mordukhovich Normals	33
3.9 Optimality conditions.....	34
4.Refernces.....	38

1. Introduction

Consider the following types of smooth extremum problems (or smooth mathematical programming problems):

$$\min_{x \in \mathbb{R}^n} f(x); \quad (P_{uc})$$

$$\min_{x \in S} f(x) \quad (P_c), \text{ where}$$

$S = \{x \mid x \in X, g_i(x) \leq 0, i=1, \dots, m, h_j(x) = 0, j=1, \dots, r\}$ called the constraint set, and $X \subseteq \mathbb{R}^n$ is any set, f, g_i ($i = 1, \dots, m$) are real-valued functions, all defined and differentiable on an open set $X \subseteq \mathbb{R}^n$, with $X \subseteq D$;

h_j ($j = 1, \dots, r < n$) are real-valued functions, all defined and continuously differentiable on D .

Problem (P_{uc}) is called unconstrained optimization problem, whereas (P_c) is constrained optimization problem.

Definition: The optimization problem in which the objective function is differentiable and functions involved in the constraints are continuously differentiable is called smooth optimization problem.

Definition : Let X be a normed vector space and $f: X \rightarrow \mathbb{R}$. We say that

1. f has a local minimum at x^0 if and only if there exists a neighborhood V of x^0 such

that $f(x^0) \leq f(x)$, for all $x \in V$.

2. f has a global minimum at x^0 over X if and only if $f(x^0) \leq f(x)$, for all $x \in X$.

1.1 Method for unconstrained smooth optimization

For (P_{uc}) , if $\nabla f(x^0) \neq 0$, then x^0 is not a minimizer of f .

Points $\nabla f(x^0) = 0$ are candidates of minimizer. (stationary points.)

If the Hessian matrix is positive definite at x^0 , then x^0 is a minimizer.

Example 1.1

$$\min f(x) = 100(x_1 - x_2)^2 + (1 - x_1)^2.$$

$$x \in \mathbb{R}^n$$

$$\text{Solution: } \nabla f(x) = 0 \Rightarrow \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} -400(x_2 - x_1^2) + 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow x^0 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ is a stationary point.}$$

Now, the Hessian matrix:

$$\begin{aligned} \nabla^2 f(x^0) &= \begin{pmatrix} f_{x_1 x_1} & f_{x_1 x_2} \\ f_{x_2 x_2} & f_{x_2 x_1} \end{pmatrix} \Big|_{x^0} \\ &= \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix} \text{ is positive definite.} \end{aligned}$$

$$\Rightarrow x^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ is a minimizer.}$$

For (P_c) , we use Lagrange function to transform the problems into unconstrained form.

That is we define

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle,$$

$$(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^r$$

$$= f(x) + \sum_{i=1}^m \lambda_i g_i + \sum_{i=1}^r \mu_i h_i \cdot \lambda_i \geq 0, \text{ provided that } \lambda \text{ and } \mu \text{ exist.}$$

Then we have new unconstrained optimization problem:

$$\begin{aligned} &\min L(x, \lambda, \mu) \\ &(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^r \end{aligned}$$

Because of the assertion, "if x^0 is a solution of $(P_{\lambda, \mu})$, then x^0 is also a solution (P_c) ",

Solving $(P_{\lambda, \mu})$ is similar to solving (P_c) .

For solving problem $(P_{\lambda, \mu})$, the well-known optimality condition called Karush-Kuhn-Tucker (KKT) necessary optimality condition helps.

Notation: For any feasible point x , the set of active inequality constraint is denoted by

$$A(x) = \{i \mid g_i(x) = 0\}.$$

Definition:- A feasible vector $x \in S$ is said to be regular if the gradients $\nabla h_j(x), j=1, 2, \dots, r$ of equality constraint and the gradient $\nabla g_i(x), i \in A(x)$ of the active inequality constraints are linearly independent.

Definition: Let E be a vector space and M be a non empty subset of E . If $f: M \rightarrow \mathbb{R} \cup \{\pm\infty\}$, we call

i) $\text{dom } f := \{x \in M \mid f(x) < +\infty\}$ the effective domain of f .

ii) $\text{epi } f = \{(x, \alpha) \in M \times \mathbb{R} \mid f(x) \leq \alpha\}$ the epigraph of f .

Further, f is said to be proper if

$$\text{dom } f \neq \emptyset \text{ and } f(x) > -\infty \text{ for each } x \in M$$

Definition: Let E be a topological vector space, M a non empty subset of E , and $f: M \rightarrow \mathbb{R} \cup \{\pm\infty\}$. The functional f is said to be lower semicontinuous (l.s.c) at $x^0 \in M$ if either $f(x^0) = -\infty$ or for $k < f(x^0)$ there exists a neighborhood U of x^0 such that $k < f(x) \forall x \in U \cap E$. The functional f is said to be lower semicontinuous (l.s.c) on M if it is l.s.c $\forall x \in E$.

1.2 Karush-Kuhn-Tucker (KKT) necessary optimality condition for constrained smooth optimization

Let x^* be a local minimum of (P_c) with f, g_i, h_i are differentiable. Assume that x^* is regular. Then there exist unique Lagrangian multipliers

$$\lambda^* = (\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*) \text{ and } \mu^* = (\mu_1^*, \mu_2^*, \dots, \mu_m^*) \text{ such that}$$

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0$$

$$\lambda_i^* g_i(x^*) = 0 \text{ for all } i \in \{1, 2, \dots, m\}, \lambda_i^* \geq 0$$

$$h_j(x^*)=0 \quad \text{for all } j=1,2,\dots,r.$$

Under suitable convexity assumption the KKT conditions are also sufficient optimality conditions.

Example 1.2

$$\min f(x) = x_1^2 + x_2^2 - 14x_1 - 6x_2 - 7$$

$$x \in S$$

$$S = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 \leq 2, x_1 + 2x_2 \leq 3\}$$

solution

$$L(x, \lambda) = f(x_1, x_2) + \lambda_1(x_1 + x_2 - 2) + \lambda_2(x_1 + 2x_2 - 3)$$

Then we consider the following optimization problem.

$$\text{Min } L(x, \lambda) \quad (p_\lambda)$$

$$(x, \lambda) \in \mathbb{R}^2 \times \mathbb{R}_+^2$$

First let us check regularity condition. Since the gradient active inequality constraints, namely

$$\{\partial_x(x_1 + x_2 - 2), \partial_x(x_1 + 2x_2 - 3)\} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}, \text{ where } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

are linearly independent, the regularity condition is satisfied. Moreover, Since f, g_1 and g_2 are convex, the KKT conditions are sufficient for (x_1, x_2) to be a solution of (p_λ) .

From KKT conditions we get:

$$(1) \frac{\partial L}{\partial x_1} = 2x_1 - 14 + \lambda_1 + \lambda_2 = 0$$

$$(5) x_1 + x_2 \leq 2$$

$$(2) \frac{\partial L}{\partial x_2} = 2x_2 - 6 + \lambda_1 + 2\lambda_2 = 0$$

$$(6) x_1 + 2x_2 \leq 3$$

$$(3) \lambda_1(x_1 + x_2 - 2) = 0$$

$$(7) \lambda_1, \lambda_2 \geq 0$$

$$(4) \lambda_2(x_1 + 2x_2 - 3) = 0$$

Solving this system of equations and inequalities by distinguishing different cases, we have the only solution $(x_1^*, x_2^*) = (3, -1)$ for (P_λ) . And by the assertion,

$(x_1^*, x_2^*) = (3, -1)$ is also a solution for the original problem.

1.3 Statement of the problem

In smooth optimization problems, as we have seen by example 1.1 and 1.2, all the functions involved were assumed to be differentiable or continuously differentiable.

Frequently, however, the assumption of classical differentiability is not always true since in many modern applications one can find functions which are not differentiable. Methods listed in section 1.1 and 1.2 will be failed if functions involved in objective function or constraints are not differentiable. Accordingly, the method of solving optimization problems without differentiability assumptions is the core problem of this project. We call such problems as non smooth optimization problems; and this project is entitled to set optimality conditions for such problems by using generalized directional derivative, sub gradient and contingent cone approach and Mordukhovich Calculus.

1.4 Objectives

The General objective of this project is to set optimality conditions for non smooth optimization problems and the specific objectives are:

- ❖ To present the relation of generalized directional derivatives and local cone approximations of sets so as to construct optimality conditions for some convex nonsmooth functions.
- ❖ To give a short survey of Mordukhovich optimality principle for non convex as well as nonsmooth functions.

2. Non smooth and convex optimization

Definition: A set X is said to be convex if $\lambda x + (1-\lambda)y \in X$ for all $x, y \in X$ and $\lambda \in (0,1)$.

Definition: A convex hull of a set X , denoted by $\text{Conv}(X)$, is the intersection of all convex sets containing X .

Definition: A function f is said to be convex if

$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ for all $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$.

As an example to this, let M be a nonempty subset of a normed vector space E . The indicator function of M $\delta_M: E \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is defined by

$$\delta_M := \begin{cases} 0 & \text{if } x \in M, \\ +\infty & \text{if } x \in E \setminus M \end{cases}$$

Then,

1. δ_M is a proper and convex function if and only if M is non empty and convex.
2. δ_M can be used as proper penalty function for most optimization problems.

2.1 Directional Derivatives and Subdifferentials for Convex Functions

Let $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ be an extended real-valued convex function and $x^0 \in \mathbb{R}^n$ be a point where f is finite.

Definition:- The directional derivative of f at x^0 in the direction of y is:

$$f'(x^0, y) = \lim_{t \downarrow 0} \frac{f(x^0 + ty) - f(x^0)}{t}$$

The directional derivative of f at x^0 exists for each direction y (possibly with values $\pm\infty$) and provides an extended real-valued function $f'(x^0, \cdot)$

Definition: The subdifferential of f at the point $x^0 \in \mathbb{R}^n$ is a set of vectors given by $\partial f(x^0) = \{u \in \mathbb{R}^n \mid u(x - x^0) \leq f(x) - f(x^0), \forall x \in \mathbb{R}^n\}$.

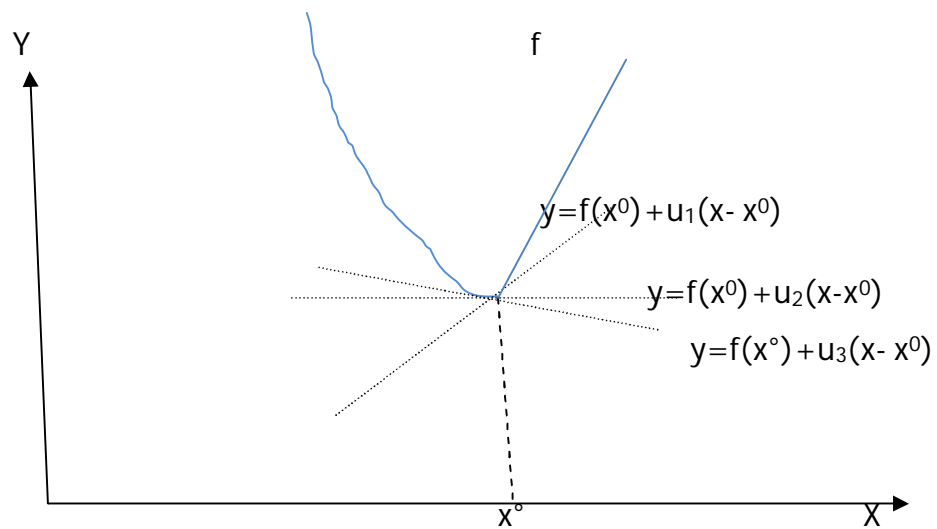


Fig. 1: example of subgradient.

A vector $u \in \partial f(x^0)$ is called the subgradient of f at x^0 .

For $f(x) = \text{Max}\{f_1(x), \dots, f_m(x)\}$ and $I(x) = \{i / f_i(x) = f(x)\}$, the index of 'active' functions at x ,

$$\partial f(x) = \text{conv} \cup_{i \in I(x)} \partial f_i(x)$$

That is, convex hull of the union subgradient of 'active' functions at x .

Moreover, if f_i 's are differentiable,

$$\partial f(x) = \text{conv}\{\nabla f_i(x) \mid i \in I(x)\}.$$

Example 2.1

Let $f(x) = \max\{-3x, \{x^2 - 4, 2x - 1\}\}$

$$f(x) = \begin{cases} -3x & \text{for } x \in [-4, 1/5] \\ 2x - 1 & \text{for } x \in [1/5, 3] \\ x^2 - 4 & (-\infty, -4] \cup [3, \infty) \end{cases}$$

Then, f is not differentiable at points $\{-4, 1/5, 3\}$. We consider these points as stationary points, and

$$\partial f(x) = \begin{cases} -3 & \text{for } x \in (-4, 1/5) \\ [-8, -3] & \text{for } x = -4 \\ [-3, 2] & \text{for } x = \frac{1}{5} \\ 2 & \text{for } x \in (\frac{1}{5}, 3) \\ [2, 6] & \text{for } x = 3 \\ 2x & \text{otherwise} \end{cases}$$

Theorem 2.1. Let f be directionally differentiable at x^0 in the direction y . f is convex if and only if

$$f(x) \geq f(x^0) + f'(x^0, x - x^0) \quad \text{for all } x^0, y \in \mathbb{R}^n.$$

proof

$$f \text{ is convex} \Leftrightarrow f(x^0 + \lambda(x - x^0)) = f(\lambda x + (1-\lambda)x^0) \leq \lambda f(x) + (1-\lambda)f(x^0)$$

$$= \lambda f(x) + f(x^0) - \lambda f(x^0)$$

$$\Leftrightarrow f(x) \geq f(x^0) + \frac{1}{\lambda} [f(x^0 + \lambda(x - x^0)) - f(x^0)]$$

Applying the limit as λ goes to zero on both sides we get $f(x) \geq f(x^0) + f'(x^0, x - x^0)$.

Conversely, assume $f(x) \geq f(x^0) + f'(x^0, x - x^0)$ for all $x^0, y \in \mathbb{R}^n$.

For all $x^0, y \in \mathbb{R}^n$, and for all $\lambda \in (0, 1)$ we obtain

$$f(x^0) \geq f(\lambda x^0 + (1-\lambda)x) + f'(\lambda x^0 + (1-\lambda)x)(x^0 - x)(1-\lambda)$$

and

$$f(x) \geq f(\lambda x^0 + (1-\lambda)x) + f'(\lambda x^0 + (1-\lambda)x)(x^0 - x)(-\lambda)$$

Since directional derivatives are linear mappings, we conclude further

$$\begin{aligned} \lambda f(x^0) + (1-\lambda)f(x) &\geq \lambda f(\lambda x^0 + (1-\lambda)x) + \lambda(1-\lambda) f'(\lambda x^0 + (1-\lambda)x)(x^0 - x) \\ &+ (1-\lambda)f(\lambda x^0 + (1-\lambda)x) - \lambda(1-\lambda) f'(\lambda x^0 + (1-\lambda)x)(x^0 - x) \\ &= f(\lambda x^0 + (1-\lambda)x). \end{aligned}$$

Consequently the function f is convex. \square

From theorem 2.1 we see that for every convex function:

$$f'(x^0)(x - x^0) \leq f(x) - f(x^0)$$

This means that the subgradient at an arbitrary point $x^0 \in \mathbb{R}^n$ is non empty.

2.1.1 Unconstrained convex and non smooth optimization

Theorem 2.2. Let f be an extended convex function and let $x^0 \in \mathbb{R}^n$ be a point where f is finite; a necessary; and sufficient condition for x^0 to be a minimum point for f is that:

$$0 \in \partial f(x^0).$$

Proof

By definition of subgradient,

$$0 \in \partial f(x^0) \Leftrightarrow f(y) \geq f(x^0), \forall y \in \mathbb{R}^n.$$

That is x^0 is a minimum point for f . \square

Example 2.2: Let $f(x) = \max\{-3x, \{x^2 - 4, 2x - 1\}\}$

find $\min_{x \in \mathbb{R}} f(x)$

Solution

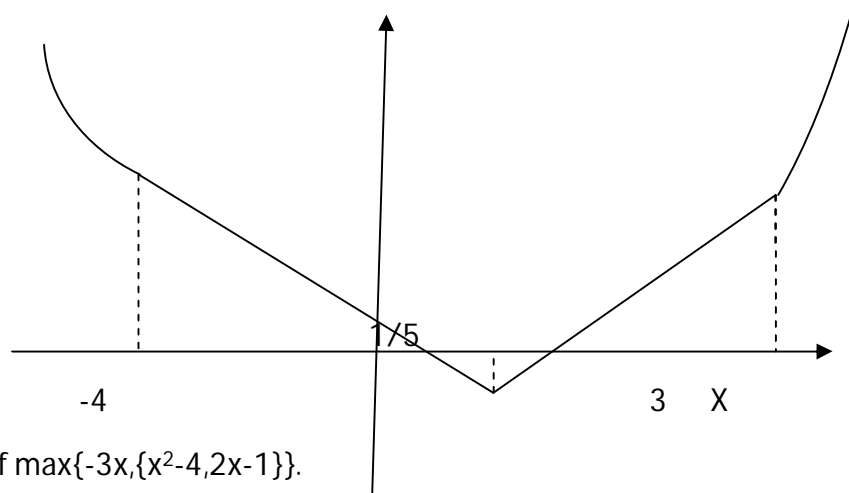


Fig. 2 : graph of $\max\{-3x, \{x^2 - 4, 2x - 1\}\}$.

We see from example 2.1 that:

The subdifferential of f containing zero occurs only at $x^* = 1/5$.

Hence $x^* = 1/5$ minimizes f . \square

Theorem 2.3. Let f be a convex function and $x^0 \in \mathbb{R}^n$ be a point where f is finite. Then x^0 is a (global) minimum point of f if and only if the following equivalent conditions hold:

i) $f'(x^0, y) \geq 0 \quad \forall y \in \mathbb{R}^n,$

ii) $0 \in \partial f(x^0).$

proof

If x^0 is a minimizer of f , then

$$\inf_{t>0} \frac{f(x^0 + ty) - f(x^0)}{t} \geq 0.$$

This implies (i) holds. Then (i) equivalent to (ii) as it follows from theorem 2.1, and (ii) holds by theorem 2.2. \square

2.1.2 Constrained non smooth and convex optimization

Now consider the convex constrained optimization problems

$$\text{Min } f(x) \quad (P_0)$$

$$x \in S,$$

where the feasible set is given by

$$S = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m\} \quad \text{and}$$

$$\text{Min } f(x) \quad (P_c)$$

$$x \in S,$$

where the feasible set is given by

$$S = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m, h_j(x) = 0, j = 1, \dots, r\}$$

Defintion:- Given $S \subseteq \mathbb{R}^n$, $S \neq \emptyset$ and $x^0 \in \text{cl}(S)$, the set:

$$T(S, x^0) = \{y \in \mathbb{R}^n \mid \exists \{x^k\} \subseteq S, \{x^k\} \rightarrow x^0, \exists \{\lambda_k\} \in \mathbb{R}, \lambda_k > 0:$$

$y = \lim_{k \rightarrow \infty} \lambda_k(x^k - x^0)\}$ is called a contingent cone to S at x^0 .

Theorem 2.4. Let f be a convex function, S be a convex set and $x^0 \in \text{int}(\text{dom}(f))$ be a feasible point. Then x^0 is a (global) minimum point of the problem (P_0) if and only if

$$f'(x^0, y) \geq 0, \forall y \in T(S, x^0).$$

proof

Let $y \in T(S, x^0)$. This implies, by definition of contingent cone, $\exists \{x^k\} \subseteq S, \{x^k\} \rightarrow x^0, \exists \{\lambda_k\} \in \mathbb{R}, \lambda_k > 0: y = \lim_{k \rightarrow \infty} \lambda_k(x^k - x^0)$. Since f is directionally differentiable, at x^0 ,

$$f(x^k) - f(x^0) = f'(x^0, x^k - x^0) + o(\|x^k - x^0\|).$$

Multiplying both sides of the last equation by λ_k and applying limit as k goes to ∞ on both sides of this equation we get:

$$\begin{aligned} \lim_{k \rightarrow \infty} \lambda_k (f(x^k) - f(x^0)) &= \lim_{k \rightarrow \infty} \lambda_k (f'(x^0, x^k - x^0)) + \lim_{k \rightarrow \infty} \lambda_k o(\|x^k - x^0\|). \\ &= \lim_{k \rightarrow \infty} \lambda_k (f'(x^0, x^k - x^0)) + 0 \quad \text{because of } \{x^k\} \rightarrow x^0 \text{ as } k \rightarrow \infty. \\ &= f'(x^0, y) \quad \text{because of } y = \lim_{k \rightarrow \infty} \lambda_k (x^k - x^0) \end{aligned}$$

But since $\lambda_k > 0$ and $f(x^k) - f(x^0) \geq 0$ as x^0 is minimize of f , we have

$\lim_{k \rightarrow \infty} \lambda_k (f(x^k) - f(x^0)) \geq 0$. Since $\lim_{k \rightarrow \infty} \lambda_k (f(x^k) - f(x^0)) = f'(x^0, y)$, we get the desired result $f'(x^0, y) \geq 0$. \square

2.2 optimality conditions

Lemma 2.1. Let $x^0 \in \text{int}(\text{dom}(f))$ be a solution of (P_0) . Let f and g be directionally differentiable at x^0 ; then the system

$$\left\{ \begin{array}{l} f'(x^0, y) < 0 \\ g'_i(x^0, y) < 0 \\ g'_i(x^0, y) \leq 0, \end{array} \right. , i \in A(x^0)$$

has no solution y in \mathbb{R}^n .

The proof of this lemma is given in [6] page 250.

Definition: A function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be affine if it has the form $f(x) = a^T x + b$ for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Similarly a function $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is said to be affine if it has the form $f(x) = Ax + b$ for some $M \times N$ matrix A and $b \in \mathbb{R}^m$. If $b = 0$, then f is said to be a linear function or a linear transformation.

Lemma 2.2. Let f_1, \dots, f_m be convex real valued functions all defined on convex set $X \subseteq \mathbb{R}^n$ and let g_1, \dots, g_r be linear affine functions on \mathbb{R}^n . If the system

$$x \in X, f_i(x) < 0, i = 1, \dots, m$$

$$g_j(x) = 0, \quad j = 1, \dots, r$$

admits no solution then there is a non-negative vector $u \in \mathbb{R}^m$ and a vector $v \in \mathbb{R}^r$ with $(u, v) \neq 0$ such that $uf(x) + vg(x) \geq 0$ for all $x \in X$.

The proof of this lemma is given in [6] page 111.

Lemma 2.3. Let f_1, \dots, f_m be convex real valued functions all defined on convex set $X \subseteq \mathbb{R}^n$. Then either the system

$$x \in X, f_i(x) < 0, i = 1, \dots, m$$

$$uf(x) \geq 0 \text{ for all } x \in X, \text{ for some } u \geq 0, u \in \mathbb{R}^m,$$

but never both.

Proof

Let x^* be solution of $f(x) < 0$. Then for any $u \geq 0, u \in \mathbb{R}^m, uf(x) < 0$.

Conversely, if the system $x \in X, f(x) < 0$, has no solution, from lemma 2.2, it follows that there exists $u \geq 0$, such that $uf(x) \geq 0$ for all $x \in X$. \square

Theorem 2.5. Let f and $g_i, i = 1, \dots, m$, be convex functions and let x^0 be a point in which all functions are finite and continuous. If x^0 is a minimum point of the problem (P_0) then there exist multipliers $u_i \geq 0, i = 0, \dots, m$, not all vanishing, such that $u_i g_i(x^0) = 0$ for $i = 1, \dots, m$, and such that the following condition holds:

$$u_0 f'(x^0, y) + \sum_{i=1}^m u_i g'_i(x^0, y) \geq 0, \forall y \in \mathbb{R}^n.$$

Proof

By lemma 2.1 we can state that there is no vector $y \in \mathbb{R}^n$ which is the solution of the system

$$\begin{cases} f'(x^0)(y) < 0 \\ g'_i(x^0)(y) < 0, & i \in A(x^0) \end{cases}$$

By lemma 2.2 we find multipliers $u_i \geq 0, i \in \{0\} \cup A(x^0)$, not all vanishing, such that

$$u_0 f'(x^0, y) + \sum_{i \in A(x^0)} u_i g'_i(x^0, y) \geq 0, \forall y \in \mathbb{R}^n$$

Setting $u_i = 0$ for $i \notin A(x^0)$ we get the desired result. \square

Definition:- Consider problem (P_c) , where f and $g_i, i = 1, \dots, m, h_j, j = 1, \dots, r$ are directionally differentiable and convex functions and let $x^0 \in S$. Then the first order constraint qualification condition is said to hold at x^0 if:

$\forall y \neq 0$ such that

$$\begin{cases} y^T g'_i(x^0) \geq 0, & i \in A(x^0) \\ y^T h'_j(x^0) = 0. \end{cases}$$

In connection with convex optimization problems one uses the important regularity condition of Slater or Slater-condition:

Let g_i be convex for each $i \in \{1, \dots, m\}$ and h_j be affine for each $j \in \{1, \dots, p\}$. The set

$$S = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$$

is said to satisfy the Slater-condition if and only if there is $x^0 \in S$ such that

$$g_i(y) < 0 \text{ for all } i \in \{1, \dots, m\}.$$

Theorem 2.6. Let f and g_i , $i = 1, \dots, m$, be convex functions and let $x^0 \in S$ be a point in which all functions are finite and continuous. Further let the Slater constraint qualification condition be fulfilled. Then x^0 is a (global) minimum point of the problem (P_0) if and only if there exist multipliers $u_i \geq 0$, $i = 1, \dots, m$, such that $u_i g_i(x^0) = 0$ for $i = 1, \dots, m$, and such that the following holds:

$$f'(x^0, y) + \sum_{i=1}^m u_i g'_i(x^0, y) \geq 0, \forall y \in \mathbb{R}^n.$$

Proof

Let x^0 be a minimum of (P_0) . By theorem 2.5 there are multipliers $u_i \geq 0$, $i = 0, 1, \dots, m$, not all vanishing, such that

$$u_0 f'(x^0, y) + \sum_{i \in A(x^0)} u_i g'_i(x^0, y) \geq 0, \forall y \in \mathbb{R}^n$$

Assuming $u_0 = 0$ we would get

$$\sum_{i \in A(x^0)} u_i g'_i(x^0, y) \geq 0, \quad \forall y \in \mathbb{R}^n.$$

Again by lemma 2.3 we see that the system $g'_i(x^0, y) < 0$, $i \in A(x^0)$

is not solvable. By assuming the Slater condition, the solvability of the system is ensured because of: $g'_i(x^0, x-x^0) \leq g_i(x) - g_i(x^0) < 0$ for all $i \in A(x^0)$, which implies

$g'_i(x^0, x-x^0) \leq g_i(x) < 0$ since $g_i(x^0) = 0$ for all $i \in A(x^0)$. This in turn implies

$g'_i(x^0, x-x^0) < 0$. This means, that the system $g'_i(x^0, x-x^0) < 0$ is solvable, which is contradiction. Therefore, the assumption $u_0 = 0$ is false. Thus $u_0 > 0$. Without loss of generality we can set $u_0 = 1$ and get the desired result. \square

3. Mordukhovich optimality conditions

3.1 Frechet derivative and subdifferentials

Definition: Let E and F be vector spaces. A functional $f: E \times F \rightarrow \mathbb{R}$ is said to be bilinear if $f(\cdot, u)$ is linear on E for each fixed $u \in F$ and $f(u, \cdot)$ is linear on F for each fixed $u \in E$.

Let $f: E \times F \rightarrow \mathbb{R}$ be a bilinear functional with the following properties:

If $f(x, u) = 0$ for each $x \in E$, then $u = 0$.

If $f(x, u) = 0$ for each $u \in F$, then $x = 0$.

Then (E, F) is dual pairs of vector spaces with respect to f .

Definition: A sequence (x_k) in E is said to be weakly convergent to $x \in E$, written $x_k \xrightarrow{w} x$ as $k \rightarrow \infty$; this means that

$\lim_{k \rightarrow \infty} \langle x^*, x_k \rangle = \langle x^*, x \rangle$ for any $x^* \in E^*$.

Analogously, A sequence (x_k^*) in E^* is said to be weak* convergent to $x^* \in E^*$, written $x_k^* \xrightarrow{w^*} x^*$ as $k \rightarrow \infty$; this means that

$\lim_{k \rightarrow \infty} \langle x_k^*, x \rangle = \langle x^*, x \rangle$ for any $x \in E$.

Definition:- Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers. We say that $\{a_n\}_{n=1}^{\infty}$ is a Cauchy sequence if for every $\epsilon > 0$ there exists a positive integer N such that

$$|a_n - a_m| < \epsilon \quad \text{for all } n, m \geq N.$$

Definition:- A metric space consists of a pair (X, d) , where X is a set and $d: X \times X \rightarrow \mathbb{R}$ is a function, called the metric or distance function, such that the following hold for all $x, y, z \in X$

1. (Symmetry) $d(x, y) = d(y, x)$
2. (Positive Definiteness) $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$
3. (Triangle Inequality) $d(x, z) \leq d(x, y) + d(y, z)$.

Definition:- A metric space (X, d) is said to be complete if every Cauchy sequence converges in X .

Definition:- A Banach space is a complete metric space.

Definition: Let X and Y be a normed vector spaces and let S be a non empty subset of X . A function $f: S \rightarrow Y$ is said to be Frechet differentiable at x^0 in the direction of y if there is continuous linear function $f'(x^0): X \rightarrow Y$ with property

$$\lim_{\|y\| \rightarrow 0} \frac{\|f(x^0+y) - f(x^0) - f'(x^0)(y)\|}{\|y\|} = 0.$$

Then $f'(x^0)$ is called the Frechet derivative (F-derivative) at x^0 .

Definition: Assume that E is a Banach space and A be a non empty subset of E , $f: E \rightarrow \mathbb{R} \cup \{\pm \infty\}$ is a proper l.s.c and $x \in \text{dom } f$. The function f is said to be Frechet subdifferentiable (F-subdifferentiable) at x^0 if there exists $x^* \in E^*$, the F-subderivative of f at x^0 , such that

$$\liminf_{y \rightarrow 0} \frac{f(x^0 + y) - f(x^0) - \langle x^*, y \rangle}{\|y\|} \geq 0$$

The set of all F-subdifferentials at x^0 is denoted by $\partial_F f(x^0)$.

Proposition 3.1

- a) If x^0 is a local minmizer of f on A then it is also a local minmizer of $f + \delta_A$ on E .
- b) If x^0 is local minmizer of f on A then $0 \in \partial_F(f + \delta_A)(x^0)$.

Proof:

a) Let x^0 is a local minmizer on A .

This implies there exist an neighborhood U of x^0 such that $f(x^0) \leq f(x)$ for all $x \in U \cap A$.

$$\text{Since } \delta_A(x) = \begin{cases} 0 & \text{for } x \in A \\ \infty & \text{for } x \notin A \end{cases}, (f + \delta_A)(x) = \begin{cases} f & \text{for } x \in A \\ \infty & \text{for } x \notin A \end{cases}$$

Therefore, $(f + \delta_A)(x^0) \leq (f + \delta_A)(x)$ for all $x \in U \cap A$.

This completes the proof □

b) First let's proof if x^0 is local minimizer of on A , then $0 \in \partial_F(f(x^0))$.

Since x^0 is local minimizer of f on A , $f(x^0) \leq f(x)$ for all $x \in U \cap A$. let $x = x^0 + y$. Then $f(x^0) \leq f(x^0 + y)$

This implies $f(x^0 + y) - f(x^0) \geq 0$.

Hence,

$$\liminf_{y \rightarrow 0} \frac{f(x^0 + y) - f(x^0) - \langle 0, y \rangle}{\|y\|} \geq 0$$

That is $0 \in \partial_F(f(x^0))$.

Since $(f + \delta_A)(x) = \begin{cases} f & \text{for } x \in A \\ \infty & \text{for } x \notin A \end{cases}$, $0 \in \partial_F(f + \delta_A)(x^0)$ if x^0 is local minimize of f on A . \square

Definition: (a) If $A \subseteq E$ is nonempty, then $A^0 = \{x^* \in E^* \mid \langle x^*, x \rangle \leq 0 \text{ for all } x \in A\}$ is a convex cone, the (negative) polar cone of A .

(b) If A is convex then $N(A, x) = (A - x)^0 = \{x^* \in E^* \mid \langle x^*, y - x \rangle \leq 0 \text{ for all } y \in A\}$ is called Normal cone to A at x .

Definition : A Banach Space E is said to be a Frechet smooth if it admits an equivalent norm that is F -differentiable on $E \setminus \{0\}$.

Definition : A Banach Space E is said to be an Asplund space if every continuous convex functional defined on non empty open convex subset D of E is F -differentiable on a dense subset of D .

3.2 Multifunction and Sequential Painlevé-Kuratowski upper limit

Definition: Let E and F be vector spaces a mapping $\phi: E \rightarrow 2^F$, which associates to $x \in E$ a (possibly empty) subset $\phi(x)$ of F is called a multifunction or a set valued mapping and is denoted by $\phi: E \rightrightarrows F$. The graph and domain of ϕ are defined respectively, by

$$\text{graph } \phi = \{(x, y) \in E \times F \mid x \in E, y \in \phi(x)\}$$

$$\text{dom } \phi = \{x \in E \mid \phi(x) \neq \emptyset\}$$

Example 3.1

Let $E = \{1,2,3\}$ and $F = \{1,2\}$ Then $2^F = \{\{1\},\{2\},\{1,2\},\emptyset\}$, $\Phi(1) = \{\{1\},\{1,2\}\}$,

$\Phi(2) = \{\{2\},\{1,2\}\}$ and $\Phi(3) = \emptyset$.

Definition : A sequence $\{x^k\}$ in a normed space X is said to be norm convergent to a point x^0 in X if for every $\epsilon > 0$ there exists a positive integer N such that

$$\|x^k - x^0\| < \epsilon \quad \text{for each } k \geq N$$

Definition: Let E be a normed vector space, $\Phi: E \rightrightarrows E^*$ be a multifunction, and $\bar{x} \in \text{Dom}\Phi$. The sequential Painlevé-Kuratowski upper limit, $s\text{Limsup}_{x \rightarrow \bar{x}} \Phi(x)$, of Φ is defined to be the set of all $x^* \in E^*$ for which there exist a sequence (x_k) in $\text{Dom}\Phi$ that is norm convergent to \bar{x} and a sequence (x_k^*) in E^* that converges to x^* such that $x_k^* \in \Phi(x_k)$ for all $k \in \mathbb{N}$.

$$s\text{Limsup}_{x \rightarrow \bar{x}} \Phi(x) = \{x^* \in E^* \mid \exists \text{ sequences } (x_k) \rightarrow \bar{x} \text{ in } \text{Dom}\Phi, (x_k^*) \xrightarrow{w^*} x^* \text{ in } E^* \text{ such that } (x_k^*) \in \Phi(x_k)\}$$

3.3 Mordukhovich Normals and sub differentials

Definition : Let A be a nonempty subset of E . If $x \in A$ and $\epsilon \geq 0$, then the set

$$(a) \hat{N}_\epsilon(A, x) := \left\{ x^* \in E^* \mid \limsup_{y \rightarrow A^x} \frac{\langle x^*, y-x \rangle}{\|y-x\|} \leq \epsilon \right\}, \text{ where } y \rightarrow A^x \text{ means } y \rightarrow x \text{ with } x \in A, \text{ is}$$

called set of ϵ -normals to A at x .

(b) If $\bar{x} \in A$, then the set

$$N_M(A, \bar{x}) := s\text{Limsup}_{x \rightarrow \bar{x}} \hat{N}_\epsilon(A, x) \\ \epsilon \downarrow 0$$

will be called Mordukhovich normal cone (M-normal cone) to A at \bar{x} . Put $N_M(A, \bar{x}) = \emptyset$ if $\bar{x} \notin A$ and $N_M(A, \bar{x}) = \{0\}$ if $\bar{x} \in \text{int}A$.

(c) $N_F(A, x) := \partial_F \delta_A(x) = \hat{N}_0(A, x) := \left\{ x^* \in E^* \mid \limsup_{y \rightarrow A^x} \frac{\langle x^*, y-x \rangle}{\|y-x\|} \leq 0 \right\}$ is called Fréchet normal cone to A at x.

Note that $x^* \in N_M(A, \bar{x})$ if and only if there exist sequences $\epsilon_k \downarrow 0$, $x_k \rightarrow \bar{x}$, $x_k^* \xrightarrow{w^*} x^*$ and such that $x_k^* \in \hat{N}_{\epsilon_k}(A, x_k)$ for all $k \in \mathbb{N}$.

Example 3.2

Let $A = \{(x, y) \in \mathbb{R}^2 \mid y \geq -|x|\}$. Then

$$N_M(A, (0,0)) = \{(x^*, x^*) \in \mathbb{R}^2 \mid x^* \leq 0\} \cup \{(x^*, -x^*) \in \mathbb{R}^2 \mid x^* \geq 0\}.$$

Lemma 3.1: Assume that $f: E \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is proper and $(x^0, \alpha^0) \in \text{epi } f$. Then

$(x^*, -\lambda) \in N_M(\text{epi } f, (x^0, \alpha^0))$ implies $\lambda \geq 0$.

Proof:

Let $(x^*, -\lambda) \in N_M(\text{epi } f, (x^0, \alpha^0))$. By definition of M-normal cone there exist sequences $\epsilon_k \downarrow 0$, $(x_k, \alpha_k) \rightarrow \text{epi } f(x^0, \alpha^0)$, $x_k^* \xrightarrow{w^*} x^*$ and $\lambda_k \rightarrow \lambda$ such that

$$(x^0, \alpha^0) \rightarrow \text{epi } f(x_k, \alpha_k) \quad \frac{\langle x_k^*, x^0 - x_k \rangle - \lambda_k (\alpha^0 - \alpha_k)}{\|(x^0, \alpha^0) - (x_k, \alpha_k)\|} \leq \epsilon_k \quad \forall k \in \mathbb{N}.$$

Choosing $x = x_k$ and letting $k \rightarrow \infty$ we get $\lambda \geq 0$. □

Definition: Let $f: E \rightarrow \mathbb{R} \cup \{\pm\infty\}$ be proper and $x \in \text{dom } f$. We call

$$(a) \partial_M f(x) := \{x^* \in E^* \mid (x^*, -1) \in N_M(\text{epi } f, (x, f(x)))\}$$

a Mordukhovich subdifferential (M-subdifferential) or basic subdifferential of f at x and

$$(b) \partial_M^\infty f(x) := \{x^* \in E^* \mid (x^*, 0) \in N_M(\text{epi } f, (x, f(x)))\}$$

a singular Mordukhovich subdifferential (singular M-subdifferential) of f at x.

Definition: let $f: E \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is proper l.s.c and $x \in \text{dom } f$ and let $\epsilon \geq 0$. The set

$$\hat{\partial}_\epsilon f(x^0) = \{x^* \in E^* \mid \liminf_{y \rightarrow x^0} \frac{f(x^0 + y) - f(x^0) - \langle x^*, x - x^0 \rangle}{\|x - x^0\|} \geq -\epsilon\}$$

is called (a Frechet) ϵ - subdifferential of f at x^0 .

Definition: let $f: E \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is proper l.s.c and $x \in \text{dom } f$ and let $\epsilon \geq 0$. The set

$$\hat{\partial}_{g\epsilon} f(x^0) = \{x^* \in E^* \mid (x^*, -1) \in \hat{N}_\epsilon(\text{epi } f, (x^0, f(x^0)))\}$$

is called the geometric ϵ - subdifferential of f at x^0 .

At $\epsilon=0$ we have $\hat{\partial}_\epsilon f(x^0) = \hat{\partial}_{g\epsilon} f(x^0) = \partial_F f(x^0)$.

Proposition 3.2

a. For any set A of E and any point $x \in A$, one has

$$N_M(A, x) = \partial_M \delta_A(x) = \partial_M^\infty \delta_A(x).$$

b. At $\epsilon=0$ we have $\hat{\partial}_\epsilon f(x^0) = \hat{\partial}_{g\epsilon} f(x^0) = \partial_F f(x^0)$.

c. $\partial_F f(x^0) = \{x^* \in E^* \mid (x^*, -1) \in N_F(\text{epi } f, (x^0, f(x^0)))\}$

d. $\partial_F f(x) \subseteq \partial_M f(x)$.

e. if x^0 is a minimizer of f on A , then

$$0 \in \partial_M (f + \delta_A)(x^0). \quad (3.3)$$

Proof

a. Let $x^* \in \partial_M \delta_A(x)$.

$$\Leftrightarrow (x^*, -1) \in N_M(\text{epi } \delta_A, (x, f(x)))$$

$$\Leftrightarrow (x^*, -1) \in N_M(A \times [0, \infty), (x, f(x))), \text{ since } \text{epi } \delta_A = A \times [0, \infty).$$

$$\Leftrightarrow x^* \in N_M(A, x).$$

Thus, $N_M(A, x) = \partial_M \delta_A(x)$

Similarly,

Let $x^* \in \partial_M^\infty \delta_A(x)$.

$$\Leftrightarrow (x^*, 0) \in N_M(\text{epi } \delta_A, (x, f(x)))$$

$$\Leftrightarrow (x^*, 0) \in N_M(A \times [0, \infty), (x, f(x))), \text{ since } \text{epi } \delta_A = A \times [0, \infty) .$$

$$\Leftrightarrow x^* \in N_M(A, x).$$

Thus, $N_M(A, x) = \partial_M^\infty \delta_A(x)$.

Hence,

$$\partial_M \delta_A(x) = N_M(A, x) = \partial_M^\infty \delta_A(x). \quad \square$$

b. Let $x^* \in \hat{\partial}_{g^\epsilon} f(x^0)$. And let $\epsilon = 0$.

$$\Leftrightarrow (x^*, -1) \in \hat{N}_0(\text{epi } f, (x^0, f(x^0)))$$

$$\Leftrightarrow x^* \in \hat{N}_0(A, x^0) = \partial_F f(x^0)$$

Thus, $\hat{\partial}_{g_0} f(x^0) = \partial_F f(x^0)$.

Similarly, for $\epsilon = 0$,

$$\hat{\partial}_0 f(x^0) = \{x^* \in E^* \mid \liminf_{y \rightarrow x^0} \frac{f(x^0 + y) - f(x^0) - \langle x^*, x - x^0 \rangle}{\|x - x^0\|} \geq 0\}$$

which is exactly the same as $\partial_F f(x^0)$.

Hence,

$$\text{at } \epsilon = 0, \text{ we have } \hat{\partial}_\epsilon f(x^0) = \hat{\partial}_{g^\epsilon} f(x^0) = \partial_F f(x^0). \quad \square$$

c. Since, $\partial_F f(x^0) = \hat{\partial}_0 f(x^0) = \hat{\partial}_{g_0} f(x^0)$ from (b) above and since

$$\hat{\partial}_{g_0} f(x^0) = \hat{\partial}_0 f(x^0) = \hat{\partial}_{g_0} f(x^0) = \{x^* \in E^* \mid (x^*, -1) \in \hat{N}_0(\text{epi } f, (x^0, f(x^0)))\},$$

$$\partial_F f(x^0) = \{x^* \in E^* \mid (x^*, -1) \in \widehat{N}_0(\text{epi } f, (x^0, f(x^0)))\}.$$

Thus, $\partial_F f(x^0) = \{x^* \in E^* \mid (x^*, -1) \in N_F(\text{epi } f, (x^0, f(x^0)))\}$. \square

d. Since $N_F(x^0)$ is a special type of Mordokhovich normal cone at $\epsilon=0$, we have

$$N_F(x^0) \subseteq N_M(x^0) \text{ for } x^0 \in A.$$

Thus, $(x^*, -1) \in N_F(\text{epi } f, (x^0, f(x^0)))$ implies $(x^*, -1) \in N_M(\text{epi } f, (x^0, f(x^0)))$ for all $x^* \in E^*$.

It follows that, $\partial_F f(x) \subseteq \partial_M f(x)$. \square

e. This follows from proposition 3.1 b and by the assertion $\partial_F f(x) \subseteq \partial_M f(x)$. \square

Definition :The proper functional $f : E \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is said to be lower regular at $x \in \text{dom } f$ if $\partial_M f(x) = \partial_F f(x)$.

3.4 Coderivatives

Definition:Let $\phi : E \rightrightarrows F$ be a closed multifunction and $(x^0, y^0) \in \text{graph } \phi$.

For any $\epsilon > 0$, the multifunction $\widehat{D}_\epsilon^* \phi(x^0, y^0) : E^* \rightrightarrows F^*$ defined by

$$\widehat{D}_\epsilon^* \phi(x^0, y^0)(y^*) = \{x^* \in E^* \mid (x^*, -y^*) \in \widehat{N}_\epsilon(\text{graph } \phi, (x^0, y^0))\} \quad \forall y^* \in F^*$$

is said to be the ϵ -coderivative of ϕ at (x^0, y^0) . In particular $D_F^* \phi(x^0, y^0) := \widehat{D}_0^* \phi(x^0, y^0)$ is called a Frechet coderivative (F-coderivative) of ϕ at (x^0, y^0) .

The multifunction $D_M^* \phi(x^0, y^0) : E^* \rightrightarrows F^*$ is defined by

$$D_M^* \phi(x^0, y^0)(y^*) = \{x^* \in E^* \mid (x^*, -y^*) \in N_M(\text{graph } \phi, (x^0, y^0))\} \quad \forall y^* \in F^*$$

is called the Mordukhovich coderivative (M-coderivative) of ϕ at (x^0, y^0) .

Definition: If E is a normed vector space, then the proper functional $f : E \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is said to be locally Lipschitz continuous, or briefly locally L-continuous, around $x \in \text{dom } f$ if there exist $\epsilon > 0$ and $\lambda > 0$ such that

$$|f(x) - f(y)| \leq \lambda \|x - y\| \quad \forall x, y \in B(x, \epsilon), \text{ where } B(x, \epsilon) = \{z \in E : \|x - z\| \leq \epsilon\}.$$

Moreover, f is called locally L -continuous on the open subset D of E if f is locally L -continuous around each $x \in D$.

Proposition 3.3:- If $f: E \rightarrow \mathbb{R}$ is locally L -continuous around $x^0 \in E$, then

$$D_M^* f(x^0)(\alpha) = \partial_M(\alpha f)(x^0) \quad \forall \alpha \in \mathbb{R}$$

Proof

Let $x^* \in \partial_M(\alpha f)(x^0)$. This implies there exist sequences $\epsilon_k \downarrow 0, x_k \rightarrow x^0, x_k^* \xrightarrow{w^*} x^*$

such that

$$\limsup_{x^0 \rightarrow x_k} \frac{\langle x_k^*, x^0 - x_k \rangle - \alpha(f(x^0) - f(x_k))}{\|(x^0, f(x^0)) - (x_k, f(x_k))\|} \leq \epsilon_k \quad \forall k \in \mathbb{N}.$$

This implies

$$(x_k^*, -\alpha) \in \widehat{N}_{\epsilon_k}(\text{graph } f, (x_k, f(x_k)))$$

Thus applying limsup as k goes to infinity to the last assertion, we get

$$(x^*, -\alpha) \in N_M(\text{graph } f, (x^0, f(x^0))) \text{ which implies } x^* \in D_M^* f(x^0)(\alpha).$$

To prove the opposite inclusion, we pick $x^* \in D_M^* f(x^0)(\alpha)$ and find sequences $\epsilon_k \downarrow 0, x_k \rightarrow x^0, x_k^* \xrightarrow{w^*} x^*$ and $\alpha_k \rightarrow \alpha$ such that

$$(x_k^*, -\alpha_k) \in \widehat{N}_{\epsilon_k}(\text{graph } f, (x_k, f(x_k))) \quad \forall k \in \mathbb{N}.$$

But since the function is Lipschitz around x^0 and from the definition of ϵ -normals, we have

$$\langle x^*, x - x_k \rangle - \langle \alpha_k, f(x) - f(x_k) \rangle \leq 2\epsilon_k(1 + L)\|x - x_k\| \text{ for } x \in x_k + \eta_k B$$

With some sequence $\eta_k \downarrow 0$, where $L > 0$ is a Lipschitz constant of f around x^0 and B is a unit ball.

The later yields

$$x_k^* \in \widehat{N}_{\hat{\epsilon}_k}(A, x_k) \text{ with } \hat{\epsilon}_k = 2\epsilon_k(1 + L) + L\|\alpha_k - \alpha\|.$$

Since $\|\alpha_k - \alpha\| \rightarrow 0$, we have $\hat{\epsilon}_k \downarrow 0$ as $k \rightarrow \infty$ and hence $x^* \in N_M(\alpha f)(x^0)$. □

3.5 Extremal principles

Definition: A_1, A_2, \dots, A_n be non empty subsets of normed vector space E . A point $x \in \bigcap_{i=1}^n A_i$ is said to be a local extremal point of the system (A_1, A_2, \dots, A_n) if there exist sequences $(z_{1k}), \dots, (z_{nk})$ in E and a neighborhood U of x such that $z_{ik} \rightarrow 0$ as $k \rightarrow \infty$ for $i=1, \dots, n$ and

$$\bigcap_{i=1}^n (A_i - z_{ik}) \cap U = \emptyset \quad \text{for all sufficiently large } k \in \mathbb{N}.$$

In this case, $(A_1, A_2, \dots, A_n, x)$ is said to be an extremal system in E .

Definition:

(a) An extremal system $(A_1, A_2, \dots, A_n, x)$ in E is said to satisfy the exact extremal principle if there exist $x_i^* \in N_M(A_i, x)$, $i=1, \dots, n$, such that

$$x_1^* + \dots + x_n^* = 0 \quad \text{and} \quad \|x_1^*\| + \dots + \|x_n^*\| = 1. \quad (3.3.2)$$

(b) An extremal system $(A_1, A_2, \dots, A_n, x)$ in E is said to satisfy the approximate extremal principle if for every $\epsilon > 0$ there exist $x_i \in A_i \cap B_E(x, \epsilon)$ and $x_i^* \in N_F(A_i, x_i) + \epsilon B_{E^*}$, $i=1, \dots, n$, such that (3.3.2) holds, where $B_E(x, \epsilon) = \{y \in E: \|x - y\| \leq \epsilon\}$ and $B_{E^*} = \{y \in E^*: \|y\| \leq 1\}$.

3.6 Sequentially normally compact sets

Definition: The set $A \subseteq E$ is said to be sequentially normally compact (SNC) at $x \in A$ if for any sequence $((\epsilon_k, x_k, x_k^*))$ in $(0, +\infty) \times A \times E^*$ one has

$$[\epsilon_k \downarrow 0, x_k \rightarrow x, x_k^* \in \widehat{N}_{\epsilon_k}(A, x_k), x_k^* \xrightarrow{w^*} 0 \text{ as } k \rightarrow \infty] \Rightarrow x_k^* \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Lemma 3.2: (a) If E is a Frechet smooth Banach space, then the appropriate extremal principle holds for any extremal system $(A_1, A_2, \dots, A_n, x)$ in E , where $n \in \mathbb{N}$ and the sets A_1, A_2, \dots, A_n are closed.

(b) Let E be an Asplund space and $(A_1, A_2, \dots, A_n, x)$ an extremal system in E . Assume that the sets A_1, A_2, \dots, A_n are closed and all but one of them are SNC at x . Then the exact extremal principle holds for $(A_1, A_2, \dots, A_n, x)$.

proof

(a) The Proof of this lemma is given in [9] page 304-306.

(b) Let ϵ_k be a sequence of positive numbers such that $\epsilon_k \downarrow 0$ as $k \rightarrow \infty$. since ,by (a), the appropriate extremal principle holds for any extremal system $(A_1, A_2, \dots, A_n, x)$ for any $k \in \mathbb{N}$ and $i = 1, \dots, n$ there exist $x_{ik} \in A_i \cap B(x, \epsilon_k)$ and $x_{ik}^* \in N_F(A_i, x_{ik}) + \epsilon B_{E^*}$ Satisfying

$$x_{1k}^* + \dots + x_{nk}^* = 0 \text{ and } \|x_{1k}^*\| + \dots + \|x_{nk}^*\| = 1 \tag{3.6.1}$$

we have $x_{ik} \rightarrow x$ as $k \rightarrow \infty$ for $i = 1, \dots, n$. since for $i = 1, \dots, n$ the sequence (x_{ik}^*) for $k \in \mathbb{N}$ is bounded in the dual of the Asplund space E , this implies the subsequence of this sequence is weak convergent to some $x_i^* \in E^*$. Since $x_{ik}^* \in \widehat{N}_\epsilon(A_i, x_{ik})$ for any k , the definition of M -normal cones shows that $x_i^* \in N_M(A_i, x)$. It is evident that $x_1^* + \dots + x_n^* = 0$. It remains to show that x_i^* are not simultaneously zero.

By hypothesis we may assume that A_1, A_2, \dots, A_{n-1} are SNC. We now suppose that $x_i^* = 0$ for $i = 1, \dots, n-1$. Since $\|x_{nk}^*\| \leq \|x_{1k}^*\| + \dots + \|x_{n-1k}^*\|$ for all $k \in \mathbb{N}$, we must conclude, letting $k \rightarrow \infty$, that $x_n^* = 0$.

But this contradicts (3.6.1) □

Definition: The proper functional $f : E \rightarrow \mathbb{R}$ is said to be sequentially normally epi-compact (SNEC) at $x \in \text{dom} f$ if $\text{epi} f$ is SNC at $(x, f(x))$.

We will also make use of the following qualification condition:

$$[x_i^* \in \partial_M^\infty f_i(x), i = 1, \dots, n \text{ and } x_1^* + \dots + x_n^* = 0] \implies x_1^* = \dots = x_n^* = 0. \tag{3.4}$$

Lemma3.3: Assume A_1, \dots, A_n are closed subsets of Frechet smooth space E , that

$x \in \bigcap_{i=1}^n A_i$ and that $[x_i^* \in N_M(A_i, x), x_1^* + \dots + x_n^* = 0] \implies x_1^* = \dots = x_n^* = 0$. If each A_i is SNC at x , then so is $A_1 \cap \dots \cap A_n$.

The proof of this lemma is given in [9] page 314.

Definition: Assume that E and F are Frechet smooth Banach spaces. A is a non empty closed sub set of $E \times F$, and $(x,y) \in A$. The set A is said to be Partially sequentially normally compact (PSNC) at (x,y) with respect to E if for any sequences $(x_k, y_k) \rightarrow A$ and $(x_k^*, y_k^*) \in N_F(A, (x_k, y_k))$, one has

$$[x_k^* \xrightarrow{W^*} 0 \text{ and } \|y_k^*\| \rightarrow 0 \text{ as } k \rightarrow \infty] \implies \|x_k^*\| \rightarrow 0 \text{ as } k \rightarrow \infty.$$

The set A is said to be strongly partially sequentially normally compact (strongly PSNC) at (x,y) with respect to E if for any sequences $(x_k, y_k) \rightarrow A^{(x,y)}$ and $(x_{ik}^*, y_{ik}^*) \xrightarrow{W^*} (x_i^*, y_i^*)$ as $k \rightarrow \infty$ with $(x_{ik}^*, y_{ik}^*) \in N_F(A_i, (x_{ik}, y_{ik}))$, $i=1,2$, one has

$$[x_k^* \xrightarrow{W^*} 0 \text{ and } y_k^* \xrightarrow{W^*} 0 \text{ as } k \rightarrow \infty] \implies \|x_k^*\| \rightarrow 0 \text{ as } k \rightarrow \infty.$$

If the set A is SNC, then it is strongly PSNC and so PSNC.

Definition: Assume that E and F are Frechet smooth Banach spaces, A_1 and A_2 are closed sub set of $E \times F$, and $(x,y) \in A_1 \cap A_2$. The system (A_1, A_2) is said to satisfy the mixed qualification condition at (x,y) with respect to F if for any sequences $(x_{ik}, y_{ik}) \rightarrow A_i^{(x,y)}$ and $(x_{ik}^*, y_{ik}^*) \in N_F(A, (x_k, y_k))$, one has

$$[x_{1k}^* + x_{2k}^* \xrightarrow{W^*} 0 \text{ and } \|y_{1k}^* + y_{2k}^*\| \rightarrow 0 \text{ as } k \rightarrow \infty] \implies (x_1^*, y_1^*) = (x_2^*, y_2^*) = 0.$$

3.7 Calculus of Mordukovich subdifferentials

Theorem 3.5: (Sum rule of M-sub differentials): Let E be a Frechet smooth Banach space, let $f_1, \dots, f_n: E \rightarrow \mathbb{R} \cup \{\pm \infty\}$ be l.s.c and let

$x^0 \in \bigcap_{i=1}^n \text{dom } f_i$. Assume that all but one f_i are SNEC at $x^0 \in \bigcap_{i=1}^n \text{dom } f_i$ and that (3.4) is satisfied. Then

$$\partial_M(f_1 + \dots + f_n)(x) \subseteq \partial_M(f_1) + \dots + \partial_M(f_n). \quad (3.6)$$

If, in addition, f_1, \dots, f_n are lower all lower regular at x , then so is $f_1 + \dots + f_n$ and (3.6) holds with equality.

The proof of theorem is given in [9] page 317-319.

3.8 Calculus of Mordukhovich Normals

We need the following constraint qualification:

$$[x_i^* \in N_M(A_i, x), x_1^* + \dots + x_n^* = 0] \Rightarrow x_1^* = \dots = x_n^* = 0 \quad (3.7)$$

Intersection rule-1 for M-normal cones: Let E be a Frechet smooth Banach space, let A_1, \dots, A_n be non empty closed subsets of E , and let $x \in \bigcap_{i=1}^n A_i$. Assume that all but one of A_i are SNC at x and that condition (3.7) is satisfied. Then

$$N_M(A_1 \cap \dots \cap A_n, x) \subseteq N_M(A_1, x) + \dots + N_M(A_n, x). \quad (3.8)$$

If in addition, A_1, \dots, A_n are all normally regular at x , then so is $A_1 \cap \dots \cap A_n$ and (3.8) holds with equality.

Definition: Let A_1 and A_2 are closed sub set of the Frechet smooth Banach spaces Z , and let $z \in A_1 \cap A_2$. The system (A_1, A_2) is said to satisfy the limiting qualification condition at z if for any sequences $z_{ik} \rightarrow z$ and $z_{ik}^* \in N_F(A_i, z_{ik})$, one has

$$[\|z_{1k}^* + z_{2k}^*\| \rightarrow 0 \text{ as } k \rightarrow \infty] \Rightarrow z_1^* = z_2^* = 0.$$

Observe that the limiting qualification condition is special case of the mixed qualification condition with respect to E when $F = \{0\}$.

$$\text{Hence } N_M(A_1, z) \cap (-N_M(A_2, z)) = \{0\}$$

is sufficient for the system to satisfy the limiting qualification condition.

Intersection rule-2 for M-normal cones: Let E and F be Frechet smooth Banach spaces, A_1 and A_2 be closed sub sets of $E \times F$, and $(x, y) \in A_1 \cap A_2$. Assume that (A_1, A_2) satisfies the limiting qualification condition, A_1 is PSNC at (x, y) with respect to E , and A_2 is strongly PSNC at (x, y) with respect to F . Then one has

$$N_M(A_1 \cap A_2, (x, y)) \subseteq N_M(A_1, (x, y)) + N_M(A_2, (x, y)). \quad (3.10)$$

3.9 Optimality conditions

Let E and F be Frechet smooth Banach spaces and $f: E \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is a proper l.s.c functional.

We consider the problem

Min $f(x)$ subject to $x \in A$, where A is closed subset of E .

We now formulate hypotheses (H_1) and a qualification condition (Q_1)

(H_1) A is a closed subset of E , $x \in A$, A is SNC at x or f is SNEC at x .

(Q_1) $\partial_M^\infty f(x) \cap (-N_M(A, x)) = \{0\}$.

Proposition 3.4: Assume that (H_1) and (Q_1) are satisfied. If x is a local minimizer of f on A , then

$$0 \in \partial_M f(x) + N_M(A, x). \quad (3.7)$$

Proof

Since $\text{epi } \delta_A = A \times [0, +\infty)$, the functional δ_A is SNEC if (and only if) the set A is SNC. Moreover, we have $\partial_M^\infty \delta_A(x) = N_M(A, x)$ (Proposition 3.2 a).

Therefore, (Q_1) implies that the condition (3.4) is satisfied for f and δ_A .

Applying theorem (3.5) to (3.3) yields the assertion. \square

We now formulate assumptions:

(H_2) For $i=1, \dots, r$ the set $A_i \subseteq E$ is closed and SNC at $x \in \bigcap_{i=1}^r A_i$.

(Q_2) $[x^* \in \partial_M^\infty f(x), x_i^* \in N_M(A_i, x), x^* + x_1^* + \dots + x_r^* = 0] \Rightarrow x^* = x_1^* = \dots = x_r^* = 0$.

Proposition 3.5 : Let the hypotheses (H_2) and the qualification condition (Q_2) be satisfied. If x is a local minimizer of f on $\bigcap_{i=1}^r A_i$, then one has

$$0 \in \partial_M f(x) + N_M(A_1, x) + \dots + N_M(A_r, x).$$

Proof. We verify the assertion for $r = 2$; it then follows for $r \geq 2$ by induction.

We want to apply Proposition 3.4 to $A := A_1 \cap A_2$. From (Q2) with $x^* := 0$

we obtain

$$N_M(A_1, x) \cap (-N(A_2, x)) = \{0\}. \quad (3.11)$$

This and (H₂) imply, by lemma 3.3, that $A_1 \cap A_2$ is SNC at x . By Intersection rule-2 for M-normal the condition (3.11) also implies that

$$N_M(A_1 \cap A_2, x) \subseteq N_M(A_1, x) + N_M(A_2, x). \quad (3.12)$$

Now we convince ourselves that condition (Q1) holds. Take any $x^* \in \partial_M^\infty f(x)$ such that $x^* \in N_M(A_1 \cap A_2, x)$. By (3.12) there exist $x_i^* \in N_M(A_i, x), i = 1, 2$, such that

$-x^* = x_1^* + x_2^*$. Hence $x^* = 0$ by (Q₂). Referring to Proposition 3.4 and (3.12) completes the proof. \square

Now consider

$$f_i(x) \leq 0, i=1,2,\dots,r$$

$$f_i(x) = 0, i= r+1,\dots,r+s, \quad (3.13)$$

$$x \in A,$$

where $f_i: E \rightarrow \mathbb{R}$ for $i= 1,\dots, r+s$.

Min $f(x)$

Subject to constraints (3.13)

(H₃) The functions f_1,\dots,f_{r+s} are continuous.

All but one of the sets $\text{epi } f, \text{epi } f_i (i=1,\dots,r), \text{graph } f_i (i=r+1,\dots,r+s),$ and A are SNC at $(x^0, f(x^0)), (x^0, 0)$ and x^0 respectively.

Theorem 3.13: Let the assumption (H₃) be satisfied. Assume that x^0 is a local minimizer of f subject to constraints (3.13).

a) There exist

$$(x^*, -\lambda) \in N_M(\text{epi } f, ((x^0, f(x^0))), \quad y^* \in N_M(A, x^0),$$

$$(x_i^*, -\lambda_i) \in N_M(\text{epi } f_i, (x^0, 0)), \quad i=1, \dots, r,$$

$$(x_i^*, -\lambda_i) \in N_M(\text{graph } f_i, (x^0, 0)), \quad i= r+1, \dots, r+s$$

Satisfying

$$x^* + x_1^* + \dots + x_{r+s}^* + y^* = 0,$$

$$\|(x^*, \lambda)\| + \|(x_1^*, \lambda_1)\| + \dots + \|(x_{r+s}^*, \lambda_{r+s})\| + \|y^*\| = 1, \quad (3.14)$$

$$\lambda_i f_i(x^0) = 0, \quad i=1, \dots, r.$$

b) Assume, in addition, that the functions f, f_1, \dots, f_{r+s} are locally L -continuous around x^0 .

Then there exist non-negative real numbers $\lambda, \lambda_1, \dots, \lambda_{r+s}$ such that $\lambda_i f_i(x^0) = 0$ for $i=1, \dots, r$ and

$$0 \in \lambda \partial_M f(x^0) + \sum_{i=r+1}^{r+s} \lambda_i (\partial_M f(x^0) \cup \partial_M (-f_i)(x^0)) + N_M(A, x^0).$$

Proof

a) Define the following subsets of $E \times \mathbb{R}^{r+s+1}$:

$$B := \{(x, \alpha, \alpha_1, \dots, \alpha_{r+s}) \mid \alpha \geq f(x)\},$$

$$B_i := \{(x, \alpha, \alpha_1, \dots, \alpha_{r+s}) \mid \alpha_i \geq f_i(x)\}, \quad i=1, \dots, r,$$

$$B_i := \{(x, \alpha, \alpha_1, \dots, \alpha_{r+s}) \mid \alpha_i = f_i(x)\}, \quad i= r+1, \dots, r+s,$$

$$B_{r+s+1} := A \times \{0\}.$$

Without loss of generality we may assume that $f(x^0) = 0$. Then $(x^0, 0) \in E \times \mathbb{R}^{r+s+1}$ is the local external point of the system of closed sets B, B_1, \dots, B_{r+s+1} . Therefore, the hypothesis (H_3) , by lemma 3.2 (b), ensures that the exact extremal principle holds for the above system, which immediately yields the assertion except for the equation $\lambda_i f_i(x^0) = 0, i=1, \dots, r$. Assume we have $f_i(x^0) < 0$ for some $i \in \{1, \dots, r\}$. Then by continuity, it

follows that $f_i(x) < 0$ for all x in the neighborhood of x^0 . Hence $(x^0, 0)$ is an interior point of $\text{epi } f_i$. It follows that $N_M(\text{epi } f_i, (x^0, 0)) = \{0\}$ and so $\lambda_i = 0$.

b) By definition of M -subdifferential and by lemma 3.1 we have

$$(x^*, -\lambda) \in N_M(\text{epi } f, (x^0, f(x^0))) \Leftrightarrow x^* \in \lambda \partial f(x^0) \text{ and } \lambda \geq 0.$$

Moreover, the definition of coderivative and proposition 3.3 imply that

$$(x^*, -\lambda) \in N_M(\text{graph } f, (x^0, f(x^0))) \Leftrightarrow x^* \in D_M^* f(x^0)(-\lambda) = \partial_M(\lambda f)(x^0).$$

Notice that $\partial_M(\lambda f)(x^0) \subseteq |\lambda|(\partial_M f(x^0) \cup \partial_M(-f)(x^0))$ for any $\lambda \in \mathbb{R}$. The assertion now follows from (a). \square

4. References

1. B. S. Mordukhovich, Variational analysis and generalized differentiation, I: Basic theory, Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol. 330, Springer, Berlin, 2006.
2. B. S. Mordukhovich, Variational analysis and generalized differentiation, II: Applications, Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol. 331, Springer, Berlin, 2006.
3. B. S. Mordukhovich, Variational analysis and generalized differentiation in optimization and Control, vol. 47, Springer Science+Business Media, New York, 2010.
4. D. P. Bertsekas, Convex analysis and optimization, Athena scientific, Belmont, Massachusetts, USA, 2003.
5. F. H. Clarke, Optimization and non smooth analysis, Society for Industrial and Applied Mathematics, Philadelphia, 1990.
6. G. Giorgi, A. Guerraggio and J. Theirfeleder, Mathematics of optimization: smooth and non smooth case, Elsevier Science and Technology, 2004.
7. J. Jahn, Introduction to the theory of nonlinear optimization, 2nd revised edition, Springer-Verlag Berlin Heidelberg, New York, 1996.
8. R.T. Rockafellar and R. J-B Wets, Variational analysis, Springer-Verlag Berlin Heidelberg, New York, 1998.
9. W. Schirotzek, Nonsmooth Analysis, Springer-Verlag Berlin Heidelberg, 2007.