OPTIMALITY CONDITIONS FOR NONSMOOTH OPTIMIZATION

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF MATHEMATICS

“In partial fulfillment of the requirements for the degree of master of science in mathematics”

By: Tadesse Birkie

Stream: Optimization

Advisor: Berhanu Guta (PhD)

Addis Ababa, Ethiopia

2017

ADDIS ABABA UNIVERSITY

DEPARTMENT OF MATHEMATICS

The undersigned hereby that they have read and recommend to the department of mathematics for acceptance of this project entitled "optimality conditions for non-smooth optimization problem" by Tadesse Birkie Betew in partial fulfillment of the requirements for the degree of master of science in mathematics.

Advisor: Dr. Berhanu Guta

Signature:……………………………..

Date: …………………………………

Examiner 1: Dr.

Signature:………………………………….

Date: ………………………………………

Examiner 2: Dr.

Signature:………………………………….

Date: ………………………………………

ADDIS ABABA UNIVERSITY

Author: Tadesse Birkie

Title: OPTIMALITY CONDITIONS FOR NON-SMOOTH OPTIMIZATION PROBLEM

Department: Mathematics

Degree: M.Sc.

Convocation: June

Year: 2017

Tadesse Birkie

Signature:……………………..

Date: ……………………………..

Acknowledgments

I would like to express my gratitude to my advisor, Dr. Berhanu Guta, for all his dedication, patience, and advice. Also, I would like to thank all my Instructors for their motivation and guidance during the past three years at Addis Ababa University, department of Mathematics. My thank also forward to my brother Aderajew Birkie and my friend kibremarkos Aklilu for all their support.

ABSTRACT

Optimization is a mathematical problem with many real world applications. The goal is to determine minimizers or maximizers of a multivariable real function, under a restriction domain. The differentiability assumptions play a vital role in nonlinear programming, because most of methods of finding the optimum point in nonlinear 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 functions 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 sub gradient of the directional derivative. 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 functions as well as functions in constrains are differentiable. The main purpose of this project focuses on finding the optimality conditions for optimizations problems without any differentiability assumptions. The sub-gradient of a directional derivative approach are used to solve nonsmooth optimization problem of convex type. we establish the existence of optimization problem specially for non-linear programing (NLP) ,We introduce “A cone Approach on the Karush-Kuhn-Tucker optimality conditions, constraint qualification” and we discussed in this paper some of the constraint qualifications as well as some relation between them and also to show and also to observe the weakest of these constraint qualifications with respect to the concept of cones and their polar.

Key words and phrases: Optimality conditions, constraint qualifications, nonlinear programming, unconstrained optimization, positive definiteness, Hessian matrix, smooth optimization, nonsmooth optimization, directional derivative, sub- gradient, sub-differentiable, polar cones, the tangent cone, Slater constraint qualification, and the Quasi-Regularity constraints.

List of Notations:

$\nabla f(x)$	*Gradient of real valued functions f*
R	*set of real numbers*
R^n	n *dimensional eculidean spaces*
$\frac{\partial f}{\partial x}$	*Partial derivative of f with respect to x.*
H(x)	*Hessian matrix of a function at x.*
L	*Lagrangian function:*
$L(.,\lambda,\mu)$	*Lagrangian* function *with Lagrange multipliers* λ *and* μ.
$<\lambda, h>$	*inner product of* λ *and h*
$f \in c'$	*f is once continuously differentiable function*
$f \in c''$	*f is twice continuously differentiable function*
p_{un}	*Unconstrained optimization problems.*
p_c	*constrained optimization problems.*
CQ	*constraint qualification*
Ω	*feasible set of the problem*
P(s)	*the polar of S*
T(x)	*the tangent cone at* $x \in \Omega$.
D(x)	*the set of first order feasible variation at a point* $x \in \Omega$.
LICQ	*Linear independence constraint qualification.*
MFCQ	*mangasarian-Fromovitz constraint qualification.*
NSO	*Nonsmooth optimization problems.*
NDO	*Nonsmooth non differentiable optimization problems*

Content

page

Introduction

Optimization is a mathematical problem with many real world applications. The goal is to determine minimizers or maximizers of a multivariable real function, under a restriction domain. A function is smooth if it is differentiable and the derivatives are continuous. Moreover specifically, this is first-order smoothness. Second –order smoothness means that second derivatives exist and are continuous, and so forth, while infinite smoothness refers to continuous derivatives of all orders. From this perspective a nonsmooth functions only has a negative description it lacks some degree of properties traditionally relied upon in analysis. One could get the impression that “nonsmooth optimization” is a subject dedicated to overcoming handicaps which have to be faced in miscellaneous circumstances where mathematical structure might be poorer than what one would like. But this is far from right.

There are many situations in operational research where one has to be Optimize a function which fails to have derivative for some values of the variables. This is what Non differentiable Optimization (NDO) or Nonsmooth optimization (NSO) deals with. For this kind of situation, new tools are required to replace standard differential calculus, and these new tools come from convex analysis.

In Nonsmooth optimization (NSO) functions don’t need to be differentiable. The general problem is that we are minimizing functions that are typically not differentiable at their minimizers this type of problems arise in many field of applications such as Economics, Mechanics, Engineering Computational Chemistry, Biology, Optimal control and Data mining.

Some of the cause of nonsmooth optimization is inherent which is original phenomenon contains various discontinuities and irregularities. It may also cause by technological that is caused by some extra technological constraints which may cause a nonsmooth dependence between variables and functions. Methodological that is some algorithms for constrained optimizations may lead to a nonsmooth problem for example, the exact penalty function method and numerical is the so called “stiff problem” which is analytically smooth but numerically unstable and behave like nonsmooth problems. This project paper that contains three main focus areas such as:

- Chapter- one, we try to discuss some basic concepts of convexity , convex function and separation properties.
- Chapter- two, discusses, on smooth optimization problems, throughout the chapter, we try to describe some basic concepts and properties of the methods for nonlinear optimization problem .we apply differential of the functions.
- Chapter - three, focuses on nonsmooth optimization problems which cannot be differentiable functions.

CHAPTER ONE

PRELIMINARY

1.1. Convex analysis and convex functions

The concept of convexity is important in the study of optimization problems. In order to decide whether an optimization problem can be solved with tools from convex analysis or not, the question about convexity must be answered. As the minimization of the objective function can be restricted to a certain set, convexity must be verified for the objective function and the constraints. We start the section about convex analysis by considering the notion of convex sets. Considering convex sets is crucial for convex functions, as there is a strong relation between them. A function is convex if the set of points lying above the graph of the functions is convex. Therefore; we first define convexity of sets.

Definition 1.1: (convex sets)

Next, is the notion of a convex set

1. A subset S $\subseteq \mathbb{R}^n$ is said to be convex, ifλx + (1-λ)y ∈S whenever x,y ∈S and 0≤ λ ≤1.

2. Let x,y∈S. Then the set of all convex combination of x and y is the set of points $\{w_x$ ∈S:$w_x = \lambda$x + (1-λ)$y \in S$,0≤ λ ≤1} ………..(1.1)

In, say, R^2, this set is exactly the line segment joining the two points x and y.

Definition 1.2: Let k⊆S. Then the set k is said to be convex provided that given two points x,y ∈ k, the set (1.1) is a subset of k.

We give some examples:

a. An interval of $[a, b] \subset \mathbb{R}$ is a convex set. To see this, let $c, d \in [a, b]$ and assume, with out loss of generality, that $c < d$. Let λ ∈(0,1). Then,

$$
\begin{aligned}
a \le c = \lambda c + (1-\lambda)c &< \lambda d + (1-\lambda)c \\
&< \lambda d + (1-\lambda)d \\
&= d \le b
\end{aligned}
$$

b. Let A be an mxn matrix, b$\in \mathbb{R}^m$, and Let s :={ x$\in \mathbb{R}^n$: Ax=b}. (The set s is just the set of all solutions of the linear equation Ax=b.) , Then

$$A((1-\lambda)\, x^1 + \lambda x^2\,) = (1-\lambda)A\, x^1 + \lambda A x^2$$

$$= (1-\lambda)b + \lambda b$$

$$= b-\lambda b + \lambda b$$

$$= b$$

C. In $\mathbb{R}^n$ the set H :={x$\in \mathbb{R}^n$:$a_1 x_1$+…+$a_n x_n$ =c} is a convex set. For any particular choices of constants a_i it is a hyper plane in $\mathbb{R}^n$. It defines equation of a generalization of the usual equation of a plane in $\mathbb{R}^3$, namely the equation ax+by+cz+d=0

To see that H is convex set, Let x^1 ,x^2 $\in$ H and define z$\in \mathbb{R}^3$ by

Z: = (1-λ) x^1+ λx^2 .Then

$$Z: = \sum_{i=1}^{n} a_i[(1-\lambda)\, x_i^1 + \lambda x_i^2]$$

$$= \sum_{i=1}^{n}(1-\lambda)\, a_i x_i^1 \ + \lambda a_i x_i^2]$$

$$= (1-\lambda) \sum_{i=1}^{n} a_i\, x_i^1 + \lambda \sum_{i=1}^{n} a_i\, x_i^2$$

$$= (1-\lambda)c + \lambda c$$

$$= c - \lambda c + \lambda c$$

$$= c$$

Proposition 1.1: If C$\subset \mathbb{R}^n$ is convex,then the cl(C), the closure of C, is also convex.

Proof:

Suppose x,y$\in$cl(C). Then there exists sequences $\{x_n\}$ and $\{y_n\}$ in C such that $x_n \to$x and $y_n \to$y as n$\to \infty$**.** For some λ , 0$\leq \lambda \leq$1, define $z_n := \lambda x_n$ + (1-λ)y_n. Then, by convexity, z_n $\in$C. Moreover $z_n \to$ λx + (1-λ)y as n$\to \infty$. Hence this point lies in cl(C).

Proposition 1.2: The intersection of any number of convex sets is convex.

Proof:

Let $\{k_\alpha\}_{\alpha\in A}$ be a family of convex sets, and Let $K:=\bigcap_{\alpha\in A} k_\alpha$.Then, for any x,y$\in k_\alpha$ for all $\alpha \in$ A and each of these sets is convex. Hence, for any $\alpha \in$ A, and $\lambda \in[0,1]$, $\lambda x + (1-\lambda)y \in k_\alpha$. Hence, $\lambda x + (1-\lambda)y \in K$.

Proposition 1.3: Let c, c_1 and c_2 be convex sets in $\mathbb{R}^n$ and let $\beta\in \mathbb{R}$, then

a, $\beta C:=\{z\in \mathbb{R}^n : z= \beta x, x\in C\}$ is convex.

b, $c_1+c_2:= \{z\in \mathbb{R}^n : z=x_1+x_2, x_1 \in c_1, x_2 \in c_2\}$ is convex.

Proof:

a, Let x,y $\in \mathbb{R}^n$ and $0\leq \lambda \leq 1$.

$\beta(\lambda x + (1-\lambda)y) = \{z\in \mathbb{R}^n : z= \beta(\lambda x + (1-\lambda)y), x, y \in C\}$

$= \{z\in \mathbb{R}^n : z= \lambda(\beta x) + (1-\lambda)(\beta y), x, y \in C\}$

$= \lambda c + (1-\lambda)c$

$=c$

b, Let $z_1,z_2 \in c_1+c_2$ and take $0\leq \lambda \leq 1$.

We take $z_1 = x_1+x_2$, with $x_1 \in c_1$, $x_2 \in c_2$ and like wise decompose $z_2 = y_1 + y_2$ with $y_1 \epsilon c_1$and$y_2 \epsilon c_2$.Then

$(1-\lambda)z_1 + \lambda z_2 = (1-\lambda)[x_1+x_2]+ \lambda[y_1+y_2]$

$= [(1-\lambda)x_1 + \lambda y_1]+ [(1-\lambda)x_2+ \lambda y_2] \in c_1+c_2$, Since the set c_1 and c_2 are convex. While, by definition, a set is convex provided all convex combinations of two points in the set is again in the set.

Definition 1.3: The **convex hull** of a set c is the intersection of all convex sets which contain the set c, denote the convex hull by con(c).

Corollary 1.1: The convex hull of a compact set in $\mathbb{R}^n$ is compact.

Proof: Let c$\subset \mathbb{R}^n$ be compact. Notice that the

$\sigma \coloneqq \{(\lambda_1, \lambda_2, \ldots, \lambda_n)\} \in \mathbb{R}^n : \sum_i^n \lambda_i = 1$ is also closed and bounded and is therefore compact.

Now, suppose that $\{v^j\}_{j=1}^{\infty} \subset$ con(c). By Caratheodory's theorem, each v^j can be written in the form: $v^{(k)} := \sum_{i=1}^{n+1} \lambda_{k,i} x^{(k,i)}$, where $\lambda_{k.i} \geq 0$, $\sum_{i=1}^{n+1} \lambda_{k,i} = 1$, and $x^{(k,i)} \in$ c. Then , since c and σ are compact, there exists a sequence, $k_1, k_2, \ldots$ such that the limits $\lim_{k\to\infty} \lambda_{k.i} = \lambda_i$ and $\lim_{k\to\infty} x^{(k,i)} = x^{(i)}$ exists for i=1,2,…,n+1. Clearly, $\lambda_i \geq 0$, $\sum_{i=1}^{n+1} \lambda_i = 1$ and $x_i \in$ c. Thus, the sequence $\{v^j\}_{j=1}^{\infty}$ has a subsequence, $\{v^{(j,k)}\}_{j=1}^{\infty}$ which converges to a point of con(c). Hence it is compact.

Definition 1.4 (convex function)

1. Let S be a non-empty convex set in $\mathbb{R}^n$, The function f:S $\to \mathbb{R}$ is said to be convex on S if $f(\lambda x + (1-\lambda)y) \leq \lambda$f(x) +(1-$\lambda$)f(y), for each x,y∈S and for each $\lambda \in [0,1]$

2. The function f is said to be strictly convex on S if the above inequality holds a strict inequality for each distinct x,y $\in S$ and for each $\lambda \in [0,1]$.
3. The function f is said to be concave (strictly concave) if –f is convex (strictly convex).

Theorem .1.1

If f is convex and f $\in c'$ over an open convex set S, then f(y)≥ f(x)+ $\nabla f(x)^T$(y-x), for all x,y ∈S…………(1)

If in addition f $\in c''$ over S, then the Hessian of f is positive semi definite, $\nabla^2 f(x) \geq 0$,

for all x∈ S

Conversely, if f $\in c'$ over S and (1) holds, or if f $\in c''$over S and $\nabla^2 f(x) \geq 0$, for all $x \in S$, then f is convex over S.

Proposition 1.5: If h(x) is convex and non-negative on a convex set X, then $h^{2}(x)$ is also convex on X.

Proof:

Suppose $x, y \in X$ and $\lambda \epsilon [0,1]$, then

$$h^2(\lambda x + (1-\lambda)y) = [h(\lambda x + (1-\lambda)y)][h(\lambda x + (1-\lambda)y)]$$

$$\leq (\lambda h(x) + (1-\lambda)h(y))(\lambda h(x) + (1-\lambda)h(y))$$

$$= \lambda h^2(x) + (1-\lambda)h^2(y) - \lambda(1-\lambda)(h(x)- h(y))$$

$$\leq \lambda h^2(x) + (1-\lambda)h^2(y)$$

Proposition 1.6: If f is a convex function on a convex set X, then the function $f^{+}(x) = \max\{f(x), 0\}$ is also convex on x.

Proof:

Suppose $x, y \in X$ and $\lambda \in [0,1]$, then

$$f^{+}(\lambda x + (1-\lambda)y) = \max\{f(\lambda x + (1-\lambda)y), 0\}$$

$$\leq \max\{\lambda f(x) + (1-\lambda)f(y), 0\} \qquad \text{(Since f is convex)}$$

$$\leq \max\{\lambda f(x), 0\} + \max\{(1-\lambda)f(y), 0\}$$ (By property of maximum function)

$$= \lambda \max\{f(x), 0\} + (1-\lambda)\max\{f(y), 0\}$$

$$= f^{+}(x) + (1-\lambda) f^{+}(y)$$

1.2. Separation properties

There are a number of results which are of crucial important in the theory of convex sets and in the theory of mathematical optimization, particularly with regard to the development of **necessary conditions**, as for example in the theory of Lagrange multipliers. We will discuss these results which will be useful to us. We will confine our attention to the finite dimension case. In order to study the separation properties of convex set. We need the notion of a hyper plane.

Definition 1.5: Let a$\in \mathbb{R}^n$, and b$\in \mathbb{R}$ and assume a$\neq 0$.Then the set

H:={x$\in \mathbb{R}^n$:$\langle a, x\rangle$ =b } is called a hyper plane with normal vector a.

We note, in passing, those hyper planes are convex sets. Indeed, if x,y∈H and $0\leq \lambda \leq 1$ we have :

$$a, \left\langle a, (1-\lambda)x + \lambda y \right\rangle = (1-\lambda)\left\langle a, x \right\rangle + \lambda\left\langle a, y \right\rangle$$

$$= (1-\lambda)b + \lambda b$$

$$= b-(\lambda)b + \lambda b$$

Each such hyper plane defines the half-spaces.

H^+ :={x$\in \mathbb{R}^n$:$\left\langle a, x \right\rangle \geq$b} , and H^- :={x$\in \mathbb{R}^n$:$\left\langle a, x \right\rangle \leq$b}

Note that $H^+ \cap H^-$ =H. These two half-spaces are closed sets, that is, they contain all their limit points. Their interiors are given by

H^{0+}:={x$\in \mathbb{R}^n$: $\left\langle a, x \right\rangle >$ b } , and H^{0-} :={x$\in \mathbb{R}^n$: $\left\langle a, x \right\rangle <$b } whether closed or open, these half-spaces are said to be generated by the hyper plane H. Each is convex set. Of critical importance to optimization problems is a group results called **separation theorems**

Definition 1.6: Let S,T$\in \mathbb{R}^n$ and Let H be a hyper plane. Then H is said to separate S from T if S lies in one closed half-space determine by H while T lies in the other closed half-spaces. In this case H is called a separation hyper plane. If S and T lie in the open half-space, then H is said to strictly separate S and T.

Theorem 1.2: Let $C \subset \mathbb{R}^n$ be convex and suppose that $y \notin$ cl(c). Then there exist an a$\in \mathbb{R}^n$ and a number $\gamma \in \mathbb{R}$ such that : $\langle a, x \rangle > \gamma$ for all x∈C and :$\langle a, y \rangle \leq \gamma$.

Proof: Let $\hat{c}$ be the projection of y on cl(c), and let $\gamma = \inf_{x \in c} ||x - y||$, that is, γ is distance from y to its projection $\hat{c}$. Note since $y \notin$ cl(c), $\gamma > 0$.

Now, choose an arbitrary x∈c and $\lambda \in (0,1)$ and form $x_\lambda := (1 - \lambda)\hat{c} + \lambda x$

Since, x∈c and $\hat{c} \in$cl(c), $x_\lambda \in$cl(c).

So, we have $||x_\lambda - y||^2 = ||(1-\lambda)\hat{c} + \lambda x - y||^2 = ||\hat{c} - \lambda\hat{c} + \lambda x - y||^2 \geq ||\hat{c} - y||^2 > 0$

But, we can write $||\hat{c} - \lambda\hat{c} + \lambda x - y||^2 = ||(\hat{c} - y) + \lambda(x - \hat{c})||^2$ and we can expand this expression using the rules of inner products.

$$0 < ||\hat{c} - y||^2 \leq \left\|(\hat{c} - y) + \lambda(x - \hat{c})\right\|^2 = \langle (\hat{c} - y) + \lambda(x - \hat{c}), (\hat{c} - y) + \lambda(x - \hat{c}) \rangle$$

$$= \langle \hat{c} - y, \hat{c} - y \rangle + \langle \hat{c} - y, \lambda(x - \hat{c}) \rangle + \langle \lambda(x - \hat{c}), \hat{c} - y \rangle + \langle \lambda(x - \hat{c}), \lambda(x - \hat{c}) \rangle$$

And from this inequality, we deduce that $2\lambda \langle (\hat{c} - y), (x - \hat{c}) \rangle + \lambda^2 \langle (x - \hat{c}), (x - \hat{c}) \rangle \geq 0$

From this last inequality, dividing both sides by 2λ and taking a limit as $\lambda \to 0^+$, we get $\langle \hat{c} - y, x - \hat{c} \rangle \geq 0$

Again, we can expand the last expression

$$\langle \hat{c} - y, \ x - \hat{c} \rangle = \langle \hat{c} - y, \ x \rangle + \langle \hat{c} - y, \ -\hat{c} \rangle \geq 0$$

By adding and subtracting y and recalling that $\|\hat{c} - y\| > 0$, we can make the following estimate, $\langle \hat{c} - y, x \rangle \geq \langle \hat{c} - y, \hat{c} \rangle = \langle \hat{c} - y, y - y + \hat{c} \rangle$

$$= \langle \hat{c} - y, y \rangle + \langle \hat{c} - y, \hat{c} - y \rangle$$

$$= \langle \hat{c} - y, y \rangle + ||\hat{c} - y||^2$$

$$> \langle \hat{c} - y, y \rangle$$

In summary, we have $\langle \hat{c} - y, x \rangle > \langle \hat{c} - y, y \rangle$ for all $x \in c$.

Finally, define $a := \hat{c} - y$. Then this last inequality reads $\langle a, x \rangle > \langle a, y \rangle$ for all $x \in c$.

CHAPTER TWO

2. SMOOTH OPTIMIZATION PROBLEMS

In this subsection, we shall mainly understand the optimality conditions for various types of extremum problems, under differentiability assumptions of the functions involved in the problems, we shall treat, separately necessary and sufficient optimality conditions.

Definition: 2.1.

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

Definition: 2.2. Let X be a normed vector space and $f: X \to \mathbb{R}$, we say that:

i. f has a local minimum at x* if and only if there exists a neighborhood X of x* such that $f(x^*) \leq f(x)$, for all $x \in X$.

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

2.1. Methods for unconstrained and constrained smooth optimization Problems.

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

$$\begin{aligned} &\text{Min } f(x) \\ &\text{s.t. } x \in \mathbb{R}^n \end{aligned} \qquad (p_{un})$$

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } x \in S \end{aligned} \qquad (p_c)$$

where, S={x: x∈ $\mathbb{R}^n$, $g_i(x) \leq 0$, i=1,…,m,h_j(x)=0, j=1,…,r }is_called_the_constraint_set_and $X \subseteq \mathbb{R}^n$ is any set, f, g_i (i=1,…,r < n) are real- valued functions, all defined and differentiable on $\mathbb{R}^n$.

Remark: Problem (p_{un}) is called unconstrained optimization problem, whereas (p_c) is constrained optimization problem.

2.2. Method for unconstrained smooth optimization

In this subsection, for unconstrained smooth optimization, we have the following conditions:

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

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

iii. If the Hessian matrix is positive definite at x^0 and if x^0 is candidate

or $\nabla^2 f(x^0) > 0$, then x^0 is a minimizer. To show this, we see the following examples.

Examples 2.1

Min f(x) = $(x_1 - 3)^2+(x_2 - 4)^2$

x∈ R^n

Solution:

$\nabla f(x)=0$

$$\Rightarrow \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 2(x_1 - 3) \\ 2(x_2 - 4) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$x^o=\begin{pmatrix} 3 \\ 4 \end{pmatrix}$ is a stationary point (critical point).

Now, the Hessian matrix:

$\nabla^2 f(x^o) = \begin{vmatrix} f_{x_1x_1} & f_{x_1x_2} \\ f_{x_2x_1} & f_{x_2x_2} \end{vmatrix}_{x^o=\binom{3}{4}}$ since, $fx_1x_2 = fx_2x_1$ is symmetric matrix.

$= \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix}$

$= 4 > 0$ is positive definite

$\therefore x^o = \binom{3}{4}$ is a minimizer.

Example 2.2

Min $f(x, y) = y^2 - 3x^2y + 2x^4$

$(x, y) \in \mathbb{R}^2$

Solution:

$\nabla f(x, y) = 0$

$$\begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Longrightarrow \begin{pmatrix} -6xy + 8x^3 \\ 2y - 3x^2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$x^o = \binom{0}{0}$ is a stationary point (critical point)

Now, the Hessian matrix:

$\nabla^2 f(x, y) = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix}_{x^o=\binom{0}{0}}$

$= \begin{vmatrix} 0 & 0 \\ 0 & 2 \end{vmatrix}$

$= 0 \geq 0$ is positive semi definite

$\therefore x^o = \binom{0}{0}$ isn't a minimizer.

Example 2.3

Min $f(x)=100(x_1 - x_2)^2+(1-x_1)^2$

$x\in \mathbb{R}^2$

Solution:

$\nabla f(x)=0$

$$\Rightarrow \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

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

Now, the Hessian matrix: $\nabla^2 f(x^o) = \begin{vmatrix} f_{x_1x_1} & f_{x_1x_2} \\ f_{x_2x_1} & f_{x_2x_2} \end{vmatrix}_{x^o=\begin{pmatrix} 1 \\ 1 \end{pmatrix}}$

$$= \begin{vmatrix} 202 & -200 \\ -200 & 200 \end{vmatrix}$$

$$= 40{,}400-40{,}000>0$$

$$=400> 0 \text{ is positive definite}$$

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

2.3. To solve constrained types of problems

In this subsection, for (p_c) we use Lagrange function to transform the problems in to unconstrained form, That is, we define

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

$$(x, \lambda, \mu) \in R^n \times R^m_+ \times R^r$$

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

Then we have new unconstrained optimization problems:

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

$$(x, \lambda, \mu) \in R^n \times R_+^m \times R^r$$

Because of the assertion, if x^o is a solution of (p_λ, μ), then x^o is a solution p_c

Solving (p_λ, μ) is similar to solving p_c

For solving problem (p_λ, μ),the most well- known optimality conditions 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|$g_i(x) = 0$}.

Definition 2.3

A feasible vector x∈S is said to be **regular**, if the gradients ∇h_j(x), j=1,…,r of equality constraint and the gradient ∇g_i(x), i ∈ A(x) of the active inequality constraints are linearly independent.

2.4. Karush-Kuhn-Tucker (KKT) necessary optimality conditions for constrained smooth optimization.

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

$\lambda^* = (\lambda^*_1, \lambda^*_2, \dots, \lambda^*_m)$

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

$\lambda^*_i g_i(x^*) = 0$ for all i∈{1,2,…,m}, $\lambda^*_i \geq 0$

$h_j(x^*) = 0$, for all j=1,2,…,r

$g_i(x^*) \leq 0$, for all 1≤ i≤m

Under suitable convex assumption the KKT conditions are also sufficient optimality condition S.

Example 2.4

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

$x\in S$

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

Solution:

To transfer Lagrange form, we have

$L(x,\lambda)=f(x_1,x_2)+\lambda_1(x_1+x_2-2)+\lambda_2(x_1+2x_2-3)\rightarrow \min$, $(x,\lambda)\in R^2\times R_+^2$

Then we have to solve the following system of equations:

Consider the following unconstrained optimization problem

Min $L(x,\lambda)$ $\quad\quad$ (p_λ)

$(x,\lambda)\in R^2\times R_+^2$

To solve this problem, First check regularity condition, whether it is linearly independent or not, since, the gradient active inequality constraints, namely,

$\{\partial_x(x_1+x_2-2),\partial_x(x_1+2x_2-3)\}=\{\binom{1}{1},\binom{1}{2}\}$

Where, $x=\binom{x_1}{x_2}$

Let $\beta,\alpha\in\mathbb{R}$

Since, $\alpha\binom{1}{1}+\beta\binom{1}{2}=\binom{0}{0}$

$\alpha=\beta=0$

$\binom{1}{1} and \binom{1}{2}$ 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 ,$\frac{\partial L}{\partial x1}$=$2x_1 - 14 + \lambda_1 + \lambda_2$=0

2 ,$\frac{\partial L}{\partial x2}$=$2x_2 - 6 + \lambda_1 + 2\lambda_2 = 0$

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

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

5 ,$x_1 + x_2 \leq 2$

6 ,$x_1 + 2x_2 \leq 3$

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

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

2.5. Necessary KKT optimality conditions

In this subsection, constrained optimization

Let f:$\mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$ for j=1,…,m and

h_i:$\mathbb{R}^n \to \mathbb{R}$ for i=1,….,r

Unconstrained problem	Constrained problem
Minimum f(x)	minimum f(x)

No restriction x$\in \mathbb{R}^n$ subject to $g_j \leq 0$, j=1,…,m

h_i (x) = 0, i=1,..,r

For constrained problem, we say that x feasible if it satisfies all the constraints of the problem, i.e. g_j (x)$\leq$ 0, j=1,…,m h_i(x)=0, i=1,…,r, In fact it is linearly constrained

Necessary optimality conditions:

Let f:$\mathbb{R}^n \to \mathbb{R}$ be continuously differentiable over $\mathbb{R}^n$

Let each $g_j : \mathbb{R}^n \to \mathbb{R}$ and each h_i:$\mathbb{R}^n \to \mathbb{R}$

Unconstrained problem	constrained problem
Minimum f(x)	minimum f(x)
No restriction x$\in \mathbb{R}^n$	subject to $g_j \leq 0$, j=1,…,m
	h_i (x)=0, i=1,..,r

Necessary optimality conditions:

If x* is an optimal solution for the problem, then x* satisfies $\nabla f(x^*)$=0

Necessary optimality conditions: Equality constrained problems

Consider the equality constrained problem,

$$\text{Min } f(x) \tag{3}$$

$$\text{s.t } h_i(x)=0,\ i=1,..,r$$

The function f and all h_i are continuously differentiable over $\mathbb{R}^n$.

The optimality conditions are specified through the use of Lagrangian functions:

Lagrangian function is given by:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{r} \lambda_i h_i(x)$$

Where $\lambda = \lambda_i, \ldots, \lambda_r$

Necessary optimality conditions:

Assuming some regularity conditions for problem (3), if x* is an optimal solution of the problem, then there exists a Lagrange multiplier $\lambda^* = \lambda_1^*, \ldots, \lambda_r^*$ such that

$\nabla_x L(x^*, \lambda^*) = 0$ $\qquad\qquad$ $\nabla f(x^*) + \sum_{i=1}^{r} \lambda_i^* \nabla h_i(x^*) = 0$

$\lambda_1 L(x^*, \lambda^*) = 0$ $\qquad\qquad$ $h_i(x^*) = 0$ for i=1,…,r

λ^* is the optimal associated with the solution x*

This condition is known as KKT conditions

Important: The KKT condition can be satisfied at a local minimum, a global minimum (solution of the problem) as well as at a saddle point.

Question: We want to determine the optimal solutions of the problem (global minima of the constrained problem) ? How can we use the KKT conditions?

Solution:

We can set up a system of linear equations using the KKT conditions:

$\nabla f(x) + \sum_{i=1}^{r} \lambda_i \nabla h_i(x) = 0$

$h_i(x) = 0$, for i=1,…,r

Here, we have n+r unknown variables (x of size n and λ of size r) and n+r equations.

We can solve the system and find the points that satisfy the equation (KKT conditions).

These points are known as stationary points (or KKT points)

The optimal solutions if any are among these points

Note: We need additional information to characterize the stationary points as global or local minim or other by using the second order information for the objective f(x) at the KKT points.

Example 2.5

Determine all the stationary points of the following constrained problem:

Min f(x)=$x_1^2+x_2^2+x_3^2$

s.t. $x_1+x_2+3x_3-2=0$

$5x_1+2x_2+x_3-5=0$,

We construct/transform the Lagrangian function for the problem

$$L(x,\lambda)=x_1^2+x_2^2+x_3^2+\lambda_l(x_1+x_2+3x_3-2)+\lambda_2(5x_1+2x_2+x_3-5)$$

We set up the equations:

1, $\frac{\partial L}{\partial x_1}=2x_1+\lambda_1+5\lambda_3=0$

2, $\frac{\partial L}{\partial x_2}=2x_2+\lambda_1+2\lambda_2=0$

3, $\frac{\partial L}{\partial x_3}=2x_3+3\lambda_1+\lambda_2=0$

4, $\frac{\partial L}{\partial \lambda_1}=x_1+x_2+3x_3-2=0$

5, $\frac{\partial L}{\partial \lambda_2}=5x_1+2x_2+x_3-5=0$

Solving this system of equations and inequalities by distinguishing different cases, we have the only solution, x=$[.8043, 0.3478, 0.2826]^T$ and $\lambda=[.0870, 0.3044]^T$, so we

have only one KKT point, namely x=$[.8043, 0.3478, 0.2826]^T$,

But, we still don't know if this is optimal or not?

We can check the Hessian of f, we will find that the Hessian is positive definite.

The Hessian is positive definite, therefore the point x* is a global minimum.

Necessary optimality conditions: inequality and equality constrained problems

Consider the following constrained problems:

Min f(x)

S.t. g_i ≤0,j=1,…,m (4)

h_i(x)=0, i=1,…,r

The functions f and all g_i are continuously differentiable over $\mathbb{R}^n$.

The optimality conditions are specified through the use of Lagrangian function

Introduce a multiplier per constraint: $\mu_j \geq 0$ for each constrained g_j ≥0 and h_i for each constraint h_i(x)=0

L(x, μ,λ)=f(x)+$\sum_{j=1}^{m} \mu_j g_j$(x)+$\sum_{i=1}^{r} \lambda_i h_i$,Where $\mu=(\mu_{1,},\ldots,\mu_m)$ and $\lambda=(\lambda_1,\ldots,\lambda_r)$

A necessary optimality conditions: Assuming some regularity conditions for problem (4)

if x* is an optimal solution of the problem, then there exists Lagrange multipliers

$\lambda^* = (\lambda_1^*,\ldots,\lambda_r^*)$ and $\mu^* = (\mu_1^*,\ldots,\mu_m^*)$ such that

$\nabla f(x^*)$+$\sum_{j=1}^{m} \mu_j^* \nabla g_j$(x*)+$\sum_{i=1}^{r} \lambda_i^* \nabla h_i$(x*)=0

g_j(x*)≤ 0, for all j=1,…,m

h_i(x)=0, for all i=1,…,r

$$\mu_j^* \geq 0, \text{ for all } j=1,\ldots,m$$

$$\mu_j^* g_j(x^*)=0 \text{ for all } j=1,\ldots,m$$

μ_j^* And λ^* are the optimal associated with the solution

This is KKT conditions.

The KKT conditions to characterize all the stationary points of the problem, and then perform some additional testing to determine the optimal solutions of the problem (global minima of the constrained problem).

Determine KKT point: we set up a system for problem (4)

$\nabla f(x)+\sum_{j=1}^{m} \mu_j \nabla g_j(x) + \sum_{i=1}^{r} \lambda_i \nabla h_i(x)=0$

$g_j(x) \leq 0$ for all j=1,…,m

$h_i(x)=0$ for all i=1,…,r

$\mu_j \geq 0$,for all j=1,…,m complementary slackness

We may solve this system n unknown variables x, μ and λ find all the points satisfying the KKT condition. These points are stationary points (KKT points) of the problem

The optimal points are among the KKT points; we need additional information to characteristic the KKT point as global or local minimum or other .we use the second order information.

Sufficient optimality condition: convex inequality constrained problem

Consider the following constrained problem

$$\text{Min } f(x) \tag{5}$$

s.t. $g_j(x) \leq 0$, j=1,…,m

Functions f and all g_j(x) are convex and continuously differentiable over R^n.

We say that the problem is convex for this convex problem, the KKT conditions are also sufficient for optimality

Assuming some additional regularity conditions for convex problem (5), x* is an optimal solution of the problem, if and only if, there exists a Lagrange multiplier. $\mu^* = \left(\mu_1^*, \ldots, \mu_m^*\right)$ such that $\nabla f(x^*) + \sum_{j=1}^{m} \mu_j^* \nabla g_j(x^*)$, $g_j(x^*) \leq 0$, for all j=1,…,m

$\mu_j^* \geq 0$,for all j=1,…,m , $\mu_j^* g_j(x^*) = 0$ for all j=1,…,m

μ^* Looking for the solutions: is the optimal associated with the solution x*

We can use the KKT condition fully recover optimal solutions of the convex problem (global minima of the constrained problem).

Looking for the solutions: we set up a KKT system for the problem (5)

$\nabla f(x) + \sum_{j}^{m} \mu_j \nabla g_j(x) + \sum_{i=1}^{r} \lambda_i \nabla h_i(x^*) = 0$

$g_j(x) \leq 0$,for all j=1,…,m

$h_i(x) = 0$ for all i=1,…,r

$\mu_j \geq 0$, for all j=1,…,m

$\mu_j g_j(x) = 0$ for all j=1,…,m complementary slackness

We solve this system in unknown variable x, μ and find all the KKT points

Important: By the assumed convexity of the problem (5) all the KKT points are global minima of the problem (optimal solutions).

CHAPTER THREE

3. Non smooth optimization problems

Definition-3.0: A non-smooth optimization (NSO) refers to the general problem of minimizing (maximizing) functions that are typically not differentiable at their minimizers (or maximizers).

Definition-3.1 (structural convex optimization)

Consider the following convex optimization problem:

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

$$\text{s.t. } x \in C$$

f(x) is a convex function. C is a closed convex subset of vector space V

Properties:

f(x) can be smooth or nonsmooth; Solving nonsmooth convex optimization problems is much harder than solving differentiable ones; For some non-smooth non convex, cases even finding a descent direction is not possible; The problem is involving linear operators.

3.1. Nonsmooth convex optimization.

In this sub-section, we consider a nonsmooth optimization problem of the form:

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

$$\text{s.t } x \in X$$

Where, Set $X \subseteq \mathbb{R}^n$ is a set of feasible solutions; Objective function $f:\mathbb{R}^n \to \mathbb{R}$ is not required to have continuous derivatives.

Indicator functions of convex sets are particularly interesting for optimization. They are used to treat unconstrained and constrained convex optimization problems in unified framework. The convex constraint set can be associated with an indicator function. In order to avoid solution outside the constraint set, indicator functions are defined different to other branches of analysis by assigning ∞ to unfeasible solutions.

Definition-3.2. **(indicator functions).** The indicator function associated with a convex set C$\in \mathbb{R}^n$ is defined by: $\delta_c(x) = \begin{cases} 0, & if\ x\epsilon C, \\ +\infty, & if\ x \notin C \end{cases}$

Then,

1. δ_c is a proper and convex function iff C is nonempty and convex.

2. δ_c can be used as proper penalty function for most optimization problems.

3.2. Sub gradient.

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

$$\text{s.t. } x \in C$$

We have the following characteristics: The function f:$\mathbb{R}^n \rightarrow \mathbb{R}$ is convex and (possibly) non-differentiable. The set C$\subseteq \mathbb{R}^n$ is nonempty, closed and convex. The optimal value f^* is finite.

Definition3.3 A vector g $\in \mathbb{R}^n$ is a sub gradient of f: $\mathbb{R}^n \rightarrow \mathbb{R}$ at $x \in \mathbb{R}^n$ when f(z)$\geq$f(x) +g^T(z-x), for all z$\in$ $dom\ f$. The set of all sub-gradients of f at x is called the sub differential of f at x and denoted by ∂f(x).

3.3. Directional Derivatives and sub differentials for convex functions.

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

Definition 3.4. The directional derivative of f at x^o in the direction of y is :

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

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

Theorem.3.1

Let x∈int(dom f) ,then the directional derivative f'(x; d) is finite for all d ∈ R^n.

In particular, we have

$$f'(x;d) = \max_{s \in \partial f(x)} s^T d$$

Proof:

When x∈int(dom f) the sub differential ∂f(x) is non-empty and compact

Using the sub differential define relation, we can see that f'(x;d)≥ s^Td, for all s∈ ∂f(x) .

Therefore, $f'(x;d) = \max_{s \in \partial f(x)} s^T d$

To show the actually equality holds, we rely on separating Hyper plane Theorem. Define

c_1={(z,w)| z∈int(dom f) ,f(z)<w},

c_2={(y,v)| y=x+αd, v=f(x)+αf'(x;d)≥ 0}, The sets are non-empty, convex, and disjoint, By the separating Hyper plane Theorem, there exists a non-zero vector (α,β) ∈ R^{n+1} such that, a^T(x+αd)+β f(x)+αf'(x;d))≤ a^Tz+βw …….(1)

For all α≥ 0, z∈dom f, and f(z)< w,we must have β≥0

We cannot have β=0, Thus, β>0 and we can divide by the relation in (1), and obtain with $\tilde{a}$=a/β, $\tilde{a}^T$(x+αd) +f(x)+ αf'(x;d) ≤ $\tilde{a}^T$z+w ……..(2)

For all α≥0, z∈dom f, and f(z)< w. Choosing α= 0 and Letting w↓ f(z), we see

$\tilde{a}^T$x+f(x)≤ $\tilde{a}^T$z+f(z),implies that f(x) -$\tilde{a}^T$ (z-x)≤f(z), for all z∈dom f

Therefore, -$\tilde{a}$ ∈ $\partial f(x)$

Letting z=x, w↓ f(z) and α=1 in (2), we obtain

$\tilde{a}^T$(x+d) +f(x)+ f'(x;d) ≤ $\tilde{a}^T$x+f(x) ,

Implies f'(x;d)≤ -$\tilde{a}^T$d, In view of f'(x;d) ≥ max s^Td

s ∈ ∂f(x)

It follows that, f'(x;d) =max s^Td

s ∈ ∂f(x)

Where the maximum is attained at the "constructed" sub gradient - $\tilde{a}$.

Definition.3.5

i, A function f is called sub differentiable at x if there exists at least one sub gradient of f at x.

ii, A function f is called sub differentiable if it is sub differentiable at all x∈dom f.

iii, A function f:$\mathbb{R}^n \rightarrow \mathbb{R}$ be convex .The sub differential of f at x$\in \mathbb{R}^n$ is a set

$\partial f(x)$ ={u$\in \mathbb{R}^n$|f(y)$\geq$ f(x) + $u^T(y-x)$ for all y $\in \mathbb{R}^n$ }

Each vector u$\in \partial f(x)$ is called a sub gradient of f at point X. The sub differential $\partial f(x)$ is a non-empty, convex and compact set., we have $\partial f(x)$= $\{\nabla f(x)\}$

3.4. Optimality conditions

3.4.1. Optimality conditions: unconstrained case

In this subsection, unconstrained optimization:

Min f(x)

The function f is convex (non-differentiable) and proper f means f(x)>-∞ for all x and domf $\neq \emptyset$

Theorem.3.2

Under this assumption, a vector x* minimizes f over $\mathbb{R}^n$ if and only if 0$\in \partial$f(x*)

Proof:

x* is optimal if and only if

f(x)≥f(x*) for all x, or equivalently f(x)≥f(x*)+0^T(x-x*) for all x $\in \mathbb{R}^n$

Thus, x* is optimal if and only if 0$\in \partial$f(x*)

Examples-3.1

1. The function f(x)=|x|

$$\text{Since, } f(x) = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$$

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

$$\partial f(0)=\begin{cases} sgn(x)\ , \ for\ x \neq 0 \\ [-1,1]\ \ , for\ x = 0 \end{cases}$$

The minimum is at x*=0, and evidently $0 \in \partial f(0)$.

2. The function f(x)=||x||

$$\partial f(x)= \begin{cases} \frac{x}{||x||}\ , \quad for\ x \neq 0 \\ \{s: \left||s|\right| \leq 1\},\ \ for\ x = 0 \end{cases}$$

Again, the minimum is at x*=0, and $0 \in \partial f(0)$.

Definition.3.6

The sub differential of f at the point $x^o \in \mathbb{R}^n$ is a set of vectors given by

$$\partial f(x^o)=\{u\in \mathbb{R}^n | u^T(x-x^o) \leq f(x) - f(x^o), \forall_x \in \mathbb{R}^n \}$$

A vector $u \in \partial f(x^o)$ is called the sub gradient of f at x^o

For $f(x)=\max\{f_1(x), \dots \dots, f_m(x)\}$ and $I(x)=\{i | f_i(x) = 0$,the index of active functions at x,

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

That is, convex hull of the union sub gradient of active functions at x.

Moreover, if f_i's are differentiable,

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

Examples.3.2

The function f(x)=max$\{x^2+2x-3, x^2-2x-3, 4\}$

Since, the function

$$f(x) = \begin{cases} x^2-2x-3, for\ x < -1 \\ 4\ , for\ x \in [-1,1] \\ x^2+2x-3\ , for\ x > 1 \end{cases}$$

$$\partial f(x) = \begin{cases} 2x-2\ , \text{for } x < -1 \\ 0, \text{for } x \in (-1,1) \\ 2x+2, \text{for } x > 1 \end{cases}$$

Then, f is not differentiable at a stationery point {-1,1}.

But, we consider these points as stationary points/critical points, and

$$\partial f(x) = \begin{cases} 2x-2, \quad for\ x < -1 \\ [-4,0]\ , for\ x = -1 \\ 0\ , for\ x \epsilon (-1,1) \\ [0,4]\ , for\ x = 1 \\ 2x+2\ , otherwise \end{cases}$$

The optimal solution set is x*=[-1,1], For every x*∈X, we have 0 ∈ ∂f(x^*)

Example.3.3

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

Since, the function is:

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

We consider the following cases to find the intervals of the indicator functions:

Case 1: -3x=2x-1

$\Rightarrow$-3x-2x=-1

$\Rightarrow$-5x=-1

$\therefore$ x=$\frac{1}{5}$

Case 2: -3x=$x^2 - 4$

$\Rightarrow x^2 + 3x - 4$=0

$\Rightarrow$(x-1)(x+4)=0

$\therefore$x=1 or x=-4

Case 3: 2x-1=$x^2 - 4$

$\Rightarrow x^2 -$ 4 -2x+1= 0

$\Rightarrow x^2$-3x+x-3=0

$\Rightarrow$(x+1)(x-3)=0

$\therefore$x=-1 or x=3

Then, f is not differentiable at a stationery point {$\frac{1}{5}$,-4,3}.

But, we consider these points as stationary points/critical points, and

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

3.4.2. Optimality conditions: constrained case

In this subsection, consider constrained optimization problem:

Min f(x)

s.t. $x \in C$

Assumption: The function f is convex (non-differentiable) and proper, the set C is non-empty, closed and convex

Theorem.3.3

If $f:\mathbb{R}^n \to \mathbb{R}$ is convex, then for all $y \in \mathbb{R}^n$, $f(y)=\max\{f(x)+u^T(y-x) \mid x \in \mathbb{R}^n, u \in \partial f(x)\}$

Theorem.3.5

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

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

Proof:

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

$$\leq \lambda f(x) + (1-\lambda) f(x^o)$$

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

$$\Leftrightarrow f(x^o + \lambda(x-x^o)) \leq \lambda f(x) + f(x^o) - \lambda f(x^o)$$

$$\Leftrightarrow f(x^o + \lambda(x-x^o)) + \lambda f(x^o) - f(x^o) \leq \lambda f(x)$$

$$\Leftrightarrow \left[\frac{f(x^0 + \lambda(x-x^0)) - f(x^0)}{\lambda}\right] + f(x^o) \leq f(x)$$

Applying the limit as λ goes to zero on both sides we get

$$f(x) \geq f(x^o) + f'(x^o, x - x^o).$$

Conversely, assume $f(x^o)+f'(x^o, x - x^o)$ for all x^o,$x \in R^n$ and for all $\lambda \in [0,1]$

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

Since, directional derivatives are linear mappings, we conclude further

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

$$= f(\lambda x^o + (1-\lambda)x)$$

Consequently, the function f is convex.

From theorem 3.5, we have seen that for every convex function

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

This means that the sub gradient at an arbitrary point $x^o \epsilon R^n$ is nonempty.

3.4.3. Unconstrained convex smooth optimization

First-order condition: A point x* is a minimizer of a convex function f if and only if f is sub differentiable at x* and, $0 \in \partial f(x^*)$

i. e ,g =0 is a sub gradient of f at x* ……………………….(*)

The condition (*) reduces to $\nabla f(x^*) = 0$ if f is differentiable at x^*

Theorem.3.6

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

Proof:

By definition of sub gradient, we have

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

$f(y) \geq f(x^o), \forall_y \in \mathbb{R}^n$

That is, x^o is a minimum point for f.

Example.3.4

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

Find min f(x)

x∈ R^n

Solution:

To determine the minimum points of f, we use the following figure

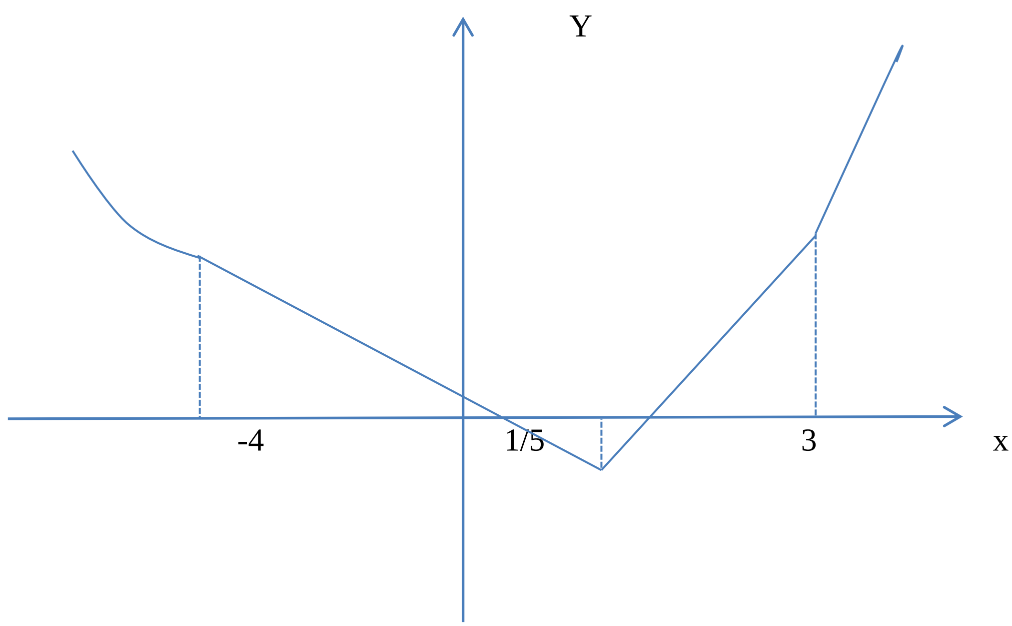

Fig 3.1 graph of max{-3x,{$x^2 - 4, 2x - 1$}}

Here, we see from example 3.3 that

The sub differential of f containing zero occurs only at x*=$\frac{1}{5}$

Hence, x*=$\frac{1}{5}$ is minimizes f.

Theorem.3.7

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

¡, $f'(x^o$,y),$\forall_y \in \mathbb{R}^n$

¡¡, 0 ∈ ∂f(x^o)

Proof:

If $x^o$is a minimum point of f, then

$$f'(x^o, y) = \lim_{t \downarrow 0} \inf \frac{f(x^0 + ty) - f(x^0)}{t} \geq 0\,, \forall_t \in \mathbb{R}^n$$

This implies (i) holds, Then(i) equivalent to (ii) as it follows from theorem 3.5 and (ii) holds by theorem 3.6.

3.4.4. constrained convex and nonsmooth optimization.

In this subsection, consider the convex constrained optimization problems

Min f(x) (p_c)

s.t: x∈ S

Where, the feasible set is given by: s={x∈ $\mathbb{R}^n$| $g_i(x) \leq 0, i = 1, \ldots, m$}, all functions are assumed to be convex.

3.4. 5. Difficulties caused by non-smoothness.

In this section, causes of non-smoothness are:

a. **Inherent** :original phenomenon contains various discontinuities and irregularities.

b. **Technological**: caused by some extra technological constraints which may cause a nonsmooth dependence between variables and functions.

c. **Methodological**: some algorithms for constrained optimization may lead to nonsmooth problems.

For examples, the exact penalty function method

d. **Numerical**: is so called “**stiff problem**” which are analytically smooth but numerically unstable and behave like nonsmooth problems.

3.4.6. Smooth problem.

Descent direction is obtained at the opposite direction of the gradient $\nabla f(x)$.

The necessary optimality conditions $\nabla f(x) = 0$.

Difference approximation can be used to approximate the gradient.

3.4.7. Non-smooth problem.

The gradient doesn’t exist every point leading to difficulties in defining the descent direction.

Gradient usually doesn’t exist at useful and may lead to serious failures.

The smooth algorithm doesn’t converge/diverges or it converges to a non-optimal point.

3.5. Optimality conditions with constraint Qualification for nonlinear programming problems solve by a cone Approach on The Karush-Kuhn-Tucker.

Under differentiability and constraint qualifications, the KKT conditions provide necessary conditions for a solution to be optimal. Under convexity, these conditions are also sufficient. If some of the functions are non-differentiable, sub differentiable versions of KKT conditions are available.

3.5.0. Nonlinear programming.

In mathematics, Non-linear programming (NLP) is the process of solving an optimization problem defined by a system of equalities and inequalities ,collectively termed constraints ,over a set of unknown real variables, along with an objective function to be maximized or minimized ,where some of the constraints or the objective functions are non-linear (not linear).

We shall study the standard NLP

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

$$\text{S.t. } g_j(x) \leq 0, \forall_j = 1, \ldots, p \qquad \text{(p)}$$

$$h_i(x) = 0, \ \forall_i = 1, \ldots, m$$

Where, the functions $f(x): \mathbb{R}^n \to \mathbb{R}$; $g_j(x): \mathbb{R}^n \to \mathbb{R}^p$ and $h_i(x): \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable.

The feasible set of the problem (p) will be denoted by Ω, i.e. $\Omega = \{x \in \mathbb{R}^n : g_j(x) \leq 0$ and $h_i(x) = 0\}$. The classical KKT condition at given $\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) + \sum_{j=1}^{p} \mu_j \nabla g_j(\bar{x}) = 0$,where $\mu_j \geq 0$, In order to have an optimal solution to the given NLP problem, the KKT necessary condition has to be satisfied. The above optimality criteria have been used to formulate algorithms that solve in(p) the presence of any constraint Qualification. These algorithms used the cones of directions of constancy. However if x solves (p), but x isn't a Kuhn-Tucker point (KKP), i.e. the KKT conditions don't hold at x. If the problem is unconstraint, then the KKT conditions reduced to $\nabla f(\bar{x}) = 0$,which is a necessary optimality conditions ;however ,this will not always be true.

Counter example 3.5.

Consider the problem (p) with $f(x): \mathbb{R}^2 \to \mathbb{R}$;$g(x): \mathbb{R}^2 \to \mathbb{R}^2$ for j=1,2 Defined by: $f(x) = x_1$ and $g(x) = (x_2 - (1 - x_1)^3, -x_2)^T$.

Solution:

Note that $X^* = (1,0)^T$, is minimizer of the problem but, the KKT conditions do not hold. In this paper we shall discuss assumptions on the constraints in order to ensure that the KKT conditions hold at a minimizer. Such an assumption is called constraint qualification (CQ). Formally, we say that the constraints h(x)=0 and g(x)$\leq$ 0 satisfy a constraint Qualification at $x^* \in \Omega$ when, given any differentiable function f minimized at x^* with respect to Ω, the KKT conditions are valid.

In general, several mathematicians obtained different constraint Qualifications. In this work, we will discuss many of them as well as some relations between them. A special interest is devoted to show the weakest such Qualification. In this context, the concept of cones and their polar will be useful.

Notation: Given $\bar{x} \in \Omega$, consider the set $A(\bar{x})$ of the inequality active constraint indices, that is,

$$A(\bar{x})=\{j:g_j(\bar{x})=0 ,1\leq j \leq p\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1^*)$$

Some important cones. In this section we present some useful cones based on the structure of the feasible set which play an important rule on the proof of the KKT theorem.

Definition -3.10 A subset C of $\mathbb{R}^n$ *is a* cone *when* td$\in$C, for all t$\geq$0 and d$\in$C

Definition-3.11 Given a set s$\in \mathbb{R}^n$, the **polar** of s, is given by: p(s)=$\{p\in \mathbb{R}^n, p^T x \leq 0$,for all x$\in$S$\}$.

Note that: For any $S\subset\mathbb{R}^n$, then p(s) is a cone and $s\subset p(p(s))$. This holds with equality if s is a closed convex cone, as established by Farka's Lemma as shown below.

Lemma-3.3 Let $C\subset\mathbb{R}^n$ be a closed convex cone, then p(p(c)) =c.

Proof.

For any x$\in$C, we have $x^T y\leq 0$, for all y$\in$p(c).

$\Rightarrow$x$\in$ p (p(c))

$\Rightarrow C\subset$ p (p(c)) ………………………..(i)

Conversely, take z∈p (p(c)) and

Let $\hat{z}=proj_c^z$ ∈C

$\Rightarrow (z-\bar{z})^T$(x-$\hat{z}$)$\le$ 0, for all x∈C.

Taking x=0 and x=2$\hat{z}$, we obtain $(z-\bar{z})^T(\hat{z})=0$

$\Rightarrow$ $(z-\bar{z})^T$(x-$\hat{z}$)$\le$ 0,for all x∈C.

$\Rightarrow$(z-$\hat{z}$)∈p(c) and since z∈p (p(c)) ,we have

$\Rightarrow (z-\bar{z})^T\hat{z}=0$

$\Rightarrow ||z-\hat{z}||^2$ =0

$\Rightarrow$||z-$\hat{z}$||=0

$\Rightarrow$z - $\hat{z}$=0

$\Rightarrow$z=$\hat{z}$ and z∈C

P (P(C))$\subset$C ……..(ii)

From (i) and (ii), we have

∴C=P (P(C))

Definition-3.12. Let s is a non-empty set in $\mathbb{R}^n$, and Let $\bar{x}\in\Omega$,The cone of feasible direction of s at $\bar{x}$ is denoted by V, and given by, V(x)={d∈ $\mathbb{R}^n$|$\bar{x}$+λ d∈S, for all λ ∈(0,δ], for some $\delta>0$},Each non-zero vector d∈V is called a feasible direction of f at $\bar{x}$.

Definition-3.13 Given a function f:$\mathbb{R}^n\to\mathbb{R}$, the cone of descent directions, at $\bar{x}$ is denoted by F, and given by F($\bar{x}$)={d∈ $\mathbb{R}^n$:f($\bar{x}$+λ d)<f($\bar{x}$), for all λ∈(0,δ], for some δ >0}**,** Each direction d∈F, is called a descent direction of f at $\bar{x}$.

The next result characterizes the descent direction and its proof follows from the derivative definition.

Lemma-3.4 Let f: $\mathbb{R}^n \rightarrow \mathbb{R}$, is differentiable function at a point $\bar{x} \in \mathbb{R}^n$, then

a. $\nabla f(\bar{x})^T d \leq 0$, for all d∈F($\bar{x}$) ……………………………(1)

b. If d∈ $\mathbb{R}^n$ satisfies $\nabla f(\bar{x})^T d<0$, and then d∈F($\bar{x}$) ………(2)

We get the set and denoted by $F(\bar{x})=\{d\in \mathbb{R}^n: \nabla f(\bar{x})^T d<0\}$

Proof of (a).

Let d∈F($\bar{x}$) and for some $\delta>0$ and for all $\lambda \in (0,\delta]$, we have

$f(\bar{x}+\lambda d)<f(\bar{x})$ i.e $f(\bar{x}+\lambda d)-f(\bar{x})<0$

$$\nabla f(\bar{x})=\lim_{\lambda \rightarrow 0} \frac{f(\bar{x}+\lambda d)-f(\bar{x})}{\lambda} \leq 0$$

$\nabla f(\bar{x})^T d\leq 0$

Proof of (b) obviously true by the definition of cone of descent directions.

Definition -3.14.

The set of first order feasible variations at a point $\bar{x} \in \Omega$, is the set $D(\bar{x})=\{d \in \mathbb{R}^n: \nabla h_i(\bar{x})^T d=0, \forall_i=1,\ldots,m$ and $\nabla g_j(\bar{x})^T d\leq 0, \forall_j \in A(\bar{x})\}$……………………….(3)

Where $A(\bar{x})$ is the active set defined by (1*).

Note that :D ($\bar{x}$) is a non-empty ,closed ,convex cone ,it is often said that this cone is a linear approximation of the feasible set .Again ,given $\bar{x} \in \Omega$, define the cone $G(\bar{x})==\{\sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x})+\sum_{j\in A(x)} \mu_j \nabla g_j(\bar{x}): \mu_j \geq 0, \forall_j \in A(\bar{x})\}$ ..…………………….(4)

Let us see some properties of this cone. For this, we shall need the classical result,named Caratheodory's Lemma.

Lemma-3.5. Let $u_1,u_2,\ldots,u_r$ are non-zero vectors in $\mathbb{R}^n$ for some $m<r$ and $x\in \mathbb{R}^n$ such that

$$x=\sum_{i=1}^{r} r_i$$

with $r_i \geq 0$, for all i$> m$. Then, there exist indices I$\subset${1,…,m} and J$\subset$ {m+1,….,r} and scalars r_i', where i$\in$ (I $\cup$J) , with $r_i' \geq 0$, for i$\in J$ such that

$$x=\sum_{i\in(\mathrm{IJ})} r' u_i \,,$$

and the vectors u_i for i$\in$ (I $\cup$ J) are linearly independent.

Proof.

Nothing to prove ,if the vectors $u_1,u_2,\ldots,u_r$ are linearly independent.

Suppose then that they are linearly independent. so

Assume that $u_1,u_2,\ldots,u_r$ are linearly independent, so that there exist scalars $\alpha_1, \alpha_2, \ldots\ldots, \alpha_r$, not all $\alpha_i = 0$, such that $\sum_{i=1}^{r} \alpha_i u_i = 0\,, for\ all$ t$\in$R, then x=$\sum_{i=1}^{r}(r_i - t\alpha_i)u_i \geq 0$, for all i>m

Define $\bar{t}$, as t of minimum absolute value that vanish one of the coefficients $(r_i - t\alpha_i)$, then X=$\sum_{i=1}^{r}(r_i - \bar{t}\alpha_i)u_i$, with$(r_i - \bar{t}\alpha_i)u_i \geq 0$, for all i>m

Therefore, x is written as a linear combination using of no more than r-1 vectors. we can repeat this process until that all vectors of the linear combination are linearly independent

Lemma-3.6. For any $\bar{x} \in \Omega$, then G($\bar{x}$) is closed convex cone.

Proof.

To prove G($\bar{x}$) which is defined in (4) is convex.

Consider A($\bar{x}$)={1,…..,q} for $x_1, x_2 \in G(\bar{x})$ and t$\in$[0,1], and then there exists $\alpha, \lambda \in R^n$ and $\mu, \beta \in R_+^q$ such that

$x_1 = \sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j(\bar{x})$ and

$x_2 = \sum_{i=1}^{m} \alpha_i \nabla h_i(\bar{x}) + \sum_{j=1}^{m} \beta_j \nabla g_j(\bar{x})$

$\Rightarrow tx_1+(1\text{-}t)x_2 = \sum_{i=1}^{m}(t\lambda_i+(1\text{-}t)\alpha_i)\nabla h_i(\bar{x}) + \sum_{j=1}^{m}(t\mu_j+(1\text{-}t)\beta_j)\nabla g_j(\bar{x})$

since,$(t\mu_j+(1\text{-}t)\beta_j \geq 0$

$\Rightarrow tx_1+(1\text{-}t)x_2 \in G(\bar{x})$

$\Rightarrow x_1, x_2 \in G(\bar{x})$, and hence , G($\bar{x}$) is convex.

To prove that $G(\bar{x})$ is closed

Consider a sequence $(s^k) \subset G(\bar{x})$ satisfying s^k-$s^* \in R^n$.

It has to be proved that $s^* \in G(\bar{x})$. For suitable matrices B and C,

we have

$$G(\bar{x})=\{\beta\lambda+Cp : p \geq 0\}, \text{ by the caratheodery's Lemma 3.5}$$

We can assume that D=(BC) has linearly independent columns, so that $D^TD=(BC)^T(BC)$ $=C^TB^TBC=C^TC$ =I, it is a non-singular matrix i.e. $|D|=|D^TD| \neq 0$

Since, $(s)^k \subset G(\bar{x})$, and then there exist $r^k = \binom{\lambda^k}{p^k}$ with $p^k \geq 0$, such that

$$(s)^k = D\, r^k \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Since, D^TD is a non-singular matrix, i.e. $|D^TD| \neq 0$

$r^k= (D^TD)^{-1}D^Ts^K$, taking the limit $k \to \infty$

We get $\binom{\lambda^*}{p^*}=r^* = \lim_{k\to\infty} r^k=(D^TD)^{-1}D^Ts^K$ with $p^* \geq 0$

Again, taking the limit $k \to \infty$, in (5) we get $\lim_{k\to\infty} s^k=s^*=D\, r^* \in G(\bar{x})$.

Hence, $G(\bar{x})$ is closed.

Lemma-3.7. For any $\bar{x} \in \Omega$, then $D(\bar{x})=P(G(\bar{x}))$.

Proof. By the lemma 3.3 and 3.6, we need to prove that $D(\bar{x})=P(G(\bar{x}))$, where $D(\bar{x})$ and $G(\bar{x})$ are defined in (3) and (4), respectively.

Consider $d \in D(\bar{x})$, and given $s \in G(\bar{x})$, and then we have

$$d^Ts=\sum_{i=1}^{m} \lambda_i d^T\nabla h_i(\bar{x})+\sum_{j\in A(\bar{x})} \mu_j d^T\nabla g_j(\bar{x}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

By definition of $D(\bar{x})$ and since $\mu_j \geq 0$

$\Rightarrow d^Ts \leq 0$, so $d \in P(G(\bar{x}))$.

Conversely,consider $d \in P(G(\bar{x}))$, i.e.$d^Ts \leq 0$, for all $S \in G(\bar{x})$.

In particular, since $\pm\nabla h_i(\bar{x})$ belongs to $G(\bar{x})$, for all i=1,…,m

We get, $d^T\nabla h_i(\bar{x}) = 0$. Further more, since $\nabla g_i(\bar{x}) \in G(\bar{x})$, for all $j \in A(\bar{x})$, we have $d^T\nabla g_j(\bar{x}) \leq 0$,

Hence,$d \in D(\bar{x})$, then $D(\bar{x})= P(G(\bar{x}))$.

3.5.1.The Tangent cone.

Definition-3.15. A vector d $\in \mathbb{R}^n$ is called a direction to $\Omega \subset \mathbb{R}^n$ from $\bar{x} \in \Omega$ when either d=0, or there exists a squence of feasible points $x^k \subset \Omega$ such that $x^k \to \bar{x}$, and also

$$\frac{x^k - \bar{x}}{||x^k - \bar{x}||} \to \frac{d}{||d||}$$

Clearly, the set T($\bar{x}$) of the tangent directions to Ω from $\bar{x}$ is a cone. This set is said to be a tangent cone.

Example 3.6

Let S={(x,y)$\in R^2$:x^2-y=0},find the tangent cone at (0,0)

Solution.

Let (x_k,y_k)→(0,0), i.e ,x_k →0 and y_k=x_k^2

$\Rightarrow \|(x_k, y_k)\| \to (0,0)$

$\Rightarrow \sqrt{x_k^2 + y_k^2} = \sqrt{x_k^2 + x_k^4} = |x_k|\sqrt{x_k^2 + 1}$

$\Rightarrow \lim_{x_{k\to 0^+}} \frac{x_k}{|x_k|\sqrt{x_k^2+1}} = \lim_{x_{k\to 0^+}} \frac{x_k}{x_k\sqrt{x_k^2+1}} = 1$, and

$\Rightarrow \lim_{x_{k\to 0^+}} \frac{y_k}{|x_k|\sqrt{x_k^2+1}} = \lim_{x_{k\to 0^+}} \frac{x_k^2}{x_k\sqrt{x_k^2+1}} = 0$, and also

$\Rightarrow \lim_{x_{k\to 0^-}} \frac{-x_k}{x_k\sqrt{x_k^2+1}} = -1$ and

$\Rightarrow \lim_{x_{k\to 0^-}} \frac{y_k}{|x_k|\sqrt{x_k^2+1}} = 0$

∴T(0,0)={(1,0),(-1,0)}.

The next lemma states that this cone is closed. However, T($\bar{x}$) is not necessarily convex.

Lemma-3.8

For any $\bar{x} \in \Omega$, then T($\bar{x}$) is closed, where T($\bar{x}$) is the tangent direction to Ω.

Proof.

Consider $d^k \subset$T($\bar{x}$) with $d^k \to$ d.To prove that d$\in$ T($\bar{x}$).

When d=0 , we get d$\in$ T($\bar{x}$)

Assume that d≠0, and suppose that without loss of generality that $d^k \neq 0$, for all k∈N. Fixed k∈N, since $d^k \in T(\bar{x})$,then there exists

$(x^{k,j})_{j\in N} \subset \Omega$, then $x^{k,j} \to \bar{x}$ and $q^{k,j} = \frac{x^{k,j}-\bar{x}}{\|x^{k,j}-\bar{x}\|} \to \frac{d}{\|d\|}$

So,There exists j_k ∈N such that $\|x^k - \bar{x}\| < \frac{1}{k}$ and $|q^k - \frac{d}{\|d^k\|}| < \frac{1}{k}$, where $x^k = (x^{k,j})_k$ and $q^k = (q^{k,j})_k$

Taking the limit k→ ∞, we get $x^k \to \bar{x}$ and also

$|q^k - \frac{d}{\|d\|}| \leq |q^k - \frac{d^k}{\|d^k\|}| + |\frac{d^k}{\|d^k\|} - \frac{d}{\|d\|}| \to 0$

Thus, $\frac{x^k-\bar{x}}{\|x^k-\bar{x}\|} = q^k \to \frac{d}{\|d\|}$, which implies d∈T($\bar{x}$)

∴, T($\bar{x}$) is closed where T($\bar{x}$) is the tangent direction to Ω.

Remark.

we have presented two different linear approximations of feasible set of at a point $\bar{x}$: the first order feasible variations cone D($\bar{x}$), and the tangent cone T($\bar{x}$).

The next result shows that T($\bar{x}$) is a subset of D($\bar{x}$).

Lemma 3.9 For any $\bar{x} \in \Omega$,and T($\bar{x}$)⊂D($\bar{x}$), where T($\bar{x}$) and D($\bar{x}$) are the tangent cone and the first order feasible variations cone at a point $\bar{x}$.

Proof.

Consider d∈ T($\bar{x}$) , d≠ 0. then there exists a squence $(x^k) \subset \Omega$ with , Such that $x^k \to \bar{x}$ And $\frac{x^k-\bar{x}}{\|x^k-\bar{x}\|} \to \frac{d}{\|d\|}$. From the smoothness of g and h. It follows that

h(x^k) $= h(\bar{x}) + \nabla h_i(\bar{x})^T(x^k - \bar{x}) + 0(\|x^k - \bar{x}\|)$ and

g(x^k) $= g(\bar{x}) + \nabla g_i(\bar{x})^T(x^k - \bar{x}) + 0(\|x^k - \bar{x}\|)$.

Since $x^k, \bar{x} \in \Omega$, we have for all i∈ $A(\bar{x})$,

$$\nabla h_i(\bar{x})^T \frac{(x^k-\bar{x})}{\|x^k-\bar{x}\|} + \frac{0(\|x^k-\bar{x}\|)}{\|x^k-\bar{x}\|} = 0 \text{ and } \nabla g_i(\bar{x})^T \frac{(x^k-\bar{x})}{\|x^k-\bar{x}\|} + \frac{0(\|x^k-\bar{x}\|)}{\|x^k-\bar{x}\|} \leq 0$$

Taking the limit k→ ∞,we get

$$\nabla h_i(\bar{x})^T \frac{d}{\|d\|} = 0 \quad \text{and} \quad \nabla g_i(\bar{x})^T \frac{d}{\|d\|} \leq 0, \text{for all i} \in A(\bar{x})$$

Thus, $d \in D(\bar{x})$,and hence we get T($\bar{x}$)⊂ D($\bar{x}$).

The converse of Lemma 3.9 is not true, as we can see in the folowing counter example.

Counter Example-3.7 Consider the functions h(x):$\mathbb{R}^2 \rightarrow \mathbb{R}$; g(x):$\mathbb{R}^2 \rightarrow \mathbb{R}$ for j=1,2 Defined by: h(x)=x_1x_2 and g(x)= -x_1-x_2 , and the point $\bar{x}$=$(0,0)^T$. Thus,

$$T(\bar{x})=\{(d_1, d_2)\in \mathbb{R}^2: d_1 \geq 0, d_2 \geq 0 \text{ and } d_1d_2 = 0\} \text{ and also}$$

$$D(\bar{x})=\{(d_1, d_2)\in \mathbb{R}^2: -d_1 - d_2 \leq 0 \text{ and } T(\bar{x})\neq D(\bar{x}).$$

Note. If T($\bar{x}$)= D($\bar{x}$) is a constraint qualification known as “Quasi-regularity”.

3.5. 2. Optimality conditions and constraint Qualification.

In this section, we prove the KKT theorem assuming the weakest qualification conditions and discuss other ones easier to be verified.

Next lemma roughly says that at a minimizer, the objective functions increases along tangent directions.

Lemma 3.10 If x*∈ Ω is a local minimizer of the problem (p), and then $\nabla f(\bar{x})^T$d≥0, for all d∈ T(x*).

Proof:

This follows directly from the relation

$0 \leq f(x^k)-f(x^*)=\nabla f(x^*)^T(x^k-x^*)+0(\|x^k-x^*\|)$,

Which is valid for $(x^k)\subset \Omega$.

Now we state the classical Karush-Kuhn-Tucker theorem.

Theorem.3.10 Let x* ∈ Ω be a local minimizer of the problem (p). and if P(T(x*))=P(D(x*)), and then there exists λ* $\in R^m$ and μ* $\in R^p$ such that $\nabla f(x^*)+\sum_{i=1}^{m} \lambda^*_{i=1} \nabla h_i(x^*) + \sum_{j=1}^{p} \mu^*_{j=1} \nabla g_j(x^*)=0$, with $\mu_j^* \geq 0$,for j=1,…,p and $\mu_j^* g_j(x^*)=0$, j=1,….,p.

Proof: Consider $x^*\in\Omega$ is a local minimizer of the problem (p) by lemma -3.10, we have

$-\nabla f(x^*)^T d\leq 0$, for all $d\in T(x^*)$. Thus, using the hypothesis and the lemma-3.7, we get

$$-\nabla f(x^*)\in p(T(x^*)) = p(D(x^*))= G(x^*).$$

This means that there exists $\lambda\in R^m$ and $\mu_j \geq 0$ for $j\in A(x^*)$,such that

$$\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla h_i(x^*) + \sum_{j\in A(x^*)}\mu_j \nabla g_j(x^*)=0.$$

Define $\lambda^*=\lambda$ and $\mu^*=\begin{cases}\mu_j, for j\in A(x^*)\\ 0, otherwise\end{cases}$, Hence, complete the proof.

Note: $P(T(x^*))$ and $P(D(x^*))$ are the polar of the tangent cone and the first order feasible variations cone respectively.

3.5.3. Constraint Qualifications.

The Kuhn-Tucker conditions are only if some regularity conditions are satisfied. These conditions are called the constraint qualification which imposes a certain restriction on constraint functions of a nonlinear programming problem. For the specific purpose of ruling out certain irregularities of the boundary of the feasible set that would available KKT conditions should be the optimal solution occurs there.

3.5.4. Quasi regular constraint qualification.

We say that the quasiregularity constraint qualification is satisfied at $\bar{x}$ when $T(\bar{x})=D(\bar{x})$.Where $T(\bar{x})$ and $D(\bar{x})$ are the tangent cone and the set of first order feasible variations cone at $\bar{x}$.

Next example shows that these conditions are not equivalent.

For example 3.8: Consider the functions $h(x):\mathbb{R}^2\to\mathbb{R}$, $g(x):\mathbb{R}^2\to\mathbb{R}$ for j=1,2 defined by

$h(x) = x_1x_2$, $g(x) = -x_1-x_2$ and the feasible point $\bar{x} = (0,0)^T$. It is easy to see that

$T(\bar{x}) = \{(d_1,d_2)\in R^2: d_1 \geq 0, d_2 \geq 0, d_1d_2=0\}$ and also

$D(\bar{x}) = \{(d_1,d_2)\in R^2: d_1 \geq 0, d_2 \geq 0\}$

and

$$p(D(\bar{x})) = P(T(\bar{x})) = \{(d_1,d_2)\in R^2: d_1 \leq 0, d_2 \leq 0\}.$$

3.5.5. Slater constraint qualification:

Regarding the problem (p), we say that the slater constraint qualification holds if h is linear and g is convex and then there exists $\bar{x} \in \Omega$, such that $h(\bar{x})=0$ and $g(\bar{x})<0$. The slater constraint is in fact, a constraint qualification.

Theorem 3.11: If the slater conditions hold and then $T(\bar{x}) = D(\bar{x})$, for all $\bar{x} \in \Omega$

Proof: Using lemma.3.9 is enough to prove that $D(\tilde{x}) \subset T(\bar{x})$

Consider an arbitrary direction $d \in D(\tilde{x})$ and $\tilde{x} \in \Omega$, it is given by slater condition.

Define; $\bar{d} = \tilde{x} - \bar{x}$ by the convexity of g_i, we have $0 > g_j(\tilde{x}) \geq g_j(\bar{x}) + \nabla g_j(\bar{x})^T \bar{d}$.

Thus, for $i \in A(x)$, $\nabla g_j(\bar{x})^T \bar{d} < 0$. Given $\lambda \in (0,1]$, define $\bar{d} = (1-\lambda)d + \lambda\bar{d}$.

3.5.6. Linear Independence constraint Qualification –LICQ.

This is the most known constraint qualification and states that the equality constraint gradients $\nabla h_i(\bar{x})$ for $i=1,\ldots\ldots,m$, and the active inequality constraint gradients $\nabla g_i(\bar{x})$, $i \in A(\bar{x})$ are linearly independent. Although easy to check, this condition is a very strong assumption.

For example 3.9: Consider the following functions:

$$\text{Min } f(x) = (x_1-3)^2+(x_2-2)^2$$

$$\text{S.t. } g_1(x) = 2x_1+x_2-6 \leq 0$$

$$g_2(x) = x_1+2x_2-6 \leq 0$$

In this problem, we have

$$\nabla f(x) = [2(x_1-3), 2(x_2-2)]$$

$$\nabla g_1(x) = [2,1]$$

$$\nabla g_2(x) = [1,2]$$

The gradients of g are linearly independent so all points are regular.

We consider the following cases.

Case-i:

For A($\bar{x}$)=∅

From KKT conditions, we get both λ_1=0 and λ_2=0, then ∇f(x)=[2(x_1-3),2(x_2-2)]=0

⇒x_1=3 and x_2=2

But, 2x_1+x_2-6=2(3)+2-6=6+2-6=2 ≠ 0

i.e. x is not feasible; it cannot be a local minimum

Case-ii:

For A($\bar{x}$)={1}

From KKT conditions, we get λ_2=0, then ∇f(x) +λ_1∇g_1(x)=[2(x_1-3),2(x_2-2)]+

λ_1[2,1]=0

x_1 =3-λ_1 and x_2=2- $\frac{\lambda_1}{2}$ - 6=0, λ_1=$\frac{4}{5}$

λ_1 and λ_2 satisfies KKT conditions and then we get x_1=3 - $\frac{4}{5}$ = $\frac{11}{5}$ and x_2 = $\frac{8}{5}$

Finally, g_2(x)=$\frac{11}{5}$ +$\frac{18}{5}$-6=$\frac{-1}{5}$ ≤ 0

Hence, KKT conditions are satisfied.

Case –iii:

For A(x)={2}

From KKT conditions, we get λ_1=0, then ∇f(x)+λ_2∇g_2(x)=[2(x_1-3),2(x_2-2)]+λ_1[1,2]

⇒x_1=3-$\frac{\lambda_2}{2}$ and x_2=2-λ_2, and the assumptions g_2(x)=0, it gives with these $x_1$and $x_2$as follows

$\Rightarrow$(3-$\frac{\lambda_2}{2}$)+2(2-λ_2)-6=0

λ_1 and λ_2 satisfies KKT conditions and then we get x_1=3- $\frac{1}{5}$ = $\frac{14}{5}$ and x_2=2- $\frac{2}{5}$ = $\frac{8}{5}$

But, g_1(x)=2($\frac{14}{5}$)+$\frac{8}{5}$-6=$\frac{1}{5} > 0$, this is violated the condition of g_1(x)≤0.

Case-iv:

For A($\bar{x}$)={1,2}

Finally, g_1(x)=2x_1 +x_2-6 and g_2(x)=x_1+2x_2-6, gives that x_1=2=x_2

But, the KKT conditions ∇f(x)+$\lambda_1\nabla g_1$(x)+$\lambda_2\nabla g_2$(x) =0

$\Rightarrow$[2(x_1-3), 2(x_2-2)]+λ_1[2,1]+λ_2[1,2]=0, gives with theses x_1 andx_2, as follows

-2+2λ_1+λ_2=0 and λ_1+2λ_2=0

$\Rightarrow\lambda_1$=$\frac{4}{3}$ and λ_2= - $\frac{2}{3}$ and since λ_1 >0 and λ_2 <0

It doesn't satisfy KKT condition.

Remark: Many problems satisfy KKT conditions without LICQ, as we can see in the following example with x*=0.

Consider the function:

Min f(x) = x_2

S.t. g_1(x) =x_1^2+$x_2 \leq 0$

g_2(x)= -x_2 ≤0

Clearly, it satisfies KKT conditions without Qualification-MFCQ.

3.5.7 Mangasarian-Fromovitz constraint qualification – MFCQ.

Another well-known condition which ensures KKT is due to Mangasarian and Fromovitz. We say that MFCQ holds at $\bar{x}$, when the equality constraint gradients are linearly independent and there exists a vector d $\in \mathbb{R}^n$ such that

$$\nabla h_i(\bar{x})^{T}d=0 \text{ for } i=1,\ldots,n \text{ and } \nabla g_j(\bar{x})^{T}d\leq 0 \text{ for all } j\in A(\bar{x}).$$

The best known necessary optimality criterion for a mathematical programming problem is the KKT optimality conditions; however, the above MFCQ condition is in a sense more general. In order for the KKT conditions to hold, one must impose a constraint qualification on the constraint of the problem. on the other hand, no such qualification need be impose on the constraints of the problem, on other hand, no such qualifications in order that the MFCQ to hold.

Moreover, the MFCQ itself can be used to drive a form of the constraint qualification for KKT conditions.

CONCLUSION

In this paper, they are appearing in applications much more than smooth optimization, solving nonsmooth optimization problems is much harder than common smooth optimization, the most efficient algorithms for solving them are first-order methods, we developed a version of the sub gradient for nonsmooth and in particular convex optimization problem, the proposed algorithm is easy to implement and also numerical results clearly demonstrate that the proposed algorithm is a significant improvement of the sub gradient method, we observe that, a KKT optimality condition for NLP has been proved the equality of the tangent cone and the polar of the first order feasible variations cone. Despite the difficulty of this property, it needs to have more readily verifiable conditions for the admittance of Lagrange multipliers. Such conditions called constraint qualifications have been investigated extensively. Some of them were discussed as: Slater, Linear Independence Gradients (LICQ), Mangasarian-Fromovitz's (MFCQ), and Quasi-Regularity conditions through which their relations are analyzed and can be summarized as LICQ implies MFCQ but the converse is not necessarily true. MFCQ implies Quasi-Regularity but the converse do not hold and that slater condition satisfies necessarily true. MFCQ implies Quasi-Regularity. Other constraint qualification is not discussed in this paper, such as Quasi –Normality condition which implies Quasi-Regularity, the constant positive linear dependence (CPLD) which is weaker than MFCQ and implies Quasi –Normality and constant rank constraint qualification (CRCQ) which shows the constraint positive linear dependence(CPLD) these are all aims to complete.

Bibliography

[1]. Abadie. J. 1967. On the Kuhn-Tucker Theorem ''Non Linear programming pp. 19 – 36" North Holland, Amsterdam

[2]. Belay Bekel, optimality conditions for nonsmooth optimization and mordukhovich subdifferenials, Addis Ababa university, 2012

[3]. Dertsekas, convex analysis and optimization, Athena Cientific, Belmot, Masschusett ,USA,2003.

[4]. F.H.Clarke,optimization and nonsmooth analysis, Society for Industrial and Applied mathematics,Philadelphia,1990.

[5].G. Giorgi ,A. Guerraggio and J.Theirfeleder,mathematics of optimization: smooth and nonsmooth case,Elsevier Scienc and Technology, 2004

[6]. Guignard.M.1967. "Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space" SIAM Journal optimal control,7: pp 232-24.

[7]. J. Jahn , introduction to the theory of nonlinear optimization ,2^{nd} revised edition ,springer-Verlag Heidelberg,New york,1996.Case,Elsevier Science and Technology,2004.

[8] .Kjeldsen.T.H.2000 "A contextualized Historical Analysis of the Kuhn-Tucker Theorem in Nonlinear programming" The impact of World war II Historria maths,4: pp 331-361.

[9]. Kuhn,H.W.andA.W.Tucker ,1951 . "Nonlinear programming" Proceedings of 2^{nd} Berkeley Symposium. Berkeley :University of California press.pp 481-492.

[10]. Qi,Land Z.,we,2000. "On the constant positive Linear Dependence Condition and its application to SQP methods SIAM Journal on Optimization ,10: pp 963-981

[11].Rodrigo,G.E.W.K.Elizabeth and A.R.,Admir.Karush- Kuhn-Tucker optimality conditions and constraint Qualifications. Department of Mathematics , Federal University of Parana.

[12]. R.T.Rockafellar and R.J-B Wets, Variational analysis,Springer-Verlag Berlin Heidelberg,New york,1998.

[13]. Tolle.J.W.and F.J.Gould,1971. “A necessary and sufficient qualification for constrained optimization. SIAM Journal On Applied Mathematics, 20: pp 164-172.

[14]. W.Schirotzek ,Nonsmooth Analysis ,Springer-Verlag Berlin Heidelberg,200