# Optimality Condition for Smooth Constrained Optimization Problems

**ADDIS ABABA UNIVERSITY**
**COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES**
**DEPARTMENT OF MATHEMATICS**

**A thesis submitted to the department of mathematics in partial fulfillment of the requirements for the degree of masters of science in mathematics**

By: **Meta Tafa Hunde**
**Stream: Optimization**
**Advisor: Berhanu Guta (PhD)**

Addis Ababa Ethiopia
August, 2020

**Acknowledgement**

First of all I would like to thank the Lord our God for all things were made by him; and without him was not any thing made that was made.Next, I would like to express my heart-felt gratitude to my respected advisor Dr. Berhanu Guta for his genuine advise and reviewing my work through giving constructive comment and I would like to thank my dear wife Hana Kelemu and all my family who facilitate suitable situations, to accomplished this project. Also I would like to thank Hiwot Tesfaye who was helped me by giving Laptop as a gift.
Finally, I would like to say thanks and respection all my spiritual brother's for sake of their praying and appreciation to strength my moral to perform this project.

**Abstract**

Mathematical optimization is the process of maximizing or minimizing an objective function by finding the best available values across a set of inputs under a restriction domain. This project focuses on finding the optimality condition for optimization problems of differentiable function. 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. For constrained Optimization problem, the objective function and the function in the constraint sets are differentiable and the well known optimality condition called Karush-Kuhn-Tucker (KKT) condition leads to find the optimum point(s) of the given optimization problem and the convetional Lagrangian approach to solving constrained optimization problems leads to optimality conditions which are either necessary or sufficient, but not both unless the underlying objective functions and functions in constraints set are also convex. The Tchebyshev norm leads to an optimality conditions which is both sufficient and necessarly without any convexity assumption.This optimality conditions can used to device a conceptually simple method for solving non-convex inequality constrained optimization problems.

**Keywords:** Convex set, convex function, constrained optimization, unconstrained Optimization, inequality constraints, equality constraints, smooth optimization, positive definiteness, Hessian matrix, non-convex optimization, equivalent optimality conditions

## Permission

This is to certify that this project entitled "**Optimality condition for smooth constrained optimization problem**" have read and recommeded for acceptance in the Departement of mathematics and also for the confirmation of the project submitted for evaluation by examiners and eventual defence in the partial fulfillment of the requirement for the Degree of Master of Science.

Advisor ———————————— Signature——————— Date———————-

Examiner 1 ———————————— Signature——————— Date———————-

Examiner 2 ———————————— Signature——————— Date———————-

## Notations

| | |
|---|---|
| $g \in c'$ | $g$ is once continuously |
| $g \in c''$ | $g$ is twice continuously differentiable function |
| NLP | Non Liniear Programming |
| $\mathbb{R}$ | The set of real numbers |
| $\mathbb{R}^n$ | $n$ dimensional eculidean space |
| $\frac{\partial g}{\partial y}$ | partial derivatives of $g$ with respect to $y$. |
| $\bigtriangledown g(y)$ | Gradient of real valued function |
| $\bigtriangledown^2 g(y)$ | Hessian of a function $g$ |
| $P_{un}$ | Unconstrained optimization problem |
| $P_c$ | Constrained optimization problem |
| $(P_<)$ | inequality constrained optimization |
| $(P_=)$ | Equality constrained optimization |
| FONC | First Order Necessary Conditions |
| SONC | Second Order Necessary Conditions |
| L | Lagrangian Function |
| $L(., \gamma, \delta)$ | Lagrangian Function with lagrange multiplier $\gamma$ and $\delta$ |
| $\langle \gamma, g \rangle$ | inner product of $\gamma$ and $g$ |

# Contents

**Chapter 1**

# 1 Preliminary

## 1.1 Convex Sets

**Definition1.1.1** (Line and Line segments in $\mathbb{R}^n$) Let $x, y \in \mathbb{R}^n$
$L = x + \theta(y - x) : \theta \in \mathbb{R}$ a line containing $x$ and $y$.
$[x, y]$=$x + \theta(y - x)$ :$0 \leq \theta \leq 1$ a closed line segment between $x$ and $y$.
=$y$+$(1 - \theta)x$ : $0 \leq \theta \leq 1$.
$(x, y)$=$x + \theta(y - x)$ :$0 < \theta < 1$ open line segment between $x$ and $y$.
=$y$+$(1 - \theta)x$ : $0 < \theta < 1$.
**Definition1.1.2** A set $S \subseteq \mathbb{R}^n$ is said to be convex if for any $x, y \in S$
$\theta x$+$(1 - \theta)y \in S$, $\forall \theta \in [0, 1]$

**Some Examples of Convex Set**

1. $\mathbb{R}^n$ is convex.

*Proof.* Let $t = (t_1, t_2, ..., t_n) \in \mathbb{R}^n$ and $s = (s_1, s_2, ..., s_n) \in \mathbb{R}^n$
we want to show that $\theta t$+$(1 - \theta)s \in \mathbb{R}^n$ $\forall \theta \in [0, 1]$
$\theta t$+$(1 - \theta)s$=$\theta(t_1, t_2, ..., t_n)$ +$(1 - \theta)$ ( $s_1, s_2, ..., s_n)$
=$(\theta t_1, \theta t_2, ..., \theta t_n)$+$((1 - \theta)s_1, (1 - \theta)s_2, ..., (1 - \theta)s_n))$
=$(\theta t_1$+$(1 - \theta)s_1$,$\theta t_2$+$(1 - \theta)s_2$, . . ., $\theta t_n$+$(1 - \theta)s_n) \in \mathbb{R}^n$
Therefore $\mathbb{R}^n$ is convex set. □

2. If A is $m \times n$ matrix and $b \in \mathbb{R}$, then the polyhedral set

$$P := \{x \in \mathbb{R}^n : Ax \leq b\}$$

is convex

*Proof.* Let$x \in \mathbb{R}^n$
Given $Ax_1 \leq b$ and $Ax_2 \leq b$.
We want to show that $(\lambda x_1 + (1 - \lambda)x_2) \in P$
$\Rightarrow A(\lambda x_1 + (1 - \lambda)x_2)$
=$\lambda Ax_1 + (1 - \lambda)Ax_2$
$\leq \lambda b + (1 - \lambda)b$
$\leq b$.
Therefore the polyhedral is convex set. □

3. Let $t_o \in \mathbb{R}^n$, and $b \geq 0$. $B(t_o, b)$=$\{t \in \mathbb{R}^n$:$\| t_1 - t_o \| \leq b\}$ is convex.

*Proof.* Let $t_1, t_2 \in \mathbb{R}^n$
Given $\|t_1 - t_o\| \leq b$ and $\|t_2 - t_o\| \leq b$

$$Let \quad t_0 = \theta t_o + (1-\theta)t_o$$

We want to show $\|\theta t_1 + (1-\theta)t_2\| \in B$
$\|\theta t_1 + (1-\theta)t_2 - (\theta t_o + (1-\theta)t_o)\|$
$=\|\theta t_1 + (1-\theta)t_2 - \theta t_o - (1-\theta)t_o\|$
$=\|\theta t_1 - \theta t_o + (1-\theta t_2) - (1-\theta t_o)\|$
$=\|\theta(t_1 - t_o) + (1-\theta)(t_2 - t_o)\|$
$\leq \|\theta(t_1 - t_o)\| + \|(1-\theta)(t_2 - t_o)\|$ by triangle inequality
$\leq \theta b + (1-\theta)b$
$\leq b$
Therefore $B(t_o, b)$ is convex. □

4. (The set of hyperplane). Let $A$ be an $m \times n$ matrix, $b \in \mathbb{R}^m$. Let $S := \{x \in \mathbb{R}^n : Ax = b\}$. (The set $S$ is just the set of solutions of the linear equation $Ax = b$) is a convex set.

*Proof.* We want to show that $((1-\lambda)x_1 + \lambda x_2) \in S, \ \forall \lambda \in [0,1]$
$A((1-\lambda)x_1 + \lambda x_2) = (1-\lambda)Ax_1 + \lambda A x_2$
$=(1-\lambda)b + \lambda b$
$=b.$ □

**Theorem 1.** *Let $K$ be a convex set and let $\lambda_1, \lambda_2, ..., \lambda_p \geq 0$ and $\sum_{i=1}^{p} \lambda_i = 1$. If $x_1, x_2, ..., x_i \in K$, then*

$$\sum_{i=1}^{p} \lambda_i x_i \in K$$

.

*Proof.* We prove the result by induction. Since $K$ is convex, the result is true, trivially, for $P = 1$ and by definition for $p = 2$. Suppose that the proposition is true for $P = r$ (induction hypothesis!) and consider the convex combination $\lambda_1 x_1, \lambda_2 x_2, ..., \lambda_{r+1} x_{r+1}$.
Define $A := \sum_{i=1}^{r} \lambda_i$. Then since $1 - A = \sum_{i=1}^{r+1} \lambda_i - \sum_{i=1}^{r} \lambda_i = \lambda_{r+1}$ we have

$$\sum_{i=1}^{p} \lambda_i x_i) + \lambda_{r+1} x_{r+1} = A(\sum_{i=1}^{r} \frac{\lambda_i}{A} x_i) + (1-A)x_{r+1}$$

Note that $\sum_{i=1}^{r} \frac{\lambda_i}{A} = 1$ and so, by induction hypothesis, $\sum_{i=1}^{r} \frac{\lambda_i}{A} x_i \in K$. Since $x_{r+1} \in K$, it follows that the right hand side is a convex combination of two points of $K$ and hence lies in $k$. □

**Theorem 2.** *Let $S_i$,$i \in I = \{1, 2, ..., n\}$ be each convex. Then their intersection $\bigcap_{i \in I} S_i$ is convex.*

*Proof.* Let $\{S_i\}_{i \in I}$ be a family of convex sets, and let $S := \bigcap_{i \in I} S_i$. Then for any $x, y \in S_i$, $\forall i \in I$,and each of these set is convex. Hence, for any $i \in I$, and $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda) y \in S_i$. Hence, $\lambda x + (1 - \lambda) y \in S$ □

**Theorem 3.** *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 sequence $\{X_n\}$ and $\{Y_n\}$ in C. such that $X_n \to x$ and $Y_n \to y$ as $n \to \infty$. For some $\lambda, 0 \leq \lambda \leq 1$, define

$$Z_n := \lambda x_n + (1 - \lambda) y_n$$

.

Then by convexity, $Z_n \in C$. Moreover $Z_n \to \lambda x_n + (1 - \lambda) y_n \quad as \quad n \to \infty$. Hence this point lies in cl(C). □

**Defintion 1.1.3** The convex hull of a set S is the intersection of all convex sets which contain the set $S$, denote the convex hull by $con(s)$.

**Theorem 4.** *The convex hull of a compact set in $\mathbb{R}^n$ is compact.*

*Proof.* Let$S \subseteq \mathbb{R}^n$ be compact . Notice that $\rho := \{\alpha_1, \alpha_2, ..., \alpha_n\} \in \mathbb{R}^n : \sum_{i=1}^{n} \alpha_i = 1$ is also closed and bounded and is therefore compact.

Now suppose $\{U_j\}_{j=1}^{\infty} \subset con(S)$. By Caratheodory's theorem, each $U_j$ can be written in the form: $U_k := \sum_{i=1}^{n+1} \alpha_{k,i} x^{(k,i)}$, where $\alpha_{k,i} \geq 0$, $\sum_{i=1}^{n+1} \alpha_{k,i} = 1$ and $x^{(k,i)} \in S$. Then, since $S$ and $\rho$ are compact, there exists $k_1, k_2 \ldots$ such that the limits $\lim k \to \infty \alpha_{k,i} = \alpha_i$ and $\lim k \to \infty x^{(k,i)} = x^{(i)}$ exists for $i = 1, 2, ..., n+1$.Clearly $\alpha_i \geq 0, \quad \sum_{i=1}^{n} \alpha_i = 1 \quad and \quad x_i \in S.$

Thus the subsequence $\{U^{(j,k)}\}_{j=1}^{\infty}$ which converges to a point of $con(s)$. Hence it is compact. □

**Definition 1.1.4** Let$S$ be a non- empty subset of $\mathbb{R}^n$ and $x_o \in S$. A non-zero vector $d \in \mathbb{R}^n$ is said to be a feasible direction of $S$ at $x_o$ if there is $\sigma > 0$ such that $x_o + \lambda d \in S, \forall \lambda \in (0, \sigma)$.
**Note**:
1. If $d$ is a feasible direction at $x_o$, so is $\alpha d$ for any $\alpha > 0$.
2. If $S$ is open or $x_o \in int(S)$, then any non-zero $d \in \mathbb{R}^n$ is feasible direction of S at $x_o$.
3. However, a feasible direction at a boundary point may not be obvious.

## 1.2 Derivatives of functions over $\mathbb{R}^n$

**Definition 1.2.1** Let $h(y) : Y \to \mathbb{R}$, where $Y \subset \mathbb{R}^n$ is open. $h(y)$ is differentiable at $y_o \in Y$, if there exists a vector $\bigtriangledown h(y)$ (called the gradient of $h(y)$ at $y_o$) such that for each $y \in Y$.
$h(y) = h(y_o) + \bigtriangledown h(y_o)^T(y - y_o) + \|y - y_o\|\beta(y_o, y - y_o)$, and $\lim_{t\to 0} \beta(y_o, t) = 0$.
$h(y)$ is differentiable on $y$ if$h(y)$ is differentiable $\forall y_o \in Y$. The gradient vector is the vector of partial derivatives:
$\bigtriangledown h(y_o)^T = (\frac{\partial h(y_o)}{\partial y_1}, \frac{\partial h(y_o)}{\partial y_2}, ..., \frac{\partial h(y_o)}{\partial y_n})$ (if each partial derivatives exists).
**Example1:** Let $h(x, y, z) = 2x^3y^2 + 3y^3z^2$. Then

$$\bigtriangledown h(x, y, z) = (6x^2y^2, 4x^3y + 9y^2z^2, 6y^3z)^T$$

**Definition 1.2.2:** The directional derivatives of $h$ at $y$, in a direction $d \in \mathbb{R}^n$ is given by
$Dh(y; d)$=$\lim_{s\to\infty} \frac{h(y+sd)-h(y)}{s}$ (if the limit exists). If so, $Dh(y; d) = \bigtriangledown h(y)^T d$
**Definition 1.2.3:** The function $h(y)$ is twice differentiable at $y_o \in Y$ if there exists a vector $\bigtriangledown h(y_o)$ and $m \times m$ symmetric matrix $\bigtriangledown^2 h(y_o)$ (called the Hessian of $h(y)$ at $y_o$ such that for each $y \in Y$
$h(y) = h(y_o) + \bigtriangledown h(y_o)^T(y - y_o) + 1/2(y - y_o \bigtriangledown^2 h(y_o)(y - y_o) + \|y - y_o\|^2\beta(y_o, y - y_o)$, and $\lim_{t\to 0} \beta(y_o, t) = 0$. $h(y)$ is twice differentiable on $y$ if $h(y)$ is twice differentiable for all $y_o \in Y$. The hessian is the matrix of second partial derivatives:

$$\bigtriangledown^2 h(y_o)_{ij} = \frac{\partial h^2(y_o)}{\partial y_i \partial y_j}$$

**Example 2:** Continuing $Example1$, we have

$$\bigtriangledown^2 f(x) = \begin{pmatrix} 12xy^2 & 12x^2y & 0 \\ 12x^2y & 4x^3 + 18yz^2 & 18y^2z \\ 0 & 18y^2z & 6y^3 \end{pmatrix}$$

**Example 3:** Consider a quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(y) = 1/2y^T By-$

$b^T y$ where $B$ is $m \times m$ symmetric matrix and $b \in \mathbb{R}^n$. Then

$$\bigtriangledown f(y) = By - b \quad and \bigtriangledown^2 f(y) = B$$

. **Definition 1.2.4:** Let $S \subseteq \mathbb{R}^n$. $g : S \rightarrow \mathbb{R}$ is said to be differentiable at $t_o$, if $\lim_{\|r\| \rightarrow \infty} \frac{\|g(t_o+r)-g(t_o)-\bigtriangledown g(t_o)^T r\|}{\|r\|} = 0$

i.e;$g(t_o + r) = g(t_o) + \bigtriangledown g(t_o)^T r + o(\|r\|)$

Or $g(t) = g(t_o) + \bigtriangledown g(t_o)^T (t - t_o) + o(\|t - t_o\|)$. If so, we have $g(t) \approx g(t_o) + \bigtriangledown g(t_o)^T (t - t_o) for \quad sufficiently \quad near \quad t_o$. (Linear or tangential approximation of $g$ near $t_o$).

**Theorem 5.** *(First Order Taylor's Theorem or MVT) If $g$ is differentiable and $t_o, t \in \mathbb{R}^n$, then $g(t) = g(t_o) + \bigtriangledown g(\bar{t})^T (t - t_o)$ for some $\bar{t} \in (t_o, t)$*

**Definition1.2.5:** Let $S \subseteq \mathbb{R}^n$. $g : S \rightarrow \mathbb{R}$ is said to be twice differentiable at $t_o \in int(S)$, if $\lim_{\|r\| \rightarrow \infty} \frac{\|g(t_o+r)-g(t_o)-\bigtriangledown g(t_o)^T r - 1/2 r^T \bigtriangledown^2 g(t_o) r\|}{\|h\|} = 0$

i.e;$g(t_o + r) = g(t_o) + \bigtriangledown g(t_o)^T r + 1/2 r^T \bigtriangledown^2 g(t_o) r + o(\|r\|)$

Or $g(t) = g(t_o) + \bigtriangledown g(t_o)^T (t - t_o) + 1/2 (t - t_o)^T \bigtriangledown^2 g(t_o)(t - t_o) + o(\|t - t_o\|^2)$. for any $t \in S$

This yields quadratic approximation of $g(t)$.

$g(t) \approx g(t_o) + \bigtriangledown g(t_o)^T (t - t_o) + 1/2 (t - t_o)^T \bigtriangledown^2 g(t_o)(t - t_o) + o(\|t - t_o\|^2)$ for $t \quad near \quad t_o$.

**Theorem 6.** *(second order Taylor's Theorem or Second order MVT) If $g$ is twice differentiable and $t_o, t \in \mathbb{R}^n$, then*

$g(t) = g(t_o) + \bigtriangledown g(t_o)^T (t - t_o) + 1/2 (t - t_o)^T \bigtriangledown^2 g(\bar{t})^T (t - t_o)$ *for some* $\bar{t} \in (t_o, t)$

## 1.3 Definiteness of Matrices

**Definition1.3.1:** Let $B$ be a square matrix of order $m$. Then matrix $B$ is said to be:

*i.* Positive definite if $y^T By > 0$ for all non-zero $y \in \mathbb{R}^n$, negative definite if $y^T By < 0$ for all non-zero $y \in \mathbb{R}^n$ and indefinite if $x^T Bx > 0$ for some $x$ and $y^T By < 0$ for some $y$.

*ii.* Positive semi-definite if $y^T By \geq 0$ for all $y \in \mathbb{R}^n$ and negative semi-definite if, $y^T By \leq 0$ for all $y \in \mathbb{R}^n$

For a symmetric matrix $B$ we can summarize the definiteness of matrix $B$ as follows: Let $B_r$ be the $r \times r$ minor sub matrix of $B$; $r = 1, 2, ..., k$.

i. $B$ is Positive definite if and only if $det(B_r) > 0 \; \forall r = 1, 2, ..., k$ or if and only if all the eigenvalue values of B are positive and Positive semi-definite if and only if $det(B_r) \geq 0 \; \forall r = 1, 2, ..., k$ or if and only if all the eigenvalues B are greater than or equal to zero.

ii. $B$ is negative definite if and only if $det(B_r) < 0 \; \forall r = 1, 2, ..., k$ or if and only if all the eigenvalue values of B are negative and negative semi-definite if and only if $det(B_r) \leq 0 \; \forall r = 1, 2, ..., k$ or if and only if all the eigenvalues B are less than or equal to zero.

**Example1:**

$B = \begin{pmatrix} 4 & 0 \\ 0 & 5 \end{pmatrix}$ is a positive definite. Since, for $y = (y_1, y_2) \neq 0$,

$$y^T By = 4y_1^2 + 5y_2^2 > 0, \forall y_1, y_2 \in \mathbb{R}^2$$

**Example2:**

$B = \begin{pmatrix} 7 & -2 \\ -2 & 1 \end{pmatrix}$ is a positive definite. Since, for $y = (y_1, y_2) \neq 0$,

$$y^T By = 7y_1^2 - 4y_1y_2 + y_2^2 = 3y_1^2 + (2y_1 - y_2)^2 > 0, \forall y_1, y_2 \in \mathbb{R}^2$$

.

## 1.4 Local, Global and Strict Optima

The ball centered at $\bar{x}$ with radius $\epsilon$ is the set $s$.

$$B(\bar{x}, \epsilon) := \{x : \|x - \bar{x}\| \leq \epsilon\}$$

Consider the following optimization problem over the set $S$:

$$P : min_x or max_x f(x)$$
$$s.t \quad x \in S$$

.

**Definition1.4.1:** $t \in S$ is local minimum of $P$ if there exits $\epsilon > 0$ such that $h(t) \leq h(s) \; \forall s \in B(\bar{t}, \epsilon) \bigcap S$.

**Definition1.4.2:** $t \in S$ is strictly local minimum of $P$ if there exits $\epsilon > 0$ such that $h(t) < h(s) \; \forall s \in B(\bar{t}, \epsilon) \bigcap S$, $t \neq s$.

**Definition1.4.3:** $t \in S$ is global minimum of $P$ if $h(t) \leq h(s) \; \forall s \in S$.

**Definition1.4.4:** $t \in S$ is strictly global minimum of $P$ if $h(t) < h(s) \; \forall s \in S, t \neq s$.

**Definition1.4.5:** $t \in S$ is local maximum of $P$ if there exits $\epsilon > 0$ such that $h(t) \geq h(s) \; \forall s \in B(\bar{t}, \epsilon) \bigcap S$.

**Definition1.4.6:** $t \in S$ is strictly local maximum of $P$ if there exits $\epsilon > 0$ such that $h(t) > h(s) \; \forall s \in B(\bar{t}, \epsilon) \bigcap S, s \neq t$.

**Definition1.4.7:** $t \in S$ is global maximum of $P$ if $h(t) \geq h(s) \; \forall s \in S$.

**Definition1.4.8:** $t \in S$ is strictly global maximum of $P$ if $h(t) > h(s) \; \forall s \in S, s \neq t$.

## 1.5 Convex function

**Definition 1.5.1:** Let $S \subseteq \mathbb{R}^n$ be a convex set. A function $h : S \to \mathbb{R}$ is said to be convex (over$S$) if for every $t_1, t_2 \in S$ and every $\beta \in [0, 1]$,

$$h(\beta t_1 + (1-\beta)t_2) \leq \beta h(t_1) + (1-\beta)h(t_2)$$

**Definition 1.5.2:** Let $S \subseteq \mathbb{R}^n$ be a convex set. A function $h : S \to \mathbb{R}$ is said to be concave (over$S$) if for every $t_1, t_2 \in S$ and every $\beta \in [0, 1]$,

$$h(\beta t_1 + (1-\beta)t_2) \geq \beta h(t_1) + (1-\beta)h(t_2)$$

**Definition 1.5.3:** Let $S \subseteq \mathbb{R}^n$ be a convex set. A function $h : S \to \mathbb{R}$ is said to be strictly convex (concave) (over$S$) if for every $t_1, t_2 \in S$ and every $\beta \in [0, 1]$,

$$h(\beta t_1 + (1-\beta)t_2) < (>)\beta h(t_1) + (1-\beta)h(t_2)$$

**Examples:**
1. Given $y_o \in \mathbb{R}^n$, let $d : \mathbb{R}^n \to \mathbb{R}$ be given by $d(y) = \|y - y_o\|$. Then $d$ is a convex function.

*Proof.* Let $y_1, y_2 \in \mathbb{R}^n$, then
$d(\beta y_1 + (1-\beta)y_2) = \|(\beta y_1 + (1-\beta)y_2 - y_o\|, for \quad every \quad \beta \in [0, 1]$
$=\|\beta y_1 + (1-\beta)y_2 - (\beta y_o + (1-\beta)y_o)\|, for \quad y_o = \beta y_o + (1-\beta)y_o$
$= \|(\beta y_1 - \beta y_o) + ((1-\beta)y_2 - (1-\beta)y_o)\|$
$= \|\beta(y_1 - y_o) + (1-\beta)(y_2 - y_o)\|$
$\leq \beta\|y_1 - y_o\| + (1-\beta)\|y_2 - y_o\|$, by triangular inequality
$= \beta d(y_1) + (1-\beta)d(y_2)$, by definition
Hence $d$ is convex function. □

2. Let $a \in \mathbb{R}^n, b \in \mathbb{R} \quad and \quad h : \mathbb{R}^n \to \mathbb{R}$ be affine function given by $h(y) = a^T y + b$. Then, $h$ is convex function.

*Proof.* To show that $h$ is convex function, let $y_1, y_2 \in \mathbb{R}^n$ and for every $\beta \in [0, 1]$, then
$h(\beta y_1 + (1-\beta)y_2) = a^T(\beta y_1 + (1-\beta)y_2) + b$

$$= a^T(\beta y_1 + (1-\beta)y_2) + (\beta b + (1-\beta)b), for \quad b = \beta b + (1-\beta)b$$

$$= \beta a^T y_1 + (1-\beta)a^T y_2 + \beta b + (1-\beta)b$$
$$= \beta(a^T y_1 + b) + (1-\beta)(a^T y_2 + b)$$

$$= \beta h(y_1) + (1-\beta)h(y_2).$$

Hence $h$ is convex function.
In fact affine function is also concave function. $\square$

3. A function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) = x^2$ is a convex function.

*Proof.* Let $x_1, x_2 \in \mathbb{R}^2$, then
$f(\lambda x_1 + (1-\lambda)x_2) = (\lambda x_1 + (1-\lambda)x_2)^2, for \quad every \lambda \in [0,1]$

$$= \lambda^2 (x_1)^2 + (1-\lambda)^2 (x_2)^2 + 2\lambda(1-\lambda)x_1 x_2$$

$$\leq \lambda^2 (x_1)^2 + (1-\lambda)^2 (x_2)^2 + \lambda(1-\lambda)[(x_1)^2 + (x_2)^2], since (x-y)^2 \geq 0 \forall x, y \in \mathbb{R} \Rightarrow x^2 + y^2 \geq 2xy$$

$$\lambda^2 (x_1)^2 + (x_2)^2 - 2\lambda (x_2)^2 + \lambda^2 (x_2)^2 + \lambda (x_1)^2 + \lambda (x_2)^2 - \lambda^2 (x_1)^2 - \lambda^2 (x_2)^2$$

$$= (x_2)^2 - \lambda (x_2)^2 + \lambda (x_1)^2$$

$$= \lambda (x_1)^2 + (1-\lambda)(x_2)^2$$

$$= \lambda f(x_1) + (1-\lambda) f(x_2)$$

Therefore$f(x) = x^2$ is convex function. $\square$

**Theorem 7.** *If $h_1, h_2 : \mathbb{R}^n \to \mathbb{R}$ are convex functions, then $H(y) = h_1(y_1) + h_2(y_2)$ is convex function.*

*Proof.* Let $y_1, y_2 \in \mathbb{R}^n$ and for every $\gamma \in [0,1]$, then
$H(\gamma y_1 + (1-\gamma)y_2) = h_1(\gamma y_1 + (1-\gamma)y_2) + h_2(\gamma y_1 + (1-\gamma)y_2)$

$$\leq \gamma h_1(y_1) + (1-\gamma)h_1(y_2) + \gamma h_2(y_1) + (1-\gamma)h_2(y_2),$$

by convexity of $h_1$ and $h_2$

$$= \gamma(h_1(y_1) + h_2(y_1)) + (1-\gamma)(h_1(y_2) + h_2(y_2))$$

$$= \gamma H(y_1) + (1-\gamma)H(y_2)$$

Therefore $h_1 + h_2$ is convex function $\square$

**Theorem 8.** *Let $S \subseteq \mathbb{R}^n$ be convex set. If $h : S \to \mathbb{R}$ is a convex function, then the level set $S_\gamma = \{y \in S : h(y) \leq \gamma\}, where \gamma \in \mathbb{R}$, is convex set.*

*Proof.* Let $y_1, y_2 \in S_\gamma, \gamma \in [0,1]$ and $y = \gamma y_1 + (1-\gamma)y_2$, then by convexity of $h$

$$\begin{aligned} h(\beta y_1 + (1-\beta)y_2) &\leq \beta h(y_1) + (1-\beta)h(y_2) \\ &\leq \beta\gamma + (1-\beta)\gamma \\ &= \gamma \\ &\Rightarrow \beta y_1 + (1-\beta)y_2 \in S_\gamma \end{aligned}$$

□

**Theorem 9.** *(Sub gradient inequality): Let $S \subseteq \mathbb{R}^n$ is convex and $h : S \to \mathbb{R}$ is differentiable. $h(x) \geq h(y) + \bigtriangledown h(y)^T(x-y)$ for every $x, y \in S$ if and only if $h$ is convex on $S$.*

*Proof.* ($\Rightarrow$) Suppose $h(x) \geq h(y) + \bigtriangledown h(y)^T(x-y), \forall x, y \in S$.
We want to show that $h$ is convex function on $S$.
Let $y_1, y_2 \in S$ and $y = \beta y_1 + (1-\beta)y_2, \beta \in [0,1]$
Then $h(y_1) \geq h(y) + \bigtriangledown h(y)^T(y_1 - y)$ and
$h(y_2) \geq h(y) + \bigtriangledown h(y)^T(y_2 - y)$
Since $y = \beta y_1 + (1-\beta)y_2$

$$\begin{aligned} = y_2 + \beta(y_1 - y_2) \Rightarrow y_1 - y &= (1-\beta)(y_1 - y_2) \\ \Rightarrow y_2 - y &= -\beta(y_1 - y_2) \end{aligned}$$

Thus by substitution,

$$h(y_1) \geq h(y) + (1-\beta) \bigtriangledown h(y)^T(y_1 - y_2)$$
$$h(y_2) \geq h(y) + (1-\beta) \bigtriangledown h(y)^T(y_1 - y_2)$$

Then multiplying the first inequality by $\beta$ and the second inequality by $(1-\beta)$ and add

$$\begin{gathered} \beta h(y_1) \geq \beta h(y) + \beta(1-\beta) \bigtriangledown h(y)^T(y_1 - y_2) \\ (1-\beta)h(y_2) \geq (1-\beta)h(y) - \beta(1-\beta) \bigtriangledown h(y)^T(y_1 - y_2) \\ \Rightarrow \beta h(y_1) + (1-\beta)h(y_2) \geq \beta h(y) + (1-\beta)h(y) \\ \beta h(y_1) + (1-\beta)h(y_2) \geq h(y) \\ h(\beta y_1 + (1-\beta)y_2) \leq \beta h(y_1) + (1-\beta)h(y_2) \end{gathered}$$

Hence $h$ is convex function on $S$.

($\Leftarrow$) Suppose $h$ is convex
$h(y+\beta(x-y)) \leq h(y)+\beta(h(x-y)), \beta \in [0,1]$ by convexity of $h$ or by assumption
$h(y+\beta(x-y)) - h(x) \leq \beta(h(x-y))$
$h(x-y) \geq \frac{(h(y+\beta(x-y))-h(y))}{\beta}$
$h(x)-h(y) \geq \frac{(h(y+\beta(x-y))-(y))}{\beta}$, so by taking $\beta \to 0^+$, we get

$$h(x)-h(y)) \geq \lim_{\beta \to 0^+)} \frac{(h(y+\beta(x-y))-h(y))}{\beta}$$

$$h(x)-h(y) \geq \bigtriangledown h(y)^T(x-y)$$

$$\Rightarrow h(x) \geq h(y) + \bigtriangledown h(y)^T(x-y)$$

□

**Theorem 10.** *Let $S \subseteq \mathbb{R}^n$ is convex set and $g : S \to \mathbb{R}$ is twice continuous differentiable. $g$ is convex on $S$ if and only if its hessian $\bigtriangledown^2 g(y)$ is $PSD$ at each $y \in S$.*

*Proof.* ($\Rightarrow$) suppose $g$ is convex on $S$.
We want show that hessian is $PSD$ i.e; $\bigtriangledown^2 g(y) \geq 0$.
$g(y+th) \geq g(y) + t \bigtriangledown g(y)^T h$ by convexity of $g$ or by theorem above for $t > 0$ and $h \in \mathbb{R}^n$. from second order Taylor's Theorem, we have
$g(y+th) = g(y) + t \bigtriangledown g(y)h + \frac{1}{2}t^2 h^T \bigtriangledown^2 g(y+\theta th)h, \theta \in (0,1)$
$g(y+th) - g(y) - t \bigtriangledown g(y)h = \frac{1}{2}t^2 h^T \bigtriangledown^2 g(y+\theta th)h$
$\Rightarrow t^2 h^T \bigtriangledown^2 g(y+\theta th)h \geq 0 \forall t > 0 and \theta \in (0,1), h \in \mathbb{R}^n$ by convexity of $g$ and by the above theorem
$h^T \bigtriangledown^2 g(y+\theta th)h \geq 0$, take $t \to 0$
$\Rightarrow h^T \bigtriangledown^2 g(y)h \geq 0, \forall h \in \mathbb{R}^n$.
Therefore $\bigtriangledown^2 g(y)$ is $PSD$
($\Leftarrow$) Suppose $\bigtriangledown^2 g(y)$ is $PSD$
We want to show that $g$ is convex.

$$(x-y)^T \bigtriangledown^2 g(y_o)(x-y) \geq 0 for \quad any \quad x, y \in S.$$

By Second Order Taylor's Theorem

$$g(x) - [g(y) + \bigtriangledown g(y)^T(x-y)] = \frac{1}{2}(x-y)^T \bigtriangledown^2 g(x)(x-y)$$

$$\Rightarrow g(x) - [g(y) + \bigtriangledown g(y)^T(x-y)] \geq 0,$$

by assumption

$$\Rightarrow g(x) \geq g(y) + \bigtriangledown g(y)^T(x-y)$$

Hence $g$ is convex by the above theorems. □

**Example:** $f(x,y,z) = x^4 + y^3 + e^z - 5y$ is convex function on $S = \{(x,y,z) \in \mathbb{R}^3 : y \geq 0\}$.
**Solution:**

$$f_x = 4x^3 \quad f_y = 3y^2 - 5 \quad f_z = e^z$$

$$f_{xx} = 12x^2 \quad f_{yx} = 0 \quad f_{zx} = 0$$

$$f_{xy} = 0 \quad f_{yy} = 6y \quad f_{zy} = 0$$

$$f_{xz} = 0 \quad f_{yz} = 0 \quad f_{zz} = e^z$$

$$\bigtriangledown^2 f(x,y,z) = \begin{bmatrix} 12x^2 & 0 & 0 \\ 0 & 6y & 0 \\ 0 & 0 & e^z \end{bmatrix}$$

Every diagonal element of $\bigtriangledown^2 f(x,y,z)$ is a non-negative over $S$. That is, $\bigtriangledown^2 f(x,y,z)$ is a diagonal matrix. Thus $\bigtriangledown^2 f(x,y,z)$ is a PSD on $S$.
Hence $f$ is convex by the above theorem.
**Example:** Find a domain (set) on which $f(x,y) = -x^3 + y^2 + y$ is convex.
**Solution:** Since

$$\bigtriangledown^2 f(x,y) = \begin{pmatrix} -6x & 0 \\ 0 & 2 \end{pmatrix}$$

$\Rightarrow \bigtriangledown^2 f(x,y) \geq 0, \forall x \in (-\infty, 0]$
Hence $f(x,y) = -x^3 + y^2 + y$ is convex on $S = \{(x,y) \in \mathbb{R}^2 : x \leq 0\}$
**Definition:**Let$f : S \to \mathbb{R}$.
1. The epigraph of $f$ is the set $epi(f) := \{(x,y) \in S \times \mathbb{R} : f(x) \leq y$ .
2. The hypograph of $f$ is the set $hyp(f) := \{(x,y) \in S \times \mathbb{R} : f(x) \geq y$.

**Theorem 11.** *Let $S \subseteq \mathbb{R}^n$ be convex and $f : S \to \mathbb{R}$ is convex function if and only if $epi(f)$ is convex set.*

*Proof.* ($\Rightarrow$) Suppose $f$ is convex function
$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)\ \forall x_1, x_2) \in S \text{ and } \lambda \in [0,1] \ldots (*)$
Let $(x_1, y_1), (x_2, y_2) \in epi(f) \Rightarrow f(x_1) \leq y_1$.

$$f(x_2) \leq y_2$$

.
From (*)
$\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2) = (x, y), \lambda \in [0,1]$

$$\Rightarrow x = \lambda x_1 + (1-\lambda)x_2$$

$$y = \lambda y_1 + (1-\lambda)y_2$$

$$f(x) = f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2), \text{ by assumption}$$
$$\leq \lambda y_1 + (1-\lambda)y_2$$
$$= y$$

$$\Rightarrow f(x) \leq y, i.e; epi(f) is convex set.$$

($\Leftarrow$) Suppose $epi(f)$ is convex set

$$(x_1, f(x_1)), (x_2, f(x_2)) \in epi(f)$$

$\Rightarrow \lambda(x_1, f(x_1)) + (1-\lambda)(x_2, f(x_2)) \in epi(f), \forall \lambda \in [0,1]$
$\Rightarrow (\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda)f(x_2)) \in epi(f)$

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

$\Rightarrow f \quad is \quad convex.$ □

**Chapter 2**

# 2 Optimality Condition for Smooth Unconstrained Optimization Problems

## 2.1 Methods of Unconstrained Problem

Unconstrained optimization problem:

$$(P) min (or max) g(y)$$

$$s.t \quad y \in Y$$

Where$y = (y_1, y_2, ..., y_n) \in \mathbb{R}^n, g(y) : \mathbb{R}^n \to \mathbb{R}$, and $Y$ is an open set
We say that $y$ is a feasible solution of $(P)$ if $y \in Y$.
**Definition 2.1:** The optimization problems in which the objectives function is differentiable and functions involved in the constraints are continuously differentiable is called smooth optimization problem.
**Definition 2.2:** A non-zero vector $d \in \mathbb{R}^n$ such that

$$g(t_o + rd) < g(t_o) \quad \forall r \in (0, \delta)$$

if there exist $\delta > 0$ then $d$ is called descent direction of $g$ at $t_o$.
**Remark:** Descent direction is a direction along which $g$ (locally) decreases.
**Definition 2.3** A non-zero vector $d \in \mathbb{R}^n$ is said to be ascent direction if,

$$g(t_o + rd) > g(t_o) \quad \forall r \in (0, \delta),$$

if there exists $\delta > 0$.

**Theorem 12.** *Suppose that $g(t)$ is differentiable at $\bar{t}$. If there is a non-zero vector $d \in \mathbb{R}^n$ such that $\bigtriangledown g(\bar{t})^T d < 0$, then for all $\delta > 0$ and sufficiently small, $g(\bar{t} + \delta d) < g(\bar{t})$ and $d$ is a descent direction of a function $g$ at $t$*

*Proof.* - Since we have:
$g(\bar{t} + \beta d) = g(\bar{t}) + \beta \bigtriangledown g(\bar{t})^T d + \beta \|d\| \gamma(\bar{t}, \beta d)$, where $\gamma(\bar{t}, \beta d) \to 0$ as $\beta \to 0$.
Rearranging, $\frac{(g(\bar{t}+\beta d) - g(\bar{t}))}{\beta} = \bigtriangledown g(\bar{t})^T d + \beta \|d\| \gamma(\bar{t}, \beta d)$
Since $\bigtriangledown g(\bar{t})^T d < 0$ and $\gamma(\bar{t}, \beta d) \to 0$ as $\beta \to 0$,

$$g(\bar{t} + \beta d) - g(\bar{t}) < 0 \quad \forall \beta > 0$$

and sufficiently and small. $\square$

**Corollary 2.1:** If $\bigtriangledown g(t_o) \neq 0$, then $- \bigtriangledown g(t_o)$ is a descent direction of $g$ at $t_o$.

*Proof.* If $\bigtriangledown g(t_o) \neq 0$ , take $d = - \bigtriangledown g(t_o)$

$$\bigtriangledown g(t_o)^T d = -\| \bigtriangledown g(t_o)\|^2 < 0$$

Therefore $d$ is a descent direction. □

**Corollary 2.2:** If $t^*$ is local minimum point of $g$, then $\bigtriangledown g(t^*)d > 0$ for every non-zero $d \in \mathbb{R}^n$ (i. e; there is no descent direction of $g$ at $t^*$).

*Proof.* By contradiction suppose $d$ is a descent direction of $g$ at $t^*$ i.e; let $g(t^*) \leq g(t) \forall t \in N_\epsilon(t^*)$

$$\Rightarrow \exists \delta > 0 \quad such \quad that \quad g(t_o + rd) \leq g(t_o), \forall r \in (0, \delta)$$

This contradicts that, if $t_o$ is local minimum of $g$, $g(t_o) \leq g(t) \forall t \in N_\epsilon(t_o)$
Therefore $d$ is not a descent direction. □

## 2.2 Classical Optimality Condition

In this sub section, we will look for conditions that are usefull to identify (local) optimal solution for the smooth unconstrained NLP,

$$min(ormax) f(x)$$

$$s.t \quad x \in \mathbb{R}^n$$

Where $f \in C^2(\mathbb{R}^n)$i.e; twice continuous differentiable.

**Theorem 13.** ***(First Order Necessary Condition)*** *Suppose for a differentiable function $g : \mathbb{R}^n \to \mathbb{R}$ at $\bar{t}$, if $\bar{t} \in \mathbb{R}^n$ is a local minimum of $g$, then $\bigtriangledown g(\bar{t}) = 0$.*

*Proof.* If it were true that $\bigtriangledown g(\bar{t}) \neq 0$, then $d = - \bigtriangledown g(\bar{t}) = 0$ would be a descent direction, where by $\bar{t}$ would not be a local minimum. □

**Note:** The converse of the above theorem 2.2 is not true.
Recall that: If $\bigtriangledown g(t_o) = 0$, then $t_o$ is called stationary (critical) point.
**Example1.** Let $f : \mathbb{R} \to \mathbb{R}, f(x) = 2x^3 - 3x^2$. Then
$Df(x) = 6x^2 - 6x = 6x(x - 1)$, which implies that the only candidates for a maximum or minimum are $x = 0$ or $x = 1$ without further conditions, however we cannot say whether these are actual maxima or minima.
**Example2.** Let $g : \mathbb{R}^2 \to \mathbb{R}, g(x, y) = 2x^3 - 2y^3 + 4xy$

Then $Df(x,y) = (6x^2 + 4y, -6y^2 + 4x)$ which implies that the only candidates for a maximum or minimum are when $6x^2 + 4y = 0$ and $-6y^2 + 4x = 0$. Solving the first equation for $y$ yields $y = \frac{-3}{2}x^2$. Plugging this into the other equation we have:

$$0 = -6y^2 + 4x = -6(\frac{-3}{2}x^2)^2 + 4x = \frac{-27}{2}x^4 + 4x = 8x - 27x^4 = x(8 - 27x^3)$$

Which implies that $x = 0$ and $x = \frac{2}{3}$ are possible solutions. Plugging the $x$ solutions into the equation for $y$ gives $y = 0$ and $y = \frac{-2}{3}$ respectively. Therefore, the only possible optima are at $(x,y) = (0,0)$ and $(x,y) = (\frac{2}{3}, \frac{-2}{3})$. Without, further conditions, however, we cannot say whether these are actual maxima or minima.

**Theorem 14.** ***(Second Order Necessary conditions)****. Suppose that $g(y)$ is twice continuous differentiable at $\bar{y} \in Y \subseteq \mathbb{R}^n$. If $\bar{y}$ is a local minimum, then $\bigtriangledown g(\bar{y}) = 0$ and $\bigtriangledown^2 g(\bar{y})$ is positive semi definite.*

*Proof.* $\bar{y}$ is local minimum $\Rightarrow \bigtriangledown g(\bar{y}) = 0$, follows first order necessary condition.
$\bar{y}$ is local minimum $\Rightarrow \bigtriangledown^2 g(\bar{y})$ is positive semi definite follows from second order Taylor's Theorem and first order necessary condition for any non zero $d \in \mathbb{R}^n$. We have

$$(g(\bar{y} + rd) = g(\bar{y}) + r \bigtriangledown g(\bar{t})^T d + \frac{1}{2} r d^T \bigtriangledown^2 g(\bar{y} + r\gamma d)d, \gamma \in (0,1)$$

$$\Rightarrow (g(\bar{y} + rd) = g(\bar{y}) + \frac{1}{2} r d^T \bigtriangledown^2 g(\bar{t} + r\gamma d)d, \gamma \in (0,1)$$

, by FONC

$$\Rightarrow \frac{1}{2} r d^T \bigtriangledown^2 g(\bar{y} + r\gamma d)d \geq 0 \forall r \in (0, \delta)$$

if there exist $\delta > 0$, Since $d$ is a descent direction i.e; $(g(\bar{y} + rd) - g(\bar{y}) < 0$
By taking the limit at $r \to 0$

$$d^T \bigtriangledown^2 g(\bar{y})d \geq 0$$

which implies that $\bigtriangledown^2 g(\bar{y})$ is $PSD$. □

**Example3.** Let $g(x,y) = x^2 + 2xy + 3y^2 - 4x - 4y - \frac{8}{3}y^3$

$$Then \bigtriangledown g(x,y) = (2x + 2y - 4, 2x + 6y - 4 - 8y^2)^T$$

and

$$\bigtriangledown^2 g(x,y) = \begin{pmatrix} 2 & 2 \\ 2 & 6 - 16y \end{pmatrix}$$

$\bigtriangledown g(x,y)=0$ has exactly two solutions:$(x,y)=(2,0)$ and $(x,y)=(\frac{3}{2},\frac{1}{2})$. But at $(x,y)=(\frac{3}{2},\frac{1}{2})$

$$\bigtriangledown^2 g(x,y)=\begin{pmatrix}2 & 2\\ 2 & -2\end{pmatrix}$$

is indefinite

Therefore the only possible candidate for a local minimum is $(x,y)=(2,0)$.

A sufficient conditions for a local optimality is a statement of the form "if $\bar{x}$ satisfies...,then $\bar{x}$ is a local minimum of $p$". Such a condition allows us to automatically declare that $\bar{x}$ is indeed a local minimum.

**Theorem 15.** ***(Second Order Sufficient Conditions)*** *Let $g$ be twice differentiable at $\bar{t}$. If $\bigtriangledown g(\bar{t})=0$ and $\bigtriangledown^2 g(\bar{t})$ is positive definite, then $\bar{t}$ is a strict local minimum.*

*Proof.* From second order Taylor's theorem, we have

$$g(t)=g(\bar{t})+\bigtriangledown g(\bar{t})^T(t-\bar{t})+\frac{1}{2}(t-\bar{t})^T\bigtriangledown^2 g(\bar{t})(t-\bar{t})+\|t-\bar{t}\|^2\gamma(\bar{t},t-\bar{t})$$

Suppose that $\bar{t}$ is not strict local minimum. Then there exists a sequence $t_k\to\bar{t}$ such that $t_k\neq\bar{t}$ and $g(t_k)\leq g(\bar{t})\forall k$ Define $d_k=\frac{t_k-\bar{t}}{\|t_k-\bar{t}\|}$, Then

$$g(t_k)=g(\bar{t})+\|t-\bar{t}\|^2(\frac{1}{2}d_k^T\bigtriangledown^2 g(\bar{t})d_k+\gamma(\bar{t},t_k-\bar{t}))$$

And so

$$\frac{1}{2}d_k^T\bigtriangledown^2 fg(\bar{t})d_k+\gamma(\bar{t},t_k-\bar{t})=\frac{(g(t_k)-g(\bar{t}))}{\|t-\bar{t}\|^2}\leq 0$$

Since$\|d_k\|=1$ for any$k$, there exists a subsequence of$\{d_k\}$ converging to some point $d$ such that $\|d\|=1$. Assume without loss of generality that $d_k\to d$. Then

$$0\geq\lim_{k\to 0}d_k^T\bigtriangledown^2 g(\bar{t})d_k+\gamma(\bar{t},t_k-\bar{t})=\frac{1}{2}d^T\bigtriangledown^2 g(\bar{t})d$$

Which is a contradiction of the positive definiteness of $\bigtriangledown^2 g(\bar{t})$ $\square$

**Note:**

*i.* If $\bigtriangledown g(\bar{t})=0$ *and* $\bigtriangledown^2 g(\bar{t})$ is negative definite, then $\bar{t}$ is a local maximum.

*ii.* If$\bigtriangledown g(\bar{t})=0$ *and* $\bigtriangledown^2 g(\bar{t})$ is positive semi definite, we cannot be sure is $\bar{t}$ is local

minimum.
**Example4.** Continuing example 3 above, we compute

$$\bigtriangledown^2 g(x,y) = \bigtriangledown^2 g(2,0) = \begin{pmatrix} 2 & 2 \\ 2 & 6 \end{pmatrix}$$

is positive definite. To see this, note that for any $t = (t_1, t_2)$ we have,

$$t^T \bigtriangledown^2 g(\bar{x})t = 4t_1^2 + 2t_1t_2 + 6t_2^2 = 3t_1^2 + (t_1 + t_2)^2 + 5t_2^2 > 0 \forall t \in \mathbb{R}^2.$$

Therefore, $\bar{x}$=$(x, y)$=$(2, 0)$ satisfies the sufficient conditions to be local minimum and so $\bar{x}$=$(x, y)$=$(2, 0)$ is a local minimum.
**Example5.**

$$min \quad g(x,y) = 10(x-y)^2 + (1-x)^2$$
$$(x,y) \in R^2$$

**Solution:** Thus we have $\bigtriangledown g(x,y) = (20(x-y) - 2(1-x), -20(x-y))^T$
$\bigtriangledown g(x,y) = 0$
$(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y})^T$=$(0,0)^T$
$\Rightarrow (x,y)^T$=$(1,1)^T$ is stationary point.
Now, the Hessian matrix,

$$\bigtriangledown^2 g(x,y) = \begin{vmatrix} g_{xx} & g_{xy} \\ g_{yx} & g_{yy} \end{vmatrix}$$

$$= \begin{vmatrix} 22 & -20 \\ -20 & 20 \end{vmatrix}$$

$$= 40 > 0$$

is positive definite. Therefore $(x,y) = (1,1)^T$ is a minimizer.
**Example6.**

$$min f(x) = y^2 - 3x^2y + 2x^4$$
$$x \in R^2$$

**Solution:** $\bigtriangledown f(x) = 0$

$$\Rightarrow (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})^T = (0,0)^T$$

$$\Rightarrow (-6xy + 8x^3, 2y - 3x^2)^T = (0,0)^T$$

$$\Rightarrow \bar{x} = (x,y)^T = (0,0)^T$$

is stationary point (or critical point)

Now, the Hessian matrix $\bigtriangledown^2 f(\bar{x}) = \bigtriangledown^2 f(0,0)^T = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix}$

$$= \begin{vmatrix} -6y + +24x^2 & -6x \\ -6x & 2 \end{vmatrix}$$

$$= \begin{vmatrix} 0 & 0 \\ 0 & 2 \end{vmatrix} = 0 \geq 0 \quad is \quad positive \quad semi \quad definite.$$

Therefore $\bar{x} = (0,0)^T$ is not a minimizer.

**Theorem 16.** ***(Convex Optimality Condition)***
*Let $g$ be twice differentiable at $\bar{t}$. If $\bigtriangledown g(\bar{t}) = 0$ and $\bigtriangledown^2 g(\bar{t})$ is positive semi definite at every $t$, then $\bar{t}$ is a global minimum.*

*Proof.* By second order Taylor's theorem, for every $t$,

$$g(t) = g(\bar{t}) + \bigtriangledown g(\bar{t})^T(t-\bar{t}) + \frac{1}{2}(t-\bar{t}^T \bigtriangledown^2 g(t_o)(t-\bar{t}) for some t_o \in (t,\bar{t})$$

$$= g(\bar{t}) + \frac{1}{2}(t-\bar{t})^T \bigtriangledown^2 g(t_o)(t-\bar{t}), \quad since \quad \bigtriangledown g(\bar{t}) = 0$$

$$\geq g(\bar{t}), \quad since \quad \bigtriangledown^2 g(\bar{t}) \quad is \quad PSD$$

Therefore $\bar{t}$ is a global minimizer. □

**Remark:** Let $g$ be twice differentiable at $\bar{t}$. If $\bigtriangledown g(\bar{t}) = 0$ and $\bigtriangledown^2 g(\bar{t})$ is indefinite, then $\bar{t}$ is a saddle point which is called saddle point condition.

**Remark: (Problem with two variables)**
The second order sufficient condition and saddle point condition can be restated for a problem with two variables as follows.
Let $g : \mathbb{R}^2 \to \mathbb{R}$ is twice differentiable at some $\bar{t} \in \mathbb{R}^2$ and suppose $\bar{t}$ is stationary point. i.e; $g_t(\bar{t}) = 0 = g_s(\bar{t})$

$$Let \quad D = \begin{vmatrix} g_{tt}(\bar{t}) & g_{ts}(\bar{t}) \\ g_{st}(\bar{t}) & g_{ss}(\bar{t}) \end{vmatrix} \quad = g_{tt}(\bar{t})g_{ss}(\bar{t}) - g_{ts}^2(\bar{t})$$

i. If $g_{tt}(\bar{t}) > 0$ and $D > 0$,then $\bar{t}$ is a local minimum.
ii. If $g_{tt}(\bar{t}) < 0$ and $D > 0$,then $\bar{t}$ is a local maximum.
iii. If $D < 0$,then $\bar{t}$ is a saddle point.

**Example7:** Find (local) minimum of the given functions (find stationary points and identify their types)

1.$f(x,y)=(x^2-y)^2+y^2-8y$
2.$h(x_1,x_2)=2x_1^2+x_2^3-12x_1-27x_2$
3.$g(x_1,x_2,x_3)=(x_1-x_2)^2+x_3^2-6x_3$
**Solution:**
1.$f(x,y)=(x^2-y)^2+y^2-8y$

$$f_x=4x(x^2-y) \quad f_y=-2x^2+4y-8$$

$$f_{xx}=12x^2-4y \quad f_{yx}=-4x$$

$$f_{xy}=-4x \quad f_{yy}=4$$

$\bigtriangledown f(x,y)=0 \Rightarrow \begin{pmatrix} 4x(x^2-y) \\ -2x^2+4y-8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \bar{x}=(-2,4) and (2,4)$ are stationary point

$$\bigtriangledown^2 f(x,y)=\begin{pmatrix} 12x^2-4y & -4 \\ -4x & 4 \end{pmatrix}$$

$\bigtriangledown^2 f(\bar{x})=\bigtriangledown^2 f(-2,4)$ and $\bigtriangledown^2 f(2,4)$ has the determinant $D=160>0$ and $D=96>0$ respectively and $f_{xx}=32>0$.
Hence, by the above remark $(-2,4)$ and $(2,4)$ are a local minimum point.
2. $h(x_1,x_2)=2x_1^2+x_2^3-12x_1-27x_2$

$$h_{x_1}=4x_1-12 \quad h_{x_2}=3x_2^2-27$$

$$h_{x_1x_1}=4 \quad h_{x_2x_1}=0$$

$$h_{x_1x_2}=0 \quad h_{x_2x_2}=6x_2$$

$\bigtriangledown h(x_1,x_2)=0 \Rightarrow \begin{pmatrix} 4x_1-12 \\ 3x_2^2-27 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \bar{x}=(x_1,x_2)=(3,-3) and (3,3)$ are stationary point

$$\bigtriangledown^2 h(x_1,x_2)=\begin{pmatrix} 4 & 0 \\ 0 & 6x_2 \end{pmatrix}$$

$\bigtriangledown^2 h(\bar{x})=\bigtriangledown^2 h(3,-3)=\begin{pmatrix} 4 & 0 \\ 0 & -18 \end{pmatrix}$
Since its determinant is $D=-72<0$, Hence $(3,-3)$ is a saddle point by saddle point condition.
$\bigtriangledown^2 h(\bar{x})=\bigtriangledown^2 h(3,3)=\begin{pmatrix} 4 & 0 \\ 0 & 18 \end{pmatrix}$
$\bigtriangledown^2 h(3,3)$ has the determinant $D=72>0$ and $h_{x_1x_1}=4>0$ Hence, by the above

remark $(3,3)$ is a local minimum point.
3. $g(x_1,x_2,x_3)=(x_1-x_2)^2+x_3^2-6x_3$

$$g_{x_1}=2(x_1-x_2)\quad g_{x_2}=-2(x_1-x_2)\quad g_{x_3}=2x_3-6$$

$$g_{x_1x_1}=2\quad g_{x_2x_1}=-2\quad g_{x_3x_1}=0$$

$$g_{x_1x_2}=-2\quad g_{x_2x_2}=2\quad g_{x_3x_2}=0$$

$$g_{x_1x_3}=0\quad g_{x_2x_3}=0\quad g_{x_3x_3}=2$$

$$\bigtriangledown g(x_1,x_2,x_3)=0\Rightarrow\begin{pmatrix}2(x_1-x_2)\\-2(x_1-x_2)\\2x_3-6\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}$$

$\Rightarrow\bigtriangledown g(\bar{x})=(x_1,x_2,x_3)=(x_1,x_2,3)$ is stationary point $\forall x_1=x_2\in\mathbb{R}$.

$$\bigtriangledown^2 g(x_1,x_2,x_3)=\begin{pmatrix}2&-2&0\\-2&2&0\\0&0&2\end{pmatrix}$$

and hence $\bigtriangledown^2 g(\bar{x})=\bigtriangledown^2 g(x_1,x_2,3)=0\geq 0$,which is a PSD.
Therefore $\bar{x}=(x_1,x_2,x_3)=(x_1,x_2,3)$ is a global minimum $\forall x_1=x_2\in\mathbb{R}$ by convex optimality condition.

**Chapter 3**

# 3 Optimality Condition for Smooth Constrained Optimization Problems

## 3.1 Methods of Constrained Problem

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

$$min \quad h(y) \qquad (P_{uc})$$
$$s.t \quad y \in R^n$$
$$min \quad f_1(y) \qquad (P_c)$$
$$s.t \quad y \in S$$

where, $S = \{y|y \in Y, g_i(y) \leq 0, i = 1, ..., m, h_j(y) = 0, j = 1, ..., r\}$ is called the constraint sets, and $Y \subseteq \mathbb{R}^n$ is any set, $f, g_i, (i = 1, \ldots, m)$ and $h_j, j = 1, ..., r$ are real valued functions, all defined and continouosly differentiable on open set $Y \subseteq \mathbb{R}^n$.
In smooth optimization problems, all the functions involved were assumed to be differentiable or continuously differentiable unless failed if functions involved in objective functions or constraints are not differentiable. But the core of this project or thesis is the method of solving optimization problems of differentiable functions and constraints. The optimality conditions for non-linearly constrained problems are important because they form the basis for algorithms for solving such problems. Given a differentiable function $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $S \subseteq \mathbb{R}^n$, then a constrained non-linear optimization problem is given as

$$(P_c) \qquad min \quad f_1(y)$$
$$s.t \quad y \in S$$

For $(P_c)$ we use Lagrange function to convert or to transform the problems into an equivalent unconstrained problem or to a problem with easy constraint so that optimal solution can be characterized using the transformed form. To convert $(P)$ into unconstrained problem we try to find a functional $f_o$ on $\mathbb{R}^n$ as follows

$$f_o : \mathbb{R}^n \to \mathbb{R} \quad Differentiable \quad and$$

$$f_o(y) = 0 \quad \forall y \in S$$

Then the Lagrangian function L for $(P)$, such that

$$L : \mathbb{R}^n \to \mathbb{R} \quad Given \quad by$$

$$L(y) = f_1(y) + f_o(y)$$

Then we define the Lagrange problem as

$$(P_L) \qquad min \quad L(y) \quad Differentiable$$

$$s.t \qquad y \in \mathbb{R}^n$$

**Lemma (Lagrange Lemma):**
i. If $t_o$ is a global minimizer of $L$ and $t_o \in S$, then $t_o$ is a solution (or optimal solution) of $(P)$.
ii. If $t_o$ is a local minimizer of $L$ and $t_o \in S$, then $t_o$ is a solution (or optimal solution) of $(P)$.

*Proof.* Given $L : \mathbb{R}^n \to \mathbb{R}$ and $t_o$ is a (local or global) minimizer of $L$. This implies that $\bigtriangledown f_1(t_o) = 0$ i.e; for a neccessary condition $t_o$, $\bigtriangledown L(t_o) = 0$, $t_o$ is a minimizer of $L$ on $\mathbb{R}^n$.
If so then,$\bigtriangledown L(t_o) = 0$ and $t_o \in S$ are necessary conditions for $t_o$ to be a minimizer of $(P)$ $\square$

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

$$A(y) = \{i : g_i(y) = 0\}$$

.
**Definition3.1:** If the gradient $\bigtriangledown h_j(y), j = 1, 2, ..., k$ of equality constraints and the gradient $\bigtriangledown g_i(y), i \in A(y)$ of the active inequality constraints are linearly independent, then a vector $x \in S$ is called regular.

## 3.2 Lagrange Method for Optimality non-linear Equality Smooth Constraints

Suppose we have the following optimization problem with equality constraints.

$$Let f_1 : \mathbb{R}^n \to \mathbb{R}, \quad be \quad differentiable \quad and$$

$$h_j(y) : \mathbb{R}^n \to \mathbb{R}, \quad be \quad differentiable \quad for \quad each \quad j \in J = \{1, 2, ..., k\}$$
$$(P =) \qquad min \quad f_1(y)$$
$$s.t \qquad y \in S := \{y \in \mathbb{R}^n : h_j(y) = 0, j \in J\}$$

.
Now, if we take

$$f_o(y) = \sum_{j=1}^{k} \gamma_j h_j(y)$$

Where $\gamma_j \in \mathbb{R}$, then $f_o(y) = 0 \quad \forall y \in S$.
So, we define the Lagrange function for $(P =)$ as follows:

$$For \quad any \quad \gamma = (\gamma_1, \gamma_2, ..., \gamma_k)^T$$

called the Lagrange multiplier.
Let $L : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ , given by

$$L(y, \gamma) = f_1(y) + \sum_{j=1}^{k} \gamma_j h_j(y)$$

Thus, the corresponding Lagrange problem for $(P =)$

$$(L_\gamma) \qquad min \quad L(y, \gamma)$$
$$s.t \qquad y \in \mathbb{R}^n, \gamma \in \mathbb{R}^k$$

**The KKT Necessary Optimality Condition for Equality Constrained Smooth Optimization Problem**

As in the unconstrained case the first order conditions are not sufficient to guarantee a local minimum. For this, we turn to the second order sufficient condition (which, as in the unconstrained case, are not necessary). For equality constrained problems we are concerned with the behavior of the Hessian of the Lagrangian, denoted by $\bigtriangledown^2_{yy} L(y, \gamma)$ at a locations where the $KKT$ conditions holds:
Let $t^*$ be optimal solution of $(P =)$ with $f_1, h_j$ are differentiable. Assume that $t^*$ is regular, then for $\gamma^* = (\gamma_1^*, \gamma_2^*, ..., \gamma_k^*)$ be unique Lagrangian multipliers, the following called dual feasibilty and primal feasibilty respectively satisfies.
$\diamond$. $\bigtriangledown f_1(t^*) + \sum_{j=1}^{k} \gamma_j \bigtriangledown h_j(t^*) = 0$,
$\diamond$. $h_j(t^*) = 0, j = 1, 2, ..., k$
**Note:**
If $t^*$ together with some $\gamma^* \in \mathbb{R}^k$ satisfies the $KKT$ conditions, $t^*$ is called a $KKT$ point for $(P =)$ and the $KKT$ conditions are only necessary conditions. But under suitable convex assumption, the $KKT$ conditions are also sufficient optimality conditions.

**Theorem 17.** *Let $t^*$ be a $KKT$ point for $(P =)$ with a corresponding Lagrange multiplier $\gamma^*$ and $L(., \gamma^*)$ be a convex function, then $t^*$ is an optimal solution of $(P =)$.*

*Proof.* $L(., \gamma^*)$ is a convex and $\bigtriangledown L_t(t^*, \gamma^*) = 0 \Rightarrow t^*$ is optimal solution of $(L_{\gamma^*})$. Additionally, $t^* \in S$. Hence by the Lagrange lemma $t^*$ is optimal solution of $(P =)$. □

**Corollary:** Consider the problem

$$(P =) : \qquad min \quad f_1(y)$$

where $f_1$ is a convex function

$$s.t \qquad h_j(y^*) = 0, j = 1, 2, ..., k.$$

If $y^*$ is a $KKT$ point of $(P =)$ with some Lagrange multipliers $(\gamma_1^*, \gamma_2^*, ..., \gamma_k^*)^T$ and and $h_{\gamma^*}(y) = \gamma_1^* h_1(y) + \gamma_2^* h_2(y) + ... + \gamma_k^* h_k(y)$ is a convex function, then $y^*$ is an optimal solution.

**Necessary KKT Optimality Condition for Equality Constraints**

In this sub-section, constrained optimization

$$Constrained \quad Problem$$

$$min f_1(y)$$

(1)

$$s.t \qquad h_j(y^*) = 0, j = 1, 2, ..., k$$

We say that $y$ is feasible if it satisfies all the constraints of the problem, i. e; $h_j(y^*) = 0, j = 1, 2, ..., k.$ In fact it is linearly independent constraints.

**Necessary Optimality Conditions**

If $y^*$ is an optimal solution for the equality problem (1), then $y^*$ satisfies $\bigtriangledown f_1(y^*) = 0$,

If the function $f_1$ and all $h_j(y)$ are continuously differentiable over$\mathbb{R}^n$ and the optimality conditions are specified through the use of Lagrangian functions.

$$L(y, \gamma) = f_1(y) + \sum_{j=1}^{k} \gamma_j h_j(y)$$

$$Where \quad \gamma = \gamma_1, \gamma_2, ..., \gamma_k$$

From dual and primal feasibilty equation we have $n+k$ unknown variables ($x$ of size $n$ and $\gamma$ of size $k$) and $n+k$ equations. We can solve the system by distinguishing different cases and find the points that satisfies the equation and these points are called a$KKT$ point or saddle points.

**Remark:**

1. The $KKT$ condition can be satisfied at the solution of the problem (extreme) as well as at a stationary points.
2. To characterize the saddle points as a local or global minimum, we use the second order for the objective function $f_1(y)$ at the $KKT$ point for more information.

**Example:** Determine all the stationary points of the following constrained problem:

$$min \quad f_1(x, y, z) = 4x^2 + 2y^2 + z^2 - 4xy$$

$$s.t \quad x + y + z = 15$$

$$2x - 2y + 2z = 20$$

**Solution:** We have

$L(y, \gamma) = 4x^2 + 2y^2 + z^2 - 4xy + \gamma_1(x + y + z - 15) + \gamma_2(2x - 2y + 2z - 20) \rightarrow$ $min \quad L(y, \gamma) \in \mathbb{R}^2 \times \mathbb{R}^2$, which is Lagrange form with $y = (x, y, z)$ and $\gamma = (\gamma_1, \gamma_2)$. Then we have solve the following system of equations:

Consider the following unconstrained optimization problem

$$minL(y, \gamma) \qquad (P_\gamma)$$

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

To solve this problem, first check the regularity conditions, whether it is linearly independent or not, since the gradient active equality constraints, namely $\partial_y(x + y + z - 15), \partial_y(2x - 2y + 2z - 20)$

$$= \{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ 2 \end{pmatrix} \}, \quad where \quad y = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Let $\alpha, \beta \in R$

$$\alpha \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 2 \\ -2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \alpha = \beta = 0$$

Thus, $\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}2\\-2\\2\end{pmatrix}$ are linearly independent. The regularity condition is satisfied.
$i.\quad \frac{\partial L}{\partial x} = 8x - 4y + \gamma_1 + 2\gamma_2 = 0$
$ii.\quad \frac{\partial L}{\partial y} = 4y - 4x + \gamma_1 - 2\gamma_2 = 0$
$iii.\quad \frac{\partial L}{\partial z} = 2z + \gamma_1 + 2\gamma_2 = 0$
$iv.\quad \frac{\partial L}{\partial \gamma_1} = x + y + z - 15 = 0$
$v.\quad \frac{\partial L}{\partial \gamma_2} = 2x - 2y + 2z - 20 = 0$
Solving this system of equations by distinguishing different cases, we have the only solution $(x, y, z) = (\frac{7}{2}, \frac{5}{2}, 9)$ and $(\gamma_1, \gamma_2) = (-7, \frac{-11}{2})$.
So we have only one $KKT$ point, namely$(x, y, z) = (\frac{7}{2}, \frac{5}{2}, 9)$. But we still don't know, if this optimal or not.
To know this we can check the convexity of the objective function $f_1$ and the constraints $h_1$ and $h_2$.
• The objective function $f_1(x, y, z) = 4x^2 + 2y^2 + z^2 - 4xy$ is a convex function, since it's Hessian is positive semi definite.
• The constraints $h_1(x, y, z) = x + y + z - 15$ and $h_2(x, y, z) = 2x - 2y + 2z - 20$ are also a convex function,since they are affine function.
Thus the $KKT$ conditions are necessary for $(x, y, z) = (\frac{7}{2}, \frac{5}{2}, 9)$ to be a solution of the problem.
Hence, $y^* = (x, y, z) = (\frac{7}{2}, \frac{5}{2}, 9)$ is optimal solution of the original problem

## 3.3 Lagrange Method for optimality non-linear inequality smooth constraints

Let $g : \mathbb{R}^n \to \mathbb{R}$, and Let $h : \mathbb{R}^n \to \mathbb{R}^r$, be differentiable functions, where

$$h(y) = (h_1(y), h_2(y), ..., h_r(y))^T$$

Thus, the Lagrange function for inequality constraints $(p <)$ is defined as follows:
For any $\gamma = (\gamma_1, \gamma_2, ..., \gamma_r)^T \in \mathbb{R}^r_+, i.e; \gamma_i \geq 0 \quad \forall i$, define

$$L(y, \gamma) = g(y) + \sum_{i=1}^{r} \gamma_i h_i(y)$$

provided that $\gamma_i$ exists, where $L$ and each $\gamma_i$ are called Lagrange function and Lagrange multipliers, respectively.
Then we have new unconstrained optimization problems:

$$min \quad L(y, \gamma) \qquad (L_\gamma)$$

$$s.t \quad (y,\gamma) \in \mathbb{R}^n \times \mathbb{R}^r_+$$

**Theorem 18.** *Let $\langle(\gamma^*, h(y^*)\rangle = 0$ for some $\gamma^* \in \mathbb{R}^r_+$. If $(y^*, \gamma^*)$ is an optimal solution of $(L_\gamma)$, for some $\gamma^* \in \mathbb{R}^r_+$ and $y^* \in S$,then $y^*$ is an optimal solution of inequality constraints problems.*

*Proof.* Since we have, $\langle\gamma^*, h(y^*)\rangle = 0$, $y^*$ is an optimal solution of $(L_\gamma)$ and $y^* \in S$. Thus we want to show $y^*$ is an optimal solution of inequality constraint problem $(P_<)$ i.e; $g(y^*) \leq g(y) \forall y$.
$L(y^*, \gamma^*) \leq L(y, \gamma^*) \forall y \in \mathbb{R}^n$, since $y^*$ is an optimal solution of $(L_\gamma)$.

$$\Rightarrow g(y^*, \gamma^*) + \langle\gamma^*, h(y^*)\rangle \leq g(y, \gamma^*) + \langle\gamma^*, h(y^*)\rangle$$

$$\Rightarrow g(y^*, \gamma^*) \leq g(y, \gamma^*), since \langle\gamma^*, h(y^*)\rangle = 0 \forall y \in \mathbb{R}^n$$
$$\Rightarrow g(y^*) \leq g(y) \forall y \in \mathbb{R}^n$$

□

**Necessary Optimality Condition for inequality Constrained Problem**
Consider the following constrained problem:

$$min \quad g(y) \qquad (2)$$

$$s.t \quad h_i(y) \leq 0, i = 1, 2, ..., r$$

The function $g$ and all $h_i$ are continuously differentiable over $\mathbb{R}^n$. The optimality conditions are specified through the use of Lagrangian function.

$$L(y,\gamma) = g(y) + \sum_{i=1}^{r} \gamma_i h_i(y)$$

$$\gamma = (\gamma_1, \gamma_2, ..., \gamma_r)^T \in \mathbb{R}^r_+, i.e; \gamma_i \geq 0 \forall i$$

Assuming some regularity conditions for problem (2), for the $KKT$ conditions, if $y^*$ is an optimal solution of the problem, then there exists a Lagrange multiplier $\gamma^* = (\gamma_1^*, \gamma_2^*, ..., \gamma_k^*) \in \mathbb{R}^r_+, i.e; \gamma_i \geq 0$ $\forall i$, we have the following necessary conditions for $y^*$ to be optimal solution of inequality constraint problem $(P_<)$.
So, the $KKT$ Conditions for (P<) called dual feasibilty, complementary slackness and primal feasibility is given as follows, respectively.

$$i. \quad \bigtriangledown g(y^*) + \sum_{i=1}^{r} \gamma_i^* \bigtriangledown h_i(y^*) = 0$$

$$\gamma_i^* \geq 0, i = 1, 2, ..., r$$

$$ii. \quad \gamma_i^* h_i(y^*) = 0, i = 1, 2, ..., r$$

$$iii. \quad h_i(y^*) \leq 0, i = 1, 2, ..., r$$

**Sufficient Optimality Condition for Convex inequality Constrained Problem**
Consider the problem (2), the function $g$ and all $h_i$ are convex and continuously differentiable over $\mathbb{R}^n$. We say that the problem is convex, for this convex problem the $KKT$ condition are also sufficient for optimality. Assuming some additional regularity conditions for convex problem (2), there exists $\gamma^* = (\gamma_1^*, \gamma_2^*, ..., \gamma_k^*) \in \mathbb{R}_+^r, i.e; \gamma_i \geq 0 \, \forall i$ if and only if $y^*$ is an optimal solution of the inequality constrained problem,such that $\triangledown g(y^*) + \sum_{i=1}^r \gamma_i^* \triangledown h_i(y^*) = 0, \gamma_i^* h_i(y^*) = 0, i = 1, 2, ..., r, h_i(y^*) \leq 0, i = 1, 2, ..., r, \gamma_i^* \geq 0, i = 1, 2, ..., r$

**Example:**
Determine all the stationary points of the following constrained problem:

$$min \quad x_1^4 + x_2^4 + 12 + 12x_1^2 + 6x_2^2 - x_1 x_2 - x_1 + x_2$$

$$s.t \quad -x_1 - x_2 \leq -6$$

$$-2x_1 + x_2 \leq -3$$

$$-x_1 \leq 0, -x_2 \leq 0$$

**Solution:** Now, we have the Lagrange form
$L(y, \gamma) = x_1^4 + x_2^4 + 12 + 12x_1^2 + 6x_2^2 - x_1 x_2 - x_1 + x_2 + \gamma_1(-x_1 - x_2 + 6) + \gamma_2(-2x_1 + x_2 + 3) \rightarrow min \quad L(y, \gamma) \in \mathbb{R}^2 \times \mathbb{R}_+^2$
Then we have solve the following system of equations:
Consider the following unconstrained optimization problem

$$min \quad L(y, \gamma) \quad (P_\gamma)$$

$$(y, \gamma) \in \mathbb{R}^2 \times \mathbb{R}_+^2$$

To solve this problem, first check the regularity conditions, whether it is linearly independent or not, since the gradient active inequality constraints, namely
$\partial_y(-x_1 - x_2 + 6), \partial_y(-2x_1 + x_2 + 3)$

$$= \{ \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix} \}, \quad where \quad y = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Let $\alpha, \beta \in R$

$$\alpha\begin{pmatrix}-1\\1\end{pmatrix}+\beta\begin{pmatrix}-2\\1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix} \Rightarrow \alpha=\beta=0$$

Thus, $\begin{pmatrix}-1\\1\end{pmatrix}, \begin{pmatrix}-2\\1\end{pmatrix}$ are linearly independent. The regularity condition is satisfied.
From $KKT$ conditions, we get:
$i.\quad \frac{\partial L}{\partial x_1} = 4x_1^3 + 24x_1 - x_2 - 1 - \gamma_1 - 2\gamma_2 = 0$
$ii.\quad \frac{\partial L}{\partial x_2} = 4x_2^3 + 12x_2 - x_1 + 1 - \gamma_1 + \gamma_2 = 0$
$iii.\quad \gamma_1, \gamma_2 \geq 0$
$iv.\quad \gamma_1(-x_1 - x_2 + 6) = 0$
$v.\quad \gamma_2(-2x_1 + x_2 + 3) = 0$
$vi.\quad -x_1 - x_2 + 6 \leq 0$
$vii.\quad -2x_1 + x_2 + 3 \leq 0$
Consider 4 cases:
$i.\gamma_1, \gamma_2 > 0$
$ii.\gamma_1 > 0, \gamma_2 = 0$
$iii.\gamma_1 = 0, \gamma_2 > 0$
$iv.\gamma_1 = \gamma_2 0$
By distinguishing different cases, of each cases we have the only solution $(x_1, x_2) = (3, 3)$ and $(\gamma_1, \gamma_2) = (\frac{34}{3}, \frac{460}{3})$ at case (i) and the other cases yield not $KKT$ point. Hence $y^* = (x_1, x_2) = (3, 3)$ is the only $KKT$ of the problem.
Know we can check whether this point is an optimal solution or not by checking the convexity of the objective function $g$ and the constraints $h_1$ and $h_2$.
• The objective function $g(x_1, x_2) = x_1^4 + x_2^4 + 12 + 12x_1^2 + 6x_2^2 - x_1x_2 - x_1 + x_2$ is a convex function, since its Hessian is positive definite.
• The constraints $h_1(x_1, x_2) = -x_1 - x_2 + 6$ and $h_2(x_1, x_2) = -2x_1 + x_2 + 3$ are also a convex function because they are affine function.
Hence, by the assertion the $KKT$ point $y^* = (x_1, x_2) = (3, 3)$ is optimal solution of the original problem.

## 3.4 Lagrange Method for Optimality non-linear Mixed Smooth Constraints

Let $g(y) = (g_1(y), g_2(y), ..., g_r(y))^T$ and $h(y) = (h_1(y), h_2(y), ..., g_r(y))^T$
We consider, the problem

$$min \quad f_1(y) \qquad (P)$$

$$s.t \quad y \in S := \{y \in \mathbb{R}^n : g_i(y) \leq 0, h_j(y) = 0\}$$

Now, we define Lagrange function for $(P)$ as follows:
For any $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_r)^T \in \mathbb{R}^r_+, i.e., \gamma_i \geq 0, \forall i$, and $\delta = (\delta_1, \delta_2, \ldots, \delta_k)^T \in \mathbb{R}^k$ define

$$L : \mathbb{R}^n \times \mathbb{R}^r_+ \times \mathbb{R}^k \to \mathbb{R} \quad by$$

$$L(y, \gamma, \delta) = f_1(y) + \langle \gamma, g(y) \rangle + \langle \delta, h(y) \rangle$$

Thus, corresponding auxiliary problem is:

$$(L_{\gamma,\delta}) \quad min \quad L(y, \gamma, \delta)$$

$$s.t \quad y \in \mathbb{R}^n for \quad any \quad \gamma \in \mathbb{R}^r_+ \quad and \quad \delta \in \mathbb{R}^k$$

. Hence, we have the following necessary condition for $y^*$ to be optimal solution of $(P)$.

• Karush-Kuhn-Tucker (KKT) conditions (for mixed constraints):

• If $y^*$ is an optimal solution of $(P)$, then for some $\gamma_i^*, \delta_j^* \in \mathbb{R}$, the followings hold:

$i. \quad \bigtriangledown f_1(y^*) + \sum_{i=1}^r \gamma_i^* \bigtriangledown g_i(y^*) + \sum_{j=1}^k \delta_j^* \bigtriangledown h_j(y^*) = 0$

$\gamma_i^* \geq 0, i = 1, 2, ..., r$

$ii. \quad \gamma_i^* g_i(y^*) = 0, i = 1, 2, ..., r$

$iii. \quad g_i(y^*) \leq 0, i = 1, 2, ..., r$

$h_j(x^*) = 0, j = 1, 2, ..., k$

**Example:**

$$minimize \quad f_1(x, y) \quad = (x-1)^2 + y - 2$$

$$s.t \quad h(x, y) = y - x - 1 = 0$$

$$g(x, y) = x + y - 2 \leq 0$$

**Solution:** $\forall (x, y) \in \mathbb{R}^2$, we have

$\bigtriangledown h(x, y) = [-1, 1]^T, \bigtriangledown g(x, y) = [1, 1]^T \quad and \quad \bigtriangledown f_1(x, y) = [2(x-1), 1]^T$

List the KKT conditions

$i. \quad \bigtriangledown f_1(x, y) + \delta \bigtriangledown h(x, y) + \gamma \bigtriangledown g(x, y) = [2x - 2 - \delta + \gamma, 1 + \delta + \gamma]^T = 0$

$ii. \quad \gamma \geq 0$

$iii. \quad \gamma(x + y - 2) = 0$

$iv. \quad x + y - 2 \leq 0$

$v. \quad y - x - 1 = 0$

• First try: $\gamma > 0$, implies $x + y - 2 = 0$ and reduces the KKT conditions to

$$2x - 2 - \delta + \gamma = 0$$

$$1 + \delta + \gamma = 0$$

$$x + y - 2 = 0$$

$$y - x - 1 = 0$$

Which has the solution $[x, y, \gamma, \delta] = [\frac{1}{2}, \frac{3}{2}, 0, -1]$. But this contradicts $\gamma > 0$
• First try: $\gamma = 0$, reduces the KKT conditions to

$$2x - 2 - \delta = 0$$

$$1 + \delta = 0$$

$$y - x - 1 = 0$$

Which has the solution $[x^*, y^*, \gamma^*, \delta^*] = [\frac{1}{2}, \frac{3}{2}, 0, -1]$
It is easy that to check the regularity conditions for $g(x, y)$ and $h(x, y)$ are linearly independent.
The Objective function $f_1(x, y)$ is a convex function, since it is positive semi definite and both the constraints $g(x, y)$ and $h(x, y)$ are affine function which is convex function and also the second order sufficient condition is satisfied at $[x^*, y^*, \gamma^*, \delta^*] = [\frac{1}{2}, \frac{3}{2}, 0, -1]$
Therefore, $(x^*, y^*) = [\frac{1}{2}, \frac{3}{2}]$ is strict local minimizer of the problem.

## 3.5 A Sufficient and Necessary Condition for Non-convex Constrained Optimization

Consider the following inequality constrained optimization problem:

$$(P_o) \quad minf_o(x)$$

$$s.t \quad f_i(x) \leq \theta_i, i = 1, 2, ..., m$$

where$X \subset \mathbb{R}^n, \theta = [\theta_1, \theta_2, ..., \theta_m] \in \mathbb{R}^m, f_1 : X \to \mathbb{R}, i = 1, 2, ..., m$ are smooth but not necessarily convex functions. Let $g(x) = (f_1(x), f_2(x), ..., f_m(x))$ the family of perturbed problems associated with problem $P_o$ is defined by

$$(P_y) \quad minf_o(x)_{x \in X}$$

$$s.t \quad g(x) \leq y$$

where the vectors $y = (y_i, i = 1, 2, ..., m)$ is a perturbation to the original problem $P_o$. When $y = \theta$ the perturbed problem reduces to the original problem $P_o$. Let the perturbation function:

$$w : \mathbb{R}^m \to \mathbb{R}$$

be defined by

$$w(y) = min\{f_o(x) : g(x) \leq y, x \in X\},$$

which has an effective domain

$$dom(w) = \{y : \exists x \in X such \quad that \quad g(x) \leq y\}$$

Clearly the perturbation function $w$ is a monotone non-increasing function of y. Define the epigraph of $w(y)$ as the set:

$$epi(w) = \{[y, y_o] : y_o \geq w(y)\} \subset \mathbb{R}^{m+1}$$

We assume the following to hold:
Assumption 1. There exists an optimal solution $x^*$ to the problem $P_o$, and the optimal cost $f_o(x^*)$ is finite.
Assumption 2. (Without loss of generality) $\theta \in \mathbb{R}+^m_+$ and $f_o(x) \geq 0 \; \forall x \in X$
Assumption 3. There exists an $x \in X$ such that $f_i(x) < \theta_i, \forall_i = 1, 2, ..., m$ and $f(x) > f(x^*)$.
The problem is, if any one of $f_i, i = 0, 1, 2, ..., m$ is not convex, then the Kuhn-Tucker necessary condition cannot guarantee that the solution obtained is indeed optimal. On the other hand, for some non-convex problems the optimal solution may not satisfy the saddle point sufficient condition.
The Lagrangian theory can be geometrically interpereted as a supporting hyperplane to the set $epi(w)$. When the functions $f_i, i = 0, 1, 2, ..., m$ are convex, then the epigraph of $w$ is convex, and a supporting hyperplane exists at every point on the perturbation function $w(y)$ which is the lower bounded of $epi(w)$. When any one of $f_i, i = 0, 1, 2, ...., m$ is not convex, the $epi(w)$ is no longer convex and a supporting hyperplane may not exists at some point of the perturbation function which is why the saddle point theoremis only suffcient but not necessary. However,since the perturbation function is monotone non-increasing, every point on the perturbation function can be supported by a shifted cone obtained by shifting the negative or that in $\mathbb{R}^{m+1}$.This simple observation begs the question, is there a sufficient and necessary condition for the optimal solution of problem $P_o$ without any convexity assumption?
**An equivalent optimality condition**
We now present a new optimality condition in terms of an unconstrained problem which is both sufficient and necessary.It has the flavor of a supporting cone for the non-convex perturbation function. Alternatively, one can also think of scalarizing the cost and constraint functions using a weighted Tchebyshev norm.

**Theorem 19.** *(A sufficient and necessary condition for optimality without convexity) Let $x^*$ be the (global) optimal solution of problem $P_o$, and let $\theta_0 = f_0(x^*)$. Then a* *solution $x^*$ solves problem $P_0$ if and only if $x^*$ solves the unconstrained problem:*

$$(P_1) \quad min_{x\in X} max_{0\leq i\leq m}\{\frac{f_i(x)}{\theta_i}\}$$

.

*Proof.* It is clear that the minimum value of the unconstrained optimization problem $P_1$ is:

$$min_{x\in X} max_{0\leq i\leq m}\{\frac{f_i(x)}{\theta_i}\} = 1$$

If $x_0$ solves problem $P_0$, then $x_0$ must be feasible , and $f_o(x_o) = f_o(x^*)$. Hence

$$max_{0\leq i\leq m}\{\frac{f_i(x)}{\theta_i}\} = 1$$

i,e; $x_0$ solves problem $P_1$.
Conversly, if $x_0$ does not solve $P_0$, then either $x_0$ is infeasible, or $x_0$ is feasible and $f_o(x_o) > f_o(x^*)$.
If $x_0$ is infeasible , then $\exists j \in \{1, 2, ..., m\}$ such that $f_j(x_o) > \theta_j$, implying that

$$max_{0\leq i\leq m}\{\frac{f_i(x)}{\theta_i}\} > 1 \qquad (*)$$

If $x_0$ is feasible, $f_o(x_o) > f_o(x^*)$ and the inequality (*) still holds. Thus in both cases , $x_0$ does not solve $P_1$. □

The above theorem leads to a conceptual simple method for solving the non-convex inequality constrained optimization problem $P_0$. Consider the following scalar function of a scalar parameter $\theta_o \in \mathbb{R}_+ \geq 0$. Note that $\theta_i, i = 1, 2, ..., m$ are fixed parameters of the problem $P_o$:

$$\phi(\theta_i) = min_{x\in X} max_{0\leq i\leq m}\{\frac{f_i(x)}{\theta_i}\}$$

**Theorem 20.** ***(Properties of the function $\phi$)***
*i.* $\theta_o < f_o(x^*) \Longrightarrow \phi(\theta_o) > 1$;
*ii.* $\theta_o > f_o(x^*) \Longrightarrow \phi(\theta_o) < 1$;
*iii.* $\theta_o = f_o(x^*) \Longleftrightarrow \phi(\theta_o) = 1$;
*iv.* $\phi(\theta_o)$ *is a monotone non-increasing function of*$\theta_i$;
*v.* $\phi(\theta_o)$ *is continuous function of* $\theta_i$.

*Proof.* i. If $x$ is feasible,i.e; $f_i(x) \leq \theta_i, \forall i$, then $f_o(x) \geq f_o(x^*)$, hence $\frac{f_o(x)}{\theta_o} > 1$, and

$$max_{0 \leq i \leq m}\{\frac{f_i(x)}{\theta_i}\} = \frac{f_o(x)}{\theta_o} > 1.$$

Else if $x$ is not feasible, i.e; $f_j(x) > \theta_j$ for some j, then

$$max_{0 \leq i \leq m}\{\frac{f_i(x)}{\theta_i}\} \geq \frac{f_j(x)}{\theta_j} > 1.$$

Thus

$$min_{x \in X} max_{0 \leq i \leq m}\{\frac{f_i(x)}{\theta_i}\} > 1.$$

ii. By Assumption 3, $if \theta_o > f_o(x^*)$, then there exists $x \in X$ such that $f_i(x) < \theta_i, i = 1, 2, ..., m$; and $\theta_o > f_o(x) > f(^x*)$ consequently

$$max_{0 \leq i \leq m}\{\frac{f_i(x)}{\theta_i}\} < 1$$

and hence

$$\phi(theta_o) = min_{x \in X} max_{0 \leq i \leq m}\{\frac{f_i(x)}{\theta_i}\} < 1$$

iii. Follows directly from Theorem 3.4.
iv. Let$\theta_o^1 > \theta_o^2$. Then

$$\frac{f_o(x)}{\theta_o^1} < \frac{f_o(x)}{\theta_o^2} \forall x \in X.$$

This implies that max $\{\frac{f_o(x)}{\theta_o^1}, \frac{f_i(x)}{\theta_i}, i = 1, 2, ..., m\} \leq max\{\frac{f_o(x)}{\theta_o^2}, \frac{f_i(x)}{\theta_i}, i = 1, 2, ..., m\}$, annd hence

$$\phi(\theta_o^1) \leq \phi(\theta_o^2).$$

v. This follows that from a result which says that if the cost function is continuous in some parameter ($\theta_o$ in this case),then the extreme value of the cost function is also continuous in the parameter. □

Because of the special property of the function $\phi$, solving the constrained optimization problem $P_o$ is now equivalent to a rather simple problem. Find the unique root of a monotone decreasing (scalar) function of a scalar variable:
$P_2$ Find $\theta_o^* = f_0(x^*)$ such that $\phi(\theta_o^*) = 1$. At first look, this looks like a trivial problems since it is almost effortless to numerically solve for the root of a monotone (strictly monotone at the root by virtue of Theorem 20 (iii)) scalar function of a scalar variable. However, in practice,this is made non-trivial by the fact that each function call of $\phi$ requires the solution of a unconstrained minimax problem, which in it self is not a trivial problem even if it is unconstrained. Nevertheless there

exist several effective methods and soft ware packages that deal with minimax optimization effectively or the mini max function in the optimization toolbox of MATLAB and these will be fully exploited. We can use any gradient-free method such as the bisection search or secant method which is sufficient. In fact, for man of our test problems, it is found that the function $\phi$ is nearly linear over a wide range about the optimal solution. This, coupled with the fact that the solution is unique, makes even unsophisticated methods like the secant method converge merely in a few iterations. Due to page constraint, we are unable to provide any example here. This will be taken up in a future if anyone interested to do as a recommendation, since in this paper it is focused only on the convex constrained functions. As a conclusion it is interesting to note that the proposed approach does not require the set X to be infinite, and hence, X can be a discrete set if necessary. In this case, the traditional Lagrangian theory is never sufficient and necessary even if all the underlying functions are convex. In practice, however, some difficulty would be expected in solving discrete mini max problems.

*Bibliography*

[1] Belay Bekele, Optimality conditions for non-smooth optimization and Mordukhovich sub-differentials, Addis Ababa university, 2012.
[2] Tadesse Birkie, Optimality conditions for non-smooth optimization, Addis Ababa university, 2012.
[3] R.Clark Robinson, Introduction to mathematical optimization, Northwestern University, copyright(c) 2013.
[4] Roberto Andreani, Gabriel Haeser, J.M.Marinez, on sequential optimality conditions for smooth constrained optimization, August 6, 2009.
[5] Joshua Wilde, revised by Isabel Tecu, Takeshi suzuki and Maria Jose Bocardi unconstrained optimization,August 13, 2013.
[6] C. J. GOH AND X,Q.YANG. A sufficient and necessary conditions for nonconvex constrained optimization, University of Western Australia, April, 1997.
[7] Jorge Nocedal StephenJ,Wright, Numerical Optimization, Second Edition, Northwestern University and University of Wisconsin,2006.
[8] Abadie.J 1967. on the Kuhn-Tucker Theorem"Non linear Programming pp.19-36" North Holland, Amasterdam.
[9] Ding Zhudu Optimization Theory and Methods non-linear Programming.
[10] Georgia Tech ECE 8823c notes by M.A Davenport and J.Romberg, updated, February 2019, Unconsrained Optimization Problem.
[11] Robert M. Freund, Optimality Conditions for Constrained Optimization Problems, February,2004, Massachusetts Institute of Technology
[12] Bazara M.S,Sherall H.D, Shetty C.M: Nonlinear programming Theory and Algorithms, ($2^{nd} edition$) 1993