"CONVEX FUNCTIONS AND OPTIMIZATION TECHINIQUES"

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS
FOR THE DEGREE OF
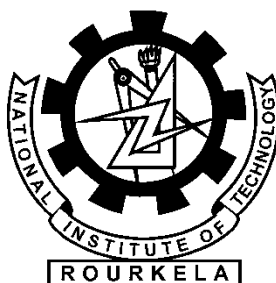
MASTER OF SCIENCE
IN
MATHEMATICS

SUBMITTED TO

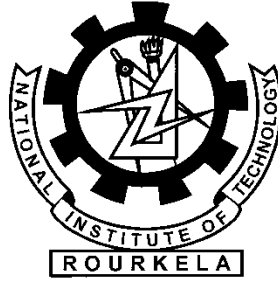NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA

BY
ARUN KUMAR GUPTA
ROLL NO. 409MA2066

UNDER THE SUPERVISION OF
Dr. ANIL KUMAR

DEPARTMENT OF MATHEMATICS
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA-769008

NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA

DECLARATION

I hereby certify that the work which is being presented in the thesis entitled "**Convex function and optimization techniques**" in partial fulfillment of the requirement for the award of the degree of Master of Science, submitted in the Department of Mathematics, National Institute of Technology, Rourkela is an authentic record of my own work carried out under the supervision of Dr. Anil Kumar. The matter embodied in this thesis has not been submitted by me for the award of any other degree.

Date: (Arun Kumar Gupta)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dr. Anil Kumar
Department of Mathematics
National Institute of Technology
Rourkela – 769 008
Orissa, India.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTER -1

Introduction:

Optimization is the process of maximizing or minimizing a desired objective function while satisfying the prevailing constraints. The optimization problems have two major divisions. One is linear programming problem and other is non-linear programming problem. But the modern game theory, dynamic programming problem, integer programming problem also part of the Optimization theory having wide range of application in modern science, economics and management. In the present work I tried to compare the solution of Mathematical programming problem by Graphical solution method and others as well as its theoretic descriptions. As we know that not like linear programming problem where multidimensional problems have a great deal of applications, non-linear programming problem mostly considered only in two variables. Therefore for nonlinear programming problems we have an opportunity to plot the graph in two dimensions and get a concrete graph of the solution space which will be a step ahead in its solutions.

Nonlinear programming deals with the problem of optimizing an objective function in the presence of equality and inequality constraints. The development of highly efficient and robust algorithms and software for linear programming, the advent of high speed computers, and the education of managers and practitioners in regard to the advantages and profitability of mathematical modeling and analysis, have made linear programming an important tool for solving problems in diverse fields. However, many realistic problems cannot be adequately represented or approximated as a linear program owing to the nature of the nonlinearity of the objective function or the nonlinearity of any of the constraints.

CHAPTER -2

Convex Functions and Generalization:

Convex and concave functions have many special and important properties. In this chapter, I introduce the important topics of convex and concave functions and develop some of their properties. For example, any local minimum of a convex function over a convex set is also a global minimum. These properties can be utilized in developing suitable optimality conditions and computational schemes for optimality problems that involve convex and concave functions.

2.1   Definitions and Basic properties:

Convex Function: A function $f: X \rightarrow R$ defined on a convex subset X of R^n is said to be **convex** on X if

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

for each $x, y \in X$ and $\lambda \in [0,1]$. The function $f: X \rightarrow R$ is called **strictly convex** on X if the above inequality is true as a strict inequality for each $x, y \in X$ and $\lambda \in [0,1]$.

Concave function: A real-valued function $f: X \rightarrow R$ defined on a convex subset X of R^n is said to be **concave** on X if

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$$

for each $x, y \in X$ and $\lambda \in [0,1]$. The function $f: X \rightarrow R$ is called **strictly concave** on X if the above inequality is true as a strict inequality for each $x, y \in X$ and $\lambda \in [0,1]$. Also, $f(x)$ is concave on $[a, b]$ if and only if the function $-f(x)$ is convex on $[a, b]$.

Now let us consider the geometric interpretation of convex and concave functions. Let x and y be two distinct points in the domain of f, and consider the point $\lambda x + (1 - \lambda)y$ with $\lambda \in [0,1]$. So for a convex function f, the value of f at points on the line segment $\lambda x + (1 - \lambda)y$, is less than or equal to the height of the chord joining the points $[x, f(x)]$ and $[y, f(y)]$. For the concave function, the chord is below the function itself. Hence, a function is both convex and concave if and only if it is affine. Figure 2.1 shows some examples of convex and concave functions.

Convex function

concave function

neither convex nor concave
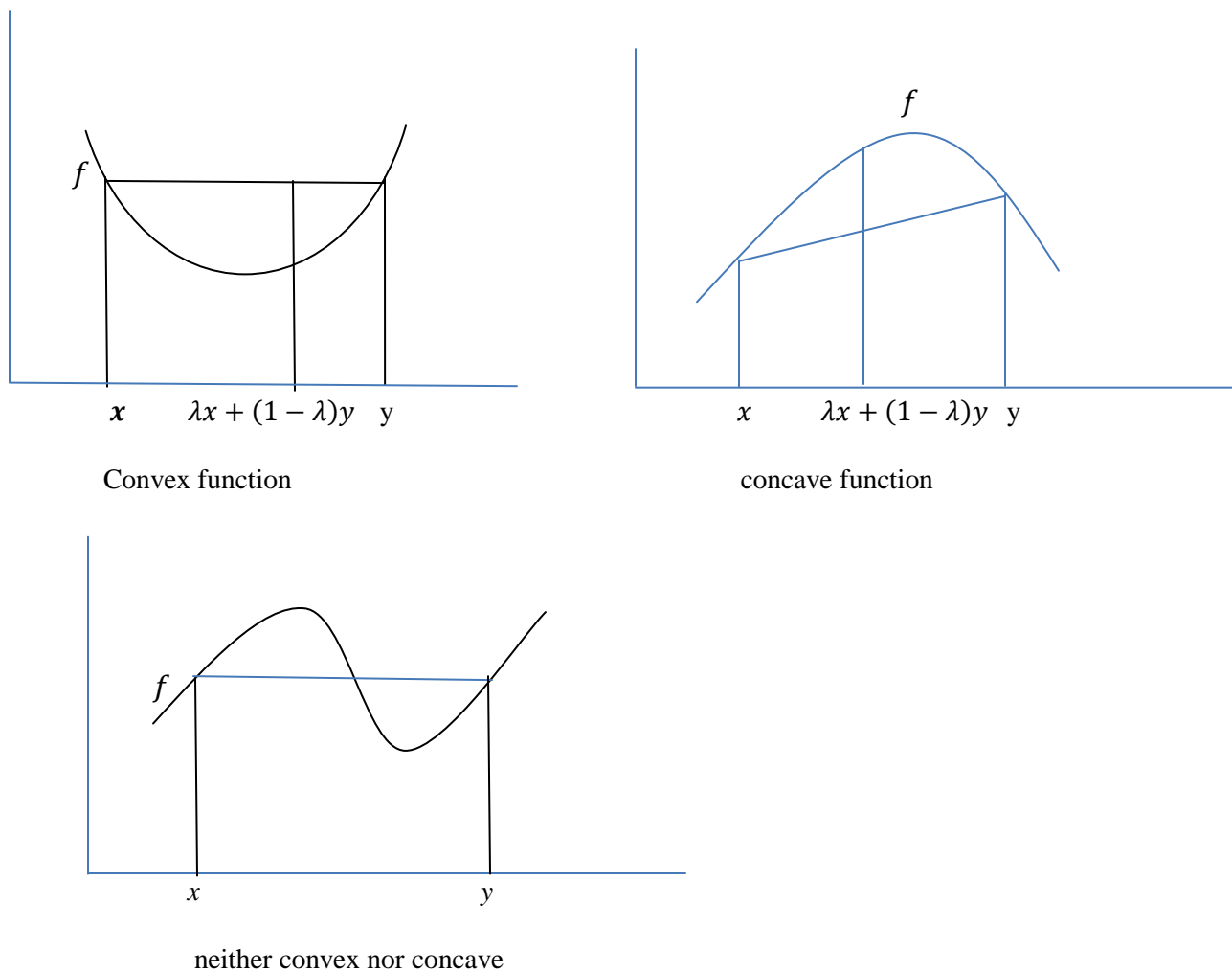
Figure 2.1 examples of convex and concave functions

The following are some examples of convex and concave functions.

Examples of convex functions:

1. Any function of the form: $f(x) = ax + b;\ a, b \in R$
2. Powers: $f(x) = x^p; p \geq 1$
3. Powers of absolute value: $f(x) = |x|^p$ on R, for p ≥ 1
4. Exponential: $f(x) = e^{ax}$, for any $a \in R$
5. Every linear transformation taking values in R is convex.
6. $f(x_1, x_2) = 2x_1^2 + x_2^2 - 2x_1x_2$
7. Every norm is a convex function, by the triangle inequality and positive homogeneity.
8. $f(x) = -sinx, on\ [0, \pi]$

Examples of Concave functions:

1. $f(x) = -x^2$
2. $f(x) = \sqrt{x}$
3. Any linear function $f(x) = ax + b; a, b \in R$
4. The function $f(x) = \sin(x)$ is concave on the interval $[0, \pi]$
5. Logarithm: $\log x$ on R_+

We give below some particularly important instances of convex functions that arise very often in practice.

1. Let $f_1, , f_2, \ldots, f_n: R^n \to R$ be convex functions. Then,
   (a) $f(x) = \sum_{j=1}^{k} c_j f_j(x)$, where $c_j > 0$ for $j = 1,2,\ldots,k$ is a convex functions.
   (b) $f(x) =$ Maximum $\{f_1(x), f_2(x), \ldots, f_k(x)\}$ is a convex function.
2. Suppose than $g: R^n \to R$ is a concave function. Let $S = \{x: g(x) > 0\}$, and define $f: S \to R^n$ as $f(x) = \frac{1}{g(x)}$. Then, f is convex over S.
3. Let $g: R \to R$ be a nondecreasing, univariate, convex function, and let $h: R^n \to R$ be a convex function. Then, the composite function $f: R^n \to R$ defined as $f(x) = g[h(x)]$ is a convex function.
4. Let $g: R^m \to R$ be a convex function, and let $h: R^n \to R^m$ be an affine function of the form $h(x) = Ax + b$, where A is a $m \times n$ matrix and b is a $m \times 1$ vector. Then, the composite function $f: R^n \to R$ defined as $f(x) = g[h(x)]$ is a convex function.

2.2 Epigraph and Hypograph of a function:

A function f on S can be described by the set $\{[x, f(x)]: x \in S\} \subset R^n$, which is referred to as the graph of the function. We can construct two sets that are related to the graph of f the epigraph, which consists of points above the graph of f, and the hypograph, which consists of points below the graph of f. These notions are discussed in the definition 2.2.1 below.

2.2.1 Definition:

Let S be a nonempty set in R^n, and let $f: S \to R$. The epigraph of f, denoted by $epi\ f$, is a subset of R^{n+1} defined by

$$epi\ f = \{(x, y): x \in S, y \in R, y \geq f(x)\} \subseteq R^{n+1}$$

And the strict epigraph of the function is denoted by $epi_s\ f$

$$epi_s\ f = \{(x, y): x \in R^n, y \in R, y > f(x)\} \subseteq R^{n+1}$$

The hypograph of a function $f: S \rightarrow R$ is denoted by $hyp\ f$, is a subset of R^{n+1} defined by

$$hyp\ f = \{(x, y): x \in R^n, y \in R, y \leq f(x)\} \subseteq R^{n+1}$$

And the strict hypograph of the function is defined by

$$hyp_s\ f = \{(x, y): x \in R^n, y \in R, y < f(x)\} \subseteq R^{n+1}$$

A function is convex if and only if its epigraph is a convex set and, equivalently, that a function is concave if and only if its hypograph is a convex set.
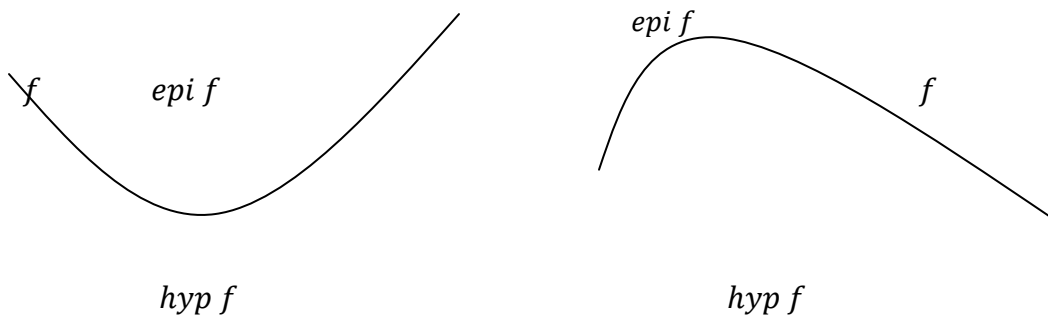
2.3   Differentiable Convex functions:

Now we focus on differentiable convex and concave functions. A function $f: S \rightarrow R$ defined on an open convex set S of R^n is said to be differentiable if the gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \frac{\partial f(x)}{\partial x_3}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$$

exists at each $x \in S$.

First order differentiable condition:

Let S be a nonempty open convex set in R^n, and let $f: S \rightarrow R$ be differentiable on S. f is convex if and only if $f(y) \geq f(x) + \nabla f(x)^T (y - x)$ for all $x, y \in S$.

Second order differentiable condition:

Let S be a nonempty set in R^n, and let $f: S \to R$. Then, f is said to be twice differentiable if there exist a vector $\nabla f(x)$, and a $n \times n$ symmetric matrix $H(x)$, called the Hessian matrix, such that

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T H(x)(y - x) \text{ for each } y \in S.$$

For twice differentiable function the Hessian matrix $H(x)$ is given comprised of the second order partial derivatives and is given by

$$H(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \equiv \begin{bmatrix} f_{11}(x) & f_{12}(x) & \cdots & f_{1n}(x) \\ f_{21}(x) & f_{22}(x) & \cdots & f_{2n}(x) \\ \cdots & \cdots & \cdots & \cdots \\ f_{n1}(x) & f_{n2}(x) & \cdots & f_{nn}(x) \end{bmatrix}$$

Examples 2.3.1

Let $f(x_1, x_2) = 2x_1 + 6x_2 - 2x_1^2 - 3x_2^2 + 4x_1 x_2$. We have

$$\nabla f(\bar{x}) = \begin{bmatrix} 2 - 4\bar{x}_1 + 4\bar{x}_2 \\ 6 - 6\bar{x}_2 + 4\bar{x}_1 \end{bmatrix} \text{ and } H(\bar{x}) = \begin{bmatrix} -4 & 4 \\ 4 & -6 \end{bmatrix}$$

Taking $\bar{x} = (0, 0)^T$, the second-order expansion of this function is given by

$$f(x_1, x_2) = (2, 6)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \frac{1}{2}(x_1, x_2)\begin{bmatrix} -4 & 4 \\ 4 & -6 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Since the given function is quadratic, no remainder term is present there, and so the above representation is exact.

$$H(\bar{x}) = \begin{bmatrix} -4 & 4 \\ 4 & -6 \end{bmatrix}$$

and so

$$-H(\bar{x}) = \begin{bmatrix} 4 & -4 \\ -4 & 6 \end{bmatrix}$$

We conclude that $-H(\bar{x})$ is positive definite and so $H(\bar{x})$ is negative definite and the function f is strictly concave.

Examples 2.3.2

Let $f(x_1, x_2) = e^{2x_1 + 3x_2}$. We have

$$\nabla f(\bar{x}) = \begin{bmatrix} 2e^{2\bar{x}_1 + 3\bar{x}_2} \\ 3e^{2\bar{x}_1 + 3\bar{x}_2} \end{bmatrix} \text{ and } H(\bar{x}) = \begin{bmatrix} 4e^{2\bar{x}_1 + 3\bar{x}_2} & 6e^{2\bar{x}_1 + 3\bar{x}_2} \\ 6e^{2\bar{x}_1 + 3\bar{x}_2} & 9e^{2\bar{x}_1 + 3\bar{x}_2} \end{bmatrix}$$

Hence, the second-order expansion of this function about the point $\bar{x} = (2, 1)^T$ is given by

$$f(\bar{x}) = e^7 + (2e^7, 3e^7)\begin{pmatrix} x_1 - 2 \\ x_2 - 1 \end{pmatrix} + \frac{1}{2}(x_1 - 2, x_2 - 1)\begin{bmatrix} 4e^7 & 6e^7 \\ 6e^7 & 9e^7 \end{bmatrix}\begin{pmatrix} x_1 - 2 \\ x_2 - 1 \end{pmatrix}$$

2.3.3 Theorem:

Let S be a nonempty set in R^n, and let $f: S \to R$ be twice differentiable on S then, f is convex on S if and only if the Hessian matrix is positive semi definite (PSD) at each point in S; that is, for any $\bar{x}$ in S, we have $x^t H(\bar{x})x \geq 0$ for all $x \in R^n$. Symmetrically, a function f is concave on S if and only if its Hessian matrix is negative semi definite (NSD) everywhere in S, that is, for any $\bar{x} \in S$, we have $x^t H(\bar{x})x \leq 0$ for all $x \in R^n$.

2.4 Generalizations of a convex function:

In this section, we present various types of functions that are similar to convex and concave functions but that share only some of their desirable properties. We'll learn that many of the results do not require the restrictive assumptions of convexity, but rather the less restrictive assumptions of quasi-convexity, pseudo-convexity.

2.4.1 Quasi-convex functions:

A function $f: S \to R$ defined on a convex subset S of a real vector space is quasi-convex if for any $x, y \in S$ and $\lambda \in [0,1]$

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

furthermore if

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

for any $x \neq y$ and $\lambda \in [0,1]$, then f is strictly quasi-convex.

The function f is said to be quasiconcave if $-f$ is quasi-convex and a strictly quasi-concave functions a function whose negative is strictly quasi-convex. Equivalently a function f is quasi-concave if

$$f(\lambda x + (1 - \lambda)y) \geq min\{f(x), f(y)\}$$

and strictly quasi-concave if

$$f(\lambda x + (1 - \lambda)y) > min\,\{f(x), f(y)\}$$

Examples of Quasi-convex functions:

- $\sqrt{|x|}$ is quasi-convex on R.
- $logx$ is quasi-linear on R_+.

2.4.2　Differentiable Quasi-convex functions:

Let S be a nonempty open convex set in R^n, and let $f: S \to R$ be differentiable on S. Then, f is quasiconvex if and only if either one of the following equivalent statements holds:

- If $x, y \in S$ and $f(x) \le f(y)$, then $\nabla f(y)^T(x - y) \le 0$.
- If $x, y \in S$ and $\nabla f(y)^T(x - y) > 0$, then $f(x) > f(y)$.

To illustrate the above theorem, let $f(x) = x^3$.
To check quasi-convexity, suppose $f(x_1) \le f(x_2) \Rightarrow x_1^3 \le x_2^3$. This is true only if $x_1 \le x_2$.

Now consider $\nabla f(x_2)(x_1 - x_2) = 3(x_1 - x_2)x_2^2$. Since $x_1 \le x_2$, $3(x_1 - x_2)x_2^2 \le 0$ therefore, $f(x_1) \le f(x_2) \Rightarrow \nabla f(x_2)(x_1 - x_2) \le 0$ and hence by the above theorem, f is quasi-convex.

Consider let $f(x_1, x_2) = x_1^3 + x_2^3$. Let $x_1 = (2, -2)^t$ and $x_2 = (1, 0)^t$. $f(x_1) = 0$ and $f(x_2) = 1$, so that $f(x_1) < f(x_2)$.

But, $\nabla f(x_2)^t(x_1 - x_2) = (3, 0)(1, -2)^t = 3$. By the necessary part of the theorem, f is not quasiconvex. This shows that the sum of two quasi-convex functions is not necessarily quasi-convex.

Propositions:

- A function which is both quasi-convex and quasi-concave is called quasi-linear.
- Every convex function is quasi-convex.
- A concave function can be quasi-convex function. For example log(x) is concave, and it is quasi-convex.
- If $f(x)$ and $g(y)$ are positive convex decreasing functions, then, $f(x)g(y)$ is quasi-convex.
- Sum of two Quasi-convex functions is not necessarily a Quasi-convex.

2.4.3   Strongly Quasi-convex function:

The concept of strong convexity extends the notion of strict convexity. Let S be a nonempty open convex set in R^n, and let $f: S \rightarrow R$. The function f is said to be strongly quasiconvex if for each $x, y \in S$, we have

$$f(\lambda x + (1 - \lambda)y) < max \{f(x), f(y)\} \text{ for each } \lambda \in (0,1), \text{ with } x \neq y.$$

The function f is said to be strongly quasi-concave if $-f$ is strongly quasi-convex. From the definition of convex function, quasi-convex function and strictly quasi-convex function, the following statements hold:

- Every strictly convex function is strongly Quasi-convex.
- Every strongly quasi-convex function is strictly quasi-convex.
- Every strongly quasi-convex function is quasi-convex.

2.4.4   Pseudo-convex functions:

The differentiable strongly (or strictly) quasi-convex functions do not share the property of convex functions, which states that if $\nabla f(\bar{x}) = 0$ at some $\bar{x}$, then $\bar{x}$ is a global minimum of f. This motivates the definition of pseudo-convex functions which share this important property with convex functions, and leads to a generalization of various derivative based optimality conditions.

2.4.5   Definition

Let S be a nonempty open convex set in R^n, and let $f: S \rightarrow R$ be differentiable on S. The function f is said to be pseudo-convex if for each $x, y \in S$ with $\nabla f(y)^T (y - x) \geq 0$ we have $f(y) \geq f(x)$; or, equivalently, if $f(y) < f(x)$, then $\nabla f(x)^T (y - x) < 0$.

The function f is said to be pseudo-concave if $-f$ is pseudo-convex.

The function f is said to be strictly pseudo-convex if, for each distinct $x, y \in S$ satisfying $\nabla f(x)^T (y - x) \geq 0$, we have $f(y) \geq f(x)$; or equivalently, if for each distinct $x, y \in S, f(y) \leq f(x)$ implies that $\nabla f(x)^T (y - x) < 0$. The function f is said to be strictly pseudo-concave if $-f$ is strictly pseudoconvex.

2.5 Relationship among various types of convex functions:

```
┌─────────────────────┐                                    ┌─────────────────────┐
│   Strictly convex    │──────────────────────────────────▶│      Convex         │
└─────────────────────┘                                    └─────────────────────┘
           │                                                          │
   Under   differentiability                              Under   differentiability
           ▼                                                          ▼
┌─────────────────────┐                                    ┌─────────────────────┐
│ Strictly pseudo-convex│─────────────────┐                │   Pseudo-convex     │
└─────────────────────┘                   │                └─────────────────────┘
           │                              ▼                           │
           ▼                                                          ▼
┌─────────────────────┐                              ┌─────────────────────┐
│ Strongly quasi-convex│────────────┐                │ Strictly quasi-convex│
└─────────────────────┘             ▼                └─────────────────────┘
                                                          Under lower semi continuity
                                                                     ▼
                              ┌─────────────────────┐
                              │    Quasi-convex      │
                              └─────────────────────┘
```

Chapter 3

Constrained Minimization:

3.1   Introduction and problem formulation:

The majority of engineering problems involve constrained minimization- that is, the task is to minimize a function subject to constraints. Constrained problem may be expressed in the following general nonlinear programming form:

$$\text{Minimize} \quad f(x)$$

$$\text{subject to} \quad g_i(x) \leq 0 \qquad i = 1,2,\dots,m \qquad\qquad (3.1.1)$$

$$\text{and} \qquad h_j(x) = 0 \qquad j = 1,2,\dots,l$$

Where $x = (x_1, x_2, \dots, x_n)^T$ is a column vector of n real-valued design variables. In equation (3.1.1), $f: R^n \to R$, is the objective or cost function, $g_i: R^n \to R$ are inequality constraints and $h_j: R^n \to R$, are equality constraints. A vector $x \in R^n$ satisfying all the constraints is called a feasible solution to the problem. The collection of all such solutions forms the feasible region. The nonlinear programming problem, then, is to find a feasible point x' such that $f(x) \geq f(x')$ for each feasible point x. Such a point x' is called an optimal solution. We define the feasible set $\boldsymbol{\Omega}$ as

$$\boldsymbol{\Omega} = \{x \in R^n : h_j(x) = 0, g_i(x) \leq 0\}$$

The phrase minimization problem refers to finding a local minimizer in the feasible set $\boldsymbol{\Omega}$.

3.2   Graphical method:

The graphical method for solving mathematical programming problem is based on a well-defined set of logical steps. Using graphical method, the given programming problem can be easily solved with a minimum amount of computational effort. We know that the simplex method is the well-studied and widely useful method for solving linear programming problem, while for the class of nonlinear programming no such widely useful method exists. Programming problems involving only two variables can easily solved graphically. The algorithms or the systematic procedure is discussed as follows:

Algorithm: The solution nonlinear programming problem by graphical method, in general, involves the following steps:

Step-1: Construct the graph of the given nonlinear programming problem.

Step-2: Identify the convex region (solution space) generated by the objective function and constraints of the given problem.

Step-3: Determine the point in the convex region at which the objective function is optimum (maximum or minimum).

Step-4: Interpret the optimum solution so obtained.

3.2.1   Solution of various kinds of problems by graphical method:

Case -1: Problems with linear objective function and nonlinear constraints:

Consider the problem

$$\text{Maximize} \quad Z = 2x_1 + 3x_2$$

$$\text{Subject to} \quad x_1^2 + x_2^2 \leq 20 \qquad\qquad (3.2.1)$$

$$x_1 x_2 \leq 8$$

$$x_1 \geq 0 , x_2 \geq 0$$

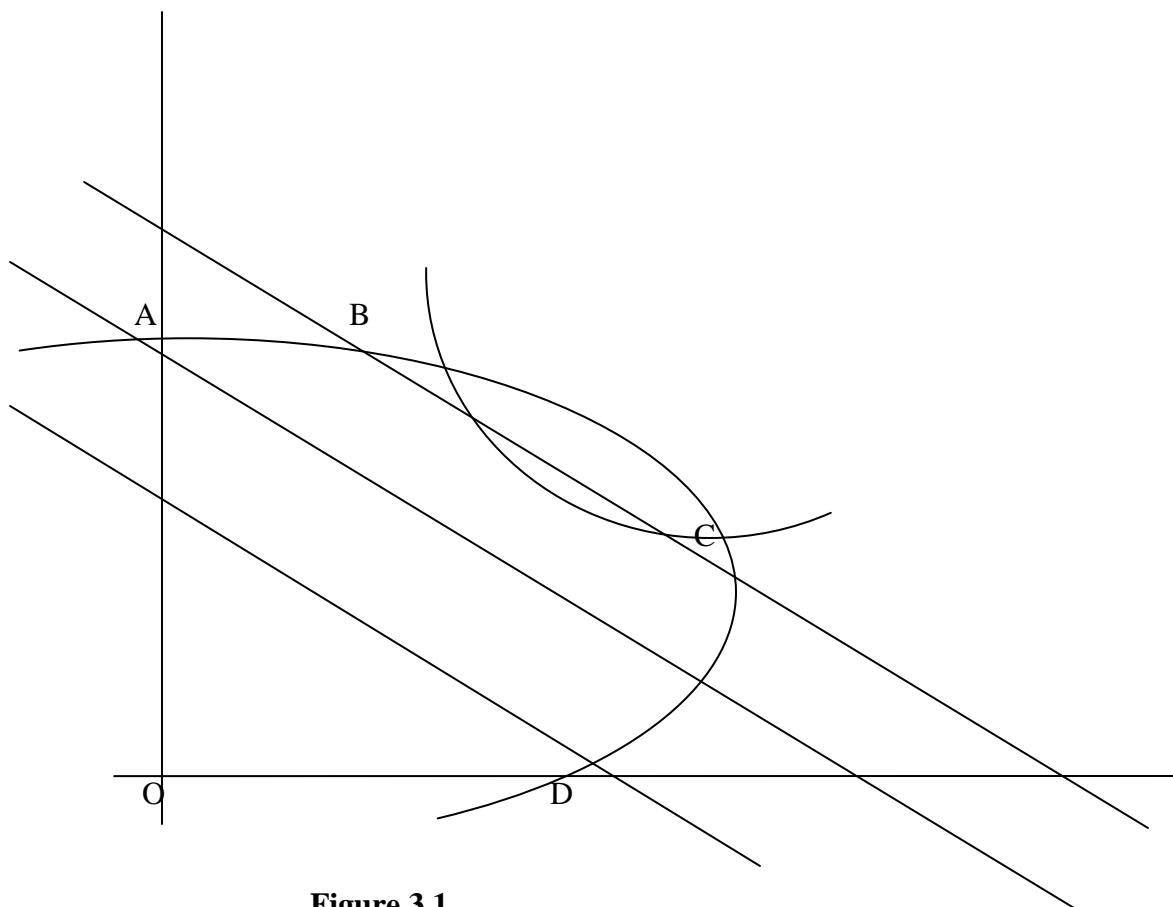We'll solve this problem by graphical method.

Figure 3.1

For this, we have to trace the graph of the constraints of the problem considering inequalities as equations in the first quadrant (since $x_1 \geq 0$, $x_2 \geq 0$). We get the above convex region OABCD. We have to find a point which maximize the value of Z and also lies in the convex region OABCD. The desired point is obtained by moving parallel to $2x_1 + 3x_2 = k$ for some k, so long as $2x_1 + 3x_2 = k$ touches the extreme boundary point of the convex region OABCD. Hence C (2, 4) gives the maximum value of Z.

$$\max Z = 2x_1 + 3x_2 = 2 \times 2 + 3 \times 4 = 16 \text{ at } x_1 = 2 \text{ and } x_2 = 4.$$

Case-2 Problems with linear objective function and nonlinear as well as linear constraints:

Consider the problem

$$\text{Maximize} \quad Z = x_1 + 2x_2$$

$$\text{Subject to} \quad x_1^2 + x_2^2 \leq 1 \qquad\qquad (3.2.2)$$

$$2x_1 + x_2 \leq 2$$

$$x_1 \geq 0 , x_2 \geq 0$$
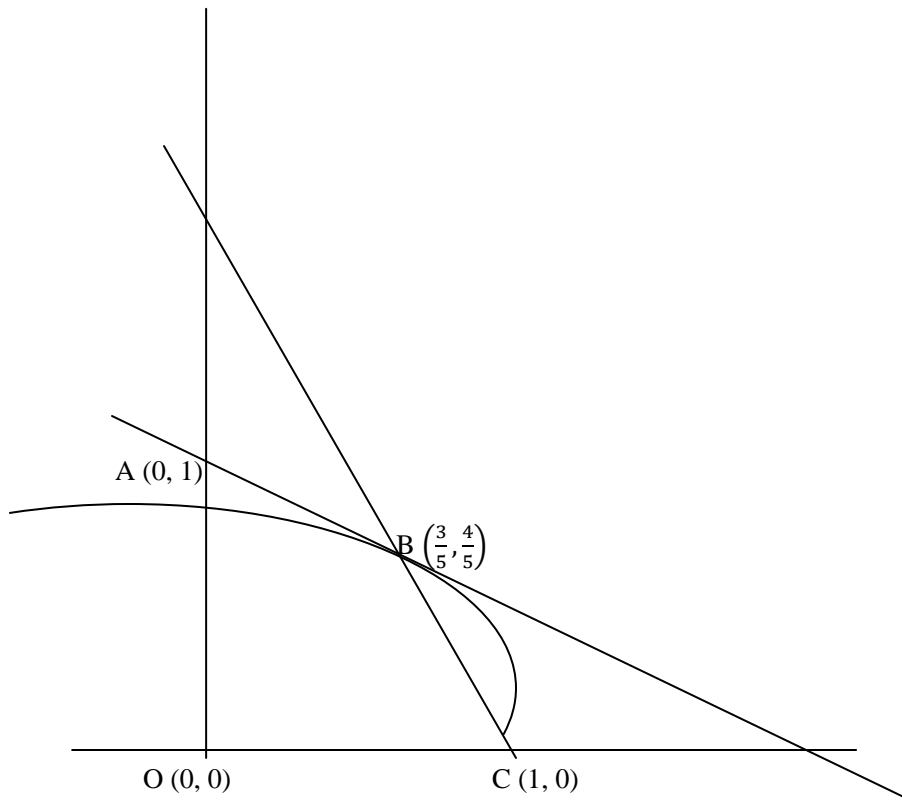
We'll solve this problem by graphical method.

Figure 3.2

For this we see that our objective function is linear and constraints are nonlinear. Constraints one is a circle of radius 1 with center at (0, 0) and constraints two is a straight line. OABC is the convex region having extreme points O (0, 0), A (0, 1), B $\left(\frac{3}{5}, \frac{4}{5}\right)$ and C (1, 0).

Max $Z = \frac{3}{5} + 2 \times \frac{4}{5} = \frac{11}{5}$ at $x_1 = \frac{3}{5}$ and $x_2 = \frac{4}{5}$.

The nonlinear programming problems having more than two variables can't be solved using graphical method. In such cases we have to use the method of Lagrange multipliers.

3.3   Lagrange multipliers method:

The basic ideas behind Lagrange multipliers can be illustrated by consider the following special case of the problem defined by Eq. (3.3.1) with three variables and only two equality constraints:

$$\text{Min } f(x_1, x_2, x_3)$$

Subject to

$$h_1(x_1, x_2, x_3) = 0 \qquad\qquad (3.3.1)$$

$$h_2(x_1, x_2, x_3) = 0$$

The feasible set $\boldsymbol{\Omega}$ is a curve in R^3, determined by the intersection of two surfaces given by h_1 and h_2. We assume that $x^* = (x_1{}^*, x_2{}^*, x_3{}^*)$ is a point of minimum in $\boldsymbol{\Omega}$ and that the functions f, h_1, h_2 have continuous first order derivatives on some open set containing x^* and $\nabla h_1(x^*), \nabla h_1(x^*)$ are linearly independent. Since $\nabla h_i(x^*)$, $i - 1,2$, are linearly independent and continuous, x_2, x_3 can be solved in term of x_1 as $x_2 = u(x_1)$ and $x_3 = v(x_1)$ in the neighborhood around x^*. Further u and v are differentiable around x^*. Hence, Eq. (3.3.1) reduces to an unconstrained minimization problem of the following form

$$\begin{matrix} min \\ x_1 \in R \end{matrix} f(x_1, u(x_1), v(x_1)) \qquad\qquad (3.3.2)$$

involving a single variable x_1. Since $x_1{}^*$ is a point of minimum of Eq. (3.3.2), we have
$$f'(x_1{}^*) = 0$$

$$\frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}\frac{du}{dx_1} + \frac{\partial f}{\partial x_3}\frac{dv}{dx_1} = 0 \qquad\qquad (3.3.3)$$

holds at $x^* = (x_1{}^*, x_2{}^*, x_3{}^*)$. Also, from the constraint equations, we have two additional relations

$$\frac{dh_i}{dx_1} = 0 , 1 \le i \le 2$$

$$\frac{\partial h_i}{\partial x_1} + \frac{\partial h_i}{\partial x_2}\frac{du}{dx_1} + \frac{\partial h_i}{\partial x_3}\frac{dv}{dx_1} = 0, \ \ 1 \le i \le 2 \qquad (3.3.4)$$

at $x^* = (x_1{}^*, x_2{}^*, x_3{}^*)$. From Eq. (3.3.3) and Eq. (3.3.4)

$$A\begin{bmatrix} 1 \\ \dfrac{du}{dx_1} \\ \dfrac{dv}{dx_1} \end{bmatrix} = 0, \quad A = [\nabla f(x^*) \quad \nabla h_1(x^*) \quad \nabla h_2(x^*)]^T$$

Since $(1, \frac{du}{dx_1}, \frac{dv}{dx_1})$ is not a null vector, we have det A= 0, which implies

$$a\nabla f(x^*) + b\nabla h_1(x^*) + c\nabla h_2(x^*) = 0 \qquad (3.3.5)$$

where a, b, c are not all zero. But $a \ne 0$ as $\nabla h_1(x^*), \nabla h_1(x^*)$ are linearly independent and hence there exist λ_1, λ_2, both not simultaneously zero, such that

$$\nabla f(x^*) + \lambda_1 \nabla h_1(x^*) + \lambda_2 \nabla h_2(x^*) = 0 \qquad (3.3.6)$$

Here λ_1, λ_2 are called Lagrange multipliers. So, in the equality constrained minimization problems, we have to find the critical point of the Lagrange function

$$L(x^*; \lambda_1, \lambda_2) = f(x^*) + \lambda_1 h_1(x^*) + \lambda_2 h_2(x^*) \qquad (3.3.7)$$

instead of the critical points of $f(x^*)$. The minimum of the Lagrangian function is also a minimum of f, because equality constraints at the critical point x^* imply that

$$\lambda_1 h_1(x^*) + \lambda_2 h_2(x^*) = 0$$

And hence

$$L(x^*; \lambda_1, \lambda_2) = f(x^*)$$

The above analysis helps us in obtaining the following generalization of the above fact the n-dimensional minimization problem with equality constraints:

$$\min f(x_1, x_2, \dots, x_n)$$

subject to $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (3.3.8)

$$h_i(x_1, x_2, \dots, x_n) = 0, 1 \le i \le m$$

Example 3.3.1 Solve the problem

$$\text{Min } x_1 + x_2$$

subject to

$$x_1^2 + x_2^2 = 1$$

Solution: The Lagrangian $L(x^*; \lambda)$ is given by

$$L(x^*; \lambda) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 1)$$

The critical points of the Lagrangian $L(x^*; \lambda)$ are given by $\nabla L(x^*; \lambda) = 0$

$$\frac{\partial L}{\partial x_1} = 1 + 2\lambda x_1 = 0 \qquad \dots (1)$$

$$\frac{\partial L}{\partial x_2} = 1 + 2\lambda x_2 = 0 \qquad \dots (2)$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_2^2 - 1 = 0 \qquad \dots (3)$$

Solving (1) and (2) and substituting in (3) we get,

$$(x_1^*; \lambda_1) = (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}; \frac{1}{\sqrt{2}})$$

$$(x_2^*; \lambda_2) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}; -\frac{1}{\sqrt{2}})$$

The Hessian matrix of $L(x^*; \lambda)$ with respect to x^* is given by

$$H_L(x^*; \lambda) = \begin{bmatrix} 2\lambda & 0 \\ 0 & 2\lambda \end{bmatrix}$$

The Hessian matrix H_L is positive definite at $(x_1^*; \lambda_1)$ and negative definite at $(x_2^*; \lambda_2)$

Hence $(x_1^*; \lambda_1)$ is a minimizer and $(x_2^*; \lambda_2)$ is a maximizer for the above problem.

Example 3.3.2

$$\text{Minimize } f(x) = x_1^2 + x_2^2 + x_3^2$$

subject to

$$x_1 + 2x_2 + 3x_3 = 7$$

$$2x_1 + 2x_2 + x_3 = \frac{9}{2}$$

Solution: The Lagrangian $L(x^*; \lambda)$ is given by

$$L(x^*; \lambda^*) = x_1^2 + x_2^2 + x_3^2 + \lambda_1(x_1 + 2x_2 + 3x_3 - 7) + \lambda_2\left(2x_1 + 2x_2 + x_3 - \frac{9}{2}\right)$$

The critical points of the Lagrangian $L(x^*; \lambda)$ are given by $\nabla L(x^*; \lambda^*) = 0$, we get

$2x_1 + \lambda_2 + 2\lambda_3 = 0$

$2x_2 + 2\lambda_1 + 2\lambda_2 = 0$

$2x_3 + \lambda_1 + \lambda_2 = 0$

$x_1 + 2x_2 + 3x_3 - 7 = 0$

$2x_1 + 2x_2 + x_3 - \frac{9}{2} = 0$

Solving these five equations, we get $\lambda_1 = -1$, $\lambda_2 = 0$ and $x_1 = \frac{1}{2}$, $x_2 = 1$, $x_3 = 3/2$.

The Hessian of $L(x^*; \lambda^*)$ with respect to x^* is $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, which is positive definite and

hence $(1/2, 1, 3/2)$ is a point of minimum.

$$\text{Min } f = x_1^2 + x_2^2 + x_3^2 = \left(\frac{1}{2}\right)^2 + 1^2 + \left(\frac{3}{2}\right)^2 = \frac{7}{2}$$

Example 3.3.3

Minimize $2x_1 + x_2$

Subject to $x_1^2 + x_2^2 - 1 = 0$

We have $L = 2x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 1)$.

The optimality conditions are

$$\frac{\partial L}{\partial x_1} = 2 + 2\lambda x_1 = 0, \quad x_1 = -\frac{1}{\lambda}$$

$$\frac{\partial L}{\partial x_2} = 1 + 2\lambda x_2 = 0, \quad x_2 = -\frac{1}{2\lambda}$$

Substituting for x (λ) into the constraint equation $h = 0$, we get $\lambda = \pm\frac{\sqrt{5}}{2}$, for each root, we obtain

$$\lambda = \frac{\sqrt{5}}{2}: \quad x_1^* = -\frac{2}{\sqrt{5}}, x_2^* = -\frac{1}{\sqrt{5}} \Rightarrow \text{Point A in fig.3.1}$$

$$\lambda = -\frac{\sqrt{5}}{2}: \quad x_1^* = \frac{2}{\sqrt{5}}, x_2^* = \frac{1}{\sqrt{5}} \Rightarrow \text{Point B in fig.1}$$
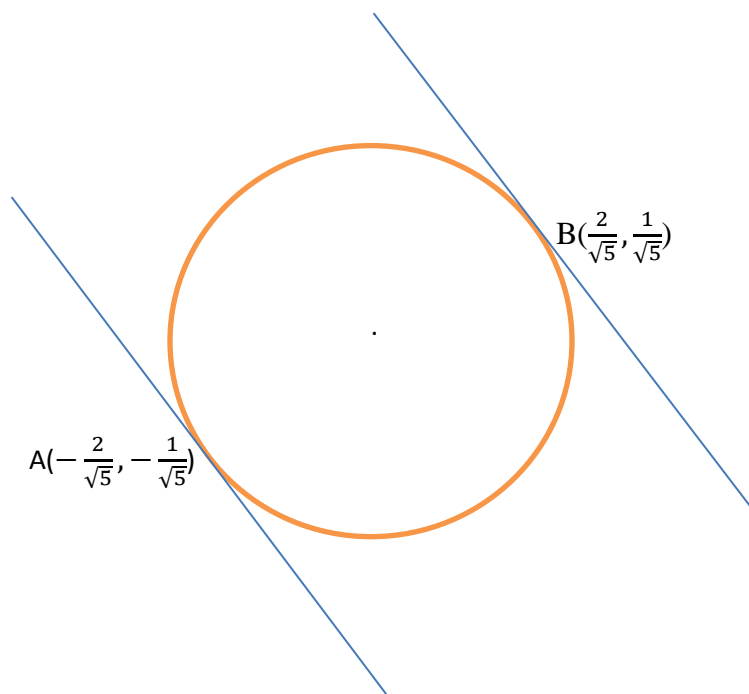
The minimum point is A.

Figure 3.3

Chapter 4

Optimality conditions:

4.1　The Fritz John (FJ) optimality conditions:

We now reduce the necessary optimality conditions to a statement in terms of gradients of the objective function and of the constraints. The resulting optimality conditions, credited to Fritz John, are given below. The Fritz John conditions are a necessary condition for a solution in nonlinear programming to be optimal.

The concept of Kuhn-tucker conditions and Duality are central discussion on constrained optimization.

4.1.1　Theorem (The Fritz John (FJ) necessary conditions):

Let X be a nonempty open set in R^n and let $f: R^n \to R$, and $g_i: R^n \to R$ for i=1,2,…,m. Consider the problem in Eq. 3.1.1 to

$$\text{Minimize} \quad f(x)$$

$$\text{subject to} \quad g_i(x) \leq 0 \qquad i = 1,2, \dots, m \qquad\qquad (4.1.1)$$

$$\text{and} \quad h_j(x) = 0 \qquad j = 1,2, \dots, l \quad , x \in X$$

Let $\bar{x}$ be a feasible solution, and denote $I = \{i: g_i(x) = 0\}$. Suppose that g_i for $i \in I$ is continuous at $\bar{x}$, that f and g_i are differentiable at $\bar{x}$, and that h_i for $i = 1,2, \dots, l$ is continuously differentiable at $\bar{x}$. If $\bar{x}$ locally solves above problems, then there exists scalars u_0, u_i for $i \in I$ and v_i for $i = 1,2, \dots, l$ such that

$$u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{x}) = 0$$

$$u_0, u_i \geq 0, \quad \text{for } i \in I \qquad\qquad (4.1.2)$$

$$u_0, u_I, v \neq (0,0,0)$$

Where u_I is the vector whose components are u_i for $i \in I$ and $v = (v_1, v_2, \dots, v_l)^T$. Furthermore, if each g_i for i L I is also differentiable at $\bar{x}$, then the Fritz John conditions can be written in the following equivalent form

$$u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{x}) = 0$$

$$u_i g_i(\bar{x}) = 0, \ i = 1, \dots \dots, m \qquad\qquad (4.1.3)$$

$$u_0, u_i \geq 0, \ i = 1, \dots \dots, m$$

$$u_0, u_I, v \neq (0,0,0)$$

In the condition of above theorem, the scalars u_0 and u_i for $i = 1, \dots \dots, m$ are usually called Langrage multipliers. The condition that $\bar{x}$ is feasible to the problem Eq. (4.1.1) is called the primal feasibility (PF) condition, whereas the requirements $u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{x}) = 0$, and $u_0, u_i \geq 0$ are sometimes called to as dual feasibility (DF) conditions. The condition $u_i g_i(\bar{x}) = 0$, for $i = 1, \dots \dots, m$ is called the complementary slackness (CS) condition. Together, the PF, DF, and the CS conditions are called the Fritz John (FJ) optimality conditions. Any point $\bar{x}$ for which there exist Lagrangian multipliers (u_0, u_i) such that $(\bar{x}, u_0, u_i)$ satisfy the FJ conditions is called a Fritz John (FJ) point.

Example 4.1.2

Minimize $\quad (x_1 - 3)^2 + (x_2 - 2)^2$

Subject to $\quad x_1^2 + x_2^2 \leq 5$

$$x_1 + 2x_2 \leq 4$$

$$-x_1 \leq 0$$

$$-x_2 \leq 0$$

The feasible region for the above problem is illustrated in figure 3.2. The Fritz -John conditions are true at the optimal point (2, 1). The set of binding constraints I at $x = (2, 1)^t$ is given by I= {1, 2}. Thus, the Lagrangain multipliers u_3 and u_4 associated with $-x_1 \leq 0$ and $-x_2 \leq 0$, respectively, are equal to zero.

$$\nabla f(x) = (-2, -2)^t, \quad \nabla g_1(x) = (4, 2)^t, \quad \nabla g_2(x) = (1, 2)^t$$
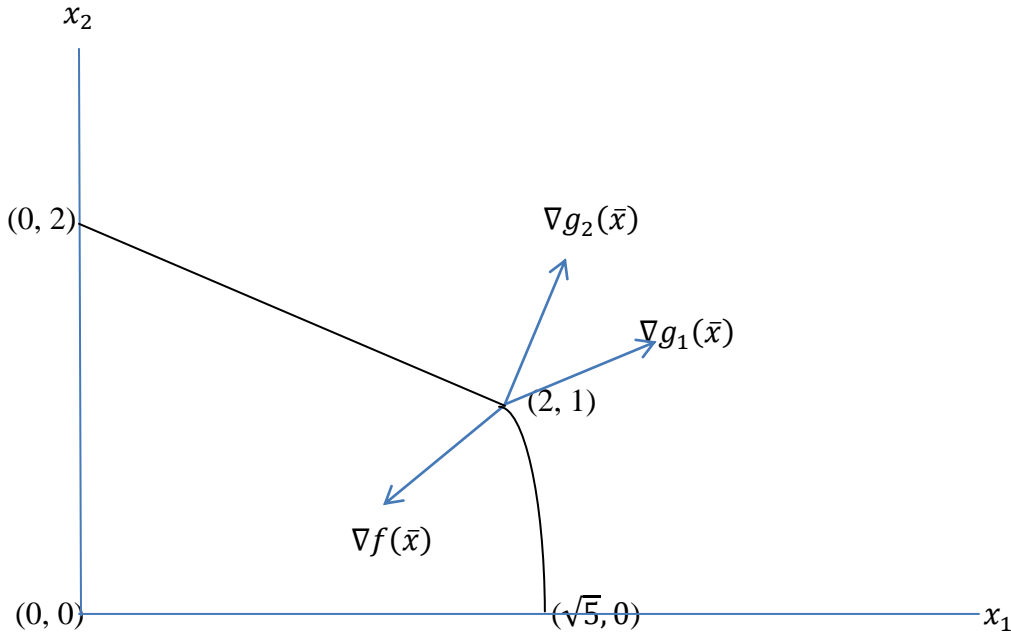
Figure 4.1

Hence, to satisfy the Fritz John conditions, we need a nonzero vector $(u_0, u_1, u_2) \geq 0$ satisfying

$$u_0 \begin{pmatrix} -2 \\ -2 \end{pmatrix} + u_1 \begin{pmatrix} 4 \\ 2 \end{pmatrix} + u_2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This implies $u_1 = \frac{u_0}{3}$ and $u_2 = \frac{2u_0}{3}$. Taking u_1 and u_2 as such for any $u_0 > 0$, we satisfy the Fritz Jhon conditions.

Let us check whether the point $x = (0,0)^t$ is a FJ point, here, the set of binding constraints is $I = \{3, 4\}$ and thus $u_1 = u_2 = 0$, Note that

$$\nabla f(x) = (-6, -4)^t \quad \nabla g_3(x) = (-1,0)^t \quad \nabla g_4(x) = (0,-1)^t$$

Also the DF condition

$$u_0 \begin{pmatrix} -6 \\ -4 \end{pmatrix} + u_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + u_4 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

holds true if and only if $u_3 = -6u_0$ and $u_4 = -4u_0$. If $u_0 > 0$, then $u_3, u_4 < 0$, contradicting the nonnegativity restrictions. If, on the other hand, $u_0 = 0$, then $u_3 = u_4 = 0$, which contradicts the stipulation that the vector (u_0, u_3, u_4) is nonzero. Thus, the Fritz John conditions do not hold true at $\hat{x} = (0,0)^t$, which also shows that the origin is not a local optimal point.

4.2   KKT optimality conditions with both Equality and Inequality constraints:

In the Fritz John conditions, the Langragian multiplier associated with the objective functions not necessarily positive. Under further assumptions on the constraint set we can show that at any local minimum, there exists a set of Langragian multiplier for which u_0 is positive. So we obtain a generalization of the KKT necessary optimality conditions.

4.2.1   KKT necessary optimality conditions:

Let X be a nonempty open set in R^n and let $f: R^n \to R$, and $g_i: R^n \to R$ for i=1,2,…,m. Consider the problem to

Minimize $f(x)$

subject to $g_i(x) \le 0$ $i = 1,2, … , m$ (4.2.1)

and $h_j(x) = 0$ $j = 1,2, … , l$, $x \in X$

Let $\bar{x}$ be a feasible solution, and denote $I = \{i: g_i(x) = 0\}$. Suppose that g_i for $i \in I$ is continuous at $\bar{x}$, that $f \, and \, g_i$ are differentiable at $\bar{x}$, and that h_i for $i = 1,2, … , l$ is continuously differentiable at $\bar{x}$. We assume that $\bar{x}$ is a regular point (gradients of active inequalities and of all the equality constraints are linearly independent). If $\bar{x}$ locally solves above problems, and each g_i for i L I is also differentiable at $\bar{x}$, then there exists scalars u_i for $i \in I$ and v_i for $i = 1,2, … , l$ exist such that

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{l} v_i \, \nabla h_i(\bar{x}) = 0$$

$$u_i \ge 0, \quad , \; i = 1, … … , m \qquad\qquad (4.2.2)$$

$$u_i g_i(\bar{x}) = 0, \; \; i = 1, … … , m$$

$$g_i \le 0, \; \; i = 1, … … , m \text{ and } h_j = 0, j = 1, 2, … , l$$

The Lagrange multipliers v_i associated with the equality constraints are unrestricted in sign–they can be positive or negative; only the multipliers associated with the inequalities have to be non-negative.

4.2.2   KKT Sufficient Conditions for Optimality:

The KKT conditions are necessary conditions: if a point $\bar{x}$ is local minimum to the problem in Eq. (4.1.1), then it should satisfy Eq. (4.2.2). Conversely, if a point $\bar{x}$ satisfies Eq. (4.2.2), it can be a local minimum or a saddle point (neither a minimum nor a maximum). A set of sufficient

conditions will now be stated which, if satisfied, will ensure that the KKT point is a strict local minimum.

Let f, g and h be twice-continuously differentiable functions. Then, the point $\bar{x}$ is a strict local minimum to Eq, (4.1.1) if there exit Lagrange multipliers μ and λ, such that

- KKT necessary conditions in Eq. (4.2.2) are satisfied at $\bar{x}$ and
- The Hessian matrix
  $$\nabla^2 L(\bar{x}) = \nabla^2 f(\bar{x}) + \sum_{i=1}^m \mu_i \nabla^2 g_i(\bar{x}) + \sum_{j=1}^k \lambda_j \nabla^2 h_j(\bar{x}) \qquad (4.2.3)$$

is positive definite on a subspace of R^n as defined by the condition:

$y^T \nabla^2 L(\bar{x}) y > 0$ for every vector $y \neq 0$ which satisfies

$\nabla h_j(\bar{x})^T y > 0, j = 1, \dots \dots, l$ and $\nabla g_i(\bar{x})^T y = 0$ for all i for which $g_i(\bar{x}) = 0$ with $\mu_i > 0$

$$(4.2.4)$$

Consider the following two remarks to understand Eq. (4.2.4):

(1) To understand what y is, consider a constraint $h(x) = 0$. This can be graphed as a surface in x-space. Then, $\nabla h(\bar{x})$ is normal to the tangent plane to the surface at $\bar{x}$. Thus, if $\nabla h(\bar{x})^t y = 0$, that means y is perpendicular to ∇h and is hence in the tangent plane. The collection of all such y's will define the entire tangent plane. This tangent plane is a subspace of the entire space R^n. Thus, the KKT sufficient conditions only require $\nabla^2 L(\bar{x})$ to be positive definite on the tangent space M defined by $M = \{y : \nabla h_j(\bar{x})^T y = 0, j = 1, \dots, l$ and $\nabla g_i(\bar{x})^T y = 0$ for all i for which $g_i(\bar{x}) = 0$ with $\mu_i > 0\}$.

(2) If $y^T \nabla^2 L(\bar{x}) y > 0$ for every vector $y \in R^n$, then the Hessian matrix $\nabla^2 L(\bar{x})$ is positive definite (on the whole space R^n). Then it will also positive definite on a (tangent) subspace of R^n.

Example 4.2.3

For the problem

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

$$\text{Subject to } 2x_1 + x_2 - 2 \leq 0 \qquad\qquad \equiv g_1$$

$$-x_1 \leq 0 \qquad\qquad \equiv g_2$$

$$-x_1 \leq 0 \qquad\qquad \equiv g_3$$

(i)   Write down the KKT conditions and find out the KKT points.
(ii)   Graph the problem and check whether the KKT point is a solution to the problem.

(iii)   Sketch the feasible "cone" at the points $(1, 0)$; that is, sketch the intersection of the descent and feasible cones.

(iv)   What is the feasible cone at the KKT point obtained in (i)?

We have $L = f + \mu_1(2x_1 + x_2 - 2) - \mu_2 x_1 - \mu_3 x_2$

The KKT conditions

$$2(x_1 - 3) + 2\mu_1 - \mu_2 = 0$$

$$2(x_2 - 3) + \mu_1 - \mu_3 = 0$$

$$\mu_1(2x_1 + x_2 - 2) = 0, \ \mu_2 x_1 = 0, \ \mu_3 x_2 = 0$$

$$\mu_1, \mu_2, \mu_3 \geq 0$$

$$2x_1 + x_2 \leq 2, \quad x_1 \geq 0, \quad x_2 \geq 0$$

Owing to the 'switching' nature of the conditions $\mu_i g_i = 0$, we need to guess which constraints are active and then proceed with solving the KKT conditions. We have the following cases:

$$g_1 = 0, \mu_2 = 0, \mu_3 = 0 \ \text{(only 1}^\text{st} \text{ constraints active)}$$

$$g_1 = 0, g_2 = 0, \mu_3 = 0 \ \text{(1}^\text{st} \text{ and 2}^\text{nd} \text{ constraints active)}$$

$$g_1 = 0, \mu_2 = 0, g_3 = 0$$

$$g_1 = 0, g_2 = 0, g_3 = 0$$

$$\mu_1 = 0, \mu_2 = 0, \mu_3 = 0$$

$$\mu_1 = 0, g_2 = 0, \mu_3 = 0$$

$$\mu_1 = 0, \mu_2 = 0, g_3 = 0$$

$$\mu_1 = 0, \ g_2 = 0, g_3 = 0$$

Now, in case-2: $g_1 = 0, g_2 = 0, \mu_3 = 0$. We get, upon using the KKT conditions,

$$x_1 = 0, x_2 = 2, \mu_1 = 2, \mu_2 = -2, \mu_3 = 0, g_1 = 0, g_2 = 0, g_3 < 0$$

Since μ_2 is negative, we have violated the KKT conditions and here our guess is incorrect. The correct guess is Case 1: $g_1 = 0, \mu_2 = 0, \mu_3 = 0$. This leads to:$x_1 = 3 - \mu_1$, $x_2 = 3 - 0.5\,\mu_1$. Substituting this into $g_1 = 0$ gives $\mu_1 = 2.8$. We can then recover $x_1 = 0.2, x_2 = 1.6$. Thus, we have the KKT point

$$x_1 = 0.2, x_2 = 1.6, \mu_1 = 2, \mu_2 = 0, \mu_3 = 0, g_1 = 0, g_2 < 0, g_3 < 0$$

The feasible cone at $x_0 = (1,0)^t$ is shown in figure below. At the point $x^t = (0.2, 1.6)^t$, we have the negative cost gradient to be $-\nabla f = (5.6, 2.8)^t = 2.8(2, 1)^t$ and the active constraint gradient to be $\nabla g_1 = (2, 1)^t$. Since these two vectors are aligned with one another, there is no descent/feasible cone (or the cone is empty).
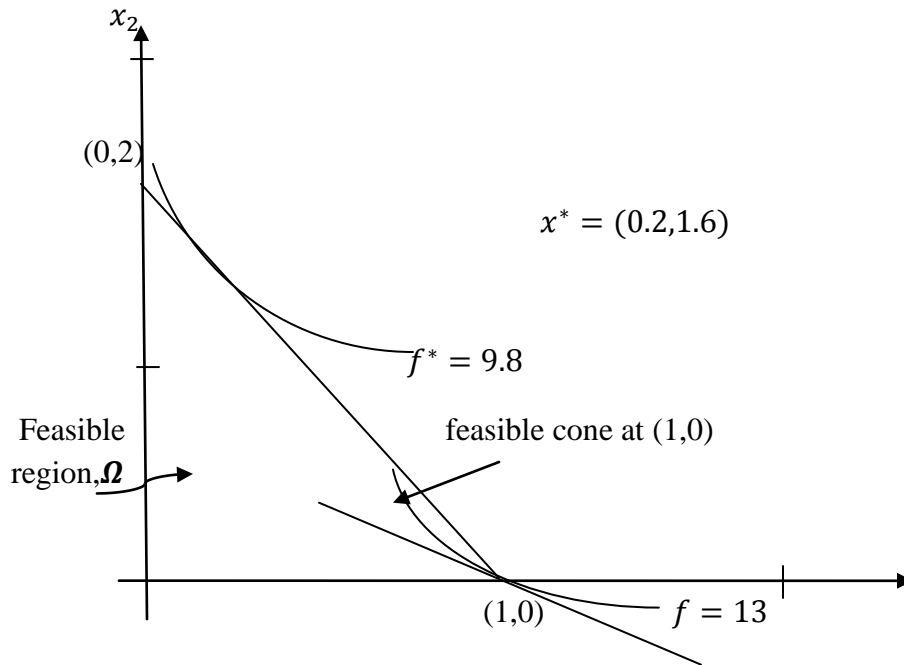

Figure 4.2

In the figure: x_2 axis, point $(0,2)$, $x^* = (0.2, 1.6)$, $f^* = 9.8$, Feasible region, Ω, feasible cone at $(1,0)$, $(1,0)$, $f = 13$.

4.3   Duality:

Consider the general nonlinear programming problem

$$\text{Minimize} \quad f(x)$$

$$\text{subject to} \quad g_i(x) \leq 0 \qquad i = 1, 2, \ldots, m \qquad\qquad (4.3.1)$$

$$\text{and} \quad h_j(x) = 0 \qquad j = 1, 2, \ldots, l$$

Where $x = (x_1, x_2, \ldots, x_n)^T$ is a column vector of n real-valued design variables. In equation (4.3.1), $f: R^n \to R$, is the objective or cost function, $g_i: R^n \to R$, are inequality constraints and $h_j: R^n \to R$, are equality constraints.

First, we assume that all functions are twice-continuously differentiable and a solution x^*, μ^*, λ^* of the 'primal' problem in Eq. (4.3.1) is a regular point satisfying the KKT necessary conditions given in the previous section. In the sufficient conditions for a minimum to Eq. (4.3.1), we only required the Hessian of the Lagrangian $\nabla^2 L(x^*, \mu^*, \lambda^*)$ be positive definite on the tangent

subspace. Now, to introduce duality, we have to make the stronger assumption that the Hessian of the Lagrangian $\nabla^2 L(x^*)$ is positive definite. This is a stronger assumption because we require positive definiteness of the Hessian at x^* on the whole space R^n and not just on a subspace. This is equivalent to requiring L to be locally convex at x^*. We can state the following equivalent theorems, under these assumptions:

(1) x^*, together with μ^* and λ^*, is solution to the primal problem of Eq.(4.3.1)
(2) x^*, μ^* and λ^* is a saddle point of the Lagrangian function $L(x, \mu, \lambda)$. That is,
$$L(x^*, \mu, \lambda) \leq L(x^*, \mu^*, \lambda^*) \leq L(x, \mu^*, \lambda^*)$$
in the neighborhood of x^*, μ^*, λ^*.
(3) x^*, μ^*, λ^* solves the dual problem

$$\text{Maximize} \quad L = f(x) + \mu^T g(x) + \lambda^T h(x) \qquad (4.3.2)$$

$$\text{subject to} \quad \nabla L = \nabla f(x) + [\nabla g(x)]\mu + [\nabla h(x)]\lambda = 0 \qquad (4.3.3)$$

$$\text{and} \quad \mu \geq 0 \qquad (4.3.4)$$

where both $\mathbf{x}$ and Lagrange multipliers μ, λ are variable and the two extrema are equal: $L(x^*, \mu^*, \lambda^*) = f(x^*)$.

From the above discussion the following two corollaries follows:

(a) Any $\mathbf{x}$ that satisfies the constraints in Eq. (4.3.1) is called primal-feasible and any x, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ that satisfies Eq. (4.3.3) and (4.3.4) is called dual-feasible. We know that any primal-feasible point provides us with a upper bound on the optimum cost:

$$f(x^*) \leq f(x) \text{ for any } x \text{ that is primal} - \text{feasible}$$

Any dual feasible point provides a lower bound on the optimum cost as

$$L(\mathbf{x}, \mu, \boldsymbol{\lambda}) \leq f(x^*) \text{ for any } x, \mu, \lambda \text{ that is dual} - \text{feasible}$$

It is important to realize that the lower bound is only on the local minimum of f in the neighborhood.

(b) If the dual objective is unbounded, then the primal has no feasible solution.

4.4   The Dual Function $\varphi(\mu, \lambda)$:

The conditions in Eq. (4.3.3), i.e. $\nabla L = \nabla f(x) + \nabla g(x)\mu + \nabla h(x)\lambda = 0$, can be viewed as equations involving two sets of variables: $\mathbf{x}$ and $[\mu, \lambda]$. The Jacobian with respect to the $\mathbf{x}$ variables of this system, is $\nabla^2 L$. Since this matrix is non-singular at x^*(by the hypothesis that it is

positive definite), we can invoke the Implicit Function Theorem which states that in a small neighborhood of $[x^*, \mu^*, \lambda^*]$, such that for μ, λ in this neighborhood, $x = x(\mu, \lambda)$ is a differentiable function of μ, λ with $\nabla L(x(\mu, \lambda), \mu, \lambda) = 0$. This gives us to treat L as an implicit function of μ, λ as $L = L(x(\mu, \lambda), \mu, \lambda)$, which is called the 'dual function' $\varphi(\mu, \lambda)$. Since $\nabla^2 L$ will remain positive definite near x^*, we can define $\varphi(\mu, \lambda)$ from the minimization problem

$$\varphi(\mu, \lambda) = minimum\ f(x) + \mu^T g(x) + \lambda^T h(x) \qquad (4.4.1)$$

Where $\mu \geq 0, \lambda$ are near μ^*, λ^*, and where the minimization is carried out with respect to x near x^*.

4.4.1   Gradient and Hessian of $\varphi(\mu, \lambda)$ in Dual Space:

The gradient of the dual function φ is

$$\frac{d\varphi}{d\mu_j} = g_j[x(\mu, \lambda)] \qquad (4.4.2)$$

$$\frac{d\varphi}{d\lambda_i} = h_i[x(\mu, \lambda)] \qquad (4.4.3)$$

The proof follows from: $\frac{d\varphi}{d\mu_j} = \frac{dL(x(\mu,\lambda),\mu,\lambda)}{d\mu_j} = \frac{\partial L}{\partial \mu_j} + \nabla L^T \frac{dx}{d\mu_j} = \frac{\partial L}{\partial \mu_j} = g_j$ in view of the fact that $\nabla L = 0$ by the very construction of φ.

The Hessian of φ in μ, λ space is given by

$$\nabla_D{}^2 \varphi = -A[\nabla^2 L(x^*)]^{-1} A^T$$

Where the rows of A are the gradient vectors of the constraints. Since we assume that x^* is a regular point, **A** has full row rank and hence the dual Hessian is negative definite.

Example: 4.4.2

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

$$\text{Subject to} \qquad 2x_1 + x_2 - 2 \leq 0 \qquad \equiv g_1$$

$$-x_1 \leq 0 \qquad \equiv g_2$$

$$-x_1 \leq 0 \qquad \equiv g_3$$

We have $L = (x_1 - 3)^2 + (x_2 - 3)^2 + \mu_1(2x_1 + x_2 - 2) - \mu_2 x_1 - \mu_3 x_2$. It is clear that $\nabla^2 L$ is positive definite. Setting $\frac{\partial L}{\partial x_1} = 0$ and $\frac{\partial L}{\partial x_2} = 0$, we obtain

$$x_1 = \frac{\mu_2}{2} - \mu_1 + 3, \qquad x_2 = \frac{\mu_3}{2} - \frac{\mu_1}{2} + 3$$

Substituting into L, we have

$$\varphi(\mu) = -\frac{5}{4}\mu_1{}^2 - \frac{\mu_2{}^2}{4} - \frac{\mu_3{}^2}{4} + \mu_1\mu_2 + \frac{\mu_1\mu_3}{2} + 7\mu_1 - 3\mu_2 - 3\mu_3$$

From Eq. (4.4.2) and eq. (4.4.3) we have

$$\frac{\partial\varphi}{\partial\mu_1} = -\frac{5}{2}\mu_1 + \mu_2 + \frac{\mu_3}{2} + 7$$

$$\frac{\partial\varphi}{\partial\mu_2} = -\frac{\mu_2}{2} + \mu_1 - 3$$

$$\frac{\partial\varphi}{\partial\mu_3} = -\frac{\mu_3}{2} + \frac{\mu_1}{2} - 3$$

$x_1 = \frac{\mu_2}{2} - \mu_1 + 3, \ x_2 = \frac{\mu_3}{2} - \frac{\mu_1}{2} + 3$

we have
$$\frac{\partial\varphi}{\partial\mu_i} = g_i$$

The dual problem is

$\max\varphi(\mu) = -\frac{5}{4}\mu_1^2 - \frac{\mu_2^2}{4} - \frac{\mu_3^2}{4} + \mu_1\mu_2 + \frac{\mu_1\mu_3}{2} + 7\mu_1 - 3\mu_2 - 3\mu_3$ with the restriction that $\mu_i \geq 0, \ i = 1, 2, 3$.

The solution is $\mu_1^* = 2 \cdot 8, \ \mu_2^* = 0, \ \mu_3^* = 0.$ Correspondingly, we obtain $x_1(\mu^*) = 0 \cdot 2$ and $x_2(\mu^*) = 1 \cdot 6$.

Chapter -5

References:

1) Nonlinear programming theory and algorithms, second edition, Mokhtar S. Bazaraa, Hanif D. Sherali, C.M. Shetty.
   John Wiley and sons publication
2) Optimization concepts and applications in engineering, Ashok D. Belegundu, Tirupathi R. Chandrupatla.
   Published by Pearson education (Singapore) Pte. Ltd.
3) Optimization theory and practice, Mohan C Joshi, Kannan M Moudgalya
   Narosa Publishing House, 2004
4) Effect of graphical method for solving mathematical programming problem, Bimal Chandra Das, Department of Textile Engineering, Daffodil International University 28 journal of Science and Technology, Vol. 4, January 2009, Dhaka, Bangladesh
5) Convex optimization, Stephen Boyd and Lieven Vandenerghe ,
   Cambridge University Press (2009)
6) Principles of Optimization Theory, C.R. Bector, S. Chandra, J. Dutta
   Narosa Publishing House
7) Convexity and optimization in finite dimension I, Josef Stoer, Christoph Witzgall
   Springer Verlag Berlin.Heidelberg.New york