INTERIOR AND EXTERIOR PENALTY METHODS TO SOLVE NONLINEAR OPTIMIZATION PROBLEMS

COLLEGE OF NATURAL SCIENCES
DEPARTMENT OF MATHEMATICS

"In Partial Fulfilment Of The Requirements For The Degree Of Master Of Science In Mathematics"

By: Kiflu Kemal
Stream: Optimization

Advisor: Berhanu Guta(PhD)

June,2017
Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY
DEPARTMENT OF MATHEMATICS

The undersigned hereby certify that they have read and recommend to the department of mathematics for acceptance of this project entitled "**Interior and exterior Penalty Method to Solve Nonlinear programming Problems**" by **Kiflu Kemal** in partial fulfilment of the requirements for the degree of Master of Science in Mathematics.

Advisor: Dr. Berhanu Guta

Signature: ____________________

Date: _________________________

Examiner 1: Dr.

Signature: ____________________

Date: _________________________

Examiner 2: Dr.

Signature: ____________________

Date: _________________________

ADDIS ABABA UNIVERSITY

Author: Kiflu Kemal

Title: Interior and Exterior Penalty Methods to Solve Nonlinear Optimization Problems

Department: Mathematics

Degree: M.Sc.

Convocation: June

Year: 2017

Permission is here with granted to Addis Ababa University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

Kiflu Kemal

Signature: ___________________

Date: ________________________

Acknowledgements

I would like to express my gratitude to my advisor, Dr. Birhanu Guta, for all his dedication, patience, and advice. Also, I would like to thank all Mathematics Instructors for their motivation and guidance during the past two years at Addis Ababa University, department of Mathematics,the Library workers. My thank also forwarded to my Brother Tilahun Blayneh and my confessor Aba Zerea Dawit who force me to join Mathematics department of Addis Ababa University and to all of my families and friends for all their invulnerable love and support.

Abstract

The methods that we describe presently, attempt to approximate a constrained optimization problem with an unconstrained one and then apply standard search techniques such as exterior penalty function method and interior penalty method to obtain solutions. The approximation is accomplished in the case of exterior penalty methods by adding a term to the objective function that prescribes a high cost for violation of the constraints. In the case of interior penalty function methods, a term is added that favors points in the interior of the feasible region over those near the boundary. For a problem with n variables and m constraints, both approaches work directly in the n-dimensional space of the variables. The discussion that follows emphasizes exterior penalty methods recognizing that interior penalty function methods embody the same principles.

Keywords: Constrained optimization, unconstrained optimization, Exterior penalty,Interior penalty(barrier) methods,Penalty Parameter,Penalty function, Penalty Term,Auxiliary function,non linear programming.

List of Notations

∇f: gradient of real valued function f
$\nabla^t f$: transpose of the gradient
$\Re$: set of real numbers
$\Re^n$: n dimensional space
$\Re^{n\times m}$: space of real $n \times m$ matrices
$\mathcal{C}$: a cone
$\frac{\partial f}{\partial x}$: partial derivative of f with respect to x
$H(x)$: Hessian matrix of a function at x
$\mathcal{L}$: Lagrangian function
$\mathcal{L}(., \lambda, \mu)$: Lagrangian function with Lagrange multipliers λ and μ
f_{μ_k}: auxiliary function for penalty methods with penalty parameter μ_k
$\alpha(x)$: penalty function
$P(x)$: barrier function
SDP:Positive Semidefinite
ϕ_{μ_k}: auxiliary function for barrier methods with penalty parameter μ_k
$< \lambda, h >$: inner product of vectors λ and h
$f \in \mathcal{C}^1$: f is once continuously differentiable function
$f \in \mathcal{C}^2$: f is twice continuously differentiable function

Table of Contents

Introduction

Since the early 1960s, the idea of replacing a constrained optimization problem by a sequence of unconstrained problems parameterized by a scalar parameter μ has played a fundamental role in the formulation of algorithms(Bertsekas,1999).
To do this replacement penalty methods have a vital role. Penalty methods approximate the solution for nonlinear constrained problem by minimizing the penalty function for a large value of μ or a smaller value ofμ.

Generally, penalty methods can be categorized in to two types, exterior penalty function methods (we can say simply penalty function methods) and interior penalty (barrier) function methods.

In exterior penalty methods some or all of the constraints are eliminate and add to the objective function a penalty term which prescribes a high cost to infeasible points. Associated with these methods is a parameter μ, which determines the severity of the penalty and as a consequence the extent to which the resulting unconstrained problem approximates the original constrained problem and can be illustrated:

$$\begin{array}{rl} \text{Minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0 \\ & x \in \Re^n \end{array} \tag{1}$$

By using exterior penalty function method the constrained optimization problem is converted in to the following unconstrained form:

$$\begin{array}{rl} \text{Minimize} & f(x) + \mu(Max\{0, g_i(x)\})^2 \\ & x \in \Re^n \end{array} \tag{2}$$

Similar to exterior penalty functions, interior penalty functions are also used to transform a constrained problem into an unconstrained problem or into a sequence of unconstrained problem. These functions set a barrier against leaving the feasible region.We can solve problem (1) by interior penalty function method. By converting it in to unconstrained problem in the following fashion:

$$\begin{array}{rl} \text{Minimize} & f(x) - \Sigma_{i=1}^{m} \dfrac{\mu}{g_i(x)}; for\{g_i(x) < 0\} \quad \text{and} \quad i = 1, \ldots, m, \\ & x \in \Re^n \end{array} \tag{3}$$

OR

$$\text{Minimize} \quad f(x) - \mu\Sigma_{i=1}^{m}\log\left[-g_i(x)\right]; for\{g_i(x) < 0\} \quad \text{and} \quad i = 1, \ldots, m, \\ x \in \Re^n \tag{4}$$

This paper considers exterior penalty function methods to find local minimizer of a nonlinear constrained problems with equality and inequality constraints and interior penalty function (barrier function) methods to solve nonlinear constrained problems with only inequality constraints locally.

In **Chapter-1** we try to discuss some basic, concepts of convex analysis and some additional preliminary concepts which help to understand the idea of the project.
Chapter-2 more explain about theories of nonlinear optimization, both unconstrained and constrained. The chapter focuses mainly on the minimization theories and basic conditions related to this optimization point of view.

Chapter-3 discusses on interior penalty function methods and exterior penalty function methods. Throughout the chapter, we try to describe some basic concepts and properties of the methods for nonlinear optimization problems. Definitions,algorithmic schemes of the respective methods,convergence theories and special properties of these methods are discussed in the chapter.

Chapter 1

Preliminary Concepts

1.1 Convex Analysis

1.2 Convex set and Convex function

The concept of convex set and convex function play an important role in the study of optimization(W.SUN,2007)

Definition 1.2.1 (Convex Sets). *Let $\mathcal{S} \subset \Re^n$ is said to be convex if the line segment joining any two points of S also belongs to S.*
In other words, if $x_1, x_2 \in S$, then $\lambda x_1 + (1-\lambda)x_2 \in S$ for each $\lambda \in [0,1]$.
***A convex combination** of a finite set of vectors $\{x_1, x_2, ..., x_n\}$ in $\Re^n$ is any vector x of the form*

$$x = \Sigma_{i=1}^{n} \alpha_i x_i, \quad \text{where} \quad \Sigma_{i=1}^{n} \alpha_i = 1, \quad \text{and} \quad \alpha_i \geq 0 \quad \text{for all} \quad i = 1, 2, \ldots, n.$$

***The convex hull** of the set S containing $\{x_1, x_2, \ldots, x_n\}$, denoted by $conv(S)$, is the set of all convex combinations of S. In other words, $x \in conv(S)$ if and only if x can be represented as a convex combination of $\{x_1, x_2, \ldots, x_n\}$.*
*If the non-negativity of the multipliers α_i for $i = 1, 2, \ldots, n$ is ignored, then the combination is said to be **an affine combination**.*
***A cone** is a non empty set $\mathcal{C}$ with the property that for all $x \in \mathcal{C}$, then $\alpha x \in \mathcal{C}$,symbolically we can write as:*

$$x \in \mathcal{C} \Rightarrow \alpha x \in \mathcal{C}, \quad \text{for all} \quad \alpha \geq 0.$$

For instance, the set $\mathcal{C} \subset \Re^2$ defined by $\{(x_1, x_2)^t | x_1 \geq 0, x_2 \geq 0\}$ is a cone in $\Re^2$.
***Note that** cones are not necessarily convex.*

For example:-

The set $\{(x_1, x_2)^t | x_1 \geq 0 \quad or \quad x_2 \geq 0\}$ *which encompasses three quarters of the two-dimensional plane is a cone, but not convex.*
The cone generated by $\{x_1, x_2, \ldots, x_n\}$ *is the set of all vectors* x *of the form*

$$x = \Sigma_{i=1}^{n} \alpha_i x_i \quad where, \quad \alpha_i \geq 0 \quad for\ all \quad i = 1, 2, \ldots, n.$$

Note that *all cones of this form are convex.*

Definition 1.2.2 (Convex Functions)**.** *Let* S *be a non empty convex set in* $\Re^n$*. As (J.Jahn,1996)defined that a function* $f : S \to \Re$ *is said to be convex for all* $x_1, x_2 \in S$ *if*

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

for each $\lambda \in [0, 1]$.

The function f *is said to be strictly convex on* S *if*

$$f[\lambda x_1 + (1 - \lambda)x_2] < \lambda f(x_1) + (1 - \lambda)f(x_2)$$

for each distinct $x_1, x_2 \in S$ *and for each* $\lambda \in (0, 1)$.

Theorem 1.2.1. ***Jensen's Inequality***
If g *is a convex function on a convex set* X *and* $x = \Sigma_{i=1}^{n} \alpha_i x_i, \quad where \quad \Sigma_{i=1}^{n} \alpha_i = 1, \quad and \quad \alpha_i \geq 0 \quad for\ all \quad i = 1, 2, \ldots, n.$*,then* $g(\Sigma_{i=1}^{n} \alpha_i x_i) \leq \Sigma_{i=1}^{n} \alpha_i g(x_i)$

Definition 1.2.3 (Concave Functions)**.** *A function* $f(x)$ *is said to be concave function over the region* S *if for any two points* $x_1, x_2 \in S$*.*
We have the function

$$f[\lambda x_1 + (1 - \lambda)x_2] \geq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

where $\lambda \in [0, 1]$.
S *is strictly concave function if*

$$f[\lambda x_1 + (1 - \lambda)x_2] > \lambda f(x_1) + (1 - \lambda)f(x_2)$$

for each distinct $x_1, x_2 \in S$ *and for each* $\lambda \in (0, 1)$.
Similarly we can describe that; If the function $-f$ *is a convex (strictly convex,uniformly convex) function on* S*,then* f *is said to be a concave (strictly concave,uniformly concave) function(W.SUN,2006)*

Lemma 1.2.1. *Let* S *be a non empty convex set in* $\Re^n$*, and let* $f : S \to \Re$ *be a convex function. Then the level set* $S_\alpha = \{x \in S : f(x) \leq \alpha\}$*, where* α *is real number, is a convex set.*

Proposition 1.2.1. *If* g *is a convex function on a convex set* X*, then the function* $g(x) = \max\{g(x), 0\}$ *is also convex on* X*.*

Proof 1.2.1. *Suppose $x, y \in X$ and $\lambda \in [0,1]$. Then*

$$\begin{aligned} g(\lambda x + (1-\lambda)y)) &= max\{g(\lambda x + (1-\lambda)y), 0\} \\ &\leq max\{\lambda g(x) + (1-\lambda)g(y), 0\}, \quad \textit{since g is convex} \\ &\leq max\{\lambda g(x), 0\} + max\{(1-\lambda)g(y), 0\}, \\ &= \lambda max\{g(x), 0\} + (1-\lambda)max\{g(y), 0\} \\ &= \lambda g(x) + (1-\lambda)g(y). \end{aligned}$$

Proposition 1.2.2. *If h is convex and non-negative on a convex set X, then h^2 is also convex on X.*

Proof 1.2.2. *Suppose $x, y \in X$ and $\lambda \in [0,1]$. Then*

$$\begin{aligned} h^2(\lambda x + (1-\lambda)y)) &= [h(\lambda x + (1-\lambda)y)][h(\lambda x + (1-\lambda)y)] \\ &\leq [\lambda h(x) + (1-\lambda)h(y)][\lambda h(x) + (1-\lambda)h(y)] \\ &= \lambda h^2(x) + (1-\lambda)h^2(y) - \lambda(1-\lambda)(h(x) - h(y))^2 \\ &\leq \lambda h^2(x) + (1-\lambda)h^2(y) \end{aligned}$$

Chapter 2

Optimization Theory and Methods

Optimization Theory and Methods is a young subject in applied mathematics,computational mathematics and operations research(W.SUN,2006)
The subject is involved in optimal solution of problems which are defined mathematically, i.e., given a practical problem, the best solution to the problem can be found from lots of schemes by means of scientific methods and tools. It involves the study of optimality conditions of the problems, the construction of model problems, the determination of algorithmic method of solution, the establishment of convergence theory of the algorithms, and numerical experiments with typical problems and real life problems.
The general form of optimization problems is

$$\begin{array}{ll} \text{Minimize} & f(x) \\ x \in X & \end{array} \tag{2.1}$$

where $x \in X$ is decision variable,$f(x) : \Re^n \rightarrow \Re$ an objective function and the set $X \in \Re^n$is the feasible set of (2.1) .Based on the description of the function f and the feasible set X the problem (2.1) can be classified as linear, quadratic, non-linear, multiple-objective problem etc.

2.1 Some Classes of Optimization Problems

2.1.1 Linear Programming

If the objective function f and the defining functions of X are linear, then (2.1) will be a linear optimization problem.
General form of a linear programming problem:

$$\begin{aligned} \text{Minimize} \quad & C^T x \\ \text{subject to} \quad & Ax = a \\ & Bx \leq b \\ & \quad x \in \Re^n \end{aligned} \tag{2.2}$$

where $f(x) = C^T x$ and $X = \{x \in \Re^n | Ax = a, Bx \leq b\}$
Under linear programming problems there are practical problems such as: linear discrete problems, transportation problems, network flow problems,etc. and we use simplex method,Big M method, Dual simplex method,Graphical method etc. to find the solution of those linear programming problems

2.1.2 Quadratic Programming

$$\begin{aligned} \text{Minimize} \quad & f(x) = \frac{1}{2} x^T Q x + q^T x + r \\ \text{subject to} \quad & Ax = a \\ & Bx \leq b \\ & \quad x \in \Re^n \end{aligned} \tag{2.3}$$

Here the objective function $f(x) = \frac{1}{2} x^T Q x + q^T x + r$ is Quadratic while the feasible set $X = \{x \in \Re^n | Ax = a, Bx \leq b\}$ is defined using linear function and r is constant.

2.1.3 Non Linear Programming Problems

The general form of a non-linear optimization problem is:

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \quad \text{for} \quad i = 1, \ldots, l, \\ & g_j(x) \leq 0, \quad \text{for} \quad j = 1, \ldots, m, \\ & \quad x \in \Re^n \end{aligned} \tag{2.4}$$

where, we assume that all the functions are smooth. The feasible set of the (NLPP) is given by $X = \{x \in \Re^n | h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, l; g_j(x) \leq 0 \quad \text{for} \quad j = 1, \ldots, m\}$. Through out this paper our interest is in solving Non-linear programming problems by classifying them primarily as unconstrained and constrained optimization problems. Particularly, if the feasible set $X = \Re^n$, the optimization problem (2.1) is called an unconstrained optimization problem where as the problems of type (2.4) are said to be constrained optimization problems. Generally Optimization problems can be classified as unconstrained optimization problem and constrained optimization problems.

2.2 Unconstrained Optimization

Unconstrained optimization problem has the following form

$$\text{Minimize} \quad f(x) \quad ,\text{subject to} \quad x \in \Re^n$$

where $f : \Re^n \to \Re$ is a given function. The first thing will be to derive some conditions, which allow to decide whether a point is a minimum or not.

Definition 2.2.1. *i. A point x^* is a local minimizer if there is a neighbourhood η of x^* such that $f(x^*) \leq f(x)$ for all $x \in \eta$.*

ii. A point x^* is a strict local minimizer (also called a strong local minimizer) if there is a neighbourhood η of x^* such that $f(x^*) < f(x)$ for all $x \in \eta$ with $x \neq x^*$.

iii. A point x^* is an isolated local minimizer if there is a neighbourhood η of x^* such that x^* is the only local minimizer in η.

iv. All isolated local minimizers are strict local minimizers.

v. We say that x^* is a global minimizer if

$$f(x^*) \leq f(x) \quad \textit{for all} \quad x \in \Re^n.$$

When the function f is smooth, there are efficient and practical ways to identify local minima. In particular, if f is twice continuously differentiable, we may be able to tell that x^* is a local minimizer (and possibly a strict local minimizer) by examining just the gradient $\nabla f(x^*)$ and the Hessian $\nabla^2 f(x^*)$.

There is no general procedure to determine whether the local minimum is really a global minimum in a non-linear optimization problem(KUMAR,2014)

Definition 2.2.2. ***Gradient****of $f : \Re^n \to \Re$ at $x^* \in \Re^n$ defined as:*

$$\begin{pmatrix} \frac{\partial f(x^*)}{\partial x_1} \\ \frac{\partial f(x^*)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

Note

1. For one variable x; $f_x(x^*)=\frac{\partial f(x^*)}{\partial x}$ is the derivative of f at x^*

$f_{xx}(x^*)=\frac{\partial^2 f(x^*)}{\partial x^2}$

- if $\frac{\partial^2 f(x^*)}{\partial x^2} > o$, then f attains its minimum.

- if $\frac{\partial^2 f(x^*)}{\partial x^2} < o$, then f attains its maximum.

- if $\frac{\partial^2 f(x^*)}{\partial x^2} = o$, then f need further investigation.

2. For two variables x_1, x_2.

$$\begin{pmatrix} \frac{\partial^2 f(x^*)}{\partial x_1^2} & \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x^*)}{\partial x_2 x_1} & \frac{\partial^2 f(x^*)}{\partial x_2^2} \end{pmatrix}$$

also called the Hessian matrix.

- if $rt - s^2 > 0$, then f attains its minimum.

- if $rt - s^2 < 0$, then f attains its maximum.

- if $rt - s^2$=0, then f need further investigation.

Where $r = \frac{\partial^2 f(x^*)}{\partial x_1^2}$, $s = \frac{\partial^2 f(x^*)}{\partial x_1 x_2}$, $t = \frac{\partial^2 f(x^*)}{\partial x_2^2}$

3.For n variables $x_1, x_2, ...x_n$ of the Hessian of the function f at x^* defined as:

$$\begin{pmatrix} \frac{\partial^2 f(x^*)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x^*)}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots \\ \frac{\partial^2 f(x^*)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_n^2} \end{pmatrix}$$

- if $|H| > 0$, then f attains its minimum.
- if $|H| < 0$, then f attains its maximum.
- if $|H| = 0$, then f need further investigation.

Definition 2.2.3. *i. A symmetric $n \times n$ matrix M is said to be positive semi definite if $x^t M x \geq 0$ for all $x \in \Re^n$. Now we can said that*

$$M \geq 0.$$

ii. We say that M is positive definite if $x^t M x > 0$ for all $x \neq 0$.

iii. When we say that M is positive (semi) definite we implicitly assume that it is symmetric.

iv. Let $M \in \Re^{n\times n}$ be a matrix. Then the eigenvalues of M are scalar λ such that

$$Mx = \lambda x, \quad \textit{where} \quad x \neq 0, \quad x \in \Re^n.$$

v. *The eigenvalues of symmetric matrices are all real numbers, while nonsymmetric matrices may have imaginary eigenvalues.*

Property 2.2.1. *If all the eigenvalues of a matrix M are positive (nonnegative), then M is positive definite (semi definite).*

Property 2.2.2. *If M is positive definite, then M^{-1} is positive definite.*

Property 2.2.3. *Let P be a symmetric $n \times n$ matrix and Q be a positive semi definite $n \times n$ matrix. Assume that $x^t Px > 0$ for all $x \neq 0$ satisfying $x^t Qx = 0$. Then there exists a scalar μ such that $P + \mu Q$ is positive definite.*

2.3 Optimality Conditions

Proposition 2.3.1 (Necessary Optimality Conditions)**.** *Assume that x^* is a local minimizer of f and $f \in C^1$ over η. Then*

$$\nabla f(x^*) = 0.$$

But this condition is not sufficient to guarantee a minimum, because it could also be a maximum or a saddle point. To ensure a minimum a second-order condition is necessary.

If in addition $f \in C^2$ over η, then

$$\nabla^2 f(x^*) \geq 0.$$

Proposition 2.3.2 (Sufficient Optimality Conditions)**.** *Let $f \in C^2$ over η, $\nabla f(x^*) = 0$, and $\nabla^2 f(x^*) > 0$, i.e., the function is locally convex in x^*. Then x^* is a strict local minimizer for f.*

Note that: if the objective function is convex, local and global minimizers are simple to characterize.

Theorem 2.3.1. *When f is convex, any local minimizer x^* is a global minimizer of f. If in addition f is differentiable, then any stationary point x^* (i.e., a point satisfying the condition $\nabla f(x^*) = 0$) is a global minimizer of f.*

Proof 2.3.1. *Suppose that x^* is a local but not a global minimizer. Then we can find a point $z \in \Re$ with $f(z) < f(x^*)$. Consider the line segment that joins x^* and z, that is*

$$x = \lambda z + (1 - \lambda)x^*, \quad \textit{for some} \quad \lambda \in (0, 1]. \tag{2.5}$$

By the convexity property for f, we have

$$f(x) \leq \lambda f(z) + (1 - \lambda)f(x^*) < f(x^*). \tag{2.6}$$

Any neighbourhood $\mathcal{N}$ of x^* contains a piece of line segment 2.5, so there will always be points $x \in \mathcal{N}$ at which 2.6 is satisfied. Hence, x^* is not a local minimizer, which contradicts the assumption. Therefore, x^* is a global minimizer.

For the second part of the theorem, suppose that x^* is not a global minimizer and choose z as above. Then, from convexity, we have

$$\begin{aligned}\nabla f(x^*)^t(z-x^*) &= \frac{d}{d\lambda} f(x^* + \lambda(z-x^*))|_{\lambda=0} \\ &= lim_{\lambda\downarrow 0}\frac{f(x^*+\lambda(z-x^*)) - f(x^*)}{\lambda} \\ &\leq lim_{\lambda\downarrow 0}\frac{\lambda f(z) + (1-\lambda)f(x^*)}{\lambda} \\ &= f(z) - f(x^*) < 0.\end{aligned} \tag{2.7}$$

Therefore, $\nabla f(x^*) \neq 0$, and so x^* is not a stationary point.

2.4 Constrained Optimization

Constrained optimization problems can be defined using an objective function and a set of constraints. The standard form of the constrained optimization problem is as follows:

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \\ & g_j(x) \leq 0 \\ & x \in \Re^n. \end{aligned} \tag{2.8}$$

where $f : \Re^n \to \Re, h_i : \Re^n \to \Re^l, g_j : \Re^n \to \Re^m$ are given continuously differentiable functions. We call f the objective function, while h is the equality constraint and g is the inequality constraint. The components of h and g are denoted by $h_1, \ldots, h_l$ and $g_1, \ldots, g_m$ respectively.
We define the feasible set (or feasible region) X to be the set of points x that satisfy the constraints; that is,

$$X = \{x | h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, l, \quad g_j(x) \leq 0 \quad \text{for} \quad j = 1, \ldots, m\}$$

So that most of the time we can rewrite 2.8 more compactly as

$$\min_{x \in X} f(x) \tag{2.9}$$

The points belonging to the feasible region are called *feasible points*.
A vector $d \in \Re^n$ is a *feasible direction* at $x \in X$ if $d \neq 0$ and $x + \alpha d \in X$ for some sufficiently small $\alpha > 0$. At a feasible point x, the inequality constraint is said to be active if $g_j(x) = 0$ and inactive if the strict inequality $g_j(x) < 0$ is satisfied.

The set $A(x) = \{i : g_j(x) = 0; j = 1, \ldots, m\}$ denotes the index set of the active (binding) inequality constraints at x.
Note that: if the set X is convex and the objective function f is convex, then 2.8 is called a convex optimization problem.

Definition 2.4.1. *Definitions of the different types of local minimizing solutions are simple extensions of the corresponding definitions for the unconstrained case, except that now we restrict consideration to the feasible points in the neighbourhood of x^*.*

i. A vector x^* is a local solution of the problem 2.9 if $x^* \in X$ and there is a neighbourhood η of x^* such that $f(x) \geq f(x^*)$ for $x \in \eta \cap X$.

ii. A vector x^* is a strict local solution (also called strong local solution) if $x^* \in X$ and there is a neighbourhood η of x^* such that $f(x) > f(x^*)$ for $x \in \eta \cap X$ with $x \neq x^*$.

iii. A vector x^* is an isolated local solution if $x^* \in X$ and there is a neighbourhood η of x^* such that x^* is the only local solution in $x \in \eta \cap X$.

Note that isolated local solutions are strict, but that the reverse is not true.

Theorem 2.4.1. *Assume that X is a convex set and for some $\epsilon > 0$ and $x^* \in X$, $f \in C^1$ over $S(x^*; \epsilon)$. Then if x^* is a local minimizer, then*

$$\nabla f(x^*)^t d \geq 0, \tag{2.10}$$

where $d = x - x^*$, feasible direction, for all $x \in X$.
If in addition f is convex over X and 2.10 holds, then x^* is a global minimizer.

Proof 2.4.1. *Let d be a feasible direction. If $\nabla f(x^*)^t d < 0$ (i.e., if d is a descent direction at x^*), then $f(x^* + \alpha d) < f(x^*)$ for all sufficiently small $\alpha > 0$ (i.e., all $\alpha \in (0, \bar{\alpha})$ for some $\bar{\alpha} > 0$). This is a contradiction since x^* is a local minimizer.*

Theorem 2.4.2 (Weierstrass's Theorem)**.** *Let X be a non empty, compact set in $\Re^n$, and let $f : X \to \Re$ be continuous on X. Then the problem min$\{f(x) : x \in X\}$ attains its minimum; that is, there is a minimizing point to this problem.*

2.4.1 Optimality Conditions for Equality Constrained Optimization

Consider the equality constrained problem

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array} \tag{2.11}$$

Where $f : \Re^n \to \Re, h : \Re^n \to \Re^l$ are given functions, and $h_1, \ldots, h_l$ are components of h.

Definition 2.4.2 (Regular Point). *Let x^* be a vector such that $h(x^*) = 0$ and, for some $\epsilon > 0, h \in C^1$ on $S(x^*; \epsilon)$. We say that x^* is a regular point if the gradients $\nabla h_1(x^*), \dots, \nabla h_l(x^*)$ are linearly independent.*

Definition 2.4.3 (Lagrangian Function). *The Lagrangian function $\mathcal{L} : \Re^{n+l} \to \Re$ for the problem 2.11 is defined by*

$$\mathcal{L}(x, \lambda) = f(x) + < \lambda, h(x) >$$

where $\lambda = (\lambda_1, \dots, \lambda_l)$ is the Lagrange multiplier of h.

Proposition 2.4.1 (Karush-Kuhn-Tucker (KKT) Necessary Conditions). *Let x^* be a local minimum for 2.11 and assume that, for some $\epsilon > 0, f \in C^1, h \in C^1$ on $S(x^*; \epsilon)$, and x^* is a regular point. Then, there exists unique vector $\lambda^* \in \Re^l$ such that*

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0. \tag{2.12}$$

If in addition, $f \in C^2, h \in C^2$ on $S(x^*; \epsilon)$, then for all $z \in \Re^n$ satisfying $\nabla h(x^*)^t z = 0$, we have

$$z^t \nabla_{xx} \mathcal{L}(x^*, \lambda^*) z \geq 0. \tag{2.13}$$

Theorem 2.4.3 (KKT Sufficient Conditions). *Let $x^* \in \Re^n$ such that $h(x^*) = 0$, and, for some $\epsilon > 0, f \in C^2, h \in C^2$ on $S(x^*; \epsilon)$. Assume that there exists vector $\lambda^* \in \Re^m$ such that*

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \tag{2.14}$$

and for every $z \neq 0$ satisfying $\nabla h(x^*)^t z = 0$, we have

$$z^t \nabla_{xx} \mathcal{L}(x^*, \lambda^*) z > 0. \tag{2.15}$$

Then x^* is a strict local minimizer for 2.11

Remark: A point is said to be KKT point if it satisfies all the KKT necessary conditions.

2.4.2 Optimality Conditions for General Constrained Optimization

Consider the constrained problem involving both equality and inequality constraints

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \\ & g_j(x) \leq 0 \\ & x \in \Re^n. \end{aligned} \tag{2.16}$$

where $f : \Re^n \to \Re, h_i : \Re^n \to \Re^l, g_j : \Re^n \to \Re^m$ are given functions and $l \leq n$. The components of h_i and g_j are denoted by $h_1, \dots, h_l$ and $g_1, \dots, g_m$ respectively. Let $A(x) = \{j : g_j(x) = 0; j = 1, \dots, m\}$ be the index set of the active (binding) inequality constraints at x.

Definition 2.4.4 (Regular Point). *Let x^* be a vector such that $h(x^*) = 0, g(x^*) \leq 0$ and, for some $\epsilon > 0, h \in C^1$ and $g \in C^1$ on $S(x^*; \epsilon)$. We say that x^* is a regular point if the gradients $\nabla h_1(x^*), \ldots, \nabla h_l(x^*)$ and $\nabla g_i(x^*)$ for $i \in A(x^*)$, are linearly independent.*

Definition 2.4.5 (Lagrangian Function). *The Lagrangian function $\mathcal{L} : \Re^{n+l+m} \rightarrow \Re$ for the problem 2.16 is defined by*

$$\mathcal{L}(x, \lambda, \mu) = f(x) + < \lambda, h(x) > + < \mu, g(x) >,$$

where $\lambda = (\lambda_1, \ldots, \lambda_l)$ and $\mu = (\mu_1, \ldots, \mu_m)$ are the Lagrange multipliers of h and g respectively.

Theorem 2.4.4 (KKT Necessary Conditions). *Let x^* be a local minimum for 2.16 and assume that, for some $\epsilon > 0, f \in C^1, h \in C^1, g \in C^1$ on $S(x^*; \epsilon)$, and x^* is a regular point. Then, there exist unique vectors $\lambda^* \in \Re^l$ and $\mu^* \in \Re^m$ such that*

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0 \tag{2.17}$$

$$\mu_i{}^* \geq 0, \quad \mu_i{}^* g_i(x^*) = 0, \quad \text{for all} \quad i = 1, \ldots, m. \tag{2.18}$$

The conditions $\mu_i{}^* g_i(x^*) = 0$, for all $i = 1, \ldots, m$, are complementarity conditions; they imply that either constraint $g_i(x^*)$ is active or $\mu_i{}^* = 0$, or possibly both. In particular, the Lagrange multipliers corresponding to inactive inequality constraints are zero.
If in addition, $f \in C^2, h \in C^2, g \in C^2$ on $S(x^*; \epsilon)$, then for all $z \in \Re^n$ satisfying $\nabla h(x^*)^t z = 0$ and $\nabla g_i(x^*)^t z = 0, i \in A(x^*)$, we have

$$z^t \nabla_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) z \geq 0. \tag{2.19}$$

Theorem 2.4.5 (KKT Sufficient Conditions). *Let $x^* \in \Re^n$ such that $h(x^*) = 0, g(x^*) \leq 0$, and, for some $\epsilon > 0, f \in C^2, h \in C^2, g \in C^2$ on $S(x^*; \epsilon)$. Assume that there exist vectors $\lambda^* \in \Re^l$ and $\mu^* \in \Re^m$ such that*

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0 \tag{2.20}$$

$$\mu_i{}^* \geq 0, \quad \mu_i{}^* g_i(x^*) = 0, \quad \text{for all} \quad i = 1, \ldots, m \tag{2.21}$$

and for every $z \neq 0$ satisfying $\nabla h(x^*)^t z = 0, \nabla g_i(x^*) \leq 0$ for all $i \in A(x^*)$, and $\nabla g(x^*)^t z = 0$, for all $i \in A(x^*)$ with $\mu_i^* > 0$, we have

$$z^t \nabla_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) z > 0. \tag{2.22}$$

Then x^* is a strict local minimizer for 2.16

Proposition 2.4.2. *Assume that f and $g_1, \ldots, g_m$ are convex and continuously differentiable functions on $\Re^n$. Let $x^* \in \Re^n$ and $\mu^* \in \Re^m$ satisfy*

$$\nabla f(x^*) + \nabla g(x^*) \mu^* = 0,$$

$$g(x^*) \leq 0, \quad \mu_j^* \geq 0, \quad \mu_j^* g_j(x^*) = 0, \quad j = 1, 2, \ldots, l.$$

Then x^* is a global minimizer.

2.5 Methods to Solve Unconstrained Optimization Problems

There are many approaches which help us to solve un constrained non linear optimization problems some of them are:
Line search method,Newton method,Steep decent method,Lagrange method,KKT conditions,Penalty methods etc.We will see Penalty method in detail on the last chapter and some of the methods are introduced below in general sense:

i. **Line Search Method**: A general approach to find an optimizer is to apply line search methods. They operate iteratively, which means they start at an initial guess x^0 for the optimizer and at each iteration compute a step, which should lead to a better solution. The algorithm terminates if an optimizer x^* satisfies certain optimality conditions. The computation of a step consists of two parts: first obtaining a search direction d_k and second determining a step length α. This results in the formula for the next iterate:

$$x^{k+1} = x^k + \alpha d_k,$$

where $k = 0, 1, \ldots$

Line search methods can be applied to unconstrained and constrained optimization problems but there are different strategies that realize the line search approach. If the objective function is convex, an appropriate line search method will find the global solution. If it is not convex, it will probably just find the local minimum next to the initial guess.

ii. **Steepest Descent Method**: This method is quite historical in the sense that it was introduced in the middle of the 19^{th} century by Cauchy. The idea of the method is to decrease the function value as much as possible in order to reach the minimum early. Thus, the question is in which direction the function decreases most. The first order Taylor expansion of f near point x in the direction d is

$$f(x + d) \approx f(x) + \nabla^t f(x)d.$$

We search for the direction

$$\min_{d \in \Re^n} \frac{\nabla^t f(x)d}{||d||},$$

which is for the Euclidean norm the negative gradient, i.e., $d = -\nabla f(x)$. That is why this method is also called gradient method. This method is one of the simplest but also one of the slowest method.

Therefore, the general iteration form of steepest descent method is given by

$$x^{k+1} = x^k - \lambda \nabla f(x^k),$$

where λ is the step length for the direction vector $-\nabla f(x^k)$.

iii. **Conjugate Gradient Method**: This class of methods can be viewed as a modification of the steepest descent method, where in order to avoid the zigzagging effect, at each iteration the direction is modified by a combination of the earlier directions:

$$d_k = -\nabla f(x^k) + \beta_k d_{k-1}$$

These corrections ensure that $d_1, \ldots, d_n$ are so-called conjugate directions. This means that there exist a matrix A such that $d_i^t A d_j = 0$, for all $i \neq j$. For instance, the coordinate directions (the unit vectors) are conjugate. Just take A as the unit matrix. The underlying idea is that A is the inverse of the Hessian. One can derive that using exact line search the optimum is reached in at most n steps for quadratic functions.

Having the direction d_k, the next iterate is calculated in the usual way

$$x^{k+1} = x^k + \lambda d_k$$

where λ is the optimal step length $arg \min_\mu f(x^k + \mu d_k)$, or its approximation. The parameter β_k can be calculated using Fletcher and Reeves formula as:

$$\beta_k = \frac{||f(x^k)||^2}{||f(x^{k-1})||^2}.$$

iv. **Newton Method**: This method is the most complex and also the fastest of the gradient methods. A problem of this method is, that it is equally attracted by all points where the gradient is zero, which can be minima, maxima and saddle points. So it is necessary that the function is locally convex (that means the Hessian matrix has to be positive definite) in order to guarantee that the computed direction is a descent direction.

Suppose we want to solve:

$$\begin{array}{cc} \text{Minimize} & f(x) \\ x \in R^n & \end{array} \tag{2.23}$$

At $x = a$, $f(x)$ can be approximated by:
$f(x) \approx h(x) := f(a) + \nabla f(a)^T (x-a) + \frac{1}{2}(x-a)^t H(a)(x-a)$,
Which is the quadratic Taylor expansion of $f(x) at x = a$.
Here $\nabla f(x)$ is the gradient of $f(x)$ and $H(x)$ is the Hessian of $f(x)$.
Notice that $h(x)$ is a quadratic function,which is minimized by solving $\nabla h(x) = 0$ Since the gradient of $h(x)$ is:
$\nabla h(x)$=$\nabla f(a) + H(a)(x-a) = 0$, we therefore are motivated to solve:

$\nabla f(a) + H(a)(x-a) = 0$, which yields
$x - a = -H(a)^{-1} \nabla f(a)$.
The direction $-H(a)^{-1} \nabla f(a)$ is called the Newton direction, or the Newton step at $x = a$.
This leads to the following algorithm for solving (2.23)
Algorithm for Newton's Method:

1. [**Initialization Step**] Give x^0, set $k \leftarrow 0$
2. [**Iterative one**] $d^k = -H(x^k)^{-1}\nabla f(x^k)$, If$d^k = 0$, then stop.
3. [**Iterative two**] Choose step-size $\alpha^k = 1$.
4. [**Stopping Criterion**] Set $x^{k+1} \leftarrow x^k + \alpha^k d^k, k \leftarrow k+1$. Go to Step1.

Note the following:

- The method assumes $H(x^k)$ is nonsingular at each iteration.
- There is no guarantee that $f(x^{k+1}) \leq f(x^k)$.
- Step 2 could be augmented by a line-search of $f(x^k + \alpha d^k)$ to find an optimal value of the step-size parameter α.

Example 2.5.1.

$$\begin{aligned} &Minimize \quad f(x) = 7x - lnx \quad for\ x > 0 \\ &x \in R^n \end{aligned} \tag{2.24}$$

Let $f(x) = 7x - lnx$. *Then* $\nabla f(x) = f'(x) = 7 - \frac{1}{x}$ *and* $H(x) = f''(x) = \frac{1}{x^2}$.*It is not hard to check that* $x^* = \frac{1}{7} = 0.142857143$ *is the unique global minimum. The Newton direction at* x *is* $d = -H(x)^{-1}\nabla f(x) = -\frac{f'(x)}{f''(x)} = -x^2(7 - \frac{1}{x}) = x - 7x^2$.
Newtons method will generate the sequence of iterates x^k *satisfying:*
$x^{k+1} = x^k + (x^k - 7(x^k)^2) = 2x^k - 7(x^k)^2$.
This is the sequences generated by the given method for different starting points.

k	x^k	x^k	x^k	x^k
0	1.0	0	0.1	0.01
1	-5.0	0	0.13	0.0193
2	-185.0	0	0.1417	0.03599257
3	$-239,945.0$	0	0.14284777	0.062916884
4	-4.0302×10^{11}	0	0.142857142	0.098124028
5	-1.1370×10^{24}	0	0.142857143	0.128849782
6	-9.0486×10^{48}	0	0.142857143	0.1414837
7	-5.7314×10^{98}	0	0.142857143	0.142843938
8	$-\infty$	0	0.142857143	0.142857142
9	$-\infty$	0	0.142857143	0.142857143
10	$-\infty$	0	0.142857143	0.142857143

Table 2.1: Newton's Iteration for Example 2.5.1

Hence,the range of quadratic convergence for Newtons method for this function happens to be $x \in (0.0, 0.2857143)$.

Chapter 3

Interior and Exterior Penalty Methods

In this section, we are concerned with exploring the computational properties of penalty function methods. We present and prove an important result that justifies using penalty function methods as a means for solving optimization problems. We also discuss some computational difficulties associated with these methods and present some techniques that should be used to overcome such difficulties. Using the special structure of the penalty function, a special purpose one-dimensional search procedure algorithm is developed. The procedure is based on Powells method for unconstrained minimization technique together with bracketing and golden section for one dimensional search. When solving a general nonlinear programming problem in which the constraints cannot easily be eliminated, it is necessary to balance the aims of reducing the objective function and staying inside or close to the feasible region, in order to induce global convergence (that is convergence to a local solution from any initial approximation). This inevitably leads to the idea of a penalty function, which is a combination of the same constraints that enables the objective function to be minimized whilst controlling constraint violations (or near constraint violations) by penalizing them. The philosophy of penalty methods is simple; you give a fine for violating the constraints and obtain approximate solutions to your original problem by balancing the objective function and a penalty term involving the constraints. By increasing the penalty, the approximate solution is forced to approach the feasible domain and hopefully, the solution of the original constrained problem. Early penalty functions were smooth so as to enable efficient techniques for smooth unconstrained optimization to be used.

3.1 The Concept Of Penalty Functions

Consider the problem

$$\text{Minimize} \quad f(x) : x \in S \tag{3.1}$$

Where $f(x)$ is continuous function on $\Re^n$ and S is a constraint set in $\Re^n$. In most applications S is defined explicitly by a number of functional constraints as illustrated on (3.2)and (3.4), Consider the following problem with single constraint $h(x) = 0$:

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \\ x \in \Re^n & \end{aligned} \tag{3.2}$$

Suppose this problem is replaced by the following unconstrained problem, where $\mu > 0$ is a large number:

$$\begin{aligned} \text{Minimize} \quad & f(x) + \mu h^2(x) \\ \text{subject to} \quad & x \in \Re^n \end{aligned} \tag{3.3}$$

We can intuitively see that an optimal solution to the above problem must have $h^2(x)$ close to zero, otherwise a large penalty term $\mu h^2(x)$ will be incurred and hence $f(x)$ $+\mu h^2(x)$ approaches to infinity which makes it difficult to minimize the unconstrained problem (Bazara,2006). Now consider the following problem with single inequality constraint $g(x) \leq 0$:

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) \leq 0 \\ & x \in \Re^n \end{aligned} \tag{3.4}$$

It is clear that the form $f(x) + \mu g^2(x)$ is not appropriate,since a penalty will be incurred when $g(x) < 0$ and $g(x) > 0$;hence a penalty is added to the objective function whether x is inside or outside the feasible region. Needless to say, a penalty is desired only if the point is not feasible, that is, if$g(x) > 0$. A suitable unconstrained problem is therefore given by:

$$\begin{aligned} \text{Minimize} \quad & f(x) + \mu(Max\{0, g(x)\})^2 \\ \text{subject to} \quad & x \in \Re^n \end{aligned} \tag{3.5}$$

Note that if $g(x) \leq 0$, then $Maximum\{0, g(x)\}^2 = 0$, and no penalty is incurred on the other hand, if $g(x) > 0$, then $Maximum\{0, g(x)\}^2 > 0$, and the penalty term $\mu g(x)$ is realized. However, it is observe that at points x where $g(x) = 0$, the forgoing objective function might not be differentiable, even though g is differentiable.
The idea of a penalty function method is to replace the type of problem (3.2),(3.4) or (3.7) by an unconstrained approximation of the form:

$$\begin{aligned} \text{Minimize} \quad & f(x) + cP(x) \\ & x \in \Re^n \end{aligned} \tag{3.6}$$

Where c is a positive constant and P is a penalty function on R^n defined as follow:

Definition 3.1.1. *A function $\mathcal{P} : \Re^n \to \Re$ is called a penalty function if P satisfies:*

i. *$P(x)$ is continuous on R^n .*

ii. *$P(x) = 0$ if and only if $x \in S$*

iii. *$P(x) > 0$ for all $x \in \Re^n | S$*

An often-used class of penalty functions for optimization problems with only inequality constraints is:

$P(x) = \sum_{i=1}^{m} (Max\{0, g(x)\})^p$, where p is a positive integer

We refer to the function $f(x) + \mu P(x)$ as an auxiliary function.
Consider the general constrained optimization problem

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, m, \\ & g_j(x) \leq 0 \quad \text{for} \quad j = 1, \ldots, l, \\ & x \in \Re^n \end{aligned} \tag{3.7}$$

Whose feasible region denoted by
$F{:=}\{x \in \Re^n | h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, l, g_j(x) \leq 0 \quad \text{for} \quad j = 1, \ldots, m, \}$.
We can approximate the optimum value of (3.7) by penalty function methods; also known as a sequence of unconstrained minimization technique(SUMT). Penalty function allows to convert a constrained optimization problem into an unconstrained problem(Manfrd H.,2011). There are many types of penalty methods. On this paper we present only about the interior penalty function methods and the exterior penalty function methods, which are branches of transformation methods they are designed to solve a kind of (3.7) problems by transforming constrained optimization problem in to an unconstrained optimization problem (or more commonly, a series of unconstrained optimization problems).On interior penalty function methods, we presume that we have given a point $x*$ that lies in the interior of the feasible region F , and we impose a very large cost on feasible points that lie ever closer to the boundary of feasible region F , thereby creating a "barrier" to exiting the feasible region.In exterior penalty function methods, the feasible region of the constrained problem is expanded from F to all of $\Re^n$, but a large penalty is added to the objective function for points that lie outside of the original feasible region.

3.2 Interior Penalty Function Methods

As described on (3.5) above the general constrained optimization problem is :

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, l, \\ & g_j(x) \leq 0 \quad \text{for} \quad j = 1, \ldots, m, \end{aligned} \tag{3.8}$$

where $f, h_1, \ldots, h_l, g_1, \ldots, g_m$, are continuous functions defined on $\Re^n$.
These methods transform the original constrained problem into unconstrained problem ; however the barrier prevent the current solution from ever leaving the feasible region. These require that the interior of the feasible sets be nonempty. Therefore, they are used with problems having only inequality constraints (there is no interior for equality constraints). Thus,once the unconstrained minimization is started from any feasible x_1, the subsequent points generated will always lie with in the feasible region, since the constraint boundaries act as barriers during the minimization process. This is the reason why the interior penalty method is also known as "barrier" method.

Definition 3.2.1 (Interior penalty Function)**.** *Consider the nonlinear inequality constrained problem*

$$\begin{array}{ll} \textit{Minimize} & f(x) \\ \textit{subject to} & g_i(x) \leq 0 \quad \textit{for} \quad i = 1, \ldots, m, \\ & x \in \Re^n. \end{array} \tag{3.9}$$

Where $f, g_1, \ldots, g_m$ are continuous functions defined on $\Re^n$.
An interior penalty function P is one that is continuous and nonnegative over the interior of $\{x|g(x) \leq 0\}$, i.e., over the set $\{x|g(x) < 0\}$, and approaches ∞ as the boundary is approached from the interior.
Let $\psi(y) \geq 0$ if $y < 0$ and $\lim_{y \to 0^-} \{\psi(y)\} \to \infty$, where ψ is a univariate function that is continuous over $\{y : y < 0\}$ and $y = g_i(x)$. Then

$$P(x) = \Sigma_{i=1}^{m} \psi[g_i(x)].$$

The most commonly used types of interior penalty functions are:

i. Inverse function: $P(x) = -\Sigma_{i=1}^{m} \frac{1}{g_i(x)}$*; for* $\{g_i(x) < 0\}$

ii. Logarithm function: $P(x) = -\Sigma_{i=1}^{m} \log[-g_i(x)]$*; for* $\{g_i(x) < 0\}$

In both cases, note that $\lim_{g_i(x) \to 0^-} P(x) \to \infty$. The auxiliary function is now

$$\phi_{\mu_k}(x) := f(x) + \mu P(x),$$

Where μ is a positive constant.

Ideally, we would like $P(x) = 0$ if $g_i(x) < 0$ and $P(x) \to \infty$ if $g_i(x) \to 0$, so that we never leave the region $\{x|g(x) \leq 0\}$. However, $P(x)$ is now discontinuous. This causes serious computational problems during the unconstrained optimization. Therefore, this ideal construction of P is replaced by the more realistic requirement that P is nonnegative and continuous over the region $\{x|g(x) < 0\}$ and that approaches infinity as the boundary is approached from the interior. Note that the barrier function at infeasible points is not

necessarily defined.
We can write the barrier problem as:

$$\begin{aligned} \text{Minimize} \quad & f(x) + \mu P(x) \\ \text{subject to} \quad & g_i(x) < 0 \quad \text{for} \quad i = 1, \ldots, m, \\ & x \in \Re^n \end{aligned} \tag{3.10}$$

From this we observe that the barrier problem itself is a constrained problem, and indeed the constraint is some what more complicated than in the original problem. The advantage of this problem, however, is that it can be solved by using an unconstrained search technique. To find the solution one starts at an initial interior point and then searches from that point using steepest descent or some other iterative descent method applicable to unconstrained problems. Thus, although the barrier problem is from a formal viewpoint a constrained problem, from a computational viewpoint it is unconstrained.

For instance, for the problem
minimize $f(x) = x$
subject to $g(x) = 5 - x \leq 0$, the barrier function is given by

$$P(x) = \frac{-1}{5 - x},$$

with $x \geq 5$ as a feasible interval for the given constrained problem.

So, the corresponding auxiliary function can be written as $\phi_{\mu_k}(x) = f(x) + \mu P(x) = x - \frac{\mu}{(5-x)}$ and its optimum solution is at

$$\frac{\partial \phi_{\mu_k}}{\partial x} = \frac{5 \pm \sqrt{25 - 4\mu}}{2}$$

Hence

$$x^* = \frac{5 \pm \sqrt{25 - 4\mu}}{2}$$

The negative value lead us to infeasibility, hence the optimum solution $x^* = \frac{5+\sqrt{25-4\mu}}{2}$ and as $\mu \to 0, x^* \to 5$.

3.2.1 Algorithmic Scheme For Interior Penalty Function Methods

[**Algorithm**]

1. [**Initialization Step**] Select a growth parameter $\gamma > 1$, a stopping parameter (tolerance) $\epsilon > 0$ and an initial value of the $\mu_1 > 0$. Choose an initial feasible solution say x^1 with $g(x^1) < 0$ and formulate the objective function $\phi_{\mu_k}(x)$. Set $k = 1$.

2. [**Iterative Step**] Starting from x^k use an unconstrained search technique to find the point that minimizes $\phi_{\mu_k}(x)$ and call it x^{k+1}, the new starting point.

3. [**Stopping Criterion**] If $||x^{k+1} - x^k|| < \epsilon$, stop with x^{k+1} an estimate of the optimal solution otherwise, put $\mu_{k+1} = \gamma\mu_k$, and formulate the new $\phi_{\mu_{k+1}}(x)$ and put $k = k+1$ and return to the iterative step.

Consider the following problem again:

Example 3.2.1.

$$\begin{aligned} &\textit{Minimize} \quad f(x) = x_1^2 + 2x_2^2 \\ &\textit{subject to} \quad g(x) = 1 - x_1 - x_2 \leq 0 \\ &\quad x \in \Re^2. \end{aligned} \tag{3.11}$$

Solution 3.2.1. *Define the barrier function*

$$P(x) = -\log[-g(x)] = -\log[x_1 + x_2 - 1]$$

The unconstrained problem is

$$\textit{Minimize}\phi_{\mu_k} = \quad x_1^2 + 2x_2^2 - \mu_k \log[x_1 + x_2 - 1].$$

The necessary conditions for the optimal solution $\nabla f(x) = 0$ *yield the following:*

$$\frac{\partial \phi_{\mu_k}}{\partial x_1} = 2x_1 - \frac{\mu_k}{(x_1 + x_2 - 1)} = 0$$

$$\frac{\partial \phi_{\mu_k}}{\partial x_2} = 4x_2 - \frac{\mu_k}{(x_1 + x_2 - 1)} = 0$$

and we get,

$$x_1 = \frac{1 \pm \sqrt{1 + 3\mu_k}}{3}$$

and

$$x_2 = \frac{1 \pm \sqrt{1 + 3\mu_k}}{6}$$

Since the negative signs lead to infeasibility, we have

$$x_1 = \frac{1 + \sqrt{1 + 3\mu_k}}{3} \quad \textit{and} \quad x_2 = \frac{1 + \sqrt{1 + 3\mu_k}}{6}$$

Starting with $\mu_1 = 1, \gamma = 0.1$ *and* $x^1 = (1, 0.5)$ *and using a tolerance of* 0.005 *(say), we have the following: Thus, this solution approach to the exact optimal solution* $x^* = (2/3, 1/3)$.

From the table we observe that every points at each iteration is in the interior of the feasible region, and the final solution itself remains in the interior.

k	μ_k	x^k	$g(x^k)$	$P(x^k)$	$\mu_k P(x^k)$	$f(x^k)$	$\phi_{\mu_k}(x^k)$
1	1	$(1.000, 0.5000)$	-0.5000	0.30103	0.3010	1.5	1.80103
2	0.1000	$(0.714, 0.357)$	-0.071	1.1487	0.1149	0.765	0.8788
3	0.0100	$(0.672, 0.336)$	-0.008	2.0969	0.02097	0.677	0.6979
4	0.0010	$(0.6672, 0.3336)$	-0.0008	3.0969	0.003097	0.668	0.6708
5	0.0001	$(0.6666, 0.3333)$	-0.0001	4.6576	0.000466	0.6667	0.6672
6	0.00001	$(0.666671, 0.333335)$	-0.0000075	5.6576	0.0000566	0.666667	0.6667

Table 3.1: Barrier Iteration for the Given Example

3.2.2 Convergence Of Interior Penalty Function Methods

We start with some μ_1 and generate a sequence of points. Let the sequence $\{\mu_k\}$ satisfy $\mu_{k+1} < \mu_k$ and $\mu_k \to 0$ as $k \to \infty$. Let x^k denote the solution to $\phi_{\mu_k}(x)$, and x^* be an optimal solution to problem 3.9. Then the following Lemma presents some basic properties of barrier methods.

Lemma 3.2.1.

i. $\phi_{\mu_k}(x^k) \geq \phi_{\mu_{k+1}}(x^{k+1})$

ii. $P(x^k) \leq P(x^{k+1})$

iii. $f(x^k) \geq f(x^{k+1})$

iv. $f(x^*) \leq f(x^k) \leq \phi_{\mu_k}(x^k)$

The above lemma is called Barrier Lemma.

Theorem 3.2.1 (Convergence Theorem)**.** *Suppose $f(x), g(x)$, and $P(x)$ are continuous functions. Let $\{x^k\}$, for $k = 1, 2, \ldots$, be a sequence of solutions of $\phi_{\mu_k}(x)$. Suppose there exists an optimal solution x^* of 3.9 for which $\eta \cap \{x|g(x) < 0\} \neq \emptyset$, where η is a neighbourhood of x^*. Then any limit point $\bar{x}$ of $\{x^k\}$ solves 3.9.*
Furthermore,$\mu_k P(x) \to 0$ as $\mu_k \to 0$.

Proof 3.2.1. *Let $\bar{x}$ be any limit point of the sequence $\{x^k\}$. From the continuity of $f(x)$ and $g(x)$, $\lim_{k\to\infty} f(x^k) = f(\bar{x})$ and $\lim_{k\to\infty} g(x^k) = g(\bar{x}) \leq 0$. Thus $\bar{x}$ is a feasible point for 3.9.*

Given any $\epsilon > 0$, there exists $\hat{x}$ such that $g(\hat{x}) < 0$ and $f(\hat{x}) \leq f(x^*) + \epsilon$. For each k,

$$f(x^*) + \epsilon + \mu_k P(\hat{x}) \geq f(\hat{x}) + \mu_k P(\hat{x}) \geq \phi_{\mu_k}(x^k).$$

Therefore, for sufficiently large k,

$$f(x^*) + 2\epsilon \geq \phi_{\mu_k}(x^k),$$

and since $\phi_{\mu_k}(x^k) \geq f(x^*)$ from (iv) of Lemma(3.2.1), then

$$f(x^*) + 2\epsilon \geq \lim_{k\to\infty} \phi_{\mu_k}(x^k) \geq f(x^*).$$

This implies that

$$\lim_{k\to\infty} \phi_{\mu_k}(x^k) = f(\bar{x}) + \lim_{k\to\infty} \mu_k P(x^k) = f(x^*).$$

We also have

$$f(x^*) \le f(x^k) \le f(x^k) + \mu_k P(x^k) = \phi_{\mu_k}(x^k).$$

Taking limits we obtain

$$f(x^*) \le f(\bar{x}) \le f(x^*).$$

From this, we have

$$f(x^*) = f(\bar{x}).$$

Hence, $\bar{x}$ is the optimal solution of the original nonlinear inequality constrained problem, 3.9. Furthermore, from

$$f(\bar{x}) + \lim_{k\to\infty} \mu_k P(x^k) = f(x^*),$$

we have

$$\lim_{k\to\infty} \mu_k P(x^k) = f(x^*) - f(\bar{x}) = 0.$$

Therefore, as $k \to \infty$, i.e., $\mu_k \to 0$, the function $\mu_k P(x^k) \to 0$ for each k. This proves the second statement of the Theorem.

3.3 Exterior Penalty Function Methods

Methods using exterior penalty functions transform a constrained problem into a single unconstrained problem or into a sequence of unconstrained problems. In these methods the constraints are placed into the objective function via a penalty parameter in a way that penalizes any violation of the constraints. It generates a sequence of infeasible points whose limit is the approximate solution of the original constrained problem(W.SUN,2006). A suitable penalty function must incur a positive penalty for infeasible points, and no penalty for feasible points. As the penalty parameter, which is used to control the impact of the additional term, takes higher values, the approximation to the solution of the original constrained problem becomes increasingly accurate.

Definition 3.3.1 (Exterior Penalty Function)**.** *Consider the optimization constrained problem*

$$\begin{aligned} \textit{Minimize} \quad & f(x) \\ \textit{subject to} \quad & h_i(x) = 0 \quad \textit{for} \quad i = 1, \ldots, m; \\ & g_j(x) \le 0 \quad \textit{for} \quad j = 1, \ldots, l; \\ & x \in X, \end{aligned} \tag{3.12}$$

where, $f, h_1, \ldots, h_l, g_1, \ldots, g_m$ are continuous functions defined on $\Re^n$, and X is a non empty set in $\Re^n$. Then, the unconstrained problem (or the penalized objective function)

$$\psi_\mu(x) := f(x, \mu) = f(x) + \mu\alpha(x) \quad \textit{for} \quad x \in \Re^n$$

is called the auxiliary function, where the penalty function $\alpha(x)$ *is defined by*

$$\alpha(x) = \Sigma_{i=1}^{m}[\max\{0, g_j(x)\}]^p + \Sigma_{i=1}^{l}|h_i(x)|^p \tag{3.13}$$

for positive integer p *and non-negative penalty parameter* μ*. If* x *is a point in the feasible region, then* $\alpha(x) = 0$ *and hence no penalty incurred. Therefore, a penalty is desired only if the point* x *is not feasible, i.e.,for a point* x *such that* $g_j(x) > 0$ *for some* $j = 1, \ldots, m$ *or* $h_i(x) \neq 0$ *for some* $i = 1, \ldots, l$*.*

Example 3.3.1.

$$\textit{minimize} \quad f(x) = x \quad \textit{such that} \quad g(x) = 5 - x \leq 0, \quad x \in \Re.$$

Note that the minimizer is $x^* = 5$*.*

Solution 3.3.1. *Let* $\alpha(x) = [\max\{g(x), 0\}]^2$ *i.e.,* $\alpha(x) = 0, \quad$ *for* $\quad x \geq 5$ *or* $\alpha(x) = (5 - x)^2, \quad$ *for* $\quad x < 5$*.*
If $\alpha(x) = 0$*, then the optimal solution to minimize* $f_\mu(x) = f(x) + \mu\alpha(x)$ *is at* $x^* = -\infty$ *and this is infeasible. So,*

$$f_\mu(x) = x + \mu(5 - x)^2.$$

Since this function is quadratic form, we can evaluate the minimizer using first derivative, $f'_\mu(x) = 1 - 2\mu(5 - x) = 0$*. This implies that* $x = 5 - \frac{1}{2\mu}$*, which converges to* $x^* = 5$ *as* $\mu \to \infty$*. Therefore, the minimizer of the original problem is* $x^* = 5$*.*

3.3.1 Algorithmic Scheme For Exterior Penalty Function Methods

Usually, we solve a sequence of problems with successively increasing μ values, i.e., for $0 < \mu_k < \mu_{k+1}$; the *optimal point* x^k for the penalized objective function $f_{\mu_k}(x)$, the sub-problem at k^{th} iteration, becomes the starting point for the next problem, where $k = 1, 2, ...$ To obtain the optimum x^k, we assumed that the penalized function has a solution for all positive values of μ_k.

[**Algorithm**]

1. [**Initialization Step**] Select a growth parameter $\gamma > 1$, a stopping parameter (tolerance) $\epsilon > 0$ and an initial value of the penalty parameter μ_1. Choose a starting point x^1 that violates at least one constraint and formulate the penalized objective function $f_{\mu_k}(x)$. Set $k = 1$.

2. [**Iterative Step**] Starting from x_k use an unconstrained search technique to find the point that minimizes $f_{\mu_k}(x)$ and call it x^{k+1}, the new starting point.

3. [**Stopping Criterion**] If $||x^{k+1} - x^k|| < \epsilon$ or the difference between two successive objective functions values is smaller than ϵ (i.e.), $|f(x^{k+1}) - f(x^k)| < \epsilon$, stop with x^k an estimate of the optimal solution; otherwise, put $\mu_{k+1} \leftarrow \gamma\mu_k$, and formulate the new $f_{\mu_{k+1}}(x)$ and put $k \leftarrow k+1$ and return to the iterative step.

Example 3.3.2.

$$\textit{Minimize} \quad f(x) = x_1^2 + 2x_2^2 \quad \textit{Subject to} \quad g(x) = 1 - x_1 - x_2 \leq 0; \quad x \in \Re^2.$$

Solution 3.3.2. *Define the penalty function* $\alpha(x) = [\max\{g(x), 0\}]^2$. *Thus*

$$\alpha(x) = 0, \quad \textit{for} \quad g(x) \leq 0 \quad \textit{and} \quad \alpha(x) = (1 - x_1 - x_2)^2, \quad \textit{for} \quad g(x) > 0$$

Then the unconstrained problem is

$$f_{\mu_k}(x) = x_1^2 + 2x_2^2 + \mu_k\alpha(x)$$

If $\alpha(x) = 0$, *then the optimal solution to minimize* $f_{\mu_k}(x) = x_1^2 + 2x_2^2 + \mu_k\alpha(x)$ *is at* $x^* = (0,0)$ *and this is infeasible. So*

$$f_{\mu_k}(x) = x_1^2 + 2x_2^2 + \mu_k(1 - x_1 - x_2)^2$$

Now, by using the necessary condition for optimality (i.e., $\nabla f_{\mu_k}(x) = 0$*) we have the following:*

$$\frac{\partial f_{\mu_k}}{\partial x_1} = 2x_1 - 2\mu_k(1 - x_1 - x_2) = 0,$$
$$\frac{\partial f_{\mu_k}}{\partial x_2} = 4x_2 - 2\mu_k(1 - x_1 - x_2) = 0.$$

This implies that

$$1 - x_1 - x_2 = \frac{x_1}{\mu_k}$$
$$1 - x_1 - x_2 = \frac{2x_2}{\mu_k}$$

From these equations, we have $x_1 = 2x_2$.
Thus $x^k = (\frac{2\mu_k}{2+3\mu_k}, \frac{\mu_k}{2+3\mu_k})$.
Starting with $\mu_1 = 1, \quad \gamma = 10 \quad$ *and* $\quad x^1 = (0,0)$ *and using a tolerance* 0.0001 *(say), we have the following on the table below:*

Thus the optimal solution is $x^* = (0.6667, 0.3333)$, *which is approximately equal to the exact optimal solution* $x^* = (2/3, 1/3)$, *with optimal value* $f(x^*) = 0.66666$.

From table 3.2 we observe that the solution reached by the penalty function method and all subsequent points are infeasible. Therefore, in applications where feasibility is strictly required, penalty methods can not be used. In such cases barrier function (interior) methods are appropriate.

k	μ_k	x^k	$g(x^k)$	$\alpha(x^k)$	$\mu_k\alpha(x^k)$	$f(x^k)$	$f_{\mu_k}(x^k)$
1	0.1	$(0.087, 0.043)$	0.87	0.7569	0.0757	0.0113	0.0869
2	1	$(0.4000, 0.2000)$	0.40	0.16	.16000	0.24000	0.4
3	10	$(0.6250, 0.3125)$	0.0625	0.0039	0.03906	0.58594	0.625
4	100	$(0.6623, 0.3311)$	0.0067	0.000044	0.004486	0.65787	0.6623
5	1000	$(0.6662, 0.3331)$	0.001	0.00000049	0.000444	0.66578	0.6662
6	10000	$(0.6666, 0.3333)$	0.0001	0.00000001	0.000044	0.66658	0.66663

Table 3.2: Penalty Iteration for Example 3.3.2

3.3.2 Convergence Of Exterior Penalty Function Methods

Consider a sequence of values $\{\mu_k\}$ with $\mu_k \uparrow \infty$ as $k \to \infty$, and let x^k be the minimizer of

$$f_{\mu_k}(x) = f(x) + \mu_k \alpha(x) \quad \text{for each} \quad k.$$

Suppose that x^* denotes any optimal solution of the original constrained problem.
The following Lemma presents some basic properties of Exterior penalty function methods:

Lemma 3.3.1. *Suppose that $f, g_1, \ldots, g_m, h_1, \ldots, h_l$ are continuous functions on $\Re^n$, and let X is non-empty set in $\Re^n$. Let α be a continuous function on $\Re^n$ given by 3.13, and suppose that for each μ_k, there is a minimizer $x^k \in X$ of $f_{\mu_k}(x) = f(x) + \mu_k\alpha(x)$. Then the following properties hold true for $0 < \mu_k < \mu_{k+1}$.*

i. $f_{\mu_k}(x^k) \leq f_{\mu_{k+1}}(x^{k+1})$

ii. $\alpha(x^k) \geq \alpha(x^{k+1})$

iii. $f(x^k) \leq f(x^{k+1})$

iv. $f(x^*) \geq f_{\mu_k}(x^k) \geq f(x^k)$

Proof 3.3.1. *i. Since $0 < \mu_k < \mu_{k+1}$ and $\alpha(x) \geq 0$, we get*

$$\mu_k\alpha(x^{k+1}) \leq \mu_{k+1}\alpha(x^{k+1}).$$

Furthermore, since x^k minimizes $f_{\mu_k}(x)$, we have

$$\begin{aligned} f_{\mu_k}(x^k) = f(x^k) + \mu_k\alpha(x^k) &\leq f(x^{k+1}) + \mu_k\alpha(x^{k+1}) \\ &\leq f(x^{k+1}) + \mu_{k+1}\alpha(x^{k+1}) \\ &= f_{\mu_{k+1}}(x^{k+1}). \end{aligned} \tag{3.14}$$

Hence,

$$f_{\mu_k}(x^k) \leq f_{\mu_{k+1}}(x^{k+1}) \tag{3.15}$$

ii. As x^{k+1} *minimizes* $f_{\mu_{k+1}}(x)$*, we have*

$$f(x^{k+1}) + \mu_{k+1}\alpha(x^{k+1}) \leq f(x^k) + \mu_{k+1}\alpha(x^k) \tag{3.16}$$

Similarly, as x_k *minimizes* $f_{\mu_k}(x)$*, we have*

$$f(x^k) + \mu_k\alpha(x^k) \leq f(x^{k+1}) + \mu_k\alpha(x^{k+1}) \tag{3.17}$$

Adding the two inequalities 3.16 and 3.17 and simplifying, we get

$$[\mu_{k+1} - \mu_k][\alpha(x^k) - \alpha(x^{k+1})] \geq 0$$

Since $\mu_{k+1} - \mu_k > 0$*, we have* $\alpha(x^k) \geq \alpha(x^{k+1})$.

iii. From inequality 3.14 we get

$$f(x^k) - f(x^{k+1}) \leq \mu_k[\alpha(x^{k+1}) - \alpha(x^k)]. \tag{3.18}$$

Since, $\alpha(x^{k+1}) - \alpha(x^k) \leq 0$ *and* $\mu_k > 0$*, we have*

$$f(x^k) - f(x^{k+1}) \leq 0.$$

This implies that $f(x^k) \leq f(x^{k+1})$

iv. To prove this, we have

$$f(x^k) \leq f(x^k) + \mu_k\alpha(x^k) \leq f(x^*) + \mu_k\alpha(x^*) = f(x^*),$$

since $\mu_k\alpha(x^k) \geq 0$ *and* $\alpha(x^*) = 0$

Theorem 3.3.1. *Consider problem 3.12 where* $f, g_1, \ldots, g_m, h_1, \ldots, h_l$ *are continuous functions on* $\Re^n$ *and* X *is non-empty set in* $\Re^n$*. Suppose that the problem has a feasible optimal solution* x^**, and let* α *be a continuous function given by 3.13. Furthermore, suppose that for each* μ_k *there exists a solution* $x^k \in X$ *to the problem to minimize* $f(x) + \mu_k\alpha(x)$ *subject to* $x \in X$*, and that* $\{x^k\}$ *is contained in a compact subset of* X*. Then, the limit* $\bar{x}$ *of any convergent subsequence of* $\{x^k\}$ *is an optimal solution to the original problem, and* $\mu_k\alpha(x^k) \to 0$ *as* $\mu_k \to \infty$.

Proof 3.3.2. *Let* $\bar{x}$ *be a limit point of* x^k*. From the continuity of the function involved,*

$$\lim_{k\to\infty} f(x^k) = f(\bar{x}).$$

Also from (iv) of lemma 3.3.1,

$$f_\mu{}^* := \lim_{k\to\infty} f_{\mu_k}(x^k) \leq f(x^*) \quad \text{and} \quad \lim_{k\to\infty} f(x^k) = f(\bar{x}) \leq f(x^*) \tag{3.19}$$

Thus, by subtracting the above two equations we get the following

$$\lim_{k\to\infty}[f_{\mu_k}(x^k) - f(x^k)] = f_\mu{}^* - f(\bar{x}).$$

This implies that

$$\lim_{k\to\infty} \mu_k \alpha(x^k) = f_\mu{}^* - f(\bar{x}). \tag{3.20}$$

(by continuity of α) which is equivalent to

$$\alpha(\bar{x}) = \lim_{k\to\infty} \frac{1}{\mu_k}[f_\mu^* - f(\bar{x}] = 0, \quad \textit{since} \quad f_\mu^* - f(\bar{x}) \quad \textit{is constant}$$

and $\frac{1}{\mu_k} \to 0$ as $k \to \infty$. Therefore, $\bar{x}$ is feasible to the original constrained problem.

Since $\bar{x}$ feasible and x^* is minimizer of the original constrained problem,

$$f(x^*) \leq f(\bar{x}) \tag{3.21}$$

Hence, from 3.19 and 3.21

$$f(x^*) = f(\bar{x})$$

Therefore, the sequence $\{x^k\}$ converge to the optimal solution of the original constrained problem.
From equation 3.20, we have $\lim_{k\to\infty} \mu_k \alpha(x^k) = f_\mu{}^* - f(\bar{x}) = 0$.
Thus, $\mu_k \alpha(x^k) \to 0$ as $\mu_k \to \infty$.

From this Theorem it follows that the optimal solution x^k to $f_{\mu_k}(x)$ can be made arbitrarily close to the feasible region as $\mu_k \to \infty$. The optimal solutions $\{x^k\}$ are generally infeasible, but as μ_k is made large, the points generated approach an optimal solution from outside the feasible region.

3.3.3 Penalty Function Methods and Lagrange Multipliers

Consider the penalty function approach to problem 3.12. The auxiliary function that we minimize is then given by

$$f_\mu(x) = f(x) + \mu \Sigma_{i=1}^m [\max\{0, g_i(x)\}]^p + \mu \Sigma_{i=1}^l |h_i(x)|^p \quad \text{for} \quad x \in \Re^n.$$

For simplicity, let us take the quadratic form, $p = 2$. Thus the above function can be rewrite as:

$$Q_\mu(x) = f(x) + \frac{\mu}{2} \Sigma_{i=1}^m [\max\{0, g_i(x)\}]^2 + \frac{\mu}{2} \Sigma_{i=1}^l [h_i(x)]^2 \quad \text{for} \quad x \in \Re^n.$$

The necessary condition for this to have a minimum is that

$$\nabla Q_\mu(x) = \nabla f(x) + \Sigma_{i=1}^m \mu \max\{0, g_i(x)\} \nabla g_i(x) + \Sigma_{i=1}^l \mu h_i(x) \nabla h_i(x) = 0. \tag{3.22}$$

Suppose that the solution to 3.22 for a fixed μ (say $\mu_k > 0$) is given by x^k. Let us also designate

$$\mu_k \max\{0, g_i(x^k)\} = u_i(\mu_k), \quad \text{for all} \quad i = 1, \ldots, m, \tag{3.23}$$

$$\mu_k h_i(x^k) = \lambda_i(\mu_k), \quad \text{for all} \quad i = 1, \ldots, l \tag{3.24}$$

so that for $\mu = \mu_k$ we may rewrite 3.22 as

$$\nabla Q_{\mu_k}(x) = \nabla f(x) + \Sigma_{i=1}^m u_i(\mu_k)\nabla g_i(x) + \Sigma_{i=1}^l \lambda_i(\mu_k)\nabla h_i(x) = 0. \tag{3.25}$$

Now consider the Lagrangian for the original problem:

$$\mathcal{L}(x, \lambda, u) = f(x) + \Sigma_{i=1}^m u_i g_i(x) + \Sigma_{i=1}^l \lambda_i h_i(x).$$

The usual KKT necessary conditions yield

$$\nabla \mathcal{L}(x, \lambda, u) = \nabla f(x) + \Sigma_{i=1}^m u_i \nabla g_i(x) + \Sigma_{i=1}^l \lambda_i \nabla h_i(x) = 0, \tag{3.26}$$

where $u_i \geq 0$ for $i = 1, \ldots, m$. Comparing 3.25 and 3.26 we can see that when we minimize the auxiliary function using $\mu = \mu_k$, the $u_i(\mu_k)$ and $\lambda_i(\mu_k)$ values given by 3.23 and 3.24 estimate the Lagrange multipliers in 3.26.

In fact it may be shown that as the penalty function method proceeds and $\mu_k \to \infty$ and x^k converges to the optimum solution x^*, which satisfies a second order sufficiency conditions, the values of $u_i(\mu_k) \to u_i^*$ and $\lambda_i(\mu_k) \to \lambda_i^*$, the optimum Lagrange multiplier values for i^{th} inequality and equality constraints respectively.

Consider the problem

$$\text{Minimize} \quad f(x) = x_1^2 + 2x_2^2 \quad \text{Subject to} \quad g(x) = 1 - x_1 - x_2 \leq 0; \quad x \in \Re^2.$$

Its Lagrangian is given by

$$\mathcal{L}(x, u) = x_1^2 + 2x_2^2 + u(1 - x_1 - x_2).$$

The KKT conditions yield

$$\frac{\partial \mathcal{L}}{\partial x_1} = 2x_1 - u = 0, \quad \frac{\partial \mathcal{L}}{\partial x_2} = 4x_2 - u = 0, \quad u(1 - x_1 - x_2) = 0.$$

From this we have $x^* = (2/3, 1/3)$ for $u > 0$ as a solution to the given problem. If $u = 0$, then the resulting value is $x = (0, 0)$ which cannot be the feasible solution since it is infeasible. Define the penalty function $\alpha(x) = [\max\{g(x), 0\}]^2$. Thus

$$\alpha(x) = 0, \quad \text{for} \quad g(x) \leq 0 \quad \text{and} \quad \alpha(x) = (1 - x_1 - x_2)^2, \quad \text{for} \quad g(x) > 0.$$

Then the unconstrained problem is

$$f_{\mu_k}(x) = x_1^2 + 2x_2^2 + \mu_k \alpha(x).$$

If $\alpha(x) = 0$, then the optimal solution to minimize $f_{\mu_k}(x) = x_1^2 + 2x_2^2 + \mu_k \alpha(x)$ is at $x^* = (0, 0)$ and this is infeasible. So

$$f_{\mu_k}(x) = x_1^2 + 2x_2^2 + \mu_k(1 - x_1 - x_2)^2$$

Now, by using the necessary condition for optimality (i.e., $\nabla f_{\mu_k}(x) = 0$) we have the following:

$$\frac{\partial f_{\mu_k}}{\partial x_1} = 2x_1 - 2\mu_k(1 - x_1 - x_2) = 0,$$
$$\frac{\partial f_{\mu_k}}{\partial x_2} = 4x_2 - 2\mu_k(1 - x_1 - x_2) = 0.$$

This implies that

$$1 - x_1 - x_2 = \frac{x_1}{\mu_k}$$
$$1 - x_1 - x_2 = \frac{2x_2}{\mu_k}.$$

From these equations, we have $x_1 = 2x_2$. Thus $x^k = (\frac{2\mu_k}{2+3\mu_k}, \frac{\mu_k}{2+3\mu_k})$.
As $\mu_k \to \infty, \quad x^k \to x^*$ and from 3.23 we have

$$u(\mu) = 2\mu[\max\{0, g(x^k)\}] = 2\mu(1 - x_1 - x_2) = 2\mu(1 - \frac{2\mu}{2+3\mu} - \frac{\mu}{2+3\mu}),$$

since $g(x) > 0$ for $u > 0$.

Thus $u(\mu) = \frac{4\mu}{2+3\mu}$, which is readily seen that

$$\lim_{\mu\to\infty} u(\mu) = \lim_{\mu\to\infty} \frac{4\mu}{2+3\mu} = 4/3 = u^*.$$

From equation 3.23 and 3.24 we observe that as $\mu_k \to \infty$ the constraint functions vanished.

Conclusion

The main idea of interior penalty functions is that an optimal solution requires that a constraint be active (i.e.,tight) so that this optimal solutions lie on the boundary between feasibility and infeasibility. Knowing this, a penalty is applied to feasible solutions when the constraint is not active, so-called "interior solutions". The basic idea, we have discussed, in exterior penalty methods is to eliminate some or all of the constraints and add to the objective function a penalty term which prescribes a high cost to infeasible points. Associated with these methods is a parameter μ, which determines the severity of the penalty and as a consequence the extent to which the resulting unconstrained problem approximates the original constrained problem. These methods are not used in cases where feasibility must be maintained, for example, if the objective function is undefined or ill-conditioned outside the feasible region. Though interior and exterior methods suffer from some computational disadvantages, in the absence of alternative software especially for no-derivative problems they are still recommended. They work well for zero order methods like Powell's method with some modifications and taking different initial points and monotonically increasing parameters. Of the two methods, the exterior penalty methods considered preferable. The primary reasons are Interior penalty methods cannot deal with equality constraints without cumbersome modifications to the basic approach, demand a feasible starting point. Finding such a point often presents formidable difficulties in and of itself and require that the search in interior penalty method never leave the feasible region. This significantly increases the computational effort associated with the line search segment of the algorithm. Finally we left for other researchers to resolve the slow rate of convergence for both methods as $\mu \to 0$ for interior penalty and as $\mu \to \infty$ for exterior penalty function to arrive at the true minima early.

References

[1] Dr.Abebe Geletu *Solving Optimization Problems using the Matlab Optimization Toolbox - a Tutorial.*TU-Ilmenau, Fakultt fr Mathematik und Naturwissenschaften,December2007.

[2] M. Bartholomew-Biggs. *Nonlinear Optimization with Financial Applications.* Kluwer Academic Publisher, Boston, 2005.

[3] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming, Theory and Algorithms.* J. Wiley and Sons, New Jersey, third edition, 2006.

[4] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts, Second edition, 1999.

[5] J.M.Borwein, A.S. Lewis. *Convex Analysis and Nonlinear Optimization.* Springer Science-Business Media, Inc., Second Edition, 2006.

[6] K. P. CHONG, S.H. ZAK. *An Introduction to Optimization.* John Wiley and Sons Inc, Canada, Second edition, 2001.

[7] Fiacco, A.V. and G.P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley and Sons, New York, 1968.

[8] Hailay Weldegiorgis Berhe.Penalty function methods using matrix laboratory (MATLAB).Departr of Mathematics, Haramaya University, Ethiopia,May2012.

[9] J. Jahn. *Introduction to the Theory of Nonlinear Optimization.* Springer-Verlag, Berlin, 1996.

[10] BIJAN KUMAR PATEL *Solution Of Some Non-Linear Programming Problems.* Departmentof Mathematics (NIT ROURKELA),2014

[11] D.G. Luenberger, Y. Ye. *Linear and Nonlinear Programming.* Springer Science+Business Media, LLC, third edition, 2008.

[12] Manfred Huber *Computational Methods*,2011.

[13] Robert M.Freund *Penalty and Barrier Methods for Constrained Optimization.* Massachusetts Institute of Technology , 2004.

[14] G. Ventura. *An augmented Lagrangian approach to essential boundary conditions in meshless methods.* International Journal for Numerical Methods in Engineering. John Wiley and Sons, Ltd. 53:825842: 2002.

[15] W.SUN, Y.X. YUAN. *Optimization Theory and Methods (Nonlinear Programming).* Springer Science+Business Media, LLC , 2006.

[16] Yibltal Y. *Lagrange Function Method and Penalty Function Method.* School of graduate studies Addis Ababa Univerisity,June,2001.