PENALTY AND AUGMENTED LAGRANGIAN METHODS TO SOLVE NONLINEAR OPTIMIZATION PROBLEMS

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES
DEPARTMENT OF MATHEMATICS

"In partial fulfilment of the requirements for the
Degree of Master of Science in Mathematics"

By: Shemels Geletaw
Stream: Optimization

Advisor: Birhanu Guta(PhD)
Addis Ababa, Ethiopia

July 5, 2016

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 "**Penalty and Augmented Lagrangian Methods to Solve Nonlinear Optimization Problems**" by **Shemels Geletaw** in partial fulfilment of the requirements for the degree of Master of Science in Mathematics.

Advisor: Dr. Birhanu Guta

Signature: _____

Date: _____

Examiner 1: Dr. Hunduma L.

Signature: _____

Date: _____

Examiner 2: Dr. Tadesse B.

Signature: _____

Date: _____

ADDIS ABABA UNIVERSITY

Author: Shemels Geletaw

Title: Penalty and Augmented Lagrangian Methods to Solve Nonlinear Optimization Problems

Department: Mathematics

Degree: M.Sc.

Convocation: June

Year: 2016

Shemels Geletaw

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 my Instructors for their motivation and guidance during the past two years at Addis Ababa University, department of Mathematics. My thank also forwarded to my family and friends for all their support and unconditional love.

Abstract

Penalty and augmented Lagrangian methods are procedures for approximating constrained optimization problems by unconstrained optimization problems to find the approximate solution of the given constrained problem. The idea of replacing a constrained optimization problem by a sequence of unconstrained problems parametrized by a scalar parameter has played a fundamental role in the formulation of algorithms. Even though, penalty methods are very natural and general; unfortunately, they suffer from a serious drawback: to approximate well the solution to constrained problem, we have to work with large penalty parameters, and this inevitably makes the problem of unconstrained minimization of the penalized objective very ill-conditioned. Their slow rates of convergence due to ill-conditioning of the associated Hessian led researchers to pursue other approaches, augmented Lagrangian methods. The main advantage of the augmented Lagrangian methods is that they allow to approximate well the solution to a constrained problem by solutions of unconstrained (and penalized) auxiliary problems without pushing the penalty parameter to infinity; as a result, the auxiliary problems remain reasonably conditioned even when we are seeking for high-accuracy solutions. In this paper we will see how all this works.

Keywords: Unconstrained optimization, constrained optimization, penalty and barrier methods, augmented Lagrangian methods.

List of Figures

List of Tables

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: auxiliary function for penalty methods with penalty parameter μ
$\alpha(x)$: penalty function
$B(x)$: barrier function
$\mathcal{F}_\mu$: auxiliary function for barrier methods
Q_μ: auxiliary function for quadratic penalty method with penalty parameter μ
$\mathcal{L}_\mu(., \lambda)$: augmented Lagrangian function with penalty parameter μ and Lagrange multiplier λ
$< \lambda, h >$: 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 parametrized by a scalar parameter μ has played a fundamental role in the formulation of algorithms, [4]. To do this replacing penalty methods and augmented Lagrangian methods have a vital roll. Penalty methods approximate the solution for nonlinear constrained problem (NCP) by minimizing the penalty function for a large value of μ.

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. 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.

Similar to penalty functions, barrier 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. If the optimal solution occurs at the boundary of the feasible region, the procedure moves from the interior to the boundary.

Even though, penalty methods are very natural and general; unfortunately, they suffer from a serious drawback: to approximate well the solution to constrained problem, we have to work with large penalty parameters, and this inevitably makes the problem of unconstrained minimization of the penalized objective very ill-conditioned. Their slow rates of convergence due to ill-conditioning of the associated Hessian led researchers to pursue other approaches, augmented Lagrangian methods.

Hestenes and Powell independently proposed the augmented Lagrangian function for equality constrained problem (ECP) which is an unconstrained function based on augmenting the Lagrangian function with a quadratic penalty term that does not require μ to go to infinity for convergence. The price that must be paid for keeping μ finite is the need to update an estimate of the Lagrange multiplier vector in each iteration. Since the first appearance of the Hestenes-Powell function, many algorithms have been proposed based on using the

augmented Lagrangian as an objective for sequential unconstrained minimization, [3].

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. It also concerns an augmented Lagrangian function method that can be used to find a minimizing local solution of the nonlinear constrained problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad h(x) = 0, \quad x \in X,$$

where $f : \Re^n \to \Re, \quad h : \Re^n \to \Re^l$ are twice-continuously differentiable functions, and X is a non empty subset of $\Re^n$. This problem format assumes that all general inequality constraints have been converted to equalities by the use of slack variables. Algorithms are built in such a way that they rely in one way or another on the unconstrained case, so that the underlying idea is to construct a closely related, unconstrained optimization problem and apply to it some of the algorithms we already have for unconstrained problems.

In **Chapter-1** we try to discus some basic concepts on matrix definiteness, concepts of convexity and theories of nonlinear optimization, both unconstrained and constrained. The chapter focuses mainly on the minimization theories and basic conditions related to this.

Chapter-2 discuses on penalty function methods and barrier function methods. Throughout the chapter, we try to describe some basic concepts and properties of the methods for nonlinear optimization problems. Definitions, convergence theories, algorithmic schemes of the respective methods, and their drawbacks are highlighted in the chapter.

Chapter-3 focuses on the general concepts of Augmented Lagrangian methods for nonlinear constrained optimization problems. Basic definitions, convergence theories, algorithmic schemes and their advantages over penalty and barrier methods are the main focussing points of the chapter.

Chapter 1

Preliminary

1.1 Definiteness of a Matrix

Definition 1.1.1 *i. A symmetric $n \times n$ matrix A will be said to be positive semidefinite if $x^t Ax \geq 0$ for all $x \in \Re^n$. In this case we write*

$$A \geq 0.$$

ii. We say that A is positive definite if $x^t Ax > 0$ for all $x \neq 0$, and write

$$A > 0.$$

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

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

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

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

Property 1.1.1 *If all the eigenvalues of a matrix A are positive (nonnegative), then A is positive definite (semidefinite).*

Property 1.1.2 *If A is positive definite, then A^{-1} is positive definite.*

Property 1.1.3 *Let P be a symmetric $n \times n$ matrix and Q be a positive semidefinite $n \times n$ 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.

Property 1.1.4 *A matrix $P = A^t A$ is positive semidefinite since $x^t A^t Ax = |Ax|^2 \geq 0$.*

1.2 Convex Analysis

The concept of convexity is of great importance in the study of optimization problems.

Definition 1.2.1 (Convex Sets) *A set S in $\Re^n$ is said to be convex if the line segment joining any two points of the set also belonging to the set.*
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 dropped, then the combination is called an affine combination.
A cone is a non empty set $\mathcal{C}$ with the property that for all $x \in \mathcal{C}$ we have

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

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

$$\{(x_1, x_2)^t | x_1 \geq 0 \quad \text{or} \quad x_2 \geq 0\}$$

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

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

Note that all cones of this form are convex.

Definition 1.2.2 (Convex Functions) *Let S be a non empty convex set in $\Re^n$. The function $f : S \to \Re$ is said to be convex on S if*

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

for each $x_1, x_2 \in S$ and for each $\lambda \in [0,1]$.
The function f is said to be strictly convex on S if the above inequality holds as a strict inequality for each distinct $x_1, x_2 \in S$ and for each $\lambda \in (0,1)$.
The function f is said to be concave (strictly concave) if $-f$ is convex (strictly convex).

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.*

Theorem 1.2.1 *If f is convex and $f \in C^1$ over an open convex set S, then*

$$f(y) \geq f(x) + \nabla f(x)^t(y - x) \quad \textit{for all} \quad x, y \in S. \tag{1.1}$$

If in addition $f \in C^2$ over S, then the Hessian of f is positive semidefinite, $\nabla^2 f(x) \geq 0$, for all $x \in S$.

Conversely, if $f \in C^1$ over S and 1.1 holds, or if $f \in C^2$ over S and $\nabla^2 f(x) \geq 0$ for all $x \in S$, then f is convex over S.

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*

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

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*

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

1.3 Optimization Theory

In this section a theoretical background for methods to solve optimization problems will be provided. The knowledge about these concepts will allow it to implement practical algorithms. Optimization problems can be classified into unconstrained optimization problem and constrained optimization problems.

1.3.1 Unconstrained Optimization

Consider the unconstrained problem

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

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 1.3.1 (Minimizing Points) *i. A point x^* is a local minimizer if there is a neighbourhood $\mathcal{N}$ of x^* such that $f(x^*) \leq f(x)$ for all $x \in \mathcal{N}$.*

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

iii. A point x^* is an isolated local minimizer if there is a neighbourhood $\mathcal{N}$ of x^* such that x^* is the only local minimizer in $\mathcal{N}$.

iv. All isolated local minimizers are strict local minimizers, but the converse is not true.

For instance the function

$$f(x) = x^4 \cos(1/x) + 2x^4$$

for $x \neq 0$ has strict local minimizers at many nearby points x_k, and we can label these points so that $x_k \to 0$ as $k \to \infty$.

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

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

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

Optimality Conditions

Proposition 1.3.1 (Necessary Optimality Conditions) *Assume that x^* is a local minimizer of f and $f \in C^1$ over $\mathcal{N}$. 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 $\mathcal{N}$, then

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

Proposition 1.3.2 (Sufficient Optimality Conditions) *Let $f \in C^2$ over $\mathcal{N}$, $\nabla f(x^*) = 0$, 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 1.3.1 *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 1.3.1 *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*

$$x = \lambda z + (1 - \lambda)x^*, \quad \text{for some} \quad \lambda \in (0, 1]. \tag{1.2}$$

By the convexity property for f, we have

$$f(x) \leq \lambda f(z) + (1 - \lambda) f(x^*) < f(x^*). \tag{1.3}$$

Any neighbourhood $\mathcal{N}$ of x^* contains a piece of line segment 1.2, so there will always be points $x \in \mathcal{N}$ at which 1.3 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

$$\begin{aligned} \nabla f(x^*)^t (z - x^*) &= \frac{d}{d\lambda} f(x^* + \lambda(z - x^*))|_{\lambda=0} \\ &= lim_{\lambda \downarrow 0} \frac{f(x^* + \lambda(z - x^*)) - f(x^*)}{\lambda} \\ &\leq lim_{\lambda \downarrow 0} \frac{\lambda f(z) + (1 - \lambda) f(x^*)}{\lambda} \\ &= f(z) - f(x^*) < 0. \end{aligned} \tag{1.4}$$

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

1.3.2 Constrained Optimization

A general formulation for nonlinear constrained optimization problems is:

$$\begin{aligned} &\text{Minimize} \quad f(x) \\ &\text{subject to} \quad h(x) = 0, \\ &g(x) \leq 0 \end{aligned} \tag{1.5}$$

where $f : \Re^n \to \Re, h : \Re^n \to \Re^l, g : \Re^n \to \Re^m$ 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,

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

so that we can rewrite 1.5 more compactly as

$$\min_{x \in X} f(x), \quad \text{or} \quad min\{f(x) : x \in X\} \tag{1.6}$$

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 $g_i(x) = 0$ and inactive if the strict inequality $g_i(x) < 0$ is satisfied.
The set $A(x) = \{i : g_i(x) = 0; i = 1, \ldots, m\}$ 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 1.5 is called a convex optimization problem.

Definition 1.3.2 (Minimizer Points) *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 1.6 if $x^* \in X$ and there is a neighbourhood $\mathcal{N}$ of x^* such that $f(x) \geq f(x^*)$ for $x \in \mathcal{N} \cap X$.

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

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

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

Theorem 1.3.2 *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*

$$\nabla f(x^*)^t d \geq 0, \tag{1.7}$$

where $d = x - x^*$, feasible direction, for all $x \in X$.
If in addition f is convex over X and 1.7 holds, then x^* is a global minimizer.

Proof 1.3.2 *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 1.3.3 (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.*

There are numerous approaches to solve a constrained optimization problems. Common methods reformulate the problem to incorporate the objective function and the constraints into a new objective function. So the constrained problem becomes similar to an unconstrained problem, which can be solved by the corresponding but slightly modified methods.

Optimality Conditions for Equality Constrained Optimization

Consider the equality constrained problem

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array} \tag{1.8}$$

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 1.3.3 (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 $\nabla h_1(x^*), \ldots, \nabla h_l(x^*)$ are linearly independent.*

Definition 1.3.4 (Lagrangian Function) *The Lagrangian function $\mathcal{L} : \Re^{n+l} \to \Re$ for the problem 1.8 is defined by*

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

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

Proposition 1.3.3 (Karush-Kuhn-Tucker (KKT) Necessary Conditions) *Let x^* be a local minimum for 1.8 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*

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0. \tag{1.9}$$

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

$$z^t \nabla_{xx} \mathcal{L}(x^*, \lambda^*) z \geq 0. \tag{1.10}$$

Theorem 1.3.4 (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*

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \tag{1.11}$$

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

$$z^t \nabla_{xx} \mathcal{L}(x^*, \lambda^*) z > 0. \tag{1.12}$$

Then x^* is a strict local minimizer for 1.8

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

Optimality Conditions for General Constrained Optimization

Consider the constrained problem involving both equality and inequality constraints

$$\begin{aligned} &\text{Minimize} \quad f(x) \\ &\text{subject to} \quad h(x) = 0, \quad g(x) \leq 0 \end{aligned} \tag{1.13}$$

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

Definition 1.3.5 (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 1.3.6 (Lagrangian Function) *The Lagrangian function $\mathcal{L} : \Re^{n+l+m} \to \Re$ for the problem 1.13 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 1.3.5 (KKT Necessary Conditions) *Let x^* be a local minimum for 1.13 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*

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0 \tag{1.14}$$

$$\mu_i{}^* \geq 0, \quad \mu_i{}^* g_i(x^*) = 0, \quad \textit{for all} \quad i = 1, \ldots, m. \tag{1.15}$$

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*

$$z^t \nabla_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) z \geq 0. \tag{1.16}$$

Theorem 1.3.6 (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*

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0 \tag{1.17}$$

$$\mu_i{}^* \geq 0, \quad \mu_i{}^* g_i(x^*) = 0, \quad \textit{for all} \quad i = 1, \ldots, m \tag{1.18}$$

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*

$$z^t \nabla_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) z > 0. \tag{1.19}$$

Then x^* *is a strict local minimizer for 1.13*

Proposition 1.3.4 *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.*

1.4 Line Search Methods

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.

[**Rate of Convergence**] It measures the performance of the given algorithm. The rates of convergence of most practical algorithms are determined by the structure of the Hessian of the Lagrangian, much like the structure of the Hessian of the objective function determines the rates of convergence for most unconstrained methods. The general form of measuring rate of convergence is:

$$||x^{k+1} - x^*|| < M||x^k - x^*||^p,$$

where $p > 1$ which is called *order of convergence* and M is some positive constant.
If $p = 1$ and $M < 1$, then it is called *linear rate of convergence*. If $p = 2$, then the convergence is called *quadratic rate of convergence*.

[Ill-Conditioning]

Conditioning is a property of the numerical problem at hand (whether it is a linear algebra problem, an optimization problem, a differential equations problem, or whatever). A problem is said to be well conditioned if its solution is not affected greatly by small perturbations to the data that define the problem. Otherwise, it is said to be ill conditioned , [11].

Suppose that $m = \lambda_1 < \lambda_2 < ... < \lambda_l = M$ be the eigenvalues of a given positive definite Hessian matrix H. Then the ratio $\frac{M}{m}$ is called the condition number of H. Problems where $\frac{M}{m}$ are "large" referred as ill-conditioned, i.e., the bigger the condition number is the more ill-conditioned H is. Well-conditioned matrices have condition numbers close to 1. When the ratio is exactly 1, it is the best case that only one step will lead to the optimal solution (i.e., there is no wrong direction). The steepest descent method converges slowly for these problems because the number of iterations is proportional to the condition number of the problem.

1.4.1 Line Search for Nonlinear Programming

Here we discus some derivative based line search methods for function $f : \Re^n \to \Re$.

i. [**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 $-\nabla f(x^k)$.

ii. [**Newton Method**] This method is the most complex and also the fastest of the gradient methods. Let us approximate the function f with its second-order Taylor expansion

$$T(x+d) = f(x) + \nabla^t f(x)d + \frac{1}{2}d^t H_f(x)d.$$

Finding the minimum of $T(x+d)$ in d can give us a new direction towards x^*. Having a positive definite Hessian H_f, the minimum is the solution of $\nabla T(x+d) = 0$. Thus, we want to solve linear equation system

$$\nabla T(x+d) = \nabla f(x) + H_f(x)d = 0$$

in d. Its solution $d = -H_f(x)^{-1}\nabla f(x)$ gives direction as well as step size. Therefore, for multivariate optimization the generalization of Newton methods is

$$x^{k+1} = x^k - H_f(x^k)^{-1}\nabla f(x^k).$$

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.

The above construction ensures that for quadratic functions the optimum (if it exists) is found in one step.

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:

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

Chapter 2

Penalty Function and Barrier Function Methods

Introduction

This chapter is devoted to the exterior penalty (or the penalty) function methods and the interior penalty (or the barrier) function methods. They are procedures for approximating constrained optimization problems by unconstrained problems. In penalty methods, the feasible region of the constrained problem is expanded 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. In barrier methods, we presume that we are given a point that lies in the interior of the feasible region, and we impose a very large cost on feasible points that lie ever closer to the boundary of feasible region, thereby creating a barrier to exiting the feasible region.

2.1 Penalty Function Methods

Methods using 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, [13]. 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 2.1.1 (Penalty Function) *Consider the nonlinear constrained problem*

$$\begin{array}{ll} \textit{Minimize} & f(x) \\ \textit{subject to} & g_i(x) \leq 0 \quad \textit{for} \quad i = 1, \ldots, m; \\ & h_i(x) = 0 \quad \textit{for} \quad i = 1, \ldots, l; \\ & x \in X, \end{array} \tag{2.1}$$

where, $f, g_1, \ldots, g_m, h_1, \ldots, h_l$ 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)

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

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

$$\alpha(x) = \Sigma_{i=1}^{m}[\max\{0, g_i(x)\}]^p + \Sigma_{i=1}^{l}|h_i(x)|^p \tag{2.2}$$

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 $g_i(x) > 0$ for some $i = 1, \ldots, m$ or $h_i(x) \neq 0$ for some $i = 1, \ldots, l$.

Example 2.1.1

$$\textit{minimize} \quad f(x) = x \quad \textit{such that} \quad g(x) = 2 - x \leq 0, \quad x \in \Re.$$

Note that the minimizer is $x^* = 2$.

Solution 2.1.1 *Let $\alpha(x) = [\max\{g(x), 0\}]^2$ i.e., $\alpha(x) = 0, \quad for \quad x \geq 2$ or $\alpha(x) = (2 - x)^2, \quad for \quad x < 2$.*
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,

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

Since this function is quadratic form, we can evaluate the minimizer using first derivative, $f'_\mu(x) = 1 - 2\mu(2 - x) = 0$. This implies that $x = 2 - \frac{1}{2\mu}$, which converges to $x^* = 2$ as $\mu \to \infty$. Therefore, the minimizer of the original problem is $x^* = 2$. Its graphical representation is shown in figure 2.1 for some values of μ.

2.1.1 Quadratic Penalty Function

The quadratic penalty method adds a multiple of the square of the violation of each constraint to the objective. As the constraints get squared the new objective function retains the properties of smoothness and differentiability. Because of its simplicity and intuitive appeal, this approach is used often in practice, although it has some important disadvantages.
Consider the problem 2.1 where, $f, g_1, \ldots, g_m, h_1, \ldots, h_l$ are continuous functions defined on

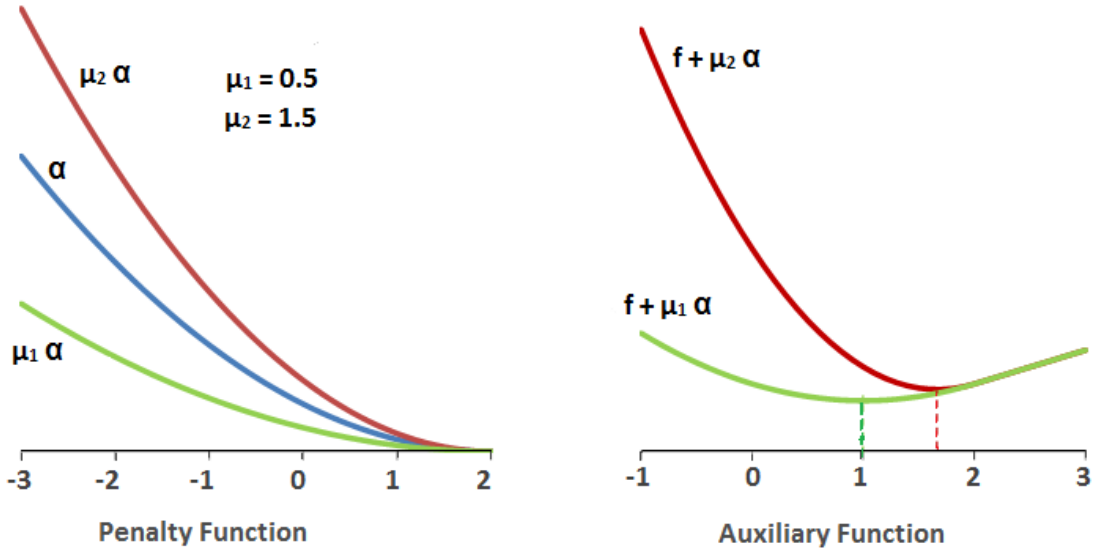

Figure 2.1: Graphical Illustration for Penalty Method

$\Re^n$, and X is a non empty set in $\Re^n$. Then, the quadratic penalized objective function is defined by

$$Q_\mu(x) := f(x) + \frac{\mu}{2}\Sigma_{i=1}^{m}[\max\{0, g_i(x)\}]^2 + \frac{\mu}{2}\Sigma_{i=1}^{l}[h_i(x)]^2 \tag{2.3}$$

For problems only with equality constraints 2.3 can be written as:

$$Q_\mu(x) := f(x) + \frac{\mu}{2}\Sigma_{i=1}^{l}|h_i(x)|^2 \tag{2.4}$$

For instance for the problem

$$\text{minimize} \quad f(x) = x^2 \quad \text{subject to} \quad x - 1 = 0,$$

the quadratic penalized objective function is:

$$Q_\mu(x) := x^2 + \frac{\mu}{2}(1 - x)^2,$$

with minimum point $x = 1$, and its approximate minimizers for different values of μ are illustrated in figure 2.2.

2.1.2 Convergence of 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 penalty methods:

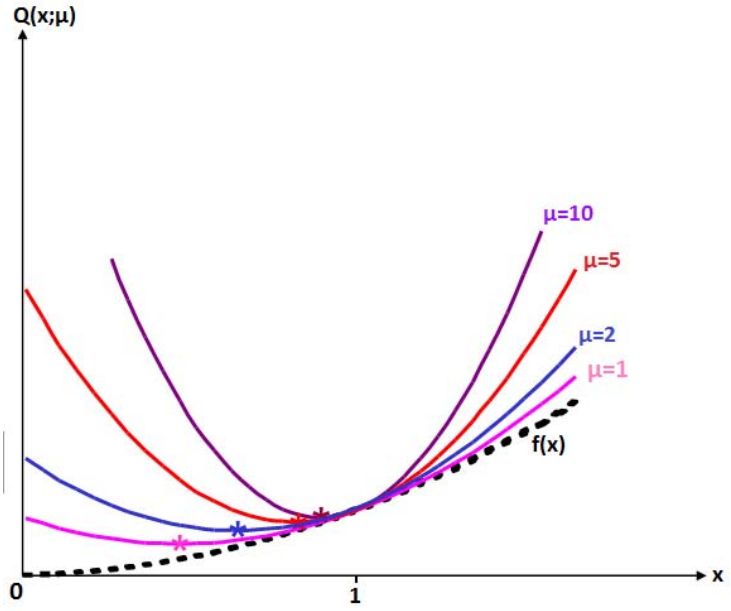

Figure 2.2: Graphical Illustration for Quadratic Penalty Method

Lemma 2.1.1 *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 2.2, 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 2.1.1 *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

$$\begin{aligned} f_{\mu_k}(x^k) = f(x^k) + \mu_k \alpha(x^k) &\leq f(x^{k+1}) + \mu_k \alpha(x^{k+1}) \\ &\leq f(x^{k+1}) + \mu_{k+1} \alpha(x^{k+1}) \\ &= f_{\mu_{k+1}}(x^{k+1}). \end{aligned} \tag{2.5}$$

Hence,

$$f_{\mu_k}(x^k) \leq f_{\mu_{k+1}}(x^{k+1}) \tag{2.6}$$

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

$$f(x^{k+1}) + \mu_{k+1}\alpha(x^{k+1}) \leq f(x^k) + \mu_{k+1}\alpha(x^k) \tag{2.7}$$

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

$$f(x^k) + \mu_k\alpha(x^k) \leq f(x^{k+1}) + \mu_k\alpha(x^{k+1}) \tag{2.8}$$

Adding the two inequalities 2.7 and 2.8 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 2.5 we get

$$f(x^k) - f(x^{k+1}) \leq \mu_k[\alpha(x^{k+1}) - \alpha(x^k)]. \tag{2.9}$$

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 2.1.1 *Consider problem 2.1 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 2.2. 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 2.1.2 *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 2.1.1,

$$f_\mu{}^* := \lim_{k\to\infty} f_{\mu_k}(x^k) \leq f(x^*) \quad \textit{and} \quad \lim_{k\to\infty} f(x^k) = f(\bar{x}) \leq f(x^*) \tag{2.10}$$

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

$$\lim_{k\to\infty} \mu_k \alpha(x^k) = {f_\mu}^* - f(\bar{x}). \tag{2.11}$$

(by continuity of α) which is equivalent to

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

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,

$$f(x^*) \leq f(\bar{x}) \tag{2.12}$$

Hence, from 2.10 and 2.12

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

Therefore, the sequence $\{x^k\}$ converge to the optimal solution of the original constrained problem.
From equation 2.11, 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.

2.1.3 Algorithmic Scheme for 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.

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

Example 2.1.2

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

Solution 2.1.2 *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 table which is the result of the MATLAB code described in the Appendix part of this paper. 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 2.1 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	$\mu_k\alpha(x^k)$	$f(x^k)$
1	1	$(0.0000, 0.0000)$	1.00000	0.00000
2	10	$(0.4000, 0.2000)$	1.60000	0.24000
3	100	$(0.6250, 0.3125)$	0.39063	0.58594
4	1000	$(0.6623, 0.3311)$	0.04386	0.65787
5	10000	$(0.6662, 0.3331)$	0.00444	0.66578
6	100000	$(0.6666, 0.3333)$	0.00044	0.66658
7	1000000	$(0.6667, 0.3333)$	0.00004	0.66666

Table 2.1: Penalty Iteration for Example 2.1.2

2.1.4 Penalty Function Methods and Lagrange Multipliers

Consider the penalty function approach to problem 2.1. 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

$$\nabla Q_\mu(x) = \nabla f(x) + \Sigma_{i=1}^m \mu \max\{0, g_i(x)\}\nabla g_i(x) + \Sigma_{i=1}^l \mu h_i(x)\nabla h_i(x) = 0. \tag{2.13}$$

Suppose that the solution to 2.13 for a fixed μ (say $\mu_k > 0$) is given by x^k. Let us also designate

$$\mu_k \max\{0, g_i(x^k)\} = u_i(\mu_k), \quad \text{for all} \quad i = 1, \ldots, m, \tag{2.14}$$

$$\mu_k h_i(x^k) = \lambda_i(\mu_k), \quad \text{for all} \quad i = 1, \ldots, l \tag{2.15}$$

so that for $\mu = \mu_k$ we may rewrite 2.13 as

$$\nabla Q_{\mu_k}(x) = \nabla f(x) + \Sigma_{i=1}^m u_i(\mu_k)\nabla g_i(x) + \Sigma_{i=1}^l \lambda_i(\mu_k)\nabla h_i(x) = 0. \tag{2.16}$$

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

$$\nabla\mathcal{L}(x, \lambda, u) = \nabla f(x) + \Sigma_{i=1}^m u_i \nabla g_i(x) + \Sigma_{i=1}^l \lambda_i \nabla h_i(x) = 0, \tag{2.17}$$

where $u_i \geq 0$ for $i = 1, \ldots, m$. Comparing 2.16 and 2.17 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 2.14 and 2.15 estimate the Lagrange multipliers in 2.17.

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 2.14 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 2.14 and 2.15 we observe that as $\mu_k \to \infty$ the constraint functions vanished.

2.1.5 Drawbacks of Penalty Function Methods

i. Difficulties Related to Ill-Conditioning

We now examine the nature of the ill-conditioning in the Hessian $\nabla^2_{xx} f_{\mu_k}(x)$. An understanding of the properties of this matrix, and the similar Hessians that arise in other penalty and barrier methods, is essential in choosing effective algorithms for the minimization problem at each iteration, [12].
Since the penalty method is based on the solution of the problem of the form

$$\text{Minimize} \quad f_{\mu_k}(x) \quad \text{subject to} \quad x \in \Re^n \tag{2.18}$$

it is natural to inquire about the degree of difficulty in solving such problems. When $f, h, g \in C^2$, the degree of difficulty for solving problem 2.18 depends on the the eigenvalue structure of the Hessian matrix $\nabla^2_{xx} f_{\mu_k}(x)$, [4].
The condition number, defined as

$$\frac{\text{maximum eigenvalue of} \quad \nabla^2 f_\mu}{\text{minimum eigenvalue of} \quad \nabla^2 f_\mu}$$

can get arbitrarily large. As a consequence of all this, the numerical solution of the Newton equation $\nabla^2 f_\mu d = -\nabla f_\mu$ is very susceptible to rounding error and the resulting search directions can be inaccurate and ineffective. Similar difficulties can occur if we attempt to minimize $f_\mu(x)$ using quasi-Newton or conjugate gradient methods, [1].

To see this, let us consider the Hessian of the quadratic penalized objective function,

$$\nabla^2_{xx} Q_{\mu_k}(x) = [\nabla^2 f(x) + \Sigma^l_{i=1} \mu_k h_i(x) \nabla^2 h_i(x)] + \mu_k \Sigma^l_{i=1} \nabla h_i(x) \nabla^t h_i(x).$$

When x is close to the minimizer of $Q_{\mu_k}(x)$, then $\mu_k h_i(x)$ approaches to the corresponding Lagrangian multiplier, λ^*, of the original constrained problem and we have the following estimation:

$$\nabla^2_{xx} Q_{\mu_k}(x) \approx \nabla^2_{xx} \mathcal{L}(x, \lambda^*) + \mu_k \Sigma^l_{i=1} \nabla h_i(x) \nabla^t h_i(x),$$

where $\nabla^2_{xx} \mathcal{L}(x, \lambda^*)$ is Hessian of the Lagrangian function with multiplier λ^* to the given constrained problem.
But $\nabla^2_{xx} \mathcal{L}(x, \lambda^*)$ is independent of μ_k and the matrix $\mu_k \Sigma^l_{i=1} \nabla h_i(x) \nabla^t h_i(x)$ with nonzero eigenvalues of order μ_k is depend on μ_k. Therefore, the matrix

$$\nabla^2_{xx} \mathcal{L}(x, \lambda^*) + \mu_k \Sigma^l_{i=1} \nabla h_i(x) \nabla^t h_i(x)$$

has some of its eigenvalues approaching a constant, while others are of order μ_k. Since μ_k is approaching ∞, the increasing ill-conditioning of $\nabla^2_{xx} Q_{\mu_k}(x)$ is apparent.

Example 2.1.3 *Consider the Hessian matrix of the Auxiliary function* $f_\mu(x) = x_1^2 + 2x_2^2 + \mu(1 - x_1 - x_2)^2$.

$$H = \begin{bmatrix} 2 + 2\mu & 2\mu \\ 2\mu & 4 + 2\mu \end{bmatrix}.$$

Suppose we want to find its eigenvalues by solving

$$|H - \lambda I| = \lambda^2 - (6 + 4\mu)\lambda + (8 + 12\mu) = 0$$

This yields

$$\lambda_1 = (3 + 2\mu) - \sqrt{4\mu^2 + 1} \quad \textit{and} \quad \lambda_2 = (3 + 2\mu) + \sqrt{4\mu^2 + 1}$$

Taking the ratio of the largest and the smallest eigenvalue yields

$$\frac{(3 + 2\mu) + \sqrt{4\mu^2 + 1}}{(3 + 2\mu) - \sqrt{4\mu^2 + 1}}.$$

It is clear that as $\mu \to \infty$, the limit of the preceding ratio (the condition number) also goes to ∞. This indicates that as the iterations proceed and we start to increase the value of μ, the Hessian of the unconstrained function that we are minimizing becomes increasingly ill-conditioned. This is a common situation and is especially problematic if we are using a method for the unconstrained optimization that requires the use of the Hessian.

Here, steepest descent is also out of the question as a possible solution method, since the rate of convergence of the method of steepest descent applied to a functional is determined by the ratio of the smallest to the largest eigenvalues of the Hessian of that functional. It follows in particular that the steepest descent method applied to f_{μ_k} converges slowly for large μ. Even Newton's method can encounter significant difficulties if μ_k is very high, and the starting point for minimizing $f_{\mu_k}(x)$ is not near a solution, [4].

In general, the poor conditioning of this system will lead to significant errors in the computed value of d, regardless of the computational technique used to solve $\nabla^2 f_\mu d = -\nabla f_\mu$.

Note that, the ill-conditioning associated with the unconstrained minimization problem 2.18 is a basic characteristic feature of penalty methods and represents the overriding factor in determining the manner in which these methods are operated.

ii. Difficulties Related to Small Value of the Penalty parameter

On the other hand, if we start the iteration from a small value of the penalty parameter, then the minimizer of the penalized objective function may diverge away from the feasible

region.
For instance, if we have the problem

$$\text{minimize} \quad -5x_1^2 + x_2^2 \quad \text{subject to} \quad x_1 = 1,$$

then the penalized objective function

$$Q_\mu(x) = -5x_1^2 + x_2^2 + \frac{\mu}{2}(x_1 - 1)^2$$

is unbounded for $\mu < 10$.
To see this let us compute the minimizer using first order necessary conditions of optimality. Thus,

$$\nabla^2_{xx} Q_\mu(x) = (-10x_1 + \mu(x_1 - 1), 2x_2)^t = (0, 0)^t.$$

This implies that $x = (\frac{\mu}{\mu - 10}, 0)$ and as $\mu \to 10$ from the left side, x is unbounded and we can not approach to the right solution $x^* = (1, 0)$. But for $\mu > 10$, as $\mu \to \infty$, we can approach to that solution.

2.2 Barrier Function Methods

These methods transform the original constrained problem into an unconstrained problem or into a sequence of unconstrained problems; however the barrier (or interior penalty) functions 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). If the optimal solution occurs at the boundary of the feasible region, the procedure moves from the interior to the boundary, [13].

Definition 2.2.1 (Barrier Function) *Consider the nonlinear inequality constrained problem*

$$\begin{array}{ll} \textit{Minimize} & f(x) \\ \textit{subject to} & g_i(x) \leq 0 \quad \textit{for} \quad i = 1, \ldots, m, \end{array} \tag{2.19}$$

where $f, g_1, \ldots, g_m, h_1, \ldots, h_l$ are continuous functions defined on $\Re^n$.
A barrier function B 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 $\phi(y) \geq 0$ if $y < 0$ and $\lim\limits_{y \to 0^-} \{\phi(y)\} = \infty$, where ϕ is a univariate function that is continuous over $\{y : y < 0\}$. Then

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

Usually,

$$B(x) = -\Sigma_{i=1}^m \frac{1}{g_i(x)} \quad \textit{or} \quad B(x) = -\Sigma_{i=1}^m \log[-g_i(x)].$$

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

$$\mathcal{F}_\mu(x) := f(x) + \frac{1}{\mu}B(x),$$

where μ *is a positive number.*

Ideally, we would like $B(x) = 0$ if $g_i(x) < 0$ and $B(x) = \infty$ if $g_i(x) = 0$, so that we never leave the region $\{x|g(x) \leq 0\}$. However, $B(x)$ is now discontinuous. This causes serious computational problems during the unconstrained optimization. Therefore, this ideal construction of B is replaced by the more realistic requirement that B 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:

$$\begin{aligned} \text{Minimize} \quad & f(x) + \frac{1}{\mu}B(x) \\ \text{subject to} \quad & g_i(x) < 0 \quad \text{for} \quad i = 1, \ldots, m, \quad x \in \Re^n. \end{aligned} \tag{2.20}$$

From this we observe that the barrier problem itself is a constrained problem, and indeed the constraint is somewhat 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 $g(x) = -x + 2 \leq 0$, the barrier function is given by

$$B(x) = \frac{-1}{-x+2},$$

with $x \geq 2$ as a feasible interval for the given constrained problem.

So, the corresponding auxiliary function can be written as $f(x) + \frac{1}{\mu}B(x) = x - \frac{1}{\mu(-x+2)}$ and its optimum solution is at $x^* = 2 - \sqrt{\frac{1}{\mu}}$, and as $\mu \to \infty, x^* \to 2$.
Figure 2.3 shows the convergence of this problem in the example for different values of μ.

2.2.1 Convergence of Barrier Function Methods

Similar to exterior 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 $\mu_k \to \infty$ as $k \to \infty$. Let x^k denote the solution to $\mathcal{F}_{\mu_k}(x)$, and x^* be an optimal solution to problem 2.19. Then the following Lemma presents some basic properties of barrier methods.

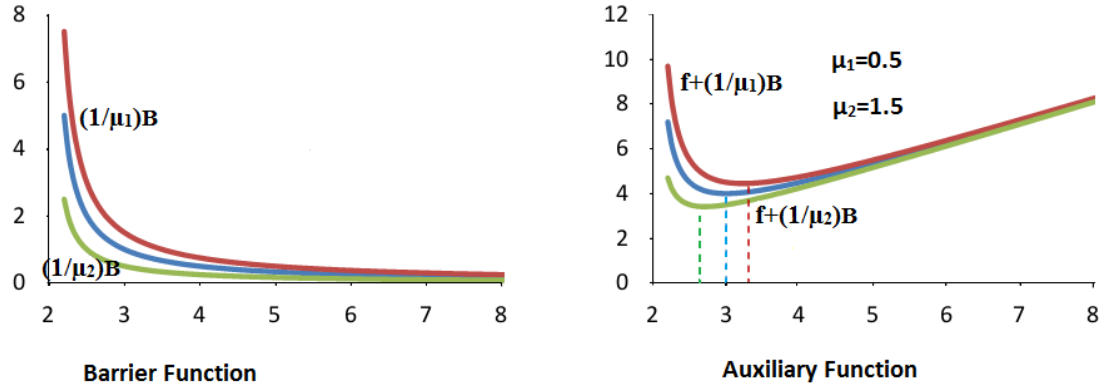

Figure 2.3: Graphical Illustration for Barrier Method

Lemma 2.2.1 (Barrier Lemma) *i.* $\mathcal{F}_{\mu_k}(x^k) \leq \mathcal{F}_{\mu_{k+1}}(x^{k+1})$

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

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

iv. $f(x^*) \leq f(x^k) \leq \mathcal{F}_{\mu_k}(x^k)$

Theorem 2.2.1 (Barrier Convergence Theorem) *Suppose* $f(x), g(x)$, *and* $B(x)$ *are continuous functions. Let* $\{x^k\}$, *for* $k = 1, 2, \ldots$, *be a sequence of solutions of* $\mathcal{F}_{\mu_k}(x)$. *Suppose there exists an optimal solution* x^* *of 2.19 for which* $\mathcal{N} \cap \{x | g(x) < 0\} \neq \emptyset$, *where* $\mathcal{N}$ *is a neighbourhood of* x^*. *Then any limit point* $\bar{x}$ *of* $\{x^k\}$ *solves 2.19. Furthermore,* $\frac{1}{\mu} B(x) \to 0$ *as* $\mu \to \infty$.

Proof 2.2.1 *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 2.19.*

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,

$$f(x^*) + \epsilon + \frac{1}{\mu_k} B(\hat{x}) \geq f(\hat{x}) + \frac{1}{\mu_k} B(\hat{x}) \geq \mathcal{F}_{\mu_k}(x^k).$$

Therefore, for k *sufficiently large,*

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

and since $\mathcal{F}_{\mu_k}(x^k) \geq f(x^*)$ *from (iv) of Lemma 2.2.1, then*

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

This implies that

$$\lim_{k\to\infty} \mathcal{F}_{\mu_k}(x^k) = f(\bar{x}) + \lim_{k\to\infty} \frac{1}{\mu_k} B(x^k) = f(x^*).$$

We also have

$$f(x^*) \leq f(x^k) + \frac{1}{\mu_k} B(x^k) = \mathcal{F}_{\mu_k}(x^k).$$

Taking limits we obtain

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

From this, we have

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

Hence, $\bar{x}$ is the optimal solution of the original nonlinear inequality constrained problem, 2.19. Furthermore, from

$$f(\bar{x}) + \lim_{k\to\infty} \frac{1}{\mu_k} B(x^k) = f(x^*),$$

we have

$$\lim_{k\to\infty} \frac{1}{\mu_k} B(x^k) = f(x^*) - f(\bar{x}) = 0.$$

Therefore, as $k \to \infty$, i.e., $\mu_k \to \infty$, the function $\frac{1}{\mu_k} B(x^k) \to 0$ for each k. This proves the second statement of the Theorem.

2.2.2 Algorithmic Scheme for Barrier 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 a starting point x^1 with $g(x^1) < 0$ and formulate the objective function $\mathcal{F}_{\mu_k}(x)$. Set $k = 1$.

2. [**Iterative Step**] Starting from x^k use an unconstrained search technique to find the point that minimizes $\mathcal{F}_{\mu_k}(x)$ 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 $\mathcal{F}_{\mu_{k+1}}(x)$ and put $k = k+1$ and return to the iterative step.

Consider the following problem again:

Example 2.2.1

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

Solution 2.2.1 *Define the barrier function*

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

The unconstrained problem is

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

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

$$\frac{\partial \mathcal{F}_{\mu_k}}{\partial x_1} = 2x_1 - \frac{1}{\mu_k(x_1 + x_2 - 1)} = 0, \quad \textit{and} \quad \frac{\partial F_{\mu_k}}{\partial x_2} = 4x_2 - \frac{1}{\mu_k(x_1 + x_2 - 1)} = 0$$

and we get, $x_1 = \frac{1 \pm \sqrt{1 + \frac{3}{\mu_k}}}{3}$ *and* $x_2 = \frac{1 \pm \sqrt{1 + \frac{3}{\mu_k}}}{6}$
Since the negative signs lead to infeasibility, we have

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

Starting with $\mu_1 = 1, \beta = 10$ *and* $x^1 = (1, 1)$ *and using a tolerance of* 0.0001 *(say), we have the following: Thus, this solution approach to the exact optimal solution* $x^* = (2/3, 1/3)$.

k	μ_k	x^k	$f(x^k)$
1	1	$(1.000000, 1.000000)$	3.000000
2	10	$(1.113417, 0.556708)$	1.859546
3	100	$(0.830267, 0.415134)$	1.034016
4	1000	$(0.722140, 0.361070)$	0.782229
5	10000	$(0.684682, 0.342341)$	0.703185
6	100000	$(0.668461, 0.331539)$	0.666676
7	1000000	$(0.668461, 0.331539)$	0.666676

Table 2.2: Barrier Iteration for the Given Example

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.

Note that the MATLAB code used to solve this problem is described in the Appendix-A of this paper.

2.2.3 Drawbacks of Barrier Function Methods

We now analyse the structure of the Hessian of the barrier function and show that as μ increases, the Hessian becomes increasingly ill-conditioned. Although our discussion focus

on the logarithmic barrier function method, the same results are true in general for other barrier methods. Consider the logarithmic barrier method,

$$\mathcal{F}_\mu(x) = f(x) + \frac{1}{\mu}\Sigma_{i=1}^m \log[-g(x)]$$

Then the Hessian is the form

$$\nabla^2_{xx}\mathcal{F}_\mu(x) = \nabla^2 f(x) + \Sigma_{i=1}^m \frac{1}{\mu g_i(x)}\nabla^2 g_i(x) - \Sigma_{i=1}^m \frac{1}{\mu g_i^2(x)}\nabla g_i(x)\nabla^t g_i(x)$$

Let x_k be the minimizer of $\mathcal{F}_{\mu_k}(x)$, and let $\lambda_i = \frac{1}{\mu g_i(x_k)}$ be the associated Lagrange multiplier estimate. Then

$$\nabla^2_{xx}\mathcal{F}_\mu(x) = [\nabla^2 f(x) + \Sigma_{i=1}^m \lambda_i \nabla^2 g_i(x)] - \Sigma_{i=1}^m \mu\lambda_i^2 \nabla g_i(x)\nabla^t g_i(x). \tag{2.21}$$

The term in the braces, $\nabla^2 f(x) + \Sigma_{i=1}^m \lambda_i \nabla^2 g_i(x)$, is an approximation to the Hessian of the Lagrangian to problem 2.19, and its condition number reflects the conditioning of the Lagrangian Hessian for 2.19. Ill-conditioning in the barrier function arises because of the final term

$$\Sigma_{i=1}^m \mu\lambda_i^2 \nabla g_i(x)\nabla^t g_i(x)$$

in equation 2.21. If the constraint is binding at the solution of problem 2.19 (and its corresponding Lagrange multiplier is not zero), $\mu\lambda_i^2$ tends to infinity as μ goes to infinity. As a result, the Hessian matrix $\nabla^2_{xx}\mathcal{F}_\mu(x)$ in general becomes increasingly ill-conditioned as the solution is approach. This structural ill-conditioning occurs even when the underline constrained problem is well conditioned.

Note 1 *Of the two methods, penalty and barrier, the penalty methods must be considered preferable. The primary reasons are as follows.*

i. barrier methods cannot deal with equality constraints without cumbersome modifications to the basic approach.

ii. barrier methods demand a feasible starting point. Finding such a point often presents formidable difficulties in and of itself.

iii. barrier methods require that the search never leave the feasible region. This significantly increases the computational effort associated with the line search segment of the algorithm.

Chapter 3

Augmented Lagrangian Penalty Function Methods

Introduction

Penalty methods discussed in the previous chapter are very natural and general; unfortunately, they suffer from a serious drawback: to approximate well the solution to constrained problem, we have to work with large penalty parameters, and this inevitably makes the problem of unconstrained minimization of the penalized objective very ill-conditioned.

Exact penalty function methods are used to alleviate the computational difficulties associated with having to take the penalty parameter to infinity in order to recover an optimal solution to the original problem. In these methods a single unconstrained minimization problem, with a reasonable sized penalty parameter can yield an optimum solution to the original problem. The *exact absolute value or* l_1 *penalty function method* and the *augmented Lagrangian penalty function method* are the two important exact penalty methods, [2].

In this chapter we discuss the augmented Lagrangian method (or multiplier penalty method) which is one of the exact penalty methods. This method is an extension of the quadratic penalty function method. The approach uses both a Lagrangian multiplier term and a penalty term in the auxiliary function. The main advantage of the method is that it allows to approximate well the solution to a constrained problem by solutions of unconstrained (and penalized) auxiliary problems without pushing the penalty parameter to infinity. This is an important advantage, since it results in elimination or at least moderation of the *ill-conditioning problem*. Moreover, we can employ efficient derivative based methods in minimizing the penalized objective function. The other advantage of the method of multiplier is that its *convergence rate* is considerably better than that of the penalty method.

3.1 Augmented Lagrangian Methods for Problems with Equality Constraints

Throughout this section we consider the problem

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, l, \end{array} \tag{3.1}$$

where $f : \Re^n \to \Re$ and $h_i(x)$ for $i = 1, \ldots, l$ are continuous functions on $\Re^n$. We assume that problem 3.1 has at least one feasible solution. Now, by perturbing the right-hand sides of the constraints from 0 to $\epsilon > 0$, where $\epsilon = (\epsilon_1, \ldots, \epsilon_l)$, it could become possible to obtain a constrained optimal solution to the original problem without letting $\mu \to \infty$. Based on this assumption, we have the penalized objective function

$$f(x) + \frac{\mu}{2}\Sigma_1^l[h_i(x) - \epsilon_i]^2 = f(x) - \Sigma_1^l \mu\epsilon_i h_i(x) + \frac{\mu}{2}\Sigma_1^l[h_i(x)]^2 + \frac{\mu}{2}\Sigma_1^l[\epsilon_i]^2.$$

Denoting $\lambda_i = -\mu\epsilon_i$ for $i = 1, \ldots, l$ and dropping the constant term $\frac{\mu}{2}\Sigma_1^l[\epsilon_i]^2$, the right-hand side of the above equation can be rewritten as

$$f(x) + \Sigma_{i=1}^l \lambda_i h_i(x) + \frac{\mu}{2}\Sigma_{i=1}^l h_i^2(x)$$

Definition 3.1.1 *For any positive penalty parameter μ the augmented Lagrangian function is defined by:*

$$\mathcal{L}_\mu(x, \lambda) := f(x) + \Sigma_{i=1}^l \lambda_i h_i(x) + \frac{\mu}{2}\Sigma_{i=1}^l h_i^2(x), \tag{3.2}$$

where λ_i is a multiplier corresponding to the constraint h_i, for $i = 1, 2, \ldots, l$.

Penalty Function Viewpoints: For a fixed multiplier vector λ, $\mathcal{L}_\mu(x, \lambda)$ correspond to the quadratic penalty for the problem

$$\begin{array}{ll} \text{Minimize} & f(x) + \Sigma_{i=1}^l \lambda_i h_i(x) \\ \text{subject to} & h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, l; \end{array} \tag{3.3}$$

with $\frac{\mu}{2}\Sigma_{i=1}^l \lambda_i |h_i(x)|^2$ as the penalty function. Note that 3.1 and 3.3 are equivalent.

Duality Viewpoints: For a penalty parameter μ, $\mathcal{L}_\mu(x, \lambda)$ correspond to the Lagrangian for the problem

$$\begin{array}{ll} \text{Minimize} & f(x) + \frac{\mu}{2}\Sigma_{i=1}^l |h_i(x)|^2 \\ \text{subject to} & h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, l; \end{array} \tag{3.4}$$

the term $\frac{\mu}{2}\Sigma_{i=1}^l |h_i(x)|^2$ tends to "convexify" the Lagrangian. For sufficiently large μ, the Lagrangian will indeed be locally convex. Here, 3.1 and 3.4 are equivalent.

These two view points lead to a rich theory and algorithmic felicity that is not present in either the pure quadratic penalty function approach or the pure Lagrangian duality-based approach.

Therefore, the augmented Lagrangian method for solving problem 3.1 proceeds by choosing λ and μ judiciously and then minimizing $\mathcal{L}_\mu(x, \lambda)$ as a function of x. The resulting x is used to choose a new λ and μ, and the process repeats.

If μ is reasonably large, minimizing $\mathcal{L}_\mu$ will tend to make $||h(x)||$ small (as for the penalty method), even if λ is somewhat arbitrary. Also, the Hessian of $\mathcal{L}_\mu$ will tend to have positive curvature in the right space and a minimizer is likely to exist.

The strategy is to check that $||h(x)||$ is suitably small after each (approximate) minimization of $\mathcal{L}_\mu$. If so, λ is updated to some value using a special strategy. If not, μ is increased and λ remains the same.

Under favourable conditions, (x, λ) approaches to the optimal KKT solution $(\bar{x}, \bar{\lambda})$ of the problem 3.1 before μ becomes too large. Assume that $\bar{x}$ be a KKT solution for the problem 3.1 corresponding to the multiplier $\bar{\lambda}$, then

$$\begin{aligned} \nabla_x \mathcal{L}_\mu(\bar{x}, \bar{\lambda}) &= \nabla f(\bar{x}) + \Sigma_{i=1}^{l} \bar{\lambda}_i \nabla h_i(\bar{x}) + \Sigma_{i=1}^{l} \mu h_i(\bar{x}) \nabla h_i(\bar{x}) \\ &= \nabla_x \mathcal{L}(\bar{x}, \bar{\lambda}) + \Sigma_{i=1}^{l} \mu h_i(\bar{x}) \nabla h_i(\bar{x}) \\ &= \nabla_x \mathcal{L}(\bar{x}, \bar{\lambda}) \\ &= 0 \end{aligned} \tag{3.5}$$

for all values of μ; whereas this was not necessarily the case with the quadratic function, unless $\nabla f(\bar{x})$ was itself zero. Hence, whereas we need to take $\mu \to \infty$ to recover $\bar{x}$ in a limiting sense using the quadratic function, it is conceivable only need to make μ large enough (under suitable regularity conditions) for the critical point $\bar{x}$ of $\mathcal{L}_\mu(\bar{x}, \bar{\lambda})$ to turn out to be its (local) minimizer. In this respect, the last term in 3.2 turns out to be local convexifier of the overall function.

From 3.5 we observe that, at the KKT solution $\bar{x}$ the multiplier in the augmented Lagrangian function is equal to the exact Lagrange multiplier of the original equality constrained problem; and also the gradient of the augmented Lagrangian function is coincides with the gradient of the Lagrangian function of 3.1 at this point.

[Basic Ideas]:

i. Optimize the augmented Lagrangian function and adjust λ, the multiplier in the augmented Lagrange function, in the hope of matching the true Lagrange multiplier λ^*;

ii. For μ large enough (finite) the unconstrained minimizer of the augmented Lagrangian coincides with the solution of the original constrained problem;

iii. At convergence, the gradient $\mu h_i(x) \nabla h_i(x)$ vanishes and we recover the standard multiplier rule.

3.1.1 Convergence of the Augmented Lagrangian Methods

Theorem 3.1.1 *Consider problem 3.1, and let the KKT solution $(\bar{x}, \bar{\lambda})$ satisfy the second order sufficiency conditions for a local minimum. Then there exists a $\bar{\mu}$ such that for $\mu \geq \bar{\mu}$, the augmented Lagrangian penalty function $\mathcal{L}_\mu(., \bar{\lambda})$, defined for some accompanying $\lambda = \bar{\lambda}$, also achieves a strict local minimum at $\bar{x}$.*

Proof 3.1.1 *We prove the result by showing that $(\bar{x}, \bar{\lambda})$ satisfies the second-order sufficient conditions to be a strict local minimizer of $\mathcal{L}_\mu(., \bar{\lambda})$ for all μ sufficiently large; that is,*

$$\nabla_x \mathcal{L}_\mu(\bar{x}, \bar{\lambda}) = 0, \quad \textit{and} \quad \textit{the Hessian of} \quad \mathcal{L}_\mu(., \bar{\lambda}) \quad \textit{is positive definite with respect to} \quad \bar{x}.$$

Since $(\bar{x}, \bar{\lambda})$ is a KKT solution to 3.1, by equation 3.5 we have $\nabla_x \mathcal{L}_\mu(\bar{x}, \bar{\lambda}) = 0$.
Let $H(\bar{x})$ be the Hessian of $\mathcal{L}_\mu(., \bar{\lambda})$. Then

$$\begin{aligned} H(\bar{x}) &= [\nabla^2 f(\bar{x}) + \Sigma_{i=1}^{l} \lambda_i \nabla^2 h_i(\bar{x}) + \mu \Sigma_{i=1}^{l} h_i(\bar{x}) \nabla^2 h_i(\bar{x})] + \mu \Sigma_{i=1}^{l} \nabla h_i(\bar{x}) \nabla^t h_i(\bar{x}) \\ &= \nabla^2_{xx} \mathcal{L}(\bar{x}, \bar{\lambda}) + \mu \Sigma_{i=1}^{l} \nabla h_i(\bar{x}) \nabla^t h_i(\bar{x}), \quad \textit{for} \quad i = 1, \ldots, l, \end{aligned} \tag{3.6}$$

where $\nabla^2 \mathcal{L}(\bar{x}, \bar{\lambda})$ is the Hessian at $x = \bar{x}$ of the Lagrangian function for problem 3.1 defined for a multiplier vector $\bar{\lambda}$. From the second order sufficiency conditions, we know that $\nabla^2 \mathcal{L}(\bar{x})$ is positive definite on the cone $C = \{d \neq 0 : \nabla h_i(\bar{x})^t d = 0, \quad \textit{for} \quad i = 1, \ldots, l\}$.

Now, on the contrary, if there does not exit a $\bar{\mu}$ such that $H(\bar{x})$ is positive definite for $\mu \geq \bar{\mu}$, then it must be the case that given any $\mu_k = k, \quad k = 1, 2, \ldots$ there exists a d_k with $||d_k|| = 1$ such that

$$d_k^t H(\bar{x}) d_k = d_k^t \nabla^2 \mathcal{L}(\bar{x}, \bar{\lambda}) d_k + k \Sigma_{i=1}^{l} [\nabla^t h_i(\bar{x}) d_k]^2 \leq 0, \tag{3.7}$$

and therefore,

$$\Sigma_{i=1}^{l} [\nabla^t h_i(\bar{x}) d_k]^2 \leq -\frac{1}{k} d_k^t \nabla^2 \mathcal{L}(\bar{x}, \bar{\lambda}) d_k \tag{3.8}$$

Since $||d_k|| = 1$ for all k, there exits a convergent subsequence for $\{d_k\}$ with limit point $\bar{d}$, where $||\bar{d}|| = 1$. From this, we observe that $\{d_k\}$ lie in a compact set (the surface of the unit sphere). Over this subsequence, since the right hand side term of 3.8 approaches zero as $k \to \infty$, we must have that $\nabla h_i(\bar{x})^t \bar{d} = 0$ for all $i = 1, \ldots, l$ for 3.8 to hold true for all k. Hence, $\bar{d} \in C$.

Moreover, since ${d_k}^t \nabla^2 \mathcal{L}(\bar{x}, \bar{\lambda}) d_k \leq 0$ for all k by 3.7, we have $\bar{d}^t \nabla^2 \mathcal{L}(\bar{x}, \bar{\lambda}) \bar{d} \leq 0$. This contradict the second order sufficiency conditions. Consequently, $H(\bar{x})$ is positive definite for μ exceeding some value $\bar{\mu}$, so by the second order KKT sufficient condition, $\bar{x}$ is a strict local minimizer for $\mathcal{L}_\mu(., \bar{\lambda})$.

Note that, in particular, if f is convex and h_i, for $i = 1, \ldots, l$, are affine, then any minimizing solution $\bar{x}$ for P also minimizes $\mathcal{L}_\mu(., \bar{\lambda})$ for all $\mu \geq 0$.

We can also use matrix properties 1.1.3 and 1.1.4 to prove this Theorem.

Example 3.1.1 *Consider the problem*

$$\begin{array}{ll} \textit{Minimize} & x_1^2 + x_2^2 \\ \textit{subject to} & x_1 + x_2 - 1 = 0 \end{array} \tag{3.9}$$

Solution 3.1.1 *The Lagrangian function is given by:*

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

By using the first order KKT necessary conditions, we have the multiplier $\lambda = -1$ and the KKT solution $x = (1/2, 1/2)$, which is the minimizer of the original problem. Furthermore, the augmented Lagrangian function for this problem is,

$$\begin{aligned} \mathcal{L}_\mu(x, \lambda) &= x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 1) + \frac{\mu}{2}(x_1 + x_2 - 1)^2 \\ &= (x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2 + \frac{\mu}{2}(x_1 + x_2 - 1)^2 + \frac{1}{2}, \end{aligned}$$

(by substituting $\lambda = -1$).

This implies that,

$$\begin{aligned} \nabla_{x_1}\mathcal{L}_\mu(x, \lambda) &= 2x_1 + \mu(x_1 + x_2 - 1) - 1 = 0, \\ \nabla_{x_2}\mathcal{L}_\mu(x, \lambda) &= 2x_2 + \mu(x_1 + x_2 - 1) - 1 = 0 \end{aligned}$$

These conditions are true for all positive values of μ at the optimal solution $x = (1/2, 1/2)$. Hence, the previous theorem is verified, i.e., without making μ infinity (or by taking large enough value) it is possible to determine the minimizer of the original constrained problem. This example verifying also the second statement of the theorem.

Remark: Without the second order sufficiency conditions of the previous theorem, there might be not exist any finite value of μ that will recover an optimum $\bar{x}$ for problem 3.1, and it might be that we need to take $\mu \to \infty$ for this to occur.

Example 3.1.2 *Consider problem P to minimize $f(x) = x_1^4 + x_1x_2$, subject to $x_2 = 0$. Its Lagrangian function is defined by $\mathcal{L}(x, \lambda) = x_1^4 + x_1x_2 + \lambda x_2$. Then*

$$\nabla\mathcal{L}(x, \lambda) = (4x_1^3 + x_2, x_1 + \lambda)^t = (0, 0)^t.$$

From this, $\bar{x} = (0, 0)^t$ is the optimal solution, and we also obtain $\bar{\lambda} = 0$ as the unique Lagrangian multiplier.

The Hessian of the Lagrangian is given by

$$\nabla^2\mathcal{L}(\bar{x}) = \nabla^2 f(\bar{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Let us check its positivity by taking a non-zero vector $d = (a, 0)^t$, which satisfies the condition $\nabla^t h(x)d = 0$ for any non-zero scalar a. Then

$$d^t\nabla^2\mathcal{L}(\bar{x})d = 0.$$

Therefore, $\nabla^2\mathcal{L}(\bar{x})$ is indefinite, the second order sufficiency condition does not hold at $(\bar{x}, \bar{\lambda})$.

Now, we have $\mathcal{L}_\mu(\bar{x}, \bar{\lambda}) = x_1^4 + x_1x_2 + \frac{\mu}{2}x_2^2$. Note that for any $\mu > 0$, the gradient of the augmented Lagrangian function,

$$\nabla\mathcal{L}_\mu(\bar{x}, \bar{\lambda}) = \begin{bmatrix} 4x_1^3 + x_2 \\ x_1 + 2\mu x_2 \end{bmatrix},$$

vanishes at $\bar{x} = (0, 0)$ and $\hat{x} = (\frac{1}{\sqrt{8\mu}}, \frac{-1}{2\mu\sqrt{8\mu}})$. Furthermore,

$$\nabla^2\mathcal{L}_\mu(x, \bar{\lambda}) = \begin{bmatrix} 12x_1^2 & 1 \\ 1 & 2\mu \end{bmatrix}.$$

We see that $\nabla^2\mathcal{L}_\mu(x, \bar{\lambda})$ is indefinite at the critical point $\bar{x}$ and hence $\bar{x}$ is not a local minimizer for any $\mu > 0$. However, $\nabla^2\mathcal{L}_\mu(x, \bar{\lambda})$ is positive define at the critical point $\hat{x}$, and thus $\hat{x}$ is the minimizer of $\mathcal{L}_\mu$ for all $\mu > 0$. Moreover, as $\mu \to \infty$, $\hat{x}$ approaches the minimizer, $(0, 0)$, of the constrained problem P.

3.1.2 Algorithmic Scheme for Augmented Lagrangian Penalty Method

We now design an algorithm that fixes the penalty parameter μ to some value $\mu_k > 0$, fixes λ at the current estimate λ^k, and performs minimization with respect to x.

Minimizing the Auxiliary Function $\mathcal{L}_\mu(., \lambda)$

The best choice here is the Newton method with line searches or the modified Newton method, when the second order information on the objective and the constraints is available. If it is not the case, we could use Quasi-Newton methods, Conjugate Gradient, etc. When solving the auxiliary problems, we should use reasonable safeguard policy in order to ensure that our minimization is indeed local.

The simplest policy here is to use the solution to the previous auxiliary problem as the starting point when solving the current one and to take specific care of line searches (e.g., to move until the closest minimum of the current auxiliary objective on the current search line). Using the previous iterate as the starting point in the new auxiliary primal problem saves a lot of computational effort.

Updating the Multiplier λ

Let $\{\lambda^k\}$ be bounded sequence and let $0 < \mu_k < \mu_{k+1}$. Then using x^k to denote the approximate minimizer of $\mathcal{L}_{\mu_k}(x, \lambda^k)$ which approaches to a KKT optimal solution $\bar{x}$ to the given equality constrained problem, we have by the optimality conditions for unconstrained minimization that

$$\begin{aligned} 0 \approx \nabla_x\mathcal{L}_{\mu_k}(x^k, \lambda^k) &= \nabla f(x^k) + \Sigma_1^l \lambda_i^k \nabla h_i(x^k) + \Sigma_1^l \mu_k h_i(x^k)\nabla h_i(x^k) \\ &= \nabla f(x^k) + \Sigma_1^l [\lambda_i^k + \mu_k h_i(x^k)]\nabla h_i(x^k). \end{aligned}$$

This implies that

$$0 = \nabla_x \mathcal{L}(\bar{x}, \bar{\lambda}) = \nabla_x \mathcal{L}_{\mu_k}(\bar{x}, \bar{\lambda}) \approx \nabla_x \mathcal{L}_{\mu_k}(x^k, \lambda^k).$$

Thus,

$$\nabla f(\bar{x}) + \Sigma_1^l \bar{\lambda}_i \nabla h_i(\bar{x}) \approx \nabla f(x^k) + \Sigma_1^l [\lambda_i^k + \mu_k h_i(x^k)] \nabla h_i(x^k),$$

where $\bar{\lambda}$ is the exact Lagrangian multiplier of problem 3.1 corresponding to the optimal solution $\bar{x}$, limit point of the sequence $\{x^k\}$. Since f and h are continuous functions,

$$\nabla f(\bar{x}) + \Sigma_1^l \bar{\lambda}_i \nabla h_i(\bar{x}) = \nabla f(\bar{x}) + \Sigma_1^l \lim_{k\to\infty} [\lambda_i^k + \mu_k h_i(x^k)] \nabla h_i(\bar{x}).$$

From this equations we have the following:

$$\Sigma_1^l \bar{\lambda}_i \nabla h_i(\bar{x}) = \Sigma_1^l \lim_{k\to\infty} [\lambda_i^k + \mu_k h_i(x^k)] \nabla h_i(\bar{x}).$$

Hence,

$$\lambda_i^k + \mu_k h_i(x^k) \approx \bar{\lambda}_i, \quad \text{for all} \quad i = 1, \ldots, l. \tag{3.10}$$

By rearranging this expression, we have that

$$h_i(x^k) \approx \frac{1}{\mu_k}(\bar{\lambda}_i - \lambda_i^k), \quad \text{for all} \quad i = 1, \ldots, l,$$

so we conclude that if λ^k is close to the optimal multiplier vector $\bar{\lambda}$, the infeasibility (or constraint violation) in x^k will be much smaller than $\frac{1}{\mu_k}$, rather than being proportional to $\frac{1}{\mu_k}$ as for the penalty method.

The relation 3.10 suggests a formula for improving our current estimate λ^k of the Lagrange multiplier vector, using the approximate minimizer x^k just calculated: We can set

$$\lambda_i^{k+1} = \lambda_i^k + \mu_k h_i(x^k), \quad \text{for all} \quad i = 1, \ldots, l. \tag{3.11}$$

Updating Penalty Parameter μ

We know that all large enough values of the penalty parameter satisfies the requirement to minimize the given augmented Lagrangian problem, but how to choose an appropriate μ in practice? Advice like "set μ something huge" definitely is of no interest: with this approach, we convert the method into the penalty one and loose all computational advantages associated with possibility to deal with moderate values of the penalty.

The main considerations to be kept in mind for selecting the penalty parameter sequence are the following, [4]:

i. μ should eventually become larger than the threshold level necessary to bring to bear the positive feature of the multiplier iteration.

ii. The initial parameter μ_0 should not be too large to the point where it causes ill-conditioning at the first unconstrained minimization.

iii. μ should not be increased too fast to the point where too much ill-conditioning is forced upon the unconstrained minimization routine too early.

iv. μ should not be increased too slowly, at least in the early minimizations, to the extent that the multiplier iteration has poor convergence rate.

The simplest way to adjust the penalty parameter is as follows: From 3.11 for appropriately large μ the quantities $||h(x^k)||$ decrease with the ratio which can be done arbitrarily close to 0, provided that μ is large enough. Let us use this observation as the driving force in our policy for adjusting the penalty parameter, and start the procedure with certain ad hoc value of μ and look whether the residuals $||h(x^k)||$ indeed reasonably decrease, say, at least by factor 0.25 each time. Until it is so, we keep μ constant; and if at certain step k the current value of $||h(x^k)||$ is too large, exceeds $0.25||h(x^k)||$, we interpret this as a signal to increase the penalty by certain factor, say, 10. Then we resolve the subproblem with this new value of the penalty and the same as before value of λ, again look whether we got required progress in constraint violation, and if not, again increase the penalty, and so on.

In general, there is no safe guideline as to what is a suitable initial value of μ, so one may have to resort to trial and error in order to determine the value of this parameter, [4].

Summery of the Algorithm

i. Choose $\mu_1 > 0$, tolerance $\epsilon > 0$, starting point x and multiplier vector λ^1, a constant $\gamma > 1$, set $k = 1$;

ii. Find an approximate minimizer x^k for the subproblem $\mathcal{L}_{\mu_k}(., \lambda^k)$;

iii. Terminate the procedure when $||\nabla_x \mathcal{L}_{\mu_k}(, \lambda^k)|| \leq \epsilon$ at $x = x^k$; x^k is an approximate KKT solution;

iv. If a convergence test for (iii) is satisfied stop, with approximate solution x^k; otherwise, go to the next step;

v. Update the multipliers using $\lambda_i^{k+1} = \lambda_i^k + \mu_k h_i(x^k), \quad \text{for all} \quad i = 1, \ldots, l$;

vi. Choose new penalty parameter (if it is necessary), $\mu_{k+1} = \gamma \mu_k$;

vii. Set starting point for the next iteration to $x^{k+1} = x^k$;

viii. Increment k by 1 and return to step ii.

3.2 Augmented Lagrangian Methods for Problems with both Equality and Inequality Constraints

The most popular augmented Lagrangian function for inequality constraints can be obtained by reducing the inequality constrained problem to an equality constrained one, with the

help of squared slack variables, and applying the equality augmented Lagrangian methods to the new problem. After some manipulation, squared slack variables are eliminated and an augmented Lagrangian without auxiliary variables arises.

Consider a nonlinear programming problem involving both equality and inequality constraints.

$$\begin{array}{rl} \text{Minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0 \quad \text{for} \quad i = 1, \ldots, m; \\ & h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, l;, \end{array} \tag{3.12}$$

where, $f, g_1, \ldots, g_m, h_1, \ldots, h_l$ are continuous functions defined on $\Re^n$. We assume that this problem has a well-defined solution x^*, which is a regular point of the constraints and which satisfies the second-order sufficiency conditions for a local minimum. It is possible to convert this problem to equality constrained problem by introducing a vector of additional variables $s = (s_1, \ldots, s_m)$. So, we can rewrite the above problem as:

$$\begin{array}{rl} \text{Minimize} & f(x) \\ \text{subject to} & g_i(x) + s_i^2 = 0 \quad \text{for} \quad i = 1, \ldots, m; \\ & h_i(x) = 0 \quad \text{for} \quad i = 1, \ldots, l;, \end{array} \tag{3.13}$$

Through this conversion we can hope to simply apply the theory for equality constrained problems to problems with inequality constraints. In order to do so we must insure that 3.13 satisfies the second-order sufficiency conditions. These conditions will not hold unless we impose a strict complementarity assumption that $u_i^* g_i(x^*) = 0$ implies the corresponding multiplier $u_i^* > 0$ as well as the usual second-order sufficiency conditions for the original problem 3.12. We have that x^* is a local minimizer of 3.12 if and only if $(x^*, s_1^*, \ldots, s_m^*)$, where $s_i^* = \sqrt{-g_i(x^*)}$ for $i = 1, \ldots, m$, is a local minimizer of problem 3.13.

Now, based on the equality constrained problem in the previous section, the augmented Lagrangian function for problem 3.13 can be written as:

$$\begin{aligned} \bar{\mathcal{L}}_\mu(x, s, \lambda, u) := f(x) + \Sigma_{i=1}^l \lambda_i h_i(x) + \frac{\mu}{2} \Sigma_{i=1}^l h_i^2(x) + \Sigma_{i=1}^m u_i [g_i(x) + s_i^2] \\ + \frac{\mu}{2} \Sigma_{i=1}^m (g_i(x) + s_i^2)^2, \end{aligned} \tag{3.14}$$

where λ_i and $u_i > 0$ are multipliers corresponding to the constraint h_i, for $i = 1, 2, \ldots, l$ and g_i, for $i = 1, 2, \ldots, m$ respectively with the penalty parameter $\mu > 0$.

This augmented Lagrangian function must be minimized with respect to (x, s) for various values of λ, u and μ. First minimizing $\bar{\mathcal{L}}_\mu(x, s, \lambda, u)$ over s_i in terms of x for each $i = 1, 2, \ldots, m$, and then minimizing the resulting expression over $x \in \Re^n$. Note that

$$\begin{aligned} \min_s \bar{\mathcal{L}}_\mu(x, s, \lambda, u) := f(x) + \Sigma_{i=1}^l \lambda_i h_i(x) + \frac{\mu}{2} \Sigma_{i=1}^l h_i^2(x) \\ + \Sigma_{i=1}^m \min_{s_i} \{ u_i [g_i(x) + s_i^2] + \frac{\mu}{2} [g_i(x) + s_i^2]^2 \} \end{aligned} \tag{3.15}$$

The minimization with respect to s_i is equivalent to

$$\min_{t_i \geq 0} \{u_i[g_i(x) + t_i] + \frac{\mu}{2}[g_i(x) + t_i]^2\}. \tag{3.16}$$

The function in braces is quadratic in t_i. Its unconstrained (global) minimizer is the scalar $\bar{t_i}$ at which the derivative is zero. Thus,

$$u_i + \mu(g_i(x) + \bar{t_i}) = 0$$

from which

$$\bar{t_i} = -[\frac{u_i}{\mu} + g_i(x)].$$

From this expression, either $\bar{t_i} \geq 0$ in which case $\bar{t_i}$ solves problem 3.16, or else the solution of problem 3.16 is $t_i^* = 0$. The solution of problem 3.16 is

$$t_i^* = \max\{0, -[\frac{u_i}{\mu} + g_i(x)]\}. \tag{3.17}$$

This implies that

$$g_i(x) + t_i^* = \max\{-\frac{u_i}{\mu}, g_i(x)\}. \tag{3.18}$$

Then, 3.15 can be rewritten as:

$$\begin{aligned} \min_s \bar{\mathcal{L}}_\mu(x, s, \lambda, u) = f(x) + \Sigma_{i=1}^l \lambda_i h_i(x) + \frac{\mu}{2}\Sigma_{i=1}^l h_i^2(x) \\ + \Sigma_{i=1}^m u_i \max\{-\frac{u_i}{\mu}, g_i(x)\} + \frac{\mu}{2}\Sigma_{i=1}^m \max^2\{-\frac{u_i}{\mu}, g_i(x)\}. \end{aligned} \tag{3.19}$$

We are thus led to the following definition of the augmented Lagrangian for 3.13

$$\begin{aligned} \mathcal{L}_\mu(x, \lambda, u) = f(x) + \Sigma_{i=1}^l \lambda_i h_i(x) + \frac{\mu}{2}\Sigma_{i=1}^l h_i^2(x) \\ + \Sigma_{i=1}^m u_i \max\{-\frac{u_i}{\mu}, g_i(x)\} + \frac{\mu}{2}\Sigma_{i=1}^m \max^2\{-\frac{u_i}{\mu}, g_i(x)\}. \end{aligned} \tag{3.20}$$

The conclusion from the preceding discussion is that the problem

$$\text{minimize} \quad \bar{\mathcal{L}}_\mu(x, s, \lambda, u) \quad \text{subject to} \quad (x, s) \in \Re^{n+m} \tag{3.21}$$

is equivalent to the problem

$$\text{minimize} \quad \mathcal{L}_\mu(x, \lambda, u) \quad \text{subject to} \quad x \in \Re^n, \tag{3.22}$$

and $[x(\lambda, u, \mu), s(\lambda, u, \mu)]$ is a solution of 3.13 if and only if $x(\lambda, u, \mu)$ is a solution of problem 3.22 and

$$s_i^2(\lambda, u, \mu) = \max\{0, -[\frac{u_i}{\mu} + g_i(x(\lambda, u, \mu))]\} \quad \text{for all} \quad i = 1, \ldots, m.$$

Since we can solve 3.22 in place of 3.21, the computation of 3.13 need not involve the additional variables $s_1, \ldots, s_m$.

The methods of augmented Lagrangian function for equality constrained problem can also be applied to problem with inequality constraint what we have discussed in this section.

Proposition 3.2.1 *Consider problem P to minimize $f(x)$ subject to $h_i(x) = 0$ for $i = 1, \ldots, l$ and $g_i(x) \leq 0$ for $i = 1, \ldots, m$, and let the KKT solution $(\bar{x}, \bar{\lambda}, \bar{u})$ satisfy the second order sufficiency conditions of problems with both equality and inequality constraints for a local minimum. Then there exists a $\bar{\mu}$ such that for $\mu \geq \bar{\mu}$, the augmented Lagrangian penalty function $\mathcal{L}_\mu(., \bar{\lambda}, \bar{u})$, defined for some accompanying $\lambda = \bar{\lambda}$ and $u = \bar{u}$, also achieves a strict local minimum at $\bar{x}$.*

In view of proposition 3.2.1, we can extend the algorithms and analysis of the previous section.

Suppose that x^k is the minimizer of the k^{th} subproblem $\mathcal{L}_{\mu_k}(., \lambda^k, u^k)$. Then the multipliers for the next step are given by:

$$\lambda_i^{k+1} = \lambda_i^k + \mu_k h(x^k) \quad \text{for} \quad i = 1, \ldots, l,$$

$$u_i^{k+1} = u_i^k + \mu_k max\{-\frac{u_i^k}{\mu_k}, g_i(x^k)\} \quad \text{for} \quad i = 1, \ldots, m.$$

Conclusion

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.

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 lies 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".

Though exterior and interior 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.

Although penalty and barrier methods met with great initial success, their slow rates of convergence due to ill-conditioning of the associated Hessian led researchers to pursue other approaches.

The augmented Lagrangian methods are among these methods which are general approach to the solution of constrained minimization problems. Their major advantages over the penalty function methods and barrier function methods are the problem conditioning and convergence rate. Generally, the methods are characterized by the following attractive properties:

- when properly implemented, the convergence rate on the multiplier variables is very satisfactory and the constraints can be satisfied with high precision;
- convergence is attained without requiring the penalty parameter to diverge;
- in the penalty method, the solution is a function of the penalty parameter. Therefore the solution depends on the penalty parameter value, being related to the allowed constraint violation. Conversely, in the augmented Lagrangian method, the solution is independent of the value of the penalty parameter.

The augmented Lagrangian function for problems with inequality constraints can be obtained by reducing the inequality constrained problem to an equality constrained one, with the help of squared slack variables, and applying the equality augmented Lagrangian procedures to the new problem.

Appendix-A

References

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

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

[3] A.D. Belegundu, T.R. Chandrupatla. *Optimization Concepts and Applications in Engineering.* Cambridge University Press, Second Edition, 2011.

[4] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York, 1982.

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

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

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

[8] E.M.T. Hendrix, B.G.Toth. *Introduction to Nonlinear and Global Optimization.* Springer Science+Business Media, LLC, 2010.

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

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

[11] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer Series in Operations Research. Springer Verlag, New York, second edition, 2006.

[12] P. Pedregal. *Introduction to Optimization.* Springer-Verlag New York, Inc, 2004.

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

[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.