

AN ABSTRACT OF THE THESIS OF

Surendra P. S. Bhatia for the degree of Doctor of Philosophy

in Electrical and Computer Engineering presented on Aug 16, 77

Title: A NEW COMPUTATIONAL METHOD FOR CONSTRAINED  
NONLINEAR PROGRAMMING

Redacted for Privacy

Abstract approved: \_\_\_\_\_  
\_\_\_\_\_ Pieter A. Frick \_\_\_\_\_

In this dissertation, a new computational method for equality constrained nonlinear programming is developed. Specifically, problems of minimizing a nonlinear objective function  $f(x)$  subject to constraints  $h(x) = 0$ , where  $f$  is a scalar valued,  $x$  is an  $n$ -vector and  $h$  is a  $m$ -vector with  $m < n$ , is considered.

The approach employed is based on the idea of seeking simultaneous satisfaction of the associated necessary conditions and the constraints. This is achieved by modifying the normal penalty method approximation to the Lagrange multipliers associated with the problem. As a result of this modification, the new method permits a solution of the constrained problem through a single minimization of an equivalent unconstrained problem, for almost any choice of the associated penalty parameter  $\rho$ .

To evaluate the effectiveness and efficiency of the new method, its performance is compared with that of Hestenes' Multiplier and Fletcher's Exact Penalty Methods in the computational solution of a number of nonlinear programming examples. An important computational advantage of the proposed method over the two aforementioned methods is that it does not require the computation of the critical value  $\rho^*$ , of the parameter  $\rho$  and provides almost unlimited freedom in its choice. The new method is shown to have superior convergence properties.

A New Computational Method for Constrained  
Nonlinear Programming

by

Surendra P. S. Bhatia

A THESIS

submitted to

Oregon State University

in partial fulfillment of  
the requirements for the  
degree of

Doctor of Philosophy

June 1978

APPROVED:

Redacted for Privacy

---

Assistant ~~Professor~~ of Electrical and Computer  
Engineering in charge of major

Redacted for Privacy

---

Head of Department of Electrical and Computer  
Engineering

Redacted for Privacy

---

Dean of Graduate School

Date thesis is presented

August 16, 1977

Typed by Lyndalu Sikes for Surendra P. S. Bhatia

## ACKNOWLEDGEMENT

I wish to express my gratitude to my major professor Dr. Pieter A. Frick for his help, keen interest and invaluable guidance throughout the research work leading to the completion of this thesis. His ideas lead the thesis topic. My sincere thanks also to Professor R. R. Mohler for his encouragement and continued interest.

I also wish to express my hearty thanks to all my friends in and around Corvallis for their invaluable moral support. In particular, I wish to thank Messers Weldon A. Lodwick, J. Chakravarty, K. V. Rao and C. S. Hsu for their time and valuable suggestions.

Last but not the least, my family. According to a culture ones family is supposed to be his own part and in accordance with another culture one is not supposed to thank himself. Hence my deep appreciation to my wife Veena and my two daughters, Shaloo and Sonia, for their patience and understanding during the many lost evenings and weekends.

## TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
II. REVIEW OF RELEVANT PAST RESEARCH	3
2.1 Optimality Conditions	3
2.2 Equality Constrained Nonlinear Programs	13
2.3 Methods of Solving Constrained Nonlinear Programs	15
III. A NEW COMPUTATIONAL METHOD FOR CONSTRAINED NONLINEAR PROGRAMMING PROBLEMS	34
3.1 Motivation	34
3.2 The Intuitive Idea Behind the Proposed Method	37
3.3 The New Method	40
3.4 Computational Implementation and Comments	45
IV. COMPUTATIONAL RESULTS	52
4.1 Constrained Minimization Algorithms	52
4.2 Experimental Conditions and Numerical Examples	58
4.3 Computational Results and Interpretation	62
V. CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH	69
BIBLIOGRAPHY	71
Appendix A - An Unconstrained Minimization Algorithm	77
Appendix B - Fibonacci Search Method for Minimization of a Single-Variable Function	78

## LIST OF TABLES

<u>Table</u>		<u>Page</u>
1.	Results obtained for the Semi-Dual Method.	63
2.	Results obtained using the Multiplier Method.	63
3.	Results obtained using the Exact Penalty Method.	63

## LIST OF ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
1.	Solutions to problem number 4.2.1 for $\rho = 0.1$ .	65
2.	Solutions to problem number 4.2.2 for $\rho = 0.1$ .	66
3.	Solutions to problem number 4.2.3 for $\rho = 0.1$ .	67
4.	Solutions to problem number 4.2.4 for $\rho = 0.1$ .	68



### Convention and Notation

The following conventions and notations are used in this dissertation:  $\mathbb{R}^n$  denotes the Euclidean space of ordered  $n$ -tuples of real numbers;  $\mathbb{R}$  denotes  $\mathbb{R}^1$ ,  $\mathbf{x} \in \mathbb{R}^n$  has components  $x_1, x_2, \dots, x_n \in \mathbb{R}$ . All vectors are assumed to be column vectors. Superscripts on vectors are used in order to distinguish different vectors. In order to avoid confusion, exponents on real numbers will be used together with parentheses.  $\langle \cdot, \cdot \rangle$  denotes the inner product in  $\mathbb{R}^n$ , defined by  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i$ .  $\|\cdot\|$  denotes the norm in  $\mathbb{R}^n$ , defined by  $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$ .  $f$  or  $f(\cdot)$  denotes a function.  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  denotes domain of  $f$  in  $\mathbb{R}^n$  and its range in  $\mathbb{R}^m$ . Given  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , if each component function has continuous partial derivatives of order  $p$ , we write  $f \in C^p$ . Given  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ , the function  $\nabla f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is called the gradient of  $f$  and is regarded as a column vector. If  $f \in C^2$ , the second derivative of  $f$ , denoted by  $\nabla^2 f$  or  $F$ , is called the Hessian of  $f$  and is defined to be the  $(n \times n)$  matrix  $\left[ \frac{\partial^2 f}{\partial x_i \partial x_j} \right]$ . If  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $f \in C^1$ , then the first derivative of  $f$  is defined as the  $(n \times m)$  matrix  $\nabla f = \left[ \frac{\partial f_i}{\partial x_j} \right]$ . If  $f \in C^2$ , the second derivative of  $f$  is a third order tensor. Since the explicit use of the second derivative is not required, we consider it as a set of  $m$  Hessians  $F_1, F_2, \dots, F_m$ , corresponding to  $m$  component functions.  $A^T$  denotes the transpose of a matrix  $A$  and

$A^{-1}$  is used to denote the inverse of  $A$ . The norm of a matrix  $A$  is taken as  $\|A\| = \sup_x \{\|Ax\| : \|x\| = 1\}$ .

Symbols:

$\forall$	All.
$\exists$	There exists.
$\triangleq$	Defined by.
$\ni$	Such that.
$\{x:P\}$	Set of points having property $P$ .
$z \in A$	$z$ belongs to $A$ .
$\cup, \cap$	Union, intersection.
$A \subset B$	$A$ contained in $B$ .
$\emptyset$	Empty set.
$\phi$	Phi.
$N_\delta(x)$	Spherical neighborhood of a point $x$ with radius $\delta$ .
$I_m$	$(m \times m)$ identity matrix.

# A NEW COMPUTATIONAL METHOD FOR CONSTRAINED NONLINEAR PROGRAMMING

## I. INTRODUCTION

A concept that appears frequently in the analysis of many complex decision problems, is that of optimization--the best way to carry out action or produce results. Using optimization concepts, a systematic approach to complex decision problems is possible and normally involves the selection of values for a number of inter-related variables, by focusing attention on a single objective designed to quantify performance and measure the quality of decisions. It is rarely possible to fully represent all the complexities of variable interactions, constraints and appropriate objectives when faced with a complex decision problem. Nevertheless an accurate representation of complexities is desirable and such a representation often results in mathematical models involving nonlinear functions. The analysis and efficient computational methods of nonlinear programs have proved to be of interest to practitioners in such diverse fields as Operation research, Management science, Engineering, Economics, Social science, System analysis and Computer science.

In this dissertation, a new computational method is developed to solve equality constrained nonlinear programming problems of the type:

minimize  $\{f(x) : x \in \mathbb{R}^n, h_i(x) = 0, i = 1, 2, \dots, m\}$  .

The material in this thesis is organized into five chapters. Chapter II contains a summary of past work, on the analysis and methods of nonlinear programming, relevant to our later discussion. The chapter contains two parts; the first part briefly describes constraints qualifications, the necessary and sufficient conditions for optimality of the general nonlinear program and the nonlinear program with equality constraints. In the second part, some numerical solution techniques for nonlinear programs are described.

Chapter III is the core of the dissertation and describes a new computational method for equality constrained nonlinear programming. The intuitive development and a convergence proof for the method is described and the algorithm is explained in detail describing the computational implementations.

Chapter IV contains computational results of applying this method to a set of nonlinear test problems. For comparison purposes the test problems are also solved using the Multiplier and the Exact Penalty methods. The computational results are illustrated by means of graphs and tables.

Chapter V contains some conclusions and suggests topics for further investigation.

## II. REVIEW OF RELEVANT PAST RESEARCH

Starting from a very general nonlinear programming problem statement, a systematic review of optimality conditions and solution methods, relevant to the development of later chapters, is provided here.

Consider the following nonlinear programming problem:

$$(A) \begin{cases} \text{minimize } f(x) \\ \text{subject to the constraints } g(x) \geq 0, \quad h(x) = 0, \end{cases}$$

where  $x \in \mathbb{R}^n$  and the real-valued functions;  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g: \mathbb{R}^n \rightarrow \mathbb{R}^p$  and  $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , here assumed to be defined on some open set  $D \subset \mathbb{R}^n$ . The functions  $f$ ,  $g$  and  $h$  can take on any form of nonlinearity, but are assumed to be continuous and differentiable on  $D$ .

The point  $x^*$  of the feasible set  $X$ , where  $X \triangleq \{x: x \in D, g(x) \geq 0, h(x) = 0\}$ , is called the local solution of the problem (A), if  $\exists$  a  $\delta > 0$  such that  $f(x) \geq f(x^*)$  for  $\forall x \in X \cap N_\delta(x^*)$ . If  $f(x) \geq f(x^*)$  for  $\forall x \in X$ , then  $x^*$  is the global solution of the problem (A).

### Section 2.1 Optimality Conditions

Let  $f, g$  and  $h \in C^1$  and define the Lagrangian associated with the problem (A), as

$$L(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x)$$

where  $\lambda \in \mathbb{R}^p$ ,  $\mu \in \mathbb{R}^m$ .

Let  $x^* \in X$ . Consider a nonzero vector  $z \in \mathbb{R}^n$  from  $x^*$ . The vector  $z$  is called a feasible direction vector if  $\exists$  a number  $\delta_1$  such that  $(x^* + \theta z) \in X \cap N_{\delta_1}(x^*)$  for all  $\theta$ ,  $0 \leq \theta < \delta_1 / \|z\|$ . Consider the set

$$Z^1(x^*) \triangleq \{z : z^T \nabla g_i(x^*) \geq 0, i \in I(x^*), z^T \nabla h_j(x^*) = 0, j=1, 2, \dots, m\},$$

where  $I(x^*) \triangleq \{i : g_i(x^*) = 0\}$  and the set

$$Z^2(x^*) \triangleq \{z : z^T \nabla f(x^*) < 0\}.$$

It is simple to show that if  $z$  is a feasible direction vector then  $z \in Z^1(x^*)$  and that if  $z \in Z^2(x^*)$  then  $\exists$  a point  $x = x^* + \theta z$ , sufficiently close to  $x^*$ , such that  $f(x) > f(x^*)$ .

In terms of  $Z^1(x^*)$  and  $Z^2(x^*)$ , we state the first-order necessary conditions for a local solution to the problem (A). The proof is contained in [15].

Theorem 2.1.1 [15] (First-order Necessary Conditions; Existence of Generalized Lagrange Multipliers)

Suppose  $x^* \in X$ . Then  $Z^1(x^*) \cap Z^2(x^*) = \emptyset$  if and only if there exist vectors  $\lambda^*$  and  $\mu^*$  such that

$$\nabla_x L(x^*, \lambda^*, \mu^*) = \nabla f(x^*) + \nabla g(x^*) \lambda^* + \nabla h(x^*) \mu^* = 0 \quad (2.1.1)$$

$$\lambda_i^* g_i(x^*) = 0 \quad (2.1.2)$$

$$\} \quad i = 1, 2, \dots, p.$$

$$\lambda_i^* \geq 0 \quad (2.1.3)$$

The condition  $Z^1(x^*) \cap Z^2(x^*) = \emptyset$  is necessary and sufficient for the existence of generalized lagrange multipliers  $(\lambda^*, \mu^*)$  such that conditions (2.1.1) - (2.1.3) hold. In general, however, the set  $Z^1(x^*) \cap Z^2(x^*)$  may not be empty [36] at the solution point  $x^*$  to the problem (A) and, therefore, the lagrangian conditions, (2.1.2) - (2.1.3), may not hold at  $x^*$ . Nonetheless, it is possible to derive "weak" necessary conditions for optimality without requiring  $Z^1(x^*) \cap Z^2(x^*)$  to be empty. Fritz John [32] and Mangasarian and Fromovitz [41] proved the following necessary conditions using a weak-lagrangian defined as

$$\tilde{L}(x, \lambda, \mu) = \lambda_0 f(x) + \lambda^T g(x) + \mu^T h(x).$$

Theorem 2.1.2 [41]:--Suppose  $f$ ,  $g$  and  $h$  are continuously differentiable on an open set containing  $X$ . If  $x^* \in X$  is a solution of the problem (A), then there exists a  $\lambda^{1*} = (\lambda_0^*, \lambda_1^*, \dots, \lambda_p^*)^T$  and a  $\mu^* \in R^m$  such that

$$\nabla_{\mathbf{x}} \tilde{L}(\mathbf{x}^*, \lambda^*, \mu^*) = \lambda_0^* \nabla f(\mathbf{x}^*) + \nabla g(\mathbf{x}^*) \lambda^* + \nabla h(\mathbf{x}^*) \mu^* = 0 \quad (2.1.4)$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0 \quad i = 1, 2, \dots, p. \quad (2.1.5)$$

$$(\lambda^{1*}, \mu^*) \neq 0, \quad \lambda^{1*} \geq 0. \quad (2.1.6)$$

The main weakness of the above theorem lies in the fact that for a suitable choice of  $\lambda_0^*$ ,  $\lambda^*$  and  $\mu^*$ , the conditions of the theorem might be satisfied for any differentiable function  $f$  whether it has a local minimum at the point or not [49].

In order to guarantee that  $Z^1(\mathbf{x}^*) \cap Z^2(\mathbf{x}^*)$  is empty, the constraining functions  $g$  and  $h$  of the problem (A) need to be qualified. Several conditions have been imposed to ensure that the set  $Z^1(\mathbf{x}^*) \cap Z^2(\mathbf{x}^*)$  is empty at the solution point. Kuhn and Tucker, in their fundamental paper [36], presented necessary conditions for inequality constrained nonlinear programs that are stronger than those of John's in Theorem 2.1.2. Restricting the constraint-functions by a regularity condition, called Constraint-qualification, they were able to ensure that the multiplier  $\lambda_0$  is indeed positive and the extended version of John's necessary conditions, equations (2.1.4) - (2.1.6), becomes equivalent to conditions (2.1.1) - (2.1.3). The type of restrictions imposed on the constraint-functions to ensure the existence of the Lagrange multipliers, at the solution of a nonlinear program, has been the subject of intensive research efforts. In



addition to the original constraint qualifications of Kuhn and Tucker, several other qualifications were proposed. For an extensive discussion on the various constraint qualifications see, for example, Bazaraa, Goode and Shetty [2] and Gould and Tolle [25]. In subsequent discussion on constraint qualifications, we follow the work of Abadie [1], Gould and Tolle [26] and Varaiya [62], as summarized in [49] and state some results relevant to our discussion.

Let  $A$  be a non-empty subset of  $\mathbb{R}^n$  and let  $x \in A$ . The cone of tangents,  $\tilde{S}(A, x)$ , of  $A$  at  $x$ , is defined as the intersection of all closed cones containing the set  $\{a-x: a \in A\}$ . The closed cone of tangents of the set  $A$  at  $x$ ,  $S(A, x)$ , is defined as

$$S(A, x) = \bigcap_{k=1}^{\infty} \tilde{S}(A \cap N_{1/k}(x), x)$$

where  $N_{1/k}(x)$  is a spherical neighborhood of  $x$  with radius  $1/k$ ,  $k$  being a natural number. The following lemma characterizes  $S(A, x)$ .

Lemma 2.1.1 [49]:

A vector  $y$  is contained in  $S(A, x)$  if and only if there exists a sequence of vectors  $\{x^k\} \subset A$ , converging to  $x$  and a sequence of non-negative numbers  $\{\alpha^k\}$  such that the sequence  $\{\alpha^k(x^k - x)\}$  converges to  $y$ .

Positively normal cone  $A'$  of a non-empty set  $A$  of  $\mathbb{R}^n$  is

defined as

$$A' = \{x: x \in \mathbb{R}^n \text{ and } x^T y \geq 0 \text{ for } \forall y \in A\} .$$

The cone of tangents and the positively normal cones play a central role in establishing strong optimality conditions as illustrated by means of the following sequence of lemmas.

Lemma 2.1.2 [49]:

Suppose  $x^0 \in X$ . The set  $Z^1(x^0) \cap Z^2(x^0)$  is empty if and only if

$$\nabla f(x^0) \in (Z^1(x^0))' .$$

Lemma 2.1.3 [49]:

Suppose  $x^*$  is a solution to the problem (A). Then

$$\nabla f(x^*) \in (S(X, x^*))' .$$

The following theorem gives a set of necessary conditions, stronger than those presented in theorem (2.1.2) and is a direct extension of the Kuhn-Tucker necessary optimality conditions [36].

Theorem 2.1.3 [49]: (Generalized Kuhn-Tucker Necessary Conditions)

Let  $x^*$  be a solution of problem (A) and suppose that

$$(Z^1(x^*))' = (S(X, x^*))' , \quad (2.1.7)$$

then there exists  $\lambda^* \in \mathbb{R}^p$  and  $\mu^* \in \mathbb{R}^m$  such that

$$\nabla f(x^*) + \nabla g(x^*) \lambda^* + \nabla h(x^*) \mu^* = 0 \quad (2.1.8)$$

$$\lambda_i^* g_i(x^*) = 0 \quad (2.1.9)$$

$$\lambda_i \geq 0 \quad \left. \begin{array}{l} \\ \\ \end{array} \right\} i = 1, \dots, p. \quad (2.1.10)$$

Condition (2.1.7) together with the lemmas 2.1.2 and 2.1.3 implies that  $Z^1(x) \cap Z^2(x)$  is empty at the solution point  $x^*$  of the problem (A) and thus is a condition weaker than requiring  $Z^1(x^*) \cap Z^2(x^*)$  to be empty. Gould and Tolle [26] have used the condition (2.1.7) as a constraint qualification and have shown that the condition is sufficient and necessary for the existence of multipliers  $\lambda^*$  and  $\mu^*$  satisfying (2.1.8) - (2.1.10). We conclude the discussion of constraint qualifications by stating two other criteria that are stronger than the one presented above, in the sense that they imply (2.1.7).

The original constraint qualification of Kuhn and Tucker [36] at a point  $x^* \in X$  requires that any vector  $z \in Z^1(x^*)$  be tangent to a once-differentiable arc, emanating from  $x^*$  and contained in the constrained set  $X$ ; that is, for each  $z \in Z^1(x^*)$  there exists a function  $\alpha$  whose domain is in the interval  $[0, \epsilon] \subset \mathbb{R}$  and range is in  $\mathbb{R}^n$ , such that  $\alpha(0) = x^*$ ,  $\alpha(\theta) \in X$  for  $0 \leq \theta \leq \epsilon$ ,  $\alpha$  is differentiable at  $\theta = 0$  and  $\frac{d\alpha(0)}{d\theta} = \beta z$ ,  $\beta > 0$ .

Mangasarian and Fromovitz [41] introduced the following constraint qualification; let  $g_i$  and  $h_j$  be differentiable and continuously differentiable, respectively, at  $x^*$ . The qualification holds if the vector  $\nabla h_j(x)$ ,  $j = 1, 2, \dots, m$  are linearly independent and there exists a vector  $y \in \mathbb{R}^n$  such that

$$\begin{aligned} y^T \nabla g_i(x^*) &< 0 && , i \in I(x^*) \\ y^T \nabla h_j(x^*) &= 0 && , j = 1, 2, \dots, m . \end{aligned}$$

Both, the Kuhn-Tucker qualification and that given by (2.1.7), have a serious disadvantage in the sense that they are not easily verifiable computationally, in contrast with the above stated qualification developed by Mangasarian and the criterion mentioned below.

Lemma 2.1.4 [15]:

A sufficient condition that Kuhn-Tucker qualification holds at a point  $x^* \in X$ , is that gradients  $\{\nabla g_i(x^*)\}$ ,  $i \in I(x^*)$ ,  $\{\nabla h_j(x^*)\}$ ,  $j = 1, 2, \dots, m$  are linearly independent.

### Second-order Optimality Conditions

We now consider optimality conditions for the problem (A) that involves second-derivatives of the functions  $f$ ,  $g$  and  $h$ ; that is, it is assumed that the functions  $f$ ,  $g$  and  $h \in C^2$ . The derivation of

the first-order conditions were based on the fact that the intersection of certain sets were empty and a similar approach was adopted by Dubovitskii and Milyutin [11] in deriving second-order conditions. In a similar derivation by Messerli and Polak [44], some general second-order necessary conditions based on the weak Lagrangian were obtained, that can be applied to problems with minimum regularity conditions. We will sacrifice generality in favour of simplicity in this summary and draw heavily from [43].

To take into account the curvature of the problem functions near a local minimum, McCormick [43] developed necessary conditions that apply to twice-differentiable functions and is summarized in theorem 2.1.4. We shall require a further constraint qualification for this result.

Second-order Constraint Qualification [43]:

Let  $x \in X$  and define

$$\hat{Z}^1(x) = \{z: z^T \nabla g_i(x) = 0, i \in I(x), z^T \nabla h_j(x) = 0, j=1, 2, \dots, m\}.$$

The second-order constraint qualification is said to hold at  $x^0 \in X$  if every nonzero  $z \in \hat{Z}^1(x^0)$  is tangent to a twice differentiable arc contained in the boundary of  $X$ ; that is, for each  $z \in \hat{Z}^1(x^0)$  there exists a twice differentiable function  $\alpha$  defined on  $[0, \epsilon] \subset \mathbb{R}$ , with range in  $\mathbb{R}^n$  such that  $\alpha(0) = x^0$ ,

$$g_i(\alpha(\theta)) = 0, i \in I(x^0), h_j(\alpha(\theta)) = 0, j = 1, 2, \dots, m$$

$$\text{for } 0 \leq \theta \leq \epsilon \text{ and } \frac{d\alpha(0)}{d\theta} = \beta z$$

for some positive  $\beta$ .

Theorem 2.1.4 [43]: (Second-order Necessary Conditions)

Let  $x^*$  be a solution of the problem (A) and suppose that there exists  $\lambda^*, \mu^*$  satisfying (2.1.8) - (2.1.10). Further suppose that the second-order constraint qualification holds at  $x^*$ . Then for  $z \neq 0$  such that  $z \in \hat{Z}^1(x^*)$ , we have

$$z^T \left[ \nabla^2 f(x^*) + \sum_{i=1}^p \lambda_i^* \nabla^2 g_i(x^*) + \sum_{j=1}^m \mu_j^* \nabla^2 h_j(x^*) \right] z \geq 0.$$

Second-order sufficient condition for optimality were derived by Hestenes [27], King [33] and McCormick [43], whose results were subsequently sharpened by Fiacco [16]. Let  $\hat{I}(x^*)$  be the set of indices  $i$  for which  $g_i(x^*) = 0$  and equations (2.1.8)-(2.1.10) are satisfied by a positive  $\lambda^*$ . Let

$$\hat{Z}^1(x^*) \triangleq \{z: z^T \nabla g_i(x^*) = 0, i \in \hat{I}(x^*), z^T \nabla g_i(x^*) \geq 0, i \in I(x^*), z^T \nabla h_j(x^*) = 0, j = 1, \dots, m\}.$$

We then have the following sufficient condition.

Theorem 2.1.5 [43]: (Second-order Sufficient Conditions)

Let  $x^* \in X$  for the problem (A). If there exist  $\lambda^*, \mu^*$  satisfying (2.1.8)-(2.1.10) and for every  $z \neq 0$  such that  $z \in \hat{Z}^1(x^*)$ , it follows that

$$z^T \left[ \nabla^2 f(x^*) + \sum_{i=1}^p \lambda_i^* \nabla^2 g_i(x^*) + \sum_{j=1}^m \mu_j^* \nabla^2 h_j(x^*) \right] z > 0$$

then  $x^*$  is a strict local minimum of the problem (A).

## Section 2.2 Equality Constrained Nonlinear Programs

A fundamental concept that provides a great deal of insight as well as simplifying the theoretical development is that of an active constraint. An inequality constraint is said to be active at a feasible point  $x$  if  $g_i(x) = 0$  and inactive if  $g_i(x) > 0$ . An active constraint at any feasible point  $x$  restrict the domain of feasibility to a neighborhood of the point, while other, inactive constraints have no influence in this neighborhood. Therefore, in studying the properties of a local solution point, attention can be restricted to active constraints. If it were known a priori which constraints were active at the solution point, the solution would be a local minimum point of the problem defined by ignoring the inactive constraints and treating all the active constraints as equality constraints. Hence with respect to, local solutions the problem could be regarded as having only equality constraints.

Inequality constraints can be transformed into an equivalent problem having equality constraints only at the expense of increasing the numbers of variables. Given  $g(x) \geq 0$ , let  $y \in R^p$  be additional variables. Then  $g(x) \geq 0$  is equivalent to  $g_i(x) - (y_i)^2 = 0$  for  $i=1, 2, \dots, p$ . Such transformations, while advantageous in some cases, can also cause considerable weakening of some results. This is, for example, the case with the Fritz John conditions as pointed out by Mangasarian and Fromovitz [41].

In view of the above discussion, in this dissertation, only the equality constraint mathematical programming problems, will be considered;

$$(B) \quad \begin{cases} \text{Minimize } f(x) \\ \text{subject to } h(x) = 0 \end{cases} \quad (2.2.1)$$

where  $x \in \mathbb{R}^n$ ,  $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f$  and  $h_i$  are real-valued functions with domain  $D = \mathbb{R}^n$ ,  $m < n$ .

### Optimality Conditions

As is expected results on the optimality conditions, for the problem (B) presented here are special cases of the results presented in the last section for the problem (A). A number of frequently used concepts are repeated here.

The Lagrangian associated with the problem (B) is given by

$$l(x, \mu) = f(x) + \mu^T h(x) \quad , \quad (2.2.2)$$

where  $\mu \in \mathbb{R}^m$  is the Lagrange multiplier vector. Let  $X = \{x: h(x)=0\}$  be the set of feasible points to the problem (B) and define [40] a Regular Point  $x$  of the constraints as a point in  $X$  such that gradients  $\nabla h_i(x)$ ,  $i=1, 2, \dots, m$ , are linearly independent.

### Theorem 2.2.1 [40]: (First-order Necessary Conditions)

Let  $x^*$  be a local minimum of the problem (B). Assume that  $x^*$  is a regular point of the constraints. Then there exists  $\mu^* \in \mathbb{R}^m$  such that

$$\nabla f(x^*) + \nabla h(x^*) \mu^* = 0 \quad (2.2.3)$$



Theorem 2.2.2 [40]: (Second-order Necessary Conditions)

Suppose that  $x^*$  is a local minimum of the problem (B) and that  $x^*$  is a regular point of the constraints. Then there exists a  $\mu^* \in \mathbb{R}^m$  such that (2.2.3) holds and

$$\nabla^2 l(x^*, \mu^*) = \nabla^2 f(x^*) + \sum_{i=1}^m \mu_i^* \nabla^2 h_i(x^*) \quad (2.2.4)$$

is positive semi-definite for all  $y \in \mathbb{R}^n$  such that  $y \neq 0$  and  $\nabla h(x^*) y = 0$ .

Theorem 2.2.3 [40]: (Second order Sufficient Conditions)

Suppose there is a point  $x^*$  satisfying  $h(x^*) = 0$  and a  $\mu^* \in \mathbb{R}^m$  such that

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

Suppose also that for all  $y \in \mathbb{R}^n$ ,  $y \neq 0$ ,  $\nabla h(x^*) y = 0$ , there holds  $y^T \nabla^2 l(x^*, \mu^*) y > 0$ . Then  $x^*$  is a strict local minimum of the problem (B).

Section 2.3 Methods of Solving Constrained Nonlinear Programs

An intuitive method for solving equality constrained problem (B) is to transform it into an unconstrained problem by elimination of  $m$  variables. This equivalent unconstrained problem can then be solved by known techniques of unconstrained minimization. Such an elimination, even when theoretically justified, often proves to be difficult if not an impossible task.

Alternatively, methods are available for solving constrained optimization problems by solving one or more "equivalent" unconstrained problems. Such methods can roughly be divided into two classes:

- a. Penalty Function Methods
- b. Lagrangian Methods.

We will describe briefly some of the methods which we need for later reference.

#### (a) Penalty Function Methods

The Penalty Function Methods provide procedure for approximating constrained optimization problems by a sequence of unconstrained problems. The intuitive idea behind all penalty methods is simple; that is, suppose we seek a minimum of a real-valued function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , on a proper subset  $D \subset \mathbb{R}^n$ . Define

$$P(x) = \begin{cases} 0 & x \in D \\ +\infty & x \notin D \end{cases}$$

and consider the augmented objective function  $G$  given by

$$G(x) = f(x) + P(x) .$$

A point  $x^*$  minimizes  $G$  if and only if it also minimizes  $f$  over  $D$ .

The penalty function,  $P(x)$ , imposes an infinite penalty on points

outside the feasible set  $D$ . Thus we transformed the constrained problem of minimizing  $f(x)$ ,  $x \in D \subset \mathbb{R}^n$ , to one of unconstrained minimization of  $G(x)$ ,  $x \in \mathbb{R}^n$ . Except perhaps in trivial cases, in practice, the unconstrained optimization of  $G(x)$  can not be carried out because of the discontinuity in  $G$  on the boundary of  $D$  and the infinite values outside  $D$ . To avoid this difficulty, the idea of a penalty parameter,  $\rho$ , was introduced as follows: let  $Q$  be a function on  $\mathbb{R}^n$  satisfying (i)  $Q$  is continuous, (ii)  $Q(x) \geq 0$  for all  $x \in \mathbb{R}^n$  and (iii)  $Q(x) = 0$  if and only if  $x \in D$ , and define the augmented objective function,  $F$ , as

$$F(x, \rho) = f(x) + \rho Q(x),$$

where  $\rho > 0$ . Instead of solving the constrained problem through a single optimization of an equivalent unconstrained problem of  $\min. G(x)$ ,  $x \in \mathbb{R}^n$ , the constrained optimum is approximated by a sequence of unconstrained optima of  $F(x, \rho)$ , generated for a sequence of values for  $\rho$ . The value of  $\rho$  is so adjusted that the corresponding sequence of unconstrained optima converges to a feasible point of the constrained problem, that also satisfies some necessary or sufficient optimality condition.

The initial suggestions for the penalty function approach came from Courant [10] and Frisch [24]. Subsequent developments involved contributions by a number of authors [7, 17, 29, 53, 59, 60,

64, 54]. A complete and rigorous treatment of the method can be found in [15]. Result relevant to later discussions are summarized below.

### Exterior-Point Method (Penalty Method)

The exterior method solves problem (A) by a sequence of unconstrained minimization problems whose optimal solutions approach the solution of problem (A) from outside the feasible set  $X$ . In the sequence of unconstrained optimizations, a penalty is imposed on every  $x \notin X$ , such that this penalty is increased from problem to problem, thereby forcing the unconstrained optima towards the feasible set.

Let  $r$  be a continuous real-valued function of  $\rho \in \mathbb{R}$  such that  $\rho^1 > \rho^2 > 0$  implies  $r(\rho^2) > r(\rho^1) > 0$  and that  $\lim_{k \rightarrow \infty} \{\rho^k\} = 0$  implies that  $\lim_{k \rightarrow \infty} \{r(\rho^k)\} = +\infty$ . Define the penalty function,  $S(x)$ , such that (i)  $S(x)$  is continuous, (ii)  $S(x) = 0$  if  $x \in X$  and (iii)  $S(x) > 0$  if  $x \notin X$ . Consider the augmented objective function for the problem (A), as

$$F(x, \rho) = f(x) + r(\rho) S(x) .$$

The exterior method consists of solving a sequence of unconstrained minimization of  $F(x, \rho^k)$  for  $k = 0, 1, \dots$ , using a strictly decreasing sequence of positive numbers  $\{\rho^k\}$ . Let  $x^{k*}$  be the minimum point of  $F(x, \rho^k)$  for a given  $\rho^k$  for some  $k$ . The following

convergence theorem [65] states that, under mild conditions on the functions of problem (A), the sequence of unconstrained minima  $\{x^{k*}\}$  has a convergent subsequence and the limit of any such subsequence is optimal for the problem (A).

Theorem 2.3.1 [65]: (Convergence Theorem; Exterior-point Method)

Suppose that the feasible set  $X$  of the problem (A) is nonempty and there exists a  $\epsilon > 0$  such that the set

$$X_\epsilon = \{x: x \in \mathbb{R}^n, g_i(x) \geq -\epsilon, i=1, 2, \dots, p; |h_j(x)| \leq \epsilon, j=1, 2, \dots, m\}$$

is compact. Also suppose that  $F(x, \rho^k)$  attains their unconstrained minima on  $\mathbb{R}^n$  for all  $k$ . If  $\{\rho^k\}$  is strictly decreasing sequence of positive numbers converging to zero; then there exists a convergent subsequence  $\{x^{k_1*}\}$  of the optimal solution to  $F(x, \rho^k)$  and the limit of any such convergent subsequence is optimal for the problem (A).

Zangwill [65] investigated conditions that imply that the functions  $F(x, \rho^k)$  attain their minima, an assumption used in the above theorem. The first condition says that  $f$  is a function such that whenever  $\|x\| \rightarrow +\infty$ , then  $f(x) \rightarrow \infty$  also. Another condition with the same implication is that the set  $X_\epsilon$  is compact for some  $\epsilon > 0$  and  $F(x, \rho)$  is a convex function of  $x$ . The convergence theorem 2.3.1 refers to a global minimum of the problem (A), and thus it seems to be a strong result. It assumes, however, that we find a

global minimum of the unconstrained problem  $F(x, \rho^k)$  for each  $k$ , which is a rather difficult task. For this reason, the method is mainly used for convex programs.

Convergence to local minima of nonconvex programs by exterior methods can also be obtained under slightly different assumptions. For a detailed discussion, we refer to Fiacco and McCormick [15]; here we shall only mention their main result. Let  $Y \subset X$  be the set of all point that are local minima to the problem (A). Suppose  $Y$  is nonempty and is an isolated compact set. Then there exists a compact set  $S$  such that  $Y$  is embedded in the interior of  $S$  and for sufficiently large  $k$ , the unconstrained minima of  $F(x, \rho^k)$  are attained in the interior of  $S$ . Moreover, if  $v^*$  is the minimum value of  $f$ , then  $\lim_{k \rightarrow \infty} \{f(x^{k*})\} = v^*$  and every limit point of any convergent subsequence of  $\{x^{k*}\}$  is in  $Y$ . The set  $S$  is related with the problem functions in the following way. For some small-enough positive  $\sigma$ ,

$$S = \{x: x \in \mathbb{R}^n, g_i(x) \geq -\epsilon, i=1, 2, \dots, p; |h_j(x)| \leq \epsilon, j=1, \dots, m; \\ f(x) \leq v^* + \sigma\}.$$

### Interior-Point Methods (Barrier Method)

The interior-point methods are applicable to the problems of the form

$$(C) \quad \min f(x), \quad \text{subject to } x \in X, \quad X \triangleq \{x: x \in \mathbb{R}^n, g(x) \geq 0\},$$

where the constraining set  $X$  has a nonempty interior that is arbitrarily close to any point of  $X$ . The method works by establishing a barrier on the boundary of the feasible set that prevents a search procedure from leaving the region. A barrier function,  $q(x)$ , is defined on the interior of  $X$  such that (i)  $q(x)$  is continuous, (ii)  $q(x) \geq 0$  and (iii)  $q(x) \rightarrow \infty$  as  $x$  approaches the boundary of  $X$  and define a real-valued function  $t$  of  $\rho \in \mathbb{R}$  such that  $\rho^1 > \rho^2 > 0$  implies  $t(\rho^1) > t(\rho^2) > 0$  and that  $\lim_{k \rightarrow \infty} \{\rho^k\} = 0$  implies that  $\lim_{k \rightarrow \infty} \{t(\rho^k)\} = 0$ . Choosing a strictly decreasing sequence of positive numbers,  $\rho^k$ , converging to zero, the augmented objective function  $G(x, \rho^k) \triangleq f(x) + t(\rho^k) q(x)$ , is minimized for each  $\rho^k$ ,  $k=0, 1, 2, \dots$ , generating a sequence of minimums  $\{x^{k*}\}$ . The method is quite analogous to the exterior-point method and virtually the same convergence properties hold for the method as for the exterior method.

Since interior-point methods require a nonempty interior of the feasible set  $X$ , the methods are not applicable to programming problems containing equality constraints. We can, however, use Mixed-Penalty methods to solve equality-inequality constrained programs. The method consists of solving a sequence of unconstrained minimization of an augmented objective function  $H(x, \rho^k, \eta^k)$

$= f(x) + t(\rho^k) q(x) + r(\eta^k) S(x)$ , where  $t$ ,  $q$ ,  $r$  and  $S$  are functions as defined for interior and exterior point methods but  $q(x)$  and  $S(x)$  are applied to inequality and equality constraints respectively. Under suitable assumptions [15], a subsequence of  $\{x^{k*}\}$  can be shown to converge to the constrained optimum.

### Parameter-Free Penalty Methods

A question on the implementation of Penalty Function Methods concerns the choice of the penalty parameter. It is necessary to decide on the initial choice and a rule to modify the parameter, to ensure a monotonically decreasing sequence that converges to zero. This problem of choosing a suitable parameter values can be avoided by modifying the penalty functions so that parameters are automatically chosen, thus modifying the method to become parameter free. A detailed description on Parameter-free Penalty Methods is given in an article by Lootsma [37]. In this reference we also mention the work of Rosenbrock [59], Huard [29, 30, 31], Bui and Huard [5], Tremolieres [61], Fiacco and McCormick [18] and Lootsma [38, 39]. Lootsma [38] has studied the convergence rate for these methods and found that they have no particular advantage over parametric penalty methods.



### Exact Penalty Method

Basically the Penalty Function Methods are not iterative in nature. The definition of  $x^{k+1*}$  in no way depends on  $x^{k*}$ . Thus if one decides to terminate the sequence at the  $N$  term corresponding to  $\rho^N$ , obtaining  $x^{N*}$ , the calculations of  $x^{ki*}$ ,  $i < N$ , is irrelevant, since  $x^{N*}$  could have been calculated directly by solving a single unconstrained optimization problem for  $\rho = \rho^N$ . This is generally the manner in which penalty methods are employed; one selects a small value of  $\rho$ , solves the corresponding unconstrained problem and takes the resulting solution as a final answer to the problem. A basic question associated with the approach is then, exactly what is a small value of  $\rho$  relative to a particular problem? Such a value may not be known in advance and consequently an initial value of  $\rho$  may produce a solution point that is not close enough to the constrained optimum, requiring a decrease in the value of  $\rho$ . This will result in performing a number of unconstrained optimization problems. The other basic difficulty associated with the penalty approach is the unfavourable structure of the constrained problem corresponding to small  $\rho$ . It turns out [40, 50] that small  $\rho$  results in an ill-conditioned Hessian of the corresponding unconstrained problem, which in turn implies slow convergence for many unconstrained optimization algorithms.

To avoid solving a sequence of unconstrained optimization problems an alternative strategy, known as the Exact Penalty Method, was developed. One of the first works dealing with exact penalty functions is due to Zangwill [65]. He shows that, for the inequality-constrained problem (C), where  $f(x)$  is convex and  $g(x)$  is concave, satisfying strong consistency conditions, there exists a  $\rho^*$  such that for a  $\rho$ ,  $\rho^* \geq \rho > 0$  the function

$$P(x, \rho) \triangleq f(x) - \frac{1}{\rho} \sum_{i=1}^p \min(0, g_i(x))$$

is exact. Thus for a proper choice of  $\rho$ , the unconstrained minimum of  $P(x, \rho)$  coincides with the constrained minimum. Pietrykowski [54] has obtained the related results for non-convex programs. The main disadvantage with the Zangwill-Pietrykowski exact penalty function is that it is not continuously differentiable. In fact the derivative fails to exist along all constraint boundaries and renders this method useless in most nonlinear programming problems, since optima in constrained problems occur, by implication, on the boundary of the feasible region. Although, special optimization methods for piece-wise differentiable exact penalty functions have been suggested [8, 9], the usefulness the piece-wise differentiable exact penalty function methods is doubtful. To date no efficient techniques have been developed to optimize such functions.

Fletcher [19, 20] and Fletcher and Lill [21] have developed differentiable exact penalty functions for the equality constrained programs. We describe their work here. Consider the problem (B) and assume that problem functions  $f, h_i, i = 1, 2, \dots, m$ , are in  $C^2$ . Define a  $(n \times m)$  matrix  $N(x)$  whose columns are the gradients of the constraints  $h_i(x), i = 1, 2, \dots, m$ . Assuming  $N(x)$  is of full rank  $m < n$ , the generalized inverse [51]  $N^+(x)$ , of  $N(x)$  is given by

$$N^+(x) = (N^T(x) N(x))^{-1} N^T(x).$$

The matrix  $P(x) = N(x) N^+(x)$  and its complementary matrix  $\hat{P}(x) = I_n - P(x)$ , projects vectors into subspaces spanned by constraint-normals and tangential to the manifold formed by intersection of the constraints, respectively. In terms of  $P$  and  $\hat{P}$ , the Kuhn-Tucker necessary conditions (2.2.3), can be written [19] as

$$\hat{P}(x) \nabla f(x) = 0.$$

Fletcher [19] has found a whole class of exact penalty functions for the program (B), a representative member of which is given by

$$\phi(x) = f(x) - h^T(x) N^+(x) f(x) + \frac{\rho}{2} h^T(x) h(x). \quad (2.3.1)$$

This function is stationary at every local constrained minimum  $x^*$  of the problem (B). It is also known that the Lagrangian,  $l(x, \mu)$ , as

given by (2.2.2), associated with the problem (B), must also be stationary at  $x^*$ , that is

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

which implies

$$\mu = -N^+(x^*) \nabla f(x^*) . \quad (2.3.2)$$

Letting  $\mu$  to be a function of  $x$  as in (2.3.2), we obtain a function  $\psi(x) = f(x) - h^T(x) N^+(x) \nabla f(x)$ , which is also stationary at every local minimum of the problem (B), but the function  $\psi(x)$  is not necessarily equal to the ordinary Lagrangian as defined by (2.2.2), where  $\mu$  is a constant vector. Comparing the functions  $\phi$  and  $\psi$ , we see that the function  $\phi$  has an additional term  $\frac{\rho}{2} h^T(x) h(x)$ , the role of which is to ensure that the function  $\phi$  has a minimum at the local solution of the problem (B) and not just a stationary point. In fact Fletcher proves the following result:

Theorem 2.3.1 [19]:

Let  $x^* \in \mathbb{R}^n$ , satisfying sufficient conditions for a strict local minimum of the problem (B). Then  $\nabla \phi(x^*) = 0$  and there exists a  $\rho^* > 0$  such that for every  $\rho \geq \rho^*$ , the matrix  $\nabla^2 \phi(x^*)$  is positive definite.

Implementation of the method requires a proper choice for  $\rho$  so that a single unconstrained optimization yields the sought minimum of the problem (B). For quadratic  $f$  and linear  $h$  any  $\rho > \|N^+ \nabla^2 f N^{T+}\|$ , derived from the requirement that  $\nabla^2 \phi$  be positive definite, would suffice. For nonlinear constraint functions, the computation of the critical parameter  $\rho^*$  is by no means simple. Fletcher and Lill [21] suggest taking  $\rho = 10 \| \nabla^2 f(x^0) \| + r$ , where  $x^0$  is the starting point for the unconstrained minimization of function  $\phi$  and  $r = 1$  is taken initially. However,  $\rho$  must be adjusted by increasing  $r$  if, during the course of minimization, a negative definite  $\nabla^2 \phi(x)$  is encountered.

The method has been extended to deal with problems of the type (A) by Fletcher [22]. The extended method essentially consists of solving the corresponding quadratic program and in limited tests [22], it is shown that once the correct set of active constraints is determined, convergence is attained in a remarkable small number of iterations. Some potential dangers of this kind of extension are discussed by Evans and Gould [12].

#### (b) Lagrangian Methods

Consider again the nonlinear programming problem (B). If  $x^*$ , the local solution of  $f(x)$  subject to  $h(x)=0$ , is the regular point of these constraints, then we know (Ref: Theorem 2.2.1) that there

exists a  $\mu^*$  such that  $(x^*, \mu^*)$  is the stationary point of the Lagrangian  $l(x, \mu) = f(x) + \mu^T h(x)$  and is a solution to the set of nonlinear equations

$$\nabla f(x) + \nabla h(x)\mu = 0 \quad \text{and} \quad h(x) = 0 .$$

Minimization of the Lagrangian by solving the corresponding necessary conditions is known as the Lagrange-Multiplier method. If  $\mu^*$  was known a priori, we could have obtained the stationary point of  $l$  by simply seeking its unconstrained minimum. The fact is that  $\mu^*$  is not known and the question is how one would determine it.

More recent methods reduce the solution of the constrained problem to sequential minimization of the Lagrangian,  $l$ , as a function of  $x$ , updating  $\mu$  from one cycle of minimization to another, in a way that approximates  $\mu^*$  at the extremum. In reference we mention the work of Everett [13], Wolfe [63], Brooks and Geoffrion [4], Falk [14], Roode [58] and will describe briefly some methods based on the augmented Lagrangian.

Consider the equality constrained program (B). The following theorem shows that it is possible to find a positive number  $c^*$  such that  $x^*$  is the unconstrained minimum of the augmented Lagrangian  $M(x, \mu, c)$ , where

$$M(x, \mu^*, c^*) = f(x) + \mu^{*T} h(x) + \frac{1}{2} c^* h^T(x) h(x) . \quad (2.3.4)$$

Theorem 2.3.2 [49]:

Suppose  $x^*$  and  $\mu^*$  satisfy the sufficiency conditions of optimality for  $x^*$  to be a strict local minimum of the problem (B). Then there exist a number  $c^* > 0$  such that for all  $c \geq c^*$ , the point  $x^*$  is a local unconstrained minimum of  $M(x, \mu^*, c)$ . Conversely, if  $h(x^0) = 0$  and  $x^0$  is an unconstrained minimum of  $M(x, \mu^0, c)$  for some  $\mu^0$ , then  $x^0$  solves problem (B).

The theorem suggest that if  $\mu^*$  and a sufficiently large  $c$  is available, then an unconstrained minimization of  $M(x, \mu^*, c)$  yields an optimal solution of the problem (B). The difficulty, of course, lies in determining the correct values of  $\mu^*$  and  $c$ . Hestenes' Method of Multipliers [28] consists of, starting with an initial estimate of multiplier  $\mu$ , updating the  $\mu$  at the end of each cycle of unconstrained minimization of  $M(x, \mu, c)$  given by (2.3.4). Suppose we choose a sufficiently large value of  $c$  and at the  $k^{\text{th}}$  cycle we have some estimate  $\mu^k$  for  $\mu$ .  $M(x, \mu^k, c)$  is minimized to give an optimal solution and at the minimum  $x^{k*}$ , we have

$$M(x^{k*}, \mu^k, c) = \nabla f(x^{k*}) + \sum_{i=1}^m [\mu_i^k + c h_i(x^{k*})] \nabla h_i(x^{k*}) = 0 .$$

The Lagrange multipliers are then updated by the formula

$$\mu^{k+1} = \mu^k + c h(x^{k*})$$

and one proceeds to the next cycle by minimizing  $M(\mathbf{x}, \mu^{k+1}, c)$ .

Independently, Powell [55] has also derived a method of multipliers for the problem (B) that is very similar to Hestenes' method. Hestenes' method was extensively studied from a computational point of view by Miele and coworkers [45, 46, 47, 48]. The method of multipliers was extended to inequality constrained programs by Bertsekas [3], Buys [6], Kort and Bertsekas [34, 35], Pierre [52], and Rockafellar [56, 57]. We describe briefly the works of Rockafellar.

Consider the programming problem (A) and define a

$$\theta(t) = \max. \{t, 0\} .$$

Then for any positive  $c$ , we have

$$\frac{1}{4c} \{ [\theta(\lambda_i - 2c g_i(\mathbf{x}))]^2 - (\lambda_i)^2 \} = \begin{cases} c(g_i(\mathbf{x}))^2 - \lambda_i g_i(\mathbf{x}) & \text{if } g_i(\mathbf{x}) \leq \frac{\lambda_i}{2c} \\ -\frac{(\lambda_i)^2}{4c} & \text{if } g_i(\mathbf{x}) \geq \frac{\lambda_i}{2c} \end{cases}$$

and the augmented Lagrangian  $M(\mathbf{x}, \lambda, \mu)$  can be defined as

$$M(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \frac{1}{4c} \sum_{i=1}^p \{ [\theta(\lambda_i - 2c g_i(\mathbf{x}))]^2 - (\lambda_i)^2 \} \tag{2.3.5}$$

$$- \sum_{j=1}^m \mu_j h_j(\mathbf{x}) + \frac{c}{2} \sum_{j=1}^m [h_j(\mathbf{x})]^2 .$$



Hestenes' method can now be easily generalized. For a given  $c > 0$  and the multipliers  $(\lambda^k, \mu^k)$ , let  $x^{k*}$  be the optimal solution of  $M(x, \lambda^k, \mu^k)$  as defined by (2.3.5). Then

$$M(x^{k*}, \lambda^k, \mu^k) = f(x^{k*}) - \sum_{i=1}^p \theta(\lambda_i^k - 2c g_i(x^{k*})) \nabla g_i(x^{k*}) \\ - \sum_{j=1}^m \{\mu_j^k - c h_j(x^{k*})\} \nabla h_j(x^{k*}) = 0 .$$

So it is reasonable to set

$$\lambda_i^{k+1} = \theta\{\lambda_i^k - 2c g_i(x^{k*})\} \quad i = 1, \dots, p.$$

$$\mu_j^{k+1} = \mu_j^k - c h_j(x^{k*}) \quad j = 1, \dots, m.$$

Rockafellar [56] has shown the convergence of the method to an optimal solution of the problem (A). In addition, he has also derived another procedure that converges to a solution of the convex program that satisfies the consistency conditions; that is, there exists a point  $x^0 \in R^n$  such that  $g(x^0) > 0$  and  $h(x^0) = 0$ .

A possible disadvantage of the Lagrangian Methods, described so far, is that the augmented lagrangian  $M$  is usually only once differentiable. This is an important distinction in the application of Newton-type algorithms which requires twice-differentiability. Mangasarian [42] has developed, for nonconvex inequality and equality

constrained program (A), a wide class of lagrangian functions which are twice differentiable everywhere. We shall describe Mangasarian's [42] work briefly.

For the nonlinear programming problem (A), the generalized Lagrangian is defined [42] as

$$\begin{aligned}
 GL(x, y, w, \alpha) = f(x) + \sum_{i=1}^p (\psi(\alpha g_i(x) + y_i)_+ - \psi(y_i)) \\
 + \sum_{j=1}^m (\psi(\alpha h_j(x) + w_j) - \psi(w_j))
 \end{aligned} \tag{2.3.6}$$

where  $\alpha > 0$ ,

$$\psi: \mathbb{R} \rightarrow \mathbb{R}, \quad \psi(\xi)_+ = \begin{cases} \psi(\xi) & \text{if } \xi \geq 0 \\ 0 & \text{if } \xi < 0 \end{cases}$$

and  $\psi$  satisfies the following conditions; (a)  $\psi$  is twice differentiable on  $\mathbb{R}$ , its second derivative,  $\psi''(\xi) > 0$  for  $\xi \neq 0$  and  $\psi''(0) = 0$ , (b)  $\psi'$  maps  $\mathbb{R}$  onto  $\mathbb{R}$  and  $\psi'(0) = 0$  and (c)  $\psi(0) = 0$ . Under these conditions imposed on  $\psi$ , it is guaranteed that for sufficiently large  $\alpha$ , a stationary point  $(x^*, y^*, w^*)$  of  $GL$  corresponds to a strict unconstrained local minimum of  $GL(x, y^*, w^*)$ .

An example of the function  $\psi$ , satisfying the above stated conditions (a), (b), and (c) except for the condition  $\psi''(0) = 0$ , is given by

$$\psi(t) = \frac{1}{2\alpha} (t)^2, \quad \alpha > 0$$

For this choice of function  $\psi$ , the Lagrangian  $GL(x, y, w, \alpha)$  becomes

$$\begin{aligned} GL(x, y, w, \alpha) = f(x) + \frac{1}{2\alpha} \sum_{i=1}^p ((\alpha g_i(x) + y_i)^2 - (y_i)^2) \\ + \frac{1}{2\alpha} \sum_{j=1}^m ((\alpha h_j(x) + w_j)^2 - (w_j)^2). \end{aligned} \tag{2.3.7}$$

Letting  $\alpha = 2c$ ,  $y_i = \lambda_i$  and  $w_j = \mu_j$ , it can be seen that the generalized Lagrangian, given by (2.3.7), is exactly Rockafellar's augmented Lagrangian given by (2.3.5), which is not twice differentiable everywhere, in contrast with the generalized Lagrangian  $GL(x, y, w, \alpha)$  based on a  $\psi$  function that satisfies all the conditions (a), (b) and (c).

### III. A NEW COMPUTATIONAL METHOD FOR CONSTRAINED NONLINEAR PROGRAMMING PROBLEMS

#### Section 3.1 Motivation

Consider the equality constrained nonlinear programming problem

$$\min. f(x) \tag{3.1.1}$$

$$B \quad \left\{ \begin{array}{l} \text{subject to constraints } h(x) = 0 \end{array} \right. \tag{3.1.2}$$

where  $x \in \mathbb{R}^n$  and  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $m < n$ .  $f$  and  $h_i$  are real-valued continuous functions and are assumed to possess continuous second partial derivatives.

Associated with the problem (B), is the Lagrangian

$$l(x, \mu) = f(x) + \mu^T h(x) \tag{3.1.3}$$

where  $\mu \in \mathbb{R}^m$  is the Lagrange multiplier vector of parameters. If  $x^*$  is the optimal solution to the problem (B), it must satisfy the necessary conditions

$$\nabla f(x^*) + \nabla h(x^*) \mu^* = 0 \tag{3.1.4}$$

together with the constraints

$$h(x^*) = 0, \tag{3.1.5}$$

provided  $x^*$  is a regular point of the constraints. This regularity condition on the constraining surface  $h(x) = 0$  at  $x^*$  implies that the  $(n \times m)$  matrix  $N(x) \triangleq \nabla h(x)$ , is of full rank at  $x^*$  and assures the existence of a  $\mu$  so that (3.1.4) holds. Defining the generalized inverse of  $N$  as  $N^+(x) = (N^T(x)N(x))^{-1}N^T(x)$ , if  $x^*$  is known, equation (3.1.4) implies that

$$\mu^* = -N^+(x^*)\nabla f(x^*) \quad (3.1.6)$$

The Penalty method to solve the problem (B) is based on the penalty function

$$(B1) \quad P(x, \rho) = f(x) + \frac{1}{2\rho} h^T(x)h(x) \quad (3.1.7)$$

where  $\rho > 0$ , is the penalty constant.

The problem of minimizing the function  $f(x)$  subject to the constraints  $h(x) = 0$  is replaced by a sequence of unconstrained minimization problems, in which  $P(x, \rho)$  of (3.1.7) is minimized with respect to  $x$  for each  $\rho > 0$ , in a sequence of  $\{\rho^k\}$ . At the end of each unconstrained minimization, theoretically, the following condition will be satisfied:

$$\nabla P(x^{\rho^*}, \rho) = \nabla f(x^{\rho^*}) + \nabla h(x^{\rho^*}) \frac{h(x^{\rho^*})}{\rho} = 0 \quad (3.1.8)$$

Since the choice of penalty parameter is arbitrary,  $x^{\rho^*}$  satisfying (3.1.8) is such that  $h(x^{\rho^*}) \neq 0$  and the point  $x^{\rho^*}$  is not feasible.

Nevertheless, if one defines the Lagrange multipliers to be

$$\mu^{\rho*} = \frac{h(x^{\rho*})}{\rho} \quad , \quad (3.1.9)$$

equation (3.1.8) reduces to

$$\nabla P(x^{\rho*}, \rho) = \nabla f(x^{\rho*}) + \nabla h(x^{\rho*}) \mu^{\rho*} \quad ,$$

meaning that the combination of  $x^{\rho*}$  and  $\mu^{\rho*}$ , thus obtained, satisfies exactly the necessary conditions (3.1.4).

To obtain the constraint satisfaction, the Penalty method employs a decreasing sequence of  $\rho$  in successive cycles of the unconstrained minimization of the penalty function  $P(x, \rho)$ . After a sufficient number of cycles, the error in the constraint satisfaction can be made as small as desired provided  $\rho$  becomes sufficiently small. Theoretically,  $h(x) = 0$  is desired at the extremum and consequently the multiplier  $\mu$  defined by (3.1.9) is identical with the multiplier  $\mu^*$  satisfying (3.1.4), which is generally nonzero, only if  $\rho = 0$ .

From a practical point of view, there are two basic difficulties associated with the implementation of Penalty methods. The method requires that the penalty function  $P(x, \rho)$  be minimized for small values of  $\rho$ . Exactly what is a small value of  $\rho$  is, relative to a problem, may not be known in advance. Consequently one has to perform a number of expensive experimental unconstrained

minimizations. Secondly, as  $\rho$  is made small to yield a good approximating problem, the Hessian of the corresponding unconstrained problem can become ill-conditioned thereby implying slow convergence for many unconstrained minimization algorithms that might be applied.

The motivation behind the proposed method is initiated by the desire to eliminate the need for solving a large number of unconstrained minimizations. This is accomplished by satisfying the necessary conditions (3.1.4) and constraints (3.1.5) simultaneously. The result is that one solves the constrained problem (B) through a single optimization of an equivalent unconstrained problem, for a single value of parameter  $\rho$ , not necessarily small.

### Section 3.2 The Intuitive Idea Behind the Proposed Method

Consider the equality constrained problem (B) and the penalty function, as given by (3.1.7)

$$P(x, \rho) = f(x) + \frac{1}{2\rho} h^T(x) h(x) . \quad (3.2.1)$$

Implementation of the Penalty method to solve the constrained problem (B), requires that one should seek the minimum,  $x^{\rho*}$ , of  $P(x, \rho)$  for decreasing values of  $\rho \rightarrow 0$ ,  $\rho > 0$ , satisfying the necessary conditions for each  $\rho$ . The process generates sequences of points  $\{x^{\rho*}\}$  and

$\left\{ \frac{h(x^{\rho*})}{\rho} \right\}$  shown in [40] to converge, as  $\rho \rightarrow 0$ , to the optimal solution  $x^*$  of the problem (B) and a corresponding optimal Lagrange multiplier  $\mu^*$  respectively.

For a fixed  $\rho > 0$ ,  $x^{\rho*}$  minimizes  $P(x, \rho)$ , if

$$\nabla P(x, \rho) = \nabla f(x) + \nabla h(x) \frac{h(x)}{\rho} = 0, \quad (3.2.2)$$

at  $x = x^{\rho*}$ . This implies that  $h(x^{\rho*}) \neq 0$ . For if  $h(x^{\rho*}) = 0$ , equation (3.2.2) implies that  $\nabla f(x^{\rho*}) = 0$  and thus  $x^{\rho*}$  is in the interior of the constraining set determined by  $h(x) = 0$ . The interior of the constraining set  $X = \{x: h(x) = 0\}$  is, however, empty. This contradicts the fact that  $x^{\rho*}$  belongs to the interior of  $X$ .

Since  $\rho \neq 0$  and  $x^{\rho*}$  is such that  $h(x^{\rho*}) \neq 0$ , defining  $q = \frac{h(x^{\rho*})}{\rho} \neq 0$ , we have a nonzero approximation to the Lagrange multiplier  $\mu$ . However, such an approximation is not a true approximation of  $\mu^*$  since  $x^{\rho*}$  is not feasible. To be able to solve the constrained problem, through a single optimization of  $P(x, \rho)$  for some given  $\rho$ , the  $m$ -vector  $q$ , approximating the optimal Lagrange vector, must be correct in such a way that  $x^{\rho*}$  is feasible for the problem (B).

At the optimum  $x^*$  of the problem (B), equation (3.1.6) implies that the associated Lagrange multiplier vector of parameters is the value of the function  $-N^+(x) \nabla f(x)$  evaluated at  $x^*$ . Assuming that



the constraint-normals are independent so that  $N$  is of full rank, the Lagrange multiplier can be considered as a continuous function  $-N^+(x) \nabla f(x)$  of  $x$ . We incorporate the two approximations to the Lagrange vector and set, for any  $x$ ,

$$q = -N^+(x) \nabla f(x) + \frac{h(x)}{\rho}. \quad (3.2.3)$$

Lemma 3.2.1;

Let  $x^0 \in \mathbb{R}^n$  be such that  $N(x^0)$  is of full rank ( $m$ ). Then there is a  $q^0 \in \mathbb{R}^m$  as defined by (3.2.3). Suppose that  $x^0$  and  $q^0$  satisfy the necessary conditions

$$0 = \nabla f(x^0) + N(x^0) q^0 \quad (3.2.4)$$

of the constrained problem (B). Then  $h(x^0) = 0$ .

Proof:-

Multiply equation (3.2.4) by  $N(x^0) N^+(x^0)$ , giving

$$0 = N(x^0) N^+(x^0) \nabla f(x^0) + N(x^0) N^+(x^0) N(x^0) q^0.$$

Since [51]  $N(x^0) N^+(x^0) N(x^0) = N(x^0)$ , this yields

$$0 = N(x^0) N^+(x^0) \nabla f(x^0) + N(x^0) q^0. \quad (3.2.5)$$

Substituting  $q^0$  from (3.2.3) in (3.2.5), yields

$$0 = N(x^0) N^+(x^0) \nabla f(x^0) + N(x^0) [-N^+(x^0) \nabla f(x^0) + (h(x^0)/\rho)]$$

or

$$0 = N(x^0) (h(x^0)/\rho)$$

Since  $N(x^0)$  is of rank  $m$ , this implies that  $h(x^0) = 0$ .

### Section 3.3 The New Method

For the equality constrained problem (B), min.  $f(x)$  subject to  $h(x) = 0$ , consider the equivalent unconstrained problem

$$(P) \min_{x, q} J(x, q, \rho) \triangleq (1/2) \{ \langle \gamma, \gamma \rangle + \langle e, e \rangle \}, \quad (3.3.1)$$

where the  $n$ -vector  $\gamma$  and  $m$ -vector  $e$  are defined as

$$\gamma(x, q) = \nabla f(x) + N(x) q \quad (3.3.2)$$

$$e(x, q) = \rho q + \rho N^+(x) \nabla f(x) - h(x). \quad (3.3.3)$$

In the following theorem, it is shown that the constrained problem (B) and the unconstrained problem (P) are equivalent, in the sense, that the extremum points of the problem (B), satisfying the necessary conditions (3.1.4) and (3.1.5), are the stationary points of the problem (P) and vice versa.

#### Theorem 3.3.1

Let  $f$  and  $h_i$ ,  $i = 1, 2, \dots, m$  be differentiable at a  $\bar{x} \in \mathbb{R}^n$  and

let  $\rho$  be a scalar. If  $(\bar{x}, \bar{\mu})$  is an extremum point of the problem (B), then  $\bar{x}$  and  $\bar{q}$  defined in (3.3.7) below, constitute a stationary point of the problem (P). Conversely, if  $(\bar{x}, \bar{q})$  is a stationary point of  $J(x, q, \rho)$ , then  $\bar{x}$  and  $\bar{\mu}$  defined by (3.3.7) below, constitute an extremum point of the constrained problem (B), provided  $\rho$  is such that  $(m \times m)$  matrix  $D_1$ , where

$$D_1 = \left[ \rho I_m - N^T(x)(F(x) + \sum_{i=1}^m H_i(x)q_i)^{-1} \{ \rho \nabla(N^+(x) \nabla f(x))^T - N(x) \} \right], \quad (3.3.4)$$

evaluated at  $\bar{x}$  and  $\bar{q}$ , is nonsingular.

Proof:-

Suppose  $(\bar{x}, \bar{\mu})$  yields an extremum to the problem (B), then

$$\nabla f(\bar{x}) + N(\bar{x}) \bar{\mu} = 0 \quad (3.3.5)$$

and

$$h(\bar{x}) = 0. \quad (3.3.6)$$

Define  $\bar{q} \in R^m$  as follows

$$\bar{q} = \bar{\mu} = -N^+(\bar{x}) \nabla f(\bar{x}). \quad (3.3.7)$$

At  $(\bar{x}, \bar{\mu})$ , it is easily verified from (3.3.3), using (3.3.7), that  $e(\bar{x}, \bar{q}) = 0$ . From (3.3.2)

$$\begin{aligned}
\gamma(\bar{x}, \bar{q}) &= \nabla f(\bar{x}) + N(\bar{x}) \bar{q} \\
&= \nabla f(\bar{x}) + N(\bar{x}) \{ -N^+(\bar{x}) \nabla f(\bar{x}) \} \\
&= [I_n - N(\bar{x}) N^+(\bar{x})] \nabla f(\bar{x}) \\
&= \hat{P}(\bar{x}) \nabla f(\bar{x})
\end{aligned}$$

where  $\hat{P}(\bar{x}) = I_n - N(\bar{x}) N^+(\bar{x})$ . Since  $(\bar{x}, \bar{\mu})$  satisfies the necessary conditions (3.3.5), which are equivalent [19] to  $\hat{P}(\bar{x}) \nabla f(\bar{x}) = 0$ , this implies that  $\gamma(\bar{x}, \bar{q}) = 0$ .

Differentiating  $J(x, q, \rho)$ , given by (3.3.1), with respect to  $x$  and  $q$ ,

$$\nabla_x J(x, q, \rho) = \nabla_x \gamma^T(x, q) \cdot \gamma(x, q) + \nabla_x e^T(x, q) \cdot e(x, q) = 0 \quad (3.3.8)$$

$$\nabla_q J(x, q, \rho) = \nabla_q \gamma^T(x, q) \cdot \gamma(x, q) + \nabla_q e^T(x, q) \cdot e(x, q) = 0, \quad (3.3.9)$$

at  $(\bar{x}, \bar{q})$  since  $\gamma(\bar{x}, \bar{q}) = 0 = e(\bar{x}, \bar{q})$ . Thus  $(\bar{x}, \bar{q})$  is the stationary point of the problem (P).

Let  $(\bar{x}, \bar{q})$  be the stationary point of  $J(x, q, \rho)$ . Then at  $(\bar{x}, \bar{q})$

$$\nabla J = \begin{pmatrix} \nabla_x J \\ \nabla_q J \end{pmatrix} = 0.$$

By equations (3.3.8) and (3.3.9), this implies that, at  $(\bar{x}, \bar{q})$ ,

$$\begin{pmatrix} \nabla_{\mathbf{x}} J \\ \nabla_{\mathbf{q}} J \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{x}\gamma}^T & \nabla_{\mathbf{x}e}^T \\ \nabla_{\mathbf{q}\gamma}^T & \nabla_{\mathbf{q}e}^T \end{pmatrix} \begin{pmatrix} \gamma \\ e \end{pmatrix} = 0 \quad (3.3.10)$$

The system of equations (3.3.10) has a trivial solution,  $\gamma(\bar{\mathbf{x}}, \bar{\mathbf{q}}) = 0 = e(\bar{\mathbf{x}}, \bar{\mathbf{q}})$ , if the  $((n+m) \times (n+m))$  matrix

$$D = \begin{pmatrix} \nabla_{\mathbf{x}\gamma}^T & \nabla_{\mathbf{x}e}^T \\ \nabla_{\mathbf{q}\gamma}^T & \nabla_{\mathbf{q}e}^T \end{pmatrix} \quad (3.3.11)$$

is nonsingular. Assuming that the matrix  $D$ , evaluated at  $(\bar{\mathbf{x}}, \bar{\mathbf{q}})$ , is nonsingular, it can be easily verified that the solution  $\gamma(\bar{\mathbf{x}}, \bar{\mathbf{q}}) = 0 = e(\bar{\mathbf{x}}, \bar{\mathbf{q}})$  to (3.3.10) implies the satisfaction of the necessary conditions (3.3.5) and the constraints (3.3.6). Thus  $(\bar{\mathbf{x}}, \bar{\mathbf{q}})$  is an extremum of the constrained problem (B).

We will examine the conditions under which matrix  $D$ , of (3.3.11) is nonsingular. Differentiating  $\gamma$  and  $e$ , given by (3.3.2) and (3.3.3), with respect to  $\mathbf{x}$  and  $\mathbf{q}$ , yields

$$\nabla_{\mathbf{x}\gamma}^T = \mathbf{F}(\mathbf{x}) + \sum_{i=1}^m H_i(\mathbf{x}) \mathbf{q}_i$$

$$\nabla_{\mathbf{q}\gamma}^T = \mathbf{N}^T(\mathbf{x})$$

$$\nabla_{\mathbf{x}e}^T = \rho \nabla_{\mathbf{x}}(\mathbf{N}^+(\mathbf{x}) \nabla f(\mathbf{x}))^T - \mathbf{N}(\mathbf{x})$$

$$\nabla_{\mathbf{q}e}^T = \rho \mathbf{I}_m ,$$

where  $F(x)$  and  $H_i(x)$  are Hessians of  $f(x)$  and  $h_i(x)$ ,  $i=1, 2, \dots, m$ , respectively and  $I_m$  is an  $(m \times m)$  identity matrix. At  $(\bar{x}, \bar{q})$ ,  $\nabla_{x\gamma}^T$  is, therefore, the Hessian of the Lagrangian  $l$ , defined by (3.1.3), associated with the constrained problem (B). Assuming that  $(n \times n)$  matrix  $\nabla_{x\gamma}^T$  is nonsingular, then the determinant of the matrix  $D$  is given by

$$|D| = |\nabla_{x\gamma}^T| \cdot |\nabla_{qe}^T - \nabla_{q\gamma}^T (\nabla_{x\gamma}^T)^{-1} \nabla_{xe}^T|. \quad (3.3.12)$$

Since  $|\nabla_{x\gamma}^T| \neq 0$ , the determinant of the matrix  $D$  is nonzero if and only if

$$|\nabla_{qe}^T - \nabla_{q\gamma}^T (\nabla_{x\gamma}^T)^{-1} \nabla_{xe}^T| \neq 0$$

or if the matrix  $D_1$ , were

$$D_1 = \left[ \rho I_m - N^T(\bar{x}) \left( F(\bar{x}) + \sum_{i=1}^m H_i(\bar{x}) \bar{q}_i \right)^{-1} \left\{ \rho \nabla_{x(N^+(\bar{x}))} \nabla f(\bar{x}) \right\}^T - N(\bar{x}) \right],$$

is nonsingular.

Hence if  $\rho$  is chosen to satisfy (3.3.4), then there is a stationary point of  $J$  such that it is also an extremum point of the constrained problem (B).

### Section 3.4 Computational Implementation and Comments

The Penalty method approximates the optimal Lagrange multiplier  $\mu^*$ , associated with the problem (B), by a vector  $q \triangleq \frac{h(x^{\rho*})}{\rho}$  for a given positive  $\rho \neq 0$ . For a given  $\rho > 0$ , the approximation  $q = h(x^{\rho*})/\rho$  of  $\mu^*$  will in general not ensure that  $x^{\rho*}$  is feasible. The key development of this dissertation is the modification of the Lagrange approximating vector  $q$  to make  $x^{\rho*}$  feasible for any  $\rho > 0$  as suggested in equation (3.2.3). Moreover, it has been shown in lemma 3.2.1 that if there is a  $x$  and  $q$  satisfying the necessary conditions (3.1.4) and the relation (3.2.3), then for almost any  $\rho$ ,  $x$  is feasible.

While lemma 3.2.1 tends to imply that any value of  $\rho$  in  $\mathbb{R}$ , except  $\rho = 0$ , will suffice, the theorem 3.3.1 further restricts the choice of  $\rho$  by preventing the matrix  $D_1$  of (3.3.4) from being singular at the optimum  $x^*$ . There are only a finite number of choices for  $\rho$  for which  $D_1(x^*)$  is singular. To see this, write the matrix  $D_1$  as

$$D_1 = \rho (I_m - A_1) + A_2 \quad ,$$

where 
$$A_1 = N^T(x) (\nabla^2 l)^{-1} \nabla (N^+(x) \nabla f(x))^T$$

$$A_2 = N^T(x) (\nabla^2 l)^{-1} N(x) \quad .$$

The determinant of  $D_1$  is a  $m^{\text{th}}$  degree polynomial in  $\rho$ , which have, at most,  $m$  real-roots. There are, therefore, at the most  $m$  possible choices,  $\rho_1, \rho_2, \dots, \rho_m$ , of  $\rho$  that would make the matrix  $D_1(x^*)$  singular. Furthermore, if matrix  $A_1$  above is an identity matrix, then  $D_1$  will be nonsingular for any  $\rho \in \mathbb{R} - \{0\}$ , provided  $|A_2| \neq 0$ . This is the case when the constrained problem (B) is a Quadratic problem; that is

$$f(x) = \frac{1}{2} x^T Q x - x^T b + c \quad \text{and} \quad h(x) = Bx - d = 0,$$

where  $Q$  is a  $(n \times n)$  symmetric nonsingular matrix,  $b, c \in \mathbb{R}^n$ ,  $B$  is a  $(m \times n)$  matrix and  $d$  is a  $m$ -vector. For this case

$$\begin{aligned} A_1 &= BQ^{-1} Q B^T \\ &= B B^T \\ &= B ((B B^T)^{-1} B)^T \\ &= B B^T (B B^T)^{-1} \\ &= B B^T (B B^T)^{-1} \\ &= I_m \end{aligned}$$

and  $|D_1| = |B Q^{-1} B^T|$  is, therefore, independent of  $\rho$ .



A proper choice of  $\rho$ , however, requires a priori knowledge of  $x^*$ . Since there are only a finite number of values for  $\rho \in \mathbb{R}$  that would make  $D_1$  singular at  $x^*$ , it is highly unlikely that an arbitrary choice for  $\rho$  would constitute such a point. If this happens, the system of equations (3.3.10), to be satisfied by the stationary point  $\gamma$  and  $e$  of the function  $J$ , may have non-trivial solutions and the unconstrained minimum of  $J$  may occur at a  $(x, q)$  where  $\gamma \neq 0 \neq e$ . Such a  $x$  and  $q$  would not satisfy the necessary conditions of the constrained problem (B). The situation may be checked by pre-setting, in computation, an iteration-number  $I_0$  so that  $\|\gamma\| \leq \epsilon$  and  $\|e\| \leq \sigma$  for some  $I \leq I_0$ , where  $\epsilon$  and  $\sigma$  are preselected positive small numbers.

On the other hand if, from theorem 3.3.1,  $\rho$  is chosen such that  $D_1$  is nonsingular then  $\gamma = 0 = e$  is the only stationary point of  $J(x, q, \rho)$ . To verify that  $\gamma = 0 = e$  constitutes a minimum of  $J$  and not only the stationary point, we compute the second derivative of  $J$

$$\nabla^2 J = \begin{pmatrix} \nabla_{xx} J & \nabla_{xq} J \\ \nabla_{qx} J & \nabla_{qq} J \end{pmatrix} .$$

Differentiating  $\nabla_x J$  and  $\nabla_q J$ , given by equations (3.3.8) and (3.3.9), with respect to  $x$  and  $q$  and obtain, at the stationary point  $\gamma = 0 = e$ ,

$$\nabla^2 J = \begin{pmatrix} \nabla_x \gamma^T & \nabla_x e^T \\ \nabla_q \gamma^T & \nabla_q e^T \end{pmatrix} \begin{pmatrix} \nabla_x \gamma^T & \nabla_q \gamma^T \\ \nabla_x e^T & \nabla_q e^T \end{pmatrix} = D D^T .$$

Since  $\nabla^2 J$  is symmetric and  $D$  is nonsingular, this implies that the second-derivative of  $J$  is positive definite at  $\gamma=0=e$  and hence the point is a minimum of  $J$ .

The proposed method seeks an optimum point  $x^*$  of the constrained problem (B) by solving an equivalent unconstrained problem (P). The point  $(x^*, q^*)$ , so sought, satisfies the necessary conditions (3.1.4) and (3.1.5) associated with the constrained problem (B) and is, therefore, only an extremum point of the problem (B). If it were known that  $x^*$  also satisfies the second-order sufficient conditions of the theorem 2.2.3, then  $x^*$  would be the minimum of the problem (B). No attempt is made to check the second-order conditions. However, if the problem (B) is a convex programming problem; that is, if  $f(x)$  is a convex function and  $h(x)$  are linear, then the necessary conditions (3.1.4) and (3.1.5) are also sufficient for  $x^*$  to be a minimum.

The Multiplier method applied to the problem (B) constructs the unconstrained problem of minimizing the augmented Lagrangian

$$M(x, \mu^*, \rho) = f(x) + \mu^{*T} h(x) + \frac{1}{2\rho} h^T(x) h(x) . \quad (3.4.1)$$

Since the optimal Lagrange multiplier,  $\mu^*$ , is not known a priori, some estimate  $\mu^k$  of  $\mu^*$  is made and  $M(x, \mu^k, \rho)$  is minimized to obtain  $x^{k*}$ . The estimate  $\mu^k$  of  $\mu^*$  is updated by the relation

$$\mu^{k+1} = \mu^k + \frac{1}{\rho} h(x^{k*}) \quad . \quad (3.4.2)$$

and the next cycle of minimization of  $M(x, \mu^{k+1}, \rho)$  is executed.

This process is repeated unless for some  $k$ ,  $x^{k*}$  is feasible, that is,  $h(x^{k*}) = 0$ , giving  $\mu^*$  by  $\mu^* = \mu^{k+1} = \mu^k$ . For such  $x^{k*}$

$$\nabla_x M(x^{k*}, \mu^k, \rho) = \nabla f(x^{k*}) + \mu^{T*} \nabla h(x^{k*}) = 0 \quad ,$$

hence  $x^{k*}$  is a constrained minimum.

Instead of following the Hestenes' scheme (3.4.2), if we estimate  $\mu^*$  by setting  $\mu = -N^+(x) \nabla f(x)$ . This would reduce  $M(x, \mu, \rho)$  of (3.4.1) to

$$M(x, \rho) = f(x) - h^T(x) N^+(x) \nabla f(x) + \frac{1}{2\rho} h^T(x) h(x) \quad ,$$

which is exactly the Fletcher's Exact Penalty function  $\phi$  of equation (2.3.1).

For  $M(x, \mu, \rho)$  given by (3.4.1)

$$\nabla_x M(x, \mu, \rho) = \nabla f(x) + \nabla h(x) \left[ \mu + \frac{1}{\rho} h(x) \right] \quad ,$$

and for  $\mu = -N^+(x) \nabla f(x)$

$$\nabla_x M(x, \rho) = \nabla f(x) + \nabla h(x) \left[ -N^+(x) \nabla f(x) + \frac{1}{\rho} h(x) \right] .$$

Defining  $q = -N^+(x) \nabla f(x) + \frac{1}{\rho} \nabla h(x)$ , if  $x$  is such that  $\nabla_x M(x, \rho) = 0$ , then by lemma 3.2.1,  $h(x) = 0$ . Defining  $\gamma$  and  $e$ , the errors in satisfying

$$\nabla f(x) + \nabla h(x) q = 0$$

and

$$q + N^+(x) \nabla f(x) - h(x)/\rho = 0 ,$$

the proposed method seeks the constrained solution by minimizing the sum of the norms of the error functions, as

$$\min_{x, q} J(x, q, \rho) = \frac{1}{2} \{ \langle \gamma, \gamma \rangle + \langle e, e \rangle \} .$$

Both the Multiplier and Exact Penalty methods restrict the choice of penalty parameter,  $\rho$ , to positive real numbers, whereas for the proposed method,  $\rho$  may take almost any value in  $\mathbb{R}$ . Moreover, for both the Multiplier and Exact Penalty methods,  $\rho$  must also be chosen below a certain threshold value  $\rho^*$  for these methods to converge. The evaluation of  $\rho^*$  or seeking satisfaction of  $\rho \leq \rho^*$  normally require some additional efforts and computation time. Hence for a given problem and an arbitrary choice of  $\rho$ , it is expected that the proposed method will consume less computing time than either the Multiplier and Exact Penalty methods.

The new method avoids solving a large number of unconstrained problems as normally required by the Penalty and Multiplier methods. It is, therefore, expected that the number of iterations required to solve a constrained problem using the new method may be significantly less than that required by the Multiplier method.

Both the new and Exact Penalty methods assume the existence of the generalized inverse  $N^+(x)$  of  $N(x)$  for all  $x$  in some neighborhood of  $x^*$ . This requires that the constraint functions,  $h(x)$ , be such that  $h_i$ ,  $i=1, 2, \dots, m$ , are linearly independent not only at  $x^*$  but also so in some neighborhood of  $x^*$ . This restricts the class of problems to which these methods may be applied. The main computational disadvantage of the new method, as with Exact Penalty methods is, that  $N^+(x)$  is to be evaluated for each  $x$ . Efficient unconstrained minimization algorithms also use derivative information, thus the implementation of these methods not only require evaluation of  $N^+(x)$  but also of  $\nabla N^+(x)$  for each. The Multiplier method, however, does not suffer from these disadvantages.

#### IV. COMPUTATIONAL RESULTS

To evaluate the effectiveness and efficiency of the proposed method, hereafter referred to as the Semi-Dual method, its performance is compared with that of Hestenes' Multiplier and Fletcher's Exact Penalty methods in the computational solution of a number of nonlinear programming problems. The Conjugate-gradient algorithm was employed to solve the resulting unconstrained optimization problems and the Fibonacci search routine was used in the step-size determination in each case. A short description of the Conjugate-gradient algorithm and Fibonacci search method is given in Appendix A and B respectively.

##### Section 4.1 Constrained Minimization Algorithms

Consider again the equality constrained nonlinear program:

$$(B) \quad \begin{cases} \min f(x) & (4.1.1) \\ \text{subject to } h(x) = 0, & (4.1.2) \end{cases}$$

where  $x \in \mathbb{R}^n$ ,  $h \in \mathbb{R}^m$ .

##### (A) Semi-Dual Method (Method-SD)

As mentioned in Chapter III, the Semi-Dual method seeks to solve the constrained problem (B) by minimizing an equivalent unconstrained problem:

$$J(x, q, \rho) = \frac{1}{2} \{ ||\gamma||^2 + ||e||^2 \} \quad (4.1.3)$$

where

$$\gamma = \nabla f(x) + N(x)q \quad (4.1.4)$$

$$e = \rho q + \rho N^+(x) \nabla f(x) - h(x) , \quad (4.1.5)$$

$x \in R^n$ ,  $m$ -vector  $q$  being an approximation to the Lagrange multiplier associated with the constrained problem. The  $n$ -vector  $\gamma$  is the error in satisfying the necessary conditions (3.2.2) corresponding to the penalty function defined by the equation (3.2.1) and  $m$ -vector  $e$  is the error in satisfying equation (3.2.3) approximating the Lagrange multiplier.

### Numerical Implementation

The following steps were used in the implementation of Semi-dual method:

- (I) The original constrained problem (B) is replaced by the unconstrained problem defined by equations (4.1.3) - (4.1.5).
- (II) A suitable value for  $\rho$  is chosen.
- (III) For the chosen value of  $\rho$ , the function  $J(x, q, \rho)$  is minimized with respect to  $(x, q)$ . The minimum of  $J(x, q, \rho)$  is achieved when

$$||\nabla_x J(x, q, \rho)||^2 \leq \epsilon_1 \text{ and } ||\nabla_q J(x, q, \rho)||^2 \leq \epsilon_2 ,$$

for preselected small values of  $\epsilon_1 > 0$  and  $\epsilon_2 > 0$ .

(B) Method of Multipliers (Method -MM)

This method is based on the construction of an augmented penalty function

$$M(x, \mu, \rho) = f(x) + \mu^T h(x) + \frac{1}{2\rho} h^T(x) h(x) , \quad (4.1.6)$$

This is obtained by adding to the penalty function  $P(x, \rho)$ , given by equation (3.1.7), a linear term in the constraint  $h(x)$  where the  $m$ -vector  $\mu$  is an approximation to the Lagrange multiplier. The use of this function was suggested by Hestenes [28] to circumvent the numerical difficulties associated with the extremely small values of the penalty constant  $\rho$ , required in normal Penalty methods. The Method-MM replaces the constrained problem (B) by a sequence of unconstrained minimization problems. In each element of the sequence (cycle), function  $M(x, \mu, \rho)$  is minimized with respect to  $x$  for a given  $\mu$  and a sufficiently small  $\rho > 0$ . Theoretically, therefore, the following necessary condition must be satisfied at the end of each cycle:

$$\nabla_x M(x, \mu, \rho) = \nabla f(x) + \nabla h(x) \left[ \mu + \frac{1}{\rho} h(x) \right] , \quad (4.1.7)$$

and  $\mu$  is updated according to the rule

$$\mu_2 = \mu_1 + \frac{1}{\rho} h(x) \quad (4.1.8)$$



where  $\mu_1$  denotes the Lagrange multiplier of the present cycle and  $\mu_2$  denotes the Lagrange multiplier of the next cycle. Equation (4.1.7) then reduces to

$$\nabla_x M(x, \mu, \rho) = \nabla f(x) + \mu_2^T \nabla h(x) = 0$$

and the combination of  $x$  and  $\mu_2$ , so obtained, satisfies the optimum condition exactly. The drive towards constrained satisfaction is supplied by changing the multiplier from cycle to cycle in accordance with the equation (4.1.8).

### Numerical Implementation

From the above considerations, the following steps for the numerical implementation of Method-MM emerges:

- (I) The original constrained problem (4.1.1), (4.1.2) is replaced by the unconstrained problem given by (4.1.6).
- (II) Initial values for  $\mu$  and  $\rho$  are chosen. An appropriate initial choice for  $\mu$  may be  $\mu = 0$ , which is equivalent to stating that the augmented penalty function (4.1.6) and the penalty function (3.1.7) are identical for the first cycle of the method.
- (III) For the given  $\mu$  and  $\rho$ , the function  $M(x, \mu, \rho)$  of (4.1.6) is minimized with respect to  $x$ . The minimum is achieved when the following stopping condition is satisfied

$$|| \nabla_{\mathbf{x}} M(\mathbf{x}, \mu, \rho) ||^2 \leq \epsilon_3 ,$$

where  $\epsilon_3$  is a preselected small positive number.

- (IV) The solution point of any given cycle is chosen as the starting point for the next cycle.
- (V) For the next cycle, the multiplier  $\mu$  is updated according to the rule of equation (4.1.8) and return to step (II).
- (VI) The algorithm is terminated when

$$|| h(\mathbf{x}) ||^2 + || \nabla_{\mathbf{x}} M(\mathbf{x}, \mu, \rho) ||^2 \leq \epsilon_4$$

for some preselected small value of  $\epsilon_4 > 0$ .

(C) Exact Penalty Method (Method-EP)

Fletcher [19] has found a whole class of exact penalty functions for the equality constrained problem (B), a representative member of which is given by

$$\phi(\mathbf{x}) = f(\mathbf{x}) - \mathbf{h}^T(\mathbf{x}) \mathbf{N}^+(\mathbf{x}) \nabla f(\mathbf{x}) + \frac{1}{2\rho} \mathbf{h}^T(\mathbf{x}) \mathbf{h}(\mathbf{x}) . \quad (4.1.9)$$

This can be easily verified that at a local solution point  $\mathbf{x}^*$  of the problem (B), the function  $\phi$  of equation (4.1.9) is stationary. If  $\mathbf{x}^*$  also satisfies the sufficient conditions for a strict local minimum, as given by theorem 2.2.3, then there exists a  $\rho^* > 0$  (see theorem 2.3.1) such that for every  $\rho \leq \rho^*$ , the Hessian of  $\phi$ ,  $\nabla^2 \phi(\mathbf{x}^*)$ , is

positive definite. Thus, for a proper choice of  $\rho$ , the solution point,  $x^*$ , of the constrained problem (B), will also minimize the function  $\phi$ .

The only problem remaining before the Exact Penalty method can be implemented, is to assign a value to  $\rho$  so that a single unconstrained minimization of the function  $\phi(x)$  would yield the sought minimum of the constrained problem (B). For quadratic  $f(x)$  and linear  $h(x)$  any choice for  $\rho$ , such that  $\rho_1 \geq \|N^+(x)\| \|\nabla^2 f(x) N^{T+}(x)\|$ , where  $\rho_1 = 1/\rho$ , would suffice [19]. For nonlinear constraints, however, the computation of  $\rho^*$  is by no means simple. Fletcher and Lill [21] suggest taking

$$\rho_1 = 10 \|\nabla^2 f(x^0)\| + r$$

where  $x^0$  is the starting point for the unconstrained minimization and  $r=1$  is taken initially. However,  $\rho_1$  must be adjusted by increasing  $r$ , if during the course of minimization, a negative definite  $\nabla^2 \phi$  is encountered.

### Numerical Implementation

The Exact Penalty method can be computationally implemented through the following steps:

- (I) The original constrained problem (4.1.1) - (4.1.2) is replaced by the unconstrained problem (4.1.9).

- (II) A suitable value for  $\rho > 0$  is chosen.
- (III) For the chosen value of  $\rho$ , the function  $\phi(x)$ , given by (4.1.9), is minimized with respect to  $x$  using an unconstrained minimization algorithm. The minimum of  $\phi(x)$  is achieved when

$$\|\nabla\phi\|^2 \leq \epsilon_5$$

for some small value of  $\epsilon_5 > 0$ .

Remark: Since the purpose of the computations was to compare the performance of the three mentioned optimization methods, the selection rule, outlined above for selecting  $\rho$  was not employed. We used, instead, the  $\rho$  values employed for the Semi-Dual method.

## Section 4.2 Experimental Conditions and Numerical Examples

Four nonlinear programming problems used in the computational experimentation are described below. All computations were performed in FORTRAN on a CDC-6600 computer. As mentioned above, a Conjugate gradient algorithm was employed in each method with a Fibonacci routine to establish the optimal step-length at each step.

### Starting Point of the Algorithm

In all examples, an arbitrary starting point of

$$x_1 = x_2 = \dots = x_n = 2.0, \quad x \in \mathbb{R}^n,$$

was selected for each method.

### Convergence of the Algorithm

Convergence of an algorithm, is attained when the norm of the cost-gradient reduces to less than a preselected small positive number. In all cases we choose  $\epsilon$  to be  $10^{-7}$ . Conversely, non-convergence of an algorithm is assumed if the number of iterations required exceeded 500.

### Test Problems

The following four numerical examples were selected [48] to test the constrained algorithms described in the section 4.1. As a bench mark, we included as the first example, a problem with linear constraints and quadratic performance index.

Example 4.2.1: Consider the problem of minimizing the function

$$f(x) = (x_1 - x_2)^2 + (x_2 + x_3 - 2)^2 + (x_4 - 1)^2 + (x_5 - 1)^2$$

subject to the constraints

$$h_1 \equiv x_1 + 3x_2 = 0$$

$$h_2 \equiv x_3 + x_4 - 2x_5 = 0$$

$$h_3 \equiv x_2 - x_5 = 0$$

The function admits the relative minimum  $f = 4.0930$  at the point

$$x_1 = -0.7674, \quad x_2 = 0.2558, \quad x_3 = 0.6279, \quad x_4 = -0.1162,$$

$$x_5 = 0.2558$$

with  $\mu_1 = 2.0465$ ,  $\mu_2 = 2.2325$ ,  $\mu_3 = -5.9534$ .

Example 4.2.2: Consider the problem of minimizing the function

$$f(x) = (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^4$$

subject to the constraints

$$h_1 \equiv x_1 (1 + (x_2)^2) + (x_3)^4 - 4 - 3\sqrt{2} = 0.$$

The function admits the relative minimum  $f = 0.0325$  at the point

$$x_1 = 1.1048, \quad x_2 = 1.1966, \quad x_3 = 1.5352, \quad \text{with } \mu_1 = -0.0107.$$

Example 4.2.3: Consider the problem of minimizing the function

$$f(x) = (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^4 + (x_4 - x_5)^4 .$$

subject to the constraints

$$h_1 \equiv x_1 + (x_2)^2 + (x_3)^3 - 2 - 3\sqrt{2} = 0$$

$$h_2 \equiv x_2 - (x_3)^2 + x_4 + 2 - 2\sqrt{2} = 0$$

$$h_3 \equiv x_1 x_5 - 2 = 0 .$$

The function admits the relative minimum  $f = 0.0787$  at the point

$$x_1 = 1.1911, \quad x_2 = 1.3626, \quad x_3 = 1.4728, \quad x_4 = 1.6350,$$

$$x_5 = 1.6790$$

with  $\mu_1 = -0.0388$ ,  $\mu_2 = -0.0167$ ,  $\mu_3 = -0.2879 \times 10^{-3}$  .

Example 4.2.4: Consider the problem of minimizing the function

$$f(x) = \log(x_3) - x_2$$

subject to the constraints

$$h_1 \equiv (x_2)^2 + (x_3)^2 - 4 = 0$$

$$h_2 \equiv x_3 - 1 \geq 0 .$$

Introducing an auxiliary variable  $x_1$ , defined by

$$x_3 = 1 + (x_1)^2$$

the previous problem is transformed to an equivalent equality constrained problem of minimizing

$$f(x) = \log(1 + (x_1)^2) - x_2$$

subject to the constraint

$$h_1 = (1 + (x_1)^2)^2 + (x_2)^2 - 4 = 0 .$$

This problem admits the relative minimum  $f = -\sqrt{3}$  at the point

$$x_1 = 0 , \quad x_2 = \sqrt{3} , \quad \text{with } \mu_1 = \frac{1}{2\sqrt{3}} .$$

### Section 4.3 Computational Results and Interpretation

The test problems described in section 4.2 were solved using the Semi-Dual method, Hestenes' method of Multiplier and Fletcher's Exact Penalty method and the computational results are presented in Tables 1, 2 and 3, respectively. Included in the numerical results of Tables 1, 2 and 3, are the number of iterations required for convergence and the actual computing time required for each test problem.



Table 1. Results obtained for the semi-dual method.

Penalty Parameter $\rho$	Example Number							
	4.2.1		4.2.2		4.2.3		4.2.4	
	No. of Iterations	CPU Time in Seconds	No. of Iterations	CPU Time in Seconds	No. of Iterations	CPU Time in Seconds	No. of Iterations	CPU Time in Seconds
0.1	33	1.241	27	1.228	62	13.284	6	0.220
0.01	34	1.316	28	1.293	53	10.799	6	0.217
0.001	32	1.244	28	1.293	52	10.420	6	0.229

Table 2. Results obtained using the multiplier method.

Penalty Parameter $\rho$	Example Number							
	4.2.1		4.2.2		4.2.3		4.2.4	
	No. of Iterations	CPU Time in Seconds	No. of Iterations	CPU Time in Seconds	No. of Iterations	CPU Time in Seconds	No. of Iterations	CPU Time in Seconds
0.1	93	1.159	78	0.983	133	3.026	14	0.138
0.01	406	3.689	130	1.531	218	4.073	187	1.522
0.001	382	3.513	(a)	-	(a)	-	(a)	-

(a) Number of iterations exceeded 500.

Table 3. Results obtained using exact penalty method.

Penalty Parameter $\rho$	Example Number							
	4.2.1		4.2.2		4.2.3		4.2.4	
	No. of Iterations	CPU Time in Seconds	No. of Iterations	CPU Time in Seconds	No. of Iterations	CPU Time in Seconds	No. of Iterations	CPU Time in Seconds
0.1	75	1.395	71 <sup>(c)</sup>	1.132	(b)	-	18	0.227
0.01	242	4.529	(a)	-	(b)	-	125	1.568
0.001	(a)	-	(b)	-	(a)	-	(b)	-

(a) Number of iterations exceeded 500.

(b) Nonconvergence

(c) Converges to a different stationary point.

A comparative study of Table 1, 2, and 3 shows that, for all test problems the Semi-Dual method converges, for all the values of  $\rho$  considered, in a smaller number of iterations than either the Multiplier or the Exact Penalty method. In some cases and for some choice of  $\rho$ , these methods do not converge at all.

In all the test-problem cases considered, the Semi-Dual approach does not exhibit a significant change in the number of iterations required for convergence for variations in the values of  $\rho$ . This is not the case with either the Multiplier or the Exact Penalty method where the iterations required for the convergence to a solution of a particular problem, depends very much on the value of  $\rho$ .

It is evident from the computational results that, depending on the problem, the Semi-Dual and Exact Penalty methods could require more actual computing time than the Multiplier method. This is largely due to the evaluation of  $N^+(x)$  and its derivative, required in both of the former methods.

Figures 1, 2, 3 and 4 show the plots of the constrained objective  $f(x)$  versus iteration for every test-problem with  $\rho = 0.1$ .

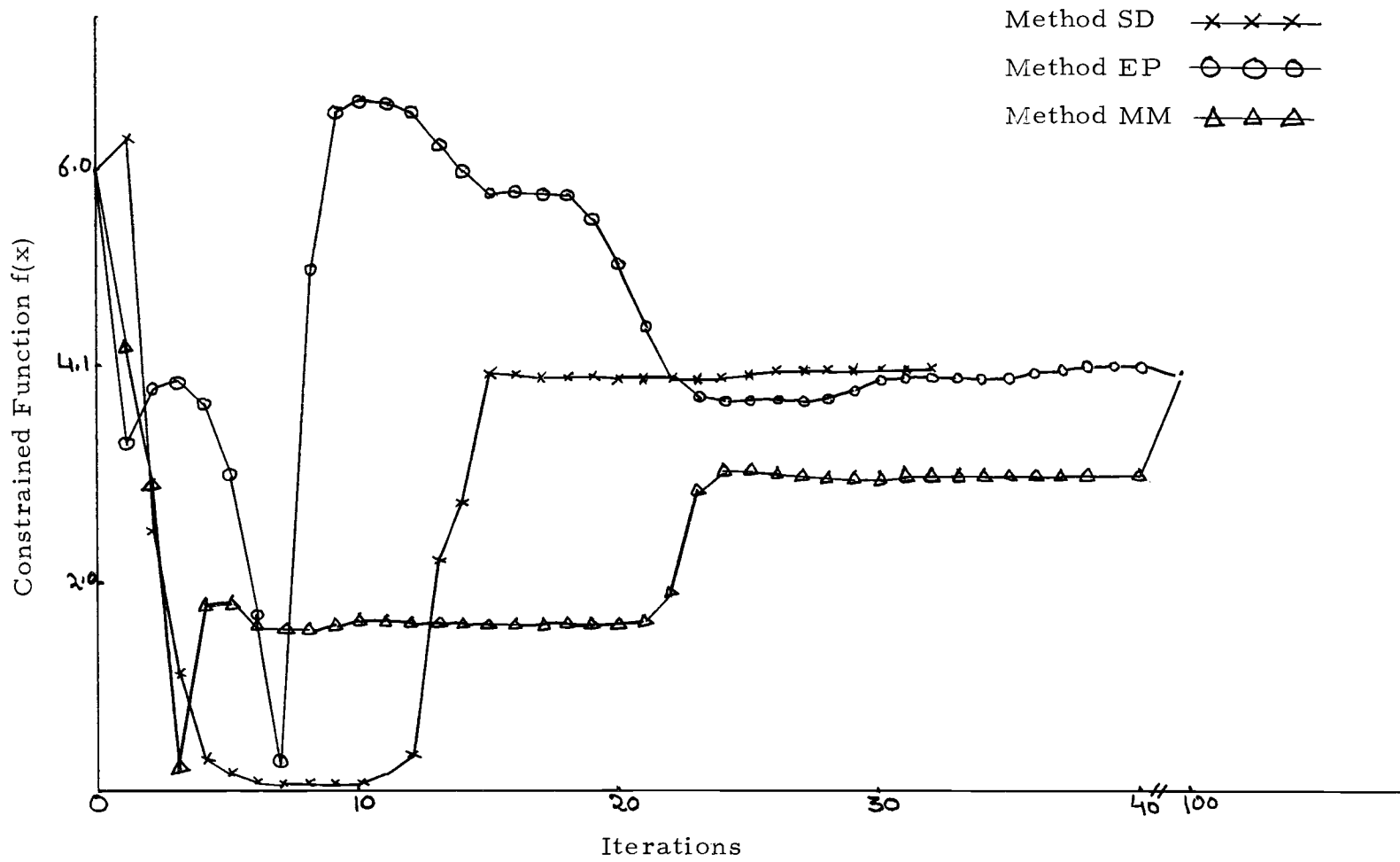


Figure 1. Solution to problem number 4.2.1 for  $\rho = 0.1$ .

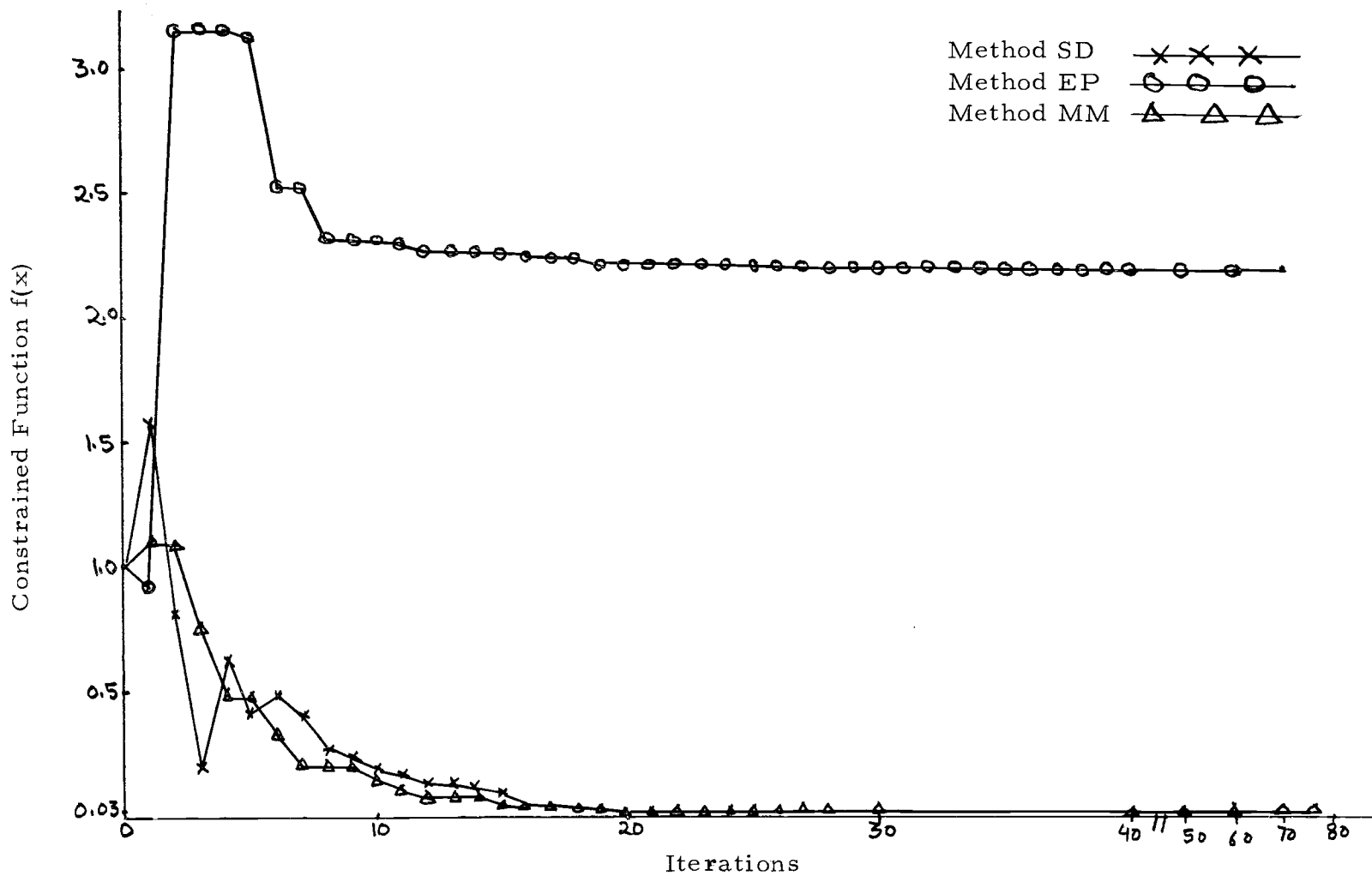


Figure 2. Solution to problem number 4.3.2 for  $\rho = 0.1$ .

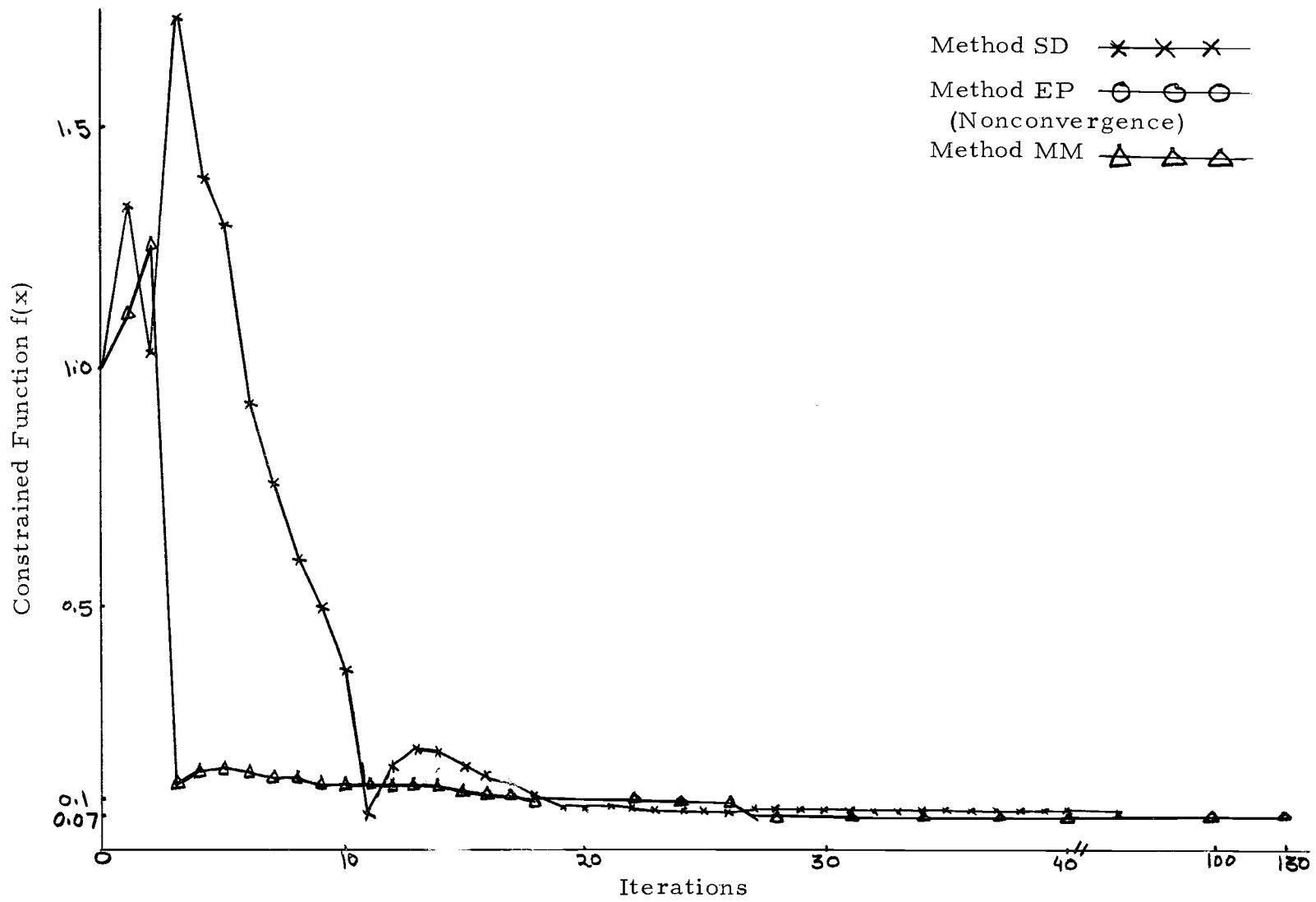


Figure 3. Solution to problem number 4.2.3 for  $\rho = 0.1$ .

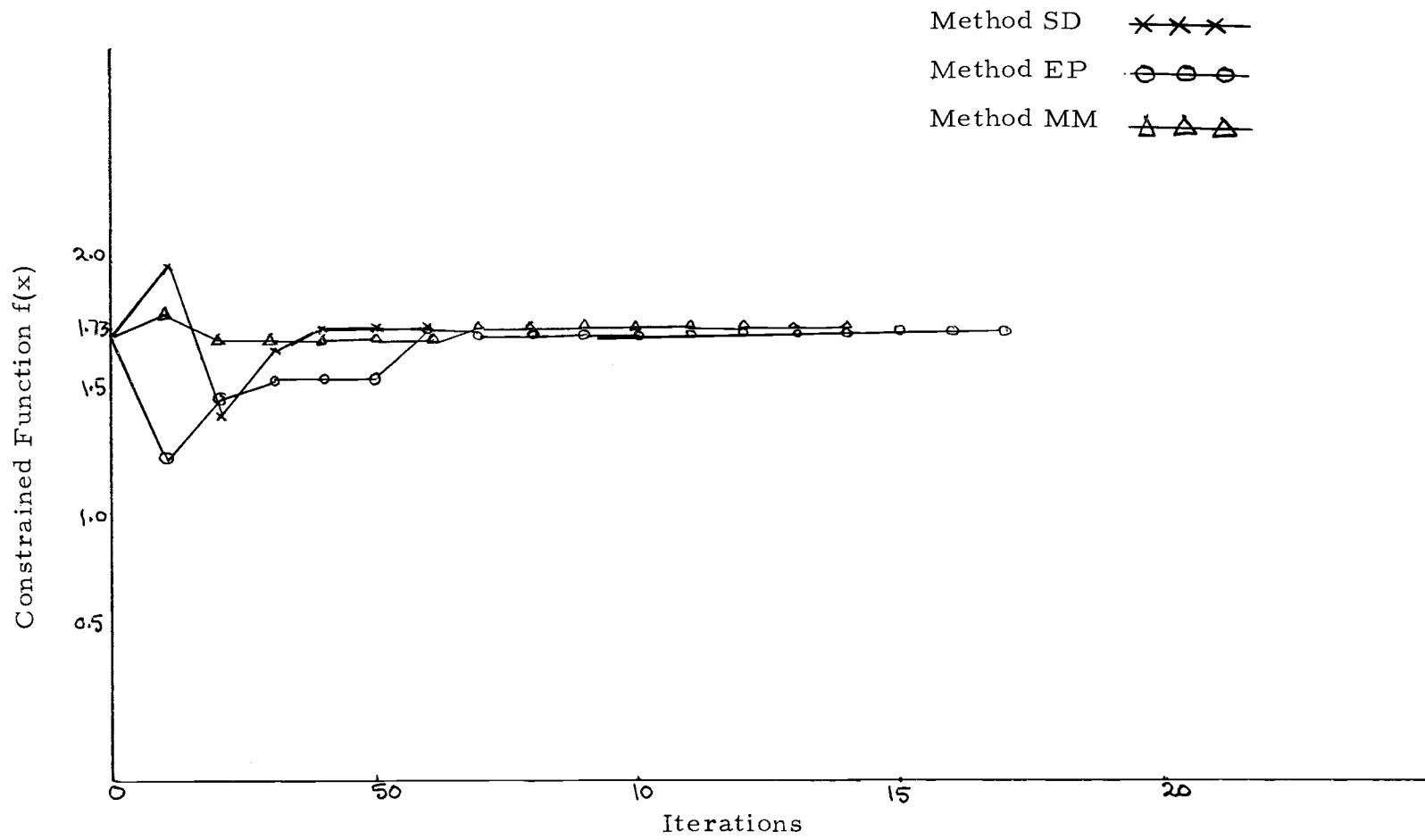


Figure 4. Solution to problem number 4.2.4 for  $\rho = 0.1$ .

## V. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

This dissertation successfully deploys the idea of simultaneously satisfying the necessary conditions and constraints associated with constrained nonlinear programs. As a result a new method is developed which avoids performing "iterations within an iteration" as required by the penalty method. The following conclusions are significant.

1. The method solves constrained problem through a single optimization of an equivalent unconstrained problem. This avoids solving a sequence of unconstrained optimization problems as required by the Penalty method.
2. The method would solve a problem almost for any choice for the penalty parameter  $\rho$ . This is in contrast with the Penalty method, which requires  $\rho$  to be positive and large and with the Multiplier and Exact Penalty methods which require  $\rho$  to be positive and large enough to be above a known threshold value  $\rho^*$  of  $\rho$ .
3. Since the proposed method does not require a large value for  $\rho$ , the Hessian of the equivalent unconstrained problem would remain well-conditioned. This avoids the problem of slow convergence caused by ill-conditioning of the Hessian associated with the Penalty method.

4. On the basis of limited numerical experiments, it is concluded that the choice of  $\rho$  does not significantly affect the number of iterations required for the convergence of the Semi-dual method. One may, therefore, select a  $\rho$  without the fear of excessively increasing the number of iterations required for the convergence.

In conclusion, the proposed Semi-dual method seems to be efficient procedure to solve equality constrained nonlinear programs. Further research and extensions are suggested in the following areas:

1. Extension of the method to general inequality-equality constrained programming problem (A).
2. Extension of the method to optimal control problems.
3. Global convergence and convergence-rate analysis of the method.
4. Development of approximation methods to determine  $N^+(x)$ .
5. Determination of sensitivity of the method with respect to penalty parameter  $\rho$ .



## BIBLIOGRAPHY

1. Abadie, J. "On the Kuhn-Tucker Theorem" in *Nonlinear Programming*, J. Abadie edited, North-Holland Publishing Co., Amsterdam, 1967.
2. Bazaraa, M. S., J. J. Goode and C. M. Shetty. "Constraint qualifications revisited", *Management Science*, 18, p. 567-573, 1972.
3. Bertsekas, D. P. "On Penalty and Multiplier methods for constrained minimization", *SIAM J. Control*, 14, 2, p. 216-235, 1976.
4. Brooks, R. and A. Geoffrion. "Finding Everett's Lagrangian multipliers by linear programming", *Operation Res.*, 14, p. 1148-1153, 1966.
5. Bui, T. L. and P. Huard. "La methode des centres dans un espaces topologique", *Numer. Math.*, 8, p. 56-67, 1966.
6. Buys, J. D. "Dual algorithm for constrained optimization problems" *Doctoral Dissertation*, University of Leiden, The Netherlands, 1972.
7. Carroll, C. W. "An operation research approach to the Economic optimization of a Kraft Pulping process." *Ph. D. Dissertation*, Institute of Paper Chemistry, Appleton, Wis., 1955.
8. Conn, A. R. "Constrained optimization using a non-differentiable penalty function" *SIAM J. Numer. Anal.*, 10, p. 760-784, 1973.
9. Conn, A. R. and T. Pietrzykowski. "A penalty function method converging directly to a constrained optimum." *RE Res. Report*, 73-11, Dept. of Combinatorics & Optimization, Univ. of Waterloo, Ontario, Canada, May, 1973.
10. Courant, R. "Variational methods for the solution of problems of equilibrium and vibrations" *Bull. Am. Math. Soc.*, 49, p. 1-23, 1943.

11. Dubovitsky, A. Y. and A. A. Milyutin. "Extremum problems in the presence of restrictions" USSR Comp. Math. and Math. Phys., 5, 3, p. 1-80, 1965.
12. Evans, J. P. and F. J. Gould. "On using equality constrained algorithms for inequality constrained problems" Math. Prog., Vol. 2, No. 3, p. 324-329, 1972.
13. Everett, H. "Generalized Lagrange multiplier method for solving problems of optimal allocation of resources" Operation Res., 11, p. 399-417, 1963.
14. Falk, J. E. "Lagrange multipliers and nonlinear programming" J. Math. Anal. Appls., 19, p. 141-159, 1967.
15. Fiacco, A. V. and G. P. McCormick, "Nonlinear Programming; Sequential Unconstrained Minimization Techniques." John Wiley and Sons, New York, 1968.
16. Fiacco, A. V. "Second-order sufficient conditions for weak and strict constrained minima" SIAM J. Appl. Math., 16, p. 105-108, 1968.
17. Fiacco, A. V. "Sequential unconstrained minimization methods for nonlinear programming." Ph.D. Dissertation, Northwestern University, Evanston, Ill., 1967.
18. Fiacco, A. V. and G. P. McCormick. "The sequential unconstrained minimization technique without parameters" Operation Res., 15, p. 820-827, 1967.
19. Fletcher, R. "A class of methods for nonlinear programming with termination and convergence properties" in Integer & Nonlinear Programming, J. Abadie edited, North-Holland Publishing Co., Amsterdam, 1970.
20. Fletcher, R. "A class of methods for nonlinear programming III, rates of convergence" in Numerical Methods for Nonlinear Optimization, F. A. Lootsma edited, Academic Press, London, 1972.
21. Fletcher, R. and S. A. Lill. "A class of methods for Nonlinear Programming II; Computational experience" in Nonlinear Programming, J. B. Rosen et al. edited, Academic Press, New York, 1970.

22. Fletcher, R. "An exact penalty function for nonlinear programming with inequalities" *Math. Prog.*, 5, p. 129-150, 1973.
23. Fletcher, R. and C. M. Reeves. "Function minimization by conjugate gradients" *Computer J.*, 7, p. 149-154, 1964.
24. Frisch, K. R. "The logarithmic potential method of convex programming" University Institute of Economics, Oslo, Norway, Memorandum, May 13, 1955.
25. Gould, F. J. and J. W. Tolle. "Geometry of Optimality conditions and constraint qualifications" *Math. Prog.*, 2 p. 1-18, 1972.
26. Gould, F. J. and J. W. Tolle. "A necessary and sufficient qualification for constrained optimization" *SIAM J. Appl. Math.*, 20, p. 164-172, 1971.
27. Hestenes, M. R. "Calculus of variation and Optimal control theory" John Wiley & Sons, 1966.
28. Hestenes, M. R. "Multiplier and Gradient Methods" *J. Opt. Theory & Appl.*, 4, p. 303-320, 1969.
29. Huard, P. "Resolution des P. M. a contraintes Non-lineaires par la Methode des centres" Note E. D. F, HR5690/3/317, 1964.
30. Huard, P. "Resolution of mathematical programming with nonlinear constraints by the method of centers" in J. Abadie edited *Nonlinear Programming*, p. 207-219, North-Holland Publishing Co., Amsterdam, 1967.
31. Huard, P. "Programmation mathematique convexe" *RIRO* 7, p. 43-59, 1968.
32. John, F. "Extremum problems with inequalities as a subsidiary conditions" in *Studies and Essays, Courant Anniversary Volume*, K. D. Friedrichs et al. edited, Interscience Publishers, New York, 1948.
33. King, R. P. "Necessary and sufficient conditions for inequality constrained extreme values" *Ind. & Eng. Chem. Fund.*, 5, p. 484-489, 1966.

34. Kort, B. W. and D. P. Bertsekas. "A new penalty function method for constrained minimization" in the Proceedings of 1972 IEEE Conference on Decision and Control, New Orleans, December, 1972.
35. Kort, B. W. and D. P. Bertsekas. "Combined Primal-Dual and Penalty methods for convex programming" SIAM J. Control, 14, 2, p. 268-294, 1976.
36. Kuhn, H. W. and A. W. Tucker. "Nonlinear Programming" in Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman edited, University of California press, Berkeley, California, p. 481-493, 1951.
37. Lootsma, F. A. "Solving constrained minimization problems" in F. A. Lootsma edited Numerical Methods for Nonlinear Optimization, Academic Press, New York, 1971.
38. Lootsma, F. A. "Constrained optimization via parameter-free penalty functions" Philips. Res. Repts., 23, p. 424-437, 1968.
39. Lootsma, F. A. "Boundary properties of penalty functions for constrained minimization" Philips Res. Repts., Suppl. No. 3, 1970.
40. Luenberger, D. G. "Introduction to Linear & Nonlinear Programming" Addison-Wesley, 1973.
41. Mangasarian, O. L. and S. Fromovitz. "The Fritz John necessary optimality conditions in the presence of equality and inequality constraints" J. Math. Anal. & Appl., 17(1), p. 37-47, 1967.
42. Mangasarian, O. L. "Unconstrained Lagrangian in Nonlinear Programming" SIAM J. Control, 5, 13, p. 772-791, 1975.
43. McCormick, G. P. "Second-order conditions for constrained minima" SIAM J. Appl. Math., 15, p. 641-652, 1967.
44. Messerli, E. J. and E. Polak. "On second-order necessary conditions of optimality" SIAM J. Control, 7, p. 272-291, 1969.

45. Miele, A., E. E. Cragg, R. R. Iyer, and A. V. Levy. "Use of the augmented penalty function in Mathematical Programming problems, Part 1" *J. Opt. Theory & Appl.*, 8, p. 115-130, 1971.
46. Miele, A., E. E. Cragg and A. V. Levy. "Use of the augmented penalty function in Mathematical Programming problems; Part 2" *J. Opt. Theory & Appl.*, 8, p. 131-153, 1971.
47. Miele, A., P. E. Moseley & E. E. Cragg. "A modification of the method of multipliers for Mathematical Programming problems" in *Techniques of Optimization*, A. V. Balakrishnan edited, Academic Press, New York, 1972.
48. Miele, A., P. E. Moseley, A. V. Levy and G. M. Coggins. "On the method of multipliers for Mathematical Programming" *J. Opt. Theory and Appl.*, 10, p. 1-33, 1972.
49. Mordecai, A. "Nonlinear Programming; Analysis and Methods" Prentice-Hall, Inc., New Jersey, 1976.
50. Murray, W. "Ill conditioning in Barrier and Penalty functions arising in constrained nonlinear programming" *Princeton Symposium on Math. Prog.*, Princeton, N. J., 1967.
51. Noble, B. "Applied linear algebra" Prentice-Hall, Englewood Cliffs, N. J., 1969.
52. Pierre, D. A. "Multiplier algorithm for nonlinear programs" paper presented at eighth Int. Symp. on Math. Prog., Stanford University, Stanford, August, 1973.
53. Pietrzykowski, T. "Application of the steepest descent method to concave programming" in the *Proceeding of the IFIP Congress*, Munich, North-Holland Publishing Co., Amsterdam, p. 185-189, 1962.
54. Pietrzykowski, T. "An exact potential method for constrained maxima" *SIAM J. Numer. Anal.*, 6, p. 299-304, 1969.
55. Powell, M. J. D. "A method for nonlinear constraints in minimization problems" in *Optimization*, R. Fletcher edited, Academic Press, London, 1969.

56. Rockafellar, R. T. "The multiplier method of Hestenes and Powell applied to convex programming" *J. Opt. Theory & Appl.*, 12, p. 555-562, 1973.
57. Rockafeller, R. T. "A dual approach to solving nonlinear programming problems by unconstrained optimization" *Math. Prog.*, 5, p. 354-373, 1973.
58. Roode, J. D. "Generalized Lagrangian functions in Mathematical Programming" Thesis, University of Leiden, The Netherlands, 1968.
59. Rosenbrock, H. H. "An automatic method for finding the greatest or least value of a function" *Computer J.*, 3, p. 175-184, 1960.
60. Rubin, H. and P. Unger. "Motion under strong constraining force" *Commun. Pure Appl. Math.*, 10, p. 65-87, 1957.
61. Tremolieres, R. "La Methode des centres a Troncature variable" Thesis, Paris, 1968.
62. Varaiya, P. "Nonlinear Programming in Banach Space" *SIAM J. Appl. Math.*, 15, p. 284-293, 1967.
63. Wolfe, P. "Methods of nonlinear programming" in *Recent Advances in Math. Prog.*, R. L. Graves et al. edited, McGraw-Hill, New York, p. 67-86, 1963.
64. Zangwill, W. I. "Nonlinear programming by sequential unconstrained maximization" Working paper 131, Center for Research in Management Science, University of California, Berkeley, 1967.
65. Zangwill, W. I. "Nonlinear programming via Penalty functions" *Management Science*, 13(5), p. 344-358, 1967.

## Appendix A

An Unconstrained Minimization Algorithm

In this appendix, a description of the Fletcher-Reeves method [23] for implementing the Conjugate-gradient algorithm is given for the unconstrained problem of minimizing  $\{f(x) : x \in \mathbb{R}^n\}$ .

Step 1: Given a  $x^0$ , compute  $g^0 = \nabla f(x^0)$  and set  $d^0 = -g^0$ .

Step 2: For  $k = 0, 1, 2, \dots, n-1$ .

Step 2.1: If  $g^k = 0$ , stop. Otherwise continue.

Step 2.2: Set  $x^{k+1} = x^k + w^k d^k$ ; where  $w^k$  minimizes  $f(x^k + w d^k)$

Step 2.3: Compute  $g^{k+1} = \nabla f(x^{k+1})$ .

Step 2.4: Unless  $k = n-1$ , set  $d^{k+1} = -g^{k+1} + \alpha^k d^k$ , where

$$\alpha^k = \frac{\langle g^{k+1}, g^{k+1} \rangle}{\langle g^k, g^k \rangle}.$$

Step 3: Replace  $x^0$  by  $x^n$  and go back to Step 1.

## Appendix B

Fibonacci Search Method for Minimization  
of a Single-Variable Function

Let  $f$  be a function of a single variable  $x$  on the interval  $[a, b]$ . Assuming  $f$  is unimodal on the interval, the Fibonacci Search method determines, approximately, the minimum of  $f$  by measuring the values of  $f$  at a certain number of points. The method successively selects a number of measurement points so that, without explicit knowledge of  $f$ , one can determine the smallest possible region of uncertainty in which the minimum must lie. Thus, if

$l_1 = b_1 - a_1$ , the initial width of uncertainty, and

$l_k$  = width of uncertainty after  $k$  measurements.

Then, if a total of  $N$  measurements are to be made, we have

$$l_k = (F_{N-k+1} / F_N) l_1,$$

where the integers  $F_k$  are the members of the Fibonacci sequence generated by the recurrence relation

$$F_k = F_{k-1} + F_{k-2}, \quad F_0 = 1 = F_1.$$



The procedure for reducing the interval of uncertainty is the following; the first two measurements are made symmetrically at a distance of  $(F_{N-1}/F_N)l_1$  from the ends of the initial interval; depending on which of these measurements are smaller value, an uncertainty interval of width  $l_2 = (F_{N-1}/F_N)l_1$  is determined. The third measurement point is placed symmetrically in the new interval of uncertainty with respect to the measurement already in the interval. The result of the third measurement gives an interval of uncertainty  $l_3 = (F_{N-2}/F_N)l_1$ . In general, each successive measurement point is placed in the current interval of uncertainty symmetrically with the point already existing in that interval. After N measurement points the interval of uncertainty is reduced to a width of  $l_N = (F_0/F_N)l_1$ . Thus the initial interval of uncertainty could be reduced to a fraction, c, of itself; that is,  $(l_N/l_1) = c$ , provided the number of measurement points, N, is so chosen such that  $(F_0/F_N) < c$ .