Nonlinear programming : augmented Lagrangian techniques for constrained minimization
by Michael James Lowe

A thesis submitted to the Graduate Faculty in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY in Electrical Engineering
Montana State University

Abstract:
The nonlinear programming problem (NLP) is presented in its standard form for a constrained minimization problem. The general method of sequential-unconstrained minimization is then given, showing how the constrained NLP problem functions may be transformed into a single unconstrained function having the same solutions as the original problem. A unique unconstrained function, incorporating both equality and inequality constraints, is formulated by augmenting the Lagrangian with weighted problem functions of the constrained NLP. First-order necessary and second-order sufficient conditions are then established for this function to possess local minima that minimize the objective function and satisfy the constraints of the NLP, thus showing the equivalence of constrained minima and unconstrained minima of the augmented Lagrangian. A multiplier algorithm based on satisfying the Kuhn-Tucker first-order conditions of optimality is presented. The algorithm consists of alternating minimization and update phases. This procedure establishes a way to select an active inequality constraint basis, such that during a minimization phase these inequality constraints exhibit a 'locked on' feature and are treated the same as equality constraints. The inequality constraints that are not locked on during a particular minimization phase are held from excessive constraint violation with the penalty terms of the augmented Lagrangian. Although the algorithm is based on first-order conditions, a theorem is presented showing that if second-order conditions are satisfied by the NLP problem functions, the algorithm converges locally for finite penalty weights. As the sequence of minimizers converge to a local minimum, the Lagrange multipliers also converge to their optimal values. These optimal multipliers provide well-known sensitivity information, and a practical interpretation of these multipliers is presented. A F0RTRAN coded computer program ALG0R that implements the multiplier algorithm is described. It generates conjugate gradient search directions by the DFP method. An efficient quadratic-convergent unidirectional search technique is presented for conducting the search. ALG0R was applied to several test functions, and the results were compared to results obtained by other well-known NLP solution algorithms. When based on the number of function calls required to obtain a desired degree of accuracy in the objective function, ALG0R was shown to be a more efficient code than the multiplier program MULTMETH, and the penalty-function codes OPTKOV, POWCON, and SUMT. The formulation of several real-world problems into the NLP format is presented, along with their solutions and interpretations of the sensitivity information possessed by the Lagrange multipliers at the solution points.

NONLINEAR PROGRAMMING: AUGMENTED LAGRANGIAN TECHNIQUES FOR CONSTRAINED MINIMIZATION

by

MICHAEL JAMES LOWE

A thesis submitted to the Graduate Faculty in partial fulfillment of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

Approved:

Paul E. Uhlrich
Head, Major Department

Donald A. Pierre
Chairman, Examining Committee

Henry L Parsons
Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

March, 1974

ACKNOWLEDGMENTS

I am indebted to Dr. Donald A. Pierre who served as my advisor, and gave considerable time and effort in the development of this work. Many of his valuable insights and suggestions have been incorporated into this thesis.

I give special thanks to Dr. Roy M. Johnson and Dr. Donald A. Rudberg for their constructive criticism and useful suggestions during all phases of the research. I acknowledge Dr. Paul E. Uhlrich and the Department of Electrical Engineering for providing the financial support and the use of the computer facilities at Montana State University. I express my sincere appreciation to Mrs. Jean Julian who typed the final draft of this thesis.

Finally, I am thankful for the understanding and encouragement of my wife, Patty, who diligently typed preliminary drafts of this work, and my daughter, Krista, who cheerfully endured and helped when she could. To my family I dedicate this thesis.

Michael James Lowe

March, 1974

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The nonlinear programming problem (NLP) is presented in its standard form for a constrained minimization problem. The general method of sequential-unconstrained minimization is then given, showing how the constrained NLP problem functions may be transformed into a single unconstrained function having the same solutions as the original problem. A unique unconstrained function, incorporating both equality and inequality constraints, is formulated by augmenting the Lagrangian with weighted problem functions of the constrained NLP. First-order necessary and second-order sufficient conditions are then established for this function to possess local minima that minimize the objective function and satisfy the constraints of the NLP, thus showing the equivalence of constrained minima and unconstrained minima of the augmented Lagrangian. A multiplier algorithm based on satisfying the Kuhn-Tucker first-order conditions of optimality is presented. The algorithm consists of alternating minimization and update phases. This procedure establishes a way to select an active inequality constraint basis, such that during a minimization phase these inequality constraints exhibit a 'locked on' feature and are treated the same as equality constraints. The inequality constraints that are not locked on during a particular minimization phase are held from excessive constraint violation with the penalty terms of the augmented Lagrangian. Although the algorithm is based on first-order conditions, a theorem is presented showing that if second-order conditions are satisfied by the NLP problem functions, the algorithm converges locally for finite penalty weights. As the sequence of minimizers converge to a local minimum, the Lagrange multipliers also converge to their optimal values. These optimal multipliers provide well-known sensitivity information, and a practical interpretation of these multipliers is presented. A FØRTRAN coded computer program ALGØR that implements the multiplier algorithm is described. It generates conjugate gradient search directions by the DFP method. An efficient quadratic-convergent unidirectional search technique is presented for conducting the search. ALGØR was applied to several test functions, and the results were compared to results obtained by other well-known NLP solution algorithms. When based on the number of

function calls required to obtain a desired degree of accuracy in the objective function, ALGØR was shown to be a more efficient code than the multiplier program MULTMETH, and the penalty-function codes OPTKOV, POWCON, and SUMT. The formulation of several real-world problems into the NLP format is presented, along with their solutions and interpretations of the sensitivity information possessed by the Lagrange multipliers at the solution points.

CHAPTER ONE

INTRODUCTION TO THE NLP AND SEQUENTIAL UNCONSTRAINED MINIMIZATION

1.1 Introduction

The solution techniques of the nonlinear programming problem (NLP) are a powerful tool in mathematical programming and real-world problem solving and decision making. The NLP arises from many apparently unrelated areas and takes on many forms as found in mathematics, the natural and physical sciences, engineering, economics, agriculture, business, and government.

In the abstract form the NLP is that of optimizing (minimizing or maximizing) some entity while satisfying side conditions or constraints. The entity to be optimized may be profit, cost, production, efficiency, consumer utility, social tension, social benefit, energy, inventory, etc., to find an optimal solution or a state of optimal operation. Constraints are often imposed that may limit the possible steps or actions that may be taken to achieve this optimum. The constraints may involve manpower, availability of space, raw materials or funds, machine capabilities, limitations imposed by time, governmental controls, behavior patterns, physical laws, etc.

The NLP presents a solid framework for problem formulation. When an NLP is explicitly stated, a solution can generally be obtained. Many situations arise, however, where only numerical data or qualitative information is available and the problem functions cannot be determined explicitly, in which case the precise calculation of optimal points is impossible.

Of course the sagacious decision maker does not blindly accept the optimal point for the NLP as the best answer for his real-world problem. The optimal point is considered only as advisory, and the assumptions and the accuracy of the NLP are scrutinized before a decision is made. In many instances, however, NLP solutions have proven themselves in practice, in which cases the optimal point is accepted and implemented almost without question.

Although the transformation of a real-world problem into an explicit NLP is largely an art, there exists a theoretical framework that facilitates problem formulation. The theory clarifies the nature of those problem formulations which are most efficiently solved, and those which have no solution. This thesis is devoted to the development, presentation, and application of this theory.

1.2 The NLP in Standard Form

In a mathematical context we are given a function f(x) which maps the variables $x \in E^n$ into scalar values, where E^n denotes n-dimensional Euclidean space. The function f(x) is called an objective function, cost function, or performance measure because f(x) is that entity to be extremized, that is, to be minimized or maximized with respect to x. Since maximization and minimization are equivalent problems (max f = min (-f)), only the latter is considered here. The point x, however, may not be chosen arbitrarily because of constraints imposed upon it. The constraints are classified as equality or inequality constraints and are defined in terms of x. Any x which satisfies all of the constraints is termed a feasible point and becomes a candidate for the optimal point. For a solvable problem there exists a non-empty set of feasible points. From this set of feasible points, a point x*, at which f achieves its minimum value, is called an optimal point.

Stated mathematically, the NLP takes the following form: Find a point x* to

$$\text{minimize } f(x) \qquad (1\text{-}1)$$

subject to the constraints

$$h_i(x) = 0, \qquad i = 1, 2, \ldots, m_1 \qquad (1\text{-}2)$$

and

$$g_j(x) \geq 0, \qquad j = 1, 2, \ldots, m_2 \qquad (1\text{-}3)$$

where x is an n-dimensional vector. The h_i's, g_j's and f are nonlinear in general and are given (explicit) functions of class C^1. Also it is assumed that the n-dimensional gradient vectors $\nabla h_i \triangleq \partial h_i/\partial x$ and $\nabla g_j \triangleq \partial g_j/\partial x$ are nonzero at those x values which satisfy $h_i(x) = 0$ and $g_j(x) = 0$.

The problem as stated above seeks the constrained absolute minimum of f. In general practice, however, constrained local minima of f are found first; the absolute minimum of f is then found by comparing all constrained local minima.

1.3 Sequential Unconstrained Minimization

Basic to all NLP solution algorithms is a search in n-space for constrained extrema. Many constrained search techniques depend heavily upon the classical theory and methods for computing unconstrained extrema.

Unconstrained search techniques often incorporate a sequence of unidirectional searches in n-space, and are based on an iterative scheme of the form[†]

$$x^{k+1} = x^k + \sigma r^k \tag{1-4}$$

where a parametric step σ is taken in a search direction r^k. The step σ is selected to obtain the minimum value of f in the r^k direction. The search direction r^k is usually chosen on the basis of information concerning f and its partial derivatives evaluated both at the point x^k and at previous iteration points. The influence of constraints imposed on x must be taken into consideration as they alter the way in which the constrained search direction is chosen. This may be a difficult task if the constraints are nonlinear. According to Davies and Swann [16][‡], there are two ways in which constraint information is incorporated in the establishment of search directions. They are as follows:

[†] The superscript notation is used generally to denote a change in value of a vector or parameter from one iteration to another. The subscript notation will be used to denote components of a vector or matrix.

[‡] References are associated with numbers in brackets and are listed alphabetically at the end of this thesis.

1. Function modification followed by unconstrained optimization.

These methods seek to define a new function that has an unconstrained optimum at the same point as the optimum of the given constrained problem. Optimization of this new function will then define the required change in the search direction.

2. Direction modification without altering the function.

Some of the methods in this category attempt to follow constraint boundaries while others try to rebound from them, so as to continue the search in the feasible region.

There are advantages and disadvantages to both types of methods which are explored in detail by McCormick [43], by Wolf [70], and in [16] where it is stated that no one method for constrained optimization is universally superior to all others and care must be exercised in choosing the appropriate one for a given situation.

The solution techniques developed in this work fall into the first category. The approach in these methods is that of transforming a given constrained minimization problem into a sequence of unconstrained minimization problems as outlined by Fiacco and McCormick [23]. This transformation is accomplished by augmenting the original cost function with weighted terms of the problem functions, to define a new objective function whose minima are unconstrained in some domain of interest. By gradually removing the effect of the constraints in the new objective function by controlled parameter values, it is possible to generate a sequence of unconstrained problems that have solutions converging to a solution of the original constrained problem.

Thus, the sequential unconstrained method of minimizing the NLP is one associated with finding a sequence of problem solutions x^m such that

$$x^m = \min_{x \in R} A(f, h, g, w^m), \qquad m = 1, 2, \ldots \qquad (1\text{-}5)$$

where A is the new objective function formed by augmenting the cost function f with weighted terms of problem functions, h and g, w^m is the controlled weighting factor, and R is the domain of x. The functions of (1-5)† are constructed so that as $m \to \infty$, a convergent sequence of

---

†Hyphenated numbers in parenthesis refer to equations by (Chapter - equation number).

unconstrained minimizers $\{x^m\}$ approaches a constrained minimum point x^* of the NLP.

As an example of this type of method, consider an augmented function $A(x,w^m)$ which incorporates the equality constraints by Courant's quadratic penalty function [13], and the inequality constraints by Carroll's inverse penalty function [11].

Define

$$A(x,\ w^m) = f(x) + w^m \sum_{i=1}^{m_1} h_i^{\,2}(x) + \frac{1}{w^m} \sum_{j=1}^{m_2} \frac{1}{g_j(x)} \qquad (1\text{-}6)$$

To use this function we require a starting point x^0 which is in the strictly feasible region R^s associated only with the inequality constraints, i.e.,

$$x^0 \in R^s \triangleq \{x |\ g_j(x) > 0, \qquad j = 1, 2, \ldots, m_2\} \qquad (1\text{-}7)$$

Because the function $A(x,w^m)$ is to be minimized, and because $A(x,w^m)$ assumes the value of $+\infty$ at the boundary of R^s, it is guaranteed that no boundary of infeasible points will be generated with respect to the inequality constraints (1-3), i.e., all points generated will remain in R^s.

The method generally proceeds as follows. Select a monotonic sequence $\{w^m\}$, where

$$\{w^m\} = \{w^m \mid w^m > 0,\ w^{m+1} \geq w^m, \text{ and } w^m \to \infty \text{ as } m \to \infty\} \qquad (1\text{-}8)$$

and compute a minimum point x^m, where

$$x^m = \min_{x \in R^s} A(x, w^m) \qquad (1\text{-}9)$$

for $m = 1, 2, \ldots$. The desirable result is that

$$\lim_{m \to \infty} x^m = x^* \qquad (1\text{-}10)$$

where x^* is a solution to the NLP. It is also well known that

$$\lim_{m \to \infty} w^m \sum_{i=1}^{m_1} h_i^2(x^m) = 0 \qquad (1\text{-}11)$$

and that

$$\lim_{m \to \infty} \frac{1}{w^m} \sum_{j=1}^{m_2} \frac{1}{g_j(x^m)} = 0 \qquad (1\text{-}12)$$

Thus

$$\lim_{m \to \infty} A(x^m, w^m) - f(x^*) = 0 \qquad (1\text{-}13)$$

In effect, the influence of the constraints on the augmented function is relaxed and finally removed in the limit, and the augmented function converges to the same minimal value as the original cost function.

The function A(x, w) of (1-6) is called a mixed interior-exterior-point penalty function. If there were no equality constraints of the form (1-2), then (1-6) would have the form

$$f(x) + \frac{1}{w} \sum_{j=1}^{m_2} \frac{1}{g_j(x)} \tag{1-14}$$

and would be called an interior-point penalty function, allowing only feasible points with respect to the inequality constraints. On the other hand, if there were no inequality constraints of the form (1-3), then (1-6) would have the form

$$f(x) + w \sum_{i=1}^{m_1} h_i^2(x) \tag{1-15}$$

and would be called an exterior-point penalty function. Although it is possible to handle inequality constraints with an exterior penalty function method, as is done in this work, equality constraints can be incorporated only through exterior-point methods. These methods allow points not feasible to the constraints, but force convergence to a feasible point in the limit as $m \to \infty$.

The development and use of sequential unconstrained methods has a long history. The mainstream of developments are reviewed in [23] and in references cited there. Important results of many of the previous and more recent methods can also be found in [1, 8, 20, 21, 22, 23, 24, 39, 42, 43, 53, 70, 72].

1.4 Motivation for Research

The usual penalty methods for solving nonlinear programming problems are subject to numerical instabilities, because the derivatives of the penalty functions or the penalty functions themselves increase without bound near the solution as computation proceeds, and are held in check by penalty weighting factors which, to insure convergence, must either increase to infinity or decrease to zero [60]. It is noted [34] that penalty function methods can become computationally ineffective if certain constraints tend to dominate the entire constraint set. Several authors [38, 41, 51, 52] confirm this fact by showing that ill-conditioning can occur in the Hessian matrix in the final phases of a penalty function search, although a method to avoid this situation is given in [41]. When problems do arise in a penalty function search, they invariably occur during the final phase of the search [54]. It is in this phase, however, that relationships between Lagrange multipliers and certain penalty terms become viable, as shown in [7, 23, 54]. In past years the idea has arisen that terminal search instabilities might be circumvented by an approach involving a Lagrangian function containing additional penalty-like terms. Most of the work in this direction has been for problems treating equality constraints. Only recently has this type of method been extended to account for inequality constraints. The convergence proofs for these new methods often pertain to a class of functions which exhibit a certain property, usually convexity, and few algorithms are posed or operational that treat mixed equality and inequality constraints of the general NLP.

Because no penalty function is found to be superior for all types of problems, this author feels that the use of Lagrangian functions and the Kuhn-Tucker theorem [23] affords promising methods of solution to be explored. Besides being viewed directly as special cases of augmented objective functions, Lagrangians are associated with every method for mathematical programming. There are extremely close connections between the Lagrangian of an NLP and the particular objective function being utilized to transform that particular problem into a sequence of unconstrained minimizations, in that the conditions of the Kuhn-Tucker theorem are directly involved in the associated Lagrangian.

As a practical aspect, Lagrange multipliers provide well-known sensitivity information [1, 72] concerning NLP solutions. While it is true that the results of a penalty function search can be used to estimate Lagrange multipliers, the estimates can be inaccurate because of the problems encountered in the final phases of the search. This author feels, therefore, that it is important that solution techniques for the NLP also solve the dual problem of constrained minimization--that of finding the associated Lagrange multipliers.

The conditions stated in this section have motivated the author to pursue the research presented in this thesis. In Chapter 2 pertinent background material is reviewed, and the theoretical structure is developed to place the multiplier algorithm of this thesis in perspective.

Sensitivity information is presented in Chapter 3, and a practical interpretation of Lagrange multipliers is given. In performing a search in n-space for a solution point to an NLP problem, consideration must be given to generating a search direction and performing a search in that direction for a minimum. Chapters 4 and 5 treat these topics in light of the Lagrangian presented here. Chapter 5 reviews the basic theory of unidirectional searches and presents a quadratic convergent search developed in this research. Chapters 1 through 5 provide the necessary development for the computer implemented algorithms presented in Chapter 6. Numerical results from a number of test problems and selected application problems are given in Chapters 7 and 8. The significant results of this research effort are reviewed in Chapter 9, along with suggestions for further research that are in line with current trends in mathematical programming.

CHAPTER TWO

CONSTRAINED MINIMIZATION BY MULTIPLIER METHODS

2.1 Introduction

In a particular situation one is usually interested in the global solutions to a mathematical programming problem. But in general, only theorems concerning local solutions can be readily proved, unless the problem has certain properties that can be exploited. One such property is that of convexity; in a convex programming problem every local solution can be proved to be a global solution. For most NLP's, however, minimization algorithms converge to various local minima depending upon initial starting conditions, and global minima are obtained by comparing the known local minima. A local minimum point is a point for which, in an open neighborhood about that point, there is no other point that both satisfies the constraints and gives a smaller value of the objective function. Points of global minima are those points of local minima which yield the smallest value of the objective function. Useful theorems can be developed stating necessary conditions and sufficient conditions that a point be a local minimum point to the general NLP, without relying upon special properties of a particular problem. This chapter is devoted to the development of such conditions with regard to an augmented Lagrangian function that is used in solving the NLP.

2.2 Review of Multiplier Techniques

Although Courant [13] formalized the use of penalty functions for nonlinear programming in 1943, it was the works of Fritz John [33] in 1948 and the fundamental theoretical paper [37] of Kuhn and Tucker in 1951 that placed nonlinear programming on a substantial theoretical foundation. The results of Kuhn and Tucker on necessary conditions and sufficient conditions characterized the solution of the nonlinear convex programming problem and gave an equivalence between this problem and the saddle point problem of the Lagrangian. It is known that the existence of a saddle point of the Lagrangian function is heavily dependent upon convexity properties of the underlying problem. In 1956 Arrow and Hurwicz [3] showed that convexity assumptions could be relaxed by using a modified Lagrangian approach. Differential gradient schemes by Arrow and Solow [4] in 1958 presented additional saddle point results in terms of a different modified Lagrangian function. King in 1966 [35], by relinquishing the convexity property and thus sacrificing the Kuhn-Tucker global properties, gave local necessary conditions for inequality constrained extreme values.

The method of multipliers for (equality constrained) nonlinear programming was independently introduced in 1968 by Hestenes [30, 31] and Powell [58]. They proposed a dual method of solution in which squares of the constraint functions are added to the Lagrangian. A series of unconstrained minimizations on the penalized Lagrangian is followed by a multiplier vector update according to a simple rule. Hestenes did not theoretically develop the idea beyond this stage, but Powell [58] showed that if second-order sufficiency conditions for optimality are satisfied, the algorithm should converge locally at a linear rate, without requiring the penalty factors to grow without bound. The main advantage of the algorithm lies in the latter property, since it provides a numerical stability that is not found in the usual penalty methods. Miele, et al. [45, 46, 48, 49, 50] have modified and improved the original methods of Hestenes and Powell and have obtained many computational results for the equality constraint case. Haarhoff and Buys [29] advanced a similar method in 1970. Their method was concerned mainly with the equality constraint case; their extension to accommodate inequality constraints was simply that of including slack-variable terms to transform inequality constraints into equality constraints. Fletcher [25, 26] advanced yet another technique related to that of Hestenes and Powell. Instead of updating the multiplier vector after a sequence of minimizations, the multiplier was adjusted continuously as the minimization on the penalized Lagrangian was performed. Suggestions were given as to how this method could be extended to include inequality constraints. Termination and convergence properties were also given in [25]. Kort and Bertsekas [36] treat both equality and inequality constraints in their multiplier methods for convex programming.

In 1970 Rockafellar [59] introduced an augmented Lagrangian for inequality constrained convex programming problems for which an unconstrained saddle point corresponded to a solution of the convex programming problem. This approach was further studied by Rockafellar in a series of papers [60, 61, 62, 63].

Arrow, Gould, and Howe [2] in 1971 studied Rockafellar's augmented Lagrangian, included in a general class of augmented Lagrangians, for nonconvex programming problems and established local saddle point properties for this class of problems. Related approaches to a different class of augmented Lagrangians has been treated by Mangasarian [44] as recently as 1973. The Lagrange multipliers of Rockafellar and Mangasarian are not constrained to be non-negative. Various Lagrangian functions cited above are listed in Appendix A.

In 1971 Pierre [54] suggested more widely applicable multiplier algorithms, accounting for both equality and inequality constraint functions. Pierre [55] introduced a particularly interesting augmented Lagrangian in August, 1973, with local convergence properties and numerical results from an operational algorithm. One of the purposes of this work is to study in detail the augmented Lagrangian of [55].

2.3 Theoretical Preliminaries

The theorems and proofs presented in this chapter for the general NLP require the objective function and the constraint functions to be continuous and to have a suitable degree of differentiability. Also the n-dimensional gradient vectors $\nabla h_i(x)$ and $\nabla g_j(x)$ are assumed to be nonzero at those points x, which satisfy $h_i(x) = 0$ and $g_j(x) = 0$.

With these requirements the following definitions are in order

Definition 1: [Continuity]

A numerical valued function f, defined on a set X in E^n, is said to be continuous at a point $x_0 \in X$ if, given any number $\varepsilon > 0$, there is a nonempty neighborhood N_{ε/x_0}, such that $|f(x) - f(x_0)| < \varepsilon$ for every point $x \in N_{\varepsilon/x_0} \cap X$. The function f is said to be continuous on X if it is continuous at each point of X.

Definition 2: [Class of differentiability]

A function f is said to be of class C^1 in an open region X in E^n if f is continuous in X and its first-order partial derivatives with respect to x, $\partial f/\partial x_i$, $i = 1, 2, \ldots, n$, are defined and continuous in X. It is said to be of class C^2 in X if it is of class C^1 and its second-order partial derivatives, $\partial^2 f/\partial x_i x_j$, $i,j = 1, 2, \ldots, n$, are defined and continuous in X.

Differentiability of higher order C^p, $p > 2$, can be defined similarly by use of Definition 2, but will not be required in this work. Well formulated multiplier algorithms make effective use of first-order necessary conditions of optimality [55]. These conditions will now be developed.

2.4 First-Order Necessary Conditions

From the classical theory of minima and maxima, a first-order necessary condition is given by the following theorem. (For proof, see Appendix B)

Theorem 1: [First-order necessary condition for an unconstrained local minimum]
A necessary condition that a function f of class C^1 have an unconstrained local minimum at a point x^* is that

$$\nabla f(x^*) = \left.\frac{\partial f}{\partial x}\right|_{x = x^*} = 0 \tag{2-1}$$

Consider again the general constrained problem restated here from Section 1.2. Find x^* and $f(x^*)$ where

$$f(x^*) = \min_{x} f(x) \tag{2-2}$$

subject to the constraints

$$h_i(x) = 0, \qquad i = 1, 2, \ldots, m_1 \tag{2-3}$$

and

$$g_j(x) \geq 0, \qquad j = 1, 2, \ldots, m_2 \qquad (2-4)$$

Let x^* be a point which satisfies (2-3) and (2-4), and let Δx denote an incremental change in x from x^*, i.e., $\Delta x = x - x^*$, where the magnitude $||\Delta x||$ is the Euclidean norm† $(\Delta x' \Delta x)^{\frac{1}{2}}$. We may then define a <u>constrained local minimum</u>.

<u>Definition 3:</u> [Constrained local minimum]

For continuous functions defined on a set X in E^n, f is said to have a constrained local minimum at x^* if there exists a real number $\varepsilon > 0$ such that

$$f(x^*) \leq f(x^* + \Delta x) \qquad (2-5)$$

for all

$$\{\Delta x | 0 < ||\Delta x|| < \varepsilon, \ h_i(x^* + \Delta x) = 0, \quad i = 1, 2, \ldots, m_1,$$
$$g_j(x^* + \Delta x) \geq 0, \quad j = 1, 2, \ldots, m_2\}$$

If (2-5) in the above definition can be replaced by

$$f(x^*) < f(x^* + \Delta x) \qquad (2-6)$$

then f is said to have a <u>strict constrained local minimum</u> at x^*.

---

†The symbol ´ is used to denote the transpose of a vector or matrix. Let C be an m x n matrix. Then C´ is the transpose of C and has dimension n x m where the $(i,j)^{th}$ entry of C´ is c_{ji}.

As a basis for understanding necessary conditions for constrained local minima, the following Lemma is of fundamental importance. (For proof, see Appendix B)

Lemma 1: [Farkas [23]. Theory of linear inequalities] Let $\{v_k\}$, $k = 0, 1, \ldots, q$, be a set of $n \times 1$ vectors. A necessary and sufficient condition that there exist nonnegative scalar values $\{s_k\}$, $k = 1, 2, \ldots, q$, such that

$$v_0 = \sum_{k=1}^{q} s_k v_k \tag{2-7}$$

is that for every vector z such that $z'v_k \geq 0$, $k = 1, 2, \ldots, q$, it follows that $z'v_0 \geq 0$.

The necessary conditions that apply to the NLP are associated with the Lagrangian L:

$$L(x, \xi, \alpha, \beta) = \xi f(x) - \sum_{i=1}^{m_1} \alpha_i h_i(x) - \sum_{j=1}^{m_2} \beta_j g_j(x) \tag{2-8}$$

where α and β are Lagrange multiplier vectors of dimension m_1 and m_2, respectively. The objective function f(x) has an associated multiplier $\xi \geq 0$. The conditions that result for $\xi \geq 0$ are those of Fritz John [1, 33] which are discussed and implemented by Pierre [56]. For the special case where $\xi = 1$, the resulting conditions are those of Kuhn and Tucker from [23] and are presented in the following development.

With $\xi = 1$, and using matrix notation, equation (2-8) is recast in the form

$$L \triangleq L(x, \alpha, \beta) = f - \alpha' h - \beta' g \tag{2-9}$$

From (2-9) the gradient of L with respect to x follows as

$$\nabla L = \nabla f - \nabla h' \alpha - \nabla g' \beta \tag{2-10}$$

Notation to be used in establishing the necessary conditions that must hold for x^* to be a local minimum is as follows: Let f^* denote $f(x^*)$, ∇f^* denote $\nabla f(x^*)$, etc., and let $\Delta x = x - x^*$ denote a sufficiently small change in x^*. Define the following sets in E^n.

$$S_1^* \triangleq \{\Delta x | \Delta x' \nabla g_j^* \geq 0, \quad j \in S_a; \ \Delta x' \nabla h_i^* = 0, \quad i = 1, 2, \ldots, m_1; \ \ \Delta x' \nabla f^* \geq 0\} \tag{2-11}$$

$$S_2^* \triangleq \{\Delta x | \Delta x' \nabla g_j^* \geq 0, \quad j \in S_a; \ \Delta x' \nabla h_i^* = 0, \quad i = 1, 2, \ldots, m_1; \ \ \Delta x' f^* < 0\} \tag{2-12}$$

and

$$S_3^* \triangleq \{\Delta x | \Delta x' \nabla g_j^* < 0, \text{ for at least one } j \in S_a;$$
$$\text{or } \Delta x' \nabla h_i^* \neq 0 \text{ for at least one } i\} \quad (2\text{-}13)$$

where x^* is assumed to satisfy (2-3) and (2-4) and where

$$S_a \triangleq \{j | g_j(x^*) = 0\} \quad (2\text{-}14)$$

Also to be used are the sets

$$S_b \triangleq \{j | g_j(x^*) > 0\} \quad (2\text{-}15)$$

$$C_a \triangleq \{j | \beta_j > 0\} \quad (2\text{-}16)$$

and

$$C_b \triangleq \{j | \beta_j = 0 \text{ and } g_j(x) \leq 0\} \quad (2\text{-}17)$$

Clearly, $S_1^* \cap S_2^* \cap S_3^*$ equals the null set Φ, and any Δx in E^n belongs to one of the three sets S_1^*, S_2^*, or S_3^*. The set S_1^* of (2-11) is of limited interest because any Δx in this set does not show obvious reductions in $f(x)$; i.e., $f(x) - f(x^*) \cong \Delta x' \nabla f^* \geq 0$. Also, the set S_3^* of (2-13) is of limited interest because any Δx in this set does not give an x point that satisfies all of the constraints. The important set is therefore S_2^* of (2-12). A necessary condition that x^* be a constrained local minimum point is that all Δx's contained in S_2^* result in x's which violate one or more of the constraints in (2-3) and (2-4). For the great majority of problems, $S_2{}^*$ contains no Δx's which violate constraints, in which case a necessary condition that x^* be a constrained local minimum point is that $S_2{}^* = \Phi$. Unfortunately, there is a restricted class of problems for which the preceding statement does not hold. This restricted class of problems includes the class of problems for which the first-order constraint qualification condition [23] is not satisfied; these special problems must be treated separately. The following theorem, which is proved in Appendix B is given for the $S_2{}^* = \Phi$ case.

<u>Theorem 2</u>: [Existence of Generalized Lagrange Multipliers]

If i) x^* satisfies constraints (2-3) and (2-4),

ii) the functions f, $\{h_i\}_{i=1}^{m_1}$, $\{g_j\}_{j=1}^{m_2}$ are of class C^1,

iii) $S_2{}^* = \Phi$

then there exist vectors α^*, β^* such that (x^*, α^*, β^*) satisfies the following conditions:

$$h_i(x) = 0 \qquad i = 1, 2, \ldots, m_1 \qquad (2\text{-}19a)$$

$$g_j(x) \geq 0, \qquad j = 1, 2, \ldots, m_2 \qquad (2\text{-}19b)$$

$$\beta_j g_j(x) = 0, \qquad j = 1, 2, \ldots, m_2 \qquad (2\text{-}19c)$$

$$\beta_j \geq 0, \qquad j = 1, 2, \ldots, m_2 \qquad (2\text{-}19d)$$

$$\nabla L = 0 \qquad (2\text{-}19e)$$

The relations (2-19) are the Kuhn-Tucker relations that constitute a necessary condition that x^* be an optimal solution to the NLP. Note that assumption iii implies that the hypothesis of Lemma 1 are satisfied by the set A at x^* where

$$A = \{\nabla f^*, \{\nabla h_i^*, i = 1, 2, \ldots, m_1\}, \{-\nabla h_i^*, i = 1, 2, \ldots, m_1\}, \{\nabla g_j^*, j \in S_a\}\} \qquad (2\text{-}20)$$

To ensure the existence of finite multipliers when applying Theorem 2, one must be able to determine if $S_2^* = \Phi$. A useful criterion that can be used to test $S_2^* = \Phi$ is given in Theorem 3, which requires the following definition.

Definition 4: [Linear independence and dependence]
The vectors $v_1, v_2, \ldots, v_q$ are said to be linearly independent if

$$c_1 v_1 + c_2 v_2 + \ldots + c_q v_q = 0 \qquad (2\text{-}21)$$

where $c_1, c_2, \ldots, c_q$ are constants, implies that $c_1 = c_2 = \ldots = c_q = 0$. Conversely, vectors $v_1, v_2, \ldots, v_q$ are said to be linearly dependent if and only if v_i can be expressed as a linear combination of v_j ($j = 1, 2, \ldots, q$; $j \neq i$).

Theorem 3: [Sufficient condition that $S_2^* = \Phi$]
A sufficient condition that $S_2^* = \Phi$ and also that the hypothesis of Lemma 1 is satisfied by the set A of (2-20) is that the gradients of the active constraints $\{\nabla h_i\}$, $i = 1, 2, \ldots, m_1$, and $\{\nabla g_j\}$, $j \in S_a$, are linearly independent at x^*. (For proof, see Appendix B)

Theorem 2 is a first-order characterization of local minima in that it involves first-order differentiability of the problem functions. It therefore does not take into account the curvature of the functions if they are nonlinear. The curvature is measured by second-order derivatives of the problem functions.

Computational advantages can be obtained by augmenting the Lagrangian with additional terms. Consider the augmented Lagrangian L_a [55]:

$$L_a \triangleq L + w_1 \sum_{i=1}^{m_1} h_i^2 + w_2 \sum_{j \in C_a} g_j^2 + w_3 \sum_{j \in C_b} g_j^2 \qquad (2\text{-}22)$$

where L is the Lagrangian of (2-9), and w_1, w_2, and w_3 are penalty weights where $0 < \{w_1, w_2, w_3\} \leq \infty$. With slight loss in generality, we could let $w_2 = w_3$ and replace the corresponding summations in (2-22) by $w_2 \sum_{j \in C} g_j^2$ where $C = C_a \bigcup C_b$. Note also that $C_a \bigcap C_b = \Phi$ and therefore the last summation in (2-22) is equivalent to $w_3 \sum_{j \notin C_a} \frac{1}{2} g_j (g_j - |g_j|)$ which is a term suggested by Pierre [54].

The gradient of L_a with respect to x is

$$\nabla L_a = \nabla f - \sum_{i=1}^{m_1} (\alpha_i - 2w_1 h_i) \nabla h_i - \sum_{j \in C_a} (\beta_j - 2w_2 g_j) \nabla g_j$$

$$- \sum_{j \in C_b} (-2w_3 g_j) \nabla g_j \qquad (2\text{-}23)$$

Consider the direction of the vectors that sum to form $-\nabla L_a$. The first term, $-\nabla f$ points in the direction of greatest incremental decrease in f [47]. A typical equality constraint vector $(\alpha_i - 2w_1h_i)\nabla h_i$ tends to force constraint satisfaction by pointing in the direction of decreasing $|h_i|$, i.e., if $h_i > 0$ and $h_i > \alpha_i/2w_1$, or if $h_i < 0$ and $h_i < \alpha_i/2w_1$. A stronger statement applies toward forcing constraint satisfaction for inequality constraint terms of the form $(\beta_j - 2w_2g_j)\nabla g_j$. Since $\beta_j > 0$ for these terms, the gradient vector points in the direction of decreasing $|g_j|$ either if $g_j < 0$ or if $g_j > \beta_j/2w_2 > 0$. Finally, the last vector terms in (2-23) always point away from infeasible areas of constraint satisfaction. Since these terms, associated with the set C_b, apply only for inequality constraints where $\beta_j = 0$ and where $g_j < 0$, these vectors point in the direction of increasing g_j's.

With the augmented Lagrangian defined in (2-22), there is a direct relationship between the constrained local minima of the original problem and the unconstrained local minima of L_a. This can be shown from the following theorem given by Pierre [56].

<u>Theorem 4</u>: [Equivalence between NLP solutions and unconstrained minima of L_a]

Given (x^*, α^*, β^*) that satisfy (2-19), and given w_1, w_2, and w_3 that are sufficiently large such that $\{0 < w_i \leq \infty\}_{i=1}^{3}$ then L_a assumes a local minimum with respect to x at x^* if and only if x^* is a constrained local minimum of f.

Proof: Let (x^*, α^*, β^*) satisfy (2-19) with $\{0 < w_i \leq \infty\}_{i=1}^{3}$. The inactive constraints at x^* are those associated with $S_b = \{j | g_j(x^*) > 0\}$. Let ε be a sufficiently small positive number such that $g_j(x^* + \Delta x) > 0$ for all $j \in S_b$ and for any Δx which satisfies $0 < ||\Delta x|| < \varepsilon$. The terms in S_b therefore will not affect the formulation of the gradient of L_a at x^*.

First, assume that $f(x^*)$ is a constrained local minimum of f, and therefore $S_2^* = \Phi$. Consider $\Delta x \in S_3^*$. Suppose the i^{th} equality constraint is violated by Δx. The term $w_1 \sum_{i=1}^{m_1} h_i^2(x^* + \Delta x)$ in (2-22) will force $L_a(x^* + \Delta x) > L_a(x^*)$ for some $w_1 \leq \infty$. Also, suppose Δx violates the j^{th} inequality constraint for some $j \in S_a$, where $\beta_j \geq 0$. Then one of the terms

$w_2 \sum_{j \in C_a} g_j^{\,2}(x^* + \Delta x)$ or $w_3 \sum_{j \in C_b} g_j^{\,2}(x^* + \Delta x)$ in (2-22) will force $L_a(x^* + \Delta x) > L_a(x^*)$ for sufficiently large w_2 or w_3 where $w_2, w_3 \leq \infty$.

Thus, we are left to consider only those nonzero $\Delta x \in S_1^*$. For any Δx in this set we have

$$L_a(x^* + \Delta x) = f(x^* + \Delta x) \geq f(x^*) = L_a(x^*) \quad (2\text{-}24)$$

from which

$$L_a(x^* + \Delta x) \geq L_a(x^*) \quad (2\text{-}25)$$

Thus, L_a assumes an unconstrained local minimum at x^* if f has a local constrained minimum at x^*.

Now assume that $L_a(x^*)$ is an unconstrained local minimum of L_a. For the constrained problem, only those Δx are allowed for which $x^* + \Delta x$ satisfies (2-3) and (2-4). For any such Δx, with $\|\Delta x\| < \varepsilon$, equation (2-22) reduces to

$$L_a(x^* + \Delta x) = f(x^* + \Delta x) - \sum_{j \in C_a} [\beta_j g_j(x^* + \Delta x) - w_2 g_j^{\,2}(x^* + \Delta x)] \quad (2\text{-}26)$$

Consider a typical term of the summation in (2-26)

$$\beta_j g_j(x^* + \Delta x) - w_2 g_j^2(x^* + \Delta x)$$

$$= [1 - (w_2/\beta_j)\ g_j\ (x^* + \Delta x)]\ \beta_j g_j(x^* + \Delta x) \quad (2\text{-}27)$$

for $\beta_j > 0$ and $w_2 < \infty$, and for sufficiently small $||\Delta x||$, it is clear that the right-hand side of (2-27) is non-negative. Thus,

$$L_a(x^* + \Delta x) \leq f(x^* + \Delta x) \quad (2\text{-}28)$$

By assumption $L_a(x^*)$ is an unconstrained local minimum of L_a, and thus,

$$f(x^*) = L_a(x^*) \leq L_a(x^* + \Delta x) \quad (2\text{-}29)$$

and from (2-28) and (2-29), therefore,

$$f(x^*) \leq f(x^* + \Delta x) \quad (2\text{-}30)$$

Thus, $f(x^*)$ is a constrained local minimum of f.

O

In the above proof if the inequalities in (2-25) and (2-30) can be replaced by strict inequalities, then in Theorem 4 it can be stated that L_a assumes a strict local minimum with respect to x at x^* if and only if x^* is a strict constrained local minimum of f.

2.5 Second-order Sufficient Conditions

Second-order sufficient conditions are developed to measure the curvature of the problem functions near a local minimum. The following developments require second-order differentiability of the problem functions. Before second-order sufficient conditions for constrained local minima are presented, the following definition and theorem are in order.

Definition 5: [Positive definitness]

A real symmetric n x n matrix M is said to be positive definite if for every nonzero vector $y \in E^n$, $y'My > 0$.

Theorem 5: [Sufficient conditions for unconstrained local minima]

Sufficient conditions that a function f of class C^2 have an unconstrained local minimum at x^* are that

$$\nabla f^* = 0 \tag{2-31}$$

and that, for all $\Delta x \neq 0$

$$\Delta x' \nabla^2 f^* \Delta x > 0 \tag{2-32}$$

If $\nabla^2 f$ is continuous at x^*, (2-32) is equivalent to the statement that

$\nabla^2 f^* = [\partial^2 f/\partial x_i \partial x_j]|_{x = x^*}$ is positive definite. (For proof, see Appendix B)

Sufficient conditions based on the Kuhn-Tucker conditions of Theorem 2 are given in the following theorem. (For a proof, see [23])

Theorem 6: [A set of sufficient conditions for a constrained local minimum]

If i) x^* satisfies (2-3) and (2-4),

ii) the functions f, $\{h_i\}_{i=1}^{m_1}$, $\{g_j\}_{j=1}^{m_2}$ are of class C^2,

iii) the gradients of the active constraints $\{\nabla h_i\}$, $i = 1, 2, \ldots, m_1$, and $\{\nabla g_j\}$, $j \in S_a$ are linearly independent at x^*,

then sufficient conditions that x^* be a constrained local minimum are that there exist vectors α^* and β^* such that (x^*, α^*, β^*) satisfies conditions (2-19) and that for every nonzero vector Δx satisfying $\Delta x' \nabla h_i^* = 0$, $i = 1, 2, \ldots, m_1$, and $\Delta x' \nabla g_j^* = 0$ for $j \in C_a$, and $\Delta x' \nabla g_j^* \geq 0$ for $j \in S_a - C_a$ at x^*, it follows that

$$\Delta x' \nabla^2 L \Delta x > 0 \qquad (2\text{-}33)$$