CONTENTS

PREFACE xiii

1 INTRODUCTION

PART I
ANALYSIS

2 CLASSICAL OPTIMIZATION—
UNCONSTRAINED AND EQUALITY CONSTRAINED
PROBLEMS

2.1 Unconstrained Extrema 9
2.2 Equality Constrained Extrema
and the Method of Lagrange 15
Exercises 24
References 25

3 OPTIMALITY CONDITIONS FOR CONSTRAINED
EXTREMA

3.1 First Order Necessary Conditions
for Inequality Constrained Extrema 28
3.2 Second Order Optimality Conditions 45

3.3 Saddlepoints of the Lagrangian 51
Exercises 56
References 60

4 CONVEX SETS AND FUNCTIONS

4.1 Convex Sets 63
4.2 Convex Functions 71
4.3 Differential Properties of Convex Functions 83
4.4 Extrema of Convex Functions 92
4.5 Optimality Conditions for Convex Programs 95
Exercises 100
References 103

5 DUALITY IN NONLINEAR CONVEX PROGRAMMING

5.1 Conjugate Functions 106
5.2 Dual Convex Programs 112
5.3 Optimality Conditions and Lagrange Multipliers 125
5.4 Duality and Optimality for Standard Convex Programs 131
Exercises 139
References 141

6 GENERALIZED CONVEXITY

6.1 Quasiconvex and Pseudoconvex Functions 145
6.2 Arcwise Connected Sets and Convex Transformable Functions 160
6.3 Local and Global Minima 172
Exercises 178
References 181

7 ANALYSIS OF SELECTED NONLINEAR PROGRAMMING PROBLEMS

7.1 Quadratic Programming 185
7.2 Stochastic Linear Programming with Separable Recourse Functions 189
7.3 Geometric Programming 196
Exercises 209
References 210

PART II

METHODS

8 ONE-DIMENSIONAL OPTIMIZATION

8.1	Newton's Method	216
8.2	Polynomial Approximation Methods	221
8.3	Direct Methods—Fibonacci and Golden Section Techniques	225
8.4	Optimal and Golden Block Search Methods	233
	Exercises	241
	References	242

9 MULTIDIMENSIONAL UNCONSTRAINED OPTIMIZATION WITHOUT DERIVATIVES: EMPIRICAL AND CONJUGATE DIRECTION METHODS

9.1	The Simplex Method	245
9.2	Pattern Search	247
9.3	The Rotating Directions Method	249
9.4	Conjugate Directions	255
9.5	Powell's Method	259
9.6	Avoiding Linearly Dependent Search Directions	265
9.7	Further Conjugate Direction-Type Algorithms	275
	Exercises	281
	References	285

10 SECOND DERIVATIVE, STEEPEST DESCENT AND CONJUGATE GRADIENT METHODS

10.1	Newton-Type and Steepest Descent Methods	288
10.2	Conjugate Gradient Methods	299
10.3	Convergence of Conjugate Gradient Algorithms	307
	Exercises	316
	References	318

11 VARIABLE METRIC ALGORITHMS

11.1	A Family of Variable Metric Algorithms	322
11.2	Quasi-Newton Methods	341

11.3 Variable Metric Algorithms without Derivatives 353
11.4 Recent Methods Based on Nonquadratic Functions 355
Exercises 364
References 367

12 PENALTY FUNCTION METHODS

12.1 Exterior Penalty Functions 372
12.2 Interior Penalty Functions 378
12.3 Parameter-Free Penalty Methods 385
12.4 Exact Penalty Functions 388
12.5 Multiplier and Lagrangian Methods 399
12.6 Some Computational Aspects of Penalty Function Methods 410
Exercises 412
References 415

13 SOLUTION OF CONSTRAINED PROBLEMS BY EXTENSIONS OF UNCONSTRAINED OPTIMIZATION TECHNIQUES

13.1 Extensions of Empirical Methods 420
13.2 Gradient Projection Algorithms for Linear Constraints 423
13.3 A Quadratic Programming Algorithm 437
13.4 Feasible Direction Methods 442
13.5 Projection and Feasible Direction Methods for Nonlinear Constraints 449
Exercises 454
References 457

14 APPROXIMATION-TYPE ALGORITHMS

14.1 Methods of Approximation Programming 461
14.2 Reduced-Gradient Algorithms 469
14.3 Cutting Plane Methods 477
14.4 Complementary Convex Programming 483
Exercises 494
References 496

AUTHOR INDEX 499

SUBJECT INDEX 504

1 INTRODUCTION

Mathematical programming is a branch of optimization theory in which a single-valued **objective function** f of n real variables $x_1, \ldots, x_n$ is minimized (or maximized), possibly subject to a finite number of **constraints**, which are written as inequalities or equations. Generally we can define a mathematical program of, say, minimization as

$$\text{(MP)} \qquad \min f(x) \tag{1.1}$$

subject to

$$g_i(x) \geq 0, \qquad i = 1, \ldots, m \tag{1.2}$$

$$h_j(x) = 0, \qquad j = 1, \ldots, p \tag{1.3}$$

where x denotes the column vector whose components are $x_1, \ldots, x_n$. In other words, (MP) is the problem of finding a vector x^* that satisfies (1.2) and (1.3) and such that $f(x)$ has a minimal—that is, optimal value. If one or more of the functions appearing in (MP) are nonlinear in x, we call it a **nonlinear program**, in contrast to a linear program, where all these functions must be linear. The study of some basic aspects of nonlinear programming is the subject of this book.

Nonlinear programming problems arise in such various disciplines as engineering, economics, business administration, physical sciences, and mathematics, or in any other area where decisions (in a broad sense) must be taken in some complex (or conflicting) situation that can be represented by a mathematical model. In order to illustrate some types of nonlinear programs, a few examples are presented below.

NONLINEAR CURVE FITTING

Suppose that in some scientific research, say in biology or physics, a certain phenomenon f is measured in the laboratory as a function of time. Also suppose that we are given a mathematical model of the phenomenon, and from the model we know that the value of f is assumed to vary with time t as

$$f(t) = x_1 + x_2 \exp(-x_3 t). \tag{1.4}$$

The purpose of the laboratory experiments is to find the unknown parameters x_1, x_2, and x_3 by measuring values of f at times $t^1, t^2, \ldots, t^M$. The decision-making process involves assigning values to the parameters, and it is reasonable to ask for those values of x_1, x_2, and x_3 that are optimal in some sense. For example, we can seek optimal values of the parameters in the least-squares sense—that is, those values for which the sum of squares of the experimental deviations from the theoretical curve is minimized. Formally, we have the nonlinear program

$$\min F(x_1, x_2, x_3) = \sum_{i=1}^{M} [f(t^i) - x_1 - x_2 \exp(-x_3 t^i)]^2. \tag{1.5}$$

Note that this is an unconstrained program that, if solved, may yield unacceptable values of the parameters. To avoid such a situation, we can impose restrictions in the form of constraints. For example, the parameter x_3 can be restricted to have a nonnegative value—that is,

$$x_3 \geq 0. \tag{1.6}$$

Also suppose that, for the particular phenomenon under consideration, the mathematical model proposed can be acceptable only if the parameters are so chosen that at $t = 0$ we have $f(0) = 1$. Hence we must add a constraint

$$x_1 + x_2 = 1. \tag{1.7}$$

Solving (1.5), subject to (1.6) and (1.7), is then a constrained nonlinear programming problem having a nonlinear objective function with linear inequality and equality constraints.

LOCATION PROBLEM

Suppose now that the location of a supply center to serve m customers having fixed spatial locations in a city must be selected. The commodity to

be supplied from the center may be electricity, water, milk, or some other goods. The criterion for selecting the location of the supply center is to minimize some distance function from the center to the customers. It may happen, for example, that we are interested in minimizing the **maximal** distance from the center to any particular customer. Since the supply of goods in this city must be along perpendicular lines (e.g., streets), the appropriate distance function is the so-called **rectangular** distance. Stating it as a mathematical model, let (x_1, x_2) denote the unknown location (coordinates) of the supply center and let (a^i, b^i) be the given location of customer i. Our problem is then

$$\min_{x_1, x_2} \{\max_{1 \le i \le m} [|a^i - x_1| + |b^i - x_2|]\}, \tag{1.8}$$

where the preceding formulation means that, first, for every possible value of (x_1, x_2) we must find that index i that maximizes the rectangular distance given between the square brackets, and, second, among all those maximal distances, depending on (x_1, x_2), we must find the smallest one. Again, if every location (x_1, x_2) is acceptable, then our problem is an unconstrained one. Sometimes, however, it is advantageous to simplify some expressions at the expense of adding extra variables and constraints. For example, define a new variable x_0 by

$$x_0 = \max_{1 \le i \le m} [|a^i - x_1| + |b^i - x_2|] \tag{1.9}$$

or

$$x_0 \ge |a^i - x_1| + |b^i - x_2|, \qquad i = 1, \ldots, m. \tag{1.10}$$

We obtain, consequently, a nonlinear program in three variables x_0, x_1, x_2:

$$\min f(x) = x_0 \tag{1.11}$$

subject to

$$g_i(x) = x_0 - |a^i - x_1| - |b^i - x_2| \ge 0, \qquad i = 1, \ldots, m. \tag{1.12}$$

The reader can easily verify that problems (1.8) and (1.11) to (1.12) are equivalent in the sense that (x_1^*, x_2^*) is an optimal solution of (1.8) if and only if (x_0^*, x_1^*, x_2^*) is an optimal solution of (1.11) and (1.12) with

$$x_0^* = |a^k - x_1^*| + |b^k - x_2^*| \tag{1.13}$$

for some k, $1 \le k \le m$. The reader can also show that, by introducing more variables, (1.8) can be transformed into a linear program.

PROCESS DESIGN

Consider the problem of manufacturing a given quantity F_B gram moles per hour of chemical product B from a feed consisting of an aqueous solution of reactant A, in a continuous stirred tank (backmix) reactor. The chemical reaction is

$$2A \rightarrow B \tag{1.14}$$

with an empirical rate equation, established in the laboratory and based on unit volume of reacting fluid,

$$-\frac{dC_A}{dt} = 8.4(C_A)^2 = 8.4[C_A^0(1 - x_A)]^2 \qquad \left(\frac{\text{g mole}}{\text{liter-hr}}\right), \tag{1.15}$$

where C_A = concentration of A in the reactor (g mole/liter)
C_A^0 = concentration of A in the feed (g mole/liter)
t = time (hours)
x_A = conversion, fraction of reactant converted into product.

Suppose that the feed solution is available at a continuous range of concentrations of A and that its unit cost p_A is given by the relation

$$p_A = 4(C_A^0)^{1.4} \qquad (\$/\text{liter}). \tag{1.16}$$

The operating cost of the continuous stirred tank reactor (CSTR) is given by

$$p_{CSTR} = 0.75(V)^{0.6} \qquad (\$/\text{hr}), \tag{1.17}$$

where V (liter) is the volume of the reactor. Assume that the product B can be sold at a price of 10 \$/g mole. Our problem is to determine the rate of feed solution F_A^0 (liter/hr), its concentration C_A^0, the volume of the reactor V, and the conversion x_A for optimum operation—that is, for maximizing total profit per hour—given by

$$p_T = 10F_B - p_A F_A^0 - p_{CSTR} \qquad (\$/\text{hr}). \tag{1.18}$$

Material balance around the reactor yields

$$F_A^0 C_A^0 = F_A^0 C_A^0(1 - x_A) - \left(\frac{dC_A}{dt}\right)V. \tag{1.19}$$

From (1.14) we get

$$\tfrac{1}{2}F_A^0 C_A^0 x_A = F_B \tag{1.20}$$

and from (1.15) and (1.19) we obtain

$$8.4C_A^0(1 - x_A)^2V - F_A^0x_A = 0. \tag{1.21}$$

Our design problem then becomes

$$\max p_T = 5F_A^0C_A^0x_A - 4(C_A^0)^{1.4}F_A^0 - 0.75(V)^{0.6} \tag{1.22}$$

subject to constraint (1.21). This is a nonlinear program in variables C_A^0, F_A^0, V, x_A in which both the objective and the constraint functions are nonlinear. Note that here the objective function is maximized.

Throughout the text we shall mainly be concerned with nonlinear programs in which the objective function is minimized. This does not represent any restriction, since every problem of the type $\max f(x)$ can be equivalently analyzed and solved by considering $\min \bar{f}(x)$, where $\bar{f}(x) = -f(x)$.

Finally, a few words on notation and terminology used in later chapters. No special symbols will be used to denote vectors. The dimension of a vector, if not specifically mentioned, should always be clear from the formula in which it appears. All vectors are assumed to be column vectors. Components of a vector will be denoted by subscripts; thus $x_1, x_2, \ldots, x_n$ are the components of the n vector x. Superscripts on vectors will be used in order to distinguish between different vectors; thus $x^1, x^2, \ldots, x^m$ are meant to be m different vectors. In order to avoid confusion, exponents on real numbers will be used together with parentheses; that is, $(a)^2$ is the square of the number a. The notation x^T will be used to indicate a row vector—that is, the transpose of a column vector. The real line, that is, the set of all real numbers, is denoted by R and the n-dimensional real Euclidean space by R^n. Vectors $x \in R^n$ are also frequently referred to as points in R^n. Specific points in R^n will frequently be written in terms of their coordinates—for example, $x^0 = (1, -2)$. The notation $x \geq 0$ means that every component of x is nonnegative; thus if $x \in R^n$, then 0 is also an n-dimensional vector, each component of which is zero. The notation $x \neq 0$ means that at least one component of x is different from zero.

Only matrices having real elements will be used. Similar to vectors, no special symbols will be used for matrices (although capital letters will generally be employed). The dimension of a matrix will also be uniquely determined from the formula in which it appears. If A is an $m \times n$ matrix (m rows and n columns), then A^T is the transpose of A, having n rows and m columns. The inverse of a matrix A will be denoted by A^{-1}—that is, $AA^{-1} = A^{-1}A = I$, where I is the identity (unit) matrix.

Norm of a vector $x \in R^n$ is defined by

$$||x|| = [(x_1)^2 + (x_2)^2 + \cdots + (x_n)^2]^{1/2}. \tag{1.23}$$

We shall frequently use the notion of neighborhoods. A set

$$N_\delta(x^0) = \{x : x \in R^n, \ ||x - x^0|| < \delta\}, \tag{1.24}$$

where δ is a positive number is called a (spherical) neighborhood of the point x^0.

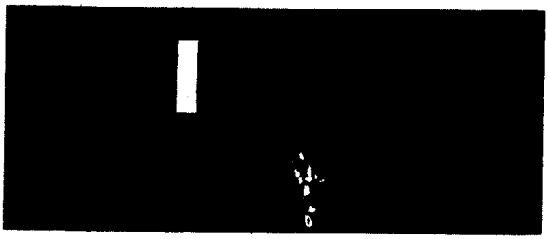
Matrix norms will be used in a few places in the text and then they will be the norms induced by vectors. Formally, if A is an $m \times n$ matrix and x is an n vector, then

$$||A|| = \sup \left\{ \frac{||Ax||}{||x||} : ||x|| \neq 0 \right\} = \sup \{||Ax|| : ||x|| = 1\}. \tag{1.25}$$

Functions will be always single valued; and as we shall see later, they can sometimes take on the values $+\infty$ or $-\infty$. In the few places where a more advanced mathematical concept is needed, we either define it in the text or, if such a definition would require extensive background material, we sacrifice completeness and refer the reader to appropriate references. Those readers unfamiliar with elementary linear algebra, real analysis, or topology may wish to consult introductory textbooks on these subjects, such as Apostol [1], Bartle [2], Hall and Spencer [3], and Noble [4].

REFERENCES

1. APOSTOL, T. M., *Mathematical Analysis*, 2nd ed., Addison-Wesley Publishing Co., Reading, Mass., 1974.
2. BARTLE, R. G., *The Elements of Real Analysis*, 2nd ed., John Wiley & Sons, New York, 1976.
3. HALL, D. W., and G. L. SPENCER, *Elementary Topology*, John Wiley & Sons, New York, 1955.
4. NOBLE, B., *Applied Linear Algebra*, Prentice-Hall, Englewood Cliffs, N.J., 1969.



The solution of a nonlinear programming problem, if not specified otherwise, consists of finding an optimal solution vector x^* and not all optimal solutions that may exist. Recognizing an optimal x^* and studying its properties form the central theme of the first part of this book, which deals with some analytic aspects of nonlinear programming problems. We shall see that if a vector x is a candidate for an optimal solution, it must satisfy certain necessary conditions of optimality. Unfortunately, however, there may be vectors other than the optimal ones that also satisfy these conditions. Consequently, necessary conditions are primarily useful in the negative sense: if a vector x does not satisfy them, it cannot be an optimal solution. To verify optimality, we may, therefore, look for sufficient conditions of optimality that, if satisfied together with the necessary ones, give a clear indication of the nature of a particular solution vector under consideration. These two types of optimality conditions constitute the first subject to be discussed in some detail. In particular, the classical method of Lagrange multipliers is extended to optimization problems with inequality constraints. In a general nonlinear program it may happen that vectors satisfying one type of optimality conditions do not satisfy the other one. Programs in which necessary optimality conditions are also sufficient are important, since a solution of a system of equations and inequalities representing, say, necessary conditions of optimality in a program under consideration is ensured as the sought optimum of that program.

A class of nonlinear programs, called convex programs and involving convex and concave functions in a certain configuration, possesses this nice property. Such convex programs are especially convenient to analyze. Associated with each such program there exist so-called dual programs, which, similar to linear programming, have some interesting theoretical properties.

Duality relations will be derived with the aid of a modern approach based on conjugate functions.

Convex programs belong to that class of nonlinear programs in which every local minimum is global. Since there are many nonconvex real-life problems, we shall also be concerned with the important question of finding more general nonconvex functions and programs for which this local-global optimum property holds. The analysis part of this text will conclude with a few selected nonlinear programs in which some of the theoretical results will be illustrated.

2 CLASSICAL OPTIMIZATION—UNCONSTRAINED AND EQUALITY CONSTRAINED PROBLEMS

The problem of finding extrema—that is, minima or maxima of real-valued functions—plays a central role in mathematical optimization. We begin the topic of extrema with the simplest case of unconstrained problems and then proceed to the subject of minima and maxima in the presence of constraints, expressed as equations. Here we shall treat the classical Lagrange multiplier theory and some necessary and sufficient conditions for extrema of differentiable functions. Treatment of these topics goes back a few centuries, hence the name "classical." In later chapters we shall discuss optimization problems in which the constraints are expressed as inequalities. All the remarkable results obtained for such problems can be classified as "modern" because they are a consequence of intensified interest in inequality constrained problems during the last two to three decades. All the "classical" results can be considered as special cases of the more general "modern" theory. We chose to present the classical results first because they can serve as a bridge between the material presented in most first- and second-year university courses of calculus or real analysis and the more advanced subject of mathematical programming. In addition, the classical theory is simpler than the modern theory in the sense that results, such as necessary and sufficient conditions for extrema, are not obscured by the more complicated requirements in the case of inequality constraints.

2.1 UNCONSTRAINED EXTREMA

Consider a real-valued function f with domain D in R^n. Then f is said to have a **local minimum** at a point $x^* \in D$ if there exists a real number $\delta > 0$ such that

$$f(x) \geq f(x^*) \tag{2.1}$$

for all $x \in D$ satisfying $\|x - x^*\| < \delta$. We define a **local maximum** in a similar way but with the sense of the inequality in (2.1) reversed. If the inequality (2.1) is replaced by a strict inequality

$$f(x) > f(x^*), \qquad x \in D, \quad x \neq x^* \tag{2.2}$$

we have a **strict** local minimum; and if the sense of the inequality in (2.2) is reversed, we have a strict local maximum. The function f has a **global minimum** (strict global minimum) at $x^* \in D$ if (2.1) [or (2.2)] holds for all $x \in D$. A similar definition holds for a **global maximum** (strict global maximum). An **extremum** is either a minimum or a maximum. Not every real function has an extremum; for example, a nonzero linear function has no extremum on R^n. It is clear from these definitions that every global minimum (maximum) of f in D is also a local minimum (maximum). The converse of this statement is, in general, false, and the reader can easily demonstrate it by examples. In later chapters we shall discuss functions, such as convex functions, that, however, have the remarkable property that every local minimum is also a global minimum.

Let $x \in D \subset R^n$ be a point where the real function f is differentiable. Recall that if a real-valued function f is differentiable at an interior point $x \in D$, then its first partial derivatives exist at x. If, in addition, the partial derivatives are continuous at x, then f is said to be continuously differentiable at x. Similarly, if f is twice differentiable at $x \in D$, then the second partial derivatives exist there. And if they are continuous at x, then f is said to be twice continuously differentiable at x. We define the **gradient** of f at x as the vector $\nabla f(x)$, given by

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \quad \ldots, \quad \frac{\partial f(x)}{\partial x_n}\right)^T. \tag{2.3}$$

Similarly, if f is twice differentiable at x, we define the **Hessian matrix** of f at x as the $n \times n$ symmetric matrix $\nabla^2 f(x)$, given by

$$\nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right], \qquad i, j = 1, \ldots, n. \tag{2.4}$$

In this section we discuss necessary and sufficient conditions for extrema of functions without constraints. We start by stating the following well-known result.

Theorem 2.1 (Necessary Condition)

Let x^* be an interior point of D at which f has a local minimum or local maximum. If f is differentiable at x^*, then

$$\nabla f(x^*) = 0. \tag{2.5}$$

This theorem will be restated and proved as part of Theorem 2.3.

Now we turn to sufficient conditions for a local extremum.

Theorem 2.2 (Sufficient Conditions)

Let x^* be an interior point of D at which f is twice continuously differentiable. If

$$\nabla f(x^*) = 0 \tag{2.6}$$

and

$$z^T \nabla^2 f(x^*) z > 0 \tag{2.7}$$

for all nonzero vectors z, then f has a local minimum at x^*. If the sense of the inequality in (2.7) is reversed, then f has a local maximum at x^*. Moreover, the extrema are strict local extrema.

This theorem can be proved by using the Taylor expansion of f and is left for the reader.

In both theorems we are utilizing the behavior of the function at x^*, the extremum. If, however, we can investigate the behavior of the function in some neighborhood of the extremum in question, we have a result that provides additional conditions for a local extremum.

Theorem 2.3

Let x^* be an interior point of D and assume that f is twice continuously differentiable on D. It is necessary for a local minimum of f at x^* that

$$\nabla f(x^*) = 0 \tag{2.8}$$

and

$$z^T \nabla^2 f(x^*) z \geq 0 \tag{2.9}$$

for all z. Sufficient conditions for a local minimum are that (2.8) holds and that for every x in some neighborhood $N_\delta(x^*)$ and for every $z \in R^n$, we have

$$z^T \nabla^2 f(x) z \geq 0. \tag{2.10}$$

If the sense of the inequalities in (2.9) and (2.10) is reversed, the theorem applies to a local maximum.

Proof. Suppose that f has a local minimum at x^*. Then

$$f(x) \geq f(x^*) \tag{2.11}$$

for all x in some neighborhood $N_\delta(x^*) \subset D$.

We can write every $x \in N_\delta(x^*)$ as $x = x^* + \theta y$, where θ is a real number and y is a vector such that $\|y\| = 1$. Hence

$$f(x^* + \theta y) \geq f(x^*) \tag{2.12}$$

for sufficiently small $|\theta|$.

For such a y, we define F by $F(\theta) = f(x^* + \theta y)$. Then (2.12) becomes

$$F(\theta) \geq F(0) \tag{2.13}$$

for all θ such that $|\theta| < \delta$.

From the Mean Value Theorem [1], we have

$$F(\theta) = F(0) + \nabla F(\lambda\theta)\theta, \tag{2.14}$$

where λ is a number between 0 and 1.

If $\nabla F(0) > 0$, then, by the continuity assumptions, there exists an $\epsilon > 0$ such that

$$\nabla F(\lambda\theta) > 0 \tag{2.15}$$

for all λ between 0 and 1 and for all θ satisfying $|\theta| < \epsilon$. Hence we can find a $\theta < 0$ such that $|\theta| < \delta$ and

$$F(0) > F(\theta), \tag{2.16}$$

a contradiction. Assuming that $\nabla F(0) < 0$ would lead to a similar contradiction. Thus

$$\nabla F(0) = y^T \nabla f(x^*) = 0. \tag{2.17}$$

But y is an arbitrary nonzero vector. Hence we must have

$$\nabla f(x^*) = 0. \tag{2.18}$$

Turning now to the second-order conditions, we have, by Taylor's theorem,

$$F(\theta) = F(0) + \nabla F(0)\theta + \tfrac{1}{2}\nabla^2 F(\lambda\theta)(\theta)^2, \qquad 1 > \lambda > 0. \tag{2.19}$$

If $\nabla^2 F(0) < 0$, then, by continuity, there exists an $\epsilon' > 0$ such that

$$\nabla^2 F(\lambda\theta) < 0 \tag{2.20}$$

for all λ between 0 and 1 and for all θ satisfying $|\theta| < \epsilon'$.

Since $\nabla F(0) = 0$, (2.20) would imply for such a θ that

$$F(\theta) < F(0), \tag{2.21}$$

a contradiction. Consequently,

$$\nabla^2 F(0) = y^T \nabla^2 f(x^*) y \geq 0. \tag{2.22}$$

Since this inequality holds for all y, subject to the arbitrary restriction on the norm of y, it must hold for all vectors z. This completes the proof of the first part of the theorem.

For the second part, suppose that (2.8) and (2.10) hold but that x^* is not a local minimum. Then there exists a $w \in N_\delta(x^*)$ such that $f(x^*) > f(w)$. Let $w = x^* + \theta y$, where $\| y \| = 1$ and $\theta > 0$. By Taylor's theorem,

$$f(w) = f(x^*) + \theta y^T \nabla f(x^*) + \tfrac{1}{2}(\theta)^2 y^T \nabla^2 f(x^* + \lambda\theta y) y, \tag{2.23}$$

where $1 > \lambda > 0$. Our assumptions lead then to

$$y^T \nabla^2 f(x^* + \lambda\theta y) y < 0, \tag{2.24}$$

contradicting (2.10), since $x^* + \lambda\theta y \in N_\delta(x^*)$. The proof for a local maximum is similar. ■

Theorem 2.2 provides sufficient conditions for a strict local extremum of f at x^*, based on the behavior of the function at that point. We shall show that it is easy to find examples of extrema for which these sufficient conditions are not satisfied. In Theorem 2.3 we have sufficient conditions for a local (not necessarily strict) extremum based on the behavior of f in a neighborhood of x^*. Finally, we present sufficient conditions for a strict local extremum, also based on a neighborhood of x^*.

Theorem 2.4

Let x^* be an interior point of D and assume that f is twice continuously differentiable. If

$$\nabla f(x^*) = 0 \tag{2.25}$$

and

$$z^T \nabla^2 f(x) z > 0 \tag{2.26}$$

for any $x \neq x^*$ in a neighborhood of x^* and for any nonzero z, then f has a strict local minimum at x^*. Reversing the sense of the inequality in (2.26) results in sufficient conditions for a strict local maximum.