C · L · A · S · S · I · C · S
IN APPLIED MATHEMATICS

4

Nonlinear Programming

Sequential Unconstrained Minimization Techniques

Anthony V. Fiacco
Garth P. McCormick

Nonlinear Programming

Sequential Unconstrained Minimization Techniques

Anthony V. Fiacco
George Washington University

Garth P. McCormick
George Washington University

siam

Society for Industrial and Applied Mathematics
Philadelphia

Library of Congress Cataloging-in-Publication Data
Fiacco, Anthony V.
Nonlinear programming : sequential unconstrained minimization techniques / Anthony V. Fiacco, Garth P. McCormick.
p. cm. - - (Classics in applied mathematics)
"This SIAM edition is an unabridged, corrected republication of the work first published by Research Analysis Corporation, McLean, Virginia" - - T.p. verso
Includes bibliographical references (p.).
ISBN 0-89871-254-8
1. Nonlinear programming. 2. Mathematical optimization.
3. Algorithms. I. McCormick, Garth P. II. Title. III. Series.
T57.8.F54 1990
519.7'6- -dc20 90-9658
CIP

Preparation of this book was undertaken in the course of research sponsored by contract DA-44-188-ARO-1 between the U.S. Army Research Office and the Research Analysis Corporation, Library of Congress Catalog Card Number: 68-30909 SBN 471 25810 5

This SIAM edition is an unabridged, corrected republication of the work first published by Research Analysis Corporation, McLean, Virginia.

Preface

The primary purpose of this book is to provide a unified body of theory on methods of transforming a constrained minimization problem into a sequence of unconstrained minimizations of an appropriate auxiliary function. The auxiliary functions considered are those that define "interior point" and "exterior point" methods which are characterized respectively according to whether the constraints are strictly satisfied by the minimizing sequence. Initial emphasis is on generality, and the central convergence theorems apply to the determination of local solutions of a nonconvex programming problem. Strong global and dual results follow for convex programming; a particularly important example is the fact that the exterior point methods do not require the Kuhn-Tucker constraint qualification [77] to ensure convergence or characterize optimality.

In addition to giving a rather comprehensive exposition of the rich theoretical foundation uncovered for this class of methods, we wish to emphasize the demonstrated practical applicability of their various realizations. This has been brought about largely by the adaptation and further development of effective computational algorithms for calculating an unconstrained minimum and by the development of special extrapolation techniques for accelerating convergence. Significant progress has also been made in the development of computational techniques that exploit the special structures characterizing large classes of problems. In addition to these efficiencies, which have been effected for the method proper, some exploration has been done in combining the present methods with other mathematical programming algorithms to obtain even more efficient composite algorithms. Various computer programs at the Research Analysis Corporation, some operational since 1961, have been constantly utilized and improved and have provided extensive computational experience.

We have also attempted to provide some historical perspective for the basic approach with an effort toward synthesis. Formal derivations and ample intuitive arguments and simple numerical examples are interjected to motivate and clarify the basic techniques.

Finally, we have attempted to recapitulate the basic supporting developments in the theory of mathematical programming in a direct and simple manner. The satisfaction of the hypotheses of Farkas' lemma [44] is the motivation for various regularity assumptions on the problem functions. This leads directly to deducing the usual necessary conditions for optimality—in particular, the existence of finite Lagrange multipliers. Our presentation is intended to clarify this point of view, which often appears to be clouded in the literature; for example, the Kuhn-Tucker constraint qualification [77] is one such assumption that has often been incorrectly regarded as being essential to characterize optimality. We proceed from those results associated with assuming continuity of the problem functions to those involving first-and second-order differentiability. Several new necessary and sufficient conditions are given for the latter.

This book is the result of several years of extensive research and computational experimentation accomplished initially at The Johns Hopkins Operations Research Office, Bethesda, Maryland, and continued at the Research Analysis Corporation, McLean, Virginia. Initial efforts led to devising a variety of heuristic gradient methods, but they generally proved too slow and unreliable. In 1961 attention was directed to an auxiliary function idea proposed by C. W. Carroll [21]. Our theoretical validation of this approach then led to numerous extensions and generalizations and an efficient computational procedure, which provided the nucleus of the framework for the developments reported here.

Chapters 1, 2, and 7 contain supporting and supplementary material and are included primarily for perspective and completeness. Most of the results of Chapter 2, with the exception of several recently developed second-order necessary and sufficient conditions for isolated and nonisolated constrained minima, are well known. The basic theoretical results for the sequential methods are contained in Chapters 3 and 4. With a few exceptions, the development proceeds in the direction of decreasing generality, from the nonconvex local results of Chapters 3-5 to the convex global results of Chapter 6. Chapter 5 is essentially an in-depth analysis of convergence properties when certain additional conditions hold. Chapter 6 is virtually a recapitulation of Chapters 3-5, the results being recast and strengthened with the important additional assumption of convexity. The computational considerations associated with unconstrained minimization algorithms are relegated entirely to Chapter 8.

In addition to giving many new and more general results, we hope that the book may provide some much-needed clarification and unification in auxiliary function methodology, which only recently has been under significant development. The results for nonconvex problems are among the few that exist in this relatively intractable area and may hopefully be followed by computational breakthroughs as well. Finally, we hope that this exposition will lead to wider recognition of the extremely rich and fertile theoretical basis and the generally proved effective computational applicability of the methods in this class of procedures.

This book is intended for use by virtually anyone involved with mathematical programming theory or computations as a comprehensive reference for the evolution, theory, and computational implementation of auxiliary function sequential unconstrained methods and a concise reference for mathematical programming theory, several other interesting and effective methods, and some of the most recent advances in the theory and implementation of methods for unconstrained minimization. It could provide a fairly complete basis for an extensive course in mathematical programming or optimization at a wide variety of levels. The nature of the approach generally makes it possible to apply classical methods of analysis, a feature not shared by most techniques and one that can be easily exploited for pedagogical purposes. A complete understanding of the proofs and developments probably requires a solid introduction to analysis and linear algebra. A great portion of the text, however, is motivational and discursive material and requires little more than a brief exposure to elementary algebra. A significant portion depends only on basic notions encountered in first-year calculus and matrix theory.

Many of the results have appeared in previous publications [47-54, 84, 85] as the theory was developed, but a considerable number of significant results appear here for the first time.

Applications of these theoretical developments to the solution of important practical nonlinear programming models are contained in another RAC Research Series book by Jerome Bracken and Garth P. McCormick, "Selected Applications of Nonlinear Programming".

We should like to thank Dr. Nicholas M. Smith for his continued personal interest and effective technical direction. Sustained support for this work has come from the Army Research Office. Professor J.B. Rosen, Dr. James E. Falk, Professor A. Charnes, Dr. Jerome Bracken, Mr. W. Charles Mylander, III, and Professor C.E. Lemke have contributed to the development of this work, particularly Professor Lemke, who reviewed an earlier version of the manuscript and gave numerous helpful suggestions. Special thanks go to Mrs. L. Zazove, who patiently typed several versions of the manuscript.

Anthony V. Fiacco
Garth P. McCormick

McLean, Virginia

Preface to the Classic Edition

The first edition of this book has long been out of print. The primary motivation for its reissue is the current interest in interior point methods for linear programming sparked by the 1984 work of N. Karmarkar. The connection of his projective scaling method with SUMT was pointed out early by Gill and others. The recent work in affine scaling, path following and primal-dual methods bears even more resemblance to our earlier work.

Our initial intent was not to develop a method for linear programming, but for the general nonlinear programming problem. The original book contains some material which was published elsewhere in the open literature, but much of it appears only in the book. In particular, the analysis concerning the trajectory of unconstrained minimizers of the logarithmic barrier function is similar to recent work and appears only here.

Many areas of research were started in this book: e.g. the relationship of penalty function theory to duality theory; use of directions of non-positive curvature to modify Newton's method when the Hessian matrix is indefinite; integration of the first and second order optimality conditions into convergence and rate of convergence analysis of algorithms; the identification of factorable functions as an important class in applied mathematics; and the beginning of the very important area of sensitivity analysis in nonlinear programming.

Except for corrections, the revision is exactly as originally published. The most important corrections are the proofs of Theorems 6 and 7. Since its original publication the term "penalty function" is now commonly called a "barrier function". This and a few other usages were kept as in the original.

The book was awarded the Lanchester Prize in Operations Research for the year 1968. We hope that readers will find it as fresh and interesting and valuable now as it was then.

Anthony V. Fiacco
Garth P. McCormick
Department of Operations Research
School of Engineering and Applied Science
George Washington University
Washington, D. C.

Symbols and Notations

x	$\equiv (x_1, \ldots, x_n)^T$, an n by 1 vector of variables.
$(E^n)^+$	the nonnegative orthant of Euclidian n-space.
$\|x\|$	$= (\sum_{j=1}^{n} x_j^2)^{1/2}$, the usual Euclidian norm.
$\nabla_x f(x^k)$,	(sometimes written ∇f^k) is the n by 1 vector whose jth element is $\partial f(x^k)/\partial x_j$.
$\nabla_{xx}{}^2 f(x^k)$,	(sometimes written $\nabla^2 f^k$) is the n by n matrix whose i, jth element is $\partial^2 f(x^k)/\partial x_i\, \partial x_j$, the Hessian of f at x^k.
Problem A,	minimize $f(x)$ subject to $g_i(x) \geq 0$, $i = 1, \ldots, m$, $h_j(x) = 0, j = 1, \ldots, p$.
R	$\equiv \{x \mid g_i(x) \geq 0,\ i = 1, \ldots, m\}$, the region defined by the inequalities of Problem A.
R°	$\equiv \{x \mid g_i(x) > 0, i = 1, \ldots, m\}$, the interior of R.
$\mathcal{L}(x, u, w)$	$\equiv f(x) - \sum_{i=1}^{m} u_i\, g_i(x) + \sum_{j=1}^{p} w_j\, h_j(x)$, the Lagrangian associated with Problem A.
$\alpha(\theta)$,	an arc in E^n parameterized by θ whose tangent at $\theta = \eta$ is denoted by $D\alpha(\eta)$ and whose vector of second derivatives is denoted by $D^2\, \alpha(\eta)$.
$A^\#$,	a pseudoinverse of the matrix A that satisfies $AA^\#A = A$.
$\text{diag}(g_i)$,	a diagonal matrix whose ith diagonal element is g_i.
$L(x, r)$	$\equiv f(x) - r \sum_{i=1}^{m} \ln g_i(x)$, the logarithmic interior point penalty function for Problem B (Problem A with no equality constraints).

$P(x, r)$ $\equiv f(x) + r \sum_{i=1}^{m} 1/g_i(x)$, the inverse interior point penalty function for Problem B (P-function).

$U(x, r)$ $\equiv f(x) + s(r)\, I(x)$, general interior unconstrained minimization function.

$T(x, t)$ $\equiv f(x) + p(t) O(x)$, general exterior unconstrained minimization function.

$V(x, r, t)$ $\equiv f(x) + s(r) I(x) + p(t) O(x)$, general mixed interior-exterior unconstrained minimization function.

$W(x, r)$ $$\equiv f(x) - r \sum_{i=1}^{m} \ln g_i(x) + \sum_{i=m+1}^{q} \frac{\{\min[0, g_i(x)]\}^2}{r},$$

the W-function, a mixed interior-exterior unconstrained minimization function for Problem M.

$x^k \to y$ the sequence $\{x^k\}$ converges (strongly, i.e., in norm) to y.

Contents

1

Introduction

1.1 STATEMENT OF THE MATHEMATICAL PROGRAMMING PROBLEM

The mathematical programming problem is to determine a vector $x^* = (x_1^*, \ldots, x_n^*)^T$ that solves the problem

$$\text{minimize } f(x) \tag{A}$$

subject to

$$g_i(x) \geq 0, \qquad i = 1, \ldots, m, \tag{1.1}$$

$$h_j(x) = 0, \qquad j = 1, \ldots, p. \tag{1.2}$$

When the problem functions f, $\{g_i\}$, and $\{h_j\}$ are all linear Problem A is called a *linear programming problem*. If any of the functions is nonlinear the problem is called a *nonlinear programming problem*. There are other terms, such as convex, concave, separable, quadratic, and factorable, which may apply to special cases of Problem A, and these will be defined later. While all the remarks in this book apply in particular to these special cases, we shall at the outset concern ourselves with problems where f, $\{g_i\}$, and $\{h_j\}$ can take on any form of nonlinearity subject only to continuity and differentiability requirements.

The following is a simple example of a nonlinear programming problem.

Example.

$$\text{minimize } f(x) = |x_1 - 2| + |x_2 - 2|$$

subject to

$$g_1(x) = x_1 - x_2^2 \geq 0,$$

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

The dashed lines in Figure 1 represent isovalue contours of the objective function; that is, points at which the objective function $f(x)$ has constant value. The feasible region is the set of points that satisfy the constraints of

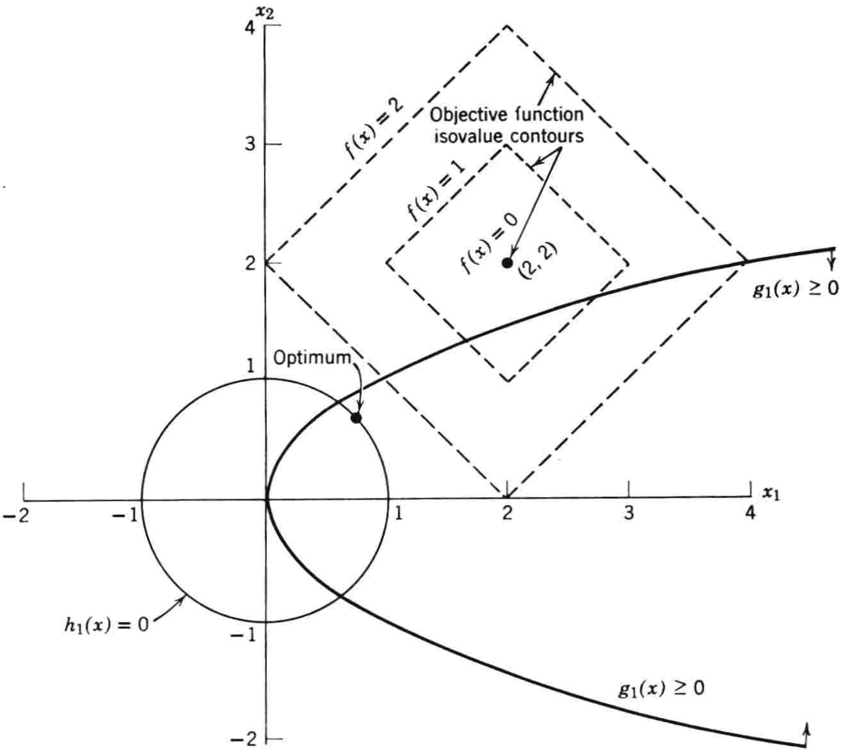

Figure 1 Nonlinear programming example.

the problem. In this example the feasible region is the arc of the circle lying within the parabola. A solution to the problem is any point in the feasible region with smallest objective function isocontour value. This is seen by inspection to be $(\sqrt{2}/2, \sqrt{2}/2)$. If the equality constraint is removed the solution is seen to be at $(2, \sqrt{2})$. If both constraints are removed the solution is at (2, 2). In the latter case (2, 2) is called an *unconstrained* minimum.

Usually the functions of Problem A are required to be continuous. Much of the theory of nonlinear programming concerns the case when the functions are continuously differentiable, or twice continuously differentiable. In these instances it is possible to prove theorems which characterize solutions to Problem A. These theorems in turn influence the development of algorithms for solving the programming problems.

Several classes of mathematical programming problems have been dealt with in recent years. We briefly mention some of these and give a number of standard references that develop theoretical results and computational methods for solving the corresponding problem. The linear programming problem has been treated extensively, and many significant results have been forthcoming, such as effective methods for problems having a particular structure. There exists an enormous literature on the subject. For basic results refer to the list of annotated references given in [96] and to the developments and additional bibliography given in [25, 26, 34, 62].

Particular results and algorithms have been obtained for "quadratic programming," where $f(x)$ is a positive semidefinite quadratic form, and the constraints are linear [7, 8, 110]. Special methods have been developed when $f(x)$ is a convex separable function and the constraints are linear [27, 88]. Special-purpose algorithms also exist for the case where $f(x)$ is convex and the constraints are linear [45, 98].

The case where $f(x)$ is convex, the $g_i(x)$ concave, and the $h_j(x)$ linear has received particular attention. When these conditions prevail (A) is called a convex programming problem. The smoothness of the problem functions makes the problem well behaved, and the convexity-concavity assumptions assure that the feasible region (set of points satisfying the constraints) is convex and, most importantly, that any local solution is also global. The basic optimality conditions for the problem were given in [77]. Numerous other contributions to the theory and development of computational algorithms have appeared, largely in the last decade [1, 6, 28, 39, 49, 50, 65, 70, 74, 99, 111, 118].

The developments mentioned above are often applicable in a "local" sense; that is, they hold if x is restricted to a suitable domain such that the requisite conditions hold in that domain. This means that some results are easily extended to apply to the characterization of *local* solutions of (A), when (A) is a nonconvex problem. The "general" nonconvex problem, where (A) is not even necessarily "locally convex" in any neighborhood of a relative minimum, has remained rather intractable. Most of the results in this important problem area have been theoretical and are quite recent [16, 17, 101, 104, 117].

The group of algorithms called sequential unconstrained minimization techniques has given rise to numerous theoretical results and effective computational procedures for solving the convex programming problem. Recent developments indicate that these results can be generalized and extended significantly, since the basic technique can be validated under very general and weak conditions. Thus a number of important results have been obtained for nonconvex programming, as well as additional generality and a finer characterization for problems having a special structure. A history of sequential unconstrained methodology is contained in the next section.

This book pursues the development of these sequential unconstrained methods.

1.2 HISTORICAL SURVEY OF SEQUENTIAL UNCONSTRAINED METHODS FOR SOLVING CONSTRAINED MINIMIZATION PROBLEMS

The Transformation Approach

The methods we shall discuss are based on transforming a given constrained minimization problem into a sequence of unconstrained minimization problems. This transformation is accomplished by defining an appropriate auxiliary function, in 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 auxiliary function by controlled changes in the value of a parameter, a sequence or family of unconstrained problems is generated that have solutions converging to a solution of the original constrained problem.

For simplicity in the present discussion, we proceed formally to sketch the basic idea. The problem is to find a solution x^* of

$$\text{minimize } f(x) \tag{B}$$

subject to

$$g_i(x) \geq 0, \qquad i = 1, 2, \ldots, m,$$

where $x \in E^n$

A typical unconstrained auxiliary function may have the form

$$\varphi[x, \lambda(t)] \equiv f(x) + \sum_{i=1}^{m} \lambda_i(t) G[g_i(x)],$$

where t is a parameter, the $\{\lambda_i(t)\}$ are weighting factors, and $G(y)$ is generally a monotonic function of y that behaves in some well-chosen manner at $y = 0$. Typical choices are either that $G(y) > 0$ for $y < 0$ and $G(y) = 0$ for $y \geq 0$, or that $G(y) \to +\infty$ as $y \to 0^+$. The former choice usually is associated with procedures that are not concerned with constraint satisfaction except at the solution, and the latter, where constraint satisfaction is enforced throughout.

When successful, the method generally proceeds computationally as follows. Select a sequence $\{t_k\}$ such that $t_k \geq 0$, and $t_k \to \infty$ as $k \to \infty$. Compute a minimum x^k of $\varphi[x, \lambda(t_k)]$ for $k = 1, 2, \ldots$. Under appropriate conditions such an x^k exists and is an unconstrained minimum of $\varphi[x, \lambda(t_k)]$. Usually the most desirable result is that $\lim_{k\to\infty} x^k = x^*$, a solution of (B). A weaker result, often adequate, is that $f(x^k) \to f(x^*)$, a minimum value of the objective function. Invariably, the result follows that

$$\lim_{k\to\infty} \sum_{i=1}^{m} \lambda_i(t_k) G[g_i(x^k)] = 0,$$

so that also

$$\lim_{k \to \infty} \varphi[x^k, \lambda(t_k)] - f(x^*) = 0;$$

that is, the modified objective function converges to the same minimal value as the original objective function. This means, in effect, that the influence of the constraints on the modified objective or auxiliary function is gradually relinquished and finally removed in the limit.

This, then, is the general idea of the approach. Its great advantage lies in the fact that the constraints need not be dealt with separately and that classical theory and modern methods for computing unconstrained extrema can be brought to bear. The theoretical difficulties are in such areas as prescribing general conditions that guarantee convergence of either the minimizing sequence or the associated modified objective function values, and in validating acceleration procedures. Computationally, obtaining rapid convergence is a central concern and depends on the efficacy of methods for unconstrained minimization and on procedures for effective extrapolation. In recent years a substantial body of theory has been established and effective computational algorithms have been implemented [49, 51].

We turn to a chronological account of the origin and development of the technique of solving a constrained problem by transforming it into a sequence of unconstrained problems. Before pursuing this relatively recent history, a note on the classical Lagrange multiplier procedure is particularly in order.

The Lagrange Multiplier Technique

Our primary interest is in sequential unconstrained methods for solving a problem of type (B). However, it should be remarked at the outset that the Lagrange multiplier technique (see [108], for example) for handling problems of type (E) [minimize $f(x)$ subject to $h_j(x) = 0$, $j = 1, \ldots, m$] is surely a technique for transforming the problem into an unconstrained problem. Note that this simply amounts to the choices $\lambda_j(t) = \lambda_j$ (constant) and $G(y) = y$ in $\varphi[x, \lambda(t)]$. In fact, by introducing appropriate slack variables, we can transform any problem of type (B) into a problem of type (E) and then formulate the associated Lagrangian problem. This procedure has been applied to variational problems [11, 63, 106].

We do not propose to delve into a detailed description of the classical Lagrange multiplier technique, or into a discussion of the relative merits of the procedure or of the computational difficulties inherent in a direct application of it to a particular example. These matters have been treated in some detail elsewhere [23, 37, 76]. We simply wish to indicate here that the Lagrangian is a classical example of the unconstrained auxiliary function approach.

Recent developments on the Lagrange multiplier technique are discussed in Sections 7.4 and 7.5. This is based on the work in [40] and [43].

Besides being viewed directly as a special case of the type of auxiliary function we are considering, it follows that the Lagrangian is inextricably associated with every method for mathematical programming, because conditions for characterizing solutions of mathematical programming problems such as (B) or (E) directly involve the associated Lagrangian. This may be viewed as a direct consequence of the fact that, assuming differentiability, a necessary condition that x^* solve (E) is that x^* be a stationary point of the associated Lagrangian for some real λ_i, constants to be determined from the requirement of stationarity and the relations $h_j(x) = 0$, $j = 1, \ldots, m$. Thus the Lagrangian will always be very much involved in our theoretical results, as much of the text will amply testify.

Owing to the additional fact that the Lagrangian may be viewed as a special case of the auxiliary unconstrained function, $\varphi[x, \lambda(t)]$, it is not surprising that we shall see extremely close connections throughout between the associated Lagrangian of a mathematical programming problem and the particular auxiliary function being utilized to transform that problem into a sequence of unconstrained minimizations.

The following historical account is intended to be indicative of the main stream of developments. Inclusions and omissions reflect the authors' point of view and, likewise, interpretations and appraisals are subjective.

Chronological Survey of Developments

In 1943 R. Courant [29] suggested studying the conditions for stationarity of $f(x) + tg^2(x)$ as $t \to \infty$, to analyze motion constrained to satisfy $g(x) = 0$ in terms of unconstrained motion. The suggestion was motivated by physical considerations and was not offered directly as a technique for solving a mathematical programming problem. The idea was apparently not rigorously pursued for over a decade.

In the interim several important theoretical developments took place. In 1951 Dantzig [33] formalized the linear programming problem and offered a first version of the simplex method. An enormous amount of effort was subsequently directed toward the development and implementation of linear programming algorithms. Shortly thereafter, in 1951, Kuhn and Tucker [77] published their results on necessary and sufficient conditions characterizing the solution of the nonlinear convex programming problem and gave an equivalence between this problem and the saddle-point problem of the Lagrangian. In 1950 Arrow and Hurwicz [5] treated this latter equivalence as well.

Arrow [4] in 1951 devised a gradient technique for approximating saddle