SOME THEORIES AND ALGORITHMS IN NONLINEAR PROGRAMMING

by

Ya-xiang Yuan
Pembroke College
University of Cambridge

December 1985

A Dissertation submitted for
the Degree of Doctor of Philosophy
at the University of Cambridge.

千里之行始于足下。

——老子

Even a journey of one thousand miles begins with a small step

-- Lao Zhi

PREFACE

All the results in this dissertation are believed to be original except where explicit reference is made to other authors. The results were worked out between January 1983 and December 1985 in the Department of Applied Mathematics and Theoretical Physics under the supervision of Professor M.J.D. Powell. The work in Sections 3.3 and 3.4 was carried out in collaboration with my supervisor, Professor M.J.D. Powell. No part of this thesis has been, or is being currently, submitted for a degree or diploma or other qualification at any other University.

ACKNOWLEDGEMENTS

I would like to record my deep gratitude to my supervisor Professor M.J.D. Powell, F.R.S. , who introduced me to Optimization, whose work has been a substantial influence on my thinking, and whose ideas have given me much inspiration. It is a great honour for me to have the opportunity to study under his supervision. I am greatly indebted to him for his constant help and encouragement. It is my pleasure to thank him for his great patience and kindness in supervising my research. Without his guidance and advice, this dissertation would not have been possible.

I also wish to thank Professor R. Fletcher of Dundee University for communicating with me and giving advice and help. I am grateful to all my former teachers in China who introduced me to mathematics.

I am under an obligation to the Chinese Academy of Science for financial support which made my studying in Cambridge possible. I am also grateful to the Vice-Chancellors Committee for the ORS award, the University of Cambridge for the OSB bursary, Pembroke College for a bursary, and D.A.M.T.P. and Pembroke College for conference funds. I wish to thank Fitzwilliam College for a Research Fellowship. It is my pleasure to thank the Education Section of the Chinese Embassy in London for travelling expenses.

CONTENTS

CHAPTER I INTRODUCTION

1.1 The Problem

The problem we consider is

$$\text{minimize} \quad f(x) \ , \qquad x \in \mathbb{R}^n \ , \tag{1.1.1}$$

subject to

$$c_i(x) = 0 \ , \quad i=1,2,\ldots,m_e; \tag{1.1.2}$$

$$c_j(x) \geq 0, \quad j=m_e+1,\ldots,m. \tag{1.1.3}$$

where $f(x)$, $c_i(x)$ $(i=1,2,\ldots m)$ are real functions in $\mathbb{R}^n$. We use the notations $E=\{1,2,\ldots,m_e\}$ and $I=\{m_e+1,\ldots,m\}$. A point x is called a **feasible point** if it satisfies (1.1.2)-(1.1.3). Let X be the set of all feasible points, which is referred as the feasible set. Since our goal is to select a x from the feasible set to minimize $f(x)$, the function $f(x)$ is called the **objective function**, and $c_i(x)$ $(i=1,2,\ldots,m)$ the **constraint functions**. If either $f(x)$ or at least one of $c_i(x)$ is nonlinear in x, problem (1.1.1)-(1.1.3) is called a nonlinear program, in contrast to a linear program, where all functions in (1.1.1)-(1.1.3) are linear. If $m>0$ (1.1.1)-(1.1.3) is called a **constrained optimization** problem, otherwise a **unconstrained** problem. x^* is called a **local solution**, or local minimizer, if there exists a small neighbourhood of it such that the inequality

$$f(x) \geq f(x^*) \tag{1.1.4}$$

holds for all feasible points x in the neighbourhood. A local solution is called a **global solution**, or **global minimizer**, if (1.1.4) holds for all feasible points x. At a solution, we say a constraint $c_i(x)$ is **active** if $c_i(x)=0$, otherwise it is **inactive**.

Nonlinear programming problems arise in many different forms and may be found in engineering, economics, physical sciences, business administration, and mathematics. The following example shows how a real decision-making problem can be represented as (1.1.1)-(1.1.3). Considering 2-dimensional Euclidean space, we want to calculate a point x^* in the curve $\{(x,y): y=(x-1)^2\}$ which is closest to the origin. Writing this as a mathematical model, we need to solve:

$$\text{minimize} \quad x_1^2 + x_2^2 \quad , \tag{1.1.5}$$

subject to

$$x_2 - (x_1-1)^2 = 0 \, . \tag{1.1.6}$$

For a unconstrained optimization problem, it is easy to see that a necessary condition for x^* being a local solution is

$$\nabla f(x^*) = 0 \quad , \tag{1.1.7}$$

if $f(x)$ is differentiable. A point satisfying (1.1.7) is also called a **stationary point**. A second order necessary condition is that the matrix $\nabla^2 f(x^*)$ be positive semidefinite, that is,

$$d^T \nabla^2 f(x^*)\, d \geq 0 \tag{1.1.8}$$

for all $d \in \mathbb{R}^n$. A sufficient condition is that x^* satisfies both (1.1.7) and

$$d^T \nabla^2 f(x^*) d > 0 \qquad (1.1.9)$$

for all nonzero $d \in \mathbb{R}^n$.

For constrained problems, a **Lagrangian function**

$$L(x,\lambda) = f(x) - \lambda^T c(x) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x) \qquad (1.1.10)$$

is defined, where $\lambda = (\lambda_1,\ldots,\lambda_m)^T \in \mathbb{R}^m$. λ_i $(i=1,\ldots,m)$ are called **multipliers**. Such a function can be derived by sensitivity analysis (Wilde, 1967), which considers small perturbations in the problem. It has been found that the function (1.1.10) is connected with Lagrange's (1760-1761) philosophy (see Wilde and Beightler, 1967). The following necessary condition, due to Kuhn and Tucker (1951), is well known:

Theorem 1.1.1 Assume that $f(x)$ and $c_i(x)$ $(i=1,2,3,\ldots)$ are continuously differentiable in an open set containing x^*. If x^* is a local solution of (1.1.1)-(1.1.3), and if $\nabla c_i(x^*)$ $(i \in I' \cup E)$ are linearly independent, where

$$I' = \left\{ i \in I, \quad c_i(x^*) = 0 \right\} , \qquad (1.1.11)$$

then there exits a $\lambda^* = (\lambda_1^*,\ldots,\lambda_m^*)^T \in \mathbb{R}^m$ such that

$$\left.\begin{aligned} &\nabla_x L(x^*, \lambda^*) = 0 \\ &c_i(x^*) = 0 \qquad i \in E \\ &c_j(x^*) \geq 0, \quad \lambda_j^* \geq 0, \quad j \in I \\ &\lambda_i^* c_i(x^*) = 0 \qquad i \in I \end{aligned}\right\} . \qquad (1.1.12)$$

The conditions (1.1.12) are called **Kuhn-Tucker conditions**. A point that satisfies (1.1.12) is called a **Kuhn-Tucker point**. The

last equation in (1.1.12) is the **complementarity condition**, which shows that any inactive constraint must have a zero multiplier. The so called **strict complementarity condition** is that $|c_i| + |\lambda_i^*| > 0$ for all i. The following results are also well known (see Fletcher, 1981):

Theorem 1.1.2 Assume that $f(x)$ and $c_i(x)$ (i=1,2,3,...) are twice continuously differentiable in an open set containing x^*. If x^* is a local solution of (1.1.1)-(1.1.3), and if ∇c_i ($i \in I' \cup E$) are linearly independent, then there exists a λ^* such that (1.1.12) holds and

$$d^T W^* d \geq 0 \qquad \text{for all } d \in S^* \ , \tag{1.1.13}$$

where

$$W^* = \nabla^2 f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2 c_i(x^*) \tag{1.1.14}$$

and

$$S^* = \Big\{ d \ \Big| \ d \neq 0, \ \ d^T \nabla c_i(x^*) = 0 \text{ if } i \in E \text{ or } \lambda_i^* > 0, \ \ d^T \nabla c_i(x^*) \geq 0 \text{ if } i \in I' \text{ and } \lambda_i^* = 0 \Big\}. \tag{1.1.15}$$

Theorem 1.1.3 Assume that $f(x)$ and $c_i(x)$ (i=1,2,3,...) are twice continuously differentiable in an open set containing x^*. If at x^* there exists a λ^* such that (1.1.12) holds and if

$$d^T W^* d > 0 \tag{1.1.16}$$

holds for all $d \in S^*$, then x^* is an isolated local solution.

The condition that $\nabla c_i (i \in I' \cup E)$ are linear independent in Theorems 1.1.1 and 1.1.2 can be relaxed. For more details, see Fletcher (1981). Optimality conditions are also studied by Bandler (1971), Ben-tal (1980), Fiacco and McCormick (1968), Guignard (1982), Hancock (1917), Mangasarian and Fromoritz (1967), Rockafellar (1970, 1981), Uzawa (1958), Wilde (1964), and many others.

A Kuhn-Tucker point x^* is a global solution if all the functions are linear. Hence it is easy to see that x^* is a solution to problem (1.1.1)-(1.1.3) if we replace f and c_i by their linear expansions at x^* respectively. Therefore we have to consider second order derivatives if we want to find a "better" point near x^*, that is, a feasible point having smaller function value f. But most methods for solving (1.1.1)-(1.1.3) do not use second order derivatives, so their tasks are no more than seeking a Kuhn-Tucker point.

1.2 Literature Survey

1.2.1 Methods for Unconstrained Problems

First we consider the case when $n = 1$. We discuss this case individually not only because it is simple and it can give us inspirations for the case when $n > 1$, but also because many methods for solving a function of several variables need to calculate the least function value along a specific direction.

Methods that use function values only are called **direct search methods**. Assume f(x) is unimodal on a finite interval, say $[a, b]$, and we want to calculate the **minimum** of the function.

The **Fibonacci search method** is given by Kiefer (1953), and he shows that the method is optimal in a certain sense. The Fibonacci search method is also studied by Johnson (1956) and Oliver and Wilde (1964). Other direct search methods include **dichotomous search** method, **golden section** method, **random search** method and **polynomial approximation** methods. For more details, see Swann (1972) and Wilde (1964).

If $f(x)$ is continuously differentiable, a simple and explicit method is the **descent direction** method. The method is as follows. Given an approximate point x, we subtract from x a small positive number δ times $\nabla f(x)$. We accept the new point if it reduces the function value, otherwise we reduce δ, and try again. Two most widely used polynomial approximation methods are a **quadratic** method (Rosenbrock, 1960) and a **cubic** method (Fletcher and Powell, 1963). These methods choose a new point by considering a quadratic or cubic approximation to the objective function and its gradient at old point(s). Changing the problem to $\nabla f(x) = 0$, we can apply the methods for nonlinear equations (Ortega and Rheinboldt, 1970).

Now we consider the general case when $n > 1$. There are two main classes of methods. The first is **line search methods** and the other is **trust region algorithms**. They are all iterative, and at the beginning of calculations, an initial point x_1 should be provided. The main ideas of both classes of methods are given below. Then we review methods that do not require derivatives, followed by those for which $\nabla f(x)$ is needed.

For line search methods, a search direction d_k is computed at the beginning of an iteration. Then a **step-length** $\alpha_k > 0$ is evaluated by some **line search** technique and the next iterate is

set by $x_{k+1} = x_k + \alpha_k d_k$. Two common line search techniques are **exact line search** and the **Armijo line search** (Armijo, 1966). An exact line search requires that

$$f(x_k + \alpha_k d_k) = \min_{\alpha>0} f(x_k + \alpha d_k) , \tag{1.2.1}$$

and Armijo line search chooses α_k such that

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k b_1 (\nabla f(x_k))^T d_k \tag{1.2.2}$$

and

$$(\nabla f(x_k + \alpha_k d_k))^T d_k \geq b_2 (\nabla f(x_k))^T d_k , \tag{1.2.3}$$

where $b_1 \in (0, 1)$, and $b_2 \in (b_1, 1)$ are positive constants (Powell, 1976c).

For trust region methods, at the beginning of an iteration a trust region is also available. An approximate problem is solved within the trust region and a trial step d_k is obtained. Then some test is applied to decide whether we should accept this step. If d_k is accepted, we let $x_{k+1} = x_k + d_k$ and continue our calculation, otherwise the trust region is reduced and we resolve the approximate problem.

Most direct search methods for 1-dimensional problems can be easily generalized to the case when $n > 1$. The **generalized Fibonacci** method is described by Box, Davies and Swann (1969). A simple method, due to Southwell (1946), is the **alternating variables** method, which consists of minimizing with respect to each independent variable in turn (see Swann, 1972, Fletcher, 1980). Shah, Buehler and Kempthorne (1961, 1964) give the **parallel**

tangents method, which can find the solution within $2n+1$ iteration if the hypersurface contours of the objective function are concentric ellipsoids. The **pattern search** method, or **Hooke-Jeeves** method, consists of two phases. One is to find a **pattern step** and the other is to make a **exploratory move** (Hooke and Jeeves, 1961). This method is modified by Box, Davies and Swann (1969). The **Rosenbrock's method**, also called **rotating coordinate method**, makes use of explorations along a set of directions (Rosenbrock, 1960). After each iteration, a new set of orthonormal search directions are defined. The method is modified by Davies, Swann and Campey (Swann, 1964, 1972). Economical methods for calculating the new search vectors are given by Palmer (1969) and Powell (1968). The **simplex method** is given by Spendley, Hext and Himsworth (1962), and is improved by Nelder and Mead (1965). The method uses $n+1$ points which form a **simplex**. At every iteration, a point which has the greatest function value is removed from the simplex, and a better point is added. A very efficient direct search method is given by Powell (1964). In **Powell's method**, at the beginning of each iteration, n linearly independent directions are available. The method searches along n directions in turn. A new direction is defined as the one from the starting point to the point obtained at the end of the n-th 1-dimensional search. Then searching along the new direction gives a new point, which is the starting point for the next iteration. One of the first n directions is replaced by the new direction and the next iteration begins. Powell's method is a special conjugate directions method. Such methods use conjugate directions $d_1, d_2, \ldots$ satisfying

$$d_i^T G d_j = 0 \quad \text{if} \quad i \neq j \ , \tag{1.2.4}$$

if the objective function is a convex quadratic function

$$f(x) = g^T x + \frac{1}{2} x^T G x \quad , \tag{1.2.5}$$

where G is a positive definite matrix. It is well known that conjugate direction methods have the nice **quadratic termination** property, that is, the solution can be found within n-steps if the objective function is (1.2.5). The first method to make use of conjugate directions seems to be given by Smith (1962), which is much more complex than Powell's method (Swann, 1972). Zangwill (1967a) points out that Powell's method could break down and he modifies the method. However, since the objective function can be closely approximated by a quadratic function near the solution, Powell's method seems to be the most efficient method that uses function values only. Methods without evaluating derivatives are also discussed by Fletcher (1965), Kowalik and Osborne (1968) and Powell (1974). (*to be continued at page 9a*)

(*continued from page 9b*)

Much more attention has been given to the methods that use first order derivatives, since they are much more efficient. A very obvious method is the steepest descent method, which takes the steepest descent direction $-\nabla f(x)$ as a search direction. This technique, first proposed by Cauchy (1847) to solve simultaneous linear equations, was exhumed a century later by Courant (1943) to deal with problems in mathematical physics (Wilde and Beightler, 1967). However the steepest descent method may converge very slowly (for example, see Fletcher, 1980).

The **conjugate gradient method** is pioneered by Fletcher and Reeves (1964), based on a method of Hestenes and Stiefel (1952) for solving linear systems. The first search direction of the method is the steepest descent direction, that is,

(*continued from line 15 of page 9*)

There is an example given by Sargon that Powell's algorithm (1964) cycles because no new directions can be obtained, and for problems of reasonable size it is a widely held view that using variable metric method with gradients approximated by finite differences is in practice better than any of the methods described (Dixon, private communication, 1986). More references are listed below:

R.P. Brent, *Algorithms for Minimization without Derivatives* (Prentice Hall Inc., Englewood Cliff. N.J., 1973)

L.C.W. Dixon, (ed.) *Optimization in Action* (Academic Press, London, 1976).

L.C.W. Dixon and G.P. Szego, (eds.) *Towards Global Optimization I* (North Holland, Amsterdam, 1975)

L.C.W. Dixon and G.P. Szego, (eds.) *Towards Global Optimization II* (North Holland, Amsterdam, 1978)

P.E. Gill and W. Murray, "Quasi-Newton methods for unconstrained optimization", *J. Ins. Math. Appl.* 9(1972) 91-108.

R. Mifflin, "A superlinearly convergent algorithm for minimization without evaluating derivatives", *Math. Prog.* 9(1975) 100-117.

F. Sloboda, "Parallel method of conjugate directions for minimizations", *Appl. Math.* 20(1975) 436-446.

F. Sloboda, "A conjugate direction method and its application", in: *Optimization Techniques* (Proc. 8th IFIP Conf., Wurzberg, 1977) *Part* 2, pp. 147-157.

G.W. Stewart, "A modification of Davidon's minimization method to accept difference approximation of derivatives", *J. Ass. Comput. Mach.* 14(1967) 72-83.

C. Sutti, "A new method for unconstrained minimisation without derivatives", in: *Towards Global Optimisation* (Proc. Workshop Univ. Cagliari, Cagliari, 1974) pp. 277-289 (1974a).

C. Sutti, "Remarks on conjugate direction methods for minimisation without derivatives", in: *Towards Global Optimisation* (Proc. Workshop Univ. Cagliari, Cagliari, 1974) pp. 290-304 (1974b).

C. Sutti, "Convergence proof of minimization algorithms for nonconvex function", *J. Opt. Theory Appl.* 23(1977) 203-210.

C. Sutti, "Nongradient methods for unconstrained optimizaiton", in: *Nonlinear Optimization* (Proc. Intern. Summer School, Univ. Bergamo, Bergamo, 1979) pp. 137-147.

C. Sutti, "Nongradient minimization methods for parallel processing computers", Part 1 and Part 2, *J. Opt. Theory Appl.* 39(1983) 465-474 and 475-488.

I am very grateful to Dr. L.C.W. Dixon for drawing my attention to the work of Brent, Mifflin, Sargon, Sloboda and Sutti.

(*to be continued at line* 16 *of page* 9)