nonlinear optimization

theory and algorithms

Edited By

L.C.W. Dixon
E. Spedicato
G.P. Szegö

BIRKHÄUSER

nonlinear optimization

theory and algorithms

(The Proceedings of an International Summer School held at University of Bergamo, Italy, in September 1979 and repeated at the Hatfield Polytechnic, England, in July 1980)

Edited by

L.C.W. Dixon

The Numerical Optimisation Centre
The School of Information Science
The Hatfield Polytechnic
England

E. Spedicato

The Department of Mathematics
University of Bergamo
Italy

BIRKHÄUSER

L.C.W. Dixon

The Numerical Optimisation Centre
The School of Information Science
The Hatfield Polytechnic
England

E. Spedicato

The Department of Mathematics
University of Bergamo
Italy

G.P. Szego
Il Rettore
University of Bergamo
Italy

Library of Congress Catalog Number 80-36665
Main entry under title:

Nonlinear optimisation, theory and algorithms.

Bibliography: p.
Includes index.
1. Mathematical optimization—Congresses.
2. Nonlinear theories—Congresses. I. Dixon, Laurence
Charles Ward. II. Spedicato, E. III. Szegö, G. P.
QA402.5.N65 519.7′6 80-36665
ISBN 3-7643-3020-1

ISBN 3-7643-3020-1

Printed in USA

PREFACE

This volume contains the lecture notes of a Summer School held at the University of Bergamo, Italy in September 1979. The school was funded by both the European Economic Community, CREST-ITG Programme on "Training in Informatics" and by the Italian C.N.R. The director was Prof. E. Spedicato of the University of Bergamo. Due to the success of that Summer School the decision was made to repeat it at Hatfield Polytechnic, England in July 1980, this time funded by the European Economic Community, CREST-ITG Programme and the English S.R.C. The director on this occasion was Dr. L.C.W. Dixon of the Numerical Optimization Center, the Hatfield Polytechnic. The editors wish to gratefully acknowledge this support, the permission of the sponsors to publish the Proceedings and the cooperation of the lecturers in submitting their contributions for publication.

The volume contains eighteeen papers. These can be divided into three groups. The first eight papers on Unconstrained Optimization include contributions from L.C.W. Dixon, C. Lemarechal, J.J. McKeown, E.Spedicato, C.Sutti and Ph.L. Toint, and cover most recent research in this area. Topics treated include convergence theory, the variable metric method, sparsity, least squares problems including the effect of large residuals, the conjugate gradient method and the special nongradient methods and methods for nondifferentiable functions. The second part on Constrained Optimization contains five papers including contributions from M.C. Bartholomew-Biggs, D.P. Bertsekas, M.J.D. Powell and K. Ritter. These cover the necessary and sufficient conditions for constrained optima, convergence and super-linear convergence on linearly constrained problems and the performance of penalty, multiplier, variable metric and recursive quadratic programming methods for the more general problem. The third part contains four contributons from F. Archetti, H.W. Kuhn, J.J. McKeown and G.P. Szegő. Topics treated include the fixed point approach, sensitivity analysis of the solution and the global optimization problem. The last paper by G.P. Szegő and G. Treccani is a mathematical appendix in which a summary of the properties of real-valued functions which are relevant to optimization is presented.

It is hoped these papers communicate the richness and diversity that exist today in this vital area of research that is being applied to more and larger practical problems every year.

L.C.W. Dixon
E. Spedicato
G.P. Szegő

CONTENTS

Lecture 1

INTRODUCTION TO NUMERICAL OPTIMISATION

by

L. C. W. Dixon
The Numerical Optimisation Centre
The Hatfield Polytechnic
England

1 The Problem

In this series of lectures we will be concerned with the problem of locating the minimum of an objective function numerically on a digital computer. We will be assuming that in any problem of interest, there is an objective function f which depends on the values of n variables x_i ; $i = 1,\ldots, n$. In any practical problem solutions of interest will have to satisfy certain constraints which we may represent as $x \in S$; then the problem could be posed.

Obtain x^{**} such that

$$\text{(i)} \quad x^{**} \in S$$

$$\text{(ii)} \quad f(x^{**}) \leq f(x) \quad \text{all} \quad x \in S \; ; \qquad (1.1)$$

such a point is termed the <u>global minimum point</u> of the problem.

However, for most of the series we will be concerned with the easier problem of locating a local minimum point x^* , such that

$$\text{(i)} \quad x^* \in S$$

$$\text{(ii)} \quad f(x^*) \leq f(x) \quad \text{all} \quad x \in N(x^*,\delta) \subset S \; ,(1.2)$$

where $N(x^*,\delta)$ defines a neighborhood of radius δ , centre x^* , i.e., $x \in N(x^*,\delta)$ if $||x - x^*|| < \delta$.

In any practical situation, the investigation of a problem involves many stages before its solution can be attempted on a computer. Let us consider a simple example.

The Machine Tool Problem: Hersom[1]

A machine tool consists of a motor P , which drives the shaft of a rotating cutting tool which rotates with velocity v , and also acts through a gearbox to move the workpiece. This is illustrated in Figure 1.

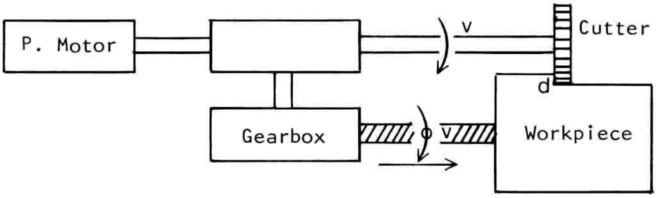


Figure 1

The objective is to cut the required metal from the workpiece as cheaply as possible. To convert this very simple problem into a numerical problem we must first determine the variables x , and express the objective function f and feasible set S in terms of them.

There are three variables in the problem; the rotational velocity v , the horizontal velocity ϕv and the depth of cut d . In this particular problem, the feasible set S is defined by a number of inequality constraints. The gearbox had a limited range -- $.005 \leq \phi \leq .02$. The motor could supply limited power $\phi^{0.8} d^{0.8} v \leq P$, and the cutting shaft snapped if the stress was too high $\phi v^2 \leq S\ max$. For ease of representation, we will assume that the depth of cut is fixed so that the optimization variables $x^T = (\phi, v)$, then

$$x \in S \quad \text{if} \quad .005 \leq \phi \leq .02$$

$$\phi^{0.8} d^{0.8} v \leq P$$

$$\phi v^2 \leq S\ max$$

$$v \geq 0 \quad .$$

Again, the cost of cutting a given volume of material is now a simple function of the variables. If R is the fixed costs with time, T_C is the cost of replacing the cutting tool, and T_L is the life of the tool, then COST = $1/(v\phi d)\ (R + T_C/T_L)$. The life of the tool is a function of its use and for one particular example, $T_L = 1/(d^{0.6}\phi^{1.7}v^3)$. The optimization problem is therefore given by

$$\underset{x \in S}{\text{Min}}\ f(x) = \frac{1}{x_1 x_2 d}(R + T_c d^{0.6} x_1^{1.7} x_2^{3})$$

$$x \in S \quad \text{if} \begin{cases} .005 \le x_1 \le .02 \\ x_1^{0.8} d^{0.8} x_2 \le P \\ x_1 x_2^2 \le S\ \max \\ x_2 \ge 0 \quad . \end{cases}$$

This is illustrated in Figure 2.

2. Descent Methods for Unconstrained Optimization

Let us simplify the problem by assuming that S consists of the complete space E^n ; the problem is then termed the unconstrained optimization problem. We will restrict our investigation into the convergence of algorithms for solving this problem to algorithms that are iterative in nature, i.e., algorithms that generate a sequence of points x^k , $k = 1,2,...$ and hope to terminate in a neighborhood of x^* in a finite time.

We will also restrict our discussions to the class of well behaved functions for minimization, namely those functions for which

(i) there exists a compact bounded level set $f(x) = V\ \max$;
(ii) $f(x)$ is bounded below; and

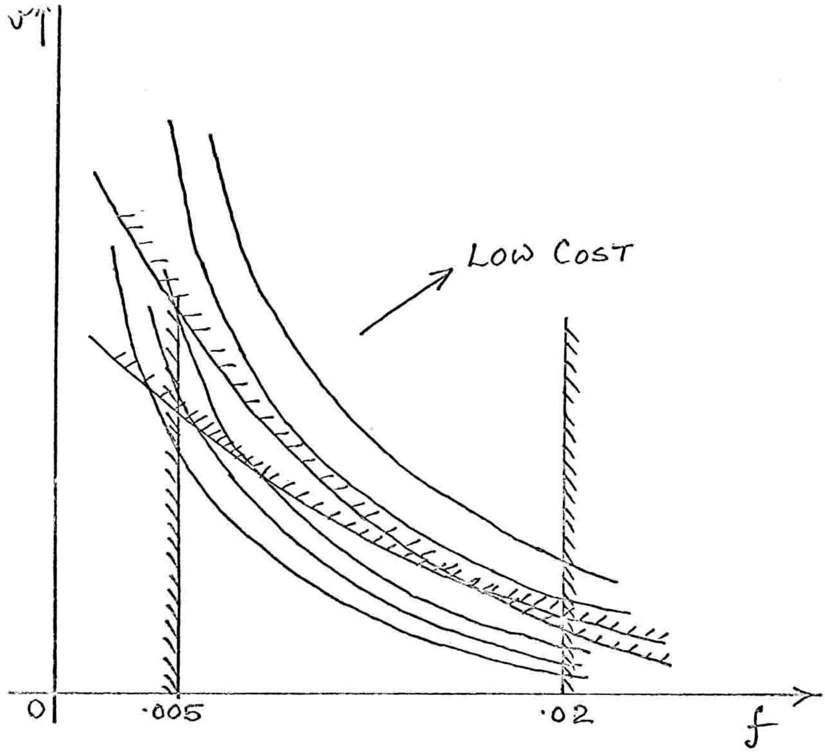


Figure 2

(iii) the Hessian G exists and is uniformly bounded in the set $f(x) \leq V \max$, i.e., $|\frac{1}{2}p^T G p| \leq M$ if $||p|| = 1$ and $f(x) \leq V \max$.

In order to be able to terminate an algorithm when it reaches a neighborhood of x^*, we must be able to recognise such a neighborhood numerically.

Theorem 2.1:

The first order necessary condition for x^* to be a local minimum point in E^n of a differentiable function $f(x)$ is that the gradient vector g or ∇f is zero at x^* where

$$g_i = \nabla f_i = \frac{\partial f}{\partial x_i} .$$

Theorem 2.2:

The second order necessary condition for x^* to be a local minimum point is E^n of a differentiable function $f(x)$ is that the Hessian matrix G or $\nabla^2 f$ is positive (semi-) definite at x^*, where

$$G_{ij} = \nabla^2 f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} .$$

Similarly,

Theorem 2.3:

If $g = 0$ at x^* and G is positive definite at x^*, then x^* is a local minimum point of the function $f(x)$.

These results follow simply from the well-known one dimensional results for a function $\phi(a)$.

Define $\phi(a) = f(x + ap)$ where p is a unit vector and note that

$$\phi'(0) = g^T p \tag{2.1}$$

and

$$\phi''(0) = p^T G p \tag{2.2}$$

and that the one-dimensional results must hold for every possible p .

We may therefore define a neighborhood of a local minimum point by $||g|| \quad E_0$ and agree to terminate an algorithm as having located such a point if

(i) $||g^k|| < E_0$; and

(ii) G^k is positive (semi-) definite. Let us term this stopping rule I (SRI).

If a point is found where $||g^k|| < E_0$ but G^k has some negative eigenvalues, then the algorithm has located a saddlepoint (or maximum).

It is normal to include rules that terminate an algorithm if it is not converging rapidly enough. Conventional additional stopping criteria are:

SRII: $k > k\ max$, an upper limit on the number of iterations, and

SRIII: $||x^k - x^{k-n}|| < E_0$, no effective progress.

The iterated algorithms we will consider will be restricted to those that construct x^{k+1} from x^k by first choosing a search direction p and then a step size a

$$x^{k+1} = x^k + ap \quad . \qquad (2.3)$$

As it is difficult to consider both choices simultaneously, we will first concentrate on the effect of the choice of p and will define a perfect line search as one in which a is chosen to minimize $\phi(a) = f(x^k + ap)$, i.e.,

$$a = \arg\min_a \phi(a) \quad . \qquad (2.4)$$

To clarify the effect of the choice of p , we will first discuss two inefficient algorithms.

3. The Univariate Search Routine

If we select the direction of search parallel to each axis in turn and perform a perfect line search, we can define algorithm A1.

(1) SELECT x^o , $k = 0$, $J = 0$

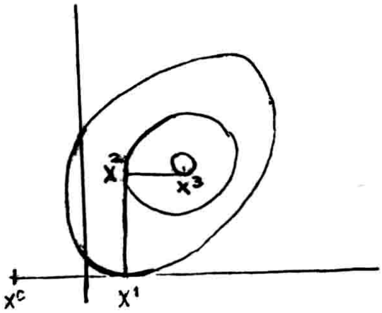
(2) IF $J = N$ SET $J = 0$;
$J = J + 1$, $k = k + 1$

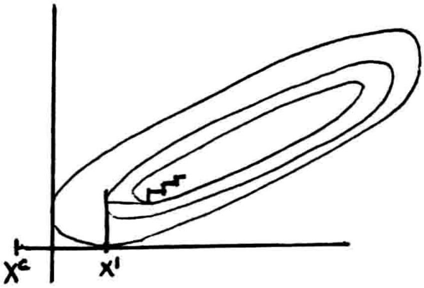
(3) CHOOSE p as unit vector along J^{th} axis

(4) $a = \arg\min f(x^k + ap)$

(5) TEST FOR TERMINATION: STOP or GO TO 2.

We can sketch its performance on two simple functions:





From the sketches we see that

(i) convergence is very rapid on nicely scaled functions, and

(ii) convergence is very poor on badly scaled ridges; indeed the step size gets so small that the search frequently terminates well away from x^* , due to the stopping criterion

$$||x^k - x^{k-n}|| < E_0 .$$

Arithmetic Confirmation

Consider a two-dimensional eclipse centered at the origin

$$f = \tfrac{1}{2}x^T \begin{pmatrix} 1 & a \\ a & b^2 \end{pmatrix} x \;;\; b^2 > a^2$$

$$= \tfrac{1}{2}(x_1^{\,2} + 2a\; x_1 x_2 + b^2 x_2^{\,2}) \quad .$$

Select a starting point $x^{(o)} = (x_1^{(o)}\ ,\ x_2^{(o)})$. Then, assuming perfect line searches,

$$x^1 = (-ax_2^{\,o}\ ,\ x_2^{\,o})$$

$$x^2 = (-ax_2^{\,o}\ ,\ d^2/b^2 x_2^{\,o})$$

and

$$x^3 = (-a^3/b^2 x_2^{\,o}\ ,\ a^2/b^2 s_x^{\,o})$$

implying

$$x^{k+2} = a^2/b^2 x^k \qquad k \geq 1 \quad .$$

This is a <u>linear rate of convergence</u>, which improves for any b as the coupling term $a \to 0$, but becomes arbitrarily poor as $a \to b$.

An even greater disadvantage than this poor rate of convergence is the fact that this algorithm may <u>limit cycle</u>, along a path with nonzero gradient. This was discovered by Powell[2] and the simplest known example is

$$f(x_1,x_2,x_3) = -x_1x_2 - x_1x_3 - x_2x_3 + (x_1 - 1)^2_+ + (-x_1 - 1)^2_+$$

$$+ (x_2 - 1)^2_+ + (-x_2 - 1)^2_+ + (x_3 - 1)^2_+ + (-x_3 - 1)^2_+$$

where

$$(x - c)^2_+ = 0 \quad \text{if} \quad x \leq c$$

$$= (x - c)^2 \quad \text{if} \quad x \geq c \quad .$$

If $x^k = (-1 - E , +1 + \frac{1}{2}E , -1 - \frac{1}{4}E)$, the subsequent steps are:

$$x^{k+1} = (+1 + E/8 , 1 + E/2 , -1 - E/4)$$

$$x^{k+2} = (1 + E/8 , -1 - E/15, -1 - E/4)$$

$$x^{k+3} = (1 + E/8 , -1 - E/16 , 1 + E/32)$$

$$x^{k+4} = (-1 - E/64 , -1 - E/16 , 1 + E/32)$$

$$x^{k+5} = (-1 - E/64 , 1 + E/128 , 1 + E/32)$$

$$x^{k+6} = (-1 - E/64 , 1 + E/128 , -1 - E/256) \quad .$$

As the point x^{k+6} is of the same form as x^k , the cycle repeats and gets closer and closer to the edges of the unit cube (+1,+1,+1), while on these edges

$$|g_1| + |g_2| + |g_3| = 2 , \quad f = -1 \quad .$$

The iteration will therefore never satisfy SRI or SRIII and will depend upon the failsafe measure SRII. The function