Effective treatment of the interpolation factor in Marquardt's nonlinear least-squares fit algorithm

T. Tabata and R. Ito

Radiation Center of Osaka Prefecture, Sakai, Osaka, Japan

A modification of Marquardt's nonlinear least-squares fit strategy is proposed. The performances of the modified and original strategies have been compared for six test problems. The results show that execution times for difficult problems are much reduced by the modification.

(Received June 1973)

The maximum neighbourhood algorithm developed by Marquardt (1963) is one of the most effective techniques for solving the problem of nonlinear least-squares fit. The present note describes a tactical improvement of this algorithm and experimental comparison of the performances of the original and modified strategy by the use of test problems.

In applying Marquardt's algorithm, the method of choosing the value of the parameter λ, which controls the interpolation of the algorithm between the steepest-descent method and Gauss-Newton method, has a considerable effect on the rate of convergence for some classes of problems. According to the recipe of Marquardt's original strategy, λ is varied from one iteration to the next by multiplying or dividing the previous value by the constant factor of 10, that is, the choice is made with a logarithmically constant step. The present strategy uses variable steps for this choice. To change the size of the step, the following five factors are provided in place of the single factor of 10: $v_1 = 1{\cdot}33$ ($\simeq 10^{1/8}$), $v_2 = 1{\cdot}78$ ($\simeq 10^{1/4}$), $v_3 = 3{\cdot}16$ ($\simeq 10^{1/2}$), $v_4 = 10$ and $v_5 = 10^2$. One of these factors v_i is selected for each iteration depending on *history* of minimisation. The *history* is defined here as a sequence of the results of comparing the sum of squared residuals r^2 with its least value so far obtained, and consists of the latest three results except at the earliest iterations. The scheme used in the present strategy to determine the value of the subscript i of the v-factor is given in **Table 1.**

Performances of the original and modified strategies have been compared for six test problems. These problems are symbolised by combinations of the numbers 1-3 and the letters a and b. For each problem, r^2 can be written in the form

$$r^2(\boldsymbol{x}) = [cf_1(\boldsymbol{x})]^2 + [f_2(\boldsymbol{x})]^2 \,,$$

where c is a constant determining the steepness of the side slopes to the curved valley $f_1(\boldsymbol{x}) = 0$. The problems denoted by the same number are different from each other only in the value of c. It is given by

$$\begin{aligned} c &= 10 \quad \text{for problems of type } a \\ &= 100 \quad \text{for problems of type } b \,. \end{aligned}$$

The functions $f_j(\boldsymbol{x})$ ($j = 1, 2$) and the starting point $\boldsymbol{x}_0$ are:

Problem 1

A parabolic valley (Rosenbrock, 1960)

$$\begin{aligned} f_1(\boldsymbol{x}) &= x_2 - x_1^2 \\ f_2(\boldsymbol{x}) &= x_1 - 1 \\ \boldsymbol{x}_0 &= (-1{\cdot}2, 1) \end{aligned}$$

Problem 2

A cubic valley

$$\begin{aligned} f_1(x) &= x_2 - (x_1^3 - x_1) \\ f_2(x) &= x_1 - 1 \\ \boldsymbol{x}_0 &= (-1{\cdot}2, 0) \end{aligned}$$

Problem 3

A circular valley (Ono, 1971)

$$\begin{aligned} f_1(\boldsymbol{x}) &= (x_1 - 1)^2 + x_2^2 - 1 \\ f_2(\boldsymbol{x}) &= x_1 - 2 \\ \boldsymbol{x}_0 &= (0, 1) \,. \end{aligned}$$

For comparing the performances, the number N of equivalent function evaluations (the total number of evaluations of the function and its derivatives; Box, 1966) and the execution time t are shown in **Table 2.** The values listed are those necessary to reduce r^2 to 10^{-5}. The derivatives were computed from analytical expressions. The computer used is a TOSBAC 3400 Model 51, whose operation times for floating point numbers are: 6·3 μsec for addition and subtraction, 14·3 μsec for multiplication and 18·5 μsec for division.

As can be seen from Table 2, the modified strategy requires appreciably smaller values of N and t than the original strategy except for problem 2a. The ineffectiveness of the present modification for this problem is due to the fact that in this case alternations of increases and decreases of the optimum value of λ occur rather rapidly along the minimisation path. In the problems of type b the value of λ can be decreased only very slowly, and the modified strategy shows excellent results for the problems of this kind, reducing the execution time by a factor of about 2-5.

Table 1 Scheme for determining the value of i of the v-factor from the previous value i_0 (D and I stand for the decrease and increase, respectively, of r^2, and DI, for example, means that r^2 was decreased and then increased at the latest two iterations.)

Conditions		
History	i_0	i
At the start	...	4
DI, ID	...	$i_0 - 1$
DDI, IDI, IID	$i_0 > 1$	$i_0 - 1$
DDD	$i_0 < 5$	$i_0 + 1$
III	$i_0 < 3$	3
All the other cases		i_0

Table 2 Comparison of the performances of the original and modified strategies

	Original strategy		*Modified strategy*	
Problem	N	t (sec)	N	t (sec)
1a	84	5·6	70	4·9
2a	70	4·5	79	5·5
3a	201	14	69	4·4
1b	528	35	173	12
2b	660	43	281	20
3b	1652	111	334	23

For the purpose of checking the effectiveness of the smallest factor of v_i, i.e. v_1, another modified strategy in which the use of v_1 is omitted has also been tested. The results have shown that the performance without v_1 is worse for all the problems other than problem 1*a*.

From these comparisons, it can be concluded that the modification with the five v-factors leads to an appreciable improvement of the maximum neighbourhood algorithm.

References

BOX, M. J. (1966). A comparison of several current optimization methods, and the use of transformations in constrained problems, *The Computer Journal*, Vol. 9, No. 1, pp. 67-77.

MARQUARDT, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters, *JSIAM*, Vol. 11, No. 2, pp. 431-441.

ONO, K. (1971). Private communication.

ROSENBROCK, H. H. (1960). An automatic method for finding the greatest or least value of a function, *The Computer Journal*, Vol. 3, No. 3, pp. 175-184.

Book review

Information Systems, COINS IV, edited by Julius T. Tou, 1974; 506 pages. (*Plenum Publishing Corporation*, $39.00.)

This book contains the proceedings of the fourth in a series of conferences held in the USA. COINS is an acronym for Computer and Information Sciences, which most will agree is a very broad topic. The theme of this conference held in late 1972 in Miami Beach was 'Information Systems'—still a rather broad field. The specific topics discussed at the conference included data base management, software development, information retrieval, pattern recognition, data analysis and urban information systems.

The book includes 24 papers, of which eight could be classified as dealing with generalised data base management topics, and of those, three discussed relational models. The quality of these eight papers is good, but most are definitely dealing with more research oriented topics in data base management. The claim on the flyleaf that the book is 'an invaluable source book for corporate management . . .' is defendable only for the paper by Everest entitled 'The Objectives of Database Management'. (Everest is the only author to prefer the germanism 'database' rather than the original 'data base'.) His paper is very readable and a useful and clear exposition of what data base management is all about.

The three relational papers include yet another tutorial on the topic, this time by C. J. Date, and two papers from authors who have attempted to implement the concepts. The first is by Whitney of General Motors Research and the other by three authors from IBM Sweden. This ends with a lament: 'to provide both advanced functions and reasonable performance is not an easy task'. True, but who should control the trade-off between the two?

An interesting, but hard to read, paper by Hardgrave addresses the problem of 'a retrieval language for tree-structured data base systems'. The author has been involved with System 2000 at the University of Texas at Austin and addresses some problems—unfortunately not widely recognised—of evaluating boolean expressions against tree-structures such as those in TDMS (and its grandchild System 200). Anyone trying to advance the state of the art in non-procedural languages bumps into this one sooner or later.

Another valuable paper is that by Sibley and Merten entitled 'Transferability and Translation of Programs and Data'. The title speaks for itself, but again the level of exposition is not for 'corporate management'.

Leaving data base management and proceeding to the rest of the book, there are three papers on the subject of document retrieval. One of these, by Crouch and Crouch, is analytic in that it looks at four 'information storage and retrieval systems' of mid-sixties vintage and assesses how they fit into a general framework. Such work is valuable and ought to be undertaken more often. So many people would rather build their own system than take the time to analyse other peoples.

A paper on management information systems in general is entitled 'EMISARI: A Management Information System Designed to Aid and Involve People'. It is written by six authors from an ominous sounding organisation called the Office of Emergency Preparedness. It is claimed that the system 'represents a major departure from conventional MIS design. It is oriented not toward data per se but rather toward the activities of the people who generate and use the data'. Studying the paper, it is not clear that the approach is so very special. It does not use a generalised DBMS and is hence a tailored system.

There are many other papers in the proceedings and different papers will interest different readers. This is the kind of book which should find a place in a university departmental library, but not many individual buyers will wish to invest $39 (US price) to have it on their own shelves.

T. W. OLLE (West Byfleet)