A SUPERLINEARLY CONVERGENT ALGORITHM FOR MINIMIZATION WITHOUT EVALUATING DERIVATIVES *

Robert MIFFLIN

Yale University, New Haven, Conn., U.S.A.

Received 27 March 1974
Revised manuscript received 5 December 1974

An algorithm for unconstrained minimization of a function of n variables that does not require the evaluation of partial derivatives is presented. It is a second order extension of the method of local variations and it does not require any exact one variable minimizations. This method retains the local variations property of accumulation points being stationary for a continuously differentiable function. Furthermore, because this extension makes the algorithm an approximate Newton method, its convergence is superlinear for a twice continuously differentiable strongly convex function.

1. Introduction

We present an algorithm for unconstrained minimization of a real-valued function $f$ defined on $\mathbb{R}^n$ that does not require the evaluation of partial derivatives of $f$. The algorithm is partly an approximate Newton method where both first and second order partial derivatives are approximated from function values and partly a method of location variations [3,18] which uses a subset of these same function values. For all of our convergence results we assume $f$ is bounded from below and continuously differentiable on $\mathbb{R}^n$ and we say that a point is stationary if the gradient of $f$ evaluated at the point is zero. The inclusion of Chernous’ko’s [3] local variations idea as given in [18] allows us to show from any starting point that any accumulation point of the algorithm sequence is stationary without assuming twice differentiability and/or convexity of $f$. When we assume $f$ is twice continuously differentiable with a positive definite matrix of second partial derivatives on

then we show that the algorithm's rate of convergence is *superlinear*, i.e.,
\[ \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \to 0 \quad \text{as } k \to \infty, \]
where \( \{x^k\} \subset \mathbb{R}^n \) is the algorithm sequence and \( x^* \in \mathbb{R}^n \) minimizes \( f \). If, in addition, the second partial derivatives of \( f \) are Lipschitz continuous, then the convergence is of *order 2*, i.e.,
\[ \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^2} \]
is bounded above for all \( k \). For these convergence rate proofs we employ ideas and results on superlinear convergence from Goldstein and Price [12] and McCormick and Ritter [17].

Other interesting nonderivative algorithms for unconstrained optimization have been proposed by Powell [20], Stewart [22], Zangwill [24], Fiacco and McCormick [7; pp. 175–178], Cullum [4], Brent [2], Greenstadt [14], Gill, Murray and Pitfield [10] and Winfield [23]. To our knowledge, no theoretical convergence rate results have been given for these methods and the only ones that have convergence proofs, other than finiteness results for strictly convex quadratic functions, are those of Cullum and Zangwill. The proofs in Cullum [4] and Zangwill [24] assume strict convexity of \( f \) as well as continuous differentiability and their theoretical algorithms, unlike ours, require some exact one-dimensional optimization subproblems, which is something that cannot be implemented. Daniel [5; pp. 206–207] has also given a convergence result for Zangwill's algorithm under somewhat weaker conditions.

In the next two sections we state and discuss the algorithm. In Sections 4 and 5 we prove stationarity and superlinear convergence, respectively, and in the concluding section we comment on work in progress.

2. The algorithm

Let \( e_i \) be the \( i \)th unit vector in \( \mathbb{R}^n \). The algorithm parameters required are positive real numbers \( \alpha, \beta, \gamma, \delta \) and \( \rho \) with \( \rho < 1 \) and \( \beta^2 < (\rho/(2n^2\gamma)) \). The following procedure due to Gill and Murray [8] is required by step 3 of the algorithm. It uses the parameter \( \delta \) which is a small positive number related to the word length of the computer being used and is chosen to avoid numerical problems resulting from division by small numbers.

*Modified Cholesky Factorization Procedure* [8]

Given an \( n \times n \) symmetric matrix \( H \) and a positive number \( \delta \), this