METHODS OF CONJUGATE DIRECTIONS
VERSUS QUASI-NEWTON METHODS

Garth P. McCORMICK *
The George Washington University, Washington, D.C., U.S.A.

and

Klaus RITTER
Rutgers University, New Brunswick, N.J., U.S.A.

Received 29 May 1971
Revised manuscript received 22 January 1972

It is shown that algorithms for minimizing an unconstrained function \( F(x), x \in \mathbb{E}^n \), which are solely methods of conjugate directions can be expected to exhibit only an \( n \) or \((n-1)\) step superlinear rate of convergence to an isolated local minimizer. This is contrasted with quasi-Newton methods which can be expected to exhibit every step superlinear convergence. Similar statements about a quadratic rate of convergence hold when a Lipschitz condition is placed on the second derivatives of \( F(x) \).

1. Introduction

It is the intent of this paper to make some statements about the relative merits of solving the problem

\[
\min F(x), \quad x \in \mathbb{E}^n,
\]

using methods of conjugate directions and quasi-Newton methods. The former methods have a long history [7] and have the property that if \( F(x) \) is a positive definite quadratic form, the global unconstrained minimizer can be located in \( n \) or fewer steps. Their use on general problems is based on the intuitive notion that in a neighborhood of

* Research was supported in part by Army Research Office, Contract Number DAHC 19-69-C-0017 and the Office of Naval Research, Contract Number N00014-71-C-0116 (NR 047-99).
an isolated local minimizer with a positive definite Hessian matrix, the function $F(x)$ behaves like a positive definite quadratic form and the rate of convergence to the local minimizer should be "fast". To the authors' knowledge, there is no paper available which makes rigorous these concepts. In fact, there is no definition of a method of conjugate directions which does not make the assumption that $F(x)$ is a quadratic function. To this end a definition of a method of conjugate directions is given. Using this definition, a theorem is proved showing the rate of convergence of conjugate direction algorithms. Next, the rate of convergence of quasi-Newton methods is established. The last section contains a comparison of the two types of methods and argues that quasi-Newton methods are superior.

2. Convergence theorems

Let $x \in \mathbb{E}^n$ and $F(x)$ be a real function. If $F(x)$ is differentiable at $x_i$ we denote its gradient by $\nabla F(x_i)$ or $g_j$, if $F(x)$ is twice differentiable at $x_i$ we denote the Hessian matrix of $F(x)$ at $x_i$ by $G(x_i)$ or $G_j$.

Suppose an algorithm for the minimization of $F(x)$ is given that generates a sequence of points $\{x_j\}$, a sequence of directions $\{s_j\}$, and a sequence of step sizes $\{\sigma_j\}$.

The algorithm is called a method of conjugate directions if the following properties hold:

(1) At each iteration $j$ the new point $x_{j+1}$ is obtained as

$$x_{j+1} = x_j - \sigma_j s_j$$

where $\sigma_j$ is the smallest local minimizer to the one-dimensional "step-size" problem:

$$\min_{\sigma \geq 0} F(x_j - \sigma s_j).$$

(2) If the sequence $\{x_j\}$ converges to a point $z$ such that $\nabla F(z) = 0$, $F(x)$ is twice continuously differentiable in a neighborhood of $z$ and $G(z) = G$ is positive definite then, for all $j$,

$$\|s_j\| = O(\|g_j\|), \quad \|g_j\| = O(\|s_j\|)$$