MINIMIZATION BY INTERPOLATION: 
A SECOND ORDER GRADIENT ALGORITHM 

by 
J. Barzilai 

November 1980 

This research was partly supported by ONR Contract N00014-75-C-0569 
with the Center for Cybernetic Studies, The University of Texas at 
Austin. Reproduction in whole or in part is permitted for any purpose 
of the United States Government.
Newton's method for finding a stationary point of $f: \mathbb{R}^n \to \mathbb{R}$ consists of the iteration

$$x_{i+1} = x_i - \left[ \nabla^2 f(x_i) \right]^{-1} \cdot \nabla f(x_i).$$

Its main attraction is its second order of convergence. However, it necessitates computation and inversion of the second order derivatives matrix.

Common minimization algorithms approximate the Hessian or its inverse by first order (i.e. gradient) information. First order information algorithms in common use, have at best superlinear rate of convergence [cf. 2].

We present a new class of algorithms which use first order information only, while maintaining quadratic convergence.

At step $i$ of the algorithm, we interpolate $f$ by a suitable interpolating function $T$, requiring

$$\begin{align*}
T(x_{i-j}) &= f(x_{i-j}) \\
\nabla T(x_{i-j}) &= \nabla f(x_{i-j}),
\end{align*}$$

and determine $x_{i+1}$ as a solution of the equation

$$\nabla T(x_{i+1}) = 0.$$

We assume that the interpolating function depends on some parameters. We further assume that the equations (1) for the parameters of $T$, and equation (2) for $x_{i+1}$ have solutions for all $i$, and that the parameters of $T$ depend on the data continuously through (1). Finally, we assume that $f$ and $T$ have continuous derivatives of order 5 near the solution.
We derive the rate of convergence of the algorithm defined by (1) and (2) by establishing a difference relation for the errors $e_i = \|x_i - x^*\|$. Here $\|\cdot\|$ is an arbitrary fixed norm. This difference relation is analogous to the one obtained in [1] for the one-dimensional case.

To this end we define a function $\Psi : \mathbb{R} \rightarrow \mathbb{R}^n$ in terms of which and the functions $f, T$ we can express the errors of a related one-dimensional interpolation problem.

We assume that a point $x^* \in \mathbb{R}^n$ which is a solution of

$$\nabla f(x^*) = 0$$

exists. Let $\Psi : \mathbb{R} \rightarrow \mathbb{R}^n$ be a curve in $\mathbb{R}^n$ through the points $x_{i-j} j = 1, 0, -1$ and $x^*$, i.e.,

$$\begin{cases}
\psi(t_{i-j}) = x_{i-j} , & j = 1, 0, -1 , \\
\psi(t^*) = x^* ,
\end{cases}$$

(4)

where the parameter $t$ is chosen so that

$$t_{i-j} = \|x_{i-j} - x^*\| , \quad t^* = \|x^* - x^*\| = 0 .$$

We will later discuss the existence of this construction. Note, however, that the construction of $\Psi$ is a part of the analysis of the properties of the algorithm, not a part of the algorithm itself.

Now define $\Theta(t) = T(\psi(t)), \phi(t) = f(\psi(t))$. Equations (1) and (4) imply

$$\Theta^{(k)}(t_{i-j}) = \phi^{(k)}(t_{i-j}) , \quad j, k = 0, 1 ,$$

which in turn implies [see 3]
where \( \eta \) is some intermediate point. Equation (10) is the basic difference relation we need (cf. [1]). Differentiating it and setting \( t=0 \), we obtain

\[
(7) \quad t_{i+1} = B_i t_i^2 - t_{i-1}^2.
\]

If the sequence \( B_i \) converges to a non-zero limit, the relation (7) implies that the sequence \( t_i \) converges to zero if \( t_0, t_1 \) are small enough, with rate of convergence which is given by the unique positive root of the indicial polynomial of (7): \( t^2 - t - 2 = 0 \), i.e. quadratically (cf. 4).

In order for the sequence \( B_i \) to converge, it is sufficient that \( \theta(5) \) and \( \phi(5) \) exist and are continuous, and \( \phi''(0) \neq 0 \). This would be the case if \( f \) has continuous derivatives of order 5 near the solution, the parameters of \( T \) depend continuously on the data, and \( T \) has continuous derivatives of order 5 for the appropriate values of the parameters. Finally, it is evident that the curve \( \psi \) can be chosen so that \( \phi''(0) = \psi' \nabla^2 f \psi \) is nonzero, e.g. by choosing

\[
\psi_k = a_k t^4 + b_k t^3 + c_k t^2 + t + d_k, \quad k = 1, \ldots, n.
\]

Note that no line search is needed in this class of algorithms, and that they may be designed to locate saddle points rather than minimum points.

A useful choice for the interpolating function \( T \) seems to be a separable sum of rational functions of the type discussed in [1].

The results in [1] for the one-dimensional case, can clearly be extended by the same device to the n-dimensional case. In particular, algorithms based on
function values only, have rates of convergence between 1.3 and 1.6; the rate of convergence is independent of the interpolating function, and inverse interpolation can be utilized for minimization. Similar results hold for the root-finding problem discussed by Traub [5]. Details of this work will appear elsewhere.
REFERENCES


Minimization by Interpolation: A Second Order Gradient Algorithm

Newton's method for finding a stationary point of \( f : \mathbb{R}^n \to \mathbb{R} \) has second order of convergence. However, it necessitates computation and inversion of the second order derivatives matrix.

We present new classes of algorithms. One of these has second order of convergence while using first order information only.
END
DATE FILMED
7-80
DTIC