AN EXAMPLE OF IRREGULAR CONVERGENCE IN SOME CONSTRAINED OPTIMIZATION METHODS THAT USE THE PROJECTED HESSIAN

Richard H. BYRD
Department of Computer Science, University of Colorado, Boulder, CO 80309, USA

Received 16 July 1984
Revised manuscript received 6 September 1984

In this paper we give examples illustrating the behavior of the Coleman-Conn horizontal vertical method and of successive quadratic programming with a Hessian approximation exact on the tangent space of the constraints. One example shows that these methods in general are not one-step superlinearly convergent.

Key words: Constrained Optimization, Reduced Hessian, Two-step convergence

Many methods for solution of optimization problems with nonlinear constraints make use of the Hessian of the Lagrangian to obtain fast local convergence. Several of these methods including those of Coleman and Conn (1982) make use of approximations to the reduced Hessian. In addition Powell (1978) analyzes a method where the Hessian approximation is accurately only on the null space of the constraint derivatives. These authors show that under reasonable conditions these methods are two-step superlinearly convergent; i.e. the sequence consisting of every other iterate is superlinearly convergent.

An obvious question is whether these methods are one-step superlinearly convergent. In this paper we give an example showing that neither the method of Coleman and Conn, nor successive quadratic programming with an accurate Hessian on that subspace is superlinearly convergent.

In order to do this we will now describe these two methods applied to a problem of the form

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad c(x) = 0
\end{align*}$$

(1)

where $f$ is a real-valued function on $\mathbb{R}^n$ and $c$ maps $\mathbb{R}^n$ to $\mathbb{R}^m$. We assume both functions are twice differentiable.

The method of successive quadratic programming (SQP), at an iterate $x_k$, has the form:

This research was supported by National Science Foundation Grant MCS-8115475.
Given \( x_k \) let \( d_k \) be the solution to
\[
\begin{align*}
\text{minimize} \quad & \nabla f(x_k)^T d + \frac{1}{2} d^T B_k d \\
\text{subject to} \quad & \nabla c(x_k)^T d = -c(x_k).
\end{align*}
\]

Then let
\[ x_{k+1} = x_k + d_k. \]

Here \( B_k \) is an \( n \times n \) matrix approximating the Hessian of the Lagrangian,
\[ \nabla^2 L(x_k, \lambda_k) = \nabla^2 f(x_k) + \sum \lambda_i \nabla^2 c_i(x_k). \]

It should be noted that if \( B_k \) is the exact Hessian and if \( \lambda \) is the vector of Lagrange multipliers to the quadratic program at \( x_{k-1} \) then this is just Newton's method on the Kuhn–Tucker conditions for problem (1).

To discuss convergence of this method, let the orthogonal projection matrix onto the null space of the constraint derivatives be denoted by
\[ P_k = I - \nabla c(x_k)(\nabla c(x_k)^T \nabla c(x_k))^{-1} \nabla c(x_k)^T, \]
and let \((x_*, \lambda_*)\) denote the Kuhn–Tucker pair for problem (1). It has been shown by Boggs, Tolle and Wang (1982) and, without the assumption of linear convergence, by Fontecilla, Steihaug, and Tapia (1983) that the sequence \( x_k \) generated by SQP is \( Q \)-superlinearly convergent if and only if
\[ \frac{\| P_k (B_k - \nabla^2 L(x_*, \lambda_*))(x_{k+1} - x_k) \|}{\| (x_{k+1} - x_k) \|} \] converges to zero.

In the case when we only know \( P_k \nabla^2 L(x_*, \lambda_*) P_k \) accurately, that is we have second derivative information only on the null space of the constraint derivatives, this result is weakened. Powell (1978) shows that, under the assumption of convergence, if the condition
\[ \frac{\| P_k (B_k - \nabla^2 L(x_*, \lambda_*)) P_k (x_{k+1} - x_k) \|}{\| (x_{k+1} - x_k) \|} \rightarrow 0 \] holds then the sequence is two-step superlinearly convergent; that is
\[ \frac{\| x_{k+1} - x_k \|}{\| x_{k-1} - x_k \|} \rightarrow 0. \]

We now consider the horizontal-vertical algorithm of Coleman and Conn. Following the usual notation let \( Z_k \) be a matrix of orthogonal columns spanning the null space of \( \nabla c(x_k)^T \), and \( Y_k \) be a matrix of orthogonal columns spanning the space of \( \nabla c(x_k) \). Of course \( Y_k \) and \( Z_k \) are not uniquely determined by \( \nabla c(x_k) \), but they may be easily computed from a QR factorization of \( \nabla c(x_k) \).