Journal of Computational Mathematics, Vol.13, No.2, 1995, 156-160.

A MODIFIED BFGS FORMULA MAINTAINING POSITIVE DEFINITENESS WITH ARMIJO-GOLDSTEIN STEPLENGTHS*

I. D. Coope
(*Department of Mathematics & Statistics, University of Canterbury, New Zealand*)
C. J. Price
(*Defence Scientific Establishment, Devonport Naval Base, Auckland , New Zealand*)

Abstract

The line search subproblem in unconstrained optimization is concerned with finding an acceptable steplength satisfying certain standard conditions. The conditions proposed in the early work of Armijo and Goldstein are sometimes replaced by those recommended by Wolfe because these latter conditions automatically allow positive definiteness of some popular quasi-Newton updates to be maintained. It is shown that a slightly modified form of quasi-Newton update allows positive definiteness to be maintained even if line searches based on the Armijo–Goldstein conditions are used.

1. Introduction

A line search method for minimizing a real function f generates a sequence $x_1, x_2, \ldots$ of points by applying the iteration

$$x_{k+1} = x_k + \alpha_k p_k, \quad k = 1, 2, \ldots . \tag{1}$$

In a quasi-Newton method the search direction p_k is chosen so that $B_k p_k = -g_k$, where B_k is (usually) a positive definite matrix and g_k denotes $\nabla f(x_k)$. For the BFGS update, (see [**?**], for example), the matrices B_k are defined by the formula

$$B_{k+1} = \left[B - \frac{Bss^T B}{s^T Bs} + \frac{yy^T}{s^T y} \right]_k , \tag{2}$$

where $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$. It is well known that if B_1 is positive definite and

$$s_k^T y_k > 0 \tag{3}$$

then all matrices $B_{k+1},\ k=1, 2, \ldots$ generated by (**??**) are positive definite. Thus p_k is a direction of descent provided only that $g_k \neq 0$. The choice of steplength, α_k, in (**??**) is crucial if the line search algorithm is to have good convergence properties. One of

* Received February 26, 1994.

the early recommendations, due to Armijo[?] and Goldstein[?], is to choose $\alpha_k > 0$ at each iteration to satisfy the conditions

$$\sigma_2 \alpha_k p_k^T g_k \leq f(x_{k+1}) - f(x_k) \leq \sigma_1 \alpha_k p_k^T g_k, \tag{4}$$

where $0 < \sigma_1 < \frac{1}{2} < \sigma_2 < 1$, (often $\sigma_2 = 1 - \sigma_1$ and $\sigma_1 = .1$ are recommended). These conditions ensure that the steplength is neither too small nor too large, and under some extra (mild) assumptions on f and the descent direction p_k the limit

$$\lim_{k \to \infty} \nabla f(x_k) = 0 \tag{5}$$

is guaranteed. Thus any limit points of the sequence $\{x_k\}$ are necessarily stationary points of f. Unfortunately, satisfaction of the Armijo/Goldstein conditions (**??**) does not automatically imply that condition (**??**) is also satisfied so that the BFGS update (**??**) may not maintain positive definiteness. In such cases, the updating of B_k could be omitted, or the line search could be continued with extra function and gradient evaluations being made until both (**??**) and (**??**) are satisfied.

An alternative to (**??**) is to use the line search conditions of Wolfe[?, ?] which require the steplength $\alpha_k > 0$ to satisfy the inequalities

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq \rho_1 \alpha_k p_k^T g_k, \\ p_k^T g_{k+1} &\geq \rho_2 p_k^T g_k, \end{aligned} \tag{6}$$

where $0 < \rho_1 < \frac{1}{2}$ and $\rho_1 < \rho_2 < 1$. Note that if $\rho_1 = \sigma_1$ then (**??**) is the same as the right hand inequality of (**??**). Again the purpose of these conditions is to ensure that the steplength is neither too large nor too small. However, condition (**??**) implies that

$$s_k^T y_k \geq (\rho_2 - 1) s_k^T g_k > 0 \tag{7}$$

so that inequality (**??**) is automatically satisfied and the BFGS updating formula can be applied with positive definiteness being maintained automatically. A disadvantage is that to test condition (**??**) requires an extra gradient evaluation at each trial value for α_k.

2. The Modified Updating Formula

The line search conditions (**??**) do allow positive definiteness to be maintained if the updating formula (**??**) is adjusted slightly. Note first that an estimate of the second directional derivative, $p_k^T [\nabla^2 f(x_k)] p_k$, is available from the quadratic polynomial, $q_k(\alpha)$, interpolating the data $q_k(0) = f(x_k)$, $q_k(\alpha_k) = f(x_{k+1})$, and $q_k'(0) = p_k^T g_k$. Thus

$$q_k(\alpha) = f(x_k) + \alpha p_k^T g_k + \tfrac{1}{2} \alpha^2 D_k \tag{8}$$

where

$$D_k = 2[(f(x_{k+1}) - f(x_k))/\alpha_k - p_k^T g_k]/\alpha_k = q_k''(\alpha). \tag{9}$$