A Modified BFGS Formula Maintaining Positive Definiteness with Armijo-Goldstein Steplengths[1]

by

I.D. Coope and C.J. Price
Department of Mathematics and Statistics,
University of Canterbury, Christchurch, New Zealand

No. 112 August, 1994

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.

Keywords: Quasi-Newton, line search, positive definiteness.

[1] Submitted to Journal of Computational Mathematics

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 [2], for example), the matrices B_k are defined by the formula

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

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 (2) 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 (1) is crucial if the line search algorithm is to have good convergence properties. One of the early recommendations, due to Armijo[1] and Goldstein[4], 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 (4) does not automatically imply that condition (3) is also satisfied so that the BFGS update (2) 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 (3) and (4) are satisfied.

An alternative to (4) is to use the line search conditions of Wolfe[5, 6] which require the steplength $\alpha_k > 0$ to satisfy the inequalities

$$f(x_{k+1}) - f(x_k) \leq \rho_1 \alpha_k p_k^T g_k, \tag{6}$$

$$p_k^T g_{k+1} \geq \rho_2 p_k^T g_k, \tag{7}$$

where $0 < \rho_1 < \frac{1}{2}$ and $\rho_1 < \rho_2 < 1$. Note that if $\rho_1 = \sigma_1$ then (6) is the same as the right hand inequality of (4). Again the purpose of these conditions is to ensure that the steplength is neither too large nor too small. However, condition (7) implies that

$$s_k^T y_k \geq (\rho_2 - 1) s_k^T g_k > 0 \tag{8}$$

so that inequality (3) is automatically satisfied and the BFGS updating formula can be applied with positive definiteness being maintained automatically. A disadvantage is that to test condition (7) requires an extra gradient evaluation at each trial value for α_k.

2 The Modified Updating Formula

The line search conditions (4) do allow positive definiteness to be maintained if the updating formula (2) 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

$$q_k(\alpha) = f(x_k) + \alpha p_k^T g_k + \tfrac{1}{2}\alpha^2 D_k \tag{9}$$

where

$$D_k = 2[(f(x_{k+1}) - f(x_k))/\alpha_k - p_k^T g_k]/\alpha_k = q_k''(\alpha). \tag{10}$$

Clearly, if the steplength α_k satisfies the conditions (4) then

$$D_k \geq 2(\sigma_2 - 1)\frac{p_k^T g_k}{\alpha_k} > 0,$$

which is consistent with a convex quadratic function along the ray $x_k + \alpha p_k$. If $f(x)$ were a quadratic function then $D_k \equiv s_k^T y_k/\alpha_k^2$ would hold and in this case the Armijo-Goldstein conditions (4) would be equivalent to the Wolfe conditions (6, 7) if

$$\rho_1 = \sigma_1, \quad \rho_2 = 2\sigma_2 - 1. \tag{11}$$

On quadratic functions the difference in slopes in moving from x_k to x_{k+1} along the direction p_k is $q_k'(\alpha_k) - q'(0) = \alpha_k D_k = \Delta_k$ where

$$\Delta_k = 2[(f(x_{k+1}) - f(x_k))/\alpha_k - p_k^T g_k] \tag{12}$$

and this difference is also given by $\Delta_k \equiv p_k^T y_k$ on quadratics. However, because $p_k^T y_k$ may be non-positive on more general functions the following modification to the standard BFGS formula may be used when the Armijo-Goldstein line search conditions are used. Let

$$z_k = \left[y + \left(\frac{\Delta - p^T y}{p^T p} \right) p \right]_k , \tag{13}$$

so that $p_k^T z_k = \Delta_k$. Then apply the standard BFGS update with z_k replacing y_k:

$$B_{k+1} = \left[B - \frac{B s s^T B}{s^T B s} + \frac{z z^T}{s^T z} \right]_k . \tag{14}$$

Positive definiteness is now maintained because $s_k^T z_k = \alpha_k p_k^T z_k = \alpha_k \Delta_k = \alpha_k^2 D_k > 0$. Moreover, the updating formula (14) is equivalent to (2) when the objective function is a strictly convex quadratic function.

3 Discussion

Inspection of the formula (13) for z_k reveals that $z_k = [I - (pp^T/p^T p)_k] y_k + (\Delta/p^T p)_k p_k$. Thus the difference in gradients that may be inconsistent with convexity is 'projected out' and replaced by information consistent with positive curvature along the direction p_k. Clearly, this modification need not be made if $s_k^T y_k > 0$. Moreover, on strictly convex quadratic functions, if the parameters ρ and σ are related through (11) then in exact arithmetic there will be no difference in algorithms using formula (14) in place of (2) provided that the same sequence of trial values is used in each line search. On more general functions differences will occur and it should be expected that these will be most apparent in the early iterations at points remote from the solution. Close to the solution f can be approximated well by a positive definite quadratic function so that there should be little difference in the two sets of line search conditions.

Limited numerical trials were performed using the following algorithm:

1. Initialization. $k = 1$. $B_k = I$.

2. Calculate $p_k = -B_k^{-1} g_k$.

3. Perform the line search, $x_{k+1} = x_k + \alpha_k p_k$, and then update B_k.

4. If the stopping conditions are not met, increment k and goto 2.

For the Wolfe conditions, the line search used was that given on pp. 34-35 of [3] with $\overline{f} = -\infty$, $\tau_1 = 4$, and $\tau_2 = \tau_3 = 0.5$. For the Armijo-Goldstein conditions, the sequence $f(x_k + \alpha_n p_k)$ is calculated for $\alpha_n = 4^n$ using $n = 0, 1, 2, \ldots$ until either (4) is satisfied, or a value α_m violating the right hand inequality in (4) is found. In the latter case an α value satisfying (4) was then found using bisection on the interval $[\alpha_{m-1}, \alpha_m]$. The general theorem on descent methods for unconstrained optimization of general functions (see theorem 2.5.1 of [3] for example) is applicable provided each B_k is positive definite, irrespective of which update is used at each iteration.

The performance of the algorithm using the Wolfe line search conditions (6), (7), with formula (2) and the Armijo-Goldstein conditions (4) with formula (14) were very similar. In the trials the parameter values $\sigma_1 = 1 - \sigma_2 = .1$ and $\rho_1 = .1$, $\rho_2 = .8$ were chosen so that the conditions (11) were stisfied. This ensured that both algorithms would behave identically on quadratics. On non-quadratics there were distinct differences. Sometimes one algorithm performed better and sometimes the other, but generally there was very little to choose between the two with perhaps a slight preference in favour of the Wolfe conditions. Thus it seems better to use the unmodified form of the BFGS update unless there are pressing reasons for preferring otherwise. One such situation when the Armijo-Goldstein conditions may be preferred is when gradient information has to be estimated by finite differences. Then the second Wolfe condition (7) is very expensive to test if more than one trial point is required in the line search. In this situation a line search based on parabolic interpolation (see [3], for example) combined with the Armijo-Goldstein conditions becomes much more attractive. If $s_k^T y_k > 0$ then the unmodified update can be used, otherwise it may be preferable to use the modified update (13), (14) rather than having to abandon the update altogether.

The modified BFGS formula in this paper automatically satisfies the "weak quasi-Newton condition"

$$s_k^T B_{k+1} s_k = 2[f(x_{k+1}) - f(x_k) - s_k^T \nabla f(x_k)],$$

and it is particularly convenient when using a line search based on Goldstein/Armijo conditions because no *extra* precautions need to be taken to maintain positive definiteness. Other ways of modifying the BFGS formula have been considered, for example, by Yuan[7]. In his approach the updated matrix satisfies the weak quasi-Newton condition

$$s_k^T B_{k+1} s_k = 2[f(x_k) - f(x_{k+1}) + s_k^T \nabla f(x_{k+1})],$$

and he includes some numerical evidence which favours the use of a modified BFGS formula when Wolfe condition line searches are used. Thus the present paper both supports and complements the work of Yuan.

Acknowledgement

The authors are grateful to an anonymous referee for comments and suggestions leading to an improved version of this paper.

References

[1] L. Armijo, *Minimization of functions having Lipschitz continuous first partial derivatives*, Pacific Journal of Mathematics, Vol. 16, pp. 1-3, 1966.

[2] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall Series in Computational Mathematics, 1983.

[3] R. Fletcher, *Practical Methods of Optimization*, 2nd Edition, John Wiley & Sons, New York, 1987.

[4] A.A. Goldstein, *On steepest descent*, SIAM journal on Control, Vol. 3, pp. 147-151, 1965.

[5] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Review, Vol. 11, pp. 226-235, 1969.

[6] P. Wolfe, *Convergence conditions for ascent methods. II: Some corrections*, SIAM Review, Vol. 13, pp. 185-188, 1971.

[7] Y. Yuan, *A modified BFGS algorithm for unconstrained optimization*, IMA J. Numer. Anal. Vol. 11, pp. 325-332, 1991.

Department of Mathematics & Statistics,
University of Canterbury, Private Bag 4800
Christchurch, NEW ZEALAND
INTERNET: i.coope@math.canterbury.ac.nz

Defence Scientific Establishment,
Devonport Naval Base,
Auckland, New Zealand.
INTERNET: chrisp@dse.govt.nz