A brief note on the Symmetric Rank-1 secant formula

Working note RAL-NA-2006–1 — Nicholas I. M. Gould

22nd February 2006

1 Introduction

We are concerned with the unconstrained minimization of a differentiable nonlinear objective function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. We are interested in iterates $\{x_k\}_{k\geq 0}$ such that $x_{k+1} = x_k + s_k$ which satisfy the secant equation $B_{k+1}s_k = y_k$, where $y_k \stackrel{\text{def}}{=} g(x_{k+1}) - g(x_k)$ and $g(x) \stackrel{\text{def}}{=} \nabla_x f(x)$, for some suitable sequence of symmetric matrices $\{B_k\}_{k\geq 0}$. The simplest of these methods is based upon the *symmetric rank-one* formula

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}. \tag{1.1}$$

This formula is often dismissed, at least for linesearch methods, since there is no guarantee that in general it will generate positive-definite updates. Recently, however, Roger Fletcher [1, Thm. 2.2] observed that if (1.1) is applied to a strictly convex quadratic function

$$f(x) = g^T x + \tfrac{1}{2} x^T H x, \tag{1.2}$$

starting with $B_0 = 0$, then the sequence of updates $\{B_k\}$ will all be positive semi-definite so long as the steps s_k are linearly independent. Indeed, B_k will be of rank k, and $B_n = H$.

In this paper we are interested on whether it is possible to improve on Fletcher's result. In particular, the fact that B_k will be singular is inconvenient if we wish to generate steps s_k according to the usual quasi-Newton step formula

$$B_k s_k = -\alpha_k g_k, \tag{1.3}$$

where $g_k \stackrel{\text{def}}{=} g(x_k)$ and the stepsize $\alpha_k > 0$ is chosen by some appropriate linesearch—in what follows, we shall simply require that $\alpha_k = 1$ is tried before more adventurous (and possibly costly) other possibilities. Instead we aim to show that (1.1) will generate positive definite approximations in the quadratic case so long as B_0 is chosen properly.

2 New results

If f is the quadratic (1.2), $y_k = H s_k$, and (1.1) becomes

$$B_{k+1} = B_k + \frac{(H - B_k)s_k s_k^T (H - B_k)^T}{s_k^T (H - B_k) s_k}.$$

Letting $E_k \stackrel{\rm def}{=} H - B_k$, we thus deduce that

$$B_{k+1} = B_k + \frac{E_k s_k s_k^T E_k^T}{s_k^T E_k s_k} \tag{2.4}$$

and

$$E_{k+1} = E_k - \frac{E_k s_k s_k^T E_k^T}{s_k^T E_k s_k} \tag{2.5}$$

so long as the formula is well defined (i.e., $s_k^T E_k s_k \neq 0$). Clearly in this case, $E_{k+1} s_k = 0$, from which it is easy to deduce by induction that

$$E_{k+1} s_j = 0 \;\text{ for all }\; 0 \leq j \leq k. \tag{2.6}$$

Thus $E_n = 0$ and $B_n = H$ if the steps $\{s_0, s_1, \ldots s_{n-1}\}$ are linearly independent.

But what of the requirement $s_k^T E_k s_k \neq 0$? The most natural case to consider is when the matrix E_k is positive semi-definite, for then $E_k s_k = 0$ whenever $s_k^T E_k s_k = 0$. If $E_k s_k = 0$, then $H s_k = B_k s_k = -g_k = -g - H x_k$ and hence $g(x_{k+1}) = g + H x_{k+1} = g + H x_k + H s_k = 0$ if s_k satisfies (1.3) with $\alpha_k = 1$. Thus so long as E_k is positive semi-definite and $s_k^T E_k s_k = 0$, it follows that x_{k+1} minimizes $q(x)$.

So now suppose that E_k is positive semi-definite. In this case we may write

$$E_k = F_k^T F_k$$

for some square F_k. Let $\mathcal{N}_k$ denote the null-space of F_k and m_k be its dimension—of course then both F_k and E_k have rank $n - m_k$. Since $F_k s_j = 0$ for all $0 \leq j \leq k-1$ from (2.6), it follows immediately that $\mathcal{S}_k = \text{span}\{s_j\}_{j=0}^k \in \mathcal{N}_k$, and thus m_k is at least as large as the rank of $\mathcal{S}_k$. Moreover,

$$E_{k+1} = F_k^T \left(I - \frac{F_k s_k s_k^T F_k^T}{\|F_k s_k\|_2^2} \right) F_k = F_k^T Q_k^T Q_k F_k = F_{k+1}^T F_{k+1}$$

where

$$F_{k+1} = Q_k F_k, \tag{2.7}$$

$Q_k = I - w_k w_k^T$ is an orthogonal projector, and $w_k = F_k s_k / \|F_k s_k\|_2$. Hence E_{k+1} is positive semi-definite whenever E_k is. Furthermore $\mathcal{N}_{k+1} \subseteq \mathcal{N}_k$ because of (2.7), and $m_{k+1} = m_k + 1$ if and only if $F_k s_k = 0$. Note that, by convention $w_k = 0$ if $F_k s_k = 0$, and thus $F_{k+1} = F_k$ in this case.

With this in mind, if we choose B_0 so that E_0 is positive semi-definite, the same will be true by induction for all E_k for $k > 0$. Moreover, the interlacing eigenvalue property [2, Thm. 8.1.8] applied to (2.4) indicates that the eigenvalues of B_k will be monotonically non-decreasing as k increases. If $\lambda_{\min}(A)$ denotes the smallest (leftmost) eigenvalue of a generic symmetric matrix A, formally we can say the following.

Theorem 2.1. Let H be symmetric positive definite, Suppose that the symmetric B_0 is chosen so that $E_0 = H - B_0$ is positive semi-definite, and that E_0 is of rank r. Then if $\{B_k\}_{k\geq 0}$ is generated according to (1.1) and $\{E_k\}_{k\geq 0}$ according to (2.5) for which $E_k s_k \neq 0$ for all $k \geq 0$, it follows that E_k will be symmetric, positive semi-definite and of rank $r - k$ for all $k \leq r$. Thus $E_r = 0$ and $B_r = H$. In addition $\lambda_{\min}(B_k) \leq \lambda_{\min}(B_{k+1})$.

The requirement that E_0 be positive semi-definite is very weak, and might be enforced by choosing $B_0 = 0$ [1, Thm. 2.2] or even $B_0 = -I$. But perhaps more usefully we have the following immediate corollary of Theorem 2.1.

Corollary 2.2. Let H be symmetric positive definite. Suppose that the symmetric B_0 and the resulting $\{B_k\}_{k\geq 0}$ and $\{E_k\}_{k\geq 0}$ are chosen as in Theorem 2.1, and that $E_k s_k \neq 0$. Suppose in addition B_0 is positive definite. Then in addition to the conclusions of Theorem 2.1, $\{B_k\}_{k\geq 0}$ will be positive definite.

In order to satisfy the requirements of Corollary 2.2, the matrix B_0 should be chosen so that $H - B_0$ is positive semi-definite and B_0 is positive definite. Although it isn't immediately clear how to do this in general, one choice is $B_0 = \lambda_{\min}(H) I$.

2.1 The Wolfe condition

Let us suppose that we satisfy the Wolfe condition

$$s_k^T g_{k+1} \geq \gamma s_k^T g_k, \tag{2.8}$$

where $0 < \beta < \gamma < 1$ and β is the constant associated with Armijo condition. In this case, using (1.3) and (2.8)

$$s_k^T (y_k - B_k s_k) = s_k^T (g_{k+1} - \gamma g_k - (1 - \gamma - \alpha_k) g_k) > 0$$

provided that $\alpha_k < 1 - \gamma$.

References

[1] R. Fletcher. A new low rank quasi-Newton update scheme for nonlinear programming. Numerical Analysis Report NA/223, Department of Mathematics, University of Dundee, Scotland, 2005.

[2] G. H. Golub and C. F. Van Loan. *Matrix computations.* Johns Hopkins University Press, Baltimore, third edition, 1996.