

An Explicit Quasi-Newton Update for Sparse Optimization Calculations

By Angelo Lucia*

Abstract. A new quasi-Newton updating formula for sparse optimization calculations is presented. It makes combined use of a simple strategy for fixing symmetry and a Schubert correction to the upper triangle of a permuted Hessian approximation. Interesting properties of this new update are that it is closed form and that it does not satisfy the secant condition at every iteration of the calculations. Some numerical results are given that show that this update compares favorably with the sparse PSB update and appears to have a superlinear rate of convergence.

1. Introduction. In sparse minimization calculations we are concerned with finding the least value of a twice continuously differentiable functional, $f: R^n \rightarrow R$, whose second-derivative matrix is sparse. Here $x \in R^n$ denotes a vector of unknown variables. It is assumed that the gradient of f , denoted by g , is available and that the sparsity pattern of the Hessian matrix, $\nabla^2 f(x)$, is fixed throughout the calculations.

When n , the number of variables, is large it is often necessary to exploit any sparsity in the Hessian matrix of f in order to save storage and computation. To date there are two quasi-Newton updates for sparse optimization calculations, the sparse Powell-symmetric-Broyden (PSB) update suggested by Toint [8] and Marwil [5] and the sparse Broyden-Fletcher-Goldfarb-Shanno (BFGS) update proposed by Shanno [7].

To briefly describe these updates we let B and \bar{B} denote successive approximations to the Hessian matrix, x and \bar{x} denote successive approximations to a local minimizer of f , and we define $s = \bar{x} - x$ and $y = g(\bar{x}) - g(x)$. We also let I be the set of index pairs for which the elements of the true Hessian matrix are known to be zero and we let $\|\cdot\|_F$ denote the Frobenius matrix norm.

It is well known (see, for example, Toint [8]) that the sparse PSB update solves the variational problem

$$\min \|\bar{B} - B\|_F$$

subject to

$$(1) \quad \bar{B}s = y,$$

$$(2) \quad \bar{B} = \bar{B}'$$

Received March 24, 1981; revised October 15, 1981 and June 21, 1982.

1980 *Mathematics Subject Classification*. Primary 65K05; Secondary 90C30.

**Current address*: Department of Chemical Engineering, Clarkson College of Technology, Potsdam, New York 13676.

and

$$(3) \quad \bar{b}_{ij} = 0, \quad \forall (i, j) \in I,$$

where Eqs. (1), (2) and (3) are the secant or quasi-Newton condition, symmetry and sparsity requirements, respectively.

To update B in this manner requires the solution of a system of linear equations in the Lagrangian multipliers for the secant condition, written as

$$(4) \quad Q\lambda = r = y - Bs.$$

This set of equations, whose coefficient matrix has the same sparsity pattern as that of the true Hessian matrix, is solved for λ and then the nonzero elements of B are updated by

$$\bar{b}_{ij} = b_{ij} + (\lambda_j s_i + \lambda_i s_j), \quad \forall (i, j) \notin I.$$

This gives a sparse, symmetric matrix \bar{B} that satisfies the secant condition. The reader is referred to the paper by Toint [8] for the details of the variational derivation of the sparse PSB.

The only other successful quasi-Newton update for sparse optimization calculations is the sparse BFGS update recently suggested by Shanno [7]. Here we first calculate a full update to B by the rule

$$B^* = B + \frac{yy'}{y's} - \frac{Bss'B}{s'Bs},$$

which is the usual BFGS formula. Then the matrix \bar{B} is calculated by solving the variational problem $\min \| \bar{B} - B^* \|_F$ subject to (1), (2) and (3). Like the sparse PSB update, the sparse BFGS requires the solution of a system of linear equations to calculate an update to B . In fact, the only difference between (4) and Shanno's system of equations is the right-hand side vector r . That is, the vector r in the sparse PSB update depends on the error in the quasi-Newton condition, whereas in Shanno's method it depends on the matrix B^* . The net result is again a sparse, symmetric matrix \bar{B} that satisfies the secant condition. The reader is referred to the paper by Shanno [7] for the details of the technique of making any nonsparse update sparse, and, in particular, the sparse BFGS update.

While these updates give reasonably good numerical results, the fact is that they require the solution of an associated linear system at each iteration. This can make them computationally expensive over the course of the minimization calculations. Furthermore, the linear system for either method is subject to ill-conditioning problems because of the way the Q matrix is constructed. Instead we prefer to update B in an explicit manner, provided this closed form update gives good results. Such a closed form update is the subject of this paper.

The rest of this paper is organized in the following way. In Section 2 we present a closed form, sparse, symmetric quasi-Newton updating formula. This update makes combined use of the Schubert [6] update, a simple similarity transformation and a different way of fixing symmetry. An interesting property of the new update, which is shared by only one other quasi-Newton update (see Dennis and Schnabel [2]), is that it does not satisfy the secant condition at each iteration. It does however satisfy this condition in the limit provided the sequence of Hessian approximations converges. In Section 3 we present some numerical results. We compare the new

updating formula to the sparse PSB and sparse BFGS updates. We find that the new update formula is comparable to the sparse PSB update. We also present numerical results that measure the error in the secant condition during the course of the minimization calculations and yet other results that measure the Dennis and Moré [1] characterization of superlinear convergence. The latter results suggest that the new updating formula may possess a superlinear rate of convergence. Finally, in Section 4, we make some closing remarks.

2. A Closed Form Update. In this section we describe a new, closed form update that makes use of the Schubert [6] update, a simple similarity transformation, and a straightforward way of fixing symmetry.

To begin, we define the vector π to be a permutation of the integers $1, 2, \dots, n$ such that $|s_{\pi_1}| \leq |s_{\pi_2}| \leq \dots \leq |s_{\pi_n}|$. This is simply the integer permutation defined by arranging the components of s in order of their ascending absolute value.

Next we permute the vectors s and y and the matrix B according to π . In particular, we define

$$v = p(s) = \left(\sum_{k=1}^n e_k e'_{\pi_k} \right) s, \quad w = p(y) = \left(\sum_{k=1}^n e_k e'_{\pi_k} \right) y$$

and

$$A = P(B) = \left(\sum_{k=1}^n e_k e'_{\pi_k} \right) B \left(\sum_{m=1}^n e_{\pi_m} e'_m \right).$$

Now we calculate a Schubert [6] correction to the upper triangle of A , denoted A_u . This gives

$$\bar{A}_u = A_u + \sum_{i=1}^n (v'_i v_i)^+ e'_i [w - Av] e'_i v'_i,$$

where v_i is the vector, formed from v , that reflects the sparsity pattern of \bar{A}_u , the superscript $+$ denotes the pseudoinverse and $'$ denotes transposition. We interpret \bar{A}_u as the least change correction (see, Dennis and Schnabel [2]) to A_u subject to the linear constraint $\bar{A}_u v = w - (A'_u - \text{diag}(A_u))v$.

To fix symmetry we simply set $\bar{a}_{ji} = \bar{a}_{ij}$ for $(i, j) \notin I$ and $j > i$. This gives

$$\bar{A} = A + \sum_{i=1}^n (v'_i v_i)^+ e'_i [w - Av] T^{-1}(e_i v'_i),$$

where T^{-1} is the linear transformation that accomplishes this symmetry operation.

Finally, to recover \bar{B} , we apply the inverse of P . Hence,

$$(5) \quad \bar{B} = B + \sum_{i=1}^n (v'_i v_i)^+ e'_i [w - Av] (T \circ P)^{-1}(e_i v'_i),$$

which is the desired result, a closed form, sparse and symmetric updating formula.

Furthermore, besides being explicit, the new update has another interesting property. It is the only published quasi-Newton formula, other than the update suggested by Dennis and Schnabel [2, p. 455], that does not satisfy the secant condition at each iteration of the calculations. To see this, observe that the linear constraint that was used in calculating a Schubert correction to A_u states simply that

$$(\bar{A}_u + A'_u - \text{diag}(A_u))v = w.$$

This clearly implies that $\bar{A}v \neq w$, or equivalently that $\bar{B}s \neq y$, when $\bar{A}_u \neq A_u$.

However, for the case when the sequence of Hessian approximations converges to say B^* , we have that $\lim A_u = \lim \bar{A}_u = A_u^*$ and in the limit

$$(A_u^* + A_u^{*'} - \text{diag}(A_u^*))v = A_u^*v = w,$$

which, in turn, implies that $B^*s = y$. So we see that although the new update does not satisfy the secant equation at each iteration of the calculations, under the assumption that the Hessian approximations converge to a finite limit, B^*s will equal y .

In the next section we present some numerical results for this new updating formula.

3. Some Numerical Results. To test the new update, we solved six benchmark problems that appear in the sparse unconstrained literature. The test problems were the quadratic operations research (QOR), general operations research (GOR), pseudo-penalty (PSP) and chained Rosenbrock (CR) functions defined in Toint [9] and the tridiagonal (TRIDIA) and extended Rosenbrock (EROSEN) functions in the paper by Shanno [7].

For all the numerical tests we used the sparse unconstrained minimization algorithm given in Toint [9] with one modification; we did not use a Hebden [4] correction to B to guarantee positive definiteness at each iteration. Instead, we used the Powell dogleg strategy (see, for example, Toint [9]) which can safely handle directions of negative curvature of the local quadratic approximations. The starting point used in step 1 of the Toint algorithm was the same as the starting point used in the paper from which the particular problem statement was taken. Also, values of $\Delta_0 = 1$ and $\epsilon = 10^{-5}$ were used in step 1 for all problems. Step 3 of the algorithm was implemented using the Yale Sparse Matrix Package—The Symmetric Codes (YSMPS) [3], which calculates a $U'DU$ decomposition of B and locally optimizes the arithmetic operations count for the Gaussian elimination process. Finally, all computer runs in this section were done on a CDC Cyber 175 computer using single-precision arithmetic unless specified otherwise.

We solved the six forementioned problems with the sparse PSB (SPSB) method, sparse BFGS (SBFGS) update, and the new updating formula. Table 1 shows the numerical results.

TABLE 1

Problem	size	no. of function calls		
		SPSB	SBFGS	new update
QOR	50	22 (3.004)	35 (4.397)	27 (2.910)
GOR	50	50 (3.986)	> 200	70 (3.914)
PSP	50	202 (12.395)	> 300	238 (10.494)
CR	25	70 (1.375)	89 (1.920)	85 (1.375)
TRIDIA	30	17 (0.843)	22 (1.052)	18 (0.785)
EROSEN	5	247 (1.238)	131 (0.994)	> 300

The symbol $> nnn$ is used to indicate that the computer run for this problem was terminated after exactly nnn function calls without reaching the required accuracy. The numbers in parentheses are the execution times in CPU seconds.

With regard to the new updating formula, we also measured the error in the secant condition, $\|y - \bar{B}s\|_2/\|s\|_2$, and the Dennis and Moré [1] characterization of superlinear convergence, $\|(B - \nabla^2 f(x^*))s\|_2/\|s\|_2$, toward the end of the calculations for all problems in Table 1. Results for the chained Rosenbrock and tridiagonal functions are shown in Tables 2 and 3. We remark that the asymptotic behavior of the characterization of superlinear convergence must be measured using double-precision arithmetic because it involves a division by $\|s\|_2$. (This is equivalent to quadruple or extended precision on IBM machines.)

TABLE 2

Asymptotic behavior of the error in the secant condition and the characterization of superlinear convergence for the CR function

iteration no.	$\ y - \bar{B}s\ _2/\ s\ _2$	$\ (B - \nabla^2 f(x^*))s\ _2/\ s\ _2$
80	$2.01 \times 10^{+1}$	$2.22 \times 10^{+1}$
81	4.17	$1.22 \times 10^{+1}$
82	9.65×10^{-1}	1.18
83	1.18×10^{-1}	1.15×10^{-1}
84	4.31×10^{-2}	1.08×10^{-2}
85	3.94×10^{-3}	7.65×10^{-3}

TABLE 3

Asymptotic behavior of the error in the secant condition and the characterization of superlinear convergence for the TRIDIA function

iteration no.	$\ y - \bar{B}s\ _2/\ s\ _2$	$\ (B - \nabla^2 f(x^*))s\ _2/\ s\ _2$
9	$6.67 \times 10^{+1}$	$9.99 \times 10^{+1}$
12	$8.02 \times 10^{+1}$	$1.65 \times 10^{+1}$
15	3.83×10^{-1}	6.28
16	2.96×10^{-2}	3.45×10^{-1}
17	3.04×10^{-3}	4.61×10^{-2}

4. Conclusions. Using a simple symmetry strategy, an elementary similarity transformation and the Schubert update, we have developed a new closed form quasi-Newton update for sparse optimization calculations. In doing so, we have avoided the need to iteratively solve a sparse symmetric linear system in order to update the Hessian approximations as in the sparse PSB and the sparse BFGS methods. We have also precluded any ill-conditioning problems associated with these linear systems. We remark that the permutation is necessary for numerical stability because without it we obtained very poor Hessian approximations. Apparently, the permutation ensures that each term $(v_i/v_j)^+$ is as small as possible, in the face of the given symmetry strategy, and thereby does not cause inordinate growth of the approximating matrices.

We have solved six benchmark problems using this closed form updating formula and compared our results with the sparse PSB and sparse BFGS updates. Based on these results we find that the closed form update is comparable to the sparse PSB method. It should be noted that the attractiveness of the proposed method compared to the sparse PSB depends on the relative cost of permuting an n vector with that of solving an $n \times n$ sparse, symmetric linear system.

An interesting property of the new updating formula is that it does not satisfy the secant condition at each iteration of the calculations. However, we have shown that, under the assumption that the Hessian approximations converge to a finite limit, the new update will satisfy the secant condition in the limit. Our numerical results verify this claim.

We have also measured the asymptotic behavior of the characterization of superlinear convergence. These numerical results show that we get a close tracking of the Newton direction toward the end of the calculations. This suggests that the rate of convergence of the new update may be Q -superlinear.

In summary, we note that our numerical results seem to indicate that a fast rate of convergence can be obtained without requiring the secant condition to be satisfied at every iteration of the calculations. This is a new result.

Finally, there is a simple and interesting Gauss-Seidel type variant of the new updating formula. If we replace A_u by \bar{A}_u in the linear constraint of the variational calculations, then that constraint becomes the usual secant condition and the updating formula is then given by

$$(6) \quad \bar{B} = B + \sum_{i=1}^n (v_i' v_i)^+ e_i' [w - (\bar{A}_u + \bar{A}_u' - \text{diag}(\bar{A}_u)) v] (T \circ P)^{-1} (e_i v_i').$$

This updating formula is still a closed form update and does not require any additional work. Furthermore, it uses updated information as soon as it is available. We remark that we actually discovered Eq. (6) before we discovered Eq. (5), but we have not tested it to any extent.

Acknowledgements. The author would like to thank L. F. Stutzman and S. J. Sidney for their help and encouragement, the referees for several helpful suggestions, and the Control Data Corporation for their financial support.

Department of Chemical Engineering
University of Connecticut
Storrs, Connecticut 06268

1. J. E. DENNIS & J. J. MORÉ, "A characterization of superlinear convergence and its application to quasi-Newton methods," *Math. Comp.*, v. 28, 1974, pp. 549–560.
2. J. E. DENNIS & R. B. SCHNABEL, "Least change secant updates for quasi-Newton methods," *SIAM Rev.*, v. 21, 1979, pp. 443–459.
3. S. C. EISENSTAT, M. C. GURSKY, M. H. SCHULTZ & A. H. SHERMAN, *Yale Sparse Matrix Package—The Symmetric Codes*, Report No. 112, Dept. of Computer Science, Yale University, New Haven, Conn., 1977.
4. M. D. HEBDEN, *An Algorithm for Minimization Using Exact Second Derivatives*, Report No. T.P. 515, A.E.R.E. Harwell, 1973.
5. E. S. MARWIL, *Exploiting Sparsity in Newton-Like Methods*, Ph.D. Thesis, Cornell University, Ithaca, N.Y., 1978.
6. L. K. SCHUBERT, "Modification of a quasi-Newton method for nonlinear equations with a sparse Jacobian," *Math. Comp.*, v. 24, 1970, pp. 27–30.
7. D. F. SHANNO, "On variable metric methods for sparse Hessians," *Math. Comp.*, v. 34, 1980, pp. 499–514.
8. PH. L. TOINT, "On sparse and symmetric matrix updating subject to a linear equation," *Math. Comp.*, v. 31, 1977, pp. 954–961.
9. PH. L. TOINT, "Some numerical results using a sparse matrix updating formula in unconstrained optimization," *Math. Comp.*, v. 32, 1978, pp. 839–851.