SIAM J. OPTIM.
Vol. 9, No. 4, pp. 1192–1204

THE QUASI-CAUCHY RELATION AND DIAGONAL UPDATING*

M. ZHU†, J. L. NAZARETH‡, AND H. WOLKOWICZ§

Dedicated to Professor John E. Dennis on the occasion of his 60th birthday

Abstract. The quasi-Cauchy (QC) relation is the weak quasi-Newton relation of Dennis and Wolkowicz [*SIAM J. Numer. Anal.*, 30 (1993), pp. 1291–1314] with the added restriction that full matrices are replaced by diagonal matrices. This relation is justified and explored and, in particular, two basic variational techniques for updating diagonal matrices that satisfy it are formulated.

For purposes of illustration, a numerical experiment is described where a diagonal updated matrix with hereditary positive definiteness is used to precondition Cauchy's steepest-descent direction. The resulting QC algorithm is shown to be significantly accelerated.

In the concluding section, the following topics are briefly discussed: additional variational principles, use of diagonal updates within other optimization algorithms together with some further numerical experience (summarized in an appendix), and an interesting connection between QC-diagonal updating and trust-region techniques.

Key words. weak-quasi-Newton, quasi-Cauchy, diagonal updating, Cauchy algorithm, steepest-descent

AMS subject classifications. 49, 90, 65

PII. S1052623498331793

1. Introduction.

We consider the problem of finding a local minimum of a smooth, unconstrained nonlinear function, namely,

$$\text{minimize}_{x\in R^n} \quad f(x). \tag{1}$$

For purposes of discussion, it is useful to identify a hierarchy of relations that can be employed within Newton and Cauchy algorithms for solving (1) (see, for example, Dennis and Schnabel [4], Bertsekas [2], and Nazareth [14] for background):

- Quasi-Newton (QN). $M_+s = y$, where the n-dimensional vector $s = x_+ - x$ denotes the step corresponding to two different points x and x_+, and $y = g_+ - g$ denotes the gradient change corresponding to the gradients g and g_+ at the two points. Assume $s^Ty > 0$. M_+ is a full $n \times n$ matrix that approximates the Hessian of f. Both s and y are used explicitly and $O(n^2)$ storage is required for the matrix M_+.

  If M is chosen to be a positive definite diagonal matrix, say D, then one can recur only the diagonal elements of M_+ in a QN update formula, for example, the BFGS, yielding an updated diagonal matrix D_+. The matrix M_+ is positive definite, and hence D_+ is also positive definite, but obviously D_+ does not satisfy the QN relation. Only $O(n)$ storage is required to store D_+. This diagonal-updating approach is used in Gill and Murray [9] and Gilbert and Lemaréchal [8].

*Received by the editors December 15, 1998; accepted for publication (in revised form) March 5, 1999; published electronically September 24, 1999.
http://www.siam.org/journals/siopt/9-4/33179.html
†Microsoft Corporation, 1 Microsoft Way, Redmond, WA 98052 (minzhu@microsoft.com).
‡Department of Pure and Applied Mathematics, Washington State University, Pullman, WA (nazareth@amath.washington.edu).
§Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada (hwolkowi@orion.math.uwaterloo.ca).

- Weak-quasi-Newton. $s^T M_+ s = s^T y$. This relation was introduced and studied by Dennis and Wolkowicz [5]. For example, one of the updates proposed in [5] is as follows:

$$M_+ = M + \frac{(s^T y - s^T M s)}{(s^T M s)^2} M s s^T M, \tag{2}$$

  where M is positive definite. The condition $s^T y > 0$ implies that M_+ is also positive definite. Again s and y are used explicitly and $O(n^2)$ storage is required.

  As in the QN case, if M is taken to be a positive definite diagonal matrix D, the foregoing formula (2) can be restricted to updating only the diagonal elements of M_+, yielding a positive definite updated matrix, say D_+. In general, D_+ does not satisfy the weak-QN relation.

  It is interesting to note that the quantity $s^T y$ in expression (2), which equals $g_+^T s - g^T s$, can be obtained directly from directional derivative differences along s that require only function values. Thus, knowledge of gradient vectors is not essential in this formula.
- Quasi-Cauchy (QC). $s^T D_+ s = s^T y$, where D_+ is required to be a diagonal matrix, i.e., the QC relation is the weak QN with matrices further restricted to be diagonal. The vectors s and y are assumed to be available. Only $O(n)$ storage is required to store the diagonal update. Additionally, we would like the matrix D_+ to be positive definite and thus able to define a metric. An obvious usage would be to precondition or rescale Cauchy's steepest-descent direction, which accounts for our choice of terminology.

  Consider the well-known Oren–Luenberger scaling matrix, namely,

$$D_+ = (s^T y / s^T s) I,$$

  where I is the identity matrix. It is interesting to note that this is precisely the unique matrix that would be obtained from the QC relation with the further restriction that the diagonal matrix is a scalar multiple of the identity matrix, i.e., the diagonal elements of the Hessian approximation D_+ are equal and the model function associated with it has contours that are hyperspheres. Thus, scaling matrices derived from the QC relation are a natural generalization of Oren–Luenberger scaling.

  As in the foregoing discussion on the weak-QN relation, the quantity $s^T y$ in the right-hand side of the QC relation can be obtained by directional derivative differences along s. Thus, explicit use of gradient vectors can be circumvented, and the resulting diagonal update can find potential use in an algorithm that requires only approximations to gradients (quasi-gradients). The QC relation and variational-based diagonal updating were originally proposed in this setting in [15], [16].

The purpose of this article is to formulate two basic techniques for diagonal updating subject to the QC relation (section 2). These are based on variational principles that are analogous to ones employed in quasi-Newton updating. The first is the analogue of the principle from which the Powell symmetric Broyden (PSB) quasi-Newton update is derived—see, for example, Dennis and Schnabel [4]. Like PSB, the diagonal update does not have the hereditary positive definiteness property. The second is based on a principle analogous to that from which the BFGS update is commonly

derived—again, see [4]. Like BFGS, the diagonal update has hereditary positive definiteness and can therefore be used to define a metric. So can its complementary form, which corresponds to DFP.

For purposes of illustration, the latter diagonal update is used to iteratively precondition (or rescale) Cauchy's steepest-descent algorithm, and the results of its numerical performance on a set of standard MINPACK-1 test problems are reported (section 3). The algorithm is shown to be significantly accelerated.

In the concluding section, we briefly discuss further variational principles; the use of diagonal updates within other optimization algorithms, in particular, the L-BFGS algorithm (some additional numerical results are summarized in an appendix); and an interesting connection with trust-region techniques.

More detail can be found in Zhu [19], [20], where a comprehensive theory of diagonal updating subject to the QC relation is developed and applied.

2. QC-diagonal updating. Suppose $D > 0$ is a positive definite diagonal matrix and D_+, which is also diagonal, is the updated version of D. We require that the updated D_+ satisfy the QC relation and that the deviation between D and D_+ is minimized under some variational principle. (Here we will use only the Frobenius matrix norm to measure the deviation.) We would like the derived update to preserve positive definiteness in a natural way, i.e., we seek well-posed metric problems such that the solution D_+, through the diagonal updating procedure, incorporates available curvature information from the step and gradient changes, as well as that contained in D. As noted earlier, a diagonal matrix uses the same computer storage as a vector so only $O(n)$ storage is required. Thus, the resulting update will have potential use in algorithms where storage is at a premium.

We now focus on two basic forms of QC-diagonal updating.

2.1. Updating D. Consider the variational problem

$$(P): \text{ minimize } ||D_+ - D||_F$$

$$\text{subject to (s.t.) } s^T D_+ s = s^T y,$$

where $s \neq 0$, $s^T y > 0$, and $D > 0$. Let

$$D_+ = D + \Lambda, \quad a = s^T D s, \quad b = s^T y. \tag{3}$$

Then the variational problem can be stated alternatively as

$$(P): \text{ minimize } \frac{1}{2}||\Lambda||_F^2$$

$$\text{s.t. } s^T \Lambda s = b - a.$$

In (P), the objective is strictly convex and the feasible set is convex. Therefore, there exists a unique solution to (P). Its Lagrangian function is

$$L(\Lambda, \mu) = \frac{1}{2}\text{tr}(\Lambda^2) + \mu(s^T \Lambda s + a - b),$$

where μ is the Lagrange multiplier associated with the constraint and tr denotes the trace operator. Differentiating with respect to the diagonal elements, setting the result to zero, and invoking the constraint $s^T \Lambda s = b - a$, we have

$$\Lambda = \frac{b-a}{\text{tr}(E^2)} E, \quad E = \text{diag}\,[s_1^2, s_2^2, \ldots, s_n^2], \tag{4}$$

where s_i is the ith element of s. When $b < a$, note that the resulting $D_+ = D + \Lambda$ is not necessarily positive definite.

The foregoing update is the counterpart of the PSB update in the quasi-Newton setting and, like the latter, it does not preserve positive definiteness. Thus it is inappropriate for use within a metric-based algorithm.

2.2. Updating $D^{1/2}$. An alternative approach to preserving positive definiteness through diagonal updating, which is the analogue of the principle used to derive the BFGS update in the quasi-Newton setting, is to update the square root or Cholesky factor $D^{1/2}$ to give the corresponding $D_+^{1/2}$ with

$$D_+^{1/2} = D^{1/2} + \Omega,$$

where Ω is chosen to

$$(FP): \text{ minimize } ||\Omega||_F \tag{5}$$

$$\text{s.t. } s^T(D^{1/2} + \Omega)^2 s = s^T y > 0.$$

The foregoing variational problem is well posed, being defined over the closed set of matrices for which the corresponding D_+ is positive semidefinite. Further, analogous to the full matrix case in standard QN updating, it always has a viable solution for which D_+ is positive definite, as we now show in the following theorem.

THEOREM 2.2.1. *Let $D > 0$, $s \neq 0$, and a, b, E be defined as in* (3) *and* (4). *There is a unique global solution Ω of (FP) which is given by*

$$\Omega = \begin{cases} 0 & \text{if } b = a, \\ -\mu^* E(I + \mu^* E)^{-1} D^{1/2} & \text{if } b \neq a, \end{cases} \tag{6}$$

where μ^* is the largest solution of the nonlinear equation $F(\mu) = b$ and

$$F(\mu) \stackrel{\text{def}}{=} s^T(D(I + \mu E)^{-2})s = \sum_{\{i: s_i \neq 0\}} \frac{d_i s_i^2}{(1 + \mu s_i^2)^2}. \tag{7}$$

Proof. In the process of the proof we will see that every expression above is well defined. Let $\Omega = \text{diag}(\omega_1, \ldots, \omega_n)$ and let ω denote the vector of diagonal elements $(\omega_1, \ldots, \omega_n)^T$. First, by a simple transformation, problem (FP) is equivalent to

$$(FP): \text{ minimize } ||\omega||_2^2 = w^T w$$

$$\text{s.t. } \omega^T E \omega + 2w^T E r = b - a,$$

where

$$r = [d_1^{1/2}, d_2^{1/2}, \ldots, d_n^{1/2}]^T.$$

When $b = a$, the global optimal solution is obviously $\omega = 0$, and hence $\Omega = 0$, which implies that $D_+ = D$ is positive definite. In the following discussion we assume that $b \neq a$. Problem (FP) has a strictly convex objective with the Hessian E of the constraint being positive semidefinite. By a theorem concerning a quadratic objective with also a quadratic constraint in [12], (FP) has a global solution. Differentiating its Lagrangian

$$L(\omega, \mu) = \omega^T \omega + \mu(\omega^T E \omega + 2w^T E r + a - b)$$

with respect to ω, where μ is the Lagrange multiplier, and setting the result to zero, we have

$$\omega_i = -\frac{\mu s_i^2 d_i^{1/2}}{(1+\mu s_i^2)}, \quad i = 1, \ldots, n.$$

Substituting these quantities into the constraint equation, we obtain

$$\begin{aligned} F(\mu) &\stackrel{\text{def}}{=} s^T (D(I+\mu E)^{-2}) s \\ &= \sum_{i=1}^{n} \frac{d_i s_i^2}{(1+\mu s_i^2)^2} \\ &= \sum_{\{i: s_i \neq 0\}} \frac{d_i}{s_i^2 (\mu + (1/s_i^2))^2} \\ &= b. \end{aligned}$$

Note that $F(\mu)$ has poles at $(-1/s_i^2)$, $i = 1, \ldots, n$. Let

$$j = \arg\max_{\{i, s_i \neq 0\}} \left(-\frac{1}{s_i^2} \right).$$

The derivative of $F(\mu)$ is

$$\frac{dF(\mu)}{\mu} = -2 \sum_{\{i: s_i \neq 0\}} \frac{r_i^2}{s_i^2 (\mu + (1/s_i^2))^3},$$

which is less than zero on the interval

$$\left(-\frac{1}{s_j^2}, +\infty \right),$$

so $F(\mu)$ is strictly decreasing in the above interval from $+\infty$ to 0. Noting that $b > 0$, we see that there is a unique solution μ^* within this interval such that $F(\mu^*) = b$. Although the behavior of $F(\mu)$ is complicated in the entire domain, solutions for $F(\mu) = b$ except μ^* are of no interest. (Note that μ^* is the largest solution.) This is because a necessary condition [12] of the solution of (FP) requires the Hessian of the Lagrangian (with respect to ω), namely, $2(I + \mu E)$, to be positive semidefinite. This is equivalent to

$$1 + \mu s_i^2 \geq 0, \quad i = 1, \ldots, n,$$

and clearly μ^* is the unique solution of $F(\mu) = b$ satisfying the above inequalities. A key observation is that $I + \mu^* E$ is positive definite, and thus μ^* is the unique global minimizer for (FP). Returning to the relationship of ω and μ, we see that

$$\Omega^* = -\mu^* E (I + \mu^* E)^{-1} D^{-1/2}$$

is the unique solution of (FP). Note also that $\forall i = 1, \ldots, n$,

$$d_i^{1/2} - \frac{\mu^* s_i^2 d_i^{1/2}}{(1+\mu^* s_i^2)} = \frac{1}{1+\mu^* s_i^2} d_i^{1/2} \neq 0,$$

so D_+ is positive definite. This completes the proof. ∎

The following is a direct result of the theorem.

COROLLARY 2.2.1. *The solution D_+ through the diagonal updating problem (FP) is positive definite and unique and is given by*

$$D_+ = \begin{cases} D & \text{if } b = a, \\ (I + \mu^* E)^{-2} D & \text{if } b \neq a. \end{cases} \tag{8}$$

Make the following definitions:

$$U = D^{-1}, \quad c = y^T U y, \quad G = [y_1^2, \ldots, y_n^2].$$

One can obtain the update that is complementary to the update in the foregoing theorem by making the following transpositions:

$$\mu \leftrightarrow \nu, \quad s \leftrightarrow y, \quad a \leftrightarrow c, \quad D \leftrightarrow U, \quad D_+ \leftrightarrow U_+.$$

This is summarized in the following result, which is based on the analogue of the variational principle from which the DFP quasi-Newton update is derived.

COROLLARY 2.2.2. *The solution U_+ through the diagonal updating problem complementary to (FP) is positive definite and uniquely given by*

$$U_+ = \begin{cases} U & \text{if } b = c, \\ (I + \nu^* G)^{-2} U & \text{if } b \neq c, \end{cases} \tag{9}$$

where ν^* is the largest solution of $H(\nu) = b$ and

$$H(\nu) \stackrel{\text{def}}{=} y^T (U(I + \nu G)^{-2}) y = \sum_{\{i : y_i \neq 0\}} \frac{u_i y_i^2}{(1 + \nu y_i^2)^2}.$$

3. Numerical illustration. An immediate application for the diagonal update of the previous section, which we use for purposes of illustration, is to dynamically scale the steepest-descent direction at each iteration of Cauchy's algorithm.

The Cauchy direction is ideal when the contours of the objective f to be minimized are hyperspheres. For a general function that is not quadratic, a preconditioning can be used to make the transformed contours closer to hyperspheres such that the efficiency of the Cauchy direction in the transformed space is enhanced. The diagonal updating is a variable preconditioning which includes the updated curvature information, and its hereditary positive definiteness is naturally maintained when the Cholesky factor is updated as shown in the previous section. An expectation that the Cauchy method will be significantly accelerated using diagonal updating is supported by our numerical results.

Our source code is written in Fortran-90, with double precision algorithmic, running on an ULTRIX DEC workstation. Purely for convenience, we implemented the complementary updates which are defined in terms of the inverse matrix U_+. The numerical experiment is done within the MINPACK-1 testing environment. Test functions are the standard unconstrained problems collected in [11], which we identify by the numbering in Table 1.

We employ a line search routine of Moré and Thuente [13] along direction, say, d, which is based on cubic interpolation and satisfies the (strong) Wolfe conditions:

$$f(x_+) \leq f(x) + \alpha \lambda g^T d, \tag{10}$$

$$|g_+^T d| \leq \beta |g^T d|, \tag{11}$$

TABLE 1
MINPACK-1 test problems.

Number	Problem name
1	Helical valley function
2	Biggs exp6 function
3	Gaussian function
4	Powell badly scaled function
5	Box 3-dimensional function
6	Variably dimensioned function
7	Watson function
8	Penalty function I
9	Penalty function II
10	Brown badly scaled function
11	Brown and Dennis function
12	Gulf research and development function
13	Trigonometric function
14	Extended Rosenbrock function
15	Extended Powell function
16	Beale function
17	Wood function
18	Chebyquad function

where $x_+ = x + \lambda d$ and the line search parameters are chosen as in [6], namely, $\alpha = 10^{-4}$, $\beta = 0.9$. The stopping criterion is also as in [6]:

$$||g(x)|| \leq 10^{-5} max\{1.0, ||x||\}. \tag{12}$$

At any iterate, say, x_+, the corresponding search direction d_+ in the methods tested is as follows:

1. Standard Cauchy. The search direction is of the form $d_+ = -g_+$.
2. Cauchy with Oren–Luenberger scaling. This scales the search direction with Oren–Luenberger scaling [7] in its complementary form,

$$d_+ = -\frac{y^T s}{y^T y} g_+,$$

   for all iterations except the first, where the initial steepest-descent search direction is employed.
3. DU-Cholesky. This implements the complementary diagonal update of Corollary 2.2.2 with $d_+ = -U_+ g_+$. In our numerical implementation, ν^* is obtained by a simple bisectional search within the interval from the largest pole of the function $H(\nu)$ to some large number on the axis such that the initial bisection condition of the endpoints is satisfied. Note that $H(0) = c$, and thus if $b > c$, then the solution $\nu^* < 0$; if $b < c$, then $\nu^* > 0$. Hence, the interval for the bisection is actually reduced with one endpoint being 0 in each case. (The cost of computing ν^* by bisection is a relatively minor portion of the algorithm. Note that more efficient reformulations and techniques, for example, Newton's method, for solving the subproblem for ν^* are possible, as discussed in the concluding section.)

The numerical comparative results are given in Table 2; it gives $nitr/nfg$, namely, the number of iterations and the number of calls for function and gradient evaluation. The symbol $*$ in the table indicates that the method takes too many iterations and is regarded as having failed to converge. The first and second columns in the table are

TABLE 2
Numerical results for diagonal updating.

Prob.	Dim.	Cauchy	Cauchy-OL	DU-Cholesky
1	3	2552/5229	431/756	370/688
2	6	24041/45488	2221/4353	1165/2120
3	3	2/4	2/6	2/6
4	2	*	*	238/1649
5	3	32535/65075	225/428	165/300
6	6	446/1001	574/877	157/274
6	8	981/2318	269/415	229/427
7	2	14/35	15/20	15/20
8	4	46282/46295	491/1386	491/1386
9	4	63/128	40/61	49/66
10	2	*	147/998	147/998
11	4	*	126/892	198/387
12	3	*	988/2506	*
13	4	76/93	35/46	67/85
13	8	134/169	109/156	80/120
14	2	1109/2248	242/558	289/701
15	4	70638/159377	2853/5081	428/827
16	2	188/377	315/471	104/167
17	4	2879/5795	1755/2347	525/1003
18	4	11/25	16/21	16/20
18	8	118/253	82/128	67/98

the numbers standing for the test problems and the problem dimensions, respectively. The remaining columns are the results for the three corresponding methods.

From the above results, we see that the Cauchy algorithms using diagonal updating are much faster than the standard Cauchy. The simple Oren–Luenberger scaling dramatically improves performance, and the DU-Cholesky diagonal update usually results in a very significant further acceleration.

One can expect similar and quite likely better performance from the diagonal update of Corollary 2.2.1 (whose quasi-Newton counterpart is the BFGS rather than the DFP update).

4. Conclusion. As noted in section 1, any QN or weak-QN update formula can be converted immediately into a diagonal-updating formula. If the original update has hereditary positive definiteness, then the associated diagonal update will retain this property. The diagonal update does not satisfy any curvature condition a priori, and the approach is therefore heuristic—in particular because a QN update does not maintain a Hessian approximation in an element-to-element sense. Nevertheless, the usefulness of this approach within optimization algorithms, when storage is at a premium, has been nicely demonstrated in the works cited earlier, namely, [8] and [9]. Let us identify it by the name *QN-diagonal updating* (and, correspondingly, *weak-QN-diagonal updating* when derived from a weak-QN formula).

In this article we have developed an alternative, variational-based approach with more solid foundations. QC-diagonal updating is an attractive theory whose appeal arises from its simplicity, its elegant solutions, and the similarity of the variational techniques employed to those of QN methods.

We conclude by briefly itemizing some broader issues involving QC-diagonal updating:

- Additional variational principles. We have used only the Frobenius norm in the variational principles of section 2. Other updates can be derived using weighted Frobenius norms, again with variational counterparts in QN updating. Furthermore, a principle based on the deviation from violation of a previous QC relation can be formulated (analogous to the derivation of the LPD QN update; see Mifflin and Nazareth [10]). For more details, see Zhu [20].

  It is also possible to extend both weak-QN-diagonal updating and QC-diagonal updating along lines that parallel work in Yuan and Byrd [18] by substituting a higher-order estimate of curvature for the quantity b in the right-hand side of the weak-QN and QC relations.
- Other applications. When proposing a new algorithmic technique, it is essential to provide a basic (level 1) numerical illustration of viability. We have done this in section 3 for an obvious application—a diagonally preconditioned Cauchy algorithm applied to a standard set of (low-dimensional) MINPACK-1 problems. A much more detailed study of QC-diagonal updating within the limited-memory BFGS algorithm is given in Zhu [20] using more practical MINPACK-2 problems of high dimension. (Some numerical results from this study are briefly summarized in the appendix.) This study has reaffirmed the usefulness of QC-diagonal updating in this setting, thus paralleling the positive experience with QN-diagonal updating mentioned above. One can also envision using a QC-diagonal update within a conjugate gradient iteration (preliminary results along these lines are also reported in [20]) and within a truncated-Newton method.
- Connections to other techniques. Suppose n is not large and evaluating a function/gradient is relatively expensive (a common assumption in nonlinear optimization). Then the cost of solving the nonlinear equation $F(\mu) = b$ in Theorem 2.2.1, which we call the QC subproblem, is essentially trivial even when it is performed by a crude unidimensional algorithm, for example, using bisection. If greater efficiency is needed, it is useful to exploit a connection between problem (*FP*) of section 2.2 and a scaled trust-region subproblem as follows. This connection is particularly ironic because the QC method developed in this article is quintessentially *metric-based*, whereas trust-region techniques are the fundamental building blocks of *model-based* approaches (for terminology see Nazareth [14]).

  Write problem (*FP*) as

$$\text{minimize } ||D_+^{1/2} - D^{1/2}||_F$$

$$\text{s.t. } s^T D_+ s = b > 0.$$

  Then using the earlier definitions

$$E = \text{diag } [s_1^2,\ s_2^2,\ \ldots\ , s_n^2\ ],$$

$$r = [d_1^{1/2},\ d_2^{1/2},\ \ldots,\ d_n^{1/2}\ ]^T$$

  and defining the vector z to be the diagonal elements of $D_+^{1/2}$, we can reexpress the variational problem as follows:

TABLE 3
MINPACK-2 test problems.

Number	Problem name	Par.
1	Elastic-Plastic Torsion	0.5D+01
2	Pressure Distribution in a Journal Bearing	0.1D+00
3	(Enneper's) Minimal Surface	0.0D+00
4	Optimal Design with Composite Materials	0.8D-02
5	Steady-State Combustion	0.1D+01
6	Homog. Superconductors: 2-D Ginzburg–Landau	0.2D+01

$$\text{minimize} \quad -r^T z + \frac{1}{2} z^T z \tag{13}$$

$$\text{s.t.} \quad z^T E z = b,$$

where $b > 0$. When E is *nonsingular* and the equality in the constraint is replaced by a $\leq$ inequality, one obtains a standard trust-region subproblem in the metric defined by $E > 0$. It is likely that many of the techniques used to solve trust-region subproblems—see, in particular, Rendl and Wolkowicz [17]—can be suitably adapted to the task of solving the QC subproblem more efficiently if desired, based on the above interpretation of (*FP*) as a *nonstandard* trust-region problem (13).

- Convergence analysis. Interesting issues remain to be addressed, in particular, the convergence of algorithms that use diagonal updating, the convergence (or not) of diagonal updates to Hessian matrices of functions when these Hessians are themselves diagonal, and the impact of diagonal updating on finite termination of associated algorithms when applied to strongly convex quadratic functions.

Appendix. Some additional numerical experience with QC-diagonal updating within a limited-memory BFGS algorithm is described briefly in this appendix. We employ the MINPACK-2 testbed—a suite of test problems, each of which comes from a real application and is representative of other commonly encountered problems.

MINPACK-2 contains problems from such diverse fields as fluid dynamics, medicine, elasticity, combustion, molecular conformation, nondestructive testing, chemical kinetics, lubrication, and superconductivity; see Averick et al. [1]. In our experiment, we consider a subset of six MINPACK-2 problems (also employed in the study of Burke and Wiegmann [3]), which are suitable for testing the behavior of unconstrained nonlinear optimization algorithms. They are summarized in Table 3. (The first two are unconstrained versions of constrained problems, and the other four are unconstrained problems.) The last column of the table denotes the default parameters for the corresponding problems as used in our testing. For a complete description of these MINPACK-2 problems, see [1].

We give a numerical comparison of the following two limited-memory BFGS algorithms, which differ only in the choice of diagonal scaling matrix used to initiate the L-BFGS upate at each iteration:

- L-BFGS-OL. The diagonal matrix is obtained in the standard way by Oren–Luenberger scaling $y^T s / y^T y$ (for notation, see section 3).
- L-BFGS-DU(C). The diagonal matrix is obtained by QC-diagonal updating of Cholesky factors.

TABLE 4
MINPACK-2, $n = 400$.

Prob.	L-BFGS-OL	L-BFGS-DU(C)	Perf.
1	35/39	33/35	+
2	83/89	68/76	+
3	21/23	28/30	−−
4	58/61	49/56	+
5	45/49	45/50	=
6	204/215	175/193	+

TABLE 5
MINPACK-2, $n = 2{,}500$.

Prob.	L-BFGS-OL	L-BFGS-DU(C)	Perf.
1	89/95	81/84	+
2	185/191	125/162	++
3	77/78	70/71	+
4	230/236	174/201	+
5	120/126	106/108	+
6	480/495	381/423	+

The retention parameter in the two L-BFGS algorithms, i.e., the number m of preserved step/gradient-change pairs over which updating is performed at each iteration, is the standard choice $m = 5$; see Gilbert and Lemaréchal [8]. The line search routine employed is that of Moré and Thuente [13], which was also used in the experiments described in section 3, with its parameters in the strong Wolfe exit conditions (10)–(11) set as follows:

$$\alpha = 10^{-3} \text{ and } \beta = 0.9. \tag{14}$$

For other implementation details, see Zhu [20].

The algorithms used the starting points and stopping criterion of [1] for all tests. Details are again given in [20].

The two limited-memory BFGS algorithms were tested on the MINPACK-2 problems in Table 3 for problems of dimensions 400, 2,500, 10,000, and 40,000; see Tables 4, 5, 6, and 7.

The test results are given in these four tables—each analogous to Table 2—corresponding to the four different choices of problem dimension. Each table reports the results for the two limited-memory BFGS algorithms. The first column records the problem names. Each entry in the second and third columns contains a pair of numbers, namely, the number of iterations and the number of function/gradient calls—the number of times the evaluation routine that returns the function value and gradient vector at a specified point is called—for the corresponding algorithm. The entries in the last column assess relative performance as follows:

- = indicates that the function/gradient counts for the two algorithm are within 5 percent of each other;
- + indicates that the function/gradient count for L-BFGS-DU(C) is better by between 5 and 15 percent;
- ++ indicates that the foregoing count for L-BFGS-DU(C) is better by more than 15 percent;
- − indicates that L-BFGS-OL is better by between 5 and 15 percent;
- −− indicates that L-BFGS-OL is better by more than 15 percent.

TABLE 6
MINPACK-2, $n = 10{,}000$.

Prob.	L-BFGS-OL	L-BFGS-DU(C)	Perf.
1	177/188	113/143	++
2	368/387	237/245	++
3	176/182	94/105	++
4	377/385	244/256	++
5	223/230	143/155	++
6	773/802	769/793	=

TABLE 7
MINPACK-2, $n = 40{,}000$.

Prob.	L-BFGS-OL	L-BFGS-DU(C)	Perf.
1	312/321	319/321	=
2	758/784	710/767	=
3	403/414	449/453	−
4	866/874	1091/1165	−−
5	415/432	312/364	++
6	1444/1502	1360/1405	+

Acknowledgments. We gratefully acknowledge some very useful feedback from the reviewers of this article, which improved it significantly.

REFERENCES

[1] B.M. AVERICK, R.G. CARTER, J.J. MORÉ, AND G. XUE, *The MINPACK-2 Test Problem Collection*, Preprint MCS-P153-0692, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 1992.

[2] D.P. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.

[3] J.V. BURKE AND A. WIEGMANN, *Notes on Limited Memory BFGS Updating in a Trust-Region Framework*, Preprint, Department of Mathematics, University of Washington, Seattle, WA, 1996.

[4] J.E. DENNIS AND R.B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice–Hall, Englewood Cliffs, NJ, 1983.

[5] J.E. DENNIS, JR. AND H. WOLKOWICZ, *Sizing and least-change secant methods*, SIAM J. Numer. Anal., 30 (1993), pp. 1291–1314.

[6] D.C. LIU AND J. NOCEDAL, *On the limited memory BFGS method for large scale optimization*, Math. Programming Ser. B, 45 (1989), pp. 503–528.

[7] D.G. LUENBERGER, *Linear and Nonlinear Programming*, 2nd. ed., Addison–Wesley, Reading, MA, 1994.

[8] J.C. GILBERT AND C. LEMARÉCHAL, *Some numerical experiments with variable-storage quasi-Newton algorithms*, Math. Programming Ser. B, 45 (1989), pp. 407–435.

[9] P.E. GILL AND W. MURRAY, *Conjugate Gradient Methods for Large-Scale Nonlinear Optimization*, Technical Report SOL 79-15, Department of Operations Research, Stanford University, Stanford, CA, 1979.

[10] R.B. MIFFLIN AND J.L. NAZARETH, *The least prior deviation quasi-Newton update*, Math. Programming, 65 (1994), pp. 247–261.

[11] J.J. MORÉ, B.S. GARBOW, AND K.E. HILLSTROM, *Testing unconstrained optimization software*, ACM Trans. Math. Software, 7 (1981), pp. 17–41.

[12] J.J. MORÉ, *Generalizations of the trust region problem*, Optim. Methods Softw., 2 (1993), pp. 189–209.

[13] J.J. MORÉ AND D.J. THUENTE, *Line search algorithms with guaranteed sufficient decrease*, ACM Trans. Math. Software, 20 (1994), pp. 286–307.

[14] J.L. NAZARETH, *The Newton-Cauchy Framework: A Unified Approach to Unconstrained Nonlinear Minimization*, Lecture Notes in Comput. Sci. 769, Springer, New York, 1994.

[15] J.L. Nazareth, *If quasi-Newton then why not quasi-Cauchy?*, SIAG/OPT Views-and-News, 6 (1995), pp. 11–14.

[16] J.L. Nazareth, *The Quasi-Cauchy Method: A Stepping Stone to Derivative-Free Algorithms*, Technical Report 95-3, Department of Pure and Applied Mathematics, Washington State University, Pullman, WA, 1995.

[17] F. Rendl and H. Wolkowicz, *A semidefinite framework for trust region subproblems with applications to large scale minimization*, Math. Programming, 77 (1994), pp. 273–300.

[18] Y. Yuan and R.H. Byrd, *Non-quasi-Newton updates for unconstrained optimization*, J. Comput. Math., 13 (1995), pp. 95–107.

[19] M. Zhu, *Limited Memory BFGS Algorithms with Diagonal Updating*, M.Sc. Project Report, School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, 1997.

[20] M. Zhu, *Techniques for Nonlinear Optimization: Principles and Practice*, Ph.D. dissertation, Department of Pure and Applied Mathematics, Washington State University, Pullman, WA, 1997.