SIAM J. OPTIMIZATION
Vol. 5, No. 1, pp. 192–218, February 1995

©1995 Society for Industrial and Applied Mathematics

010

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

AN OPTIMAL POSITIVE DEFINITE UPDATE FOR SPARSE HESSIAN MATRICES*

R. FLETCHER†

Abstract. A Hessian update is described that preserves sparsity and positive definiteness and satisfies a minimal change property. The update reduces to the BFGS update in the dense case and generalises a recent result in [*SIAM J. Numer. Anal.*, 26 (1989), pp. 727–739] relating to the Byrd and Nocedal measure function. A surprising outcome is that a sparsity projection of the inverse Hessian plays a major role. It is shown that the Hessian itself can be recovered from this information under mild assumptions.

The update is computed by solving a concave programming problem derived by using the Wolfe dual. The Hessian of the dual is important and plays a similar role to the matrix Q that arises in the sparse PSB update of Toint [*Math. Comp.*, 31 (1977), pp. 954–961]. This matrix is shown to satisfy the same structural and definiteness conditions as Toint's matrix. The update has been implemented for tridiagonal systems and some numerical experiments are described. These experiments indicate that there is potential for a significant reduction in the number of quasi-Newton iterations, but that more development is needed to obtain an efficient implementation. Solution of the variational problem by primal methods is also discussed and provides an interesting application of generalized elimination. The possibility of instability and nonexistence of a positive definite update raised by Sorensen [*Math. Programming Study*, 18 (1982), pp. 135–159] is still a difficulty and some remedies are discussed.

Key words. sparse matrix update, positive definite matrix, BFGS formula

AMS subject classifications. 65K, 90C

1. Background and introduction.

This paper primarily relates to quasi-Newton line search methods for finding a solution x^* of the unconstrained optimization problem

$$\text{minimize} \quad f(x) \qquad x \in \mathbb{R}^n. \tag{1.1}$$

These methods generate a sequence of iterates $\{x^{(k)}\}$, $k = 1, 2, \ldots$ by

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)}, \tag{1.2}$$

where $s^{(k)}$ is the current search direction and $\alpha^{(k)}$ is a step chosen to approximately minimize $f(x)$. The methods require the gradient vector $g(x) = \nabla f(x)$ to be available, but not the Hessian matrix $G(x) = \nabla^2 f(x)$. The latter is approximated by a symmetric matrix $B^{(k)}$, which is initially arbitrary and is updated after each iteration. $B^{(k)}$ is used to determine the search direction by solving the system

$$B^{(k)} s^{(k)} = -g^{(k)}, \tag{1.3}$$

where $g^{(k)}$ denotes $g(x^{(k)})$. $B^{(k)}$ is required to be positive definite (written $B^{(k)} > 0$), which implies that $s^{(k)}$ is a descent direction and ensures a reduction in $f(x)$ in the line search.

*Received by the editors November 6, 1992; accepted for publication (in revised form) December 14, 1993. This paper was presented at the Scottish Computational Mathematics Symposium, Strathclyde University, September 14, 1992.

†Department of Mathematics and Computer Science, University of Dundee, DD1 4HN, Scotland, United Kingdom (fletcher@mcs.dundee.ac.uk).

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

An important feature of such methods is the *updating formula* that is used to incorporate new information into $B^{(k)}$. After each iteration difference vectors

$$\delta^{(k)} = x^{(k+1)} - x^{(k)}, \quad \gamma^{(k)} = g^{(k+1)} - g^{(k)} \tag{1.4}$$

can be calculated. It follows from the Taylor series that

$$\gamma^{(k)} = \bar{G}^{(k)}\delta^{(k)}, \tag{1.5}$$

where

$$\bar{G}^{(k)} = \int_0^1 G(x^{(k)} + \theta\delta^{(k)})\, d\theta \tag{1.6}$$

is the average Hessian matrix along the step. By analogy with (1.5), $B^{(k+1)}$ is chosen to satisfy

$$\gamma^{(k)} = B^{(k+1)}\delta^{(k)}, \tag{1.7}$$

known as the quasi-Newton condition. The most popular updating formula is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula $B^{(k+1)} = \text{bfgs}(B^{(k)}, \delta^{(k)}, \gamma^{(k)})$ where

$$\text{bfgs}(B, \delta, \gamma) = B - \frac{B\delta\delta^T B}{\delta^T B\delta} + \frac{\gamma\gamma^T}{\delta^T\gamma} \tag{1.8}$$

and it is clear that (1.7) is satisfied. Moreover, if $B^{(k)} > 0$, then $B^{(k+1)}$ is positive definite if and only if

$$\delta^{(k)T}\gamma^{(k)} > 0. \tag{1.9}$$

The scalar $\delta^{(k)T}\gamma^{(k)}$ represents from (1.5) the component of the average Hessian matrix along $\delta^{(k)}$. It is easily possible to ensure that this condition holds.

A significant result due to Goldfarb [6] is that if $H^{(k)}$ denotes $B^{(k)-1}$, then the correction $E = H^{(k+1)} - H^{(k)}$ in the BFGS formula satisfies a minimum property with respect to a weighted Frobenius norm of the form

$$\|E\|_W^2 = \|W^{\frac{1}{2}}EW^{\frac{1}{2}}\|_F^2 = (\text{trace}(EWEW))^{\frac{1}{2}}, \tag{1.10}$$

where $W > 0$ and $W\delta^{(k)} = \gamma^{(k)}$. This result can be interpreted as showing that $H^{(k)}$ is changed by the minimum amount (in the sense of (1.10)) required to satisfy (1.7) and symmetry. This ensures that previous information accumulated in $B^{(k)}$ is disturbed as little as possible. The well-known Davidon–Fletcher–Powell (DFP) formula can also be interpreted in a similar way. Another formula that satisfies a minimum correction property is the Powell-symmetric-Broyden (PSB) formula

$$B^{(k+1)} = B^{(k)} + \frac{\eta\delta^T + \delta\eta^T}{\delta^T\delta} - \frac{\eta^T\delta}{(\delta^T\delta)^2}\delta\delta^T, \tag{1.11}$$

where $\eta = \gamma - B^{(k)}\delta$ and where δ and γ denote $\delta^{(k)}$ and $\gamma^{(k)}$, respectively. In this case it is the Frobenius norm ($W = I$ in (1.10)) of the correction to $B^{(k)}$ that is minimized (subject to (1.7) and symmetry). Unfortunately the PSB update does not generally preserve $B^{(k)} > 0$ and practical experience has been disappointing. This is thought to be due to the lack of certain affine invariance properties that hold for the BFGS and DFP formulae. More detail about all the above subject matter is given, for example, in [4].

Quasi-Newton methods become less attractive when n is very large because of the storage and computational requirements associated with large dense matrices. However, it is often the case that the Hessian is a sparse matrix and it is attractive to look for updating formulae that preserve the same sparsity in $B^{(k)}$. Thus we express the sparsity conditions on B as

$$B_{ij} = 0 \quad \forall\,(i,j) \in \mathcal{S}, \tag{1.12}$$

where $\mathcal{S}$ is a set of pairs of integers in the range $[1 : n]$. Because of symmetry it is assumed that $(i,j) \in \mathcal{S}$ if and only if $(j,i) \in \mathcal{S}$. It is also assumed that $G(x)$ satisfies (1.12) for all $x \in \mathbb{R}^n$. The complementary set of index pairs not in $\mathcal{S}$ is denoted by $\mathcal{S}^\perp$. Because we are concerned with positive definite matrices, it is assumed that

$$(i,i) \in \mathcal{S}^\perp \quad i = 1, 2, \ldots, n. \tag{1.13}$$

For such problems it is fruitful to determine a minimum correction update formula that is constrained by (1.12) in addition to the quasi-Newton condition and symmetry. In a seminal paper, Toint [10] shows that it is reasonably straightforward to compute a minimum correction to $B^{(k)}$ in the Frobenius norm subject to these conditions. To present Toint's update, we define the projection operator $\mathcal{G}(M)\ :\ \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ by

$$\mathcal{G}(M)_{ij} = \begin{cases} 0 & (i,j) \in \mathcal{S}\ , \\ M_{ij} & (i,j) \in \mathcal{S}^\perp. \end{cases} \tag{1.14}$$

This has been colourfully dubbed the *gangster operator* since it shoots holes in M according to the sparsity pattern defined by (1.12). Toint also introduces the notation

$$\delta_{[i]} = \mathcal{G}(\delta e_i^T) e_i, \tag{1.15}$$

where e_i denotes the unit vector that is column i of I. (The use of subscript $[i]$ is due to Coleman [2] in his very readable monograph on large sparse optimization.) Toint [10] shows that the resulting minimum correction satisfies

$$B^{(k+1)} = B^{(k)} + \mathcal{G}(\delta\lambda^T + \lambda\delta^T), \tag{1.16}$$

where $\lambda \in \mathbb{R}^n$ is obtained by solving the linear system

$$Q\lambda = r \tag{1.17}$$

in which $r = \gamma - B^{(k)}\delta$. It follows from (1.16) and (1.7) that column i of the matrix Q is defined by

$$Qe_i = \delta_i \delta_{[i]} + \delta_{[i]}^T \delta_{[i]} e_i. \tag{1.18}$$

It is easily shown that Q is symmetric positive semidefinite and that $\mathcal{G}(Q) = Q$ (i.e. Q satisfies the sparsity conditions (1.12)). In addition, Q is positive definite if

$$\delta_{[i]} \neq 0 \quad i = 1, 2, \ldots, n, \tag{1.19}$$

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

in which case sparse LDL^T factors of Q are calculated and (1.17) can readily be solved to obtain λ. If Q is singular then $\delta_{[i]} = 0$ for some i. However, it then follows from (1.5) that $\gamma_i = 0$ and hence $r_i = 0$. Thus (1.17) is consistent and can be solved by deleting row and column i from Q (ignoring the effects of round-off error).

As with the dense PSB update, the condition $B^{(k)} > 0$ is not preserved by this update, and likewise practical performance has not been outstanding. In view of this, it is natural to inquire what happens when the sparsity conditions (1.12) are included in the minimum correction property that defines the BFGS or DFP formula. Unfortunately, as Toint [10] points out, the use of a weighted Frobenius norm leads to formulae that are intractable in both cases (see, also, [11] and [2] for more details). However, Toint [11] proves that if

$$\delta_i \neq 0 \quad i = 1, 2, \ldots, n, \tag{1.20}$$

if G is irreducible, and if $\delta^T\gamma > 0$, then there does exist a symmetric update that preserves positive definiteness. The condition $\delta^T\gamma > 0$ is clearly required, else (1.7) would imply $\delta^T B\delta \leq 0$, contradicting $B > 0$. The assumption of irreducibility (that is, G cannot be reduced to block diagonal form by a symmetric permutation) is not a serious restriction because if G is reducible then (1.1) can be decomposed into two or more problems which can be solved separately. It is assumed throughout what follows that G is irreducible. On the other hand, (1.20) is critical and Sorensen [9] shows that a positive definite update may not exist if $\delta_i = 0$ for some i, and that serious growth in B can occur in a neighbourhood of this situation. We return to these points later in the paper.

More recent research into sparse updates has avoided the requirement that $B^{(k)}$ should be positive definite. Most promising has been the approach of Griewank and Toint (e.g., [7]) based on the partially separable optimization problem

$$\text{minimize} \quad f(x) = \sum_{i=1}^{m} f_i(x), \tag{1.21}$$

in which each *element function* $f_i(x)$ depends on only a few of the components of x. Then the Hessian of $f(x)$ can be decomposed into a sum of Hessians of the $f_i(x)$, the nontrivial submatrices of which can be treated as dense matrices. Similar remarks apply for the gradient vector. Griewank and Toint suggest that these element Hessian submatrices are approximated by the use of dense updating techniques. There are, however, some difficulties that must be overcome. The element Hessian submatrices may not be positive definite so it is not possible to rely on the analogue of the condition $\delta^T\gamma > 0$ holding for each submatrix. Thus the BFGS formula cannot be used, and Toint uses the symmetric rank-one formula in the Harwell Subroutine Library code VE08. Consequently, the overall Hessian approximation obtained by summing the submatrix approximations is also not generally positive definite. Also the possibility of zero in the denominator of the updates must be allowed for. Another aspect is that the user must specify the decomposition (1.21) to the code and this is not always convenient. These remarks are not intended to disparage the method, which has been very successful in practice, but they do indicate that if an effective sparse positive definite update were available, then many of these difficulties would be circumvented.

This paper makes a contribution to this objective by providing an update that preserves sparsity and positive definiteness and satisfies a minimal change property. The update reduces to the BFGS update in the dense case. However, the amount of

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

computation required to compute the update is not trivial (although improvements here may well be possible) and the difficulties noted by Sorensen are still present. The new update arises from a recent observation of Fletcher [5] that the BFGS and DFP formulae can be derived by a variational argument using the measure function

$$\psi(A) = \mathrm{trace}(A) - \ln\det(A), \tag{1.22}$$

where ln denotes the natural logarithm. This function is introduced by Byrd and Nocedal [1] in the convergence analysis of quasi-Newton methods. The function is strictly convex on the set of positive definite matrices and is minimized by $A = I$. The function becomes unbounded as A becomes singular or infinite and so acts as a barrier function that keeps A positive definite. A suitable variational property is to minimize $\psi(H^{(k)}B)$ since, in the absence of any constraints, the solution is just $B = H^{(k)-1}$. Introducing the constraints (1.7) and (1.12) leads to an update in which $H = B^{-1}$ stays close to $H^{(k)}$ in some sense. The objective function can also be expressed as

$$\psi(H^{(k)}B) = \psi(BH^{(k)}) = \psi({H^{(k)}}^{1/2}B{H^{(k)}}^{1/2})$$

using the properties of the trace and determinant. In the sparse case it is also noted that $\psi(H^{(k)}B)$ can be computed from only B and $\mathcal{G}(H^{(k)})$ and the full matrix $H^{(k)}$ is not required.

A theorem extending the BFGS result in [5] to include the sparsity conditions (1.12) is set out in §2. Necessary conditions related to a rank-two correction of the form $\delta\lambda^T + \lambda\delta^T$ are derived. However, there is also a surprising outcome in that the matrix $\mathcal{G}(H)$ is seen to play a major role. The issue of whether $\mathcal{G}(H)$ determines B, and how this calculation can be carried out, is seen to be fundamental to the update. It is shown that the update can be determined by solving a nonlinear system $r(\lambda) = 0$ involving the residual of the quasi-Newton condition. Unfortunately, the DFP formula cannot be generalised in the same way.

Issues concerning the determination of B from $\mathcal{G}(H)$ are considered in §3. A simplifying assumption is made that the sparsity pattern specified by $\mathcal{S}$ is such that elements that fill in during the calculation of LDL^T factors of B are in $\mathcal{S}^{\perp}$. In the notation of Duff, Erisman, and Reid [3] this can be expressed as

$$\mathcal{G}(L\backslash L^T) = L\backslash L^T. \tag{1.23}$$

Another way of expressing this is that B does not fill in with respect to $\mathcal{S}$ when factored. This assumption is not very restrictive in practice since factors of B are required in (1.3) to determine the search direction, which necessitates using a data structure for B that allows for fill-in. With this assumption it is shown that B is readily determined from $\mathcal{G}(H)$. The inverses of certain submatrices of H, referred to as Markowitz submatrices, are shown to play an important role.

In §4 the solution of the system $r(\lambda) = 0$ is considered. It is shown that $r(\lambda)$ is the gradient of a concave programming problem derived by using the Wolfe dual and this enables the solution of $r(\lambda) = 0$ to be undertaken in a reliable way. The Jacobian $Q(\lambda)$ of this system is important and plays a similar role to the matrix Q that arises in the linear system (1.17) in Toint's sparse PSB update. The structure of Q is analysed in detail in §5, and is shown to satisfy the same structural and definiteness conditions (when (1.23) holds) as for Toint's matrix. Thus the nonlinear system can be solved by a few iterations of analogous complexity to (1.17).

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The update has been implemented for tridiagonal systems and some numerical experiments are described in §6. These experiments indicate that there is the potential for a significant reduction in the number of quasi-Newton iterations, but that more development is needed to obtain an efficient implementation. Section 7 discusses the possibility of using primal algorithms to determine the update and provides an interesting application of generalized elimination. This leads to Toint's result on the existence of a positive definite update subject to (1.20). The issue of stability raised by Sorensen is discussed in §8, together with other points of interest including a conjecture related to partially separable updates. Directions for further research are suggested.

2. A variational result.

The main aim of this section is to extend Theorem 2.1 of [5] to include the sparsity conditions (1.12).

THEOREM 2.1. *Let $B^{(k)}$ be positive definite and consider the solution of the variational problem*

$$\begin{aligned} &(2.1) \qquad && \underset{B>0}{\textit{minimize}} && \psi(H^{(k)}B) \\ &(2.2) && \textit{subject to} && B^T = B, \\ &(2.3) && && B\delta = \gamma, \\ &(2.4) && && B_{ij} = 0 \quad \forall\, (i,j) \in \mathcal{S}. \end{aligned}$$

If a solution exists it is characterised by the existence of $\lambda \in \mathbb{R}^n$ such that

$$\mathcal{G}(H) = \mathcal{G}(H^{(k)} + \lambda\delta^T + \delta\lambda^T), \tag{2.5}$$

where H denotes B^{-1}.

Proof. If a solution to the variational problem exists, it satisfies $B > 0$. Because the remaining constraints in the problem are linear, constraint qualification holds and first order conditions obtained by the method of Lagrange multipliers are necessary for a solution. A suitable Lagrangian function is

$$\begin{aligned} \mathcal{L}(B,\Lambda,\lambda,\Pi) &= \tfrac{1}{2}\psi(H^{(k)}B) + \operatorname{trace}(\Lambda^T(B^T - B)) + \lambda^T(B\delta - \gamma) + \tfrac{1}{2}\operatorname{trace}(\Pi^T B) \\ &= \tfrac{1}{2}(\operatorname{trace}(H^{(k)}B) - \ln\det H^{(k)} - \ln\det B) + \operatorname{trace}(\Lambda^T(B^T - B)) \\ &\quad + \lambda^T(B\delta - \gamma) + \tfrac{1}{2}\operatorname{trace}(\Pi^T B), \end{aligned}$$

where Λ, λ, and Π are Lagrange multipliers for (2.2), (2.3), and (2.4) respectively. Without loss of generality it can be assumed that Λ is strictly lower triangular and Π is symmetric. Because $B_{ij} = 0$ does not apply for $(i,j) \in \mathcal{S}^\perp$, it follows that

$$\mathcal{G}(\Pi) = 0. \tag{2.6}$$

To solve the first order conditions it is necessary to find B, Λ, λ, and Π to satisfy (2.2), (2.3), (2.4), and the equations $\partial\mathcal{L}/\partial B_{ij} = 0$. Using the identity $\partial B/\partial B_{ij} = e_i e_j^T$ and Lemma 1.4 of [5], it follows that

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial B_{ij}} = 0 &= \tfrac{1}{2}(\operatorname{trace}(H^{(k)}e_i e_j^T) - (B^{-1})_{ji}) + \operatorname{trace}(\Lambda^T(e_j e_i^T - e_i e_j^T)) \\ &\quad + \lambda^T e_i e_j^T \delta + \tfrac{1}{2}\operatorname{trace}(\Pi^T e_i e_j^T) \\ &= \tfrac{1}{2}((H^{(k)})_{ji} - (B^{-1})_{ji}) + \Lambda_{ji} - \Lambda_{ij} + (\lambda\delta^T)_{ij} + \tfrac{1}{2}\Pi_{ij}. \end{aligned}$$

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Transposing and adding, using the symmetry of $H^{(k)}$, B and Π, gives

$$H^{(k)} - B^{-1} + \lambda\delta^T + \delta\lambda^T + \Pi = 0$$

or

$$H = H^{(k)} + \lambda\delta^T + \delta\lambda^T + \Pi. \tag{2.7}$$

Then (2.5) is deduced directly from (2.6) and (2.7). □

Although the derivation of this result is straightforward, the outcome came as a major surprise to me for the following reason. If B is an irreducible sparse matrix then B^{-1} is generally dense. It is therefore most unexpected to find that B is determined by $\mathcal{G}(H)$ (i.e., by zeroing elements of B^{-1} in accordance with the sparsity pattern of B).

The result for the dense case given in [5] corresponds to $\Pi = 0$ and shows that the optimum H matrix involves a rank-two correction of $H^{(k)}$. In that case it is possible to directly solve for H using $B\delta = \gamma$ and the resulting update is the BFGS formula. Because the resulting B matrix is positive definite and ψ is a convex function it is possible to deduce that it solves the variational problem.

When the sparsity conditions (2.4) are included, solution of (2.7) is no longer straightforward because of the additional term Π, and a finite calculation to determine λ does not appear to be possible. However, an iterative approach along the following lines can be envisaged. The following sequence of operations defines r as a function of λ ($\mathbb{R}^n \to \mathbb{R}^n$).

```
Given $\lambda$,
calculate $\mathcal{G}(H)$ from (2.5),
find $B > 0$ such that $\mathcal{G}(B^{-1}) = \mathcal{G}(H)$,
calculate $r := B\delta - \gamma$.
```

The update is determined by finding λ such that

$$r(\lambda) = 0, \tag{2.8}$$

which is a nonlinear system of n equations in n variables.

This discussion raises a number of interesting and complex issues that are addressed in the rest of the paper. First is the question as to whether $\mathcal{G}(H)$ does indeed determine $B > 0$ and whether the outcome is unique. It is hopeful that $\mathcal{G}(H)$ contains the same number of nonzero elements as B. A solution is given to this question in §3 in the case that assumption (1.23) holds. An illustration of the calculation is given in §6 for the case of tridiagonal matrices. However, when assumption (1.23) does not hold, then the question is as yet unresolved, although it is shown that the condition $(i, i) \in \mathcal{S}^{\perp}$ is necessary.

Given that $B > 0$ is well determined, the next question is that of whether a solution to the variational problem, and hence to (2.8), does exist. The contributions of Toint and Sorensen provide a complete answer to this question as discussed in §1. Toint's result can be seen as a consequence of the presentation on primal algorithms set out in §7 and the issues are discussed in more detail in §8.

Another question relates to the practicality of solving (2.8) reliably and rapidly. At first sight this is not promising since a nonlinear system might be as hard to solve as the original problem (1.1). However, it is shown in §§4 and 5 that there are features

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

that enable (2.8) to be solved effectively when assumption (1.23) holds and a solution exists. Preliminary numerical experiments described in §6 indicate that the extra expense of solving (2.8) is compensated for by a significant reduction in the number of line searches required to solve (1.1).

Before following up these questions, it is worth remarking that another result in [5] regarding the DFP update does not carry over conveniently to the sparse case. Extending the corollary to Theorem 2.1 in [5] gives the variational problem

$$\begin{aligned} &(2.9) && \underset{H>0}{\text{minimize}} && \psi(B^{(k)}H), \\ &(2.10) && \text{subject to} && H^T = H \\ &(2.11) && && H\gamma = \delta, \\ &(2.12) && && B_{ij} = 0 \quad \forall\,(i,j) \in \mathcal{S}. \end{aligned}$$

After proceeding analogously to Theorem 2.1 here, and setting $\partial\mathcal{L}/\partial H_{ij} = 0$, it follows that

$$B = B^{(k)} + \lambda\gamma^T + \gamma\lambda^T + B\Pi B. \tag{2.13}$$

It is possible to pre and post multiply by H and operate with $\mathcal{G}(.)$ to eliminate Π, but the resulting term $\mathcal{G}(HB^{(k)}H)$ appears to make further progress unlikely.

3. Determination of B from $\mathcal{G}(H)$. The results of this section are a consequence of assumption (1.23) discussed in §1 that $\mathcal{S}$ is chosen such that there is no fill-in when the LDL^T factors of B are calculated. The main result shows that B is well determined by $\mathcal{G}(H)$ and provides the basis of an algorithm for computing the LDL^T factors of B. This algorithm is shown to be particularly efficient when the sparsity pattern of B is formed from dense overlapping blocks on the diagonal.

A lemma is required that shows the effect on the inverse when bordering a partitioned matrix with a rank-one term having some sparsity.

LEMMA 3.1. *Consider symmetric matrices partitioned conformally as*

$$A = \begin{bmatrix} 0 & 0^T & 0^T \\ 0 & A_{11} & A_{12} \\ 0 & A_{21} & A_{22} \end{bmatrix} + \begin{pmatrix} 1 \\ a/\alpha \\ 0 \end{pmatrix} \begin{pmatrix} \alpha & a^T & 0^T \end{pmatrix}, \quad X = \begin{bmatrix} \xi & x_1^T & x_2^T \\ x_1 & X_{11} & X_{12} \\ x_2 & X_{21} & X_{22} \end{bmatrix} \tag{3.1}$$

($A_{21} = A_{12}^T$, etc.) in which

$$\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} > 0 \tag{3.2}$$

is positive definite and

$$\begin{bmatrix} \xi & x_1^T \\ x_1 & X_{11} \end{bmatrix} > 0 \tag{3.3}$$

is positive definite. Then a necessary and sufficient condition for $A = X^{-1}$ is that both

$$\begin{bmatrix} \xi & x_1^T \\ x_1 & X_{11} \end{bmatrix} \begin{pmatrix} \alpha \\ a \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{3.4}$$

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

and

$$x_2 = -A_{22}^{-1} A_{21} x_1 \tag{3.5}$$

hold. Moreover α, a, *and* x_2 *are determined uniquely by* ξ, x_1, *and* A_{11}, A_{12}, *and* A_{22}.

Proof. First of all it is readily established from the properties of the inverse in (3.2) that

$$A_{12} A_{22}^{-1} = -X_{11}^{-1} X_{12} \tag{3.6}$$

and

$$X_{11}^{-1} = A_{11} - A_{12} A_{22}^{-1} A_{21}. \tag{3.7}$$

Also the solution of (3.4) can be expressed as

$$\begin{pmatrix} \alpha \\ a \end{pmatrix} = \begin{pmatrix} 1 \\ -X_{11}^{-1} x_1 \end{pmatrix} / (\xi - x_1^T X_{11}^{-1} x_1). \tag{3.8}$$

To prove the main result, we form the product

$$XA = \begin{bmatrix} \xi & x_1^T \\ x_1 & X_{11} \\ x_2 & X_{21} \end{bmatrix} \begin{pmatrix} \alpha \\ a \end{pmatrix} (1 \quad a^T/\alpha \quad 0^T) + \begin{bmatrix} 0 & x_1^T A_{11} + x_2^T A_{21} & x_1^T A_{12} + x_2^T A_{22} \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}.$$

Now $A = X^{-1}$ if and only if $XA = I$, which is seen to require that both (3.4) and (3.5) hold. Conversely, if (3.4) and (3.5) hold, then it follows that

$$x_2 \alpha + X_{21} a = (x_2 - X_{21} X_{11}^{-1} x_1) \alpha = 0$$

from (3.5), (3.6), and (3.8) and

$$A_{11} x_1 + A_{12} x_2 + a/\alpha = (A_{11} - A_{12} A_{22}^{-1} A_{21} - X_{11}^{-1}) x_1 = 0$$

from (3.5), (3.7), and (3.8), showing that $XA = I$. Thus the main result is established. It is clear from (3.4) and (3.5) and the existence of the relevant inverses that α, a, and x_2 are determined uniquely by ξ, x_1, and A_{11}, A_{12}, and A_{22}. □

Some other items of terminology are introduced to simplify the description of the main result. When factors $B = LDL^T$ (L unit lower triangular, D diagonal, $B \in \mathbb{R}^{n \times n}$) are calculated by Gaussian elimination, then it is convenient to refer to

$$B^{(i)} = B - \sum_{j=1}^{i-1} l_j d_j l_j^T = \begin{bmatrix} 0 & 0 \\ 0 & B_{22}^{(i)} \end{bmatrix} \tag{3.9}$$

as the ith reduced matrix in the calculation. Here l_i denotes the ith column of L and d_i the ith diagonal element of D.

DEFINITION. *Let sparse factors* $B = LDL^T$ *exist. The ith* Markowitz submatrix *of any matrix* M *of the same dimension as* B *is defined to be the submatrix obtained by selecting elements of* M *corresponding to the structural nonzero elements of* $l_i l_i^T$.

The definition is illustrated by the following matrices:

$$\begin{bmatrix}
\ast & & \ast & & & \ast & & & \\
& \ast & & & \ast & & \ast & \ast & \\
\ast & & \ast & & & \ast & \ast & & \ast \\
& & & \ast & & \ast & & & \ast \\
& \ast & & & \ast & & \ast & \ast & \\
\ast & & \ast & \ast & & \ast & \ast & & \ast \\
& \ast & \ast & & \ast & \ast & \ast & \ast & \ast \\
& \ast & & & \ast & & \ast & \ast & \ast \\
& & \ast & \ast & & \ast & \ast & \ast & \ast
\end{bmatrix}, \qquad
\begin{bmatrix}
& & & & & & & & \\
& & & & & & & & \\
& & \bullet & & & \bullet & \bullet & & \bullet \\
& & & \ast & & \ast & & & \ast \\
& & & & \ast & & \ast & \ast & \\
& & \bullet & \ast & & \ast & \ast & & \ast \\
& & \bullet & & \ast & \ast & \ast & \ast & \ast \\
& & & & \ast & & \ast & \ast & \ast \\
& & \bullet & \ast & & \ast & \ast & \ast & \ast
\end{bmatrix},$$

Matrix B satisfying (1.23) Reduced matrix $B^{(3)}$

$$\begin{bmatrix}
& & & & & & & & \\
& & & & & & & & \\
& & \star & & & \star & \star & & \star \\
& & & & & & & & \\
& & & & & & & & \\
& & \star & & & \star & \star & & \star \\
& & \star & & & \star & \star & & \star \\
& & & & & & & & \\
& & \star & & & \star & \star & & \star
\end{bmatrix}.$$

Expanded Markowitz submatrix $M_{[3]}$

The $\bullet$ elements in $B^{(3)}$ are those that determine the vector l_3 and the diagonal element d_3. If M is any $n \times n$ matrix, then $M_{[3]}$ is the matrix obtained by selecting the $\star$ elements of M and zeroing the others. The pattern of these elements is related to the possible fill-in caused by the $\bullet$ elements when processing the reduced matrix $B^{(3)}$ (hence the term Markowitz submatrix). It is convenient to extend Toint's [10] notation by using $M_{[i]}$ to refer to the expanded ($n \times n$) ith Markowitz submatrix of M as in the above figure, and M_i to refer to the (dense) submatrix itself. The requirement that no fill-in is caused when factorizing B is equivalent to requiring that the pattern of nonzeros of B is that obtained from the overlay of all the matrices $B_{[i]}$.

The main theorem of this section can now be proved. The matrices B and H are assumed to be symmetric and the sparsity pattern of the Markowitz submatrices is that induced by the structure of the LDL^T factorization of B. The theorem shows that B is uniquely determined by $\mathcal{G}(H)$ and provides a construction in which the factors of B are expressed directly in terms of the Markowitz submatrices $H_{[i]}$. To do this the generalized inverse notation $H^{+}_{[i]}$ is used. Because H_i is assumed to be positive definite, $H^{+}_{[i]}$ consists simply of the elements of the matrix H_i^{-1} scattered into the positions corresponding to nonzero elements of $H_{[i]}$.

THEOREM 3.1. *Let assumption* (1.23) *hold. If for* $i = 1, 2, \ldots, n$ *the Markowitz submatrices* H_i *are positive definite, then the ith diagonal element of* D *and the ith column of* L *are defined by*

$$d_i = e_i^T H^{+}_{[i]} e_i, \qquad l_i = H^{+}_{[i]} e_i / d_i. \tag{3.10}$$

Moreover, B *is uniquely determined by* $\mathcal{G}(H)$ *and* B *is positive definite.*

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. Consider the reduced matrix $B^{(i)}$ in the calculation of $B = LDL^T$ and partition

$$B^{(i)} = \begin{bmatrix} 0 & 0 \\ 0 & B_{22}^{(i)} \end{bmatrix}, \qquad H = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix}$$

as indicated by (3.9). The notation $H_{11}^{(i)}$, etc. just indicates a different partitioning of the same H matrix as i changes. The theorem only assumes that $\mathcal{G}(H)$ is known, and shows that the unknown elements of H can be deduced. Also from (3.9) it follows that

$$B^{(i)} = B^{(i+1)} + l_i d_i l_i^T, \tag{3.11}$$

where, because of assumption (1.23), column l_i has the same sparsity structure as column i of B and hence as column i of $H_{[i]}$.

The main step is to prove by induction that

$$B_{22}^{(i)} = H_{22}^{(i)-1} > 0 \text{ and is determined uniquely by } H_i, \dots, H_n. \tag{3.12}$$

When $i = n$, $H_n > 0$ is just a scalar and $B_{22}^{(n)} = H_n^{-1}$ determines $B_{22}^{(n)}$ uniquely. Now we assume that (3.12) is true with i replaced by $i+1$ and deduce that (3.12) itself is true. $H_{22}^{(i)}$ is obtained by bordering $H_{22}^{(i+1)}$ with some elements from column i of H. To apply Lemma 3.1 to $H_{22}^{(i)}$, it can be assumed without loss of generality that a symmetric row and column permutation is made to $H_{22}^{(i)}$ so that the elements of the Markowitz submatrix H_i occur in adjacent rows and columns. Thus we can identify H_i with the matrix

$$\begin{bmatrix} \xi & x_1^T \\ x_1 & X_{11} \end{bmatrix} \tag{3.13}$$

in Lemma 3.1. Note that the vector x_2 in (3.1) represents the unknown elements from H that occur in the border. The positive definiteness of (3.13) follows from the same assumption about H_i. Moreover because the structure of l_i is prescribed, and matches that of column i of $H_{[i]}$, we can identify the subdiagonal part of l_i with $\binom{a}{0}$ and also $d_i = \alpha$. In addition we have

$$B_{22}^{(i+1)} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad H_{22}^{(i+1)} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}.$$

It follows from (3.12) with i replaced by $i+1$ that these matrices satisfy condition (3.2) of Lemma 3.1. Thus the lemma can be invoked and it follows that a necessary and sufficient condition to obtain $B_{22}^{(i)} = H_{22}^{-1}$ is that (3.4) and (3.5) hold. It follows from (3.4) and the identification of H_i with (3.13) that l_i and d_i are defined by (3.10). Equation (3.5) determines the unknown elements of H represented by x_2, although these are not required to compute the factors of B. Because of (3.10) and H_i being positive definite, it follows that $d_i > 0$ and hence by induction that $B_{22}^{(i)}$ is positive definite. It follows from the lemma that d_i and l_i are uniquely defined by $B_{22}^{(i+1)}$ and the first column of H_i. Hence by induction $B_{22}^{(i)}$ depends on $H_i, \dots, H_n$ and (3.12) is established.

Finally for $i = 1$, $B = B^{(1)} = H^{-1}$ and the LDL^T factors of B have been determined uniquely from $H_1, \dots, H_n$. Because of assumption (1.23), the elements of the Markowitz submatrix are available in $\mathcal{G}(H)$ to make the calculation. □

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Note that this proof does not require a prior assumption that H is positive definite. Rather this can be deduced as a consequence of the positive definiteness of the Markowitz submatrices H_i and the inductive argument.

It can be observed that the construction calculates the LDL^T factors of B rather than B itself. This is convenient as B is subsequently used to solve systems. It is also possible to express

$$B = LDL^T = \sum_{i=1}^{n} l_i d_i l_i^T = \sum_{i=1}^{n} \frac{H_{[i]}^{+} e_i e_i^T H_{[i]}^{+}}{e_i^T H_{[i]}^{+} e_i}. \tag{3.14}$$

This form is particularly useful for the purposes of §5.

In the case that the pattern of nonzeros in B consists of overlapping dense diagonal blocks then a particularly efficient algorithm is determined. Special cases of such patterns are the symmetric band matrices of arbitrary bandwidth. The columns of D and L are determined in the order $1, 2, \ldots, n$, and for each i, USU^T factors of H_i are available, where U is upper triangular and $S > 0$ is diagonal. This corresponds to having factorized H_i taking pivots in the reverse order. Then $H_i^{-1} = U^{-T} S^{-1} U^{-1}$ and it is readily observed that $d_i = S_{11}^{-1}$ and the nonzero part of l_i is obtained by solving the system $U^T x = e_1$. Moreover, advantage can be taken of the overlap in the Markowitz submatrices of H in the following way. When i is incremented, the first row and column of U and S are deleted. If i changes to move into the next overlapping block, then the remaining part of the USU^T factors is bordered by elements of a new submatrix. When factorising this submatrix, a low rank matrix is added into the old USU^T factors that can be updated efficiently by the use of square root free Givens' rotations.

It may be that a similar approach can be used when the pattern of nonzeros is less regular, but the organization is likely to be more complex. It is hoped to investigate this aspect in the near future. An observation of some interest is that Duff, Erisman, and Reid [3, §12.7] reference a method of calculation whereby $\mathcal{G}(A^{-1})$ can be calculated from the LU factors of A under the same assumptions about fill-in. The construction given in Theorem 3.1 can be regarded as reversing this calculation in the symmetric case.

Finally, the case when assumption (1.23) does *not* hold is of some interest. The above scheme no longer applies because the elements of the Markowitz submatrix H_i are not always available to calculate column i of $B_{22}^{(i)}$. However, B and $\mathcal{G}(H)$ still have the same number of nonzero elements and it is conjectured that B will remain well determined if H is a positive definite matrix. For a general unsymmetric matrix A, A is not always well determined by $\mathcal{G}(A^{-1})$ as the following example shows.

$$A = \begin{bmatrix} a & b \\ c & 0 \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} 0 & 1/c \\ 1/b & -a/bc \end{bmatrix}, \quad \mathcal{G}(A^{-1}) = \begin{bmatrix} 0 & 1/c \\ 1/b & 0 \end{bmatrix}.$$

Here we have $\mathcal{S} = \{(2,2)\}$ (relaxing the condition $(i,i) \in \mathcal{S}^{\perp}$). When the $(2,2)$ element of A^{-1} is deleted, then all reference to a is lost and A is not uniquely defined by $\mathcal{G}(A^{-1})$.

4. Solving the system $r(\lambda) = 0$.

This section considers the practicality of finding a reliable and efficient method for solving the nonlinear system $r(\lambda) = 0$ generated by the prototype algorithm that precedes (2.8). The method we consider is based on Newton's method with iterates $\lambda^{(t)}$, $t = 1, 2, \ldots$, and an iteration formula

$$Q^{(t)} \Delta\lambda^{(t)} = -r(\lambda^{(t)}), \tag{4.1}$$

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $Q^{(t)}$ denotes the Jacobian of $r(\lambda)$ evaluated at $\lambda^{(t)}$. At first sight this is not promising since a nonlinear system might be at least as hard to solve as the main problem (1.1), but it turns out that there are some simplifying features that can be exploited.

It is observed in [5] that the problem posed in (2.1), (2.2), and (2.3) is a convex programming problem and significant progress is made by examining the Wolfe dual (see, for example, [4]). The Wolfe dual is

$$\text{(4.2)} \qquad \underset{B,\Lambda,\lambda,\Pi}{\text{maximize}} \quad \mathcal{L}(B,\Lambda,\lambda,\Pi)$$

$$\text{(4.3)} \qquad \text{subject to} \quad \nabla_B \mathcal{L} = 0,$$

and it has been shown in Theorem 2.1 that (4.3) implies that B $(= H^{-1})$ is defined by (2.5). Section 3 gives a scheme for calculating a positive definite matrix B from $\mathcal{G}(H)$ and this enables the variable B in the dual to be expressed as a function $B(\lambda)$. The terms involving Λ and Π are eliminated by virtue of the symmetry and sparsity of $B(\lambda)$, giving rise to the more simple dual problem

$$\text{(4.4)} \qquad \begin{aligned} &\underset{\lambda}{\text{maximize}} && \phi(B(\lambda),\lambda) = \tfrac{1}{2}\psi(H^{(k)}B) + \lambda^T(B\delta - \gamma) \\ &\text{subject to} && B(\lambda) > 0. \end{aligned}$$

The chain rule gives

$$\frac{d\phi}{d\lambda_k} = \sum_{(i,j)\in\mathcal{S}^{\perp}} \frac{\partial\phi}{\partial B_{ij}} \frac{\partial B_{ij}}{\partial\lambda_k} + \frac{\partial\phi}{\partial\lambda_k}$$

and $\partial\phi/\partial B_{ij} = 0$ because of the stationary property of the Lagrangian in Theorem 2.1. It follows from (4.4) that

$$\text{(4.5)} \qquad \nabla_\lambda \phi = B\delta - \gamma = r$$

so that the nonlinear system $r(\lambda) = 0$ that occurs in (2.8) is seen to arise from the stationary point condition for (4.4). It follows that the Jacobian $Q(\lambda)$ of $r(\lambda)$ is the Hessian of $\phi(\lambda)$ and hence is a symmetric matrix. In fact, it is shown in §5 that Q is a negative semidefinite matrix (and usually negative definite) when $B > 0$, so the simplified dual is a concave programming problem. Thus ϕ provides an objective function with which to measure progress when solving (2.8), and enables a line search in λ to be used. It is only necessary to ensure that λ is chosen so that all the Markowitz submatrices of $H^{(k)} + \lambda\delta^T + \delta\lambda^T$ are positive definite. Initially $\lambda^{(1)} = 0$ is suitable, and subsequently the condition imposes an upper limit on the step in the line search.

A result of particular importance is that the matrix Q has the same sparsity structure as B if assumption (1.23) holds. Together with the negative semidefinite property, this enables the Newton step (4.1) to be calculated efficiently, using the same data structure as for B, without the need to allow for fill-in or pivoting. The situation is analogous to the solution of $Q\lambda = r$ in Toint's sparse PSB update. This gives some reason to hope that the cost of solving (2.8) may not be prohibitive. It might be expected that as the outer iterates $x^{(k)}$ approach the solution, fewer iterations of the inner iteration would hopefully be required. Some preliminary numerical experience in this respect is outlined in §6. Detailed expressions for calculating Q and results about the structure of Q are given in §5.

5. Properties of the Jacobian $Q(\lambda)$. This section sets out the properties of the Jacobian matrix $Q(\lambda)$ of the nonlinear system $r(\lambda) = 0$ generated by the

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

algorithm preceding (2.8). General results with respect to symmetry and negative semidefiniteness are established without the need for assumption (1.23). A necessary and sufficient condition for Q to be negative definite is given, analogous to a result of Toint [10]. It is also shown that if assumption (1.23) holds, then the sparsity structure of Q is identical to that of B, and an expression is given from which Q can be calculated. Finally, an example is provided that shows that if (1.23) does not hold, then Q may be less sparse than B. The significance of these results has already been discussed in §4.

In formulating some general results about Q without the need for assumption (1.23), it is necessary to assume that B is positive definite and is well determined by $\mathcal{G}(H)$. Precisely what minimal set of conditions are required to assure this situation is as yet an open question. To establish the first results in this section, it is assumed that B has been determined from $\mathcal{G}(H^{(k)} + \delta\lambda^T + \lambda\delta^T)$ in such a way as to be a continuously differentiable function of λ, and the effect of a perturbation

$$\lambda \rightarrow \lambda + \varepsilon z \tag{5.1}$$

($\varepsilon \in \mathbb{R}$, $z \in \mathbb{R}^n$) is considered. This induces a perturbation ΔB in B that maintains the sparse structure of B. A perturbation ΔH in H is also induced that generally affects all the elements of H. It follows that the derivative with respect to ε of the perturbation in H is

$$\dot{H} = \lim_{\varepsilon \rightarrow 0} \Delta H/\varepsilon = \mathcal{G}(z\delta^T + \delta z^T) + \Omega, \tag{5.2}$$

where Ω allows for the variation of the elements of H that are zeroed by $\mathcal{G}(H)$. Consequently,

$$\mathcal{G}(\Omega) = 0, \qquad \Omega^T = \Omega, \tag{5.3}$$

and Ω depends upon z. An expression for the derivative of the perturbation in B can be established by differentiating through the equation $BH = I$, giving

$$\dot{B} = -B\dot{H}B. \tag{5.4}$$

In the case that z is a unit vector e_k, the derivative

$$\begin{aligned} \frac{\partial H}{\partial \lambda_k} &= \mathcal{G}(e_k\delta^T + \delta e_k^T) + \Omega_k \\ &= e_k\delta_{[k]}^T + \delta_{[k]}e_k^T + \Omega_k \end{aligned} \tag{5.5}$$

is obtained, making use of Toint's notation described in §1. Similarly, Ω_k depends upon k and satisfies (5.3), although its value is unlikely to be readily available. An expression for $\partial B/\partial\lambda_k$ that follows from (5.4) is

$$\frac{\partial B}{\partial \lambda_k} = -B(e_k\delta_{[k]}^T + \delta_{[k]}e_k^T + \Omega_k)B. \tag{5.6}$$

A consequence of the sparsity of ΔB is that $\partial B/\partial\lambda_k$ has the same sparsity pattern as B. In effect, the Ω_k matrix in (5.6) is determined by the need to achieve this outcome (there are just enough free parameters).

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Turning now to r, it is convenient to use the sparsity and symmetry of B to write

$$r_j = e_j^T B\delta - \gamma_j = e_j^T B\delta_{[j]} - \gamma_j = \tfrac{1}{2}\,\mathrm{trace}((e_j\delta_{[j]}^T + \delta_{[j]}e_j^T)B) - \gamma_j, \tag{5.7}$$

and hence

$$\frac{\partial r_j}{\partial \lambda_k} = \tfrac{1}{2}\,\mathrm{trace}\left((e_j\delta_{[j]}^T + \delta_{[j]}e_j^T)\frac{\partial B}{\partial \lambda_k}\right). \tag{5.8}$$

Moreover, because of the sparsity of $\partial B/\partial\lambda_k$ and (5.3) it follows that

$$\mathrm{trace}\left(\Omega_j \frac{\partial B}{\partial \lambda_k}\right) = 0. \tag{5.9}$$

Equations (5.6)–(5.9) can be incorporated to give

$$\frac{\partial r_j}{\partial \lambda_k} = -\tfrac{1}{2}\,\mathrm{trace}\{(e_j\delta_{[j]}^T + \delta_{[j]}e_j^T + \Omega_j)B(e_k\delta_{[k]}^T + \delta_{[k]}e_k^T + \Omega_k)B\}, \tag{5.10}$$

which shows that the matrix $Q = [\partial r_j/\partial\lambda_k]$ is symmetric.

Next the issue of definiteness of Q is considered and we need to examine z^TQz for some $z \neq 0$. It follows by the chain rule that

$$\sum_{k=1}^{n} Q_{jk}z_k = \sum_{k=1}^{n} \frac{\partial r_j}{\partial \lambda_k}\frac{d\lambda_k}{d\varepsilon} = \frac{dr_j}{d\varepsilon} = \dot{r}_j$$

and hence $z^TQz = z^T\dot{r}$. If λ is perturbed as in (5.1) then it follows from (5.2) and (5.4) that

$$\dot{B} = -B(\mathcal{G}(z\delta^T + \delta z^T) + \Omega)B \tag{5.11}$$

and, as above, $\dot{B}$ has the same sparsity pattern as B. Because of this we can write

$$z^T\dot{r} = z^T\dot{B}\delta = \tfrac{1}{2}\,\mathrm{trace}\{(\mathcal{G}(z\delta^T + \delta z^T))\dot{B}\} \tag{5.12}$$

and hence

$$z^TQz = -\tfrac{1}{2}\,\mathrm{trace}\{(\mathcal{G}(z\delta^T + \delta z^T) + \Omega)B(\mathcal{G}(z\delta^T + \delta z^T) + \Omega)B\}, \tag{5.13}$$

using the fact that $\mathrm{trace}(\Omega\dot{B}) = 0$. Using a weighted Frobenius norm

$$\|A\|_W = (\mathrm{trace}(AWAW))^{1/2}, \qquad W > 0,$$

(5.13) can be expressed as

$$z^TQz = -\tfrac{1}{2}\|\mathcal{G}(z\delta^T + \delta z^T) + \Omega\|_B^2 \leq 0, \tag{5.14}$$

which proves that the matrix Q is negative semidefinite.

The next general result is to show that Toint's condition

$$\delta_{[j]} \neq 0, \qquad j = 1, 2, \ldots, n \tag{5.15}$$

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

is necessary and sufficient for Q to be negative definite. If $\delta_{[j]} = 0$ for some j, then it follows from (5.7) that $r_j = -\gamma_j$ is independent of B and hence of λ. Thus row (and column) j of Q is zero and Q is singular. For the converse result, let Q be singular so that there exists some $z \neq 0$ such that $z^TQz = 0$. It is then a consequence of (5.14) that

$$\mathcal{G}(z\delta^T + \delta z^T) + \Omega = 0. \tag{5.16}$$

It follows directly from the diagonal elements that $z_j\delta_j = 0, \quad j = 1, 2, \ldots, n$. Let $z_j \neq 0$ for some j and hence $\delta_j = 0$. Row j of (5.16) implies that

$$z_j\delta_k + \delta_j z_k = 0 \qquad \forall\, (j,k) \in \mathcal{S}^\perp$$

and it follows that $\delta_{[j]} = 0$, which contradicts (5.15).

The singularity of Q when $\delta_{[j]} = 0$ is unlikely to cause difficulties in practice. It is assumed that the vector γ is faithful to the sparsity pattern of the true Hessian G, that is it can be expressed as

$$\gamma = \bar{G}\delta, \tag{5.17}$$

where $\bar{G}$ is the averaged Hessian matrix (1.6). It then follows that $\delta_{[j]} = 0$ implies $\gamma_j = 0$ and hence $r_j = 0$, so that the Newton system (4.1) is consistent. Thus in exact arithmetic a solution with $\Delta\lambda_j = 0$ can be computed and there is no difficulty. Inexact arithmetic poses possible difficulties due to round-off error, but it is hoped that these can be handled adequately by the use of tolerances.

The results so far derived in this section are not very useful for computation because they involve the unknown matrices Ω_k. In the case that assumption (1.23) holds, it is possible to use (3.14) to derive an expression from which Q can be calculated. This expression also enables the sparsity structure of Q to be determined. It follows from (3.14) using $H^+_{[i]}\delta = H^+_{[i]}\delta_{[i]}$ that

$$r_j = \left(\sum_{i=1}^{n} \frac{e_j^T H^+_{[i]} e_i e_i^T H^+_{[i]} \delta_{[i]}}{e_i^T H^+_{[i]} e_i} \right) - \gamma_j. \tag{5.18}$$

In fact, the sum in (5.18) need only be taken over those i for which j is in the scope of $[i]$ (that is j is one of the nonzero rows in the expanded Markowitz submatrix with index $[i]$), because otherwise $e_j^T H^+_{[i]} = 0$. Now let k be in the scope of $[i]$. It follows from (2.5) and the definition of $H_{[i]}$ that

$$\frac{\partial H_{[i]}}{\partial \lambda_k} = e_k\delta^T_{[i]} + \delta_{[i]}e_k^T \tag{5.19}$$

and hence

$$\frac{\partial H^+_{[i]}}{\partial \lambda_k} = -H^+_{[i]}(e_k\delta^T_{[i]} + \delta_{[i]}e_k^T)H^+_{[i]} \tag{5.20}$$

by the same argument as for (5.4). If k is not in the scope of $[i]$ then

$$\frac{\partial H_{[i]}}{\partial \lambda_k} = \frac{\partial H^+_{[i]}}{\partial \lambda_k} = 0. \tag{5.21}$$

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

It now follows from (5.18) and (5.21) that

$$\frac{\partial r_j}{\partial \lambda_k} = \sum_{i=1}^{n} \left\{ \frac{e_j^T \frac{\partial H_{[i]}^+}{\partial \lambda_k} e_i e_i^T H_{[i]}^+ \delta_{[i]} + e_j^T H_{[i]}^+ e_i e_i^T \frac{\partial H_{[i]}^+}{\partial \lambda_k} \delta_{[i]}}{e_i^T H_{[i]}^+ e_i} - \frac{e_j^T H_{[i]}^+ e_i e_i^T H_{[i]}^+ \delta_{[i]} e_i^T \frac{\partial H_{[i]}^+}{\partial \lambda_k} e_i}{(e_i^T H_{[i]}^+ e_i)^2} \right\}, \tag{5.22}$$

where the sum is taken over those i for which both j and k are in the scope of $[i]$. After substituting (5.20) and denoting $H_{[i]}^+ = M$, $H_{[i]}^+\delta_{[i]} = v$ and $\delta_{[i]}^T H_{[i]}^+ \delta_{[i]} = \mu$, the ith term in (5.22) can be rearranged as

$$\left(\frac{2M_{ij}M_{ik}}{M_{ii}^2} - \frac{M_{jk}}{M_{ii}}\right) v_i^2 - \frac{(M_{ij}v_k + M_{ik}v_j)}{M_{ii}} v_i - \frac{\mu M_{ij} M_{ik}}{M_{ii}}. \tag{5.23}$$

The symmetry with respect to $j \leftrightarrow k$ can readily be observed. Because the ith term only contributes to the sum if both j and k are in the scope of $[i]$, it follows that the ith term has the same sparsity pattern as $B_{[i]}$, and hence Q has the same sparsity pattern as B. Equations (5.22) and (5.23) provide a (rather complicated) formula from which Q can be evaluated. The expression can be simplified a little by using the fact that $d_i = M_{ii}$ and $l_{ji} = M_{ij}/d_i$ derived from the LDL^T factors of B.

It is also possible to give an example which shows that Q may be less sparse than B if assumption (1.23) does not hold. Consider the matrix

$$B = \begin{bmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{bmatrix}, \tag{5.24}$$

which is sparse on the reverse diagonal and fills in in the (2,3) position when factorized. Consider the computation of $Q_{41} = \partial r_4/\partial\lambda_1$. We first need (5.5), which can be written

$$\frac{\partial H}{\partial \lambda_1} = \begin{bmatrix} 2\delta_1 & \delta_2 & \delta_3 & \omega_1 \\ \delta_2 & & \omega_2 & \\ \delta_3 & \omega_2 & & \\ \omega_1 & & & \end{bmatrix}. \tag{5.25}$$

Then $\partial B/\partial\lambda_1$ can be calculated using (5.6) and (5.24), and ω_1 and ω_2 are chosen to make the reverse diagonal zero. This gives rise to the system

$$\begin{bmatrix} 8 & 1 \\ 1 & 8 \end{bmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \begin{pmatrix} 2\delta_2 + 2\delta_3 \\ -\delta_1 + 2\delta_2 + 2\delta_3 \end{pmatrix}. \tag{5.26}$$

Then (5.8) gives

$$Q_{41} = \partial r_4/\partial\lambda_1 = (\delta_2 + \delta_3)(\delta_2 + \delta_3 - 4\omega_1 - 4\omega_2) + 2\delta_4\omega_2. \tag{5.27}$$

Clearly from (5.26), ω_2 is not zero, and the presence of the $2\delta_4\omega_2$ term in (5.27) ensures that there are cases in which Q_{41} is not zero. In general, Q is a dense matrix. However, if the sparsity of the (2,3) element is relaxed and B is treated as a band matrix, then the ω_2 parameter is removed and ω_1 is chosen to zero the (4,1) element of (5.6). This

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

provides the equation $\delta_2 + \delta_3 = 4\omega_1$ and it follows from (5.27) that $Q_{41} = 0$, so that Q is also a band matrix.

6. Numerical experiments. In this section some numerical experiments are described that are designed to test the effectiveness of both the sparse updates in a line search quasi-Newton method and the inner Newton iteration based on (4.1). The experiments are limited to the relatively simple case of a tridiagonal Hessian matrix. The computations have been carried out on a SUN SPARCstation SLC in single precision.

In the case of a tridiagonal Hessian matrix, we may write

$$\mathcal{G}(H) = \begin{bmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & b_2 & a_3 & \ddots & \\ & & \ddots & \ddots & b_{n-1} \\ & & & b_{n-1} & a_n \end{bmatrix}. \tag{6.1}$$

Clearly assumption (1.23) holds when factorizing B, and the Markowitz submatices are the 2×2 blocks on the diagonal, and the final 1×1 block. It is readily verified from Theorem 3.1 that the LDL^T factors of B are given by

$$L = \begin{bmatrix} 1 & & & & \\ -b_1/a_2 & 1 & & & \\ & -b_2/a_3 & 1 & & \\ & & \ddots & \ddots & \\ & & & -b_{n-1}/a_n & 1 \end{bmatrix}, \tag{6.2}$$

$$D = \begin{bmatrix} a_2/\Delta_1 & & & & \\ & a_3/\Delta_2 & & & \\ & & \ddots & & \\ & & & a_n/\Delta_{n-1} & \\ & & & & a_n^{-1} \end{bmatrix}, \tag{6.3}$$

where $\Delta_i = a_i a_{i+1} - b_i^2$. The elements of Q are given by (5.23), which simplifies considerably in this case.

Two test problems of variable dimensions have been used in the experiments. One is the *boundary value problem*

$$\text{minimize} \quad f(x) = \tfrac{1}{2} x^T T x - e_n^T x - h^2 \sum (\kappa \cos x_i + 2x_i), \tag{6.4}$$

where $h = 1/(n+1)$ and

$$T = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2 \end{bmatrix}.$$

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

An initial point $x_i^{(1)} = ih$, $i = 1, 2, \dots, n$ has been used. Choosing $\kappa = 0$ gives rise to a quadratic function that is useful for testing purposes. The problem is otherwise nonquadratic and the value $\kappa = 1$ has been used. The second test problem is the *chained Rosenbrock problem*

$$\text{(6.5)} \qquad \text{minimize} \quad f(x) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2,$$

where n is even. The solution of this problem is $x^* = (1, 1, \dots, 1)^T$ and the usual initial point is $x^{(1)} = (-1.2, 1, -1.2, 1, \dots, -1.2, 1)^T$. However, this initial point sometimes leads to the location of a local minimum that exists in the vicinity of $x_1 = -1$ with $f(x) \simeq 4$. Thus the initial point $x^{(1)} = 0$ has been used that avoids these difficulties and does not appear to make the problem easier.

The new sparse update (spqn) is implemented in a very crude way. A standard quasi-Newton code is used with two-sided Wolfe–Powell conditions in the line search (e.g., [4]) using the parameters $\rho = 0.01$ and $\sigma = 0.1$. The inner (dual) iteration is implemented with an Armijo-type line search with cutback factor 0.25. The initial step in the dual line search is either 1, if feasible ($B > 0$), or otherwise 0.9 of the distance to the step, which would make $H_{ii} = 0$. If B does not become positive definite then the Armijo cutback is used. The inner iteration is terminated when the predicted increase in ϕ is less than $10^{-7}n$. A comparison is made with other methods that do not take advantage of the tridiagonal structure, including a standard BFGS code, an implementation of Nocedal's low storage method [8] based on five stored difference pairs, and an implementation of the Polak–Ribiere conjugate gradient method. The same line search is used for all these methods. Results are also given for Newton's method (exact Hessian) and the LANCELOT method which do take advantage of the tridiagonal structure. The line search in Newton's method uses the parameter $\sigma = 0.9$.

The outcome of these tests is set out in Tables 1–4. All the methods are able to solve the problems, although less accuracy is obtained by the conjugate gradient method, particularly for the larger problems. In all cases the methods that make use of the tridiagonal structure of the Hessian obtain significantly better results. The most marked difference is for the $n = 100$ boundary value problem, both for $\kappa = 0$ and $\kappa = 1$. These problems are quadratic or nearly so, and once a good approximation to the Hessian is obtained then the problem is effectively solved. It is clear that the new update enables the Hessian to be approximated very rapidly because of the relatively few elements in the sparse matrix that must be determined. The same is true for the LANCELOT code. In passing, it is also interesting to notice that Nocedal's limited memory method is comparable with BFGS, even for near quadratic problems. The chained Rosenbrock problem is highly nonlinear and the improvement obtained by the new update is less spectacular. A possible interpretation of this is that the local Hessian matrix for the problem changes markedly as the iterations proceed towards the solution. The new update is likely to get a good estimate of the local Hessian more quickly than say BFGS, but both methods are efficient at revising this estimate as the local Hessian changes.

The results for the LANCELOT code on the chained Rosenbrock problem (I am indebted to Nick Gould for providing these results.) lie between those for the new update and for Newton's method. The LANCELOT code assumes that the objective function has the form

$$\text{(6.6)} \qquad f(x) = \sum_i g_i \left(\textstyle\sum_j f_j(x) + a_i^T x - b_i \right).$$

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 1
Results for boundary value problem $n = 10$.

Method	κ	f^*	Number of iterations	Function evaluations	Gradient evaluations
spqn	0	−0.552217	5	10	10
	1	−0.615442	7	12	12
bfgs	0	−0.552216	6	12	11
	1	−0.615442	11	21	20
nocedal	0	−0.552216	7	12	12
	1	−0.615441	11	23	22
PR-cg	0	−0.552217	9	24	12
	1	−0.615442	13	34	19
lancelot	0	−0.552216	1	2	2
	1	−0.615441	5	6	6
newton	0	−0.552216	2	3	3
	1	−0.615441	2	3	3

TABLE 2
Results for boundary value problem $n = 100$.

Method	κ	f^*	Number of iterations	Function evaluations	Gradient evaluations
spqn	0	−0.506503	3	10	9
	1	−0.514007	5	15	14
bfgs	0	−0.506503	50	68	64
	1	−0.514002	48	71	62
nocedal	0	−0.506501	48	133	111
	1	−0.514000	48	149	123
PR-cg	0	−0.506498	48	78	61
	1	−0.513997	49	101	75
lancelot	0	−0.506502	1	2	2
	1	−0.514005	2	3	3
newton	0	−0.506503	2	3	3
	1	−0.514007	2	3	3

This is referred to as *group partial separability* and it extends the ideas of Griewank and Toint referred to in (1.21). Using group functions that are squares enables the quadratic parts of (6.4) and (6.5) to be specified exactly. Thus the only nonlinear terms that need to be approximated are the $\cos x_i$ term in (6.4) and the x_i^2 term in the first bracket of (6.5). This enables the LANCELOT code to approach more closely the performance of Newton's method, whilst only requiring first derivatives of the nonlinear functions $g_i(\alpha)$ and $f_j(x)$. The new update given in this paper requires less information from the user than LANCELOT, only requiring first derivatives of $f(x)$, along with the sparsity pattern of G. Thus the comparative performance of the

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 3
Results for chained Rosenbrock problem $n = 10$.

Method	f^*	Number of iterations	Function evaluations	Gradient evaluations
spqn	$1.4_{10}-9$	37	91	78
bfgs	$3.0_{10}-8$	52	123	104
nocedal	$3.2_{10}-8$	53	109	102
PR-cg	$2.0_{10}-7$	107	218	186
lancelot	$4.5_{10}-21$	34	35	30
newton	$3.9_{10}-9$	24	29	27

TABLE 4
Results for chained Rosenbrock problem $n = 100$.

Method	f^*	Number of iterations	Function evaluations	Gradient evaluations
spqn	$1.0_{10}-10$	290	727	648
bfgs	$2.7_{10}-8$	426	920	802
nocedal	$5.3_{10}-8$	479	881	871
PR-cg	$3.0_{10}-6$	652	1107	1083
lancelot	$1.7_{10}-13$	227	228	191
newton	$1.8_{10}-10$	161	190	182

new update is very much what might reasonably have been hoped for.

On the other hand, the cost of calculating the new update is significant and dominates the computation time for test problems such as these that are readily evaluated. Even when $x^{(k)}$ is close to x^*, it is observed that about four iterations are required to solve the dual problem (4.4) and up to a dozen or more on the early iterates of the outer problem. The inner iteration is solved to high accuracy and the second order convergence associated with Newton's method is observed, which gives confidence in the correctness of the calculations. However, a lower accuracy solution to the dual might be more effective overall. The Armijo line search also shows up rather poorly, particularly when the singularity on the boundary of $B > 0$ comes into play. A special purpose line search for the dual would undoubtedly have improved the overall performance. This would involve determining precisely the step to the boundary in the dual line search. Because a rank-two update of each Markowitz submatrix H_i is involved, the solution of a quadratic equation for each distinct Markowitz submatrix is required (even for a general sparsity pattern). Another possibility for improving the performance of the dual iteration is to seek some quick method for estimating a good initial value $\lambda^{(1)}$ rather than the value $\lambda^{(1)} = 0$ used here.

It is disappointing not to have observed that only one dual step is required when $x^{(k)}$ is asymptotically close to x^*. I had expected that this would be the case as the Hessian approximation converges. Possibly the outer iteration converges before the phenomenon becomes apparent.

One referee correctly points out that the number of function and gradient calls for bfgs, nocedal and sqpn in the tables could be appreciably improved by using a weaker

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

tolerance (e.g., $\sigma = 0.9$) in the line search, albeit at the expense of some increase in the number of iterations. The best choice of σ for any particular problem depends on the cost of evaluating the function and gradient. However, I would agree that $\sigma = 0.9$ is probably a better choice for the default option for these methods.

7. Primal algorithms. In this section the possibility of using primal algorithms to solve the problem in (2.1)–(2.4) is considered. If the positive definite constraint is inactive, then this problem is one with linear constraints on the elements of B and a nonquadratic objective function. Generalised elimination techniques (e.g., [4]) can therefore be used to provide a reduced unconstrained minimization problem for which there are various possible methods of solution. The basis vectors calculated in the elimination process are shown to have a particularly nice interpretation.

It is convenient to express $B = B^{(k)} + E$, where E is the change in $B^{(k)}$. Then the symmetry and sparsity constraints (2.2) and (2.3) can be immediately satisfied by choosing only independent nonzero elements in E as the unknowns. For example, a 4×4 tridiagonal matrix can be expressed as

$$E = \begin{bmatrix} x_1 & x_5 & & \\ x_5 & x_2 & x_6 & \\ & x_6 & x_3 & x_7 \\ & & x_7 & x_4 \end{bmatrix}. \tag{7.1}$$

The system $(B^{(k)} + E)\delta = \gamma$ is then rearranged as an underdetermined system

$$A^T x = b, \tag{7.2}$$

where $b = \gamma - B^{(k)}\delta$, and where $b \in \mathbb{R}^n$, $x \in \mathbb{R}^\tau$, and τ is the number of independent unknowns in E. In generalised elimination, A is bordered by an arbitrary matrix V so that $[A\ V]$ is nonsingular, and the inverse

$$[A\ V]^{-T} = [Y\ Z] \tag{7.3}$$

is used to define the matrices Y (having the same dimensions as A) and Z. Then the feasible region of (7.2) can be parametrized by

$$x = Yb + Zy, \tag{7.4}$$

where Yb is a feasible point of (7.2) and Zy represents an arbitrary correction in null space of A. The reduced optimization problem can therefore be expressed in terms of the vector y, which has $\tau - n$ components.

In this particular application the columns of Y and Z can be regarded as elementary $n \times n$ matrices Y_i and Z_i by scattering their elements according to the sparsity pattern of E. A particularly convenient form is obtained if the columns of V are unit vectors with unit element in positions corresponding to off-diagonal elements of E. This approach is only numerically stable[1]

if the elements δ_i $i = 1, 2, \ldots, n$ are not close to zero. To illustrate this construction, (7.1) can be rearranged in the form (7.2) as

$$A^T x = \begin{bmatrix} \delta_1 & & & & \delta_2 & & \\ & \delta_2 & & & \delta_1 & \delta_3 & \\ & & \delta_3 & & & \delta_2 & \delta_4 \\ & & & \delta_4 & & & \delta_3 \end{bmatrix} x = b. \tag{7.5}$$

[1] There are other more stable constructions for V based on Gaussian elimination with pivoting or the use of QR factors; see [4].

Using the above choice of V,

$$[A\ V] = \begin{bmatrix} \delta_1 & & & & & & \\ & \delta_2 & & & & & \\ & & \delta_3 & & & & \\ & & & \delta_4 & & & \\ \delta_2 & \delta_1 & & & 1 & & \\ & \delta_3 & \delta_2 & & & 1 & \\ & & \delta_4 & \delta_3 & & & 1 \end{bmatrix}. \tag{7.6}$$

It is readily verified from (7.3) that $Y_i = e_i\delta_i^{-1}e_i^T \quad i = 1, 2, 3, 4$, and

$$Z_1 = \begin{bmatrix} -\delta_2\delta_1^{-1} & 1 & & \\ 1 & -\delta_1\delta_2^{-1} & & \\ & & 0 & \\ & & & 0 \end{bmatrix}, \quad Z_2 = \begin{bmatrix} 0 & & & \\ & -\delta_3\delta_2^{-1} & 1 & \\ & 1 & -\delta_2\delta_3^{-1} & \\ & & & 0 \end{bmatrix},$$

and

$$Z_3 = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & -\delta_4\delta_3^{-1} & 1 \\ & & 1 & -\delta_3\delta_4^{-1} \end{bmatrix}. \tag{7.7}$$

In general, matrices Y_i are used to construct a feasible correction, $E^{(k)}$ say, by

$$E^{(k)} = \sum Y_i b = \operatorname{diag}(b_i/\delta_i). \tag{7.8}$$

Moreover, it is readily verified that $Z_i\delta = 0$ illustrating the null space property of the Z_i matrices. Thus a parametrization of feasible matrices in (2.2)–(2.4) is given by

$$B(y) = B^{(k)} + E^{(k)} + \sum Z_i y_i. \tag{7.9}$$

For more general sparsity patterns a similar outcome is obtained and there is one Z_i matrix, having a similar structure to (7.7), for each off-diagonal element of E.

This construction is essentially given by Toint [11]. Assuming that $\delta_i \neq 0 \quad i = 1, 2, \ldots, n$, that B is irreducible, and that $\delta^T\gamma > 0$, Toint proves that a positive semi-definite matrix T with $\operatorname{rank}(T) = n - 1$ can be constructed from a linear combination of the matrices Z_i. It then follows that $\hat{B} = B^{(k)} + E^{(k)} + \alpha T$ is positive definite for sufficiently large α. This proves the existence of a positive definite matrix B that satisfies the constraints (2.2)–(2.4). A consequence of Toint's result is that the primal always has a solution under these conditions. The implications of this result are discussed in more detail in §8.

If a primal algorithm is to be used, then it is advantageous if the initial iterate $B^{(1)}$ is a positive definite matrix. This enables ψ to be used as a merit function to force convergence. Also the barrier function property of ψ ensures that all subsequent iterates stay positive definite. A possible choice of $B^{(1)}$ is the matrix $\hat{B}$ referred to in the previous paragraph.

In constructing an efficient algorithm, it is important to get at least first derivatives of the reduced objective function $\psi(B(y))$ derived from (2.1) and (7.9). Differentiating ψ with respect to B_{ij} as in §2 and using the chain rule, it follows that

$$\frac{\partial\psi}{\partial y_i} = \operatorname{trace}(Z_i(H^{(k)} - H)). \tag{7.10}$$

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Because $\mathcal{G}(Z_i) = Z_i$, it follows that only $\mathcal{G}(H^{(k)} - H)$ is required in (7.10). As mentioned in §3, Duff, Erisman, and Reid [3] indicate that the elements of $\mathcal{G}(H)$ can be determined efficiently from factors of B if assumption (1.23) holds. Thus calculation of the reduced gradient of ψ is not unduly expensive. Note also that the form of (7.10) confirms the characterisation result given in (2.5) since $\operatorname{trace}(\mathcal{G}(\delta\lambda^T + \lambda\delta^T)Z_i) = 0$ showing that if B is derived from (2.5) then it is a stationary point of the primal (and a solution if B is positive definite).

The next stage is to look at the Hessian matrix of $\psi(y)$. It follows using (5.4) that

$$\frac{\partial^2\psi}{\partial y_j \partial y_i} = -\operatorname{trace}\left(Z_i \frac{\partial H}{\partial y_j}\right) = \operatorname{trace}(Z_i H Z_j H). \tag{7.11}$$

However, it seems unlikely that the determination of this matrix and its use in a primal Newton method will be profitable. First, the reduced primal has $\tau - n$ variables and this could be significantly larger than n. Also the reduced Hessian may not be all that sparse. Perhaps the most likely option is to calculate an approximate solution of the update problem, using a few steps of preconditioned conjugate gradients with a diagonal or perhaps tridiagonal matrix derived from (7.11) as preconditioner.

8. Stability issues and discussion.

This section takes up the issue of the numerical stability of the sparse positive definite update. An example of Sorensen [9] highlights a potential difficulty of sparse positive definite updates. A possible solution to these difficulties is suggested. A conjecture relating the new update to partially separable optimization is discussed and possibilities for further work are suggested.

The assumption $\delta_i \neq 0$ $i = 1, 2, \ldots, n$ used by Toint (see §7) is not there to simplify the proof, but is symptomatic of a serious difficulty that can arise when B is sparse and $\delta_i = 0$. This is made clear by Sorensen [9] who essentially cites the following example in which B is required to solve

$$\begin{bmatrix} a & b & \\ b & c & * \\ & * & * \end{bmatrix} \begin{pmatrix} -1 \\ \varepsilon \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}. \tag{8.1}$$

When $\varepsilon \neq 0$ the first equation implies $b = (a+1)/\varepsilon$ and $a > 0$ is required for positive definiteness. Thus b grows without limit as ε goes to zero. Moreover, the inequality $ac > b^2$ implies that ac increases like ε^{-2} so the rate of growth is quadratic. If $\varepsilon = 0$, then the only solution has $a = -1$ and there does not exist a positive definite update. Yet $\delta^T\gamma = 1 > 0$.

Both types of algorithm described in this paper fail on this example when $\varepsilon = 0$. The primal problem has no feasible point so cannot be started. Because the primal is infeasible, so the dual is unbounded, and the dual iteration is seen to cause both λ and $\phi(\lambda)$ to increase without bound. At the same time both $B(\lambda)$ and $H(\lambda)$ increase without bound. The most growth is seen to occur in the ith diagonal element of B. In addition, both algorithms exhibit ill conditioning as $\varepsilon \to 0$. In the primal, the matrices Y_i and Z_i become arbitrarily large giving similarly large B matrices in the calculation. The dual iteration again exhibits large growth in λ, B, H, and $\psi(\lambda)$.

These phenomena can be seen in the following example derived from (8.1). Let δ and γ be as in (8.1), let $B^{(1)} = I$, and let $B^{(2)}$ be tridiagonal. A short calculation using (7.10) indicates that the solution has the form

$$(8.2)\qquad \begin{aligned} B &= \begin{bmatrix} 3 & 4\varepsilon^{-1} & \\ 4\varepsilon^{-1} & 8\varepsilon^{-2} & -4\varepsilon^{-1} \\ & -4\varepsilon^{-1} & 6 \end{bmatrix} (1+O(\varepsilon^2)), \\ H &= \begin{bmatrix} 16\varepsilon^{-2} & -12\varepsilon^{-1} & -8\varepsilon^{-2} \\ -12\varepsilon^{-1} & 9 & 6\varepsilon^{-1} \\ -8\varepsilon^{-2} & 6\varepsilon^{-1} & 4\varepsilon^{-2} \end{bmatrix} (1+O(\varepsilon^2)) \end{aligned}$$

with $\lambda = (-8\varepsilon^{-2},\ 4\varepsilon^{-1},\ 2\varepsilon^{-2})^T(1+O(\varepsilon^2))$. The increasing growth as $\varepsilon \to 0$ is readily observed.

Some possible ways of avoiding such difficulties are now discussed. It is important to realise that such a situation is only likely to occur when the average Hessian matrix $\bar{G}$ in (1.6) is indefinite or ill conditioned. Because of (1.5), $\bar{G}$ is always feasible in the primal and, if $\bar{G}$ is positive definite and well-behaved, then $\delta^T\gamma/\delta^T\delta$ is not close to zero. It follows that if $B^{(k)}$ is well-behaved, then there is no possibility of serious growth in $\psi(B^{(k)}H)$ and hence $B^{(k+1)}$ is well-behaved. Even if $\bar{G}$ is indefinite then a satisfactory update may yet be obtained (this happens for example in the dense case). Thus it may be that these difficulties arise relatively infrequently and an ad hoc solution may be adequate. However, when difficulties do arise then their effects are severe due to the ε^{-2} growth. In addition it is likely that no such growth is evident in $\bar{G}$ when, for example, it is indefinite. Thus any heuristic should avoid letting B and H grow large. One possibility is to skip the update if some δ_i is close to zero, but this precludes a useful update if $\bar{G}$ is well-behaved. Another possibility is to make no change to row/column i of B. This may decouple two parts of the matrix (as for (8.1)) and require $\delta^T\gamma > 0$ for each part. It may also cause organisational problems by changing the effective sparsity pattern. The solution that currently appeals the most is simply to impose an upper limit on the size of B and H, and to abort the dual iteration if either of these upper limits are exceeded. It is worth pointing out that if an upper limit on the trace of both B and H is imposed, then it is readily deduced that *cond*(B) is bounded and it follows (see [4], p.31) that the resulting quasi-Newton method is globally convergent. Of course, the user does not find it easy to set such upper limits and too small a value might preclude superlinear convergence. Further experience of these situations would be valuable and might lead to improved heuristics.

As it stands, the dual iteration (4.1) does not have much in common with the calculation involved in the BFGS update (1.8). For example in the dense case, a unit step of (4.1) does not provide the BFGS correction. Experience with some simple cases indicates that it might be possible to express the update as

$$(8.3)\qquad B_{(i)}^{(k+1)} = \mathrm{bfgs}(B_{(i)}^{(k)}, \delta_{[i]}, \gamma_{(i)}).$$

The matrices $B_{(i)}$ derive from an additive decomposition of B corresponding to distinct Markowitz submatrices (that is $B = \sum B_{(i)}$). Terms in which a Markowitz submatrix is a submatrix of another Markowitz submatrix would be excluded. A corresponding decomposition $\gamma_{(i)}$ for which $\sum \gamma_{(i)} = \gamma$ is also required. The updates in (8.3) require that the scalar products

$$(8.4)\qquad \delta_{[i]}^T\gamma_{(i)} > 0$$

are all positive. The existence of vectors $\gamma_{(i)}$ for which this holds is related to Toint's condition that $\delta_i \neq 0 \quad i = 1, 2, \ldots, n$. For example it is clear that this property

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

cannot be attained in (8.1) when $\varepsilon = 0$. The outcome in (8.3) would have the flavour of a partially separable update but with the blocks being determined by the result of fill-in in the LDL^T factors rather than being prescribed by the user. At present, however, it is not clear how the decomposition of γ would be determined.

A related problem to the one considered in this paper arises if the structural constraints on B are expressed as

$$B_{ij} = \beta_{ij} \quad \forall\, (i,j) \in \mathcal{S}, \tag{8.5}$$

where the β_{ij} are known values of the true Hessian that are independent of x but might be nonzero. For example, in the boundary value problem (6.4), one would require $B_{i,i+1} = -1$. To some extent this problem can be transformed by taking the product $\bar{B}\delta$ over to the other side of the quasi-Newton equation (2.3), where $\bar{B}$ denotes the matrix with entries β_{ij} for $(i,j) \in \mathcal{S}$ and zero otherwise. Condition (2.5) is then valid for this modified problem, but there is a difference in that B must be positive definite and not $B - \bar{B}$. Also the possibility of instability or nonexistence of the update may be compounded as there are fewer elements to adjust. For example, in the case of (6.4) the modified problem decomposes into distinct 1×1 diagonal blocks and an update is immediately determined, which may or may not correspond to a positive definite B. On the other hand there is the possibility of determining the unknown elements more rapidly, and the argument used earlier in the section regarding $\bar{G}$ might indicate that a successful update will often be obtained. Again some practical experience is called for.

In summary, it is felt that an update of some potential interest has been suggested in this paper, although further development is required before the idea can be incorporated into production software. Possible areas of future work include the following.

- Better implementation of the dual line search;
- Implementation and numerical experience for band matrices and more general sparsity patterns (overlapping blocks, skyline, random (subject to (1.23)));
- Low accuracy solution of dual;
- Heuristics for alleviating ill-conditioning when some δ_i is close to zero;
- Primal methods;
- Alternative ways of computing the update such as (8.3);
- Relationship to partially separable approach;
- Theory of obtaining B from $\mathcal{G}(H)$ when (1.23) does not hold;
- Superlinear convergence of $x^{(k)} \to x^*$;
- Experience with $B_{ij} = \beta_{ij}$ problems;
- Use in an NLP context.

Most of these have already been discussed at some point in this paper.

REFERENCES

[1] R. H. Byrd and J. Nocedal, *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal., 26 (1989), pp. 727–739.

[2] T. F. Coleman, *Large Sparse Numerical Optimization*, Lecture Notes in Computer Science 165, Springer-Verlag, Berlin, 1984.

[3] I. S. Duff, A. M. Erisman, and J. K. Reid, *Direct Methods for Sparse Matrices*, Clarendon Press, Oxford, 1986.

[4] R. Fletcher, *Practical Methods of Optimization*, 2nd. ed., John Wiley, Chichester, 1987.

[5] ______ , *A new result for quasi-Newton formulae*, SIAM J. Optim., 1 (1991), pp. 18–21.

[6] D. Goldfarb, *A family of variable metric methods derived by variational means*, Maths. Comput., 24 (1970), pp. 23–26.

[7] A. Griewank and Ph. L. Toint, *Partitioned variable metric updates for large structured optimization problems*, Numer. Math., 39 (1982), pp. 429–448.

[8] J. Nocedal, *Updating quasi-Newton matrices with limited storage*, Math. Comput., 35 (1980), pp. 773–782.

[9] D. C. Sorensen, *Collinear scaling and sequential estimation in sparse optimization algorithms*, Math. Programming Study, 18 (1982), pp. 135–159.

[10] Ph. L. Toint, *On sparse and symmetric updating subject to a linear equation*, Maths. Comput., 31 (1977), pp. 954–961.

[11] ______ , *A note on sparsity exploiting quasi-Newton methods*, Math. Programming, 21 (1981), pp. 172–181.

Downloaded 11/08/13 to 216.165.95.73. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php