AMO - Advanced Modeling and Optimization, Volume 11, Number 1, 2009

DIRECTION-SET UPDATES FOR DERIVATIVE-FREE OPTIMIZATION *

I.D. COOPE
DEPARTMENT OF MATHEMATICS & STATISTICS
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND
IAN.COOPE@CANTERBURY.AC.NZ

D. HUTCHINSON
SCHOOL OF COMPUTING
UNIVERSITY OF LEEDS, ENGLAND
DH@COMP.LEEDS.AC.UK

Abstract. The updating of direction sets in direct search methods for unconstrained optimization is examined. Both weak and strong quasi-Newton updates are considered together with other simple quadratic interpolation conditions. Efficient and numerically stable techniques are described for implementing the appropriate updates. The updating schemes are applicable to both line search and trust region algorithms as well as some newer grid-based methods for derivative-free optimization and the updates considered can usually be calculated in $\mathcal{O}(n^2)$ arithmetic operations.

Key words. Direction-set updates, quasi-Newton method, updating formulas, conjugate directions, derivative-free optimization, minimum norm corrections.

AMS subject classifications. 49M30, 65F99, 65K05, 90C53, 90C56.

1. Introduction.

The derivation of quasi-Newton updating formulas by variational means has been a topic of interest to many authors since the first approach by Greenstadt in [10]. Examples can be found in [4], [7], [8], [11], [12], [13], [15], [16], and the references therein. In most cases a quasi-Newton formula is derived by seeking a smallest correction in some norm to the current estimate of the Hessian matrix of second derivatives (or its inverse) of the function to be minimized. In this paper the updating of direction sets in direct search methods for unconstrained minimization is examined in cases where the underlying method can be interpreted as a conjugate direction, quasi-Newton method or more general grid-based method. Both weak and strong quasi-Newton updates are considered together with other simple quadratic interpolation conditions. Efficient and numerically stable techniques are described for implementing the appropriate updates. The updating schemes are applicable to both line search and trust region algorithms for derivative-free optimization and enable the new direction set to be calculated in $\mathcal{O}(n^2)$ arithmetic operations.

Early in the development of quasi-Newton methods low rank corrections to estimates of the inverse Hessian were used in algorithms for the unconstrained minimization of $f(x), x \in \mathbf{R}^n$. Typically, the iteration

$$x_{k+1} = x_k + \alpha_k p_k, \quad k = 1, 2, \ldots,$$

is applied from an initial approximation x_1 where

$$p_k = -H_k \nabla f(x_k),$$

for a symmetric matrix H_k, intended to approximate $[\nabla^2 f(x_k)]^{-1}$ in some sense. In the absence of any better approximation it is usual to make the choice $H_1 = I$. The line search parameter α_k (steplength) is then chosen automatically at each iteration

* **AMO - Advanced Modeling and Optimization. ISSN: 1841-4311**

and the matrix H_k is updated to reflect information obtained in the course of the current iteration. Later implementations updated Choleski factors of $B_k = H_k^{-1}$. Probably the most widely recommended update is the well-known BFGS update

$$B_{k+1} = \left[B - \frac{Bss^T B}{s^T Bs} + \frac{yy^T}{s^T y} \right]_k ,$$

followed closely by the rank-one update

$$B_{k+1} = \left[B + \frac{(y - Bs)s^T}{(y - Bs)^T s} \right]_k .$$

Here $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ and the notation $[\,\cdot\,]_k$ means everything inside the bracket has an iteration subscript $_k$. The rank-one update need not preserve positive definiteness for line search algorithms so trust-region methods are sometimes recommended when this update is used. At each iteration of a trust-region method a model problem of the following form is solved (see [3], for example).

$$\min \left\{ \tfrac{1}{2} p^T B_k p + p^T \nabla f(x_k) : \quad |\|p\| - \Delta_k| \le \rho \Delta_k \right\} , \tag{1.1}$$

where B_k is an approximation to $\nabla^2 f(x_k)$, Δ_k is the trust-region radius and $0 < \rho < 1$ is a relative tolerance (typically $\rho = 0.1$). More recently direct search methods have had a resurgence - motivated by improved techniques for establishing good convergence properties. A good survey of progress in this area is given in [14] and in view of the renewed interest in direction set methods it seems timely to revisit the updating problem.

2. The weak quasi-Newton condition. Perhaps the simplest quasi-Newton condition that can be imposed on the symmetric matrix B_{k+1} is that it matches the second derivative information provided by quadratic interpolation along a single direction p. Therefore, the problem considered first is to find a matrix B^+ closest in some sense to a given symmetric positive definite matrix B such that

$$p^T B^+ p = d, \tag{2.1}$$

where here, as in the remainder of this paper, we drop iteration subscripts k and denote updated items with a superscript '+'. In most minimum-norm correction schemes a matrix E of minimum norm is sought as an additive correction to B so that $B^+ = B + E$ satisfies the required interpolation conditions. The Frobenius matrix norm is convenient because it usually leads to a linear sub-problem that is easily solved but it has the disadvantage of giving poor accuracy in directions with little curvature. This can be alleviated by using weighted Frobenius norms (see e.g. [13]) but we prefer to adopt an approach more closely resembling that taken in [7]. Specifically, if S is an invertible matrix whose columns are conjugate with respect to the matrix B scaled such that $S^T BS = I$ then

$$B = S^{-T} S^{-1}$$

and a correction matrix E is required so that

$$B^+ = S^{-T}[I + E]S^{-1} = B + S^{-T} ES^{-1} \tag{2.2}$$

satisfies (2.1). Now it is appropriate to use the Frobenius norm to measure the size of the correction because it is perfectly scaled relative to the identity matrix. Equation

(2.1) is sometimes called the "weak quasi-Newton condition" because it is a scalar version of the more usual (strong) quasi-Newton condition $B^+p = y$ obtained by premultiplying by p^T and writing $d = p^T y$. In a derivative-free optimization algorithm d may be provided by interpolating three function values obtained at distinct points along a line through the current point x and parallel to the vector p. For example, the formula

$$d = f(x-p) - 2f(x) + f(x+p)$$

may be used when the interpolation points are equally spaced because this is consistent with the second order directional derivative information $p^T[\nabla^2 f(x)]p = d$ when f is a quadratic function. The problem of choosing a smallest correction is easily solved when p is represented as a linear combination of the columns of S so that

$$p = Sh$$

for some known vector h. Then the matrix E can be determined by solving the simple problem

$$\min\left\{\|E\|_F^2 : \quad h^T E h = d - h^T h, \quad E = E^T\right\}. \tag{2.3}$$

The solution is

$$E = \eta[hh^T] \tag{2.4}$$

where η is the scalar

$$\eta = \frac{d - h^T h}{(h^T h)^2}.$$

Instead of updating B to B^+ it is possible to update S directly. First notice that B^+, defined by equations (2.2), (2.4), is positive definite, if and only if $d > 0$ since $\det[I + \eta hh^T] = 1 + \eta h^T h = d/(h^T h)$. If B^+ is not positive definite then it cannot be represented in the (real) form $[SS^T]^{-1}$ so only the case where $d > 0$ is appropriate. In this case, it suffices to let S^+ be the matrix

$$S^+ = S[I + \beta hh^T] = S + \beta ph^T, \tag{2.5}$$

where the scalar β is chosen so that

$$[I + \eta hh^T]^{-1} = [I + \beta hh^T]^2.$$

Writing $\gamma = -\eta/(1 + \eta h^T h)$, so that $[I + \eta hh^T]^{-1} = [I + \gamma hh^T]$, it can be seen that the two possible choices for β are given by the roots of the quadratic equation

$$(h^T h)\beta^2 + 2\beta - \gamma = 0,$$

Choosing the root smallest in modulus, and taking care to avoid cancellation error, the formula for β is

$$\beta = \frac{\gamma}{1 + \sqrt{1 + \gamma h^T h}}.$$

Therefore, S can be updated in $\mathcal{O}(n^2)$ multiplications and additions. If S is a triangular matrix then triangularity can also be restored to S^+ in $\mathcal{O}(n^2)$ multiplications by taking advantage of the rank-one nature of the correction (see, for example, [9]). Simple though this approach is, there is, perhaps, a better way which is considered now because both approaches provide useful insight into more complicated situations where several conditions are required to be satisfied simultaneously by the updated matrix.

Problem (2.3) is unchanged in theory if S is replaced by SQ, where Q is any orthogonal matrix. To see this, notice that

$$B^+ = S^{-T}[I+E]S^{-1} = S^{-T}[QQ^T + E]S^{-1} = [SQ]^{-T}[I+F][SQ]^{-1}$$

where $F = Q^T EQ$. But $\|F\|_F = \|E\|_F$, therefore, the problem of minimizing $\|E\|_F^2$ in problem (2.3) is equivalent to minimizing $\|F\|_F^2$ with S replaced by SQ. Replacing E by QFQ^T in (2.3) the equivalent problem is:

$$\min\left\{\|F\|_F^2 : \quad (Q^T h)^T F(Q^T h) = d - h^T h, \quad F = F^T\right\}. \tag{2.6}$$

The solution to this problem is very easy if Q is chosen so that $Q^T h = \theta e_1$, where $\theta^2 = h^T h$ and e_1 is the first column of the identity matrix. This is achieved by letting Q be an appropriate Householder matrix. Then the problem is trivially to

$$\min\left\{\|F\|_F^2 : \quad \theta^2 F_{11} = d - \theta^2, \quad F = F^T\right\},$$

which has solution $F_{11} = d/\theta^2 - 1$, and all other entries of F are zero. Thus $I+F$ is the identity matrix except that its first entry is replaced by d/θ^2 and the updated matrix S^+ is obtained simply by multiplying the first column of the matrix SQ by $|\theta|/\sqrt{d}$. The product SQ takes only $\mathcal{O}(n^2)$ multiplications because of the special form of the Householder matrix. The excellent numerical stability properties of operating with orthogonal matrices and simple column scaling are thus obtained in this approach but the first approach may be preferred if maintaining triangularity of the factor S^+ is required. The latter approach shows also that if p is any column of S (or a scalar multiple) then the solution to problem (2.3) is obtained by simply rescaling this column. This extends readily to rescaling any number of columns. That is, if each column $s_j, j = 1, 2, \ldots, n$, of the matrix S is used to obtain corresponding second directional derivative estimates d_j, then the problem

$$\min\{\|E\| : \; s_j^T S^{-T}[I+E]S^{-1} s_j = d_j, \; j = 1, 2, \ldots, n, \; E = E^T\}, \tag{2.7}$$

is solved (in terms of updating S) by simply scaling each column of S so that

$$s_j^+ = s_j/\sqrt{d_j} \tag{2.8}$$

because

$$S^{-1} s_j = e_j. \tag{2.9}$$

This seems to be a quite natural approach. In the next section the case where both problems (2.3) and 2.7) are combined to provide $n+1$ conditions is considered.

3. Directional derivatives and the weak quasi-Newton condition. Frequently, in derivative-free optimization algorithms it happens that central difference formulas are used to estimate a gradient vector or first directional derivatives along a set of directions. At negligible extra cost this also provides second directional derivatives (see, for example, [5], [6]). Often these are then used to define a search direction p so that $n+1$ directional derivatives may have been estimated by quadratic interpolation at the end of each iteration of the optimization algorithm. Specifically we suppose that, if G denotes the true second derivative matrix (constant if $f(x)$ is quadratic), then estimates $d_j \approx s_j^T G s_j,\ j = 1, 2, \ldots, n$ and $d \approx p^T G p$, are available. In this section we assume that the information is consistent with a positive definite second derivative matrix. Equation (2.8) shows that we can rescale each column of S as soon as the second directional derivatives are estimated so we suppose $d_j = 1,\ j = 1, 2, \ldots, n$, and and we require to calculate B^+ (implicitly by updating S) by solving the problem

$$\min\{\|E\| :\ s_j^T B^+ s_j = 1,\ j = 1, 2, \ldots, n,\ p^T B^+ p = d\}, \tag{3.1}$$

where, as before, $B^+ = S^{-T}[I + E]S^{-1}$. Again writing $p = Sh$ and using equation (2.9) the problem is

$$\min\{\|E\| :\ e_j^T E e_j = 0,\ j = 1, 2, \ldots, n,\ h^T E h = d - h^T h, \}, \tag{3.2}$$

where we have chosen not to include the symmetry constraint $E = E^T$ in (3.1), (3.2), anticipating that the solution will be symmetric automatically. The solution is

$$E = \mu[hh^T - \mathrm{diag}(hh^T)] \tag{3.3}$$

where μ is the scalar

$$\mu = \frac{d - h^T h}{(h^T h)^2 - \sum h_i^4}.$$

This formula for μ is well-defined provided that h has at least two non-zero components because the denominator is then guaranteed to be positive. If h has only one non-zero component, say $h_k \neq 0$, then there is a potential inconsistency in the constraints $s_k^T B^+ s_k = d_k$ and $p^T B^+ p = d$. This may be unlikely to happen in a practical algorithm but there may be potential instabilites when all but one component of h is tiny. In such a case a simple remedy is to drop the offending constraint. This simplifies the problem to one of re-scaling each of the directions as in equation (2.8). Notice that, if $n = 2$, then there are only 3 entries of the symmetric matrix B^+ to calculate so the exact second derivative matrix is obtained (provided $h_1 \neq 0$ and $h_2 \neq 0$) by imposing the $n+1$ constraints in (3.1). As in Section 2, the next step is to update S using the form

$$[B^+]^{-1} = [S^+ S^{+T}] = S[I - \mu\mathrm{diag}(\mathrm{hh^T}) + \mu\mathrm{hh^T}]^{-1}\mathrm{S^T}. \tag{3.4}$$

Usually this can be achieved in $\mathcal{O}(n^2)$ multiplications because the matrix inside the brackets is a rank-1 correction to a diagonal matrix but the method of calculation is deferred to Section 5.

4. Directional derivatives and the strong quasi-Newton condition.

In this section the case is considered where the strong quasi-Newton condition is imposed on B^+ through the equation

$$B^+ p = y, \tag{4.1}$$

simultaneously with second directional derivative conditions

$$s_i^T B^+ s_i = d_i, \quad i = 1, \ldots, n. \tag{4.2}$$

As before the matrix B^+ is represented in the form

$$B^+ = S^{-T}[I+E]S^{-1}$$

and the symmetric correction matrix, E, of minimum Frobenius norm is sought subject to satisfying, if possible, the conditions (4.1) and (4.2). Therefore, the problem to be solved is

$$\min \left\{ \|E\|_F^2 : \quad E_{ii} = d_i - 1, \quad Eh = z - h, \quad E = E^T \right\}. \tag{4.3}$$

Here, h is the vector satisfying $p = Sh$, and $z = S^T y$. If the (strong) quasi-Newton condition is written in the form $\frac{1}{2}[E + E^T]h = z - h$ then there is no need to include the symmetry condition - it will be satisfied automatically. Following the approach of Greenstadt [10, 11, 12, 13] the solution to the quadratic programming problem (4.3) is readily shown to have the form

$$I + E = D - \mathrm{diag}(\lambda h^T + h\lambda^T) + \lambda h^T + h\lambda^T, \tag{4.4}$$

where $D = \mathrm{diag}(d_1, \ldots, d_n)$. It is easy to verify that (4.2) is satisfied by (4.4) for any choice of the vector λ and it remains to choose $\lambda \in R^n$ to satisfy the equation $Eh = z - h$ or equivalently,

$$[D - \mathrm{diag}(\lambda h^T + h\lambda^T) + \lambda h^T + h\lambda^T]h = z. \tag{4.5}$$

The vector λ can be isolated in Equation (4.5) by noticing that $(h\lambda^T)h = (hh^T)\lambda$ and that

$$\mathrm{diag}(\lambda h^T + h\lambda^T)h = 2\,\mathrm{diag}(hh^T)\lambda$$

which allows Equation (4.5) to be re-written as

$$(h^T h)\lambda - 2\,\mathrm{diag}(hh^T)\lambda + (hh^T)\lambda = z - Dh,$$

or, letting $\hat{h}$ denote the unit vector, $\hat{h} = h/\|h\|$, the matrix form of the system of linear equations defining λ is

$$[I - 2\,\mathrm{diag}(\hat{h}\hat{h}^T) + \hat{h}\hat{h}^T]\lambda = (z - Dh)/(h^T h). \tag{4.6}$$

The following theorem provides a simple condition on h which guarantees that the linear system (4.6) is solvable uniquely for λ.

Theorem 4.1. *Let $\hat{u} \in R^n$ be a unit vector, and let* $\mathrm{nnz}(\cdot)$ *denote the number of nonzero components in the vector $(\cdot)$. Then the matrix*

$$A = I - 2\,\mathrm{diag}(\hat{u}\hat{u}^T) + \hat{u}\hat{u}^T, \tag{4.7}$$

is positive definite iff $\mathrm{nnz}(\hat{u}) > 2$*, and positive semi-definite, with* $\mathrm{rank}(A) = n - 1$*, if* $\mathrm{nnz}(\hat{u}) \leq 2$.

Proof.

If $\mathrm{nnz}(\hat{u}) = 1$ with, say, $\hat{u}_j = \pm 1$ then A is the identity matrix except that the jth diagonal entry is zero. Therefore, A is singular (positive semi-definite with rank $n - 1$).

If $\mathrm{nnz}(\hat{u}) = 2$ with, say, $\hat{u}_j = \cos\theta, \quad \hat{u}_k = \sin\theta$, then A is the identity matrix except for the 2×2 positive semi-definite submatrix,

$$\begin{bmatrix} a_{jj} & a_{jk} \\ a_{kj} & a_{kk} \end{bmatrix} = \begin{bmatrix} \sin^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \cos^2\theta \end{bmatrix}$$

so again A is singular (positive semi-definite with rank $n-1$).

If $\mathrm{nnz}(\hat{u}) \geq 3$ then consider the diagonal matrix

$$I - 2\,\mathrm{diag}(\hat{u}\hat{u}^T). \tag{4.8}$$

This matrix has at most one non-positive diagonal entry because $\hat{u}$ has at least three nonzero components. If all diagonal entries are positive then A is the sum of a positive definite (diagonal) matrix and a rank one positive semi-definite matrix and is, therefore, positive definite. Otherwise, the diagonal matrix (4.8) has exactly one non-positive diagonal entry, say its jth diagonal entry. In this case A can be rewritten:

$$A = D_j + \hat{u}\hat{u}^T - 2\hat{u}_j^2 e_j e_j^T, \tag{4.9}$$

where D_j is the diagonal matrix (4.8) except that its jth diagonal entry is 1, that is

$$[D_j]_{ii} = 1 - 2\hat{u}_i^2 > 0, \quad i \neq j; \quad [D_j]_{jj} = 1,$$

and e_j denotes the jth column of the identity matrix. Since D_j is positive definite, this form shows that A has at most one non-positive eigenvalue (interlacing eigenvalue theorem). Therefore, A is positive-definite if and only if A has a positive determinant. Now $\det(D_j) > 0$, and

$$\det(A) = \det(D_j)\det(I + D_j^{-1}\hat{u}\hat{u}^T - 2\hat{u}_j^2 e_j e_j^T), \tag{4.10}$$

because $D_j^{-1}e_j = e_j$. Using the well-known result (see [1], for example) that

$$\det(I + xy^T + wz^T) = (1 + y^T x)(1 + z^T w) - (z^T x)(y^T w)$$

and

$$1 - \hat{u}_j^2 = \sum_{i \neq j} \hat{u}_i^2 > 0,$$

equation(4.10) becomes

$$\begin{aligned}
\det(A) &= \det(D_j)\{(1 + \hat{u}^T D_j^{-1}\hat{u})(1 - 2\hat{u}_j^2) + 2\hat{u}_j^4\} \\
&= \det(D_j)\left(1 - \hat{u}_j^2 + (1 - 2\hat{u}_j^2)\sum_{i \neq j}\frac{\hat{u}_i^2}{1 - 2\hat{u}_i^2}\right) \\
&= \det(D_j)\left(\sum_{i \neq j}\frac{2\hat{u}_i^2(1 - \hat{u_i}^2 - \hat{u}_j^2)}{1 - 2\hat{u}_i^2}\right) \\
&> 0.
\end{aligned}$$

Therefore, A is positive definite iff $\mathrm{nnz}(\hat{\mathrm{u}}) \geq 3$ and positive semi-definite with rank $n-1$ if $\mathrm{nnz}(\hat{\mathrm{u}}) < 3$.

□

Corollary 4.2. *The necessary and sufficient condition for equation (4.6) to have a unique solution is that* $\mathrm{nnz}(\mathrm{h}) \geq 3$.

The above result is not surprising because the cases where $\mathrm{nnz}(\mathrm{h}) < 3$ include those where $n = 2$ or $n = 1$. If $n = 2$ then taking symmetry into account there are 3 entries of E to be estimated but equations (4.1) and (4.2) provide 4 scalar conditions.

Similarly, when $n = 1$ there are 2 scalar conditions for the one entry of E. If $n = 3$ and no component of h is zero then there are six scalar conditions for the six unknown entries defining the symmetric matrix E. The exact second derivative matrix is then obtained by the calculation described above. Theorem 4.1 also shows that for larger values of n equations (4.1) and (4.2) may be inconsistent when the data used to define the correction E lies in a subspace of dimension less than three. More importantly it provides an easy test for what is referred to in [4] as "poisedness."

To illustrate this let $h_1 = 1,\ h_2 = 10^{-8} = h_3$, and let $\hat{u} = h/\|h\|$. If $\hat{u}$ and the matrix A defined by equation (4.7) are calculated using standard IEEE double precision arithmetic then (using MATLAB) the calculated matrix, float(A), say, produces a negative calculated eigenvalue (again using MATLAB) and the condition number of A exceeds 10^{15}. Theorem 4.1 states that the exact matrix A is positive definite. Clearly, it would be unwise to attempt to calculate λ in equation (4.6) using such an ill-conditioned matrix. This is easily avoided by applying a threshold policy. For example, if the number of components of $\hat{u}$ that have absolute value greater than, say, $\tau = 10^{-5}$ is less than three the conditions (4.1) and (4.2) are potentially inconsistent and one or more should be discarded. The simplest ones to discard are those in the set (4.2) for which $|\hat{h}_j| < \tau$ because these are possibly in conflict with the information provided in the vector quasi-Newton condition (4.1). Fortunately, this strategy should not require frequent loss of information when n is large because it is unlikely that there will be too many tiny components in $\hat{h}$ and the condition number of A can be expected to be small. In limited numerical trials the calculated condition number rarely exceeded single figures.

Of course, in practice it is not necessary and certainly not desirable to form the matrix A in order to solve equation (4.6). At worst A is just a rank-2 correction to a positive definite diagonal matrix and for large n it is most likely just a rank-1 correction to a positive definite diagonal matrix. Therefore, all the updating formulas considered so far are expressible in terms of vectors and diagonal matrices that can be calculated in $\mathcal{O}(n)$ floating point operations. In the next Section it is shown that the update formulas can be applied to S stably and efficiently.

5. Updating S efficiently. All the updates derived in this paper can be applied to the matrix S in $\mathcal{O}(n^2)$ multiplications and additions. Already this has been established for the weak quasi-Newton condition in Section 2. For the update (3.4) we first note that we can only apply the update if

$$I - \mu \mathrm{diag}(\mathrm{hh}^{\mathrm{T}}) + \mu \mathrm{hh}^{\mathrm{T}} \tag{5.1}$$

is positive definite. If the diagonal matrix $D_\mu = [I - \mu \mathrm{diag}(\mathrm{hh}^{\mathrm{T}})]$ is also positive definite (which is always the case when μ is negative) then

$$I - \mu \mathrm{diag}(\mathrm{hh}^{\mathrm{T}}) + \mu \mathrm{hh}^{\mathrm{T}} = \mathrm{D}_\mu^{\frac{1}{2}}[\mathrm{I} + \mu \tilde{\mathrm{h}}\tilde{\mathrm{h}}^{\mathrm{T}}]\mathrm{D}_\mu^{\frac{1}{2}}, \qquad \tilde{\mathrm{h}} = \mathrm{D}_\mu^{-\frac{1}{2}}\mathrm{h}.$$

Then the update on S can be completed using the technique described in Section 2 for the weak quasi-Newton update. If D_μ is not positive definite but the matrix (5.1) is, then we proceed as in Section 4 to write (5.1) as a rank-2 correction to a positive-definite diagonal matrix following the procedure described by equations (4.7)-(4.9). This is then treated as two sequential updates. Either way we are left with an update of the form $\tilde{S}[I + \eta \tilde{h}\tilde{h}^T]^{-1}\tilde{S}^T$ which can be handled in the same way as the simple update described in Section 2, although it may be necessary to apply an update of the form (2.5) twice. These techniques can be applied to the more complicated update

(4.4) by re-writing the trailing terms of (4.4) as $\lambda h^T + h\lambda^T = uu^T - vv^T$ (see [2] for details). Then we are left with two symmetric rank-1 changes and we can apply the techniques already described.

6. Concluding remarks. Several direction set updating schemes for derivative-free optimization have been described. The examples considered in this paper are far from exhaustive. Many other interpolation conditions might be included in place of or in addition to those already considered. Of course the updating scheme forms only a small part of the overall optimization algorithm and it remains to be seen whether the ideas presented here will lead to more efficient or robust algorithms for derivative-free optimization. In a forthcoming paper we will report on numerical trials for a derivative-free trust-region algorithm for unconstrained optimization that makes direct use of the updates described here.

Acknowledgement: This work was initiated whilst the first author was Visiting Professor at the University of Leeds. Support from the School of Computing and in particular from Professors Ken Brodlie and Roger Boyle is gratefully acknowledged.

REFERENCES

[1] D.S. Bernstein. *Matrix Mathematics.* Princeton University Press, Princeton, N.J., 2005.

[2] K.W. Brodlie, A.R. Gourlie, and J. Greenstadt. Rank-one and rank-two corrections to positive definite matrices expressed in product form. *J. Inst. Math. Applics.*, 11:73–82, 1973.

[3] A.R. Conn, N.I.M. Gould, and Ph.L. Toint. *Trust-Region Methods.* MPS-SIAM Series on optimization, Philadelphia, PA, 2000.

[4] A.R. Conn, K. Scheinberg, and Ph.L. Toint. On the convergence of derivative-free methods for unconstrained optimization. In M.D. Buhmann and A. Iserles, editors, *Approximation theory and optimization*, pages 83–108, Cambridge, 1997. Cambridge University Press.

[5] I.D. Coope. A conjugate direction implementation of the BFGS algorithm with automatic scaling. *J. Austral. Math. Soc. Ser. B*, 31:122–134, 1989.

[6] I.D. Coope and C.J. Price. Positive bases in numerical optimization. *Computational Optimization & Applications*, 21:169–175, 2002.

[7] J.E. Dennis and R.B. Schnabel. Least change secant methods for quasi-Newton methods. *SIAM Review*, 21:443–459, 1979.

[8] D. Goldfarb. A family of variable metric methods derived by variational means. *Math. Comp.*, 24:23–26, 1970.

[9] G.H. Golub and C.F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.

[10] J. Greenstadt. Variations on variable-metric methods. *Math. Comp.*, 24:1–22, 1970.

[11] J. Greenstadt. A quasi-Newton method with no derivatives. *Math. Comp.*, 26:145–166, 1972.

[12] J. Greenstadt. Revision of a derivative-free quasi-Newton method. *Math. Comp.*, 32:201–221, 1978.

[13] J. Greenstadt. A quasi-quasi-Newton method for generating quasi-Choleski factors. *SIAM J. Sci. Stat. Comput.*, 6:425–437, 1985.

[14] T.G. Kolda, R.M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003.

[15] M.J.D. Powell. Direct search algorithms for optimization calculations. *Acta Numerica*, 7:287–336, 1998.

[16] Ph.L. Toint. On sparse and symmetric updating subject to a linear equation. *Maths. Comp.*, 31:954–961, 1977.