

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

A sparse quasi newton update derived variationally with a nondiagonally weighted frobenius norm

Toint, Ph L.

Published in:
Mathematics Of Computation

DOI:
[10.1090/S0025-5718-1981-0628706-6](https://doi.org/10.1090/S0025-5718-1981-0628706-6)

Publication date:
1981

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for pulished version (HARVARD):
Toint, PL 1981, 'A sparse quasi newton update derived variationally with a nondiagonally weighted frobenius norm', *Mathematics Of Computation*, vol. 37, no. 156, pp. 425-433. <https://doi.org/10.1090/S0025-5718-1981-0628706-6>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A Sparse Quasi-Newton Update Derived Variationally With a Nondiagonally Weighted Frobenius Norm

By Ph. L. Toint

Abstract. The problem of symmetric sparse updating is considered from a variational point of view and a new class of sparse symmetric quasi-Newton updating formulae is derived. This class results from the use of a nondiagonally weighted Frobenius norm. The computation of the update involves only one positive definite and symmetric linear system that has the same sparsity pattern as the problem itself.

1. Introduction. There has been, in the past few years, a growing interest in deriving square matrix updating formulae that preserve any sparsity pattern present in the original matrix. These updates are usually required to possess the quasi-Newton property, i.e., the updated matrix has to satisfy a linear equation that makes it useful as an approximation to some other unknown matrix. This is the case in the field of nonlinear systems of equations where the approximated matrix is the system's Jacobian, and in minimization of nonlinear functions where the approximated matrix is the second derivative of the objective function. In the last case, the updated matrix is also required to be symmetric.

A number of updates are now well known in the case when no sparsity condition applies, both in nonlinear equations (Broyden [1], Ortega and Rheinboldt [10]) and in unconstrained minimization (Davidon [2], Powell [11], Huang [8]). Updates that preserve sparsity appeared more recently, first in nonlinear equations (Schubert [13]) and, somewhat later, in optimization (Toint [16], [18], Marwil [9], Shanno [14]). Their theoretical behavior has also been investigated in some recent papers (Toint [17], Dennis and Schnabel [4], Powell [12], Dennis and Walker [5]).

A fruitful approach to these updates (both sparse and nonsparse) is the so-called "variational approach". Let us consider the symmetric updating problem (i.e., the optimization updating problem) and assume that A is the matrix we want to update so that the updated matrix A^* satisfies, for given nonzero vectors x and w in \mathbf{R}^n ,

$$(1) \quad A^*x = w,$$

$$(2) \quad A^* = (A^*)^T,$$

It is clear that conditions (1) and (2) alone do not uniquely determine A^* , except when A^* is required to be a diagonal matrix. In general, we may fix the remaining degrees of freedom by remembering that there may be some valuable information in A already and, consequently, by requiring A^* to be as close as possible to A in

Received February 8, 1980; revised July 21, 1980 and January 28, 1981.

1980 *Mathematics Subject Classification.* Primary 65F30; Secondary 15A24.

Key words and phrases. Nonlinear optimization, sparsity, matrix updating.

© 1981 American Mathematical Society
0025-5718/81/0000-0163/\$03.25

some matrix norm. The most used matrix norm (see Dennis and Moré [3]) is the weighted Frobenius norm defined by

$$(3) \quad \|M\|_W \triangleq \{\text{Tr}(W^{-1}MW^{-1}M^T)\}^{1/2}$$

for $M \in \mathbf{R}^{n \times n}$ and W a symmetric positive definite matrix. Expression (3) is also equivalent to

$$(4) \quad \|M\|_W = \left\{ \sum_{i,j=1}^n [W^{-1/2}MW^{-1/2}]_{ij}^2 \right\}^{1/2}$$

when M is symmetric, which reduces to the usual Frobenius norm if one chooses W as the identity matrix. This variational approach leads to some of the classical dense updates (Greenstadt [7], Goldfarb [6], Dennis and Moré [3]) and the sparse update derived in [16] also belongs to the same class. When nonlinear systems are considered, condition (2) may be dropped, yielding other classical formulae ([1], [13]).

However, the weighting matrix W is of vital importance in unconstrained minimization because there is some natural scaling associated with the problem. For if the second derivative matrix at the optimum point, H^* say, is positive definite, then the scaling $(H^*)^{-1/2}M(H^*)^{-1/2}$ of (4) gives a problem where the resulting matrix approaches the identity as M approaches H^* , which is highly desirable. Of course, H^* is unknown in practice. Nevertheless, one can exploit its existence by using any positive definite matrix that contains some information about H^* , a good candidate being A , the actual approximation to the hessian.

Unfortunately, it has not been possible until now to use that strategy in the sparse case, and all the formulae that have been proposed so far use a weighting matrix W that is diagonal whatever the sparsity pattern of A . This choice of W results in a generally poor scaling of A^* due to the fact that the eigenvectors of W are always the vectors of the canonical basis, while those of the true hessian may be quite different.

It is the purpose of this short paper to propose an updating procedure that uses a nondiagonal weighting matrix W while preserving sparsity as well as symmetry and the quasi-Newton equation (1). W will not be allowed to be a general symmetric and positive definite matrix, but we will require W to have a special form that will make a practical updating procedure possible. It is worth noting that the resulting new class of sparse updates includes a generalization of the well-known DFP formula to the sparse case. This is because the form that will be assumed for W will allow the equation $Wx = w$ to be satisfied (see [12] for a more explicit discussion).

Section 2 presents a more detailed formulation of the problem as well as a practical updating procedure. Section 3 deals with the formal derivation of the new class, while Section 4 contains some concluding remarks.

2. Problem Formulation and Description of the Updating Procedure. Assume that A is an $n \times n$ sparse symmetric matrix of real numbers. Assume moreover that the sparsity conditions that apply to A are described by the conditions

$$(5) \quad A_{ij} = 0, \quad (i, j) \in I,$$

where I is a set of ordered pairs of integers in the range 1 to n . Assume also that, for $i = 1, \dots, n$,

$$(6) \quad (i, i) \in J,$$

where J is the set of ordered pairs of integers in the considered range that do not belong to I . Condition (6) states that no sparsity condition applies to the diagonal, which is not a very heavy restriction since it is quite natural in the unconstrained optimization problem (see Toint [17] for a detailed comment on this question). We also assume that conditions (5) are consistent with the symmetry of A , i.e., that, if $(i, j) \in I$, then $(j, i) \in I$ also.

This paper is concerned with the problem of finding a correction matrix E such that the conditions

$$(7) \quad A^* = A + E,$$

$$(8) \quad A_{ij}^* = 0, \quad (i, j) \in I,$$

hold together with conditions (1) and (2) and such that $\|A^* - A\|_W$ is minimum, where W is a given symmetric positive definite matrix and where $\|\cdot\|_W$ is defined by (3) or (4).

As stated in the introduction, we require W to have a special form; namely, we require W to be a symmetric positive definite rank two modification of the identity matrix whose inverse W^{-1} can be written in the form

$$(9) \quad W^{-1} = I + \alpha xx^T + \beta(xy^T + yx^T)$$

for x given by (1) and any $\alpha, \beta \in \mathbf{R}$ and $y \in \mathbf{R}^n$. We denote by \mathbf{M} the class of interest, i.e.,

$$(10) \quad \mathbf{M} = \{W \in \mathbf{R}^{n \times n} \mid W \text{ is positive definite and (9) is satisfied}\}.$$

It is important to observe that the BFGS correction to the identity matrix

$$(11) \quad W = I - \frac{xx^T}{\|x\|^2} + \frac{ww^T}{x^T w}$$

belongs to \mathbf{M} if and only if $x^T y > 0$, the parameters α, β , and y in (9) being given by

$$(12) \quad \alpha = \frac{\|y\|^2 + x^T w}{(x^T w)^2}, \quad \beta = \frac{-1}{x^T w}, \quad y = w.$$

We now formulate our problem: we seek the correction E that, for a given $W \in \mathbf{M}$, minimizes

$$(13) \quad \frac{1}{2} \|E\|_W^2$$

subject to

$$(14) \quad Ex = r,$$

$$(15) \quad E = E^T,$$

$$(16) \quad E_{ij} = 0, \quad (i, j) \in I,$$

where

$$(17) \quad r \stackrel{\Delta}{=} w - Ax$$

is the residual vector. [N.B. If there is no sparsity ($I = \emptyset$) and if the equation $Wx = w$ holds, it results from [4] that this method gives the DFP correction to A .]

In order to describe the updating procedure, we define the vectors $x(i)$ ($i = 1, \dots, n$) by the relations

$$(18) \quad x(i)_j \stackrel{\Delta}{=} \begin{cases} x_j, & (i, j) \in J, \\ 0, & (i, j) \in I, \end{cases}$$

as well as the operator $P_W^{\mathbf{B}}[\cdot]$ that projects a given matrix onto the linear manifold \mathbf{B} w.r.t. the weighted Frobenius norm induced by W as defined in (3). One important particular member of this projector family is the unweighted Frobenius sparsity projector (sometimes called the "gangster operator"). If we let

$$(19) \quad \mathbf{S} = \{M \in \mathbf{R}^{n \times n} \mid M_{ij} = 0 \text{ if and only if } (i, j) \in I\},$$

then this projector applied to any matrix M yields

$$(20) \quad P_I^{\mathbf{S}}[M]_{ij} = \begin{cases} M_{ij}, & (i, j) \in J, \\ 0, & (i, j) \in I. \end{cases}$$

In general, if

$$(21) \quad \mathbf{A} \stackrel{\Delta}{=} \{M \in \mathbf{R}^{n \times n} \mid M = M^T \text{ and } Mx = y\}$$

and

$$(22) \quad \mathbf{V} \stackrel{\Delta}{=} \mathbf{A} \cap \mathbf{S},$$

the matrix A^* is defined by the equation

$$(23) \quad A^* = P_W^{\mathbf{V}}[A]$$

because of (7) and (13)–(17). It is proved in Section 3 that A^* can be calculated by the following updating procedure:

(i) Calculate the matrix \hat{N} from

$$(24) \quad \hat{N} = P_I^{\mathbf{S}}[N],$$

where

$$(25) \quad N = \beta(ry^T + yr^T) + \beta^2(r^T y)(xy^T + yx^T) + \beta^2(r^T x)yy^T.$$

(ii) Solve the linear system of equations in z given by

$$(26) \quad Qz = r + \hat{N}x,$$

where Q has the elements

$$(27) \quad Q_{ij} = x(i)_j x(j)_i + \delta_{ij} \|x(i)\|^2 \quad (i, j = 1, \dots, n)$$

(δ_{ij} is the Kronecker symbol).

(iii) Then build the correction matrix

$$(28) \quad E = P_I^{\mathbf{S}}[zx^T + xz^T] - \hat{N}.$$

It is important to observe that the matrix Q defined in (27) has the same sparsity pattern as A and is symmetric. Moreover, Q is positive definite if and only if

$$(29) \quad \|x(i)\| \neq 0, \quad i = 1, \dots, n,$$

as proved in Toint [16]. This implies that z is well defined (and relatively easy to compute by using available sparse symmetric solvers) provided that (29) holds.

One can also observe in (i)–(iii) that the obtained formula does not depend on the value of the parameter α of (9).

Finally, A^* is defined by (7) and clearly satisfies the sparsity conditions (8).

In the case when some of the vectors $x(i)$ are zero, we proceed as in the case $W = I$ (see [16]), i.e., we reduce the size of the linear system (26) by deleting the j th row and column of (26) whenever $j \in K = \{i \mid \|x(i)\| = 0\}$. The j th row and column of E are also set to zero.

3. Formal Derivation of the Update. This section presents the formal solution of (13)–(16).

Assume first that no $\|x(i)\|$ is zero and observe that, in light of (16), (14) may be rewritten as

$$(30) \quad \sum_j E_{ij}x(i)_j = r_i \quad (i = 1, \dots, n).$$

As in [16], we now perform the change of variables $2E = B + B^T$ and transform the constrained problem (13)–(16) into the equivalent formulation

$$\min_{\frac{1}{8}} \|B + B^T\|_W^2$$

subject to

$$(B + B^T)x = 2r, \quad B_{ij} + B_{ji} = 0, \quad (i, j) \in I.$$

This new problem has the Lagrangian

$$\begin{aligned} L &= \frac{1}{8} \text{Tr} [W^{-1}(B + B^T)W^{-1}(B^T + B)] \\ &\quad - \sum_{i=1}^n \lambda_i \left[\sum_{j=1}^n (B_{ij} + B_{ji})x(i)_j - 2r_i \right] - \sum_{(i,j) \in I} \gamma_{ij}(B_{ij} + B_{ji}) \\ &= \frac{1}{8} \text{Tr} [W^{-1}BW^{-1}B] + \frac{1}{8} \text{Tr} [W^{-1}B^TW^{-1}B^T] \\ &\quad + \frac{1}{4} \text{Tr} [W^{-1}BW^{-1}B^T] - \sum_{i=1}^n \lambda_i \left[\sum_{j=1}^n (B_{ij} + B_{ji})x(i)_j - 2r_i \right] \\ &\quad - \sum_{(i,j) \in I} \gamma_{ij}(B_{ij} + B_{ji}), \end{aligned}$$

where λ_i ($i = 1, \dots, n$) and γ_{ij} ($(i, j) \in I$) are the Lagrange parameters associated with the constraints.

As in [7], we observe that

$$\begin{aligned} \frac{\partial}{\partial B_{km}} [\text{Tr}(BC)] &= \frac{\partial}{\partial B_{km}} \left(\sum_{i,j} B_{ij}C_{ji} \right) = C_{mk}, \\ \frac{\partial}{\partial B_{km}} [\text{Tr}(C^TB)] &= \frac{\partial}{\partial B_{km}} \left(\sum_{ij} C_{ji}B_{ji} \right) = C_{km}. \end{aligned}$$

We now use these relations to compute $\partial L / \partial B_{ij}$. In this calculation, the matrix B that appears in the expressions $W^{-1}BW^{-1}B$, $W^{-1}B^TW^{-1}B^T$, and $W^{-1}BW^{-1}B^T$ can be treated formally as being different in each of its two occurrences. For

example, we obtain

$$\begin{aligned} & \frac{\partial}{\partial B_{ij}} [\text{Tr}(W^{-1}BW^{-1}B)] \\ &= \frac{\partial}{\partial B_{ij}^{(1)}} [\text{Tr}(W^{-1}B^{(1)}W^{-1}B^{(2)})] + \frac{\partial}{\partial B_{ij}^{(2)}} [\text{Tr}(W^{-1}B^{(1)}W^{-1}B^{(2)})] \\ &= \frac{\partial}{\partial B_{ij}^{(1)}} [\text{Tr}(B^{(1)}W^{-1}B^{(2)}W^{-1})] + \frac{\partial}{\partial B_{ij}^{(2)}} [\text{Tr}(W^{-1}B^{(1)}W^{-1}B^{(2)})] \\ &= (W^{-1}B^{(2)}W^{-1})_{ji} + (W^{-1}B^{(1)}W^{-1})_{ji} = 2(W^{-1}B^T W^{-1})_{ij} \end{aligned}$$

since

$$\frac{\partial B_{ij}^{(1)}}{\partial B_{ij}} = \frac{\partial B_{ij}^{(2)}}{\partial B_{ij}} = 1.$$

Hence, differentiation with respect to B_{ij} yields

$$\frac{\partial L}{\partial B_{ij}} = 0 = \begin{cases} \frac{1}{2} [W^{-1}(B + B^T)W^{-1}]_{ij} - [\lambda_i x(i)_j + \lambda_j x(j)_i], & (i, j) \in J, \\ \frac{1}{2} [W^{-1}(B + B^T)W^{-1}]_{ij} - (\gamma_{ij} + \gamma_{ji}), & (i, j) \in I. \end{cases}$$

Rewriting these equations in terms of the symmetric correction E , we now obtain

$$(31) \quad 0 = \begin{cases} (W^{-1}EW^{-1})_{ij} - [\lambda_i x(i)_j + \lambda_j x(j)_i], & (i, j) \in J, \\ (W^{-1}EW^{-1})_{ij} - (\gamma_{ij} + \gamma_{ji}), & (i, j) \in I. \end{cases}$$

Symmetry has thus been enforced by our change of variables without the need of including supplementary Lagrange multipliers. Using the definition of W^{-1} , (9) and (14) yield

$$W^{-1}EW^{-1} = E + \beta(cx^T + xc^T) + \beta^2(c^T y)xx^T + N,$$

where $c \stackrel{\Delta}{=} Ey$ and where N is given by

$$(32) \quad \begin{aligned} N = & \alpha(rx^T + xr^T) + \beta(ry^T + yr^T) + [\alpha\beta(r^T x) + \beta^2(r^T y)](xy^T + yx^T) \\ & + [\alpha^2(r^T x) + 2\alpha\beta(r^T y)]xx^T + \beta^2(r^T x)yy^T. \end{aligned}$$

Hence, using (31), (16) and (24),

$$(33) \quad \begin{aligned} P_I^S [W^{-1}EW^{-1}] &= P_I^S [\lambda x^T + x\lambda^T] \\ &= E + \beta P_I^S [cx^T + xc^T] + \beta^2(c^T y)P_I^S [xx^T] + \hat{N} \end{aligned}$$

and defining

$$(34) \quad d \stackrel{\Delta}{=} \lambda - \beta c,$$

one obtains

$$(35) \quad E = P_I^S [dx^T + xd^T] - \beta^2(c^T y)P_I^S [xx^T] - \hat{N}.$$

Applying now (14) and observing that, because of (27),

$$(36) \quad P_I^S [px^T + xp^T]x = Qp$$

for any $p \in \mathbb{R}^n$, we have

$$(37) \quad r = Ex = Qd - \beta^2(c^T y)P_I^S [xx^T]x - \hat{N}x,$$

and, since Q is nonsingular, this yields

$$(38) \quad dQ^{-1}(r + \hat{N}x) + \beta^2(c^T y)Q^{-1}P_I^S[xx^T]x.$$

Now

$$(39) \quad P_I^S[xx^T]x = \frac{1}{2}Qx$$

by (36), and thus (38) becomes, in light of (26),

$$(40) \quad d = z + \frac{\beta^2}{2}(c^T y)x.$$

Replacing d given by (40) in (35), we observe that the terms in $\beta^2(c^T y)$ cancel out, yielding (28) where N is given by (32). Indeed, because of (40),

$$(41) \quad dx^T + xd^T = zx^T + xz^T + \beta^2(c^T y)xx^T.$$

We now show the remarkable fact that the value of α does not influence the value of E and, consequently, that N may be chosen as (25). We first rewrite (28) as

$$(42) \quad E = P_I^S[Q^{-1}rx^T + xr^TQ^{-1}] + P_I^S[Q^{-1}\hat{N}xx^T + xx^T\hat{N}Q^{-1} - N].$$

Because, by (32),

$$(43) \quad \begin{aligned} Q^{-1}\hat{N}x &= \alpha Q^{-1}P_I^S[rx^T + xr^T]x + \beta Q^{-1}P_I^S[ry^T + yr^T]x \\ &+ [\alpha\beta(r^T x) + \beta^2(r^T y)]Q^{-1}P_I^S[xy^T + yx^T]x \\ &+ [\alpha^2(r^T x) + 2\alpha\beta(r^T y)]Q^{-1}P_I^S[xx^T]x + \beta^2(r^T x)Q^{-1}P_I^S[yy^T]x, \end{aligned}$$

we obtain, after some manipulations and the use of (36) and (39), that all the terms in α cancel out, yielding

$$(44) \quad \begin{aligned} E &= P_I^S[Q^{-1}rx^T + xr^TQ^{-1}] \\ &+ \beta P_I^S[Q^{-1}P_I^S(ry^T + yr^T)xx^T + xx^T P_I^S(ry^T + yr^T)Q^{-1} - ry^T - yr^T] \\ &+ \beta^2(r^T x)P_I^S[Q^{-1}P_I^S(yy^T)xx^T + xx^T P_I^S(yy^T)Q^{-1} - yy^T]. \end{aligned}$$

Since this expression does not depend on α , we may choose $\alpha = 0$ in (32), and we therefore obtain (25) and we are done.

We now turn to the case where one of the $x(i)$, $x(k)$ say, is the zero vector. Then clearly the k th component of $\hat{N}x$ is also zero and r_k should be zero. If this is not the case, then the problem is inconsistent. This situation may occur because of incorrect sparsity constraints or rounding errors, and cannot be corrected in the present calculation. Observe that, by (17), we have $y_k = r_k = x_k = 0$ and consequently the k th row and column of \hat{N} are zero. Moreover, the remaining components of r , namely r_i ($i \neq k$), are independent of the k th row and column of E . We therefore satisfy the quasi-Newton condition (14) by setting the k th row and column of E to zero. This procedure may be repeated for each k such that $x(k)$ is zero, and the resulting reduced problem is nonsingular.

4. Concluding Remarks. At this point, some remarks are of interest:

- (a) All that is needed to obtain the correction E is to solve the $n \times n$ linear system (26) and to build the matrix N . The linear system is sparse, symmetric, and positive definite so that special purpose direct factorization algorithms (based on the minimum degree rule, for example) or conjugate gradients can be used to solve

it. The choice of method may depend on the sparsity structure and the amount of fill-in that is to be expected. A nice feature of the procedure is that the size of the linear system to be solved is not greater than the one that occurs from the straight choice $W = I$.

(b) The matrix N (or \hat{N}) never needs to be stored since it can easily be assembled elementwise when needed.

(c) The correction matrix E is no longer the "sparsification of a rank two matrix" as in the case when $W = I$. Instead, it is easy to see that E is the "sparsification a rank four matrix", that is to say the sparsity constrained entries of E may be replaced by suitable numbers so that the resulting matrix is of rank four. However, E itself is generally of full rank.

(d) The identity matrix appearing in (9) may be replaced by any suitable positive multiple, and (9) becomes

$$(45) \quad W^{-1} = \rho I + \alpha x x^T + \beta (x y^T + y x^T) \quad (\rho > 0).$$

It is straightforward to verify that, in this case, Eqs. (25), (26) and (28) become respectively

$$(46) \quad \bar{N} = \beta \rho (r y^T + y r^T) + \beta^2 (r^T y) (x y^T + y x^T) + \beta^2 (r^T x) y y^T,$$

$$(47) \quad Qz = \rho^2 r + P_I^S [\bar{N}] x,$$

$$(48) \quad E = \frac{1}{\rho^2} P_I^S [z x^T + x z^T - \bar{N}].$$

This may be useful in order to preserve the invariance under scaling of the matrix W (for a discussion of this aspect, see Shanno and Phua [15]).

(e) Although the choice (12) yields the equation

$$(49) \quad Wx = w$$

that is recommended by Powell in a recent paper [12], this equation is not used directly in the present derivation.

Numerical experiments with the proposed update are in progress. It is believed that the new technique will improve the stability of the approximation to the hessian matrix with respect to scaling when compared with the usual choice $W = I$.

5. Acknowledgements. The author wishes to thank Professor M. J. D. Powell for many helpful discussions and A. Bihain and Ch. Duchateau for valuable comments on a first draft of the paper.

Department of Mathematics
Facultes Universitaires de Namur
Rempart de la Vierge
B-5000 Namur, Belgium

1. C. G. BROYDEN, "A class of methods for solving nonlinear simultaneous equations," *Math. Comp.*, v. 19, 1965, pp. 577-593.

2. W. C. DAVIDON, *Variable Metric Method for Minimization*, Report #ANL-5990, (Rev.) A.N.L. Research and Development Report, 1959.

3. J. E. DENNIS & J. MORE, "Quasi-Newton methods, motivation and theory," *SIAM Rev.*, v. 19, 1977, pp. 46-89.

4. J. E. DENNIS & R. B. SCHNABEL, *Least Change Secant Updates for Quasi-Newton Methods*, Report #TR78-344, Dept. of Computer Science, Cornell Univ., 1978.
5. J. E. DENNIS & H. F. WALKER, *Convergence Theorems for Least Change Secant Update Methods*, Report #TR476-(141-171-163)-2, Dept. of Mathematical Science, Rice University, Houston, 1979.
6. D. GOLDFARB, "A family of variable metric methods derived by variational means," *Math. Comp.*, v. 24, 1970, pp. 23–26.
7. J. GREENSTADT, "Variations on variable metric methods," *Math. Comp.*, v. 24, 1970, pp. 1–22.
8. H. Y. HUANG, "Unified approach to quadratically convergent algorithms for function minimization," *J. Optim. Theory Appl.*, v. 5, 1970, pp. 405–423.
9. E. S. MARWIL, *Exploiting Sparsity in Newton-Like Methods*, Ph. D. Thesis, Cornell Univ., 1978.
10. J. M. ORTEGA & W. C. RHEINOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
11. M. J. D. POWELL, "A new algorithm for unconstrained optimization," *Nonlinear Programming* (J. B. Rosen, O. L. Mangasarian and K. Ritter, Eds.), Academic Press, New York, 1970, pp. 31–65.
12. M. J. D. POWELL, *Quasi-Newton Formulae for Sparse Second Derivative Matrices*, Report #DAMTP 1979/NA7, Dept. of Applied Mathematics and Theoretical Physics, Cambridge University (G.B.), 1979.
13. L. K. SCHUBERT, "Modification of a quasi-Newton method for nonlinear equations with a sparse Jacobian," *Math. Comp.*, v. 24, 1970, pp. 27–30.
14. D. F. SHANNO, *On Variable-Metric Methods for Sparse Hessians, The New Method*, Report #MIS TR 27, Dept. of Management Information Systems, University of Arizona, Tucson, 1978.
15. D. F. SHANNO & K. H. PHUA, "Matrix conditioning and nonlinear optimization," *Math. Programming*, v. 14, 1978, pp. 149–160.
16. PH. L. TOINT, "On sparse and symmetric updating subject to a linear equation," *Math. Comp.*, v. 31, 1977, pp. 954–961.
17. PH. L. TOINT, "On the superlinear convergence of an algorithm for solving a sparse minimization problem," *SIAM J. Numer. Anal.*, v. 16, 1979.
18. PH. L. TOINT, "A note about sparsity exploiting quasi-Newton updates," *Math. Programming*, v. 21, 1981.