NBER WORKING PAPER SERIES

REPRESENTING SYMMETRIC RANK 2 UPDATES

David Gay*

Working Paper No. 124

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

February 1976

Preliminary

*NBER Computer Research Center. Research supported in part by National Science Foundation Grants DCR75-10143 and MCS76-00324 to the National Bureau of Economic Research, Inc.

Abstract

Various quasi-Newton methods periodically add a symmetric "correction" matrix of rank at most 2 to a matrix approximating some quantity A of interest (such as the Hessian of an objective function). In this paper we examine several ways to express a symmetric rank 2 matrix Δ as the sum of rank 1 matrices. We show that it is easy to compute rank 1 matrices Δ_1 and Δ_2 such that $\Delta = \Delta_1 + \Delta_2$ and $||\Delta_1|| + ||\Delta_2||$ is minimized, where $||\cdot||$ is any inner product norm. Such a representation recommends itself for use in those computer programs that maintain A explicitly, since it should reduce cancellation errors and/or improve efficiency over other representations. In the common case where Δ is indefinite, a choice of the form $\Delta_1 = \Delta_2^T = xy^T$ appears best. This case occurs for rank 2 quasi-Newton updates Δ exactly when Δ may be obtained by symmetrizing some rank 1 update; such popular updates as the DFP, BFGS, PSB, and Davidon's new optimally conditioned update fall into this category.

Contents

1. Introduction

Various quasi-Newton methods maintain an approximation $A \in \mathbb{R}^{n\times n}$ to some $n\times n$ matrix of interest. In the case of unconstrained minimization, for example, A might approximate the Hessian of the objective function or its inverse (see [Dennis & Moré, 1974]). Such quasi-Newton methods periodically determine a new approximation $\bar{A} = A + \Delta$, which is required to satisfy a "quasi-Newton equation" of the form $\bar{A}w = b$ (i.e. $\Delta w = z \equiv b - Aw$) for certain $w, b \in \mathbb{R}^n$ determined by the method. Frequently A, $\bar{A}$, and hence Δ are required to be symmetric and Δ is chosen to have rank at most 2. In this case A and $\bar{A}$ are also often required to be positive definite. It may be desirable to maintain such matrices A in the factored form $A = LL^T$ and to explicitly update only the factor L: this assures that A has no negative eigenvalues. Brodlie, Gourlay, & Greenstadt [1973] have shown under these conditions that $\bar{A}$ may be expressed as $\bar{A} = (I + qp^T)A(I + pq^T)$ (for certain n-vectors $p, q \in \mathbb{R}^n$) if and only if $\Delta = \bar{A} - A$ may be expressed in the form $\Delta = uu^T - vv^T$, which is possible for such popular quasi-Newton updates as the BFGS and DFP; Davidon [1975] recommends that such a factored representation be used with his optimally conditioned update. (See Davidon's [1975] paper for another view of such factored representations.)

Despite the above, it may sometimes be desirable to maintain A explicitly. Moreover, there exist situations in which the approximation A must be allowed to be indefinite; this happens, for instance, in certain algorithms for solving the nonlinear least squares problem, in which a quasi-Newton approximation is made to only one part of the Hessian [Dennis, 1975]. Thus it is natural to ask how symmetric matrices $\Delta \in \mathbb{R}^{n\times n}$ of rank ≤ 2 may be represented as sums of rank 1 matrices (outer products) and to compare such representations with an eye toward efficiency and accuracy in computer implementations.

If Δ has rank 1 and satisfies $\Delta w = z \neq 0$, then it is easily seen that $w^T z \neq 0$ and $\Delta = \frac{zz^T}{w^T z}$. While the symmetric rank 1 update (SR1) can also be written in any of the forms considered below for rank 2 updates, the simple representation $\Delta = \frac{zz^T}{w^T z}$ would appear to be the most efficient (and accurate). (Of course, any of the standard, normally rank 2 updates may degenerate to the SR1 under certain conditions; in algorithms using such updates, detecting degeneracy may be difficult and it may well be most efficient to always use the rank 2 form of the update.) We therefore consider below various ways to represent a symmetric, rank 2 matrix Δ as the sum of two or three matrices of rank 1.

This paper is organized as follows. The next section presents some background material. Section three considers the common case where Δ has eigenvalues of opposite sign (i.e. is indefinite). Section four briefly considers the semidefinite case (where Δ has two eigenvalues of the same sign) and section five examines "asymmetric" representations $\Delta = \Delta_1 + \Delta_2$ with Δ_1, Δ_1^T, Δ_2, and Δ_2^T all distinct. Finally, section six discusses applications to some quasi-Newton updates and section seven presents our conclusions.

2. Background

We shall have occasion below to refer to both the standard inner product $y^T x$ and a possibly nonstandard one

$$(1) \qquad \langle x,y \rangle = y^T M x$$

defined for $x,y \in \mathbb{R}^n$ by a positive definite (hence symmetric) $n \times n$ matrix $M \in \mathbb{R}^{n \times n}$. We shall also refer both to the corresponding inner product vector norm $||\cdot||$ defined by

(2) $$||x|| = \sqrt{\langle x,x \rangle}$$

and to the matrix norm which this induces. Reasonable choices for M may include I (the $n \times n$ identity matrix) and A^{-1}, if Δ is used in the update $\bar{A} = A + \Delta$ with A positive definite; we shall say more about the choice of M in section 7.

We shall find it useful to classify rank 2 matrices Δ by the signs of their nonzero eigenvalues. Hence we state:

(3) Lemma If $\Delta = \Delta^T \in \mathbb{R}^{n \times n}$ has rank 2, then Δ may be expressed in the form

(4a), (4b), (4c)
$$\Delta = \begin{cases} uu^T - vv^T \\ uu^T + vv^T \\ -(uu^T + vv^T) \end{cases},$$

where u and v are linearly independent, if and only if Δ has

$$\begin{cases} \text{(a) one positive and one negative} \\ \text{(b) two positive} \\ \text{(c) two negative} \end{cases} \quad \text{eigenvalue(s).}$$

Hence Δ may be expressed in exactly one of the forms (4). It is possible to choose u and v in (4) so that $\langle u,v \rangle = 0$.

Proof (Cf Lemma 1 of [Brodlie, Gourlay, & Greenstadt, 1973]): It is easily seen that Δ may be expressed in the form (4) if and only if Δ is respectively (a) indefinite, (b) positive semidefinite, or (c) negative semidefinite.

It remains to show that $\langle u,v\rangle = 0$ is possible. If L is any matrix (e.g., a Cholesky factor of M) such that $M = L^TL$ and x,y are (orthogonal) eigenvectors of

$$L\Delta L^T \text{ such that } L\Delta L^T = \begin{cases} \text{(a)} & xx^T - yy^T \\ \text{(b)} & xx^T + yy^T \\ \text{(c)} & -(xx^T + yy^T) \end{cases},$$

then (4) holds for $u = L^{-1}x$ and $v = L^{-1}y$. ■

In comparing expressions for Δ, we shall employ the following easily proved lemma.

(5) Lemma With Δ as above, if u and v span the column space of Δ, then there exist unique scalers $\mu, \nu, \xi \in \mathbb{R}$ such that $\Delta = \mu uu^T + \nu vv^T + \xi(uv^T + vu^T)$. ■

3. Eigenvalues of Opposite Sign

As we shall see presently, many quasi-Newton updates generate a Δ with one positive and one negative eigenvalue. Hence we shall first examine the case where

$$\Delta = uu^T - vv^T \tag{6}$$

for some linearly independent $u, v \in \mathbb{R}^n$. There are many ways to express Δ in the form (6), as the next lemma shows:

(7) Lemma If (6) holds, then

$$\Delta = \bar{u}\bar{u}^T - \bar{v}\bar{v}^T \tag{8}$$

if and only if there exist $\Theta \in \mathbb{R}$ and $\sigma = \pm 1$ such that

(9a) $\bar{u} = (\sec\theta)\,u + (\tan\theta)\,v$ and

(9b) $\sigma\bar{v} = (\tan\theta)\,u + (\sec\theta)\,v.$

Proof: ($\Longrightarrow$) If (6) and (8) hold, then $\bar{u}$ and $\bar{v}$ must be linear combinations of u and v, say $\bar{u} = \alpha u + \beta v$ and $\bar{v} = \gamma u + \delta v$. Since

$$\Delta = uu^T - vv^T = (\alpha u + \beta v)(\alpha u + \beta v)^T - (\gamma u + \delta v)(\gamma u + \delta v)^T$$

$$= (\alpha^2 - \gamma^2)uu^T - (\delta^2 - \beta^2)vv^T + (\alpha\beta - \gamma\delta)(uv^T + vu^T),$$

lemma (5) implies

(10a,b,c) $\alpha^2 = 1 + \gamma^2$, $\delta^2 = 1 + \beta^2$, and $\alpha\beta = \delta\gamma$.

From (10a,b) we have $\gamma = \sigma_1\sqrt{\alpha^2-1}$ and $\delta = \sigma_2\sqrt{\beta^2+1}$ for the $\sigma_1, \sigma_2 = \pm 1$ of appropriate sign. From (10c) we obtain $\alpha\beta = \sigma_1\sigma_2\sqrt{(\alpha\beta)^2 + \alpha^2 - \beta^2 - 1}$, so $\alpha^2 = \beta^2 + 1$ and hence there exists $\theta \in R$ such that $\alpha = \sec\theta$ and $\beta = \tan\theta$. Inserting these into the above expressions for γ and δ and using (10c), we obtain $\gamma = \sigma\tan\theta$ and $\delta = \sigma\sec\theta$, where $\sigma = \pm 1$, whence (9) follows.
($\Longleftarrow$) Conversely, if (6) and (9) hold, then it is easily verified that (8) also holds. ∎

It seems to be fairly well known that a matrix of the form $xy^T + yx^T$ has one positive and one negative eigenvalue. We can, however, say more than this:

(11) Lemma. If $\Delta = \Delta^T \in \mathbb{R}^{n\times n}$ has rank 2, then Δ has one negative and one positive eigenvalue if and only if

(12) $\Delta = xy^T + yx^T$

for some linearly independent $x, y \in \mathbb{R}^n$. Moreover, if (12) holds, then x and y are essentially unique: if $\Delta = \bar{x}\bar{y}^T + \bar{y}\bar{x}^T$, then there exists $\tau \in \mathbb{R}$, $\tau \neq 0$, such that either

(13) $\quad (\bar{x} = \tau x$ and $\bar{y} = y/\tau)$ or $(\bar{x} = \tau y$ and $\bar{y} = x/\tau)$.

Proof If Δ has eigenvalues of opposite sign, then by Lemma (3) there exist linearly independent $u, v \in \mathbb{R}^n$ such that $\Delta = uu^T - vv^T$: then (12) holds with

(14) $\quad x = u + v$ and $y = \frac{u-v}{2}$.

Conversely, if (12) holds, then $u = \frac{||x||y + ||y||x}{\sqrt{2||x|| \, ||y||}}$

$v = \frac{||x||y - ||y||x}{\sqrt{2||x|| \, ||y||}}$ are such that $\Delta = uu^T - vv^T$ (and $\langle u,v \rangle = 0$), whence Δ has eigenvalues of opposite signs. Finally, if $\Delta = xy^T + yx^T = \bar{x}\bar{y}^T + \bar{y}\bar{x}^T$, then $\bar{x}$ and $\bar{y}$ must be linear combinations of x and y, say $\bar{x} = \alpha x + \beta y$ and $\bar{y} = \tau x + \delta y$. From Lemma (5) we obtain $\alpha\gamma = \beta\delta = 0$ and $\alpha\beta + \beta\gamma = 1$, whence the essential uniqueness (13) follows. ■

Generalizing Powell's [1970c] derivation of the "Powell-symmetric-Broyden" (PSB) update, Dennis [1972] showed that a large family of symmetric rank 2 quasi-Newton updates could be obtained by symmetrizing nonsymmetric rank 1 updates. Thus, given nonzero $w, z \in \mathbb{R}^n$, we may obtain a symmetric rank 2 $\Delta \in \mathbb{R}^{n\times n}$ which satisfies the quasi-Newton equation.

(15) $\quad \Delta w = z$

by starting with the rank 1 update $\Delta_0 = zd^T$ determined by some specified $d \in R^n$ with $d^T w = 1$ and generating the sequence $\Delta_1, \Delta_2, \Delta_3, \ldots$ in which $\Delta_{2j+1} = \frac{1}{2}(\Delta_{2j} + \Delta_{2j}^T)$ and $\Delta_{2j+2} = \Delta_{2j+1} + (z - \Delta_{2j+1}w)d^T$ for $j = 0,1,2, \ldots$. Dennis [1972] has shown that $\lim_{i\to\infty} \Delta_i = \Delta$, where

(16) $\qquad \Delta = zd^T + dz^T - (z^Tw)dd^T.$

Such well known updates as the DFP and BFGS have the form (16) for the properly chosen d. The next lemma is of interest because it characterizes those rank 2 updates Δ which may be expressed in the form (16), a form convenient for certain proof techniques (see, e.g., [Broyden, Dennis, & Moré, 1973]):

(17) <u>Lemma</u>: If $\Delta = \Delta^T \varepsilon \mathbb{R}^n$ has rank 2 and (15) holds, then Δ may be expressed in the form (16) with z and d linearly independent if and only if Δ has eigenvalues of opposite sign.

<u>Proof</u>: If (16) holds, then (12) holds with $x = d$ and $y = z - (\frac{z^Tw}{2})d$, whence Δ has eigenvalues of opposite sign by Lemma (11). Conversely, if Δ has eigenvalues of opposite sign, then by Lemma (11) there exist $x, y \varepsilon \mathbb{R}^n$ such that (12) holds. From (15) we have

(18) $\qquad 0 \neq z = \Delta w = (y^Tw)x + (x^Tw)y$

and we may assume without loss of generality that $x^Tw \neq 0$. Setting $d = \frac{x}{x^Tw}$, we find $d^Tw = 1$ and (using (18))

$$zd^T + dz^T - (z^Tw)dd^T = [(y^Tw)x + (x^Tw)y]\frac{x^T}{x^Tw} + \frac{x}{x^Tw}[(y^Tw)x + (x^Tw)y]^T - [2(x^Tw)(y^Tw)][\frac{x}{x^Tw}][\frac{x}{x^Tw}]^T$$

$$= xy^T + yx^T = \Delta. \blacksquare$$

From the above proof, we see that there are usually exactly two choices for d in (16), there being but one in the exceptional case where $\Delta = zd^T + dz^T$ (ie. where $x^Tw = 0$ or $y^Tw = 0$ with x, y as in (12)).

We have seen that there are various ways to express a symmetric indefinite matrix Δ of rank 2 as the sum of two or three rank 1 matrices. For explicit floating point computations involving Δ, it would appear more efficient and probably more accurate (as there is less chance for roundoff to occur) to use one of the forms (6) or (12) expressing $\Delta = \Delta_1 + \Delta_2$ as the sum of two rank 1 matrices Δ_1 and Δ_2. Form (6) offers a one dimensional family of possibilities, while (12) offers essentially just one. To compare all these possibilities, it seems reasonable to examine $||\Delta_1|| + ||\Delta_2||$, since minimizing this sum should hopefully tend to minimize cancellation error in some sense. (We shall have more to say about the choice of inner product norm (2) below.) Note that the vector and induced matrix norms $||\cdot||$ are related by $||xy^T|| = ||x||\ ||y||$. Using Lemma (7) and the connection (14) between (6) and (12), it is easy to prove:

(19) Lemma: If $\Delta = uu^T - vv^T = \bar{u}\bar{u}^T - \bar{v}\bar{v}^T = xy^T + yx^T$, where u and v are linearly independent with $\langle u,v\rangle = 0$, then

$$||\bar{u}\bar{u}^T|| + ||\bar{v}\bar{v}^T|| \geq ||uu^T|| + ||vv^T|| = ||xy^T|| + ||yx^T|| = ||u||^2 + ||v||^2. \blacksquare$$

Thus, no matter what inner product norm $||\cdot||$ is used to measure the representations (6) and (12) as described above, form (12) rates exactly as well as the best representation of form (6). (Note that the best representation of form (6) depends heavily on the inner product $\langle\cdot,\cdot\rangle$; indeed, given any u,v such that (6) holds, it is possible to find many inner products with respect to which u and v are optimal in the above sense.)

4. The Semidefinite Case

We next briefly consider the case where the rank 2 matrix $\Delta = \Delta^T \in \mathbb{R}^{n\times n}$ has two eigenvalues of the same sign. Since the other case is similar, we assume both nonzero eigenvalues are positive. Just as in the case of mixed signs, there are many ways to express $\Delta = \Delta_1 + \Delta_2$ as the sum of two symmetric rank 1 matrices. This time, however, $||\Delta_1|| + ||\Delta_2||$ depends only on the inner product norm $||\cdot||$ and not on the particular choice of Δ_1 and Δ_2:

(20) Lemma: If $u, v \in \mathbb{R}^n$ are linearly independent, then

(21) $$\Delta = uu^T + vv^T = \bar{u}\bar{u}^T + \bar{v}\bar{v}^T$$

if and only if there exist $\theta \in \mathbb{R}$ and $\sigma = \pm 1$ such that

(22a) $$\bar{u} = (\cos\theta)u - (\sin\theta)v \qquad \text{and}$$

(22b) $$\sigma\bar{v} = (\sin\theta)u + (\cos\theta)v.$$

Moreover, if L is any real matrix such that $L^T L = M$ (with M in (1)), then

(23) $$||uu^T|| + ||vv^T|| = ||\bar{u}\bar{u}^T|| + ||\bar{v}\bar{v}^T|| = \mathrm{trace}(L\Delta L^T).$$

Proof: The equivalence of (21) and (22) follows from reasoning similar to that in the proof of Lemma (7). From (22) we obtain

(24) $$||uu^T|| + ||vv^T|| = ||\bar{u}\bar{u}^T|| + ||\bar{v}\bar{v}^T|| = ||u||^2 + ||v||^2.$$

Let x and y be eigenvectors of $L\Delta L^T$ (such that $x^T y = 0$), scaled so that $L\Delta L^T = xx^T + yy^T$: then $u = L^{-1}x$ and $v = L^{-1}y$ have $<u,v> = 0$ and satisfy (21), and $||u||^2 + ||v||^2 = x^T x + y^T y = \mathrm{trace}\,(L\Delta L^T)$, whence (23) follows form (24). ■

5. Asymmetric Representations

Heretofore we have considered expressing symmetric rank 2 matrices Δ as the sum $\Delta_1 + \Delta_2$ of rank 1 matrices which either are symmetric or else satisfy $\Delta_1^T = \Delta_2$. These expressions require only two vectors to express both Δ_1 and Δ_2 in outer product form. There also exist choices for Δ_1 and Δ_2 which neither are symmetric nor satisfy $\Delta_1^T = \Delta_2$. Indeed, we may state:

(25) Lemma: If $u, v \in \mathbb{R}^n$ are linearly independent and $\sigma = \pm 1$, then

(26) $$\Delta = uu^T + \sigma vv^T = pq^T + rs^T,$$

if and only if there exist $\theta, \tau \in \mathbb{R}$ such that either

(27a) $pq^T = \hat{p}\hat{q}^T$ and $rs^T = \hat{r}\hat{s}^T$ or

(27b) $pq^T = \hat{r}\hat{s}^T$ and $rs^T = \hat{p}\hat{q}^T$, where

(27c) $$\hat{p}\hat{q}^T = ([\cos\theta]u - \sigma[\sin\theta]v)([\cos\theta - \tau\sin\theta]u - [\sin\theta + \tau\cos\theta]v)^T$$

and

(27d) $$\hat{r}\hat{s}^T = ([\sin\theta + \tau\cos\theta]u + \sigma[\cos\theta - \tau\sin\theta]v)([\sin\theta]u + [\cos\theta]v)^T.$$

Moreover, if $\langle u,v\rangle = 0$ and (27) holds, then

(28) $$||pq^T|| + ||rs^T|| \geq ||u||^2 + ||v||^2,$$

with equality in (28) if and only if $\tau = 0$.

Proof: The proof of (26)$\Longleftrightarrow$(27) is similar to that of Lemma (7). To show (28), consider $\Phi(\theta,\tau) \equiv ||pq^T|| + ||rs^T||$, where p, q, r, s are given by (27). It is straightforward to verify that $\frac{\partial\Phi}{\partial\tau}(\theta,0) = 0$ and $\frac{\partial^2\Phi}{\partial\tau^2}(\theta,\bar{\tau}) > 0$

for all $\bar{\tau} \in \mathbb{R}$, whence $\Phi(\Theta,\tau) \geq \Phi(\Theta,0)$, with equality only for $\tau = 0$. But $\Phi(\Theta,0) = ||u||^2 + ||v||^2$ for all $\Theta \in \mathbb{R}$. ■

If $\tau = 0$ and $\sigma = 1$ in (27), then pq^T and rs^T are both symmetric and (27) reduces to the expression described in Lemma (20). On the other hand, if $\tau = 0$ but $\sigma = -1$, then (26) becomes

$$(29) \qquad \Delta = ([\cos\Theta]u + [\sin\Theta]v)([\cos\Theta]u - [\sin\Theta]v)^T + \\ + ([\sin\Theta]u - [\cos\Theta]v)([\sin\Theta]u + [\cos\Theta]v)^T,$$

which involves four pairwise independent vectors unless Θ is an integral multiple of $\pi/4$. For Θ an integral multiple of $\pi/2$, (29) reduces to (6), and for Θ an odd multiple of $\pi/4$, (29) reduces to (12). Other choices of Θ apparently boast no practical advantages.

6. Application to Quasi-Newton Updates

As previously remarked, it may improve efficiency and reduce some cancellation errors to express rank 2 quasi-Newton updates $\bar{A} = A + \Delta$ as the sum $\Delta = \Delta_1 + \Delta_2$ of two rank 1 matrices. This is readily done for some of the popular updates. For example, the direct DFP and inverse BFGS formulae satisfy the quasi-Newton equation $\bar{A}w = b$ by choosing $\bar{A} = A + \frac{zb^T + bz^T}{b^Tw} - \frac{(z^Tw)bb^T}{(b^Tw)^2}$, where $z = b - Aw$; in this case $\bar{A} = A + xy^T + yx^T$ for $x = \frac{b}{b^Tw}$ and $y = z - (\frac{z^Tw}{2})\frac{b}{b^Tw} = (1 - \frac{z^Tw}{2b^Tw})b - Aw$. The direct PSB update, $\bar{A} = A + \frac{zw^T + wz^T}{w^Tw} - \frac{(z^Tw)ww^T}{(w^Tw)^2}$, is readily expressed in a similar form.

And the inverse DFP and direct BFGS updates, $\bar{A} = A + \frac{bb^T}{b^Tw} - \frac{(Aw)(Aw)^T}{w^TAw}$, are already expressed as the sum of two rank 1 matrices, though they may also be expressed in the form $\bar{A} = A + xy^T + yx^T$ with $x = b + \left(\frac{b^Tw}{w^TAw}\right)^{1/2} Aw$ and $y = \frac{b - \left(\frac{b^Tw}{w^TAw}\right)^{1/2} Aw}{2b^Tw}$; the latter form permits the calculation of at least the diagonal elements of $\Delta = \bar{A} - A$ with a smaller bound on the absolute error. For more information on the various updates, see [Dennis & Moré, 1974] and the references cited therein.

Davidon's [1975] optimally conditioned (OC) update is not so easy to express as the sum of two rank 1 matrices. In this case

(30a) $$\Delta = \frac{bb^T}{\beta} - \frac{(Aw)(Aw)^T}{\alpha} + \alpha\phi\left(\frac{b}{\beta} - \frac{Aw}{\alpha}\right)\left(\frac{b}{\beta} - \frac{Aw}{\alpha}\right)^T$$

where $\alpha = w^TAw$, $\beta = w^Tb$, $\gamma = b^TA^{-1}b$, and

(30b) $$\phi = \begin{cases} \dfrac{\beta(\gamma-\beta)}{\alpha\gamma-\beta^2} & \text{if} \quad \beta < \dfrac{2\alpha\gamma}{\alpha+\gamma} \\[2ex] \dfrac{\beta}{\beta-\alpha} & \text{otherwise [SR1]} . \end{cases}$$

If $\beta \geq \frac{2\alpha\gamma}{\alpha+\gamma}$, then Davidon's OC update reduces to the symmetric rank 1 (SR1) update discussed in the introduction. The other case is more complex, but we may approach it in the following general way.

Suppose $s, t \in \mathbb{R}^n$ and

(31) $$\Delta = \sigma ss^T + \tau tt^T + \xi(st^T + ts^T)$$

with $\sigma, \tau, \xi \in R$ and $\tau \neq 0$. Then

(32) $$\Delta = \left(\frac{\sigma\tau-\xi^2}{\tau}\right)ss^T + \tau\left(\frac{\xi}{\tau}s + t\right)\left(\frac{\xi}{\tau}s + t\right)^T.$$

Thus if $\sigma\tau \leq \xi^2$ we find that (12) holds, i.e. $\Delta = xy^T + yx^T$, with

(33a) $$x = \left(\frac{\xi + \sqrt{\xi^2 - \sigma\tau}}{2}\right)s + \left(\frac{\tau}{2}\right)t \quad \text{and}$$

(33b) $$y = \left(\frac{\xi - \sqrt{\xi^2 - \sigma\tau}}{\tau}\right)s + t.$$

In the case of (30) we find that (31) holds with

(34a,b) $$s = Aw, \quad t = b,$$

(34c) $$\xi = \frac{-\phi}{\beta} = \frac{\beta-\gamma}{\alpha\gamma-\beta^2},$$

(34d) $$\tau = \left(1 + \frac{\alpha\phi}{\beta}\right)/\beta = \frac{1 - \alpha\xi}{\beta},$$

and $\sigma = \frac{\phi-1}{\alpha} = \frac{-(\beta\xi+1)}{\alpha}$. Since (30) is only used when $\beta > 0$ and since the positive definiteness of A insures $\alpha > 0$, $\gamma > 0$, and $\alpha\gamma > \beta^2$, we have $\tau > 0$ and

(34e) $$\xi^2 - \sigma\tau = \frac{1 + \xi(\beta-\alpha)}{\alpha\beta} = \frac{2\alpha\gamma - \beta(\alpha+\gamma)}{\alpha\beta(\alpha\gamma-\beta^2)},$$

whence $\xi^2 > \sigma\tau$ if and only if test (30b) does not select the SR1 update. (Note that (32) and Lemma (3) imply the possibly surprising fact that the Δ given by (31) with $\tau \neq 0$ and s, t linearly independent has one negative and one positive eigenvalue if and only if $\xi^2 > \sigma\tau$.) Using (34) and (33), it is thus easy to compute Davidon's OC update (30) in the form $\Delta = xy^T + yx^T$.

Some quasi-Newton methods, such as Powell's [1970a,b] MINFA and Dennis & Mei's [1975] MINOP (which, like Davidon's OCOPTR, uses the OC update) require that both A and A^{-1} be maintained. If the (direct) DFP is used to update A, then the corresponding update for A^{-1} has the form of the (direct) BFGS, and vice versa. If the OC update is used for A, then the corresponding update for A^{-1} is also optimally conditoned in the same sense, whence (using the above notation)

$$\bar{A}^{-1} = A^{-1} + \frac{ww^T}{\beta} - \frac{(A^{-1}b)(A^{-1}b)^T}{\gamma} + \gamma\bar{\phi}\left(\frac{w}{\beta} - \frac{A^{-1}b}{\gamma}\right)\left(\frac{w}{\beta} - \frac{A^{-1}b}{\gamma}\right)^T$$

with $$\bar{\phi} = \begin{cases} \dfrac{\beta(\alpha-\beta)}{\alpha\gamma-\beta^2} = \phi + \dfrac{\beta(\alpha-\gamma)}{\alpha\gamma-\beta^2} & \text{if } \beta < \dfrac{2\alpha\gamma}{\alpha+\gamma} \\ \dfrac{\beta}{\beta-\gamma} & \text{otherwise [SR1].} \end{cases}$$

Thus we may interchange α and γ in (34c,d), determine x and y by (33) with $s = A^{-1}b$ and $t = w$, and compute $\bar{A}^{-1} = A^{-1} + xy^T + yx^T$. In the case of a general Δ as in (31), if A and $\bar{A} = A + \Delta$ are nonsingular, then

$$\begin{aligned}\bar{A}^{-1} = A^{-1} &- \delta^{-1}[\sigma - (\xi^2 - \sigma\tau)(t^TA^{-1}t)]\,(A^{-1}s)(A^{-1}s)^T - \\ &- \delta^{-1}[\tau - (\xi^2 - \sigma\tau)(s^TA^{-1}s)]\,(A^{-1}t)(A^{-1}t)^T - \\ &- \delta^{-1}[\xi + (\xi^2 - \sigma\tau)(s^TA^{-1}t)]\,[(A^{-1}s)(A^{-1}t)^T + (A^{-1}t)(A^{-1}s)^T],\end{aligned}$$

where $$\delta = 1 + \sigma(s^TA^{-1}s) + \tau(t^TA^{-1}t) - (\xi^2-\sigma\tau)[(s^TA^{-1}s)(t^TA^{-1}t) - (s^TA^{-1}t)^2] + \\ + 2\xi(s^TA^{-1}t).$$

This may be expressed as A^{-1} plus two rank 1 matrices by a device like (32).

7. Conclusion

In the common case where A and $\bar{A}$ are positive definite, it may seem desirable to choose $M = A^{-1}$ in (1) and express the update $\bar{A} = A + \Delta = A + \Delta_1 + \Delta_2$ in a form (with Δ_1 and Δ_2 of rank 1) which minimizes $||\Delta_1|| + ||\Delta_2||$ with respect to the inner product norm (2). Such a norm seems quite in keeping with Davidon's [1959, 1975] notion of a "variable metric" determined by the current approximation A. It should select $\Delta_j = p_j q_j^T$ $(j = 1,2)$ which make smaller changes to A in eigendirections corresponding to small eigenvalues in the sense that $|(r^T p_j)(s^T q_j)|$ is reduced for unit eigenvectors r,s of A corresponding to small eigenvalues. Lemmas (19) and (25) establish the remarkable fact that when Δ is indefinite (as is the case for what appear to be the most frequently used updates), a choice of the form $\Delta_1 = xy^T$ and $\Delta_2 = yx^T$ minimizes $||\Delta_1|| + ||\Delta_2||$ no matter what inner product (1) defines the norm (2). We have seen that such representations can be readily programmed. In the opposite case where Δ is semidefinite, (32) shows how Δ may be conveniently represented in the form $\pm\Delta = uu^T + vv^T$; Lemmas (20) and (25) show that any such choice of u and v minimizes $||\Delta_1|| + ||\Delta_2||$. (Footnote: if $\sigma = \tau = 0$ in (31) with s,t linearly independent, then Δ is indefinite; thus if Δ has two eigenvalues of the same sign, then at least one of σ,τ must be nonzero, and by possibly interchanging s and t we may arrange that $\tau \neq 0$ and thus that (32) makes sense.)

8. References

Brodlie, K. W.; Gourlay, A. R.; & Greenstadt, J. (1973), "Rank-one and Rank-two Corrections to Positive Definite Matrices Expressed in Product Form", J. Inst. Maths. Applics 11, pp 73-82.

Broyden, C. G.; Dennis, J.E.; & Moré, J. J. (1973), "On the Local and Superlinear Convergence of Quasi-Newton Methods", J. Inst. Math. Appl. 12, pp. 223-245.

Davidon, W. C. (1959), "Variable Metric Method for Minimization", Argonne Nat. Labs. report ANL-5990 Rev.

Davidon, W. C. (1975), "Optimally Conditioned Optimization Algorithms Without Line Searches", Math. Prog. 9, pp. 1-30.

Dennis, J.E. (1972), "On Some Methods Based on Broyden's Secant Approximation to the Hessian", in Numerical Methods for Nonlinear Optimization, edited by F. A. Lootsma, Academic Press, London.

Dennis, J. E. (1975), private communication.

Dennis, J. E.; & Mei, H. H. W. (1975), "An Unconstrained Optimization Algorithm Which Uses Function and Gradient Values", technical report TR75-246 of the Computer Science Dept., Cornell University, Ithaca, New York 14850.

Dennis, J. E.; & Moré, J. J. (1974), "Quasi-Newton Methods, Motivation and Theory", technical report TR74-217 of the Computer Science Dept., Cornell University, Ithaca, New York 14850.

Powell, M. J. D. (1970a), "A New Algorithm for Unconstrained Optimization", in Nonlinear Programming, edited by J. B. Rosen, O. L. Mangasarian, K. Ritter; Academic Press, New York.

Powell, M. J. D. (1970b), "A FORTRAN Subroutine for Unconstrained Minimization, Requiring First Derivatives of the Objective Function", A.E.R.E., Harwell, report R.6469.

Powell, M. J. D. (1970c), "A New Algorithm for Unconstrained Optimization", in Nonlinear Programming, edited by J. B. Rosen, O. L. Mangasarian & K. Ritter; Academis Press, New York.

NATIONAL BUREAU OF ECONOMIC RESEARCH, INC.
575 TECHNOLOGY SQUARE, CAMBRIDGE, MASSACHUSETTS 02139
(617) 661-8788
NEW YORK
NEW HAVEN
CAMBRIDGE

October, 1976

A couple of comments are appropriate to my Working Paper on "Representing Symmetric Rank 2 Updates". First, §6 may overemphasize the application of earlier results to updating matrices A which must always be positive definite and which hence might better be kept in the factored form $A = L^T L$ or $A = L^T D L$, where L is a triangular matrix and D is a diagonal matrix with positive diagonal elements. Second, (33) should be replaced in practice by:

$$\text{If } \xi \geq 0 \text{ then } \begin{cases} x = \left(\dfrac{\sigma}{\xi + \sqrt{\xi^2 - \sigma\tau}}\right)s + t \\ y = \left(\dfrac{\xi + \sqrt{\xi^2 - \sigma\tau}}{2}\right)s + t \end{cases}$$

$$\text{else } (\xi < 0) \begin{cases} x = s + \left(\dfrac{\tau}{\xi - \sqrt{\xi^2 - \sigma\tau}}\right)t \\ y = \left(\dfrac{\sigma}{2}\right)s + \left(\dfrac{\xi - \sqrt{\xi^2 - \sigma\tau}}{2}\right)t. \end{cases}$$

David M Gay