MATHEMATICS OF COMPUTATION, VOLUME 24, NUMBER 111, JULY, 1970

Optimal Conditioning of Quasi-Newton Methods*

By D. F. Shanno and P. C. Kettler

Abstract. Quasi-Newton methods accelerate gradient methods for minimizing a function by approximating the inverse Hessian matrix of the function. Several papers in recent literature have dealt with the generation of classes of approximating matrices as a function of a scalar parameter. This paper derives necessary and sufficient conditions on the range of one such parameter to guarantee stability of the method. It further shows that the parameter effects only the length, not the direction, of the search vector at each step, and uses this result to derive several computational algorithms. The algorithms are evaluated on a series of test problems.

I. Introduction. Quasi-Newton methods for minimizing a function $f(x)$, x an n-vector, are iterative accelerated gradient methods which use past computational history to approximate the inverse of the Hessian matrix of the function. This is accomplished by selecting an initial approximation $H^{(0)}$ to the inverse Hessian, as well as an initial estimate $x^{(0)}$ to the minimum of $f(x)$, and then finding at each step $\alpha^{(K)}$, the scalar parameter which minimizes $f(x^{(K)} - \alpha^{(K)}H^{(K)}g^{(K)})$. Letting $s^{(K)} = -H^{(K)}g^{(K)}$, $\sigma^{(K)} = \alpha^{(K)}s^{(K)}$, $x^{(K+1)} = x^{(K)} + \sigma^{(K)}$, $g(x) = \nabla f(x)$, $g^{(K)} = g(x^{(K)})$, and $y^{(K)} = g^{(K+1)} - g^{(K)}$, $H^{(K)}$ is then updated by

$$H^{(K+1)} = H^{(K)} + D^{(K)}, \tag{1}$$

where $D^{(K)}$ is the correction matrix. As has been shown by Fletcher and Powell [3], n-step convergence to the minimum of a positive definite quadratic form is achieved when $D^{(K)}$ satisfies

$$D^{(K)}y^{(K)} = \sigma^{(K)} - H^{(K)}y^{(K)}. \tag{2}$$

In a previous paper [6], a class of correction matrices $D^{(K)}$ satisfying (2) was generated by a scalar parameter to be

$$D^{(K)} = t\,\frac{\sigma^{(K)}\sigma^{(K)\prime}}{\sigma^{(K)\prime}y^{(K)}} + \frac{((1-t)\sigma^{(K)} - H^{(K)}y^{(K)})((1-t)\sigma^{(K)} - H^{(K)}y^{(K)})'}{((1-t)\sigma^{(K)} - H^{(K)}y^{(K)})'y^{(K)}}. \tag{3}$$

It was then shown in [6] that if $t > (\alpha^{(K)} - 1)/\alpha^{(K)}$, positive definiteness of $H^{(K)}$ implies positive definiteness for $H^{(K+1)}$. Further, it was shown that the smallest eigenvalue of $H^{(K+1)}$ was a nondecreasing function of increasing t, and hence that the condition of $H^{(K)}$ improved as $t \to \infty$.

Section II of this paper will show that $s^{(K+1)} = -H^{(K+1)}g^{(K+1)}$ can be represented as

$$s^{(K+1)} = \phi^{(K)}(t)r^{(K)} \tag{4}$$

Received August 28, 1969, revised January 22, 1970.

AMS Subject Classifications. Primary 6530; Secondary 6550, 9058.

Key Words and Phrases. Function minimization, quasi-Newton methods, variable metric methods, gradient search, steepest descent methods, stability of search methods, conditioning of search method, Hessian matrix inverse approximations, quadratic convergence.

* This work was sponsored in part by a grant from the General Electric Foundation to the University of Chicago.

where $\phi^{(K)}(t)$ is a scalar function of t and $r^{(K)}$ a vector independent of t. Further, it will be shown that $\phi^{(K)}(t)$ assumes all values from $(-\infty, \infty)$, and that for all t such that $\phi^{(K)}(t) > 0$, $H^{(K+1)}$ is positive definite.

The chief significance of (4) lies in the fact that if one were able to choose t properly, the search for $\alpha^{(K)}$ at each step could be eliminated. A number of choices for t were tried with the aim of minimizing the total number of function evaluations necessary to minimize f. Section III contains the results of these trials, together with the reasons for the various choices, and an accelerated means for determining $\alpha^{(K)}$ when the t chosen is not optimal.

Finally, some numerical accuracy difficulties which were encountered in an attempt to choose optimal t will be documented and their significance discussed.

II. Representation of $H^{(K+1)}g^{(K+1)}$ as a Function of t. An exact representation of $H^{(K+1)}$ as a function of t is derived by combining (1) with (3) to yield

$$\begin{aligned} H^{(K+1)} &= H^{(K)} + t\,\frac{\sigma^{(K)}\sigma^{(K)\prime}}{\sigma^{(K)\prime}y^{(K)}} \\ &\quad + \frac{((1-t)\sigma^{(K)} - H^{(K)}y^{(K)})((1-t)\sigma^{(K)} - H^{(K)}y^{(K)})'}{((1-t)\sigma^{(K)} - H^{(K)}y^{(K)})'y^{(K)}}. \end{aligned} \tag{5}$$

We note, as in [6], that a necessary condition for $\alpha^{(K)}$ to minimize $f(x)$ along $f(x^{(K)} + \alpha^{(K)}s^{(K)})$ is that

$$df/d\alpha^{(K)} = g^{(K+1)\prime}s^{(K)} = g^{(K+1)\prime}\sigma^{(K)} = 0. \tag{6}$$

It is now possible to use (5) and (6) to show:

THEOREM 1. *Let* $a = g^{(K)\prime}H^{(K)}g^{(K)}$, $b = g^{(K+1)\prime}H^{(K)}g^{(K+1)}$. *Then*

$$H^{(K+1)}g^{(K+1)} = \phi^{(K)}(t)(aH^{(K)}g^{(K+1)} + bH^{(K)}g^{(K)}), \tag{7}$$

where

$$\phi^{(K)}(t) = \frac{(\alpha^{(K)}t - \alpha^{(K)} + 1)}{(\alpha^{(K)}t - \alpha^{(K)} + 1)a + b}. \tag{8}$$

Proof. Since $y^{(K)} = g^{(K+1)} - g^{(K)}$, (6) and (5) yield

$$\begin{aligned} H^{(K+1)}g^{(K+1)} &= H^{(K)}g^{(K+1)} \\ &\quad - \frac{g^{(K+1)}H^{(K)}g^{(K+1)}}{((1-t)\sigma^{(K)} - H^{(K)}y^{(K)})'y^{(K)}}((1-t)\sigma^{(K)} - H^{(K)}y^{(K)}). \end{aligned} \tag{9}$$

Now by the definitions of $\sigma^{(K)}$ and $H^{(K)}y^{(K)}$ and (6) we get

$$\begin{aligned} &((1-t)\sigma^{(K)} - H^{(K)}y^{(K)})'y^{(K)} \\ &\quad = (-1 + \alpha^{(K)} - \alpha^{(K)}t)g^{(K)\prime}H^{(K)}g^{(K)} - g^{(K+1)\prime}H^{(K)}g^{(K+1)}. \end{aligned} \tag{10}$$

Substituting (10) for the denominator in (9), cross-multiplying and collecting terms yields

$$\begin{aligned} H^{(K+1)}g^{(K+1)} &= \frac{-(\alpha^{(K)}t - \alpha^{(K)} + 1)aH^{(K)}g^{(K+1)} - (\alpha^{(K)}t - \alpha^{(K)} + 1)bH^{(K)}g^{(K)}}{-(\alpha^{(K)}t - \alpha^{(K)} + 1)a - b} \\ &= \frac{(\alpha^{(K)}t - \alpha^{(K)} + 1)}{(\alpha^{(K)}t - \alpha^{(K)} + 1)a + b}(aH^{(K)}g^{(K+1)} + bH^{(K)}g^{(K)}). \quad \text{Q.E.D.} \end{aligned} \tag{11}$$

We wish to show that $\phi^{(K)}(t)$ has the properties mentioned previously.

Note that since $H^{(K)}$ is assumed to be positive definite, $g^{(K)\prime}H^{(K)}g^{(K)}$ and $g^{(K+1)\prime}H^{(K)}g^{(K+1)} > 0$ unless either $g^{(K)}$ or $g^{(K+1)} = 0$, at which point the algorithm is terminated. Also, since $H^{(K)}$ is positive definite, $\alpha^{(K)}$ is positive. Hence

$$\frac{d\phi^{(K)}(t)}{dt} = \frac{\alpha^{(K)}b}{((\alpha^{(K)}t - \alpha^{(K)} + 1)a + b)^2} > 0. \tag{12}$$

and solving for zeroes of the numerator and denominator of $\phi^{(K)}(t)$ yields

$$\phi^{(K)}(t) > 0 \quad \text{for} \quad t > \frac{\alpha^{(K)} - 1}{\alpha^{(K)}} \quad \text{or} \quad t < \frac{\alpha^{(K)} - 1}{\alpha^{(K)}} - \frac{b}{\alpha^{(K)}a}. \tag{13}$$

In order to show that $\phi^{(K)}(t)$ can assume any value on $(-\infty, \infty)$, simply solve the equation $\phi^{(K)}(t) = s$ for any s.

We now prove the assertion that for $\phi^{(K)}(t) > 0$, $H^{(K)}$ positive definite implies $H^{(K+1)}$ positive definite. For this we first require the following lemma.

LEMMA 1. *$H^{(K)}$ positive definite implies $g^{(K+1)\prime}H^{(K+1)}g^{(K+1)} > 0$ if and only if $\phi^{(K)}(t) > 0$.*

Proof. From (2) and (3), $H^{(K+1)}y^{(K)} = \sigma^{(K)}$, so by (6),

$$g^{(K+1)\prime}H^{(K+1)}g^{(K+1)} = g^{(K+1)\prime}H^{(K+1)}g^{(K)}. \tag{14}$$

Applying (14) to (11) yields

$$g^{(K+1)\prime}H^{(K+1)}g^{(K+1)} = \phi^{(K)}(t)ab. \tag{15}$$

Now since $H^{(K)}$ is positive definite, a and b are > 0, so the lemma is proved.

THEOREM 2. *If $H^{(K)}$ is positive definite $H^{(K+1)}$ is positive definite if and only if $\phi^{(K)}(t) > 0$.*

Proof. Since $H^{(K)}$ is positive definite, any set of n vectors which are mutually $H^{(K)}$ orthogonal span E^n. Since $g^{(K)} \neq 0$, $g^{(K+1)} \neq 0$, and $g^{(K)\prime}H^{(K)}g^{(K+1)} = 0$, let $g^{(K)}$, $g^{(K+1)}$, and any $n - 2$ vectors $z_1, \cdots, z_{n-2}$ which are mutually $H^{(K)}$ orthogonal and which satisfy $z_i'H^{(K)}g^{(K)} = 0$, $z_i'H^{(K)}g^{(K+1)} = 0$, be a basis for E^n. Now let ξ be any arbitrary vector. We wish to consider $\xi H^{(K+1)}\xi$. Since $z_1, \cdots, z_{n-2}, g^{(K)}, g^{(K+1)}$ are a basis for E^n, we can write ξ as

$$\xi = \sum_{i=1}^{n-2} a_i z_i + a_{n-1}g^{(K)} + a_n g^{(K+1)}. \tag{16}$$

Now substitution into (5) shows that the $H^{(K)}$ orthogonality of the z's to each other and to $g^{(K)}, g^{(K+1)}$ guarantees that $z_i'H^{(K+1)}z_j = 0$ for $i \neq j$ and that $z_i'H^{(K+1)}g^{(K)} = 0$, $z_i'H^{(K+1)}g^{(K+1)} = 0$. Hence by (5) and (16),

$$\xi'H^{(K+1)}\xi = \sum_{i=1}^{n-2} a_i^2 z_i'H^{(K)}z_i + a_{n-1}^2 g^{(K)\prime}H^{(K+1)}g^{(K)} + 2a_{n-1}a_n g^{(K)\prime}H^{(K+1)}g^{(K+1)}$$
$$+ a_n^2 g^{(K+1)\prime}H^{(K+1)}g^{(K+1)},$$

and from (14),

$$\xi'H^{(K+1)}\xi = \sum_{i=1}^{n-2} a_i^2 z_n'H^{(K)}z_i + a_{n-1}^2 g^{(K)\prime}H^{(K+1)}g^{(K)} + (2a_{n-1}a_n + a_n^2)g^{(K+1)\prime}H^{(K+1)}g^{(K+1)}. \tag{17}$$

We now use again the fact that $H^{(K+1)}y^{(K)} = \sigma^{(K)}$ to show

$$g^{(K)\prime}H^{(K+1)}g^{(K)} = \alpha^{(K)}g^{(K)\prime}H^{(K)}g^{(K)} + g^{(K+1)\prime}H^{(K+1)}g^{(K+1)}. \tag{18}$$

Hence (17) becomes

$$\begin{aligned}\xi' H^{(K+1)}\xi = \sum_{i=1}^{n-2} a_i^2 z_i' H^{(K)} z_i + a_{n-1}^2 \alpha^{(K)} g^{(K)\prime} H^{(K)} g^{(K)} \\ + (a_{n-1} + a_n)^2 g^{(K+1)\prime} H^{(K+1)} g^{(K+1)}.\end{aligned} \tag{19}$$

Applying Lemma 1 to (19) gives the desired result.

Thus for any t in the range defined by (13), $H^{(K+1)}$ is positive definite. Hence, choice of t in this range maintains stability while scaling $s^{(K+1)}$ to any length desired. The following section deals with some choices of t in this range.

III. Choosing the Parameter *t*. Section II shows that $s^{(K+1)}$ can be written as

$$\begin{aligned} s^{(K+1)} &= -H^{(K+1)}g^{(K+1)} = -\phi^{(K)}(t)r^{(K)}, \quad \text{where} \\ r^{(K)} &= g^{(K)\prime}H^{(K)}g^{(K)}H^{(K)}g^{(K+1)} + g^{(K+1)\prime}H^{(K)}g^{(K+1)}H^{(K)}g^{(K)}. \end{aligned} \tag{20}$$

Having chosen a t, and hence $s^{(K+1)}$, we are then faced with the problem of determining $\alpha^{(K+1)}$. In general, the closer $\alpha^{(K+1)}$ is to 1, the fewer functional evaluations are necessary to determine the optimal value of $\alpha^{(K+1)}$. But since $\alpha^{(K+1)}$ is determined by the length of $s^{(K+1)}$, $\alpha^{(K+1)}$ is in fact a function of t. Thus the proper choice of t would yield $\alpha^{(K+1)} = 1$ at each step.

The problem here is attempting to determine the magnitude of the step-size to the minimum along $s^{(K+1)}$. Since f is assumed to be a nonlinear function, no analytic expression for the step-size can generally be obtained. The best which can be achieved is an estimate to the parameter t, and to this end several algorithms are tried.

The first algorithm tried is the algorithm for $t = \infty$ developed in [6]. The rationale for this in view of the developments of this paper can be derived from the following argument:

Expand $f(x)$ in a Taylor series about $x^{(K)}$ to yield

$$\begin{aligned} f(x^{(K+1)}) &= f(x^{(K)}) + (x^{(K+1)} - x^{(K)})'g^{(K)} \\ &\quad + \tfrac{1}{2}(x^{(K+1)} - x^{(K)})'T^{(K)}(x^{(K+1)} - x^{(K)}) \\ &= f(x^{(K)}) + \sigma^{(K)\prime}g^{(K)} + \tfrac{1}{2}\sigma^{(K)\prime}T^{(K)}\sigma^{(K)} \\ &= f(x^{(K)}) + \alpha^{(K)}s^{(K)\prime}g^{(K)} + \tfrac{1}{2}\alpha^{(K)2}s^{(K)\prime}T^{(K)}s^{(K)}. \end{aligned} \tag{21}$$

Differentiating (21) with respect to $\alpha^{(K)}$ yields

$$s^{(K)\prime}g^{(K)} + \alpha^{(K)}s^{(K)\prime}T^{(K)}s^{(K)} = 0. \tag{22}$$

or

$$\alpha^{(K)} = \frac{g^{(K)\prime}H^{(K)}g^{(K)}}{g^{(K)\prime}H^{(K)}T^{(K)}H^{(K)}g^{(K)}}. \tag{23}$$

Now if $H^{(K)}g^{(K)} = T^{(K)-1}g^{(K)}$, then $\alpha^{(K)} = 1$. Hence we want to choose $H^{(K)}g^{(K)}$ as close to $T^{(K)-1}g^{(K)}$ as possible. Fletcher and Powell have shown that as $x^{(K)}$ approaches a region in which f is essentially a quadratic function, $H^{(K)}$ will tend to $T^{(K)-1}$. In general, the rationale for choosing $t = \infty$ derives from the fact that $g^{(K+1)}$ and $g^{(K)}$ are $H^{(K)}$ orthogonal, hence the step-sizes in these orthogonal directions have no discernible relationship to each other. However, computational experience may have generated a reasonable step-size in the direction of $H^{(K)}g^{(K+1)}$. Thus we wish to keep the eigenvalue of $H^{(K+1)}$ in the direction of $g^{(K+1)}$ as close to that of $H^{(K)}$ in the direction of $g^{(K+1)}$ as possible.

As in Lemma 1,

$$\begin{aligned} g^{(K+1)'}H^{(K+1)}g^{(K+1)} &= g^{(K)'}H^{(K+1)}g^{(K+1)} \\ &= \frac{(\alpha^{(K)}t - \alpha^{(K)} + 1)g^{(K)'}H^{(K)}g^{(K)}g^{(K+1)'}H^{(K)}g^{(K+1)}}{(\alpha^{(K)}t - \alpha^{(K)} + 1)g^{(K)'}H^{(K)}g^{(K)} + g^{(K+1)'}H^{(K)}g^{(K+1)}} \end{aligned} \tag{24}$$

and hence

$$\lim_{t\to\infty} g^{(K+1)'}H^{(K+1)}g^{(K+1)} = g^{(K+1)'}H^{(K)}g^{(K+1)}, \tag{25}$$

yielding precisely the desired result.

Another, and perhaps more important, property of $t = \infty$ arises from the problem of minimizing the condition number of the matrix $H^{(K+1)}$. This problem is alluded to in [6]. Here the desire is to minimize $\mu(A) = ||A||\ ||A^{-1}||$. Letting $||\cdot||$ be the spectral norm, $\mu(A) = \lambda_n/\lambda_1$, where λ_n and λ_1 are the largest and smallest eigenvalues of A respectively.

To attack this problem for $H^{(K+1)}$, we see from (19) and (15) that for any vector ξ',

$$\begin{aligned} \xi' H^{(K+1)}\xi &= \sum_{i=1}^{n-2} a_i^2 z_i' H^{(K)} z_i + a_{n-1}^2\alpha^{(K)}g^{(K)'}H^{(K)}g^{(K)} \\ &\quad + (a_{n-1} + a_n)^2\psi^{(K)}(t)g^{(K+1)'}H^{(K)}g^{(K+1)}, \end{aligned} \tag{26}$$

where $\psi^{(K)}(t) = a\phi^{(K)}(t)$. As this is true for any vector ξ, it is true for $r_n(t)$ and $r_1(t)$, where $r_n(t)$ and $r_1(t)$ are the eigenvectors corresponding to λ_n and λ_1 and normalized so that $||r_n(t)|| = ||r_1(t)|| = 1$. The quantity which we wish to minimize is then

$$\mu(H^{(K+1)}) = \frac{r_n(t)'H^{(K+1)}r_n(t)}{r_1(t)'H^{(K+1)}r_1(t)}. \tag{27}$$

Now by (26) and (27), certainly as $\psi^{(K)}(t) \to \infty$, $r_n(t) \to g^{(K+1)}/||g^{(K+1)}||$ and $\lambda_n \to \infty$. Also, in this case, the component of $r_1(t)$ in the direction of $g^{(K+1)} \to 0$, and λ_1 remains finite. Hence $\mu(H^{(K+1)}) \to \infty$ as $\psi^{(K)}(t) \to \infty$. Also, as $\psi^{(K)}(t) \to 0$, $r_1(t) \to g^{(K+1)}/||g^{(K+1)}||$ and $\lambda_1 \to 0$. Again, here the component of $r_n(t)$ in the direction of $g^{(K+1)} \to 0$, and $\lambda_n > 0$. Here again $\mu(H^{(K+1)}) \to \infty$.

Thus we wish to keep $\psi^{(K)}(t)$ bounded away from 0 and ∞. Now by (8),

$$\psi^{(K)}t = a\phi^{(K)}(t) = \frac{(\alpha^{(K)}t - \alpha^{(K)} + 1)a}{(\alpha^{(K)}t - \alpha + 1)a + b}. \tag{28}$$

Differentiation of (28) shows that $d\psi^{(K)}(t)/dt = 0$ for $t = \infty$, and that this is an inflection point of $\psi^{(K)}(t)$.

As $t = \infty$ corresponds to $\psi^{(K)}(t) = 1$, this clearly avoids the problem of $\psi^{(K)}(t) \to 0$ or $\psi^{(K)}(t) \to \infty$. Further, since $t = \infty$ is an inflection point of $\psi^{(K)}(t)$, it would appear

TABLE I

		$t = \infty$		$t = \alpha^{(k)}$		$t = \frac{2\alpha^{(K)} - 1}{\alpha^{(K)}}$		Contracting Norm		Constant Norm	
Function	Initial Est.	Iter.	Eval.	Iter.	Eval.	Iter.	Eval.	Iter.	Eval.	Iter.	Eval.
Sum of Two Exponentials	(5, 0)	9	47	8	46	9	53	10	62	11	54
Sum of Two Exponentials	(0, 0)	10	36	10	37	10	38	11	38	10	38
Sum of Two Exponentials	(0, 20)	9	27	9	28	10	28	9	31	9	29
Sum of Two Exponentials	(2.5, 10)	6	19	6	18	6	19	6	31	6	19
Sum of Two Exponentials	(5, 20)	8	31	8	33	8	36	9	32	8	32
Rosenbrock	(1, −1.2)	13	45	13	45	14	54	19	104	13	53
Rosenbrock	(2, −2)	15	59	17	67	18	67	20	88	18	64
Rosenbrock	(−3.635, 5.621)	22	88	28	112	27	113	25	116	26	99
Rosenbrock	(6.39, −.221)	19	61	19	79	16	57	21	84	13	52
Rosenbrock	(1.489, −2.547)	19	71	19	71	16	73	19	73	19	70
Wood	(−3, −1, −3, −1)	21	97	20	83	21	81	25	129	19	70
Weibull	(5, .15, 2.5)	19	99	19	98	20	90	31	239	19	96
Weibull	(250, .3, 5)	33	111	38	137	42	187	(a)		38	158
Weibull	(100, 3, 12.5)	24	138	25	130	25	139	(a)		25	132

(a) Exceeded allotted computer time without converging that the deterioration of the condition number μ is symmetric about $\psi^{(K)}(t) = 1$, and hence $t = \infty$ is in some sense an optimal choice to minimize $\mu(H^{(K+1)})$ at each step.

Another method which suggested itself takes into account the result [4], that as $x^{(K)}$ approaches the minimum, Newton's method converges quadratically, i.e. if $\hat{x}$ is the minimum

$$(29) \qquad ||x^{(K+1)} - \hat{x}|| \leqq M\, ||x^{(K)} - \hat{x}||^2.$$

This suggested using $||\sigma^{(K)}||$ as an estimate of $||x^{(K)} - x||$ and t was chosen so that $||s^{(K+1)}|| = ||\sigma^{(K)}||^2$. This proved computationally unsatisfactory (see Section IV), so was modified by choosing t so that $||s^{(K+1)}|| = ||\sigma^{(K)}||$.

Two other methods, both directly sensitive to $\alpha^{(K)}$, were also tried. The first was simply $t = \alpha^{(K)}$, which has no rational basis, but appears to be reasonably satisfactory. The last was $t = (2\alpha^{(K)} - 1)/\alpha^{(K)}$, which is obtained by the composite Fletcher-Powell scaling discussed in [6].

The results of testing all five methods on a series of test problems are discussed in Section IV. Two further points are necessary to round out this section.

First, in the two methods which chose t so that $||s^{(K+1)}|| = ||\sigma^{(K)}||^2$ and $||s^{(K+1)}|| = ||\sigma^{(K)}||$, the method proved numerically unstable for $t < (\alpha^{(K)} - 1)/\alpha^{(K)} - b/a\alpha^{(K)}$. This arises from the fact that it is computationally unfeasible to force $g^{(K+1)\prime}H^{(K)}g^{(K)} = 0$, but rather only $|g^{(K+1)\prime}H^{(K)}g^{(K)}| < \delta$, where δ is rather crude. This instability can be eliminated by refining δ, but numerical experiments with this showed that the number of function evaluations increases. Hence t was restricted to the range $t > (\alpha^{(K)} - 1)/\alpha^{(K)}$, and if $t < (\alpha^{(K)} - 1)/\alpha^{(K)} - b/a\alpha^{(K)}$ was indicated, $t = \infty$ was substituted.

Finally, the cubic quadrature technique devised by Davidon [2], was used to locate $\alpha^{(K)}$ at each step after two points were found at which $df/d\alpha^{(K)} < 0$ and $df/d\alpha^{(K)} > 0$. In an attempt to expedite the α search when no point had yet been found at which $df/d\alpha^{(K)} > 0$, rather than simply doubling $\alpha^{(K)}$ as suggested by Davidon, a new approximation was found as follows:

Assume $f(x) = \frac{1}{2}x'Ax + x'b$. Then by (23),

$$(30) \qquad \alpha^{(K)} = -s^{(K)\prime}g^{(K)}/(s^{(K)\prime}As^{(K)}).$$

Now

$$(31) \qquad \begin{aligned} f(x^{(K)} + s^{(K)}) - f(x^{(K)}) &= x^{(K)\prime}As^{(K)} + \tfrac{1}{2}s^{(K)\prime}As^{(K)} + b's^{(K)} \\ &= s^{(K)\prime}g^{(K)} + \tfrac{1}{2}s^{(K)\prime}As^{(K)}, \end{aligned}$$

hence

$$(32) \qquad s^{(K)\prime}As^{(K)} = 2(f(x^{(K)} + s^{(K)}) - f(s^{(K)}) - s^{(K)\prime}g^{(K)}).$$

(30) and (32) combine to yield

$$(33) \qquad \alpha^{(K)} = \frac{-s^{(K)\prime}g^{(K)}}{2(f(x^{(K)} + s^{(K)}) - f(x^{(K)}) - s^{(K)\prime}g^{(K)})}.$$

Thus if $df/d\alpha^{(K)} < 0$ for $\alpha^{(K)} = 1$ and $\alpha^{(K)}$ yielded by (33) satisfies $\alpha^{(K)} > 1$, this $\alpha^{(K)}$ is tried. Otherwise $\alpha^{(K)} = 2$ is tried. Comparison of the results in Section IV with the results in [6] verifies that this saves a fair number of function evaluations.

IV. Computational Results. The five methods for selecting t discussed in Section III were tested on the four test problems documented in [6]. As previous testing included the straight Fletcher-Powell and Barnes-Rosen [1], [5] techniques, they are not included here. Previous tests have shown both to be substantially inferior to four of the five methods tested here.

The results of the tests are summarized in Table 1. As in [6], Iter. designates the number of times $H^{(K)}$ is updated, and Eval. the number of function evaluations. The choice of t which yields $||s^{(K+1)}|| = ||\sigma^{(K)}||^2$ is designated as the contracting norm version, while the version which yields $||s^{(K+1)}|| = ||\sigma^{(K)}||$ is designated as the constant norm version.

It is clear from the table that the contracting norm version is markedly inferior to the other four versions, apparently because the quadratic convergence criterion does not begin to take effect until the last few iterations. The other four versions seem remarkably similar, with a slight edge going to the $t = \infty$ version. The reduction in total function evaluations for $t = \infty$ and $t = (2\alpha^{(K)} - 1)/\alpha^{(K)}$ from [6] is due to the improved $\alpha^{(K)}$ search documented in Section III. Note, however, that for the initial estimates (250, .3, 5) for the Weibull function, $t = (2\alpha^{(K)} - 1)/\alpha^{(K)}$ is inferior to this choice of t without the new algorithm for determining $\alpha^{(K)}$. This verifies that ill-conditioned problems are very sensitive to all phases of the computational technique.

In general, the results of this testing appear to justify using $t = \infty$ as a reasonable choice for t generally. It may, however, prove somewhat better to combine this algorithm with the constant norm algorithm, using $t = \infty$ until steps become small, then switching to the constant norm. Computational experience verifies that this may accelerate convergence to some degree.

The University of Chicago
Chicago, Illinois 60637

1. J. G. P. Barnes, "An algorithm for solving non-linear equations based on the secant method," *Comput. J.*, v. 8, 1965, pp. 66–72. MR **31** #5330.

2. W. C. Davidon, *Variable Metric Method for Minimization*, Argonne National Laboratory Report ANL-5990, November 1959.

3. R. Fletcher & M. J. D. Powell, "A rapidly convergent descent method for minimization," *Comput. J.*, v. 6, 1963/64, pp. 163–168. MR **27** #2096.

4. A. Ralston, *A First Course in Numerical Analysis*, McGraw-Hill, New York, 1965, p. 332. MR **32** #8479.

5. E. M. Rosen, "A review of quasi-Newton methods in nonlinear equation solving and unconstrained optimization," Nat. Conference of the ACM, *Proceedings of the Twenty-First Conference*, Thompson Book Co., Washington, D.C., 1966, pp. 37–41.

6. D. F. Shanno, *Conditioning of Quasi-Newton Methods for Function Minimization*, Center for Math. Studies in Business and Economics, Department of Economics and Graduate School of Business, University of Chicago, Report 6910, March 1969; *Math. Comp.*, v. 24, 1970, pp. 647–656.