Numer. Math. 18, 289–297 (1972)

Derivative Free Analogues of the Levenberg-Marquardt and Gauss Algorithms for Nonlinear Least Squares Approximation

KENNETH M. BROWN*
IBM Philadelphia Scientific Center and University of Minnesota

J. E. DENNIS, JR.**
Cornell University

Received August 7, 1970

Abstract. In this paper we give two derivative-free computational algorithms for nonlinear least squares approximation. The algorithms are finite difference analogues of the Levenberg-Marquardt and Gauss methods. Local convergence theorems for the algorithms are proven. In the special case when the residuals are zero at the minimum, we show that certain computationally simple choices of the parameters lead to quadratic convergence. Numerical examples are included.

1. Introduction

Among the techniques used for unconstrained optimization of a nonlinear objective function in the least squares sense, two of the most popular have been Gauss' method (also called the Gauss-Newton method) [1, 14] and the Levenberg-Marquardt algorithm [1, 11, 12, 14]. In this paper we give derivative free (finite difference) analogues of these methods and prove local convergence for the algorithms. We show how to select parameters in such a way that quadratic convergence is guaranteed for any function which is zero at the minimum. Results of computer experiments show how the methods compare to other currently used techniques.

2. Description of the Methods

Let $f_1, \ldots, f_M$ be real valued functions of N real unknowns. We consider the problem of finding an $x^* \in E^N$ which minimizes

$$\phi(x) = \|F(x)\|_2^2 = \sum_{i=1}^{M} |f_i(x)|^2.$$

($F(x)$ is often the residual vector obtained when approximating a data set by a nonlinear form.) The minimum, x^*, is contained among the zeros of $1/2\, \nabla\phi(x) = J(x)^T F(x)$, where $\nabla\phi(x)$ denotes the gradient of ϕ and $J(x)$ is the Jacobian matrix of F.

* On leave 1970–71 at Yale University

** The work of this author was supported in part by the National Science Foundation under Grant GJ-844.

Let x_0 be an initial approximation to a minimum point of ϕ. Generate a sequence of approximations to this minimum point by

$$x_{n+1} = x_n - [\mu_n I + J(x_n)^T J(x_n)]^{-1} J(x_n)^T F(x_n)$$

where μ_n is a sequence of nonnegative real constants. We shall refer to this as the Levenberg-Marquardt or L.M. iteration [1, 11, 12, 14].

The special case in which $\mu_n = 0$ for all n is called Gauss', or the Gauss-Newton, method [1, 14].

The L.M. method requires the evaluation of the Jacobian of F at each iteration point. Powell [17] replaces the L.M. iteration with a technique in which Broyden's single rank update [5] is used to approximate $J(x_n)$. We will use the more obvious approach of approximating $J(x_n)$ by the corresponding matrix of difference quotients.

Let h be a real number and let $\Delta F(x, h)$ denote the matrix whose i-th, j-th element is given by

$$f_i(x + hu_j) - f_i(x),$$

where u_j is the j-th unit vector. We consider the following finite difference analogue of the Levenberg-Marquardt algorithm (f.d.L.M.)

$$\begin{aligned} x_{n+1} &= x_n - [\mu_n I + h_n^{-2} \Delta F(x_n, h_n)^T \Delta F(x_n, h_n)]^{-1} h_n^{-1} \Delta F(x_n, h_n)^T F(x_n) \\ &= x_n - h_n [h_n^2 \mu_n I + \Delta F(x_n, h_n)^T \Delta F(x_n, h_n)]^{-1} \Delta F(x_n, h_n)^T F(x_n). \end{aligned}$$

The finite difference analogue of Gauss' method (f.d.G.) results when $\mu_n = 0$ for each n.

In the sequel we show that if μ_n and $|h_n|$ are $O(\|F(x_n)\|_2)$ then f.d.L.M. converges quadratically whenever $\|F(x_n)\|_2 \underset{n}{\to} 0$; moreover, we show that whenever $|h_n|$ is $O(\|F(x_n)\|_2)$ and $\|F(x_n)\|_2 \underset{n}{\to} 0$, f.d.G. is quadratically convergent.

We assume that $h^{-1} \Delta F(x, h) \to J(x)$ as $h \to 0$ so $0^{-1} \Delta F(x, 0) \equiv J(x)$ is a notational convenience.

3. Convergence Results

It is well known (see [12] and 14]) that the nonnegative sequence $\{\mu_n\}$ can be chosen so that $\phi(x_n)$ forms a monotone decreasing sequence. If we make some standard assumption, for example, that the level sets of ϕ are bounded and we are careful in our choice of $\{\mu_n\}$ then some subsequence of $\{x_n\}$ is convergent to an extreme point, say x^*, of ϕ. If the method is locally convergent, as soon as some term of the convergent subsequence gets sufficiently close to x^*, the entire sequence will converge to it.

Lemma 1. Let D_0 be an open convex set and let $K \geqq 0$. If, for all $x, y \in D_0$,

$$\|J(x) - J(y)\|_2 \leqq K \|x - y\|_2,$$

then the following inequalities hold for every $x, y \in D_0$:

i) $$\|J(x)^T - J(y)^T\|_2 \leqq K \|x - y\|_2$$

ii) there exist nonnegative constants C and C' such that

$$\|J(x) - J(y)\|_1 \leqq C \cdot K \|x - y\|_1$$

and

$$\|A\|_2 \leqq C'\|A\|_1$$

for any real rectangular matrix A.

iii) $$\|F(x)-F(y)-J(y)(x-y)\|_1 \leqq (CK/2)\|x-y\|_1^2.$$

Proof. A proof of (i) follows from the fact that $\|A\|_2=\|A^T\|_2$ for any rectangular matrix A. This in turn follows from page 54, [22].

Statement (ii) follows from a straightforward application of the equivalence of the ℓ_1 and ℓ_2 vector norms.

Statement (iii) can be found in [7].

The following result is inspired by and closely related to the work in [3].

Theorem 1. Let $J(x^*)^T F(x^*)=0$ for some $x^*\in D_0$ and let

$$\|J(x)-J(y)\|_2 \leqq K\|x-y\|_2 \quad \text{for} \quad x, y\in D_0.$$

If $K\|F(x^*)\|_2$ is a strict lower bound for the spectrum of $J(x^*)^T J(x^*)$ and $\{\mu_n\}$ is any bounded nonnegative sequence of real numbers, then there is an $\eta>0$ such that if every $|h_n|<\eta$ and $h_n \underset{n}{\to} 0$, the f.d.L.M. method converges locally to x^*.

Proof. For each $x\in D_0$, let $\lambda(x, h)$ denote the least eigenvalue of

$$h^{-2}\Delta F(x, h)^T \Delta F(x, h).$$

For $|h|$ sufficiently small and x in some smaller neighborhood (than D_0) of x^*, λ is a nonnegative jointly continuous function of x and h. Since by hypothesis

$$\lambda(x^*, 0) > K\|F(x^*)\|_2 \geqq 0,$$

we can find $\eta'>0$ and D_1, an open convex neighborhood of x^*, such that $\bar{D}_1 \subset D_0$ and for $x\in\bar{D}_1$, $|h|\leqq\eta'$, $\lambda(x, h)>\lambda'=K\|F(x^*)\|_2+(\lambda(x^*, 0)-K\|F(x^*)\|_2)/2$. Let B denote the uniform bound on

$$\|h^{-1}\Delta F(x, h)^T\|_2 \quad \text{for} \quad x\in\bar{D}_1, \quad h\in[-\eta', \eta'].$$

(B exists by continuity.) Now set

$$e_n \equiv \|x_n - x^*\|_2,$$

and

$$L_n \equiv \mu_n I + h_n^{-2}\Delta F(x_n, h_n)^T \Delta F(x_n, h_n)$$

and notice that the smallest eigenvalue of L_n is $\mu_n+\lambda(x_n, h_n) \geqq \mu_n+\lambda'>0$. Hence L_n^{-1} exists and its ℓ_2 norm is bounded by $(\mu_n+\lambda')^{-1}$; moreover x_{n+1} (obtained from the f.d.L.M. iteration) exists and, for some $\theta\in(0, 1)$ we can write:

$$\begin{aligned}
e_{n+1} &\equiv \|x_{n+1}-x^*\|_2 \\
&\leqq (\mu_n+\lambda')^{-1}\{\|L_n(x)-h_n^{-1}\Delta F(x_n, h_n)^T J(x_n+\theta(x^*-x_n))\|_2 \\
&\quad \cdot\|x_n-x^*\|_2+\|J(x^*)-h_n^{-1}\Delta F(x_n, h_n)\|_2\cdot\|F(x^*)\|_2\} \\
&\leqq (\mu_n+\lambda')^{-1}\{[\mu_n+B\|h_n^{-1}\Delta F(x_n, h_n)-J(x_n)\|_2 \\
&\quad +B\|J(x_n)-J(x_n+\theta(x^*-x_n))\|_2]\cdot e_n+[\|J(x^*)-J(x_n)\|_2 \\
&\quad +\|J(x_n)-h_n^{-1}\Delta F(x_n, h_n)\|_2]\|F(x^*)\|_2\}.
\end{aligned}$$

It is shown in [7] that

$$\|h_n^{-1} \Delta F(x_n, h_n) - J(x_n)\|_1 \leqq C K |h_n|/2.$$

Hence by Lemma 1

$$\begin{aligned} e_{n+1} \leqq & (\mu_n + \lambda')^{-1} (\mu_n + B K e_n + B C' C K |h_n|/2 + K \|F(x^*)\|_2) e_n \\ & + (\mu_n + \lambda')^{-1} C' C K \|F(x^*)\|_2 |h_n|/2. \end{aligned} \tag{1}$$

Now choose $r > 0$ and $\eta'' > 0$ so small that $N(x^*, r) \subset D_1$, $\eta'' < \eta'$ and

$$\lambda(x^*, 0) - K\|F(x^*)\|_2 > 2 B K r + B C' C K \eta''.$$

This insures that $\delta < 1$, where

$$\delta \equiv \sup_n (\mu_n + \lambda')^{-1} (\mu_n + B K r + B C' C K \eta''/2 + K \|F(x^*)\|_2).$$

Finally select $\eta < \eta''$ such that

$$\eta \leqq \frac{2(1-\delta)\lambda' r}{C' C K \|F(x^*)\|_2}, \quad \text{for} \quad \|F(x^*)\|_2 \neq 0,$$

or

$$\eta \leqq \frac{2(\lambda' - B K r)}{B C' C K}, \quad \text{if} \quad \|F(x^*)\|_2 = 0.$$

Now let us assume that for $n \geqq 0$, we have $e_n \leqq r$ and $|h_n| \leqq \eta$. From (1) we obtain

$$e_{n+1} \leqq \delta e_n + (1-\delta) r \leqq r.$$

This shows that the f.d.L.M. iteration is locally well-defined. The convergence of $\{x_n\}$ to x^*, as long as $h_n \to 0$, follows directly from the discussion on page 41 of Traub [21].

Remark 1. For computational convenience, if some estimate of $F(x^*)$ is known, the μ_n and h_n may be chosen to be $O(\|F(x_n) - F(x^*)\|_{any})$. Local convergence can be proven for this choice of the parameters; moreover, as we shall see later, this choice guarantees quadratic convergence whenever $F(x^*) = 0$.

Corollary 1. Under the hypotheses of the precious theorem, the L.M. iteration converges locally for any bounded nonnegative sequence $\{\mu_n\}$ and the convergence is of order at least one.

Proof. The L.M. iteration is the special case of the f.d.L.M. iteration with $h_n = 0$ for every n. Hence η can be taken to be zero and from the proof of the previous theorem, $e_{n+1} \leqq \delta e_n$ and so the convergence is at least linear.

If $F(x^*) = 0$, the situation simplifies considerably.

Theorem 2. Let F satisfy the differentiability assumptions of Theorem 1 in a neighborhood D_0 of x^*, a zero of F. If $\{\mu_n\}$ is a bounded nonnegative sequence, there exists an $\eta > 0$ such that if $|h_n| \leqq \eta$, f.d.L.M. converges locally. If, in addition, $|h_n|$ and μ_n are $O(\|F(x_n)\|_{any})$ the method is quadratically convergent.

Proof. The proof is the same as the proof of Theorem 1 until equation (1). Eq. (1) now becomes

$$e_{n+1} \leqq (\mu_n + \lambda')^{-1} (\mu_n + B K e_n + B C' C K |h_n|/2) e_n. \tag{1'}$$

The local convergence is trivial. To see that the convergence is quadratic, note that since there is a uniform upper bound on $\|J(x)\|_2$ for $\|x-x^*\|_2 \leqq r$,

$$\|F(x_n)\|_2 = \|F(x_n) - F(x^*)\|_2 \leqq \sup_x \|J(x)\|_2 \|x_n - x^*\|_2$$

where $x \in [x_n, x^*] \subset \bar{N}(x^*, r)$. Hence, using the above and the order equivalence of norms, $|h_n|$ and μ_n are $O(e_n)$ and so from (1'), $e_{n+1} = O(e_n^2)$.

Corollary 2. Under the hypotheses of Theorem 2, the L.M. iteration converges quadratically.

Remark 2. Theorems 1 and 2 and Corollaries 1 and 2 hold for the f.d.G. and Gauss methods respectively by merely setting $\mu_n = 0$ in each statement and proof.

4. Numerical Results

In some twenty numerical experiments run, the f.d.L.M. algorithm behaved almost identical to its analytic counterpart (the L.M. method). When convergence occurred, both algorithms converged at the same rate; when one of the algorithms diverged (or failed to converge in the prescribed number of iterations, 100) its analogue usually did likewise. The same behavior was observed for the f.d.G. algorithm and Gauss' method. These results answer in the affirmative sense one of the questions posed by Bard [1]. We give a portion of the computational investigations below and show a comparison with other methods based on the Box survey paper [4]. All computations were done in APL\360 on an IBM 360/50. (The APL\360 system performs all calculations in double precision.)

Although Theorem 2 dictates that h_n and μ_n be chosen as $O(\|F(x_n)\|_{any})$, these choices must be tempered in the actual implementation of the algorithm to prevent an absurd choice of h_n relative to x_n; e.g., suppose $\|x_0\|_\infty = 0.001$ but $\|F(x_0)\|_\infty = 1000$. In the computer program the algorithm was implemented as follows:

$$x_{n+1} = x_n - [\mu_n I + J(x_n, \boldsymbol{h}_n)^T J(x_n, \boldsymbol{h}_n)]^{-1} J(x_n, \boldsymbol{h}_n)^T F(x_n),$$

where the i-th, j-th element of $J(x_n, \boldsymbol{h}_n)$ is given by

$$\frac{f_i(x_n + \boldsymbol{h}_n^j u_j) - f_i(x_n)}{\boldsymbol{h}_n^j},$$

where again u_j denotes the j-th unit column vector and where μ_n and $\boldsymbol{h}_n$ were chosen according to the rules:

$$\mu_n = c \cdot \|F(x_n)\|_\infty,$$

$$c = \begin{cases} 10 & \text{whenever } \quad 10 \leqq \|F(x_n)\|_\infty \\ 1 & \text{whenever } \quad 1 < \|F(x_n)\|_\infty < 10 \\ 0.01 & \text{whenever } \quad \|F(x_n)\|_\infty \leqq 1; \end{cases}$$

$$\boldsymbol{h}_n^j = \min\{\|F(x_n)\|_{any}, \delta_n^j\},$$

where $\boldsymbol{h}_n$ and δ_n are vectors of length N and the superscript j denotes the j-th component of these vectors and where, moreover,

$$\delta_n^j = \begin{cases} 10^{-9} & \text{if } |x_n^j| < 10^{-6} \\ 0.001 \times |x_n^j| & \text{if } 10^{-6} \leqq |x_n^j|; \end{cases}$$

that is, the components, x_n^j, of x_n are incremented by (possibly) different amounts, $\boldsymbol{h}_n^j$, as opposed to using the uniform increment h_n of Theorem 2. Asymptotically, however, the definition of the $\boldsymbol{h}_n^j$ assures us that the conditions of the theorem will be met as the zero x^* is approached.

Remark 3. The selection procedure for c (and hence μ_n) is motivated by the fact that the method of descent [14] has global convergence properties not held by Gauss' method. When one is far away from the solution μ_n is chosen to be large in order to weight the descent part of the correction. As the iterates proceed toward the solution μ_n is decreased to weight the Gauss part of the correction. When we are far from the solution we are interested in stability; when we are close, stability is guaranteed by the local convergence results and we strive for rapidity of convergence.

Example.

$$\phi(x_1, x_2, x_3) = \sum_p [(e^{-x_1 p} - e^{-x_2 p}) - x_3(e^{-p} - e^{-10p})]^2 \tag{2}$$

where the summation is over the values $p = 0.1\ (0.1)\ 1.0$. This problem has a zero residual at $(1, 10, 1)$, $(10, 1, -1)$ and whenever $x_1 = x_2$ with $x_3 = 0$.

Case A. The minimum of $\phi(x_1, x_2, 1)$ is to be found; i.e. x_3 is held fixed at $x_3 = 1$.

Table A. *Number of function evaluations required to reduce ϕ to less than 10^{-5} (Case A: 2 dimensions)*

Method	Starting point x_0 and $\phi(x_0)$				
	(0, 0) 3.064	(0, 20) 2.087	(5, 0) 19.588	(5, 20) 1.808	(2.5, 10) 0.808
Swann [20]	78	57	231	53	6
Rosenbrock [18]	96	68	103	103	109
Nelder and Mead [13], Spendley, Hext and Himsworth [19]	41	43	39	39	40
Powell (1964) [15]	64	51	84	77	23
Fletcher and Reeves [9]	93	144	102	84	21
Davidon [6], Fletcher and Powell [8]	51	48	45	63	27
Powell (1965) [16]	38	22	46	29	12
Barnes [2]	27	15	F	36	24
Greenstadt I [10, p. 14]	73	47	F	49	12
Greenstadt I [10, p. 16]	46	53	F	54	17
Gauss	F	F	F	F	16
F.D.G.	F	F	F	F	16
L.M.	22	25	25	31	16
F.D.L.M.	22	25	25	31	16

Case B. The minimum of $\phi(x_1, x_2, x_3)$ as given in (2) is to be found.

Table B. *Number of function evaluations required to reduce ϕ to less than 10^{-5} (Case B: 3 dimensions)*

Method	Starting point x_0 and $\phi(x_0)$								
	(0, 20, 1) 2.087	(2.5, 10, 10) 275.881	(0, 0, 10) 306.401	(0, 10, 1) 1.885	(0, 10, 10) 213.673	(0, 10, 20) 1031.154	(0, 20, 0) 9.706	(0, 20, 10) 209.280	(0, 20, 20) 1021.655
Swann [20]	448	F	F	313	F	F	25	F	F
Rosenbrock [18]	347	281	200	198	292	350	273	460	246
Nelder and Mead [13], Spendley, Hext and Himsworth [19]	119	128	112	73	110	307	79	164	315
Powell (1964) [15]	78	F	F	56	F	F	19	F	F
Fletcher and Reeves [9]	564	112	104	56	92	92	344	608	188
Davidon [6], Fletcher and Powell [8]	92	68	144	24	148	140	96	148	140
Powell (1965) [16]	34	N.R.	N.R.	29	28	28	43	46	33
Barnes [2]	42	N.R.	N.R.	15	48	N.R.	37	51	59
Greenstadt I [10, p. 17]	58	31	46*	26	64	81	8*	84	92
Greenstadt I [10, p. 18]	54	45	32*	17	101	90	4*	103	106
Gauss	21	21(a)	F	17	17	17	21	21	21
F.D.G.	21	21(a)	13(b)	17	17	17	21	21	21
L.M.	41	33	41(c)	17	41	93	41	61	109
F.D.L.M.	41	33	41(c)	17	41	93	41	61	109

Notes concerning the tables.

1. The starting points used are given in the tables.

2. For the Gauss, f.d.G., L.M. and f.d.L.M. methods, the number of evaluations was arrived at by counting "one" each time F was evaluated and "N" each time $J(x)$ or $J(x, h)$ was evaluated.

3. Unless noted, convergence was to the extremum (1, 10) in Table A and to (1, 10, 1) in Table B. The exceptions are identified as:

* convergence to another solution,

(a) convergence to (10, 1, −1)

(b) convergence to (0.338..., 0.338..., 2.9E−13)

(c) convergence to (1.43..., 1.43..., 7.14E−13).

4. The notations "F" and "N.R." mean failed and not reported respectively.

Remark 4. The convergence obtained by the Gauss and f.d.G. methods to the extremum (10, 1, −1) is interesting because Box states [4, p. 69] that this optimum has never been found with any combination of the eight methods and all starting guesses which he tried.

Remark 5. In order to see just how close the f.d.L.M. algorithm approximates the L.M. method when one uses the criteria for selecting h_n and μ_n given above, we list below the error vectors $\|x_n - x^*\|_\infty$ produced by each method for the starting guess (0, 10, 20).

F.D.L.M.	L.M.
1.85810 2805 E 01	1.85810 2806 E 01
1.81716 4122 E 01	1.81716 4391 E 01
1.77806 1116 E 01	1.77806 2546 E 01
1.74218 6983 E 01	1.74219 1531 E 01
1.71019 2030 E 01	1.71019 9933 E 01
1.68023 4391 E 01	1.68024 2170 E 01
1.65076 3651 E 01	1.65076 9860 E 01
1.62143 6158 E 01	1.62144 0580 E 01
1.59219 0436 E 01	1.59219 3053 E 01
1.56301 4585 E 01	1.56301 5403 E 01
1.31709 9176 E 01	1.31708 9027 E 01
1.08275 9847 E 01	1.08273 8312 E 01
8.62326 5064 E 00	8.62295 3369 E 00
6.58882 6077 E 00	6.58844 2804 E 00
4.76392 5553 E 00	4.76350 3998 E 00
3.19750 9939 E 00	3.19709 3467 E 00
1.94335 0326 E 00	1.94299 0431 E 00
1.15671 7605 E 00	1.15668 7275 E 00
7.70363 2711 E−01	7.71008 0718 E−01
7.07093 2946 E−01	7.07718 6346 E−01
3.16189 0351 E−01	3.16772 2772 E−01
2.75553 9832 E−02	2.77907 5637 E−02
2.03121 1656 E−04	2.10392 4422 E−04
1.26532 6821 E−08	1.43025 5381 E−08

Note how nicely the asymptotic quadratic convergence predicted by Theorem 2 is borne out by this example.

Conclusions.

F.d.G. may be used instead of Gauss' method and f.d.L.M. instead of the L.M. method with little, if any, difference in the qualitative results. The L.M. and f.d.L.M. methods appear more stable than the Gauss and f.d.G. methods; however, when the latter do converge they seem to be much faster than the L.M. and f.d.L.M. algorithms.

References

1. Bard, Y.: Comparison of gradient methods for the solution of nonlinear parameter estimation problems. SIAM J. Numer. Anal. 7, 157–186 (1970).
2. Barnes, J. G. P.: An algorithm for solving nonlinear equations based on the secant method. The Computer Journal **8**, 66–72 (1965).
3. Boggs, P. T., Dennis, J. E., Jr.: Function minimization by descent along a difference approximation to the gradient. To appear.
4. Box, M. J.: A comparison of several current optimization methods, and the use of transformations in constrained problems. The Computer Journal **9**, 67–77 (1966).
5. Broyden, C. G.: A class of methods for solving nonlinear simultaneous equations. Math. Comp. **19**, 577–593 (1965).
6. Davidon, W. C.: Variable metric methods for minimization. A. E. C. Research and Development Report, ANL-5990. (1959) (Rev.).

7. Dennis, J. E., Jr.: On the convergence of Newtonlike methods. To appear in: Numerical methods for nonlinear algebraic equations, ed. P. Rabinowitz. London: Gordon and Breach 1970.
8. Fletcher, R., Powell, M. J. D.: A rapidly convergent descent method for minimization. The Computer Journal **6**, 163–168 (1963).
9. — Reeves, C. M.: Function minimization by conjugate gradients. The Computer Journal **7**, 149–154 (1964).
10. Greenstadt, J.: Variations of variable-metric methods. Math. Comp. **24**, 1–22 (1970).
11. Levenberg, K.: A method for the solution of certain nonlinear problems in least squares. Quart. Appl. Math. **2**, 164–168 (1944).
12. Marquardt, D. W.: An algorithm for least squares estimation of nonlinear parameters. SIAM J. Appl. Math. **11**, 431–441 (1963).
13. Nelder, J. A., Mead, R.: A simplex method for function minimization. The Computer Journal **7**, 308–313 (1965).
14. Ortega, J. M., Rheinboldt, W. C.: Iterative solution of nonlinear equations in several variables. Chapt. 8. New York: Academic Press 1970.
15. Powell, M. J. D.: An efficient method of finding the minimum of a function of several variables without calculating derivatives. The Computer Journal **7**, 155–162 (1964).
16. — A method for minimizing a sum of squares of nonlinear functions without calculating derivatives. The Computer Journal **7**, 303–307 (1965).
17. — A hybrid method for nonlinear equations. U.K.A.E.R.E. Technical Paper No. 364, January. 1969.
18. Rosenbrock, H. H.: An automatic method for finding the greatest or least value of a function. The Computer Journal **3**, 175–184 (1960).
19. Spendley, W., Hext, G. R., Himsworth, F. R.: Sequential applications of simplex designs in optimisation and evolutionary operation. Technometrics **4**, 441–461 (1962).
20. Swann, W. H.: Report on the development of a new direct searching method of optimisation. I.C.I. Ltd., Central Instrument Laboratory Research Note 64/3 (1964).
21. Traub, J. F.: Iterative methods for the solution of equations. Englewood Cliffs: Prentice-Hall 1964.
22. Wilkinson, J. H.: The algebraic eigenvalue problem, p. 54. Clarendon: Oxford 1965.

Kenneth M. Brown
Associate Prof. of Computer Science
Institute of Technology
University of Minnesota
Minneapolis, Minnesota 55455
USA

J. E. Dennis, Jr.
Associate Prof. of Computer Science
Cornell University
Ithaca, New York 14850
USA