Research Report Number 71 - 7 #8

Department of Computer Science

YALE UNIVERSITY

New Haven, Connecticut 06520

July 1971

ON THE SECOND ORDER CONVERGENCE OF BROWN'S DERIVATIVE - FREE METHOD FOR SOLVING SIMULTANEOUS NONLINEAR EQUATIONS

by

Kenneth M. Brown and J. E. Dennis, Jr.†

†Department of Computer Science
Cornell University
Ithaca, New York

ON THE SECOND ORDER CONVERGENCE OF BROWN'S DERIVATIVE - FREE METHOD FOR SOLVING SIMULTANEOUS NONLINEAR EQUATIONS

KENNETH M. BROWN[†] AND J. E. DENNIS, JR.[‡]

Abstract. Consider the problem of solving $F(x)=0$, a system of N real, e.g. transcendental, nonlinear equations in N real unknowns. Brown [2], [4] has given a derivative-free, Newton-like method for solving such a system. In [3] second order convergence is proved for the analytic form of this method (requiring explicit derivatives); however, the analytic form requires N^2 derivative and N function evaluations per iterative step, the same computational effort required by Newton's method (usual or derivative-free form). On the other hand, the derivative-free algorithm requires only $N^2/2 + 3N/2$ function evaluations per iterative step; moreover, there is a corresponding savings in storage -- from $N^2 + N$ locations to $N^2/2 + 3N/2$ locations. In this paper we give a constructive method for choosing the increment, h, in the first difference quotients which are used in the derivative-free method. Based upon this choice, we are able to prove second order convergence under hypotheses no more restrictive than those needed for Newton's method, namely: in a vicinity of a root, x^*, the Jacobian matrix of F has continuous entries and at x^* this matrix is nonsingular. Results of computational experiments are presented; the algorithm is particularly effective on Rosenbrock's function [14] and several nonlinear economics problems [16].

---

† Department of Computer Science, Yale University, New Haven, Connecticut 06520. Now at Department of Computer, Information, and Control Sciences, University of Minnesota, Minneapolis, Minnesota 55455. The work of this author was supported in part by the Office of Naval Research Contract NR 044-401. A portion of this work forms part of this author's thesis for the Ph.D. degree in Computer Science at Purdue University, written under the direction of Professor Samuel D. Conte.

‡ Department of Computer Science, Cornell University, Ithaca, New York 14850. The work of this author was supported in part by the National Science Foundation under Grant GJ-844.

CONVERGENCE OF BROWN'S METHOD FOR SIMULTANEOUS NONLINEAR EQUATIONS

1. Introduction.

In this paper we consider the system

$$(1.1)\qquad \begin{aligned} f_1(x_1, x_2, \ldots, x_N) &= 0 \\ f_2(x_1, x_2, \ldots, x_N) &= 0 \\ \ldots\ldots\ldots\ldots\ldots \\ f_N(x_1, x_2, \ldots, x_N) &= 0 , \end{aligned}$$

or in vector notation as

$$(1.2)\qquad F(x) = 0 .$$

Here we assume that each f_i is real-valued and continuously differentiable and that the x_i are real; typically we may have N real, transcendental equations in N real unknowns. The problem of solving such a system of nonlinear equations falls conveniently into three subproblems, namely a) proceeding from perhaps poor initial estimates in some regular fashion into a region of local convergence; b) using a rapidly convergent, computationally efficient and stable algorithm local to the root; and c) obtaining further solutions - different from those previously found - of the system (see Brown and Gearhart

[5]). We shall concentrate our efforts on b).

In this paper we analyze an algorithm proposed by Brown [1] for solving (1.1). The method is a Newton-like iteration based upon Gaussian elimination; it is derivative-free and has a built-in partial pivoting effect to help control rounding errors. Experimentally, the method has shown stability and rapid convergence in a vicinity of a solution; here we show how to guarantee second order convergence for the method by proper parameter selection. In §2 we describe the method algorithmically and establish the notation needed for the convergence analysis. The local, second order convergence of the method is proved in §3 under hypotheses no more restrictive than those needed for proving the convergence of Newton's method. In §4 we give computer results obtained by implementing a new FORTRAN program based on the method; comparisons are made with some of the better recent techniques as well as with the classical Newton's method.

2. Description of the Method.

Given a vector x^n which is an approximation to the solution x^* of (1.1), Newton's method is based on expanding the ~~entire~~ function vector f about the point x^n, retaining only the linear terms in this expansion as an approximation to f, equating this linear system to zero (since, if x^n is close to x^*, at points x in a neighborhood of x^n: $f(x) \sim f(x^*) = 0$), and taking the solution of the linear system to be the next iterate, x^{n+1}. The difficulty with this approach is that all equations are treated simultaneously; i.e., there is no attempt made to utilize information contained in the first few equations in later ones.

Brown [3] approached the problem by working with one equation at a time: expand the <u>first</u> function f_1 in a Taylor series expansion about x^n, truncate to linear terms and equate to zero; solve for that variable, say x_j, associated with the partial derivative of largest absolute value, say $\frac{\partial f_1(x^n)}{\partial x_j}$, as a function (necessarily a <u>linear</u> function) of the other $N-1$ variables. Now consider the second equation; in that equation replace the variable x_j with the linear function just obtained -- this replaces the second equation by an equation having just N-1 unknowns. Again expand f_2, truncate, set to zero and solve for one variable as a linear combination of the rest. Continue in this fashion eliminating one variable per equation until the <u>Nth</u> equation which will then involve just one unknown. Do a single (one dimensional) Newton step on this <u>Nth</u> equation and take the result to be one component of x^{n+1}; finally, back-solve the system of linear relationships built up to get the remaining N-1 components of x^{n+1}.

In addition to using the exact partial derivative expressions in the Taylor series expansions, Brown has shown how to approximate these partials by first difference quotients in such a way as to effect a savings of about one-half in the number of functions values needed per iteration and storage locations used relative to Newton's method. We shall show how to guarantee second order convergence for this derivative free method by a computationally simple choice of parameters.

The following notation will be used.

$$x \equiv (x_1, \ldots , x_N)^T .$$

$$x^n \equiv (x_1^n, \ldots , x_N^n)^T ,$$

where the superscript n denotes the nth iteration. Let ε be real and let u_j denote the jth unit column vector. For $x \in E^N$, denote by $T_j x$, the j-vector $(x_1, \ldots, x_j)^T$ obtained by truncating the last $N-j$ components of x. Now if g is a real function of k variables, let $\Delta g(T_k x,\varepsilon)$ stand for the k-dimensional row vector whose jth component is defined by the equation $\Delta g(T_k x,\varepsilon)\ T_k u_j = g(T_k(x + \varepsilon u_j)) - g(T_k(x))$. If $\varepsilon = 0$ then replace $\varepsilon^{-1}\Delta g(T_k x,\varepsilon)$ by $\nabla g(T_k x)$, the gradient vector. Another useful convention is that when f, g and h are real functions of k, $k+1$ and $k+2$ variables respectively $\langle T_k x, f, g, h\rangle^T$ will denote the vector of length $k+3$

$$(T_k x,\ f(T_k x),\ g(T_k x, f(T_k x)),\ h(T_k x, f(T_k x), g(T_k x, f(T_k x))))^T.$$

We will often use this notation with two, or more than three functions.

We now define the algorithm formally with (I) being the derivative free method and (II) denoting the form of the method which uses the exact derivative expressions.

CONVERGENCE OF BROWN'S METHOD FOR SIMULTANEOUS NONLINEAR EQUATIONS

(I) In order to obtain x^{n+1} from x^n and $\varepsilon^n \neq 0$, one proceeds as follows:

Define $g_1 \equiv f_1$ and form $\Delta g_1(x^n;\varepsilon^n)$. Without loss of generality, assume that $\| \Delta g_1(x^n;\varepsilon^n) \|_\infty = |\Delta g_1(x^n;\varepsilon^n)u_N|$ and define

$$b_N(T_{N-1}x) = x_N^n - (\Delta g_1(x^n;\varepsilon^n)u_N)^{-1}[T_{N-1}\Delta g_1(x^n;\varepsilon^n)T_{N-1}(x-x^n) + \varepsilon^n g_1(x^n)].$$

In general, given the functions $g_1, \ldots, g_k, b_N, b_{N-1}, \ldots, b_{N-k+1}$ define $g_{k+1}(T_{N-k}x) \equiv f_{k+1} < T_{N-k}x, b_{N-k+1}, \ldots, b_N >$, assume without loss of generality that

$$\| \Delta g_{k+1}(T_{N-k}x^n;\varepsilon^n) \|_\infty = |\Delta g_{k+1}(T_{N-k}x^n;\varepsilon^n)T_{N-k}\, u_{N-k}| \text{ and set}$$

$$b_{N-k}(T_{N-k-1}x) = x_{N-k}^n - (\Delta g_{k+1}(T_{N-k}x^n;\varepsilon^n)T_{N-k}\, u_{N-k})^{-1}$$

$$\cdot [T_{N-k-1}\Delta g_{k+1}(T_{N-k}x^n;\varepsilon^n)T_{N-k-1}(x - x^n)$$

$$+ \varepsilon^n g_{k+1}(T_{N-k}x^n)] .$$

Proceed by induction for $k = 1, 2, \ldots, N-1$ and notice that b_1 is a constant. Set

$$x_1^{n+1} = b_1 = x_1^n - (g_N(T_1(x^n + \varepsilon^n u_1) - g_N(T_1x^n))^{-1} \varepsilon^n g_N(T_1(x^n))$$

and $x^{n+1} = \langle x_1^{n+1}, b_2, \ldots, b_N \rangle^T$.

(II) In order to obtain x^{n+1} from x^n and $\varepsilon^n = 0$, one proceeds as follows:

Define $g_1 \equiv f_1$ and form $\nabla g_1(x^n)$. Without loss of generality assume that $\| \nabla g_1(x^n) \|_\infty = |\nabla g_1(x^n) u_N|$ and define

$$b_N(T_{N-1}x) = x_N^n - (\nabla g_1(x^n)u_N)^{-1}[\{T_{N-1}\nabla g_1(x^n)T_{N-1}(x-x^n) + g_1(x^n)\}].$$

Proceed by analogy with (I) and the above and set

$$x_1^{n+1} = b_1 = x_1^n - \left(\frac{dg_N}{dx_1}(T_1 x^n) \right)^{-1} g_N(T_1(x^n));$$

$$x^{n+1} = \langle x_1^{n+1}, b_2, \ldots, b_N \rangle^T.$$

We will show in the next section that (I) and (II) are consistent.

Remark 2.1. If F is a linear system, (I) and (II) reduce to transverse Gaussian elimination with partial (column) pivoting and, if the coefficient matrix is nonsingular, x^1 is the root regardless of the choice of x^0 and ε^0.

Remark 2.2. The reader will observe that whereas (II) requires the same number of evaluations and storage locations as Newton's method, (I) requires only

$$\sum_{k=2}^{N+1} k = \frac{N^2}{2} + \frac{3N}{2}$$

function evaluations per iterative step and $\left(\frac{N^2}{2} + \frac{3N}{2} \right)$ storage locations.

Revised

CONVERGENCE OF BROWN'S METHOD FOR SIMULTANEOUS NONLINEAR EQUATIONS

3. Convergence Results.

In this section we will prove that the method is well defined and has the same local convergence properties as Newton's method. We will work with two basic sets of assumptions on F.

The weak hypothesis. Let x^* be a zero of F, $R > 0$, and the Jacobian of F be continuous in $\bar{S}(x^*; R) \equiv \{x \in E^N: \|x - x^*\|_\infty \leq R\}$ and nonsingular at x^*.

The strong hypothesis. Let $K \geq 0$ and assume that, in addition to the weak hypothesis, F satisfies the property that

$$\|J(x) - J(x^*)\|_\infty \leq K\|x - x^*\|_\infty , \quad \text{for } \|x - x^*\|_\infty \leq R.$$

The goal of this section is the following theorem.

THEOREM If F satisfies the weak hypothesis then there exist positive numbers r, ε such that if $x^0 \in S(x^*;r)$ and $\{\varepsilon^n\}$ is bounded in modulus by ε, Brown's method (I) for nonlinear systems applied to F generates a sequence $\{x^n\}$ which converges to x^*. Moreover, if F satisfies the strong hypothesis and $\{\varepsilon^n\}$ is $O(\{|f_1(x^n)|\})$, then the convergence is at least second order.

Remark. The condition for quadratic convergence of Newton's method with difference quotient approximations in place of partial derivatives is $\{\varepsilon^n\} = O(\{\|F(x^n)\|_\infty\})$ [9], [15]. Clearly the requirement in the theorem is more stringent and the $\|F(x^n)\|$ requirement would suffice, as would any requirement which implies $O(\{\|x^n - x^*\|\}) = \{\varepsilon^n\}$. We use the $|f_1(x^n)|$ requirement because it is computationally convenient in the implementation of the method.

Proof. The proof consists of three basic parts. First we show that under the weak hypothesis there exist $R' > 0$ and $\varepsilon' > 0$ such that if $x^n \in S(x^*; R')$ and $|\varepsilon^n| \le \varepsilon'$, then one iteration of the method can be carried out and x^{n+1} exists. In the second part we prove that positive numbers $R'' \le R'$ and $\varepsilon'' \le \varepsilon'$ exist such that $x^0 \in S(x^*; R'')$ and $|\varepsilon^0| \le \varepsilon''$ imply that the iteration is a sequence of contractive mappings with uniformly bounded contractivity and hence converges. In the third part of the proof we show that under the strong hypothesis, the contractivity of each iteration function is bounded by a sequence uniformly proportional, in n, to the current error and thus the convergence is quadratic.

CONVERGENCE OF BROWN'S METHOD FOR SIMULTANEOUS NONLINEAR EQUATIONS

The first part of the proof, which is given immediately below, is very tedious and unenlightening. We recommend that the reader allow his intuition to convince him of the assertion and assume that the authors have insured its validity.

Part i. Clearly the g and b functions depend implicitly on the point x^n and the value ε^n as well as on the explicit variables indicated in section 2. Brown [3] has shown in detail that if x^n, ε^n are taken as $x^*, 0$, then the fact that $J(x^*)$ is nonsingular guarantees that $\nabla g_i(x^*) \neq 0$ for $i = 1, \dots, N$. This amounts to the well-known fact that Gaussian elimination with partial pivoting can be carried out on the nonsingular matrix $J(x^*)$.

Let us think of $g_1(x)$ as $g_1(x^n; \varepsilon^n)(x)$. Since g_1 is defined in terms of f_1 and x, it is entirely independent of the implicit variables x^n and ε^n. Hence $g_1(x^n; \varepsilon^n)(x)$ is continuous in $(x^n;\varepsilon^n)$ and satisfies the same differentiability assumptions as f_1 in the variable x. Here, of course, x^n and $x \in \bar{S}(x^*; R)$. Furthermore, $\nabla[g_1(x^n; \varepsilon^n)](x)u_i = \frac{\partial g_1}{\partial x_i}(x)$ is independent of, and hence continuous in, $(x^n; \varepsilon^n)$ as well as x.

Since from [3] some $\frac{\partial g_1(x^*)}{\partial x_j} \neq 0$, by continuity there is an $R_1 > 0$ and an $\varepsilon_1 > 0$ such that if $||x^n - x^*||_\infty < R_1$, $||x - x^*||_\infty < R_1$ and $|\varepsilon^n| < \varepsilon_1$ then $\frac{\partial[g_1(x^n;\varepsilon^n)]}{\partial x_j}(x) \neq 0$. Let us now consider $\Delta g_1(x^n;\varepsilon^n)$. By the mean value theorem, for each i between 1 and N and for some $\xi_i \in (x^n, x^n+\varepsilon^n u_i)$, $\Delta g_1(x^n;\varepsilon^n)u_i = \frac{\partial[g_1(x^n;\varepsilon^n)]}{\partial x_i}(\xi_i)$. Hence it is not hard to see that Δg_1 is continuous in $(x^n;\varepsilon^n)$. Furthermore, since $||\xi_j - x^*||_\infty \leq R_1$, at least this component of $\Delta g_1(x^n;\varepsilon^n)$ is not zero. It is consistent to assume $j = N$. Thus, for $||x^n - x^*||_\infty < R_1$ and $|\varepsilon^n| < \varepsilon_1$, $b_N(x^n;\varepsilon^n)(T_{N-1}x)$ is defined, continuous in $(x^n;\varepsilon^n)$ and affine in $T_{N-1}x$. By inspection, $b_N(x^*;0)(T_{N-1}x^*) = x_N^*$, and so by continuity, given any $\eta > 0$, there exist numbers $R_1'(\eta)$ and $\varepsilon_1'(\eta)$ no larger than R_1 and ε_1 respectively such that if $||x^n - x^*||_\infty < R_1'(\eta)$, $b_N(x^n;\varepsilon^n)(T_{N-1}x)$ is defined and $|b_N(x^n;\varepsilon^n)(T_{N-1}x) - x_N^*| < \eta$.

$g_2(x^n;\varepsilon^n)(T_{N-1}x)$ is formally defined as $f_2(\langle T_{N-1}x,b_N\rangle)$, which makes it clear that g_2 is continuous in $(x^n;\varepsilon^n)$ and $T_{N-1}x$ as long as $\|x^n - x^*\|_\infty < R_1'(R)$, $\|T_{N-1}(x - x^*)\|_\infty < R_1'(R)$ and $|\varepsilon^n| < \varepsilon_1'(R)$. If we formally differentiate $g_2(x^n;\varepsilon^n)$ with respect to the $N-1$ explicit variables we obtain an $N-1$ tuple whose ith coordinate is

$$\frac{\partial[g_2(x^n;\varepsilon^n)]}{\partial x_i}(T_{N-1}x) = \frac{\partial f_2}{\partial x_i}\langle T_{N-1}x,b_N\rangle + \frac{\partial f_2}{\partial x_N}\langle T_{N-1}x,b_N\rangle\frac{\partial b_N}{\partial x_i}(T_{N-1}x)$$

$$= \frac{\partial f_2}{\partial x_i}\langle T_{N-1}x,b_N\rangle + \frac{\partial f_2}{\partial x_N}\langle T_{N-1}x,b_N\rangle\frac{\Delta g_1(x^n;\varepsilon^n)u_i}{\Delta g_1(x^n;\varepsilon^n)u_N} .$$

Thus, as long as $\|x^n - x^*\|_\infty \leq R_1'(R)$, $\|T_{N-1}(x - x^*)\|_\infty \leq R_1'(R)$ and $|\varepsilon^n| \leq \varepsilon_1'(R)$, then $\nabla[g_2(x^n;\varepsilon^n)](T_{N-1}x)$ exists and is continuous in $(x^n;\varepsilon^n)$ and $(T_{N-1}x)$. Now, as in the previous step we use the result from [3] that for some $i \leq N-1$, $\frac{\partial[g_2(x^*;0)]}{\partial x_i}(T_{N-1}x^*) \neq 0$, together with continuity to insure the existence of $R_2 > 0$ and $\varepsilon_2 > 0$ such that $R_2 \leq R_1'(R)$ and $\varepsilon_2 \leq \varepsilon_1'(R)$ and, in fact, for $\|x^n-x^*\|_\infty < R_2$, $\|T_{N-1}(x - x^*)\|_\infty < R_2$ and $|\varepsilon^n| < \varepsilon_2$, it follows that $\Delta g_2(x^n;\varepsilon^n) \neq 0$. It is entirely consistent to assume that $i = N-1$. Thus, for such $(x^n;\varepsilon^n)$, $b_{N-1}(x^n;\varepsilon^n)$ is defined, continuous in $(x^n;\varepsilon^n)$ and affine in $T_{N-2}x$. Again, by inspection, $b_{N-1}(x^*;0)(T_{N-2}x^*) = x^*_{N-1}$ and so for any $\eta > 0$ there exist numbers $R'_2(\eta) \leq R_2$ and $\varepsilon'_2(\eta) \leq \varepsilon_2$ such that for $\|x^n - x^*\|_\infty < R'_2(\eta) > \|T_{N-2}(x - x^*)\|_\infty$ and $|\varepsilon^n| < \varepsilon'_2(\eta)$, $|b_{N-1}(x^n;\varepsilon^n)(T_{N-2}x) - x^*_{N-1}| < \eta$.

Choose $R' = R'_N(R'_{N-1}(\ldots(R'_1(R))\ldots))$, $\varepsilon' = \varepsilon'_N(R'_{N-1}(\ldots(R'_1(R))\ldots))$ and let $\|x^n - x^*\|_\infty < R'$, $|\varepsilon^n| < \varepsilon'$. Then, since the R'_i and ε'_i are chosen by the above process, all the g and b functions are defined in terms of the implicit variables $(x^n;\varepsilon^n)$. Furthermore, $|x_1^{n+1} - x_1^*| = |b_1 - x_1^*| < R'_{N-1}(R'_{N-2}(\ldots(R'_1(R))\ldots))$ so $\|x^n - x^*\|_\infty < R'_{N-1}(\ldots(R'_1(R))\ldots) > \|T_1(x^{n+1} - x^*)\|_\infty$ and $|\varepsilon^n| < \varepsilon_N \leq \varepsilon'_{N-1}(R'_{N-1}(\ldots(R'_1(R))\ldots))$ imply that $|x_2^{n+1} - x_2^*| = |b_2(T_1x^{n+1}) - x^*| < R'_{N-2}(\ldots(R'_1(R))\ldots)$.

Hence $\|T_2(x^{n+1}-x^*)\|_\infty = \max\{|x_2^{n+1} - x_2^*|,|x_1^{n+1} - x_1^*|\} \leq R'_{N-2}(\ldots(R'_1(R))\ldots)$,

since $R'_{N-1}(\ldots) \leq R_{N-1} \leq R'_{N-2}(\ldots)$, and so x_3^{n+1} is defined, etc.

Clearly this leads to x^{n+1} is defined and, in fact, $x^{n+1} \in S(x^*;R)$,

as long as $x^n \in S(x^*;R')$ and $|\varepsilon^n| < \varepsilon'$.

Part ii. Let $x^n \in S(x^*;R')$ and $|\varepsilon^n| < \varepsilon'$. Each g_i is continuously differentiable and so there exist functions $\rho_1, \ldots, \rho_N$ such that if $\|T_{N-i+1}(x - x^*)\|_\infty < R' > \|T_{N-i+1}(y - x^*)\|_\infty$, then

$$g_i(T_{N-i+1}x) - g_i(T_{N-i+1}y) - \nabla g_i(T_{N-i+1}y)T_{N-i+1}(x - y) = \rho_i(T_{N-i+1}x,T_{N-i+1}y)$$

and $\rho_i(T_{N-i+1}x,T_{N-i+1}y) / \|T_{N-i+1}(x - y)\|_\infty \to 0$ as $\|T_{N-i+1}(x - y)\|_\infty \to 0$.

Now ρ_i depends implicitly on $(x^n;\varepsilon^n)$, since g_i does, as well as on the explicit variables $T_{N-i+1}x$, $T_{N-i+1}y$. Obviously ρ_i is continuous in the explicit variables since the defining equation is, but we showed in Part i that the defining equation and hence ρ_i is also continuous in $(x^n;\varepsilon^n)$. First we note that

$$|b_N(T_{N-1}x)-x_N^*| = |x_N^* - x_N^n + (\Delta g_1(x^n;\varepsilon^n)u_N)^{-1}$$

$$\cdot\ [T_{N-1}\Delta g_1(x^n;\varepsilon^n)T_{N-1}(x - x^n) + \quad g_1(x^n)]|$$

$$\leq |x_N^* - x_N^n + (\Delta g_1(x^n;\varepsilon^n)u_N)^{-1}$$

$$\cdot\ [T_{N-1}\Delta g_1(x^n;\varepsilon^n)T_{N-1}(x - x^n) + \quad \nabla g_1(x^n)(x^n - x^*)]|$$

$$+ |\rho_1(x^*,x^n)| \qquad \cdot\ |\Delta g_1(x^n;\varepsilon^n)u_N|^{-1} .$$

Now, by the mean value theorem, $\Delta g_1(x^n;\varepsilon^n)u_N = \dfrac{\partial g_1}{\partial x_N}(\xi)$ for some $\xi \in (x^n, x^n + \varepsilon^n u_N)$. This is in the region where we assumed in Part i (without loss of generality) that this partial doesn't vanish and is hence bounded below by some number $1/b$ independent of $(x^n;\varepsilon^n)$. Thus the second term on the right hand side is bounded by $b|\rho_1(x^*,x^n)|$. In order to bound the first term on the right hand side, we add and subtract $(\Delta g_1(x^n;\varepsilon^n)u_N)^{-1}\ \Delta g_1(x^n;\varepsilon^n)(x^n - x^*)$ inside the absolute value. Rearranging terms we obtain

$$|x_N^* - x_N^n + (\Delta g_1(x^n;\varepsilon^n)u_N)^{-1}(\Delta g_1(x^n;\varepsilon^n)u_N)(x_N^n - x_N^*)$$

$$+ (\Delta g_1(x^n;\varepsilon^n)u_N)^{-1} (\nabla g_1(x^n) - \Delta g_1(x^n;\varepsilon^n))(x^n - x^*)$$

$$+ (\Delta g_1(x^n;\varepsilon^n)u_N)^{-1} T_{N-1}\Delta g_1(x^n;\varepsilon^n)T_{N-1}(x - x^n + x^n - x^*)|$$

$$\leq |\Delta g_1(x^n;\varepsilon^n)u_N|^{-1}$$

$$\cdot(\varepsilon^n)^{-1} \sum_{i=1}^{N} |\rho_1(x^n + \varepsilon^n u_i, x^n)| \, \|x^n - x^*\|_\infty + \|T_{N-1}(x - x^*)\|_1 .$$

We have used $|\Delta g_1(x^n;\varepsilon^n)u_N| = \|\Delta g_1(x^n;\varepsilon^n)\|_\infty$ as well as the Hölder inequaltiy for $p = 1$, $q = \infty$. Now combine the first inequality with earlier results to obtain

(3.1) $$|b_N(T_{N-1}x)-x_N^*| \leq \|T_{N-1}(x - x^*)\|_1 + b\|x^n - x^*\|_\infty$$

$$\cdot \sum_{i=1}^{N} |\rho_1(x^n + \varepsilon^n u_i,x^n) \,/\, \varepsilon^n| + b|\rho_1(x^*,x^n)|.$$

Without loss of generality we can appeal to Part i to assume that

$b^{-1} \geq |\Delta g_2(x^n;\varepsilon^n)u_{N-1}|$ uniformly for $\|x^n - x^*\| < R'$, $|\varepsilon^n| < \varepsilon'$.

CONVERGENCE OF BROWN'S METHOD FOR SIMULTANEOUS NONLINEAR EQUATIONS

We begin exactly as above.

$$|b_{N-1}(T_{N-2}x)-x^*_{N-1}| = |x^*_{N-1} - x^n_{N-1} + (\Delta g_2(x^n;\varepsilon^n)u_{N-1})^{-1}$$

$$\cdot [T_{N-2}\Delta g_2(x^n;\varepsilon^n)T_{N-2}(x - x^n) + g_2(T_{N-1}x^n)]| .$$

Remember that $g_2(T_{N-1}x^*) \neq 0$ and so the situation is slightly more complicated than before when $g_1(T_N x^*) = f_1(x^*) = 0$. We handle this as follows:

$$g_2(T_{N-1}x^n) = g_2(T_{N-1}x^n) - g_2(T_{N-1}x^*) + f_2[T_{N-1}x^*,b_N] - f_2(x^*)$$

$$= \nabla g_2(T_{N-1}x^n)T_{N-1}(x^n - x^*) - \rho_2(T_{N-1}x^*,T_{N-1}x^n)$$

$$+ \frac{\partial f_2}{\partial x_N}(\xi)(b_N(T_{N-1}x^*) - x^*_N) .$$

Of course $\xi \in (< T_{N-1}x^*,b_N > , x^*)$ and so its existence depends on this interval being of length no more than R'. From (3.1),

$$|b_N(T_{N-1}x^*) - x^*_N| \leq b\|x^n - x^*\|_\infty \sum_{i=1}^{N} |\rho_1(x^n + \varepsilon^n u_i,x^n) / \varepsilon^n| + b\rho_1(x^*,x^n)$$

which can be made arbitrarily small, and hence less than R' by taking

$\|x^n - x^*\|_\infty$ small. Select b' such that b' is a uniform upper bound on all the elements of $J(x)$ for $x \in \bar{S}(x^*;R')$. (We are really only concerned with the transverse strict lower triangular part of $J(x)$.)

At this point, split the right hand side of the inequality after substituting for $g_2(T_{N-1}x^n)$ and adding and subtracting $(\Delta g_2(x^n;\varepsilon^n)u_{N-1})^{-1}\ \Delta g_2(x^n;\varepsilon^n)T_{N-1}(x^n - x^*)$ and obtain as before,

$$|b_{N-1}(T_{N-2}x)-x^*_{N-1}| \leq |x^*_{N-1}-x^n_{N-1}+(\Delta g_2(x^n;\varepsilon^n)u_{N-1})^{-1}(\Delta g_2(x^n;\varepsilon^n)u_{N-1})(x^n_{N-1}-x^*_{N-1})$$

$$+ (\Delta g_2(x^n;\varepsilon^n)u_{N-1})^{-1}\ \ (\nabla g_2(T_{N-1}x^n) - \Delta g_2(x^n;\varepsilon^n))T_{N-1}(x^n-x^*)$$

$$+ (\Delta g_2(x^n;\varepsilon^n)u_{N-1})^{-1}\ T_{N-2}\Delta g_2(x^n;\varepsilon^n)T_{N-2}(x^n-x^*+x-x^n)|$$

$$+ b|\rho_2(T_{N-1}x^*,T_{N-1}x^n)| + bb'|b_N(T_{N-1}x^*) - x^*_N|$$

$$\leq \|T_{N-2}(x-x^*)\|_1 + b\|T_{N-1}(x^n-x^*)\|_\infty \sum_{i=1}^{N-1} |\rho_2(T_{N-1}(x^n+\varepsilon^n u_i),T_{N-1}x^n)/\varepsilon^n|$$

$$+ b|\rho_2(T_{N-1}x^*,T_{N-1}x^n)|$$

$$+ b^2b'\{\|x^n-x^*\|_\infty \sum_{i=1}^{N} |\rho_1(x^n+\varepsilon^n u_i,x^n)\ /\ \varepsilon^n| + |\rho_1(x^*,x^n)|\} .$$

$$(3.2)\ |b_{N-1}(T_{N-2}x)-x^*_{N-1}| \le \|T_{N-2}(x-x^*)\|_1$$

$$+ \sum_{j=0}^{1} b^{2-j}[(b')^{1+j}\|T_{N-j}(x^n-x^*)\|_\infty$$

$$\cdot \sum_{i=1}^{N-j} |\rho_{j+1}(T_{N-j}(x^n+\varepsilon^n u_i),T_{N-j}x^n)/\varepsilon^n| + |\rho_{j+1}(T_{N-j}x^*,T_{N-j}x^n)|\,].$$

There is no additional difficulty in establishing the general case,

$$(3.3)\ |b_{N-p}(T_{N-p-1}x)-x^*_{N-p}| \le \|T_{N-p-1}(x-x^*)\|_1$$

$$+ \sum_{j=0}^{p} b^{p-j+1}[(b')^{p-j}\|T_{N-j}(x^n-x^*)\|_\infty$$

$$\cdot \sum_{i=1}^{N-j} |\rho_{j+1}(T_{N-j}(x^n+\varepsilon^n u_i),T_{N-j}x^n)/\varepsilon^n| + |\rho_{j+1}(T_{N-j}x^n,T_{N-j}x^*)|\,]$$

for $0 \le p \le N-1$.

We know enough about the ρ functions to allow us to conclude that for any $\eta > 0$, there are positive numbers $R(\eta) \le R'$ and $\varepsilon(\eta) \le \varepsilon'$ such that for $|\varepsilon^n| < \varepsilon(\eta)$ and any $j \ge 0$, $|\rho_{j+1}(T_{N-j}(x^n+\varepsilon^n u_i),T_{N-j}x^n)/\varepsilon^n| < \eta$; and for $\|x^n - x^*\|_\infty < R(\eta)$ and any $j \ge 0$, $|\rho_{j+1}(T_{N-j}x^*,T_{N-j}x^n)| \le \eta\|T_{N-j}(x^n-x^*)\|_\infty$.

Hence there is a constant C, independent of x, x^n, ε^n such that for $\|x^n - x^*\|_\infty < R(\eta)$ and $|\varepsilon^n| < \varepsilon(\eta)$ we can simplify (3.3) to the following form.

$$(3.4)\quad |b_{N-p}(T_{N-p-1}x) - x^*_{N-p}| \le \|T_{N-p-1}(x - x^*)\|_1 + CN(N+1)\eta\,\|x^n - x^*\|_\infty .$$

Remember that b_1 is a constant function whose value is x_1^{n+1} and by (3.4), $p=N-1$,

$$|x_1^* - x_1^{n+1}| \le CN(N+1)\eta\,\|x^n - x^*\|_\infty ;$$

and by (3.4) with $T_1x = T_1x^{n+1}$,

$$|x_2^* - x_2^{n+1}| \le 2\,CN(N+1)\eta\,\|x^n - x^*\|_\infty .$$

Clearly then,

$$\|x^* - x^{n+1}\|_\infty \le 2^{N-1}N(N+1)C\eta\,\|x^n - x^*\|_\infty .$$

Choose $\eta < [2^{N-1}N(N+1)C]^{-1}$ and set $R'' \equiv R(\eta)$ and $\varepsilon'' \equiv \varepsilon(\eta)$ and the proof of *Part ii* is complete.

Part iii. Let the strong hypothesis hold and let i be any index between 1 and N.

$$\|\nabla f_i(x) - \nabla f_i(x^*)\|_1 \leq \max_{1\leq j\leq N} \|\nabla f_j(x) - \nabla f_j(x^*)\|_1 = \|J(x) - J(x^*)\|_\infty \leq K\|x - x^*\|_\infty .$$

Notice also that $|b_i(T_{i-1}x) - b_i(T_{i-1}y)| \leq (i-1)\|T_{i-1}(x - y)\|_\infty$ follows readily from the definition of b_i and the maximum component assumption on $\Delta g_{N-i+1}(x^n;\varepsilon^n)$. Let $1 \leq j \leq i$ and by the chain rule

$$\frac{\partial g_i}{\partial x_j}(T_{N-i+1}x) = \frac{\partial f_i}{\partial x_j}(< T_{N-i+1}x, b_{N-i+2}, \ldots, b_N >)$$

$$+ \sum_{k=N-i+2}^{N} \frac{\partial f_i}{\partial b_k}(< T_{N-i+1}x, b_{N-i+2}, \ldots, b_N >)\frac{\partial b_k}{\partial x_j}(< T_{N-i+1}x, \ldots, b_{k-1} >).$$

Now $\dfrac{\partial b_k}{\partial x_j}$ is a constant so

$$\|\nabla g_i(T_{N-i+1}x) - \nabla g_i(T_{N-i+1}x^*)\|_1 \leq \|T_{N-i+1}(\nabla f_i(x) - \nabla f_i(x^*))\|_1$$

$$+ \sum_{k=N-i+2}^{N} \Big| \frac{\partial f_i}{\partial b_k}(< T_{N-i+1}x, b_{N-i+2}, \ldots, b_N >)\frac{\partial b_k}{\partial x_j}$$

$$- \frac{\partial f_i}{\partial b_k}(< T_{N-i+1}x^*, b_{N-i+2}, \ldots, b_{k-1} >)\frac{\partial b_k}{\partial x_j} \Big| .$$

Inside each term of the sum, add and subtract $\frac{\partial f_i}{\partial b_k}(x^*)\frac{\partial b_k}{\partial x_j}$.

Rearrange the terms and use the fact that all the b-partials are less than or equal to one in absolute value. The following inequalities result:

$$\|\nabla g_i(T_{N-i+1}x) - \nabla g_i(T_{N-i+1}x^*)\|_1$$

$$\le \|\nabla f_i(x) - \nabla f_i(x^*)\|_1 + \|\nabla f_i(< T_{N-i+1}x,\dots,b_N >) - \nabla f_i(x^*)\|_1$$

$$+ \|\nabla f_i(x^*) - \nabla f_i(< T_{N-i+1}x^*,\dots,b_N >)\|_1$$

$$\le K\|x - x^*\|_\infty + K\| < T_{N-i+1}x,\dots,b_N > - x^*\|_\infty$$

$$+ K\|x^* - < T_{N-i+1}x^*,\dots,b_N > \|_\infty \; .$$

We can use (3.4), $0 \le p \le i-2$ to bound these last two terms.

$$|x^*_{N-i+2} - b_{N-i+2}(T_{N-i+1}x)| \le \|T_{N-i+1}(x - x^*)\|_1 + CN(N+1)\eta||x^n-x^*||_\infty .$$

$$|x^*_{N-i+3} - b_{N-i+3}(< T_{N-i+1}x, b_{N-i+2} >)| \le \|T_{N-i+1}(x - x^*)\|_1 + |x^*_{N-i+2} - b_{N-i+2}(T_{N-i+1}x)|$$

$$+ CN(N+1)\eta||x^n-x^*||_\infty$$

$$\le 2\|T_{N-i+1}(x-x^*)\|_1 + 2CN(N+1)\eta||x^n-x^*||_\infty .$$

$$\vdots \qquad\qquad \vdots$$

$$|x^*_{N-i+j} - b_{N-i+j}(< T_{N-i+1}x, b_{N-i+2}, \ldots, b_{N-i+j-1} >)|$$

$$\leq 2^{j-2} \|T_{N-i+1}(x - x^*)\|_1 + 2^{j-2} CN(N+1)\eta \| x^n - x^* \|_\infty .$$

Hence

$$\|\nabla g_i(T_{N-i+1}x) - \nabla g_i(T_{N-i+1}x^*)\|_1 \leq 2K\| x-x^* \|_\infty + K2^{i-1} \| T_{N-i+1}(x-x^*) \|_1$$

$$+ 2K2^{i-1} CN(N+1)\eta \| x^n - x^* \|_\infty .$$

But since the ℓ_1 and ℓ_∞ norms are equivalent, we can pick constants Q and Q' such that the following inequality holds for every $i=1,\ldots,N$:

$$\|\nabla g_i(T_{N-i+1}x) - \nabla g_i(T_{N-i+1}x^*)\|_1 \leq Q\|x - x^*\|_\infty + Q'\eta\|x^n - x^*\|_\infty \quad .$$

At this point we wish to reexamine $\rho_{j+1}(T_{N-j}(x^n+\varepsilon^n u_i), T_{N-j}x^n / \varepsilon^n$ and $\rho_{j+1}(T_{N-j}x^n, T_{N-j}x^*)$ for $i \leq N-j$.

We can write

$$\rho_{j+1}(T_{N-j}(x^n+\varepsilon^n u_i),T_{N-j}x^n) = \int_0^1 [\nabla g_{j+1}(T_{N-j}(x^n+t\varepsilon^n u_i)) - \nabla g_{j+1}(T_{N-j}x^n)]\varepsilon^n T_{N-j}u_i dt$$

$$= \int_0^1 [\nabla g_{j+1}(T_{N-j}(x^n+t\varepsilon^n u_i)) - \nabla g_{j+1}(T_{N-j}x^*)]\varepsilon^n T_{N-j}u_i dt$$

$$+ [\nabla g_{j+1}(T_{N-j}x^*) - \nabla g_{j+1}(T_{N-j}x^n)]\varepsilon^n T_{N-j}u_i \quad .$$

Hence,

$$|\rho_{j+1}(T_{N-j}(x^n+\varepsilon^n u_i),T_{N-j}x^n)|$$

$$\leq |\varepsilon^n| \cdot Q \cdot (\|x^n-x^*\|_\infty + |\varepsilon^n|) + |\varepsilon^n| Q'\eta\|x^n-x^*\|_\infty + |\varepsilon^n| (Q+Q'\eta)\|x^n-x^*\|_\infty$$

and so there is a constant Q'' such that

$$|\rho_{j+1}(T_{N-j}(x^n+\varepsilon^n u_i),T_{N-j}x^n) \;/\; \varepsilon^n| \leq Q''\|x^n - x^*\|_\infty + Q|\varepsilon^n| \;.$$

If we choose $|\varepsilon^n| = 0(|f_1(x^n)|) = 0(|f_1(x^n) - f_1(x^*)|) = 0(\|x^n - x^*\|_\infty)$,

then we may as well assume Q'' was chosen such that

$|\rho_{j+1}(T_{N-j}(x^n+\varepsilon^n u_i),T_{N-j}x^n) \;/\; \varepsilon^n| \leq Q''\|x^n - x^*\|_\infty$. Now write

$$\rho_{j+1}(T_{N-j}x^n,T_{N-j}x^*)$$

$$= \int_0^1 [\nabla g_{j+1}(T_{N-j}(x^n + t(x^*-x^n)) - \nabla g_{j+1}(T_{N-j}x^*)] \cdot T_{N-j}(x^*-x^n)dt \;;$$

thus $|\rho_{j+1}(T_{N-j}x^n, T_{N-j}x^*)| \leq (Q + Q'\eta)\|x^n - x^*\|_\infty^2$. This allows us to redevelop (3.4) as follows:

(3.5) $|x^*_{N-p} - b_{N-p}(T_{N-p-1}x)| \leq \|T_{N-p-1}(x - x^*)\|_1 + C'\|x^n - x^*\|_\infty^2$.

From whence we obtain

$$|x_1^* - x_1^{n+1}| \leq C'\|x^n - x^*\|_\infty^2 ,$$

$$|x_2^* - x_2^{n+1}| \leq 2C'\|x^n - x^*\|_\infty^2 ,$$

$$\vdots$$

etc., until

$$\|x^* - x^{n+1}\|_\infty \leq 2^{N-1}C'\|x^* - x^n\|_\infty^2$$

and so the method is at least second order, since $2^{N-1}C'$ is independent of n.

This completes the proof.

4. Numerical Results.

Example 4.1. The following are Powell's equations [13] derived from Rosenbrock's function [14].

$$f_1(x_1,x_2) \equiv 10(x_2 - x_1^2) = 0$$

$$f_2(x_1,x_2) \equiv 1 - x_1 = 0$$

The "standard" starting guess (-1.2, 1.0) was used and the results are given in Table 4.1.

Table 4.1.

Method	Final $\|F\|$	Number of (Equivalent) Evaluations of the Function Vector F
Newton	7.5×10^{-8}	12
Broyden (I) [6]	4.8×10^{-10}	59
Broyden (II) [6]	2.5×10^{-10}	39
Brown (I)	zero	7
Brown (II)	zero	9

Remark 4.1. Rosenbrock's function [14]

$$\phi(x_1,x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \tag{4.1}$$

consists of a steep-sided parabolic valley whose single minimum occurs at $x_1 = x_2 = 1$. C.G. Broyden (in correspondence) proposed that a starting guess of (-0.8, 1.0) on the other side of the valley might provide a good challenge for method (I). We are happy to report a final norm of 6.5×10^{-7} in only 5 (equivalent) function vector evaluations for the proposed starting guess.

When we solved the system in the reverse order

$$f_2 = 0$$

$$f_1 = 0$$

we obtained convergence in just one iteration from the standard starting guess (-1.2, 1.0). This example supports a good general strategy to follow when using Brown's methods (I) and (II) namely, always preorder the system of equations so that the linear (or most nearly linear) equations come first and then the remaining equations become progressively more nonlinear -- as measured, say, by their degree.

Broyden has pointed out in [7] that his new methods, also [7], will produce the exact solution (1.0, 1.0) from the standard starting guess in just three iterations.

Example 4.2. Rosenbrock's function (4.1) has been used as a standard ("tough") example to test many function minimization algorithms. One way of minimizing a function, ϕ, of N variables is to locate the zeros of the associated gradient system, for, as is well known, any local minima of ϕ must occur among the zeros of $\nabla\phi$. We take this approach with Rosenbrock's function (4.1); hence, we seek the zeros of

$$f_1(x_1,x_2) \equiv 2(x_1 - 1) - 400x_1(x_2 - x_1^2) = 0$$

$$f_2(x_1,x_2) \equiv 200(x_2 - x_1^2) = 0 .$$

Again we take $x^0 = (-1.2, 1.0)$. The results are given in Table 4.2.

The problem was run for $N = 5, 10, 15$ and 20 with the starting vector for all cases being a vector having 0.5 in each component. Both Brown (I) and Brown (II) converged in each case to the root, a vector all of whose components are 1.0. For $N = 5$ Newton's method converged to the root given approximately by (-.579, -.579, -.579, -.579, 8.90). We note that the failure of Newton's method on this problem is root attributable to singularities of the Jacobian matrix, since the Jacobian matrix is nonsingular at the starting guess and at the two roots. The results are given in Table 4.3. In the table "diverged" means that $\|x^n\|_\infty \to \infty$ whereas "converged" means that each component of x^{n+1} agreed with the corresponding component of x^n to 15 significant digits and $\|f(x^{n+1})\|_{\ell_2} < 10^{-15}$; moreover, conv. ≡ converged, div. ≡ diverged, and its. ≡ iterations.

Table 4.2.

Method	Final Value of ϕ	Number of (Equivalent) Evaluations of the Function Vector F
Powell [13]	$< 1.0 \times 10^{-4}$	70
Stewart [18]	1.7×10^{-6}	132
Broyden [7]		
a) "O.M.D."	failed	> 500
b) "O.M.U."	failed	> 500
c) New $\lambda = .05$ (version 1)	failed	> 500
d) New $\lambda = .1$ (version 1)	failed	> 500
e) New $\lambda = .2$ (version 1)	$< 1.0 \times 10^{-12}$	158
f) New $\lambda = .05$ (version 2)	$< 1.0 \times 10^{-12}$	480
g) New $\lambda = .1$ (version 2)	$< 1.0 \times 10^{-12}$	184
h) New $\lambda = .2$ (version 2)	$< 1.0 \times 10^{-12}$	188
Brown (I)	$< 1.3 \times 10^{-11}$	53

Example 4.3. In order to illustrate how Brown's methods (I) and (II) capitalize on the preordering strategy given in Remark 4.1, we consider the following example from [3]

$$f_i(x) \equiv -(N+1) + 2x_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} x_j \,, \qquad i = 1, \ldots, N-1$$

$$f_N(x) \equiv -1 + \prod_{j=1}^{N} x_j \,.$$

Table 4.3.

N	Newton's Method	Brown (I)	Brown (II)
5	conv. (18 its.)	conv. (7 its.)	conv. (6 its.)
10	div., $\|x^1\| \sim 10^3$	conv. (8 its.)	conv. (7 its.)
15	div., $\|x^1\| \sim 10^5$	conv. (8 its.)	conv. (8 its.)
20	div., $\|x^1\| \sim 10^6$	conv. (8 its.)	conv. (8 its.)

Example 4.4. This system is due to Freudenstein and Roth [11]:

$$f_1(x_1,x_2) \equiv -13 + x_1 + ((-x_2 + 5)x_2 - 2)x_2 = 0$$

$$f_2(x_1,x_2) \equiv -29 + x_1 + ((x_2 + 1)x_2 - 14)x_2 = 0 .$$

The starting guess used was $x^0 = (15, -2)$. The solution is at (5,4). The results are given in Table 4.4.

Table 4.4.

Method	Result
Newton	converged in 42 iterations
Broyden's I [6, p. 591]	diverged
Broyden's II [6, p. 591]	diverged
Broyden [7]	diverged
Damped Newton (discrete form) [17]	diverged
Brown (I) and (II)	converged in 10 iterations

Example 4.5. This example is a macroeconomic model due to Christensen [8]. It entails a system of 19 simultaneous equations, ten of which are linear. Our colleague, R.M. Bass of the Office of Emergency Preparedness, Washington, D.C., has used our program to solve these equations. With "good" starting guesses he obtained convergence for all 39 time periods using an average of only 3.5 iterations per time period. With poorer starting guesses he obtained convergence for 22 of the 39 time periods with an average of 5.4 iterations per time period (when convergence was obtained). Dr. Bass had originally solved these equations by minimizing $\sum_{i=1}^{19} f_i^2$, using the method of Fletcher and Powell [10]. Approximately 500 iterations were needed to reduce the sum of squares to 5×10^{-3}.

Remark 4.2. As the Rosenbrock example (contrast example 4.1 with 4.2) and Bass' experience confirm experimentally, it is ridiculous to complicate the problem of solving simultaneous nonlinear equations unnecessarily; specifically, <u>do not solve a nonlinear system by attempting to minimize</u> $\sum f_i^2$!

Example 4.6. Scarf [16] has given an elegant method for finding the fixed points of a mapping which takes the unit simplex into itself; i.e. he has given a constructive proof of Brouwer's fixed point theorem. Scarf's technique turns out to be remarkably easy to implement on a digital computer. The technique applies directly to a nonlinear model of a pure trade economy. Scarf's algorithm is an example of a good technique for attacking the first subproblem of solving nonlinear equations: getting into a region of local convergence from perhaps poor initial estimates (see §1). We coupled Brown (I) with Scarf's algorithm and tested it on a ten dimensional pure trade model with the following results: the time needed to solve the problem was reduced from 4.6 minutes (when using "pure" Scarf) to just 16 seconds when employing the hybrid technique of using Scarf's algorithm to get an initial guess and then switching over to Brown (I).

Acknowledgements. The first author wishes to thank Professor Samuel D. Conte, his thesis advisor, for his abundant encouragement and support, and, in particular, for many helpful discussions relative to portions of this research. This author also wishes to thank Professor Peter Henrici who guided and stimulated the author's early work in numerical analysis and who proposed research along the present lines. Further thanks are due to Jim Ortega, Joe Traub and Louis Rall for their early encouragement of this research. We are most appreciative of the help given us by Ron Bass and Herb Scarf relative to examples 4.5 and 4.6 and to David Jordan, Richard Tenney and Roy Weber for their assistance with the numerical experiments.

omit

Ken, This seems a little like blackmail.

REFERENCES

1. Brown, K.M. (1966) A Quadratically Convergent Method for Solving Simultaneous Nonlinear Equations, Ph. D. Diss., Purdue Univ., Lafayette, Indiana.

2. Brown, K.M. (1967) "Solution of Simultaneous Nonlinear Equations," Comm. ACM Vol. 10, pp. 728-729.

3. Brown, K.M. (1969) "A Quadratically Convergent Newton-like Method Based Upon Gaussian Elimination," SIAM J. on Num. Anal., Vol. 6, pp. 560-569.

4. Brown, K.M. and S.D. Conte (1967) "The Solution of Simultaneous Nonlinear Equations," Proc. 22nd Nat. Conf., ACM, pp. 111-114.

5. Brown, K.M. and W.B. Gearhart (1971) "Deflation Techniques for the Calculation of Further Solutions of a Nonlinear System," Numer. Math., Vol. 16, pp. 334-342.

6. Broyden, C.G. (1965) "A Class of Methods for Solving Nonlinear Simultaneous Equations," Math. Comp., Vol. 19, pp. 577-593.

7. Broyden, C.G. (1969) "A New Method of Solving Simultaneous Nonlinear Equations," The Computer Journal, Vol. 12, No. 1, pp. 94-99.

8. Christensen, L.R. (1970) "Tax Policy and Investment Expenditures in a Model of General Equilibrium," American Economic Review, Vol. 2, pp. 18-22.

9. Dennis, J.E. Jr. (1970) "On the Convergence of Newton-like Methods," Numerical Methods for Nonlinear Algebraic Equations, P. Rabinowitz, ed., Gordon and Breach, London.

10. Fletcher, R. and M.J.D. Powell (1963) "A Rapidly Convergent Descent Method for Minimization," Computer Journal, Vol. 6, pp. 163-168.

REFERENCES (cont.)

11. Freudenstein, F. and B. Roth (1963) "Numerical Solutions of Systems of Nonlinear Equations," JACM, Vol. 10, pp. 550-556.

12. Ortega, J.M. and W.C. Rheinboldt (1970) Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York.

13. Powell, M.J.D. (1965) "A Method for Minimizing a Sum of Squares of Nonlinear Functions Without Calculating Derivatives," Computer Journal, Vol. 7, pp. 303-307.

14. Rosenbrock, H.H. (1960) "An Automatic Method for Finding the Greatest or Least Value of a Function," Computer Journal, Vol. 3, pp. 175-184.

15. Samanskii, V. (1967) "On a Modification of the Newton Method (Russian)," Ukrain. Mat. Z., Vol. 19, pp. 133-138.

16. Scarf, H. (1967) "The Approximation of Fixed Points of a Continuous Mapping," SIAM J. Appl. Math., Vol. 15, No. 5, pp. 1328-1343.

17. Späth, H. (1967) "The Damped Taylor's Series Method for Minimizing a Sum of Squares and for Solving Systems of Nonlinear Equations," CACM, Vol. 10, pp. 726-728.

18. Stewart, G.W. III (1967) "A Modification of Davidon's Minimization Method to Accept Difference Approximations of Derivatives," JACM, Vol. 14, No. 1, pp. 72-83.