SIAM J. NUMER ANAL.
Vol. 17, No. 6, December 1980

© 1980 Society for Industrial and Applied Mathematics
0036-1429/80/1706-0008 $01.00/0

PARAMETER SELECTION FOR NEWTON-LIKE METHODS APPLICABLE TO NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS*

R. E. BANK† AND D. J. ROSE‡

Dedicated to Professor Garrett Birkhoff on the occasion of his seventieth birthday

Abstract. We discuss two Newton-like methods for solving the systems which arise from nonlinear partial differential equations and nonlinear networks. Under appropriate conditions both methods provide global and quadratic convergence provided parameters controlling the iteration are chosen correctly.

1. Introduction.

Consider the nonlinear system of equations

$$g(u) = 0, \tag{1.1}$$

where $g = (g_1, g_2, \cdots, g_n)^T$ and $u = (u_1, u_2, \cdots, u_n)^T$. Our particular interest here is in applications where the system (1.1) arises from suitable discretization of elliptic partial differential equations or from an equilibrium electrical network. For example, $g(u)$ may have the form

$$g(u) = Au + f(u), \tag{1.2a}$$

$$g(u) = A(u)u + f(u), \tag{1.2b}$$

or

$$g(u) = AC(A^T u) + f(u). \tag{1.2c}$$

Under appropriate conditions on the matrices A (or $A(u)$) and the nonlinear mappings $f(u)$ and $C(x)$, each $g(u)$ of (1.2) is a homeomorphism from $R^n \to R^n$, and hence the system (1.1) possesses a unique solution. The mapping in (1.2a) arises naturally from the mildly nonlinear Poisson equation $-\Delta u + f(x, y, u) = 0$, for example, while (1.2b) arises from a PDE of the form

$$-\nabla \cdot a(x, y, u)\nabla u + f(x, y, u) = 0.$$

The mapping (1.2c) represents the equilibrium potential of an electrical network when A is interpreted as the reduced arc-vertex incidence matrix of the network and $C(x)$ is interpreted as a vector of nonlinear conductivity functions on the arcs. See [6], [9], [10], [11] for further discussions of these applications.

In order to solve the nonlinear system (1.1), we will consider two methods related to Newton's method, which proceeds from the iterate u^k to u^{k+1} by solving the linear system

$$g'(u^k)x = -g(u^k) \tag{1.3a}$$

and setting

$$u^{k+1} = u^k + x. \tag{1.3b}$$

Here $g'(u)$ is the Jacobian of the mapping $g(u)$. The methods we discuss modify (1.3), producing a change $u^{k+1} - u^k = y$ at each step to assure global convergence of the iteration. Furthermore, the change y has the property that $y \to x$ of (1.3) as $k \to \infty$. This

* Received by the editors January 17, 1980.

† University of Texas, Austin, Texas 78712. The work of this author was partially supported by the U.S. Office of Naval Research under Contract N00014-80-C-0645.

‡ Bell Laboratories, Murray Hill, New Jersey 07974.

implies (see [4, p. 51]) that $u^k \to u^*$ with $g(u^*) = 0$ at least superlinearly; we will show that each method converges quadratically.

As in (1.3a), the methods we present require solving a linear system at each iteration involving $g'(u^k)$. We will thus assume the global existence of $g'(u)$ as well as the uniform positive definite condition

$$x^T g'(u)x \geqq k_1 x^T x, \tag{1.4}$$

for all x and u in R^n. If $g'(u)$ is also continuous on R^n, then $g(u)$ is *uniformly monotone* and g is a homeomorphism from $R^n \to R^n$ (see [8, §§ 5.4, 6.4]). While these global conditions can be weakened and localized, such assumptions simplify our exposition and correspond loosely to the requirement that the original PDE be elliptic and well-posed for any set of appropriate boundary conditions.

In applications the matrices $g'(u^k)$ are large and sparse, and the associated linear systems require careful attention computationally. While we do not discuss in detail any particular linear equations solver, our computations and those of our colleagues have shown that recent specialized solvers such as fast direct methods, multigrid iteration, sparse direct factorization, and vectorized SOR can be quite effective in this context. That is, achieving quadratic convergence of the overall iteration more that compensates for the time spent in the linear equations solver.

In § 2 we consider the case when $g'(u)$ is also symmetric. Then g is a *gradient mapping*; i.e., there exists a real valued function $V: R^n \to R^1$ such that $V'(u) = g^T(u)$. The condition (1.4) then implies that $V(u)$ is *uniformly convex* and *coercive* in the sense that $V(u) \to +\infty$ as $\|u\| \to \infty$.[1] Hence $V(u)$ has a unique minimum at $u = u^*$ with $g(u^*) = 0$. See [8, §§ 3.4, 4.1–4.4].

In the gradient mapping case, a natural modification to Newton's method uses (1.3a) while replacing (1.3b) by the *damped* change

$$u^{k+1} = u^k + t_k x, \qquad t_k > 0. \tag{1.5}$$

We show how to choose the damping parameters t_k in order to provide global and quadratic convergence of the sequence u^k. Furthermore, we require only the evaluation of $V'(u) = g^T(u)$ rather than the evaluation of $V(u)$ as discussed in [4, § 6], [8, § 8.3, 14.2–14.4] and [10, p. 45]. This is an advantage when $g(u)$ arises from a PDE or a nonlinear network.

Section 3 discusses a method for solving (1.1) under the general (nonsymmetric) uniformly positive definite condition (1.4). The method can be motivated by considering an implicit scheme for solving the autonomous system of ordinary differential equations

$$du/dt + g(u(t)) = 0, \qquad u(0) = u^0.$$

This leads to a Newton-like method which replaces (1.3a) by the linear system

$$(I/s_k + g'(u^k))x = -g(u^k), \tag{1.6}$$

computing u^{k+1} as in (1.3b). The parameters s_k correspond to the "time steps" in the discrete ODE scheme. Such methods are also known as continuation methods ([10, § 10.4]) or embedding methods ([7]). Again our aim is to determine the sequence or parameters such that $u^k \to u^*$ quadratically for any u^0.

In § 4 we present some numerical experiments which illustrate and compare the algorithms discussed in §§ 2 and 3. A nonlinear system (1.1) arises from a finite element

[1] We use the 2-norm throughout; $\|x\| = (x^T x)^{1/2}$.

discretization of the nonlinear elliptic PDE

$$-\Delta u + e^{u-v} - e^{w-u} = k(x, y), \tag{1.7}$$

with standard boundary conditions. Here $u(x, y)$ and $k(x, y)$ are defined on a rectangular region of R^2 as are given functions v and w. Equation (1.7) arises in semiconductor device modeling, and is more generally part of a coupled system of elliptic PDE's,

$$-\Delta u + e^{u-v} - e^{w-u} = k(x, y), \tag{1.8}$$

$$-\nabla \cdot (\mu_n e^{u-v} \nabla v) = 0, \tag{1.9}$$

$$-\nabla \cdot (\mu_p e^{w-u} \nabla w) = 0. \tag{1.10}$$

In this application the functions u, v, and w are functions of $(x, y) \in R^2$, as are the known functions μ_n, μ_p, and k; $k(x, y)$ is a doping profile of the device. Equation (1.8) is a nonlinear Poisson equation for the electrostatic potential, while (1.9) and (1.10) are continuity equations for quasi-Fermi levels of electrons and holds, respectively. Equations (1.8)–(1.10) are posed on bounded regions of R^2, often unions of rectangles, with appropriate standard elliptic boundary conditions. The numerical solution of (1.8)–(1.10), using parts of the methodology presented here, will be presented elsewhere; see [2] for a preliminary discussion.

2. Gradient mappings. Throughout this section we will assume that our mapping g is a gradient mapping, and that the associated variational functional $V: R^n \to R^1$ with

$$V'(u) = g^T(u) \tag{2.1}$$

satisfies the coerciveness condition

$$V(u) \to +\infty \quad \text{as } \|u\| \to \infty, \tag{2.2}$$

and is strictly convex on R^n; i.e.,

$$V(tu + (1-t)z) < tV(u) + (1-t)V(z) \tag{2.3}$$

for $u \neq z$ and any $t \in (0, 1)$. As we noted in § 1, the uniform positive definite condition (1.4) insures that (2.2) and (2.3) hold, implying a unique u^* minimizing $V(u)$ with $g(u^*) = 0$.

For given u and x consider the halfline

$$w(t) = u + tx, \qquad t \in [0, \infty). \tag{2.4}$$

The functional $V(w(t))$ is clearly strictly convex on this convex subset and has a unique minimum at some t^*. The *directional derivative*, dV/dt, along the line $w(t)$ is given by

$$dV(w(t))/dt \equiv V'(t) = V'(w(t))x = g^T(w(t))x. \tag{2.5}$$

If x is chosen so that $g^T(u)x < 0$, then $t^* > 0$, where t^* satisfies

$$V'(t^*) = g^T(w(t^*))x = 0. \tag{2.6}$$

Hence V can be decreased, certainly for any $0 < t \leqq t^*$. Suppose $g(u) \neq 0$ and x satisfies

$$M(u)x = -g(u), \tag{2.7}$$

for positive definite $M(u)$. Then (2.5) implies $V'(0) < 0$, and x is called a *descent direction*. Examples are the *steepest descent* direction when $M(u) = I$, and the *Newton* direction when $M(u) = g'(u)$. A convergence result for such descent methods is given in the following

PROPOSITION 1. *Let M be a continuous mapping from R^n to the set of $n \times n$ matrices such that $M(u)$ is positive definite. Suppose the gradient mapping $g(u)$ with V as in* (2.2) *and* (2.3) *is continuously differentiable. Given u^0, consider the iterates*

$$u^{k+1} = u^k + t_k x, \tag{2.8}$$

where

$$M(u^k)x = -g(u^k). \tag{2.9}$$

Then there exist closed intervals $I_k = [\alpha, \beta_k]$, $\alpha > 0$, $\beta_k > 0$, containing $t_k^* = t^*$ of (2.6) *(with $u = u^k$) such that if $t_k \in I_k$ for all $k \geqq 0$, the u^k converge to u^* with $g(u^*) = 0$.*

Proof. Since $M(u)$ and $g'(u)$ are continuous and positive definite on the compact convex set $D = \{u \mid V(u) \leqq V(u^0)\}$, $M(u)$ and $g'(u)$ are uniformly positive definite on D; i.e., there are constants c_i and k_i, $i = 1, 2$, such that

$$c_1 y^T y \leqq y^T M(u) y \leqq c_2 y^T y \tag{2.10a}$$

and

$$k_1 y^T y \leqq y^T g'(u) y \leqq k_2 y^T y, \tag{2.10b}$$

for all $y \in R^n$ and $u \in D$. We show that (2.10) implies that the t_k^* are uniformly bounded away from zero. To see this, let $w(t)$ and x be as in (2.4) and (2.7) respectively. Writing $V(t)$ for $V(w(t))$ and using a Taylor expansion about $u = w(0)$, we obtain for $t > 0$ with $w(t) \in D$,

$$V(t) - V(0) = t g^T(w(0))x + (t^2/2) x^T g'(w(s)) x, \tag{2.11}$$

for $s \in (0, t)$. Substituting for $g(u)$ from (2.7) and using (2.10) gives the bounds

$$p_1(t) x^T x \leqq V(t) - V(0) \leqq p_2(t) x^T x, \tag{2.12}$$

where the p_i are parabolas given by

$$p_1(t) = -c_2 t + k_1 t^2/2 \tag{2.13a}$$

and

$$p_2(t) = -c_1 t + k_2 t^2/2; \tag{2.13b}$$

see Fig. 1.

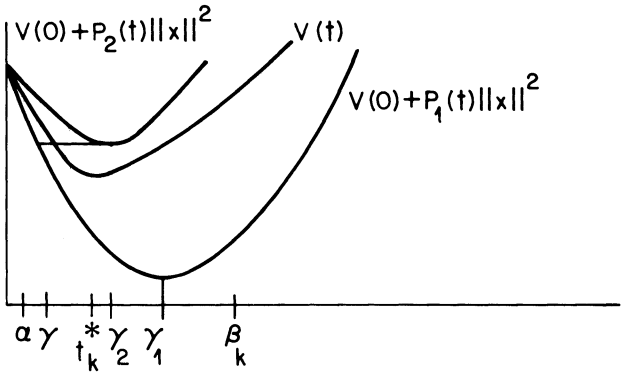

FIG. 1

Now $p_1(t)$ and $p_2(t)$ are minimized at $\gamma_1 = c_2/k_1$ and $\gamma_2 = c_1/k_2$, respectively. Let $\gamma = \min\{t | p_1(t) = p_2(\gamma_2)\}$; $0 < \gamma \leqq \gamma_2$ and $V(\gamma) < V(0)$ for $g(u) \neq 0$ by the strict convexity of $V(t)$. Hence $\gamma = \gamma(k_1, k_2, c_1, c_2) < t^*$ where $V'(t^*) = 0$.

Let α be any fixed constant in $(0, \gamma]$ and, for each iterate u^k, let $\beta_k > \alpha$ be the unique t such that

$$V(u^k + tx) = V(u^k + \alpha x), \tag{2.14}$$

with x as in (2.9). Clearly $t_k^* \in I_k = [\alpha, \beta_k]$ and, from (2.12),

$$V(w(t_k)) - V(u^k) \leqq V(w(\alpha)) - V(u^k) \leqq p_2(\alpha)\|x\|^2. \tag{2.15}$$

It remains to show that the iteration defined by (2.8) and (2.9) with $t_k \in I_k$ converges to u^* with $g(u^*) = 0$. Supposing $g(u^k) \neq 0$ for all k, (2.15) implies that all $u^k \in D$ and the sequence $V_k = V(u^k)$ is strictly decreasing. Since $V(u)$ is bounded below on D, $V_k \to \mu$ as $k \to \infty$. Compactness of D and the continuity and strict convexity of V imply a unique cluster point u^* of the sequence u^k such that $V(u^*) = \mu$ and $\lim u^k = u^*$.

Finally using (2.15) with $u^{k+1} = w(t_k)$ and $c(\alpha) = -p_2(\alpha) > 0$ shows that

$$\|x\| \leqq [(V_k - V_{k+1})/c(\alpha)]^{1/2}. \tag{2.16}$$

Then, since $\|g(u^k)\| \leqq \|M(u^k)\|\,\|x\|$, and $\|M(u)\| \leqq c_3$ for some $c_3 > 0$ on the compact set D,

$$\|g(u^k)\| \leqq c_4[V_k - V_{k+1}]^{1/2}, \qquad c_4 = c_3 c(\alpha)^{-1/2}. \tag{2.17}$$

This implies $g(u^*) = \lim g(u^k) = 0$. □

The essential content of Proposition 1 is known, although our self-contained treatment, especially regarding the intervals I_k, will prove useful below. For a considerably more general treatment necessitating notions not needed here see [8, Chaps. 8, 14].

Recall that for continuously differentiable $g(u)$, the uniform positive definite condition (1.4) is equivalent to uniform monotonicity [8, pp. 141–142]; i.e.,

$$(g(u) - g(v))^T(u - v) \geqq k_1(u - v)^T(u - v). \tag{2.18}$$

Then, using the Cauchy-Schwarz inequality, we obtain $\|u - v\| \leqq \|g(u) - g(v)\|/k_1$, implying

$$\|u - u^*\| \leqq \|g(u)\|/k_1. \tag{2.19}$$

Combining (2.17) and (2.19) gives the bound

$$\|u^k - u^*\| \leqq (c_4/k_1)[V_k - V_{k+1}]^{1/2}. \tag{2.20}$$

As a final remark, note that (2.17) implies that given any $\varepsilon > 0$ there is some $k = l$ such that $\|g(u^l)\| < \varepsilon$. Hence if the descent strategy is used as a subiteration, we can always attain a norm-decreasing major iteration; i.e., given $\alpha_k < 1$ we can attain

$$\|g(u^{k+1})\| < \alpha_k \|g(u^k)\| \tag{2.21}$$

by decreasing V in an inner iteration to go from k to $k + 1$. An application of this simple observation will be given later.

We turn now to a detailed discussion of the Newton descent method described by (2.8) and (2.9) replaced by

$$g'(u^k)x = -g(u^k). \tag{2.22}$$

Our goal is to choose the seps t_k consistent with Proposition 1 while maintaining at least superlinear convergence. Recall that $u^k \to u^*$ *superlinearly* if there is a sequence $\alpha_k \to 0$ such that

$$\|u^{k+1} - u^*\| \leqq \alpha_k \|u^k - u^*\|; \tag{2.23}$$

$u^k \to u^*$ *quadratically* if

$$\|u^{k+1} - u^*\| \leqq \beta \|u^k - u^*\|^2, \tag{2.24}$$

for some $\beta > 0$ and $k > K_1$. We begin by showing that when $u^k \to u^*$ in Proposition 1, the sequence $t_k^* \to 1$. Our statement and proof strengthen considerably the result claimed in [5].

THEOREM 1. *Let $g(u)$ be as in Proposition* 1, *with u^k and x as in* (2.8) *and* (2.22), *respectively. Then for any sequence t_k such that $u^k \to u^*$, the sequence $t_k^* \to 1$. Furthermore, if there exists a Lipschitz constant $L > 0$ such that*

$$\|g'(u) - g'(u^*)\| \leqq L\|u - u^*\|, \tag{2.25}$$

for u in some set $S = \{u \mid \|u - u^*\| \leqq r\}$, then, when $u^k \in S$,

$$|t_k^* - 1| < K_2 \|u^k - u^*\| \tag{2.26}$$

for $K_2 > 0$.

Proof. Let $w(t) = u^k + tx$, and write $V(t)$ for $V(w(t))$. We will show that for any fixed $t \in (0, 1)$,

$$V'(t) + V'(1) < 0 \quad \text{as } k \to \infty, \tag{2.27}$$

and, similarly, for any fixed $t \in (1, 2)$,

$$V'(t) + V'(1) > 0 \quad \text{as } k \to \infty. \tag{2.28}$$

Supposing (2.27) and (2.28), let $\varepsilon > 0$ be arbitrarily small and take $t = 1 - \varepsilon$ in (2.27) and $t = 1 + \varepsilon$ in (2.28). Then, since $V'(t)$ is strictly increasing, $V'(1-\varepsilon) < V'(1) < V'(1+\varepsilon)$, while (2.27) and (2.28) imply $V'(1-\varepsilon) < -V'(1) < V'(1+\varepsilon)$. Thus $V'(1-\varepsilon) < 0 < V'(1+\varepsilon)$ showing $|t_k^* - 1| < \varepsilon$. Hence establishing (2.27) and (2.28) will imply $t_k^* \to 1$ as $k \to \infty$. Let S_k be the compact convex set, $S_k = \{u \mid \|u - u^*\| \leqq 2k_1^{-1}\|g(u^k)\| + \|u^k - u^*\|\}$, k_1 as in (2.10b), and

$$\eta_k = \max_{u \in S_k} \|g'(u) - g'(u^*)\|. \tag{2.29}$$

Note that $w(t) \in S_k$ for $t \in (0, 2)$, since $w(t) - u^* = (u^k - u^*) + t[g'(u^k)]^{-1} g(u^k)$ and $g'(u)$ is symmetric.

To establish (2.27) and (2.28) recall $V'(t) = g^T(w(t))x$. Expanding $g(w(t))$ for $w(t) \in S_k$ [8, p. 71] shows

$$g(w(t)) = g(u) + t \int_0^1 g'(w(st))x \, ds. \tag{2.30}$$

Combining (2.22) and (2.30) and rearranging gives

$$V'(t) = x^T\Big(((t-1)/2) \int_0^1 [g'(u) + g'(w(st))] \, ds + ((t+1)/2) \int_0^1 [g'(w(st)) - g'(u) \, ds\Big)x. \tag{2.31}$$

Adding and subtracting $g'(u^*)$ in the second integrand of (2.31) and then combining (2.29) and (2.31) twice yields

$$V'(t)+V'(1)\leqq((t-1)k_1+\eta_k(t+3))x^Tx \tag{2.32}$$

for $t\in(0,1)$, and similarly

$$V'(t)+V'(1)\geqq((t-1)k_1-\eta_k(t+3))x^Tx \tag{2.33}$$

for $t\in(1,2)$. Since $\eta_k\to 0$ as $k\to\infty$, (2.31) and (2.32) establish (2.27) and (2.28) when

$$\eta_k<|t-1|k_1/(t+3). \tag{2.34}$$

To establish the second statement, take $\varepsilon=|t-1|=(5/k_1)\eta_k$; then,

$$|t_k^*-1|<(5/k_1)\eta_k. \tag{2.35}$$

From (2.29) and the definition of S_k we obtain

$$\begin{aligned}\eta_k&=\max_{y\in S_k}\|g'(y)-g'(u^*)\| \\ &\leqq\max_{y\in S_k}L\|y-u^*\|\leqq L(2k_1^{-1}\|g(u^k)\|+\|u^k-u^*\|) \\ &\leqq K_3\|u^k-u^*\|,\end{aligned} \tag{2.36}$$

where $K_3=L(2(k_2/k_1)+1)$, noting that $\|g(u^k)\|=\|g(u^k)-g(u^*)\|\leqq k_2\|u-u^*\|$. Thus for all k with $S_k\subset S$, (2.35) and (2.36) establish (2.26) with $K_2=5K_3/k_1$. □

COROLLARY 1. *Let $g(u)$, u^k, and x be as above and $t_k=t_k^*$. Then $u^k\to u^*$ quadratically.*

Proof. Since $u^k\in S$ of Theorem 1 eventually, there is no loss in assuming (2.25) and (2.26) for all u^k. Let $B_k=(1/t_k^*)g'(u^k)$; then $u^{k+1}=u^k-B_k^{-1}g(u^k)$. We show

$$\|B_k-g'(u^*)\|\leqq K_4\|u^k-u^*\|, \tag{2.37}$$

which establishes quadratic convergence; see [4, pp. 52–53]. Recall from Proposition 1 that $t_k^*>\gamma$; thus using (2.26) we get

$$\begin{aligned}\|B_k-g'(u^*)\|&=(1/t_k^*)\|(g'(u^k)-g'(u^*))+(1-t_k^*)g'(u^*)\| \\ &\leqq\gamma^{-1}(L\|u^k-u^*\|+K_2\|u^k-u^*\|\,\|g'(u^*)\|) \\ &\leqq K_4\|u^k-u^*\|\end{aligned}$$

for $K_4=\gamma^{-1}(L+K_2\|g'(u^*)\|)$. □

While the choice of $t_k=t_k^*$ provides quadratic convergence, it requires an approximate solution of the single nonlinear equation

$$h(t)=x^Tg(u^k+tx)=0. \tag{2.38}$$

For example, if a single step of Newton's method for (2.37) is used starting at t_{k-1}, we obtain a sequence

$$t_k=t_{k-1}-x^Tg(u^k+t_{k-1}x)/(x^Tg'(u^k+t_{k-1}x)x). \tag{2.39}$$

However such a method requires two function evaluations of g and g' per iteration and may not ensure descent initially. Eventually, since $t_k=1$ is a good step, it may be preferable to determine when "eventually" holds. Taking $t_k=1$ when $V'(1)<0$ is satisfactory, but this test will fail if t_k^* approaches 1 from below. Also for small k, t_k^* could be much larger than 1, indicating an "over-damped" step as in the following

Example. Let $g(u) = Au + 2\sinh u$ where A is $n \times n$ with $\|A\| = 1$. Choose $u^0 = 50 \cdot 1$ where $\cdot 1 = (1, 1, \cdots, 1)^T$. Then $g(50 \cdot 1) \doteq e^{50} \cdot 1$ and $g'(50 \cdot 1) = \operatorname{diag}(e^{50}, \cdots, e^{50}) + A$ so $[g'(50 \cdot 1)]^{-1} \doteq \operatorname{diag}(e^{-50}, \cdots e^{-50})$. Thus x of (3.7) satisfies $x \doteq -(\cdot 1)$, and clearly it is important to overdamp. Note also that $V'(u^0 + tx)$, $t = 1$, is $V'(1) \doteq -ne^{49}$ and that if $t_{-1} = 1$ in (2.39), $t_0 \doteq 2$.

We now describe a step choice strategy, using condition (2.27) as a test, which has several desirable features including quadratic convergence of the u^k. The idea is to allow a step $t_k = 1$ if $V'(1) \leqq 0$ and $V'(2) > 0$ or if $V'(1) > 0$ is not too large.

Let $a \in (\frac{1}{2}, 1)$; given u^k and x consider the following algorithm for selecting t_k.

Algorithm step

(1) if $V'(1) \leqq 0$
(2) then search (1);
(3) else if $V'(1) + V'(a) < 0$
(4) then return (1)
(5) else search (a).

Algorithm search (t)

(6) if $t < 1$
(7) then while $V'(t) > 0$ $t \leftarrow t/2$
(8) else while $V'(2t) \leqq 0$ $t \leftarrow 2t$
(9) return (t).

The step t_k is the argument of the function return. If (2) or (5) execute, the algorithm *search* produces a $t_k \in [1, \infty)$ or $(0, a]$, respectively, satisfying

$$t_k^*/2 < t_k \leqq t_k^*. \tag{2.40}$$

This particularly simple search strategy could be replaced by a number of variants; for example, the factors $\frac{1}{2}$ in (7) and 2 in (8) could be replaced by factors $\alpha < 1$ and $\beta > 1$ respectively. Alternatively, (7) and (8) could incorporate a more sophisticated root solving routine. In the application presented in § 4, we have found it advantageous to increase or decrease t in algorithm *search* exponentially, and to begin the search in the vicinity of the previously chosen t_k.

The interesting case in algorithm *step* is when (3) and (4) execute. In this case $V'(1) > 0$ and $V'(a) < 0$ with $V'(1) + V'(a) < 0$. Then

$$\begin{aligned} V(1) - V(2a-1) &= \int_{2a-1}^{a} V'(s)\,ds + \int_{a}^{1} V'(s)\,ds \\ &< (1-a)(V'(a) + (1)) < 0, \end{aligned} \tag{2.41}$$

since V' is strictly increasing. Thus $V(1) < V(2a-1)$, and

$$0 < 2a - 1 < a < t_k^* < 1; \tag{2.42}$$

taking $t_k = 1$ still forces descent. Furthermore for any $\alpha \in (0, \min(2a-1, \gamma/2))$, each t_k produced by algorithm *step* has $t_k \in I_k$ and convergence of u_k to u^* is assured by Proposition 1. This conclusion is generally valid for any algorithm *search* returning a t_k with $V'(t_k) \leqq 0$ such that $t_k > f(\gamma) > 0$ and f is independent of k. We may then choose any $\alpha \in (0, \min(2a-1), f(\gamma))$ for Proposition 1. Finally, note that the proof of Theorem 1 shows that for all k sufficiently large (2.27) holds with $t = a$ and hence algorithm *step* returns $t_k = 1$ in line (4) or in line (9). We summarize in

COROLLARY 2. *Let $g(u)$, u^k, and x be as in Corollary 1 with t_k determined by algorithm step (with, perhaps, a fancier search as discussed). Then $u^k \to u^*$ quadratically.*

Asymptotically the cost of a careful implementation of algorithm *step* over the straightforward implementation of Newton's method is an unnecessary evaluation of $V'(a)$ or $V'(2)$, since $g(u+x)$ of $V'(1)$ must be computed any time $t_k = 1$ before convergence is attained. If the additional array to hold $g(u+tx)$ for $t=a$ or $t=2$ is unavailable, $g(u+x)$ can be recomputed, thus requiring a total of three g evaluations.

The additional g evaluation(s) as $k \to \infty$ motivated our use of the Newton-descent algorithm as a subiteration in the norm-decreasing outer iteration (2.21). This is easily implemented by inserting a line at the beginning of algorithm *step*:

(0) if $\|g(u^k+x)\| < \alpha_{k'}\|g(u^{k'})\|$ return (1).

Here the indices k and k' count the inner (descent) iterations and outer iterations, respectively. Asymptotically, since $t_k = 1$ produces quadratic convergence, we have

$$\|g(u^{k+1})\| \leqq \beta(k_2/k_1)\|g(u^k)\|^2, \qquad k > K_1, \tag{2.43}$$

using (2.19), (2.24) and $\|g(u^{k+1})\| \leqq k_2\|u^{k+1}-u^*\|$. Thus, for k sufficiently large, $t_k = 1$ will be returned in line (0) for any sequence $1 > \alpha_{k'} > \beta(k_2/k_1)\|g(u^{k'})\|$. We have found that choosing $\alpha_{k'}$ constant for all k' is effective.

3. Uniformly monotone mappings. In this section we examine a general uniformly monotone mapping $g(u)$ which is continuously differentiable on R^n and hence satisfies

$$x^T g'(u)x \geqq k_1 x^T x, \qquad x, u \in R^n. \tag{3.1}$$

Since $g'(u)$ is not assumed to be symmetric no $V(u)$ as in § 2 exists. The method we will discuss can be motivated by considering the autonomous system of ODE's,

$$\begin{aligned} &du/dt + g(u(t)) = 0, \\ &u(0) = u^0, \end{aligned} \tag{3.2}$$

which is integrated numerically from u^0 to the steady state $u^* = \lim_{t\to\infty} u(t)$ with $g(u^*) = 0$. Such methods are known as continuation or embedding (homotopy) methods, and are useful analytically as well as algorithmically; see [7] and [10, §§ 5.3, 7.5, 10.4]. However, following the curve $u(t)$ can lead to initial slow convergence, and our analysis establishing quadratic convergence will abandon the ODE (3.2).

Suppose (3.2) is discretized by the standard backward Euler implicit formula

$$(v^k - v^{k-1})/s_{k-1} + g(v^k) = 0, \tag{3.3}$$

with v^0 given. Here $s_k = t_{k+1} - t_k$ and v^k represents an approximation to $u(t_k)$. Equation (3.2) is a nonlinear equation for v^k; formally, one step of Newton's method applied to (3.3) (or, equivalently, replacing $g(v^k) \leftarrow g(v^{k-1}) + g'(v^{k-1})(v^k - v^{k-1})$) leads to

$$M_{k-1}(v^k - v^{k-1}) = -g(v^{k-1}), \tag{3.4}$$

where

$$M_{k-1} = I/s_{k-1} + g'(v^{k-1}),$$

The iteration (3.4) may be viewed in two ways. Given v^0 and $g(v)$ as in (3.1), (3.4) is a Newton-like iteration which is well defined at each step. Taking this point of view we will investigate the convergence of the iterates v^k to the steady state u^* with $g(u^*) = 0$.

The other view considers the iteration (3.4) as approximately integrating the ODE (3.2). In this view we will investigate the errors $\varepsilon_k = v^k - u(t_k)$, where $u(t)$ satisfies (3.2). If such errors can be kept small throughout the integration, then the v^k will eventually be close to u^*. In contrast the analysis of the discrete iteration (3.4) will show that it may

not be necessary to follow the trajectory $u(t)$ very accurately to insure rapid convergence of $v^k \to u^*$.

Let D be the convex compact set

$$D = \{u \mid \|u - u^*\| \leqq (2\|g(v^0)\|)/k_1\}, \tag{3.5}$$

for given v^0 and k_1 as in (3.1). In addition to (3.1), we assume

$$\|g'(u) - g'(v)\| \leqq L\|u - v\|, \qquad u, v \in D. \tag{3.6}$$

We will show that there is a sequence of parameters $s_k > 0$ such that the iterates v^k of (3.4) have $\|g(v^k)\| < \|g(v^{k-1})\|$ and $\|g(v^k)\| \to 0$. Then, since

$$\|M_{k-1}^{-1}\| \leqq s_{k-1}/(1 + k_1 s_{k-1}) < 1/k_1, \tag{3.7}$$

by (3.1) and the Cauchy-Schwarz inequality, $\|v^k - v^{k-1}\| \leqq (1/k_1)\|g(v^{k-1})\|$, and hence all $v^k \in D$. In addition $v^k \to u^*$ by (2.19) (which did not require symmetric g').

We now determine a recurrence relation for the $\|g_k\|$, abbreviating arguments as in $g_k = g(v^k)$. Expanding $g(v^k)$ gives

$$g_k = g_{k-1} + g'_{k-1}(v^k - v^{k-1}) + \int_0^1 [g'(v^{k-1} + t(v^k - v^{k-1})) - g'(v^{k-1})](v^k - v^{k-1})\, dt$$

$$= (1/s_{k-1})M_{k-1}^{-1} g_{k-1} + \int_0^1 [g'(v^{k-1} + t(v^k - v^{k-1})) - g'(v^{k-1})](v^k - v^{k-1})\, dt,$$

by using (3.4) and rearranging. Equations (3.4) and (3.7) imply

$$\|g_k\| \leqq (1/(1 + k_1 s_{k-1}))\|g_{k-1}\| + (L/2)\|v^k - v^{k-1}\|^2$$

$$\leqq (1/(1 + k_1 s_{k-1}))\|g_{k-1}\| + (L/2)(s_{k-1}/(1 + k_1 s_{k-1}))^2 \|g_{k-1}\|^2,$$

or

$$\|g_k\| \leqq (1/(1 + k_1 s_{k-1}))\|g_{k-1}\|[1 + ((L/2)s_{k-1}^2/(1 + k_1 s_{k-1}))\|g_{k-1}\|]. \tag{3.8}$$

Define $\gamma_k = s_k\|g_k\|$; then from (3.8) we obtain

$$\gamma_k \leqq (s_k/s_{k-1})(1 + k_1 s_{k-1})^{-2}\gamma_{k-1}(1 + k_1 s_{k-1} + (L/2)s_{k-1}\gamma_{k-1}). \tag{3.9}$$

Suppose the sequence s_k is chosen from s_0 such that $\gamma_k = \delta$ for all k (e.g., $s_1 = s_0\|g_0\|/\|g_1\|$). Then, rewriting (3.9),

$$(s_k/s_{k-1}) \geqq (1 + k_1 s_{k-1})^2/(1 + k_1 s_{k-1} + (L/2)\delta s_{k-1}). \tag{3.10}$$

We wish to choose s_0 such that the sequence s_k increases sufficiently such that $\|g_k\| = \delta/s_k \to 0$. Certainly (3.10) implies that

$$(s_k/s_{k-1}) > (1 + \alpha k_1 s_{k-1}), \qquad \alpha \in (0, 1) \tag{3.11}$$

if

$$((1 + k_1 s_{k-1})^2/(1 + k_1 s_{k-1} + (L/2)\delta s_{k-1})) > (1 + \alpha k_1 s_{k-1}). \tag{3.12}$$

Equation (3.12) implies a restriction on δ; namely

$$\delta < (2k_1/L)(1 - \alpha)((1 + k_1 s_{k-1})/(1 + \alpha k_1 s_{k-1})). \tag{3.13}$$

For example, taking $\alpha = \frac{1}{2}$, (3.13) and hence (3.11) holds if

$$\delta \leqq k_1/L, \tag{3.14}$$

or

$$s_0 \leqq k_1/(L\|g_0\|). \tag{3.15}$$

Summarizing, we obtain

THEOREM 2. *Let g satisfy* (3.1) *and* (3.6) *and the iterates v^k be defined by* (3.4) *where the s_k are chosen such that*

$$\delta = s_k\|g_k\| \quad \textit{for all } k \geqq 0. \tag{3.16}$$

If the sequence v^k converges to u^* it converges quadratically. Furthermore, given v^0 and s_0 as in (3.15), *the resulting sequence $v^k \to u^*$.*

Proof. For the first statement, note that there is no loss in assuming that the converging sequence v^k has $v^k \in D$ for all $k \geqq 0$. Then (3.8) and (3.16) imply

$$\|g_k\| \leqq K_5\|g_{k-1}\|^2, \quad \text{where } K_5 = (1/k_1\delta) + (L/2k_1^2). \tag{3.17}$$

This inequality and (2.19) show

$$\|v^k - u^*\| \leqq K_5(k_2^2/k_1)\|v^{k-1} - u^*\|, \tag{3.18}$$

where $k_2 = \max_{u\in D}\|g(u)\|$. The second statement follows immediately from (3.11) (with $\alpha = \frac{1}{2}$) which implies $\|g_k\| \to 0$ monotonically. □

The existential element of Theorem 2 regarding s_0 differs from the analysis of the previous section. We pause to make several remarks.

Theorem 2 implies the existence of an s_0 such that

$$\|g_k\| \leqq \alpha_{k-1}\|g_{k-1}\| < \|g_{k-1}\|, \tag{3.19}$$

where $\alpha_0 < 1$ and $\alpha_k \leqq \alpha_{k-1}$ for $k \geqq 1$. Algorithmically, we choose some $\delta > 0$ ($\delta = 1$, say); if (3.19) fails for some k, that v^k is rejected, δ is reduced, and the iteration is restarted at v^{k-1}. While this strategy ultimately produces (quadratic) convergence, initial convergence can be slow. Note that K_5 of (3.17) is inversely proportional to δ. This suggests making δ as large as possible while still ensuring convergence. We have found it useful to increase δ if the α_k of (3.19) are not decreasing fast enough. Care must be taken, however, to avoid thrashing of the changing δ's. (See our discussion in § 4).

In finite element discretizations it is convenient to consider the iteration

$$(B/s_{k-1} + g'_{k-1})(v^k - v^{k-1}) = -g(v^{k-1}), \tag{3.20}$$

where B is symmetric and positive definite, in place of (3.4). Let $h(w) = B^{-1/2}g(v)$, $v = B^{-1/2}w$ and $w^0 = B^{1/2}v^0$. Then (3.20) is equivalent to

$$(I/s_{k-1} + h'_{k-1})(w^k - w^{k-1}) = -h(w^{k-1}), \tag{3.21}$$

and $h'(w) = B^{-1/2}g'(v)B^{-1/2}$.

Suppose $g'(u)$ in Theorem 2 is also symmetric. Then there is a $V(u)$ as in § 2. Equations (2.5) and (2.30) with M_k as in (3.4) imply

$$V'(1) = -x^T M_k x + \int_0^1 x^T g'(w(t))x\, dt. \tag{3.22}$$

Applying (3.6) and (3.7) we see

$$V'(1) \leqq (x^T x/s_k)((L\delta/2k_1) - 1) \leqq 0 \tag{3.23}$$

if $\delta \leqq 2k_1/L$, which is less restrictive than (3.14). Hence the condition on s_0 implying convergence in Theorem 2 implies the iteration (3.4) is a descent method; c.f. [8, 14.4.6].

As a final remark, note that if, in our discussion of (3.8)–(3.15), g_0 satisfies the Kantorovich-like condition

$$\|g_0\| < 2\alpha k_1^2/L, \qquad \alpha \in (0, 1),$$

then using (3.8) in the limit as $s_k \to +\infty$ gives $\|g_k\| \leqq (L/2k_1^2)\|g_{k-1}\|^2$, and

$$\|g_k\| < (2k_1^2/L)\alpha^{2^k}. \tag{3.24}$$

This means that Newton's method ($s_k = +\infty$) converges to u^* for all $v^0 \in D$.

We now examine the extent to which the iteration (3.4) with $v_0 = u(0)$ follows the solution $u(t)$ of (3.2) with $u(0) = u^0$. Some useful properties of (4.1) are summarized in the following.

PROPOSITION 2. *Let $u(t)$ and $v(t)$ satisfy* (3.2) *with $g(u)$ as in* (3.1). *Then*

$$\|g(u(t))\| \leqq e^{-k_1 t}\|g(u^0)\|, \tag{3.25}$$

$$\|u(t) - u^*\| \leqq (\|g(u^0)\|/k_1)e^{-k_1 t}, \tag{3.26}$$

$$\|u(t) - v(t)\| \leqq \|u(0) - v(0)\|e^{-k_1 t}. \tag{3.27}$$

Let $\{t_i\}$ be a set of mesh points with $t_0 = 0$ and $s_k = t_{k+1} - t_k$. Consider the ODE's defined by

$$u_0'(t) = -g(u_0(t)), \qquad u_0(0) = u^0, \tag{3.28a}$$

and for $k > 0$,

$$u_k'(t) = -g(u_k(t)), \qquad t \geqq t_k, \tag{3.28b}$$

$$u_k(t_k) = u_{k-1}(t_k) + \varepsilon_k. \tag{3.28c}$$

Then

$$\begin{aligned}\|u_k(t_k) - u_0(t_k)\| &= \Big\| \sum_{j=1}^{k} (u_j(t_k) - u_{j-1}(t_k)) \Big\| \\ &\leqq \sum_{j=1}^{k} \|u_j(t_k) - u_{j-1}(t_k)\| \leqq \sum_{j=1}^{k} \|\varepsilon_j\| e^{-k_1(t_k - t_j)} \\ &\leqq \|\varepsilon_k\| + e^{-k_1 t_k} \sum_{j=1}^{k-1} (\|\varepsilon_j\|/s_j) e^{k_1 t_j} s_j \\ &\leqq \|\varepsilon_k\| + \max_{1 \leqq j \leqq k-1} (\|\varepsilon_j\|/s_j) k_1^{-1},\end{aligned}$$

from (3.27), and

$$\sum_{j=1}^{k-1} e^{k_1 t_j} s_j \leqq \int_{-\infty}^{t_k} e^{k_1 s}\, ds = k_1^{-1} e^{k_1 t_k}.$$

Given the iteration (3.4) with $v^0 = u_0(t)$ we identify the v^k with $u_k(t_k)$; hence, in view of (3.28b) and (3.28c), ε_k measures the extent to which v_k fails to satisfy the ODE $u_{k-1}' = -g(u_{k-1})$ at $t = t_k$; i.e., ε_k is a local truncation error. Summarizing (3.28c) and the above calculation we have

$$\|v^k - u_0(t_k)\| \leqq \|\varepsilon_k\| + \max_{1 \leqq j \leqq k-1} (\|\varepsilon_j\|/s_j) k_1^{-1}, \tag{3.29a}$$

where

$$\varepsilon_k = v^k - u_{k-1}(t_k). \tag{3.29b}$$

In order to bound the right-hand side of (3.29a) we rewrite (3.4) using (3.29b) to obtain

$$M_{k-1}(\varepsilon_k + (u_{k-1}(t_k) - u_{k-1}(t_{k-1}))) = -g(v^{k-1}),$$

since $v^{k-1} = u_{k-1}(t_{k-1})$. Thus,

$$\begin{aligned} M_{k-1}\varepsilon_k &= -(I/s_{k-1} + g'(v^{k-1}))(u_{k-1}(t_k) - u_{k-1}(t_{k-1})) - g(v^{k-1}) \\ &= -u'_{k-1}(t_{k-1}) - (s_{k-1}/2)u''_{k-1}(\theta_k) \\ &\quad - g'(v^{k-1})s_{k-1}u'_{k-1}(\eta_k) - g(v^{k-1}) \\ &= -(s_{k-1}/2)g'(u_{k-1}(\theta_k))g(u_{k-1}(\theta_k)) \\ &\quad + s_{k-1}g'(v^{k-1})g(u_{k-1}(\eta_k)). \end{aligned}$$

where θ_k, $\eta_k \in (t_{k-1}, t_k)$, with standard Taylor expansions. Let

$$k_2 = \max_{u \in D} \|g'(u)\|, \tag{3.30}$$

with D as in (3.5). Then

$$\|\varepsilon_k\| \leqq (\tfrac{3}{2})k_2(1 + s_{k-1}k_1)^{-1}s_{k-1}^2\|g(v^{k-1})\|, \tag{3.31}$$

since $\max\{\|g(u_{k-1}(\theta_k)\|, \|g(u_{k-1}(\eta_k)\|\} \leqq g\|(v^{k-1})\|$ by (3.25). If, as before, the s_k are defined such that

$$s_k\|g(v^k)\| = \delta \quad \text{for all } k \geqq 0, \tag{3.32}$$

and

$$s_{k-1}/s_k < 1, \tag{3.33}$$

then

$$\|\varepsilon_k\| \leqq (\tfrac{3}{2})(k_2/k_1)\delta \tag{3.34a}$$

and

$$\max_{1 \leqq j \leqq k-1} (\|\varepsilon_j\|/s_j)k_1^{-1} \leqq (\tfrac{3}{2})(k_2/k_1)\delta, \tag{3.34b}$$

using (3.31) with $k = j$. Finally, (3.29a) implies

$$\|v^k - u_0(t_k)\| \leqq 3(k_2/k_1)\delta, \tag{3.35}$$

and we have shown

PROPOSITION 3. *Let g and v^k be as in Theorem 2 and let $u_0(t)$ satisfy* (3.28a). *Then there is an $s_0 > 0$ such that* (3.35) *holds for s_k chosen as in* (3.32).

Note that condition (3.33) is weaker than (3.11) and leads to a slightly weaker restriction on s_0 than (3.14). Combining (3.35) and (3.26) shows that

$$\|v^k - u^*\| \leqq (3k_2s_0 + e^{-k_1t_k})(\|g(v^0)\|/k_1), \tag{3.36}$$

which indicates that the iterates v^k, while following the curve $u_0(t)$, can be made arbitrarily close to u^* as $k \to \infty$ by choosing s_0 sufficiently small. However (3.36) grossly misrepresents the convergence of $v^k \to u^*$, mainly since the use of the triangle

inequality is not sharp here. Indeed, in view of the convergence statement in Theorem 2, it is preferable to combine (3.35) and $\|v^k - u^*\| \leqq \|g(v^k)\|/k_1$, and sharpen (3.35) as

$$\|v^k - u_0(t_k)\| \leqq \min\{3(k_2/k_1)\delta,\ 1/k_1(\|g_k\| + e^{-k_1 t_k}\|g_0\|)\}. \tag{3.37}$$

4. Numerical experiments. In this section we discuss some computations illustrating the behavior of the algorithms described in §§ 2 and 3. We consider the PDE (1.7) with v and w fixed. The "t-method" of § 2 and the "s-method" of § 3 were implemented to be compatible with a code using piecewise linear finite elements and a multi-level iterative scheme for solving the resulting linear system ([1], [3]).

For both methods our convergence criteria were

$$(\|u^k - u^{k-1}\|/\|u^k\|) \leqq \mathrm{TOL} \tag{4.1a}$$

and

$$(\|u^k - u^{k-1}\|/\|u^{k-1} - u^{k-2}\|) \leqq \tfrac{1}{2}. \tag{4.1b}$$

Note that (4.1a) alone may be inadequate in the early states of the computation since (4.1a) can be satisfied for sufficiently small s_k or t_k. In our computations function and Jacobian evaluations and related assembly processing required over half of the computing time. The details of our implementations have thus been aimed at minimizing function and Jacobian evaluations.

The t-method was implemented along the lines suggested in § 2. As in line (0) of algorithm *step*, $t_k = 1$ is used if

$$\|g(u^k + x)\| \leqq .5\|g(u^k)\|. \tag{4.2}$$

When (4.2) is satisfied only one function is required for iteration k. We chose $a = \frac{3}{4}$ in line (3). In algorithm *search* we used the conveniently implemented sequences

$$t_k^j = (\tfrac{3}{4})^j(\tfrac{1}{4})^{(j^2-j)/2}, \qquad j \geqq 1, \tag{4.3a}$$

in line (7), and

$$t_k^j = (\tfrac{5}{4})^j(4)^{(j^2-j)/2}, \qquad j \geqq 1, \tag{4.3b}$$

in line (8). When $t_{k-1} < .1$ (or $t_{k-1} > 10$), the search begins at $10t_{k-1}$ (or $t_{k-1}/10$) rather than at t_k^1 of (4.3); t_k^j of (4.3) for $j \geqq 2$ are modified appropriately.

For our implementation of the s-method we chose the initial $\delta = 1$. If in any iteration

$$\|g(v^{k+1})\| \geqq \|g(v^k)\|, \tag{4.4}$$

v^{k+1} is rejected, δ is reduced by a factor of .1, and the iteration is restarted at v^k. The s-method is inefficient if δ is unnecessarily small. Thus, if the decreasing sequence $\|g_k\|$ does decrease "fast enough," we increase δ. In particular, if after the lth ($l \geqq 5$) consecutive accepted iterate

$$s_l \leqq 2^{l-4} s_{l-1}, \tag{4.5}$$

then δ is increased by a factor of 10.

Our example is a p-n junction; the equation has the form (1.7). The domain and boundary conditions are shown in Fig. 2.

The doping profile, $k(x, y)$, and the solution, $u(x, y)$, change over several orders of magnitude in a very small region near the junction. There is also a notable singularity due to the change in boundary conditions along the upper boundary. A relatively

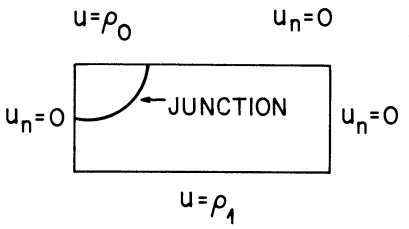

FIG. 2. *p-n junction problem with boundary conditions.*

inaccurate $u^0(x, y)$ was chosen to be piecewise constant reflecting the values ρ_0 and ρ_1 and their locations.

The *p-n* junction problem was first solved on a relatively coarse nonuniform mesh with $n = 50$. The comparative results for the *t*-method and the *s*-method are summarized in Tables I and II, respectively. The *t*-method used 12 g' evaluations and 22 g evaluations while the *s*-method required 24 of each. Superlinear convergence is evident in the final 3-5 iterations, and the norm-decreasing character of both methods is apparent. In the *s*-method an attempted increase in δ after iteration 13 caused rejection in iteration 14.

The finite element code allows the option of adaptive mesh refinement and a multilevel iteration scheme. We used this feature starting from the same initial mesh with $n = 50$. The results are summarized in Table III.

Except for $n = 50$, $t_k = 1$ was used for all iterations in the *t*-method, and there were no rejections in the *s*-method.

The adaptive refinement procedure is designed to refine the mesh in those regions where the error is relatively large. Note that more iterations are used in solving the nonlinear equations on the coarsest mesh, since good initial choices for u^0 and s_0 are subsequently available. It would be advantageous to relax the fixed tolerance of 10^{-6} in the finer meshes and use only 1–2 iterations there.

Acknowledgments. The authors are grateful to J. McKenna for suggesting the semiconductor modeling equations, and to J. Dennis for a thoughtful critique of an earlier draft of the paper.

TABLE I
t-method, TOL = 10^{-6}, $\|g_0\| = 1.1 \cdot 10^7$; evals ≡ *evaluations of* $g(u)$

k	t_k	$\|u^k - u^{k-1}\|/\|u^k\|$	$\|g_k\|$	evals
1	6.6 (−3)	3.7 (1)	1.13 (7)	4
2	9.3 (−3)	3.4 (1)	1.12 (7)	3
3	1.3 (−2)	1.9 (1)	1.11 (7)	3
4	1.8 (−2)	6.0	1.08 (7)	3
5	1.4 (−1)	1.9	9.5 (6)	2
6	1	5.5 (−1)	3.1 (6)	1
7	1	9.0 (−2)	6.6 (5)	1
8	1	5.5 (−2)	8.0 (4)	1
9	1	9.0 (−3)	3.6 (3)	1
10	1	3.7 (−4)	1.1 (1)	1
11	1	1.1 (−6)	9.7 (−5)	1
12	1	1.6 (−11)	—	—

TABLE II
s-method, TOL = 10^{-6}, $\|g_0\| = 1.1 \cdot 10^7$

k	δ	$\|u^k - u^{k-1}\|/\|u^k\|$	$\|g_k\|$
1	1	7.6 (−2)	3.4 (6)
2	1	2.0 (−1)	7.5 (5)
3	1	6.9 (−1)	3.4 (5)
4	1	4.3 (−1)	2.3 (5)
5	1	3.0 (−1)	3.8 (4)
6		rejected	
7	.1	1.4 (−1)	9.7 (3)
8		rejected	
9	.01	5.4 (−2)	6.7 (3)
10	.01	6.5 (−2)	6.2 (3)
11	.01	6.0 (−2)	5.7 (3)
12	.01	5.5 (−2)	4.9 (3)
13	.01	4.7 (−2)	3.9 (3)
		$l = 5$ in (4.5), $\delta \leftarrow 10\delta$;	
14		$\delta = .1$ causes rejection	
15	.01	3.7 (−2)	2.8 (3)
16	.01	2.4 (−2)	1.8 (3)
17	.01	1.5 (−2)	1.1 (3)
18	.01	1.3 (−2)	5.7 (2)
19	.01	1.1 (−2)	2.3 (2)
20	.01	6.5 (−3)	5.5 (1)
21	.01	2.1 (−3)	4.1
22	.01	1.7 (−4)	2.5 (−2)
23	.01	1.1 (−6)	9.9 (−7)
24	.01	4.2 (−11)	—

TABLE III
Use of adaptive refinement, TOL = 10^{-6}.

n	t-method iterations	s-method iterations
50	12	24
57	5	7
80	4	5
95	7	7
173	6	6
179	4	4

REFERENCES

[1] R. E. BANK AND T. DUPONT, *An optimal order process for solving elliptic finite element equations*, Math. Comp., to appear.

[2] R. E. BANK AND D. J. ROSE, *Solving Nonlinear Systems of Equations Arising from Semiconductor Device Modelling*, unpublished internal memorandum, 1980.

[3] R. E. BANK AND A. H. SHERMAN, PLTMG *Users' Guide*, CNA 152, Center for Numerical Analysis, University of Texas, 1979.

[4] J. E. DENNIS AND J. J. MORÉ, *Quasi-Newton methods, motivation and theory*. SIAM Rev., 19 (1977), pp. 46–89.

[5] M. N. YAKOVLEV, *Algorithms for minimizing strictly convex functionals*. Zh. Vychisl. Mat. i Mat. Fiz., 7 (1967), pp. 429–431.

[6] G. H. MEYER, *The numerical solution of quasilinear elliptic equations*, in Numerical Solution of Systems of Nonlinear Algebraic Equations, G. Byrne and C. Hall, eds., Academic Press, New York, 1973, pp. 27–61.

[7] ———, *On solving nonlinear equations with a one-parameter operator embedding*, this Journal, (1968), pp. 739–752.

[8] J. M. ORTEGA AND W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[9] T. A. PORSCHING, *On the origins and numerical solution of some sparse nonlinear systems*, in Sparse Matrix Computations, J. Bunch and D. Rose, eds., Academic Press, New York, 1976, pp. 439–453.

[10] W. C. RHEINBOLDT, *Methods for Solving Systems of Nonlinear Equations*, CBMS Regional Conference Series in Applied Mathematics No. 14, Society for Industrial and Applied Mathematics, Philadelphia PA, 1974.

[11] D. J. ROSE, *Symmetric Elimination on Sparse Positive Definite Systems and the Potential Flow Newwork Problem*, PhD thesis, Harvard University, Cambridge, MA, 1970.