JOURNAL OF INTEGRAL EQUATIONS
AND APPLICATIONS
Volume 3, Number 4, Fall 1991

MESH INDEPENDENCE OF NEWTON-LIKE METHODS FOR INFINITE DIMENSIONAL PROBLEMS

C.T. KELLEY AND E.W. SACHS

ABSTRACT. Globally convergent modifications of Newton's method, such as the Armijo rule, can be applied to infinite dimensional problems and their discretizations. We show that if the construction of the discretizations is done properly, then the convergence behavior of the iteration is the same for the discrete problems as it is for the infinite-dimensional problem. Basic to these results is the use of the concept of discrete convergence as a tool to measure the performance of algorithms and a new setting of Banach spaces with incomplete metrics, for example, norms generated by continuous inner products. The motivating problems are integral equations with continuous kernels. This result extends to the globally convergent case results of Allgower, Böhmer, Potra, and Rheinboldt, and the authors. In addition, we strengthen the previous results on mesh independence of quasi-Newton methods. Numerical results are reported that illustrate the results.

1. Introduction. Many methods for solution of nonlinear equations in $\mathbf{R}^n$ depend on inner product information. Quasi-Newton methods, such as Broyden's method [**6**] or the BFGS method [**7, 10, 11, 27**] use inner products to construct approximations to Jacobian and Hessian matrices. Globally convergent modifications of Newton's method, such as line searches or trust region strategies, use inner product norms to test for sufficient decrease and for computation of gradients and steepest descent directions. When such methods are applied to discretizations of equations in Banach spaces these inner products impose a Hilbert space structure on the discretized problem that may not be appropriate for the infinite dimensional problem. Other, more general, global convergence methods may use merit functions that satisfy crucial

Key words and phrases. Quasi-Newton methods, Armijo rule, integral equations.
AMS(MOS) *Mathematics Subject Classification.* Primary 45G10, 65H10.
The research of the first author was supported by NSF Grant No. DMS-8601139, NSF Grant No. DMS-8900410, NSF Grant No. INT-8800560, AFOSR Grant Nos. AFOSR-ISSA-860074, AFOSR-ISSA-890044, and AFOSR Grant No. AFOSR-89-0124.
The research of the second author was supported by the Deutsche Forschungsgemeinschaft.

estimates in norms other than that of the Banach space in which the problem is posed. The questions that arise are how the performance of the iteration on the discrete problems is related to those of similar methods on the infinite dimensional problem and how the use of the incompatible norms affects the convergence rate of the iterates, especially as the discretization is refined.

In previous work ([**16, 15, 18, 19**, and **21**]) such questions were investigated for various Newton-like methods. In [**1**] and [**2**] the results were on Newton's method, where additional norms are not introduced by the method, but the issue of how the convergence properties of the infinite dimensional problem affect those of the discrete problem is still of interest. All these results are concerned with local convergence rates. The purpose of this paper is to show how a globalization strategy, the Armijo rule for line searches [**3**] performs in this context and to show how the estimates using possibly incompatible norms can be used to advantage for the discrete problems. These results are also new in the context of mesh-independence for local convergence of quasi-Newton methods. In particular, we are able to generalize the statement of the mesh-independence principle by incorporating more general merit functions and consider the case in which the spaces on which the various approximate problems are defined are not the same. The results are obtained by an extension of the work in [**19, 25**, and **26**]. Basic to this analysis is the notion of discrete convergence [**28**] and its application to merit functions for global convergence algorithms.

A special case of our results is the situation where the Banach space is endowed with a continuous inner product. The merit function is the inner product of the nonlinear residual with itself, and the resulting method is the classical Armijo rule.

The work in this paper, like that in [**19**], is motivated by nonlinear integral equations. The results apply, however, to pointwise methods such as those discussed in [**17, 18, 20**, and **22**]. Although our results are stated in more generality, we begin with an example of a nonlinear integral equation to motivate the results that follow. Let $\mathcal{K}$ denote the nonlinear operator on $C[0,1]$ defined by

$$\mathcal{K}(u)(x) = \int_0^1 k(x, y, u(y))\, dy \tag{1.1}$$

and consider the equation

$$\mathcal{F}(u)(x) = u(x) - \mathcal{K}(u)(x) = 0. \tag{1.2}$$

We assume that k and $\partial k/\partial u$ are continuous functions on $[0,1] \times [0,1] \times \mathbf{R}$ and Lipschitz continuous in the third argument. With these assumptions, $\mathcal{K}$ is a completely continuous map on $C[0,1]$. We assume that a solution, $u^* \in C[0,1]$, to (1.2) exists and that $\mathcal{F}'(u^*)$ is a nonsingular linear operator on $C[0,1]$. Note that for $u \in C[0,1]$, $\mathcal{F}'(u)$ is a bounded operator on $L^p[0,1]$ for $1 \le p \le \infty$. If $u \in C[0,1]$ the L^2 adjoint of $\mathcal{F}'(u)$, $\mathcal{F}'(u)^*$, is also a well-defined operator from $C[0,1]$ to $C[0,1]$ and is given by

$$\mathcal{F}'(u)^* v(x) = v(x) - \int_0^1 \frac{\partial k}{\partial u}(x,y,u(y))v(y)\,dy.$$

Hence, given $u \in C[0,1]$ sufficient near u^*, both the Newton step, $-\mathcal{F}'(u)^{-1}\mathcal{F}(u)$, and the L^2 steepest descent direction for the functional, $||\mathcal{F}(u)||_2^2/2$, $-\mathcal{F}'(u)^*\mathcal{F}(u)$, are defined. We will abstract this use of a Hilbert space steepest descent direction for a problem posed in a Banach space later in the paper.

We discretize the integral in (1.2) by a quadrature rule. We index the rules by $N = 1, 2, \ldots$, with increasing N denoting a more accurate quadrature rule. At level N the quadrature nodes will be denoted by $\{x_i^N\}_{i=1}^{m_N}$, where m_N is the number of nodes in the rule, and the quadrature weights will be denoted by $\{w_i^N\}_{i=1}^{m_N}$. We assume that, for all $u \in C[0,1]$,

$$\lim_{N\to\infty} \sum_{i=1}^{m_N} u(x_i^N)w_i^N = \int_0^1 u(x)\,dx \quad \text{and} \quad \sum_{i=1}^{m_N} w_i^N = 1. \tag{1.3}$$

For $u \in C[0,1]$ and $N \ge 1$ define $\mathcal{K}_N$ by

$$\mathcal{K}_N(u)(x) = \sum_{j=1}^{m_N} k(x, x_j^N, u(x_j^N))w_j^N. \tag{1.4}$$

By (1.3), $\mathcal{K}_N$ converges strongly to $\mathcal{K}$ in the sense that

$$\lim_{N\to\infty} \mathcal{K}_N(u) = \mathcal{K}(u),$$

in the norm of $C[0,1]$ for all $u \in C[0,1]$. We will approximate $\mathcal{F}$ by $\hat{\mathcal{F}}_N = I - \hat{\mathcal{K}}_N$. If $\hat{u}^N$ is a solution of the approximate equation,

$$\hat{\mathcal{F}}_N(u)(x) = u(x) - \hat{\mathcal{K}}_N(u)(x) = 0, \tag{1.5}$$

then the vector $u^N \in \mathbf{R}^{m_N}$ with components $u_{N,i} = \hat{u}^N(x_i^N)$ satisfies the finite dimensional system of nonlinear equations

$$u_{N,i} = \sum_{j=1}^{m_N} k(x_i^N, x_j^N, u_{N,j}) w_j^N = 0. \tag{1.6}$$

If we let $\Pi_N : C([0,1] \to \mathbf{R}^{m_N}$ be the map that takes $u \in C[0,1]$ into the vector with components $u(x_i^N)$ and let $P_N : \mathbf{R}^{m_N} \to C[0,1]$ be some interpolation operator, (1.6) may be written, noting that $\Pi_N P_N = I$ on $\mathbf{R}^{m_N}$, as

$$\mathcal{F}_N u^N = \Pi_N \hat{\mathcal{F}}_N P_N u^N = 0.$$

Note that if u^N is the solution to (1.6), the solution, $\hat{u}^N$, to (1.5) is

$$\hat{u}^N(x) = \sum_{j=1}^{m_N} k(x, x_j^N, u_{N,i}) w_j^N.$$

Hence, either (1.6) or (1.5) may be regarded as an approximation to (1.2).

The discussion above illustrates a theme of the paper. To have the quadrature rule approximation converge to the integral, we must require that the integrands be continuous. Hence, it is reasonable to pose our problem in the Banach space of continuous functions. However, in order to discuss global convergence and use the concepts of descent direction, it is natural to take gradients in a Hilbert space setting. The assumptions that we make in Section 2 are motivated by consideration of these issues in the specific case of (1.2).

If one wants to solve either (1.2) or (1.6), by, e.g., Newton's method, the rate of local convergence is given by

$$||e_+|| \leq \frac{M\gamma}{2} ||e_c||^2. \tag{1.7}$$

In (1.7) e_c and e_+ denote the errors for the old and new iterates, γ is the Lipschitz constant for $\partial k / \partial u$ with respect to its third argument, M

is a bound for $||\mathcal{F}'(u)^{-1}||$ in a sufficiently small neighborhood of u^*, and the norm, $||\cdot||$, is arbitrary. The choice of norm only affects the radius of the ball about the solution for which (1.7) holds. In particular, if a max norm is used (L^∞ for the infinite dimensional problem and ℓ^∞ for the finite dimensional problems), the radius of the ball for which (1.7) holds depends only on continuity properties of k and its derivatives, and not on whether the infinite dimensional problem or any of the finite dimensional problems are being considered.

The disadvantages of Newton's method are that for each iterate a full matrix must be computed, stored and factored. For fine meshes with many quadrature points the cost can be prohibitive. For this reason, methods which approximate the Jacobian are often preferred, even though more iterates may be required. One such method is described in [**4**] and [**5**]. The approach is to form an approximate Jacobian from the solution of a discrete problem on a coarser mesh. The coarse mesh problem can be solved with Newton's method at a cost that is negligible in comparison with a fine mesh evaluation of the nonlinear integral term. Storage and factorization of matrices need only be done for the coarse mesh problem. Convergence of this method, as was pointed out in [**16**], can be accelerated by use of Broyden's method and extended to some problems having singular Fréchet derivative at the solution. The use of Broyden's method, which requires a continuous inner product for its formulation, is one point where norms other than the L^∞ norm enter into the algorithms. This issue is also discussed in [**19**].

Following the notation in [**19**], we discuss algorithms in terms of solution of a nonlinear equation, $F(u) = 0$, on a Banach space, Z, with norm $||\cdot||_Z$. F could be any of $\mathcal{F}$ or $\mathcal{F}_N$, for example. We assume that a solution, u^*, exists and that $F'(u^*)$ is nonsingular. In order to discuss the quasi-Newton methods and some of the globally convergent algorithms, we will assume that Z has a continuous inner product, which we denote by $(\cdot,\cdot)$. We do not assume that Z is complete in the norm induced by the inner product. We let X denote the completion of Z in the inner product norm, and let $||\cdot||_X$ denote this norm. For example, in the max norm or $\mathbf{R}^{m_N}$ with ℓ^∞ norm, the inner product, the L^2 inner product or an ℓ^2 inner product with quadrature weights, and X could be L^2 or $\mathbf{R}^{m_N}$ with the ℓ^2 norm.

Quasi-Newton methods maintain not only an approximate solution, u, but an approximate Fréchet derivative, B. For local convergence, the algorithms make the transition from a current pair of approximations, (u_c, B_c) to $(u^*, F'(u^*))$, to a new pair, (u_+, B_+), through the following steps, assuming $F(u_c)$ has been completed.

- Solve $B_c s = -F(u_c)$.
- $u_+ = u_c + s$.
- Complete $F(u_+)$ and terminate if sufficiently small.
- Update B_c to form B_+.

The particular quasi-Newton method is determined by the update formula in the last step. The method we will consider in this paper is Broyden's method [**6**] where

$$B_+ = B_c + \frac{(y - B_c s) \otimes s}{||s||_X^2}. \tag{1.8}$$

In (1.8) $\otimes$ denotes the Hilbert space tensor product induced by $(\cdot, \cdot)$, and $y = F(u_+) - F(u_c)$. In finite dimension, if B_0 and u_0 are sufficiently near $F'(u^*)$ and u^* respectively, the Broyden iterates exist (i.e., B_n is always nonsingular) and converge q-superlinearly to u^* with respect to $|| \cdot ||_X$, then

$$\lim_{n \to \infty} \frac{||u_{n+1} - u^*||_X}{||u_n - u^*||_X} = 0. \tag{1.9}$$

In the case of an infinite dimensional space (1.9) also holds provided that, in addition to the assumptions needed in finite dimension, $B_0 - F'(u^*)$ is compact [**12**]. The problem that arises for problems like (1.2) is with the relation between the X norm convergence promised by the theory and the fact that the strong convergence of the approximations to our nonlinear function holds only in the topology of Z. This makes it difficult to describe the convergence behavior of the iterates in a way that is independent of the level of the discretization. This issue has been resolved in [**19, 21**] under certain assumptions. We will discuss some of those results in Section 2.

Similar issues related to inner products arise in global modifications of Newton-like iterative methods. In this paper we consider line search methods that seek sufficient decrease in a merit function, g, that is defined on Z and that is an increasing function of $||u - u^*||$. A typical merit function is $g(u) = (F(u), F(u))/2$, where $(\cdot,\cdot)$ is a continuous inner product on Z. As this particular example is very natural and we have already used the inner product structure in the description of Broyden's method, we consider that example first. We describe the transition from (u_c, B_c) to (u_+, B_+) as follows:

- Solve $B_c p = -F(u_c)$.
- If $(F'(u_c)^* F(u_c), p) > 0$, use a different B_c (e.g., $B_c = F'(u_c)$).
- Find σ such that $||F(u_c + \sigma p)||_X$ is sufficient less than $||F(u_c)||_X$; set $s = \sigma p$.
- $u_+ = u_c + s$.
- Compute $F(u_+)$ and terminate if sufficiently small.
- Update B_c to form B_+.

In the above algorithm, the first role of the inner product is to decide if p is a descent direction for $||F||_X$ in the second step. In that step the Hilbert space adjoint of $F'(u_c)$, which we denote by $F'(u_c)^*$ is used as well and we assume throughout this paper that $F'(u_c)$ is defined on Z. This assumption is certainly valid if F' is the sum of the identity with an integral operator with a continuous kernel. Note that the second step may be omitted if p is the Newton step for the equation $F(u) = 0$ or the Hilbert space steepest descent step $-F'(u_c)^* F(u_c)$ for minimizing $g(u) = ||F(u)||_X^2/2$. Moreover, the inner product, $(F'(u_c)^* F(u_c), p) > 0$, may be approximated with only one additional function evaluation as

$$(F(u_c), F(u_c + \delta p) - F(u_c))/\delta,$$

or as

$$(g(u_c + \delta p) - g(u_c))/\delta,$$

for a sufficiently small δ.

The third step looks for an appropriate step length. There are various methods for finding this step length. Our choice is the Armijo rule [**3**] which aims to decrease a merit function, g, by a sufficient amount. The choice of g in the algorithm described above is

$$g(u) = ||F(u)||_X^2/2. \tag{1.10}$$

The Armijo rule, as we formulate it in this paper, depends on parameters, $\mu \in (0,1)$, $\rho > 0$, and $q \in (0,1)$. Given a descent direction, p, first find α so that

$$\alpha > \frac{\rho g(u_c)}{||p||_X^2}, \tag{1.11}$$

then find the smallest integer, $j \geq 0$, such that

$$g(u_c) - g(u_c + \alpha q^j p) > \alpha \mu q^j g(u_c), \tag{1.12}$$

and then set

$$u_+ = u_c + \alpha q^j p. \tag{1.13}$$

For definiteness we will use

$$\alpha = \max\left\{1, \frac{1.1\rho g(u_c)}{||p||_X^2}\right\}$$

in this paper.

As was the case with quasi-Newton methods, the theorems for global convergence [**9, 13, 24**] assert convergence for the special case $g(u) = ||F(u)||_X^2/2$ in the norm of X. Under certain assumptions, the conclusions are that either $||F(u_n)||_X \to 0$ or $(F'(u_n)^* F(u_n), s_n)/||s_n||_X^2 \to 0$.

In this paper we consider Armijo-type line search schemes based on decreasing a merit function, g, by a sufficient amount. In [**25, 26**] merit functions of the form $\phi(F(u))$ where ϕ is convex, but not necessarily smooth, were considered. In our situation, where analysis of mesh-independent convergence properties is the goal, we restrict our attention to smooth merit functions, g, that satisfy certain assumptions. We give these assumptions in Section 2.

We begin Section 2 with a description of the setting in which we will work and review the concept of discrete convergence. We then give the assumptions on the nonlinearity and the merit functions that we need, motivating them with the particular example of inner product based merit functions and integral equations with continuous kernels on $Z = C[0,1]$. In Section 3 we state and prove our mesh independence results, and in Section 4 we consider a numerical example.

2. Notation and preliminaries. The notation in this paper is motivated by the example in Section 1. In this section we introduce notation and motivate the work in Section 3 by considering the special setting of the classical Armijo rule as described in Section 1. With this algorithm in mind, we summarize the properties of the merit function given by (1.10) as motivation for the more general setting discussed in Section 3.

Let Z be a Banach space. We are concerned with solution of a nonlinear equation, $F(u) = 0$ on Z. Our solution will be approximated by solution of approximate equations on different spaces. We use assumptions that are different from those in [**19**], but the conclusions of that paper will follow from the assumptions given here.

Let $\{Z_N\}_{N=1}^{\infty}$ be a sequence of Banach spaces. We approximate $F(u) = 0$ by $F_N(u) = 0$ on Z_N. For consistency of notation, we let $Z_\infty = Z$ and $F_\infty = F$. We assume that continuous linear maps, $\Pi_N : Z \to Z_N$ and $P_N : Z_N \to Z$, exist and are bounded uniformly in N. These maps will define the convergence properties of the approximation.

In order to formulate the fundamental assumptions on the approximates we use the concept of discrete convergence, cf. Stummel [**28**]. We assume that

$$\text{(2.1)} \qquad \begin{aligned} &\text{(a)} \quad ||\Pi_N u||_{Z_N} \to ||u||_Z, \quad \text{for all } u \in Z. \\ &\text{(b)} \quad P_N \Pi_N \to I, \quad \text{strongly in } Z. \end{aligned}$$

We say that a sequence, $\{u_N\}$, with $u_N \in Z_N$ converges in the discrete sense if

$$||u_N - \Pi_N u||_{Z_N} \to 0.$$

We denote this by

$$u_N \xrightarrow{d} u.$$

We say that a sequence of (possibly nonlinear) continuous operators, $\{A_N\}$, is discretely convergent to a continuous operator A on Z if

$$A_N(u_N) \xrightarrow{d} A(u)$$

whenever $u_N \xrightarrow{d} u$. We denote this by

$$A_N \xrightarrow{d} A.$$

Finally, we say that a sequence of functionals, $\{g_N\}$, with g_N defined on Z_N, converges discretely to g if $u_N \stackrel{d}{\to} u$ implies that $g_N(u_N) \to g(u)$. We denote this as

$$g_N \stackrel{d}{\longrightarrow} g.$$

The algorithms we consider use a sequence of approximations to F, $\{F_N\}$, defined on Z_N, and a sequence of approximations to a merit function g, $\{g_N\}$. For convenience, we identify g_∞ with g. We assume that for $1 \le N \le \infty$ there are norms $||\cdot||_{X_N}$ defined on Z_N and C_G, $r > 0$, such that there is $||\cdot||_{X_N} \le C_Z||\cdot||_{Z_N}$ for all N and, for all $u \in Z_N$,

$$g_N(u) \le C_G||F_N(u)||_{X_N}^r. \tag{2.2}$$

We let X_N denote the completion of Z_N with respect to the X_N norm. Consistently with the notation above, we let $X = X_\infty$. In the case where X_N is finite dimensional, X_N and Z_N differ only in the norms, but in the infinite dimensional setting X_N may properly contain Z_N. It is useful here to reconsider the problem (1.2) and the approximations discussed in Section 1. In this case one could use $Z = C[0,1]$ with the max norm; and $P_N\Pi_N \to I$ strongly in Z. To use the merit function, $||F||_2^2/2$, one would use $X = L^2$ and let the X_N norms be discrete L^2 norms. Note that $P_N\Pi_N \to I$ is not true in X, in fact, Π_N is not defined on $X = L^2$.

Our fundamental assumptions on the spaces X_N are that

$$\begin{aligned} &\text{(a)}\ \ ||\cdot||_{X_N} \stackrel{d}{\longrightarrow} ||\cdot||_X, \\ &\text{(b)}\ \ \text{there is } C_X > 0 \text{ such that, for all } N \text{ and } u \in Z_N, \\ &\qquad\qquad ||P_N u||_X \le C_X||u||_{X_N}. \end{aligned} \tag{2.3}$$

Assumption (2.3)(a) relates the norms, $||\cdot||_{X_N}$, to the topology of Z_N in a way that is crucial to the mesh independence results of this paper. In the model problem discussed in Section 1, Z_N is $\mathbf{R}^{m_N}$ with the ℓ_∞ norm and $||\cdot||_{X_N} = (\cdot,\cdot)_N$ is an approximation to the L^2 inner product based on a quadrature rule. In that case, (2.3)(a) clearly holds. However, (2.3)(a) would not hold if we used $Z = X = L^2[0,1]$ as point evaluation is not defined on L^2. If one wanted to consider the situation $Z = X$, the approximate integration rule that defines the inner product could not be based on point evaluations, but would have to be a projection method, and Π_N would be taken as a projection. Stummel [**28**] calls (2.1)(b) a consistency condition and the uniform boundedness condition, (2.3)(b), a stabilizing condition.

For the approximating mappings, we assume

$$\text{(2.4)}\qquad \begin{aligned} &\text{(a)}\quad g_N \xrightarrow{d} g,\\ &\text{(b)}\quad F_N \xrightarrow{d} F. \end{aligned}$$

In case of (1.10) and $g_N(u) = ||F_N(u)||^2_{X_N}/2$ the relation (2.4)(a) follows from (2.4)(b) and (2.3)(a).

We assume that a root, u^* of F exists in Z and that $F'(u^*)$ is nonsingular. Let

$$\text{(2.5)}\qquad \mathcal{T}_N(\beta) = \{u \in Z_N \mid ||F_N(u)||_{Z_N} \leq \beta\}$$

For purposes of simplicity, we will assume that $\mathcal{T}(\beta)$ is connected for all β sufficiently small. This means that u^* is the only root of F. In general, when F has several roots, our mesh independence results hold if we restrict attention to a connected component of $\mathcal{T}(\beta)$ and assume that all iterates remain in that component. We let

$$\mathcal{C}_N(\beta) = \overline{\text{co}\mathcal{T}_N(\beta)}.$$

We assume that the maps $\{F_N\}$ have uniform continuity and boundedness properties. In particular, we assume that there are β^Z, M_S^Z, M^Z, and γ_D^Z such that, for all N and $u, v \in \mathcal{C}_N(\beta^Z)$,

$$\text{(2.6)}\qquad \begin{aligned} &\text{(a)}\quad ||u||_{Z_N} \leq M_S^Z,\\ &\text{(b)}\quad ||F'_N(u) - F'_N(v)||_{\mathcal{L}(Z_N)} \leq \gamma_F^Z ||u - v||_{Z_N}, \text{ and}\\ &\text{(c)}\quad \sup_N (||F'_N(u)||_{\mathcal{L}(Z_N)}, ||(F'_N(u))^{-1}||_{\mathcal{L}(Z_N)}) \leq M^Z. \end{aligned}$$

The Kantorovich theorem and assumptions (2.1), (2.3), (2.4) and (2.6) allow us to conclude that for N sufficiently large, F_N has a root, $u^N \in Z_N$, and that $P_N u^N \to u^*$ in Z. In addition, there is $\gamma_Z^F < M^Z$ such that

$$||F_N(u) - F_N(v)||_{Z_N} \leq \gamma_F^Z ||u - v||_{Z_N},$$

for all $u, v \in \mathcal{C}_N(\beta^Z)$.

Note that if $u^N \xrightarrow{d} u$ then $P_N u_N \to u$ in the norm of Z by (2.1)(b). Hence, (2.6)(b) implies that

$$\lim_{N\to\infty} P_N F_N(\Pi_N u) = F(u).$$

In [**19**] all the spaces $\{Z_N\}$ were taken to be identical and the inner products, $(\cdot,\cdot)_N$, changed with N. The results in this paper are more general.

Our assumptions on the sequence of maps $\{F_N\}$ are essentially the uniform continuity and differentiability requirements for convergence of Newton's method in a mesh independent way in the topology of Z. These assumptions are different than those in [**19**] in that the functions $\{F_N\}$ are required to have uniform Lipschitz continuity properties and existence of the sequence of roots, $\{u^N\}$, need not be assumed. In order to consider inner product based algorithms, we must make assumptions similar to those in (2.6) using the X_N norms. These are, in fact, the critical assumptions for the theory in this paper as it applies to the classical Armijo rule.

We assume that there are $\beta > 0$, $\gamma_F > 0$, $\gamma_D > 0$, $M_S > 0$, and $M > 0$ such that the maps $\{F_N\}$ and the sets

$$\mathcal{S}_N(\beta) = \{u \in Z_N \mid ||F_N(u)||_{X_N} \le \beta\}, \tag{2.7}$$

$$\mathcal{D}_N(\beta) = \overline{\mathrm{co}\mathcal{S}_N(\beta)}, \tag{2.8}$$

satisfy, for all $N \ge 1$ and $u, v \in \mathcal{D}_N(\beta)$,

$$\begin{aligned} &\text{(a)}\quad ||u||_{X_N} \le M_S,\\ &\text{(b)}\quad ||F_N'(u) - F_N'(v)||_{\mathcal{L}(X_N)} \le \gamma_D ||u-v||_{X_N}, \text{ and}\\ &\text{(c)}\quad \sup_N(||F_N'(u)||_{\mathcal{L}(X_N)}, ||(F_N'(u))^{-1}||_{\mathcal{L}(X_N)}) \le M. \end{aligned} \tag{2.9}$$

In (2.8) the closure is taken in the topology of Z_N. As in (2.6), (2.9)(c) implies that there is $\gamma_F \le M$ such that, for $u, v \in \mathcal{D}_N(\beta)$,

$$||F_N(u) - F_N(v)||_{X_N} \le \gamma_F ||u - v||_{X_N}.$$

Implicit in the assumptions listed above is the statement that $F'_N(u)$ and its inverse can be extended in a continuous way to X_N for all $u \in Z_N$ and that $F'(u)Z \subset Z$ for all $u \in Z$. As above, if F has more than one root, the results in this paper still hold if we restrict attention to a sequence of connected components of $\mathcal{S}_N(\beta)$, that contain roots $u^N \stackrel{d}{\to} u^*$, and assume that iterates at all levels remain in these components.

To motivate our assumptions on the merit function, note that if the X_N norms are discrete L^2 norms and g_N is given by

$$g_N(u) = ||F_N(u)||_{X_N}^2/2$$

then our assumptions imply

(a) There is a γ_g such that, for all N and $u, v, w \in Z_N$,

$$|(g'_N(u) - g'_N(v))w| \leq ||w||_{X_N}\gamma_g||u - v||_{X_N}.$$

(b) There is $\eta > 0$, independent of N such that

$$g_N(u) \geq \eta||u - u^N||_{X_N}^2, \text{ for all } u \in \mathcal{D}_N(\beta).$$

(c) $g_N(u^N) = 0$.

In fact, we may use $\eta = 1/M$ and $\gamma_g = M\gamma_F + \beta\gamma_D$. We generalize this to incorporate other smooth merit functions, such as $g(u) = ||F(u)||_{L^p}^p$ for example. We assume

(2.10)

(a) There are δ and γ_g such that, for all N and $u, v, w \in Z_N$, $|(g'_N(u) - g'_N(v))w| \leq ||w||_{X_N}\gamma_g||u - v||_{X_N}^{\delta}$.

(b) There are $\nu, \eta > 0$, independent of N such that $g_N(u) \geq \eta||u - u^N||_{X_N}^{\nu}$, for all $u \in \mathcal{D}_N(\beta)$.

(c) $g_N(u^N) = 0$.

Note that our assumptions exclude nonsmooth merit functions. The choice $g(u) = ||F(u)||_Z$, for example, in the case $Z = C[0,1]$, is excluded. Mesh independence results of the type presented in this paper and global convergence results that give explicit rate information need smoothness on the merit function. Choices of g such as $g(u) = ||F(u)||_{L^r}^r$, for $r > 1$, are included. In the case of (1.2) and the approximations in Section 1, such a choice for g would make $X = L^r$ and the X_N norms discrete L^r norms. As we pointed out before, use of the quadrature rule requires $Z = C[0,1]$ with the max norm, and the merit function requires a space $X \neq Z$. Also, $P_N \Pi_N \to I$ fails in the X norm.

The Armijo rule drives the merit function to its minimum value via an approximate line search on the ray from the current iterate along a descent direction, i.e., a direction p such that $g'(u)p < 0$. In this paper we will restrict our descent directions to those of the form $p = -B^{-1}F(u)$, even though our analysis applies to descent directions that do not come from Newton like methods. The reason for this restriction is that we want to relate the method for construction of the operator B to the mesh-independence properties of the iteration. If, in the case $Z = C[0,1]$, we set $g(u) = g(u : r) = ||F(u)||_{L^r}^r / r$ for some $2 \leq r < \infty$, then g is smooth and one direction satisfying $g'(u)p < 0$ is clearly the Newton direction, with $B = F'(u)$. Note that, with this choice of p, B and B^{-1} are bounded in both the Z and X norms. Hence, if the initial iterate is in Z, all subsequent iterates will be in Z as well. Another descent direction, the steepest descent direction for $g(u) = g(u : 2) = ||F(u)||_{L^2}^2/2$, uses $B = (F'^*)^{-1}$. This operator is not automatically guaranteed to be bounded in Z, but is if, say, $F' - I$ is an integral operator with a continuous kernel. Either of these choices of descent direction satisfy, for some $\tau > 0$,

$$g'(u)p \leq -\tau g(u). \tag{2.11}$$

For example, in the case of the Newton step and $g(u) = g(u : 2)$, $\tau = 2$ and for the steepest descent direction, assuming either that it is defined in Z or that we are working in X, $\tau = 2M^{-2}$. Note that these estimates do not depend on N in any way and that we have suppressed mention of N in this paragraph.

We can now state our generalized Armijo rule. The rule will be the same for every N and hence we suppress the subscripts on Z, X, and G. Let g be a merit function that satisfies (2.2) and (2.10). As in Section 1 we have parameters, $\mu \in (0,1)$, $\rho > 0$, and $q \in (0,1)$. Given a descent direction, $p \in Z$, that satisfies (2.11), find α such that

$$\alpha > \rho g(u)^{1/\delta} ||p||_X^{-(1+1/\delta)}, \tag{2.12}$$

then find the smallest integer, $j \geq 0$, such that

$$g(u_c) - g(u_c + \alpha q^j p) > \alpha \mu q^j g(u_c), \tag{2.13}$$

and then set

$$u_+ = u_c + \alpha q^j p. \tag{2.14}$$

The condition in (2.12) is a generalization of that in (1.11) for $\delta \neq 1$. For definiteness we use

$$\alpha = \max\{1, 1.1 \rho g(u)^{1/\delta} ||p||_X^{-(1+1/\delta)}\}. \tag{2.15}$$

Note the dependence of the method on the Hölder continuity properties of g as described by (2.10)(a).

We will compare the convergence properties of iterative methods by the number of iterates required to drive the X norm of F to a small value. For $N > 1$ and $u_\ell^N \to u^N$, we define

$$\ell_N(\varepsilon) = \min\{\ell \mid g_N(u_\ell^N) < \varepsilon\}, \tag{2.16}$$

where g_N is a merit function satisfying (2.10).

In [**19**] we considered the local convergence behavior of Broyden's method as the level, N, varied. In that paper we used $g_N = || \cdot ||_{X_N}$ where $|| \cdot ||_{X_N}$ was an inner product norm. If we denote the Broyden update at level N as

$$B_{n+1}^N = B_n^N + \frac{(y - B_n^N) \otimes_N s}{||s||_{X_N}^2}$$

where $\otimes_N$ is the X_N tensor product, we were able to show that, under certain assumptions on the initial derivative approximations, that for any $\varepsilon > \delta > 0$, there was an N_0 so that if $N > N_0$, then

$$\ell_\infty(\varepsilon + \delta) \leq \ell_N(\varepsilon) \leq \ell_\infty(\varepsilon). \tag{2.17}$$

The key point in [**19**] was that the rules for computation of the iterates at each level produced a convergent sequence of iterates, $\{u_n^N\}$, as $N \to \infty$. In this paper we strengthen this result to show that (2.17) holds for general merit functions, g_N, that satisfy (2.10) and holds for globally convergent algorithms.

A second issue raised in [**19**] is the use of Nyström interpolation to convert a good X-norm estimate of the solution into a good Z-norm estimate. The assumptions necessary for this are, first of all, that for all N,

$$F_N = I + K_N, \quad \text{where } K_N : X_N \to Z_N, \tag{2.18}$$

and that the maps, K_N be uniformly continuous in the sense that there is a $\delta_K \in C[0, \infty)$ with $\delta_K(0) = 0$ such that

$$||K_N(u) - K_N(v)||_{Z_N} \le \delta_K(||u - v||_{X_N}), \tag{2.19}$$

independently of N. These assumptions are certainly valid if $F_N = \mathcal{F}_N$ and $K_N = \mathcal{K}_N$ are the operators considered in Section 1. If (2.18) and (2.19) hold and if $\tilde{u}^N$ is near u^N in the X_N norm, $K(\tilde{u}^N)$ will be near u^N in the Z_N norm. This observation holds equally well for the sequences generated by the globally convergent methods considered here.

3. Global convergence results. In this section we prove two results on the generalized Armijo rule given by (2.15), (2.13), and (2.14). The first gives rates on convergence of g_N to zero that do not depend on N. The second, the mesh independence result, says that for almost all values of the parameter, μ, in the Armijo rule, mesh independence holds.

The first results will depend only on the constants in the assumptions in (2.9), (2.10), (2.11), and indirectly on the fact that the descent directions remain in Z. This latter assumption is implied, in the case of the Newton direction, by (2.6). As is typical in analysis of such methods (see, for example, [**24**] or [**13**]) we require a lemma which gives a lower bound on the steplengths.

Lemma 3.1. *Assume that* (2.6), (2.9), *and* (2.10) *hold. Assume that there is τ such that for all $u \in \mathcal{D}(\beta)$, p is chosen so that* (2.11) *holds. Let $\mu < \tau$ and $\sigma = \alpha q^j$ be the steplength taken by the Armijo rule from $u = u_c$ in the direction p. Then there is $\tilde{c}$ such that*

$$\sigma \ge \tilde{c} g(u)^{1/\delta} ||p||_X^{-(1+1/\delta)}. \tag{3.1}$$

Proof. If $j = 0$, then by (2.15) we may choose $\tilde{c} = 1$. If $j \geq 1$, then we must have

$$g\left(u + \frac{\sigma p}{q}\right) > \left(1 - \frac{\sigma\mu}{q}\right) g(u)$$

and hence, with (2.10)(a) and (2.11),

$$\begin{aligned}\mu\sigma g(u) &> -q(g(u + \frac{\sigma p}{q}) - g(u)) = -q\int_0^1 g'(u + \frac{t\sigma p}{q})\frac{\sigma p}{q}\,dt \\ &= -\sigma g'(u)p + \sigma\int_0^1 (g'(u) - g'(u + \frac{t\sigma p}{q}))p\,dt \\ &\geq \sigma\tau g(u) - \gamma_g(\sigma||p||_X)^{1+\delta}/(q^\delta(1+\delta)).\end{aligned} \tag{3.2}$$

Hence

$$\sigma^\delta \geq \frac{(\tau - \mu)q^\delta(1+\delta)g(u)}{\gamma_g||p||_X^{1+\delta}}$$

which implies the result with

$$\tilde{c} = \left(\frac{(\tau - \mu)q^\delta(1+\delta)}{\gamma_g}\right)^{1/\delta}. \qquad \square$$

From the lower bound, (3.1), a uniform rate of global convergence can be obtained if the descent directions are related to the errors. We state this condition in terms of the operators, $\{B_n\}$.

Theorem 3.2. *Let the assumptions of Lemma* 3.1 *hold. Assume that* $\nu \geq 2$ *in* (2.10)(b), *and* $\delta \leq 1$. *Let* $\{u_n\}$ *be the sequence of iterates produced with the Armijo rule, with* $u_0 \in \mathcal{D}(\beta)$, *and* $p_n = -B_n^{-1}F(u_n)$. *Assume that* (2.11) *holds for* $p = p_n$ *and* $u = u_n$ *for all* n. *Moreover, assume that there is* M_B *such that, for all* $u \in Z$ *and* $n \geq 0$,

$$||B_n^{-1}u||_X \leq M_B||u||_X. \tag{3.3}$$

Then $g(u_n) \to 0$ *and satisfies*

$$g(u_{n+1}) \leq (1 - \zeta g(u_n)^{\xi - 1})g(u_n), \tag{3.4}$$

where $\xi = (1 + \delta^{-1})(1 - \nu^{-1})$.

Proof. By the Armijo rule, (2.13), and Lemma 3.1 we have

$$g(u_n) - g(u_{n+1}) \geq \sigma\mu g(u_n) \geq \tilde{c}\mu(g(u_n)/||p_n||_X)^{1+1/\delta}.$$

By (3.3)

$$||p_n||_X \leq M_B||F(u_n)||_X,$$

and by (2.9), letting $e_n = u_n - u^*$,

$$||F(u_n)||_X \leq M||e_n||_X.$$

Hence, by (2.10)(b),

$$||p_n||_X \leq M_B M||e_n||_X \leq M_B M\eta^{-1/\nu}g(u_n)^{1/\nu}.$$

If we let

$$\zeta = \tilde{c}\mu\left(\frac{\eta^{1/\nu}}{M_B M}\right)^{1+1/\delta},$$

then

$$\tilde{c}\mu\left(\frac{g(u_n)}{||p_n||_X}\right)^{1+1/\delta} \geq \zeta g(u_n)^{\xi}.$$

This completes the proof. □

Theorem 3.2 allows one to obtain q-linear convergence of the sequence $\{q(u_n)\}$ with q-factor $1-\zeta$ in the case $\xi = 1$, which follows from, say, $\nu = 2$ and $\delta = 1$. If $\xi > 1$ we obtain a weaker result.

Corollary 3.3. *Let the hypotheses of Theorem* 3.2 *hold. Then there is* $\tilde{\zeta}$ *such that*

$$g(u_n) \leq \tilde{\zeta}n^{1/(\xi-1)}. \tag{3.5}$$

Proof. Let $\phi(r) = \zeta r^{\xi}$ and define $\{r_n\}$ by $r_0 = g(u_0)$ and

$$r_{n+1} = r_n - \phi(r_n), \quad n \geq 0.$$

Clearly, $g(u_n) \leq r_n$ for all $n \geq 0$. Note that if

$$\Phi(r) = \int_{r_0}^{r} \frac{dt}{\phi(t)} = \frac{r^{1-\xi} - r_0^{1-\xi}}{\zeta(1-\xi)}$$

then, since $\{r_n\}$ is a decreasing sequence and ϕ is an increasing function, we have for all $i \geq 0$,

$$\begin{aligned}\Phi(r_{i+1}) - \Phi(r_i) &= \int_{r_i}^{r_{i+1}} \phi(t)^{-1}\, dt \leq \int_{r_i}^{r_{i+1}} \phi(r_i)^{-1}\, dt \\ &= (r_{i+1} - r_i)\phi(r_i)^{-1} = -1.\end{aligned}$$

Summing from $i = 0$ to $i = n$ implies $\Phi(r_n) \leq -n$ and

$$r_0^{1-\xi} - r_n^{1-\xi} \leq -n\zeta(\xi - 1).$$

Therefore,

$$r_n \leq (r_0^{1-\varepsilon} + \zeta(\xi - 1)n)^{1/(1-\xi)}.$$

This completes the proof with $\hat{\zeta} = (\zeta(\xi - 1))^{1/(1-\xi)}$. □

The conclusions of this section are, so far, that the rate of convergence can be estimated in terms of the operator bounds and continuity properties shared by all the problems, $F_N(u^N) = 0$, and all the merit functions, g_N. The important assumptions here were (2.6), (2.9), (2.10), and (2.11). The mesh independence result we want, namely (2.17), requires continuity properties of the individual iterates. This continuity will follow from the relationship between the problems at the various levels, (2.4), assumptions on the maps B_n and B_n^N, and requirements that the iterates produced by the Armijo rule for $N = \infty$, which converge in X, remain in Z if the initial iterate is in Z.

Our assumptions on the sequence $\{u_n\}$ and the maps $\{B_n\}$ are that

(3.6)
(a) The assumptions of Theorem 3.2 hold;
(b) $u_n \in Z$ for all n;
(c) There is $\hat{M}$ such that $||(B_n)^{-1}||_{\mathcal{L}(Z)} \leq \hat{M}$ for all n.

Assumption (3.6)(a) implies, in particular, that the sequence $\{u_n\}$ exists. Clearly, the Newton and steepest descent sequences satisfy such an assumption in the case $g(u) = ||F(u)||_{L^2}^2$ for the case of integral equations of the form (1.1) if k is sufficiently smooth and uniformly bounded. Assumption (3.6)(b) holds for iterates generated by Newton's method or steepest descent since Z is left invariant by $F'(u)$ and $F'(u)^*$ because both are sums of the identity and integral operators with continuous kernels. Assumption (3.6)(b) holds for Broyden iterates as well if B_0 leaves Z invariant since B_n differs from B_{n-1} by a rank one integral operator with a continuous kernel. Assumption (3.6)(c) will hold for the case of integral equations of the form (1.1) with k sufficiently smooth and uniformly bounded and iterates given by either steepest descent or Newton if (3.6)(a)–(b) hold as, in that case, $I-F$ is a continuous map from L^2 to C and F' a continuous map from L^2 to the space of bounded operators on C. In addition, quasi-Newton sequences such as Broyden's method can also be made to satisfy this assumption if the method is restarted by Newton's method or steepest descent if (2.11) fails or the quasi-Newton operator becomes ill-conditioned.

The assumption on the properties of the sequences $\{B_n^N\}$ is one of continuous dependence on previous data. This assumption is clearly satisfied in the case of the steepest descent method, Newton's method, or discretizations of quasi-Newton methods for integral equations if the quadrature rule and associated inner product satisfy our hypotheses.

$$\begin{aligned}&\text{For any } k>0, \text{ if } u_i^N \stackrel{d}{\to} u_i \text{ and}\\ &(B_i^N)^{-1} \stackrel{d}{\to} B_i^{-1} \text{ for all } 0 \le i < k, \text{ then}\\ &(B_k^N)^{-1} \stackrel{d}{\to} B_k^{-1}.\end{aligned} \tag{3.7}$$

Theorem 3.4. *Assume that* (2.4), (3.6), *and* (3.7) *hold. Then, except for at most countably many values of the parameter,* μ, *in the Armijo rule,* $\mu > 0$, *taken sufficiently small, if* $u_0^N \stackrel{d}{\to} u_0$ *and* $(B_0^N)^{-1} \stackrel{d}{\to} B_0^{-1}$, *then* $u_n^N \stackrel{d}{\to} u_n$ *for all* $n > 0$.

Proof. By (3.7), it clearly suffices to prove the result for $n = 1$. Our proof is in two stages. First, we show that the descent directions converge. Then we show that there is a countable exceptional set, E_1, such that for $\mu \notin E_1$ the choice of the integer, $j_1^N(\mu)$, selected by the Armijo rule for iterate 1 at level N, is the same as $j_1^\infty(\mu)$ for all but finitely many N. Once this second stage is complete, the result follows by taking as the countable set $\cup_\ell E_\ell$.

Note that $F_N(u_0^N) \xrightarrow{d} F(u_0)$ by (2.4)(b) and the assumption that $u_0^N \xrightarrow{d} u$. Hence, as $(B_0^N)^{-1} \xrightarrow{d} B_0^{-1}$, by assumption,

$$p^N = -(B_0^N)^{-1} F_N(u_0^N) \xrightarrow{d} p = -B_0^{-1} F(u_0).$$

This implies that if the α^N is defined by (2.15), $\alpha^N \to \alpha^\infty$ and hence

$$\alpha^N p^N \xrightarrow{d} \alpha^\infty p.$$

For any $j \geq 0$, (2.4)(a) and (b) imply that

$$g_N(u_0^N + q^j \alpha^N p^N) \to g(u + q^j \alpha^\infty p). \tag{3.8}$$

Hence, if

$$g_N(u_0^N + q^j \alpha^N p^N) < (1 - \mu \alpha^\infty q^j) g(u_0^N) \tag{3.9}$$

holds for $N = \infty$, (3.9) must hold for all but finitely many N. Conversely, if (3.9) holds for all but finitely many $N < \infty$, (3.9) can fail to hold for $N = \infty$ for at most a single exceptional value of μ, μ_j^1. We complete the proof by setting $E_1 = \cup_j \{\mu_j^1\}$, where the union is taken only over those values of j for which an exceptional value of μ exists. □

As was the case in [**19**] and [**21**], convergence of the sequence $\{u_n^N\}$ as $N \to \infty$ implies that Theorem 3.4 implies that (2.17) holds. We state this as a corollary.

Corollary 3.5. *Let the assumptions of Theorem* 3.4 *hold. Then, except for at most countably many values of the parameter, μ, in the Armijo rule,* (2.17) *holds for any $\varepsilon > \delta > 0$ and N sufficiently large.*

Globally convergent variants of quasi-Newton methods generally require that a decision be made on acceptance or rejection of a proposed step. This choice could be made differently even for nearby problems and it is this potential discontinuity that accounts for the countable set of exceptional values of μ for which the mesh-independence condition, (2.17) fails. The main result, Theorem 3.4, is quite natural in light of this discrete part of the algorithm.

Finally, we note that the merit function that determines the termination criterion in the definition of ℓ_N, (2.16), need not be the same as that in the global convergence criterion. For example, termination could be based on an L^∞ or C^k norm, while the merit function in the Armijo rule could be an L^p norm of F. The proof that (2.17) holds would remain the same, but the possibility that $\ell_N(\varepsilon) = \ell_\infty(\varepsilon) = \infty$ would have to be allowed.

4. Numerical examples. In this section we report on observations based on the Chandrasekhar equation [**8**]. We observe the results predicted by the theory in this paper with $F = \mathcal{F} : C[0,1] \to C[0,1]$ where

$$\mathcal{F}(H)(x) = H(x) - \frac{1}{1 - L(H)(x)} = 0, \tag{4.1}$$

where

$$L(H)(x) = \frac{c}{2}\int_0^1 \frac{xH(y)}{x+y}\, dy,$$

and $c \in [0,1]$ is a parameter. We seek a solution, $H \in C[0,1]$. For $c \neq 0, 1$ there are two solutions to (4.1) [**23, 14**], and at each of these, the Fréchet derivative is nonsingular at the solution. One of the solutions, characterized by analyticity in c, is of physical interest. The computations reported here found that physical solution and the results were compared with the tables in [**8**] to verify that fact. We let the quadrature rule for $N = 0$ be a composite 4 point Gauss rule with two subintervals and for $N \geq 1$ a composite 20 point Gauss rule. Our notation is that $m_0 = 8$ and $m_N = 20N$ for $N \geq 1$.

With this approximation, and taking Π_N as linear interpolation, the assumptions of all the results in Section 3 clearly hold. The computations were done in FORTRAN on an Alliant FX/40 running Alliant Concentrix 4.1.0. The Alliant was purchased with a DURIP grant from the AFOSR.

For the computations reported here, we used $c = .5$. The parameters in the Armijo Rule were $\mu = \rho = 10^{-4}$ and $q = .5$. We use the merit function $g(u) = ||\mathcal{F}(u)||_2^2/2$. Therefore, $\delta = 1$ and $\nu = 2$ in (2.10). Both Newton's method and Broyden's method computations were done. The initial iterate was

$$H_0(x) = 65\sin(20x).$$

For Broyden's method, the descent condition, (2.11), was tested with $\tau = 10^{-2}$ and the approximation

$$(g(H + \varepsilon p) - g(H))/\varepsilon \approx (\nabla g_N(H), p)_N,$$

TABLE 4.1. Newton's method.

	$N=0$		$N=1$		$N=4$		$N=32$	
n	j_n^N	$\|\mathcal{F}_N(u_n^N)\|$	j_n^N	$\|\mathcal{F}_N(u_n^N)\|$	j_n^N	$\|\mathcal{F}_N(u_n^N)\|$	j_n^N	$\|\mathcal{F}_N(u_n^N)\|$
1	0	.4038	3	$.4452D+2$	3	$.4452D+2$	3	$.4452D+2$
2	0	$.6360D-2$	0	.4312	0	.4312	0	.4312
3	0	$.1098D-5$	0	$.4626D-2$	0	$.4626D-2$	0	$.4626D-2$
4	0	$.2829D-13$	0	$.5205D-6$	0	$.5205D-6$	0	$.5205D-6$
5			0	$.6348D-14$	0	$.6327D-14$	0	$.6300D-14$

TABLE 4.2. Broyden's method.

	$N=0$		$N=1$		$N=4$		$N=32$	
n	j_n^N	$\|\mathcal{F}_N(u_n^N)\|$	j_n^N	$\|\mathcal{F}_N(u_n^N)\|$	j_n^N	$\|\mathcal{F}_N(u_n^N)\|$	j_n^N	$\|\mathcal{F}_N(u_n^N)\|$
1	0	$.1347D+1$	0	.7200	0	.7200	0	.7200
2	0	$.1334D+1$	0	.6069	0	.6069	0	.6069
3	0	$.1332D+1$	0	.6005	0	.6005	0	.6005
4	0	$.1274D+1$	0	.5571	0	.5571	0	.5571
5	0	.1453	0	$.9016D-1$	0	$.9021D-1$	0	$.9021D-1$
6	0*	$.8643D-3$	1	$.7285D-1$	1	$.7284D-1$	1	$.7284D-1$
7	0	$.6306D-5$	0*	$.2864D-3$	0*	$.2863D-3$	0*	$.2863D-3$
8	0	$.1138D-8$	0	$.6458D-6$	0	$.6454D-6$	0	$.6454D-6$
9			0	$.1340D-9$	0	$.1339D-9$	0	$.1339D-9$

with $\varepsilon = 10^{-6}$. When (2.11) failed, the Fréchet derivative was computed and Broyden's method was restarted. The initial approximation to the Fréchet derivative was

$$B_0 v(x) = v(x) + 75\int_0^1 v(y)\,dy.$$

These approximations were selected to exercise the globally convergent modification of Newton's and Broyden's method. In the tables above we tabulate, for $N = 0, 1, 4, 32$, the iterate number n, the number of stepsize reductions needed, j_n^N, and the value of $||\mathcal{F}_N||_{H_N}$. One can clearly see the mesh independence, not only in the convergence rates but also in the decision to reject the Broyden direction and restart with a Newton step. This happens at iterate 7 for all N, and we indicate this with an asterisk next to j_n^N. The iteration was terminated when $||\mathcal{F}_N||_{H_N} < 10^{-7}$.

REFERENCES

1. E.L. Allgower and K. Böhmer, *Application of the mesh independence principle to mesh refinement strategies*, SIAM J. Numer. Anal. **24** (1987), 1335–1351.

2. E.L. Allgower, K. Böhmer, F.A. Potra, and W.C. Rheinboldt, *A mesh-independence principle for operator equations and their discretizations*, SIAM J. Numer. Anal. **23** (1986), 160–169.

3. L. Armijo, *Minimization of functions having Lipschitz-continuous first partial derivatives*, Pacific J. Math. **16** (1966), 1–3.

4. K. Atkinson, *Iterative variants of the Nyström method for the numerical solution of integral equations*, Numer. Math. **22** (1973), 17–31.

5. H. Brakhage, *Über die numerische Behandlung von Integralgleichungen nach der Quadraturformelmethode*, Numer. Math. **2** (1960), 183–196.

6. C.G. Broyden, *A class of methods for solving simultaneous equations*, Math. Comp. **19** (1965), 577–593.

7. ———, *A new double-rank minimization algorithm*, Amer. Math. Soc. Notices, **16** (1969), 670.

8. S. Chandrasekhar, *Radiative transfer*, Dover, New York, 1960.

9. J.E. Dennis and R.B. Schnabel, *Numerical methods for nonlinear equations and unconstrained optimization*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

10. R. Fletcher, *A new approach to variable metric methods*, Comput. J. **13** (1970), 317–322.

11. D. Goldfarb, *A family of variable metric methods derived by variational means*, Math. Comp. **24** (1970), 23–26.

12. A Griewank, *The local convegence of Broyden-like methods on Lipschitzian problmes in Hilbert space*, SIAM J. Numer. Anal. **24** (1987), 684–705.

13. W.A. Gruver and E. Sachs, *Algorithmic methods in optimal control*, Pitman, London, 1980.

14. C.T. Kelley, *Solution of the Chandrasekhar H-equation by Newton's method*, J. Math. Phys. **21** (1980), 1625–1628.

15. ———, *A fast two-grid method for matrix H-equations*, Transport Theory of Statist. Phys. **18** (1989), 185–204.

16. ———, *Operator prolongation methods for nonlinear equations*, in *Computational solution of nonlinear systems of equations*, (E.L. Allgower and K. Georg, eds.), Lectures in Applied Mathematics, vol. **26** Amer. Math. Soc., Providence, RI, 1990, 359–388.

17. C.T. Kelley and J.I. Northrup, *Pointwise quasi-Newton methods and some applications*, in *Distributed parameter systems*, (F. Kappel, K. Kunisch, and W. Schappacher, eds.), New York, Springer-Verlag, 1987, 167–180.

18. ———, *A pointwise quasi-Newton method for integral equations*, SIAM J. Numer. Anal. **25** (1988), 1138–1155.

19. C.T. Kelley and E.W. Sachs, *Broyden's method for approximation solution of nonlinear integral equations*, J. Integral Equations **9** (1985), 25–44.

20. ———, *A quasi-Newton method for elliptic boundary value problems*, SIAM J. Numer. Anal. **24** (1987), 516–531.

21. ———, *Quasi-Newton methods and unconstrained optimal control problems*, SIAM J. Control Optim. **25** (1987), 1503–1517.

22. ———, *A pointwise quasi-Newton method for unconstrained optimal control problems*, Numer. Math. **55** (1989), 159–176.

23. T.W. Mullikin, *Some probability distributions for neutron transport in a half space*, J. Appl. Probab. **5** (1968), 357–374.

24. J. Ortega and W. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.

25. E. Sachs, *Global convergence of quasi-Newton type algorithms for some nonsmooth optimization problems*, JOTA **40** (1983), 201–219.

26. ———, *Convergence rates of quasi-Newton methods for some nonsmooth optimization problems*, SIAM J. Control Optim. **23** (1985), 401–418.

27. D.F. Shanno, *Conditioning of quasi-Newton methods for function minimization*, Math. Comp. **24** (1970), 647–657.

28. F. Stummel, *Diskrete Konvergenz linearer Operatoren*, Math. Ann. **190** (1970), 45–92.

Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205

Universität Trier, FB IV-Mathematik, Postfach 3825, 5500 Trier, West Germany