

MULTILEVEL ALGORITHMS FOR CONSTRAINED COMPACT FIXED POINT PROBLEMS *

C.T. KELLEY[†] AND E.W. SACHS[‡]

Abstract. In this paper we extend the multilevel algorithm of Atkinson and Brakhage for compact fixed point problems and the projected Newton method of Bertsekas to create a fast multilevel algorithm for parabolic boundary control problems having bound constraints on the control. We also extend results from finite dimension on constraint identification. Our approach permits both adaptive integration in time and inexact evaluation of the cost functional.

Key words. projected Newton method, constraint identification, collective compactness, parabolic optimal control problems

AMS(MOS) subject classifications. 45G10, 47H17, 49K20, 49M15, 65H10, 65K10

1. Introduction. In this paper, an expanded version of [15], we consider fast algorithms for solution of nonlinear equations that can be expressed in the form

$$(1.1) \quad u(t) = \mathcal{P}(\mathcal{K}(u))(t)$$

where \mathcal{K} is a completely continuous map from $L^\infty(\Omega)$ to $C(\Omega)$ for some bounded $\Omega \subset \mathbb{R}^n$ and \mathcal{P} is the map on $C(\Omega)$ given by

$$(1.2) \quad \mathcal{P}(u)(t) = \begin{cases} u_{min}(t), & \text{if } u(t) \leq u_{min}(t), \\ u(t), & \text{if } u_{min}(t) \leq u(t) \leq u_{max}(t), \\ u_{max}(t), & \text{if } u(t) \geq u_{max}(t), \end{cases}$$

for given u_{min} and u_{max} in $C(\Omega)$. The particular algorithm we consider is a generalization and synthesis of the Atkinson-Brakhage multi-level algorithm for compact fixed point problems [3], [6], and the projected Newton method of Bertsekas [5] for bound constrained minimization problems.

A paradigm for problems of the form (1.1) is the Urysohn equation

$$(1.3) \quad \mathcal{K}(u)(t) = \int_{\Omega} k(t, s, u(s)) ds.$$

Maps that are not easily expressible in this way, however, are the real target. In particular we wish to develop an algorithm general enough to be applicable to boundary control problems for partial differential equations. The algorithms and assumptions in this paper provide fast local convergence for problems with continuous controls in one space dimension.

Our methods differ from previous multi-level approaches for such problems [10] in that the smoothing requirements on the approximate Fréchet derivatives at the various levels are relaxed, the results on identification of active indices from [5] can

* Version of December 7, 1994

[†] North Carolina State University, Department of Mathematics, Center for Research in Scientific Computation, Box 8205, Raleigh, N. C. 27695-8205. The research of this author was supported by Air Force Office of Scientific Research grant #AFOSR-FQ8671-9101094, National Science Foundation grants #DMS-9024622 and #INT-8800560, North Atlantic Treaty Organization grant #CRG 920067, and an allocation of computing resources from the North Carolina Supercomputing Center.

[‡] Universität Trier, FB IV – Mathematik, Postfach 3825, 5500 Trier, Federal Republic of Germany. The research of this author was supported by the Volkswagen-Stiftung.

be extended, and the algorithm is a direct approximation of the projected Newton method and therefore quasi-Newton methods can be used to accelerate the convergence.

We are motivated by constrained parabolic optimal control problems in one space dimension. One such problem, a constrained version of the problem considered in [16], is to minimize

$$(1.4) \quad f(u) = \frac{1}{2} \int_0^1 (y(u; T, x) - z(x))^2 dx + \frac{\alpha}{2} \int_0^T u^2(t) dt,$$

where $\alpha > 0$ is given and $y(t, x) = y(u; t, x)$ is the solution to the nonlinear parabolic problem

$$(1.5) \quad \begin{aligned} y_t(t, x) &= y_{xx}(t, x), & 0 < x < 1, & \quad 0 < t < T, \\ y(0, x) &= y_0(x), & 0 < x < 1, \\ y_x(t, 0) &= 0, & y_x(t, 1) &= g(y(t, 1)) + u(t), & 0 < t < T. \end{aligned}$$

In this problem u is allowed to vary over the set

$$(1.6) \quad \mathcal{U} = \{u \in L^\infty([0, T]) \mid u_{\min}(t) \leq u(t) \leq u_{\max}(t), \text{ for a. e. } t \in [0, T]\}$$

and the nonlinear function g is assumed to satisfy

$$(1.7) \quad g \in C^2(\mathbb{R}), \quad g', g'' \in L^\infty(\mathbb{R}).$$

Such problems arise in metallurgy, for example [21].

The gradient of f in $L^2([0, T])$ is

$$(1.8) \quad (\nabla f(u))(t) = \alpha u(t) + d(t, 1),$$

where $d(t, x)$ is the solution of the adjoint problem

$$(1.9) \quad \begin{aligned} -d_t(t, x) &= d_{xx}(t, x), & 0 < x < 1, & \quad 0 < t < T \\ d(T, x) &= y(T, x) - z(x), & 0 < x < 1, \\ d_x(t, 0) &= 0, & d_x(t, 1) &= g'(y(t, 1))d(t, 1), & 0 < t < T. \end{aligned}$$

We let \mathcal{K} be the map that takes u into $-d(t, 1)/\alpha$. It is known [20] that \mathcal{K} is completely continuous from $L^\infty([0, T])$ to $C([0, T])$, (and hence a completely continuous map on $C([0, T])$), and in fact is a continuous map from $L^p([0, T])$ to $C([0, T])$ for $p > 2$. Standard techniques in optimization [4] imply that a necessary condition for u^* to be a solution of the problem given by (1.4), (1.5), and (1.6), is that

$$u^* = \mathcal{P}(\mathcal{K}(u^*)).$$

In [16] we considered the unconstrained problem with \mathcal{U} replaced by $C([0, T])$ and used DASSL [7] to perform the integration in time. Use of such an adaptive time integrator required weaker smoothing assumptions than used in [10] on the maps that take discrete versions of u into those of d . Under these relaxed smoothing assumptions we showed how the Atkinson-Brakhage algorithm could be implemented with appropriate finite difference gradients to obtain fast convergence.

The purpose of this paper is to merge the work in [16] with the ideas in [5] to design a fast algorithm for constrained optimal control problems of the type described above. We extend the projected Newton method to the abstract setting of constrained

compact fixed point problems. We generalize the convergence results and the results on identification of the intervals on which $u(t)$ attains its bounds. This latter result is an extension of some of the results in [17]. Once the analysis is complete we can apply the ideas in [16] directly to the problem given by (1.4)–(1.6) and produce a fast algorithm. The algorithm here differs from that proposed in [16] in that numerical Jacobians are not computed on coarse grids. Instead, GMRES [19] iteration is used to solve the coarse mesh linearized problems needed by the Atkinson-Brakhage iteration and a projected form of the Newton-GMRES iteration [8] is used to solve the coarse mesh problem itself. This modification, suggested for the first time in [12], makes the analysis of the algorithm proposed here simpler than the one from [16]. We conclude the paper with a report on some numerical results for constrained optimal control problems.

In the remainder of this section we briefly describe the projected Newton iteration of Bertsekas and our proposed algorithm for (1.1). We do not discuss the line search used in [5] to ensure global convergence since the focus of this paper is fast algorithms for local convergence. We take the position that the method from [5] is sufficient to obtain convergence from distant initial iterates on coarse meshes and that the solution so obtained can be interpolated to provide a good initial iterate on finer meshes.

The problem considered in [5] is to minimize a function f defined on a box in R^N

$$\mathcal{U}_N = \{u \in R^N \mid u_{min}^i \leq u^i \leq u_{max}^i, 1 \leq i \leq N\}.$$

Here u^i denotes the i th component of the vector $u \in R^N$ and $u_{min}, u_{max} \in R^N$ are given. If f is differentiable there is a solution u^* and each solution satisfies the necessary condition

$$(1.10) \quad u = \mathcal{P}(u - \nabla f(u))$$

where, similarly to (1.2) for the continuous problem,

$$(\mathcal{P}u)^i = \begin{cases} u_{min}^i, & \text{if } u^i \leq u_{min}^i, \\ u^i, & \text{if } u_{min}^i \leq u^i \leq u_{max}^i, \\ u_{max}^i, & \text{if } u^i \geq u_{max}^i. \end{cases}$$

The second order sufficient condition for optimality of u^* is that (1.10) holds and that the matrix H given by

$$H = P_A + P_I \nabla^2 f(u^*) P_I,$$

be positive definite. Here P_A is the projection

$$(P_A u)^i = \begin{cases} u^i, & \text{if } (u^*)^i = u_{max}^i \text{ and } (\nabla f(u^*))^i < 0 \\ u^i, & \text{if } (u^*)^i = u_{min}^i \text{ and } (\nabla f(u^*))^i > 0 \\ 0, & \text{otherwise.} \end{cases}$$

and $P_I = I - P_A$.

The iteration proposed in [5] is of the form

$$(1.11) \quad u_{n+1} = \mathcal{P}(u_n - \alpha_n H_n^{-1} \nabla f(u_n))$$

where

$$H_n = P_{A,n} + P_{I,n} \nabla^2 f(u_n) P_{I,n},$$

α_n is selected by an Armijo type rule, and $P_{A,n}$ and $P_{I,n}$ are approximations to P_A and P_I respectively. The construction of $P_{A,n}$ and $P_{I,n}$ in [5] ensures that in the iteration $P_{A,n} = P_A$ and $\alpha_n = 1$ for n sufficiently large. After the active set has been identified (when $P_{A,n} = P_A$) the iteration reduces to Newton's method on the inactive set and therefore local quadratic convergence holds. Crucial to the analysis in [5] is identification of the active set after finitely many iterations. This allows one to reduce the analysis of the limiting behavior of the iteration to that for Newton's method.

For problems with a continuum of constraints, such as the ones under consideration in this paper, identification of the active set after finitely many iterations is unlikely. To use the ideas of [5] one must change the the algorithm to take this into consideration and change the analysis as well. In § 2 we introduce notation and discuss the assumptions needed to make the estimate

$$\|u - u^*\|_X \leq K \|u - \mathcal{P}(\mathcal{K}(u))\|_X$$

for u sufficiently near u^* and some constant K . This estimate is trivial in the finite dimensional case if f is sufficiently smooth, u is an element of the sequence of iterates, and the active set has already been identified. In the infinite dimensional case considered here attention must be paid to the size of sets in which activity of the constraints is unclear. In § 3 we show how the algorithm in [5] can be extended to take the continuum of constraints into account and prove a local convergence result. The results in these sections require only a modest smoothing assumption. We show how the Atkinson-Brakhage algorithm can be used to create a fast algorithm and present some numerical results in § 4. The design of the fast algorithm requires a stronger compactness assumption than the results in the earlier sections.

2. The Basic Estimate. In this section we state our assumptions on the non-linearity and show that for continuous u the error in the solution is proportional to the size of the nonlinear residual in a neighborhood of the solution. We work in the space of continuous functions on Ω , $C = C(\Omega)$ and also on $X = L^\infty(\Omega)$, where $\Omega \subset R^d$. We assume that \mathcal{K} is a smooth map on X and seek to solve the constrained compact fixed point problem

$$(2.1) \quad \mathcal{F}(u) = u - \mathcal{P}(\mathcal{K}(u))$$

where \mathcal{P} is given by (1.2). We note that since \mathcal{K} maps L^p to $C = C(\Omega)$, the point evaluation implicit in \mathcal{P} is defined and therefore \mathcal{F} is a well defined map on X . The spaces X and C are both given the sup norm, which we denote by $\|\cdot\|_X$, and C is a closed subspace of X .

In keeping with the application to optimal control we will denote points in Ω by t . For $t \in \Omega$ we let U be given by

$$U(t) = \{u \in X \mid u_{min}(t) \leq u(t) \leq u_{max}(t)\},$$

where $u_{min}, u_{max} \in C$, and define the two point set

$$\partial U(t) = \{u_{min}(t), u_{max}(t)\}.$$

If $S \subset \Omega$ we let $S^c = \Omega \setminus S$ be the complement of S in Ω . If S and T are subsets of Ω we denote the symmetric difference by

$$S\Delta T = (S \cup T) \setminus (S \cap T).$$

Throughout this paper χ_S will denote the characteristic function of the set S and μ will denote Lebesgue measure on R^d . We will let $\mathcal{L}(X, Y)$ be the Banach space of bounded linear maps from a Banach space X to a Banach space Y and let $\mathcal{COM}(X)$ be the space of compact linear maps on X .

2.1. Differentiability, Smoothing, and Nonsingularity Assumptions. We make the following differentiability assumption on \mathcal{K} .

ASSUMPTION 2.1. *There is a solution $u^* \in C$ to (2.1) and a neighborhood $\mathcal{N} \subset X$ of u^* in which the \mathcal{K} is Lipschitz continuously Fréchet differentiable in X and in C .*

For future use we let γ be the Lipschitz constant for the map \mathcal{K} and γ_1 the Lipschitz constant for the maps \mathcal{K}' .

Our smoothing assumptions are

ASSUMPTION 2.2. *There is $p \in [1, \infty)$ such that the family of maps $\{\mathcal{K}'(u)\}$, where $u \in \mathcal{N}$ is a uniformly bounded subset of $\mathcal{L}(L^p(\Omega), C)$. Let*

$$(2.2) \quad \sup_{u \in \mathcal{N}} \|\mathcal{K}'(u)\|_{\mathcal{L}(L^p(\Omega), C)} = M_K.$$

An immediate consequence of these assumptions is

PROPOSITION 2.1. *For all measurable $S, T \subset \Omega$ and $u \in \mathcal{N}$*

$$\|\mathcal{K}'(u)(\chi_S - \chi_T)\|_C \leq M_K \mu(S \Delta T)^{1/p}.$$

We define active and inactive sets for u^* by

$$(2.3) \quad A^* = \{t \mid u^*(t) \in \partial U(t)\} \text{ and } I^* = \{t \mid u^*(t) \in \text{int}(U)(t)\}.$$

In (2.3) $\text{int}(U)(t)$ is defined to be the interval

$$\text{int}(U)(t) = (u_{\min}(t), u_{\max}(t)).$$

For a measurable subset S of Ω define

$$\mathcal{K}_S(u) = \chi_S \mathcal{K}(u^* + \chi_S(u - u^*))$$

and

$$\mathcal{G}_S(u) = u - \mathcal{K}_S(u)$$

Clearly we have

PROPOSITION 2.2. *For all measurable $I \subset \Omega$ \mathcal{G}_I is Lipschitz continuously Fréchet differentiable as a map on X and L^p . If I is closed, \mathcal{G}_I is Lipschitz continuously Fréchet differentiable as a map on $C(I)$. Moreover there is M_1 , independent of I , such that*

$$(2.4) \quad \|\mathcal{G}'_I(u)\|_{\mathcal{L}(C(I))}, \text{ (if } I \text{ is closed)}, \|\mathcal{G}'_I(u)\|_{\mathcal{L}(X)}, \|\mathcal{G}'_I(u)\|_{\mathcal{L}(L^p)} \leq M_1$$

for all $u \in \mathcal{N}$. In Proposition 2.2, the action of $\mathcal{G}'_I(u)$ on $w \in C(I)$ is understood to be given by extension of w to zero on $[0, T] \setminus I$, application of $\mathcal{G}'_I(u)$ to that extension, and restriction to I .

The next assumption is needed to make the extension of the projected Newton method described in [5] converge quadratically.

ASSUMPTION 2.3. *$\mathcal{G}'_{I^*} = I - \chi_{I^*} \mathcal{K}'(u^*) \chi_{I^*}$ is a nonsingular map on X and on L^p . There are η_1 and M_2 such that for every $u \in \mathcal{N}$ and every measurable $I \subset \Omega$*

such that $\|u - u^*\|_X < \eta_1$ and $\mu(I\Delta I^*) < \eta_1$ the map $\mathcal{G}'_I(u)$ is a nonsingular map on C , on X , and on L^p , and

$$(2.5) \quad \|\mathcal{G}'_I(u)^{-1}\|_{\mathcal{L}(C)}, \|\mathcal{G}'_I(u)^{-1}\|_{\mathcal{L}(X)}, \|\mathcal{G}'_I(u)^{-1}\|_{\mathcal{L}(L^p)} \leq M_2.$$

The most simple analog of the iteration (1.11) from [5] is

$$(2.6) \quad u_{n+1} = \mathcal{P}(u_n - A_n \mathcal{F}(u_n))$$

where A_n is an approximation of $\mathcal{G}'_{I^*}(u^*)^{-1}$. Construction of an extension of (1.11) that has good L^∞ convergence properties requires a more complex iteration than that given in (2.6). Fundamental to the local convergence analysis of any Newton-like algorithm is an bound of the error in terms of the size of \mathcal{F} . The remainder of this section is devoted to such an estimate.

2.2. Assumptions and Results on the sets A^* and I^* . The next two lemmas are simple consequences of the Lipschitz continuity of \mathcal{K} and the convexity of U . Before stating them we recall some more notation. If $S \subset R^k$ for some k and $t \in R^k$ we denote the distance from t to S by

$$\text{dist}(t, S) = \inf\{s \in S \mid \|t - s\|_{R^k}\}.$$

Note that

$$\mathcal{G}(u) = \mathcal{G}_\Omega(u) = u - \mathcal{K}(u).$$

For $u \in C$ define

$$(2.7) \quad \Omega_\delta(u) = \{t \mid |\mathcal{G}(u)(t)| \geq \delta\}, \text{ and } \Omega_\delta^* = \Omega_\delta(u^*) = \{t \mid |\mathcal{G}(u^*)(t)| \geq \delta\}.$$

We require the following trivial lemma.

LEMMA 2.3. $\Omega_\delta^* \subset A^*$ for all $\delta > 0$.

Proof. Since $u^*(t) = \mathcal{P}(\mathcal{K}(u^*))(t)$ either $\mathcal{G}(u^*)(t) = 0$ or $u^*(t) = \mathcal{P}(\mathcal{K}(u^*))(t) \neq (\mathcal{K}(u^*))(t)$. Hence if $t \in \Omega_\delta^*$ then $u^*(t)$ is the image under \mathcal{P} of a point outside of U and therefore $u^*(t) \in \partial U$. This means that $t \in A^*$ as asserted. \square

Recall that we denote by γ the Lipschitz constant for \mathcal{K} in the set \mathcal{N} , therefore $\bar{\gamma} = 1 + \gamma$ is the Lipschitz constant for \mathcal{G} in \mathcal{N} . We have

LEMMA 2.4. Assume that Assumption 2.1 holds. There is $\sigma_0 > 0$ such that if $u \in C$ satisfies

$$(2.8) \quad \|u - u^*\|_X < \sigma \leq \sigma_0,$$

and $\delta > \bar{\gamma}\sigma$ then

$$\Omega_\delta(u) \subset \Omega_{\delta - \bar{\gamma}\sigma}^* \subset A^*.$$

Proof. Let $\sigma_0 < \delta/\bar{\gamma}$ be small enough so that that (2.8) implies that $u \in \mathcal{N}$. Note that for $t \in \Omega_\delta(u)$,

$$|\mathcal{G}(u^*)(t)| \geq |\mathcal{G}(u)(t)| - \bar{\gamma}\sigma \geq \delta - \bar{\gamma}\sigma > 0.$$

Hence $t \in \Omega_{\delta - \bar{\gamma}\sigma}^* \subset A^*$ and the proof is complete. \square

ASSUMPTION 2.4. *There is $\nu \in (0, 1)$ such that $u_{max}(t) \geq u_{min}(t) + \nu$ for all $t \in \Omega$. A^* is the closure of a finite union of disjoint open components. There is $c_0 > 0$ such that for all $\delta > 0$ the sets*

$$E_\delta = \{t \in R^d \mid \text{dist}(t, \partial A^*) < \delta\}$$

are uniformly bounded in measure by

$$(2.9) \quad \mu(E_\delta) \leq c_0 \delta^d.$$

On each component of A^* either $u = u_{max}$ or $u = u_{min}$.

Moreover, there is c_1 such that

$$(2.10) \quad \begin{aligned} |\mathcal{G}(u^*)(t)| &\geq c_1 \text{dist}(t, \partial A^*) \text{ for all } t \in A^* \text{ and} \\ \text{dist}(t, A^*) &\leq c_1^{-1} \text{dist}(u^*(t), \partial U(t)) \text{ for all } t \in I^*. \end{aligned}$$

Note that (2.9) and the assertion that on each component of A^* either $u = u_{max}$ or $u = u_{min}$ follow from the previous parts of the assumption if $d = 1$. For $d > 1$ they may be viewed as regularity conditions on the boundary of A^* .

LEMMA 2.5. *Let Assumptions 2.1 and 2.4 hold. Then there are c_2 and δ_1 such that if $\delta \leq \delta_1$ then $\Omega_\delta^* \subset A^*$ and*

$$(2.11) \quad \mu(A^* \setminus \Omega_\delta^*) \leq c_2 \delta^d.$$

Proof. Let

$$\delta_1 = \min(\delta_0, \delta_0/c_1)$$

where δ_0 and c_1 are from Assumption 2.4. $\Omega_\delta^* \subset A^*$ by Lemma 2.3. If $t \in A^* \setminus \Omega_\delta^*$ then $|\mathcal{G}(u^*)(t)| < \delta$ and $u^*(t) \in \partial U(t)$. Hence, by (2.10) from Assumption 2.4 for $\delta \leq \delta_1 \leq \delta_0$

$$\text{dist}(t, \partial A) < \delta/c_1.$$

This implies that

$$A^* \setminus \Omega_\delta^* \subset E_{\delta/c_1}$$

and therefore, by (2.9) from Assumption 2.4,

$$\mu(A^* \setminus \Omega_\delta^*) \leq c_0 c_1^{-d} \delta^d.$$

This completes the proof with $c_2 = c_0 c_1^{-d}$. \square

Let $\sigma_1 = \min(\sigma_0, \delta_1/\bar{\gamma})/2$. Under all of our assumptions, for a given $u \in C$ such that $\|u - u^*\|_X < \sigma \leq \sigma_1$ we have by Lemma 2.4 that

$$\Omega_{\delta_1}(u) \subset \Omega_{\delta_1/2}^* \subset A^*.$$

Define

$$(2.12) \quad \bar{I} = \{t \mid \mathcal{K}(u)(t) \in \text{int}(U)(t)\} \cap I^*.$$

Since $A^* \cap I^* = \emptyset$, $\Omega_{\delta_2}^* \subset A^*$, and $\bar{I} \subset I^*$, we can decompose Ω into the disjoint union

$$(2.13) \quad \Omega = \Omega_{\delta_2}^* \cup \bar{I} \cup R(u)$$

where $\delta_2 = \delta_1/2$ and

$$(2.14) \quad R(u) = \Omega \setminus (\bar{I} \cup \Omega_{\delta_2}^*).$$

As a trivial corollary to Lemma 2.5 we obtain the estimate

COROLLARY 2.6. *Let Assumptions 2.1 and 2.4 hold. If $u \in C$ and $\|u - u^*\|_X < \sigma_1$ then*

$$\mu((R(u) \cup \bar{I}) \setminus I^*) = \mu(A^* \setminus \Omega_{\delta_2}^*) \leq c_2 \delta_2^d.$$

The next lemma will enable us to estimate the size of $R(u)$.

LEMMA 2.7. *Let Assumption 2.4 hold. There is c_3 such that for all $\delta > 0$*

$$(2.15) \quad \mu(\{t \mid \text{dist}(\mathcal{K}(u^*)(t), \partial U(t)) < \delta\}) \leq c_3 \delta^d.$$

Proof. Let

$$S = \{t \mid \text{dist}(\mathcal{K}(u^*)(t), \partial U(t)) < \delta\}.$$

If $t \in S \cap A^*$ then

$$\text{dist}(\mathcal{K}(u^*)(t), \partial U(t)) = \text{dist}(\mathcal{K}(u^*)(t), \mathcal{P}(\mathcal{K}(u^*)(t))) = |\mathcal{G}(u^*)(t)| < \delta$$

and hence, by (2.10) from Assumption 2.4, $\text{dist}(t, \partial A^*) < \delta/c_1$. Hence $S \cap A^* \subset E_{\delta/c_1}$.

If $t \in S \cap I^*$ then (2.10) from Assumption 2.4 implies that $\text{dist}(t, A^*) < c_1^{-1} \delta$ and therefore $S \cap I^* \subset E_{\delta/c_1}$. Hence $S \subset E_{\delta/c_1}$ and therefore

$$\mu(S) \leq \mu(E_{\delta/c_1}) \leq c_0 c_1^{-d} \delta^d.$$

This completes the proof with $c_3 = c_0 c_1^{-d}$. \square

LEMMA 2.8. *Let Assumptions 2.1 and 2.4 hold. Then if $u \in C$ is such that $\|u - u^*\|_X \leq \sigma_1$ and $R(u)$ is defined by (2.14) then*

$$(2.16) \quad \mu(R(u)) \leq c_3 \delta_2^d.$$

Proof. If $t \in R(u)$ then $|\mathcal{G}(u^*)(t)| < \delta_2$ because $R(u) \cap \Omega_{\delta_2}^* = \emptyset$. We divide the remainder of the proof into two cases. $t \in R(u)$ implies that $|\mathcal{G}(u^*)(t)| < \delta_2$ and that either (1) $\mathcal{K}(u^*)(t) \notin \text{int}(U(t))$ or (2) $\mathcal{K}(u^*)(t) \in \text{int}(U(t))$ and $\mathcal{K}(u^*)(t) \notin \text{int}(U(t))$.

In case (1) $|\mathcal{G}(u^*)(t)| < \delta_2$ implies that $\text{dist}(\mathcal{K}(u^*)(t), \partial U(t)) < \delta_2$. In case (2), Lipschitz continuity implies that

$$|\mathcal{K}(u)(t) - \mathcal{K}(u^*)(t)| \leq \gamma \sigma_1$$

and hence, since $\mathcal{K}(u^*)(t) \in \text{int}(U(t))$ in case (2),

$$\text{dist}(\mathcal{K}(u^*)(t), \partial U(t)) \leq \gamma \sigma_1 \leq \bar{\gamma} \sigma_1 \leq \delta_2.$$

These estimates imply that

$$R(u) \subset \{t \mid \text{dist}(\mathcal{K}(u^*)(t), \partial U(t)) < \delta_2\}$$

and therefore, by Lemma 2.7,

$$\mu(R(u)) \leq c_3 \delta_2^d.$$

This completes the proof. \square

COROLLARY 2.9. *Let Assumptions 2.1 and 2.4 hold. Then if $u \in X$ is such that $\|u - u^*\|_X \leq \sigma_1$*

$$\mu(\bar{I}\Delta I^*) \leq (c_2 + c_3)\delta_2^d.$$

LEMMA 2.10. *Let Assumption 2.1 hold. Then if $u \in X$ is such that $\|u - u^*\|_X < \sigma_1$ then*

$$\mathcal{P}(\mathcal{K}(u)(t)) = u^*(t)$$

for all $t \in \Omega_{\delta_2}^*$, and

$$|u(t) - u^*(t)| \leq \|\mathcal{F}(u)\|_X$$

for all $t \in \Omega_{\delta_2}^*$.

Proof. Let $t \in \Omega_{\delta_2}^* \subset A^*$. Without loss of generality we may assume that $u^*(t) = u_{max}(t)$ and $\mathcal{K}(u^*)(t) \geq u_{max}(t) + \delta_2$. Since $\|u - u^*\| < \sigma_1$

$$\mathcal{K}(u)(t) \geq \mathcal{K}(u^*)(t) - \gamma\sigma_1 \geq u_{max}(t) + \delta_2 - \bar{\gamma}\sigma_1 > u_{max}(t).$$

Hence $\mathcal{P}(\mathcal{K}(u))(t) = u_{max}(t) = u^*(t)$. Therefore

$$|u(t) - u^*(t)| = |u(t) - \mathcal{P}(\mathcal{K}(u)(t))| = |\mathcal{F}(u)(t)| \leq \|\mathcal{F}(u)\|_X.$$

This completes the proof. \square

2.3. The Main Estimate. Now we reduce δ_1 if necessary so that

$$(2.17) \quad (c_2 + c_3)\delta_2^d < \eta_1$$

where η_1 is the bound in Assumption 2.3. This has the effect of reducing δ_2 and σ_1 . By Corollary 2.9 this implies that $\mathcal{G}'_{\bar{I}}(u)$ satisfies (2.5).

We can now formally state and prove the main result in this section

THEOREM 2.11. *Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. There are $\sigma_2 > 0$ and $K > 0$ such that if $u \in C$ and $\|u - u^*\|_X \leq \sigma_2$ then*

$$\|u - u^*\|_X \leq K\|\mathcal{F}(u)\|_X.$$

Proof. We let $\sigma_2 = \sigma_1$ for now. We will reduce σ_2 as the proof progresses. Let $\sigma^* = \|u - u^*\|_X \leq \sigma_2$ and let $\|\mathcal{F}(u)\|_X = \epsilon$. We will estimate σ^* in terms of ϵ in the course of the proof in a way that can be applied to show $\sigma^* = O(\epsilon)$.

Without loss of generality we assume that $\epsilon < \sigma^*$ as if that is not the case the lemma holds with $K = 1$. We decompose the projected gradient map into three parts

$$u - \mathcal{P}(\mathcal{K}(u)) = \chi_{\bar{I}}(u - \mathcal{P}(\mathcal{K}(u))) + \chi_{R(u)}(u - \mathcal{P}(\mathcal{K}(u))) + \chi_{\Omega_{\delta_2}^*}(u - \mathcal{P}(\mathcal{K}(u))).$$

Since $\mathcal{K}(u) = u^*$ on $\Omega_{\delta_2}^*$ by Lemma 2.10 and $\mathcal{PK}(u) = \mathcal{K}(u)$ on \bar{I} by definition we have

$$u - \mathcal{PK}(u) = \chi_{\bar{I}}(u - \mathcal{K}(u)) + \chi_{R(u)}(u - \mathcal{P}(\mathcal{K}(u))) + \chi_{\Omega_{\delta_2}^*}(u - u^*).$$

Let $e = u - u^*$, $e_I = \chi_{\bar{I}}e$, $e_R = \chi_{R(u)}e$, and $e_\delta = \chi_{\Omega_{\delta_2}^*}e$. Therefore $\|e_\delta\|_X \leq \|\mathcal{F}(u)\|_X = \epsilon$ by Lemma 2.10. To analyze the sizes of e_I and e_R we note that \mathcal{K} is a Lipschitz continuous map from $L^p(\Omega)$ to X with Lipschitz constant M_K , where M_K is the bound on $\|\mathcal{K}'\|_{\mathcal{L}(L^p(\Omega), X)}$ from (2.2). Hence

$$\|\mathcal{K}(u) - \mathcal{K}(u^* + e_I)\|_X \leq \gamma\|e_\delta\|_X + M_K\|e_R\|_{L^p} \leq \gamma\epsilon + M_K\|e_R\|_{L^p}.$$

We write $\chi_{\bar{I}}(u - \mathcal{K}(u))$ as

$$(2.18) \quad \chi_{\bar{I}}(u - \mathcal{K}(u)) = \chi_{\bar{I}}(u - \mathcal{K}(u^* + e_I)) + \zeta_1(u) = \chi_{\bar{I}}\mathcal{G}_{\bar{I}}(u) + \zeta_1(u),$$

where

$$\zeta_1(u) = \chi_{\bar{I}}(\mathcal{K}(u^* + e_I) - \mathcal{K}(u^* + e_I + e_R + e_\delta)).$$

Recalling that

$$\|\mathcal{K}'(u)\|_{\mathcal{L}(L^p, X)} \leq M_K$$

for all $u \in \mathcal{N}$ we have, since $\|e_R\|_{L^p} \leq \|e_R\|_X \mu(R(u))^{1/p}$,

$$(2.19) \quad \begin{aligned} \|\zeta_1(u)\|_X &\leq \gamma\|e_\delta\|_X + M_K\|e_R\|_{L^p} \\ &\leq \gamma\epsilon + M_K\mu(R(u))^{1/p}\|e_R\|_X \leq \gamma\epsilon + M_K\mu(R(u))^{1/p}\sigma^*. \end{aligned}$$

Hence, by (2.5),

$$(2.20) \quad \|e_I\|_X \leq M_2\|\zeta_1(u)\|_X \leq M_2(\gamma\epsilon + M_K c_3^d \delta_2^{d/p} \sigma^*).$$

For convenience we rewrite (2.20) as

$$(2.21) \quad \|e_I\|_X \leq c_4(\epsilon + \delta_2^{d/p} \sigma^*),$$

where $c_4 = M_2 \max(\gamma, M_K c_3^d)$.

Now reduce δ_1 if necessary so that

$$c_4 \delta_2^{d/p} < \frac{1}{8\gamma} \text{ and } M_K (c_3 \delta_2^d)^{1/p} < \frac{1}{8}.$$

At this point we have shown that if δ_1 is sufficiently small then $\|e_\delta\| \leq \|\mathcal{F}(u)\|$

$$\|e_I\| \leq c_4\|\mathcal{F}(u)\| + \frac{\sigma^*}{8\gamma}.$$

We will obtain a similar estimate for $\|e_R\|_X$ and then apply these estimates to obtain the conclusion $\|e\|_X = O(\epsilon)$.

Let $\xi = \chi_{R(u)}\mathcal{F}(u)$. Let $e_T = e_I + e_\delta$, we have

$$\|e_T\|_X \leq (1 + c_4)\epsilon + \frac{\sigma^*}{8\gamma}.$$

Since $u^* = \mathcal{P}(\mathcal{K}(u^*))$ we may write

$$e_R = \xi + \zeta_2(u)$$

where

$$\zeta_2(u) = \chi_{R(u)}(\mathcal{P}(\mathcal{K}(u^* + e_T + e_R)) - \mathcal{P}(\mathcal{K}(u^*))).$$

Since $\|e_R\|_X \leq \sigma^*$,

$$\begin{aligned} \|\zeta_2(u)\|_X &\leq \gamma\|e_T\|_X + \|\mathcal{P}(\mathcal{K}(u^* + e_R)) - \mathcal{P}(\mathcal{K}(u^*))\|_X \\ &\leq \gamma((1 + c_4)\epsilon + \sigma^*/(8\gamma)) + M_K\mu(R(u))^{1/p}\|e_R\|_X \\ (2.22) \quad &\leq \gamma(1 + c_4)\epsilon + \sigma^*/8 + M_K(c_3\delta_2^d)^{1/p}\sigma^* \\ &\leq \gamma(1 + c_4)\epsilon + \sigma^*/4. \end{aligned}$$

Here we use the fact that \mathcal{P} has Lipschitz constant 1 because it is a projection onto a convex set.

Noting that $\|\xi\|_X \leq \|\mathcal{F}(u)\|_X = \epsilon$ we have that

$$\|e_R\|_X \leq \epsilon + \|\zeta_2(u)\|_X \leq K_2\epsilon + \sigma^*/4,$$

where $K_2 = 1 + \gamma(1 + c_4)$. Therefore

$$\|e\|_X = \sigma^* \leq \|e_R\|_X + \|e_T\|_X \leq (1 + c_4 + K_2)\epsilon + 3\sigma^*/4,$$

and

$$\sigma^*/4 \leq (1 + c_4 + K_2)\epsilon.$$

This completes the proof with $K = 4(1 + c_4 + K_2)$. \square

3. The Algorithm. In this section we describe our Newton like iteration in broad terms. The details of an efficient implementation will be presented in § 4. Our first task is to formulate the projected Newton iteration and analyze its convergence properties. Following that it is easy to describe the class of algorithms that we implement.

Assume that the assumptions of Theorem 2.11 hold. Let $u_c \in C$ be such that $\|e_c\| < \sigma_2$, where $e_c = u_c - u^*$. We let

$$\|\mathcal{F}(u_c)\|_X = \delta_c.$$

Let $\bar{p} \in (0, 1)$. We define the sets

$$\begin{aligned} A_c &= \{t \mid \mathcal{P}(\mathcal{K}(u_c))(t) \neq \mathcal{K}(u_c)(t), |\mathcal{G}(u_c)(t)| \geq \delta_c^{\bar{p}}\} \\ (3.1) \quad I_c &= \{t \mid \mathcal{P}(\mathcal{K}(u_c))(t) = \mathcal{K}(u_c)(t)\} \\ R_c &= \{t \mid \mathcal{P}(\mathcal{K}(u_c))(t) \neq \mathcal{K}(u_c)(t), |\mathcal{G}(u_c)(t)| < \delta_c^{\bar{p}}\} \end{aligned}$$

Note that the point evaluations required to determine the sets A_c , I_c , and R_c are well defined since $u_c \in C$. Note also that I_c is a closed set because u_{min} and u_{max} are continuous.

Using (2.6) as an iteration would, as we shall see, not produce iterates that converge in the uniform norm. We use an evaluation of \mathcal{K} to remedy this. We propose the iteration

$$(3.2) \quad \begin{aligned} u^{1/3} &= \chi_{A_c} \mathcal{P}\mathcal{K}(u_c) + \chi_{I_c \cup R_c} u_c \\ u^{2/3} &= u^{1/3} - \mathcal{G}'_{I_c}(u^{1/3})^{-1} \mathcal{F}(u^{1/3}) \\ u_+ &= \mathcal{P}(\mathcal{K}(u^{2/3})) \end{aligned}$$

In the final form of the algorithm we advocate here, we will replace

$$\mathcal{G}'_{I_c}(u^{1/3})^{-1} \mathcal{F}(u^{1/3})$$

with an approximation of the action of $\mathcal{G}'_{I_c}(u^{1/3})^{-1}$ on $\mathcal{F}(u^{1/3})$ that can be evaluated efficiently. We note here that even though the intermediate iterate $u^{1/3}$ is not continuous on $[0, T]$, it is everywhere defined and continuous on the set $I_c \cup R_c$, and in particular on the closed set I_c , and so methods for approximation of integral operators on the space of continuous functions are applicable to the approximation of the action of $\mathcal{G}'_{I_c}(u^{1/3})^{-1}$. We defer the construction of this approximation to § 4 and in this section consider (3.2) and estimate $e^{2/3}$ in the L^p norm, where p is the exponent in Assumption 2.2. That estimate leads directly to a uniform estimate for $e_+ = u_+ - u^*$ by Assumption 2.2.

LEMMA 3.1. *Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Let σ_2 and K be the constants from Theorem 2.11. Let $u_c \in X$ be such that*

$$(3.3) \quad \|e_c\|_X \leq \min\{\sigma_2, (1/2 + \gamma/K)^{-1/(1-\bar{p})}/\gamma, \nu/\gamma\}$$

then $A_c \subset A^*$ and $u^{1/3} = u^*$ on A_c .

Proof. Since $\delta_c \leq \gamma\|e_c\|_X$, (3.3) implies that

$$(3.4) \quad \delta^{1-\bar{p}} < (1/2 + \gamma/K)^{-1}.$$

Let $t \in A_c$, since

$$|\mathcal{G}(u^*)(t)| \geq |\mathcal{G}(u_c)(t)| - \gamma\|e_c\| \geq \delta_c^{\bar{p}} - \gamma\delta_c/K \geq \delta_c/2$$

by (3.4). Hence $A_c \subset A^*$ by Lemma 2.3.

To complete the proof, note that if $t \in A_c$ then $\mathcal{K}(u)(t) \in \partial U(t)$. As

$$|u^*(t) - \mathcal{P}(\mathcal{K}(u^*))(t)| = |\mathcal{P}(\mathcal{K}(u))(t) - \mathcal{P}(\mathcal{K}(u^*))(t)| \leq \gamma\|e_c\|_X < \nu$$

we must have $u^*(t) = \mathcal{P}(\mathcal{K}(u))(t) = u^{1/3}(t)$ for $t \in A_c$. This completes the proof. \square

The set R_c is small. This is made precise by the following lemma.

LEMMA 3.2. *Let Assumptions 2.1, 2.2, 2.3, 2.4, and equation (3.3) hold. Then if $u_c \in X$ is such that*

$$(3.5) \quad \|e_c\|_X \leq (\gamma K)^{-1/(1-\bar{p})}/\gamma$$

then

$$\mu(R_c) \leq c_3 3^d \delta_c^{\bar{p}/d},$$

where c_3 is the constant from Lemma 2.7.

Proof. By (3.3) $\|e_c\|_X < 1/\gamma$. Note that if $t \in R_c$ then $\mathcal{P}(\mathcal{K}(u_c))(t) \in \partial U(t)$ and also

$$\begin{aligned} |\mathcal{K}(u_c)(t) - \mathcal{P}(\mathcal{K}(u_c))(t)| &\leq |\mathcal{G}(u_c)(t)| + |\mathcal{F}(u_c)(t)| \\ &\leq \delta_c^{\bar{p}} + \delta_c \leq 2\delta_c^{\bar{p}}. \end{aligned}$$

The last estimate follows from the assumption that $\|e_c\|_X < 1/\gamma$ which implies that $\delta_c < 1$. Therefore

$$\begin{aligned} R_c &\subset \{t \mid \text{dist}(\mathcal{K}(u_c)(t), \partial U(t)) < 2\delta_c^{\bar{p}}\} \\ &\subset \{t \mid \text{dist}(\mathcal{K}(u^*)(t), \partial U(t)) < 2\delta_c^{\bar{p}} + \bar{\gamma}\|e_c\|\} \\ &\subset \{t \mid \text{dist}(\mathcal{K}(u^*)(t), \partial U(t)) < 2\delta_c^{\bar{p}} + \bar{\gamma}K\delta_c\} \\ &\subset \{t \mid \text{dist}(\mathcal{K}(u^*)(t), \partial U(t)) < 3\delta_c^{\bar{p}}\}. \end{aligned}$$

The last estimate above follows from (3.5). The conclusion of the lemma is a direct application of Lemma 2.7. \square

A_c will serve as the approximation to the active set. An immediate corollary of Lemmas 3.1 and Lemma 3.2 is a theorem on identification of the active set.

THEOREM 3.3. *Let Assumptions 2.1, 2.2, 2.3, 2.4, and equation (3.3) hold. Assume that $\delta_c < 1$. Then there is c_A such that*

$$\mu(A^* \setminus A_c) \leq c_A \delta_c^{\bar{p}/d}.$$

Proof. If $t \in A^* \setminus A_c$ then either $t \in R_c$ or $t \in I_c \cap A^*$. If $t \in I_c \cap A^*$ then $\mathcal{G}(u_c)(t) = \mathcal{F}(u_c)(t)$ and Assumption 2.4 implies

$$\begin{aligned} c_1 \text{dist}(t, \partial A^*) &\leq |\mathcal{G}(u^*)(t)| \leq |\mathcal{G}(u_c)(t)| + \gamma\|e_c\|_X \\ &= |\mathcal{F}(u_c)(t)| + \gamma\|e_c\|_X \leq \delta_c(1 + \gamma/K). \end{aligned}$$

This implies that

$$I_c \cap A^* \subset E_{c_1^{-1}\delta_c(1+\gamma/K)}$$

and therefore

$$\mu(I_c \cap A^*) \leq c_0(c_1^{-1}\delta_c(1 + \gamma/K))^d.$$

This completes the proof as $A^* \setminus A_c \subset R_c \cup (I_c \cap A^*)$ with

$$c_A = c_3 + c_0(c_1^{-1}(1 + \gamma/K))^d.$$

\square

The reader should compare the following lemma with (2.18) and (2.19), which used a similar decomposition of Ω to obtain a similar result.

LEMMA 3.4. *Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Let u_c satisfy the assumptions of Lemma 3.2. Then $|e^{1/3}(t)| \leq |e_c(t)|$ for all $t \in \Omega$ and there are $w_c \in X$ and $K_w > 0$ such that*

$$\|w_c\|_{L^p(\Omega)} \leq K_w \|e_c\|_X \delta_c^{\bar{p}/(pd)}$$

and

$$\mathcal{F}(u^{1/3}) = \chi_{I_c} \mathcal{G}_{I_c}(u^* + \chi_{I_c} e^{1/3}) - w_c.$$

Proof. By Lemma 3.1 $\chi_{A_c} \mathcal{F}(u^{1/3}) = \chi_{A_c} e^{1/3} = 0$. Hence

$$\begin{aligned} \mathcal{F}(u^{1/3}) &= u^{1/3} - \mathcal{P}(\mathcal{K}(u^* + \chi_{I_c \cup R_c} e_c)) \\ &= \chi_{I_c \cup R_c}(u^{1/3} - \mathcal{P}(\mathcal{K}(u^* + \chi_{I_c \cup R_c} e_c))). \end{aligned}$$

By Lemma 3.2

$$\|\chi_{R_c} \mathcal{F}(u^{1/3})\|_{L^p} \leq c_5 \|e_c\|_X \delta_c^{\bar{p}/(pd)}$$

where

$$c_5 = (c_3 3^d)^{1/p}.$$

Similarly

$$\|\mathcal{K}(u^* + \chi_{I_c} e_c + \chi_{R_c} e_c) - \mathcal{K}(u^* + \chi_{I_c} e_c)\|_X \leq M_K \|e_c\|_X \mu(R_c)^{1/p} \leq c_6 \|e_c\|_X \delta_c^{\bar{p}/(pd)}$$

where $c_6 = M_K c_5$. Hence if we let

$$w_c = \chi_{R_c} \mathcal{F}(u^{1/3}) + \chi_{I_c} (\mathcal{K}(u^* + \chi_{I_c} e_c + \chi_{R_c} e_c) - \mathcal{K}(u^* + \chi_{I_c} e_c)),$$

note that that for any $f \in X$, $\|f\|_{L^p} \leq \mu(\Omega)^{1/p} \|f\|_X$, and set $K_w = (1 + \mu(\Omega)^{1/p} M_K) c_5$ the proof is complete. \square

A trivial corollary of the proof is

THEOREM 3.5. *Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Let $u_c \in C$ satisfy the assumptions of Lemma 3.2 and assume that $\|e_c\|_{L^p}, \|e_c\|_X < 1$. Then there is K_P such that*

$$(3.6) \quad \|e^{2/3}\|_{L^p} \leq K_P \|e_c\|_X^{1+(\bar{p}/pd)}$$

and there is K_N such that

$$(3.7) \quad \|e_+\|_X \leq K_N \|e_c\|_X^{1+(\bar{p}/pd)}$$

Proof. By Lemma 3.4,

$$u^{2/3} = u^{1/3} - \mathcal{G}'_{I_c}(u^{1/3})^{-1} \mathcal{G}_{I_c}(u^* + \chi_{I_c} e^{1/3}) - \mathcal{G}'_{I_c}(u^{1/3})^{-1} w_c.$$

By Assumption 2.3, $\|\mathcal{G}'_{I_c}(u^{1/3})^{-1}\|_{L^p} \leq M_2$ and hence by Lemma 3.4

$$\begin{aligned} \|\mathcal{G}'_{I_c}(u^{1/3})^{-1} w_c\|_{L^p} &\leq M_2 K_w \|e_c\|_X \delta_c^{\bar{p}/(pd)} \\ &\leq M_2 K_w \bar{\gamma}^{\bar{p}/(pd)} \|e_c\|_X^{1+(\bar{p}/pd)}. \end{aligned}$$

By standard estimates for Newton iterates and Assumption 2.3

$$\|e^{1/3} - \mathcal{G}'_{I_c}(u^{1/3})^{-1} \mathcal{G}_{I_c}(u^* + \chi_{I_c} e^{1/3})\|_X \leq M_2 \bar{\gamma} \|e^{1/3}\|_X / 2.$$

Hence

$$\|e^{1/3} - \mathcal{G}'_{I_c}(u^{1/3})^{-1} \mathcal{G}_{I_c}(u^* + \chi_{I_c} e^{1/3})\|_{L^p} \leq \mu(\Omega)^{1/p} M_2 \bar{\gamma} \|e^{1/3}\|_X / 2.$$

The first assertion therefore holds with

$$K_P = M_2 K_w \bar{\gamma}^{\bar{p}/(p^d)} + \mu(\Omega)^{1/p} M_2 \bar{\gamma} / 2.$$

The second assertion follows from Assumption 2.2 with

$$K_N = M_K K_P.$$

This completes the proof. \square

The algorithms we implement replace $\mathcal{G}'_{I_c}(u_c)^{-1}$ in (3.2) with an approximation. The behavior of these algorithms is described by the next theorem, which is a direct consequence of Theorem 3.5.

THEOREM 3.6. *Let the assumptions of Theorem 3.5 hold and let u_+ be defined by*

$$(3.8) \quad \begin{aligned} u^{1/3} &= \chi_{A_c} \mathcal{P}\mathcal{K}(u_c) + \chi_{I_c \cup R_c} u_c \\ u^{2/3} &= u^{1/3} - B_c^{-1} \mathcal{F}(u^{1/3}) \\ u_+ &= \mathcal{P}(\mathcal{K}(u^{2/3})) \end{aligned}$$

where,

$$\|(B_c^{-1} - \mathcal{G}'_{I_c}(u^{1/3})^{-1})\|_{\mathcal{L}(X, L^p)} \leq \rho_c.$$

Then there is $K_C > 0$ such that

$$(3.9) \quad \|e_+\|_X \leq K_N \|e_c\|_X^{1+(\bar{p}/p^d)} + K_C \rho_c \|e_c\|_X.$$

Proof. This follows directly from Theorem 3.5 with $K_C = M_K \gamma$. To see this note that

$$\begin{aligned} \|e^{2/3}\|_{L^p} &\leq K_P \|e_c\|_X^{1+(\bar{p}/p^d)} + \rho \|\mathcal{F}(u^{1/3})\|_X \\ &\leq K_P \|e_c\|_X^{1+(\bar{p}/p^d)} + \gamma \rho_c \|e_c\|_X. \end{aligned}$$

\square

Finally, we give an inexact variant of Theorem 3.5. The analysis is like that of the finite dimensional case [9].

THEOREM 3.7. *Let the assumptions of Theorem 3.5 hold and let u_+ be defined by*

$$(3.10) \quad \begin{aligned} u^{1/3} &= \chi_{A_c} \mathcal{P}\mathcal{K}(u_c) + \chi_{I_c \cup R_c} u_c \\ u^{2/3} &= u^{1/3} + \tilde{s} \\ u_+ &= \mathcal{P}(\mathcal{K}(u^{2/3})) \end{aligned}$$

where $\tilde{s} \in X \cap C(I_c)$ and

$$(3.11) \quad \|\mathcal{G}'_{I_c}(u^{1/3})\tilde{s} + \mathcal{F}(u^{1/3})\|_{L^p} \leq \rho_c \|\mathcal{F}(u^{1/3})\|_X.$$

Then there is $K_I > 0$ such that

$$(3.12) \quad \|e_+\|_X \leq K_N \|e_c\|_X^{1+(\bar{p}/p^d)} + K_I \rho_c \|e_c\|_X.$$

In particular, if $\rho_c = \rho$ is independent of $u_c \in \mathcal{N}$, we have q -linear convergence and if $\rho_n \rightarrow 0$ as $n \rightarrow \infty$ the convergence is q -superlinear.

Proof. Note that

$$\begin{aligned} e^{2/3} &= e^{1/3} - \mathcal{G}'_{I_c}(u^{1/3})^{-1} \mathcal{F}(u^{1/3}) \\ &\quad - \mathcal{G}'_{I_c}(u^{1/3})^{-1} (\mathcal{G}'_{I_c}(u^{1/3})\tilde{s} + \mathcal{F}(u^{1/3})) \end{aligned}$$

Hence, as in the proof of Theorem 3.5

$$\begin{aligned} \|e^{2/3}\|_{L^p} &\leq K_P \|e_c\|_X^{1+(\bar{p}/p^d)} + M_2 \rho_c \|\mathcal{F}(u^{1/3})\|_X \\ &\leq K_P \|e_c\|_X^{1+(\bar{p}/p^d)} + M_2 \gamma \rho_c \|e_c\|_X \end{aligned}$$

Therefore

$$\begin{aligned} \|e_+\|_X &\leq M_K \|e^{2/3}\|_{L^p} \\ &\leq M_K (K_P \|e_c\|_X^{1+(\bar{p}/p^d)} + M_2 \gamma \rho_c \|e_c\|_X). \end{aligned}$$

Setting $K_I = M_K M_2 \gamma$ completes the proof. \square

4. Implementation and an Example. As in [16] we construct the operators B_c using a collectively compact sequence of approximations to \mathcal{K} . Recall that a family of maps $\{K_\alpha\}_{\alpha \in \Lambda} \subset \mathcal{L}(X, Y)$ is collectively compact if

$$\cup_{\alpha \in \Lambda} K_\alpha B$$

is precompact in Y for every bounded set $B \subset X$.

We incorporate an estimate of the set I_c using a ‘‘coarse mesh’’ approximation. We consider a sequence of approximations $\{\mathcal{K}_m\}$ to \mathcal{K} . We refer to the equation $u - \mathcal{P}(\mathcal{K}_m(u)) = 0$ as the equation for level m . Such approximations can be obtained, for example, by replacing the integral in (1.3) by a quadrature rule. Typically the problem at level m is equivalent to a finite dimensional problem of dimension N_m , say. Our compactness assumption is

ASSUMPTION 4.1. *The sequence of maps \mathcal{K}_m converge strongly to \mathcal{K} in \mathcal{N} and the family $\{\mathcal{K}'_m(u)\}_{m \leq \infty, u \in \mathcal{N}}$ is uniformly Lipschitz continuous as a family of maps from \mathcal{N} to $\mathcal{L}(L^p, C)$ and is a collectively compact subset of $\mathcal{L}(X, C)$.*

In Assumption 4.1, $\mathcal{K}_\infty = \mathcal{K}$. In the case of smooth nonlinearities, [2], [13], Assumption 4.1, solvability of $\mathcal{F}(u) = 0$, uniform Lipschitz continuity of the family $\{\mathcal{K}'_m(u)\}_{m \leq \infty, u \in \mathcal{N}}$ and nonsingularity of $\mathcal{F}'(u^*)$, would imply that $u^m = \mathcal{K}_m(u^m)$ would have a solution for m sufficiently large and that $u^m \rightarrow u^*$ in X . In the nonsmooth case considered here we must assume that the approximate problems have solutions which converge to u^* .

ASSUMPTION 4.2. *For m sufficiently large there is a solution to the fixed point problem $u^m = \mathcal{K}(u^m) \in C$. u^m converges uniformly to u^* .*

Our algorithm, like the original Atkinson-Brakhage algorithm, [3], [6], or multigrid methods of the second kind [11], begins with a coarse level l for which the problem at that level is solved. The Atkinson-Brakhage approach uses low level information to construct an approximate inverse B_l^L which is used to solve the problem at a higher level L . Our general scheme may be described in terms of the transition from a solution u^L at level L to a solution u^{L+1} for level $L+1$. The initial iterate at level $L+1$ is $u_0^{L+1} = u^L$. We apply the iteration (3.2) with $B_c = B_l^{L+1}$ until $\|\mathcal{F}_{L+1}(u^{L+1})\|_X \leq \epsilon_{L+1}$ where ϵ_{L+1} is an estimate for truncation error at level $L+1$. For a sufficiently fine coarse mesh one would hope to require only one iterate for each fine level L . Any necessary transfer of information between grids between levels can be done through Nyström interpolation [3] or directly by polynomial interpolation [14], [16].

The remaining issue is the construction of the maps B_l^L . We denote

$$\mathcal{G}_{l,S}(u) = I - \chi_S \mathcal{K}'_l(u) \chi_S,$$

and let

$$I_l = \{t \mid \mathcal{K}(u^l) = \mathcal{P}(\mathcal{K}(u^l))\}.$$

We describe the action of $(B_l^L)^{-1}$ on a function v by the basic formula

$$(4.1) \quad (B_l^L)^{-1}v = v + \chi_{I_c} \chi_{I_l} \mathcal{G}'_{l,I_l}(u^l)^{-1} \chi_{I_l} \chi_{I_c} \mathcal{K}'_L(u_c) \chi_{I_c} v.$$

In (4.1) u_c is the current iterate in at level L for the computation of u^L . Assumptions 2.2, 4.1, and 4.2 imply that B_l^L will satisfy the assumptions of Theorem 3.5 with ρ independent of L if the index of the lowest level, l , is sufficiently large. As with the original Atkinson-Brakhage iteration, (4.1) is an approximation to a two term Neumann series expansion of $(\mathcal{G}'_L(u_c))^{-1}$.

In [16] the levels corresponded to piecewise linear approximations of the functions and the equations at the levels were equivalent to nonlinear equations for the values of the piecewise linear functions at the nodal points. The action of the fine mesh derivative $w = \mathcal{K}'(u_c)v$ was approximated by a difference quotient. The action of $(I - \mathcal{K}'_l(u^l))^{-1}$ was approximated by forming $I - \mathcal{K}'_l(u^l)$ as a matrix by numerically differentiating in coordinate directions, forming an LU factorization of the resulting matrix and storing the factors. Then w was projected onto the coarse mesh, the equation $(I - \mathcal{K}'_l(u^l))z = w$ was solved, and z extended to the fine mesh. The final approximation was $B_c^{-1}v = v - z$. Of course, a similar technique could be used for the constrained problems considered here, setting

$$w = \chi_{I_l} \chi_{I_c} \mathcal{K}'_L(u_c) \chi_{I_c} v,$$

forming a matrix equivalent of \mathcal{G}'_{l,I_l} and factoring it, and then using that factorization to compute $z = (\mathcal{G}'_{l,I_l})^{-1}w$. Instead, we follow the technique of [12] and use a GMRES [19] iteration with a discrete L^2 inner product, to solve a finite dimensional problem associated with the linear system

$$(4.2) \quad \mathcal{G}'_{l,I_l}(u^l)z = w$$

and then to use interpolation to approximate the solution to (4.2). The matrix-vector products that are required by GMRES are done with a forward difference approximation to the action of the Fréchet derivative. As our assumptions on the family $\{\mathcal{K}_l\}$

will guarantee that the condition number of $I - \mathcal{K}'_l(u^l)$ is bounded independently of l , the results in [18] guarantee that the behavior of a GMRES iteration will be mesh independent in the sense of [1].

The theory in [16] allowed for inaccuracy in the evaluation of \mathcal{K}_L , which can be introduced, say, by the control of relative and absolute errors in adaptive integration codes or ODE codes like DASSL. In [16] we showed how one should adjust this accuracy to take into account the expected error $\|u^L - u^*\|_X$. That analysis can be incorporated here in a direct way in the evaluation of the difference approximations to directional derivatives $\mathcal{K}(u)'w$ that are required by both the Atkinson-Brakhage algorithm in the approximation of \mathcal{K}'_L at the fine mesh level and the solution of (4.2) by Newton-GMRES iteration. In [16] the analysis was made complicated by the numerical Jacobian and additional accuracy was needed in the coarse mesh function evaluations. The use of GMRES eliminates some of that complexity. As in [16] the use of inaccurate function evaluations introduces an absolute error in the iteration. Making this absolute error smaller as the iteration progresses means that the convergence rate of the algorithm becomes r-linear.

Assume that the compact fixed point maps, \mathcal{K}_m , can be evaluated to a given absolute accuracy τ_m , with smaller values of τ_m resulting in a increase of the cost of the nonlinear function evaluation. In the case of the method of lines solution of a parabolic partial differential equation, for example, τ_m would be the absolute error tolerance given to the routine that performs the integration in time. Letting $h_m = O(\sqrt{\tau_m})$ we approximate the action of $\mathcal{K}'_m(u)$ on a vector w by the forward difference map

$$A_m(u, w) = \begin{cases} 0, & w = 0 \\ \|w\|_X \frac{\mathcal{K}(u+h_m w/\|w\|_X) - \mathcal{K}(u)}{h_m}, & w \neq 0. \end{cases}$$

$A_m(u, w)$, though nonlinear in w , approximates $\mathcal{K}'_m(u)w$ up to a relative error of $O(h_m)$. We have

$$(4.3) \quad \|A_m(u, w) - \mathcal{K}'_m(u)w\|_X \leq \frac{\gamma}{2} h_m \|w\|_X.$$

Hence $\mathcal{G}'_{l,I_l}(u^l)z$, which is used in the construction of the Atkinson-Brakhage approximate inverse, would be approximated by

$$G_l(u^l, z) = I - \chi_{I_l} A_l(u^l, \chi_{I_l} z).$$

The effect of this approximation is that in the evaluation of the action of the Atkinson-Brakhage approximate inverse given by (4.1) $w = \mathcal{K}'_L(u_c)\chi_{I_c}v$ is approximated by $A_L(u_c, v)$ and an error of $O(h_L\|v\|_X)$ is made. Then $z = \mathcal{G}'_{l,I_l}(u^l)^{-1}\chi_{I_l}\tilde{w}$ is approximated by \tilde{z} , where \tilde{z} is computed by using GMRES to solve $\mathcal{G}'_{l,I_l}(u^l)\tilde{z} = \chi_{I_l}\tilde{w}$ with $G_l(u^l, z)$ used in the matrix-vector product routine whenever $\mathcal{G}'_{l,I_l}(u^l)z$ is requested and an approximate L^2 inner product used in the GMRES routine itself. We use the zero function as the initial iterate to GMRES and request a reduction in the residual by a factor of $\tilde{\rho}$. Following [8] we assume that the GMRES routine returns a vector that satisfies

$$(4.4) \quad \|G_l(u^l, \tilde{z}) - \chi_{I_l}\tilde{w}\|_{2,l} \leq \tilde{\rho}\|\chi_{I_l}\tilde{w}\|_{2,l}.$$

In (4.4) $\|\cdot\|_{2,l}$ is the norm associated with the approximate L^2 norm for level l . In our examples from control problems, $\|\cdot\|_{2,l} = \|\cdot\|_{L^2}$ and the functions at level l

are piecewise linear. In the example from integral equations the approximate inner product is given by a quadrature rule. In either case there are $C_\infty > 0$ and $N_l \rightarrow \infty$ such that

$$(4.5) \quad N_l^{-1} \|\cdot\|_X \leq \|\cdot\|_{2,l} \leq C_\infty \|\cdot\|_X.$$

We can use (4.4) and (4.5) to obtain a uniform estimate

$$(4.6) \quad \|G_l(u^l, \tilde{z}) - \chi_{I_l} \tilde{w}\|_X \leq \tilde{\rho} N_l C_\infty \|\chi_{I_l} \tilde{w}\|_X.$$

Using (4.3) we have

$$\|G_l(u^l, \tilde{z}) - \mathcal{G}'_l(u^l) \tilde{z}\|_X \leq \frac{\gamma h_l}{2} \|\tilde{z}\|_X$$

and hence by Assumption 2.3, if $h_l \leq 2/(M_2 \gamma)$,

$$\|\tilde{z}\|_X \leq \frac{M_2 \|G_l(u^l, \tilde{z})\|_X}{1 - M_2 \gamma h_l / 2} \leq \frac{M_2}{1 - M_2 \gamma h_l / 2} \|\chi_{I_l} \tilde{w}\|_X (1 + \tilde{\rho} N_l C_\infty).$$

Hence if

$$(4.7) \quad \tilde{\rho} N_l C_\infty < 1 \text{ and } h_l \leq 1/(M_2 \gamma)$$

we have

$$\|\tilde{z}\|_X \leq 4M_2 \|\chi_{I_l} \tilde{w}\|_X.$$

Summarizing, if (4.7) holds then

$$\|\mathcal{G}'_l(u^l) \tilde{z} - \chi_{I_l} \tilde{w}\|_X \leq (\tilde{\rho} N_l C_\infty + 2M_2 \gamma h_l) \|\chi_{I_l} \tilde{w}\|_X.$$

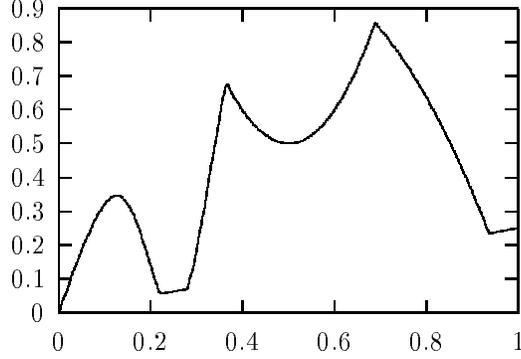
This leads us to the rule of thumb that $h_l = O(\tilde{\rho} N_l)$ as a guide to selection of the accuracy required by a coarse mesh function evaluation. Also we get insight into the tolerances $\tilde{\rho}$ and $\tau_l = O(h_l)$ from (4.7).

If a Newton-GMRES [8] iteration is used the Theorem 3.7 and analysis above provides guidance in the choice of h_c as a function of N_c and ρ_c , where ρ_c is the factor in (3.11). In the computations reported below, we set $\rho_c = \tilde{\rho} = .001$ with a view toward making $\tilde{\rho} N_l C_\infty$ and ρ_n small.

The overall effect of the GMRES approximations is to introduce a relative error of $O(\tilde{\rho} N_l + h_l + h_L)$ into the approximate Newton iteration and if $\tilde{\rho} N_l + h_l + h_L$ is sufficiently small the overall rate of r-linear convergence will be preserved. Approximating \mathcal{F}_L will introduce an absolute error of τ_L and accordingly τ_L should be reduced as the grids are refined to approximate the expected truncation error. Note that it is not necessary to approximate $\mathcal{K}_L(u_c)$, which is used in the evaluation of \mathcal{F}_L , and the perturbations used in the approximation of \mathcal{K}'_L to the same accuracy. We use an accuracy of τ_l for the perturbed evaluation since that does not change the size of the relative error in B_l^{-1} .

As in [16] for an example we consider maps \mathcal{K} defined by

$$\mathcal{K}(u) = -d(t, 1)/\alpha,$$

FIG. 4.1. Plot of u^* 

where for $x \in (0, 1)$ and $t \in (0, T)$,

$$\begin{aligned}
 -d_t &= d_{xx} + f_2(t, x), & d(T, x) &= y(T, x) - z(x), \\
 d_x(t, 0) &= h_l(t), & d_x(t, 1) &= h_r(t), \\
 (4.8) \quad y_t &= y_{xx} + f_1(t, x), & y(0, x) &= y_0(x), \\
 y_x(t, 0) &= g_l(t), & y_x(t, 1) &= g_r(y(t, 1)) + q(t) + u(t).
 \end{aligned}$$

where the right hand sides and boundary values can be used to specify a solution to the problem. Discretization in space was by piecewise linear finite elements. The mesh at level L consisted of the $L + 1$ points $\{i/L\}_{i=0}^L$. The functions \mathcal{K}_L were evaluated by using DASSL to solve the systems of ordinary differential equations obtained by discretizing (4.8) in space and setting the absolute error to τ_L . The sets A_c , R_c , and I_c were specified as in (3.1) with $\bar{p} = 1/2$.

In the computations we set the right hand sides and boundary conditions so that the solution to (4.8) was

$$\begin{aligned}
 y(t, x) &= (1 + xt)/(1 + x^2t) \text{ and} \\
 d(t, x) &= \alpha(1 - x(2x(1 - t) - 1)^2)(1 - \exp(-(t - 1/4)^2)),
 \end{aligned}$$

with

$$u^*(t) = \mathcal{P}(d(t, 1)/\alpha) = \mathcal{P}((1 - (1 - 2t)^2)(1 - \exp(-(t - 1/4)^2))).$$

We imposed constraints given by

$$u_{max}(t) = 1 + 10(t - .5)^2, \text{ and } u_{min}(t) = t/4.$$

The solution to (1.1) is plotted in Figure 4.1. The solution is Lipschitz continuous but not differentiable. Hence we would expect piecewise linear approximations to be first order accurate.

We have constructed the examples in such a way that the solution u^* does not depend on α . The reason for this was to isolate the ill conditioning that arises from

small values of α in the operator and to make it possible to directly compare results for different values of α . We report results for $\alpha = 1, .5, .25, .1$. Since the Lipschitz constants of \mathcal{K} are proportional to $1/\alpha$ and the solution of the linear equation for the step becomes very sensitive to high frequency perturbations of the solution for small α , one would expect that finer coarse meshes and more accurate initial iterates will be required for the smaller values of α . This is the case for both the Atkinson-Brakhage and Newton-GMRES forms of the algorithm.

The maps \mathcal{K}_L were constructed exactly as reported in [16] and the absolute accuracy of the function evaluation at level L was $\tau_L = .001/L$, consistent with the expected first order accuracy. In the forward differencing we used $h_l = .1/\sqrt{L}$. In the computations we used a coarse mesh of $N_l = l + 1$ points. Consistently with our requirement that $h_l = O(\tilde{\rho}N_l)$ we set $\tilde{\rho} = .01$. The iteration at the coarse mesh was terminated when $\|\mathcal{F}_l(u)\|_\infty < \epsilon_l = 10^{-4}$. The iteration at a higher mesh level L was terminated if $\|\mathcal{F}_L(u)\|_\infty < \epsilon_L = .01/L$. For both methods considered, Projected Newton-GMRES and the Atkinson-Brakhage iteration, only one outer iterate was required for termination at the higher mesh levels. After that the mesh spacing was halved and the iteration continued. The initial iterate was

$$u_0 = \mathcal{P}(u(t) + \delta \sin(t)(t - .5)t).$$

δ controls the size of the initial iterate error and was reduced as α is reduced. The number of points in the coarse mesh N_l is increased as α is reduced.

In the tables that follow we report the norm ν_n of \mathcal{F}_L at each iterate n and for the final iterate ($n=1$) at each level the ratio ν_1/ν_0 and the number of GMRES iterates I_G required for (3.11) to hold in the case of Newton-GMRES or needed to solve (4.2) in the case of the Atkinson-Brakhage iteration. In the header for each table we report α , δ , l , and the computation time T_C in seconds.

All computations were done on the CRAY Y-MP at the North Carolina Supercomputing Center running UNICOS 6.0. All codes were written in CRAY FORTRAN cft77 version 5.0.0.0. Computation times were taken from the output of the CRAY hardware performance monitor.

We report on two methods of solution. The first, reported in Tables 4.1, 4.2, 4.3, and 4.4, is a direct Newton-GMRES approach which satisfies (3.11) with a GMRES iteration, using $\rho_c = .001$.

In Tables 4.5, 4.6, 4.7, and 4.8, we report the results of the Atkinson-Brakhage algorithm using $\tilde{\rho} = .001$.

In both methods a fine mesh function evaluation was used to test for termination. A more efficient approach would be to see if the function norm at the next finer mesh has the predicted size of half the initial evaluation at the previous mesh. This would not avoid the fine mesh function evaluation to test for termination at the finest and ultimate mesh. Hence at least four fine mesh function evaluations per level were done, one to compute $\mathcal{K}(u_c)$, one to compute $\mathcal{K}(u^{1/3})$, one to compute $u_+ = \mathcal{P}(\mathcal{K}(u^{2/3}))$, and one to test for termination. The Newton-GMRES required one additional fine mesh evaluation for each inner iterate, while the Atkinson-Brakhage required a low accuracy (*i. e.* using τ_l instead of τ_L) fine mesh function evaluation to compute $\mathcal{K}'_L(u_c)$ via a difference in (4.1) and at most a few coarse mesh evaluations for the GMRES iteration. This accounts for the advantage in the Atkinson-Brakhage method, which, as the tables show, executes in roughly 60% of the time of Newton-GMRES. Note that for the computations reported in the tables that required more GMRES iterations in the inner iteration of Newton-GMRES, the advantage of the Atkinson-Brakhage

TABLE 4.1
Projected Newton-GMRES.

$\alpha = 1.0, \delta = .1, l = 20, T_C = 669$

L	n	I_G	ν_n	ν_n/ν_{n-1}
20	0		0.23E-01	
	1	3	0.26E-03	0.11E-01
40	0		0.70E-01	
	1	3	0.11E-02	0.15E-01
80	0		0.38E-01	
	1	3	0.39E-03	0.10E-01
160	0		0.14E-01	
	1	3	0.59E-05	0.42E-03
320	0		0.63E-02	
	1	3	0.15E-04	0.25E-02
640	0		0.45E-02	
	1	3	0.24E-05	0.54E-03
1280	0		0.28E-02	
	1	3	0.63E-05	0.23E-02
2560	0		0.68E-03	
	1	3	0.31E-07	0.45E-04
5120	0		0.42E-03	
	1	3	0.47E-06	0.11E-02

TABLE 4.2
Projected Newton-GMRES.

$\alpha = .5, \delta = .1, l = 40, T_C = 745$

L	n	I_G	ν_n	ν_n/ν_{n-1}
40	0		0.20E-01	
	1	3	0.53E-03	0.27E-01
80	0		0.37E-01	
	1	4	0.11E-02	0.30E-01
160	0		0.18E-01	
	1	3	0.11E-03	0.63E-02
320	0		0.66E-02	
	1	3	0.54E-04	0.81E-02
640	0		0.45E-02	
	1	3	0.17E-03	0.38E-01
1280	0		0.31E-02	
	1	3	0.16E-04	0.51E-02
2560	0		0.65E-03	
	1	3	0.15E-05	0.23E-02
5120	0		0.42E-03	
	1	4	0.58E-07	0.14E-03

TABLE 4.3
Projected Newton-GMRES.

$$\alpha = .25, \delta = .05, l = 80, T_C = 651$$

L	n	I_G	ν_n	ν_n/ν_{n-1}
80	0		0.18E-01	
	1	3	0.14E-02	0.74E-01
	2	2	0.21E-04	0.16E-01
160	0		0.15E-01	
	1	3	0.63E-03	0.42E-01
320	0		0.59E-02	
	1	3	0.49E-03	0.83E-01
640	0		0.50E-02	
	1	4	0.13E-03	0.26E-01
1280	0		0.29E-02	
	1	3	0.72E-04	0.25E-01
2560	0		0.74E-03	
	1	3	0.49E-05	0.67E-02
5120	0		0.42E-03	
	1	3	0.16E-05	0.37E-02

TABLE 4.4
Projected Newton-GMRES.

$$\alpha = .1, \delta = .025, l = 640, T_C = 730$$

L	n	I_G	ν_n	ν_n/ν_{n-1}
640	0		0.22E-01	
	1	3	0.68E-02	0.33E+00
	2	2	0.29E-03	0.43E-01
1280	0		0.26E-02	
	1	4	0.35E-03	0.14E+00
2560	0		0.85E-03	
	1	3	0.96E-04	0.11E+00
5120	0		0.48E-03	
	1	3	0.27E-05	0.55E-02

method was larger. Having said that, the nested iteration form of Newton-GMRES is still a fast algorithm for compact fixed point problems since the number of inner iterations required at each mesh level is bounded independently of the mesh size.

REFERENCES

- [1] E. L. ALLGOWER, K. BÖHMER, F. A. POTRA, AND W. C. RHEINOLDT, *A mesh-independence principle for operator equations and their discretizations*, SIAM J. Numer. Anal., 23 (1986), pp. 160–169.
- [2] P. M. ANSELONE, *Collectively Compact Operator Approximation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1971.
- [3] K. E. ATKINSON, *Iterative variants of the Nyström method for the numerical solution of integral equations*, Numer. Math., 22 (1973), pp. 17–31.

TABLE 4.5
Atkinson-Brakhage iteration.

$$\alpha = 1.0, \delta = .1, l = 20, T_C = 391$$

L	n	I_G	ν_n	ν_n/ν_{n-1}
20	0		0.23E-01	
	1	3	0.26E-03	0.11E-01
40	0		0.70E-01	
	1	3	0.82E-03	0.12E-01
80	0		0.38E-01	
	1	3	0.17E-03	0.45E-02
160	0		0.14E-01	
	1	2	0.88E-04	0.63E-02
320	0		0.63E-02	
	1	2	0.19E-04	0.30E-02
640	0		0.45E-02	
	1	3	0.14E-05	0.31E-03
1280	0		0.28E-02	
	1	3	0.11E-05	0.38E-03
2560	0		0.69E-03	
	1	3	0.52E-06	0.75E-03
5120	0		0.42E-03	
	1	3	0.14E-06	0.34E-03

TABLE 4.6
Atkinson-Brakhage iteration.

$$\alpha = .5, \delta = .1, l = 40, T_C = 444$$

L	n	I_G	ν_n	ν_n/ν_{n-1}
40	0		0.20E-01	
	1	3	0.53E-03	0.27E-01
80	0		0.37E-01	
	1	3	0.12E-02	0.32E-01
160	0		0.18E-01	
	1	3	0.74E-03	0.42E-01
320	0		0.76E-02	
	1	3	0.14E-03	0.18E-01
640	0		0.48E-02	
	1	3	0.38E-04	0.79E-02
1280	0		0.29E-02	
	1	3	0.14E-04	0.50E-02
2560	0		0.65E-03	
	1	3	0.44E-05	0.67E-02
5120	0		0.42E-03	
	1	3	0.26E-05	0.61E-02

TABLE 4.7
Atkinson-Brakhage iteration.

$$\alpha = .25, \delta = .05, l = 80, T_C = 428$$

L	n	I_G	ν_n	ν_n/ν_{n-1}
80	0		0.18E-01	
	1	3	0.14E-02	0.74E-01
	2	2	0.21E-04	0.16E-01
160	0		0.15E-01	
	1	3	0.43E-03	0.28E-01
320	0		0.67E-02	
	1	3	0.25E-02	0.37E+00
640	0		0.69E-02	
	1	2	0.21E-03	0.30E-01
1280	0		0.30E-02	
	1	2	0.24E-03	0.82E-01
2560	0		0.60E-03	
	1	2	0.57E-05	0.96E-02
5120	0		0.42E-03	
	1	2	0.70E-06	0.17E-02

TABLE 4.8
Atkinson-Brakhage iteration.

$$\alpha = .1, \delta = .025, l = 440, T_C = 446$$

L	n	I_G	ν_n	ν_n/ν_{n-1}
640	0		0.22E-01	
	1	3	0.68E-02	0.33E+00
	2	2	0.29E-03	0.43E-01
1280	0		0.26E-02	
	1	3	0.15E-03	0.60E-01
2560	0		0.67E-03	
	1	4	0.24E-04	0.35E-01
5120	0		0.42E-03	
	1	4	0.83E-05	0.20E-01

- [4] D. B. BERTSEKAS, *On the Goldstein-Levitin-Polyak gradient projection method*, IEEE Trans. Autom. Control, (1976), pp. 174–184.
- [5] ———, *Projected Newton methods for optimization problems with simple constraints*, SIAM J. Control Optim., 20 (1982), pp. 221–246.
- [6] H. BRAKHAGE, *Über die numerische Behandlung von Integralgleichungen nach der Quadraturformelmethode*, Numer. Math., 2 (1960), pp. 183–196.
- [7] K. E. BRENAN, S. L. CAMPBELL, AND L. R. PETZOLD, *The Numerical Solution of Initial Value Problems in Differential-Algebraic Equations*, Elsevier, New York, 1989.
- [8] P. N. BROWN AND Y. SAAD, *Hybrid Krylov methods for nonlinear systems of equations*, SIAM J. Sci. Stat. Comp., 11 (1990), pp. 450–481.
- [9] R. DEMBO, S. EISENSTAT, AND T. STEihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408.
- [10] W. HACKBUSCH, *On the fast solving of parabolic boundary control problems*, SIAM J. Control Optim., 17 (1979), pp. 231–244.
- [11] ———, *Multi-Grid Methods and Applications*, vol. 4 of Springer Series in Computational Math-

- ematics, Springer-Verlag, New York, 1985.
- [12] M. HEINKENSCHLOSS, C. T. KELLEY, AND H. T. TRAN, *Fast algorithms for nonsmooth compact fixed point problems*. SIAM J. on Numerical Analysis, to appear, 1992.
 - [13] L. KANTOROVICH AND G. AKILOV, *Functional Analysis in Normed Spaces*, Pergamon Press, New York, 1964.
 - [14] C. T. KELLEY, *Operator prolongation methods for nonlinear equations*, in Computational Solution of Nonlinear Systems of Equations, E. L. Allgower and K. Georg, eds., vol. 26 of AMS Lectures in Applied Mathematics, American Mathematical Society, Providence, RI, 1990, pp. 359–388.
 - [15] C. T. KELLEY AND E. W. SACHS, *Multilevel algorithms for constrained optimal control problems*. proceedings of Copper Mountain Conference on Iterative Methods, Copper Mountain, Co., 1992.
 - [16] ———, *Fast algorithms for compact fixed point problems with inexact function evaluations*, SIAM J. on Sci. Stat. Comp., 12 (1991), pp. 725–742.
 - [17] ———, *Mesh independence of the gradient projection method for optimal control problems*, SIAM J. Control and Optimization, 30 (1992), pp. 477–493.
 - [18] N. M. NACHTIGAL, S. C. REDDY, AND L. N. TREFETHEN, *How fast are nonsymmetric matrix iterations*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 778–795.
 - [19] Y. SAAD AND M. SCHULTZ, *GMRES a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comp., 7 (1986), pp. 856–869.
 - [20] E. SACHS, *A parabolic control problem with a boundary condition of the Stefan-Boltzmann type*, ZAMM, 58 (1978), pp. 443–449.
 - [21] J. P. YVON, *Contrôle optimal d'un four industriel*, tech. rep., University of Paris, 1973.