PSEUDO-TRANSIENT CONTINUATION FOR NONSMOOTH NONLINEAR EQUATIONS *

K. R. KAVANAGH† AND C. T. KELLEY *

Abstract. Pseudo-transient continuation is a Newton-like iterative method for computing steady-state solutions of differential equations in cases where the initial data is far from a steady state. The iteration mimics a temporal integration scheme, with the time step being increased as steady state is approached. The iteration is an inexact Newton iteration in the terminal phase.

In this paper we show how steady-state solutions to certain ordinary and differential algebraic equations with nonsmooth dynamics can be computed with the method of pseudo-transient continuation. An example of such a case is a discretized partial differential equation with a Lipschitz continuous, but non-differentiable, constitutive relation as part of the nonlinearity. In this case we can approximate a generalized derivative with a difference quotient.

The existing theory for pseudo-transient continuation requires Lipschitz continuity of the Jacobian. Newton-like methods for nonsmooth equations have been globalized by trust-region methods, smooth approximations, and splitting methods in the past, but these approaches require problem-specific components in an algorithm. The method in this paper addresses the nonsmoothness directly.

Key words. Pseudo-transient Continuation, Nonlinear equations, Semismooth functions, Clarke differential

AMS subject classifications. 65H10, 65H20, 65L05,

1. Introduction. In this paper we show how pseudo-transient continuation (Ψtc) can be used to solve a class of nonsmooth nonlinear equations. Ψtc is a predictor-corrector method for efficient integration of a time-dependent differential equation to steady state. The objective of the method is not temporal accuracy, but rather to resolve the transient behavior of the solution until the iteration is close to steady state, and then to increase the "time step" and transition to a fast Newton-like method.

In this paper we extend the theoretical convergence results of [7, 18] to problems with certain nonsmooth nonlinearities and, thereby, partially explain the results reported in [8, 10]. We also show how generalized derivatives can be approximated by finite differences, and how those approximate derivatives can be used effectively both in locally convergent iterations, such as those which arise in temporal integration, and in the context of Ψtc . This aspect of the work is motivated by several papers on simulation of unsaturated flow, [10, 14, 15, 24, 30, 31], in which Lipschitz continuous spline approximations to the non-Lipschitz continuous van Geneuchten and Mualem [25, 33] constitutive laws are used. These nonsmooth functions are then differentiated with finite differences as if they were smooth. The results in this paper explain the success reported in those papers. Another aspect of the paper is an extension of the local results in [9, 21, 27, 28].

Ψtc methods are particularly appropriate for the types of nonsmooth nonlinearities which we discuss in this paper. Traditional methods for globalizing iterative methods for nonlinear equa-

tions, such as line searches, can fail as commonly implemented in practice for both smooth and nonsmooth equations [7, 8, 18]. The existing global convergence results for nonsmooth nonlinear equations are either based on line searches for a inexact Newton formulation [9, 22], a sequence of smooth approximations [4, 29], or explicit treatment of the nonsmoothness [20]. Only the latter admits approximation of the generalized Jacobian by a difference, Ψtc allows one to deal with the nonsmoothness directly and exploits the dynamics to guarantee convergence to x^*, the solution one wants.

In the remainder of this introductory section we review the relevant results from nonsmooth analysis (§ 2.1) and Ψtc (§ 2.2). Then we describe the setting for the new results. In § 3 we show how finite difference approximations of generalized Jacobians affect the local convergence of inexact Newton methods for nonsmooth problems. We use those results in § 4, where we state and prove our local and global convergence results for Ψtc . We present a numerical example in § 5.

Some extensions of our results to infinite dimensions are possible, using ideas from [5, 11, 12, 19, 32] if the appropriate compactness conditions hold. These extensions will be explored in a subsequent paper.

2. Previous results. In this section we review the prior results about Ψtc and nonsmooth analysis that we will need for this paper.

In this paper the norm will be the a scaled discrete l^2 norm on R^N,

$$\|w\| = \frac{1}{\sqrt{N}}\|w\|_2,$$

unless stated otherwise. The ball of radius ϵ about a point $x \in R^N$ will be denoted

$$\mathcal{B}(x, \epsilon) = \{z \mid \|x - z\| < \epsilon\}.$$

As is standard, we will let x^* denote solution of $F(x) = 0$, and

$$e = x - x^*$$

the error. We will let $(x)_i$ denote the ith component of the vector x.

2.1. Nonsmooth Analysis. In this section we review the concepts from nonsmooth analysis [6, 23] which we will need for our convergence results. We then state the local convergence result from [21, 28] which extend in this paper.

Let $F : R^N \to R^N$ be locally Lipschitz continuous. This implies that F is Fréchet differentiable almost everywhere, and that the directional derivatives

$$F'(x : w) = \lim_{h \to 0} \frac{F(x + hw) - F(x)}{h}$$

exist for all $x, w \in R^N$.

We let D_F denote the set of points where F is Fréchet differentiable. The generalized Jacobian [6] of F at $u \in R^N$ is the set

(2.1) $$\partial F(x) = \text{co}\left\{\lim_{x_j \to x; x_j \in D_F} F'(x_j)\right\},$$

where co denotes the convex hull.

We will consider extensions of the Newton-like iteration

$$(2.2) \qquad x_{n+1} = x_n - V_n^{-1} F(x_n)$$

where, $V_n \in \partial F(x_n)$, and, as is standard x_n is the current approximation to a solution x^* and x_{n+1} the new approximation. Our results will be stated in terms of an inexact formulation,

$$(2.3) \qquad x_{n+1} = x_n + s,$$

where

$$(2.4) \qquad \|V_n s + F(x_n)\| \leq \eta_n \|F(x_n)\|,$$

and $V_n \in \partial F(x_n)$.

The most important concept is that of semismoothness [23, 28].

DEFINITION 2.1. *F is semismooth at* $x \in R^N$ *if F is locally Lipschitz continuous and for all* $w \in R^N$, *the limit*

$$(2.5) \qquad \lim_{V \in \partial F(x+tw'), w' \to w, t \downarrow 0} \{V w'\}$$

exists.

Semismoothness is a useful concept [5, 28, 32] in the proofs of convergence and local convergence rates of the iteration (2.2). In the standard theory for Lipschitz continuously differentiable F, local quadratic convergence follows from nonsingularity of the Jacobian $F'(x^*)$ at the solution and the Lipschitz continuity of F'. In the nonsmooth case, one must prove that the Newton iteration is well defined and quantify the degree of nonsmoothness to obtain convergence rates.

Lemma 2.2, taken from [27], and Lemma 2.4, are the results that are needed to prove local superlinear convergence of (2.2).

LEMMA 2.2. *F is semismooth at* $x \in R^N$ *if and only if*

$$(2.6) \qquad \lim_{w \to 0, V \in \partial F(x+w)} \frac{\|F(x+w) - F(x) - V w\|}{\|w\|} = 0.$$

To obtain convergence rates one needs a stronger condition than semismoothness [28].

DEFINITION 2.3. *F is semismooth of order* p *at* x *if for all* $w \in R^N$ *and* $V \in \partial F(x+w)$

$$(2.7) \qquad F(x+w) - F(x) - V w = O(\|w\|^{1+p})$$

as $w \to 0$.

LEMMA 2.4. *Let F be semismooth,* $F(x^*) = 0$, *and assume that all matrices in* $\partial F(x^*)$ *are nonsingular. Then there are* M *and* Δ *such that if* $x \in \mathcal{B}(x^*, \Delta)$ *and* $V \in \partial F(x)$, *then* $\|V^{-1}\| \leq M$.

These results have been used to prove several convergence theorems [9, 21, 27, 28] for (2.2) and (2.3). Theorem 2.5 is a combination of the local convergence results, and is the basis for the new algorithms in this paper.

THEOREM 2.5. *Let* $F : R^N \to R^N$ *with* $F(x^*) = 0$. *Assume that F is semismooth at* x^*. *Then there are* $\bar{\eta}, \bar{\delta}, K > 0$ *such that if* $x_0 \in \mathcal{B}(x^*, \bar{\delta})$ *and* $\eta_n \leq \bar{\eta}$ *then the inexact Newton iteration (2.3) converges to* x^* *and*

$$\|e_{n+1}\| \leq K \eta_n \|e_n\| + o(\|e_n\|).$$

Moreover, if F is semismooth of order p at x^*, then

$$\|e_{n+1}\| \leq K(\eta_n \|e_n\| + \|e_n\|^{1+p}).$$

In previous work [11,16,19,20] on nonsmooth nonlinear equations in function spaces and their discretizations, we used properties of the solution to isolate a smooth component of the nonlinearity. Each problem required a different approach, and all assumed that the nonsmooth component was small. In those papers, mesh-independent convergence results were obtained, and standard implementations of matrix-free Newton-Krylov methods worked well.

The formulation we consider in this paper is different. Here one does not have to explicitly split the operator into smooth and nonsmooth parts, a significant advantage in complicated applications [30]. However, we know of no general proofs of mesh-independent convergence rates, a problem also mentioned in [32]. In fact, the numerical results in § 5 show mesh-dependent performance of the iteration, especially in the mid-range. Mesh-dependent convergence was also reported in [4]. In § 5.3 we illustrate this phenomenon and show how a nested iteration can overcome it.

Numerical differentiation, a topic we consider in § 3.1, is a simple matter if the smooth and nonsmooth parts of the nonlinearity are split. Here, we can prove accuracy only if one is differentiating in coordinate directions, and only then for special classes of operators. Since the directions in a matrix-free Newton-Krylov solver are not predictable, our results do not apply to those methods.

2.2. Pseudo-Transient Continuation. The objective of Ψtc , as we present it here, is to find the steady state solution of the semi-explicit index-one differential algebraic equation (DAE)

$$(2.8) \qquad D \left(\begin{array}{c} u \\ v \end{array} \right)' = - \left(\begin{array}{c} f(u,v) \\ g(u,v) \end{array} \right) = -F(x), \quad x(0) = x_0.$$

Here $x = (u,v)^T \in C([0,\infty], R^{N_1+N_2})$. The functions $u : [0,\infty] \to R^{N_1}$ and $v : [0,\infty] \to R^{N_2}$ are to be found. The differential variables u and the algebraic variables v are clearly separated in the semi-explicit case where

$$D = \left(\begin{array}{cc} D_{11} & 0 \\ 0 & 0 \end{array} \right),$$

where D_{11} is a nonsingular scaling matrix. A good general reference for DAEs is [3].

We assume that the initial data for (2.8) are consistent (*i. e.* $g(u(0), v(0)) = 0$) and seek the solution x^* to $F(x^*) = 0$ that satisfies

$$\lim_{t \to \infty} x(t) = x^*.$$

If (2.8) is a discretization in space of a PDE, and the initial data is far from the desired steady state, the application of a conventional method, such as a line search [17], to the time-independent equation

$$F(x) = 0,$$

may fail to converge. Possible failure modes [8] are stagnation of the iteration at a singularity of F', the Jacobian of F, or or finding a solution other than x^*.

We formulate Ψtc as

(2.9)
$$x_{n+1} = x_n - (\delta_n^{-1}D + F'(x_n))^{-1}F(x_n).$$

In (2.9), $\{\delta_n\}$ is adjusted to efficiently find the steady state solution rather than to enforce temporal accuracy.

The convergence results in [7, 18] assume that the time step was updated with "switched evolution relaxation" (SER) [26], *i. e.*

(2.10)
$$\delta_n = \max\left(\delta_{n-1}\frac{\|F(x_{n-1})\|}{\|F(x_n)\|}, \delta_{max}\right).$$

In [7] we proved convergence for smooth F under the assumptions that the DAE has index one in a certain uniform sense, has a global solution in time, and that the solution converges to a steady state. The convergence result for the exact $\eta = 0$ case is

(2.11)
$$\|x_{n+1} - x^*\| = O(\|x_n - x^*\|(\delta_{max}^{-1} + \|x_n - x^*\|))$$

as $n \to \infty$.

In this paper we relax the smoothness assumptions and consider the iteration

(2.12)
$$x_{n+1} = x_n + s,$$

where
(2.13)
$$\|(\delta_n^{-1}D + V(x_n))s + F(x_n)\| \le \eta_n\|F(x_n)\|$$

where $V(x_n)$ is near to the set $\partial F(x_n)$ the sense that

(2.14)
$$V(x_n) \in \mathcal{D}(x_n, C, h), \text{ for some small } h.$$

where

(2.15) $\mathcal{D}(x_n, C, h) = \{V \mid \|V - \bar{V}\| \le Ch, \text{for some } \bar{V} \in \partial F(\bar{x}) \text{ and } \|V - \bar{V}\| \le Ch\}.$

The sense in which $V(x_n)$ is close to $\partial F(x_n)$ is technical because $\partial F(x)$ is not continuous in x. The requirement that $V(x_n) \in \mathcal{D}(x_n, C, h)$ is, in a sense, a requirement that a combination of the forward and backward error be small.

In [7, 8, 10, 18] we report on compuational results that show that both the local and global phases of the Ψtc iteration perform as (2.11) predicts even if the nonlinearity is nonsmooth [7,8,10] and the derivative is approximated by differencing [10, 13, 14, 30].

3. Local Convergence Theory. In this section we analyze the accuracy of a finite difference approximation of a generalized Jacobian in the case where the nonsmoothness arises from a substitution operator. The approximation is accurate in a combined forward and backward sense, and this affects not only the convergence speed of an inexact Newton iteration, but also the limiting accuracy.

3.1. Finite difference approximations. The results in this paper are motivated in part by our experience with nonsmooth nonlinear substitution operators. A substitution operator on R^N has the form

(3.1) $$\Phi(x) = (\phi(x_1), \ldots, \phi(x_N))^T$$

where $\phi : R \to R$. The generalized Jacobian of Φ is the set of diagonal matrices

$$\partial\Phi(x) = (\partial\phi(x_1), \ldots, \partial\phi(x_N))^T.$$

In this section, we consider maps that are compositions of smooth maps with substitution operators. Let

$$G(x) = S(\Phi(x))$$

where S is differentiable. The smoothness of S and the definition (2.1) of ∂G imply that

$$\partial G(x) = S'(\Phi(x))\partial\Phi(x)$$

where S' is the Jacobian of S. Of interest here is the approximation of ∂G with a finite difference approximation using the coordinate directions.

Let $\partial_h^F G(x)$ be the matrix whose ith column is

$$\frac{G(x + h1_i) - G(x)}{h},$$

where 1_i is the unit vector in the ith coordinate direction. We show that the forward difference $\partial_h^F F$ approximates $\partial G(x)$ in the sense described by (2.15).

THEOREM 3.1. *Let* $\phi : R \to R$ *be Lipschitz continuous and differentiable except at finitely many points* $\{\xi_i\}_{i=1}^M$. *Let* S *be Lipschitz continuously differentiable in* R^N. *Then there is* $C > 0$ *such that for all* h *sufficiently small*

(3.2) $$\partial_h^F G(x) \in \mathcal{D}(x, C, h).$$

Proof. We begin by showing that it suffices to prove the result for scalar functions. Differentiability of S and the Lipschitz continuity of ϕ imply that

$$G(x + h1_i) - G(x) = S'(\Phi(x))(\Phi(x + h1_i) - \Phi(x)) + O(h^2)$$

$$= S'(\Phi(\bar{x}))(\Phi(x + h1_i) - \Phi(x)) + O(h^2)$$

for all $\bar{x}$ such that $\|\bar{x} - x\| \le h$. Hence we need only prove the result for substitution operators.

Since Φ is a substitution operator, the ith component of $\Phi(x + h1_i) - \Phi(x)$ is

$$\phi((x)_i + h) - \phi((x)_i),$$

and we need only consider scalar functions. Now let,

$$h < \min_{i,j} \|\xi_i - \xi_j\|_\infty,$$

then at most one ξ is in the interval $[(x)_i, (x)_i + h]$. If ϕ is differentiable in the interval $[(x)_i, (x)_i + h]$ then

$$\frac{\phi((x)_i + h) - \phi((x)_i)}{h} = \phi'((x)_i) + O(h)$$

and we let the ith component of $\bar{x}$ be $(x)_i$.

Now assume that $\xi_j \in [(x)_i, (x)_i + h]$ for some j. Let $\phi'_+(\xi_j)$ and $\phi'_-(\xi_j)$ be the right and left handed derivatives at ξ_j

$$\phi'_\pm(\xi_j) = \lim_{h \to 0} \frac{\phi(\xi_j \pm h) - \phi(\xi_j))}{\pm h}.$$

Let $\xi_j - (x)_i = \nu h$, for $\nu \in [0, 1]$. Since

$$\phi((x)_i + h) - \phi((x)_i) = \phi(\xi_j + (1 - \nu)h) - \phi(\xi_j - \nu h)$$

$$= \phi(\xi_j + (1 - \nu)h) - \phi(\xi_j) + \phi(\xi_j) - \phi(\xi_j - \nu h)$$

$$= (1 - \nu)\phi'_+(\xi_j) + \nu\phi'_-(\xi_j) + O(h^2).$$

Since

$$(1 - \nu)\phi'_+(\xi_j) + \nu\phi'_-(\xi_j) \in \partial\phi(\xi_j)$$

for all $\nu \in [0, 1]$, the proof is complete with $\bar{x}_i = \xi_j$. $\square$

A similar result holds for central differences. Let $\partial_h^C G(x)$ be the matrix whose ith column is

$$\frac{G(x + h1_i) - G(x - h1_i)}{2h}.$$

If S is Lipschitz continuously twice differentiable and ϕ is piecewise Lipschitz continuously twice differentiable, then the statement of Theorem 3.1 with

$$\|V - \bar{V}\| \le Ch$$

in (2.15) replaced by

(3.3) $$\|V - \bar{V}\| \le Ch^2.$$

3.2. Local Convergence. If the generalized Jacobian is approximated by a finite differece, one cannot expect asymptotic convergence, because the accuracy in the terminal phase of the iteration will be limited by the accuracy in the derivative. We quantify this in Theorem 3.2, which extends the existing local convergence theorems for inexact Newton methods for semi-smooth equations. The new assumption that $V(x) \in \mathcal{D}(x, C, h)$ is motivated by the results in § 3.1.

THEOREM 3.2. *Assume that F is semismooth at x^*, $F(x^*) = 0$, and that all matrices in $\partial F(x^*)$ are nonsingular. Assume that there is $C > 0$ such that*

(3.4) $$V(x) \in \mathcal{D}(x, C, h)$$

for all x sufficiently near x^*.

Then there is δ such that if $x_0 \in \mathcal{B}(x^*, \delta)$, $\{\eta_n\}$ and h are sufficiently small, then the iteration

(3.5) $$x_{n+1} = x_n + s,$$

where

(3.6)
$$\|V(x_n)s + F(x_n)\| \le \eta \|F(x_n)\|,$$

converges to x^*. *Moreover, there is* $K > 0$ *such that*

(3.7)
$$\|e_{n+1}\| \le K((\eta_n + h)\|e_n\| + h) + o(\|e_n\|),$$

or, if F *is is semismooth of order* p *at* x^*, *then*

(3.8)
$$\|e_{n+1}\| \le K((\eta_n + h)\|e_n\| + \|e_n\|^{1+p} + h).$$

Proof. The plan of the proof is to compare x_{n+1} with the Newton iteration from $\bar{x}_n$, where $\bar{x}_n$ is the point specified in the definition of $\mathcal{D}$. We can then apply Theorem 2.5.

Let δ and h be small enough so that

(3.9)
$$\|V^{-1}\| \le M, \text{ for all } u \in \mathcal{B}(u^*, h + \delta),$$

which we can do by Lemma 2.4. We assume that $x_n \in \mathcal{B}(x^*, \delta)$ and will, reducing δ and h if necessary, show that $x_{n+1} \in \mathcal{B}(x^*, \delta)$ and that

By assumption, there are $\bar{x}_n \in \mathcal{B}(x_n, h)$ and $\bar{V}_n \in \partial F(\bar{x}_n)$ such that

$$\|V(x_n) - \bar{V}_n\| \le Ch.$$

Hence the step s is nearly an inexact Newton step from $\bar{x}_n$.

By (3.9), for h sufficiently small,

$$\|V(x_n)^{-1}\| \le 1/(M^{-1} - Ch) \le 2M$$

and hence

$$\|s\| \le 2M(\eta_n + 1)\|F(x_n)\|.$$

Therefore,

$$\|\bar{V}_n s + F(\bar{x}_n)\| \le \|V(x_n)s + F(\bar{x}_n)\| + Ch\|s\|$$

(3.10)
$$\le \|V(x_n)s + F(x_n)\| + \|F(x_n) - F(\bar{x}_n)\| + Ch\|s\|$$

$$\le \eta_n\|F(x_n)\| + Lh + Ch(2M(1 + \eta_n)\|F(x_n)\|,$$

where L is the Lipschitz constant of F. Since

$$\|F(x_n)\| \le \|F(\bar{x}_n)\| + Lh,$$

we may set

$$K_0 = 1 + 2L + 2MC(1 + \eta_n) \le 1 + 2L + 4MC,$$

and obtain

(3.11)
$$\|\bar{V}_n s + F(\bar{x}_n)\| \le K_0((\eta_n + h)\|F(\bar{x}_n)\| + h).$$

The inexact Newton condition (3.6) and (3.11) imply that

$$x_{n+1} = \bar{x}_n + \bar{V}_n^{-1}(F(\bar{x}_n) + r_n)$$

with

$$\|r_n\| \le K_0((\eta_n + h)\|F(\bar{x}_n)\| + h).$$

Hence,

$$\|e_{n+1}\| = \|\bar{e}_n + \bar{V}_n^{-1}F(\bar{x}_n) + \bar{V}_n^{-1}r_n\|$$

$$\le MK_0((\eta_n + h)\|F(\bar{x}_n)\| + h) + o(\|\bar{e}_n\|).$$

If F is semismooth of order p at x^*, Theorem 2.5 implies that there is $K_1 > 0$ such that

$$\|e_{n+1}\| = \|\bar{e}_n + \bar{V}_n^{-1}F(\bar{x}_n) + \bar{V}_n^{-1}r_n\|$$

$$\le K_1\|\bar{e}_n\|^{1+p} + MK_0(\eta_n\|F(\bar{x}_n)\| + h).$$

Since $\|F(\bar{x}_n)\| \le L\|\bar{e}_n\|$ and $\|\bar{e}_n\| \le \|e_n\| + h$, we obtain (3.8) with

$$K = 2K_1 + MK_0(1 + L),$$

and complete the proof. □

3.3. Optimal choice of h. If $\bar{x}_n \ne x_n$, then the estimate (3.7) and (3.8) do not imply convergence, but stagnation once the error is $O(h)$. This is analogous to convergence results [17] for Newton's method when there are errors, such as floating point roundoff, in the evaluation of F. In this case, however, h is larger than floating point roundoff, and we can combine Theorems 3.1 and 3.2 to estimate the optimal choice of h.

Suppose that F is piecewise C^1 (and hence semismooth of order 1 [23]) and can be evaluated up to an absolute error of ϵ_F. If we incorporate the error in F into the result of Theorem 3.1 in the standard way [17] we obtain

$$\partial_h^F G(x) \in \mathcal{D}(x, C', h')$$

where $h' = O(h + \epsilon_F/h)$. Then the estimate (3.8) becomes

$$(3.12) \qquad \|e_{n+1}\| \le K((\eta_n + h + \epsilon_F/h)\|e_n\| + \|e_n\|^2 + h).$$

If we solve the equation for the step exactly, then $\eta_n = 0$. In that case, if $\|e_n\| = O(h^{1/2})$, then

$$(3.13) \qquad \|e_{n+1}\| = O\left(\frac{\epsilon_F}{h^{1/2}} + h\right).$$

The term on the right of (3.13) is minimized when

$$(3.14) \qquad h = O(\epsilon_F^{2/3}).$$

If, for example, $\epsilon_F \approx 10^{-15}$ is double precision floating point roundoff, (3.14) would say that the best results would be obtained if $h \approx 10^{-10}$, rather that 10^{-8} as a conventional analysis would predict. We provide numerical evidence for this in § 5.2.

4. Convergence of Ψtc . The analysis of Ψtc in this paper follows the pattern of [7, 18], considering the iteration in two phases. For phase one, the initial or global phase, we show that Ψtc is a consistent convergent scheme for integration of the DAE. The scheme will be first order if F is semismooth of order 1, order $1 + p$ if F is semismooth of order $p < 1$, and convergent, but with no order, if F is merely semismooth.

From the analysis of the global phase we will conclude that, for sufficiently small δ_0, the iteration will approach x^*. For the second local phase of the iteration, we show that if x is near x^* and $\{\delta_n\}$ is bounded away from zero, then $\delta_n \to \delta_{max}$ and then the terminal phase of convergence can be described by the results in § 3.

The analysis of the local phase does not depend on the dynamics, and we will defer the detailed assumptions on the DAE until § 4.2.

4.1. Local Phase. We consider the local phase first, as we did in [7, 18], in order to establish targets for the integration in the global phase. We seek to find ϵ_L so that if $x_0 \in \mathcal{B}(x^*, \epsilon_L)$ and $\{\delta_n\}$ is bounded away from zero, then $\{x_n\}$ and $\{\delta_n\}$ in (2.12) satisfy $x_n \to x^*$ and $\delta_n \to \delta_{max}$ or

The local convergence rates in the terminal phase depend on

ASSUMPTION 4.1. *F is semismooth at x^*. There are $C, h, \beta, \epsilon_L > 0$ such that for all $x \in \mathcal{B}(x^*, \epsilon_L)$ and all $\delta > 0$*

$$\|(D + \delta V(x))^{-1}D\| \le 1/(1 + \beta\delta),$$

and

$$V(x) \in \mathcal{D}(x, C, h).$$

THEOREM 4.1. *Let the assumptions of Theorem 3.2 and Assumption 4.1 hold. Let $\{\delta_n\}$ be given by (2.10). Then there are C_T and ϵ_T such that if $\{\eta_n\}$ is sufficiently small and $x_0 \in \mathcal{B}(x^*, \epsilon_T)$, then either $\inf_n \delta_n = 0$ or $\delta_n \to \delta_{max}$, the Ψtc iteration converges, and, for n sufficiently large*

(4.1) $$\|e_{n+1}\| \le C_T((\eta_n + \delta_n^{-1} + h)\|e_n\| + h) + o(\|e_n\|)$$

or, if F is semismooth of order p,

(4.2) $$\|e_{n+1}\| \le C_T(\|e_n\|^{1+p} + (\eta_n + \delta_n^{-1} + h)\|e_n\| + h).$$

Proof. We assume that x_0 is near enough to x^* so that the conclusions of Theorem 3.2 hold. If $x_n \in \mathcal{B}(\epsilon_T)$, then following the proof of Theorem 3.2,

$$e_{n+1} = e_n - (\delta_n^{-1}D + \bar{V}_n)^{-1}F(\bar{x}_n) + r_n$$

where

$$\|r_n\| = O((\eta_n + h)\|F(\bar{x}_n)\| + h).$$

Semismoothness and our assumptions imply that

$$F(\bar{x}_n) - \bar{V}_n e_n = O(h) + o(\|e_n\|)$$

and hence

$$e_{n+1} = e_n - (\delta_n^{-1}D + \bar{V}_n)^{-1}\bar{V}_n e_n + R_n$$

$$= (\delta_n^{-1}D + \bar{V}_n)^{-1}\delta_n^{-1}D e_n + R_n,$$

where

$$R_n = O((\eta_n + h)\|F(\bar{x}_n)\| + h) + o(\|e_n\|).$$

If $\delta_n > \delta^*$ for all n, then Assumption 4.1 implies that

$$\|(\delta_n^{-1}D + \bar{V}_n)^{-1}\delta_n^{-1}D\| < 1/(1 + \beta\delta^*).$$

This implies that the iteration is q-linearly convergent, and hence $\delta_n \to \delta_{max}$ and $x_n \to x^*$.

The completion of the proof for large δ_n is a direct consequence of the Theorem 3.2, since the inexact Newton condition

$$\|(\delta_n^{-1}D + V(x_n))s + F(x_n)\| \leq \eta_n\|F(x_n)\|$$

implies that there is C_h such that

$$\|V(x_n)s + F(x_n)\| \leq (\eta_n + C_h h)\|F(x_n)\| + \delta_n^{-1}\|D_{11}\|\|s\|,$$

and then C and η_n in (3.10) can be replaced by $C + \|D_{11}\|$ and $\eta_n + \delta_n^{-1} + C_h h$. This implies convergence if δ_{max} is sufficiently large. □

4.2. Global Phase. In the analysis of the global phase we must assume that the Ψtc iteration is, for small δ, a stable explicit method for the DAE (2.8). To do this we must assume that the DAE is consistent and has index one. In the smooth case, one can express this in terms of the nonsingularity of g_v, the Jacobian of g with respect to the algebraic variables. In the nonsmooth case, however, one must take the limit in (2.1) in all components together. This means that the index assumption is more technical, using the nonsingularity of the matrix pencil $\delta^{-1}D + V(x)$ in part 5 of Assumption 4.2.

We assume that $V(x) \in \mathcal{D}(x, C, h)$, for a sufficiently small h. We decompose operators $V \in \partial F$ into blocks

$$(4.3) \qquad V(x) = \begin{pmatrix} V_{uu} & V_{uv} \\ V_{vu} & V_{vv} \end{pmatrix},$$

where $V_{uu} \in \partial_u f, \dots, V_{vv} \in \partial_v g$.

With this in mind we can formulate our assumptions on the dynamics. Define a neighborhood of the trajectory from x_0

$$(4.4) \qquad S(\epsilon) = \{z \mid \inf_{t \geq 0} \|z - x(t)\| \leq \epsilon\}.$$

ASSUMPTION 4.2. $g(u_0, v_0) = 0$, i. e. *the initial values (u_0, v_0) are consistent.*
There are $\epsilon_G \in (0, \epsilon_T/2)$, where ϵ_T is the radius from Theorem 4.1, such that
1. *F is semismooth in $S(\epsilon_G)$.*
2. *$V_{vv}(x)$ is nonsingular for all $x \in S(\epsilon_G)$, and there is M_V such that $\|V_{vv}(x)^{-1}\| \leq M_V$ for all $x \in S(\epsilon_S)$.*

3. *For all $z_0 \in S(\epsilon_S)$, the solution of $Dz' = -F(z), \quad z(0) = z_0$ exists, $z(t) \in S(\epsilon_G)$ for all t, and $\lim_{t \to \infty} z(t) = x^*$.*

4. *$V(x) \in \mathcal{D}(x, C, h)$ for all $x \in S(\epsilon_G)$.*

5. *Moreover, there are $M_D, M_I > 0$ such that for all $\delta > 0$,*
    (a) $(\delta^{-1}D + V(x))$ is nonsingular for all $x \in S(\epsilon_G)$,
    (b) $\|(\delta^{-1}D + V(x))\| \leq M_D$ for all $x \in S(\epsilon_G)$, and
    (c) $\|(\delta^{-1}D + V(x))^{-1}\| \leq M_I$ for all $x \in S(\epsilon_G)$.

We analyze the global phase by simply showing that the global truncation error of the scheme

$$x_{n+1} = x_n - (\delta_n^{-1}D + V(x_n))^{-1}F(x_n),$$

is of order p, ie

$$\|x_n - x(t_n)\| = O(\delta^p)$$

where $\delta = max_{0 \leq m \leq n}\delta_n$. This will imply, similarly to [7, 18], that the Ψtc iteration will correctly track the solution until x_n is in the ball of local convergence required by Theorem 4.1.

We will use a simple consequence of pth order semismoothness.

LEMMA 4.2. *Let Assumption 4.2 hold. Let $x(t)$ be the solution to (2.8). Let $\delta > 0$ and let*

$$\sigma = \begin{pmatrix} \sigma_u \\ \sigma_v \end{pmatrix} = x(t + \delta) - x(t).$$

Then, for δ, h sufficiently small,

(4.5)
$$(\delta^{-1}D + V(x(t)))\sigma = -F(x(t)) + O(h) + o(\delta),$$

and if F is semismooth of order p,

(4.6)
$$(\delta^{-1}D + V(x(t)))\sigma = -F(x(t)) + O(\delta^{1+p} + h),$$

uniformly in t.

Proof. In the interests of brevity, We will give the proof for $h = 0$ and F semismooth of order p. The analysis for $h > 0$ and semismooth F follows the outlines of the proofs of Theorems 3.2 and 4.1.

Write $x(t) = (u(t), v(t))^T$. By integrating the DAE (2.8), we see that u is a Lipschitz continuous function of t. The nonsingularity of V_{uu} implies that v is also a Lipschitz continuous function of t. This Lipschitz continuity implies that

$$\|\sigma\| = O(\delta).$$

Integrate (2.8) over the interval $[t, t + \delta]$ and use the Lipschitz continuity of F to obtain

(4.7)
$$D\sigma = -\int_t^{t+\delta} F(x(\tau)) \, d\tau = -\delta F(x(t + \delta)) + O(\delta^2),$$

uniformly in t.

By the definition of semismoothness with $x = x(t+\delta)$ and $w = -\sigma$ we have, for δ sufficiently small,

$$F(x(t+\delta)) \;\; = F(x(t)) + V(x(t))\sigma + O(\|\sigma\|^{1+p})$$

(4.8)

$$= F(x(t)) + V(x(t))\sigma + O(\delta^{1+p}).$$

The estimate (4.8) is uniform in t because the set $\{x(t) \,|\, t \geq 0\}$ is compact.

Hence, multiplying (4.8) by delta and substituting into (4.7),

$$(D + \delta V(x(t)))\sigma = -\delta F(x(t)) + O(\delta^{2+p})$$

and the proof is complete. □

Lemma 4.2 will imply convergence of Ψtc in the same way as in the smooth case [7]. The objective is to show that for δ_0 sufficiently small, the Ψtc iteration remains in the tube $S(\epsilon_G)$ We state the result and sketch the proof.

THEOREM 4.3. *Let Assumptions 4.1 and 4.2 hold. Then if δ_0, $\{\eta_n\}$, and h are sufficiently small, and $\{\delta_n\}$ is bounded from below. then $x_n \to x^*$ and (4.1) holds. If F is semismooth of order p, then (4.2) holds.*

Proof. Let

$$t_n = \sum_{n=0}^{n-1} \delta_n.$$

The SER formula implies that

$$\delta_n \leq \delta_0/\|F(x_n)\|.$$

For now assume that F is semismooth of order p and that $\eta_n = 0$ for all n. Our assumptions imply that there is T such that $x(t) \in \mathcal{B}(x^*, \epsilon_T/2)$, the ball of local convergence from Theorem 4.1, for all $t \geq T$. Lemma 4.2 implies that for small δ, the Ψtc iteration is an accurate integrator for (2.8) in the sense that

(4.9) $$\|x_n - x(t_n)\| = O(\delta^p + nh),$$

where $\delta = \max_{0 \leq k \leq n} \delta_k$. Hence, we can select δ_0 and h such that $x_n \in S(\epsilon_G)$ until $t_n > T$.

If F is semismooth and $\{\eta_n\}$ is non-zero, then (4.9) becomes (see [18])

(4.10) $$\|x_n - x(t_n)\| = O\left(nh + \sum_{j=0}^{n} \eta_j \delta_j\right) + o(1), \text{ as } \delta \to 0 ,$$

and the convergence result still holds if, say $\eta_n = O(\delta_0)$. □

5. Numerical Example. We illustrate the results with a simple one-dimensional example taken from [1, 2, 4]. This example is sufficient to illustrate the convergence results in this paper, and allows us to refine the grids to a degree that was not possible in the two and three dimensional results that motivated this paper [10, 14, 15, 24, 30, 31].

We use direct methods to compute the Newton step in this section, so $\eta_n = 0$. In all but § 5.2, we compute $V \in \partial F(x)$ analytically, so $h = 0$ in those computations.

This example, taken from [4], is a Lipschitz reformulation of the boundary value problem [1,2]

$$-u_{zz} + \lambda \max(0, u)^p = 0, \; z \in (0, 1),$$

with boundary data

$$u(0) = u(1) = 0.$$

and $p \in (0, 1)$.

The reformulation adds a new variable

$$v = \begin{cases} u^p & \text{if } u \geq 0 \\ u & \text{if } u < 0 \end{cases}$$

to obtain a Lipschitz continuous elliptic-algebraic system, $F(x) = 0$, where $x = (u, v)^T$ and

(5.1) $$F(x) = \begin{pmatrix} f(u, v) \\ g(u, v) \end{pmatrix} = \begin{pmatrix} -u_{zz} + \lambda \max(0, v) \\ u - \omega(v) \end{pmatrix} = 0$$

where

$$\omega(v) = \begin{cases} v^{1/p} & \text{if } v \geq 0 \\ v & \text{if } v < 0 \end{cases}$$

We report results on a pseudo-transient continuation (Ψtc) approach. We use the method from [7] which is designed for differential algebraic equations (DAEs).

The reason we formulate the problem with DAE (rather than ODE) dynamics is that the pseudo-time variable should not be added to both equations in (5.1), but only the first. The reason for this is that the true time-dependent system is

$$u_t = u_{zz} - \lambda \max(0, u)^p.$$

and that the auxiliary variable v is used only to make the nonlinearity Lipschitz continuous. One might think that an ODE formulation would work equally well, but, in fact, the ODE formulation, which does not model the physics, failed to converge in our testing.

We discretize the problem with a central differences, using a difference increment of δ_z. The nonsmooth nonlinearity is a substitution operator, and its generalized Jacobian is a set of diagonal matrices.

We report on several computations with $p = .1$ and $\lambda = 200$. This choice leads to a large "dead core" [1, 2], a region in which the solution vanishes. We plot the solution in Figure 5.1.

FIG. 5.1. *Solution*
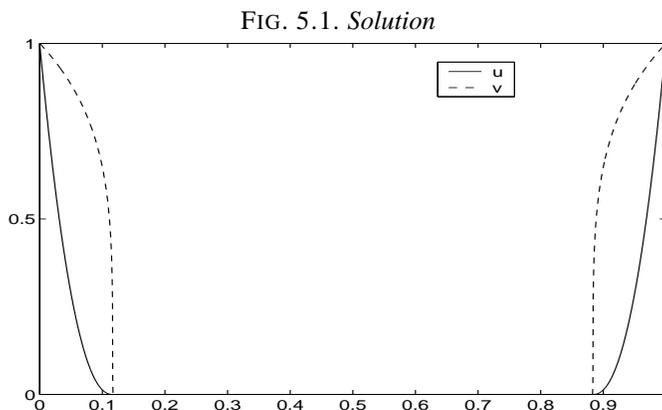
$\delta_0 = 1$ and $\delta_{max} = 10^6$ for all the computations. We terminate the nonlinear iteration when either

(5.2) $$\|F(x_n)\|/\|F(x_0)\| < 10^{-13} \text{ or } \|s_n\| < 10^{-10},$$

where $s_n = x_{n+1} - x_n$. In the tables we see the superlinear convergence clearly in the reduction in the norms of the steps, this is consistent with the estimate $s_n = -e_n + o(\|e_n\|)$ which follows from local superlinear convergence. The superlinear convergence is less visible in the residual norms, because the generalized Jacobians become more ill-conditioned as the mesh is refined. The residual norms begin to stagnate after a reduction of 10^{12}.

5.1. Exact Compuation of the Generalized Jacobian. For the results in this section we compute the generalized Jacobian analytically. If we let L_{δ_z} be the discretized Laplacian, we can write

(5.3) $$F(x) = \begin{pmatrix} f(u,v) \\ g(u,v) \end{pmatrix} = \begin{pmatrix} -L_{\delta_z} u \\ u - v - max(0, v^{1/p}) \end{pmatrix} + \begin{pmatrix} \lambda \\ 1 \end{pmatrix} max(0,v),$$

and use the known result for the scalar function $\max(0, v)$

$$\partial \max(0, v) = \begin{cases} 0, & \text{if } v < 0 \\ [0,1], & \text{if } v = 0 \\ 1, & \text{if } v > 0, \end{cases}$$

we obtain

(5.4) $$\partial F = \begin{pmatrix} -L_{\delta_z} & 0 \\ 1 & -1 - (1/p)\max(0, v^{(1-p)/p}) \end{pmatrix} + \begin{pmatrix} 0 & \lambda \\ 0 & 1 \end{pmatrix} \partial \max(0,v).$$

The calculations in this section used $V(x_n) \in \partial F(x_n)$. We may use any choice from the set-valued map $\partial(0, v)$ and we choose $W \in \partial F$ using

$$\begin{Bmatrix} 0, & \text{if } v \le 0 \\ 1, & \text{if } v > 0, \end{Bmatrix} \in \partial \max(0, v).$$

With this choice, the W_{vv} is nonsingular. Had we used 1 when $v = 0$, the W_{vv} would be singular at $v = 0$.

In Figure 5.2 we plot the norms of the steps and nonlinear residuals together with the growth of δ for a mesh of width $\delta_z = 1/2048$. δ grows smoothly in the early phase of the iteration and reaches its maximum rapidly. The superlinear convergence is clearly visible in the curve for the norms of the steps. The Jacobian of the nonlinear residual has a condition number of $O(1/h^2)$, and reflects the error less accurately.

5.2. Compuation of the Generalized Jacobian by Differences. For the results in this section we compute the generalized Jacobian with several choices of difference. The results were similar for all the meshes. In Figures 5.2 and 5.3 we plot residual and step norm histories for
- analytic generalized Jacobian (Exact),
- forward differences, increment 10^{-8} (F-8), and
- forward differences, increment 10^{-10} (F-10),

for a mesh of width $\delta_z = 1/2048$ and 20 iterations. In this way we can clearly see the point at which the iteration stagnates. As we predicted in § 3.3, the iteration is more accurate when the difference increment is $10^{-10} \approx \epsilon^{2/3}$ than it is with the standard choice of $10^{-8} \approx \epsilon^{1/2}$.

FIG. 5.2. *Analytic Generalized Jacobian*
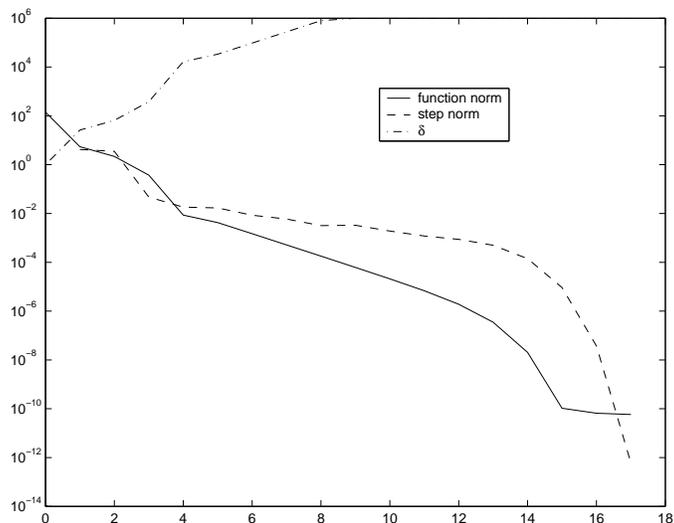
FIG. 5.3. *Norms of the Steps and Residuals.*
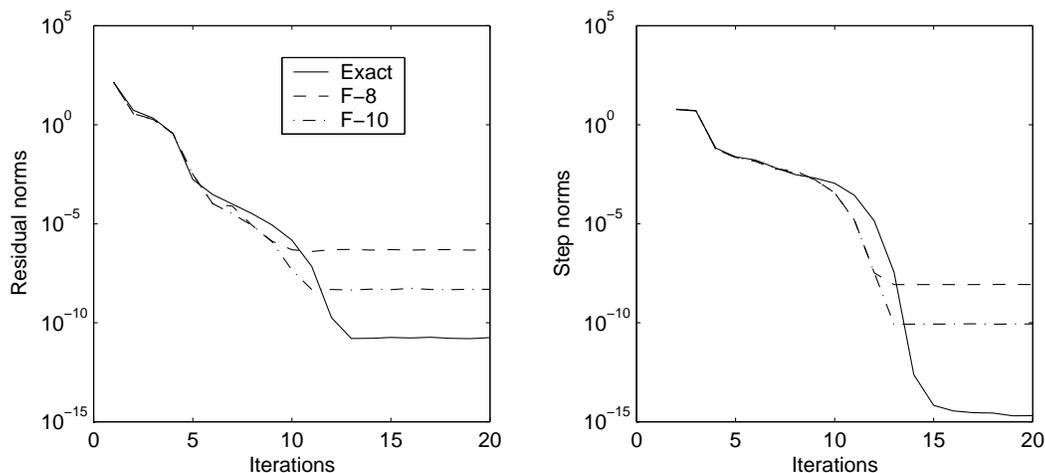
5.3. Mesh Dependence and Nested Iteration. We used the analytic ∂F (5.4) in the computations reported in this section.

In Figure 5.4 we plot the progress of the iteration for mesh sizes of $1/128$, $1/512$, and $1/2048$, terminating the iteration when $\|s\| < 10^{-13}$. In this way we can examine the dependence of the convergence on the mesh width. While the convergence in the early phase is identical for the three meshes, and superlinear in the terminal phase, the global convergence becomes slower as the mesh is refined.

Nested iteration or grid sequencing means to solve the problem to high precision on a coarse mesh, interpolate to a finer mesh in such a way that the interpolation error can be corrected with a few (eg one) iterations, and to continue this until one has a solution on a target, finest mesh. We set $\delta = 10^6$ for the finer meshes, under the assumption that we are in the locally convergent phase of the iteration.

FIG. 5.4. *Mesh-dependence of Convergence*
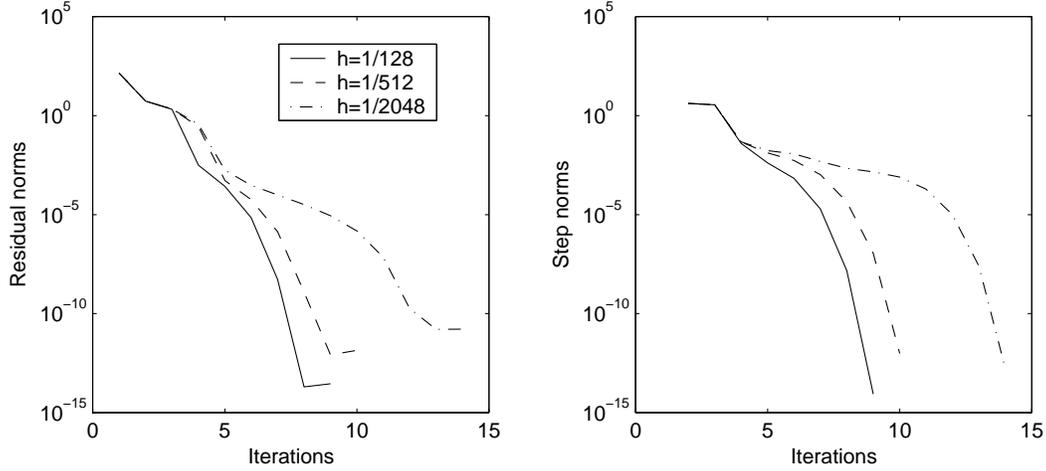
For this example, one would hope not only to eliminate the mesh-dependency in the iteration history that are visible in Figure 5.4, but also to approximate the solution up to truncation error at each level.

This was a successful strategy. However, the results must be interpreted in light of the continuity properties of the solution. u^* is Lipschitz continuously differentiable on $[0, 1]$, but if $p < 1/2$, $v^* = (u^*)^p$ is not. This means that linear interpolation will not approximate v^* to second order if $p < 1/2$. To address this we interpolate u from the coarse to fine mesh with linear interpolation, and then compute v as

$$v = max(0, u)^p.$$

In tables 5.1 and 5.2 we report the residual and step norms on a sequence of meshes $\{2^{-n}\}_{n=6}^{11}$ for $p = .1$ and $p = .5$. The initial steps at each mesh reflect both the error in the initial iterate and the truncation error in the interpolation.

The iterations for both values of p show that we have recovered mesh independence in the sense that the iteration requires a roughly constant number of steps to terminate at each level. The tables for $p = 1/2$ clearly show second order convergence. The interpolation error for $p = .1$ is visible in the sizes of the initial steps.

TABLE 5.1
Step norms: $p = .1$, nested iteration

$n \backslash \delta_z$	1/64	1/128	1/256	1/512	1/1024	1/2048
0	4.20e+00	2.02e-02	1.02e-02	5.72e-03	3.45e-03	3.61e-03
1	3.53e+00	1.13e-02	1.23e-03	1.13e-03	2.14e-03	6.16e-04
2	3.91e-02	8.95e-04	1.56e-04	2.15e-04	1.58e-04	1.37e-05
3	4.11e-03	6.44e-05	2.19e-06	6.18e-06	7.24e-05	7.14e-08
4	6.89e-04	3.26e-07				2.23e-12
5	1.94e-05					
6	1.47e-08					

TABLE 5.2
Step norms: $p = .5$, nested iteration

$n\backslash\delta_z$	1/64	1/128	1/256	1/512	1/1024	1/2048
0	1.32e+00	1.52e-03	3.87e-04	9.74e-05	2.44e-05	6.13e-06
1	5.29e-01	4.37e-05	7.89e-06	1.39e-06	2.47e-07	4.25e-08
2	5.20e-03	1.10e-06	9.73e-08	4.75e-08		4.21e-09
3	6.59e-05		3.83e-09			1.76e-11
4	2.73e-05					
5	9.74e-08					

REFERENCES

[1] A. K. AZIZ, A. B. STEPHENS, AND M. SURI, *Numerical methods for reaction-diffusion problems with non-differentiable kinetics*, Numer. Math., 53 (1988), pp. 1–11.

[2] J. W. BARRETT AND R. M. SHANAHAN, *Finite element approximation of a model reaction - diffusion problem with a non-Lipschitzian nonlinearity*, Numer. Math., 59 (1991), pp. 217–242.

[3] K. E. BRENAN, S. L. CAMPBELL, AND L. R. PETZOLD, *The Numerical Solution of Initial Value Problems in Differential-Algebraic Equations*, no. 14 in Classics in Applied Mathematics, SIAM, Philadelphia, 1996.

[4] X. CHEN, *A superlinearly and globally convergent method for reaction and diffusion problems with a non-lipschitzian operator*, Computing [suppl], (2001), pp. 79–90.

[5] X. CHEN, Z. NASHED, AND L. QI, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM Journal on Numerical Analysis, 38 (2001), pp. 1200–1216.

[6] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Wiley–Interscience, New York, 1983.

[7] T. COFFEY, C. T. KELLEY, AND D. E. KEYES, *Pseudo-transient continuation and differential-algebraic equations*, Tech. Rep. CRSC-TR02-18, North Carolina State University, Center for Research in Scientific Computation, July 2002. to appear in SIAM J. Sci. Comp.

[8] T. S. COFFEY, R. J. MCMULLAN, C. T. KELLEY, AND D. S. MCRAE, *Globally convergent algorithms for nonsmooth nonlinear equations in computational fluid dynamics*, J. Comp. Appl. Math., 152 (2003), pp. 69–81.

[9] F. FACCHINEI, A. FISCHER, AND C. KANZOW, *Inexact newton methods for semismooth equations with applications to variational inequality problems*, in Nonlinear Optimization and Applications, G. D. Pillo and F. Giannessi, eds., New York, 1996, Plenum Press, pp. 125–139.

[10] M. W. FARTHING, C. E. KEES, T. COFFEY, C. T. KELLEY, AND C. T. MILLER, *Effi cient steady-state solution techniques for variably saturated groundwater flow*, Tech. Rep. CRSC-TR02-31, North Carolina State University, Center for Research in Scientific Computation, November 2002. To appear in Advances in Water Resources.

[11] M. HEINKENSCHLOSS, C. T. KELLEY, AND H. T. TRAN, *Fast algorithms for nonsmooth compact fi xed point problems*, SIAM J. Numer. Anal., 29 (1992), pp. 1769–1792.

[12] M. HINTERMÜLLER, K. ITO, AND K. KUNISCH, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.

[13] K. R. KAVANAGH, *Nonsmooth Nonlinearities in Applications from Hydrology*, PhD thesis, North Carolina State University, Raleigh, North Carolina, 2003.

[14] K. R. KAVANAGH, C. T. KELLEY, R. C. BERGER, J. P. HALLBERG, AND S. E. HOWINGTON, *Nonsmooth nonlinearities and temporal integration of Richards' equation*, in Computational Methods in Water Resources XIV, Vol. 2, S. M. Hassanizadeh, R. J. Schotting, W. G. Gray, and G. F. Pinder, eds., Amsterdam, 2002, Elsevier, pp. 947–954.

[15] C. E. KEES AND C. T. MILLER, *Higher order time integration methods for two-phase flow*, Advances in Water Resources, 25 (2002), pp. 159–177.

[16] C. T. KELLEY, *Identification of the support of nonsmoothness*, in Large Scale Optimization: State of the Art, W. W. Hager, D. W. Hearn, and P. Pardalos, eds., Boston, 1994, Kluwer Academic Publishers B.V., pp. 192–205.

[17] ———, *Iterative Methods for Linear and Nonlinear Equations*, no. 16 in Frontiers in Applied Mathematics, SIAM, Philadelphia, 1995.

[18] C. T. KELLEY AND D. E. KEYES, *Convergence analysis of pseudo-transient continuation*, SIAM J. Numer. Anal., 35 (1998), pp. 508–523.

[19] C. T. KELLEY AND E. W. SACHS, *Multilevel algorithms for constrained compact fixed point problems*, SIAM J. Sci. Comp., 15 (1994), pp. 645–667.

[20] ———, *A trust region method for parabolic boundary control problems*, SIAM J. Optim., 9 (1999), pp. 1064–1081.

[21] J. MARTINEZ AND L. QI, *Inexact newton methods for solving nonsmooth equations*, 1995.

[22] J. M. MARTINEZ, *Quasi-Newton methods for solving underdetermined nonlinear simultaneous equations*, J. Comp. Appl. Math., 34 (1991), pp. 171–190.

[23] R. MIFFLIN, *Semismooth and semiconvex functions in constrained optimization*, Siam Journal on Control and Optimization, 15 (1977), pp. 959–972.

[24] C. T. MILLER, G. A. WILLIAMS, C. T. KELLEY, AND M. D. TOCCI, *Robust solution of Richards' equation for non-uniform porous media*, Water Resources Research, 34 (1998), pp. 2599–2610.

[25] Y. MUALEM, *A new model for predicting the hydraulic conductivity of unsaturated porous media*, Water Resources Research, 12 (1976), pp. 513–522.

[26] W. MULDER AND B. V. LEER, *Experiments with implicit upwind methods for the Euler equations*, J. Comp. Phys., 59 (1985), pp. 232–246.

[27] J. S. PANG AND L. QI, *Nonsmooth equations: motivation and algorithms*, SIAM J. Optim., 3 (1993), pp. 443–645.

[28] L. QI AND J. SUN, *A nonsmooth version of newton's method*, Mathematical Programming, 68 (1993), pp. 353–367.

[29] L. Q. QI AND X. J. CHEN, *A globally convergent successive approximation method for severely nonsmooth equations*, SIAM Journal on Control and Optimization, 33 (1995), pp. 402–418.

[30] K. STAHELI, J. H. SCHMIDT, AND S. SWIFT, *Guidelines for Solving Groundwater Problems with ADH*, January 1998.

[31] M. D. TOCCI, C. T. KELLEY, C. T. MILLER, AND C. E. KEES, *Inexact Newton methods and the method of lines for solving Richards' equation in two space dimensions*, Computational Geosciences, 2 (1998), pp. 291–310.

[32] M. ULBRICH, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim., 13 (2003), pp. 805–842.

[33] M. T. VAN GENUCHTEN, *Predicting the hydraulic conductivity of unsaturated soils*, Soil Science Society of America Journal, 44 (1980), pp. 892–898.