Pseudo-transient Continuation for Nonsmooth Nonlinear Equations

C. T. Kelley

Department of Mathematics
Center for Research in Scientific Computation
North Carolina State University
Raleigh, North Carolina, USA
Joint with K. R. Fowler (Kavanagh).

Workshop on Solution Methods for Large Scale Nonlinear Problems

Livermore, CA

August 7, 2003

Supported by NSF, ARO, DOEd.

NCSU Students/Postdocs

- PhD Students
  - Mike Tocci (1998, Mathworks)
  - Lea Jenkins (2000, Clemson)
  - Todd Coffey (2002, Sandia)
  - Katie Fowler (2003, Clarkson)
  - Jill Reese (200?, HA381)
- Postdoc: Chris Kees (HA 337)

Outline

- What is Ψtc ?
- Motivation
  - Unsaturated flow; nonsmooth constitutive laws (Tocci, Jenkins, Kavanagh, Kees, Howington, Farthing, Miller, K.)
  - CFD (Coffey, Keyes, McRae, McMullan, K.)
- Local convergence
  - What do you use for a Taylor expansion?
  - Quality of derivative approximation.
- Numerical example

What is Ψtc ?

- Find steady state solutions of

$$x' = F(x)$$

- Mimic temporal integration.
  Grow the time step in the terminal phase.
- Addresses failure mode of Newton-Armijo.
- Avoids non-physical results.

Newton-Armijo: the obvious thing

$$x_{n+1} = x_n + s \text{ where } s = \lambda d$$

and

$$\|F'(x_n)d + F(x_n)\| \le \eta_n \|F(x_n)\|.$$

You pick λ such that

$$\|F(x_n + \lambda d)\| \le (1 - \alpha\lambda)\|F(x_n)\|$$

where usually $\alpha = 10^{-4}$.

Newton-Armijo bottom line

If F is smooth and the computation of d and λ succeed, then either

- **BAD:** the iteration is unbounded, *i. e.* $\limsup \|x_n\| = \infty$,
- **BAD:** the derivatives tend to singularity, *i. e.* $\limsup \|F'(x_n)^{-1}\| = \infty$, or
- **GOOD:** the iteration converges to a solution x^* in the terminal phase, $\lambda = 1$, and

$$\|x_{n+1} - x^*\| = O(\|x_n - x^*\|\eta_n + \|x_n - x^*\|^2).$$

So what's the problem?

- Stagnation at singularity of F' really happens.
  - steady flow $\longrightarrow$ shocks in CFD
- Non-physical results
  - fires go out
  - negative concentrations
- Nonsmooth nonlinearities
  - are not uncommon: flux limiters, constitutive laws
  - globalization is harder
  - finite diff directional derivatives may be wrong

Ψtc is one way to fix it.

DAE Dynamics: semi-explicit

$x = (u, v)^T$ and

$$D \begin{pmatrix} u \\ v \end{pmatrix}' = -\begin{pmatrix} f(u,v) \\ g(u,v) \end{pmatrix} = -F(x), \quad x(0) = x_0,$$

where

$$D = \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix}, \text{ nonsingular .}$$

Differential variables u. Algebraic variables v.

Ψtc for smooth problems

$$x_{n+1} = x_n + s_n$$

where

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

and (SER)

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

Theory for smooth problems

Joint with Todd Coffey, David Keyes

If

- $Dx' = -F(x)$; $x(0) = x_0$; $x^* = \lim_{t\to\infty} x(t)$.
- DAE has uniform index one (g_{vv} nonsingular near $x(t)$).
- x^* is stable steady state.
- δ_0 is sufficiently small.
- Update δ_n with SER.

Then $x_n \to x^*$ and local convergence is what you'd expect from inexact Newton.

That’s nice, but . . .

Not all nonlinearities are smooth.

- Slope limiters in CFD
- Non-differentiable constitutive laws.
  e. g. Groundwater flow in the unsaturated zone.
- Nonsmooth reaction models (see example).

Nonsmooth Calculus

Assume that F is Lipschitz continuous on R^N. Then F is differentiable almost everywhere.
The generalized Jacobian (Clarke) at x is

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

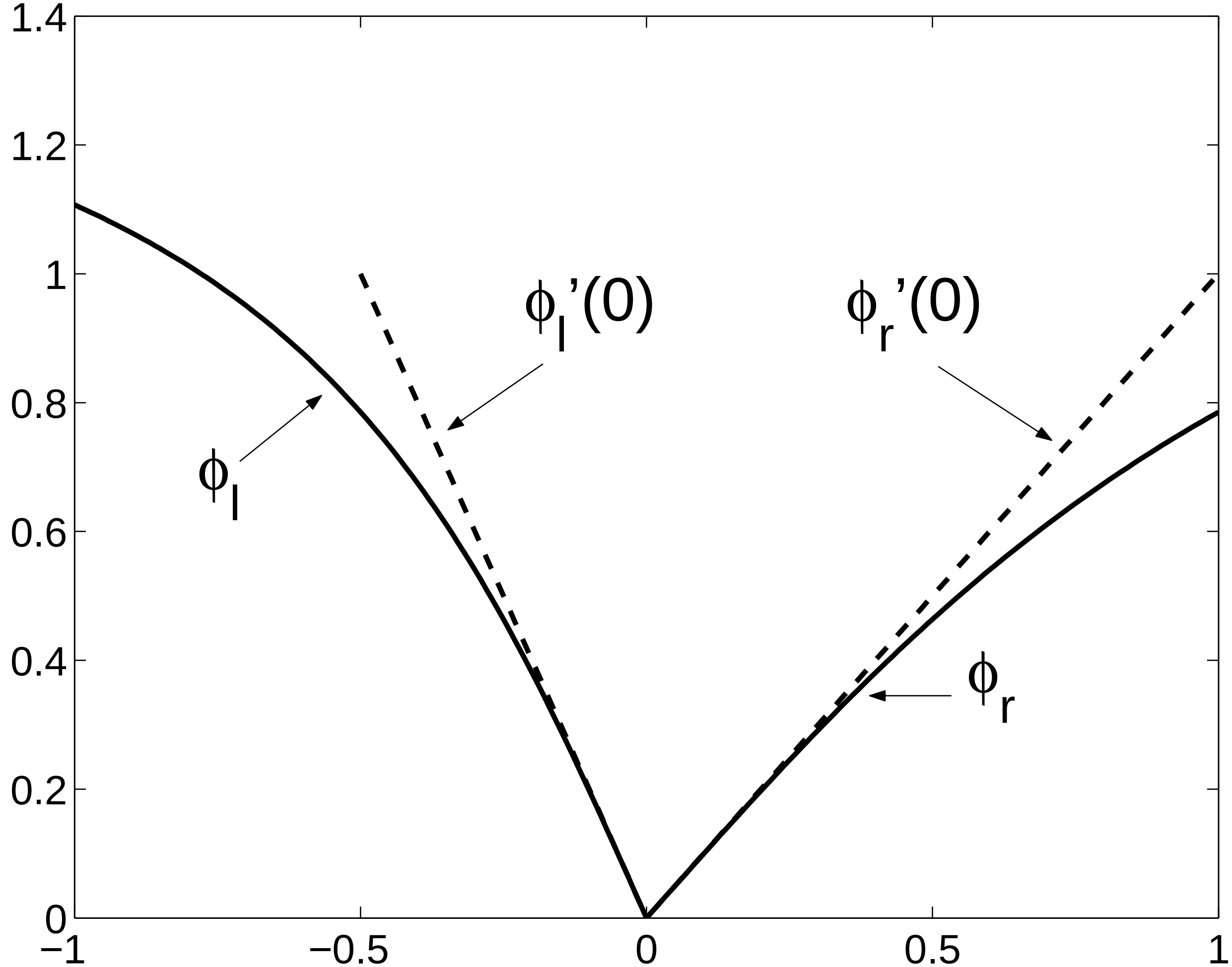
You'd like to replace Newton's method with

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

where $V_n \in \partial F(x_n)$.
Does that work? How do you compute V_n?

Piecewise smooth function: $\phi = \phi_l + \phi_r$

$\partial\phi(0) = [\phi_l(0), \phi_r(0)]$, a **set**.



Difference approximations

Scalar functions

$$\partial_h \phi(x) = \frac{\phi(x+h) - \phi(x)}{h}$$

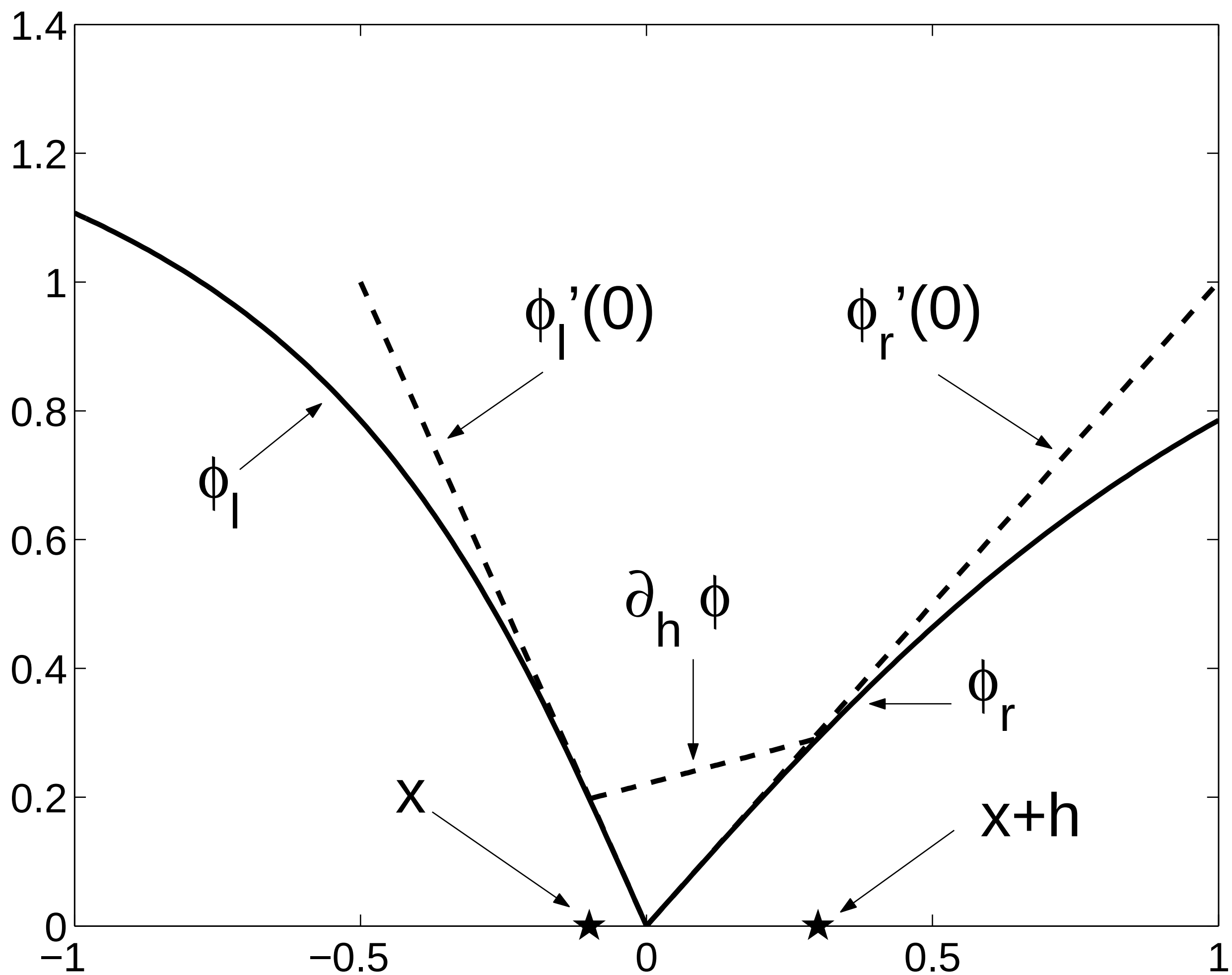
For Lipschitz functions:

$$\partial_h \phi(x) \in \partial \phi(\bar{x}) + O(h)$$

where $|x - \bar{x}| \leq h$.
Same story for scalar constitutive laws in PDEs.
If you differentiate in coordinate directions!

Difference approximation accuracy

$\phi_l'(0) + O(h) \leq \partial_h \phi(x) \leq \phi_r'(0) + O(h)$, so $\partial_h \phi(x) \in \partial \phi(0) + O(h)$



Semismoothness

A Lipschitz function F is semismooth (Mifflin, Pang, Qi) if

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

and semismooth of order p at z if

$$F(z+w) - F(z) - Vw = O(\|w\|^{1+p})$$

for all $w \in R^N$ and $V \in \partial F(x+w)$ as $w \to 0$.
What you need for local convergence of Newton's method.
Piecewise smooth functions are semismooth of order 1.

Why semismoothness?

If

- F semismooth of order p,
- $F(x^*) = 0$, and
- everything in $\partial F(x^*)$ uniformly nonsingular,
- x_c near x^*,

then if

$$x_+ = x_c - V^{-1}F(x_c), \text{ where } V \in \partial F(x_c),$$

you get fast local convergence

$$\|x_+ - x^*\| = O(\|x_c - x^*\|^{1+p}).$$

Convergence Proof, $e = x - x^*$

Semismoothness ($z = x^*, w = e_c, z + w = x_c$) implies

$$F(x_c) - Ve_c = O(\|e_c\|^{1+p})$$

Subtract x^* from both sides of

$$x_+ = x_c - V^{-1}F(x_c),$$

to get

$$e_+ = e_c - V^{-1}F(x_c) = e_c - e_c + O(\|e_c\|^{1+p}) = O(\|e_c\|^{1+p}).$$

Things get more complicated if x_c is far from x^*.
Armijo may fail.

Formulation of Ψtc

$$x_{n+1} = x_n + s_n$$

where

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

and

$$V(x) \in \partial F(\overline{x}) + O(h), \|x - \overline{x}\| \le h.$$

Local Convergence: $e_n = x_n - x^* \to 0$

Once close, grow the time step and get fast convergence.
If $\delta_{max} = \infty$

- F semismooth of order 1.
- $F(x^*) = 0$. Everything in $\partial F(x^*)$ nonsingular.
- $\|(D + \delta V(x))^{-1} D\| \leq 1/(1 + \beta\delta)$, for all $\delta > 0$.
- x_0 sufficiently near x^*.
- h sufficiently small.

then $\delta_n \to \infty$ and

$$\|e_{n+1}\| = O(\|e_n\|^2 + (\eta_n + \delta_n^{-1})\|e_n\| + h).$$

Early stagnation comes from the difference.

Stable steady state

$$Dx' = F(x), x(0) = x_0$$

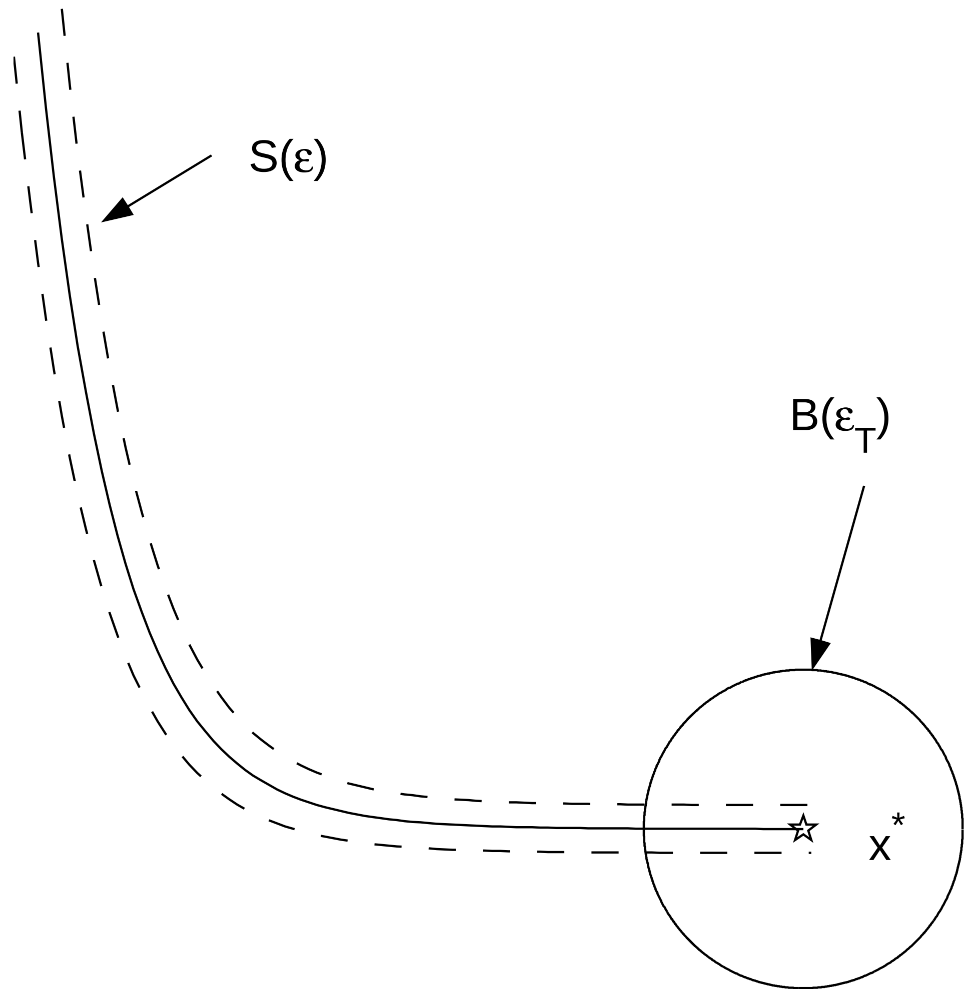
has consistent initial data and

$$x(t) \rightarrow x^* \text{ as } t \rightarrow \infty$$

Iteration confined to a neighborhood of the trajectory $x(t) : t > 0$.

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

Iterations hug trajectory



Assumptions in $S(\varepsilon)$

- $(\delta^{-1}D + V(x))$ is uniformly bounded and well-conditioned.
- $V_{vv}(x)$ is uniformly bounded and well conditioned.

We partition V consistently with F.

$$F(x) = \begin{pmatrix} f(u,v) \\ g(u,v) \end{pmatrix} \text{ and } V(x) = \begin{pmatrix} V_{uu} & V_{uv} \\ V_{vu} & V_{vv} \end{pmatrix}.$$

Bottom line: $x_n \to x^*$, $\delta_n \to \infty$, everything works.

Example

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

$$z \in (0,1), u(0) = u(1) = 0,$$

where $p \in (0,1)$.
Reformulate as a DAE to make the nonlinearity Lipschitz.
Let

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

Reformulation

Set $x = (u,v)^T$ and solve

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

The nonlinearity is

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

DAE Dynamics

$$D\begin{pmatrix} u \\ v \end{pmatrix}' = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}' = \begin{pmatrix} u' \\ 0 \end{pmatrix}$$

$$= -\begin{pmatrix} f(u,v) \\ g(u,v) \end{pmatrix} = -F(x), \quad x(0) = x_0,$$

Why not ODE dynamics?

Original time-dependent problem is

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

Applying Ψtc to

$$v_t = u - \omega(v)$$

rather than using $u - \omega(v) = 0$ as an algebraic constraint

- adds non-physical time dependence,
- changes the problem, and
- doesn't work.

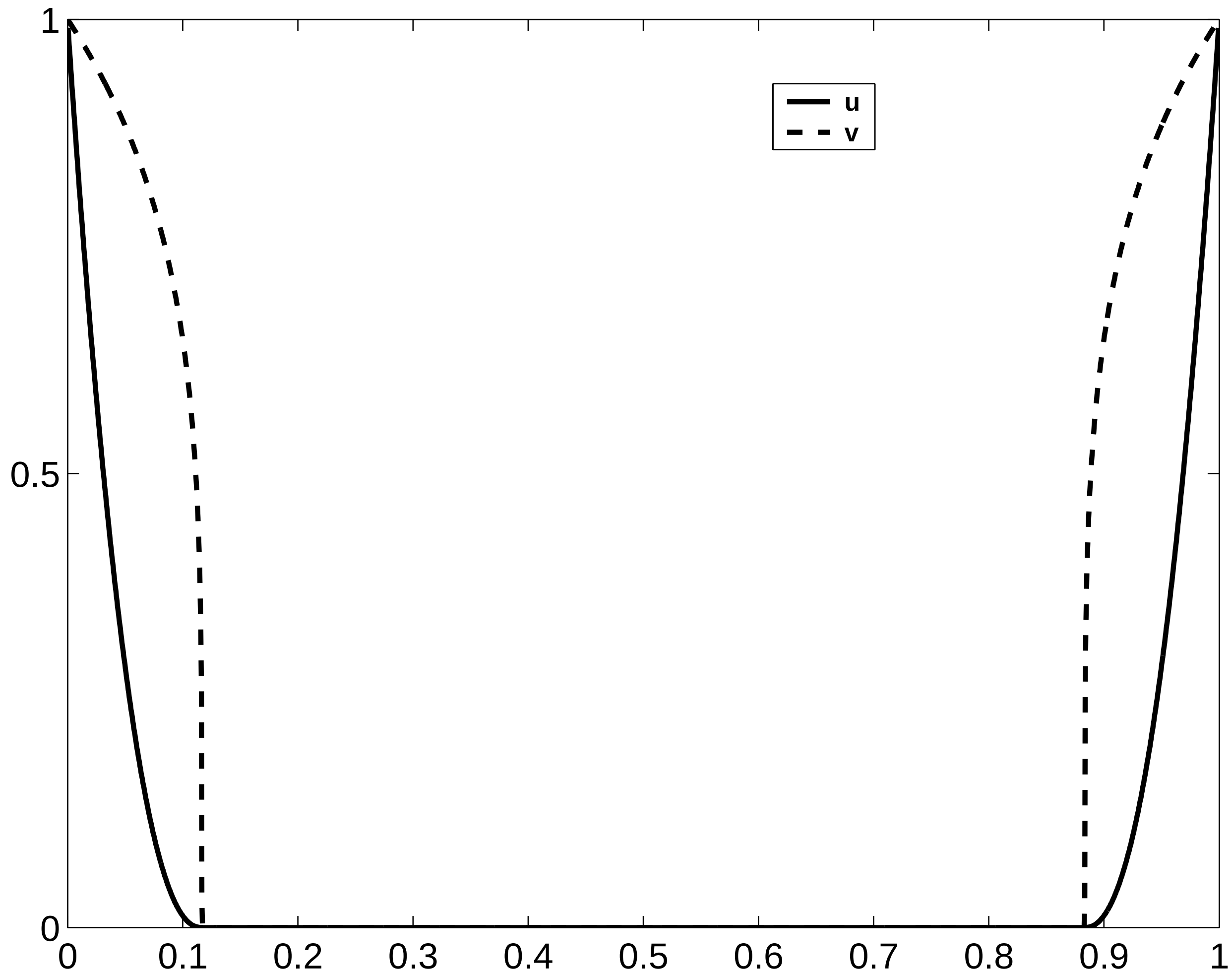
Parameters

- $p = .1$ and $\lambda = 200$. Leads to "dead core".
- $\delta_0 = 1.0$, $\delta_{max} = 10^6$.
- Spatial mesh size $\delta_z = 1/2048$; discrete Laplacian L_{δ_z}
- Terminate nonlinear iteration when either

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

Step is an accurate estimate of error (semismoothness).

Solution



Analytic ∂F

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

Since

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

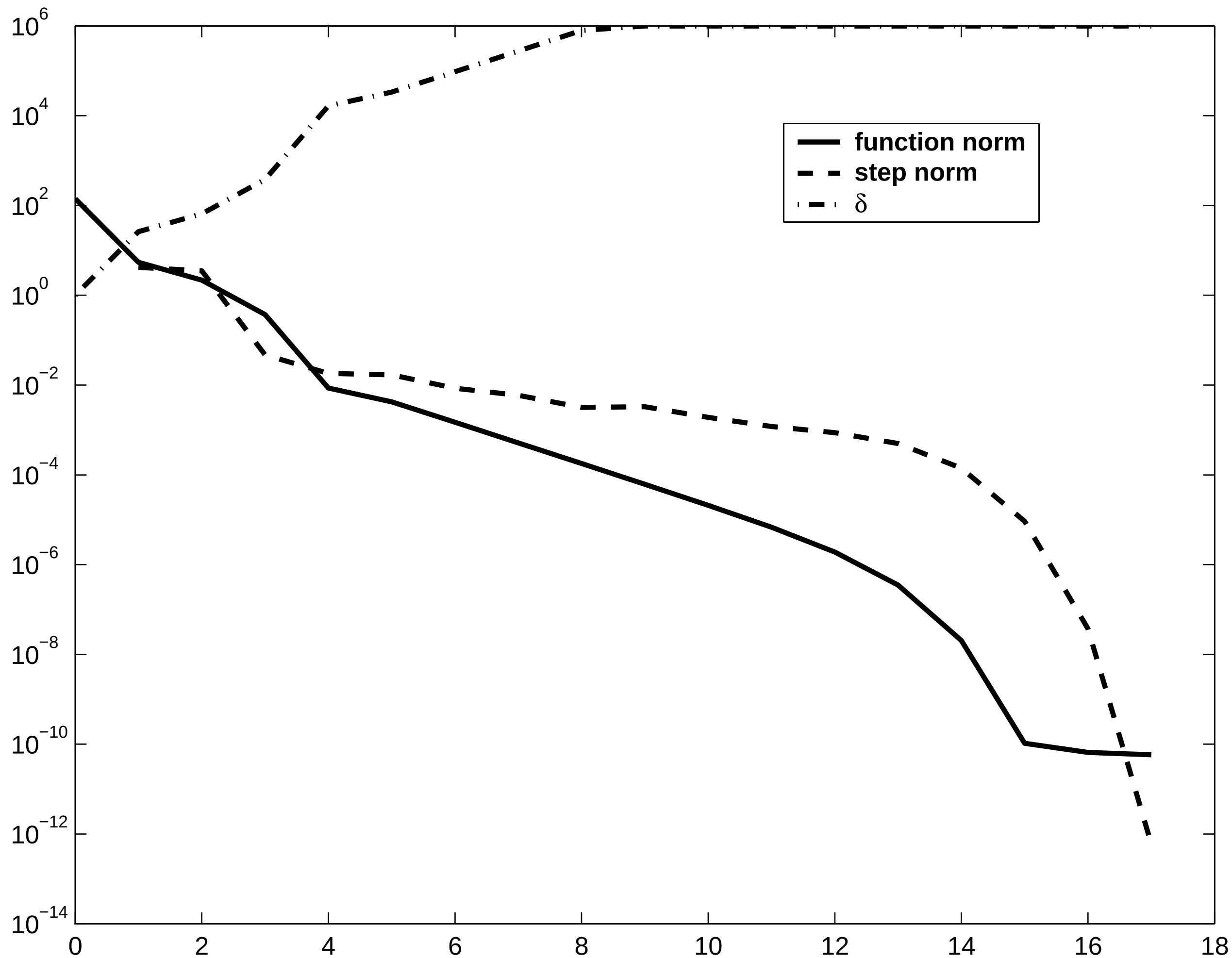
we get ...

∂F

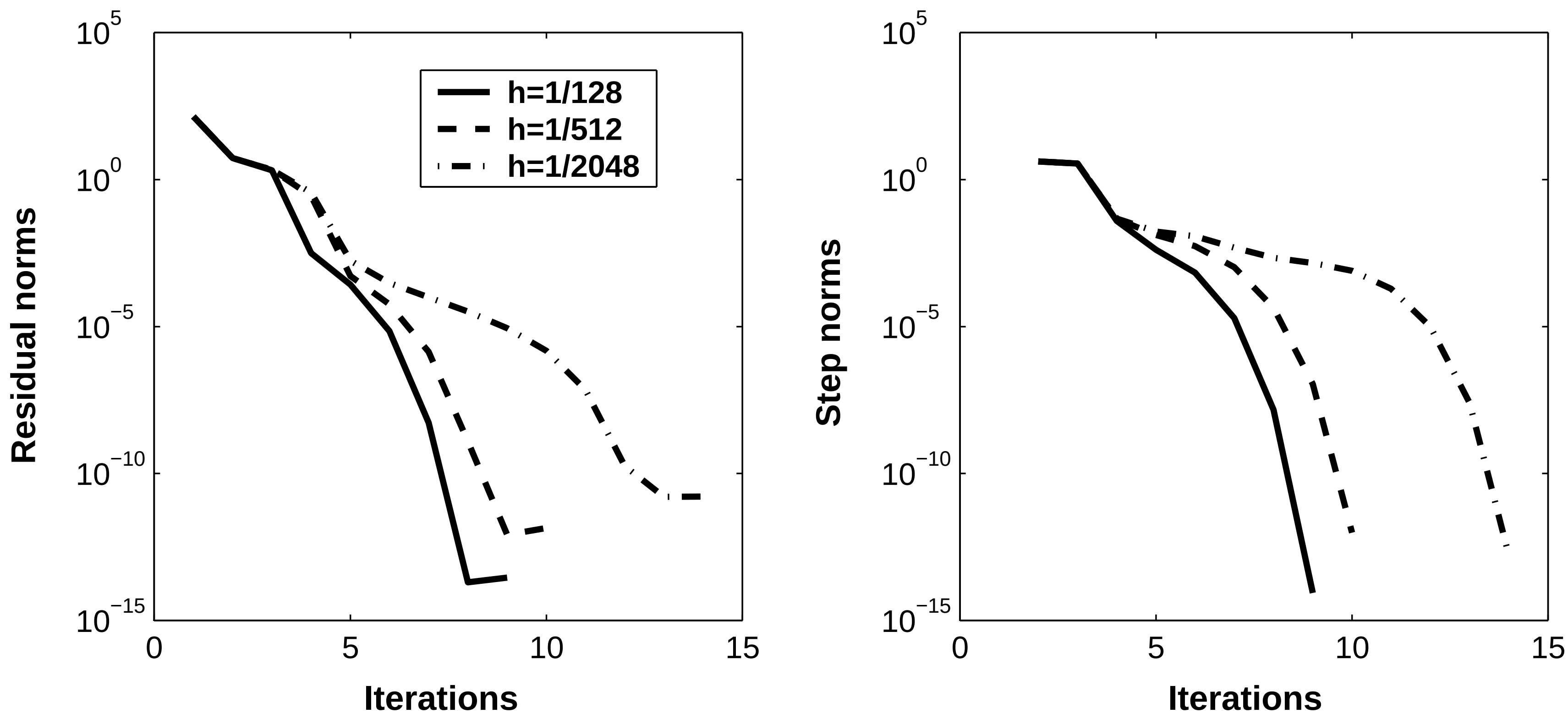
$$\partial F = \begin{pmatrix} -L_{\delta_z} & 0 \\ 1 & -1-(1/p)\max(0, v^{(1-p)/p}) \end{pmatrix}$$

$$+\begin{pmatrix} 0 & \lambda \\ 0 & 1 \end{pmatrix} \partial \max(0, v).$$

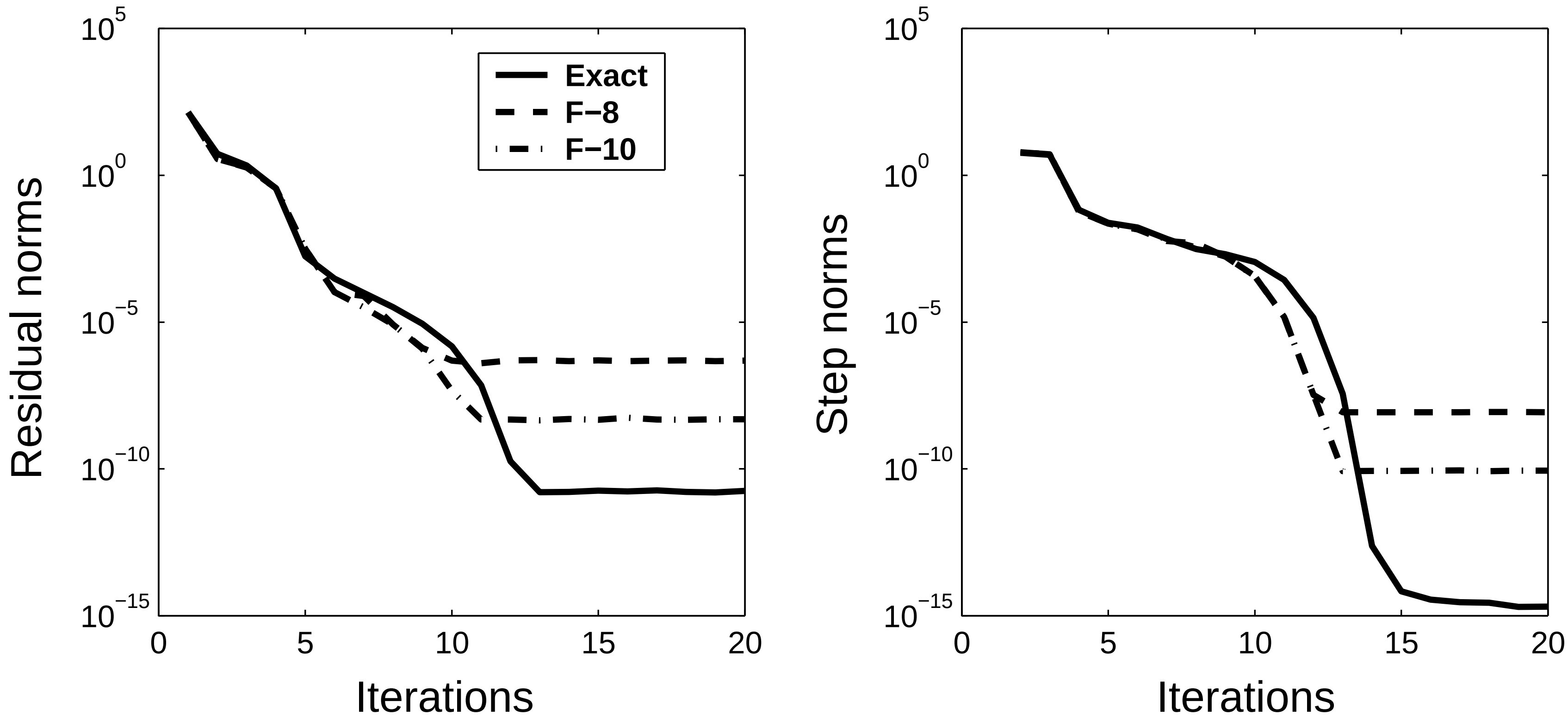
Convergence



Mesh Dependence



Forward Difference ∂F



Optimal difference increment

Let ε_F be floating point roundoff. Include this and

$$V(x) \in \partial F(\overline{x}) + O(h + \varepsilon_F / h)$$

So, if $\|e_n\| = \sqrt{h}$, then

$$\begin{aligned} e_{n+1} &= O((h + \varepsilon_F / h)\|e_n\| + \|e_n\|^2 + h) \\ &= O\left(\frac{\varepsilon_F}{h^{1/2}} + h\right) \end{aligned}$$

which is minimized if $h = O(\varepsilon_F^{2/3}) \approx 10^{-10}$ in IEEE.

Conclusions

- Ψtc can help if Newton-Armijo fails
- Generalized derivatives can be used in Ψtc
- Difference approximations work well with care
  Scalar functions and substitution operators
  Differentiate in coordinate directions
- Finite-difference Newton-Krylov needs more structure (in the works)
- Stagnation tied to difference increment
- Many applications