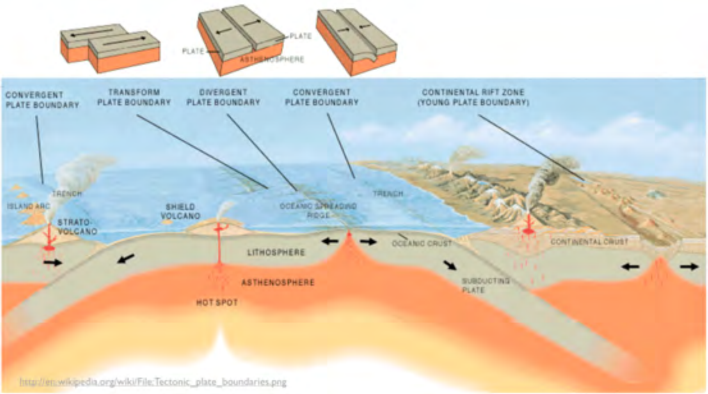
On adaptive methods in heterogeneous media

Jed Brown jed@jedbrown.org (ANL and CU Boulder)
Collaborators: Mark Adams (LBL), Matt Knepley (UChicago), Dave May (ETH), Laetitia Le Pourhiet (UPMC)

National University of Singapore, 2015-02-13

This talk:
http://59A2.org/files/20150213-AdaptHeterogeneous.pdf

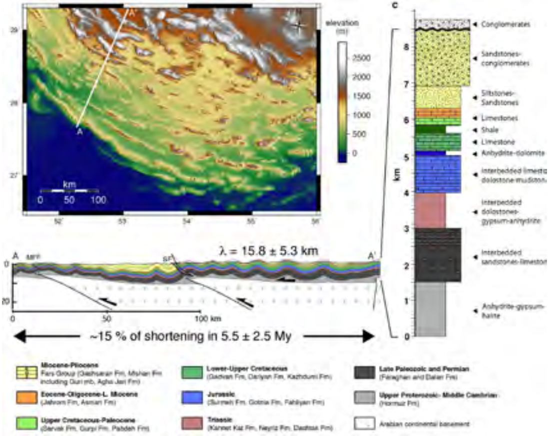
Regional scale geodynamic processes



- Buoyancy and topography drive flow
- Large variation in length scales
- Small scales influence large scales
- Complex rheology (research)
- Material failure (faults)
- Post-failure deformation
- Thermomechanically coupled

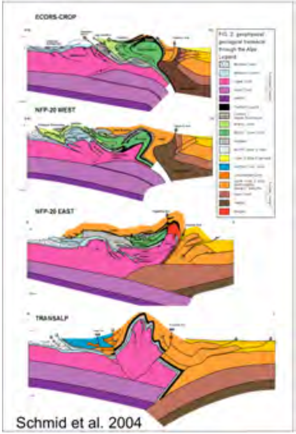
Geology is complicated

Zagros Mountains



[Yamato et al (2011)]

European Alps



Schmid et al. 2004

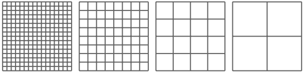
- Ductile folding
- Discontinuous material properties

- Inherently 3D
- Faulting

Continental rifting

Rifting Video

Multigrid Preliminaries



Multigrid is an $O(n)$ method for solving algebraic problems by defining a hierarchy of scale. A multigrid method is constructed from:

1. a sequence of discretizations
   - coarser approximations of problem, same or different equations
   - constructed algebraically or geometrically
2. intergrid transfer operators
   - residual restriction I_h^H (fine to coarse)
   - state restriction $\hat{I}_h^H$ (fine to coarse)
   - partial state interpolation I_H^h (coarse to fine, 'prolongation')
   - state reconstruction $\mathbb{I}_H^h$ (coarse to fine)
3. Smoothers (S)
   - correct the high frequency error components
   - Richardson, Jacobi, Gauss-Seidel, etc.
   - Gauss-Seidel-Newton or optimization methods
   - Compatible Monte Carlo, . . .

τ formulation of Full Approximation Scheme (FAS)

- classical formulation: "coarse grid *accelerates* fine grid $\searrow\nearrow$
- τ formulation: "fine grid feeds back into coarse grid" $\nearrow\searrow$
- To solve $Nu = f$, recursively apply

$$\begin{aligned}
\text{pre-smooth} \quad & \tilde{u}^h \leftarrow S^h_{\text{pre}}(u^h_0, f^h) \\
\text{solve coarse problem for } u^H \quad & N^H u^H = \underbrace{I^H_h f^h}_{f^H} + \underbrace{N^H \hat{I}^H_h \tilde{u}^h - I^H_h N^h \tilde{u}^h}_{\tau^H_h} \\
\text{correction and post-smooth} \quad & u^h \leftarrow S^h_{\text{post}}\left(\tilde{u}^h + I^h_H(u^H - \hat{I}^H_h \tilde{u}^h), f^h\right)
\end{aligned}$$

I^H_h	residual restriction	$\hat{I}^H_h$	solution restriction
I^h_H	solution interpolation	$f^H = I^H_h f^h$	restricted forcing
$\{S^h_{\text{pre}}, S^h_{\text{post}}\}$	smoothing operations on the fine grid		

- At convergence, $u^{H*} = \hat{I}^H_h u^{h*}$ solves the τ-corrected coarse grid equation $N^H u^H = f^H + \tau^H_h$, thus τ^H_h is the "fine grid feedback" that makes the coarse grid equation accurate.
- τ^H_h is *local* and need only be recomputed where it becomes stale.
- Interpretation by Achi Brandt in 1977, many tricks followed

Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

- 2-dimensional model problem for power-law fluid cross-section

$$-\nabla\cdot\left(|\nabla u|^{\mathfrak{p}-2}\nabla u\right) - f = 0, \qquad 1 \le \mathfrak{p} \le \infty$$

  Singular or degenerate when $\nabla u = 0$

- Regularized variant

$$-\nabla\cdot(\eta\nabla u) - f = 0$$

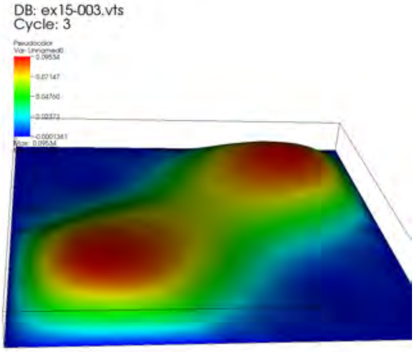
$$\eta(\gamma) = (\varepsilon^2 + \gamma)^{\frac{\mathfrak{p}-2}{2}} \qquad \gamma(u) = \frac{1}{2}|\nabla u|^2$$

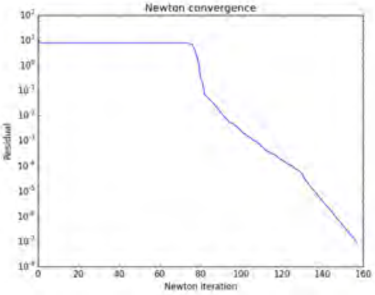
- Friction boundary condition on one side of domain

$$\nabla u \cdot n + A(x)\,|u|^{q-1}\,u = 0$$

Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

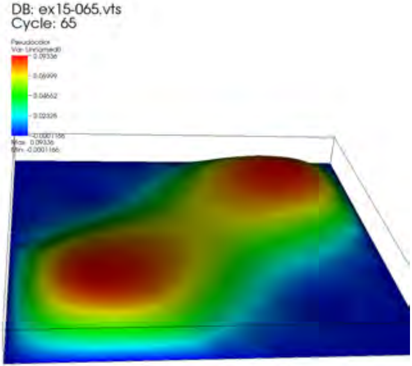
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

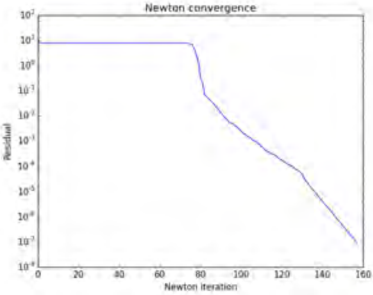




Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

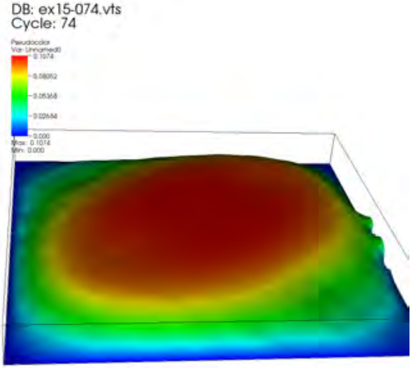
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

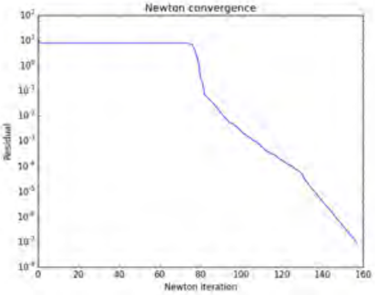




Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

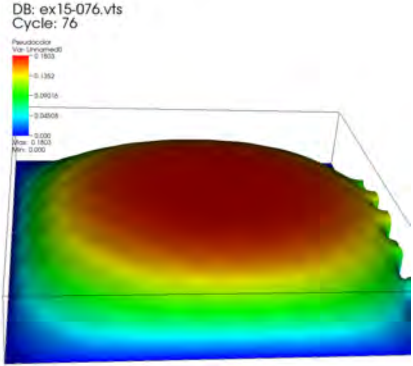
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

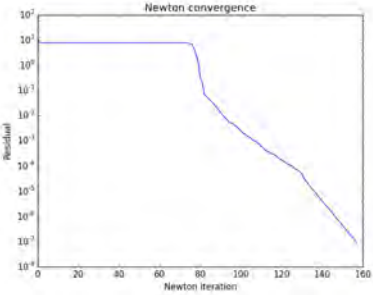




Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

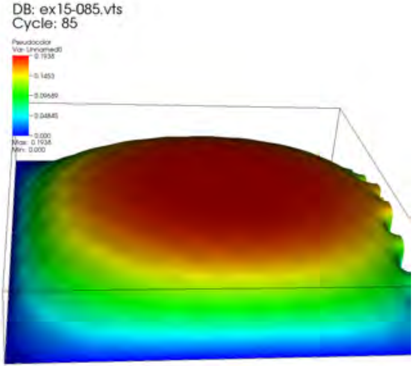
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

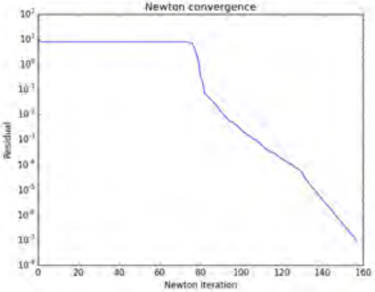




Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

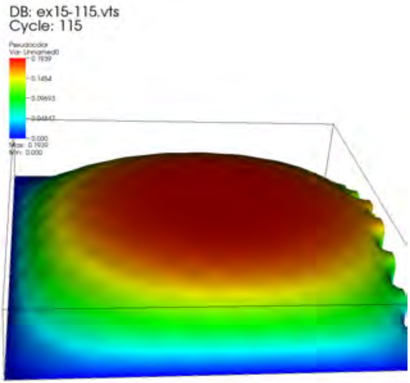
- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners

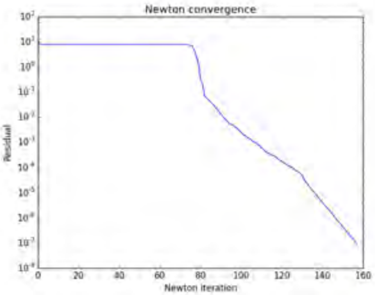




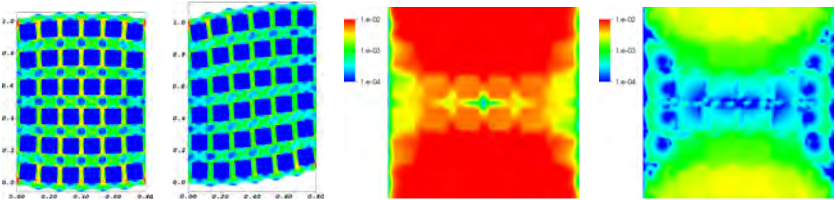
Model problem: $\mathfrak{p}$-Laplacian with slip boundary conditions

- $\mathfrak{p} = 1.3$ and $q = 0.2$, checkerboard coefficients $\{10^{-2}, 1\}$
- Friction coefficient $A = 0$ in center, 1 at corners





τ corrections

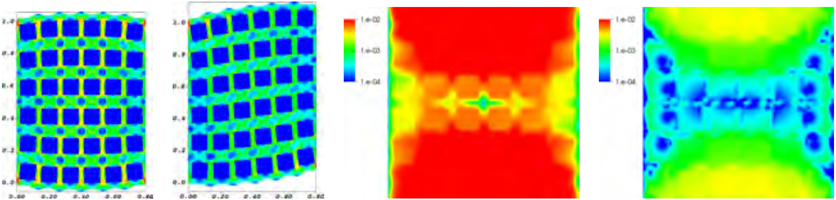


- Plane strain elasticity, $E = 1000, \nu = 0.4$ inclusions in $E = 1, \nu = 0.2$ material, coarsen by 3^2.
- Solve initial problem everywhere and compute $\tau_h^H = A^H \hat{I}_h^H u^h - I_h^H A^h u^h$
- Change boundary conditions and solve FAS coarse problem

$$N^H \acute{u}^H = \underbrace{I_h^H \acute{f}^h}_{\acute{f}^H} + \underbrace{N^H \hat{I}_h^H \tilde{u}^h - I_h^H N^h \tilde{u}^h}_{\tau_h^H}$$

- Prolong, post-smooth, compute error $e^h = \acute{u}^h - (N^h)^{-1} \acute{f}^h$

τ corrections



- Plane strain elasticity, $E = 1000, \nu = 0.4$ inclusions in $E = 1, \nu = 0.2$ material, coarsen by 3^2.
- Solve initial problem everywhere and compute $\tau_h^H = A^H \hat{I}_h^H u^h - I_h^H A^h u^h$
- Change boundary conditions and solve FAS coarse problem

$$N^H \acute{u}^H = \underbrace{I_h^H \acute{f}^h}_{\acute{f}^H} + \underbrace{N^H \hat{I}_h^H \tilde{u}^h - I_h^H N^h \tilde{u}^h}_{\tau_h^H}$$

- Prolong, post-smooth, compute error $e^h = \acute{u}^h - (N^h)^{-1} \acute{f}^h$
- Coarse grid *with* τ is nearly $10\times$ better accuracy

τ adaptivity: an idea for heterogeneous media

- Applications with localized nonlinearities
  - Subduction, rifting, rupture/fault dynamics
  - Carbon fiber, biological tissues, fracture
  - **Phase-field models for fracture**
  - **Crystal growth in irregular media**
- Adaptive methods fail for heterogeneous media
  - Rocks are rough, solutions are not “smooth”
  - Cannot build accurate coarse space without scale separation
- τ adaptivity
  - Fine-grid work needed everywhere at first
  - Then τ becomes accurate in nearly-linear regions
  - Only visit fine grids in “interesting” places: active nonlinearity, drastic change of solution

Comparison to nonlinear domain decomposition

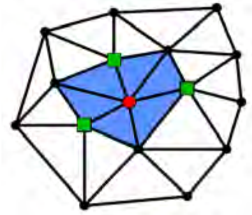
- ASPIN (Additive Schwarz preconditioned inexact Newton)
  - Cai and Keyes (2003)
  - More local iterations in strongly nonlinear regions
  - Each nonlinear iteration only propagates information locally
  - Many real nonlinearities are activated by long-range forces
    - locking in granular media (gravel, granola)
    - binding in steel fittings, crack propagation
  - Two-stage algorithm has different load balancing
    - Nonlinear subdomain solves
    - Global linear solve
- τ adaptivity
  - Minimum effort to communicate long-range information
  - Nonlinearity sees effects as accurate as with global fine-grid feedback
  - Fine-grid work always proportional to “interesting” changes

Nonlinear and matrix-free smoothing

- matrix-based smoothers require global linearization
- nonlinearity often more efficiently resolved locally
- nonlinear additive or multiplicative Schwarz
- nonlinear/matrix-free is good if

$$C = \frac{(\text{cost to evaluate residual at one "point"}) \cdot N}{(\text{cost of global residual})} \sim 1$$

  - finite difference: $C < 2$
  - finite volume: $C \sim 2$, depends on reconstruction
  - finite element: $C \sim$ number of vertices per cell
- larger block smoothers help reduce C
- additive correction (Jacobi/Chebyshev/multi-stage)
  - global evaluation, as good as $C = 1$
  - but, need to assemble corrector/scaling
  - need spectral estimates or wave speeds



Plan: ruthlessly eliminate communication

- Eliminate, not "aggregate and amortize"

Why?

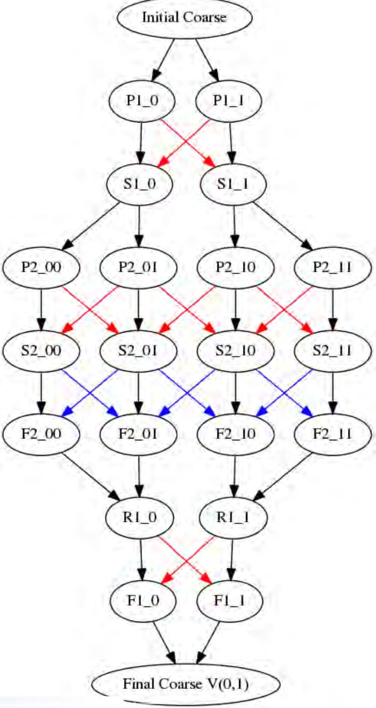
- Enables pruning unnecessary work
- More scope for dynamic load balance
- Tolerance for high-frequency load imbalance
  - From irregular computation or hardware error correction
- Local recovery despite global coupling

Requirements

- Must retain optimal convergence with good constants
- Flexible, robust, and debuggable

Low communication MG

- red arrows can be removed by τ-FAS with overlap
- blue arrows can also be removed, but then algebraic convergence stalls when discretization error is reached
- no simple way to check that discretization error is obtained
- if fine grid state is not stored, use compatible relaxation to complete prolongation P
- "Segmental refinement" by Achi Brandt (1977)
- 2-process case by Brandt and Diskin (1994)

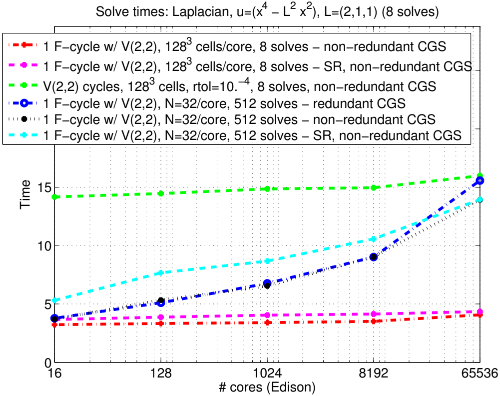


Segmental refinement: no horizontal communication

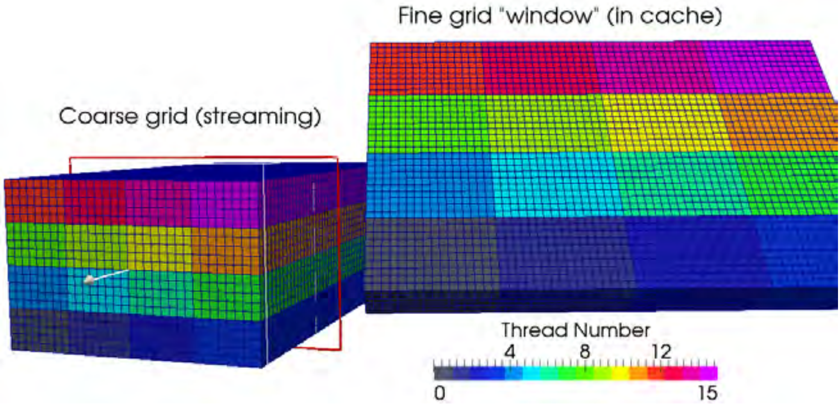
- Adams, Brown, Knepley, Samtaney, arXiv:1406.7808
- 27-point second-order stencil, manufactured analytic solution
- 5 SR levels: 16^3 cells/process local coarse grid
- Overlap $=$ Base $+ (L - \ell)$Increment
  - Implementation requires even number of cells—round down.
- FMG with $V(2,2)$ cycles

Table: $\|e_{SR}\|_\infty / \|e_{FMG}\|_\infty$

	Base		
Increment	1	2	3
1	1.59	2.34	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00



Reducing memory bandwidth



- Sweep through “coarse” grid with moving window
- Zoom in on new slab, construct fine grid “window” in-cache
- Interpolate to new fine grid, apply pipelined smoother (*s*-step)
- Compute residual, accumulate restriction of state and residual into coarse grid, expire slab from window

Arithmetic intensity of sweeping visit

- Assume 3D cell-centered, 7-point stencil
- 14 flops/cell for second order interpolation
- $\geq$ 15 flops/cell for fine-grid residual or point smoother
- 2 flops/cell to enforce coarse-grid compatibility
- 2 flops/cell for plane restriction
- assume coarse grid points are reused in cache
- Fused visit reads u^H and writes $\hat{I}_h^H u^h$ and $I_h^H r^h$
- Arithmetic Intensity

$$\frac{\overbrace{15}^{\text{interp}} + \overbrace{2\cdot(15+2)}^{\text{compatible relaxation}} + \overbrace{2\cdot 15}^{\text{smooth}} + \overbrace{15}^{\text{residual}} + \overbrace{2}^{\text{restrict}}}{3\cdot \texttt{sizeof(scalar)} / \underbrace{2^3}_{\text{coarsening}}} \gtrsim 30 \tag{1}$$

- Still $\gtrsim$ 10 with non-compressible fine-grid forcing

Regularity

Accuracy depends on operator regularity

- Even with regularity, we can only converge up to discretization error, unless we add a *consistent* fine-grid residual evaluation
- Visit fine grid with some overlap, but patches do not agree exactly in overlap
- Need decay length for high-frequency error components (those that restrict to zero) that is bounded with respect to grid size
- Required overlap J is proportional to the number of cells to cover decay length
- Can enrich coarse space along boundary, but causes loss of coarse-grid sparsity
- Brandt and Diskin (1994) has two-grid LFA showing $J \lesssim 2$ is sufficient for Laplacian
- With L levels, overlap $J(k)$ on level k,

$$2J(k) \geq s(L-k+1)$$

  where s is the smoothness order of the solution or the discretization order (whichever is smaller)

Other uses of segmental refinement

- Compression of solutions, local decompression, resilience
- Transient adjoints
  - Adjoint model runs backward-in-time, needs state from solution of forward model
  - Status quo: hierarchical checkpointing
  - Memory-constrained and requires computing forward model multiple times
  - If forward model is stiff, each step has global dependence
  - Compression via τ-FAS accelerates recomputation, can be local
- Visualization and analysis
  - Targeted visualization in small part of domain
  - Interesting features emergent so can't predict where to look

Outlook

- τ adaptivity: benefits of AMR without fine-scale smoothness
- Coarse-centric restructuring is a major interface change
- Nonlinear smoothers (and discretizations)
  - Smooth in neighborhood of “interesting” fine-scale features
  - Which discretizations can provide efficient matrix-free smoothers?
- Weakening data dependencies enables dynamic load balancing
- Reliability of error estimates for refreshing τ
  - We want a coarse indicator for whether τ needs to change
  - Phase fields can provide such information
- Exploit structure or explain why it is not exploitable

Nonlinear deflation for finding multiple distinct solutions

Patrick Farrell, Ásgeir Birkisson, Simon Funke arXiv:1410.5620

- Find a solution $F(u^*) = 0$
- Deflate system using $\eta(u; u^*) = \|u - u^*\|$ or variants

$$G(u) = \frac{F(u)}{\eta(u; u^*)} = 0$$

- Jacobian has structure

$$\nabla_u G(u) = \frac{\nabla_u F(u)}{\eta(u; u^*)} - \underbrace{\frac{F(u)}{\eta^2(u; u^*)} \eta'(u; u^*)}_{\text{rank 1, dense}}$$

- Apply Jacobian matrix-free, precondition using Woodbury formula
- Application to Allen-Cahn, Yamabe, Navier-Stokes, and others
- Preconditioning using PETSc's GAMG; number of iterations constant for each solution
- Extension to nonlinear eigenproblems?

FAS-EIS for eigenproblems

Cohen, Kronik, Brandt, *Locally Refined Multigrid Solution of the All-Electron Kohn-Sham Equation*, 2013; Brandt, *Multiscale calculation of many eigenfunctions*, 2003; Livne 2001

- Proposes techniques for fast solution of self-consistent Kohn-Sham
- Exact Interpolation Scheme (EIS)
  - Adapts coarse basis functions to represent eigenfunctions
  - Orthogonality preserved by interpolation: $O(NK + K^3)$ or better
  - Can exploit localization
  - Related to bootstrap AMG
- Multiscale Eigen-Basis
  - K eigenvalues in $O(N \log K)$
  - Transform to/from eigenbasis in $O(N \log K)$
  - Works in 1D, generalization hard

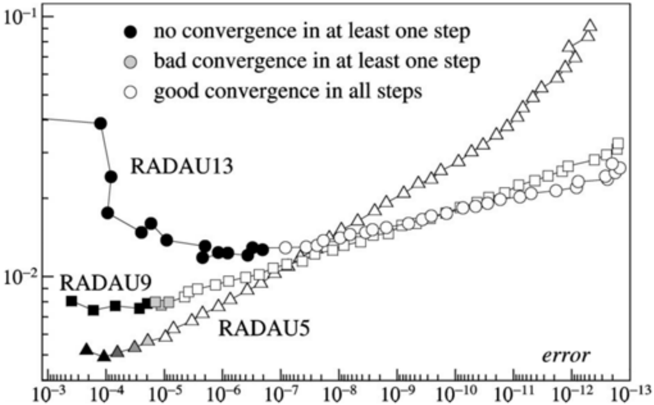
What is performance?

- Cost-Accuracy tradeoff
- Versatility with respect to external requirements

What is performance?

- **Cost-Accuracy** tradeoff
- **Versatility** with respect to external requirements

Work-precision diagram: *de rigueur* in ODE community



[Hairer and Wanner (1999)]

- Tests discretization, adaptivity, algebraic solvers, implementation
- No reference to number of time steps, number of grid points, etc.

Edison, SuperMUC, Titan