Adaptive regularization strategies for nonlinear PDE
Coarse mesh solvers for quasilinear convection diffusion

Sara Pollock

Texas A&M Department of Mathematics

SIAM Conference on Computational Science and Engineering, March 14, 2015
Initial observations on finite element approximation: $-\text{div}(\kappa(u)\nabla u) - f = 0$

- Nonmonotone problems with *internal layers*: Newton iterations may not converge.
- Coarse-mesh solutions display nonphysical oscillations typical of linear convection-dominated problems. Linearized form:
  
  $$B(u; w, v) = (\kappa(u)\nabla w, \nabla v) + (\kappa'(u)w\nabla u, \nabla v)$$

- From a coarse mesh, adaptively refining for local features is not sufficient.
  - *efficiency* is a *local* problem,
  - *stability* is a *global* problem.

- The coarse mesh approximation to a smooth problem may not be smooth. The Jacobian may be ill-conditioned and indefinite.  
  
  $$\kappa(u) \Downarrow$$
Recent results

Recent papers:


Main points:

- Newton-like method that stabilizes the path from the initial guess to a steady state solution.
- Fitting the solver into an adaptive method constructing an adaptive mesh and sequence of iterates through the coarse mesh, pre-asymptotic and asymptotic regimes.
- Appropriate stopping criteria to choose transitional solutions in the pre-asymptotic regime that may be used to with error indicators to refine the mesh.
Problem 1: quasilinear diffusion:

\[ g(u) = -\text{div}(\kappa(u)\nabla u) - f = 0 \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega. \]

Problem 2: quasilinear convection diffusion:

\[ g(u) = -\text{div}(\kappa(u)\nabla u) + b(u) \cdot \nabla u - f = 0 \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega. \]

If \( \kappa(s) \in C^2(\mathbb{R}) \) satisfies \( \kappa(s) \geq \alpha \geq 0 \) and \( \kappa(s), \kappa'(s) \) and \( \kappa''(s) \), and there is a solution \( u \in H^1_0(\Omega) \cap W^{2+\varepsilon}_2(\Omega), \varepsilon > 0 \) with \( g'(u) : H^1_0(\Omega) \to H^{-1}(\Omega) \) an isomorphism then both problems have isolated solutions. For problem 1 There is a unique \( u \in W^{1,p}(\Omega) \).

Local uniqueness and approximation properties of the finite element solution hold assuming the mesh is fine enough.

The coarse mesh problem may be considered noisy and ill-posed.

**Adaptive method**: Assume the initial mesh is not fine enough.

1. Construct a sequence of efficient adaptive partitions over which the problem is well-resolved and stable.
2. Construct sequence of transitional solutions to determine mesh refinement and to start a Newton-like iteration on the next partition.
Newton-like updates: Tikhonov and $\Psi^{tc}$

**Newton update:** linearize $g(x) = 0 \Rightarrow g'(x^n)w^n = -g(x^n), \quad x^{n+1} = x^n + w^n$.

**Pseudo-transient continuation:** Discretize $\partial x / \partial t + g(x) = 0$ with backward Euler:

$$(\Delta t_n^{-1} I + g'(x^n))w^n = g(x^n), \quad x^{n+1} = x^n + w^n.$$  \hspace{1cm} (1)

**Tikhonov regularization** of the linearization with positive semidefinite $R$:

minimize $G_\alpha(w) = \|g'(x^n)w + g(x^n)\|^2 + \alpha\|Rw\|^2$, yields iteration

$$(\alpha_n R^T R + g'(x^n)^T g'(x^n))w^n = -g'(x^n)^T g(x^n), \quad x^{n+1} = x^n + w^n,$$

which coincides with the pseudo-transient continuation stabilization of $\partial(Rx)/\partial t + g(x) = 0$ with $\alpha = 1/\Delta t$ using the normal equations.

Or without the normal equations:

$$(\alpha_n R + g'(x^n))w^n = -g(x^n), \quad x^{n+1} = x^n + w^n.$$  \hspace{1cm} (2)

**Common idea:** stabilize the Jacobian by penalizing against changes between iterates.

**Good news:** Iteration (1) stabilizes the linearization when the Jacobian is indefinite, which it often is on the coarse mesh.

**Good news:** The method is general: $R$ may be chosen to regularize only bad dof.

**Bad news:** Loss of sparsity, condition, question of criteria for switching from (1) to (2).
Tikhonov regularization

minimize $G_{\alpha}(w) = \|g'(x^n)w + g(x^n)\|_{L^2}^2 + \alpha \|Rw\|_{L^2}^2$,

yields iteration

$$(\alpha_n R^T R + g'(x^n)^T g'(x^n))w^n = -g'(x^n)^T g(x^n) \quad x^{n+1} = x^n + w^n,$$

or for positive $g'(x^n)$ the Larentiev iteration

$$(\alpha_n R + g'(x^n))w^n = -g(x^n) \quad x^{n+1} = x^n + w^n.$$  

Choice of selective regularization/penalty term $R$

- *A priori* information about the solution, e.g., smoothness, admissible frequencies, can be exploited.
- Reduce corrections nonphysical features, e.g., sharp spikes.
- Allow smooth corrections.
- Define $R$ locally:
  - dof in regions of low curvature in null ($R$)
  - dof in regions of high curvature far from null ($R$)

*Example:* $R$ based on the Laplacian, regularized dof selected by local error indicators, e.g., the jump term in a residual based indicator.
Improved method: continuation point of view

(1) Replace backward Euler discretization with Newmark update

\[ x^{n+1} - x^n = \Delta t_n \left\{ (1 - \gamma) \dot{x}^n + \gamma \dot{x}^{n+1} \right\}, \quad \dot{x} = \partial x / \partial t. \]

Reason: for \( \gamma > 1 \), controllable increase in numerical dissipation.

**Problem:** Resulting iteration still requires use of normal equations to stabilize ill-conditioned indefinite Jacobians.

(2) Stabilize the Newmark update by \( \sigma \)-splitting

\[ x^{n+1} - x^n = \Delta t_n \left\{ \sigma (1 - \gamma) \dot{x}^n + (1 - \sigma)(1 - \gamma) \dot{x}^n + \sigma \gamma \dot{x}^{n+1} + (1 - \sigma) \gamma \dot{x}^{n+1} \right\}, \]

and linearize the \( (1 - \sigma) \) fraction by \( R(\dot{x}^{n+1} - \dot{x}^n) = g(x^{n+1}) - g(x^n) = g'(\bar{x}) w^n \) for fixed \( \bar{x} \).

Resulting iteration:

\[ \left( \alpha_n R + \gamma \left\{ (1 - \sigma) g'(\bar{x}) + \sigma g'(x^n) \right\} \right) w^n = -g(x^n), \quad x^{n+1} = x^n + w^n, \]

**Good news:** Stabilizes iteration by controlling sign of small eigenvalues without use of normal equations.

**Bad news:** Newmark update decreases rate of convergence from quadratic to \( q \)-linear with rate \( q = 1 - 1/\gamma \).
Basic algorithm

Start with initial $x^0, \gamma, \sigma_0$ and $K_0$. On partition $\mathcal{T}_k$, $k = 0, 1, 2, \ldots$

1. Compute $R, g'(\bar{x})$.
2. Set $\alpha_0 = \|g(x^0)\|$.
3. While Exit criteria are not met on iteration $n - 1$:
   
   (a) Set $\sigma = \max\{\sigma_0, 1 - \|g(x^n)\|/K_0\}$
   (b) Solve $(\alpha_n R + \gamma\{\sigma g'(x^n) + (1 - \sigma)g'(\bar{x})\})\Delta x^n = g(x^n)$.
   (c) Update $x^{n+1} = x^n + \Delta x^n$.
   (d) Update $\alpha_n$.
4. Update $\gamma$ for partition $\mathcal{T}_{k+1}$.

$\alpha$-update: For local convergence: $\alpha_n \leq \|g(x^n)\|$, so long as residual is decreasing.

Set $\beta_0 = 1$. For $n \geq 1$: $\alpha_n = \beta_n\|g(x^n)\|$, with $\hat{\beta}_n = \frac{\|g(x^n)\|}{\|g(x^{n-1})\|}$.

$\beta_n$-correction: If $\|g(x^n)\| < \|g(x^{n-1})\| \Rightarrow \beta_{n-1}/2 \leq \hat{\beta}_n \leq 1$.
If $\|g(x^n)\| > \|g(x^{n-1})\| \Rightarrow \hat{\beta}_n \leq 2\beta_{n-1}$
Exit strategy

Given a user set tolerance $tol$, accepted rate of convergence $q_{acc} = 1 - 1/(M\gamma)$ for some constant $M$, e.g., $M = 2$, and a maximum number of iterations either chosen as a constant or based on the predicted rate of convergence, exit the solver on partition $T_k$ after calculating iterate $x^{n+1}$ when one of the conditions holds.

**Exit Criteria:**

1. $\|g(x^{n+1})\| \leq tol$.
2. (a) $\|g(x^{n+1})\| < \|g(x^n)\|$, AND
   (b) $\|g(x^n)\| < \min\{\|g(x^0)\|, \|g(x_{k-1})\|\}$, AND
   (c) $\|g(x^{n+1})\|/\|g(x^n)\| < q_{acc}$, AND
   (d) $\|g(x^{n+1})\|/\|g(x^n)\| > \|g(x^n)\|/\|g(x^{n-1})\|$. ***
3. Maximum number of iterations exceeded.

$\gamma$-update: if the predicted error reduction rate is achieved, $\gamma$ should be reduced; and if the iteration fails, more stability is needed and $\gamma$ should be increased.

Based on the above exit criteria:

1. If $\gamma_k > 1$ set $1 \leq \gamma_{k+1} < \gamma_k$.
2. If $\|g(x^{n+1})\|/\|g(x^n)\| \leq 1 - 1/\gamma_k$, or within tolerance, set $1 \leq \gamma_{k+1} < \gamma_k$.
3. If $\|g(x^{n+1})\|/\|g(x^n)\| > q_{acc}$, set $\gamma_{k+1} > \gamma_k$. 

S. Pollock 8/15
Texas A&M Mathematics
CSE15 March 14, 2015
Updated marking strategy

Three possible outcomes on exit from the solver:

(1) The iterate \( x_k \) has residual \( \|g(x_k)\| \leq \text{tol} \).
   \[ \implies \text{Refine based on local error indicators.} \]

(2) The iterate \( x_k \) has residual \( \|g(x_k)\| > \text{tol} \).
   \[ \implies \text{Split the refinement between} \]
   \[ \bullet \text{local error indicators: largest error;} \]
   \[ \bullet \text{coarsest set of elements, chosen by largest indicator.} \]

(3) The solver failed and \( x_k \) is reset to zero.
   \[ \implies \text{Refine the coarsest set of elements.} \]

Example: Dörfler marking strategy. \( \theta = \theta_C + \theta_F \) and the marked sets
\( M_F \subset T_k \) and \( M_C \subset T_k \) are chosen by sets of least cardinality with
\[ \sum_{T \in M_F} \eta_T^2 \geq \theta_F \sum_{T \in T_k} \eta_T^2, \quad \sum_{T \in M_C} \eta_T^2 \geq \theta_C \sum_{T \in T_k} \eta_T^2, \] marked set: \( M = M_F \cup M_C \).

\[ \theta_C = \Phi(\theta) = \theta \left( \frac{1}{2} + \frac{1}{\pi} \arctan \left( \|g(x_k)\|/100 - \pi/2 \right) \right). \]

Case (1): \( \theta_F = \theta \) and \( \theta_C = 0 \). Case (2): \( \theta_F \to \theta \) as \( \|g(x_k)\| \to 0 \).
Model problem with concentric layers

Quasilinear diffusion problem on $\Omega = (0, 1)^2$.

$$-\text{div}(\kappa(u) \nabla u) = f \quad \text{in} \quad \Omega = (0, 1), \quad u = 0 \quad \text{on} \quad \partial \Omega.$$ 

$$\kappa(s) = k + \frac{1}{((\varepsilon + (s-a)^2) + \frac{1}{((\varepsilon + (s-c)^2)}, \quad \varepsilon = 6 \times 10^{-4}, \quad a = 0.5, \quad c = 0.8 \quad \text{and} \quad k = 1.$$ 

$f(x,y)$ chosen so the exact solution $u(x,y) = \sin(\pi x) \sin(\pi y)$. Parameters $a$ and $c$ control the position of the internal layers. The initial mesh has 144 elements. The Dörfler parameter $\theta = 0.6$. 

Figure: Solution snapshots from pre-asymptotic regime after 25, 30 and 35 adaptive refinements.
Numerics: pre-asymptotic mesh partitions

\[ \int \kappa(u) \nabla u \cdot \nabla v = \int f v \]

\[ \kappa(u) = 1 + \frac{1}{6 \times 10^{-4} + (u - 0.5)^2} + \frac{1}{6 \times 10^{-4} + (u - 0.8)^2}. \]

Figure: Mesh after 25, 30 and 35 adaptive refinements.
### Numerics: Error Reduction

<table>
<thead>
<tr>
<th>iter</th>
<th>|g(u_k)|</th>
<th>γ</th>
<th>dof</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>604.982</td>
<td>13</td>
<td>877</td>
</tr>
<tr>
<td>25</td>
<td>345.184</td>
<td>13</td>
<td>911</td>
</tr>
<tr>
<td>30</td>
<td>93.2727</td>
<td>9</td>
<td>1092</td>
</tr>
<tr>
<td>35</td>
<td>12.3063</td>
<td>3</td>
<td>2138</td>
</tr>
<tr>
<td>37</td>
<td>1.41396e-11</td>
<td>1</td>
<td>3661</td>
</tr>
</tbody>
</table>

**Left:** Residual from the weak form, parameter \(\gamma\), and degrees of freedom in the pre-asymptotic regime.

**Right:** \(H^1\) error (below) and error estimator (above) vs. \(n^{-1/2}\), with \(n\) the number of elements.

**Observe:** The number of elements refined at each iteration increases through the pre-asymptotic regime due to increased equidistribution of the error.
Model problem with convection and diffusion layers

\[-\text{div}(\kappa(u)\nabla u) + b(u) \cdot \nabla u = f \quad \text{in } \Omega = (0, 1), \quad u = 0 \quad \text{on } \partial \Omega.\]

\[\kappa(s) = k + \frac{1}{((\varepsilon + (s - a)^2) \varepsilon = 6 \times 10^{-4}, \quad a = 0.5, \quad \text{and } k = 1.\]

\[b(u) = \left( \frac{2}{((\varepsilon + (s - c)^2)}, \frac{1}{((\varepsilon + (s - c)^2)} \right)^T, \quad c = 0.6.\]

\[f(x, y) \text{ chosen so the exact solution } u(x, y) = \sin(\pi x) \sin(\pi y). \quad \text{The initial mesh has 144 elements. The Dörfler parameter } \theta = 0.6.\]

Figure: Solution snapshots from pre-asymptotic regime after 36, 39 and 42 adaptive refinements.
\[ \int \kappa(u) \nabla u \cdot \nabla v + b(u) \cdot \nabla u = \int f v \]

Above: Mesh after 36, 39 and 42 adaptive refinements.

Right: Residual from the weak form, parameter \( \gamma \), and degrees of freedom in the pre-asymptotic regime.

<table>
<thead>
<tr>
<th>iter</th>
<th>( | g(u_k) | )</th>
<th>( \gamma )</th>
<th>dof</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>447.078</td>
<td>18</td>
<td>1220</td>
</tr>
<tr>
<td>36</td>
<td>21.724</td>
<td>7</td>
<td>1381</td>
</tr>
<tr>
<td>39</td>
<td>9.716</td>
<td>7</td>
<td>1500</td>
</tr>
<tr>
<td>42</td>
<td>3.340</td>
<td>3</td>
<td>1722</td>
</tr>
<tr>
<td>44</td>
<td>4.488e-10</td>
<td>1</td>
<td>2073</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Current directions

- Targeted regularization strategy with detailed analysis of parameters.

- Generalization to inner iteration for inexact linear solves.

- Nonlinear solver and regularization strategy for quasilinear reaction-diffusion describing complete sets of solutions (W. Rundell, SP).

- Application of nonlinear solver to generalized multiscale method for high contrast semilinear problem. (G. Li, SP).

Acknowledgements

William Rundell, Texas A&M University
Main references


- Xu, J., Two-grid discretization techniques for linear and nonlinear PDEs, SINUM, 33, 5, 1759–1777, 1996