SAND2025-04681C

An Adaptive Finite Element Trust-Region Method for Regularized PDE-Constrained Optimization

H. Antil, **R. Baraldi**, R. Khandelwal, D. Kouri

ECOM 2025

George Mason University

April 17, 2025

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2025 –

Problem Formulation

Goal: Develop efficient algorithms to solve the regularized nonsmooth optimization problem,

$$\boxed{\min_{z \in \mathcal{Z}} F(z) := f(z) + \phi(z).} \tag{1}$$

- $\mathcal{Z}$ is a Hilbert space with $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$;
- $\phi : \mathcal{Z} \to [-\infty, \infty]$ is proper, closed, and convex, but may be nonsmooth;
- $f : \mathcal{Z} \to \mathbb{R}$ has Lipschitz continuous gradients on an open set containing $\operatorname{dom} \phi$;
- $F > -\infty$, bounded below on $\operatorname{dom} \phi$.

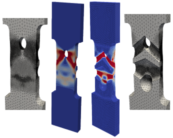
Key Requirements

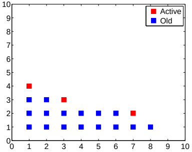
1. **Large-scale Problems**: Rapid convergence, mesh independence, and matrix free;
2. **Leverage Inexactness**: Converges even for inexact computations.

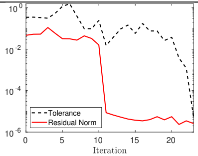
$f(z)$ may be nonconvex and often impossible to compute exactly.

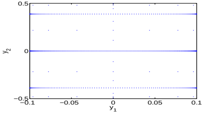
- Stems from discretization, iterative procedures, adaptive model reduction, surrogate models, iterative linear and nonlinear solves, etc.

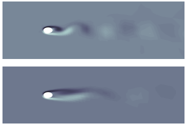
Adaptive Finite Elements	Adaptive Quadrature	Adaptive Compression











Nonsmooth Regularizers

$\phi(z)$ are typically sparsity-inducing and temper model complexity.

- **Sparse regularization**: $\mathcal{Z} \hookrightarrow L^1(\Omega)$, $\beta \in L^\infty(\Omega)$, $\beta \geq 0$ a.e., and

$$\phi(z) = \int_\Omega \beta(\omega)|z(\omega)|\mathrm{d}\omega.$$

  Applications: Optimal control, data-science, learning, basis-pursuit.

- **Total Variation**: $\mathcal{Z} \hookrightarrow BV(\Omega)$, $\beta \in L^\infty(\Omega)$, $\beta \geq 0$ a.e,

$$\phi(z) = \int_\Omega \beta(\omega)|\nabla z(\omega)|\mathrm{d}\omega.$$

  Applications: Image processing, digital image correlation, topology optimization.

- **Convex Constraints**: $C \subset \mathcal{Z}$ nonempty, closed and convex,

$$\phi(z) = \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{otherwise.} \end{cases}$$

  Applications: Optimal control, inverse problems, optimal design.

Outline

Goal: Incorporate adaptive mesh refinement into inexact trust-region framework [4]

- Trust Region Algorithm
  - Inexactness conditions
- Error Estimates for AFEM
- Numerical Example

Basic Algorithm

Algorithm 1: Nonsmooth Trust-Region Method

Data: $z_0 \in \operatorname{dom}\phi$, $\Delta_0 > 0$, $0 < \eta_1 < \eta_2 < 1$, and $0 < \gamma_1 \le \gamma_2 < 1 \le \gamma_3$

for $k = 1, 2, \ldots$ **do**

Model Selection: Select subproblem model f_k of f near z_k.

Step Computation: Compute $z_{k+1} \in \mathcal{Z}$ that approximately solves

$$\min_{z \in \mathcal{Z}} \{ m_k(z) := f_k(z) + \phi(z) \} \quad \text{subject to} \quad \|z - z_k\| \le \Delta_k.$$

Evaluate Objective: Compute $\mathrm{cred}_k \approx \mathrm{ared}_k$.

Step Acceptance: Compute ratio of computed and predicted reduction

$$\rho_k := \frac{\mathrm{cred}_k}{m_k(z_k) - m_k(z_{k+1})} < \eta_1 \quad \Rightarrow \quad z_{k+1} \leftarrow z_k.$$

Update Trust-Region Radius: $\Delta_{k+1} \in \begin{cases} [\gamma_1 \Delta_k, \gamma_2 \Delta_k], & \text{if } \rho_k < \eta_1 \\ [\gamma_2 \Delta_k, \Delta_k], & \text{if } \rho_k \in [\eta_1, \eta_2) \\ [\Delta_k, \infty), & \text{if } \rho_k > \eta_2. \end{cases}$

end

(Inexactness) Challenges

Recall: infinite-dimensional optimization function and gradient evaluations are often impossible to compute without discretization, leading to inexactness.

1. Ensure subproblem step yields decrease - existence of a Cauchy point ([1, 3, 2]);
2. Handle nonsmooth subproblems efficiently - subproblem solvers ([3]);
3. Ensure convergence with f_k, g_k inexact - adaptive FE.
   - Assume approximations of objective f_k and gradient g_k have the form:

$$f_k(z,\tau) : \mathcal{Z} \times [0,+\infty) \to \mathbb{R}, \quad f_k(z,0) = f(z)$$
$$g_k(z,\tau) : \mathcal{Z} \times [0,+\infty) \to \mathcal{Z}, \quad g_k(z,0) = \nabla f(z).$$

   - Trust-region algorithm will 1) controls error in evaluations via satisfying 2) computable error estimates.
   - **Algorithm selects tolerance**, and we must find grid whose error estimates are below given tolerance.

Inexactness Conditions

Objective function approximates the decrease

$$\mathrm{cred}_k \approx \mathrm{ared}_k := (f(z_k) + \phi(z_k) - (f(z_{k+1}) + \phi(z_{k+1}))$$

so for pred_k predicted model decrease, the objective bound (with $\lim_{k\to+\infty} \theta_k = 0$) is

$$\boxed{|\mathrm{ared}_k - \mathrm{cred}_k| \leq \kappa_{\mathrm{obj}} \left[\eta \min\{\mathrm{pred}_k, \theta_k\}\right]^{\zeta},} \tag{2}$$

The **gradient** bound is, for $\kappa_{\mathrm{grad}} > 0$ and $h_k := t_0^{-1} \|\mathrm{Prox}_{t_0\phi}(z_k - t_0 g_k) - z_k\|_{\mathcal{Z}}$:

$$\boxed{\|g_k - \nabla f(z_k)\| \leq \kappa_{\mathrm{grad}} \min\{\Delta_k, h_k\} \quad \forall\, k.} \tag{3}$$

Conclusion: There exists iterative procedures that satisfy (2) & (3) in finite time.

Computable A Posteriori Error Estimates

Consider full space (for a moment):

$$\min_{u\in\mathcal{U},z\in\mathcal{Z}} F(u,z) := f(u,z) + \phi(z) \quad \text{subject to} \quad c(u,z) = 0. \tag{4}$$

with Lagrangian

$$L(u,z,\lambda) = f(u,z) + \langle \lambda, c(u,z)\rangle_{\mathcal{U},\mathcal{U}^*}\,.$$

For given mesh parameter $h > 0$, let $\eta_h^k := (\eta_h^{c,k}, \eta_h^{L_u,k}, \eta_h^{L_z,k})$ with $\eta_h^{i,k} \to 0$ as $h \to 0$ where $i \in \{c, L_u, L_z\}$ such that

$$\|L_z(u_h^k, z_h^k, \lambda_h^k)\|_{\mathcal{Z}^*} \lesssim \underbrace{\|g_{k,h}\|_{\mathcal{Z}_h^*} + \|c_h(u_h,z_h)\|_{\mathcal{U}_h^*}}_{\eta_h^{L_z,k}} + \|L_z(u_h^k, z_h^k, \lambda_h^k)\|_{\mathcal{Z}_h^*}\,,$$

$$\|L_u(u_h^k, z_h^k, \lambda_h^k)\|_{\mathcal{U}^*} \lesssim \underbrace{\|g_{k,h}\|_{\mathcal{Z}_h^*} + \|c_h(u_h,z_h)\|_{\mathcal{U}_h^*}}_{\eta_h^{L_u,k}} + \|L_u(u_h^k, z_h^k, \lambda_h^k)\|_{\mathcal{U}_h^*}\,,$$

$$\|c(u_h^k, z_h^k)\|_{\mathcal{U}^*} \lesssim \underbrace{\|g_{k,h}\|_{\mathcal{Z}_h^*}}_{\eta_h^{c,k}} + \|c_h(u_h^k, z_h^k)\|_{\mathcal{U}_h^*}\,.$$

Adaptive Finite Elements

Goals: Use inexact state/adjoint estimates to adaptively refine finite element mesh.

Our cost function is the usual regularized data misfit with L^1 penalty

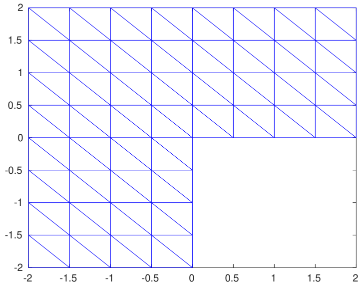
$$\min_{z\in\mathcal{Z}} \int_\Omega \left\{ \frac{1}{2}\|u(z)-u_d\|^2_{L^2(\Omega)} + \dots \right.$$

$$\left. \dots \frac{\alpha}{2}\|z\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|z\|_{L^1(\Omega)} \right\} \mathrm{d}x$$

for $\mathcal{U} \equiv H^1_0(\Omega), \mathcal{Z} \equiv L^2(\Omega)$ and

$$\begin{aligned} -\Delta u = z, &\quad x \in \Omega \subset \mathbb{R}^2 \\ u = 0, &\quad x \in \partial\Omega \end{aligned}$$

$\alpha = 10^{-4}$, $\beta = 10^{-2}$, Q2 discretization for states and P0 for controls; stopping tolerance 10^{-6}.



Initial mesh for nonconvex domain in Poisson problem.

Error Estimators for Poisson

We introduce the following the computable error estimators:

$$\eta_h^{c,k} := \sum_{T\in\mathcal{T}_h} h_T^2 \| -\Delta u_h^k - z_h\|_{L^2(T)}^2 + \frac{1}{2}\sum_{e\in\mathcal{E}_h} \| | [\frac{\partial u_h^k}{\partial n}] | |_{L_2(e)}^2,$$

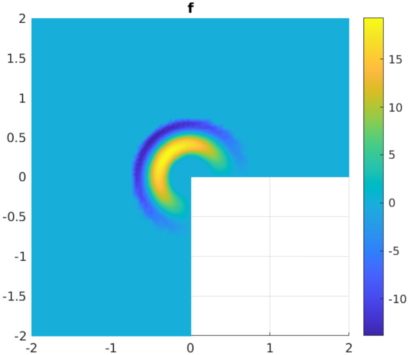
$$\eta_h^{L_u,k} := \sum_{T\in\mathcal{T}_h} h_T^2 \| u_h^k - u_d - \Delta\lambda_h^k\|_{L^2(T)}^2 + \frac{1}{2}\sum_{e\in\mathcal{E}_h} | | [\frac{\partial \lambda_h^k}{\partial n}] | |_{L_2(e)}^2,$$

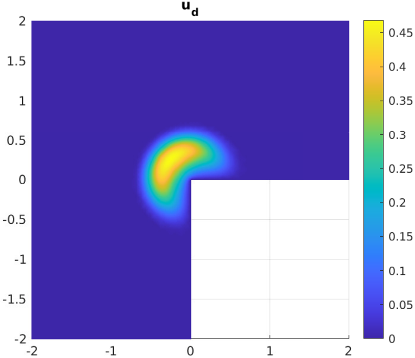
yields the following adaptive mesh refinement criteria:

$$\eta_h^{c,k} \lesssim \min\{\|\nabla f_h^k\|_{\mathcal{Z}_h^*} + \|c_h(u_h^k, z_h^k)\|_{\mathcal{U}_h^*}, \tau\},$$
$$\eta_h^{L_u,k} \lesssim \min\{\|\nabla f_h^k\|_{\mathcal{Z}_h^*} + \|c_h(u_h^k, z_h^k)\|_{\mathcal{U}_h^*} + \|L_u(u_h, z_h, \lambda_h)\|_{\mathcal{U}_h^*}, \tau\},$$

where $\tau > 0$ is a given tolerance.

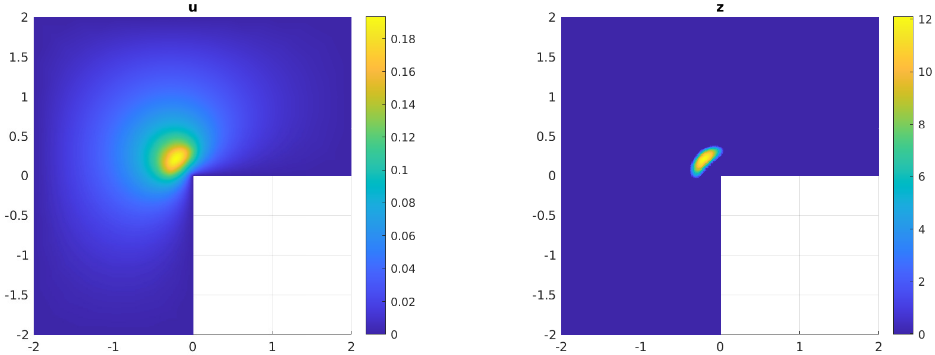
Adaptive Finite Elements: True Solution





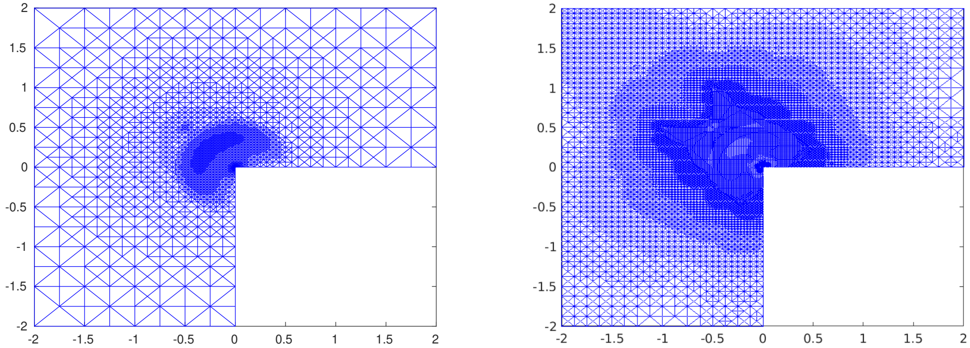
True solution: $u_d(x) = (x_1 - x_1^2) \cdot (x_2 - x_2^2)$ and $f(x) = 2(x_1 - x_1^2) + 2(x_2 - x_2^2)$

Adaptive Finite Elements: $u(z)$ and z



Final state (left) and control (right).

Adaptive Finite Elements: Grid Comparison



Q2 grid (left) and Q1 grid (right).

Conclusions

- Nonsmooth problems are extremely prevalent in the literature and require careful treatment.
- Moreover, numerical solutions require expensive approximations.
  - Objectives and gradients can only be compute inexactly.
- Nonsmooth trust-region is provably convergent even while adaptively changing the FE grid.
- Next: topology optimization problem.

Acknowledgements

Questions?
Thank you-

- Organizers: Harbir Antil, Drew Kouri, Denis Ridzal.
- Collaborators: H.A., D.K., Rohit Khandelwal.

References

[1] R. J. Baraldi and D. P. Kouri. A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations. *Mathematical Programming*, 1(201):559–598, 2023.

[2] R. J. Baraldi and D. P. Kouri. Local convergence analysis of an inexact trust-region method for nonsmooth optimization. *Optimization Letters*, 18:663–680, 2024.

[3] R. J. Baraldi and D. P. Kouri. Efficient subproblem solvers for a nonsmooth trust-region method. *Computational Optimization and Applications*, 90:193–226, 2025.

[4] R. J. Baraldi, R. Kandelwahl, D. P. Kouri, and H. Antil. An adaptive finite element method for nonsmooth optimal control. *in preparation*, 2025.

Inexact Gradient Algorithm

Assume that we have an approximation $\bar{g} : \mathcal{Z} \times [0, +\infty) \to \mathcal{Z}$ that satisfies $C_{\mathrm{grad}} \geq 0$ such that

$$\bar{g}(z, 0) = \nabla f(z) \quad \text{and} \quad \|\nabla f(z) - \bar{g}(z, \tau)\| \leq C_{\mathrm{grad}}\tau \quad \forall \tau \in \mathbb{R}_+. \tag{6}$$

Algorithm 2: Inexact Gradient Computation

Require: Constant $\bar{\kappa}_{\mathrm{grad}} > 0$, a tolerance $\mu_{\mathrm{grad}} > 1$, current iterate $z_k \in \mathcal{Z}$, and Δ_k.

  Set $\tau_k^- \leftarrow \bar{\kappa}_{\mathrm{grad}} \Delta_k$

  Compute $h_k \leftarrow r_0^{-1} \left\| \mathrm{Prox}_{r_0\phi}(z_k - r_0 \bar{g}(z_k, \tau_k^-)) - z_k \right\|$

  Set $\tau_k^+ \leftarrow \bar{\kappa}_{\mathrm{grad}} \min\{h_k, \Delta_k\}$

  while $\mu_{\mathrm{grad}} \tau_k^+ < \tau_k^-$ **do**

    Compute $h_k \leftarrow r_0^{-1} \left\| \mathrm{Prox}_{r_0\phi}(z_k - r_0 \bar{g}(z_k, \tau_k^+)) - z_k \right\|$

    Set $\tau_k^- \leftarrow \tau_k^+$ and $\tau_k^+ \leftarrow \bar{\kappa}_{\mathrm{grad}} \min\{h_k, \Delta_k\}$

  end while

If $h(z_k) > 0$, Algorithm 2 terminates finitely and (3) holds with $\kappa_{\mathrm{grad}} = \mu_{\mathrm{grad}} \bar{\kappa}_{\mathrm{grad}} C_{\mathrm{grad}}$.

Inexact Function Algorithm

Assume $\tau_k^{\mathrm{obj}} \leq \mu_{\mathrm{obj}}\tau_{k+1}^{\mathrm{obj}}$, then $\kappa_{\mathrm{obj}} = (1+\mu_{\mathrm{obj}})\bar{\kappa}_{\mathrm{obj}} C_{\mathrm{obj}}$ since

$$|\mathrm{ared}_k - \mathrm{cred}_k| \leq (1+\mu_{\mathrm{obj}})\bar{\kappa}_{\mathrm{obj}} C_{\mathrm{obj}}[\eta \min\{\mathrm{pred}_k, \theta_k\}]^{\zeta}.$$

Algorithm 3: Inexact Objective Function Computation

Require: Constants $\bar{\kappa}_{\mathrm{obj}} > 0$ and $\mu_{\mathrm{obj}} \geq 1$, the current objective function tolerance τ_k, the current iterate $x_k \in X$ and approximate objective value $v_k = \bar{f}(x_k, \tau_k)$, the new iterate $z_{k+1} \in X$, and the predicted reduction pred_k.

Set $\tau_{k+1} \leftarrow \bar{\kappa}_{\mathrm{obj}}[\eta \min\{\mathrm{pred}_k, \theta_k\}]^{\zeta}$

if $\mu_{\mathrm{obj}}\tau_{k+1} < \tau_k$ **then**

  Compute $v_k \leftarrow \bar{f}(x_k, \tau_{k+1})$

end if

Compute $v_{k+1} \leftarrow \bar{f}(z_{k+1}, \tau_{k+1})$ and set $\mathrm{cred}_k = (v_k + \phi(z_k)) - (v_{k+1} + \phi(z_{k+1}))$

Recall: $h_k := r_0^{-1} \|\mathrm{Prox}_\phi(z_k - r_0 g_k) - z_k\|_{\mathcal{Z}}$

Theorem (Algorithm Convergence)

Let $\{z_k\}$ be the sequence of iterates generated

$$\liminf_{k\to\infty} h_k = 0 \quad \Rightarrow \quad \liminf_{k\to\infty} r_0^{-1} \|\mathrm{Prox}_\phi(z_k - r_0 \nabla f(z_k)) - z_k\|_{\mathcal{Z}} = 0. \tag{7}$$

Note: Permits unbounded model curvature.

Given $\varepsilon > 0$ and bounded model curvature, then Trust-Region Algorithm satisfies $h_k \leq \min\{\varepsilon, 1\}$ in at most $\mathcal{O}(\varepsilon^{-2})$ iterations.

Note: This is a worst-case bound; we find much better performance in practice.

Proximal Operator

Definition (Moreau-Yosida Envelope)

For a proper, lower semicontinuous function $\phi : \mathcal{Z} \to \overline{\mathbb{R}}$ and a parameter $t > 0$, the *Moreau envelope* $e_{t\phi} : \mathcal{Z} \to \mathbb{R}$ and the *proximal operator* $\mathrm{Prox}_{t\phi} : \mathcal{Z} \to \mathcal{Z}$ are defined by

$$e_{t\phi}(y) := \inf_{z \in \mathcal{Z}} \left\{ \frac{1}{2t} \|z - y\|^2 + \phi(z) \right\}, \tag{8a}$$

$$\mathrm{Prox}_{t\phi}(y) := \operatorname*{argmin}_{z \in \mathcal{Z}} \left\{ \frac{1}{2t} \|z - y\|^2 + \phi(z) \right\}. \tag{8b}$$

- Interpretation: extension of cost function to minimizing ϕ and near y.
- Utility: many proximal operators have analytic solutions (can be relaxed somewhat);
  - **L^1-Norm**: $\mathcal{Z} = L^2(\Omega)$, $\phi(z) = \beta \|z\|_{L^1(\Omega)} \Rightarrow \mathrm{Prox}_{t\phi} = \mathrm{sign}(z) \odot (|z| - t\beta)_+$.
  - **ReLU**: $\mathcal{Z} = \mathbb{R}$, $\phi(z) = \max\{0, z\} \Rightarrow \mathrm{Prox}_{t\phi} = \max\{z - t, \min\{0, z\}\}$.

First-Order Necessary Optimality Conditions

Definition (Local Minimizer)

$\bar{z} \in \mathcal{Z}$ is a **local minimizer** of $(f+\phi)$ if there exists a neighborhood U of $\bar{z}$ on which $(f+\phi)(\bar{z}) \leq (f+\phi)(z) \quad \forall z \in U$. Additionally, $\bar{z}$ satisfies

$$\boxed{-\nabla f(\bar{z}) \in \partial\phi(\bar{z}) \quad \Leftrightarrow \quad \bar{z} = \mathrm{Prox}_{t\phi}(\bar{z} - t\nabla f(\bar{z}))\, \forall t > 0.} \tag{9}$$

Lemma (Generalized Gradient & Stationary Point)

If $\bar{z} \in \mathcal{Z}$ is a stationary point of $(f+\phi)$, then $h(\bar{z}) = 0$ where

$$h(z) := t^{-1}\left\|\mathrm{Prox}_{t\phi}(z - t\nabla f(z)) - z\right\|$$

for arbitrary fixed $t > 0$.