SAND2023-12929C

An Inexact Trust-Region Algorithm for Nonsmooth Nonconvex Optimization

Drew P. Kouri, Robert J. Baraldi

SIAM Conference on Computational Science and Engineering
Amsterdam, The Netherland

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology &
owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Secu
DE-NA0003525.

Problem Formulation

Goal: Develop an efficient algorithm to solve the **nonsmooth optimization problem**,

$$\min_{x \in H} \; f(x) + \phi(x).$$

- H is a Hilbert space with inner product $(\cdot, \cdot)$ and associated norm $\|\cdot\|$;
- $\phi : H \to [-\infty, +\infty]$ is proper, closed and convex, but may be **nonsmooth**;
- $f : H \to \mathbb{R}$ has Lipschitz continuous gradients on an open set containing $\mathrm{dom}\phi$;
- $F := f + \phi$ is bounded below on $\mathrm{dom}\phi$.

Problem Formulation

Goal: Develop an efficient algorithm to solve the **nonsmooth optimization problem**,

$$\min_{x \in H} f(x) + \phi(x).$$

- H is a Hilbert space with inner product $(\cdot, \cdot)$ and associated norm $\|\cdot\|$;
- $\phi : H \to [-\infty, +\infty]$ is proper, closed and convex, but may be **nonsmooth**;
- $f : H \to \mathbb{R}$ has Lipschitz continuous gradients on an open set containing domϕ;
- $F := f + \phi$ is bounded below on domϕ.

Key Requirements of Algorithm

1. **Large-Scale Problems:** Rapid convergence, mesh independence, and matrix free.
2. **Leverage Inexactness:** Converges even when f and ∇f are computed inexactly via adaptive discretization, reduced-order modelling, compression, etc.

Motivating Application

Sparse Control

Goal: Determine a control z that produces a state close to w and that has ***small support***.

Given a domain $\Omega \subset \mathbb{R}^d$, a target state $w \in L^2(\Omega)$, bounds $a \le 0 \le b$ a.e., and penalty parameters $\alpha, \beta \ge 0$,

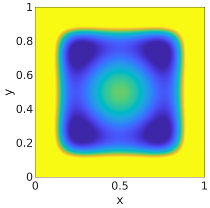
$$\min_{z \in L^2(\Omega)} \int_\Omega |S(z) - w|^2(x)\,\mathrm{d}x + \frac{\alpha}{2}\int_\Omega |z|^2(x)\,\mathrm{d}x + \beta \int_\Omega |z|(x)\,\mathrm{d}x$$

$$\text{subject to} \qquad a \le z \le b \quad \text{a.e.,}$$

where $S(z) = u \in H_0^1(\Omega)$ solves

$$\begin{aligned} -\Delta u + u^3 &= z \quad \text{in } \Omega \\ u &= 0 \quad \text{in } \partial\Omega \end{aligned}$$

Optimal Control

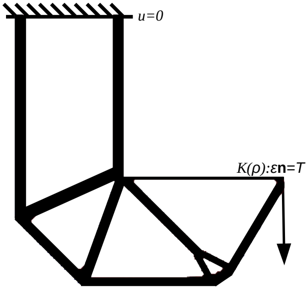


Challenges: Objective function is **nonsmooth**, **nonconvex**, and **expensive**.

Motivating Application

Elastic Topology Optimization

Goal: Determine a ***binary*** ρ that is maximally stiff and that satisfies the volume constraint.



Given a domain $\Omega \subset \mathbb{R}^d$ and a volume fraction $v \in (0,1)$,

$$\min_{\rho \in L^2(\Omega)} \int_{\Gamma_t} T(x) \cdot [S(\rho)](x)\, dx$$

$$\text{subject to} \quad \int_\Omega \rho(x)\, dx \le v|\Omega|, \quad 0 \le \rho \le 1 \ \text{ a.e.},$$

where $S(\rho) = u \in (H^1(\Omega))^d$ solves

$$\begin{aligned} -\nabla \cdot (K(\rho) : \varepsilon) = 0, \quad \varepsilon = \tfrac{1}{2}(\nabla u + \nabla u^\top) \quad & \text{in } \Omega \\ K(\rho) : \varepsilon \mathbf{n} = T \quad & \text{on } \Gamma_t \\ u = 0 \quad & \text{on } \Gamma_d \end{aligned}$$

Challenges: Objective function is **expensive** and highly **nonconvex** due to material models like the **S**olid **I**sotropic **M**aterial with **P**enalization (SIMP).

Motivating Applications

Sparse Control:

1. **Subgradient and Bundle Methods:** **Convergence is generally slow** (e.g., sublinear).

Motivating Applications

Sparse Control:

1. **Subgradient and Bundle Methods:** **Convergence is generally slow** (e.g., sublinear).
2. **Proximal Gradient Methods:** Robust yet **slow convergence** on highly nonconvex problems.

Motivating Applications

Sparse Control:

1. **Subgradient and Bundle Methods:** **Convergence is generally slow** (e.g., sublinear).
2. **Proximal Gradient Methods:** Robust yet **slow convergence** on highly nonconvex problems.
3. **Proximal Newton Methods:** Rapid convergence, but require **nonstandard prox computations**.

Motivating Applications

Sparse Control:

1. **Subgradient and Bundle Methods:** **Convergence is generally slow** (e.g., sublinear).
2. **Proximal Gradient Methods:** Robust yet **slow convergence** on highly nonconvex problems.
3. **Proximal Newton Methods:** Rapid convergence, but require **nonstandard prox computations**.

Topology Optimization:

1. **Optimality Criterion Method:** **Heuristic** fixed-point iteration that **converges slowly**.

Motivating Applications

Sparse Control:

1. **Subgradient and Bundle Methods: Convergence is generally slow** (e.g., sublinear).
2. **Proximal Gradient Methods:** Robust yet **slow convergence** on highly nonconvex problems.
3. **Proximal Newton Methods:** Rapid convergence, but require **nonstandard prox computations**.

Topology Optimization:

1. **Optimality Criterion Method: Heuristic** fixed-point iteration that **converges slowly**.
2. **Method of Moving Asymptotes:** Seq. rational opt. approach is inherently **finite dimensional**.

Motivating Applications

Sparse Control:

1. **Subgradient and Bundle Methods:** **Convergence is generally slow** (e.g., sublinear).
2. **Proximal Gradient Methods:** Robust yet **slow convergence** on highly nonconvex problems.
3. **Proximal Newton Methods:** Rapid convergence, but require **nonstandard prox computations**.

Topology Optimization:

1. **Optimality Criterion Method:** **Heuristic** fixed-point iteration that **converges slowly**.
2. **Method of Moving Asymptotes:** Seq. rational opt. approach is inherently **finite dimensional**.
3. **Augmented Lagrangian:** Inner-outer iteration can be **overly expensive**.

Motivating Applications

Sparse Control:

1. **Subgradient and Bundle Methods:** **Convergence is generally slow** (e.g., sublinear).
2. **Proximal Gradient Methods:** Robust yet **slow convergence** on highly nonconvex problems.
3. **Proximal Newton Methods:** Rapid convergence, but require **nonstandard prox computations**.

Topology Optimization:

1. **Optimality Criterion Method:** **Heuristic** fixed-point iteration that **converges slowly**.
2. **Method of Moving Asymptotes:** Seq. rational opt. approach is inherently **finite dimensional**.
3. **Augmented Lagrangian:** Inner-outer iteration can be **overly expensive**.
4. **Interior Points:** Nonconvexity can lead to **expensive inertia correction**.

Motivating Applications

Sparse Control:

1. **Subgradient and Bundle Methods: Convergence is generally slow** (e.g., sublinear).
2. **Proximal Gradient Methods:** Robust yet **slow convergence** on highly nonconvex problems.
3. **Proximal Newton Methods:** Rapid convergence, but require **nonstandard prox computations**.

Topology Optimization:

1. **Optimality Criterion Method: Heuristic** fixed-point iteration that **converges slowly**.
2. **Method of Moving Asymptotes:** Seq. rational opt. approach is inherently **finite dimensional**.
3. **Augmented Lagrangian:** Inner-outer iteration can be **overly expensive**.
4. **Interior Points:** Nonconvexity can lead to **expensive inertia correction**.

It can be extremely difficult to incorporate inexactness in these methods!

Nonsmooth Trust Regions

Require: An initial guess x_1, initial trust-region radius $\Delta_1 > 0$, $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 \le \gamma_2 < 1$

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

  Model Selection: Choose a subproblem model f_k of f near x_k

  Step Computation: Compute x_k^+ that *approximately* solves

$$\min_{x \in H} \{m_k(x) := f_k(x) + \phi(x)\} \qquad \text{subject to} \qquad \|x - x_k\| \le \Delta_k$$

  Evaluate Objective: Compute the actual reduction $\text{ared}_k := F(x_k) - F(x_k^+)$

  if $\rho_k := \frac{\text{ared}_k}{m_k(x_k) - m_k(x_k^+)} < \eta_1$ **then**

    $x_{k+1} \leftarrow x_k$ and $\Delta_{k+1} \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k]$

  else

    $x_{k+1} \leftarrow x_k^+$

    if $\rho_k < \eta_2$ **then**

      $\Delta_{k+1} \in [\gamma_2 \Delta_k, \Delta_k]$

    else

      $\Delta_{k+1} \in [\Delta_k, \infty)$

    end if

  end if

end for

Nonsmooth Trust Regions

Subproblem

Trust-Region Subproblem: At each iteration, we approximately solve

$$\min_{x\in H} \{m_k(x) := f_k(x) + \phi(x)\} \qquad \text{subject to} \qquad \|x - x_k\| \leq \Delta_k,$$

where $\Delta_k > 0$ is the radius and $f_k : H \to \mathbb{R}$ is a model of the f near the iterate x_k.

Example: Perhaps the most common model f_k is the quadratic Taylor model

$$f_k(x) = (g_k, x - x_k) + \frac{1}{2}(B_k(x - x_k), x - x_k),$$

where $g_k \approx \nabla f(x_k)$ and B_k encapsulates curvature information, e.g., $B_k = \nabla^2 f(x_k)$ or an approximation thereof (e.g., quasi-Newton).

Nonsmooth Trust Regions

Approximate Subproblem Solution

Recall: TR methods use a ***Cauchy point*** to measure **sufficient decrease** of the trial iterate x_k^+.

Nonsmooth Trust Regions

Recall: TR methods use a ***Cauchy point*** to measure **sufficient decrease** of the trial iterate x_k^+.

We generalize the ***Cauchy point*** to nonsmooth problems using the ***proximal gradient path***

$$x_k^{\text{cp}} = p_k(t_k) \qquad \text{where} \qquad p_k(t) := \text{prox}_{t\phi}(x_k - tg_k),$$

where the ***proximity operator*** is given by

$$\text{prox}_{t\phi}(x) := \operatorname*{arg\,min}_{y \in H} \left\{ \tfrac{1}{2t} \|y - x\|^2 + \phi(y) \right\}.$$

Nonsmooth Trust Regions

Approximate Subproblem Solution

Recall: TR methods use a ***Cauchy point*** to measure **sufficient decrease** of the trial iterate x_k^+.

We generalize the ***Cauchy point*** to nonsmooth problems using the ***proximal gradient path***

$$x_k^{\mathrm{cp}} = p_k(t_k) \qquad \text{where} \qquad p_k(t) := \mathrm{prox}_{t\phi}(x_k - tg_k),$$

where the ***proximity operator*** is given by

$$\mathrm{prox}_{t\phi}(x) := \operatorname*{arg\,min}_{y\in H} \left\{ \tfrac{1}{2t}\|y-x\|^2 + \phi(y) \right\}.$$

We require that the step length t_k satisfies both

1. **Trust-Region Feasibility:** $\|x_k^{\mathrm{cp}} - x_k\| \le \nu_1 \Delta_k$
2. **Sufficient Decrease:** $m_k(x_k^{\mathrm{cp}}) - m_k(x_k) \le \mu_1[(g_k, x_k^{\mathrm{cp}} - x_k) + \phi(x_k^{\mathrm{cp}}) - \phi(x_k)]$

and at least one of the following conditions:

$$t_k \ge \nu_2 t_k' \qquad \text{or} \qquad t_k \ge \nu_3,$$

where t_k' satisfies

$$m_k(p_k(t_k')) - m_k(x_k) \ge \mu_2[(g_k, p_k(t_k') - x_k) + \phi(p_k(t_k')) - \phi(x_k)] \qquad \text{or} \qquad \|p_k(t_k') - x_k\| \ge \nu_4\Delta_k.$$

Nonsmooth Trust Regions

- **GCP Computation:** Can compute x_k^{cp} with **finitely many evals** of $p_k(t)$.

Nonsmooth Trust Regions

- **GCP Computation:** Can compute x_k^{cp} with **finitely many evals** of $p_k(t)$.
- **Consequence of GCP:** There exists a trial iterate x_k^+ that satisfies

$$\|x_k^+ - x_k\| \le \nu_{\text{rad}} \Delta_k, \quad \nu_{\text{rad}} \ge \nu_1$$
$$m_k(x_k) - m_k(x_k^+) \ge \mu_3 [m_k(x_k) - m_k(x_k^{\text{cp}})], \quad 0 < \mu_3 \le 1.$$

Nonsmooth Trust Regions

- **GCP Computation:** Can compute x_k^{cp} with **finitely many evals** of $p_k(t)$.
- **Consequence of GCP:** There exists a trial iterate x_k^+ that satisfies

$$\|x_k^+ - x_k\| \le \nu_{\text{rad}}\Delta_k, \quad \nu_{\text{rad}} \ge \nu_1$$
$$m_k(x_k) - m_k(x_k^+) \ge \mu_3[m_k(x_k) - m_k(x_k^{\text{cp}})], \quad 0 < \mu_3 \le 1.$$

- **Trial Iterate Requirements:** Avoid GCP computation by ensuring that x_k^+ satisfies

$$\begin{aligned} &\|x_k^+ - x_k\| \le \nu_{\text{rad}}\Delta_k \\ &m_k(x_k) - m_k(x_k^+) \ge \kappa_{\text{fcd}} h_k \min\left\{\frac{h_k}{1+\omega_k}, \Delta_k\right\}, \end{aligned} \tag{FCD}$$

where $h_k := \|p_k(r_0) - x_k\|/r_0$ for fixed $r_0 > 0$ and $\omega_k \ge 0$ measures the curvature of f_k.

Nonsmooth Trust Regions

Require: An initial guess x_1, initial trust-region radius $\Delta_1 > 0$, $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 \le \gamma_2 < 1$

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

  Model Selection: Choose a subproblem model f_k of f near x_k

  Step Computation: Compute x_k^+ that *approximately* solves

$$\min_{x \in H} \{ m_k(x) := f_k(x) + \phi(x) \} \qquad \text{subject to} \qquad \|x - x_k\| \le \Delta_k$$

  Evaluate Objective: Compute the actual reduction $\text{ared}_k := F(x_k) - F(x_k^+)$

  if $\rho_k := \frac{\text{ared}_k}{m_k(x_k) - m_k(x_k^+)} < \eta_1$ **then**

    $x_{k+1} \leftarrow x_k$ and $\Delta_{k+1} \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k]$

  else

    $x_{k+1} \leftarrow x_k^+$

    if $\rho_k < \eta_2$ **then**

      $\Delta_{k+1} \in [\gamma_2 \Delta_k, \Delta_k]$

    else

      $\Delta_{k+1} \in [\Delta_k, \infty)$

    end if

  end if

end for

Nonsmooth Trust Regions

Inexact Algorithm

Require: An initial guess x_1, initial trust-region radius $\Delta_1 > 0$, $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 \le \gamma_2 < 1$

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

  Model Selection: Choose a subproblem model f_k of f near x_k

  Step Computation: Compute a trial iterate x_k^+ that satisfies (FCD)

  Evaluate Objective: Compute the actual reduction $\mathrm{ared}_k := F(x_k) - F(x_k^+)$

  if $\rho_k := \frac{\mathrm{ared}_k}{m_k(x_k) - m_k(x_k^+)} < \eta_1$ **then**

    $x_{k+1} \leftarrow x_k$ and $\Delta_{k+1} \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k]$

  else

    $x_{k+1} \leftarrow x_k^+$

    if $\rho_k < \eta_2$ **then**

      $\Delta_{k+1} \in [\gamma_2 \Delta_k, \Delta_k]$

    else

      $\Delta_{k+1} \in [\Delta_k, \infty)$

    end if

  end if

end for

Nonsmooth Trust Regions

Inexact Algorithm

Require: An initial guess x_1, initial trust-region radius $\Delta_1 > 0$, $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 \le \gamma_2 < 1$

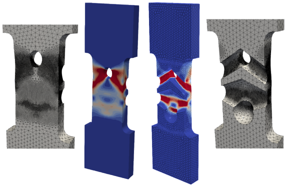
```
for k = 1, 2, ... do
    Model Selection: Choose a subproblem model f_k of f near x_k          Inexact!
    Step Computation: Compute a trial step x_k^+ that satisfies (FCD)
    Evaluate Objective: Evaluate the computed reduction cred_k ≈ ared_k    Inexact!
    if ρ_k := ared_k / (m_k(x_k) − m_k(x_k^+)) < η_1 then
        x_{k+1} ← x_k and Δ_{k+1} ∈ [γ_1 Δ_k, γ_2 Δ_k]
    else
        x_{k+1} ← x_k^+
        if ρ_k < η_2 then
            Δ_{k+1} ∈ [γ_2 Δ_k, Δ_k]
        else
            Δ_{k+1} ∈ [Δ_k, ∞)
        end if
    end if
end for
```

Nonsmooth Trust Regions

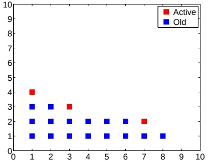
In infinite-dimensional optimization, the objective function and its gradient are often **impossible** to compute without discretization, iteration, etc., leading to **inexactness**.

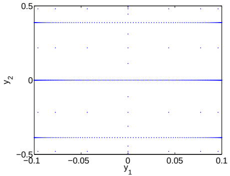
In infinite-dimensional optimization, the objective function and its gradient are often **impossible** to compute without discretization, iteration, etc., leading to **inexactness**.

Adaptive Finite Elements

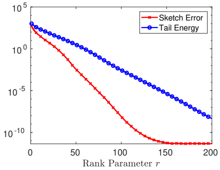


Adaptive Quadrature





Adaptive Compression



Nonsmooth Trust Regions

In infinite-dimensional optimization, the objective function and its gradient are often **impossible** to compute without discretization, iteration, etc., leading to **inexactness**.

When evaluating the of reduction of the objective function, we approximate

$$\mathrm{cred}_k \approx \mathrm{ared}_k := (f(x_k) + \phi(x_k)) - (f(x_{k+1}) - \phi(x_{k+1})),$$

where cred_k satsifies:

$$\exists\, \kappa_{\mathrm{obj}} > 0, \quad \zeta > 1, \quad \eta < \min\{\eta_1, 1 - \eta_2\}, \quad \text{and} \quad \theta_k \searrow 0 \quad \text{such that}$$
$$|\mathrm{ared}_k - \mathrm{cred}_k| \le \kappa_{\mathrm{obj}}[\eta \min\{m_k(x_k) - m_k(x_{k+1}), \theta_k\}]^{\zeta} \quad \forall\, k.$$

Nonsmooth Trust Regions

In infinite-dimensional optimization, the objective function and its gradient are often **impossible** to compute without discretization, iteration, etc., leading to **inexactness**.

When evaluating the of reduction of the objective function, we approximate

$$\mathrm{cred}_k \approx \mathrm{ared}_k := (f(x_k) + \phi(x_k)) - (f(x_{k+1}) - \phi(x_{k+1})),$$

where cred_k satsifies:

$$\exists\, \kappa_{\mathrm{obj}} > 0, \quad \zeta > 1, \quad \eta < \min\{\eta_1, 1 - \eta_2\}, \quad \text{and} \quad \theta_k \searrow 0 \quad \text{such that}$$
$$|\mathrm{ared}_k - \mathrm{cred}_k| \le \kappa_{\mathrm{obj}}[\eta \min\{m_k(x_k) - m_k(x_{k+1}), \theta_k\}]^{\zeta} \quad \forall\, k.$$

We also require that the model gradient g_k must satisfy:

$$\exists\, \kappa_{\mathrm{grad}} > 0 \qquad \text{such that} \qquad \|\nabla f(x_k) - g_k\| \le \kappa_{\mathrm{grad}} \min\{h_k, \Delta_k\} \quad \forall\, k.$$

Convergence Theory

Under the stated assumptions, the iterates produced by the TR algorithm satisfy

$$\liminf_{k\to\infty} h_k = 0 \qquad \Longrightarrow \qquad \liminf_{k\to\infty} h(x,t) = 0 \quad \forall t > 0,$$

where $\quad h_k := \frac{1}{r_0}\|\mathrm{prox}_{r_0\phi}(x_k - r_0 g_k) - x_k\| \quad$ and $\quad h(x,t) := \frac{1}{t}\|\mathrm{prox}_{t\phi}(x - t\nabla f(x)) - x\|.$

Finite Termination: $\quad \forall \tau > 0 \quad \exists K_\tau \in \mathbb{N}$ such that $\quad h_{K_\tau} \le \tau h_1.$

Tikhonov Regularization: If $f(x) = f_0(x) + \frac{\alpha}{2}\|x - x_0\|^2$, where $\alpha > 0$, $x_0 \in H$, ∇f_0 is **completely continuous** and $r_0 \ge \alpha^{-1}$, then any **weak accumulation point** of $\{x_k\}$ is a **critical point** of $f + \phi$. See, e.g., **sparse control**.

Convergence Theory

Under the stated assumptions, the iterates produced by the TR algorithm satisfy

$$\liminf_{k\to\infty} h_k = 0 \qquad \Longrightarrow \qquad \liminf_{k\to\infty} h(x,t) = 0 \quad \forall t > 0,$$

where $\quad h_k := \frac{1}{r_0}\|\mathrm{prox}_{r_0\phi}(x_k - r_0 g_k) - x_k\| \quad$ and $\quad h(x,t) := \frac{1}{t}\|\mathrm{prox}_{t\phi}(x - t\nabla f(x)) - x\|.$

Finite Termination: $\quad \forall \tau > 0 \quad \exists K_\tau \in \mathbb{N} \quad$ such that $\quad h_{K_\tau} \le \tau h_1.$

Strong Local Convergence: Suppose f is **strongly convex** on a convex set $U \subseteq H$ with $U \cap \mathrm{dom}\phi \neq \emptyset$ and $\exists K_0 \in \mathbb{N}$ such that $x_k \in U$ for $k \ge K_0$. If $\exists \bar{x} \in U$ satisfying $h(\bar{x},t) = 0 \; \forall t > 0$, then $x_k \to \bar{x}$. That is, $\{x_k\}$ **converges strongly to a critical point**.

Convergence Theory

Under the stated assumptions, the iterates produced by the TR algorithm satisfy

$$\liminf_{k\to\infty} h_k = 0 \quad \Longrightarrow \quad \liminf_{k\to\infty} h(x,t) = 0 \quad \forall t > 0,$$

where $\quad h_k := \frac{1}{r_0}\|\mathrm{prox}_{r_0\phi}(x_k - r_0 g_k) - x_k\|\quad$ and $\quad h(x,t) := \frac{1}{t}\|\mathrm{prox}_{t\phi}(x - t\nabla f(x)) - x\|.$

Finite Termination: $\quad \forall \tau > 0 \quad \exists K_\tau \in \mathbb{N}\quad$ such that $\quad h_{K_\tau} \le \tau h_1.$

Convergence Rates: Further, suppose f_k is a quadratic Taylor model and $\nabla^2 f$ is Lipschitz.

1. If $\tau_k \to 0$, then x_k converges ***superlinearly***.
2. If $\tau_k \le \tau h_k^{1+\alpha}$ for $\tau > 0$ and $\alpha \ge 0$, then x_k converges ***quadratically***.

Requires additional assumptions on subproblem solver, see Bobby Baraldi's talk (MS252).

For our numerical results, we compute trial iterates using **spectral proximal gradient**.

Baraldi & Kouri, A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations, Math. Prog., 2022.
Kouri, A matrix-free trust-region Newton algorithm for convex-constrained optimization, Opt. Letters, 2022.

Numerical Results

Goals: **1. Comparison of TR method with modern nonsmooth methods.**
2. Demonstration of mesh independence for TR method.

Let $\Omega = (0,1)^2$, $w \equiv -1$, $a \equiv -25$, $b \equiv 25$, $\alpha = 10^{-4}$ and $\beta = 10^{-2}$, and consider

$$\min_{z \in L^2(\Omega)} \int_\Omega |S(z) - w|^2(x)\,\mathrm{d}x + \frac{\alpha}{2}\int_\Omega |z|^2(x)\,\mathrm{d}x + \beta \int_\Omega |z|(x)\,\mathrm{d}x$$
$$\text{subject to} \qquad a \le z \le b \quad \text{a.e.},$$

where $S(z) = u \in H_0^1(\Omega)$ solves

$$-\Delta u + u^3 = z \quad \text{in } \Omega$$
$$u = 0 \quad \text{in } \partial\Omega$$

Discretization: P1 FEM for state variables and piecewise constant for controls.

Problem Size: 131,072 control degrees of freedom.

Numerical Results

Sparse Control

method	iter	fval	grad	hess	phi	prox	time (s)	TR speedup*
TR	4	5	5	39	57	142	22.88	1.0000
PG	59	149	60	0	149	209	498.56	21.79
SPG	30	46	31	0	46	62	168.26	7.35
R2	106	107	46	0	107	153	368.27	16.10
nmAPG	93	194	186	0	194	196	1018.66	44.52
iPiano	103	240	104	0	104	344	816.96	35.71
FISTA	141	430	283	0	430	290	1532.58	66.98
PANOC	83	285	108	0	272	287	948.04	41.44
ZeroFPR	21	70	43	0	45	93	247.39	10.81

Proximal Gradient Methods **Accelerated Methods** **Proximal Quasi-Newton Methods**

*TR speedup is the ratio of the wallclock time for TR divided by the times for the other methods.

Numerical Results

Sparse Control

τ_{op}	1e-4				1e-6				1e-8			
mesh	iter	npde	lpde	prox	iter	npde	lpde	prox	iter	npde	lpde	prox
64x64	3	4	56	80	5	6	108	129	7	8	186	181
128x128	3	4	54	79	4	5	79	102	6	7	129	151
256x256	3	4	56	80	5	6	108	129	6	7	133	153
512x512	3	4	54	78	5	6	102	123	6	7	127	147

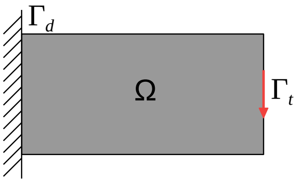
Trust-region algorithm demonstrates **mesh independence** with respect to the number of iterations and the number of PDE solves!

Requires only modest additional computational work to achieve tight tolerances!

Numerical Results

Elastic Topology Optimization

Goals: **1. Comparison of TR method with modern projected and AL methods.**
2. Demonstration of TR inexactness control for 3D problems.



Let $\Omega = (0,2)\times(0,1)^d$, $d = 1,2$, and $v = 0.4$, and consider

$$\min_{\rho\in L^2(\Omega)} \int_{\Gamma_t} T(x)\cdot[S(\rho)](x)\,\mathrm{d}x$$

$$\text{subject to} \quad \int_\Omega \rho(x)\,\mathrm{d}x = v|\Omega|, \quad 0 \le \rho \le 1 \;\text{ a.e.,}$$

where $S(\rho) = u \in (H^1(\Omega))^{d+1}$ solves

$$\begin{aligned} -\nabla\cdot(K(\rho):\varepsilon) &= 0 && \text{in } \Omega \\ \varepsilon &= \frac{1}{2}(\nabla u + \nabla u^\top) && \text{in } \Omega \\ K(\rho):\varepsilon\mathbf{n} &= T && \text{on } \Gamma_t \\ u &= 0 && \text{on } \Gamma_d \end{aligned}$$

Numerical Results

2D Elastic Topology Optimization

Formulation: SIMP power $p = 3$ with Helmholtz filtering (radius$=$ 0.1).

Discretization: Q1 FEM for displacement variables and piecewise constant for density.

Problem Size: 26,880 density degrees of freedom.

method	iter	fval	grad	hess	proj	time(s)	TR speedup*
TR	9	10	10	236	1200	16.49	1.0000
LMTR	33	34	31	418	391	32.42	1.9660
PQN	126	235	127	0	4972	164.49	9.9751
SPG	84	90	85	0	170	52.36	3.1753
AL-TR	9	52	51	1153	0	61.98	3.7586
AL-LMTR	11	276	263	4368	0	280.77	17.0267

Projected Newton-Type Methods **Spectral Projected Gradient** **AL Methods**

*TR speedup is the ratio of the wallclock time for TR divided by the times for the other methods.

Numerical Results

3D Elastic Topology Optimization

Formulation: SIMP power $p = 3$ with Helmholtz filtering (radius$=0.1$).

Discretization: Q1 FEM for displacement variables and piecewise constant for density.

Problem Size: 221,184 density degrees of freedom.

Inexact Solves: Solve using CG with AMG preconditioning.

- **Helmholtz Filter:** Requires $\sim$8 iterations to achieve the relative error of $\sim 10^{-12}$

  $\Longrightarrow$ Considered to be **exact**.

- **Elasticity Equations:** Trust-region algorithm controls accuracy of linear solver.

k	$F(x_k)$	h_k	$\|x_k - x_{k-1}\|$	Δ_k	fval	grad	hess	proj	obj tol	grad tol
0	1.0000	4.017e-2	---	1e1	1	1	0	3	1.000e-2	1.000e-2
1	0.8157	1.927e-2	1.000e1	1e2	2	2	12	44	1.000e-2	1.000e-2
2	0.4716	1.279e-2	5.420e1	1e3	3	3	25	75	1.000e-2	1.000e-2
3	0.4144	6.280e-3	1.260e1	1e4	4	4	39	103	4.632e-3	1.000e-2
4	0.1600	3.101e-3	1.990e2	1e4	5	5	52	132	1.000e-2	1.000e-2
5	0.1300	1.226e-3	1.085e2	1e5	6	6	65	161	2.970e-3	1.000e-2
6	0.1262	1.242e-5	6.044e1	1e6	7	7	78	190	3.539e-4	1.000e-2
7	0.1254	6.590e-6	5.821e1	1e7	8	8	91	220	6.971e-5	6.590e-3
8	0.1251	3.221e-6	3.599e1	1e8	9	9	104	249	1.942e-5	3.221e-3

Conclusions:

- **Numerical solution** of infinite-dimensional problems requires **expensive approximations**
- Often, the objective function and its gradient can only be computed **inexactly**
- Nonsmooth trust region is **provably convergent** even with **inexact computations**
- **We can efficiently compute a trial step using the spectral proximal gradient method**
- SPG trust-region subproblem solver is **matrix free**, but may **require** many prox computations
  Future: Can we incorporate inexact prox computations? Can we handle nonconvex ϕ?
- Nonsmooth trust-region method **outperforms** existing nonsmooth methods!

References:

- R. J. Baraldi & D. P. Kouri, Local Convergence Analysis of an Inexact Trust-Region Method for Nonsmooth Optimization, Submitted, 2023.
- R. J. Baraldi & D. P. Kouri, A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations, Math Programming, 2022.
- D. P. Kouri, A matrix-free trust-region Newton algorithm for convex-constrained optimization, Optimization Letters, 2022.
- D. P. Kouri & D. Ridzal, Inexact trust-region methods for PDE-constrained optimization, Frontiers in PDE-Constrained Optimization, 2018.