10116C

Sandia National Laboratories

An Inexact Trust-Region Algorithm for Nonsmooth Nonconvex Optimization

Drew P. Kouri, Robert J. Baraldi

7th International Conference on Continuous Optimization

Bethlehem, PA

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

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

SAND2022-7792 PE

Problem Formulation

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

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

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

Problem Formulation

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

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

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

Examples:

1. **Sparse Regularization:** $H = L^2(\Omega)$ and $\phi(x) = \beta\|x\|_{L^1(\Omega)}$ with $\beta > 0$.
2. **Convex Constraints:** $\phi(x) = 0$ if $x \in C$ and $\phi(x) = +\infty$ otherwise.

Problem Formulation

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

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

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

Examples:

1. **Sparse Regularization:** $H = L^2(\Omega)$ and $\phi(x) = \beta \|x\|_{L^1(\Omega)}$ with $\beta > 0$.
2. **Convex Constraints:** $\phi(x) = 0$ if $x \in C$ and $\phi(x) = +\infty$ otherwise.

Notation: $\mathrm{prox}_{\gamma\phi}(x)$ for $x \in H$ and $\gamma > 0$ is the **proximity operator** given by

$$\mathrm{prox}_{\gamma\phi}(x) := \arg\min_{y \in H} \left\{ \frac{1}{2\gamma} \|x - y\|_H^2 + \phi(y) \right\}.$$

In example 2, $\mathrm{prox}_{\gamma\phi}(x) = \mathrm{proj}_C(x)$ is the **metric projection** of x onto C.

Motivating Application

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 $a \leq 0 \leq b$ a.e., and penalty parameters $\alpha, \beta \geq 0$,

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

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

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

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

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

1. **Subgradient and Bundle Methods:** Iterates x_{k+1} solve the optimization problem

$$\min_{x\in H} \frac{t_k}{2}\|x - x_k\|_H^2 + \sup_{j\in I_k}\{f(y_j) + \phi(y_j) + (\nabla f(y_j) + \eta_j, x - y_j)_H\},$$

where $t_k \geq 0$ and $\eta_j \in \partial\phi(y_j)$. Typically, **convergence is slow** (e.g., sublinear).

Sparse Control

Existing Methods

1. **Subgradient and Bundle Methods:** Iterates x_{k+1} solve the optimization problem

$$\min_{x\in H} \frac{t_k}{2}\|x-x_k\|_H^2 + \sup_{j\in I_k}\{f(y_j)+\phi(y_j)+(\nabla f(y_j)+\eta_j, x-y_j)_H\},$$

   where $t_k \geq 0$ and $\eta_j \in \partial\phi(y_j)$. Typically, **convergence is slow** (e.g., sublinear).

2. **Proximal Gradient Methods:** Iterates x_{k+1} solve the optimization problem

$$\min_{x\in H} (\nabla f(x_k), x-x_k)_H + \frac{1}{2\gamma_k}\|x-x_k\|_H^2 + \phi(x) \quad \Longleftrightarrow \quad x_{k+1} = \mathrm{prox}_{\gamma_k\phi}(x_k - \gamma_k\nabla f(x_k)).$$

   PG methods are robust, but **slow**. Can use acceleration (Nesterov) or momentum (heavy balls).

Sparse Control

1. **Subgradient and Bundle Methods:** Iterates x_{k+1} solve the optimization problem

$$\min_{x\in H}\ \frac{t_k}{2}\|x-x_k\|_H^2+\sup_{j\in I_k}\{f(y_j)+\phi(y_j)+(\nabla f(y_j)+\eta_j,x-y_j)_H\},$$

where $t_k\geq 0$ and $\eta_j\in\partial\phi(y_j)$. Typically, **convergence is slow** (e.g., sublinear).

2. **Proximal Gradient Methods:** Iterates x_{k+1} solve the optimization problem

$$\min_{x\in H}\ (\nabla f(x_k),x-x_k)_H+\frac{1}{2\gamma_k}\|x-x_k\|_H^2+\phi(x)\quad\iff\quad x_{k+1}=\mathrm{prox}_{\gamma_k\phi}(x_k-\gamma_k\nabla f(x_k)).$$

PG methods are robust, but **slow**. Can use acceleration (Nesterov) or momentum (heavy balls).

3. **Proximal Newton-Type Methods:** Iterates x_{k+1} solve the optimization problem

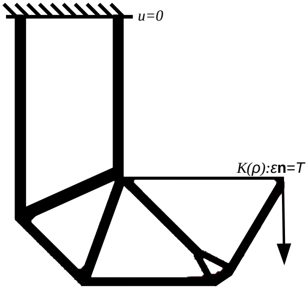
$$\min_{x\in H}\ (\nabla f(x_k),x-x_k)_H+\frac{1}{2}(B_k(x-x_k),x-x_k)_H+\phi(x),$$

where $B_k\in L(X)$ approximates the Hessian of f. PN methods require **positive definite B_k** (e.g., convexity) and **nonstandard/nontrivial prox computations**.

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 $v \in (0,1)$,

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

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

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).

Elastic Topology Optimization

1. **Optimality Criterion Method:** A **heuristic** fixed-point iteration that is related to a projected gradient method.

   Bendsøe & Kikuchi, Generating optimal topologies in structural design using a homogenization method, CMAME, 1988.

Elastic Topology Optimization

1. **Optimality Criterion Method:** A **heuristic** fixed-point iteration that is related to a projected gradient method.

   Bendsøe & Kikuchi, Generating optimal topologies in structural design using a homogenization method, CMAME, 1988.

2. **Method of Moving Asymptotes:** A sequential convex optimization approach that uses rational approximations of the objective and constraints. The dual subproblem is commonly solved using nonlinear CG. This method is inherently **finite dimensional**.

   Svanberg, The method of moving asymptotes—A new method for structural optimization, IJNME, 1987.

Elastic Topology Optimization

1. **Optimality Criterion Method:** A **heuristic** fixed-point iteration that is related to a projected gradient method.

   Bendsøe & Kikuchi, Generating optimal topologies in structural design using a homogenization method, CMAME, 1988.

2. **Method of Moving Asymptotes:** A sequential convex optimization approach that uses rational approximations of the objective and constraints. The dual subproblem is commonly solved using nonlinear CG. This method is inherently **finite dimensional**.

   Svanberg, The method of moving asymptotes—A new method for structural optimization, IJNME, 1987.

3. **Augmented Lagrangian:** Robust, yet minimizing the penalty function at each iteration can be expensive.

Elastic Topology Optimization

1. **Optimality Criterion Method:** A **heuristic** fixed-point iteration that is related to a projected gradient method.

   Bendsøe & Kikuchi, Generating optimal topologies in structural design using a homogenization method, CMAME, 1988.

2. **Method of Moving Asymptotes:** A sequential convex optimization approach that uses rational approximations of the objective and constraints. The dual subproblem is commonly solved using nonlinear CG. This method is inherently **finite dimensional**.

   Svanberg, The method of moving asymptotes—A new method for structural optimization, IJNME, 1987.

3. **Augmented Lagrangian:** Robust, yet minimizing the penalty function at each iteration can be expensive.

4. **Interior Points:** Primal-dual line-search methods have been used succesfully. However, nonconvexity can lead to expensive inertia correction.

Elastic Topology Optimization

1. **Optimality Criterion Method:** A **heuristic** fixed-point iteration that is related to a projected gradient method.

   Bendsøe & Kikuchi, Generating optimal topologies in structural design using a homogenization method, CMAME, 1988.

2. **Method of Moving Asymptotes:** A sequential convex optimization approach that uses rational approximations of the objective and constraints. The dual subproblem is commonly solved using nonlinear CG. This method is inherently **finite dimensional**.

   Svanberg, The method of moving asymptotes—A new method for structural optimization, IJNME, 1987.

3. **Augmented Lagrangian:** Robust, yet minimizing the penalty function at each iteration can be expensive.

4. **Interior Points:** Primal-dual line-search methods have been used succesfully. However, 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_0, initial trust-region radius $\Delta_0 > 0$, $0 < \eta_1 < \eta_2 < 1$ 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+1} that *approximately* solves

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

Evaluate Objective: Compute the actual reduction $\text{ared}_k := F(x_k) - F(x_{k+1})$

Step Acceptance: Compute the ratio of actual and predicted reduction:

$$\rho_k := \frac{\text{ared}_k}{m_k(x_k) - m_k(x_{k+1})} < \eta_1 \qquad \Longrightarrow \qquad x_{k+1} \leftarrow x_k$$

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

end for

Nonsmooth Trust Regions

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

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

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

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

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

Recall: The ***Cauchy point*** is used to determine if iterate x_{k+1} has produced **sufficient reduction** of the model m_k — **Need a generalization for nonsmooth problems!**

Nonsmooth Trust Regions

Recall: The ***Cauchy point*** is used to determine if iterate x_{k+1} has produced **sufficient reduction** of the model m_k — **Need a generalization for nonsmooth problems!**

The ***generalized Cauchy point*** is a point along the proximal gradient path

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

Nonsmooth Trust Regions

Recall: The ***Cauchy point*** is used to determine if iterate x_{k+1} has produced **sufficient reduction** of the model m_k — **Need a generalization for nonsmooth problems!**

The ***generalized Cauchy point*** is a point along the proximal gradient path

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

that satisfies both

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

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

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

Nonsmooth Trust Regions

Require: An initial step length $\gamma > 0$ and positive parameters $0 < \beta_{\text{dec}} < 1 < \beta_{\text{inc}}$

```
if k = 1 then
    Set t̄ = γ
else
    Set t̄ = t_{k−1}
end if
if (1) and (2) are satisfied at t_k = t̄ then
    Compute the largest ℓ ∈ ℕ such that t_k = t̄ β_inc^ℓ satisfies (1) and (2)
else
    Compute the smallest ℓ ∈ ℕ such that t_k = t̄ β_dec^ℓ satisfies (1) and (2)
end if
```

$$\|p_k(t_k) - x_k\|_H \le \nu_1 \Delta_k \tag{1}$$

$$m_k(p_k(t_k)) - m_k(x_k) \le \mu_1[(g_k, p_k(t_k) - x_k)_H + \phi(p_k(t_k)) - \phi(x_k)] \tag{2}$$

Nonsmooth Trust Regions

Consequence of GCP: There exists an iterate x_{k+1} that satisfies

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

Proof: Take $x_{k+1} = x_k^{\text{cp}}$, computed using previous algorithm.

Nonsmooth Trust Regions

Consequence of GCP: There exists an iterate x_{k+1} that satisfies

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

Proof: Take $x_{k+1} = x_k^{\text{cp}}$, computed using previous algorithm.

The GCP computation requires **repeated** evaluation of the proximity operator!

Can avoid GCP computation by computing steps that satisfy

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

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

Nonsmooth Trust Regions

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

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

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

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

where cred_k satsifies:

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

We also require that the model gradient g_k must satisfy:

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

Nonsmooth Trust Regions

Require: An initial guess x_0, initial trust-region radius $\Delta_0 > 0$, $0 < \eta_1 < \eta_2 < 1$ 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 . . . . **Inexact!**

Step Computation: Compute a trial step x_{k+1} that satisfies (FCD)

Evaluate Objective: Evaluate the computed reduction $\mathrm{cred}_k \approx \mathrm{ared}_k$ . . . . **Inexact!**

Step Acceptance: Compute the ratio of computed and predicted reduction:

$$\rho_k := \frac{\mathrm{cred}_k}{m_k(x_k) - m_k(x_{k+1})} < \eta_1 \qquad \Longrightarrow \qquad x_{k+1} \leftarrow x_k$$

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

end for

Convergence Theory

Recall: $h_k := \frac{1}{r_0}\|\operatorname{prox}_{r_0\phi}(x_k - r_0 g_k) - x_k\|_H$

> Under the stated assumptions, the iterates produced by the TR algorithm satisfy
>
> $$\liminf_{k\to\infty} h_k = 0 \qquad \Longrightarrow \qquad \liminf_{k\to\infty} \frac{1}{r_0}\|\operatorname{prox}_{r_0\phi}(x_k - r_0\nabla f(x_k)) - x_k\|_H = 0.$$

Note: This result permits unbounded model curvature.

Convergence Theory

Recall: $h_k := \frac{1}{r_0}\|\operatorname{prox}_{r_0\phi}(x_k - r_0 g_k) - x_k\|_H$

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

$$\liminf_{k\to\infty} h_k = 0 \qquad \Longrightarrow \qquad \liminf_{k\to\infty} \frac{1}{r_0}\|\operatorname{prox}_{r_0\phi}(x_k - r_0\nabla f(x_k)) - x_k\|_H = 0.$$

Note: This result permits unbounded model curvature.

Application: If the smooth objective function f has the form

$$f(x) = f_0(x) + \frac{\alpha}{2}\|x - x_0\|_H^2,$$

where $\alpha > 0$, $x_0 \in H$, ∇f_0 is **completely continuous** and $r_0 \geq \alpha^{-1}$, then any **weak accumulation point** of $\{x_k\}$ is a **critical point** of $f + \phi$. See, e.g., **sparse control**.

Recall: ∇f_0 is **completely continuous** if $y_k \rightharpoonup y$ implies $\nabla f_0(y_k) \to \nabla f_0(y)$.

Spectral Proximal Gradient Subproblem Solver

Model: For the SPG subproblem solver, we employ the models

$$f_k(x) = \tfrac{1}{2}(B_k(x-x_k), x-x_k)_H + (g_k, x-x_k)_H \quad \text{and} \quad \phi_k(x) = \begin{cases} \phi(x) \text{ if } \|x-x_k\|_H \le \Delta_k \\ +\infty \text{ otherwise} \end{cases}$$

SPG Iteration: $x_{k,\ell+1} = x_{k,\ell} + \alpha_\ell s_\ell$ where $s_\ell = \mathrm{prox}_{\lambda_\ell \phi_k}(x_{k,\ell} - \lambda_\ell \nabla f_k(x_{k,\ell})) - x_{k,\ell}$

1. Start with $x_{k,0} = x_k^{\mathrm{cp}}$ to ensure fraction of Cauchy decrease (FCD)
2. Compute the step length α_ℓ by minimizing the quadratic upper bound

$$t \mapsto f_k(x_{k,\ell} + ts_\ell) + t[\phi_k(x_{k,\ell} + s_\ell) - \phi_k(x_{k,\ell})] + \phi_k(x_{k,\ell})$$

3. Compute the safeguarded spectral step length λ_ℓ as

$$\lambda_\ell := \max\left\{\lambda_{\min}, \min\left\{\lambda_{\max}, \frac{(s_{\ell-1}, s_{\ell-1})}{(B_k s_{\ell-1}, s_{\ell-1})}\right\}\right\}$$

Birgin, et al., Nonmonotone spectral projected gradient methods on convex sets, SIOPT, 2000.

Kouri, A matrix-free trust-region Newton algorithm for convex-constrained optimization, Opt. Letters, 2022.

Spectral Proximal Gradient Subproblem Solver

Proximity Operator for ϕ_k: The proximity operator of ϕ_k applied to $x \in H$ is given by

$$\operatorname{prox}_{\gamma\phi_k}(x) = \begin{cases} \operatorname{prox}_{\gamma\phi}(x) & \text{if } \|\operatorname{prox}_{\gamma\phi}(x) - x_k\|_H \le \Delta_k \\ \operatorname{prox}_{t^\star\gamma\phi}(x_k + t^\star(x - x_k)) & \text{otherwise} \end{cases},$$

where $t^\star \in [0,1]$ is any $t \in [0,1]$ that satisfies

$$\psi_k(t) := \|\operatorname{prox}_{t\gamma\phi}(x_k + t(x - x_k)) - x_k\|_H - \Delta_k = 0.$$

Here, ψ_k is nondecreasing and continuous on $[0,1]$ with $\psi_k(0) < 0$ and $\psi_k(1) > 0$.

Can compute $\operatorname{prox}_{\gamma\phi_k}(x)$ by applying, e.g., Brent's method to $\psi_k(t)$.

Numerical Results

Sparse Control

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

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

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)	speedup*
TR	3	4	4	26	36	80	17.2037	1.0000
PG	32	92	33	0	92	125	267.5140	15.5498
SPG	21	31	22	0	31	44	100.3323	5.8320
R2	60	61	25	0	61	86	180.1468	10.4714
nmAPG	43	86	86	0	86	88	409.1166	23.7807
iPiano	60	154	61	0	61	215	477.4582	27.7532
FISTA	54	169	109	0	169	116	542.4645	31.5319
PANOC	95	381	178	0	368	383	1151.7171	66.9459
ZeroFPR	44	139	89	0	92	185	437.2350	25.4152

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

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

Numerical Results

Sparse Control

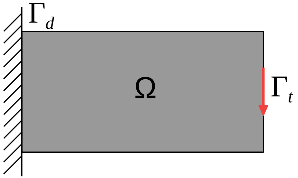
τ_{op}	1e-4				1e-6				1e-8			
mesh	iter	npde	lpde	prox	iter	npde	lpde	prox	iter	npde	lpde	prox
64x64	3	4	56	80	5	6	108	129	7	8	186	181
128x128	3	4	54	79	4	5	79	102	6	7	129	151
256x256	3	4	56	80	5	6	108	129	6	7	133	153
512x512	3	4	54	78	5	6	102	123	6	7	127	147

Algorithm demonstrates **mesh independences** with respect to the number of iterations and the number of PDE solves!

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

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)	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**

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

Numerical Results

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**.

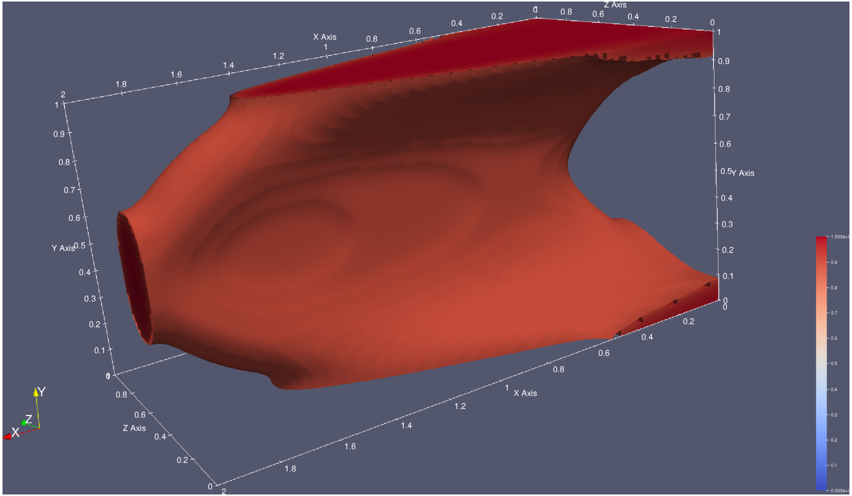
- **Elasticity Equations:** Trust-region algorithm controls accuracy of linear solver.

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	---	20	1	1	0	4	1.000e-2	1.000e-2
1	0.7156	1.771e-2	2.000e1	50	2	2	28	96	1.000e-2	1.000e-2
2	0.4393	6.788e-3	5.000e1	50	3	3	55	204	1.000e-2	1.000e-2
3	0.3168	2.853e-3	5.000e1	125	4	4	82	405	1.000e-2	1.000e-2
4	0.1654	8.805e-4	1.250e2	125	5	5	109	639	1.000e-2	8.802e-3
5	0.1255	2.066e-5	1.250e2	125	6	6	143	707	1.000e-2	2.066e-4
6	0.1247	2.713e-6	6.272e1	312.5	7	7	171	765	1.461e-4	2.713e-5

Numerical Results

Elastic Topology Optimization

Filtered Density: 0.9 Countour



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, A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations, submitted, 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.