SAND2025-02697C

An Inexact Trust-Region Algorithm for Nonsmooth Risk-Averse Optimization

Drew P. Kouri

SIAM Conference on Computational Science and Engineering
Fort Worth, Texas

March 6, 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.

Problem Formulation

Goal: Develop an efficient algorithm to solve the **risk-averse optimization problem**,

$$\min_{x\in X}\ f(x)+\mathcal{R}(F(x))+\phi(x).$$

- X and Y are Hilbert spaces and $\mathcal{R}:Y\to\mathbb{R}$ is a coherent risk measure;
- $\phi:X\to(-\infty,+\infty]$ is proper, closed and convex, but may be **nonsmooth**;
- $f:X\to\mathbb{R}$ has Lipschitz continuous gradients on an open set containing dom ϕ;
- $F:X\to Y$ has Lipschitz continuous Jacobians on an open set containing dom ϕ;
- $J:=f+\mathcal{R}\circ F+\phi$ is bounded below.

Problem Formulation

Goal: Develop an efficient algorithm to solve the **risk-averse optimization problem**,

$$\boxed{\min_{x\in X}\sup_{\theta\in\mathfrak{A}} f(x) + (\theta, F(x))_Y + \phi(x).}$$

- X and Y are Hilbert spaces and $\mathfrak{A} \subset Y$ is nonempty, closed, convex and bounded;
- $\phi : X \to (-\infty, +\infty]$ is proper, closed and convex, but may be **nonsmooth**;
- $f : X \to \mathbb{R}$ has Lipschitz continuous gradients on an open set containing dom ϕ;
- $F : X \to Y$ has Lipschitz continuous Jacobians on an open set containing dom ϕ;
- $J := f + \sigma_{\mathfrak{A}} \circ F + \phi$ is bounded below—$\sigma_{\mathfrak{A}}(y) := \sup_{\theta\in\mathfrak{A}}(\theta, y)_Y$ is the *support function*.

Problem Formulation

Goal: Develop an efficient algorithm to solve the **risk-averse optimization problem**,

$$\min_{x\in X}\sup_{\theta\in\mathfrak{A}}\; f(x)+(\theta,F(x))_Y+\phi(x).$$

- X and Y are Hilbert spaces and $\mathfrak{A}\subset Y$ is nonempty, closed, convex and bounded;
- $\phi : X\to(-\infty,+\infty]$ is proper, closed and convex, but may be **nonsmooth**;
- $f : X\to\mathbb{R}$ has Lipschitz continuous gradients on an open set containing dom ϕ;
- $F : X\to Y$ has Lipschitz continuous Jacobians on an open set containing dom ϕ;
- $J := f+\sigma_{\mathfrak{A}}\circ F+\phi$ is bounded below—$\sigma_{\mathfrak{A}}(y) := \sup_{\theta\in\mathfrak{A}}(\theta,y)_Y$ is the *support function*.

Examples

1. **Risk Averse/Distributionally Robust:** $Y=L^2(\Omega,\mathcal{F},\mathbb{P})$ and $\mathfrak{A}$ is a subset of pdfs.
2. **Convex Constraints:** ϕ is the indicator function of a nonempty, closed, convex set.
3. **Sparse Regularization:** ϕ is the L^1-norm or other nonsmooth regularizer.

Problem Formulation

Goal: Develop an efficient algorithm to solve the **risk-averse optimization problem**,

$$\min_{x\in X}\sup_{\theta\in\mathfrak{A}}\; f(x) + (\theta, F(x))_Y + \phi(x).$$

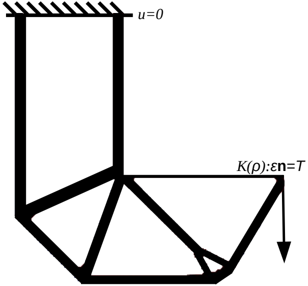
- X and Y are Hilbert spaces and $\mathfrak{A} \subset Y$ is nonempty, closed, convex and bounded;
- $\phi : X \to (-\infty, +\infty]$ is proper, closed and convex, but may be **nonsmooth**;
- $f : X \to \mathbb{R}$ has Lipschitz continuous gradients on an open set containing dom ϕ;
- $F : X \to Y$ has Lipschitz continuous Jacobians on an open set containing dom ϕ;
- $J := f + \sigma_{\mathfrak{A}} \circ F + \phi$ is bounded below—$\sigma_{\mathfrak{A}}(y) := \sup_{\theta\in\mathfrak{A}}(\theta, y)_Y$ is the *support function*.

$$\begin{aligned} \sigma_{\mathfrak{A}}(y) = \mathcal{R}(y) &= (1-\lambda)\mathbb{E}[y] + \lambda\,\mathrm{AVaR}_p(y), \qquad 0 \le \lambda,\ p \le 1 \\ \mathrm{AVaR}_p(y) &= \min_t\{t + \tfrac{1}{1-p}\mathbb{E}[y-t]_+\} \\ \mathfrak{A} &= \{\theta \,|\, \mathbb{E}[\theta] = 1,\ \ (1-\lambda) \le \theta \le (1-\lambda) + \tfrac{\lambda}{1-p}\ \text{ a.s.}\}. \end{aligned}$$

Motivating Application

Elastic Topology Optimization

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



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

$$\min_{\rho \in L^2(D)} \mathcal{R}\left(\int_{\Gamma_t} T_{\xi}(x) \cdot [S(\rho)](x)\,\mathrm{d}x\right)$$

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

where $S(\rho) = u : \Xi \to (H^1(D))^d$ solves the weak form of

$$\begin{aligned} -\nabla \cdot (K(\rho) : \varepsilon) = 0, \quad \varepsilon = \tfrac{1}{2}(\nabla u + \nabla u^\top) \qquad & \text{in } D, \text{ a.s.} \\ K(\rho) : \varepsilon \mathbf{n} = T_{\xi} \qquad & \text{on } \Gamma_t, \text{ a.s.} \\ u = 0 \qquad & \text{on } \Gamma_d, \text{ a.s.} \end{aligned}$$

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

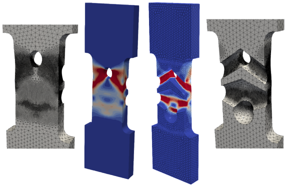
Key Algorithm Requirements

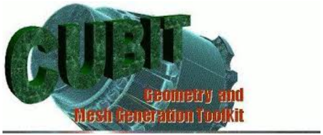
1. **Large-Scale Problems:** Rapid convergence, mesh independence, and matrix free.

Key Algorithm Requirements

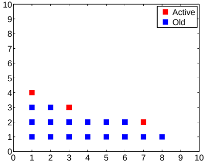
1. **Large-Scale Problems:** Rapid convergence, mesh independence, and matrix free.
2. **Leverage Inexactness:** Converges even when f, ∇f, F, and F' are computed inexactly via adaptive discretization, reduced-order modelling, compression, etc.

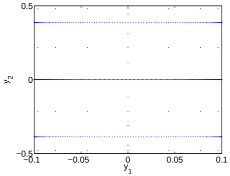
Adaptive Finite Elements



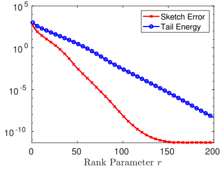


Adaptive Quadrature





Adaptive Compression



Nonsmooth Trust Regions

Basic Algorithm

```
Input: An initial guess $x_1$, initial radius $\Delta_1 > 0$, $\tau_1$, $\alpha$, $\beta > 0$, $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 \le \gamma_2 < 1$
for $k = 1, 2, \ldots$ do
    Model Selection: Choose a smooth local models $q_k$ of $f$ and $\ell_k$ of $F$ around $x_k$
    Step Computation: Compute $x_k^+$ that approximately solves
        $\min_{x \in X} \{m_k(x) := q_k(x) + \sigma_{\mathfrak{A}}(\ell_k(x)) + \phi(x)\}$   subject to   $\|x - x_k\| \le \Delta_k$
    Evaluate Objective: Compute the actual reduction $\mathrm{ared}_k \approx J(x_k) - J(x_k^+)$
    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

Subproblem

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

$$\min_{x\in X} \{m_k(x) := q_k(x) + \underbrace{\sigma_{\mathfrak{A}}(\ell_k(x)) + \phi(x)}_{=:\psi_k(x)}\} \qquad \text{subject to} \qquad \|x - x_k\| \le \Delta_k,$$

where $\Delta_k > 0$ is the trust-region radius.

Nonsmooth Trust Regions

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

$$\min_{x\in X} \{m_k(x) := q_k(x) + \underbrace{\sigma_{\mathfrak{A}}(\ell_k(x)) + \phi(x)}_{=:\psi_k(x)}\} \qquad \text{subject to} \qquad \|x - x_k\| \le \Delta_k,$$

where $\Delta_k > 0$ is the trust-region radius.

Trust-Region Models: We choose q_k and ℓ_k based on quadratic models of the Lagrangian

$$L(x,\theta) := f(x) + (\theta, F(x))_Y$$

akin to *sequential quadratic programming*. Given the k-th iterate (x_k, θ_k), we choose

$$q_k(x) := \tfrac{1}{2}(B_k(x-x_k), x-x_k)_X + (g_k, x-x_k)_X + q_k(0) \quad \text{and} \quad \ell_k(x) := A_k(x-x_k) + b_k,$$

where $B_k \approx L_{xx}(x_k, \theta_k)$, $g_k \approx \nabla f(x_k)$, $A_k \approx F'(x_k)$ and $b_k \approx F(x_k)$.

Nonsmooth Trust Regions

Subproblem

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

$$\min_{x\in X} \{m_k(x) := q_k(x) + \underbrace{\sigma_{\mathfrak{A}}(\ell_k(x)) + \phi(x)}_{=:\psi_k(x)}\} \qquad \text{subject to} \qquad \|x - x_k\| \le \Delta_k,$$

where $\Delta_k > 0$ is the trust-region radius.

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

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

where $P_{\mathfrak{A}}$ denotes the projection onto $\mathfrak{A}$ and the ***proximity operator*** is given by

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

Baraldi & Kouri, A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations, Math. Prog., 2022.

Nonsmooth Trust Regions

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

$$\min_{x\in X} \{m_k(x) := q_k(x) + \underbrace{\sigma_{\mathfrak{A}}(\ell_k(x)) + \phi(x)}_{=:\psi_k(x)}\} \qquad \text{subject to} \qquad \|x - x_k\| \le \Delta_k,$$

where $\Delta_k > 0$ is the trust-region radius.

How do we evaluate $\mathrm{prox}_{t\psi_k}$? In general, **no analytical expression** for $\mathrm{prox}_{t\psi_k}$ exists! Solve the dual problem

$$\max_{\theta\in\mathfrak{A}} \underbrace{\min_{v\in X}\{\tfrac{1}{2t}\|v - x\|^2 + (\theta, \ell_k(v))_Y + \phi(v)\}}_{=:d_k(\theta)} \quad\Longrightarrow\quad v_k(\theta) := \mathrm{prox}_{t\phi}(x - tA_k^*\theta).$$

1. If $\dim Y = 1$, maximize d_k using Golden Section, Brent's, etc.
2. Otherwise, maximize d_k using spectral projected gradient, i.e., $\nabla d_k(\theta) = \ell_k(v_k(\theta))$.

Nonsmooth Trust Regions

Generalized Cauchy Point

We set $x_k^{\text{cp}} = p_k(t_k)$, where the step length t_k satisfies both

1. **Trust-Region Feasibility:**

$$\|x_k^{\text{cp}} - x_k\| \le \nu_1 \Delta_k$$

2. **Sufficient Decrease:**

$$m_k(x_k^{\text{cp}}) - m_k(x_k) \le \mu_1[(g_k, x_k^{\text{cp}} - x_k) + \psi_k(x_k^{\text{cp}}) - \psi_k(x_k)]$$

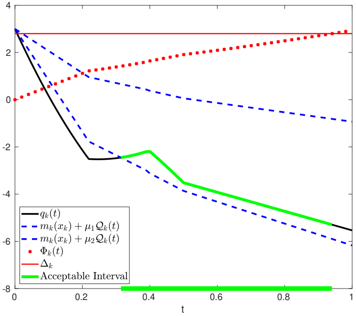
and at least one of the following conditions:

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

where t_k' either satisfies

$$m_k(p_k(t_k')) - m_k(x_k) \ge \mu_2[(g_k, p_k(t_k') - x_k) + \psi_k(p_k(t_k')) - \psi_k(x_k)]$$

$$\text{or} \qquad \|p_k(t_k') - x_k\| \ge \nu_4 \Delta_k .$$



The interval of **acceptable** t_k is depicted by the **green** line on the horizontal axis.

Baraldi & Kouri, A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations, Math. Prog., 2022.

Nonsmooth Trust Regions

Generalized Cauchy Point

We set $x_k^{\text{cp}} = p_k(t_k)$, where the step length t_k satisfies both

1. **Trust-Region Feasibility:**

$$\|x_k^{\text{cp}} - x_k\| \le \nu_1 \Delta_k$$

2. **Sufficient Decrease:**

$$m_k(x_k^{\text{cp}}) - m_k(x_k) \le \mu_1[(g_k, x_k^{\text{cp}} - x_k) + \psi_k(x_k^{\text{cp}}) - \psi_k(x_k)]$$

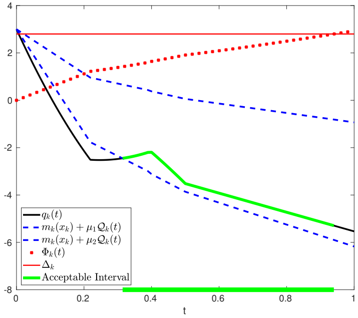
and at least one of the following conditions:

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

where t_k' either satisfies

$$m_k(p_k(t_k')) - m_k(x_k) \ge \mu_2[(g_k, p_k(t_k') - x_k) + \psi_k(p_k(t_k')) - \psi_k(x_k)]$$

$$\text{or} \qquad \|p_k(t_k') - x_k\| \ge \nu_4 \Delta_k.$$



The interval of **acceptable** t_k is depicted by the **green** line on the horizontal axis.

We can compute x_k^{cp} with **finitely many evaluations** of the proximity operator using a backtracking or bidirectional proximal search.

Baraldi & Kouri, A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations, Math. Prog., 2022.

Nonsmooth Trust Regions

Generalized Cauchy Point

- Computing the Cauchy point x_k^{cp} requires **finitely many evaluations** of $p_k(t)$.

---

[1] Baraldi & Kouri, Efficient proximal subproblem solvers for a nonsmooth trust-region method, COAP, 2025.

Nonsmooth Trust Regions

Generalized Cauchy Point

- Computing the Cauchy point x_k^{cp} requires **finitely many evaluations** of $p_k(t)$.
- We compute a trial iterate x_k^+ that produces at least a **fraction of Cauchy decrease**, i.e.,

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

using spectral projected gradient, truncated nonlinear CG, etc. subproblem solvers.[1]

---

[1] Baraldi & Kouri, Efficient proximal subproblem solvers for a nonsmooth trust-region method, COAP, 2025.

Nonsmooth Trust Regions

Generalized Cauchy Point

- Computing the Cauchy point x_k^{cp} requires **finitely many evaluations** of $p_k(t)$.
- We compute a trial iterate x_k^+ that produces at least a **fraction of Cauchy decrease**, i.e.,

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

  using spectral projected gradient, truncated nonlinear CG, etc. subproblem solvers.[1]
- The trial iterate x_k^+ satisfies the so-called **fraction of Cauchy decrease condition**

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

  where $h_k := \|p_k(t_k) - x_k\|/t_k$ for $t_k \in [t_{\min}, t_{\max}]$.

---

[1] Baraldi & Kouri, Efficient proximal subproblem solvers for a nonsmooth trust-region method, COAP, 2025.

Nonsmooth Trust Regions

Subproblem Model Selection

For the subproblem model, we employ quadratic and linear approximations of f and F, respectively, i.e.,

$$q_k(x) := \frac{1}{2}(B_k(x - x_k), x - x_k)_X + (g_k, x - x_k)_X + q_k(0) \qquad \text{and} \qquad \ell_k(x) := A_k(x - x_k) + b_k,$$

and require that the model components b_k, g_k and A_k satisfy:

$$\begin{aligned} &\exists\, \kappa_{\text{grad}} > 0 \quad \text{such that} \quad \|\nabla f(x_k) - g_k\|_X \le \kappa_{\text{grad}} \min\{h_k, \Delta_k\} \quad \forall\, k \\ &\exists\, \kappa_{\text{val}} > 0 \quad \text{such that} \quad \|b_k - F(x_k)\|_Y \le \kappa_{\text{val}} \min\{h_k, \Delta_k^2\} \quad \forall\, k \\ &\exists\, \kappa_{\text{jac}} > 0 \quad \text{such that} \quad \|A_k - F'(x_k)\|_{\mathcal{L}(X,Y)} \le \kappa_{\text{jac}} \min\{h_k, \Delta_k\} \quad \forall\, k. \end{aligned}$$

Nonsmooth Trust Regions

Subproblem Model Selection

For the subproblem model, we employ quadratic and linear approximations of f and F, respectively, i.e.,

$$q_k(x) := \frac{1}{2}(B_k(x - x_k), x - x_k)_X + (g_k, x - x_k)_X + q_k(0) \qquad \text{and} \qquad \ell_k(x) := A_k(x - x_k) + b_k,$$

and require that the model components b_k, g_k and A_k satisfy:

$$\begin{aligned} &\exists\, \kappa_{\text{grad}} > 0 && \text{such that} && \|\nabla f(x_k) - g_k\|_X \le \kappa_{\text{grad}} \min\{h_k, \Delta_k\} \quad \forall\, k \\ &\exists\, \kappa_{\text{val}} > 0 && \text{such that} && \|b_k - F(x_k)\|_Y \le \kappa_{\text{val}} \min\{h_k, \Delta_k^2\} \quad \forall\, k \\ &\exists\, \kappa_{\text{jac}} > 0 && \text{such that} && \|A_k - F'(x_k)\|_{\mathcal{L}(X,Y)} \le \kappa_{\text{jac}} \min\{h_k, \Delta_k\} \quad \forall\, k. \end{aligned}$$

Can compute model quantities in *finitely* many iterations!

Nonsmooth Trust Regions

Inexact Objective Values

We also handle inexactness when evaluating the objective function. We replace the *actual reduction* ared_k with the *computed reduction*

$$\mathrm{cred}_k \approx \mathrm{ared}_k := (f(x_k) + \sigma_{\mathfrak{A}}(F(x_k)) + \phi(x_k)) - (f(x_k^+) + \sigma_{\mathfrak{A}}(F(x_k^+)) + \phi(x_k^+)),$$

which is requirec to satisfies:

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

Convergence Theory

Main Result: If $t_k \in [t_{\min}, t_{\max}]\ \forall\, k$, then the iterates produced by the TR algorithm satisfy

$$\liminf_{k\to\infty} \left\{h_k := \tfrac{1}{t_k}\|\mathrm{prox}_{t_k\psi_k}(x_k - t_k g_k) - x_k\|_X\right\} = 0$$

$$\implies \quad \liminf_{k\to\infty} \left\{\hat{h}_k(t) := \tfrac{1}{t}\|\mathrm{prox}_{t\hat{\psi}_k}(x_k - t_k\nabla f(x_k)) - x_k\|_X\right\} = 0 \quad \text{for fixed } t > 0,$$

$$\text{where} \quad \hat{\psi}_k(x) := \sigma_{\mathfrak{A}}(F'(x_k)(x - x_k) + F(x_k)) + \phi(x).$$

Finite Termination: $\forall\,\tau > 0 \quad \exists\, K_\tau \in \mathbb{N}$ such that $h_{K_\tau} \le \tau h_1$ and $\eta_{K_\tau} \le \tau \eta_1$.

Convergence Theory

Main Result: If $t_k \in [t_{\min}, t_{\max}]\ \forall\, k$, then the iterates produced by the TR algorithm satisfy

$$\liminf_{k\to\infty} \left\{ h_k := \frac{1}{t_k}\|\mathrm{prox}_{t_k\psi_k}(x_k - t_k g_k) - x_k\|_X \right\} = 0$$

$$\Longrightarrow \quad \liminf_{k\to\infty} \left\{ \hat{h}_k(t) := \frac{1}{t}\|\mathrm{prox}_{t\hat{\psi}_k}(x_k - t_k\nabla f(x_k)) - x_k\|_X \right\} = 0 \quad \text{for fixed } t > 0,$$

$$\text{where} \quad \hat{\psi}_k(x) := \sigma_{\mathfrak{A}}(F'(x_k)(x - x_k) + F(x_k)) + \phi(x).$$

Finite Termination: $\forall\, \tau > 0 \quad \exists\, K_\tau \in \mathbb{N}$ such that $h_{K_\tau} \le \tau h_1$ and $\eta_{K_\tau} \le \tau \eta_1$.

Strong Convergence: There exists a subsequence $\mathcal{K}$ on which any **strong accumulation point** $\bar{x}$ of $\{x_k\}_{\mathcal{K}}$ is a **stationary point**. See, e.g., **finite dimensions**.

Convergence Theory

Main Result: If $t_k \in [t_{\min}, t_{\max}]\ \forall\, k$, then the iterates produced by the TR algorithm satisfy

$$\liminf_{k\to\infty} \left\{h_k := \frac{1}{t_k}\|\mathrm{prox}_{t_k\psi_k}(x_k - t_k g_k) - x_k\|_X\right\} = 0$$

$$\implies \quad \liminf_{k\to\infty} \left\{\hat{h}_k(t) := \frac{1}{t}\|\mathrm{prox}_{t\hat{\psi}_k}(x_k - t_k\nabla f(x_k)) - x_k\|_X\right\} = 0 \quad \text{for fixed } t > 0,$$

where $\quad \hat{\psi}_k(x) := \sigma_{\mathfrak{A}}(F'(x_k)(x - x_k) + F(x_k)) + \phi(x).$

Finite Termination: $\quad \forall\, \tau > 0 \quad \exists\, K_\tau \in \mathbb{N} \quad$ such that $\quad h_{K_\tau} \le \tau h_1 \quad$ and $\quad \eta_{K_\tau} \le \tau\eta_1.$

Tikhonov Regularization: If $f(x) = f_0(x) + \frac{\alpha}{2}\|x\|_X^2$, $\alpha > 0$, and F, F' and ∇f_0 are **completely continuous**, then there exists a subsequence $\mathcal{K}$ on which any **weak accumulation point** $\bar{x}$ of $\{x_k\}_{\mathcal{K}}$ is a **stationary point**. Moreover, if f_0 is α_0-weakly convex with $\alpha_0 \in (0, \alpha)$, then $x_k \to \bar{x}$ on $\mathcal{K}$.

Numerical Results

Goals: **1. Compare TR method with Primal-Dual Risk Minimization.**
2. Demonstrate TR method with inexact nonlinear PDE solves.

Let $\Omega = (0,1)$, $\Xi = [0,1]^4$ and $\alpha = 10^{-3}$, and consider

$$\min_{z\in L^2(\Omega)} \mathcal{R}\left(\frac{1}{2}\|S(z)-1\|^2_{L^2(\Omega)}\right) + \frac{\alpha}{2}\|z\|^2_{L^2(\Omega)}$$

where $S(z) = u : \Xi \to H^1(\Omega)$ solves the weak form of Burger's equation

$$\begin{aligned} -\nu(\xi)\partial_{xx}u + u\partial_x u &= f(\xi) + z \quad && \text{in } \Omega \text{ a.s.} \\ u(0) = d_0(\xi), \quad u(1) &= d_1(\xi) && \text{a.s.} \end{aligned}$$

Numerical Results

Risk-Averse Control of Burger's Equation

Discretization: Piecewise linear FEM for states and piecewise constant for controls.

Problem Size: 256 control degrees of freedom and 10,000 Monte Carlo samples.

Risk Measure: $\mathcal{R} = (1-\lambda)\mathbb{E} + \lambda\,\mathrm{AVaR}_p$ with $\lambda = 0.75$ and $p = 0.9$.

Stopping Tolerances: 10^{-8} (primal) and 10^{-6} (dual)

method	iter	nfval	ngrad	nhess
TR	6	7	7	102
PD-Risk	8	79	73	128
Bundle	69	182	182	---

TR method required 11x fewer nonlinear PDE solves than PD-Risk and 26x fewer than bundle!

Numerical Results

Risk-Averse Control of Burger's Equation

PDEs solved using damped Newton's method with TR-controlled relative tolerance.

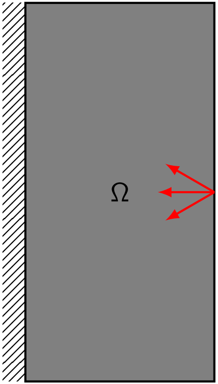
Accurate PDE Solves: Relative Residual Tolerance = $10^{-4}\sqrt{\epsilon_{\text{mach}}}$

Inexact PDE Solves: Tolerance Scaling Parameter $\kappa_{\text{obj}} = \kappa_{\text{grad}} = \kappa_{\text{jac}} = \kappa_{\text{val}} = 1$

k	$J(x_k)$	h_k	Δ_k	$\|x_k^+ - x_k\|$	val tol	grad tol	itsp
0	2.3839e-2	3.7618e-3	1.0000e+1	---	---	---	---
1	1.3854e-2	1.7076e-3	2.5000e+1	7.8350e-1	---	---	15
2	1.2456e-2	1.6069e-3	6.2500e+1	4.2606e-1	---	---	15
3	1.2243e-2	9.3483e-4	1.5625e+2	1.7313e-1	---	---	15
4	1.1990e-2	7.2187e-4	3.9062e+2	4.9320e-1	---	---	15
5	1.1980e-2	1.1940e-4	9.7656e+2	3.4825e-2	---	---	15
6	1.1977e-2	9.7123e-5	2.4414e+3	6.7072e-2	---	---	15
0	2.3424e-2	4.3324e-3	1.0000e+1	---	1.000e-2	4.332e-3	---
1	1.4550e-2	2.1675e-3	2.5000e+1	6.1175e-1	1.657e-4	2.167e-3	15
2	1.2337e-2	9.2366e-4	6.2500e+1	7.1732e-1	3.826e-5	9.237e-4	15
3	1.2063e-2	5.8787e-4	1.5625e+2	3.1769e-1	3.914e-6	5.879e-4	15
4	1.2008e-2	3.0646e-4	3.9062e+2	1.3349e-1	6.617e-7	3.065e-4	15
5	1.1979e-2	7.4265e-5	9.7656e+2	1.5709e-1	3.335e-7	7.426e-5	15

Numerical Results

Goal: **Demonstrate rapid convergence on a highly nonconvex problem.**



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

$$\min_{\rho \in L^2(\Omega)} \; \mathcal{R}\left(\int_{\Gamma_t} T_\xi(x) \cdot [S(\rho)](x)\,\mathrm{d}x \right)$$

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

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

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

Numerical Results

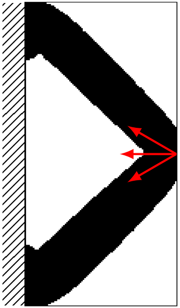
Formulation: Cubic SIMP with Helmholtz filtering (radius$=0.1$).

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

Problem Size: 7,500 density degrees of freedom and 1,000 Monte Carlo samples.

Risk Measure: $\mathcal{R} = \mathrm{AVaR}_p$ with $p = 0.9$.

k	$J(x_k)$	h_k	Δ_k	$\|x_k^+ - x_k\|$	nfval	ngrad	nhess	itsp
0	4.9617e+0	1.6622e+0	1.0000e+1	---	1	1	0	---
1	3.4673e+0	1.1176e+0	2.5000e+1	2.6884e+0	2	2	17	15
2	2.6872e+0	4.9080e-1	6.2500e+1	2.3967e+0	3	3	35	15
3	1.9243e+0	2.0510e-1	1.5625e+2	9.5495e+0	4	4	53	15
4	6.4746e-1	6.0598e-2	1.5625e+2	5.7689e+1	5	5	71	15
5	4.9045e-1	6.5601e-4	1.5625e+2	4.6057e+1	6	6	89	15
6	4.7909e-1	1.5554e-4	3.9063e+2	2.1277e+1	7	7	107	15
7	4.7721e-1	8.9386e-5	9.7656e+2	1.3371e+1	8	8	111	1



On average, required 19 SPG iterations to compute $\mathrm{prox}_{t\psi_k}$ to a tolerance of 10^{-10}!

Numerical Results

Beyond Risk-Averse Optimization

Sparse Control: Let $\Omega = (0, 0.6) \times (0, 0.2)$, Ω_b, $\Omega_o \subset \Omega$, $a = -10$, $b = 10$, $w = 0.2$, $\alpha = 10^{-4}$, $\beta = 10^{-2}$, $\kappa = 0.25$, $\gamma = 1.45$ and consider

$$\min_{a \le z \le b} \max\left\{0,\, w - \frac{1}{|\Omega_o|}\int_{\Omega_o} S(z)\,\mathrm{d}x\right\} + \frac{\alpha}{2}\|z\|_2^2 + \beta\|z\|_1$$

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

$$\begin{aligned} -\kappa\Delta u + \gamma u^3 &= 12\chi_{\Omega_b} && \text{in } \Omega \\ u &= 0 && \text{on } \Gamma_d = [0, 0.6] \times \{0\} \\ \kappa\nabla u \cdot n &= 0 && \text{on } \partial\Omega \setminus \Gamma_d. \end{aligned}$$

dim	2,400	9,600	38,400	153,600
iter	3	3	3	3
nhess	31	40	51	51
av-prox	20.5	20.1	10.7	20.1

Variational Inequalities: Bingham flow in a cylindrical pipe[1] satisfies a VI of the 2nd kind that admits the dual problem

$$\min_{z \in L^2(\Omega)^2} \frac{1}{2}\int_\Omega \mu\nabla S(z) \cdot \nabla S(z)\,\mathrm{d}x$$
$$\text{subject to} \quad |z(x)|_2 \le g \quad \text{for a.a. } x \in \Omega$$

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

$$\int_\Omega \{\mu\nabla u \cdot \nabla v - fv + z \cdot \nabla v\}\,\mathrm{d}x = 0 \quad \forall v \in H_0^1(\Omega).$$

Note: In this application, $\mathfrak{A} = \{0\}$ and $F(z) \equiv 0$.

dim	65,536	262,144	1,048,576	4,194,304
iter	3	3	3	3
lsolve	57	35	39	39
nproj	102	68	69	69

TR algorithm exhibits mesh independence!

---

[1] de los Reyes & González, Path following methods for steady laminar Bingham flow in cylindrical pipes, ESAIM: M2AN, 2009.

Conclusions:

- **Numerical solution** of infinite-dimensional problems requires **expensive approximations**
- Often, the objective function and its gradient can only be computed **inexactly**
- Nonsmooth risk-averse trust region is **provably convergent** even with **inexact computations**
- SPG trust-region subcomponent solvers are **matrix free**, but may **require** many iterations
  Future: Can we incorporate inexact prox computations? Can we handle nonconvex ϕ?
- **What can we say about the local convergence rate?**

References:

- D. P. Kouri & T. M. Surowiec, A Primal-Dual Algorithm for Risk Minimization, Math Programming, 2022.
- R. J. Baraldi & D. P. Kouri, A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations, Math Programming, 2023.
- 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.

Rapid Optimization Library

`rol.sandia.gov`

Trilinos package for **large-scale optimization**. Uses: optimal design, optimal control and inverse problems in engineering applications; mesh optimization; image processing.

RAPID OPTIMIZATION LIBRARY

Numerical optimization made practical:
Any application, any hardware, any problem size.

- **Modern optimization algorithms.**
- **Maximum HPC hardware utilization.**
- **Special programming interfaces for simulation-based optimization.**
- **Optimization under uncertainty.**

- Hardened, production-ready algorithms for **unconstrained, equality-constrained, inequality-constrained and nonsmooth optimization**.
- Novel algorithms for **optimization under uncertainty** and **risk-averse optimization**.
- Unique capabilities for optimization-guided **inexact and adaptive computations**.
- Geared toward **maximizing HPC hardware utilization** through direct use of application data structures, memory spaces, linear solvers and nonlinear solvers.
- Special interfaces for **engineering applications**, for streamlined and efficient use.
- Rigorous **implementation verification**: finite difference and linear algebra checks.
- **Hierarchical and custom** (user-defined) algorithms and stopping criteria.