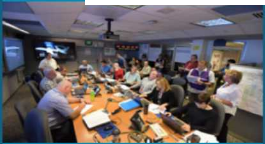
SAND2018-6961C



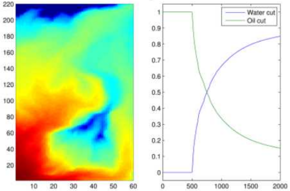
Smoothing Techniques for Risk-Averse PDE-Constrained Optimization

Drew P. Kouri, Thomas M. Surowiec

23rd International Symposium on Mathematical Programming
Bordeaux, France

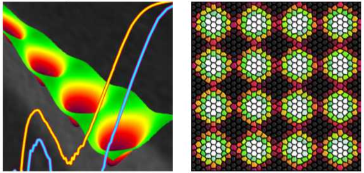
Motivating Applications

Reservoir Optimization



$$v = -\mathbf{K}\lambda(s)\nabla p, \quad \nabla \cdot v = q$$
$$\phi\, \partial_t s + \nabla \cdot (f(s)v) = \widehat{q}$$

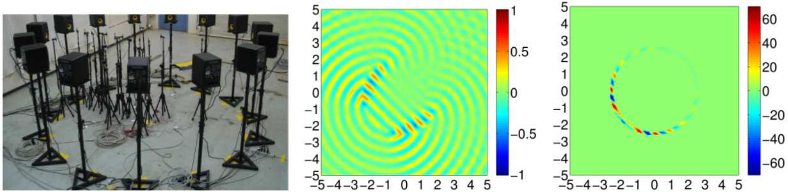
Superconductor Vortex Pinning



Courtesy Argonne National Laboratory

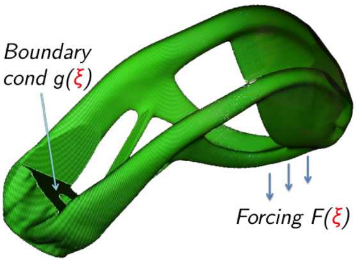
$$\gamma(\partial_t + i\mu)\psi = \epsilon\psi - |\psi|^2\psi + (\nabla - i\mathbf{A})^2\psi$$
$$\mathbf{J} = \mathrm{Im}(\bar{\psi}(\nabla - i\mathbf{A})\psi) - (\partial_t\mathbf{A} + \nabla\mu), \quad \nabla \cdot \mathbf{J} = 0$$

Direct Field Acoustic Testing



$$-\Delta u - \kappa^2(1 + \sigma\epsilon)^2 u = z$$

Topology Optimization and Additive Manufacturing



Given $V_0 \in (0,1)$ compute a density that solves:

$$\min_{0 \le z \le 1} \mathcal{R}\left(\int_D F \cdot S(z)\,dx + \int_{\Gamma_t} t \cdot S(z)\,dx\right)$$

subject to $\int_D z(x)\,dx \le V_0|D|$, where $S(z) = u$ solves the **linear elasticity equations**

$$\begin{aligned} -\nabla \cdot (\mathbf{E}(z) : \epsilon u) &= F, && \text{in } D, \text{ a.s.} \\ \epsilon u &= \frac{1}{2}(\nabla u + \nabla u^\top), && \text{in } D, \text{ a.s.} \\ \epsilon u n &= t, && \text{on } \Gamma_t, \text{ a.s.} \\ u &= g, && \text{on } \Gamma_d, \text{ a.s.} \end{aligned}$$

- Uncertain external forces (loads) and boundary conditions.
- Uncertain internal forces, e.g., residual stresses due to AM.
- Uncertain material properties (porosity, etc.) due to AM.
- Compute light-weight designs that minimize the probability of structural failure.

Optimization Under Uncertainty

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, Z be a reflexive Banach space, and $\mathcal{X} := L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $1 \le p < \infty$. We consider the optimization problem

$$\min_{z \in Z_{\mathrm{ad}}} \{\mathcal{R}(F(z)) + \wp(z)\}$$

where $\mathcal{R} : \mathcal{X} \to (-\infty, \infty]$ is a ***measure of risk***,

$F : Z \to \mathcal{X}$ is the ***uncertain objective function***,

$\wp : Z \to \mathbb{R}$ is a ***deterministic objective function***, and

$Z_{\mathrm{ad}} \subseteq Z$ is a **closed, convex** set of ***decision variables***.

Modeling Risk Preference

What is risk? *Possibility of loss or injury* (Merriam Webster)

...In our optimization problem, $F(z)$ is a **risk**!

We **cannot** directly minimize $F(z) + \wp(z) \in \mathcal{X} := L^p(\Omega, \mathcal{F}, \mathbb{P})$

...How should we quantify our risk?

- **Traditional Stochastic Programming:** Minimize *on average*

$$\mathcal{R}(F(z)) = \mathbb{E}[F(z)].$$

- **Risk-Averse Stochastic Programming:** Model *risk preferences*

$$\mathcal{R}(F(z)) = \mathbb{E}[F(z)] + c\mathbb{E}[(F(z) - \mathbb{E}[F(z)])_+^p]^{1/p}.$$

- **Probabilistic Optimization:** Minimize the *probability of loss*

$$\mathcal{R}(F(z)) = \mathbb{P}(F(z) > \tau).$$

- **Stochastic Orders:** Model risk preference with a *benchmark* Y

$$\mathbb{P}(F(z) \le x) \le \mathbb{P}(Y \le x) \quad \forall x \in \mathbb{R}.$$

Coherent Measures of Risk

$\mathcal{R} : \mathcal{X} \to (-\infty, \infty]$ is a **coherent** measure of risk if it satisfies

(R1) **Subadditivity:** $\mathcal{R}(X + X') \le \mathcal{R}(X) + \mathcal{R}(X')$

(R2) **Monotonicity:** $X \ge X'$ a.s. $\implies \mathcal{R}(X) \ge \mathcal{R}(X')$

(R3) **Translation Equivariance:** $\mathcal{R}(X + t) = \mathcal{R}(X) + t, \quad \forall t \in \mathbb{R}$

(R4) **Positive Homogeneity:** $\mathcal{R}(tX) = t\mathcal{R}(X), \quad \forall t > 0$

Note: {(R1) + (R4) $\implies$ convexity} and {convexity + (R4) $\implies$ (R1)}

Examples of **coherent** risk measures with $X \in \mathcal{X}$:

- Risk Neutral: $\mathcal{R}(X) = \mathbb{E}[X]$
- Mean Plus Semideviation: $\mathcal{R}(X) = \mathbb{E}[X] + c\mathbb{E}[(X - \mathbb{E}[X])_+^p]^{1/p}, \; c \in (0,1)$
- Conditional Value-at-Risk: $\mathcal{R}(X) = \inf \{t + (1-\beta)^{-1}\mathbb{E}[(X-t)_+]\}, \; \beta \in (0,1)$

Ph. Artzner, F. Delbaen, J.-M. Eber and D. Heath, *Coherent measures of risk.* Math. Finance, 1999.

Coherent Measures of Risk

Some Good and *Not* So Good Properties?

Biconjugate Representation: Recall $\mathcal{R}^*(\vartheta) = \sup_X\{\mathbb{E}[\vartheta X] - \mathcal{R}(X)\}$

- If $\mathcal{R}$ is proper, **convex** and lsc

$$\iff \quad \mathcal{R}(X) = \sup\,\{\mathbb{E}[\vartheta X] - \mathcal{R}^*(\vartheta) \mid \vartheta \in \mathrm{dom}(\mathcal{R}^*)\}$$

- If $\mathcal{R}$ is **translation equivariant** and **monotonic**

$$\iff \quad \mathrm{dom}(\mathcal{R}^*) \subseteq \{\vartheta \in \mathcal{X}^* \mid \mathbb{E}[\vartheta] = 1,\ \vartheta \geq 0 \text{ a.s.}\}$$

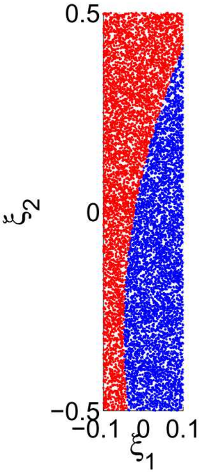
- If $\mathcal{R}$ is **positive homogeneous**

$$\iff \quad \mathcal{R}(X) = \sup_{\vartheta\in\mathrm{dom}(\mathcal{R}^*)} \mathbb{E}[\vartheta X]$$

Optimal $\vartheta^\star \in \mathrm{dom}(\mathcal{R}^*)$ are called ***risk identifiers***

Example (Conditional Value-at-Risk): $\mathcal{R} = \mathrm{CVaR}_\beta$

$$\mathrm{dom}(\mathcal{R}^*) = \left\{\vartheta \in \mathcal{X}^* \mid \mathbb{E}[\vartheta] = 1,\ 0 \leq \vartheta \leq \frac{1}{1-\beta} \text{ a.s.}\right\}$$



Differentiability: If $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is **coherent**, then $\mathcal{R}$ is **Fréchet differentiable**

$$\iff \quad \exists\,\vartheta \in \mathcal{X}^* \text{ with } \vartheta \geq 0 \text{ a.s., } \mathbb{E}[\vartheta] = 1, \text{ and } \mathcal{R}(X) = \mathbb{E}[\vartheta X] \text{ for all } X \in \mathcal{X}$$

Is Nondifferentiability *Really* an Issue?

Example: Optimal control of Burger's equation using CVaR

— **Problem size is small:** 1D spatial domain, 4D stochastic domain

— PDE is nonlinear $\implies$ Objective function is not convex

— CVaR risk measure quantifies *tail weight* and is **not** differentiable

Application of a *nonconvex nonsmooth* optimization algorithm:

β	0.1	0.5	0.9
# iter	9,740	10,035	10,128

Required $\mathcal{O}(10^8)$ nonlinear and $\mathcal{O}(10^8)$ linearized PDE solves!

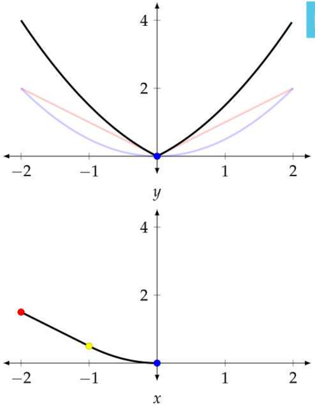
Application of smoothed $\mathcal{R}$ with globalized Newton's method:

Required $\mathcal{O}(10^6)$ nonlinear and $\mathcal{O}(10^7)$ linearized PDE solves!

Solving real world problems is intractable without . . .

- Better **nonsmooth** optimization algorithm or **differentiable** $\mathcal{R}$
- **Adaptive/variable fidelity** approximation in physical and stochastic space
- In optimization, accuracy is **not** required far from a solution

Epi-Regularized Risk Measures



Let $\mathcal{R}$, $\Phi : \mathcal{X} \to (-\infty, \infty]$ satisfy:

1. $\mathcal{R}$, Φ are proper, closed and convex
2. $\mathrm{dom}\,\mathcal{R}^* \subseteq \mathrm{dom}\,\Phi^*$
3. $(\mathrm{dom}\,\mathcal{R}^* - \mathrm{dom}\,\Phi^*)$ contains a neighborhood of 0

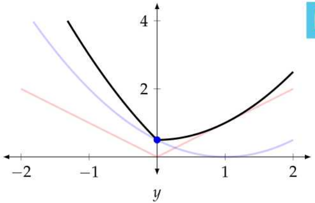
The **epi-regularization** of $\mathcal{R}$ is given by

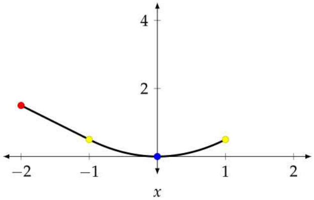
$$\mathcal{R}_\varepsilon(X) := \inf_{Y \in \mathcal{X}} \left\{ \mathcal{R}(X - Y) + \varepsilon \Phi(Y/\varepsilon) \right\}, \quad \varepsilon > 0$$

Properties of $\mathcal{R}_\varepsilon^\Phi$:

1. $(\mathcal{R}_\varepsilon^\Phi)^*(\vartheta) = \mathcal{R}^*(\vartheta) + \varepsilon\Phi^*(\vartheta)$ with $\mathrm{dom}(\mathcal{R}_\varepsilon^\Phi)^* = \mathrm{dom}\,\mathcal{R}^* \cap \mathrm{dom}\,\Phi^* = \mathrm{dom}\,\mathcal{R}^*$
2. $-\varepsilon\Phi(0) \le \mathcal{R}(X) - \mathcal{R}_\varepsilon^\Phi(X) \le \varepsilon\Phi^*(\vartheta) \quad \forall\,\vartheta \in \partial\mathcal{R}(X)$
3. **Coherent Risk:** $\mathcal{R}_\varepsilon^\Phi$ is convex, translation equivariant and monotonic
4. **Coherent Risk:** $\mathcal{R}_\varepsilon^\Phi$ is **not** positively homogeneous $\implies$ $\mathcal{R}_\varepsilon^\Phi$ is **not** coherent

Epi-Regularized Risk Measures





Let $\mathcal{R}$, $\Phi : \mathcal{X} \to (-\infty, \infty]$ satisfy:

1. $\mathcal{R}$, Φ are proper, closed and convex
2. $\mathrm{dom}\,\mathcal{R}^* \subseteq \mathrm{dom}\,\Phi^*$
3. $(\mathrm{dom}\,\mathcal{R}^* - \mathrm{dom}\,\Phi^*)$ contains a neighborhood of 0

The **epi-regularization** of $\mathcal{R}$ is given by

$$\mathcal{R}_\varepsilon(X) := \inf_{Y \in \mathcal{X}} \{\mathcal{R}(X - Y) + \varepsilon\Phi(Y/\varepsilon)\}, \quad \varepsilon > 0$$

Differentiability of $\mathcal{R}_\varepsilon^\Phi$:

1. If Φ^* is *strictly convex* on $\mathrm{dom}\,\mathcal{R}^*$, then $\mathcal{R}_\varepsilon^\Phi$ is *Hadamard differentiable*
2. If, in addition, Φ^* is *weak* closed* and satisfies

$$\theta_k \rightharpoonup^* \theta \quad \text{in } \mathcal{X}^* \quad \text{and} \quad \Phi^*(\theta_k) \to \Phi^*(\theta) \quad \Longrightarrow \quad \theta_k \to \theta \quad \text{in } \mathcal{X}^*,$$

then $\mathcal{R}_\varepsilon^\Phi$ is *continuously Fréchet differentiable*

Example: Optimized Certainty Equivalents

Let $u(t) = -v(-t)$ is a normalized, concave utility function and define $\mathcal{R}(X) = \inf_t\{t + \mathbb{E}[v(X-t)]$ and $\Phi(X) = \mathbb{E}[\phi(X)]$, then

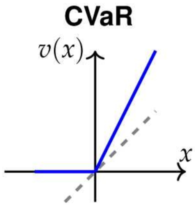
$$\mathcal{R}_\varepsilon^\Phi(X) = \inf_{Y\in\mathcal{X}}\left\{\inf_{t\in\mathbb{R}}\{t + \mathbb{E}[v(X-Y-t)]\} + \varepsilon\Phi(Y/\varepsilon)\right\}$$
$$= \inf_{t\in\mathbb{R}}\left\{t + \inf_{Y\in\mathcal{X}}\mathbb{E}\left[v(X-Y-t) + \varepsilon\phi(Y/\varepsilon)\right]\right\}$$

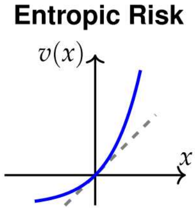
Decomposability of $\mathcal{X}$ ensures that

$$\mathcal{R}_\varepsilon^\Phi(X) = \inf_{t\in\mathbb{R}}\left\{t + \mathbb{E}\left[\inf_{y\in\mathbb{R}}\{v(X-y-t) + \varepsilon\phi(y/\varepsilon)\}\right]\right\}$$

The inner infimum is the **infimal convolution** of $v(x)$ with $\phi(x)$

$$\Longrightarrow \quad \mathcal{R}_\varepsilon^\Phi(X) = \inf_{t\in\mathbb{R}}\{t + \mathbb{E}[v_\varepsilon^\phi(X-t)]\}!$$





Epi-Regularized Risk Measures

Results: If $\Phi(0) < \infty$ and $\Phi(1) < \infty$, then

- $\mathcal{R}_{\varepsilon_n}$ *Mosco converges* to $\mathcal{R}$ for $\varepsilon_n \downarrow 0$
- $\mathcal{R}_{\varepsilon_n}$ *pointwise converges* to $\mathcal{R}$ for $\varepsilon_n \downarrow 0$
- If $\Phi(0) \leq 0$, then $\mathcal{R}_\varepsilon^\Phi$ *monotonically increases* to $\mathcal{R}$ as $\varepsilon \downarrow 0$

Recall: $G_n : \mathcal{X} \to (-\infty, \infty]$ *Mosco converges* to $G : \mathcal{X} \to (-\infty, \infty]$ if

1. For all $X_n \rightharpoonup X$ in $\mathcal{X}$, we have
$$\liminf_{n\to\infty} G_n(X_n) \geq G(X)$$
2. For all $X \in \mathcal{X}$, there exists $X_n \to X$ in $\mathcal{X}$ such that
$$\limsup_{n\to\infty} G_n(X_n) \leq G(X)$$

$\implies$ Minimizers of G_n converge to minimizers of G (cf. Γ-convergence)

$$\min_{z\in Z_{\rm ad}}\{J(z) := \mathcal{R}(F(z)) + \wp(z)\} \quad \rightsquigarrow \quad \min_{z\in Z_{\rm ad}}\{J_\varepsilon(z) := \mathcal{R}_\varepsilon^\Phi(F(z)) + \wp(z)\}$$

- **Consistency of Minimizers:**

  Let $z_n \in Z_{\rm ad}$ denote a minimizer of J_{ε_n} with $\varepsilon_n \downarrow 0$.

  If F is *completely continuous*

  $$z \rightharpoonup z^\star \quad \text{in } Z \qquad \Longrightarrow \qquad F(z) \to F(z^\star) \quad \text{in } \mathcal{X}.$$

  Then any *weak limit point* of $\{z_n\}$ minimizes F over $Z_{\rm ad}$.

- **Consistency of First-Order Stationary Points:**

  Let $z_n \in Z_{\rm ad}$ denote a $1^{\rm st}$-order stationary point of J_{ε_n} with $\varepsilon_n \downarrow 0$, i.e.,

  $$0 \in \partial_C J_{\varepsilon_n}(z_n) + \mathcal{N}_{Z_{\rm ad}}(z_n).$$

  If ∇F is *completely continuous*

  $$z \rightharpoonup z^\star \quad \text{in } Z \qquad \Longrightarrow \qquad \nabla F(z) \to \nabla F(z^\star) \quad \text{in } L^p(\Omega, \mathcal{F}, \mathbb{P}; Z^*).$$

  Then any *weak limit point*, z, of $\{z_n\}$ satisfies $\quad 0 \in \partial_C J(z) + \mathcal{N}_{Z_{\rm ad}}(z)$.

Epi-Regularized Risk Measures

Let the following conditions hold:

- $z \in Z_{\text{ad}}$ is a minimizer of J and $z_\varepsilon \in Z_{\text{ad}}$ is a minimizer of J_ε
- F and $\wp$ are continuously Fréchet differentiable
- Either ($h_\varepsilon := z_\varepsilon - z$)
  1. F is twice continuously Fréchet differentiable and $\exists\, K \in \mathcal{X}$ s.t. $K \in (0, \infty)$ a.s. and

$$K\|z_\varepsilon - z\|_Z^2 \le \int_0^1 \int_0^1 \nabla^2 F(z + \tau t h_\varepsilon)(h_\varepsilon, h_\varepsilon)\mathrm{d}\tau \mathrm{d}t \text{ a.s.}$$

  2. $\wp$ is twice continuously Fréchet differentiable and $\exists\, L > 0$ such that

$$L\|z_\varepsilon - z\|_Z^2 \le \int_0^1 \int_0^1 \nabla^2 \wp(z + \tau t h_\varepsilon)(h_\varepsilon, h_\varepsilon)\mathrm{d}\tau \mathrm{d}t$$

Then, $\exists\, \vartheta \in \partial\mathcal{R}(F(z))$ and $c > 0$ such that

$$c\|z_\varepsilon - z\|_Z^2 \le \varepsilon(\Phi(0) + \Phi^*(\vartheta))$$

Epi-Regularized CVaR with $\Phi(X) = \frac{1}{2}\mathbb{E}[X^2] + \mathbb{E}[X] \implies \Phi(0) = 0$ **and** $\Phi^*(\vartheta) \le \frac{\beta^2}{2(1-\beta)}$

Nonconvex Stochastic Programming in $\mathbb{R}^n$

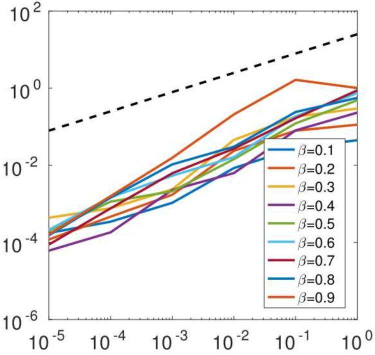
Let $\gamma > 0$ and $\bar{z} \in [-1,1]^n$ be fixed. Consider the optimization problem

$$\min_{z\in\mathbb{R}^n} \left\{ \mathcal{R}\left(\frac{1}{2}(1-\tanh(\eta\xi^\top z))\right) + \frac{\gamma}{2} z^\top z \right\}$$

where ξ is uniformly distributed on $[0,1]^n$ and $\eta = \mathrm{sgn}(\xi^\top \bar{z})$.

Epi-regularization Error in Optimal Solutions:

- $\mathcal{R}(X) = \mathrm{CVaR}_\beta[X]$ with $\beta \in \{0.1, 0.2, \ldots, 0.9\}$
- $\Phi(X) = \frac{1}{2}\mathbb{E}[X^2] + \mathbb{E}[X]$ with $\varepsilon \in \{10^{-5}, 10^{-4}, \ldots, 10^{0}\}$



Risk-Averse PDE-Constrained Optimization

Let $D_o \subseteq D \subset \mathbb{R}^2$ and $\alpha > 0$. We consider the optimization problem

$$\min_{z \in Z_{\mathrm{ad}}} \quad \frac{1}{2}\mathrm{CVaR}_\beta\left(\int_{D_o} (1 - S(z;x,\xi))_+^2 \,\mathrm{d}x\right) + \frac{\alpha}{2}\int_D z(x)^2 \,\mathrm{d}x$$

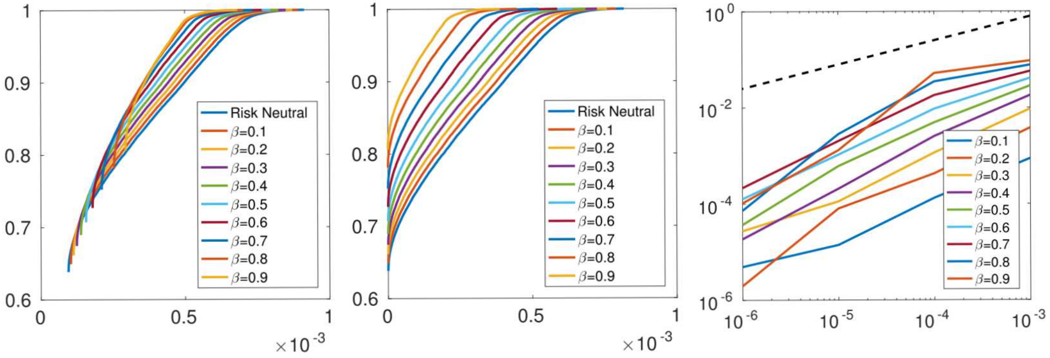
where $S(z) = u \in L^q(\Omega, \mathcal{F}, \mathbb{P}; H^1(D))$ solves the weak form of

$$\begin{aligned} -\kappa(\xi)\Delta u(\xi) + c(\xi)u(\xi) &= B(\xi)z + f && \text{in } D, \text{ a.s.} \\ \kappa(\xi)\nabla u(\xi)\cdot n &= 0 && \text{on } \partial D, \text{ a.s.} \end{aligned}$$

Coefficients: $\kappa(\xi) = 2.5 \times 10^{\xi_1}$, $c(\xi) = 1.45 \times 10^{\xi_2}$, $r(\xi) = 10^{\xi_3}$, and $\zeta(\xi) = B(\xi)z \in L^2(\Omega, \mathcal{F}, \mathbb{P}; H^1(D))$ solves the weak form of

$$\begin{aligned} -r(\xi)\Delta\zeta(\xi) + \zeta(\xi) &= z && \text{in } D, \text{ a.s.} \\ r(\xi)\nabla\zeta(\xi)\cdot n &= 0 && \text{on } \partial D, \text{ a.s.} \end{aligned}$$

Risk-Averse PDE-Constrained Optimization



Left: CDF of full objective. **Center:** CDF of random objective. **Right:** Control errors.

Conclusions:

- **Numerical solution** of risk-averse PDE-optimization is **expensive**
- Most **coherent risk measures** are **not** Fréchet differentiable
- Use **infimal convolution** to **smooth** risk measures
- Appropriate assumptions ensure smoothed risk **is** Fréchet differentiable
- Proved **consistency** of minimizers and first-order stationary points

Open Problems:

- Path-following methods for smoothed risk measures
- Algorithms for **nonsmooth**, **nonconvex** risk-averse PDE-optimization
- Analysis and algorithms for **probabilistic** PDE-optimization
- Analysis and algorithms for **dominance-constrained** PDE-optimization