Theory and Algorithms for PDE-Constrained Optimization under Uncertainty

SAND2015-0459PE

Drew Kouri

Optimization and Uncertainty Quantification
Sandia National Laboratories, Albuquerque, New Mexico
dpkouri@sandia.gov

January 26, 2016

Duke University, Durham, NC.

Joint work with T. Surowiec (HUBerlin), M. Heinkenschloss (Rice),

D. Ridzal (SNL), and B. van Bloemen Waanders (SNL).

Funding support: ASC.

Outline

Problem Formulation

Known Probability Distribution

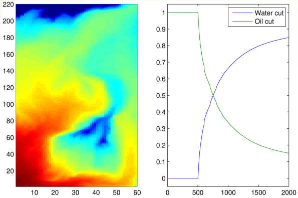
- Risk-Neutral Optimization
- Risk-Averse Optimization

Unknown Probability Distribution

Conclusions

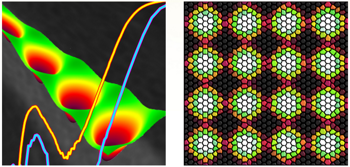
Motivation

Reservoir Optimization



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

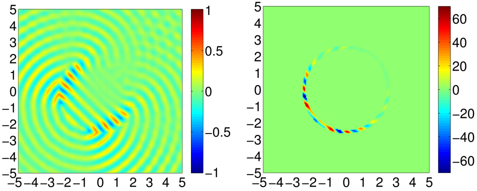
Superconductor Vortex Pinning



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

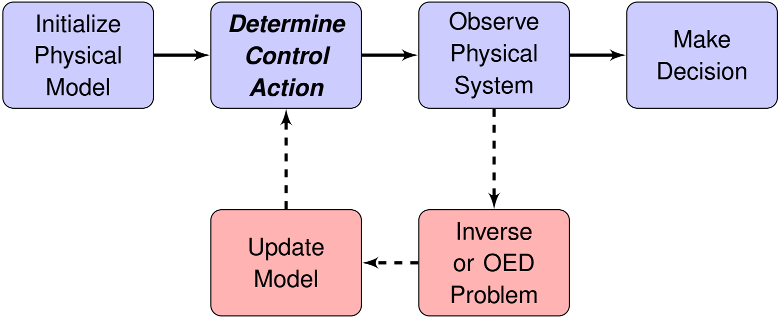


$$-\Delta u - \kappa^2(1+\sigma\epsilon)^2 u = z$$

Sandia National Laboratories

Optimization Problem Formulation

Goal: Control uncertainty rather than quantify uncertainty.



We implement the control prior to observing the state.
Control is deterministic.

Deterministic Optimization of PDEs

Optimal Control: Given $\alpha > 0$, $\Omega_o \subseteq \Omega$, $\Omega_c \subseteq \Omega$, and $w \in L^2(\Omega_o)$.

$$\min_{z \in \mathcal{Z}} J(z) \equiv \frac{1}{2} \int_{\Omega_o} ((U(z))(x) - w(x))^2 \, \mathrm{d}x + \frac{\alpha}{2} \int_{\Omega_c} z^2(x) \, \mathrm{d}x$$

where $U(z) = u \in H^1(\Omega)$ solves the **weak form** of

$$\begin{aligned} -\nabla \cdot (\kappa \nabla u) + N(u) &= z && \text{in } \Omega \\ u &= g, && \text{on } \partial\Omega. \end{aligned}$$

Topology Optimization: Given $0 < V_0 < 1$ and $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$.

$$\min_{z \in \mathcal{Z}} J(z) \equiv \int_{\Omega} F(x) \cdot (U(z))(x) \, \mathrm{d}x \quad \text{s.t.} \quad 0 \le z \le 1, \quad \int_{\Omega} z(x) \, \mathrm{d}x \le V_0 |\Omega|$$

where $U(z) = u \in H^1(\Omega)^d$ solves the **weak form** of

$$\begin{aligned} -\nabla \cdot (\mathbf{E}(z) : \epsilon(u)) &= F, && \text{in } \Omega \\ \epsilon(u) &= \frac{1}{2}(\nabla u + \nabla u^\top), && \text{in } \Omega \\ u &= g, && \text{on } \partial\Omega. \end{aligned}$$

Sandia National Laboratories

General PDE-Constrained Optimization

Let $\mathcal{U}$ and $\mathcal{Z}$ be reflexive Banach spaces, $\mathcal{C}$ be a Banach space, $f : \mathcal{U} \times \mathcal{Z} \to \mathbb{R}$, and $e : \mathcal{U} \times \mathcal{Z} \to \mathcal{C}$. Consider

$$\min_{z \in \mathcal{Z}_{\text{ad}}} J(z) = f(U(z), z)$$

where $U(z) = u \in \mathcal{U}$ solves the **weak form** PDE

$$e(u,z) = 0 \quad \text{and} \quad \mathcal{Z}_{\text{ad}} \subseteq \mathcal{Z}.$$

Assumptions:

- For each $z \in \mathcal{Z}_{\text{ad}}$, $e(u,z) = 0$ is well posed, i.e.,
  - $\exists!\, U(z) = u \in \mathcal{U}$ such that $e(U(z), z) = 0$;
  - $\exists\, c > 0$ indpendent of z such that $\|U(z)\|_{\mathcal{U}} \leq c(\|z\|_{\mathcal{Z}} + 1)$.
- e is sequentially weakly continuous.
- f is sequentially weakly lower semicontinuous (lsc).
- $\mathcal{Z}_{\text{ad}}$ is convex, closed and bounded – or –
  $\mathcal{Z}_{\text{ad}} = \mathcal{Z}$ and J is coercive, i.e., $\lim_{\|z\|_{\mathcal{Z}} \to \infty} J(z) = \infty$.

Result: There exists a minimizer to J in $\mathcal{Z}_{\text{ad}}$.

Differentiability

Assumptions:

- f and e are k-times continuously Fréchet differentiable.
- For each $z \in \mathcal{Z}_{\mathrm{ad}}$, $e_u(U(z), z)$ has a bounded inverse.

Result: J is k-times continuously Fréchet differentiable in $\mathcal{Z}_{\mathrm{ad}}$.

Gradient Evaluation:

- Compute $U(z) = u \in \mathcal{U}$, the weak solution of the **state** equation:

$$e(u, z) = 0.$$

- Compute $\Lambda(z) = \lambda \in \mathcal{C}^*$, the weak solution of the **adjoint** equation:

$$e_u(U(z), z)^{-*}\lambda + f_u(U(z), z) = 0.$$

- Evaluate **gradient**: $\nabla J(z) = f_z(U(z), z) + e_u(U(z), z)^* \Lambda(z)$.

Example

Optimal Control: Given $\alpha > 0$, $\Omega_o \subseteq \Omega$, $\Omega_c \subseteq \Omega$, and $w \in L^2(\Omega_o)$.

$$\min_{z \in \mathcal{Z}} J(z) \equiv \frac{1}{2} \int_{\Omega_o} ((U(z))(x) - w(x))^2 \, \mathrm{d}x + \frac{\alpha}{2} \int_{\Omega_c} z^2(x) \, \mathrm{d}x$$

where $U(z) = u \in H^1(\Omega)$ solves the **weak form** of

$$-\nabla \cdot (\kappa \nabla u) + N(u) = z, \quad \text{in } \Omega$$
$$u = g, \quad \text{on } \partial\Omega.$$

Gradient Evaluation:

- Compute $U(z) = u \in H^1(\Omega)$, the weak solution of the **state** equation:

$$-\nabla \cdot (\kappa \nabla u) + N(u) = z, \quad \text{in } \Omega$$
$$u = g, \quad \text{on } \partial\Omega.$$

- Compute $\Lambda(z) = \lambda \in H^1(\Omega)$, the weak solution of the **adjoint** equation:

$$-\nabla \cdot (\kappa \nabla \lambda) + N_u(u)^* \lambda = -(u - \bar{w}), \quad \text{in } \Omega$$
$$\lambda = 0, \quad \text{on } \partial\Omega.$$

- Evaluate **gradient**: $\nabla J(z) = \alpha z - \Lambda(z)$.

Optimization of PDEs with Uncertain Inputs

Optimal Control: Given $\alpha > 0$, $\Omega_o \subseteq \Omega$, $\Omega_c \subseteq \Omega$, and $w \in L^2(\Omega_o)$.

$$\min_{z \in \mathcal{Z}} \; J(z) \equiv \frac{1}{2}\mathcal{R}\left[\int_{\Omega_o} ((U(z))(\xi, x) - w(x))^2 \,\mathrm{d}x\right] + \frac{\alpha}{2}\int_{\Omega_c} z^2(x)\,\mathrm{d}x$$

where $U(z) = u : \Xi \to H^1(\Omega)$ solves the **weak form** of

$$-\nabla \cdot (\epsilon(\xi)\nabla u(\xi)) + N(u(\xi), \xi) = z, \qquad \text{in } \Omega, \text{ a.s.}$$

$$u(\xi) = g(\xi), \qquad \text{on } \partial\Omega, \text{ a.s.}$$

Topology Optimization: Given $0 < V_0 < 1$ and $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$.

$$\min_{z \in \mathcal{Z}} \; J(z) \equiv \mathcal{R}\left[\int_{\Omega} F(\xi, x) \cdot (U(z))(\xi, x)\,\mathrm{d}x\right] \quad \text{s.t.} \quad 0 \le z \le 1, \quad \int_{\Omega} z(x)\,\mathrm{d}x \le V_0|\Omega|$$

where $U(z) = u : \Xi \to H^1(\Omega)^d$ solves the **weak form** of

$$-\nabla \cdot (\mathbf{E}(z) : \epsilon(u(\xi)) = F(\xi), \qquad \text{in } \Omega, \text{ a.s.}$$

$$\epsilon(u(\xi) = \frac{1}{2}(\nabla u(\xi) + \nabla u(\xi)^\top), \qquad \text{in } \Omega, \text{ a.s.}$$

$$u(\xi) = g(\xi), \qquad \text{on } \partial\Omega, \text{ a.s.}$$

General PDE-Optimization under Uncertainty

Let $(\Xi, \mathcal{F}, P)$ be a complete probability space with $\Xi \subseteq \mathbb{R}^M$. Consider

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} J(z) = \mathcal{R}(f((U(z))(\xi), z, \xi))$$

where $U(z) = u \in L^p_P(\Xi; \mathcal{U})$ solves the **weak form** PDE

$$e(u, z, \xi) = 0 \quad \text{and} \quad \mathcal{Z}_{\mathrm{ad}} \subseteq \mathcal{Z}.$$

Assumptions:

- For each $z \in \mathcal{Z}_{\mathrm{ad}}$ and $\gamma \in \Xi$, $e(u, z, \xi) = 0$ is well posed, i.e.,
  - $\exists!\, U(z) = u \in L^p_P(\Xi; \mathcal{U})$ such that $e(U(z), z, \xi) = 0$;
  - $\exists\, c > 0$ indpendent of z and $\xi \in \Xi$ such that $\|U(z)\|_{\mathcal{U}} \leq c(\|z\|_{\mathcal{Z}} + 1)$.
- e is a.s. sequentially weakly continuous.
- f is a.s. sequentially weakly lsc and $f((U(z))(\xi), z, \xi) \in L^q_P(\Xi)$.
- $\mathcal{Z}_{\mathrm{ad}}$ is convex, closed and bounded – or –
  $\mathcal{Z}_{\mathrm{ad}} = \mathcal{Z}$ and $z \mapsto f((U(z))(\xi), z, \xi)$ is a.s. i.e.,
  $\lim_{\|z\|_{\mathcal{Z}} \to \infty} f((U(z))(\xi), z, \xi) = \infty$.

Sandia National Laboratories

Risk Measures, $\mathcal{R}$

Assumptions:

- $\mathcal{R} : L^q_P(\Xi) \to \mathbb{R} \cup \{+\infty\}$
  see, Rockafellar, Uryasev, Shapiro, Dentcheva, Ruszczynski, . . .
- $\mathcal{R}$ is convex, lsc and satisfies $\mathcal{R}(C) = C$ for all constants C;
- $\mathcal{R}$ is monotonic, i.e., if $X_1 \geq X_2$ a.s., then $\mathcal{R}(X_1) \geq \mathcal{R}(X_2)$.

Result: There exists a minimizer to J in $\mathcal{Z}_{\text{ad}}$.

Risk Neutral v.s. Risk Averse

- **Risk Neutral:** $\mathcal{R} = \mathbb{E}$.
  - Optimal solution minimizes on average.
- **Risk Averse:** $\mathcal{R}(X) > \mathbb{E}[X] \quad \forall$ nonconstant $X \in L^q_P(\Xi)$.
  - More conservative than $\mathcal{R} = \mathbb{E}$.
  - Can minimize measures of deviation and/or tail events.

Sandia National Laboratories

Known v.s. Unknown Probability Distribution

Known Probability Distribution:

- $\Xi \subseteq \mathbb{R}^M$ is known and P has Lebesgue density $\rho : \Xi \to [0, \infty)$.
- Enables UQ techniques including gPC and sampling.
- All analysis performed in $L^p_\rho(\Xi)$ instead of $L^p_P(\Xi)$.

Unknown Probability Distribution:

- Must determine optimal solutions that are robust to unknown pdf.
- Use data to estimate pdf (i.e., experimental data or inverted coefficients).
- Reformulate optimization problem into a minimax problem

$$\min_{z \in \mathcal{Z}_{\text{ad}}} \sup_{P \in \mathcal{A}} \mathbb{E}_P[f((U(z)(\xi), z, \xi)]$$

- $\mathcal{A}$ is called the *ambiguity* set and is defined with data, i.e., moment matching.
- Must discretize the probability measures $P \in \mathcal{A}$.
- May also require specialized optimization algorithms.

Sandia National Laboratories

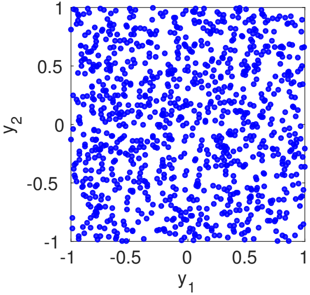
Sample-based Approximation

Monte Carlo

- Handles general Ξ and ρ;
- Error is independent of dim.;
- Error is probabilistic with rate

$$\mathbb{E}[\text{error}] = \mathcal{O}(Q^{-\frac{1}{2}});$$

- Use QMC and variance reduction;
- Weights are positive ($\omega_k = Q^{-1}$).

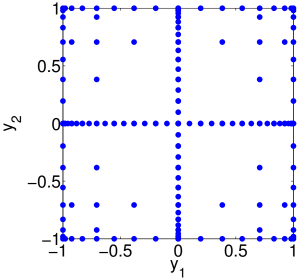


Quadrature

- Often requires TP Ξ and ρ;
- Error is dependent on dim.;
- Sparse grids lessen affect of dim.;
- Regularity accelerates conv., e.g.,

$$\text{error} = \mathcal{O}(Q^{-r}\log(Q)^{(M-1)(r+1)});$$

- SG weights are pos. and neg.



Sandia National Laboratories

Outline

Problem Formulation

Known Probability Distribution

- Risk-Neutral Optimization
- Risk-Averse Optimization

Unknown Probability Distribution

Conclusions

Outline

Problem Formulation

Known Probability Distribution

Risk-Neutral Optimization

Risk-Averse Optimization

Unknown Probability Distribution

Conclusions

Risk-Neutral Optimization Problem

$$\min_{z\in\mathcal{Z}_{\rm ad}} J(z) \equiv \mathbb{E}[f(U(z),z,\cdot)] = \int_\Xi \rho(\xi) f((U(z))(\xi),z,\xi)\,\mathrm{d}\xi$$

- There exists minimizers of J in $z \in \mathcal{Z}_{\rm ad}$.
- J is k-times continuously Fréchet differentiable if f and e are for all $\xi \in \Xi$.

Since pdf is known, we replace $\mathbb{E}$ with a quadrature approximation:

$$J(z) \approx J_Q(z) \equiv \sum_{k=1}^{Q} \omega_k f((U(z))(\xi_k),z,\xi_k)$$

- There exists minimizers of J_Q in $z \in \mathcal{Z}_{\rm ad}$ if $\mathcal{Z}_{\rm ad}$ is bounded.
- J_Q is k-times continuously Fréchet differentiable if f and e are for all $\xi \in \Xi$.

Require $(U(z))(\xi_k) = u_k \in \mathcal{U}$, $k = 1,\ldots,Q$, that solve the **weak form** PDEs,

$$e(u_k,z,\xi_k) = 0, \quad k = 1,\ldots,Q.$$

- Decoupled system of PDEs can be solved concurrently.
- Use favorite numerical PDE technique to solve deterministic PDEs.
- Convergence depends on quad. rule and regularity of state and adjoint w.r.t. ξ.

Sandia National Laboratories

Numerical Optimization for Risk-Neutral Problems

- Efficient Numerical Method
  - Gradient computation requires two PDE solves per quad. point;
  - High accuracy or large $\dim(\Xi) \implies$ large Q;
  - In optimization, accuracy **not** required far from a solution.
- Accurate Characterization of the Random Field
  - Use adaptive sparse grids to exploit anisotropy in random fields (Gerstner and Griebel 1998, Ma and Zabaras 2009, Agarwal and Aluru 2009, Webster et al.);
  - Use adaptive finite elements to accurately resolve PDE solution (Carstensen 2005, Becker et al. 2007).
- Trust Regions
  - Globally convergent opt. algorithm (Powell 1975, Sorensen 1982);
  - Allows for inexact gradients and objective functions (Carter 1989, Heinkenschloss and Vicente 2001, Ziems and Ulbrich 2011);
  - Natural framework for model management (Alexandrov et al. 1998, Dennis and Torczon 1996).

Trust-Region Algorithm

Given: z_0, $m_0(s) \approx J(z_0 + s)$, $J_0 \approx J$, $\delta_0 > 0$, and $\text{gtol} > 0$.
While $\|\nabla m_k(s)\|_{\mathcal{Z}} > \text{gtol}$

1. **Model Update:** Choose a new $m_k(s) \approx J(z_k + s)$.
2. **Step Computation:** Approximate a solution, s_k, to the subproblem

$$\min_{s \in \mathcal{Z}} \; m_k(s) \quad \text{subject to} \quad \|s\|_{\mathcal{Z}} \leq \delta_k.$$

3. **Objective Update:** Choose a new $J_k(z) \approx J(z)$.
4. **Step Acceptance:** Compute

$$\varrho_k = \frac{J_k(z_k) - J_k(z_k + s_k)}{m_k(0) - m_k(s_k)}.$$

   If $\varrho_k \geq \eta \in (0, 1)$, then $z_{k+1} = z_k + s_k$ else $z_{k+1} = z_k$.

5. **Trust Region Update:** Choose a new trust-region radius, δ_{k+1}.

EndWhile

Trust-Region Algorithm

Given: z_0, $m_0(s) \approx J(z_0 + s)$, $J_0 \approx J$, $\delta_0 > 0$, and $\mathrm{gtol} > 0$.
While $\|\nabla m_k(s)\|_{\mathcal{Z}} > \mathrm{gtol}$

1. **Model Update:** Choose a new $m_k(s) \approx J(z_k + s)$. ← **ADAPTIVITY**

2. **Step Computation:** Approximate a solution, s_k, to the subproblem

$$\min_{s \in \mathcal{Z}} \; m_k(s) \quad \text{subject to} \quad \|s\|_{\mathcal{Z}} \le \delta_k.$$

3. **Objective Update:** Choose a new $J_k(z) \approx J(z)$. ← **ADAPTIVITY**

4. **Step Acceptance:** Compute

$$\varrho_k = \frac{J_k(z_k) - J_k(z_k + s_k)}{m_k(0) - m_k(s_k)}.$$

   If $\varrho_k \ge \eta \in (0,1)$, then $z_{k+1} = z_k + s_k$ else $z_{k+1} = z_k$.

5. **Trust Region Update:** Choose a new trust-region radius, δ_{k+1}.

EndWhile

Sandia National Laboratories

Inexact Gradients and Objective Functions

Kouri, Heinkenschloss, Ridzal, and van Bloemen Waanders

Inexact Gradients

There exists $c > 0$ independent of k such that

$$\|\nabla m_k(0) - \nabla J(z_k)\|_{\mathcal{Z}} \le c \min\{\|\nabla m_k(0)\|_{\mathcal{Z}}, \delta_k\}$$

(Carter 1989, Heinkenschloss and Vicente 2001).

Inexact Objective Functions

There exists $K > 0$, $\omega \in (0,1)$, and $\theta(z,s) \to 0$ as $r \to 0$ such that

$$|(J(z_k) - J(z_k + s_k)) - (J_k(z_k) - J_k(z_k + s_k))| \le K\theta(z_k, s_k)$$
$$\theta(z_k, s_k)^\omega \le \eta \min\{(m_k(0) - m_k(s_k)), r_k\}.$$

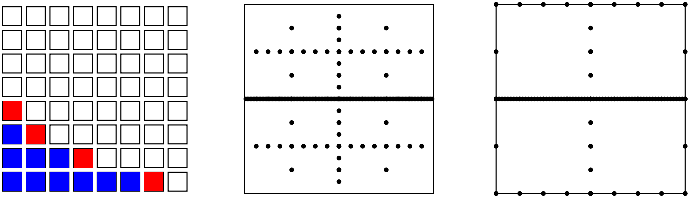
Here, $\eta > 0$ is tied to algorithmic parameters and $\lim_{k\to\infty} r_k = 0$.
(Carter 1989, Ziems and Ulbrich 2013).

- **Cannot** compute $J(z_k)$ and $\nabla J(z_k)$;
- Control *a posteriori* errors using **adaptive sparse grids**.

Sparse Grids and Adaptivity

Gerstner and Griebel 2003

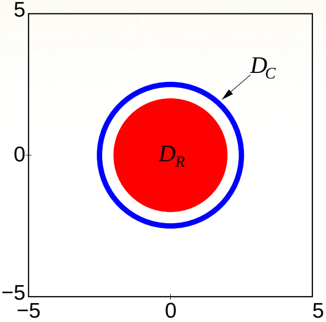
- **1D Operators:** For $k = 1, \ldots, M$, $\mathbb{E}_k^0 \equiv 0$ and
$$\Delta_k^i \equiv \mathbb{E}_k^i - \mathbb{E}_k^{i-1} \quad \text{where} \quad \mathbb{E}_k^i(g) \xrightarrow{i \to \infty} \int_{\Xi_k} \rho_k(\xi) g(\xi) \mathrm{d}\xi$$
- **Sparse-Grid Operator:** For an index set $\mathcal{I} \subset \mathbb{N}^M$,
$$\mathbb{E}_{\mathcal{I}} \equiv \sum_{\mathbf{i} \in \mathcal{I}} (\Delta_1^{i_1} \otimes \cdots \otimes \Delta_M^{i_M})$$
- **Admissibility:** $\mathbf{i} \in \mathcal{I}$ and $\mathbf{i} \geq \mathbf{j} \implies \mathbf{j} \in \mathcal{I}$
- **Error:** Given the index set $\mathcal{I} \subset \mathbb{N}^M$, the error is
$$\mathbb{E} - \mathbb{E}_{\mathcal{I}} = \sum_{\mathbf{i} \notin \mathcal{I}} (\Delta_1^{i_1} \otimes \cdots \otimes \Delta_M^{i_M})$$
- **Adaptivity:** Pick $\mathbf{i} \notin \mathcal{I}$ s.t. $\mathcal{I} \cup \{\mathbf{i}\}$ admissible and $\Delta_1^{i_1} \otimes \cdots \otimes \Delta_M^{i_M}$ “large”



Direct Field Acoustic Testing (DFAT)

Larkin and Whalen

- **Physical Domain:** $\Omega = (-5,5)^2$
- **Parameter Space:** $\Xi = [-\sqrt{3},\sqrt{3}]^M$
- **Probability Measure:**
  $\rho(\xi)\mathrm{d}\xi = \left(2\sqrt{3}\right)^{-M}\mathrm{d}\xi$
- **Stochastic Material:** $\epsilon(\xi,x)$
  KL expansion of Matérn covariance
- **Desired State:** $\theta = \frac{\pi}{4}$, $k = 10$
  $\bar{w}(x) = \exp\left(i\left((k\cos\theta)x_1 + (k\sin\theta)x_2\right)\right)$



Let $\alpha > 0$ and $\tau = 0.1$. Consider the optimal control problem

$$\min_{z\in L^2(D;\mathbb{C})} \frac{1}{2}\sigma\left[\int_{D_R}(u(z;\xi,x)-\bar{w}(x))\overline{(u(z;\xi,x)-\bar{w}(x))}\,\mathrm{d}x\right] + \frac{\alpha}{2}\int_{D_C} z(x)\overline{z(x)}\,\mathrm{d}x$$

where $u = u(z) \in L^2_\rho(\Xi; H^1(\Omega;\mathbb{C}))$ solves

$$-\Delta u(\xi,x) - k^2(1+\tau\epsilon(\xi,x))^2 u(\xi,x) = z(x) \qquad \forall\,(\xi,x)\in\Xi\times\Omega$$

$$\frac{\partial u}{\partial n}(\xi,x) = iku(\xi,x) \qquad \forall\,(\xi,x)\in\Xi\times\partial\Omega.$$

Results: Risk Neutral

$\alpha = 10^{-4}$

dim	Algorithm	PDE Solves	CP_{grad}	CP_{obj}
20	Grad. Adapt. TR	1,136,784	1,405	120,401
	Obj. Adapt. TR	122,331	1,509	2,933
40	Grad. Adapt. TR	16,327,120	1,445	1,804,001
	Obj. Adapt. TR	128,051	1,549	2,973

Table : Computational cost of the classical trust-region algorithm applied to the Helmholtz example with $M \in \{20, 40\}$ and $\alpha = 10^{-4}$.

Outline

Problem Formulation

Known Probability Distribution

- Risk-Neutral Optimization
- Risk-Averse Optimization

Unknown Probability Distribution

Conclusions

Sandia National Laboratories

Motivation - Control Uncertainty

Optimal control should be **"risk averse."** For example:

- Reduce **variance** or **deviation**

$$\mathbb{E}[(X-\mathbb{E}[X])^2] \qquad \text{or} \qquad \mathbb{E}[(X-\mathbb{E}[X])_+^q]^{\frac{1}{q}}$$

  e.g. *reduce uncertainty and variability in controlled system.*

- Control **rare events**, **tail probabilities**, or **quantiles**

$$\Pr[X \le t] \qquad \text{or} \qquad \mathrm{VaR}_\beta[X] = \inf\{\, t \in \mathbb{R} \,:\, \Pr[X \le t] \ge \beta \,\}$$

  e.g. *reduce failure regions and certify reliability.*

- Minimize over **quantiles**

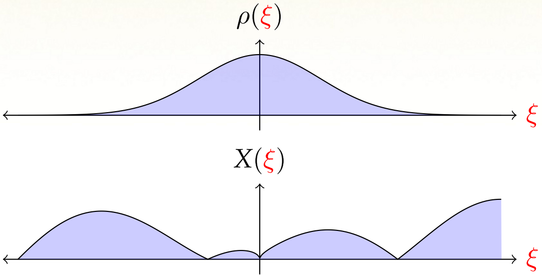
$$\mathrm{CVaR}_\beta[X] = \frac{1}{1-\beta}\int_{X \ge \mathrm{VaR}_\beta[X]} X(\xi)\rho(\xi)\,\mathrm{d}\xi = \mathbb{E}[X \,|\, X \ge \mathrm{VaR}_\beta[X]]$$

  e.g. *minimize over undesirable events.*

Risk Measures

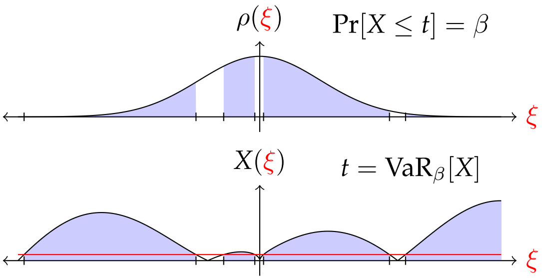
RISK NEUTRAL:

$\mathcal{R}(X) = \mathbb{E}[X]$



CONDITIONAL VALUE-AT-RISK:

$\mathcal{R}(X) = \mathrm{CVaR}_\beta[X]$



Sandia National Laboratories

Risk Measures

Shapiro, Dentcheva, Ruszczynski, Rockafellar, Uryasev, . . .

$\mathcal{R} : L^q \to \mathbb{R} \cup \{\infty\}$ is a **monetary** risk measure if for $X,\, X_1,\, X_2 \in L^q_\rho(\Xi)$

- **Monotonicity:** $X_1 \geq X_2$ a.e. $\implies \mathcal{R}(X_1) \geq \mathcal{R}(X_2)$
- **Translation Equivariance:** $\mathcal{R}(X+t) = \mathcal{R}(X) + t, \quad \forall t \in \mathbb{R}$

$\mathcal{R}$ is a **convex** risk measure if

- $\mathcal{R}$ is a monetary risk measure
- **Convexity:** $\mathcal{R}(tX_1 + (1-t)X_2) \leq t\mathcal{R}(X_1) + (1-t)\mathcal{R}(X_2), \quad \forall t \in [0,1]$

$\mathcal{R}$ is a **coherent** risk measure if

- $\mathcal{R}$ is a convex risk measure
- **Positive Homogeneity:** $\mathcal{R}(tX) = t\mathcal{R}(X), \quad \forall t > 0.$

Examples of coherent risk measures with $X \in= L^q_\rho(\Xi)$:

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

Sandia National Laboratories

Duality Theory of Risk Measures

By the Fenchel-Moreau Theorem, if $\mathcal{R}$ is a **convex** risk measure, then

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

where $\mathcal{R}^*$ is conjugate to $\mathcal{R}$, i.e., $\mathcal{R}^*(\vartheta) = \sup_{X \in \mathrm{dom}(\mathcal{R})} \{\mathbb{E}[\vartheta X] - \mathcal{R}(X)\}$.

Moreover, if $\mathcal{R}$ is a **coherent** risk measure, then

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

$\mathrm{dom}(\mathcal{R}^*)$ is the **risk envelope** $\implies$ related to the **ambiguity set** for unknown pdf.

Example (Conditional Value-at-Risk):

$$\mathcal{R}(X) = \mathrm{CVaR}_\beta[X] = \inf_t \left\{ t + (1-\beta)^{-1}\mathbb{E}[(X-t)_+] \right\} = \sup_{\vartheta \in \mathrm{dom}(\mathcal{R}^*)} \mathbb{E}[\vartheta X]$$

where $\mathrm{dom}(\mathcal{R}^*) = \left\{ \vartheta \in (L^q_\rho(\Xi))^* \ : \ \vartheta(\xi) \in \left[0, \frac{1}{1-\beta}\right] \ \rho\text{-a.e.}, \ \int_\Xi \vartheta(\xi)\rho(\xi)\,\mathrm{d}\xi = 1 \right\}.$

Sandia National Laboratories

The Risk Quadrangle

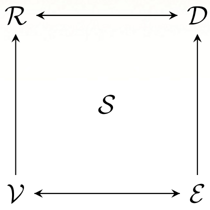
Optimization

Risk $\mathcal{R}$: Measures overall "hazard"

$$\mathcal{R}(X) = \mathbb{E}[X] + \mathcal{D}(X) = \min_t\{t + \mathcal{V}(X-t)\}.$$

Regret $\mathcal{V}$: Measures ones "displeasure"

$$\mathcal{V}(X) = \mathbb{E}[X] + \mathcal{E}(X).$$



Estimation

Deviation $\mathcal{D}$: Measures "non-constantancy"

$$\mathcal{D}(X) = \mathcal{R}(X) - \mathbb{E}[X] = \min_t\{\mathcal{E}(X-t)\}.$$

Error $\mathcal{E}$: Measures proximity to zero

$$\mathcal{E}(X) = \mathcal{V}(X) - \mathbb{E}[X].$$

Statistic: $\mathcal{S}(X) = \arg\min_t\{t + \mathcal{V}(X-t)\} = \arg\min_t\{\mathcal{E}(X-t)\}.$

Example: Given $\eta > 0$,

$$\mathcal{E}(X) = \eta\|X\|_{L^2_\rho(\Xi)} \implies \mathcal{V}(X) = \mathbb{E}[X] + \eta\|X\|_{L^2_\rho(\Xi)} \quad \text{and} \quad \mathcal{S}(X) = \mathbb{E}[X]$$

$$\implies \mathcal{D}(X) = \eta\|X - \mathbb{E}[X]\|_{L^2_\rho(\Xi)} = \eta\sigma(X)$$

$$\implies \mathcal{R}(X) = \mathbb{E}[X] + \eta\|X - \mathbb{E}[X]\|_{L^2_\rho(\Xi)} = \mathbb{E}[X] + \eta\sigma(X).$$

The Expectation Quadrangle

Let the **scalar regret function**, $v : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$, be convex, lsc, and satisfy

$$v(0) = 0, \quad \text{but} \quad v(x) > x \quad \forall x \neq 0.$$

We define the **scalar error function** as $e(x) = v(x) - x$.

- e is convex and lsc;
- $e(0) = 0$, but $e(x) > 0$ for all $x \neq 0$.

The corresponding regret and error measures are

$$\mathcal{V}(X) = \mathbb{E}[v(X)] \quad \text{and} \quad \mathcal{E}(X) = \mathbb{E}[e(X)], \quad \text{respectively.}$$

Example (Quantile-based Quadrangle): Given $0 < \beta < 1$, define

$$v(x) = \frac{1}{1-\beta}(x)_+ \quad \Longrightarrow \quad e(x) = \frac{\beta}{1-\beta}(x)_+ + (-x)_+$$

This gives rise to the quadrangle

$$\mathcal{R}(X) = \min_t \left\{ t + \frac{1}{1-\beta}\mathbb{E}[(X-t)_+] \right\} = \mathrm{CVaR}_\beta[X]$$

$$\mathcal{D}(X) = \mathrm{CVaR}_\beta[X - \mathbb{E}[X]] = \beta\text{-CVaR deviation of } X$$

$$\mathcal{E}(X) = \text{normalized Koenker-Bassett error}$$

$$\mathcal{S}(X) = \mathrm{VaR}_\beta[X] = \beta\text{-quantile of } X.$$

Complications with Risk Measures

- $\mathcal{V}$ (or v) are lsc $\implies$ $\mathcal{R}$ may not be differentiable.
- Derivative-based optimization algorithms may not apply.
- If v is not differentiable, quad. approx. may not converge.

Our Approach

- Smooth $\mathcal{V}$ or $\mathcal{E}$ to improve differentiability of objective function;
- Smoothing may help ensure convergence of quad. approx.;
- Use Newton-type algorithms to solve smoothed problem;
- Must quantify error committed by smoothing;
- Must perform continuation on smoothing parameter.

Example: Smoothed CVaR

Given $\varepsilon > 0$, we consider approximations of $(\cdot)^+$ given by

$$(x)^+_\varepsilon = \int_{-\infty}^{x} \mathbf{G}_\varepsilon(\tau)\,\mathrm{d}\tau \quad \text{where} \quad \mathbf{G}_\varepsilon(x) = \int_{-\infty}^{x} \frac{1}{\varepsilon}\delta\left(\frac{\tau}{\varepsilon}\right)\,\mathrm{d}\tau$$

and $\delta : \mathbb{R} \to \mathbb{R}$ satisfies

1. $\delta \in C(\mathbb{R})$ and let $0 < K < \infty$ s.t. $\sup_x |\delta(x)| \le K$;
2. $\delta(x) \ge 0 \;\forall\, x$ and $\int_{-\infty}^{\infty} \delta(x)\,\mathrm{d}x = 1$;
3. Either $\int_{-\infty}^{\infty} \delta(x)x\,\mathrm{d}x \le 0$ or $\int_{-\infty}^{0} \delta(x)|x|\,\mathrm{d}x = 0$;
4. $\mathrm{supp}(\delta)$ is connected.

Results:

- $x \mapsto (x)^+_\varepsilon$ is nondecreasing and convex;
- $x \mapsto (x)^+_\varepsilon$ is at least twice continuously differentiable;
- $-\varepsilon\Delta_2 \le (x)^+_\varepsilon - (x)^+ \le \varepsilon\Delta_1$ where

$$\Delta_1 = \int_{-\infty}^{0} \delta(x)|x|\,\mathrm{d}x \quad \text{and} \quad \Delta_2 = \max\left\{0, \int_{-\infty}^{\infty} \delta(x)x\,\mathrm{d}x\right\}$$

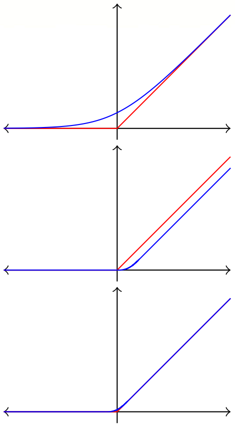
Smoothed Plus Function

Examples:

$$(x)^+_{\varepsilon,1} = x + \varepsilon \log\left(1 + \exp\left(\frac{-x}{\varepsilon}\right)\right)$$

$$(x)^+_{\varepsilon,2} = \begin{cases} 0 & \text{if } x \leq 0 \\ \left(\frac{x^3}{\varepsilon^2} - \frac{x^4}{2\varepsilon^3}\right) & \text{if } x \in (0, \varepsilon) \\ x - \frac{\varepsilon}{2} & \text{if } x \geq \varepsilon \end{cases}$$

$$(x)^+_{\varepsilon,3} = \left(x + \frac{\varepsilon}{2}\right)^+_{\varepsilon,2}$$



Smoothed CVaR

$F_\varepsilon^\beta(t,X) = t + \frac{1}{1-\beta}\mathbb{E}[(X-t)_\varepsilon^+]$ and $\mathcal{R}_\varepsilon^\beta[X] = \inf\{F_\varepsilon^\beta(t,X) \,:\, t\in\mathbb{R}\}$

Results:

- There exists $c>0$ s.t. $\left|\mathcal{R}_\varepsilon^\beta[X] - \mathrm{CVaR}_\beta[X]\right| \le \frac{c}{1-\beta}\varepsilon$ for all $X\in L_\rho^1(\Xi)$.
- Smoothed CVaR, $\mathcal{R}_\varepsilon^\beta$, is a convex risk measure.
- $X\mapsto F_\varepsilon^\beta(t,X) : L_\rho^1(\Xi)\to\mathbb{R}$ is Hadamard differentiable.
- $X\mapsto F_\varepsilon^\beta(t,X) : L_\rho^2(\Xi)\to\mathbb{R}$ is twice continuously Fréchet differentiable.
- $t\mapsto F_\varepsilon^\beta(t,X) : \mathbb{R}\to\mathbb{R}$ is twice continuously differentiable.

PDE-Optimization Problem

$$J_\varepsilon(t,z) \equiv t + \frac{1}{1-\beta}\mathbb{E}\left[(f(U(z),z),\cdot))_\varepsilon^+\right]$$

Results:

- **Differentiability:** $J_\varepsilon(t,z)$ is Hadamard differentiable w.r.t. z and continuously differentiable w.r.t. t.
- **Differentiability:** If $u=u(z)\in L_\rho^p(\Xi;\mathcal{U})$ with $p\ge 4$ then $J_\varepsilon(t,z)$ is twice continuously Fréchet differentiable.
- **Convergence Rate:** Suppose $(t_\varepsilon, z_\varepsilon)$ is a minimizer of $J_\varepsilon(t,z)$ and (t^*,z^*) is a minimizer of $J(t,z)$. Then, $(|t^*-t_\varepsilon|^2 + \|z^*-z_\varepsilon\|_{\mathcal{Z}}^2)^{\frac{1}{2}} \le C\varepsilon^{\frac{1}{2}}$.

Sandia National Laboratories

Optimal Control of 1D Elliptic Equation

Let $\alpha = 10$, $\Omega_o = \Omega_c = \Omega = (-1, 1)$, and $\bar{u} \equiv 1$ and consider

$$\underset{z \in L^2(-1,1)}{\text{minimize}} \quad J(z) = \frac{1}{2}\text{CVaR}_\beta\left[\int_{-1}^{1} (u(\cdot, x; z) - 1)^2 \, \mathrm{d}x\right] + \frac{\alpha}{2}\int_{-1}^{1} z(x)^2 \, \mathrm{d}x$$

where $u = u(z) \in L^2_\rho(\Xi; H^1_0(0, 1))$ solves the weak form of

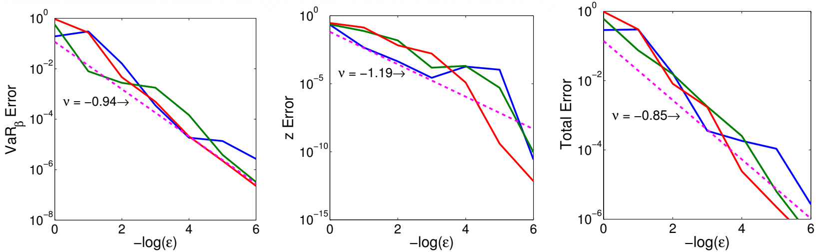
$$\begin{aligned} -\partial_x \left(\epsilon(\xi, x)\partial_x u(\xi, x)\right) &= f(\xi, x) + z(x) && (\xi, x) \in \Xi \times \Omega, \\ u(\xi, -1) = 0, \quad u(\xi, 1) &= 0 && \xi \in \Xi. \end{aligned}$$

$\Xi = [-0.1, 0.1] \times [-0.5, 0.5]$ is endowed with the uniform density $\rho(\xi) \equiv 5$, and the random field coefficients are

$$\epsilon(\xi, x) = 0.1\chi_{(-1,\xi_1)} + 10\chi_{(\xi_1,1)}, \quad \text{and} \quad f(\xi, x) = \exp(-(x - \xi_2)^2).$$

Primal Results

Sample Approximation: Level 8 Gauss-Patterson sparse grids.



Optimization Algorithm: Trust region with truncated CG.

		$\log(\varepsilon)$						
		0	-1	-2	-3	-4	-5	-6
	0.05	12(11)	18(5)	15(4)	6(11)	4(16)	24(3)	22(2)
β	0.5	11(9)	4(22)	4(21)	6(13)	16(4)	5(8)	5(8)
	0.95	14(12)	14(11)	5(17)	5(19)	3(9)	4(7)	4(5)

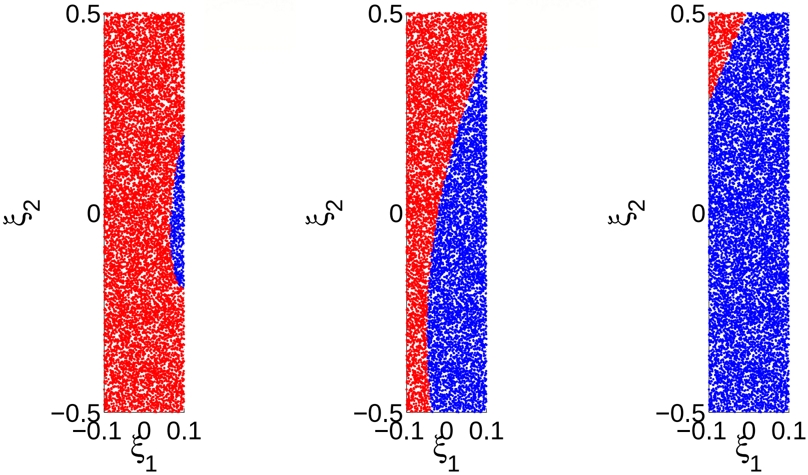
Continuation: # TR iterations (average # CG iterations)

Sandia National Laboratories

Dual Results

Sample Approximation: Monte Carlo with 10,000 samples.

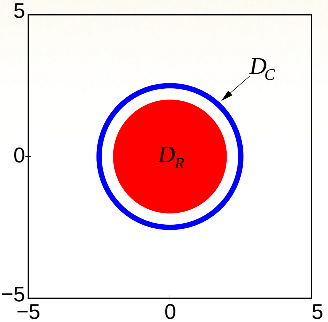
$\beta = 0.05$ $\beta = 0.5$ $\beta = 0.95$



$\vartheta^* = 0$ and $\vartheta^* = (1-\beta)^{-1}$

Direct Field Acoustic Testing (DFAT)

- **Physical Domain:** $D = (-5,5)^2$
- **Parameter Space:** $\Xi = [-\sqrt{3},\sqrt{3}]^M$
- **Probability Measure:**
  $\rho(\xi)\mathrm{d}\xi = \left(2\sqrt{3}\right)^{-M}\mathrm{d}\xi$
- **Stochastic Material:** $\epsilon(\xi,x)$
  KL expansion of Matérn covariance
- **Desired State:** $\theta = \frac{\pi}{4}$, $k = 10$
  $\bar{w}(x) = \exp\left(i\left((k\cos\theta)x_1 + (k\sin\theta)x_2\right)\right)$



Let $\alpha > 0$ and $\vartheta = 0.1$. Consider the optimal control problem

$$\min_{z\in L^2(D;\mathbb{C})} \frac{1}{2}\sigma\left[\int_{D_R}(u(z;\xi,x)-\bar{w}(x))\overline{(u(z;\xi,x)-\bar{w}(x))}\,\mathrm{d}x\right] + \frac{\alpha}{2}\int_{D_c} z(x)\overline{z(x)}\,\mathrm{d}x$$

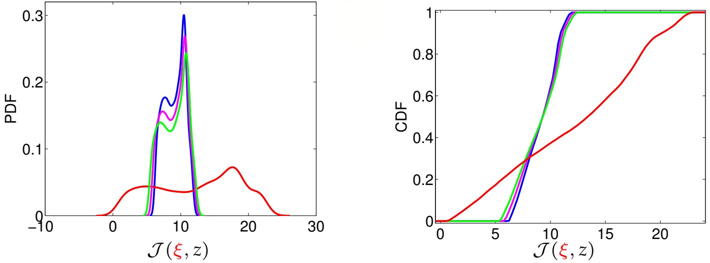
where $u = u(z) \in L^2_\rho(\Xi;H^1(D;\mathbb{C}))$ solves

$$\begin{aligned} -\Delta u(\xi,x) - k^2(1+\vartheta\epsilon(\xi,x))^2 u(\xi,x) &= z(x) && \forall\,(\xi,x)\in\Xi\times D \\ \frac{\partial u}{\partial n}(\xi,x) &= iku(\xi,x) && \forall\,(\xi,x)\in\Xi\times\partial D. \end{aligned}$$

Sandia National Laboratories

Results: $M = 6$, $\gamma = 5$, and $\alpha = 10^{-4}$

$$\mathcal{J}(\xi, z) = \int_{D_R} (u(z;\xi,x) - \bar{w}(x))\overline{(u(z;\xi,x) - \bar{w}(x))}\,\mathrm{d}x.$$



- Mean value: $\xi \leftarrow \mathbb{E}[\xi]$
- Mean plus CVaR: $\sigma[X] = \frac{1}{2}\mathbb{E}[X] + \frac{1}{2}\mathrm{CVaR}_{0.1}[X]$
- Risk neutral: $\sigma[X] = \mathbb{E}[X]$
- CVaR: $\sigma[X] = \mathrm{CVaR}_{0.1}[X]$

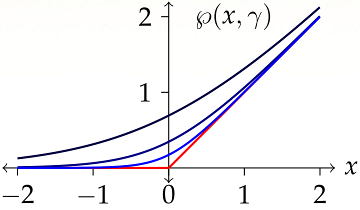
Mean plus CVaR Value-at-Risk: $t = 6.59073$

CVaR Value-at-Risk: $t = 6.90629$

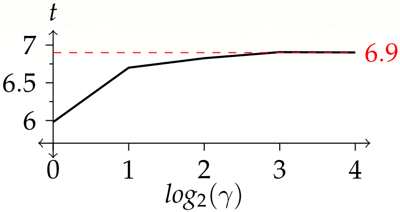
Sandia National Laboratories

Results: The Effect of Smoothing - CVaR

Smoothed Plus Function



VaR Approximation



γ	$\lVert z\rVert_{\mathcal{Z}}$	Abs. Err.	Rate	t	Abs. Err.	Rate
1	37.8369	-	-	5.9792	-	-
2	37.0495	1.2177	-	6.7004	0.7212	-
4	36.7416	0.7111	0.7760	6.8258	0.1254	2.5239
8	36.6653	0.4396	0.6939	6.9066	0.0808	0.6341

Theoretical convergence rate is $\frac{1}{2}$ (Kouri and Surowiec).

Sandia National Laboratories

Outline

Problem Formulation

Known Probability Distribution

Risk-Neutral Optimization

Risk-Averse Optimization

Unknown Probability Distribution

Conclusions

Distributionally Robust PDE-Optimization

Recall: $(\Xi, \mathcal{F})$ is a measurable space and prob. measure is *unknown*.

- $\mathcal{M}$ denotes the Banach space of regular Borel measures on $\mathcal{F}$.
- $\mathcal{M}^+ \subset \mathcal{M}$ is the set of positive measures, i.e.,

$$\mu \in \mathcal{M}^+ \quad \Longrightarrow \quad \mu(V) \geq 0 \quad \forall\, V \in \mathcal{F}.$$

- **Ambiguity Set:** $\mathcal{A} \subset \mathcal{M}$ defined by data. For example:
  - **Moment Matching:** Given generalized moment data $m_1, \ldots, m_N$,

$$\mathcal{A} = \left\{ P \in \mathcal{M}^+ \;:\; P(\Xi) = 1,\; \int_\Xi \psi_i(\xi)\,\mathrm{d}P(\xi) = m_i,\; i = 1, \ldots, N \right\}.$$

  - **Φ-Divergence (e.g., Kullback-Leibler):** Given an estimated prob. measure P_0 and $\epsilon > 0$,

$$\mathcal{A} = \left\{ P \in \mathcal{M}^+ \;:\; P(\Xi) = 1,\; D_\Phi(P, P_0) \leq \epsilon \right\}.$$

- **Distributionally-robust (a.k.a. data-driven) optimization problem:**

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} \sup_{P \in \mathcal{A}} \int_\Xi f((U(z))(\xi), z, \xi)\,\mathrm{d}P(\xi).$$

Sandia National Laboratories

Measure Discretization

General Approach:

1. Let $\{\varphi_i\}_{i=1}^n$ be a partition of unity and $\mu \in \mathcal{M}$ be any measure.
2. Define the "localized" measures

$$\mu_i(V) = \int_V \varphi_i(\xi)\, \mathrm{d}\mu(\xi).$$

3. Note $\mu(\Xi) = \mu_1(\Xi) + \ldots + \mu_n(\Xi)$.
4. Define the projection operators $\Pi_n : C(\Xi) \to \text{span}\{\varphi_1, \ldots, \varphi_n\}$ as

$$\Pi_n y = \sum_{i=1}^n \mu_i(\Xi)^{-1} \int_\Xi y(\xi)\, \mathrm{d}\mu_i(\xi)\ \varphi_i \quad \forall y \in C(\Xi) \tag{3}$$

   and $\Lambda_n : \mathcal{M} \to \text{span}\{\mu_1, \ldots, \mu_n\}$ as

$$\Lambda_n \nu = \sum_{i=1}^n \mu_i(\Xi)^{-1} \int_\Xi \varphi_i(\xi)\, \mathrm{d}\nu(\xi)\ \mu_i \quad \forall \nu \in \mathcal{N}, \tag{4}$$

5. **Lemma:** Π_n is the adjoint of Λ_n.
6. **Lemma:** Λ_n is invariant on the space of prob. measures.

Sandia National Laboratories

Measure Discretization

Piecewise Constants:

1. Let $\{V_i\}_{i=1}^n$ be a tesselation of Ξ and define $\varphi_i = \chi_{V_i}$.
2. The "localized" measures are

$$\mu_i(V) = \mu(V \cap V_i).$$

3. The projection operator $\Pi_n : C(\Xi) \to \text{span}\{\varphi_1, \dots, \varphi_n\}$ is

$$\Pi_n y = \sum_{i=1}^n \mu(V_i)^{-1} \int_{V_i} y(\xi)\,\mathrm{d}\mu(\xi)\ \chi_{V_i} \quad \forall\, y \in C(\Xi) \tag{5}$$

and $\Lambda_n : \mathcal{M} \to \text{span}\{\mu_1, \dots, \mu_n\}$ is

$$\Lambda_n \nu = \sum_{i=1}^n \mu(V_i)^{-1} \nu(V_i) \mu_i \quad \forall\, \nu \in \mathcal{N}, \tag{6}$$

4. **Theorem:** Suppose V_i are convex, bounded, and Lipschitz, and $\mu \in \mathcal{M}$. Then $\exists c > 0$ only depending on M such that

$$\|\nu - \Lambda_n \nu\|_{W^{1,\infty}(\Xi)^*} \leq c \sum_{i=1}^n \left(1 + \frac{|\mu|(V_i)}{|\mu(V_i)|}\right) |\nu|(V_i)\ \mathrm{diam}(V_i).$$

Sandia National Laboratories

Example

Suppose $\Xi = [0,1]$ and P has pdf

$$\text{pdf}(\xi) = \frac{\beta}{1-e^{-\beta}} e^{-\beta\xi} \quad \text{for} \quad \beta > 0.$$

Approx. P using piecewise constant projection and μ set to the uniform prob. measure:

$$\text{approx-pdf}(\xi) = \sum_{i=1}^{n} \frac{(e^{-\beta a_{i-1}} - e^{-\beta a_i})}{(1-e^{-\beta})(a_i - a_{i-1})} \chi_{[a_{i-1},a_i]}(\xi).$$

β	n	Error	Sum W. Diam.	Max. Diam.	Max. W. Diam.
1	10	3.592×10^{-2}	1.438×10^{-1}	2.518×10^{-1}	5.899×10^{-2}
	100	3.740×10^{-3}	1.496×10^{-2}	4.269×10^{-2}	1.471×10^{-3}
	1000	3.751×10^{-4}	1.501×10^{-3}	6.089×10^{-3}	2.733×10^{-5}
	10000	3.750×10^{-5}	1.500×10^{-4}	7.955×10^{-4}	4.404×10^{-7}
10	10	2.282×10^{-1}	1.304×10^{-1}	7.572×10^{-1}	1.010×10^{-1}
	100	3.053×10^{-2}	1.451×10^{-2}	5.328×10^{-1}	8.191×10^{-3}
	1000	3.551×10^{-3}	1.502×10^{-3}	3.133×10^{-1}	5.424×10^{-4}
	10000	3.763×10^{-4}	1.517×10^{-4}	1.300×10^{-1}	2.710×10^{-5}
100	10	3.076×10^{-1}	1.226×10^{-1}	9.758×10^{-1}	1.194×10^{-1}
	100	4.128×10^{-2}	1.327×10^{-2}	9.531×10^{-1}	1.261×10^{-2}
	1000	5.022×10^{-3}	1.348×10^{-3}	9.301×10^{-1}	1.247×10^{-3}
	10000	5.899×10^{-4}	1.360×10^{-4}	9.072×10^{-1}	1.224×10^{-4}

Optimization Algorithms

$$J(z) = \sup_{P \in \mathcal{A}} \int_{\Xi} f((U(z))(\xi), z, \xi)\, \mathrm{d}P(\xi)$$ **may not be differentiable!**

- $\mathcal{Z}$ Hilbert + $\nabla^2 f(U(z), z, \cdot)$ bounded $\implies$ $J(z)$ is *proximally subdiff.*
  - **Analytic Definition:** $\zeta \in \mathcal{Z}^*$ is a *proximal subgradient* if $\exists\, \sigma$ and η such that $\forall\, y \in \mathcal{Z}$ with $\|z - y\|_{\mathcal{Z}} \leq \eta$,

    $$J(y) \geq J(z) + \langle \zeta, y - z \rangle_{\mathcal{Z}^*, \mathcal{Z}} - \sigma \|y - z\|_{\mathcal{Z}}^2.$$

  - **Geometric Definition:** J is *locally supported* by a quadratic.
  - **Example:** $-|x|$ is not proximally subdifferentiable at $x = 0$.
  - **Optimality:** If $z \in \mathcal{Z}$ minimize J then 0 is a proximal subgradient.
- Cannot use derivative-based optimization algorithms.
- Subgradient descent and bundle methods converge *sublinearly*.
- **Expensive PDEs** $\implies$ **Need rapid optimization algorithms.**

Sandia National Laboratories

Outline

Problem Formulation

Known Probability Distribution

- Risk-Neutral Optimization
- Risk-Averse Optimization

Unknown Probability Distribution

Conclusions

Conclusions:

- **Risk Neutral:**
  - Can efficiently solve using adaptive sparse grids and trust regions.
- **Risk Averse:**
  - Risk measures often not differentiable;
  - Define smooth risk measures using the risk quadrangle;
  - Can use Newton's method/quad. and can prove error bounds.
- **Unknown Distribution:**
  - Incorporate data into distributionally-robust opt. formulation;
  - Objective func. not differentiable;
  - Nonsmooth optimization algorithms converge slowly.

Future Work:

- **Risk measures:** Develop error indicators and use locally adaptive sparse grids with trust-region algorithm.
- **Unknown distribution:** Develop opt. algorithm that exploits structures inherent to PDE-constrained problems.
- Incorporate **(buffered) probabilistic objectives and constraints** to control *tail-probabilities* and *rare events* (Rockafellar, Uryasev, Royset, Shapiro, Henrion, Kibzun, ...)