A Data-Driven Approach to PDE-Cons SAND2016-4352PE
under Uncertainty

Drew Kouri

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

May 14, 2016

West Coast Optimization Meeting, University of Washington, Seattle, WA.

Funding support: ASC, DARPA.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Outline

Problem Formulation

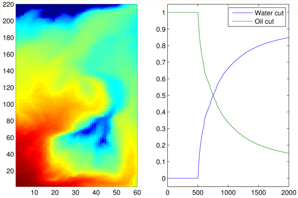
Known Probability Distribution

Unknown Probability Distribution

Conclusions

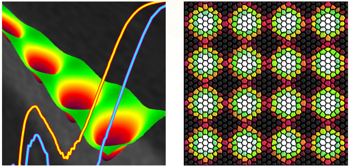
Motivation

Reservoir Optimization



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

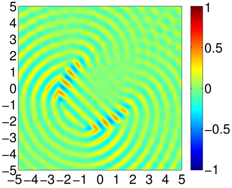
Superconductor Vortex Pinning

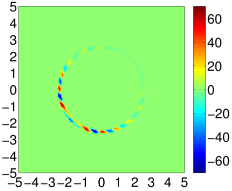


Courtesy Argonne National Laboratory

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

Direct Field Acoustic Testing





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

Sandia National Laboratories

Optimization of PDEs with Uncertain Inputs

Optimal Control: Given $\alpha > 0$, $D_o \subseteq D$, $D_c \subseteq D$, and $w \in L^2(D_o)$.

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

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

$$\begin{aligned} -\nabla\cdot(\epsilon(\xi)\nabla u(\xi)) + N(u(\xi), \xi) &= \chi_{D_c} z, && \text{in } D, \text{ a.s.} \\ u(\xi) &= g(\xi), && \text{on } \partial D, \text{ a.s.} \end{aligned}$$

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

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

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

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

Sandia National Laboratories

General PDE-Optimization under Uncertainty

$(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space and $\xi : \Omega \to \Xi$ is a random variable. Consider

$$\min_{z \in Z_{\mathrm{ad}}} J(z) = \mathcal{R}(F(z))$$

where $F(z) := f(S(z)(\xi), z, \xi))$ and $S(z)(\xi) = u : \Omega \to U$ solves the PDE

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

Assumptions on PDE Solution Map $S(z)$:

- U is a reflexive Banach space and Z is a Hilbert spaces.
- For each $z \in Z_{\mathrm{ad}}$, $e(u, z, \xi) = 0$ is well posed, i.e.,
  - $\exists!\, S(z)(\xi) : \Omega \to U$ such that $e(S(z), z, \xi) = 0$ for all z;
  - $\exists\, 0 < c(\xi) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ independent of z such that

$$\|S(z)(\xi)\|_U \leq c(\xi)(\|z\|_Z + 1).$$

- The map $S : Z \to L^p(\Omega, \mathcal{F}, \mathbb{P}; U)$ is completely continuous, i.e.,

$$z_n \rightharpoonup z \text{ in } Z \quad \Longrightarrow \quad S(z_n) \to S(z) \text{ in } L^p(\Omega, \mathcal{F}, \mathbb{P}; U).$$

Not satisfied for topology optimization problem!

Sandia National Laboratories

Existence of Minimizers

Assumptions: $F : Z_{\text{ad}} \to L^q(\Omega, \mathcal{F}, \mathbb{P})$;

- $F(z)$ has the lower semicontinuity property that if $z_n \rightharpoonup z$ then

$$\liminf_{n\to\infty} \mathbb{E}[\vartheta F(z_n)] \geq \mathbb{E}[\vartheta F(z)]$$

  for all $\vartheta \in (L^q(\Omega, \mathcal{F}, \mathbb{P}))^*$ satisfying $\vartheta \geq 0$ $\mathbb{P}$-a.e.;
- Z_{ad} is convex, closed and bounded – or –
  $Z_{\text{ad}} = Z$ and $F(z)$ has the coercivity propery that
  $\exists\, r > 0$ and coercive $\varphi : Z \to \mathbb{R} \cup \{+\infty\}$, such that

$$\|z\|_Z \geq r \quad \Longrightarrow \quad F(z) \geq \varphi(z) \text{ a.s.}$$

Assumptions: $\mathcal{R} : L^q(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R} \cup \{+\infty\}$

- $\mathcal{R}$ is convex and lower semicontinuous
- $\mathcal{R}$ satisfies $\mathcal{R}(C) = C$ for all constants C;
- $\mathcal{R}$ is monotonic, i.e., if $X \geq X'$ a.s., then $\mathcal{R}(X) \geq \mathcal{R}(X')$.

Result: There exists a minimizer of $J(z) = \mathcal{R}(F(z))$ in Z_{ad}.

Numerical Cost Surrogates, $\mathcal{R}$

$F(z) = f(S(z)(\xi), z, \xi)$ is a random variable with q finite moments.
No ordering on $L^q(\Omega, \mathcal{F}, \mathbb{P}) \implies$ cannot minimize $F(z)$!

- **Traditional Stochastic Programming:** Minimize $F(z)$ *on average* (called risk neutral),
$$\mathcal{R}(F(z)) = \mathbb{E}[F(z)].$$
- **Risk-Averse Stochastic Programming:** Incorporate *risk preferences* in $\mathcal{R}$; for example,
$$\mathcal{R}(F(z)) = \mathbb{E}[F(z)] + c\mathbb{E}[(F(z) - \mathbb{E}[F(z)])_+^q]^{1/q}.$$
- **Probabilistic Optimization:** Minimize the probability that $F(z)$ exceeds a threshold τ,
$$\mathcal{R}(F(z)) = \mathbb{P}(F(z) > \tau).$$
- **Ambiguous Stochastic Programming:** $\mathfrak{A}$ is an admissible set of probability distributions and
$$\mathcal{R}(F(z)) = \sup_{P \in \mathfrak{A}} \mathbb{E}_P[F(z)].$$

Known v.s. Unknown Probability Distribution

Traditional Stochastic Programming: Minimize $F(z)$ on average,

$$\min_{z \in Z_{\mathrm{ad}}} \mathbb{E}_{\mathbb{P}}[F(z))].$$

Known Probability Distribution:

- $\xi : \Omega \to \Xi \subseteq \mathbb{R}^M$ has **known** Lebesgue density $\rho : \Xi \to [0, \infty)$.
- Analysis in $L^q_\rho(\Xi)$ and objective function is **smooth** if F is.
- Enables polynomial approx. including PC and quadrature.

Unknown Probability Distribution:

- $\mathbb{P} \circ \xi^{-1}$ is **unknown** or is estim. from **incomplete**/**contaminated** data.
- Must determine minimizers that are **robust** to unknown distribution.
- Formulate optimization problem as the min-max problem

$$\min_{z \in Z_{\mathrm{ad}}} \sup_{P \in \mathfrak{A}} \mathbb{E}_P[F(z)].$$

- Numerical solution $\implies$ Must discretize the probability measures $P \in \mathfrak{A}$.

Outline

Choosing a Risk Measure

Controlling Uncertainty and the Risk Quadrangle

- Reduce **variability** of optimized system:

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

- Control **rare events**, reduce **failure regions**, and certify **reliability**:

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

- Minimize over **undesirable events**:

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

Opt. Under Uncertainty

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

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

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

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

$$\begin{tikzcd}
\mathcal{R} \arrow[rr, leftrightarrow] \arrow[drr, phantom, "\mathcal{S}"] & & \mathcal{D} \\
\mathcal{V} \arrow[u] \arrow[rr, leftrightarrow] & & \mathcal{E} \arrow[u]
\end{tikzcd}$$

Statistical Estimation

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

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

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

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

Classification of Risk Measures

Artzner, Delbaen, Eber, Heath, Shapiro, Rockafellar, Uryasev, ...

$\mathcal{R} : L^q_\rho(\Xi) \to \mathbb{R} \cup \{\infty\}$ is **coherent** if

- **Convexity:** $\mathcal{R}(tX + (1-t)X') \le t\mathcal{R}(X) + (1-t)\mathcal{R}(X'), \quad \forall t \in [0,1]$
- **Monotonicity:** $X \ge X'$ a.e. $\implies \mathcal{R}(X) \ge \mathcal{R}(X')$
- **Translation Equivariance:** $\mathcal{R}(X+t) = \mathcal{R}(X) + t, \quad \forall t \in \mathbb{R}$
- **Positive Homogeneity:** $\mathcal{R}(tX) = t\mathcal{R}(X), \quad \forall t > 0$

$\mathcal{R}$ is **law invariant** if

$$\mathbb{P}(X \le t) = \mathbb{P}(X' \le t) \ \ \forall t \in \mathbb{R} \quad \implies \quad \mathcal{R}(X) = \mathcal{R}(X')$$

Examples of **law-invariant coherent** risk measures with $X \in L^q_\rho(\Xi)$:

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

Coherent Risk As Distributionally Robust

- If $\mathcal{R}$ is **convex** and lsc, then
$$\mathcal{R}(X) = \sup\left\{\mathbb{E}[\vartheta X] - \mathcal{R}^*(\vartheta) \;:\; \vartheta \in \mathrm{dom}(\mathcal{R}^*)\right\}.$$
- If $\mathcal{R}$ is **translation equivariant** and **monotonic**, then
$$\mathrm{dom}(\mathcal{R}^*) \subseteq \left\{\vartheta \in (L^q_\rho(\Xi))^* \;:\; \mathbb{E}[\vartheta] = 1,\; \vartheta \geq 0 \; \rho\text{-a.e.}\right\}$$
- If $\mathcal{R}$ is **positive homogeneous**, then
$$\mathcal{R}(X) = \sup_{\vartheta \in \mathrm{dom}(\mathcal{R}^*)} \mathbb{E}[\vartheta X].$$

Coherent risk $\implies$ **disitributionally robust** with $\mathfrak{A} = \mathrm{dom}(\mathcal{R}^*)$.

Example (Conditional Value-at-Risk):

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

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

Outline

Problem Formulation

Known Probability Distribution

Unknown Probability Distribution

Conclusions

Sandia National Laboratories

Distributionally Robust PDE-Optimization

Žáčková, Rogosinksy, Shapiro, Gaivoronski, Kuhn, ...

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

- $\mathfrak{M}$ is the Banach space of regular Borel measures on $\mathcal{B}$, i.e.,

$$C(\Xi)^* \cong \mathfrak{M}.$$

- $\mathfrak{M}^+ \subset \mathfrak{M}$ is the set of positive measures, i.e.,

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

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

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

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

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

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

$$\min_{z \in Z_{\mathrm{ad}}} \sup_{P \in \mathfrak{A}} \int_\Xi f(S(z)(\xi), z, \xi)\, \mathrm{d}P(\xi).$$

Example: Φ-Divergence

Ben-Tal, Teboulle, Bayraksan, Love, Shapiro, . . .

Supppose

(i) A **nominal** probability measure P_0 is given,

(ii) The random variable $X \in L^q(\Xi, \mathcal{B}, P_0)$, and

(iii) $\Phi : \mathbb{R} \to [0, \infty]$ is convex lower semicontinuous satisfying

$$\Phi(1) = 0 \quad \text{and} \quad \Phi(x) = \infty \quad \forall\, x < 0.$$

Define, for fixed $\epsilon > 0$,

$$\mathfrak{A} = \{\vartheta \in (L^q(\Xi, \mathcal{B}, P_0))^* \;:\; \mathbb{E}_{P_0}[\vartheta] = 1,\ \vartheta \geq 0,\ \mathbb{E}_{P_0}[\Phi(\vartheta)] \leq \epsilon\}\,.$$

Then $\quad \mathcal{R}(X) = \sup_{\vartheta \in \mathfrak{A}} \mathbb{E}_{P_0}[\vartheta X] = \inf_{\lambda \geq 0,\, \mu} \{\lambda\epsilon + \mu + \mathbb{E}_{P_0}[(\lambda\Phi)^*(X - \mu)]\}$

is a **law-invariant coherent** risk measure!

Example (Kullback-Leibler Divergence): $\Phi(x) = x\ln(x) - x + 1,\, x \geq 0$

$$\mathcal{R}(X) = \inf_{\lambda > 0} \left\{ \lambda c + \lambda \ln \mathbb{E}_{P_0}\left[e^{X/\lambda}\right] \right\}.$$

Measure Approximation

For General Ambiguity Sets, $\mathfrak{A}$

Approach:

1. Let $\{\varphi_i\}_{i=1}^n$ be a partition of unity on Ξ and $\mu \in \mathfrak{M}$ 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 $\Pi_n^\mu : C(\Xi) \to \text{span}\{\varphi_1, \ldots, \varphi_n\}$ as
$$\Pi_n^\mu y = \sum_{i=1}^n \mu_i(\Xi)^{-1} \int_\Xi y(\xi)\,\mathrm{d}\mu_i(\xi)\;\varphi_i \quad \forall\, y \in C(\Xi)$$
and $\Lambda_n^\mu : \mathfrak{M} \to \text{span}\{\mu_1, \ldots, \mu_n\}$ as
$$\Lambda_n^\mu \nu = \sum_{i=1}^n \mu_i(\Xi)^{-1} \int_\Xi \varphi_i(\xi)\,\mathrm{d}\nu(\xi)\;\mu_i \quad \forall\, \nu \in \mathfrak{N},$$

Sandia National Laboratories

Approximation Properties

- **Lemma:** If $\mu \in \mathfrak{M}^+$ is σ-finite, then $\Lambda_n^\mu \nu$ is absolutely continuous with respect to μ with density

$$f_n^\mu[\nu] = \sum_{i=1}^n \mu_i(\Xi)^{-1} \int_\Xi \varphi_i(\xi)\,\mathrm{d}\nu(\xi)\varphi_i \quad \text{for any } \nu \in \mathfrak{M}.$$

- **Lemma:** Λ_n^μ is invariant on the space of probability measures.
- **Lemma:** Π_n^μ is the adjoint of Λ_n^μ.
- **Theorem:** Let $V_i = \mathrm{supp}(\varphi_i)$ and $\|\cdot\|_{u,V_i}$ denote the uniform norm on V_i. Then, there exists $c_i > 0$ such that

$$|\langle \nu - \Lambda_n^\mu \nu, y\rangle_{\mathfrak{M},C(\Xi)}| \leq \sum_{i=1}^n c_i \left\{ \int_\Xi \sqrt{\varphi_i(\xi)}\,\mathrm{d}|\nu|(\xi) \right\} \inf_{\bar{y}_i \in \mathbb{R}} \|\sqrt{\varphi_i}(y - \bar{y}_i)\|_{u,V_i}.$$

Measure Approximation

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 $\Pi_n^\mu : C(\Xi) \to \mathrm{span}\{\varphi_1, \ldots, \varphi_n\}$ is

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

and $\Lambda_n^\mu : \mathfrak{M} \to \mathrm{span}\{\mu_1, \ldots, \mu_n\}$ is

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

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

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

Example — Voronoi Tesselation

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

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

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}

Approximation and Optimization Algorithms

Given an arbitrary $\mu \in \mathfrak{M}^+$ with $\mu(\Xi) = 1$, we approximate

$$J(z) = \sup_{P\in\mathfrak{A}} \int_\Xi f(S(z)(\xi), z, \xi)\,\mathrm{d}P(\xi)$$

using our measure discretization, i.e.,

$$J_n(z) = \sup_{p\in\mathfrak{A}_n} \sum_{i=1}^n \frac{p_i}{\mu_i(\Xi)} \int_\Xi f(S(z)(\xi), z, \xi)\,\mathrm{d}\mu_i(\xi), \quad \mathfrak{A}_n = \left\{ p \in \mathbb{R}^n \,:\, \sum_{i=1}^n \frac{p_i}{\mu_i(\Xi)}\mu_i \in \mathfrak{A} \right\}.$$

- **Theorem (Piecewise Constants):** If $\xi \mapsto f(S(z)(\xi), z, \xi) \in W^{1,\infty}(\Xi)$ and z_n minimizes J_n defined on a family of tesselations $\{V_{ni}\}_{i=1}^n$ satisfying

  $$\lim_{n\to\infty} \sup_{P\in\mathfrak{A}} \sum_{i=1}^n P(V_{ni})\mathrm{diam}(V_{ni}) = 0.$$

  Then, z_n has a w-converging subsequence and the w-limit minimizes J.
- J and J_n may **not** be differentiable!
  - J and J_n are *Fréchet subdifferentiable.*
  - Compute value and subgradient using linear/convex optimization.
  - **Cannot** use derivative-based optimization algorithms.
  - Subgradient descent and bundle methods converge *sublinearly*.
- **Expensive PDEs $\Longrightarrow$ Need rapid optimization algorithms.**

Example — Moment Matching

Let $\psi_i : \Xi \to \mathbb{R}$ be $\mathcal{B}$-measurable functions and $m_i \in \mathbb{R}$ for $i = 1, \ldots, N$

$$\mathfrak{A} = \left\{ P \in \mathfrak{M}^+ \;:\; P(\Xi) = 1, \quad \begin{array}{l} \int_\Xi \psi_i(\xi)\,\mathrm{d}P(\xi) = m_i,\; i = 1, \ldots, N_e \\ \int_\Xi \psi_i(\xi)\,\mathrm{d}P(\xi) \le m_i,\; i = N_e + 1, \ldots, N \end{array} \right\}.$$

Theorem (Rogosinsky): If $\mathfrak{A} \neq \emptyset$, then for each $z \in Z$ there exists ξ_i and $p_i \ge 0$ with $p_1 + \cdots + p_{N+1} = 1$ such that

$$\sup_{P \in \mathfrak{A}} \int_\Xi f(S(z)(\xi), z, \xi)\,\mathrm{d}P(\xi) = \sum_{i=1}^{N+1} p_i f(S(z)(\xi_i), z, \xi_i)$$

Approximation: Localized measures μ_j

$$\mathfrak{A}_n = \left\{ p \in \mathbb{R}^n \;:\; \sum_{j=1}^n p_j = 1, \quad \begin{array}{l} \sum_{j=1}^n \frac{p_j}{\mu_j(\Xi)} \int_\Xi \psi_i(\xi)\,\mathrm{d}\mu_j(\xi) = m_i,\; i = 1, \ldots, N_e \\ \sum_{j=1}^n \frac{p_j}{\mu_j(\Xi)} \int_\Xi \psi_i(\xi)\,\mathrm{d}\mu_j(\xi) \le m_i,\; i = N_e + 1, \ldots, N \end{array} \right\}.$$

Theorem (Kouri): If $\mathfrak{A}_n \neq \emptyset$, then for each $z \in Z$ there exists $p_i \ge 0$ with at most $\min\{n, N+1\}$ nonzero such that $p_1 + \cdots + p_{N+1} = 1$ and

$$\sup_{q \in \mathfrak{A}_n} \sum_{j=1}^n \frac{q_j}{\mu_j(\Xi)} \int_\Xi f(S(z)(\xi), z, \xi)\,\mathrm{d}\mu_j(\xi) = \sum_{j=1}^{N+1} \frac{p_j}{\mu_j(\Xi)} \int_\Xi f(S(z)(\xi), z, \xi)\,\mathrm{d}\mu_j(\xi).$$

Sandia National Laboratories

Example — Moment Matching

Optimal Control of 1D Elliptic Equation

Let $\alpha = 10^{-4}$, $D_o = D_c = D = (-1, 1)$, and $w \equiv 1$ and consider

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

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

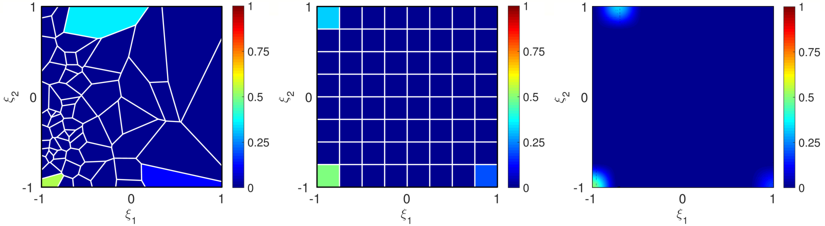
$$\begin{aligned} -\partial_x \left(\epsilon(\xi, x)\partial_x u(\xi, x)\right) &= f(\xi, x) + z(x) \qquad && (\xi, x) \in \Xi \times D, \\ u(\xi, -1) = 0, \quad u(\xi, 1) &= 0 && \xi \in \Xi. \end{aligned}$$

$\Xi = [-0.1, 0.1] \times [-0.5, 0.5]$, the true distribution is a tensor product of truncated exponentials, and the random field coefficients are

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

Example — Moment Matching

$$P(\Xi) = 1, \quad \int_\Xi \xi_1 \, \mathrm{d}P(\xi) \approx -0.537, \quad \text{and} \quad \int_\Xi \xi_2 \, \mathrm{d}P(\xi) \approx -0.313$$



- **Left:** Voronoi ($n = 64$) with 1000 MC samples per cell.
- **Center:** Uniform ($n = 64$) with level 4 sparse grids.
- **Right:** C^2 parition of unity ($n = 64$) with level 4 sparse grids, i.e., shifted/scaled tensor products of

$$\theta(x) = \begin{cases} 4x^2(3 - 4x) & \text{if } 0 < x \le \frac{1}{2} \\ 4(x-1)^2(4x-1) & \text{if } \frac{1}{2} < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Example — Moment Matching

	n	Obj. Val.	Center	Prob.	Center	Prob.	Center	Prob.
Voronoi 1000	16	0.13457	$(-0.864, -0.893)$	0.435	$(-0.634, 0.841)$	0.328	$(0.195, -0.848)$	0.237
	64	0.13777	$(-0.882, -0.933)$	0.540	$(-0.331, 0.849)$	0.346	$(0.467, -0.909)$	0.114
	256	0.14056	$(-0.981, -0.983)$	0.605	$(0.116, 0.922)$	0.351	$(0.330, -0.960)$	0.044
	1024	0.14133	$(-0.126, -0.987)$	0.484	$(-0.916, 0.988)$	0.342	$(-0.939, -0.994)$	0.174
	4096	0.14207	$(-0.978, -0.997)$	0.368	$(-0.813, 0.988)$	0.343	$(0.350, -0.991)$	0.289
Square $\ell = 4$	16	0.13221	$(-0.750, -0.750)$	0.709	$(-0.750, 0.750)$	0.150	$(0.750, 0.750)$	0.142
	64	0.13779	$(-0.857, -0.875)$	0.496	$(-0.875, 0.875)$	0.321	$(0.875, -0.875)$	0.193
	256	0.14058	$(-0.063, -0.938)$	0.457	$(-0.938, 0.938)$	0.333	$(-0.938, -0.938)$	0.210
	1024	0.14194	$(-0.969, -0.969)$	0.438	$(-0.969, 0.969)$	0.338	$(0.906, -0.969)$	0.223
	4096	0.14286	$(-1.000, -1.000)$	0.433	$(-0.968, 1.000)$	0.342	$(1.000, -1.000)$	0.225
C^2 $\ell = 4$	16	0.13444	$(-1.000, -1.000)$	0.696	$(1.000, 1.000)$	0.164	$(-1.000, 1.000)$	0.140
	64	0.13953	$(-1.000, -1.000)$	0.501	$(-0.714, 1.000)$	0.329	$(1.000, -1.000)$	0.170
	256	0.14154	$(-1.000, -1.000)$	0.663	$(0.867, 1.000)$	0.231	$(-1.000, 1.000)$	0.106
	1024	0.14244	$(-1.000, -1.000)$	0.441	$(-0.935, 1.000)$	0.340	$(1.000, -1.000)$	0.218
	4096	0.14286	$(-1.000, -1.000)$	0.433	$(-0.968, 1.000)$	0.342	$(1.000, -1.000)$	0.225
	$\star$	0.15640	$(-0.995, -0.996)$	0.657	$(0.432, 1.000)$	0.323	$(-0.993, 0.999)$	0.019

$\star$ Computed using Gaivoronski's stochastic descent algorithm for moment matching.

Example — CVaR

Optimal Control of 1D Elliptic Equation

Let $\alpha = 10$, $\Omega_o = \Omega_c = \Omega = (-1,1)$, and $w \equiv 1$ and consider

$$\operatorname*{minimize}_{z \in L^2(-1,1)} \quad J(z) = \frac{1}{2}\mathcal{R}\left[\int_{-1}^{1} (S(z)(\cdot,x) - 1)^2 \,\mathrm{d}x\right] + \frac{\alpha}{2}\int_{-1}^{1} z(x)^2 \,\mathrm{d}x$$

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

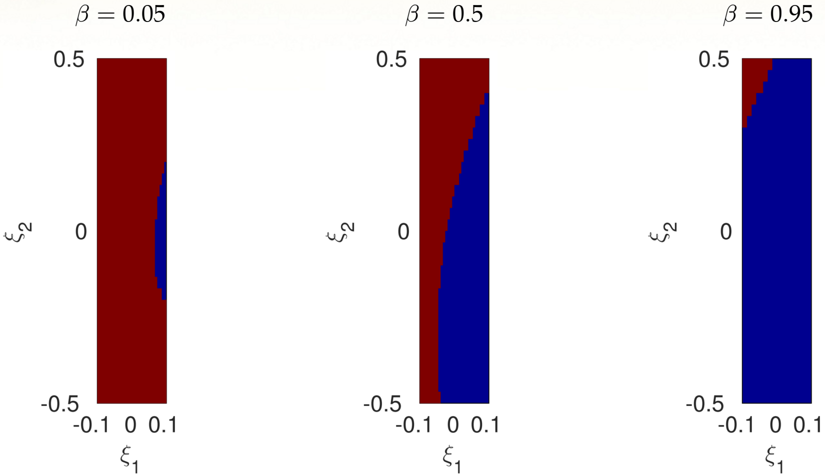
$$\begin{aligned} -\partial_x\left(\epsilon(\xi,x)\partial_x u(\xi,x)\right) &= f(\xi,x) + z(x) & (\xi,x) \in \Xi\times\Omega, \\ u(\xi,-1) = 0, \quad u(\xi,1) &= 0 & \xi \in \Xi. \end{aligned}$$

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

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

Example — CVaR

Discretization: Uniform ($n = 900$) with level 4 sparse grids.



$$\mathfrak{A}_n = \left\{ p \in \mathbb{R}^n \,:\, \sum_{i=1}^{n} p_i = 1,\ 0 \le p_i \le \frac{\mu(V_i)}{1-\beta},\ i = 1, \ldots, n \right\}$$

Outline

Problem Formulation

Known Probability Distribution

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 primal and bidual formulations;
  - Can use Newton's method/quad. and can prove error bounds.
- **Unknown Distribution:**
  - Incorporate data into distributionally-robust opt. formulation;
  - Objective function 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 with adaptive tessellation and sampling that exploits PDE constraint.
- Incorporate **(buffered) probabilistic objectives and constraints** to control *tail-probabilities* and *rare events* (Rockafellar, Uryasev, Royset, Shapiro, Henrion, Kibzun, ...)

Sandia National Laboratories