Optimization of Elliptic PDEs with Uncertain Inputs: Basic Theory and Numerical Stability

Thomas M. Surowiec

Philipps-Universität Marburg

Joint with work with
Drew P. Kouri
Sandia National Laboratories,
Albuquerque, New Mexico

Werner Römisch
Humboldt-Universität zu Berlin

Mario Hoffhues
Philipps-Universität Marburg,

June 19, 2020

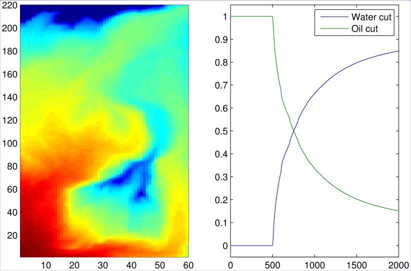
Deutsche Forschungsgemeinschaft

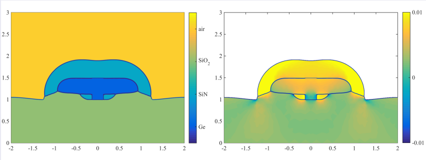
Philipps Universität Marburg

Uncertainty in Science and Engineering

Origins of Uncertainty

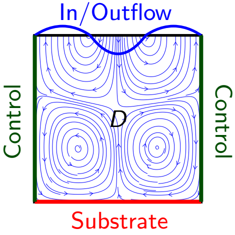
- Uncertainty is pervasive in all science and engineering applications!
- Incorporating uncertainty in physical models often leads to parametric systems of PDEs.
- Estimated/quantified random inputs.





- Simulation $\Rightarrow$ Optimization: Typically nonconvex, ∞-dim. optimization problems.
- After discretization: Extremely large-scale nonlinear programs.

- Reservoir Optimization based on work by J.E. Aarnes, K.–A. Lie (2004), generated by D. P. Kouri
- Optimal Gi-on-Si Semiconductor Laser Profile: L. Adam, M. Hintermüller, D. Peschka, T. M. S. (2019)

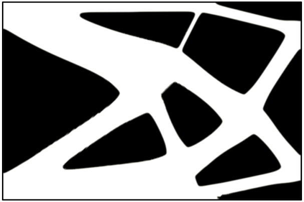


Optimization of Chemical Vapor Deposition

$$\min_{z\in\mathcal{Z}_{\mathrm{ad}}} \frac{1}{2}\mathcal{R}\left(\int_D (\nabla \times V(z))\,\mathrm{d}x\right) + \frac{\gamma}{2}\int_{\Gamma_c} |z|^2\,\mathrm{d}x$$

where $(V(z), P(z), T(z)) = (v, p, \tau)$ solves

$$\begin{aligned} -\nu\nabla^2 v + (v\cdot\nabla)v + \nabla p + \eta\tau g &= 0 \quad \text{in } D\\ -\kappa\Delta\tau + v\cdot\nabla\tau &= 0 \quad \text{in } D\\ \kappa\nabla\tau\cdot n + h(z-\tau) &= 0 \quad \text{on } \Gamma_c \end{aligned}$$



Topology Optimization of Elastic Structures

$$\min_{0\le z\le 1} \mathcal{R}\left(\int_D F\cdot S(z)\,\mathrm{d}x\right) \quad \text{s.t.} \quad \int_D z\,\mathrm{d}x \le V_0|D|,$$

where $S(z) = u$ solves

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

- CVD based on work by K. Ito, S.S. Ravindran (1998), generated by D. P. Kouri
- Optimal Truss Design in: P. Farrell, I. Papadopoulos, T. M. S. (2020)

Mitigating an Airborne Pollutant in Steady State

- Physical domain $D = [0,1]^2$, $p_k \in D$ fixed control locations
- Space of pollutant concentrations $U = H^1(D)$.
- **Uncertainties**: Diffusivity $\epsilon(\omega)$, wind $\mathbb{V}(\omega)$, sources $f(\omega)$.

Forward Problem

- **Model** Transport of pollutant with advection-diffusion eq.:

$$
\begin{aligned}
-\nabla \cdot (\epsilon(\omega)\nabla u) + \mathbb{V}(\omega) \cdot \nabla u &= f(\omega) - Bz && \text{in } D, \text{ a.s.} && \text{(1a)}\\
u &= 0 && \text{on } \Gamma_d = \{0\} \times (0,1), \text{ a.s.} && \text{(1b)}\\
\epsilon(\omega)\nabla u \cdot n &= 0 && \text{on } \partial D \setminus \Gamma_d, \text{ a.s.} && \text{(1c)}
\end{aligned}
$$

- **Goal** Steady state pollutant concentration should be minimized.
- **Decision** Where to inject chemicals that dissolve pollutant?
- **Decision variable** $z \in \mathbb{R}^9$, where $B \in \mathcal{L}(Z, L^2(D))$ is given by

$$
Bz = \sum_{k=1}^{9} z_k \exp\left(-\frac{(x-p_k)^\top (x-p_k)}{2\sigma^2}\right), \quad \sigma = 0.05.
$$

Mitigating an Airborne Pollutant in Steady State

A Stochastic Optimization Problem

- The target optimization problem is

$$\min_{z\in\mathcal{Z}_{\mathrm{ad}}} \mathcal{R}\left(\frac{\kappa_s}{2}\int_D S(z)^2\,\mathrm{d}x\right) + \wp(z), \quad \kappa_s > 0. \tag{2}$$

- $S(z) = u : \Omega \to U$ solves the random PDE (1).
- Admissible controls $\mathcal{Z}_{\mathrm{ad}} = \left\{z \in \mathbb{R}^9 \;:\; 0 \le z \le 1\right\}$.
- $\wp$ is the control cost:

$$\wp(z) = \kappa_c \|z\|_1 = \kappa_c \sum_{k=1}^{9} z_k, \quad \kappa_c > 0, \quad z \in \mathcal{Z}_{\mathrm{ad}}.$$

- $\mathcal{R}$ is a numerical surrogate to model risk preferences.

Mitigating an Airborne Pollutant in Steady State

How NOT to solve the problem when searching for a **robust** solution

- Replace all stochastic terms by their mean values ("mean value problem")
- Set $\mathcal{R} = \mathbb{E}$ ("risk neutral problem").
- Why not?

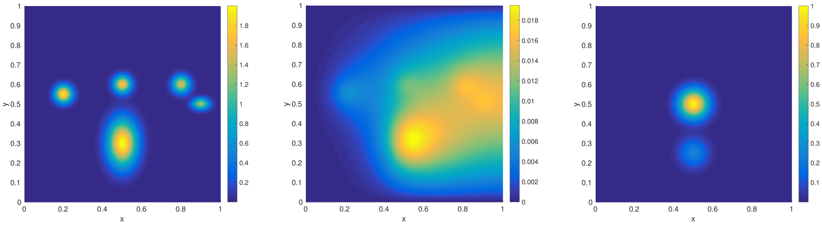


Figure: MV of sources f (l.), uncontrolled state evaluated at the MV of the random inputs (m.), optimal control for the MV problem (r.).

Mitigating an Airborne Pollutant in Steady State

How NOT to solve the problem when searching for a **robust** solution

- Replace all stochastic terms by their mean values ("mean value problem")
- Set $\mathcal{R} = \mathbb{E}$ ("risk neutral problem");
  -Optimal z should perform well **on average**.
- Why not?

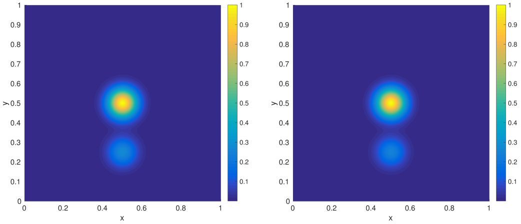


Figure: Optimal risk neutral z (l.), optimal z for mean value problem (r.).

Risk Measures

- $\mathcal{R}$ (hopefully!) yields a solution that properly handles the true uncertainty.
- $\mathcal{R}$ shapes the **distribution** of the **random variable** $\mathcal{J}(S(z)) : \Omega \to \mathbb{R}$.
- Examples:
  - $\mathcal{R}(X) = \mathbb{E}[X]$ (risk neutral)
  - $\mathcal{R}(X) = \alpha\mathbb{E}[X] + (1-\alpha)\mathbb{V}[X]$ (mean + variance)
  - $\mathcal{R}(X) = \alpha\mathbb{E}[X] + (1-\alpha)\mathrm{CVaR}_\beta[X]$ (mean + tail expectation)
- In many interesting cases: $\mathcal{R}$ is a **nonsmooth** convex functional.

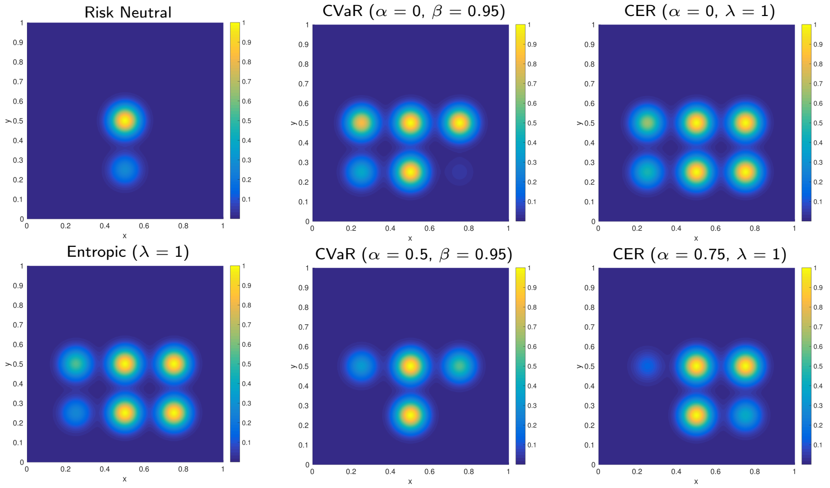
We consider risk-averse PDE-constrained optimization problems of the form:

$$\inf_{z \in \mathcal{Z}_{\mathrm{ad}}} \mathcal{R}(\mathcal{J}(S(z))) + \wp(z) \tag{RA}$$

- $\mathcal{R}$ risk measure, $\mathcal{J}$ random objective, S solution of PDE, $\wp$ deterministic.

Does any of this *really* make a difference?

Does any of this *really* make a difference?



Does any of this *really* make a difference?

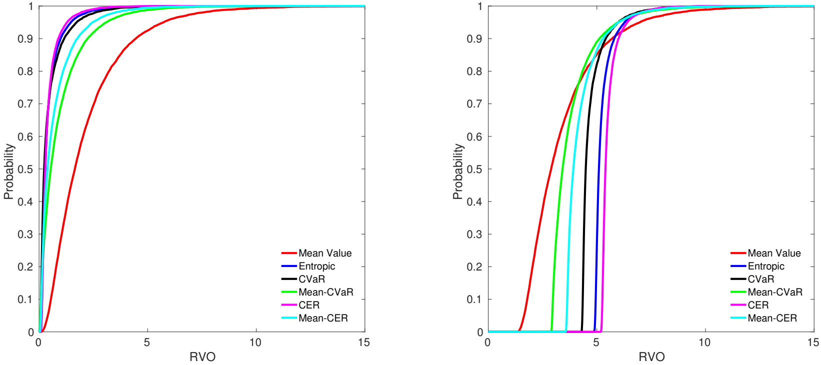


Figure: CDFs of $\mathcal{J}(S(z^\star))$. Distributions of $\mathcal{J}(S(z^\star))$ (l.). Distributions of $\mathcal{J}(S(z^\star)) + \wp(z^\star)$ representing total cost (r.)

PDE-Optimization under Uncertainty

$$\min_{z \in \mathcal{Z}_{\text{ad}}} \{F(z) := \mathcal{R}(\mathcal{J}(S(z))) + \wp(z)\}$$

Spaces

- **Probability Space:** $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F} \subseteq 2^\Omega$ and $\mathbb{P} : \mathcal{F} \to [0,1]$
- **Deterministic State Space:** U is a reflexive Banach space
- **Stochastic State Space:** $\mathcal{U} := L^q(\Omega, \mathcal{F}, \mathbb{P}; U)$ with $q \in [1, \infty]$
- **Random Objective Space:** $\mathcal{X} := L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [1, \infty)$
- **Control Space:** Z is a reflexive Banach space and
  $\emptyset \neq \mathcal{Z}_{\text{ad}} \subseteq Z$ is closed and convex

Functions

- $\mathcal{R} : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ is a **numerical surrogate** for cost
- $\mathcal{J} : \mathcal{U} \to \mathcal{X}$ is the **random variable objective function**
- $S : Z \to \mathcal{U}$ is the **random field PDE solution**
- $\wp : Z \to \mathbb{R} \cup \{\infty\}$ is the **control cost/penalty**

Example — Risk-Averse Optimal Control

Let $D = (0,1)^2$ and $\alpha > 0$. We consider

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} \quad \mathcal{R}\left(\frac{1}{2}\int_D (S(z;x,\omega) - w(x))^2 \,\mathrm{d}x\right) + \frac{\alpha}{2}\int_D z(x)^2 \,\mathrm{d}x$$

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

$$\begin{aligned} -\nabla \cdot (\kappa(\omega)\nabla u(\omega)) + \lambda(\omega)(\gamma u^3(\omega) + u(\omega)) &= B(\omega)z + f(\omega) && D, \\ \kappa(\omega)\nabla u(\omega) \cdot n &= 0 && \partial D, \end{aligned}$$

$\mathbb{P}$-almost surely (a.s.)

Operators

$$\langle \mathbf{A}u, v\rangle := \mathbb{E}\left[\int_D \kappa(\omega)\nabla u \cdot \nabla v \,\mathrm{d}x\right] \qquad \langle \mathbf{N}(u), v\rangle := \mathbb{E}\left[\int_D \lambda(\omega)(\gamma u^3 + u)v \,\mathrm{d}x\right]$$

$$\langle \mathbf{B}z, v\rangle := \mathbb{E}\left[\int_D (B(\omega)z)v \,\mathrm{d}x\right] \qquad \langle \mathbf{b}, v\rangle := \mathbb{E}\left[\int_D f(\omega)v \,\mathrm{d}x\right]$$

PDE-Optimization under Uncertainty: Assumptions

Solution Mapping and State Equation

$$S(z) = u \quad \text{solves} \quad \mathbf{A}u + \mathbf{N}(u) - \mathbf{B}z - \mathbf{b} = 0$$

(A1) $z \mapsto S(z)$ is completely continuous and continuously F-diff.

(A2) $\mathcal{J} : \mathcal{U} \to \mathcal{X}$ is continuously Fréchet differentiable

Adjoint State Mapping and Adjoint Equation

$$\Lambda(z) = \lambda \quad \text{solves} \quad (\mathbf{A} + \mathbf{N}'(S(z)))^* \lambda = -\nabla \mathcal{J}(S(z))$$

(A3) $z \mapsto \mathbf{B}^* \Lambda(z)$ is completely continuous

(A4) $\wp : Z \to \mathbb{R} \cup \{\infty\}$ is proper, closed and convex

Assumptions may require some effort to verify, cf.

- Kouri, D. P. and Surowiec, T.M. *Risk-Averse optimal control of semilinear elliptic PDEs*, to appear in ESAIM COCV: DOI: https://doi.org/10.1051/cocv/2019061.
- Kouri, D. P. and Surowiec, T.M. *Existence and optimality conditions for risk-averse PDE-constrained optimization*, SIAM/ASA J. Uncertainty Quantification 6, 2 (2018), 787–815.
- Kouri, D. P. and Surowiec, T.M. *Risk-averse PDE-constrained optimization using the conditional value-at-risk*, SIAM J. Optimization., 26(1) (2016), pp. 365-396.

Implications

The standing assumptions provide us with

- Existence of optimal solutions $z^\star$
- Optimality conditions
- A framework for function-space-based algorithms

Some Recent Advances in Numerical Methods

Risk-Neutral Case: $\mathcal{R} = \mathbb{E}$

- Stochastic Approximation/Stochastic Gradients:
  - Krumscheid, Martin, Nobile (2018*),
  - Martin, Nobile, Tsilifis (2020),
  - Geiersbach, Pflug (2019),
- Trust-Region with Adaptive Sparse Grids:
  - Kouri, Heinkenschloss, Ridzal, van Bloemen Waanders (2014)
  - Kouri, Heinkenschloss, Ridzal, van Bloemen Waanders (2013)
- Newton with adaptive Monte Carlo Sample:
  - Van Barel, Vandewalle (2019)

Risk-Averse Case: $\mathcal{R} = \text{CVaR}_\beta$, semideviation, etc.

- Smoothing + Trust Region: Kouri, TMS (2016, 2019a, 2018)
- Primal-dual: Kouri, TMS (2019b*)
- Interior-Point Garreis, TMS, Ulbrich M. (2019*)
- Splitting method: Angershausen, Clason (2019*)
- Smoothing + LS-Newton-CG: Heinkenschloss, Markowski (2019*)

$* =$ Preprints.

Outline

- Stability in Stochastic Optimization
- Challenges in the Hilbert Space Setting
- Qualitative Stability for a General Model
- Quantitate Stability for a General Model
- Application to PDE-Constrained Optimization under Uncertainty

What do we mean by "Stability"?

What do we mean by "Stability"?

A News Vendor Problem (cf. Shapiro, Dentcheva, Ruszczynski (2009))

- How much to order of x (deterministic) to satisfy demand d (stochastic)?
- Cost: $c > 0$ per unit
- Additional: Backorder cost $b > c$, holding cost $h \geq 0$ per unit.
- A stochastic optimization problem:

$$\min_{x \geq 0} \{ f(x) := \mathbb{E}_\xi [cx + b \max\{0, d(\xi) - x\} + h \max\{0, x - d(\xi)\}] \}$$

- "Maximize (avg.) performance of x by minimizing expected cost."

What do we mean by "Stability"?

A News Vendor Problem (cf. Shapiro, Dentcheva, Ruszczynski (2009))

- How much to order of x (deterministic) to satisfy demand d (stochastic)?
- Cost: $c > 0$ per unit
- Additional: Backorder cost $b > c$, holding cost $h \geq 0$ per unit.
- A stochastic optimization problem:

$$\min_{x \geq 0} \{ f(x) := \mathbb{E}_\xi [cx + b \max\{0, d(\xi) - x\} + h \max\{0, x - d(\xi)\}] \}$$

- "Maximize (avg.) performance of x by minimizing expected cost."
- Assume $d \sim U(1,2)$. Then

$$x^\star = H^{-1}(\kappa) = 1 + \kappa \text{ where } \kappa = \frac{b - c}{b + h}$$

and H^{-1} is the quantile function for d.

What do we mean by "Stability"?

A News Vendor Problem (cf. Shapiro, Dentcheva, Ruszczynski (2009))

- What if the distribution of d is **not known** or **estimated from data**?

What do we mean by "Stability"?

A News Vendor Problem (cf. Shapiro, Dentcheva, Ruszczynski (2009))

- What if the distribution of d is **not known** or **estimated from data**?
- Given: $d_1, \dots, d_N$ (scenarios) with probabilities $p_1, \dots, p_N$.
- We must consider instead:

$$\min_{x \geq 0} \{ \widehat{f}_N(x) := \sum_{i=1}^{N} p_i (cx + b \max\{0, d_i - x\} + h \max\{0, x - d_i\}) \}$$

- This is now a nonsmooth optimization problem!
- Continuous case: $b = 2, c = 1, h = 1 \Rightarrow x^\star = 4/3$
- Does the solution of the discrete problem $\widehat{x}_N^\star$ approach $4/3$ as $N \to +\infty$?

What do we mean by "Stability"?

- Q: Does $\widehat{x}_N^\star \to x^\star = 4/3$ as $N \to +\infty$?
- A: **Yes, but very slowly***.

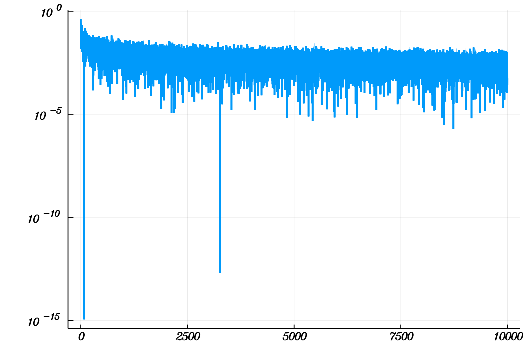


Figure: Sample size N versus error $|x^\star - \widehat{x}_N^\star|$.

$*$: $\widehat{x}_N^\star$ calculated for each N via projected subgradients with fixed step size 0.075 up to max 100 iterations.

What do we mean by "Stability"?

- Ω is a Polish space, $\mathcal{F}$ the associated Borel σ-algebra, $\mathcal{P}(\Omega)$ the set of Borel probability measures on Ω.
- Θ is a real separable Hilbert space, $\Theta_{\mathrm{ad}} \subset \Theta$ nonempty, closed, convex.
- $f : \Theta \times \Omega \to \overline{\mathbb{R}}$ is from a class of relevant integrands.

Optimal Value: $$\nu(\mathbb{P}) = \inf_{\theta \in \Theta_{\mathrm{ad}}} \int_\Omega f(\theta,\omega)\, \mathrm{d}\mathbb{P}(\omega).$$

Optimal Solutions: $$\theta_{\mathbb{P}} \in \operatorname*{argmin}_{\theta \in \Theta_{\mathrm{ad}}} \int_\Omega f(\theta,\omega)\, \mathrm{d}\mathbb{P}(\omega)$$

We would like statements of the type:

- If $\mathbb{P}$ is replaced by $\mathbb{P}_N$ such that $\mathbb{P}_N \Rightarrow \mathbb{P}$, does it hold that

$$\nu(\mathbb{P}_N) \to \nu(\mathbb{P}) \text{ and } \|\theta_{\mathbb{P}_N} - \theta\|_\Theta \to 0?$$

- If so, can we quantify this convergence somehow?

A (Very) Brief History

Qualitative and **quantitative stability analysis** has been developed for *n*-**dimensional parameter spaces** Θ and nontrivial constraint sets since at least the 1980s:

- Wets 1979, Solis & Wets 1981, 1983
- Dupačová 1983a, 1983b, 1984a, 1984b, 1987;
- Dupačová & Wets 1986, 1988

As noted in Dupačová & Wets 1988, "there is substantial statistical literature dealing with the questions broached here" notably

- Wald 1949,
- Huber 1967

(the latter exclude constraints, require smoothness, ...)

- See also the "more recent" surveys by Shapiro 2003 and Pflug 2003.
- Our perspective is based largely on Rachev & Römisch 2002.
- None of this work has been developed for ∞-dim. stochastic optimization

Return to PDE-Constrained Optimization

How does any of this relate to
PDE-Constrained Optimization under Uncertainty?

Return to PDE-Constrained Optimization

- $D \subset \mathbb{R}^m$ open, bounded domain, ∂D is Lipschitz.
- $V = H_0^1(D)$ with inner product $(u, v)_V = (\nabla u, \nabla v)_{L^2(D)}$.
- $V^* = H^{-1}(D)$ topological dual of V with dual pairing $\langle \cdot, \cdot \rangle$.
- $H = L^2(D)$ with inner product $(u, v)_H$.
- Ξ is a metric space, $\mathbb{P} \in \mathcal{P}(\Xi)$ the set of all Borel probability measures.

Return to PDE-Constrained Optimization

- $D \subset \mathbb{R}^m$ open, bounded domain, ∂D is Lipschitz.
- $V = H_0^1(D)$ with inner product $(u, v)_V = (\nabla u, \nabla v)_{L^2(D)}$.
- $V^* = H^{-1}(D)$ topological dual of V with dual pairing $\langle \cdot, \cdot \rangle$.
- $H = L^2(D)$ with inner product $(u, v)_H$.
- Ξ is a metric space, $\mathbb{P} \in \mathcal{P}(\Xi)$ the set of all Borel probability measures.

- Consider the bilinear form $a(\cdot, \cdot; \xi) : V \times V \to \mathbb{R}$ defined by

$$a(u, v; \xi) = \int_D \sum_{i,j=1}^{n} b_{ij}(x, \xi) \frac{\partial u(x)}{\partial x_i} \frac{\partial v(x)}{\partial x_j} \mathrm{d}x \quad (\xi \in \Xi).$$

- $b_{ij} : D \times \Xi \to \mathbb{R}$ are measurable on $D \times \Xi$ and there exist $L > \gamma > 0$:

$$\gamma \sum_{i=1}^{n} y_i^2 \leq \sum_{i,j=1}^{n} b_{ij}(x, \xi) y_i y_j \leq L \sum_{i=1}^{n} y_i^2 \quad (\forall y \in \mathbb{R}^n)$$

for a.e. $x \in D$ and $\mathbb{P}$-a.e. $\xi \in \Xi$. (i.e, each b_{ij} is essentially bounded.)

Return to PDE-Constrained Optimization

- We consider the optimization problem: Minimize the functional

$$
\begin{aligned}
\mathcal{J}(u,z) \quad &:= \quad \frac{1}{2}\int_{\Xi}\int_{D} |u(x,\xi)-\widetilde{u}_d(x)|^2 \,\mathrm{d}x\mathrm{d}\mathbb{P}(\xi) + \frac{\alpha}{2}\int_{D}|z(x)|^2\,\mathrm{d}x \\
&= \quad \frac{1}{2}\mathbb{E}_{\mathbb{P}}[\|u-\widetilde{u}_d\|_H^2] + \frac{\alpha}{2}\|z\|_H^2
\end{aligned}
$$

subject to $z \in Z_{\mathrm{ad}} \subset H$ (a closed convex bounded set), where u satisfies

$$
a(u,v;\xi) = \int_D (z(x)+g(x,\xi))v(x)dx \quad \text{for } \mathbb{P}\text{-a.e. } \xi \in \Xi,
$$

for all test functions $v \in C_0^\infty(D)$. We also assume $\alpha > 0$, $\widetilde{u}_d \in H$, and $g : D \times \Xi \to \mathbb{R}$ is measurable on $D \times \Xi$ and at least square integr. wrt D.

Deriving the Reduced Objective Functional

We derive a class of integrands $f : Z \times \Xi \to \mathbb{R}$ to motivate the general framework.

Deriving the Reduced Objective Functional

We derive a class of integrands $f : Z \times \Xi \to \mathbb{R}$ to motivate the general framework.

- For each $\xi \in \Xi$ define $A(\xi) : V \to V^\star$ via the Riesz Representation Thm:

$$\langle A(\xi)u, v\rangle = a(u, v; \xi) \quad (u, v \in V).$$

- $A(\xi)$ is linear, uniformly positive definite (with $\gamma > 0$) and uniformly bounded (with $L > 0$)
- The random PDE may be written in the form

$$A(\xi)u = z + g(\xi) \quad (\mathbb{P}\text{-a.e. } \xi \in \Xi).$$

Deriving the Reduced Objective Functional

We derive a class of integrands $f : Z \times \Xi \to \mathbb{R}$ to motivate the general framework.

- The inverse mapping $A(\xi)^{-1} : V^\star \to V$ exists and is linear, uniformly positive definite (with $\frac{1}{L}$) and uniformly bounded (with $\frac{1}{\gamma}$).

Deriving the Reduced Objective Functional

We derive a class of integrands $f : Z \times \Xi \to \mathbb{R}$ to motivate the general framework.

- The inverse mapping $A(\xi)^{-1} : V^\star \to V$ exists and is linear, uniformly positive definite (with $\frac{1}{L}$) and uniformly bounded (with $\frac{1}{\gamma}$).

Important Details:

- $J : V \to V^\star$ is the duality mapping: $\langle Ju, v\rangle = (u, v)_V \quad (u, v \in V)$.
- For $b \in V^\star$, $t > 0$ consider: $K_t(\xi)u = u - tJ^{-1}(A(\xi)u - b) \quad (v \in V)$.
- $K_t(\xi)$ is a contraction of V with constant $\kappa(t)$,

$$0 < \kappa(t) = \sqrt{1 - 2\gamma t + L^2 t^2} < 1 \quad \text{iff} \quad t \in \left(0, \frac{2\gamma}{L^2}\right).$$

- $u^\star \in V : K_t(\xi)u^\star = u^\star$ iff $A(\xi)u^\star = b$ and $u^\star \in \mathbb{B}\left(0, \frac{t}{1-\kappa(t)}\|b\|_\star\right)$ in V.

The fixed point arguments will be helpful later!

Deriving the Reduced Objective Functional

We derive a class of integrands $f : Z \times \Xi \to \mathbb{R}$ to motivate the general framework.

- We consider the integrand

$$
\begin{aligned}
f(z,\xi) &= \frac{1}{2}\left\|A(\xi)^{-1}(z+g(\xi)) - \widetilde{u}_d\right\|_H^2 + \frac{\alpha}{2}\|z\|_H^2 \\
&= \frac{1}{2}\left\|A(\xi)^{-1}z - (\widetilde{u}_d - A(\xi)^{-1}g(\xi))\right\|_H^2 + \frac{\alpha}{2}\|z\|_H^2 \\
&= \frac{1}{2}\left\|A(\xi)^{-1}z - u_d(\xi)\right\|_H^2 + \frac{\alpha}{2}\|z\|_H^2
\end{aligned}
$$

  for any $z \in Z_{\mathrm{ad}}$ and $\xi \in \Xi$,
- This provides us with an ∞-dimensional stochastic optimization problem:

$$
\min \left\{ F(z) = \int_{\Xi} f(z,\xi)\mathrm{d}\mathbb{P}(\xi) : z \in Z_{\mathrm{ad}} \right\}. \tag{3}
$$

Challenges in the Hilbert Space Setting

- ∞-dim. stochastic optimization problems arise in PDE-constrained optimization under uncertainty and functional data analysis.
- Θ is typically a space of functions, e.g., $L^2(D)$, $H^1(D)$, ...
- Even seemingly simple contraints sets such as

$$Z_{\mathrm{ad}} := \left\{ z \in L^2(D) \,|\, -1 \leq z(x) \leq 1 \quad \text{a.e. } x \in D \right\}$$

have **empty interior** and are **not compact wrt the norm topology**.

- Simple looking objectives such as

$$F(z) = \frac{1}{2}\mathbb{E}_{\mathbb{P}}\left[\left\| A(\xi)^{-1} z - u_d(\xi) \right\|_H^2 \right] + \frac{\alpha}{2}\|z\|_H^2$$

are only **continuous wrt the norm topology**, but many times we are forced to work with the **weak topology** on Z.

These issues rule out a direct translation of the n-dim arguments to the ∞-dim. setting.

An Abstract Problem Class

- Ω is a Polish space, $\mathcal{F}$ Borel σ-algebra, $\mathcal{P}(\Omega)$ Borel probability measures.
- Θ is a real separable Hilbert space, $\Theta_{\mathrm{ad}} \subset \Theta$ nonempty, closed, convex.
- (V,H) is a (real) rigged Hilbert space and $\theta_d \in H$.
- For $\theta \in \Theta$ and $\omega \in \Omega$, define

$$\Sigma(\omega)\theta := S(\omega)\theta - s(\omega),$$

  where $S(\omega) : \Theta \to V$ is bounded and linear in θ, ind. of ω, $s(\omega) \in H$.
- Define

$$f(\theta, \omega) := \frac{1}{2}\|\Sigma(\omega)\theta - \theta_d\|_H^2 = \frac{1}{2}\|S(\omega)\theta - (\theta_d + s(\omega))\|_H^2$$

  and assume $f(\theta, \cdot) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ for every $\theta \in \Theta_{\mathrm{ad}}$ and any $\mathbb{P} \in \mathcal{P}(\Omega)$.

An Abstract Problem Class

For $\alpha > 0$, we consider the optimization problems

$$\inf_{\theta \in \Theta_{\mathrm{ad}}} F(\theta) := \frac{1}{2}\mathbb{E}_{\mathbb{P}}[f(\theta)] + \frac{\alpha}{2}\|\theta\|_{\Theta}^2. \tag{4}$$

Proposition

Problem (4) admits a unique solution $\theta_{\mathbb{P}} \in \Theta_{\mathrm{ad}}$.

- F is finite, strongly convex, and lower semicontinuous, hence, weakly lower semicontinuous on the weakly compact set Θ_{ad}.
- In fact, we also have

$$\|\theta - \theta_{\mathbb{P}}\|_{\Theta}^2 \leq \frac{8}{\alpha}(F(\theta) - F(\theta_{\mathbb{P}})) \quad (\forall \theta \in \Theta_{\mathrm{ad}}).$$

Probability Metrics

How can we measure the distance between probability measures?

Probability Metrics

How can we measure the distance between probability measures?

- Ξ a metric space, $\{\mathbb{P}_N\} \subset \mathcal{P}(\Xi)$.
- $\mathbb{P}_N$ converges weakly to $\mathbb{P}$ in $\mathcal{P}(\Xi)$ iff

$$\lim_{N\to\infty} \mathbb{E}_{\mathbb{P}_N}[f] = \lim_{N\to\infty} \int_\Xi f(\xi)\mathrm{d}\mathbb{P}_N(\xi) = \int_\Xi f(\xi)\mathrm{d}\mathbb{P}(\xi) = \mathbb{E}_{\mathbb{P}}[f].$$

  for all bounded continuous functions $f \in C_b(\Xi, \mathbb{R})$.
- The topology of weak convergence is metrizable if Ξ is separable.

Probability Metrics

How can we measure the distance between probability measures?

- Ξ a metric space, $\{\mathbb{P}_N\} \subset \mathcal{P}(\Xi)$.
- $\mathbb{P}_N$ converges weakly to $\mathbb{P}$ in $\mathcal{P}(\Xi)$ iff

$$\lim_{N\to\infty} \mathbb{E}_{\mathbb{P}_N}[f] = \lim_{N\to\infty} \int_{\Xi} f(\xi)\mathrm{d}\mathbb{P}_N(\xi) = \int_{\Xi} f(\xi)\mathrm{d}\mathbb{P}(\xi) = \mathbb{E}_{\mathbb{P}}[f].$$

  for all bounded continuous functions $f \in C_b(\Xi, \mathbb{R})$.
- The topology of weak convergence is metrizable if Ξ is separable.

Zolotarev ζ-distance on $\mathcal{P}(\Xi)$ (Zolotarev 83):

$$d_{\mathfrak{F}}(\mathbb{P}, \mathbb{Q}) = \sup_{f\in\mathfrak{F}} \left| \int_{\Xi} f(\xi)\mathrm{d}\mathbb{P}(\xi) - \int_{\Xi} f(\xi)\mathrm{d}\mathbb{Q}(\xi) \right|,$$

where $\mathfrak{F}$ is a family of real-valued Borel measurable functions on Ξ.

Whether convergence wrt $d_{\mathfrak{F}}$ implies or is implied by weak convergence depends on the richness and on analytical properties of $\mathfrak{F}$.

Probability Metrics

The right tool for the right job...

A number of important probability metrics are of the form $d_{\mathfrak{F}}$:

- bounded Lipschitz metric (metrizes top'y of weak convergence)
- Fortet-Mourier type metrics:

Probability Metrics

The right tool for the right job...

A number of important probability metrics are of the form $d_{\mathfrak{F}}$:

- bounded Lipschitz metric (metrizes top'y of weak convergence)
- Fortet-Mourier type metrics: For $p \in [1, \infty)$, let $\mathfrak{F} = \mathcal{F}_p(\Omega)$, where

$$\mathcal{F}_p(\Omega) := \Big\{ f : \Omega \to \mathbb{R} \,\Big|\, |f(\omega_1) - f(\omega_2)| \leq \max\{1, d(\omega_1, \omega_0)^{p-1}, d(\omega_2, \omega_0)^{p-1}\} d(\omega_1, \omega_2) \quad \forall \omega_1, \omega_2 \in \Omega \Big\},$$

where d is the metric on Ω.

Probability Metrics

The right tool for the right job...

A number of important probability metrics are of the form $d_{\mathfrak{F}}$:

- bounded Lipschitz metric (metrizes top'y of weak convergence)
- Fortet-Mourier type metrics: For $p \in [1, \infty)$, let $\mathfrak{F} = \mathcal{F}_p(\Omega)$, where

$$\mathcal{F}_p(\Omega) := \Big\{ f : \Omega \to \mathbb{R} \,\big|\, |f(\omega_1) - f(\omega_2)| \leq \max\{1, d(\omega_1, \omega_0)^{p-1}, d(\omega_2, \omega_0)^{p-1}\} d(\omega_1, \omega_2) \quad \forall \omega_1, \omega_2 \in \Omega \Big\},$$

  where d is the metric on Ω.
- Rachev (1991): $\mathbb{P}_N \to \mathbb{P}$ weakly in $\mathcal{P}(\Omega)$ iff $d_{\mathcal{F}_p(\Omega)}(\mathbb{P}_N, \mathbb{P}) \to 0$.

Probability Metrics

The right tool for the right job...

A number of important probability metrics are of the form $d_{\mathfrak{F}}$:

- bounded Lipschitz metric (metrizes top'y of weak convergence)
- Fortet-Mourier type metrics: For $p \in [1, \infty)$, let $\mathfrak{F} = \mathcal{F}_p(\Omega)$, where

$$\mathcal{F}_p(\Omega) := \Big\{ f : \Omega \to \mathbb{R} \,\Big|\, \ |f(\omega_1) - f(\omega_2)| \leq \max\{1, d(\omega_1, \omega_0)^{p-1}, d(\omega_2, \omega_0)^{p-1}\} d(\omega_1, \omega_2) \quad \forall \omega_1, \omega_2 \in \Omega \Big\},$$

  where d is the metric on Ω.
- Rachev (1991): $\mathbb{P}_N \to \mathbb{P}$ weakly in $\mathcal{P}(\Omega)$ iff $d_{\mathcal{F}_p(\Omega)}(\mathbb{P}_N, \mathbb{P}) \to 0$.
- Why not Wasserstein?

$$d_{\mathcal{F}_p(\Omega)}(\mathbb{P}, \mathbb{Q}) \leq \left(1 + \int_\Omega d(\omega_0, \omega)^p \, \mathrm{d}\mathbb{P}(\omega) + \int_\Omega d(\omega_0, \omega)^p \, \mathrm{d}\mathbb{Q}(\omega) \right)^{\frac{p-1}{p}} W_p(\mathbb{P}, \mathbb{Q})$$

Probability Metrics

The right tool for the right job...

Lemma (Topsøe 67)

Let $\mathfrak{F}$ be uniformly bounded and $\mathbb{P}(\{\xi \in \Xi : \mathfrak{F}$ is not equicontinuous at $\xi\}) = 0$ holds. Then $\mathfrak{F}$ is a $\mathbb{P}$-uniformity class, i.e., weak convergence of $(\mathbb{P}_N)$ to $\mathbb{P}$ implies

$$\lim_{N \to \infty} d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) = 0.$$

- Compared with classical probability metrics we consider a much smaller family $\mathfrak{F}$:
$$\mathfrak{F} = \{f(\theta, \cdot) : \theta \in \Theta_{\mathrm{ad}}\}.$$
- This leads to a semi-metric, which we call problem-based or minimal information (m.i.) distance and $\mathfrak{F}$ the m.i. family, respectively.

A Qualitative Stability Result

Theorem

Suppose we have $\{\mathbb{P}_N\}$, $\mathbb{P}$ *with* $\mathbb{P}_N, \mathbb{P} \in \mathcal{P}(\Omega)$ *such that* $d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) \to 0$, *where* $\mathfrak{F}$ *is any set of extended real-valued Borel measurable functions on* Ω *that contain the m.i. family:*

$$\{f(\theta, \cdot) : \theta \in \Theta_{\mathrm{ad}}\} \subset \mathfrak{F}.$$

Then $\theta_{\mathbb{P}_N} \to \theta_{\mathbb{P}}$ *strongly in* Θ *as* $N \to +\infty$.

A Qualitative Stability Result

Theorem

Suppose we have $\{\mathbb{P}_N\}$, $\mathbb{P}$ *with* $\mathbb{P}_N, \mathbb{P} \in \mathcal{P}(\Omega)$ *such that* $d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) \to 0$, *where* $\mathfrak{F}$ *is any set of extended real-valued Borel measurable functions on* Ω *that contain the m.i. family:*

$$\{f(\theta, \cdot) : \theta \in \Theta_{\rm ad}\} \subset \mathfrak{F}.$$

Then $\theta_{\mathbb{P}_N} \to \theta_{\mathbb{P}}$ *strongly in* Θ *as* $N \to +\infty$.

- In fact, if we restrict the probability measures to

$$\mathcal{P}_p(\Omega) := \left\{ \mathbb{P} \in \mathcal{P}(\Omega) \,\middle|\, \int_\Omega d(\omega_0, \omega)^p \mathrm{d}\mathbb{P}(\omega) < +\infty \right\},$$

and assume there exists $p \in [1, \infty)$ and some $L > 0$ such that

$$\{\frac{1}{L} f(\theta, \cdot) : \theta \in \Theta_{\rm ad}\} \subset \mathcal{F}_p(\Omega) = \mathfrak{F}$$

then $\mathcal{P}_p(\Omega) \ni \mathbb{Q} \mapsto \theta_{\mathbb{Q}} \in \Theta_{\rm ad}$ is continuous wrt weak conv. on $\mathcal{P}_p(\Omega)$.

A Quantitative Stability Result

The qualitative results indicate that the solutions $\theta_{\mathbb{P}_N}$ will converge to a solution $\theta_{\mathbb{P}}$ if $\mathfrak{F}$ is a $\mathbb{P}$-uniformity class, but they do not provide us with a rate of convergence.

A Quantitative Stability Result

The qualitative results indicate that the solutions $\theta_{\mathbb{P}_N}$ will converge to a solution $\theta_{\mathbb{P}}$ if $\mathfrak{F}$ is a $\mathbb{P}$-uniformity class, but they do not provide us with a rate of convergence.

Theorem

Under the standing assumptions and with $\mathfrak{F} = \{f(\theta,\cdot) : \theta \in \Theta_{\mathrm{ad}}\}$, we obtain the following estimates for the optimal values $\nu(\mathbb{P})$ and solutions $\theta_{\mathbb{P}}$ of (3):

Lipschitz $\quad |\nu(\mathbb{Q}) - \nu(\mathbb{P})| \leq d_{\mathfrak{F}}(\mathbb{Q},\mathbb{P})$

Hölder (1/2) $\quad \|\theta_{\mathbb{Q}} - \theta_{\mathbb{P}}\|_H \leq 2\sqrt{\frac{2}{\alpha}}\, d_{\mathfrak{F}}(\mathbb{Q},\mathbb{P})^{\frac{1}{2}}$

for any $\mathbb{Q} \in \mathcal{P}(\Omega)$.

Return Again to PDE-Constrained Optimization

The results of the previous theorem clearly hold for our PDE-constrained problems. It remains to determine when $\mathfrak{F}$ is a $\mathbb{P}$-uniformity class.

Return Again to PDE-Constrained Optimization

The results of the previous theorem clearly hold for our PDE-constrained problems. It remains to determine when $\mathfrak{F}$ is a $\mathbb{P}$-uniformity class.

- The integrand is given by

$$f(z,\xi) \quad = \quad \frac{1}{2}\left\|A(\xi)^{-1}z - u_d(\xi)\right\|_H^2 + \frac{\alpha}{2}\|z\|_H^2$$

for any $z \in Z_{\mathrm{ad}}$ and $\xi \in \Xi$.

Return Again to PDE-Constrained Optimization

The results of the previous theorem clearly hold for our PDE-constrained problems. It remains to determine when $\mathfrak{F}$ is a $\mathbb{P}$-uniformity class.

- The integrand is given by

$$f(z,\xi) \quad = \quad \frac{1}{2}\left\|A(\xi)^{-1}z - u_d(\xi)\right\|_H^2 + \frac{\alpha}{2}\|z\|_H^2$$

  for any $z \in Z_{\mathrm{ad}}$ and $\xi \in \Xi$.
- $u^\star = A^{-1}(\xi)z$ iff $K_t(\xi)u^\star = u^\star$, where $K_t(\xi)$ is a contraction of V.
- Since $K_t(\xi)u = u - tJ^{-1}(A(\xi)u - b(\xi))$, if A and b are Lipschitz in ξ, then $u^\star(\xi)$ is Lipschitz as well! cf. e.g., Dontchev & Rockafellar (2014).

Return Again to PDE-Constrained Optimization

Theorem

Suppose

- *The coefficient functions* $b_{ij}(x,\cdot)$, $i,j = 1,\dots,n$, *and fixed RHS term* $g(x,\cdot)$ *are Lipschitz continuous on* Ξ *uniformly with respect to* $x \in D$.
- $g \in L^\infty(\Xi, \mathcal{F}, \mathbb{P}; V^\star)$

Then

- *The m.i. family* $\mathfrak{F} = \{f(z,\cdot) : z \in Z_{\mathrm{ad}}\}$ *is uniformly bounded and equi-Lipschitz continuous on* Ξ.
- $\mathfrak{F}$ *is a* $\mathbb{P}$*-uniformity class.*
- *The family* $\{f(\cdot,\xi) : \xi \in \Xi\}$ *is Lipschitz continuous on each bounded subset of* H *(with constant independent of* ξ*).*

A Priori Error Estimate

Solving

$$\min\ \mathbb{E}_{\mathbb{P}}[f(z)] + \frac{\alpha}{2}\|z\|_H^2 \text{ over } z \in Z_{\mathrm{ad}}$$

numerically requires

1. An Approximation $\mathbb{P}_N$ of $\mathbb{P}$.
2. Finite-Dimensional Approximations of H and V.

1. Let $\mathbb{P}_N$ be a finite-sample-based approximation of $\mathbb{P}$
2. Discretize V, H using finite elements with mesh parameter $h > 0$.

This leads to finite-dimensional approximations of the type:

$$\min\ \frac{1}{2}\sum_{i=1}^{N} \pi_i[\|(A_i^h)^{-1}(z^h) - u_{d,i}^h\|_{L^2(D)}^2] + \frac{\alpha}{2}\|z^h\|_{L^2(D)}^2 \text{ over } z^h \in Z_{\mathrm{ad}}^h.$$

A Priori Error Estimate

This leads to finite-dimensional approximations of the type:

$$\min \ \frac{1}{2}\sum_{i=1}^{N}\pi_i[\|(A_i^h)^{-1}(z^h) - u_{d,i}^h\|_{L^2(D)}^2] + \frac{\alpha}{2}\|z^h\|_{L^2(D)}^2 \text{ over } z^h \in Z_{\mathrm{ad}}^h.$$

where

- A_i^h arises from $A(\xi)$ using $\xi = \xi_i$ and replacing V by a finite-dimensional subspace V_h with nodal basis associated with a FEM.
- $u_{d,i}^h = (A_i^h)^{-1}(g_i^h - \widetilde{u}_{d,h})$.
- Z_{ad}^h approximates Z_{ad} using a FEM nodal basis for H.

A Priori Error Estimate

What rate can we expect for $\|z^h_{\mathbb{P}_N} - z_{\mathbb{P}_N}\|_H$?

A Priori Error Estimate

What rate can we expect for $\|z^h_{\mathbb{P}_N} - z_{\mathbb{P}_N}\|_H$?

Using e.g.,

- P0-FEM (see e.g., Falk 1973) or
- P1-FEM, variational discretization (see e.g. monographs Hinze, Pinnau, Ulbrich, Ulbrich (2009) or Antil & Leykekhman (2018) ref's therein):

$$\|z^h_{\mathbb{P}_N} - z_{\mathbb{P}_N}\|_H \leq C_N h^q \quad \omega \in \Omega.$$

  where $q \in (0, 3/2]$ depends on the regularity of D, B, g, $\underline{a}$, $\overline{a}$, and $\widetilde{u}_d$.
- $C_N \geq 0$ potentially depends on N
- Ideal case: $q = 3/2$; for P0-FEM: $q = 1$; otherwise: $q \in (1, 3/2)$.

A Priori Error Estimate

Combining this with the stability analysis yields

$$\|z_{\mathbb{P}} - z^h_{\mathbb{P}_N}\|_{L^2(D)} \leq 2\sqrt{2}\alpha^{-1/2} d_{\mathfrak{F}}(\mathbb{P}, \mathbb{P}_N)^{1/2} + C_N h^q.$$

A Priori Error Estimate

Combining this with the stability analysis yields

$$\|z_{\mathbb{P}} - z^h_{\mathbb{P}_N}\|_{L^2(D)} \leq 2\sqrt{2}\alpha^{-1/2} d_{\mathfrak{F}}(\mathbb{P}, \mathbb{P}_N)^{1/2} + C_N h^q.$$

Then $z^h_{\mathbb{P}_N} \to z_{\mathbb{P}}$ as $N \to +\infty$ and $h \downarrow 0$ provided $\{f(z,\cdot) \,|\, z \in Z_{\mathrm{ad}}\} = \mathfrak{F}$ is a $\mathbb{P}$-uniformity class and $\mathbb{P}_N \to \mathbb{P}$ weakly.

Open problem: Obtaining a rate on $d_{\mathfrak{F}}(\mathbb{P}, \mathbb{P}_N) \to 0$ via empirical process theory.
Open problem: Risk-averse problems, general objectives, nonlinear PDEs

Numerical Experiments

Find optimal forcing term z by solving

$$\min_{z\in\mathcal{Z}} \left\{\mathbb{E}\left(\frac{1}{2}\int_D (S(z)-u_d)^2\,\mathrm{d}x\right)+\frac{\alpha}{2}\int_D z^2\,\mathrm{d}x\right\}$$

where
$$-1 \le z(x) \le 1 \text{ a.e. } x \text{ in } D.$$

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

$$\begin{aligned}
-\nu(\xi)\partial^2_{xx}u &= f(\xi)+z && \text{in } D, \text{ a.s.}\\
[u(\xi)](0) &= d_0(\xi), && \text{a.s.}\\
[u(\xi)](1) &= d_1(\xi) && \text{a.s.}
\end{aligned}$$

We set

- $u_d(x) := \sin(50x/\pi)$, $\alpha = 10^{-3}$, $D = (0,1)$, $\mathcal{Z} = L^2(0,1)$
- ν, f, d_0, d_1: Random viscosity, forcing, boundary values.
- $\nu(\xi_1) = 10^{2\xi_1-3}$, $f(\xi_2) = 10^{-2}(2\xi_2-1)$, $d_0(\xi_3) = 1+10^{-3}(2\xi_3-1)$, $d_1(\xi_4) = 10^{-3}*(2\xi_4+1)$
- $\xi_1,\dots,\xi_4 \sim U(0,1)$.

Numerical Experiments

- Replace $\mathbb{P}$ by a sample-based Monte-Carlo approximation $\mathbb{P}_N$.
- State and control variables are discretized via P1-FE on a uniform grid.
- We solve the fully discrete problem via a semismooth Newton iteration.
- Control updates require solution of a (reduced) square linear system.
- Expectation of the reduced Hessian only implicitly available.
- This requires three sparse linear system solves for each sample.
- CG tolerance is set to 1e-8.
- SSN stops when discrete L^2-norm of residual drops below 1e-8.

Numerical Experiments

- Replace $\mathbb{P}$ by a sample-based Monte-Carlo approximation $\mathbb{P}_N$.
- State and control variables are discretized via P1-FE on a uniform grid.
- We solve the fully discrete problem via a semismooth Newton iteration.
- Control updates require solution of a (reduced) square linear system.
- Expectation of the reduced Hessian only implicitly available.
- This requires three sparse linear system solves for each sample.
- CG tolerance is set to 1e-8.
- SSN stops when discrete L^2-norm of residual drops below 1e-8.

- The iteration count for SSN is (as expected) mesh independent.
- Surprisingly, the iteration count is unaffected by the sample size N.
- The average number of SSN steps was 4.
- The average number CG iterations was 18.

Numerical Experiments: Solution Statistics

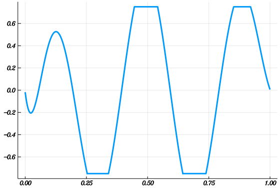


Figure: (r)

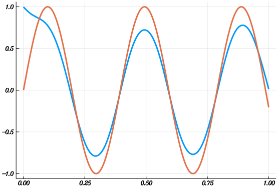


Figure: (ℓ)

(ℓ) Optimal solution $z^\star$ $(a = -0.75,\ b = 0.75,\ N = 500,\ h = 1/(1023), \alpha = 1\text{e-}3)$

(r) Average State $\mathbb{E}_{\mathbb{Q}_M}[S(z^\star)]$ (blue) vs. u_d (orange)

(Q_M empirical measure, $M = 1000$ **(out of sample!)**)

Numerical Experiments: Solution Statistics

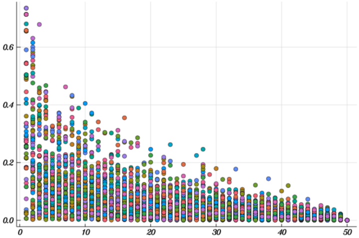


Figure: (ℓ)

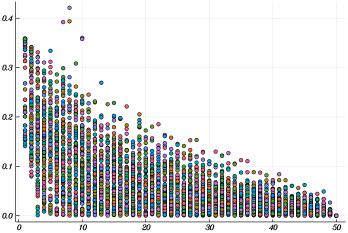


Figure: (ℓ)

(ℓ) $\|z^h_{\mathbb{P}_m} - z^h_{\mathbb{P}_N}\|_H$ ($a = -1.0$, $b = 1.0$, $h = 1/(1023)$, $N = 50$, $m = 1, \ldots, N$, 100 test runs)

(r) $|\nu^h(\mathbb{P}_m) - \nu^h(\mathbb{P}_N)|$ ($a = -0.75$, $b = 0.75$, $h = 1/(1023)$, $N = 50$, $m = 1, \ldots, N$, 100 test runs)

Conclusion & Outlook

- PDE-constrained optimization offers a powerful framework for modeling optimal decision making problems in engineering and the natural sciences.

Conclusion & Outlook

- PDE-constrained optimization offers a powerful framework for modeling optimal decision making problems in engineering and the natural sciences.
- Both the inputs as well as the models themselves are a source of uncertainty, noise, and ambiguity.

Conclusion & Outlook

- PDE-constrained optimization offers a powerful framework for modeling optimal decision making problems in engineering and the natural sciences.
- Both the inputs as well as the models themselves are a source of uncertainty, noise, and ambiguity.
- PDE-constrained optimization under uncertainty seeks to remedy these issues and provide solutions that are robust and resilient to outlier events.

Conclusion & Outlook

- PDE-constrained optimization offers a powerful framework for modeling optimal decision making problems in engineering and the natural sciences.
- Both the inputs as well as the models themselves are a source of uncertainty, noise, and ambiguity.
- PDE-constrained optimization under uncertainty seeks to remedy these issues and provide solutions that are robust and resilient to outlier events.
- Sampling/discretization/observation necessitate the replacement of the underlying probability measure $\mathbb{P}$ by some approximation $\mathbb{P}_N$.

Conclusion & Outlook

- PDE-constrained optimization offers a powerful framework for modeling optimal decision making problems in engineering and the natural sciences.
- Both the inputs as well as the models themselves are a source of uncertainty, noise, and ambiguity.
- PDE-constrained optimization under uncertainty seeks to remedy these issues and provide solutions that are robust and resilient to outlier events.
- Sampling/discretization/observation necessitate the replacement of the underlying probability measure $\mathbb{P}$ by some approximation $\mathbb{P}_N$.
- The method of probability metrics offers a means of quantifying the convergence of solutions $\theta_{\mathbb{P}_N}$ to the true solution $\theta_{\mathbb{P}}$ in the large-sample/large-data limit.

Some open problems:
- Sharp estimates on $d_{\mathfrak{F}}(\mathbb{P}, \mathbb{P}_N)$ without enlarging $\mathfrak{F}$
- Stability for risk-averse problems
- Dynamic problems

Thank You!

https://github.com/thomas-surowiec/cmai_talk_2020

- Slides
- Full bibliography
- Julia code for the examples