An Introduction to Risk-Averse PDE-Constrained Optimization: Theory, Numerical Solution, and Open Problems

Thomas M. Surowiec

Philipps-Universität Marburg
FB12 Mathematik und Informatik

CMAI Short Course 2021
Part I: Modeling and Basic Theory
June 18, 2021

Deutsche Forschungsgemeinschaft

Philipps Universität Marburg

Figure: Drew P. Kouri, Sandia National Laboratories

Overview of Part I

Section 1:
Introducing Uncertainty into PDE-Constrained Optimization Problems

The Basic Workflow of PDE-Constrained Optimization

The basic workflow from modeling to numerical solution is as follows.

- Modeling: PDE, additional constraints, and objective function.
- Theory: control-to-state map, existence and optimality conditions.
- Algorithms: function space-based methods (optimize-then-discretize).
- Numerical Solution: discretize (FD, FEM, wavelets, NN,...) and solve.

The Basic Workflow of PDE-Constrained Optimization

The basic workflow from modeling to numerical solution is as follows.

- Modeling: PDE, additional constraints, and objective function.
- Theory: control-to-state map, existence and optimality conditions.
- Algorithms: function space-based methods (optimize-then-discretize).
- Numerical Solution: discretize (FD, FEM, wavelets, NN,...) and solve.

Incorporating uncertainty into this workflow adds several new tasks:

- Modeling: where/how to include random inputs, model risk preference.
- Theory: measurability, integrability, & differentiability: an "extra step".
- Algorithms: sample uncertainty as-you-go versus before-you-go.
- Numerical Solution: when to stop, how to interpret the solution.

What is the right PDE model for my application?

PDE models for applications in the natural sciences can be complicated.

Example (Optimization of Optoelectronics)

We need a model that accounts for...

- ...elasticity of the structure (linear elasticity)
- ...optical properties (Helmholtz and photon number equation)
- ...electronic properties (van Roosbroeck)

coupled by the layout of the materials Ge, Si, SiN, SiO_4, air (decision variables).

What is the right PDE model for my application?

PDE models for applications in the natural sciences can be complicated.

Example (Optimization of Optoelectronics)

We need a model that accounts for...

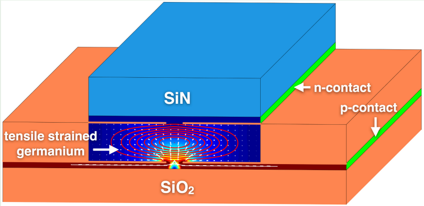
- ...elasticity of the structure (linear elasticity)
- ...optical properties (Helmholtz and photon number equation)
- ...electronic properties (van Roosbroeck)

coupled by the layout of the materials Ge, Si, SiN, SiO_4, air (decision variables).

The objective function should simultaneously ensure...

- ...tensile strain inside Ge-region is maximized
- ...bulk of support of (at least) first eigenmode

coincide.



What is the right PDE model for my application?

Topology Optimization

Optimize material distribution to maximize gain via band gaps

Elasticity

$$\mathrm{div}\Big(\mathbb{C} : (e(u) - e_0) + \sigma_0\Big) = 0 \quad (1d)$$

Doping Optimization

optimize doping to maximize gain / minimize threshold currents

Upscaling

band edges: $E_c(e), E_v(e)$

optical gain: $g(\xi; \omega, e)$

Optoelectronic PDE Model

van Roosbroeck

$$\begin{aligned} -\nabla \cdot \varepsilon \nabla \psi &= q(C + p - n) \\ -\nabla \cdot \mu_n n \nabla \phi_n &= +R \\ -\nabla \cdot \mu_p p \nabla \phi_p &= -R \end{aligned} \quad (1a)$$

Helmholtz Equation

$$\left[\nabla^2 + \frac{\omega^2}{c^2}\varepsilon_r(\xi; \omega, e)\right]\Theta = \beta^2 \Theta \quad (1b)$$

Photon Number Equation

$$\dot{S} = v_g(2\mathrm{Im}\beta - \alpha_c)S + \dot{S}_{sp} \quad (1c)$$

Example (Optimization of Optoelectronics)

- Theoretically, we know the physical effects of the topology (layout).
- What is the simplest effective model?

What is the right PDE model for my application?

Example (Optimization of Optoelectronics)

- What is the simplest effective model?
- Optimizing only considering elasticity and first eigenmode yields:

What is the right PDE model for my application?

Example (Optimization of Optoelectronics)

- What is the simplest effective model?
- Optimizing only considering elasticity and first eigenmode yields:

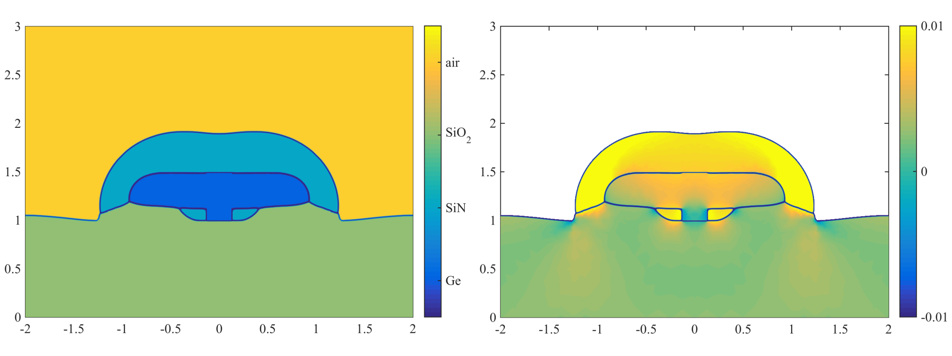
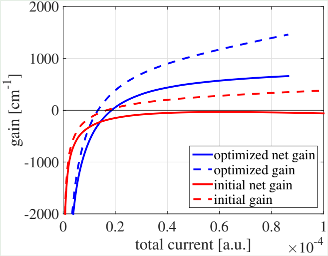


Figure: Optimal material layout (l.) and its corresponding strain field (r.).

What is the right PDE model for my application?

Example (Optimization of Optoelectronics)

- Simulations of the drift-diffusion system indicate success!



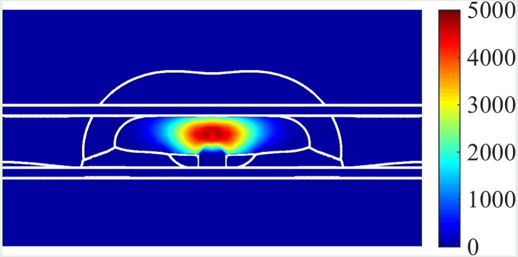


Figure: current-gain characteristics of initial and optimized device (l.) modal gain $g|\Theta|^2$ [cm^{-1}] optimized design (r.)

It's good to know the "true" model, but a simplification might be enough!

What is the right PDE model for my application?

Musings

Even if I know the "true" model, do I...

- ...know all the real input parameters?
- ...have I estimated them from data?

What is the right PDE model for my application?

Musings

Even if I know the "true" model, do I...

- ...know all the real input parameters?
- ...have I estimated them from data?

- What if I don't fully believe the "true" model?
- Can I use a simpler model by replacing the inputs with random parameters and learning their distributions from data?
- We don't wish to quantify uncertainty, but make optimal decision in the face of uncertainty.

These thoughts lead to the focus of our course:
Optimizing PDEs with Random Inputs.

What kind of restrictions on my decisions should I expect?

Aside from the PDEs, bound constraints present further difficulties.

Domain $D \subset \mathbb{R}^n$ open, bounded; $a, b, T : D \to \mathbb{R}$ $a < b$; $\Phi : Z \to \mathbb{R}$, $\gamma \in \mathbb{R}$.

- Control constraints: **decisions** $z : D \to \mathbb{R}$ must fulfill

$$a(x) \leq z(x) \leq b(x) \text{ for a.e. } x \in D \qquad \text{or} \qquad \Phi(z) \leq \gamma.$$

- State constraints: **solutions** u of PDE $u : D \to \mathbb{R}$ must fulfill

$$u(x) \leq T \text{ for a.e. } x \in D.$$

- Existence, uniqueness, etc. of Lagrange multipliers not always guaranteed.
- Sometimes the multipliers are only signed measures $\mu : \mathcal{P}(D) \to [0, \infty]$.

Example

- Topology opt.: $z \sim$ material density, $a \equiv 0$, $b \equiv 1$, $\Phi(z) = \int_D z \, \mathrm{d}x$.
- State constraint: T max. allowable deflection, temperature, current, etc.

What kind of restrictions on my decisions should I expect?

After incorporating uncertainty:

- The decision/design/control $z : D \to \mathbb{R}$ is made in anticipation of uncertainty and should be deterministic.
- The state u will be a random element $u : (\Omega, \mathcal{F}, \mathbb{P}) \to X(D)$, where $X(D)$ is some space of functions $v : D \to \mathbb{R}$.

Example (Stochastic State Constraints)

- Option A: $p \in (0, 1)$

$$\varphi(u) := \mathbb{P}(\{\omega \in \Omega \,|\, u(x, \omega) \leq T(x) \text{ a.e. } x \in D\}) \geq p.$$

- Option B: $p = 1$

$$\mathbb{P}(\{\omega \in \Omega \,|\, u(x, \omega) \leq T(x) \text{ a.e. } x \in D\}) = 1.$$

- Option A is mathematically hard: continuity, differentiability of φ nontrivial
- Option B is closer to deterministic case. May be too restrictive.
- Plenty of ongoing work, see reference list.

What exactly would I like to optimize?

Example (Optimization of Optoelectronics)

- $\varphi \sim$ layout of materials, $\boldsymbol{u} \sim$ material displacement, $\Theta \sim$ first eigenmode.
- $\int_\Omega j(\varphi, \Theta)\mathrm{tr}e(\boldsymbol{u})\ \mathrm{d}x \sim$ force overlap of Ge, high tensile strain, $\mathrm{supp}\,\Theta$
- $\alpha f_{\mathrm{GL}}(\varphi, \varepsilon) \sim$ regularizes material boundaries
- $J(\varphi, \Theta, \boldsymbol{u}) := \int_\Omega j(\varphi, \Theta)\mathrm{tr}e(\boldsymbol{u})\ \mathrm{d}x + \alpha f_{\mathrm{GL}}(\varphi, \varepsilon)$.

What exactly would I like to optimize?

Example (Optimization of Optoelectronics)

- $\varphi \sim$ layout of materials, $\boldsymbol{u} \sim$ material displacement, $\Theta \sim$ first eigenmode.
- $\int_\Omega j(\varphi, \Theta)\mathrm{tr}\varepsilon(\boldsymbol{u})\,\mathrm{d}x \sim$ force overlap of Ge, high tensile strain, $\operatorname{supp}\Theta$
- $\alpha f_{\mathrm{GL}}(\varphi, \varepsilon) \sim$ regularizes material boundaries
- $J(\varphi, \Theta, \boldsymbol{u}) := \int_\Omega j(\varphi, \Theta)\mathrm{tr}\varepsilon(\boldsymbol{u})\,\mathrm{d}x + \alpha f_{\mathrm{GL}}(\varphi, \varepsilon)$.

Example (Optimal Control: Tracking-Type Functionals)

- Minimize distance of u to u_d (desired state) with minimal cost $\alpha > 0$.

$$J(u, z) := \frac{1}{2}\|u - u_d\|^2_{L^2(D)} + \frac{\alpha}{2}\|z\|^2_{L^2(D)}.$$

What exactly would I like to optimize?

Example (Optimization of Optoelectronics)

- $\varphi \sim$ layout of materials, $\boldsymbol{u} \sim$ material displacement, $\Theta \sim$ first eigenmode.
- $\int_\Omega j(\varphi, \Theta)\mathrm{tre}(\boldsymbol{u})\,\mathrm{d}x \sim$ force overlap of Ge, high tensile strain, $\mathrm{supp}\,\Theta$
- $\alpha f_{\mathrm{GL}}(\varphi, \varepsilon) \sim$ regularizes material boundaries
- $J(\varphi, \Theta, \boldsymbol{u}) := \int_\Omega j(\varphi, \Theta)\mathrm{tre}(\boldsymbol{u})\,\mathrm{d}x + \alpha f_{\mathrm{GL}}(\varphi, \varepsilon)$.

Example (Optimal Control: Tracking-Type Functionals)

- Minimize distance of u to u_d (desired state) with minimal cost $\alpha > 0$.

$$J(u, z) := \frac{1}{2}\|u - u_d\|^2_{L^2(D)} + \frac{\alpha}{2}\|z\|^2_{L^2(D)}.$$

Example (Minimal Compliance in Topology Optimization)

- Find a material density $z : D \to \mathbb{R}$ that minimizes compliance

$$J(z) := (F, S(z))_{L^2(D)}$$

- F fixed force in the bulk; traction forces g on $\Gamma_N \subset \partial D$ also possible.

New: How can I model my risk preferences in this setting?

If we include uncertainty in the PDE, then we will be confronted with optimization problems of the type

$$\min_{z \in Z_{\mathrm{ad}}} \mathcal{J}(S(z))(\omega) + \rho(z)$$

- $z \in Z_{\mathrm{ad}}$ decision variables, designs, controls, etc. (**deterministic**)
- $z \mapsto S(z)$ solution of the random PDE. (**stochastic**)
- $\mathcal{J}$ objective. (either **deterministic** or **stochastic**)
- ρ cost or regularization term.

Since $\mathcal{J}(S(z))(\omega)$ is a random variable, this problem doesn't make sense yet.

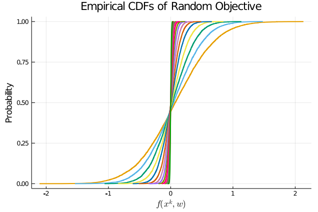
What does it mean if my objective is a random variable?

What can I hope to achieve?

New: How can I model my risk preferences in this setting?

Example (A Simple Example)

- $f(x,\omega) := \frac{1}{2}(\alpha_1(\omega)x_1^2 + \alpha_2(\omega)x_2^2) - (\beta_1(\omega)x_1 + \beta_2(\omega)x_2)$.
- $\alpha_i \sim U(0,1)$, $\beta_i \sim N(0,1)$ for $i = 1,2$.
- The stochastic optimization problem
$$\min_{x\in\mathbb{R}^2} \mathbb{E}[f(x)]$$
has unique solution $(x_1^\star, x_2^\star) = (0,0)$. Thus, $f(x^\star,\cdot)$ is a degenerate r.v.
- Solving iteratively, the sample cdf's converge as expected:



New: How can I model my risk preferences in this setting?

- This is an extreme example.
- But it illustrates the point: We choose the **numerical surrogate for risk** that **shapes the distribution** of the random variable

$$X_{z^\star}(\omega) := \mathcal{J}(S(z^\star))(\omega) + \rho(z^\star)$$

  in a desired way.
- For example, if $X_z \geq a$ for all $z \in \mathcal{Z}_{\mathrm{ad}}$, then ideally the length of the interval $[a, F^{-1}_{X_{z^\star}}(0.95)]$ is as small as possible.
- $F^{-1}_{X_{z^\star}}(0.95)$ is the upper 95% quantile of $X_{z^\star}$.

$$F^{-1}_{X_{z^\star}}(0.95) := \inf\{\alpha : \mathbb{P}(X_{z^\star} \leq \alpha) \geq 0.95\}$$

We go into more detail below in Section 2.
For now, we will consider the objectives $\mathcal{R}[\mathcal{J}(S(z))]$ and specify $\mathcal{R}$ later.

What do I know about the forward problem?

In general, the PDE-constraint can be viewed as

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

where

- $e : U \times Z \to W$,
- U is the state space,
- Z is the control space,
- W is typically a less regular space, e.g., the dual space U^*.

Much of the literature works under the assumption that

- There exists a continuous, differentiable mapping $z \mapsto S(z)$ such that

  $$e(S(z), z) = 0,$$

  so the PDE can be treated implicitly.
- S is often called the "**control-to-state mapping.**"
- **Reduced space approach**

What do I know about the forward problem?

- $D \subset \mathbb{R}^n$ open, bounded set with Lipschitz boundary Γ.
- $f \in L^2(D)$, $g \in H^{-1/2}(\Gamma)$, $u_0 \in H^1(\Gamma)$, $\eta > 0$.

Example (Linear Elliptic Boundary Value Problem (Neumann))

There exists a **unique solution** $u \in H^1(D)$ that solves the weak form of

$$-\Delta u + u = f \text{ in } D$$
$$\partial_{\boldsymbol{n}} u = g \text{ on } \Gamma$$

The mapping $(g, f) \mapsto u$ is **bounded and linear.**

What do I know about the forward problem?

- $D \subset \mathbb{R}^n$ open, bounded set with Lipschitz boundary Γ.
- $f \in L^2(D)$, $g \in H^{-1/2}(\Gamma)$, $u_0 \in H^1(\Gamma)$, $\eta > 0$.

Example (Linear Elliptic Boundary Value Problem (Neumann))

There exists a **unique solution** $u \in H^1(D)$ that solves the weak form of

$$\begin{aligned} -\Delta u + u &= f \text{ in } D \\ \partial_{\boldsymbol{n}} u &= g \text{ on } \Gamma \end{aligned}$$

The mapping $(g, f) \mapsto u$ is **bounded and linear.**

Example (Nonlinear Boundary Value Problem (Allen-Cahn))

There exist **solutions** $u \in H^1(D)$ of

$$\begin{aligned} -\Delta u + u - u^2 + u^3 &= f \text{ in } D \\ \partial_{\boldsymbol{n}} u &= \eta(u - u_0) \text{ on } \Gamma \end{aligned}$$

What do I know about the forward problem?

- Without $S(z)$, theory and algorithmic approaches are more difficult.
- A minimal regularity assumption requires continuous Fréchet derivatives $e_u(u^\star, z^\star), e_y(u^\star, z^\star)$ exist at a solution pair $(u^\star, z^\star)$ and

$$e'(u^\star, z^\star) \text{ is surjective }.$$

- **Full space approach**.

In either case, the workflow becomes

- Does there exist a well-defined control-to-state mapping $S(z)$?
- What regularity does u have? (e.g. u is in $H^1(D)$? $H^2(D)$?)
- Is S differentiable? What PDE does $w = S'(u)h$ solve?

We consider more problems in the Examples section below.

What do I know about the forward problem? (New)

After adding random inputs, the general PDE-constraint can be viewed as

$$e(u, z; \omega) = 0 \quad \mathbb{P}\text{-a.s.}.$$

This adds a few questions to the workflow (assume reduced space approach):

- Is $u : (\Omega, \mathcal{F}) \to U$ (strongly) **measurable**?
- Is $u : (\Omega, \mathcal{F}) \to U$ **integrable or essentially bounded**?

The regularity, continuity, and differentiability questions now need to be posed in **Lebesgue-Bochner spaces** $L^p(\Omega, \mathcal{F}, \mathbb{P}; V)$.

These spaces are typically considered in the context of deterministic parabolic and hypebolic PDEs with time interval $[0, T]$ replacing Ω.

A Model Forward Problem

Domain and Inputs

- (**Physical Domain:**) $D \subset \mathbb{R}^n$ is an open bound subset with sufficiently smooth boundary $\Gamma \subset \mathbb{R}^n$.
- (**Inputs:**) $A : D \to \mathbb{S}^{n \times n}$, $f : D \to \mathbb{R}$ are measurable mappings.

Data Assumptions

- (**Differential Operator:**) For $x \in D$, $A(x) \in \mathbb{S}^{n \times n}$ satisfies the usual boundedness and uniform ellipticity conditions.
- (**Forcing Term**[a]**:**) f is square-integrable on D, i.e., $f \in L^2(D)$.

[a] Can be relaxed to, e.g., $f \in H^1(D)^*$ or $f \in H^{-1}(D)$.

Deterministic Problem

Find $u : D \to \mathbb{R}$ such that

$$\begin{aligned} -\mathrm{div}\,(A(x)\nabla u(x)) &= f(x), && \text{for } x \in D, \\ u(s) &= 0, && \text{for } s \in \Gamma. \end{aligned} \tag{1}$$

A Model Forward Problem

Deterministic Solution Space, Weak Form

1. For $u \in C_c^\infty(D)$, define the norm $\|\cdot\|_U$ by

$$\|u\|_U^2 := \int_D \nabla u(x) \cdot \nabla v(x).$$

Recall: The closure of $C_c^\infty(D)$ w.r.t. $\|\cdot\|_U$ is a real separable Hilbert space, usually denoted by $H_0^1(D)$; here, often by U.

2. For $u, v \in U$, define

$$a(u,v) := \int_D A(x)\nabla u(x) \cdot \nabla v(x) \mathrm{d}x, \quad L(v) := \int_D f(x)v(x)\mathrm{d}x.$$

3. Consider weak/distributional/variational problem associated with (1): Find $u \in U$ such that

$$a(u,v) = L(v), \quad \forall v \in U. \tag{2}$$

A simple application of the Lax-Milgram Lemma shows that (2) admits a unique solution.

A Stochastic Forward Problem

Adding Random Inputs

- (**Random Domain:**) Ω is a sample space, $\mathcal{F} \subset \mathcal{P}(\Omega)$ a σ-algebra, $\mathbb{P} : \mathcal{F} \to [0,1]$ a probability measure such that $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space.
- (**Random Inputs:**) $A : \Omega \to L^\infty(D, \mathbb{S}^{n\times n})$, $f : \Omega \to H^{-1}(D)$ measurable mappings.

A Parametric Problem

Find $u : \Omega \to H_0^1(D)$ such that

$$\begin{aligned} -\mathrm{div}\,(A(\omega,x)\nabla u(\omega,x)) &= f(\omega,x), \quad && \text{for } x \in D, \\ u(\omega,s) &= 0, && \text{for } s \in \Gamma. \end{aligned} \tag{3}$$

holds for all $\omega \in \Omega$.

A Stochastic Forward Problem

Bochner Spaces

- $L^q(\Omega, \mathcal{F}, \mathbb{P}; U)$ ($q \in [1, \infty)$) is the space of all (strongly) measurable functions $y : \Omega \to U$:

$$\mathbb{E}_{\mathbb{P}}[\|y\|_U^q] < +\infty.$$

- $L^\infty(\Omega, \mathcal{F}, \mathbb{P}; U)$ is the space of all bounded (strongly) measurable functions $y : \Omega \to U$:

$$\operatorname{ess\,sup}_{\omega \in \Omega} \|y(\omega)\|_U < +\infty.$$

A Stochastic Forward Problem

Bochner Spaces

- $L^q(\Omega, \mathcal{F}, \mathbb{P}; U)$ ($q \in [1, \infty)$) is the space of all (strongly) measurable functions $y : \Omega \to U$:
$$\mathbb{E}_{\mathbb{P}}[\|y\|_U^q] < +\infty.$$
- $L^\infty(\Omega, \mathcal{F}, \mathbb{P}; U)$ is the space of all bounded (strongly) measurable functions $y : \Omega \to U$:
$$\operatorname{ess\,sup}_{\omega \in \Omega} \|y(\omega)\|_U < +\infty.$$

Variational Form (Stochastic Case)

- Let $\mathcal{U} := L^2(\Omega, \mathcal{F}, \mathbb{P}; U)$, $a : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$, $L : \mathcal{U} \to \mathbb{R}$ be defined by
$$\mathbf{a}(u, v) = \mathbb{E}\left[\int_D A(\omega, x)\nabla u(\omega, x) \cdot \nabla v(\omega, x)\mathrm{d}x\right], \quad \mathbf{L}(v) = \mathbb{E}\left[\int_D f(\omega, x)v(\omega, x)\mathrm{d}x\right].$$
- Find $u \in \mathcal{U}$ such that
$$\mathbf{a}(u, v) = \mathbf{L}(v) \quad \forall v \in \mathcal{U}. \tag{4}$$
- If $f \in L^2(\Omega, \mathcal{F}, \mathbb{P}; U^*)$, and **a** is $\mathcal{U}$-coercive, then there exists a unique solution $u \in \mathcal{U}$.

A Stochastic Forward Problem: Remarks

- It can be shown that $u \in \mathcal{U}$ solves (4) if and only if $u(\omega) \in U$ solves (3) w.p.1.
- $u : \Omega \to U$ is measurable and inherits the integrability provided by A and f.
- With more structure on A and f, $u : \Omega \to U$ may even be continuous or smooth.
- For optimization or optimal control, we might consider the problem

$$\mathbf{a}(u, v) = \langle \mathbf{B}(\cdot)z, v\rangle + \mathbf{L}(v) \quad \forall v \in \mathcal{U},$$

  where $B : \Omega \to \mathcal{L}(Z, \mathcal{U}^*)$ is, e.g., bounded and measurable in Ω.
- This defines a continuously Fréchet differentiable control-to-state mapping $z \mapsto S(z)$.

Challenges: Nonlinear and Evolution Equations

Remarks

- The classical theory yields not only existence and uniqueness of a solution, but also measurability, integrability, etc. for the linear case.
- The situation is more challenging for nonlinear elliptic PDE, e.g.,

$$\begin{aligned} -\xi(\omega)\Delta u + N(u,\omega) &= f, \text{ in } D, \text{ a.e. } \omega \in \Omega, \\ \xi(\omega)\partial_{\mathbf{n}} u &= z, \text{ on } \Gamma, \text{ a.e. } \omega \in \Omega, \end{aligned}$$

  as there is no general means of obtaining measurability and integrability.
- Problems involving time-dependence, e.g.,

$$\begin{aligned} \frac{\partial u}{\partial t} - \xi_t(\omega)\Delta u &= f, \text{ in } D \times (0,T), \text{ a.e. } \omega \in \Omega, \\ \xi_t(\omega)\partial_{\mathbf{n}} u &= z, \text{ on } \Gamma \times (0,T), \text{ a.e. } \omega \in \Omega, \\ u(0) &= u_0 \end{aligned}$$

  present an even greater challenge (ξ_t stochastic process). (OPEN PROBLEM)!

Why do I think this is a hard problem?

- If z is not a static decision variable, but time-dependent, then other than $z(0)$, the controls need to be adapted to the underlying stochastic process ξ_t.
- As more information is revealed in time, $z(t)$ will be dependent on the realization of stochasticity in the time $0 < s < t$.
- This usually leads to **dynamic programming** problems, which even in finite dimensions suffer from the curse of dimensionality.
- Now, z depends on **time** t, **uncertainty** ω, **space** $x \in D$ or $s \in \Gamma$. This only deepens the curse.
- **Open loop** and **multistage** perspectives are also possible, as has been done in stochastic programming for decades, but what does it mean in this context?

Why do I think this is a hard problem?

- If z is not a static decision variable, but time-dependent, then other than $z(0)$, the controls need to be adapted to the underlying stochastic process ξ_t.
- As more information is revealed in time, $z(t)$ will be dependent on the realization of stochasticity in the time $0 < s < t$.
- This usually leads to **dynamic programming** problems, which even in finite dimensions suffer from the curse of dimensionality.
- Now, z depends on **time** t, **uncertainty** ω, **space** $x \in D$ or $s \in \Gamma$. This only deepens the curse.
- **Open loop** and **multistage** perspectives are also possible, as has been done in stochastic programming for decades, but what does it mean in this context?

Example (A thought experiment)

- Suppose $z^\star$ should give us an optimal policy for drug administration based on a stochastic tumor growth model.
- If $z_t^\star(\omega) \in L^2(\Gamma)$ is the optimal control, how do I know what to do?
- Should I (assuming I could) blindly administer whatever a sample path of this process tells me?

Does my mathematical model possess a solution?

- In the deterministic setting, we are usually confronted with problems of the type:

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} \mathcal{J}(z) \qquad \text{(Reduced Space)}$$

or

$$\min_{(u,z) \in U \times \mathcal{Z}_{\mathrm{ad}}} \{ J(u,z) \mid e(u,z) = 0 \}. \qquad \text{(Full Space)}$$

- We work with academic models of real phenomena, we should know that a solution $z^\star$ exists.

Does my mathematical model possess a solution?

- In the deterministic setting, we are usually confronted with problems of the type:

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} \mathcal{J}(z) \qquad \text{(Reduced Space)}$$

or

$$\min_{(u,z) \in U \times \mathcal{Z}_{\mathrm{ad}}} \{ J(u,z) \mid e(u,z) = 0 \}. \qquad \text{(Full Space)}$$

- We work with academic models of real phenomena, we should know that a solution $z^\star$ exists.
- (**Reduced Space**)
  -Prove that $\mathcal{J} : \mathcal{Z}_{\mathrm{ad}} \to \mathbb{R}$ is weakly lower semicontinuous,
  -$\mathcal{Z}_{\mathrm{ad}}$ nonempty, weakly sequentially closed,
  -$\mathrm{lev}_\alpha \mathcal{J} \cap \mathcal{Z}_{\mathrm{ad}}$ is weakly sequentially compact for some $\alpha \in \mathcal{J}(z_0)$ with $z_0 \in \mathcal{Z}_{\mathrm{ad}}$.

Does my mathematical model possess a solution?

- In the deterministic setting, we are usually confronted with problems of the type:

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} \mathcal{J}(z) \qquad \text{(Reduced Space)}$$

or

$$\min_{(u,z) \in U \times \mathcal{Z}_{\mathrm{ad}}} \{ J(u,z) \mid e(u,z) = 0 \} . \qquad \text{(Full Space)}$$

- We work with academic models of real phenomena, we should know that a solution $z^\star$ exists.
- (**Reduced Space**)
  -Prove that $\mathcal{J} : \mathcal{Z}_{\mathrm{ad}} \to \mathbb{R}$ is weakly lower semicontinuous,
  -$\mathcal{Z}_{\mathrm{ad}}$ nonempty, weakly sequentially closed,
  -$\mathrm{lev}_\alpha \mathcal{J} \cap \mathcal{Z}_{\mathrm{ad}}$ is weakly sequentially compact for some $\alpha \in \mathcal{J}(z_0)$ with $z_0 \in \mathcal{Z}_{\mathrm{ad}}$.
- (**Full Space**)
  - Do essentially the same but over feasible set $\{(u,z) \in U \times \mathcal{Z}_{\mathrm{ad}} \mid e(u,z) = 0\}$

Does my mathematical model possess a solution?

- In the deterministic setting, we are usually confronted with problems of the type:

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} \mathcal{J}(z) \qquad \text{(Reduced Space)}$$

or

$$\min_{(u,z) \in U \times \mathcal{Z}_{\mathrm{ad}}} \{ J(u,z) \mid e(u,z) = 0 \}. \qquad \text{(Full Space)}$$

- We work with academic models of real phenomena, we should know that a solution $z^{\star}$ exists.
- (**Reduced Space**)
  -Prove that $\mathcal{J} : \mathcal{Z}_{\mathrm{ad}} \to \mathbb{R}$ is weakly lower semicontinuous,
  -$\mathcal{Z}_{\mathrm{ad}}$ nonempty, weakly sequentially closed,
  -$\mathrm{lev}_{\alpha} \mathcal{J} \cap \mathcal{Z}_{\mathrm{ad}}$ is weakly sequentially compact for some $\alpha \in \mathcal{J}(z_0)$ with $z_0 \in \mathcal{Z}_{\mathrm{ad}}$.
- (**Full Space**)
  - Do essentially the same but over feasible set $\{(u,z) \in U \times \mathcal{Z}_{\mathrm{ad}} \mid e(u,z) = 0\}$
- (**Stochastic Case**)
  - Reduced Space: Weak lsc of $z \mapsto \mathcal{R}[\mathcal{J}(S(z))]$ requires extra steps and assumptions.
  - Full Space: Potentially more challenging due to compactness issues...I haven't tried.

How can I characterize the solutions?

- As in all branches of optimization, we characterize solutions using optimality conditions.
- There are several ways of doing this for PDE-constrained optimization depending on whether the reduced space or full space approach is considered.

How can I characterize the solutions?

- As in all branches of optimization, we characterize solutions using optimality conditions.
- There are several ways of doing this for PDE-constrained optimization depending on whether the reduced space or full space approach is considered.
- (**Reduced Space**)
  -Prove that $\mathcal{J} : \mathcal{Z} \to \mathbb{R}$ is Gâteaux differentiable,
  -Assuming $\mathcal{Z}_{\mathrm{ad}}$ nonempty, closed, and convex, we have

$$\mathcal{J}'(z^\star)(z - z^\star) \geq 0 \quad \forall z \in \mathcal{Z}_{\mathrm{ad}}.$$

  - Using adjoints (see page 32), "unfold" this into a set of of equations and inequalities in $z^\star, u^\star = S(z^\star)$ and $\lambda^\star$ (adjoint state).
  - Generally requires **implicit function theorem** to differentiate S at $z^\star$.

How can I characterize the solutions?

- As in all branches of optimization, we characterize solutions using optimality conditions.
- There are several ways of doing this for PDE-constrained optimization depending on whether the reduced space or full space approach is considered.
- (**Reduced Space**)
  -Prove that $\mathcal{J} : \mathcal{Z} \to \mathbb{R}$ is Gâteaux differentiable,
  -Assuming $\mathcal{Z}_{\mathrm{ad}}$ nonempty, closed, and convex, we have

$$\mathcal{J}'(z^\star)(z - z^\star) \geq 0 \quad \forall z \in \mathcal{Z}_{\mathrm{ad}}.$$

  - Using adjoints (see page 32), "unfold" this into a set of of equations and inequalities in $z^\star, u^\star = S(z^\star)$ and $\lambda^\star$ (adjoint state).
  - Generally requires **implicit function theorem** to differentiate S at $z^\star$.
- (**Full Space**)
  - Uses optimization theory in Banach spaces.
  - Requires **constraint qualifications** (as usual in optimization) to guarantee existence of Lagrange multipliers, e.g., $e'(u^\star, z^\star)$ **is surjective**.

How can I solve the optimization problem numerically?

As mentioned above, there are two approaches

- Reduced space approach
- Full space approach

We highlight the main points for the **deterministic reduced space** approach.

The full space approach is especially important for PDEs with non-unique solutions, e.g., stationary Allen-Cahn (p. 17 above).

Reduced space approach

- The PDE is now implicitly satisfied.
- There may still exist control and state constraints.
- The PDE-constraint is formulated in function spaces U, Z, W, which are typically Hilbert spaces; in some settings general Banach spaces.
- Efficient algorithms need to be "aware" of the original function spaces:
  1. Inner products, dual norms, etc. should be implemented with the proper discrete counterpart.
  2. Gradients/Hessians of the objective, constraints, and solution operators should be calculated using the associated discrete Riesz mappings.
- This ensures (usually) **mesh independent** behavior and proper **scaling** by mesh refinements.

Reduced space approach

Example (Gradient Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Given $z \in Z$, calculate **state** $u = S(z)$ by solving $e(u, z) = 0$.

Reduced space approach

Example (Gradient Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Given $z \in Z$, calculate **state** $u = S(z)$ by solving $e(u, z) = 0$.
- Given u, solve **adjoint equation** for $\lambda = P(z)$: $e_u(u, z)^* \lambda = -J'_u(u, z)$.

Reduced space approach

Example (Gradient Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Given $z \in Z$, calculate **state** $u = S(z)$ by solving $e(u, z) = 0$.
- Given u, solve **adjoint equation** for $\lambda = P(z)$: $e_u(u, z)^* \lambda = -J_u'(u, z)$.
- Given z, u, λ calculate the full gradient

$$\nabla \mathcal{J}(z) = e_z(u, z)^* \lambda + \nabla J_u(u, z).$$

Reduced space approach

Example (Gradient Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Given $z \in Z$, calculate **state** $u = S(z)$ by solving $e(u, z) = 0$.
- Given u, solve **adjoint equation** for $\lambda = P(z)$: $e_u(u, z)^* \lambda = -J'_u(u, z)$.
- Given z, u, λ calculate the full gradient

$$\nabla \mathcal{J}(z) = e_z(u, z)^* \lambda + \nabla J_u(u, z).$$

- The latter might require an additional solve to ensure that the true discrete gradient is being using. (More in Part II.)

Reduced space approach

Example (Gradient Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Given $z \in Z$, calculate **state** $u = S(z)$ by solving $e(u, z) = 0$.
- Given u, solve **adjoint equation** for $\lambda = P(z)$: $e_u(u, z)^* \lambda = -J_u'(u, z)$.
- Given z, u, λ calculate the full gradient

$$\nabla \mathcal{J}(z) = e_z(u, z)^* \lambda + \nabla J_u(u, z).$$

- The latter might require an additional solve to ensure that the true discrete gradient is being using. (More in Part II.)
- In the linear case, we only require the solution of several (typically) **sparse structured linear** systems to determine $\nabla \mathcal{J}$.

Reduced space approach

Example (Gradient Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Given $z \in Z$, calculate **state** $u = S(z)$ by solving $e(u, z) = 0$.
- Given u, solve **adjoint equation** for $\lambda = P(z)$: $e_u(u, z)^* \lambda = -J'_u(u, z)$.
- Given z, u, λ calculate the full gradient

$$\nabla \mathcal{J}(z) = e_z(u, z)^* \lambda + \nabla J_u(u, z).$$

- The latter might require an additional solve to ensure that the true discrete gradient is being using. (More in Part II.)
- In the linear case, we only require the solution of several (typically) **sparse structured linear** systems to determine $\nabla \mathcal{J}$.
- Without the use of **adjoints**, the gradient would contain large dense matrices associated with the solution operators S, P.

Reduced space approach

Example (Hessian Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Define the Lagrangian $L(u, z, \lambda)$:

$$L(u, z, \lambda) = J(u, z) + \langle e(u, z), \lambda \rangle_{W, W^*}$$

Reduced space approach

Example (Hessian Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Define the Lagrangian $L(u, z, \lambda)$:

$$L(u, z, \lambda) = J(u, z) + \langle e(u, z), \lambda \rangle_{W, W^*}$$

- Given $z \in Z$, calculate $u = S(z)$ by solving $e(u, z) = 0$.

Reduced space approach

Example (Hessian Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Define the Lagrangian $L(u, z, \lambda)$:

$$L(u, z, \lambda) = J(u, z) + \langle e(u, z), \lambda \rangle_{W, W^*}$$

- Given $z \in Z$, calculate $u = S(z)$ by solving $e(u, z) = 0$.
- Given u, solve the adjoint equation for $\lambda = P(z)$: $e_u(u, z)^* \lambda = -J'_u(u, z)$

Reduced space approach

Example (Hessian Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Define the Lagrangian $L(u, z, \lambda)$:

$$L(u, z, \lambda) = J(u, z) + \langle e(u, z), \lambda \rangle_{W, W^*}$$

- Given $z \in Z$, calculate $u = S(z)$ by solving $e(u, z) = 0$.
- Given u, solve the adjoint equation for $\lambda = P(z)$: $e_u(u, z)^* \lambda = -J'_u(u, z)$
- Given u, z, direction v, solve the **linearized state equation** for w:

$$e_u(u, z)w = e_z(u, z)v.$$

Reduced space approach

Example (Hessian Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Define the Lagrangian $L(u, z, \lambda)$:

$$L(u, z, \lambda) = J(u, z) + \langle e(u, z), \lambda \rangle_{W, W^*}$$

- Given $z \in Z$, calculate $u = S(z)$ by solving $e(u, z) = 0$.
- Given u, solve the adjoint equation for $\lambda = P(z)$: $e_u(u, z)^* \lambda = -J'_u(u, z)$
- Given u, z, direction v, solve the **linearized state equation** for w:

$$e_u(u, z) w = e_z(u, z) v.$$

- Given u, z, λ, w, v solve the **adjoint equation** for p:

$$e_u(u, z)^* p = L''_{uu}(u, z, \lambda) w - L''_{uz}(u, z, \lambda) v.$$

Reduced space approach

Example (Hessian Computation of $\mathcal{J}(S(z)) = J(S(z), z)$)

- Define the Lagrangian $L(u, z, \lambda)$:

$$L(u, z, \lambda) = J(u, z) + \langle e(u, z), \lambda \rangle_{W, W^*}$$

- Given $z \in Z$, calculate $u = S(z)$ by solving $e(u, z) = 0$.
- Given u, solve the adjoint equation for $\lambda = P(z)$: $e_u(u, z)^* \lambda = -J'_u(u, z)$
- Given u, z, direction v, solve the **linearized state equation** for w:

$$e_u(u, z)w = e_z(u, z)v.$$

- Given u, z, λ, w, v solve the **adjoint equation** for p:

$$e_u(u, z)^* p = L''_{uu}(u, z, \lambda)w - L''_{uz}(u, z, \lambda)v.$$

- Given u, z, λ, w, v, p yields

$$\nabla^2 \mathcal{J}(z)v = e_z(u, z)^* p - \nabla_{zu} L(u, z, \lambda)w + \nabla_{zz} L(u, z, \lambda)v.$$

New: How should I treat uncertainty numerically?

There are essentially two possibilities (assume for now $\mathcal{R} = \mathbb{E}$)

- Sample-before-you-go: **Replace** underlying $\mathbb{P}$ by **approximation** $\mathbb{P}_N$ solve

$$\min_{z \in Z_{\mathrm{ad}}} \int_\Omega \mathcal{J}(S(z))(\omega) \mathrm{d}\mathbb{P}_N(\omega) = \sum_{i=1}^{N} \pi_i \mathcal{J}(S(z))(\omega^i)$$

  where π_i are given weights for the samples $i = 1, \ldots, N$.

New: How should I treat uncertainty numerically?

There are essentially two possibilities (assume for now $\mathcal{R} = \mathbb{E}$)

- Sample-before-you-go: **Replace** underlying $\mathbb{P}$ by **approximation** $\mathbb{P}_N$ solve

$$\min_{z \in Z_{\mathrm{ad}}} \int_\Omega \mathcal{J}(S(z))(\omega)\mathrm{d}\mathbb{P}_N(\omega) = \sum_{i=1}^{N} \pi_i \mathcal{J}(S(z))(\omega^i)$$

  where π_i are given weights for the samples $i = 1, \ldots, N$.

- Sample-while-you-go: For fixed step sizes $\gamma_1, \ldots, \gamma_N$ approximate the solution $z^\star$ using **stochastic gradients**

$$G_i := S'(z_i, \omega^i)^* \nabla \mathcal{J}(S(z^i, \omega^i)) \quad i = 1, \ldots, N$$

  with a sequence $\{z^i\}$ given by (for example)

$$z^{i+1} = \mathrm{Proj}_{Z_{\mathrm{ad}}}(z^i - \gamma^i G_i).$$

- Batches and second-order information can be incorporated.

New: How should I treat uncertainty numerically?

We discuss the pros and cons in detail in Part II of the course.

Brief Comments I

- "Sample-before-you-go" covers all manner of empirical approximations: Monte Carlo, Quasi-Monte Carlo, Deterministic Quadrature, ...
- Sometimes called Sample Average Approximation (SAA)
- Yields a deterministic PDE-constrained problem that can be solved with existing approaches.
- **These are not optimization algorithms.**

Brief Comments II

- "Sample-while-you-go" has its origins in Stochastic Approximation.
- Many variants with different step size rules, half-steps, extrapolations, etc.
- Immensely popular in machine learning
- **These are (typically first-order) algorithms.**

New: How should I treat uncertainty numerically?

- Knowing **when** to stop with **sample-before-you-go** is easy: Use the usual stopping criteria from nonlinear programming.
- E.g. Check the relative change in the residual of the first-order system: Define

$$\mathrm{res}_k := z_k - \mathrm{Proj}_{\mathcal{Z}_{\mathrm{ad}}}(z_k - \mathbb{E}_{\mathbb{P}_N}[\mathcal{J}(S(z_k))])$$

and stop when

$$\|\mathrm{res}_k\| \le \tau_{\mathrm{rel}}\|\mathrm{res}_0\| + \tau_{\mathrm{abs}},$$

where τ_{rel} and τ_{abs} are absolute and relative tolerances, respectively.

New: How should I treat uncertainty numerically?

- Knowing **when** to stop with **sample-before-you-go** is easy: Use the usual stopping criteria from nonlinear programming.
- E.g. Check the relative change in the residual of the first-order system: Define

$$\mathrm{res}_k := z_k - \mathrm{Proj}_{\mathcal{Z}_{\mathrm{ad}}}(z_k - \mathbb{E}_{\mathbb{P}_N}[\mathcal{J}(S(z_k))])$$

  and stop when

$$\|\mathrm{res}_k\| \le \tau_{\mathrm{rel}}\|\mathrm{res}_0\| + \tau_{\mathrm{abs}},$$

  where τ_{rel} and τ_{abs} are absolute and relative tolerances, respectively.
- Knowing **when** to stop the **sample-as-you-go** algorithms is more difficult.
- The basic convergence theory usually only provides statements on the objective functions. More comments in Part II.

New: How should I treat uncertainty numerically?

- Knowing **when** to stop with **sample-before-you-go** is easy: Use the usual stopping criteria from nonlinear programming.
- E.g. Check the relative change in the residual of the first-order system: Define

$$\mathrm{res}_k := z_k - \mathrm{Proj}_{\mathcal{Z}_{\mathrm{ad}}}(z_k - \mathbb{E}_{\mathbb{P}_N}[\mathcal{J}(S(z_k))])$$

  and stop when

$$\|\mathrm{res}_k\| \leq \tau_{\mathrm{rel}}\|\mathrm{res}_0\| + \tau_{\mathrm{abs}},$$

  where τ_{rel} and τ_{abs} are absolute and relative tolerances, respectively.
- Knowing **when** to stop the **sample-as-you-go** algorithms is more difficult.
- The basic convergence theory usually only provides statements on the objective functions. More comments in Part II.
- In both settings, the "solution" $z^\star(\mathbb{P}_N)$ is dependent on the realization of random processes.
- Ideally, we would know what happens to $z^\star(\mathbb{P}_N)$ as $N \to +\infty$. More comments in Part II.

Example: A Contaminant Mitigation Problem

Find optimal placement of mitigating factors z by solving:

$$\min_{z \in \mathcal{Z}_{\text{ad}}} \left\{ \mathcal{R}\left[\frac{\kappa_s}{2} \int_D S(z)^2 \, \mathrm{d}x \right] + \kappa_c \|z\|_1 \right\}$$

where $\kappa_s = 10^5$, $\kappa_c = 1$ and $S(z) = u : \Omega \to H^1(D)$ solves the weak form of

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

- $D = (0,1)^2$ is the physical domain, $(\Omega, \mathcal{F}, \mathbb{P})$ complete probability space
- $\mathcal{Z}$ is the control space, e.g., $L^2(D)$ or $\mathbb{R}^n$; $\mathcal{Z}_{\text{ad}} = \{z \in \mathcal{Z} \,|\, 0 \le z \le 1\}$.
- u is the advected pollutant.
- $\mathcal{R} : \mathcal{X} \to \overline{\mathbb{R}}$ is a numerical surrogate for "risk", i.e., a risk measure.

Random inputs: $\epsilon, \mathbb{V}, f$ permeability, wind, sources of contaminent.

Example: Topology Optimization

Find an optimal material distribution $z^\star$ that minimizes compliance by solving

$$\min_{z\in\mathcal{Z}_{\text{ad}}} \mathcal{R}\left[\int_D F(\cdot)\cdot S(z)\,\mathrm{d}x\right] + \wp(z)$$

where $S(z) = u$ solves

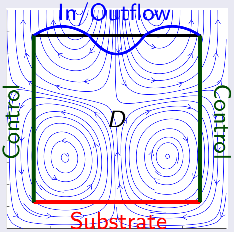
$$\begin{aligned} -\nabla\cdot(\mathbf{E}(\omega)(z) : \epsilon u) &= F(\omega) && \text{in } D \\ \epsilon u &= \tfrac{1}{2}(\nabla u + \nabla u^\top) && \text{in } D \\ u &= g(\omega) && \text{on } \partial D \end{aligned}$$

and the material density $z \in \mathcal{Z}_{\text{ad}}$ fulfills

- $z : D \to \mathbb{R}$.
- $z(x) \in [0,1]$ a.e. on D ($z = 0$ "no material", $z = 1$ "material").
- $\int_D z\,\mathrm{d}x \le V_0|D|$ (volume fraction).

Random inputs: Linear elastic isotropic material with uncertain Lamé coefficients **E** traction forces g bulk forces F.

Example: Optimization of Chemical Vapor Deposition



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

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

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

- Find an equilibrium boundary temperature $z : \Gamma_c \to \mathbb{R}$ that minimizes the vorticity in CVD reactor.
- V velocity, P pressure, T temperature.
- Possible random inputs: kinematic viscosity ν, thermal expansion coefficient η, thermal conductivity κ, heat transfer coefficient due to rugosity h.

Section 2:
Risk-Averse Decision Making

Robust Optimization

Motivation

- An **uncertainty region** Ω is known, but no data for statistical estimation is available.
- An **uncertainty region** Ω is known and data is available, but there is no room for error.
- An **uncertainty region** Ω is known and data is available, but the $\dim \Omega$ is intractable.
- The functional $\mathcal{R}$ should then take the form

$$\mathcal{R}[X] := \sup_{\omega \in \Omega} X(\omega).$$

Robust Optimization

Motivation

- An **uncertainty region** Ω is known, but no data for statistical estimation is available.
- An **uncertainty region** Ω is known and data is available, but there is no room for error.
- An **uncertainty region** Ω is known and data is available, but the $\dim \Omega$ is intractable.
- The functional $\mathcal{R}$ should then take the form

$$\mathcal{R}[X] := \sup_{\omega \in \Omega} X(\omega).$$

- The robust PDE-constrained optimization problem takes the general form:

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} \left\{ \wp(z) + \sup_{\omega \in \Omega} \mathcal{J}(S(z))(\omega) \right\} \tag{5}$$

- A nonsmooth, possibly nonconvex, ∞-dimensional problem.
- Existing approaches transform (5) into a mathematical program with equilibrium constraints (MPEC). MPECs with PDE operators are very hard to solve numerically.

Probability and Stochastic Dominance Constraints

Motivation

- A **complete probability space** $(\Omega, \mathcal{F}, \mathbb{P})$ is available.
- A **benchmark** decision, design, stationary control $z_d \in Z$ or objective value c is given.

Probability and Stochastic Dominance Constraints

Motivation

- A **complete probability space** $(\Omega, \mathcal{F}, \mathbb{P})$ is available.
- A **benchmark** decision, design, stationary control $z_d \in Z$ or objective value c is given.
- The risk-averse PDE-constrained optimization problem might take the general form:

$$\min_{z \in \mathcal{Z}_{\text{ad}}} \{\wp(z) + \mathbb{E}\left[\mathcal{J}(S(z))\right] : \mathbb{P}\left\{\mathcal{J}(S(z)) \le \mathcal{J}(S(z_d))\right\} \ge p\} \quad p \in (0,1). \tag{6}$$

Probability and Stochastic Dominance Constraints

Motivation

- A **complete probability space** $(\Omega, \mathcal{F}, \mathbb{P})$ is available.
- A **benchmark** decision, design, stationary control $z_d \in Z$ or objective value c is given.
- The risk-averse PDE-constrained optimization problem might take the general form:

$$\min_{z \in \mathcal{Z}_{\text{ad}}} \{\wp(z) + \mathbb{E}\left[\mathcal{J}(S(z))\right] : \mathbb{P}\left\{\mathcal{J}(S(z)) \leq \mathcal{J}(S(z_d))\right\} \geq p\} \quad p \in (0,1). \tag{6}$$

- "Find $z^\star \in \mathcal{Z}_{\text{ad}}$ that performs well on average such that the random variable $\omega \mapsto \mathcal{J}(S(z^\star))(\omega) - \mathcal{J}(S(z_d))(\omega)$ is non-positive with probability p."

Probability and Stochastic Dominance Constraints

Motivation

- A **complete probability space** $(\Omega, \mathcal{F}, \mathbb{P})$ is available.
- A **benchmark** decision, design, stationary control $z_d \in Z$ or objective value c is given.
- The risk-averse PDE-constrained optimization problem might take the general form:

$$\min_{z \in \mathcal{Z}_{\text{ad}}} \{\wp(z) + \mathbb{E}[\mathcal{J}(S(z))] : \mathbb{P}\{\mathcal{J}(S(z)) \leq \mathcal{J}(S(z_d))\} \geq p\} \quad p \in (0,1). \tag{6}$$

- "Find $z^\star \in \mathcal{Z}_{\text{ad}}$ that performs well on average such that the random variable $\omega \mapsto \mathcal{J}(S(z^\star))(\omega) - \mathcal{J}(S(z_d))(\omega)$ is non-positive with probability p."

Many variants available:

- Replace $\mathcal{J}(S(z_d))$ with a constant c or choose more than one/all $p \in (0,1)$.
- Compare tails of $\mathcal{J}(S(z))$ to $\mathcal{J}(S(z_d))$ over a range of values.
- The constraint function $\varphi(z) := \mathbb{P}\{\mathcal{J}(S(z)) \leq c\}$ is highly nontrivial.

Minimizing the Expectation

Optimization with Risk Measures

- The risk-averse PDE-constrained optimization problem takes the general form:

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} \{\wp(z) + \mathcal{R}[\mathcal{J}(S(z))]\} \tag{7}$$

- $\mathcal{R}$ should "shape" the distribution of the objective function at an optimal solution $z^\star$.

Traditional Approach: $\mathcal{R} = \mathbb{E}$ ("risk neutral case")

- Optimize to achieve best performance **on average**.
- Q: What could possibly go wrong?
- A: Does not account for potentially catastrophic tail events.
- A: Typically consider: $\nu\mathbb{E} + (1-\nu)\mathcal{R}$ with $\mathcal{R}$ being something other than $\mathbb{E}$, $\nu \in (0, 1)$, instead.

Mitigating Risk using Risk Measures

Traditional Mean-Var Approach: $\mathcal{R} = \nu\mathbb{E} + (1-\nu)\mathbb{V}$

- Maximize **average performance** vs. **minimize variance** $\mathbb{V}$
- Q: What could possibly go wrong?
- A: $\mathbb{V}$ may penalize favorable outlier situations: We are happy if $\mathcal{J}(S(z))(\omega) << \mathcal{J}(S(y))(\omega)$.
- $\mathbb{V}$ is not monotone w.r.t. the partial order on $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Thus, it could happen that

$$\mathcal{J}(S(z))(\omega) \leq \mathcal{J}(S(y))(\omega), \ \mathbb{P}\text{-a.e.}\omega \in \Omega$$

  but $\mathbb{V}(\mathcal{J}(S(z))) > \mathbb{V}(\mathcal{J}(S(y)))$.
- This is not favorable for optimization.

Mitigating Risk using Risk Measures

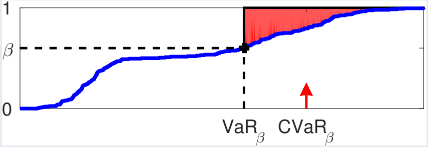
Minimize the β-Quantile: $\mathcal{R}[X] = \inf\{\tau : \mathbb{P}(X \leq \tau) \geq \beta\}$

- Also know as **Value-at-Risk** (confidence/risk level $\beta \in (0,1)$).
- Minimize the lowest value τ such that with probability β, $\mathcal{J}(S(z))$ does not exceed the value τ.
- Measures risky (catastrophic) events.
- Q: What could possibly go wrong?
- A: VaR does not account for the size of the tail.
- A: VaR is not subadditive.
- Obviously difficult to minimize in general.

Mitigating Risk using Risk Measures

Minimize the β-Average VaR: $\mathcal{R}[X] := \frac{1}{1-\beta}\int_\beta^1 \mathrm{VaR}_\alpha[X]\mathrm{d}\alpha$

- Many names: Excess Loss, Mean Shortfall, Average VaR, Tail VaR.
- We use **Conditional Value-at-Risk** (CVaR).
- Minimize the expectation of the tail above the β-quantile.



- CVaR is positively homogeneous, subadditive, monotone w.r.t. the usual partial order, and "**translation equivariant**": $\mathrm{CVaR}[X+c] = \mathrm{CVaR}[X] + c$ for any constant $c \in \mathbb{R}$.
- CVaR has a convenient form for optimization:

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

Basic Properties and Reformulations

- A risk measure $\mathcal{R} : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to \overline{\mathbb{R}}$ that is
  -**convex,**
  -**positively homogeneous,**
  -**monotonic**
  -**translation equivariant** (p. 46)
  is said to be **coherent**.
- CVaR_β is coherent.

- Coherent risk measures are distributionally robust wrt the nominal measure $\mathbb{P}$.
- What does that mean and why should I care?

Distributional Robustness and CVaR

- Return to risk neutral case $\mathcal{R} = \mathbb{E}$.
- $\mathbb{P}$ is unknown, but an iid sample is available.
- We could use the **empirical probability measure** $\mathbb{P}_N$ and consider

$$\min_{z \in \mathcal{Z}_{\text{ad}}} \wp(z) + \frac{1}{N} \sum_{i=1}^{N} \mathcal{J}(S(z))(\omega^i).$$

- This is a risk-neutral formulation.
- We could incorporate risk aversion by defining a set of measures

$$\mathfrak{A} \subset \mathcal{P}(\Omega),$$

where $\mathcal{P}(\Omega)$ is the set of all Borel probability measures over $(\Omega, \mathcal{F})$, that contains $\mathbb{P}_N$.

- A robust data-driven formulation would then be

$$\min_{z \in \mathcal{Z}_{\text{ad}}} \wp(z) + \sup_{\mathbb{Q} \in \mathfrak{A}} \mathbb{E}_{\mathbb{Q}} \left[\mathcal{J}(S(z)) \right].$$

Distributional Robustness and CVaR

What is the "correct" set $\mathfrak{A}$?

Examples

- Let $\mathfrak{F}$ be a subset of integrands $f : \Omega \to \mathbb{R}$.
- For $\varepsilon > 0$

$$\mathfrak{A} := \left\{ \mathbb{Q} \in \mathcal{P}(\Omega) : \sup_{f \in \mathfrak{F}} |\mathbb{E}_{\mathbb{P}_N}[f] - \mathbb{E}_{\mathbb{Q}}[f]| \leq \varepsilon \right\}$$

- Depending on the set $\mathfrak{F}$ this is, e.g., a Wasserstein-1, Fortet-Mourier, bounded Lipschitz, or minimal information metric ball.
- Without further insight, it is hard to see how this connects back to the theory of risk measures.
- The metrics do have the advantage that the support (atoms) can be moved.
- But the more tractable metrics, e.g., Wasserstein, are defined with integrands that have very little to do with the original integrand $\mathcal{J}(S(z))$.

Distributional Robustness and CVaR

- Using convex analysis, it can be shown in general that

$$\mathrm{CVaR}_\beta[X] = \sup_{\vartheta \in \mathfrak{A}} \mathbb{E}_{\mathbb{P}}[\vartheta X],$$

where $\mathfrak{A} = \left\{ \vartheta \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \;\middle|\; \mathbb{E}_{\mathbb{P}}[\vartheta] = 1, \quad 0 \le \vartheta \le \frac{1}{1-\beta} \quad \mathbb{P}\text{-a.s.} \right\}$

- For $\mathbb{P} = \mathbb{P}_N$ we have $\mathfrak{A} = \left\{ \vartheta \in \mathbb{R}^N \;\middle|\; \sum_{i=1}^N \frac{\vartheta_i}{N} = 1, \quad 0 \le \vartheta_i \le \frac{1}{1-\beta} \quad i = 1, \ldots, N \right\}$.
- This implies

$$\mathrm{CVaR}_\beta[X] = \sup_{\vartheta} \left\{ \frac{1}{N} \sum_{i=1}^N \vartheta_i X(\omega^i) : \sum_{i=1}^N \frac{\vartheta_i}{N} = 1, \quad 0 \le \vartheta_i \le \frac{1}{1-\beta} \quad i = 1, \ldots, N \right\}$$

- CVaR can therefore be used in a data-driven distributionally robust setting.
- The original weights on $\mathbb{P}_N$ are readjusted to ϑ_i / N.
- Our risk preference, expressed by CVaR_β, determines a meaningful set of probability measures.

Section 3:
Existence of Solutions and Optimality Conditions

Outline

- We formulate a canonical example based on Sections 1 and 2.
- There are several necessary structural assumptions on the nature of the uncertainty.
- We prove the existence of an optimal solution and derive optimality conditions.
- The emphasis is on understanding the structure of the proofs for extension to larger classes.

A Canonical Example

Risk-Averse PDE-Constrained Optimization

- Under the standing assumptions, we formulate our model problem:

$$\min \left\{ f(z) := \mathrm{CVaR}_\beta \left[\frac{1}{2} \int_D |S(z) - u_d|^2 \mathrm{d}x \right] + \frac{\alpha}{2} \|z\|_Z^2 \text{ over } z \in \mathcal{Z}_{\mathrm{ad}} \right\}, \tag{8}$$

where $\mathcal{Z}_{\mathrm{ad}} \subset Z$ is a nonempty, closed, and convex set and $S(z) = u$ is the unique solution to

$$\text{Find } u \in \mathcal{U} : \mathbb{E}\left[\int_D A \nabla u \cdot \nabla v \mathrm{d}x \right] = \mathbb{E}[\langle Bz + f, v \rangle_{U^*, U}], \quad \forall v \in \mathcal{U}.$$

- How do we prove existence of an optimal solution?
- CVaR is clearly non-smooth. How do we derive optimality conditions?

Technical Assumptions

We need to make a number of technical assumptions. Come back later when you have time...

Technical Assumptions

We need to make a number of technical assumptions. Come back later when you have time...

- $(\Omega, \mathcal{F}, \mathbb{P})$ complete probability space, $D \subset \mathbb{R}^n$ open, bounded with Lipschitz boundary Γ.
- $A : \Omega \to L^\infty(D)$, $\xi, \zeta \in \mathbb{R}^n$:
$$A_{ij}(\omega)\xi_i\xi_j \geq c(\omega)\|\xi\|^2_{\mathbb{R}^n}$$
$$|A_{ij}(\omega)\xi_i\zeta_j| \leq C(\omega)\|\xi\|_{\mathbb{R}^n}\|\zeta\|_{\mathbb{R}^n}$$
for some $C, c \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$: $C \geq c > 0$, $c^{-1} \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.
- $f \in \mathcal{U}^* = L^2(\Omega, \mathcal{F}, \mathbb{P}; U^*)$, but could be **more** regular, e.g., $f \in L^r(\Omega, \mathcal{F}, \mathbb{P}; U^*)$ with $r \geq 2$.
- For $\mathbb{P}$-a.e. $\omega \in \Omega$, $B(\omega) \in \mathcal{L}(Z, U^*)$.
- For any $z \in Z$, there exists a random variable $\kappa_B \in \mathcal{L}^q(\Omega, \mathcal{F}, \mathbb{P})$, $\kappa_B > 0$:
$$\|B(\cdot)z\|_{U^*} \leq \kappa_B(\cdot)\|z\|_Z, \quad \mathbb{P}\text{-a.e.}$$
- $B(\omega)$ is **completely continuous** for $\mathbb{P}$-a.e. $\omega \in \Omega$

Existence of an Optimal Solution

Proposition

The optimization problem (8) admits a (unique) solution $z^\star$.

Existence of an Optimal Solution

Proposition

The optimization problem (8) *admits a (unique) solution* $z^\star$.

Proof

- Since $\mathcal{Z}_{\text{ad}} \neq \emptyset$ and f is defined everywhere on Z, there exists an infimizing sequence $\{z_k\} \subset \mathcal{Z}_{\text{ad}}$. The infimum is finite since $f \geq 0$.

Existence of an Optimal Solution

Proposition

The optimization problem (8) *admits a (unique) solution* $z^\star$.

Proof

- Since $\mathcal{Z}_{\text{ad}} \neq \emptyset$ and f is defined everywhere on Z, there exists an infimizing sequence $\{z_k\} \subset \mathcal{Z}_{\text{ad}}$. The infimum is finite since $f \geq 0$.
- For some $z_0 \in \mathcal{Z}_{\text{ad}}$ and $k_0 \in \mathbb{N}$ sufficiently large, we have $\{z_k\}_{k \geq k_0} \subset \{z \in Z \mid f(z) \leq f(z_0)\}$.

Existence of an Optimal Solution

Proposition

The optimization problem (8) *admits a (unique) solution* $z^\star$*.*

Proof

- Since $\mathcal{Z}_{\text{ad}} \neq \emptyset$ and f is defined everywhere on Z, there exists an infimizing sequence $\{z_k\} \subset \mathcal{Z}_{\text{ad}}$. The infimum is finite since $f \geq 0$.
- For some $z_0 \in \mathcal{Z}_{\text{ad}}$ and $k_0 \in \mathbb{N}$ sufficiently large, we have $\{z_k\}_{k \geq k_0} \subset \{z \in Z \mid f(z) \leq f(z_0)\}$.
- Since, for all k, $\text{CVaR}_\beta[\frac{1}{2}\|S(z_k) - u_d\|^2] \geq 0$, we have $\|z_k\|_Z \leq \sqrt{\frac{2}{\alpha} f(z_0)}$

Existence of an Optimal Solution

Proposition

The optimization problem (8) *admits a (unique) solution* $z^\star$.

Proof

- Since $\mathcal{Z}_{\text{ad}} \neq \emptyset$ and f is defined everywhere on Z, there exists an infimizing sequence $\{z_k\} \subset \mathcal{Z}_{\text{ad}}$. The infimum is finite since $f \geq 0$.
- For some $z_0 \in \mathcal{Z}_{\text{ad}}$ and $k_0 \in \mathbb{N}$ sufficiently large, we have $\{z_k\}_{k \geq k_0} \subset \{z \in Z \mid f(z) \leq f(z_0)\}$.
- Since, for all k, $\text{CVaR}_\beta[\frac{1}{2}\|S(z_k) - u_d\|^2] \geq 0$, we have $\|z_k\|_Z \leq \sqrt{\frac{2}{\alpha} f(z_0)}$
- Hence, $\{z_k\}$ is bounded. Furthermore, as Z is reflexive, there exists $\{z_{k_l}\} \subset \{z_k\}$ and $z^\star \in Z$ such that $z_{k_l} \rightharpoonup z^\star$ (weakly).

Existence of an Optimal Solution

Proposition

The optimization problem (8) *admits a (unique) solution* $z^\star$.

Proof

- Since $\mathcal{Z}_{\text{ad}} \neq \emptyset$ and f is defined everywhere on Z, there exists an infimizing sequence $\{z_k\} \subset \mathcal{Z}_{\text{ad}}$. The infimum is finite since $f \geq 0$.
- For some $z_0 \in \mathcal{Z}_{\text{ad}}$ and $k_0 \in \mathbb{N}$ sufficiently large, we have $\{z_k\}_{k \geq k_0} \subset \{z \in Z \mid f(z) \leq f(z_0)\}$.
- Since, for all k, $\text{CVaR}_\beta[\frac{1}{2}\|S(z_k) - u_d\|^2] \geq 0$, we have $\|z_k\|_Z \leq \sqrt{\frac{2}{\alpha} f(z_0)}$
- Hence, $\{z_k\}$ is bounded. Furthermore, as Z is reflexive, there exists $\{z_{k_l}\} \subset \{z_k\}$ and $z^\star \in Z$ such that $z_{k_l} \rightharpoonup z^\star$ (weakly).
- Due to the equivalence of weak and strong closure for convex sets, $z^\star \in \mathcal{Z}_{\text{ad}}$. In remains to argue that $z^\star$ is in fact a minimizer.

Existence of an Optimal Solution

Proof (continued)

- Next, consider that $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \to 0$ for $\mathbb{P}$-a.e. ω.

Existence of an Optimal Solution

Proof (continued)

- Next, consider that $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \to 0$ for $\mathbb{P}$-a.e. ω.
- Let $w \in U$ be arbitrary. Since $\{z_k\}$ is bounded, there exists some $M > 0$, independent of ω, k, and w such that

$$|\langle B(\omega)(z_{k_l} - z^\star), w\rangle_{U^*,U}| \leq \|B(\omega)(z_{k_l} - z^\star)\|_{U^*}\|w\|_U \\ \leq \kappa_B(\omega)\|z_{k_l} - z^\star\|_Z\|w\|_U \leq \kappa_B(\omega)M\|w\|_U.$$

Existence of an Optimal Solution

Proof (continued)

- Next, consider that $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \to 0$ for $\mathbb{P}$-a.e. ω.
- Let $w \in U$ be arbitrary. Since $\{z_k\}$ is bounded, there exists some $M > 0$, independent of ω, k, and w such that

$$|\langle B(\omega)(z_{k_l} - z^\star), w\rangle_{U^*,U}| \leq \|B(\omega)(z_{k_l} - z^\star)\|_{U^*}\|w\|_U \leq \kappa_B(\omega)\|z_{k_l} - z^\star\|_Z\|w\|_U \leq \kappa_B(\omega)M\|w\|_U.$$

Taking the supremum over all $w \in U$ with $\|w\|_U = 1$ yields $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \leq \kappa_B(\omega)M$.

Existence of an Optimal Solution

Proof (continued)

- Next, consider that $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \to 0$ for $\mathbb{P}$-a.e. ω.
- Let $w \in U$ be arbitrary. Since $\{z_k\}$ is bounded, there exists some $M > 0$, independent of ω, k, and w such that

$$|\langle B(\omega)(z_{k_l} - z^\star), w\rangle_{U^*,U}| \leq \|B(\omega)(z_{k_l} - z^\star)\|_{U^*}\|w\|_U \\ \leq \kappa_B(\omega)\|z_{k_l} - z^\star\|_Z\|w\|_U \leq \kappa_B(\omega)M\|w\|_U.$$

  Taking the supremum over all $w \in U$ with $\|w\|_U = 1$ yields $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \leq \kappa_B(\omega)M$.
- It follows from Lebesgue's theorem that $B(\cdot)(z_{k_l} - z) \to 0$ strongly in $\mathcal{U}^*$.

Existence of an Optimal Solution

Proof (continued)

- Next, consider that $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \to 0$ for $\mathbb{P}$-a.e. ω.
- Let $w \in U$ be arbitrary. Since $\{z_k\}$ is bounded, there exists some $M > 0$, independent of ω, k, and w such that

$$|\langle B(\omega)(z_{k_l} - z^\star), w\rangle_{U^*,U}| \le \|B(\omega)(z_{k_l} - z^\star)\|_{U^*}\|w\|_U \\ \le \kappa_B(\omega)\|z_{k_l} - z^\star\|_Z\|w\|_U \le \kappa_B(\omega)M\|w\|_U.$$

  Taking the supremum over all $w \in U$ with $\|w\|_U = 1$ yields $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \le \kappa_B(\omega)M$.
- It follows from Lebesgue's theorem that $B(\cdot)(z_{k_l} - z) \to 0$ strongly in $\mathcal{U}^*$.
- In light of this, we deduce:

$$\begin{aligned} \underline{c}\|u_{k_l} - u^\star\|_{\mathcal{U}}^2 &\le \mathbf{a}(u_{k_l} - u^\star, u_{k_l} - u^\star) \\ &= \mathbb{E}[\langle B(z_{k_l} - z^\star, u_{k_l} - u^\star\rangle_{U^*,U}] \\ &\le \|B(\cdot)(z_{k_l} - z^\star)\|_{\mathcal{U}^*}\|u_{k_l} - u^\star\|_{\mathcal{U}}. \end{aligned}$$

Existence of an Optimal Solution

Proof (continued)

- Next, consider that $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \to 0$ for $\mathbb{P}$-a.e. ω.
- Let $w \in U$ be arbitrary. Since $\{z_k\}$ is bounded, there exists some $M > 0$, independent of ω, k, and w such that

$$|\langle B(\omega)(z_{k_l} - z^\star), w\rangle_{U^*,U}| \leq \|B(\omega)(z_{k_l} - z^\star)\|_{U^*}\|w\|_U \\ \leq \kappa_B(\omega)\|z_{k_l} - z^\star\|_Z\|w\|_U \leq \kappa_B(\omega)M\|w\|_U.$$

  Taking the supremum over all $w \in U$ with $\|w\|_U = 1$ yields $\|B(\omega)(z_{k_l} - z^\star)\|_{U^*} \leq \kappa_B(\omega)M$.
- It follows from Lebesgue's theorem that $B(\cdot)(z_{k_l} - z) \to 0$ strongly in $\mathcal{U}^*$.
- In light of this, we deduce:

$$\begin{aligned} \underline{c}\|u_{k_l} - u^\star\|_{\mathcal{U}}^2 &\leq \mathbf{a}(u_{k_l} - u^\star, u_{k_l} - u^\star) \\ &= \mathbb{E}[\langle B(z_{k_l} - z^\star, u_{k_l} - u^\star\rangle_{U^*,U}] \\ &\leq \|B(\cdot)(z_{k_l} - z^\star)\|_{\mathcal{U}^*}\|u_{k_l} - u^\star\|_{\mathcal{U}}. \end{aligned}$$

- Hence, $u_{k_l} = S(z_{k_l}) \to S(z^\star) = u^\star$ strongly in $\mathcal{U}$

Existence of an Optimal Solution

Proof (continued)

- $(J \circ S) : Z \to L^1(\Omega, \mathcal{F}, \mathbb{P})$ is convex, continuously Fréchet differentiable, completely continuous.

Existence of an Optimal Solution

Proof (continued)

- $(J \circ S) : Z \to L^1(\Omega, \mathcal{F}, \mathbb{P})$ is convex, continuously Fréchet differentiable, completely continuous.
- Since CVaR_β is positively homogeneous and subadditive, it is convex.

Existence of an Optimal Solution

Proof (continued)

- $(J \circ S) : Z \to L^1(\Omega, \mathcal{F}, \mathbb{P})$ is convex, continuously Fréchet differentiable, completely continuous.
- Since CVaR_β is positively homogeneous and subadditive, it is convex.
- CVaR_β is finite on $L^1(\Omega, \mathcal{F}, \mathbb{P})$. It follows that CVaR_β is continuous on $L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Existence of an Optimal Solution

Proof (continued)

- $(J \circ S) : Z \to L^1(\Omega, \mathcal{F}, \mathbb{P})$ is convex, continuously Fréchet differentiable, completely continuous.
- Since CVaR_β is positively homogeneous and subadditive, it is convex.
- CVaR_β is finite on $L^1(\Omega, \mathcal{F}, \mathbb{P})$. It follows that CVaR_β is continuous on $L^1(\Omega, \mathcal{F}, \mathbb{P})$.
- Combining
  - The complete continuity of the composite function $(\mathrm{CVaR}_\beta \circ J \circ S)$

Existence of an Optimal Solution

Proof (continued)

- $(J \circ S) : Z \to L^1(\Omega, \mathcal{F}, \mathbb{P})$ is convex, continuously Fréchet differentiable, completely continuous.
- Since CVaR_β is positively homogeneous and subadditive, it is convex.
- CVaR_β is finite on $L^1(\Omega, \mathcal{F}, \mathbb{P})$. It follows that CVaR_β is continuous on $L^1(\Omega, \mathcal{F}, \mathbb{P})$.
- Combining
  - The complete continuity of the composite function $(\mathrm{CVaR}_\beta \circ J \circ S)$
  - The weak lower semicontinuity of $\frac{\alpha}{2}\|\cdot\|_Z^2$,

Existence of an Optimal Solution

Proof (continued)

- $(J \circ S) : Z \to L^1(\Omega, \mathcal{F}, \mathbb{P})$ is convex, continuously Fréchet differentiable, completely continuous.
- Since CVaR_β is positively homogeneous and subadditive, it is convex.
- CVaR_β is finite on $L^1(\Omega, \mathcal{F}, \mathbb{P})$. It follows that CVaR_β is continuous on $L^1(\Omega, \mathcal{F}, \mathbb{P})$.
- Combining
  - The complete continuity of the composite function $(\text{CVaR}_\beta \circ J \circ S)$
  - The weak lower semicontinuity of $\frac{\alpha}{2}\|\cdot\|_Z^2$,

  we how shown that $f + i_{\mathcal{Z}_{\text{ad}}}$ is
  - weakly inf-compact : The sublevel set with upper bound $f(z_0)$ is weakly sequentially compact.
  - weakly lower-semicontinuous on Z.

Existence of an Optimal Solution

Proof (continued)

- $(J \circ S) : Z \to L^1(\Omega, \mathcal{F}, \mathbb{P})$ is convex, continuously Fréchet differentiable, completely continuous.
- Since CVaR_β is positively homogeneous and subadditive, it is convex.
- CVaR_β is finite on $L^1(\Omega, \mathcal{F}, \mathbb{P})$. It follows that CVaR_β is continuous on $L^1(\Omega, \mathcal{F}, \mathbb{P})$.
- Combining
  - The complete continuity of the composite function $(\mathrm{CVaR}_\beta \circ J \circ S)$
  - The weak lower semicontinuity of $\frac{\alpha}{2}\|\cdot\|_Z^2$,

  we how shown that $f + i_{\mathcal{Z}_{\text{ad}}}$ is
  - weakly inf-compact : The sublevel set with upper bound $f(z_0)$ is weakly sequentially compact.
  - weakly lower-semicontinuous on Z.
- Hence, $z^\star$ is an optimal solution.

$$f(z^\star) \geq \inf_{z \in \mathcal{Z}_{\text{ad}}} f(z) = \lim_{k \to +\infty} f(z_{k_l}) = \liminf_{l \to +\infty} f(z_{k_l}) \geq f(z^\star).$$

- f strongly convex implies uniqueness of solution.

Existence of an Optimal Solution

Important Points

- This is a standard argument from the calculus of variations with a slight twist due to the composite functional $\mathcal{R} \circ J \circ S$.

Existence of an Optimal Solution

Important Points

- This is a standard argument from the calculus of variations with a slight twist due to the composite functional $\mathcal{R} \circ J \circ S$.
- Convexity of $\mathcal{Z}_{\text{ad}}$ or at least weak closedness.

Existence of an Optimal Solution

Important Points

- This is a standard argument from the calculus of variations with a slight twist due to the composite functional $\mathcal{R} \circ J \circ S$.
- Convexity of $\mathcal{Z}_{\text{ad}}$ or at least weak closedness.
- Form of f (e.g., CVaR is monotone) leads to weak inf-compactness.

Existence of an Optimal Solution

Important Points

- This is a standard argument from the calculus of variations with a slight twist due to the composite functional $\mathcal{R} \circ J \circ S$.
- Convexity of $\mathcal{Z}_{\text{ad}}$ or at least weak closedness.
- Form of f (e.g., CVaR is monotone) leads to weak inf-compactness.
- Z must be at least reflexive. For function-space-based algorithms Z should be a Hilbert space.

Existence of an Optimal Solution

Important Points

- This is a standard argument from the calculus of variations with a slight twist due to the composite functional $\mathcal{R} \circ J \circ S$.
- Convexity of $\mathcal{Z}_{\text{ad}}$ or at least weak closedness.
- Form of f (e.g., CVaR is monotone) leads to weak inf-compactness.
- Z must be at least reflexive. For function-space-based algorithms Z should be a Hilbert space.
- We need some form of compactness beyond what the Sobolev embedding theorems might give us for the deterministic solutions.

Existence of an Optimal Solution

Important Points

- This is a standard argument from the calculus of variations with a slight twist due to the composite functional $\mathcal{R} \circ J \circ S$.
- Convexity of $\mathcal{Z}_{\mathrm{ad}}$ or at least weak closedness.
- Form of f (e.g., CVaR is monotone) leads to weak inf-compactness.
- Z must be at least reflexive. For function-space-based algorithms Z should be a Hilbert space.
- We need some form of compactness beyond what the Sobolev embedding theorems might give us for the deterministic solutions.
- This was done by assuming pointwise complete continuity of B. Together with the standing assumptions, this actually amounts to a stronger property for B.

Existence of an Optimal Solution

Important Points

- This is a standard argument from the calculus of variations with a slight twist due to the composite functional $\mathcal{R} \circ J \circ S$.
- Convexity of $\mathcal{Z}_{\mathrm{ad}}$ or at least weak closedness.
- Form of f (e.g., CVaR is monotone) leads to weak inf-compactness.
- Z must be at least reflexive. For function-space-based algorithms Z should be a Hilbert space.
- We need some form of compactness beyond what the Sobolev embedding theorems might give us for the deterministic solutions.
- This was done by assuming pointwise complete continuity of B. Together with the standing assumptions, this actually amounts to a stronger property for B.
- We work with the case $q = 2$ even though more regularity is possible depending on B and f.

Existence of an Optimal Solution

Important Points

- This is a standard argument from the calculus of variations with a slight twist due to the composite functional $\mathcal{R} \circ J \circ S$.
- Convexity of $\mathcal{Z}_{\text{ad}}$ or at least weak closedness.
- Form of f (e.g., CVaR is monotone) leads to weak inf-compactness.
- Z must be at least reflexive. For function-space-based algorithms Z should be a Hilbert space.
- We need some form of compactness beyond what the Sobolev embedding theorems might give us for the deterministic solutions.
- This was done by assuming pointwise complete continuity of B. Together with the standing assumptions, this actually amounts to a stronger property for B.
- We work with the case $q = 2$ even though more regularity is possible depending on B and f.
- Continuity of the risk measure follows from convex analysis.

Main Components

Derivation in the Abstract Setting

- Let $z^\star$ be the optimal solution, $z \in \mathcal{Z}_{\text{ad}}$, $\tau \in (0,1)$.
- Then by optimality of $z^\star$ and convexity, we have

$$f(z^\star) \le f(z^\star + \tau(z - z^\star)) \Rightarrow 0 \le \frac{f(z^\star + \tau(z - z^\star)) - f(z^\star)}{\tau}$$

- Passing to the limit as $\tau \downarrow 0$ yields

$$0 \le f'(z^\star; z - z^\star) \quad z \in \mathcal{Z}_{\text{ad}}$$

provided f is directionally differentiable.

Main Components

Derivation in the Concrete Setting

- We need only investigate the difference quotients

$$\frac{1}{2\tau}\left[\mathrm{CVaR}_\beta[\|S(z^\star+\tau(z-z^\star))-u_d\|^2]-\mathrm{CVaR}_\beta[\|S(z)-u_d\|^2]\right],$$
$$\frac{\alpha}{2\tau}\left[\|z^\star+\tau(z-z^\star)\|_Z^2-\|z^\star\|_Z^2\right]$$

Main Components

Derivation in the Concrete Setting

- We need only investigate the difference quotients

$$\frac{1}{2\tau}\left[\mathrm{CVaR}_\beta[\|S(z^\star+\tau(z-z^\star))-u_d\|^2]-\mathrm{CVaR}_\beta[\|S(z)-u_d\|^2]\right],$$
$$\frac{\alpha}{2\tau}\left[\|z^\star+\tau(z-z^\star)\|_Z^2-\|z^\star\|_Z^2\right]$$

- The latter term is unproblematic and yields (as expected):

$$\alpha(z^\star,z-z^\star)_Z+\tau\frac{\alpha}{2}\|z-z^\star\|_Z^2.$$

Main Components

Derivation in the Concrete Setting

- We need only investigate the difference quotients

$$\frac{1}{2\tau}\left[\mathrm{CVaR}_\beta[\|S(z^\star+\tau(z-z^\star))-u_d\|^2]-\mathrm{CVaR}_\beta[\|S(z)-u_d\|^2]\right],$$
$$\frac{\alpha}{2\tau}\left[\|z^\star+\tau(z-z^\star)\|_Z^2-\|z^\star\|_Z^2\right]$$

- The latter term is unproblematic and yields (as expected):

$$\alpha(z^\star,z-z^\star)_Z+\tau\frac{\alpha}{2}\|z-z^\star\|_Z^2.$$

- Since $(J\circ S)$ is continuously Fréchet differentiable from Z into $L^1(\Omega,\mathcal{F},\mathbb{P})$, we only need a differentiability concept for CVaR_β that is strong enough to allow from a chain rule.

Main Components

Derivation in the Concrete Setting

- We need only investigate the difference quotients

$$\frac{1}{2\tau}\left[\mathrm{CVaR}_\beta[\|S(z^\star+\tau(z-z^\star))-u_d\|^2]-\mathrm{CVaR}_\beta[\|S(z)-u_d\|^2]\right],$$
$$\frac{\alpha}{2\tau}\left[\|z^\star+\tau(z-z^\star)\|_Z^2-\|z^\star\|_Z^2\right]$$

- The latter term is unproblematic and yields (as expected):

$$\alpha(z^\star,z-z^\star)_Z+\tau\frac{\alpha}{2}\|z-z^\star\|_Z^2.$$

- Since $(J\circ S)$ is continuously Fréchet differentiable from Z into $L^1(\Omega,\mathcal{F},\mathbb{P})$, we only need a differentiability concept for CVaR_β that is strong enough to allow from a chain rule.
- It can be shown in that

$$\mathcal{R}'[X;H]=\sup_{\vartheta\in\partial\mathcal{R}[X]}\mathbb{E}[\vartheta X],$$

where $\mathcal{R}'[X;H]$ is the **Hadamard directional derivative** and $\partial\mathcal{R}[X]$ is the **subdifferential**.

Primal First-Order Optimality Conditions

Combining these observations, we can readily prove the following result.

Proposition

Under the standing assumptions, the following first-order necessary (and sufficient) optimality condition holds for a solution $z^\star$:

$$\sup_{\vartheta \in \partial \mathcal{R}(J(S(z^\star)))} \mathbb{E}[(S(z^\star) - u_d, S'(z^\star)(z - z^\star))_{L^2}\, \vartheta] + \alpha(z^\star, z - z^\star)_Z \geq 0, \ \forall z \in \mathcal{Z}_{\mathrm{ad}}.$$

Dual First-Order Optimality Conditions

- It's hard to make use of the primal optimality condition.
- Introducing the bounded linear operators **A**, **B** and functional **f**, where

$$\langle \mathbf{A}w, v\rangle = \mathbf{a}(w, v), \quad \langle \mathbf{B}z, v\rangle = \mathbb{E}_{\mathbb{P}}[\langle B(\cdot)z, v(\cdot)\rangle], \quad \langle \mathbf{f}, v\rangle = \mathbb{E}_{\mathbb{P}}[\langle f(\cdot), v(\cdot)\rangle],$$

for $w, v \in \mathcal{U}, z \in Z$, we can "unfold" the primal system into a dual optimality system.

Dual First-Order Optimality Conditions

- It's hard to make use of the primal optimality condition.
- Introducing the bounded linear operators **A**, **B** and functional **f**, where

$$\langle \mathbf{A}w, v\rangle = \mathbf{a}(w, v), \quad \langle \mathbf{B}z, v\rangle = \mathbb{E}_{\mathbb{P}}[\langle B(\cdot)z, v(\cdot)\rangle], \quad \langle \mathbf{f}, v\rangle = \mathbb{E}_{\mathbb{P}}[\langle f(\cdot), v(\cdot)\rangle],$$

  for $w, v \in \mathcal{U}, z \in Z$, we can "unfold" the primal system into a dual optimality system.
- If $z^\star$ is an optimal solution, then there exist $u^\star, \lambda^\star, \vartheta^\star$ such that

$$\begin{aligned} (\alpha z^\star + \mathbb{E}[\mathbf{B}^*\lambda^\star\vartheta^\star], z - z^\star)_Z \geq 0,\ \forall z \in \mathcal{Z}_{\mathrm{ad}} \\ \mathcal{R}[Y] - \mathcal{R}[\mathcal{J}(u^\star)] - \mathbb{E}[\vartheta^\star(Y - \mathcal{J}(u^\star))] \geq 0,\ \forall Y \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \\ \mathbf{A}u^\star = \mathbf{B}z^\star + \mathbf{f} \\ \mathbf{A}^*\lambda^\star = u^\star - u_d. \end{aligned}$$

- The optimal control is the projection of $-\frac{1}{\alpha}\mathbb{E}[\vartheta^\star B^*\lambda^\star]$ onto $\mathcal{Z}_{\mathrm{ad}}$.
- **Without stochasticity**, this is what we would normally expect.
- **With stochasticity** and risk-aversion, we have the risk-adjusted average of the adjoint states.

Summary Part I

- PDE-constrained optimization under uncertainty presents a number of exciting challenges from theory to computation.
- In addition to the usual workflow, several additional considerations appear including: measurability issues, modeling risk preferences, and stochastic optimization algorithms for infinite-dimensional problems.
- Faced with uncertainty, we need a way to obtain solution that are resilient to outlier events. We choose to do this with risk measures from management theory.
- The problems should be numerically tractable, the optimization model intuitive, and the solutions plausible.
- We sketched the main ideas for proving existence of solutions and deriving optimality conditions. This is somewhat more difficult than the standard setting due to the nonsmoothness of CVaR_β.
- Part II will be dedicated to algorithms and the numerical solution of the canonical problem.

References: Papers on optimization of optoelectronic devices

These non-exhaustive lists are meant to give you a head start on the literature!

L. Adam, M. Hintermüller, D. Peschka, and T.M. Surowiec.
Optimization of a multiphysics problem in semiconductor laser design.
SIAM Journal on Applied Mathematics, 79(1):257–283, (2019).

L. Adam, M. Hintermüller, and T.M. Surowiec.
A semismooth Newton method with analytical path-following for the H^1-projection onto the Gibbs simplex.
IMA Journal of Numerical Analysis 39(3):1276–1295, (2019).

L. Adam, M. Hintermüller, and T.M. Surowiec.
A PDE-constrained optimization approach for topology optimization of strained photonic devices.
Optimization and Engineering 19(3), 521–557, (2018).

D. Peschka, N. Rotundo, and M. Thomas.
Doping optimization for optoelectronic devices.
Optical and Quantum Electronics, 50(3):125, (2018).

D. Peschka, M. Thomas, A. Glitzky, R. Nürnberg, K. Gärtner, M. Virgilio, S. Guha, Th. Schroeder, G. Capellini, and Th. Koprucki.
Modeling of edge-emitting lasers based on tensile strained germanium microstrips.
IEEE Photonics Journal, 7(3):1–15, (2015).

References: Books on PDE-constrained optimization

Lions, J.-L.
Optimal control of systems governed by partial differential equations.
Springer-Verlag, New York-Berlin, 1971.

Tröltzsch, F.
Optimal control of partial differential equations. Theory, methods and applications.
American Mathematical Society, Providence, RI, 2010.

Borzì, A. and Schulz, V.
Computational optimization of systems governed by partial differential equations.
Computational Science & Engineering, 8. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2012.

De los Reyes, J.C.
Numerical PDE-constrained optimization.
SpringerBriefs in Optimization. Springer, Cham, 2015.

Hinze, M., Pinnau, R., Ulbrich, M., and Ulbrich, S.
Optimization with PDE constraints.
Mathematical Modelling: Theory and Applications, 23. Springer, New York, 2009

Antil, A., Kouri, D.P., Lacasse, M.-D., and Ridzal, D. (eds.)
Frontiers in PDE-constrained optimization.
Papers based on the workshop held at the Institute for Mathematics and its Applications, Minneapolis, MN, June 610, 2016. The IMA Volumes in Mathematics and its Applications, 163. Springer, New York, 2018.

References: Books on stochastic optimization

Birge, J. R. and Louveaux, F.,
Introduction to stochastic programming.
Springer-Verlag, New York, 1997.

Shapiro, A., Dentcheva, D. and Ruszczyński, A..
Lectures on Stochastic Programming: Modeling and Theory.
SIAM, Philadelphia, 2009.

Ben-Tal, A., El Ghaoui, L., Nemirovski, A.
Robust optimization.
Princeton University Press, Princeton, NJ, 2009.

Prékopa, A.
Stochastic programming.
Mathematics and its Applications, 324. Kluwer Academic Publishers Group, Dordrecht, 1995.

Lan, G.
First-order and stochastic optimization methods for machine learning.
Springer, Cham, 2020.

Ruszczyński, A. and Shapiro, A. (eds.)
Handbooks in Operations Reseach & Management Science, Vol. 10 Elsevier Science B.V., 2003.

References: Books on optimization theory and PDEs

Attouch, H., Buttazzo, G. and Michaille, G.
Variational analysis in Sobolev and BV spaces,
Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006.

Bonnans, J. F. and Shapiro, A.
Perturbation Analysis of Optimization Problems.
Springer Verlag, Berlin, Heidelberg, New York, 2000.

Evans, L.C.
Partial differential equations.
American Mathematical Society, Providence, RI, 2010.

Gilbarg, D. and Trudinger, N. S.
Elliptic partial differential equations of second order.
Reprint of the 1998 edition. Classics in Mathematics. Springer-Verlag, Berlin, 2001.

Grisvard, P.
Elliptic problems in nonsmooth domains.
Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.

Luenberger, D.G.
Optimization by vector space methods.
John Wiley & Sons, Inc., New York-London-Sydney, 1969.

References: Books on numerical methods for PDEs

BRENNER, S.C. AND SCOTT, L.R.
The mathematical theory of finite element methods.
Springer, New York, 2008.

ERN, A. AND GUERMOND, J.-L.
Theory and practice of finite elements.
Springer-Verlag, New York, 2004.

BRAESS, D.
Finite elements. Theory, fast solvers, and applications in elasticity theory.
Cambridge University Press, Cambridge, 2007.

BARTELS, S.
Numerical methods for nonlinear partial differential equations.
Springer, Cham, 2015.

References: Applications and numerics of random PDE

I. Babuška, R. Tempone, and G. E. Zouraris.
Galerkin finite element approximations of stochastic elliptic partial differential equations.
SIAM J. Numer. Anal., 42 (2004), pp. 800–825.

I. Babuška, R. Tempone, and G. E. Zouraris.
Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation.
Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 1251–1294.

I. Babuška, F. Nobile, and R. Tempone.
A stochastic collocation method for elliptic partial differential equations with random input data.
SIAM Rev., 52 (2010), pp. 317–355.

L. Bonfiglio, P. Perdikaris, S. Brizzolara, and G. Karniadakis.
A multi-fidelity framework for investigating the performance of super-cavitating hydrofoils under uncertain flow conditions.
in 19th AIAA Non-Deterministic Approaches Conference, American Institute of Aeronautics and Astronautics, 2017.

L. J. Durlofsky and Y. Chen.
Uncertainty Quantification for Subsurface Flow Problems Using Coarse-Scale Models,
Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 163–202.

G. E. Karniadakis, C.-H. Su, D. Xiu, D. Lucor, C. Schwab, and R. A. Todor.
Generalized polynomial chaos solution for differential equations with random inputs.
Tech. Report 2005-01, Seminar for Applied Mathematics, ETH Zurich, Zurich, Switzerland, 2005.

B. Khoromskij and C. Schwab.
Tensor-structured galerkin approximation of parametric and stochastic elliptic PDEs.
SIAM J. Sci. Comput., 33 (2011), pp. 364–385.

References: Applications and numerics of random PDE

F. Nobile, R. Tempone, and C. G. Webster
An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data.
SIAM J. Numer. Anal., 46 (2008), pp. 2411–2442.

F. Nobile, R. Tempone, and C. G. Webster
A sparse grid stochastic collocation method for partial differential equations with random input data.
SIAM J. Numer. Anal., 46 (2008), pp. 2309–2345.

G. Stadler, M. Gurnis, C. Burstedde, L. C. Wilcox, L. Alisic, and O. Ghattas.
The dynamics of plate tectonics and mantle flow: From local to global scales.
Science, 329 (2010), pp. 1033–1038.

C. Schwab and C. J. Gittelson.
Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs.
Acta Numerica, 20 (2011), pp. 291–467.

D. M. Tartakovsky, A. Guadagnini, and M. Riva.
Stochastic averaging of nonlinear flows in heterogeneous porous media.
Journal of Fluid Mechanics, 492 (2003), pp. 47–62.

D. Xiu and J. S. Hesthaven.
High-order collocation methods for differential equations with random inputs.
SIAM J. Sci. Comput., 27 (2005), pp. 1118–1139.

D. Xiu and G. E. Karniadakis.
Modeling uncertainty in flow simulations via generalized polynomial chaos.
J. Comput. Phys., 187 (2003), pp. 137–167.

References: Risk Neutral PDE-Constrained Optimization $\mathcal{R} = \mathbb{E}$

A. A. Ali, E. Ullmann, and M. Hinze,
Multilevel Monte Carlo analysis for optimal control of elliptic PDEs with random coefficients.
SIAM/ASA J. Uncertain. Quantif. 5 (2017), no. 1, 466492.

A. Borzì, V. Schulz, C. Schillings, and G. von Winckel.
On the treatment of distributed uncertainties in PDE-constrained optimization.
GAMM Mitteilungen, 33 (2010), pp. 230–246.

A. Borzì and G. von Winckel.
A POD framework to determine robust controls in PDE optimization.
Comput. Vis. Sci., 14 (2011), pp. 91–103.

P. Chen and A. Quarteroni.
Weighted reduced basis method for stochastic optimal control problems with elliptic PDE constraint
SIAM/ASA Journal on Uncertainty Quantification, 2 (2014), pp. 364–396.

P. Chen, A. Quarteroni, and G. Rozza.
Stochastic optimal robin boundary control problems of advection-dominated elliptic equations.
SIAM Journal on Numerical Analysis, 51 (2013), pp. 2700–2722.

M. H. Farshbaf-Shaker, R. Henrion, and D. Hömberg.
Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization.
Set-Valued Var. Anal., 26(4):821–841, 2018.

S. Garreis and M. Ulbrich.
Constrained optimization with low-rank tensors and applications to parametric problems with PDEs.
SIAM J. Sci. Comput. 39 (2017), no. 1, A25A54.

References: Risk Neutral PDE-Constrained Optimization $\mathcal{R} = \mathbb{E}$

C. Geiersbach, C. and G. Ch. Pflug.
Projected stochastic gradients for convex constrained problems in Hilbert spaces.
SIAM J. Optim. 29 (2019), no. 3, 20792099.

C. Geiersbach and W. Wollner.
A stochastic gradient method with mesh refinement for PDE-constrained optimization under uncertainty.
SIAM J. Sci. Comput. 42 (2020), no. 5, A2750A2772.

C. Geiersbach and W. Wollner.
Optimality conditions for convex stochastic optimization problems in Banach spaces with almost sure state constraint.
WIAS Preprint No. 2755, DOI: 10.20347/WIAS.PREPRINT.2755, 2020.

M. Hoffhues, W. Römisch, T. M. Surowiec
On Quantitative Stability in Infinite-Dimensional Optimization under Uncertainty
Optimization Letters. https://doi.org/10.1007/s11590-021-01707-2

D. P. Kouri.
A multilevel stochastic collocation algorithm for optimization of PDEs with uncertain coefficients.
SIAM/ASA Journal on Uncertainty Quantification, 2 (2014), pp. 55–81.

D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders.
A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty.
SIAM Journal on Scientific Computing, 35 (2013), pp. A1847–A1879.

W. Römisch and T. .M. Surowiec.
Asymptotic Properties of Monte Carlo Methods in Elliptic PDE-Constrained Optimization under Uncertainty
Preprint arXiv:2106.06347 https://arxiv.org/abs/2106.06347

References: Risk Neutral PDE-Constrained Optimization $\mathcal{R} = \mathbb{E}$

D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders.
Inexact objective function evaluations in a trust-region algorithm for PDE-constrained optimization under uncertainty.
SIAM Journal on Scientific Computing, 36 (2014), pp. A3011–A3029.

J. Milz
Sample average approximations of strongly convex stochastic programs in Hilbert spaces.
Preprint arXiv:2104.05114v1.

N. Petra, H. Zhu, G. Stadler, T. J. R. Hughes, and O. Ghattas.
An inexact Gauss-Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model.
Journal of Glaciology, 58 (2012).

V. Schulz and C. Schillings.
On the nature and treatment of uncertainties in aerodynamic design.
AIAA Journal, 47 (2009), pp. 646–654.

H. Tiesler, R. M. Kirby, D. Xiu, and T. Preusser.
Stochastic collocation for optimal control problems with stochastic PDE constraints.
SIAM Journal on Control and Optimization, 50 (2012), pp. 2659–2682.

References: Risk Averse PDE-Constrained Optimization

A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas.
Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations.
SIAM/ASA J. Uncertainty Quantification, 5(1), (2017) pp. 1166–1192.

H. Antil, D. .P. Kouri, and J. Pfefferer.
Risk-Averse Control of Fractional Diffusion with Uncertain Exponent.
SIAM J. Control Optim., 59(2), (2021) 11611187.

S. Garreis, T. M. Surowiec, and M. Ulbrich.
An interior-point approach for solving risk-averse PDE-constrained optimization problems with Coherent Risk Measures
To appear in SIAM Journal on Optimization

M. Heinkenschloss, B. Kramer, T. Takhtaganov, and K. Willcox.
Conditional-value-at-risk estimation via reduced-order models.
SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 1395–1423.

O. Lass and S. Ulbrich.
Model order reduction techniques with a posteriori error control for nonlinear robust optimization governed by partial differential equations.
SIAM J. Sci. Comput., 39 (2017), pp. S112–S139.

J. Milz and M. Ulbrich.
An Approximation Scheme for Distributionally Robust PDE-Constrained Optimization.
Preprint No. IGDK-2020-09, Technische Universitt Mnchen, 2020.

P. Kolvenbach, O. Lass, and S. Ulbrich.
An approach for robust PDE-constrained optimization with application to shape optimization of electrical engines and of dynamic elastic structures under uncertainty.
Optim. Eng., 19 (2018), pp. 697–731.

References: Risk Averse PDE-Constrained Optimization

D. P. Kouri
A measure approximation for distributionally robust PDE-constrained optimization problems.
SIAM Journal on Numerical Analysis, 55 (2017), pp. 3147–3172.

D. P. Kouri and T. M. Surowiec.
Risk-averse PDE-constrained optimization using the conditional value-at-risk.
SIAM Journal on Optimization 26, 1 (2016), 365–396.

D. P. Kouri and T. M. Surowiec.
A primal-dual algorithm for risk minimization
To appear in Mathematical Programming Ser. A

D. P. Kouri and T. M. Surowiec.
Epi-Regularization of Risk Measures
Mathematics of Operations Research 45, 2 (2020), 774–795

D. P. Kouri and T. M. Surowiec.
Risk-averse optimal control of semilinear elliptic PDEs
ESAIM: Control, Optimisation and Calculus of Variations 26, 53 (2020).

D. P. Kouri and T. M. Surowiec.
Existence and optimality conditions for risk-averse PDE-constrained optimization
SIAM/ASA J. Uncertainty Quantification 6, 2 (2018), 787–815.

Z. Zou, D. P. Kouri, and W. Aquino.
An adaptive local reduced basis method for solving PDEs with uncertain inputs and evaluating risk.
Comput. Methods Appl. Mech. Engrg., 345 (2019), pp. 302–322.