An Approximation Scheme for Distributionally Robust Nonlinear Optimization with Applications to PDE-Constrained Problems under Uncertainty

Michael Ulbrich

Technical University of Munich

joint work with Johannes Milz

IGDK 1754

One World Optimization Seminar

April 12, 2021

Outline

- Distributionally robust nonlinear optimization problem (DROP)
- Tractability (in an NLP sense) via quadratic (Taylor) expansion of cost function and constraints w.r.t. random parameters

  → Approximated DROP
- Resulting cost and constraint functions are sums of value functions of a semidefinite program and a trust region problem
- Construction of smoothing functions for both value functions

  → Smoothed approximated DROP
- Analysis of DROP as well as of approximated and smoothed DROP
- Continuation method that drives smoothing parameters to zero
- Convergence results
- DROP with PDE constraints
- Application to DRO for steady and unsteady Burgers equation

Distributionally Robust Nonlinear Optimization

Distributionally Robust NLP:

$$\min_{x \in X} \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)] \quad \text{s.t.} \quad \sup_{P \in \mathcal{P}} \mathbb{E}_P[g_i(x,\xi)] \le 0 \qquad \text{(DROP)}$$

- $\mathcal{P}$ ambiguity set of probability distributions on Ξ ($= \mathbb{R}^p$ in the following)

  → DRO targets at "robustifying" against distributional uncertainty
- in this talk mostly $X = \mathbb{R}^n$; later X, is a Hilbert space of controls
- $f : X \times \Xi \to \mathbb{R}$ parametric cost function
- $g_i : X \times \Xi \to \mathbb{R}$ parametric constraint functions, $i \in [m] := \{1, \dots, m\}$
- we could have further "deterministic" constraints $x \in X_{\text{ad}}$ with $X_{\text{ad}} \subset X$ closed

Distributionally Robust Nonlinear Optimization

Distributionally Robust NLP:

$$\min_{x\in X} \sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)] \quad \text{s.t.} \quad \sup_{P\in\mathcal{P}} \mathbb{E}_P[g_i(x,\xi)] \le 0 \qquad \text{(DROP)}$$

- $\mathcal{P}$ ambiguity set of probability distributions on Ξ ($= \mathbb{R}^p$ in the following)
  → DRO targets at "robustifying" against distributional uncertainty
- in this talk mostly $X = \mathbb{R}^n$; later X, is a Hilbert space of controls
- $f : X \times \Xi \to \mathbb{R}$ parametric cost function
- $g_i : X \times \Xi \to \mathbb{R}$ parametric constraint functions, $i \in [m]$

We are particularly interested in cases where f (and possibly g_i) is *complicated* (e.g., only *implicitly given*).

Example: Optimal control problems

- x is the control
- $f(x,\xi) = J(S(x,\xi),x,\xi)$, where S is the control-to-state operator, i.e.:
  State $y(\xi) = S(x,\xi)$ corresponding to $x \in U$ solves parametric state equation

$$e(y(\xi),x,\xi) = 0, \quad \xi \in \Xi.$$

Distributionally Robust Nonlinear Optimization (2)

- DROP is a robust version of a

  Stochastic Optimization Problem (SOP):

$$\min_{x \in X} \mathbb{E}_P[f(x,\xi)] \quad \text{s.t} \quad \mathbb{E}_P[g_i(x,\xi)] \le 0, \quad 1 \le i \le m,$$

  where distribution P is known and fixed
- Thus, SOP corresponds to a DROP with $\mathcal{P} = \{P\}$
- Robust Optimization Problems with uncertainty set $\mathcal{U} \subset \Xi$

$$\min_{x \in X} \sup_{\xi \in \mathcal{U}} f(x,\xi) \quad \text{s.t} \quad \sup_{\xi \in \mathcal{U}} g_i(x,\xi) \le 0, \quad 1 \le i \le m,$$

  can be viewed as DROPs with $\mathcal{P} = \{\delta_\xi : \xi \in \mathcal{U}\}$

Distributionally Robust Nonlinear Optimization (3)

- We explain the main ideas for the following

  Distributionally Robust Optimization Problem (DROP)

$$\min_{x \in X} \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)]$$

- Distributionally robust constraints can be handled by the same techniques

We present

- A sampling-free tractable approximation of $x \mapsto \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)]$
- A numerical strategy for solving the resulting approximated DROP which combines a smoothing approach with a continuation method and allows to use standard derivative-based NLP solvers

- Milz and M.U., SIOPT, 2020
  → DRO in finite dimensions; convergence of sequence of stationary points
- Milz and M.U., 2020, submitted to SICON, under revision
  → DRO with PDE constraints

Some Literature on Optimization under Uncertainty

Robust optimization ($\min_x \sup_{\xi \in \mathcal{U}} f(x, \xi)$, $\mathcal{U} \subset \Xi \Leftrightarrow \mathcal{P} = \{\delta_\xi : \xi \in \mathcal{U}\}$)

- Ben-Tal, Nemirovski 1998; Ben-Tal, El Ghaoui, Nemirovski 2009; Diehl, Bock, Kostina 2006; Hale, Zhang 2007; de Gournay, Allaire, Jouve 2008; Sichau, Ulbrich 2012; Ben-Tal, den Hertog 2014; Lass, Ulbrich 2017; Kolvenbach, Lass, Ulbrich 2018; Alla, Hinze, Kolvenbach, Lass, Ulbrich 2019; …

Distributionally robust optimization

- moment constraints: Scarf 1957; Shapiro, Kleywegt 2002; Popescu 2007; Delage, Ye 2010; Goh, Sim 2010; Wiesemann, Kuhn, Sim 2014; …
- distance to reference measure: Pflug, Wozabal 2007; Gao, Kleywegt 2016; Shapiro 2017; Esfahani, Kuhn 2018; …

Risk averse PDE-constrained optimization

- Borzì, von Winckel 2009; Conti, Held, Pach, Rumpf, Schultz 2011; Kouri et al. (2013); Chen, Quarteroni 2014; Kouri 2017; Alexanderian, Petra, Stadler, Ghattas 2017; Kouri, Surowiec 2018; Chen, Villa, Ghattas 2018; Kouri, Shapiro 2018; Van Barel, Vandewalle 2019; Kouri, Surowiec 2020; Garreis, Surowiec, Ulbrich 2021; …

Ambiguity Set

- Moment constraints and entropic dominance modeling confidence regions; cf. Delage, Ye 2010; So 2011; Chen, Sim, Xu 2019:

$$\mathcal{P} = \{P : \ \|\bar{\Sigma}^{-\frac{1}{2}}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 \le \Delta,$$

(trust-region for mean)

$$\bar{\Sigma}_0 \preccurlyeq \mathrm{Cov}_P[\xi] \preccurlyeq \bar{\Sigma}_1,$$

(semidefinite box constraints on covariance)

$$\ln \mathbb{E}_P\left[\exp\left(y^T(\xi - \mathbb{E}_P[\xi])\right)\right] \le y^T \bar{\Sigma}_1 y, \quad \forall y \in \mathbb{R}^p$$

(implies existence of all finite moments)$\}$,

where $\Delta > 0$, $\bar{\Sigma} \succ 0$, and $0 \preccurlyeq \bar{\Sigma}_0 \preccurlyeq \bar{\Sigma} \preccurlyeq \bar{\Sigma}_1$ in $\mathbb{S}^p$
($\mathbb{S}^p$ sym. $p \times p$-matrices, $\preccurlyeq$ Löwner partial order on $\mathbb{S}^p$)

- $\mathcal{P}$ contains the following set of normal distributions:

$$\{ P = N(\mu, \Sigma) : \quad \|\bar{\Sigma}^{-\frac{1}{2}}(\mu - \bar{\mu})\|_2 \le \Delta, \quad \bar{\Sigma}_0 \preccurlyeq \Sigma \preccurlyeq \bar{\Sigma}_1 \}.$$

- $\mathcal{P}$ is weak* closed and tight

Approximation Scheme for DROP

We require that $f(x,\cdot)$ is sufficiently smooth for all $x \in X$.

Our approximation scheme consists of the following steps:

1. $f(x,\cdot)$ is approximated by 2nd order Taylor's expansion $Q(x,\cdot)$ about $\bar{\mu}$:

$$Q(x,\xi) := f(x,\bar{\mu}) + \nabla_\xi f(x,\bar{\mu})^T(\xi-\bar{\mu}) + \tfrac{1}{2}(\xi-\bar{\mu})^T\nabla_{\xi\xi} f(x,\bar{\mu})(\xi-\bar{\mu}),$$

Alternatively, we could use any other quadratic approximation $Q(x,\cdot)$ of $f(x,\cdot)$.

2. We assume and use: $\Psi(x) = \sup_{P\in\mathcal{P}} \mathbb{E}_P[Q(x,\xi)] \approx \sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)]$

→ Approximated DROP: $\min_{x\in X} \Psi(x)$ with $\Psi(x) = \sup_{P\in\mathcal{P}} \mathbb{E}_P[Q(x,\xi)]$

3. Important observation: Ψ is sum of two tractable value functions
4. We construct a smoothing function $\tilde{\Psi}$ for Ψ → tractable smoothed DROPs
5. Smoothing parameters are driven to 0 within a continuation method
6. Study convergence of continuation sequence to solution of approx. DROP

Approximation Scheme for DROP

We require that $f(x,\cdot)$ is sufficiently smooth for all $x \in X$.

Our approximation scheme consists of the following steps:

1. $f(x,\cdot)$ is approximated by 2nd order Taylor's expansion $Q(x,\cdot)$ about $\bar{\mu}$:

$$Q(x,\xi) := f(x,\bar{\mu}) + \nabla_\xi f(x,\bar{\mu})^T(\xi-\bar{\mu}) + \tfrac{1}{2}(\xi-\bar{\mu})^T\nabla_{\xi\xi} f(x,\bar{\mu})(\xi-\bar{\mu}),$$

Alternatively, we could use any other quadratic approximation $Q(x,\cdot)$ of $f(x,\cdot)$.

2. We assume and use: $\Psi(x) = \sup_{P\in\mathcal{P}} \mathbb{E}_P[Q(x,\xi)] \approx \sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)]$

→ Approximated DROP: $\min_{x\in X} \Psi(x)$ with $\Psi(x) = \sup_{P\in\mathcal{P}} \mathbb{E}_P[Q(x,\xi)]$

3. Important observation: Ψ is sum of two tractable value functions
4. We construct a smoothing function $\tilde{\Psi}$ for Ψ → tractable smoothed DROPs
5. Smoothing parameters are driven to 0 within a continuation method
6. Study convergence of continuation sequence to solution of approx. DROP

Taylor expansions also used for robust / risk-measure-based optimization in, e.g.,

Rockafellar, Royset 2010; Ben-Tal, den Hertog 2014; Alexanderian, Petra, Stadler, Ghattas 2017; Lass, Ulbrich 2017; Chen, Villa, Ghattas 2018; Kolvenbach, Lass, Ulbrich 2018

Tractable but Nonsmooth Optimal Value Functions

- Since $\xi \mapsto Q(x,\xi)$ is quadratic, we obtain that

$$\Psi(x) = \sup_{P\in\mathcal{P}} \mathbb{E}_P[Q(x,\xi)] = \Psi_{\mathrm{SDP}}(x) + \Psi_{\mathrm{TRP}}(x)$$

with the semidefinite program (SDP)

$$\Psi_{\mathrm{SDP}}(x) = \max_{\Sigma\in\mathbb{S}^p} \left\{ \tfrac{1}{2}\nabla_{\xi\xi} f(x,\bar{\mu}) \bullet \Sigma : \quad \bar{\Sigma}_0 \preccurlyeq \Sigma \preccurlyeq \bar{\Sigma}_1 \right\}$$

and the nonconvex trust-region problem (TRP)

$$\Psi_{\mathrm{TRP}}(x) = \max_{\mu\in\mathbb{R}^p} \left\{ Q(x,\mu) : \quad \|\bar{\Sigma}^{-\frac{1}{2}}(\mu - \bar{\mu})\|_2 \le \Delta \right\}.$$

($A \bullet B = \mathrm{trace}[A^T B]$ Frobenius inner product)

We obtain value functions defined by semidefinite programs and trust-region problems which are computationally tractable, however, nonsmooth.

Continuation Method

- We consider the approximated DROP

$$\min_{x\in X} \Psi(x), \quad \text{where} \quad \Psi(x) = \sup_{P\in\mathcal{P}} \mathbb{E}_P[Q(x,\xi)] = \Psi_{\text{SDP}}(x) + \Psi_{\text{TRP}}(x)$$

- Let $\tilde{\Psi}_{\text{SDP}}(\cdot;\tau)$ and $\tilde{\Psi}_{\text{TRP}}(\cdot;\nu,\eta)$ be smoothing functions for Ψ_{SDP} and Ψ_{TRP}, respectively, with smoothing parameters $\tau,\nu,\eta > 0$.

Continuation method

Choose $\rho_1 = (\tau_1,\nu_1,\eta_1) \in \mathbb{R}^3_{++}$ and $x^0 \in X$.

For $k = 1, 2, \ldots$

1. Compute approx. stationary point x^k (starting from x^{k-1}) of smoothed DROP

$$\min_{x\in X} \underbrace{\tilde{\Psi}_{\text{SDP}}(x;\tau_k) + \tilde{\Psi}_{\text{TRP}}(x;\nu_k,\eta_k)}_{=:\tilde{\Psi}(x;\rho_k)}$$

2. Choose $0 < \rho_{k+1} \le \rho_k$ (componentwise), $\rho_{k+1} \ne \rho_k$ such that $\rho_k \to 0$

- Method based on Chen, Nashed, Qi 2000; Chen 2012
- Distrib. robust constraints will be approximated and smoothed in the same way
- We will present global convergence results

Smoothing Functions

- Our smoothing will involve $s = 3$ parameters, collected in a vector $\rho \in \mathbb{R}^3_{++}$
- $\tilde{\phi} : X \times \mathbb{R}^s_{++} \to \mathbb{R}$ is a smoothing function of the C^0-function $\phi : X \to \mathbb{R}$ if
  - $\tilde{\phi}(\cdot;\rho)$ is continuously differentiable for every $\rho \in \mathbb{R}^s_{++}$
  - There exists $\gamma : \mathbb{R}^m_{++} \to \mathbb{R}_+$ with $\lim_{\rho \to 0^+} \gamma(\rho) = 0$ such that:

$$|\tilde{\phi}(x;\rho) - \phi(x)| \leq \gamma(\rho) \quad \forall\ x \in X,\ \rho \in \mathbb{R}^m_{++}.$$

- Let $X = \mathbb{R}^n$ and ϕ locally Lipschitz with smoothing function $\tilde{\phi}$.

  Then gradient consistency holds if, for all x, we have

$$S_{\tilde{\phi}}(x) = \partial\phi(x) \ \ \text{(Clarke's differential)},$$

  where

$$S_{\tilde{\phi}}(x) = \text{conv}\{z \in \mathbb{R}^n : \exists\ \mathbb{R}^n \times \mathbb{R}^m_{++} \ni (x_k, \rho_k) \to (x, 0),\ \nabla_x \tilde{\phi}(x_k;\rho_k) \to z\}.$$

Note: Our results on smoothing functions will hold in H-spaces X, while our gradient consistency results will refer to $X = \mathbb{R}^n$.

Smoothing Function for Value Function of SDP

- We can apply the following explicit formula (cf. Xu, Sun, Qi 2011):

$$\Psi_{\mathrm{SDP}}(x) = \max\left\{ \tfrac{1}{2}\nabla_{\xi\xi} f(x,\bar{\mu}) \bullet \Sigma : \quad \bar{\Sigma}_0 \preccurlyeq \Sigma \preccurlyeq \bar{\Sigma}_1 \right\}$$

$$= \tfrac{1}{2}\nabla_{\xi\xi} f(x,\bar{\mu}) \bullet \bar{\Sigma}_0 + \tfrac{1}{2}\sum_{i=1}^{p} \max\{0, \lambda_i(B(x))\}$$

  with $B(x) = (\bar{\Sigma}_1 - \bar{\Sigma}_0)^{\frac{1}{2}} \nabla_{\xi\xi} f(x,\bar{\mu}) (\bar{\Sigma}_1 - \bar{\Sigma}_0)^{\frac{1}{2}}$, $\lambda_i(A) = i$th eigenvalue of A
- Entropy fct. $\ln(1 + e^{z/\tau})$, $\tau > 0$, is used to analytically approximate $\max\{0, z\}$
- Gradient consistent smoothing function $\tilde{\Psi}_{\mathrm{SDP}}$ is obtained by replacing $\max\{0, \cdot\}$ with $\ln(1 + e^{\cdot/\tau})$; proof via theory of spectral functions (Lewis 1996)

- Error estimate uniformly in $x \in X$: For all $(x, \tau) \in X \times (0, \infty)$:

$$\Psi_{\mathrm{SDP}}(x) \le \tilde{\Psi}_{\mathrm{SDP}}(x;\tau) \le \Psi_{\mathrm{SDP}}(x) + \tau p$$

(p = parameter dimension)

Evaluations of $x \mapsto \tilde{\Psi}_{\mathrm{SDP}}(x;\tau)$ and $x \mapsto \nabla_x \tilde{\Psi}_{\mathrm{SDP}}(x;\tau)$ are computationally tractable and are based on an eigendecomposition of $B(x)$

Smoothing Function for Value Function of TRP I

Consider maps $g : X \to \mathbb{R}^p$ and $H : X \to \mathbb{S}^p$.

We now show that a smoothing function for

$$v_{\text{TRP}}(x) = \max_{s\in\mathbb{R}^p} \left\{ \tfrac{1}{2} s^T H(x) s + g(x)^T s : \quad \tfrac{1}{2}\|s\|_2^2 \le \tfrac{1}{2}\Delta^2 \right\} \qquad \text{(TRP)}$$

is provided by the value function of the following lifted TRP:

$$\tilde{v}_{\text{TRP}}(x;\nu,\eta) = \max_{\tilde{s}\in\mathbb{R}^{p+2}} \left\{ \tfrac{1}{2} \tilde{s}^T \tilde{H}_\eta(x) \tilde{s} + \tilde{g}_\nu(x)^T \tilde{s} : \quad \tfrac{1}{2}\|\tilde{s}\|_2^2 \le \tfrac{1}{2}\Delta^2 \right\}, \qquad \text{(LTRP)}$$

where $(\nu,\eta) > 0$,

$$\tilde{H}_\eta(x) = \begin{bmatrix} H(x) & & \\ & 0 & \\ & & E(H(x);\eta) \end{bmatrix} \in \mathbb{S}^{p+2}, \quad \tilde{g}_\nu(x) = \begin{bmatrix} g(x) \\ \sqrt{2\nu} \\ \sqrt{2\nu} \end{bmatrix} \in \mathbb{R}^{p+2},$$

and

$$E(H;\eta) = \lambda_{\max}(H) + \eta \ln \sum_{i=1}^{p} \exp \frac{\lambda_i(H) - \lambda_{\max}(H)}{\eta} \approx \lambda_{\max}(H) \qquad \text{(entropy fct.)}$$

Smoothing Function for Value Function of TRP II

We motivate the approximation property

$$\boxed{\tilde{v}_{\mathrm{TRP}}(x;\nu,\eta) \approx v_{\mathrm{TRP}}(x)}$$

- For brevity, we omit the dependence on x

Smoothing Function for Value Function of TRP II

We motivate the approximation property

$$\tilde{v}_{\mathrm{TRP}}(x;\nu,\eta) \approx v_{\mathrm{TRP}}(x)$$

- For brevity, we omit the dependence on x
- Strong duality for TRPs (cf. Stern, Wolkowicz 1995) implies that

$$v_{\mathrm{TRP}} = \max\left\{\tfrac{1}{2}s^T H s + g^T s : \quad \tfrac{1}{2}\|s\|_2^2 \le \tfrac{1}{2}\Delta^2\right\}$$

Smoothing Function for Value Function of TRP II

We motivate the approximation property

$$\tilde{v}_{\mathrm{TRP}}(x;\nu,\eta) \approx v_{\mathrm{TRP}}(x)$$

- For brevity, we omit the dependence on x
- Strong duality for TRPs (cf. Stern, Wolkowicz 1995) implies that

$$\begin{aligned} v_{\mathrm{TRP}} &= \max\left\{\tfrac{1}{2}s^T Hs + g^T s : \quad \tfrac{1}{2}\|s\|_2^2 \le \tfrac{1}{2}\Delta^2\right\} \\ &= \inf\left\{d(\lambda) \;:\; \lambda - \lambda_{\max}(H) > 0, \quad \lambda \ge 0\right\} \end{aligned} \qquad \text{(strong duality)}$$

where $d(\lambda) = \frac{1}{2}g^T(\lambda I - H)^{-1}g + \frac{1}{2}\Delta^2\lambda \quad \text{if} \quad \lambda - \lambda_{\max}(H) > 0$

Smoothing Function for Value Function of TRP II

We motivate the approximation property

$$\boxed{\tilde{v}_{\mathrm{TRP}}(x;\nu,\eta) \approx v_{\mathrm{TRP}}(x)}$$

- For brevity, we omit the dependence on x
- Strong duality for TRPs (cf. Stern, Wolkowicz 1995) implies that

$$\begin{aligned} v_{\mathrm{TRP}} &= \max\left\{\tfrac{1}{2}s^T H s + g^T s : \quad \tfrac{1}{2}\|s\|_2^2 \le \tfrac{1}{2}\Delta^2\right\} \\ &= \inf\left\{d(\lambda) \;:\; \lambda - \lambda_{\max}(H) > 0, \quad \lambda \ge 0\right\} && \text{(strong duality)} \\ &\approx \inf\left\{d(\lambda) \;:\; \lambda - E(H;\eta) > 0, \quad \lambda \ge 0\right\} && \text{(smooth max. eigenvalue fct.)} \end{aligned}$$

where $d(\lambda) = \frac{1}{2}g^T(\lambda I - H)^{-1}g + \frac{1}{2}\Delta^2\lambda \quad \text{if} \quad \lambda - \lambda_{\max}(H) > 0$

Smoothing Function for Value Function of TRP II

We motivate the approximation property

$$\tilde{v}_{\text{TRP}}(x;\nu,\eta) \approx v_{\text{TRP}}(x)$$

- For brevity, we omit the dependence on x
- Strong duality for TRPs (cf. Stern, Wolkowicz 1995) implies that

$$\begin{aligned} v_{\text{TRP}} &= \max\left\{\tfrac{1}{2}s^T H s + g^T s : \quad \tfrac{1}{2}\|s\|_2^2 \le \tfrac{1}{2}\Delta^2\right\} \\ &= \inf\ \{d(\lambda)\ :\ \lambda - \lambda_{\max}(H) > 0, \quad \lambda \ge 0\} && \text{(strong duality)} \\ &\approx \inf\ \{d(\lambda)\ :\ \lambda - E(H;\eta) > 0, \quad \lambda \ge 0\} && \text{(smooth max. eigenvalue fct.)} \\ &\approx \min_{\lambda}\ d(\lambda) + \frac{\nu}{\lambda} + \frac{\nu}{\lambda - E(H;\eta)} && \text{(reciprocal barrier)} \end{aligned}$$

where $d(\lambda) = \frac{1}{2}g^T(\lambda I - H)^{-1}g + \frac{1}{2}\Delta^2\lambda \quad \text{if} \quad \lambda - \lambda_{\max}(H) > 0$

Smoothing Function for Value Function of TRP II

We motivate the approximation property

$$\tilde{v}_{\text{TRP}}(x;\nu,\eta) \approx v_{\text{TRP}}(x)$$

- For brevity, we omit the dependence on x
- Strong duality for TRPs (cf. Stern, Wolkowicz 1995) implies that

$$\begin{aligned}
v_{\text{TRP}} &= \max\left\{\tfrac{1}{2}s^T Hs + g^T s : \quad \tfrac{1}{2}\|s\|_2^2 \le \tfrac{1}{2}\Delta^2\right\} \\
&= \inf\left\{d(\lambda) \ : \ \lambda - \lambda_{\max}(H) > 0, \quad \lambda \ge 0\right\} && \text{(strong duality)} \\
&\approx \inf\left\{d(\lambda) \ : \ \lambda - E(H;\eta) > 0, \quad \lambda \ge 0\right\} && \text{(smooth max. eigenvalue fct.)} \\
&\approx \min_{\lambda} \ d(\lambda) + \frac{\nu}{\lambda} + \frac{\nu}{\lambda - E(H;\eta)} && \text{(reciprocal barrier)} \\
&= \max\left\{\tfrac{1}{2}\tilde{s}^T \tilde{H}_\eta \tilde{s} + \tilde{g}_\nu{}^T \tilde{s} \ : \ \tfrac{1}{2}\|\tilde{s}\|_2^2 \le \tfrac{1}{2}\Delta^2\right\} && \text{(strong duality)} \\
&= \tilde{v}_{\text{TRP}}(\nu,\eta),
\end{aligned}$$

where $d(\lambda) = \frac{1}{2}g^T(\lambda I - H)^{-1}g + \frac{1}{2}\Delta^2\lambda \quad \text{if} \quad \lambda - \lambda_{\max}(H) > 0$

Smoothing Function for Value Function of TRP III

- The lifted TR-problem (LTRP) has a unique optimal solution
- Sensitivity analysis and Danskin's theorem imply that $x \mapsto \tilde{v}_{\mathrm{TRP}}(x;\nu,\eta)$ is as smooth as $x \mapsto g(x)$ and $x \mapsto H(x)$ are

- Error estimate uniformly in $x \in X$: For all $(x,\eta,\nu) \in X \times (0,\infty)^2$:

$$v_{\mathrm{TRP}}(x) \le \tilde{v}_{\mathrm{TRP}}(\eta,\nu) \le v_{\mathrm{TRP}}(x) + 2\sqrt{2\nu}\Delta + \frac{1}{2}\Delta^2\eta \ln p.$$

- We can prove that $\tilde{v}_{\mathrm{TRP}}$ is a gradient consistent smoothing function for v_{TRP}

- Lifted trust-region problem (LTRP) can be solved efficiently
  → Typically, the Moré-Sorensen algorithm takes ≤ 10 iterations (Moré, Sorensen 1983)
- Derivative of $x \mapsto \tilde{v}_{\mathrm{TRP}}(x;\nu,\eta)$ is inexpensive to obtain (Danskin 1966)

Smoothing Function $\tilde{\Psi}_{\text{TRP}}$ for Ψ_{TRP}

Recall

$$\Psi_{\text{TRP}}(x) = \max_{\mu\in\mathbb{R}^p}\left\{ Q(x,\mu) : \quad \|\bar{\Sigma}^{-\frac{1}{2}}(\mu-\bar{\mu})\|_2 \le \Delta \right\}, \quad \text{where}$$

$$Q(x,\xi) = f(x,\bar{\mu}) + \nabla_\xi f(x,\bar{\mu})^T(\xi-\bar{\mu}) + \tfrac{1}{2}(\xi-\bar{\mu})^T\nabla_{\xi\xi}f(x,\bar{\mu})(\xi-\bar{\mu}),$$

Making the substitutions

$$g(x) = \bar{\Sigma}^{\frac{1}{2}}\nabla_\xi f(x,\bar{\mu}), \quad H(x) = \bar{\Sigma}^{\frac{1}{2}}\nabla_{\xi\xi}f(x,\bar{\mu})\bar{\Sigma}^{\frac{1}{2}} \quad (\text{and } s = \bar{\Sigma}^{-\frac{1}{2}}(\xi-\bar{\mu}))$$

in the definitions of v_{TRP} and $\tilde{v}_{\text{TRP}}(\cdot;\nu,\eta)$, we obtain

$$\Psi_{\text{TRP}}(x) = v_{\text{TRP}}(x)|_{\text{substitution}}$$

and we can construct a smoothing function for Ψ_{TRP} via

$$\tilde{\Psi}_{\text{TRP}}(x;\nu,\eta) = f(x,\bar{\mu}) + \tilde{v}_{\text{TRP}}(x;\nu,\eta)|_{\text{substitution}}$$

Smoothed Approximated DRO Problem

- We started with the DRO problem:

$$\min_{x \in X} \sup_{P \in \mathcal{P}} \mathbb{E}_P[f(x,\xi)]$$

- We approximated $f(x,\xi)$ by quadratic Taylor expansion w.r.t. ξ:

$$\min_{x \in X} \sup_{P \in \mathcal{P}} \mathbb{E}_P[Q(u,\xi)]$$

- This was shown to be equivalent to

$$\min_{x \in X} \Psi_{\text{SDP}}(x) + \Psi_{\text{TRP}}(x)$$

- We then approximated Ψ_{SDP} and Ψ_{TRP} by smoothing functions:
  Smoothed Approximated DRO Problem:

$$\min_{x \in X} \tilde{\Psi}_{\text{SDP}}(x;\tau) + \tilde{\Psi}_{\text{TRP}}(x;\nu,\eta)$$

- If $f(\cdot,\xi)$, $\nabla_\xi f(\cdot,\xi)$, $\nabla_{\xi\xi} f(\cdot,\xi)$ are C^q, then this problem has a C^q-objective
- We thus can apply standard optimization solvers

Extension to Distributionally Robust Constraints

Distributionally Robust NLP:

$$\min_{x\in\mathbb{R}^n} \sup_{P\in\mathcal{P}} \mathbb{E}_P[f(x,\xi)] \quad \text{s.t.} \quad \sup_{P\in\mathcal{P}} \mathbb{E}_P[g_i(x,\xi)] \le 0 \quad (i \in [m]) \qquad \text{(DROP)}$$

For each g_i, we proceed as we did for f:

- g_i is approximated by a function $Q^{g_i}(x,\xi)$ that is quadratic w.r.t. ξ (we use Taylor expansion about $\bar{\mu}$)
- Approximated DROP:

$$\min_{x\in\mathbb{R}^n} \sup_{P\in\mathcal{P}} \mathbb{E}_P[Q(x,\xi)] \quad \text{s.t.} \quad \sup_{P\in\mathcal{P}} \mathbb{E}_P[Q^{g_i}(x,\xi)] \le 0 \quad (i \in [m])$$

- There holds $\sup_{P\in\mathcal{P}} \mathbb{E}_P[Q^{g_i}(x,\xi)] = \Psi^{g_i}_{\text{SDP}}(x) + \Psi^{g_i}_{\text{TRP}}(x)$
- We construct smoothing functions $\tilde{\Psi}^{g_i}_{\text{SDP}}(x;\tau)$ and $\tilde{\Psi}^{g_i}_{\text{TRP}}(x;\nu,\eta)$
- Smoothed Approximated DROP:

$$\min_{x\in\mathbb{R}^n} \tilde{\Psi}_{\text{SDP}}(x;\tau) + \tilde{\Psi}_{\text{TRP}}(x;\nu,\eta) \quad \text{s.t.} \quad \tilde{\Psi}^{g_i}_{\text{SDP}}(x;\tau) + \tilde{\Psi}^{g_i}_{\text{TRP}}(x;\nu,\eta) \le 0 \quad (i \in [m])$$

- Standard NLP solvers are applicable to the smoothed approximated DROP

Convergence of Smoothing Method for Approximated DROP

$$F(x) := \Psi_{\mathrm{SDP}}(x) + \Psi_{\mathrm{TRP}}(x), \qquad \tilde{F}(x;\rho) := \tilde{\Psi}_{\mathrm{SDP}}(x;\tau) + \tilde{\Psi}_{\mathrm{TRP}}(x;\nu,\eta),$$
$$G_i(x) := \Psi^{g_i}_{\mathrm{SDP}}(x) + \Psi^{g_i}_{\mathrm{TRP}}(x), \qquad \tilde{G}_i(x;\rho) := \tilde{\Psi}^{g_i}_{\mathrm{SDP}}(x;\tau) + \tilde{\Psi}^{g_i}_{\mathrm{TRP}}(x;\nu,\eta).$$

Approximated DROP:

$$\min_{x\in\mathbb{R}^n} F(x) \quad \text{s.t.} \quad G_i(x) \le 0 \quad (i \in [m]) \qquad \text{(ADROP)}$$

KKT conditions for ADROP:

$$0 \in \partial F(\bar{x}) + \sum_i \bar{\lambda}_i \partial G_i(\bar{x}),$$
$$\bar{\lambda}_i \ge 0, \quad G_i(\bar{x}) \le 0, \quad \bar{\lambda}_i G_i(\bar{x}) = 0 \quad (i \in [m])$$

Smoothed approximated DROP ($\rho = \rho_k$):

$$\min_{x\in\mathbb{R}^n} \tilde{F}(x;\rho_k) \quad \text{s.t.} \quad \tilde{G}_i(x;\rho_k) \le 0 \quad (i \in [m]) \qquad \text{(SADROP)}$$

We compute ε_k-KKT-pairs (x^k, λ^k) of (SADROP), required to satisfy

$$\|\nabla_x \tilde{F}(x^k;\rho_k) + \sum_i \lambda_i^k \nabla_x \tilde{G}_i(x^k;\rho_k)\|_\infty \le \varepsilon_k,$$
$$|\min\{-G_i(x^k;\rho_k), \lambda_i^k\}| \le \varepsilon_k \quad (i \in [m])$$

Convergence Result

Feasibility:

If (ADROP) has a strictly feasible point, then there exists $\delta > 0$ such that, for all $\rho_k \in (0, \delta]^3$, (SADROP) is feasible.

Convergence to KKT-points:

Let $\mathbb{R}_{++} \ni \varepsilon_k \to 0$, $\mathbb{R}^3_{++} \ni \rho_k \to 0$ and consider a sequence (x^k, λ^k) of ε_k-KKT-pairs (x^k, λ^k) of (SADROP).

Then every accumulation point $(\bar{x}, \bar{\lambda})$ of (x^k, λ^k) is a KKT-pair of (ADROP).

It is also possible to prove results about convergence of global solutions.

Preliminary Remarks on Numerical Performance

In Milz, M.U., SIOPT 2020, we numerically compare several approaches for solving unconstrained approximated DROPs.

DROPs are generated from unconstrained testset (Moré, Garbow, Hillstrom) via

$$f(x,\xi) = \hat{f}(x+\xi) \qquad (\hat{f} = \text{cost function given in the testset})$$

We compare

- our continuation method, using IPOPT with BFGS-updates
- nonconvex bundle method MPBNGC (Mäkelä, Karmitsa, Wilppu 2016)
- PENLAB (Fiala, Kocvara, Stingl 2013) applied to a nonlinear SDP reformulation of the approximated DROP

Our method is competitive and needs lowest no. of higher order derivative evals

We can use any NLP solver and our approach can handle additional distributionally robust (and other) constraints quite easily

I postpone numerical experiments to the end, where I will present results for PDE-constrained DROPs

Distributionally Robust Optimization with PDEs

Distributionally Robust Reduced Optimal Control Problem:

$$\min_{u\in U_{\mathrm{ad}}} \sup_{P\in\mathcal{P}} \mathbb{E}_P[\hat{J}(u,\xi)] \qquad \text{(DROPDE)}$$

- $\mathcal{P}$ ambiguity set of probability distributions on Ξ (here: $\Xi = \mathbb{R}^p$)
- $U_{\mathrm{ad}} \subset U$ set of admissible controls, U a Hilbert space
- $\hat{J} : U \times \Xi \to \mathbb{R}$ reduced parametric cost function, defined by

$$\hat{J}(u,\xi) = J(S(u,\xi), u, \xi),$$

- $J : Y \times U \times \Xi \to \mathbb{R}$ parametric cost function,
- $S : U \times \Xi \to Y$ unique solution operator of a parametric PDE:

  Given $u \in U$, the parametric state $y(\xi) = S(u,\xi)$ solves

$$e(y(\xi), u, \xi) = 0, \quad \xi \in \Xi,$$

  where $e : Y \times U \times \Xi \to Z$ and Y, Z Banach spaces

PDE-Constrained DROP: Assumptions

Assumptions on control-to-state operator S:

- PDE has uniquely defined solution operator $S : U_{\text{ad}} \times \Xi \to Y$
- $S(\cdot, \xi)$ is weakly-weakly continuous on U_{ad} for all $\xi \in \Xi$
- $S(u, \cdot)$ is continuous for all $u \in U_{\text{ad}}$

Assumptions on cost function J:

- $J : V_Y \times U_{\text{ad}} \times \Xi \to \mathbb{R}$ is continuous with $V_Y \supset S(U_{\text{ad}}, \Xi)$
- $J(\cdot, \cdot, \xi)$ is weakly lower semicontinuous for all $\xi \in \Xi$
- There exists $\gamma \in \mathbb{R}$ with $\hat{J}(u, \xi) \geq \gamma$ for all $u \in U_{\text{ad}}$, $\xi \in \Xi$
- $\hat{J}(u, \cdot)$ is uniformly integrable w.r.t. $\mathcal{P}$ for all $u \in U_{\text{ad}}$:

$$\lim_{t\to\infty} \sup_{P \in \mathcal{P}} \mathbb{E}_P\left[|\hat{J}(u,\xi)|\, 1_{|\hat{J}(u,\cdot)| \geq t}(\xi)\right] = 0$$

- $\{u \in U_{\text{ad}} : \mathbb{E}_P[\hat{J}(u, \xi)] \leq \eta \ \forall\, P \in \mathcal{P}\}$ is nonempty and bounded for some $\eta \in \mathbb{R}$

(similar assumptions can be found in Kouri, Surowiec 2018)

PDE-Constrained DROP: Existence Results

Assumptions on control-to-state operator S:

- PDE has uniquely defined solution operator $S : U_{\text{ad}} \times \Xi \to Y$
- $S(\cdot, \xi)$ is weakly-weakly continuous for all $\xi \in \Xi$ on U_{ad}
- $S(u, \cdot)$ is continuous for all $u \in U_{\text{ad}}$

Assumptions on cost function J:

- $J : V_Y \times U_{\text{ad}} \times \Xi \to \mathbb{R}$ is continuous with $V_Y \supset S(U_{\text{ad}}, \Xi)$
- $J(\cdot, \cdot, \xi)$ is weakly lower semicontinuous for all $\xi \in \Xi$
- There exists $\gamma \in \mathbb{R}$ with $\hat{J}(u, \xi) \geq \gamma$ for all $u \in U_{\text{ad}}$, $\xi \in \Xi$
- $\hat{J}(u, \cdot)$ is uniformly integrable w.r.t. $\mathcal{P}$ for all $u \in U_{\text{ad}}$:

$$\lim_{t \to \infty} \sup_{P \in \mathcal{P}} \mathbb{E}_P[|\hat{J}(u, \xi)| \, 1_{|\hat{J}(u,\cdot)| \geq t}(\xi)] = 0$$

- $\{u \in U_{\text{ad}} : \mathbb{E}_P[\hat{J}(u, \xi)] \leq \eta \ \forall \, P \in \mathcal{P}\}$ is nonempty and bounded for some $\eta \in \mathbb{R}$

Theorem: Let the Assumptions hold. Then:

1. $\sup_P \mathbb{E}_P[\hat{J}(\cdot, \xi)]$ is finite-valued and weakly lower semicontinuous on U_{ad}
2. If U_{ad} is closed and convex, then the DROP has an optimal solution
3. For each $u \in U_{\text{ad}}$, there exists $P^* \in \mathcal{P}$ with $\mathbb{E}_{P^*}[\hat{J}(u, \xi)] = \sup_{P \in \mathcal{P}} \mathbb{E}_P[\hat{J}(u, \xi)]$

Approximated and Smoothed PDE-Constrained DROP: Existence and Convergence

Assumptions:

- PDE solution operator $S : U_{\mathrm{ad}} \times B_\epsilon(\bar{\mu}) \to Y$ is uniquely defined
- $\hat{J}(u, \cdot)$ is twice differentiable at $\bar{\mu}$ for all $u \in U_{\mathrm{ad}}$
- $\hat{J}(\cdot, \bar{\mu})$ is weakly lower semicontinuous on U_{ad}
- $\nabla_\xi \hat{J}(\cdot, \bar{\mu})$ and $\nabla_{\xi\xi} \hat{J}(\cdot, \bar{\mu})$ are weakly continuous on U_{ad}
- U_{ad} is nonempty, closed, convex; U_{ad} is bounded or Ψ is coercive on U_{ad}

Remark: Conditions on $\hat{J}$ can be translated to assumptions on J, S, and/or e

Theorem: Let the Assumptions hold. Then:

1. Ψ and $\tilde{\Psi}(\cdot; \rho_k)$ are weakly lower semicontinuous on U_{ad} for all $\rho_k \in \mathbb{R}^3_{++}$
2. The approximated DROP has an optimal solution
3. The smoothed approximated DROP has an optimal solution for all $\rho_k \in \mathbb{R}^3_{++}$
4. If $\mathbb{R}^3_{++} \ni \rho_k \to 0$ and (u_k) is a corresponding solution sequence of the smoothed approximated DROP, then $(u_k) \subset U_{\mathrm{ad}}$ is bounded and every weak accumulation point of (u_k) is a solution of the approximated DROP

Computation of Quadratic Approximation and Derivatives

- UFL and FEniCS are used to compute derivatives similar to dolfin-adjoint
- Formulas for derivatives are provided in Kolvenbach, Lass, Ulbrich 2018
- We combine sensitivity and adjoint approaches to compute derivatives
- For not too high parameter dim. p, we compute the full matrix $\nabla_{\xi\xi}\hat{J}(u,\bar{\mu})$
  - Allows to compute $\nabla_{\xi\xi}\hat{J}(u,\bar{\mu}) \bullet \bar{\Sigma}_0$ and eigendecomposition of $\nabla_{\xi\xi}\hat{J}(u,\bar{\mu})$
  - Note: Products $\nabla_{\xi\xi}\hat{J}(u,\bar{\mu})s_\xi$ suffice if iterative methods are used

Number of PDE solves

objective	$\Psi_{\text{SDP}}(u) + \Psi_{\text{TRP}}(u)$	1 state, p lin., $p+1$ adj.
smoothed obj.	$\tilde{\Psi}_{\text{SDP}}(u;\tau) + \tilde{\Psi}_{\text{TRP}}(u;\nu,\eta)$	1 state, p lin., $p+1$ adj.
derivative	$\nabla_u\tilde{\Psi}_{\text{SDP}}(u;\tau) + \nabla_u\tilde{\Psi}_{\text{TRP}}(u;\nu,\eta)$	p lin., p adj.

($\rho = (\tau,\nu,\eta)$ smoothing parameters)

DRO for Steady Burgers Equation

Consider

$$\min_{u\in U}\ \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\tfrac{1}{2}\|S(u,\xi)-y_d\|^2_{L^2(D)}\right]+\tfrac{\alpha}{2}\|u\|^2_{L^2(D)},$$

where $y(\xi)=S(u,\xi)\in Y=H^1(D)$ solves the weak form of the

Steady Burgers Equation:

$$\begin{aligned} -a(\xi)y_{xx}(x,\xi)+y(x,\xi)y_x(x,\xi) &= b(\xi)+u(x), \qquad x\in D,\quad \xi\in\Xi \\ y(0,\xi) &= d_0(\xi), \\ y(1,\xi) &= d_1(\xi) \end{aligned}$$

- $D=(0,1),\ U=L^2(D),\ \alpha=10^{-3},\ y_d=1\in L^2(D)$
- $a,b:\Xi=\mathbb{R}^4\to\mathbb{R},\quad a(\xi)=10^{\xi_1-2}>0,\quad b(\xi)=0.01\,\xi_2$
- $d_0,\ d_1:\Xi\to\mathbb{R},\ d_0(\xi)=1+10^{-3}\xi_3,\ d_1(\xi)=10^{-3}\xi_4$

- Problem based on Kouri, Heinkenschloss, Ridzal, van Bloemen Waanders 2013

DRO for Steady Burgers Equation (2)

- Existence of solutions to Burgers equation (ξ fixed) shown in Volkwein 1997
- If $a(\xi) > 0$ is sufficiently large, then the solution $y(\xi)$ is unique
- There exists a solution operator S that is measurable w.r.t. $\xi \in \Xi$
- $\|S(u,\cdot)\|_Y^r$ is uniformly integrable w.r.t. $\mathcal{P}$ for all $r \geq 1$
- Hence, many types of reduced cost functions $\hat{J}$ (in particular, tracking type functionals as in the test problem) are uniformly integrable w.r.t. $\mathcal{P}$
- The PDE operator $(y, u, \xi) \mapsto e(y, u, \xi)$ is smooth on $H^1(D) \times H^{-1}(D) \times \Xi$
- $e_y(y, u, \xi)$ is boundedly invertible, hence the implicit function theorem yields that, locally, S is smooth

$\rightarrow$ Combining all this, we can verify the assumptions of our theory (at least if S is unique, e.g., if a is sufficiently large)

DRO for Steady Burgers Equation (3)

Discretization

- Continuous piecewise linear finite elements on uniform grid with $N = 2000$ elements as in Kouri, Heinkenschloss, Ridzal, van Bloemen Waanders 2013

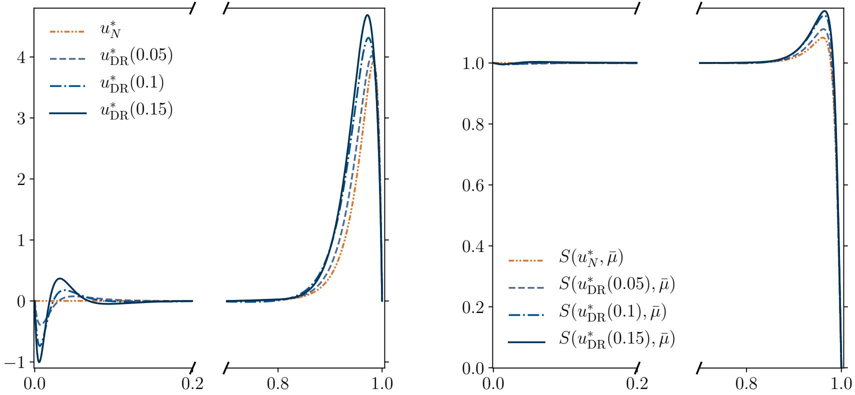
Application of continuation method

- $u^0 = 0$ and $(\tau_1, \nu_1, \eta_1) = 10^{-2}(1, 10^{-2}, 1)$.
- Update rule $(\tau_{k+1}, \nu_{k+1}, \eta_{k+1}) = 10^{-1}(\tau_k, 10^{-1}\nu_k, \eta_k)$.
- We use moola with LBFGS, Wolfe line search and termination tolerance $< 10^{-4}$ for each inner iteration (Schwedes, Ham, Funke, Piggott 2017)
- Termination if $\eta_k < 10^{-4}$ (three outer iterations)
- Trust-region problems are solved with Moré-Sorensen algorithm algorithm taking ≤ 5 iterations; cf. Moré, Sorensen 1983

Numerical Results for Steady Burgers Equation

- Compare stat. point $u^*_{\mathrm{DR}}(\Delta)$ of approx. DROP $\min_{u \in U} \sup_{P \in \mathcal{P}(\Delta)} \mathbb{E}_P[Q(u,\xi)]$

  with stat. point u^*_N of the nominal problem $\min_{u \in U} \hat{J}(u, \bar{\mu})$
- $\mathcal{P}(\Delta) = \{\, P : \|\mathbb{E}_P[\xi]\|_2 \le \Delta, \quad 0 \preccurlyeq \mathrm{Cov}_P[\xi] \preccurlyeq \Delta \mathrm{I} \,\}, \quad \bar{\mu} = 0$



Numerical Results for Steady Burgers Equation (2)

- $\mathcal{P}(0.1) = \{ P : \|\mathbb{E}_P[\xi]\|_2 \le 0.1, \quad 0 \preccurlyeq \mathrm{Cov}_P[\xi] \preccurlyeq 0.1\,\mathrm{I} \}$, $\bar{\mu} = 0$.

Table: Iteration history of continuation method ($\tilde{\Psi}(u;\rho) = \tilde{\Psi}_{\mathrm{SDP}}(u;\nu) + \tilde{\Psi}_{\mathrm{TRP}}(u;\tau,\eta)$).

k	$\tilde{\Psi}(u^k;\rho_k)$	$\|\nabla_u \tilde{\Psi}(u^k;\rho_k)\|_U$	#iter	$\frac{\|u^k-u^{k-1}\|_U}{1+\|u^{k-1}\|_U}$	$\#\tilde{\Psi}(u^k;\rho_k)$	$\#\nabla_u\tilde{\Psi}(u^k;\rho_k)$
1	7.97059e-03	6.13993e-05	18	8.24726e-01	21	21
2	4.71019e-03	9.30584e-05	9	7.27281e-02	11	11
3	4.54354e-03	8.85734e-05	3	3.23832e-03	5	5

Table: Empirical “performance” of u_N^* and $u_{\mathrm{DR}}^*(0.1)$.

control	$\mathrm{E}^m(u)$	$\mathrm{SD}^m(u)$	$Q^m_{0.80}(u)$	$Q^m_{0.95}(u)$	$Q^m_{0.99}(u)$
u_N^*	5.27694e-03	3.36866e-03	8.68155e-03	1.12073e-02	1.19278e-02
u_{DR}^*	5.01929e-03	2.70053e-03	7.68191e-03	9.81026e-03	1.04263e-02

$$\mathrm{E}^m(u) \approx \max_{1\le i\le m} \mathbb{E}_{P_i}[\hat{J}(u,\xi)], \qquad \text{(max. expectation – estimate of obj. val. of DROP)}$$

$$\mathrm{SD}^m(u) \approx \max_{1\le i\le m} \mathrm{SD}_{P_i}[\hat{J}(u,\xi)], \qquad \text{(max. standard deviation)}$$

$$Q^m_\beta(u) \approx \max_{1\le i\le m} \mathrm{VaR}_{P_i,\beta}(\hat{J}(u,\xi)), \text{(max. } \beta\text{-quantile)}$$

$P_i = N(\mu_i, \sigma_i^2 I) \in \mathcal{P}$, μ_i uniformly distr. on $\{\mu : \|\mu\|_2 \le 0.1\}$, σ_i^2 on $\{\sigma^2 : 0 \le \sigma^2 \le 0.1\}$, $m = 10$.

DRO for Unsteady Burgers Equation

Consider

$$\min_{u \in U} \sup_{P \in \mathcal{P}} \mathbb{E}_P\big[\tfrac{1}{2}\|S(u,\xi) - y_d\|^2_{L^2(D\times I)}\big] + \tfrac{\alpha}{2}\big(\|u_1\|^2_{L^2(I)} + \|u_2\|^2_{L^2(I)}\big),$$

where the state $y(\xi) = S(u,\xi) \in Y$ solves the weak form of the

Unsteady Burgers Equation:

$$\begin{aligned} y_t(x,t,\xi) - a(\xi)y_{xx}(x,t,\xi) + y(x,t,\xi)y_x(x,t,\xi) &= b(t,\xi), \quad (x,t) \in D \times I, \quad \xi \in \Xi,\\ y(x,0,\xi) &= y_0(x,\xi), \quad x \in D,\\ y_x(0,t,\xi) = u_1(t), \quad y_x(1,t,\xi) &= u_2(t), \quad t \in I \end{aligned}$$

- $D = (0,1), \ \ U = L^2(D) \times L^2(D), \ \ \alpha = 10^{-2}$
- $Y = W(I) = \{v \in L^2(I; H^1(D)) : v_t \in L^2(I; H^1(D)^*)\}$
- $a(\xi) = 10^{\xi_1 - 1}, \ \ b(t,\xi) = 0.01\,\xi_4 t$
- $y_0(x,\xi) = (1-10\xi_2)x^2(1+10\xi_3 - x)(1-x), \ \ \xi \in \Xi = \mathbb{R}^4$
- Problem based on Büskens, Griesse 2006

DRO for Unsteady Burgers Equation (2)

- Theory for the deterministic case can be found in Volkwein 2001
- For all $u \in U$, the state equation has a unique solution $y(\xi) = S(u, \xi)$
- The (weak) state equation operator is smooth w.r.t. $(y, \mu, \xi) \in Y \times U \times \mathbb{R}^4$
- $e_y(y, u, \xi)$ is boundedly invertible, hence the implicit function theorem yields that the solution operator S is smooth
- The weak-weak continuity of $S(\cdot, \xi)$ can be inferred from the weak closedness of $\{(y, u) : e(y, u, \xi) = 0\}$ and the estimate $\|S(u, \xi)\|_Y \leq C(\xi)(1 + \|u\|_U)$
- $J(\cdot, \cdot, \xi)$ is weakly lower semicontinuous and J is bounded below
- Analyzing $C(\xi)$ shows that $\|S(u, \xi)\|_Y^2$ is uniformly integrable w.r.t. $\mathcal{P}$
- Hence, $\hat{J}$ is uniformly integrable w.r.t. $\mathcal{P}$
- We also can show that $\hat{J}(\cdot, \xi)$ is weakly lower semicontinuous and $\nabla_\xi \hat{J}(\cdot, \xi)$, $\nabla_{\xi\xi} \hat{J}(\cdot, \xi)$ are weakly continuous

→ Combining all this, we can verify the assumptions of our theory

DRO for Unsteady Burgers Equation (3)

Discretization

- Implicit Euler scheme in time on a uniform mesh of $(0,1)$ with 100 steps.
- Piecewise linear finite elements in space on a uniform mesh with 100 elements.

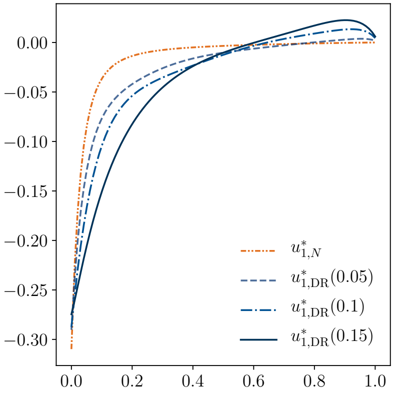
Application of continuation method

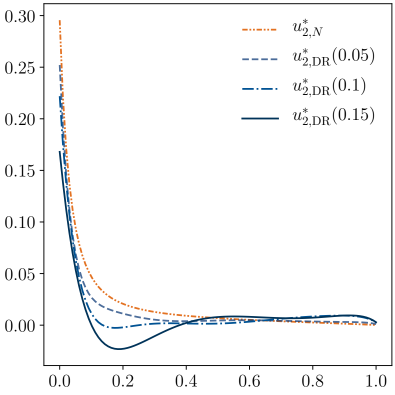
- We choose $u^0 = u_{\mathrm{N}}^*$ (stationary control of nominal problem), $(\tau_1, \nu_1, \eta_1) = 10^{-2}(1, 10^{-2}, 1)$.
- Update rule $(\tau_{k+1}, \nu_{k+1}, \eta_{k+1}) = 10^{-1}(\tau_k, 10^{-1}\nu_k, \eta_k)$
- We use scipy with LBFGS; termination tolerance $< 10^{-6}$ for inner iterations
- Termination if $\eta_k < 10^{-4}$ (three outer iterations)

Numerical Results for Unsteady Burgers Equation

- Compare stat. point $u^*_{\mathrm{DR}}(\Delta)$ of approx. DROP $\min_{u\in U} \sup_{P\in\mathcal{P}(\Delta)} \mathbb{E}_P[Q(u,\xi)]$

  with stat. point u^*_N of the nominal problem $\min_{u\in U} \hat{J}(u,\bar{\mu})$
- $\mathcal{P}(\Delta) = \{ P : \|\mathbb{E}_P[\xi]\|_2 \le \Delta, \quad 0 \preccurlyeq \mathrm{Cov}_P[\xi] \preccurlyeq \Delta \mathrm{I} \}, \quad \bar{\mu} = 0$





Numerical Results for Unsteady Burgers Equation

- $\mathcal{P}(0.01) = \{ P : \|\mathbb{E}_P[\xi]\|_2 \le 0.01, \quad 0 \preccurlyeq \mathrm{Cov}_P[\xi] \preccurlyeq 0.01\,\mathrm{I} \}$, $\bar{\mu} = 0$.

Table: Iteration history of continuation method ($\tilde{\Psi}(u;\rho) = \tilde{\Psi}_{\mathrm{SDP}}(u;\tau) + \tilde{\Psi}_{\mathrm{TRP}}(u;\nu,\eta)$).

k	$\tilde{\Psi}(u^k;\rho_k)$	$\|\nabla_u \tilde{\Psi}(u^k;\rho_k)\|_U$	#iter	$\frac{\|u^k - u^{k-1}\|_U}{1+\|u^{k-1}\|_U}$	$\#\tilde{\Psi}(u^k;\rho_k)$	$\#\nabla_u \tilde{\Psi}(u^k;\rho_k)$
1	9.71222e-03	7.95245e-04	22	2.71162e-03	26	26
2	8.30158e-03	7.45890e-03	16	3.05599e-04	20	20
3	8.17309e-03	3.15171e-03	3	3.16757e-05	7	7

Table: Empirical "performance" of u_N^* and $u_{\mathrm{DR}}^*(0.01)$.

control	$\mathrm{E}^m(u)$	$\mathrm{SD}^m(u)$	$Q^m_{0.80}(u)$	$Q^m_{0.95}(u)$	$Q^m_{0.99}(u)$
u_N^*	7.06471e-03	1.25907e-02	9.09762e-03	3.38810e-02	5.98617e-02
u_{DR}^*	6.56620e-03	1.14197e-02	8.45055e-03	3.06941e-02	5.42306e-02

Conclusions

- Developed an approximation scheme for DRO using quadratic (Taylor's) expansion w.r.t. ξ of parametric objective function and constraints
- Designed continuation algorithm based on smoothing methods
- Proved convergence result for continuation method
- Considered PDE-constrained DRO
- Proved existence of optimal solutions of the DROP, the approximated DROP and the smoothed DROP
- Showed numerical results

Literature I

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):1166–1192, 2017.

A. Alla, M. Hinze, P. Kolvenbach, O. Lass, and S. Ulbrich. A certified model reduction approach for robust parameter optimization with PDE constraints. *Adv. Comput. Math.*, 45 (3):1221–1250, 2019.

A. Ben-Tal and D. den Hertog. Hidden conic quadratic representation of some nonconvex quadratic optimization problems. *Math. Program.*, 143(1-2):1–29, 2014.

A. Ben-Tal and A. Nemirovski. Robust Convex Optimization. *Math. Oper. Res.*, 23(4): 769–805, 1998.

A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009.

A. Borzìand G. von Winckel. Multigrid methods and sparse-grid collocation techniques for parabolic optimal control problems with random coefficients. *SIAM J. Sci. Comput.*, 31: 2172–2192, 2009.

C. Büskens and R. Griesse. Parametric sensitivity analysis of perturbed PDE optimal control problems with state and control constraints. *J. Optim. Theory Appl.*, 131(1):17–35, 2006.

P. Chen and A. Quarteroni. Weighted reduced basis method for stochastic optimal control problems with elliptic pde constraint. *SIAM/ASA J. Uncertainty Quantification*, 2(1): 364–396, 2014.

Literature II

P. Chen, U. Villa, and O. Ghattas. Taylor approximation for PDE-constrained optimization under uncertainty: Application to turbulent jet flow. *Proc. Appl. Math. Mech.*, 18(1): e201800466, 2018.

X. Chen. Smoothing methods for nonsmooth, nonconvex minimization. *Math. Program.*, 134 (1):71–99, 2012.

X. Chen, Z. Nashed, and L. Qi. Smoothing methods and semismooth methods for nondifferentiable operator equations. *SIAM J. Numer. Anal.*, 38(4):1200–1216, 2000.

Z. Chen, M. Sim, and H. Xu. Distributionally robust optimization with infinitely constrained ambiguity sets. *Oper. Res.*, 2019.

S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz. Risk averse shape optimization. *SIAM J. Control Optim.*, 49(3):927–947, 2011.

J. M. Danskin. The theory of max-min, with applications. *SIAM J. Appl. Math.*, 14(4): 641–664, 1966.

F. de Gournay, G. Allaire, and F. Jouve. Shape and topology optimization of the robust compliance via the level set method. *ESAIM Control. Optim. Calc. Var.*, 14(1):43–70, 2008.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.*, 58(3):595–612, 2010.

Literature III

M. Diehl, H. G. Bock, and E. Kostina. An approximation technique for robust nonlinear optimization. *Math. Program.*, 107(1):213–230, 2006.

P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.*, 171(1):115–166, 2018.

R. Gao and A. J. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *Preprint arXiv:1604.02199*, 2016.

S. Garreis, T. M. Surowiec, and M. Ulbrich. An interior-point approach for solving risk-averse PDE-constrained optimization problems with coherent risk measures. *SIAM J. Optim.*, 31: 1–29, 2021.

J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Oper. Res.*, 58:902–917, 2010.

E. T. Hale and Y. Zhang. Case studies for a first-order robust nonlinear programming formulation. *J. Optim. Theory Appl.*, 134(1):27–45, 2007.

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(3):697–731, 2018.

D. P. Kouri. A measure approximation for distributionally robust PDE-constrained optimization problems. *SIAM J. Numer. Anal.*, 55(6):3147–3172, 2017.

Literature IV

D. P. Kouri and A. Shapiro. Optimization of PDEs with uncertain inputs. In H. Antil, D. P. Kouri, M.-D. Lacasse, and D. Ridzal, editors, *Frontiers in PDE-Constrained Optimization*, volume 163 of *The IMA Volumes in Mathematics and its Applications*, pages 41–81. Springer, New York, NY, 2018.

D. P. Kouri and T. M. Surowiec. Existence and Optimality Conditions for Risk-Averse PDE-Constrained Optimization. *SIAM/ASA J. Uncertainty Quantification*, 6(2):787–815, 2018.

D. P. Kouri and T. M. Surowiec. Epi-regularization of risk measures. *Math. Oper. Res.*, 45: 774–795, 2020.

D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. van Bloemen Waanders. A trust-region algorithm with adaptive stochastic collocation for pde optimization under uncertainty. *SIAM J. Sci. Comput.*, 35(4):A1847–A1879, 2013.

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(5):S112–S139, 2017.

A. S. Lewis. Derivatives of spectral functions. *Math. Oper. Res.*, 21(3):576–588, 1996.

J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM J. Sci. and Stat. Comput.*, 4(3):553–572, 1983.

Literature V

G. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quant. Finance*, 7(4):435–442, 2007.

I. Popescu. Robust mean-covariance solutions for stochastic optimization. *Oper. Res.*, 55(1): 98–112, 2007.

R. T. Rockafellar and J. O. Royset. On buffered failure probability in design and optimization of structures. *Reliab. Eng. Syst. Saf.*, 95:499–510, 2010.

H. E. Scarf. A min-max solution of an inventory problem. Technical report, RAND Corporation, Santa Monica, California, 1957.

T. Schwedes, D. A. Ham, S. W. Funke, and M. D. Piggott. *Mesh Dependence in PDE-Constrained Optimisation*. SpringerBriefs in Mathematics of Planet Earth. Springer, Cham, 2017.

A. Shapiro. Distributionally robust stochastic programming. *SIAM J. Optim.*, 27(4): 2258–2275, 2017.

A. Shapiro and A. Kleywegt. Minimax analysis of stochastic problems. *Optim. Methods Softw.*, 17(3):523–542, 2002.

A. Sichau and S. Ulbrich. A Second Order Approximation Technique for Robust Shape Optimization. In *Uncertainty in Mechanical Engineering*, volume 104 of *Applied Mechanics and Materials*, pages 13–22. Trans Tech Publications, 1 2012.

Literature VI

A. M.-C. So. Moment inequalities for sums of random matrices and their applications in optimization. *Math. Program.*, 130(1):125–151, 11 2011.

R. J. Stern and H. Wolkowicz. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. *SIAM J. Optim.*, 5(2):286–313, 1995.

A. Van Barel and S. Vandewalle. Robust optimization of PDEs with random coefficients using a multilevel Monte Carlo method. *SIAM/ASA J. Uncertainty Quantification*, 7(1):174–202, 2019.

S. Volkwein. *Mesh-Independence of an Augmented Lagrangian-SQP Method in Hilbert Spaces and Control Problems for the Burgers Equation*. Dissertation, Department of Mathematics, Technical University of Berlin, Berlin, 1997. URL `https://imsc.uni-graz.at/volkwein/diss.ps`.

S. Volkwein. Second-order conditions for boundary control problems of the Burgers equation. *Control Cybernet.*, 30:249–278, 2001.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Oper. Res.*, 62(6):1358–1376, 2014.

Y. Xu, W. Sun, and L. Qi. A feasible direction method for the semidefinite program with box constraints. *Appl. Math. Lett.*, 24(11):1874–1881, 2011.