Tutorial II
PDE-Constrained Optimization Under Uncertainty

Matthias Heinkenschloss

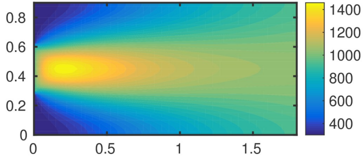
Department of Computational Applied Mathematics and Operations Research
Rice University, Houston, Texas
heinken@rice.edu

April 17, 2025

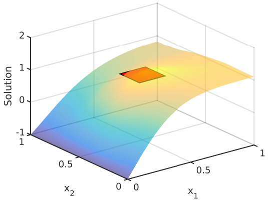
East Coast Optimization Meeting 2025
'Optimization and Digital Twins'
April 17 - 18, 2025
George Mason University, Arlington, Virginia (USA)

Combustion (nonlinear convection-diffusion-reaction)

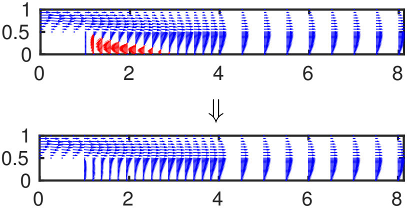
(Buffoni,Willcox,2010)



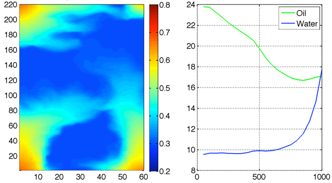
Linear advection-diffusion



Flow control (Navier-Stokes)



Reservoir management



Optimization Algorithms

- Consider problems with many control/design variables
  $\rightarrow$ gradient-based methods
- Deterministic PDE-constrained optimization:
  - Objective function evaluation: Solve (nonlinear) PDE.
  - Gradient evaluation: Solve linear adjoint PDE.
  - Hessian-times-vector evaluation: Solve 2 linear PDEs.
- PDE-constrained optimization under uncertainty:
  - Incorporate all realizations of uncertainties.
  - Risk measure.
  - Computational cost multiplies with number of samples.
- Many optimization approaches (stochastic gradient to SAA).
- Many variations for each optimization approach (discretization,...)
- Theory of many approaches limited to risk-neutral formulations, and convex problems (linear PDEs).
- Will focus on risk-averse optimization (AVaR), SAA, and Newton-type methods to solve SAA subproblems.

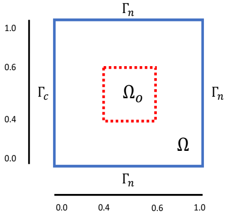
Model Problem

- $\boldsymbol{\xi}$ RV with values $\xi \in \varXi \subset \mathbb{R}^2$
- For given value of ξ and control u, y solves

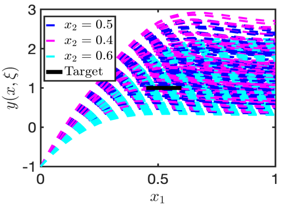
$$
\begin{aligned}
-\nabla \cdot \big(\kappa(x,\xi)\nabla y(x;u,\xi)\big) + [1\ 0]^T \cdot \nabla y(x;u,\xi) &= f(x,\xi), && x \in \varOmega \\
y(x;u,\xi) &= u(x,\xi), && x \in \varGamma_c \\
\big(\kappa(x,\xi)\nabla y(x;u,\xi)\big) \cdot n(x) &= 0, && x \in \varGamma_n.
\end{aligned}
$$

- Spatially discretized variables, $\mathbf{u}, \mathbf{y}$, consider QoI

$$
F(\mathbf{u},\xi) = \left\| \big(\mathbf{C}\mathbf{y}(\mathbf{u},\xi) - \mathbf{1}\big)_+ \right\|_2
$$



Problem domain



No control

Outline

Risk-Averse Optimization

Smoothing Methods

Reducing Cost of Gradient and Hessian Computations

Reducing Cost of Function Evaluations

Risk-Averse Optimization Framework

- $\boldsymbol{\xi}$ is continuous RV with values $\xi \in \varXi \subset \mathbb{R}^M$, probability density function ρ
- State $y(u, \boldsymbol{\xi})$ is random field, and QoI $F(u, \boldsymbol{\xi})$ is RV
- Scalarize the QoI, introduce risk measure $\mathcal{R} : L^p_\rho(\varXi) \to \mathbb{R}$

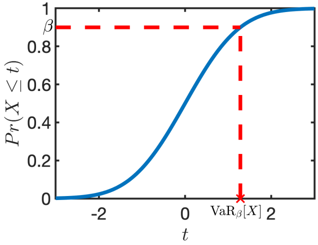
$$\min_u \; \mathcal{R}\big[F(u, \cdot)\big] + \mathcal{P}(u),$$

  for our applications, $F(u, \xi)$ will depend on solution to PDE
- In risk neutral optimization, $\mathcal{R} = \mathbb{E}$
- Consider $\mathcal{R}$ = Average Value-at-Risk (AVaR)
  (also known as Conditional Value-at-Risk (CVaR)).

Value-at-Risk (VaR)

- Let $X : \Xi \to \mathbb{R}$ be RV describing loss (previously was $F(u, \xi)$)
- For probability level $\beta \in (0, 1)$, define

$$\text{VaR}_\beta[X] := \inf_{t \in \mathbb{R}} \{t : Pr(X > t) \leq 1 - \beta\}$$



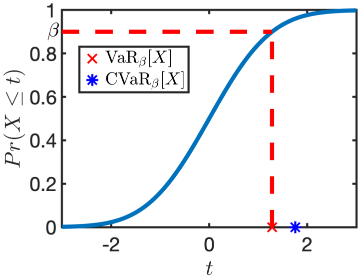
$\text{VaR}_\beta[X]$ for $X \sim \mathcal{N}(0, 1)$

- Problems with VaR:
  - Does not capture magnitude of loss
  - Sensitive to small changes in probability β
  - $\text{VaR}_\beta[F(u, \cdot)]$ not guaranteed convex in u even if F is

Average Value-at-Risk (AVaR)

- Given probability level $\beta \in (0,1)$, for continuous cdf AVaR is:

$$\text{AVaR}_\beta[X] = \mathbb{E}\big[X : X \geq \underset{\beta}{\text{VaR}}[X]\big]$$



$\text{AVaR}_\beta[X]$ for $X \sim \mathcal{N}(0,1)$

- Equivalent to minimization [Rockafellar and Uryasev, 2000]

$$\text{AVaR}_\beta[X] = \min_{t \in \mathbb{R}} \; t + (1-\beta)^{-1}\mathbb{E}\big[(X-t)_+\big]$$

- Captures failure magnitude, $\text{AVaR}_\beta[F(u,\cdot)]$ convex in u if F is

Illustration of Risk Measures

Test three optimal control problems

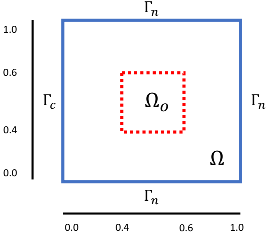
$$\min_{u} \ \mathbb{E}\big[F(u,\cdot)\big] + \frac{10^{-1}}{2}\|u\|^2_{H^1(\Gamma_c)}$$

$$\min_{u} \ \mathsf{AVaR}_\beta[F(u,\cdot)] + \frac{10^{-1}}{2}\|u\|^2_{H^1(\Gamma_c)}$$

$$\min_{u} \ \frac{1}{2}\|u\|^2_{H^1(\Gamma_c)} \quad \text{s.t.} \ \ \mathsf{AVaR}_\beta[F(u,\cdot)] \le \tau$$

for Advection Diffusion example

$$F(\mathbf{u},\xi) = \Big\|\big(\mathbf{C}\mathbf{y}(\mathbf{u},\xi) - \mathbf{1}\big)_+\Big\|_2$$



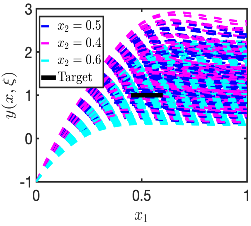
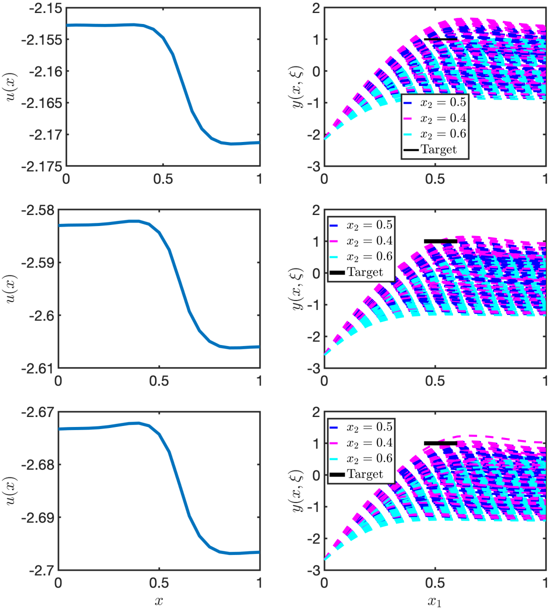


Illustration of Risk Measures, $N = 100$

$$\min_u \; \mathbb{E}[F(u,\cdot)] + \|u\|^2_{H^1}$$

$$\min_u \; \mathsf{AVaR}_{0.9}[F(u,\cdot)] + \|u\|^2_{H^1}$$

$$\min_u \; \|u\|^2_{H^1}, \quad \text{s.t.} \quad \mathsf{AVaR}_{0.9}[F(u,\cdot)] \le \frac{1}{N}$$



Challenges of Risk-Averse Optimization

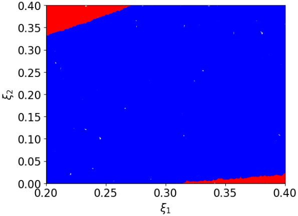
Two major challenges associated with solving:

$$\min_u \text{AVaR}_\beta\left[F(u,\cdot)\right] + \mathcal{P}(u)$$
$$= \min_{u,t}\ t + (1-\beta)^{-1}\mathbb{E}\left[\left(F(u,\cdot) - t\right)_+\right] + \mathcal{P}(u)$$

1. Non-smoothness from $(\cdot)_+$
2. Solution requires sampling, many samples vanish due to $(\cdot)_+$

# Samples	$\mathrm{AVaR}_\beta[X]$	%nonzero
1,000	2.52e-02	5.10
10,000	2.54e-02	5.01
20,000	2.54e-02	5.00

AVaR estimates for QoI of model problem, $\beta = 0.95$.



$\mathbb{G}_\beta = \{\xi : X(\xi) > \text{VaR}_\beta[X]\}$
risk-region

Risk-Region Depends on Control

Example: Boundary Control of Navier-Stokes

Find boundary control z on $\Gamma_c = \{0\} \times (0, 0.5)$ to minimize

$$F(u, \xi) = \int_{\Omega_o} \left(\partial_{x_1} y_2(\xi, u)(x) - \partial_{x_2} y_1(\xi, u)(x)\right)^2 dx$$

(measure of recirculation in $\Omega_o = [1, 4] \times [0, 0.5]$), where velocity $\vec{y} = [y_1, y_2]^T$ and pressure p solve Navier-Stokes equation

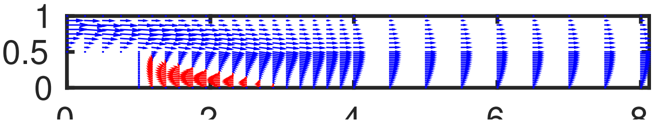
$$\begin{aligned} -\nu(\xi)\Delta\vec{y}(\xi) + (\vec{y}(\xi) \cdot \nabla)\vec{y}(\xi) + \nabla p(\xi) &= 0 \\ \nabla \cdot \vec{y}(\xi) &= 0 \end{aligned}$$

$\vec{y}(\xi) = \vec{0}$ on Γ_d, $\qquad \nu(\xi)\dfrac{\partial \vec{y}(\xi)}{\partial n} - \vec{n}p(\xi) = \vec{0}$ on Γ_{out},

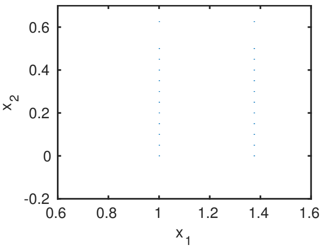
$\vec{y}(\xi) = [u, 0]^T$ on Γ_c, $\qquad \vec{y}(\xi) = \vec{y}_{in}(\xi)$ on Γ_{in}.

Uncertainty in viscosity and inflow velocities.

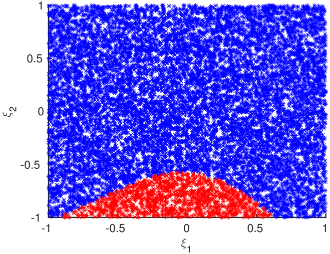
Uncontrolled flow at sample



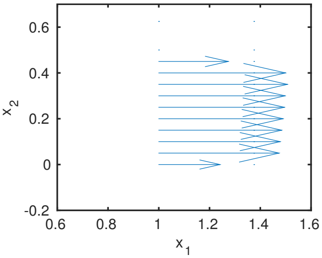
Sample Average Approx., $N = 10^3$ samples, $\mathrm{AVaR}_{\beta=0.9}$



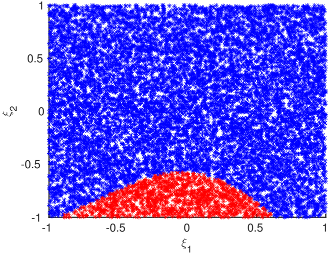
Initial control $u \equiv 0$



Risk region $\widehat{\mathbb{G}}_\beta(u)$ initial control

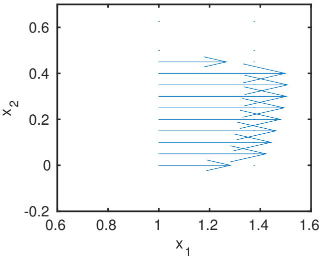


Control u iteration 10

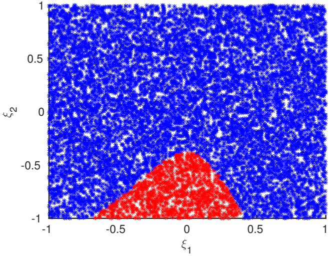


Risk region $\widehat{\mathbb{G}}_\beta(u)$ iteration 10

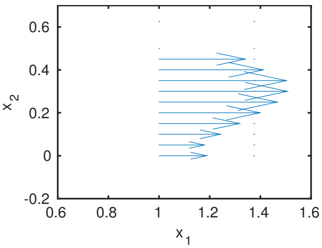
Sample Average Approx., $N = 10^3$ samples, $\text{AVaR}_{\beta=0.9}$



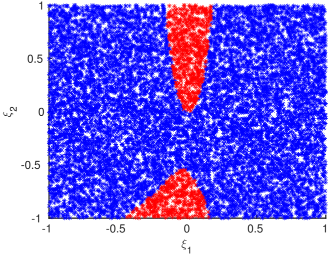
Control u iteration 20



Risk region $\widehat{\mathbb{G}_\beta}(u)$ iteration 20



Control u iteration 30



Risk region $\widehat{\mathbb{G}_\beta}(u)$ iteration 30

Challenges Extends to Other Risk-Averse Optimization

Similar challenges apply to other risk measures, e.g.

- Optimized certainty equivalents, [Ben-Tal and Teboulle, 2007], [Garreis et al., 2021]

$$\mathcal{R}[X] := \inf_{t \in \mathbb{R}} \left\{ t + \mathbb{E}\big[v(X - t)\big] \right\},$$

  where $v(0) = 0$ and $v(x) > x$ is a "scalar regret function"

- Buffered Probability of Failure [Rockafellar and Royset, 2010], [Mafusalov et al., 2018]

$$\begin{aligned} \text{bPoF}_\tau(X) &= \underbrace{Pr(X > \tau)}_{\text{Prob. of failure}} + \underbrace{Pr(X \in [\text{VaR}_\beta[X], \tau])}_{\text{Prob. of near failure}}, \\ &= \min_{a \geq 0} \mathbb{E}\Big[\big(a(X - \tau) + 1\big)_+\Big], \end{aligned}$$

  for a given failure threshold τ

Outline

Addressing Non-smoothness

- AVaR

$$\begin{aligned} &\min_u \ \mathsf{AVaR}_\beta\big[F(u,\cdot)\big] + \mathcal{P}(u) \\ &= \min_{u,t} \ t + (1-\beta)^{-1}\mathbb{E}\Big[\big(F(u,\cdot)-t\big)_+\Big] + \mathcal{P}(u) \end{aligned}$$

- Finite-dimensional setting, subproblem of SAA

$$\min_{u,t} \ t + \frac{1}{N}\sum_{i=1}^{N}(1-\beta)^{-1}\big(F(u,\xi_i)-t\big)_+ + \mathcal{P}(u)$$

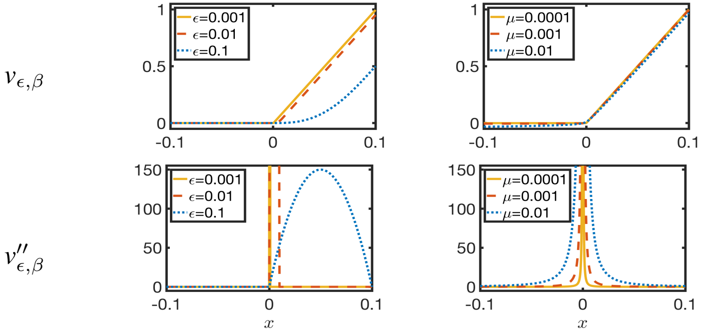
- Direct smoothing methods to approximate $(1-\beta)^{-1}(\cdot)_+$ [Kouri and Surowiec, 2016], [Basova et al., 2011], [Markowski, 2022]
- Reformulation as inequality constrained program
  - Log-barrier method, interior point method [Garreis et al., 2021]
  - Epi-regularization [Kouri and Surowiec, 2020]
  - Method of multipliers [Kouri and Surowiec, 2022]
- These optimization approaches ultimately require solution of a sequence of smoothed subproblems.
  Approximate minimization via (semi-smooth) Newton's method.

Smoothed Subproblems

- Can show approaches (direct smoothing, log-barrier, augmented Lagrangian) lead to problem of the form

$$\min_{u,t} \; J_\beta^\epsilon(u,t) := t + \frac{1}{N}\sum_{i=1}^{N} v_{\epsilon,\beta}\big(F(u,\xi_i) - t\big) + \mathcal{P}(u),$$

where $v_{\epsilon,\beta} : \mathbb{R} \to \mathbb{R}$ differs for each method



Conv. smoothing — Log barrier

Newton-CG for Smoothed CVaR

- Consider Newton-CG method for approximate solution of

$$\min_{u,t} J_\beta^\epsilon(u,t) = t + \frac{1}{N}\sum_{i=1}^{N} v_{\epsilon,\beta}\big(F(u,\xi_i) - t\big) + \alpha\mathcal{P}(u).$$

- Requires:
  - first- and second-order derivative information of J_β^ϵ for quadratic model at kth iterate (u_k, t_k),
  - approximate minimization of quadratic model via conjugate gradient (CG) method to generate search direction, and
  - line search or trust-region for globalization.

Quadratic Model

- At the kth optimization iteration, generate step s^k by approximately minimizing quadratic model $m_k(s_u, s_t)$:

$$
\begin{aligned}
m_k(s_u, s_t) = \frac{1}{2} \begin{bmatrix} s_u \\ s_t \end{bmatrix}^T \begin{bmatrix} \nabla_{uu} J_\beta^\epsilon(u_k, t_k) & \nabla_{ut} J_\beta^\epsilon(u_k, t_k) \\ \nabla_{tu} J_\beta^\epsilon(u_k, t_k) & \nabla_{tt} J_\beta^\epsilon(u_k, t_k) \end{bmatrix} \begin{bmatrix} s_u \\ s_t \end{bmatrix} \\
+ \begin{bmatrix} \nabla_u J_\beta^\epsilon(u_k, t_k) \\ \nabla_t J_\beta^\epsilon(u_k, t_k) \end{bmatrix}^T \begin{bmatrix} s_u \\ s_t \end{bmatrix}
\end{aligned}
$$

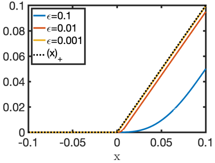
Derivatives in Quadratic Model

$$
\begin{aligned}
\nabla_u J_\beta^\epsilon(u,t) &= \dots \\
\nabla_t J_\beta^\epsilon(u,t) &= 1 - \frac{1}{N}\sum_{i=1}^N v'_{\epsilon,\beta}\big(F(u,\xi_i) - t\big) \\
\nabla_{uu} J_\beta^\epsilon(u,t) &= \dots \\
\nabla_{ut} J_\beta^\epsilon(u,t) &= -\frac{1}{N}\sum_{i=1}^N v''_{\epsilon,\beta}\big(F(u,\xi_i) - t\big)\nabla_u F(u,\xi_i) \\
&= \nabla_{tu} J_\beta^\epsilon(u,t)^T \\
\nabla_{tt} J_\beta^\epsilon(u,t) &= \frac{1}{N}\sum_{i=1}^N v''_{\epsilon,\beta}\big(F(u,\xi_i) - t\big)
\end{aligned}
$$

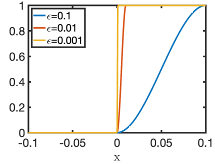
- Hessian is rank-deficient if $v''_{\epsilon,\beta}\big(F(u,\xi_i) - t\big) = 0$ for all ξ_i

Recall structure of smoothing function (specific example shown)

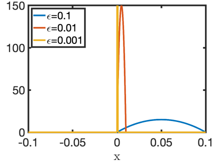
$$v_{\epsilon,1}(x) = \begin{cases} 0, & x \le 0 \\ \frac{x^3}{\epsilon^2} - \frac{x^4}{2\epsilon^3}, & x \in (0,\epsilon) \\ x - \frac{\epsilon}{2}, & x \ge \epsilon \end{cases}$$



$$v'_{\epsilon,1}(x) = \begin{cases} 0, & x \le 0 \\ \frac{3x^2}{\epsilon^2} - \frac{2x^3}{\epsilon^3}, & x \in (0,\epsilon) \\ 1, & x \ge \epsilon \end{cases}$$



$$v''_{\epsilon,1}(x) = \begin{cases} 0, & x \le 0 \\ \frac{6x}{\epsilon^2} - \frac{6x^2}{\epsilon^3}, & x \in (0,\epsilon) \\ 0, & x \ge \epsilon \end{cases}$$



Rank-Deficient Quadratic Model

- If $F(u, \xi_i) - t \notin (0, \epsilon)$ for all $i = 1, \dots, N$,

$$
\begin{aligned}
v'_{\epsilon,\beta}\big(F(u,\xi_i) - t\big) &= 0, \quad \text{for } \xi_i \ \text{ s.t. } F(u,\xi_i) - t \le 0,\\
v'_{\epsilon,\beta}\big(F(u,\xi_i) - t\big) &= 1, \quad \text{for } \xi_i \ \text{ s.t. } F(u,\xi_i) - t \ge \epsilon,\\
v''_{\epsilon,\beta}\big(F(u,\xi_i) - t\big) &= 0, \quad \text{for all } \xi_1, \dots, \xi_N.
\end{aligned}
$$

- Quadratic model becomes

$$
\begin{aligned}
m(s_u, s_t) = &\frac{1}{2} \begin{bmatrix} s_u \\ s_t \end{bmatrix}^T \begin{bmatrix} \nabla_{uu} J_\beta^\epsilon(u,t) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} s_u \\ s_t \end{bmatrix} \\
&+ \begin{bmatrix} \nabla_u J_\beta^\epsilon(u,t) \\ 1 - \frac{1}{N}\big|\{i : F(u,\xi_i) - t \ge \epsilon\}\big| \end{bmatrix}^T \begin{bmatrix} s_u \\ s_t \end{bmatrix},
\end{aligned}
$$

- Model is unbounded unless $\big|\{i : F(u,\xi_i) - t \ge \epsilon\}\big| = N$
  - Line Search Newton: long, poor-quality line searches
  - Trust Region Newton: rejection of iterate and shrinking of trust radius

Modification of Line Search Newton-CG

- Introduce a computationally inexpensive sub-step
- Minimize J_β^ϵ w.r.t. t first to obtain intermediate point $(u_k, t_{k+1/2})$
- Since $\nabla_t J_\beta^\epsilon(u_k, t_{k+1/2}) = 0$, rank deficient Hessian no longer causes unboundedness of quadratic model

$$
\begin{aligned}
m_{k+1/2}(s_u, s_t) = \frac{1}{2} \begin{bmatrix} s_u \\ s_t \end{bmatrix}^T \begin{bmatrix} \nabla_{uu} J_\beta^\epsilon(u_k, t_{k+1/2}) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} s_u \\ s_t \end{bmatrix} \\
+ \begin{bmatrix} \nabla_u J_\beta^\epsilon(u_k, t_{k+1/2}) \\ 0 \end{bmatrix}^T \begin{bmatrix} s_u \\ s_t \end{bmatrix}
\end{aligned}
$$

Modified Line Search Newton-CG Algorithm

Input: Initial iterate (u_0, t_0).
Output: Approximate min $(u_\epsilon^*, t_\epsilon^*)$ of smoothed CVaR objective $J_\beta^\epsilon(u,t)$.

for $k = 0, 1, \ldots$ **do**
    Solve $\min\limits_t \; J_\beta^\epsilon(u_k, t)$ for $t_{k+1/2}$.
    Build quadratic model $m_{k+1/2}$ centered around $(u_k, t_{k+1/2})$.
    Approximately minimize $m_{k+1/2}$ using truncated-CG with tolerance
$$\mathrm{tol}_k^{\mathrm{CG}} = \min\left\{ \|\nabla J_\beta^\epsilon(u_k, t_{k+1/2})\|^2, 0.01 \|\nabla J_\beta^\epsilon(u_k, t_{k+1/2})\| \right\}$$
    to obtain search direction $s^k = [s_u^k, s_t^k]^T$.
    Set $u_{k+1} = u_k + \alpha_k^{\mathsf{N}} s_u^k$, $t_{k+1} = t_{k+1/2} + \alpha_k^{\mathsf{N}} s_t^k$, α_k^{N} obtained via line search.
end for

Easy to implement; computationally inexpensive to execute.

Convergence proof for line search and trust-region method extends.

Numerical Results

Apply Modification with Reduction to advection diffusion example

$$\min_{u,t}\ t + \frac{1}{N}\sum_{i=1}^{N} v_\epsilon\big(F(u,\xi_i) - t\big) + \frac{10^{-2}}{2}\big\|u\big\|^2_{H^1(\Gamma_c)},$$

where $y(x;u,\xi)$, $\xi \in \mathbb{R}^5$ solves

$$\begin{aligned} -\nabla \cdot \big(\kappa(x,\xi)\nabla y(x;u,\xi)\big) + c(x,\xi)\cdot\nabla y(x;u,\xi) &= f(x,\xi), && x \in \Omega \\ y(x;u,\xi) &= u(x,\xi), && x \in \Gamma_c \\ \big(\kappa(x,\xi)\nabla y(x;u,\xi)\big)\cdot n(x) &= 0, && x \in \Gamma_n, \end{aligned}$$

with quadratic QoI

$$F(u,\xi) = \frac{1}{2}\int_{\Omega_o}\big(y(x;u,\xi) - 1\big)^2 dx$$

Cost per Newton Iteration for PDE-Constrained Problems

$$\min_{u,t}\ t + \frac{1}{(1-\beta)N}\sum_{i=1}^{N} v_\epsilon\big(F(u,\xi_i) - t\big) + \frac{10^{-2}}{2}\|u\|^2_{H^1(\Gamma_c)},$$

- For a fixed sample, discretized PDE and QoI are:

$$\mathbf{A}(\xi)\mathbf{y}(\xi) + \mathbf{B}\mathbf{u} = \mathbf{f}(\xi)$$

$$F(\mathbf{u},\xi) = \frac{1}{2}\mathbf{y}(\mathbf{u},\xi)^T\mathbf{M}\mathbf{y}(\mathbf{u},\xi) + \mathbf{q}^T\mathbf{y}(\mathbf{u},\xi)$$

- Function evaluation requires N state solves
- Gradient requires N linear PDE solves

$$\begin{aligned}\nabla_u F(\mathbf{u},\xi) &= y_u^T(\mathbf{u},\xi)\nabla_y F(\mathbf{u},\xi)\\ &= \mathbf{B}^T\Big(\underbrace{-\mathbf{A}(\xi)^{-T}\big(\mathbf{M}\mathbf{y}+\mathbf{q}\big)}_{\text{linear PDE solve}}\Big)\end{aligned}$$

- Hessian-vector requires $2N$ linear PDE solves

$$\begin{aligned}\nabla_{uu} F(\mathbf{u},\xi)\mathbf{z} &= y_u^T(\mathbf{u},\xi)\nabla_{yy} F(\mathbf{u},\xi) y_u(\mathbf{u},\xi)\mathbf{z}\\ &= \mathbf{B}^T\mathbf{A}(\xi)^{-T}\mathbf{M}\mathbf{A}(\xi)^{-1}\mathbf{B}\mathbf{z}\end{aligned}$$

Plain vs Modified for Advection Diffusion

Trust Region Newton-CG on smoothed CVaR for sequence of decreasing smoothing parameters, $\beta = 0.9$; each subproblem solved to tol. $0.01 * \epsilon$.

	Standard		Modified	
N	iter	PDEs	iter	PDEs
100	41(1)	30,400	11(2)	10,600
500	61(1)	231,000	16(2)	75,000
1,000	51(1)	382,000	16(2)	148,000
5,000	43(1)	1,595,000	16(2)	770,000

Cost per iteration : N linear state PDEs (function)

$+N$ linear adjoint PDEs (gradient)

$+2N$ linear 2nd order adjoint PDEs (Hessian)

Plain vs Modified Distributed Control of Burgers' Eqn.

Trust Region Newton-CG on smoothed CVaR for sequence of decreasing smoothing parameters, $\beta = 0.9$; each subproblem solved to tol. $0.01 * \epsilon$.

	Standard			Modified		
N	iter	N PDEs	L PDEs	iter	N PDEs	L PDEs
100	42(2)	5,400	29,800	8(5)	1,400	10,200
500	44(2)	27,500	157,000	9(4)	7,500	56,500
1,000	40(2)	51,000	290,000	9(5)	15,000	119,000
5,000	42(2)	275,000	1,515,000	9(5)	75,000	605,000

Cost per iteration : N nonlinear state PDEs (function)
$+N$ linear adjoint PDEs (gradient)
$+2N$ linear 2nd order adjoint PDEs (Hessian)

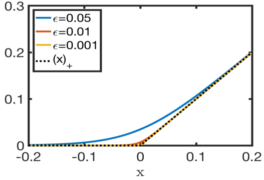
Outline

Reducing Cost of Gradient and Hessian Computations

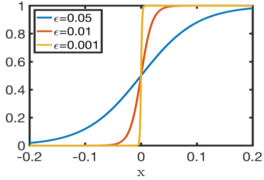
- Further reduce PDE solves in gradient and Hessian for

$$J_\epsilon(u,t) = t + \frac{1}{N}\sum_{i=1}^{N} v_\epsilon\big(F(u,\xi_i) - t\big) + \mathcal{P}(u)$$

- Derivatives of smoothing functions nearly vanish for many samples



v_ϵ



v'_ϵ

- Approximate u gradient with subset $\mathbb{G}_\epsilon[u,t] \subset \{\xi_1,\ldots,\xi_N\}$

$$\widetilde{\nabla_u J_\epsilon}(u,t) = \frac{1}{N}\sum_{\xi\in\mathbb{G}_\epsilon[u,t]} v'_\epsilon\big(F(u,\xi) - t\big)\nabla_u F(u,\xi_i) + \nabla\mathcal{P}(u)$$

- Use theory for trust region method with inexact gradients [Heinkenschloss and Vicente, 2002], [Kouri et al., 2014]

Inexact Gradient Condition

- Assume QoIs computed and ordered:

$$F(u,\xi_1) \geq \ldots \geq F(u,\xi_N)$$

- Use first j samples for gradient estimate $\widetilde{\nabla J_\epsilon}(u,t)$:

$$\widetilde{\nabla_u J_\epsilon}(u,t) = \frac{1}{N}\sum_{i=1}^{j} v'_\epsilon\big(F(u,\xi_i) - t\big)\nabla_u F(u,\xi_i) + \nabla \mathcal{P}(u)$$

- Use exact t gradient, requires no extra computation

$$\widetilde{\nabla_t J_\epsilon}(u,t) = 1 - \frac{1}{N}\sum_{i=1}^{N} v'_\epsilon\big(F(u,\xi_i) - t\big)$$

- Inexact gradient condition requires that at each iteration

$$\left\|\widetilde{\nabla J_\epsilon}(u,t) - \nabla J_\epsilon(u,t)\right\| \leq \gamma \min\left\{\left\|\widetilde{\nabla J_\epsilon}(u,t)\right\|, \Delta\right\}$$

for trust region radius Δ, $\gamma > 0$ independent of iteration

Estimate Gradient Error

- Assume $\|\nabla_u F(u,\xi)\| \le A$ for all u, ξ

$$\|\widetilde{\nabla_u J_\epsilon}(u,t) - \nabla_u J_\epsilon(u,t)\| = \frac{1}{N}\left\|\sum_{i=j+1}^{N} v'_\epsilon\big(F(u,\xi_i)-t\big)\nabla_u F(u,\xi_i)\right\|$$
$$\le \frac{1}{N}\sum_{i=j+1}^{N} \left|v'_\epsilon\big(F(u,\xi_i)-t\big)\right| \left\|\nabla_u F(u,\xi_i)\right\|$$
$$\le \frac{A}{N}\sum_{i=j+1}^{N} \left|v'_\epsilon\big(F(u,\xi_i)-t\big)\right|,$$

- Iteratively increase $j = 1, 2, \ldots$ until gradient condition is met

$$\frac{1}{N}\sum_{i=j+1}^{N} \left|v'_\epsilon\big(F(u,\xi_i)-t\big)\right| \le \mathrm{rtol}\cdot\min\left\{\|\widetilde{\nabla J_\epsilon}(u,t)\|, \Delta\right\}$$

Algorithm with Efficient Gradient Computations

Input: Approximate gradient tolerances rtol, atol
Output: Approximate gradient $\widetilde{\nabla_u J_\epsilon}(u,t)$
Compute and reorder QoIs s.t.

$$F(u,\xi_1) \geq \ldots \geq F(u,\xi_N)$$

Set $j = 0$
Set $\widetilde{\nabla_u J_\epsilon}(u,t) = 0$
$\mathrm{graderr} = \frac{1}{N}\sum_{i=j+1}^N \left| v_\epsilon'\big(F(u,\xi_i) - t\big)\right|$
while $\mathrm{graderr} > \mathrm{rtol} \cdot \min\{\|\widetilde{\nabla_u J_\epsilon}(u,t)\|, \mathrm{atol}\}$ **do**
  $j = j + 1$
  Compute and store $\nabla_u F(u,\xi_j)$
  Compute $\widetilde{\nabla_u J_\epsilon}(u,t) = \frac{1}{N}\sum_{i=1}^j v_\epsilon'\big(F(u,\xi_i) - t\big)\nabla_u F(u,\xi_i)$
  Compute $\mathrm{graderr} = \frac{1}{N}\sum_{i=j+1}^N \left| v_\epsilon'\big(F(u,\xi_i) - t\big)\right|$
end while

Plain vs Modified for Control of Advection Diffusion

Trust Region Newton-CG on smoothed CVaR for sequence of decreasing smoothing parameters, $\beta = 0.9$; each subproblem solved to tol. $10^{-2} * \epsilon$.

	Standard		Modified		Modified w. Red.	
N	iter	PDEs	iter	PDEs	iter	PDEs
100	41(1)	30,400	11(2)	10,600	10(2)	5,277
500	61(1)	231,000	16(2)	75,000	16(2)	30,176
1,000	51(1)	382,000	16(2)	148,000	19(2)	64,076
5,000	43(1)	1,595,000	16(2)	770,000	19(2)	326,381

Cost per iteration : N linear state PDEs (function)
$+N$ linear adjoint PDEs (gradient)
$+2N$ linear 2nd order adjoint PDEs (Hessian)

Plain vs Modified for Control of Advection Diffusion

Trust Region Newton-CG on smoothed CVaR for sequence of decreasing smoothing parameters, $\beta = 0.9$; each subproblem solved to tol. $10^{-2} * \epsilon$.

	Standard		Modified		Modified w. Red.	
N	iter	PDEs	iter	PDEs	iter	PDEs
100	41(1)	30,400	11(2)	10,600	10(2)	5,277
500	61(1)	231,000	16(2)	75,000	16(2)	30,176
1,000	51(1)	382,000	16(2)	148,000	19(2)	64,076
5,000	43(1)	1,595,000	16(2)	770,000	19(2)	326,381

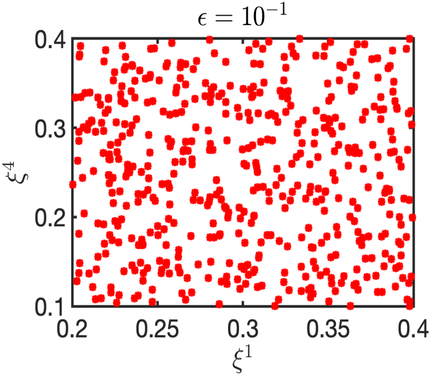
Cost per iteration : N linear state PDEs (function)
$+j$ linear adjoint PDEs (gradient)
$+2j$ linear 2nd order adjoint PDEs (Hessian)

Modifed with Reduction in Derivatives, $N = 500$

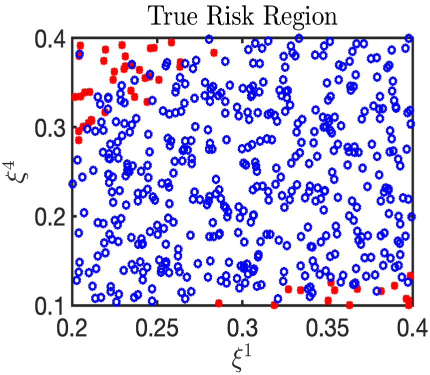
Trust Region Newton-CG on smoothed CVaR for sequence of decreasing smoothing params., $N = 500$, $\beta = 0.9$; each subprob. solved to tol. $10^{-2} * \epsilon$.

ϵ	iter	∇J_ϵ^*	RR	PDEs
1e-1	2(2)	1.8e-04	1.00	8,965
1e-2	1(2)	4.8e-04	1.00	4,985
1e-3	4(2)	1.9e-05	0.26	6,265
1e-4	2(3)	3.1e-08	0.11	2,459
1e-5	2(3)	5.4e-09	0.10	2,469
1e-6	5(2)	1.1e-09	0.10	5,033
Total	16(2)		0.43	30,176

Modifed with Reduction in Derivatives, $N = 500$



$\{\xi_1, \dots, \xi_j\}$



$\{\xi_i : F(u, \xi_i) > \mathsf{VaR}_\beta[g(u, \xi_i)]\}$

Plain vs Modified for Control of Burgers'

Trust Region Newton-CG on smoothed CVaR for sequence of decreasing smoothing parameters, $\beta = 0.9$; each subprob. solved to tol. $10^{-2} * \epsilon$.

	Standard			Modified			Modified w. Red.		
N	iter	NPDEs	LPDEs	iter	NPDEs	LPDEs	iter	NPDEs	LPDEs
100	42(2)	5,400	29,800	8(5)	1,400	10,200	9(5)	1,500	5,542
500	44(2)	27,500	157,000	9(4)	7,500	56,500	10(5)	8,000	31,077
1,000	40(2)	51,000	290,000	9(5)	15,000	119,000	10(5)	16,000	61,889
5,000	42(2)	275,000	1,515,000	9(5)	75,000	605,000	9(5)	75,000	255,381

Cost per iteration : N linear state PDEs (function)
$+N$ linear adjoint PDEs (gradient)
$+2N$ linear 2nd order adjoint PDEs (Hessian)

Plain vs Modified for Control of Burgers'

Trust Region Newton-CG on smoothed CVaR for sequence of decreasing smoothing parameters, $\beta = 0.9$; each subprob. solved to tol. $10^{-2} * \epsilon$.

	Standard			Modified			Modified w. Red.		
N	iter	NPDEs	LPDEs	iter	NPDEs	LPDEs	iter	NPDEs	LPDEs
100	42(2)	5,400	29,800	8(5)	1,400	10,200	9(5)	1,500	5,542
500	44(2)	27,500	157,000	9(4)	7,500	56,500	10(5)	8,000	31,077
1,000	40(2)	51,000	290,000	9(5)	15,000	119,000	10(5)	16,000	61,889
5,000	42(2)	275,000	1,515,000	9(5)	75,000	605,000	9(5)	75,000	255,381

Cost per iteration : N linear state PDEs (function)

$+j$ linear adjoint PDEs (gradient)

$+2j$ linear 2nd order adjoint PDEs (Hessian)

2D Boundary Control of Navier-Stokes

Trust Region Newton-CG on smoothed CVaR for sequence of decreasing smoothing parameters, $N = 500$, $\beta = 0.9$; each subprob. solved to tol $10^{-2} * \epsilon$

	Modified w/ Reduction				
ϵ	iter	$\nabla \widehat{J}^*_\epsilon$	RR	N PDE	L PDE
1e-1	7(11)	1.8e-03	0.94	4,000	82,791
1e-2	4(16)	1.0e-04	0.92	2,500	64,408
1e-3	6(13)	1.4e-06	0.30	3,500	29,380
1e-4	6(24)	3.9e-08	0.15	3,500	21,628
1e-5	6(12)	3.6e-07	0.10	3,500	8,892
Total	29(15)		0.40	17,000	207,099

Outline

Sample Reduction in Function Evaluations

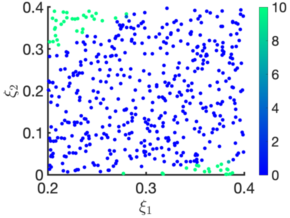
- Previous work requires N state solves for function evaluation

$$\|\nabla J_\epsilon(u,t) - \widetilde{\nabla J_\epsilon}(u,t)\| \le \frac{1}{N} \sum_{\xi \notin \mathbb{G}} \left| v'_\epsilon\big(F(u,\xi) - t\big) \right|$$

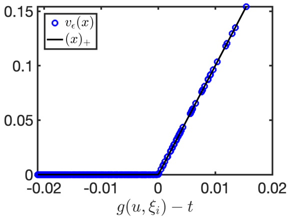
N state PDEs (function)

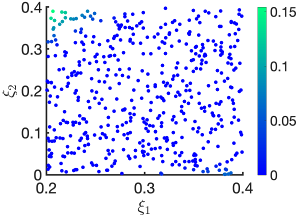
$+ \quad |\mathbb{G}|$ linear adjoint PDEs (gradient)

$+ \quad 2|\mathbb{G}| \cdot$ #CG linear PDEs (Hessian-vector)



- Next: use inexpensive QoI estimates from reduced order models (ROMs) to do sample reduction in function evaluation





Literature Review

Surrogate models in CVaR estimation with continuous expectation

- [Heinkenschloss et al., 2018], [Heinkenschloss et al., 2020] identify risk regions using ROMs
- [Zou et al., 2019] discretize uncertainty space using Voronoi cells, use ROM within each cell

Surrogate models in optimization under uncertainty

- Risk neutral optimization: [Zahr et al., 2019], [Kouri et al., 2014]
- [Yang and Gunzburger, 2017], replace full-order Helmholtz equation with POD ROM; limited error analysis

Here: Extend ideas in [Heinkenschloss et al., 2018], [Heinkenschloss et al., 2020] to optimization with discrete expectation, balance errors in a trust-region method as in [Kouri et al., 2014], [Zahr et al., 2019]

Newton TR with Inexact Function and Gradient

- Assume don't have direct access to function or gradient
- Approximate function at kth iteration $\widetilde{J}_\epsilon^k(u,t)$
- For global convergence, need to bound errors in gradient and function approximations [Kouri et al., 2014]
- Use QoI estimates F_r in function approximation

$$\left|F(u,\xi) - F_r(u,\xi)\right| \leq \delta_r(u,\xi)$$

- Investigate two approaches of using ROM to obtain function estimation $\widetilde{J}_\epsilon$ within a specified error tolerance, i.e.

$$|J_\epsilon(u,t) - \widetilde{J}_\epsilon(u,t)| \leq \mathrm{tol}$$

- Incorporate inexact function evaluations into TR framework with convergence guarantee

Two Approaches for Using ROMs

Approach 1

- Given a ROM of pre-specified accuracy
- Use ROM to estimate risk region such that function tolerance is satisfied
- Evaluate FOM in risk region
- Works with any surrogate that has error bound–don't need to refine surrogate

Approach 2

- Replace all FOM evaluations with ROM evaluations
- Requires ROM to be sufficiently accurate
- Only works if surrogate can be refined based on function tolerance

Will explore Approach 1

Compute F_r using projection based ROMs computed using Reduced Basis with greedy sampling (from N samples given in SAA)

Approach 1: Use FOM in Subset of Samples

Goal: estimate $J_\epsilon(u,t) = t + \frac{1}{N}\sum_{i=1}^{N} v_\epsilon\big(F(u,\xi_i) - t\big)$

- Use F_r to find region $\mathbb{G}_r[u,t]$ in which to evaluate g

$$\widetilde{J}_\epsilon(u,t) := t + \frac{1}{N}\sum_{\xi\in\mathbb{G}_r} v_\epsilon\big(F(u,\xi) - t\big)$$

- Error in approximation only due to truncation of sum

$$\big|J_\epsilon(u,t) - \widetilde{J}_\epsilon(u,t)\big| = \frac{1}{N}\Big|\sum_{\xi\notin\mathbb{G}_r} v_\epsilon\big(F(u,\xi) - t\big)\Big| \leq \mathrm{UB}(F_r,\delta_r)$$

- Given v_ϵ monotonic, upper bound error in function by

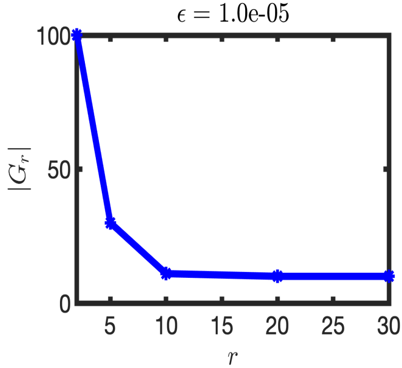
$$\mathrm{UB}(F_r,\delta_r) := \frac{1}{N}\sum_{\xi\notin\mathbb{G}_r} \max\Big\{\big|v_\epsilon\big(F_r(u,\xi) - t - \delta_r(\xi)\big|, \big|v_\epsilon\big(F_r(u,\xi) - t + \delta_r(\xi)\big|\Big\}$$

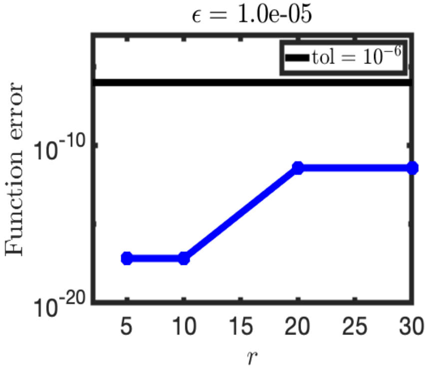
- Quality of F_r impacts amount of sample reduction, i.e. $|\mathbb{G}_r|$

Effect of ROM size on Function Approximation 1

For given ROM size r, find set $|\mathbb{G}_r|$ such that

$$\Big|\underbrace{J_\epsilon(u,t)}_{\sum_{i=1}^N} - \underbrace{\widetilde{J}_\epsilon(u,t)}_{\sum_{\xi\in\mathbb{G}_r}}\Big| \le \underbrace{\text{UB}(F_r,\delta_r)}_{\sum_{\xi\notin\mathbb{G}_r}} \le 10^{-6}$$





As r increases, $|\mathbb{G}_r|$ decreases

As $|\mathbb{G}_r|$ decreases,
UB increases, but stays below tol

Can reduce # FOMs even with low quality ROM

Distributed Control of Burgers' Equation, $N = 100$

ϵ	Newt(CG)	FOMs
10^{-1}	5(9)	600
10^{-2}	9(8)	1,000
10^{-3}	15(21)	1,600
10^{-4}	5(14)	600
10^{-5}	1(4)	200
10^{-6}	0(-)	100
Total	35(56)	4,100

Newton TR with exact function and gradient.

ϵ	Newt(CG)	avg $\lvert\mathbb{G}_k^{\text{grad}}\rvert$	avg $\lvert\mathbb{G}_k^{\text{func}}\rvert$	FOMs	ROMs
10^{-1}	5(9)	100	98	603	1,100
10^{-2}	9(8)	60	79	1,055	1,900
10^{-3}	14(20)	57	66	1,479	2,900
10^{-4}	5(14)	17	16	226	1,100
10^{-5}	1(4)	11	10	57	300
10^{-6}	0(-)	10	-	23	100
Total	34(55)			3,443	7,400

Newton TR with inexact function and gradient (Approach 1).

Conclusions

- Risk averse optimization of PDE constrained optimization challenging.
- Risk measures introduce non-smoothness in integrand, only samples in small subdomain of parameter region contribute.
- Many optimization approaches (direct smoothing, log-barrier, augmented Lagrangian) require solution of sequence of smoothed problems.
- Use Newton's method, but risk-averse objective introduces rank-deficiency, which leads to poor steps and inefficient method. Introduced easy to implement, comput. inexpensive modification to remove difficulties from inconsistent quadratic models.
- Introduced sample reduction by allowing inexact gradients and Hessians.
- Preliminary work on sampling strategy based on reduced order models to further reduce sample size for objective evaluation.
- More details in [Markowski, 2022].

Literature I

[Basova et al., 2011] Basova, H. G., Rockafellar, R. T., and Royset, J. O. (2011).
A computational study of the buffered failure probability in reliability-based design optimization.
In Faber, M., Koehler, J., and Kazuyoshi, N., editors, *Applications of Statistics and Probability in Civil Engineering*, pages 43–48. CRC Press, Boca Raton, DOI: 10.1201/b11332-24, http://dx.doi.org/10.1201/b11332-24.

[Ben-Tal and Teboulle, 2007] Ben-Tal, A. and Teboulle, M. (2007).
An old-new concept of convex risk measures: the optimized certainty equivalent.
Math. Finance, 17(3):449–476, DOI: 10.1111/j.1467-9965.2007.00311.x, https://doi.org/10.1111/j.1467-9965.2007.00311.x.

[Garreis et al., 2021] Garreis, S., Surowiec, T., and Ulbrich, M. (2021).
An Interior-Point Approach for Solving Risk-Averse PDE-Constrained Optimization Problems with Coherent Risk Measures.
SIAM J. Optim., 31(1):1–29, DOI: 10.1137/19M125039X, https://doi.org/10.1137/19M125039X.

[Heinkenschloss et al., 2020] Heinkenschloss, M., Kramer, B., and Takhtaganov, T. (2020).
Adaptive reduced-order model construction for conditional value-at-risk estimation.
SIAM/ASA J. Uncertain. Quantif., 8(2):668–692, DOI: 10.1137/19M1257433, https://doi.org/10.1137/19M1257433.

Literature II

[Heinkenschloss et al., 2018] Heinkenschloss, M., Kramer, B., Takhtaganov, T., and Willcox, K. (2018).
Conditional-value-at-risk estimation via reduced-order models.
SIAM/ASA J. Uncertain. Quantif., 6(4):1395–1423, DOI: 10.1137/17M1160069, https://doi.org/10.1137/17M1160069.

[Heinkenschloss and Vicente, 2002] Heinkenschloss, M. and Vicente, L. N. (2002).
Analysis of inexact trust-region SQP algorithms.
SIAM J. Optim., 12(2):283–302, DOI: 10.1137/S1052623499361543, https://doi.org/10.1137/S1052623499361543.

[Kouri et al., 2014] Kouri, D. P., Heinkenschloss, M., Ridzal, D., and van Bloemen Waanders, B. G. (2014).
Inexact objective function evaluations in a trust-region algorithm for PDE-constrained optimization under uncertainty.
SIAM J. Sci. Comput., 36(6):A3011–A3029, DOI: 10.1137/140955665, http://doi.org/10.1137/140955665.

[Kouri and Surowiec, 2016] Kouri, D. P. and Surowiec, T. M. (2016).
Risk-averse PDE-constrained optimization using the Conditional Value-at-Risk.
SIAM J. Optim., 26(1):365–396, DOI: 10.1137/140954556, http://doi.org/10.1137/140954556.

Literature III

[Kouri and Surowiec, 2020] Kouri, D. P. and Surowiec, T. M. (2020).
Epi-regularization of risk measures.
Math. Oper. Res., 45(2):774–795, DOI: 10.1287/moor.2019.1013, https://doi.org/10.1287/moor.2019.1013.

[Kouri and Surowiec, 2022] Kouri, D. P. and Surowiec, T. M. (2022).
A primal-dual algorithm for risk minimization.
Math. Program., 193(1, Ser. A):337–363, DOI: 10.1007/s10107-020-01608-9, https://doi.org/10.1007/s10107-020-01608-9.

[Mafusalov et al., 2018] Mafusalov, A., Shapiro, A., and Uryasev, S. (2018).
Estimation and asymptotics for buffered probability of exceedance.
European J. Oper. Res., 270(3):826–836, DOI: 10.1016/j.ejor.2018.01.021, https://doi.org/10.1016/j.ejor.2018.01.021.

[Markowski, 2022] Markowski, M. (2022).
Efficient Solution of Smoothed Risk-Averse PDE-Constrained Optimization Problems.
PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, TX, https://hdl.handle.net/1911/113220.

Literature IV

[Rockafellar and Royset, 2010] Rockafellar, R. T. and Royset, J. O. (2010).
On buffered failure probability in design and optimization of structures.
Reliability Engineering & System Safety, 95(5):499 – 510, DOI: 10.1016/j.ress.2010.01.001, http://dx.doi.org/10.1016/j.ress.2010.01.001.

[Rockafellar and Uryasev, 2000] Rockafellar, R. T. and Uryasev, S. (2000).
Optimization of conditional value-at-risk.
The Journal of Risk, 2(2):21–41, DOI: 10.21314/JOR.2000.038.

[Yang and Gunzburger, 2017] Yang, H. and Gunzburger, M. (2017).
Algorithms and analyses for stochastic optimization for turbofan noise reduction using parallel reduced-order modeling.
Comput. Methods Appl. Mech. Engrg., 319:217–239, DOI: 10.1016/j.cma.2017.02.030, https://doi.org/10.1016/j.cma.2017.02.030.

[Zahr et al., 2019] Zahr, M. J., Carlberg, K. T., and Kouri, D. P. (2019).
An Efficient, Globally Convergent Method for Optimization Under Uncertainty Using Adaptive Model Reduction and Sparse Grids.
SIAM/ASA J. Uncertain. Quantif., 7(3):877–912, DOI: 10.1137/18M1220996, https://doi.org/10.1137/18M1220996.

Literature V

[Zou et al., 2019] Zou, Z., Kouri, D. P., and Aquino, W. (2019).

An adaptive local reduced basis method for solving PDEs with uncertain inputs and evaluating risk.

Comput. Methods Appl. Mech. Engrg., 345:302–322, DOI:
10.1016/j.cma.2018.10.028,
https://doi.org/10.1016/j.cma.2018.10.028.