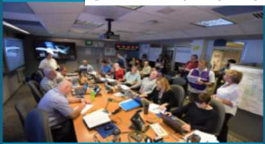
SAND2019-8675C

A Primal-Dual Algorithm for Large-Scale Risk Minimization



Drew P. Kouri, Thomas M. Surowiec

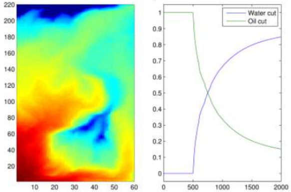
International Conference on Stochastic Programming

Trondheim, Norway

Motivating Applications

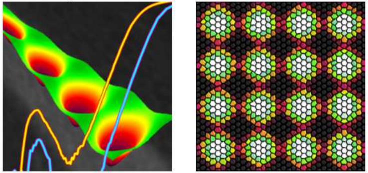
Simulation Constrained Optimization

Reservoir Optimization



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

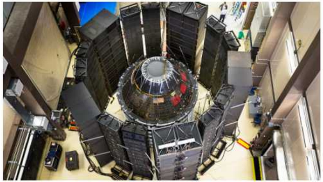
Superconductor Vortex Pinning



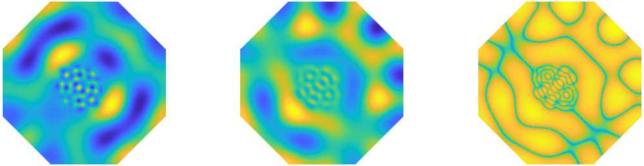
Courtesy Argonne National Laboratory

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

Direct Field Acoustic Testing



https://blogs.nasa.gov/orion/2016/03/07/engineers-test-new-acoustics-method-on-flown-orion/



$$\Delta u + \kappa^2(1+\sigma\epsilon)^2 u = 0, \qquad \nabla u \cdot n + i\chi\kappa\gamma u = i\rho_0\omega(1-\chi)z$$

Simulation Constrained Optimization

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let U and Z be reflexive Banach spaces and let Y be a Banach space. We consider the optimization problem

$$\min_{z \in Z_{\mathrm{ad}}} \{\mathcal{R}(f(S(z), \xi)) + \wp(z)\} \quad \text{where} \quad u = S(z, \xi) \quad \text{solves} \quad e(u, z, \xi) = 0 \quad \text{a.s.}$$

Here $\xi : \Omega \to \Xi$ are **random parameters (i.e., boundary/initial conditions, etc.)**

- $f : U \times \Xi \to \mathbb{R}$ is a ***state (simulation variable) objective function***,
- $\wp : Z \to \mathbb{R}$ is a ***control (design, etc.) objective function***,
- $e : U \times Z \times \Xi \to Y$ is the ***simulation constraint***, and
- $Z_{\mathrm{ad}} \subseteq Z$ is a **closed, convex** set of ***decision variables***.

Computational Cost to Evaluate $F(z) := f(S(z, \xi), \xi)$ and Its Derivatives:

Value	$e(u, z, \xi) = 0$	
Gradient	$e(u, z, \xi) = 0$	$e_u(u, z, \xi)^* \lambda = -f_u(u, \xi)$
HessVec	$e(u, z, \xi) = 0$	$e_u(u, z, \xi)^* \lambda = -f_u(u, \xi)$
	$e_u(u, z, \xi)s = -e_z(u, z, \xi)v$	$e_u(u, z, \xi)^* p = L_{uu}(u, z, \lambda, \xi)s - L_{uz}(u, z, \lambda, \xi)v$

Modeling Risk Preference

Choose Your Own Adventure

What is risk and how should we quantify it?

Risk (noun)**:** *Possibility of loss or injury* (Merriam Webster)

Optimistic Problem Formulations

- **Risk-Neutral Approach:**
  Minimize *on average*

$$\mathcal{R}(F(z)) = \mathbb{E}[F(z)].$$

- **Reliability Approach:**
  Minimize *probability of loss*

$$\mathcal{R}(F(z)) = \mathbb{P}(F(z) > x).$$

Conservative Problem Formulations

- **Risk-Averse Approach:**
  Model *risk preferences*

$$\mathcal{R}(F(z)) = \mathbb{E}[F(z)] + \mathcal{D}(F(z)).$$

- **Buffered Approach:**
  Minimize *buffered probability*

$$\mathcal{R}(F(z)) = \mathrm{bPOE}_x(F(z)).$$

Modeling Risk Preference

Choose Your Own Adventure

What is risk and how should we quantify it?

Risk (noun)**:** *Possibility of loss or injury* (Merriam Webster)

Optimistic Problem Formulations

- **Risk-Neutral Approach:**
  Minimize *on average*

$$\mathcal{R}(F(z)) = \mathbb{E}[F(z)].$$

- **Reliability Approach:**
  Minimize *probability of loss*

$$\mathcal{R}(F(z)) = \mathbb{P}(F(z) > x).$$

Conservative Problem Formulations

- **Risk-Averse Approach:**
  Model *risk preferences*

$$\mathcal{R}(F(z)) = \mathbb{E}[F(z)] + \mathcal{D}(F(z)).$$

- **Buffered Approach:**
  Minimize *buffered probability*

$$\mathcal{R}(F(z)) = \mathrm{bPOE}_x(F(z)).$$

Average Value-at-Risk and Buffered Probability

A **risk measure** is any $\mathcal{R} : \mathcal{X} \to (-\infty, \infty]$ such that $\mathcal{R}(C) = C$ for all $C \in \mathbb{R}$.

For example, $\mathcal{R}(X) = \mathbb{E}[X]$, $\mathcal{R}(X) = \mathbb{E}[X] + \mathbb{E}[|X - \mathbb{E}[X]|^p]^{1/p}$, or $\mathcal{R}(X) = \mathrm{AVaR}_\beta(X)$.

The **Average Value-at-Risk** is the *average of the* $(1-\beta) \times 100\%$ *largest scenarios*:

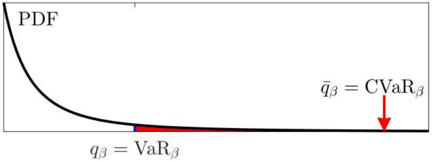
$$\mathrm{AVaR}_\beta(X) = \frac{1}{1-\beta} \int_\beta^1 q_\alpha(X)\, \mathrm{d}\alpha = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\beta} \mathbb{E}[\max\{0, X - t\}] \right\}.$$

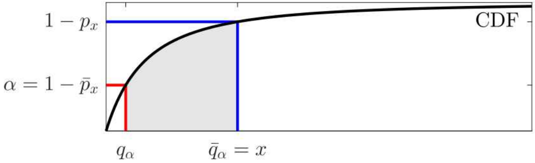
The **Buffered Probability** is the *probability that* $\mathrm{AVaR}_\beta(X)$ *exceeds a threshold* x:

$$\mathrm{bPOE}_x(X) = 1 - \beta \quad \text{where } \beta \text{ solves} \quad \mathrm{AVaR}_\beta(X) = x,$$

which can be computed by solving the convenient 1D convex optimization problem

$$\mathrm{bPOE}_x(X) = \min_{t \geq 0} \; \mathbb{E}[\max\{0, t(X - x) + 1\}].$$





Coherent Measures of Risk

Ph. Artzner et al., Mathematical Finance, 9(3), 1999.

$\mathcal{R} : \mathcal{X} \to (-\infty, \infty]$ is a **coherent** measure of risk if it satisfies

(R1) **Subadditivity:** $\mathcal{R}(X + X') \le \mathcal{R}(X) + \mathcal{R}(X')$

(R2) **Monotonicity:** $X \ge X' \text{ a.s.} \implies \mathcal{R}(X) \ge \mathcal{R}(X')$

(R3) **Translation Equivariance:** $\mathcal{R}(X + t) = \mathcal{R}(X) + t, \quad \forall t \in \mathbb{R}$

(R4) **Positive Homogeneity:** $\mathcal{R}(tX) = t\mathcal{R}(X), \quad \forall t > 0$

Convexity: {(R1) + (R4) $\implies$ convexity} and {convexity + (R4) $\implies$ (R1)}

Dual Representation: $\mathcal{R}(X) = \sup_{\vartheta \in \mathrm{dom}\,\mathcal{R}^*} \mathbb{E}[\vartheta X], \quad \mathrm{dom}\,\mathcal{R}^* \subseteq \{\theta \in \mathcal{X}^* \,|\, \mathbb{E}[\theta] = 1, \ \theta \ge 0 \text{ a.s.}\}$

Examples of risk measures that are **coherent**:

- Mean-Plus-Semideviation: $\mathcal{R}(X) = \mathbb{E}[X] + c\mathbb{E}[\max\{0, X - \mathbb{E}[X]\}^p]^{1/p}, \ c \in (0,1)$
- Average Value-at-Risk: $\mathcal{R}(X) = \inf_t\{t + (1-\beta)^{-1}\mathbb{E}[\max\{X - t, 0\}]\}, \ \beta \in (0,1)$

Examples of risk measures that are **not coherent**:

- Mean-Deviation: $\mathcal{R}(X) = \mathbb{E}[X] + \mathbb{E}[|X - \mathbb{E}[X]|^p]^{1/p}$ **Violates (R2)!**
- Entropic Risk: $\mathcal{R}(X) = \log \mathbb{E}[\exp X]$ **Violates (R4)!**

Is Nondifferentiability *Really* an Issue?

Result: If $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is **coherent**, then $\mathcal{R}$ is **Fréchet differentiable**

$$\iff \quad \exists\, \vartheta \in \mathcal{X}^* \text{ with } \vartheta \geq 0 \text{ a.s., } \mathbb{E}[\vartheta] = 1, \text{ and } \mathcal{R}(X) = \mathbb{E}[\vartheta X] \text{ for all } X \in \mathcal{X}$$

Nonsmooth, nonconvex, & stochastic simulation-constrained optimization:

— Algorithms for **nonsmooth, nonconvex** problems often converge **(sub)linearly**!

— Evaluating the **cost function** requires **simulations for every sample**!

— Evaluating **(sub)gradients** requires additional **linear solves for every sample**!

A small, nonconvex, & nonsmooth example: AVaR minimization of Burger's equation:

— **Bundle Method:** Required $\mathcal{O}(10^8)$ nonlinear and $\mathcal{O}(10^8)$ linearized solves.

— **Smoothing + Newton:** Required $\mathcal{O}(10^6)$ nonlinear and $\mathcal{O}(10^7)$ linearized solves.

Solving real world problems is intractable without . . .

- Better **nonsmooth** optimization algorithm or **differentiable** $\mathcal{R}$;
- **Adaptive/variable fidelity** approximation in physical and stochastic space;
- In optimization, accuracy is **not** required far from a solution.

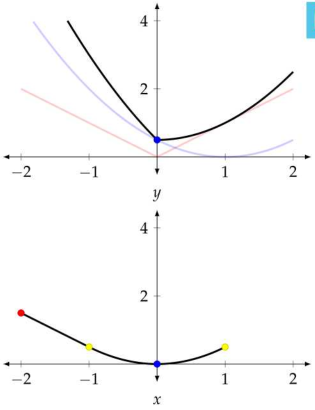
Epi-Regularized Risk Measures

The **epi-regularization** of $\mathcal{R}$ is given by

$$\mathcal{R}_\varepsilon^\Phi(X) := \inf_{Y\in\mathcal{X}} \left\{\mathcal{R}(X-Y) + \varepsilon\Phi(Y/\varepsilon)\right\}, \quad \varepsilon > 0$$

where $\mathcal{R}$, $\Phi : \mathcal{X} \to (-\infty, \infty]$ satisfy:

1. $\mathcal{R}$, Φ are proper, closed and convex
2. $\mathrm{dom}\,\mathcal{R}^* \subseteq \mathrm{dom}\,\Phi^* \implies \mathrm{dom}(\mathcal{R}_\varepsilon^\Phi)^* = \mathrm{dom}\,\mathcal{R}^*$
3. $(\mathrm{dom}\,\mathcal{R}^* - \mathrm{dom}\,\Phi^*)$ contains a neighborhood of 0



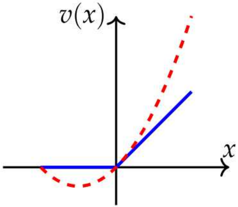
Properties of $\mathcal{R}_\varepsilon^\Phi$:

1. $|\mathcal{R}(X) - \mathcal{R}_\varepsilon^\Phi(X)| = \mathcal{O}(\varepsilon)$
2. $\mathcal{R}$ **coherent** $\implies$ $\mathcal{R}_\varepsilon^\Phi$ is a ***convex*** risk measure, but is **not** coherent
3. If Φ^* is *strictly convex* on $\mathrm{dom}\,\mathcal{R}^*$, then $\mathcal{R}_\varepsilon^\Phi$ is *Hadamard differentiable*
4. If, in addition, Φ is a *potential*, then $\mathcal{R}_\varepsilon^\Phi$ is *continuously differentiable*

Epi-Regularized Risk Measures

Let $v : \mathbb{R} \to \mathbb{R}$ and $\phi : \mathbb{R} \to \mathbb{R}$ be convex and define

$$\mathcal{R}(X) = \mathbb{E}[v(X)] \qquad \text{and} \qquad \Phi(X) = \mathbb{E}[\phi(X)].$$

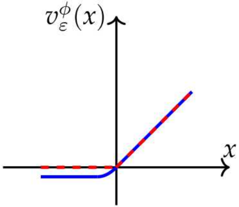


The **decomposability** of $\mathcal{X}$ ensures that

$$\mathcal{R}_\varepsilon^\Phi(X) = \inf_{Y \in \mathcal{X}} \{\mathbb{E}[v(X-Y)] + \varepsilon \mathbb{E}[\phi(Y/\varepsilon)]\}$$

$$= \mathbb{E}\left[\inf_{y \in \mathbb{R}} \{v(X-y) + \varepsilon\phi(y/\varepsilon)\}\right] = \mathbb{E}[v_\varepsilon^\phi(X)]$$

where the inner infimum, v_ε^ϕ, is the **infimal convolution** of v with ϕ!



Example: If $v(x) = \max\{0, x\}$ and $\phi(x) = \frac{1}{2}x^2 + x$, then

$$v_\varepsilon^\phi(x) = \begin{cases} -\frac{\varepsilon}{2} & \text{if } x \le -\varepsilon \\ \frac{1}{2\varepsilon}x^2 + x & \text{if } -\varepsilon < x < 0 \\ x & \text{if } x \ge 0 \end{cases}.$$

Primal-Dual Risk Minimization

We consider the optimization problem:

$$\min_{w \in W_{\mathrm{ad}}} \{g(w) + \Psi(G(w))\} \tag{P}$$

- W_{ad} is a closed, convex subset of the reflexive Banach space W,
- $g : W \to \mathbb{R}$ is weakly lower semicontinuous,
- $G : W \to \mathcal{X} := L^2(\Omega, \mathcal{F}, \mathbb{P})$ is weak-to-strong continuous,
- $\Psi : \mathcal{X} \to \mathbb{R}$ is convex, monotonic and positively homogeneous,
- $\exists\, \gamma \in \mathbb{R}$ such that $\{w \in W_{\mathrm{ad}} \,|\, g(w) + \Psi(G(w)) \le \gamma\}$ is nonempty and bounded.

Consequences: Problem (P) **has a solution** and Ψ is continuous, subdifferentiable and

$$\Psi(X) = \sup_{\theta \in \mathfrak{A}} \mathbb{E}[\theta X] \quad \forall\, X \in \mathcal{X} \quad \text{where} \quad \mathfrak{A} := \partial\Psi(0) \subseteq \{\theta \in \mathcal{X} \,|\, \theta \ge 0 \text{ a.s.}\}$$

$$\implies \quad \min_{w \in W_{\mathrm{ad}}} \{g(w) + \Psi(G(w))\} = \min_{w \in W_{\mathrm{ad}}} \sup_{\theta \in \mathfrak{A}} \{\ell(w, \theta) := g(w) + \mathbb{E}[\theta G(w)]\}.$$

Notation: Let $K := \sup_{\theta \in \mathfrak{A}} \|\theta\|_{\mathcal{X}}$ denote the Lipschitz modulus of Ψ at $X = 0$.

Primal-Dual Risk Minimization

Motivated by the method of multipliers, we define the *generalized augmented Lagrangian*

$$L(w,\lambda,r) := \max_{\theta\in\mathfrak{A}} \left\{ g(w) + \mathbb{E}[\theta G(w)] - \frac{1}{2r}\mathbb{E}[(\lambda-\theta)^2] \right\}, \quad r > 0.$$

Relation to Epi-Regularization: As a consequence of convex duality,

$$L(w,\lambda,r) = g(w) + \min_{Y\in\mathcal{X}} \left\{ \Psi(G(w)-Y) + \mathbb{E}[\lambda Y] + \frac{r}{2}\mathbb{E}[Y^2] \right\} = g(w) + \Psi^{\Phi}_{1/r}(G(w))$$

where $\Phi(Y) = \mathbb{E}[\lambda Y] + \frac{1}{2}\mathbb{E}[Y^2] \quad \Longrightarrow \quad 0 \le \Psi(X) - \Psi^{\Phi}_{1/r}(X) \le K^2/r$ for all $X \in \mathcal{X}$.

Consequences: $L(\cdot,\lambda,r)$ is continuously differentiable with derivative given by

$$\nabla_w L(w,\lambda,r) = \mathbf{P}_{\mathfrak{A}}(rG(w)+\lambda).$$

$L(w,\cdot,r)$ is also continuously differentiable with derivative given by

$$\nabla_\lambda L(w,\lambda,r) = (\mathbf{P}_{\mathfrak{A}}(rG(w)+\lambda)-\lambda)/r.$$

Primal-Dual Risk Minimization

We can rewrite the *generalized augmented Lagrangian* in the more revealing form

$$L(w,\lambda,r) = g(w) + \mathbb{E}[\lambda G(w)] + \frac{r}{2}\mathbb{E}[G(w)^2] - \frac{1}{2r}\mathbb{E}[\{(\mathbf{Id} - \mathbf{P}_{\mathfrak{A}})(rG(w)+\lambda)\}^2].$$

Equality Constraints ($G(w)=0$): Let Ψ be the *indicator function* of $\{0\}$, then $\mathfrak{A}=\mathcal{X}$ and

$$L(w,\lambda,r) = g(w) + \mathbb{E}[\lambda G(w)] + \frac{r}{2}\mathbb{E}[G(w)^2].$$

Inequality Constraints ($G(w)\le 0$): Let Ψ be the *indicator function* of $\{X\in\mathcal{X} \mid X\le 0 \text{ a.s.}\}$, then $\mathfrak{A}=\{\theta\in\mathcal{X}\mid\theta\ge 0 \text{ a.s.}\}$ and

$$L(w,\lambda,r) = g(w) + \frac{1}{2r}\mathbb{E}[\max\{0, rG(w)+\lambda\}^2] - \frac{1}{2r}\mathbb{E}[\lambda^2].$$

Primal-Dual Risk Minimization

Initialize: Given $\lambda_0 \in \mathfrak{A}$ and $r_0 > 0$.

While("Not Converged")

1. Find $w_{k+1} \in W_{\text{ad}}$ that *approximately* minimizes $L(\cdot, \lambda_k, r_k)$.
2. Set $\lambda_{k+1} = \mathbf{P}_{\mathfrak{A}}(r_k G(w_{k+1}) + \lambda_k)$.
3. Update r_{k+1}.

EndWhile

Practical Implementation: If W is a Hilbert space, then **"Converged"** could mean

$$\|w_{k+1} - \mathbf{P}_{W_{\text{ad}}}(w_{k+1} - \nabla_w L(w_{k+1}, \lambda_k, r_k))\|_W \le \tau_w \qquad \text{and} \qquad \|\lambda_k - \lambda_{k+1}\|_{\mathcal{X}} \le \tau_\lambda.$$

Moreover, we can update $r_{k+1} = \rho_r r_k$ for some $\rho_r > 0$ if $\|\lambda_k - \lambda_{k+1}\|_{\mathcal{X}} > \tau_{\lambda,k}$ with $\tau_{\lambda,k} > 0$.

Primal-Dual Risk Minimization

AVaR, bPOE, mean-plus-semideviation, ...

Example: Suppose $\Psi(X) = \mathbb{E}[\max\{0, X\}]$, then $\mathfrak{A} = \{\theta \in \mathcal{X} \mid 0 \le \theta \le 1 \text{ a.s.}\}$.

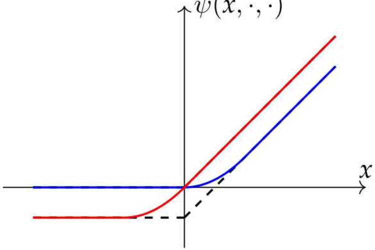
While("Not Converged")

1. Find $w_{k+1} \in W_{\mathrm{ad}}$ that *approximately* solves

$$\min_{w \in W_{\mathrm{ad}}} \{g(w) + \mathbb{E}[\psi(G(w), \lambda_k, r_k)]\}.$$

2. Set $\lambda_{k+1} = \max\{0, \min\{1, r_k G(w_{k+1}) + \lambda_k\}\}$.
3. Update r_{k+1}.

EndWhile



$$\psi(x,t,r) = \begin{cases} -\frac{1}{2r}t^2 & \text{if } rx + t < 0 \\ \frac{r}{2}x^2 + tx & \text{if } 0 \le rx + t \le 1 \\ \frac{1}{r}\{(rx+t) - \frac{1}{2}(t^2+1)\} & \text{if } 1 < rx + t \end{cases}$$

Note: Equivalent to the **method of multipliers** applied to the smooth reformulation

$$\min_{w \in W_{\mathrm{ad}},\, s, \eta \in \mathcal{X}} \{g(w) + \mathbb{E}[\eta]\} \quad \text{subject to} \quad G(w) - \eta + s = 0, \;\; \eta \ge 0, \;\; s \ge 0 \;\text{ a.s.}$$

1. **Primal Variables:** Let $\epsilon_k \to \epsilon^\star \geq 0$ and $r_k \to r^\star \leq \infty$. If the iterates $\{w_k\} \subset W_{\mathrm{ad}}$ are ϵ_k-minimizers, then **any weak accumulation point of $\{w_k\}$ is a $(\epsilon^\star + \frac{K^2}{r^\star})$-minimizer of the original problem**.

2. **Dual Variables:** If, in addition, $\{\epsilon_k\}$ satisfies

$$\epsilon_k = \frac{\gamma_k^2}{2r_k}, \quad \sum_{k=0}^{\infty} \gamma_k < \infty, \quad \text{and} \quad \gamma_k \geq 0,$$

then **the dual variables $\{\lambda_k\}$ converge weakly to a maximizer of the dual problem**.

3. **Primal Variables:** Let $\epsilon_k \to 0$ and $r_k \to \infty$, and suppose g and G are continuously differentiable. If the iterates $\{w_k\} \subset W_{\mathrm{ad}}$ are ϵ_k-stationary points then **any weak accumulation point of $\{w_k\}$ is a stationary point of the original problem**.

Elliptic 1d: $D = (-1, 1)$, $\alpha = 10$, $Z = Z_{\text{ad}} = L^2(D)$

$$\min_{z \in Z_{\text{ad}}} \mathcal{R}\left(\frac{1}{2}\int_D (S(z) - 1)^2 \,\mathrm{d}x\right) + \frac{\alpha}{2}\int_D z^2 \,\mathrm{d}x$$

where $u = S(z)$ solves

$$\begin{aligned} -\partial_x(\epsilon(\xi)\partial_x u(\xi)) &= f(\xi) + z && \text{in } D \text{ a.s.} \\ [u(\xi)](-1) = 0, \quad [u(\xi)](1) &= 0 && \text{a.s.} \end{aligned}$$

Elliptic 2d: $D = (0, 1)^2$, $\alpha = 10^{-5}$, $Z = \mathbb{R}^9$, $Z_{\text{ad}} = \{z \in Z \,|\, 0 \le z \le 1\}$

$$\min_{z \in Z_{\text{ad}}} \mathcal{R}\left(\frac{1}{2}\int_D S(z)^2 \,\mathrm{d}x\right) + \alpha\|z\|_1$$

where $u = S(z)$ solves

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

Numerical Examples

Burgers: $D = (0,1)$, $\alpha = 10^{-3}$, $Z = Z_{\mathrm{ad}} = L^2(D)$

$$\min_{z \in Z_{\mathrm{ad}}} \mathcal{R}\left(\frac{1}{2}\int_D (S(z)-1)^2\,\mathrm{d}x\right) + \frac{\alpha}{2}\int_D z^2\,\mathrm{d}x$$

where $u = S(z)$ solves

$$-\nu(\xi)\partial_{xx}u(\xi) + u(\xi)\partial_x u(\xi) = f(\xi) + z \qquad \text{in } D \text{ a.s.}$$

$$[u(\xi)](0) = d_0(\xi), \quad [u(\xi)](1) = d_1(\xi) \qquad \text{a.s.}$$

Risk Measures:

Mean-Plus-Semideviation	$\mathcal{R}(X) = \mathbb{E}[X] + c\mathbb{E}[\max\{0, X - \mathbb{E}[X]\}]$
Mean-Plus-Semideviation-From-Target	$\mathcal{R}(X) = \mathbb{E}[X] + c\mathbb{E}[\max\{0, X - t\}]$
Average Value-at-Risk	$\mathcal{R}(X) = \lambda\mathbb{E}[X] + (1-\lambda)\mathrm{AVaR}_\beta(X)$
Buffered Probability	$\mathcal{R}(X) = \inf_{t\geq 0}\mathbb{E}[\max\{0, t(X-x)+1\}]$

Numerical Results

Comparison with Nonsmooth Bundle Method (Schramm and Zowe)

example	risk	PD Algorithm iter	PD Algorithm nfval	PD Algorithm ngrad	PD Algorithm subiter	Bundle iter	Bundle neval
elliptic 1d	MPSD	7	14	14	7	37	530
	MPSDFT	7	11	11	4	28	427
	CVAR	7	23	23	16	37	240
	BPOE	7	66	59	33	---	---
elliptic 2d	MPSD	5	15	15	5	---	---
	MPSDFT	6	21	20	8	---	---
	CVAR	9	99	57	31	---	---
	BPOE	10	123	72	47	---	---
burgers	MPSD	14	35	30	21	362	395
	MPSDFT	11	23	23	12	329	361
	CVAR	11	63	63	52	369	466
	BPOE	11	179	129	76	---	---

Between 7 and 38 fold reduction in computational work!

Conclusions:

- **Numerical solution** of risk-averse simulation-constrained optimization is **expensive**
- Most **coherent risk measures** are **not** continuously differentiable
- We can use the **infimal convolution** to **smooth** risk measures
- Appropriate assumptions ensure smoothed risk **is** continuously differentiable
- Generalized method of multipliers solves a sequence of **smooth, epi-regularized** subproblems
- Proved **convergence** of approximate minimizers and first-order stationary points
- Numerical examples suggest $\sim$**10–40x improvement** compared to bundle method

References:

- D. P. Kouri, T. M. Surowiec. A primal-dual algorithm for risk minimization. Submitted. 2018.
- D. P. Kouri, T. M. Surowiec. Epi-regularization of risk measures. *To appear in* Mathematics of Operations Research, 2019.
- D. P. Kouri, T. M. Surowiec. Existence and optimality conditions for risk-averse PDE-constrained optimization. SIAM/ASA Journal on Uncertainty Quantification, 6(2):787–815, 2018.
- D. P. Kouri, T. M. Surowiec. Risk-averse PDE-constrained optimization using the conditional value-at-risk. SIAM Journal on Optimization, 26(1):365–396, 2016.