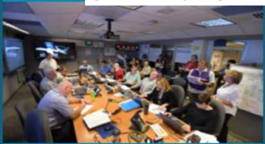
Sandia National Laboratories

SAND2018-12893PE



# A Primal-Dual Algorithm for Large-Scale Risk Minimization

**Drew P. Kouri**

Center for Applied Mathematics Colloquium

Cornell University, Ithaca NY

November 16, 2018

DEFENSE SCIENCES OFFICE

U.S. DEPARTMENT OF ENERGY

NNSA

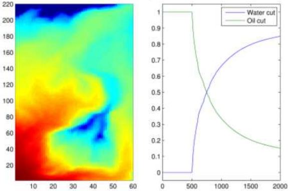
Sandia National Laboratories is a multimission
Administration under contract DE-NA0003525.

# Outline

1. Motivating Applications
2. Problem Formulation and Quantifying Risk
3. Coherent Risk Measures and Nondifferentiability
4. Smooth Risk Measures via the Infimal Convolution
5. Primal-Dual Risk Minimization
6. Numerical Results
7. Conclusions

# Motivating Applications

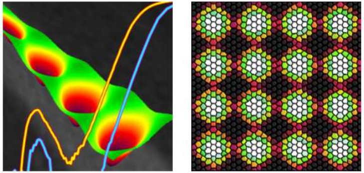
### Reservoir Optimization



$$v = -\mathbf{K}\lambda(s)\nabla p, \quad \nabla \cdot v = q$$

$$\phi\,\partial_t s + \nabla \cdot (f(s)v) = \widehat{q}$$

### Superconductor Vortex Pinning



Courtesy Argonne National Laboratory

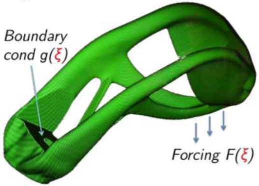
$$\gamma(\partial_t + i\mu)\psi = \epsilon\psi - |\psi|^2\psi + (\nabla - i\mathbf{A})^2\psi$$

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

### Direct Field Acoustic Testing

$$-\Delta u - \kappa^2(1 + \sigma\epsilon)^2 u = z$$

### Design for Additive Manufacturing



$$-\nabla \cdot (\mathbf{E}(z) : \varepsilon(u)) = F, \quad \varepsilon(u) = \tfrac{1}{2}(\nabla u + \nabla u^\top)$$

# Simulation Constrained Optimization

Let $U$ and $Z$ be reflexive Banach spaces and let $Y$ be a Banach space. We consider the optimization problem

$$\boxed{\min_{z\in Z_{\mathrm{ad}},\, u\in U} \{f(u)+\wp(z)\} \quad \text{subject to} \quad e(u,z)=0}$$

where $f : U \to \mathbb{R}$ is a ***state objective function***,

$\wp : Z \to \mathbb{R}$ is a ***control objective function***,

$Z_{\mathrm{ad}} \subseteq Z$ is a **closed, convex** set of ***decision variables***, and

$e : U \times Z \to Y$ is the ***simulation constraint***.

If $e(u,z)=0$ has a unique solution $S(z)\in U$ for each $z\in Z_{\mathrm{ad}}$, then we can solve the *reduced* optimization problem

$$\boxed{\min_{z\in Z_{\mathrm{ad}}} \{F(z)+\wp(z)\} \quad \text{where} \quad F(z)=f(S(z)).}$$

# Simulation Constrained Optimization

Consider the abstract (deterministic) minimization problem

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

**Objective Evaluation** **(1 solve)**: To evaluate $F(z) := f(S(z))$, one **must solve**

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

**Gradient Evaluation** **(2 solves)**: To evaluate $\nabla F(z)$, one **must solve**

$$e(u,z) = 0 \quad \text{and} \quad e_u(S(z),z)^*\lambda = -\nabla f(S(z)).$$

**Hessian-Times-A-Vector** **(4 solves)**: To evaluate $\nabla^2 F(z)v$, one **must solve**

$$e(u,z) = 0, \quad e_u(S(z),z)^*\lambda = -\nabla f(S(z)), \quad e_u(S(z),z)s = -e_z(S(z),z)v,$$
$$\text{and} \quad e_u(S(z),z)^*p = L_{uu}(S(z),z,\lambda)s - L_{uz}(S(z),z,\lambda)v$$

where $L(u,z,\lambda) = f(u) + \langle\lambda, e(u,z)\rangle$.

# Optimization Under Uncertainty

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $Z$ be a reflexive Banach space, and $\mathcal{X} := L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $1 \le p < \infty$. We consider the optimization problem

$$\boxed{\min_{z \in Z_{\mathrm{ad}}} \{\mathcal{R}(F(z)) + \wp(z)\}}$$

where $\mathcal{R} : \mathcal{X} \to (-\infty, \infty]$ is a ***quantification of risk***,

$F : Z \to \mathcal{X}$ is the ***uncertain objective function***,

$\wp : Z \to \mathbb{R}$ is a ***deterministic objective function***, and

$Z_{\mathrm{ad}} \subseteq Z$ is a **closed, convex** set of ***decision variables***.

# Modeling Risk Preference

Choose Your Own Adventure

**What is risk?** *Possibility of loss or injury* (Merriam Webster)

. . . In our optimization problem, $F(z)$ is a **risk**!

We **cannot** directly minimize $F(z) + \wp(z) \in \mathcal{X} := L^p(\Omega, \mathcal{F}, \mathbb{P})$

. . . How should we quantify our risk?

## Traditional Problem Formulations

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

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

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

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

## Modern 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*

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

# Quantifying Risk and Controlling Uncertainty

- Reduce **variability** of optimized system:

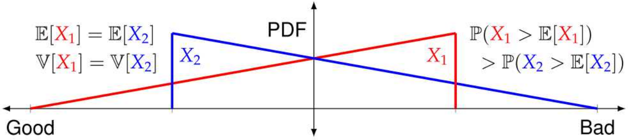
$$\mathbb{E}[(X - \mathbb{E}[X])^2] \qquad \text{or} \qquad \mathbb{E}[\max\{0, X - \mathbb{E}[X]\}^p]^{1/p}$$

- Control **rare events**, reduce **failure**, and certify **reliability**:

$$\sup X \qquad \text{or} \qquad \mathrm{VaR}_\beta(X) = q_\beta(X) = \inf\{\, t \in \mathbb{R} \mid \mathbb{P}(X \le t) \ge \beta \,\}$$

- Minimize statistics (e.g., average) of **undesirable events**:

$$\mathrm{CVaR}_\beta(X) = \textit{average of the } (1-\beta) \times 100\% \textit{ largest scenarios}$$



# Conditional Value-at-Risk and Buffered Probability

A **risk measure** is any $\mathcal{R} : \mathcal{X} \to (-\infty, \infty]$ such that $X \in \mathcal{X}$ and $\mathcal{R}(X)$ have the same units.

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

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

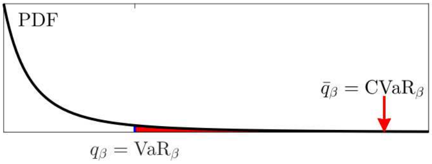
$$\mathrm{CVaR}_\beta(X) = \frac{1}{1-\beta} \int_\beta^1 q_\alpha(X)\,\mathrm{d}\alpha = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\beta} \mathbb{E}[\max\{0, X - t\}] \right\}.$$

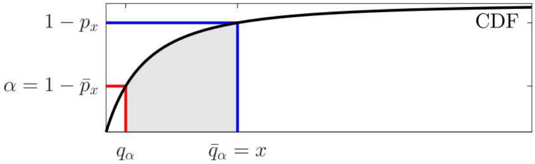
The **Buffered Probability** is the *probability that* $\mathrm{CVaR}_\beta(X)$ *exceeds a threshold* $x$:

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

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

Definition and Examples

$\mathcal{R} : \mathcal{X} \to (-\infty, \infty]$ 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:** $\mathcal{R}(tX) = t\mathcal{R}(X), \quad \forall t > 0$

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

Examples of **coherent** risk measures with $X \in \mathcal{X}$:

- Risk Neutral: $\mathcal{R}(X) = \mathbb{E}[X]$
- Mean Plus Semideviation: $\mathcal{R}(X) = \mathbb{E}[X] + c\mathbb{E}[\max\{0, X - \mathbb{E}[X]\}^p]^{1/p}, \; c \in (0, 1)$
- Conditional Value-at-Risk: $\mathcal{R}(X) = \mathrm{CVaR}_\beta(X), \; \beta \in (0, 1)$

**Ph. Artzner, F. Delbaen, J.-M. Eber and D. Heath,** *Coherent measures of risk.* Math. Finance, 1999.

# Convex and Variational Analysis

Let $V$ be a normed vector space with topological dual space $V^*$ and let $\Phi : V \to [-\infty, \infty]$.

1. $\Phi$ is **proper** if $\Phi(v) > -\infty \; \forall\, v \in V$ and $\exists\, v_0 \in V$ such that $\Phi(v_0) < \infty$.
2. The **epigraph** of $\Phi$ is the set
$$\operatorname{epi}\Phi := \{\, (t, v) \in \mathbb{R} \times V \mid \Phi(v) \le t \,\}.$$
3. $\Phi$ is **closed** (lsc) if its epigraph is closed in the product topology on $\mathbb{R} \times V$.
4. The **effective domain** of $\Phi$ is the set
$$\operatorname{dom}\Phi := \{\, v \in V \mid \Phi(v) < \infty \,\}.$$
5. The **Fenchel conjugate** of $\Phi$, denoted $\Phi^* : V^* \to [-\infty, \infty]$, is
$$\Phi^*(\vartheta) := \sup_{v \in V}\{\langle \vartheta, v\rangle_{V^*,V} - \Phi(v)\} = \sup_{v \in \operatorname{dom}\Phi}\{\langle \vartheta, v\rangle_{V^*,V} - \Phi(v)\}.$$
6. **Fenchel-Moreau Theorem:** If $\Phi$ is proper, closed and convex, then $\Phi = \Phi^{**}$.

# Coherent Measures of Risk

Some Good and *Not* So Good Properties?

**Biconjugate Representation:** Recall $\mathcal{R}^*(\vartheta) = \sup_X\{\mathbb{E}[\vartheta X] - \mathcal{R}(X)\}$

- If $\mathcal{R}$ is proper, closed and **convex**

$$\iff \qquad \mathcal{R}(X) = \sup\left\{\mathbb{E}[\vartheta X] - \mathcal{R}^*(\vartheta) \;\middle|\; \vartheta \in \operatorname{dom}\mathcal{R}^*\right\}$$

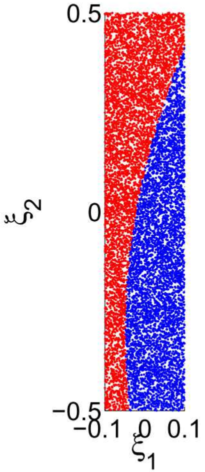
- If $\mathcal{R}$ is **translation equivariant** and **monotonic**

$$\iff \qquad \operatorname{dom}\mathcal{R}^* \subseteq \{\vartheta \in \mathcal{X}^* \mid \mathbb{E}[\vartheta] = 1,\ \vartheta \geq 0 \text{ a.s.}\}$$

- If $\mathcal{R}$ is **positive homogeneous**

$$\iff \qquad \mathcal{R}(X) = \sup_{\vartheta\in\operatorname{dom}\mathcal{R}^*} \mathbb{E}[\vartheta X]$$

$\operatorname{dom}\mathcal{R}^*$ is the ***risk envelope*** and optimal $\vartheta^\star \in \operatorname{dom}\mathcal{R}^*$ are ***risk identifiers***



**Example (Conditional Value-at-Risk):** $\mathcal{R} = \mathrm{CVaR}_\beta$

$$\operatorname{dom}\mathcal{R}^* = \left\{\vartheta \in \mathcal{X}^* \;\middle|\; \mathbb{E}[\vartheta] = 1,\ 0 \leq \vartheta \leq \frac{1}{1-\beta} \text{ a.s.}\right\}$$

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

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

# Is Nondifferentiability *Really* an Issue?

**Example:** Optimal control of Burger's equation using CVaR

- **Problem size is small:** 1D spatial domain, 4D stochastic domain
- PDE is nonlinear $\Longrightarrow$ Objective function is not convex
- CVaR risk measure quantifies *tail weight* and is **not** differentiable

**Application of the *nonconvex bundle* method by Schramm and Zowe:**

| $\beta$ | 0.1 | 0.5 | 0.9 |
|---|---|---|---|
| # iter | 9,740 | 10,035 | 10,128 |

**Required** $\mathcal{O}(10^8)$ nonlinear and $\mathcal{O}(10^8)$ linearized PDE solves!

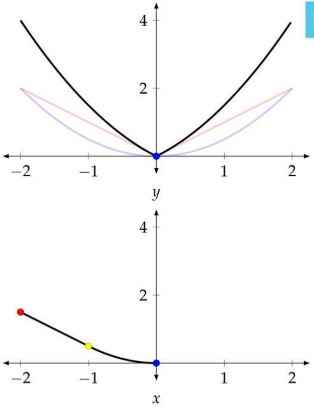
**Application of smoothed $\mathcal{R}$ with globalized Newton's method:**

**Required** $\mathcal{O}(10^6)$ nonlinear and $\mathcal{O}(10^7)$ linearized PDE 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



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

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

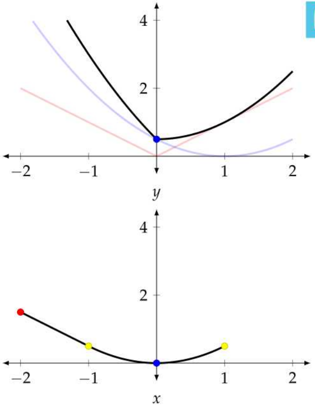
The **epi-regularization** of $\mathcal{R}$ is given by

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

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

1. $(\mathcal{R}_\varepsilon^\Phi)^*(\vartheta) = \mathcal{R}^*(\vartheta) + \varepsilon \Phi^*(\vartheta)$ with $\operatorname{dom}(\mathcal{R}_\varepsilon^\Phi)^* = \operatorname{dom} \mathcal{R}^* \cap \operatorname{dom} \Phi^* = \operatorname{dom} \mathcal{R}^*$
2. $-\varepsilon \Phi(0) \leq \mathcal{R}(X) - \mathcal{R}_\varepsilon^\Phi(X) \leq \varepsilon \Phi^*(\vartheta) \quad \forall \vartheta \in \partial \mathcal{R}(X)$
3. **Coherent Risk:** $\mathcal{R}_\varepsilon^\Phi$ is convex, translation equivariant and monotonic
4. **Coherent Risk:** $\mathcal{R}_\varepsilon^\Phi$ is **not** positively homogeneous $\implies$ $\mathcal{R}_\varepsilon^\Phi$ is **not** coherent

# Epi-Regularized Risk Measures



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

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

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

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

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

1. If $\Phi^*$ is *strictly convex* on $\operatorname{dom} \mathcal{R}^*$, then $\mathcal{R}_\varepsilon^\Phi$ is *Hadamard differentiable*
2. If, in addition, $\Phi^*$ is *weak* closed* and satisfies

$$\theta_k \rightharpoonup^* \theta \quad \text{in } \mathcal{X}^* \quad \text{and} \quad \Phi^*(\theta_k) \to \Phi^*(\theta) \quad \Longrightarrow \quad \theta_k \to \theta \quad \text{in } \mathcal{X}^*,$$

then $\mathcal{R}_\varepsilon^\Phi$ is *continuously Fréchet differentiable*

# Example: Optimized Certainty Equivalents

Let $u(t) = -v(-t)$ is a normalized, concave utility function and define $\mathcal{R}(X) = \inf_t\{t + \mathbb{E}[v(X-t)]\}$ and $\Phi(X) = \mathbb{E}[\phi(X)]$, then

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

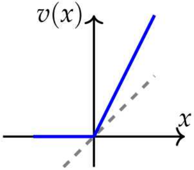
**Decomposability** of $\mathcal{X}$ ensures that

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

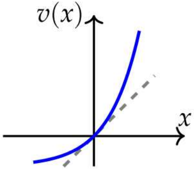
The inner infimum is the **infimal convolution** of $v(x)$ with $\phi(x)$

$$\boxed{\implies \quad \mathcal{R}_\varepsilon^\Phi(X) = \inf_{t\in\mathbb{R}}\{t + \mathbb{E}[v_\varepsilon^\phi(X-t)]\}!}$$

**CVaR**



**Entropic Risk**

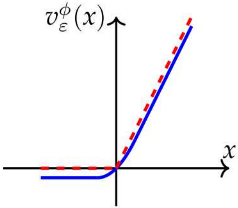


# Example: Optimized Certainty Equivalents

**CVaR:** $v(x) = \frac{1}{1-\beta}\max\{0, x\}$ and $\phi(x) = \frac{1}{2}x^2 + x$

$$\Longrightarrow \quad {\color{blue}\mathcal{R}}(X) = \mathrm{CVaR}_\beta(X)$$

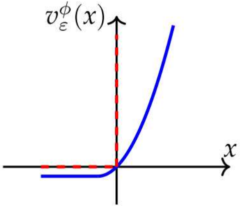
$$v_\varepsilon^\phi(x) = \begin{cases} -\frac{\varepsilon}{2} & \text{if } x \le 0 \\ \frac{1}{2\varepsilon}x^2 + x & \text{if } -\varepsilon < x < \frac{\varepsilon\beta}{1-\beta} \\ \frac{1}{1-\beta}\left(x - \frac{\varepsilon\beta^2}{2(1-\beta)}\right) & \text{if } x \ge \frac{\varepsilon\beta}{1-\beta}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



**Worst-Case Scenario:** $v(x) = \delta_{(-\infty,0]}(x)$ and $\phi(x) = \frac{1}{2}x^2 + x$

$$\Longrightarrow \quad {\color{blue}\mathcal{R}}(X) = \operatorname{ess\,sup} X$$

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



# Primal-Dual Risk Minimization

We consider the optimization problem:

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

- $W_{\mathrm{ad}}$ 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 $\Psi$ is continuous, subdifferentiable and

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

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

**Notation:** Let $K := \sup_{\theta\in\mathfrak{A}} \|\theta\|_{\mathcal{X}}$ denote the Lipschitz modulus of $\Phi$ 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 Fréchet 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 Fréchet 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 convenient 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 $\Psi$ 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 ($G(w) \leq 0$):** Let $\Psi$ be the *indicator function* of $\{X \in \mathcal{X} \,|\, X \leq 0 \text{ a.s.}\}$, then $\mathfrak{A} = \{\theta \in \mathcal{X} \,|\, \theta \geq 0 \text{ a.s.}\}$ and

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

# Primal-Dual Risk Minimization

The Algorithm

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

While("Not Converged")
    1. Find $w_{k+1} \in W_{\mathrm{ad}}$ 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

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

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

1. **Primal Variables:** Let $\epsilon_k \to \epsilon^\star \geq 0$ and $r_k \to r^\star \leq \infty$. If the iterates $\{w_k\} \subset W_{\mathrm{ad}}$ satisfy

$$L(w_{k+1}, \lambda_k, r_k) - \epsilon_k \leq \min_{w \in W_{\mathrm{ad}}} L(w, \lambda_k, r_k),$$

then **any weak accumulation point**, $w^\star \in W_{\mathrm{ad}}$, of $\{w_k\}$ satisfies

$$g(w^\star) + \Psi(G(w^\star)) - \left(\epsilon^\star + \tfrac{K^2}{r^\star}\right) \leq \min_{w \in W_{\mathrm{ad}}} \{g(w) + \Psi(G(w))\}.$$

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.**

# Primal-Dual Risk Minimization

3. **Primal Variables:** Let $\epsilon_k \to 0$ and $r_k \to \infty$, and suppose $g$ and $G$ are continuously Fréchet differentiable. If the iterates $\{w_k\} \subset W_{\mathrm{ad}}$ satisfy

$$\langle \nabla_w L(w_{k+1}, \lambda_k, r_k), w - w_{k+1} \rangle_{W^*,W} \geq -\epsilon_k \|w - w_{k+1}\|_W \quad \forall\, w \in W_{\mathrm{ad}},$$

then **any weak accumulation point**, $w^\star \in W_{\mathrm{ad}}$, of $\{w_k\}$ satisfies:
$\exists\, \lambda^\star \in \partial\Psi(G(w^\star))$ such that

$$\langle \nabla g(w^\star) + \mathbb{E}[\lambda^\star G'(w^\star)], w - w^\star \rangle_{W^*,W} \geq 0 \quad \forall\, w \in W_{\mathrm{ad}}.$$

# Primal-Dual Risk Minimization

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

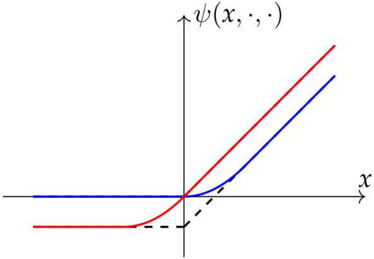
**Risk Measures:** This case covers CVaR, bPOE, mean-plus-semideviation, ...

**Smooth Reformulation:** We can reformulate our optimization problem as

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

Applying the **method of multipliers to the equality constraint** and **explicitly minimizing** for $\eta \ge 0$ and $s \ge 0$ is **equivalent** to our algorithm!

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



# Primal-Dual Risk Minimization

**Recall:** $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is a *coherent* measure of risk if

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

$$\implies \quad \mathcal{R}(X) = \sup_{\theta \in \mathfrak{D}} \mathbb{E}[\theta X] \quad \text{where} \quad \mathfrak{D} = \partial\mathcal{R}(0) \subset \{\theta \in \mathcal{X} \,|\, \mathbb{E}[\theta] = 1, \;\; \theta \ge 0 \text{ a.s.}\}.$$

**Decompose:** $\mathfrak{D} = \mathfrak{A} \cap \{\theta \in \mathcal{X} \,|\, \mathbb{E}[\theta] = 1\}$ where $\mathbf{P}_{\mathfrak{A}}$ is *easy* to compute, then

$$\mathcal{R}(X) = \sup_{\theta \in \mathfrak{A}} \inf_{t \in \mathbb{R}} \{\mathbb{E}[\theta X] + t(1 - \mathbb{E}[\theta])\} = \inf_{t \in \mathbb{R}} \left\{ t + \sup_{\theta \in \mathfrak{A}} \mathbb{E}[\theta(X - t)] \right\} = \inf_{t \in \mathbb{R}} \{t + \Psi(X - t)\}.$$

**Note:** The second equality holds under appropriate *regularity* conditions such as

$$\exists\, \delta > 0 \quad \text{such that} \quad (-\delta, \delta) \subseteq \{\mathbb{E}[\theta] - 1 \,|\, \theta \in \mathfrak{A}\}.$$

# Numerical Examples

Problem Descriptions

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

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

where $u = S(z)$ solves

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

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

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

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

Problem Descriptions

**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

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

**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** | $\mathcal{R}(X) = \mathbb{E}[X] + c\mathbb{E}[\max\{0, X - t\}]$ |
| **Conditional Value-at-Risk** | $\mathcal{R}(X) = \lambda\mathbb{E}[X] + (1-\lambda)\mathrm{CVaR}_\beta(X)$ |
| **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)

| | | PD Algorithm | | | | Bundle | |
|---|---|---|---|---|---|---|---|
| example | risk | iter | nfval | ngrad | subiter | iter | 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 PDE-optimization is **expensive**
- Most **coherent risk measures** are **not** continuously differentiable
- Use **infimal convolution** to **smooth** risk measures
- Appropriate assumptions ensure smoothed risk **is** continuously differentiable
- Generlized method of multipliers to solve 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. Submitted. 2017.
- 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.