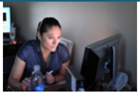
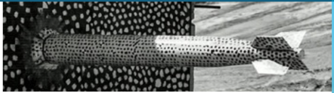
-4393C

National Laboratories

An Augmented Lagrangian Approach for Risk-Averse PDE-Constrained Optimization with State Constraints





Drew P. Kouri

Harbir Antil, Denis Ridzal, Thomas M. Surowiec

SIAM Conference on Uncertainty Quantification

Atlanta, GA

U.S. DEPARTMENT OF ENERGY

NNSA

SAND2022-#### PE

Problem Statement

Goal: Develop efficient algorithms to solve the **risk-averse optimization problem**,

$$\min_{x\in X} \mathcal{R}[f(x)] \quad \text{subject to} \quad g(x)=0, \quad Tx \in C := C_1 \cap \ldots \cap C_m.$$

- X and Y are Banach spaces and Z is a Hilbert space;
- $T \in L(X,Z)$ with injective T^* and $C_i \subset Z$ is closed, convex and boundedly regular;
- $f : X \to L^2(\Omega,\mathcal{F},\mathbb{P})$ and $g : X \to Y$ are continuously differentiable;
- $\mathcal{R} : L^2(\Omega,\mathcal{F},\mathbb{P}) \to \mathbb{R}$ is convex, monotonic and positively homogeneous.

Problem Statement

Goal: Develop efficient algorithms to solve the **risk-averse optimization problem**,

$$\min_{x\in X} \mathcal{R}[f(x)] \quad \text{subject to} \quad g(x) = 0, \quad Tx \in C := C_1 \cap \ldots \cap C_m.$$

- X and Y are Banach spaces and Z is a Hilbert space;
- $T \in L(X, Z)$ with injective T^* and $C_i \subset Z$ is closed, convex and boundedly regular;
- $f : X \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $g : X \to Y$ are continuously differentiable;
- $\mathcal{R} : L^2(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is convex, monotonic and positively homogeneous.

Consequence: $\mathcal{R}$ is **continuous**, **subdifferentiable** and

$$\mathcal{R}[F] = \sup_{\theta\in\mathfrak{A}} \mathbb{E}[\theta F] \quad \text{where} \quad \mathfrak{A} := \partial\mathcal{R}[0] \subseteq \{\theta \in L^2(\Omega, \mathcal{F}, \mathbb{P}) \,|\, \theta \geq 0 \text{ a.s.}\}$$

$$\implies \quad \min_{x\in X} \mathcal{R}[f(x)] = \min_{x\in X} \sup_{\theta\in\mathfrak{A}} \{\ell(x, \theta) := \mathbb{E}[\theta f(x)]\}.$$

Motivation

PDE-constrained optimization (optimal control):

$$\min_{u_\xi,\, z} \mathcal{R}[f(u_\xi, z, \xi)] \qquad \text{subject to} \qquad g(u_\xi, z, \xi) = 0, \quad T_1 u_\xi \in C_1, \quad T_2 z \in C_2.$$

Motivation

PDE-constrained optimization (optimal control):

$$\min_{u_\xi,\, z} \mathcal{R}[f(u_\xi, z, \xi)] \qquad \text{subject to} \qquad g(u_\xi, z, \xi) = 0, \quad T_1 u_\xi \in C_1, \quad T_2 z \in C_2.$$

- Given a control z, the PDE $g(u_\xi, z, \xi) = 0$ is expensive to solve for the state $u_\xi = u_\xi(z)$.

  Solve PDE constraint gradually using, e.g., trust-region SQP (no nonlinear solves).

Motivation

PDE-constrained optimization (optimal control):

$$\min_{u_\xi,\, z} \mathcal{R}[f(u_\xi, z, \xi)] \qquad \text{subject to} \qquad g(u_\xi, z, \xi) = 0, \quad T_1 u_\xi \in C_1, \quad T_2 z \in C_2.$$

- Given a control z, the PDE $g(u_\xi, z, \xi) = 0$ is expensive to solve for the state $u_\xi = u_\xi(z)$.
  Solve PDE constraint gradually using, e.g., trust-region SQP (no nonlinear solves).
- Linear (KKT-like) systems in SQP take advantage of iterative solvers and good preconditioners.
  Use matrix-free SQP to exploit inexact linear system solves; also mesh adaptivity, etc.[†,††]

[†] **Heinkenschloss, Ridzal (2014)**, *A matrix-free trust-region SQP method for equality constrained optimization*, SIOPT.
[††] **Ziems, Ulbrich (2011)**, *Adaptive multilevel inexact SQP methods for PDE-constrained optimization*, SIOPT.

Motivation

PDE-constrained optimization (optimal control):

$$\min_{u_\xi,\, z} \mathcal{R}[f(u_\xi, z, \xi)] \qquad \text{subject to} \qquad g(u_\xi, z, \xi) = 0, \quad T_1 u_\xi \in C_1, \quad T_2 z \in C_2.$$

- Given a control z, the PDE $g(u_\xi, z, \xi) = 0$ is expensive to solve for the state $u_\xi = u_\xi(z)$.
  Solve PDE constraint gradually using, e.g., trust-region SQP (no nonlinear solves).
- Linear (KKT-like) systems in SQP take advantage of iterative solvers and good preconditioners.
  Use matrix-free SQP to exploit inexact linear system solves; also mesh adaptivity, etc.[†,††]
- **Catch:** SQP with inexact linear system solves cannot directly handle general inequality constraints or nonsmooth objective functions.
  Penalize $Tx \in C$ explicitly and smooth $\mathcal{R}$ using augmented Lagrangian.

[†] **Heinkenschloss, Ridzal (2014)**, *A matrix-free trust-region SQP method for equality constrained optimization*, SIOPT.
[††] **Ziems, Ulbrich (2011)**, *Adaptive multilevel inexact SQP methods for PDE-constrained optimization*, SIOPT.

Motivation

PDE-constrained optimization (optimal control):

$$\min_{u_\xi,\, z} \mathcal{R}[f(u_\xi, z, \xi)] \qquad \text{subject to} \qquad g(u_\xi, z, \xi) = 0, \quad T_1 u_\xi \in C_1, \quad T_2 z \in C_2.$$

- Given a control z, the PDE $g(u_\xi, z, \xi) = 0$ is expensive to solve for the state $u_\xi = u_\xi(z)$.

  Solve PDE constraint gradually using, e.g., trust-region SQP (no nonlinear solves).
- Linear (KKT-like) systems in SQP take advantage of iterative solvers and good preconditioners.

  Use matrix-free SQP to exploit inexact linear system solves; also mesh adaptivity, etc.[†,††]
- **Catch:** SQP with inexact linear system solves cannot directly handle general inequality constraints or nonsmooth objective functions.

  Penalize $Tx \in C$ explicitly and smooth $\mathcal{R}$ using augmented Lagrangian.
- Control and state multipliers have different regularity, e.g., L^2 for controls and measures for states, resulting in vastly different scales, which can lead to strong mesh dependence for NLP methods.

  Use separate penalties and multiplier estimates for control and state constraints.

[†] **Heinkenschloss, Ridzal (2014)**, *A matrix-free trust-region SQP method for equality constrained optimization*, SIOPT.
[††] **Ziems, Ulbrich (2011)**, *Adaptive multilevel inexact SQP methods for PDE-constrained optimization*, SIOPT.

PDE-Constrained Optimization is Expensive

Full Space

$$\min_{\substack{T_1 u_n \in C_1 \\ T_2 z \in C_2}} \sum_{n=1}^{N} w_n f(u_n, z, \xi_n) \quad \text{subject to} \quad g(u_n, z, \xi_n) = 0$$

- **Numerical solution is severely limited due to memory!**
- PDE solution variables are treated as optimization variables.
- Must store each PDE solution variable u_n.
- Often need to store one Lagrange multiplier per ξ_n.

Reduced Space

$$\min_{\substack{T_1 S_n(z) \in C_1 \\ T_2 z \in C_2}} \sum_{n=1}^{N} w_n f(S_n(z), z, \xi_n) \quad \text{where} \quad g(S_n(z), z, \xi_n) = 0$$

- **Numerical solution is severely limited due to computation!**
- Objective evaluation requires the solution to $g(u_\xi, z, \xi_n) = 0$ for each ξ_n.
- Gradient evaluation requires an additional linearized solve per ξ_n.
- Hessian-times-a-vector requires two additional linearized solves per ξ_n.

PDE-Constrained Optimization is Expensive

Full Space

$$\min_{\substack{T_1 u_n \in C_1 \\ T_2 z \in C_2}} \sum_{n=1}^{N} w_n f(u_n, z, \xi_n)$$

$$\text{subject to} \quad g(u_n, z, \xi_n) = 0$$

- Numerical solution is severely limited due to memory!
- PDE solution variables are optimization variables.
- Must store each PDE solution variable u_n.
- Often need to store one Lagrange multiplier per ξ_n.

Requires $\mathcal{O}(MN)$ storage!

N is typically $\mathcal{O}(10^3)$

M can be $\geq 10^9$

Reduced Space

$$\min_{\substack{T_1 S_n(z) \in C_1 \\ T_2 z \in C_2}} \sum_{n=1}^{N} w_n f(S_n(z), z, \xi_n)$$

$$\text{where} \quad g(S_n(z), z, \xi_n) = 0$$

- **Numerical solution is severely limited due to computation!**
- Objective evaluation requires the solution to $g(u_\xi, z, \xi_n) = 0$ for each ξ_n.
- Gradient evaluation requires an additional linearized solve per ξ_n.
- Hessian-times-a-vector requires two additional linearized solves per ξ_n.

PDE-Constrained Optimization is Expensive

Full Space

$$\min_{\substack{T_1 u_n \in C_1 \\ T_2 z \in C_2}} \sum_{n=1}^{N} w_n \, f(u_n, z, \xi_n)$$

$$\text{subject to} \quad g(u_n, z, \xi_n) = 0$$

Requires $\mathcal{O}(MN)$ storage!

N is typically $\mathcal{O}(10^3)$

M can be $\geq 10^9$

Reduced Space

$$\min_{\substack{T_1 S_n(z) \in C_1 \\ T_2 z \in C_2}} \sum_{n=1}^{N} w_n \, f(S_n(z), z, \xi_n)$$

$$\text{where} \quad g(S_n(z), z, \xi_n) = 0$$

$\mathcal{O}(N)$ nonlinear solves!

Additional $\mathcal{O}(N)$ linear solves for derivative computations

Modeling Risk Preference

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

... In our optimization problem, $f(u_\xi, z, \xi)$ is a **risk**!

We **cannot** directly minimize $f(u_\xi, z, \xi) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$

... How should we **quantify the risk**?

Optimistic Formulations

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

$$\mathcal{R}[X] = \mathbb{E}[X].$$

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

$$\mathcal{R}[X] = \mathbb{P}(X > x).$$

Conservative Formulations

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

$$\mathcal{R}[X] = \mathbb{E}[X] + \mathcal{D}[X].$$

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

$$\mathcal{R}[X] = \text{bPOE}_x(X).$$

Modeling Risk Preference

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

... In our optimization problem, $f(u_\xi, z, \xi)$ is a **risk**!

We **cannot** directly minimize $f(u_\xi, z, \xi) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$

... How should we **quantify the risk**?

Optimistic Formulations

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

$$\mathcal{R}[X] = \mathbb{E}[X].$$

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

$$\mathcal{R}[X] = \mathbb{P}(X > x).$$

Conservative Formulations

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

$$\mathcal{R}[X] = \mathbb{E}[X] + \mathcal{D}[X].$$

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

$$\mathcal{R}[X] = \mathrm{bPOE}_x(X).$$

Coherent Measures of Risk

$\mathcal{R} : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to (-\infty, \infty]$ is a **coherent** measure of risk if it satisfies

(R1) Convexity: $\mathcal{R}[tX + (1-t)X'] \le t\mathcal{R}[X] + (1-t)\mathcal{R}[X'], \quad \forall t \in [0,1]$

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

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)!**

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

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

Artzner, Delbaen, Eber, Heath (1999), *Coherent measures of risk*, Math Finance.

Coherent Measures of Risk

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

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

$$\iff \quad \mathcal{R}[X] = \sup\{\mathbb{E}[\vartheta X] - \mathcal{R}^*[\vartheta] \mid \vartheta \in \mathrm{dom}(\mathcal{R}^*)\}$$

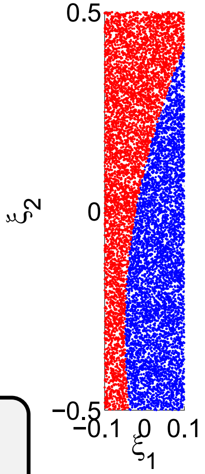
- If $\mathcal{R}$ is **translation equivariant** and **monotonic**

$$\iff \quad \mathrm{dom}(\mathcal{R}^*) \subseteq \{\vartheta \in \mathcal{X}^* \mid \mathbb{E}[\vartheta] = 1,\ \vartheta \ge 0 \text{ a.s.}\}$$

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

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

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



Differentiability: The **coherent** risk measure $\mathcal{R}$ is **Fréchet differentiable**

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

Risk-Averse Augmented Lagrangian

Motivated by the *Primal-Dual Risk Minimization*[†] and *ALESQP*[††] algorithms, we define

$$L(x,\lambda,r) := \max_{\substack{\mu_0\in\mathfrak{A}\\ \mu_i\in Z}} \left\{ \mathbb{E}[\mu_0 f(x)] - \frac{1}{2r_0}\mathbb{E}[(\lambda_0-\mu_0)^2] + \sum_{i=1}^m (\mu_i, Tx)_Z - I^*_{C_i}(\mu_i) - \frac{1}{2r_i}\|\lambda_i-\mu_i\|_Z^2 \right\}$$

$$= \widehat{\mathcal{R}}(f(x),\lambda_0,r_0) + \sum_{i=1}^m \frac{1}{2r_i}\|\Lambda_i(x,\lambda_i,r_i)\|_Z^2 - \frac{1}{2r_i}\|\lambda_i\|_Z^2$$

where $\Lambda_i(x,\lambda,r) := r((r^{-1}\lambda + Tx) - \mathbf{P}_{C_i}(r^{-1}\lambda+Tx))$.

[†] **Kouri, Surowiec (2021)**, *A primal-dual algorithm for risk minimization*, Math Prog.
[††] **Antil, Kouri, Ridzal (2021)**, *ALESQP: An augmented Lagrangian equality-constrained SQP method for optimization with general constraints*, Submitted.

Risk-Averse Augmented Lagrangian

Motivated by the *Primal-Dual Risk Minimization*† and *ALESQP*†† algorithms, we define

$$L(x,\lambda,r) := \max_{\substack{\mu_0\in\mathfrak{A}\\ \mu_i\in Z}} \left\{ \mathbb{E}[\mu_0 f(x)] - \frac{1}{2r_0}\mathbb{E}[(\lambda_0-\mu_0)^2] + \sum_{i=1}^m (\mu_i, Tx)_Z - I_{C_i}^*(\mu_i) - \frac{1}{2r_i}\|\lambda_i-\mu_i\|_Z^2 \right\}$$

$$= \widehat{\mathcal{R}}(f(x),\lambda_0,r_0) + \sum_{i=1}^m \frac{1}{2r_i}\|\Lambda_i(x,\lambda_i,r_i)\|_Z^2 - \frac{1}{2r_i}\|\lambda_i\|_Z^2$$

where $\Lambda_i(x,\lambda,r) := r((r^{-1}\lambda + Tx) - \mathbf{P}_{C_i}(r^{-1}\lambda+Tx))$.

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

$$\widehat{\mathcal{R}}(f(x),\lambda,r) = \min_{F\in L^2(\Omega,\mathcal{F},\mathbb{P})} \left\{ \mathcal{R}[f(x)-F] + \mathbb{E}[\lambda F] + \frac{r}{2}\mathbb{E}[F^2] \right\} = \mathcal{R}^{\Phi}_{1/r}[f(x)]$$

where $\Phi(F) = \mathbb{E}[\lambda F] + \frac{1}{2}\mathbb{E}[F^2] \quad \Longrightarrow \quad 0 \le \mathcal{R}[F] - \mathcal{R}^{\Phi}_{1/r}[F] \le K^2/r$.

† **Kouri, Surowiec (2021)**, *A primal-dual algorithm for risk minimization*, Math Prog.
†† **Antil, Kouri, Ridzal (2021)**, *ALESQP: An augmented Lagrangian equality-constrained SQP method for optimization with general constraints*, Submitted.

Risk-Averse Augmented Lagrangian

Motivated by the *Primal-Dual Risk Minimization*† and *ALESQP*†† algorithms, we define

$$L(x,\lambda,r) := \max_{\substack{\mu_0\in\mathfrak{A}\\ \mu_i\in Z}} \left\{ \mathbb{E}[\mu_0 f(x)] - \frac{1}{2r_0}\mathbb{E}[(\lambda_0-\mu_0)^2] + \sum_{i=1}^m (\mu_i, Tx)_Z - I^*_{C_i}(\mu_i) - \frac{1}{2r_i}\|\lambda_i-\mu_i\|_Z^2 \right\}$$

$$= \widehat{\mathcal{R}}(f(x),\lambda_0,r_0) + \sum_{i=1}^m \frac{1}{2r_i}\|\Lambda_i(x,\lambda_i,r_i)\|_Z^2 - \frac{1}{2r_i}\|\lambda_i\|_Z^2$$

where $\Lambda_i(x,\lambda,r) := r((r^{-1}\lambda + Tx) - \mathbf{P}_{C_i}(r^{-1}\lambda+Tx))$.

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

$$\widehat{\mathcal{R}}(f(x),\lambda,r) = \min_{F\in L^2(\Omega,\mathcal{F},\mathbb{P})} \left\{ \mathcal{R}[f(x)-F] + \mathbb{E}[\lambda F] + \frac{r}{2}\mathbb{E}[F^2] \right\} = \mathcal{R}^{\Phi}_{1/r}[f(x)]$$

where $\Phi(F) = \mathbb{E}[\lambda F] + \frac{1}{2}\mathbb{E}[F^2] \quad \Longrightarrow \quad 0 \le \mathcal{R}[F] - \mathcal{R}^{\Phi}_{1/r}[F] \le K^2/r$.

Differentiability: $x \mapsto L(x,\lambda,r)$ is Fréchet differentiable with **Lipschitz continuous gradient**.

† **Kouri, Surowiec (2021)**, *A primal-dual algorithm for risk minimization*, Math Prog.
†† **Antil, Kouri, Ridzal (2021)**, *ALESQP: An augmented Lagrangian equality-constrained SQP method for optimization with general constraints*, Submitted.

Risk-Averse Augmented Lagrangian

Require: $\lambda_0^{(0)} \in \mathfrak{A}$, $\lambda_i^{(0)} \in Z$, $r_0^{(0)} > 0$, $r_i^{(0)} > 0$, $\nu_i > 0$, and $\gamma_i \in (0, \frac{1}{2})$

1: **while** "Not Converged" **do**

2: Find $x^{(k)} \in X$ that *approximately* solves

$$\min_{x \in X} \; L(x, \lambda^{(k)}, r^{(k)}) \quad \text{subject to} \quad g(x) = 0$$

3: Update penalty parameters $r_0^{(k+1)}$ and $r_i^{(k+1)}$

4: Update risk identifier estimates

$$\lambda_0^{(k+1)} = \begin{cases} \mathbf{P}_{\mathfrak{A}}(r_0^{(k)} f(x^{(k)}) + \lambda_0^{(k)}) & \text{if } \|\mathbf{P}_{\mathfrak{A}}(r_0^{(k)} f(x^{(k)}) + \lambda_0^{(k)}) - \lambda_0^{(k)}\|_2 \le r_0^{(k)} \tau_0^{(k)} \\ \lambda_0^{(k)} & \text{otherwise} \end{cases}$$

5: Update Lagrange multiplier estimates

$$\lambda_i^{(k+1)} = \begin{cases} \Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) & \text{if } \|\Lambda_i(x^{(k)}, \lambda_i^{(k)}, r_i^{(k)}) - \lambda_i^{(k)}\| \le \nu_i (r_i^{(k+1)})^{\gamma_i} \\ \lambda_i^{(k)} & \text{otherwise} \end{cases}$$

6: **end while**

Risk-Averse Augmented Lagrangian

Use **composite-step SQP** to produce $x^{(k)}$ and Lagrange multiplier $\zeta^{(k)}$ that satisfy

$$\|L_x'(x^{(k)}, \lambda^{(k)}, r^{(k)}) + g'(x^{(k)})^* \zeta^{(k)}\|_{X^*} \le \varepsilon^{(k)} \qquad \text{and} \qquad \|g(x^{(k)})\|_Y \le \delta^{(k)}.$$

- **Composite-step SQP** is matrix-free to handle extreme-scale problems.
- Main computational work is the repeated solution of the **augmented system**

$$\begin{pmatrix} I_{X,X^*} & g'(x_j)^* \\ g'(x_j) & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}.$$

- Handles inexact linear system solves[†], mesh adaptivity[††], etc. to improve efficiency.
- Use linearized PDE solvers to precondition iterative augmented system solves[†††].

[†] **Heinkenschloss, Ridzal (2014)**, *A matrix-free trust-region SQP method for equality constrained optimization*, SIOPT.
[††] **Ziems, Ulbrich (2011)**, *Adaptive multilevel inexact SQP methods for PDE-constrained optimization*, SIOPT.
[†††] **Kouri, Ridzal (2018)**, *Inexact trust-region methods for PDE-constrained optimization*, IMA.

Risk-Averse Augmented Lagrangian

Risk Penalty Parameter:

if $\|\mathbf{P}_{\mathfrak{A}}(f(x^{(k)}),\lambda_0^{(k)},r_0^{(k)}) - \lambda_0^{(k)}\|_2 > r_0^{(k)}\tau_0^{(k)}$
then
$r_0^{(k+1)} = \eta_0 r_0^{(k)}$
$\theta_0^{(k+1)} = \min\{1/r_0^{(k+1)},\theta_0\}$
$\tau_0^{(k+1)} = \tau_0^{(0)}(\theta_0^{(k+1)})^{\alpha_0}$
else
$r_0^{(k+1)} = r_0^{(k)}$
$\theta_0^{(k+1)} = \min\{1/r_0^{(k+1)},\theta_0\}$
$\tau_0^{(k+1)} = \tau_0^{(k)}(\theta_0^{(k+1)})^{\beta_0}$
end if

Constraint Penalty Parameters:

if $\|\Lambda_i(x^{(k)},\lambda_i^{(k)},r_i^{(k)}) - \lambda_i^{(k)}\| > r_i^{(k)}\tau_i^{(k)}$
then
$r_i^{(k+1)} = \eta_i r_i^{(k)}$
$\theta_i^{(k+1)} = \min\{1/r_i^{(k+1)},\theta_i\}$
$\tau_i^{(k+1)} = \tau_i^{(0)}(\theta_i^{(k+1)})^{\alpha_i}$
else
$r_i^{(k+1)} = r_i^{(k)}$
$\theta_i^{(k+1)} = \min\{1/r_i^{(k+1)},\theta_i\}$
$\tau_i^{(k+1)} = \tau_i^{(k)}(\theta_i^{(k+1)})^{\beta_i}$
end if

- These penalty parameter updates are used in LANCELOT†.
- Constraint penalties are updated in unison after L iterations (L large) to ensure feasibility.
- Updates based on infeasibility: $d_{C_i}(Tx) \le \frac{1}{r}\|\Lambda_i(x,\lambda,r) - \lambda\|_Z \le d_{C_i}(Tx) + \frac{1}{r}\|\lambda\|_Z$.

† **Conn, Gould, Toint (1991)**, *A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds*, SINUM.

Numerical Examples

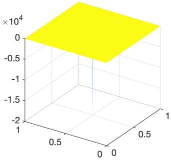
We consider the PDE-constrained optimal control problem

$$\min_{\substack{z\in L^2(D)\\ u_\xi\in H_0^1(D)\cap C_0(D)}} \quad \mathcal{R}\left(\frac{1}{2}\int_D (u_\xi+1)^2\,\mathrm{d}x\right)+\frac{\alpha}{2}\int_D z^2\,\mathrm{d}x$$

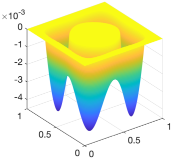
$$\text{subject to} \qquad -10\le z\le 10, \quad u_\xi \ge \psi$$

$$-\Delta u_\xi + u_\xi^3 = f_\xi + z \quad \text{in } D \text{ a.s.}$$

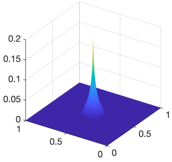
$$u_\xi = 0 \quad \text{on } \partial D \text{ a.s.}$$



ψ multiplier



z_a multiplier



z_b multiplier

where $D=(0,1)^2$, $\alpha=10^{-3}$, $\xi_i\sim\mathcal{N}(0,1)$ for $i=1,\ldots,200$,

$$f_\xi(x)=\sqrt{2}\sum_{i=1}^{100}\frac{\sin((i-\frac{1}{2})\pi x_1)}{(i-\frac{1}{2})\pi}\xi_{2i}+\frac{\sin((i-\frac{1}{2})\pi x_2)}{(i-\frac{1}{2})\pi}\xi_{2i-1}$$

$$\psi(x)=-\tfrac{2}{3}+\tfrac{1}{2}\min\{x_1+x_2,\min\{1+x_1-x_2,\min\{1-x_1+x_2,2-x_1-x_2\}\}\}$$

Risk Measures:

Mean-Plus-Semideviation $\quad \mathcal{R}[F]=\mathbb{E}[F]+c\,\mathbb{E}[\max\{0,F-\mathbb{E}[F]\}]$

Conditional Value-at-Risk $\quad \mathcal{R}[F]=(1-\lambda)\mathbb{E}[F]+\lambda\,\mathrm{CVaR}_\beta(F)$

Numerical Results

100 Monte Carlo Samples, Tolerances: $\varepsilon_* = \delta_* = \tau_* = \sigma_* = 10^{-6}$

	MPSD ($c = 0.8$)						
mesh	AL	SQP	CG	normg	grad-lag	feas	dual-risk
32x32	17	52	113	2.50e-15	3.67e-07	1.75e-07	0.00e+00
64x64	18	65	141	1.68e-15	9.09e-11	5.32e-08	0.00e+00
128x128	20	60	126	5.90e-15	1.64e-11	1.42e-08	0.00e+00

	CVaR ($\beta = 0.8,\ \lambda = 0.75$)						
mesh	AL	SQP	CG	normg	grad-lag	feas	dual-risk
32x32	18	63	158	4.62e-16	4.73e-08	1.70e-08	0.00e+00
64x64	20	65	168	1.49e-15	2.07e-08	3.70e-08	0.00e+00
128x128	20	67	177	5.92e-15	1.23e-07	6.29e-09	0.00e+00

$\texttt{normg} = \|g(x^{(k)})\|_Y \qquad \texttt{grad-lag} = \|L'_x(x^{(k)}, \lambda^{(k)}, r^{(k)}) + g'(x^{(k)})^* \zeta^{(k)}\|_{X^*}$

$\texttt{feas} = \max_i d_{C_i}(Tx^{(k)}) \qquad \texttt{dual-risk} = \mathbb{E}[(\lambda_0^{(k)} - \mathbf{P}_{\mathfrak{A}}(r_0^{(k)} f(x^{(k)}) + \lambda_0^{(k)}))^2]^{1/2} / r_0^{(k)}$

We observe that the AL, SQP, and CG iterations are nearly mesh independent!

Numerical Results

64×64 **Mesh, Tolerances:** $\varepsilon_* = \delta_* = \tau_* = \sigma_* = 10^{-6}$

	MPSD $(c = 0.8)$						
nsamp	AL	SQP	CG	normg	grad-lag	feas	dual-risk
100	18	65	141	1.68e-15	9.10e-11	5.32e-08	0.00e+00
200	17	101	226	6.35e-16	1.26e-10	7.90e-07	0.00e+00
400	18	81	174	5.64e-16	1.45e-09	1.46e-08	3.36e-07

	CVaR $(\beta = 0.8,\ \lambda = 0.75)$						
nsamp	AL	SQP	CG	normg	grad-lag	feas	dual-risk
100	20	65	168	1.49e-15	2.07e-08	3.70e-07	0.00e+00
200	17	91	221	2.43e-15	8.64e-08	8.76e-07	0.00e+00
400	20	96	236	3.34e-16	4.08e-08	5.94e-09	0.00e+00

$\texttt{normg} = \|g(x^{(k)})\|_Y \qquad \texttt{grad-lag} = \|L'_x(x^{(k)}, \lambda^{(k)}, r^{(k)}) + g'(x^{(k)})^* \zeta^{(k)}\|_{X^*}$

$\texttt{feas} = \max_i d_{C_i}(Tx^{(k)}) \qquad \texttt{dual-risk} = \mathbb{E}[(\lambda_0^{(k)} - \mathbf{P}_{\mathfrak{A}}(r_0^{(k)} f(x^{(k)}) + \lambda_0^{(k)}))^2]^{1/2} / r_0^{(k)}$

We observe that the AL, SQP, and CG iterations are nearly sample-size independent!

Conclusions:

- **Numerical solution** of stochastic PDE-constrained optimization is **expensive**.
- Numerical solution is complicated by **nonsmooth risk measures** and **state constraints**.
- Augmented Lagrangian **penalizes** the state/control constraints and **smooths** the risk measures.
- PDE is **solved gradually** using trust-region SQP, avoiding complications with nonlinear solvers, etc.
- Numerical examples suggest nearly **mesh/sample-size independent** performance for **nonsmooth state-constrained** problems!

References:

- **Antil, Kouri, Ridzal (2021)**, *ALESQP: An augmented Lagrangian equality-constrained SQP method for optimization with general constraints*, Submitted.
- **Kouri, Surowiec (2021)**, *A primal-dual algorithm for risk minimization*, Math Prog.
- **Kouri, Surowiec (2020)**, *Epi-regularization of risk measures*, MOR.
- **Kouri, Shapiro (2018)**, *Optimization of PDEs with uncertain inputs*, IMA.
- **Kouri, Ridzal (2018)**, *Inexact trust-region methods for PDE-constrained optimization*, IMA.
- **Kouri, Surowiec (2016)**, *Risk-averse PDE-constrained optimization using the conditional value-at-risk*, SIOPT.
- **Heinkenschloss, Ridzal (2014)**, *A matrix-free trust-region SQP method for equality constrained optimization*, SIOPT.