Primal-dual proximal splitting for risk-averse optimal control

Sebastian Angerhausen   Christian Clason

Department of Mathematics and Scientific Computing, University of Graz

Workshop “PDE Constrained Optimization”
Essen, July 29, 2025

Motivation: robust optimization

optimal control problem

$$\begin{cases} \min\limits_{u \in U_{\text{ad}}} \frac{1}{2}\|y - y^d\|^2 + \frac{\gamma}{2}\|u\|^2 \\ -\operatorname{div}(a \operatorname{grad} y) = u + f & \text{in } \Omega \\ y = g & \text{in } \partial\Omega \end{cases}$$

Motivation: robust optimization

random optimal control problem

$$\begin{cases} \min\limits_{u \in U_{\mathrm{ad}}} \dfrac{1}{2}\|y(\xi) - y^d(\xi)\|^2 + \dfrac{\gamma}{2}\|u\|^2 \\ \qquad -\operatorname{div}(a(\xi)\,\operatorname{grad} y) = u + f(\xi) \quad \text{in } \Omega \\ \qquad\qquad\qquad\qquad y = g(\xi) \quad \text{in } \partial\Omega \end{cases}$$

- uncertain data, target $\rightsquigarrow$ parametrized optimal control problem

Motivation: robust optimization

random optimal control problem

$$
\begin{cases}
\min_{u \in U_{\text{ad}}} \mathbb{E}\left[\frac{1}{2}\|y(\xi) - y^d(\xi)\|^2\right] + \frac{\gamma}{2}\|u\|^2 \\
-\operatorname{div}(a(\xi) \operatorname{grad} y) = u + f(\xi) \quad \text{in } \Omega \\
y = g(\xi) \quad \text{in } \partial\Omega
\end{cases}
$$

- uncertain data, target $\rightsquigarrow$ parametrized optimal control problem
- objective value random $\rightsquigarrow$ minimize expectation

Motivation: robust optimization

risk-averse optimal control problem

$$\begin{cases} \min_{u\in U_{\mathrm{ad}}} \mathcal{R}\left[\frac{1}{2}\|y(\xi)-y^d(\xi)\|^2\right]+\frac{\gamma}{2}\|u\|^2 \\ \qquad -\operatorname{div}(a(\xi)\operatorname{grad} y)=u+f(\xi) \quad \text{in } \Omega \\ \qquad y=g(\xi) \quad \text{in } \partial\Omega \end{cases}$$

- uncertain data, target $\rightsquigarrow$ parametrized optimal control problem
- objective value random $\rightsquigarrow$ minimize expectation
- expectation undervalues expensive but rare events
- $\rightsquigarrow$ replace expectation with risk-averse measure

Motivation: risk measure

risk-averse control problem

$$\min_{u \in U_{\text{ad}}} \mathcal{R}[J(u,\xi)]$$

Goal:

- minimize $J(u,\xi)$ for almost all ξ
- w.l.o.g. $J(u,\xi) \leq 0$ for almost all ξ (almost sure constraints)
- infeasible $\rightsquigarrow \mathbb{P}[J(u,\xi) > 0] \leq \alpha$ (probabilistic constraints)
- nonconvex $\rightsquigarrow$ find convex dominating risk measure

Motivation: risk measure

Goal: convex majorant for $\mathbb{P}[Z > 0] \le \alpha$

Idea:

$$\mathrm{VaR}_\alpha[Z] := \inf\{t > 0 : \mathbb{P}[Z \le t] \ge 1 - \alpha\} = \inf\{t > 0 : \mathbb{P}[Z > t] \le \alpha\}$$

- Value-at-Risk: largest objective that can occur with probability $1 - \alpha$
- $\rightsquigarrow \mathbb{P}[Z > 0] \le \alpha$ iff $\mathrm{VaR}_\alpha[Z] \le 0$
- $\mathbb{P}[Z > 0] = \mathbb{E}[\mathbb{1}_{(0,\infty)}(Z)]$ (expectation of characteristic function)
- still not convex...
- $\rightsquigarrow$ replace characteristic function by piecewise linear function

Motivation: risk measure

Goal: convex majorant for $\mathbb{P}[Z > 0] \le \alpha$

For any $\gamma > 0$,

$$\mathbb{P}[Z > 0] = \mathbb{E}\left[\mathbb{1}_{(0,\infty)}(Z)\right] \le \mathbb{E}\left[\max\{0, 1 + \gamma Z\}\right] = \gamma \mathbb{E}\left[\max\{0, \gamma^{-1} + Z\}\right]$$

Then:

$$\inf_{\gamma > 0} \gamma \mathbb{E}\left[\max\{0, \gamma^{-1} + Z\}\right] - \alpha \le 0$$

implies $\mathrm{VaR}_\alpha[Z] \le 0$ and $\mathbb{P}[Z > 0] \le \alpha$ (inf not larger than 0)

Equivalent:

$$\inf_{t < 0} t + \alpha^{-1} \mathbb{E}\left[\max\{0, Z - t\}\right] \le 0$$

Note: $t > 0$ implies $t + \alpha^{-1} \mathbb{E}\left[\max\{0, Z - t\}\right] \ge t > 0$

Motivation: risk measure

Goal: convex majorant for $\mathbb{P}[Z > 0] \leq \alpha$

Conditional Value-at-Risk

$$\mathrm{CVaR}_\alpha[Z] := \inf_{t\in\mathbb{R}} t + \alpha^{-1}\mathbb{E}\left[\max\{0, Z - t\}\right]$$

- for $\alpha \to 0$, $\mathrm{CVaR}_\alpha[Z] \leq 0 \to Z \preceq 0$ almost surely (formally)
- $\rightsquigarrow$ risk-averse control problem

$$\min_{u\in U_{\mathrm{ad}}} \mathrm{CVaR}_\alpha\left[J(u, \xi)\right]$$

Goal: efficient numerical methods

Convex risk measures

Notation:

- probability space $(\Xi, \mathcal{A}, \mathbb{P})$
- random variable $Z \in L^p(\Xi)$, $1 \leq p < \infty$ (w.r.t. measure $\mathbb{P}$)

$\mathcal{R} : L^p(\Xi) \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ convex risk measure if

1. convex
2. monotone: $\mathcal{R}(Z) \geq \mathcal{R}(Z')$ if $Z \succeq Z'$
3. translation invariant: $\mathcal{R}(Z + a) = \mathcal{R}(Z) + a$ for $a \in \mathbb{R}$ and $Z \in L^p(\Xi)$

$\mathcal{R} : L^p(\Xi) \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ coherent risk measure if also

4. positively homogeneous: $\mathcal{R}(tZ) = t\mathcal{R}(Z)$ for $t > 0$ and $Z \in L^p(\Xi)$

$\Rightarrow \mathcal{R}(0) = 0$ and $\mathcal{R}(a) = a$

Convex risk measures: dual characterization

- $\mathcal{R} : L^p(\Xi) \to \overline{\mathbb{R}}$ proper, convex, lsc
- $\mathcal{R}^* : L^q(\Xi) \to \overline{\mathbb{R}}$ Fenchel conjugate
- $\Pi := \operatorname{dom} \mathcal{R}^*$ risk envelope

Then:

- monotone iff $Z^* \succeq 0$ for all $Z^* \in \Pi$
- translation invariant iff $\mathbb{E}[Z^*] = 1$ for all $Z^* \in \Pi$
- positively homogeneous iff

$$\mathcal{R}(Z) = \sup_{Z^* \in \operatorname{dom} \mathcal{R}^*} \langle Z^*, Z \rangle_{L^p(\Xi)} = \sup_{Z^* \in \operatorname{dom} \mathcal{R}^*} \mathbb{E}[Z^* Z]$$

$\rightsquigarrow$ $\mathcal{R}$ coherent implies

- $\Pi \subseteq \Delta_q := \{Z^* \in L^q(\Xi) : \mathbb{E}[Z^*] = 1, \quad Z^* \succeq 0\}$
- $\mathcal{R}^* = \delta_\Pi$

Convex risk measures: examples

1. $\mathcal{R}[Z] = \mathbb{E}[Z]$ is coherent with risk envelope

$$\Pi := \{Z^* \in L^\infty(\Xi) : Z^*(\xi) = 1 \text{ almost surely}\}$$

2. $\mathcal{R}[Z] = \operatorname{ess\,sup}_{\xi \in \Xi} Z(\xi)$ is coherent with risk envelope

$$\Pi := \{Z^* \in L^\infty(\Xi) : Z^* \succeq 0, \mathbb{E}[Z^*] = 1\}$$

3. $\mathcal{R}[Z] = \mathrm{CVaR}_\alpha[Z]$ is coherent with risk envelope

$$\Pi := \left\{Z^* \in L^\infty(\Xi) : 0 \preceq Z^* \preceq \alpha^{-1}, \mathbb{E}[Z^*] = 1\right\}$$

(truncated Gibbs simplex)

Convex risk measures: optimality conditions

- $\mathcal{R} : L^p(\Xi) \to \overline{\mathbb{R}}$ coherent risk measure

$$\partial\mathcal{R}(Z) = \arg\max_{Z^* \in \Pi} \langle Z^*, Z \rangle_{L^p(\Xi)}$$

$\rightsquigarrow$ subdifferential weighted expectation

- $J(u, \xi)$ continuously differentiable for almost every $\xi \in \Xi$

Solution $\bar{u} \in U_{\text{ad}}$ to $\min_{u \in U_{\text{ad}}} \mathcal{R}[J(u, \xi)]$ satisfies

$$\begin{cases} \bar{Z}^* \in \partial\mathcal{R}[J(\bar{u}, \xi)] \\ \bar{p} = J'(\bar{u}; \cdot) \quad \text{almost surely,} \\ 0 \in \mathbb{E}[\bar{p}\bar{Z}^*] + N_{U_{\text{ad}}}(\bar{u}) \end{cases}$$

Convex risk measures: optimality conditions

- $\mathcal{R} : L^p(\Xi) \to \overline{\mathbb{R}}$ coherent risk measure
- $J(u,\xi)$ continuously differentiable for almost every $\xi \in \Xi$

Solution $\bar u \in U_{\text{ad}}$ to $\min_{u\in U_{\text{ad}}} \mathcal{R}[J(u,\xi)]$ satisfies

$$\begin{cases} \bar Z^* \in \partial\mathcal{R}[J(\bar u,\xi)] \\ \bar p = J'(\bar u;\cdot) \quad \text{almost surely,} \\ 0 \in \mathbb{E}[\bar p \bar Z^*] + N_{U_{\text{ad}}}(\bar u) \end{cases}$$

$(\bar u, \bar Z^*)$ is saddle point of

$$\min_{u\in U_{\text{ad}}} \sup_{Z^*\in\Pi} \mathbb{E}\left[J(u,\xi)Z^*(\xi)\right]$$

Primal-dual proximal splitting

Saddle point problem

$$\min_{u \in U_{\mathrm{ad}}} \sup_{Z \in \Pi} \mathbb{E}\left[J(u,\xi)Z(\xi)\right]$$

Algorithm

$$u^{k+1} = \mathrm{proj}_{U_{\mathrm{ad}}} u^k - \tau_k J'(u^k)^* Z^k$$
$$Z^{k+1} = \mathrm{proj}_{\Pi}(Z^k + \sigma_k J(u^{k+1}))$$

- $\sigma_k, \tau_k > 0$ step sizes
- alternating direction minimization (not always convergent)

Primal-dual proximal splitting

Saddle point problem

$$\min_{u \in U_{\mathrm{ad}}} \sup_{Z \in \Pi} \mathbb{E}\left[J(u,\xi) Z(\xi)\right]$$

Algorithm

$$\begin{aligned}
u^{k+1} &= \operatorname{proj}_{U_{\mathrm{ad}}} u^k - \tau_k J'(u^k)^* Z^k \\
\hat{u}^{k+1} &= 2u^{k+1} - u^k \\
Z^{k+1} &= \operatorname{proj}_{\Pi}(Z^k + \sigma_k J(\hat{u}^{k+1}))
\end{aligned}$$

- $\sigma_k, \tau_k > 0$ step sizes
- primal-dual proximal splitting (proximal point method for (u, Z))
- problem: $J(u^k), J'(u^k) \in L^p(\Xi)$ (scenarios if Ξ finite)

Primal-dual proximal splitting: stochastic optimization

Goal: efficient algorithm for stochastic optimal control

Idea 1: dual-primal proximal splitting

$$
\begin{aligned}
Z^{k+1} &= \mathrm{proj}_{\Pi}(Z^k + \sigma_k J(u^k)) \\
\hat{Z}^{k+1} &= 2Z^{k+1} - Z^k, \\
u^{k+1} &= \mathrm{proj}_{U_{\mathrm{ad}}} u^k - \tau_k J'(u^k)^* \hat{Z}^{k+1}
\end{aligned}
$$

$\rightsquigarrow$ reuse state equation solves for $J'(u^k, \xi)$

Idea 2: mini-batch updates only for subset $A_k \in \mathcal{A}$

$$
\begin{aligned}
J(u) &\rightsquigarrow J_k(u) := \mathbb{1}_{A_k} J(u) + (1 - \mathbb{1}_{A_k}) J_{k-1}(u^{k-1}) \\
J'(u)^* &\rightsquigarrow J_k'(u)^* := J'(u)^* \circ \mathbb{1}_{A_k} + J_{k-1}'(u^{k-1})^* \circ (1 - \mathbb{1}_{A_k})
\end{aligned}
$$

Primal-dual proximal splitting: convergence

Analysis based on [Clason, Valkonen '20], Chap. 8.1, [Combettes, Pesquet '15]

Consider iteration for $w^k = (u^k, Z^k)$ in implicit form

$$0 \in W_k H(w^{k+1}) + D_k(w^k + 1) + M_k(w^{k+1} - w^k)$$

- W_k step size operator (diagonal)
- H optimality conditions $0 \in H(\bar{w})$
- D_k discrepancy operator: linearization error, mini-batch
- M_k local preconditioner (decouples proximal point method)

Challenge: mini-batches A_k random $\rightsquigarrow$ inclusion a.s., filtration

Primal-dual proximal splitting: convergence

Central tool: abstract convergence result (formal)

Theorem

Assume

- $\mathcal{Z}$ *nonempty*
- $\{R_k\}_{k\in\mathbb{N}}$ *uniformly bounded in operator norm*
- $\{w_k\}_{k\in\mathbb{N}}$ *stochastic quasi-Fejér monotone: for all* $z \in \mathcal{Z}$,

$$\frac{1}{2}\mathbb{E}\left[\|w_{k+1} - z\|^2_{R_{k+1}}\right] + \Delta_k(z) \leq \frac{1}{2}\|w_k - z\|^2_{R_k} + \lambda_k(z) \quad a.s.$$

- *weak accumulation points of* $\{w_k\}_{k\in\mathbb{N}}$ *belong to* $\mathcal{Z}$,
- $R_{n_k}w \to Rw$ *for all* w, *convergent subsequences* $\{w^{n_k}\}_{k\in\mathbb{N}}$

Then $w_k \rightharpoonup \bar{w} \in \mathcal{Z}$ *almost surely*

Proof: combine standard (Opial) argument with Robbins–Siegmund lemma

Primal-dual proximal splitting: convergence

Theorem

Asssume

- J' *is locally Lipschitz, satisfies three-point inequality*
- U_{ad}, Π *bounded* ($\rightsquigarrow \{(u^k, Z^k)\}_{k\in\mathbb{N}}$ *bounded)*
- *choice of mini-batches satisfies*

$$\mathbb{E}\left[\operatorname*{ess\,sup}_{\xi\in\Xi}(1 - \mathbb{1}_{A_k})\right] \leq Mk^{-3} \quad a.s.$$

(*$\rightsquigarrow$ decay rate for expectation that* $\mathbb{P}[A_k^C] > 0$)

- *step size conditions (very technical)*

Then $(u^k, Z^k) \rightharpoonup (\bar{u}, \bar{Z})$ *saddle point*

Proof: apply abstract theorem to $R_k = T_k M_k$ with T_k testing operator (standard)

Risk-averse optimal control: examples

$$\min_{u \in U_{\text{ad}}} \text{CVaR}_\alpha[J(u,\xi)]$$

- $\Xi = \{\xi_1, \dots, \xi_S\}$ i.i.d. scenarios (here: $S = 1000$)
- $J(u,\xi) = \frac{1}{2}\|y(u) - y^d\|^2 + \frac{\gamma}{2}\|u\|^2$
- $y(u)$ solves random PDE
  1. elliptic with random jump diffusion
  2. Burgers' equation with random coefficients

- projection on Π_α not closed form $\rightsquigarrow$ algorithm, warm starts
- mini-batch: $A_k \subset \Xi$ with ξ_k selected with probability

$$p_k = \begin{cases} q_k & \text{if } k < M^{1/3} \\ \max\{q_k, (1 - Mk^{-3})^{1/S}\} & \text{if } k \geq M^{1/3} \end{cases}$$

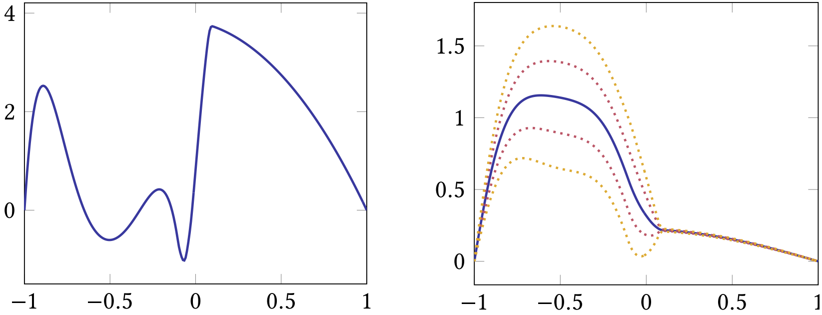
for $M > 0$, $q_k \in [0,1]$

Risk-averse optimal control: elliptic example

$$
\begin{aligned}
-\mathrm{div}(a(\xi)\nabla y) &= f(\xi) + u \qquad \text{in } \Omega \\
y &= 0 \qquad \text{on } \partial\Omega
\end{aligned}
$$

- $\Omega = (-1, 1)$
- $a(x, \xi) = 0.1\mathbb{1}_{(-1,\xi_1]}(x) + 10\mathbb{1}_{(\xi_1,1)}$
- $f(x, \xi) = \exp(-(x - \xi_2)^2)$
- $\xi_1 \sim \mathcal{U}((-.1, .1))$
- $\xi_2 \sim \mathcal{U}((-0.5, 0.5))$
- $y^d \equiv 1$
- $U_{\text{ad}} = \{u : -10 \le u(x) \le 10\}$
- $u^0 = 0, Z^0 = 0$, step sizes constant (estimated by power method)

Risk-averse optimal control: elliptic example

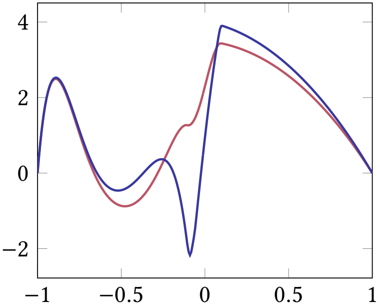


(a) optimal control

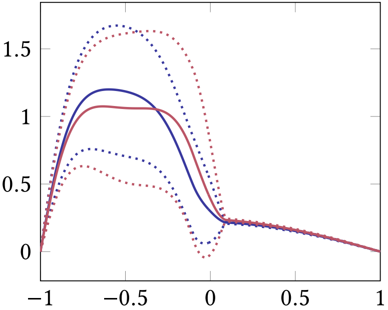
(b) optimal state mean $\pm\,\sigma$ and 2σ

Figure: risk-averse control, state for $\alpha = 0.1$

Risk-averse optimal control: elliptic example



(a) optimal controls



(b) optimal states mean $\pm 2\sigma$

Figure: risk-averse ($\alpha = 0.01$) vs. risk-neutral control

Risk-averse optimal control: elliptic example

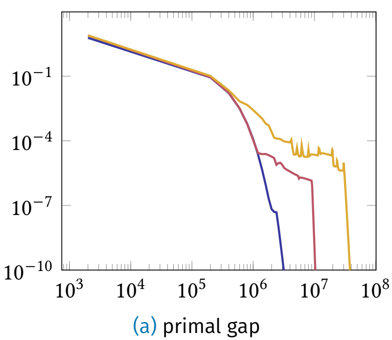


Figure: convergence w.r.t. PDE solves for $\alpha = 0.99$, $\alpha = 0.5$, $\alpha = 0.01$

Risk-averse optimal control: elliptic example

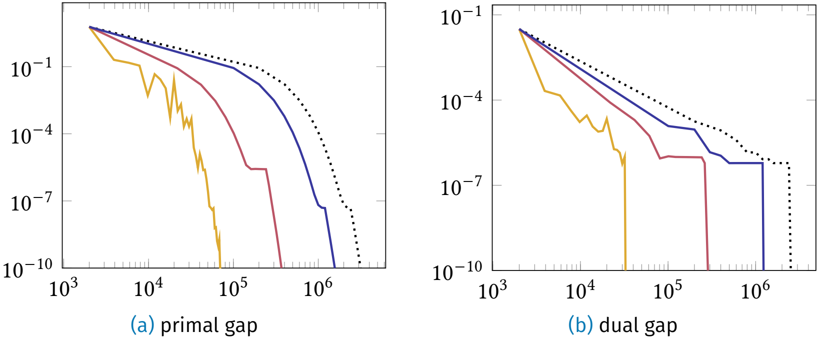


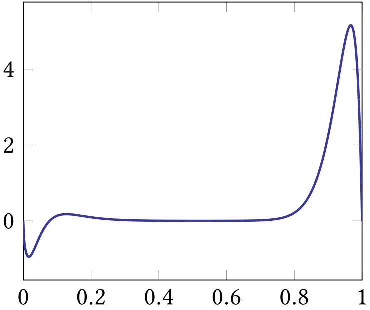
Figure: convergence w.r.t. PDE solves for $\alpha = 0.99$ and minibatch: $q_k = 0.5$, $q_k = 0.1$, $q_k = 0.01$, $M = 10^{20}$

Risk-averse optimal control: Burgers' equation

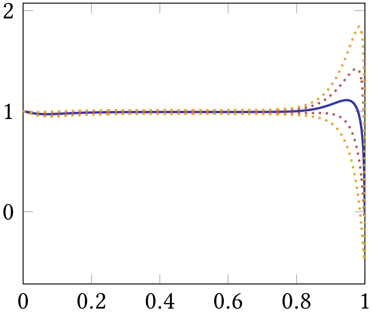
$$-\nu(\xi)\Delta y + y \operatorname{grad} y = f(\xi) + u \quad \text{in } (0,1)$$
$$y(0) = g_0(\xi), \quad y(1) = g_1(\xi)$$

- $\nu(\xi) = 10^{\xi_1 - 2}$
- $f(\xi) = 0.01\xi_2$
- $g_1(\xi) = 1 + 0.001\xi_3$
- $g_1(\xi) = 0.001\xi_4$
- $\xi_i \sim \mathcal{U}([-1,1])$ i.i.d.
- $y^d \equiv 1$
- $U_{\text{ad}} = \{u : -10 \leq u(x) \leq 10\}$
- $u^0 = 0, Z^0 = 0$, step sizes constant (estimated by power method)

Risk-averse optimal control: Burgers example



(a) optimal control



(b) optimal state mean $\pm\ \sigma$ and 2σ

Figure: risk-averse control, state for $\alpha = 0.1$

Risk-averse optimal control: Burgers example

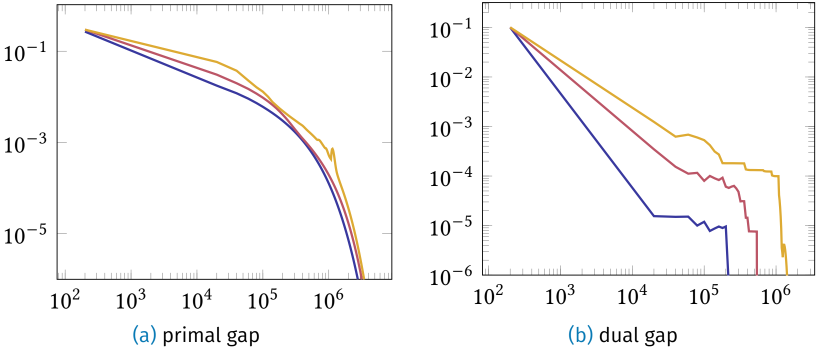
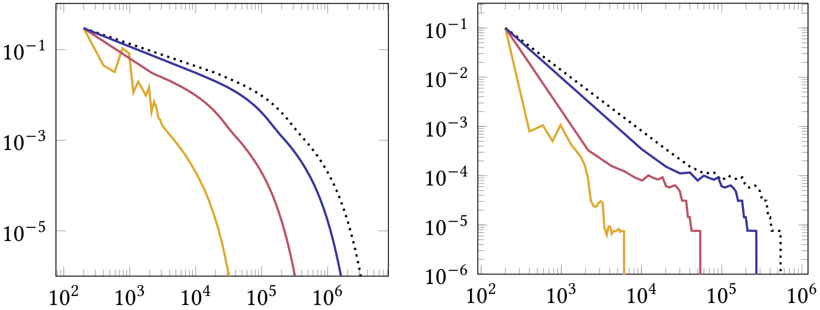


Figure: convergence w.r.t. PDE solves for $\alpha = 0.99$, $\alpha = 0.5$, $\alpha = 0.01$

Risk-averse optimal control: Burgers example



(a) primal gap (b) dual gap

Figure: convergence w.r.t. PDE solves for $\alpha = 0.99$ and minibatch: $q_k = 0.5$, $q_k = 0.1$, $q_k = 0.01$, $M = 10^{20}$

Conclusion

Risk-averse optimal control problems:

- can be formulated using coherent risk measures
- can be solved numerically using primal-dual splitting
- involve a lot of PDE solves...
- $\rightsquigarrow$ stochastic primal-dual splitting ($\approx 100\times$ speedup)

Outlook:

- acceleration (use strongly convex control costs)
- adaptive mini-batches (e.g., probability based on residual)
- application to probabilistic constraints
- other risk measures

Dissertation: `https://doi.org/10.17185/duepublico/78165`
Code: `https://zenodo.org/records/7121224`