Learning to reflect: On data-driven approaches to stochastic control

12$^{\text{th}}$ Bachelier World Congress of the Bachelier Finance Society – Rio de Janeiro

Lukas Trottner
based on joint work with Sören Christensen, Asbjørn Holk Thomsen and Claudia Strauch

10 July 2024

Aarhus University Kiel University

AARHUS UNIVERSITY

Framework for data-driven stochastic optimal control

- consider a d-dimensional ergodic diffusion

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,$$

  with stationary density π
- we assume that the drift b is unknown
- what challenges arise from this uncertainty when we want to optimally control the process and how can they be solved in a data-driven way?
- concrete control problems considered in the literature:
  1. impulse controls in 1D (Christensen, Strauch (2023); Christensen, Dexheimer, Strauch (2024+))
  2. reflection controls (singular) (Christensen, Strauch, T. (2024); Christensen, Holk Thomsen, T. (2024+))
- common theme: long-term average costs only depend on π and σ $\rightsquigarrow$ given observations of the (un)controlled process, first estimate π and then estimate optimal control as an M-estimator

Framework for data-driven stochastic optimal control

- consider a d-dimensional ergodic diffusion

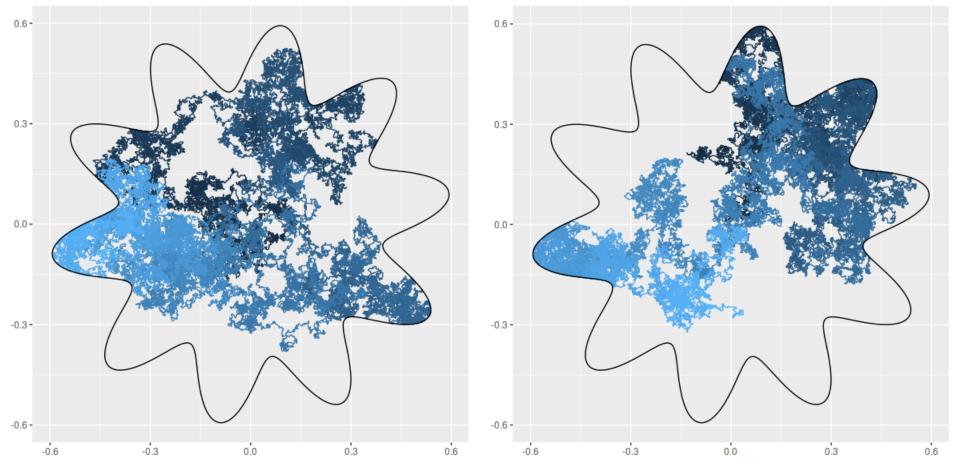
$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,$$

  with stationary density π
- we assume that the drift b is unknown
- what challenges arise from this uncertainty when we want to optimally control the process and how can they be solved in a data-driven way?
- concrete control problems considered in the literature:
  1. impulse controls in 1D (Christensen, Strauch (2023); Christensen, Dexheimer, Strauch (2024+))
  2. reflection controls (singular) (Christensen, Strauch, T. (2024); Christensen, Holk Thomsen, T. (2024+))
- common theme: long-term average costs only depend on π and $\sigma \rightsquigarrow$ given observations of the (un)controlled process, first estimate π and then estimate optimal control as an M-estimator

Problem

Exploration vs. exploitation

Reflected diffusions



For simplicity, assume that $b = -\nabla V$ for a gradient Lipschitz potential $V : \mathbb{R}^d \to \mathbb{R}$ and $\sigma = \sqrt{2}\mathbb{I}_d$, i.e.,

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t.$$

Let $D \subset \mathbb{R}^d$ be a sufficiently smooth bounded domain. Normally reflected process in D:

$$\mathrm{d}X_t^D = -\nabla V(X_t^D)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t + n(X_t^D)\,\mathrm{d}\underbrace{L_t^D}_{\text{local time at boundary } \partial D}.$$

Reflection control problem

For simplicity, assume that $b = -\nabla V$ for a gradient Lipschitz potential $V : \mathbb{R}^d \to \mathbb{R}$ and $\sigma = \sqrt{2}\mathbb{I}_d$, i.e.,

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t.$$

Let $D \subset \mathbb{R}^d$ be a sufficiently smooth bounded domain. Normally reflected process in D:

$$\mathrm{d}X_t^D = -\nabla V(X_t^D)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t + n(X_t^D)\,\mathrm{d}\underbrace{L_t^D}_{\text{local time at boundary } \partial D}.$$

Costs up to time T:

$$J_T(D) := \int_0^T c(X_s^D)\,\mathrm{d}s + \kappa L_T^D, \quad c : \mathbb{R}^d \to \mathbb{R}_+, \kappa > 0.$$

Example in 1D: interest rate model with central bank intervention

Long-term average costs: For $\tilde{\pi}(x) := \mathrm{e}^{-V(x)}, \tilde{\pi}(D) = \int_D \tilde{\pi}, \pi_D = \tilde{\pi}/\tilde{\pi}(D)$,

$$J(D) := \lim_{T\to\infty} \frac{1}{T}\mathbb{E}^\mu[J_T(D)] = \int_D c(x)\pi_D(x)\,\mathrm{d}x + \kappa \int_{\partial D} \pi_D(x)\,\mathcal{H}^{d-1}(\mathrm{d}x).$$

Optimisation objective: for a given domain class Θ determine

$$D^* \in \operatorname*{arg\,min}_{D\in\Theta} J(D).$$

For known dynamics we therefore arrive at a shape optimisation problem.

Costs up to time T:

$$J_T(D) := \int_0^T c(X_s^D)\,\mathrm{d}s + \kappa L_T^D, \quad c : \mathbb{R}^d \to \mathbb{R}_+, \kappa > 0.$$

Example in 1D: interest rate model with central bank intervention

Long-term average costs: For $\tilde{\pi}(x) := \mathrm{e}^{-V(x)}, \tilde{\pi}(D) = \int_D \tilde{\pi}, \pi_D = \tilde{\pi}/\tilde{\pi}(D)$,

$$J(D) := \lim_{T\to\infty} \frac{1}{T}\mathbb{E}^\mu[J_T(D)] = \int_D c(x)\pi_D(x)\,\mathrm{d}x + \kappa \int_{\partial D} \pi_D(x)\,\mathcal{H}^{d-1}(\mathrm{d}x).$$

Optimisation objective: for a given domain class Θ determine

$$D^* \in \operatorname*{arg\,min}_{D\in\Theta} J(D).$$

For known dynamics we therefore arrive at a shape optimisation problem.

Central statistical observation

X is ergodic iff $\tilde{\pi}(\mathbb{R}^d) < \infty$, in which case $\pi = \tilde{\pi}/\tilde{\pi}(\mathbb{R}^d)$ and

$$\pi_D(x) = \pi(x)/\pi(D), \quad x \in D.$$

- assume that we observe a full path $(X_t)_{t\in[0,T]}$ of the uncontrolled process
- assume sufficient regularity and ergodicity assumptions on X and that π has anisotropic Hölder regularity of order $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d) \in (1, \mathfrak{b}]^d$

⇝ we can determine a fully data-driven kernel estimator $\hat{\pi}_T$ such that

$$\mathbb{E}^{\pi}\big[\|\hat{\pi}_T - \pi\|_\infty\big] \lesssim \Psi_{d,\overline{\boldsymbol{\beta}}}(T), \quad \overline{\boldsymbol{\beta}} = \Big(\frac{1}{d}\sum_{i=1}^{d}\frac{1}{\beta_i}\Big)^{-1},$$

with minimax optimal rate $\Psi_{d,\overline{\boldsymbol{\beta}}}(T)$

Proposition (Christensen, Strauch, T. (2024); Christensen, Holk, T. (2024+))

Let $\hat{\pi}_T^* := \hat{\pi}_T \vee \underline{\pi}$, where $\pi \geq \underline{\pi}$ on $B(0,\overline{\lambda})$. Let Θ be a family of domains s.t. $B(0,\underline{\lambda}) \subset D \subset B(0,\overline{\lambda})$ and $\mathcal{H}^{d-1}(\partial D) \leq \Lambda$ for any $D \in \Theta$. For $\widehat{D}_T \in \arg\min_{D\in\Theta} J(D, \hat{\pi}_T^*)$, it holds for a warm start that

$$\mathbb{E}^{\mu}\big[J(\widehat{D}_T) - J(D^*)\big] \lesssim \Psi_{d,\overline{\boldsymbol{\beta}}}(T).$$

Learning the optimal reflection boundary

- assume that we observe a full path $(X_t)_{t\in[0,T]}$ of the uncontrolled process
- assume sufficient regularity and ergodicity assumptions on X and that π has anisotropic Hölder regularity of order $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d) \in (1, \mathfrak{b}]^d$

⇝ we can determine a fully data-driven kernel estimator $\hat{\pi}_T$ such that

$$\mathbb{E}^{\pi}\big[\|\hat{\pi}_T - \pi\|_\infty\big] \lesssim \Psi_{d,\overline{\boldsymbol{\beta}}}(T), \quad \overline{\boldsymbol{\beta}} = \Big(\frac{1}{d}\sum_{i=1}^{d}\frac{1}{\beta_i}\Big)^{-1},$$

with minimax optimal rate $\Psi_{d,\overline{\boldsymbol{\beta}}}(T)$

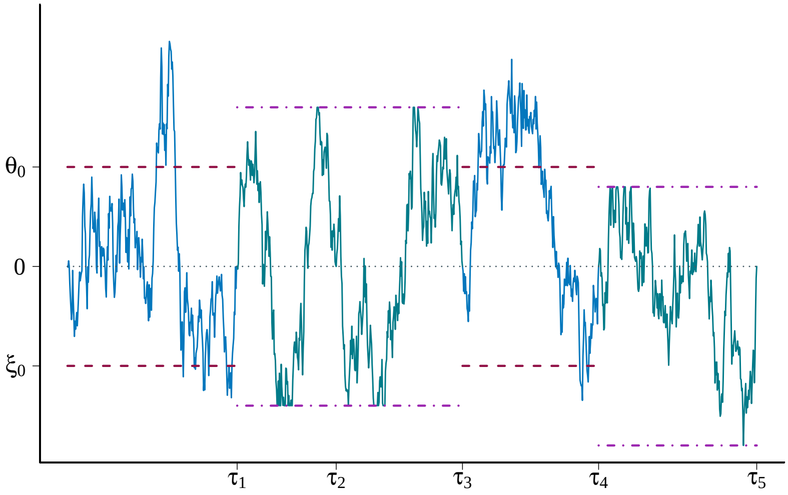
Proposition (Christensen, Strauch, T. (2024); Christensen, Holk, T. (2024+))

Let $\hat{\pi}_T^* := \hat{\pi}_T \vee \underline{\pi}$, where $\pi \geq \underline{\pi}$ on $B(0, \overline{\lambda})$. Let Θ be a family of domains s.t. $B(0, \underline{\lambda}) \subset D \subset B(0, \overline{\lambda})$ and $\mathcal{H}^{d-1}(\partial D) \leq \Lambda$ for any $D \in \Theta$. For $\widehat{D}_T \in \arg\min_{D\in\Theta} J(D, \hat{\pi}_T^*)$, it holds for a warm start that

$$\mathbb{E}^{\mu}\big[J(\widehat{D}_T) - J(D^*)\big] \lesssim \Psi_{d,\overline{\boldsymbol{\beta}}}(T).$$

⇝ this gives a bound on the simple regret only

⇝ how can we use this to determine strategies that overcome exploration vs. exploitation tradeoff with sublinear regret rate?


θ_0
0
ξ_0
τ_1
τ_2
τ_3
τ_4
τ_5

Theorem (Christensen, Strauch, T. (2024)[1]; Christensen, Holk, T. (2024+)[2])

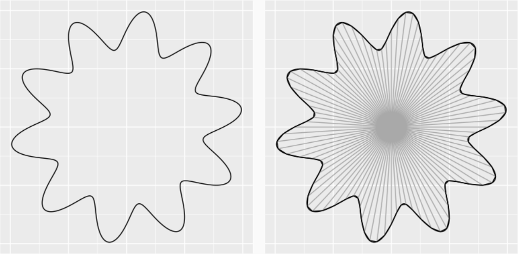
There exists a purely data-driven episodic domain learning strategy $\widehat{Z}$ such that the expected regret per time unit satisfies

$$\frac{1}{T}\mathbb{E}\Big[\int_0^T c(X_t^{\widehat{Z}})\,\mathrm{d}t + \kappa L_T^{\widehat{Z}}\Big] - J(D^*) \lesssim \begin{cases} \frac{\sqrt{\log T}}{T^{1/3}}, & d = 1, \\ \Big(\frac{(\log T)^2}{T}\Big)^{\frac{1}{3}}, & d = 2, \\ \Big(\frac{\log T}{T}\Big)^{\frac{\overline{\beta}}{3\overline{\beta}+d-2}}, & d \geq 3. \end{cases}$$

[1] Strauch, Christensen and Trottner (2024). Learning to reflect: A unifying approach to data-driven control strategies. *Bernoulli*

[2] Christensen, Holk Thomsen and Trottner (forthcoming). Data-driven rules for multidimensional reflection problems. *SIAM/ASA J. Uncert. Quantif.*

- as target domains Θ only allow strongly star-shaped sets at 0 (appropriate when continuous costs c are minimal close to the origin)
- for $D \in \Theta$ consider polytope approximation $\widetilde{D}_N$ such that for a sufficiently large number N of spanning points $J(D) \approx J(\widetilde{D}_N) = \tilde{J}(r_1, r_2, \dots, r_N)$
- we derive explicit formulas for $\nabla \tilde{J}(\boldsymbol{r})$, making gradient-based optimisation methods accessible



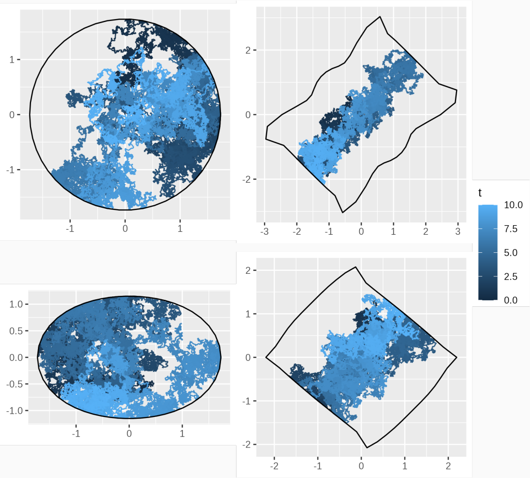


Figure 1: Simulated optimal shapes and corresponding path realizations of reflected processes. Top left: Brownian motion with norm cost. Top right: Ornstein–Uhlenbeck process with norm cost. Bottom left: Brownian motion with skewed cost. Bottom right: Ornstein–Uhlenbeck process with skewed cost.

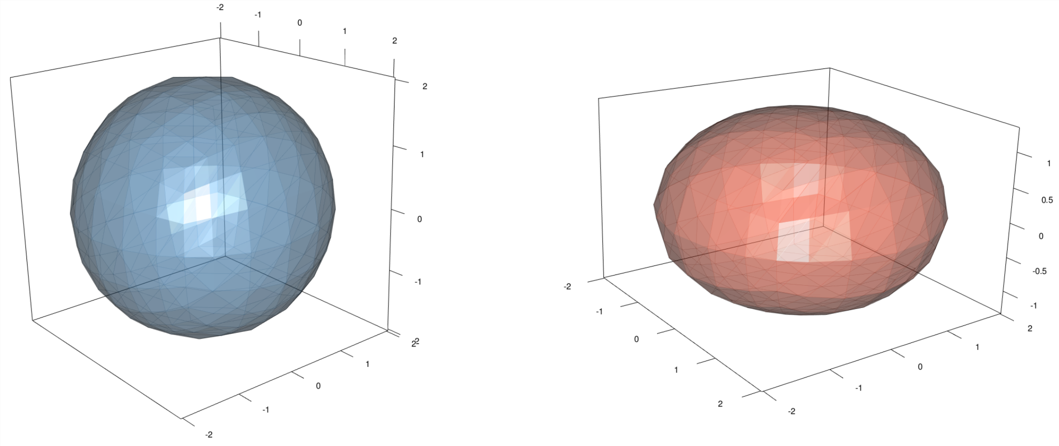


Figure 2: Optimised shapes for Brownian motion with reflection cost $\kappa = 1$ and cost function $c = |\cdot|$ (left) and $c(x, y, z) = \sqrt{x^2 + 5y^2 + z^2}$ (right).

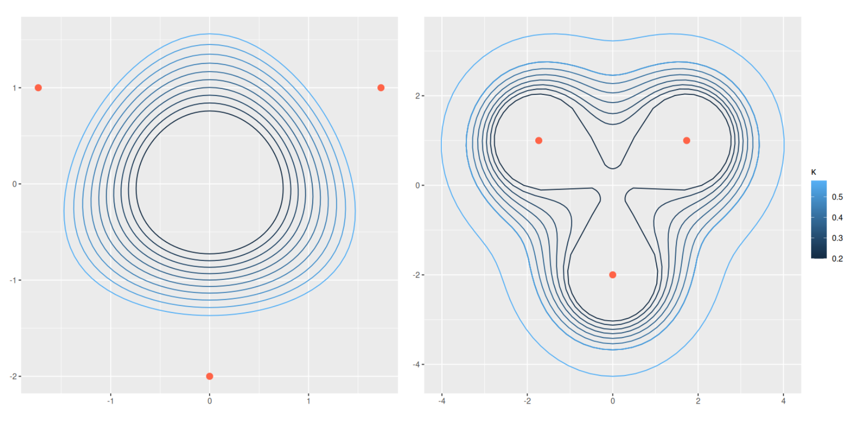


Figure 3: For each κ, we plot the optimized reflection boundaries, where π is a mixture of three Gaussians with means at the points marked in red. Left: Norm cost function, $c = |\cdot|$. Right: Cost function $c(x) = \min\{|x - \mu_1|, |x - \mu_2|, |x - \mu_3|\}$.

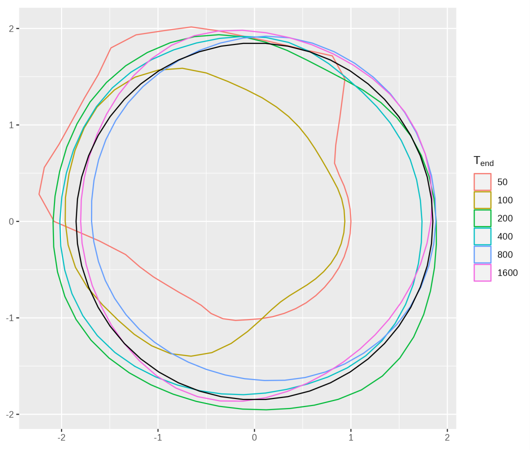


Figure 4: Estimates of the optimal shape (black) using kernel estimates after increasing periods of exploration. Notably, after only $T = 150$, the estimated optimal shape has an associated cost only 0.61% higher than the true optimum.

Summary

- we study singular control problems for ergodic diffusion processes (not part of the talk but of the paper: and Lévy processes) in presence of uncertainty on the characteristics
- our data-driven solutions are based on nonparametric adaptive estimation of quantities that characterize the optimal control policy
- the exploration-exploitation tradeoff is overcome by an appropriate separation of the timeline into exploration and exploitation phases
- we derive non-asymptotic regret rates from the minimax optimal sup-norm convergence rates of our estimators

Summary

- we study singular control problems for ergodic diffusion processes (not part of the talk but of the paper: and Lévy processes) in presence of uncertainty on the characteristics
- our data-driven solutions are based on nonparametric adaptive estimation of quantities that characterize the optimal control policy
- the exploration-exploitation tradeoff is overcome by an appropriate separation of the timeline into exploration and exploitation phases
- we derive non-asymptotic regret rates from the minimax optimal sup-norm convergence rates of our estimators

Thank you for your attention!