Learning to reflect – Data-driven solutions to singular control problems

21st INFORMS APS Conference

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

30 June 2023

Aarhus University Kiel University

Outline

1. A singular control problem for scalar ergodic diffusions
2. Data-driven approach to singular control
3. Extension to higher dimension

A singular control problem for scalar ergodic diffusions

Framework (1D)

regular 1-dim. Itô diffusion

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

with assumptions that guarantee an invariant density

$$\rho(x) \;\coloneqq\; \frac{1}{C\sigma^2(x)} \exp\left(2\int_0^x \frac{b(y)}{\sigma^2(y)} dy\right),$$

and ergodicity in the sense $\mathbb{P}(X_t \in \mathrm{d}x) \xrightarrow[t\to\infty]{\mathrm{TV}} \rho(x)\,\mathrm{d}x$.

Framework (1D)

- Singular control: $Z = (U_t, D_t)_{t \geqslant 0}$, U, D non-decreasing, right-continuous and adapted,

$$dX_t^Z = b(X_t^Z)\,dt + \sigma(X_t^Z)\,dW_t + dU_t - dD_t.$$

For reflection controls (U^{ξ}, D^{θ}) we have $X_t^Z \in [\xi, \theta]$ for all t

Framework (1D)

- Singular control: $Z = (U_t, D_t)_{t \geqslant 0}$, U, D non-decreasing, right-continuous and adapted,

$$\mathrm{d}X_t^Z = b(X_t^Z)\,\mathrm{d}t + \sigma(X_t^Z)\,\mathrm{d}W_t + \mathrm{d}U_t - \mathrm{d}D_t.$$

  For reflection controls (U^{ξ}, D^{θ}) we have $X_t^Z \in [\xi, \theta]$ for all t

- c continuous, nonnegative running cost function, $q_u, q_d > 0$.
  Minimize

$$\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\Big[\int_0^T c(X_s^Z)\,\mathrm{d}s + q_u U_T + q_d D_T\Big],$$

Solution for singular control problem

For each (ξ, θ), the corresponding reflection strategy has value

$$C(\xi,\theta) = \frac{1}{\int_{\xi}^{\theta} \rho(x)\,\mathrm{d}x} \left(\int_{\xi}^{\theta} c(x)\rho(x)\,\mathrm{d}x + \frac{q_u \sigma^2(\xi)}{2}\rho(\xi) + \frac{q_d \sigma^2(\theta)}{2}\rho(\theta) \right).$$

Solution for singular control problem

For each (ξ, θ), the corresponding reflection strategy has value

$$C(\xi, \theta) = \frac{1}{\int_\xi^\theta \rho(x)\,dx} \left(\int_\xi^\theta c(x)\rho(x)\,dx + \frac{q_u \sigma^2(\xi)}{2}\rho(\xi) + \frac{q_d \sigma^2(\theta)}{2}\rho(\theta) \right).$$

Theorem (Alvarez (2018))

Under some assumptions, the optimal value for the singular problem is given by

$$C^*_{\text{sing}} = \min_{(\xi,\theta)} C(\xi, \theta).$$

and the reflection strategy for the minimizer (ξ^*, θ^*) is optimal.

Data-driven approach to singular control

Questions

Central Assumption in Stochastic Control

The dynamics of the underlying process is known.

What to do if this is not the case?

- Which are the relevant *characteristics* of X to *estimate* approximately optimal boundaries?
- How does controlling the process *influence* the estimation?

Estimator

Crucial characteristics: b (assume σ to be known).

Estimator

Crucial characteristics: b (assume σ to be known).

Singular problem: $C^*_{\text{sing}} = \min_{(\xi,\theta)\in[-B,-1/B]\times[1/B,B]} C(\xi,\theta)$.

$$C(\xi,\theta) = \frac{1}{\int_\xi^\theta \rho(x)\,\mathrm{d}x}\left(\int_\xi^\theta c(x)\rho(x)\,\mathrm{d}x + \frac{q_u\sigma^2(\theta)}{2}\rho(c) + \frac{q_d\sigma^2(\xi)}{2}\rho(\theta)\right)$$

Estimator

Crucial characteristics: b (assume σ to be known).

Singular problem: $C^*_{\text{sing}} = \min_{(\xi,\theta)\in[-B,-1/B]\times[1/B,B]} C(\xi,\theta)$.

$$C(\xi,\theta) = \frac{1}{\int_\xi^\theta \rho(x)\,dx}\left(\int_\xi^\theta c(x)\rho(x)\,dx + \frac{q_u\sigma^2(\theta)}{2}\rho(c) + \frac{q_d\sigma^2(\xi)}{2}\rho(\theta)\right)$$

Plug-in estimator: If $\widehat{\rho}_T$ is an estimator of ρ and we know $\rho \geqslant \underline{\rho} > 0$ on $[-B,B]$, then for $\widehat{\rho}^*_T \coloneqq \widehat{\rho}_T \vee \underline{\rho}/2$ set

$$\widehat{C}_T(\xi,\theta) \coloneqq \frac{1}{\int_\xi^\theta \widehat{\rho}^*_T(x)\,dx}\left(\int_\xi^\theta c(x)\widehat{\rho}^*_T(x)\,dx + \frac{q_u\sigma^2(\xi)}{2}\widehat{\rho}^*_T(\xi) + \frac{q_d\sigma^2(\theta)}{2}\widehat{\rho}^*_T(\theta)\right),$$

$$\widehat{(\xi,\theta)}_T \in \operatorname*{arg\,min}_{(\xi,\theta)\in[-B,-1/B]\times[1/B,B]} \widehat{C}_T(\xi,\theta)$$

Estimator

Crucial characteristics: b (assume σ to be known).

Singular problem: $C^*_{\text{sing}} = \min_{(\xi,\theta)\in[-B,-1/B]\times[1/B,B]} C(\xi,\theta)$.

$$C(\xi,\theta) = \frac{1}{\int_\xi^\theta \rho(x)\,\mathrm{d}x}\left(\int_\xi^\theta c(x)\rho(x)\,\mathrm{d}x + \frac{q_u\sigma^2(\theta)}{2}\rho(c) + \frac{q_d\sigma^2(\xi)}{2}\rho(\theta)\right)$$

Plug-in estimator: If $\widehat{\rho}_T$ is an estimator of ρ and we know $\rho \geqslant \underline{\rho} > 0$ on $[-B,B]$, then for $\widehat{\rho}^*_T \coloneqq \widehat{\rho}_T \vee \underline{\rho}/2$ set

$$\widehat{C}_T(\xi,\theta) \coloneqq \frac{1}{\int_\xi^\theta \widehat{\rho}^*_T(x)\,\mathrm{d}x}\left(\int_\xi^\theta c(x)\widehat{\rho}^*_T(x)\,\mathrm{d}x + \frac{q_u\sigma^2(\xi)}{2}\widehat{\rho}^*_T(\xi) + \frac{q_d\sigma^2(\theta)}{2}\widehat{\rho}^*_T(\theta)\right),$$

$$\widehat{(\xi,\theta)}_T \in \operatorname*{arg\,min}_{(\xi,\theta)\in[-B,-1/B]\times[1/B,B]} \widehat{C}_T(\xi,\theta)$$

If we have a deterministic bound $\|\widehat{\rho}_T\|_\infty \leqslant c(T)$,

$$\mathbb{E}_b\left[C(\widehat{(\xi,\theta)}_T) - C^*_{\text{sing}}\right] \leqslant 2\mathbb{E}_b\Big[\max_{(\xi,\theta)\in[-B,-1/B]\times[1/B,B]}\left|C(\xi,\theta) - \widehat{C}_T(\xi,\theta)\right|\Big]$$

$$\lesssim \mathbb{E}_b\left[\|\widehat{\rho}_T - \rho\|_{L^\infty([-B,B])}\right] + c(T)\mathbb{P}_b\left(\|\widehat{\rho}_T - \rho\|_{L^\infty([-B,B])} \geqslant \underline{\rho}/2\right).$$

$\rightsquigarrow$ need non-asymptotic sup-norm concentration rates for appropriate nonparametric estimator $\widehat{\rho}_T$

Concentration of kernel density estimator

Let

$$\widehat{\rho}_T(x) \coloneqq \frac{1}{Th_T}\int_0^T K\left(\frac{x - X_t}{h_T}\right) \mathrm{d}t$$

be a kernel estimator for ρ.

Proposition (Christensen, Strauch, T. (2023+))

Suppose that

1. b, σ are Lipschitz and $0 < \underline{\sigma} \leqslant \sigma(x) \leqslant \overline{\sigma} < \infty$ for all x;
2. for some $\gamma, A > 0$, $\operatorname{sgn}(x)b(x) \leqslant -\gamma$ if $|x| > A$;
3. $\rho_b \in C^1(\mathbb{R})$ with Hölder continuous derivative.

Then, given a compactly supported and symmetric probability density K and the bandwidth choice $h_T \sim (\log T)^2/\sqrt{T}$ we have

$$\mathbb{E}_b^0\left[\|\widehat{\rho}_T - \rho\|^p_{L^\infty(D)}\right]^{1/p} \in \mathrm{O}\left(\sqrt{\frac{\log T}{T}}\right),$$

for any $p \geqslant 1$ and any open, bounded domain D.

Recall from before:

$$\mathbb{E}_b^0\left[C(\widehat{(\xi,\theta)}_T) - C^*_{\text{sing}}\right] \lesssim \mathbb{E}_b^0\left[\|\widehat{\rho}_T - \rho\|_{L^\infty([-B,B])}\right] + \frac{\sqrt{T}}{(\log T)^2}\mathbb{P}_b^0\left(\|\widehat{\rho}_T - \rho\|_{L^\infty([-B,B])} \geqslant \underline{\rho}/2\right).$$

Corollary (Christensen, Strauch, T. (2023+))

Given the previous assumptions on X, it holds

$$\mathbb{E}_b^0\left[C(\widehat{(\xi,\theta)}_T) - C^*_{\text{sing}}\right] \in \mathrm{O}\left(\sqrt{\frac{\log T}{T}}\right).$$

Exploration vs. exploitation

Naïve idea:

- estimate the optimal boundary based on the controlled process
- use the strategy based on the estimated boundary

Exploration vs. exploitation

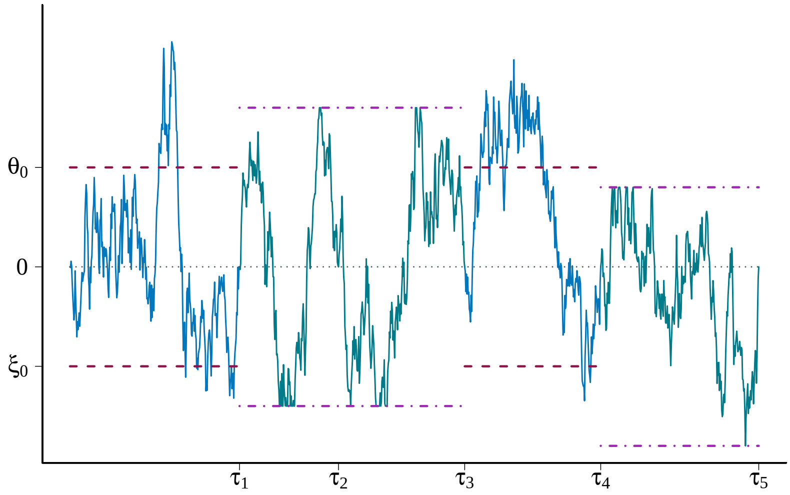
Naïve idea:

- estimate the optimal boundary based on the controlled process
- use the strategy based on the estimated boundary

Problem

Exploration vs. Exploitation!

Regret given explorations vs. exploitation tradeoff



Theorem (Christensen, Strauch, T. (2023+))

If we consider a data-driven reflection strategy $\widehat{Z}$ s.t. the time S_T spent in exploration periods until time T is of order $S_T \approx T^{2/3}$, then the expected regret per time unit,

$$\frac{1}{T}\mathbb{E}_b^0\Big[\int_0^T c(X_s^{\widehat{Z}})\,\mathrm{d}s + q_u U_T^{\widehat{Z}} + q_d D_T^{\widehat{Z}}\Big] - C^*_{\mathrm{sing}},$$

is of order $\mathrm{O}\big(\sqrt{\log T}\; T^{-1/3}\big)$.

Extension to higher dimension

The multivariate case

- Let now $d \geqslant 2$ and consider an ergodic d-simensional Langevin diffusion

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

  with C^2 potential $V\colon \mathbb{R}^d \to \mathbb{R}$ and invariant density $\rho \asymp \exp(-V(\cdot))$;
- normally reflected process in domain D of class C^2

$$\mathrm{d}X_t^D = -\nabla V(X_t^D) + \sqrt{2}\,\mathrm{d}W_t + n(X_t^D)\,\mathrm{d}L_t^D,$$

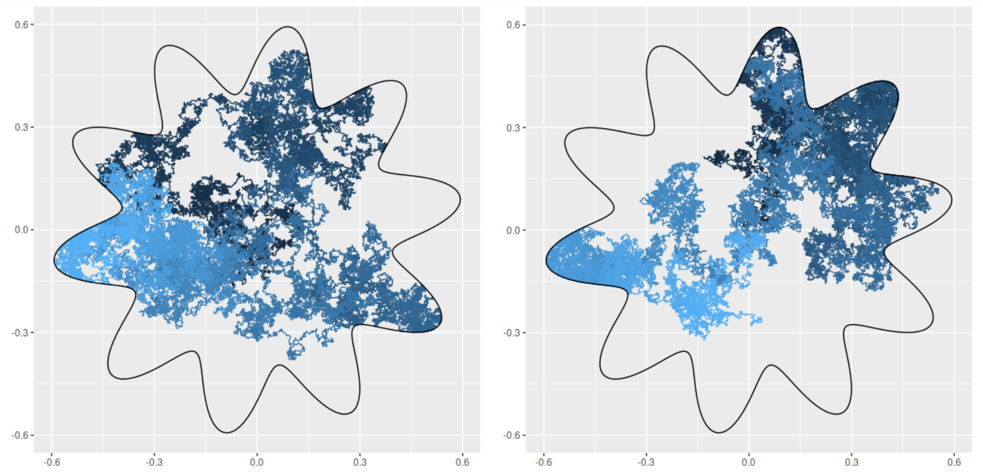
  where n is the inward unit normal vector of D and $L_\cdot^D$ is local time of X on ∂D
- control problem: minimize

$$C(D) \coloneqq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\Big[\int_0^T c(X_t^D)\,\mathrm{d}t + \kappa L_T^D\Big]$$

  over appropriate class of domains D
- numerically, this is, e.g., tractable for star shaped sets, i.e., domains D given by

$$D = \big\{r(q)q : q \in S^{d-1}\big\}, \quad r\colon S^{d-1} \to (0,\infty) \text{ smooth.}$$

Reflected diffusions on star shaped sets



A formula for the ergodic costs

Theorem (Christensen, Holk, T. (2023+))

If D is a non-empty, bounded domain of class C^2, then

$$C(D) = C(D, \rho) = \frac{1}{\int_D \rho(x)\,\mathrm{d}x}\left(\int_D c(x)\rho(x)\,\mathrm{d}x + \kappa \int_{\partial D} \rho(y)\,\mathcal{H}^{d-1}(\mathrm{d}y)\right),$$

where $\mathcal{H}^{d-1}$ is the $(d-1)$-dimensional Hausdorff measure.

Learning the optimal reflection boundary

Multivariate kernel density estimator:

$$\widehat{\rho}_{\boldsymbol{h},T}(x) \coloneqq \frac{1}{\prod_{i=1}^{d} h_i} \int_0^T \mathbb{K}((x - X_t)/\boldsymbol{h})\,\mathrm{d}t, \quad \mathbb{K}(x) \coloneqq \prod_{i=1}^{d} K(x_i),\, x/\boldsymbol{h} \coloneqq (x_i/h_i)_{i=1,\ldots,d}.$$

Results from Strauch (2018) show that if X satisfies both a Poincaré inequality and a Nash inequality, then under anisotropic β-Hölder smoothness assumptions on ρ and gsufficient order of K, there exists an adaptive bandwith choice $\widehat{\boldsymbol{h}}_T$ such that

$$\mathbb{E}\Big[\big\|\widehat{\rho}_{\widehat{\boldsymbol{h}}_T,T} - \rho\big\|_\infty^p\Big]^{1/p} \lesssim \Psi_{d,\beta}(T) \coloneqq \begin{cases} \frac{\log T}{\sqrt{T}}, & d = 2, \\ \left(\frac{\log T}{T}\right)^{\frac{\overline{\beta}}{2\overline{\beta}+d-2}}, & d \geqslant 3, \end{cases} \quad \text{where } \overline{\beta} = \Big(\frac{1}{d}\sum_{i=1}^{d}\frac{1}{\beta_i}\Big)^{-1}.$$

Proposition (Christensen, Holk, T. (2023+))

Let $\widehat{\rho}_T^* \coloneqq (\widehat{\rho}_T \wedge 2\overline{\rho}) \vee \underline{\rho}/2$, where for some $\underline{\lambda}, \overline{\lambda} > 0$, $\rho \in [\underline{\rho}, \overline{\rho}]$ on $B(0,\overline{\lambda}) \setminus B(0,\underline{\lambda})$. Let Θ be a family of domains s.t. $B(0,\underline{\lambda}) \subset D \subset B(0,\overline{\lambda})$ and $\mathcal{H}^{d-1}(\partial D) \leqslant \Lambda$ for any $D \in \Theta$. Then for $\widehat{D}_T \in \arg\min_{D\in\Theta} C(D, \widehat{\rho}^*_{\widehat{\boldsymbol{h}}_T,T})$, it holds

$$\mathbb{E}\big[C(\widehat{D}_T, \rho) - \min_{D\in\Theta} C(D,\rho)\big] \lesssim \Psi_{d,\beta}(T).$$

Summary

- we study singular control problems for ergodic diffusion processes in presence of uncertainty on the drift
- our data-driven solutions are based on nonparametric adaptive estimation of the invariant density
- in the one-dimensional case, the exploration-exploitation tradeoff is overcome by separating the timeline into exploration and exploitation phases of random length
- we derive non-asymptotic regret rates from the minimax optimal sup-norm convergence rates of the invariant density estimator

Summary

- we study singular control problems for ergodic diffusion processes in presence of uncertainty on the drift
- our data-driven solutions are based on nonparametric adaptive estimation of the invariant density
- in the one-dimensional case, the exploration-exploitation tradeoff is overcome by separating the timeline into exploration and exploitation phases of random length
- we derive non-asymptotic regret rates from the minimax optimal sup-norm convergence rates of the invariant density estimator

Thank you for your attention!