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

ISOR Colloquium – University of Vienna

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

8 April 2024

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 dimensions
4. Data-driven optimal control for Lévy processes

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

$$\pi(x) \ := \ \frac{1}{C\sigma^2(x)} \exp\left(2\int_0^x \frac{b(y)}{\sigma^2(y)}\,\mathrm{d}y\right),$$

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

Framework (1D)

- Singular control: $Z = (U_t, D_t)_{t\geq 0}$, 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

Framework (1D)

- Singular control: $Z = (U_t, D_t)_{t\geq 0}$, 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

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

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

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

Solution for singular control problem

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

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

Theorem (Alvarez (2018))

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

$$V_{\text{sing}} = \min_{(\xi,\theta)} C(\xi, \theta).$$

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?

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

Estimator

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

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

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

Estimator

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

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

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

Plug-in estimator: If $\hat{\pi}_T$ is an estimator of π and we know $\pi \geq \underline{\pi} > 0$ on $[-B, B]$, then for $\hat{\pi}_T^* := \hat{\pi}_T \vee \underline{\pi}$ set

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

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

Estimator

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

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

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

Plug-in estimator: If $\hat{\pi}_T$ is an estimator of π and we know $\pi \geq \underline{\pi} > 0$ on $[-B, B]$, then for $\hat{\pi}_T^* := \hat{\pi}_T \vee \underline{\pi}$ set

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

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

Then,

$$\mathbb{E}_b\left[C(\widehat{(c,d)}_T) - V_{\text{sing}}\right] \leq 2\mathbb{E}_b\Big[\max_{(\xi,\theta)\in[-B,-1/B]\times[1/B,B]}\left|C(\xi,\theta) - \widehat{C}_T(\xi,\theta)\right|\Big] \lesssim \mathbb{E}_b\left[\|\hat{\pi}_T - \pi\|_{L^\infty([-B,B])}\right].$$

$\rightsquigarrow$ need non-asymptotic sup-norm rates for an appropriate nonparametric estimator $\hat{\pi}_T$

Concentration of kernel density estimator

Let

$$\hat{\pi}_T(x) := \frac{1}{Th_T}\int_0^T K\Big(\frac{x - X_t}{h_T}\Big)\,\mathrm{d}t$$

be a kernel estimator for π.

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

Suppose that

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

$$\mathbb{E}_b^0\Big[\|\hat{\pi}_T - \pi\|^p_{L^\infty(D)}\Big]^{1/p} \in \mathrm{O}\Bigg(\sqrt{\frac{\log T}{T}}\Bigg),$$

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

Combining

$$\mathbb{E}_b^0\left[C(\widehat{(c,d)}_T) - V_{\text{sing}}\right] \lesssim \mathbb{E}_b^0\left[\|\hat{\pi}_T - \pi\|_{L^\infty([-B,B])}\right]$$

and

$$\mathbb{E}_b^0\left[\|\hat{\pi}_T - \pi\|_{L^\infty([-B,B])}\right] \in \mathrm{O}\left(\sqrt{\frac{\log T}{T}}\right)$$

we obtain:

Corollary (Christensen, Strauch, T. 2024+))

Given the previous assumptions on X, it holds

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

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

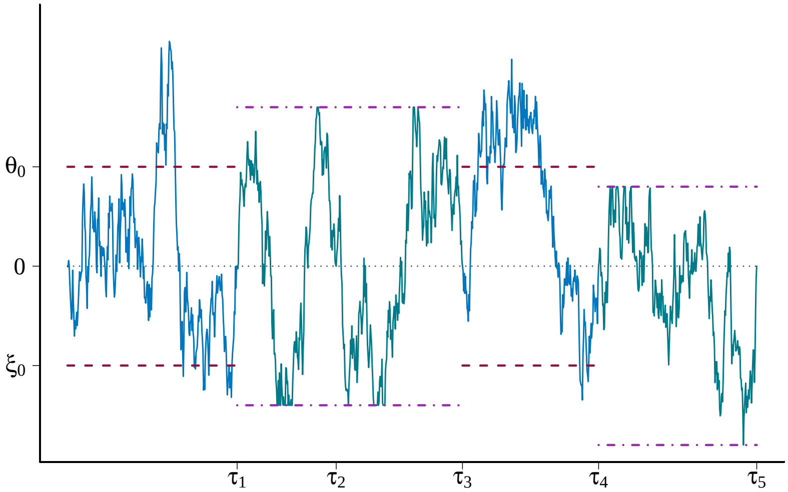
Naïve idea:

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

Problem

Exploration vs. Exploitation!

Strategy to overcome exploration vs. exploitation dilemma



Regret bound

Theorem (Christensen, Strauch, T. (2024+))

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,

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

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

Extension to higher dimensions

The multivariate case

- Let now $d \geq 2$ and consider a 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 $V : \mathbb{R}^d \to \mathbb{R}$;
- if e^{-V} is integrable, then X is ergodic with invariant density $\pi \propto \mathrm{e}^{-V}$;
- 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

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

over appropriate class of domains D ($c \geq$ costs inside domain, $\kappa > 0$ costs at the boundary)

A formula for the ergodic costs

Let $\mathbf{D}$ be a class of domains such that for any $D \in \mathbf{D}$ we have $\inf_{x,y\in\overline{D}} p_1^D(x,y) > 0$.

Theorem (Christensen, Holk, T. (2024+))

Let $\mathcal{H}^{d-1}$ be the $(d-1)$-dimensional Hausdorff measure. For any $x \in D \in \mathbf{D}$, it holds that

$$C(D) = \frac{1}{\int_D \mathrm{e}^{-V(y)}\,\mathrm{d}y}\Big(\int_D c(y)\mathrm{e}^{-V(y)}\,\mathrm{d}y + \kappa\int_{\partial D} \mathrm{e}^{-V(y)}\,\mathcal{H}^{d-1}(\mathrm{d}y)\Big).$$

and

$$\mathbb{E}^x\Big[\Big|\frac{1}{T}\Big(\int_0^T c(X_t^D)\,\mathrm{d}t + \kappa L_T^D\Big) - C(D)\Big|\Big] \lesssim_D \frac{1}{\sqrt{T}}.$$

If $\mathrm{e}^{-V} \in L^1(\mathbb{R}^d)$, then in particular

$$C(D) = C(D,\pi) = \frac{1}{\int_D \pi(y)\,\mathrm{d}y}\Big(\int_D c(y)\pi(y)\,\mathrm{d}y + \kappa\int_{\partial D} \pi(y)\,\mathcal{H}^{d-1}(\mathrm{d}y)\Big).$$

- let D be strongly star-shaped at 0 $\rightsquigarrow \partial D = \{r(q)q \,:\, q \in S^{d-1}\}$ for some radial function $r\,:\, S^{d-1} \to (0,\infty)$
- for N points $\{q_i\}_{i=1}^N \subset S^{d-1}$ consider the polytope $\widetilde{D}$ with vertices $\{p_i\}_{i=1}^N = \{r(q_i)q_i\}_{i=1}^N \rightsquigarrow \widetilde{D}$ can be split into N simplices $\{S_I\}_{I\in\mathcal{J}}$ with facets $\{F_I\}_{I\in\mathcal{J}}$ opposite the origin
- for $\boldsymbol{r} = \{r_i\}_{i=1}^N = \{r(q_i)\}_{i=1}^N$ we have

$$J(D) \approx J(\widetilde{D}) \equiv J(\boldsymbol{r}) = \frac{1}{\sum_{I\in\mathcal{J}} \int_{S_I} \mathrm{e}^{-V(x)}\,\mathrm{d}x} \sum_{I\in\mathcal{J}} \left(\int_{S_I} c(x)\mathrm{e}^{-V(x)}\,\mathrm{d}x + \kappa \int_{F_I} \mathrm{e}^{-V(x)}\,\mathcal{H}^{d-1}(\mathrm{d}x) \right)$$

- we derive explicit expressions for $\frac{\partial J(\boldsymbol{r})}{\partial r_i}$ based on which we propose a gradient descent algorithm to optimize ergodic costs in the class of star-shaped domains

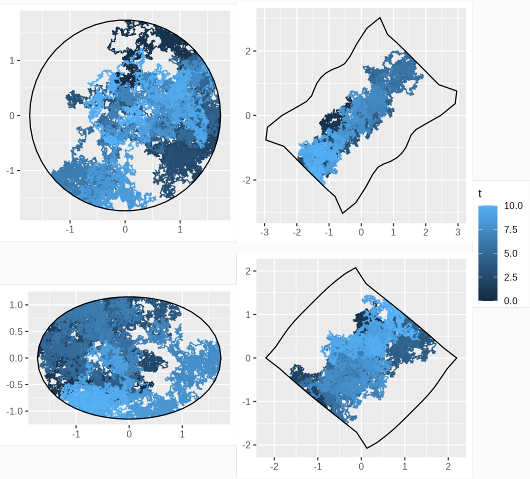


Figure 1: Simulated optimal shapes and corresponding path realizations of reflected processes.

Learning the optimal reflection boundary

Multivariate kernel density estimator:

$$\hat{\pi}_{\boldsymbol{h},T}(x) := \frac{1}{\prod_{i=1}^{d} h_i} \int_0^T \mathbb{K}((x - X_t)/\boldsymbol{h})\,\mathrm{d}t, \quad \mathbb{K}(x) := \prod_{i=1}^{d} K(x_i), \quad x/\boldsymbol{h} := (x_i/h_i)_{i=1,\dots,d}.$$

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

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

Multivariate kernel density estimator:

$$\hat{\pi}_{\boldsymbol{h},T}(x) := \frac{1}{\prod_{i=1}^d h_i} \int_0^T \mathbb{K}((x - X_t)/\boldsymbol{h})\,\mathrm{d}t, \quad \mathbb{K}(x) := \prod_{i=1}^d K(x_i), \quad x/\boldsymbol{h} := (x_i/h_i)_{i=1,\dots,d}.$$

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

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

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

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 $\widehat{D}_T \in \arg\min_{D\in\Theta} C(D, \hat{\pi}^*_{\hat{\boldsymbol{h}}_T,T})$, it holds for a warm start that

$$\mathbb{E}^\mu\Big[C(\widehat{D}_T,\pi) - \min_{D\in\Theta} C(D,\pi)\Big] \lesssim \Psi_{d,\boldsymbol{\beta}}(T).$$

Numerical implementation

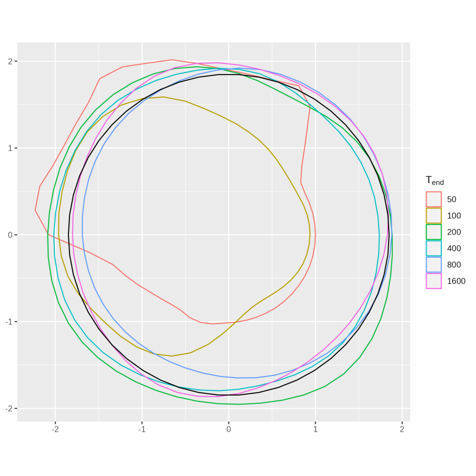


Figure 2: Estimates of the optimal shape (black) using kernel estimates after increasing periods of exploration. Notably, after only $T = 150$, the estimated optimal shape has an associated cost only 0.61% higher than the true optimum.

Episodic domain learning

- X does not hit points $\rightsquigarrow$ construction of stochastic exploration/exploitation intervals as in the one-dimensional case not feasible
- instead: alternate between exploration/exploitation intervals with deterministic lengths a_i and b_i (+ small stochastic fluctuation for exploitation lengths to make sure that the process is inside of proposed reflection domain)

Theorem (Christensen, Holk, T. (2024+))

If we choose exploration lengths $a_i \asymp 2^i$ and exploitation lengths $b_i \asymp a_i/\Psi_{d,\beta}(a_i)$, then a strategy that reflects the process in the i-th exploitation interval at a boundary estimated based on data collected in the i-th exploration interval, yields a regret per time unit of order

$$\frac{1}{T}\mathbb{E}[\widetilde{C}_T] - C(D^*) \leq \begin{cases} \left(\frac{(\log T)^2}{T}\right)^{\frac{1}{3}}, & d = 2, \\ \left(\frac{\log T}{T}\right)^{\frac{\overline{\beta+1}}{3\overline{\beta+1}+d-2}}, & d \geq 3. \end{cases}$$

Data-driven optimal control for Lévy processes

An optimal control problem for Lévy processes

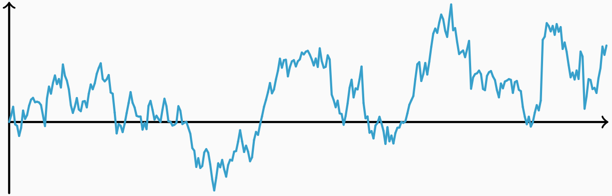
- ξ upward regular Lévy process on $\mathbb{R}$, $\mathbb{E}^0[\xi_1] \in (0, \infty)$
- for impulse controls $S = (\tau_n, \zeta_n)_{n \in \mathbb{N}}$

$$\xi_t^S = \xi_t - \sum_{n : \tau_n \leq t} (\xi_{\tau_n,-}^S - \zeta_n)$$

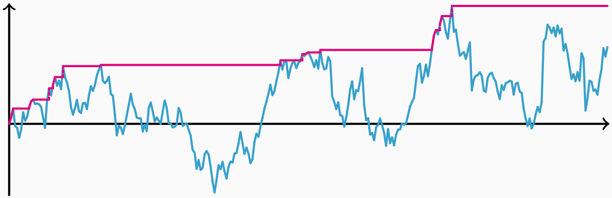
and for a given value function γ solve

$$v^* := \sup_S \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}^x \Big[\sum_{n : \tau_n \leq T} \big(\gamma(\xi_{\tau_n,-}^S) - \gamma(\zeta_n) \big) \Big]$$

Wiener–Hopf factorisation (path picture)

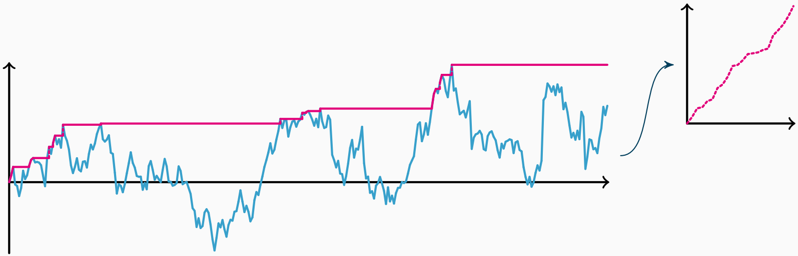


X, a Lévy process



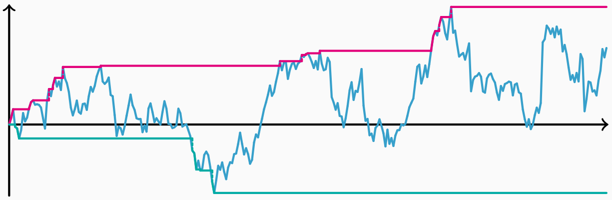
- X, a Lévy process
- $\bar{X}_t = \sup_{s \leq t} X_s$, the running supremum

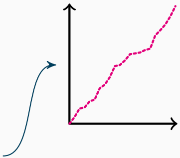
Wiener–Hopf factorisation (path picture)



- X, a Lévy process
- $\bar{X}_t = \sup_{s \le t} X_s$, the running supremum
- H_t, the ascending ladder height process: suprema 'stitched together'

Wiener–Hopf factorisation (path picture)



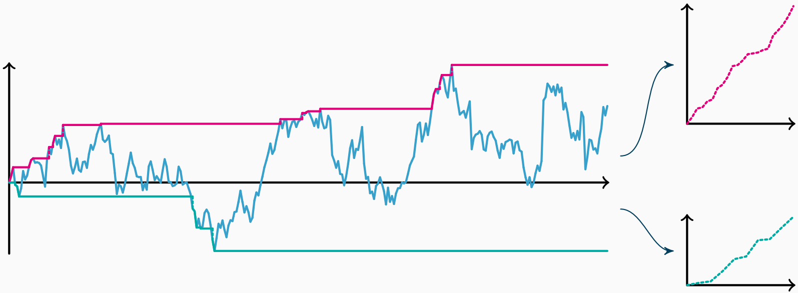


—— X, a Lévy process

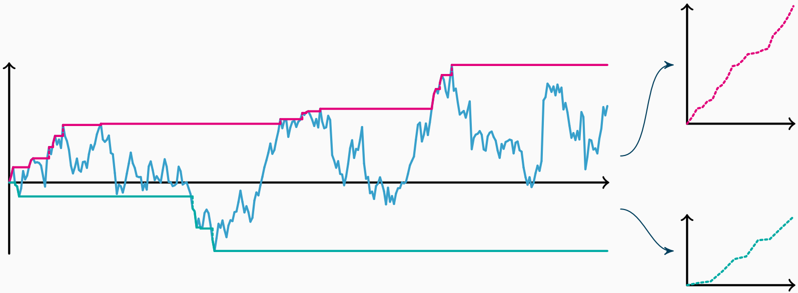
—— $\bar{X}_t = \sup_{s \le t} X_s$, the running supremum

········ H_t, the ascending ladder height process: suprema ‘stitched together’

········



- X, a Lévy process
- $\bar{X}_t = \sup_{s \le t} X_s$, the running supremum
- H_t, the ascending ladder height process: suprema 'stitched together'
- $\widehat{H}_t$, the descending ladder height process



- X, a Lévy process
- $\bar{X}_t = \sup_{s \le t} X_s$, the running supremum
- H_t, the ascending ladder height process: suprema ‘stitched together’
- $\widehat{H}_t$, the descending ladder height process

$H, \widehat{H}$ are subordinators (increasing Lévy processes), possibly killed

Solution for known dynamics

- essential process determining optimal solution: ascending ladder height process $H_t = \xi_{L_t^{-1}}$, where $(L_t)_{t\geq 0}$ is local time at supremum of ξ
- Reason: for scaling of L s.t. $\mathbb{E}^0[\xi_1] = \mathbb{E}^0[H_1]$ the long term average reward when reflecting in x is given by

$$d_H\gamma'(x) + \int_0^\infty (\gamma(x+y) - \gamma(x))\,\Pi_H(\mathrm{d}y) = \mathcal{A}_H\gamma(x) = \lim_{\varepsilon\downarrow 0} \frac{\mathbb{E}^x[\gamma(\xi_{T_{x+\varepsilon}})] - \gamma(x)}{\mathbb{E}^x[T_{x+\varepsilon}]}$$

- essential process determining optimal solution: ascending ladder height process $H_t = \xi_{L_t^{-1}}$, where $(L_t)_{t\geq 0}$ is local time at supremum of ξ
- Reason: for scaling of L s.t. $\mathbb{E}^0[\xi_1] = \mathbb{E}^0[H_1]$ the long term average reward when reflecting in x is given by

$$d_H\gamma'(x) + \int_0^\infty (\gamma(x+y) - \gamma(x))\,\Pi_H(\mathrm{d}y) = \mathcal{A}_H\gamma(x) = \lim_{\varepsilon\downarrow 0} \frac{\mathbb{E}^x[\gamma(\xi_{T_{x+\varepsilon}})] - \gamma(x)}{\mathbb{E}^x[T_{x+\varepsilon}]}$$

Theorem (Christensen, Sohr (2020))

Let $f := \mathcal{A}_H\gamma$ be unimodal with maximizer θ^* (+ technical assumptions). Then $v^* = f(\theta^*)$ and reflecting in θ^* is optimal.

Data-driven strategy

1. Construct nonparametric estimator $\hat{f}_T$ of $f = \mathcal{A}_H \gamma$ given data $(\xi_t)_{t \in [0,T]}$

Data-driven strategy

1. Construct nonparametric estimator $\hat{f}_T$ of $f = \mathcal{A}_H \gamma$ given data $(\xi_t)_{t \in [0,T]}$
2. Use $\hat{\theta}_T := \arg\max_x \hat{f}_T(x)$ as an estimator of the optimal reflection boundary

Data-driven strategy

1. Construct nonparametric estimator $\hat{f}_T$ of $f = \mathcal{A}_H \gamma$ given data $(\xi_t)_{t\in[0,T]}$
2. Use $\hat{\theta}_T := \arg\max_x \hat{f}_T(x)$ as an estimator of the optimal reflection boundary
3. Analyze sup-norm estimation rates of $\hat{f}_T$ to determine regret of the strategy, since for $\theta^* \in D$

$$\mathbb{E}^0[f(\theta^*) - f(\hat{\theta}_T)] \leq 2\mathbb{E}^0[\|\hat{f}_T - f\|_{L^\infty(D)}].$$

Data-driven strategy

1. Construct nonparametric estimator $\hat{f}_T$ of $f = \mathcal{A}_H\gamma$ given data $(\xi_t)_{t\in[0,T]}$
2. Use $\hat{\theta}_T := \arg\max_x \hat{f}_T(x)$ as an estimator of the optimal reflection boundary
3. Analyze sup-norm estimation rates of $\hat{f}_T$ to determine regret of the strategy, since for $\theta^* \in D$

$$\mathbb{E}^0[f(\theta^*) - f(\hat{\theta}_T)] \leq 2\mathbb{E}^0[\|\hat{f}_T - f\|_{L^\infty(D)}].$$

Statistical challenge

How can we build an estimator of $\mathcal{A}_H\gamma$ although local time L cannot be observed?

- Integration by parts reveals

$$\mathcal{A}_H\gamma(x) = d_H\gamma'(x) + \int_{0+}^{\infty}(\gamma(x+y)-\gamma(x))\,\Pi_H(\mathrm{d}y) = \int_0^{\infty}\eta\gamma'(x+y)\,\mu(\mathrm{d}y),$$

where $\eta = \mathbb{E}^0[\xi_1]$ and

$$\mu(\mathrm{d}y) = \frac{1}{\mathbb{E}^0[\xi_1]}\Big(d_H\delta_0(\mathrm{d}y) + \Pi_H((y,\infty))\,\mathrm{d}y\,\mathbf{1}_{(0,\infty)}(y)\Big), \quad y \geq 0.$$

- Integration by parts reveals

$$\mathcal{A}_H\gamma(x) = d_H\gamma'(x) + \int_{0+}^{\infty}(\gamma(x+y)-\gamma(x))\,\Pi_H(\mathrm{d}y) = \int_0^{\infty}\eta\gamma'(x+y)\,\mu(\mathrm{d}y),$$

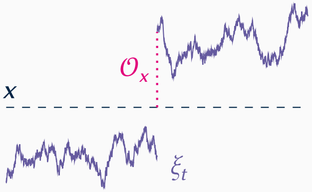
where $\eta = \mathbb{E}^0[\xi_1]$ and

$$\mu(\mathrm{d}y) = \frac{1}{\mathbb{E}^0[\xi_1]}\Big(d_H\delta_0(\mathrm{d}y) + \Pi_H((y,\infty))\,\mathrm{d}y\,\mathbf{1}_{(0,\infty)}(y)\Big), \quad y \geq 0.$$

- Let

$$\mathcal{O}_x = \xi_{T_x} - x, \quad x \geq 0,$$

be the overshoot of ξ over a level x.



- Integration by parts reveals

$$\mathcal{A}_H\gamma(x) = d_H\gamma'(x) + \int_{0+}^{\infty}(\gamma(x+y)-\gamma(x))\,\Pi_H(\mathrm{d}y) = \int_0^{\infty}\eta\gamma'(x+y)\,\mu(\mathrm{d}y),$$

where $\eta = \mathbb{E}^0[\xi_1]$ and

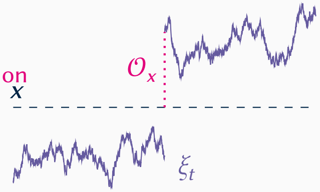
$$\mu(\mathrm{d}y) = \frac{1}{\mathbb{E}^0[\xi_1]}\Big(d_H\delta_0(\mathrm{d}y) + \Pi_H((y,\infty))\,\mathrm{d}y\,\mathbf{1}_{(0,\infty)}(y)\Big), \quad y \geq 0.$$

- Let

$$\mathcal{O}_x = \xi_{T_x} - x, \quad x \geq 0,$$

be the overshoot of ξ over a level x.

- $(\mathcal{O}_x)_{x\geq 0}$ is an $\mathbb{R}_+$-valued Feller process and μ is its invariant distribution



Construction of an estimator

- Integration by parts reveals

$$\mathcal{A}_H\gamma(x) = d_H\gamma'(x) + \int_{0+}^{\infty} (\gamma(x+y) - \gamma(x))\,\Pi_H(\mathrm{d}y) = \int_0^{\infty} \eta\gamma'(x+y)\,\mu(\mathrm{d}y),$$

  where $\eta = \mathbb{E}^0[\xi_1]$ and

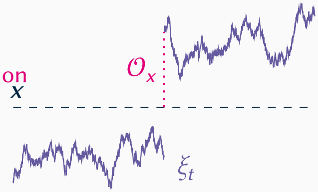
$$\mu(\mathrm{d}y) = \frac{1}{\mathbb{E}^0[\xi_1]}\Big(d_H\delta_0(\mathrm{d}y) + \Pi_H((y,\infty))\,\mathrm{d}y\,\mathbf{1}_{(0,\infty)}(y)\Big), \quad y \geq 0.$$

- Let

$$\mathcal{O}_x = \xi_{T_x} - x, \quad x \geq 0,$$

  be the overshoot of ξ over a level x.
- $(\mathcal{O}_x)_{x\geq 0}$ is an $\mathbb{R}_+$-valued Feller process and μ is its invariant distribution
- natural spatial estimator

$$\tilde{f}_Y(x) := \frac{1}{Y}\int_0^Y \eta\gamma'(x + \mathcal{O}_y)\,\mathrm{d}y,$$



- Integration by parts reveals

$$\mathcal{A}_H\gamma(x) = d_H\gamma'(x) + \int_{0+}^{\infty}(\gamma(x+y)-\gamma(x))\,\Pi_H(\mathrm{d}y) = \int_0^{\infty}\eta\gamma'(x+y)\,\mu(\mathrm{d}y),$$

where $\eta = \mathbb{E}^0[\xi_1]$ and

$$\mu(\mathrm{d}y) = \frac{1}{\mathbb{E}^0[\xi_1]}\Big(d_H\delta_0(\mathrm{d}y) + \Pi_H((y,\infty))\,\mathrm{d}y\,\mathbf{1}_{(0,\infty)}(y)\Big), \quad y \geq 0.$$

- Let

$$\mathcal{O}_x = \xi_{T_x} - x, \quad x \geq 0,$$

be the overshoot of ξ over a level x.
- $(\mathcal{O}_x)_{x\geq 0}$ is an $\mathbb{R}_+$-valued Feller process and μ is its invariant distribution
- natural spatial estimator

$$\tilde{f}_Y(x) := \frac{1}{Y}\int_0^Y \eta\gamma'(x+\mathcal{O}_y)\,\mathrm{d}y,$$

- temporal estimator:

$$\hat{f}_T(x) := \tilde{f}_{\xi_T}(x)\mathbf{1}_{(0,\infty)}(\xi_T) = \frac{1}{\xi_T}\int_0^{\xi_T}\eta\gamma'(x+\mathcal{O}_y)\,\mathrm{d}y\,\mathbf{1}_{(0,\infty)}(\xi_T)$$

Regret bound

- path integrals of the Markov process $(\mathcal{O}_x)_{x\geq 0}$ concentrate nicely if the process is exponentially ergodic, that is,

$$\|\mathbb{P}^y(\mathcal{O}_x \in \cdot) - \mu\|_{\mathrm{TV}} \lesssim V(y)\mathrm{e}^{-\kappa x}$$

- this is demonstrated in Döring and T. (2023)[1] under natural tail and regularity assumptions on Π

Theorem (Christensen, Strauch, T. (2024+))

Assume $\theta^* \in D$ and let $\hat{\theta}_T = \arg\max_{x \in D} \hat{f}_T(x)$. If $\Pi \sim \mathrm{Leb}$ and has an exponential moment, it holds that

$$\mathbb{E}^0\big[f(\theta^*) - f(\hat{\theta}_T)\big] \in \mathrm{O}\Big(\sqrt{\frac{\log T}{T}}\Big).$$

[1] L. Döring and L. Trottner (2023). Stability of overshoots of Markov additive processes. *Ann. Appl. Prob.*

S. Christensen and C. Strauch. **“Nonparametric learning for impulse control problems—Exploration vs. exploitation”.** In: *Ann. Appl. Prob.* 33.2 (2023), pp. 1569 –1587.

S. Christensen, C. Strauch, and L. Trottner. **“Learning to reflect: A unifying approach for data-driven stochastic control strategies”.** In: *Bernoulli* (to appear).

S. Christensen, A. H. Thomsen, and L. Trottner. **Data-driven rules for multidimensional reflection problems.** arXiv:2311.06639. 2023.

L. Döring and L. Trottner. **“Stability of overshoots of Markov additive processes”.** In: *Ann. Appl. Probab.* 33.6B (2023), pp. 5413–5458. ISSN: 1050-5164,2168-8737.

C. Strauch. **“Adaptive invariant density estimation for ergodic diffusions over anisotropic classes”.** In: *Ann. Statist.* 46.6B (2018), pp. 3451–3480. ISSN: 0090-5364.

Summary

- we study singular control problems for ergodic diffusion processes 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
- for diffusions, the exploration-exploitation tradeoff is overcome by separating 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 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
- for diffusions, the exploration-exploitation tradeoff is overcome by separating 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!