

Learning to reflect – On data driven approaches to singular stochastic optimal control

Nonparametric Inference in the Age of AI – Seoul National University

Lukas Trottnner

based on joint works with [Sören Christensen](#), [Asbjørn Holk Thomsen](#) and [Claudia Strauch](#)

25 February 2025

[University of Birmingham](#) [Kiel University](#) [Aarhus University](#) [Heidelberg University](#)



UNIVERSITY OF
BIRMINGHAM

Framework for data-driven stochastic optimal control

- consider a d -dimensional diffusion

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t,$$

- we assume that the drift b is **unknown**
- which 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 (AOAP, 2023); Christensen, Dexheimer, Strauch (2023+))
 2. **reflection controls (singular)** (Christensen, Strauch, T. (Bernoulli, 2024); Christensen, Holk Thomsen, T. (JUQ, 2024))

Framework for data-driven stochastic optimal control

- consider a d -dimensional diffusion

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t,$$

- we assume that the drift b is **unknown**
- which 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 (AOAP, 2023); Christensen, Dexheimer, Strauch (2023+))
 2. **reflection controls (singular)** (Christensen, Strauch, T. (Bernoulli, 2024); Christensen, Holk Thomsen, T. (JUQ, 2024))

Challenge

Exploration vs. exploitation

Reflection control problem

- consider a d -dimensional Langevin diffusion

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dW_t;$$

if ergodic: stationary density $\pi \propto \exp(-V(\cdot))$

- we play the following game:

- the aim is to keep the process close to a target state, say 0, at minimal long run costs
- normally reflect the process in a domain D that we are free to choose:

$$dX_t^D = -\nabla V(X_t^D) dt + \sqrt{2} dW_t + n(X_t^D) dL_t^D, \quad \text{where } L_t^D = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^t \mathbf{1}_{(\partial D)_\varepsilon}(X_s^D) ds$$

- costs:

$$J_T(D) = \underbrace{\int_0^T c(X_t^D) dt}_{c \text{ increasing in } |x|} + \underbrace{\kappa L_T^D}_{\text{reflection costs}}$$

Reflection control problem

- consider a d -dimensional Langevin diffusion

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dW_t;$$

if ergodic: stationary density $\pi \propto \exp(-V(\cdot))$

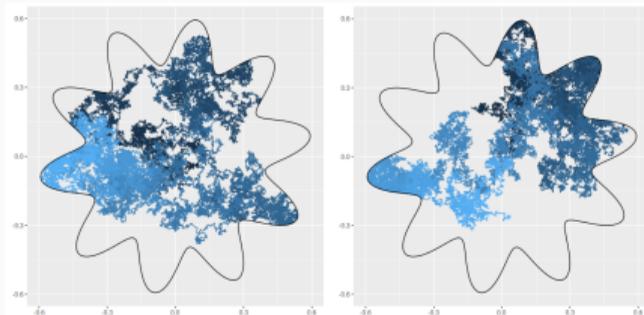
- we play the following game:

- the aim is to keep the process close to a target state, say 0, at minimal long run costs
- normally reflect the process in a domain D that we are free to choose:

$$dX_t^D = -\nabla V(X_t^D) dt + \sqrt{2} dW_t + n(X_t^D) dL_t^D, \quad \text{where } L_t^D = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^t \mathbf{1}_{(\partial D)_\varepsilon}(X_s^D) ds$$

- costs:

$$J_T(D) = \underbrace{\int_0^T c(X_t^D) dt}_{c \text{ increasing in } |x|} + \underbrace{\kappa L_T^D}_{\text{reflection costs}}$$



Reflection control problem

- consider a d -dimensional Langevin diffusion

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dW_t;$$

if ergodic: stationary density $\pi \propto \exp(-V(\cdot))$

- we play the following game:

- the aim is to keep the process close to a target state, say 0, at minimal long run costs
- normally reflect the process in a domain D that we are free to choose:

$$dX_t^D = -\nabla V(X_t^D) dt + \sqrt{2} dW_t + n(X_t^D) dL_t^D, \quad \text{where } L_t^D = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^t \mathbf{1}_{(\partial D)_\varepsilon}(X_s^D) ds$$

- costs:

$$J_T(D) = \underbrace{\int_0^T c(X_t^D) dt}_{c \text{ increasing in } |x|} + \underbrace{\kappa L_T^D}_{\text{reflection costs}}$$

- Ergodic optimal control:** for an admissible domain class Θ determine

$$D^* \in \arg \min_{D \in \Theta} \underbrace{\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}[J_T(D)]}_{=: J(D)} \quad (\rightsquigarrow \text{shape optimisation problem})$$

- Data-driven optimal control:** If V is unknown, determine an estimator \hat{D} of D^* based on observations of the (controlled) process

Ergodic costs

- let D be a class of C^2 -domains such that for any $D \in D$ we have $\inf_{x,y \in \bar{D}} p_1^D(x,y) > 0$ for bicontinuous transition densities p_t^D
- for any $D \in D$, X^D is ergodic with invariant density

$$\pi_D(x) = \frac{\exp(-V(x))}{\int_D \exp(-V(x))} (= \pi(x)/\pi(D) \text{ if free diffusion is ergodic})$$

Theorem

For any $D \in D$, it holds that

$$J(D) = \int_D c(x)\pi_D(x) dx + \kappa \int_{\partial D} \pi_D(x) \mathcal{H}_{d-1}(dx).$$

and

$$\mathbb{E}^x \left[\left| \frac{1}{T} \left(\int_0^T c(X_t^D) dt + \kappa L_T^D \right) - J(D) \right| \right] \lesssim_D \frac{1}{\sqrt{T}}, \quad x \in D.$$

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

$$J(D) = J(D, \pi) = \frac{1}{\int_D \pi(y) dy} \left(\int_D c(y)\pi(y) dy + \kappa \int_{\partial D} \pi(y) \mathcal{H}^{d-1}(dy) \right).$$

Invariant density estimation

Multivariate kernel density estimator:

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

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

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

Learning the optimal reflection boundary

Proposition

Let $\hat{\pi}_T^* := \hat{\pi}_{\mathbf{h}_T, T} \vee \underline{\pi}$, where $\pi \geq \underline{\pi}$ on $B(0, \bar{\lambda})$. Let Θ be a family of domains s.t. $B(0, \underline{\lambda}) \subset D \subset B(0, \bar{\lambda})$ and $\mathcal{H}^{d-1}(\partial D) \leq \Lambda$ for any $D \in \Theta$. For

$$\hat{D}_T \in \arg \min_{D \in \Theta} J(D, \hat{\pi}_T^*),$$

it holds for a warm start μ that

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

Learning the optimal reflection boundary

Proposition

Let $\hat{\pi}_T^* := \hat{\pi}_{\mathbf{h}_T, T} \vee \underline{\pi}$, where $\pi \geq \underline{\pi}$ on $B(0, \bar{\lambda})$. Let Θ be a family of domains s.t. $B(0, \underline{\lambda}) \subset D \subset B(0, \bar{\lambda})$ and $\mathcal{H}^{d-1}(\partial D) \leq \Lambda$ for any $D \in \Theta$. For

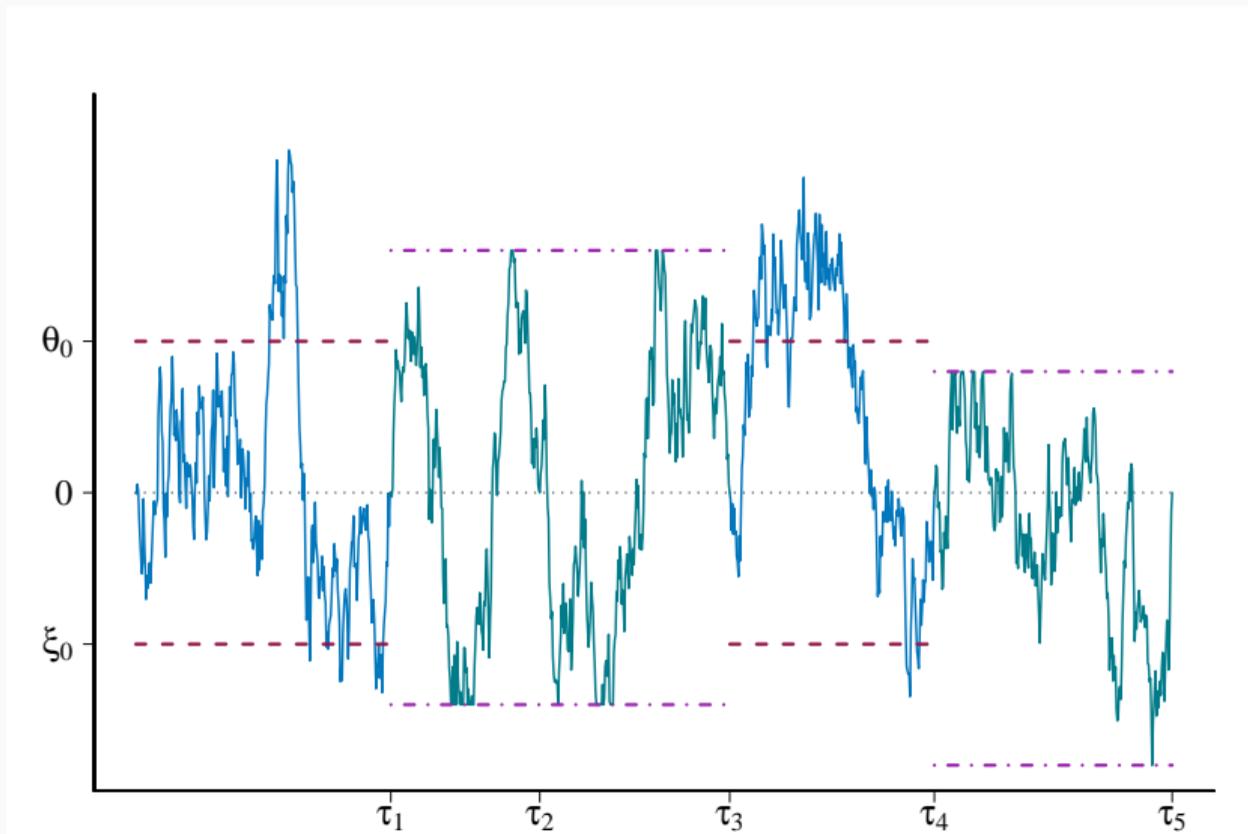
$$\hat{D}_T \in \arg \min_{D \in \Theta} J(D, \hat{\pi}_T^*),$$

it holds for a warm start μ that

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

- ↪ 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?

Episodic domain learning in 1D



Regret bound for episodic domain learning

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

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

$$\frac{1}{T} \mathbb{E} \left[\int_0^T c(X_t^{\hat{Z}}) dt + \kappa L_{\hat{Z}} \right] - J(D^*) \lesssim \begin{cases} \frac{\sqrt{\log T}}{T^{1/3}}, & d = 1, \\ \left(\frac{(\log T)^2}{T} \right)^{\frac{1}{3}}, & d = 2, \\ \left(\frac{\log T}{T} \right)^{\frac{\bar{\beta}}{3\bar{\beta} + d - 2}}, & d \geq 3. \end{cases}$$

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

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

Regret bound for episodic domain learning

Theorem

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

$$\frac{1}{T} \mathbb{E} \left[\int_0^T c(X_t^{\hat{Z}}) dt + \kappa L \frac{\hat{Z}}{T} \right] - J(D^*) \lesssim \begin{cases} \frac{\sqrt{\log T}}{T^{1/3}}, & d = 1, \\ \left(\frac{(\log T)^2}{T} \right)^{\frac{1}{3}}, & d = 2, \\ \left(\frac{\log T}{T} \right)^{\frac{\bar{\beta}}{3\bar{\beta} + d - 2}}, & d \geq 3. \end{cases}$$

- **1D case:** for S_T the (random) exploration time and N_T the number of exploration intervals until time T , choose a strategy such that for some $m, M > 0$,

$$\mathbb{P}(T^{-2/3} S_T \leq m) \lesssim T^{-1/3} \quad \text{and} \quad \limsup_{T \rightarrow \infty} T^{-2/3} \mathbb{E}[N_T] \leq M$$

- if $(c_n)_{n \in \mathbb{N}}$ is a binary sequence with $c_n = 0$ if n -th period is exploration, this is satisfied provided that for some $a > 0$

$$n^{2/3} \leq |\{j \leq n : c_j = 0\}| \leq n^{2/3} + a.$$

Regret bound for episodic domain learning

Theorem

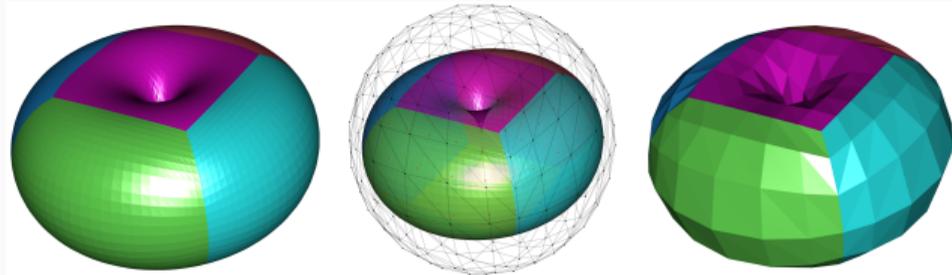
There exists a purely data-driven episodic domain learning strategy \hat{Z} such that the **expected regret per time unit** satisfies

$$\frac{1}{T} \mathbb{E} \left[\int_0^T c(X_t^{\hat{Z}}) dt + \kappa L \hat{Z}_T \right] - J(D^*) \lesssim \begin{cases} \frac{\sqrt{\log T}}{T^{1/3}}, & d = 1, \\ \left(\frac{(\log T)^2}{T} \right)^{\frac{1}{3}}, & d = 2, \\ \left(\frac{\log T}{T} \right)^{\frac{\bar{\beta}}{3\bar{\beta} + d - 2}}, & d \geq 3. \end{cases}$$

- **multivariate case**: X does not hit points for $d \geq 2 \rightsquigarrow$ construction of stochastic exploration/exploitation intervals as in the one-dimensional case not feasible
- instead: alternate between exploration/exploitation intervals with **deterministic** lengths $a_i \asymp 2^i$ and exploitation lengths $b_i \asymp a_i / \Psi_{d, \beta}(a_i)$ (+ asymptotically negligible stochastic fluctuation for exploitation lengths to make sure that the process is inside of proposed reflection domain)
- for technical reasons estimated reflection domain in i -th exploitation interval calculated only from data in i -th exploration interval

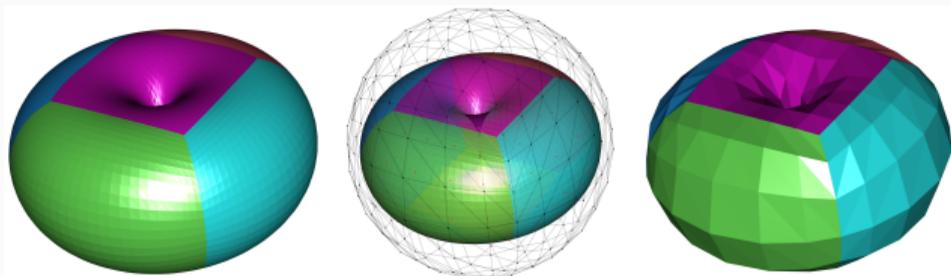
Numerical shape optimisation

- as target domains Θ only allow **strongly star-shaped** sets at 0 (appropriate when continuous costs c are minimised close to the origin) $\rightsquigarrow \partial D = \{r(q)q : q \in S^{d-1}\}$ for some radial function $r : S^{d-1} \rightarrow (0, \infty)$
- for N points $\{q_i\}_{i=1}^N \subset S^{d-1}$ consider the polytope \tilde{D} with vertices $\{p_i\}_{i=1}^N = \{r(q_i)q_i\}_{i=1}^N \rightsquigarrow \tilde{D}$ can be split into N simplices $\{S_l\}_{l \in \mathcal{J}}$ with facets $\{F_l\}_{l \in \mathcal{J}}$ opposite the origin



Numerical shape optimisation

- as target domains Θ only allow **strongly star-shaped** sets at 0 (appropriate when continuous costs c are minimised close to the origin) $\rightsquigarrow \partial D = \{r(q)q : q \in S^{d-1}\}$ for some radial function $r : S^{d-1} \rightarrow (0, \infty)$
- for N points $\{q_i\}_{i=1}^N \subset S^{d-1}$ consider the polytope \tilde{D} with vertices $\{p_i\}_{i=1}^N = \{r(q_i)q_i\}_{i=1}^N \rightsquigarrow \tilde{D}$ can be split into N simplices $\{S_l\}_{l \in \mathcal{J}}$ with facets $\{F_l\}_{l \in \mathcal{J}}$ opposite the origin

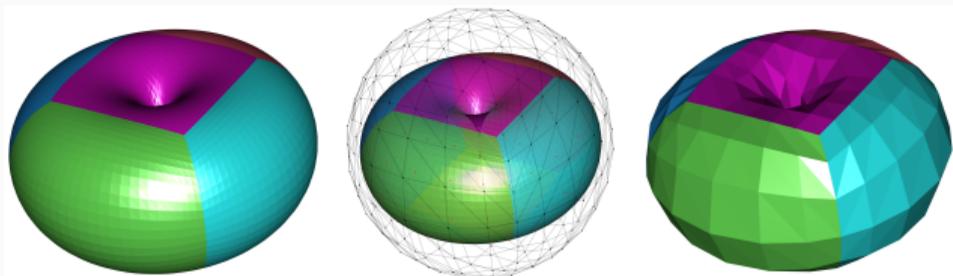


- for $\mathbf{r} = \{r_i\}_{i=1}^N = \{r(q_i)\}_{i=1}^N$ we have

$$J(D) \approx J(\tilde{D}) \equiv J(\mathbf{r}) = \frac{1}{\sum_{l \in \mathcal{J}} \int_{S_l} e^{-V(x)} dx} \sum_{l \in \mathcal{J}} \left(\int_{S_l} c(x) e^{-V(x)} dx + \kappa \int_{F_l} e^{-V(x)} \mathcal{H}^{d-1}(dx) \right)$$

Numerical shape optimisation

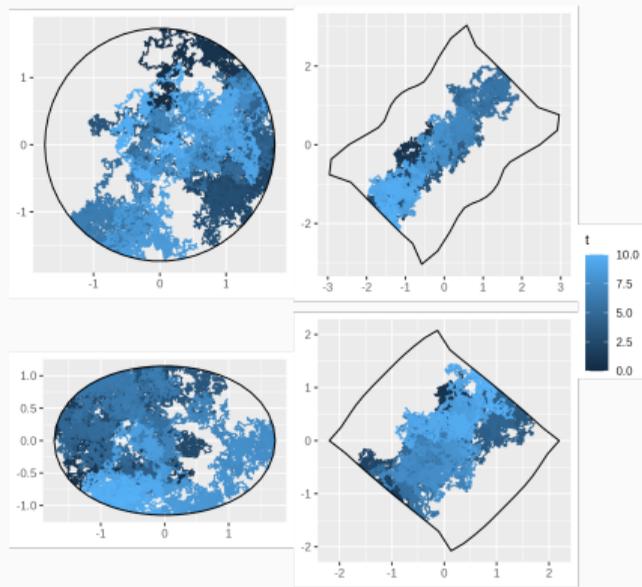
- as target domains Θ only allow **strongly star-shaped** sets at 0 (appropriate when continuous costs c are minimised close to the origin) $\rightsquigarrow \partial D = \{r(q)q : q \in S^{d-1}\}$ for some radial function $r : S^{d-1} \rightarrow (0, \infty)$
- for N points $\{q_i\}_{i=1}^N \subset S^{d-1}$ consider the polytope \tilde{D} with vertices $\{p_i\}_{i=1}^N = \{r(q_i)q_i\}_{i=1}^N \rightsquigarrow \tilde{D}$ can be split into N simplices $\{S_l\}_{l \in \mathcal{J}}$ with facets $\{F_l\}_{l \in \mathcal{J}}$ opposite the origin



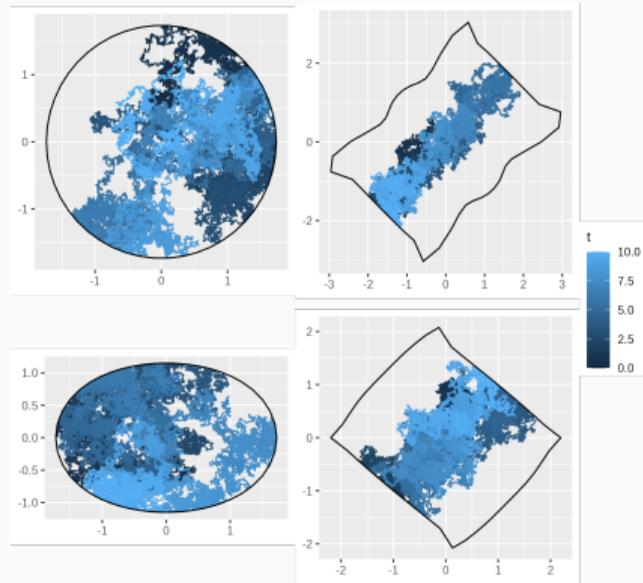
- for $\mathbf{r} = \{r_i\}_{i=1}^N = \{r(q_i)\}_{i=1}^N$ we have

$$J(D) \approx J(\tilde{D}) \equiv J(\mathbf{r}) = \frac{1}{\sum_{l \in \mathcal{J}} \int_{S_l} e^{-V(x)} dx} \sum_{l \in \mathcal{J}} \left(\int_{S_l} c(x) e^{-V(x)} dx + \kappa \int_{F_l} e^{-V(x)} \mathcal{H}^{d-1}(dx) \right)$$

- we derive explicit expressions for $\nabla J(\mathbf{r})$ to employ a **gradient descent algorithm** for shape optimisation

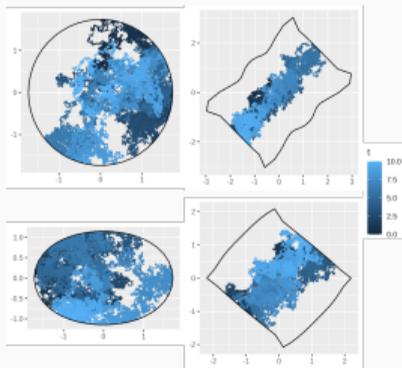


Simulated optimal shapes and corresponding path realizations of reflected processes. Top left: Brownian motion with norm cost. Top right: Ornstein–Uhlenbeck process with norm cost. Bottom left: Brownian motion with skewed cost. Bottom right: Ornstein–Uhlenbeck process with skewed cost.



	Brownian motion	Ornstein–Uhlenbeck
norm cost function	2.22 (2.31)	1.18 (1.15)
skewed cost function	2.83 (2.91)	1.66 (1.74)

Table 1: Average realized costs vs. expected average long term costs (in brackets)



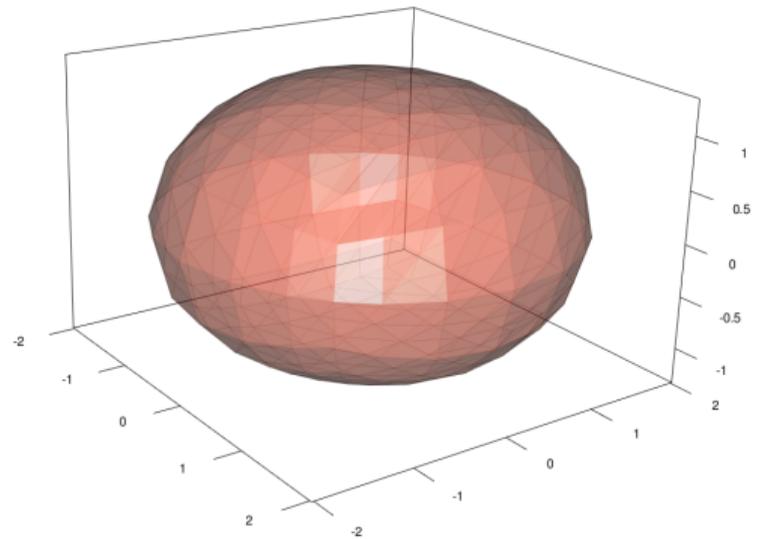
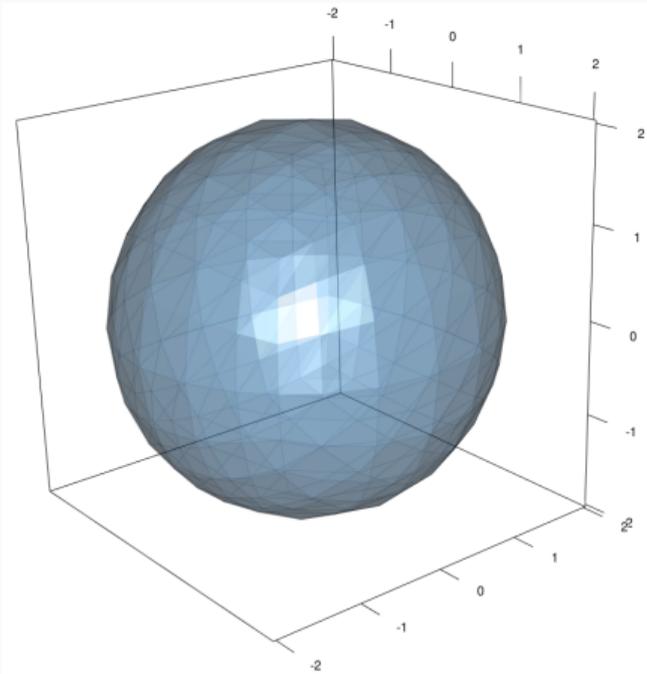
- Simulation of reflected diffusion (Słomiński, SPA 1994): simulate proposal

$$X_{(n+1)\Delta}^{\text{prop}} = X_{n\Delta} - \nabla V(X_{n\Delta})\Delta + \sqrt{2\Delta}\xi_{n+1}, \quad (\xi_i)_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d),$$

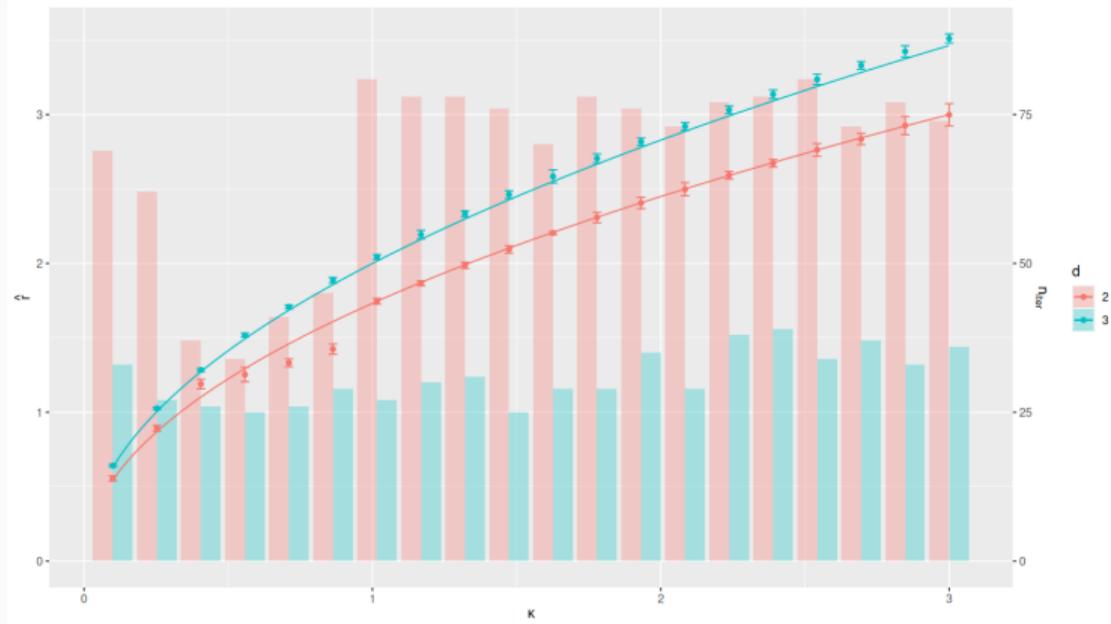
then set

$$X_{(n+1)\Delta} = \text{Proj}_D(X_{(n+1)\Delta}^{\text{prop}}), \quad L_{(n+1)\Delta} = L_{n\Delta} + |X_{(n+1)\Delta}^{\text{prop}} - X_{(n+1)\Delta}|$$

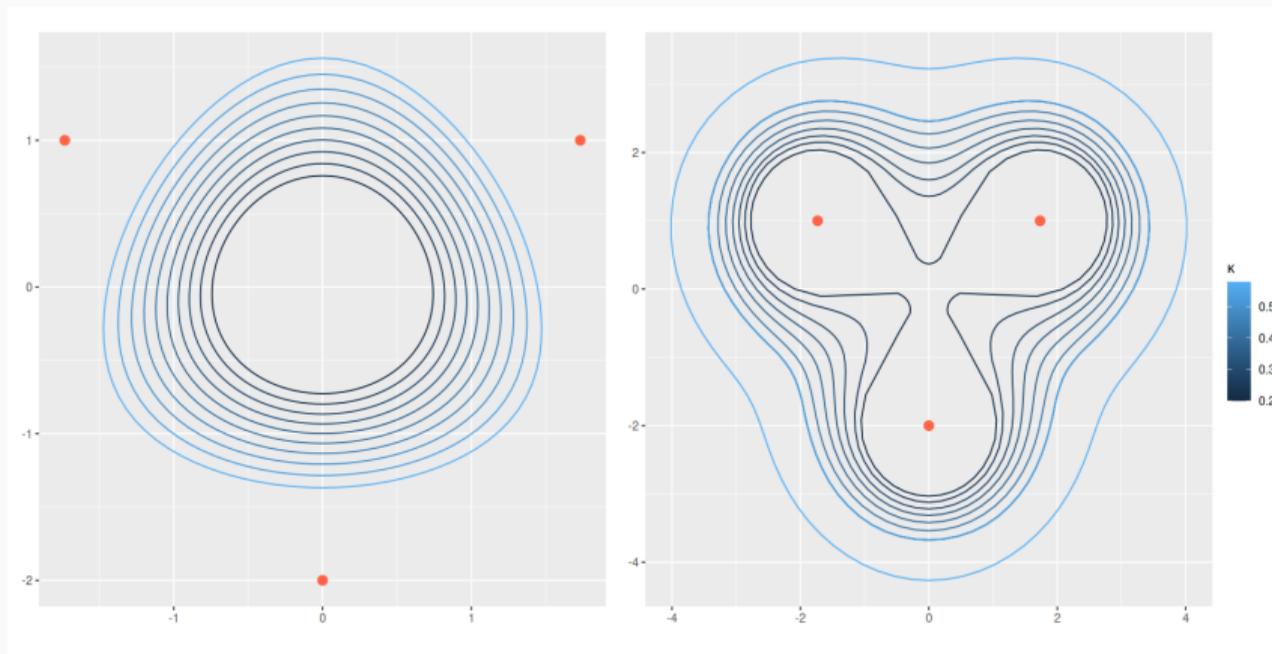
- this works well for polyhedral domains D in low dimensions because projection can be simulated efficiently
- Fishman et al. (NeurIPS, 2023) demonstrate weak convergence of Metropolis approximation and Rejection approximation of reflected Brownian motion
- this is motivated by denoising reflected diffusion models (Lou and Ermon, ICML 2023), see also Holk, Strauch and T. (2024+) for a first statistical analysis



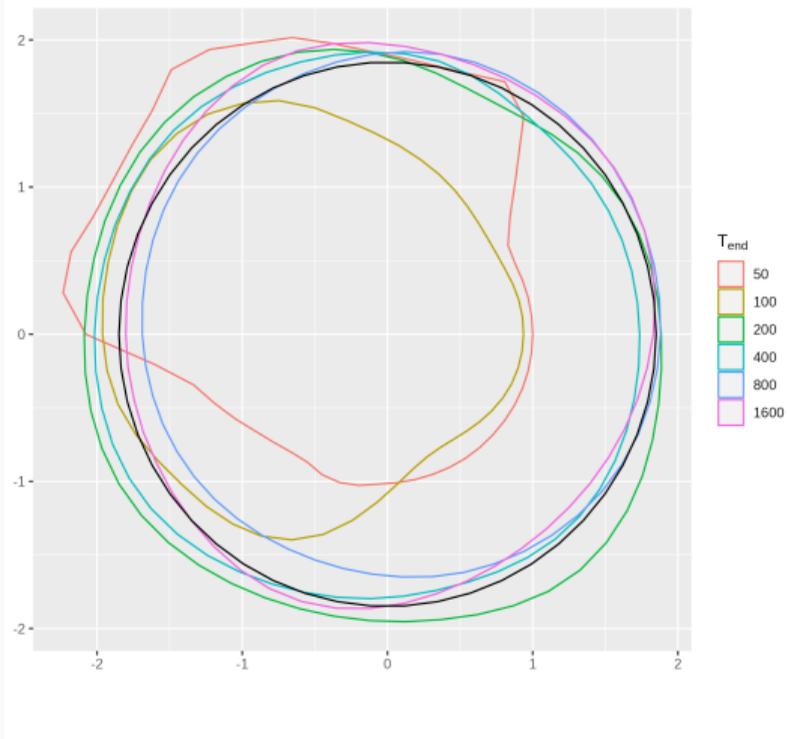
Optimised shapes for Brownian motion with reflection cost $\kappa = 1$ and cost function $c = \|\cdot\|$ (left) and $c(x, y, z) = \sqrt{x^2 + 5y^2 + z^2}$ (right).



For each value of κ , we use the BFGS algorithm (using the built-in R implementation `optim`) to find an approximate optimal shape. To not bias the results towards a ball, we initialize the algorithm with $r_i = 1 + \frac{1}{2} U_i$, where $U_i \sim \text{Unif}[-1, 1]$ for $i = 1, \dots, N$ ($N \approx 200$). Once the approximate optimal values $\hat{r}_1, \hat{r}_2, \dots, \hat{r}_N$ are found, we plot the mean of these along with error bars with height of their standard deviation. For reference we draw a curve of the theoretical optimal radius $r^* = \sqrt{(d+1)\kappa}$. Finally, we also add a bar-plot illustrating the number of iterations of the BFGS algorithm were needed to compute the shapes.



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 $c(x) = \min\{|x - \mu_1|, |x - \mu_2|, |x - \mu_3|\}$.



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.

References

- S. Christensen, N. Dexheimer, and C. Strauch. **Data-driven optimal stopping: A pure exploration analysis**. 2023. arXiv: 2312.05880 [math.ST].
- 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* 30.3 (2024), pp. 2074–2101.
- S. Christensen, A. H. Thomsen, and L. Trottner. **“Data-driven rules for multidimensional reflection problems”**. In: *SIAM/ASA J. Uncertain. Quantif.* 12.4 (2024), pp. 1240–1272.
- N. Fishman et al. **“Metropolis Sampling for Constrained Diffusion Models”**. In: *Advances in Neural Information Processing Systems*. Vol. 36. 2023, pp. 62296–62331.
- A. Holk, C. Strauch, and L. Trottner. **Statistical guarantees for denoising reflected diffusion models**. 2024. arXiv: 2411.01563 [math.ST].
- L. Słomiński. **“On approximation of solutions of multidimensional SDEs with reflecting boundary conditions”**. In: *Stochastic Process. Appl.* 50.2 (1994), pp. 197–219.
- C. Strauch. **“Adaptive invariant density estimation for ergodic diffusions over anisotropic classes”**. In: *Ann. Statist.* 46.6B (2018), pp. 3451–3480.

References

- S. Christensen, N. Dexheimer, and C. Strauch. **Data-driven optimal stopping: A pure exploration analysis**. 2023. arXiv: 2312.05880 [math.ST].
- 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* 30.3 (2024), pp. 2074–2101.
- S. Christensen, A. H. Thomsen, and L. Trottner. **“Data-driven rules for multidimensional reflection problems”**. In: *SIAM/ASA J. Uncertain. Quantif.* 12.4 (2024), pp. 1240–1272.
- N. Fishman et al. **“Metropolis Sampling for Constrained Diffusion Models”**. In: *Advances in Neural Information Processing Systems*. Vol. 36. 2023, pp. 62296–62331.
- A. Holk, C. Strauch, and L. Trottner. **Statistical guarantees for denoising reflected diffusion models**. 2024. arXiv: 2411.01563 [math.ST].
- L. Słomiński. **“On approximation of solutions of multidimensional SDEs with reflecting boundary conditions”**. In: *Stochastic Process. Appl.* 50.2 (1994), pp. 197–219.
- C. Strauch. **“Adaptive invariant density estimation for ergodic diffusions over anisotropic classes”**. In: *Ann. Statist.* 46.6B (2018), pp. 3451–3480.

Thank you for your attention!