

Data-driven rules for multidimensional reflection problems

Sören Christensen*

Asbjørn Holk Thomsen[†]Lukas Trottner[†]

November 14, 2023

Abstract

Over the recent past data-driven algorithms for solving stochastic optimal control problems in face of model uncertainty have become an increasingly active area of research. However, for singular controls and underlying diffusion dynamics the analysis has so far been restricted to the scalar case. In this paper we fill this gap by studying a multivariate singular control problem for reversible diffusions with controls of reflection type. Our contributions are threefold. We first explicitly determine the long-run average costs as a domain-dependent functional, showing that the control problem can be equivalently characterized as a shape optimization problem. For given diffusion dynamics, assuming the optimal domain to be strongly star-shaped, we then propose a gradient descent algorithm based on polytope approximations to numerically determine a cost-minimizing domain. Finally, we investigate data-driven solutions when the diffusion dynamics are unknown to the controller. Using techniques from nonparametric statistics for stochastic processes, we construct an optimal domain estimator, whose static regret is bounded by the minimax optimal estimation rate of the unreflected process' invariant density. In the most challenging situation, when the dynamics must be learned simultaneously to controlling the process, we develop an episodic learning algorithm to overcome the emerging exploration-exploitation dilemma and show that given the static regret as a baseline, the loss in its sublinear regret per time unit is of natural order compared to the one-dimensional case.

2020 MSC: Primary 93E35, 68T05; secondary 49Q10, 60J60, 62M05

key words: Stochastic singular control, reinforcement learning, exploration vs. exploitation, reversible diffusions, shape optimization, nonparametric statistics

1. Introduction and problem formulation

Our underlying processes are Langevin diffusions X on \mathbb{R}^d for $d \geq 2$, which is a well studied class of reversible diffusion processes with drift of potential form. For a C^2 -function $V : \mathbb{R}^d \rightarrow \mathbb{R}$ and a d -dimensional Brownian motion W , X solves the SDE

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

We consider a basic class of stochastic control problems with a clear interpretation: X is interpreted as the position of a particle, which we want to be close to a target state, 0, say. The distance is measured by a locally bounded function $f : \mathbb{R}^d \rightarrow [0, \infty)$ in such a way that $f(X_t)$ stand for the costs associated with being away from the target state. The decision maker can now control the process by choosing a nonempty bounded domain (= open, connected set) $D \subset \mathbb{R}^d$ of

*Kiel University, Department of Mathematics, Heinrich-Hecht-Platz 6, 24118 Kiel, Germany.

Email: christensen@math.uni-kiel.de

[†]Aarhus University, Department of Mathematics, Ny Munkegade 118, 8000 Aarhus C, Denmark.

Email: a.holk@math.au.dk/trottner@math.au.dk

class \mathcal{C}^2 and normally reflecting the process at ∂D . We denote the resulting normally reflected processes by X^D , which is given as the solution to

$$dX_t^D = -\nabla V(X_t^D) dt + \sqrt{2} dW_t + n(X_t^D) dL_t^D,$$

where n is the unit inward normal vector of D at $x \in \partial D$ and L^D denotes the local time of X^D on ∂D , that is, a non-decreasing one-dimensional process with continuous paths that increases only when $X^D \in \partial D$. We assume that controlling the process is associated with costs proportional to L^D (with proportionality factor $\kappa > 0$), so that for each D , the total costs associated with D until time t are

$$\int_0^t f(X_s^D) ds + \kappa L_t^D.$$

Here, we consider a long-term-average criterion and our problem consists of minimizing

$$J(D) := \liminf_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}^x \left[\int_0^t f(X_s^D) ds + \kappa L_t^D \right] \quad (1.1)$$

over all admissible D . The limit is independent of the initial value $x \in D$ due to the ergodic nature of the reflected diffusion, which will be made precise in Theorem 2.4. For known characteristics of the underlying process, the problem we consider here is closely connected to singular stochastic control problems. We discuss this in more detail in Section 1.2 below.

1.1. Contributions

In this paper, we first address the question how the solution to the problem (1.1) can be meaningfully characterized when the characteristics of the underlying process are known. Our main contribution here is Theorem 2.4 which shows under minimal assumptions that $J(D)$ is the L^1 -limit of the average costs and is explicitly given by

$$J(D) = \frac{1}{\int_D e^{-V(y)} dy} \left(\int_D f(y) e^{-V(y)} dy + \kappa \int_{\partial D} e^{-V(y)} \mathcal{H}^{d-1}(dy) \right), \quad (1.2)$$

where \mathcal{H}^{d-1} denotes the $(d-1)$ -dimensional Hausdorff measure. The formula is interesting in that the control problem has been transformed into a shape optimization problem.

In Section 3 we give a numerical approach to minimizing $J(D)$ based on approximating D by polytopes and then applying a gradient descent method. To this end, we derive explicit formulas (Theorem 3.1) under the assumption that D is star-shaped and then illustrate the method on problems with underlying Brownian motions and Ornstein–Uhlenbeck processes.

Finally, in Section 4 we show that the approach we present provides a suitable basis for addressing the problem in the context of model-based reinforcement learning. More precisely, we show that the problem can be solved when the drift function $b = -\nabla V$ is unknown to the decision maker, so that the control has to be purely data driven. Our approach is based on estimating the stationary density of the uncontrolled process ρ nonparametrically and specifying the exact rate with techniques from nonparametric statistics for diffusion processes (Theorem 4.4). We use this via (1.2) to estimate the optimal boundary based on a path of the uncontrolled process and obtain that the resulting static regret has the same sublinear rate (Proposition 4.6). In the more practically relevant situation of simultaneous optimization and data collection, we face an exploration vs. exploitation dilemma. The previous results together with an episodic learning approach lead to a proof of a sublinear regret rate (Theorem 4.7 and Corollary 4.8).

1.2. Related literature

Stochastic singular control problems are a class of stochastic control problems that have been extensively studied, see [23, 40] for textbook treatments. They arise in various applications such as inventory management in operations research, control of queueing networks, portfolio selection with transaction costs in finance, equity issuance in insurance mathematics, position control in engineering, or optimal harvesting in biology. The actions of the agents affect the state and the costs in proportion to the size of the action. This structure implies that the optimal controls usually have the following general form: The space is divided into a region where action is required and a region D where no action is required. The optimal control then reflects the controlled process at the boundary of the no-action region to keep it inside. Therefore, the problem class is closely related to reflected SDEs [35, 43], as observed for a variety of examples, see e.g., [14, 25, 30, 45] (although this connection is difficult to be made precise in full generality, see the discussion in [6]). Thus, stochastic singular control can typically be reduced to finding the optimal no-action region D and the optimal reflection direction at the boundary of D . However, finding the solution is usually only possible explicitly in a few examples, typically with underlying one-dimensional diffusions, see the following section. Characterizing optimal controls becomes much harder when the problem has more than one dimension. Some characterizations of the optimal solutions in special multidimensional cases can be found in [13, 18, 21, 22, 31]. Thus, the problem (1.1) can be viewed as an optimization problem over a class of strategies typically relevant to general singular control problems. We restrict this, however, by the fact that only normal reflections are admissible. This assumption may be justifiable in some cases from the real world problem being modeled. However, we will also see below that in subclasses of problems, such as the radially symmetric case, this is not a restriction at all. In what follows, we will see that the assumption of normal reflection simplifies the problem to the point where a deeper analysis can be performed.

There is a fair amount of literature on the numerical treatment of singular control problems. The methods range from approaches based on a discretization with discrete Markov chains [33] over an approximation of the solution of the corresponding Hamilton–Jacobi–Bellman equation [32], approaches using linear programming [49] up to finite element approximations [48]. However, all methods have limited applicability, especially in higher dimensions, and care must be taken in the exact implementation. In this paper, we present a new approach that is structurally different from the existing ones.

As described above, the approach we have taken allows for data-driven control when the drift of the process is unknown. This question falls into the currently fast growing field of model based reinforcement learning (RL), where the agent does not know the system parameters and learns them by interacting with the environment and getting feedback. The agent chooses policies based on the current parameter estimation and tries to minimize the regret, which is the gap between the expected reward of the best policy and the actual reward achieved. Many discrete-time RL problems have been studied, where sublinear regret bounds have been obtained for different scenarios, such as bandit problems, tabular Markov decision problems, and linear quadratic (LQ) problems [15, 26, 41]. However, for continuous-time RL problems, it is well known that time discretization does not work satisfactorily in standard approaches such as Deep-Q learning [47]. Therefore, the concrete model must be included here. Most of the previous works only propose algorithms, and only a few analyze their regrets, mostly for LQ problems. In particular, [20] proved an asymptotic sublinear regret for regularized least-squares algorithms in an ergodic continuous-time LQ problem, but without giving the exact order of the bound. Recently, [5, 24] generalized the least-squares algorithms to finite-time horizon episodic settings and gave non-asymptotic regret bounds. These works assume a parametric structure of the problem. On the other hand, [39] considered propagator models and combined exploration and exploitation

schemes to achieve sublinear regrets with high probability. Closest to this paper are the articles [11, 12], in which singular and impulse control problems with a nonparametric statistical diffusion structure are considered. However, these papers make heavy use of the one-dimensional structure of the underlying processes. This leads to the strategy already being described by one or two values, so only these need to be learned. One of the main contributions of this work is that we can give exact sublinear rates of regression even for the case where the optimal strategy is only infinite-dimensional parameterizable. To the best of our knowledge, this work is the first to provide such results in this context.

2. Optimal reflection as a shape optimization problem for known characteristics

As detailed above, for our approach it is central to find explicit expressions for $J(D)$ for known drift. Before we get to that, we start here with a brief discussion of the one-dimensional special case, since it serves as the main motivation. The main observation for obtaining explicit solutions for all singular control problems for underlying linear diffusions is that the values $J(D)$ can be found (semi-)explicitly in terms of speed measure and scale function. In this linear case, solving (1.1) therefore boils down to a standard optimization problem, see [1, 10] and the references therein. In the one dimensional ergodic case discussed here, this is realized by the fact that the stationary density of the reflected diffusion is just the conditional density of the uncontrolled diffusion. The key observation in this section is that this also holds in the multivariate case with underlying Langevin diffusions. This turns out to be key to establishing (1.2).

2.1. Ergodicity of the reflected Langevin diffusions

It is well-known that for an (uncontrolled) Langevin diffusion as introduced above, if e^{-V} is integrable, then X has a stationary density given by

$$\rho_{\mathbb{R}^d}(x) := \rho(x) := c^{-1} \exp(-V(x)), \text{ where } c = c_{\mathbb{R}^d}(V) = \int_{\mathbb{R}^d} e^{-V(u)} du.$$

For general diffusions, knowing the distribution on the whole state space does not give any information about the stationary distributions of the corresponding diffusions with reflection in a subdomain D . A main observation for our approach is that for Langevin diffusions, this is different. Indeed, in this particular situation one can obtain the stationary distribution on D by conditioning:

Lemma 2.1. *For a bounded domain $D \subset \mathbb{R}^d$ of class \mathcal{C}^2 , the density ρ_D given by*

$$\rho_D(x) = c_D^{-1} \exp(-V(x)), \quad x \in D, \text{ where } c_D = \int_D e^{-V(u)} du,$$

is a stationary density of the normally reflected process X^D .

This result is well known. For the convenience of the reader, a proof is provided in Appendix A.

It will be important for our purposes to have sufficiently fast convergence of the reflected diffusion to equilibrium. To this end, we will assume that $D \subset \mathbb{R}^d$ is a bounded domain of class \mathcal{C}^2 that is sufficiently nice to guarantee that the Markov process $(X^D, (\mathbb{P}^x)_{x \in \bar{D}})$ has transition densities $(p_t^D(x, y))_{t > 0, x, y \in \bar{D}}$ —that is, $\mathbb{E}^x[f(X_t^D)] = \int_{\bar{D}} p_t^D(x, y) f(y) dy$ for any bounded measurable function $f : \bar{D} \rightarrow \mathbb{R}$ and $x \in \bar{D}$ —such that for any $t > 0$, p_t^D is continuous on $\bar{D} \times \bar{D}$ and we have the minorization property

$$\inf_{x, y \in \bar{D}} p_1^D(x, y) \geq \delta, \tag{2.1}$$

for some $\delta > 0$, where necessarily $\delta\lambda(D) \in (0, 1)$. Let us denote by \mathbf{D} the class of bounded domains $D \subset \mathbb{R}^d$ of class \mathcal{C}^2 such that these assumptions hold for the process reflected in ∂D . These can be verified under mild assumptions on the boundary ∂D by general results on fundamental solutions of parabolic PDEs with Neumann boundary conditions. For instance, continuity of the transition densities and (2.1) are ensured whenever ∂D is the union of a finite number of hypersurfaces of class \mathcal{C}^3 , cf. [27, p.166]. These assumptions now guarantee uniqueness of the invariant distribution and exponential ergodicity of the process.

Lemma 2.2. *Let $D \in \mathbf{D}$ with transition minorization as in (2.1). Then, the unique stationary distribution π_D of the normally reflected diffusion X^D is given by*

$$\pi_D(dx) = \rho_D(x) dx, \quad x \in \bar{D},$$

and X^D is uniformly ergodic, satisfying the bound

$$\|P_t^D(x, \cdot) - \pi_D\|_{\text{TV}} \leq \frac{2}{1-\delta\lambda(D)} e^{\log(1-\delta\lambda(D))t}, \quad x \in \bar{D}, \quad t > 0.$$

The proof is deferred to Appendix A. As a consequence we have the following rate in the ergodic theorem.

Corollary 2.3. *Let $D \in \mathbf{D}$. There exists a constant $C = C(D) > 0$ such that for any $h \in L^\infty(\bar{D})$ and $x \in \bar{D}$ it holds*

$$\frac{1}{t} \mathbb{E}^x \left[\left| \int_0^t (h(X_s^D) - \pi_D(h)) ds \right| \right] \leq \frac{C \|h\|_{L^\infty(\bar{D})}}{\sqrt{t}}.$$

Again, the proof can be found in Appendix A.

2.2. Solution of the ergodic control problem

Given the ergodicity assumptions from the previous subsection we can now fully characterize the ergodic average expected costs $J(D)$ from (1.1) in terms of the invariant distribution of the reflected diffusion X^D . In fact, we will show more: we prove that the bias of the average costs vanishes linearly in time and that their stochastic fluctuation measured in terms of the L^1 -deviation from $J(D)$ vanishes at square-root rate.

Theorem 2.4. *Let $D \in \mathbf{D}$. Then, there exist constants $C(D), C'(D) > 0$ that depend on D but are independent of $x \in \bar{D}$ and $t \geq 1$ such that*

$$\mathbb{E}^x \left[\left| \frac{1}{t} \left(\int_0^t f(X_s^D) ds + \kappa L_t^D \right) - \left(\int_D f(y) \rho_D(y) dy + \kappa \int_{\partial D} \rho_D(y) \mathcal{H}^{d-1}(dy) \right) \right| \right] \leq \frac{C(D)}{\sqrt{t}}, \quad (2.2)$$

and

$$\left| \mathbb{E}^x \left[\frac{1}{t} \left(\int_0^t f(X_s^D) ds + \kappa L_t^D \right) \right] - \left(\int_D f(y) \rho_D(y) dy + \kappa \int_{\partial D} \rho_D(y) \mathcal{H}^{d-1}(dy) \right) \right| \leq \frac{C'(D)}{t}, \quad (2.3)$$

where \mathcal{H}^{d-1} denotes the $(d-1)$ -dimensional Hausdorff measure. In particular,

$$J(D) = \int_D f(y) \rho_D(y) dy + \kappa \int_{\partial D} \rho_D(y) \mathcal{H}^{d-1}(dy).$$

We will need the following basic result for the proof, which we include here for the lack of a precise reference.

Lemma 2.5. *Let $O \subset \mathbb{R}^d$ be a bounded open set of class \mathcal{C}^k for some $k \geq 2$ and let n be the unit inward normal vector on ∂O . Then there exists a function $\varphi \in \mathcal{C}^k(\mathbb{R}^d)$ such that $\nabla \varphi = n$ on ∂O .*

Proof. For a set $A \subset \mathbb{R}^d$ let $d(x, A) := \inf\{|x - y| : y \in A\}$ be the distance function to A and denote by $A^\varepsilon := \{y \in \mathbb{R}^d : |x - y| < \varepsilon \text{ for some } x \in A\}$ its ε -fattening for $\varepsilon > 0$. Let

$$s(x) := \begin{cases} d(x, \partial O), & x \in O, \\ -d(x, \partial O), & x \in \mathbb{R}^d \setminus O, \end{cases}$$

be the signed distance to the boundary ∂O . By [29, Theorem 3] there exists an open neighborhood U of ∂O such that $s \in \mathcal{C}^k(U)$, and in particular ∇s is well-defined on ∂O . Moreover, we have $\nabla s = n$ on ∂O . Indeed, for given $y \in \partial O$, by the above there exists $r > 0$ such that $s|_{B(y,r)} : B(y,r) \rightarrow \mathbb{R}$ is a \mathcal{C}^k function and clearly $s|_{B(y,r)}^{-1}(\{0\}) = \partial O \cap B(y,r)$. By the smoothness of ∂O we have $s(y + hn(y)) = h$ for $h \geq 0$ small enough, whence

$$\langle \nabla s(y), n(y) \rangle = \nabla_{n(y)} s(y) = \lim_{h \rightarrow 0} \frac{s(y + hn(y)) - s(y)}{h} = 1, \quad (2.4)$$

and therefore $\nabla s(y) \neq 0$. The usual argument via the implicit function theorem now shows that $\nabla s(y)$ is orthogonal to the hyperplane tangent to ∂O at y . That is, $\nabla s(y) \propto n(y)$ and we conclude from (2.4) that indeed $\nabla s(y) = n(y)$. Now, since $s \in \mathcal{C}^k(U)$, we must have $s \in \mathcal{C}^k((\partial O)^\varepsilon)$ for some $\varepsilon > 0$, and since $\text{cl}((\partial O)^\varepsilon/2) \subset (\partial O)^\varepsilon$ is compact, it follows from the classical Whitney extension theorem [50] that there exists a function $\varphi \in \mathcal{C}^k(\mathbb{R}^d)$ such that $\varphi = s$ on $\text{cl}((\partial O)^\varepsilon/2)$. In particular, on ∂O it holds $\nabla \varphi = \nabla s = n$, and we conclude that φ has the desired properties. ■

Proof of Theorem 2.4. We only prove the claim on $L^1(\mathbb{P}^x)$ -convergence at square root rate in (2.2), the statement on convergence in expectation at linear rate in (2.3) follows from similar, but easier considerations. We first have as in the proof of Lemma 2.1, letting A be the differential operator from (A.1),

$$\int_D Af(x) \pi_D(dx) = - \int_{\partial D} \langle \nabla f(x), n(x) \rangle \rho_D(x) \mathcal{H}^{d-1}(dx), \quad f \in \mathcal{C}^2(\mathbb{R}^d).$$

In particular, since D is of class \mathcal{C}^2 , by Lemma 2.5 we may choose $\varphi \in \mathcal{C}^2(\mathbb{R}^d)$ such that $\nabla \varphi(x) = n(x)$ for $x \in \partial D$, whereby we find

$$\int_D A\varphi(x) \pi_D(dx) = - \int_{\partial D} \rho_D(x) \mathcal{H}^{d-1}(dx).$$

Now by Itô's formula for $t \geq 0$ almost surely,

$$\begin{aligned} \varphi(X_t^D) - \varphi(X_0^D) &= \int_0^t A\varphi(X_s^D) ds + \int_0^t \langle \nabla \varphi(X_s^D), n(X_s^D) \rangle dL_s^D + \int_0^t \langle \nabla \varphi(X_s^D), dW_s \rangle \\ &= \int_0^t A\varphi(X_s^D) ds + L_t^D + \int_0^t \langle \nabla \varphi(X_s^D), dW_s \rangle, \end{aligned}$$

where we used

$$\int_0^t \langle \nabla \varphi(X_s^D), n(X_s^D) \rangle dL_s^D = \int_{\{s \leq t : X_s^D \in \partial D\}} \langle \nabla \varphi(X_s^D), n(X_s^D) \rangle dL_s^D = \int_{\{s \leq t : X_s^D \in \partial D\}} dL_s^D = L_t^D.$$

Note also that since $\nabla \varphi$ is continuous, it is bounded on D , and hence $\int_0^t \langle \nabla \varphi(X_s^D), dW_s \rangle$ is a L^2 -martingale. Combining these, we find for any $x \in \bar{D}$,

$$\mathbb{E}^x \left[\left| \frac{1}{t} L_t^D - \int_{\partial D} \rho_D(x) \mathcal{H}^{d-1}(dx) \right| \right]$$

$$\begin{aligned}
&= \mathbb{E}^x \left[\left| \frac{1}{t} \left(\varphi(X_t^D) - \varphi(X_0^D) - \int_0^t (A\varphi(X_s^D) - \pi_D(A\varphi)) \, ds - \int_0^t \langle \nabla \varphi(X_s^D), dW_s \rangle \right) \right| \right] \\
&\leq 2 \frac{\|\varphi\|_{L^\infty(D)}}{t} + \mathbb{E}^x \left[\left| \frac{1}{t} \int_0^t (A\varphi(X_s^D) - \pi_D(A\varphi)) \, ds \right| \right] + \frac{1}{t} \mathbb{E}^x \left[\left| \int_0^t \langle \nabla \varphi(X_s^D), dW_s \rangle \right| \right] \\
&\leq \frac{\|\varphi\|_{L^\infty(D)} + \|A\varphi\|_{L^\infty(D)}}{\sqrt{t}} + \frac{1}{t} \mathbb{E}^x \left[\left| \int_0^t \langle \nabla \varphi(X_s^D), dW_s \rangle \right| \right] \\
&\leq \frac{\|\varphi\|_{L^\infty(D)} + \|A\varphi\|_{L^\infty(D)}}{\sqrt{t}} + \frac{1}{t} \left(\mathbb{E}^x \left[\int_0^t |\nabla \varphi(X_s^D)|^2 \, ds \right] \right)^{1/2} \\
&\leq \frac{\|\varphi\|_{L^\infty(D)} + \|A\varphi\|_{L^\infty(D)} + \|\nabla \varphi\|_{L^\infty(D)}}{\sqrt{t}}
\end{aligned}$$

where we used Corollary 2.3 for the second inequality and Hölder inequality together with Itô-isometry for the third inequality. Using this, another application of Corollary 2.3 for the continuous cost component yields

$$\mathbb{E}^x \left[\left| \frac{1}{t} \left(\int_0^t f(X_s^D) \, ds + \kappa L_t^D \right) - \left(\int_D f(y) \rho_D(y) \, dy + \kappa \int_{\partial D} \rho_D(y) \mathcal{H}^{d-1}(dy) \right) \right| \right] \leq \frac{C}{\sqrt{t}},$$

as claimed. \blacksquare

Putting pieces together, we obtain a formula for $J(D)$ which is just based on the stationary density ρ of the uncontrolled process if the latter is ergodic.

Corollary 2.6. *For any $D \in \mathbf{D}$, it holds that*

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

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

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

We point out that this result is a multidimensional version of [1, Lemma 2.1].

3. Numerical optimization

Corollary 2.6 shows that, for known $\kappa > 0, V$ and f , our problem for known dynamics boils down to minimizing the functional

$$J : D \mapsto \frac{1}{\int_D e^{-V(x)} \, dx} \left(\int_D f(x) e^{-V(x)} \, dx + \kappa \int_{\partial D} e^{-V(x)} \mathcal{H}^{d-1}(dx) \right)$$

over a suitable set of bounded domains $D \subset \mathbb{R}^d$. We are therefore faced with a shape optimization problem. For a general overview on shape optimization we refer to [16]. To approach our particular problem numerically, we restrict ourselves to bounded domains that are *strongly starshaped at 0*. Specifically, we assume that for any D the boundary is given by

$$\partial D = \{r(q)q : q \in S^{d-1}\},$$

for some suitably smooth radial function $r : S^{d-1} \rightarrow (0, \infty)$ on the d -sphere S^{d-1} . Rather than optimizing over all such functions, we discretize the problem by considering $N \in \mathbb{N}$ points placed uniformly (in a suitable sense) on the sphere, say $\{q_i\}_{i=1}^N \subset S^{d-1}$, and then approximating

any star-shaped set D by the polytope \tilde{D} with vertices $\{p_i\}_{i=1}^N := \{r(q_i)q_i\}_{i=1}^N$. The polytope \tilde{D} has N facets and hence can naturally be split into N simplices, each with a vertex at the origin and the remaining vertices given by $p_{i_1}, p_{i_2}, \dots, p_{i_d}$ for some appropriate set of indices i_1, i_2, \dots, i_d . Denote for such a set of indices $I = \{i_j\}_{j=1}^d$ by F_I the facet of the simplex opposite the origin, and by S_I the simplex itself. Letting \mathcal{I} be the family of such sets of indices, we arrive at the following approximation of the objective function

$$J(D) \approx J(\tilde{D}) = \frac{1}{\sum_{I \in \mathcal{I}} \int_{S_I} e^{-V(x)} dx} \sum_{I \in \mathcal{I}} \left(\int_{S_I} f(x) e^{-V(x)} dx + \kappa \int_{F_I} e^{-V(x)} \mathcal{H}^{d-1}(dx) \right). \quad (3.1)$$

Note that this approximation depends on r (and hence on D) only through the N lengths $\{r_i\}_{i=1}^N := \{r(q_i)\}_{i=1}^N$, and hence we may consider J as a function of N variables, $J(\mathbf{r}) := J(\tilde{D})$. Gradient based optimization schemes can now be used, since

$$\frac{\partial J(\mathbf{r})}{\partial r_i} = \frac{1}{C} \sum_{I \in \mathcal{I}_i} \left(\frac{\partial}{\partial r_i} \left(\int_{S_I} f(x) e^{-V(x)} dx + \kappa \int_{F_I} e^{-V(x)} \mathcal{H}^{d-1}(dx) \right) - J(\mathbf{r}) \frac{\partial}{\partial r_i} \left(\int_{S_I} e^{-V(x)} dx \right) \right), \quad (3.2)$$

where $C = \sum_{I \in \mathcal{I}} \int_{S_I} e^{-V(x)} dx$, and $\mathcal{I}_i = \{I \in \mathcal{I} : i \in I\}$. To evaluate these expressions, we use the following theorem.

Theorem 3.1. *Let $q_1, \dots, q_d \in S^{d-1}$ and $r_1, \dots, r_d > 0$ be given and let $p_i = r_i q_i$ for $i = 1, \dots, d$. Denote by S the simplex in \mathbb{R}^d spanned by the origin and the points p_1, \dots, p_d and by F the facet of S opposite the origin. Finally, let P denote the $d \times d$ matrix whose i 'th column is p_i and P_{-1} the $d \times (d-1)$ matrix whose i 'th column is $p_{i+1} - p_1$. Then we have the following for $g \in C(\mathbb{R}^d, \mathbb{R})$:*

$$\int_S g(x) dx = |P| \int_0^1 \int_{(0,1)^{d-1}} g(r\eta(\mathbf{t})) \psi(\mathbf{t}) r^{d-1} dt dr \quad (3.3)$$

$$\int_F g(x) \mathcal{H}^{d-1}(dx) = \sqrt{|P_{-1}^T P_{-1}|} \int_{(0,1)^{d-1}} g(\eta(\mathbf{t})) \psi(\mathbf{t}) dt \quad (3.4)$$

where $\psi(t_1, \dots, t_{d-1}) = \prod_{i=1}^{d-2} t_i^{d-1-i}$, and

$$\eta(t_1, \dots, t_{d-1}) = (1 - t_1)p_1 + t_1(1 - t_2)p_2 + \dots + \left(\prod_{i=1}^{d-2} t_i \right) (1 - t_{d-1})p_{d-1} + \left(\prod_{i=1}^{d-1} t_i \right) p_d.$$

Furthermore, for $i = 1, \dots, d$, we have

$$\frac{\partial}{\partial r_i} \int_S g(x) dx = \frac{1}{r_i} |P| \int_{(0,1)^{d-1}} g(\eta_i(\mathbf{t})) \hat{\psi}(\mathbf{t}) dt, \quad (3.5)$$

where $\hat{\psi}(t_1, \dots, t_{d-1}) = (1 - t_1)\psi(t_1, \dots, t_{d-1})$, and η_i denotes η after swapping p_1 and p_i . Finally, if also $g \in C^1(\mathbb{R}^d, \mathbb{R})$,

$$\begin{aligned} \frac{\partial}{\partial r_i} \int_F g(x) \mathcal{H}^{d-1}(dx) &= \frac{1}{2} \text{tr} \left((P_{-1}^T P_{-1})^{-1} \frac{\partial}{\partial r_i} (P_{-1}^T P_{-1}) \right) \int_{(0,1)^{d-1}} g(\eta(\mathbf{t})) \psi(\mathbf{t}) dt \\ &\quad + \sqrt{|P_{-1}^T P_{-1}|} \int_{(0,1)^{d-1}} \langle \nabla g(\eta(\mathbf{t})), q_i \rangle \hat{\psi}(\mathbf{t}) dt. \end{aligned} \quad (3.6)$$

Proof. We first note that η is simply repeated linear interpolation. That is, if $L(t; x, y) := x + t(y - x)$ for $t \in (0, 1)$ and $x, y \in \mathbb{R}^d$, we have

$$\eta(t_1, \dots, t_{d-1}) = L\left(t_1; p_1, L\left(t_2; p_2, L\left(t_3; p_3 \dots\right)\right)\right).$$

As such, for any point $x \in \text{int } F$, there exists a unique $\mathbf{t} \in (0, 1)^d$ such that $\eta(\mathbf{t}) = x$, and similarly for any $y \in \text{int } S$, a unique $r \in (0, 1)$ such that $r\eta(\mathbf{t}) = y$. As such, to verify (3.3) and (3.4), we only need to show that the functions

$$(r, \mathbf{t}) \mapsto |P|\psi(\mathbf{t})r^{d-1}, \quad \text{and} \quad \mathbf{t} \mapsto \sqrt{|P_{-1}^T P_{-1}|}\psi(\mathbf{t}),$$

denote the Jacobian and Gramian of $(r, \mathbf{t}) \mapsto r\eta(\mathbf{t})$ and η , respectively. To this end, we introduce the following notation: let $t_d \equiv 0$, $T_1 = 1$ and set $T_i = \prod_{j=1}^{i-1} t_j$ for $i = 2, \dots, d$. Then we may write $\eta(t_1, \dots, t_{d-1}) = \sum_{i=1}^d T_i(1 - t_i)p_i$, whereby

$$\frac{\partial}{\partial t_i} r\eta(t_1, \dots, t_{d-1}) = r \left(\frac{\eta(t_1, \dots, t_{d-1}) - \sum_{j=1}^{i-1} T_j(1 - t_j)p_j - T_i p_i}{t_i} \right), \quad i = 1, \dots, d-1,$$

while of course $\frac{\partial}{\partial r} r\eta = \eta$. From this it follows by determinant properties that the Jacobian of $(r, \mathbf{t}) \mapsto r\eta(\mathbf{t})$ is given by

$$|\mathcal{J}(r, \mathbf{t})| = r^{d-1} \left(\prod_{i=1}^{d-1} \frac{1}{t_i} \right) \left(\prod_{i=1}^d T_i \right) |P| = |P|\psi(\mathbf{t})r^{d-1},$$

showing (3.3) as desired.

To show (3.4), we note that since F lies in a $d-1$ -dimensional hyperplane, say H , we may embed it in \mathbb{R}^{d-1} by an isometry $\Psi : H \rightarrow \mathbb{R}^{d-1}$ with $\Psi(p_1) = 0$. Specifically, Ψ can be constructed as $x \mapsto \Psi'(A(x - p_1))$, where A is the rotation matrix such that the d 'th coordinate of Ax is 0 for all $x \in H$, and $\Psi' : \mathbb{R}^d \rightarrow \mathbb{R}^{d-1}$ simply discards the last coordinate. Then, integrating a function g over F with respect to \mathcal{H}^{d-1} is equivalent to integrating $g \circ \Psi^{-1}$ over $\Psi(F)$ with respect to the $d-1$ -dimensional Lebesgue measure. Now, since $\Psi(F)$ by the above construction of Ψ is a simplex in \mathbb{R}^{d-1} consisting of the origin and $d-1$ other points, say p'_1, \dots, p'_{d-1} , it follows by the above,

$$\int_{F'} g \circ \Psi^{-1}(x) dx = |P'| \int_0^1 \int_{(0,1)^{d-2}} g \circ \Psi^{-1}(r\eta'(\mathbf{t})) \psi'(\mathbf{t}) r^{d-1} dt dr,$$

where η' similarly is linear interpolation between p'_1, \dots, p'_{d-1} , ψ' is the $d-1$ -dimensional equivalent of ψ and P' is the $(d-1) \times (d-1)$ matrix whose i 'th column is p'_i . By some elementary substitutions and renaming of variables, we may write in an abuse of notation $\Psi^{-1}(r\eta'(\mathbf{t})) = \eta(\mathbf{t})$ and $\psi'(\mathbf{t})r^{d-1} = \psi(\mathbf{t})$. Finally, to find $|P'|$, we see

$$|P'| = (d-1)! \text{Vol}_{d-1} F' = (d-1)! \text{Vol}_{d-1} F = \sqrt{|P_{-1}^T P_{-1}|},$$

where we use that Ψ is an isometry and hence preserves volumes.

To show (3.5), let $i \in \{1, \dots, d\}$ be fixed and consider now for some small h the simplex S' with a vertex at the origin and at the points $p_1, \dots, p_{i-1}, \frac{r_i+h}{r_i} p_i, p_{i+1}, \dots, p_d$. Since either $S \subset S'$ or $S' \subset S$, the symmetric difference $S \triangle S'$ is another simplex with vertices at p_1, \dots, p_d and $\frac{r_i+h}{r_i} p_i$. Assume without loss of generality that $S \subset S'$. Shifting the coordinate system so that p_i lies at the origin, we get a simplex with a vertex at the origin and at the points $p_1 - p_i, p_2 - p_i, \dots, \frac{h}{r_i} p_i, \dots, p_d - p_i$. Note that by properties of the determinant, we get

$$\left| \begin{bmatrix} p_1 - p_i & p_2 - p_i & \cdots & \frac{h}{r_i} p_i & \cdots & p_d - p_i \end{bmatrix} \right| = \frac{h}{r_i} |P|.$$

Using this and (3.3), we find that

$$\int_{S'} g(x) dx - \int_S g(x) dx = \int_{S' \setminus S} g(x) dx = \frac{h}{r_i} |P| \int_0^1 \int_{(0,1)^{d-1}} g(r\eta_{i,h}(\mathbf{t}) + p_i) r^{d-1} \psi(\mathbf{t}) dt dr,$$

where, using the same notation as earlier,

$$\eta_{i,h}(t_1, \dots, t_{d-1}) = T_i(1-t_i) \frac{h}{r_i} p_i + \sum_{j \neq i} T_j(1-t_j)(p_j - p_i).$$

Dividing by h and letting $h \rightarrow 0$ (implicitly using dominated convergence and the continuity of g), we thus find

$$\frac{\partial}{\partial r_i} \int_S g(x) dx = \frac{1}{r_i} |P| \int_0^1 \int_{(0,1)^{d-1}} g(r\eta_{i,0}(\mathbf{t}) + p_i) r^{d-1} \psi(\mathbf{t}) d\mathbf{t} dr. \quad (3.7)$$

Noting that

$$\sum_{j \neq i} T_j(1-t_j) = \sum_{j \neq i}^{d-1} (T_j - T_{j+1}) + T_d = T_1 - (T_i - T_{i+1}) = 1 - T_i(1-t_i),$$

we find

$$\begin{aligned} r\eta_{i,0}(t_1, \dots, t_{d-1}) + p_i &= r \left(\sum_{j \neq i} T_j(1-t_j)(p_j - p_i) \right) + p_i \\ &= r \left(\sum_{j \neq i} T_j(1-t_j)p_j - (1 - T_i(1-t_i))p_i \right) + p_i \\ &= r\eta(t_1, \dots, t_{d-1}) + (1-r)p_i, \end{aligned}$$

which together with (3.7) yields

$$\frac{\partial}{\partial r_i} \int_S g(x) dx = \frac{1}{r_i} |P| \int_0^1 \int_{(0,1)^{d-1}} g(r\eta(\mathbf{t}) + (1-r)p_i) r^{d-1} \psi(\mathbf{t}) d\mathbf{t} dr.$$

At this point we remark that the ordering of the points p_1, \dots, p_d is arbitrary and has no influence on the value of the above integrals. As such, we may swap the places of two points, say p_1 and p_i , and thus replacing η by η_i in the above integral. From this, we see

$$r\eta_i(t_1, \dots, t_{d-1}) + (1-r)p_i = (1-rt_1)p_i + rt_1(1-t_2)p_2 + \dots + rt_1 \left(\prod_{j=2}^{d-1} t_j \right) p_d = \eta_i(rt_1, \dots, t_{d-1}).$$

Thus, making the substitution $u = rt_1$ in the above integral, we get

$$\begin{aligned} \int_0^1 \int_{(0,1)^{d-1}} g(r\eta(\mathbf{t}) + (1-r)p_i) r^{d-1} \psi(\mathbf{t}) d\mathbf{t} dr &= \int_0^1 \int_{(0,1)^{d-2}} \int_0^r g(\eta_i(u, \mathbf{t})) \psi(u, \mathbf{t}) du d\mathbf{t} dr \\ &= \int_{(0,1)^{d-2}} \int_0^1 g(\eta(u, \mathbf{t})) \psi(u, \mathbf{t}) (1-u) du d\mathbf{t} \\ &= \int_{(0,1)^{d-1}} g(\eta_i(\mathbf{t})) \widehat{\psi}(\mathbf{t}) d\mathbf{t}, \end{aligned}$$

which shows (3.5). Finally, to show (3.6), we have by Jacobi's formula,

$$\frac{\partial}{\partial r_i} |P_{-1}^T P_{-1}| = |P_{-1}^T P_{-1}| \operatorname{tr} \left((P_{-1}^T P_{-1})^{-1} \frac{\partial}{\partial r_i} (P_{-1}^T P_{-1}) \right),$$

whereby

$$\frac{\partial}{\partial r_i} \int_F g(x) \mathcal{H}^{d-1}(dx) = \frac{1}{2} \operatorname{tr} \left((P_{-1}^T P_{-1})^{-1} \frac{\partial}{\partial r_i} (P_{-1}^T P_{-1}) \right) \int_{(0,1)^{d-1}} g(\eta(\mathbf{t})) \psi(\mathbf{t}) d\mathbf{t}$$

$$\begin{aligned}
& + \sqrt{|P_{-1}^T P_{-1}|} \frac{\partial}{\partial r_i} \left(\int_{(0,1)^{d-1}} g(\eta_i(\mathbf{t})) \psi(\mathbf{t}) \, d\mathbf{t} \right) \\
& = \frac{1}{2} \operatorname{tr} \left((P_{-1}^T P_{-1})^{-1} \frac{\partial}{\partial r_i} (P_{-1}^T P_{-1}) \right) \int_{(0,1)^{d-1}} g(\eta(\mathbf{t})) \psi(\mathbf{t}) \, d\mathbf{t} \\
& \quad + \sqrt{|P_{-1}^T P_{-1}|} \int_{(0,1)^{d-1}} \langle \nabla g(\eta_i(\mathbf{t})), \mathbf{q}_i \rangle \widehat{\psi}(\mathbf{t}) \, d\mathbf{t},
\end{aligned}$$

where we in the last equality used $\frac{\partial}{\partial r_i} \eta_i(t_1, \dots, t_{d-1}) = (1 - t_1) \mathbf{q}_i$. \blacksquare

When the dynamics of the process are known, that is we have access to the potential V , the theorem allows us to find explicit expressions for the right hand side of (3.1) and (3.2) given N points on the sphere and therefore to numerically optimize within the space of polytope approximations to star-shaped sets via gradient descent. For the two-dimensional case, strikingly simple expressions can be derived from this and, given sufficient regularity of the star-shaped domain, the approximation rate in (3.1) is at most of order $1/N$.

Corollary 3.2. *Let $r \in \mathcal{C}^2([0, 2\pi], [r_{\min}, r_{\max}])$ with $0 < r_{\min} \leq r_{\max} < \infty$ be some periodic radial function, and denote by $D \subset \mathbb{R}^2$ the set with $\partial D = r([0, 2\pi])$. For any $N \in \mathbb{N}$, let $r_i = r(\frac{2i\pi}{N})$, $\mathbf{q}_i = (\cos \frac{2i\pi}{N}, \sin \frac{2i\pi}{N})$ and $\mathbf{p}_i = r_i \mathbf{q}_i$ and denote by $\tilde{D} \subset \mathbb{R}^2$ the simplex with vertices $\{\mathbf{p}_i\}_{i=1}^N$. Finally, for $i = 1, \dots, N$ and $t \in (0, 1)$, let $\eta_i^+(t) = \mathbf{p}_i + t(\mathbf{p}_{i+1} - \mathbf{p}_i)$ and $\eta_i^-(t) = \mathbf{p}_i + t(\mathbf{p}_{i-1} - \mathbf{p}_i)$, where we identify $\mathbf{p}_0 = \mathbf{p}_N$ and $\mathbf{p}_{N+1} = \mathbf{p}_1$. Then, there exists a constant $K \geq 0$ such that*

$$|J(D) - J(\tilde{D})| \leq \frac{K}{N},$$

where for $\tilde{\rho} = e^{-V}$ we have the explicit representations

$$J(\tilde{D}) = \frac{1}{C} \sum_{i=1}^N \left(\sin \frac{2\pi}{N} r_i r_{i+1} \int_0^1 \int_0^1 (f\tilde{\rho})(r\eta_i^+(t)) r \, dr \, dt + \kappa |p_{i+1} - p_i| \int_0^1 \tilde{\rho}(\eta_i^+(t)) \, dt \right),$$

and

$$\frac{\partial J(\tilde{D})}{\partial r_i} = \frac{1}{C} \int_0^1 (\psi^+(t) \tilde{\rho}(\eta_i^+(t)) + \psi^-(t) \tilde{\rho}(\eta_i^-(t))) \, dt,$$

where

$$\psi^\pm(t) = \left(\sin \frac{2\pi}{N} r_{i\pm 1} (f(\eta_i^\pm(t)) - J(\tilde{D})) - \kappa |p_{i\pm 1} - p_i| \langle \nabla V(\eta_i^\pm(t)), \mathbf{q}_i \rangle \right) (1-t) + \frac{\kappa (r_i - \cos \frac{2\pi}{N} r_{i\pm 1})}{|p_i - p_{i\pm 1}|},$$

and

$$C = \sum_{i=1}^N \sin \frac{2\pi}{N} r_i r_{i+1} \int_0^1 \int_0^1 \tilde{\rho}(r\eta_i^+(t)) r \, dr \, dt.$$

Proof. We show only the first claim, the rest is an immediate consequence of Theorem 3.1. To this end, we assume for convenience that $N \geq 4$ and introduce the following notation: for $\theta \in [0, 2\pi]$ and any radial function $\hat{r}: [0, 2\pi] \rightarrow (0, \infty)$ which is \mathcal{C}^2 almost everywhere, let $p(\theta) = (\cos \theta, \sin \theta)$ and $\widehat{r}(\theta) = (\hat{r}(\theta), \hat{r}'(\theta))$ when the derivative exists. Also, for any continuous function $g: \mathbb{R}^2 \rightarrow \mathbb{R}$, write

$$I(\widehat{r}, g) := \int_0^{2\pi} \int_0^{\widehat{r}(\theta)} g(sp(\theta)) s \, ds \, d\theta, \quad \text{and} \quad S(\widehat{r}, g) := \int_0^{2\pi} g(\widehat{r}(\theta)p(\theta)) |\widehat{r}(\theta)| \, d\theta.$$

It is then straightforward to check that

$$J(D) = \frac{I(r, f\tilde{\rho}) + \kappa S(r, \tilde{\rho})}{I(r, \tilde{\rho})}.$$

Letting $\tilde{r} : [0, 2\pi] \rightarrow [r_{\min}, r_{\max}]$ denote the radial function corresponding to \tilde{D} , we then have

$$|J(D) - J(\tilde{D})| = \frac{1}{I(r, \tilde{\rho})} \left| (I(r, f\tilde{\rho}) - I(\tilde{r}, f\tilde{\rho})) + \kappa(S(r, \tilde{\rho}) - S(\tilde{r}, \tilde{\rho})) + J(\tilde{D})(I(r, \tilde{\rho}) - I(\tilde{r}, \tilde{\rho})) \right|.$$

By our assumptions on r , it follows that $I(r, \tilde{\rho})$ is bounded from below and $J(\tilde{D})$ from above. As such, to show the desired convergence rate, we only need to show that for any continuous $g : \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$|I(r, g) - I(\tilde{r}, g)| \vee |S(r, \tilde{\rho}) - S(\tilde{r}, \tilde{\rho})| \leq \frac{K_1}{N} \quad (3.8)$$

for some $K_1 \geq 0$. In fact, by dominating with $\sup_{x \in B(0, r_{\max})} |g(x)|$, it suffices to show (3.8) with $g \equiv 1$. This follows if we show for $i = 1, \dots, N$

$$|r(\theta) - \tilde{r}(\theta)| \vee |r'(\theta) - \tilde{r}'(\theta)| \leq \frac{K_2}{N}, \quad \theta \in \left(\frac{2(i-1)\pi}{N}, \frac{2i\pi}{N} \right) \quad (3.9)$$

for some $K_2 \geq 0$ independent of i . Indeed, if this is the case, we have first,

$$|I(r, 1) - I(\tilde{r}, 1)| \leq \sum_{i=1}^N \int_{\frac{2(i-1)\pi}{N}}^{\frac{2i\pi}{N}} \left| \int_{\tilde{r}(\theta)}^{r(\theta)} s \, ds \right| d\theta \leq r_{\max} \sum_{i=1}^N \int_{\frac{2(i-1)\pi}{N}}^{\frac{2i\pi}{N}} |r(\theta) - \tilde{r}(\theta)| d\theta \leq \frac{2\pi r_{\max} K_2}{N},$$

and similarly, since $|r - \tilde{r}| = \sqrt{(r - \tilde{r})^2 + (r' - \tilde{r}')^2} \leq \sqrt{2} \frac{K_2}{N}$,

$$\begin{aligned} |S(r, \tilde{\rho}) - S(\tilde{r}, \tilde{\rho})| &\leq \sum_{i=1}^N \int_{\frac{2(i-1)\pi}{N}}^{\frac{2i\pi}{N}} (|r(\theta)| |\tilde{\rho}(r(\theta)p(\theta)) - \tilde{\rho}(\tilde{r}(\theta)p(\theta))| + \tilde{\rho}(\tilde{r}(\theta)p(\theta)) ||r(\theta)| - |\tilde{r}(\theta)||) d\theta \\ &\leq K_3 \sum_{i=1}^N \int_{\frac{2(i-1)\pi}{N}}^{\frac{2i\pi}{N}} (|r(\theta) - \tilde{r}(\theta)| + |r(\theta) - \tilde{r}(\theta)|) d\theta \leq \frac{2\pi(1 + \sqrt{2})K_2 K_3}{N}, \end{aligned}$$

where

$$K_3 = \left(\left(\sup_{\theta \in [0, 2\pi]} |r(\theta)| \sup_{x \in \text{cl } B(0, r_{\max})} |\nabla \tilde{\rho}(x)| \right) \vee \sup_{x \in \text{cl } B(0, r_{\max})} \tilde{\rho}(x) \right)$$

whereby we may choose $K_1 = 2\pi K_2 (r_{\max} \vee (1 + \sqrt{2})K_3)$. To verify (3.9), we first find by some straightforward trigonometry, identifying $r_0 = r_N$,

$$\tilde{r}(\theta) = \sum_{i=1}^N \frac{r_i (m_i \cos \frac{2i\pi}{N} - \sin \frac{2i\pi}{N})}{m_i \cos \theta - \sin \theta} \mathbf{1}_{\left[\frac{2(i-1)\pi}{N}, \frac{2i\pi}{N} \right)}(\theta), \quad \text{where} \quad m_i = \frac{r_i \sin \frac{2i\pi}{N} - r_{i-1} \sin \frac{2(i-1)\pi}{N}}{r_i \cos \frac{2i\pi}{N} - r_{i-1} \cos \frac{2(i-1)\pi}{N}}.$$

Here, when $r_i \cos \frac{2i\pi}{N} = r_{i-1} \cos \frac{2(i-1)\pi}{N}$ we set $m_i = \infty$ and treat \tilde{r} in a limiting sense. Now, for $i \in \{1, \dots, N\}$ and $\theta \in \left(\frac{2(i-1)\pi}{N}, \frac{2i\pi}{N} \right)$, let $\varphi_i(\theta)$ denote the i 'th summand in $\tilde{r}(\theta)$. Let also $\psi_i(\theta) = \frac{m_i \sin \theta + \cos \theta}{m_i \cos \theta - \sin \theta}$, such that $\varphi_i'(\theta) = \varphi_i(\theta) \psi_i(\theta)$. Further, since $\psi_i'(\theta) = 1 + \psi_i(\theta)^2$, we find also $\varphi_i''(\theta) = \varphi_i(\theta)(1 + 2\psi_i(\theta)^2)$, whereby to bound \tilde{r}, \tilde{r}' and \tilde{r}'' (when the latter two exist), we need only bound φ_i and ψ_i for all i . Clearly, we have $\varphi_i(\theta) \leq r_{\max}$, and since $\psi_i'(\theta) > 0$, we must have $|\psi_i(\theta)| \leq |\psi_i(\frac{2(i-1)\pi}{N})| \vee |\psi_i(\frac{2i\pi}{N})|$. Some simple, albeit tedious, applications of trigonometric identities show that

$$\psi_i\left(\frac{2(i-1)\pi}{N}\right) = -\frac{r_{i-1} - r_i \cos \frac{2\pi}{N}}{r_i \sin \frac{2\pi}{N}} \quad \text{and} \quad \psi_i\left(\frac{2i\pi}{N}\right) = \frac{r_i - r_{i-1} \cos \frac{2\pi}{N}}{r_{i-1} \sin \frac{2\pi}{N}},$$

and since e.g. $|r_i - r_{i-1} \cos \frac{2\pi}{N}| \leq \cos \frac{2\pi}{N} |r_i - r_{i-1}| + r_i(1 - \cos \frac{2\pi}{N})$, we find by the Lipschitz-continuity of r and cosine,

$$|\psi_i(\theta)| \leq \frac{2\pi}{Nr_{\min}} \left(\frac{1}{\tan \frac{2\pi}{N}} \sup_{\hat{\theta} \in [0, 2\pi)} |r'(\hat{\theta})| + r_{\max} \right) \leq \frac{\sup_{\hat{\theta} \in [0, 2\pi)} |r'(\hat{\theta})|}{r_{\min}}.$$

Here we use that $\frac{x}{\tan x} \in [0, 1]$ for $0 \leq x \leq \frac{\pi}{2}$, and hence $\frac{2\pi}{N \tan \frac{2\pi}{N}} \in [0, 1]$ for $N \geq 4$. Since this bound depends neither on i nor θ , we find for all $\theta \in [0, 2\pi) \setminus \{\frac{2i\pi}{N}\}_{i=1}^N$

$$\tilde{r}(\theta) \leq r_{\max}, \quad |\tilde{r}'(\theta)| \leq \frac{r_{\max}}{r_{\min}} \sup_{\hat{\theta} \in [0, 2\pi)} |r'(\hat{\theta})|, \quad \text{and} \quad |\tilde{r}''(\theta)| \leq r_{\max} \left(1 + 2 \left[\frac{1}{r_{\min}} \sup_{\hat{\theta} \in [0, 2\pi)} |r'(\hat{\theta})| \right]^2 \right).$$

Thus we have by Lipschitz-continuity for $i \in \{1, \dots, N\}$ and $\theta \in [\frac{2(i-1)\pi}{N}, \frac{2i\pi}{N})$, since $r(\frac{2(i-1)\pi}{2}) = r_{i-1} = \tilde{r}(\frac{2(i-1)\pi}{N})$

$$|r(\theta) - \tilde{r}(\theta)| \leq |r(\theta) - r_{i-1}| + |r_{i-1} - \tilde{r}(\theta)| \leq \frac{2\pi(1 + \frac{r_{\max}}{r_{\min}}) \sup_{\hat{\theta} \in [0, 2\pi)} |r'(\hat{\theta})|}{N}$$

Furthermore, since \tilde{r} and r agree on $\{\frac{2i\pi}{N}\}_{i=1}^N$, it follows by the mean value theorem that there exist $\theta^*, \hat{\theta}^* \in (\frac{2(i-1)\pi}{N}, \frac{2i\pi}{N})$, such that $r'(\theta^*) = \tilde{r}'(\hat{\theta}^*)$. Thus we also have

$$\begin{aligned} |r'(\theta) - \tilde{r}'(\theta)| &\leq |r'(\theta) - r'(\theta^*)| + |\tilde{r}'(\hat{\theta}^*) - \tilde{r}'(\theta)| \\ &\leq \frac{2\pi(\sup_{\hat{\theta} \in [0, 2\pi)} |r''(\hat{\theta})| + r_{\max}(1 + 2(\frac{1}{r_{\min}} \sup_{\hat{\theta} \in [0, 2\pi)} |r'(\hat{\theta})|)^2))}{N}, \end{aligned}$$

and the proof is finished by setting

$$K_2 = 2\pi \left(\left((1 + \frac{r_{\max}}{r_{\min}}) \sup_{\hat{\theta} \in [0, 2\pi)} |r'(\hat{\theta})| \right) \vee \left(\sup_{\hat{\theta} \in [0, 2\pi)} |r''(\hat{\theta})| + r_{\max}(1 + 2(\frac{1}{r_{\min}} \sup_{\hat{\theta} \in [0, 2\pi)} |r'(\hat{\theta})|)^2) \right) \right).$$

■

This now implies for $d = 2$ that for suitable admissible domain families $\Theta \ni D$, the infimum of $D \mapsto J(D)$ over Θ is well approximated by the infimum over the polytope approximations.

Corollary 3.3. *Let $d = 2$ and let Θ be the family of domains D that are strongly starhaped at 0 and are identified by C^2 periodic radial functions $r_D : [0, 2\pi] \rightarrow (0, \infty)$ such that for some global constants $\underline{\lambda}, \bar{\lambda}, \Lambda$,*

$$\underline{\lambda} \leq r_D \leq \bar{\lambda} \quad \text{and} \quad \max_{\theta \in [0, 2\pi]} (|r'_D(\theta)| + |r''_D(\theta)|) \leq \Lambda.$$

Then, letting \tilde{D}_N be the polytope approximation of D from Corollary 3.2, it holds

$$|\inf_{D \in \Theta} J(\tilde{D}_N) - \inf_{D \in \Theta} J(D)| \leq 1/N.$$

Proof. For given $\varepsilon > 0$ choose $D^\varepsilon \in \Theta$ s.t. $J(D^\varepsilon) \leq \inf_{D \in \Theta} J(D) + \varepsilon$. Then, by Corollary 3.2,

$$\inf_{D \in \Theta} J(\tilde{D}_N) - \inf_{D \in \Theta} J(D) \leq \inf_{D \in \Theta} J(\tilde{D}_N) - J(D^\varepsilon) + \varepsilon \leq J(\tilde{D}_N) - J(D^\varepsilon) + \varepsilon \leq \sup_{D \in \Theta} (J(\tilde{D}_N) - J(D)) + \varepsilon \leq \frac{1}{N} + \varepsilon.$$

Similarly, letting $D^\varepsilon \in \Theta$ such that $J(\tilde{D}_N) \leq \inf_{D \in \Theta} J(D) + \varepsilon$, we have,

$$\inf_{D \in \Theta} J(D) - \inf_{D \in \Theta} J(\tilde{D}_N) \leq J(D^\varepsilon) - J(\tilde{D}_N) + \varepsilon \leq \sup_{D \in \Theta} |J(\tilde{D}_N) - J(D)| + \varepsilon \leq \frac{1}{N} + \varepsilon.$$

Taking together both bounds and letting $\varepsilon \downarrow 0$ therefore gives the result. ■

In the simple case of a pure Brownian motion, corresponding to $e^{-V} \equiv 1$, and given radially symmetric costs $f = |\cdot|$, one expects the optimal reflection boundary to be a sphere centered at 0. Optimizing the corresponding cost functional over the space of such balls only, gives the optimization problem a parametric structure that can be easily solved analytically to reveal the optimal ball to be $D^* = B(0, r^*)$, where $r^* = \sqrt{(d+1)\kappa}$. It is now interesting to test our method with regard to two questions: does the numerical optimization over the more general class of star-shaped domains support our intuition by identifying a ball as the optimal reflection domain, and if so, do we obtain a good approximation of the optimal radius $r^* = \sqrt{(d+1)\kappa}$ as well? The result for different choices of κ is visualized in Figure 3.1, giving an affirmative answer to both questions.

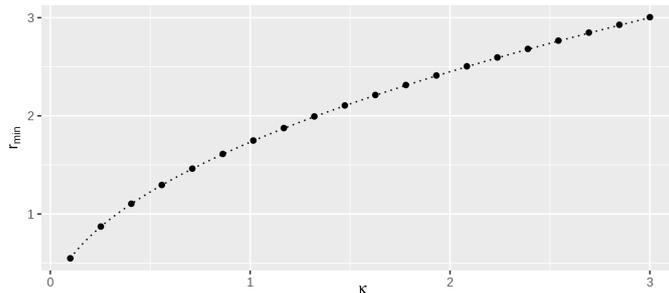


Figure 3.1: For each value of κ , we use a quasi-Newton method for finding the optimal shape according to the gradients above. Each time, a ball is indeed identified as the optimal shape and we plot its radius as a function of κ . Finally the dotted line represents the optimal line $r^* = \sqrt{3\kappa}$.

Our method is also well-equipped for handling more challenging non-symmetric situations, where it is hard to make an educated guess on the optimal shape. As such, we also test the method on reflected Ornstein–Uhlenbeck processes with strong correlation, in particular the processes

$$dX_t^D = AX_t^D dt + \sqrt{2} dW_t + n(X_t^D) dL_t^D, \quad A = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}^{-1},$$

as well as a skewed cost-function, namely $f(x, y) = \sqrt{x^2 + 5y^2}$. The found approximately optimal shapes can be seen in Figure 3.2. Here, for each shape we take $N = 50$, $\kappa = 1$ and use a quasi-Newton method to find the optimal shape with starting values $r_i = 1$ for $i = 1, \dots, 50$. The found optimal shapes correspond to, from top-left to bottom-right, Brownian motion with norm cost, Ornstein–Uhlenbeck process with norm cost, Brownian motion with skewed cost and Ornstein–Uhlenbeck process with skewed cost. Finally, Figure 3.2 also shows simulations of the above reflected processes in these approximately optimal shapes to assess the convergence of the realized costs towards the theoretical objective function. To do this, we use the usual Euler scheme, where we then project the process onto D whenever it leaves the domain. The distance of this projection is then the associated simulated increase in local time, cf. [44]. We simulate the relevant processes with time-steps of 10^{-4} until time $T = 100$ (but plot only until $T = 10$ for visual clarity). The average realized cost in comparison to the expected average long run costs are given in Table 3.1.

4. Learning the optimal boundary

We now turn to the challenging situation, when the dynamics of the unconstrained Langevin diffusion are unknown, which makes it impossible to set the optimization algorithm from the

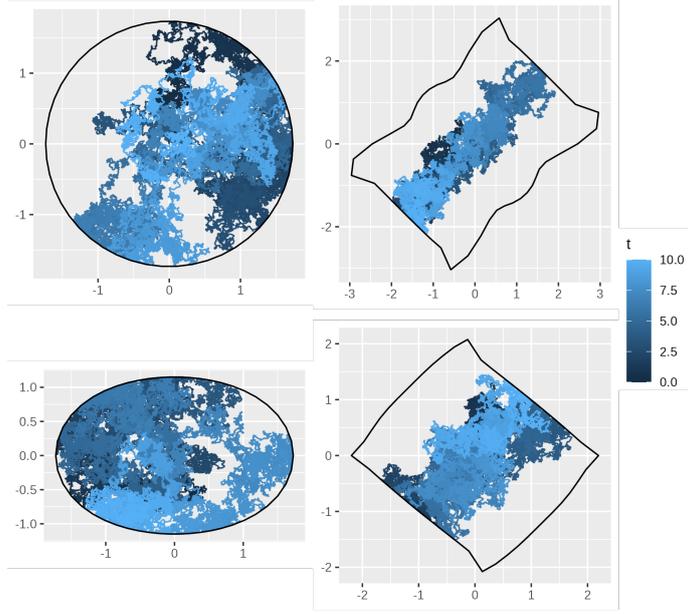


Figure 3.2: Simulated optimal shapes and corresponding path realizations of reflected processes.

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

Table 3.1: Average realized costs vs. expected average long term costs (in brackets)

previous section into motion, without feeding it information based on collected data first. As apparent from the explicit form of the cost functional given in Corollary 2.6, a natural data-driven reflection procedure can be based on a plug-in approach, provided that we have an efficient estimator of (functionals of) the invariant density of the unconstrained Langevin diffusion at our disposal.

4.1. Adaptive nonparametric estimation of the invariant density

As in the scalar case discussed in [11], we employ a kernel estimator of the invariant density, whose sup-norm risk given appropriate conditions on the diffusion coefficients is well-understood in a general context by now. In the following, we will concentrate on a class of potentials V s.t. the process satisfies certain functional inequalities. This setting is quite natural given the reversible nature of Langevin diffusions and is studied in [46], where minimax optimal estimation rates for a Lepski type *adaptive* kernel estimator are established under *anisotropic* Hölder smoothness assumptions on the invariant density. To recall these results, some preliminary definitions are necessary.

Let π be the invariant distribution of X with density $\rho \propto \exp(-V)$ and let L be the $L^2(\pi)$ -generator of X with domain $\mathcal{D}(L)$. Since $-L$ is self-adjoint and nonnegative on the Hilbert space $L^2(\pi)$ endowed with the inner product $\langle f, g \rangle_\pi := \int fg d\pi$, we may define $\sqrt{-L}$ via spectral calculus and note that for any $f \in \mathcal{D}(\sqrt{-L})$, we have $\|\sqrt{-L}f\|_\pi^2 = -\langle Lf, f \rangle_\pi$, which for $f \in \mathcal{C}_c^2(\mathbb{R}^d)$ is equal to $\|\nabla f\|_\pi^2$. We treat diffusion models satisfying the following conditions.

Definition 4.1. (PI) X satisfies a *Poincaré inequality* with constant C_P if, for any $f \in \mathcal{D}(\sqrt{-L})$,

$$\text{Var}_\pi(f) := \pi(f^2) - \pi(f)^2 \leq C_P \|\sqrt{-L}f\|_\pi^2.$$

(NI) X satisfies a *Nash inequality* with constants C_N^1, C_N^2 if, for any $f \in \mathcal{D}(\sqrt{-L})$,

$$\|f\|_\pi^{d+2} \leq \left(C_N^1 \|f\|_\pi^2 + C_N^2 \|\sqrt{-L}f\|_\pi^2 \right)^{d/2} \|f\|_{L^1(\pi)}^2.$$

Denote by $\Sigma(C_P, C_N^1, C_N^2)$ the class of potentials $V : \mathbb{R}^d \rightarrow \mathbb{R}^d$ s.t. the corresponding Langevin diffusion satisfies (PI) and (NI).

Remark 4.2. A Nash inequality is called *tight* if $C_N^1 = 1$, in which case it implies a *logarithmic Sobolev inequality* and therefore also a Poincaré inequality.

The combination of a Poincaré inequality with a Nash inequality is particularly attractive from a statistical point of view. A Poincaré inequality is equivalent to exponential ergodicity in $L^2(\pi)$. More precisely, for any $f \in L^2(\pi)$ with $\pi(f) = 0$, $\|P_t f\|_\pi \leq \exp(-t/C_P) \|f\|_\pi$, which enforces a fast mixing behavior of the diffusion. A Nash inequality on the other hand is equivalent to *ultracontractivity* expressed through the heat kernel bound $\|P_t\|_{L^1(\pi) \rightarrow L^\infty(\pi)} \leq t^{-d/2}$ for $t \in (0, 1]$, cf. [4, Theorem 6.3.1]. The key observation for the statistical approach is that the combination of both yields tight variance bounds for path integrals $\int_0^t f(X_s) ds$ of functionals $f \in L^2(\pi)$, which allows for efficient control of the stochastic fluctuations of kernel estimators.

To control the bias, we impose anisotropic Hölder regularity conditions on the invariant density.

Definition 4.3. Let $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d) \in (0, \infty)^d$, $\boldsymbol{\mathcal{L}} = (\mathcal{L}_1, \dots, \mathcal{L}_d) \in (0, \infty)^d$. A function $h : \mathbb{R}^d \rightarrow \mathbb{R}$ is said to belong to the anisotropic Hölder class $\mathcal{H}_d(\boldsymbol{\beta}, \boldsymbol{\mathcal{L}})$ if, for all $i = 1, \dots, d$,

$$\begin{aligned} \|D_i^k g\|_\infty &\leq \mathcal{L}_i, \quad k = 1, \dots, \lfloor \beta_i \rfloor, \\ \|D_i^{\lfloor \beta_i \rfloor} g(\cdot + te_i) - D_i^{\lfloor \beta_i \rfloor} g(\cdot)\|_\infty &\leq \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor}, \quad t \in \mathbb{R}, \end{aligned}$$

where $\lfloor \beta \rfloor$ denotes the largest integer strictly smaller than $\beta > 0$. Denote by

$$\mathbb{H}_d(\boldsymbol{\beta}, \boldsymbol{\mathcal{L}}) = \mathbb{H}_d(\boldsymbol{\beta}, \boldsymbol{\mathcal{L}}; C_\infty, C_P, C_N^1, C_N^2),$$

the set of invariant densities $\rho_V \in \mathcal{H}_d(\boldsymbol{\beta} + \mathbf{1}, \boldsymbol{\mathcal{L}})$ s.t. $\|\rho_V\|_\infty \leq C_\infty$ and $V \in \Sigma(C_P, C_N^1, C_N^2)$.

Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a symmetric Lipschitz kernel function with $\text{supp}(K) \subset [-1/2, 1/2]$ and $\int K(x) dx = 1$. We say that K is of order $\ell \in \mathbb{N}$ if $\int x^m K(x) dx = 0$ for any $m = 0, 1, \dots, \ell$. For $h > 0$ we let $K_h(\cdot) := h^{-1}K(\cdot/h)$ and for $\mathbf{h}, \boldsymbol{\eta} \in (0, \infty)^d$ we set

$$\mathbb{K}_{\mathbf{h}}(x) := \prod_{i=1}^d K_{h_i}(x_i), \quad x \in \mathbb{R}^d,$$

and

$$\mathbb{K}_{\mathbf{h}} \star \mathbb{K}_{\boldsymbol{\eta}}(x) := \prod_{i=1}^d K_{h_i} \star K_{\eta_i}(x_i), \quad x \in \mathbb{R}^d.$$

We now define the following kernel estimators given a continuous record $(X_t)_{t \in [0, T]}$ s.t. $\rho_V \in \mathbb{H}(\boldsymbol{\beta}, \boldsymbol{\mathcal{L}})$:

$$\widehat{\rho}_{\mathbf{h}, T}(x) := \frac{1}{T} \int_0^T \mathbb{K}_{\mathbf{h}}(x - X_s) ds, \quad \widehat{\rho}_{\mathbf{h}, \boldsymbol{\eta}, T}(x) := \frac{1}{T} \int_0^T \mathbb{K}_{\mathbf{h}} \star \mathbb{K}_{\boldsymbol{\eta}}(x - X_s) ds, \quad x \in \mathbb{R}^d.$$

In order to efficiently estimate ρ_V via $\widehat{\rho}_{\mathbf{h}, T}$, the bandwidth \mathbf{h} has to be carefully chosen to achieve an optimal balance between bias and variance of the kernel estimator. If the Hölder smoothness parameter $\boldsymbol{\beta}$ is unknown, the bias cannot be evaluated directly, which poses the fundamental

challenge to design a fully data-driven/adaptive bandwidth selection procedure to obtain a rate-optimal but possibly random bandwidth $\widehat{\mathbf{h}}_T$.

As it turns out, in dimension $d = 2$ this problem is significantly simplified since in this case a tight variance bound of the kernel estimator only depends logarithmically on the bandwidth. Consequently, the smoothness independent and deterministic bandwidth choice

$$\widehat{\mathbf{h}}_T \sim T^{-1/2}(1, 1),$$

yields the optimal sup-norm estimation rate $\log T / \sqrt{T}$ given that the order of K is chosen large enough. In dimension $d \geq 3$, the situation is significantly more involved and the bandwidth is chosen according to the following Lepski type selection rule:

Let $q \geq 1$ and the set of candidate bandwidths $\mathcal{H}_T = \mathcal{H}_T^{(q)}$ be given by

$$\mathcal{H}_T := \left\{ \mathbf{h} \in (0, 1]^d : T \mathfrak{a}_\circ^2 \geq \prod_{j=1}^d h_j^{(2/d)-1} \log T \right\}.$$

Here, $\mathfrak{a}_\circ = \mathfrak{a}_\circ^{(q)} := (2\Lambda)^{-2}$, where $\Lambda = \Lambda^{(q)} := \gamma_{2q}^\circ(d, \|K\|_\infty; C_P, C_N^1, C_N^2)$ and the functional γ_p° , $p \geq 1$, is defined in [46, Remark 3.5] based on the functional γ_p in [34]. Denote

$$\bar{\zeta}_T := 2 \left(1 \vee \sup_{\mathbf{h} \in \mathcal{H}_T} \left\| T^{-1} \int_0^T |\mathbb{K}_{\mathbf{h}}(X_s - \cdot)| ds \right\|_\infty \right),$$

and set

$$\widehat{\Delta}_T(\mathbf{h}) := \sup_{\boldsymbol{\eta} \in \overline{\mathcal{H}}_T} \left\{ \left[\|\widehat{\rho}_{\mathbf{h}, \boldsymbol{\eta}} - \widehat{\rho}_{\boldsymbol{\eta}}\|_\infty - \lambda \widehat{A}_T(\boldsymbol{\eta}) \right]_+ \right\},$$

where $\overline{\mathcal{H}}_T \subset \mathcal{H}_T$ is the dyadic grid in \mathcal{H}_T , $\lambda = \lambda^{(q)} := (1 \vee \|K\|_{L^1(\lambda)}^d) \Lambda$, and

$$\widehat{A}_T(\mathbf{h}) := \prod_{i=1}^d h_i^{1/d-1/2} \sqrt{\frac{\bar{\zeta}_T \log T}{T}}.$$

We now specify the bandwidth $\widehat{\mathbf{h}}_T = \widehat{\mathbf{h}}_T^{(q)}$ by

$$\widehat{\Delta}_T(\widehat{\mathbf{h}}_T) + \lambda \widehat{A}_T(\widehat{\mathbf{h}}_T) = \inf_{\mathbf{h} \in \mathcal{H}_T} \left\{ \widehat{\Delta}_T(\mathbf{h}) + \lambda \widehat{A}_T(\mathbf{h}) \right\}.$$

Finally, for $q \geq 1$, $d \geq 2$ and $\widehat{\mathbf{h}}_T = \widehat{\mathbf{h}}_T^{(q)}$ given as above, we set

$$\widehat{\rho}_{\widehat{\mathbf{h}}_T, T}(x) = \frac{1}{T} \int_0^T \mathbb{K}_{\widehat{\mathbf{h}}_T}(x - X_s) ds, \quad x \in \mathbb{R}^d.$$

According to the discussion in [46, Section 3] on the two-dimensional case and [46, Theorem 3.4], we now have the following uniform sup-norm estimation result.

Theorem 4.4. *Suppose $\boldsymbol{\beta} \in (0, \mathfrak{b}]^d$ for some $\mathfrak{b} \in \mathbb{N} \cap [2, \infty)$ and let K have order $\mathfrak{b} + 1$. Then, for any $q \geq 1$ and $\mathcal{L} \in (0, \infty)^d$ it holds*

$$\sup_{\rho_V \in \mathbb{H}_d(\boldsymbol{\beta}, \mathcal{L})} \left(\mathbb{E}^x \left[\|\widehat{\rho}_{\widehat{\mathbf{h}}_T, T} - \rho_V\|_\infty^q \right] \right)^{1/q} = \mathcal{O}(\Psi_{d, \boldsymbol{\beta}}(T)),$$

where for the harmonic mean smoothness $\overline{\boldsymbol{\beta} + \mathbf{1}} := (d^{-1} \sum_{i=1}^d \frac{1}{\beta_i + 1})^{-1}$, the rate $\Psi_{d, \boldsymbol{\beta}}$ is specified by

$$\Psi_{d, \boldsymbol{\beta}}(T) := \begin{cases} \frac{\log T}{\sqrt{T}}, & d = 2, \\ \left(\frac{\log T}{T} \right)^{\frac{\overline{\boldsymbol{\beta} + \mathbf{1}}}{2\overline{\boldsymbol{\beta} + \mathbf{1}} + d - 2}}, & d \geq 3. \end{cases}$$

4.2. Data-driven estimation of the optimal reflection boundary

In this section we consider a set of domains $\Theta \subset \mathbf{D}$ such that the set of minimizers $\arg \min_{D \in \Theta} J(D)$ is well-defined and let $D^* \in \arg \min_{D \in \Theta} J(D)$. Our data-driven procedure to determine reflection domains \widehat{D} whose average costs are close to the optimal ergodic costs $J(D^*)$ uses the following assumption.

Assumption 4.5. (i) For some constants $\underline{\lambda}, \bar{\lambda}, \Lambda$ it holds $B(0, \underline{\lambda}) \subset D^* \subset B(0, \bar{\lambda})$ and $\mathcal{H}^{d-1}(\partial D^*) \leq \Lambda$;

(ii) we are given information on two constants $\underline{\rho}, \bar{\rho}$ such that $\underline{\rho} \leq \inf_{B(0, \bar{\lambda})} \rho \leq \sup_{B(0, \bar{\lambda})} \rho \leq \bar{\rho}$.

Accordingly, we define the truncated invariant density estimator $\widehat{\rho}_{T,q}^*$ based on data $(X_t)_{t \in [0, T]}$ of the uncontrolled diffusion process as

$$\widehat{\rho}_{T,q}^*(x) := (\widehat{\rho}_{T,q}(x) \wedge 2\bar{\rho}) \vee \underline{\rho}/2, \quad x \in \mathbb{R}^d,$$

with $\widehat{\rho}_{T,q}$ specified as the adaptive invariant density estimator from the previous subsection with bandwidth choice $\mathbf{h} = \widehat{\mathbf{h}}_T^{(q)}$ for $q \geq 1$. Moreover, we let $\Theta(\underline{\lambda}, \bar{\lambda}, \Lambda) \subset \Theta$ be the subfamily of reflection domains satisfying Assumption 4.5.(i). Let

$$\widehat{J}_{T,q}(D) := \frac{1}{\int_D \widehat{\rho}_{T,q}^*(x) dx} \left(\int_D f(y) \widehat{\rho}_{T,q}^*(y) dy + \kappa \int_{\partial D} \widehat{\rho}_{T,q}^*(y) \mathcal{H}^{d-1}(dy) \right), \quad D \in \Theta,$$

be the estimator of the asymptotic costs associated to the reflection domain D and define the reflection domain estimator

$$\widehat{D}_{T,q} \in \arg \min_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} \widehat{J}_{T,q}(D).$$

Here, we must assume that $\Theta \subset \mathbf{D}$ is a metrizable space that is sufficiently nice to allow a measurable choice of $\widehat{D}_{T,q}$ considered as a random mapping into the Borel space associated to Θ . We now have the following concentration result for the simple regret.

Proposition 4.6. Suppose that $X_0 \sim \mu$, where $\mu \ll \pi$ with $\|\frac{d\mu}{d\pi}\|_{L^q(\pi)} < \infty$ for some $q \in (1, \infty]$. Then, for any $p \geq 1$, given the assumptions from Theorem 4.4 we have the regret bound

$$\mathbb{E}^\mu [|J(D^*) - J(\widehat{D}_{T,p\bar{q}})|^p]^{1/p} \leq C \|\frac{d\mu}{d\pi}\|_{L^q(\pi)}^{1/p} \Psi_{d,\beta}(T),$$

where $\bar{q} := q/(q-1)$ is the conjugate Hölder exponent of q and C depends on κ, p, q, f and the constants from Assumption 4.5.

Proof. For fixed $D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)$, write (in obvious notation)

$$J(D) = A(D)/B(D), \quad \widehat{J}_{T,p\bar{q}}(D) = \widehat{A}_T(D)/\widehat{B}_T(D).$$

Using

$$B(D) \wedge \widehat{B}_T(D) \geq \int_{B(0, \underline{\lambda})} \rho(x) \wedge \widehat{\rho}_{T,p\bar{q}}^*(x) dx \geq \frac{\underline{\rho}}{2} \lambda(B(0, \underline{\lambda})) =: \omega,$$

it follows

$$\begin{aligned} |J(D) - \widehat{J}_{T,p\bar{q}}(D)| &\leq \frac{A(D)}{B(D)\widehat{B}_T(D)} |\widehat{B}_T(D) - B(D)| + \frac{|\widehat{A}_T(D) - A(D)|}{\widehat{B}_T(D)} \\ &\leq \frac{\|f\|_{L^1(\pi)} + \kappa\bar{\rho}\Lambda}{\omega^2} \|\rho - \widehat{\rho}_{T,p\bar{q}}^*\|_{L^\infty(B(0, \bar{\lambda}))} + \frac{\|f\|_{L^1(B(0, \bar{\lambda}))} + \kappa\Lambda}{\omega} \|\rho - \widehat{\rho}_{T,p\bar{q}}^*\|_{L^\infty(B(0, \bar{\lambda}))}. \end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{E}^\mu \left[\sup_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} |J(D) - \widehat{J}_{T, p\bar{q}}(D)|^p \right]^{1/p} &\leq C(\kappa, \underline{\lambda}, \bar{\lambda}, \underline{\rho}, \bar{\rho}, p, f) \mathbb{E}^\mu \left[\|\rho - \widehat{\rho}_{T, p\bar{q}}^*\|_{L^\infty(B(0, \bar{\lambda}))}^p \right]^{1/p} \\
&\leq C(\kappa, \underline{\lambda}, \bar{\lambda}, \underline{\rho}, \bar{\rho}, p, f) \|\frac{d\mu}{d\pi}\|_{L^q(\pi)}^{1/p} \mathbb{E}^\pi \left[\|\rho - \widehat{\rho}_{T, p\bar{q}}^*\|_{L^\infty(B(0, \bar{\lambda}))}^{p\bar{q}} \right]^{1/p\bar{q}},
\end{aligned} \tag{4.1}$$

where we used Hölder inequality twice for the last line. The expectation can be bounded as follows:

$$\begin{aligned}
\mathbb{E}^\pi \left[\|\rho - \widehat{\rho}_{T, p\bar{q}}^*\|_{L^\infty(B(0, \bar{\lambda}))}^{p\bar{q}} \right]^{1/p\bar{q}} &\leq \mathbb{E}^\pi \left[\|\rho - \widehat{\rho}_{T, p\bar{q}}\|_{L^\infty(B(0, \bar{\lambda}))}^{p\bar{q}} \right]^{1/p\bar{q}} + \mathbb{E}^\pi \left[\|\rho - \widehat{\rho}_{T, p\bar{q}}^*\|_{L^\infty(B(0, \bar{\lambda}))}^{p\bar{q}} \mathbf{1}_{\{\|\widehat{\rho}_{T, p\bar{q}} - \rho\|_\infty > \underline{\rho}/2\}} \right]^{1/p\bar{q}} \\
&\leq \mathbb{E}^\pi \left[\|\rho - \widehat{\rho}_{T, p\bar{q}}\|_{L^\infty(B(0, \bar{\lambda}))}^{p\bar{q}} \right]^{1/p\bar{q}} + 2\bar{\rho} \left(\mathbb{P}^\pi \left(\|\widehat{\rho}_{T, p\bar{q}} - \rho\|_\infty > \underline{\rho}/2 \right) \right)^{1/p\bar{q}} \\
&\leq (1 + 4\bar{\rho}/\underline{\rho}) \mathbb{E}^\pi \left[\|\rho - \widehat{\rho}_{T, p\bar{q}}\|_{L^\infty(B(0, \bar{\lambda}))}^{p\bar{q}} \right]^{1/p\bar{q}} \\
&\lesssim (1 + 4\bar{\rho}/\underline{\rho}) \Psi_{d, \beta}(T).
\end{aligned}$$

For the first inequality we used that on $\{\|\widehat{\rho}_{T, p\bar{q}} - \rho\|_\infty \leq \underline{\rho}/2\}$ it holds $\widehat{\rho}_{T, p\bar{q}}|_{B(0, \bar{\lambda})} \equiv \widehat{\rho}_{T, p\bar{q}}^*|_{B(0, \bar{\lambda})}$. The last two lines follow from the Markov inequality and Theorem 4.4. Plugging this bound into (4.1) we find

$$\mathbb{E}^\mu \left[\sup_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} |J(D) - \widehat{J}_{T, p\bar{q}}(D)|^p \right]^{1/p} \lesssim \|\frac{d\mu}{d\pi}\|_{L^q(\pi)}^{1/p} \Psi_{d, \beta}(T). \tag{4.2}$$

Finally, since $D^* \in \arg \min_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} J(D)$ and $\widehat{D}_{T, p\bar{q}} \in \arg \min_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} \widehat{J}_{T, p\bar{q}}(D)$, we have

$$\mathbb{E}^\mu \left[|J(D^*) - J(\widehat{D}_{T, p\bar{q}})|^p \right]^{1/p} \leq 2 \mathbb{E}^\mu \left[\sup_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} |J(D) - \widehat{J}_{T, p\bar{q}}(D)|^p \right]^{1/p},$$

which in combination with (4.2) yields the claim. \blacksquare

This result may be interpreted in two different ways. On the one hand, it shows for the generic situation, where the controller has access to a separate diffusion data sample and uses it in online estimation of an optimal reflection boundary, that the regret vanishes at the nonparametric estimation rate. On the other hand, it demonstrates that a simple explore-then-commit strategy, where we first estimate an optimal set for T time units and afterwards exploit by reflecting the process at the estimated boundaries, yields a regret bounded by $\Psi_{d, \beta}(T)$.

As a proof of concept, we apply the above methodology to simulated data. In particular, we consider the Ornstein–Uhlenbeck process X in \mathbb{R}^2 governed by

$$dX_t = -\frac{X_t}{10} dt + \sqrt{2} dW_t,$$

and simulate data from this model until time T_{end} for increasing values of T_{end} , corresponding to increasing periods of exploration. For each simulation, we then estimate the invariant density ρ (here a normal density) via a discretized version of the kernel density estimator as well as its gradient $\nabla \rho$ via the gradient of the kernel estimator. Finally, setting $f = |\cdot|$ and $\kappa = 1$, we use these estimates along with Corollary 3.2 to find approximately optimal star-shaped polygons, and plot them in Figure 4.2 along with the numerically approximated optimal shape, i.e. the one found using Corollary 3.2 with the true density ρ .

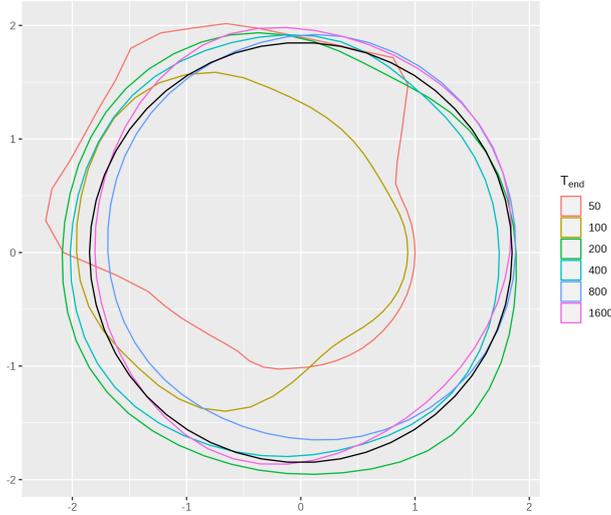


Figure 4.1: Left: Estimates of the optimal shape (black) using kernel estimates after increasing periods of exploration. Notably, after only $T = 150$, the estimated optimal shape has an associated cost only 0.61% higher than the true optimum.

Episodic domain learning Aiming now at strategies with sublinear regret rates without any simplifying assumptions on the data-collection mechanism, we face a classical exploration vs. exploitation dilemma in light of the necessity to simultaneously control the process and estimate its dynamics over time. The key bound from Proposition 4.6 will allow us to do so.

Our episodic learning algorithm separates the time-line into exploration and exploitation phases. Let T_i be the start of the i -th exploration period, where we let the diffusion run freely without reflection and S_i be the start of the i -th exploitation period, where we reflect the process according to an estimate of the optimal reflection boundary based on past observations of the exploration process. We always start with an exploration period, i.e. $T_1 = 0$, and then alternate between exploration and exploitation periods. We denote by $\tau_i = S_i - T_i$ the length of the i -th exploration period and by $\sigma_i = T_{i+1} - S_i$ the length of the i -th exploitation period.

Contrary to the scalar diffusion case in [12], the multivariate diffusion does not hit points, which makes it difficult to introduce an appropriate life-cycle decomposition of the exploration process that allows for elegant renewal theoretic arguments in the analysis. Instead we choose sequences $(a_i) \subset [1, \infty)^{\mathbb{N}}$, $(b_i) \subset [1, \infty)^{\mathbb{N}}$ and simply let $S_i = \inf\{t \geq T_i + a_i : \tilde{X}_t \in \text{cl}(B(0, \underline{\lambda}))\}$ and $T_{i+1} = S_i + b_i$, where $(\tilde{X}_t)_{t \geq 0}$ denotes the process that is controlled according to the above strategy. This implies that the i -th exploitation length is deterministically given by $\sigma_i = b_i$ and due to the strongly recurrent behavior of the process the i -th exploration length is relatively close to a_i with high probability. Moreover, the strategy makes sure that at the start of an exploitation period we have $\tilde{X}_{S_i} \in \text{cl} B(0, \underline{\lambda}) \subset \bar{D}$ for any $D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)$. For the estimator of the reflection boundary \hat{D}_i in the i -th exploitation period, we only take into account the observations gathered from the last exploration period on the time interval $[T_i, T_i + a_i]$ by letting

$$\tilde{\rho}_i^*(x) := (\hat{\rho}_{i,2}(x) \wedge 2\bar{\rho}) \vee \underline{\rho}/2, \quad x \in \mathbb{R}^d,$$

where $\hat{\rho}_{i,2}$ is the adaptive invariant density estimator based on the diffusion data $(\tilde{X}_t)_{t \in [T_i, T_i + a_i]}$ for

the parameter choice $q = 2$ in the construction of the stochastic bandwidth, and then set

$$\widehat{D}_i \in \arg \min_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} \frac{1}{\int_D \widehat{\rho}_i^*(y) dy} \left(\int_D f(y) \widehat{\rho}_i^*(y) dy + \kappa \int_{\partial D} \widehat{\rho}_i^*(y) \mathcal{H}^{d-1}(dy) \right).$$

Let $\widetilde{\mathbb{F}} = (\widetilde{\mathcal{F}}_t)_{t \geq 0}$ be the filtration generated by the controlled process \widetilde{X} and set $\widetilde{C}_{a,b} := \int_a^b f(\widetilde{X}_s) ds + \kappa(\widetilde{L}_b - \widetilde{L}_a)$ as the costs on the time interval $[a, b]$ associated to \widetilde{X} , where \widetilde{L} is the local time on the reflection boundaries during the exploitation phases and is set equal to zero during the exploration phases. We also let $C_{a,b}(x, D)$ be the costs on the time interval $[a, b]$ associated to a Langevin diffusion $Z^{x,D}$ that is driven by a Brownian motion independent of $\widetilde{\mathbb{F}}$ and that is reflected in D and is started in x . Denote by $\tau(x, D)$ its first hitting time of $\text{cl}(B(0, \underline{\lambda}))$. Furthermore, we set $n(T) := \min\{i \in \mathbb{N} : \sum_{j=1}^i (a_j + b_j) \geq T\}$ and note that

$$n(T) \geq \min \left\{ i \in \mathbb{N} : \sum_{j=1}^i (\tau_j + \sigma_j) \geq T \right\} = \min\{i \in \mathbb{N} : T_{i+1} \geq T\} = \min\{i \in \mathbb{N} : T_i \geq T\} - 1,$$

which in particular implies that $S_{n(T)+1} > T_{n(T)+1} \geq T$.

For technical reasons, we assume that the potential V satisfies the following drift condition: for some constants $r, M > 0$ it holds

$$\forall |x| \geq M : \langle \nabla V(x), x/|x| \rangle \geq r. \quad (4.3)$$

Due to the specific structure of the generator, it is well known, see e.g. [3], that the Langevin diffusion then satisfies a Poincaré inequality (PI) and that its generator has a Lyapunov function $V \geq 1$ that is locally bounded, is for some $a, R > 0$ given by $V(x) = \exp(a|x|)$ for all $|x| \geq R$, and satisfies $\pi(V) < \infty$. This implies, see [19, Theorem 5.2, Theorem 7.2], that X is V -uniformly ergodic in the sense that for some constant $b > 0$,

$$\sup_{|g| \leq V} |P_t g(x) - \pi(g)| \lesssim V(x) \exp(-bt), \quad x \in \mathbb{R}^d, \quad (4.4)$$

and that for any set C s.t. $\lambda(C) > 0$, it holds

$$\mathbb{E}^x \left[\int_0^{\tau_C} V(X_s) ds \right] \leq c(C)V(x), \quad x \in \mathbb{R}^d, \quad (4.5)$$

where $c(C)$ is a constant depending on C and τ_C is the first hitting time of C (we note that a much stronger statement than (4.5) is true, but this won't be needed in our context). We also need some assumptions on the set of viable reflection domains $\Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)$ that allow sufficient uniform bounds in the following. More precisely, we assume the constants $C'(D)$ appearing in Theorem 2.4 to be uniformly bounded in D , that is,

$$\sup_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} C'(D) < \infty. \quad (4.6)$$

Note that this assumption boils down to uniform lower bounds on the transition densities, cf. Lemma 2.2, and uniform bounds on the maxima of the functions φ_D and their partial derivatives described in Lemma 2.5, i.e., certain uniform regularity assumptions on the boundaries. Moreover, we also require a uniform upper bound on the transition densities in the form

$$\sup_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} \sup_{t \geq 1, x, y \in \overline{D}} p_t^D(x, y) < \infty. \quad (4.7)$$

By continuity of p_t^D , it follows from the Gaussian upper bound on $p_t^D(x, y)$ given in [42, Corollary 6.15] for a.e. $(x, y) \in D^2$ that this holds pointwise, i.e., $\sup_{t \geq 1, x, y \in \bar{D}} p_t^D(x, y) < \infty$ for any $D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)$. Verifying (4.6) and (4.7) is highly problem specific and is a difficult task with the tools available in the literature when $\Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)$ is infinite. Still, these assumptions are not unreasonable provided that appropriate uniform regularity conditions on the boundaries are in force. With this technical preparation we can prove our final theorem.

Theorem 4.7. *Suppose that the non-reflected Langevin diffusion satisfies (4.3) and (NI) and that its invariant density satisfies $\rho_V \in \mathcal{H}_d(\boldsymbol{\beta} + \mathbf{1}, \mathcal{L})$. Assume also that $f \in L^2(\pi)$, $f \leq V$, that the initial distribution of the first exploration phase μ satisfies $\mu \ll \pi$ and $d\mu/d\pi \in L^2(\pi)$ and that (4.6) and (4.7) hold. Then, the average regret per time unit is bounded by*

$$\frac{1}{T} \mathbb{E}[\tilde{C}_{0,T}] - J(D^*) \lesssim \frac{1}{T} \left(\sum_{i=1}^{n(T)} a_i + \sum_{i=1}^{n(T)} b_i \Psi_{d,\beta}(a_i) \right).$$

Proof. Without loss of generality, let $T \geq 1$. Using that the costs are nonnegative, we have

$$\begin{aligned} \mathbb{E}[\tilde{C}_{0,T}] &= \mathbb{E} \left[\sum_{T_i \leq T} \tilde{C}_{T_i, S_i \wedge T} \right] + \mathbb{E} \left[\sum_{S_i \leq T} \tilde{C}_{S_i, T_{i+1} \wedge T} \right] \\ &\leq \sum_{i=1}^{n(T)} \mathbb{E}[\tilde{C}_{T_i, S_i}] + \sum_{i=1}^{n(T)} \mathbb{E}[\tilde{C}_{S_i \wedge T, T_{i+1} \wedge T}] \\ &\leq \sum_{i=1}^{n(T)} (\mathbb{E}[C_{0, a_i}(\tilde{X}_{T_i}, \mathbb{R}^d)] + \mathbb{E}[C_{0, \tau(\tilde{X}_{T_i+a_i}, \mathbb{R}^d)}(\tilde{X}_{T_i+a_i}, \mathbb{R}^d)]) + \sum_{i=1}^{n(T)-1} \mathbb{E}[C_{0, b_i}(\tilde{X}_{S_i}, \hat{D}_i)] \\ &\quad + \mathbb{E}[C_{0, (T-S_{n(T)}) \vee 1}(\tilde{X}_{S_{n(T)}}, \hat{D}_{n(T)})]. \end{aligned} \tag{4.8}$$

We start by bounding the second term, associated to the exploitation periods until the $(n(T) - 1)$ -th episode. By conditioning on $\tilde{\mathcal{F}}_{S_i}$, we see that

$$\begin{aligned} & \left| \mathbb{E}[C_{0, b_i}(\tilde{X}_{S_i}, \hat{D}_i)] - b_i \mathbb{E}[J(\hat{D}_i)] \right| \\ & \leq b_i \int_{\Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} \int_{\text{cl} B(0, \underline{\lambda})} \left| \mathbb{E}^x \left[\frac{1}{b_i} \int_0^{b_i} f(X_s^D) ds + \kappa L_{b_i}^D - J(D) \right] \right| \mathbb{P}(\tilde{X}_{S_i} \in dx, \hat{D}_i \in dD) \\ & \leq \sup_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} C'(D) < \infty, \end{aligned} \tag{4.9}$$

where the last two lines follow from Theorem 2.4 and (4.6). Observe now that on the previous data collection interval $[T_i, T_i + a_i]$ the process \tilde{X} is equal in law to the Langevin diffusion with potential V started according to the law $\mathbb{P}_{\tilde{X}_{T_i}}$. For $i = 1$, the latter has, by assumption on μ , a Radon–Nikodym derivative w.r.t. the invariant distribution π that lies in $L^2(\pi)$. Extend the transition densities p_t^D from \bar{D}^2 to $\mathbb{R}^d \times \mathbb{R}^d$ by setting $p_t^D(x, y) = 0$ for $x, y \notin \bar{D}$. For $i \geq 2$, we then observe that for any $y \in \text{cl}(B(0, \bar{\lambda}))$ we have

$$\begin{aligned} \frac{d\mathbb{P}(\tilde{X}_{T_i} \in \cdot)}{d\lambda}(y) &= \int_{\Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} \int_{\text{cl} B(0, \underline{\lambda})} p_{b_i}^D(x, y) \mathbb{P}(\tilde{X}_{S_{i-1}} \in dx, \hat{D}_i \in dD) \\ &\leq \sup_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} \sup_{t \geq 1, (x, y) \in \bar{D}^2} p_t^D(x, y) < \infty, \end{aligned}$$

where we used (4.7). Since the Lebesgue density of π is bounded away from zero on $B(0, \bar{\lambda})$, this now implies

$$\sup_{i \in \mathbb{N}} \left\| \frac{d\mathbb{P}(\tilde{X}_{T_i} \in \cdot)}{d\pi} \right\|_{L^2(\pi)} < \infty. \tag{4.10}$$

Proposition 4.6 therefore yields that

$$\mathbb{E}[|J(D^*) - J(\widehat{D}_i)|] \lesssim \Psi_{d,\beta}(a_i).$$

In summary, the above estimates and the triangle inequality therefore yield

$$\sum_{i=1}^{n(T)-1} \mathbb{E}[C_{0,b_i}(\widetilde{X}_{S_i}, \widehat{D}_i)] \lesssim n(T) + \sum_{i=1}^{n(T)-1} b_i(\Psi_{d,\beta}(a_i) + J(D^*)). \quad (4.11)$$

We turn to the last summand in (4.8). Similarly to (4.9) we see that

$$\left| \mathbb{E}[C_{0,(T-S_{n(T)})\vee 1}(\widetilde{X}_{S_{n(T)}}, \widehat{D}_{n(T)})] - \mathbb{E}[(T - S_{n(T)}) \vee 1] J(\widehat{D}_{n(T)}) \right| \leq \sup_{D \in \Theta(\underline{\lambda}, \bar{\lambda}, \Lambda)} C'(D) < \infty, \quad (4.12)$$

and using (4.10) and Proposition 4.6 we have

$$\mathbb{E}[(T - S_{n(T)}) \vee 1] |J(D^*) - J(\widehat{D}_{n(T)})| \leq b_{n(T)} \mathbb{E}[|J(D^*) - J(\widehat{D}_{n(T)})|] \lesssim b_{n(T)} \Psi_{d,\beta}(a_{n(T)}). \quad (4.13)$$

Combining (4.11), (4.12) and (4.13) we finally arrive at

$$\begin{aligned} & \sum_{i=1}^{n(T)-1} \mathbb{E}[C_{0,b_i}(\widetilde{X}_{S_i}, \widehat{D}_i)] + \mathbb{E}[C_{0,(T-S_{n(T)})\vee 1}(\widetilde{X}_{S_{n(T)}}, \widehat{D}_{n(T)})] \\ & \leq n(T) + \sum_{i=1}^{n(T)} b_i \Psi_{d,\beta}(a_{n(T)}) + J(D^*) \left(\sum_{i=1}^{n(T)-1} b_i + \mathbb{E}[(T - S_{n(T)}) \vee 1] \right) \\ & \leq n(T) + TJ(D^*) + \sum_{i=1}^{n(T)} b_i \Psi_{d,\beta}(a_i). \end{aligned} \quad (4.14)$$

Let us now treat the exploration periods. Recall that since the unreflected diffusion satisfies a Poincaré inequality, we have exponentially fast convergence of its semigroup in $L^2(\pi)$. Since moreover $f \in L^2(\pi)$, the combined statements of Theorem 3.1 and Corollary 3.2 in [9] imply that for $\widetilde{f} := f - \mu(f)$ and $g := -\int_0^\infty P_s \widetilde{f} ds$, we have $g \in \mathcal{D}(L)$ and

$$\left\| \int_0^t (f - \pi(f))(X_s) ds \right\|_{L^2(\mathbb{P}^\pi)}^2 \leq Ct \|\sqrt{-L}g\|_\pi^2,$$

for some constant C that is independent of $t \geq 0$. Using (4.10) and Hölder inequality therefore implies

$$\begin{aligned} \left| \mathbb{E}[C_{0,a_i}(\widetilde{X}_{T_i}, \mathbb{R}^d)] - a_i \pi(f) \right| &= \left| \mathbb{E} \left[\mathbb{E}^{\widetilde{X}_{T_i}} \left[\int_0^{a_i} (f - \pi(f))(X_s) ds \right] \right] \right| \\ &\leq \int_{\mathbb{R}^d} \mathbb{E}^x \left[\left\| \int_0^{a_i} (f - \pi(f))(X_s) ds \right\| \frac{d\mathbb{P}(\widetilde{X}_{T_i} \in \cdot)}{d\pi}(x) \pi(dx) \right] \\ &\leq \left\| \frac{d\mathbb{P}(\widetilde{X}_{T_i} \in \cdot)}{d\pi} \right\|_{L^2(\pi)} \left\| \int_0^{a_i} (f - \pi(f))(X_s) ds \right\|_{L^2(\mathbb{P}^\pi)} \\ &\leq C' \sqrt{a_i}. \end{aligned}$$

It therefore follows by the triangle inequality that

$$\sum_{i=1}^{n(T)} \mathbb{E}[C_{0,a_i}(\widetilde{X}_{T_i}, \mathbb{R}^d)] \lesssim \sum_{i=1}^{n(T)} a_i. \quad (4.15)$$

Furthermore, using (4.5) and the assumption $f \lesssim V$, we can write

$$\mathbb{E}[C_{0,\tau}(\tilde{X}_{T_i+a_i}, \mathbb{R}^d)(\tilde{X}_{T_i+a_i}, \mathbb{R}^d)] = \int_{\mathbb{R}^d} \mathbb{E}^x \left[\int_0^{\tau_{\text{cl}(B(0,\bar{\lambda}))}} f(X_s) ds \right] \mathbb{P}(\tilde{X}_{T_i+a_i} \in dx) \leq \mathbb{E}[V(\tilde{X}_{T_i+a_i})].$$

Now since $\tilde{X}_{T_i+a_i}$ has the same law as X_{a_i} with X started according to the law $\mathbb{P}_{\tilde{X}_{T_i}}$ which, by construction, is supported on $\text{cl}(B(0,\bar{\lambda}))$, it follows from (4.4) that

$$\begin{aligned} \mathbb{E}[V(\tilde{X}_{T_i+a_i})] &\leq \sup_{x \in \text{cl}(B(0,\bar{\lambda}))} \mathbb{E}^x[V(X_{a_i})] \leq \pi(V) + \sup_{x \in \text{cl}(B(0,\bar{\lambda}))} |P_{a_i}V(x) - \pi(V)| \\ &\leq \pi(V) + C\|V\|_{L^\infty(\text{cl}(B(0,\bar{\lambda})))} < \infty, \end{aligned}$$

for some constant independent of $i \in \mathbb{N}$. This allows us to conclude that

$$\sum_{i=1}^{n(T)} \mathbb{E}[C_{0,\tau}(\tilde{X}_{T_i+a_i}, \mathbb{R}^d)(\tilde{X}_{T_i+a_i}, \mathbb{R}^d)] \leq (\pi(V) + C\|V\|_{L^\infty(\text{cl}(B(0,\bar{\lambda})))})n(T). \quad (4.16)$$

Taking together the bounds (4.14), (4.15) and (4.16) now shows that the average regret per time unit is bounded by

$$\frac{1}{T} \mathbb{E}[\tilde{C}_{0,T}] - J(D^*) \lesssim \frac{1}{T} \left(\sum_{i=1}^{n(T)} a_i + \sum_{i=1}^{n(T)} b_i \Psi_{d,\beta}(a_i) \right).$$

■

As a consequence, we obtain the following explicit rates for a strategy that doubles exploration times and chooses subsequent exploitation times inverse proportionally to the nonparametric estimation rate.

Corollary 4.8. *Let $a_i = 2^i$ and $b_i = a_i/\Psi_{d,\beta}(a_i)$ for $i \in \mathbb{N}$. Given the assumptions of Theorem 4.7 it holds*

$$\frac{1}{T} \mathbb{E}[\tilde{C}_{0,T}] - J(D^*) \lesssim \begin{cases} \left(\frac{(\log T)^2}{T} \right)^{\frac{1}{3}}, & d = 2, \\ \left(\frac{\log T}{T} \right)^{\frac{\beta+1}{3\beta+1+d-2}}, & d \geq 3, \end{cases}$$

for the corresponding reflection strategy with exploration periods of length a_i and exploitation periods of length b_i .

Proof. Let $d \geq 3$ and denote $\alpha := \overline{\beta+1}/(2\overline{\beta+1}+d-2)$. Let $T \in [\sum_{i=1}^n (a_i + b_i), \sum_{i=1}^{n+1} (a_i + b_i)]$. Then, $T \sim 2^{n(\alpha+1)}/n^\alpha$ and

$$\sum_{i=1}^{n(T)} a_i + \sum_{i=1}^{n(T)} b_i \Psi_{d,\beta}(a_i) \leq 2 \sum_{i=1}^n a_i \sim 2^n,$$

which, by Theorem 4.7 yields

$$\frac{1}{T} \mathbb{E}[\tilde{C}_{0,T}] - J(D^*) \lesssim (n2^{-n})^\alpha = (nn^\alpha 2^{-n(\alpha+1)})^{\alpha/(\alpha+1)} \sim (\log T/T)^{\alpha/(\alpha+1)} = (\log T/T)^{\frac{\beta+1}{3\beta+1+d-2}}.$$

The claim for $d = 2$ is proved analogously. ■

The rate loss of the strategy's regret per time unit relative to the static regret from Proposition 4.6 provides a natural analogue to the regret bounds from [12, Theorem 2.5] in the one-dimensional case, even though the construction of the strategies substantially differs. Let us also remark that *doubling tricks* in the strategy design that make sure that the number of episodes at

time T is of order $\log T$, are commonly encountered in algorithms with verifiable optimal regret rates for undiscounted reinforcement learning problems. For instance, the popular UCRL2 algorithm in [2] recomputes policies as soon as the occurrence of a state-action pair has doubled. Such strategies have the drawback of choosing suboptimal policies for arbitrarily long periods of time (here, not reflecting at all), see [8]. Recently, [7] have proposed the *regret of exploration* as an appropriate measure to capture such inefficiencies. It is an interesting and challenging question for future work to adapt reflection strategies based on the nonparametric approach advocated in this paper to such finer-grained performance measures.

A. Remaining proofs

Proof of Lemma 2.1. We show this using [28, Theorem 1]. Since we consider only smooth sets D and normal reflections, the set of test functions \mathcal{H} as defined in [28] simplifies to

$$\mathcal{H} = \{f \in \mathcal{C}_c^2(\bar{D}) \oplus \mathbb{R} : \langle n(x), \nabla f(x) \rangle \geq 0 \text{ for } x \in \partial D\}.$$

Letting $\pi(dx) = \rho_D(x)dx$, we note $\pi(\partial D) = 0$, and thus it suffices to check

$$\int_D Af(x) \pi(dx) \leq 0, \quad f \in \mathcal{H},$$

where A is the differential operator given by

$$Af(x) = -\langle \nabla V(x), \nabla f(x) \rangle + \Delta f(x), \quad f \in \mathcal{C}^2(\mathbb{R}^d), x \in \mathbb{R}^d. \quad (\text{A.1})$$

To this end, let $f \in \mathcal{H}$ be given, and we find using the divergence theorem

$$\begin{aligned} \int_D Af(x) \pi_D(dx) &= \int_D -\langle \nabla V(x), \nabla f(x) \rangle \rho_D(x) dx + \int_D \Delta f(x) \rho_D(x) dx \\ &= \int_D \langle \nabla f(x), \nabla \rho_D(x) \rangle dx + \int_D \Delta f(x) \rho_D(x) dx \\ &= \left(- \int_{\partial D} \langle \nabla f(x), n(x) \rangle \rho_D(x) \mathcal{H}^{d-1}(dx) - \int_D \Delta f(x) \rho_D(x) dx \right) \\ &\quad + \int_D \Delta f(x) \rho_D(x) dx \\ &= - \int_{\partial D} \langle \nabla f(x), n(x) \rangle \rho_D(x) \mathcal{H}^{d-1}(dx), \end{aligned}$$

where \mathcal{H}^{d-1} denotes the $(d-1)$ -dimensional Hausdorff measure. Since $f \in \mathcal{H}$, we have $\langle \nabla f, n \rangle \geq 0$, and hence $\int_D Af(x) \pi_D(dx) \leq 0$ as desired. \blacksquare

Proof of Lemma 2.2. Since p_t^D is continuous on \bar{D}^2 , the process X^D is a strong Feller process and, if we denote by λ_D the restriction of the Lebesgue measure to \bar{D} , (2.1) ensures that

$$P_1^D(x, \cdot) \geq \delta \lambda_D, \quad x \in \bar{D}. \quad (\text{A.2})$$

Consequently, the compact state space \bar{D} is small and hence petite in the sense of [38]. Using [36, Theorem 1.1], it follows that X^D is Harris recurrent. This implies that X^D has a unique invariant distribution π_D , s.t. $\pi_D(dx) = \rho_D(x) dx$ follows from Lemma 2.1. Moreover, the Doeblin recurrence property (A.2) implies by [37, Theorem 16.2.4] uniform ergodicity of the 1-skeleton with explicit constants, that is

$$\|P_n^D(x, \cdot) - \pi_D\|_{\text{TV}} \leq 2e^{\log(1-\delta\lambda(D))n}, \quad x \in \bar{D}.$$

By stationarity of π_D it is easy to see that $\|P_t^D(x, \cdot) - \pi_D\|_{\text{TV}}$ is decreasing in t , whence,

$$\|P_t^D(x, \cdot) - \pi_D\|_{\text{TV}} \leq \|P_{\lfloor t \rfloor}^D(x, \cdot) - \pi_D\|_{\text{TV}} \leq 2e^{\log(1-\delta\lambda(D))\lfloor t \rfloor} \leq \frac{2}{1-\delta\lambda(D)} e^{\log(1-\delta\lambda(D))t}.$$

■

Proof of Corollary 2.3. Due to the exponential ergodicity of X^D , it follows from [17, Corollary 3.3] that

$$\frac{1}{t} \mathbb{E}^{\pi_D} \left[\left| \int_0^t (h(X_s^D) - \pi_D(h)) \, ds \right| \right] \lesssim \frac{\|h\|_{L^\infty(\bar{D})}}{\sqrt{t}}. \quad (\text{A.3})$$

To get the result for deterministic initial condition $X_0^D = x \in \bar{D}$, we note that for $t \geq 1$ and $\tilde{h} := h - \pi_D(h)$ and $g_t(y) := t^{-1} \mathbb{E}^y \left[\left| \int_0^{t-\sqrt{t}} h(X_s^D) \, ds \right| \right]$, we obtain

$$\begin{aligned} & \left| \mathbb{E}^x \left[\left| \frac{1}{t} \int_0^t \tilde{h}(X_s^D) \, ds \right| \right] - \mathbb{E}^{\pi_D} \left[\left| \frac{1}{t} \int_0^t \tilde{h}(X_s^D) \, ds \right| \right] \right| \\ & \leq 2\|\tilde{h}\|_{L^\infty(\bar{D})} \frac{1}{\sqrt{t}} + \left| \mathbb{E}^x \left[\left| \frac{1}{t} \int_0^{t-\sqrt{t}} \tilde{h}(X_{s+\sqrt{t}}^D) \, ds \right| \right] - \mathbb{E}^{\pi_D} \left[\left| \frac{1}{t} \int_0^{t-\sqrt{t}} \tilde{h}(X_{s+\sqrt{t}}^D) \, ds \right| \right] \right| \\ & = 2\|\tilde{h}\|_{L^\infty(\bar{D})} \frac{1}{\sqrt{t}} + |\mathbb{E}^x[g_t(X_{\sqrt{t}})] - \pi_D(g_t)| \\ & \leq 2\|\tilde{h}\|_{L^\infty(\bar{D})} \left(\frac{1}{\sqrt{t}} + \|P_{t^{1/2}}^D(x, \cdot) - \pi_D\|_{\text{TV}} \right) \\ & \lesssim \|h\|_{L^\infty(\bar{D})} \frac{1}{\sqrt{t}}, \end{aligned} \quad (\text{A.4})$$

where we used the Markov property and stationarity of π_D for the second line, $\|g_t\|_\infty \leq \|\tilde{h}\|_{L^\infty(\bar{D})}$ for the third line and Lemma 2.2 for the last line. The claim now follows from combining (A.3) and (A.4) with the triangle inequality. ■

Acknowledgements AHT is grateful for financial funding by Villum Synergy Programme, project no. 50099. LT gratefully acknowledges financial support of Carlsberg Foundation Young Researcher Fellowship grant CF20-0640.

References

- [1] L. H. Alvarez. *A Class of Solvable Stationary Singular Stochastic Control Problems*. 2018. arXiv: [1803.03464](https://arxiv.org/abs/1803.03464).
- [2] P. Auer, T. Jaksch, and R. Ortner. “Near-optimal Regret Bounds for Reinforcement Learning”. In: *Advances in Neural Information Processing Systems*. Vol. 21. 2008.
- [3] D. Bakry, P. Cattiaux, and A. Guillin. “Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré”. In: *J. Funct. Anal.* 254.3 (2008), pp. 727–759. DOI: [10.1016/j.jfa.2007.11.002](https://doi.org/10.1016/j.jfa.2007.11.002).
- [4] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Vol. 348. Grundlehren der mathematischen Wissenschaften. Springer, Cham, 2014, pp. xx+552. DOI: [10.1007/978-3-319-00227-9](https://doi.org/10.1007/978-3-319-00227-9).
- [5] M. Basei, X. Guo, A. Hu, and Y. Zhang. “Logarithmic regret for episodic continuous-time linear-quadratic reinforcement learning over a finite-time horizon”. In: *Journal of Machine Learning Research* 23.178 (2022), pp. 1–34.

- [6] F. Boetius and M. Kohlmann. “Connections between optimal stopping and singular stochastic control”. In: *Stochastic Process. Appl.* 77.2 (1998), pp. 253–281. DOI: [10.1016/S0304-4149\(98\)00049-0](https://doi.org/10.1016/S0304-4149(98)00049-0).
- [7] V. Boone and B. Gaujal. “The Regret of Exploration and the Control of Bad Episodes in Reinforcement Learning”. In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. PMLR, 2023, pp. 2824–2856.
- [8] A. N. Burnetas and M. N. Katehakis. “Optimal adaptive policies for Markov decision processes”. In: *Math. Oper. Res.* 22.1 (1997), pp. 222–255. DOI: [10.1287/moor.22.1.222](https://doi.org/10.1287/moor.22.1.222).
- [9] P. Cattiaux, D. Chafaï, and A. Guillin. “Central limit theorems for additive functionals of ergodic Markov diffusions processes”. In: *ALEA Lat. Am. J. Probab. Math. Stat.* 9.2 (2012), pp. 337–382.
- [10] S. Christensen, E. Mordecki, and F. O. Eguren. *Two sided ergodic singular control and mean field game for diffusions*. 2023. arXiv: [2306.09263](https://arxiv.org/abs/2306.09263).
- [11] S. Christensen and C. Strauch. “Nonparametric learning for impulse control problems—Exploration vs. exploitation”. In: *Ann. Appl. Prob.* 33.2 (2023), pp. 1569–1587. DOI: [10.1214/22-AAP1849](https://doi.org/10.1214/22-AAP1849).
- [12] S. Christensen, C. Strauch, and L. Trottner. “Learning to reflect: A unifying approach for data-driven stochastic control strategies”. In: *Bernoulli* (to appear).
- [13] M. H. A. Davis and M. Zervos. “A pair of explicitly solvable singular stochastic control problems”. In: *Appl. Math. Optim.* 38.3 (1998), pp. 327–352. DOI: [10.1007/s002459900094](https://doi.org/10.1007/s002459900094).
- [14] T. De Angelis, G. Ferrari, and J. Moriarty. “A solvable two-dimensional degenerate singular stochastic control problem with nonconvex costs”. In: *Math. Oper. Res.* 44.2 (2019), pp. 512–531. DOI: [10.1287/moor.2018.0934](https://doi.org/10.1287/moor.2018.0934).
- [15] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. “On the sample complexity of the linear quadratic regulator”. In: *Found. Comput. Math.* 20.4 (2020), pp. 633–679. DOI: [10.1007/s10208-019-09426-y](https://doi.org/10.1007/s10208-019-09426-y).
- [16] M. C. Delfour and J.-P. Zolésio. *Shapes and geometries*. Second. Vol. 22. Advances in Design and Control. Metrics, analysis, differential calculus, and optimization. SIAM, Philadelphia, PA, 2011, pp. xxiv+622. DOI: [10.1137/1.9780898719826](https://doi.org/10.1137/1.9780898719826).
- [17] N. Dexheimer, C. Strauch, and L. Trottner. “Adaptive invariant density estimation for continuous-time mixing Markov processes under sup-norm risk”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 58.4 (2022), pp. 2029–2064. DOI: [10.1214/21-aihp1235](https://doi.org/10.1214/21-aihp1235).
- [18] J. Dianetti and G. Ferrari. “Multidimensional singular control and related Skorokhod problem: Sufficient conditions for the characterization of optimal controls”. In: *Stochastic Process. Appl.* 162 (2023), pp. 547–592. DOI: [10.1016/j.spa.2023.05.006](https://doi.org/10.1016/j.spa.2023.05.006).
- [19] D. Down, S. P. Meyn, and R. L. Tweedie. “Exponential and uniform ergodicity of Markov processes”. In: *Ann. Probab.* 23.4 (1995), pp. 1671–1691.
- [20] T. E. Duncan, L. Guo, and B. Pasik-Duncan. “Adaptive continuous-time linear quadratic Gaussian control”. In: *IEEE Trans. Automat. Control* 44.9 (1999), pp. 1653–1662. DOI: [10.1109/9.788532](https://doi.org/10.1109/9.788532).
- [21] S. Federico, G. Ferrari, and P. Schuhmann. “A singular stochastic control problem with interconnected dynamics”. In: *SIAM J. Control Optim.* 58.5 (2020), pp. 2821–2853. DOI: [10.1137/19M1296288](https://doi.org/10.1137/19M1296288).
- [22] S. Federico, G. Ferrari, and P. Schuhmann. “Singular control of the drift of a Brownian system”. In: *Appl. Math. Optim.* 84 (2021), S561–S590. DOI: [10.1007/s00245-021-09779-3](https://doi.org/10.1007/s00245-021-09779-3).

- [23] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*. 2nd ed. Vol. 25. Stochastic Modelling and Applied Probability. Springer, New York, 2006, pp. xviii+429.
- [24] X. Guo, A. Hu, and Y. Zhang. *Reinforcement learning for linear-convex models with jumps via stability analysis of feedback controls*. 2021. arXiv: [2104.09311](https://arxiv.org/abs/2104.09311).
- [25] X. Guo and P. Tomecek. “A class of singular control problems and the smooth fit principle”. In: *SIAM J. Control Optim.* 47.6 (2008), pp. 3076–3099. DOI: [10.1137/070685336](https://doi.org/10.1137/070685336).
- [26] J. He, D. Zhou, and Q. Gu. “Logarithmic regret for reinforcement learning with linear function approximation”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 4171–4180.
- [27] S. Itô. *Diffusion equations*. Vol. 114. Translations of Mathematical Monographs. American Mathematical Society, Providence, RI, 1992, pp. x+225. DOI: [10.1090/mmono/114](https://doi.org/10.1090/mmono/114).
- [28] W. Kang and K. Ramanan. “Characterization of stationary distributions of reflected diffusions”. In: *Ann. Appl. Probab.* 24.4 (2014), pp. 1329–1374. DOI: [10.1214/13-AAP947](https://doi.org/10.1214/13-AAP947).
- [29] S. G. Krantz and H. R. Parks. “Distance to C^k hypersurfaces”. In: *J. Differential Equations* 40.1 (1981), pp. 116–120. DOI: [10.1016/0022-0396\(81\)90013-9](https://doi.org/10.1016/0022-0396(81)90013-9).
- [30] L. Kruk. “Optimal policies for n -dimensional singular stochastic control problems. I. The Skorokhod problem”. In: *SIAM J. Control Optim.* 38.5 (2000), pp. 1603–1622. DOI: [10.1137/S0363012998347535](https://doi.org/10.1137/S0363012998347535).
- [31] L. Kruk. “Optimal policies for n -dimensional singular stochastic control problems. II. The radially symmetric case. Ergodic control”. In: *SIAM J. Control Optim.* 39.2 (2000), pp. 635–659. DOI: [10.1137/S0363012998347547](https://doi.org/10.1137/S0363012998347547).
- [32] S. Kumar and K. Muthuraman. “A numerical method for solving singular stochastic control problems”. In: *Oper. Res.* 52.4 (2004), pp. 563–582. DOI: [10.1287/opre.1030.0107](https://doi.org/10.1287/opre.1030.0107).
- [33] H. J. Kushner. “Numerical methods for stochastic control problems in continuous time”. In: *SIAM J. Control Optim.* 28.5 (1990), pp. 999–1048.
- [34] O. Lepski. “Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure”. In: *Ann. Statist.* 41.2 (2013), pp. 1005–1034. DOI: [10.1214/13-AOS1109](https://doi.org/10.1214/13-AOS1109).
- [35] P.-L. Lions and A.-S. Sznitman. “Stochastic differential equations with reflecting boundary conditions”. In: *Comm. Pure Appl. Math.* 37.4 (1984), pp. 511–537. DOI: [10.1002/cpa.3160370408](https://doi.org/10.1002/cpa.3160370408).
- [36] S. P. Meyn and R. L. Tweedie. “Generalized resolvents and Harris recurrence of Markov processes”. In: *Doeblin and modern probability (Blaubeuren, 1991)*. Vol. 149. Contemp. Math. Amer. Math. Soc., Providence, RI, 1993, pp. 227–250. DOI: [10.1090/conm/149/01273](https://doi.org/10.1090/conm/149/01273).
- [37] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Second. Cambridge University Press, Cambridge, 2009, pp. xxviii+594. DOI: [10.1017/CB09780511626630](https://doi.org/10.1017/CB09780511626630).
- [38] S. P. Meyn and R. L. Tweedie. “Stability of Markovian processes. II. Continuous-time processes and sampled chains”. In: *Adv. in Appl. Probab.* 25.3 (1993), pp. 487–517. DOI: [10.2307/1427521](https://doi.org/10.2307/1427521).
- [39] E. Neuman and Y. Zhang. *Statistical Learning with Sublinear Regret of Propagator Models*. 2023. arXiv: [2301.05157](https://arxiv.org/abs/2301.05157).
- [40] B. Øksendal and A. Sulem. *Applied stochastic control of jump diffusions*. Third. Universitext. Springer, Cham, 2019, pp. xvi+436. DOI: [10.1007/978-3-030-02781-0](https://doi.org/10.1007/978-3-030-02781-0).

- [41] I. Osband and B. Van Roy. “Model-based reinforcement learning and the eluder dimension”. In: *Advances in Neural Information Processing Systems 27* (2014).
- [42] E. M. Ouhabaz. *Analysis of heat equations on domains*. Vol. 31. London Mathematical Society Monographs Series. Princeton University Press, Princeton, NJ, 2005, pp. xiv+284.
- [43] A. Pilipenko. *An introduction to stochastic differential equations with reflection*. Vol. 1. Universitätsverlag Potsdam, 2014.
- [44] L. Słomiński. “On approximation of solutions of multidimensional SDEs with reflecting boundary conditions”. In: *Stochastic Process. Appl.* 50.2 (1994), pp. 197–219. DOI: [10.1016/0304-4149\(94\)90118-X](https://doi.org/10.1016/0304-4149(94)90118-X).
- [45] H. M. Soner and S. E. Shreve. “Regularity of the value function for a two-dimensional singular stochastic control problem”. In: *SIAM J. Control Optim.* 27.4 (1989), pp. 876–907. DOI: [10.1137/0327047](https://doi.org/10.1137/0327047).
- [46] C. Strauch. “Adaptive invariant density estimation for ergodic diffusions over anisotropic classes”. In: *Ann. Statist.* 46.6B (2018), pp. 3451–3480. DOI: [10.1214/17-AOS1664](https://doi.org/10.1214/17-AOS1664).
- [47] C. Tallec, L. Blier, and Y. Ollivier. “Making deep q-learning methods robust to time discretization”. In: *International Conference on Machine Learning*. PMLR, 2019, pp. 6096–6104.
- [48] M. G. Vieten and R. H. Stockbridge. “Convergence of finite element methods for singular stochastic control”. In: *SIAM J. Control Optim.* 56.6 (2018), pp. 4336–4364. DOI: [10.1137/17M1155119](https://doi.org/10.1137/17M1155119).
- [49] M. G. Vieten and R. H. Stockbridge. “On the solution structure of infinite-dimensional linear problems stemming from singular stochastic control problems”. In: *SIAM J. Control Optim.* 58.6 (2020), pp. 3363–3388. DOI: [10.1137/19M1297415](https://doi.org/10.1137/19M1297415).
- [50] H. Whitney. “Analytic extensions of differentiable functions defined in closed sets”. In: *Trans. Amer. Math. Soc.* 36.1 (1934), pp. 63–89. DOI: [10.2307/1989708](https://doi.org/10.2307/1989708).