NUMERICAL INTEGRATION ON BOUNDED DOMAINS VIA LEARNED REFLECTED DIFFUSIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study high-dimensional integration on bounded convex domains. Given a non-negative integrand r, the goal is to estimate $Z = \int_{\mathcal{D}} r(x)\,\mathrm{d}x$ on $\mathcal{D} \subset \mathbb{R}^d$. Standard Quasi-Monte Carlo does not adapt its evaluation points to the integrand, while effective importance sampling requires a proposal that captures its concentration. Learned diffusion samplers can supply such a proposal, but existing methods are formulated on $\mathbb{R}^d$ and do not respect the domain constraint $x \in \mathcal{D}$. We instead build the proposal from reflected dynamics that stay in $\mathcal{D}$ by construction. A reflected diffusion progressively transforms the target distribution toward the uniform distribution on $\mathcal{D}$, and a neural score model is trained to reverse this process. A change-of-measure formula for reflected diffusions yields a training objective without additional boundary terms and an unbiased importance-sampling estimator of Z. For non-smooth integrands, where reparameterization gradients may be undefined or uninformative, we derive a REINFORCE estimator of the training gradient using Gaussian proposal densities before reflection, without evaluating reflected transition densities or differentiating r. We evaluate on numerous synthetic targets and real-world problems, demonstrating strong performance of the proposed method, as well as favorable scaling with the dimension of the problem and the number of points used for the Monte Carlo estimate.

1 INTRODUCTION

We consider the problem of estimating high-dimensional integrals over bounded constrained domains. Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded convex domain, and let $r : \mathcal{D} \to \mathbb{R}_{\geq 0}$ be measurable and satisfy $\int_{\mathcal{D}} r(x)\,\mathrm{d}x < \infty$. Our goal is to estimate the integral

$$Z = \int_{\mathcal{D}} r(x)\,\mathrm{d}x. \tag{1}$$

Integrals of this form are a computational bottleneck across the sciences. In statistical physics and chemistry, free energies and thermodynamic observables are logarithms of such constants, and learned samplers have been used to estimate them for many-body systems and lattice field theories (Noé et al., 2019; Wirnsberger et al., 2020; Nicoli et al., 2021). In computer graphics, every pixel of a rendered image is an integral over light paths, parameterized so that the integration domain is the unit cube (Müller et al., 2019). In computational finance, the price of a path-dependent derivative is an expectation over a discretized asset path, which the standard inverse-CDF construction turns into an integral over $[0, 1]^d$ with d the number of monitoring dates (Glasserman, 2004; Dick et al., 2013). In Bayesian statistics, the model evidence is an integral of an unnormalized posterior (Kloek & Van Dijk, 1978; Kass & Raftery, 1995), confined to a bounded region whenever the parameters are subject to simplex constraints or under compactly supported priors. These tasks give rise to integration problems on bounded domains, where the location of integrand mass is unknown in advance. This motivates samplers that respect the domain constraints and adapt to where the mass lies, including when it concentrates near faces or corners, or when the integrand vanishes on parts of the domain.

Existing estimators differ in how much they adapt to $r(x)$. Standard Quasi-Monte Carlo (Niederreiter, 1992; Dick et al., 2013, QMC) spreads a deterministic point set over the cube without reference to the integrand. Bayesian quadrature (O'Hagan, 1991; Rasmussen & Ghahramani, 2003) adapts by placing a Gaussian process prior on $r(x)$, but at a cost cubic in the number of evaluations (Briol

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al., 2019). Importance sampling relies on a proposal distribution whose choice determines the estimator's variance (Hammersley & Handscomb, 1964). Neural proposals learn that proposal from evaluations of $r(x)$ alone, either with normalizing flows (Rezende & Mohamed, 2015; Müller et al., 2019) or with diffusion samplers (Zhang & Chen, 2022; Vargas et al., 2023; Berner et al., 2024). The latter avoids the invertibility constraints that limit flow architectures (Kong & Chaudhuri, 2020). Diffusion samplers, however, are formulated on $\mathbb{R}^d$ with a Gaussian reference process throughout the existing work and do not respect the constraint $x \in \mathcal{D}$. A more thorough discussion of prior work on numerical integration methods is presented in Appendix A.

To design a learnable diffusion-based proposal that respects the domain constraint, we draw inspiration from diffusion models (Song et al., 2021) for constrained generative modeling. Existing approaches fall into two groups. The first changes the geometry in which the process evolves: mirror diffusion models map $\mathcal{D}$ onto $\mathbb{R}^d$ through the gradient of a strictly convex potential and run a standard diffusion in the resulting dual space (Liu et al., 2023), while log-barrier diffusions equip the interior of $\mathcal{D}$ with the Riemannian metric induced by the Hessian of a barrier function, under which the process remains in the interior (Fishman et al., 2023). The second group keeps the Euclidean geometry of $\mathcal{D}$ and confines trajectories by reflecting them at the boundary, using reflected Brownian motion as the noising process (Fishman et al., 2023; Lou & Ermon, 2023). In this work, we follow the second route and adapt reflected diffusions from generative modeling to sampling, where no samples from π are available and the reverse dynamics must be learned from evaluations of r alone.

We outline our contributions as follows:

- We propose *reflected diffusion samplers* (RDS) for integration on bounded domains, extending control-based diffusion samplers (Zhang & Chen, 2022; Vargas et al., 2023; Berner et al., 2026) to constrained $\mathcal{D}$. Our construction uses reflected Brownian motion with a time-dependent noise schedule. When initialized from the uniform distribution ν on $\mathcal{D}$, this process preserves uniform marginals (Loper, 2020) and defines a tractable reference path measure for the integral (1). Girsanov's theorem (Le Gall, 2016) then transfers this representation to equivalent controlled reflected diffusions through path-space importance weights. We show that the unique zero-variance ideal proposal corresponds to the same reflected noising dynamics initialized from the normalized target $\pi(x) = r(x)/Z$, and propose to learn its reversal with a neural network-parameterized drift. The learned reversal by construction also acts as the sampler from π.
- We design a training objective for the learned reversal to be the KL divergence to the ideal proposal. For smooth targets, its gradients can be computed via the reparameterization trick. It includes a running control cost with a terminal potential and no boundary term, closely resembling the optimal control objective of Zhang & Chen (2022). We also derive a REINFORCE gradient estimator with a leave-one-out baseline, which uses only the pre-reflection Gaussian transition densities and does not require gradients of $r(x)$, applying to targets that are non-smooth or vanish to zero on part of $\mathcal{D}$, where reparameterization gradients are uninformative. Integrands that vanish on part of the domain are handled by adding a positive offset and correcting the resulting integral estimate.
- We provide rigorous theoretical derivation of all components of the proposed methodology, adapting existing theoretical results to the considered setting of the variance exploding reflected diffusion samplers. In addition, we prove a total-variation accuracy bound for the designed estimator, which is a reflected version of the bound presented in Guo et al. (2026) for diffusion samplers in $\mathbb{R}^d$.
- We evaluate on Genz integrands (Genz, 1984), bounded synthetic targets, Feynman parameter integral (Bonciani et al., 2016) and Asian option pricing (L'Ecuyer, 2009) against Quasi-Monte Carlo, Bayesian quadrature and normalizing flow importance sampling baselines. The experiments demonstrate strong performance of the proposed method, as well as favorable scaling with the dimension of the problem and the number of points used for the estimate. In addition, we compare against PIS (Zhang & Chen, 2022) proposal by mapping $\mathcal{D}$ onto $\mathbb{R}^d$, and demonstrate that the learned RDS proposal achieves lower integration errors against its unconstrained counterpart.

2 Background on Reflected SDE

Denote $\overline{\mathcal{D}}$ and $\partial\mathcal{D}$ to be the closure and boundary of $\mathcal{D}$, and $\mathrm{vol}(\mathcal{D})$ to be its volume. Consider a reflected SDE on $\mathcal{D}$ with drift $f : \overline{\mathcal{D}} \times [0,1] \to \mathbb{R}^d$ and scalar diffusion coefficient $g : [0,1] \to \mathbb{R}$ (Pilipenko, 2014):

$$\mathrm{d}X_t = f(X_t, t)\,\mathrm{d}t + g(t)\,\mathrm{d}W_t + \mathrm{d}L_t. \tag{2}$$

Here W_t is a standard d-dimensional Brownian motion. Under the smooth-boundary assumptions used for the theoretical results, a solution of SDE (2) has the form (X, ℓ), where ℓ is continuous, nondecreasing, starts from zero, and increases only when $X_t \in \partial\mathcal{D}$. Throughout the manuscript, the associated vector reflection term is the derived process $L_t := \int_0^t \mathbf{n}_{\text{in}}(X_s)\, \mathrm{d}\ell_s$, where $\mathbf{n}_{\text{in}}$ is the inward unit normal. Thus $\mathrm{d}L_t$ pushes X inward at boundary times. The formal weak- and strong-solution definitions are given in Section B.1.

Let $\mu_t := \text{Law}(X_t)$ be the marginal law of X_t. Suppose that, for $t > 0$, μ_t admits a density p_t and that these densities have the time, space, and boundary-trace regularity needed for the following identities to hold. Then p_t evolves according to the Fokker–Planck equation with a Neumann boundary condition (Schuss, 2015, Section 2.6.2–2.6.3):

$$\begin{cases} \partial_t p_t = \operatorname{div}\left(-p_t f + \dfrac{g^2(t)}{2}\nabla_x p_t\right), & x \in \mathcal{D},\ t \in (0,1], \\ \mathbf{n}(x)^\top \left(p_t f - \dfrac{g^2(t)}{2}\nabla_x p_t\right) = 0, & x \in \partial\mathcal{D},\ t \in (0,1], \end{cases} \tag{3}$$

where $\mathbf{n}(x)$ is the outward unit normal to $\partial\mathcal{D}$. The boundary condition prevents probability mass from crossing the boundary, so the density evolves entirely within $\mathcal{D}$. The initial law is imposed separately through $\mu_t \Rightarrow \mu_0$ as $t \downarrow 0$, which follows from path continuity. Even when μ_0 has a density p_0, this does not make (3) valid at $t = 0$ without additional initial regularity and boundary compatibility. Without the stated density and regularity, the forward equation is instead interpreted in its weak measure-valued form.

3 Reflected Diffusion Samplers

In this section, we present the main theoretical results and the derivation of the proposed method. Each statement points to a fully quantified statement and proof in the Appendix. The two distributions used in this section are the uniform distribution and the normalized target:

$$\nu(x) := \operatorname{vol}(\mathcal{D})^{-1}, \qquad \pi(x) := \frac{r(x)}{Z}, \qquad x \in \overline{\mathcal{D}}. \tag{4}$$

Path-measure and training theoretical results assume measurable r bounded above and away from zero. The full-interval score representation requires additional regularity. We later discuss zero- and negative-valued targets. See Appendix B.1 for notation and conventions.

3.1 Reference process of the reflected diffusion

Our goal is to construct an importance-sampling estimator for (1) using stochastic dynamics that remain inside the constrained domain $\mathcal{D}$. To this end, we first introduce a simple reference process whose path measure will serve as a base measure with respect to which a proposal is defined. Ideally, this reference process should satisfy two properties. First, it should remain inside $\overline{\mathcal{D}}$ by construction. Second, its marginal distribution should be known and easy to sample from. We obtain both properties with the reflected variance-exploding[1] SDEs (Lou & Ermon, 2023, RVE SDEs). Whenever the SDE below is well posed, each initial law determines a probability measure on the trajectory space $C = C([0,1]; \overline{\mathcal{D}})$, which we call a path measure. For noise scales $0 < \sigma_0 \ll \sigma_1$, define the variance-exploding schedule

$$\sigma(t) := \sigma_0^{1-t}\sigma_1^t, \qquad g(t) := \left(\frac{\mathrm{d}}{\mathrm{d}t}\sigma(t)^2\right)^{1/2} = \sigma(t)\left(2\log\frac{\sigma_1}{\sigma_0}\right)^{1/2}. \tag{5}$$

For any initial distribution μ_0, the corresponding RVE SDE is

$$\mathrm{d}X_t = g(t)\,\mathrm{d}W_t + \mathrm{d}L_t, \qquad X_0 \sim \mu_0, \quad 0 \le t \le 1. \tag{6}$$

As the reference process, we use the RVE SDE (6) initialized from the uniform distribution $\mu_0 = \nu$, and denote the induced path measure by $\overrightarrow{\mathbb{P}}$. The following proposition justifies the use of the

[1] Here, "variance-exploding" refers to the noise schedule inherited from the unconstrained SDE, not to the marginal variance of the reflected process, which remains bounded because $\mathcal{D}$ is bounded. We follow the terminology of Lou & Ermon (2023).

RVE SDE (6) as a noising process by showing that its marginals remain explicitly known when initialized from the uniform distribution, while, for an arbitrary initial distribution, the terminal law is approximately uniform.

Proposition 3.1.1. *For the reflected noising process in* (6)*, under the assumptions of Proposition C.3.1, the uniform distribution is invariant:*

$$X_0 \sim \nu \quad \Longrightarrow \quad X_t \sim \nu, \qquad 0 \le t \le 1.$$

Moreover, fix any initial probability distribution μ_0 on $\overline{\mathcal{D}}$. For each $n \in \mathbb{N}$, let $X^{(n)}$ solve (6) *with initial law μ_0 and noise schedule $g_n(\cdot)$ corresponding to scales $0 < \sigma_{0,n} < \sigma_{1,n}$, and denote its terminal law by $\mu_1^{(n)} := \mathrm{Law}(X_1^{(n)})$. Suppose that $\int_0^1 g_n^2(t)\,\mathrm{d}t = \sigma_{1,n}^2 - \sigma_{0,n}^2 \underset{n\to\infty}{\longrightarrow} \infty$. Then*

$$\|\mu_1^{(n)} - \nu\|_{\mathrm{TV}} \underset{n\to\infty}{\longrightarrow} 0.$$

Full formulation of Proposition 3.1.1 with proofs is available in Proposition C.3.1. In practice one should choose $\sigma_0 \ll \sigma_1$ to control the discrepancy between the terminal distribution μ_1 of X_1 and the uniform distribution ν.

Consequently, since the initial marginal of $\overrightarrow{\mathbb{P}}$ is uniform, (1) can be written as an expectation under the reference path measure $Z = \mathbb{E}_{\overrightarrow{\mathbb{P}}}[r(X_0)/\nu(X_0)]$. When used directly, it corresponds to uniform Monte Carlo. This estimator can have large variance when the mass of $r(x)$ is concentrated in a small region of $\mathcal{D}$, since most samples from the uniform distribution then make only a negligible contribution to the estimate. This motivates replacing the reference path measure with a more suitable proposal. Generally, any path measure $\overrightarrow{\mathbb{Q}}$, which is equivalent to $\overrightarrow{\mathbb{P}}$ ($\overrightarrow{\mathbb{P}} \ll \overrightarrow{\mathbb{Q}}$ and $\overrightarrow{\mathbb{Q}} \ll \overrightarrow{\mathbb{P}}$), gives the importance weight

$$W_{\overrightarrow{\mathbb{Q}}}(X) := \frac{r(X_0)}{\nu(X_0)} \frac{\mathrm{d}\overrightarrow{\mathbb{P}}}{\mathrm{d}\overrightarrow{\mathbb{Q}}}(X), \qquad Z = \mathbb{E}_{\overrightarrow{\mathbb{Q}}}[W_{\overrightarrow{\mathbb{Q}}}(X)]. \tag{7}$$

Thus, every admissible proposal $\overrightarrow{\mathbb{Q}}$ yields an unbiased importance-sampling estimator, and the remaining question is how to choose $\overrightarrow{\mathbb{Q}}$ to reduce its variance, which is justified in the following proposition. Let $\overrightarrow{\mathbb{Q}}^\star$ denote a path measure of the RVE SDE (6) initialized from a target π:

$$\overrightarrow{\mathbb{Q}}^\star: \quad \mathrm{d}X_t = g(t)\,\mathrm{d}W_t + \mathrm{d}L_t, \qquad X_0 \sim \pi, \quad 0 \le t \le 1. \tag{8}$$

Proposition 3.1.2. *Under the assumptions of Proposition E.1.1, $\overrightarrow{\mathbb{Q}}^\star$ is the unique zero-variance proposal among path measures equivalent to $\overrightarrow{\mathbb{P}}$, i.e., $W_{\overrightarrow{\mathbb{Q}}^\star}(X) = Z$ with probability one for $X \sim \overrightarrow{\mathbb{Q}}^\star$.*

The full formal statement of Proposition 3.1.2 with proofs is available in Proposition E.1.1. The conditions and proof of existence and pathwise uniqueness of $\overrightarrow{\mathbb{P}}$ and $\overrightarrow{\mathbb{Q}}^\star$ can be found in Lemma C.2.1. Sampling $X_0 \sim \pi$ is intractable in general, so we instead consider the time reversal of the process (8), which we discuss in the next section.

3.2 Time-reversal of the optimal proposal

Let p_t denote the marginal density of the forward process (8), initialized from the target π. We apply the time-reversal map $Y_t := X_{1-t}$ to both the optimal forward process and the reference process, defining their reversed path laws by

$$X \sim \overrightarrow{\mathbb{Q}}^\star \Longrightarrow Y \sim \overleftarrow{\mathbb{Q}}^\star, \qquad X \sim \overrightarrow{\mathbb{P}} \Longrightarrow Y \sim \overleftarrow{\mathbb{P}}. \tag{9}$$

Thus, under $\overleftarrow{\mathbb{Q}}^\star$, trajectories start from p_1 and end at π, and $\frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{P}}}(Y) = \frac{\pi(Y_1)}{\nu(Y_1)}$. The reversed reference law $\overleftarrow{\mathbb{P}}$ is characterized by the reflected SDE

$$\mathrm{d}Y_t = g(1-t)\,\mathrm{d}\overline{W}_t + \mathrm{d}\overline{L}_t, \qquad Y_0 \sim \nu, \quad 0 \le t \le 1, \tag{10}$$

where $\overline{W}_t$ is a Brownian motion with respect to the reverse-time filtration, and $\overline{L}_{\bar{t}} := \int_0^{\bar{t}} \mathbf{n}_{\mathrm{in}}(Y_s)\,\mathrm{d}\bar{\ell}_s$. Lemma C.3.2 shows that (10) has a unique solution law, with strong existence and pathwise uniqueness. The optimal reverse law $\overleftarrow{\mathbb{Q}}^\star$ admits the SDE representation

$$\mathrm{d}Y_t = g^2(1-t)\nabla \log p_{1-t}(Y_t)\,\mathrm{d}t + g(1-t)\,\mathrm{d}\overline{W}_t + \mathrm{d}\overline{L}_t, \qquad Y_0 \sim p_1, \quad 0 \le t \le 1. \tag{11}$$

Lemma C.4.1 establishes the weak representation (11) and, under additional conditions, its uniqueness. In practice, p_1 is replaced by ν, while the unknown score $\nabla \log p_t$ is approximated by a neural network $s_\theta : \overline{\mathcal{D}} \times [0,1] \to \mathbb{R}^d$. This defines the learned reverse law $\overleftarrow{\mathbb{Q}}_\theta$ through

$$\overleftarrow{\mathbb{Q}}_\theta : \quad \mathrm{d}Y_t = g^2(1-t)s_\theta(Y_t, 1-t)\,\mathrm{d}t + g(1-t)\,\mathrm{d}\overline{W}_t + \mathrm{d}\overline{L}_t, \qquad Y_0 \sim \nu, \quad 0 \le t \le 1. \tag{12}$$

Theorem E.2.1 gives existence and uniqueness of (12) under explicit continuity and spatial-Lipschitz assumptions on s_θ.

3.3 Training objective and estimator accuracy

We train $\overleftarrow{\mathbb{Q}}_\theta$ by minimizing its reverse KL divergence to the ideal reversed proposal $\overleftarrow{\mathbb{Q}}^\star$.

Theorem 3.3.1. *Under the assumptions of Theorem E.2.1, for a realization of* (12) *driven by Brownian motion* $\overline{W}$*,*

$$\log \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}{\mathrm{d}\overleftarrow{\mathbb{P}}}(Y) = \int_0^1 g(1-t)s_\theta(Y_t, 1-t)^\top \mathrm{d}\overline{W}_t + \frac{1}{2}\int_0^1 g^2(1-t)\|s_\theta(Y_t, 1-t)\|_2^2\,\mathrm{d}t, \quad \overleftarrow{\mathbb{Q}}_\theta\text{-a.e.} \tag{13}$$

Consequently, the path KL admits a representation

$$\mathsf{KL}(\overleftarrow{\mathbb{Q}}_\theta || \overleftarrow{\mathbb{Q}}^\star) = \mathcal{L}(\theta) + \log Z, \tag{14}$$

where

$$\mathcal{L}(\theta) := \mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}\left[\frac{1}{2}\int_0^1 g^2(1-t)\|s_\theta(Y_t, 1-t)\|_2^2\,\mathrm{d}t + \log\frac{\nu(Y_1)}{r(Y_1)}\right]. \tag{15}$$

The full formal statement and proof are available in Theorem E.2.1. This result shows that minimising $\mathcal{L}$ is equivalent to minimising $\mathsf{KL}(\overleftarrow{\mathbb{Q}}_\theta || \overleftarrow{\mathbb{Q}}^\star)$ up to the constant $\log Z$, which is not needed for training.

Remark 3.3.2. Since $\overleftarrow{\mathbb{Q}}_\theta$ starts from ν, whereas $\overleftarrow{\mathbb{Q}}^\star$ starts from p_1, the chain rule for KL gives

$$\mathsf{KL}(\overleftarrow{\mathbb{Q}}_\theta || \overleftarrow{\mathbb{Q}}^\star) = \mathsf{KL}(\nu || p_1) + \int_{\mathcal{D}} \mathsf{KL}(\overleftarrow{\mathbb{Q}}_\theta(\cdot \mid Y_0 = y) || \overleftarrow{\mathbb{Q}}^\star(\cdot \mid Y_0 = y))\nu(y)\,\mathrm{d}y.$$

Consequently, even perfect matching of the conditional dynamics leaves the irreducible term $\mathsf{KL}(\nu||p_1)$, which is determined by the noising schedule.

The objective in (15) has the same structure as the standard control-based losses used in diffusion samplers (Zhang & Chen, 2022; Vargas et al., 2023). It combines a quadratic running cost with a terminal potential that favors states of high target density. Despite the reflected dynamics, no additional boundary term appears in the loss. After training, trajectories from $\overleftarrow{\mathbb{Q}}_\theta$ give the importance-sampling estimator

$$\widehat{Z} = \frac{1}{N}\sum_{i=1}^N \frac{r(Y_1^{(i)})}{\nu(Y_1^{(i)})}\frac{\mathrm{d}\overleftarrow{\mathbb{P}}}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}\big(Y^{(i)}\big), \qquad Y^{(i)} \sim \overleftarrow{\mathbb{Q}}_\theta, \tag{16}$$

where $\mathrm{d}\overleftarrow{\mathbb{P}}/\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta$ is calculated using (13). If the learned law equals the ideal reverse law and the noised initial distribution is exactly uniform, the weight is the constant Z and $Y_1 \sim \pi$, so the learned process can also be viewed as a sampler for the normalized target.

Closeness of the learned and ideal path laws controls the relative error of the normalizing-constant estimate. We quantify this relation for the continuous estimator (16) on $[0,1]$.

Theorem 3.3.3. *Under the assumptions of Theorem E.3.1, fix* θ *and let* $\widehat{Z}$ *be the estimator* (16)*, computed from* $N \ge 1$ *paths, each with law* $\overleftarrow{\mathbb{Q}}_\theta$*. Then* $\mathbb{E}[\widehat{Z}] = Z$ *and, for every* $\varepsilon > 0$*,*

$$\Pr\left(\left|\frac{\widehat{Z}}{Z} - 1\right| \ge \varepsilon\right) \le \frac{2\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}}}{\varepsilon}. \tag{17}$$

The full statement and proof appear in Theorem E.3.1. Pinsker's inequality connects this total-variation bound to either KL direction:

$$2\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}} \le \sqrt{2\min\{\mathsf{KL}(\overleftarrow{\mathbb{Q}}^\star || \overleftarrow{\mathbb{Q}}_\theta), \mathsf{KL}(\overleftarrow{\mathbb{Q}}_\theta || \overleftarrow{\mathbb{Q}}^\star)\}}. \tag{18}$$

Algorithm 1 Training reflected diffusion samplers

Input: Initial parameters θ, target r, domain $\mathcal{D}$, number of steps K, time discretization $\{t_k\}_{k=0}^{K-1}$, variance increments $\{g_k^2\}_{k=0}^{K-1}$, batch size B, number of iterations M
Define $\phi_k(y;\theta), \psi(y)$ from (21) and $\widetilde{q}_k(\widetilde{y} \mid y;\theta)$ from (23)
for $m = 0, \dots, M-1$ **do**
  Sample $Y_0^{(i)} \overset{\text{i.i.d.}}{\sim} \nu, \quad i = 1, \dots, B$
  for $k = 0, \dots, K-1$ **do**
    Sample $\xi_k^{(i)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d), \quad i = 1, \dots, B$
    $\widetilde{Y}_{k+1}^{(i)} \leftarrow Y_k^{(i)} + g_k^2 s_\theta(Y_k^{(i)}, 1-t_k) + g_k \xi_k^{(i)}, \quad Y_{k+1}^{(i)} \leftarrow \mathrm{refl}_{\mathcal{D}}(\widetilde{Y}_{k+1}^{(i)})$
  end for
  if algorithm is RDS-RP **then**
    $\mathcal{L}_{\mathrm{RP}}(\theta) \leftarrow \frac{1}{B}\sum_{i=1}^{B}\left[\sum_{k=0}^{K-1}\phi_k(Y_k^{(i)};\theta) + \psi(Y_K^{(i)})\right]$
    $\theta \leftarrow \mathrm{OptStep}\,(\theta, \nabla_\theta \mathcal{L}_{\mathrm{RP}}(\theta))$
  else
    Detach augmented trajectories $(Y_0, \widetilde{Y}_1, \dots, \widetilde{Y}_K, Y_K)$
    Compute $c_k^{(i)}$ from (25) and set $b_k^{(i)} \leftarrow \frac{1}{B-1}\sum_{\ell \neq i} c_k^{(\ell)}$
    $\mathcal{L}_{\mathrm{RF}}(\theta) \leftarrow \frac{1}{B}\sum_{i=1}^{B}\sum_{k=0}^{K-1}\left[\phi_k(Y_k^{(i)};\theta) + (c_k^{(i)} - b_k^{(i)})\,\log \widetilde{q}_k(\widetilde{Y}_{k+1}^{(i)} \mid Y_k^{(i)};\theta)\right]$
    $\theta \leftarrow \mathrm{OptStep}\,(\theta, \nabla_\theta \mathcal{L}_{\mathrm{RF}}(\theta))$
  end if
end for

The training optimizes $\mathsf{KL}(\overleftarrow{\mathbb{Q}}_\theta || \overleftarrow{\mathbb{Q}}^\star)$ via Theorem 3.3.1. Under additional score assumptions, Theorem E.3.2 expresses the other term $\mathsf{KL}(\overleftarrow{\mathbb{Q}}^\star || \overleftarrow{\mathbb{Q}}_\theta)$ as the initial-law mismatch $\mathsf{KL}(p_1||\nu)$ plus an integrated score error, which brings additional mathematical insight about the nature of this bound.

3.4 Practical implementation

In practice, we simulate (12) using a discrete variance-exploding (VE) noise schedule together with the reflection operator used in reflected diffusion models (Lou & Ermon, 2023). Let $\Delta = \{0 = t_0 < \cdots < t_K = 1\}$ be a time discretization, set $|\Delta| := \max_{0 \le k < K}(t_{k+1} - t_k)$ and

$$\overline{\sigma}_k := \sigma(1 - t_k), \qquad g_k^2 := \overline{\sigma}_k^2 - \overline{\sigma}_{k+1}^2, \qquad k = 0, \dots, K-1, \tag{19}$$

where $\sigma(\cdot)$ is defined in (5). Starting from $Y_0 \sim \nu$, at each step $k = 0, \dots, K-1$ we first sample the unconstrained update and then apply a reflection operator

$$\begin{aligned} \widetilde{Y}_{k+1} &= Y_k + g_k^2 s_\theta(Y_k, 1-t_k) + g_k \xi_k, \qquad \xi_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d), \\ Y_{k+1} &= \mathrm{refl}_{\mathcal{D}}(\widetilde{Y}_{k+1}), \end{aligned} \tag{20}$$

where $\mathrm{refl}_{\mathcal{D}} : \mathbb{R}^d \to \overline{\mathcal{D}}$ denotes the measurable, deterministic, domain-specific reflection map. Under standard regularity assumptions, the scheme converges weakly to the reflected SDE as $|\Delta| \to 0$ (Lou & Ermon, 2023).

Reparameterized gradients. For the given discretization, define the running and terminal costs

$$\phi_k(y;\theta) := \frac{1}{2} g_k^2 \|s_\theta(y, 1-t_k)\|_2^2, \qquad \psi(y) := \log \frac{\nu(y)}{r(y)}. \tag{21}$$

Then the finite-step version of the objective (15) admits the following representation:

$$\mathcal{L}_{\mathrm{RP}}(\theta) = \mathbb{E}_{\tau \sim q(\cdot;\theta)}\left[\sum_{k=0}^{K-1}\phi_k(Y_k;\theta) + \psi(Y_K)\right], \tag{22}$$

where $q(\cdot;\theta)$ is the distribution induced by the model over trajectories in discretized time. We compute gradients using the reparameterization trick, expressing each stochastic transition as a deterministic function of the current state and an independent Gaussian noise variable, and backpropagating through the simulated trajectory. Algorithm 1 summarizes the resulting training procedure.

REINFORCE gradients. The reparameterization gradient used above relies on differentiating the terminal cost through the sampled state. For non-smooth targets this derivative may be undefined, while on regions where the target is constant it vanishes and cannot capture changes in probability mass across the region boundaries. We therefore use a score-function gradient estimator, inspired by the REINFORCE technique (Williams, 1992), that differentiates the trajectory distribution instead. Before reflection is applied, the unconstrained transitions have Gaussian density

$$\widetilde{q}_k(\widetilde{Y}_{k+1} \mid Y_k; \theta) = \mathcal{N}\Big(\widetilde{Y}_{k+1}; Y_k + g_k^2 s_\theta(Y_k, 1 - t_k), g_k^2 I_d\Big). \tag{23}$$

To calculate the true density of a reflected transition, one has to compute an infinite sum of Gaussian densities corresponding to pre-images of the reflection map, see Lou & Ermon (2023). However, we show that the gradient of (22) can be expressed without evaluating computationally expensive reflected transition densities. We augment the trajectory with the pre-reflection states, $\widetilde{\tau} := (Y_0, \widetilde{Y}_1, Y_1, \ldots, \widetilde{Y}_K, Y_K)$, and denote its law by $\widetilde{q}(\cdot; \theta)$. Then, $\nabla_\theta \mathcal{L}_{\mathrm{RP}}(\theta)$ admits a simplified form.

Theorem 3.4.1. *Under the assumptions of Theorem E.4.2, the gradient of* (22) *satisfies*

$$\nabla_\theta \mathcal{L}_{RP}(\theta) = \mathbb{E}_{\widetilde{\tau} \sim \widetilde{q}(\cdot;\theta)} \left[\sum_{k=0}^{K-1} \nabla_\theta \phi_k(Y_k; \theta) + \sum_{k=0}^{K-1} c_k \, \nabla_\theta \log \widetilde{q}_k(\widetilde{Y}_{k+1} \mid Y_k; \theta) \right], \tag{24}$$

where

$$c_k := \sum_{j=k+1}^{K-1} \phi_j(Y_j; \theta) + \psi(Y_K). \tag{25}$$

Here ∇_θ denotes differentiation with the state held fixed. Full formal statement of Theorem 3.4.1 with proofs is available in Theorem E.4.2. This result allows the gradient to be computed using only the tractable pre-reflection Gaussian transitions. In particular, no reflected transition density or derivative of the reflection map is required. To reduce the variance of the score-function term, for a batch of B independent trajectories we use the leave-one-out (Kool et al., 2019) baseline $b_k^{(i)}$ and replace $c_k^{(i)}$ in by $c_k^{(i)} - b_k^{(i)}$. We summarize the training procedure using the REINFORCE gradient estimator in Algorithm 1.

Evaluation. For a batch of N trajectories generated by the reflected scheme (20), the final importance-sampling estimator (16) under a time discretization corresponds to

$$\widehat{Z} = \frac{1}{N} \sum_{i=1}^{N} \frac{r(Y_K^{(i)})}{\nu(Y_K^{(i)})} \exp\left(- \sum_{k=0}^{K-1} \left[g_k s_\theta(Y_k^{(i)}, 1 - t_k)^\top \xi_k^{(i)} + \frac{1}{2} g_k^2 \| s_\theta(Y_k^{(i)}, 1 - t_k) \|_2^2 \right] \right). \tag{26}$$

Under the assumption that a discretized step of (10) preserves the uniform distribution on $\mathcal{D}$, which holds for $\mathcal{D}$ considered in our experiments, the discrete-time estimate (26) remains unbiased. The formal proof and discussion are available in Proposition E.5.1 and Remark E.5.2.

4 Experiments

Benchmarks. We evaluate our method on bounded-domain integration problems defined on the unit hypercube $\overline{\mathcal{D}} = [0, 1]^d$, using the same reflection operator as in Lou & Ermon (2023). We consider Genz Continuous and Gaussian Peak to test smooth and non-smooth interior concentration, Genz Product Peak for sharply localized factorized structure, Genz Corner Peak for mass concentrated at a boundary corner, and Genz Discontinuous for discontinuous support with a zero-valued region (Genz, 1984; Kaarnioja, 2025). GMM Cube tests multimodal exploration, Stripe Cube tests concentration near boundary faces, and Funnel Cube tests adaptation to strongly varying conditional scales. Finally, we consider two real world integrals - Elliptic Two-Loop and Asian option. The Elliptic Two-Loop integral is a finite five-dimensional Feynman parameter integral over the unit cube (Bonciani et al., 2016; Borowka et al., 2018). And the Asian option benchmark (L'Ecuyer, 2009) provides a high-dimensional problem with strong coordinate dependence, a non-smooth payoff, and a large zero-valued region. Complete definitions and parameter settings are given in Appendix F.1.

Baselines. We compare the reparameterization and reinforced variants of our method, denoted RDS-RP and RDS-RF, respectively, against Path Integral Sampler (Zhang & Chen, 2022, PIS),

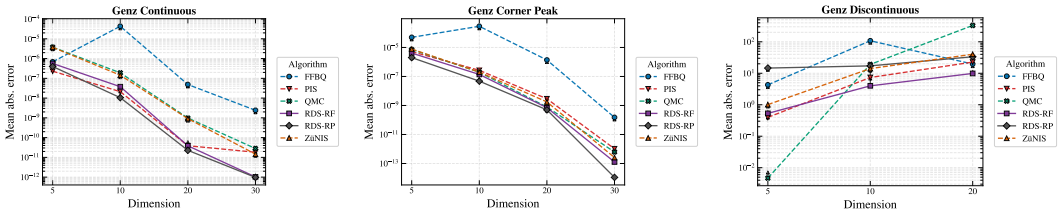


Figure 1: Mean absolute error as a function of dimension on Genz test integrals for the number of samples $N = 2^{20}$. Error bars show variability across independent evaluation runs.

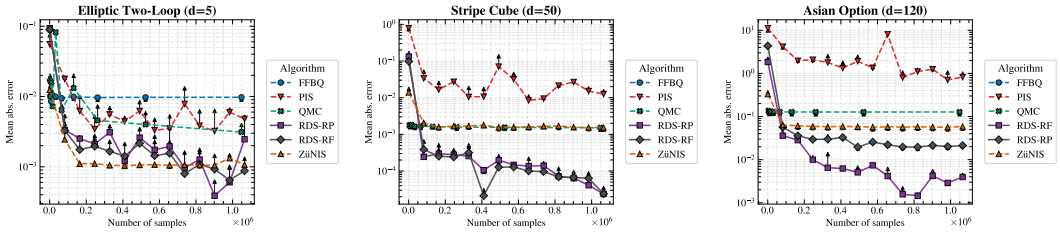
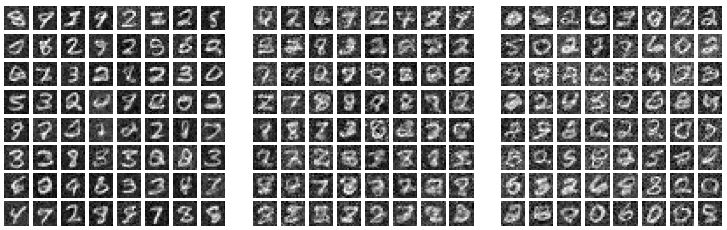


Figure 2: Mean absolute error as a function of the number of samples used for estimation on representative benchmarks. Results for FFBQ are missing on some plots due to memory limitations of this algorithm in large dimensions.



RDS-RP RDS-RF PIS

Method	$\mathcal{W}_2 \downarrow$	ELBO $\uparrow$
RDS-RP	**145.40** ±0.76	**−154.50** ±0.08
RDS-RF	161.87 ±0.63	−163.75 ±0.42
PIS	185.56 ±0.33	−172.16 ±0.67

Metrics

Figure 3: Sample visualizations and metrics on the Digits benchmark (Blessing et al., 2024).

Quasi-Monte Carlo (Niederreiter, 1992; Dick et al., 2013), Fast Fourier Bayesian Quadrature (Warren & Ramos, 2024, FFBQ), and the normalizing flow importance sampling library ZüNIS (Deutschmann & Götz, 2024). For PIS, we map $z \in \mathbb{R}^d$ to $x \in [0,1]^d$ with the sigmoid transformation $x = \sigma(z)$ and train on the change-of-variables target $\tilde{r}(z) = r(\sigma(z))\,|\det J_\sigma(z)|$, which preserves the original integral. We report the absolute normalizing-constant error $\Delta Z = |\widehat{Z} - Z|$, and show mean and standard deviation across independent evaluation runs. Full experimental details are in Appendix F.2.

Overall, RDS performs well across a broad range of targets with substantially different geometry. In particular, it remains competitive on smooth, multimodal, boundary-concentrated, and high-dimensional problems, indicating that the learned proposal is not specialized to a particular class of integrands. The strong performance across both the dimension and target-evaluation sweeps also shows that the method can effectively adapt its sampling distribution to the regions that contribute most to the integral. Figure 1 illustrates the dependence of the estimation error on dimension, while Figure 2 compares methods at matched sample budgets. The full results are given in Appendix F.3.

4.1 Problems with smooth target functions

RDS-RP is closely related to PIS, as both optimize a diffusion proposal using pathwise reparameterization gradients. The main difference is that RDS-RP evolves directly on the constrained domain and handles the boundary by reflection. This is particularly helpful for targets such as Genz Corner Peak and Stripe Cube, where a substantial amount of mass may lie close to the boundary. In PIS, the transformation to an unconstrained space sends such points far into the tails, which makes these regions harder to represent accurately. This is consistent with the stronger performance of RDS-RP on boundary-concentrated benchmarks. Besides normalizing-constant estimation, we inspect samples generated by the learned reverse process. Figure 8 shows representative RDS samples for

several cube-supported targets. These visualizations provide a complementary check that the learned dynamics capture the geometry of the target rather than only producing accurate estimates of Z.

4.2 Problems with zero-valued target regions

The path-measure and REINFORCE results require measurable $r(x)$ bounded above and away from zero, but several benchmarks vanish on subsets of $\mathcal{D}$. For these targets we use $r_\varepsilon(x) := r(x) + \varepsilon$ with a small $\varepsilon > 0$. Since $Z_\varepsilon = Z + \varepsilon \operatorname{vol}(\mathcal{D})$, we estimate Z_ε and report the corrected estimate $\widehat{Z} = \widehat{Z}_\varepsilon - \varepsilon \operatorname{vol}(\mathcal{D})$, where $\operatorname{vol}(\mathcal{D}) = 1$ for the unit cube.

Zero-valued regions are particularly challenging for optimal-control methods trained with reparameterization: it is not applicable to discontinuous functions. When the target is locally constant, $\nabla_x \log r(x)$ provides no information about how probability mass should be moved, while at the boundary of such regions the derivative may be undefined. This issue is present in both Genz Discontinuous and the Asian option benchmark, whose targets vanish over subsets of the domain. On these problems, RDS-RF consistently outperforms the reparameterized RDS-RP variant, demonstrating the advantage of differentiating the trajectory distribution rather than the terminal target itself. The samples for Genz Discontinuous in Figure 4 illustrate this behavior particularly clearly. RDS-RF successfully concentrates probability mass inside the non-zero-valued region and recovers the sharp support geometry despite the absence of informative local gradients outside it. This improved adaptation also translates into more accurate estimates of Z on both benchmarks.

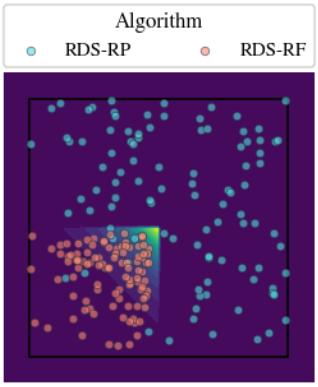


Figure 4: Visual sample comparison between RDS-RP and RDS-RF on Genz Discontinuous.

4.3 High-Dimensional Sampling

In addition, to demonstrate the benefits of learning a reflected diffusion proposal (RDS) on a bounded domain in comparison to an unconstrained diffusion proposal (PIS) from another angle, we consider the high-dimensional Digits benchmark from the diffusion sampling literature (Blessing et al., 2024). The target $\widetilde{r}(z)$ is a NICE (Dinh et al., 2014) normalizing flow model trained on downsampled 14×14 images of MNIST (LeCun et al., 1998). While the images naturally lie in a unit hypercube, the benchmark was designed for $\mathbb{R}^d$ sampling, so the normalizing flow was trained on images preprocessed with a shifted logit transform. Mirroring the previously described setup for PIS, to define an equivalent target with the same value of Z on $\overline{\mathcal{D}} = [0,1]^d$, we train RDS on $r(x) = \widetilde{r}(\sigma^{-1}(x))\,|\det J_{\sigma^{-1}}(x)|$. As discussed in Section 3.3, our learned model also acts as a sampler for the normalized target $\pi = r/Z$. And, since the ground truth value of Z is unknown, we compare RDS against PIS using sampling metrics, namely ELBO, which lower-bounds the value of $\log Z$, and $\mathcal{W}_2$ to the set of ground truth samples (see Blessing et al. (2024) and Appendix F for more details).

RDS-RP and PIS in essence have an identical KL training target, and the gradients are identically computed through reparameterization. We parameterize both models with neural networks of the same architecture and use the same number of discretization steps, so the comparison targets the only crucial difference that lies in the underlying diffusion dynamics: Brownian motion on $\mathbb{R}^d$ and reflected Brownian motion on $[0,1]^d$. Figure 3 presents the comparison results, demonstrating that RDS-RP and RDS-RF outperform PIS across both metrics.

5 Conclusion

We presented reflected diffusion samplers for integration over bounded domains. The construction replaces the Gaussian reference process of existing diffusion samplers with a reflected variance-exploding process that leaves the uniform distribution on $\mathcal{D}$ invariant, where trajectories never leave the domain, and a Girsanov change of measure for reflected diffusions supplies both the training objective and the importance weights used to estimate Z. Across a number of benchmarks, in our experiments the developed approach supplies accurate estimates of the integral. While the proposed approach is currently limited to non-negative targets, one can consider a decomposition $\int_{\mathcal{D}} r(x)\,\mathrm{d}x = \int_{\mathcal{D}} r^+(x)\,\mathrm{d}x - \int_{\mathcal{D}} r^-(x)\,\mathrm{d}x$, then estimate both parts as discussed in Section 4.2. We leave the analysis and evaluation of this adaptation for future work.

AI USE STATEMENT

In this work, we used generative AI tools to formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, design or provide feedback on research methodology or experiments, summarize or analyze existing literature, identify relevant literature, draft parts of a research paper, edit a research paper to improve readability. We have not used generative AI tools to help develop theoretical models or conceptual frameworks, propose or refine hypotheses, create or edit software code, implement methods, assist with translation, clean and reformat dataset, support qualitative and thematic data analysis, interpret results, create or modify scientific figures or images, format references, propose a title or keywords for a research paper. We have reviewed all AI-assisted work. All suggested existing literature references were manually examined by the authors. All parts of the text that were edited or written by AI tools were reviewed and validated by the authors to ensure technical accuracy and contextual appropriateness. All parts of mathematical claims and proofs that were written with the assistance of AI tools were rigorously manually checked for technical accuracy and mathematical correctness by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide complete statements and proofs of the theoretical results in the appendix, including the assumptions used for the reflected SDEs, path-space change of measure, estimator guarantees, and REINFORCE gradient estimator. Section 3.4 gives the discretization and training procedure used in practice, while the experimental appendix specifies the benchmark definitions, transformations to bounded domains, evaluation protocols, model architectures, optimization settings, and other hyperparameters used in the experiments. We also report the reference values used for normalizing-constant evaluation and describe how all reported metrics are computed. The source code for the implementation of the proposed methods and experiments is provided in the supplementary material.

REFERENCES

Tara Akhound-Sadegh, Jarrid Rector-Brooks, Joey Bose, Sarthak Mittal, Pablo Lemos, Cheng-Hao Liu, Marcin Sendera, Siamak Ravanbakhsh, Gauthier Gidel, Yoshua Bengio, Nikolay Malkin, and Alexander Tong. Iterated denoising energy matching for sampling from boltzmann densities. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=gVjMwLDFoQ.

Matthieu Alfaro, Danielle Hilhorst, and Hiroshi Matano. The singular limit of the allen–cahn equation and the fitzhugh–nagumo system. *Journal of Differential Equations*, 245(2):505–565, 2008.

Julius Berner, Lorenz Richter, and Karen Ullrich. An optimal control perspective on diffusion-based generative modeling. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=oYIjw37pTP.

Julius Berner, Lorenz Richter, Marcin Sendera, Jarrid Rector-Brooks, and Nikolay Malkin. From discrete-time policies to continuous-time diffusion samplers: Asymptotic equivalences and faster training. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=xLE3xJUuDO. Expert Certification.

Denis Blessing, Xiaogang Jia, Johannes Esslinger, Francisco Vargas, and Gerhard Neumann. Beyond ELBOs: A large-scale evaluation of variational methods for sampling. *International Conference on Machine Learning (ICML)*, 2024.

Roberto Bonciani, Vittorio Del Duca, Hjalte Frellesvig, Johannes M. Henn, Francesco Moriello, and Vladimir A. Smirnov. Two-loop planar master integrals for higgs $\to$ 3 partons with full heavy-quark mass dependence. *Journal of High Energy Physics*, 2016(12), December 2016. ISSN 1029-8479. doi: 10.1007/jhep12(2016)096. URL http://dx.doi.org/10.1007/JHEP12(2016)096.

S. Borowka, G. Heinrich, S. Jahn, S.P. Jones, M. Kerner, J. Schlenk, and T. Zirke. pysecdec: A toolbox for the numerical evaluation of multi-scale integrals. *Computer Physics Communications*,

222:313–326, January 2018. ISSN 0010-4655. doi: 10.1016/j.cpc.2017.09.015. URL http://dx.doi.org/10.1016/j.cpc.2017.09.015.

François-Xavier Briol, Chris J Oates, Mark Girolami, Michael A Osborne, and Dino Sejdinovic. Probabilistic integration. *Statistical Science*, 34(1):1–22, 2019.

Krzysztof Burdzy, Zhen-Qing Chen, and John Sylvester. The heat equation and reflected brownian motion in time-dependent domains. *The Annals of Probability*, 32(1B):775–804, 2004.

Patrick Cattiaux. Time reversal of diffusion processes with a boundary condition. *Stochastic processes and their Applications*, 28(2):275–292, 1988.

Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Stochastic control liaisons: Richard sinkhorn meets gaspard monge on a schrodinger bridge. *Siam Review*, 63(2):249–313, 2021.

Nicolas Deutschmann and Niklas Götz. Accelerating hep simulations with neural importance sampling. *Journal of High Energy Physics*, 2024(3):83, 2024.

Josef Dick, Frances Y Kuo, and Ian H Sloan. High-dimensional integration: the quasi-monte carlo way. *Acta Numerica*, 22:133–288, 2013.

Laurent Dinh, David Krueger, and Yoshua Bengio. NICE: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

Nic Fishman, Leo Klarner, Valentin De Bortoli, Emile Mathieu, and Michael John Hutchinson. Diffusion models for constrained domains. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=xuWTFQ4VGO. Expert Certification.

Robert L Foote. Regularity of the distance function. *Proceedings of the American Mathematical Society*, 92(1):153–155, 1984.

Alan Genz. Testing multidimensional integration routines. In *Proc. of international conference on Tools, methods and languages for scientific and engineering computation*, pp. 81–94, 1984.

Paul Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer, 2004.

Tom Gunter, Michael A Osborne, Roman Garnett, Philipp Hennig, and Stephen J Roberts. Sampling for inference in probabilistic models with fast bayesian quadrature. *Advances in neural information processing systems*, 27, 2014.

Wei Guo, Molei Tao, and Yongxin Chen. Complexity analysis of normalizing constant estimation: from jarzynski equality to annealed importance sampling and beyond. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=96fJALwotm.

J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods*. Methuen & Co. London; John Wiley & Sons, New York, 1964.

Aaron J Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Daniel S. Levine, Brandon M Wood, Bin Hu, Brandon Amos, Brian Karrer, Xiang Fu, Guan-Horng Liu, and Ricky T. Q. Chen. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=6Eg1OrHmg2.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Asbjørn Holk, Claudia Strauch, and Lukas Trottner. Statistical guarantees for denoising reflected diffusion models. *Journal of Machine Learning Research*, 27(116):1–50, 2026.

Vesa Kaarnioja. Explicit solutions of genz test integrals. *Applied Mathematics Letters*, 163:109444, 2025.

Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

Teun Kloek and Herman K Van Dijk. Bayesian estimates of equation system parameters: an application of integration by monte carlo. *Econometrica: Journal of the Econometric Society*, pp. 1–19, 1978.

Zhifeng Kong and Kamalika Chaudhuri. The expressive power of a class of normalizing flow models. In *International conference on artificial intelligence and statistics*, pp. 3599–3609. PMLR, 2020.

Wouter Kool, Herke van Hoof, and Max Welling. Buy 4 REINFORCE samples, get a baseline for free! In *ICLR 2019 Workshop: Deep Reinforcement Learning Meets Structured Prediction*, 2019. URL https://openreview.net/forum?id=r1lgTGL5DE.

Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*, volume 274. Springer, 2016.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Guan-Horng Liu, Tianrong Chen, Evangelos Theodorou, and Molei Tao. Mirror diffusion models for constrained and watermarked generation. *Advances in Neural Information Processing Systems*, 36: 42898–42917, 2023.

Jackson Loper. Uniform ergodicity for brownian motion in a bounded convex set. *Journal of Theoretical Probability*, 33(1):22–35, 2020.

Aaron Lou and Stefano Ermon. Reflected diffusion models. In *International Conference on Machine Learning*, pp. 22675–22701. PMLR, 2023.

Pierre L'Ecuyer. Quasi-monte carlo methods with applications in finance. *Finance and Stochastics*, 13(3):307–349, 2009.

Laurence Illing Midgley, Vincent Stimper, Gregor N. C. Simm, Bernhard Schölkopf, and José Miguel Hernández-Lobato. Flow annealed importance sampling bootstrap. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=XCTVFJwS9LJ.

Thomas Müller, Brian McWilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. Neural importance sampling. *ACM Transactions on Graphics (ToG)*, 38(5):1–19, 2019.

Radford M Neal. Annealed importance sampling. *Statistics and computing*, 11(2):125–139, 2001.

Kim A Nicoli, Christopher J Anders, Lena Funcke, Tobias Hartung, Karl Jansen, Pan Kessel, Shinichi Nakajima, and Paolo Stornati. Estimation of thermodynamic observables in lattice field theories with deep generative models. *Physical review letters*, 126(3):032001, 2021.

Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*. SIAM, 1992.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

Anthony O'Hagan. Bayes–hermite quadrature. *Journal of statistical planning and inference*, 29(3): 245–260, 1991.

Katharina Ott, Michael Tiemann, Philipp Hennig, and François-Xavier Briol. Bayesian numerical integration with neural networks. In *Uncertainty in Artificial Intelligence*, pp. 1606–1617. PMLR, 2023.

Andrey Pilipenko. *An introduction to stochastic differential equations with reflection*. Universitätsverlag Potsdam, 2014.

Carl Edward Rasmussen and Zoubin Ghahramani. Bayesian monte carlo. *Advances in neural information processing systems*, pp. 505–512, 2003.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

Zeev Schuss. *Brownian dynamics at boundaries and interfaces*, volume 758. Springer, 2015.

Marcin Sendera, Minsu Kim, Sarthak Mittal, Pablo Lemos, Luca Scimeca, Jarrid Rector-Brooks, Alexandre Adam, Yoshua Bengio, and Nikolay Malkin. Improved off-policy training of diffusion samplers. *Advances in Neural Information Processing Systems*, 37:81016–81045, 2024.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Gunnar Taraldsen. Optimal learning from the doob-dynkin lemma. *arXiv preprint arXiv:1801.00974*, 2018.

Francisco Vargas, Will Sussman Grathwohl, and Arnaud Doucet. Denoising diffusion samplers. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=8pvnfTAbu1f.

Francisco Vargas, Shreyas Padhy, Denis Blessing, and Nikolas Nüsken. Transport meets variational inference: Controlled monte carlo diffusions. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=PP1rudnxiW.

Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

Houston Warren and Fabio Ramos. Fast fourier bayesian quadrature. In *International Conference on Artificial Intelligence and Statistics*, pp. 4555–4563. PMLR, 2024.

RJ Williams. On time-reversal of reflected brownian motions. In *Seminar on Stochastic Processes, 1987*, pp. 265–276. Springer, 1987.

Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256, 1992.

Peter Wirnsberger, Andrew J Ballard, George Papamakarios, Stuart Abercrombie, Sébastien Racanière, Alexander Pritzel, Danilo Jimenez Rezende, and Charles Blundell. Targeted free energy estimation via learned mappings. *The Journal of Chemical Physics*, 153(14), 2020.

Qinsheng Zhang and Yongxin Chen. Path integral sampler: A stochastic control approach for sampling. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=_uCb2ynRu7Y.

A Related work

Quasi-Monte Carlo (QMC) builds the estimator $\widehat{Z}_N = \frac{1}{N}\sum_i r(x_i)$ avoiding randomness. It places the points on a deterministic set designed to evenly cover $[0,1]^d$ using a lattice or a digital net (Niederreiter, 1992; Dick et al., 2013). Independent points clump together and leave gaps by chance. An evenly spread point set avoids both, and for smooth integrands this yields a substantially faster error decay (Dick et al., 2013). Since the point set is fixed in advance and does not depend on r, every region of $\mathcal{D}$ receives equal attention regardless of where the mass of r lies. Unlike the following methods, QMC cannot concentrate effort where it is needed, which can lead to unsatisfactory estimates in a more general case.

Bayesian quadrature (BQ) places a Gaussian process prior on the integrand r and conditions it on evaluations of r at a set of points (O'Hagan, 1991; Rasmussen & Ghahramani, 2003). Since

integration is linear, the resulting posterior over r induces a Gaussian posterior over Z itself, whose mean serves as the estimate and whose variance quantifies how much the collected evaluations constrain Z. This last property is the method's main attraction: it reports how much of Z remains undetermined, and the evaluation points can be chosen adaptively to reduce that uncertainty most quickly (Gunter et al., 2014). The cost is that the estimate is only as good as the prior: the convergence rate depends on the integrand being well captured by the Gaussian process, and the posterior variance depends only on the evaluation points and the covariance function, not on the observed values of r, so the reported uncertainty is trustworthy only when the prior is well specified (Briol et al., 2019). Conditioning on n evaluations also costs $O(n^3)$, which limits the method to comparatively small evaluation budgets and low dimensions.

Importance sampling (IS) introduces a proposal density q on $\mathcal{D}$, and estimates (1) by

$$\widehat{Z}_N \;=\; \frac{1}{N}\sum_{i=1}^{N}\frac{r(x_i)}{q(x_i)}, \qquad x_i \sim q. \tag{27}$$

The estimator is unbiased, $\mathbb{E}_q[\widehat{Z}_N] = Z$, and its variance is

$$\mathrm{Var}_q\left(\widehat{Z}_N\right) \;=\; \frac{Z^2}{N}\chi^2(\pi\|q), \tag{28}$$

where $\chi^2(\pi\|q)$ is chi-square divergence. The minimiser q is therefore the normalized target π itself, which attains $\mathrm{Var}_q(\widehat{Z}_N) = 0$ (Hammersley & Handscomb, 1964). Sampling from π is exactly the intractable problem, so in practice the proposal is only an approximation, and the choice is itself difficult: whenever q underweights a region carrying appreciable mass of π, the ratio π/q becomes large there and (28) degrades. However, IS originally does not use the expressive power of neural networks that have recently led to prominent results in many areas. This is also a limitation of QMC and BQ, and the following algorithms address this.

Normalizing flows (Dinh et al., 2014, NF) are one of the early approaches to generative modeling using a training dataset. They are easily extended to sampling, and Midgley et al. (2023) suggests Annealed Importance Sampling (Neal, 2001) powered with normalizing flows on unbounded domains. Müller et al. (2019) target the integration problem (1) itself rather than sampling, and train a flow q_θ on the unit cube directly from unnormalized evaluations of the integrand. Training minimises $\mathrm{KL}(\pi\|q_\theta)$, whose gradient is a log-likelihood weighted by the importance ratio r/q_θ under samples drawn from q_θ itself, so no dataset is required. The trained flow is then used as the proposal q in (27). The authors apply their method to light-transport simulation, where each pixel of a rendered image is an integral over the space of light paths. Building on similar ideas, normalizing flows have also been applied for free energy estimation in physical and chemical systems (Noé et al., 2019). A common limitation of NF is the expressive power of the used neural architectures (Kong & Chaudhuri, 2020).

Compared to NF, **diffusion samplers** (Zhang & Chen, 2022; Vargas et al., 2023; Berner et al., 2024; Sendera et al., 2024; Akhound-Sadegh et al., 2024; Havens et al., 2025) are not limited to invertible architectures with computable determinant. This is a family of diffusion-based sampling approaches sharing the same underlying principle as diffusion models (Song et al., 2021): a noising process transforms the target distribution into an easy-to-sample prior. The training objective is then to learn a generative process that starts from the prior and reverses the noising process, terminating in the distribution of interest. Given a noising process and a learned generative process, these models also present a natural way to estimate the normalizing constant of the underlying distribution via importance sampling (Zhang & Chen, 2022; Guo et al., 2026). In contrast to diffusion models, no training dataset is available in this problem formulation, and the training objective instead makes use of the unnormalized density. Path Integral Sampler (PIS; Zhang & Chen, 2022) derives its loss function from stochastic optimal control theory. It introduces an easy-to-evaluate base process that is reweighted at the terminal time to match the target distribution, and the training loss is the KL divergence between the generative process and the reweighted base process. Thanks to the Girsanov theorem, this divergence is easily evaluated. Denoising Diffusion Samplers (DDS; Vargas et al., 2023) arrive at a training objective of the same form, motivated from a different perspective. Similarly to Ho et al. (2020), DDS introduces a noising process and trains a generative process as its time reversal. Diffusion samplers are derived and applied for problems where r is positive on all of $\mathbb{R}^d$, while our work extends PIS and DDS to bounded domains and problems where r may take values of zero.

B Formal notation and assumptions

B.1 Notation and conventions

General notation. The domain $\mathcal{D} \subset \mathbb{R}^d$ is bounded and convex. Its closure, boundary, and Lebesgue volume are $\overline{\mathcal{D}}$, $\partial\mathcal{D}$, and $\mathrm{vol}(\mathcal{D})$. For topological spaces E and F, we denote by $C(E;F)$ the set of continuous maps from E to F and abbreviate $C(E) := C(E;\mathbb{R})$. When E is compact and F is metric, this space carries the uniform topology induced by metric $d_\infty(f,g) = \sup_{x\in E} d_F(f(x),g(x))$ unless stated otherwise. For $E \subseteq \mathbb{R}^d$ and $f : E \to \mathbb{R}^m$,

$$\mathrm{Lip}_{x\in E}(f) := \sup_{x\neq y\in E} \frac{\|f(x)-f(y)\|_2}{\|x-y\|_2}. \tag{29}$$

Forward and reverse time are t and $\bar{t}$. Forward and reverse processes are denoted by X and Y. Symbols used in only one proof are defined locally.

For probability measures on a common measurable space, $\mu_1 \ll \mu_0$ denotes absolute continuity, and $\mu_1 \ll \mu_0 \ll \mu_1$ denotes equivalence, meaning mutual absolute continuity. We also use

$$\|\mu_1-\mu_0\|_{\mathrm{TV}} := \sup_E |\mu_1(E)-\mu_0(E)|, \tag{30}$$

$$\mathsf{KL}(\mu_1||\mu_0) := \int \log\left(\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}\right) \mathrm{d}\mu_1, \tag{31}$$

$$\chi^2(\mu_1\|\mu_0) := \int \left(\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}-1\right)^2 \mathrm{d}\mu_0. \tag{32}$$

The last two are $+\infty$ when $\mu_1 \not\ll \mu_0$. The same notation applies to densities through their induced measures. The χ^2-divergence and the variances of the integrable importance weights below are understood as extended nonnegative quantities and may equal $+\infty$.

If a continuous process X is defined under a construction probability $\mathbf{P}$, we denote its induced path law on $C([0,1];\overline{\mathcal{D}})$ by $\mathrm{Law}(X) = X_\#\mathbf{P}$. We write ζ for a generic path, with $\zeta(t)$ its value at time t, while X_t denotes the corresponding coordinate of a random path X.

For a real-valued continuous semimartingale X, a level $a \in \mathbb{R}$, and time $t \geq 0$, we write $\mathrm{LT}_t^a(X)$ for the right local time of X at a accumulated up to t, using the version that is right-continuous in a. Informally, it measures the quadratic variation accumulated while X lies immediately to the right of a. Its Tanaka characterization and one-sided occupation approximation are given by Le Gall (2016, Propositions 9.2 and 9.9).

Finally, we set

$$A(t) := \int_0^t g(s)^2 \,\mathrm{d}s. \tag{33}$$

Domain related definitions. We use the following concrete meaning of boundary regularity. We say that $\partial\mathcal{D}$ is C^1 near $z \in \partial\mathcal{D}$ if there are an open neighborhood U_z of z and a function $b_z \in C^1(\mathsf{U}_z)$ such that

$$\mathcal{D}\cap\mathsf{U}_z = \{x\in\mathsf{U}_z : b_z(x) < 0\}, \qquad \partial\mathcal{D}\cap\mathsf{U}_z = \{x\in\mathsf{U}_z : b_z(x) = 0\},$$

and $\nabla b_z(x) \neq 0$ on $\partial\mathcal{D}\cap\mathsf{U}_z$. Any such b_z is called a *local defining function*. The nonvanishing-gradient condition makes its zero set a C^1 hypersurface, by the implicit function theorem, and the gradient is perpendicular to that hypersurface.

The defining function is not unique, but the oriented unit normal is. Indeed, at z the gradients of any two local defining functions are nonzero and perpendicular to the same tangent hyperplane, so they are scalar multiples. Because both functions are negative on the interior side, that scalar is positive. Consequently, we define the outward and inward unit normals by

$$\mathbf{n}_{\mathrm{out}}(z) := \frac{\nabla b_z(z)}{\|\nabla b_z(z)\|_2}, \qquad \mathbf{n}_{\mathrm{in}}(z) := -\mathbf{n}_{\mathrm{out}}(z). \tag{34}$$

These definitions do not depend on the choice of b_z. Assumption B.2.2 imposes the stronger C^∞ version of this condition at every boundary point.

Solutions of reflected SDEs. Fix $T > 0$ and a continuous reflected process $X = (X_t)_{0\le t\le T}$ in $\overline{\mathcal{D}}$. A *scalar regulator* for X is the real-valued boundary-push process $\ell = (\ell_t)_{0\le t\le T}$: almost surely, the map $t \mapsto \ell_t$ is continuous and nondecreasing, $\ell_0 = 0$, and its Stieltjes measure is supported on the boundary-contact times,

$$\int_0^t \mathbb{1}_{\{X_s\notin\partial\mathcal{D}\}}\,\mathrm{d}\ell_s = 0, \qquad 0 \le t \le T.$$

The associated vector reflection term is $L_t := \int_0^t \mathbf{n}_{\mathrm{in}}(X_s)\,\mathrm{d}\ell_s$. Fix a Borel drift $f : \overline{\mathcal{D}} \times [0,T] \to \mathbb{R}^d$, a deterministic $g \in L^2([0,T])$, and an initial law μ on $\overline{\mathcal{D}}$. A *weak solution* of the form (2) on $[0,T]$ is a tuple

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbf{P}; \xi_0, W, X, \ell)$$

such that the filtered probability space satisfies the usual conditions, ξ_0 is an $\mathcal{F}_0$-measurable $\overline{\mathcal{D}}$-valued random variable with law μ, W is a d-dimensional Brownian motion relative to $(\mathcal{F}_t)$, and X and ℓ are continuous $(\mathcal{F}_t)$-adapted processes. They satisfy $X_0 = \xi_0$, $X_t \in \overline{\mathcal{D}}$, the preceding regulator conditions, $\int_0^T \|f(X_s, s)\|_2\,\mathrm{d}s < \infty$ almost surely, and, almost surely, for every $t \le T$,

$$X_t = X_0 + \int_0^t f(X_s, s)\,\mathrm{d}s + \int_0^t g(s)\,\mathrm{d}W_s + \int_0^t \mathbf{n}_{\mathrm{in}}(X_s)\,\mathrm{d}\ell_s.$$

We call the full tuple a weak realization and the law of X its weak solution law.

For a strong solution, we are given $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbf{P}; \xi_0, W)$, where the filtered probability space satisfies the usual conditions, ξ_0 is an $\mathcal{F}_0$-measurable $\overline{\mathcal{D}}$-valued random variable with law μ, and W is a d-dimensional Brownian motion relative to $(\mathcal{F}_t)$. A *strong solution* for these data is a continuous adapted pair (X, ℓ) satisfying the state, regulator, integrability, and integral-equation conditions in the weak-solution definition above, including $X_0 = \xi_0$, and such that, for every $t \le T$,

$$\sigma(X_s, \ell_s : 0 \le s \le t) \subseteq \sigma(\xi_0, W_s : 0 \le s \le t) \vee \mathcal{N}_{\mathbf{P}}.$$

Here $\mathcal{A} \vee \mathcal{B} := \sigma(\mathcal{A} \cup \mathcal{B})$ for sigma-fields $\mathcal{A}$ and $\mathcal{B}$, and $\mathcal{N}_{\mathbf{P}}$ consists of the $\mathbf{P}$-null events and their complements. Thus the condition allows changes on null events, but no additional randomness: both X and ℓ up to time t are determined by the prescribed initial variable and Brownian path up to time t.

This is the normal-reflection specialization of Pilipenko (2014, Definitions 2.2.2 and 2.3.1 and Remark 2.2.4). The strong formulation is the one used in the source's discussion preceding Theorem 2.5.4. The source keeps ξ_0 separate and requires $X_0 = \xi_0$. We retain that convention for both weak and strong solutions. Since our SDE displays prescribe only an initial law, the notation $X_0 \sim \mu$ is shorthand for including a variable $\xi_0 \sim \mu$ in the realization and imposing $X_0 = \xi_0$.

A displayed *law definition* is the primary specification of a path measure. If it is given by an SDE, existence and uniqueness in law must be established separately. An *SDE characterization* instead describes an already-defined law: weak existence identifies that law as one solution law, whereas pathwise uniqueness, strong existence, and uniqueness in law are additional conclusions only when explicitly stated.

B.2 Assumptions for theoretical results

Assumption B.2.1. The domain $\mathcal{D} \subset \mathbb{R}^d$ is bounded and convex.

Assumption B.2.1 is the standing domain condition throughout the appendix.

The integration problem and the algorithmic construction can be formulated without assuming a smooth boundary. The local C^1 condition in Section B.1 is enough to define the unit normal. Some of our theoretical arguments, however, invoke standard results on normal reflection, Neumann heat equations, and reflected time reversal under stronger boundary regularity (Holk et al., 2026; Pilipenko, 2014; Burdzy et al., 2004; Cattiaux, 1988; Williams, 1987). We therefore isolate the following C^∞ condition as an additional theory-only assumption rather than include it in the general problem description.

Assumption B.2.2. The boundary of $\mathcal{D}$ is C^∞: for every $z \in \partial\mathcal{D}$, there are a neighborhood U_z of z and a function $b_z \in C^\infty(\mathsf{U}_z)$ such that

$$\begin{gathered}\mathcal{D} \cap \mathsf{U}_z = \{x \in \mathsf{U}_z : b_z(x) < 0\},\ \partial\mathcal{D} \cap \mathsf{U}_z = \{x \in \mathsf{U}_z : b_z(x) = 0\},\\ \nabla b_z(x) \ne 0 \ \text{ for } x \in \mathsf{U}_z \cap \partial\mathcal{D}.\end{gathered} \tag{35}$$

Assumption B.2.2 is stronger than the local C^1 condition in Section B.1, so the unit normals are defined at every boundary point and can be used in the vector reflection term.

A density is determined only Lebesgue-almost everywhere: changing it on a null set leaves its law unchanged, but makes pointwise boundary values and $\nabla \log p_t$ ambiguous. Let $k_u(y, x)$ denote the transition density of unit normally reflected Brownian motion after elapsed time $u > 0$. For $t > 0$ we choose the representative

$$p_t(x) := \int_{\mathcal{D}} k_{A(t)}(y, x)\, \pi(y)\, \mathrm{d}y.$$

The schedule makes the elapsed time in this kernel strictly positive. Under Assumption B.2.2, Burdzy et al. (2004, Theorems 2.4(i) and 2.5) show that the kernel is strictly positive and jointly C^2 in its spatial variables up to the boundary at positive elapsed times. Compactness therefore permits differentiation under the integral, so this is a strictly positive $C^2(\overline{\mathcal{D}})$ representative of the marginal density. The static-domain specialization of Burdzy et al. (2004, Theorem 2.9) also gives its homogeneous Neumann boundary condition. This choice is needed for the pointwise reverse drift in Lemma C.4.1. Throughout, a score on $\overline{\mathcal{D}}$ means its stated continuous extension from the interior, and at $t = 0$ the representative is $p_0 = \pi$.

The integration problem (1) permits r to vanish. Some results use the following stronger target assumption.

Assumption B.2.3. The target r admits a Borel measurable extension to $\overline{\mathcal{D}}$, denoted by r as well, and there are constants $0 < m_r \le M_r < \infty$ such that

$$m_r \le r(x) \le M_r, \qquad x \in \overline{\mathcal{D}}.$$

We henceforth use this extension and $\pi = r/Z$ on $\overline{\mathcal{D}}$ as well. The two-sided bounds directly make r and the normalized density π bounded above and away from zero on $\overline{\mathcal{D}}$. Proposition E.1.1, Theorem E.2.1, and Theorem E.4.2 invoke this assumption. Extending these results to a target with zeros requires a separate support and absolute-continuity analysis and is not asserted here.

B.3 Score and learned-control assumptions

Let p_t be the forward marginal density from the target-started law (8).

Assumption B.3.1. For every $t \in [0, 1]$, the chosen representative p_t is strictly positive on $\overline{\mathcal{D}}$, and its interior score admits an extension such that

$$\big[(t, x) \mapsto \nabla \log p_t(x)\big] \in C\big([0, 1] \times \overline{\mathcal{D}}; \mathbb{R}^d\big).$$

Assumption B.3.2.

$$\sup_{t \in [0,1]} \mathrm{Lip}_{x \in \overline{\mathcal{D}}}\big(\nabla \log p_t(x)\big) < \infty. \tag{36}$$

Note that these assumptions impose additional regularity on π. Indeed, since $p_0 = \pi$, Assumption B.3.1 requires $\nabla \log \pi$ to exist in $\mathcal{D}$ and to admit a continuous extension to $\overline{\mathcal{D}}$. Moreover, the positive-time densities satisfy the homogeneous Neumann condition, so joint continuity down to $t = 0$ forces

$$\mathbf{n}_{\mathrm{out}}^{\top} \nabla \log \pi = 0 \qquad \text{on } \partial\mathcal{D}.$$

Under Assumption B.2.3, $\nabla \log \pi = -\nabla V$. Thus this requires the interior gradient of V to extend continuously to $\overline{\mathcal{D}}$ and to satisfy $\mathbf{n}_{\mathrm{out}}^{\top} \nabla V = 0$ on $\partial\mathcal{D}$. Assumption B.3.2 additionally requires ∇V to be Lipschitz on $\overline{\mathcal{D}}$. The two assumptions are stronger than these target-only consequences: they also require the positive-time scores to converge jointly to the target score as $t \downarrow 0$ and to remain uniformly spatially Lipschitz over the full time interval. None of these score conditions follows from $V \in C(\overline{\mathcal{D}})$ alone.

The preceding assumptions concern the exact score and hence the ideal reverse law. We separately impose the following regularity conditions on the learned field used in the implemented law.

Assumption B.3.3.

$$\big[(t, x) \mapsto s_\theta(x, t)\big] \in C\big([0, 1] \times \overline{\mathcal{D}}; \mathbb{R}^d\big).$$

Since $[0,1] \times \overline{\mathcal{D}}$ is compact, joint continuity makes s_θ bounded and Borel.

Assumption B.3.4.

$$\sup_{t\in[0,1]} \mathrm{Lip}_{x\in\overline{\mathcal{D}}}\big(s_\theta(x,t)\big) < \infty.$$

Assumption B.3.4 is used to prove uniqueness for the learned SDE.

C Existence and reversal of the reflected laws

The arguments in this section separate two tasks. First, the reflected-SDE and Skorokhod results of Pilipenko (2014, Theorems 2.5.1–2.5.3) and the Yamada–Watanabe discussion preceding Pilipenko (2014, Theorem 2.5.4) provide well-posedness for equations with prescribed coefficients. We use them to construct the zero-drift forward law and the continuous learned law, and to upgrade the ideal reverse equation to a unique strong solution when its score drift is spatially Lipschitz. The required domain hypotheses are verified in Lemma C.1.1.

Second, well-posedness of an equation with a prescribed score drift does not by itself identify that equation with the reversal of the forward law. We first define the ideal reverse law by reversing the already constructed target-started forward process. The time-homogeneous reflected-reversal statement of Holk et al. (2026, Section 2, equation (4)), followed by the local clock calculation in Lemma C.4.1, identifies this particular reversal as a weak solution of the scheduled score SDE on $[0,1)$. Score continuity is needed only to extend this representation to the closed endpoint, while spatial Lipschitzness supplies the subsequent uniqueness upgrade.

C.1 Geometric prerequisites for reflected SDEs

The following lemma supplies a technical prerequisite for the existence and uniqueness arguments in this section. It verifies that the standing domain assumptions imply the geometric conditions required by the reflected Skorokhod and SDE results used below.

For $x \in \partial\mathcal{D}$ and $\rho > 0$, define the inward normal sets used in Pilipenko (2014, Section 2.5) by

$$\mathsf{N}_{x,\rho} := \{n \in \mathbb{R}^d : \|n\|_2 = 1,\ B(x-\rho n, \rho) \cap \mathcal{D} = \varnothing\}, \qquad \mathsf{N}_x := \bigcup_{\rho>0} \mathsf{N}_{x,\rho}.$$

The uniform exterior-sphere condition requires a radius $\rho_{\mathrm{ext}} > 0$ such that $\mathsf{N}_x = \mathsf{N}_{x,\rho_{\mathrm{ext}}} \neq \varnothing$ for every $x \in \partial\mathcal{D}$. The uniform-cone condition in Pilipenko (2014, Theorem 2.5.2) requires constants $\delta_{\mathrm{c}} > 0$ and $\alpha \in [0,1)$ such that, for every $x \in \partial\mathcal{D}$, some unit vector v_x satisfies, for every $y \in B(x,\delta_{\mathrm{c}}) \cap \partial\mathcal{D}$,

$$\{z \in \mathbb{R}^d : \langle z-y, v_x\rangle \geq \alpha\|z-y\|_2\} \cap B(x,\delta_{\mathrm{c}}) \subset \overline{\mathcal{D}}. \tag{37}$$

The closure in (37) is essential because the cone contains its boundary vertex y.

Lemma C.1.1. *Under Assumptions B.2.1 and B.2.2, $\mathcal{D}$ satisfies the uniform exterior-sphere and uniform-cone conditions just defined.*

Proof. We verify the two conditions separately. For the exterior-sphere condition, we place a ball tangent to the boundary at x on the exterior side of its supporting hyperplane. Convexity will keep the whole domain outside that ball. We then check that smoothness leaves no other possible direction in N_x. This second step is needed for the equality $\mathsf{N}_x = \mathsf{N}_{x,\rho_{\mathrm{ext}}}$ in the definition.

Fix $x \in \partial\mathcal{D}$. The inward unit normal $\mathbf{n}_{\mathrm{in}}(x)$ is unique by Assumption B.2.2. The supporting hyperplane at x leaves a convex domain entirely on its inward side, so

$$\langle y - x, \mathbf{n}_{\mathrm{in}}(x)\rangle \geq 0, \qquad y \in \overline{\mathcal{D}}.$$

For any $\rho > 0$, the point $x - \rho\mathbf{n}_{\mathrm{in}}(x)$ is the center of a radius-ρ ball tangent at x and placed on the exterior side. For every $y \in \overline{\mathcal{D}}$,

$$\|y - (x - \rho\mathbf{n}_{\mathrm{in}}(x))\|_2^2 = \|y-x\|_2^2 + 2\rho\langle y-x, \mathbf{n}_{\mathrm{in}}(x)\rangle + \rho^2 \geq \rho^2.$$

Thus every point of the domain is at least ρ from the center, and hence $B(x-\rho\mathbf{n}_{\mathrm{in}}(x),\rho) \cap \mathcal{D} = \varnothing$ for every $\rho > 0$.

It remains to show that no other unit vector belongs to N_x. Suppose that $n \in \mathsf{N}_{x,\rho}$ for some $\rho > 0$. Disjointness from the open ball gives the following inequality first for $y \in \mathcal{D}$ and then, by continuity and density of $\mathcal{D}$ in $\overline{\mathcal{D}}$, for every $y \in \overline{\mathcal{D}}$:

$$2\rho\langle n, y - x\rangle + \|y - x\|_2^2 \geq 0.$$

To identify the direction n, let γ be any smooth boundary curve with $\gamma(0) = x$. Substituting $y = \gamma(t)$ and dividing by $t > 0$ gives $\langle n, \gamma'(0)\rangle \geq 0$ in the limit $t \downarrow 0$. Approaching through $t < 0$ reverses the inequality and gives $\langle n, \gamma'(0)\rangle \leq 0$. Hence $\langle n, \gamma'(0)\rangle = 0$. Therefore n is orthogonal to every tangent vector at x. The smooth boundary has a one-dimensional normal space, so n is either $\mathbf{n}_{\text{in}}(x)$ or $-\mathbf{n}_{\text{in}}(x)$. To select the sign, a local defining function b at x gives

$$b\big(x + \varepsilon\mathbf{n}_{\text{in}}(x)\big) = -\varepsilon\|\nabla b(x)\|_2 + o(\varepsilon) < 0$$

for all sufficiently small $\varepsilon > 0$. Hence $x + \varepsilon\mathbf{n}_{\text{in}}(x) \in \mathcal{D}$. Substituting these interior points in the same inequality gives

$$2\rho\varepsilon\langle n, \mathbf{n}_{\text{in}}(x)\rangle + \varepsilon^2 \geq 0.$$

Dividing by ε and letting $\varepsilon \downarrow 0$ gives $\langle n, \mathbf{n}_{\text{in}}(x)\rangle \geq 0$. Thus $n = \mathbf{n}_{\text{in}}(x)$, and

$$\mathsf{N}_x = \mathsf{N}_{x,\rho} = \{\mathbf{n}_{\text{in}}(x)\} \qquad (\rho > 0).$$

Any fixed $\rho_{\text{ext}} > 0$ therefore verifies the uniform exterior-sphere condition.

For the cone condition, set $\alpha := 1/2$. Around each boundary point x_0, choose a local defining function b and use the inward normal at x_0 as a fixed cone axis. After shrinking the neighborhood, continuity of ∇b makes its variation at most $\|\nabla b(x_0)\|_2/4$. Consequently, if a displacement h satisfies the cone inequality $\langle h, v_{x_0}\rangle \geq \alpha\|h\|_2$, then the defining function decreases strictly along h: the fixed-gradient term is at most $-\|\nabla b(x_0)\|_2\|h\|_2/2$, whereas the gradient variation contributes at most $\|\nabla b(x_0)\|_2\|h\|_2/4$. Compactness of the boundary will then make the neighborhood radius uniform. The value $1/2$ is only a convenient fixed aperture. Any $\alpha \in (1/4, 1)$ would work with the same gradient-variation bound.

Fix $x_0 \in \partial\mathcal{D}$ and choose a defining function b from Assumption B.2.2, so locally $\mathcal{D} = \{b < 0\}$ and $\nabla b(x_0) \neq 0$. Since $\nabla b(x_0)$ points outward, set

$$m := \|\nabla b(x_0)\|_2, \qquad v_{x_0} := -\frac{\nabla b(x_0)}{m}.$$

Thus v_{x_0} points inward. By continuity of ∇b, after shrinking the defining-function neighborhood we may choose $\rho_{x_0} > 0$ such that $B(x_0, 2\rho_{x_0})$ lies in that neighborhood and

$$\sup_{w \in B(x_0, 2\rho_{x_0})} \|\nabla b(w) - \nabla b(x_0)\|_2 \leq \frac{m}{4}.$$

This bound says that throughout the neighborhood, the gradient remains close to the fixed outward direction at x_0.

Let $y \in B(x_0, \rho_{x_0}) \cap \partial\mathcal{D}$ and suppose that $z \in B(x_0, \rho_{x_0})$ satisfies

$$\langle z - y, v_{x_0}\rangle \geq \alpha\|z - y\|_2.$$

In other words, the displacement from y to z lies in a cone pointing in the fixed inward direction v_{x_0}. Write $h := z - y$. The segment from y to z remains in $B(x_0, \rho_{x_0})$. At every point w on that segment,

$$\begin{aligned}\langle \nabla b(w), h\rangle &\leq \langle \nabla b(x_0), h\rangle + \frac{m}{4}\|h\|_2 \\ &= -m\langle v_{x_0}, h\rangle + \frac{m}{4}\|h\|_2 \leq -m\left(\alpha - \frac{1}{4}\right)\|h\|_2 = -\frac{m}{4}\|h\|_2.\end{aligned}$$

The first term represents movement in the ideal inward direction at x_0. The gradient variation can cost at most $m\|h\|_2/4$, so the net directional derivative remains negative. Since $b(y) = 0$, the fundamental theorem of calculus gives

$$b(z) - b(y) = \int_0^1 \langle \nabla b(y + th), h\rangle \, \mathrm{d}t \leq -\frac{m}{4}\|h\|_2.$$

Hence $b(z) < 0$ when $z \neq y$, while for $z = y$ the point already lies in $\partial\mathcal{D}$. Thus every such z belongs to $\overline{\mathcal{D}}$.

The radius ρ_{x_0} constructed so far may depend on x_0. To obtain one radius for the whole boundary, use that boundedness of $\mathcal{D}$ makes $\partial\mathcal{D}$ compact. The sets $B(x_0, \rho_{x_0}/2)$ cover the boundary. Choose a finite subcover with centers $x_1, \ldots, x_N$ and set

$$\delta_{\mathrm{c}} := \min_{1 \le i \le N} \frac{\rho_{x_i}}{2} > 0.$$

For each $x \in \partial\mathcal{D}$, choose i such that $x \in B(x_i, \rho_{x_i}/2)$ and set $v_x := v_{x_i}$. Then $B(x, \delta_{\mathrm{c}}) \subset B(x_i, \rho_{x_i})$. Therefore both the cone vertex y and the point z remain in the neighborhood where the preceding local argument applies. It applies simultaneously to every $y \in B(x, \delta_{\mathrm{c}}) \cap \partial\mathcal{D}$ and every point in the cone with vertex y, axis v_x, aperture parameter α, and intersection with $B(x, \delta_{\mathrm{c}})$. This is exactly (37). □

C.2 Forward reflected path law

Lemma C.2.1. *Under Assumptions B.2.1 and B.2.2, the forward RVE SDE* (6)*, initialized from any probability law on $\overline{\mathcal{D}}$, has a pathwise-unique strong solution on* $[0, 1]$. *Hence $\overrightarrow{\mathbb{P}}$ and $\overrightarrow{\mathbb{Q}}^\star$ are well defined.*

Proof. The drift in (6) is zero and its diffusion coefficient $g(t)I_d$ is independent of the state. The schedule (5) is continuous and bounded on $[0, 1]$. Extend it constantly after time 1. Thus the coefficients are bounded and continuous on $[0, \infty) \times \overline{\mathcal{D}}$ and uniformly Lipschitz in space, with spatial Lipschitz constant zero. Assumptions B.2.1 and B.2.2, through Lemma C.1.1, give the uniform exterior-sphere and uniform-cone conditions. Hence Pilipenko (2014, Theorem 2.5.3) gives weak existence, while the uniform exterior-sphere condition, Pilipenko (2014, Theorem 2.5.1) give pathwise uniqueness, and Pilipenko (2014, Exercise 2.5.3) gives the equality of the scalar regulators. The Yamada–Watanabe discussion immediately preceding Pilipenko (2014, Theorem 2.5.4) then gives a unique strong solution and uniqueness in law. This proves the claim on $[0, 1]$. No density assumption on the initial law is used. □

C.3 Uniform marginals and reversal of the reference law

Let $P_u(x, \mathrm{d}y)$ be the transition kernel of unit normally reflected Brownian motion on $\overline{\mathcal{D}}$. We use $P_u f(x)$ for its action on functions and μP_u for its action on measures.

Proposition C.3.1. *Suppose Assumptions B.2.1 and B.2.2 hold, and let X follow* (6)*. If $X_0 \sim \nu$, then $X_t \sim \nu$ for every $t \in [0, 1]$. If $X_0 \sim \mu_0$ is arbitrary, then for every fixed $t \in (0, 1]$,*

$$A(t) \longrightarrow \infty \quad \Longrightarrow \quad \|\mathrm{Law}(X_t) - \nu\|_{\mathrm{TV}} \longrightarrow 0 \tag{38}$$

along any family of noise schedules. The limit varies the schedule, not t.

Proof. The two domain assumptions have different roles. Bounded convexity is the hypothesis used for the stationarity and uniform-ergodicity conclusions of Loper (2020, Lemma 1), whereas smoothness keeps (6) within the manuscript's single-valued inward-normal and unique-law framework. Moreover, (5) gives $g \in C([0, 1])$ with $\inf_{[0,1]} g > 0$, so the clock A is finite and strictly increasing. For $u \in [0, A(1)]$, set $R_u := X_{A^{-1}(u)}$. The time-changed martingale $\int_0^{A^{-1}(u)} g(s)\,\mathrm{d}W_s$ has bracket uI_d, so it is Brownian. The increasing clock preserves the boundary support of the scalar regulator $\ell^R_u := \ell_{A^{-1}(u)}$. Hence R is unit normally reflected Brownian motion and

$$X_t = R_{A(t)}, \qquad 0 \le t \le 1. \tag{39}$$

This constructs only the segment $0 \le u \le A(1)$. Under Assumptions B.2.1 and B.2.2, Lemma C.1.1 gives the two domain conditions required by Pilipenko (2014, Theorem 2.5.3). Applied to the bounded continuous coefficients given by drift zero and diffusion coefficient I_d, the cited theorem supplies a global weak unit reflected-Brownian solution. The uniform exterior-sphere part of the same lemma verifies the domain hypothesis for pathwise uniqueness in Pilipenko (2014, Theorem 2.5.1). Combining weak existence and pathwise uniqueness, the Yamada–Watanabe discussion immediately

preceding Pilipenko (2014, Theorem 2.5.4) gives a unique strong solution and uniqueness in law. Uniqueness in law identifies the constructed segment with the restriction of this global solution. By Loper (2020, Lemma 1(1)), the global solution is a homogeneous Markov process with transition kernel P_u. Hence, conditionally on R_0, the law of R_u is $P_u(R_0, \cdot)$. The limit $u \to \infty$ below concerns this global Markov family, not an extension of the finite realization. If $X_0 \sim \nu$, stationarity in Loper (2020, Lemma 1(1)) proves the first claim. For a general μ_0, the uniform ergodicity in Loper (2020, Lemma 1(2)) gives

$$\|\mu_0 P_u - \nu\|_{\mathrm{TV}} \le \sup_{x \in \overline{\mathcal{D}}} \|P_u(x, \cdot) - \nu\|_{\mathrm{TV}} \longrightarrow 0 \qquad (u \to \infty). \tag{40}$$

For fixed physical time t, (39) gives $\mathrm{Law}(X_t) = \mu_0 P_{A(t)}$. Substituting $u = A(t)$ into (40) therefore proves (38) whenever a family of noise schedules makes $A(t) \to \infty$. Thus this is a limit in accumulated clock time, not a limit $t \to \infty$. □

Lemma C.3.2. *Suppose Assumptions B.2.1 and B.2.2 hold. The reflected SDE*

$$\mathrm{d}Y_{\bar{t}} = g(1 - \bar{t})\, \mathrm{d}\overline{W}_{\bar{t}} + \mathbf{n}_{\mathrm{in}}(Y_{\bar{t}})\, \mathrm{d}\bar{\ell}_{\bar{t}}, \qquad Y_0 \sim \nu, \quad 0 \le \bar{t} \le 1,$$

admits a strong solution, has pathwise uniqueness, and is unique in law. Its unique solution law is the reversed reference law $\overleftarrow{\mathbb{P}}$ defined in (9)*. The Brownian motion $\overline{W}$ and scalar regulator $\bar{\ell}$ in a realization of this law need not be literal pathwise reversals of the forward drivers.*

Proof. Here the two domain assumptions first enter through Lemma C.2.1. Bounded convexity is used again in Loper (2020, Lemma 1), while $g > 0$ makes the forward and reverse clock changes invertible. Lemma C.2.1 makes $\overrightarrow{\mathbb{P}}$ well defined. Let $X \sim \overrightarrow{\mathbb{P}}$ and set $Y_{\bar{t}} := X_{1-\bar{t}}$. By (9), $\mathrm{Law}(Y) = \overleftarrow{\mathbb{P}}$ by definition. By the definition of $\overrightarrow{\mathbb{P}}$, $X_0 \sim \nu$. Equation (39) gives a unit normally reflected Brownian motion R, with $R_0 = X_0 \sim \nu$, such that

$$X_t = R_{A(t)}, \qquad 0 \le t \le 1.$$

Thus, for a Brownian motion W' and a scalar regulator ℓ', the unit-clock process satisfies

$$\mathrm{d}R_u = \mathrm{d}W'_u + \mathbf{n}_{\mathrm{in}}(R_u)\, \mathrm{d}\ell'_u, \qquad R_0 \sim \nu, \quad 0 \le u \le A(1).$$

By Loper (2020, Lemma 1(1)), the initial law $R_0 \sim \nu$ makes R stationary. We first reverse R on the unit-clock interval $[0, A(1)]$ by setting

$$\widehat{R}_u := R_{A(1)-u}, \qquad 0 \le u \le A(1).$$

By the reversibility conclusion of Loper (2020, Lemma 1(1)),

$$\mathrm{Law}\big((\widehat{R}_u)_{0 \le u \le A(1)}\big) = \mathrm{Law}\big((R_u)_{0 \le u \le A(1)}\big).$$

Indeed, repeated detailed balance reverses every finite-dimensional Markov product, and path continuity upgrades equality of finite-dimensional distributions to equality of laws on $C([0, A(1)]; \overline{\mathcal{D}})$.

We next return from unit-clock time u to reverse physical time $\bar{t}$. Define

$$\overline{A}(\bar{t}) := A(1) - A(1 - \bar{t}) = \int_0^{\bar{t}} g(1 - s)^2\, \mathrm{d}s, \qquad 0 \le \bar{t} \le 1.$$

Since $A(0) = 0$ and A is increasing, $\overline{A}(0) = 0$, $\overline{A}(1) = A(1)$, and $\overline{A}$ maps $[0, 1]$ increasingly onto $[0, A(1)]$. Then

$$Y_{\bar{t}} = X_{1-\bar{t}} = R_{A(1-\bar{t})} = R_{A(1)-\overline{A}(\bar{t})} = \widehat{R}_{\overline{A}(\bar{t})}.$$

Applying the deterministic time change $u = \overline{A}(\bar{t})$ to both processes in the preceding path-law identity therefore gives

$$\mathrm{Law}\big((Y_{\bar{t}})_{0 \le \bar{t} \le 1}\big) = \mathrm{Law}\big((R_{\overline{A}(\bar{t})})_{0 \le \bar{t} \le 1}\big).$$

We now time-change the unit-clock SDE displayed above. Define

$$Z_{\bar{t}} := R_{\overline{A}(\bar{t})}, \qquad M_{\bar{t}} := W'_{\overline{A}(\bar{t})}, \qquad \bar{\ell}_{\bar{t}} := \ell'_{\overline{A}(\bar{t})}.$$

Because $\overline{A}'(\bar{t}) = g(1-\bar{t})^2$, the continuous local martingale M has bracket $\overline{A}(\bar{t})I_d$. Hence

$$\overline{W}_{\bar{t}} := \int_0^{\bar{t}} g(1-s)^{-1}\,\mathrm{d}M_s$$

has bracket $\bar{t}I_d$ and is Brownian by the characterization in Le Gall (2016, Theorem 5.12), while stochastic-integral associativity from Le Gall (2016, Theorem 5.6) gives

$$M_{\bar{t}} = \int_0^{\bar{t}} g(1-s)\,\mathrm{d}\overline{W}_s.$$

The same increasing time change preserves the boundary support of the scalar regulator. Consequently,

$$\mathrm{d}Z_{\bar{t}} = g(1-\bar{t})\,\mathrm{d}\overline{W}_{\bar{t}} + \mathbf{n}_{\mathrm{in}}(Z_{\bar{t}})\,\mathrm{d}\bar{\ell}_{\bar{t}}, \qquad Z_0 \sim \nu.$$

Since Z has path law $\overleftarrow{\mathbb{P}}$ by the preceding law identity, this proves that $\overleftarrow{\mathbb{P}}$ is a weak solution law of the displayed SDE.

It remains to prove strong existence and pathwise uniqueness. In the normal-reflection notation of Pilipenko (2014, Section 2.5), take drift $a = 0$ and diffusion vectors

$$b_k(\bar{t}, x) := g(1-\bar{t})e_k, \qquad k \in \{1, \dots, d\}.$$

On $[0,1]$ these coefficients are bounded and continuous, and their spatial Lipschitz constants are zero. Extending them constantly after time 1 preserves these properties. Under Assumptions B.2.1 and B.2.2, Lemma C.1.1 gives the exterior-sphere condition required by Pilipenko (2014, Theorem 2.5.1), which gives pathwise uniqueness. Together with the weak solution just constructed, the Yamada–Watanabe discussion immediately preceding Pilipenko (2014, Theorem 2.5.4) gives a unique strong solution and uniqueness in law. Since the constructed weak solution has law $\overleftarrow{\mathbb{P}}$, that uniqueness identifies $\overleftarrow{\mathbb{P}}$ as the unique solution law. □

C.4 Ideal reverse path law

Lemma C.4.1. *Suppose Assumptions B.2.1 and B.2.2 hold. Let $X \sim \overrightarrow{\mathbb{Q}}^\star$, let p_t be the density of X_t, and set $Y_{\bar{t}} := X_{1-\bar{t}}$ for $0 \le \bar{t} \le 1$. Then the path law $\overleftarrow{\mathbb{Q}}^\star := \mathrm{Law}((Y_{\bar{t}})_{0\le\bar{t}\le 1})$ is well defined.*

If Assumption B.3.1 also holds, then $\overleftarrow{\mathbb{Q}}^\star$ is a weak solution law of (11) *on* $[0,1]$. *If, in addition, Assumption B.3.2 holds, then* (11) *has pathwise uniqueness, a unique strong solution, and uniqueness in law. In particular, $\overleftarrow{\mathbb{Q}}^\star$ is its unique solution law.*

Proof. Step 1: define the law. Assumptions B.2.1 and B.2.2 are used here to invoke Lemma C.2.1, which constructs $X \sim \overrightarrow{\mathbb{Q}}^\star$ with continuous paths. Thus $Y_{\bar{t}} := X_{1-\bar{t}}$ is continuous on $[0,1]$ and $\overleftarrow{\mathbb{Q}}^\star = \mathrm{Law}(Y)$ is well defined.

Step 2: identify the SDE on $[0,1)$. Set $T := A(1)$. By (39), there is a unit normally reflected Brownian motion R, started from the density π, with

$$X_t = R_{A(t)}.$$

At this stage we need only a weak-solution representation of the already defined reversed path: we must exhibit a filtered extension carrying a Brownian motion and an adapted scalar normal regulator for which the reverse SDE holds. Pathwise uniqueness and construction from a prescribed Brownian motion are addressed separately in Step 4.

Our domain is open, connected, bounded, and has C^∞ boundary. In the source notation of Holk et al. (2026, Equation (6)), take the conductivity $f_{\mathrm{H}} \equiv 1/2$ in their constrained forward model. Then $\nabla f_{\mathrm{H}} = 0$, $\sqrt{2f_{\mathrm{H}}}I_d = I_d$, and their conormal coefficient is $\nu_{\mathrm{H}} = f_{\mathrm{H}}\mathbf{n}_{\mathrm{in}} = \mathbf{n}_{\mathrm{in}}/2$. Thus their forward model is unit normally reflected Brownian motion, and its vector reflection term is $\mathbf{n}_{\mathrm{in}}\,\mathrm{d}(\ell^{\mathrm{H}}/2)$, which has exactly our inward-normal form. The reflected time-reversal statement in Holk et al. (2026, Equation (4)) and its subsequent specialization to the conductivity model therefore apply to the law of R. Under $f_{\mathrm{H}} \equiv 1/2$, the source's specialized reverse drift $\nabla f_{\mathrm{H}} + 2f_{\mathrm{H}}\nabla \log q_{T-u}$ reduces to $\nabla \log q_{T-u}$.

If q_u denotes the density of R_u and $\widehat{R}_u := R_{T-u}$, the cited result provides an enlargement of the reversed natural filtration carrying a Brownian motion $B^{\leftarrow}$ and the reversed boundary clock $\overline{\ell}^{\mathrm{H}}$ such that, for $0 \le u < T$,

$$\mathrm{d}\widehat{R}_u = \nabla \log q_{T-u}(\widehat{R}_u)\,\mathrm{d}u + \mathrm{d}B^{\leftarrow}_u + \frac{1}{2}\mathbf{n}_{\mathrm{in}}(\widehat{R}_u)\,\mathrm{d}\overline{\ell}^{\mathrm{H}}_u.$$

Consequently, the law of the actual reversed path $\widehat{R}$ is a weak solution law of the displayed SDE on $[0,T)$. The cited result makes no uniqueness assertion for these coefficients. Set

$$\lambda^{\leftarrow}_u := \frac{1}{2}\overline{\ell}^{\mathrm{H}}_u.$$

Then the last term is $\mathbf{n}_{\mathrm{in}}(\widehat{R}_u)\,\mathrm{d}\lambda^{\leftarrow}_u$. This forced rescaling only converts scalar local-time conventions. It does not choose or change the reflected path.

Define the reverse clock

$$\overline{A}(\bar{t}) := T - A(1-\bar{t}) = \int_0^{\bar{t}} g(1-s)^2\,\mathrm{d}s.$$

Then $Y_{\bar{t}} = \widehat{R}_{\overline{A}(\bar{t})}$ and $q_{A(t)} = p_t$. Since $\overline{A}'(\bar{t}) = g(1-\bar{t})^2$, the standard deterministic time change provides a Brownian motion $\overline{W}$ such that

$$\mathrm{d}B^{\leftarrow}_{\overline{A}(\bar{t})} = g(1-\bar{t})\,\mathrm{d}\overline{W}_{\bar{t}}.$$

Set $\overline{\ell}_{\bar{t}} := \lambda^{\leftarrow}_{\overline{A}(\bar{t})}$. The time-changed boundary term is therefore $\int_0^{\bar{t}} \mathbf{n}_{\mathrm{in}}(Y_s)\,\mathrm{d}\overline{\ell}_s$. Since $\overline{A}$ is increasing, $\overline{\ell}$ is supported on the boundary times of Y. Substituting $T - \overline{A}(s) = A(1-s)$ in the source drift therefore gives

$$Y_{\bar{t}} = Y_0 + \int_0^{\bar{t}} g(1-s)^2 \nabla \log p_{1-s}(Y_s)\,\mathrm{d}s + \int_0^{\bar{t}} g(1-s)\,\mathrm{d}\overline{W}_s + \int_0^{\bar{t}} \mathbf{n}_{\mathrm{in}}(Y_s)\,\mathrm{d}\overline{\ell}_s, \qquad 0 \le \bar{t} < 1. \tag{41}$$

No score assumption is used here: for $\bar{t} < 1$, the remaining forward clock $A(1-\bar{t})$ is positive, and the representative fixed above supplies the required positive-time score.

Step 3: include $\bar{t} = 1$. Equation (41) has so far been proved only for $\bar{t} < 1$. To obtain the SDE on $[0,1]$, we check that every term has a limit as $\bar{t} \uparrow 1$. The possible obstruction is the score as the forward time $1-\bar{t}$ approaches 0.

Assumption B.3.1 gives

$$K := \sup_{(t,x)\in[0,1]\times\overline{\mathcal{D}}} \|\nabla \log p_t(x)\|_2 < \infty.$$

Hence the drift integral converges. The Brownian motion has its standard continuous extension to 1, and its stochastic integral does too because $g(1-\cdot) \in L^2([0,1])$. The reversed boundary clock supplied by the cited result is continuous on $[0,T]$. Therefore $\overline{\ell}$ from Step 2 extends continuously to $[0,1]$ with the same monotonicity and boundary-support properties. Thus the scalar regulator also extends correctly. Since Y is continuous, passing to the limit in (41) proves (11) on $[0,1]$. Since $Y_{\bar{t}} = X_{1-\bar{t}}$, its law is $\overleftarrow{\mathbb{Q}}^{\star}$, with $Y_0 \sim p_1$ and $Y_1 \sim p_0 = \pi$.

Step 4: prove uniqueness. Set

$$b(\bar{t},x) := g(1-\bar{t})^2 \nabla \log p_{1-\bar{t}}(x), \qquad \sigma(\bar{t},x) := g(1-\bar{t}) I_d.$$

Assumption B.3.1 makes b bounded and continuous, while Assumption B.3.2 gives

$$\sup_{\bar{t}\in[0,1]} \mathrm{Lip}_{x\in\overline{\mathcal{D}}}\, b(\bar{t},x) \le \|g\|_\infty^2 \sup_{t\in[0,1]} \mathrm{Lip}_{x\in\overline{\mathcal{D}}}\big(\nabla \log p_t\big) < \infty.$$

The schedule makes σ bounded, continuous, and independent of x. After extending the coefficients constantly beyond time 1, Assumptions B.2.1 and B.2.2, through Lemma C.1.1, give the uniform exterior-sphere condition required by Pilipenko (2014, Theorem 2.5.1), which gives pathwise uniqueness. Together with the weak solution proved in Step 3, the Yamada–Watanabe discussion immediately preceding Pilipenko (2014, Theorem 2.5.4) gives a unique strong solution and uniqueness in law. Hence $\overleftarrow{\mathbb{Q}}^{\star}$ is the unique solution law of (11). □

D GIRSANOV CHANGE OF DRIFT FOR REFLECTED DIFFUSIONS

This section proves the path-law change of measure used in Theorem E.2.1. We first apply ordinary Girsanov theory on a filtered probability space carrying a reflected solution. We then show that the resulting density depends only on the observed path, so it descends to a Radon–Nikodym derivative between coordinate path laws. The second step is where reflection requires additional work. Theorem D.1.6 combines the two steps, after which we specialize it to the learned and reference diffusions.

D.1 CONSTRUCTION AND THE PATH DENSITY

Lemma D.1.1. *Fix $T > 0$. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbf{P})$ satisfy the usual conditions. Suppose it carries an $\mathcal{F}_0$-measurable $\overline{\mathcal{D}}$-valued random variable ξ_0, an $\mathbb{R}^d$-valued Brownian motion W relative to $(\mathcal{F}_t)_{0\le t\le T}$, a continuous adapted $\overline{\mathcal{D}}$-valued process X with $X_0 = \xi_0$, and a continuous adapted nondecreasing process ℓ with $\ell_0 = 0$. Set $L_t := \int_0^t \mathbf{n}_{\mathrm{in}}(X_s)\,\mathrm{d}\ell_s$, and suppose*

$$\mathrm{d}X_t = b_t\,\mathrm{d}t + g(t)\,\mathrm{d}W_t + \mathrm{d}L_t, \qquad 0 \le t \le T, \tag{42}$$

where b is progressively measurable with $\int_0^T \|b_t\|_2\,\mathrm{d}t < \infty$ almost surely, $g \in L^2([0,T])$ is deterministic, and $\mathrm{d}\ell_s$ is supported on $\{s : X_s \in \partial\mathcal{D}\}$. Let ϑ be a bounded predictable $\mathbb{R}^d$-valued process and set

$$\mathcal{E}_t^{\vartheta} := \exp\left(\int_0^t \vartheta_s^{\top}\,\mathrm{d}W_s - \frac{1}{2}\int_0^t \|\vartheta_s\|_2^2\,\mathrm{d}s\right). \tag{43}$$

Then $\mathcal{E}^{\vartheta}$ is a uniformly integrable $\mathbf{P}$-martingale, and $\mathrm{d}\widetilde{\mathbf{Q}} := \mathcal{E}_T^{\vartheta}\,\mathrm{d}\mathbf{P}$ defines a probability equivalent to $\mathbf{P}$. Under $\widetilde{\mathbf{Q}}$,

$$\widetilde{W}_t := W_t - \int_0^t \vartheta_s\,\mathrm{d}s$$

is Brownian relative to $(\mathcal{F}_t)_{0\le t\le T}$, the initial law of X is unchanged, and the same solution pair (X, ℓ) satisfies the stated regulator conditions and

$$\mathrm{d}X_t = (b_t + g(t)\vartheta_t)\,\mathrm{d}t + g(t)\,\mathrm{d}\widetilde{W}_t + \mathrm{d}L_t. \tag{44}$$

Proof. By Le Gall (2016, Section 5.1.3(iv)), $\int_0^{\cdot} \vartheta_s^{\top}\mathrm{d}W_s$ is a continuous local martingale. Boundedness of ϑ gives $\mathbb{E}_{\mathbf{P}}\exp(\frac{1}{2}\int_0^T \|\vartheta_s\|_2^2\,\mathrm{d}s) < \infty$. The finite-horizon form of Le Gall (2016, Theorem 5.23), obtained by stopping at T, therefore makes $\mathcal{E}^{\vartheta}$ a uniformly integrable martingale. Le Gall (2016, Theorem 5.22) makes $\widetilde{W}$ a continuous local martingale under $\widetilde{\mathbf{Q}}$. Its quadratic covariations are unchanged, so Le Gall (2016, Theorem 5.12) makes it Brownian. Substitution gives (44). The probability change does not alter the paths (X, ℓ) or their adaptedness, and equivalence preserves the regulator conditions. Finally, since $\xi_0 = X_0$ is $\mathcal{F}_0$-measurable by hypothesis, $\mathbb{E}_{\mathbf{P}}[\mathcal{E}_T^{\vartheta} \mid \mathcal{F}_0] = \mathcal{E}_0^{\vartheta} = 1$, so the law of the initial variable ξ_0 is unchanged. □

Let $\omega \in \Omega$ denote a construction-space outcome and let $\zeta \in C([0,T]; \overline{\mathcal{D}})$ denote a generic path. The observed path associated with ω is $X(\omega)$. Our goal is to identify a Borel functional z on path space such that, for every bounded Borel functional F,

$$\begin{aligned}\mathbb{E}_{\widetilde{\mathbb{Q}}}[F] &= \int F(\zeta)\,\widetilde{\mathbb{Q}}(\mathrm{d}\zeta)\\ &= \int z(\zeta)F(\zeta)\,\mathbb{P}(\mathrm{d}\zeta) = \mathbb{E}_{\mathbb{P}}[zF].\end{aligned} \tag{45}$$

This is the defining identity for $z = \mathrm{d}\widetilde{\mathbb{Q}}/\mathrm{d}\mathbb{P}$. Lemma D.1.1 initially gives only

$$\begin{aligned}\int F(\zeta)\,\widetilde{\mathbb{Q}}(\mathrm{d}\zeta) &= \int_\Omega F(X(\omega))\,\widetilde{\mathbf{Q}}(\mathrm{d}\omega)\\ &= \int_\Omega \mathcal{E}_T^{\vartheta}(\omega)F(X(\omega))\,\mathbf{P}(\mathrm{d}\omega).\end{aligned}$$

The final expression is perfectly well defined, but it is still an expectation on the construction space. We cannot yet rewrite it as an expectation under $\mathbb{P}$, because $\mathcal{E}_T^\vartheta$ need not be a function of the path X.

Conditioning converts this construction-space weight into one that retains only the information contained in X:

$$\begin{aligned}&\int_\Omega \mathcal{E}_T^\vartheta(\omega) F(X(\omega))\,\mathbf{P}(\mathrm{d}\omega)\\ &\quad= \int_\Omega \Big(\mathbb{E}_\mathbf{P}[\mathcal{E}_T^\vartheta F(X) \mid \sigma(X)]\Big)(\omega)\,\mathbf{P}(\mathrm{d}\omega)\\ &\quad= \int_\Omega \Big(\mathbb{E}_\mathbf{P}[\mathcal{E}_T^\vartheta \mid \sigma(X)]\Big)(\omega) F(X(\omega))\,\mathbf{P}(\mathrm{d}\omega).\end{aligned}$$

The first equality is the tower property. The second holds because $F(X)$ is $\sigma(X)$-measurable and can therefore be pulled out of the inner conditional expectation. The conditioning is introduced precisely because pushforward through X discards the remaining construction-space information: the conditional expectation is the $\sigma(X)$-measurable weight that preserves all averages against bounded functionals of X.

Since $X_\#\mathbf{P} = \mathbb{P}$ is a probability measure and $\mathcal{E}_T^\vartheta \geq 0$, the conditional-expectation form of the Doob–Dynkin lemma in Taraldsen (2018, Lemma 3), applied with $\Gamma = \mathcal{E}_T^\vartheta$ and $Y = X$, provides a Borel path functional z such that

$$\Big(\mathbb{E}_\mathbf{P}[\mathcal{E}_T^\vartheta \mid \sigma(X)]\Big)(\omega) = z(X(\omega)) \quad \text{for } \mathbf{P}\text{-almost every } \omega \in \Omega.$$

The definition of the pushforward law $\mathbb{P} = X_\#\mathbf{P}$ now gives

$$\int_\Omega z(X(\omega))F(X(\omega))\,\mathbf{P}(\mathrm{d}\omega) = \int z(\zeta)F(\zeta)\,\mathbb{P}(\mathrm{d}\zeta) = \mathbb{E}_\mathbb{P}[zF].$$

Together with the preceding identities, this is (45) and therefore proves

$$\frac{\mathrm{d}\widetilde{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} = z \qquad \mathbb{P}\text{-almost surely.}$$

There is no reason in general for $z(X(\omega))$ to equal $\mathcal{E}_T^\vartheta(\omega)$. If the latter is not determined by X, two outcomes of the construction can have the same observed path but different values of the proposed likelihood. Writing $\mathrm{d}\widetilde{\mathbb{Q}}/\mathrm{d}\mathbb{P}(X(\omega)) = \mathcal{E}_T^\vartheta(\omega)$ would then be ill-defined on coordinate path space.

In our case (X, ℓ) is supplied as the solution pair. After passage to coordinate path space only X remains, and the definition of a solution does not by itself show that ℓ is a function of X. Establishing this is the purpose of the following recovery argument. For $L_t := \int_0^t \mathbf{n}_{\mathrm{in}}(X_s)\,\mathrm{d}\ell_s$, the reflected SDE gives

$$W_t = \int_0^t g(s)^{-1}\big(\mathrm{d}X_s - b(X_s, s)\,\mathrm{d}s - \mathrm{d}L_s\big).$$

This does not reconstruct W from X until ℓ, and hence L, has been reconstructed from X. Let us first see how this can be done in one dimension. The following half-line calculation isolates the idea and fixes the factor dictated by our right-local-time convention. This result is presented for completeness and better motivation of further results, and is also available at Pilipenko (2014, Theorem 1.3.1).

Lemma D.1.2. *Fix $T > 0$. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbf{P})$ satisfy the usual conditions. Suppose it carries an $\mathbb{R}$-valued Brownian motion W relative to $(\mathcal{F}_t)_{0\leq t\leq T}$ and continuous adapted processes X and ℓ, with $X_0 \geq 0$ being $\mathcal{F}_0$-measurable, such that*

$$X_t = X_0 + W_t + \ell_t, \qquad 0 \leq t \leq T, \tag{46}$$

where $X_t \in [0, \infty)$, ℓ is nondecreasing with $\ell_0 = 0$, and $\mathrm{d}\ell$ is supported on $\{t : X_t = 0\}$. Then, almost surely, simultaneously for $0 \leq t \leq T$,

$$\ell_t = \frac{1}{2}\,\mathrm{LT}_t^0(X) = \lim_{n\to\infty} \frac{n}{2}\int_0^t \mathbb{1}_{\{0\leq X_s\leq 1/n\}}\,\mathrm{d}s. \tag{47}$$

Consequently, both ℓ and $W = X - X_0 - \ell$ have Borel versions as functions of the observed path X.

Proof. Note that the state space is the half-line, its boundary is the point zero, and the term $+\ell_t$ pushes the process inward, toward positive values, whenever reflection is needed. Here $[0, 1/n]$ is the interior boundary layer of width $1/n$. The one-sided interval appears because X can approach the boundary only from inside $[0, \infty)$ and $\mathrm{LT}^0(X)$ is the right local time at zero.

The finite-variation process ℓ does not contribute to quadratic variation, so $\langle X \rangle_t = t$. The zero-level consequence of the occupation-density formula in Le Gall (2016, Corollary 9.7) therefore gives

$$\int_0^t \mathbb{1}_{\{X_s=0\}} \, \mathrm{d}s = 0.$$

Apply the negative-part Tanaka identity in Le Gall (2016, Proposition 9.2) at level zero. Since $X \geq 0$, it yields

$$\frac{1}{2} \mathrm{LT}_t^0(X) = \int_0^t \mathbb{1}_{\{X_s=0\}} \, \mathrm{d}X_s = \int_0^t \mathbb{1}_{\{X_s=0\}} \, \mathrm{d}W_s + \ell_t.$$

The stochastic integral vanishes: its quadratic variation is $\int_0^t \mathbb{1}_{\{X_s=0\}} \, \mathrm{d}s = 0$. This proves the first equality in (47). Since $\mathrm{d}\langle X \rangle_s = \mathrm{d}s$, Le Gall (2016, Proposition 9.9) gives the second equality.

For each n, the approximating process on the right-hand side of (47) is a Borel function of the path X. Its pointwise limit is the continuous nondecreasing path ℓ, so the convergence is uniform on $[0, T]$. Defining the resulting path functional by this uniform limit when it exists, and by the zero path otherwise, gives a Borel version $\mathfrak{l}$ of ℓ. Now define

$$\mathfrak{w}(\zeta)(t) := \zeta(t) - \zeta(0) - \mathfrak{l}(\zeta)(t).$$

This is a Borel path functional, and $W = \mathfrak{w}(X) = X - X_0 - \ell$ almost surely. □

Thus, on the half-line, the observed path first determines its cumulative boundary push and then determines its Brownian coordinate.

The higher-dimensional argument uses the same idea after reducing the boundary behavior to one dimension. For the sake of completeness, we provide full construction and proof of the argument. Similar arguments are present in Holk et al. (2026, Eq. 3), but the exact form differs from the one obtained in our Lemma D.1.4 due to differences in problem formulation.

We choose a smooth function φ that vanishes precisely on $\partial\mathcal{D}$ and whose gradient there is the inward unit normal. Tanaka's formula applied to the nonnegative scalar process $\varphi(X)$ recovers the scalar boundary regulator from the local time of $\varphi(X)$ at zero. For $L_t := \int_0^t \mathbf{n}_{\mathrm{in}}(X_s) \, \mathrm{d}\ell_s$, this also recovers L. Substituting it into the reflected SDE then recovers W. The next three lemmas make the boundary coordinate, regulator recovery, and Brownian-coordinate recovery precise.

Lemma D.1.3. *Suppose $\mathcal{D}$ is bounded and has C^∞ boundary. There is $\varphi \in C^\infty(\mathbb{R}^d)$ such that*

$$\varphi \geq 0 \text{ on } \overline{\mathcal{D}}, \qquad \{x \in \overline{\mathcal{D}} : \varphi(x) = 0\} = \partial\mathcal{D}, \qquad \nabla\varphi = \mathbf{n}_{\mathrm{in}} \text{ on } \partial\mathcal{D}. \tag{48}$$

Proof. Define the signed distance from the boundary, with the sign chosen positive inside $\mathcal{D}$, by

$$d_{\mathrm{s}}(x) := \begin{cases} \mathrm{dist}(x, \partial\mathcal{D}), & x \in \overline{\mathcal{D}}, \\ -\mathrm{dist}(x, \partial\mathcal{D}), & x \notin \overline{\mathcal{D}}. \end{cases}$$

Theorem 1 and the preliminary lemma of Foote (1984) give a neighborhood of the compact smooth hypersurface $\partial\mathcal{D}$ with a unique nearest-point projection, on which the unsigned distance is smooth away from $\partial\mathcal{D}$. In every local defining chart, Foote (1984, Remark 2) shows that assigning a sign according to the two sides extends the distance smoothly across the hypersurface. With the sign positive in $\mathcal{D}$, every local construction is the restriction of the same globally defined function d_{s}, so no gluing is needed: the local neighborhoods can be united to give an open neighborhood N of $\partial\mathcal{D}$ on which d_{s} is smooth. Moreover, for $z \in \partial\mathcal{D}$ and all sufficiently small $s > 0$, the point $z + s\mathbf{n}_{\mathrm{in}}(z)$ lies in $\mathcal{D}$ and has z as its nearest boundary point. Consequently

$$d_{\mathrm{s}}\big(z + s\mathbf{n}_{\mathrm{in}}(z)\big) = s.$$

Since d_{s} vanishes on $\partial\mathcal{D}$, its derivative there annihilates every tangent vector. The last identity gives its derivative in the inward-normal direction. Hence

$$\nabla d_{\mathrm{s}}(z) = \mathbf{n}_{\mathrm{in}}(z), \qquad z \in \partial\mathcal{D}. \tag{49}$$

It remains only to extend this local coordinate smoothly away from the boundary. Compactness of $\partial\mathcal{D}$ allows us to choose $\varepsilon > 0$ such that $\{x : \operatorname{dist}(x, \partial\mathcal{D}) < 3\varepsilon\} \subset N$, and choose a nondecreasing function $\rho \in C^\infty(\mathbb{R})$ satisfying

$$\rho(s) = s \quad (|s| \le \varepsilon), \qquad \rho(s) = 2\varepsilon \quad (s \ge 2\varepsilon), \qquad \rho(s) = -2\varepsilon \quad (s \le -2\varepsilon).$$

This is the static version of the modified signed-distance construction in Alfaro et al. (2008, Section 5.1, equations (5.2)–(5.3)). Writing $N_{3\varepsilon} := \{x : \operatorname{dist}(x, \partial\mathcal{D}) < 3\varepsilon\}$, define

$$\varphi(x) := \begin{cases} \rho(d_{\mathrm{s}}(x)), & x \in N_{3\varepsilon}, \\ 2\varepsilon, & x \in \mathcal{D} \setminus N_{3\varepsilon}, \\ -2\varepsilon, & x \in \mathbb{R}^d \setminus (\overline{\mathcal{D}} \cup N_{3\varepsilon}). \end{cases}$$

Near the boundary of $N_{3\varepsilon}$, the first expression is already the corresponding constant, so these definitions give a function in $C^\infty(\mathbb{R}^d)$. Monotonicity of ρ and $\rho(s) = s$ near zero show that $\varphi > 0$ in $\mathcal{D}$ and $\varphi = 0$ on $\partial\mathcal{D}$. Finally, (49) and $\rho'(0) = 1$ give $\nabla\varphi = \mathbf{n}_{\mathrm{in}}$ on $\partial\mathcal{D}$. □

Lemma D.1.4. *Fix $T > 0$, let φ be as in Lemma D.1.3, and let $g \in C([0,T])$ satisfy $\inf_{0\le t\le T}|g(t)| > 0$. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbf{P})$ satisfy the usual conditions. Suppose it carries an $\mathbb{R}^d$-valued Brownian motion W relative to $(\mathcal{F}_t)_{0\le t\le T}$, a continuous adapted $\overline{\mathcal{D}}$-valued process X with X_0 being $\mathcal{F}_0$-measurable, a bounded progressively measurable process b, and a continuous adapted nondecreasing process ℓ with $\ell_0 = 0$. Suppose*

$$X_t = X_0 + \int_0^t b_s \,\mathrm{d}s + \int_0^t g(s)\,\mathrm{d}W_s + L_t, \qquad L_t = \int_0^t \mathbf{n}_{\mathrm{in}}(X_s)\,\mathrm{d}\ell_s, \tag{50}$$

where $\mathrm{d}\ell$ is supported on $\{t : X_t \in \partial\mathcal{D}\}$. Then, almost surely, X spends zero Lebesgue time on the boundary and

$$\ell_t = \frac{1}{2}\,\mathrm{LT}_t^0(\varphi(X)) = \lim_{n\to\infty} \frac{n}{2}\int_0^t \mathbb{1}_{\{0\le\varphi(X_s)\le 1/n\}} g(s)^2 \|\nabla\varphi(X_s)\|_2^2 \,\mathrm{d}s \quad (0 \le t \le T). \tag{51}$$

Consequently, there are fixed Borel maps

$$\mathfrak{l} : C([0,T];\overline{\mathcal{D}}) \to C([0,T];\mathbb{R}), \qquad \mathfrak{L} : C([0,T];\overline{\mathcal{D}}) \to C([0,T];\mathbb{R}^d)$$

such that $\ell = \mathfrak{l}(X)$ and $L = \mathfrak{L}(X)$ almost surely.

Proof. Set $Y_t^\partial := \varphi(X_t) \ge 0$. This is the boundary-distance coordinate. The multidimensional Itô formula in Le Gall (2016, Theorem 5.10) gives

$$\mathrm{d}Y_t^\partial = \langle\nabla\varphi(X_t), b_t\rangle\,\mathrm{d}t + g(t)\langle\nabla\varphi(X_t), \mathrm{d}W_t\rangle + \frac{1}{2}g(t)^2\Delta\varphi(X_t)\,\mathrm{d}t + \mathrm{d}\ell_t. \tag{52}$$

Indeed, $\mathrm{d}\ell$ is supported on the boundary, where $\nabla\varphi = \mathbf{n}_{\mathrm{in}}$. The quadratic variation is

$$\langle Y^\partial\rangle_t = \int_0^t g(s)^2\|\nabla\varphi(X_s)\|_2^2\,\mathrm{d}s.$$

The occupation-density formula in Le Gall (2016, Corollary 9.7) assigns zero quadratic variation to the single level $Y^\partial = 0$. Since $\|\nabla\varphi\|_2 = 1$ on the boundary and $\inf_t|g(t)| > 0$,

$$\int_0^T \mathbb{1}_{\{X_s\in\partial\mathcal{D}\}}\,\mathrm{d}s = 0. \tag{53}$$

Apply the negative-part Tanaka identity in Le Gall (2016, Proposition 9.2) to the nonnegative process Y^∂. It gives

$$\frac{1}{2}\,\mathrm{LT}_t^0(Y^\partial) = \int_0^t \mathbb{1}_{\{Y_s^\partial=0\}}\,\mathrm{d}Y_s^\partial.$$

Inserting (52), the restricted martingale has zero quadratic variation, both ordinary-time terms vanish by (53) because their integrands are bounded, and the regulator term equals ℓ_t. This proves the first equality in (51). Le Gall (2016, Proposition 9.9) gives the second.

Measurability. For a path $\zeta \in C([0,T];\overline{\mathcal{D}})$, let

$$\mathrm{I}_n(\zeta)_t := \frac{n}{2}\int_0^t \mathbb{1}_{\{0\le\varphi(\zeta(s))\le 1/n\}} g(s)^2 \|\nabla\varphi(\zeta(s))\|_2^2 \,\mathrm{d}s.$$

Each I_n is a Borel map into $C([0,T];\mathbb{R})$, and every resulting time path is nondecreasing. Define $\mathrm{I}(\zeta)$ as their uniform limit when they are uniformly Cauchy, and as the zero path otherwise. This is a Borel map. On the probability-one event in (51), the approximants converge pointwise to the continuous nondecreasing path ℓ. Such convergence is uniform on a compact interval. Hence $\mathrm{I}(X) = \ell$. Finally define

$$\mathfrak{L}(\zeta)_t := \int_0^t \nabla\varphi(\zeta(s))\,\mathrm{d}\mathrm{I}(\zeta)_s.$$

Deterministic Riemann–Stieltjes sums show that $\mathfrak{L}$ is Borel, and the boundary support of $\mathrm{d}\ell$ gives $\mathfrak{L}(X) = L$. □

Lemma D.1.5. *In Lemma D.1.4, suppose in addition that $b_t = b(X_t,t)$ for some $b \in C(\overline{\mathcal{D}}\times[0,T];\mathbb{R}^d)$. Then the paths L and W are almost surely Borel functions of X. Moreover, for every $v \in C(\overline{\mathcal{D}}\times[0,T];\mathbb{R}^d)$, the terminal Itô integral*

$$\int_0^T v(X_t,t)^\top \mathrm{d}W_t \tag{54}$$

has a Borel version on $C([0,T];\overline{\mathcal{D}})$.

Proof. The preceding lemma already gives $L = \mathfrak{L}(X)$. Put

$$B_t(X) := \int_0^t b(X_s,s)\,\mathrm{d}s, \qquad S_t(X) := X_t - L_t - B_t(X).$$

Then $S_t = X_0 + \int_0^t g(s)\,\mathrm{d}W_s$. For the schedule in (5), $h := g^{-1}$ is C^1, and integration by parts gives the explicit pathwise formula

$$W_t = h(t)S_t - h(0)S_0 - \int_0^t S_s\,\mathrm{d}h(s).$$

For a general continuous nonvanishing g, stochastic-integral associativity from Le Gall (2016, Theorem 5.6) gives

$$W_t = \int_0^t g(s)^{-1}\,\mathrm{d}S_s. \tag{55}$$

Le Gall (2016, Proposition 5.9) approximates this integral by deterministic left-point sums. A single diagonal almost-sure subsequence over rational times and continuity give one Borel version of the full path W: define the continuous extension on the Borel set where it exists, and use the zero path elsewhere. The same argument applied to $v(X_t,t)$ gives a Borel version of (54). The selected subsequence, and hence the Borel version, may depend on the fixed weak law and on v. No version simultaneous over all laws and integrands is asserted. □

We can now combine the standard construction-space change with the recovery lemmas. Existence and uniqueness in law are assumed here rather than derived: their only role is to make the two SDEs denote unambiguous path laws and to identify the weak solution produced by Girsanov. An analogous result for a slightly different setting is available at Holk et al. (2026, Appendix A).

Theorem D.1.6. *Let $\mathcal{D} \subset \mathbb{R}^d$ be bounded and convex with C^∞ boundary. Let $\alpha,\beta \in C(\overline{\mathcal{D}}\times[0,1];\mathbb{R}^d)$, let $g \in C([0,1])$ satisfy $\inf_t|g(t)| > 0$. In the two equations below, $L^{\mathbb{P}}$ and $L^{\mathbb{Q}}$ denote the vector reflection terms derived from the scalar regulators in their respective weak realizations:*

$$\mathrm{d}X_t = \alpha(X_t,t)\,\mathrm{d}t + g(t)\,\mathrm{d}W_t^{\mathbb{P}} + \mathrm{d}L_t^{\mathbb{P}}, \qquad \mathrm{d}X_t = \beta(X_t,t)\,\mathrm{d}t + g(t)\,\mathrm{d}W_t^{\mathbb{Q}} + \mathrm{d}L_t^{\mathbb{Q}} \tag{56}$$

Suppose these equations have a common initial law μ, admit weak solutions, and have unique path laws $\mathbb{P}$ and $\mathbb{Q}$ on $C([0,1];\overline{\mathcal{D}})$. Then $\mathbb{P} \ll \mathbb{Q} \ll \mathbb{P}$. Let $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbf{Q};\xi_0,W^{\mathbb{Q}},X,\ell)$ be any weak realization of the second SDE in (56)*. Then, $\mathbf{Q}$-almost surely,*

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(X) = \exp\left(-\int_0^1 u(X_t,t)^\top \mathrm{d}W_t^{\mathbb{Q}} + \frac{1}{2}\int_0^1 \|u(X_t,t)\|_2^2\,\mathrm{d}t\right), \qquad u := \frac{\alpha-\beta}{g}. \tag{57}$$

Equivalently, on any weak realization of the first SDE with construction probability $\mathbf{P}$*,* $\mathbf{P}$*-almost surely,*

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(X) = \exp\left(-\int_0^1 u(X_t,t)^\top \mathrm{d}W_t^{\mathbb{P}} - \frac{1}{2}\int_0^1 \|u(X_t,t)\|_2^2\,\mathrm{d}t\right). \tag{58}$$

Proof. Start from an arbitrary weak realization $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbf{P}; \xi_0, W^{\mathbb{P}}, X, \ell)$ of the first SDE. The function $u = (\alpha - \beta)/g$ is bounded and continuous on the compact cylinder. By the definition of a weak realization, X and ℓ are continuous and adapted, $\xi_0 = X_0$ is $\mathcal{F}_0$-measurable, and $W^{\mathbb{P}}$ is Brownian relative to the construction filtration. Thus the hypotheses of Lemma D.1.1 hold with $\vartheta_t = -u(X_t,t)$. Under the resulting equivalent probability,

$$W_t^{\mathbb{Q}} = W_t^{\mathbb{P}} + \int_0^t u(X_s,s)\,\mathrm{d}s$$

is Brownian, and $(\Omega, \mathcal{F}, (\mathcal{F}_t), \widetilde{\mathbf{Q}}; \xi_0, W^{\mathbb{Q}}, X, \ell)$ is a weak realization of the second SDE. Its initial law is still μ, and uniqueness in law identifies its coordinate law with $\mathbb{Q}$.

Lemmas D.1.4 and D.1.5 show that ℓ, hence $L^{\mathbb{P}}$, as well as $W^{\mathbb{P}}$ and the stochastic integral in (58) have Borel versions determined by X. Thus the construction-space density equals $z(X)$ for a Borel path functional z. Hence, for every bounded Borel path functional F,

$$\mathbb{E}_{\mathbb{Q}}[F] = \mathbb{E}_{\mathbb{P}}[zF].$$

This is exactly the statement that $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = z$ on coordinate path space. Strict positivity of z gives equivalence.

To obtain the formula on an arbitrary realization of the second SDE, its weak-solution properties provide the same adaptedness, scalar-regulator, Brownian-motion, and $\mathcal{F}_0$-measurability hypotheses. Apply the same argument in the opposite direction with $\vartheta_t = u(X_t,t)$. The changed drift is $\beta + gu = \alpha$, so uniqueness in law identifies the changed coordinate law with $\mathbb{P}$. Recovery and pushforward therefore give

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(X) = \exp\left(\int_0^1 u(X_t,t)^\top \mathrm{d}W_t^{\mathbb{Q}} - \frac{1}{2}\int_0^1 \|u(X_t,t)\|_2^2\,\mathrm{d}t\right).$$

Taking the reciprocal proves (57) on every such realization. □

E Formal statements and proofs for other main-text results

E.1 Optimal path proposal and terminal tilt

Proposition E.1.1. *Suppose Assumptions B.2.1, B.2.2, and B.2.3 hold. Then*

$$\frac{\mathrm{d}\overrightarrow{\mathbb{Q}}^\star}{\mathrm{d}\overrightarrow{\mathbb{P}}}(X) = \frac{\pi(X_0)}{\nu(X_0)}. \tag{59}$$

Reversing both path laws gives

$$\frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{P}}}(Y) = \frac{\pi(Y_1)}{\nu(Y_1)}.$$

For every proposal law $\overrightarrow{\mathbb{Q}}$ *equivalent to* $\overrightarrow{\mathbb{P}}$*, the importance weight in* (7) *satisfies*

$$W_{\overrightarrow{\mathbb{Q}}}(X) = Z\frac{\mathrm{d}\overrightarrow{\mathbb{Q}}^\star}{\mathrm{d}\overrightarrow{\mathbb{Q}}}(X), \qquad \mathrm{Var}_{\overrightarrow{\mathbb{Q}}}(W_{\overrightarrow{\mathbb{Q}}}) = Z^2\chi^2(\overrightarrow{\mathbb{Q}}^\star \| \overrightarrow{\mathbb{Q}}).$$

Consequently, $\overrightarrow{\mathbb{Q}}^\star$ *is the unique variance-minimizing proposal and* $W_{\overrightarrow{\mathbb{Q}}^\star} = Z$ *almost surely.*

Proof. The laws $\overrightarrow{\mathbb{P}}$ and $\overrightarrow{\mathbb{Q}}^\star$ have the same reflected dynamics and differ only in their initial distributions, which are ν and π, respectively. Conditional on $X_0 = x$, their trajectory laws therefore coincide. Formally, the deterministic-start solution map supplies a common measurable version of

these conditional path laws. To turn this observation into a path-law identity, let $F : C \to \mathbb{R}$ be bounded and Borel measurable. Conditioning on X_0 gives

$$\begin{aligned}\mathbb{E}_{\overrightarrow{\mathbb{Q}}^\star}[F(X)] &= \int_{\mathcal{D}} \mathbb{E}_{\overrightarrow{\mathbb{Q}}^\star}[F(X) \mid X_0 = x]\ \pi(x)\,\mathrm{d}x \\ &= \int_{\mathcal{D}} \mathbb{E}_{\overrightarrow{\mathbb{P}}}[F(X) \mid X_0 = x]\ \frac{\pi(x)}{\nu(x)}\,\nu(x)\,\mathrm{d}x \\ &= \mathbb{E}_{\overrightarrow{\mathbb{P}}}\left[F(X)\frac{\pi(X_0)}{\nu(X_0)}\right].\end{aligned}$$

Since this holds for every bounded Borel F, it proves (59). Assumption B.2.3 directly makes π/ν strictly positive and bounded above and away from zero on $\overline{\mathcal{D}}$, so $\overrightarrow{\mathbb{P}}$ and $\overrightarrow{\mathbb{Q}}^\star$ are equivalent.

Under path reversal, $Y_{\bar{t}} = X_{1-\bar{t}}$ and therefore $X_0 = Y_1$. Applying the preceding expectation identity to the reversed-path functional gives, for every bounded Borel F,

$$\mathbb{E}_{\overleftarrow{\mathbb{Q}}^\star}[F(Y)] = \mathbb{E}_{\overleftarrow{\mathbb{P}}}\left[F(Y)\frac{\pi(Y_1)}{\nu(Y_1)}\right].$$

Thus the terminal factor changes the entire reverse path law from $\overleftarrow{\mathbb{P}}$ to $\overleftarrow{\mathbb{Q}}^\star$, whose initial and terminal marginals are p_1 and π.

Now let $\overrightarrow{\mathbb{Q}}$ be any path law equivalent to $\overrightarrow{\mathbb{P}}$. Using $r = Z\pi$, (7), (59), and the Radon–Nikodym chain rule gives

$$W_{\overrightarrow{\mathbb{Q}}}(X) = \frac{r(X_0)}{\nu(X_0)}\frac{\mathrm{d}\overrightarrow{\mathbb{P}}}{\mathrm{d}\overrightarrow{\mathbb{Q}}}(X) = Z\frac{\mathrm{d}\overrightarrow{\mathbb{Q}}^\star}{\mathrm{d}\overrightarrow{\mathbb{P}}}(X)\frac{\mathrm{d}\overrightarrow{\mathbb{P}}}{\mathrm{d}\overrightarrow{\mathbb{Q}}}(X) = Z\frac{\mathrm{d}\overrightarrow{\mathbb{Q}}^\star}{\mathrm{d}\overrightarrow{\mathbb{Q}}}(X).$$

In particular, $\mathbb{E}_{\overrightarrow{\mathbb{Q}}}[W_{\overrightarrow{\mathbb{Q}}}] = Z$. Allowing infinite variance,

$$\mathrm{Var}_{\overrightarrow{\mathbb{Q}}}(W_{\overrightarrow{\mathbb{Q}}}) = \mathbb{E}_{\overrightarrow{\mathbb{Q}}}\left[(W_{\overrightarrow{\mathbb{Q}}} - Z)^2\right] = Z^2\mathbb{E}_{\overrightarrow{\mathbb{Q}}}\left[\left(\frac{\mathrm{d}\overrightarrow{\mathbb{Q}}^\star}{\mathrm{d}\overrightarrow{\mathbb{Q}}} - 1\right)^2\right] = Z^2\chi^2(\overrightarrow{\mathbb{Q}}^\star \| \overrightarrow{\mathbb{Q}}).$$

The right-hand side is nonnegative and is zero exactly when $\mathrm{d}\overrightarrow{\mathbb{Q}}^\star/\mathrm{d}\overrightarrow{\mathbb{Q}} = 1$ almost surely, that is, exactly when $\overrightarrow{\mathbb{Q}} = \overrightarrow{\mathbb{Q}}^\star$. For this proposal, $W_{\overrightarrow{\mathbb{Q}}^\star} = Z$ almost surely. □

E.2 Continuous training KL

Theorem E.2.1. *Suppose Assumptions B.2.1, B.2.2, B.2.3, B.3.3, and B.3.4 hold. Then the reflected SDEs* (10) *and* (12) *have unique strong solutions and are unique in law. Their unique path laws are* $\overleftarrow{\mathbb{P}}$ *and* $\overleftarrow{\mathbb{Q}}_\theta$*, respectively, and the training identity* (15) *holds.*

Proof. Lemma C.3.2 proves that the reference equation (10) admits a strong solution, has pathwise uniqueness, and is unique in law, with unique solution law $\overleftarrow{\mathbb{P}}$.

For the learned equation, in the normal-reflection notation of Pilipenko (2014, Section 2.5), take

$$a(\bar{t}, y) := g(1-\bar{t})^2 s_\theta(y, 1-\bar{t}), \qquad b_k(\bar{t}, y) := g(1-\bar{t})e_k, \quad k \in \{1, \dots, d\}.$$

Assumption B.3.3, compactness of $[0,1] \times \overline{\mathcal{D}}$, and continuity of the schedule make these coefficients bounded and continuous. Assumption B.3.4 and boundedness of g make them uniformly Lipschitz in space. Extend the coefficients constantly after time 1, these properties are preserved. Under Assumptions B.2.1 and B.2.2, Lemma C.1.1 gives the uniform exterior-sphere and uniform-cone conditions required by Pilipenko (2014, Theorem 2.5.3), which gives weak existence. Its exterior-sphere conclusion is the domain hypothesis of Pilipenko (2014, Theorem 2.5.1), which gives pathwise uniqueness. The Yamada–Watanabe discussion immediately preceding Pilipenko (2014, Theorem 2.5.4) then gives a unique strong solution and uniqueness in law. This defines $\overleftarrow{\mathbb{Q}}_\theta$ and, together with Lemma C.3.2, verifies the existence-and-uniqueness-in-law hypotheses of Theorem D.1.6.

The continuity and positive lower bound of the schedule in (5), together with the coefficient continuity just checked, verify the remaining hypotheses of Theorem D.1.6. Both $\overleftarrow{\mathbb{P}}$ and $\overleftarrow{\mathbb{Q}}_\theta$ start from ν, so the theorem applies to this pair. Specifically, take reference drift $\alpha = 0$, learned drift $\beta(y, \bar{t}) = g(1-\bar{t})^2 s_\theta(y, 1-\bar{t})$, and diffusion coefficient $g(1-\bar{t})$. Then $u = -g(1-\bar{t})s_\theta$, so under $\overleftarrow{\mathbb{Q}}_\theta$,

$$\log \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}{\mathrm{d}\overleftarrow{\mathbb{P}}}(Y) = \int_0^1 g(1-\bar{t}) s_\theta(Y_{\bar{t}}, 1-\bar{t})^\top \mathrm{d}\overline{W}_{\bar{t}} + \frac{1}{2}\int_0^1 g(1-\bar{t})^2 \|s_\theta(Y_{\bar{t}}, 1-\bar{t})\|_2^2 \,\mathrm{d}\bar{t}. \tag{60}$$

The theorem is not applied directly to $\overleftarrow{\mathbb{Q}}_\theta$ and $\overleftarrow{\mathbb{Q}}^\star$, whose initial distributions are ν and p_1. Here $\overline{W}$ is Brownian under the construction probability inducing $\overleftarrow{\mathbb{Q}}_\theta$. Its integrand is predictable because Y is adapted and continuous, and Assumption B.3.3, compactness, and boundedness of g make it bounded. The stochastic integral is therefore a square-integrable martingale and has mean zero. Proposition E.1.1 gives the terminal tilt. Using $\overleftarrow{\mathbb{P}}$ as the intermediate law, the Radon–Nikodym chain rule gives

$$\frac{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}(Y) = \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}{\mathrm{d}\overleftarrow{\mathbb{P}}}(Y)\frac{\nu(Y_1)}{\pi(Y_1)}.$$

Taking logarithms and expectation under $\overleftarrow{\mathbb{Q}}_\theta$ yields

$$\mathsf{KL}(\overleftarrow{\mathbb{Q}}_\theta \| \overleftarrow{\mathbb{Q}}^\star) = \mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}\left[\log \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}{\mathrm{d}\overleftarrow{\mathbb{P}}}(Y) + \log \frac{\nu(Y_1)}{\pi(Y_1)}\right].$$

The strict positivity in Assumption B.2.3 gives $\log(\nu/\pi) = \log(\nu/r) + \log Z$, while its two-sided bounds make this terminal logarithm bounded. This proves (15). □

E.3 Normalizing-constant accuracy for the continuous reflected sampler

Theorem E.3.1 gives a total-variation accuracy bound for the normalizing-constant estimator (16). It is a simplified reflected version of Guo et al. (2026, Theorem 5), formulated for continuous ideal and learned path laws on $[0, 1]$.

In that argument, the formal derivation of the mixed path-density formula introduces stationary Wiener measure as an intermediate reference. This is the Brownian path measure with Lebesgue marginals. The use of a Wiener reference also appears in the stochastic-control formulations of Chen et al. (2016, Section III) and Chen et al. (2021, Section 4.4). A general forward–backward factorization through a common reference is given by Vargas et al. (2024, Proposition 2.2 and Appendix E.1).

We instead use $\overleftarrow{\mathbb{P}}$, the path law of a deterministic time change of stationary normally reflected Brownian motion initialized from the uniform distribution ν. This is a probability measure whose marginals remain uniform throughout $[0, 1]$, as shown in Lemma C.3.2. Proposition E.1.1 and Theorem D.1.6 provide the two density factors relative to $\overleftarrow{\mathbb{P}}$ needed for the comparison.

Theorem E.3.1. *Suppose Assumptions B.2.1, B.2.2, B.2.3, B.3.3, and B.3.4 hold. Fix θ and an integer $N \geq 1$, and let $Y^{(1)}, \ldots, Y^{(N)}$ be paths, each with law $\overleftarrow{\mathbb{Q}}_\theta$. Let $\widehat{Z}$ be the estimator* (16)*. Expectations and probabilities below are over these sampled paths. Then $\mathbb{E}[\widehat{Z}] = Z$ and, for every $\varepsilon > 0$,*

$$\Pr\left(\left|\frac{\widehat{Z}}{Z} - 1\right| \geq \varepsilon\right) \leq \frac{2\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}}}{\varepsilon}. \tag{61}$$

In particular, for every $\varepsilon > 0$ and $\delta \in (0, 1)$,

$$\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}} \leq \frac{\varepsilon\delta}{2} \quad \Longrightarrow \quad \Pr\left(\left|\frac{\widehat{Z}}{Z} - 1\right| \geq \varepsilon\right) \leq \delta. \tag{62}$$

Proof. Assumption B.2.3 and boundedness of $\mathcal{D}$ give $0 < Z < \infty$ and make π/ν bounded above and away from zero. Together with Assumptions B.2.1 and B.2.2, these are the hypotheses of

Proposition E.1.1, which gives $\overleftarrow{\mathbb{Q}}^\star \ll \overleftarrow{\mathbb{P}} \ll \overleftarrow{\mathbb{Q}}^\star$ and

$$\frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{P}}}(Y) = \frac{\pi(Y_1)}{\nu(Y_1)}.$$

Under the same domain and target assumptions, Assumptions B.3.3 and B.3.4 allow Theorem E.2.1 to give the unique laws $\overleftarrow{\mathbb{P}}$ and $\overleftarrow{\mathbb{Q}}_\theta$. The Girsanov comparison in its proof establishes $\overleftarrow{\mathbb{Q}}_\theta \ll \overleftarrow{\mathbb{P}} \ll \overleftarrow{\mathbb{Q}}_\theta$ and the density formula (60). Define

$$F(Y) := \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}(Y) = \frac{r(Y_1)}{Z\nu(Y_1)} \frac{\mathrm{d}\overleftarrow{\mathbb{P}}}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}(Y), \qquad \overleftarrow{\mathbb{Q}}_\theta\text{-almost surely.} \tag{63}$$

Since $\overleftarrow{\mathbb{Q}}^\star$ is a probability measure,

$$\mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}[F(Y)] = \int \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta} \mathrm{d}\overleftarrow{\mathbb{Q}}_\theta = \int \mathrm{d}\overleftarrow{\mathbb{Q}}^\star = 1.$$

Linearity of expectation therefore gives

$$\mathbb{E}\left[\frac{\widehat{Z}}{Z}\right] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}[F(Y)] = 1,$$

which proves unbiasedness. Then, we have the following bound:

$$\mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}\left[\left|\frac{\widehat{Z}}{Z} - 1\right|\right] = \mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}\left[\left|\frac{1}{N}\sum_{i=1}^{N}(F(Y^{(i)}) - 1)\right|\right] \le \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}[|F(Y^{(i)}) - 1|] = \mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}[|F(Y^{(1)}) - 1|]. \tag{64}$$

Under the total-variation convention in Section B.1,

$$\mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}[|F(Y) - 1|] = \int \left|\frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta} - 1\right| \mathrm{d}\overleftarrow{\mathbb{Q}}_\theta = 2\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}}.$$

Substituting this identity into (64) yields

$$\mathbb{E}\left[\left|\frac{\widehat{Z}}{Z} - 1\right|\right] \le 2\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}}. \tag{65}$$

Finally, Markov's inequality gives

$$\Pr\left(\left|\frac{\widehat{Z}}{Z} - 1\right| \ge \varepsilon\right) \le \frac{\mathbb{E}[|\widehat{Z}/Z - 1|]}{\varepsilon} \le \frac{2\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}}}{\varepsilon}.$$

This proves (61). If $\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}} \le \varepsilon\delta/2$, then $2\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}}/\varepsilon \le \delta$, proving (62). □

The following theorem expresses the ideal-to-learned KL in terms of the initial-law mismatch and score error.

Theorem E.3.2. *Let Z be the normalizing constant defined in* (1)*. Suppose $Z > 0$ and Assumptions B.2.1, B.2.2, B.3.3, B.3.4, B.3.1, and B.3.2 hold. Fix θ. Then*

$$\mathsf{KL}(\overleftarrow{\mathbb{Q}}^\star \| \overleftarrow{\mathbb{Q}}_\theta) = \mathsf{KL}(p_1 \| \nu) + \frac{1}{2}\int_0^1 g(t)^2 \mathbb{E}_{X\sim p_t}\left[\|s_\theta(X,t) - \nabla \log p_t(X)\|_2^2\right] \mathrm{d}t. \tag{66}$$

Proof. Step 1: separate the initial-law change from the drift change. The standing integrability of r in (1), together with $Z > 0$, makes $\pi = r/Z$ a probability density. Assumptions B.2.1 and B.2.2 together with Assumptions B.3.1 and B.3.2 are precisely the hypotheses of the full-interval uniqueness conclusion in Lemma C.4.1. Hence $\overleftarrow{\mathbb{Q}}^\star$ is the unique law of (11). Theorem D.1.6 requires

a common initial law, so we cannot apply it directly to $\overleftarrow{\mathbb{Q}}^\star$ and $\overleftarrow{\mathbb{Q}}_\theta$: the former starts from p_1 and the latter from ν.

To handle this difference, let $\overleftarrow{\mathbb{Q}}_\theta^{p_1}$ be the law of the learned SDE (12) started from p_1. The well-posedness part of the proof of Theorem E.2.1 uses only Assumptions B.2.1, B.2.2, B.3.3, and B.3.4, and applies to every prescribed initial law. It therefore gives unique laws $\overleftarrow{\mathbb{Q}}_\theta$ and $\overleftarrow{\mathbb{Q}}_\theta^{p_1}$ for initial distributions ν and p_1. These laws have the same conditional trajectory law given their initial point: for a fixed initial value they solve the same uniquely solvable SDE.

For every bounded Borel path functional H, conditioning on the initial point and using this common conditional law gives

$$\begin{aligned}\mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta^{p_1}}[H(Y)] &= \int_{\mathcal{D}} \mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}[H(Y) \mid Y_0 = y]\, p_1(y)\,\mathrm{d}y \\ &= \int_{\mathcal{D}} \mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}[H(Y) \mid Y_0 = y] \frac{p_1(y)}{\nu(y)}\,\nu(y)\,\mathrm{d}y \\ &= \mathbb{E}_{\overleftarrow{\mathbb{Q}}_\theta}\left[H(Y)\frac{p_1(Y_0)}{\nu(Y_0)}\right].\end{aligned}$$

This is the defining identity for

$$\frac{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta^{p_1}}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}(Y) = \frac{p_1(Y_0)}{\nu(Y_0)}. \tag{67}$$

For the probability density π, the positive-time density properties in Section B.2 give a continuous, strictly positive p_1 on the compact closure. Thus p_1/ν and its reciprocal are bounded, and its logarithm is integrable.

It remains to compare $\overleftarrow{\mathbb{Q}}_\theta^{p_1}$ with $\overleftarrow{\mathbb{Q}}^\star$. These laws now have the same initial law p_1, diffusion coefficient $g(1-\bar{t})$, and normal reflection. Their drifts are

$$\alpha(y,\bar{t}) = g(1-\bar{t})^2 s_\theta(y, 1-\bar{t}), \qquad \beta(y,\bar{t}) = g(1-\bar{t})^2 \nabla \log p_{1-\bar{t}}(y).$$

Assumptions B.3.3 and B.3.1, compactness, and the schedule (5) make both drifts continuous and bounded. The domain conditions and the common nonvanishing diffusion coefficient therefore allow Theorem D.1.6 to compare these two laws.

Define the drift difference in diffusion coordinates by

$$e(y,\bar{t}) := \frac{\beta(y,\bar{t}) - \alpha(y,\bar{t})}{g(1-\bar{t})} = g(1-\bar{t})\big(\nabla \log p_{1-\bar{t}}(y) - s_\theta(y, 1-\bar{t})\big).$$

In Theorem D.1.6, the field $u = (\alpha - \beta)/g(1-\bar{t})$ is therefore $-e$. Its formula, written under a weak realization of $\overleftarrow{\mathbb{Q}}^\star$, gives

$$\log \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta^{p_1}}(Y) = \int_0^1 e(Y_{\bar{t}},\bar{t})^\top \,\mathrm{d}\overline{W}_{\bar{t}}^\star + \frac{1}{2}\int_0^1 \|e(Y_{\bar{t}},\bar{t})\|_2^2\,\mathrm{d}\bar{t}, \tag{68}$$

where $\overline{W}^\star$ is the Brownian motion in that ideal-law realization. In particular, the positive half-energy term is the one in the density formula under the numerator law $\overleftarrow{\mathbb{Q}}^\star$.

We can now combine the two changes by the Radon–Nikodym chain rule:

$$\frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta} = \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta^{p_1}} \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta^{p_1}}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}.$$

Taking logarithms and substituting (67) and (68) yields

$$\log \frac{\mathrm{d}\overleftarrow{\mathbb{Q}}^\star}{\mathrm{d}\overleftarrow{\mathbb{Q}}_\theta}(Y) = \log \frac{p_1(Y_0)}{\nu(Y_0)} + \int_0^1 e(Y_{\bar{t}},\bar{t})^\top \,\mathrm{d}\overline{W}_{\bar{t}}^\star + \frac{1}{2}\int_0^1 \|e(Y_{\bar{t}},\bar{t})\|_2^2\,\mathrm{d}\bar{t}. \tag{69}$$

Step 2: take the ideal-law expectation. KL is the expectation of this log density under its numerator law. The initial term is

$$\mathbb{E}_{\overleftarrow{\mathbb{Q}}^\star}\left[\log \frac{p_1(Y_0)}{\nu(Y_0)}\right] = \int_{\mathcal{D}} p_1(y) \log \frac{p_1(y)}{\nu(y)}\,\mathrm{d}y = \mathsf{KL}(p_1||\nu),$$

because $Y_0 \sim p_1$ under $\overleftarrow{\mathbb{Q}}^\star$. The field e is continuous on the compact domain $\overline{\mathcal{D}} \times [0,1]$, so there is a finite constant M_e with $\|e(y,\bar{t})\|_2 \le M_e$. Hence $\mathbb{E}_{\overleftarrow{\mathbb{Q}}^\star} \int_0^1 \|e(Y_{\bar{t}},\bar{t})\|_2^2\,\mathrm{d}\bar{t} \le M_e^2 < \infty$. The integrand is also predictable. By Le Gall (2016, Eq. (5.13)) and the subsequent moment identities, its Itô integral is therefore a square-integrable martingale started at zero. In particular,

$$\mathbb{E}_{\overleftarrow{\mathbb{Q}}^\star}\left[\int_0^1 e(Y_{\bar{t}},\bar{t})^\top\,\mathrm{d}\overline{W}^\star_{\bar{t}}\right] = 0, \qquad \mathbb{E}_{\overleftarrow{\mathbb{Q}}^\star}\int_0^1 \|e(Y_{\bar{t}},\bar{t})\|_2^2\,\mathrm{d}\bar{t} \le M_e^2 < \infty.$$

Taking $\overleftarrow{\mathbb{Q}}^\star$-expectations in (69) therefore gives

$$\mathsf{KL}(\overleftarrow{\mathbb{Q}}^\star||\overleftarrow{\mathbb{Q}}_\theta) = \mathsf{KL}(p_1||\nu) + \frac{1}{2}\mathbb{E}_{\overleftarrow{\mathbb{Q}}^\star}\int_0^1 \|e(Y_{\bar{t}},\bar{t})\|_2^2\,\mathrm{d}\bar{t}.$$

Finally, $Y_{\bar{t}} \sim p_{1-\bar{t}}$ under the ideal law. Tonelli's theorem allows us to move the expectation inside the nonnegative energy integral, and the substitution $t = 1 - \bar{t}$ returns to forward time:

$$\begin{aligned}\mathbb{E}_{\overleftarrow{\mathbb{Q}}^\star}\int_0^1 \|e(Y_{\bar{t}},\bar{t})\|_2^2\,\mathrm{d}\bar{t} &= \int_0^1 g(1-\bar{t})^2\mathbb{E}_{\overleftarrow{\mathbb{Q}}^\star}\left[\|\nabla \log p_{1-\bar{t}}(Y_{\bar{t}}) - s_\theta(Y_{\bar{t}},1-\bar{t})\|_2^2\right]\,\mathrm{d}\bar{t}\\ &= \int_0^1 g(t)^2\mathbb{E}_{X\sim p_t}\left[\|\nabla \log p_t(X) - s_\theta(X,t)\|_2^2\right]\,\mathrm{d}t.\end{aligned}$$

Substituting this identity proves (66). □

Remark E.3.3. Training uses $\mathsf{KL}(\overleftarrow{\mathbb{Q}}_\theta||\overleftarrow{\mathbb{Q}}^\star)$ because its expectation is under the learned law, from which we can generate trajectories. The other direction, $\mathsf{KL}(\overleftarrow{\mathbb{Q}}^\star||\overleftarrow{\mathbb{Q}}_\theta)$, is useful for describing approximation error. Equation (66) separates the mismatch between p_1 and ν from the error in the learned score. That score error is averaged under the ideal marginals p_t, which are fixed as θ changes. This gives an interpretable sufficient condition for accurate estimation, but not a directly usable training loss: the true scores and ideal-law expectations are generally unavailable.

The two KLs compare the same continuous path laws on $[0,1]$, but neither direction generally bounds the other. Pinsker's inequality gives

$$2\|\overleftarrow{\mathbb{Q}}^\star - \overleftarrow{\mathbb{Q}}_\theta\|_{\mathrm{TV}} \le \sqrt{2\min\{\mathsf{KL}(\overleftarrow{\mathbb{Q}}^\star||\overleftarrow{\mathbb{Q}}_\theta), \mathsf{KL}(\overleftarrow{\mathbb{Q}}_\theta||\overleftarrow{\mathbb{Q}}^\star)\}}. \tag{70}$$

Combining (70) with (61) yields the KL accuracy bounds in either direction. In particular, for $\varepsilon > 0$ and $\delta \in (0,1)$, either divergence being at most $\varepsilon^2\delta^2/2$ suffices for the confidence guarantee (62).

Remark E.3.4. It can be noted that there is no dependence on N in bound (61). Dependence on N can be recovered by using the variance. For independent sampled paths, when $\chi^2(\overleftarrow{\mathbb{Q}}^\star\|\overleftarrow{\mathbb{Q}}_\theta) < \infty$, the weight identity (63) and independence give $\mathrm{Var}_{\overleftarrow{\mathbb{Q}}_\theta^{\otimes N}}(\widehat{Z}/Z) = \chi^2(\overleftarrow{\mathbb{Q}}^\star\|\overleftarrow{\mathbb{Q}}_\theta)/N$. Chebyshev's inequality then yields, for every $\varepsilon > 0$,

$$\mathbb{P}_{\overleftarrow{\mathbb{Q}}_\theta^{\otimes N}}\left(\left|\frac{\widehat{Z}}{Z} - 1\right| \ge \varepsilon\right) \le \frac{\chi^2(\overleftarrow{\mathbb{Q}}^\star\|\overleftarrow{\mathbb{Q}}_\theta)}{N\varepsilon^2}. \tag{71}$$

This gives an explicit dependence on N, but requires control of the χ^2 divergence, which small KL alone does not provide.

E.4 Estimating gradients of the objective via REINFORCE

Throughout this subsection, $\nabla_\theta f(y;\theta)$ denotes the gradient with respect to the explicit parameter dependence of f, with the state argument y held fixed. Dependence of sampled states on θ is accounted for by the score-function terms. We use detached trajectories when constructing an autodifferentiable surrogate loss.

Lemma E.4.1. *Let $X' \mid X = x \sim p(\cdot \mid x;\theta)$, where $p(x' \mid x;\theta)$ is differentiable with respect to θ, and let $Y = T(X')$, where T is a deterministic reflection map. Under regularity conditions, we have*

$$\nabla_\theta \log q(y \mid x;\theta) = \mathbb{E}\left[\nabla_\theta \log p(X' \mid x;\theta) \mid Y = y,\ X = x\right]. \tag{72}$$

Consequently, for any integrable function $h : \mathcal{D} \to \mathbb{R}$,

$$\mathbb{E}\left[h(Y)\nabla_\theta \log q(Y \mid X;\theta)\right] = \mathbb{E}\left[h(Y)\nabla_\theta \log p(X' \mid X;\theta)\right]. \tag{73}$$

Proof. Since $Y = T(X')$ is a deterministic map of X', its conditional density can be represented as

$$q(y \mid x;\theta) = \int \delta(T(x') - y)\, p(x' \mid x;\theta)\, \mathrm{d}x', \tag{74}$$

where δ denotes the Dirac delta. Therefore, differentiating (74) with respect to θ gives

$$\begin{aligned}\nabla_\theta q(y \mid x;\theta) &= \int \delta(T(x') - y)\, \nabla_\theta p(x' \mid x;\theta)\, \mathrm{d}x' \\ &= \int \delta(T(x') - y)\, p(x' \mid x;\theta) \nabla_\theta \log p(x' \mid x;\theta)\, \mathrm{d}x'. \end{aligned} \tag{75}$$

Dividing by $q(y \mid x;\theta)$, we obtain

$$\begin{aligned}\nabla_\theta \log q(y \mid x;\theta) &= \frac{\int \delta(T(x') - y) p(x' \mid x;\theta) \nabla_\theta \log p(x' \mid x;\theta)\, \mathrm{d}x'}{q(y \mid x;\theta)} \\ &= \mathbb{E}\left[\nabla_\theta \log p(X' \mid x;\theta) \mid Y = y,\ X = x\right], \end{aligned} \tag{76}$$

where the last equality follows from the Bayes formula for $p(x' \mid y, x;\theta)$. Then for any integrable function $h : \mathcal{D} \to \mathbb{R}$, we have

$$\begin{aligned}&\mathbb{E}\left[h(Y)\nabla_\theta \log p(X' \mid X;\theta)\right] = \mathbb{E}\left[\mathbb{E}\left[h(Y)\nabla_\theta \log p(X' \mid X;\theta) \mid Y, X\right]\right] \\ &= \mathbb{E}\left[h(Y)\, \mathbb{E}\left[\nabla_\theta \log p(X' \mid X;\theta) \mid Y, X\right]\right] = \mathbb{E}\left[h(Y)\nabla_\theta \log q(Y \mid X;\theta)\right]. \end{aligned} \tag{77}$$

□

Theorem E.4.2. *Suppose Assumptions B.2.1 and B.2.3 hold. The gradient of* (22) *satisfies*

$$\nabla_\theta \mathcal{L}_{RP}(\theta) = \mathbb{E}_{\widetilde{\tau}\sim\widetilde{q}(\cdot;\theta)}\left[\sum_{k=0}^{K-1} \nabla_\theta \phi_k(Y_k;\theta) + \sum_{k=0}^{K-1} c_k\, \nabla_\theta \log \widetilde{q}_k(\widetilde{Y}_{k+1} \mid Y_k;\theta)\right], \tag{78}$$

where

$$c_k := \sum_{j=k+1}^{K-1} \phi_j(Y_j;\theta) + \psi(Y_K). \tag{79}$$

Proof. Assumption B.2.3 makes $\psi = \log(\nu/r)$ bounded and Borel measurable. For brevity, define $\mu(\tau;\theta) := \sum_{j=0}^{K-1} \phi_j(Y_j;\theta) + \psi(Y_K)$. The density of the reflected trajectory factorizes as

$$q(\tau;\theta) = \nu(Y_0) \prod_{k=0}^{K-1} q_k(Y_{k+1} \mid Y_k;\theta). \tag{80}$$

Since ν does not depend on θ, differentiating $\mathcal{L}_{\mathrm{RP}}(\theta) = \mathbb{E}_{\tau\sim q(\cdot;\theta)}[\mu(\tau;\theta)]$ and applying the log-derivative identity gives

$$\nabla_\theta \mathcal{L}_{\mathrm{RP}}(\theta) = \mathbb{E}_{\tau\sim q(\cdot;\theta)}\left[\sum_{j=0}^{K-1} \nabla_\theta \phi_j(Y_j;\theta) + \sum_{k=0}^{K-1} \mu(\tau;\theta)\nabla_\theta \log q_k(Y_{k+1} \mid Y_k;\theta)\right]. \tag{81}$$

For a fixed k, define

$$\mu(\tau;\theta) = \sum_{j=0}^{k} \phi_j(Y_j;\theta) + c_k, \tag{82}$$

where $c_k = \sum_{j=k+1}^{K-1} \phi_j(Y_j;\theta) + \psi(Y_K)$. Let $\mathcal{F}_k = \sigma(Y_0,\dots,Y_k)$. The sum $\sum_{j=0}^{k} \phi_j(Y_j;\theta)$ is $\mathcal{F}_k$-measurable, while the transition score has zero conditional mean:

$$\begin{aligned}\mathbb{E}\left[\nabla_\theta \log q_k(Y_{k+1} \mid Y_k;\theta) \middle| \mathcal{F}_k\right] &= \int q_k(y \mid Y_k;\theta)\nabla_\theta \log q_k(y \mid Y_k;\theta)\,\mathrm{d}y \\ &= \nabla_\theta \int q_k(y \mid Y_k;\theta)\,\mathrm{d}y = 0. \end{aligned} \tag{83}$$

Consequently, by using a tower property of conditional expectation

$$\mathbb{E}\left[\mu(\tau;\theta)\nabla_\theta \log q_k(Y_{k+1} \mid Y_k;\theta)\right] = \mathbb{E}\left[c_k \nabla_\theta \log q_k(Y_{k+1} \mid Y_k;\theta)\right]. \tag{84}$$

It remains to replace the reflected transition score by the score of the tractable pre-reflection Gaussian transition. According to (20), the pre-reflection state is sampled as $\widetilde{Y}_{k+1} \sim \widetilde{q}_k(\cdot \mid Y_k;\theta)$, after which the reflected state is obtained by $Y_{k+1} = \mathrm{refl}_{\mathcal{D}}(\widetilde{Y}_{k+1})$. Define $h_k(Y_{k+1}) := \mathbb{E}[c_k \mid Y_{k+1}]$. By the Markov structure of the discretized process, once Y_{k+1} is known, the future reflected trajectory $(Y_{k+2},\dots,Y_K)$ is conditionally independent of the particular preimage $\widetilde{Y}_{k+1}$. Hence, $\mathbb{E}[c_k \mid Y_k, \widetilde{Y}_{k+1}] = h_k(Y_{k+1})$. Using the tower property,

$$\mathbb{E}_{\widetilde{\tau}\sim\widetilde{q}(\cdot;\theta)}\left[c_k \nabla_\theta \log \widetilde{q}_k(\widetilde{Y}_{k+1} \mid Y_k;\theta)\right] = \mathbb{E}_{\widetilde{\tau}\sim\widetilde{q}(\cdot;\theta)}\left[h_k(Y_{k+1})\nabla_\theta \log \widetilde{q}_k(\widetilde{Y}_{k+1} \mid Y_k;\theta)\right]. \tag{85}$$

Applying Lemma E.4.1 to h_k with $X = Y_k$, $X' = \widetilde{Y}_{k+1}$ and $Y = Y_{k+1}$ yields

$$\begin{aligned}&\mathbb{E}_{\widetilde{\tau}\sim\widetilde{q}(\cdot;\theta)}\left[h_k(Y_{k+1})\nabla_\theta \log \widetilde{q}_k(\widetilde{Y}_{k+1} \mid Y_k;\theta)\right] \\ &= \mathbb{E}_{\tau\sim q(\cdot;\theta)}\left[h_k(Y_{k+1})\nabla_\theta \log q_k(Y_{k+1} \mid Y_k;\theta)\right] \\ &= \mathbb{E}_{\tau\sim q(\cdot;\theta)}\left[c_k \nabla_\theta \log q_k(Y_{k+1} \mid Y_k;\theta)\right]. \end{aligned} \tag{86}$$

Substituting (84) and (86) into (81), and observing that $\widetilde{q}$ pushes forward to q after discarding the pre-reflection states, we obtain

$$\nabla_\theta \mathcal{L}_{\mathrm{RP}}(\theta) = \mathbb{E}_{\widetilde{\tau}\sim\widetilde{q}(\cdot;\theta)}\left[\sum_{k=0}^{K-1} \nabla_\theta \phi_k(Y_k;\theta) + \sum_{k=0}^{K-1} c_k \nabla_\theta \log \widetilde{q}_k(\widetilde{Y}_{k+1} \mid Y_k;\theta)\right],$$

which proves the claim. □

E.5 Constructing unbiased estimate after discretization

Proposition E.5.1. *Fix a discretization Δ as in Section 3.4. Let $(Y_k)_{k=0}^K$ be the Markov chain obtained by discretizing the reference dynamics associated with $\overleftarrow{\mathbb{P}}$, with $Y_0 \sim \nu$ and*

$$\widetilde{Y}_{k+1} = Y_k + g_k \xi_k, \qquad Y_{k+1} = \mathrm{refl}_{\mathcal{D}}(\widetilde{Y}_{k+1}), \qquad \xi_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d).$$

Assume that each transition of this reference chain leaves ν invariant. Then, for any fixed θ, the estimator (26)*, computed from trajectories generated by* (20)*, is unbiased:*

$$\mathbb{E}[\widehat{Z}] = Z.$$

Proof. Let $\widetilde{Q}_{\theta,\Delta}$ denote the law of the augmented trajectory

$$\widetilde{\tau} = (Y_0, \widetilde{Y}_1, Y_1, \dots, \widetilde{Y}_K, Y_K)$$

generated by (20), and let $\widetilde{P}_\Delta$ denote the corresponding reference law obtained by setting $s_\theta \equiv 0$ in (20). Their pre-reflection transition densities are

$$\widetilde{q}_k(\widetilde{y} \mid y;\theta) = \mathcal{N}\Big(\widetilde{y}; y + g_k^2 s_\theta(y, 1-t_k), g_k^2 I_d\Big), \qquad \widetilde{p}_k(\widetilde{y} \mid y) = \mathcal{N}(\widetilde{y}; y, g_k^2 I_d).$$

Since the two laws have the same initial distribution and use the same deterministic reflection map,

$$\frac{\mathrm{d}\widetilde{P}_\Delta}{\mathrm{d}\widetilde{Q}_{\theta,\Delta}}(\widetilde{\tau}) = \prod_{k=0}^{K-1} \frac{\widetilde{p}_k(\widetilde{Y}_{k+1} \mid Y_k)}{\widetilde{q}_k(\widetilde{Y}_{k+1} \mid Y_k;\theta)}.$$

Using the Gaussian representation in (20) gives

$$\log \frac{\widetilde{p}_k(\widetilde{Y}_{k+1} \mid Y_k)}{\widetilde{q}_k(\widetilde{Y}_{k+1} \mid Y_k; \theta)} = -g_k s_\theta(Y_k, 1-t_k)^\top \xi_k - \frac{1}{2} g_k^2 \|s_\theta(Y_k, 1-t_k)\|_2^2.$$

Thus the exponential factor in (26) is exactly $\mathrm{d}\widetilde{P}_\Delta / \mathrm{d}\widetilde{Q}_{\theta,\Delta}$. By assumption, reflected transition for the reference process preserves ν. Since $Y_0 \sim \nu$, it follows inductively that $Y_K \sim \nu$ under P_Δ. Therefore,

$$\mathbb{E}_{P_\Delta}\left[\frac{r(Y_K)}{\nu(Y_K)}\right] = \int_{\mathcal{D}} \frac{r(y)}{\nu(y)} \nu(y)\,\mathrm{d}y = \int_{\mathcal{D}} r(y)\,\mathrm{d}y = Z.$$

For one trajectory, the importance weight in (26) therefore satisfies

$$\mathbb{E}_{\widetilde{Q}_{\theta,\Delta}}\left[\frac{r(Y_K)}{\nu(Y_K)} \frac{\mathrm{d}\widetilde{P}_\Delta}{\mathrm{d}\widetilde{Q}_{\theta,\Delta}}\right] = \mathbb{E}_{\widetilde{P}_\Delta}\left[\frac{r(Y_K)}{\nu(Y_K)}\right] = \int_{\mathcal{D}} \frac{r(y)}{\nu(y)} \nu(y)\,\mathrm{d}y = Z.$$

Averaging N trajectories preserves the expectation, which proves $\mathbb{E}[\widehat{Z}] = Z$. □

Remark E.5.2. The invariance assumption in Proposition E.5.1 holds for the coordinate-wise folding reflection on $\overline{\mathcal{D}} = [0,1]^d$ used in our experiments. It is enough to consider one coordinate. Let φ_g denote the density of $\mathcal{N}(0, g^2)$. For $x, y \in [0,1]$, the transition density of $Y = \mathrm{refl}_{[0,1]}(x + g\xi)$, where $\xi \sim \mathcal{N}(0,1)$, is

$$K_g(y \mid x) = \sum_{m \in \mathbb{Z}} \left[\varphi_g(y - x + 2m) + \varphi_g(y + x + 2m)\right].$$

Since $\varphi_g(z) = \varphi_g(-z)$, reindexing $m \mapsto -m$ shows that

$$K_g(y \mid x) = K_g(x \mid y).$$

As K_g is a Markov kernel,

$$\int_0^1 K_g(x \mid y)\,\mathrm{d}x = 1,$$

and therefore, by symmetry,

$$\int_0^1 K_g(y \mid x)\,\mathrm{d}x = 1.$$

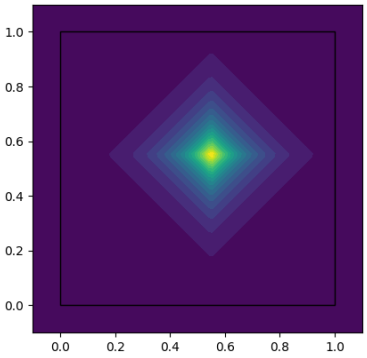
Hence the uniform distribution on $[0,1]$ is invariant under every reference transition. Since both the Gaussian noise and the reflection act coordinate-wise, the d-dimensional transition kernel factorizes across coordinates, and $\nu = \mathrm{Unif}([0,1]^d)$ is invariant as well.

F Experimental details

F.1 Benchmarks

Genz integrals. The Genz family provides a standard collection of numerical integration test functions, used in literature (Ott et al., 2023; Warren & Ramos, 2024), designed to exhibit distinct forms of difficulty. We consider five out of six Genz integrals - we exclude the periodic integrand as it takes negative values, which does not fit our study. All Genz targets used here admit closed-form normalization constants Kaarnioja (2025), which allows to produce an exact reference value for any integration result. The five Genz targets span substantially different forms of concentration. Continuous and Gaussian Peak provide smooth interior modes, Product Peak has a factorized rational structure, Corner Peak concentrates mass at a boundary corner and Discontinuous combines a hard support constraint with an exponential trend within the support. The dimensional sweep from 5 to 30 further tests whether the sampler can preserve accurate mass allocation as the number of coordinates increases.

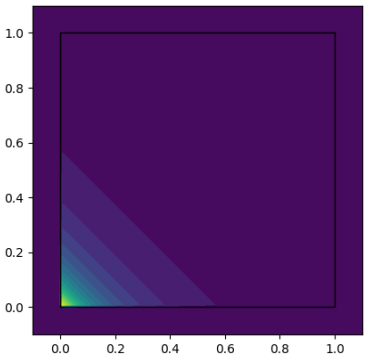
For the Genz experiments, several dimensions are considered, which are specified in the description of the corresponding problem. The concentration parameters are varied with dimension in order to prevent the benchmark from becoming either nearly uniform or excessively degenerate as d increases. Unless otherwise stated, scalar parameters are shared across dimensions, so that the resulting target is symmetric with respect to coordinate permutations.



Continuous Genz. The Continuous Genz target is defined by

$$r(x) = \exp\left(-\sum_{k=1}^{d} a_k |x_k - u_k|\right), \qquad x \in [0,1]^d.$$

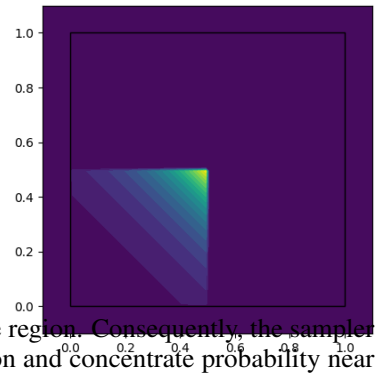
The parameter u_k specifies the location of the peak along dimension k, while a_k controls its concentration. For all coordinates we take $u_k = 0.55$ and the scalar concentration parameter is chosen as $a_k = a \in \{8.0, 5.0, 4.0, 3.0\}$ for $d \in \{5, 10, 20, 30\}$, respectively. Geometrically, the target has a sharp non-smooth peak at $u = (u_k)_{k=1}^d$ and decays exponentially in the ℓ_1 distance from it, producing diamond-shaped level sets with axis-aligned corners. As the dimension grows, its mass becomes increasingly concentrated in a small region of the domain, making accurate exploration and integration progressively more challenging.



Corner Peak Genz. The Corner Peak target is given by

$$r(x) = \left(1 + \sum_{k=1}^{d} a_k x_k\right)^{-(d+1)}, \qquad x \in [0,1]^d.$$

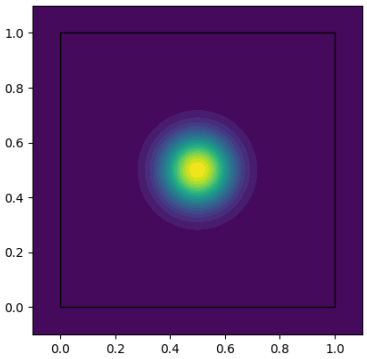
The density is largest at the corner $x = 0$ and decays as the weighted sum of the coordinates increases. For all coordinates we take $a_k = a \in \{0.50, 0.25, 0.10, 0.08\}$ for $d \in \{5, 10, 20, 30\}$, respectively. Because the dominant probability mass is associated with a boundary corner rather than an interior mode, this benchmark probes the ability of the sampler to discover and represent boundary-concentrated distributions. The progressively smaller a_k values compensate for the increasing dimensionality and preserve a meaningful level of concentration across the dimensional sweep.



Discontinuous Genz. The Discontinuous Genz is defined by

$$r(x) = \begin{cases} \exp\left(\sum_{k=1}^{d} a_k x_k\right), & x_k \le u_k \ \forall k, \\ 0, & \text{otherwise.} \end{cases}$$

We set $u_k = 0.5$ and use $a_k = a \in \{5.0, 4.0, 3.0, 3.0\}$ for $d \in \{5, 10, 20, 30\}$. respectively. This benchmark combines two distinct difficulties. First, the support is restricted to the subcube $[0, 0.5]^d$, which introduces a sharp discontinuity at every boundary of the feasible region. Second, within the support the density increases exponentially with the coordinates, so that the mode is located at the upper corner $(0.5, \ldots, 0.5)^\top$ of the active region. Consequently, the sampler must simultaneously identify a discontinuous support restriction and concentrate probability near the boundary of that restricted support. In the implementation, we add a small positive correction of $\varepsilon = 10^{-10}$ to make a target reward strictly positive and then subtract $\varepsilon \, \mathrm{vol}(\mathcal{D})$ in the final estimate.

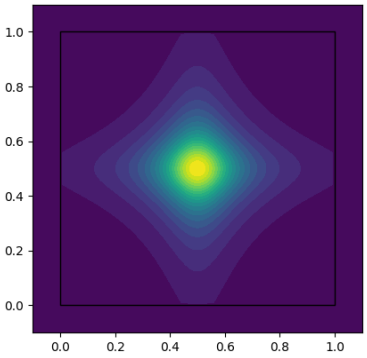


Gaussian Peak Genz. The Gaussian Peak Genz is defined by

$$r(x) = \exp\left(-\sum_{k=1}^{d} a_k^2 (x_k - u_k)^2\right), \qquad x \in [0,1]^d.$$

We use $u_k = 0.5$ for all coordinates and set $a_k = a \in \{8.0, 5.0, 4.0, 3.0\}$ for $d \in \{5, 10, 20, 30\}$, respectively.

The resulting distribution is a sharply concentrated, smooth peak centred at the middle of the hypercube. Relative to the Continuous target, the Gaussian Peak is globally differentiable and has rapidly decaying tails away from its mode, making the pair useful for distinguishing the effect of nonsmoothness from that of concentration alone.

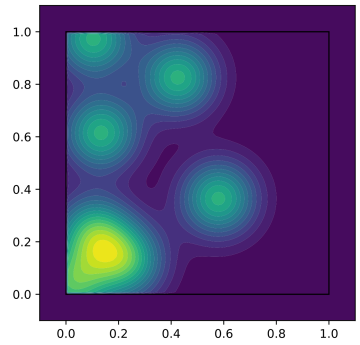


Product Peak Genz. The Product Peak Genz target is given by

$$r(x) = \prod_{k=1}^{d} \frac{1}{a_k^{-2} + (x_k - u_k)^2}, \qquad x \in [0,1]^d.$$

We take $u_k = 0.5$ and $a_k = a \in \{8.0, 5.0, 4.0, 4.0\}$ for $d \in \{5, 10, 20, 30\}$, respectively.

The function factorizes across dimensions, but each factor exhibits a sharply localized peak around $u_k = 0.5$. In the implementation, we rescale $r(x)$ by the maximum reward for numerical stability and adjust the analytic normalizing constant accordingly.

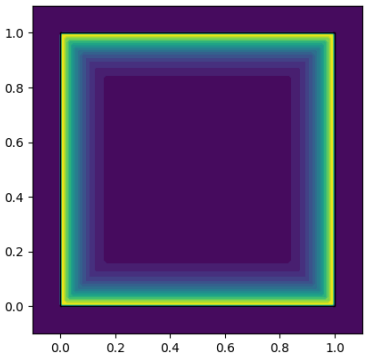


GMM Cube. The GMM Cube target is a bounded multimodal distribution constructed by placing a mixture of five isotropic Gaussian components inside the unit hypercube. For a d-dimensional target

$$r(x) = \frac{1}{5} \sum_{j=1}^{5} \mathcal{N}(x; \mu_j, \sigma^2 I_d), \qquad x \in [0,1]^d,$$

with $\sigma = 0.1$. The component locations are sampled once from the interior region $[0.01, 0.99]^d$, thereby ensuring all modes remain separated from the boundary. The target is subsequently restricted to the unit cube. This construction produces a genuinely multimodal bounded target in which the sampler must discover several well-separated regions of high probability rather than a single dominant mode.

Because the Gaussian mixture is initially defined on the full Euclidean space and then restricted to the cube, its normalization constant is the probability that a draw from the unconstrained mixture falls inside $[0,1]^d$. This quantity is estimated in the implementation by Monte Carlo sampling from the unconstrained mixture. The target therefore tests both multimodal exploration and the ability to account for truncation by the bounded domain.



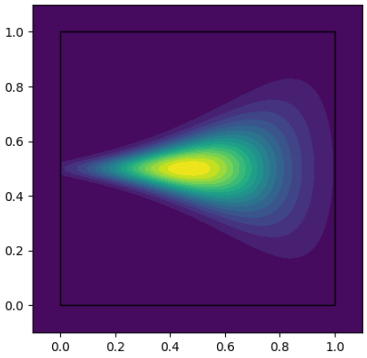
Stripe Cube. The Stripe Cube target is designed to concentrate probability near the boundary of the unit hypercube while retaining a small positive density throughout its interior. For a point $x \in [0,1]^d$, define the distance to the closest boundary face by $\delta(x) = \min_i \{x_i, 1 - x_i\}$. We map this distance to a normalized boundary score $s(x) = 1 - 2\delta(x)$, which equals zero at the centre of the cube and one on the boundary. The target density is then

$$r(x) = \alpha + (\beta - \alpha) s^{\gamma}(x), \qquad x \in [0,1]^d,$$

where $\alpha = 10^{-10}, \beta = 1$ and $\gamma = 8$. Hence the density rises steeply as a point approaches any boundary face, producing a set of high-probability boundary "stripes" around the cube while retaining a numerically nonzero floor in its interior.

Whereas the Corner Peak benchmark concentrates mass near a single corner, Stripe Cube assigns high density collectively to all boundary faces. The resulting target therefore contains a large, geometrically thin high-density region whose total structure becomes increasingly complex with dimension. The normalization constant is available analytically

$$Z = \alpha + (\beta - \alpha) \frac{d}{d + \gamma}.$$



Funnel Cube. The Funnel Cube target is a bounded analogue of the classical funnel distribution and is designed to test adaptation to strongly varying conditional scales. The first coordinate follows a Gaussian density,

$$x_1 \sim \mathcal{N}(0.6, 0.22^2),$$

restricted to the interval $[0, 1]$. Conditional on x_1, the remaining $d - 1$ coordinates are independent Gaussian variables centered at $\tilde{\mu} = 0.5$ with a standard deviation that changes exponentially across the first coordinate:

$$\log \sigma(x_1) = \log \sigma_{\min} + x_1 \left(\log \sigma_{\max} - \log \sigma_{\min}\right),$$

where $\sigma_{\min} = 0.02$ and $\sigma_{\max} = 0.30$. The target density can therefore be written as

$$r(x) = \mathcal{N}(x_1; 0.6, 0.22^2) \prod_{k=2}^{d} \mathcal{N}\Big(x_k; 0.5, \sigma^2(x_1)\Big), \qquad x \in [0, 1]^d.$$

The defining difficulty of this benchmark is the strong dependence of the conditional width on x_1. Samples associated with one region of the first coordinate lie in a very narrow tube around 0.5, whereas samples associated with another region can occupy a considerably wider cross-section. The target therefore tests whether a learned sampler can represent pronounced heteroscedasticity rather than simply locating isolated modes. Its normalization constant is computed by reducing the d-dimensional integral to a one-dimensional numerical integration over x_1, where the probability that every remaining coordinate lies in $[0, 1]$ is evaluated analytically.

Asian option. We consider Asian option pricing as an application-oriented high-dimensional benchmark with a nonsmooth payoff and a strong temporal dependence structure, following the standard construction L'Ecuyer (2009). The implementation is based on a discretised Black–Scholes asset path. A point $x = (x_1, \ldots, x_d) \in [0, 1]^d$ is first transformed into independent standard-normal shocks through the inverse Gaussian cumulative distribution function, $\epsilon_i = \Phi^{-1}(x_i)$. For monitoring times $0 = t_0 < t_1 < \cdots < t_d = T$, the Brownian path is then constructed as

$$W_{t_j} = \sum_{i=1}^{j} \sqrt{t_i - t_{i-1}}\, \epsilon_i,$$

and the corresponding asset price at time t_j is

$$S_{t_j} = S_0 \exp\left[\left(r - \frac{1}{2}\sigma^2\right) t_j + \sigma W_{t_j}\right].$$

The default parameters are $S_0 = 100$, $r = \log(1.09)$, $\sigma = 0.2$, and $T = \frac{120}{365}$. The reward is given by the discounted arithmetic-average call payoff

$$r(x) = e^{-rT} \max(0, \frac{1}{d} \sum_{i=1}^{d} S_{t_i} - K),$$

where K is the strike price. The arithmetic average couples the payoff to the entire simulated trajectory, while the positive-part operator introduces a nonsmooth boundary and makes the reward exactly zero over a substantial portion of the input domain.

We consider two principal configurations from Example 3 (L'Ecuyer, 2009). The easy problem has dimension $d = 10$, uses a starting monitoring period of 111 days, a strike of $K = 90$, and has a ground-truth option value $Z = 13.008$. The hard problem increases the dimension to $d = 120$, starts monitoring at period 1, retains the strike $K = 90$, and has $Z = 11.207$.

The difference between the two settings therefore changes both the number of Gaussian increments used to represent the asset path and the portion of the maturity interval entering the arithmetic average. The hard configuration consequently presents a substantially higher-dimensional integration problem with a longer temporal dependence structure, providing a stringent test of the sampler's scalability.

As for the Genz Discontinuous benchmark, we add a small constant $\varepsilon = 10^{-10}$ to the reward to make it strictly positive. We remove this artificial contribution from the final integral estimate by subtracting $\varepsilon \operatorname{vol}(\mathcal{D})$.

Digits. We adapt the Digits benchmark from Blessing et al. (2024), where the target is a NICE normalizing-flow density (Dinh et al., 2014) trained on downsampled 14×14 MNIST images. The original target is a normalized continuous density p_{NICE} on $\mathbb{R}^{196}$. To obtain a bounded-domain integration problem, we apply the coordinate-wise logit transformation and define the cube target by the corresponding change of variables,

$$r(x) = p_{\mathrm{NICE}}(\mathrm{logit}(x)) \prod_{j=1}^{196} \frac{1}{x_j(1-x_j)}.$$

This transformation preserves the normalizing constant, so $Z = 1$. Samples generated on the cube can be mapped back through the logit transformation and reshaped into 14×14 grayscale images. Beyond integration, this benchmark provides a challenging high-dimensional sampling task with complex learned structure, allowing us to test whether RDS can recover meaningful samples from the target distribution itself.

Elliptic Two-Loop. The Elliptic Two-Loop target is a Feynman-parameter integral arising in collider phenomenology. It corresponds to the planar two-loop master integral $I^{A}_{1,1,0,1,1,1,1,0,0}$ with full heavy-quark mass dependence, which contributes to the next-to-leading-order corrections to Higgs-plus-jet production (Bonciani et al., 2016). This sector is distinguished by the appearance of elliptic functions in its analytic solution.

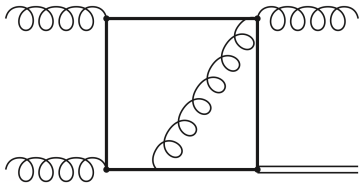


Figure 5: Two-loop six-propagator graph giving rise to elliptic functions (Borowka et al., 2018).

Using the propagator notation of Bonciani et al. (2016), the six active propagators are

$$D_1 = k_1^2 - m^2, \qquad D_2 = (k_1 + p_1 + p_2)^2 - m^2, \qquad D_4 = (k_2 + p_1 + p_2)^2 - m^2,$$
$$D_5 = (k_1 + p_1)^2 - m^2, \qquad D_6 = (k_1 - k_2)^2, \qquad D_7 = (k_2 - p_3)^2 - m^2.$$

We evaluate the integral at the Euclidean kinematic point

$$s = (p_1 + p_2)^2 = -\tfrac{4}{3}, \qquad t = (p_1 + p_3)^2 = -\tfrac{16}{5}, \qquad p_4^2 = -\tfrac{100}{39}, \qquad m^2 = 1,$$

used in the numerical benchmark of Borowka et al. (2018).

Feynman parametrization gives an integral over the five-dimensional simplex $\{\alpha_i \geq 0,\ \sum_{i=1}^{6} \alpha_i = 1\}$. To express it on the unit hypercube, we use the stick-breaking transformation

$$\alpha_i = z_i \prod_{j<i}(1 - z_j), \qquad i = 1, \ldots, 5, \qquad \alpha_6 = \prod_{j=1}^{5}(1 - z_j),$$

for $z \in [0,1]^5$, whose Jacobian is

$$J(z) = \prod_{j=1}^{4}(1 - z_j)^{5-j}.$$

In four spacetime dimensions, the corresponding cube target is

$$r(z) = \left(-\frac{s}{m^2}\right)^{3/2} \frac{J(z)}{\mathcal{F}(\alpha(z))^2}, \qquad z \in [0,1]^5,$$

where $\mathcal{F}$ is the second Symanzik polynomial. The Euclidean kinematics make the target real and non-negative.

The benchmark combines a non-factorized Feynman integrand with nontrivial boundary structure, providing an application-oriented test beyond the synthetic targets considered above. Its normalization is known analytically and, in the normalization used here, is

$$Z = 0.247074199140732131068066,$$

in agreement with the numerical and analytic values reported by Borowka et al. (2018).

F.2 Evaluation protocol and training details

Evaluation protocol. We compare methods across different budgets corresponding to the number of points used for the integral estimation. For neural network-based methods, namely PIS, ZüNIS and RDS, for a fixed number of points used for evaluation, the same number of points is used for training. For these methods we use a batch size of 512 in all experiments. We report the mean absolute error $|\widehat{Z} - Z|$ over independent runs as a function of the number of points. For all experiments we present mean and std of the error across 3 random seeds.

Quasi-Monte Carlo. As a non-adaptive baseline, we use randomized quasi-Monte Carlo (QMC) with scrambled Sobol sequences L'Ecuyer (2009). For a budget $N = 2^m$, we generate N points $x^{(1)}, \dots, x^{(N)} \in [0,1]^d$ from a scrambled Sobol sequence and estimate

$$\widehat{Z}_{\mathrm{QMC}} = \frac{1}{N} \sum_{i=1}^{N} r(x^{(i)}). \tag{87}$$

Independent random scramblings are used across runs. Using powers of two preserves the digital-net structure of the Sobol construction. QMC does not involve any optimization, and its target-evaluation budget is exactly N. We use SciPy implementation Virtanen et al. (2020).

Fast Fourier Bayesian Quadrature. We use the public implementation of Fast Fourier Bayesian Quadrature (FFBQ) by Warren & Ramos (2024). FFBQ approximates the integrand with a Gaussian process and computes the resulting Bayesian-quadrature estimate using scalable convolution operators. In our experiments, we fit a low-rank GP with an RBF kernel using 5,000 uniformly sampled training points and use an independent set of 5,000 points for quadrature evaluation. We employ the sparse QMC convolution operator provided by the authors. Low-rank GP was chosen, because of the scalability issues with dimension of FFBQ. The GP training and evaluation sets are fixed across experiments, while we vary the number N of QMC samples used to approximate the final convolution integral.

Path Integral Sampler. We include the Path Integral Sampler (PIS) (Zhang & Chen, 2022), an on-policy diffusion sampler derived from a stochastic-control formulation of sampling. To apply PIS on the bounded domain, we instead use a change of variables between $\mathbb{R}^d$ and $[0,1]^d$. Specifically, for $z \in \mathbb{R}^d$, we set $x = \sigma(z)$, where $\sigma(\cdot)$ is a sigmoid function. Accordingly, PIS is trained on the transformed reward $\widetilde{r}(z) = r(\sigma(z)) \left|\det J_\sigma(z)\right|$, which preserves the original integral by the change-of-variables formula. The control is parameterized by an MLP with four hidden layers of width 256. We train PIS with Adam using a learning rate of 10^{-3}. Each diffusion trajectory is discretized using 64 time steps. The control is trained using the standard path-space objective consisting of the accumulated quadratic control cost and the terminal target-density contribution. The resulting estimate of Z is obtained using the path-integral importance weights associated with the learned controlled process. For PIS, we used a public implementation from Blessing et al. (2024).

ZüNIS. We use the public implementation of ZüNIS (Deutschmann & Götz, 2024). The proposal is parameterized by a bounded piecewise-quadratic coupling flow with 8 coupling layers, 10 bins per scalar transformation, and conditioner networks with two hidden layers of width 256. We optimize the proposal with the reverse-KL objective using Adam with learning rate 10^{-3} and global gradient-norm clipping at 1. The first batch is sampled uniformly from the cube, while all subsequent batches are drawn from the current flow proposal. Each evaluated survey batch is used for a single optimization update. At evaluation time, we draw fresh samples from the learned proposal and estimate the integral using importance sampling,

$$\widehat{Z}_{\mathrm{ZuNIS}} = \frac{1}{N} \sum_{i=1}^{N} \frac{r(x^{(i)})}{q_\theta(x^{(i)})}, \qquad x^{(i)} \sim q_\theta. \tag{88}$$

RDS-RP and RDS-RF. Both variants use the same reflected diffusion and differ only in the gradient estimator used for training. The score s_θ is parameterized by a 4-layer MLP with hidden width 256. We use $\sigma_0 = 10^{-3}$ and $\sigma_1 = 1$, and discretize the diffusion into 64 reflected steps for integration tasks. Both models are trained on-policy with Adam using the same optimization hyperparameters and global gradient-norm clipping at 1.

For RDS-RP, gradients are propagated through the simulated trajectory using the reparameterized Gaussian increments and the coordinate-wise reflection map. Since reflection on the cube is piecewise linear, it is differentiable almost everywhere and can be handled directly by automatic differentiation. RDS-RF instead uses the score-function estimator described in Section 3.4. The sampled states are treated as stop-gradient quantities, while the dependence on the proposal parameters is captured through the scores of the Gaussian transitions and their corresponding cost-to-go terms. Thus, RDS-RP and RDS-RF use the same proposal family, discretization, and training objective, providing a direct comparison between pathwise and score-function optimization.

Our implementations are based upon the published codebase of Blessing et al. (2024).

Digits experiments. For the Digits benchmark, we use the PIS architecture and remaining training hyperparameters following Blessing et al. (2024). Both PIS and RDS are discretized using 128 diffusion steps and trained with Adam using a learning rate of 5×10^{-4}, batch size of 2048 and $40,000$ iterations. For RDS, we set the terminal noise scale to $\sigma_1 = 0.5$.

We evaluate sample quality using the 2-Wasserstein distance and the evidence lower bound (ELBO), following Blessing et al. (2024). Since the NICE target admits direct sampling, we draw an independent set of reference samples from the target and compare them with samples generated by each method. For RDS, samples generated on $[0,1]^{196}$ are first mapped back to the original $\mathbb{R}^{196}$ space using the coordinate-wise logit transformation. We compute the entropy-regularized empirical 2-Wasserstein distance between the generated and reference samples using the squared Euclidean transport cost and regularization $\varepsilon = 10^{-3}$.

The ELBO can be computed directly from the same path-space importance weights used for normalizing-constant estimation. For a trajectory $\widetilde{\tau}$ generated by RDS, let

$$w(\widetilde{\tau}) := \frac{r(Y_K)}{\nu(Y_K)} \exp\left(- \sum_{k=0}^{K-1} \left[g_k s_\theta(Y_k, 1-t_k)^\top \xi_k + \frac{1}{2} g_k^2 \| s_\theta(Y_k, 1-t_k) \|_2^2 \right] \right),$$

so that $\mathbb{E}[w(\widetilde{\tau})] = Z$ by Proposition E.5.1. The corresponding path-space ELBO is

$$\mathrm{ELBO} := \mathbb{E}[\log w(\widetilde{\tau})] \le \log \mathbb{E}[w(\widetilde{\tau})] = \log Z,$$

where the inequality follows from Jensen's inequality. We estimate it using

$$\widehat{\mathrm{ELBO}} = \frac{1}{N} \sum_{i=1}^{N} \log w(\widetilde{\tau}^{(i)}).$$

For PIS, we analogously use its path-space importance weights. Higher ELBO and lower Wasserstein distance indicate better agreement with the target distribution.

F.3 Full experimental results

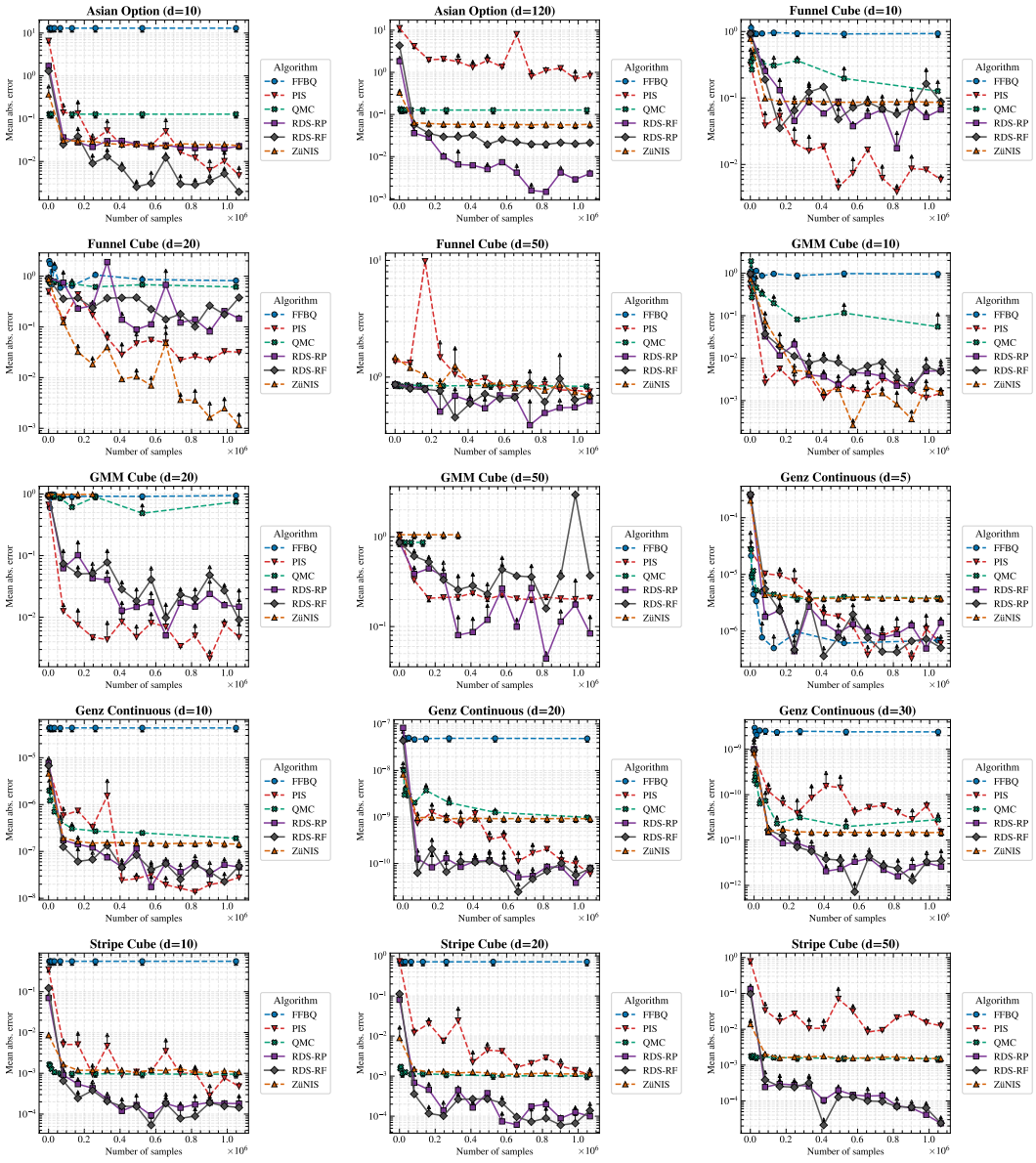


Figure 6: Mean absolute error as a function of the number of samples used for estimation across the considered integration benchmarks and dimensionalities.

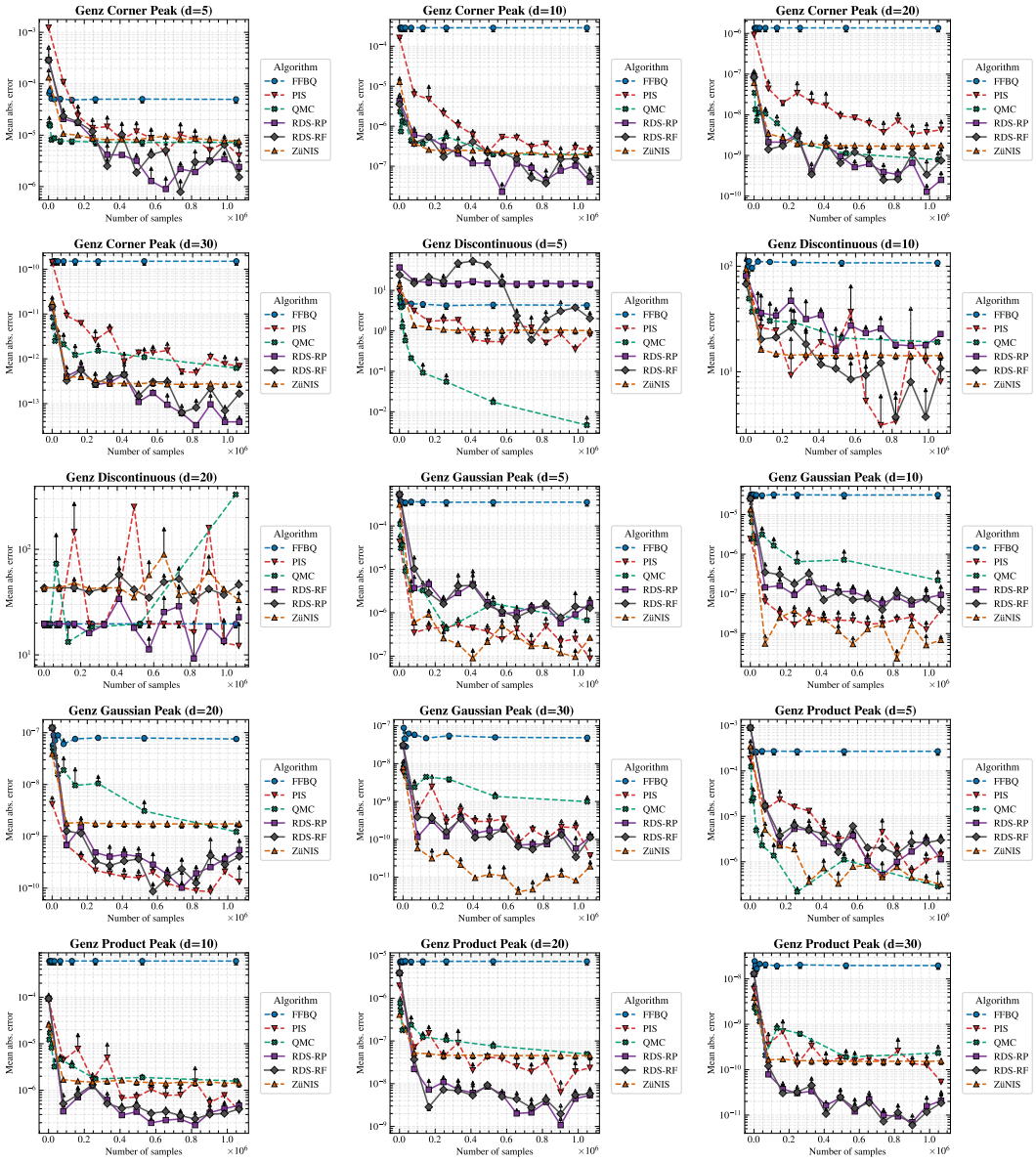
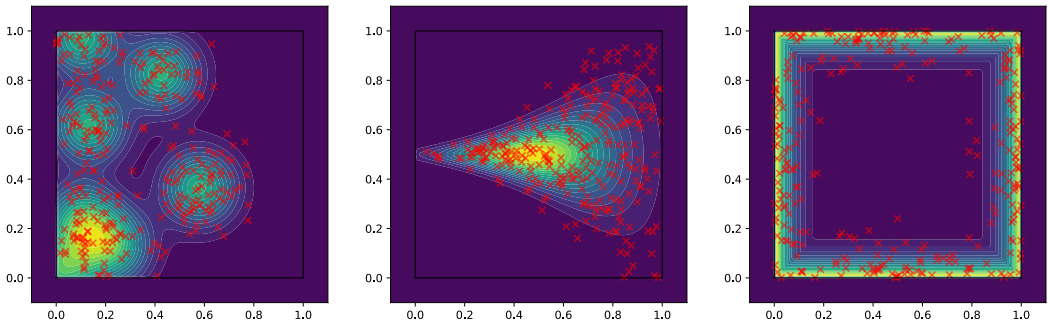


Figure 7: Mean absolute error as a function of the number of samples used for estimation across the considered integration benchmarks and dimensionalities (continued).

Table 1: Mean absolute error of the normalizing-constant estimate using $N = 2^{20}$ samples. Values are mean ± standard deviation. Best result in each row is shown in bold.

Target	d	FFBQ	PIS	QMC	RDS-RF	RDS-RP	ZüNIS
Genz Continuous	5	6.84e−7 ± 1.87e−7	**2.31e−7 ± 1.53e−7**	3.80e−6 ± 9.57e−8	5.66e−7 ± 4.46e−7	3.99e−7 ± 3.00e−7	3.73e−6 ± 5.46e−8
	10	4.32e−5 ± 3.58e−8	2.12e−8 ± 1.15e−8	1.90e−7 ± 4.35e−8	3.78e−8 ± 9.14e−9	**1.05e−8 ± 8.67e−9**	1.44e−7 ± 1.72e−9
	20	4.86e−8 ± 7.84e−10	3.89e−11 ± 2.60e−11	9.81e−10 ± 9.46e−11	3.86e−11 ± 4.78e−11	**2.28e−11 ± 2.57e−11**	9.14e−10 ± 2.56e−12
	30	2.42e−9 ± 2.00e−11	1.82e−11 ± 5.50e−12	2.81e−11 ± 1.62e−11	1.03e−12 ± 6.53e−13	**9.71e−13 ± 3.39e−13**	1.46e−11 ± 2.36e−13
Genz Gaussian Peak	5	3.52e−4 ± 4.76e−7	1.56e−7 ± 1.37e−7	6.70e−7 ± 9.57e−7	1.04e−6 ± 1.66e−6	1.11e−6 ± 1.59e−6	**4.21e−8 ± 1.51e−8**
	10	3.12e−5 ± 8.43e−8	9.71e−9 ± 3.57e−9	2.20e−7 ± 2.32e−7	2.35e−8 ± 1.50e−8	1.59e−8 ± 1.15e−8	**3.25e−9 ± 4.64e−10**
	20	7.55e−8 ± 6.82e−9	**4.05e−11 ± 2.96e−11**	1.22e−9 ± 7.95e−10	1.20e−10 ± 1.53e−10	7.44e−11 ± 1.08e−10	1.73e−9 ± 1.17e−11
	30	4.84e−8 ± 1.02e−9	4.48e−10 ± 3.23e−10	9.95e−10 ± 6.31e−10	6.25e−11 ± 2.09e−11	3.14e−11 ± 1.96e−11	**6.76e−12 ± 7.24e−12**
Genz Discontinuous	5	4.29e0 ± 3.19e−2	3.97e−1 ± 1.09e−1	**4.68e−3 ± 4.32e−3**	5.32e−1 ± 2.02e−1	1.46e1 ± 4.71e−1	1.01e0 ± 1.49e−2
	10	1.08e2 ± 1.26e0	7.34e0 ± 9.01e0	1.91e1 ± 2.07e0	**3.99e0 ± 4.77e0**	1.72e1 ± 3.93e0	1.41e1 ± 5.23e−2
	20	1.97e1 ± 2.34e−6	2.27e1 ± 7.47e0	3.31e2 ± 5.27e2	**9.87e0 ± 4.00e0**	3.34e1 ± 4.00e0	4.02e1 ± 1.24e0
Genz Corner Peak	5	4.98e−5 ± 8.43e−8	5.18e−6 ± 2.19e−6	7.25e−6 ± 1.89e−8	4.10e−6 ± 3.45e−6	**2.00e−6 ± 1.10e−6**	7.77e−6 ± 2.41e−7
	10	2.89e−4 ± 2.78e−7	2.66e−7 ± 4.40e−8	1.98e−7 ± 8.42e−9	1.38e−7 ± 6.37e−8	**4.76e−8 ± 6.61e−8**	1.99e−7 ± 1.38e−8
	20	1.37e−6 ± 2.87e−9	2.90e−9 ± 2.36e−9	7.85e−10 ± 6.42e−10	7.07e−10 ± 3.33e−10	**5.00e−10 ± 1.18e−10**	1.76e−9 ± 5.48e−11
	30	1.50e−10 ± 1.61e−13	1.03e−12 ± 3.86e−13	6.20e−13 ± 6.15e−13	1.27e−13 ± 4.52e−14	**1.10e−14 ± 5.46e−15**	2.57e−13 ± 2.60e−15
Genz Product Peak	5	2.70e−4 ± 4.66e−7	10.00e−7 ± 3.09e−7	2.85e−7 ± 4.24e−7	4.23e−6 ± 8.18e−7	1.06e−6 ± 5.65e−7	**1.81e−7 ± 4.19e−8**
	10	5.99e−4 ± 3.81e−7	4.10e−7 ± 9.22e−8	1.58e−6 ± 1.95e−7	4.05e−7 ± 1.49e−7	**3.98e−7 ± 1.16e−7**	1.40e−6 ± 1.64e−8
	20	7.29e−6 ± 4.77e−8	1.47e−8 ± 5.30e−9	4.96e−8 ± 5.98e−9	1.27e−9 ± 7.77e−10	**9.17e−10 ± 7.95e−10**	4.50e−8 ± 3.74e−10
	30	1.97e−8 ± 1.83e−10	2.76e−10 ± 2.09e−10	2.33e−10 ± 1.55e−10	1.00e−11 ± 4.85e−12	**8.56e−12 ± 2.27e−12**	1.52e−10 ± 1.51e−12
GMM Cube	10	9.59e−1 ± 1.81e−2	1.44e−3 ± 1.29e−3	5.49e−2 ± 7.75e−2	1.85e−3 ± 1.43e−3	4.81e−3 ± 1.47e−3	**3.02e−4 ± 1.68e−4**
	20	9.44e−1 ± 3.27e−3	**2.24e−3 ± 1.78e−3**	7.45e−1 ± 2.32e−1	4.17e−2 ± 4.77e−2	1.18e−2 ± 5.76e−3	9.84e−1 ± 0.00e0
	50	-	1.98e−1 ± 6.14e−3	8.65e−1 ± 7.30e−8	3.00e−1 ± 8.00e−2	**1.68e−2 ± 2.11e−2**	1.05e0 ± 7.30e−8
Stripe Cube	10	5.56e−1 ± 1.67e−4	7.02e−4 ± 1.86e−4	9.50e−4 ± 8.50e−6	1.08e−4 ± 1.99e−5	**6.80e−5 ± 6.73e−5**	1.01e−3 ± 5.63e−5
	20	7.14e−1 ± 5.51e−5	2.09e−3 ± 1.06e−3	1.01e−3 ± 1.29e−5	**1.44e−4 ± 1.17e−4**	1.70e−4 ± 7.08e−5	1.07e−3 ± 6.51e−5
	50	-	1.17e−2 ± 4.98e−3	1.51e−3 ± 3.67e−5	**6.12e−5 ± 5.64e−5**	8.10e−5 ± 8.71e−5	1.53e−3 ± 1.32e−5
Funnel Cube	10	9.33e−1 ± 4.96e−3	**3.72e−3 ± 2.28e−3**	1.27e−1 ± 7.26e−2	4.76e−2 ± 4.57e−3	1.16e−1 ± 8.30e−2	8.55e−2 ± 1.92e−4
	20	8.15e−1 ± 1.54e−1	1.79e−2 ± 8.68e−3	6.13e−1 ± 1.53e−1	1.99e−1 ± 9.58e−2	8.78e−2 ± 4.73e−2	**9.32e−4 ± 6.00e−4**
	50	-	7.41e−1 ± 5.28e−2	8.34e−1 ± 2.37e−2	**6.02e−1 ± 8.33e−2**	6.50e−1 ± 2.45e−2	6.90e−1 ± 1.26e−2
Asian Option	10	1.30e1 ± 5.49e−3	1.25e−2 ± 4.73e−3	1.27e−1 ± 2.43e−5	**1.28e−3 ± 3.81e−4**	2.15e−2 ± 3.60e−3	2.44e−2 ± 1.29e−3
	120	-	7.04e−1 ± 5.81e−2	1.28e−1 ± 1.93e−4	**1.06e−3 ± 1.47e−3**	2.15e−2 ± 1.48e−3	5.66e−2 ± 3.69e−4
Elliptic Two-Loop	5	9.74e−3 ± 5.71e−5	2.51e−3 ± 2.58e−4	3.12e−3 ± 1.35e−3	**3.05e−4 ± 2.12e−4**	7.32e−4 ± 9.99e−5	1.02e−3 ± 1.03e−5



GMM Cube Funnel Cube Stripe Cube

Figure 8: Samples generated by RDS-RP model.