A Trust Region Approach for Learning Schrödinger Bridges

Takeshi Koshizuka*
RIKEN AIP
Kyushu University
takeshi.koshizuka@riken.jp

Denis Blessing
Karlsruhe Institute of Technology
denis.blessing@kit.edu

Max Zimmer, Sebastian Pokutta
Zuse Institute Berlin
Technische Universität Berlin
{zimmer, pokutta}@zib.de

Lorenz Richter*
dida
Zuse Institute Berlin
richter@zib.de

Abstract

We study the numerical approximation of Schrödinger Bridges in the setting where samples from the target distribution are unavailable and only an unnormalized target density is given. In this regime, the widely used iterative proportional fitting (IPF) procedure (and its variants) often exhibits severe numerical instabilities, stemming from ill-conditioned underlying control problems and from the fact that the optimal solution may lie far from the initialization. To address these issues, we propose a trust region variant of IPF that constrains successive updates to remain at a fixed Kullback-Leibler divergence from the current iterate. This restriction significantly improves numerical stability. Moreover, we show that the resulting method admits an interpretation as a mirror descent scheme with an adaptively chosen step size. We demonstrate the effectiveness of our approach on several challenging sampling tasks.

1 Introduction

We study the problem of sampling from a target distribution

$$p_{\mathrm{target}} = \frac{\exp\left(-\mathcal{E}\right)}{\mathcal{Z}}, \tag{1}$$

where $\mathcal{E} \in C(\mathbb{R}^d, \mathbb{R})$ is an energy function that can be evaluated pointwise and $\mathcal{Z} = \int_{\mathbb{R}^d} \exp\left(-\mathcal{E}(x)\right) \mathrm{d}x$ is the normalizing constant. Our approach relies on transporting samples from a simple distribution to the target via a stochastic process. In particular, we study Schrödinger bridges, i.e., stochastic measure transport with minimal control costs, mathematically equivalent to entropy-regularized optimal transport [29]. While Schrödinger bridges have been extensively studied in data-driven settings, where samples from the target distribution are available, their application to pure sampling tasks remains less explored. In the sampling regime, where only the energy $\mathcal{E}$ is available, the problem becomes significantly more challenging, e.g., due to unknown locations of the modes of the distribution.

Existing approaches for solving Schrödinger bridges often rely on iterative schemes, such as iterative proportional fitting (IPF) or Markovian projections [19, 46]. However, applying these to sampling introduces difficult optimization landscapes characterized by a delicate exploration-exploitation tradeoff and a notorious tendency toward mode-collapse, often implying severe numerical instabilities. This stems from two primary sources: the ill-conditioning of the underlying control problems and the

*Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

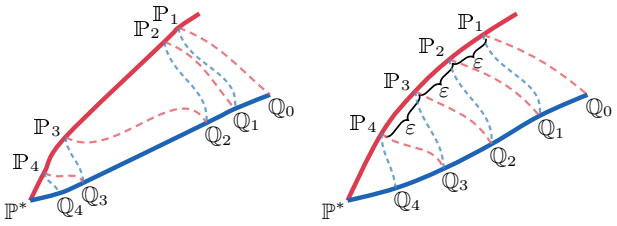


Figure 1: The classical iterative proportional fitting (IPF) algorithm for solving Schrödinger bridges proceeds by alternating projections, producing sequences of path measures $(\mathbb{P}_i)_{i\geq 0}$ and $(\mathbb{Q}_i)_{i\geq 0}$ that are guaranteed to converge to the optimal measure $\mathbb{P}^*$. However, as illustrated on the left, successive measures $\mathbb{P}_i$ may be separated by arbitrarily large distances, which can complicate optimization and lead to numerical instabilities. To address this issue, we propose the *Trust Region Schrödinger Bridge Sampler*, which enforces equidistant updates in terms of the Kullback-Leibler divergence.

substantial distance often found between the initialization and the optimal solution. The classical IPF algorithm can be viewed as an alternating KL projection between boundary constraints. Given an intermediate path measure $\mathbb{Q}_i$, the next iterate is obtained by projecting onto a constraint $\mathcal{C}$:

$$\mathbb{P}_{i+1} = \operatorname*{arg\,min}_{\mathbb{P}\in\mathcal{C}} D_{\mathrm{KL}}(\mathbb{P}|\mathbb{Q}_i). \tag{2}$$

It can be shown that this corresponds to a mirror descent scheme with a fixed step size of 1 [28]. By relaxing this parameter, one can consider a generalized update rule with step size γ, akin to the discretization of continuous Sinkhorn flows,

$$\mathbb{P}_{i+1} \approx \mathbb{P}_i^{1-\gamma}(\mathrm{Proj}_{\mathcal{C}}\,\mathbb{Q}_i)^{\gamma}, \tag{3}$$

where $\mathrm{Proj}_{\mathcal{C}}$ denotes the KL projection in (2). While this generalized view suggests that the step size γ could technically be varied to stabilize training, practical guidelines for selecting this parameter are non-existent.

To address these limitations, we extend the recently introduced Adjoint Schrödinger Bridge Sampler (ASBS) [30] along two complementary axes. First, we generalize its unit-step forward matching update to arbitrary mirror-descent step sizes and select the step size adaptively through a path space KL trust region. Second, we incorporate the importance weight that the forward KL objective requires but ASBS omits. Combining these extensions yields the *Trust Region Schrödinger Bridge Sampler* (TR-SBS), which mitigates mode collapse while retaining the simulation-efficient matching structure of ASBS. Our main contributions are:

1. We propose TR-SBS, a γ-IPF scheme whose step size is selected by a path space KL trust region, adapted to the moving target of IPF (Proposition 3.1).
2. We derive a generalized adjoint-matching objective that accommodates arbitrary step sizes and incorporates the importance weight that the cross-entropy objective requires (Proposition 3.3).
3. We establish a convergence guarantee for the resulting trust region SB iteration (Proposition 3.4) and empirically demonstrate that adaptive step size selection and importance weighting play complementary roles in mitigating mode collapse (Section 4).

1.1 Related work

The Schrödinger bridge problem dates back to the seminal work of [44] and has since been studied extensively in probability theory and stochastic control; see, e.g., [15, 19, 29, 42]. More recently, the iterative proportional fitting (IPF) algorithm for numerical approximation has been revisited from an optimization perspective, and its equivalence to a mirror descent scheme has been analyzed in [3, 28]. In the past few years, Schrödinger bridges have emerged as a powerful framework for generative modeling, providing a unifying viewpoint that connects stochastic processes, optimal transport, and diffusion-based methods. Representative contributions include [12, 16, 22, 50, 53, 54], all of which focus on data-driven settings where samples from the target distribution are available. Subsequent research has extended these foundations in several directions. For instance, [46] proposed diffusion Schrödinger bridge matching, while [37] investigated mixture-based transport constructions. A further unifying perspective was introduced by [31].

Schrödinger bridges have also been explored in data-free settings, where the goal is sampling from a target distribution; see, e.g., [6, 27]. In parallel, diffusion-based sampling methods employing

alternative transport maps have led to substantial performance improvements, as demonstrated in [4, 5, 11, 23, 48, 51, 56]. These developments were further unified and generalized in [40, 52], with extensions to the underdamped setting studied in [8]. Related control-matching approaches based on fixed-point iterations have been investigated in [13, 17, 26, 30, 33]. Among these works, ASBS [30] is most closely related to ours. ASBS implements the unit-step IPF update through an unweighted adjoint-matching fixed-point objective. We generalize this construction to arbitrary step sizes, incorporate the importance weight that the forward KL objective requires, and select the step size adaptively through a KL trust region. ASBS has itself been generalized by the Bridge Matching Sampler [9], which casts such matching objectives as fixed-point iterations based on Nelson's relation and yields a single, stable objective.

Finally, we want to mention that trust region methods have a long history as robust optimization algorithms; see [14] for an overview. These methods have been extended to probability distributions, particularly in reinforcement learning, see, e.g., [1, 35, 39, 45, 49]. A treatment of trust regions for stochastic optimal control and path space measures has recently been suggested in [7], where the target path measure is fixed. In contrast, the target of our forward update changes after every backward IPF step, which calls for a moving-target formulation and an SB-specific reformulation; we detail the differences in Appendix A.8.

2 Schrödinger bridges

A Schrödinger bridge between p_{prior} and p_{target} is defined as a stochastic process

$$\mathrm{d}X_t^u = \sigma(t)u(X_t^u, t)\mathrm{d}t + \sigma(t)\vec{\mathrm{d}}W_t, \qquad X_0^u \sim p_{\mathrm{prior}} \tag{4}$$

with minimal control costs in the sense of

$$\min_{u\in\mathcal{U}} D_{\mathrm{KL}}(\vec{\mathbb{P}}^u|\vec{\mathbb{P}}^{\mathrm{ref}}) = \min_{u\in\mathcal{U}} \mathbb{E}\left[\frac{1}{2}\int_0^T \|u(X_t^u,t)\|^2\mathrm{d}t\right] \tag{5}$$

such that $X_T^u \sim p_{\mathrm{target}}$ (see Appendix A for remarks on the notation). This problem can be approached by additionally considering the reverse-time SDE

$$\mathrm{d}Y_t^v = -\sigma(t)v(Y_t^v, t)\mathrm{d}t + \sigma(t)\overleftarrow{\mathrm{d}}W_t, \qquad Y_T^v \sim p_{\mathrm{target}}. \tag{6}$$

As for instance shown in [29], the solution is connected to the *Schrödinger system*

$$\varphi_t(x) = \int_{\mathbb{R}^d} p^{\mathrm{ref}}_{T|t}(y|x)\varphi_T(y)\mathrm{d}y, \qquad \widehat{\varphi}_t(x) = \int_{\mathbb{R}^d} p^{\mathrm{ref}}_{t|0}(x|y)\widehat{\varphi}_0(y)\mathrm{d}y \tag{7}$$

with coupled boundary conditions $\varphi_0(x)\widehat{\varphi}_0(x) = p_{\mathrm{prior}}(x)$ and $\varphi_T(x)\widehat{\varphi}_T(x) = p_{\mathrm{target}}(x)$, where $p^{\mathrm{ref}}_{T|t}$ is the transition density of the reference process. The functions φ and $\widehat{\varphi}$ are called *Schrödinger potentials* and the optimal controls solving the Schrödinger bridge problem (and its time-reversal) are then given by

$$u^* = \sigma^\top \nabla \log \varphi, \qquad v^* = \sigma^\top \nabla \log \widehat{\varphi}. \tag{8}$$

In this case, Y^{v^*} is the time reversal of X^{u^*}, i.e., $\vec{\mathbb{P}}^{u^*} = \overleftarrow{\mathbb{P}}^{v^*}$.

2.1 Iterative proportional fitting

Solving for the optimal control functions (8) is non-trivial, in particular due to the coupled boundary conditions (7). A classical approach is based on iteratively solving so-called *Schrödinger half-bridges*, in which only one of the time marginals of the process is fixed. The resulting *iterative proportional fitting* (IPF) procedure alternates updates of the forward and backward controls as

$$\begin{aligned} u_{i+1} &= \operatorname*{arg\,min}_{u\in\mathcal{U}} \Big\{ D_{\mathrm{KL}}(\vec{\mathbb{P}}^u|\overleftarrow{\mathbb{P}}^{v_i}) : \vec{\mathbb{P}}_0^u = p_{\mathrm{prior}} \Big\}, \\ v_{i+1} &= \operatorname*{arg\,min}_{v\in\mathcal{U}} \Big\{ D_{\mathrm{KL}}(\overleftarrow{\mathbb{P}}^v|\vec{\mathbb{P}}^{u_{i+1}}) : \overleftarrow{\mathbb{P}}_T^v = p_{\mathrm{target}} \Big\}. \end{aligned} \tag{9}$$

Since in our setting we do not have samples from p_{target}, it is convenient to replace the reverse KL in the second equation by a forward KL, yielding

$$v_{i+1} = \operatorname*{arg\,min}_{v\in\mathcal{U}} \Big\{ D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}}|\overleftarrow{\mathbb{P}}^v) : \overleftarrow{\mathbb{P}}_T^v = p_{\mathrm{target}} \Big\}. \tag{10}$$

As shown in Corollary E.2 of [52], replacing the reverse KL with the forward KL in the backward update does not alter the minimizing sequence (u_i, v_i). Moreover, under regularity assumptions, the IPF iterates converge to the Schrödinger bridge in total variation as $i \to \infty$ [16].

2.2 IPF as mirror descent

In [28] it is shown that the IPF iterates stated in (9) can be equivalently formulated by a mirror descent scheme on path space,

$$u_{i+1} = \arg\min_{u \in \mathcal{U}} \left\{ \langle \delta F(\vec{\mathbb{P}}^{u_i}), \vec{\mathbb{P}}^{u} - \vec{\mathbb{P}}^{u_i} \rangle + \frac{D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u} | \vec{\mathbb{P}}^{u_i})}{\gamma_i} \right\}, \tag{11}$$

when choosing $F(\vec{\mathbb{P}}^{u_i}) = D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_i}_T | p_{\mathrm{target}})$ and the step size $\gamma_i = 1$. Here, $\delta F(\vec{\mathbb{P}}^{u_i})$ denotes the first variation of the terminal loss functional F, evaluated at the current iterate $\vec{\mathbb{P}}^{u_i}$. Thus, each update balances a linearized descent step that drives the terminal marginal toward the target distribution p_{target} against a regularization term that penalizes deviations from the current path space dynamics $\vec{\mathbb{P}}^{u_i}$ via the KL divergence. Further, as also shown in [28], we can integrate the step sizes $\gamma_i \in (0, 1]$ into the classical IPF formulation (9), yielding

$$u_{i+1} = \arg\min_{u \in \mathcal{U}} \left\{ D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u} | \overleftarrow{\mathbb{P}}^{v_i}) + \frac{1-\gamma_i}{\gamma_i} D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u} | \vec{\mathbb{P}}^{u_i}) \right\} \tag{12}$$

under the constraint $\vec{\mathbb{P}}^{u}_0 = p_{\mathrm{prior}}$ for the u-updates, while the v-updates remain the same. In practice, it is natural to ask which step size to choose in order to stabilize the procedure. We shall answer this question in the following section.

3 Trust region based Schrödinger bridges

Inspired by [7], we propose to stabilize the u-updates in the IPF procedure (9) by incorporating a trust region constraint. The purpose of this constraint is to control the magnitude of each update and to ensure that successive control iterates remain within a prescribed neighborhood of the previous iterate. Concretely, we restrict deviations, measured in terms of the KL divergence, from the previous control u_i by a radius $\varepsilon > 0$, and define

$$u_{i+1} = \arg\min_{u \in \mathcal{U}} D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u} | \overleftarrow{\mathbb{P}}^{v_i}) \;\text{ s.t. }\; D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u} | \vec{\mathbb{P}}^{u_i}) \leq \varepsilon, \tag{13}$$

again subject to the half-bridge constraint $\vec{\mathbb{P}}^{u}_0 = p_{\mathrm{prior}}$. We call the iteration that alternates the backward IPF half-step with the forward update (13) *trust region IPF* (TR-IPF) and state it in path space in Algorithm 1 (Appendix A.6). TR-SBS, which we describe in Section 3.1, learns the steps of TR-IPF from samples. Its convergence, speed, and stability are analyzed in Section 3.2.

In practice, under suitable regularity assumptions, the constrained optimization problem (13) can be addressed via a relaxed Lagrangian formulation. Introducing a Lagrange multiplier $\lambda \geq 0$, we define the augmented objective

$$\mathcal{L}_i^{\mathrm{TR}}(u, \lambda) := D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u} | \overleftarrow{\mathbb{P}}^{v_i}) + \lambda \big(D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u} | \vec{\mathbb{P}}^{u_i}) - \varepsilon \big), \tag{14}$$

which leads to the saddle-point problem

$$\max_{\lambda \geq 0} \min_{u \in \mathcal{U}} \mathcal{L}_i^{\mathrm{TR}}(u, \lambda). \tag{15}$$

We note that the KL divergence is convex in the path measures. Moreover, the value function $\lambda \mapsto \min_{u \in \mathcal{U}} \mathcal{L}_i^{\mathrm{TR}}(u, \lambda)$ is concave in λ, since it is given as the pointwise infimum of affine functions of λ; see, e.g., [7]. Under standard regularity assumptions, the saddle-point problem (15) thus admits a unique solution, which we denote by (u_{i+1}, λ_i). For later analysis, we introduce the cost-to-go functional of the trust region control (TRC) problem, where it is crucial to restrict the evaluation to the future path segments. Let $\vec{\mathbb{P}}^{u}_{[t,T]}$ and $\overleftarrow{\mathbb{P}}^{v}_{[t,T]}$ denote the path measures restricted to the time interval $[t, T]$. We define[2]

$$\mathcal{L}_i^{\mathrm{TRC}}(u, \lambda; x, t) := \mathbb{E}_{\vec{\mathbb{P}}^{u}} \left[\log \frac{\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}} + \lambda \log \frac{\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}} \middle| X_t = x \right]. \tag{16}$$

[2]Note that the normalization constant $\mathcal{Z}$ from p_{target} in the Radon-Nikodym derivative $\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]} / \mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}$ is merely an additive constant independent of u and γ and can therefore be ignored in optimization.

By definition, this functional accumulates the running and terminal costs only from time t to T. Observe that taking the expectation at $t = 0$ recovers the original Lagrangian up to a constant shift:

$$\mathbb{E}_{X_0 \sim p_{\text{prior}}}\left[\mathcal{L}_i^{\text{TRC}}(u, \lambda; X_0, 0)\right] = \mathcal{L}_i^{\text{TR}}(u, \lambda) + \lambda\varepsilon. \tag{17}$$

Since the term $\lambda\varepsilon$ is constant with respect to u, minimizing $\mathcal{L}_i^{\text{TR}}(u, \lambda)$ is equivalent to minimizing the expected costs defined by (17).

Comparing $\mathcal{L}_i^{\text{TRC}}$ with the generalized Sinkhorn update (12), we observe that the Lagrange multiplier can be identified with the IPF step size via $\lambda = \frac{1-\gamma}{\gamma}$. Enforcing the constraint $D_{\text{KL}}(\vec{\mathbb{P}}^u | \vec{\mathbb{P}}^{u_i}) \leq \varepsilon$ induces an adaptive choice of the effective step size γ_i at each iteration i. Further, we note that $\mathcal{L}_i^{\text{TRC}}$ is again a cost functional of a classical unconstrained optimal control problem, see the proof of Proposition 3.2 in Appendix A.

The following result shows that the new iterates $u_{i+1} = u_{i+1}^{\lambda_i}$, obtained by solving (15) for the optimal Lagrange multiplier λ_i, can be derived from the previous iterates u_i by reweighting with respect to the current target $\overleftarrow{\mathbb{P}}^{v_i}$. We refer to Appendix A for the proof.

Proposition 3.1 (Geometric interpretation of the trust region update)**.** *Let* $u_{i+1}^{\lambda} := \arg\min_{u \in \mathcal{U}} \mathcal{L}_i^{\text{TR}}(u, \lambda)$ *be the solution at iteration* i*. For any* $\lambda \geq 0$*, the path measure* $\vec{\mathbb{P}}^{u_{i+1}^{\lambda}}$ *can be interpreted as the projection of a geometric average of measures. In particular, it holds that*

$$\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}^{\lambda}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X) \propto \left(\frac{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X)\right)^{\frac{1}{1+\lambda}} = \left(\frac{p_{\text{target}}}{\vec{\mathbb{P}}_T^{u_i}}(X_T)\right)^{\frac{1}{1+\lambda}} =: w_i(X)^{\frac{1}{1+\lambda}}. \tag{18}$$

The proportionality constant depends only on X_0 *and is fixed by the constraint* $\vec{\mathbb{P}}_0^{u_{i+1}^{\lambda}} = p_{\text{prior}}$*; explicitly,* $\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}^{\lambda}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X) = w_i(X)^{\frac{1}{1+\lambda}} / \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}\left[w_i^{\frac{1}{1+\lambda}} \,\middle|\, X_0\right]$.

We can further derive an explicit formula for the value functions associated with the iterations in the trust region Schrödinger bridge problem. The proof is again stated in Appendix A.

Proposition 3.2 (Optimality for trust region Schrödinger bridge problem)**.** *Let the value function be defined as the minimal cost-to-go conditioned on* $X_t^u = x$*:*

$$V_{i+1}^{\lambda}(x, t) := \inf_{u \in \mathcal{U}} \mathcal{L}_i^{\text{TRC}}(u, \lambda; x, t). \tag{19}$$

Then, the value function satisfies

$$V_{i+1}^{\lambda}(x, t) = -(1+\lambda)\log \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}\left[\left(\frac{\mathrm{d}\vec{\mathbb{P}}_{[t,T]}^{u_i}}{\mathrm{d}\overleftarrow{\mathbb{P}}_{[t,T]}^{v_i}}(X)\right)^{-\frac{1}{1+\lambda}} \middle| X_t = x\right], \tag{20}$$

where the Radon-Nikodym derivative restricted to $[t, T]$ *is given explicitly in Lemma A.2. Moreover, the optimal control admits the representation* $u_{i+1}^{\lambda} = \frac{\lambda}{1+\lambda}u_i - \frac{1}{1+\lambda}v_i - \frac{1}{1+\lambda}\sigma^\top \nabla\left(V_{i+1}^{\lambda} - \log \vec{\mathbb{P}}_t^{u_i}\right)$.

A key insight from Proposition 3.2 is that it provides an explicit representation of the value function V_{i+1}^{λ} for arbitrary λ, up to expectation terms that can be efficiently approximated via Monte Carlo sampling. This representation allows us to identify the optimal dual variable λ_i by maximizing the associated Lagrangian dual function. Specifically, substituting the optimal control u_{i+1}^{λ} into the objective yields the dual function

$$\mathcal{L}_i^{\text{Dual}}(\lambda) := \mathcal{L}_i^{\text{TR}}(u_{i+1}^{\lambda}, \lambda) = \mathbb{E}_{X_0 \sim p_{\text{prior}}}[V_{i+1}^{\lambda}(X_0, 0)] - \lambda\varepsilon. \tag{21}$$

Its first term depends on the trajectories only through the weights w_i in (18), which also enter the matching loss (25). A Monte Carlo estimate of $\mathcal{L}_i^{\text{Dual}}(\lambda)$ only requires trajectories of the current forward process X^{u_i} and energy evaluations $\mathcal{E}(X_T)$, but neither gradients of $\mathcal{E}$ nor differentiation through the simulation (Algorithm 3). The optimal dual

$$\lambda_i = \arg\max_{\lambda \geq 0} \mathcal{L}_i^{\text{Dual}}(\lambda) \tag{22}$$

is the maximizer of a concave one-dimensional function and can therefore be computed with standard scalar optimization routines.

In principle, one could then define the next control as $u_{i+1} = u_{i+1}^{\lambda_i}$ using the characterization in Proposition 3.2. However, this requires Monte Carlo estimates of the value function and its gradient $\nabla V_{i+1}^{\lambda_i}$ at every state x, which is infeasible in high-dimensional state spaces. We therefore learn a parameterized approximation of u_{i+1}, as described in the next section.

3.1 Learning the constrained optimal control

In this section, we describe how the optimal control is learned at each iteration. To this end, we introduce the cross-entropy loss defined via the forward KL divergence [34]

$$\mathcal{L}_i^{\mathrm{CE}}(u) := D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}} | \vec{\mathbb{P}}^{u}). \tag{23}$$

This objective becomes tractable by means of importance sampling, since trajectories X^{u_i} can be generated from the previous iterate:

$$\mathcal{L}_i^{\mathrm{CE}}(u) = \mathbb{E}_{\vec{\mathbb{P}}^{u_i}} \left[\left(\log \frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u}}(X^{u_i}) \right) \frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X^{u_i}) \right]. \tag{24}$$

In general, importance reweighting introduces exponential terms that may lead to large variance and numerical instabilities, particularly in high-dimensional settings; see, e.g., [24]. Omitting the importance weights leads to a fixed-point type iteration, as in ASBS [30], whose objective no longer corresponds to a proper divergence and only uses the gradient of the target rather than the target itself. Gradient information alone does not determine the relative weights of well-separated modes (mode blindness), which is problematic for multimodal targets [36].

In our setting, however, the trust region keeps the variance of the log-weights at about 2ε for small ε (Section 3.2; see also Appendix A.7 and, for the weights, [7, Appendix I.3]), which allows us to retain the principled importance reweighting. At the same time, the cross-entropy loss admits a least-squares matching representation [18].

Proposition 3.3 (Generalized adjoint matching representation of the cross-entropy loss)**.** *Up to a positive factor and an additive constant, the cross-entropy loss* (23) *can be equivalently expressed as*

$$\mathcal{L}_i^{\mathrm{F}}(u) = \mathbb{E}_{\vec{\mathbb{P}}^{u_i}_{0,T} \vec{\mathbb{P}}^{\mathrm{ref}}_{t|0,T}} \left[\left\| u(X_t,t) + \frac{v_i(X_T,T) + \sigma^\top \nabla \mathcal{E}(X_T)}{1+\lambda_i} - \frac{\lambda_i}{1+\lambda_i} u_i(X_T,T) \right\|^2 \frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X) \right], \tag{25}$$

where the importance weight $\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} \propto \left(\frac{\mathrm{d}\vec{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}\right)^{\frac{1}{1+\lambda_i}}$ *is given in Proposition 3.1,* $\mathcal{E}(x) = -\log p_{\mathrm{target}}(x) - \log \mathcal{Z}$ *denotes the energy function, the expectation is over endpoints* $(X_0, X_T) \sim \vec{\mathbb{P}}^{u_i}_{0,T}$*,* $t \sim \mathcal{U}(0,T)$*, and* $X_t \sim \vec{\mathbb{P}}^{\mathrm{ref}}_{t|0,T}(\cdot \mid X_0, X_T)$*, the reference process conditioned on these endpoints. For a time-dependent* σ*, the terms after* $u(X_t,t)$ *in* (25) *are evaluated with* $\sigma = \sigma(T)$ *and multiplied by* $\sigma(t)^\top \sigma(T)^{-\top}$ *(Appendix A.5).*

The objective (25) thus retains the simulation-efficient matching structure of [30] while remaining a proper divergence that incorporates target information through the importance weight; without this weight, one obtains fixed-point iterations such as ASBS, and with an inactive KL constraint ($\lambda_i = 0$, $\gamma_i = 1$), TR-SBS reduces to ASBS-IW, i.e., ASBS with the importance weight (Figure 2 and Appendix A.9).

To complete the IPF scheme (9), it remains to perform the v-updates. We employ a generalized matching objective based on [30]. A key distinction from standard bridge matching approaches [37, 38, 46], which rely on ground-truth target samples $X_T \sim p_{\mathrm{target}}$, is that our formulation utilizes samples $X_T \sim \vec{\mathbb{P}}_T^{u_{i+1}}$ generated by the current forward iterate. Specifically, we define the loss as

$$\mathcal{L}_i^{\mathrm{B}}(v) = \mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}_{0,T} \vec{\mathbb{P}}^{\mathrm{ref}}_{t|0,T}} \left[\zeta_t \left\| v(X_t,t) - \sigma^\top \nabla \log p^{\mathrm{ref}}_{t|0}(X_t|X_0) \right\|^2 \right], \tag{26}$$

where $\zeta_t = (1+\sigma(t)^2(T-t)/t)^{-1}$ is a time-dependent weighting factor, cf. [46], and the expectation is taken as in (25) but with endpoints from $\vec{\mathbb{P}}^{u_{i+1}}_{0,T}$ and without importance weight. Its minimizer is the backward IPF half-step (Lemma A.4). The overall training procedure of our Trust Region Schrödinger Bridge Sampler (TR-SBS) is summarized in Algorithm 2 in Appendix B.5. The diffusion scale $\sigma_{\max}$ and the computational cost are discussed in Appendices B.7 and C.4.

3.2 Convergence of TR-IPF: speed and stability

We now analyze how the trust region affects convergence in the idealized setting in which each step of TR-SBS is computed exactly, i.e., without sampling and optimization errors, so that TR-SBS

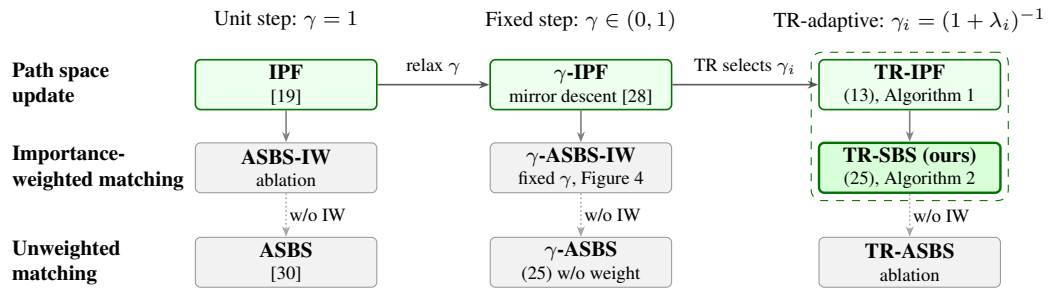


Figure 2: **Overview of IPF-based samplers.** The columns differ in how the mirror descent step size of the forward update is chosen. The rows show the path space update, which reweights the current iterate $\vec{\mathbb{P}}^{u_i}$ by w_i^γ (Proposition 3.1), its implementation by matching objectives learned from samples of $\vec{\mathbb{P}}^{u_i}$ with this importance weight (IW), and the variants without it.

reduces to TR-IPF; we discuss deviations from this setting in Section 5. To this end, let $\mathbb{P}^*$ minimize $F(\mathbb{P}) \coloneqq D_{\mathrm{KL}}(\mathbb{P}_T | p_{\mathrm{target}})$ over path measures with $\mathbb{P}_0 = p_{\mathrm{prior}}$. As stated in Section 2.2, IPF is mirror descent on F with unit step size, so TR-IPF is mirror descent with the adaptive step size γ_i. This yields the following result, which we prove in Appendix A.6.

Proposition 3.4 (Convergence of TR-IPF)**.** *Let $(\vec{\mathbb{P}}^{u_i})_{i \geq 0}$ be generated by TR-IPF (Algorithm 1) and $\ell_i \coloneqq \log \frac{\mathrm{d}\vec{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}$. Assume that the constraint is active, $D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}} | \vec{\mathbb{P}}^{u_i}) = \varepsilon$, and that $\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}}(\ell_i) \leq M$ for all i. Then $F(\vec{\mathbb{P}}^{u_{i+1}}) \leq F(\vec{\mathbb{P}}^{u_i})$ for all i, and, at leading order for small ε,*

$$\gamma_i \gtrsim \sqrt{2\varepsilon / M}, \qquad F(\vec{\mathbb{P}}^{u_I}) - F(\mathbb{P}^*) \lesssim \frac{D_{\mathrm{KL}}(\mathbb{P}^* | \vec{\mathbb{P}}^{u_0}) \sqrt{M}}{I \sqrt{2\varepsilon}}. \tag{27}$$

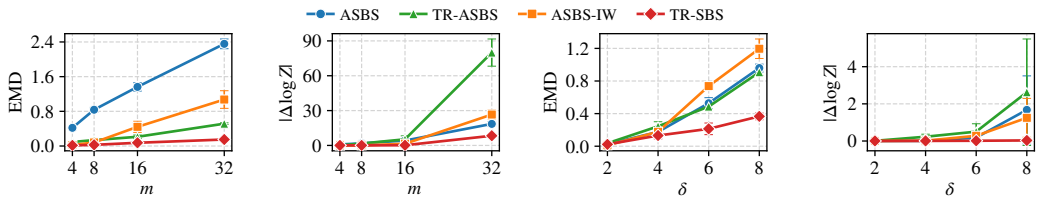
Both inequalities in (27) rest on the small-step expansion $\varepsilon = D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}} | \vec{\mathbb{P}}^{u_i}) \approx \gamma_i^2 \overline{V}_i / 2$, where $\overline{V}_i \coloneqq \mathbb{E}_{X_0 \sim p_{\mathrm{prior}}}[\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}}(\ell_i \,|\, X_0)] \leq M$, which determines the step size $\gamma_i \approx \sqrt{2\varepsilon / \overline{V}_i}$ as well as the variance of the log-weights.

Speed. With a fixed step size γ, the gap after I iterations is at most $D_{\mathrm{KL}}(\mathbb{P}^* | \vec{\mathbb{P}}^{u_0}) / (\gamma I)$ (Lemma A.5). The trust region acts as a step size of at least $\sqrt{2\varepsilon / M}$, which gives (27), i.e., $I = \mathcal{O}(\varepsilon^{-1/2})$ iterations for a given gap; the step grows automatically when $\overline{V}_i$ is small, e.g., as $\vec{\mathbb{P}}^{u_i}$ approaches the target.

Stability. By Proposition 3.1, $\gamma_i \ell_i$ is, up to a function of X_0, the logarithm of the importance weight of the update. Its conditional variance, averaged over X_0, is thus $\gamma_i^2 \overline{V}_i \approx 2\varepsilon$ at leading order. With a fixed step size, it is $\gamma^2 \overline{V}_i$, which becomes large when $\vec{\mathbb{P}}^{u_i}$ puts mass where the target is small, e.g., in early iterations, so that the importance weights degenerate. The radius ε thus trades stability for speed. Conversely, if $\overline{V}_i$ is very large, the step size $\gamma_i \approx \sqrt{2\varepsilon / \overline{V}_i}$ becomes so small that the learned updates stall, as we observe in some of our experiments. Bounding γ_i from below (Appendix B.4) restores progress but gives up the variance control, since then $\gamma_i^2 \overline{V}_i > 2\varepsilon$; the trust region alone does not resolve this dilemma between stagnation and degenerate weights.

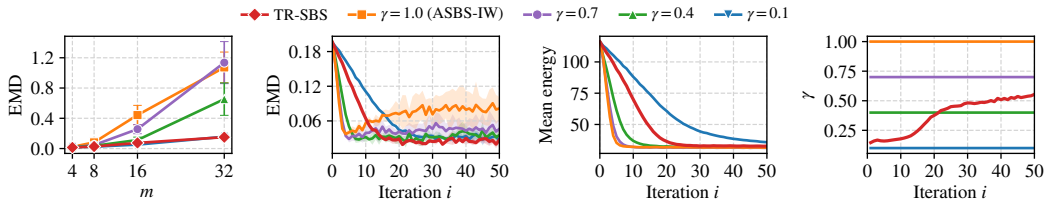
4 Experiments

We evaluate our proposed method on high-dimensional sampling tasks designed to rigorously test its mode-seeking capabilities and scalability. Specifically, we consider a high-dimensional Many-Well (MW) potential characterized by exponentially many modes, a Gaussian Mixture Model (GMM) that affords precise control over mode placement, and alanine dipeptide, a molecular benchmark with several metastable conformations, one of which is rare. On MW, we report the marginal Earth Mover's Distance (EMD) and the error of the log-normalizing constant ($|\Delta \log Z|$), and on GMM, the Sinkhorn divergence (SD) [20] and the total variation distance (TVD) of the mixture weights; definitions and further metrics are given in Appendices B and C.



(a) EMD (vs. m) (b) $|\Delta \log Z|$ (vs. m) (c) EMD (vs. δ) (d) $|\Delta \log Z|$ (vs. δ)

Figure 3: **Results on Many-Well potential.** We evaluate robustness to increasing modes and robustness to mode distance. **(a, b)** Performance vs. number of modes. **(c, d)** Performance vs. mode distance. Our proposed method (TR-SBS, red) achieves the lowest estimation error and marginal distance for large m and δ, handling both combinatorial complexity and high energy barriers.



(a) EMD (vs. m) (b) EMD (c) Mean energy (d) Gamma

Figure 4: **Ablation study on the step size γ.** Legend is shared across all subplots. **(a)** Adaptive TR-SBS matches the robustness of conservative $\gamma = 0.1$. **(b, c)** Dynamics reveal the trade-off: Small $\gamma = 0.1$ is stable (b) but slow (c). TR-SBS converges efficiently (c) while maintaining stability (b). **(d)** The adaptive γ starts conservative to prevent collapse and increases to accelerate convergence.

To isolate the contributions of the two ingredients of TR-SBS, the trust region and the importance weight, we compare **TR-SBS** with the original **ASBS** [30] and two ablations that each add one of them to ASBS (Figure 2): **ASBS-IW**, which adds only the importance weight, i.e., minimizes the forward KL divergence with unit step size, and **TR-ASBS**, which selects the step size by the trust region without importance weighting. On GMM, we additionally compare with two diffusion-based samplers outside the ASBS family, **PIS** [56] and **DIS** [5] (Appendix B.6).

Qualitative visualizations and further results are provided in Appendix C, including a ϕ^4 scalar field theory model (Appendix C.5) and the Lennard-Jones cluster LJ-13 (Appendix C.7).

4.1 Many-Well potential

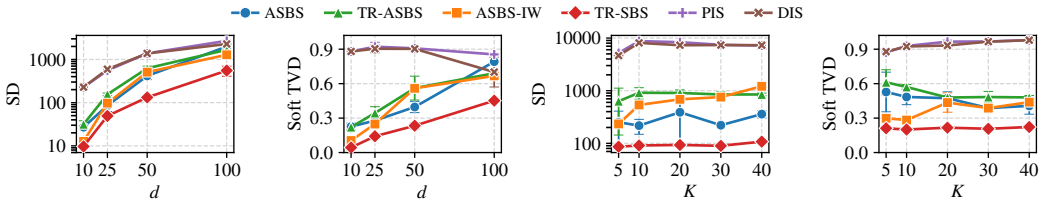
To evaluate the sampler's ability to escape local minima and cross energy barriers, we employ the MW potential [40, 55]. The energy function $\mathcal{E}$ is given by

$$\mathcal{E}(x) = \sum_{i=1}^{m} (x_i^2 - \delta)^2 + \frac{1}{2} \sum_{j=m+1}^{d} x_j^2. \tag{28}$$

Here, the first m dimensions represent double-well potentials, creating 2^m modes.

Results. For $d = 64$ and $m \in \{4, \ldots, 32\}$ double-well dimensions (Figure 3 (a, b)), **TR-SBS** (red) performs best in both metrics, which shows that combining the trust region with importance weighting is essential for accurate global sampling in multimodal landscapes. For $d = 8$, $m = 5$, and barrier distances $\delta \in \{2, \ldots, 8\}$ (Figure 3 (c, d)), TR-SBS attains the lowest marginal EMD and log-partition error for $\delta \geq 4$, i.e., it is robust to large separations and balances the probability mass between distant modes.

Effect of the mirror descent step size $\gamma = (1 + \lambda)^{-1}$ We compare TR-SBS with γ-ASBS-IW (Figure 2) using fixed step sizes $\gamma \in \{1.0, 0.7, 0.4, 0.1\}$, where $\gamma = 1$ is ASBS-IW, on MW with $d = 64$, analyzing both robustness to mode complexity (m) and training dynamics ($m = 8$); see Figure 4.



(a) SD (vs. d) (b) TVD (vs. d) (c) SD (vs. K) (d) TVD (vs. K)

Figure 5: **Results on GMM** (mean $\pm$ standard deviation over three seeds). **(a, b)** Performance vs. dimension $d \in \{10, 25, 50, 100\}$ with $K = 10$. **(c, d)** Performance vs. number of components $K \in \{5, 10, 20, 30, 40\}$ with $d = 10$. TR-SBS (red) attains the lowest SD and TVD in every setting. PIS and DIS use 128 instead of 200 time steps and a different training budget (Appendix B.6).

Stability: Figure 4a shows that aggressive fixed step sizes ($\gamma \geq 0.4$) fail to prevent mode collapse as combinatorial complexity increases ($m \to 32$). A conservative step size ($\gamma \approx 0.1$) is required to maintain low error, a level of robustness matched by our adaptive TR-SBS.

Efficiency vs. Stability: Since each outer iteration uses $N_{\mathrm{FC}} N_{\mathrm{new}}$ target-gradient evaluations, the iteration axis in Figure 4 is also proportional to their cumulative number, which scales as $\mathcal{O}(I N_{\mathrm{FC}} N_{\mathrm{new}})$. Figure 4c shows that a small step $\gamma = 0.1$ results in slow convergence of the mean energy $\mathbb{E}[\mathcal{E}(X_T)]$, whereas TR-SBS (red) converges more than twice as fast. However, Figure 4b shows that aggressive steps converge to a higher error, while TR-SBS attains the low final error of conservative steps. As shown in Figure 4d, the adaptive γ starts small (≈ 0.15) for initial stability and increases as the training stabilizes, as expected from Section 3.2. Finally, we refer to an ablation study on the impact of the maximum noise scale in Appendix C.4.

4.2 Gaussian mixture model (GMM)

We verify the scalability and robustness of the sampler using a GMM, $p_{\mathrm{target}}(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, s^2 I_d)$ with $s = \ln(1 + e)$, means μ_k sampled uniformly from $[-L, L]^d$, and random mixing weights with $\max_k \pi_k / \min_k \pi_k \leq 3$. We use the TVD between the true and empirical mixing weights as a primary metric for detecting mode collapse, alongside the SD to assess geometric fidelity. For each run, we evaluate the checkpoint with the lowest SD during training on $4{,}000$ independent samples; all settings are given in Appendix B.1.

Results. For $K = 10$, $L = 7$, and $d \in \{10, 25, 50, 100\}$, Figure 5 (a, b) shows that TR-SBS attains the lowest SD and TVD in every dimension, and its advantage grows with d (e.g., TVD 0.23 vs. at least 0.40 for the ASBS variants at $d = 50$). Neither importance weighting with unit step size (ASBS-IW) nor the trust region without importance weighting (TR-ASBS) is consistently better than ASBS. Only their combination attains the lowest error for all d, which suggests that the two components are complementary. PIS and DIS concentrate on few components in this setting (TVD ≥ 0.70 for all d). For $d = 10$, widely separated components ($L = 40$), and $K \in \{5, 10, 20, 30, 40\}$ (Figure 5 (c, d)), TR-SBS keeps the TVD at about 0.2 for all K, while all other methods have a TVD of at least 0.28 and an SD at least twice as large. The error of ASBS-IW grows with K, which shows that importance-weighted updates need the small steps enforced by the trust region. Figure 8 in Appendix C.2 visualizes the samples for $K = 40$.

4.3 Alanine dipeptide

Finally, we sample the Boltzmann distribution of alanine dipeptide in internal coordinates ($d = 60$), as in [30, 32] (Appendix B.4). Its conformations are characterized by the two backbone torsion angles ϕ and ψ, i.e., the rotations around the N-C_α and C_α-C bonds, whose joint distribution is shown as a Ramachandran plot in Figure 6. Besides its dominant conformations, the molecule has a rare metastable conformation, C7ax, which holds only about 0.17% of the probability mass (dashed box). Transitions into the region $\phi > 0$, which contains C7ax, are a classical rare event problem in molecular simulation [10], and even flows trained on molecular dynamics (MD) samples struggle to model this weakly populated mode [32]. Missing it cannot be corrected by reweighting the final samples and biases estimates such as conformational free energies. [30] report that ASBS misses such

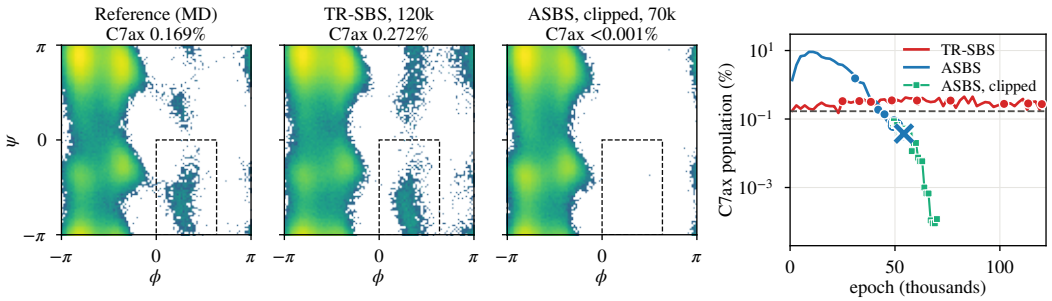


Figure 6: Alanine dipeptide. Left: Ramachandran plots of the reference, of TR-SBS after 120,000 epochs, and of ASBS with clipped corrector targets after 70,000 epochs (Appendix C.6); the dashed box marks C7ax. Right: C7ax population over training (lines: evaluations of 2,000 samples, averaged over 2,000 epochs; dots: 10^7 samples, open dots from ASBS resumed at 48,500 epochs, see Appendix C.6; squares: ASBS with clipped corrector targets, 10^7 samples; cross: divergence of ASBS without clipping; dashed: reference).

Table 1: Alanine dipeptide (10^7 samples): ASBS before (Appendix C.6) and after its divergence, ASBS with clipped corrector targets, and TR-SBS at the end of training. After the divergence, 19% of the ASBS samples have the wrong chirality. Last row: from [30], with a Gaussian prior and another budget (Appendix B.4).

Method	Epoch	$\mathrm{KL}_\phi \downarrow$	$\mathrm{KL}_\psi \downarrow$	$W_2 \downarrow$	C7ax (%)
ASBS (before divergence)	54k	0.0457	0.0276	0.574	0.034
ASBS (diverged)	58.5k	0.2967	0.3786	1.784	0.193
ASBS, clipped	70k	0.0252	0.0406	0.832	<0.001
TR-SBS	120k	0.0144	0.0037	0.246	0.272
ASBS [30]	–	0.02 ± 0.00	0.01 ± 0.00	0.25 ± 0.01	–

low-density conformations and suggest importance weighting in training as a remedy, which TR-SBS uses. To make it possible for both methods to reach C7ax at all, we use the same prior for both, which spreads a small fraction of its samples along ϕ (Appendix B.4). Following [30], we report KL divergences of the torsion marginals and W_2 on (ϕ, ψ), and in addition the C7ax population. Figure 6 shows that ASBS first overpopulates C7ax by a factor of about 50, then drains it to about a fifth of the reference population, and diverges between 54,000 and 54,500 epochs, also when training is resumed from an earlier checkpoint; afterwards, its samples are far from the target, e.g., 19% of them have the wrong chirality (Table 1). With corrector targets clipped in addition to the clipping used for both methods, ASBS does not diverge within 70,000 epochs but drains C7ax further, to 0.0001% (Appendix C.6). TR-SBS, whose updates are bounded by the trust region, keeps the C7ax population close to the reference throughout training and trains stably for 120,000 epochs.

5 Conclusion and limitations

We proposed a trust region approach for computing Schrödinger bridges from an unnormalized target density. Building on the mirror descent interpretation of IPF, it is, to our knowledge, the first method to select the IPF step size adaptively, via a KL constraint on path space that locally controls the variance of the importance weights. On Many-Well, GMM, and alanine dipeptide, it mitigates the mode collapse of ASBS.

The trust region radius ε, however, trades stability for speed (Section 3.2): a small trust region stabilizes the updates but requires more outer iterations, so that TR-SBS can need several times more target-gradient evaluations than ASBS, which converges faster but can collapse to a subset of the modes (Appendix B.7). Moreover, note that our theoretical analysis concerns the idealized iteration: Proposition 3.4 and the variance argument are local and assume exact updates, whereas TR-SBS learns each step from samples, normalizes and clips the weights, and, on alanine dipeptide and LJ-13, bounds γ_i from below and changes the matching anchor (Appendices B.4 and B.5.2). Like ASBS, it is also sensitive to the number of training epochs per update and to the diffusion scale (Appendix C.4). Adapting the radius during training and analyzing the iteration with learned updates are natural next steps.

Acknowledgments and Disclosure of Funding

This work was supported by JST CREST Grant Number JPMJCR24Q5 and JST ASPIRE Grant Number JPMJAP2329, Japan, and by the Horizon Europe Marie Skłodowska-Curie Actions Staff Exchanges project 101131557 (REMODEL).

References

[1] A. Abdolmaleki, R. Lioutikov, N. Lau, L. P. Reis, J. Peters, and G. Neumann. Model-based relative entropy stochastic search. *Advances in Neural Information Processing Systems*, 28, 2015.

[2] S.-i. Amari. *Information geometry and its applications*. Springer, 2016.

[3] P.-C. Aubin-Frankowski, A. Korba, and F. Léger. Mirror descent with relative smoothness in measure spaces, with application to Sinkhorn and EM. *Advances in Neural Information Processing Systems*, 35:17263–17275, 2022.

[4] J. Berner, L. Richter, M. Sendera, J. Rector-Brooks, and N. Malkin. From discrete-time policies to continuous-time diffusion samplers: Asymptotic equivalences and faster training. *Transactions on Machine Learning Research*, 2026.

[5] J. Berner, L. Richter, and K. Ullrich. An optimal control perspective on diffusion-based generative modeling. *Transactions on Machine Learning Research*, 2024.

[6] E. Bernton, J. Heng, A. Doucet, and P. E. Jacob. Schrödinger bridge samplers. *arXiv preprint arXiv:1912.13170*, 2019.

[7] D. Blessing, J. Berner, L. Richter, C. Domingo-Enrich, Y. Du, A. Vahdat, and G. Neumann. Trust region constrained measure transport in path space for stochastic optimal control and inference. In *Advances in Neural Information Processing Systems*, volume 38, pages 165462–165510, 2025.

[8] D. Blessing, J. Berner, L. Richter, and G. Neumann. Underdamped diffusion bridges with applications to sampling. In *International Conference on Learning Representations*, 2025.

[9] D. Blessing, L. Richter, J. Berner, E. Malitskiy, and G. Neumann. Bridge matching sampler: Scalable sampling via generalized fixed-point diffusion matching. In *International Conference on Machine Learning*, pages 8496–8541. PMLR, 2026.

[10] P. G. Bolhuis, C. Dellago, and D. Chandler. Reaction coordinates of biomolecular isomerization. *Proceedings of the National Academy of Sciences*, 97(11):5877–5882, 2000.

[11] J. Chen, L. Richter, J. Berner, D. Blessing, G. Neumann, and A. Anandkumar. Sequential controlled Langevin diffusions. In *International Conference on Learning Representations*, 2025.

[12] T. Chen, G.-H. Liu, and E. A. Theodorou. Likelihood training of Schrödinger bridge using forward-backward SDEs theory. In *International Conference on Learning Representations*, 2022.

[13] J. Choi, Y. Chen, M. Tao, and G.-H. Liu. Non-equilibrium annealed adjoint sampler. In *Advances in Neural Information Processing Systems*, volume 38, pages 90956–90988, 2025.

[14] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust region methods*. SIAM, 2000.

[15] P. Dai Pra. A stochastic control approach to reciprocal diffusion processes. *Applied Mathematics and Optimization*, 23(3):313–329, 1991.

[16] V. De Bortoli, J. Thornton, J. Heng, and A. Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. *Advances in neural information processing systems*, 34:17695–17709, 2021.

[17] C. Domingo-Enrich, M. Drozdzal, B. Karrer, and R. T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *International Conference on Learning Representations*, 2025.

[18] C. Domingo-Enrich, J. Han, B. Amos, J. Bruna, and R. T. Q. Chen. Stochastic optimal control matching. *Advances in Neural Information Processing Systems*, 37:112459–112504, 2024.

[19] R. Fortet. Résolution d'un système d'équations de M. Schrödinger. *Journal de mathématiques pures et appliquées*, 19(1-4):83–105, 1940.

[20] A. Genevay, G. Peyré, and M. Cuturi. Learning generative models with Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pages 1608–1617. PMLR, 2018.

[21] V. L. Ginzburg and L. D. Landau. On the theory of superconductivity. In *On superconductivity and superfluidity: a scientific autobiography*, pages 113–137. Springer, 2009.

[22] G. A. Gottwald and S. Reich. Localized Schrödinger bridge sampler. *Journal of Computational Physics*, 548:114583, 2026.

[23] R. Gruhlke, J. Berner, D. Sommer, and L. Richter. Tensor train diffusion: Leveraging low-rank structures for high-dimensional score-based sampling. In *International Conference on Machine Learning*, pages 36905–36941. PMLR, 2026.

[24] C. Hartmann and L. Richter. Nonasymptotic bounds for suboptimal importance sampling. *SIAM/ASA Journal on Uncertainty Quantification*, 12(2):309–346, 2024.

[25] C. Hartmann, L. Richter, C. Schütte, and W. Zhang. Variational characterization of free energy: Theory and algorithms. *Entropy*, 19(11):626, 2017.

[26] A. J. Havens, B. K. Miller, B. Yan, C. Domingo-Enrich, A. Sriram, D. S. Levine, B. M. Wood, B. Hu, B. Amos, B. Karrer, X. Fu, G.-H. Liu, and R. T. Q. Chen. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. In *International Conference on Machine Learning*, pages 22204–22237. PMLR, 2025.

[27] J. Huang, Y. Jiao, L. Kang, X. Liao, J. Liu, and Y. Liu. Schrödinger-Föllmer sampler. *IEEE Transactions on Information Theory*, 71(2):1283–1299, 2025.

[28] M. R. Karimi, Y.-P. Hsieh, and A. Krause. Sinkhorn flow as mirror flow: A continuous-time framework for generalizing the Sinkhorn algorithm. In *International Conference on Artificial Intelligence and Statistics*, pages 4186–4194. PMLR, 2024.

[29] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014.

[30] G.-H. Liu, J. Choi, Y. Chen, B. K. Miller, and R. T. Q. Chen. Adjoint Schrödinger bridge sampler. In *Advances in Neural Information Processing Systems*, volume 38, pages 15673–15708, 2025.

[31] G.-H. Liu, Y. Lipman, M. Nickel, B. Karrer, E. A. Theodorou, and R. T. Q. Chen. Generalized Schrödinger bridge matching. In *International Conference on Learning Representations*, 2024.

[32] L. I. Midgley, V. Stimper, G. N. C. Simm, B. Schölkopf, and J. M. Hernández-Lobato. Flow annealed importance sampling bootstrap. In *International Conference on Learning Representations*, 2023.

[33] J. Nam, B. Máté, A. P. Toshev, M. Kaniselvan, R. Gómez-Bombarelli, R. T. Q. Chen, B. Wood, G.-H. Liu, and B. K. Miller. Enhancing diffusion-based sampling with molecular collective variables. In *International Conference on Learning Representations*, 2026.

[34] N. Nüsken and L. Richter. Solving high-dimensional Hamilton-Jacobi-Bellman PDEs using neural networks: perspectives from the theory of controlled diffusions and measures on path space. *Partial Differential Equations and Applications*, 2(4):48, 2021.

[35] F. Otto, P. Becker, N. A. Vien, H. C. Ziesche, and G. Neumann. Differentiable trust region layers for deep reinforcement learning. In *International Conference on Learning Representations*, 2021.

[36] R. Ouyang, L. Grenioux, and J. M. Hernández-Lobato. A diffusive classification loss for learning energy-based generative models. In *International Conference on Machine Learning*, pages 95197–95252. PMLR, 2026.

[37] S. Peluchetti. Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling. *Journal of Machine Learning Research*, 24(374):1–51, 2023.

[38] S. Peluchetti. Non-denoising forward-time diffusions. *arXiv preprint arXiv:2312.14589*, 2023.

[39] J. Peters, K. Mülling, and Y. Altun. Relative entropy policy search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 24, pages 1607–1612, 2010.

[40] L. Richter and J. Berner. Improved sampling via learned diffusions. In *International Conference on Learning Representations*, 2024.

[41] B. Rosenstein and D. Li. *Ginzburg–Landau Theory of Condensates*. Cambridge University Press, 2021.

[42] L. Rüschendorf. Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, 23(4):1160–1174, 1995.

[43] V. G. Satorras, E. Hoogeboom, and M. Welling. E(n) equivariant graph neural networks. In *International Conference on Machine Learning*, 2021.

[44] E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse*, pages 144–153, 1931.

[45] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz. Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897. PMLR, 2015.

[46] Y. Shi, V. De Bortoli, A. Campbell, and A. Doucet. Diffusion Schrödinger bridge matching. *Advances in Neural Information Processing Systems*, 36:62183–62223, 2023.

[47] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[48] J. Sun, J. Berner, L. Richter, M. Zeinhofer, J. Müller, K. Azizzadenesheli, and A. Anandkumar. Dynamical measure transport and neural PDE solvers for sampling. *arXiv preprint arXiv:2407.07873*, 2024.

[49] D. Thalmeier, H. J. Kappen, S. Totaro, and V. Gómez. Adaptive smoothing for path integral control. *Journal of Machine Learning Research*, 21(191):1–37, 2020.

[50] F. Vargas. Machine-learning approaches for the empirical Schrödinger bridge problem. Technical report, University of Cambridge, Computer Laboratory, 2021.

[51] F. Vargas, W. Grathwohl, and A. Doucet. Denoising diffusion samplers. In *International Conference on Learning Representations*, 2023.

[52] F. Vargas, S. Padhy, D. Blessing, and N. Nüsken. Transport meets variational inference: Controlled Monte Carlo diffusions. In *The Twelfth International Conference on Learning Representations*, 2024.

[53] F. Vargas, P. Thodoroff, A. Lamacraft, and N. Lawrence. Solving Schrödinger bridges via maximum likelihood. *Entropy*, 23(9):1134, 2021.

[54] G. Wang, Y. Jiao, Q. Xu, Y. Wang, and C. Yang. Deep generative learning via Schrödinger bridge. In *International conference on machine learning*, pages 10794–10804. PMLR, 2021.

[55] H. Wu, J. Köhler, and F. Noé. Stochastic normalizing flows. *Advances in neural information processing systems*, 33:5933–5944, 2020.

[56] Q. Zhang and Y. Chen. Path integral sampler: A stochastic control approach for sampling. In *International Conference on Learning Representations*, 2022.

Contents of the appendix

A Proofs and additional propositions **16**

A.1 Notation . . . 16
A.2 Auxiliary results . . . 16
A.3 Proofs of Propositions 3.1 and 3.2 . . . 18
A.4 Stochastic optimal control formulation for SB based on the reciprocal property . . . 23
A.5 Proof of the main results in Section 3.1 . . . 26
A.6 TR-IPF and its convergence . . . 27
A.7 Remarks on the variance of importance weights . . . 30
A.8 Relation to trust region methods on path space [7] . . . 31
A.9 Relation to ASBS and limiting cases . . . 31
A.10 Anchoring the matching target at the regressed time . . . 31

B Experiment details **32**

B.1 Gaussian mixture model (GMM) . . . 33
B.2 Many-Well potential (MW) . . . 34
B.3 Ginzburg-Landau (ϕ^4 scalar field theory) model . . . 35
B.4 Alanine dipeptide . . . 35
B.5 Implementation details and hyperparameters . . . 37
B.6 PIS and DIS baselines . . . 40
B.7 Computational cost . . . 41

C Additional experimental results **42**

C.1 Additional metrics on GMM . . . 42
C.2 Visualization on GMM ($d = 10$, $K = 40$) . . . 42
C.3 Visualization on high-barrier Many-Well ($d = 8, m = 5, \delta = 8$) . . . 42
C.4 Ablation study . . . 43
C.5 ϕ^4 scalar field theory . . . 45
C.6 Alanine dipeptide . . . 46
C.7 Lennard-Jones cluster (LJ-13) . . . 47

A Proofs and additional propositions

In this section we provide the proofs for statements in the main part of the article and provide additional propositions.

A.1 Notation

We denote with $\vec{\mathbb{P}}^u \in \mathcal{P}$ a path space measure related to a forward stochastic differential equation with control function $u \in \mathcal{U}$, where $\mathcal{P}$ is the set of probability measures on the space of continuous functions $C([0,T],\mathbb{R}^d)$. Likewise, $\overleftarrow{\mathbb{P}}^v \in \mathcal{P}$ corresponds to a reverse-time SDE with control $v \in \mathcal{U}$. We denote by $\mathbb{P}_t$ the time marginal at time $t \in [0,T]$, by $\mathbb{P}_{0,T}$ the joint law of the endpoints (X_0, X_T), and by $\mathbb{P}_{[t,T]}$ the time restriction to the interval $[t,T]$. The set $\mathcal{U} \subset C(\mathbb{R}^d \times [0,T], \mathbb{R}^d)$ denotes the space of admissible controls. With ∇ we denote the gradient w.r.t. the spatial variable x. In the proofs we allow for a base drift f, i.e., $\vec{\mathbb{P}}^u$ is the law of $\mathrm{d}X_t = (f + \sigma u)(X_t, t)\,\mathrm{d}t + \sigma(t)\vec{\mathrm{d}}W_t$ and $\overleftarrow{\mathbb{P}}^v$ the law of $\mathrm{d}Y_t = (f - \sigma v)(Y_t, t)\,\mathrm{d}t + \sigma(t)\overleftarrow{\mathrm{d}}W_t$; the main text corresponds to $f = 0$. The diffusion coefficient $\sigma(t)$ is invertible and does not depend on the state. We omit function arguments, such as the time argument of σ, whenever they are clear from the context. $\mathbb{P}^{\mathrm{ref}} = \vec{\mathbb{P}}^0$ denotes the uncontrolled reference process and $\mathbb{P}^{\mathrm{ref}}_{t|0,T}$ its bridge, i.e., the law of X_t given (X_0, X_T). In conditional expectations given $X_t = x$, such as the cost-to-go $\mathcal{L}_i^{\mathrm{TRC}}(u, \lambda; x, t)$ and the value functions, the forward path measures $\vec{\mathbb{P}}^u_{[t,T]}$ and $\vec{\mathbb{P}}^{u_i}_{[t,T]}$ denote the laws of the forward SDE on $[t,T]$ with the common initial law $\vec{\mathbb{P}}^{u_i}_t$, so that $\frac{\mathrm{d}\vec{\mathbb{P}}^u_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}$ has no boundary term at time t; we write $p_t := \vec{\mathbb{P}}^{u_i}_t$.

Expectations. $\mathbb{E}_{\mathbb{P}}$ denotes the expectation of a path functional $X \mapsto G(X)$ under $X \sim \mathbb{P}$, and $\mathbb{E}_{\mathbb{P}}[\,\cdot \mid X_t = x]$ the conditional expectation given the state at time t. For endpoint-based objectives such as (25) and (26), $\mathbb{E}_{\mathbb{P}_{0,T}\mathbb{P}^{\mathrm{ref}}_{t|0,T}}$ denotes sampling $(X_0, X_T) \sim \mathbb{P}_{0,T}$, $t \sim \mathcal{U}(0,T)$, and $X_t \sim \mathbb{P}^{\mathrm{ref}}_{t|0,T}(\cdot \mid X_0, X_T)$, with the time integral normalized accordingly. Unless stated otherwise, a Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(X)$ is evaluated along the trajectory X that is integrated in the surrounding expectation.

A.2 Auxiliary results

We collect the auxiliary identities used in the proofs below, stated in our notation with proofs or proof sketches. They correspond to results from [18, 28, 30, 40].

Lemma A.1 (Terminal density ratio)**.** *Let $\mathbb{P}, \mathbb{Q} \in \mathcal{P}$ with $\mathbb{Q}_T \ll \mathbb{P}_T$ and $\mathbb{Q}(\cdot \mid X_T = y) = \mathbb{P}(\cdot \mid X_T = y)$ for $\mathbb{P}_T$-a.e. y. Then $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(X) = \frac{\mathrm{d}\mathbb{Q}_T}{\mathrm{d}\mathbb{P}_T}(X_T)$. In particular, if v_i is an exact backward IPF half-step for u_i, i.e., $\overleftarrow{\mathbb{P}}^{v_i} \in \arg\min\{D_{\mathrm{KL}}(\mathbb{Q}|\vec{\mathbb{P}}^{u_i}) : \mathbb{Q}_T = p_{\mathrm{target}}\}$, then*

$$\frac{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X) = \frac{p_{\mathrm{target}}}{\vec{\mathbb{P}}^{u_i}_T}(X_T). \tag{29}$$

Proof. For bounded measurable G, disintegration with respect to X_T gives $\mathbb{E}_{\mathbb{Q}}[G] = \int_{\mathbb{R}^d} \mathbb{E}_{\mathbb{P}}[G \mid X_T = y]\,\frac{\mathrm{d}\mathbb{Q}_T}{\mathrm{d}\mathbb{P}_T}(y)\,\mathbb{P}_T(\mathrm{d}y) = \mathbb{E}_{\mathbb{P}}\big[G\,\frac{\mathrm{d}\mathbb{Q}_T}{\mathrm{d}\mathbb{P}_T}(X_T)\big]$, which is the first claim. For the second claim, the chain rule of the KL divergence yields $D_{\mathrm{KL}}(\mathbb{Q}|\vec{\mathbb{P}}^{u_i}) = D_{\mathrm{KL}}(\mathbb{Q}_T|\vec{\mathbb{P}}^{u_i}_T) + \mathbb{E}_{\mathbb{Q}_T}\big[D_{\mathrm{KL}}(\mathbb{Q}(\cdot|X_T) \,|\, \vec{\mathbb{P}}^{u_i}(\cdot|X_T))\big]$. Under the constraint $\mathbb{Q}_T = p_{\mathrm{target}}$, the first term is fixed and the second term vanishes if and only if $\mathbb{Q}(\cdot|X_T) = \vec{\mathbb{P}}^{u_i}(\cdot|X_T)$. This minimizer is the law of the reverse-time SDE with the backward drift of the time reversal of X^{u_i}, started at p_{target}, and is therefore of the form $\overleftarrow{\mathbb{P}}^v$. The first claim then gives (29). By [52, Corollary E.2], the forward KL version of the half-step used in Section 3 has the same minimizer. □

Lemma A.2 (Forward-backward path likelihood ratio, [40, Proposition 2.3 and Remark A.1])**.** *Let $X \sim \vec{\mathbb{P}}^u$ and $t \in [0,T]$. Then, $\vec{\mathbb{P}}^u$-almost surely,*

$$\log \frac{\mathrm{d}\overleftarrow{\mathbb{P}}^v_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^u_{[t,T]}}(X) = \log \frac{p_{\mathrm{target}}(X_T)}{\vec{\mathbb{P}}^u_t(X_t)} - \int_t^T \sigma^{-1}(f+\sigma u)\cdot \sigma^{-1}\vec{\mathrm{d}}X_s + \frac{1}{2}\int_t^T \|\sigma^{-1}(f+\sigma u)\|^2 \mathrm{d}s + \int_t^T \sigma^{-1}(f-\sigma v)\cdot \sigma^{-1}\overleftarrow{\mathrm{d}}X_s - \frac{1}{2}\int_t^T \|\sigma^{-1}(f-\sigma v)\|^2 \mathrm{d}s, \tag{30}$$

where the integrands are evaluated at (X_s, s). For $f = 0$ this reads

$$\log \frac{\mathrm{d}\overleftarrow{\mathbb{P}}^v_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^u_{[t,T]}}(X) = \log \frac{p_{\mathrm{target}}(X_T)}{\vec{\mathbb{P}}^u_t(X_t)} + \frac{1}{2}\int_t^T \big(\|u\|^2 - \|v\|^2\big)\mathrm{d}s - \int_t^T u\cdot\sigma^{-1}\vec{\mathrm{d}}X_s - \int_t^T v\cdot\sigma^{-1}\overleftarrow{\mathrm{d}}X_s. \tag{31}$$

Moreover, writing $X = X^u$ with driving Brownian motion W,

$$\log \frac{\mathrm{d}\vec{\mathbb{P}}^u_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^v_{[t,T]}}(X^u) = \log \frac{\vec{\mathbb{P}}^u_t(X^u_t)}{p_{\mathrm{target}}(X^u_T)} + \int_t^T \Big(\frac{1}{2}\|u+v\|^2 + \nabla\cdot(\sigma v - f)\Big)\mathrm{d}s + \int_t^T (u+v)\cdot \mathrm{d}W_s. \tag{32}$$

Proof sketch. Discretizing both processes with the Euler-Maruyama scheme, the ratio of the backward and forward Gaussian transition densities yields Riemann-Itô sums with the forward drift at the left and the backward drift at the right endpoints, which converge to (30); see [4, Lemma B.7]. For (32), convert the backward integral into a forward one, $\int_t^T b\cdot\sigma^{-1}\overleftarrow{\mathrm{d}}X_s = \int_t^T b\cdot\sigma^{-1}\vec{\mathrm{d}}X_s + \int_t^T \nabla\cdot(\sigma b)\,\mathrm{d}s$ with $b = \sigma^{-1}(f-\sigma v)$, and insert $\sigma^{-1}\vec{\mathrm{d}}X^u_s = \sigma^{-1}(f+\sigma u)\,\mathrm{d}s + \mathrm{d}W_s$. □

Lemma A.3 (Conditional expectation matching)**.** *Let Y be a random vector, Z a random variable, and $\zeta(Z) > 0$ a weight with $\mathbb{E}[\zeta(Z)\|Y\|^2] < \infty$. For every measurable u with $\mathbb{E}[\zeta(Z)\|u(Z)\|^2] < \infty$,*

$$\mathbb{E}\big[\zeta(Z)\|u(Z) - Y\|^2\big] = \mathbb{E}\big[\zeta(Z)\|u(Z) - \mathbb{E}[Y|Z]\|^2\big] + \mathbb{E}\big[\zeta(Z)\|Y - \mathbb{E}[Y|Z]\|^2\big]. \tag{33}$$

Hence both objectives differ by a constant independent of u, and both are minimized by $u(Z) = \mathbb{E}[Y|Z]$.

Proof. Expanding $u(Z) - Y = (u(Z) - \mathbb{E}[Y|Z]) + (\mathbb{E}[Y|Z] - Y)$, the cross term $2\mathbb{E}\big[\zeta(Z)(u(Z) - \mathbb{E}[Y|Z])\cdot\mathbb{E}[\mathbb{E}[Y|Z] - Y \mid Z]\big]$ vanishes by the tower property. □

Lemma A.4 (Backward matching loss and the IPF half-step, cf. [28, 30])**.** *Let $\mathbb{P} := \vec{\mathbb{P}}^{u_{i+1}}$ be reciprocal (cf. Proposition A.7), i.e., $\mathbb{P}$ has the bridges $\mathbb{P}^{\mathrm{ref}}_{\cdot|0,T}$ and the endpoint density $\mathbb{P}_{0,T}(y,z) = \widehat{\varphi}_0(y)\, p^{\mathrm{ref}}_{T|0}(z|y)\,\varphi_T(z)$, and set $\widehat{\varphi}_t(x) := \int_{\mathbb{R}^d} p^{\mathrm{ref}}_{t|0}(x|y)\,\widehat{\varphi}_0(y)\,\mathrm{d}y$. Then the minimizer of the backward loss* (26) *over all measurable v is $v^*(x,t) = \sigma^\top \nabla \log \widehat{\varphi}_t(x)$. The reverse-time SDE* (6) *with v^* has the same backward transition kernels as the time reversal of $\mathbb{P}$; hence $\overleftarrow{\mathbb{P}}^{v^*}(\cdot|X_T) = \mathbb{P}(\cdot|X_T)$, i.e., v^* is the exact backward IPF half-step of Lemma A.1.*

Proof sketch. Since $\mathbb{P}$ is reciprocal, $\mathbb{P}_t = \widehat{\varphi}_t\varphi_t$ with $\varphi_t(x) := \int_{\mathbb{R}^d} p^{\mathrm{ref}}_{T|t}(z|x)\,\varphi_T(z)\,\mathrm{d}z$, and the conditional law of X_0 given $X_t = x$ has density proportional to $\widehat{\varphi}_0(y)\, p^{\mathrm{ref}}_{t|0}(x|y)$. By Lemma A.3, the minimizer of (26) is $\sigma^\top \mathbb{E}_{\mathbb{P}}[\nabla_x \log p^{\mathrm{ref}}_{t|0}(x|X_0) \mid X_t = x] = \sigma^\top \nabla \log \widehat{\varphi}_t(x)$. By Nelson's relation, the time reversal of $\mathbb{P}$ has the backward drift $f - \sigma\sigma^\top \nabla \log \widehat{\varphi}_t = f - \sigma v^*$, so it shares the backward transition kernels and the conditional law given X_T with $\overleftarrow{\mathbb{P}}^{v^*}$; cf. [28, 30]. □

Lemma A.5 (One-step inequality of KL mirror descent)**.** *Let* $F(\mathbb{P}) \coloneqq D_{\mathrm{KL}}(\mathbb{P}_T|p_{\mathrm{target}})$ *and* $\mathcal{A} \coloneqq \{\mathbb{P} \in \mathcal{P} : \mathbb{P}_0 = p_{\mathrm{prior}}\}$*. For* $\mathbb{P}_i \in \mathcal{A}$ *and* $\gamma \in (0, 1]$*, let* $\mathbb{P}_{i+1} \in \mathcal{A}$ *be defined by*

$$\frac{\mathrm{d}\mathbb{P}_{i+1}}{\mathrm{d}\mathbb{P}_i}(X) = \frac{e^{-\gamma g(X)}}{\mathbb{E}_{\mathbb{P}_i}[e^{-\gamma g} \mid X_0]}, \qquad g(X) \coloneqq \log \frac{\mathbb{P}_{i,T}}{p_{\mathrm{target}}}(X_T). \tag{34}$$

Then, for every $\mathbb{Q} \in \mathcal{A}$ *for which the quantities below are finite,*

$$\gamma\big(F(\mathbb{P}_{i+1}) - F(\mathbb{Q})\big) \le D_{\mathrm{KL}}(\mathbb{Q}|\mathbb{P}_i) - D_{\mathrm{KL}}(\mathbb{Q}|\mathbb{P}_{i+1}). \tag{35}$$

In particular, choosing $\mathbb{Q} = \mathbb{P}_i$ *shows* $F(\mathbb{P}_{i+1}) \le F(\mathbb{P}_i)$*.*

Proof. The function g is the first variation $\delta F(\mathbb{P}_i)$ up to an additive constant; we write $\langle g, \mathbb{P} - \mathbb{Q}\rangle \coloneqq \mathbb{E}_{\mathbb{P}}[g] - \mathbb{E}_{\mathbb{Q}}[g]$. *Three-point identity.* Let $Z(x) \coloneqq \mathbb{E}_{\mathbb{P}_i}[e^{-\gamma g} \mid X_0 = x]$. Since all measures in $\mathcal{A}$ have the initial marginal p_{prior}, for every $\mathbb{Q} \in \mathcal{A}$ we have $D_{\mathrm{KL}}(\mathbb{Q}|\mathbb{P}_i) - D_{\mathrm{KL}}(\mathbb{Q}|\mathbb{P}_{i+1}) = \mathbb{E}_{\mathbb{Q}}\big[\log \frac{\mathrm{d}\mathbb{P}_{i+1}}{\mathrm{d}\mathbb{P}_i}\big] = -\gamma\mathbb{E}_{\mathbb{Q}}[g] - \mathbb{E}_{p_{\mathrm{prior}}}[\log Z]$. For $\mathbb{Q} = \mathbb{P}_{i+1}$, this gives $D_{\mathrm{KL}}(\mathbb{P}_{i+1}|\mathbb{P}_i) = -\gamma\mathbb{E}_{\mathbb{P}_{i+1}}[g] - \mathbb{E}_{p_{\mathrm{prior}}}[\log Z]$. Subtracting both identities yields

$$\gamma\langle g, \mathbb{P}_{i+1} - \mathbb{Q}\rangle = D_{\mathrm{KL}}(\mathbb{Q}|\mathbb{P}_i) - D_{\mathrm{KL}}(\mathbb{Q}|\mathbb{P}_{i+1}) - D_{\mathrm{KL}}(\mathbb{P}_{i+1}|\mathbb{P}_i). \tag{36}$$

Relative smoothness and convexity. For every $\mathbb{P}$ with $F(\mathbb{P}) < \infty$, a direct computation gives $F(\mathbb{P}) - F(\mathbb{P}_i) - \langle g, \mathbb{P} - \mathbb{P}_i\rangle = D_{\mathrm{KL}}(\mathbb{P}_T|\mathbb{P}_{i,T})$, which lies in $[0, D_{\mathrm{KL}}(\mathbb{P}|\mathbb{P}_i)]$ by the data processing inequality. Applying the upper bound to $\mathbb{P} = \mathbb{P}_{i+1}$ and the lower bound to $\mathbb{P} = \mathbb{Q}$ and subtracting, we obtain $F(\mathbb{P}_{i+1}) - F(\mathbb{Q}) \le \langle g, \mathbb{P}_{i+1} - \mathbb{Q}\rangle + D_{\mathrm{KL}}(\mathbb{P}_{i+1}|\mathbb{P}_i)$. Multiplying by γ and inserting the three-point identity,

$$\gamma\big(F(\mathbb{P}_{i+1}) - F(\mathbb{Q})\big) \le D_{\mathrm{KL}}(\mathbb{Q}|\mathbb{P}_i) - D_{\mathrm{KL}}(\mathbb{Q}|\mathbb{P}_{i+1}) - (1-\gamma)D_{\mathrm{KL}}(\mathbb{P}_{i+1}|\mathbb{P}_i), \tag{37}$$

and the last term is non-positive since $\gamma \le 1$. For $\mathbb{Q} = \mathbb{P}_i$, the right-hand side of (35) equals $-D_{\mathrm{KL}}(\mathbb{P}_i|\mathbb{P}_{i+1}) \le 0$. □

A.3 Proofs of Propositions 3.1 and 3.2

Proof of Proposition 3.1. First note that, for a fixed Lagrange multiplier $\lambda \ge 0$, it holds

$$u_{i+1}^{\lambda} = \operatorname*{arg\,min}_{u \in \mathcal{U}} \mathcal{L}_i^{\mathrm{TR}}(u, \lambda) = \operatorname*{arg\,min}_{u \in \mathcal{U}} \left\{ \frac{1}{1+\lambda} \mathcal{L}_i^{\mathrm{TR}}(u, \lambda) \right\}. \tag{38}$$

Then, we can compute

$$\begin{aligned}
u_{i+1}^{\lambda} &= \operatorname*{arg\,min}_{u \in \mathcal{U}} \left\{ \frac{1}{1+\lambda} D_{\mathrm{KL}}(\vec{\mathbb{P}}^u|\overleftarrow{\mathbb{P}}^{v_i}) + \frac{\lambda}{1+\lambda} D_{\mathrm{KL}}(\vec{\mathbb{P}}^u|\vec{\mathbb{P}}^{u_i}) \right\} && \text{(39a)} \\
&= \operatorname*{arg\,min}_{u \in \mathcal{U}} \mathbb{E}\left[\frac{1}{1+\lambda} \log \frac{\mathrm{d}\vec{\mathbb{P}}^u}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}(X^u) + \frac{\lambda}{1+\lambda} \log \frac{\mathrm{d}\vec{\mathbb{P}}^u}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X^u) \right] && \text{(39b)} \\
&= \operatorname*{arg\,min}_{u \in \mathcal{U}} \mathbb{E}\left[\log \left(\left(\frac{\mathrm{d}\vec{\mathbb{P}}^u}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}(X^u) \right)^{\frac{1}{1+\lambda}} \left(\frac{\mathrm{d}\vec{\mathbb{P}}^u}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X^u) \right)^{\frac{\lambda}{1+\lambda}} \right) \right] && \text{(39c)} \\
&= \operatorname*{arg\,min}_{u \in \mathcal{U}} \mathbb{E}\left[\log \frac{\mathrm{d}\vec{\mathbb{P}}^u}{\mathrm{d}\left((\overleftarrow{\mathbb{P}}^{v_i})^{\frac{1}{1+\lambda}} (\vec{\mathbb{P}}^{u_i})^{\frac{\lambda}{1+\lambda}} \right)}(X^u) \right] && \text{(39d)} \\
&= \operatorname*{arg\,min}_{u \in \mathcal{U}} D_{\mathrm{KL}}\left(\vec{\mathbb{P}}^u \middle| (\overleftarrow{\mathbb{P}}^{v_i})^{\frac{1}{1+\lambda}} (\vec{\mathbb{P}}^{u_i})^{\frac{\lambda}{1+\lambda}} \right). && \text{(39e)}
\end{aligned}$$

Let $\mathbb{G}$ denote the normalized geometric average $\mathbb{G} \propto (\overleftarrow{\mathbb{P}}^{v_i})^{\frac{1}{1+\lambda}} (\vec{\mathbb{P}}^{u_i})^{\frac{\lambda}{1+\lambda}}$, i.e., $\frac{\mathrm{d}\mathbb{G}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} \propto \big(\frac{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}\big)^{\frac{1}{1+\lambda}}$. The minimization is over path measures with $\vec{\mathbb{P}}_0^u = p_{\mathrm{prior}}$. By the chain rule of the KL divergence with respect to X_0,

$$D_{\mathrm{KL}}(\vec{\mathbb{P}}^u|\mathbb{G}) = D_{\mathrm{KL}}(p_{\mathrm{prior}}|\mathbb{G}_0) + \mathbb{E}_{X_0 \sim p_{\mathrm{prior}}}\big[D_{\mathrm{KL}}(\vec{\mathbb{P}}^u(\cdot|X_0) \,|\, \mathbb{G}(\cdot|X_0))\big], \tag{40}$$

where the first term does not depend on u. Hence the minimizer satisfies $\vec{\mathbb{P}}^{u^\lambda_{i+1}}(\cdot|X_0) = \mathbb{G}(\cdot|X_0)$, i.e., it is the geometric average with its initial marginal replaced by p_{prior}. Since $\vec{\mathbb{P}}^{u_i}_0 = p_{\text{prior}}$, this gives

$$\frac{\mathrm{d}\vec{\mathbb{P}}^{u^\lambda_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X) = \frac{\mathrm{d}\mathbb{G}(\cdot|X_0)}{\mathrm{d}\vec{\mathbb{P}}^{u_i}(\cdot|X_0)}(X) = \frac{\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X)\right)^{\frac{1}{1+\lambda}}}{\mathbb{E}_{\vec{\mathbb{P}}^{u_i}}\left[\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}\right)^{\frac{1}{1+\lambda}} \middle| X_0\right]}, \tag{41}$$

which is the first identity of the statement together with its normalization. That this measure is induced by a control $u^\lambda_{i+1} \in \mathcal{U}$ follows from Proposition 3.2. The second equality follows from Lemma A.1, which relates the Radon-Nikodym derivative of path measures to the likelihood ratio of their terminal marginals. This completes the proof. □

Proof of Proposition 3.2. We remark that while the main text considers SDEs without base drift, here we establish the result for a more general case with drift $(f + \sigma u)$. The setting in the main text is recovered by setting $f = 0$.

The proof consists of two parts. First, we establish a lower bound for the objective functional $\mathcal{L}^{\mathrm{TRC}}_i$ using Jensen's inequality. Second, we explicitly construct a control $u^*(= u^\lambda_{i+1})$ that attains this lower bound, thereby proving that the lower bound coincides with the value function defined as $V^\lambda_{i+1}(x,t) = \inf_u \mathcal{L}^{\mathrm{TRC}}_i(u,\lambda;x,t)$.

Part 1: Lower bound via Jensen's inequality. For any control $u \in \mathcal{U}$, we expand the Lagrangian defined in Equation (16). Using the chain rule for Radon-Nikodym derivatives, we have:

$$\mathcal{L}^{\mathrm{TRC}}_i(u,\lambda;x,t) = \mathbb{E}_{\vec{\mathbb{P}}^u}\left[\log\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}}\frac{\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}\right) + \lambda\log\frac{\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}\middle| X^u_t = x\right] \tag{42a}$$

$$= \mathbb{E}_{\vec{\mathbb{P}}^u}\left[-(1+\lambda)\log\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}}\right)^{-\frac{1}{1+\lambda}} - (1+\lambda)\log\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]}}\middle| X^u_t = x\right]. \tag{42b}$$

Applying Jensen's inequality to the strictly convex function $\phi(y) = -\log y$, we obtain the following lower bound:

$$\mathcal{L}^{\mathrm{TRC}}_i(u,\lambda;x,t) \geq -(1+\lambda)\log\mathbb{E}_{\vec{\mathbb{P}}^u}\left[\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}}\right)^{-\frac{1}{1+\lambda}}\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]}}\middle| X^u_t = x\right] \tag{43a}$$

$$= -(1+\lambda)\log\mathbb{E}_{\vec{\mathbb{P}}^{u_i}}\left[\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}}\right)^{-\frac{1}{1+\lambda}}\middle| X^{u_i}_t = x\right]. \tag{43b}$$

Note that the right-hand side is independent of u. Thus, $V^\lambda_{i+1}(x,t)$ is bounded from below by this quantity.

Part 2: Attainability via martingale comparison. We now construct the optimal control u^* that minimizes the objective. The lower bound in Equation (43a) is attained if and only if Jensen's inequality becomes an equality. This requires the random variable inside the logarithm to be constant almost surely with respect to the optimal path measure $\vec{\mathbb{P}}^{u^*}$.

Let us denote this state-dependent constant by $C(x,t)$. By definition, at the optimum, the value of the objective function becomes:

$$\inf_{u\in\mathcal{U}}\mathcal{L}^{\mathrm{TRC}}_i(u,\lambda;x,t) = -(1+\lambda)\log\mathbb{E}_{\vec{\mathbb{P}}^{u^*}}[C(x,t)] = -(1+\lambda)\log C(x,t). \tag{44}$$

Since the value function $V^\lambda_{i+1}(x,t)$ is defined as this minimal cost, we establish the fundamental relationship:

$$V^\lambda_{i+1}(x,t) = -(1+\lambda)\log C(x,t) \quad\Longleftrightarrow\quad \log C(x,t) = -\frac{V^\lambda_{i+1}(x,t)}{1+\lambda}. \tag{45}$$

Now, writing the condition "inside of log equals C" explicitly:

$$\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}}\right)^{-\frac{1}{1+\lambda}} \frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u^*}_{[t,T]}} = C(x,t). \tag{46}$$

Taking the logarithm and rearranging terms:

$$\log \frac{\mathrm{d}\vec{\mathbb{P}}^{u^*}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}} = -\frac{1}{1+\lambda} \log \frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}} - \log C(x,t). \tag{47}$$

We identify u^* by matching the martingale parts (coefficients of the Brownian motion $\mathrm{d}W_s$) on both sides of (46), which holds for every lower limit $s \in [t,T]$ in place of t (with x replaced by X_s), so that both sides are processes in s. For the first term on the RHS, we recall the explicit representation of the log-likelihood ratio from Lemma A.2, see (32):

$$\log \frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}} = \log \frac{p_t(X_t^{u_i})}{p_{\mathrm{target}}(X_T^{u_i})} + \int_t^T \left(\frac{1}{2}\|u_i + v_i\|^2 + \nabla \cdot (\sigma v_i - f)\right) \mathrm{d}s + \int_t^T (u_i + v_i) \cdot \mathrm{d}W_s, \tag{48}$$

where $p_s \coloneqq \vec{\mathbb{P}}^{u_i}_s$. As a process in the lower limit s, the stochastic integral contributes the martingale part $-(u_i + v_i) \cdot \mathrm{d}W_s$, and the boundary term $\log p_s(X_s)$ contributes $\sigma^\top \nabla \log p_s \cdot \mathrm{d}W_s$. Similarly, by the Girsanov theorem, the LHS is given by:

$$\log \frac{\mathrm{d}\vec{\mathbb{P}}^{u^*}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}} = \frac{1}{2}\int_t^T \|u^* - u_i\|^2 \mathrm{d}s + \int_t^T (u^* - u_i) \cdot \mathrm{d}W_s. \tag{49}$$

For the normalizing constant term $\log C(x,t)$, applying Itô's formula to $V_{i+1}^\lambda(X_s, s)$ under the dynamics $\mathrm{d}X_s = (f + \sigma u^*)\mathrm{d}s + \sigma \mathrm{d}W_s$:

$$\mathrm{d}(\log C) = -\frac{1}{1+\lambda}\Bigg[\underbrace{\left(\partial_t V_{i+1}^\lambda + (f + \sigma u^*) \cdot \nabla V_{i+1}^\lambda + \frac{1}{2}\operatorname{Tr}(\sigma\sigma^\top \nabla^2 V_{i+1}^\lambda)\right)}_{\text{drift part}} \mathrm{d}s$$
$$+ \underbrace{\nabla V_{i+1}^\lambda \cdot \sigma \mathrm{d}W_s}_{\text{martingale part}}\Bigg]. \tag{50}$$

Equating the coefficients of the martingale part ($\mathrm{d}W_s$) across the equation yields:

$$(u^* - u_i) = -\frac{1}{1+\lambda}(u_i + v_i) - \frac{1}{1+\lambda}\sigma^\top \nabla\big(V_{i+1}^\lambda - \log p_t\big). \tag{51}$$

Solving for u^*, we obtain the optimal control law:

$$u^* = \frac{\lambda}{1+\lambda}u_i - \frac{1}{1+\lambda}v_i - \frac{1}{1+\lambda}\sigma^\top \nabla\big(V_{i+1}^\lambda - \log p_t\big). \tag{52}$$

By construction, this control u^* eliminates the martingale part of the log-likelihood ratio in (46), rendering the term inside the expectation in (43a) deterministic (equal to $C(x,t)$). Consequently, we have

$$V_{i+1}^\lambda(x,t) = -(1+\lambda)\log \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}\left[\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}}\right)^{-\frac{1}{1+\lambda}} \,\middle|\, X_t^{u_i} = x\right] \tag{53}$$

□

Remark A.6 (Relation to [28] and the HJB equation). The explicit formulas derived in Proposition 3.2 are equivalent to Theorem 4.2 of [28] when v_i is the exact backward half-step. The main difference lies in the derivation: whereas our proof is probabilistic, relying on Jensen's inequality and martingale matching, [28] obtain the same result analytically via the associated Hamilton-Jacobi-Bellman (HJB)

equation. For completeness, we present below an alternative proof in our notation following this approach.

More precisely, their method applies a logarithmic (Fleming or Cole-Hopf) transformation [28, Proposition C.1], which in our notation reads $V_{i+1}^{\lambda} = -(1+\lambda)\log\Psi$, to linearize the nonlinear HJB equation associated with $\mathcal{L}_i^{\mathrm{TRC}}$. The resulting function Ψ admits a Feynman-Kac representation, which, after a Girsanov change of measure, coincides with our expectation formula in Equation (20). This connection is also implicitly reflected in our probabilistic proof: matching the *drift terms* ($\mathrm{d}s$ terms) in (46), rather than the martingale terms used to identify u^*, recovers precisely the same HJB equation.

Alternate Proof of Proposition 3.2 via the HJB equation. We provide an alternative proof of Proposition 3.2 based on the HJB equation, following the approach of [28] but expressed in our notation.

The proof proceeds in three main steps. First, we decompose the objective functional to isolate the dynamic component of the control problem. Second, we derive the associated nonlinear HJB equation and linearize it using a logarithmic transformation. Finally, we solve the resulting linear PDE via the Feynman-Kac formula and apply a change of measure to recover the variational representation [25].

Part 1: Derive the SOC formulation by using Girsanov theorem We first derive the SOC formulation of the objective functional:

$$\mathcal{L}_i^{\mathrm{TRC}}(u,\lambda;x,t) = \mathbb{E}_{\vec{\mathbb{P}}^u}\left[\log\frac{\mathrm{d}\vec{\mathbb{P}}^u_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}} + \lambda\log\frac{\mathrm{d}\vec{\mathbb{P}}^u_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}\,\middle|\, X_t^u = x\right]. \tag{54}$$

We expand the terms using Lemma A.2 and the standard Girsanov theorem, substituting the following expressions:

$$\log\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}} = \log\frac{p_t(X_t^{u_i})}{p_{\mathrm{target}}(X_T^{u_i})} + \int_t^T\left(\frac{1}{2}\|u_i+v_i\|^2 + \nabla\cdot(\sigma v_i - f)\right)\mathrm{d}s + \int_t^T (u_i+v_i)\cdot\mathrm{d}W_s, \tag{55}$$

$$\log\frac{\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}} = \frac{1}{2}\int_t^T \|u-u_i\|^2\mathrm{d}s + \int_t^T (u-u_i)\cdot\mathrm{d}W_s. \tag{56}$$

Substituting these into $\mathcal{L}_i^{\mathrm{TRC}}$ and collecting the terms, we obtain:

$$\begin{aligned}
&\mathcal{L}_i^{\mathrm{TRC}}(u,\lambda;x,t)\\
&= \mathbb{E}_{\vec{\mathbb{P}}^u}\Bigg[\tfrac{1+\lambda}{2}\int_t^T\left(\left\|u + \left(\tfrac{1}{1+\lambda}v_i - \tfrac{\lambda}{1+\lambda}u_i\right)\right\|^2 + \tfrac{\lambda}{(1+\lambda)^2}\|v_i+u_i\|^2 + \tfrac{2}{1+\lambda}\nabla\cdot(\sigma v_i - f)\right)\mathrm{d}s\\
&\qquad + \log\frac{p_t(X_t^u)}{p_{\mathrm{target}}(X_T^u)}\,\Bigg|\, X_t^u = x\Bigg].
\end{aligned} \tag{57}$$

Since $\log p_t(X_t^u)$ is constant conditioned on $X_t^u = x$, we can separate it from the optimization objective. We define the dynamic cost functional $\widetilde{J}(u;x,t)$ and the potential term $h(x,s)$ as follows:

$$\mathcal{L}_i^{\mathrm{TRC}}(u,\lambda;x,t) = \log p_t(x) + \widetilde{J}(u;x,t), \tag{58}$$

$$\begin{aligned}
\widetilde{J}(u;x,t) \coloneqq \mathbb{E}_{\vec{\mathbb{P}}^u}\Bigg[&\int_t^T \tfrac{1+\lambda}{2}\left(\left\|u + \left(\tfrac{1}{1+\lambda}v_i - \tfrac{\lambda}{1+\lambda}u_i\right)\right\|^2 + h(X_s^u,s)\right)\mathrm{d}s\\
&- \log p_{\mathrm{target}}(X_T^u)\,\Bigg|\, X_t^u = x\Bigg],
\end{aligned} \tag{59}$$

$$h(x,s) \coloneqq \frac{\lambda}{(1+\lambda)^2}\|v_i+u_i\|^2 + \frac{2}{1+\lambda}\nabla\cdot(\sigma v_i - f). \tag{60}$$

Part 2: The nonlinear PDE for the value function We denote the value function for the dynamic part of the cost as $\widetilde{V}(x,t) \coloneqq \inf_u \widetilde{J}(u;x,t)$. Based on the running cost derived in Part 1, $\widetilde{V}$ must satisfy the Hamilton-Jacobi-Bellman (HJB) equation:

$$\frac{\partial \widetilde{V}}{\partial t} + \min_u \left\{ (f+\sigma u)\cdot\nabla\widetilde{V} + \frac{1}{2}\operatorname{Tr}(\sigma\sigma^\top \nabla^2 \widetilde{V}) + \tfrac{1+\lambda}{2}\left(\left\| u - \left(\tfrac{\lambda}{1+\lambda}u_i - \tfrac{1}{1+\lambda}v_i \right) \right\|^2 + h \right) \right\} = 0, \tag{61}$$

with the terminal condition $\widetilde{V}(x,T) = -\log p_{\text{target}}(x)$.

The minimizer u^* of the expression inside the curly braces can be found by taking the gradient with respect to u and setting it to zero. The relevant terms for minimization are $\sigma u \cdot \nabla\widetilde{V} + \frac{1+\lambda}{2}\|u - (\frac{\lambda}{1+\lambda}u_i - \frac{1}{1+\lambda}v_i)\|^2$. The first-order condition yields:

$$\sigma^\top \nabla\widetilde{V} + (1+\lambda)\left(u^* - \left(\tfrac{\lambda}{1+\lambda}u_i - \tfrac{1}{1+\lambda}v_i \right) \right) = 0. \tag{62}$$

Solving for u^*, we obtain the optimal control:

$$u^* = \left(\tfrac{\lambda}{1+\lambda}u_i - \tfrac{1}{1+\lambda}v_i \right) - \frac{1}{1+\lambda}\sigma^\top \nabla\widetilde{V}. \tag{63}$$

Substituting this optimal control u^* back into the HJB equation, the drift term and the quadratic running cost term combine as follows:

$$\text{Drift term:} \quad (f+\sigma u^*)\cdot\nabla\widetilde{V} = \left(f + \sigma\left(\tfrac{\lambda}{1+\lambda}u_i - \tfrac{1}{1+\lambda}v_i \right) \right)\cdot\nabla\widetilde{V} - \frac{1}{1+\lambda}\|\sigma^\top\nabla\widetilde{V}\|^2, \tag{64}$$

$$\text{Quad term:} \quad \frac{1+\lambda}{2}\left\| -\frac{1}{1+\lambda}\sigma^\top\nabla\widetilde{V} \right\|^2 = \frac{1}{2(1+\lambda)}\|\sigma^\top\nabla\widetilde{V}\|^2. \tag{65}$$

Summing these simplifies the nonlinear gradient term to $-\frac{1}{2(1+\lambda)}\|\sigma^\top\nabla\widetilde{V}\|^2$. Thus, we arrive at the following nonlinear PDE for $\widetilde{V}$:

$$\frac{\partial \widetilde{V}}{\partial t} + \left(f + \sigma\left(\tfrac{\lambda}{1+\lambda}u_i - \tfrac{1}{1+\lambda}v_i \right) \right)\cdot\nabla\widetilde{V} + \frac{1}{2}\operatorname{Tr}(\sigma\sigma^\top\nabla^2\widetilde{V}) - \frac{1}{2(1+\lambda)}\|\sigma^\top\nabla\widetilde{V}\|^2 + \frac{1+\lambda}{2}h = 0. \tag{66}$$

Part 3: Linearization and solution via the Feynman-Kac formula To solve the nonlinear PDE (66), we apply the Cole-Hopf transformation with a scaling factor matching the nonlinearity. Define a new function $E(x,t)$ such that:

$$\widetilde{V}(x,t) = -(1+\lambda)\log E(x,t). \tag{67}$$

The derivatives transform as follows:

$$\frac{\partial\widetilde{V}}{\partial t} = -(1+\lambda)\frac{1}{E}\frac{\partial E}{\partial t}, \quad \nabla\widetilde{V} = -(1+\lambda)\frac{\nabla E}{E}, \quad \nabla^2\widetilde{V} = -(1+\lambda)\frac{\nabla^2 E}{E} + (1+\lambda)\frac{\nabla E(\nabla E)^\top}{E^2}. \tag{68}$$

Substituting these into Equation (66), the nonlinear gradient terms cancel out perfectly:

$$\text{Hessian part:} \ \frac{1}{2}\operatorname{Tr}\left(\sigma\sigma^\top(1+\lambda)\frac{\nabla E(\nabla E)^\top}{E^2} \right) = \frac{1+\lambda}{2}\frac{\|\sigma^\top\nabla E\|^2}{E^2}, \tag{69}$$

$$\text{Nonlinear part:} \ -\frac{1}{2(1+\lambda)}\left\| \sigma^\top\left(-(1+\lambda)\frac{\nabla E}{E} \right) \right\|^2 = -\frac{1+\lambda}{2}\frac{\|\sigma^\top\nabla E\|^2}{E^2}. \tag{70}$$

Thus, multiplying the remaining terms by $-\frac{E}{1+\lambda}$, we obtain the linear PDE for $E(x,t)$:

$$\frac{\partial E}{\partial t} + \left(f + \sigma\left(\tfrac{\lambda}{1+\lambda}u_i - \tfrac{1}{1+\lambda}v_i \right) \right)\cdot\nabla E + \frac{1}{2}\operatorname{Tr}(\sigma\sigma^\top\nabla^2 E) - \frac{1}{2}hE = 0. \tag{71}$$

The terminal condition transforms to $E(x,T) = \exp\left(-\frac{\widetilde{V}(x,T)}{1+\lambda} \right) = (p_{\text{target}}(x))^{\frac{1}{1+\lambda}}$.

This is a backward Kolmogorov equation with a potential term. Let $\widetilde{u}_i \coloneqq \frac{\lambda}{1+\lambda}u_i - \frac{1}{1+\lambda}v_i$. By the Feynman-Kac formula, the solution is given by:

$$E(x,t) = \mathbb{E}_{\vec{\mathbb{P}}^{\widetilde{u}_i}}\left[\exp\left(-\frac{1}{2}\int_t^T h(X_s,s)\mathrm{d}s \right)(p_{\text{target}}(X_T))^{\frac{1}{1+\lambda}} \,\middle|\, X_t = x \right]. \tag{72}$$

Part 4: Explicit calculation of the expectation We now change the measure in the expectation from $\vec{\mathbb{P}}^{\widetilde{u}_i}$ to $\vec{\mathbb{P}}^{u_i}$. The drift difference is $\widetilde{u}_i - u_i = -\frac{1}{1+\lambda}(u_i + v_i)$. The Girsanov density is:

$$\frac{\mathrm{d}\vec{\mathbb{P}}^{\widetilde{u}_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} = \exp\left(-\frac{1}{1+\lambda}\int_t^T (u_i + v_i)\cdot \mathrm{d}W_s^{u_i} - \frac{1}{2(1+\lambda)^2}\int_t^T \|u_i + v_i\|^2 \mathrm{d}s\right). \tag{73}$$

Substituting this into the Feynman-Kac formula for $E(x,t)$, we combine the exponents:

$$E(x,t) = \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}\Bigg[\exp\Bigg(-\tfrac{1}{1+\lambda}\Big\{\int_t^T \Big(\tfrac{1}{2}\|u_i + v_i\|^2 + \nabla\cdot(\sigma v_i - f)\Big)\mathrm{d}s \\ + \int_t^T (u_i + v_i)\cdot \mathrm{d}W_s^{u_i} - \log p_{\mathrm{target}}(X_T)\Big\}\Bigg)\,\Bigg|\, X_t = x\Bigg]. \tag{74}$$

The term inside the brackets is exactly the log-likelihood ratio (55), excluding the initial term $\log p_t(x)$. Thus, substituting $E(x,t)$ back into

$$V_{i+1}^{\lambda}(x,t) \coloneqq \inf_{u\in\mathcal{U}} \mathcal{L}_i^{\mathrm{TRC}}(u,\lambda;x,t) = \log p_t(x) + \inf_{u\in\mathcal{U}} \widetilde{J}(u;x,t) = \log p_t(x) + \widetilde{V}(x,t) \tag{75a}$$

$$= \log p_t(x) - (1+\lambda)\log E(x,t), \tag{75b}$$

the $\log p_t(x)$ terms cancel out, yielding the desired result:

$$V_{i+1}^{\lambda}(x,t) = -(1+\lambda)\log \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}\left[\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}}\right)^{-\frac{1}{1+\lambda}}\,\middle|\, X_t^{u_i} = x\right]. \tag{76}$$

Moreover, since $\widetilde{V} = V_{i+1}^{\lambda} - \log p_t$, the optimal control obtained in Part 2 reads $u^* = \frac{\lambda}{1+\lambda}u_i - \frac{1}{1+\lambda}v_i - \frac{1}{1+\lambda}\sigma^\top\nabla\big(V_{i+1}^{\lambda} - \log p_t\big)$, as stated in Proposition 3.2. □

A.4 Stochastic optimal control formulation for SB based on the reciprocal property

Proposition A.7 (Preservation of the reciprocal property)**.** *Assume that the path measure at iteration i, denoted by $\vec{\mathbb{P}}^{u_i}$, is a reciprocal process with respect to the reference measure $\mathbb{P}^{\mathrm{ref}}$. Specifically, assume it admits the factorization:*

$$\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}}{\mathrm{d}\mathbb{P}^{\mathrm{ref}}}(X) = \exp\left(\log\widehat{\varphi}_{i-1}(X_0) + \log\varphi_i(X_T)\right) = \widehat{\varphi}_{i-1}(X_0)\varphi_i(X_T). \tag{77}$$

If v_i is the exact backward half-step for u_i (Lemma A.1), the updated measure $\vec{\mathbb{P}}^{u_{i+1}}$ obtained via the update rule (12) *remains a reciprocal process.*

Proof. By Proposition 3.1, the Radon-Nikodym derivative of the updated measure is given by the multiplicative update:

$$\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X) = \frac{\Psi_i(X_T)}{c_i(X_0)}, \qquad \Psi_i(X_T) \coloneqq \left(\frac{p_{\mathrm{target}}(X_T)}{\vec{\mathbb{P}}_T^{u_i}(X_T)}\right)^{\frac{1}{1+\lambda_i}}. \tag{78}$$

Note that Ψ_i depends solely on the terminal state X_T, and the normalization $c_i(X_0) \coloneqq \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[\Psi_i(X_T)\,|\,X_0]$, which ensures $\vec{\mathbb{P}}_0^{u_{i+1}} = p_{\mathrm{prior}}$, solely on the initial state X_0. Multiplying this density by the hypothesis for step i, we obtain:

$$\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\mathbb{P}^{\mathrm{ref}}}(X) = \frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X)\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}}{\mathrm{d}\mathbb{P}^{\mathrm{ref}}}(X) \tag{79a}$$

$$= \frac{\Psi_i(X_T)}{c_i(X_0)}\left[\widehat{\varphi}_{i-1}(X_0)\varphi_i(X_T)\right] \tag{79b}$$

$$= \frac{\widehat{\varphi}_{i-1}(X_0)}{c_i(X_0)}\left[\varphi_i(X_T)\Psi_i(X_T)\right]. \tag{79c}$$

By defining the updated potentials as $\widehat{\varphi}_i \coloneqq \widehat{\varphi}_{i-1}/c_i$ and $\varphi_{i+1} \coloneqq \varphi_i\cdot\Psi_i$, the measure $\vec{\mathbb{P}}^{u_{i+1}}$ retains the form of a reciprocal process in the same class as $\mathbb{P}^{\mathrm{ref}}$. □

Remark A.8. In our algorithm, the controls are initialized as $u_0 = v_0 = 0$, which implies $\vec{\mathbb{P}}^{u_0} = \mathbb{P}^{\mathrm{ref}}$. Since the reference measure is trivially reciprocal (with $\widehat{\varphi}_{-1} = \varphi_0 \equiv 1$), it follows by induction that $\vec{\mathbb{P}}^{u_i}$ is a reciprocal process for all $i \geq 0$. For $i = 0$, $v_0 = 0$ is in general not the backward half-step, but for $f = 0$, Lemma A.2 gives $\frac{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_0}}{\mathrm{d}\vec{\mathbb{P}}^{u_0}}(X) = \frac{p_{\mathrm{target}}(X_T)}{p_{\mathrm{prior}}(X_0)}$, so that the proof applies with $\Psi_0 = p_{\mathrm{target}}^{\gamma_0}$ up to a function of X_0. Crucially, Proposition A.7 holds for any value of the dual variable λ_i. Therefore, the preservation of the reciprocal structure is guaranteed even in our Trust Region IPF scheme where the step size $\gamma_i = \frac{1}{1+\lambda_i}$ is adaptively determined at each iteration i. Finally, we note that this result constitutes the continuous path-measure counterpart of Lemma 2 in [28].

Proposition A.9 (Equivalence of Reciprocal and Trust Region SOC)**.** *Consider the optimization at iteration i. First, assume that the previous forward control u_i is a reciprocal control associated with a potential φ_i. Second, let the backward control v_i be fixed, corresponding to the backward potential $\widehat{\varphi}_i$ via $v_i = \sigma^\top \nabla \log \widehat{\varphi}_i$.*

We restrict the search for the updated forward control u to the class of controls of reciprocal form, i.e., h-transforms of the reference process with a terminal function,

$$\mathcal{U}_{\mathrm{REC}} \coloneqq \left\{ u = \sigma^\top \nabla \log \varphi_t \,\middle|\, \varphi \colon \mathbb{R}^d \to (0,\infty),\ \varphi_t(x) \coloneqq \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}[\varphi(X_T) \,|\, X_t = x] \right\}, \tag{80}$$

so that $\frac{\mathrm{d}\vec{\mathbb{P}}^u}{\mathrm{d}\vec{\mathbb{P}}^{\mathrm{ref}}} = a(X_0)\,\varphi(X_T)$ for a function a of X_0 alone. The potential φ is determined by u up to a positive factor. For any $u \in \mathcal{U}_{\mathrm{REC}}$, let φ denote its associated forward potential. The Reciprocal SOC objective function is defined as:

$$\mathcal{L}_i^{\mathrm{REC}}(u, \lambda; x, t) \coloneqq \mathbb{E}_{\vec{\mathbb{P}}^u}\left[\frac{1}{2} \int_t^T \|u(X_s^u)\|^2 \mathrm{d}s + \Psi_i(X_T^u, u) \,\middle|\, X_t^u = x \right], \tag{81}$$

where the terminal cost Ψ_i utilizes the potentials of u and u_i:

$$\Psi_i(y, u) \coloneqq \log \widehat{\varphi}_i(y) + \mathcal{E}(y) + \lambda \log \frac{\varphi(y)}{\varphi_i(y)}. \tag{82}$$

Then, for any $u \in \mathcal{U}_{\mathrm{REC}}$, the following identity holds:

$$\mathcal{L}_i^{\mathrm{TRC}}(u, \lambda; x, t) = \mathcal{L}_i^{\mathrm{REC}}(u, \lambda; x, t) - \lambda \log \varphi_t(x) + C(x, t), \tag{83}$$

where $C(x,t)$ is independent of u. In particular, if φ is normalized such that $\varphi_t(x) = 1$, which leaves u unchanged, then $\mathcal{L}_i^{\mathrm{TRC}}$ and $\mathcal{L}_i^{\mathrm{REC}}$ differ by a constant.

Proof. First, we justify the ansatz. According to Proposition A.7, if the current iterate $\vec{\mathbb{P}}^{u_i}$ is a reciprocal process, then the optimal solution for the trust region control (TRC) problem is also a reciprocal process. Therefore, it is theoretically justified to restrict the search space to the reciprocal class. For $u \in \mathcal{U}_{\mathrm{REC}}$, conditionally on $X_t = x$, the density of $\vec{\mathbb{P}}^u_{[t,T]}$ with respect to the reference process is $\varphi(X_T)/\varphi_t(x)$ (Appendix A.1), and $\mathbb{E}_{\vec{\mathbb{P}}^u}[\frac{1}{2}\int_t^T \|u\|^2 \mathrm{d}s \,|\, X_t = x] = \mathbb{E}_{\vec{\mathbb{P}}^u}[\log \varphi(X_T) \,|\, X_t = x] - \log \varphi_t(x)$ by the Girsanov theorem.

Recall the definition of the TRC loss function:

$$\mathcal{L}_i^{\mathrm{TRC}}(u, \lambda; x, t) = \mathbb{E}_{\vec{\mathbb{P}}^u}\left[\log \frac{\mathrm{d}\vec{\mathbb{P}}^u_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}} + \lambda \log \frac{\mathrm{d}\vec{\mathbb{P}}^u_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}} \,\middle|\, X_t^u = x \right]. \tag{84}$$

For the first term, we utilize the identity relating the KL divergence to the potentials and control cost [30, Eq. (59)]. Note that the terminal density of the backward process $\overleftarrow{\mathbb{P}}^{v_i}$ matches the target, i.e., $q_T = p_{\mathrm{target}} = \mathcal{Z}^{-1} e^{-\mathcal{E}}$. Thus, we obtain:

$$\begin{aligned} &\mathbb{E}_{\vec{\mathbb{P}}^u}\left[\log \frac{\mathrm{d}\vec{\mathbb{P}}^u_{[t,T]}}{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}_{[t,T]}} \,\middle|\, X_t = x \right] \\ &\quad = \mathbb{E}_{\vec{\mathbb{P}}^u}\left[\frac{1}{2} \int_t^T \|u(s, X_s^u)\|^2 \mathrm{d}s + \log \widehat{\varphi}_i(X_T^u) + \mathcal{E}(X_T^u) + C_1(x) \,\middle|\, X_t = x \right], \end{aligned} \tag{85}$$

where $C_1(x) = \log p_t(x) - \log \widehat{\varphi}_i(x) + \log \mathcal{Z}$, with $p_t = \vec{\mathbb{P}}^{u_i}_t$ (Appendix A.1), represents terms that depend only on the initial state x (and the fixed backward process), which are constant with respect to the optimization variable u.

For the second term (the regularization term), we decompose the Radon-Nikodym derivative via the reference measure. Since both $\vec{\mathbb{P}}^{u}_{[t,T]}$ and $\vec{\mathbb{P}}^{u_i}_{[t,T]}$ start from p_t at time t and have the densities $\varphi(X_T)/\varphi_t(x)$ and $\varphi_i(X_T)/\varphi_i(x,t)$ with respect to the reference process given $X_t = x$, we have:

$$\begin{aligned}\mathbb{E}_{\vec{\mathbb{P}}^u}\left[\log \frac{\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}\,\middle|\, X_t = x\right] &= \mathbb{E}_{\vec{\mathbb{P}}^u}\left[\log\left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{\mathrm{ref}}_{[t,T]}}\Big/\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{\mathrm{ref}}_{[t,T]}}\right)\,\middle|\, X_t = x\right] \\ &= \mathbb{E}_{\vec{\mathbb{P}}^u}\left[\log\frac{\varphi(X_T)}{\varphi_i(X_T)} + C_2(x)\,\middle|\, X_t = x\right], \end{aligned} \tag{86}$$

where $C_2(x) = \log\varphi_i(x,t) - \log\varphi_t(x)$.

Substituting (85) and (86) into the objective (84), and grouping all terms independent of u into $C(x,t) \coloneqq C_1(x) + \lambda\log\varphi_i(x,t)$, we arrive at the Reciprocal SOC formulation:

$$\mathcal{L}^{\mathrm{TRC}}_i(u,\lambda;x,t) = \mathcal{L}^{\mathrm{REC}}_i(u,\lambda;x,t) - \lambda\log\varphi_t(x) + C(x,t). \tag{87}$$

This concludes the proof. □

Remark A.10. This proposition provides the specific realization of the general relationship indicated in Eq. (17) for our Trust Region Control objective $\mathcal{L}^{\mathrm{TRC}}_i$. The formulation of $\mathcal{L}^{\mathrm{REC}}_i$ implicitly leverages the optimality condition that the solution must belong to the reciprocal class. Consequently, minimizing $\mathcal{L}^{\mathrm{REC}}$ over potentials normalized by $\varphi_t(x) = 1$ is equivalent to minimizing $\mathcal{L}^{\mathrm{TRC}}_i$. A crucial advantage of this reformulation is that the trust region regularization term $D_{\mathrm{KL}}(\vec{\mathbb{P}}^u|\vec{\mathbb{P}}^{u_i})$, which originally involves path-dependent integration, collapses into a purely terminal cost term $\lambda\log(\varphi(X_T)/\varphi_i(X_T))$ under the reciprocal assumption, up to the term $-\lambda\log\varphi_t(x)$, which only fixes the normalization of φ.

Proposition A.11 (Optimality for the reciprocal SOC problem)**.** *Let $\vec{\mathbb{P}}^{u_i}$ be reciprocal with forward potential φ_i, i.e., $u_i = \sigma^\top\nabla\log\varphi_i$, and let $v_i = \sigma^\top\nabla\log\widehat{\varphi}_i$ be the exact backward half-step for u_i (Lemma A.4). Define*

$$\mathscr{V}^{\lambda}_{i+1}(x,t) \coloneqq -\log\mathbb{E}_{\mathbb{P}^{\mathrm{ref}}}\left[e^{-G(X_T)}\,\middle|\, X_t = x\right], \tag{88}$$

$$G \coloneqq \tfrac{1}{1+\lambda}\big(\log\widehat{\varphi}_i(\cdot,T) + \mathcal{E}\big) - \tfrac{\lambda}{1+\lambda}\log\varphi_i. \tag{89}$$

Then the updated control $u^{\lambda}_{i+1} = \arg\min_{u\in\mathcal{U}}\mathcal{L}^{\mathrm{TR}}_i(u,\lambda)$ of Proposition 3.1 lies in $\mathcal{U}_{\mathrm{REC}}$ and admits the representation

$$u^{\lambda}_{i+1}(x,t) = -\sigma^\top\nabla\mathscr{V}^{\lambda}_{i+1}(x,t). \tag{90}$$

Proof. Write $\mathscr{V}^{\lambda}_{i+1}(x,t) = -\log Z(x,t)$ with $Z(x,t) \coloneqq \mathbb{E}_{\mathbb{P}^{\mathrm{ref}}}[e^{-G(X_T)}\,|\,X_t = x]$. Since $\vec{\mathbb{P}}^{u_i}$ is reciprocal and v_i is its exact backward half-step, $\vec{\mathbb{P}}^{u_i}_T = \varphi_i\widehat{\varphi}_i(\cdot,T)$ (Lemma A.4), so that the weight in Proposition 3.1 reads $w_i = p_{\mathrm{target}}/\vec{\mathbb{P}}^{u_i}_T = e^{-\mathcal{E}}/(\mathcal{Z}\varphi_i\widehat{\varphi}_i(\cdot,T))$. By Proposition 3.1 and the reciprocal factorization $\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}}{\mathrm{d}\mathbb{P}^{\mathrm{ref}}} = a_i(X_0)\varphi_i(X_T)$,

$$\frac{\mathrm{d}\vec{\mathbb{P}}^{u^{\lambda}_{i+1}}}{\mathrm{d}\mathbb{P}^{\mathrm{ref}}}(X) = \frac{a_i(X_0)}{c_i(X_0)}\varphi_i(X_T)\,w_i(X_T)^{\frac{1}{1+\lambda}} = \widetilde{a}(X_0)\,e^{-G(X_T)}, \tag{91}$$

where $c_i(X_0) = \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[w_i^{\frac{1}{1+\lambda}}\,|\,X_0]$ and $\widetilde{a}$ collects all factors depending on X_0 alone, using $1 - \frac{1}{1+\lambda} = \frac{\lambda}{1+\lambda}$. Hence $\vec{\mathbb{P}}^{u^{\lambda}_{i+1}}$ is the Doob h-transform of the reference process with terminal function e^{-G}, i.e., $u^{\lambda}_{i+1} \in \mathcal{U}_{\mathrm{REC}}$ with potential $\varphi = e^{-G}$ and $\varphi_t = Z(\cdot,t)$, and its control is $u^{\lambda}_{i+1} = \sigma^\top\nabla\log Z = -\sigma^\top\nabla\mathscr{V}^{\lambda}_{i+1}$. Equivalently, $\mathscr{V}^{\lambda}_{i+1}$ solves the HJB equation $\partial_t\mathscr{V} + f\cdot\nabla\mathscr{V} + \frac{1}{2}\operatorname{Tr}(\sigma\sigma^\top\nabla^2\mathscr{V}) - \frac{1}{2}\|\sigma^\top\nabla\mathscr{V}\|^2 = 0$ with terminal condition $\mathscr{V}(\cdot,T) = G$, i.e., it is the value function of the SOC problem with running cost $\frac{1}{2}\|u\|^2$ and terminal cost G, which is the terminal cost Ψ_i of Proposition A.9 evaluated at its own optimal potential $\varphi = e^{-G}$. □

A.5 Proof of the main results in Section 3.1

Proof of Proposition 3.3. The derivation proceeds in three steps: identifying the optimal drift via the value function, applying the matching property of conditional expectations, and performing a change of measure to the sampling distribution.

Part 1: Characterization of the optimal drift. Let u_{i+1} denote the optimal control minimizing the forward KL divergence $\mathcal{L}_i^{\mathrm{CE}}(u) = D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}} | \vec{\mathbb{P}}^{u})$. Equivalently, minimizing the KL divergence corresponds to minimizing the expected L^2 distance between the drifts under the optimal path measure $\vec{\mathbb{P}}^{u_{i+1}}$:

$$\mathcal{L}_i^{\mathrm{CE}}(u) = \mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}\left[\frac{1}{2}\int_0^T \|u(X_t,t) - u_{i+1}(X_t,t)\|^2 \mathrm{d}t\right]. \tag{92}$$

From Proposition A.11, the optimal control is given by $u_{i+1}(x,t) = -\sigma^\top \nabla \mathscr{V}_{i+1}^{\lambda_i}(x,t)$. The value function admits the Feynman-Kac representation:

$$\mathscr{V}_{i+1}^{\lambda_i}(x,t) = -\log \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}\left[\exp\left(-G(X_T)\right) \mid X_t = x\right], \tag{93}$$

where the effective terminal potential $G(X_T)$ is defined based on Eq. (88) as:

$$G(X_T) \coloneqq \frac{1}{1+\lambda_i}\left(\log \widehat{\varphi}_i(X_T) + \mathcal{E}(X_T)\right) - \frac{\lambda_i}{1+\lambda_i}\log \varphi_i(X_T). \tag{94}$$

We explicitly derive the expression for the optimal control. Recall that the value function is given by $\mathscr{V}_{i+1}^{\lambda}(x,t) = -\log Z(x,t)$, where $Z(x,t)$ serves as the local normalizing constant (partition function) defined by the conditional expectation under the reference measure:

$$Z(x,t) \coloneqq \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}\left[\exp\left(-G(X_T)\right) \mid X_t = x\right]. \tag{95}$$

The optimal control is related to the gradient of the value function by $u_{i+1}(x,t) = -\sigma^\top \nabla \mathscr{V}_{i+1}^{\lambda_i}(x,t)$. Applying the chain rule to the logarithmic term $\nabla(-\log Z) = -\frac{\nabla Z}{Z}$, and using that, for $f = 0$, $X_T = X_t + \int_t^T \sigma(s)\,\mathrm{d}W_s$ under the reference process, so that ∇_x commutes with the conditional expectation, we have:

$$\begin{aligned}
u_{i+1}(x,t) &= -\sigma^\top \nabla\left(-\log \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}[e^{-G(X_T)} \mid X_t = x]\right) && \text{(96a)}\\
&= -\sigma^\top\left(-\frac{\nabla \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}[e^{-G(X_T)} \mid X_t = x]}{\mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}[e^{-G(X_T)} \mid X_t = x]}\right)\\
&= \sigma^\top \frac{1}{Z(x,t)} \nabla \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}\left[e^{-G(X_T)} \mid X_t = x\right] && \text{(96b)}\\
&= \sigma^\top \frac{1}{Z(x,t)} \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}\left[\nabla\left(e^{-G(X_T)}\right) \mid X_t = x\right] && \text{(96c)}\\
&= \sigma^\top \frac{1}{Z(x,t)} \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}\left[e^{-G(X_T)}\left(-\nabla G(X_T)\right) \mid X_t = x\right] && \text{(96d)}\\
&= -\sigma^\top \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}\left[\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{\mathrm{ref}}_{[t,T]}} \nabla G(X_T) \mid X_t = x\right] && \text{(96e)}\\
&= -\sigma^\top \mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}\left[\nabla G(X_T) \mid X_t = x\right] && \text{(96f)}
\end{aligned}$$

Here, the expectation is taken with respect to the optimal bridge process (which shares the same bridge dynamics as the reference), conditioned on the current state. Computing the gradient $-\sigma^\top \nabla G(X_T)$ explicitly using the relations $v_i = \sigma^\top \nabla \log \widehat{\varphi}_i$ and $u_i = \sigma^\top \nabla \log \varphi_i$, we define the target vector field Y_T:

$$Y_T \coloneqq -\sigma^\top \nabla G(X_T) = -\frac{1}{1+\lambda_i}\left(v_i(X_T) + \sigma^\top \nabla \mathcal{E}(X_T)\right) + \frac{\lambda_i}{1+\lambda_i} u_i(X_T). \tag{97}$$

Here, $\sigma = \sigma(t)$ is evaluated at the time t of the regressed control, and $v_i(X_T) \coloneqq \sigma(t)^\top \nabla \log \widehat{\varphi}_i(X_T)$ and $u_i(X_T) \coloneqq \sigma(t)^\top \nabla \log \varphi_i(X_T)$. These are the controls $v_i(X_T, T)$ and $u_i(X_T, T)$ at time T multiplied by $\sigma(t)^\top \sigma(T)^{-\top}$, which is the identity for constant σ; in (25), $v_i(X_T, T)$ and $u_i(X_T, T)$ are to be read in this sense. Our implementation uses a scalar $\sigma(t)$, parametrizes the controls as $u = \sigma u_\theta$ and $v = \sigma v_\phi$ (Appendix B.5.1), and regresses $u_\theta(X_t, t)$ onto $\sigma(t)^{-1} Y_T = -\nabla G(X_T)$, which contains no time-dependent factor. Thus, the optimal control is the projection of this target: $u_{i+1}(X_t, t) = \mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}[Y_T \mid X_t]$.

Part 2: Conditional flow matching. Substituting u_{i+1} back into the objective $\mathcal{L}_i^{\mathrm{CE}}(u)$, we obtain:

$$\mathcal{L}_i^{\mathrm{CE}}(u) = \mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}\left[\frac{1}{2}\int_0^T \|u(X_t) - \mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}[Y_T \mid X_t]\|^2 \mathrm{d}t\right]. \tag{98}$$

Invoking the *matching property* of conditional expectations (Lemma A.3, cf. [18]), minimizing $\|u - \mathbb{E}[Y|X]\|^2$ is equivalent to minimizing $\|u - Y\|^2$ up to a constant independent of u (the variance of Y given X). Therefore:

$$\begin{aligned}\mathcal{L}_i^{\mathrm{CE}}(u) &= \mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}\left[\frac{1}{2}\int_0^T \|u(X_t) - Y_T\|^2 \mathrm{d}t\right] + C \\ &= \mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}\left[\frac{1}{2}\int_0^T \left\|u(X_t) + \frac{v_i(X_T) + \sigma^\top \nabla \mathcal{E}(X_T)}{1+\lambda_i} - \frac{\lambda_i}{1+\lambda_i} u_i(X_T)\right\|^2 \mathrm{d}t\right] + C. \end{aligned} \tag{99}$$

Part 3: Change of measure. Finally, to evaluate this expectation using samples from the current policy $\vec{\mathbb{P}}^{u_i}$, we apply a change of measure. From Proposition 3.1, see (41), the Radon-Nikodym derivative (the importance weight) is computed by:

$$\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X) = \left(\frac{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X)\right)^{\frac{1}{1+\lambda_i}} \Bigg/ \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}\left[\left(\frac{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}\right)^{\frac{1}{1+\lambda_i}} \,\middle|\, X_0\right]. \tag{100}$$

Thus, the path expectation $\mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}$ can be implemented by sampling endpoints (X_0, X_T) from $\vec{\mathbb{P}}^{u_i}$, re-weighting them, and then sampling the bridge path from $\mathbb{P}^{\mathrm{ref}}_{t|0,T}$. Note that we can rely on the reference bridge because $\vec{\mathbb{P}}^{u_i}$ belongs to the reciprocal class. Substituting the importance weight into Equation (99) yields the objective (25). □

A.6 TR-IPF and its convergence

Algorithm 1 states TR-IPF as an idealized iteration on path measures. With $\lambda_i = 0$ for all i, it reduces to IPF (9), and with a fixed $\lambda_i = \lambda$ to the γ-IPF update (12) with $\gamma = (1+\lambda)^{-1}$. TR-SBS (Algorithm 2) implements each step with samples: the forward update by the matching objective (25), the backward half-step by (26), and the multiplier by the dual objective (21).

Algorithm 1 Trust region IPF (TR-IPF), idealized iteration on path measures.

Require: Prior p_{prior}, unnormalized target p_{target}, initial control u_0, trust region radius $\varepsilon > 0$, number of iterations I

1: **for** $i = 0, \dots, I-1$ **do**

2: **Backward half-step:** $\overleftarrow{\mathbb{P}}^{v_i} = \arg\min\{D_{\mathrm{KL}}(\mathbb{Q}|\vec{\mathbb{P}}^{u_i}) : \mathbb{Q}_T = p_{\mathrm{target}}\}$, so that $\frac{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X) = w_i(X) \coloneqq \frac{p_{\mathrm{target}}}{\vec{\mathbb{P}}^{u_i}_T}(X_T)$ (Lemma A.1).

3: **Step size:** solve the saddle-point problem (15) for λ_i and set $\gamma_i = (1+\lambda_i)^{-1}$.

4: **Forward update:** $\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X) = w_i(X)^{\gamma_i} / \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[w_i^{\gamma_i} \mid X_0]$ (Proposition 3.1), which solves (13).

5: **return** $\vec{\mathbb{P}}^{u_I}$

We restate Proposition 3.4 with its assumptions in full.

Proposition A.12 (Convergence rate of TR-IPF; detailed statement of Proposition 3.4)**.** *Let $(\vec{\mathbb{P}}^{u_i})_{i\geq 0}$ be the forward path measures generated by TR-IPF (Algorithm 1). TR-SBS generates them under the idealized assumption that each of its steps (the forward and backward matching problems and the Lagrange multiplier) is computed exactly and that $u_0 = v_0 = 0$, so that all iterates are reciprocal (Proposition A.7); since v_0 is in general not the backward half-step for u_0, this holds from the first backward update on, i.e., with u_1 in place of u_0. Let p_{prior} and p_{target} be the initial and target distributions, respectively, and define the objective function as $F(\mathbb{P}) := D_{\mathrm{KL}}(\mathbb{P}_T|p_{\mathrm{target}})$. Let $\mathcal{A} := \{\mathbb{P} \mid \mathbb{P}_0 = p_{\mathrm{prior}}\}$ denote the admissible class of path measures satisfying the initial boundary condition, and let $\mathbb{P}^* \in \arg\min_{\mathbb{P}\in\mathcal{A}} F(\mathbb{P})$ be the optimal global path measure.*

Suppose that the trust region constraint is active for every iteration $i = 0, \ldots, I-1$, i.e., it tightly holds that $D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}}|\vec{\mathbb{P}}^{u_i}) = \varepsilon$ for a fixed trust region radius $\varepsilon > 0$. For each i, let $\overleftarrow{\mathbb{P}}^{v_i}$ be the backward path measure at iteration i, and define the log-likelihood ratio $\ell_i := \log(\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}/\mathrm{d}\vec{\mathbb{P}}^{u_i})$. Assume that there exists a constant $M > 0$ such that the variance of the update direction is bounded: $\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}}(\ell_i) \leq M$ for all $i \geq 0$.

Then $F(\vec{\mathbb{P}}^{u_{i+1}}) \leq F(\vec{\mathbb{P}}^{u_i})$ for all i, and the gap after $I \geq 1$ steps is bounded at leading order by

$$F(\vec{\mathbb{P}}^{u_I}) - F(\mathbb{P}^*) = \min_{1\leq i\leq I} \left\{F(\vec{\mathbb{P}}^{u_i}) - F(\mathbb{P}^*)\right\} \lesssim \frac{D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_0})\sqrt{M}}{I\sqrt{2\varepsilon}}. \tag{101}$$

Consequently, to achieve an optimization gap of at most $\Delta > 0$, the required number of outer iterations I is $\mathcal{O}\left(\frac{D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_0})\sqrt{M}}{\Delta\sqrt{\varepsilon}}\right)$.

Proof. The proof proceeds in three main steps.

Part 1: TR-IPF as mirror descent By Proposition 3.3, the actual forward control loss function is equivalent to $D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}}|\vec{\mathbb{P}}^{u})$ up to a constant independent of u. Furthermore, Proposition 3.1 shows that the target measure $\vec{\mathbb{P}}^{u_{i+1}}$ satisfies the relation $\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} \propto \left(\frac{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}\right)^{\gamma_i}$, where $\gamma_i = \frac{1}{1+\lambda_i}$ and the normalization is conditional on X_0, see (41). Given this relation, optimizing the loss function

$$\min_{u\in\mathcal{U}} D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}}|\vec{\mathbb{P}}^{u}) \tag{102}$$

is equivalent to the optimization of

$$\min_{u\in\mathcal{U}} \left\{D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u}|\overleftarrow{\mathbb{P}}^{v_i}) + \frac{1-\gamma_i}{\gamma_i} D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u}|\vec{\mathbb{P}}^{u_i})\right\}, \tag{103}$$

which exactly corresponds to the forward control update of the γ-IPF. Furthermore, $\vec{\mathbb{P}}^{u_{i+1}}$ matches the minimizer of the exact population-level problem:

$$\min_{\mathbb{P}\in\mathcal{A}} \left\{D_{\mathrm{KL}}(\mathbb{P}|\overleftarrow{\mathbb{P}}^{v_i}) + \frac{1-\gamma_i}{\gamma_i} D_{\mathrm{KL}}(\mathbb{P}|\vec{\mathbb{P}}^{u_i})\right\}. \tag{104}$$

This exact matching is constructively shown in Proposition 3.2, guaranteeing that the optimal solution parameterized by the control u coincides perfectly with the optimal solution in the general path measure space $\mathbb{P}$.

On the other hand, by Lemma A.4, the optimization of

$$\mathcal{L}_i^B(v) = \mathbb{E}_{\vec{\mathbb{P}}_{0,T}^{u_{i+1}}\mathbb{P}_{t|0,T}^{\mathrm{ref}}}\left[\zeta_t \left\|v(X_t,t) - \sigma^\top \nabla \log p_{t|0}^{\mathrm{ref}}(X_t|X_0)\right\|^2\right] \tag{105}$$

yields the exact backward half-step v_{i+1} for u_{i+1}, since $\vec{\mathbb{P}}^{u_{i+1}}$ is reciprocal (Proposition A.7). Hence Lemma A.1 gives $\ell_i(X) = \log \frac{p_{\mathrm{target}}}{\vec{\mathbb{P}}_T^{u_i}}(X_T)$ for all $i \geq 1$. Therefore, the overall alternating optimization procedure constitutes the γ-IPF, and the forward update coincides with the mirror descent update (34) with $\mathbb{P}_i = \vec{\mathbb{P}}^{u_i}$, $g = -\ell_i$, and the adaptively chosen step size $\gamma = \gamma_i = \frac{1}{1+\lambda_i} \in (0, 1]$, cf. Section 2.2 and [28].

Part 2: Lower bound on the adaptive step size. The geometric form of the generalized IPF update implies that the next path measure is obtained by exponentially tilting the current path measure toward $\overleftarrow{\mathbb{P}}^{v_i}$. Then the update can be written as $\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} = \exp\{\gamma_i \ell_i - \psi_i(\gamma_i; X_0)\}$, where $\psi_i(\gamma; x) := \log \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[\exp(\gamma \ell_i) \mid X_0 = x]$ is the conditional log-partition function. Thus, for fixed i, the curve

$$\gamma \mapsto \mathbb{P}_{i,\gamma}, \qquad \frac{\mathrm{d}\mathbb{P}_{i,\gamma}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} = \exp\{\gamma \ell_i - \psi_i(\gamma; X_0)\}, \tag{106}$$

defines, conditionally on X_0, a one-dimensional exponential family on path space, with base measure $\vec{\mathbb{P}}^{u_i}(\cdot|X_0)$, natural parameter γ, and sufficient statistic ℓ_i. In particular, we have $\mathbb{P}_{i,0} = \vec{\mathbb{P}}^{u_i}$ and $\mathbb{P}_{i,\gamma_i} = \vec{\mathbb{P}}^{u_{i+1}}$, and all $\mathbb{P}_{i,\gamma}$ have the initial marginal p_{prior}.

By the standard local quadratic approximation of the Kullback-Leibler divergence in information geometry [2], the KL divergence between two nearby points in a smooth parametric statistical model is, to second order, one half of the squared parameter displacement measured in the Fisher information metric: $D_{\mathrm{KL}}(P_{\theta+d\theta}|P_\theta) = \frac{1}{2} d\theta^\top \mathcal{I}(\theta) d\theta + o(\|d\theta\|^2)$. Applying this fact to the family $\{\mathbb{P}_{i,\gamma}\}_\gamma$ at $\gamma = 0$ conditionally on X_0 and averaging over $X_0 \sim p_{\mathrm{prior}}$ gives

$$D_{\mathrm{KL}}(\mathbb{P}_{i,\gamma_i}|\mathbb{P}_{i,0}) = \frac{\gamma_i^2}{2}\mathcal{I}_i(0) + o(\gamma_i^2), \tag{107}$$

where the Fisher information along this exponential family is given by the averaged conditional variance of the sufficient statistic, $\mathcal{I}_i(0) = \mathbb{E}_{X_0 \sim p_{\mathrm{prior}}}\big[\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}}(\ell_i \mid X_0)\big] \le \mathrm{Var}_{\vec{\mathbb{P}}^{u_i}}(\ell_i)$. Consequently, the KL divergence can be expanded as

$$D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}}|\vec{\mathbb{P}}^{u_i}) = \frac{\gamma_i^2}{2}\mathbb{E}_{X_0}\big[\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}}(\ell_i \mid X_0)\big] + o(\gamma_i^2). \tag{108}$$

Under the active trust region constraint, we have $D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}}|\vec{\mathbb{P}}^{u_i}) = \varepsilon$. Using the uniform variance bound $\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}}(\ell_i) \le M$ and neglecting the higher-order term in the small-step (small-ε) regime, we obtain at leading order:

$$\varepsilon \lesssim \frac{\gamma_i^2}{2} M. \tag{109}$$

Therefore, we establish the asymptotic lower bound for the adaptive step size as:

$$\gamma_i \gtrsim \sqrt{\frac{2\varepsilon}{M}} =: \underline{\gamma}_\varepsilon. \tag{110}$$

Part 3: Applying and telescoping the descent lemma. By Part 1, we can apply Lemma A.5 with $\mathbb{P}_i = \vec{\mathbb{P}}^{u_i}$, $\gamma = \gamma_i \le 1$, and $\mathbb{Q} = \mathbb{P}^*$. This yields the following bound on the objective progress with respect to the global optimal path measure $\mathbb{P}^*$:

$$\gamma_i\big(F(\vec{\mathbb{P}}^{u_{i+1}}) - F(\mathbb{P}^*)\big) \le D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_i}) - D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_{i+1}}). \tag{111}$$

Summing this inequality over the outer iterations from $i = 0$ to $I - 1$, the right-hand side telescopes. Using the nonnegativity of the Kullback-Leibler divergence ($D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_I}) \ge 0$), we obtain the accumulated bound:

$$\sum_{i=0}^{I-1} \gamma_i\big(F(\vec{\mathbb{P}}^{u_{i+1}}) - F(\mathbb{P}^*)\big) \le D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_0}) - D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_I}) \le D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_0}). \tag{112}$$

Part 4: Best-iterate rate. Let $\Delta_I := \min_{1 \le i \le I}\Big\{F(\vec{\mathbb{P}}^{u_i}) - F(\mathbb{P}^*)\Big\}$ be the minimal objective gap over the first I updates. By definition, we have $F(\vec{\mathbb{P}}^{u_{i+1}}) - F(\mathbb{P}^*) \ge \Delta_I$ for all $0 \le i < I$. Substituting this into the accumulated sum from the telescoping bound yields:

$$\Delta_I \sum_{i=0}^{I-1} \gamma_i \le \sum_{i=0}^{I-1} \gamma_i\big(F(\vec{\mathbb{P}}^{u_{i+1}}) - F(\mathbb{P}^*)\big) \le D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_0}). \tag{113}$$

From the local quadratic approximation in Part 2, we have the uniform lower bound for the step size at leading order in the small-ε regime: $\gamma_i \gtrsim \underline{\gamma}_\varepsilon = \sqrt{2\varepsilon/M}$. Thus, the accumulated step size is bounded from below by $\sum_{i=0}^{I-1} \gamma_i \gtrsim I\sqrt{2\varepsilon/M}$.

Combining this with the telescoping bound gives

$$\Delta_I \left(I\sqrt{\frac{2\varepsilon}{M}} \right) \lesssim D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_0}). \tag{114}$$

Rearranging this inequality, we obtain the final best-iterate convergence rate:

$$\min_{1\le i\le I} \left\{ F(\vec{\mathbb{P}}^{u_i}) - F(\mathbb{P}^*) \right\} = \Delta_I \lesssim \frac{D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_0})\sqrt{M}}{I\sqrt{2\varepsilon}}. \tag{115}$$

Since $F(\vec{\mathbb{P}}^{u_{i+1}}) \le F(\vec{\mathbb{P}}^{u_i})$ by Lemma A.5, the minimum is attained at $i = I$, so that the bound also holds for the last iterate. To ensure that the optimization gap on the right-hand side is at most $\Delta > 0$, it suffices to require

$$I \ge \frac{D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_0})\sqrt{M}}{\Delta\sqrt{2\varepsilon}}. \tag{116}$$

Hence, the required number of outer iterations satisfies

$$I = \mathcal{O}\left(\frac{D_{\mathrm{KL}}(\mathbb{P}^*|\vec{\mathbb{P}}^{u_0})\sqrt{M}}{\Delta\sqrt{\varepsilon}} \right). \tag{117}$$

This completes the proof. □

A.7 Remarks on the variance of importance weights

Recall from Section 3.1 that, at iteration i of the Trust Region Schrödinger Bridge Sampler (TR-SBS), the next control is obtained by minimizing the cross-entropy objective

$$\mathcal{L}_i^{\mathrm{CE}}(u) := D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}}|\vec{\mathbb{P}}^{u}). \tag{118}$$

To make this objective tractable, we employ importance sampling. Since we can generate trajectories X^{u_i} from the previous iterate, we rewrite

$$\mathcal{L}_i^{\mathrm{CE}}(u) = \mathbb{E}_{\vec{\mathbb{P}}^{u_i}} \left[\left(\log \frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u}}(X^{u_i}) \right) \frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X^{u_i}) \right]. \tag{119}$$

In general, importance reweighting introduces exponential terms that may lead to large variance and numerical instabilities, especially in high-dimensional settings; see, e.g., [24]. In our case, however, the trust region constraint

$$D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u}|\vec{\mathbb{P}}^{u_i}) \le \varepsilon \tag{120}$$

(cf. (13)) implicitly controls the variance of the importance weights. This connection is made precise by the inequality

$$\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}} \left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} \right) \ge \exp\Big(D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}}|\vec{\mathbb{P}}^{u_i}) \Big) - 1. \tag{121}$$

Hence, if the KL divergence grows, the variance of the importance weights is forced to increase exponentially. Controlling the KL divergence is therefore a necessary condition to avoid variance explosion.

Empirically, we observe that enforcing this constraint is sufficient to keep the variance small in practice. We stress, however, that the following argument is local: it relies on a second-order expansion around $\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} \approx 1$ and does not provide a global bound on the variance. For small values of ε, as typically chosen in applications, we consistently find

$$\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}} \left(\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} \right) \approx 2\varepsilon, \tag{122}$$

see Appendix I.3 in [7]. This behavior follows at leading order in ε from the exponential family argument in Part 2 of the proof of Proposition A.12. By Proposition 3.1, the weight $W := \frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} =$

$\exp\{\gamma_i \ell_i - \psi_i(\gamma_i; X_0)\}$ satisfies $\mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[W \mid X_0] = 1$, and expanding in γ_i gives $W - 1 = \gamma_i\big(\ell_i - \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[\ell_i \mid X_0]\big) + \mathcal{O}(\gamma_i^2)$. Hence

$$\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}}(W) = \gamma_i^2 \overline{V}_i + o(\gamma_i^2) = 2D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}} | \vec{\mathbb{P}}^{u_i}) + o(\gamma_i^2) \approx 2\varepsilon, \tag{123}$$

with $\overline{V}_i = \mathbb{E}_{X_0 \sim p_{\mathrm{prior}}}[\mathrm{Var}_{\vec{\mathbb{P}}^{u_i}}(\ell_i \mid X_0)]$ as in Section 3.2, where we used $D_{\mathrm{KL}}(\vec{\mathbb{P}}^{u_{i+1}} | \vec{\mathbb{P}}^{u_i}) = \frac{\gamma_i^2}{2}\overline{V}_i + o(\gamma_i^2)$ from Part 2. The same expansion shows that the log-weight $\log W$ has variance 2ε at leading order. Both statements are local: the neglected terms involve higher cumulants of ℓ_i, which can be large when $\vec{\mathbb{P}}^{u_i}$ is far from $\overleftarrow{\mathbb{P}}^{v_i}$.

A.8 Relation to trust region methods on path space [7]

Our trust region formulation is inspired by [7], who constrain consecutive updates of a stochastic optimal control problem in KL divergence on path space. The present work differs in the following three respects.

Moving-target trust region formulation (Proposition 3.1). In [7], the updates are optimized toward a fixed target path measure. In contrast, the target $\overleftarrow{\mathbb{P}}^{v_i}$ of our forward update changes after each backward IPF step. We formulate the trust region update for this moving-target sequence and interpret the resulting dual variable as an adaptive γ-IPF step size $\gamma_i = (1 + \lambda_i)^{-1}$.

SB-specific stochastic-control reformulation (Propositions 3.2 and A.9). By exploiting the reciprocal structure of the IPF iterates, we reformulate the path space trust region update as an equivalent terminal-cost stochastic optimal control problem. This reformulation leads to the importance-weighted generalized adjoint-matching objective of Proposition 3.3, which can be evaluated tractably using endpoints sampled from the previous iterate.

Trust region IPF interpretation and convergence analysis (Proposition 3.4). We interpret the trust region iteration as a γ-IPF, or mirror descent, procedure with adaptive step size and establish a convergence guarantee for the resulting moving-target SB sequence.

A.9 Relation to ASBS and limiting cases

Figure 2 places TR-SBS among related IPF-based samplers. Moving from the first to the second row implements the path space update with samples of $\vec{\mathbb{P}}^{u_i}$, which requires the Radon-Nikodym weight to change the measure to the desired next iterate; without this weight (third row), one obtains fixed-point iterations such as ASBS (Section 3.1). Moving from left to right changes how the mirror-descent step size is selected. ASBS-IW and TR-ASBS isolate the two ingredients of TR-SBS and serve as ablations in Section 4. ASBS itself arises as a limiting case: when the KL constraint in (13) is inactive, complementary slackness gives $\lambda_i = 0$ and hence $\gamma_i = 1$, so TR-IPF reduces to IPF, TR-ASBS to ASBS, and TR-SBS to ASBS-IW. As $\varepsilon \to 0$, the update vanishes, with $\lambda_i \to \infty$ and $\gamma_i \to 0$.

A.10 Anchoring the matching target at the regressed time

On LJ-13 and alanine dipeptide, we replace the anchor $u_i(X_T, T)$ in (25) by $u_i(X_t, t)$, the previous control at the time and state of the regressed control (Appendix B.4). The following lemma compares a single update with the original and with the modified anchor, starting from the same iterate $\vec{\mathbb{P}}^{u_i}$, which we assume to be obtained from $u_0 = 0$ by exact updates as in Proposition 3.3. We write $\gamma_i = 1/(1+\lambda_i)$; for a time-dependent σ, $u_i(X_T, T)$ and $v_i(X_T, T)$ are multiplied by $\sigma(t)^\top \sigma(T)^{-\top}$, as in (25) (Appendix A.5).

Lemma A.13 (Anchor at the regressed time)**.** *Let*

$$\widetilde{\mathcal{L}}_i^{\mathrm{F}}(u) \coloneqq \mathbb{E}_{\vec{\mathbb{P}}_{0,T}^{u_i} \vec{\mathbb{P}}_{t|0,T}^{\mathrm{ref}}}\left[\left\| u(X_t,t) + \gamma_i\big(v_i(X_T,T) + \sigma^\top \nabla \mathcal{E}(X_T)\big) - (1-\gamma_i)\, u_i(X_t,t)\right\|^2 \frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X)\right]$$

denote the loss (25) *with the anchor $u_i(X_T, T)$ replaced by $u_i(X_t, t)$, and define the importance-weighted adjoint matching target $\widetilde{u}_{i+1}(x,t) \coloneqq -\mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}[v_i(X_T,T) + \sigma^\top \nabla \mathcal{E}(X_T) \mid X_t = x]$. Then:*

1. *The previous control satisfies $u_i(x,t) = \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[u_i(X_T,T) \mid X_t = x]$, whereas the minimizer of* (25) *is $u_{i+1}(x,t) = (1-\gamma_i)\,\mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}[u_i(X_T,T) \mid X_t = x] + \gamma_i\, \widetilde{u}_{i+1}(x,t)$.*

2. *The minimizer of* $\widetilde{\mathcal{L}}_i^{\mathrm{F}}$ *is* $(1-\gamma_i)\,u_i + \gamma_i\,\widetilde{u}_{i+1}$.

3. *For* $\gamma_i = 1$*, both minimizers equal* $\widetilde{u}_{i+1}$.

4. *For every* $\gamma_i \in (0,1]$*, if* $\vec{\mathbb{P}}^{u_i}$ *is the Schrödinger bridge, both minimizers equal* u_i*, i.e., the Schrödinger bridge is a fixed point of both iterations.*

Proof. 1. For $i = 0$, $u_0 = 0$ and the first identity is trivial. For $i \geq 1$, $\vec{\mathbb{P}}^{u_i}$ results from an exact update, so that, as shown in Appendix A.5 for $\vec{\mathbb{P}}^{u_{i+1}}$, it is the reference process tilted by a terminal potential,

$$\frac{\mathrm{d}\vec{\mathbb{P}}^{u_i}_{[t,T]}}{\mathrm{d}\vec{\mathbb{P}}^{\mathrm{ref}}_{[t,T]}} = \frac{\varphi_i(X_T)}{\varphi_i(X_t,t)}, \tag{124}$$

with $\varphi_i(x,t) = \mathbb{E}_{\vec{\mathbb{P}}^{\mathrm{ref}}}[\varphi_i(X_T) \mid X_t = x]$ and $u_i = \sigma^\top \nabla \log \varphi_i$. The computation in Appendix A.5, with $e^{-G(X_T)}$ and $Z(x,t)$ replaced by $\varphi_i(X_T)$ and $\varphi_i(x,t)$, gives $u_i(x,t) = \sigma^\top \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[\nabla \log \varphi_i(X_T) \mid X_t = x] = \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[u_i(X_T,T) \mid X_t = x]$. The second identity is $u_{i+1}(x,t) = \mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}[Y_T \mid X_t = x]$ (Appendix A.5) with Y_T split into its two terms.

2. By Part 3 of the proof of Proposition 3.3, the weighted expectation in (25), and hence in $\widetilde{\mathcal{L}}_i^{\mathrm{F}}$, is an expectation under $\vec{\mathbb{P}}^{u_{i+1}}$. By the matching property (Lemma A.3), the minimizer of $\widetilde{\mathcal{L}}_i^{\mathrm{F}}$ is therefore the conditional expectation of its regression target given X_t under $\vec{\mathbb{P}}^{u_{i+1}}$. Since $u_i(X_t,t)$ is a function of X_t, this conditional expectation is $(1-\gamma_i)\,u_i(X_t,t) + \gamma_i\,\widetilde{u}_{i+1}(X_t,t)$.

3. For $\gamma_i = 1$, the anchor has weight zero in both losses.

4. If $\vec{\mathbb{P}}^{u_i}$ is the Schrödinger bridge, its terminal marginal is p_{target}, so that the backward half-step does not change it, $\overleftarrow{\mathbb{P}}^{v_i} = \vec{\mathbb{P}}^{u_i}$. By Proposition 3.1, the importance weight is then 1 and $\vec{\mathbb{P}}^{u_{i+1}} = \vec{\mathbb{P}}^{u_i}$, so that the two conditional expectations of the anchor in statements 1 and 2 coincide and equal u_i. Moreover, since $\vec{\mathbb{P}}^{u_{i+1}} = \overleftarrow{\mathbb{P}}^{v_i}$, Proposition 3.3 with $\lambda_i = 0$ shows that $\widetilde{u}_{i+1}$ is the control of the update with $\gamma_i = 1$, whose path measure is $\overleftarrow{\mathbb{P}}^{v_i} = \vec{\mathbb{P}}^{u_i}$; hence $\widetilde{u}_{i+1} = u_i$, and both minimizers equal $(1-\gamma_i)\,u_i + \gamma_i\,u_i = u_i$. □

The modification thus replaces the conditional expectation of the anchor under the updated path measure by that under the previous one. Away from the Schrödinger bridge, the two differ in general, so that $\widetilde{\mathcal{L}}_i^{\mathrm{F}}$ is not the cross-entropy loss (23). Moreover, the minimizer in statement 2 need not be of the form $\sigma^\top \nabla \log \varphi$ with φ as in the proof, so that the next iterate need not be the reference process tilted by a terminal potential. The lemma then does not apply to the next update, and sampling X_t from the reference bridge is no longer exact. The iterates with the modified anchor are therefore not those of TR-IPF, and our analysis in Section 3.2 does not cover them. Statement 4 is not affected: an iteration started at the Schrödinger bridge stays there.

We nevertheless consider the modification reasonable, for three reasons. First, the Schrödinger bridge remains a fixed point (statement 4). Second, by statement 1, the minimizers of the two losses differ by $(1-\gamma_i)\big(\mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}} - \mathbb{E}_{\vec{\mathbb{P}}^{u_i}}\big)[u_i(X_T,T) \mid X_t]$. This difference is caused only by the tilt $\big(p_{\mathrm{target}}/\vec{\mathbb{P}}^{u_i}_T\big)^{\gamma_i}(X_T)$, up to normalization, that distinguishes $\vec{\mathbb{P}}^{u_{i+1}}$ from $\vec{\mathbb{P}}^{u_i}$, so that it vanishes at the Schrödinger bridge and is small when this tilt is close to constant. Since it equals $\gamma_i \operatorname{Cov}_{\vec{\mathbb{P}}^{u_i}}\big(u_i(X_T,T), \log w_i(X_T) \,\big|\, X_t\big) + \mathcal{O}(\gamma_i^2)$, it is, however, of the same order in γ_i as the update itself, so that the modified iteration differs from TR-IPF already at first order. Third, the modification removes one source of estimation error of the regression target from finite samples (Appendix B.4): in (25), the previous control enters the target only through its terminal values $u_i(X_T,T)$, whose conditional expectation $\mathbb{E}_{\vec{\mathbb{P}}^{u_{i+1}}}[u_i(X_T,T) \mid X_t]$ has to be estimated by the regression, whereas in $\widetilde{\mathcal{L}}_i^{\mathrm{F}}$, it enters directly at the time and state of the regressed control. Empirically, TR-SBS fails on LJ-13 without the two modifications of Appendix B.4 (Appendix C.7).

B Experiment details

In this section, we provide a comprehensive description of the experimental setup, evaluation metrics, and implementation details for the four benchmarks used to evaluate the mode-seeking capabilities

and scalability of our method: the Gaussian Mixture Model (GMM), the Many-Well (MW) potential, the Ginzburg-Landau (GL) model, and alanine dipeptide.

B.1 Gaussian mixture model (GMM)

To assess the ability to capture multiple isolated modes in high-dimensional spaces, we conduct evaluations under two distinct settings: one focusing on scalability with respect to dimensionality, and the other on robustness to an increasing number of modes.

B.1.1 Problem setup

The target density $p(x)$ is defined as a mixture of K multivariate Gaussian components in $\mathbb{R}^d$:

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k). \tag{125}$$

The component means are sampled uniformly from a hypercube $\mu_k \sim \mathrm{Uniform}([-L, L]^d)$. For the covariances, we use isotropic matrices $\Sigma_k = \sigma^2 I_d$ with $\sigma = \mathrm{Softplus}(1.0) \approx 1.31$ across all experiments. In both settings, the mixing weights π_k are *non-uniform* with a maximum probability ratio $\max_k \pi_k / \min_k \pi_k$ of 3.0. The target (means and weights) is generated once with a fixed seed and shared by all methods and training seeds.

Experimental configurations The specific configurations for the target distribution and the prior distribution differ between the two experiments to create appropriate difficulty levels for each task.

1. **Scalability with dimensionality:** We fix $K = 10$ and vary $d \in \{10, 25, 50, 100\}$. The means are sampled with $L = 7$. The prior distribution is set to $\mathcal{N}(0, 5.0^2 I_d)$.
2. **Robustness to increasing modes:** We fix $d = 10$ and vary $K \in \{5, 10, 20, 30, 40\}$. To accommodate a larger number of modes without excessive overlap, we use a wide range $L = 40$. The prior distribution covers this region with $\mathcal{N}(0, 30^2 I_d)$.

Training and model selection For each method and setting, we train with three seeds. The ASBS-family methods are trained for at most 150 outer iterations, and PIS and DIS for 40,000 gradient steps (Appendix B.6). During training, we periodically compute the SD to the target and keep the checkpoint with the lowest value. This checkpoint is then evaluated on 4,000 independent samples with a fixed evaluation seed, and we report the mean and sample standard deviation of these evaluations over the three seeds. Note that this model selection uses samples from the target, which are available for this synthetic benchmark.

B.1.2 Evaluation metrics

We employ the following metrics to evaluate the recovery of the multi-modal distribution:

- **Total variation distance (TVD):** Measures the L_1 discrepancy between the true mixing weights π and the empirical weights $\widehat{\pi}$ estimated from N samples.

$$\mathrm{TVD}(\pi, \widehat{\pi}) = \frac{1}{2} \sum_{k=1}^{K} |\widehat{\pi}_k - \pi_k| \tag{126}$$

  To compute the empirical weights $\widehat{\pi}_k$ specifically for the soft assignment metric, we utilize the posterior responsibility γ_{ik} of component k for each generated sample x_i. This is calculated using the ground-truth parameters as follows:

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)}. \tag{127}$$

  The soft empirical weight is then derived by averaging these responsibilities over the batch:

$$\widehat{\pi}_k = \frac{1}{N} \sum_{i=1}^{N} \gamma_{ik}. \tag{128}$$

- **Sinkhorn divergence:** To assess geometric fidelity beyond categorical mass assignment, we compute the Sinkhorn Divergence $S_\varepsilon(\mu, \nu)$ between the true and empirical distributions ($\varepsilon = 10^{-3}$).

For the final evaluation, we use $N = 4{,}000$ generated samples and the same number of ground-truth samples. In Appendix C.1, we additionally report the 2-Wasserstein distance (W_2), the sliced 2-Wasserstein distance, and the normalized effective sample size of the raw path weights $\frac{\mathrm{d}\tilde{\mathbb{P}}^v}{\mathrm{d}\tilde{\mathbb{P}}^u}$ (raw path ESS).

B.2 Many-Well potential (MW)

To evaluate the sampler's ability to navigate high-dimensional landscapes with a combinatorial number of modes and cross high energy barriers, we employ the Many-well potential. We conduct two sets of experiments: one focusing on the robustness to the number of modes (increasing m), and the other on the robustness to the geometric distance between modes (increasing δ).

B.2.1 Problem setup

The energy function $\mathcal{E}(x)$ for a configuration $x \in \mathbb{R}^d$ is composed of non-convex double-well potentials in the first m dimensions and a harmonic potential in the remaining $d - m$ dimensions:

$$\mathcal{E}(x) = \sum_{i=1}^{m} (x_i^2 - \delta)^2 + \frac{1}{2} \sum_{j=m+1}^{d} x_j^2, \tag{129}$$

where m denotes the number of active double-well dimensions and δ controls the location of the minima ($x_i = \pm\sqrt{\delta}$). The total number of symmetric global minima is 2^m.

Experimental configurations We configure the experiments as follows to test specific capabilities of the sampler:

1. **Robustness to increasing modes (varying m):** We fix the total dimensionality at $d = 64$ and the barrier parameter at $\delta = 4.0$. We vary the number of active double-wells $m \in \{4, 8, 16, 32\}$. At $m = 32$, the landscape contains 2^{32} modes, posing an extreme challenge for global mass discovery.
2. **Robustness to mode distance (varying δ):** We fix the dimensionality at $d = 8$ with $m = 5$ active wells (32 modes). We vary the well parameter $\delta \in \{2, 4, 6, 8\}$. Increasing δ expands the geometric distance between modes and heightens the energy barriers, testing the sampler's ability to escape local minima.

B.2.2 Evaluation metrics

To quantitatively assess the sample quality, we leverage the separability of the Many-Well potential, which allows us to compute ground-truth statistics analytically or via high-precision numerical integration.

- **Marginal Earth Mover's distance (EMD):** We compute the squared 2-Wasserstein distance, i.e., the optimal transport cost with squared Euclidean cost, between the empirical and ground-truth marginal distributions for each dimension j. We report the mean ($\mathrm{EMD}_{\mathrm{mean}}$) across all d dimensions.
- **Log-Partition Function Estimation Error ($\Delta_{\log Z}$):** To evaluate the global mass balance across the 2^m symmetric modes, we define the estimation error as the deviation of the estimated partition function $\widehat{Z}$ from the ground truth Z_{true}:

$$\Delta_{\log Z} = \left| \log \widehat{Z} - \log Z_{\mathrm{true}} \right|. \tag{130}$$

  The estimator $\widehat{Z}$ is derived by exploiting the symmetry of the potential. We define a specific quadrant $B = \{x \in \mathbb{R}^d : x_1, \ldots, x_m > 0\}$, which theoretically contains exactly 2^{-m} of the total probability mass. Using the empirical mass of this quadrant $\widehat{\pi}_B$, we estimate the global constant as $\widehat{Z} = Z_{\mathrm{true}} \cdot (2^m \widehat{\pi}_B)^{-1}$.

The ground-truth constant Z_{true} is obtained analytically due to the separability of the energy function:

$$\log Z_{\text{true}} = m \log Z_{DW}(\delta) + (d - m) \log \sqrt{2\pi}. \tag{131}$$

Here, $Z_{DW}(\delta)$ denotes the partition function of the 1D double-well component, which we calculate via high-precision numerical integration:

$$Z_{DW}(\delta) = \int_{-\infty}^{\infty} \exp\left(-(u^2 - \delta)^2\right) du. \tag{132}$$

In practice, this integral is approximated using numerical quadrature over a sufficiently large bounded interval to ensure negligible truncation error.

B.3 Ginzburg-Landau (ϕ^4 scalar field theory) model

To evaluate our method on a physically motivated problem with strong spatial correlations and continuous phase transitions, we employ the discretized Ginzburg-Landau model.

B.3.1 Problem setup

The target distribution is defined by the energy functional:

$$\mathcal{E}(x) = \frac{1}{2} \sum_{i=1}^{d} (x_i - x_{i+1})^2 + \frac{1}{2} \sum_{i=1}^{m} (x_i^2 - \delta)^2 + \frac{1}{2} \sum_{i=m+1}^{d} x_i^2, \tag{133}$$

where d is the total dimensionality, m is the number of active double-well dimensions, and δ dictates the location of the minima. We apply periodic boundary conditions such that $x_{d+1} = x_1$, with a coupling strength of 1.

Based on our configuration, we use $d = 16$, $m = 5$, and $\delta = 6$.

B.3.2 Evaluation metrics

Unlike the Gaussian Mixture Model or the Many-Well potential, the Ginzburg-Landau energy lacks a tractable normalizing constant and ground-truth samples. Therefore, we evaluate performance through the following intrinsic metrics:

- **Normalized effective sample size (ESS):** To measure the statistical efficiency and stability of the sampling, we calculate the normalized ESS derived from the importance weights $w = \frac{\mathrm{d}\vec{\mathbb{P}}^v}{\mathrm{d}\vec{\mathbb{P}}^u}(X^u)$. The values are normalized to the range $[0, 1]$, where a higher value indicates that the trust region constraint successfully maintained low-variance updates.
- **Symmetry violation (mirror marginal EMD):** Since the target energy obeys $\mathcal{E}(x) = \mathcal{E}(-x)$, the predictive distribution should ideally satisfy $p(x) = p(-x)$. To quantify this, we compute the 2-Wasserstein distance between the marginals of the predictive distribution $\vec{\mathbb{P}}_T^u$ and of its own reflection $-\vec{\mathbb{P}}_T^u$, instead of comparing with ground-truth samples as in the MW experiments. We report the mean EMD across all dimensions ($\overline{W}_2(X_T, -X_T)$). This allows us to assess global mode balance and symmetry preservation in the absence of ground-truth data.

B.4 Alanine dipeptide

B.4.1 Problem setup

We sample the Boltzmann distribution of alanine dipeptide (22 atoms) at 300 K, modeled with the Amber ff96 force field and the OBC implicit solvent model in OpenMM, as in [32]. Following [30, 32], the sampler operates in internal coordinates (bond lengths, bond angles, and dihedral angles; $d = 60$ after removing the six external degrees of freedom). The target density includes the Jacobian determinant of the coordinate transform. We use 10^6 reference samples from molecular dynamics. Alanine dipeptide is chiral: the four substituents of C_α can be arranged in two mirror-image forms, L and D, which cannot be converted into each other without breaking bonds. The energy is the same for

both forms, so the target also has modes of the D-form, while the reference samples contain only the L-form. We determine the form of a sample from the sign of $(x_{\mathrm{N}}-x_{\mathrm{C}_\alpha})\cdot\big((x_{\mathrm{C}}-x_{\mathrm{C}_\alpha})\times(x_{\mathrm{C}_\beta}-x_{\mathrm{C}_\alpha})\big)$, where $x_{\mathrm{C}_\alpha}\in\mathbb{R}^3$ is the Cartesian position of C_α and x_{N}, x_{C}, and x_{C_β} are those of the nitrogen, the carbonyl carbon, and the methyl carbon bonded to it; this sign is the same for all reference samples.

Prior. Both methods use the prior $(1-w)\,\mathcal{N}(0,I)+w\,\rho_\phi$ with $w=0.02$. In ρ_ϕ, the Gaussian component along the direction of a rigid rotation about the torsion ϕ is replaced by a uniform distribution over one full turn of ϕ, cut at $\phi=2.2$, where the target density is negligible. Thus, 2% of the initial states cover all values of ϕ, including the C7ax region $0\le\phi\le 2$, while the bulk of the prior is unchanged. This prior differs from the Gaussian prior $\mathcal{N}(0,0.25\,I)$ of [30], with which ASBS essentially does not find C7ax (see their Ramachandran plot).

Hyperparameters. The controls u and v are time-conditional MLPs with Fourier time features, 8 layers, and 256 hidden units, whose outputs are clipped componentwise to $[-100,100]$. We use $\sigma_{\mathrm{prior}}=1$, $\sigma_{\min}=10^{-3}$, $\sigma_{\max}=0.5$, $N_t=200$, $I=40$, $N_{\mathrm{FC}}=2{,}000$, $N_{\mathrm{BC}}=1{,}000$, and $L=100$, i.e., $120{,}000$ epochs in total, and Adam with learning rate 10^{-4} and gradient-norm clipping at 100. Training runs data-parallel on 4 NVIDIA A40 GPUs with, per GPU, batch size $B=500$, buffer size $N_{\mathrm{buf}}=40{,}000$, and $N_{\mathrm{new}}=1{,}000$. TR-SBS uses $\varepsilon=0.2$, $c_{\mathrm{IW}}=10$, $N_{\mathrm{state}}=500$ per GPU, i.e., $2{,}000$ initial states for the dual problem, and $N_{\mathrm{path}}=20$. ASBS uses the same settings with the losses of [30], i.e., adjoint matching at a single time $t\sim\mathcal{U}(0,T)$ and corrector matching at $t=T$ only, without importance weights and with $\gamma_i=1$. For both methods, samples with D-chirality are discarded before they enter the buffers, and $\nabla\mathcal{E}(X_T)$ in the regression target is clipped to norm 100. We report one training run (seed 0) per method. Compared with the alanine dipeptide setup of [30], which uses the prior $\mathcal{N}(0,0.25\,I)$, $I=15$, $N_{\mathrm{FC}}=4{,}000$, $N_{\mathrm{BC}}=2{,}000$ ($90{,}000$ epochs), $1{,}000$ new samples per epoch, and a buffer of 10^4 samples, we use a wider prior that also covers C7ax, more and shorter stages, and more new samples per epoch and larger buffers on four GPUs. The target, the coordinates, the network, the noise schedule, and the evaluation metrics are the same.

Modifications of TR-SBS. On alanine dipeptide, we use two modifications. First, the step size is bounded from below, $\gamma_i\ge 0.05$. Second, in (25), we replace the anchor $u_i(X_T,T)$ by $u_i(X_t,t)$, the previous control at the same time and state as the regressed control. In (25), the previous control enters the regression target only through its terminal values $u_i(X_T,T)$, so that its values at intermediate times and states are determined by the regression as conditional expectations given X_t. For small γ_i, this term dominates the target, as its weight is $\lambda_i/(1+\lambda_i)=1-\gamma_i$, whereas the update itself is of order γ_i. Estimation errors of these conditional expectations, due to finite samples or to a learned control that is not exactly consistent with its own terminal values, then enter every update with a weight close to one and can accumulate over many small steps. With the modified anchor, the previous control enters the regression target at the time and state of the regressed control, so that only the term of order γ_i has to be estimated. For $\lambda_i=0$ (ASBS), the anchor drops out, so ASBS is unaffected; for large λ_i, the regression target approaches the previous control at every time, so that small steps leave the control unchanged. Lemma A.13 (Appendix A.10) characterizes the resulting update: its minimizer is a convex combination of the previous control and the importance-weighted adjoint matching target, it agrees with the minimizer of (25) for $\gamma_i=1$, and the Schrödinger bridge remains a fixed point. Away from the Schrödinger bridge, however, the modified loss is not exactly the cross-entropy loss, and our analysis in Section 3.2 applies to the unmodified loss.

B.4.2 Evaluation metrics

Following [30], we report

- the KL divergences $D_{\mathrm{KL}}(p_{\mathrm{ref}}|p_{\mathrm{model}})$ between the marginal histograms (200 bins on $[-\pi,\pi]$, with 10^{-5} added to the bin probabilities) of the backbone torsions ϕ and ψ and of the remaining torsions γ_1, γ_2, γ_3,
- W_2 between 10^4 model and 10^4 reference samples of (ϕ,ψ), with the Euclidean distance between the angles.

In addition, we report the populations of the rare conformations C7ax and α_L, i.e., the fraction of all samples that have L-chirality and satisfy $0\le\phi\le 2$ and $\psi<0$ (C7ax) or $\psi\ge 0$ (α_L). The reference populations are 0.169% and 0.049%. Except for W_2, all model metrics use 10^7 samples. Except for

Algorithm 2 Trust Region Schrödinger Bridge Sampler (TR-SBS), overview (see Algorithm 3 for details).

Require: Initial controls $u_0, v_0 = \mathbf{0}$, number of iterations I

1: **for** $i = 0, \dots, I-1$ **do**

2: Sample trajectories $X^{u_i} \sim \vec{\mathbb{P}}^{u_i}$ by integrating the forward SDE (4), and compute terminal controls $u_i(X_T^{u_i}, T)$ and the Radon-Nikodym derivative $w = \frac{\mathrm{d}\vec{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X^{u_i})$ as in (31).

3: Construct buffer $\mathcal{B} = \{(X_0^{u_i}, X_T^{u_i}), u_i(X_T^{u_i}, T), w\}$.

4: Compute the Lagrange multiplier $\lambda_i = \arg\max_{\lambda \in \mathbb{R}_+} \mathcal{L}_i^{\mathrm{Dual}}(\lambda)$ via (21), estimated with N_{path} trajectories $X^{u_i} \sim \vec{\mathbb{P}}^{u_i}(\cdot \mid X_0)$ for each of N_{state} initial states $X_0 \sim p_{\mathrm{prior}}$.

5: Compute $u_{i+1} = \arg\min_u \mathcal{L}_i^{\mathrm{F}}(u)$ using (25), buffer $\mathcal{B}$, and the weights $\frac{\mathrm{d}\vec{\mathbb{P}}^{u_{i+1}}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}} \propto w^{\frac{1}{1+\lambda_i}}$ as in (18).

6: Compute $v_{i+1} = \arg\min_v \mathcal{L}_i^{\mathrm{B}}(v)$ via (26) with endpoints $(X_0, X_T) \sim \vec{\mathbb{P}}_{0,T}^{u_{i+1}}$.

7: **return** control u_I solving the Schrödinger Bridge

ASBS after its divergence (19%), at most 0.022% of the samples of all reported checkpoints have D-chirality.

B.5 Implementation details and hyperparameters

The code is available at `https://github.com/take-koshizuka/tr_sbs`.

In this section, we detail the implementation of TR-SBS, covering the underlying SDEs, numerical integration schemes, network architectures, and data management strategies. The detailed pseudocode is provided in Algorithm 3. All experiments were conducted on an internal computing cluster. The training and evaluation for each configuration were performed using a single NVIDIA A40 GPU with 48GB of VRAM, except for the training on alanine dipeptide, which used four (Appendix B.4).

B.5.1 Geometric noise schedule and SDE formulation

We adopt the Variance Exploding (VE) SDE framework [47] as our reference dynamics, where the base drift is zero ($f(x,t) := 0$). The time-dependent diffusion coefficient $\sigma(t)$ is that of the VE SDE with the geometric noise levels $\sigma_{\min}(\sigma_{\max}/\sigma_{\min})^s$, $s \in [0,1]$, of [47], run in reverse time $s = 1-t$, so that the noise is largest at the prior ($t=0$) and smallest at the target ($t=1$):

$$\sigma(t) \coloneqq \sigma_{\min} \left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)^{1-t} \sqrt{2 \ln \frac{\sigma_{\max}}{\sigma_{\min}}}. \tag{137}$$

Forward and backward dynamics The Schrödinger bridge problem is formulated as finding a control policy u that minimizes transport cost while satisfying boundary conditions. The forward process X_t^u and the corresponding reverse-time backward process Y_t^v are governed by:

$$\mathrm{d}X_t^u = \sigma(t) u(X_t^u, t)\mathrm{d}t + \sigma(t)\vec{\mathrm{d}}W_t, \qquad X_0^u \sim p_{\mathrm{prior}}, \tag{138}$$

$$\mathrm{d}Y_t^v = -\sigma(t) v(Y_t^v, t)\mathrm{d}t + \sigma(t)\overleftarrow{\mathrm{d}}W_t, \qquad Y_T^v \sim p_{\mathrm{target}}, \tag{139}$$

where $u \in \mathcal{U}$ is the forward control and v is the backward control. The objective is to minimize the KL divergence with respect to the reference path measure (where $u \equiv 0$). Except for alanine dipeptide (Appendix B.4) and LJ-13 (Appendix C.7), we used a zero-mean Gaussian distribution as the prior distribution $p_{\mathrm{prior}} = \mathcal{N}(\mathbf{0}, \sigma_{\mathrm{prior}}^2 \mathbf{I})$.

Parameterization We parameterize the controls using neural networks u_θ and v_ϕ. To stabilize training and align with the scale of the diffusion coefficient, we scale the network outputs by $\sigma(t)$:

$$u(x,t) = \sigma(t) u_\theta(x,t), \tag{140}$$

$$v(x,t) = \sigma(t) v_\phi(x,t). \tag{141}$$

Substituting these into the SDEs, the effective drift terms become $\sigma(t)^2 u_\theta(X_t, t)$ and $-\sigma(t)^2 v_\phi(Y_t, t)$, ensuring that the learnable drift magnitudes remain comparable to the noise level throughout the process.

Algorithm 3 Trust Region Schrödinger Bridge Sampler (TR-SBS)

Require: Sample-able source $X_0 \sim p_{\text{prior}}$, differentiable target log density $\log p_{\text{target}}(x)$ *(with* $\mathcal{E}(x) = -\log p_{\text{target}}(x)$*)*, neural networks for forward control u and backward control v.
Require: Replay buffers $\mathcal{B}_u, \mathcal{B}_v$ with max capacity N_{buf}, number of new samples per iteration N_{new}.
Require: Number of outer loop I, numbers of inner epochs $N_{\text{FC}}, N_{\text{BC}}$, batch size B, number of gradient steps L.
Require: A uniform time grid of $[0, T]$ given by $\mathcal{T} = \{t_0, \dots, t_{N_t}\}$.
Initialize $u_0, v_0 \coloneqq 0, \quad \mathcal{B}_u \leftarrow \varnothing, \quad \mathcal{B}_v \leftarrow \varnothing$.
for $i = 0, 1, \dots, I-1$ **do** ▷ outer-loop index

Optimize multiplier parameter λ: ▷▷ Solving dual problem
Sample N_{state} initial states $X_0 \sim p_{\text{prior}}$.
For each X_0, simulate N_{path} trajectories $\{X_t^{u_i}\}_{t\in\mathcal{T}}$ under the current control u_i.
Solve for $\lambda_i = \arg\max_{\lambda>0} \widehat{\mathcal{L}}_i^{\text{Dual}}(\lambda)$ via Brent (scipy.optimize) using the Monte Carlo estimate of:

$$\widehat{\mathcal{L}}_i^{\text{Dual}}(\lambda) = (1+\lambda)\,\mathbb{E}_{X_0}\left[-\log \mathbb{E}_{X_t^{u_i}|X_0}\left[\left(\frac{p_{\text{prior}}(X_0)\prod_{n=0}^{N_t-1} p(X_{t_{n+1}}^{u_i} \mid X_{t_n}^{u_i})}{p_{\text{target}}(X_T^{u_i})\prod_{n=0}^{N_t-1} q(X_{t_n}^{u_i} \mid X_{t_{n+1}}^{u_i})}\right)^{-\frac{1}{1+\lambda}}\right]\right] - \lambda\varepsilon$$

where p and q are forward and backward transition kernels, respectively.

Forward control (FC) update:
Initialize $u \leftarrow u_i$.
for $j = 1, 2, \dots, N_{\text{FC}}$ **do** ▷▷ Forward Control (FC) matching
  Update replay buffer $\mathcal{B}_u \leftarrow \text{POPULATEBUFFER}(u_i, v_i, \mathcal{B}_u, N_{\text{new}}, N_{\text{buf}})$.
  Sample a mini-batch $\{(X_0^{(k)}, X_T^{(k)}, w^{(k)})\}_{k=1}^B$ from $\mathcal{B}_u$.
  Sample time $t \sim \mathcal{U}(0,T)$ and bridge state $X_t^{(k)} \sim p^{\text{ref}}_{t|0,T}(\cdot|X_0^{(k)}, X_T^{(k)})$ for each pair $(X_0^{(k)}, X_T^{(k)})$.
  Take L gradient steps with respect to the FC objective $\mathcal{L}_i^{\text{F}}(u)$ (Equation (25)), with weights $w^{\frac{1}{1+\lambda_i}}$ normalized and clipped as described in Appendix B.5.2:

$$\mathcal{L}_i^{\text{F}}(u) = \mathbb{E}_{X_{0,T}^{(k)}\sim\mathcal{B}_u} \sum_{\tau\in\{0,t,T\}} \mathbb{E}_{X_\tau^{(k)}|X_{0,T}^{(k)}}\left[w^{\frac{1}{1+\lambda_i}}\left\|u(X_\tau^{(k)},\tau) - \sigma(\tau)\sigma(T)^{-1}\, m_i^*(X^{(k)})\right\|^2\right], \tag{134}$$

$$m_i^*(X^{(k)}) \coloneqq -\frac{1}{1+\lambda_i}\left(v_i(X_T^{(k)},T) + \sigma(T)^\top \nabla\mathcal{E}(X_T^{(k)})\right) + \frac{\lambda_i}{1+\lambda_i} u_i^{(k)}(X_T^{(k)},T). \tag{135}$$

  Update $u_{i+1} \leftarrow u$.

Backward control (BC) update:
Initialize $v \leftarrow v_i$.
for $j = 1, 2, \dots, N_{\text{BC}}$ **do** ▷▷ Backward Control (BC) matching
  Update replay buffer $\mathcal{B}_v \leftarrow \text{POPULATEBUFFER}(u_{i+1}, v_i, \mathcal{B}_v, N_{\text{new}}, N_{\text{buf}})$.
  Sample a mini-batch $\{(X_0^{(k)}, X_T^{(k)}, \cdot)\}_{k=1}^B$ from $\mathcal{B}_v$ and discard unused fields.
  Sample time $t \sim \mathcal{U}(0,T)$ and bridge state $X_t^{(k)} \sim p^{\text{ref}}_{t|0,T}(\cdot \mid X_0^{(k)}, X_T^{(k)})$ for each pair $(X_0^{(k)}, X_T^{(k)})$.
  Minimize the BC objective $\mathcal{L}_i^{\text{B}}(v)$ (with $\zeta_\tau = (1+\sigma^2(1-\tau)/\tau)^{-1}$):

$$\mathcal{L}_i^{\text{B}}(v) = \mathbb{E}_{(X_0^{(k)}, X_T^{(k)})\sim\mathcal{B}_v} \sum_{\tau\in\{t,T\}} \mathbb{E}_{X_\tau^{(k)}|X_{0,T}^{(k)}}\left[\zeta_\tau\left\|v(X_\tau^{(k)},\tau) - \sigma(\tau)^\top \nabla \log p^{\text{ref}}_{\tau|0}(X_\tau^{(k)} \mid X_0^{(k)})\right\|^2\right]$$

  Update $v_{i+1} \leftarrow v$.

end for
Return: Kinetic-optimal drift $u^* \approx u_I$

Algorithm 4 PopulateBuffer: Auxiliary routine for replay buffer construction

Routine POPULATEBUFFER($u_i, v_i, \mathcal{B}, N_{\text{new}}, N_{\text{buf}}$)
Sample N_{new} trajectories $(X_0, \dots, X_T^{u_i})$ from the current model u_i.
for $\ell = 1, \dots, N_{\text{new}}$ **do**
    Compute $w^{(\ell)} = w(X^{u_i,(\ell)})$ via path-integral:

$$\log w^{(\ell)} = \log \frac{\mathrm{d}\overleftarrow{\mathbb{P}}^{v_i}}{\mathrm{d}\vec{\mathbb{P}}^{u_i}}(X^{u_i,(\ell)}) \approx -\log\left(\frac{p_{\text{prior}}(X_0^{(\ell)}) \prod_{n=0}^{N_t-1} p(X_{t_{n+1}}^{u_i,(\ell)} \mid X_{t_n}^{u_i,(\ell)})}{p_{\text{target}}(X_T^{u_i,(\ell)}) \prod_{n=0}^{N_t-1} q(X_{t_n}^{u_i,(\ell)} \mid X_{t_{n+1}}^{u_i,(\ell)})}\right). \tag{136}$$

    $\mathcal{B} \leftarrow \mathcal{B} \cup \{(X_0^{(\ell)}, X_T^{(\ell)}, w^{(\ell)})\}$.
if $|\mathcal{B}| > N_{\text{buf}}$ **then**
    Remove the oldest $|\mathcal{B}| - N_{\text{buf}}$ elements from $\mathcal{B}$ (FIFO).
end if
return $\mathcal{B}$.

B.5.2 Numerical integration and trajectory weighting

To ensure numerical stability and accuracy, especially in regimes with limited discretization steps, we employ an exact integration scheme for the noise variance rather than a simple Euler-Maruyama approximation.

Exact integration of diffusion variance For a discretization grid $\mathcal{T} = \{t_0, \dots, t_{N_t}\}$, we compute the exact accumulated variance $\bar{\sigma}_k^2$ over each interval $[t_k, t_{k+1}]$ analytically:

$$\bar{\sigma}_k^2 := \int_{t_k}^{t_{k+1}} \sigma_s^2 \, ds = \sigma_{\max}^2 \left[\left(\frac{\sigma_{\min}}{\sigma_{\max}}\right)^{2t_k} - \left(\frac{\sigma_{\min}}{\sigma_{\max}}\right)^{2t_{k+1}}\right]. \tag{142}$$

This formulation captures the precise energy of the noise injected into the system, avoiding discretization errors associated with approximating $\sigma_t^2 \Delta t$ when σ_t varies rapidly.

State update rule Using the calculated variance $\bar{\sigma}_k^2$, the forward state update follows a modified Euler-Maruyama scheme:

$$X_{t_{k+1}} = X_{t_k} + \sigma(t_k) u(X_{t_k}, t_k) \Delta t_k + \bar{\sigma}_k \cdot \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, I), \tag{143}$$

where $\Delta t_k = t_{k+1} - t_k$. This ensures that the stochastic term remains consistent with the underlying continuous-time process.

Computation of importance weights The computation of the Radon-Nikodym derivative $w(\mathbf{X})$ is crucial for the trust region update. Consistent with Algorithm 4, the log-weight for a trajectory $\mathbf{X} = (X_{t_0}, \dots, X_{t_{N_t}})$ is given by:

$$\log w(\mathbf{X}) \approx \log p_{\text{target}}(X_T) + \sum_{k=0}^{N_t-1} \log q(X_{t_k} \mid X_{t_{k+1}}) - \log p_{\text{prior}}(X_0) - \sum_{k=0}^{N_t-1} \log p(X_{t_{k+1}} \mid X_{t_k}). \tag{144}$$

Here, the transition probabilities p (forward) and q (backward) are defined by the Gaussian transitions derived from the exact variance integration:

$$\log p(X_{t_{k+1}} \mid X_{t_k}) = -\frac{1}{2\bar{\sigma}_k^2} \left\| X_{t_{k+1}} - X_{t_k} - \sigma(t_k) u(X_{t_k}, t_k) \Delta t_k \right\|^2 + Z_k, \tag{145}$$

$$\log q(X_{t_k} \mid X_{t_{k+1}}) = -\frac{1}{2\bar{\sigma}_k^2} \left\| X_{t_k} - X_{t_{k+1}} - \sigma(t_{k+1}) v(X_{t_{k+1}}, t_{k+1}) \Delta t_k \right\|^2 + Z_k. \tag{146}$$

Note that the normalization constant Z_k cancels out in the likelihood ratio.

Normalization and clipping of the importance weights In the forward matching loss, we use the tempered weights $w^{\frac{1}{1+\lambda_i}}$. For numerical stability, the tempered log-weights are shifted such that the weights have unit mean over each training mini-batch, and the resulting weights are clipped at a

maximum value c_{IW} (Table 2). The batch normalization replaces the normalization $\mathbb{E}_{\vec{\mathbb{P}}^{u_i}}[w^{\frac{1}{1+\lambda_i}} \mid X_0]$ in Proposition 3.1, which cannot be estimated from a buffer that contains a single trajectory per initial state; the dual objective in Algorithm 3, in contrast, uses N_{path} trajectories per initial state and estimates the conditional expectation directly. In addition, our implementation multiplies the log-weights $\log w$ of each batch of N_{new} new trajectories by $r = \min\{|\widehat{V}_0|/|\overline{\log w}|, 1\} \in (0, 1]$ before they enter the buffer, where $\overline{\log w}$ is their mean and $\widehat{V}_0$ is the Monte Carlo estimate of $\mathcal{L}_i^{\mathrm{Dual}}(0)$ in Algorithm 3, both computed with the unnormalized target. This heuristic additionally tempers the weights in the loss to $w^{r/(1+\lambda_i)}$. On alanine dipeptide, $r < 0.1$ in six of the first seven of the 40 outer iterations, so that the weights are almost uniform there, and $r \approx 0.99$ afterwards.

B.5.3 Network architecture

We parameterize both $u_\theta(t, x)$ and $v_\phi(t, x)$ using a **Time-Conditional MLP** designed to handle high-dimensional inputs.

- **Embedding:** The time t is mapped via *sinusoidal embeddings* with frequency components, while the input state x is projected using a learnable linear layer. These embeddings are summed to form the input to the backbone.
- **Backbone:** A 4-layer Multilayer Perceptron (MLP) with a hidden dimension of 256 and **GELU** activations (8 layers on alanine dipeptide, Appendix B.4).
- **Output clipping:** To ensure numerical stability, each component of the network output is clipped to $[-100, 100]$.

B.5.4 Replay buffers and optimization

Buffer management In our implementation, we maintain separate replay buffers, $\mathcal{B}_u$ and $\mathcal{B}_v$, dedicated to the Forward Control (FC) and Backward Control (BC) updates, respectively. Additionally, the optimization of the Lagrange multiplier λ (the dual problem) is performed using a fresh set of $N_{\mathrm{state}} \times N_{\mathrm{path}}$ trajectories, starting from N_{state} initial states sampled from the prior, at each outer iteration. We adopted a First-In-First-Out (FIFO) update rule where only a portion of the buffer (N_{new} samples) is refreshed at each iteration. As noted in [30], this mechanism renders the algorithm a *hybrid method between on-policy and off-policy* training.

This design choice allows for a fair comparison with the Adjoint Schrödinger Bridge Sampler (ASBS) [30]. However, we emphasize that our TR-SBS formulation is inherently more amenable to off-policy configurations due to the explicit incorporation of importance sampling weights and the trust region constraint, which theoretically allows for reduced bias when reusing samples from previous policies.

Optimization settings We use the Adam optimizer with a constant learning rate of 1×10^{-4} for training both control networks across all tasks.

Table 2 provides a unified summary of the hyperparameters across all five experimental settings. To ensure a fair comparison, we evaluate the ASBS-family baselines under the same settings, except for the diffusion scale $\sigma_{\max}$ on GMM, which was chosen per method.

B.6 PIS and DIS baselines

We compare with the Path Integral Sampler (PIS) [56] and the time-reversed diffusion sampler (DIS) [5] on GMM. Both are trained by minimizing the reverse KL divergence (ELBO) on path space. Their control network uses the same architecture as the forward control of the ASBS-family methods (4 layers, width 256, output norm clipped at 100) and does not take the score $\nabla \log p_{\mathrm{target}}$ as an input, so it has the same number of parameters as the ASBS controller at a given dimension. PIS starts from the zero initial state and DIS from the Gaussian prior of the respective setting. Both use 128 time steps with a fixed cosine noise schedule with minimum value 0.1 and maximum value $\sigma_{\max}$, where $\sigma_{\max} = 128$ (PIS) and 2 (DIS) for the mode suite and $\sigma_{\max} = 32$ (PIS) and 4 (DIS) for the dimension suite. We train for 40,000 steps with Adam (learning rate 5×10^{-4}, batch size 1,000, gradient clipping at 1.0) and use the same model selection and evaluation protocol as for the other methods (Appendix B.1).

Table 2: Unified hyperparameters for the GMM, MW, and GL experiments (alanine dipeptide: Appendix B.4; LJ-13: Appendix C.7). For MW, values separated by a slash (e.g., $0.1/0.2$) indicate that both values were evaluated for each specific target configuration, and the one yielding better performance was selected.

Parameter	**GMM (Scale d)**	**GMM (Robust K)**	**MW (Scale m)**	**MW (Robust δ)**	**GL (Periodic)**
σ_{prior}	5.0	30.0	1.0	1.0	1.0
$\sigma_{\max}$	$1/2/1/2^{\ddagger}$	$1/2/2/4^{\ddagger}$	0.2	$0.1/0.2$	0.2
$\sigma_{\min}$	0.1	0.1	0.01	0.01	0.01
ε	0.1	0.3	0.3	0.3	0.3
c_{IW}	10	10	1	1	1
I	$\leq 150^{\dagger}$	$\leq 150^{\dagger}$	50	35	80
N_{FC}	150	150	200	100	150
N_{BC}	50	50	50	100	150
L	100	100	100	100	100
N_t	200	200	200	200	200
B	1000	1000	1000	1000	1000
N_{buf}	50000	50000	50000	50000	50000
N_{new}	10000	10000	10000	10000	10000
N_{state}	5000	5000	5000	5000	5000
N_{path}	40	40	40	40	40

† Maximum number of outer iterations; the reported checkpoint is selected as described in Appendix B.1.
‡ For ASBS / TR-ASBS / ASBS-IW / TR-SBS, respectively.

Differences in budget. The ASBS-family methods use 200 time steps, whereas PIS and DIS use 128; the numbers of optimizer updates and target evaluations also differ. The comparison in Figure 5 therefore compares the completed experiments and does not claim equal total compute. A comparison with matched discretization and budget, as well as PIS and DIS results on MW, is left for future work.

B.7 Computational cost

Evaluations per outer iteration. In each outer iteration, the forward update populates the buffer with N_{new} new trajectories in each of its N_{FC} epochs and evaluates the target gradient $\nabla\mathcal{E}(X_T)$ at their end points. This gives $N_{\text{FC}}N_{\text{new}}$ target-gradient evaluations for both ASBS and TR-SBS. TR-SBS additionally evaluates $\mathcal{E}(X_T)$ for the same trajectories to compute the importance weights, and $N_{\text{state}}N_{\text{path}}$ times for the dual objective, but requires no further gradient evaluations. The backward update does not evaluate the target. With the values in Table 2, e.g., MW with varying δ uses $N_{\text{FC}}N_{\text{new}} = 100 \times 10{,}000 = 10^6$ target-gradient evaluations and $N_{\text{state}}N_{\text{path}} = 2 \times 10^5$ additional energy evaluations per outer iteration.

Number of outer iterations. The total cost scales as $\mathcal{O}(IN_{\text{FC}}N_{\text{new}})$, and the outer iterations needed to converge differ between the methods. On MW with $d = 8$ and $\delta = 8$ (Figure 11c), the mean energy indicates that ASBS has essentially converged after about $I = 6$ outer iterations, although to a mode-collapsed solution, whereas TR-SBS with $\varepsilon = 0.3$ reaches convergence after about $I = 32$ outer iterations. TR-SBS thus requires approximately $32/6 \approx 5.3$ times more target-gradient evaluations than ASBS in this setting. Figure 4 provides a complementary view: larger fixed step sizes converge faster in energy but to a worse solution, while a conservative fixed step size is stable but slow.

Hardware. All GMM runs, including PIS and DIS, use a single NVIDIA A40 GPU; the hardware of the other experiments is described in Appendix B.5. We do not report wall-clock times, since the runs were executed on a shared cluster. On alanine dipeptide, with the same hardware for both methods (Appendix B.4), one epoch of TR-SBS took about 1.5 times as long as one epoch of ASBS.

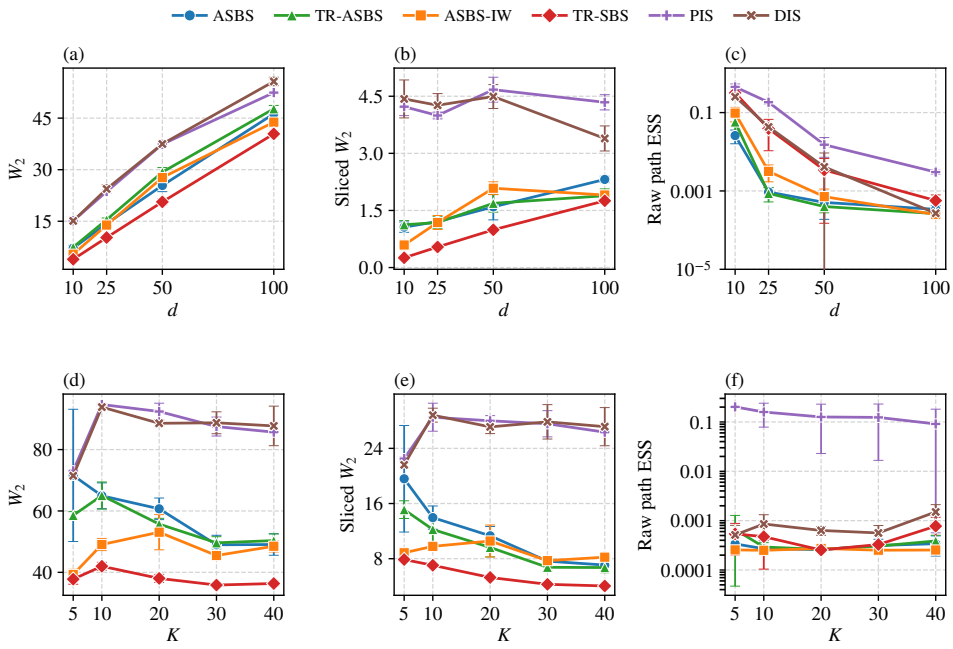


Figure 7: **Additional metrics on GMM** (mean $\pm$ standard deviation over three seeds). (a)-(c): $K = 10$, varying d. (d)-(f): $d = 10$, varying K. Lower is better for W_2 (a, d) and sliced W_2 (b, e); higher is better for the raw path ESS (c, f).

C Additional experimental results

C.1 Additional metrics on GMM

Figure 7 reports further metrics for the GMM experiment of Section 4.2, using the same checkpoints and evaluation protocol (Appendix B.1). The raw path ESS is the normalized effective sample size of the untempered path weights $\frac{\mathrm{d}\vec{\mathbb{P}}^v}{\mathrm{d}\vec{\mathbb{P}}^u}(X)$, $X \sim \vec{\mathbb{P}}^u$, evaluated with the learned controls.

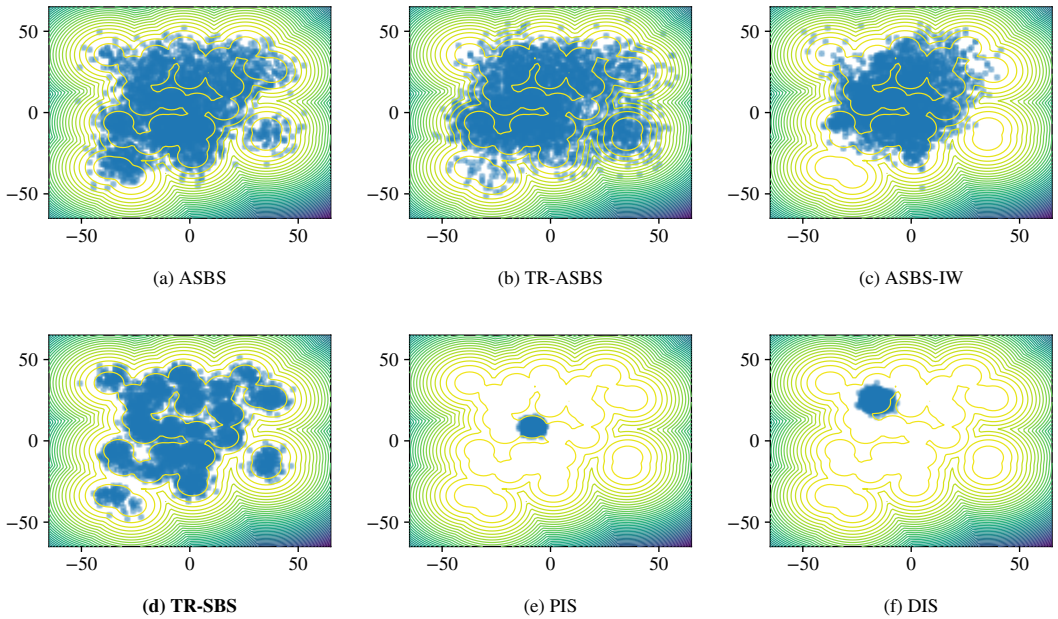
C.2 Visualization on GMM ($d = 10$, $K = 40$)

Figure 8 shows the first two coordinates of the 4,000 evaluation samples of the selected checkpoint (seed 0) for each method in the setting of Section 4.2 with $d = 10$ and $K = 40$, together with contours of the target marginal log density. PIS and DIS collapse onto a single component. ASBS-IW concentrates on the central components and misses the outlying ones in the lower left and on the right. ASBS and TR-ASBS reach most components, including the outlying ones, but place considerable mass between them and scatter samples into low-density regions. TR-SBS concentrates its samples on the high-density regions, including the outlying components, with clearer gaps between them, which is consistent with its lower SD and TVD in Figure 5 (c, d). Since the figure shows a two-dimensional projection of a ten-dimensional target, it complements but does not replace this quantitative comparison.

C.3 Visualization on high-barrier Many-Well ($d = 8, m = 5, \delta = 8$)

We provide a detailed visual analysis of the Many-Well potential with a high energy barrier ($\delta = 8$) in $d = 8$ dimensions. We compare four methods: ASBS, ASBS-IW, TR-ASBS (Ablation), and TR-SBS (Proposed) under two noise regimes.

- **Failure at small noise ($\sigma_{\max} = 0.1$):** Figure 9 illustrates the results with a conservative noise scale. Across all methods (a-d), the sampler fails to cross the high energy barrier. The marginal histograms (dim 0-2) show that the samples do not reach the wells at $\pm\sqrt{\delta}$ but remain between them and the barrier, resulting in *underfitting*. The energy distributions



(a) ASBS (b) TR-ASBS (c) ASBS-IW

(d) TR-SBS (e) PIS (f) DIS

Figure 8: **Samples on GMM with** $d = 10$ **and** $K = 40$ (first two coordinates, seed 0). Each panel shows the 4,000 evaluation samples of the checkpoint with the lowest SD during training. Blue points show generated samples; contours show the target marginal log density. All panels use identical contours and coordinate limits and show every sample. PIS and DIS use 128 instead of 200 time steps (Appendix B.6).

(rightmost column) are consistently shifted towards higher values compared to the ground truth, confirming that no method can reach the optimal solution.

- **Success with TR-SBS at Large Noise ($\sigma_{\max} = 0.2$):** Figure 10 presents the results under an aggressive noise scale designed to facilitate exploration.
  - **ASBS** (a) suffers from severe mode collapse across almost all dimensions. On the other hand,
  - **ASBS-IW** (b) exhibits *inconsistent recovery across dimensions*. While importance weighting enables it to capture both modes, it misbalances their mass in some dimensions (dim 1 and 2). This suggests that relying solely on importance weighting is insufficient to consistently correct the proposal bias.
  - **TR-ASBS** (c) manages to prevent total collapse but lacks fidelity. It fails to fit the distribution sharply.
  - **TR-SBS:** In stark contrast, **TR-SBS** (d) demonstrates superior performance. The marginal histograms accurately reproduce the symmetric bimodal structure of the ground truth, and the energy distribution overlaps closely with the target. This demonstrates that the combination of the trust region (for stability) and Importance Weighting (for precision) is essential for consistently navigating this challenging landscape.

C.4 Ablation study

Noise. Whereas ε controls the magnitude of each path space update (Section 3.2), the diffusion scale $\sigma_{\max}$ primarily controls exploration and barrier crossing: too little diffusion leads to stagnation, while larger diffusion facilitates barrier crossing but may increase finite-sample instability. The interaction of the two parameters is therefore indirect and problem-dependent.

We perform ablation studies to validate the impact of key hyperparameters and demonstrate the efficacy of the proposed trust region mechanism in managing the trade-off between efficiency and stability. In addition to the sample quality metrics utilized in the preceding subsection, we introduce the sample mean of the energy, $\mathbb{E}[\mathcal{E}(X_T)]$, to measure the convergence speed (specifically, the required number of iterations I) of the IPF update.

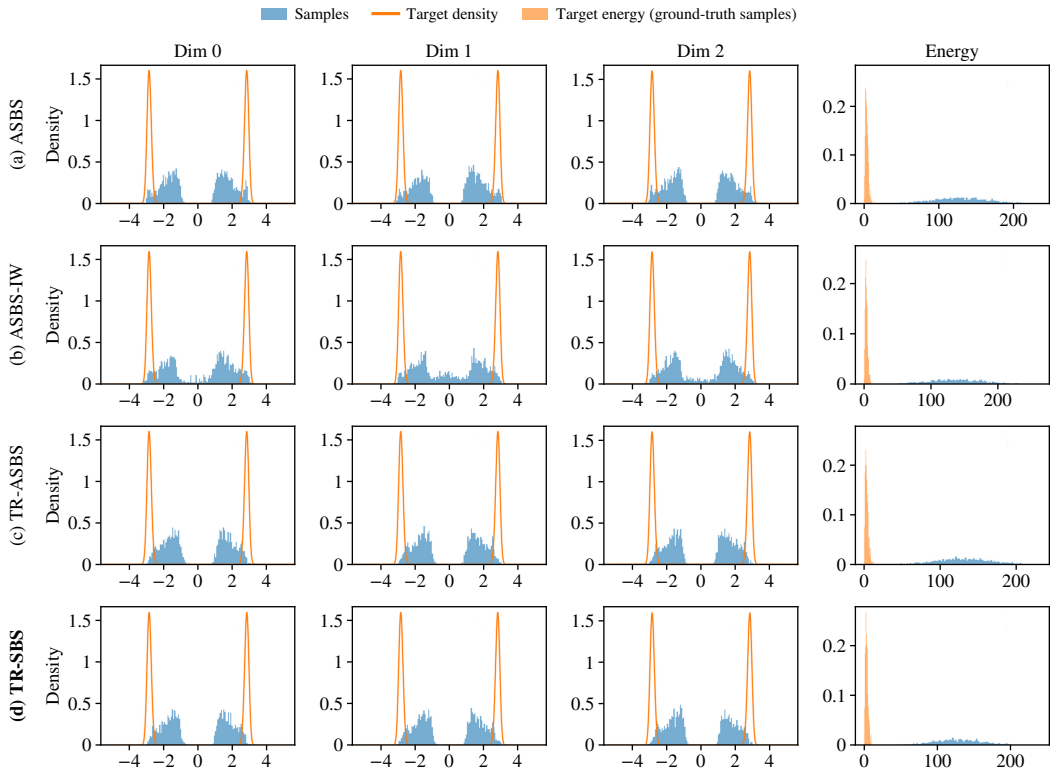


Figure 9: **Visualization at small noise** ($\sigma_{\max} = 0.1$)**.** Rows (a)–(d) show the four methods; columns show the marginal histograms of dimensions 0–2 and the energy distribution. Due to insufficient noise, **all methods** (a-d) fail to reach the wells (underfitting).

Impact of the maximum noise scale ($\sigma_{\max}$) A larger $\sigma_{\max}$ yields a smoother optimization landscape, making the problem more tractable and significantly accelerating convergence. However, in the context of IPF which approximates the objective using finite samples, an excessively large $\sigma_{\max}$ induces mode collapse. We demonstrate that **TR-SBS** allows for aggressive smoothing (large $\sigma_{\max}$) without sacrificing mode diversity. We analyze this trade-off using the MW potential ($d = 8$), analyzing both robustness to mode distance (δ) and training dynamics ($\delta = 8$). Figure 11 presents the results.

Stability: While a large $\sigma_{\max}$ aids optimization speed, it introduces instability in the finite-sample regime of standard ASBS, triggering severe mode collapse. Figure 11a shows that for ASBS with $\sigma_{\max} = 0.2$ (blue line), the Marginal EMD explodes as the task difficulty increases ($\delta \to 8$). This collapse is further evidenced in the training dynamics in Figure 11b, where the error for ASBS rises uncontrollably. In contrast, **TR-SBS** with $\sigma_{\max} = 0.2$ (red line) maintains a low EMD comparable to conservative settings. Figure 11d shows a histogram of the first coordinate slice of MW. Under large noise ($\sigma_{\max} = 0.2$), standard ASBS (top right) collapses to a single mode. In contrast, TR-SBS (bottom right) successfully captures the bimodal structure even with the aggressive noise scale, thanks to the trust region constraint.

Efficiency vs. Stability: Figure 11c tracks the averaged energy during training. When $\sigma_{\max}$ is small (0.1), both methods stagnate at a high-energy plateau, trapped in local optima. Increasing $\sigma_{\max}$ accelerates the convergence of both, reaching lower energy states by smoothing the landscape.

Sensitivity to the numbers of inner epochs (N_{FC}, N_{BC}) As inherited from ASBS, the numbers of forward and backward matching epochs per outer iteration are important hyperparameters. Table 3 reports TR-SBS on MW ($d = 8$, $m = 5$, $\delta = 4$) for different values. Too few epochs in either update degrade the result, while too many epochs increase the cost and can overfit; in particular, the error of the log-normalizing constant deteriorates markedly away from $N_{\text{FC}} = N_{\text{BC}} = 100$.

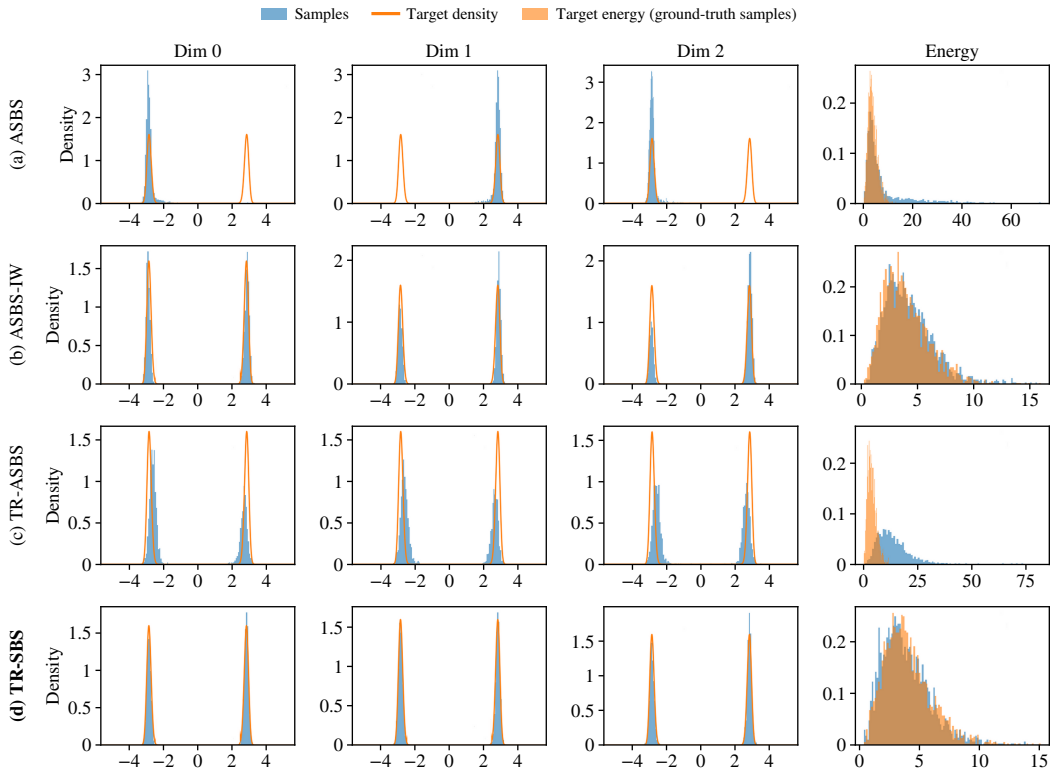


Figure 10: **Visualization at large noise ($\sigma_{\max} = 0.2$).** Rows and columns as in Figure 9. Aggressive noise facilitates exploration but causes instability. **(a)** ASBS suffers from mode collapse, and **(b)** ASBS-IW misbalances the modes. **(c)** TR-ASBS prevents total collapse but lacks fidelity. **(d)** Only **TR-SBS** successfully captures the global structure and energy distribution.

Table 3: Sensitivity of TR-SBS to N_{FC} and N_{BC} on MW ($d = 8$, $m = 5$, $\delta = 4$). The setting $100/100$ is the main result averaged over three seeds; the other settings use a single seed.

N_{FC}	N_{BC}	Marginal EMD ↓	$\lvert\Delta \log Z\rvert$ ↓
10	100	0.294	200.694
100	10	0.215	2.455
100	100	**0.131**	**0.006**
100	300	0.191	546.959
300	100	0.212	99.551

C.5 ϕ^4 scalar field theory

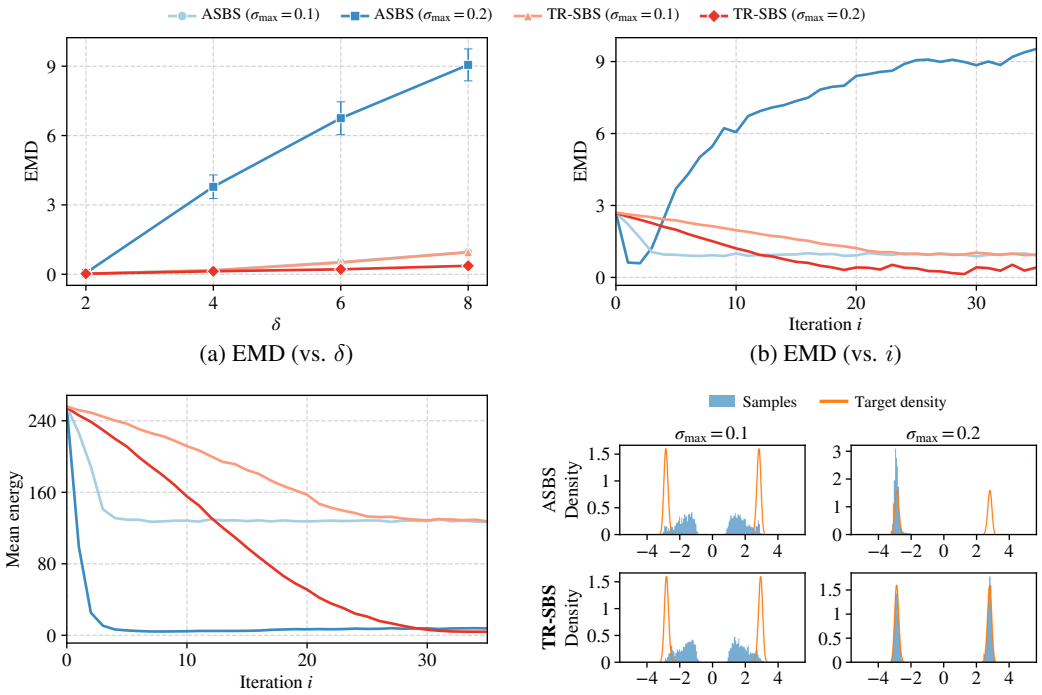
We evaluate our approach on the discretized Ginzburg-Landau model (ϕ^4 scalar field theory), a standard model of phase transitions and symmetry breaking [21, 41], with energy

$$\mathcal{E}(x) = \frac{1}{2}\sum_{i=1}^{d}(x_i - x_{i+1})^2 + \frac{1}{2}\sum_{i=1}^{m}(x_i^2 - \delta)^2 + \frac{1}{2}\sum_{i=m+1}^{d} x_i^2, \tag{147}$$

with periodic boundary conditions $x_{d+1} = x_1$.

Table 4: Results on the discretized Ginzburg-Landau model.

Method	ESS ↑	$\overline{W}_2$ ↓
ASBS	0.005	1.78
TR-SBS	**0.445**	**0.15**



(a) EMD (vs. δ) (b) EMD (vs. i)

(c) Mean energy (vs. i) (d) Visualization (dim 0)

Figure 11: **Ablation study on the maximum noise scale $\sigma_{\max}$.** Legend is shared across subplots (a-c). **(a)** Large $\sigma_{\max} = 0.2$ causes mode collapse in ASBS (blue) as barrier distance δ increases, whereas **TR-SBS** (red) remains robust. **(b)** Training dynamics confirm that ASBS diverges in EMD while TR-SBS remains stable. **(c)** A larger $\sigma_{\max}$ leads to faster convergence. **(d)** Histograms confirm that ASBS collapses to a single mode under large noise, while TR-SBS maintains the global structure.

We use $d = 16$, $m = 5$, and $\delta = 6$; all other settings are in Appendix B.3. The bistable coordinates induce multiple metastable configurations, while the nearest-neighbor term creates strong spatial correlations across lattice sites. Since the energy satisfies $\mathcal{E}(x) = \mathcal{E}(-x)$, the target distribution obeys $p(x) = p(-x)$. We therefore evaluate global mode balance using $\overline{W}_2(X_T, -X_T)$ together with the normalized effective sample size (ESS). Table 4 shows that TR-SBS substantially improves both weight stability and symmetry recovery over ASBS.

C.6 Alanine dipeptide

Table 5 reports all metrics of Appendix B.4 for TR-SBS and ASBS at similar stages of training, and Figure 12 the populations of the rare conformations over training. Early in training, ASBS strongly overpopulates the region $\phi > 0$, which contains C7ax and α_L. It then drains the region $\phi > 0$: at 48,500 epochs, it attains a lower KL_ϕ than TR-SBS but already underpopulates C7ax by a factor of two. The training of ASBS diverged between 54,000 and 54,500 epochs. To evaluate ASBS closer to the divergence, we resumed it from its checkpoint at 48,500 epochs with the same settings and newly filled buffers, saving a checkpoint every 500 epochs. The resumed run drains C7ax further, to 0.034% at 54,000 epochs, about a fifth of the reference, and diverges in the same interval, between 54,000 and 54,500 epochs, in both runs at the beginning of a stage of adjoint matching. After the divergence, the samples are far from the target: at 58,500 epochs of the original run, 19% have the wrong chirality, and all KL divergences and W_2 are larger than at any earlier checkpoint. Between 48,500 and 60,000 epochs, 15% of the optimization steps of the original run were skipped because the loss or its gradient was non-finite. In a third run, also resumed from 48,500 epochs, we additionally clip the norm of the regression target of corrector matching per sample at 100, the bound at which the outputs of the networks are clipped componentwise and the gradient of the energy in the terminal cost is clipped in norm, and discard samples with an energy above 1,000 before they enter the buffers, which affects fewer than one in 10^5 samples. This run does not diverge within 70,000 epochs, but it drains C7ax further, to 0.0001% at 70,000 epochs, about 1/1,400 of the reference (Table 5

and Figure 12); between 58,000 and 70,000 epochs, its KL_ϕ decreases, but its KL_ψ and W_2 increase. Its Ramachandran plot at 70,000 epochs is shown in Figure 6. TR-SBS overpopulates C7ax by a factor between 1.6 and 2.1 at all checkpoints evaluated with 10^7 samples, while its overpopulation of α_L decreases from a factor of about seven to three, and KL_ϕ and KL_ψ decrease by more than an order of magnitude over training. Apart from the diverged ASBS checkpoint and TR-SBS at 33,000 epochs, the KL divergences of γ_2 and γ_3 are nearly identical for all checkpoints.

Table 5: Alanine dipeptide: metrics (10^7 samples) of TR-SBS and ASBS with the same prior at similar stages of training and at the end of training. Only some checkpoints of each run were saved and evaluated with 10^7 samples, so paired rows show the closest such checkpoints, whose epochs differ by up to 6,500. ASBS diverges between 54,000 and 54,500 epochs; the row at 54,000 epochs is from the run resumed at 48,500 epochs, and the row at 58,500 epochs is after the divergence. The rows "ASBS, clipped" are from the run resumed at 48,500 epochs with clipped corrector targets (Appendix C.6). Reference populations: C7ax 0.169%, α_L 0.049%.

Method	Epoch	KL_ϕ	KL_ψ	KL_{γ_1}	KL_{γ_2}	KL_{γ_3}	W_2	C7ax (%)	α_L (%)
ASBS	31k	0.2784	0.0243	0.0032	0.0158	0.0263	0.6659	1.5352	1.2509
TR-SBS	33k	0.1862	0.0306	0.0193	0.0210	0.0300	0.3205	0.3342	0.3646
ASBS	48.5k	0.0256	0.0132	0.0036	0.0159	0.0262	0.4922	0.0788	0.0837
TR-SBS	48.5k	0.0570	0.0093	0.0074	0.0162	0.0264	0.2700	0.3500	0.2259
ASBS, resumed	54k	0.0457	0.0276	0.0032	0.0157	0.0262	0.5744	0.0338	0.0746
ASBS (diverged)	58.5k	0.2967	0.3786	0.8315	0.0333	0.0423	1.7838	0.1927	0.3708
TR-SBS	120k	0.0144	0.0037	0.0016	0.0159	0.0263	0.2460	0.2724	0.1360
ASBS, clipped	58k	0.0470	0.0205	0.0037	0.0160	0.0262	0.5793	0.0115	0.0156
TR-SBS	60.5k	0.0326	0.0069	0.0029	0.0159	0.0261	0.2956	0.3467	0.2001
ASBS, clipped	70k	0.0252	0.0406	0.0025	0.0161	0.0263	0.8319	0.0001	0.0001
TR-SBS	76.5k	0.0207	0.0055	0.0018	0.0160	0.0261	0.2595	0.3458	0.1463

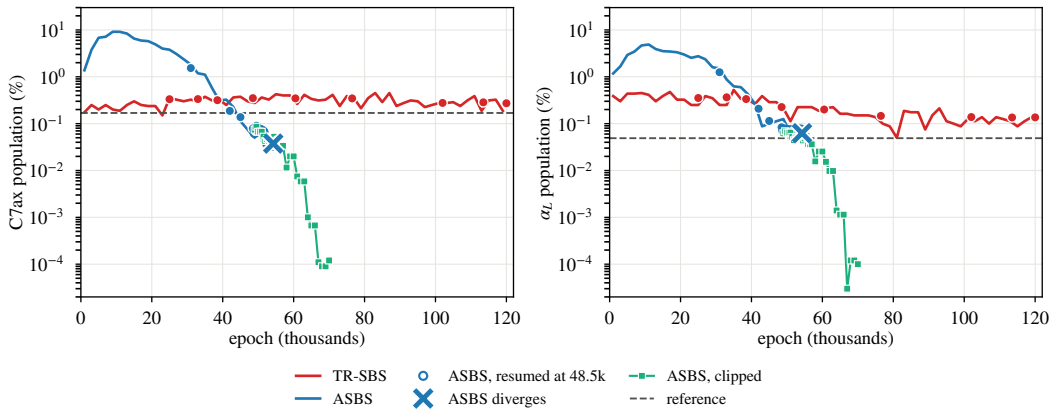


Figure 12: Alanine dipeptide: populations of C7ax (left) and α_L (right) over training for TR-SBS and ASBS with the same prior. Lines: evaluations of 2,000 samples, averaged over 2,000 epochs; dots: 10^7 samples, open dots from ASBS resumed at 48,500 epochs; squares: ASBS with clipped corrector targets, resumed at 48,500 epochs, 10^7 samples, up to 70,000 epochs; dashed: reference; cross: divergence of ASBS without clipping, after which its populations are not shown (see Table 5).

C.7 Lennard-Jones cluster (LJ-13)

We compare TR-SBS with ASBS on the 13-particle Lennard-Jones cluster LJ-13 in three dimensions ($d = 39$) with the harmonic confinement of [30], using 10^4 reference samples. Both methods use the same settings: EGNN controls [43] with 3 layers and 64 hidden units, a mean-free Gaussian prior with scale 2, $\sigma_{\min} = 10^{-3}$, $\sigma_{\max} = 0.5$, $N_t = 200$, $N_{\mathrm{FC}} = 200$, $N_{\mathrm{BC}} = 50$, $L = 100$, $B = 200$, $N_{\mathrm{buf}} = 50{,}000$, $N_{\mathrm{new}} = 1{,}000$, and Adam with learning rate 10^{-4} on one NVIDIA A40 GPU. These are fewer integration steps and a smaller network than in [30]. TR-SBS uses $\varepsilon = 0.2$, $c_{\mathrm{IW}} = 5$,

$N_{\text{state}} = 500$, $N_{\text{path}} = 20$, and the two modifications of Appendix B.4. ASBS is trained for $12{,}500$ epochs and TR-SBS for $19{,}125$ epochs; for both, we evaluate the checkpoint with the lowest mean energy of the generated samples during training, which requires no reference samples. Following [30], we report W_2 between $2{,}000$ model and $2{,}000$ reference samples, where the cost between two configurations is computed after optimal relabeling of the particles and rigid alignment, and W_2 between their energies, $\mathcal{E}(\cdot)W_2$, averaged over three trials.

Table 6 shows that TR-SBS performs on par with ASBS: ASBS attains a slightly lower W_2 and TR-SBS a slightly lower $\mathcal{E}(\cdot)W_2$, and both attain a W_2 close to and an $\mathcal{E}(\cdot)W_2$ below the values reported in [30]. Two independent sets of reference samples have $W_2 = 1.32$, which is the value expected for exact samples of this size. Without the two modifications of Appendix B.4, TR-SBS fails on LJ-13 ($W_2 = 5.24$, with a positive median energy).

Table 6: LJ-13: mean $\pm$ standard deviation over three evaluation trials of $2{,}000$ samples.

Method	$W_2 \downarrow$	$\mathcal{E}(\cdot)W_2 \downarrow$
ASBS	1.571 ± 0.002	0.94 ± 0.07
TR-SBS	1.585 ± 0.001	0.82 ± 0.11
ASBS, reported in [30]	1.59 ± 0.03	1.99 ± 1.01
Reference vs. reference samples	1.32 ± 0.01	0.39 ± 0.04

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [N/A].
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for [N/A]).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will also be asked to include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While [Yes] is generally preferable to [No], it is perfectly acceptable to answer [No] provided a proper justification is given (e.g., error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering [No] or [N/A] is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the contribution (a trust region variant of IPF for sampling) and the scope is fully supported by the theoretical formulations in Section 3 and the numerical experiments in Section 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

Justification: We discuss the limitations in Section 5, including the additional outer iterations and energy evaluations (Appendix B.7), the sensitivity to N_{FC}, N_{BC}, and $\sigma_{\max}$ (Table 3 and Appendix C.4), the local nature of the variance argument, and the missing validation on large-scale molecular systems.

Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All theoretical results (Propositions 3.1 to 3.3 and Proposition 3.4) are formally stated together with their assumptions, and complete proofs, including self-contained auxiliary lemmas, are provided in Appendix A.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Complete details, including SDE formulations, network architectures, evaluation metrics, and hyperparameter tables, are provided in Appendix B to ensure full reproducibility.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code is available at `https://github.com/take-koshizuka/tr_sbs`, and full implementation details and pseudocode are provided in Appendix B.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.

- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All training details, including the optimizer, batch sizes, iterations, and target distributions, are fully documented in Appendix B and Table 2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The results presented in Figures 3 to 5 include mean values and error bars computed across multiple independent random seeds (three seeds and sample standard deviations for GMM, Appendix B.1). Single-seed ablations are marked as such (Table 3).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Detailed information regarding the compute resources, including hardware specifications (NVIDIA A40 GPU with 48GB VRAM), is provided in Appendix B.5, and the numbers of target evaluations are discussed in Appendix B.7.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: This work focuses on foundational theoretical and algorithmic advancements in machine learning and strictly adheres to the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This work is fundamental research on stochastic optimal control and sampling from unnormalized densities. It does not have direct, immediate societal impacts or foreseeable negative applications.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.

- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper introduces theoretical sampling algorithms and evaluates them on synthetic mathematical benchmarks, which do not pose a high risk for misuse.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The empirical evaluations are based entirely on standard, reproducible mathematical distributions and synthetic target potentials.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not introduce new datasets or large-scale models.

    Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The research does not involve crowdsourcing or human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The research does not involve human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not used as a component of the core methodology in this research.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.

- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.