Learning Optimal Flows for Non-Equilibrium Importance Sampling

Yu Cao
Courant Institute of Mathematical Sciences, New York University
yucaoyc@outlook.com

Eric Vanden-Eijnden
Courant Institute of Mathematical Sciences, New York University
eve2@cims.nyu.edu

Abstract

Many applications in computational sciences and statistical inference require the computation of expectations with respect to complex high-dimensional distributions with unknown normalization constants, as well as the estimation of these constants. Here we develop a method to perform these calculations based on generating samples from a simple base distribution, transporting them by the flow generated by a velocity field, and performing averages along these flowlines. This non-equilibrium importance sampling (NEIS) strategy is straightforward to implement, and can be used for calculations with arbitrary target distributions. On the theory side we discuss how to tailor the velocity field to the target and establish general conditions under which the proposed estimator is a perfect estimator, with zero-variance. We also draw connections between NEIS and approaches based on mapping a base distribution onto a target via a transport map. On the computational side we show how to use deep learning to represent the velocity field by a neural network and train it towards the zero variance optimum. These results are illustrated numerically on benchmark examples (with dimension up to 10), where we show that, after training the velocity field, the variance of the NEIS estimator by up to 6 orders of magnitude lower than that of a vanilla estimator. We also compare the performances of NEIS with those of Neal's annealed importance sampling (AIS).

1 Introduction

Given a potential function $U_1 : \Omega \to \mathbb{R}$ on the domain $\Omega \subseteq \mathbb{R}^d$, the main goal of this paper is to evaluate

$$\mathcal{Z}_1 := \int_\Omega e^{-U_1(x)}\,\mathrm{d}x. \tag{1}$$

The calculations of such integrals arise in many applications from several scientific fields. For instance, $\mathcal{Z}_1$ is known as the partition function in statistical physics [22], where it is used to characterize the thermodynamic properties of a system with energy U_1, and as the evidence in Bayesian statistics, where it is used for model selection [13].

When the dimension d of the domain Ω is large, standard numerical quadrature methods are inapplicable to (1) and the method of choice to estimate $\mathcal{Z}_1$ is Monte-Carlo sampling [23, 8]. This requires expressing $\mathcal{Z}_1$ as an expectation, which can be done e.g. by realizing that

$$\mathcal{Z}_1 = \mathbb{E}_0\big[e^{-U_1}/\rho_0\big], \tag{2}$$

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

where $\mathbb{E}_0$ denotes expectation with respect to the probability density function $\rho_0 > 0$. If ρ_0 is both known (i.e. we can evaluate it pointwise in Ω, normalization factor included) and simple to sample from, we can build an estimator for $\mathcal{Z}_1$ by replacing the expectation on the right hand side of (2) by the empirical average of e^{-U_1}/ρ_0 over samples drawn from ρ_0. Unfortunately, finding a density ρ_0 that has the two properties above is hard: unless ρ_0 is well-adapted to e^{-U_1}, the estimator based on (2) is terrible in general, with a standard deviation that is typically much larger than its mean or even infinite. A similar issue arises if we want to estimate the expectation $\mathbb{E}_0 f$ of some function $f : \Omega \to \mathbb{R}$, and the two problems are in fact connected when $f > 0$ since the second reduces to (2) for $U_1 = -\log(f\rho_0)$.

These difficulties have prompted the development of importance sampling strategies [15] whose aim is to produce estimators with a reasonably low variance for $\mathcal{Z}_1$ or $\mathbb{E}_0 f$. These include for example umbrella sampling [44, 42], replica exchange (aka parallel tempering) [15, 24], nested sampling [38, 39], etc., in which the estimation of $\mathcal{Z}_1$ is factorized into the calculation of several expectations of the type (2), but with better properties, that can then be recombined using thermodynamic integration [18] or Bennett acceptance ratio method [6].

Complementary to these equilibrium techniques, non-equilibrium sampling strategies have also been introduced for the calculation of (1). For example, Neal's annealed importance sampling (AIS) [27] based on the Jarzynski equality [16, 17, 1] calculates $\mathcal{Z}_1$ using properly weighted averages over sequences of states evolving from samples from ρ_0, without requiring that the kernel used to generate these states be in detailed-balance with respect to either ρ_0, or $\rho_1 := e^{-U_1}/\mathcal{Z}_1$, or any density interpolating between these two. Instead the weight factors are based on the probability distribution of the sequence of states in path space. Other non-equilibrium sampling strategies in this vein include bridge and path sampling [14], and sequential Monte Carlo (SMC) sampling [25, 2].

In this paper, we analyze another non-equilibrium importance sampling (NEIS) method, originally introduced in [34]. NEIS is based on generating samples from a simple base density ρ_0, then propagating them forward and backward in time along the flowlines of a velocity field, and computing averages along these trajectories—the basic idea of the method is to use the flow induced by this velocity field to sweep samples from ρ_0 through regions in Ω that contribute most to the expectation. As shown in [34] and recalled below, this procedure leads to consistent estimators for the calculation of $\mathcal{Z}_1$ or $\mathbb{E}_0 f$ via a generalization of (2). One advantage of the method, which is a rare feature among importance sampling strategies, is that it leads to estimators that always have lower variance than the vanilla estimator based on (2) [34]. The question we investigate in this paper is how low their variance can be made, both in theory and in practice. Our **main contributions** are:

- Under mild assumptions on U_1 and ρ_0, we show that if the NEIS velocity field is the gradient of a potential that satisfies a Poisson equation, the NEIS estimator for $\mathcal{Z}_1$ has zero variance.
- Under the same assumptions, we show that this optimal flow can be used to construct a perfect transport map from ρ_0 to ρ_1. This allows us to compare NEIS with importance sampling strategies involving transport maps like normalizing flows (NF) that have recently gained popularity [33, 20, 31], and highlight some potential advantages of the former over the latter.
- On the practical side, we derive variational problems for the optimal velocity field in NEIS, and show how to solve these problems by approximating the velocity by a neural network and optimizing its parameters using deep learning training strategies, similar to what is done with neural ODE [9].
- We illustrate the feasibility and usefulness of this approach by testing it on numerical examples. First we consider Gaussian mixtures in up to 10 dimensions. In this context, we show that training the velocity used in NEIS allows to reduce the variance of a vanilla estimator using a standard Gaussian distribution as ρ_0 by up to 6 orders of magnitude. Second we study Neal's 10 dimensional funnel distribution [28, 2] for which the variance of the vanilla importance sampling method is infinity; training a linear dynamics with 2 parameters in NEIS can lead to an estimator with empirical variance less than 1. In these examples we also show that after training NEIS leads to estimators with lower variance than AIS [27].

Related works. The idea of transporting samples from ρ_0 to lower the variance of the vanilla estimator based on (2) is also at the core of importance sampling strategies using normalizing flows (NF) [41, 40, 33, 20, 31, 47, 46, 30, 26]. The type of transport used in NF-based method is however different in nature from the one used in NEIS. With NF, one tries to construct a map that transforms each sample from ρ_0 into a sample from the target $\rho_1 = e^{-U_1}/\mathcal{Z}_1$. In contrast, NEIS uses samples

from ρ_0 as initial conditions to generate trajectories, and uses the data along these entire trajectories to build an estimator. Intuitively, this means that samples likely on ρ_0 must become likely on ρ_1 sometime along these trajectories rather than at a given time specified beforehand, which is easier to enforce.

NEIS bears similarities with Neal's AIS [27], except that in NEIS the sampling is done once from ρ_0 to generate deterministic trajectories to gather data for the estimator, whereas AIS uses random trajectories. There are some methods based on AIS that optimize the transition kernel: for instance, stochastic normalizing flows (SNF) proposed in [47] incorporates NF between annealing steps; and annealed flow transport (AFT) in [2] combines NF with the sequential Monte Carlo method to provide optimized flow transport. These approaches require learning several maps along the annealed transition, whereas the NEIS herein only needs to learn a single flow dynamics.

A time-discrete version of NEIS termed NEO was proposed in [43]. The current implementation of NEO iterates on a map that needs to be prescribed beforehand, but this map could perhaps be optimized using a strategy similar to the one proposed here.

From a practical standpoint, the idea of optimizing the velocity field in NEIS using a neural network approximation for this field can be viewed as an application of neural ODEs [9] that uses the variance of the NEIS estimator as the objective function to minimize. The nature of this objective poses specific challenges in the training procedure, which we investigate here.

Notations. For symmetry, we will denote $\rho_0(x) = e^{-U_0(x)}$ with $U_0 = -\log \rho_0 : \Omega \to \mathbb{R}$ and $\mathcal{Z}_0 = \int_\Omega e^{-U_0(x)}\,\mathrm{d}x = 1$. We denote a d-dimensional vector filled with zeros as $\mathbf{0}_d$ and the $d \times d$ identity matrix as $\mathbb{I}_d$. $\langle \cdot, \cdot \rangle$ is the Euclidean inner product in $\mathbb{R}^d$. We assume that the domain Ω is either an open and connected subset of $\mathbb{R}^d$ with smooth boundary or a d-dimensional torus (without boundary). We denote by $\mathscr{N}(\mu, \Sigma)$ the multivariate Gaussian density with mean μ and covariance matrix Σ. For two functions $f, g : D \to \mathbb{R}$ where D is a domain of interest, the notation $f \lesssim g$ means that there exists a constant $C > 0$ such that $f(x) \leq Cg(x)$ for any $x \in D$. Suppose $\boldsymbol{T} : \Omega \to \mathbb{R}^d$ is a map and μ is a distribution, then the pushforward distribution of μ by the map $\boldsymbol{T}$ is denoted as $\boldsymbol{T}\#\mu$. The notation $|\cdot|$ is the usual ℓ_2 norm for vectors and $\|\cdot\|$ is the matrix norm or functional norm.

2 Flow-based NEIS method

Here we recall the main ingredients of the non-equilibrium importance sampling (NEIS) method proposed in [34]. Let $\boldsymbol{b} : \Omega \to \mathbb{R}^d$ be a velocity field which we assume belongs to the vector space

$$\mathfrak{B} := \Big\{ \boldsymbol{b} \in C^\infty\big(\overline{\Omega}, \mathbb{R}^d\big) \;\Big|\; \boldsymbol{b} \cdot \boldsymbol{n}|_{\partial\Omega} = 0,\ \sup_{x \in \Omega} |\nabla \boldsymbol{b}(x)| < \infty \Big\}, \tag{3}$$

where $\boldsymbol{n}$ is the normal vector at the boundary $\partial\Omega$. Define the associated flow map $\boldsymbol{X}_t : \Omega \to \Omega$ via

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{X}_t(x) = \boldsymbol{b}\left(\boldsymbol{X}_t(x)\right), \qquad \boldsymbol{X}_0(x) = x, \tag{4}$$

and let $\mathcal{J}_t(x)$ be the Jacobian of this map:

$$\mathcal{J}_t(x) := \left|\det\left(\nabla_x \boldsymbol{X}_t(x)\right)\right| \equiv \exp\left(\int_0^t \nabla \cdot \boldsymbol{b}\left(\boldsymbol{X}_s(x)\right)\,\mathrm{d}s \right). \tag{5}$$

Finally, let us denote

$$\mathcal{F}_t^{(k)}(x) := e^{-U_k(\boldsymbol{X}_t(x))} \mathcal{J}_t(x) \tag{6}$$

for $k \in \{0, 1\}$, $x \in \Omega$ and $t \in \mathbb{R}$. NEIS is based on the following result, proven in Appendix B.1:

Proposition 2.1. *If $\boldsymbol{b} \in \mathfrak{B}$, then for any $-\infty < t_- < t_+ < \infty$, we have*

$$\mathcal{Z}_1 = \mathbb{E}_0 \mathcal{A}_{t_-, t_+}, \tag{7}$$

where

$$\mathcal{A}_{t_-, t_+}(x) := \int_{t_-}^{t_+} \frac{\mathcal{F}_t^{(1)}(x)}{\int_{t-t_+}^{t-t_-} \mathcal{F}_s^{(0)}(x)\,\mathrm{d}s}\,\mathrm{d}t. \tag{8}$$

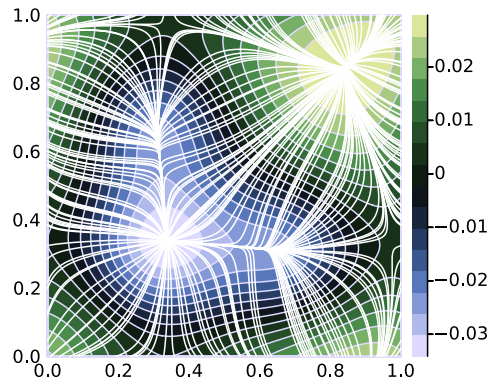


Figure 1: Contour plot of V and flowlines of $\boldsymbol{b} = \nabla V$, where V solves the Poisson's equation (11) with $\mathcal{D} = 1$, assuming that $\rho_0 = 1$ and ρ_1 is a mixture density with 3 modes; see (54). With this $\boldsymbol{b} = \nabla V$, we have $\mathcal{A}(x) = \mathcal{Z}_1$ for almost all $x \in [0,1]^2$.

In addition, if

$$\lim_{\substack{t_- \to -\infty \\ t_+ \to \infty}} \mathcal{A}_{t_-,t_+}(x) = \mathcal{A}(x) := \frac{\int_{\mathbb{R}} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_{\mathbb{R}} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t} \tag{9}$$

exists for almost all $x \sim \rho_0$, then

$$\mathcal{Z}_1 = \mathbb{E}_0 \mathcal{A}. \tag{10}$$

When $\boldsymbol{b} = \mathbf{0}_d$, or $t_- \uparrow 0$ and $t_+ \downarrow 0$, (7) reduces to (2). The aim, however, is to choose $\boldsymbol{b}$ so that the estimator based on (7) has a lower variance than the one based (2): we will show below that this can indeed be done. For now, note that Jensen's inequality implies that an estimator based on (10) for *any* $\boldsymbol{b}$ has lower variance than the one based (2); see [34] or Proposition H.3 below for details.

Note also that, if one allows the magnitude of the flow $\boldsymbol{b}$ to be arbitrarily large, the finite-time NEIS will behave like the infinite-time NEIS; such a relation will be discussed and elaborated in Appendix B.4.

Finally note that the estimator (10) based on (9) is invariant with respect to the parametrization of the flowlines generated by the dynamics $\boldsymbol{b}$, as shown by the following result proved in Appendix B.2:

Proposition 2.2 (An invariance property)**.** *Suppose $\boldsymbol{b}, \alpha\boldsymbol{b} \in \mathfrak{B}$, where $\alpha \in C^\infty(\Omega, \mathbb{R})$ satisfying $\inf_{x\in\Omega} \alpha(x) > 0$. Then the fields $\boldsymbol{b}$ and $\alpha\boldsymbol{b}$ generate the same flowlines, and $\mathcal{A}_{\boldsymbol{b}} = \mathcal{A}_{\alpha\boldsymbol{b}}$ where $\mathcal{A}_{\boldsymbol{b}}$ and $\mathcal{A}_{\alpha\boldsymbol{b}}$ are the function defined in* (9) *using $\boldsymbol{b}$ and $\alpha\boldsymbol{b}$, respectively.*

3 Optimal NEIS

The NEIS estimator for (10) is unbiased no matter what $\boldsymbol{b}$ is. However, its performance relies on the choice of $\boldsymbol{b}$. Therefore, a natural question is to find the field $\boldsymbol{b}$ that achieves the largest variance reduction. The next result shows that an optimal $\boldsymbol{b}$ exists that leads to a zero-variance estimator:

Proposition 3.1 (Existence of zero-variance dynamics)**.** *Assume that $\Omega = [0,1]^d$ and $U_0, U_1 \in C^\infty(\Omega, \mathbb{R})$. Let $\mathcal{D} : \Omega \to (0,\infty)$ be some smooth positive function with $\inf_{x\in\Omega} \mathcal{D}(x) > 0$, and suppose that $V \in C^\infty(\Omega)$ solves the following Poisson's equation on Ω*

$$\nabla \cdot (\mathcal{D}\nabla V) = \rho_1 - \rho_0, \qquad \textit{with} \quad \int_\Omega V(x)\,\mathrm{d}x = 0. \tag{11}$$

If the solution V is a Morse function, then $\boldsymbol{b} = \nabla V$ is a zero-variance velocity field: that is, if we use it to define (9)*, we have*

$$\int_{\mathbb{R}} \rho_0(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\,\mathrm{d}s = \int_{\mathbb{R}} \rho_1(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\,\mathrm{d}s, \qquad \textit{for almost all } x \sim \rho_0, \tag{12}$$

and as a result

$$\mathcal{A}(x) = \mathcal{Z}_1 \quad \textit{for almost all } x \sim \rho_0. \tag{13}$$

This proposition is proven in Appendix E where we also make a connection between (11) and Beckmann's transportation problem. We stress that the optimal $\boldsymbol{b}$ specified in Proposition 3.1 is not unique (see Proposition E.9): however, we show below in Proposition 4.1 that, under certain conditions, all local minima of the variance (viewed as a functional of $\boldsymbol{b}$) are global minima. We also note that the assumption that the solution is a Morse function is mostly a technicality, as discussed in Appendix E.3. Similarly, we consider the torus in Proposition 3.1 for simplicity mainly; we expect that the proposition will hold in general when Ω has compact closure or even when $\Omega = \mathbb{R}^d$, see Appendix E for examples including that of Gaussian mixture distributions.

For illustration, the contour plot of V and the flowlines of $\boldsymbol{b} = \nabla V$ are shown in Figure 1 in a simple example in a two-dimension torus where $\rho_0 = 1$ and ρ_1 is a mixture density with 3 modes; their explicit expressions are given in (54); in this example, we solved (11) numerically with $\mathcal{D} = 1$, see Appendix E.5 for more details. Some other examples where the zero-variance dynamics is explicit are discussed in Appendix F.

Connection to transport maps and normalizing flows. The zero-variance dynamics provides a transport map $\boldsymbol{T}$ from ρ_0 to ρ_1, as shown in:

Proposition 3.2 (Existence of a perfect generator)**.** *Suppose $\mathcal{D} = 1$ for simplicity. Under the same assumption as in Proposition 3.1, let V be the Morse function solving* (11) *and $\boldsymbol{b} = \nabla V$ the associated zero-variance dynamics. Then there exists a continuously differentiable function $\varkappa$ (defined almost everywhere on Ω) such that*

$$\int_{-\infty}^{0} \rho_0(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\,\mathrm{d}s = \int_{-\infty}^{\varkappa(x)} \rho_1(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\,\mathrm{d}s. \tag{14}$$

Furthermore, the map $\boldsymbol{T}(x) := \boldsymbol{X}_{\varkappa(x)}(x)$ is a transport map from ρ_0 to ρ_1, i.e., $\boldsymbol{T}\#\rho_0 = \rho_1$.

The proof is given in Appendix G.1: Note that we consider again $\boldsymbol{b} = \nabla V$ on the torus for technical simplicity: the statement of the proposition should hold in general for a zero-variance dynamics $\boldsymbol{b}$. The solution of (14) is particularly simple in one-dimension, where we can take $\boldsymbol{b}(x) = 1$, and straightforwardly verify that $\varkappa(x) = \boldsymbol{T}(x) - x$ with

$$\boldsymbol{T}(x) = F_1^{-1}(F_0(x)) \qquad \text{where} \quad F_i(x) = \int_{-\infty}^{x} \rho_i(y)dy, \;\; i = 0, 1. \tag{15}$$

We also illustrate the statement of Proposition 3.2 via numerical examples in Appendix G.2.

To avoid confusion, we stress that we will *not* use the transport map $\boldsymbol{T}$ of Proposition 3.2 in the examples below. Indeed, using this map would require identifying $\varkappa$, which introduces an unnecessary additional calculation which we can avoid using the NEIS estimator directly. In addition, the NEIS estimator will likely have better properties than those based on transport maps, as we can we can think of NEIS as using a time-parameterized family of transport maps rather than a single one. In particular, the variance of the NEIS estimator will be small if samples likely on ρ_0 become likely on ρ_1 *sometime* along the NEIS trajectories, rather than at the same fixed time for all samples. The former seems easier to fulfill than the latter. For example, in one-dimension, the NEIS estimator has zero variance for any $\boldsymbol{b}$ bounded away from zero, whereas building a transport map from ρ_0 to ρ_1 is already nontrivial in that simple case since it requires solving (15).

4 Variational formulations

The Poisson equation (11) admits a variational formulation:

$$\min_V \int_\Omega \left(\tfrac{1}{2}|\nabla V|^2\mathcal{D} + V(\rho_1 - \rho_0)\right)\,\mathrm{d}x. \tag{16}$$

If $\mathcal{D}$ is chosen to be a probability density function (for example $\mathcal{D} = \rho_0$ or $\mathcal{D} = \frac{1}{2}(\rho_1 + \rho_0)$) the two terms in the objective in (16) are expectations which can be estimated via sampling (using e.g. direct sampling for the expectation with respect to ρ_0 and NEIS for the one with respect to ρ_1). This means

that we can in principle use an MCMC estimator of (16) as empirical loss, and minimize it over all V in some parametric class. Here however, we will follow a different strategy that allows us to directly parametrize $\boldsymbol{b}$ instead of V (i.e. relax the requirement that $\boldsymbol{b} = \nabla V$) and simply use the variance of the estimator as objective function.

Specifically, since we quantify the performance of the estimators based on (7) and (10) by their variance, we can view these quantities as functionals of $\boldsymbol{b}$ that we wish to minimize. Since the estimators are unbiased, these objectives are

$$\begin{aligned} \mathrm{Var}_{t_-,t_+}(\boldsymbol{b}) &= \mathcal{M}_{t_-,t_+}(\boldsymbol{b}) - \mathcal{Z}_1^2, \qquad \text{(finite-time);} \\ \mathrm{Var}(\boldsymbol{b}) &= \mathcal{M}(\boldsymbol{b}) - \mathcal{Z}_1^2, \qquad \text{(infinite-time),} \end{aligned} \tag{17}$$

where we defined the second moments $\mathcal{M}_{t_-,t_+}(\boldsymbol{b}) := \mathbb{E}_0\big[|\mathcal{A}_{t_-,t_+}|^2\big]$ and $\mathcal{M}(\boldsymbol{b}) := \mathbb{E}_0\big[|\mathcal{A}|^2\big]$. With the finite-time objective, we know that with $\boldsymbol{b} = \mathbf{0}_d$, (7) reduces to (2). Therefore minimizing $\mathcal{M}_{t_-,t_+}(\boldsymbol{b})$ over $\boldsymbol{b}$ by gradient descent starting from $\boldsymbol{b}$ near $\mathbf{0}_d$ will necessarily produce a better estimator: while we cannot guarantee that the variance of this optimized estimator will be zero, the experiments conducted below indicate that it can be a several order of magnitude below that of the vanilla estimator.

For the infinite-time objective, we know that for any $\boldsymbol{b}$, (9) leads to an estimator with a lower variance than the one based on (2) [34]. Minimizing $\mathrm{Var}(\boldsymbol{b})$ over $\boldsymbol{b}$ using gradient descent leads to a local minimum; the next result shows that all such local minima are global:

Proposition 4.1 (Global minimum)**.** *Under some technical assumptions listed in Proposition H.1, if $\boldsymbol{b}_* \in \mathfrak{B}$ is a local minimum of Var$(\boldsymbol{b})$ where the functional derivative of Var$(\boldsymbol{b})$ with respect to $\boldsymbol{b}$ vanishes, i.e. δVar$(\boldsymbol{b}_*)/\delta\boldsymbol{b} = \mathbf{0}_d$ on Ω, then $\boldsymbol{b}_*$ is a global minimum and Var$(\boldsymbol{b}_*) = 0$.*

The expression of the functional derivative $\delta\mathrm{Var}(\boldsymbol{b}_*)/\delta\boldsymbol{b}$ is given in Proposition D.1. The technical assumptions under which Proposition 4.1 hold are explained in Appendix A and the proof is given in Appendix H.

5 Training towards the optimal $\boldsymbol{b}$

Here we discuss how to use deep learning techniques to find the optimal $\boldsymbol{b}$; these techniques will be illustrated on numerical examples in Section 6. Some technical details are deferred to Appendix I.

Objective. We use the finite-time objective $\mathcal{M}_{t_-,t_+}(\boldsymbol{b})$ in (7) with $t_- \in [-1, 0], t_+ = t_- + 1$. Two natural choices are $t_- = 0$ and $t_- = -1/2$, which will be used below in the numerical experiments. This leads to no loss of generality *a priori* since in the training scheme we put no restriction on the magnitude that $\boldsymbol{b}$ can reach, and with large $\boldsymbol{b}$ the flow line can travel large distance even during $t \in [-1, 1]$ (the range of integration in s, t in (8)); see the discussion in Appendix B.4 for more details. In practice we use a time-discretized version of (8) with $2N_t$ discretization points, and use the standard Runge-Kutta scheme of order 4 (RK4) to integrate the ODE (4) over $t \in [-1, 1]$ using uniform time step ($\Delta t = 1/N_t$). We note that this numerical discretization introduces a bias. However, this bias can be systematically controlled by changing the time step or using higher order integrator. In our experiments, we observed that the RK4 integrator led to negligible errors, see Table 5.

Neural architecture. In our experiments, we either parameterize $\boldsymbol{b}$ by a deep neural network directly, or we assume that $\boldsymbol{b}$ is a gradient field,

$$\boldsymbol{b}(x) = \nabla V(x) \qquad \text{(gradient form),}$$

and parameterize the potential V by a deep neural network. We always use an ℓ-layer neural network with width m for all inner layers; therefore, from now on, we simply refer the neural network structure by a pair (ℓ, m); see Appendix I.2 for more details. When we parametrize the potential function V instead of $\boldsymbol{b}$, the only difference is that the output dimension of the neural network becomes 1 instead of d. The activation function is chosen as the *softplus* function, which is a smooth version of ReLU that gave better empirical results compared to the sigmoid function. At initialization the neural parameters were randomly generated. Theoretical results about the gradient of $\mathcal{M}_{t_-,t_+}(\boldsymbol{b})$ with respect to parameters are given in Appendix C and corresponding numerical implementations are explained in Appendix I.

Direct training method. We minimize $\mathcal{M}_{t_-,t_+}(\boldsymbol{b})$ with respect to the parameters in the neural network using stochastic gradient descent (SGD) in which we evaluate the loss and its gradiant empirically using mini-batches of data drawn from ρ_0 at every iteration step.

Assisted training method. When local minima of U_1 are far away from the local minimum of U_0 at $x = 0$, the direct training method by sampling data from ρ_0 and minimizing $\mathcal{M}_{t_-,t_+}(\boldsymbol{b})$ fails, because the flowlines may not reach the importance region of ρ_1 due to poor initialization of $\boldsymbol{b}$. More specifically, if along almost all trajectories, $e^{-U_1(\boldsymbol{X}_s(x))} \approx 0$ for $s \in [-1, 1]$, then with large probability $\mathcal{A}_{t_-,t_+}(x) \approx 0$ where $x \sim \rho_0$; as a result the empirical variance of the estimator can be extremely small if the number of samples is small, while the true variance could be extremely large. Such a *mode collapse* phenomenon is common in rare event simulations.

To get around this difficulty, recall that ideally we would like to find a dynamics $\boldsymbol{b}$ such that $\mathcal{A}$ is approximately a constant function in the infinite-time case. That is, if $\boldsymbol{b}$ is a zero-variance dynamics, then $\mathcal{A}$ is a constant function and $\mathbb{E}_p\big[(\mathcal{A} - (\mathbb{E}_p\mathcal{A})^2)\big] = 0$ for *any* distribution p, and we are not constrained to use the base distribution ρ_0 and minimize the functional $\boldsymbol{b} \to \mathrm{Var}(\boldsymbol{b})$ in (17). Motivated by this idea, we use an assisted training scheme in which, at iteration i of SGD, the loss function is

$$\mathbb{E}_{p_i}\Big[\big(\mathcal{A}_{t_-,t_+} - \mathbb{E}_{p_i}\mathcal{A}_{t_-,t_+}\big)^2\Big]. \tag{18}$$

Here $\mathbb{E}_{p_i}$ denotes expectation with respect to the probability density p_i defined as

$$p_i = (1 - c_i)\rho_0 + c_i \boldsymbol{Z}\#\rho_0, \qquad c_i = \max\big\{c - i\frac{c}{\upsilon L}, 0\big\}, \tag{19}$$

where $\upsilon \in (0, 1)$ controls the number of steps during which the training is assisted, $c \in (0, 1)$, L is the total number of training steps, and $\boldsymbol{Z} := \boldsymbol{Z}_{t=1}$ is the time-1 map of the ODE $\dot{\boldsymbol{Z}}_t(x) = -\varsigma\nabla U_1\big(\boldsymbol{Z}_t(x)\big)$ with $\boldsymbol{Z}_{t=0}(x) = x$ and $\varsigma > 0$ is a parameter. In essence, using (18) means that, for the first υL training steps, there is a probability c_i that the data $x \sim \rho_0$ are replaced by $\boldsymbol{Z}(x)$, so that the training method can better explore important regions near local minima of U_1. Subsequently, the assistance is turned off so that some subtle adjustment can be made. If some samples from $\rho_1 = e^{-U_1}/\mathcal{Z}_1$ were available beforehand, we could equivalently replace $\boldsymbol{Z}\#\rho_0$ in (19) by the empirical distribution over these samples. We emphasize that the assisted training method is only used to guide the training initially and the NEIS estimator for $\mathcal{Z}_1$ is unbiased.

6 Numerical experiments

We consider three benchmark examples to illustrate the effectiveness of NEIS assisted with training. The first two examples involve Gaussian mixtures, for which we use NEIS with $t_- = 0$; the third example is Neal's funnel distribution, for which we use NEIS with $t_- = -1/2$. In all examples, we compare the performance of NEIS with those of annealed importance sampling (AIS) [27]; the number of transition steps in AIS is denoted as K and we refer to this method as AIS-K below; for more details see Appendix I.1. For the comparison, we choose to record the query costs to U_1 and ∇U_1 as the measurement of computational complexity, which connects to the framework in the *theory of information complexity* (see e.g., [29]). The runtime could depend on coding, machine condition, etc., whereas query complexity more or less only depends on the computational problem (U_0, U_1 and $\boldsymbol{b}$) itself; for most examples of interest, U_0 is simple whereas U_1 and its derivatives will be expensive to compute.

For simplicity, we always use as base density $\rho_0(x) = (2\pi)^{-d/2}e^{-\frac{1}{2}|x|^2}$. We remark that a better choice of ρ_0 (i.e., more adapted to ρ_1) can significantly improve the sampling performance; our ρ_0 is precisely used to validate the performance of NEIS in situations where ρ_0 is *not* well chosen. It would be interesting to study how to adapt the choice of ρ_0 for easier training in NEIS, but this is left for future investigations.

When presenting results, we rescale the estimates so that the exact value is $\mathcal{Z}_1 = 1$ for all examples. More implementation details about training are deferred to Appendix I.2. All trainings and estimates of $\mathcal{Z}_1$ are conducted on a laptop with CPU i7-12700H; we use 15 threads at maximum. The runtime of training ranges approximately between 40$\sim$80 seconds for Gaussian mixture (2D), 9.5$\sim$12 minutes for Gaussian mixture (10D); for the Funnel distribution (10D), the runtime is around 25 minutes for a generic linear ansatz and around 2 minutes for a two-parameter ansatz. Appendix J includes

additional figures about training. When computing the gradient of the variance with respect to parameters, we use an integration-based method when $t_- = -1/2$ (for the convenience of numerical implementation) and use an ODE-based method when $t_- = 0$ (for higher accuracy); details about these two approximation methods are given in Appendix I.3 and Appendix I.4 respectively. The codes are accessible on `https://github.com/yucaoyc/NEIS`.

An asymmetric 2-mode Gaussian mixture in 2D. As a first illustration, we consider an asymmetric 2-mode Gaussian mixture

$$e^{-U_1} \propto \frac{1}{5}\mathcal{N}(\lambda \boldsymbol{e}_1, \sigma_1^2 \mathbb{I}_d) + \frac{4}{5}\mathcal{N}(-\lambda \boldsymbol{e}_2, \sigma_1^2 \mathbb{I}_d) \tag{20}$$

where $\boldsymbol{e}_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, $\boldsymbol{e}_2 = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, $\sigma_1 = \sqrt{0.1}$, $\lambda = 5$. With this choice of parameters, the variance of the vanilla estimator based on (2) is approximately 1.85×10^6. We use NEIS with $t_- = 0$ and set the time step to $\Delta t = 1/50$ for ODE discretization during both training and estimation of $\mathcal{Z}_1$. We train over $L = 50$ SGD steps using the loss (18) by imposing bias in the first 60% the training period only (i.e. with $\upsilon = 0.6$). The evolution of the variance during the training is shown in Figure 7 in Appendix; the best optimized flow has a variance of about 1, as opposed to 10^6 for the vanilla estimator. Since p_i in (19) is quite different from ρ_0 during the assisted learning period, it may happen that the empirical variance significantly exceeds the variance of the vanilla importance sampling; this does not contradict with Proposition H.3 below. As seen in Figure 7, at the end of the assisted period, the variance is already quite small and in most cases, the variance continues to reduce as $\boldsymbol{b}$ gets further optimized.

After training, we estimate $\mathcal{Z}_1$ using NEIS with the optimized flow and compare its performance with AIS-10 and AIS-100. We first record the query cost for training and then set a total number of queries to $U_1, \nabla U_1$ as budgets. Given the query budget, we estimate $\mathcal{Z}_1$ using each method for 10 times, leading to the results given in Table 5 below. When we determine the estimation cost of NEIS, we deduct the query cost of training from the total query budget for fairer comparison. Note that NEIS uses less total queries to produce more accurate estimate of $\mathcal{Z}_1$: in particular, the standard deviation of estimating $\mathcal{Z}_1$ by NEIS method is around 1 magnitude smaller than AIS-100. Moreover, the bias from ODE discretization appears to be negligible. Figure 2 show an optimized flow and also provide a visual comparison of NEIS with AIS under fixed query budget; more comparison using various ansatzes or architectures can be found in Figures 11 and 12.

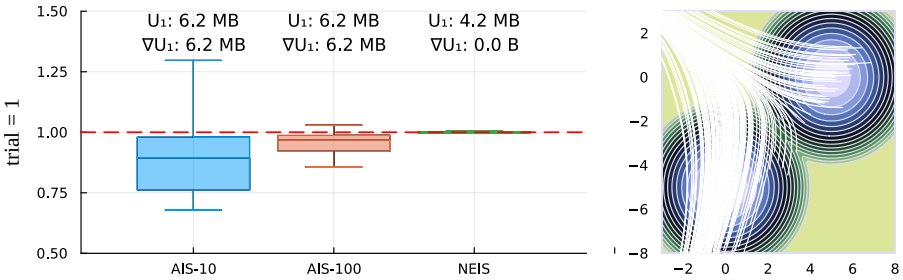
A symmetric 4-mode Gaussian mixture in 10D. Next we consider a symmetric 4-mode Gaussian mixture in $d = 10$ dimension with

$$e^{-U_1(x)} \propto \sum_{i=1}^{4} \mathcal{N}(\mu_i, \Sigma) \tag{21}$$

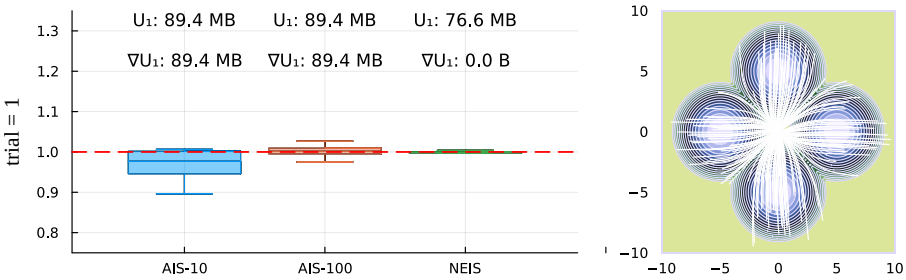
where the vector $\mu_i = \left[\, \lambda\cos\left(\frac{i\pi}{2}\right)\ \lambda\sin\left(\frac{i\pi}{2}\right)\ 0\ 0\ \cdots\ 0 \,\right]$ and $\Sigma = \mathrm{Diag}\left[\, \sigma_1^2\ \sigma_1^2\ \sigma_2^2\ \sigma_2^2\ \cdots\ \sigma_2^2 \,\right]$ is a diagonal matrix. The parameters are $d = 10$, $\sigma_1 = \sqrt{0.1}$, $\sigma_2 = \sqrt{0.5}$ and $\lambda = 5$. With this choice of parameters, the variance of the vanilla estimator based on (2) is approximately 2.15×10^6. We use NEIS with $t_- = 0$ and the time step $\Delta t = 1/60$ is used for ODE discretization during both training and estimation of $\mathcal{Z}_1$. We show the training result in Figure 8. The variance reduces to about 10 after 60 SGD steps for the gradient ansatz (here we only considered this ansatz as it produces more promising empirical results).

Similar to the last example, we compare NEIS using the optimized $\boldsymbol{b}$ with AIS, under fixed query budgets; see Table 5. The best result from NEIS has an estimator with the standard deviation less than $1/3$ of the one from AIS-100, even though NEIS uses less query cost. This comparison suggests that AIS-100 needs more than 9 times more resources than NEIS with optimized flow in order to achieve similar precision and the cost spent on training indeed pays off if we require an accurate estimate of $\mathcal{Z}_1$ (meaning less fluctuation for Monte Carlo estimates). Moreover, this table also shows that the bias from ODE discretization is negligible.

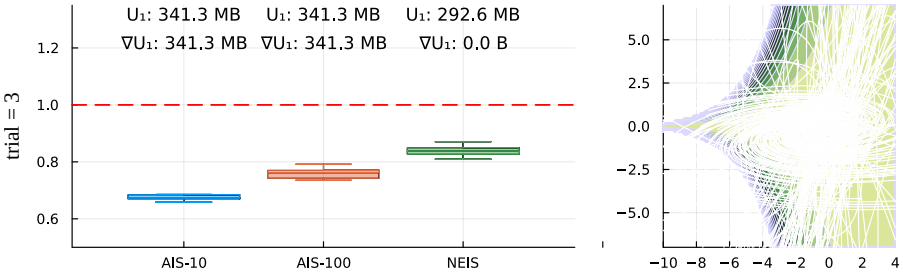
Figure 2 shows a particular optimized flow: as can be seen, the mass near the origin flows towards four local minima of U_1, as we would intuitively expect. More optimized flows and comparisons can be found in Figurer 13 in Appendix.



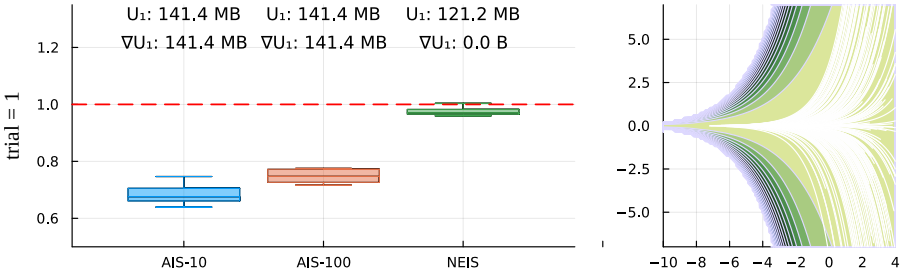
(a) Asymmetric Gaussian mixture (2D): gradient ansatz, $\ell = 2$, $m = 20$



(b) Symmetric Gaussian mixture (10D): gradient ansatz, $\ell = 2$, $m = 30$



(c) Funnel distribution (10D): generic linear ansatz



(d) Funnel distribution (10D): two-parameter ansatz

Figure 2: Selective comparison results for various models. Left panels: estimates of $\mathcal{Z}_1$ by AIS and NEIS with optimized flow under the fixed query budget shown above the panels; we repeat these calculations 10 times and show boxplots of these 10 estimates for each method. The trial number refers to the index for randomly chosen initialization. The query numbers refer to the queries used for each estimate of $\mathcal{Z}_1$ for each method; 1MB $= 10^6$. The dashed red lines show the exact value $\mathcal{Z}_1 = 1$. Right panels: streamlines of optimized flows atop the contours of U_1, both projected into the x_1-x_2 plane for the two 10D examples. Full comparison and figures can be found in Figures 11,12,13, and 14 in the Appendix.

A funnel distribution in 10D. We consider the following 10D funnel distribution studied in [2, 28]: for the state $x = [x_1, x_2, \ldots, x_{10}] \in \mathbb{R}^{10}$,

$$x_1 \sim \mathscr{N}(0, 9), \qquad x_i \sim \mathscr{N}(0, e^{x_1}), \qquad 2 \le i \le d.$$

For numerical stability, we consider the above funnel distribution restricted to a unit ball centered at the origin with radius 25. Instead of heuristically parameterizing the dynamics via neural-networks, we consider a generic linear ansatz and a two-parametric linear ansatz:

$$\boldsymbol{b}(x) = W_1 x + b_1, \qquad W_1 \in \mathbb{R}^{10\times 10}, b_1 \in \mathbb{R}^{10}, \tag{22a}$$

$$\boldsymbol{b}(x) = -[\beta, \alpha x_2, \alpha x_3, \ldots, \alpha x_{10}], \qquad \alpha, \beta \in \mathbb{R}. \tag{22b}$$

The generic linear ansatz can be regarded as a neural network without inner layers. With (22a), we drawn the entries in the matrix W_1 randomly and we set $b_1 = \mathbf{0}_{10}$ initially, and we use the assisted training method; with (22b), we set $\alpha = \beta = 2$ initially, and we use the direct training method. In both cases, we choose the finite-NEIS scheme with $t_- = -1/2$. We notice that the asymmetric choice $t_- = 0$ can also leads into more optimal dynamics, but its performance is not as competitive as the symmetric case $t_- = -1/2$. It is very likely that such a difference is due to the structure of funnel distribution: each coordinate x_i ($1 \le i \le 10$) has mean 0 and therefore, a symmetric version can probably better weight the contribution from both forward and backward flows.

The training results are shown in Figures 9 and 10 in Appendix: we can observe that the error and variance are overall decreasing during the training and the parameters α, β tend to increase with a similar speed. We use the same protocol as in the two previous examples to compare NEIS with AIS. As can be be observed in Table 5, the two-parameter ansatz (22b) gives the best estimate; the generic linear ansatz (22a) is not as competitive (probably due to over-parameterization), but it still outperforms the AIS-100, under fixed query budget. Figure 2 shows these optimized flows; more results can be found in Figure 14 in the Appendix. The apparent gap between estimates and the ground truth in Figure 2 (or see Table 5) comes from insufficient sample size.

7 Conclusion and outlook

In this work, we revisited the NEIS strategy proposed in [34] and analyzed its capabilities, both from theoretical and computational standpoints. Regarding the former, we showed that NEIS leads to a zero-variance estimator for a velocity field $\boldsymbol{b} = \nabla V$ with a potential V that satisfies a certain Poisson equation with the difference between the target and the base density as source. Moreover, a zero-variance dynamics can be used to construct a transport map from ρ_0 to ρ_1.velocity In turn, we highlighted the connection and difference between NEIS and importance sampling strategies based on the normalizing flows (NF).

On the computational side, we showed that the variance of the NEIS estimator can be used as objective function to train the velocity field $\boldsymbol{b}$. This training procedure can be performed in practice by approximating the velocity field by a neural network, and optimizing the parameters in this network using SGD, similar to what is done in the context of neural ODE but with a different objective. Our numerical experiments showed that this strategy is effective and can lower the variance of a vanilla estimator for $\mathcal{Z}_1$ by several orders of magnitude.

While the numerical examples we used in the present paper are somewhat academic, the results suggest that NEIS has potential in more realistic settings. In order to explore other applications, it would be interesting to investigate how to best parametrize $\boldsymbol{b}$ (e.g., less parameters and non-stiff energy landscape with respect to these parameters) and how to best initiate the training procedure. It would also be interesting to ask whether we can improve the performance of NEIS by optimizing certain parameters in the base density ρ_0 in concert with $\boldsymbol{b}$. The answers to these questions are probably model specific and are left for future work.

Acknowledgment

We would like to thank Jonathan Weare and Fang-Hua Lin for helpful discussions, and the anonymous referees for their useful comments and suggestions. The work of EVE is supported by the National Science Foundation under awards DMR-1420073, DMS-2012510, and DMS-2134216, by the Simons Collaboration on Wave Turbulence, Grant No. 617006, and by a Vannevar Bush Faculty Fellowship.

References

[1] Christophe Andrieu, James Ridgway, and Nick Whiteley. Sampling normalizing constants in high dimensions using inhomogeneous diffusions, 2018. arXiv: 1612.07583.

[2] Michael Arbel, Alex Matthews, and Arnaud Doucet. Annealed flow transport Monte Carlo. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 318–330. PMLR, 2021.

[3] Larry Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pac. J. Math.*, 16(1):1–3, 1966.

[4] Michèle Audin and Mihai Damian. *Morse Theory and Floer Homology*. Springer London, 2014. doi: 10.1007/978-1-4471-5496-9.

[5] Martin Beckmann. A continuous model of transportation. *Econometrica*, 20(4):643–660, 1952.

[6] Charles H Bennett. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.*, 22(2):245–268, 1976.

[7] O. Bräunling. Fourier series on the n-dimensional torus, 2004. `https://www.uni-math.gwdg.de/schick/teach/jvnfour1.pdf`.

[8] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press LLC, 2011.

[9] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[10] Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural ODEs. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[11] Weinan E, Chao Ma, and Lei Wu. The Barron space and the flow-induced function spaces for neural network models, 2019. arXiv: 1906.08039.

[12] Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, 2010.

[13] F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *Mon. Not. R. Astron. Soc.*, 384(2):449–463, 2008. doi: 10.1111/j.1365-2966.2007.12353.x.

[14] Andrew Gelman and Xiao-Li Meng. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.*, 13(2):163–185, 1998. doi: 10.1214/ss/1028905934. Number: 2.

[15] Charles J. Geyer. Importance sampling, simulated tempering, and umbrella sampling. In Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, editors, *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

[16] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56(5):5018–5035, 1997. doi: 10.1103/PhysRevE.56.5018.

[17] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78(14): 2690–2693, 1997. doi: 10.1103/PhysRevLett.78.2690.

[18] John G Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3(5):300–313, 1935.

[19] Achim Klenke. *Probability Theory: A Comprehensive Course*. 2014.

[20] Ivan Kobyzev, Simon J. D. Prince, and Marcus A. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 1–1, 2020. doi: 10.1109/TPAMI.2020.2992934.

[21] Peter D. Lax. Change of variables in multiple integrals II. *Am. Math. Mon.*, 108(2):115–119, 2001. doi: 10.2307/2695524.

[22] Evgenii Mikhailovich Lifshitz and Lev Petrovich Pitaevskii. *Statistical Physics: Theory of the Condensed State*, volume 9. Elsevier, 2013.

[23] Jun S Liu. *Monte Carlo Strategies in Scientific Computing*, volume 10. Springer, 2001.

[24] Jianfeng Lu and Eric Vanden-Eijnden. Methodological and computational aspects of parallel tempering methods in the infinite swapping limit. *J. Stat. Phys.*, 2019. doi: 10.1007/s10955-018-2210-y.

[25] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, 68(3):411–436, 2006. doi: 10.1111/j.1467-9868.2006.00553.x.

[26] Thomas Müller, Brian Mcwilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. Neural importance sampling. *ACM Trans. Graph.*, 38(5):145:1–145:19, 2019. doi: 10.1145/3341156.

[27] Radford M. Neal. Annealed importance sampling. *Stat. Comput.*, 11(2):125–139, 2001. doi: 10.1023/A:1008923215028.

[28] Radford M. Neal. Slice sampling. *Ann. Stat.*, 31(3), 2003. doi: 10.1214/aos/1056562461.

[29] Erich Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis*. Lecture Notes in Mathematics. Springer-Verlag, 1988. doi: 10.1007/BFb0079792.

[30] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457), 2019. doi: 10.1126/science.aaw1147.

[31] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.*, 22(57):1–64, 2021.

[32] Philipp Petersen. Neural network theory, 2020.

[33] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 1530–1538, Lille, France, 2015.

[34] Grant M. Rotskoff and Eric Vanden-Eijnden. Dynamical computation of the density of states and Bayes factors using nonequilibrium importance sampling. *Phys. Rev. Lett.*, 122(15):150602, 2019. doi: 10.1103/PhysRevLett.122.150602.

[35] Walter Rudin. *Principles of Mathematical Analysis*. 1976.

[36] Filippo Santambrogio. A dacorogna-moser approach to flow decomposition and minimal flow problems. *ESAIM: ProcS*, 45:265–274, 2014. doi: 10.1051/proc/201445027.

[37] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*. Birkhäuser, Cham, 2015.

[38] John Skilling. Nested sampling. *AIP Conf. Proc.*, 735(1):395–405, 2004. doi: 10.1063/1.1835238.

[39] John Skilling. Nested sampling for general Bayesian computation. *Bayesian Anal.*, 1(4):833–859, 2006. doi: 10.1214/06-BA127.

[40] E. G. Tabak and Cristina V. Turner. A family of nonparametric density estimation algorithms. *Commun. Pure Appl. Math.*, 66(2):145–164, 2013. doi: 10.1002/cpa.21423.

[41] Esteban G. Tabak and Eric Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. *Commun. Math. Sci.*, 8(1):217–233, 2010.

[42] Erik H. Thiede, Brian Van Koten, Jonathan Weare, and Aaron R. Dinner. Eigenvector method for umbrella sampling enables error analysis. *J. Chem. Phys.*, 145(8):084115, 2016. doi: 10.1063/1.4960649.

[43] Achille Thin, Yazid Janati, Sylvain Le Corff, Charles Ollion, Arnaud Doucet, Alain Durmus, Eric Moulines, and Christian Robert. NEO: Non equilibrium sampling on the orbit of a deterministic transform, 2021. arXiv:2103.10943.

[44] Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[45] Sharan Vaswani, Aaron Mishkin, Issam Laradji, Mark Schmidt, Gauthier Gidel, and Simon Lacoste-Julien. Painless stochastic gradient: Interpolation, line-search, and convergence rates. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[46] Peter Wirnsberger, Andrew J. Ballard, George Papamakarios, Stuart Abercrombie, Sébastien Racanière, Alexander Pritzel, Danilo Jimenez Rezende, and Charles Blundell. Targeted free energy estimation via learned mappings. *J. Chem. Phys.*, 153(14):144112, 2020. doi: 10.1063/5.0018903.

[47] Hao Wu, Jonas Köhler, and Frank Noe. Stochastic normalizing flows. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 5933–5944, 2020.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] It can be found in `https://github.com/yucaoyc/NEIS`.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] Details can be found in Section 5, Section 6 or Appendix I.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] We alternatively use the boxplot for multiple independent experiments in Section 6.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] see Section 6.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A] Our examples do not involve special data and we only use standard Julia packages that are widely available. The dependence of the code will be explained in README.md in the code (`https://github.com/yucaoyc/NEIS`).
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A The functional space for the infinite-time case

A.1 Assumptions

For simplicity, we make:

Assumption A.1. (i) The domain Ω is either

- an open and connected subset of $\mathbb{R}^d$ with smooth boundary,
- or a d-dimensional torus (without boundary).

(ii) $U_0,\ U_1 \in C^\infty(\Omega, \mathbb{R})$.

(iii) $\mathcal{Z}_0 := \int_\Omega e^{-U_0} = 1$, $\mathcal{Z}_1 < \infty$ and $\mathrm{Var}^{(\max)} \in (0, \infty)$, where

$$\mathrm{Var}^{(\max)} := \mathbb{E}_{\rho_0}[e^{-2(U_1 - U_0)}] - (\mathcal{Z}_1)^2$$

is the variance for the vanilla importance sampling method.

Notations

The open ball around x with radius r is denoted as $B_r(x) := \big\{y \in \Omega \;\big|\; |y - x| < r\big\}$. For any subset $D \subset \Omega$, let $\tau_D^+(x)$ and $\tau_D^-(x)$ be the first hitting times to the boundary ∂D in the forward and backward directions, respectively. More specifically,

$$\begin{aligned} \tau_D^+(x) &:= \inf\big\{t \geq 0 : \boldsymbol{X}_t(x) \in \partial D\big\}, \\ \tau_D^-(x) &:= \sup\big\{t \leq 0 : \boldsymbol{X}_t(x) \in \partial D\big\}. \end{aligned} \tag{23}$$

For later convenience, let us denote

$$\mathcal{B}(x) := \int_{-\infty}^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t, \tag{24}$$

and thus

$$\mathcal{A}(x) \overset{(9)}{=} \frac{\int_{-\infty}^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\mathcal{B}(x)}.$$

A.2 Preliminaries

Lemma A.2. *If $\boldsymbol{b} \in \mathfrak{B}$ in* (3)*, then for any $t, s \in \mathbb{R}$, $x \in \Omega$ and $k = 0, 1$,*

$$\mathcal{J}_t\big(\boldsymbol{X}_s(x)\big) = \mathcal{J}_{t+s}(x)/\mathcal{J}_s(x), \qquad \mathcal{F}_t^{(k)}\big(\boldsymbol{X}_s(x)\big) = \mathcal{F}_{t+s}^{(k)}(x)/\mathcal{J}_s(x). \tag{25}$$

This lemma can be verified easily by definition and thus the proof is omitted. A simple consequence of (25) is the following result that we also state without proof:

Lemma A.3. *For any $s \in \mathbb{R}$ and $x \in \Omega$,*

$$\mathcal{B}\big(\boldsymbol{X}_s(x)\big) = \frac{\mathcal{B}(x)}{\mathcal{J}_s(x)}, \qquad \mathcal{A}\big(\boldsymbol{X}_s(x)\big) = \mathcal{A}(x), \tag{26}$$

provided that these terms are well-defined.

A.3 The functional space

For the infinite-time case, we consider the following family of vector fields denoted as $\mathfrak{B}_\infty$.

Definition A.4. $\mathfrak{B}_\infty$ is a set that contains all $\boldsymbol{b} \in \mathfrak{B}$ such that $\Omega \backslash \mho(\boldsymbol{b})$ has Lebesgue measure zero, where $\mho(\boldsymbol{b}) \subset \Omega$ is the collection of points x at which the functions

$$z \to \int_0^\infty \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t, \qquad z \to \int_{-\infty}^0 \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t \tag{27}$$

are continuous on a local neighborhood of x for $k = 0, 1$.

We use the notation $\mho(\boldsymbol{b})$ because in general, such a subset depends on the choice of $\boldsymbol{b}$. The main reason behind the above definition is that we need functions in (27) to behave nicely almost everywhere. The integrability of $t \to \mathcal{F}_t^{(k)}(x)$ depends on the long-term behavior of the flow, which is not easy to characterize in general; thus we simply include the integrability into the assumption. However, we can indeed expect that the above conditions should hold for most interesting examples, e.g.:

Example A.5. If $U_0(x) = \frac{|x|^2}{2} + \frac{d}{2}\ln(2\pi)$ and $U_1(x) = \frac{|x|^2}{2\sigma^2}$ (so that both ρ_0 and ρ_1 are Gaussian densities), one optimal flow is $\boldsymbol{b}(x) = x$ (cf. Appendix F below). For this choice, we can direct compute that when $z \neq \mathbf{0}_d$, $\boldsymbol{X}_t(z) = e^t z$ and hence

$$\int_0^\infty \mathcal{F}_t^{(1)}(z)\,\mathrm{d}t = \frac{1}{2}\Big(\frac{|z|^2}{2\sigma^2}\Big)^{-d/2}\int_{\frac{|z|^2}{2\sigma^2}}^\infty s^{\frac{d}{2}-1}e^{-s}\,\mathrm{d}s,$$

$$\int_{-\infty}^0 \mathcal{F}_t^{(1)}(z)\,\mathrm{d}t = \frac{1}{2}\Big(\frac{|z|^2}{2\sigma^2}\Big)^{-d/2}\int_0^{\frac{|z|^2}{2\sigma^2}} s^{\frac{d}{2}-1}e^{-s}\,\mathrm{d}s,$$

and similar expressions hold for $\mathcal{F}^{(0)}$ by letting $\sigma = 1$ in above equations. Then clearly, $\mho(\boldsymbol{b}) = \Omega\backslash\{\mathbf{0}_d\}$ and such a dynamics belongs to $\mathfrak{B}_\infty$. However, the constant function $\boldsymbol{b} = \mathbf{0}_d \notin \mathfrak{B}_\infty$: we can easily verify that e.g., for this dynamics, $\int_0^\infty \mathcal{F}_t^{(1)}(z)\,\mathrm{d}t = \infty$ for any z and hence $\mho(\mathbf{0}_d) = \emptyset$; see also Lemma A.7 below.

Here are a few immediate properties of the set $\mho(\boldsymbol{b})$:

Lemma A.6. *We have:*

(i) $\mho(\boldsymbol{b})$ *is an open subset of* Ω.

(ii) The set $\mho(\boldsymbol{b})$ *is closed under the evolution of the dynamics* $\boldsymbol{b}$*, i.e., if* $x^\star \in \mho(\boldsymbol{b})$*, then* $\boldsymbol{X}_t(x^\star) \in \mho(\boldsymbol{b})$ *for any* $t \in \mathbb{R}$.

(iii) If $x^\star \in \overline{\mho(\boldsymbol{b})}$*, then* $\boldsymbol{X}_t(x^\star) \in \overline{\mho(\boldsymbol{b})}$ *for any* $t \in \mathbb{R}$*, where* $\overline{\mho(\boldsymbol{b})}$ *is the closure of* $\mho(\boldsymbol{b})$ *in* $\mathbb{R}^d$.

(iv) $x \to \int_{-\infty}^\infty \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t$ *is continuous on* $\mho(\boldsymbol{b})$.

(v) $(x,t) \to \int_t^\infty \mathcal{F}_s^{(k)}(x)\,\mathrm{d}s$ *is continuous on* $\mho(\boldsymbol{b}) \times \mathbb{R}$.

Proof. The part (i) follows easily from the definition of $\mho(\boldsymbol{b})$ and the part (iv) also trivially holds.

For the part (ii), notice that if $y = \boldsymbol{X}_t(x)$, then

$$\begin{aligned}\int_0^\infty \mathcal{F}_s^{(k)}(y)\,\mathrm{d}s &= \int_0^\infty \mathcal{F}_s^{(k)}\big(\boldsymbol{X}_t(x)\big)\,\mathrm{d}s \overset{(25)}{=} \Big(\int_0^\infty \mathcal{F}_{t+s}^{(k)}(x)\,\mathrm{d}s\Big)/\mathcal{J}_t(x)\\ &= \Big(\int_0^\infty \mathcal{F}_s^{(k)}(x)\,\mathrm{d}s - \int_0^t \mathcal{F}_s^{(k)}(x)\,\mathrm{d}s\Big)/\mathcal{J}_t(x)\\ &= \Big(\int_0^\infty \mathcal{F}_s^{(k)}\big(\boldsymbol{X}_{-t}(y)\big)\,\mathrm{d}s - \int_0^t \mathcal{F}_s^{(k)}\big(\boldsymbol{X}_{-t}(y)\big)\,\mathrm{d}s\Big)/\mathcal{J}_t\big(\boldsymbol{X}_{-t}(y)\big).\end{aligned}$$

Since $(x,t) \to \boldsymbol{X}_t(x)$, $(x,t) \to \mathcal{J}_t(x)$ are continuous by the smoothness assumption of $\boldsymbol{b}$, it is clear that $\mathcal{J}_t\big(\boldsymbol{X}_{-t}(y)\big)$ is continuous on Ω. The continuity of $y \to \int_0^t \mathcal{F}_s^{(k)}\big(\boldsymbol{X}_{-t}(y)\big)\,\mathrm{d}s$ also trivially holds on Ω. Next, the continuity of $y \to \int_0^\infty \mathcal{F}_s^{(k)}\big(\boldsymbol{X}_{-t}(y)\big)\,\mathrm{d}s$ in a local neighborhood of $\boldsymbol{X}_t(x^\star)$ comes from the assumption that $x^\star \in \mho(\boldsymbol{b})$. Thus, $y \to \int_0^\infty \mathcal{F}_s^{(k)}(y)\,\mathrm{d}s$ is continuous in a local neighborhood of $\boldsymbol{X}_t(x^\star)$, if $x^\star \in \mho(\boldsymbol{b})$. The other case for $y \to \int_{-\infty}^0 \mathcal{F}_s^{(k)}(y)\,\mathrm{d}s$ can be similarly verified.

The part (iii) is an immediate consequence of part (ii). Let us prove it by contradiction. Assume that the conclusion in part (iii) does not hold, then there exists $x^\star \in \overline{\mho(\boldsymbol{b})}$ such that $y^\star := \boldsymbol{X}_t(x^\star) \notin \overline{\mho(\boldsymbol{b})}$ for some $t \in \mathbb{R}$. By the part (ii), we know this $x^\star \notin \mho(\boldsymbol{b})$ and thus $x^\star \in \partial\mho(\boldsymbol{b})$ is in the boundary. As $\overline{\mho(\boldsymbol{b})}$ is closed and $y^\star \notin \overline{\mho(\boldsymbol{b})}$, there must exist an $\epsilon > 0$ such that for any $y \in \mathbb{R}^d$ with $|y - y^\star| < \epsilon$,

we have $y \notin \overline{\mho(\boldsymbol{b})}$. By the smoothness assumption on $\boldsymbol{b}$, we know there exists a $\delta > 0$ such that for any $x \in B_\delta(x^\star)$, we have $|\boldsymbol{X}_t(x) - y^\star| < \epsilon$ outside of $\overline{\mho(\boldsymbol{b})}$. Since $x^\star$ is in the boundary of $\mho(\boldsymbol{b})$, we must be able to find an $x \in \mho(\boldsymbol{b})$ such that $x \in B_\delta(x^\star)$ and by part (ii), $\boldsymbol{X}_t(x) \in \mho(\boldsymbol{b})$. Thus, we reach a contradiction.

For the part (v),

$$\int_t^\infty \mathcal{F}_s^{(k)}(x)\,\mathrm{d}s = \int_0^\infty \mathcal{F}_{t+s}^{(k)}(x)\,\mathrm{d}s \overset{(25)}{=} \mathcal{J}_t(x) \int_0^\infty \mathcal{F}_s^{(k)}\big(\boldsymbol{X}_t(x)\big)\,\mathrm{d}s.$$

Because $(x,t) \to \boldsymbol{X}_t(x)$ is continuous (by the smoothness of $\boldsymbol{b} \in \mathfrak{B}$) and $z \to \int_0^\infty \mathcal{F}_s^{(k)}(z)\,\mathrm{d}s$ is continuous in a local neighborhood of $\boldsymbol{X}_t(x) \in \mho(\boldsymbol{b})$ by the part (ii), it is then clear that $\int_0^\infty \mathcal{F}_s^{(k)}\big(\boldsymbol{X}_t(x)\big)\,\mathrm{d}s$ is continuous with respect to (x,t), and hence the conclusion follows easily. □

A notable family of points that are excluded from $\mho(\boldsymbol{b})$ are points at which $\int_{-\infty}^\infty \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t = \infty$. For instance, these include stationary points of $\boldsymbol{b}$ and periodic orbits, which we state as:

Lemma A.7. *Given a vector field $\boldsymbol{b} \in \mathfrak{B}$, we have $x \notin \mho(\boldsymbol{b})$,*

(i) if x is a stationary point of $\boldsymbol{b}$ (i.e., $\boldsymbol{b}(x) = \boldsymbol{0}_d$), or

(ii) if x is on a periodic orbit.

Proof. If x is a stationary point of $\boldsymbol{b}$, the trajectory is $\boldsymbol{X}_t(x) = x$ for all $t \in \mathbb{R}$. Then it is obvious that $\int_{-\infty}^\infty \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t = e^{-U_k(x)} \int_{-\infty}^\infty e^{(\boldsymbol{\nabla}\cdot\boldsymbol{b})(x)t}\,\mathrm{d}t = \infty$, which establishes the part (i). For the part (ii), let us assume that the period is τ. If $c := \int_0^\tau \nabla\cdot\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s \geq 0$, then

$$\begin{aligned}
\int_0^\infty \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t &= \int_0^\infty e^{-U_k\left(\boldsymbol{X}_t(x)\right) + \int_0^t \boldsymbol{\nabla}\cdot\boldsymbol{b}\left(\boldsymbol{X}_s(x)\right)\,\mathrm{d}s}\,\mathrm{d}t \\
&\geq \int_0^\infty e^{-\max_{0\leq s\leq\tau} U_k\left(\boldsymbol{X}_s(x)\right) + c\lfloor t/\tau\rfloor + \int_{\tau\lfloor t/\tau\rfloor}^{t-\tau\lfloor t/\tau\rfloor} \nabla\cdot\boldsymbol{b}\left(\boldsymbol{X}_s(x)\right)\,\mathrm{d}s}\,\mathrm{d}t \\
&\geq C\int_0^\infty e^{c(t/\tau-1)}\,\mathrm{d}t = \infty,
\end{aligned}$$

where the constant $C = e^{-\max_{0\leq s\leq\tau} U_k\left(\boldsymbol{X}_s(x)\right)} e^{\min_{0\leq s\leq\tau}\int_0^s \nabla\cdot\boldsymbol{b}\left(\boldsymbol{X}_r(x)\right)\,\mathrm{d}r}$. Therefore, $\int_{-\infty}^\infty \mathcal{F}_t^{(k)}(x) = \infty$. If $c := \int_0^\tau \nabla\cdot\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s < 0$, we can similarly show that $\int_{-\infty}^0 \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t = \infty$, and the same conclusion holds. □

A.4 Perturbation of the dynamics

Next we investigate the following question: given $\boldsymbol{b} \in \mathfrak{B}_\infty$ and $\delta\boldsymbol{b} \in C_c^\infty(\Omega, \mathbb{R}^d)$, can we guarantee that $\boldsymbol{b} + \epsilon\delta\boldsymbol{b} \in \mathfrak{B}_\infty$ for small enough ϵ? Such a conclusion trivially holds for the finite-time case; however, more underlying structures are needed for the infinite-time case, due to the fact that the long-term behavior of the flow $\boldsymbol{b}$ can sensitively depend on the local perturbation $\delta\boldsymbol{b}$. This question is probably unavoidable in order to understand the mathematical structure of $\mathfrak{B}_\infty$.

Notation: We shall use the notation $\boldsymbol{X}_t(\cdot)$ to represent the flow map under $\boldsymbol{b}$, and use $\boldsymbol{X}_t^\epsilon(\cdot)$ to represent the flow map under $\boldsymbol{b}^\epsilon := \boldsymbol{b} + \epsilon\delta\boldsymbol{b}$ in this section below. Moreover, we shall use $\mathcal{F}_\cdot^{(k,\epsilon)}(\cdot)$ to denote the function defined in (6) for the dynamics $\boldsymbol{b}^\epsilon$ and the Jacobian of the flow $\mathcal{J}_{(\cdot)}^\epsilon(\cdot)$ is similarly defined.

Definition A.8 (*b*-stability)**.** Given $\boldsymbol{b} \in \mathfrak{B}_\infty$:

(a) (For an open bounded set). A nonempty open bounded set $D \subset \mho(\boldsymbol{b})$ is said to be ***b**-stable* if:

(i) There exists a point $x^\star \in D$ and $\zeta \in (0,1)$ such that

$$|\boldsymbol{b}(x^\star)| > 0, \qquad |\boldsymbol{b}(x) - \boldsymbol{b}(x^\star)| < \zeta|\boldsymbol{b}(x^\star)|, \qquad \forall x \in D; \tag{28}$$

(ii) For any $x \in D$, the trajectory $\{\boldsymbol{X}_t(x)\}_{t\in\mathbb{R}}$ intersects with the boundary ∂D at exactly two points.

(b) (For a point). A point $x \in \mho(\boldsymbol{b})$ is said to be $\boldsymbol{b}$-stable if there exists a neighborhood $B_\epsilon(x)$ such that the region $B_\epsilon(x)$ is $\boldsymbol{b}$-stable. Otherwise, the point x is said to be $\boldsymbol{b}$-unstable.

The assumption in the part (i) ensures that the trajectory $t \to \boldsymbol{X}_t(x)$ will leave this region D within a finite amount of time; see Lemma A.9 below. The part (ii) is used to ensure that the trajectory is not (infinitely) recurrent to D; once the trajectory leaves D, it will not return to D again.

Lemma A.9. *Suppose $\boldsymbol{b} \in \mathfrak{B}$ and $D \subset \Omega$ is open and bounded. If there exists a point $x^\star \in D$ and $\zeta \in (0,1)$ such that*

$$|\boldsymbol{b}(x^\star)| > 0, \qquad |\boldsymbol{b}(x) - \boldsymbol{b}(x^\star)| < \zeta|\boldsymbol{b}(x^\star)|, \qquad \forall x \in D;$$

then for any $x \in D$, we have

$$\tau_D^+(x) - \tau_D^-(x) \le \frac{\textit{Diameter}(D)}{(1-\xi)|\boldsymbol{b}(x^\star)|} < \infty.$$

Proof. Consider the quantity $h_t(x) := \big\langle \boldsymbol{b}(x^\star), \boldsymbol{X}_t(x) - x^\star \big\rangle$ for $x \in D$. Then when $t \in \big(\tau_D^-(x), \tau_D^+(x)\big)$,

$$\frac{\mathrm{d}}{\mathrm{d}t} h_t(x) = \big\langle \boldsymbol{b}(x^\star), \boldsymbol{b}(\boldsymbol{X}_t(x)) - \boldsymbol{b}(x^\star) + \boldsymbol{b}(x^\star) \big\rangle \ge (1-\zeta)|\boldsymbol{b}(x^\star)|^2 > 0.$$

Therefore,

$$\begin{aligned}(1-\xi)|\boldsymbol{b}(x^\star)|^2\big(\tau_D^+(x) - \tau_D^-(x)\big) &\le \int_{\tau_D^-(x)}^{\tau_D^+(x)} \frac{\mathrm{d}}{\mathrm{d}t} h_t(x)\,\mathrm{d}t \\ &= h_{\tau_D^+(x)}(x) - h_{\tau_D^-(x)}(x) \le |\boldsymbol{b}(x^\star)|\ \mathrm{Diameter}(D).\end{aligned}$$

Then the conclusion can be immediately obtained. □

We now state the main result of this section:

Proposition A.10. *Suppose $\boldsymbol{b} \in \mathfrak{B}_\infty$ and $x^\star \in \mho(\boldsymbol{b})$ is $\boldsymbol{b}$-stable. Then there exists an open bounded neighborhood of $x^\star$, denoted as $D \subset \mho(\boldsymbol{b})$, such that for an arbitrary $\delta\boldsymbol{b} \in C_c^\infty(D, \mathbb{R}^d)$, there exists an $\epsilon_0 > 0$ and $\boldsymbol{b} + \epsilon\delta\boldsymbol{b} \in \mathfrak{B}_\infty$ for any $\epsilon \in (0, \epsilon_0)$.*

Proof. The main idea is that if the path $\{\boldsymbol{X}_t(x)\}_{t\in\mathbb{R}}$ passes through D, then a small perturbation within a bounded time period does not affect the long-term behavior; if the path does not pass through D, then $\boldsymbol{b}^\epsilon = \boldsymbol{b}$ along this path and therefore, $\int_0^\infty \mathcal{F}_t^{(k,\epsilon)}(x)\,\mathrm{d}t = \int_0^\infty \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t$ and hence, the continuity is also preserved locally around x.

Step (I): Setup and the choice of D.

Since $x^\star \in \mho(\boldsymbol{b})$, we know $|\boldsymbol{b}(x^\star)| > 0$ by Lemma A.7. Moreover, we can find a small $\boldsymbol{b}$-stable ball $B_\theta(x^\star)$ by Definition A.8 with a parameter $\zeta \in (0,1)$ in (28). Then consider the following cross-section within $B_\theta(x^\star)$

$$S := \big\{y :\ |y - x^\star| < \theta/2, \qquad \langle y - x^\star, \boldsymbol{b}(x^\star)\rangle = 0\big\},$$

and define the streamtube $\mathfrak{T}$ passing through S as

$$\mathfrak{T} := \big\{\boldsymbol{X}_t(y) :\ y \in S,\ t \in \mathbb{R}\big\}.$$

It is not hard to see that $\mathfrak{T}$ is an open subset of $\mho(\boldsymbol{b})$ by Lemma A.6. Then let us choose D as an arbitrary open ball around $x^\star$ such that

$$D \subset \mathfrak{T} \cap B_{\theta/2}(x^\star).$$

Next let us consider an arbitrary $\delta\boldsymbol{b} \in C_c^\infty(D, \mathbb{R}^d)$ and from now on, we shall fix $\delta\boldsymbol{b}$. It is easy to verify that $\boldsymbol{b} + \epsilon\delta\boldsymbol{b} \in \mathfrak{B}$ for any $\epsilon \in \mathbb{R}$. The non-trivial part is to check that $\Omega\backslash\mho(\boldsymbol{b}^\epsilon)$ has measure zero for sufficiently small ϵ and hence $\boldsymbol{b}^\epsilon \in \mathfrak{B}_\infty$.

Step (II): Choice of ϵ_0.

Let us pick

$$\epsilon_0 = \min\Big\{\frac{|\boldsymbol{b}(x^\star)|}{1+|\delta\boldsymbol{b}(x^\star)|},\ \frac{(\zeta^\star-\zeta)|\boldsymbol{b}(x^\star)|}{2\|\delta\boldsymbol{b}\|_{L^\infty(D)}+\zeta^\star|\delta\boldsymbol{b}(x^\star)|}\Big\} > 0, \tag{29}$$

for any $\zeta^\star \in (\zeta, 1)$. The main motivation is that we need (28) to hold for $\boldsymbol{b}^\epsilon$ as well. Indeed, if $\epsilon \le \epsilon_0$,

$$|\boldsymbol{b}^\epsilon(x^\star)| = |\boldsymbol{b}(x^\star)+\epsilon\delta\boldsymbol{b}(x^\star)| \ge |\boldsymbol{b}(x^\star)| - \epsilon|\delta\boldsymbol{b}(x^\star)| \overset{(29)}{\ge} \frac{|\boldsymbol{b}(x^\star)|}{1+|\delta\boldsymbol{b}(x^\star)|} > 0,$$

and for any $x \in B_\theta(x^\star)$,

$$\begin{aligned}|\boldsymbol{b}^\epsilon(x)-\boldsymbol{b}^\epsilon(x^\star)| &\le |\boldsymbol{b}(x)-\boldsymbol{b}(x^\star)| + \epsilon|\delta\boldsymbol{b}(x)-\delta\boldsymbol{b}(x^\star)| \\ &\overset{(28)}{\le} \zeta|\boldsymbol{b}(x^\star)| + \epsilon(2\|\delta\boldsymbol{b}\|_{L^\infty(D)}) \\ &\overset{(29)}{<} \zeta^\star|\boldsymbol{b}(x^\star)| - \zeta^\star\epsilon|\delta\boldsymbol{b}(x^\star)| \\ &\le \zeta^\star|\boldsymbol{b}^\epsilon(x^\star)|.\end{aligned}$$

As discussed above, this property ensures that any trajectory $\big\{\boldsymbol{X}_t^\epsilon(x)\big\}_{t\in\mathbb{R}}$ with $x \in B_\theta(x^\star)$ will pass through the boundary $\partial B_\theta(x^\star)$, and at the same time, since $B_\theta(x^\star)$ is $\boldsymbol{b}$-stable, the condition (ii) in Definition A.8 ensures that such a trajectory only intersects with the boundary $\partial B_\theta(x^\star)$ at exactly 2 points.

Let us denote

$$\tau := \sup_{\epsilon\in[0,\epsilon_0)} \frac{\text{Diameter}(D)}{(1-\xi^\star)|\boldsymbol{b}^\epsilon(x^\star)|} \le \frac{\text{Diameter}(D)\big(1+|\delta\boldsymbol{b}(x^\star)|\big)}{(1-\xi^\star)|\boldsymbol{b}(x^\star)|} < \infty.$$

By Lemma A.9, we know

$$\boldsymbol{X}_t^\epsilon(x) \notin B_\theta(x^\star), \qquad \forall x \in B_\theta(x^\star),\ \forall\epsilon\in[0,\epsilon_0),\ \forall t \in (-\infty,-\tau]\cup[\tau,\infty). \tag{30}$$

Step (III): Prove that $\Omega\backslash\mho(\boldsymbol{b}^\epsilon)$ has measure zero for any $\epsilon \in (0,\epsilon_0)$.

We will prove that for any $\epsilon \in (0,\epsilon_0)$,

$$\mathfrak{T} \subset \mho(\boldsymbol{b}^\epsilon), \qquad \mho(\boldsymbol{b})\backslash\overline{\mathfrak{T}} \subset \mho(\boldsymbol{b}^\epsilon).$$

It could be observed that the boundary $\partial\mathfrak{T}$ has measure zero: the boundary contains flowlines from a hyper-surface with dimension $d-2$, that is, $\partial\mathfrak{T}$ is a set of flowlines passing through $\big\{y:\ |y-x^\star| = \theta/2,\ \langle y-x^\star, \boldsymbol{b}(x^\star)\rangle = 0\big\}$. Hence, provided that the above equation holds, we immediately know that

$$\begin{aligned}\Omega\backslash\mho(\boldsymbol{b}^\epsilon) \equiv \mho(\boldsymbol{b}^\epsilon)^c \subset \big(\mathfrak{T}\cup(\mho(\boldsymbol{b})\backslash\overline{\mathfrak{T}})\big)^c &= \mathfrak{T}^c\cap(\mho(\boldsymbol{b})\cap\overline{\mathfrak{T}}^c)^c = \mathfrak{T}^c\cap(\mho(\boldsymbol{b})^c\cup\overline{\mathfrak{T}}) \\ &= (\mathfrak{T}^c\cap\mho(\boldsymbol{b})^c)\cup(\mathfrak{T}^c\cap\overline{\mathfrak{T}}) \subset \mho(\boldsymbol{b})^c\cup\partial\mathfrak{T} = \big(\Omega\backslash\mho(\boldsymbol{b})\big)\cup\partial\mathfrak{T}\end{aligned}$$

has measure zero, where the superscript c means set complement. As a remark, from now on, we shall fix $\epsilon \in (0,\epsilon_0)$.

Proof of $\mathfrak{T} \subset \mho(\boldsymbol{b}^\epsilon)$. Let us pick an arbitrary $x \in \mathfrak{T}$ and we shall prove that $z \to \int_0^\infty \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t$ is continuous locally near x. Similarly, we can verify that $z \to \int_{-\infty}^0 \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t$ is locally continuous near x. Therefore, $x \in \mho(\boldsymbol{b}^\epsilon)$ and thus $\mathfrak{T} \subset \mho(\boldsymbol{b}^\epsilon)$.

Next we return to verify that $z \to \int_0^\infty \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t$ is continuous locally near x. We claim that there exists a $y \in S\cup D \in B_{\theta/2}(x^\star)$ and $s \in \mathbb{R}$ such that $\boldsymbol{X}_s^\epsilon(y) = x$. To prove this, we need to discuss two cases:

- Suppose $\big\{\boldsymbol{X}_t^\epsilon(x)\big\}_{t\in\mathbb{R}}$ never enters D. Because $\boldsymbol{b}^\epsilon = \boldsymbol{b}$ on $\Omega\backslash D$, we know $\boldsymbol{X}_t^\epsilon(x) = \boldsymbol{X}_t(x)$ for any $t \in \mathbb{R}$. By the definition of the streamtube $\mathfrak{T}$, there exists a $y \in S$ and $s \in \mathbb{R}$ such that $y = \boldsymbol{X}_{-s}(x) = \boldsymbol{X}_{-s}^\epsilon(x)$, which immediately gives $\boldsymbol{X}_s^\epsilon(y) = x$.

- Suppose $\{\boldsymbol{X}_t^\epsilon(x)\}_{t\in\mathbb{R}}$ enters D at some time. Then the above conclusion follows easily.

Because $\boldsymbol{b}^\epsilon$ is smooth, for small enough δ, we can ensure that $B_\delta(x) \subset \mathfrak{T}$ and also $\boldsymbol{X}_{-s}^\epsilon(z) \in B_{\theta/2}(x^\star)$ for any $z \in B_\delta(x)$. By (30),

$$\boldsymbol{X}_{t-s}^\epsilon(z) = \boldsymbol{X}_t^\epsilon\big(\boldsymbol{X}_{-s}^\epsilon(z)\big) \notin B_\theta(x^\star), \qquad \forall z \in B_\delta(x),\ \forall t \geq \tau. \tag{31}$$

We divide the proof of continuity of $z \to \int_0^\infty \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t$ into two cases:

(a) If $\tau \leq s$, then we already know for any point $z \in B_\delta(x)$, we have $\boldsymbol{X}_t^\epsilon(z) \notin B_\theta(x^\star)$ for $t \geq 0$. Recall that $\boldsymbol{b} = \boldsymbol{b}^\epsilon$ outside of $B_\theta(x^\star)$. Hence $\boldsymbol{X}_t^\epsilon(z) = \boldsymbol{X}_t(z)$ for any $t \geq 0$, and

$$\int_0^\infty \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t = \int_0^\infty \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t,$$

which is continuous on a neighbor of x by $x \in \mho(\boldsymbol{b})$.

(b) If $\tau > s$, then by (31), we know $\boldsymbol{X}_t^\epsilon(\boldsymbol{X}_{\tau-s}^\epsilon(z)) \notin B_\theta(x^\star)$ for any $z \in B_\delta(x)$ and $t \geq 0$ and in particular, $\boldsymbol{X}_{\tau-s}^\epsilon(z) \notin B_\theta(x^\star)$ for any $z \in B_\delta(x)$. Let us rewrite

$$\begin{aligned}
\int_0^\infty \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t &= \int_0^{\tau-s} \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t + \int_{\tau-s}^\infty \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t \\
&= \int_0^{\tau-s} \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t + \int_0^\infty \mathcal{F}_{t+(\tau-s)}^{(k,\epsilon)}\big(z\big)\,\mathrm{d}t \\
&\overset{(25)}{=} \int_0^{\tau-s} \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t + \mathcal{J}_{\tau-s}^\epsilon(z) \int_0^\infty \mathcal{F}_t^{(k,\epsilon)}\big(\boldsymbol{X}_{\tau-s}^\epsilon(z)\big)\,\mathrm{d}t.
\end{aligned}$$

Since $\boldsymbol{b}^\epsilon$ is smooth, apparently $z \to \int_0^{\tau-s} \mathcal{F}_t^{(k,\epsilon)}(z)\,\mathrm{d}t$ and $z \to J_{\tau-s}^\epsilon(z)$ are continuous on Ω. The continuity of $z \to \int_0^\infty \mathcal{F}_t^{(k,\epsilon)}\big(\boldsymbol{X}_{\tau-s}^\epsilon(z)\big)\,\mathrm{d}t$ locally near x can be exactly proved in the same way as the Part (a) above for the new point $\boldsymbol{X}_{\tau-s}^\epsilon(x)$.

One small technical result to verify is that $\boldsymbol{X}_{\tau-s}^\epsilon(x) \in \mho(\boldsymbol{b})$ in order to apply the same argument from the Part (a). Note that $\boldsymbol{X}_{\tau-s}^\epsilon(x) = \boldsymbol{X}_\tau^\epsilon(y) = \boldsymbol{X}_{\tau-\widetilde{\tau}}^\epsilon\big(\boldsymbol{X}_{\widetilde{\tau}}^\epsilon(y)\big)$ where

$$\widetilde{\tau} := \inf\big\{t \geq 0 \;\big|\; \boldsymbol{X}_t^\epsilon(y) \in \partial B_\theta(x^\star)\big\}.$$

Since $\boldsymbol{X}_{\widetilde{\tau}}^\epsilon(y) \in \partial B_\theta(x^\star)$, we know $\boldsymbol{X}_{\widetilde{\tau}}^\epsilon(y) \in \mho(\boldsymbol{b})$ (e.g., by choosing a small enough θ). Outside of $B_\theta(x^\star)$, we know $\boldsymbol{b}^\epsilon = \boldsymbol{b}$ and thus $\boldsymbol{X}_{\tau-\widetilde{\tau}}^\epsilon\big(\boldsymbol{X}_{\widetilde{\tau}}^\epsilon(y)\big) = \boldsymbol{X}_{\tau-\widetilde{\tau}}\big(\boldsymbol{X}_{\widetilde{\tau}}^\epsilon(y)\big) \in \mho(\boldsymbol{b})$ by Lemma A.6.

Proof of $\mho(\boldsymbol{b})\backslash\overline{\mathfrak{T}} \subset \mho(\boldsymbol{b}^\epsilon)$. Let us consider an arbitrary point $x \in \mho(\boldsymbol{b})\backslash\overline{\mathfrak{T}}$. Note that $\boldsymbol{b}^\epsilon = \boldsymbol{b}$ outside of $D \subset \mathfrak{T}$. For a local neighborhood $B_\delta(x)$ outside of $\overline{\mathfrak{T}}$, we also know for any $y \in B_\delta(x)$, $\boldsymbol{X}_t^\epsilon(y) = \boldsymbol{X}_t(y) \notin \overline{\mathfrak{T}}$ for $t \in \mathbb{R}$, by both the definition of $\mathfrak{T}$ and the construction $\boldsymbol{b}^\epsilon = \boldsymbol{b}$ outside of $\mathfrak{T}$. By the same argument as in the Part (a) above, it could be readily shown that $x \in \mho(\boldsymbol{b}^\epsilon)$ and thus $\mho(\boldsymbol{b})\backslash\overline{\mathfrak{T}} \subset \mho(\boldsymbol{b}^\epsilon)$. □

B Supplementary material for Section 2

B.1 Proof of Proposition 2.1

We first prove the finite-time case. As $\boldsymbol{b} \in \mathfrak{B}$, we know that $\int_{t_-}^{t_+} \mathcal{F}_{-s}^{(0)}(x)\,\mathrm{d}s \equiv \int_{t_-}^{t_+} e^{-U_0\left(\boldsymbol{X}_{-s}(x)\right)} \mathcal{J}_{-s}(x)\,\mathrm{d}s < \infty$ by the continuity assumption of $\boldsymbol{b}$ and Assumption A.1. Then

$$\begin{aligned}
\mathcal{Z}_1 &= \int_\Omega e^{-U_1(x)} \frac{\int_{t_-}^{t_+} \mathcal{F}_{-t}^{(0)}(x)\,\mathrm{d}t}{\int_{t_-}^{t_+} \mathcal{F}_{-s}^{(0)}(x)\,\mathrm{d}s}\,\mathrm{d}x \\
&= \int_{t_-}^{t_+} \int_\Omega e^{-U_1(x)} \frac{\mathcal{F}_{-t}^{(0)}(x)}{\int_{t_-}^{t_+} \mathcal{F}_{-s}^{(0)}(x)\,\mathrm{d}s}\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned}$$

By the change of variables $x = \boldsymbol{X}_t(\widetilde{x})$, we have

$$
\begin{aligned}
\mathcal{Z}_1 &= \int_{t_-}^{t_+} \int_\Omega e^{-U_1\left(\boldsymbol{X}_t(\widetilde{x})\right)} \frac{\mathcal{F}_{-t}^{(0)}(\boldsymbol{X}_t(\widetilde{x}))}{\int_{t_-}^{t_+} \mathcal{F}_{-s}^{(0)}(\boldsymbol{X}_t(\widetilde{x}))\,\mathrm{d}s} \mathcal{J}_t(\widetilde{x})\,\mathrm{d}\widetilde{x}\,\mathrm{d}t \\
&\overset{(25)}{=} \int_{t_-}^{t_+} \int_\Omega e^{-U_1\left(\boldsymbol{X}_t(\widetilde{x})\right)} \frac{\mathcal{F}_0^{(0)}(\widetilde{x})}{\int_{t_-}^{t_+} \mathcal{F}_{t-s}^{(0)}(\widetilde{x})\,\mathrm{d}s} \mathcal{J}_t(\widetilde{x})\,\mathrm{d}\widetilde{x}\,\mathrm{d}t \\
&= \int_{t_-}^{t_+} \int_\Omega e^{-U_1\left(\boldsymbol{X}_t(\widetilde{x})\right)} \mathcal{J}_t(\widetilde{x}) \frac{\rho_0(\widetilde{x})}{\int_{t_-}^{t_+} \mathcal{F}_{t-s}^{(0)}(\widetilde{x})\,\mathrm{d}s}\,\mathrm{d}\widetilde{x}\,\mathrm{d}t \\
&= \mathbb{E}_0\Big[\int_{t_-}^{t_+} \frac{\mathcal{F}_t^{(1)}(\cdot)}{\int_{t-t_+}^{t-t_-} \mathcal{F}_s^{(0)}(\cdot)\,\mathrm{d}s}\,\mathrm{d}t\Big] \equiv \mathbb{E}_0\big[\mathcal{A}_{t_-,t_+}\big].
\end{aligned}
\tag{32}
$$

Note that as the integrand is non-negative, switching the order of time integration and space integration is justified by Fubini–Tonelli theorem.

The proof of Proposition 2.1 for the infinite-time case is essentially the same as the finite-time case, except the followings:

- We need to replace the domain Ω in (32) by $\mho(\boldsymbol{b})$ defined in Definition A.4.
- We need the continuity of $x \to \int_{-\infty}^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t$ in order to use Theorem 2 in [21] to get the first line in (32). A generalization with discontinuity should be possible by e.g., considering piecewise continuous $\boldsymbol{b}$. However, we shall not explore this further in this work. As a remark, the map $x \to \boldsymbol{X}_t(x)$ being bijective (due to the nature of ODE flows) on $\overline{\mho(\boldsymbol{b})}$ was proved in Lemma A.6 (iii), when applying [21, Theorem 2].

B.2 Proof of Proposition 2.2

Consider

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t(x) &= \boldsymbol{b}\big(\boldsymbol{X}_t(x)\big), \qquad \boldsymbol{X}_0(x) = x \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{Z}_t(x) &= \alpha\big(\boldsymbol{Z}_t(x)\big)\boldsymbol{b}\big(\boldsymbol{Z}_t(x)\big), \qquad \boldsymbol{Z}_0(x) = x.
\end{aligned}
$$

To prove the first conclusion, we need to verify that the trajectory $\big\{\boldsymbol{X}_t(x)\big\}_{t\in\mathbb{R}} = \big\{\boldsymbol{Z}_t(x)\big\}_{t\in\mathbb{R}}$. From now on, let us fix x and introduce a scalar-valued function θ by $\theta_t := \int_0^t \alpha\big(\boldsymbol{Z}_s(x)\big)\,\mathrm{d}s$ (i.e., $\frac{\mathrm{d}}{\mathrm{d}t}\theta_t = \alpha\big(\boldsymbol{Z}_t(x)\big)$. By taking the time derivative, it is not hard to verify that $\boldsymbol{Z}_t(x) = \boldsymbol{X}_{\theta_t}\big(x\big)$ as both satisfy the same ODE. Of course, θ also depends on x but we shall not explicitly specify this dependence for simplicity of notations. Thus, the trajectory $t \to \boldsymbol{Z}_t(x)$ is the same as $t \to \boldsymbol{X}_t(x)$ under time rescaling specified by θ.

Next we shall prove the following lemma, which immediately leads into the second result in Proposition 2.2.

Lemma B.1. *Suppose $g : \Omega \to \mathbb{R}$ is a non-negative continuous function. For any $x \in \Omega$,*

$$
\begin{aligned}
&\int_{-\infty}^{\infty} g\big(\boldsymbol{Z}_t(x)\big) \exp\Big(\int_0^t \nabla\cdot(\alpha\boldsymbol{b})\big(\boldsymbol{Z}_s(x)\big)\,\mathrm{d}s\Big)\,\mathrm{d}t \\
&\quad = \frac{1}{\alpha(x)} \int_{-\infty}^{\infty} g\big(\boldsymbol{X}_t(x)\big) \exp\Big(\int_0^t \big(\nabla\cdot\boldsymbol{b}\big)\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\Big)\,\mathrm{d}t.
\end{aligned}
$$

Proof. Because α is strictly positive, $\theta_\cdot : \mathbb{R} \to \mathbb{R}$ is bijective, and by the inverse function theorem

$$
\frac{\mathrm{d}}{\mathrm{d}t}\theta_t^{-1} = \frac{1}{\alpha\big(\boldsymbol{Z}_{\theta_t^{-1}}(x)\big)} = \frac{1}{\alpha\big(\boldsymbol{X}_t(x)\big)}.
$$

Then by the change of variables $\widetilde{t} = \theta_t$ and $\widetilde{s} = \theta_s$,

$$\begin{aligned}
&\int_{-\infty}^{\infty} g\big(\boldsymbol{Z}_t(x)\big) \exp\Big(\int_0^t \nabla\cdot(\alpha\boldsymbol{b})\big(\boldsymbol{Z}_s(x)\big)\,\mathrm{d}s\Big)\,\mathrm{d}t \\
=&\int_{-\infty}^{\infty} g\big(\boldsymbol{X}_{\theta_t}(x)\big) \exp\Big(\int_0^t \nabla\cdot(\alpha\boldsymbol{b})\big(\boldsymbol{X}_{\theta_s}(x)\big)\,\mathrm{d}s\Big)\,\mathrm{d}t \\
=&\int_{-\infty}^{\infty} g\big(\boldsymbol{X}_{\widetilde{t}}(x)\big)\frac{1}{\alpha\big(\boldsymbol{X}_{\widetilde{t}}(x)\big)} \exp\Big(\int_0^{\widetilde{t}} \boldsymbol{\nabla}\cdot(\alpha\boldsymbol{b})\big(\boldsymbol{X}_{\widetilde{s}}(x)\big)\frac{1}{\alpha\big(\boldsymbol{X}_{\widetilde{s}}(x)\big)}\,\mathrm{d}\widetilde{s}\Big)\,\mathrm{d}\widetilde{t}.
\end{aligned}$$

It is then sufficient to show that $\psi_1(t) = \psi_2(t)$ for any $t \in \mathbb{R}$, where

$$\begin{aligned}
\psi_1(t) &:= \frac{1}{\alpha\big(\boldsymbol{X}_t(x)\big)} \exp\Big(\int_0^t \boldsymbol{\nabla}\cdot(\alpha\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\frac{1}{\alpha\big(\boldsymbol{X}_s(x)\big)}\,\mathrm{d}s\Big), \\
\psi_2(t) &:= \frac{1}{\alpha(x)} \exp\Big(\int_0^t (\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\Big).
\end{aligned}$$

It is easy to observe that $\psi_1(0) = \psi_2(0)$. Let us consider the time derivative of ψ_1

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\psi_1(t) &= \psi_1(t)\Big(-\frac{1}{\alpha(\boldsymbol{X}_t(x))}\big\langle\nabla\alpha(\boldsymbol{X}_t(x)), \boldsymbol{b}(\boldsymbol{X}_t(x))\big\rangle + \nabla\cdot(\alpha\boldsymbol{b})\big(\boldsymbol{X}_t(x)\big)\frac{1}{\alpha\big(\boldsymbol{X}_t(x)\big)}\Big) \\
&= \psi_1(t)\big(\nabla\cdot\boldsymbol{b}\big)\big(\boldsymbol{X}_t(x)\big).
\end{aligned}$$

It is clear that $\psi_2(t)$ satisfies the same ODE and thus $\psi_1 = \psi_2$. □

B.3 Remarks on the discrete-time analogy of (7)

Let us briefly explain how (7) connects to the method in [43] by time discretization. Suppose $N_- = t_-/h$ and $N_+ = t_+/h$, where $h \ll 1$ is the time step size; for simplicity of notation, let us assume that N_- and N_+ are simply integers. By discretizing the time integration for the finite-time NEIS scheme in (7),

$$\begin{aligned}
\mathbb{E}_{x\sim\rho_0}\Big[\int_{t_-}^{t_+} \frac{\mathcal{F}_t^{(1)}(x)}{\int_{t-T_+}^{t-T_-}\mathcal{F}_s^{(0)}(x)\,\mathrm{d}s}\,\mathrm{d}t\Big] &\approx \mathbb{E}_{x\sim\rho_0}\Big[\sum_{k=N_-}^{N_+} \frac{e^{-U_1\left(\boldsymbol{X}_{kh}(x)\right)}\mathcal{J}_{kh}(x)}{\sum_{j=k-N_+}^{k-N_-} e^{-U_0\left(\boldsymbol{X}_{jh}(x)\right)}\mathcal{J}_{jh}(x)}\Big] \\
&= \mathbb{E}_{x\sim\rho_0}\Big[\sum_{k=N_-}^{N_+} \frac{e^{-(U_1-U_0)\left(\boldsymbol{T}^{-k}(x)\right)}\rho_0\big(\boldsymbol{T}^{-k}(x)\big)\,\mathcal{J}_{\boldsymbol{T}^{-k}}(x)}{\sum_{j=k-N_+}^{k-N_-}\rho_0\big(\boldsymbol{T}^{-j}(x)\big)\mathcal{J}_{\boldsymbol{T}^{-j}}(x)}\Big],
\end{aligned}$$

where we denote $\boldsymbol{T}(\cdot) = \boldsymbol{X}_{-h}(\cdot)$, and $\mathcal{J}_{\boldsymbol{T}^{-j}}(x) := \Big|\det\Big(\nabla\boldsymbol{T}^{-j}(x)\Big)\Big|$ is the Jacobian for the map $\boldsymbol{T}^{-j}$. Then by choosing $N_- = -K$ and $N_+ = 0$, we have

$$\mathbb{E}_{x\sim\rho_0}\Big[\sum_{k=0}^{K} e^{-(U_1-U_0)\left(\boldsymbol{T}^k(x)\right)} w_k(x)\Big], \qquad w_k(x) = \frac{\rho_0\big(\boldsymbol{T}^k(x)\big)\mathcal{J}_{\boldsymbol{T}^k}(x)}{\sum_{j=-k}^{-k+K}\rho_0\big(\boldsymbol{T}^{-j}(x)\big)\mathcal{J}_{\boldsymbol{T}^{-j}}(x)},$$

which are Eqs. (8) and (10) in arXiv v1 of [43].

B.4 Remark on the relation between finite-time and infinite-time NEIS

In what follows, we briefly elaborate on the relation between the finite-time and infinite-time NEIS schemes. Suppose we fix $-\infty < t_- < 0 < t_+ < \infty$ and consider a fixed valid flow $\boldsymbol{b}$ for the infinite-time NEIS, i.e.,

$$\mathcal{A}^{\boldsymbol{b}}(x) \equiv \frac{\int_{\mathbb{R}} \exp\big(-U_1(\boldsymbol{X}_t(x)) + \int_0^t \nabla\cdot\boldsymbol{b}(\boldsymbol{X}_r(x))\,\mathrm{d}r\big)\,\mathrm{d}t}{\int_{\mathbb{R}} \exp\big(-U_0(\boldsymbol{X}_t(x)) + \int_0^t \nabla\cdot\boldsymbol{b}(\boldsymbol{X}_r(x))\,\mathrm{d}r\big)\,\mathrm{d}t}$$

is well-defined for almost all $x \in \Omega$, where the state $\boldsymbol{X}_t(x)$ solves the ODE $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t(x) = \boldsymbol{b}(\boldsymbol{X}_t(x))$ for any $x \in \Omega$. The superscript in $\mathcal{A}^{\boldsymbol{b}}$ is used to emphasize that it is the estimator for this particular flow $\boldsymbol{b}$ and similar notations are used below.

Next we consider a family of rescaled flow $\boldsymbol{b}^\alpha$ parameterized by $\alpha > 0$, defined as

$$\boldsymbol{b}^\alpha(x) = \alpha \boldsymbol{b}(x), \qquad \forall x \in \Omega.$$

Let us study its estimator for the finite-time NEIS:

$$\mathcal{A}^{\boldsymbol{b}^\alpha}_{t_-,t_+}(x) \overset{(8)}{=} \int_{t_-}^{t_+} \frac{\exp\big(-U_1(\boldsymbol{Z}_t(x)) + \int_0^t \nabla\cdot \boldsymbol{b}^\alpha(\boldsymbol{Z}_r(x))\,\mathrm{d}r\big)}{\int_{t-t_+}^{t-t_-} \exp\big(-U_0(\boldsymbol{Z}_s(x)) + \int_0^s \nabla\cdot \boldsymbol{b}^\alpha(\boldsymbol{Z}_r(x))\,\mathrm{d}r\big)\,\mathrm{d}s}\,\mathrm{d}t, \tag{33}$$

where $\boldsymbol{Z}_t(x)$ solves the ODE $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{Z}_t(x) = \boldsymbol{b}^\alpha(\boldsymbol{Z}_t(x))$ for any $x \in \Omega$. From Appendix B.2, we already know that $\boldsymbol{Z}_t(x) = \boldsymbol{X}_{\alpha t}(x)$ for any $x \in \Omega$ and $t \in \mathbb{R}$. By change of time variables $t = \widetilde{t}/\alpha$, $s = \widetilde{s}/\alpha$ and $r = \widetilde{r}/\alpha$, we have

$$\begin{aligned}
\mathcal{A}^{\boldsymbol{b}^\alpha}_{t_-,t_+}(x) &= \int_{\alpha t_-}^{\alpha t_+} \frac{\exp\big(-U_1(\boldsymbol{X}_{\widetilde{t}}(x)) + \int_0^{\widetilde{t}} \nabla\cdot \boldsymbol{b}(\boldsymbol{X}_{\widetilde{r}}(x))\,\mathrm{d}\widetilde{r}\big)}{\int_{\widetilde{t}-\alpha t_+}^{\widetilde{t}-\alpha t_-} \exp\big(-U_0(\boldsymbol{X}_{\widetilde{s}}(x)) + \int_0^{\widetilde{s}} \nabla\cdot \boldsymbol{b}(\boldsymbol{X}_{\widetilde{r}}(x))\,\mathrm{d}\widetilde{r}\big)\,\mathrm{d}\widetilde{s}}\,\mathrm{d}\widetilde{t} \\
&= \int_{\mathbb{R}} \frac{\chi_{[\alpha t_-,\alpha t_+]}(\widetilde{t})\,\exp\big(-U_1(\boldsymbol{X}_{\widetilde{t}}(x)) + \int_0^{\widetilde{t}} \nabla\cdot \boldsymbol{b}(\boldsymbol{X}_{\widetilde{r}}(x))\,\mathrm{d}\widetilde{r}\big)}{\int_{\widetilde{t}-\alpha t_+}^{\widetilde{t}-\alpha t_-} \exp\big(-U_0(\boldsymbol{X}_{\widetilde{s}}(x)) + \int_0^{\widetilde{s}} \nabla\cdot \boldsymbol{b}(\boldsymbol{X}_{\widetilde{r}}(x))\,\mathrm{d}\widetilde{r}\big)\,\mathrm{d}\widetilde{s}}\,\mathrm{d}\widetilde{t}.
\end{aligned}$$

For any $\widetilde{t} \in \mathbb{R}$, as $\alpha \to \infty$, the integrand in the last equation pointwisely converges to

$$\frac{\exp\big(-U_1(\boldsymbol{X}_{\widetilde{t}}(x)) + \int_0^{\widetilde{t}} \nabla\cdot \boldsymbol{b}(\boldsymbol{X}_{\widetilde{r}}(x))\,\mathrm{d}\widetilde{r}\big)}{\int_{\mathbb{R}} \exp\big(-U_0(\boldsymbol{X}_{\widetilde{s}}(x)) + \int_0^{\widetilde{s}} \nabla\cdot \boldsymbol{b}(\boldsymbol{X}_{\widetilde{r}}(x))\,\mathrm{d}\widetilde{r}\big)\,\mathrm{d}\widetilde{s}}.$$

If we heuristically swap the order of taking the limit $\alpha \to \infty$ and the integral with respect to $\widetilde{t}$ (which should generally hold for most examples), we end up with an identity:

$$\lim_{\alpha\to\infty} \mathcal{A}^{\boldsymbol{b}^\alpha}_{t_-,t_+}(x) = \int_{\mathbb{R}} \frac{\exp\big(-U_1(\boldsymbol{X}_{\widetilde{t}}(x)) + \int_0^{\widetilde{t}} \nabla\cdot \boldsymbol{b}(\boldsymbol{X}_{\widetilde{r}}(x))\,\mathrm{d}\widetilde{r}\big)}{\int_{\mathbb{R}} \exp\big(-U_0(\boldsymbol{X}_{\widetilde{s}}(x)) + \int_0^{\widetilde{s}} \nabla\cdot \boldsymbol{b}(\boldsymbol{X}_{\widetilde{r}}(x))\,\mathrm{d}\widetilde{r}\big)\,\mathrm{d}\widetilde{s}}\,\mathrm{d}\widetilde{t} \equiv \mathcal{A}^{\boldsymbol{b}}(x).$$

The above relation heuristically justifies that due to the space-time rescaling, in the limit of large magnitude of the flow (i.e., $\alpha \to \infty$ above), it does not matter how t_-, t_+ are chosen as long as $t_- < 0 < t_+$. In particular, if the flow $\boldsymbol{b}$ is a zero-variance dynamics, i.e., $\mathcal{A}^{\boldsymbol{b}}(x) = \mathcal{Z}_1$ for $x \in \Omega$ almost surely, then in the finite-time NEIS, the flow $\boldsymbol{b}^\alpha = \alpha\boldsymbol{b}$ (with $\alpha \gg 1$) should be approximately a zero-variance dynamics for the finite-time scheme. The finite-time NEIS scheme may not have explicit analytical results about zero-variance dynamics in the same way as the infinite-time NEIS scheme; however, due to the above discussed relation, the finite-time version still possesses the ability to handle and learn an approximately zero-variance dynamics, which is good enough in practice, e.g., during training the optimal flow in Section 5.

C The first-order perturbation of the variance for the finite-time scheme

Here we study how the variance (or equivalently, the second moment) of the estimator depends on $\boldsymbol{b}$, since the performance of the finite-time NEIS scheme (7) largely depends on this choice. More specifically, in the following proposition, we study how the second moment changes under a small perturbation of $\boldsymbol{b}$. The expression (34) below will be useful for training optimal dynamics in Section 5 (see also Appendix I for details).

Proposition C.1. *Suppose $\boldsymbol{b} \in \mathfrak{B}$ and for any perturbation $\delta\boldsymbol{b} \in C_c^\infty(\Omega, \mathbb{R}^d)$, denote $\boldsymbol{b}^\epsilon := \boldsymbol{b} + \epsilon\delta\boldsymbol{b}$. Then,*

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{M}_{t_-,t_+}(\boldsymbol{b}+\epsilon\delta\boldsymbol{b})\Big|_{\epsilon=0} \\
&= 2\mathbb{E}_{x\sim\rho_0}\Bigg[\mathcal{A}_{t_-,t_+}(x)\,\Big(\int_{t_-}^{t_+} \frac{\mathcal{G}_t^{(1)}(x)\int_{t-t_+}^{t-t_-}\mathcal{F}_s^{(0)}(x)\,\mathrm{d}s - \mathcal{F}_t^{(1)}(x)\int_{t-t_+}^{t-t_-}\mathcal{G}_s^{(0)}(x)\,\mathrm{d}s}{\Big(\int_{t-t_+}^{t-t_-}\mathcal{F}_s^{(0)}(x)\,\mathrm{d}s\Big)^2}\,\mathrm{d}t\Big)\Bigg],
\end{aligned} \tag{34}$$

where for $k \in \{0,1\}$, we define

$$\mathcal{G}_t^{(k)}(x) := \mathcal{F}_t^{(k)}(x) \left(\begin{aligned} &\big\langle -\nabla U_k\big(\boldsymbol{X}_t(x)\big), \boldsymbol{Y}_t(x)\big\rangle + \int_0^t \big\langle \nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big), \boldsymbol{Y}_s(x)\big\rangle \,\mathrm{d}s \\ &\qquad\qquad + \int_0^t (\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s \end{aligned} \right), \tag{35}$$

and $\boldsymbol{Y}_t(x)$ is the solution of the following ODE:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{Y}_t(x) = \nabla\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\boldsymbol{Y}_t(x) + \delta\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big), \qquad \boldsymbol{Y}_0(x) = 0. \tag{36}$$

The expression of the functional derivative $\frac{\delta\mathcal{M}_{t_-,t_+}(\boldsymbol{b})}{\delta\boldsymbol{b}}$ (not presented in this work) for the finite-time case can be derived in the same way as Lemma D.8 below for the infinite-time case. However, it appears that the expression of $\frac{\delta\mathcal{M}_{t_-,t_+}(\boldsymbol{b})}{\delta\boldsymbol{b}}$ is too complicated to provide useful analytical results.

Proof. Let us perturb $\boldsymbol{b}$ by $\epsilon\delta\boldsymbol{b}$ where $\epsilon \ll 1$. Let us consider

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t(x) = \boldsymbol{b}\big(\boldsymbol{X}_t(x)\big), \\ &\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t^\epsilon(x) = \boldsymbol{b}^\epsilon\big(\boldsymbol{X}_t^\epsilon(x)\big) \equiv (\boldsymbol{b} + \epsilon\delta\boldsymbol{b})\big(\boldsymbol{X}_t^\epsilon(x)\big), \end{aligned}$$

with a fixed initial condition $\boldsymbol{X}_0(x) = \boldsymbol{X}_0^\epsilon(x) = x$. For a small ϵ, we can expect that $\boldsymbol{X}_t(x) \approx \boldsymbol{X}_t^\epsilon(x)$ and also we know $\boldsymbol{X}_t^0(x) \equiv \boldsymbol{X}_t(x)$. Define $\mathcal{F}_t^{(k,\epsilon)}(x)$ for the dynamics $\boldsymbol{b}^\epsilon$ in the same way as in (6), namely,

$$\mathcal{F}_t^{(k,\epsilon)}(x) := \exp\Big(-U_k\big(\boldsymbol{X}_t^\epsilon(x)\big) + \int_0^t (\nabla\cdot\boldsymbol{b}^\epsilon)\big(\boldsymbol{X}_s^\epsilon(x)\big)\,\mathrm{d}s\Big).$$

By these notations,

$$\mathcal{M}_{t_-,t_+}(\boldsymbol{b}^\epsilon) = \mathbb{E}_{x\sim\rho_0}\left[\Big(\int_{t_-}^{t_+} \frac{\mathcal{F}_t^{(1,\epsilon)}(x)}{\int_{t-t_+}^{t-t_-}\mathcal{F}_s^{(0,\epsilon)}(x)\,\mathrm{d}s}\,\mathrm{d}t\Big)^2\right].$$

Then we take the derivative of $\mathcal{M}_{t_-,t_+}(\boldsymbol{b}^\epsilon)$ with respect to ϵ:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{M}_{t_-,t_+}(\boldsymbol{b}^\epsilon) \\ &= 2\mathbb{E}_{x\sim\rho_0}\left[\begin{aligned} &\Big(\int_{t_-}^{t_+} \frac{\mathcal{F}_t^{(1,\epsilon)}(x)}{\int_{t-t_+}^{t-t_-}\mathcal{F}_s^{(0,\epsilon)}(x)\,\mathrm{d}s}\,\mathrm{d}t\Big) \\ &\quad\times\Big(\int_{t_-}^{t_+} \frac{\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{F}_t^{(1,\epsilon)}(x)\int_{t-t_+}^{t-t_-}\mathcal{F}_s^{(0,\epsilon)}(x)\,\mathrm{d}s - \mathcal{F}_t^{(1,\epsilon)}(x)\int_{t-t_+}^{t-t_-}\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{F}_s^{(0,\epsilon)}(x)\,\mathrm{d}s}{\Big(\int_{t-t_+}^{t-t_-}\mathcal{F}_s^{(0,\epsilon)}(x)\,\mathrm{d}s\Big)^2}\,\mathrm{d}t\Big) \end{aligned}\right]. \end{aligned}$$

Next, we need to compute $\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{F}_t^{(k,\epsilon)}(x)$. Let us first consider the perturbation to the trajectory. Let $\boldsymbol{Y}_t^\epsilon(x) := \frac{\mathrm{d}}{\mathrm{d}\epsilon}\big(\boldsymbol{X}_t^\epsilon(x)\big)$ and then

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{Y}_t^\epsilon(x) &= \frac{\mathrm{d}}{\mathrm{d}\epsilon}\Big((\boldsymbol{b} + \epsilon\delta\boldsymbol{b})\big(\boldsymbol{X}_t^\epsilon(x)\big)\Big) \\ &= \nabla\boldsymbol{b}\big(\boldsymbol{X}_t^\epsilon(x)\big)\boldsymbol{Y}_t^\epsilon(x) + \delta\boldsymbol{b}\big(\boldsymbol{X}_t^\epsilon(x)\big) + \epsilon\Big(\nabla\delta\boldsymbol{b}\big(\boldsymbol{X}_t^\epsilon(x)\big)\Big)\boldsymbol{Y}_t^\epsilon(x), \\ \boldsymbol{Y}_0^\epsilon(x) &= 0. \end{aligned}$$

When $\epsilon = 0$, $\boldsymbol{Y}_t^0(x)$ is the solution to

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{Y}_t^0(x) = \nabla\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\boldsymbol{Y}_t^0(x) + \delta\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big), \qquad \boldsymbol{Y}_0^0(x) = 0.$$

Now we are ready to explicitly write down $\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{F}_t^{(k,\epsilon)}(x)$. It is straightforward to derive that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{F}_t^{(k,\epsilon)}(x)\\
&= \mathcal{F}_t^{(k,\epsilon)}(x)\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Big(-U_k\big(\boldsymbol{X}_t^\epsilon(x)\big)+\int_0^t \nabla\cdot(\boldsymbol{b}+\epsilon\delta\boldsymbol{b})\big(\boldsymbol{X}_s^\epsilon(x)\big)\,\mathrm{d}s\Big)\\
&= \mathcal{F}_t^{(k,\epsilon)}(x)\begin{pmatrix}\big\langle -\nabla U_k\big(\boldsymbol{X}_t^\epsilon(x)\big),\boldsymbol{Y}_t^\epsilon(x)\big\rangle\\ \quad+\int_0^t\Big\langle\nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_s^\epsilon(x)\big),\boldsymbol{Y}_s^\epsilon(x)\Big\rangle+(\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s^\epsilon(x)\big)\,\mathrm{d}s\\ \quad+\epsilon\int_0^t\Big\langle\nabla(\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s^\epsilon(x)\big),\boldsymbol{Y}_s^\epsilon(x)\Big\rangle\,\mathrm{d}s\end{pmatrix}.
\end{aligned}$$

When we let $\epsilon=0$, we have

$$\begin{aligned}
\mathcal{G}_t^{(k)}(x) &:= \frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{F}_t^{(k,\epsilon)}(x)\Big|_{\epsilon=0}\\
&= \mathcal{F}_t^{(k)}(x)\begin{pmatrix}\big\langle -\nabla U_k\big(\boldsymbol{X}_t(x)\big),\boldsymbol{Y}_t^0(x)\big\rangle\\ \quad+\int_0^t\big\langle\nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big),\boldsymbol{Y}_s^0(x)\big\rangle+(\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\end{pmatrix}.
\end{aligned}$$

Finally, we arrive at the conclusion by combining previous results and dropping the superscript in $\boldsymbol{Y}_t^0(x)$ for simplicity. □

D The first-order perturbation of the variance for the infinite-time scheme

The goal of this section is to derive the functional derivative of $\mathcal{M}(\boldsymbol{b})$ with respect to $\boldsymbol{b}$, denoted as $\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}$, defined as follows: for any $\delta\boldsymbol{b}\in C_c^\infty(\Omega,\mathbb{R}^d)$ such that $\boldsymbol{b}+\epsilon\delta\boldsymbol{b}\in\mathfrak{B}_\infty$ for small enough ϵ, we have

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{M}(\boldsymbol{b}+\epsilon\delta\boldsymbol{b})\Big|_{\epsilon=0}=\int_\Omega\Big\langle\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}},\delta\boldsymbol{b}\Big\rangle. \tag{37}$$

Since $\mathcal{M}$ and Var only differ by a constant (which is independent of $\boldsymbol{b}$), it is apparent that $\frac{\delta\mathrm{Var}(\boldsymbol{b})}{\delta\boldsymbol{b}}\equiv\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}$.

Proposition D.1. *The functional derivative $\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}:\Omega\to\mathbb{R}^d$ has the following form*

$$\begin{aligned}
&\frac{\delta Var(\boldsymbol{b})}{\delta\boldsymbol{b}}(x)\equiv\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}(x)\\
&=\frac{2\nabla\mathcal{A}(x)}{\mathcal{B}(x)}\left(\int_0^\infty\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^0\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t-\int_{-\infty}^0\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_0^\infty\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\right).
\end{aligned}\tag{38}$$

Remark D.2. The proof of the last formula is slightly formal: for instance, conditions on $\boldsymbol{b}$ to ensure the existence of $\nabla\mathcal{A}$ are not discussed.

Recall the expression of $\mathcal{B}$ from (24). The proof of Proposition D.1 is given in Appendix D.2: it relies on a few explicit formulas, that we state first.

D.1 Some explicit formulas

We need some explicit formulas for $\boldsymbol{Y}_t(x)$ (36) and $\mathcal{G}_t^{(k)}(x)$ (35). We notice that $\mathcal{G}^{(k)}$ depends on $\boldsymbol{Y}_\cdot(\cdot)$ and $\delta\boldsymbol{b}$ linearly, and $\boldsymbol{Y}_t(x)$ also depends on $\delta\boldsymbol{b}$ linearly. Therefore, the first step is to rewrite the expression of $\boldsymbol{Y}_t(x)$ more explicitly in terms of $\delta\boldsymbol{b}$.

Lemma D.3. *Suppose the dynamics $\boldsymbol{b}\in\mathfrak{B}$ and $\delta\boldsymbol{b}\in C_c^\infty(\Omega,\mathbb{R}^d)$. Then we have*

$$\boldsymbol{Y}_t(x)=\int_0^t\boldsymbol{C}_{t,s}(x)\;\delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s,\qquad\forall x\in\Omega. \tag{39}$$

The kernel $\boldsymbol{C}_{t,s}(x) \in \mathbb{R}^{d\times d}$ has the following form

$$\boldsymbol{C}_{t,s}(x) = \begin{cases} \exp_{\mathcal{T}_{\leftarrow}}\Big(\int_s^t \nabla \boldsymbol{b}\big(\boldsymbol{X}_r(x)\big)\,\mathrm{d}r\Big), & \text{if } t \geq s \geq 0, \\ \Big(\exp_{\mathcal{T}_{\leftarrow}}\Big(\int_t^s \nabla \boldsymbol{b}\big(\boldsymbol{X}_r(x)\big)\,\mathrm{d}r\Big)\Big)^{-1}, & \text{if } t \leq s \leq 0, \end{cases} \tag{40}$$

where $\exp_{\mathcal{T}_{\leftarrow}}$ is the chronological time-ordered operator exponential.

Proof. By plugging the ansatz (39) into (36), we immediately know that $\boldsymbol{C}_{s,s}(x) = \mathbb{I}_d$ for all $s \in \mathbb{R}$ and

$$\partial_t \boldsymbol{C}_{t,s}(x) = \nabla \boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\boldsymbol{C}_{t,s}(x). \tag{41}$$

This linear ODE has an explicit solution as in (40), by introducing time-ordered operator exponential. □

Next, we shall rewrite $\mathcal{G}_t^{(k)}(x)$.

Lemma D.4. *We can rewrite $\mathcal{G}_t^{(k)}(x)$ as follows*

$$\mathcal{G}_t^{(k)}(x) = \mathcal{F}_t^{(k)}(x)\Big(\int_0^t \Big\langle \boldsymbol{V}_{t,s}^{(k)}(x), \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\Big\rangle + (\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\Big), \tag{42}$$

where $\boldsymbol{V}_{t,s}^{(k)}(x)$ is defined as

$$\boldsymbol{V}_{t,s}^{(k)}(x) := -\boldsymbol{C}_{t,s}(x)^T\nabla U_k\big(\boldsymbol{X}_t(x)\big) + \int_s^t \boldsymbol{C}_{r,s}(x)^T\nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_r(x)\big)\,\mathrm{d}r. \tag{43}$$

Proof. Recall the expression of $\mathcal{G}_t^{(k)}(x)$ from (35). After plugging (39), we have

$$\begin{aligned}
\mathcal{G}_t^{(k)}(x) &\overset{(35)}{=} \mathcal{F}_t^{(k)}(x)\begin{pmatrix} \big\langle -\nabla U_k\big(\boldsymbol{X}_t(x)\big), \boldsymbol{Y}_t(x)\big\rangle + \int_0^t \big\langle \nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big), \boldsymbol{Y}_s(x)\big\rangle\,\mathrm{d}s \\ + \int_0^t (\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s \end{pmatrix} \\
&\overset{(39)}{=} \mathcal{F}_t^{(k)}(x)\begin{pmatrix} -\int_0^t \Big\langle \boldsymbol{C}_{t,s}(x)^T\nabla U_k\big(\boldsymbol{X}_t(x)\big), \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\Big\rangle\,\mathrm{d}s + \int_0^t (\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s \\ + \int_0^t\int_0^s \Big\langle \boldsymbol{C}_{s,r}(x)^T\nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big), \delta\boldsymbol{b}\big(\boldsymbol{X}_r(x)\big)\Big\rangle\,\mathrm{d}r\,\mathrm{d}s \end{pmatrix} \\
&= \mathcal{F}_t^{(k)}(x)\begin{pmatrix} -\int_0^t \Big\langle \boldsymbol{C}_{t,s}(x)^T\nabla U_k\big(\boldsymbol{X}_t(x)\big), \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\Big\rangle\,\mathrm{d}s + \int_0^t (\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s \\ + \int_0^t \Big\langle \int_s^t \boldsymbol{C}_{r,s}(x)^T\nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_r(x)\big)\,\mathrm{d}r, \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\Big\rangle\,\mathrm{d}s \end{pmatrix} \\
&= \mathcal{F}_t^{(k)}(x)\Big(\int_0^t \big\langle \boldsymbol{V}_{t,s}^{(k)}(x), \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\big\rangle + (\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\Big).
\end{aligned}$$

□

Then we present a few properties, which will be useful when computing the functional derivative $\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}$. The following lemma shows how $\boldsymbol{C}_{t,-s}(\cdot)$ and $\boldsymbol{V}_{t,-s}^{(k)}(\cdot)$ change under the dynamical evolution $\boldsymbol{X}_s(\cdot)$.

Lemma D.5. *When $t \leq -s \leq 0$ or $t \geq -s \geq 0$,*

$$\boldsymbol{C}_{t,-s}\big(\boldsymbol{X}_s(x)\big) = \boldsymbol{C}_{t+s,0}(x), \qquad \boldsymbol{V}_{t,-s}^{(k)}\big(\boldsymbol{X}_s(x)\big) = \boldsymbol{V}_{t+s,0}^{(k)}(x). \tag{44}$$

Proof. We first consider the term $\boldsymbol{C}_{t,-s}\big(\boldsymbol{X}_s(x)\big)$. When $t \geq -s \geq 0$,

$$\begin{aligned}\boldsymbol{C}_{t,-s}\big(\boldsymbol{X}_s(x)\big) &= \exp_{\mathcal{T}_{\leftarrow}}\Big(\int_{-s}^{t} \nabla \boldsymbol{b}\big(\boldsymbol{X}_r\big(\boldsymbol{X}_s(x)\big)\big)\,\mathrm{d}r\Big)\\ &= \exp_{\mathcal{T}_{\leftarrow}}\Big(\int_{0}^{t+s} \nabla \boldsymbol{b}(\boldsymbol{X}_r(x))\,\mathrm{d}r\Big) = \boldsymbol{C}_{t+s,0}(x).\end{aligned}$$

The case for $t \leq -s \leq 0$ can be similarly verified. Recall from (43) that

$$\begin{aligned}&\boldsymbol{V}_{t,-s}^{(k)}\big(\boldsymbol{X}_s(x)\big)\\ &\overset{(43)}{=} -\boldsymbol{C}_{t,-s}\big(\boldsymbol{X}_s(x)\big)^T \nabla U_k\Big(\boldsymbol{X}_t\big(\boldsymbol{X}_s(x)\big)\Big) + \int_{-s}^{t} \boldsymbol{C}_{r,-s}\big(\boldsymbol{X}_s(x)\big)^T \nabla(\nabla\cdot\boldsymbol{b})\Big(\boldsymbol{X}_r\big(\boldsymbol{X}_s(x)\big)\Big)\,\mathrm{d}r\\ &= -\boldsymbol{C}_{t+s,0}(x)^T \nabla U_k\big(\boldsymbol{X}_{t+s}(x)\big) + \int_{0}^{t+s} \boldsymbol{C}_{r,0}(x)^T \nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_r(x)\big)\,\mathrm{d}r\\ &\overset{(43)}{=} \boldsymbol{V}_{t+s,0}^{(k)}(x),\end{aligned}$$

where to get the third line, we use the above formula (44) about $\boldsymbol{C}_{t,-s}\big(\boldsymbol{X}_s(x)\big)$. □

The following lemma connects $\boldsymbol{C}_{t,0}$ and $\boldsymbol{V}_{t,0}^{(k)}$ with gradients.

Lemma D.6. *For any $x \in \Omega$ and $t \in \mathbb{R}$, we have*

$$\nabla_x \boldsymbol{X}_t(x) := \Big[\tfrac{\partial(\boldsymbol{X}_t(x))_i}{\partial x_j}\Big]_{i,j} = \boldsymbol{C}_{t,0}(x), \tag{45}$$

$$\mathcal{F}_t^{(k)}(x)\boldsymbol{V}_{t,0}^{(k)}(x) = \nabla_x \mathcal{F}_t^{(k)}(x), \qquad \textit{for } k \in \{0,1\}. \tag{46}$$

As a consequence, $\det\big(\boldsymbol{C}_{t,0}(x)\big) = \mathcal{J}_t(x)$.

Proof. We fix an index $1 \leq j \leq d$ and consider the dynamics $\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\boldsymbol{X}}_t^{\epsilon}(x) = \boldsymbol{b}\big(\widetilde{\boldsymbol{X}}_t^{\epsilon}(x)\big)$, $\widetilde{\boldsymbol{X}}_0^{\epsilon}(x) = x + \epsilon \boldsymbol{e}_j$ where $\boldsymbol{e}_j$ is a vector with the jth element to be one, and all other elements to be zero. Clearly, $\widetilde{\boldsymbol{X}}_t^{0}(x) = \boldsymbol{X}_t(x)$.

Let $\widetilde{\boldsymbol{Y}}_t^{\epsilon}(x) := \frac{\mathrm{d}}{\mathrm{d}\epsilon}\widetilde{\boldsymbol{X}}_t^{\epsilon}(x)$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\boldsymbol{Y}}_t^{\epsilon}(x) = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\boldsymbol{b}\big(\widetilde{\boldsymbol{X}}_t^{\epsilon}(x)\big) = \nabla\boldsymbol{b}\big(\widetilde{\boldsymbol{X}}_t^{\epsilon}(x)\big)\widetilde{\boldsymbol{Y}}_t^{\epsilon}(x), \qquad \widetilde{\boldsymbol{Y}}_0^{\epsilon}(x) = \boldsymbol{e}_j.$$

Moreover, when $\epsilon = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\boldsymbol{Y}}_t^{0}(x) = \nabla\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\widetilde{\boldsymbol{Y}}_t^{0}(x), \qquad \widetilde{\boldsymbol{Y}}_0^{0}(x) = \boldsymbol{e}_j,$$

whose solution is simply j^{th} column of $\boldsymbol{C}_{t,0}(x)$. Besides, the jth column of $\nabla_x \boldsymbol{X}_t(x)$ is given by

$$\lim_{\epsilon\to 0}\frac{\boldsymbol{X}_t(x+\epsilon\boldsymbol{e}_j) - \boldsymbol{X}_t(x)}{\epsilon} = \lim_{\epsilon\to 0}\frac{\widetilde{\boldsymbol{X}}_t^{\epsilon}(x) - \widetilde{\boldsymbol{X}}_t^{0}(x)}{\epsilon} = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\widetilde{\boldsymbol{X}}_t^{\epsilon}(x)\Big|_{\epsilon=0} = \widetilde{\boldsymbol{Y}}_t^{0}(x).$$

By combining above results, we easily know that $\nabla_x \boldsymbol{X}_t(x) = \boldsymbol{C}_{t,0}(x)$.

Next, for $k \in \{0,1\}$ and any $x \in \Omega$,

$$\begin{aligned}\nabla\mathcal{F}_t^{(k)}(x) &= \mathcal{F}_t^{(k)}(x)\Big(-\big(\nabla_x\boldsymbol{X}_t(x)\big)^T\nabla U_k\big(\boldsymbol{X}_t(x)\big) + \int_0^t \big(\nabla_x\boldsymbol{X}_s(x)\big)^T\nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\Big)\\ &= \mathcal{F}_t^{(k)}(x)\Big(-\boldsymbol{C}_{t,0}(x)^T\nabla U_k\big(\boldsymbol{X}_t(x)\big) + \int_0^t \boldsymbol{C}_{s,0}(x)^T\nabla(\nabla\cdot\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\Big)\\ &\overset{(43)}{=} \mathcal{F}_t^{(k)}(x)\boldsymbol{V}_{t,0}^{(k)}(x).\end{aligned}$$

□

D.2 Proof of Proposition D.1

We list without proof the following result for the infinite-time case, which can be derived in the same way as Proposition C.1.

Lemma D.7. *Let $\boldsymbol{b}^\epsilon := \boldsymbol{b} + \epsilon\delta\boldsymbol{b}$. Suppose that for small enough ϵ, we have $\boldsymbol{b}^\epsilon \in \mathfrak{B}_\infty$. Then*

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{M}(\boldsymbol{b}+\epsilon\delta\boldsymbol{b})\Big|_{\epsilon=0} = 2\mathbb{E}_{x\sim\rho_0}\Big[\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\Big(\int_{-\infty}^{\infty}\mathcal{G}_t^{(1)}(x)\,\mathrm{d}t - \mathcal{A}(x)\int_{-\infty}^{\infty}\mathcal{G}_t^{(0)}(x)\,\mathrm{d}t\Big)\Big],$$

where $\mathcal{G}_t^{(k)}(x)$ is defined in (35) *for $k = 0, 1$.*

Lemma D.8. *The functional derivative of the second moment for the infinite-time case is*

$$\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}(x) = 2\Big(\int_{-\infty}^{\infty}\mathcal{F}_s^{(0)}(x)\boldsymbol{\mathcal{S}}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big) - \nabla_x\Big(\mathcal{F}_s^{(0)}(x)\mathcal{G}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)\Big)\,\mathrm{d}s\Big), \tag{47}$$

where

$$\boldsymbol{\mathcal{S}}_s^{\infty}(x) := \begin{cases} \dfrac{\mathcal{A}(x)}{\mathcal{B}(x)}\displaystyle\int_s^{\infty}\mathcal{F}_t^{(1)}(x)\boldsymbol{V}_{t,s}^{(1)}(x)\,\mathrm{d}t - \dfrac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_s^{\infty}\mathcal{F}_t^{(0)}(x)\boldsymbol{V}_{t,s}^{(0)}(x)\,\mathrm{d}t, & \text{if } s > 0; \\ -\dfrac{\mathcal{A}(x)}{\mathcal{B}(x)}\displaystyle\int_{-\infty}^{s}\mathcal{F}_t^{(1)}(x)\boldsymbol{V}_{t,s}^{(1)}(x)\,\mathrm{d}t + \dfrac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{-\infty}^{s}\mathcal{F}_t^{(0)}(x)\boldsymbol{V}_{t,s}^{(0)}(x)\,\mathrm{d}t, & \text{if } s < 0. \end{cases}$$

$$\mathcal{G}_s^{\infty}(x) := \begin{cases} \dfrac{\mathcal{A}(x)}{\mathcal{B}(x)}\displaystyle\int_s^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \dfrac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_s^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t, & \text{if } s > 0; \\ -\dfrac{\mathcal{A}(x)}{\mathcal{B}(x)}\displaystyle\int_{-\infty}^{s}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t + \dfrac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{-\infty}^{s}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t, & \text{if } s < 0. \end{cases}$$

When $s = 0$, $\boldsymbol{\mathcal{S}}_0^{\infty}(\cdot), \mathcal{G}_0^{\infty}(\cdot)$ are not specified above, because they will not affect the functional derivative $\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}$ by changing values at a single point.

Proof. In Lemma D.7, we need to simplify the term $\int_{-\infty}^{\infty}\mathcal{G}_t^{(k)}(x)\,\mathrm{d}t$. By plugging the formula of $\mathcal{G}^{(k)}$ from (42), we have

$$\begin{aligned}
&\int_{-\infty}^{\infty}\mathcal{G}_t^{(k)}(x)\,\mathrm{d}t \\
&\overset{(42)}{=} \int_{-\infty}^{\infty}\mathcal{F}_t^{(k)}(x)\Big(\int_0^t\Big\langle\boldsymbol{V}_{t,s}^{(k)}(x), \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\Big\rangle + (\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\Big)\,\mathrm{d}t \\
&= \int_0^{\infty}\int_s^{\infty}\mathcal{F}_t^{(k)}(x)\Big\langle\boldsymbol{V}_{t,s}^{(k)}(x), \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\Big\rangle + \mathcal{F}_t^{(k)}(x)(\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}t\,\mathrm{d}s \\
&\qquad - \int_{-\infty}^{0}\int_{-\infty}^{s}\mathcal{F}_t^{(k)}(x)\Big\langle\boldsymbol{V}_{t,s}^{(k)}(x), \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\Big\rangle + \mathcal{F}_t^{(k)}(x)(\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}t\,\mathrm{d}s.
\end{aligned}$$

By plugging the last equation into Lemma D.7 and with straightforward simplification, we can verify that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{M}(\boldsymbol{b}+\epsilon\delta\boldsymbol{b})\Big|_{\epsilon=0} \\
&= 2\mathbb{E}_{x\sim\rho_0}\Big[\int_{-\infty}^{\infty}\Big\langle\boldsymbol{\mathcal{S}}_s^{\infty}(x), \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\Big\rangle\,\mathrm{d}s + \int_{-\infty}^{\infty}\mathcal{G}_s^{\infty}(x)(\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\Big] \\
&= 2\int_{-\infty}^{\infty}\int_{\Omega}\rho_0(x)\Big(\Big\langle\boldsymbol{\mathcal{S}}_s^{\infty}(x), \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\Big\rangle + \mathcal{G}_s^{\infty}(x)(\nabla\cdot\delta\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\Big)\,\mathrm{d}x\,\mathrm{d}s \\
&\overset{\widetilde{x}=\boldsymbol{X}_s(x)}{=} 2\int_{-\infty}^{\infty}\int_{\Omega}\Big(\rho_0\big(\boldsymbol{X}_{-s}(\widetilde{x})\big)\Big\langle\boldsymbol{\mathcal{S}}_s^{\infty}\big(\boldsymbol{X}_{-s}(\widetilde{x})\big), \delta\boldsymbol{b}(\widetilde{x})\Big\rangle\mathcal{J}_{-s}(\widetilde{x}) \\
&\qquad\qquad + \rho_0\big(\boldsymbol{X}_{-s}(\widetilde{x})\big)\mathcal{G}_s^{\infty}\big(\boldsymbol{X}_{-s}(\widetilde{x})\big)(\boldsymbol{\nabla}\cdot\delta\boldsymbol{b})(\widetilde{x})\mathcal{J}_{-s}(\widetilde{x})\Big)\,\mathrm{d}\widetilde{x}\,\mathrm{d}s
\end{aligned}$$

$$\overset{(6)}{=} 2\int_{-\infty}^{\infty}\int_{\Omega}\mathcal{F}_{-s}^{(0)}(x)\Big\langle \boldsymbol{\mathcal{S}}_s^{\infty}\big(\boldsymbol{X}_{-s}(x)\big),\delta\boldsymbol{b}(x)\Big\rangle - \Big\langle \nabla\Big(\mathcal{F}_{-s}^{(0)}(x)\mathcal{G}_s^{\infty}\big(\boldsymbol{X}_{-s}(x)\big)\Big),\delta\boldsymbol{b}(x)\Big\rangle\,\mathrm{d}x\,\mathrm{d}s.$$

The integration by parts in the last line is valid because $\delta\boldsymbol{b}$ vanishes at the boundary $\partial\Omega$. By comparing the last equation with (37), we can immediately obtain (47) after straightforward simplifications. □

Then we need to simplify $\boldsymbol{\mathcal{S}}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)$ and $\mathcal{G}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)$.

Lemma D.9. *$\boldsymbol{\mathcal{S}}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)$ and $\mathcal{G}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)$ has the following form*

$$\boldsymbol{\mathcal{S}}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big) = \begin{cases} -\dfrac{\mathcal{A}(x)}{\mathcal{B}(x)}\displaystyle\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t + \dfrac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\displaystyle\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t, & \text{if } s>0,\\ \dfrac{\mathcal{A}(x)}{\mathcal{B}(x)}\displaystyle\int_{0}^{\infty}\nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \dfrac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\displaystyle\int_{0}^{\infty}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t, & \text{if } s<0.\end{cases}$$

$$\mathcal{G}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big) = \begin{cases} -\dfrac{\mathcal{A}(x)}{\mathcal{B}(x)}\displaystyle\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t + \dfrac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\displaystyle\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t, & \text{if } s>0,\\ \dfrac{\mathcal{A}(x)}{\mathcal{B}(x)}\displaystyle\int_{0}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \dfrac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\displaystyle\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t, & \text{if } s<0.\end{cases}$$

Proof. When $s>0$,

$$\begin{aligned}
&\boldsymbol{\mathcal{S}}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)\\
=& -\frac{\mathcal{A}\big(\boldsymbol{X}_s(x)\big)}{\mathcal{B}\big(\boldsymbol{X}_s(x)\big)}\int_{-\infty}^{-s}\mathcal{F}_t^{(1)}\big(\boldsymbol{X}_s(x)\big)\boldsymbol{V}_{t,-s}^{(1)}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}t\\
&\quad+\frac{\Big(\mathcal{A}\big(\boldsymbol{X}_s(x)\big)\Big)^2}{\mathcal{B}\big(\boldsymbol{X}_s(x)\big)}\int_{-\infty}^{-s}\mathcal{F}_t^{(0)}\big(\boldsymbol{X}_s(x)\big)\boldsymbol{V}_{t,-s}^{(0)}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}t\\
\overset{(25),(26),(44)}{=}& -\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{-\infty}^{-s}\mathcal{F}_{t+s}^{(1)}(x)\boldsymbol{V}_{t+s,0}^{(1)}(x)\,\mathrm{d}t + \frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{-\infty}^{-s}\mathcal{F}_{t+s}^{(0)}(x)\boldsymbol{V}_{t+s,0}^{(0)}(x)\,\mathrm{d}t\\
=& -\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\boldsymbol{V}_{t,0}^{(1)}(x)\,\mathrm{d}t + \frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\boldsymbol{V}_{t,0}^{(0)}(x)\,\mathrm{d}t\\
\overset{(46)}{=}& -\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t + \frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t
\end{aligned}$$

which is clearly independent of s from this expression. Similarly, when $s<0$,

$$\begin{aligned}
&\boldsymbol{\mathcal{S}}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)\\
=& \frac{\mathcal{A}\big(\boldsymbol{X}_s(x)\big)}{\mathcal{B}\big(\boldsymbol{X}_s(x)\big)}\int_{-s}^{\infty}\mathcal{F}_t^{(1)}\big(\boldsymbol{X}_s(x)\big)\boldsymbol{V}_{t,-s}^{(1)}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}t\\
&\quad-\frac{\Big(\mathcal{A}\big(\boldsymbol{X}_s(x)\big)\Big)^2}{\mathcal{B}\big(\boldsymbol{X}_s(x)\big)}\int_{-s}^{\infty}\mathcal{F}_t^{(0)}\big(\boldsymbol{X}_s(x)\big)\boldsymbol{V}_{t,-s}^{(0)}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}t\\
=& \frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{-s}^{\infty}\mathcal{F}_{t+s}^{(1)}(x)\boldsymbol{V}_{t+s,0}^{(1)}(x)\,\mathrm{d}t - \frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{-s}^{\infty}\mathcal{F}_{t+s}^{(0)}(x)\boldsymbol{V}_{t+s,0}^{(0)}(x)\,\mathrm{d}t\\
=& \frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{0}^{\infty}\mathcal{F}_t^{(1)}(x)\boldsymbol{V}_{t,0}^{(1)}(x)\,\mathrm{d}t - \frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\boldsymbol{V}_{t,0}^{(0)}(x)\,\mathrm{d}t\\
=& \frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{0}^{\infty}\nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{0}^{\infty}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t,
\end{aligned}$$

which is again independent of s. We can similarly simplify $\mathcal{G}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)$. □

By plugging the formula in Lemma D.9 into (47),

$$
\begin{aligned}
&\int_{-\infty}^{\infty} \mathcal{F}_s^{(0)}(x)\mathcal{G}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s \\
=&\int_0^{\infty} \mathcal{F}_s^{(0)}(x)\,\mathrm{d}s\,\Big(-\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t+\frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\Big) \\
&\quad+\int_{-\infty}^{0} \mathcal{F}_s^{(0)}(x)\,\mathrm{d}s\,\Big(\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{0}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t-\frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\Big) \\
=&\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\Big(\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{0}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t-\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big).
\end{aligned}
$$

Besides,

$$
\begin{aligned}
&\int_{-\infty}^{\infty} \mathcal{F}_s^{(0)}(x)\boldsymbol{\mathcal{S}}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s \\
=&-\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{0}^{\infty}\mathcal{F}_s^{(0)}(x)\,\mathrm{d}s\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t+\frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{0}^{\infty}\mathcal{F}_s^{(0)}(x)\,\mathrm{d}s\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \\
&\quad+\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{-\infty}^{0}\mathcal{F}_s^{(0)}(x)\,\mathrm{d}s\int_{0}^{\infty}\nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t-\frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{-\infty}^{0}\mathcal{F}_s^{(0)}(x)\,\mathrm{d}s\int_{0}^{\infty}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t.
\end{aligned}
$$

By combining previous results,

$$
\begin{aligned}
&\frac{1}{2}\times\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}(x) \\
\overset{(47)}{=}&\int_{-\infty}^{\infty}\mathcal{F}_s^{(0)}(x)\boldsymbol{\mathcal{S}}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s-\nabla\Big(\int_{-\infty}^{\infty}\mathcal{F}_s^{(0)}(x)\mathcal{G}_{-s}^{\infty}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s\Big) \\
=&-\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t+\frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \\
&+\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{0}^{\infty}\nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t-\frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{0}^{\infty}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \\
&-\nabla\Big(\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\Big)\Big(\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{0}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t-\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big) \\
&-\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\nabla\Big(\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{0}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t-\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big) \\
=&\frac{\big(\mathcal{A}(x)\big)^2}{\mathcal{B}(x)}\Big(\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t-\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{0}^{\infty}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\Big) \\
&-\nabla\Big(\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\Big)\Big(\int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{0}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t-\int_{0}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big) \\
&+\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\Big(\int_{0}^{\infty}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t-\int_{-\infty}^{0}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{0}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big).
\end{aligned}
$$

It can be directly verify that

$$
\nabla\big(\frac{\mathcal{A}}{\mathcal{B}}\big)(x)=\frac{\mathcal{A}(x)}{\mathcal{B}(x)}\Big(\frac{\int_{-\infty}^{\infty}\nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}-2\frac{\int_{-\infty}^{\infty}\nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}\Big).
$$

By plugging this into the expression of the functional derivative and dividing both sides by $\frac{\mathcal{A}}{\mathcal{B}}$,

$$
\frac{\mathcal{B}(x)}{2\mathcal{A}(x)}\times\frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}(x)
$$

$$
\begin{aligned}
&= \frac{\int_{-\infty}^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\mathcal{B}(x)}\Big(\int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t - \int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\Big)\\
&- \frac{\int_{-\infty}^{\infty} \nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}\Big(\int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}\,\mathrm{d}t\Big)\\
&+ 2\frac{\int_{-\infty}^{\infty} \nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}\Big(\int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big)\\
&+ \Big(\int_0^{\infty} \nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_{-\infty}^{0} \nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big).
\end{aligned}
$$

We keep the terms involving $\int \nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t$ untouched and we only try to simplify terms involving $\int \nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t$. The coefficient for $\int_0^{\infty} \nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t$ is

$$
\begin{aligned}
&\frac{1}{\mathcal{B}(x)}\begin{pmatrix} -\int_{-\infty}^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t + 2\int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t \\ -2\int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t + \int_{-\infty}^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t \end{pmatrix}\\
&=\frac{1}{\mathcal{B}(x)}\Big(\int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big).
\end{aligned}
$$

Similarly, the coefficient for $\int_{-\infty}^{0} \nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t$ is

$$
\begin{aligned}
&\frac{1}{\mathcal{B}(x)}\begin{pmatrix} \int_{-\infty}^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t + 2\int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t \\ -2\int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_{-\infty}^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t \end{pmatrix}\\
&=\frac{1}{\mathcal{B}(x)}\Big(\int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\frac{\mathcal{B}(x)}{2\mathcal{A}(x)} \times \frac{\delta\mathcal{M}(\boldsymbol{b})}{\delta\boldsymbol{b}}(x)\\
&= \Big(\int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big)\times\\
&\qquad\qquad \Big(\frac{\int_{-\infty}^{\infty} \nabla\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t} - \frac{\int_{-\infty}^{\infty} \nabla\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}\Big)\\
&= \Big(\int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_0^{\infty} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^{0} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t\Big)\big(-\nabla\ln(\mathcal{A})(x)\big).
\end{aligned}
$$

After straightforward simplification, we obtain (38). □

E Proof of Proposition 3.1 and discussion about its assumptions and implications

In this section, we prove Proposition 3.1, and discuss its assumptions (in particular the Morse function condition) as well as some of it implications (including the settings that go beyond the ones in the proposition). In particular, we solve the Poisson equation (11) when ρ_0 is the standard Gaussian density in $\mathbb{R}^d$ and ρ_1 is the density of a Gaussian mixture distribution.

E.1 Proof of Proposition 3.1 when $\mathcal{D} = 1$

We proceed in three steps:

Step 1: We shall first establish the limiting behavior of the dynamics.

More specifically, the trajectory $t \to \boldsymbol{X}_t(x)$ will converge to a local maximum of V in the forward direction and converge to a local minimum in the backward direction, except at a set of points with measure zero.

Lemma E.1. *For any $x \in \Omega$, we have $\lim_{|t|\to\infty} |\boldsymbol{b}(\boldsymbol{X}_t(x))| = 0$.*

Proof. We only need to show one direction $t \to \infty$ and the other case follows similarly. If this does not hold, then there exists $\epsilon > 0$ and a monotone increasing sequence $\{t_k\}_{k=1}^{\infty}$ such that $\big|\boldsymbol{b}\big(\boldsymbol{X}_{t_k}(x)\big)\big| \geq \epsilon$ and $\lim_{k\to\infty} t_k = \infty$. Consider

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\big|\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\big|^2 &= \frac{\mathrm{d}}{\mathrm{d}t}\big|\nabla V\big(\boldsymbol{X}_t(x)\big)\big|^2 \\ &= 2\Big\langle \nabla V\big(\boldsymbol{X}_t(x)\big), \nabla^2 V\big(\boldsymbol{X}_t(x)\big)\nabla V\big(\boldsymbol{X}_t(x)\big)\Big\rangle \geq -2C\big|\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\big|^2,\end{aligned}$$

where $C := \sup_{x\in\Omega}\big\|\nabla^2 V(x)\big\| < \infty$. By Grönwall's inequality,

$$\big|\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\big| \geq \big|\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\big|e^{-C(t-s)},$$

for all $t \geq s \geq 0$. Without loss of generality, we can ensure that $t_k - t_{k-1} \geq 1$. Hence,

$$\begin{aligned}V\big(\boldsymbol{X}_t(x)\big) - V\big(\boldsymbol{X}_0(x)\big) &= \int_0^t \big|\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\big|^2\,\mathrm{d}s \geq \sum_{k=1}^{\infty}\chi_{[0,t]}(t_{k+1})\int_{t_k}^{t_k+1}\epsilon e^{-C(t-t_k)}\,\mathrm{d}t \\ &= \sup\{k : t_{k+1} \leq t\}\frac{\epsilon\big(1-e^{-C}\big)}{C},\end{aligned}$$

which will diverge to infinity as $t \to \infty$. This contradicts with the boundedness of V and thus the assumption does not hold. As a remark, the notation $\chi_A()$ is an indicator function for the set A. □

Lemma E.2. *Suppose x is not on the stable or unstable manifold of a saddle point. Then the trajectory $\{\boldsymbol{X}_t(x)\}_{t\in\mathbb{R}}$ must converge to a local maximum of V in the forward direction and a local minimum of V in the backward direction.*

Proof. Since the torus $\Omega = [0,1]^d$ is bounded, the trajectory $\{\boldsymbol{X}_t(x)\}_{t\geq 0}$ must be bounded and there exists an increasing sequence $\{t_k\}_{k=1}^{\infty}$ such that $\big\{\boldsymbol{X}_{t_k}(x)\big\}_{k=1}^{\infty}$ is convergent by Bolzano-Weierstrass theorem and let us denote the limit as $x^\star$. By Lemma E.1, we know $\boldsymbol{b}(x^\star) = \boldsymbol{0}_d$ and thus $x^\star$ is a critical point. By the assumption, $x^\star$ is not a saddle point nor a local minimum, that is, $x^\star$ must be a local maximum of V. By the assumption that V is a Morse function, the critical point $x^\star$ has a non-degenerate Hessian. After the trajectory enters its basin of attraction (containing an open ball around $x^\star$), the trajectory $\{\boldsymbol{X}_t(x)\}_{t\in\mathbb{R}}$ will eventually converge to $x^\star$. The backward direction can be proved in a similar way. □

Lemma E.3. *Under the same assumption as in Lemma E.2, we know $\int_0^\infty \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t < \infty$ and $\int_{-\infty}^0 \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t < \infty$. In particular, $\lim_{|t|\to\infty}\mathcal{F}_t^{(k)}(x) = 0$ and $\lim_{|t|\to\infty}\mathcal{J}_t(x) = 0$.*

Proof. Without loss of generality, we only consider the forward branch. Since V is assumed to be a Morse function, the Hessian $\nabla^2 V(x^\star) < 0$ is non-degenerate and $0 > \operatorname{tr}\big(\nabla^2 V(x^\star)\big) = \Delta V(x^\star)$. Therefore,

$$\lim_{t\to\infty}\boldsymbol{\nabla}\cdot\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big) = \boldsymbol{\nabla}\cdot\boldsymbol{b}(x^\star) = \Delta V(x^\star) < 0. \tag{48}$$

Since $\rho_k = e^{-U_k}/\mathcal{Z}_k$ are bounded on the torus, we know

$$\int_0^\infty \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t = \int_0^\infty e^{-U_k\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)\,\mathrm{d}t \leq C\int_0^\infty \mathcal{J}_t(x)\;\mathrm{d}t = C\int_0^\infty e^{\int_0^t \boldsymbol{\nabla}\cdot\boldsymbol{b}\left(\boldsymbol{X}_s(x)\right)\mathrm{d}s}\,\mathrm{d}t,$$

where $C := \sup_{x\in\Omega} \max\{e^{-U_0(x)}, e^{-U_1(x)}\} < \infty$ herein. From (48), there exists $\beta > 0$ and $\tau > 0$ such that $\boldsymbol{\nabla}\cdot\boldsymbol{b}(\boldsymbol{X}_s(x)) \le -\beta$ for all $s \ge \tau$, then if $t \ge \tau$,

$$\int_0^t \boldsymbol{\nabla}\cdot\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s \le \int_0^\tau \boldsymbol{\nabla}\cdot\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s - \beta(t-\tau),$$

and therefore,

$$\int_0^\infty \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t \le C\Big(\int_0^\tau e^{\int_0^t \boldsymbol{\nabla}\cdot\boldsymbol{b}\left(\boldsymbol{X}_s(x)\right)}\,\mathrm{d}t + \int_\tau^\infty e^{\int_0^\tau \boldsymbol{\nabla}\cdot\boldsymbol{b}\left(\boldsymbol{X}_s(x)\right)\,\mathrm{d}s} e^{-\beta(t-\tau)}\,\mathrm{d}t\Big) < \infty.$$

In particular, when $t \ge \tau$,

$$\mathcal{J}_t(x) \le e^{\int_0^\tau \boldsymbol{\nabla}\cdot\boldsymbol{b}\left(\boldsymbol{X}_s(x)\right)\,\mathrm{d}s} e^{-\beta(t-\tau)},$$

which converges to zero exponentially fast as $t \to \infty$. The same conclusion holds for $\mathcal{F}_t^{(k)}(x) \equiv e^{-U_k\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)$ when $t \to \infty$. □

Step 2: We verify that $\boldsymbol{b} = \nabla V$ is a zero-variance dynamics.

We need to show that $\int_{-\infty}^\infty (\rho_1 - \rho_0)\big(\boldsymbol{X}_t(x)\big)\mathcal{J}_t(x)\,\mathrm{d}t = 0$ almost everywhere on Ω.

Under the same assumption as Lemma E.2, let us consider

$$\begin{aligned}\int_{-\infty}^\infty (\rho_1 - \rho_0)\big(\boldsymbol{X}_t(x)\big)\mathcal{J}_t(x)\,\mathrm{d}t &= \int_{-\infty}^\infty \Delta V\big(\boldsymbol{X}_t(x)\big) e^{\int_0^t \Delta V\left(\boldsymbol{X}_s(x)\right)\,\mathrm{d}s}\,\mathrm{d}t \\ &= \int_{-\infty}^\infty \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}_t(x)\,\mathrm{d}t \\ &= \lim_{t\to\infty}\mathcal{J}_t(x) - \lim_{t\to-\infty}\mathcal{J}_t(x) = 0.\end{aligned}$$

The last line comes from Lemma E.3. The validity of the above equation almost everywhere on Ω will be explained in Step 3.

Step 3: We prove that $\boldsymbol{b} = \nabla V \in \mathfrak{B}_\infty$ in the sense of Definition A.4, i.e., such a gradient ascent dynamics is a valid one for the infinite-time NEIS scheme.

If we can find two open subsets D_1, D_2 such that $x \to \int_0^\infty \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t$ is continuous on D_1, $x \to \int_{-\infty}^0 \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t$ is continuous on D_2, and both $\Omega\backslash D_1$ and $\Omega\backslash D_2$ have Lebesgue measure zero, then clearly $\mho(\boldsymbol{b}) \supset D_1 \cap D_2$ and $\boldsymbol{b} \in \mathfrak{B}_\infty$.

Due to the symmetric role of forward and backward branches of trajectories, it is then sufficient to prove the following lemma.

Lemma E.4. *There exists an open subset $D \subset \Omega$ such that $x \to \int_0^\infty \mathcal{F}_t^{(k)}(x)$ is continuous and $\Omega\backslash D$ has measure zero.*

Proof. Let us denote the local maxima of V as $\mathscr{X}_1, \mathscr{X}_2, \cdots, \mathscr{X}_r$. The index $r < \infty$ because V is a Morse function and Ω is compact. Since $\nabla^2 V(\mathscr{X}_i) < 0$ for $1 \le i \le r$, there exists a local neighborhood $B_{\delta_i}(\mathscr{X}_i)$ such that $\lim_{t\to\infty}\boldsymbol{X}_t(z) = \mathscr{X}_i$ if $z \in B_{\delta_i}(\mathscr{X}_i)$. Hence, it is not hard to characterize the basin of attraction of $\mathscr{X}_i$

$$O_i = \Big\{\boldsymbol{X}_t(x) :\ x \in B_{\delta_i}(\mathscr{X}_i),\ t \le 0\Big\}$$

which is open. Then define an open subset $D := \cup_{i=1}^r O_i$. By Lemma E.2, we know $\Omega\backslash D$ has Lebesgue measure zero, since there is only a finite number of critical points and the stable/unstable manifold of saddle points has measure zero. Next, we still need to verify $z \to \int_0^\infty \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t$ is continuous at an arbitrary point $x \in D$. Since $D = \cup_{i=1}^r O_i$ and O_i are open and disjoint, it is sufficient to verify this conclusion for $x \in O_i$ for an arbitrary index i.

By the smoothness of V, there exists a local neighborhood $O_\delta := \big\{z \in O_i :\ V(\mathscr{X}_i) - \delta < V(z) \le V(\mathscr{X}_i)\big\}$ such that $\frac{(\boldsymbol{\nabla}\cdot\boldsymbol{b})(z)}{(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\mathscr{X}_i)} \in (\frac{1}{2}, \frac{3}{2})$ for every $z \in O_\delta$. Let $M = \sup\big\{\max\{e^{-U_1(x)}, e^{-U_0(x)}\} :$

$x \in \Omega\}$. We know $M < \infty$ because Ω is compact and U_0, U_1 are smooth. Define $\tau := \inf\{t \geq 1 : \boldsymbol{X}_t(x) \in O_\delta\}$. Due to the smoothness of $\boldsymbol{b}$, for any integer $j \geq 2$, we can choose a small neighborhood $B_{\delta_j}(x)$ such that $\boldsymbol{X}_t(z) \in O_\delta$ for all $t \geq j\tau$ and for all $z \in B_{\delta_j}(x)$ (note that O_δ is automatically a trapping region of $\boldsymbol{b}$ by construction). Then when $z \in B_{\delta_j}(x)$,

$$\begin{aligned}
\int_{j\tau}^{\infty} \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t &\leq \int_{j\tau}^{\infty} M\mathcal{J}_t(z)\,\mathrm{d}t \\
&\leq \int_{j\tau}^{\infty} M\mathcal{J}_{j\tau}(z) e^{(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\mathscr{X}_i)\frac{1}{2}(t-j\tau)}\,\mathrm{d}t \\
&\leq 2M\mathcal{J}_{j\tau}(z)\frac{1}{-(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\mathscr{X}_i)} \\
&= \frac{2M}{-(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\mathscr{X}_i)}\Big(\mathcal{J}_{j\tau}(z) - \mathcal{J}_{j\tau}(x) + \mathcal{J}_{j\tau}(x)\Big).
\end{aligned}$$

For an arbitrary $\epsilon > 0$, by Lemma E.3, we can pick j large enough such that $\mathcal{J}_{j\tau}(x) < \frac{-(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\mathscr{X}_i)}{2M}\frac{\epsilon}{6}$. Next we can accordingly pick δ_j small enough such that $\mathcal{J}_{j\tau}(z) - \mathcal{J}_{j\tau}(x) < \frac{-(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\mathscr{X}_i)}{2M}\frac{\epsilon}{6}$ for all $z \in B_{\delta_j}(x)$. In this way, we can ensure that

$$\int_{j\tau}^{\infty} \mathcal{F}_t^{(k)}(z) \leq \frac{\epsilon}{6} + \frac{\epsilon}{6} = \frac{\epsilon}{3}, \qquad \forall z \in B_{\delta_j}(x). \tag{49}$$

Furthermore, due to the smoothness of $\boldsymbol{b}$, we can choose δ_j (possibly even smaller) so that

$$\Big|\int_0^{j\tau} \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t - \int_0^{j\tau} \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t\Big| \leq \frac{\epsilon}{3}, \qquad \forall z \in B_{\delta_j}(x). \tag{50}$$

The continuity of $z \to \int_0^{j\tau} \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t$ can be easily established due to the differentiability of $z \to \mathcal{F}_t^{(k)}(z)$. By combining previous results, for each $\epsilon > 0$, we can find a δ_j such that for any $z \in B_{\delta_j}(x)$,

$$\begin{aligned}
&\Big|\int_0^{\infty} \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t - \int_0^{\infty} \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t\Big| \\
\leq &\Big|\int_0^{j\tau} \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t - \int_0^{j\tau} \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t\Big| + \int_{j\tau}^{\infty} \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t + \int_{j\tau}^{\infty} \mathcal{F}_t^{(k)}(x)\,\mathrm{d}t \\
\overset{(49),(50)}{\leq} &\frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

This proves the continuity of $z \to \int_0^{\infty} \mathcal{F}_t^{(k)}(z)\,\mathrm{d}t$ at the point x. □

E.2 A remark about the general case

We shall prove that $\boldsymbol{b} = \mathcal{D}\nabla V$ is a zero-variance dynamics where V solves (11).

Notice that $\mathcal{J}_t(x) = e^{\int_0^t \nabla\cdot(\mathcal{D}V)\left(\boldsymbol{X}_s(x)\right)\,\mathrm{d}s}$ and

$$\begin{aligned}
\int_{-\infty}^{\infty} \big(\rho_1(\boldsymbol{X}_t(x)) - \rho_0(\boldsymbol{X}_t(x))\big)\mathcal{J}_t(x)\,\mathrm{d}t &= \int_{-\infty}^{\infty} \nabla\cdot(\mathcal{D}\nabla V)(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}_t(x)\,\mathrm{d}t = \lim_{t\to\infty}\mathcal{J}_t(x) - \lim_{t\to-\infty}\mathcal{J}_t(x).
\end{aligned}$$

As long as $\mathcal{J}_t(x)$ vanishes when $|t| \to \infty$, then such a dynamics $\boldsymbol{b} = \mathcal{D}\nabla V$ is indeed a zero-variance dynamics.

For a point $x \in \Omega$, suppose the gradient ascent trajectory under ∇V will converge to a (non-degenerate) local maximum of V, denoted as $x^\star$; by Proposition 2.2, the trajectory initiated from x under $\boldsymbol{b} = \mathcal{D}\nabla V$ is the same and $\boldsymbol{X}_t(x) \to x^\star$ as $t \to \infty$ under the flow $\boldsymbol{b}$. Since $\mathcal{D}$ is strictly

positive, it also does not change the concavity of local extreme points: when $x^\star$ is a local maximum of V (with $\nabla V(x^\star) = \mathbf{0}_d$ and $\nabla^2 V(x^\star) < 0$), then

$$\nabla \boldsymbol{b}(x^\star) = \nabla V(x^\star)\nabla \mathcal{D}(x^\star)^T + \mathcal{D}(x^\star)\nabla^2 V(x^\star) = \mathcal{D}(x^\star)\nabla^2 V(x^\star) < 0,$$

which implies that as $t \to \infty$,

$$\nabla \cdot \boldsymbol{b}\big(\boldsymbol{X}_t(x)\big) \to \nabla \cdot \boldsymbol{b}(x^\star) = \operatorname{tr}\big(\nabla \boldsymbol{b}(x^\star)\big) < 0.$$

By the same argument as in the case $\mathcal{D} = 1$ (i.e. Lemma E.3), we can establish the validity that $\mathcal{J}_t(x) \to 0$ as $|t| \to \infty$.

E.3 A remark about the existence of Morse function

In Poisson's equation (11), a Morse function V does not always exist for an arbitrary smooth density function ρ_1, e.g., when $\rho_1 = \rho_0$, $\mathcal{D} = 1$, we know $V = 0$ is the solution of (11) but $V = 0$ is not a Morse function. However, since Morse functions are dense in $C^\infty(\Omega, \mathbb{R})$ [4], we can always find a Morse function such that the dynamics $\boldsymbol{b} = \nabla V$ behaves almost like a zero-variance dynamics, which is summarized in the next proposition.

Proposition E.5. *Suppose $\Omega = [0,1]^d$ is a torus and $U_0, U_1 \in C^\infty(\Omega, \mathbb{R})$. Without loss of generality, assume $\mathcal{Z}_0 = \mathcal{Z}_1 = 1$. For any $\epsilon \in (0,1)$, there exists a Morse function V such that the dynamics $\boldsymbol{b} = \nabla V$ provides an estimator $1 - \epsilon \le \mathcal{A}(x) \le 1 + \epsilon$ for almost all $x \in \Omega$ in the infinite-time NEIS method. Consequently, the variance Var$(\boldsymbol{b}) \le \epsilon^2$.*

Proof. Denote $\theta := \inf\{\rho_0(x) : x \in \Omega\} \equiv e^{-\sup\{U_0(x):\, x \in \Omega\}} > 0$ since U_0 is smooth and Ω is compact. Since both $\rho_k = e^{-U_k}$ are smooth for $k = 0, 1$, one could approximate $\rho_1 - \rho_0$ by trigonometric polynomials $T_N(x) = \sum_{|\mu|_\infty \le N, \mu \ne \mathbf{0}_d} a_\mu e^{i2\pi\langle \mu, x\rangle}$ such that

$$\|(\rho_1 - \rho_0) - T_N\|_{C^0(\Omega)} < \frac{\epsilon\theta}{2},$$

where $a_\mu = \int_\Omega e^{-i2\pi\langle \mu, x\rangle}(\rho_1 - \rho_0)(x)\,\mathrm{d}x \in \mathbb{C}$ are Fourier coefficients, $\mu \in \mathbb{Z}^d$ and $N \in \mathbb{N}$; see e.g., [7, Theorem 16]. Let $\Psi_N(x) = \sum_{|\mu|_\infty \le N, \mu \ne \mathbf{0}_d} \frac{a_\mu}{-4\pi^2|\mu|^2} e^{i2\pi\langle \mu, x\rangle} \in C^\infty(\Omega, \mathbb{R})$. It is clear that $\Delta\Psi_N = T_N$. As Morse functions are dense, we can find a Morse function V such that $\|V - \Psi_N\|_{C^2(\Omega)} < \frac{\epsilon\theta}{2}$ [4, Proposition 1.2.4], and in particular, $\|\Delta V - \Delta\Psi_N\|_{C^0(\Omega)} < \frac{\epsilon\theta}{2}$. Therefore,

$$\|\Delta V - (\rho_1 - \rho_0)\|_{C^0(\Omega)} \le \|\Delta V - \Delta\Psi_N\|_{C^0(\Omega)} + \|\Delta\Psi_N - (\rho_1 - \rho_0)\|_{C^0(\Omega)} \le \epsilon\theta. \tag{51}$$

By Proposition 3.1, we know that $\boldsymbol{b} = \nabla V$ is a zero-variance dynamics for $\widetilde{\rho}_1 := \rho_0 + \Delta V$. The Proposition 3.1 is proved under the assumption that densities are positive smooth functions for convenience and it is straightforward to verify that it also holds if ρ_1 is an arbitrary smooth function in Proposition 3.1. In particular, using the same argument in Appendix E.1 Step 2, we have for almost all $x \sim \rho_0$,

$$\frac{\int_{\mathbb{R}} \widetilde{\rho}_1(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{\mathbb{R}} \rho_0(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t} = 1 + \frac{\int_{\mathbb{R}} \Delta V(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{\mathbb{R}} \rho_0(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t} = 1 + \frac{\mathcal{J}_t(x)|_{t=-\infty}^{t=\infty}}{\int_{\mathbb{R}} \rho_0(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t} = 1. \tag{52}$$

Hence,

$$\begin{aligned}
|\mathcal{A}(x) - 1| &\overset{(9)}{=} \left|\frac{\int_{\mathbb{R}} \rho_1(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{\mathbb{R}} \rho_0(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t} - 1\right| \\
&= \left|\frac{\int_{\mathbb{R}} \widetilde{\rho}_1(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{\mathbb{R}} \rho_0(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t} + \frac{\int_{\mathbb{R}} (\rho_1 - \widetilde{\rho}_1)(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{\mathbb{R}} \rho_0(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t} - 1\right| \\
&\overset{(52)}{=} \left|\frac{\int_{\mathbb{R}} (\rho_1 - \widetilde{\rho}_1)(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{\mathbb{R}} \rho_0(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t}\right| \\
&\le \frac{\int_{\mathbb{R}} |(\rho_1 - \widetilde{\rho}_1)(\boldsymbol{X}_t(x))|\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{\mathbb{R}} \rho_0(\boldsymbol{X}_t(x))\mathcal{J}_t(x)\,\mathrm{d}t}
\end{aligned}$$

$$\overset{(51)}{\leq} \frac{\int_{\mathbb{R}} \epsilon\theta\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{\mathbb{R}} \theta\mathcal{J}_t(x)\,\mathrm{d}t} = \epsilon.$$

Since we assumed $\mathcal{Z}_1 = 1$,

$$\mathrm{Var}(\boldsymbol{b}) = \mathbb{E}_0[|\mathcal{A}|^2] - \big(\mathcal{Z}_1\big)^2 = \mathbb{E}_0|\mathcal{A} - \mathcal{Z}_1|^2 \leq \epsilon^2.$$

□

E.4 Solution of Poisson's equation (11) for Gaussian mixtures

Lemma E.6. *One solution of the Poisson's equation $\Delta V = Ce^{-\frac{|x-\mu|^2}{2\sigma^2}}$ with $d \geq 2$ on $\Omega = \mathbb{R}^d$ is $V(x) = f(|x-\mu|)$, where the function $f : \mathbb{R}^+ \to \mathbb{R}$ has the derivative*

$$f'(r) = C2^{d/2-1}\sigma^d r^{1-d} \int_0^{\frac{r^2}{2\sigma^2}} t^{d/2-1}e^{-t}\,\mathrm{d}t \equiv C2^{d/2-1}\sigma^d r^{1-d}\Gamma\big(d/2, \frac{r^2}{2\sigma^2}\big),$$

where $\Gamma\big(a, x\big) := \int_0^x t^{a-1}e^{-t}\,\mathrm{d}t$ is the lower incomplete gamma function.

Proof. Without loss of generality, let $\mu = \mathbf{0}_d$. Then a natural radial solution is given as $V(x) = f(|x|)$ for some scalar-valued function f. The above Poisson's equation becomes

$$f''(|x|) + f'(|x|)\frac{d-1}{|x|} = Ce^{-\frac{|x|^2}{2\sigma^2}}.$$

By some straightforward computation,

$$f'(r) = Cr^{-(d-1)} \int_0^r s^{d-1}e^{-s^2/(2\sigma^2)}\,\mathrm{d}s = C2^{d/2-1}\sigma^d r^{1-d} \underbrace{\int_0^{\frac{r^2}{2\sigma^2}} t^{d/2-1}e^{-t}\,\mathrm{d}t}_{\equiv\Gamma\left(d/2, \frac{|r|^2}{2\sigma^2}\right)}.$$

□

Proposition E.7. *Suppose V solves the following Poisson's equation on $\Omega = \mathbb{R}^d$*

$$\Delta V = \rho_1 - \rho_0$$

where

$$\rho_0(x) = \frac{1}{\sqrt{2\pi}^d} \exp\Big(-\frac{|x|^2}{2}\Big)$$

$$\rho_1(x) = \sum_{i=1}^n \omega_i \frac{1}{\sqrt{2\pi\sigma_i^2}^d} \exp\Big(-\frac{|x-\mu_i|^2}{2\sigma_i^2}\Big), \qquad n \in \mathbb{N},\ \mu_i \in \mathbb{R}^d, \sigma_i \in \mathbb{R}^+,\ \forall 1 \leq i \leq n,$$

and $\mu_i \neq \mu_j$ if $i \neq j$. Then one solution for the gradient flow dynamics $\boldsymbol{b} = \nabla V$ is given as

$$\boldsymbol{b}(x) = 2^{-1}\pi^{-d/2}\Big(\sum_{i=1}^n \omega_i|x-\mu_i|^{-d}\Gamma\big(d/2, \frac{|x-\mu_i|^2}{2\sigma_i^2}\big)(x-\mu_i) - |x|^{-d}\Gamma\big(d/2, \frac{|x|^2}{2}\big)x\Big). \quad (53)$$

Proof. We just need to apply the last lemma and the formula $\nabla V(x) = \frac{f'(|x|)}{|x|}x$ if $V(x) = f(|x|)$. □

Proposition E.8. *Suppose $\boldsymbol{b}$ is given in (53). Then*

$$\lim_{\left(\max_{i=1}^n \sigma_i\right)\to 0} \boldsymbol{b}(x)$$

$$= \begin{cases} 2^{-1}\pi^{-d/2}\Big(\sum_{i\neq j} \omega_i|x-\mu_i|^{-d}\Gamma(d/2)(x-\mu_i) - |x|^{-d}\Gamma\big(d/2, \frac{|x|^2}{2}\big)x\Big), & \text{if } x = \mu_j, \\ 2^{-1}\pi^{-d/2}\Big(\sum_{i} \omega_i|x-\mu_i|^{-d}\Gamma(d/2)(x-\mu_i) - |x|^{-d}\Gamma\big(d/2, \frac{|x|^2}{2}\big)x\Big), & \text{otherwise}. \end{cases}$$

The limiting dynamics $x \to \lim_{\left(\max_{i=1}^n \sigma_i\right)\to 0} \boldsymbol{b}(x)$ is continuous on the region $\mathbb{R}^d \backslash \{\mu_i\}_{i=1}^n$.

Proof. If $x = \mu_j$, then

$$\boldsymbol{b}(x) = \frac{1}{2\pi^{d/2}}\Big(\sum_{i\neq j}\omega_i|x-\mu_i|^{-d}\Gamma\big(d/2, \frac{|x-\mu_i|^2}{2\sigma_i^2}\big)(x-\mu_i) - |x|^{-d}\Gamma\big(d/2, \frac{|x|^2}{2}\big)x\Big).$$

When $\sigma_i \to 0$ for all i,we know $\Gamma\big(d/2, \frac{|\mu_j-\mu_i|^2}{2\sigma_i^2}\big) \to \Gamma(d/2)$ when $i \neq j$ and hence we have the above result. Similarly, we can obtain the expression when $x \neq \mu_j$ for any j. □

E.5 Example: Poisson's equation yields a zero-variance dynamics

The example for Figure 1 is

$$\begin{aligned} \rho_0(x) &= e^{-U_0(x)} = 1, \\ \rho_1(x) \propto e^{-U_1(x)} &= \frac{\phi\big(x - \left[\begin{smallmatrix}0.3\\0.3\end{smallmatrix}\right]\big) + \phi\big(x - \left[\begin{smallmatrix}0.7\\0.3\end{smallmatrix}\right]\big) + \phi\big(x - \left[\begin{smallmatrix}0.3\\0.7\end{smallmatrix}\right]\big)}{3}, \\ \phi(x) &= e^{2\cos(2\pi x_1)+2\cos(2\pi x_2)}. \end{aligned} \tag{54}$$

The periodic boundary condition in Proposition 3.1 helps to ease the technicalities in proving that the gradient dynamics $\boldsymbol{b} = \nabla V$ from solving the Poisson's equation (11) is a zero-variance dynamics, by removing the effect from the boundary $\partial\Omega$. The same conclusion, however, should hold if V solves the Poisson's equation with Neumann boundary condition:

$$\Delta V = \rho_1 - \rho_0, \qquad \nabla V \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega, \tag{55}$$

where $\boldsymbol{n}$ is the normal vector of the boundary $\partial\Omega$. We consider the same model (54) and $\Omega = (0,1)^2$. The potential V and flowlines of $\boldsymbol{b} = \nabla V$ are visualized in Figure 3 and we can numerically verify that $\mathcal{A}(x) = \mathcal{Z}_1$ for almost all $x \in \Omega$.

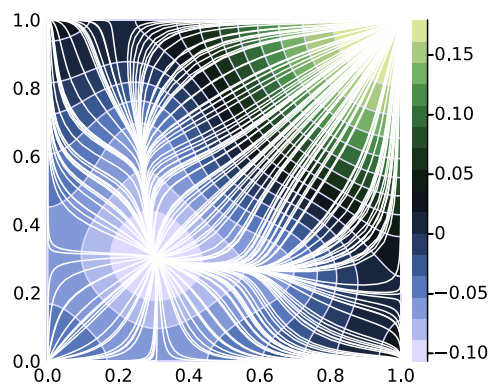


Figure 3: Contour plot of V and flowlines of $\boldsymbol{b} = \nabla V$ for the model (54) on the domain $\Omega = (0,1)^2$ with Neumann boundary condition.

E.6 Non-uniqueness of zero-variance dynamics

Recall from Proposition 2.2 that there are certain degrees of freedom to choose the dynamics: for a given $\boldsymbol{b}_1$, if we choose $\boldsymbol{b}_2 = \alpha\boldsymbol{b}_1$ where $\alpha \in C^\infty(\Omega, \mathbb{R})$ is strictly positive, then this function α can be absorbed into the time rescaling and it does not affect the variance of the sampling scheme. However, even if we remove this parameterization redundancy, zero-variance dynamics may still not be unique, e.g., due to the geometric rotational symmetry.

Proposition E.9 (Non-uniqueness)**.** *For given ρ_0 and ρ_1, there might exist more than one zero-variance dynamics (let us say $\boldsymbol{b}_1, \boldsymbol{b}_2$) but there is no scalar-valued function α such that $\boldsymbol{b}_2 = \alpha\boldsymbol{b}_1$.*

Proof. We construct an example to prove the non-uniqueness: let $d = 2$, $U_0(x) = |x|^2/2 + \ln(2\pi)$ and U_1 be given by

$$\exp\big(-U_1(x)\big) = \exp\big(-U_0(x)\big) + \frac{1}{2\pi} x_1 x_2 \exp\big(-x_1^4 - x_2^4\big).$$

We can easily verify that $|x|e^{-x^4} < \frac{3}{4}e^{-x^2/2}$ for any $x \in \mathbb{R}$. Then we know $\frac{7}{16}e^{-U_0(x)} < e^{-U_1(x)} < \frac{25}{16}e^{-U_0(x)}$. Therefore, U_1 is clearly well-defined and $\mathcal{Z}_1 = 1$. For the dynamics $\boldsymbol{b}(x) = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$ with $v_1^2 + v_2^2 > 0$, we have $\mathcal{J}_t(x) = 1$ for any $x \in \Omega, t \in \mathbb{R}$, and

$$\begin{aligned}
&\int_{-\infty}^{\infty} e^{-U_1\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t \\
=&\int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t \\
&\qquad + \frac{1}{2\pi}\int_{-\infty}^{\infty} (x_1 + v_1 t)(x_2 + v_2 t) \exp\big(-(x_1 + v_1 t)^4 - (x_2 + v_2 t)^4\big)\,\mathrm{d}t.
\end{aligned}$$

When either $v_1 = 0$ or $v_2 = 0$, we can easily verify that

$$\frac{\int_{-\infty}^{\infty} e^{-U_1\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t} = 1, \qquad \forall x \in \mathbb{R}^2.$$

Therefore, the variance is zero for two dynamics with orthogonal directions $\boldsymbol{b} = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and $\boldsymbol{b} = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$. It is clear that there is no scalar-valued function α such that $\boldsymbol{b}_2 = \alpha \boldsymbol{b}_1$, and thus the non-uniqueness is established. □

E.7 Connection to the Beckmann's problem.

The Poisson's equation (11) with $\mathcal{D} = 1$ is the Euler-Lagrange equation associated with

$$\min \int_{\Omega} |\boldsymbol{b}(x)|^p \,\mathrm{d}x \quad \text{subject to} \quad \boldsymbol{\nabla} \cdot \boldsymbol{b} = \rho_1 - \rho_0, \tag{56}$$

when $p = 2$. The variational problem in (56) is known as Beckman's problem of continuous transportation [5]; when $p = 1$, it is also related to optimal transport in W_1 Wasserstein distance [36, 37].

F Explicitly solvable zero-variance dynamics

In this section, we provide some examples with explicitly solvable zero-variance dynamics. Throughout this section, we consider $\Omega = \mathbb{R}^d$.

Table 1: Examples with explicitly solvable zero-variance dynamics for the infinite-time case with the domain $\Omega = \mathbb{R}^d$. By Proposition 2.2, given a zero-variance dynamics $\boldsymbol{b}$, any dynamics of the form $\alpha\boldsymbol{b}$ for some scalar-valued positive function α is also a zero-variance dynamics. In this table, we have removed such a degree of freedom.

Dimension	U_0 and U_1	$\boldsymbol{b}$	Details
$d = 1$	arbitrary	$\boldsymbol{b}(x) = 1$	Section F.2
general d	$U_0(x) = \lvert x\rvert^2/2 + \frac{d}{2}\ln(2\pi)$ $U_1(x) = (x - \varpi)^T \Sigma^{-1}(x - \varpi)/2$	$\boldsymbol{b}(x) = \Lambda x + v$ with $\Lambda = \ln\big(\Sigma^{-1/2}\big)$, $v = -\big(\mathbb{I}_d - \Sigma^{1/2}\big)^{-1} \ln\big(\Sigma^{-1/2}\big)\varpi$.	Section F.3
general d	ρ_0 and ρ_1 has the same marginal distribution on the orthogonal subspace of $\{cv : c \in \mathbb{R}\}$	$\boldsymbol{b}(x) = v$	Section F.4

F.1 Some general properties

Given a $\boldsymbol{b} \in \mathfrak{B}$ and the distribution ρ_0, we study the family of U_1 such that $\boldsymbol{b}$ is a zero-variance dynamics. Given an ODE flow map $\boldsymbol{X}_\tau(\cdot)$ based on the dynamics $\boldsymbol{b}$, let us introduce

$$U^\tau(x) := U_0\big(\boldsymbol{X}_{-\tau}(x)\big) - \log\big(\mathcal{J}_{-\tau}(x)\big). \tag{57}$$

Then $\rho \propto e^{-U^\tau}$ is the push-forward distribution of the flow map $\boldsymbol{X}_\tau(\cdot)$, i.e., $\rho = \big(\boldsymbol{X}_\tau(\cdot)\big)\#\rho_0$. The family of distributions that can be written as a linear combination of such pushforward distributions can be characterized by

$$\mathfrak{F} := \Big\{U :\ e^{-U} \in \text{Span}\big\{e^{-U^\tau}\big\}_{\tau\in\mathbb{R}}\Big\}.$$

We have:

Proposition F.1. *For every $U_1 \in \mathfrak{F}$, the variance Var($\boldsymbol{b}$) (if well-defined) is exactly zero, i.e., if the distribution $\rho_1 \propto e^{-U_1}$ is a linear combination of push-forward distributions by the ODE flows maps, then $\mathcal{Z}_1$ can be estimated with zero-variance by the infinite-time NEIS scheme.*

This proposition is proven in Section F.5 below. In words it says that, if we can learn a perfect neural ODE such that $\rho_1 = \boldsymbol{X}_\tau(\cdot)\#\rho_0$ for some τ, then such a dynamics is also the optimal one (i.e., zero-variance dynamics) for the infinite-time NEIS scheme. Conversely, if $U \notin \mathfrak{F}$, then is it still possible that Var($\boldsymbol{b}$) $= 0$? The answer is positive:

Proposition F.2. *The exists a dynamics $\boldsymbol{b} \in \mathfrak{B}$ and a $\rho_1 \propto e^{-U_1}$ such that $\mathcal{Z}_1$ can be computed with zero-variance but ρ_1 does not need to be a linear combination of $\big\{\boldsymbol{X}_\tau(\cdot)\#\rho_0\big\}_{\tau\in\mathbb{R}}$ (namely, $U_1 \notin \mathfrak{F}$).*

This proposition is proven in Section F.6 below.

F.2 Flows for the 1D case

Let us consider $\boldsymbol{b} = 1$. Then we can compute $\mathcal{Z}_1$ via the infinite-time NEIS scheme with zero-variance for arbitrary potentials U_0 and U_1. This could be verified via direct computation: $\mathcal{J}_t(x) = 1$ and $\boldsymbol{X}_t(x) = x + t$ for any $t, x \in \mathbb{R}$, and thus for an arbitrary x,

$$\mathcal{A}(x) = \frac{\int_{-\infty}^{\infty} e^{-U_1\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)\,\mathrm{d}t} = \frac{\int_{-\infty}^{\infty} e^{-U_1(x+t)}\,\mathrm{d}t}{\int_{-\infty}^{\infty} e^{-U_0(x+t)}\,\mathrm{d}t} = \mathcal{Z}_1.$$

Therefore, the variance is exactly zero.

Another perspective to understand this comes from Proposition F.1. For $\boldsymbol{b} = 1$, we know $e^{-U^\tau(x)} = e^{-U_0(x-\tau)}$ in (57). By Proposition F.1, if the potential U_1 can be expressed as follows

$$e^{-U_1} = \int_{-\infty}^{\infty} f(\tau)e^{-U^\tau}\ d\tau = (e^{-U_0} * f),$$

then $\mathcal{Z}_1$ can be computed with zero-variance, where f is a tempered distribution and $*$ means the convolution. Then it is sufficient to show the existence of such a f for a generic U_1.

For a given potential U_1, we can solve the above equation for f using Fourier transform; more specifically,

$$f = \mathscr{F}^{-1}\Big(\mathscr{F}(e^{-U_1})/\mathscr{F}(e^{-U_0})\Big),$$

where $\mathscr{F}^{(-1)}$ are (inverse) Fourier transform.

F.3 Linear flows for Gaussian distributions

Proposition F.3. *Suppose $U_0(x) = |x|^2/2 + \frac{d}{2}\ln(2\pi)$ and $U_1(x) = (x-\varpi)^T\Sigma^{-1}(x-\varpi)/2$, where the covariance matrix Σ is non-degenerate. A zero-variance linear dynamics is $\boldsymbol{b}(x) = \Lambda x + v$ with*

$$\Lambda = \ln\Big(\Sigma^{-1/2}\Big), \qquad v = -\Big(\mathbb{I}_d - \Sigma^{1/2}\Big)^{-1}\ln\Big(\Sigma^{-1/2}\Big)\varpi. \tag{58}$$

Before presenting detailed proofs, let us make a few remarks about zero-variance dynamics in (58):

- If we further let $\Sigma = (1-\epsilon)\mathbb{I}_d$ where $\epsilon \ll 1$ is an asymptotic parameter, then the above choice (58) can be approximated as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t(x) = \Lambda \boldsymbol{X}_t(x) + v \approx \frac{\epsilon}{2}\boldsymbol{X}_t(x) - \big(1+\frac{\epsilon}{4}\big)\varpi + \mathcal{O}(\epsilon^2).$$

  The leading order dynamics $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t(x) \approx -\varpi$ is consistent with the parallel velocity case below in Section F.4.

- For the above Gaussian case, the dynamics (58) can be regarded as the gradient flow dynamics of the following quadratic potential

$$V(x) = \frac{1}{2}\Big(x - \big(\mathbb{I}_d - \Sigma^{1/2}\big)^{-1}\varpi\Big)^T \ln\Big(\Sigma^{-1/2}\Big)\Big(x - \big(\mathbb{I}_d - \Sigma^{1/2}\big)^{-1}\varpi\Big).$$

  Indeed, it is not hard to guess that the optimal dynamics might have a linear form in order to transport Gaussians. It is natural to guess that $\boldsymbol{b} = -\nabla U_1$ or $\boldsymbol{b} = -\nabla(U_1 - U_0)$ might be zero-variance dynamics, but it could be verified that neither of them are zero-variance dynamics.

Proof of Proposition F.3. Suppose we consider a family of linear dynamics

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t(x) = \Lambda \boldsymbol{X}_t(x) + v, \qquad \boldsymbol{X}_0(x) = x.$$

Then $\boldsymbol{X}_t(x) = e^{\Lambda t}x + \Lambda^{-1}\big(e^{\Lambda t} - \mathbb{I}_d\big)v$ and $(\nabla\cdot\boldsymbol{b})(x) = \mathrm{tr}(\Lambda)$. Therefore, $U^{-\tau}$ defined in (57) has the following form for any $\tau \in \mathbb{R}$,

$$U^{-\tau}(x) = U_0\Big(e^{\Lambda\tau}x + \Lambda^{-1}(e^{\Lambda\tau} - \mathbb{I}_d)v\Big) - \mathrm{tr}(\Lambda)\tau.$$

By Proposition F.1, the variance is zero if

$$U_1(x) = U^{-1}(x) + C$$

where C is some constant. This condition can be simplified as

$$(x-\varpi)^T\Sigma^{-1}(x-\varpi)/2 = U_0\Big(e^{\Lambda}x + \Lambda^{-1}(e^{\Lambda} - \mathbb{I}_d)v\Big) - \mathrm{tr}(\Lambda) + C.$$

Thus, we just need to ensure

$$-\Lambda^{-1}(\mathbb{I}_d - e^{-\Lambda})v = \varpi, \qquad e^{\Lambda^T}e^{\Lambda} = \Sigma^{-1},$$

by matching the order of x and, more specifically, we can choose Λ and v as in (58). □

F.4 Flows with parallel velocity

As we have mentioned, in the 1D case, the choice $\boldsymbol{b} = 1$ gives a zero-variance estimator for the infinite-time NEIS scheme. A straightforward generalization is to consider the following parallel velocity case

$$\boldsymbol{b}(x) = \alpha(x)v,$$

where $v \in \mathbb{R}^d$ and α is a positive scalar-valued function. Due to Proposition 2.2, it suffices to consider $\boldsymbol{b}(x) = v$. As we can always rotate the coordinate without affecting partition functions, without loss of generality, let us assume $v = \boldsymbol{e}_1$ for simplicity, where $\boldsymbol{e}_1$ is a vector with the first element to be 1 and zeros otherwise. For an arbitrary initial proposal x, only the first coordinate x_1 is changing under the dynamics $\boldsymbol{b} = \boldsymbol{e}_1$. The estimator essentially works like the 1D case above:

$$\mathcal{A}(x) = \frac{\int_{-\infty}^{\infty} e^{-U_1\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)\,\mathrm{d}t} = \frac{\int_{-\infty}^{\infty} e^{-U_1(q,x_2,x_3,\cdots,x_d)}\,\mathrm{d}q}{\int_{-\infty}^{\infty} e^{-U_0(q,x_2,x_3,\cdots,x_d)}\,\mathrm{d}q} = \mathcal{Z}_1 \times \frac{\widetilde{\rho}_1(x_2,x_3,\cdots,x_d)}{\widetilde{\rho}_0(x_2,x_3,\cdots,x_d)},$$

where $\widetilde{\rho}_k$ is the marginal distribution of ρ_k for the subspace $\mathbb{R}^{d-1}$ by tracing out the first coordinate. Therefore, the dynamics $\boldsymbol{b} = \boldsymbol{e}_1$ is a zero-variance dynamics iff $\widetilde{\rho}_0 = \widetilde{\rho}_1$.

Proposition F.4. *Suppose $\boldsymbol{b}(x) = \alpha(x)v$ where $\alpha \in C^\infty(\mathbb{R}^d,\mathbb{R})$ with $\inf_{x\in\Omega}\alpha(x) > 0$ and $v \in \mathbb{R}^d$. Such a dynamics $\boldsymbol{b}$ gives a zero-variance estimator iff ρ_0 and ρ_1 have the same marginal distribution on the orthogonal space of $\big\{cv:\ c\in\mathbb{R}\big\}$.*

F.5 Proof of Proposition F.1

Lemma F.5. *Fix the potential U_0 and a dynamics $\boldsymbol{b} \in \mathfrak{B}$. If $\frac{\int_{-\infty}^{\infty} e^{-U_k\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t} = C_k$ is a constant function for $k \in \{2,3\}$, then any mixture of U_2 and U_3, given below, also ensures that $\frac{\int_{-\infty}^{\infty} e^{-U\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t}$ is a constant function, as long as U is a valid potential function:*

$$U = -\log\big(\omega_2 e^{-U_2} + \omega_3 e^{-U_3}\big), \qquad \omega_2, \omega_3 \in \mathbb{R}.$$

Proof. We can easily observe that

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-U\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t &= \int_{-\infty}^{\infty} \omega_2 e^{-U_2(\boldsymbol{X}_t(x))}\mathcal{J}_t(x) + \omega_3 e^{-U_3(\boldsymbol{X}_t(x))}\mathcal{J}_t(x)\,\mathrm{d}t \\ &= (C_2\omega_2 + C_3\omega_3)\int_{-\infty}^{\infty} e^{-U_0(\boldsymbol{X}_t(x))}\mathcal{J}_t(x)\,\mathrm{d}t.\end{aligned}$$

□

Lemma F.6. *For the distribution $\rho_1 \propto e^{-U^\tau + C}$, the partition function $\mathcal{Z}_1$ can be computed with zero-variance.*

Proof. We only need to verify the case $C = 0$. Then

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-U^\tau\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t &= \int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_{t-\tau}(x)\right)} \mathcal{J}_{-\tau}\big(\boldsymbol{X}_t(x)\big)\mathcal{J}_t(x)\,\mathrm{d}t \\ &\overset{(25)}{=} \int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_{t-\tau}(x)\right)} \frac{\mathcal{J}_{t-\tau}(x)}{\mathcal{J}_t(x)}\mathcal{J}_t(x)\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_{t-\tau}(x)\right)} \mathcal{J}_{t-\tau}(x)\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t.\end{aligned}$$

This means $\boldsymbol{b}$ is a zero-variance dynamics for the distribution ρ_1. □

Proof of Proposition F.1. Combine the above two lemmas. □

F.6 Proof of Proposition F.2

Consider a 2D example with $U_0(x) = \frac{|x|^2}{2} + \ln(2\pi)$ and $\boldsymbol{b} = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$. Then $\mathcal{J}_t(x) = 1$, and $U^\tau(x) = \frac{(x_1-\tau)^2 + x_2^2}{2} + \ln(2\pi)$ where $x = \left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right]$. Let us consider

$$e^{-U_1(x)} := e^{-U^1(x)} + \epsilon x_1 e^{-x_1^4 - x_2^4},$$

where $\epsilon := \frac{1}{2\pi}\big(2\sup_{y\in\mathbb{R}} |y| e^{-y^4 + (y-1)^2/2} \sup_{y\in\mathbb{R}} e^{-y^4 + y^2/2}\big)^{-1} > 0$. It could be straightforwardly verified that U_1 is a well-defined potential and

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-U_1\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t &= \int_{-\infty}^{\infty} e^{-U^1\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t + \epsilon\int_{-\infty}^{\infty} (x_1+t) e^{-(x_1+t)^4 - x_2^4}\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} e^{-U^1\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} e^{-U_0\left(\boldsymbol{X}_t(x)\right)} \mathcal{J}_t(x)\,\mathrm{d}t.\end{aligned}$$

The last equality holds by Lemma F.6. Therefore, $\boldsymbol{b}$ is a zero-variance dynamics for ρ_0 and ρ_1. However, $U_1 \notin \mathfrak{F}$ because $\int_{-\infty}^{\infty} f(\tau) e^{-U^\tau(x)}\,\mathrm{d}\tau$ must be a separable function, whereas U_1 is not.

G Proof of Proposition 3.2 and more discussions

G.1 Proof of Proposition 3.2

We proceed in three steps:

Step 1: Let us first assume the existence of $\varkappa \in C^1(D, \mathbb{R})$ satisfying (14), where D is an open subset of Ω and $\Omega \backslash D$ has measure zero. We shall verify that $\boldsymbol{T}\#\rho_0 = \rho_1$ almost everywhere.

From (14), let us replace x by $\boldsymbol{X}_\epsilon(x)$,

$$0 \overset{(25)}{=} \frac{1}{\mathcal{J}_\epsilon(x)} \Big(\int_{-\infty}^{0} \rho_0(\boldsymbol{X}_{s+\epsilon}(x)) \mathcal{J}_{s+\epsilon}(x) \, \mathrm{d}s - \int_{-\infty}^{\varkappa(\boldsymbol{X}_\epsilon(x))} \rho_1(\boldsymbol{X}_{s+\epsilon}(x)) \mathcal{J}_{s+\epsilon}(x) \, \mathrm{d}s \Big).$$

By straightforward simplification,

$$0 = \int_{-\infty}^{0+\epsilon} \rho_0(\boldsymbol{X}_s(x)) \mathcal{J}_s(x) \, \mathrm{d}s - \int_{-\infty}^{\varkappa(\boldsymbol{X}_\epsilon(x))+\epsilon} \rho_1(\boldsymbol{X}_s(x)) \mathcal{J}_s(x) \, \mathrm{d}s.$$

By taking the derivative with respect to ϵ at $\epsilon = 0$,

$$\begin{aligned} \rho_0(x) &= \rho_1(\boldsymbol{X}_{\varkappa(x)}(x)) \mathcal{J}_{\varkappa(x)}(x) \big(1 + \langle \nabla \varkappa(x), \boldsymbol{b}(x) \rangle\big) \\ &= \rho_1(\boldsymbol{T}(x)) \mathcal{J}_{\varkappa(x)}(x) \big(1 + \langle \nabla \varkappa(x), \boldsymbol{b}(x) \rangle\big). \end{aligned}$$

To show that $\boldsymbol{T}\#\rho_0 = \rho_1$, we need to verify that $\rho_0(x) = \rho_1(\boldsymbol{T}(x)) \mathcal{J}_{\boldsymbol{T}}(x)$ where $\mathcal{J}_{\boldsymbol{T}}(x) = \big| \det \big(\nabla_x \boldsymbol{X}_{\varkappa(x)}(x) \big) \big|$. Therefore, it remains to prove that

$$\det \big(\nabla_x \boldsymbol{X}_{\varkappa(x)}(x) \big) = \mathcal{J}_{\varkappa(x)}(x) \big(1 + \big\langle \nabla \varkappa(x), \boldsymbol{b}(x) \big\rangle\big). \tag{59}$$

By the direct computation,

$$\nabla_x \boldsymbol{X}_{\varkappa(x)}(x) \overset{(45)}{=} \boldsymbol{C}_{\varkappa(x),0}(x) + \boldsymbol{b}\big(\boldsymbol{X}_{\varkappa(x)}(x)\big) \big(\nabla \varkappa(x)\big)^T.$$

By the matrix determinant lemma,

$$\begin{aligned} \det \big(\nabla_x \boldsymbol{X}_{\varkappa(x)}(x) \big) &= \Big(1 + \Big\langle \nabla \varkappa(x), \big(\boldsymbol{C}_{\varkappa(x),0}(x)\big)^{-1} \boldsymbol{b}\big(\boldsymbol{X}_{\varkappa(x)}(x)\big) \Big\rangle \Big) \det \big(\boldsymbol{C}_{\varkappa(x),0}(x) \big) \\ &= \Big(1 + \Big\langle \nabla \varkappa(x), \boldsymbol{b}(x) \Big\rangle \Big) \mathcal{J}_{\varkappa(x)}(x), \end{aligned}$$

where we used (40) and Lemma D.6 to get the second line. The last equation verifies (59) and therefore, $\boldsymbol{T}\#\rho_0 = \rho_1$ almost everywhere.

Step 2: We shall explain D and establish the existence of $\varkappa \in C^1(D, \mathbb{R})$.

Suppose we denote the local minima of V as $\mathscr{X}_1, \mathscr{X}_2, \cdots, \mathscr{X}_a$, and local maxima of V as $\mathscr{Y}_1, \mathscr{Y}_2, \cdots, \mathscr{Y}_b$, where $a, b \in \mathbb{N}$. Then in the proof of Proposition 3.1, we have already mentioned that

$$D := \Big\{ x \in \Omega \,\Big|\, \lim_{t \to -\infty} \boldsymbol{X}_t(x) = \mathscr{X}_i, \ \lim_{t \to \infty} \boldsymbol{X}_t(x) = \mathscr{Y}_j, \text{ for some } i, j \Big\}$$

is an open subset of $\mho(\boldsymbol{b})$ and $\Omega \backslash D$ has measure zero, due to the assumption that V is a Morse function.

Consider the following function $L(x, t) : D \times \mathbb{R} \to \mathbb{R}$, defined as

$$L(x, t) := \int_{-\infty}^{0} \rho_0(\boldsymbol{X}_s(x)) \mathcal{J}_s(x) \, \mathrm{d}s - \int_{-\infty}^{t} \rho_1(\boldsymbol{X}_s(x)) \mathcal{J}_s(x) \, \mathrm{d}s.$$

We can observe that

- $\lim_{t \to -\infty} L(x, t) = \int_{-\infty}^{0} \rho_0(\boldsymbol{X}_s(x)) \mathcal{J}_s(x) \, \mathrm{d}s > 0.$

•

$$\lim_{t\to\infty} L(x,t) = \int_{-\infty}^{0} \rho_0(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\,\mathrm{d}s - \int_{-\infty}^{\infty} \rho_1(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\,\mathrm{d}s$$
$$= -\int_{0}^{\infty} \rho_0(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\,\mathrm{d}s < 0$$

where the second equality comes from the fact that $\boldsymbol{b} = \nabla V$ is a zero-variance dynamics; see e.g., (12).

- With fixed x, the function $t \to L(x,t)$ is continuously differentiable and is strictly monotonically decreasing.

These imply that for each $x \in D$, there exists a unique $\varkappa(x)$ such that $L\big(x, \varkappa(x)\big) = 0$ by the intermediate value theorem. Therefore, $\varkappa$ is well-defined via (14).

Next, we need to prove that such a function $\varkappa \in C^1(D, \mathbb{R})$, which can be immediately obtained by the implicit function theorem [35], provided that we can prove $L \in C^1(D \times \mathbb{R}, \mathbb{R})$. Due to the smoothness assumption on ρ_0 and ρ_1, it is clear that $\partial_t L(x,t) = -\rho_1(\boldsymbol{X}_t(x))\mathcal{J}_t(x) < 0$ exists and is continuous. Therefore, the task becomes to prove that $\nabla_x L(x,t)$ exists and is continuous. Since it is clear that $\int_0^t \rho_1(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\,\mathrm{d}s$ is continuously differentiable with respect to x, it is then sufficient to prove that

$$G_k(x) := \int_{-\infty}^{0} \rho_k(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\,\mathrm{d}s$$

is continuously differentiable for $k \in \{0,1\}$.

Step 3: Prove that $G_k \in C^1(D, \mathbb{R})$.

In Appendix E.1, we have proved that G_k is continuous; see Lemma E.4 in particular. Next, we first verify that G_k is differentiable and then verify that ∇G_k is also continuous.

Part 1: G_k is differentiable.

We want to verify that G_k is differentiable and in particular

$$\nabla_x G_k(x) = \int_{-\infty}^{0} \nabla_x\big(\rho_k(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\big)\,\mathrm{d}s. \tag{60}$$

Let us consider an arbitrary $x \in D$. Without loss of generality, suppose its limit in the backward branch is $\mathscr{X}_1 = \lim_{t\to-\infty} \boldsymbol{X}_t(x)$. Let us focus on a local neighborhood $B_\delta(x)$ such that $\lim_{t\to-\infty} \boldsymbol{X}_t(z) = \mathscr{X}_1$ for all $z \in B_\delta(x)$. Such a small δ exists because $\mathscr{X}_1$ is a strict local minimum and $\boldsymbol{b} = \nabla V$ is smooth from the assumption that V is a Morse function. In order to verify that G_k is differentiable (i.e., (60)), by the Leibniz rule (see e.g., [19, Theorem 6.28]), it is sufficient to prove that there exists an integrable function $Q : (-\infty, 0] \to \mathbb{R}$ such that

$$\big|\nabla_z\big(\rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)\big)\big| \le Q(s), \qquad \forall s \in (-\infty, 0],\ \forall z \in B_\delta(x). \tag{61}$$

By direct computation,

$$\begin{aligned}
&\nabla_z\big(\rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)\big)\\
=&\rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)\Big(-(\nabla\boldsymbol{X}_s(z))^T\nabla U_k(\boldsymbol{X}_s(z)) + \int_0^s \big(\nabla\boldsymbol{X}_r(z)\big)^T\nabla(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\boldsymbol{X}_r(z))\,\mathrm{d}r\Big)\\
\overset{(45)}{=}&\rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)\Big(-(\boldsymbol{C}_{s,0}(z))^T\nabla U_k(\boldsymbol{X}_s(z)) + \int_0^s \big(\boldsymbol{C}_{r,0}(z)\big)^T\nabla(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\boldsymbol{X}_r(z))\,\mathrm{d}r\Big).
\end{aligned} \tag{62}$$

Since the domain Ω is assumed to be a torus, we know ρ_k, ∇U_k and $\nabla(\boldsymbol{\nabla}\cdot\boldsymbol{b})$ are uniformly bounded on the domain Ω. Therefore,

$$\big|\nabla_z\big(\rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)\big)\big| \lesssim \mathcal{J}_s(z)\Big(\|\boldsymbol{C}_{s,0}(z)\| + \int_s^0 \|\boldsymbol{C}_{r,0}(z)\|\,\mathrm{d}r\Big). \tag{63}$$

Recall from (5) and (40) that

$$\partial_s \mathcal{J}_s(z) = (\boldsymbol{\nabla}\cdot\boldsymbol{b})(\boldsymbol{X}_s(z))\mathcal{J}_s(z), \qquad \mathcal{J}_0(z) = 1;$$
$$\partial_s \boldsymbol{C}_{s,0}(z) = \nabla\boldsymbol{b}(\boldsymbol{X}_s(z))\boldsymbol{C}_{s,0}(z), \qquad \boldsymbol{C}_{0,0}(z) = \mathbb{I}_d.$$

Recall that $\boldsymbol{X}_s(z) \to \mathscr{X}_1$ as $s \to -\infty$; the value $\boldsymbol{\nabla}\cdot\boldsymbol{b}(\mathscr{X}_1)$ and the hessian matrix $\nabla\boldsymbol{b}(\mathscr{X}_1)$ are both strictly positive, which imply that $\mathcal{J}_s(x)$ and $\|\boldsymbol{C}_{s,0}(z)\|$ are both decaying as $s \to -\infty$.

Let us denote $\upsilon > 0$ as the smallest eigenvalue of $\nabla\boldsymbol{b}(\mathscr{X}_1)$. For a given ϵ with $0 < \epsilon < \min\big\{\boldsymbol{\nabla}\cdot\boldsymbol{b}(\mathscr{X}_1), \upsilon\big\}$, let us define

$$\mathcal{E} = \big\{z \in \Omega : \boldsymbol{\nabla}\cdot\boldsymbol{b}(z) > \nabla\cdot\boldsymbol{b}(\mathscr{X}_1) - \epsilon,\ \nabla\boldsymbol{b}(z) > (\upsilon-\epsilon)\mathbb{I}_d\big\},$$

which is an open neighborhood of $\mathscr{X}_1$. We can find a subset of $\mathcal{E}$, denoted as $\mathcal{E}_{\text{trap}}$, such that $\mathcal{E}_{\text{trap}}$ is a trapping region of the dynamics $\boldsymbol{b}$ for the backward branch. Thus we can find a negative time τ (which possibly depends on ϵ) such that $\boldsymbol{X}_s(x) \in \mathcal{E}_{\text{trap}}$ for all $s \le \tau$. If we choose a δ small enough, then we can even ensure that

$$\boldsymbol{X}_s(z) \in \mathcal{E}_{\text{trap}}, \qquad \forall s \le \tau, \forall z \in B_\delta(x), \tag{64}$$

due to the smoothness of $\boldsymbol{b}$ and the construction that $\mathcal{E}_1$ is a trapping region. Therefore, when $s \le \tau$, $\mathcal{J}_s(z)$ decays to zero exponentially fast as $s \to -\infty$ with a rate at least $\boldsymbol{\nabla}\cdot\boldsymbol{b}(\mathscr{X}_1) - \epsilon$; similarly, the matrix norm $\|\boldsymbol{C}_{s,0}(z)\|$ also decays to zero exponentially fast as $s \to -\infty$ with a rate at least $\upsilon - \epsilon$. Therefore, on the region $(-\infty, 0] \times B_\delta(x)$, we can readily obtain

$$\mathcal{J}_s(z) \lesssim e^{\left(\boldsymbol{\nabla}\cdot\boldsymbol{b}(\mathscr{X}_1)-\epsilon\right)s}, \qquad \|\boldsymbol{C}_{s,0}(z)\| \lesssim e^{\left(\upsilon-\epsilon\right)s}. \tag{65}$$

By plugging the above estimates into (63), we know that

$$\big|\nabla_z\big(\rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)\big)\big| \lesssim \mathcal{J}_s(z) \lesssim e^{\left(\boldsymbol{\nabla}\cdot\boldsymbol{b}(\mathscr{X}_1)-\epsilon\right)s}. \tag{66}$$

The function $e^{\left(\boldsymbol{\nabla}\cdot\boldsymbol{b}(\mathscr{X}_1)-\epsilon\right)s}$ is integrable on $(-\infty, 0]$ and this serves as the function Q needed in (61) with some multiplicative constant.

Part 2: ∇G_k is continuous.

Let us denote $H(z,s) := \nabla_z\big(\rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)\big)$. From (62), we can observe that $\nabla_z H(z,s)$ is continuous with respect to z with

$$\begin{aligned}\nabla_z H(z,s) =& \frac{H(z,s)\big(H(z,s)\big)^T}{\rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)} + \\ & \rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)\nabla_z\Big(-(\boldsymbol{C}_{s,0}(z))^T\nabla U_k(\boldsymbol{X}_s(z))\Big) + \\ & \rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)\int_0^s \nabla_z\Big(\big(\boldsymbol{C}_{r,0}(z)\big)^T\nabla(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\boldsymbol{X}_r(z))\Big)\,\mathrm{d}r.\end{aligned}$$

Lemma G.1. *For a given smooth vector field $W \in C^\infty(\Omega, \mathbb{R}^d)$, there exists a constant C such that for any $z \in B_\delta(x)$ and $s \in (-\infty, 0]$, we have*

$$\big\|\nabla_z\big(\boldsymbol{C}_{s,0}(z)^T W(\boldsymbol{X}_s(z))\big)\big\| \le Ce^{(\upsilon-\epsilon)s}.$$

By this lemma and the estimates in (66),

$$\begin{aligned}\|\nabla_z H(z,s)\| &\lesssim \Big\|\frac{H(z,s)\big(H(z,s)\big)^T}{\rho_k(\boldsymbol{X}_s(z))\mathcal{J}_s(z)}\Big\| + \mathcal{J}_s(z)e^{(\upsilon-\epsilon)s} + \mathcal{J}_s(z)\int_s^0 e^{(\upsilon-\epsilon)r}\,\mathrm{d}r \\ &\lesssim \mathcal{J}_s(z) \lesssim e^{\left(\boldsymbol{\nabla}\cdot\boldsymbol{b}(\mathscr{X}_1)-\epsilon\right)s},\end{aligned}$$

which readily leads into the continuity of ∇G_k based on (60).

Part 3: Proof of Lemma G.1

By direct computation

$$
\begin{aligned}
&\nabla_{z_j}\Big(\boldsymbol{C}_{s,0}(z)^T W(\boldsymbol{X}_s(z))\Big)_i \\
&= \sum_l \nabla_{z_j}\Big(\big(\boldsymbol{C}_{s,0}(z)\big)_{l,i} W_l(\boldsymbol{X}_s(z))\Big) \\
&\overset{(45)}{=} \sum_l \Big(\nabla_{z_j}\big(\boldsymbol{C}_{s,0}(z)\big)_{l,i}\Big) W_l(\boldsymbol{X}_s(z)) + \sum_{l,m} \big(\boldsymbol{C}_{s,0}(z)\big)_{l,i}\big(\nabla W(\boldsymbol{X}_s(z))\big)_{l,m}\big(\boldsymbol{C}_{s,0}(z)\big)_{m,j}.
\end{aligned}
$$

Since W is assumed to be smooth on the torus Ω, we know W and ∇W are uniformly bounded. Besides, previously, we know that $\|\boldsymbol{C}_{s,0}(z)\| \lesssim e^{\big(\upsilon-\epsilon\big)s}$. Therefore, we only need to prove that for any $1 \le \ell \le d$,

$$
\|\partial_{z_\ell}\boldsymbol{C}_{s,0}(z)\| \lesssim e^{(\upsilon-\epsilon)s}. \tag{67}
$$

From (41), we have

$$
\begin{aligned}
&\partial_s\big(\partial_{z_\ell}\boldsymbol{C}_{s,0}(z)\big) = \nabla \boldsymbol{b}(\boldsymbol{X}_s(z))\big(\partial_{z_\ell}\boldsymbol{C}_{s,0}(z)\big) + \boldsymbol{S}(s,z), \qquad \big(\partial_{z_\ell}\boldsymbol{C}_{0,0}(z)\big) = \boldsymbol{0}_{d\times d}, \\
&\big(\boldsymbol{S}(s,z)\big)_{i,j} = \sum_{n,m}(\partial_{z_i,z_n,z_m}V)(\boldsymbol{X}_s(z))\big(\boldsymbol{C}_{s,0}(z)\big)_{m,\ell}\big(\boldsymbol{C}_{s,0}(z)\big)_{n,j}.
\end{aligned} \tag{68}
$$

Since $\|\boldsymbol{C}_{s,0}(z)\| \lesssim e^{(\upsilon-\epsilon)s}$ from (65), the source term $\|\boldsymbol{S}(s,z)\|$ also decays exponentially fast with rate $2(\upsilon-\epsilon)$ as $s \to -\infty$, namely,

$$
\|\boldsymbol{S}(s,z)\| \lesssim e^{2(\nu-\epsilon)s}, \qquad \forall s \in (-\infty, 0], z \in B_\delta(x). \tag{69}
$$

By rewriting (68) in the integral form and by (40), we have

$$
\big(\partial_{z_\ell}\boldsymbol{C}_{s,0}(z)\big) = -\int_s^0 \boldsymbol{C}_{s,0}(z)\big(\boldsymbol{C}_{r,0}(z)\big)^{-1}\boldsymbol{S}(r,z)\,\mathrm{d}r = -\int_s^0 \boldsymbol{C}_{s,r}(z)\boldsymbol{S}(r,z)\,\mathrm{d}r.
$$

To prove Lemma G.1, we only need to consider the case $s \ll 0$. Suppose we consider $s \le \tau$ only; recall the role of τ in (64). Then we could obtain that when $s \le r \le \tau$

$$
\|\boldsymbol{C}_{s,r}(z)\| = \left\|\exp_{\mathcal{T}_\to}\Big(-\int_s^r \nabla \boldsymbol{b}(\boldsymbol{X}_u(z))\,\mathrm{d}u\Big)\right\| \overset{(64)}{\le} e^{-(\nu-\epsilon)(r-s)}, \tag{70}
$$

where $\exp_{\mathcal{T}_\to}$ is the antichronological time-ordered operator exponential. We can separate the above integral form using τ and obtain the following estimates: when $s \le \tau$,

$$
\begin{aligned}
\|\partial_{z_\ell}\boldsymbol{C}_{s,0}(z)\| &\le \left\|\int_s^\tau \boldsymbol{C}_{s,r}(z)\boldsymbol{S}(r,z)\,\mathrm{d}r\right\| + \left\|\int_\tau^0 \boldsymbol{C}_{s,r}(z)\boldsymbol{S}(r,z)\,\mathrm{d}r\right\| \\
&\le \int_s^\tau \|\boldsymbol{C}_{s,r}(z)\boldsymbol{S}(r,z)\|\,\mathrm{d}r + \int_\tau^0 \left\|\boldsymbol{C}_{s,0}(z)\big(\boldsymbol{C}_{r,0}(z)\big)^{-1}\boldsymbol{S}(r,z)\right\|\mathrm{d}r \\
&\le \int_s^\tau \underbrace{\|\boldsymbol{C}_{s,r}(z)\|}_{\text{use (70)}}\underbrace{\|\boldsymbol{S}(r,z)\|}_{\text{use (69)}}\,\mathrm{d}r + \underbrace{\|\boldsymbol{C}_{s,0}(z)\|}_{\text{use (65)}}\int_\tau^0 \underbrace{\left\|\big(\boldsymbol{C}_{r,0}(z)\big)^{-1}\right\|}_{=\mathcal{O}(1)}\cdot\underbrace{\|\boldsymbol{S}(r,z)\|}_{\text{use (69)}}\,\mathrm{d}r \\
&\lesssim \int_s^\tau e^{-(\upsilon-\epsilon)(r-s)}e^{2(\upsilon-\epsilon)r}\mathrm{d}r + e^{(\upsilon-\epsilon)s}\underbrace{\int_\tau^0 e^{2(\upsilon-\epsilon)r}\mathrm{d}r}_{=\mathcal{O}(1)} \\
&\lesssim e^{(\upsilon-\epsilon)s} + e^{(\upsilon-\epsilon)s} \lesssim e^{(\upsilon-\epsilon)s}.
\end{aligned}
$$

This verifies (67) for any $z \in B_\delta(x)$ and $s \le \tau$. When $s \in [-\tau, 0]$, we can simply choose the prefactor large enough so that (67) holds, as τ is a finite value. Hence, Lemma G.1 is verified.

G.2 Examples

We elaborate on Proposition 3.2 by concrete examples. For these examples, we estimate $\varkappa$ from (14) either analytically or numerically, and then we validate Proposition 3.2 by comparing ρ_1 and the empirical distribution of $\boldsymbol{X}_{\varkappa(x)}(x)$ where $x \sim \rho_0$.

G.2.1 Gaussian examples in 1D

For the case $U_0(x) = \frac{|x|^2}{2} + \frac{1}{2}\ln(2\pi)$ and $U_1(x) = \frac{|x-\omega|^2}{2\sigma^2}$ with $\sigma < 1$, from Table 1, we already know that a zero-variance dynamics is $\boldsymbol{b}(x) = x - \frac{\omega}{1-\sigma}$. Then $\boldsymbol{X}_t(x) = e^t x - (e^t - 1)\frac{\omega}{1-\sigma}$ for any $t, x \in \mathbb{R}$ and $\mathcal{J}_t(x) = e^t$ for any $t \in \mathbb{R}$. By direct computation

$$\frac{\int_{-\infty}^{\theta} \rho_1(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\mathrm{d}s}{\int_{-\infty}^{0} \rho_0(\boldsymbol{X}_s(x))\mathcal{J}_s(x)\mathrm{d}s} = -\frac{\mathrm{erf}(\frac{\omega}{\sqrt{2}(1-\sigma)}) + \mathrm{erf}\left(\frac{-e^{\theta}(\omega+x(\sigma-1)+\omega\sigma)}{\sqrt{2}\sigma(\sigma-1)}\right)}{\mathrm{erf}(\frac{x}{\sqrt{2}}) + \mathrm{erf}\left(\frac{\omega}{\sqrt{2}(\sigma-1)}\right)}.$$

By solving $\varkappa$ in (14) using the last equation, we have

$$\varkappa(x) = \log(\sigma), \qquad \forall x \neq \frac{\omega}{1-\sigma},$$

which is independent of the state x.

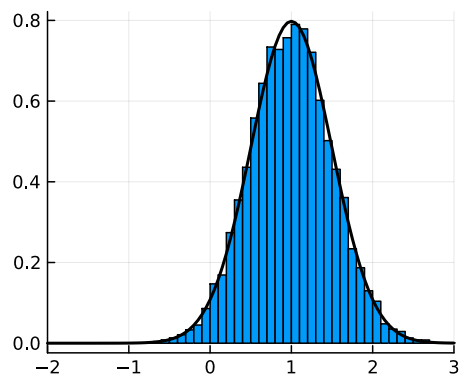


Figure 4: This figure shows the histogram of sample points of $\boldsymbol{X}_{\varkappa(x)}(x)$ (blue) and the distribution ρ_1 (black) for the 1D Gaussian example in Appendix G.2.1 with $\omega = 1$ and $\sigma = 0.5$.

G.2.2 Three-mode mixture on a 2D torus

We consider the model (54) on the torus $\Omega = [0,1]^2$ and recall that the zero-variance dynamics has been shown in Figure 1. Then the contour plot of $\varkappa$ is visualized in Figure 5. Moreover, in the same figure, the empirical distributions of $\boldsymbol{X}_{\varkappa(x)}(x)$ and contour plots of ρ_1 are provided, which numerically verifies Proposition 3.2.

G.2.3 An example on $(0,1)^2$ with Neumann boundary condition

We consider the model

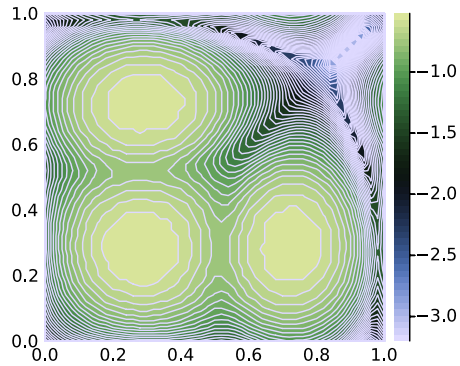
$$\begin{gathered} V(x) = \gamma\cos(2\pi x_1)\cos(2\pi x_2), \qquad \rho_0(x) = 1, \qquad \rho_1(x) = \rho_0(x) + \Delta V(x), \\ x = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right], \qquad \gamma = \frac{0.45}{4\pi^2}, \end{gathered} \tag{71}$$

on the domain $\Omega = (0,1)^2$ and the potential V automatically solves the Poisson's equation with Neumann boundary condition (see (55)) by the above construction. Apart from verifying that $\boldsymbol{b} = \nabla V$ is a zero-variance dynamics numerically, we can also observe that Proposition 3.2 holds in this case; see Figure 6.

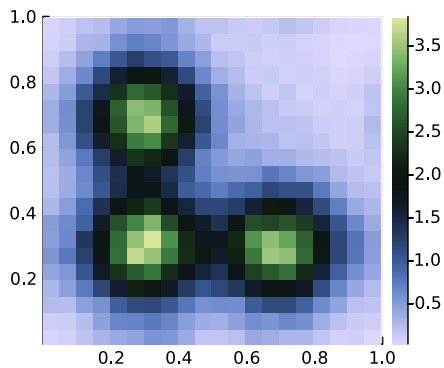
H Proof of Proposition 4.1

Below is the full version of Proposition 4.1.

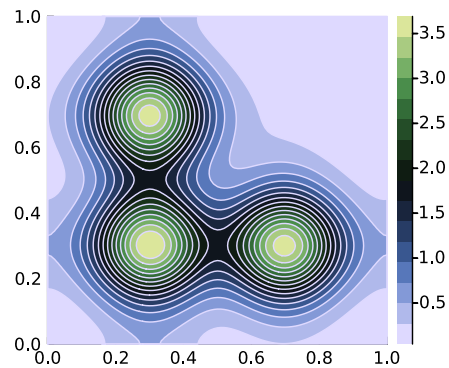
Proposition H.1 (Local minimum)**.** *Assume that*



(a) $\varkappa$



(b) histogram of samples of $\boldsymbol{X}_{\varkappa(x)}(x)$ where $x \sim \rho_0$



(c) ρ_1

Figure 5: The figure shows the time map $\varkappa$ defined via (14) and it also numerically verifies that ρ_1 is the distribution of $\boldsymbol{X}_{\varkappa(x)}(x)$ with $x \sim \rho_0$, for the model (54) on the torus $\Omega = [0,1]^2$ (with periodic boundary condition).

(i) (Local minimum). $\boldsymbol{b} \in \mathfrak{B}_\infty$ is a (non-trivial) local minimum of Var, namely, Var$(\boldsymbol{b}) <$ Var$^{(\max)}$ and if there is a perturbation $\delta\boldsymbol{b} \in C_c^\infty(\Omega, \mathbb{R}^d)$ such that $\boldsymbol{b} + \epsilon\delta\boldsymbol{b} \in \mathfrak{B}_\infty$ for sufficiently small ϵ, then Var$(\boldsymbol{b} + \epsilon\delta\boldsymbol{b}) \geq$ Var$(\boldsymbol{b})$.

(ii) (Continuity assumption). The functional derivative $\frac{\delta \mathrm{Var}(\boldsymbol{b})}{\delta \boldsymbol{b}}$ is continuous and $\frac{\delta \mathrm{Var}(\boldsymbol{b})}{\delta \boldsymbol{b}} = \mathbf{0}_d$ on Ω. In particular, $\nabla\mathcal{A}$ exists and is continuous on Ω.

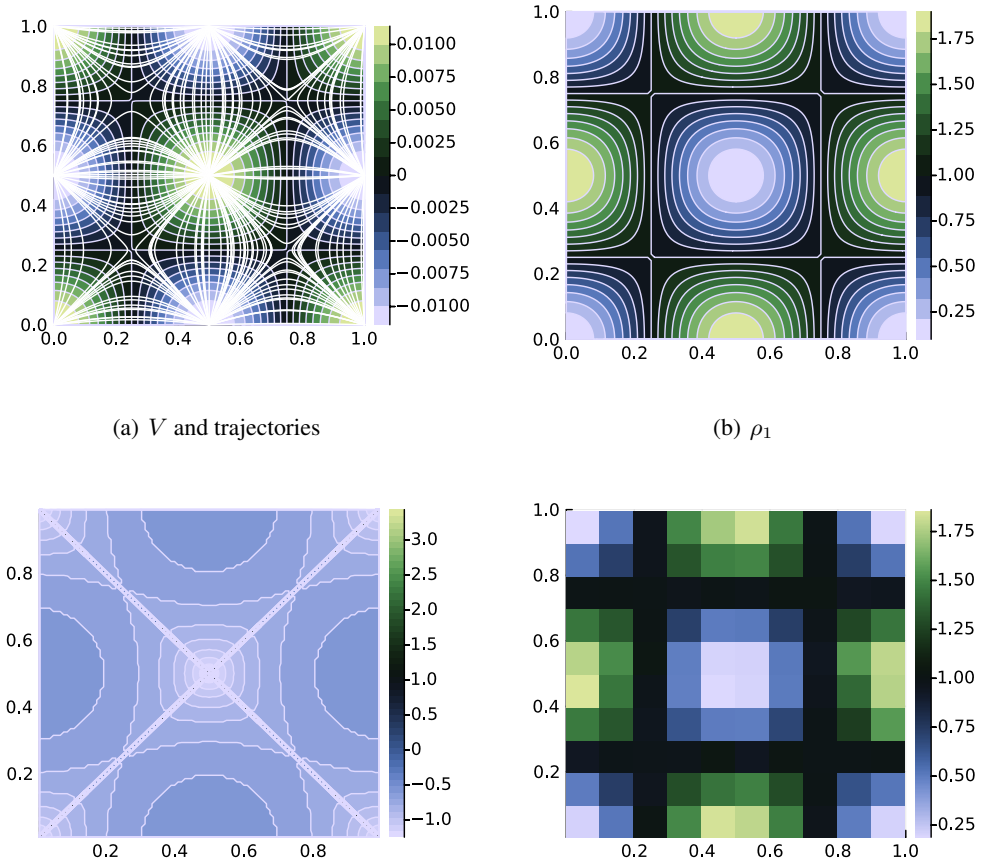
(iii) (Technical assumptions). $\mho(\boldsymbol{b}) = \Omega$ (see Definition A.4) and the set of $\boldsymbol{b}$-unstable points (see Definition A.8) has Lebesgue measure zero. Moreover, the set $\left\{x \in \Omega : \nabla(U_1 - U_0)(x) = \mathbf{0}_d\right\}$ has Lebesgue measure zero.

Then $\boldsymbol{b}$ is a zero-variance dynamics for the infinite-time NEIS scheme, i.e., Var$(\boldsymbol{b}) = 0$.

Recall the formula of $\frac{\delta \mathrm{Var}(\boldsymbol{b})}{\delta \boldsymbol{b}}$ from (38):

$$\frac{\delta \mathrm{Var}(\boldsymbol{b})}{\delta \boldsymbol{b}}(x) = \frac{2\nabla\mathcal{A}(x)}{\mathcal{B}(x)} \left(\int_0^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^0 \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_{-\infty}^0 \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^\infty \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t \right).$$

Here is a sketch of the main idea behind the proof. If $\nabla\mathcal{A} = \mathbf{0}_d$ on the domain Ω, then $\mathcal{A}$ is a constant function. Hence, $\boldsymbol{b}$ provides a zero-variance estimator for the infinite-time NEIS scheme and it must be a global minimum of the functional Var as well. If the other term $\int_0^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^0 \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t -$



(a) V and trajectories

(b) ρ_1

(c) $\varkappa$ in symlog scale

(d) histogram of samples of $\boldsymbol{X}_{\varkappa(x)}(x)$ where $x \sim \rho_0$

Figure 6: This figure visualizes the model (71) on the domain $\Omega = (0,1)^2$ (with Neumann boundary condition). In (a), we show the potential V and sample trajectories under the dynamics $\boldsymbol{b} = \nabla V$; in (b), we show the distribution ρ_1. In (c), we present the time map $\varphi \circ \varkappa$ where $\varkappa$ is defined via (14) and the rescaling function $\varphi(z) := \operatorname{sign}(z)\log(1+|z|)$ is the symlog function. In (d), we show a histogram of samples of $\boldsymbol{X}_{\varkappa(x)}(x)$ where $x \sim \rho_0$. The panel (d) resembles the panel (b), which numerically verifies that ρ_1 is the distribution of $\boldsymbol{X}_{\varkappa(x)}(x)$ with $x \sim \rho_0$.

$\int_{-\infty}^{0} \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^{\infty} \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t = 0$ locally on an open subset, then it could be shown that this is equivalent to $\langle \nabla(U_1 - U_0)(x), \boldsymbol{b}(x)\rangle = 0$ (see Lemma H.4 and Lemma H.5). Such a dynamics $\boldsymbol{b}$ should not be optimal, because $\boldsymbol{b}$ is perpendicular to the gradient of potential difference and such a $\boldsymbol{b}$ does not explore the local structure (cf. Proposition H.3). This intuition leads into the following idea: if $\boldsymbol{b}$ is a local minimum of Var, then we should only have $\nabla \mathcal{A} = \mathbf{0}_d$ almost everywhere on Ω; otherwise, we should be able to perturb $\boldsymbol{b}$ so that the dynamics $\boldsymbol{b}$ can better explore the landscapes of U_0 and U_1 and the variance can be further reduced; see Proposition H.6.

Remark H.2. As we work on the domain Ω only, we shall consider the topological space for the domain Ω instead of $\mathbb{R}^d$ from now on.

H.1 A characterization of the global maximum

Proposition H.3. *If the dynamics $\boldsymbol{b} \in \mathfrak{B}_\infty$, then*

$$\mathit{Var}(\boldsymbol{b}) \le \mathit{Var}^{(\max)}. \tag{72}$$

The equality can be achieved iff

$$\big\langle \boldsymbol{b}(x), \nabla(U_1 - U_0)(x)\big\rangle = 0, \qquad \forall x \in \Omega. \tag{73}$$

Proof. We only need to prove that $\mathcal{M}(\boldsymbol{b}) \le \mathbb{E}_{\rho_0}[e^{-2(U_1-U_0)}]$ and the equality is achieved iff (73) holds.

By Jensen's inequality,

$$\begin{aligned}\Big(\frac{\int_{-\infty}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}\Big)^2 &= \Big(\frac{\int_{-\infty}^{\infty} e^{-U_1\left(\boldsymbol{X}_t(x)\right)+U_0\left(\boldsymbol{X}_t(x)\right)}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}\Big)^2\\ &\le \frac{\int_{-\infty}^{\infty} e^{-2(U_1-U_0)\left(\boldsymbol{X}_t(x)\right)}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}.\end{aligned}$$

By taking the expectation $\mathbb{E}_{x\sim\rho_0}\big[\,\cdot\,\big]$ for both sides and by (10) for new potentials $\widetilde{U}_1 = 2U_1 - U_0$ and $\widetilde{U}_0 = U_0$, we immediately have the inequality:

$$\begin{aligned}\mathcal{M}(\boldsymbol{b}) &= \mathbb{E}_{x\sim\rho_0}\Big[\Big(\frac{\int_{-\infty}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}\Big)^2\Big]\\ &\le \mathbb{E}_{x\sim\rho_0}\Big[\frac{\int_{-\infty}^{\infty} e^{-2(U_1-U_0)\left(\boldsymbol{X}_t(x)\right)}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}\Big]\\ &\overset{(10)}{=} \frac{\int_{\Omega} e^{-2U_1+U_0}}{\int_{\Omega} e^{-U_0}} = \mathbb{E}_{x\sim\rho_0}\big[e^{-2(U_1-U_0)}\big].\end{aligned}$$

Next in order to achieve the equality, we need the equality to hold in the above Jensen's inequality:

$$t \to e^{-\left(U_1-U_0\right)\left(\boldsymbol{X}_t(x)\right)} \text{ is a constant function.}$$

By taking derivative with respect to t, we immediately know that

$$\big\langle \nabla(U_1 - U_0)(x), \boldsymbol{b}(x)\big\rangle = 0, \qquad \rho_0\text{-almost surely.}$$

By the continuity assumption on U_1, U_0 and $\boldsymbol{b}$, we obtain (73).

Conversely, when (73) holds,

$$\frac{\mathrm{d}}{\mathrm{d}t}(U_1 - U_0)\big(\boldsymbol{X}_t(x)\big) = \big\langle \nabla(U_1 - U_0)(\boldsymbol{X}_t(x)), \boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\big\rangle = 0.$$

Hence, $U_1\big(\boldsymbol{X}_t(x)\big) - U_0\big(\boldsymbol{X}_t(x)\big) = U_1(x) - U_0(x)$ for any $t \in \mathbb{R}$. Then

$$\frac{\int_{-\infty}^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t} = e^{-U_1(x)+U_0(x)}\frac{\int_{-\infty}^{\infty} e^{-U_0(\boldsymbol{X}_t(x))}\mathcal{J}_t(x)\,\mathrm{d}t}{\int_{-\infty}^{\infty} e^{-U_0(\boldsymbol{X}_t(x))}\mathcal{J}_t(x)\,\mathrm{d}t} = e^{-U_1(x)+U_0(x)}.$$

Hence, the equality in (72) holds under the condition (73). □

H.2 Some observations about the functional derivative

We need some simplified understanding of the condition $\int_0^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t = \int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_0^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t$ arising from $\frac{\delta\mathrm{Var}(\boldsymbol{b})}{\delta\boldsymbol{b}} = \mathbf{0}_d$, which are presented in the following two lemmas.

Lemma H.4. $\int_0^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_{-\infty}^{0}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t = \int_{-\infty}^{0}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t\int_0^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t$ *is equivalent to*

$$\frac{\int_0^{\infty}\mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_0^{\infty}\mathcal{F}_t^{(0)}(x)\,\mathrm{d}t} = \mathcal{A}(x). \tag{74}$$

Proof.

$$\int_0^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^0 \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t = \int_{-\infty}^0 \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^\infty \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t$$

$$\iff \frac{\int_{-\infty}^0 \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}{\int_0^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t} = \frac{\int_{-\infty}^0 \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_0^\infty \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t} \qquad \text{(add 1 to both sides)}$$

$$\iff \frac{\int_{-\infty}^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}{\int_0^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t} = \frac{\int_{-\infty}^\infty \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_0^\infty \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}$$

$$\iff \frac{\int_0^\infty \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_0^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t} = \mathcal{A}(x).$$

□

The following result provides a simplified characterization of the equality (74).

Lemma H.5.

(i) Suppose the condition in (74) *holds for any x in an open set D. Then*

$$\langle \nabla(U_1 - U_0)(x), \boldsymbol{b}(x)\rangle = 0, \qquad \forall x \in D. \tag{75}$$

(ii) Conversely, if (75) *holds for $D = \Omega$, then $\boldsymbol{b}$ is a global maximum of $\mathcal{M}$ (as well as Var).*

Proof. The part (ii) immediately follows from Proposition H.3. Next we prove the part (i). From previous results, the condition is that

$$\frac{\int_0^\infty \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_0^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t} = \mathcal{A}(x), \qquad \forall x \in D.$$

After replacing x by $\boldsymbol{X}_s(x) \in D$ in the above equation and by (25), we have

$$\mathcal{A}(x) \overset{(26)}{=} \mathcal{A}\big(\boldsymbol{X}_s(x)\big) = \frac{\int_0^\infty \mathcal{F}_t^{(1)}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}t}{\int_0^\infty \mathcal{F}_t^{(0)}\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}t} = \frac{\int_s^\infty \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t}{\int_s^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t}, \ \ \forall\, x \in D,\ s \in \big(\tau_D^-(x), \tau_D^+(x)\big),$$

where $\tau_D^+(x)$ and $\tau_D^-(x)$ defined in (23) are hitting times for the forward and backward branches to the boundary of D. Note that the right hand side of the last equation depends on s, whereas the left hand side does not. Let us take the derivative with respect to s and with straightforward simplifications, we obtain

$$\mathcal{F}_s^{(1)}(x)/\mathcal{F}_s^{(0)}(x) = \mathcal{A}(x), \qquad \forall\, x \in D,\ s \in \big(\tau_D^-(x), \tau_D^+(x)\big).$$

By (6), we know $\exp\Big(U_0\big(\boldsymbol{X}_s(x)\big) - U_1\big(\boldsymbol{X}_s(x)\big)\Big) = \mathcal{A}(x)$. Again, the left hand side depends on s whereas the right hand side does not. So we take the derivative with respect to s again and obtain $\big\langle \nabla(U_1 - U_0)\big(\boldsymbol{X}_s(x)\big), \boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\big\rangle = 0$ for any $x \in D$ and $s \in \big(\tau_D^-(x), \tau_D^+(x)\big)$. Then (75) follows immediately by choosing $s = 0$. □

H.3 A weaker version

Lemma H.5 leads into the following intuition: if there is a certain open subset D in which $\langle \nabla(U_1 - U_0), \boldsymbol{b}\rangle = 0$, then such a dynamics $\boldsymbol{b}$ should not be a local minimum of Var, because such a $\boldsymbol{b}$ cannot explored the landscape structure of U_1 on D. This intuition is more rigorously formulated in the following proposition.

Proposition H.6. *Consider a dynamics $\boldsymbol{b} \in \mathfrak{B}_\infty$. Suppose $D \subset \mho(\boldsymbol{b})$ is nonempty and open. Let $K := \Omega \backslash D$. Assume that*

(i) For any $x \in D$, we have $\langle \boldsymbol{b}(x), \nabla(U_1 - U_0)(x)\rangle = 0$.

(ii) For any $t \in \mathbb{R}$ and $x \in D$, we have $\boldsymbol{X}_t(x) \notin K^\circ$, i.e., trajectories from D are confined inside $\overline{D}$.

(iii) There exists a $\boldsymbol{b}$-stable point $x^\star \in D$ such that $\nabla(U_1 - U_0)(x^\star) \neq \mathbf{0}_d$.

Then such a dynamics $\boldsymbol{b}$ must not be a local minimum of the second moment $\mathcal{M}$ (as well as the variance).

Proof. We proceed in two steps.

Step (I): The first goal is to find a smooth function $\delta\boldsymbol{b} \in C_c^\infty(D, \mathbb{R}^d)$ and $\epsilon_0 > 0$ such that

$$\rho_0(E) > 0 \text{ where } E := \big\{x \in D \,\big|\, \big\langle \delta\boldsymbol{b}(x), \nabla(U_1 - U_0)(x)\big\rangle \neq 0\big\} \subset \operatorname{supp}(\delta\boldsymbol{b}); \tag{76a}$$

$$\boldsymbol{b} + \epsilon\delta\boldsymbol{b} \in \mathfrak{B}_\infty, \qquad \forall \epsilon \in (0, \epsilon_0); \tag{76b}$$

$$\operatorname{dist}\big(E, \partial D\big) > 0. \tag{76c}$$

By the assumption (iii) of this proposition and Proposition A.10, we know there is an open ball $B_\lambda(x^\star) \subset D$ such that for any $\delta\boldsymbol{b} \in C_c^\infty\big(B_\lambda(x^\star), \mathbb{R}^d\big)$, $\boldsymbol{b} + \epsilon\delta\boldsymbol{b} \in \mathfrak{B}_\infty$ for small enough ϵ and thus (76b) is satisfied. It is clear that we can easily choose λ small enough so that (76c) holds.

Next, the task is to find a smooth function $\delta\boldsymbol{b}$ supported on $B_\lambda(x^\star)$ such that (76a) holds. Since $\nabla(U_1 - U_0)(x^\star) \neq \mathbf{0}_d$ and U_0, U_1 are assumed to be smooth, we can choose λ small enough such that

$$|\nabla(U_1 - U_0)(x) - \nabla(U_1 - U_0)(x^\star)| \leq \frac{1}{4}|\nabla(U_1 - U_0)(x^\star)|, \qquad \forall x \in B_\lambda(x^\star).$$

It is well-known that

$$\varphi(x) := \begin{cases} e^{-\frac{1}{1-|x|^2}}, & \text{if } |x| < 1; \\ 0, & \text{if } |x| \geq 1, \end{cases}$$

is a smooth function compactly supported on $B_1(0)$. Then let us consider

$$\delta\boldsymbol{b}_1(x) := \varphi\big(\frac{x - x^\star}{\lambda_1}\big)\nabla(U_1 - U_0)(x), \qquad \text{where } \lambda_1 \in (0, \lambda).$$

It is clear that $\delta\boldsymbol{b}_1$ is compactly supported on $B_{\lambda_1}(x^\star) \subset B_\lambda(x^\star)$ and for any $x \in B_{\lambda_1}(x^\star)$, we have $\langle \delta\boldsymbol{b}_1(x), \nabla(U_1 - U_0)(x)\rangle > 0$ so that (76a) clearly holds. Next, we still need to further smooth out $\delta\boldsymbol{b}_1$ (see e.g., Appendix C of [12]) by introducing

$$\delta\boldsymbol{b}_2(x) := \int_{\mathbb{R}^d} \varphi_\varepsilon(x - y)\delta\boldsymbol{b}_1(y)\,\mathrm{d}y,$$

where $\varphi_\varepsilon(x) := \frac{1}{\varepsilon^d}\varphi(x/\varepsilon)$. Note that we can easily extend $\delta\boldsymbol{b}_1$ to $\mathbb{R}^d$ by letting $\delta\boldsymbol{b}_1 = \mathbf{0}_d$ outside of $B_{\lambda_1}(x^\star)$ so that $\delta\boldsymbol{b}_1$ can be well-defined on $\mathbb{R}^d$. By choosing $0 < \varepsilon < \lambda - \lambda_1$, we can ensure that the smooth function $\delta\boldsymbol{b}_2$ is compactly supported on $B_\lambda(x^\star)$. It is also not hard to show that (76a) still holds for $\delta\boldsymbol{b}_2$: for any $x \in B_{\lambda_1}(x^\star)$,

$$\begin{aligned}
\big\langle \delta\boldsymbol{b}_2(x), \nabla(U_1 - U_0)(x)\big\rangle &= \int_{\mathbb{R}^d} \varphi_\varepsilon(x - y)\big\langle \delta\boldsymbol{b}_1(y), \nabla(U_1 - U_0)(x)\big\rangle\,\mathrm{d}y \\
&= \int_{\mathbb{R}^d} \varphi_\varepsilon(x - y)\varphi\big(\frac{y - x^\star}{\lambda_1}\big)\big\langle \nabla(U_1 - U_0)(y), \nabla(U_1 - U_0)(x)\big\rangle\,\mathrm{d}y \\
&\geq \frac{7}{16}|\nabla(U_1 - U_0)(x^\star)|^2 \int_{\mathbb{R}^d} \varphi_\varepsilon(x - y)\varphi\big(\frac{y - x^\star}{\lambda_1}\big)\,\mathrm{d}y > 0.
\end{aligned}$$

In summary, $\delta\boldsymbol{b}_2$ constructed above satisfies all requirements.

Step (II): We prove that $\boldsymbol{b}$ is not a local minimum by showing that $\mathcal{M}(\boldsymbol{b} + \epsilon\delta\boldsymbol{b}) < \mathcal{M}(\boldsymbol{b})$ for any $\epsilon \in (0, \epsilon_0)$, where $\delta\boldsymbol{b}$ satisfies all conditions in the Step (I).

By the construction of $\delta\boldsymbol{b}$, we know $\boldsymbol{b}^\epsilon := \boldsymbol{b} + \epsilon\delta\boldsymbol{b}$ does not change the velocity field at ∂D, and therefore, $\boldsymbol{X}_t^\epsilon(x) \notin K^\circ$ for any $x \in D$ still holds for the dynamics $\boldsymbol{b}^\epsilon$ i.e., trajectories from D do not enter K°. As an immediately consequence, trajectories $t \to \boldsymbol{X}_t^\epsilon(x)$ with $x \in K^\circ$ will not enter D (as ODE trajectories are reversible). The slightly technical part is to consider trajectories $t \to \boldsymbol{X}_t^\epsilon(x)$ with $x \in \partial D \equiv \partial K$, where the trajectory $t \to \boldsymbol{X}_t^\epsilon(x)$ evolves under the dynamics $\boldsymbol{b}^\epsilon$. Let us consider two disjoint sets:

$$\widetilde{D} := \big\{x \in D \cup \partial D \equiv \Omega\backslash K^\circ \;\big|\; \boldsymbol{X}_t^\epsilon(x) \notin K^\circ,\ \forall t \in \mathbb{R}\big\} \supset D, \qquad \widetilde{K} := \Omega\backslash\widetilde{D} \subset K.$$

By such a construction, we can observe that for any trajectory $t \to \boldsymbol{X}_t^\epsilon(x)$ with $x \in \widetilde{K}$ (or $x \in \widetilde{D}$), it must remain inside $\widetilde{K}$ (or $\widetilde{D}$). In other words, the flows within $\widetilde{D}$ and $\widetilde{K}$ are completely separated from each other. Moreover, because $\delta\boldsymbol{b}$ is only supported on $E \subset D$ which is completely inside D by (76c), we know the above definitions of these two sets $\widetilde{D}$ and $\widetilde{K}$ are independent of $\epsilon \in [0, \epsilon_0)$.

Let us use $\mathcal{A}$ to denote the function defined in (24) corresponding to the dynamics $\boldsymbol{b}$ and use $\mathcal{A}_\epsilon$ to denote the one corresponding to the perturbed dynamics $\boldsymbol{b}^\epsilon$. Recall the assumption (i) that $\langle \boldsymbol{b}(x), \nabla(U_1 - U_0)(x)\rangle = 0$ for all $x \in D$. By the fact that $\partial D = \partial K$, and by the continuity of $\boldsymbol{b}$, ∇U_0, and ∇U_1, we know

$$\big\langle \boldsymbol{b}(x), \nabla(U_1 - U_0)(x)\big\rangle = 0, \qquad \forall x \in \partial D \cup D \equiv \Omega\backslash K^\circ.$$

For any trajectory $t \to \boldsymbol{X}_t(x)$ with $x \in \widetilde{D}$, we can easily show that $e^{-(U_1-U_0)\left(\boldsymbol{X}_t(x)\right)} = e^{-(U_1-U_0)(x)}$ for all $t \in \mathbb{R}$, and thus we have $\mathcal{A}(x) = e^{-(U_1-U_0)(x)}$ for any $x \in \widetilde{D}$ (the same calculation, in fact, has been shown in the proof of Proposition H.3). Hence,

$$\begin{aligned}\mathcal{M}(\boldsymbol{b}) &= \mathbb{E}_{\rho_0}\Big[\chi_{\widetilde{K}}(\cdot)\big(\mathcal{A}(\cdot)\big)^2\Big] + \mathbb{E}_{\rho_0}\Big[\chi_{\widetilde{D}}(\cdot)\big(\mathcal{A}(\cdot)\big)^2\Big] \\ &= \mathbb{E}_{\rho_0}\Big[\chi_{\widetilde{K}}(\cdot)\big(\mathcal{A}(\cdot)\big)^2\Big] + \mathbb{E}_{\rho_0}\Big[\chi_{\widetilde{D}}(\cdot)e^{-2(U_1-U_0)(\cdot)}\Big],\end{aligned}$$

where $\chi_A(\cdot)$ is an indicator function for a set A.

Next we consider the trajectory $t \to \boldsymbol{X}_t^\epsilon(x)$ with $x \in \widetilde{K}$. Since such a trajectory never enters $\widetilde{D} \supset D$ and $\boldsymbol{b}^\epsilon = \boldsymbol{b}$ on $\Omega\backslash D$, we know $\boldsymbol{X}_t^\epsilon(x) = \boldsymbol{X}_t(x)$ for any $t \in \mathbb{R}$ and $x \in \widetilde{K}$, and thus $\mathcal{A} = \mathcal{A}_\epsilon$ on $\widetilde{K}$ for any $\epsilon \in [0, \epsilon_0)$. Hence, $\mathbb{E}_{\rho_0}\Big[\chi_{\widetilde{K}}(\cdot)\big(\mathcal{A}(\cdot)\big)^2\Big] = \mathbb{E}_{\rho_0}\Big[\chi_{\widetilde{K}}(\cdot)\big(\mathcal{A}_\epsilon(\cdot)\big)^2\Big]$. By the same argument as in Proposition H.3 (by treating $\widetilde{D}$ as the domain),

$$\mathcal{M}(\boldsymbol{b}^\epsilon) - \mathcal{M}(\boldsymbol{b}) = \mathbb{E}_{\rho_0}\Big[\chi_{\widetilde{D}}(\cdot)\big(\mathcal{A}_\epsilon(\cdot)\big)^2\Big] - \mathbb{E}_{\rho_0}\Big[\chi_{\widetilde{D}}(\cdot)e^{-2(U_1-U_0)(\cdot)}\Big] \le 0,$$

where the equality is achieved only if $\langle \boldsymbol{b}^\epsilon, \nabla(U_1 - U_0)\rangle = 0$ on $\widetilde{D}$. Note that on $\widetilde{D}$,

$$\langle \boldsymbol{b}^\epsilon, \nabla(U_1 - U_0)\rangle = \langle \boldsymbol{b} + \epsilon\delta\boldsymbol{b}, \nabla(U_1 - U_0)\rangle = \epsilon\langle \delta\boldsymbol{b}, \nabla(U_1 - U_0)\rangle.$$

However, due to the fact that $\langle \delta\boldsymbol{b}, \nabla(U_1 - U_0)\rangle \neq 0$ for some open subset of D with strictly positive ρ_0-measure (as constructed in the Step (I)), the equality $\mathcal{M}(\boldsymbol{b}^\epsilon) = \mathcal{M}(\boldsymbol{b})$ cannot be achieved and thus

$$\mathcal{M}(\boldsymbol{b}^\epsilon) < \mathcal{M}(\boldsymbol{b}).$$

Since we can find a local perturbation $\delta\boldsymbol{b}$ such that $\mathcal{M}(\boldsymbol{b}^\epsilon) < \mathcal{M}(\boldsymbol{b})$ for any $\epsilon \in (0, \epsilon_0)$, then $\boldsymbol{b}$ must not be a local minimum of $\mathcal{M}$. □

H.4 Proof of Proposition H.1

By Proposition D.1, we know either

$$\nabla\mathcal{A}(x) = \mathbf{0}_d, \text{ or } \int_0^\infty \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_{-\infty}^0 \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t - \int_{-\infty}^0 \mathcal{F}_t^{(0)}(x)\,\mathrm{d}t \int_0^\infty \mathcal{F}_t^{(1)}(x)\,\mathrm{d}t = 0.$$

Define

$$K := \big\{x \in \Omega :\ \nabla\mathcal{A}(x) = \mathbf{0}_d\big\},$$

which is a closed subset of Ω by the continuity assumption on $\nabla\mathcal{A}$.

Hence, $D := \Omega\backslash K$ is open and by Lemma H.5 (i), we know $\langle \boldsymbol{b}, \nabla(U_1 - U_0)\rangle = 0$ on D. Here are a few cases to discuss.

Case (I): $K^\circ = \emptyset$.

If $K^\circ = \emptyset$, then we claim that $\boldsymbol{b}$ must be a global maximum and this contradicts with the assumption that $\text{Var}(\boldsymbol{b}) < \text{Var}^{(\max)}$. It is not hard to observe that $K^\circ = \emptyset$ implies that $\overline{D} = \Omega$. By continuity, we know $\langle \boldsymbol{b}, \nabla(U_1 - U_0)\rangle = 0$ on $\overline{D} = \Omega$. Then by Lemma H.5 part (ii), $\boldsymbol{b}$ must be a global maximum.

Case (II): $D = \emptyset$.

If $D = \emptyset$ (i.e., $K = \Omega$), then it is clear that $\mathcal{A}$ is a constant function, and thus the variance $\text{Var}(\boldsymbol{b}) = 0$. This means $\boldsymbol{b}$ is a zero-variance dynamics.

Case (III): $K^\circ \neq \emptyset$ and $D \neq \emptyset$

In order to use Proposition H.6, we need to deal with the case that some trajectories $t \to \boldsymbol{X}_t(x)$ for $x \in D$ might enter K°. Let us introduce

$$S = \Big\{x \in D : \ \big\{\boldsymbol{X}_t(x)\big\}_{t\in\mathbb{R}} \cap K^\circ \neq \emptyset\Big\}, \tag{77}$$

which essentially contains all points in D whose trajectories enter K° at some time. We can easily show that S is open: because $\boldsymbol{b}$ is assumed to be smooth, the trajectories are continuous under a small perturbation for initial states. Then the new disjoint sub-regions to consider are $\widetilde{K} := \overline{K^\circ \cup S}$ and $\widetilde{D} := \Omega\backslash\widetilde{K}$. We collect some facts for clarity:

- $\widetilde{K}^\circ = K^\circ \cup S \neq \emptyset$;
- $\widetilde{D} \subset D$, which immediately implies that $\langle \boldsymbol{b}, \nabla(U_1 - U_0)\rangle = 0$ on $\widetilde{D}$.
- If $x \in \widetilde{D}$, then $\boldsymbol{X}_t(x)$ must not enter $\widetilde{K}^\circ$. Indeed, if not, then we have either $t \to \boldsymbol{X}_t(x)$ entering K° (which contradicts with $x \notin S$), or $\boldsymbol{X}_t(x)$ entering S (which still means $\boldsymbol{X}_t(x)$ will enter K° due to the reversibility of deterministic trajectories).

We need to discuss two cases:

(a) Firstly, let us consider $\widetilde{D} = \emptyset$, i.e., $\widetilde{K} = \Omega$ and thus $\widetilde{K}^\circ = \Omega$ by the assumption that Ω is open in the topology of the space $\mathbb{R}^d$. The connectivity granted by the definition of S in (77) implies that we can divide $\widetilde{K}^\circ$ into a countable number of sub-regions on which the function $\mathcal{A}$ is a constant. More specifically, for $x \in \Omega$, define

$$R_x := \big\{y \in \Omega : \ \exists\gamma. \in C([0,1],\mathbb{R}^d), \gamma_0 = x, \gamma_1 = y, \ \mathcal{A}(\gamma_t) = \mathcal{A}(x), \ \forall t \in [0,1]\big\}.$$

Obviously, $x \in R_x$. By the invariance of $\mathcal{A}$ under the dynamical flow (see (26)) and the definition of S (77), we have $\cup_{x\in K^\circ} R_x \supset \widetilde{K}^\circ = \Omega$, which implies that $\cup_{x\in K^\circ} R_x = \Omega$.

- Suppose $x, y \in K^\circ$ and $R_x \cap R_y \neq \emptyset$, then there exists a $z \in R_x \cap R_y$ such that z connects to both x and y via a continuous path and thus $\mathcal{A}(x) = \mathcal{A}(z) = \mathcal{A}(y)$. It is then clear that $R_x = R_y$ via treating z as a bridge. Therefore, $\cup_{x\in K^\circ} R_x$ can be simplified as $\cup_{x\in E} R_x$ where $\{R_x\}_{x\in E}$ are disjoint and $E \subset K^\circ$.
- Next, we can show that E is countable. Since K° is open and $\nabla\mathcal{A}(x) = \mathbf{0}_d$ on K, we can easily verify that for each $x \in K^\circ$, there exists a local neighbor $B_\delta(x) \subset R_x$ and due to the fact that $\mathbb{Q}^d$ is dense and countable, E is at most countable.

To summarize, we have

$$\bigcup_{x\in E} R_x = \Omega,$$

where $\{R_x\}_{x\in E}$ are disjoint and E is countable.

As $\mathcal{A}$ is a constant function on R_x by the definition of $R_{(\cdot)}$, $\mathcal{A}$ is a step function on Ω. By the continuity of $\mathcal{A}$ from the assumption, and also Ω is a connected open domain, we readily know $\mathcal{A}$ must be a constant function on Ω instead, and such a $\boldsymbol{b}$ must be a zero-variance dynamics.

(b) Next, let us consider the case $\widetilde{D} \neq \emptyset$. By the assumption that $\nabla(U_1 - U_0) = \mathbf{0}_d$ only on a set with Lebesgue measure zero, we know there must exist $y \in \widetilde{D}$ such that $\nabla(U_1 - U_0)(y) \neq \mathbf{0}_d$. By the assumption that $\boldsymbol{b}$-unstable points has Lebesgue measure zero, there must exist a $\boldsymbol{b}$-stable point $x^\star$ (around y) such that $\nabla(U_1 - U_1)(x^\star) \neq \mathbf{0}_d$. Then Proposition H.6 tells us that $\boldsymbol{b}$ must not be a local minimum, which contradicts with the assumption.

I Supplementary material for numerical experiments

I.1 Details about AIS.

For AIS method, we use equally-spaced temperature distribution, i.e., $\pi_k \propto \rho_0^{1-\beta_k}\rho_1^{\beta_k}$, where $\beta_k = k/K$ for $0 \leq k \leq K$; for each transition step, we use Metropolis-adjusted Langevin algorithm with time step 0.1 to generate the chain.

More specifically, suppose $M_j(\cdot,\cdot)$ is a transition kernel which leaves π_j invariant, then

$$\mathcal{Z}_1 = \mathbb{E}\Big[e^{-\sum_{j=1}^{K}(\beta_j - \beta_{j-1})\big(U_1(x_{j-1}) - U_0(x_{j-1})\big)}\Big], \tag{78}$$

where $x_j \sim M_j(x_{j-1},\cdot)$ and $x_0 \sim \rho_0$ [27, 2]. To implement such a transition kernel, we use Metropolis-adjusted Langevin algorithm: suppose τ (which is chosen as 0.1 as a prescribed parameter) is the time step, then let $\widetilde{x}_j := x_{j-1} + \tau\nabla\log\pi_j(x_{j-1}) + \sqrt{2\tau}\xi_j$, where ξ_j are *i.i.d.* d-dimensional standard normal random variables; the state $\widetilde{x}_j$ is accepted with a rate

$$\min\Big\{1, \frac{\pi_j(\widetilde{x}_j)q(x_{j-1} \mid \widetilde{x}_j)}{\pi_j(x_{j-1})q(\widetilde{x}_j \mid x_{j-1})}\Big\},$$

where $q(x' \mid x) = \exp\big(-\frac{1}{4\tau}|x' - x - \tau\nabla\log\pi_j(x)|^2\big)$ coming from the transition probability for the Langevin step.

For each j, inside the Metropolis-Hasting correction term, we need 2 queries to ∇U_1 (i.e., $\nabla U_1(x_{j-1})$ and $\nabla U_1(\widetilde{x}_j)$ hidden inside the computation of $\nabla\log\pi_j$ for the acceptance rate), and we need two queries to U_1 when computing $U_1(x_{j-1})$ (inside (78) and $\pi_j(x_{j-1})$) and $U_1(\widetilde{x}_j)$ (inside $\pi_j(\widetilde{x}_j)$).

I.2 More implementation details

Neural network architecture. We use the following ℓ-layer neural network [11, 32] to parameterize the dynamics $\boldsymbol{b} : \Omega \to \mathbb{R}^d$ during training:

$$x \to W_\ell\big(f_{\ell-1} \circ \cdots f_2 \circ f_1(x)\big) + b_\ell,$$

where ℓ is the layer depth (one output layer and $\ell - 1$ hidden layers), $f_j(\cdot) = \eta(W_j(\cdot) + b_j)$ for $j = 1, 2, \cdots, \ell - 1$, η is the activation function, $W_j \in \mathbb{R}^{n_j} \times \mathbb{R}^{n_{j+1}}$, $b_j \in \mathbb{R}^{n_{j+1}}$ for $j = 1, 2, \cdots, \ell$. When we choose $n_j = m$ for all $2 \leq j \leq \ell$, we refer such a neural network as (ℓ, m)-architecture which is mentioned in Section 5.

More specifically, let us take $\ell = 2$ as an example: an $(\ell, m) = (2, m)$ architecture for a generic ansatz for $\boldsymbol{b} : \Omega \to \mathbb{R}^d$ refers to the following parameterization

$$\boldsymbol{b}(x) = W_2\eta(W_1 x + b_1) + b_2, \qquad \text{(generic ansatz)},$$

where weights $W_1 \in \mathbb{R}^d \times \mathbb{R}^m$, $W_2 \in \mathbb{R}^m \times \mathbb{R}^d$ and bias vectors $b_1 \in \mathbb{R}^m$, $b_2 \in \mathbb{R}^d$.

For the gradient-form ansatz, as we essentially need to parameterize a potential $V : \Omega \to \mathbb{R}$, the (ℓ, m)-architecture for V refers to the following choice when $\ell = 2$,

$$V(x) = W_2\eta(W_1 x + b_1)$$

where $W_2 \in \mathbb{R}^m \times \mathbb{R}$, $W_1 \in \mathbb{R}^d \times \mathbb{R}^m$, $b_1 \in \mathbb{R}^m$. The bias vector in the output layer is chosen as zero (i.e., $b_2 = 0$) because b_2 is a redundant parameter after taking the gradient. The dynamics $\boldsymbol{b}$ in the gradient form refers to $\boldsymbol{b} = \nabla V$.

Initialization and trial repetition. When we parameterize the flow via neural networks, weights and biases in the neural network are randomly generated. Therefore we consider two (or three) independent trials (associated with different random initializations of $\boldsymbol{b}$) for the same neural network architecture characterized by a pair (ℓ, m).

Optimization algorithm. We use the SGD algorithm to optimize parameters for $\boldsymbol{b}$. The Armijo line search algorithm (see e.g. [3, 45]) can used to find the learning rate; in practice, we notice that solving $\dot{\vartheta} = -\frac{\nabla_\vartheta \mathcal{M}_{t_-,t_+}(\boldsymbol{b}_\vartheta)}{|\nabla_\vartheta \mathcal{M}_{t_-,t_+}(\boldsymbol{b}_\vartheta)|}$ with a relatively large learning rate also works well for numerical examples considered in Section 6, and it is used for training in Section 6; ϑ are trainable parameters and $\boldsymbol{b}_\vartheta$ is the flow parameterized by ϑ. The way to approximate the loss function $\mathcal{M}_{t_-,t_+}(\boldsymbol{b}) \equiv \mathbb{E}_0\big[|\mathcal{A}_{t_-,t_+}|^2\big]$ (see Section 4) and in particular its gradient with respect to parameters in $\boldsymbol{b}$ will be explained in Appendix I.3 and I.4.

I.3 An integration-based forward propagation method to compute the estimator and its gradient

To estimate $\mathcal{A}_{t_-,t_+}(x)$ in (8) with $t_+ = t_- + 1$ and $t_- \in [-1, 0]$, the most straightforward approach is to compute $\mathcal{F}_t^{(k)}(x)$ at time grid points and then employ some integration scheme like the trapezoidal quadrature method.

More specifically, we first discretize the time interval $[-1, 1]$ by $2N_t + 1$ equally-spaced points $t_m := \frac{1}{N_t} m$ where $-N_t \le m \le N_t$. Then we use classical ODE integration schemes (we use RK4) to propagate the dynamics $\boldsymbol{X}_t(x)$ both forward and backward in time to estimate the following quantities:

$$\hat{U}_{0,m} = U_0\big(\boldsymbol{X}_{t_m}(x)\big), \quad \hat{U}_{1,m} = U_1\big(\boldsymbol{X}_{t_m}(x)\big), \quad \hat{D}_m = \boldsymbol{\nabla} \cdot \boldsymbol{b}\big(\boldsymbol{X}_{t_m}(x)\big).$$

Then

$$\mathcal{F}_{t_m}^{(k)}(x) \approx e^{-\hat{U}_{k,m} + \frac{1}{N_t}\mathrm{sign}(m)\mathcal{Q}_{\mathrm{Tr}}(\hat{D}_0, \hat{D}_1, \cdots, \hat{D}_m)}$$

where

$$\mathcal{Q}_{\mathrm{Tr}}(\hat{D}_0, \hat{D}_1, \cdots, \hat{D}_m) = \sum_{j=0}^{m} \hat{D}_j - \frac{1}{2}\big(\hat{D}_0 + \hat{D}_m\big)$$

is the trapezoidal quadrature scheme. Then similarly, we can use the trapezoidal quadrature scheme to estimate $\int_{t_j - t_+}^{t_j - t_-} \mathcal{F}_s^{(0)}(x)\,\mathrm{d}s$ given the values $\mathcal{F}_{t_m}^{(0)}(x)$ for $-N_t \le m \le N_t$ and the time t_j with $t_j \in [t_-, t_+]$. After we have approximated values for $\frac{\mathcal{F}_{t_j}^{(1)}(x)}{\int_{t_j - t_+}^{t_j - t_-} \mathcal{F}_s^{(0)}(x)\,\mathrm{d}s}$, the trapezoidal quadrature scheme is utilized again to approximate $\mathcal{A}_{t_-,t_+}(x)$. The computational cost is mostly dominated by the ODE integration, e.g., propagating $\boldsymbol{X}_t(x)$ or evaluating $\boldsymbol{\nabla} \cdot \boldsymbol{b}(\boldsymbol{X}_s(x))$ in general. Therefore, this straightforward integration-based method has linear computational cost with respect to N_t and is thus expected to be optimal. Using the same principle, we can estimate the gradient of the second moment with respect to parameters in the dynamics (see (34)) during the training.

I.4 An ODE-based forward propagation method to compute the estimator and its gradient for $t_- = 0$, $t_+ = 1$

To simplify notations, let us introduce

$$\mathscr{B}_t(x) := \int_{t-t_+}^{t-t_-} \mathcal{F}_s^{(0)}(x)\,\mathrm{d}s, \text{ and thus, } \mathcal{A}_{t_-,t_+}(x) \overset{(8)}{=} \int_{t_-}^{t_+} \frac{\mathcal{F}_t^{(1)}(x)}{\mathscr{B}_t(x)}\,\mathrm{d}t, \tag{79}$$

which implicitly depends on $\boldsymbol{b}$. Let us denote $\alpha_t(x) := \int_{t_-}^{t} \frac{\mathcal{F}_s^{(1)}(x)}{\mathscr{B}_s(x)}\,\mathrm{d}s$.

I.4.1 ODE dynamics to compute the estimator

Then the estimator $\mathcal{A}_{0,1}$ is simply $\alpha_1(x)$. To compute $\mathcal{A}_{0,1}$, we simply need to run the following ODE:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\alpha_t(x) = e^{-U_1\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)/\mathscr{B}_t(x), & \alpha_0(x) = 0, \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathscr{B}_t(x) = e^{-U_0\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x) - e^{-U_0\left(\boldsymbol{X}_t^{\text{lag}}(x)\right)}\mathcal{J}_t^{\text{lag}}(x), & \mathscr{B}_0(x) = \mathscr{B}_1^R(x), \\ \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t(x) = \boldsymbol{b}\big(\boldsymbol{X}_t(x)\big), & \boldsymbol{X}_0(x) = x, \\ \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t^{\text{lag}}(x) = \boldsymbol{b}\big(\boldsymbol{X}_t^{\text{lag}}(x)\big), & \boldsymbol{X}_0^{\text{lag}}(x) = \boldsymbol{X}_1^R(x), \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}_t(x) = \boldsymbol{\nabla}\cdot\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\mathcal{J}_t(x), & \mathcal{J}_0(x) = 1, \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}_t^{\text{lag}}(x) = \boldsymbol{\nabla}\cdot\boldsymbol{b}\big(\boldsymbol{X}_t^{\text{lag}}(x)\big)\mathcal{J}_t^{\text{lag}}(x), & \mathcal{J}_0^{\text{lag}}(x) = \mathcal{J}_1^R(x), \end{cases} \tag{80}$$

where $\boldsymbol{X}_t^{\text{lag}}(x) := \boldsymbol{X}_{t-1}(x)$ and $\mathcal{J}_t^{\text{lag}}(x) := \mathcal{J}_{t-1}(x)$. Moreover, in order to obtain the initial condition, we shall run the dynamics backward in time, i.e., simulating the following ODE on $[0,1]$,

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}_t^R(x) = -\boldsymbol{b}\big(\boldsymbol{X}_t^R(x)\big), & \boldsymbol{X}_0^R(x) = x, \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}_t^R(x) = -\boldsymbol{\nabla}\cdot\boldsymbol{b}\big(\boldsymbol{X}_t^R(x)\big)\mathcal{J}_t^R(x), & \mathcal{J}_0^R(x) = 1, \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathscr{B}_t^R(x) = e^{-U_0\left(\boldsymbol{X}_t^R(x)\right)}\mathcal{J}_t^R(x), & \mathscr{B}_0^R(x) = 0, \end{cases} \tag{81}$$

where the superscript R means the reversed process. As a remark, the auxiliary backward ODE (81) has dimension $d+2$ and the ODE (80) has dimension $2d+4$.

I.4.2 ODE dynamics to compute the gradient

Denote ϑ as a vector containing parameters in $\boldsymbol{b}$ and let N_p be the number of parameters. In what follows, we also vectorize all quantities involving ϑ, e.g., for each parameter ϑ_j, there is a corresponding $\boldsymbol{Y}_t^{(j)}(x)$ in (36) and we simply use the notation $\boldsymbol{Y}_t(x)$ to be a matrix whose j^{th} column is $\boldsymbol{Y}_t^j(x)$ to save notations; the same convention applies to other quantities.

Next, we shall similarly re-write the expression $\nabla_\vartheta\mathcal{A}_{0,1}$ (implicitly given in $\nabla_\vartheta\mathcal{M}_{0,1}(\boldsymbol{b})$ (34)) in terms of outputs from an ODE. Using the same idea, let us introduce

$$\begin{cases} \boldsymbol{D}_t(x) := \int_0^t \frac{\mathcal{G}_r^{(1)}(x)\int_{r-1}^r \mathcal{F}_s^{(0)}(x)\,\mathrm{d}s - \mathcal{F}_r^{(1)}(x)\int_{r-1}^r \mathcal{G}_s^{(0)}(x)\,\mathrm{d}s}{\left(\int_{r-1}^r \mathcal{F}_s^{(0)}(x)\,\mathrm{d}s\right)^2}\,\mathrm{d}r \\ \boldsymbol{g}_t(x) := \int_{t-1}^t \mathcal{G}_s^{(0)}(x)\,\mathrm{d}s, \\ \boldsymbol{L}_t(x) := \int_0^t \nabla_\vartheta(\nabla_x\cdot\boldsymbol{b})\big(\boldsymbol{X}_s(x)\big)\,\mathrm{d}s, \\ \boldsymbol{H}_t(x) := \int_0^t \Big(\nabla(\boldsymbol{\nabla}\cdot\boldsymbol{b})(\boldsymbol{X}_s(x))\Big)^T\boldsymbol{Y}_s(x)\,\mathrm{d}s, \end{cases}$$

and then we can rewrite quantities involved inside $\nabla_\vartheta\mathcal{M}_{0,1}(\boldsymbol{b})$ as follows:

$$\text{the evolution of } \boldsymbol{X}_t(x), \boldsymbol{X}_t^{\text{lag}}(x), \mathcal{J}_t(x),\ \mathcal{J}_t^{\text{lag}}(x),\ \mathscr{B}_t(x),\ \alpha_t(x) \text{ in (80)},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{D}_t(x) = \begin{pmatrix} \frac{e^{-U_1\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)\big(-\nabla U_1\big(\boldsymbol{X}_t(x)\big)^T\boldsymbol{Y}_t(x) + \boldsymbol{H}_t(x) + \boldsymbol{L}_t(x)\big)}{\mathscr{B}_t(x)} \\ -\frac{e^{-U_1\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)\boldsymbol{g}_t(x)}{\big(\mathscr{B}_t(x)\big)^2} \end{pmatrix},\ \boldsymbol{D}_0(x) = \boldsymbol{0}_{N_p},$$

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{g}_t(x) &= \begin{pmatrix} e^{-U_0\left(\boldsymbol{X}_t(x)\right)}\mathcal{J}_t(x)\big(-\nabla U_0\big(\boldsymbol{X}_t(x)\big)^T\boldsymbol{Y}_t(x) \\ \qquad\qquad + \boldsymbol{H}_t(x) + \boldsymbol{L}_t(x)\big) \\ -e^{-U_0\left(\boldsymbol{X}_t^{\mathrm{lag}}(x)\right)}\mathcal{J}_t^{\mathrm{lag}}(x)\big(-\nabla U_0\big(\boldsymbol{X}_t^{\mathrm{lag}}(x)\big)^T\boldsymbol{Y}_t^{\mathrm{lag}}(x) \\ \qquad\qquad + \boldsymbol{H}_t^{\mathrm{lag}}(x) + \boldsymbol{L}_t^{\mathrm{lag}}(x)\big) \end{pmatrix}, && \boldsymbol{g}_0(x) = \boldsymbol{g}_1^R(x), \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{H}_t(x) &= \Big(\nabla(\boldsymbol{\nabla}\cdot\boldsymbol{b})\big(\boldsymbol{X}_t(x)\big)\Big)^T\boldsymbol{Y}_t(x), && \boldsymbol{H}_0(x) = \boldsymbol{0}_{N_p}, \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{H}_t^{\mathrm{lag}}(x) &= \Big(\nabla(\boldsymbol{\nabla}\cdot\boldsymbol{b})\big(\boldsymbol{X}_t^{\mathrm{lag}}(x)\big)\Big)^T\boldsymbol{Y}_t^{\mathrm{lag}}(x), && \boldsymbol{H}_0^{\mathrm{lag}}(x) = \boldsymbol{H}_1^R(x), \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{L}_t(x) &= \nabla_\vartheta(\nabla_x\cdot\boldsymbol{b})\big(\boldsymbol{X}_t(x)\big), && \boldsymbol{L}_0(x) = \boldsymbol{0}_{N_p}, \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{L}_t^{\mathrm{lag}}(x) &= \nabla_\vartheta(\nabla_x\cdot\boldsymbol{b})\big(\boldsymbol{X}_t^{\mathrm{lag}}(x)\big), && \boldsymbol{L}_0^{\mathrm{lag}}(x) = \boldsymbol{L}_1^R(x), \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{Y}_t(x) &= \nabla\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big)\boldsymbol{Y}_t(x) + \nabla_\vartheta\boldsymbol{b}\big(\boldsymbol{X}_t(x)\big), && \boldsymbol{Y}_0(x) = \boldsymbol{0}_{d\times N_p}, \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{Y}_t^{\mathrm{lag}}(x) &= \nabla\boldsymbol{b}\big(\boldsymbol{X}_t^{\mathrm{lag}}(x)\big)\boldsymbol{Y}_t^{\mathrm{lag}}(x) + \nabla_\vartheta\boldsymbol{b}\big(\boldsymbol{X}_t^{\mathrm{lag}}(x)\big), && \boldsymbol{Y}_0^{\mathrm{lag}}(x) = \boldsymbol{Y}_1^R(x),
\end{aligned}
$$

where the initial conditions come from solving an auxiliary backward equation on $[0,1]$, given as follows:

$$
\begin{aligned}
&\text{the evolution of } \boldsymbol{X}_t^R(x), \mathcal{J}_t^R(x), \mathscr{B}_t^R(x) \text{ in (81)}, \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{g}_t^R(x) &= e^{-U_0(\boldsymbol{X}_t^R(x))}\mathcal{J}_t^R(x)\begin{pmatrix} -\nabla U_0\big(\boldsymbol{X}_t^R(x)\big)^T\boldsymbol{Y}_t^R(x) \\ +\boldsymbol{H}_t^R(x) + \boldsymbol{L}_t^R(x) \end{pmatrix}, && \boldsymbol{g}_0^R(x) = \boldsymbol{0}_{N_p}, \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{H}_t^R(x) &= -\Big(\nabla(\boldsymbol{\nabla}\cdot\boldsymbol{b})\big(\boldsymbol{X}_t^R(x)\big)\Big)^T\boldsymbol{Y}_t^R(x), && \boldsymbol{H}_0^R(x) = \boldsymbol{0}_{N_p}, \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{L}_t^R(x) &= -\nabla_\vartheta(\nabla_x\cdot\boldsymbol{b})\big(\boldsymbol{X}_t^R(x)\big), && \boldsymbol{L}_0^R(x) = \boldsymbol{0}_{N_p}, \\
\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{Y}_t^R(x) &= -\nabla\boldsymbol{b}\big(\boldsymbol{X}_t^R(x)\big)\boldsymbol{Y}_t^R(x) - \nabla_\vartheta\boldsymbol{b}\big(\boldsymbol{X}_t^R(x)\big), && \boldsymbol{Y}_0^R(x) = \boldsymbol{0}_{d\times N_p}.
\end{aligned}
$$

As a remark, we combine the auxiliary states $\boldsymbol{Y}_t(x)$ for all parameters together, so that all three terms $\boldsymbol{Y}_t(x)$, $\boldsymbol{Y}_t^{\mathrm{lag}}(x)$, $\boldsymbol{Y}_t^R(x)$ have the dimension $d \times N_p$. Below is a table that summarizes the dimension of all components.

Table 2: This table summarizes all variables and their corresponding dimensions.

notations	dimensions
$\mathcal{J}_t(x)$, $\mathcal{J}_t^{\mathrm{lag}}(x)$, $\mathscr{B}_t(x)$, $\alpha_t(x)$, $\mathcal{J}_t^R(x)$, $\mathscr{B}_t^R(x)$	1
$\boldsymbol{X}_t(x)$, $\boldsymbol{X}_t^{\mathrm{lag}}(x)$, $\boldsymbol{X}_t^R(x)$	d
$\boldsymbol{D}_t(x)$, $\boldsymbol{g}_t(x)$, $\boldsymbol{H}_t(x)$, $\boldsymbol{H}_t^{\mathrm{lag}}(x)$, $\boldsymbol{L}_t(x)$, $\boldsymbol{L}_t^{\mathrm{lag}}(x)$, $\boldsymbol{g}_t^R(x)$, $\boldsymbol{L}_t^R(x)$, $\boldsymbol{H}_t^R(x)$	N_p
$\boldsymbol{Y}_t(x)$, $\boldsymbol{Y}_t^{\mathrm{lag}}(x)$, $\boldsymbol{Y}_t^R(x)$	$d \times N_p$

In the end, the outputs we need are

$$
\mathcal{A}_{0,1}(x) = \alpha_1(x), \quad \nabla_\vartheta\mathcal{A}_{0,1}(x) = \boldsymbol{D}_1(x), \quad \text{so that } \nabla_\vartheta\mathcal{M}_{0,1}(\boldsymbol{b}) = 2\mathbb{E}_{\rho_0}\big[\alpha_1(x)\boldsymbol{D}_1(x)\big].
$$

I.5 A discussion on the backward propagation for differentiation

In the formulas (34) and (35), a crucial step is to evaluate

$$
\Phi(t,x) := \big\langle\nabla\varphi(\boldsymbol{X}_t(x)), \boldsymbol{Y}_t(x)\big\rangle, \quad \varphi = U_k, \text{ or } \boldsymbol{\nabla}\cdot\boldsymbol{b}.
$$

for time $t \in [-T, T]$ where $T = t_+ - t_-$. For any fixed t, this value can be efficiently measured by the adjoint equation with backward propagation proposed in [9]. For instance, let us consider $t > 0$

and

$$\Phi(t,x) \overset{(39)}{=} \Big\langle \nabla\varphi\big(\boldsymbol{X}_t(x)\big), \int_0^t \boldsymbol{C}_{t,s}(x)\, \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big)\, \mathrm{d}s \Big\rangle$$
$$= \int_0^t \Big\langle \underbrace{\boldsymbol{C}_{t,s}(x)^T \nabla\varphi(\boldsymbol{X}_t(x))}_{=:\mathscr{A}(s,x)}, \delta\boldsymbol{b}\big(\boldsymbol{X}_s(x)\big) \Big\rangle\, \mathrm{d}s$$

Moreover, it can be easily verified by the definition (40) that for $s \in [0,t]$,

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{A}(s,x) = -\nabla\boldsymbol{b}^T\big(\boldsymbol{X}_s(x)\big)\mathscr{A}(s,x), \quad \mathscr{A}(t,x) = \nabla\varphi\big(\boldsymbol{X}_t(x)\big).$$

The above two equations are Eqs. (4) & (5) in [9]. This can eliminate the need to simulate $\boldsymbol{Y}_t(x)$, whose memory cost scales like $\mathcal{O}(dN_p)$. However, in the finite-time NEIS scheme, we need to estimate $\Phi(t,x)$ not only for a fixed t but also for $t \in [-1,1]$. Therefore, the computational cost scales like $\mathcal{O}(N_t^2)$ where N_t is the number of time-discretization (or say the depth of the flow map in the normalizing flow context). As a comparison, the forward propagation method uses more memory but could be computationally cheaper as it only needs to visit the whole trajectory for $\mathcal{O}(1)$ times (see the table below).

Table 3: A comparison between the forward and backward propagation in computing the derivative of $\mathcal{M}_{0,1}(\boldsymbol{b})$ with respect to parameters; see (34) and (35). The notation N_p is the number of parameters for the training and N_t is the number of grid points in time-discretization.

Method	Key difference	Memory	Computational cost
Backward propagation	simulate $\mathscr{A}(s,x)$ instead of $\boldsymbol{Y}_s(x)$	$\mathcal{O}(N_p)$	$\mathcal{O}(N_t^2)$
Forward propagation	simulate $\boldsymbol{Y}_s(x)$	$\mathcal{O}(dN_p)$	$\mathcal{O}(N_t)$

I.6 A discussion on query complexity

The query complexities to estimate $\mathcal{Z}_1$ by both AIS and NEIS are summarized in the Table 4 below:

Table 4: A summary of query complexities to U_1 when estimating $\mathcal{Z}_1$ for various methods. K is the transition step in AIS; N is the number of time steps ($\Delta t = 1/N$) for either integration-based or ODE-based discretization; s is the order of ODE integration schemes. The integration-based method does not depend on s as the query to U_1 is achieved during trapezoidal integral (see Appendix I.3). See the AIS-K algorithm and relevant analysis in Appendix I.1

	AIS-K	ODE-based discretization (80)	integration-based discretization (see Appendix I.3)
U_1	$2K$	sN	$2N$
∇U_1	$2K$	0	0

J More training and comparisons results

Table 5: Comparison of NEIS with AIS: we include training and estimation query costs for AIS and NEIS, as well as statistics for 10 independent estimates of $\mathcal{Z}_1$; for each method, we first set the query cost to obtain one approximated value of $\mathcal{Z}_1$ and then accordingly choose the sample size; we repeatedly estimate $\mathcal{Z}_1$ for 10 times and report the mean and std of these 10 estimates (in the form of mean $\pm$ std). For NEIS, we consider multiple random initializations for training and therefore, there are multiple estimates about $\mathcal{Z}_1$ in the last row in each panel (either 2 or 3 estimates). Estimates in AIS, however, simply refer to estimating results from independent experiments. The exact value $\mathcal{Z}_1 = 1$ and 1 MB $= 10^6$; the best result using AIS is colored in blue and the best result using NEIS is colored in green. The layer number $\ell = 2$ for both generic and gradient ansatz.

Asymmetric 2-mode Gaussian mixture (2D):

	AIS-10	AIS-100	NEIS (Generic, $m = 20$)	NEIS (Generic, $m = 30$)	NEIS (Grad, $m = 20$)	NEIS (Grad, $m = 30$)
training cost	N/A	N/A	U_1: 2 MB ∇U_1: 2.1 MB			
cost per estimate	U_1: 6.2 MB ∇U_1: 6.2 MB		U_1: 4.2 MB ∇U_1: 0			
10 estimates	0.932 ± 0.466	1.033 ± 0.071	1.011 ± 0.013	0.997 ± 0.016	0.997 ± 0.007	0.999 ± 0.012
	1.068 ± 0.499	1.031 ± 0.166	0.997 ± 0.009	1.001 ± 0.011	1.000 ± 0.008	1.003 ± 0.007

Symmetric 4-mode Gaussian mixture (10D):

	AIS-10	AIS-100	NEIS (Grad, $m = 30$)	NEIS (Grad, $m = 40$)
training cost	N/A	N/A	U_1: 11.5 MB ∇U_1: 12.8 MB	
cost per estimate	U_1: 89.4 MB ∇U_1: 89.4 MB		U_1: 76.6$\sim$76.8 MB ∇U_1: 0	
10 estimates	0.987 ± 0.074	1.005 ± 0.019	0.998 ± 0.004	0.998 ± 0.011
	0.997 ± 0.052	1.000 ± 0.014	1.001 ± 0.005	0.996 ± 0.008

Funnel distribution (10D):

	AIS-10	AIS-100	NEIS (linear (22a))	NEIS (two-parameter (22b))
training cost	N/A	N/A	U_1: 40.4 MB ∇U_1: 48.8 MB	U_1: 20.2 MB ∇U_1: 20.2 MB
cost per estimate	U_1: 341.3 MB ∇U_1: 341.3 MB		U_1: 293 MB ∇U_1: 0	U_1: 121.2 MB ∇U_1: 0
10 estimates	0.681 ± 0.022	0.767 ± 0.035	0.816 ± 0.089	0.984 ± 0.036
			0.847 ± 0.054	
			0.890 ± 0.171	

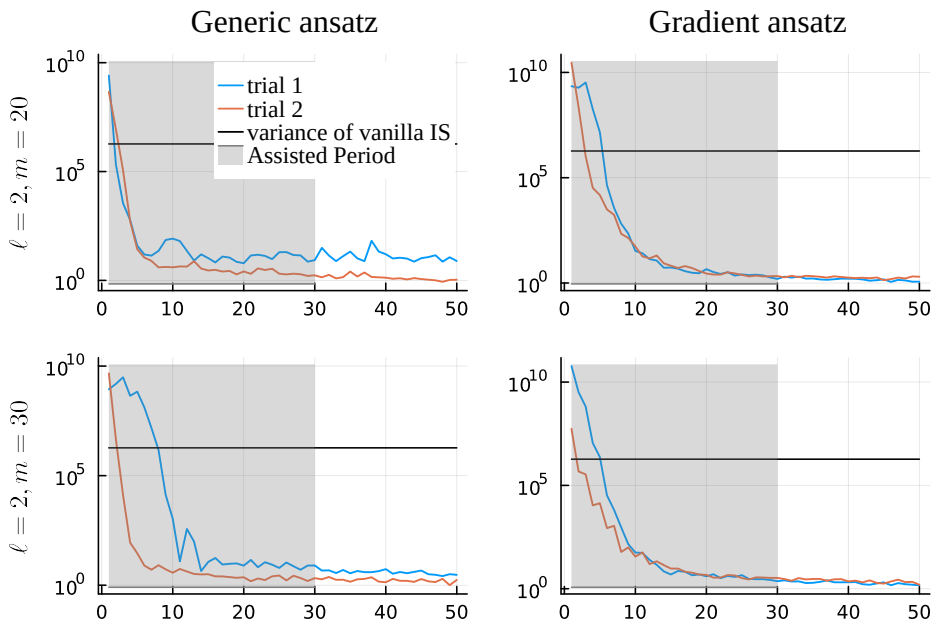


Figure 7: Asymmetric 2-mode Gaussian mixture in 2D: variance against SGD steps for the asymmetric 2-mode Gaussian mixture (20) in two trial runs. (8) was discretized with time step $\frac{1}{50}$. A mini-batch of sample size 200 was used throughout the training. The biasing parameters used in (19) are $\upsilon = 0.6$, $c = 0.1$ and $\varsigma = 1$ (see (19) and the follow-up paragraph).

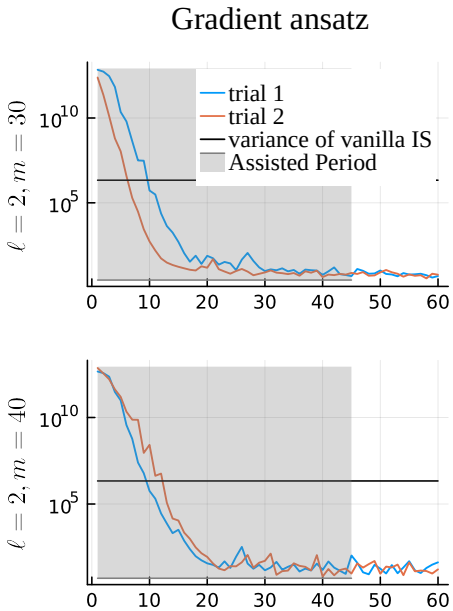


Figure 8: Symmetric 4-mode Gaussian mixture in 10D: variance as a function of SGD training step in 2 trial runs, with gradient form ansatz for $\boldsymbol{b}$. (8) was discretized with time step $\frac{1}{60}$. The sample size of the mini-batch was 800 during the training. The biasing parameters are $\upsilon = 0.75$, $c = 0.3$ and $\varsigma = 1$ (see (19) and the follow-up paragraph).

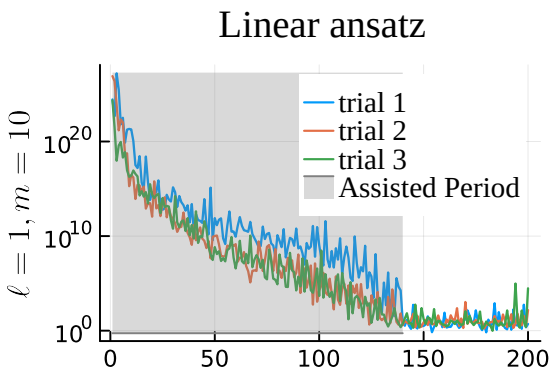


Figure 9: Funnel distribution in 10D: we consider the finite-time NEIS scheme with $t_- = -\frac{1}{2}$ and use the linear ansatz (22a). During the training, the integral inside $\mathcal{A}_{-\frac{1}{2},\frac{1}{2}}$ was discretized with time step $\frac{1}{100}$, and the mini-batch sample size was 10^3. The biasing parameters are $\upsilon = 0.7$, $c = 0.3$ and $\varsigma = 1$ (see (19) and the follow-up paragraph).

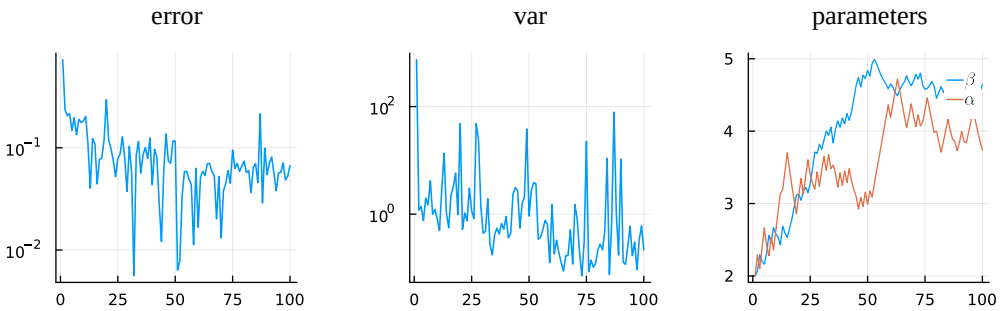
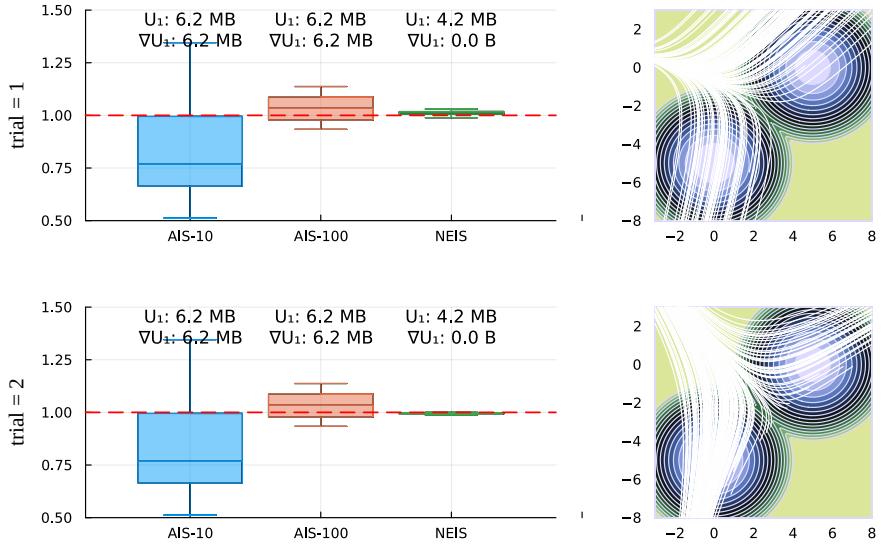
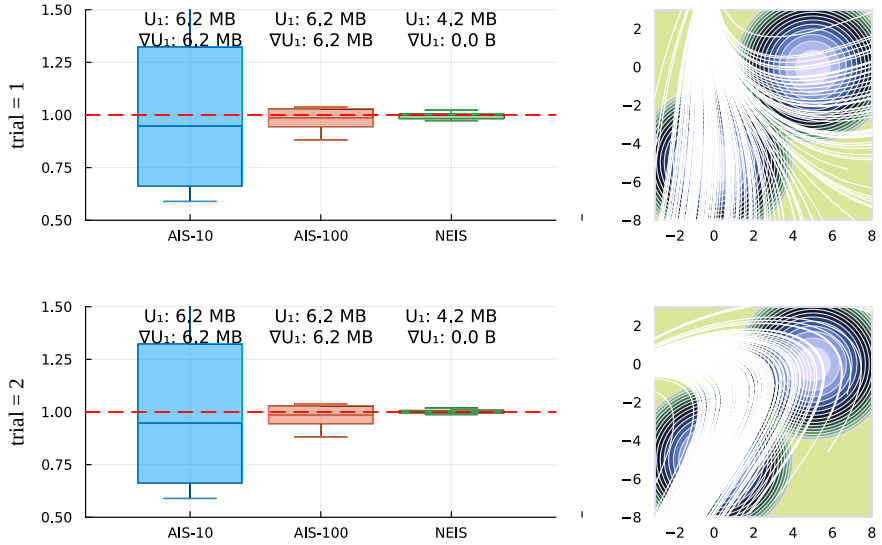


Figure 10: Funnel distribution in 10D: we consider the finite-time NEIS scheme with $t_- = -\frac{1}{2}$ and use the two-parameter ansatz (22b). During the training, the integral inside $\mathcal{A}_{-\frac{1}{2},\frac{1}{2}}$ was discretized with time step $\frac{1}{100}$, and the mini-batch sample size was 10^3. The direct training method is used.

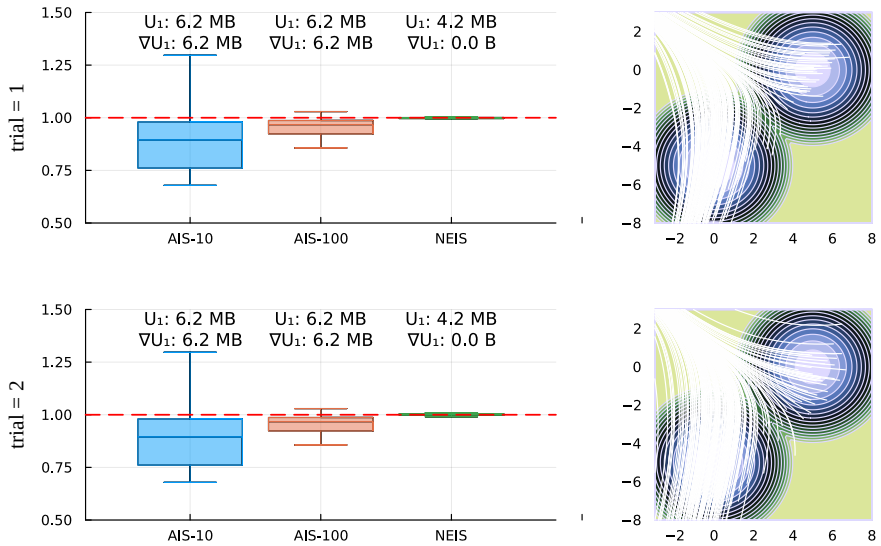


(a) Generic ansatz, $\ell = 2$, $m = 20$

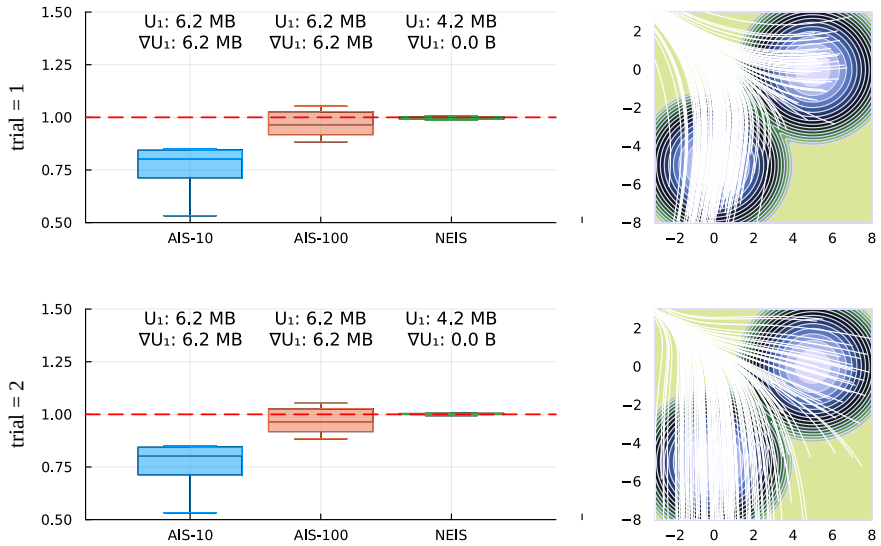


(b) Generic ansatz, $\ell = 2$, $m = 30$

Figure 11: Asymmetric 2-mode Gaussian mixture in 2D: we consider the **generic ansatz** with $\ell = 2$ as the architecture for training. In the left part, we show a comparison of NEIS using optimized flow with AIS, under fixed query budget: we estimate $\mathcal{Z}_1$ using the above mentioned query budget for each method and then repeat the experiment 10 times; we show a boxplot of these 10 independent estimates for each method. In the right part, we plot a contour of U_1 together with streamlines of optimized flows. For each architecture, we use two random initializations for training, which refer to two trials above; the estimates using AIS are reused within each panel.

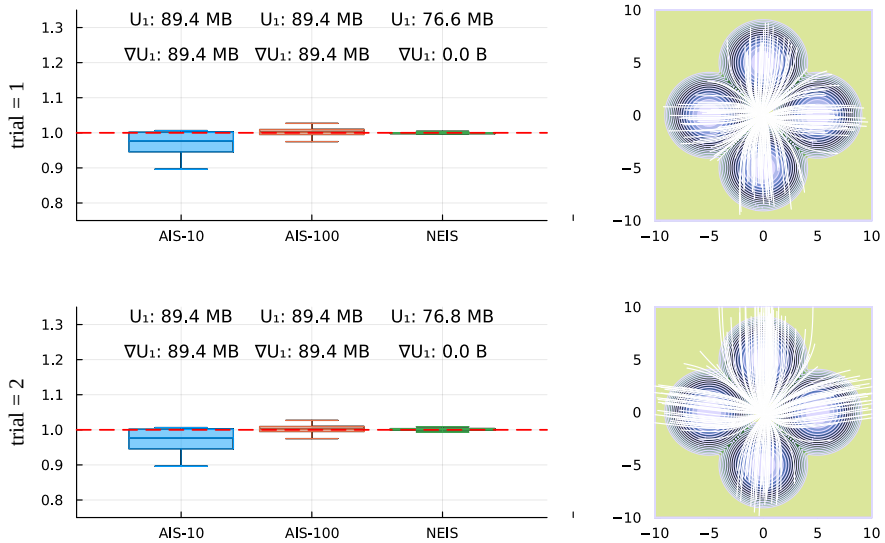


(a) Gradient ansatz, $\ell = 2$, $m = 20$

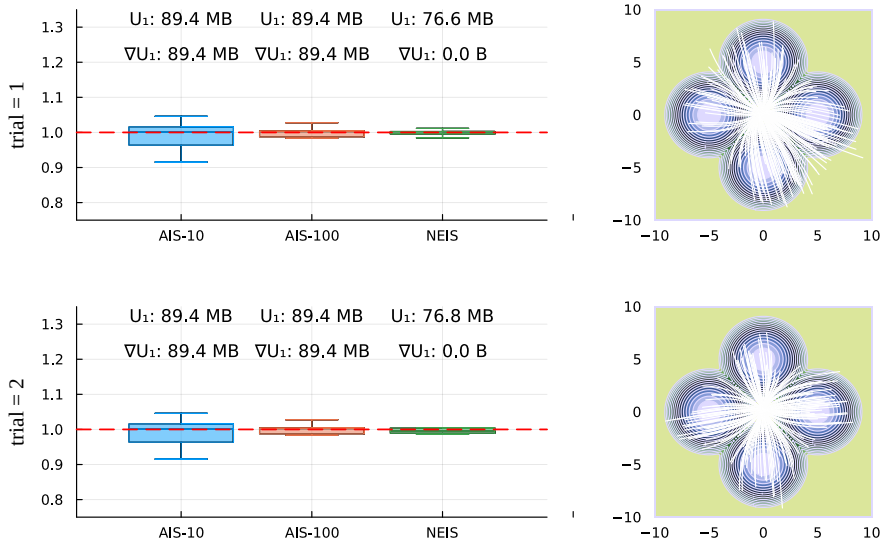


(b) Gradient ansatz, $\ell = 2$, $m = 30$

Figure 12: Asymmetric 2-mode Gaussian mixture in 2D: we consider the **gradient ansatz** with $\ell = 2$ as the architecture for training. In the left part, we show a comparison of NEIS using optimized flow with AIS, under fixed query budget: we estimate $\mathcal{Z}_1$ using the above mentioned query budget for each method and then repeat the experiment 10 times; we show a boxplot of these 10 independent estimates for each method. In the right part, we plot a contour of U_1 together with streamlines of optimized flows. For each architecture, we use two random initializations for training, which refer to two trials above; the estimates using AIS are reused within each panel.

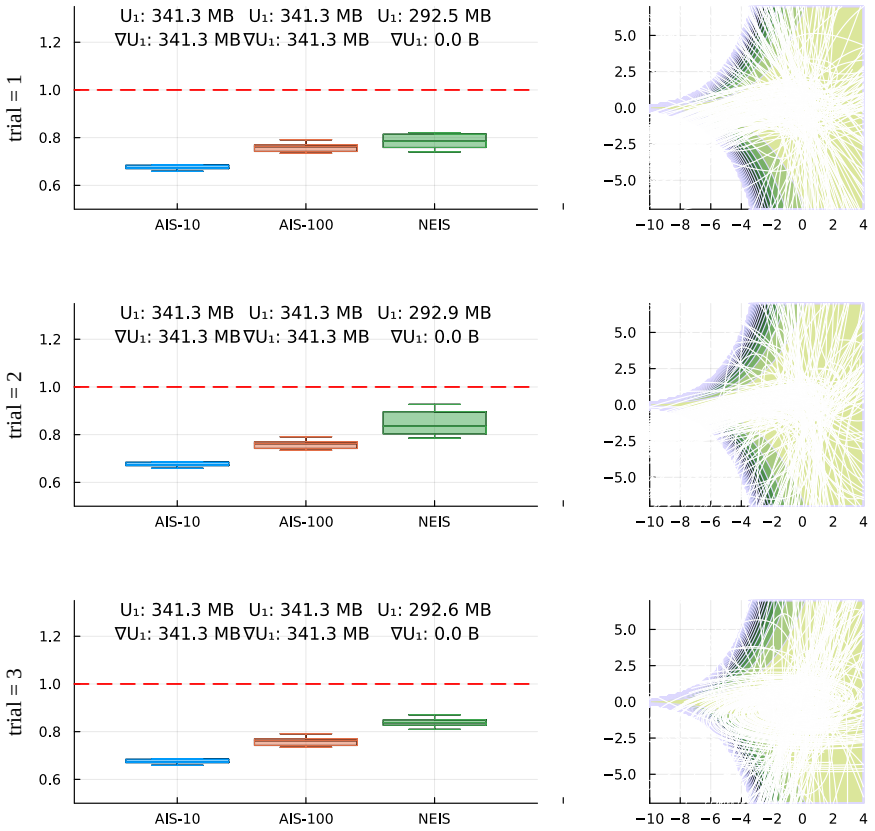


(a) Gradient ansatz, $\ell = 2$, $m = 30$

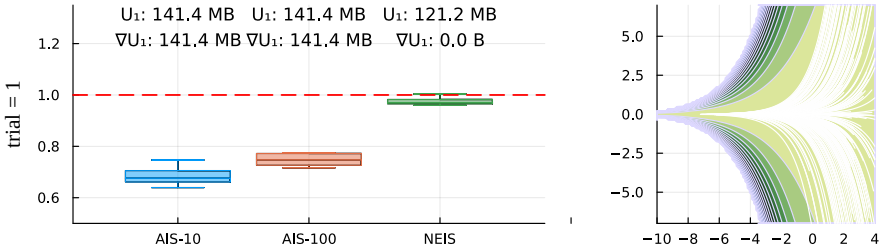


(b) Gradient ansatz, $\ell = 2$, $m = 40$

Figure 13: Symmetric 4-mode Gaussian mixture in 10'D: we consider the **gradient ansatz** with $\ell = 2$ as the architecture for training. In the left part, we show a comparison of NEIS using optimized flow with AIS, under fixed query budget: we estimate $\mathcal{Z}_1$ using the above mentioned query budget for each method and then repeat the experiment 10 times; we show a boxplot of these 10 independent estimates for each method. In the right part, we plot a contour of projected U_1 (more specifically the function $(x_1, x_2) \to U_1\big([\,x_1\ x_2\ 0 \cdots\,]\big)$) together with streamlines of optimized flows projected to the x_1-x_2 plane. For each architecture, we use two random initializations for training, which refer to two trials above; the estimates using AIS are reused within each panel.



(a) Linear ansatz (22a)



(b) Two-parameter ansatz (22b)

Figure 14: Funnel distribution in 10D: we consider the **generic linear ansatz** and a **two-parameter ansatz** as the architecture for training. In the left part, we show a comparison of NEIS using optimized flow with AIS, under fixed query budget: we estimate $\mathcal{Z}_1$ using the above mentioned query budget for each method and then repeat the experiment 10 times; we show a boxplot of these 10 independent estimates for each method. In the right part, we plot the contour of projected ρ_1 in log10 scale (more specifically we plot the function $(x_1, x_2) \to \log_{10}\left(\rho_1([\,x_1\ x_2\ 0\cdots\,])\right)$) together with streamlines of optimized flows projected to the x_1-x_2 plane. For the linear ansatz, we use three random initializations for training, which refer to three trials above; for the two-parameter ansatz, we only use a particular initialization (see the paragraph below (22b)), which refers to the only trial in part (b); the estimates using AIS are reused within each panel.