CONTACT WASSERSTEIN GEODESICS FOR NON-CONSERVATIVE SCHRÖDINGER BRIDGES

Anonymous authors
Paper under double-blind review

ABSTRACT

The Schrödinger Bridge provides a principled framework for modeling stochastic processes between distributions; however, existing methods are limited by energy-conservation assumptions, which constrains the bridge's shape preventing it from model varying-energy phenomena. To overcome this, we introduce the *non-conservative generalized Schrödinger bridge* (NCGSB), a novel, energy-varying reformulation based on contact Hamiltonian mechanics. By allowing energy to change over time, the NCGSB provides a broader class of real-world stochastic processes, capturing richer and more faithful intermediate dynamics. By parameterizing the Wasserstein manifold, we lift the bridge problem to a tractable geodesic computation in a finite-dimensional space. Unlike computationally expensive iterative solutions, our *contact Wasserstein geodesic* (CWG) is naturally implemented via a ResNet architecture and relies on a non-iterative solver with near-linear complexity. Furthermore, CWG supports guided generation by modulating a task-specific distance metric. We validate our framework on tasks including manifold navigation, molecular dynamics predictions, and image generation, demonstrating its practical benefits and versatility.

1 INTRODUCTION

Inferring the stochastic process that most likely generates a set of sparse observations is a fundamental challenge, e.g., in cellular dynamics (Yeo et al., 2021; Zhang et al., 2024; Moon et al., 2019), meteorology (Franzke et al., 2015), and economics (Kazakevičius et al., 2021; Huang et al., 2024). Here, the target is not merely the distributions of observed data, but rather the underlying dynamics of cell populations, weather patterns, or economic phenomena, enabling reconstruction of missing intermediate states and predicting the systems' future evolution.

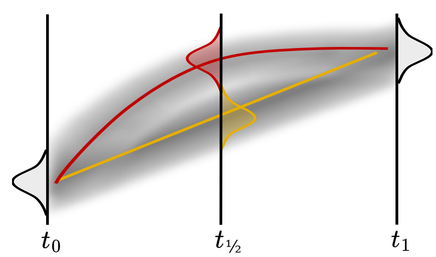

Figure 1: Probability paths obtained under energy-conserving constraints (—) and without such constraints (—, corresponding to energy-decreasing paths). This relaxation increases modeling flexibility in applications where distributions at intermediate time steps are of interest.

The *Schrödinger Bridge* (SB, Schrödinger (1931)) is a powerful mathematical framework to address this. SB seeks the most likely stochastic path between marginals (i.e., observations), while being close to a reference process, typically Brownian motion. This offers a general stochastic optimal control perspective that encompasses both *Optimal Transport* (OT, Vargas et al. (2021)) and generative approaches such as diffusion models (Ho et al., 2020; Chen et al., 2024), which can be interpreted as optimal bridges with a Gaussian initial marginal (Bortoli et al., 2021). Unfortunately, current SB solvers operate on an infinite probability space and rely on iterative forward–backward stochastic simulations or progressive refinement of the reference dynamics. This leads to complex and costly optimizations, limiting adoption.

Solutions provided by the SB preserve the distribution's energy throughout the full stochastic path. Here, energy is understood as a combination of *kinetic energy*, which reflects how fast samples move across the probability manifold, and *potential energy* from the underlying landscape. This energy preservation constrains the shape of the bridge and excludes stochastic paths with varying energy profiles, such as dissipative behaviors commonly encountered in real-world physical systems, e.g., storms gradually losing intensity in weather forecasting.

This paper provides a novel mathematical generalization of the SB to model non-conservative systems and develops near-linear time algorithms to realize the framework. We build on the geometric perspective of the SB, which casts it as a flow governed by Hamiltonian dynamics on the Wasserstein probability space (Sec. 3). Extending the Hamiltonian system to a contact Hamiltonian (Zadra, 2023), we propose a more general energy-varying formulation of the SB problem: **The Non-Conservative Generalized SB** (NCGSB, Sec. 4). To make computations tractable, we introduce the **Contact Wasserstein Geodesic** (CWG), which solves the NCGSB problem by casting it as a geodesic computation, reshaping its cost functional into a Riemannian metric whose induced distance is minimized. Discretizing the geodesic leads to geodesic segments that match standard residual blocks. We show how this leads to an efficient solver that avoids outer iteration loops and achieves near-linear complexity in both dimensionality and batch size. Additionally, our approach allows for metric modulation, enabling guided generation tailored to the task's specifications (Sec. 5). We demonstrate our approach on benchmarks and tasks such as LiDAR manifold navigation, molecular dynamics predictions, and image-based reconstruction of systems such as sea-surface temperature and robotic pick-and-place (Sec. 6).

In summary, **we contribute**: (1) a novel non-conservative formulation of the Schrödinger Bridge problem that models a wider range of real-world physical stochastic processes; (2) the introduction of the Contact Wasserstein Geodesic (CWG) framework, a general geometric solver compatible with all Schrödinger Bridge variants, enabling efficient and scalable computation; (3) a guided generation methodology based on the modulation of the metric associated with CWG.

2 Related Work

Schrödinger bridges. The SB problem imposes no constraints on the probabilistic path beyond matching the endpoint marginals, limiting its applicability when intermediate observations exist. It also does not permit including known physical laws governing the system's dynamics. The *multi-marginal Schrödinger Bridge* (mmSB) treats intermediate observations (Theodoropoulos et al., 2025) as constraints, which enables reconstruction of continuous dynamics without piecewise approximations. In contrast, the *Generalized Schrödinger Bridge* (GSB) adds a state cost, allowing potential energy functionals to be minimized along the probability path. This let us to model mean-field interactions (Gaitonde et al., 2021; Ruthotto et al., 2020), conservative forces (Philippidis et al., 1979; Noé et al., 2020), or geometric priors (Chen & Lipman, 2024; Liu et al., 2018).

Non-conservative Schrödinger bridge formulations. The SB problem assumes constant energy, preventing it from modeling non-conservative systems. The *momentum SB* augments the state space with velocity (Theodoropoulos et al., 2025; Chen et al., 2023), allowing damping to be incorporated (Blessing et al., 2025; Sterling et al., 2025). This, however, doubles the state space and increases computational cost. Other extensions replace the Brownian reference process with the *Ornstein–Uhlenbeck (OU)* process (Orland, 2025; Zhang & Stumpf, 2025), introducing a non-conservative prior for the target dynamics. The OU process defines a mean distribution with a curl-free component that drives convergence and a divergence-free component that induces rotation. Yet, this formulation lacks a mechanism for energy dissipation in the rotational dynamics. **We introduce** a more general energy-varying framework in which dissipation naturally emerges across all components of the dynamics, while only requiring a scalar augmentation of the state space.

Schrödinger bridge solvers. Matching-based iterative approaches (Shi et al., 2023; Gushchin et al., 2024; Peluchetti, 2023) have gained popularity by improving the scalability and robustness of traditional *Iterative Proportional Fitting (IPF)* algorithms (Kullback, 1968; Léonard, 2013) through Markovian projections, thereby avoiding the need for full trajectory storage. However, for GSB, restrictive assumptions like Gaussian probability paths limit expressivity (Liu et al., 2024; Tang et al., 2025), while for mmSB, global trajectory consistency is compromised due to the piecewise nature of the approach and its sensitivity to the initial choice of the reference process (Shen et al., 2025). To overcome these limitations, the stochastic dynamics can be learned indirectly by leveraging the analytical optimality conditions of the SB problem (Vargas et al., 2021; Chen et al., 2022). This has been shown to scale more effectively to both the GSB (Liu et al., 2022; Buzun et al., 2025) and the mmSB problem (Theodoropoulos et al., 2025; Hong et al., 2025; Chen et al., 2023), due to the symmetries inherent in these optimality conditions (App. A). However, these methods are limited by a classical SB solver bottleneck: the computational overhead of their iterative nature, which alternates between forward–backward passes or repeated dynamics integration and reference updates. **We propose** a cheaper non-iterative solver that scales nearly linearly with both dimensionality and sample size.

Schrödinger bridge guided generation. The SB framework naturally extends to conditional generation, where the marginals and the transition path depend on additional parameters or objectives (Shi et al., 2022). Model guidance techniques (Song et al., 2023; Guo et al., 2024) introduce a guidance term into the stochastic process, typically derived from the gradient of a loss function. This approach steers the flow locally and resembles a gradient-based form of optimal control. Alternatively, Raja et al. (2025) employs a global optimal control perspective. Their approach generates full trajectories and chooses the one minimizing a task-specific action functional. However, their method is deterministic and yields a single optimal path rather than a posterior distribution over paths. Unlike previous guidance methods, **we propose** a hybrid approach to guided generation within the NCGSB framework. By embedding a task-specific loss into the potential, we reshape the Riemannian metric so that the resulting geodesic reflects the guidance objective, allowing the learned dynamics to align with the desired outcome.

3 Preliminaries

The Wasserstein manifold. Before formally introducing the SB problem, we define its domain. Let $\mathcal{P}^+(\mathcal{M})$ denote the space of smooth, positive density functions supported on a manifold $\mathcal{M}$. Each element $\rho \in \mathcal{P}^+(\mathcal{M})$ is a function $\rho(x) : \mathcal{M} \to \mathbb{R}^+$ satisfying $\int_{\mathcal{M}} \rho(x)\, dx = 1$. The density dynamics is represented by a time-dependent family $\{\rho^t\}_{t\in\mathbb{R}^+} \subset \mathcal{P}^+(\mathcal{M})$. The infinitesimal variation of the density at time t is the time derivative $\partial_t \rho^t(x)$, which lies in the tangent space $\mathcal{T}_{\rho^t}\mathcal{P}^+(\mathcal{M})$. The collection of all such tangent spaces forms the tangent bundle $\mathcal{T}\mathcal{P}^+(\mathcal{M})$. When equipped with the Wasserstein metric, $\mathcal{P}^+(\mathcal{M})$ becomes a Riemannian manifold (Ambrosio et al., 2005). The corresponding metric tensor is defined as,

$$g^{\mathcal{W}_2}(\partial_t\rho^t, \partial_t\rho^t) = \int_{\mathcal{M}} \partial_t\rho^t(x)(-\Delta_{\rho^t})^\dagger \partial_t\rho^t(x)\, \rho^t(x)\, dx, \tag{1}$$

where $(-\Delta_{\rho^t})^\dagger$ is the inverse of the weighted Laplacian operator $\Delta_{\rho^t} = -\nabla_x \cdot (\rho^t\, \nabla_x)$ (Chow et al., 2020), inducing an inner product on $\mathcal{T}\mathcal{P}^+(\mathcal{M})$ and a distance $d_{\mathcal{W}_2}(\rho_a, \rho_b)$ for $\rho_a, \rho_b \in \mathcal{P}^+(\mathcal{M})$. The minimum-length curve ρ^t connecting the two distributions ρ_a, ρ_b is called a *geodesic*.

Multi-marginal generalized Schrödinger bridge (mmGSB). Given two endpoint densities $\rho_a, \rho_b \in \mathcal{P}^+(\mathcal{M})$, the SB problem (Schrödinger, 1931; Schrödinger, 1932) seeks the most probable interpolating density path ρ^t. This minimizes the Kullback–Leibler divergence w.r.t. a reference process ρ^t_{ref}, typically Brownian motion. The SB problem is equivalent to a stochastic optimal control setting (Dai Pra, 1991), which minimizes the cost required to transport a set of diffusing particles from an initial distribution ρ_a to a target distribution ρ_b. This dynamic reformulation of the OT problem (Benamou & Brenier, 2000; Chen et al., 2014) has solutions corresponding to geodesics on the Wasserstein manifold $\mathcal{P}^+(\mathcal{M})$. These trajectories are straight, since the classical SB problem assumes that particles dynamics are unaffected by external potential functions U. This assumption, however, limits our ability to model complex real-world physical systems.

Intermediate observations represented by marginal distributions $\{\rho_m\}_{m=1}^M$ at specific time steps $\{t_m\}_{m=1}^M$, can also be incorporated as additional constraints (Chen et al., 2023; Tang et al., 2025). This leads to the mmGSB problem,

$$\min_{v^t} J(v^t, \rho^t) = \int_0^1 \int_{\mathcal{M}} \left(\frac{1}{2}\|v^t(x)\|^2 + U(x)\right) \rho^t(x)\, dx\, dt; \tag{2a}$$

$$\text{s.t. } \partial_t\rho^t(x) + \nabla_x \cdot \big(\rho^t(x)\, v^t(x)\big) = \varepsilon\Delta_x\rho^t(x); \tag{2b}$$

$$\rho^0 = \rho_a,\ \rho^1 = \rho_b,\ \rho^{t_m} = \rho_m,\ \forall m \in 1, \ldots, M. \tag{2c}$$

Here, the density evolution ρ^t is governed by the Fokker–Planck equation (2b), which generalizes Brownian motion by incorporating a deterministic drift term v^t alongside a stochastic diffusion term scaled by ε, to satisfy the boundary conditions in equation (2c). This drift v^t acts as the control variable and ensures that the probability path interpolates between the given boundary marginals. The objective functional represents the kinetic energy associated with the drift and quantifies the deviation from the (uncontrolled) reference stochastic process.

Wasserstein Hamiltonian flows and geodesics. A convenient solution to the mmGSB problem (2) is to specify analytical optimality conditions (Sec. 2). These take the form of a Wasserstein Hamiltonian Flow (Chow et al., 2020), which describes a probability distribution evolving according to Hamiltonian dynamics. This evolution lies on planes tangent to $\mathcal{P}^+(\mathcal{M})$, specifically on the cotangent bundle $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$, the dual of $\mathcal{T}\mathcal{P}^+(\mathcal{M})$, and it is governed by the derivatives of a scalar Hamiltonian function H. However, their integration remains computationally expensive (Buzun et al., 2025; Wu et al., 2025), so we propose a geometric reformulation that results in a significant simplification. To this end, we introduce Proposition 1, a standard result from differential geometry (App. B.2), which is instrumental in lifting these equations to geodesics on $\mathcal{P}^+(\mathcal{M})$.

Proposition 1. *Let the optimality conditions of the mmGSB problem (2) be expressed in Hamiltonian form, yielding the optimal bridge $\rho^t(x)$. Then, $\rho^t(x)$ can be viewed as a geodesic connecting the marginals in equation 2c w.r.t. the modified Riemannian metric g_J, known as the Jacobi metric (Abraham & Marsden, 2008).*

To access the Jacobi metric and determine the corresponding geodesic, we first derive the Hamiltonian optimality conditions of the mmGSB problem (2) using Lagrange multipliers (Cui et al., 2024). This introduces a potential function $S^t(x)$, whose gradient defines the drift via $v^t(x) = \nabla_x S^t(x)$. The potential enforces the dynamic constraint (2b) within the cost functional (2a), whose first variation yields the Hamiltonian optimality conditions,

$$\partial_t \rho^t(x) = \partial_S H(\cdot) = -\nabla_x \cdot \big(\rho^t(x) \nabla_x S^t(x)\big); \tag{3a}$$

$$\partial_t S^t(x) = -\partial_\rho H(\cdot) = -\tfrac{1}{2}\|\nabla_x S^t(x)\|^2 + \tfrac{1}{2}\varepsilon^2 \partial_\rho I(\rho^t(x)) + U(x), \tag{3b}$$

with the corresponding Hamiltonian, $H(\rho^t, S^t) = \mathcal{K}(\rho^t, S^t) + \mathcal{F}(\rho^t)$, defined as the sum of a kinetic energy $\mathcal{K}$ and a potential energy $\mathcal{F} = -U - I$, dependent on the potential function U and the Fisher information I. A detailed derivation of these dynamics and the full Hamiltonian function is in App. C.1. To handle boundary conditions (2c), the Hamiltonian dynamics (3) are typically integrated backward in time, where the solution at each intermediate point (ρ^{t_m}, S^{t_m}) serves as the initial condition for the next segment (Theodoropoulos et al., 2025). The potential function $S^t(x)$ in equation 3a is linked to the infinitesimal density variation $\partial_t \rho^t \in \mathcal{T}\mathcal{P}^+(\mathcal{M})$ via the weighted Laplacian operator Δ_{ρ^t}, introduced through the Wasserstein metric (1). This connection establishes a correspondence between the tangent bundle $\mathcal{T}\mathcal{P}^+(\mathcal{M})$ and the cotangent bundle $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$, where $S^t(x)$ naturally resides, and where the Hamiltonian dynamics of (ρ^t, S^t) unfold.

By Proposition 1, the Hamiltonian dynamics (3) corresponds to a geodesic flow on the underlying Wasserstein manifold $\mathcal{P}^+(\mathcal{M})$, which minimizes the Jacobi metric $g_{\mathrm{J}} = (H - \mathcal{F})\, g^{\mathcal{W}_2}$. The original metric $g^{\mathcal{W}_2}$ (1) accounts only for the kinetic energy of the transport map $\mathcal{K}$ by measuring distances between distributions. In contrast, the Jacobi metric g_{J} also includes the potential energy $\mathcal{F}$, which is maximized to attain values $\mathcal{F} \approx H$. Consequently, computing the geodesic between marginals under this metric is equivalent to solving the mmGSB problem (2).

4 The Non-Conservative Generalized Schrödinger Bridge

Non-conservative formulation. The solution to the GSB problem (2) assumes a constant energy function H, and restricts the drift v^t to depend solely on the potential energy $\mathcal{F}$. This limits the model's flexibility in representing dynamics that cannot be described by a conservative potential, which reduces its ability to capture real-world processes involving energy dissipation and external interactions. To overcome this, we introduce the *non-conservative generalized Schrödinger bridge* (NCGSB), which allows for time-varying energy systems. To do so, we reformulate the cost functional J as the time integral of a new scalar state z^t, representing the *Lagrangian action*, whose evolution depends recursively on itself. The NCGSB problem is formulated as follows,

$$\min_{v^t} J(v^t, \rho^t) = \int_0^1 \partial_t z^t \, dt; \tag{4a}$$

$$\text{s.t. } \partial_t z^t = \int_{\mathcal{M}} \left(\frac{1}{2}\|v^t(x)\|^2 + U(x)\right) \rho^t(x)\, dx - z^t; \tag{4b}$$

$$\partial_t \rho^t(x) + \nabla_x \cdot \big(\rho^t(x)\, v^t(x)\big) = \varepsilon \Delta_x \rho^t(x); \tag{4c}$$

$$\rho^0 = \rho_a,\ \rho^1 = \rho_b,\ \rho^{t_m} = \rho_m,\ \forall m \in 1, \dots, M. \tag{4d}$$

The objective in equation 4a is no longer to minimize a static quantity, but rather a time-varying state z^t. Its dynamics (4b) depend explicitly on its current value. This recursive structure endows the system with a form of memory, as its evolution is influenced by the entire trajectory, implicitly encoded in z^t. Because non-conservative forces are path-dependent, augmenting the system's state space with the scalar z^t allows their effects to be modeled, enabling the system's energy to vary over time. By relaxing the implicit energy-conservation constraint of the GSB problem, our approach enhances the model's flexibility and improves the quality of the resulting optimal solution.

Guided NCGSB. NCGSB (4) can be extended to the guided generation setting by introducing a guiding function f, which steers the generative process toward desired conditions at any chosen time (Song et al., 2023; Guo et al., 2024). For a given time t_s, the guidance is expressed as $y = f(x^{t_s})$, with $x^{t_s} \sim \rho^{t_s}$. To enforce this form, the bridge ρ^t is steered according to $\rho^t(x|y) = \frac{1}{Z}\,\rho^t(x)\,e^{-\|y-f(\bar{x}^{t_s})\|^2}$, where Z is a normalization constant and $\bar{x}^{t_s}$ denotes a sample from the predicted guided distribution $\bar{\rho}^{t_s}$ conditioned on the current distribution ρ^t. By Bayes' rule, the dynamics of the guided bridge $\rho^t(x|y)$ acquire an additional guidance term via the drift v^t, determined by $\|y - f(\bar{x}^{t_s})\|^2$. To perform a guided generation that enforces the constraint $y = f(x^{t_s})$ while preserving the underlying data manifold, we incorporate y into the Lagrangian action constraint (4b) as (see App. C.3 for details),

$$\partial_t z^t = \int_{\mathcal{M}} \left(\frac{1}{2}\|v^t(x)\|^2 + U(x) + \|y - f(\bar{x}^{t_s})\|^2 \right) \rho^t(x)\,dx \;-\; z^t. \tag{5}$$

Wasserstein contact Hamiltonian flows and geodesics. Analogously to mmGSB (2), understanding the dynamics of the optimality conditions in NCGSB (4) is essential for reformulating it as a geodesic computation. As detailed in App. C.2, we propose to leverage the contact Hamiltonian formalism (Kholodenko, 2013), an extension of classical Hamiltonian mechanics to non-conservative systems (App. B.1), to model the dynamics of the NCGSB optimality conditions as Wasserstein contact Hamiltonian flows. This generalizes Prop. 1, since the contact Hamiltonian dynamics defines a geodesic but on the extended space $\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$ (Udrişte, 2000; Testa et al., 2025). The contact Hamiltonian optimality conditions are,

$$\partial_t \rho^t(x) = \partial_S H(\cdot) = -2\nabla_x \cdot \big(\rho^t(x)\nabla_x S^t(x)\big), \tag{6a}$$

$$\begin{aligned}\partial_t S^t(x) = \partial_\rho H(\cdot) - S^t(x)\partial_z H(\cdot) = -\tfrac{1}{2}\|\nabla_x S^t(x)\|^2 + \tfrac{1}{2}\varepsilon^2 \partial_\rho I\big(\rho^t(x)\big) + U(x)&\\ + z^t + 2\varepsilon \log 2\rho^t,&\end{aligned} \tag{6b}$$

$$\begin{aligned}\partial_t z^t = S^t(x)\partial_S H(\cdot) - H(\cdot) = \int_{\mathcal{M}} \left(\frac{1}{2}\|\nabla_x S^t(x)\|^2 + U(x) \right) \rho^t(x)\,dx + \frac{1}{2}\varepsilon^2 I(\rho^t)&\\ - 2\int_{\mathcal{M}} \varepsilon\big(\log 2\rho^t(x) - 1\big)\rho^t(x)\,dx - z^t.&\end{aligned} \tag{6c}$$

The corresponding contact Hamiltonian function is defined as, $H(\rho^t, S^t, z^t) = \mathcal{K}(\rho^t, S^t) + \mathcal{F}(\rho^t) + \mathcal{B}(\rho^t) + z^t$. This differs from its conservative counterpart in two ways. First, its explicit dependence on z^t allows the total energy to vary over time. Second, the potential energy is augmented by an entropy term, $\mathcal{B}(\rho^t) = 2\int_{\mathcal{M}} \varepsilon(\log 2\rho^t(x) - 1)\rho^t(x)\,dx$, producing an additional diffusion in the dynamics. As previously mentioned, for guided generation, an additional potential energy term $\|y - f(\bar{x}^{t_s})\|^2$ can here be introduced to steer the flow. Geometrically, the dynamics of (ρ^t, S^t, z^t) can be interpreted as a flow on the cotangent bundle of the Wasserstein manifold, augmented by the scalar state z^t. That is, the dynamics unfold on the space $\mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$.

The contact Hamiltonian flow evolving on the extended phase space $\mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$, and interpolating between the marginal densities, induces a geodesic on the augmented manifold $\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$. This geodesic minimizes a Jacobi metric $\tilde{g}_{\mathrm{J}} = (H - \mathcal{F} - \mathcal{B})\, g^{\mathcal{W}_2}$, which generalizes the classical Wasserstein metric by incorporating the potential energy of the contact Hamiltonian function. Computing the geodesic under the Jacobi metric $\tilde{g}_{\mathrm{J}}$ corresponds to NCGSB (4). Unlike the conservative case, the contact Hamiltonian H is no longer constant along the flow, allowing the total energy to vary over time. This introduces an additional degree of freedom that can be leveraged to shape the system's energy along the path over $\mathcal{P}^+(\mathcal{M})$. This is the reason that the geodesic (ρ^t, H^t) is defined on the extended space $\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$.

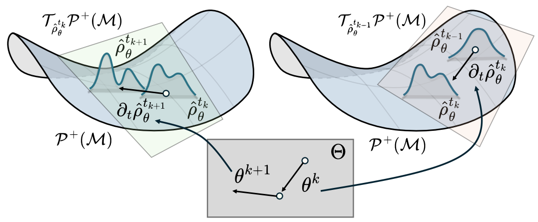

Figure 2: Visualization of the ResNet transformation. Two successive pushforwards $\rho_\theta^{t_{k-1}} \to \rho_\theta^{t_k} \to \rho_\theta^{t_{k+1}}$ on $\mathcal{P}^+(\mathcal{M})$ are shown as local updates $\partial_t \rho_\theta^{t_k}$, $\partial_t \rho_\theta^{t_{k+1}}$ on tangent spaces. Each update is parameterized by $\theta^k, \theta^{k+1} \in \Theta$, defining local coordinates on $\mathcal{TP}^+(\mathcal{M})$. This coordinate system is not unique.

5 Contact Wasserstein Geodesics (CWG)

ResNet resembles a discrete geodesic. Our objective is to compute a geodesic ρ^t on $\mathcal{P}^+(\mathcal{M})$, induced by the contact Hamiltonian dynamics, that is constrained to pass through a set of observed marginals $\{\rho_a, \rho_m, \rho_b\}$ (i.e., discretized distributions along the probability path). These constraints naturally lead to a discretized parameterization of ρ^t, where the overall density transformation is modeled as a composition of maps, each connecting a pair of consecutive observations. A ResNet is ideally suited for this problem, as its sequential block structure directly mirrors this piecewise, compositional nature of the approximated geodesic. Let λ be a fixed reference measure on $\mathcal{P}^+(\mathcal{M})$ (e.g., a standard Gaussian or uniform distribution). We define a $(K+1)$-block ResNet as follows,

$$T_{\{\theta^k\}_{k=0}^K} = T_{\theta^K} \circ \cdots \circ T_{\theta^1} \circ T_{\theta^0}, \tag{7}$$

with parameters $\{\theta^k\}_{k=0}^K \in \Theta$. The process begins by sampling an initial batch of points $x^s \sim \lambda$, that is pushed forward through the first block to obtain $x^{t_0} = T_{\theta^0}(x^s)$. Then, the parameters θ^0 are optimized such that the resulting pushforward reference measure approximates the initial marginal $\rho_\theta^{t_0} \approx \rho_a$. Thereafter, each subsequent block k pushes forward the sample via $x^{t_{k+1}} = T_{\theta^{k+1}}(x^{t_k}),\ x^{t_k} \sim \rho_\theta^{t_k}$. The full pushforward map induces,

$$\rho_\theta^{t_{k+1}} = (T_{\theta^{k+1}})_\# \rho_\theta^{t_k} = \rho_\theta^{t_k}\big(T^{-1}_{\theta^{k+1}}(x^{k+1})\big)\ \det\big[\nabla_x T^{-1}_{\theta^{k+1}}(x^{k+1})\big]. \tag{8}$$

Starting from the reference measure λ, the ResNet parameters $\{\theta^k\}_{k=0}^K$ define a sequence of discrete probability transitions $\{\partial_t \rho_\theta^{t_k}\}_{k=0}^K$, which in turn specify the discrete family of densities $\{\rho_\theta^{t_k}\}_{k=0}^K$. Geometrically, the discretizations $\{\partial_t \rho_\theta^{t_k}, \rho_\theta^{t_k}\}_{k=0}^K$, provided by the ResNet, approximate $(\rho^t, \partial_t \rho^t) \in \mathcal{TP}^+(\mathcal{M})$, which can be seen as inducing a mapping from $\mathcal{TP}^+(\mathcal{M})$ onto the parameter space Θ (Fig. 5). As stated in Proposition 2, the existence of such a map endows the finite-dimensional space Σ, where the parameterized densities $\rho_\theta^{t_k}$ reside, with the geometric properties of $\mathcal{P}^+(\mathcal{M})$. This lifting enables faster and tractable computations for SB problems.

Proposition 2. *Approximate the evolution of the density $\rho^t \in \mathcal{P}^+(\mathcal{M})$ by a series of K smooth parametrized pushforwards T_{θ^k}, with θ^k belonging to a finite-dimensional space Θ. If each pushforward T_{θ^k} is an immersion $T_{\theta^k} : \Theta \to \mathcal{TP}^+(\mathcal{M})$, then the parameter space Θ can be endowed with a Riemannian structure via the pullback of the Wasserstein metric $g^{\mathcal{W}_2}$. Consequently, the contact Hamiltonian dynamics on $\mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$ can be represented in the reduced phase space $\Theta^* \times \mathbb{R}$, with the associated geodesic on $\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$ projected onto $\Sigma \times \mathbb{R}$ (see App. D.1).*

Proposition 2 allows us to transform the geodesic computation from the infinite-dimensional $\mathcal{P}^+(\mathcal{M})$ to a geodesic on a finite-dimensional parameterized space Σ , such that the resulting geodesic flow on $\Sigma \times \mathbb{R}$ evolves under the pullback of the Jacobi metric $T_\theta^* \tilde{g}_{\mathrm{J}} = \Phi^{t_k} T_\theta^* g^{\mathcal{W}_2}$, where the scalar factor $\Phi^{t_k} = H(\rho^{t_k}, S^{t_k}, z^{t_k}) - \mathcal{F}(\rho^{t_k}) - \mathcal{B}(\rho^{t_k})$, encodes the potential energy. Specifying the time evolution of H^{t_k} determines a unique parameterized bridge on Σ. This formulation enables a tractable computation of geodesic flows to solve the NCGSB problem. Although different parameterizations $\{\theta^k\}_{k=0}^K$ may define distinct coordinate systems on Σ, the geodesics solutions remain equivalent and share the same length (Syrota et al., 2025).

Geodesic computation. The *contact Wasserstein geodesic* (CWG) corresponds to the discrete path $\{\rho_\theta^{t_k}\}_{k=0}^K$, that approximates the NCGSB solution. This is trained to reconstruct the available marginals while minimizing the geodesic energy under the pullback Jacobi metric $T_\theta^* \tilde{g}_{\mathrm{J}} = \Phi^{t_k} T_\theta^* g^{\mathcal{W}_2}$. The initial and final marginals, ρ_a and ρ_b, are enforced at the path endpoints, corresponding to the ResNet outputs at times t_0 and t_K. All available intermediate marginals ρ^m must appear at time points matching the ResNet discretization for the condition to be enforced.

CWG training happens in two stages (Alg. 1 in App. D.2): (1) we optimize the first ResNet block to match the initial marginal ρ_a, and (2) we find the optimal path by minimizing the loss,

$$\ell = \underbrace{d^2_{\mathcal{W}_2}(\rho_\theta^{t_K}, \rho_b)}_{\text{Terminal marginal}} + \underbrace{\sum_{m=1}^{M} d^2_{\mathcal{W}_2}(\rho_\theta^{t_{k_m}}, \rho_m)}_{\text{Intermediate marginals}} + \underbrace{\sum_{k=1}^{K} \Phi^{t_k}\, d^2_{\mathcal{W}_2}(\rho_\theta^{t_k}, \rho_\theta^{t_{k-1}})}_{\text{Energy minimization}} . \tag{9}$$

Here $d_{\mathcal{W}_2}$ denotes the Wasserstein-2 distance between probability distributions. In practice, this distance is approximated using empirical estimators based on samples drawn from the distributions. The complexity of this method is $\mathcal{O}\big(NK(T_{\text{sh}} + d(LW + \log N))\big)$, scaling linearly in dimension d and nearly linearly in batch size N (see App. D.3), rather than exponentially and quadratically (Hong et al., 2025). Unlike Chen et al. (2023); Shen et al. (2025), our CWG avoids costly iteration loops and is only weakly affected by the number of marginals.

Guided contact Wasserstein geodesics. In the conditional setting, the Lagrangian action dynamics (4b) in NCGSB (4) is augmented as in equation 5. Here, the scaling factor $\Phi^{t_k} = H(\rho^{t_k}, S^{t_k}, z^{t_k}) - \mathcal{F}(\rho^{t_k}) - \mathcal{B}(\rho^{t_k})$ of the pullback Jacobi metric is augmented with the guidance term $\|y - f(\bar{x}^{t_s})\|^2$, to enforce the constraint $y = f(\bar{x}^{t_s})$ at time t_s of the generative process. Under the ResNet parameterization, the desired distribution is approximated by $x^{t_{k_s}} \approx \bar{x}^{t_s}$ at time step t_{k_s}, and the Jacobi metric is modified as $\tilde{g}'_{\text{J}} = \big(\Phi^{t_k} + f(x^{t_{k_s}})\big)\, g^{\mathcal{W}_2}$, with $x^{t_{k_s}} \sim \rho_\theta^{t_{k_s}}$. This penalizes geodesics crossing undesired regions at t_{k_s}. The loss for the guided optimization is

$$\ell = d^2_{\mathcal{W}_2}(\rho_\theta^{t_K}, \rho_b) + \sum_{m=1}^{M} d^2_{\mathcal{W}_2}(\rho_\theta^{t_{k_m}}, \rho_m) + \sum_{k=1}^{K} \big(\Phi^{t_k} + f(x^{t_{k_s}})\big)\, d^2_{\mathcal{W}_2}(\rho_\theta^{t_k}, \rho_\theta^{t_{k-1}}) + d'^2_{\mathcal{W}_2}(\rho_\theta^{t_{t_s}}, \rho_s) \tag{10}$$

where the modified distance $d'^2_{\mathcal{W}_2}$ measures deviations between the generated distribution $\rho_\theta^{t_{t_s}}$ and the intermediate marginal ρ_s at t_s, while incorporating the penalty for samples x_s that violate the guidance constraint $y = f(x_s)$, c.f. App. E.1. In practice, this loss is optimized through a fine-tuning procedure applied to a model initially trained without any guidance.

Proof of concept. We demonstrate our framework on a 2D distribution-matching task and guided generation setting using the Two-Moons and Checkerboard benchmarks (Holderrieth & Erives, 2025). These lack intermediate marginals $\{\rho_m\}_{m=1}^{M}$, and the initial distribution ρ_a coincides with the reference distribution λ. Hence, only the second step of Alg. 1 is needed. Figure 3 shows that our method successfully generates the target distributions, and steers the generation to samples confined to a subset of the target space (here, the upper half). This guided behavior is achieved via the term $\|y - f(\bar{x}^{t_s})\|^2$, with $t_s = 1$, f measuring 2D sample positions, and y defining the admissible region.

6 Results

We benchmark our approach against four established baselines summarized in Table 1. Further details of the experimental setups are provided in App. F.1.

LiDAR manifold navigation. First, we tackle a standard GSB task: computing a bridge evolving on a geometric manifold. We use the LiDAR scan of Mount Rainier (OpenTopography, 2025) as the reference surface, and we aim to connect two marginals while remaining on the manifold and favoring low-altitude regions. These conditions are encoded into the potential function U (see App. F.2). In this experiment, we do not model a physical system but instead compute an optimal

Method	GSB	mmSB	Energy variation	Image Gen.	Guided Gen.
DSBM (Shi et al., 2023)	✗	✗	✗	✓	✗
GSBM (Liu et al., 2024)	✓	✗	✗	✓	✗
SBIRR (Shen et al., 2025)	✗	✓	✗	✗	✗
DM-SB (Chen et al., 2023)	✗	✓	✗	✗	✗
CWG (ours)	✓	✓	✓	✓	✓

Table 1: Comparison of our CWG with baselines designed to address various SB variants and types of problems.

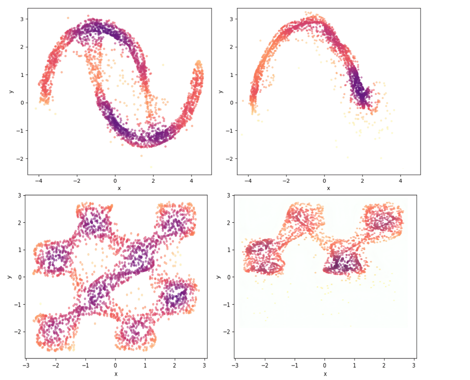

Figure 3: Two-Moons (top) and Checkerboard (bottom) benchmarks with guided variants (right).

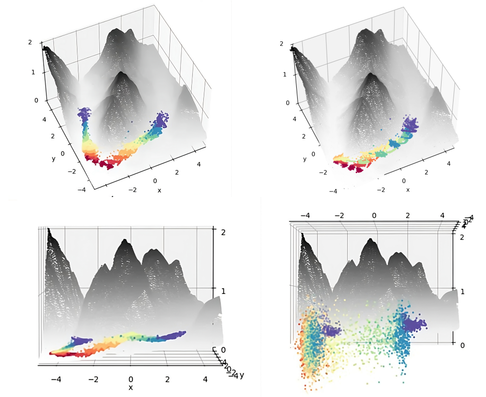

Figure 4: LiDAR Manifold Navigation: CWG before and after guidance (top), CWG vs GSBM (bottom).

transport map between the marginals under a conservative setting. Unlike our approach, DSBM and GSBM iteratively fit a deterministic path between the marginals, falling short on representing a posterior distribution. As a result, there is no guarantee that a Gaussian path remains on the manifold (Fig. 4). This leads to substantially higher-energy paths (Table 2), while *our approach finds lower-energy solutions and converges significantly faster*. Furthermore, our method uniquely supports guided generation, illustrated here by steering the probabilistic path to the right side of the mount (Fig. 4; see App. F.2 for quantitative results).

Table 2: Bridge energy J (↓) and training time (tt) (↓) in LiDAR Manifold Navigation.

Metric	CWG (ours)	GSBM	DSBM
J	$\mathbf{1.95}_{\pm\mathbf{0.07}}$	$4.74_{\pm 0.10}$	$17.29_{\pm 0.14}$
tt (s)	$\mathbf{280}_{\pm\mathbf{20}}$	$1570_{\pm 50}$	$1340_{\pm 50}$

Table 3: Wasserstein error at validation (↓) and training time (tt) (↓) in Single Cell Sequencing.

Metric	CWG (ours)	DM-SB	SBIRR
$d_{\mathcal{W}_2}(x^{t_1})$	$\mathbf{1.11}_{\pm\mathbf{0.06}}$	$2.25_{\pm 0.01}$	$1.92_{\pm 0.02}$
$d_{\mathcal{W}_2}(x^{t_3})$	$\mathbf{0.33}_{\pm\mathbf{0.02}}$	$1.64_{\pm 0.03}$	$1.86_{\pm 0.02}$
tt (s)	$\mathbf{710}_{\pm\mathbf{30}}$	$38120_{\pm 1100}$	$1740_{\pm 40}$

Table 4: FID scores at validation steps (↓) and training time (tt) (↓) in Sea Prediction (2020-2024).

Metric	CWG (ours)	DM-SB	SBIRR
FID(x^{t_1})	$\mathbf{121.5}_{\pm\mathbf{5.6}}$	$160.7_{\pm 4.5}$	$242.3_{\pm 9.9}$
FID(x^{t_3})	$\mathbf{159.5}_{\pm\mathbf{7.4}}$	$185.5_{\pm 7.1}$	$235.8_{\pm 10.4}$
tt (s)	$\mathbf{1030}_{\pm\mathbf{50}}$	$73600_{\pm 3200}$	$19100_{\pm 900}$

Table 5: FID score (↓) and training time (tt) (↓) in Robotic Task Reconstruction.

Metric	CWG (ours)	GSBM	DSBM
FID	$\mathbf{18.83}_{\pm\mathbf{0.66}}$	$40.23_{\pm 1.95}$	$149.78_{\pm 0.81}$
tt (s)	$\mathbf{1090}_{\pm\mathbf{40}}$	$91100_{\pm 8000}$	$27400_{\pm 2500}$

Single cell sequencing. Next, we reconstruct stem cell differentiation dynamics from a series of isolated cellular snapshots. We use the Embryoid Body (EB) dataset from Moon et al. (2019), which tracks cell state progression across five developmental stages $[t_0, t_1, t_2, t_3, t_4]$. Cell differentiation is fundamentally a non-conservative biological process (Zeevaert et al., 2020; Kinney et al., 2014) and the ability to model energy-varying bridges is essential. To evaluate generalization in regions with no available data, we split the dataset into a training set $[t_0, t_2, t_4]$ and a validation set $[t_1, t_3]$. Accordingly, the former contains the distributions $\{\rho_a, \rho_{m_2}, \rho_b\}$, while the latter contains $\{\rho_{m_1}, \rho_{m_3}\}$. The geometry of the training distributions is encoded in the potential function U, which penalizes paths that stray from the observed data manifold. Minimizing U ensures the learned bridge remains close to the data manifold, enabling effective generalization. The combination of the data manifold guidance and an energy-varying bridge allows our approach to outperform other mmSB baselines in both reconstruction accuracy and computation time. Quantitative results are reported in Table 3, with additional details and an ablation study on the importance of energy variation provided in App. F.3.

Image generation. We also demonstrate our framework's applicability to image generation tasks. Given a sequence of images capturing the time evolution of physical phenomena, the model's objective is to predict realistic intermediate frames at unobserved time steps. To ensure the generated frames remain faithful to the underlying data distribution, we introduce a potential function U, that penalizes deviations from the learned data manifold. Specifically, for samples $x^t \sim \rho^t$, $U(x^t)$ is defined as the reconstruction error, obtained via a VAE (Song & Itti, 2025). Details on the energy behavior and extended results are provided in App. F.4 and App. F.5.

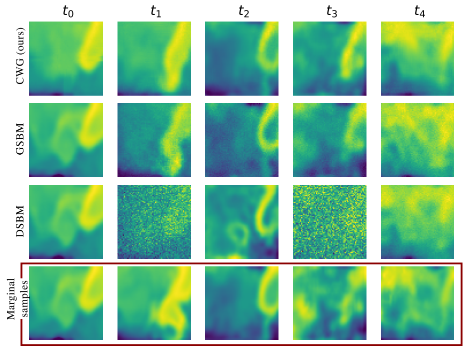

Figure 5: Predictions from CWG (ours, top), GSBM (middle), and DSBM (bottom). The red row shows marginal samples.

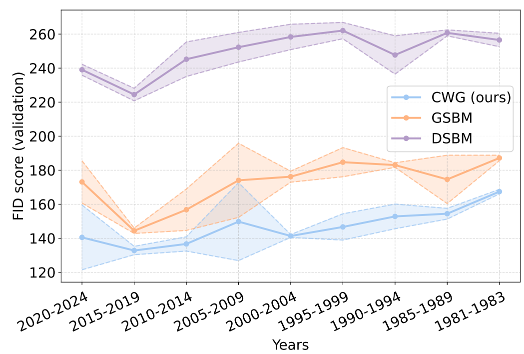

Figure 6: FID scores at the validation time steps for three methods, evaluated over 9 tests from 1981–2024. Our CWG is significantly lower than the baselines.

Specifically, we use the NOAA OISST v2 High Resolution Dataset (Huang et al., 2021), which provides daily sea surface temperature averages over multiple years, and the BridgeData V2 (Walke et al., 2023) dataset, for robotic manipulation tasks. For sea temperature prediction, we group data from 1981–2024 into five-year intervals. Using heatmaps from January, May, and September (i.e., $[t_0, t_2, t_4]$), our method predicts the temperature profiles of March and July (i.e., $[t_1, t_3]$). Our CWG produces cleaner, more accurate predictions than the baselines (Fig. 5). Since our framework operates efficiently in probability space and is not constrained by energy conservation, it achieves these results with an order of magnitude less computation (Table 4 for 2020-2024; Fig. 6 for all years).

In the robotic task reconstruction, our model generates realistic intermediate frames connecting the initial and final states of a robot's reaching motion (Fig. 7), and demonstrates consistently robust performance, outperforming baselines in image quality (Table 5). Moreover, Fig. 8 showcases guided generation, where our model successfully steers the placing motion task toward a target location on the left side of the table. This is achieved with only a minimal drop in image quality, maintaining a clear advantage over competing methods (Table 6).

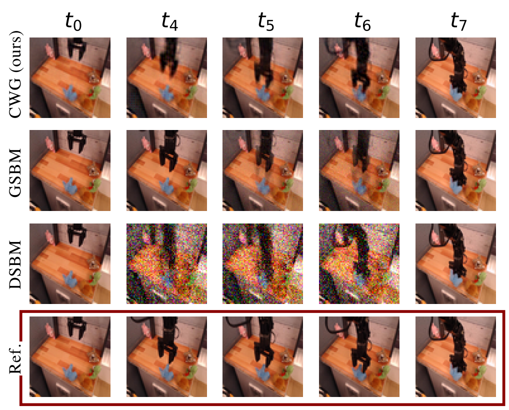

Figure 7: Reconstructions from CWG (top), GSBM (middle), and DSBM (bottom). Red row shows the reference.

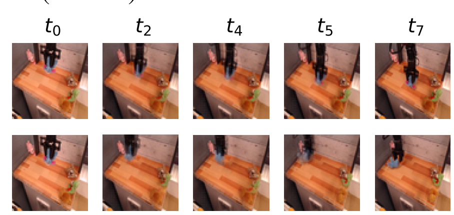

Figure 8: CWG outputs before (top) vs. after guidance (place the item left).

Metric	Standard	Guidance
Centroid	$35.8_{\pm 11.1}$	$22.3_{\pm 2.9}$
FID	$19.52_{\pm 0.78}$	$23.77_{\pm 1.94}$

Table 6: Item centroid position (px) and FID before vs. after guidance.

7 Conclusion

Our work is motivated by the need to model intermediate time steps of Schrödinger bridges (SBs), arising from the underlying dynamics of the observed physical system. As standard SBs conserve energy across time, they cannot meaningfully encode such dynamics. To counter this, we introduced the *non-conservative generalized Schrödinger bridge* (NCGSB), which extends the usual Hamiltonian to its non-conservative counterpart, the *contact Hamiltonian*, allowing energy to vary. We show that NCGSB is equivalent to geodesics on *contact Wasserstein manifolds*. This link leads to a non-iterative and near-linear time algorithm for computing the non-conservative bridge, which can practically be realized by a ResNet-like construction, easing its implementation. We show that these theoretical contributions lead to a SB framework that is not only more expressive but also significantly faster than existing approaches, as validated by the significant improvements achieved across a range of diverse tasks.

REFERENCES

Ralph Abraham and Jerrold E Marsden. *Foundations of mechanics*. Number 364. American Mathematical Soc., 2008.

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005. URL https://doi.org/10.1007/b137080.

Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. Latent space oddity: on the curvature of deep generative models. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=SJzRZ-WCZ.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000. URL https://doi.org/10.1007/s002110050002.

Quentin Bertrand, Joey Bose, Alexandre Duplessis, Marco Jiralerspong, and Gauthier Gidel. On the stability of iterative retraining of generative models on their own data. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=JORAfH2xFd.

Denis Blessing, Julius Berner, Lorenz Richter, and Gerhard Neumann. Underdamped diffusion bridges with applications to sampling. In *International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=Q1QTxFm0Is.

Christian Blohmann. Lagrangian field theory. 2024. Lecture Notes.

Roberto Borelli, Agostino Dovier, and Federico Fogolari. Data structures and algorithms for k-th nearest neighbours conformational entropy estimation. *Biophysica*, 2(4):340–352, 2022. ISSN 2673-4125. doi: 10.3390/biophysica2040031. URL https://www.mdpi.com/2673-4125/2/4/31.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. URL https://openreview.net/forum?id=9BnCwiXB0ty.

Alessandro Bravetti, Hans Cruz, and Diego Tapias. Contact Hamiltonian mechanics. *Annals of Physics*, 376:17–39, 2017.

Nazar Buzun, Daniil Shlenskii, Maxim Bobrin, and Dmitry V. Dylov. Hota: Hamiltonian framework for optimal transport advection, 2025. URL https://arxiv.org/abs/2507.17513.

Minshuo Chen, Song Mei, Jianqing Fan, and Mengdi Wang. An overview of diffusion models: Applications, guided generation, statistical rates and optimization. *arXiv preprint arXiv:2404.07771*, 2024. URL https://arxiv.org/abs/2404.07771.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=g7ohDlTITL.

Tianrong Chen, Guan-Horng Liu, and Evangelos Theodorou. Likelihood training of Schrödinger bridge using forward-backward SDEs theory. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=nioAdKCEdXB.

Tianrong Chen, Guan-Horng Liu, Molei Tao, and Evangelos Theodorou. Deep momentum multi-marginal Schrödinger bridge. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2023. URL https://openreview.net/forum?id=ykvvv0gc4R.

Xiwen Chen, Wenhui Zhu, Peijie Qiu, Hao Wang, Huayu Li, ZIHAN LI, Yalin Wang, Aristeidis Sotiras, and Abolfazl Razi. FIC-TSC: Learning time series classification with Fisher information constraint. In *International Conference on Machine Learning (ICML)*, 2025. URL https://openreview.net/forum?id=Dr8msCnFYw.

Yongxin Chen, Tryphon Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint, 2014. URL https://arxiv.org/abs/1412.4430.

Shui-Nee Chow, Wuchen Li, and Haomin Zhou. Wasserstein Hamiltonian flows. *Journal of Differential Equations*, 268(3):1205–1219, 2020. URL https://doi.org/10.1016/j.jde.2019.08.046.

Jianbo Cui, Shu Liu, and Haomin Zhou. Stochastic Wasserstein Hamiltonian flows. *Journal of Dynamics and Differential Equations*, 36(4):3885–3921, 2024. URL https://doi.org/10.1007/s10884-023-10264-4.

Paolo Dai Pra. A stochastic control approach to reciprocal diffusion processes. *Applied mathematics and Optimization*, 23(1):313–329, 1991. URL https://doi.org/10.1007/BF01442404.

M. do Carmo. *Riemannian Geometry*. Birkhäuser Basel, 1992.

Jean Feydy. Geometric data analysis, beyond convolutions. *Applied Mathematics*, 3, 2020.

Christian L. E. Franzke, Terence J. O'Kane, Judith Berner, Paul D. Williams, and Valerio Lucarini. Stochastic climate theory and modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 6(1): 63–78, 2015. URL https://doi.org/10.1002/wcc.318.

Jason Gaitonde, Jon Kleinberg, and Éva Tardos. Polarization in geometric opinion dynamics. In *ACM Conference on Economics and Computation (EC)*, pp. 499–519, 2021. URL https://doi.org/10.1145/3465456.3467633.

Hansjörg Geiges. A brief history of contact geometry and topology. *Expositiones Mathematicae*, 19(1):25–53, 2001. ISSN 0723-0869. URL https://doi.org/10.1016/S0723-0869(01)80014-1.

Hansjörg Geiges. An introduction to contact topology. *Cambridge Studies in Advanced Mathematics*, 109, 2008.

Yingqing Guo, Hui Yuan, Yukang Yang, Minshuo Chen, and Mengdi Wang. Gradient guidance for diffusion models: An optimization perspective. In *Neural Information Processing Systems (NeurIPS)*, 2024. URL https://openreview.net/forum?id=X1QeUYBXke.

Nikita Gushchin, Sergei Kholkin, Evgeny Burnaev, and Alexander Korotin. Light and optimal Schrödinger bridge matching. In *International Conference on Machine Learning (ICML)*, pp. 17100–17122. PMLR, 2024. URL https://openreview.net/forum?id=EWJn6hfZ4J.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems (NeurIPS)*, 33:6840–6851, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

Peter Holderrieth and Ezra Erives. Introduction to flow matching and diffusion models, 2025. URL https://diffusion.csail.mit.edu/.

Wanli Hong, Yuliang Shi, and Jonathan Niles-Weed. Trajectory inference with smooth Schrödinger bridges. In *International Conference on Machine Learning (ICML)*, 2025. URL https://openreview.net/forum?id=GHyvvWu1XC.

Boyin Huang, Chunying Liu, Viva Banzon, Eric Freeman, Garrett Graham, Bill Hankins, Tom Smith, and Huai-Min Zhang. Improvements of the daily optimum interpolation sea surface temperature (doisst) version 2.1. *Journal of Climate*, 34(8):2923–2939, 2021.

Hongbin Huang, Minghua Chen, and Xiao Qiao. Generative learning for financial time series with irregular and scale-invariant patterns. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=CdjnzWsQax.

Rytis Kazakevičius, Aleksejus Kononovicius, Bronislovas Kaulakys, and Vygintas Gontis. Understanding the nature of the long-range memory phenomenon in socioeconomic systems. *Entropy*, 23(9):1125, 2021. URL https://doi.org/10.3390/e23091125.

Arkady L Kholodenko. *Applications of contact geometry and topology in physics*. World Scientific, 2013.

Melissa A Kinney, Rabbia Saeed, and Todd C McDevitt. Mesenchymal morphogenesis of embryonic stem cells dynamically modulates the biophysical microtissue niche. *Scientific reports*, 4(1):4290, 2014. URL https://doi.org/10.1038/srep04290.

Solomon Kullback. Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39(4):1236–1243, 1968. URL https://www.jstor.org/stable/2239692.

Flavien Léger and Wuchen Li. Hopf–cole transformation via generalized Schrödinger bridge problem. *Journal of Differential Equations*, 274:788–827, 2021. URL https://doi.org/10.1016/j.jde.2020.10.029.

David Levin. The approximation power of moving least-squares. *Mathematics of computation*, 67 (224):1517–1531, 1998.

Guan-Horng Liu, Tianrong Chen, Oswin So, and Evangelos Theodorou. Deep generalized Schrödinger bridge. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. URL https://openreview.net/forum?id=fp33Nsh0O5.

Guan-Horng Liu, Yaron Lipman, Maximilian Nickel, Brian Karrer, Evangelos Theodorou, and Ricky T. Q. Chen. Generalized Schrödinger bridge matching. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=SoismgeX7z.

Zhiyu Liu, Bo Wu, and Hai Lin. A mean field game approach to swarming robots control. In *American Control Conference (ACC)*, pp. 4293–4298, 2018. URL https://doi.org/10.23919/ACC.2018.8431807.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport, 2013. URL https://arxiv.org/abs/1308.0215.

Kevin R Moon, David Van Dijk, Zheng Wang, Scott Gigante, Daniel B Burkhardt, William S Chen, Kristina Yim, Antonia van den Elzen, Matthew J Hirn, Ronald R Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37(12):1482–1492, 2019. URL https://doi.org/10.1038/s41587-019-0336-3.

Frank Noé, Alexandre Tkatchenko, Klaus-Robert Müller, and Cecilia Clementi. Machine learning for molecular simulation. *Annual review of physical chemistry*, 71(1):361–390, 2020. URL https://doi.org/10.1146/annurev-physchem-042018-052331.

OpenTopography. Southwest flank of mt. rainier, wa, 2025. URL https://doi.org/10.5069/G9PZ56R1. Distributed by OpenTopography. Accessed 2025-08-11.

Henri Orland. Schrödinger bridges for systems of interacting particles. *Physical Review E*, 112(1): 014115, 2025.

Stefano Peluchetti. Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling. *Journal of Machine Learning Research (JMLR)*, 24(374):1–51, 2023. URL https://www.jmlr.org/papers/v24/23-0527.html.

Chris Philippidis, Chris Dewdney, and Basil J Hiley. Quantum interference and the quantum potential. *Il Nuovo Cimento B (1971-1996)*, 52(1):15–28, 1979.

Sanjeev Raja, Martin Šípka, Michael Psenka, Tobias Kreiman, Michal Pavelka, and Aditi S Krishnapriyan. Action-minimization meets generative modeling: Efficient transition path sampling with the Onsager-Machlup functional. In *International Conference on Machine Learning (ICML)*, 2025. URL https://openreview.net/forum?id=QwoGfQzuMa.

Lars Ruthotto, Stanley J Osher, Wuchen Li, Levon Nurbekyan, and Samy Wu Fung. A machine learning framework for solving high-dimensional mean field game and mean field control problems. *Proceedings of the National Academy of Sciences*, 117(17):9183–9193, 2020. URL https://doi.org/10.1073/pnas.1922204117.

Erwin Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. In *Annales de l'institut Henri Poincaré*, volume 2, pp. 269–310, 1932. URL http://eudml.org/doc/78968.

Erwin Schrödinger. Über die umkehrung der naturgesetze. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse*, pp. 144–153, 1931. URL https://doi.org/10.1002/ange.19310443014.

Yunyi Shen, Renato Berlinghieri, and Tamara Broderick. Multi-marginal Schrödinger bridges with iterative reference refinement. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2025. URL https://openreview.net/forum?id=VcwZ3gtYFY.

Yuyang Shi, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. Conditional simulation using diffusion Schrödinger bridges. In *Uncertainty in Artificial Intelligence (UAI)*, pp. 1792–1802. PMLR, 2022. URL https://proceedings.mlr.press/v180/shi22a.html.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger bridge matching. In *Advances on Neural Information Processing Systems (NeurIPS)*, 2023. URL https://openreview.net/forum?id=qy07OHsJT5.

Ana Cannas da Silva. Lectures on symplectic geometry. *Lecture Notes in Mathematics*, 1764, 2001.

Hae Jin Song and Laurent Itti. Riemannian-geometric fingerprints of generative models. In *2nd Beyond Euclidean Workshop: Hyperbolic and Hyperspherical Learning for Computer Vision*, 2025. URL https://openreview.net/forum?id=00LViVLqkN.

Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *International Conference on Machine Learning (ICML)*, 2023. URL https://proceedings.mlr.press/v202/song23k.html.

Benjamin Sterling, Chad Gueli, and Mónica F. Bugallo. Critically-damped higher-order Langevin dynamics. *arXiv preprint arXiv:2506.21741*, 2025. URL https://arxiv.org/abs/2506.21741.

Stas Syrota, Yevgen Zainchkovskyy, Johnny Xi, Benjamin Bloem-Reddy, and Søren Hauberg. Identifying metric structures of deep latent variable models. In *International Conference on Machine Learning (ICML)*, 2025. URL https://openreview.net/forum?id=FHzkr8FRie.

Sophia Tang, Yinuo Zhang, Alexander Tong, and Pranam Chatterjee. Branched Schrödinger bridge matching. *arXiv preprint arXiv:2506.09007*, 2025. URL https://arxiv.org/abs/2506.09007.

Andrea Testa, Søren Hauberg, Tamim Asfour, and Leonel Rozo. Geometric contact flows: Contactomorphisms for dynamics and control. In *International Conference on Machine Learning (ICML)*, 2025. URL https://openreview.net/forum?id=v2nQ1e78Rc.

Panagiotis Theodoropoulos, Augustinos D Saravanos, Evangelos A Theodorou, and Guan-Horng Liu. Momentum multi-marginal Schrödinger bridge matching. *arXiv preprint arXiv:2506.10168*, 2025. URL https://arxiv.org/abs/2506.10168.

Sahand Tokasi and Peter Pickl. Symplectic reduction of classical mechanics on shape space. *Foundations of Physics*, 52(5):106, 2022.

Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics. In *International Conference on Machine Learning (ICML)*, pp. 9526–9536. PMLR, 2020. URL https://proceedings.mlr.press/v119/tong20a.html.

Constantin Udrişte. Geometric dynamics. *Southeast Asian Bulletin of Mathematics*, 24(2):313–322, 2000.

Francisco Vargas, Pierre Thodoroff, Austen Lamacraft, and Neil Lawrence. Solving Schrödinger bridges via maximum likelihood. *Entropy*, 23(9):1134, 2021. URL https://doi.org/10.3390/e23091134.

Homer Rich Walke, Kevin Black, Tony Z. Zhao, Quan Vuong, Chongyi Zheng, Philippe Hansen-Estruch, Andre Wang He, Vivek Myers, Moo Jin Kim, Max Du, Abraham Lee, Kuan Fang, Chelsea Finn, and Sergey Levine. Bridgedata v2: A dataset for robot learning at scale. In *Conference on Robot Learning (CoRL)*, 2023. URL https://openreview.net/forum?id=f55MlAT1Lu.

Hao Wu, Shu Liu, Xiaojing Ye, and Haomin Zhou. Parameterized Wasserstein Hamiltonian flow. *SIAM Journal on Numerical Analysis*, 63(1):360–395, 2025. URL https://doi.org/10.1137/23M159281X.

Grace Hui Ting Yeo, Sachit D Saksena, and David K Gifford. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nature communications*, 12(1):3222, 2021. URL https://doi.org/10.1038/s41467-021-23518-w.

Federico Zadra. *Topics in contact Hamiltonian systems: analytical and numerical perspectives*. PhD thesis, University of Groningen, 2023.

Kira Zeevaert, Mohamed H Elsafi Mabrouk, Wolfgang Wagner, and Roman Goetzke. Cell mechanics in embryoid bodies. *Cells*, 9(10):2270, 2020. URL https://doi.org/10.3390/cells9102270.

Jiaqi Zhang, Erica Larschan, Jeremy Bigness, and Ritambhara Singh. scNODE: generative model for temporal single cell transcriptomic data prediction. *Bioinformatics*, 40(Supplement_2):ii146–ii154, 2024. URL https://doi.org/10.1093/bioinformatics/btae393.

Stephen Y. Zhang and Michael PH. Stumpf. Learning non-equilibrium diffusions with Schrödinger bridges: from exactly solvable to simulation-free. *arXiv preprint arXiv:2505.16644*, 2025. URL https://arxiv.org/abs/2505.16644.

A Extended State-of-the-art on Schrödinger Bridge Solvers

Existing methodologies for addressing the Schrödinger Bridge problem can be broadly divided into two main categories, depending on their solution strategy: those that directly fit the stochastic dynamics on the probabilistic manifold, and those that leverage the analytic optimality conditions of the problem to solve it. We review them in the sequel.

Schrödinger Bridge Solvers via Dynamics Parametrization. Traditional SB solvers often use Iterative Proportional Fitting (IPF) (Kullback, 1968; Léonard, 2013), which alternates forward and backward updates to successively match the initial and terminal marginals. However, IPF is computationally expensive, as it stores full trajectories, and it suffers from error accumulation, numerical instability, and reliance on strong priors (Vargas et al., 2021; Gushchin et al., 2024). Recent matching-based approaches (Shi et al., 2023; Gushchin et al., 2024; Peluchetti, 2023) improve scalability and robustness by learning time-reversed drifts via Markovian projections, circumventing the need for full trajectory storage and mitigating the IPF discretization errors. While Liu et al. (2024) extended this idea to the GSB setting, their assumption of Gaussian probability paths limits the model's expressivity. To overcome this, Tang et al. (2025) proposed branched dynamics using Gaussian mixtures. This enables more flexible path structures but at the expense of higher computational cost. For mmSB, iterative reference refinement with piecewise SB interpolation (Shen et al., 2025) suffers from inconsistencies in global trajectory construction due to its piecewise nature and shows high sensitivity to the choice of initial reference process. Alternatively, Tong et al. (2020) proposed a continuous normalizing flow for deterministic interpolation, removing noise from the reference process but preventing the construction of a true probabilistic bridge.

Schrödinger Bridge Solvers via Optimality Conditions. The optimality conditions of the SB problem take the form of dynamical equations on a Hamiltonian phase space, driven by dual potential functions (Chow et al., 2020). These conditions allow the exact dynamics to be recovered via integration and provide a flexible framework for generalization through modifications of the Hamiltonian. Furthermore, the state-dependent nature of the Hamiltonian framework offers a natural way to obtain Markovian approximations of a stochastic process. Vargas et al. (2021) and Chen et al. (2022) leveraged this view and solved the SB problem using control- and likelihood-based approaches, both employing iterative forward-backward updates on the Hamiltonian dynamics. Liu et al. (2022) extended this idea to the GSB setting, although without convergence guarantees. However, a critical bottleneck of these and derived methods is that a tractable integration of optimality conditions relies on iterative updates of a reference process. For example, Buzun et al. (2025) improved stability by directly modeling the dual potential and minimizing residuals of the Hamiltonian conditions, yet iterative updates incur significant computational overhead and may destabilize training due to the dependence on self-generated samples (Bertrand et al., 2024). This issue persists in mmSB settings (Theodoropoulos et al., 2025; Hong et al., 2025). Even when belief propagation is used to reduce time complexity (Chen et al., 2023), scaling to high dimensions remains poor. Therefore, while leveraging the optimality conditions offers clear advantages, it remains essential to develop computationally efficient, non-iterative algorithms with favorable scaling properties.

B Extended Preliminaries on Differential Geometry

B.1 Hamiltonian and Contact Hamiltonian Dynamics

Hamiltonian and contact Hamiltonian dynamics are governed by specific energy constraints that can be analyzed via differential geometry as flows on specialized manifolds. Hamiltonian dynamics is energy-conserving and evolves on a *symplectic manifold*. Contact Hamiltonian dynamics is more general, allowing for variable energy levels, and takes place on a *contact manifold*. Their formal definitions and key differences are discussed next.

Symplectic and Contact Structures. Let $\mathcal{M}$ be a smooth compact manifold, and let $\mathcal{T}_x\mathcal{M}$ denote the tangent space at $x \in \mathcal{M}$. The collection of all the tangent spaces identifies the tangent bundle $\mathcal{TM} = \cup_{x\in\mathcal{M}}\mathcal{T}_x\mathcal{M}$. A vector field $X : \mathcal{M} \to \mathcal{TM}$ assigns a tangent vector v to each point $x \in \mathcal{M}$. The set of all the vector fields over $\mathcal{TM}$ is denoted as $\Gamma(\mathcal{TM})$. A differential 1-form $\alpha : \mathcal{TM} \to \mathbb{R}$ is a smooth map field acting on vectors of the tangent bundle. For a smooth function $f : \mathcal{M} \to \mathbb{R}$, the 1-form $\alpha = df$ generalizes the gradient from Euclidean spaces. Specifically, df measures the variation of f under an infinitesimal displacement on $\mathcal{M}$. This displacement is locally described by a starting point x and a direction v, such that $(x, v) \in \mathcal{TM}$. Alternatively, it can be globally expressed by a vector field X. The variation of f along the vector field X is given by $df(X)$. This variation is independent of the choice of reference frame. To preserve this invariance, df must transform covariantly with X. Consequently, the 1-form $\alpha = df$ resides in the cotangent bundle $\mathcal{T}^*\mathcal{M}$, the dual space to $\mathcal{TM}$. The symplectic and contact structures provide two distinct mechanisms for associating a 1-form to a vector field, thereby establishing connections between the tangent and cotangent bundles. By considering the dynamics governed by the vector field and the scalar function defining the 1-form, a relationship between these elements emerges, as illustrated in Figure 9.

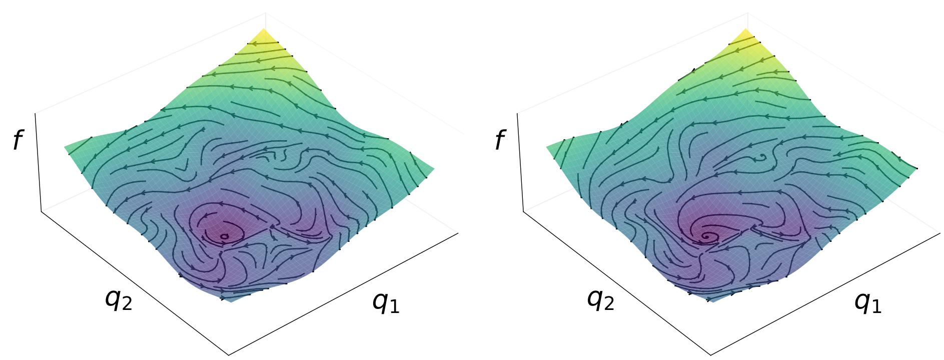

Figure 9: The same scalar function f, associated with the 1-form $\alpha = df$, gives rise to two distinct vector fields under the symplectic (left) and contact (right) geometric structures. The streamlines of these vector fields are illustrated on a representation of the state manifold. In symplectic geometry, the streamlines are tangent to the level curves of f, representing isoenergetic trajectories where f remains constant, thus describing the dynamics of conservative systems. In contrast, in contact geometry, a single flow line can traverse different energy levels.

Symplectic Geometry. A differential 2-form $\omega : \mathcal{TM} \times \mathcal{TM} \to \mathbb{R}$ is a skew-symmetric, bilinear, and smooth field of maps acting on pairs of tangent vectors. A 2-form is called *symplectic* if it is both closed ($d\omega = 0$) and non-degenerate. The symplectic form lacks the properties required to define an inner product. However, it still establishes a fundamental relation between differential 1-forms and vector fields: Given a 1-form df, the symplectic form ω uniquely determines a vector field X_f that is tangent to the level sets of f. This relation is defined by,

$$df(X) = \omega(X_f, X), \quad \forall X \in \Gamma(\mathcal{TM}). \tag{11}$$

By definition, f remains constant along the flow of X_f, which in turn preserves the symplectic form ω, i.e., $\mathcal{L}_{X_f}\omega = 0$ where $\mathcal{L}_{X_f}$ denotes the Lie derivative (Silva, 2001). In this framework, the function f is interpreted as a conserved energy, or equivalently, as a Hamiltonian H. The symplectic structure thereby endows $\mathcal{M}$ with a natural geometric framework for formulating Hamiltonian dynamics (Tokasi & Pickl, 2022). The pair $(\mathcal{M}, \omega)$ is referred to as a symplectic manifold. Notably, the non-degeneracy of ω implies that $\mathcal{M}$ must be even-dimensional.

Contact Geometry. While symplectic manifolds provide a geometric framework for modeling the dynamics of conservative systems in classical mechanics, a more general approach is required to describe non-conservative systems. This is addressed by contact manifolds, the odd-dimensional counterparts of symplectic manifolds (Geiges, 2001; Bravetti et al., 2017). A *contact manifold* is defined as $(\mathcal{M}, \eta)$, where $\mathcal{M}$ is an odd-dimensional smooth manifold, and η is a non-degenerate 1-form known as the contact form (Geiges, 2008). The contact form satisfies the *maximal non-integrability* condition, meaning that the top-degree differential form $\eta \wedge (d\eta)^d \neq 0$ is nowhere vanishing on $\mathcal{M}$. This form is constructed by taking the exterior product of η with the d-fold wedge product of its exterior derivative $d\eta$, i.e.,

$$(d\eta)^d = \underbrace{d\eta \wedge \cdots \wedge d\eta}_{d \text{ times}}. \tag{12}$$

The $(2d+1)$-form defines a volume form on $\mathcal{M}$, ensuring that the hyperplanes $\ker(\eta) \subset \mathcal{TM}$, constraining the dynamics on the contact manifold, do not form a foliation, i.e., they do not partition the manifold into lower-dimensional submanifolds (Geiges, 2001; 2008). Geometrically, this means that the contact distribution imposes *non-holonomic constraints*: it restricts the admissible directions of motion at each point without confining the dynamics to a fixed submanifold or energy level. This property is crucial for modeling systems where energy can change over time, enabling constraints on energy behavior without enforcing conservation.

Like symplectic geometry, contact geometry connects scalar functions to vector fields, enabling the description of dynamical systems (Zadra, 2023). Given an energy function $H : \mathcal{M} \to \mathbb{R}$, the dynamics on a contact manifold are defined by a contact Hamiltonian vector field X_H, as follows,

$$dH(X) = d\eta(X_H, X) - \mathcal{L}_{X_H}\eta(X),\ \forall X \in \Gamma(\mathcal{TM}). \tag{13}$$

Unlike symplectic geometry, where dynamics are confined to energy-preserving flows along the level sets of the Hamiltonian, contact geometry allows for an additional component of motion. Specifically, the dynamics on a contact manifold are not restricted to the term $d\eta(X_H, X)$, which lies tangent to the level sets of H, but also include a transverse component $\mathcal{L}_{X_H}\eta(X)$, arising from the non-degeneracy of the contact form. Consequently, while in symplectic geometry the symplectic form ω is strictly preserved, contact geometry allows the contact form η to be preserved only up to a scaling factor $a \in \mathbb{R}$ (Bravetti et al., 2017).

B.2 Riemannian and Jacobi Metrics

The Jacobi metric g_J is a rescaled version of a Riemannian metric g that allows Hamiltonian dynamics on the cotangent bundle $\mathcal{T}^*\mathcal{M}$ to be represented as geodesics on the Riemannian manifold $(\mathcal{M}, g)$. The construction is detailed below.

The Riemannian Metric. Let $\mathcal{M}$ be a smooth compact manifold. A Riemannian metric $g : \mathcal{TM} \times \mathcal{TM} \to \mathbb{R}$ is a smooth, symmetric, and positive-definite bilinear field of maps defined on pairs of vectors in the tangent bundle. This enables the introduction of an inner product on the tangent spaces of the manifold, allowing us to measure distances and curve lengths. For a smooth curve $x(t) : [t_0, t_1] \to \mathcal{M}$, the length l w.r.t. the metric g is $l = \int_{t_0}^{t_1} \sqrt{g(\dot{x}(t), \dot{x}(t))}dt$, where $\dot{x}(t) \in \mathcal{T}_{x(t)}\mathcal{M}$ is the vector tangent to the curve at $x(t)$. The curve minimizing this length between two points $x(t_0)$ and $x(t_1)$ on $\mathcal{M}$ is called a *geodesic*. Geodesics generalize straight lines in Euclidean space to curved spaces, representing the shortest paths in the geometry induced by g.

The Jacobi Metric. The geodesic flow $x(t)$ on a Riemannian manifold $(\mathcal{M}, g)$ lifts to the joint evolution of coordinates $(x(t), \alpha(x(t), \dot{x}(t))$ on the cotangent bundle $\mathcal{T}^*\mathcal{M}$ (Abraham & Marsden, 2008). This extended dynamics is governed by an energy function $H(x, \alpha) : \mathcal{T}^*\mathcal{M} \to \mathbb{R} = g^{-1}(\alpha, \alpha)$, which remains constant along the flow. A reparameterization $ds = \sqrt{H}dt$ links the trajectory of the integrated dynamical system at time t on $\mathcal{T}^*\mathcal{M}$ with the length of the corresponding geodesic on $\mathcal{M}$. This framework reveals a fundamental connection between geodesic flows and Hamiltonian dynamics in the special case where the Hamiltonian consists solely of a kinetic energy term. The cotangent bundle $\mathcal{T}^*\mathcal{M}$ is naturally equipped with a symplectic structure, making it a symplectic manifold $(\mathcal{T}^*\mathcal{R}, \omega)$.

This formulation can be further generalized by introducing a potential energy function into the Hamiltonian, given by $H(x, \alpha) = g^{-1}(\alpha, \alpha) + \mathcal{F}(x)$. In this setting, the geodesic structure underlying the Hamiltonian flow is determined by the Jacobi metric,

$$g_J = \big(H - \mathcal{F}(x)\big)\, g, \tag{14}$$

which rescales the original metric g by a position-dependent conformal factor (Abraham & Marsden, 2008). The corresponding time reparameterization takes the form $ds = \sqrt{H - \mathcal{F}(x)}\, dt$, restoring the interpretation of the trajectory as a geodesic with respect to the metric g_J (Udrişte, 2000).

C Insights on the Schrödinger Bridge

C.1 Hamiltonian Structure of the Generalized Schrödinger Bridge

This part presents the derivation of the Hamiltonian structure of the mmGSB problem (2), introduced in section (3), obtained via the method of Lagrange multipliers. To transform the constrained optimization problem into an unconstrained one, we incorporate the Fokker–Planck equation, scaled by the Lagrange multiplier S^t, into the original running cost $\mathcal{L}$ (i.e., the Lagrangian), as follows,

$$J(v^t, \rho^t, S^t) = \int_0^1 \mathcal{L}(v^t, \rho^t, S^t)\, dt; \tag{15}$$

$$\mathcal{L}(v^t, \rho^t, S^t) = \int_{\mathcal{M}} \left(\frac{1}{2}\|v^t(x)\|^2 + U(x) \right) \rho^t(x)\, dx \\ + \int_{\mathcal{M}} S^t(x) \underbrace{\left(\partial_t \rho^t(x) + \nabla_x \cdot (\rho^t(x) v^t(x)) - \varepsilon \Delta \rho^t(x)\right)}_{\text{Fokker-Planck equation}} dx. \tag{16}$$

The optimality conditions resulting from the extremization of the cost functional J in equation (15) follow from the Euler–Lagrange equations, generalized to the setting of classical field theory (Blohmann, 2024). In this framework, the arguments of the Lagrangian $\mathcal{L}$, in equation (16), are viewed as smooth fields defined over space and time. By setting to zero the variations of $\mathcal{L}$ with respect to these fields, we obtain the stationarity conditions for J. For a generic field $\psi^t(x)$, the corresponding Euler–Lagrange equation takes the form,

$$d_\psi \mathcal{L} = \partial_\psi \mathcal{L} + \partial_t(\partial_{\partial_t \psi} \mathcal{L}) + \nabla_x \cdot (\partial_{\nabla_x \psi} \mathcal{L}) + \Delta_x(\partial_{\Delta_x} \mathcal{L}) = 0. \tag{17}$$

Applying equation (17) to equation (16) for the fields v^t, ρ^t, and S^t, we obtain the following system of optimality conditions,

$$d_v \mathcal{L} = v^t(x)\rho^t(x) - \rho^t(x)\nabla_x S^t(x) = 0 \implies v^t(x) = \nabla_x S^t(x), \tag{18a}$$

$$d_\rho \mathcal{L} = \frac{1}{2}\|v^t(x)\|^2 - \partial_t S^t(x) - \nabla_x S^t(x) \cdot v^t(x) - \varepsilon \Delta_x S^t(x) + U(x) = 0, \tag{18b}$$

$$d_S \mathcal{L} = \partial_t \rho^t(x) + \nabla_x \cdot (\rho^t(x) v(x)) - \varepsilon \Delta_x \rho^t(x) = 0. \tag{18c}$$

Substituting the expression for the optimal velocity from equation (18a) into equations (18b) and (18c), we obtain the following Hamiltonian system,

$$\partial_t \rho^t(x) = \partial_S H(\cdot) = -\nabla_x \cdot (\rho^t(x)\nabla_x S^t(x)) + \varepsilon \Delta \rho^t(x), \tag{19a}$$

$$\partial_t S^t(x) = -\partial_\rho H(\cdot) = -\frac{1}{2}\|\nabla_x S^t(x)\|^2 - \varepsilon \Delta_x S^t(x) + U(x), \tag{19b}$$

with the corresponding Hamiltonian function,

$$H(\rho^t, S^t) = \frac{1}{2}\int_{\mathcal{M}} \|\nabla_x S^t(x)\|^2 \rho^t(x)\, dx - \int_{\mathcal{M}} U(x)\rho^t(x)\, dx + \varepsilon \int_{\mathcal{M}} S^t(x)\Delta_x \rho^t(x)\, dx. \tag{20}$$

The Hamiltonian system (19) can be reformulated in a linear and decoupled form by applying the Hopf–Cole coordinate transformation (Léger & Li, 2021; Chow et al., 2020), derived from the generating function F,

$$F(\rho^t, S^t) = S^t(x)\rho^t(x) - \varepsilon \rho^t(x)(\log \rho^t(x) - 1), \tag{21}$$

which yields the transformed coordinates,

$$\hat{\rho}^t(x) = \partial_S F(\cdot) = \rho^t(x), \tag{22a}$$

$$\hat{S}^t(x) = -\partial_\rho F(\cdot) = S^t(x) - \varepsilon \log \rho^t(x). \tag{22b}$$

This transformation preserves the Hamiltonian structure of the dynamics, as it is compatible with the underlying formulation. The Hopf–Cole transformation is well established in the literature, not only for simplifying the mathematical form of the equations, but also for enabling the design of efficient numerical integration schemes (Léger & Li, 2021). In our context, it is particularly advantageous for obtaining a Hamiltonian with separable kinetic and potential energy components. The Hamiltonian system (19) then becomes,

$$\partial_t \hat{\rho}^t(x) + \nabla_x \cdot (\hat{\rho}^t(x)\nabla_x \hat{S}^t(x)) = 0, \tag{23a}$$

$$\partial_t \hat{S}^t(x) + \varepsilon \frac{1}{\hat{\rho}^t(x)} \partial_t \hat{\rho}^t(x) = -\frac{1}{2}\left\|\nabla_x \hat{S}^t(x) + \varepsilon \nabla_x \log \hat{\rho}^t(x)\right\|^2 \\ - \varepsilon \Delta_x \hat{S}^t(x) - \varepsilon^2 \Delta_x \log \hat{\rho}^t(x) + U(x). \tag{23b}$$

Expanding the squared norm in equation (23b) and substituting $\partial_t \rho^t(x)$ from equation (23a) yields,

$$\begin{aligned}\partial_t \hat{S}^t(x) - \varepsilon \frac{1}{\hat{\rho}^t(x)} \nabla_x \cdot (\hat{\rho}^t(x) \nabla_x \hat{S}^t(x)) = & -\frac{1}{2}\|\nabla_x \hat{S}^t(x)\|^2 - \frac{1}{2}\varepsilon^2 \|\nabla_x \log, \hat{\rho}^t(x)\|^2 \\ & - \varepsilon \nabla_x \hat{S}^t(x) \nabla_x \log \hat{\rho}^t(x) - \varepsilon \Delta_x \hat{S}^t(x) \\ & - \varepsilon^2 \Delta_x \log \hat{\rho}^t(x) + U(x).\end{aligned} \tag{24}$$

Using the identity,

$$\varepsilon \frac{1}{\hat{\rho}^t(x)} \nabla_x \cdot (\hat{\rho}^t(x) \nabla_x \hat{S}^t(x)) = \varepsilon \Delta_x \hat{S}^t(x) + \varepsilon \nabla_x \hat{S}^t(x) \nabla_x \log \hat{\rho}^t(x), \tag{25}$$

equation (24) simplifies to,

$$\partial_t \hat{S}^t(x) = -\frac{1}{2}\|\nabla_x \hat{S}^t(x)\|^2 - \varepsilon^2 \frac{1}{2} \|\nabla_x \log \hat{\rho}^t(x)\|^2 - \varepsilon^2 \Delta_x \log \hat{\rho}^t(x) + U(x). \tag{26}$$

This form reveals the emergence of the Fisher information, defined as,

$$I(\hat{\rho}^t(x)) = \int_{\mathcal{M}} \|\nabla_x \log \hat{\rho}^t(x)\|^2 \hat{\rho}^t(x)\, dx, \tag{27a}$$

$$\partial_{\hat{\rho}} I(\hat{\rho}^t(x)) = -2\Delta_x \log \hat{\rho}^t(x) - \|\nabla_x \log \hat{\rho}^t(x)\|^2. \tag{27b}$$

Thus, equations (23a) and (26) admit the linear decoupled Hamiltonian formulation described in equation (3),

$$\begin{aligned}\partial_t \hat{\rho}^t(x) &= \partial_{\hat{S}} H(\cdot) = -\nabla_x \cdot (\hat{\rho}^t(x) \nabla_x \hat{S}_t(x)); \\ \partial_t \hat{S}^t(x) &= -\partial_{\hat{\rho}} H(\cdot) = -\frac{1}{2}\|\nabla_x \hat{S}^t(x)\|^2 + \frac{1}{2}\varepsilon^2 \partial_{\hat{\rho}} I(\hat{\rho}^t(x)) + U(x),\end{aligned}$$

governed by the Hamiltonian function,

$$H(\hat{\rho}^t, \hat{S}^t) = \underbrace{\frac{1}{2}\int_{\mathcal{M}} \|\nabla_x \hat{S}^t(x)\|^2 \hat{\rho}^t(x)\, dx}_{\text{Kinetic energy } \mathcal{K}} \underbrace{- \int_{\mathcal{M}} U(x) \hat{\rho}^t(x)\, dx - \frac{1}{2}\varepsilon^2 I(\hat{\rho}^t(x))}_{\text{Potential energy } \mathcal{F}} \tag{28}$$

In this formulation, the Fisher information contributes to the potential energy and encodes the effect of stochastic diffusion. Minimizing the Fisher information term promotes smoothness in the density and steers the Hamiltonian flow toward the target distribution ρ^1, providing greater robustness due to the regularizing effect of diffusion. This mechanism has been studied in the literature and employed in control applications for its regularization properties (Chen et al., 2025).

C.2 Contact Hamiltonian Structure of the Non-Conservative GSB

Here we derive the contact Hamiltonian formulation of the NCGSB problem, introduced in equation (4) and discussed in Section 4. The structure of the derivation closely mirrors that of the GSB in Appendix C.1, with one key distinction: the Lagrangian $\mathcal{L}$ now depends explicitly on the accumulated action z^t. Specifically, the augmented cost functional and Lagrangian are given by,

$$J(v^t, \rho^t, S^t, z^t) = \int_0^1 \mathcal{L}(v^t, \rho^t, S^t, z^t)\, dt; \tag{29}$$

$$\begin{aligned}\mathcal{L}(v^t, \rho^t, S^t, z^t) = & \int_{\mathcal{M}} \left(\frac{1}{2}\|v^t(x)\|^2 + U(x)\right) \rho^t(x)\, dx - z^t \\ & + \int_{\mathcal{M}} S^t(x) \left(\partial_t \rho^t(x) + \nabla_x \cdot (\rho^t(x) v^t(x)) - \varepsilon \Delta \rho^t(x)\right) dx.\end{aligned} \tag{30}$$

Since $\mathcal{L}$ depends explicitly on the evolving action z^t, the problem lies outside the scope of classical variational calculus. Instead, it fits within the framework of non-conservative variational principles, where the cost functional J evolves dynamically with the system. This is naturally addressed by the Herglotz variational principle, which extends the Euler–Lagrange equations to systems with dissipative effects. The optimality conditions obtained from the variations of $\mathcal{L}$, namely, the Herglotz-type

Euler–Lagrange equations, for a generic field argument $\psi^t(x)$ in this case take the form,

$$d_\psi \mathcal{L} = \partial_\psi \mathcal{L} + \partial_t(\partial_{\partial_t \psi}\mathcal{L}) + \nabla_x \cdot (\partial_{\nabla_x \psi}\mathcal{L}) + \Delta_x(\partial_{\Delta_x}\mathcal{L}) - \partial_z \mathcal{L}\, \partial_{\partial_t \psi}\mathcal{L} = 0. \tag{31}$$

Applying equation (31) for the fields v^t, ρ^t, and S^t, and recovering the dynamics of z^t from the optimization problem (4), we obtain the following system of optimality conditions,

$$d_v \mathcal{L} = v^t(x)\rho^t(x) - \rho^t(x)\nabla_x S^t(x) = 0 \implies v^t(x) = \nabla_x S^t(x), \tag{32a}$$

$$d_\rho \mathcal{L} = \frac{1}{2}\|v^t(x)\|^2 - \partial_t S^t(x) - \nabla_x S^t(x)\cdot v^t(x) - \varepsilon \Delta_x S^t(x) + U(x) - S^t(x) = 0, \tag{32b}$$

$$d_S \mathcal{L} = \partial_t \rho^t(x) + \nabla_x \cdot (\rho^t(x) v(x)) - \varepsilon \Delta_x \rho^t(x) = 0, \tag{32c}$$

$$\partial_t z^t - \int_{\mathcal{M}} \left(\frac{1}{2}\|v^t(x)\|^2 + U(x)\right)\rho^t(x)\, dx - z^t = 0. \tag{32d}$$

Compared to the optimality conditions for the GSB problem presented in equations (18), the set (32) includes an additional term, $-S^t(x)$, in equation (32b), which accounts for the dissipation term. By substituting the expression for the optimal velocity from equation (32a) into equations (32c), (32b), and equation (32d), we obtain the system of contact Hamiltonian dynamics,

$$\partial_t \rho^t(x) = \partial_S H(\cdot) = -\nabla_x \cdot (\rho^t(x)\nabla_x S^t(x)) + \varepsilon \Delta_x \rho^t(x), \tag{33a}$$

$$\partial_t S^t(x) = \partial_\rho H(\cdot) - S^t(x)\partial_z H(\cdot) = -\frac{1}{2}\|\nabla_x S^t(x)\|^2 - \varepsilon \Delta_x S^t(x) + U(x) - S^t(x), \tag{33b}$$

$$\partial_t z^t = S^t(x)\partial_S H(\cdot) - H(\cdot) = \int_{\mathcal{M}} \left(\frac{1}{2}\|\nabla_x S^t(x)\|^2 + U(x)\right)\rho^t(x)\, dx + z^t, \tag{33c}$$

with the associated contact Hamiltonian function given by,

$$H(\rho^t, S^t, z^t) = \frac{1}{2}\int_{\mathcal{M}} \|\nabla_x S^t(x)\|^2 \rho^t(x)\, dx - \int_{\mathcal{M}} U(x)\rho^t(x)\, dx + \varepsilon \int_{\mathcal{M}} S^t(x)\Delta_x \rho^t(x)\, dx - z^t. \tag{34}$$

In this case as well, it is beneficial to derive a decoupled and linearized representation of the dynamics. To this end, we perform a coordinate transformation from the original variables (ρ^t, S^t, z^t) to an alternative canonical set $(\hat{\rho}^t, \hat{S}^t, \hat{z}^t)$, while preserving the contact structure. This is achieved via a contact transformation generated by a generating function F, defined as,

$$F(\rho^t, S^t) = \frac{1}{2} S^t(x)\rho^t(x) - \varepsilon \rho^t(x)\left(\log \rho^t(x) - 1\right) - \frac{1}{2} z^t. \tag{35}$$

which yields the transformed coordinates,

$$\hat{\rho}^t(x) = \partial_S F(\cdot) = \frac{1}{2}\rho^t(x), \tag{36a}$$

$$\hat{S}^t(x) = \partial_\rho F(\cdot) - S^t(x)\, \partial_z F(\cdot) = S^t(x) - \varepsilon \log \rho^t(x), \tag{36b}$$

$$\hat{z}^t = S^t(x)\, \partial_S F(\cdot) - F(\cdot) = z^t + \varepsilon \rho^t(x)\left(\log \rho^t(x) - 1\right). \tag{36c}$$

The coefficients $\frac{1}{2}$ in equation (35) are required to eliminate exactly the nonlinear coupling term $S^t(x)\Delta_x \rho^t(x)$ from the Hamiltonian function. Indeed, by substituting the new coordinates (36) into the dynamical system (33), we obtain the following transformed contact Hamiltonian dynamics,

$$\begin{aligned}
\partial_t \hat{\rho}^t(x) &= \partial_{\hat{S}} H(\cdot) = -2\nabla_x \cdot (\hat{\rho}^t(x)\nabla_x \hat{S}^t(x)), \\
\partial_t \hat{S}^t(x) &= \partial_{\hat{\rho}} H(\cdot) - \hat{S}^t(x)\partial_{\hat{z}} H(\cdot) = -\frac{1}{2}\|\nabla_x \hat{S}^t(x)\|^2 + \frac{1}{2}\varepsilon^2 \partial_{\hat{\rho}} I(\hat{\rho}^t(x)) + U(x) \\
&\qquad + \hat{z}^t + 2\varepsilon \log 2\hat{\rho}^t, \\
\partial_t \hat{z}^t &= \hat{S}^t(x)\partial_{\hat{S}} H(\cdot) - H(\cdot) = \int_{\mathcal{M}} \left(\frac{1}{2}\|\nabla_x \hat{S}^t(x)\|^2 + U(x)\right)\hat{\rho}^t(x)\, dx + \frac{1}{2}\varepsilon^2 I(\hat{\rho}^t) \\
&\qquad - 2\int_{\mathcal{M}} \varepsilon(\log 2\hat{\rho}^t(x) - 1)\hat{\rho}^t(x)\, dx - \hat{z}^t,
\end{aligned}$$

with the associated contact Hamiltonian function given by,

$$H(\hat{\rho}^t, \hat{S}^t, \hat{z}^t) = \underbrace{\frac{1}{2}\int_{\mathcal{M}} \|\nabla_x \hat{S}^t(x)\|^2 \hat{\rho}^t(x)\, dx}_{\text{Kinetic energy } \mathcal{K}} + \underbrace{\hat{z}^t}_{\text{Non-conservative potential}} \underbrace{- \int_{\mathcal{M}} U(x)\hat{\rho}^t(x)\, dx - \frac{1}{2}\varepsilon^2 I(\hat{\rho}^t)}_{\text{Potential energy } \mathcal{F}} \underbrace{+2\int_{\mathcal{M}} \varepsilon(\log 2\hat{\rho}^t(x) - 1)\hat{\rho}^t(x)\, dx}_{\text{Entropy } \mathcal{B}}. \tag{37}$$

C.3 Guided Schrödinger Bridge

We consider the bridge ρ^t, computed between the terminal marginals ρ_a and ρ_b, using any variant of the SB problem (e.g., GSB, mmSB, NCGSB). This process can be modified to enforce desired conditions y at any chosen time t_s defined as,

$$y = f(x^{t_s}), \quad x^{t_s} \sim \rho^{t_s}, \tag{38}$$

while preserving the underlying data manifold. Conditioning in this way modifies the probability flow ρ^t (Guo et al., 2024) as,

$$\rho^t(x \mid y) = \frac{1}{Z}\,\rho^t(x)\, e^{-\|y - f(\bar{x}^{t_s})\|^2}, \tag{39}$$

where Z is a normalization constant, and $\bar{x}^{t_s}$ denotes a sample from the predicted prescribed distribution $\bar{\rho}^{t_s}$ conditioned on the current distribution ρ^t. This weight biases the generation toward samples that satisfy the desired property y. In a dynamical setting, we perform this conditioning by incorporating a control term G^t into the Fokker–Planck dynamics,

$$\partial_t \rho^t(x) + \nabla_x \cdot \left[\rho^t(x)\left(v^t(x) + G^t(x)\right)\right] = \varepsilon \Delta_x \rho^t(x). \tag{40}$$

A naive choice such as $G^t(x) \propto \nabla_{x^t} f(x^t)$ often drives the dynamics off the data manifold, producing unrealistic samples far from it and the target distribution ρ_b. Instead, Bayes' rule provides the correct structure of the guidance term. The gradient of the conditional log-likelihood decomposes as,

$$\nabla_{x^t} \log \rho^t(x^t \mid y) = \nabla_{x^t} \log \rho^t(x^t) + \underbrace{\nabla_{x^t} \log \rho^t(y \mid x^t)}_{\text{estimated by } G^t}. \tag{41}$$

Substituting the conditional form from equation (39) yields,

$$G^t(x) = \nabla_{x^t} \log e^{-\|y - f(\bar{x}^{t_s})\|^2} = -\nabla_{x^t} \|y - f(\bar{x}^{t_s})\|^2. \tag{42}$$

Specifically, in the NCGSB framework (4), this guidance is incorporated directly into the drift v^t by adding the penalty $\|y - f(\bar{x}^{t_s})\|^2$ to the Lagrangian in the action dynamics constraint (4b),

$$\begin{aligned}
\min_{v^t} \quad & J(v^t, \rho^t) = \int_0^1 \partial_t z^t\, dt, \\
\text{s.t.} \quad & \partial_t z^t = \int_{\mathcal{M}} \left(\frac{1}{2}\|v^t(x)\|^2 + U(x) + \|y - f(\bar{x}^{t_s})\|^2\right)\rho^t(x)\, dx - z^t, \\
& \partial_t \rho^t(x) + \nabla_x \cdot \left(\rho^t(x)\, v^t(x)\right) = \varepsilon \Delta_x \rho^t(x), \\
& \rho^0 = \rho_a, \quad \rho^1 = \rho_b, \quad \rho^{t_m} = \rho_m, \quad \forall\, m \in \{1, \ldots, M\}.
\end{aligned} \tag{43}$$

Here, the inclusion of $\|y - f(\bar{x}^{t_s})\|^2$ in the Lagrangian produces the desired $-\nabla_{x^t}\|y - f(\bar{x}^{t_s})\|^2$ correction in the drift, while preserving the Schrödinger bridge structure and constraints.

D ResNet Parameterization for Discrete Geodesics

D.1 Geometric Interpretation

As stated in the main paper, our objective is to compute a geodesic ρ^t on $\mathcal{P}^+(\mathcal{M})$, induced by the contact Hamiltonian dynamics, that is constrained to pass through a set of observed marginals

$\{\rho_a, \rho_m, \rho_b\}$ (i.e., discretized distributions along the probability path). These constraints naturally lead to a discretized parameterization of ρ^t, where the overall density transformation is modeled as a composition of maps, each connecting a pair of consecutive observations. In this context, a ResNet architecture is ideally suited for this problem, as its sequential block structure directly mirrors this piecewise, compositional nature of the approximated geodesic.

Geometrically, each parameterized pushforward defines a vector $\partial_t \rho_\theta^{t_k}$ in the tangent space of $\rho_\theta^{t_{k-1}}$, representing its change rate. The pair $(\rho_\theta^{t_{k-1}}, \partial_t \rho_\theta^{t_k})$ thus corresponds to a point on the tangent bundle $\mathcal{TP}^+(\mathcal{M})$, with parameters $\theta^k \in \Theta$ representing one of the possible coordinate charts for this update. As such, the parameter space Θ forms a finite–dimensional subspace of $\mathcal{TP}^+(\mathcal{M})$ (see Fig. 5). The block transformation defines a smooth immersion $T_{\theta^k} : \Theta \to \mathcal{TP}^+(\mathcal{M})$ with full-rank Jacobian $\nabla_x T_{\theta^k}$, ensuring the pullback of the Wasserstein metric $g^{\mathcal{W}_2}$ to Θ, denoted $T_\theta^* g^{\mathcal{W}_2}$, is well-defined and induces a Riemannian structure. This Riemannian metric identifies Θ with its dual Θ^* via the standard tangent–cotangent isomorphism (do Carmo, 1992). Consequently, the contact Hamiltonian dynamics on $\mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$ can be equivalently expressed in the reduced phase space $\Theta^* \times \mathbb{R}$ (Wu et al., 2025). At the same time, the Wasserstein manifold is approximated by the finite-dimensional submanifold $\mathcal{P}_\theta^+(\mathcal{M})$, whose tangent space is $\mathcal{TP}_\theta^+(\mathcal{M}) = \Theta$.

D.2 Training Algorithm

Algorithm 1 Training the Contact Wasserstein Geodesic (CWG) Framework

Input: Dataset: samples from marginals $x_{a,\{i,j\}} \sim \rho_a,\ x_{b,\{i,j\}} \sim \rho_b,\ x_{n,\{i,j\}} \sim \{\rho_n\}_{n=1}^N$.
Output: A trained ResNet $T_{\{\theta^k\}_{k=0}^K}$.

Part I: Initialization to Match the Initial Marginal

```
for i = 1 to E do                                   ▷ Epoch loop
    for j = 1 to B do                               ▷ Batch loop
        x^s ~ λ                                     ▷ Sample from reference distribution
        x^{t_0} = T_{θ^0}(x^s)                      ▷ Apply initial block
        min_{θ^0} d^2_{W_2}(x^{t_0}, x_{a,{i,j}})   ▷ Match initial marginal
    end for
end for
```

Part II: Geodesic Optimization

```
for i = 1 to E do
    for j = 1 to B do
        x^s ~ λ                                                   ▷ Sample from reference distribution
        {x^{t_0}, x^{t_1}, ..., x^{t_K}} = T_{{θ^k}_{k=0}^K}(x^s) ▷ Full ResNet transformation
        min_{θ\θ^0} ℓ({x^{t_k}}_{k=0}^K, x_{b,{i,j}}, x_{n,{i,j}}) ▷ Minimize geodesic loss (equation (9))
    end for
end for
```

D.3 Time Complexity and Practical Considerations

At each iteration of the geodesic optimization of Alg. 1, we sample N points $x^s \in \mathcal{M}$ of dimension d from the reference distribution λ and pass them through the ResNet. Assuming its architecture consists of K blocks, each being an MLP of L layers and hidden dimension W, the computational cost of this forward pass is $\mathcal{O}(N\,d\,K\,L\,W)$. The loss function ℓ in equation (9) requires $M + K$ evaluations of the Wasserstein distance between sample batches, and K evaluations of the factor Φ^{t_k}. For the Wasserstein distance, we employ the `geomloss` library (Feydy, 2020), which uses the Sinkhorn algorithm with time complexity $\mathcal{O}(N(d + T_{\text{sh}}))$, where T_{sh} is the number of Sinkhorn iterations until convergence (Feydy, 2020). To evaluate Φ^{t_k}, we apply its complete definition,

$$\Phi^{t_k} = H^{t_k} + \int_{\mathcal{M}} \Big[U\big(x^{t_k}\big) - 2\varepsilon\big(\log(2\rho_\theta^{t_k}) - 1\big)\Big]\, \rho_\theta^{t_k}\, dx + \tfrac{1}{2}\varepsilon^2 I(\rho_\theta^{t_k}), \tag{44}$$

to a batch of samples. As discussed in Sec. 3, the Fisher information $I(\rho_\theta^{t_k})$ in the potential energy originates from the entropy regularization in the SB formulation and does not require explicit computation. Its effect is implicitly captured by the entropy-regularized Wasserstein distance. Therefore, evaluating Φ^{t_k} reduces to computing the scalar functions H and U, along with estimating the entropy term $\log(2\rho_\theta^{t_k})\,\rho_\theta^{t_k}$, which is the computational bottleneck. This term can be estimated using a k-NN entropy estimator with time complexity $\mathcal{O}(d\,N\log N)$ (Borelli et al., 2022). Considering all components, and given that $K \geq M$, the overall time complexity becomes,

$$\mathcal{O}(N\,d\,K\,L\,W)+\mathcal{O}\big(N(K+M)(d+T_{\text{sh}})\big)+\mathcal{O}(K\,d\,N\log N) \approx \mathcal{O}\Big(NK\big(T_{\text{sh}}+d(LW+\log N)\big)\Big). \tag{45}$$

Our method demonstrates highly favorable scaling properties, offering a significant advantage over existing approaches. Notably, its computational complexity scales linearly with data dimensionality d and nearly linearly with the batch size N. This stands in stark contrast to existing methods like (Hong et al., 2025), which scale quadratically in N and potentially exponentially in d. Furthermore, our model's performance is only weakly influenced by the number of marginals and it circumvents the expensive outer iteration loops required for convergence in methods like (Chen et al., 2023; Shen et al., 2025).

E Implementation Details

E.1 Empirical Approximation of the Wasserstein-2 Distance

Let ρ_c and ρ_d be two probability distributions from which we draw batches of samples $\{x_{c,i}\}_{i=1}^N \sim \rho_c$ and $\{x_{d,j}\}_{j=1}^M \sim \rho_d$, respectively. The Wasserstein-2 distance, denoted by $d_{\mathcal{W}_2}(\rho_c, \rho_d)$, measures the minimal cost of transporting mass between these two distributions. To approximate this distance empirically, we first construct a cost matrix $C \in \mathbb{R}^{N\times M}$, where each entry,

$$C_{ij} = \|x_{c,i} - x_{d,j}\|^2, \tag{46}$$

represents the squared Euclidean distance between sample $x_{c,i}$ and sample $x_{d,j}$. A transport plan is then defined as a matrix $\pi \in \mathbb{R}_+^{N\times M}$ that assigns how much mass to move from each $x_{c,i}$ to each $x_{d,j}$, minimizing the total transport cost weighted by C. To avoid degenerate solutions where all mass is concentrated on a few points, entropy regularization is introduced, encouraging smoother and more distributed transport plans. For this computation, we employ the `SamplesLoss` function from the `geomloss` library (Feydy, 2020), with parameters resumed in Table 7.

In the conditional generation setting, the probability flow $\rho^t(x \mid y)$ is conditioned on the feature $y = f(x^{t_s})$, with $x^{t_s} \sim \rho_s$, to ensure that the generated samples satisfy y at time step t_s. Therefore, when comparing the prescribed distribution ρ_s with a marginal ρ_m, a modified Wasserstein-2 distance $d'_{\mathcal{W}_2}(\rho_s, \rho_m)$ incorporating the feature penalty is used. This distance is defined via the cost matrix

$$C_{ij} = \|x_i^{t_s} - x_{m,j}\|^2 + \|y - f(x_{m,j})\|^2, \tag{47}$$

which penalizes transport plans assigning mass to samples $x_{m,j}$ inconsistent with the conditioning feature y.

F Extended Results

F.1 Experimental Setup

The LiDAR Manifold Navigation and Cell Sequencing experiments were conducted on a machine equipped with 13th Gen Intel® Core™ i7-13850HX CPUs. The Image Generation experiment was run on a system with an NVIDIA GeForce RTX5090 GPU (32GB VRAM, CUDA12.9, driver version 575.64.03). Results for the LiDAR Manifold Navigation (Table 2), Single Cell Sequencing (Table 3), Sea Temperature Prediction (Table 4 and Appendix F.4), and Robot Task Reconstruction (Table 5) are based on a total of ten evaluations obtained from training runs with different initial conditions. The tables report the mean and standard deviation of these distributions.

The ResNet architecture varies depending on the task. For the LiDAR Manifold Navigation and Cell Sequencing experiments, each block is an MLP that processes the output of the previous block and generates an update, which is added to the input with a step size τ: $x \leftarrow x + \tau\,\text{block}(x)$. Details of this architecture are provided in Table 9. In contrast, for the Image Generation experiment, the input consists of images, and each block is implemented as a 2D U-Net. Details of this architecture are provided in Table 10.

Parameter	Value
Entropy	0.05
Euclidean norm order	2
Scaling	0.7

Table 7: `SampleLoss` parameters.

Parameter	Lidar	Cell	Sea	Robot
# samples N	1000	1000	200	100
Weight w_b	10	10	100	10
Weight w_m	–	10	100	–
Weight w_g	1	1	1	1

Table 8: Loss function (9) parameters for the experiments.

Component	ResNet	
	Lidar	**Cell**
Number of blocks K	20	5
Layers per block	3	3
Layer hidden size	30	128
Step size τ	0.1	0.1
Input dimension d	3	5

Table 9: Configuration of the ResNet

Component	Image ResNet	
	Sea	**Robot**
Number of blocks K	5	8
Input channels	1	3
Output channels	1	3
Layers per block	1	2
Downsampling blocks	2 (32 and 64 channels)	
Upsampling blocks	2 (32 and 64 channels)	
Step size τ	1	1
Input dimension d	4096	12288

Table 10: Configuration of the Image ResNet.

During training, the weights $\{w_b, w_m, w_g\}$ balancing the loss terms in equation (9),

$$\ell = w_b\, d^2_{\mathcal{W}_2}(\rho_\theta^{t_K}, \rho_b) + w_m \sum_{m=1}^{M} d^2_{\mathcal{W}_2}(\rho_\theta^{t_{k_m}}, \rho_m) + w_g \sum_{k=1}^{K} \Phi^{t_k}\, d^2_{\mathcal{W}_2}(\rho_\theta^{t_k}, \rho_\theta^{t_{k-1}}),$$

along with the number of samples N used for Wasserstein distance estimation, are experiment-specific and summarized in Table 8. The GPU memory consumption for the two image generation experiments, comparing our CWG method with the baseline approaches, is reported in Table 11. Due to the explicit handling of the full probability distribution (albeit in discretized form) within the ResNet architecture, CWG exhibits particularly high memory requirements. In contrast, methods such as GSBM and DSBM model only the drift component and subsequently integrate the dynamics. While this makes them significantly more demanding in terms of computation time, they are more memory-efficient than CWG.

Methodology	CWG	GSBM	DSBM
Sea Temperature (MB)	25200 ± 200	9600 ± 300	9000 ± 200
Robot Task (MB)	16100 ± 200	12500 ± 300	10200 ± 200

Table 11: Comparison of GPU memory consumption across the methods evaluated in the Image Generation experiment. CWG shows a decrease in memory usage for the Robotic Task Reconstruction experiments, whereas the other methods exhibit an increase due to the reduced batch size used in this training (Table 8).

F.2 LiDAR Manifold Navigation

The LiDAR dataset (OpenTopography, 2025) consists of point clouds contained in the domain $[-5, 5]^3 \subset \mathbb{R}^3$. The objective of the experiment is to construct a bridge across the data manifold for connecting two distributions while avoiding regions of high elevation and remaining closely aligned with the manifold structure. The initial distribution ρ_a is composed by a mixture of 4 Gaussian distributions, the target distribution ρ_b is composed of 2 Gaussians on the two sides of the mountain.

The manifold shape is incorporated in the problem through the potential function U, inherited from the baseline (Liu et al., 2024),

$$\int_{\mathcal{M}} U(x)\rho^t(x)\,dx = \int_{\mathcal{M}} \big(U_{\text{manifold}}(x) + U_{\text{height}}(x)\big)\rho^t(x)\,dx,$$
$$U_{\text{manifold}}(x) = w_{\text{manifold}}\|\psi(x) - x\|^2, \quad U_{\text{height}}(x) = w_{\text{height}}\|\psi^{(z)}(x)\|^2. \tag{48}$$

Here, $\psi(x)$ denotes the projection of a point x onto an approximate tangent plane, estimated from its p nearest neighbors on the data manifold. The notation $\psi^{(z)}(x)$ refers to the z-coordinate of $\psi(x)$, i.e., the height of the fitted plane. The weights w_{manifold} and w_{height} control the relative importance of the two terms in the potential function. We now detail the construction of $\psi(x)$. Let $N_p(x) = \{x_1^l, \ldots, x_p^l\}$ denote the set of p nearest neighbors of $x \in \mathbb{R}^3$ in the dataset. To approximate the local tangent plane, we employ a moving least-squares (MLS) procedure (Levin, 1998). Specifically, the plane parameters (a, b, c) are obtained by solving,

$$\min_{a,b,c} \frac{1}{p}\sum_{i=1}^{p} w(x, x_i^l)\left(a x_i^{l(x)} + b x_i^{l(y)} + c - x_i^{l(z)}\right)^2, \tag{49}$$

where the superscripts indicate coordinates and the weights are defined as,

$$w(x, x_i^l) = \exp\left(-\tfrac{\|x - x_i^l\|}{\gamma}\right), \tag{50}$$

with γ being a scaling parameter. Given the fitted plane, the projection operator $\psi(x)$ is defined as,

$$\psi(x) = x - \frac{x^\top n + c}{\|n\|^2}\,n, \quad n = [a \;\; b \;\; -1]^\top, \tag{51}$$

where n denotes the plane's normal vector. Differentiation through ψ naturally restricts gradients to this tangent plane, thereby ensuring that optimization of the state cost U evolves within the geometry of the data manifold. The values of the parameters used in the computation of the projection operator $\psi(x)$ are provided in Table 12.

In the guided generation setting, the feature function f, defines as follows,

$$f(x_b) = \mathrm{ReLU}(w_x x_b^{(x)} - b_x) + \mathrm{ReLU}(w_y x_b^{(y)} - b_y), \quad x_b \sim \rho_b \tag{52}$$

is used to penalize samples from the terminal marginal distribution ρ_b on the left side of the mountain. The parameters used in this experiment are listed in Table 13, and the results of guidance fine-tuning are reported in Table 14. The training time (tt) metric shows that guiding the generation requires only 10.7% of the original training time.

Table 12: Potential function U parameters for the LiDAR Manifold Navigation experiment.

Parameter	Value
Weight manifold w_{manifold}	5
Weight height w_{height}	1
Spatial scaling γ	0.1
# neighbor points p	20

Table 13: Parameters for the feature function f (52), used for guided generation in the LiDAR Manifold Navigation experiment.

Parameter	Value
Weight (x) w_x	-1
Bias (x) b_x	0
Weight (y) w_y	1
Bias (y) b_y	0

F.3 Single Cell Sequencing

The Embryoid Body (EB) stem cell differentiation dataset (Moon et al., 2019) captures cell state progression across five developmental stages $[t_0, t_1, t_2, t_3, t_4]$ over a 27-day period. Snapshots were collected at five discrete time intervals: $t_0 \in [0, 3]$, $t_1 \in [6, 9]$, $t_2 \in [12, 15]$, $t_3 \in [18, 21]$, and $t_4 \in [24, 27]$. These stages involve significant structural changes, with cells moving and reorganizing within increasingly stiff tissue while consuming and releasing mechanical energy (Zeevaert et al.,

Table 14: Bridge energy J ($\downarrow$) with penalty f in the LiDAR Manifold Navigation task, reported for our CWG method before and after guidance fine-tuning.

Metric	before	after
J	$12.31_{\pm 0.18}$	$\mathbf{2.49_{\pm 0.43}}$
tt (s)	$\mathbf{280_{\pm 20}}$	$310_{\pm 25}$

2020; Kinney et al., 2014). Consequently, the resulting dynamics exhibit energy dissipation and are better described by the NCGSB framework than by energy-conserving models. In this experiment, we evaluate the framework's ability to generalize to regions with no available data by dividing the dataset into a training set $[t_0, t_2, t_4]$ and a validation set $[t_1, t_3]$. The geometry of the data manifold is incorporated into the NCGSB problem through a potential function U, defined as,

$$U(x^t) = \frac{1}{N_1} \sum_{i=1}^{N_1} \left[\frac{1}{N_2} \sum_{j=1}^{N_2} e^{\frac{1}{\gamma} \|x_j^t - x_i\|^2} (x_j^t - x_i)^2 \right]^{-1}, \tag{53}$$

where $x_j^t \sim \rho^t$ denotes a sample from the posterior distribution, and $x_i \sim \{\rho_a, \rho_m, \rho_b\}$ are samples from the marginal distributions. N_1 and N_2 indicate the number of samples taken from ρ^t and $\{\rho_a, \rho_m, \rho_b\}$, respectively, while γ represents a spatial scaling parameter. The exponential term acts as a kernel-like weight that reduces the influence in the inverse summation $\sum_{i=1}^{N_1}[\cdot]^{-1}$ of the data points far from the bridge. Globally, the potential function U measures the distance of the bridge ρ^t from the available data $\{\rho_a, \rho_m, \rho_b\}$ (the training set) and its minimization guides the construction of a bridge that stays close to the known manifold while generalizing effectively to regions without data (the validation set). While other approaches leveraging this dataset (Tong et al., 2020; Shen et al., 2025) first embed the data into a 100-dimensional feature space using principal component analysis (PCA) and then restrict the analysis to the first five dimensions, this procedure excessively linearizes and flattens the data manifold, making navigation trivial and eliminating the need for intermediate marginals (Shen et al., 2025). To better preserve the manifold's geometry, we instead apply the PHATE algorithm (Moon et al., 2019) to the 100-dimensional representation, producing a 5-dimensional nonlinear embedding that more faithfully captures the original structure.

In the results reported in Table 3, the scalar Hamiltonian function H^{t_k} in the pullback Jacobi metric, $T_\theta^* \tilde{g}_{\mathrm{J}} = \left(H^{t_k} - \mathcal{F}(\rho^t) - \mathcal{B}(\rho^t)\right) T_\theta^* g^{\mathcal{W}_2}$ is linearly varied from an initial value $H^{t_0} = 1$ to a final value $H^{t_K} = 0.82$. This law is treated as a hyperparameter of the methodology. All ten experiments reported in Table 3 are conducted under this Hamiltonian behavior. A comparison with the conservative case, where $H^{t_0} = 1$ is kept constant, is provided in Table 16. The superior performance of the energy-varying case indicates that the problem is inherently non-conservative.

Table 15: Single Cell Sequencing Parameters.

Parameter	Value
Spatial scaling γ	0.3
# bridge samples N_1	1000
# marginal samples N_2	3000

Table 16: Wasserstein error at validation ($\downarrow$) and training time (tt) ($\downarrow$) in the ablation study comparing the energy-varying (e-v) and energy-conserving (e-c) versions of CWG on the Single Cell Sequencing task.

Metric	CWG (e-v)	CWG (e-c)
$d_{\mathcal{W}_2}(x^{t_1})$	$\mathbf{1.11_{\pm 0.06}}$	$1.31_{\pm 0.06}$
$d_{\mathcal{W}_2}(x^{t_3})$	$\mathbf{0.33_{\pm 0.02}}$	$0.49_{\pm 0.03}$
tt (s)	$710_{\pm 30}$	$710_{\pm 30}$

F.4 Sea Temperature Prediction

The NOAA OISST v2 High Resolution Dataset (Huang et al., 2021) is a long-term Climate Data Record that integrates observations from multiple platforms (satellites, ships, buoys, and Argo floats) into a global gridded product. For this experiment, we use daily averages of sea surface temperature in the Gulf of Mexico between 1981 and 2024, represented as 64×64 single-channel images. We cluster the measurements over five-year periods and select five representative months to define five time frames: January (t_0), March (t_1), May (t_2), July (t_3), and September (t_4). Each month corresponds to a distribution of images, denoted as

$\{\rho_a$ (January), ρ_{m_1} (March), ρ_{m_2} (May), ρ_{m_3} (July), ρ_b (September)$\}$. A sample from one of these distributions is a heatmap of the Gulf's temperature for a specific day in the corresponding month of the specified five-year period. The goal of this test is to evaluate our method's ability to interpolate across missing time frames, generating realistic temperature maps for months without data. To this end, we partition the dataset into a training set $\{t_0, t_2, t_4\}$ and a validation set $\{t_1, t_3\}$, and assess the quality of predictions on the held-out months. To encourage generalization beyond the training data, we introduce a potential function U that penalizes deviations from the learned data manifold. Building on the approach of Song & Itti (2025), where generative models are evaluated by measuring the distance between their outputs and a geometric manifold of real images learned by a VAE, we adopt a similar strategy. Specifically, we use a state-of-the-art VAE architecture, with parameters listed in Table 17, to learn the manifold of the training images in our dataset. The potential function $U(x^t)$, for samples $x^t \sim \rho^t$, is then defined as the squared distance between a bridge sample x^t and its VAE-projected reconstruction $\tilde{x}^t = \text{VAE}(x^t)$: $U(x^t) = \|x^t - \tilde{x}^t\|^2$. Results for each five-year periods are presented below.

In these results, the scalar Hamiltonian function H^{t_k} in the pullback Jacobi metric, $T_\theta^* \tilde{g}_{\text{J}} = \left(H^{t_k} - \mathcal{F}(\rho^t) - \mathcal{B}(\rho^t)\right) T_\theta^* g^{\mathcal{W}_2}$ is linearly varied from an initial value $H^{t_0} = 1$ to a final value $H^{t_K} = 1.36$. This law is treated as a hyperparameter of the methodology, and all ten experiments reported in the following tables were conducted under this Hamiltonian behavior. The higher final energy observed to be beneficial for the modeling of the warmer months can be associated with the increased thermodynamical entropy of these cases, resulting in generally more diverse samples and larger variation in the data manifold. By introducing higher energy, we encourage the model to effectively capture this diversification, traversing regions of the data manifold that, in other contexts Arvanitidis et al. (2018), are considered uncertain and typically avoided.

Table 17: Architecture of the ConvVAE used in the Sea Temperature Prediction experiment. All Conv2D and ConvTranspose2D layers use kernel size 4, stride 2, and padding 1, followed by ReLU activations (except the last layer, which uses Sigmoid).

Stage	**Layer (channels)**	**Output size**
Input	Single-channel image	$1 \times 64 \times 64$
Encoder	Conv2D (1→32)	$32 \times 32 \times 32$
	Conv2D (32→64)	$64 \times 16 \times 16$
	Conv2D (64→128)	$128 \times 8 \times 8$
	Flatten	8192
Latent space	Linear $\to \mu$	5
	Linear $\to \log\sigma^2$	5
Decoder	Linear → reshape	$128 \times 8 \times 8$
	ConvT2D (128→64)	$64 \times 16 \times 16$
	ConvT2D (64→32)	$32 \times 32 \times 32$
	ConvT2D (32→1), Sigmoid	$1 \times 64 \times 64$

Table 18: FID scores at training and validation steps (↓), and training time (tt) (↓) in Sea Temperature (2020–2024).

Metric	**CWG**	**GSBM**	**DSBM**
FID(x^{t_0})	$41.56_{\pm 1.89}$	–	–
FID(x^{t_1})	$\mathbf{121.47_{\pm 5.61}}$	$160.68_{\pm 4.54}$	$242.26_{\pm 9.94}$
FID(x^{t_2})	$51.51_{\pm 5.52}$	$54.47_{\pm 5.88}$	$56.51_{\pm 5.78}$
FID(x^{t_3})	$\mathbf{159.53_{\pm 7.38}}$	$185.54_{\pm 7.11}$	$235.83_{\pm 10.44}$
FID(x^{t_4})	$61.48_{\pm 6.79}$	$59.14_{\pm 5.52}$	$58.39_{\pm 6.13}$
tt (s)	$\mathbf{1030_{\pm 50}}$	$73600_{\pm 3200}$	$19100_{\pm 900}$

Table 19: FID scores at training and validation steps (↓) in Sea Temperature (2015–2019).

Metric	CWG	GSBM	DSBM
FID(x^{t_0})	$42.96_{\pm 2.07}$	–	–
FID(x^{t_1})	$\mathbf{130.33_{\pm 5.94}}$	$145.97_{\pm 6.12}$	$220.76_{\pm 8.61}$
FID(x^{t_2})	$58.72_{\pm 5.46}$	$62.13_{\pm 5.69}$	$65.47_{\pm 6.16}$
FID(x^{t_3})	$135.25_{\pm 6.73}$	$142.82_{\pm 7.27}$	$228.14_{\pm 9.26}$
FID(x^{t_4})	$63.02_{\pm 6.14}$	$59.73_{\pm 5.86}$	$61.29_{\pm 6.05}$

Table 20: FID scores at training and validation steps (↓) in Sea Temperature (2010–2014).

Metric	CWG	GSBM	DSBM
FID(x^{t_0})	$45.19_{\pm 2.27}$	–	–
FID(x^{t_1})	$\mathbf{132.47_{\pm 6.58}}$	$168.93_{\pm 6.24}$	$255.36_{\pm 10.19}$
FID(x^{t_2})	$60.57_{\pm 6.08}$	$59.79_{\pm 5.98}$	$61.89_{\pm 6.23}$
FID(x^{t_3})	$140.84_{\pm 6.01}$	$144.60_{\pm 6.83}$	$235.08_{\pm 9.07}$
FID(x^{t_4})	$54.92_{\pm 5.23}$	$51.63_{\pm 5.55}$	$50.27_{\pm 5.94}$

Table 21: FID scores at training and validation steps (tt) (↓) Sea Temperature (2005–2009).

Metric	CWG	GSBM	DSBM
FID(x^{t_0})	$47.28_{\pm 2.43}$	–	–
FID(x^{t_1})	$\mathbf{172.69_{\pm 7.62}}$	$195.83_{\pm 8.04}$	$260.97_{\pm 10.92}$
FID(x^{t_2})	$56.08_{\pm 5.52}$	$59.57_{\pm 5.87}$	$62.03_{\pm 6.29}$
FID(x^{t_3})	$\mathbf{126.87_{\pm 6.94}}$	$152.26_{\pm 6.63}$	$243.58_{\pm 10.39}$
FID(x^{t_4})	$66.43_{\pm 6.58}$	$63.17_{\pm 6.34}$	$64.86_{\pm 6.63}$

Table 22: FID scores at training and validation steps (↓) in Sea Temperature (2000–2004).

Metric	CWG	GSBM	DSBM
FID(x^{t_0})	$49.83_{\pm 2.64}$	–	–
FID(x^{t_1})	$\mathbf{140.48_{\pm 7.93}}$	$172.96_{\pm 8.34}$	$250.86_{\pm 11.08}$
FID(x^{t_2})	$56.73_{\pm 5.36}$	$56.97_{\pm 5.79}$	$59.64_{\pm 6.10}$
FID(x^{t_3})	$\mathbf{142.18_{\pm 7.01}}$	$179.42_{\pm 6.98}$	$265.78_{\pm 10.67}$
FID(x^{t_4})	$68.29_{\pm 6.97}$	$65.91_{\pm 6.68}$	$64.73_{\pm 6.91}$

Table 23: FID scores at training and validation steps (↓) in Sea Temperature (1995–1999).

Metric	CWG	GSBM	DSBM
FID(x^{t_0})	$51.79_{\pm 2.71}$	–	–
FID(x^{t_1})	$\mathbf{138.92_{\pm 7.35}}$	$176.14_{\pm 7.03}$	$257.26_{\pm 10.93}$
FID(x^{t_2})	$60.19_{\pm 6.01}$	$63.48_{\pm 6.12}$	$66.07_{\pm 6.57}$
FID(x^{t_3})	$\mathbf{154.37_{\pm 8.14}}$	$193.28_{\pm 8.47}$	$266.86_{\pm 11.34}$
FID(x^{t_4})	$69.57_{\pm 7.02}$	$66.83_{\pm 6.69}$	$66.18_{\pm 6.94}$

Table 24: FID scores at training and validation steps (↓) in Sea Temperature (1990–1994).

Metric	CWG	GSBM	DSBM
FID(x^{t_0})	$54.87_{\pm 2.93}$	–	–
FID(x^{t_1})	$\mathbf{145.59_{\pm 7.82}}$	$184.37_{\pm 7.46}$	$258.97_{\pm 11.26}$
FID(x^{t_2})	$72.38_{\pm 6.42}$	$78.69_{\pm 6.71}$	$82.13_{\pm 7.04}$
FID(x^{t_3})	$\mathbf{160.08_{\pm 7.97}}$	$181.67_{\pm 8.19}$	$236.46_{\pm 11.59}$
FID(x^{t_4})	$73.42_{\pm 7.19}$	$70.68_{\pm 6.87}$	$69.83_{\pm 7.09}$

Table 25: FID scores at training and validation steps (↓) in Sea Temperature (1985–1989).

Metric	CWG	GSBM	DSBM
FID(x^{t_0})	$56.97_{\pm 3.08}$	–	–
FID(x^{t_1})	$\mathbf{151.27_{\pm 8.13}}$	$188.79_{\pm 7.68}$	$262.47_{\pm 11.64}$
FID(x^{t_2})	$66.89_{\pm 6.31}$	$69.19_{\pm 6.47}$	$71.23_{\pm 7.08}$
FID(x^{t_3})	$157.58_{\pm 8.73}$	$160.37_{\pm 8.91}$	$258.96_{\pm 11.78}$
FID(x^{t_4})	$66.02_{\pm 7.41}$	$62.87_{\pm 7.16}$	$61.59_{\pm 7.34}$

Table 26: FID scores at training and validation steps (↓) in Sea Temperature (1981–1983).

Metric	CWG	GSBM	DSBM
FID(x^{t_0})	$59.97_{\pm 3.20}$	–	–
FID(x^{t_1})	$\mathbf{166.08_{\pm 8.91}}$	$188.79_{\pm 9.02}$	$252.59_{\pm 12.03}$
FID(x^{t_2})	$69.29_{\pm 6.65}$	$71.68_{\pm 6.74}$	$73.87_{\pm 7.26}$
FID(x^{t_3})	$\mathbf{168.73_{\pm 8.46}}$	$185.57_{\pm 8.08}$	$260.47_{\pm 11.92}$
FID(x^{t_4})	$77.83_{\pm 7.68}$	$74.97_{\pm 7.34}$	$74.29_{\pm 7.57}$

F.5 Robotic Task Reconstruction

BridgeData V2 (Walke et al., 2023) is a large and diverse dataset of robotic manipulation behaviors, designed to advance research in scalable robot learning. In this experiment, our goal is to reconstruct the full video of a robot performing manipulation tasks while training the network only on images from the beginning and end of the sequence, interpreted as samples from the endpoint distributions ρ_a and ρ_b. No intermediate marginals ρ_m are used. Following the Sea Temperature Prediction experiment, we introduce a potential function U that penalizes deviations from the learned data manifold. This manifold is learned from the initial and final frames of the videos, and the penalty encourages plausible intermediate frames consistent with these distributions. We employ a state-of-the-art VAE architecture, with parameters listed in Table 27, to model the image manifold. The potential function $U(x^t)$ for samples $x^t \sim \rho^t$ is defined as the squared distance between a bridge sample x^t and its VAE reconstruction $\tilde{x}^t = \text{VAE}(x^t)$: $U(x^t) = \|x^t - \tilde{x}^t\|^2$. Figure 7 presents snapshots of the reconstructions produced by our CWG method compared to the baselines. In this experiment, the Hamiltonian function H^{t_k} is held constant, as varying it yields no apparent benefit.

In the guided generation setting, we define the feature function f as

$$f(x_b) = \mathrm{ReLU}(w_c c_b^{(x)} - b_c), \tag{54}$$

where $c_b^{(x)}$ denotes the (x)-coordinate pixel position of the centroid corresponding to the target location of the item placed by the robot, extracted from the image x_b sampled from the reference marginal ρ_b. We impose a penalty f on this image (parameters of this function are available in Table 28) so that samples corresponding to placements in undesired locations are discouraged, while those leading to desirable targets are favored.

The centroid extraction is performed by applying morphological opening and closing operations from the OpenCV library to remove noise and refine object boundaries, followed by color-based masking to isolate the object of interest.

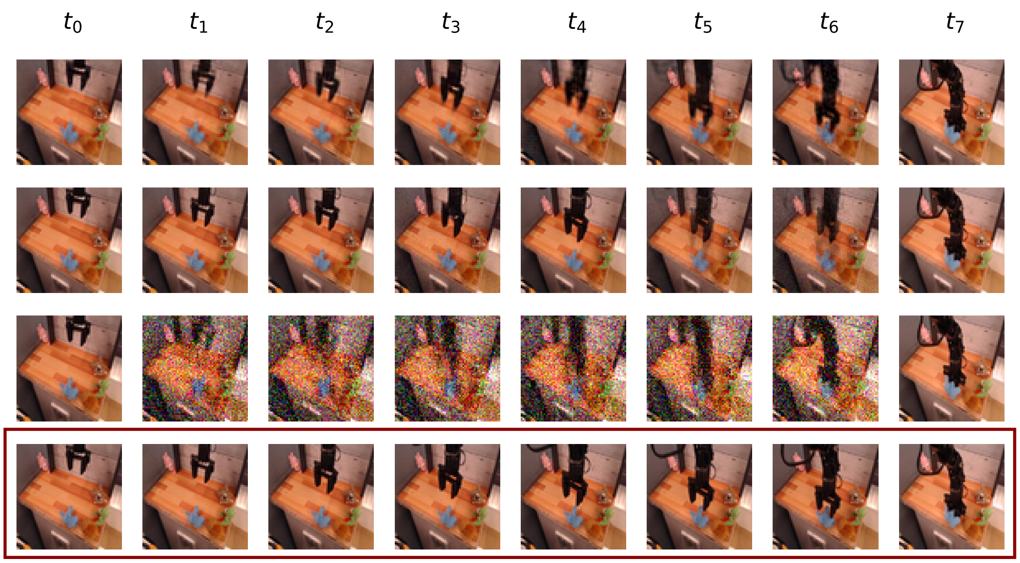

Figure 10: Reconstructions from CWG (top), GSBM (middle), and DSBM (bottom) in the Robot Task Reconstruction experiment. Red row shows the reference.

Table 27: Architecture of the ConvVAE used in the Robot Task Reconstruction experiment. All Conv2D and ConvTranspose2D layers use kernel size 4, stride 2, and padding 1, followed by ReLU activations (except the last layer, which uses Sigmoid).

Stage	Layer (channels)	Output size
Input	Single-channel image	$3 \times 64 \times 64$
Encoder	Conv2D (3→32)	$32 \times 32 \times 32$
	Conv2D (32→64)	$64 \times 16 \times 16$
	Conv2D (64→128)	$128 \times 8 \times 8$
	Flatten	8192
Latent space	Linear $\rightarrow \mu$	2
	Linear $\rightarrow \log \sigma^2$	2
Decoder	Linear → reshape	$128 \times 8 \times 8$
	ConvT2D (128→64)	$64 \times 16 \times 16$
	ConvT2D (64→32)	$32 \times 32 \times 32$
	ConvT2D (32→3), Sigmoid	$3 \times 64 \times 64$

Table 28: Parameters for the feature function f (54), used for guided generation in the Robot Task Reconstruction experiment.

Parameter	Value
Weight w_c	1
Bias b_c	30