The Schrödinger Bridge between Gaussian Measures has a Closed Form

Charlotte Bunne[*] ETH Zürich

Ya-Ping Hsieh[*] ETH Zürich

Marco Cuturi Apple[‡]

Andreas Krause ETH Zürich

Abstract

The static optimal transport (OT) problem between Gaussians seeks to recover an optimal map, or more generally a coupling, to morph a Gaussian into another. It has been well studied and applied to a wide variety of tasks. Here we focus on the dynamic formulation of OT, also known as the *Schrödinger bridge* (SB) problem, which has recently seen a surge of interest in machine learning due to its connections with diffusion-based generative models. In contrast to the static setting, much less is known about the dynamic setting, even for Gaussian distributions. In this paper, we provide *closed-form expressions for SBs between Gaussian measures*. In contrast to the static Gaussian OT problem, which can be simply reduced to studying convex programs, our framework for solving SBs requires significantly more involved tools such as Riemannian geometry and generator theory. Notably, we establish that the solutions of SBs between Gaussian measures are themselves Gaussian processes with explicit mean and covariance kernels, and thus are readily amenable for many downstream applications such as generative modeling or interpolation. To demonstrate the utility, we devise a new method for modeling the evolution of single-cell genomics data and report significantly improved numerical stability compared to existing SB-based approaches.

1 Introduction

The *Schrödinger bridge* (SB) (Léonard, 2013; Chen et al., 2021), alternatively known as the *dynamic* entropy-regularized optimal transport (OT), has recently received significant attention from the machine learning community. In contrast to the classical *static* OT where one seeks a coupling between measures that minimizes the average cost

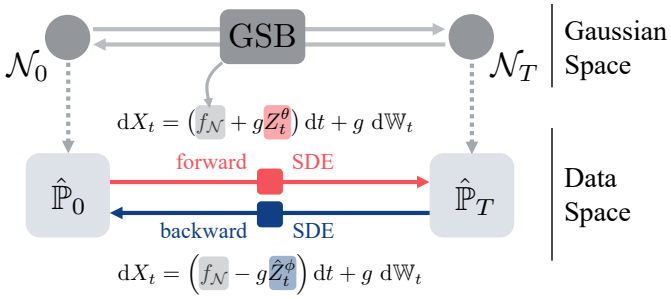


Figure 1: Solving the SB problem between $\hat{\mathbb{P}}_0$ and $\hat{\mathbb{P}}_1$ is notoriously difficult, because it requires learning the time-dependent drifts of two SDEs that respect the desired marginals, and a random initialization for these drifts is usually extremely far from satisfying that constraint. We propose a data-dependent procedure that relies first on Gaussian approximations of the data measures, which provide a closed-form drift $f_{\mathcal{N}}$ in (29) (the Gaussian Schrödinger bridge (GSB)). We show that this facilitates the training of forward/backward drifts $\hat{Z}_t^{\theta}, \hat{Z}_t^{\phi}$.

(Villani, 2009; Peyré and Cuturi, 2019), the goal of SBs is to find the optimal *stochastic processes* that evolve a given measure into another. As such, SBs are particularly suitable for learning complex continuous-time systems, and have been successfully applied to a wide range of applications such as sampling (Bernton et al., 2019; Huang et al., 2021b), generative modeling (Chen et al., 2022; De Bortoli et al., 2021b; Wang et al., 2021), molecular biology (Holdijk et al., 2022), and mean-field games (Liu et al., 2022).

Despite of these impressive achievements, a common limitation of the existing works is that the SBs are typically solved in a purely numerical fashion. In sharp contrast, it is well-known that many important OT problems for *Gaussian* measures admit *closed-form* solutions, and the advantages of such solutions are numerous: they have inspired new learning methods (Rabin et al., 2011; Vayer et al., 2019; Bonneel et al., 2015), they can serve as the ground truth for evaluating numerical schemes (Janati et al., 2020), and they have lead to the discovery of a new geometry that is both rich in theory and application (Takatsu, 2010).

[*]Equal contributions. [‡]Work done while at Google. Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

The goal of our paper is to continue this pursuit of closed-form solutions and thereby extending these advantages to SB-based learning methods. For an overview of the method, see Fig. 1. To this end, we make the following **contributions**:

1. As our central result, we derive the closed-form expressions for Gaussian Schrödinger bridges (GSBs), i.e., SBs between Gaussian measures. This is a challenging task for which all existing techniques fail, and thus we need to resort to a number of new ideas from entropic OT, Riemannian geometry, and generator theory; see Section 3.
2. We extend the deep connection between geometry and Gaussian OT to Gaussian Schrödinger bridges. In particular, our results can be seen as a vast generalization of the classical Bures-Wasserstein geodesics between Gaussian measures (Takatsu, 2010; Bhatia et al., 2019), which is the foundation of many computational methods (Chewi et al., 2020; Altschuler et al., 2021; Han et al., 2021).
3. Via a simple Gaussian approximation on real *single-cell genomics* data, we numerically demonstrate that many benefits of the closed-form expressions in static OT immediately carry over to SB-based learning methods: We report improved numerical stability and tuning insensitivity when trained on benchmark datasets, which ultimately lead to an overall better performance.

2 Preliminaries on Gaussian Optimal Transport Problems

Throughout this paper, let $\xi \sim \mathcal{N}(\mu, \Sigma)$ and $\xi' \sim \mathcal{N}(\mu', \Sigma')$ denote two given Gaussian random variables. By abusing the notation, we will continue to denote the measures of these Gaussians by ξ and ξ', respectively. We will also denote by $\Pi(\xi, \xi')$ the set of all their couplings.

2.1 Static Gaussian Optimal Transport

The *static* entropy-regularized OT between Gaussians refers to the following minimization problem (Peyré and Cuturi, 2019):

$$\min_{\pi \in \Pi(\xi,\xi')} \int \|x - x'\|^2 \mathrm{d}\pi(x, x') + 2\sigma^2 D_{\mathrm{KL}}(\pi \| \xi \otimes \xi'), \quad (1)$$

where $\xi \otimes \xi'$ denotes the product measure of ξ and ξ', and $\sigma \geq 0$ is a regularization parameter. When $\sigma = 0$, (1) reduces to the classical 2-Wasserstein distance between ξ and ξ' (Villani, 2009), whose closed-form solution is classical (Dowson and Landau, 1982; Olkin and Pukelsheim, 1982). The case for general σ is more involved, and an analytical expression was only recently found (Bojilov and Galichon, 2016; del Barrio and Loubes, 2020; Janati et al., 2020; Mallasto et al., 2021): Setting

$$D_\sigma := (4\Sigma^{\frac{1}{2}}\Sigma'\Sigma^{\frac{1}{2}} + \sigma^4 I)^{\frac{1}{2}}, \quad C_\sigma := \frac{1}{2}(\Sigma^{\frac{1}{2}} D_\sigma \Sigma^{-\frac{1}{2}} - \sigma^2 I), \quad (2)$$

then the solution $\pi^\star$ to (1) is itself a Gaussian:

$$\pi^\star \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu' \end{bmatrix}, \begin{bmatrix} \Sigma & C_\sigma \\ C_\sigma^\top & \Sigma' \end{bmatrix} \right). \quad (3)$$

2.2 Dynamic Gaussian Optimal Transport

In the literature, (1) is commonly referred to as the *static* OT formulation, since it merely asks *where* the mass should be transported to (i.e., $\pi(x, x')$ dictates how much mass at x should be transported to x'). In contrast, the more general problem of *dynamic* Gaussian OT seeks to answer *how* the mass the should be transported:

$$\min_{\rho_0 = \xi,\ \rho_1 = \xi'} \mathbb{E}\left[\int_0^1 \frac{1}{2}\|v_t\|^2 + \frac{\sigma^4}{8} \|\nabla \log \rho_t\|^2 \, \mathrm{d}t \right]. \quad (4)$$

Here, the minimization is taken over all pairs (ρ_t, v_t) where ρ_t is an absolutely continuous curve of measures (Ambrosio et al., 2006), and $v_t : \mathbb{R}^d \to \mathbb{R}^d$ is such that the continuity equation holds:

$$\partial_t \rho_t = -\nabla_x \cdot (\rho_t v_t), \quad (5)$$

where $(\nabla_x \cdot v_t)(x) := \sum_{i=1}^d \frac{\partial}{\partial x_i} v_t^i(x)$ denotes the divergence operator with respect to the x variable. It can be shown that, if $\rho_t^\star$ is the optimal curve for (4), then the joint distribution of the end marginals $(\rho_0^\star, \rho_1^\star)$ coincides with (3), hence the interpretation of $\rho_t^\star$ as the optimal *trajectory* in the space of measures (Chen et al., 2016; Gentil et al., 2017; Chen et al., 2021; Gentil et al., 2020).

To our knowledge, the only work that has partially addressed the closed-form solution of (4) is Mallasto et al. (2021), whose results are nonetheless insufficient to cover important applications such as generative modeling. In Section 5, we will derive a vast generalization of the results in Mallasto et al. (2021) and provide a detailed comparison in Sections 3–4.

3 The Gaussian Schrödinger Bridge Problem and Analysis Overview

The purpose of this section is to introduce the core objectives in our paper, the *Gaussian Schrödinger bridges*, and establish their connection to the Gaussian OT problems in Section 2. To help the reader navigate our somewhat technical proofs in Sections 4–5, we illustrate in Section 3.2 the high-level challenges as well as our new techniques for solving Gaussian Schrödinger bridges.

3.1 Schrödinger Bridges as Dynamic Entropy-Regularized Optimal Transport

Let ν, ν' be two given measures and let $\mathbb{Q}_t$ be an arbitrary stochastic process. In its most generic form, the Schrödinger bridge refers to the following constrained KL-minimization problem over all stochastic processes $\mathbb{P}_t$ (Léonard, 2013; Chen et al., 2021):

$$\min_{\mathbb{P}_0=\nu,\ \mathbb{P}_1=\nu'} D_{\mathrm{KL}}(\mathbb{P}_t \| \mathbb{Q}_t). \tag{6}$$

In practice, ν and ν' typically arise as the (empirical) *marginal* distributions of a complicated continuous-time dynamics observed at the starting and end times, and $\mathbb{Q}_t$ is a "prior process" representing our belief of the dynamics before observing any data. The solution $\mathbb{P}_t^\star$ to (6) is thus interpreted as the best dynamics that conforms to the prior belief $\mathbb{Q}_t$ while respecting the data marginals ($\mathbb{P}_0^\star = \nu, \mathbb{P}_1^\star = \nu'$).

In this paper, we will consider a general class of $\mathbb{Q}_t$'s that includes most existing processes in the machine learning applications of SBs. Specifically, with some initial condition Y_0, we will take $\mathbb{Q}_t$ to be the measure of the linear *stochastic differential equation* (SDE):

$$\mathrm{d}Y_t = (c_t Y_t + \alpha_t)\,\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t := f_t\,\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t. \tag{7}$$

Here, $c_t : \mathbb{R}^+ \to \mathbb{R}$, $\alpha_t : \mathbb{R}^+ \to \mathbb{R}^d$, and $g_t : \mathbb{R}^+ \to \mathbb{R}^+$ are smooth functions. In this case, SBs can be seen as generalized dynamical OT between two (not necessarily Gaussian) measures:

Theorem 1. *Consider the Schrödinger bridge problem with Y_t as the reference process:*

$$\min_{\mathbb{P}_0=\nu,\ \mathbb{P}_1=\nu'} D_{\mathrm{KL}}(\mathbb{P}_t \| Y_t). \tag{8}$$

Then (8) is equivalent to

$$\inf_{(\rho_t, v_t)} \mathbb{E}\left[\int_0^1 \frac{\|v_t\|^2}{2g_t^2} + \frac{g_t^2}{8}\|\nabla \log \rho_t\|^2 - \frac{1}{2}\langle f_t, \nabla \log \rho_t\rangle\,\mathrm{d}t\right] \tag{9}$$

where the infimum is taken all pairs (ρ_t, v_t) such that $\rho_0 = \nu, \rho_1 = \nu'$, ρ_t absolutely continuous, and

$$\partial_t \rho_t = -\nabla_x \cdot (\rho_t (f_t + v_t)). \tag{10}$$

The proof of Theorem 1, which we defer to Appendix A, is a straightforward extension of the argument in (Léonard, 2013; Chen et al., 2016; Gentil et al., 2017) which establishes the equivalence when Y_t is a reversible Brownian motion, i.e., $f_t \equiv 0, g_t \equiv \sigma$, and Y_0 follows the Lebesgue measure.[1]

[1]The reversible Brownian motion is a technical construct to simplify the computations. For our purpose, one can think of $Y_0 \sim \xi$ instead of the Lebesgue measure, and our results still hold verbatim.

3.2 The Gaussian Schrödinger Bridge Problem

The central goal of our paper is to derive the closed-form solution of SBs when the marginal constraints are Gaussians $\xi \sim \mathcal{N}(\mu, \Sigma)$, $\xi' \sim \mathcal{N}(\mu', \Sigma')$. Namely, we are interested in the following class of the SBs, termed Gaussian Schrödinger bridges:

$$\min_{\mathbb{P}_0=\xi,\ \mathbb{P}_1=\xi'} D_{\mathrm{KL}}(\mathbb{P}_t \| Y_t). \tag{GSB}$$

To emphasize the dependence on the reference SDE, we will sometimes call (GSB) the Y_t-GSB.

Technical challenges; related work. In order to analyze (GSB), we first notice that the objective in (9) becomes $\sigma^{-2}\,\mathbb{E}\left[\int_0^1 \frac{1}{2}\|v_t\|^2 + \frac{\sigma^4}{8}\|\nabla \log \rho_t\|^2\,\mathrm{d}t\right]$ for $\sigma\mathbb{W}_t$-GSBs. Up to a constant factor, this is simply (4), so Theorem 1 reduces to the well-known fact that $\sigma\mathbb{W}_t$-GSBs are a reformulation of the dynamic Gaussian OT (Léonard, 2013; Chen et al., 2016; Gentil et al., 2017).

At first sight, this might suggest that one can extend existing tools in Gaussian OT to analyze GSBs. Unfortunately, the major difficulty of tackling GSBs is that these existing tools are fundamentally insufficient for the generalized objective (9). To be more precise, there exist three prominent frameworks for studying Gaussian OT problems:

- **Convex analysis:** An extremely fruitful observation in the field is that many Gaussian OT instances can be reduced to a *convex* program, for which one can import various convex techniques such as KKT or fixed-point arguments. This is the case for static Gaussian OT (1), both when $\sigma = 0$ (Dowson and Landau, 1982; Olkin and Pukelsheim, 1982; Bhatia et al., 2019) and $\sigma > 0$ (Janati et al., 2020). Furthermore, in the case of $\sigma = 0$, the solution to the dynamic formulation (4) can be recovered from the static one via a simple linear interpolation (McCann, 1997).
- **Ad hoc computations:** When $\sigma > 0$ in (4), the problem is no longer reducible to a convex program (Léonard, 2013; Chen et al., 2021). In this case, the only technique we are aware of is the ad hoc approach of (Mallasto et al., 2021), which manages to find a closed form for (4) (and thus $\sigma\mathbb{W}_t$-GSBs) through a series of brute-force computations.
- **Control theory:** On a related note, in a series of papers, Chen et al. (2015, 2016, 2019) exploit the deep connection between $\sigma\mathbb{W}_t$-GSBs and control theory to study the *existence* and *uniqueness* of the solutions. Although a variety of new optimality conditions are derived in these works, they are all expressed in terms differential equations with coupled initial conditions, and it is unclear whether solving these differential equations is an easier task than (GSB) itself. In particular, no closed-form, even for $\sigma\mathbb{W}_t$-GSBs, can be found therein.

By Theorem 1, GSBs are more general than (4) and thus irreducible to convex programs, so there is no hope for the convex route. As for ad hoc computations, the time-dependent f_t and g_t terms in (9) present a serious obstruction for generalizing the approach of Mallasto et al. (2021) to Y_t-GSBs when $f_t \neq 0$ or g_t is not constant; this is exemplified by the convoluted expressions in our Theorem 3, which hopefully will convince the reader that they are beyond any ad hoc guess. Finally, the control-theoretic view has so far fallen short of producing closed-form solutions even for $\sigma\mathbb{W}_t$-GSBs, so it is essentially irrelevant for our purpose.

To conclude, in order to find an analytic expression for general GSBs, we will need drastically different techniques.

Our approach. To overcome the aforementioned challenges, in Section 4, we will first develop a principled framework for analyzing the closed-form expressions of $\sigma\mathbb{W}_t$-GSBs, i.e., (4). Unlike the ad hoc approach of Mallasto et al. (2021) which is very specific to Brownian motions, our analysis reveals the general role played by the *Lyapunov operator* (see (14)) on covariance matrices, thereby essentially reducing the solutions of GSBs to solving a matrix equation. This route is enabled via yet another equivalent formulation of (4), namely the action minimization problem on the *Bures-Wasserstein geometry*, which has recently emerged as a rich source for inspiring new computational methods (Chewi et al., 2020; Altschuler et al., 2021; Han et al., 2021). In Section 5, we show how the insight gained from our geometric framework in Section 4 can be easily adapted to GSBs with general reference processes, which ultimately leads to the full resolution of (GSB).

4 The Bures-Wasserstein Geometry of $\sigma\mathbb{W}_t$-Gaussian Schrödinger Bridges

This section illustrates the simple geometric intuition that underlies the somewhat technical proof of our main result (cf. Theorem 3). After briefly reviewing the action minimization problems on Euclidean spaces in Section 4.1, we present the main observation in Section 4.2: $\sigma\mathbb{W}_t$-GSBs are but action minimization problems on the Bures-Wasserstein manifolds, which can be tackled by following a standard routine in physics.

4.1 A Brief Review on Action Minimization Problems

Consider the following *action minimization* problem with fixed endpoints $x, x' \in \mathbb{R}^d$:

$$\min_{x(0)=x, x(1)=x'} \int_0^1 \frac{1}{2}\|\dot{x}(t)\|^2 - U(x(t))\,\mathrm{d}t, \qquad (11)$$

where the minimum is taken over all piecewise smooth curves. A celebrated result in physics asserts that the optimal curve for (11) satisfies the *Euler-Lagrange* equation:

$$\ddot{x}(t) = -\nabla U(x(t)), \quad x(0) = x, \quad x(1) = x'. \qquad (12)$$

In particular, when $U \equiv 0$, (12) reduces to $\ddot{x} \equiv 0$, i.e., $x(t)$ is a straight line connecting x and x'.

More generally, one can consider (11) on any *Riemannian manifold*, provided that the Euclidean norm $\|\cdot\|$ in (11) is replaced by the corresponding Riemannian norm. In this case, the Euler-Lagrange equation (12) still holds, with $\ddot{x}$ and ∇U replaced with their Riemannian counterparts (Villani, 2009).

4.2 $\sigma\mathbb{W}_t$-GSBs as Action Minimization Problems

We begin with the following simple observation. Based on the seminal work by Otto (2001), Gentil et al. (2020) show that SBs between two arbitrary measures can be formally understood as an action minimization problem of the form (11) on an *infinite*-dimensional manifold. Since we have restricted the measures in (GSB) to be Gaussian, and since Gaussian measures are uniquely determined by their means and covariances, Gentil et al. (2020) strongly suggests a *finite*-dimensional geometric interpretation of $\sigma\mathbb{W}_t$-GSBs. The main result in this section, Theorem 2 below, makes this link precise.

The proper geometry we need is the *Bures-Wasserstein manifold* (Takatsu, 2010; Bhatia et al., 2019) defined as follows. Consider the space of covariance matrices (i.e., symmetric positive definite matrices) of dimension d, which we denote by $\mathbb{S}^d_{++}$, and consider its natural tangent space as the space of symmetric matrices:

$$\mathcal{T}_\Sigma \mathbb{S}^d_{++} := \{U \in \mathbb{R}^{d\times d} : U^\top = U\}. \qquad (13)$$

A notion that will play a pivotal role is the so-called *Lyapunov operator*: For any $\Sigma \in \mathbb{S}^d_{++}$ and $U \in \mathcal{T}_\Sigma \mathbb{S}^d_{++}$, we define $\mathcal{L}_\Sigma[U]$ to be the symmetric solution to the equation

$$A: \quad \Sigma A + A\Sigma = U. \qquad (14)$$

It is shown in Takatsu (2010) that the Lyapunov operator defines a geometry on $\mathbb{S}^d_{++}$, known as the *Bures-Wasserstein geometry*: For any two tangent vectors $U, V \in \mathcal{T}_\Sigma \mathbb{S}^d_{++}$, the operation

$$\langle U, V\rangle_\Sigma := \frac{1}{2}\operatorname{tr}\mathcal{L}_\Sigma[U]V \qquad (15)$$

satisfies all the axioms of the Riemannian metric; additional background on the Bures-Wasserstein geometry can be found in Appendix B.1.

We are now ready to state the main result of the section. Let $\|\cdot\|_\Sigma$ be the induced norm of $\langle\cdot,\cdot\rangle_\Sigma$. Fix $\sigma > 0$ and let $\mathbb{W}_t$ be a reversible Brownian motion. Consider the following special case of (GSB):

$$\min_{\mathbb{P}_0=\mathcal{N}(0,\Sigma),\ \mathbb{P}_1=\mathcal{N}(0,\Sigma')} D_{\mathrm{KL}}(\mathbb{P}_t\|\sigma\mathbb{W}_t). \qquad (16)$$

Then we have:

Theorem 2. *The minimizer of* (16) *(and hence* (4)*) coincides with the solution of the action minimization problem:*

$$\min_{\Sigma_0=\Sigma,\Sigma_1=\Sigma'} \int_0^1 \frac{1}{2}\|\dot{\Sigma}_t\|^2_{\Sigma_t} - \mathcal{U}_\sigma(\Sigma_t)\,\mathrm{d}t \tag{17}$$

where $\mathcal{U}_\sigma(\Sigma_t) := -\frac{\sigma^4}{8}\operatorname{tr}\Sigma_t^{-1}$ *and the minimum is taken over all piecewise smooth curves in* $\mathbb{S}^d_{++}$*. In particular, the minimizer of* (16) *solves the Euler-Lagrange equation in the Bures-Wasserstein geometry:*

$$\nabla_{\dot{\Sigma}_t}\dot{\Sigma}_t = -\operatorname{grad}\mathcal{U}_\sigma(\Sigma_t), \quad \Sigma_0 = \Sigma, \quad \Sigma_1 = \Sigma', \tag{18}$$

where $\nabla_{\dot{\Sigma}_t}\dot{\Sigma}_t$ *denotes the Riemannian acceleration and* grad *the Riemannian gradient in the Bures-Wasserstein sense.*

An important implication. As alluded to in Section 3, the solution curve to (4) or (16) is not new; it is derived in Mallasto et al. (2021) via a strenuous and rather unenlightening calculation:

$$\Sigma_t := \bar{t}^2\Sigma + t^2\Sigma' + t\cdot\bar{t}\big(C_\sigma + C_\sigma^\top + \sigma^2 I\big). \tag{19}$$

Here, $\bar{t} := 1 - t$ and C_σ is defined in (2). However, the interpretation of (19) as the minimizer of (17) is new and suggests a principled avenue towards the closed-form solution of $\sigma\mathbb{W}_t$-GSBs: solve the Euler-Lagrange equation (18). Inspecting the formulas for $\nabla_{\dot{\Sigma}_t}\dot{\Sigma}_t$ and $\operatorname{grad}\mathcal{U}_\sigma(\Sigma_t)$ (see (B.5) and (B.6)), one can further reduce (18) to computing the Lyapunov operator $\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]$, which presents the bottleneck in the proof of Theorem 2 as there is, in general, no closed form for the matrix equation (14). To this end, our main contribution is the following technical Lemma:

Lemma 1. *Define the matrix* $\tilde{S}_t$ *to be:*

$$\tilde{S}_t := t\Sigma' + \bar{t}C_\sigma - \bar{t}\Sigma - tC_\sigma^\top + \frac{\sigma^2}{2}(\bar{t} - t)I. \tag{20}$$

Then $\tilde{S}_t^\top\Sigma_t^{-1}$ *is symmetric.*

Armed with Lemma 1, it is straightforward to verify that $\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t] = \tilde{S}_t^\top\Sigma_t^{-1}$, i.e., $\tilde{S}_t^\top\Sigma_t^{-1}$ is symmetric and satisfies:

$$\tilde{S}_t^\top\Sigma_t^{-1}\cdot\Sigma_t^{-1} + \Sigma_t^{-1}\cdot\Sigma_t^{-1}\tilde{S}_t = \tilde{S}_t^\top + \tilde{S}_t = \dot{\Sigma}_t \tag{21}$$

which is more or less equivalent to the original Euler-Lagrange equation (18); we defer the details to Appendix B.2.

To conclude, in contrast to the purely technical approach of Mallasto et al. (2021), our Theorem 2 provides a geometric and conceptually clean solution for $\sigma\mathbb{W}_t$-GSBs: Compute the Lyapunov operator $\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]$ via verifying the symmetry of the matrix in Lemma 1. *It turns out that this technique can be readily extended to general GSBs*, and therefore serves as the foundation for the proof of our main result; see Section 5.

Remark. It is interesting to note that the matrix $\tilde{S}_t$ in (20) is itself *not* symmetric. Other consequences of Theorem 2 that might be of independent interest can be found in Appendix B.3. We also note that, when $\sigma = 0$, the solution to (17) is simply the Wasserstein geodesic between Gaussian measures, whose formula is well-known (Dowson and Landau, 1982; Takatsu, 2010). However, as explained in Section 3, the case of $\sigma > 0$ requires a completely different analysis since, unlike when $\sigma = 0$, it is not reducible to a convex program. This leads to the significantly more involved proofs of Theorem 2 and of (19) in Mallasto et al. (2021).

5 Closed-Form Solutions of General Gaussian Schrödinger Bridges

We now present the closed-form solutions of general GSBs.

5.1 Linear Stochastic Differential Equations

We need the following background knowledge on the linear SDE Y_t. Let $\tau_t := \exp\Big(\int_0^t c_s\,\mathrm{d}s\Big)$. Then the solution to (7) is (Platen and Bruti-Liberati, 2010):

$$Y_t = \tau_t\bigg(Y_0 + \int_0^t \tau_s^{-1}\alpha_s\,\mathrm{d}s + \int_0^t \tau_s^{-1}g_s\,\mathrm{d}\mathbb{W}_s\bigg). \tag{22}$$

Another crucial fact in our analysis is that Y_t is a *Gaussian process given* Y_0, and is thus characterized by the first two moments. Using the independent increments of $\mathbb{W}_t$ and Itô's isometry (Protter, 2005), we compute:

$$\mathbb{E}[Y_t\,|\,Y_0] = \tau_t\bigg(Y_0 + \int_0^t \tau_s^{-1}\alpha_s\,\mathrm{d}s\bigg) =: \eta(t) \tag{23}$$

and, for any $t' \geq t$,

$$\begin{aligned}\mathbb{E}\Big[(Y_t - \eta(t))(Y_{t'} - \eta(t'))^\top\,\Big|\,Y_0\Big] \\ = \bigg(\tau_t\tau_{t'}\int_0^t \tau_s^{-2}g_s^2\,\mathrm{d}s\bigg)I =: \kappa(t,t')I.\end{aligned} \tag{24}$$

5.2 Main Result

We now present the main result of our paper. With the important application of diffusion-based models in mind, we will not only derive solution curves as in (19) but also their SDE representations.

Let $\xi = \mathcal{N}(\mu_0,\Sigma_0)$ and $\xi' = \mathcal{N}(\mu_1,\Sigma_1)$ be two arbitrary Gaussian distributions in (GSB), and let D_σ, C_σ be as defined in (2).

Theorem 3. *Denote by* $\mathbb{P}_t$ *the solution to Gaussian*

SDE WITH $\alpha_t \equiv 0$	SETTING	$\kappa(t,t')$	$\sigma_\star^2$	r_t	$\bar{r}_t$	ρ_t	$\zeta(t)$
BM	$\substack{c_t\equiv 0\\ g_t\equiv\omega\in\mathbb{R}^+}$	$\omega^2 t$	ω^2	t	$1-t$	t	0
VESDE	$\substack{c_t\equiv 0\\ g_t=\sqrt{\dot{\mathrm{q}}(t)}}$	$\mathrm{q}(t)$	$\mathrm{q}(1)$	$\frac{\mathrm{q}(t)}{\mathrm{q}(1)}$	$1-\frac{\mathrm{q}(t)}{\mathrm{q}(1)}$	$\frac{\mathrm{q}(t)}{\mathrm{q}(1)}$	0
VPSDE	$-2c_t=g_t^2$	$\tau_{t'}\left(\tau_t^{-1}-\tau_t\right)$	$\tau_1^{-1}-\tau_1$	$\frac{\tau_t^{-1}-\tau_t}{\tau_1^{-1}-\tau_1}$	$\tau_1\left(\frac{\tau_t}{\tau_1}-\frac{\tau_t^{-1}-\tau_t}{\tau_1^{-1}-\tau_1}\right)$	$\frac{\tau_t^{-1}\left(\tau_t^{-1}-\tau_t\right)}{\tau_1^{-1}\left(\tau_1^{-1}-\tau_1\right)}$	0
SUB–VPSDE	$\frac{g_t^2}{-2c_t}=1-\tau_t^4$	$\tau_t\tau_{t'}\left(\tau_t^{-1}-\tau_t\right)^2$	$\tau_1\left(\tau_1^{-1}-\tau_1\right)^2$	$\frac{\tau_t}{\tau_1}\cdot\left(\frac{\tau_t^{-1}-\tau_t}{\tau_1^{-1}-\tau_1}\right)^2$	$\tau_t\left(1-\left(\frac{\tau_t^{-1}-\tau_t}{\tau_1^{-1}-\tau_1}\right)^2\right)$	$\left(\frac{\tau_t^{-1}-\tau_t}{\tau_1^{-1}-\tau_1}\right)^2$	0
SDE WITH $\alpha_t \not\equiv 0$	SETTING	$\kappa(t,t')$	$\sigma_\star^2$	r_t	$\bar{r}_t$	ρ_t	$\zeta(t)$
OU/VASICEK	$\substack{c_t\equiv-\lambda\in\mathbb{R}\\ \alpha_t\equiv\mathbf{v}\in\mathbb{R}^d\\ g_t\equiv\omega\in\mathbb{R}^+}$	$\frac{\omega^2 e^{-\lambda t'}\sinh\lambda t}{\lambda}$	$\frac{\omega^2\sinh\lambda}{\lambda}$	$\frac{\sinh\lambda t}{\sinh\lambda}$	$\substack{\sinh\lambda t\coth\lambda t\\ -\sinh\lambda t\coth\lambda}$	$\substack{e^{-\lambda(1-t)}\\ \cdot\frac{\sinh\lambda t}{\sinh\lambda}}$	$\frac{\mathbf{v}}{\lambda}\left(1-e^{-\lambda t}\right)$
α_t-BDT	$\substack{c_t\equiv 0\\ g_t\equiv\omega\in\mathbb{R}^+}$	$\omega^2 t$	$\omega^2 1$	t	$1-t$	t	$\int_0^t \alpha_s\,\mathrm{d}s$

Table 1: Examples of reference SDEs and the corresponding solutions of GSBs. All relevant functions in the Table are either introduced in Section 5.1 or (25).

Schrödinger bridges (GSB). *Set*

$$r_t := \frac{\kappa(t,1)}{\kappa(1,1)}, \quad \bar{r}_t := \tau_t - r_t\tau_1, \quad \sigma_\star := \sqrt{\tau_1^{-1}\kappa(1,1)},$$
$$\zeta(t) := \tau_t\int_0^t \tau_s^{-1}\alpha_s\,\mathrm{d}s, \rho_t := \frac{\int_0^t \tau_s^{-2}g_s^2\,\mathrm{d}s}{\int_0^1 \tau_s^{-2}g_s^2\,\mathrm{d}s},$$
$$P_t := \dot{r}_t(r_t\Sigma_1 + \bar{r}_t C_{\sigma_\star}), \quad Q_t := -\dot{\bar{r}}_t(\bar{r}_t\Sigma_0 + r_t C_{\sigma_\star}),$$
$$S_t := P_t - Q_t^\top + \left[c_t\kappa(t,t)(1-\rho_t) - g_t^2\rho_t\right]I. \tag{25}$$

Then the following holds:

1. *The solution $\mathbb{P}_t$ is a Markov Gaussian process whose marginal variable $X_t \sim \mathcal{N}(\mu_t, \Sigma_t)$, where*

$$\mu_t := \bar{r}_t\mu_0 + r_t\mu_1 + \zeta(t) - r_t\zeta(1), \tag{26}$$
$$\Sigma_t := \bar{r}_t^2\Sigma_0 + r_t^2\Sigma_1 + r_t\bar{r}_t\left(C_{\sigma_\star} + C_{\sigma_\star}^\top\right) + \kappa(t,t)(1-\rho_t)I. \tag{27}$$

2. *X_t admits a closed-form representation as the SDE:*

$$\mathrm{d}X_t = f_{\mathcal{N}}(t, X_t)\,\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t \tag{28}$$

where

$$f_{\mathcal{N}}(t,x) := S_t^\top\Sigma_t^{-1}(x-\mu_t) + \dot{\mu}_t. \tag{29}$$

Moreover, the matrix $S_t^\top\Sigma_t^{-1}$ is symmetric.

As in Theorem 2, the key step in the proof of Theorem 3 is to recognize the symmetry of the matrix $S_t^\top\Sigma_t^{-1}$ where S_t, defined in (25), simply becomes the $\tilde{S}_t$ in Lemma 1 (up to an additive factor of $\frac{\sigma^2\bar{t}}{2}I$) for $\sigma\mathbb{W}_t$-GSBs. Although this can be directly verified via generalizing Lemma 1, the computation becomes quite tedious, so our proof of Theorem 3 will follow a slightly different route. In any case, given the symmetry of $S_t^\top\Sigma_t^{-1}$, the proof simply boils down to a series of straightforward calculations; see Appendix C.

Closed forms for conditional distributions. In many practical applications such as generative modeling, a requirement to employ the SDE representation of GSBs in (28) is that its *conditional distributions* given the initial points can be computed efficiently. As an immediate corollary of Theorem 3, we obtain the following closed-form expressions for these conditional distributions.

Corollary 1. *Let $X_t \sim \mathbb{P}_t$ be the the solution to* (GSB). *Then the conditional distribution of X_t given end points has a simple solution: $X_t|X_0 = x_0 \sim \mathcal{N}\left(\mu_{t|0}, \Sigma_{t|0}\right)$, where*

$$\mu_{t|0} = \bar{r}_t x_0 + r_t\left(\mu_1 + C_{\sigma_\star}^\top\Sigma_0^{-1}(x_0-\mu_0)\right) + \zeta(t) - r_t\zeta(1), \tag{30}$$
$$\Sigma_{t|0} = r_t^2\left(\Sigma_1 - C_\sigma^\top\Sigma_0^{-1}C_\sigma\right) + \kappa(t,t)(1-\rho_t)I. \tag{31}$$

Similarly, $X_t|X_1 = x_1 \sim \mathcal{N}\left(\mu_{t|1}, \Sigma_{t|1}\right)$, where

$$\mu_{t|1} = r_t x_1 + \bar{r}_t\left(\mu_0 + C_{\sigma_\star}\Sigma_1^{-1}(x_1-\mu_1)\right) + \zeta(t) - r_t\zeta(1), \tag{32}$$
$$\Sigma_{t|1} = \bar{r}_t^2\left(\Sigma_0 - C_\sigma\Sigma_1^{-1}C_\sigma^\top\right) + \kappa(t,t)(1-\rho_t)I. \tag{33}$$

Examples of GSBs. Our framework captures most popular reference SDEs in the machine learning literature as well as other mathematical models in financial engineering; see Table 1. A non-exhaustive list includes:

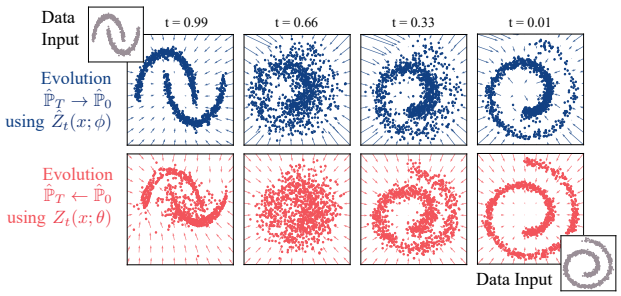


Figure 2: Illustration of the time-dependent drifts learned by GSBFLOW with VE SDE for two toy marginal distributions. *Top.* Evolution of $\hat{\mathbb{P}}_1$ (moons) $\to \hat{\mathbb{P}}_0$ (spiral) via backward policy $\hat{Z}_t^\phi(x)$. *Bottom.* Evolution of $\hat{\mathbb{P}}_0$ (spiral) $\to \hat{\mathbb{P}}_1$ (moons) via forward policy $Z_t^\theta(x)$.

- The basic *Brownian motion* (BM) and the *Ornstein–Uhlenbeck* (OU) processes, both widely adopted as the reference process for SB-based models (De Bortoli et al., 2021a,b; Lavenant et al., 2021; Vargas et al., 2021; Wang et al., 2021). We also remark that, even though (27) is known for BM (Mallasto et al., 2021), what is crucial in these applications is the SDE presentation (28), which is new even for BM.
- The *variance exploding SDEs* (VESDEs), which underlies the training of *score matching with Langevin dynamics* for diffusion-based generative modeling (Huang et al., 2021a; Song and Ermon, 2019; Song et al., 2021).
- The *variance preserving SDEs* (VPSDEs), which can be seen as the continuous limit of *denoising diffusion probabilistic models* (Ho et al., 2020; Sohl-Dickstein et al., 2015; Song et al., 2021), another important class of algorithms for diffusion-based generative modeling.
- The *sub-VPSDEs* proposed by (Song et al., 2021), which are motivated by reducing the variance of VPSDEs.
- Several important SDEs in financial engineering, such as the *Vasicek model* (which generalizes OU processes) and the *constant volatility* α_t*-Black–Derman–Toy* (BDT) *model* (Platen and Bruti-Liberati, 2010).

6 Empirical Evaluation

The purpose of our experiments is to demonstrate that, by leveraging moment information, GSBFLOW is significantly more stable compared to other SB-based objectives, especially when moving beyond the *generative* setting where $\hat{\mathbb{P}}_1$ is a simple Gaussian. Indeed, while performing competitively in the generative setting ($\mathcal{N}_0 \to \hat{\mathbb{P}}_1$), our method *outperforms* when modeling the evolution of two complex distributions ($\hat{\mathbb{P}}_0 \to \hat{\mathbb{P}}_1$), the most general and ambitious setting to estimate a bridge. This is demonstrated on synthetic data as well as a task from molecular biology concerned with modeling the dynamics of cellular systems, i.e., single-cell genomics (Macosko et al., 2015; Frangieh et al., 2021; Kulkarni et al., 2019).

Table 2: Evaluation of predictive performance w.r.t. the entropy-regularized Wasserstein distance W_ε (Cuturi, 2013) of GSBFLOW and baselines on generating different single-cell datasets (using 3 runs).

Method	**Tasks**	
	Wasserstein Loss $W_\varepsilon \downarrow$	
	Moon et al. (2019)	Schiebinger et al. (2019)
Song et al. (2021)		
VESDE	20.83 ± 0.18	40.81 ± 0.42
sub-VPSDE	$\mathbf{19.96 \pm 0.58}$	48.15 ± 3.38
GSBFLOW (*ours*)		
VESDE	25.18 ± 0.10	$\mathbf{27.85 \pm 0.68}$

6.1 Synthetic Dynamics

Before conducting the single-cell genomics experiments, we first test GSBFLOW on a synthetic setting. Our first task involves recovering the stochastic evolution of two-dimensional synthetic data containing two interleaving half circles ($\hat{\mathbb{P}}_1$) into a spiral ($\hat{\mathbb{P}}_0$). Fig. 2 shows the trajectories learned by GSBFLOW based on the VESDE (see Table 1 and Appendix D.5.1).

While it is sufficient to parameterize only a single policy ($\hat{Z}_t^\phi(x)$) in generative modeling, the task of learning to evolve $\hat{\mathbb{P}}_0$ into $\hat{\mathbb{P}}_1$ requires one to recover *both* vector fields $\hat{Z}_t^\phi(x)$ and $Z_t^\theta(x)$. As demonstrated in Fig. 2, GSBFLOW is able to successfully learn both policies $Z_t^\theta(x)$ and $\hat{Z}_t^\phi(x)$ and reliably recovers the corresponding targets of the forward and backward evolution. While initializing the reference process through the closed-form SB between the Gaussian approximations of both synthetic datasets provides good results, the power of GSBFLOW becomes evident in more complex applications which we tackle next.

6.2 Single-Cell Dynamics

Modern single-cell profiling technologies are able to provide rich feature representations (e.g., gene expression) of *individual* cells at any development state. A crucial issue that arises with such profiling methods is their destructive nature: Measuring a cell requires destroying it and thus a cell cannot be measured twice. As a result, independent samples are collected at each snapshot, with no access to ground-truth single-cell trajectories throughout time, resulting in challenging, *unaligned*, datasets. Recovering cellular dynamics from such unaligned snapshots, i.e., $\hat{\mathbb{P}}_0$ to $\hat{\mathbb{P}}_1$, has, however, extremely important scientific and biomedical relevance (Kulkarni et al., 2019). For example, it determines our understanding on how and why tumor cells evade cancer therapies (Frangieh et al., 2021) or

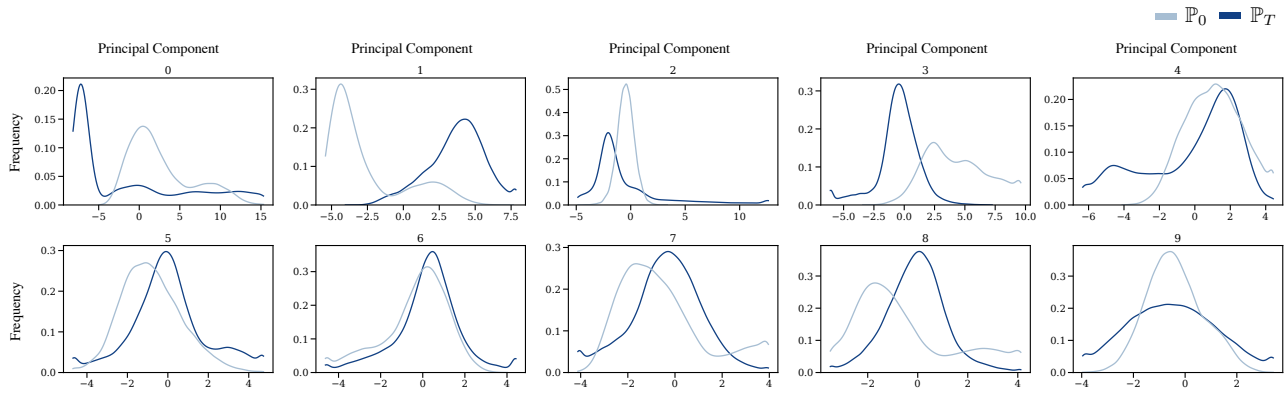


Figure 3: The expression levels of the first 10 principal components from the dataset by Schiebinger et al. (2019).

unveils mechanisms of cell differentiation and development (Schiebinger et al., 2019). Following related work, in particular previous methods based on optimal transport (Schiebinger et al., 2019; Bunne et al., 2021, 2022; Tong et al., 2020), the task is thus to learn the stochastic process that described the evolution of single cells from $\hat{\mathbb{P}}_0$ to $\hat{\mathbb{P}}_1$.

6.2.1 Experimental Setup

Single-cell genomics via SBs. Let us consider the evolution of a gene, for which we can collect the empirical distributions $\hat{\mathbb{P}}_0, \hat{\mathbb{P}}_1$ of its expression levels at the times $t = 0, 1$ (Schiebinger et al., 2019; Moon et al., 2019). Our goal is to two-fold:

1. To solve the **generative modeling** problem, i.e., to generate $\hat{\mathbb{P}}_0$ or $\hat{\mathbb{P}}_1$ from a standard Gaussian noise, and
2. to **evolve** $\mathbb{P}_0 \to \mathbb{P}_1$ or $\mathbb{P}_1 \to \mathbb{P}_0$, i.e., to recover a stochastic process $\mathbb{P}_t$ satisfying $\mathbb{P}_0 = \hat{\mathbb{P}}_0, \mathbb{P}_1 = \hat{\mathbb{P}}_1$.

Although there are numerous algorithms for generative modeling, to our knowledge, the only framework that can simultaneously solve both tasks is the SB-based scheme recently proposed in (Chen et al., 2022). In order to apply this framework, one has to choose a prior process Y_t, which is taken by the authors to be the high-performing VESDE and sub-VPSDE. These SB-based methods, as well as several standard generative modeling algorithms (Ho et al., 2020; Sohl-Dickstein et al., 2015; Song et al., 2021; Huang et al., 2021a; Song and Ermon, 2019; Song et al., 2021) for the first task, constitute strong baselines for our experiments.

Our choice of Y_t; the GSBFLOW. Instead of directly diving into the numerical solution of SBs as in Chen et al. (2022), we first empirically verify that the distributions $\hat{\mathbb{P}}_0, \hat{\mathbb{P}}_1$ in single-cell genomics are typically close to *non-standard* Gaussian distributions: See Fig. 3 for the canonical dataset (Schiebinger et al., 2019) and Fig. 5 in Appendix D.1 for the same phenomenon on another standard benchmark (Moon et al., 2019).

Since the solutions of SBs are Lipschitz in terms of $\hat{\mathbb{P}}_0, \hat{\mathbb{P}}_1$ (Carlier et al., 2022), a reasonable approximation to the original SB objective is to replace $\hat{\mathbb{P}}_0, \hat{\mathbb{P}}_1$ by Gaussians with matching moments. This results in a GSB problem which can be solved in closed form by our Theorem 3. Intuitively, if we denote an existing prior process by Y_t and the solution of its corresponding GSB by X_t, then X_t presents a more appealing prior process than Y_t since it carries the moment information of $\hat{\mathbb{P}}_0$ and $\hat{\mathbb{P}}_1$, whereas Y_t is completely data-oblivious.

Motivated by these observations, we propose a simple modification of the framework in Chen et al. (2022): Replace the prior process Y_t by its GSB approximation and keep everything else the same. The resulting scheme, which we term the GSBFLOW, learns a pair of forward $Z_t^\theta(x)$ and backward parametrized drifts $\hat{Z}_t^\phi(x)$ that progressively transport samples from $\hat{\mathbb{P}}_0 \to \hat{\mathbb{P}}_1$ and $\hat{\mathbb{P}}_1 \to \hat{\mathbb{P}}_0$, respectively. The full algorithm is presented in Appendix D.2 for completeness.

6.2.2 Results

We investigate the ability of GSBFLOW to generate cell populations $\hat{\mathbb{P}}_1$ from noise $\mathcal{N}_0$ ($\mathcal{N}_0 \to \hat{\mathbb{P}}_1$, Fig. 4a, b) on the the canonical datasets (Moon et al., 2019; Schiebinger et al., 2019); as well as to predict the dynamics of single-cell genomics ($\hat{\mathbb{P}}_0 \to \hat{\mathbb{P}}_1$, Fig. 4c) (Moon et al., 2019), i.e., the inference of cell populations $\hat{\mathbb{P}}_1$ resulting from the developmental process of an initial cell population $\hat{\mathbb{P}}_0$, with the goal of learning individual dynamics, identify ancestor and descendant cells. Details on datasets and experimental design can be found in Appendices D.4–D.5. The evaluation is conducted on the first 20 or 30 components of the PCA space of the > 1500 highly differentiable genes (see Figs. 7–8).

We evaluate the quality of the generated cellular states through the entropy-regularized Wasserstein distance W_ε (see Table 2) and by visualizing the first two principal components (PC), see Fig. 4a, b. GSBFLOW performs competitively on reconstructing embryoid body differentiation landscapes (Moon et al., 2019), and outperforms

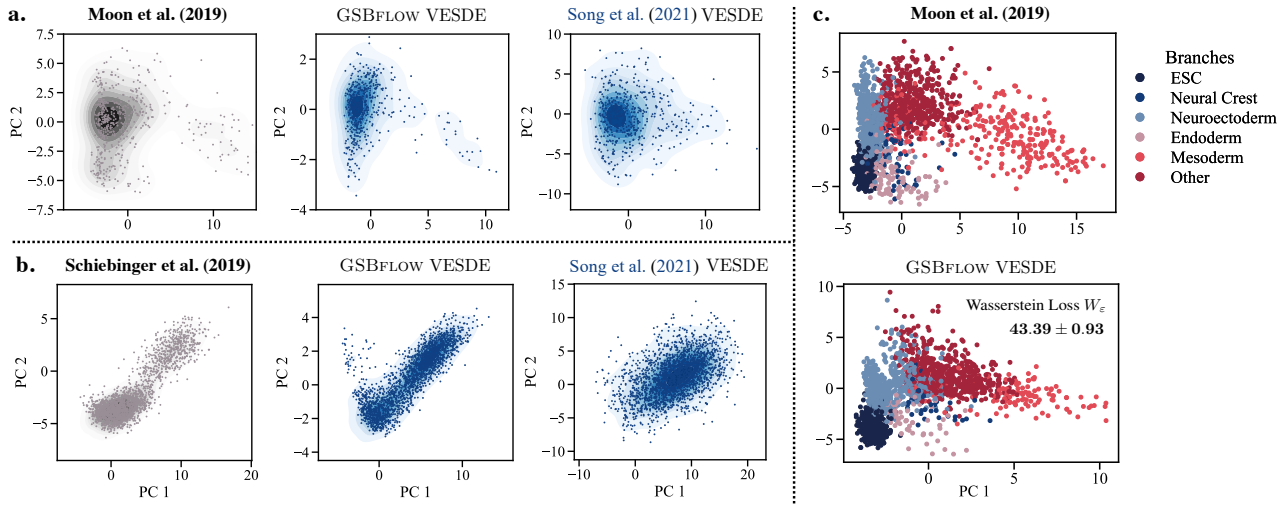


Figure 4: **a.-b.** Visual evaluation of the ability of our method to model the **generation** of data from **a.** Moon et al. (2019) and **b.** Schiebinger et al. (2019). Density plots are visualized in 2D PCA space and show generated data points using either GSBFLOW (our method) or the procedure in Song et al. (2021). **c.** Evaluation of GSBFLOW's ability to model the entire **evolution** of a developmental process of Moon et al. (2019), visualized by the data and GSBFLOW predictions colored by the lineage branch class.

score-based generative models baselines on the iPSC reprogramming task (Schiebinger et al., 2019) as quantified by W_ε between data and predictions. Further, we analyze GSBFLOW's ability to predict the temporal evolution of embryoid body differentiation (Moon et al., 2019), where cells measured at day 1 to 3 serve as samples of $\hat{\mathbb{P}}_0$, while $\hat{\mathbb{P}}_1$ is constructed from samples between day 12 to 27. As no ground truth trajectories are available in the data, we compare the predicted evolution to the data and compare how well the heterogeneity of lineage (Fig. 4c, upper panel) or sublineage branches (Fig. 6a) is captured. Fig. 4c (lower panel) and Fig. 6b thereby closely resemble the data (see W_ε in Fig. 4c) and thus demonstrate GSBFLOW's ability to learn cell differentiation into various lineages and to capture biological heterogeneity on a more macroscopic level.

7 Conclusion and Future Work

We derive closed-form solutions of GSBs, an important class of dynamic OT problems. Our technique originates from a deep connection between Gaussian OT and the Bures-Wasserstein geometry, which we generalize to the case of general SB problems. Numerically, we demonstrate that our new closed forms inspire a simple modification of existing SB-based numerical schemes, which can however lead to significantly improved performance.

Limitation of our framework. In a broader context, we hope our results can serve as the inspiration for more learning algorithms, much like how existing closed-form solutions of Gaussian OT problems have contributed to the machine learning community. We thus acknowledge a severe limitation of our closed-form solutions: These formulas require matrix inversions, which might face scalability issues for high-dimensional data. In addition, existing matrix inversion algorithms are typically extremely sensitive to the condition number, and thus our formulas are not as useful for ill-conditioned data. Lifting these constraints to facilitate further applications, such as to image datasets, is an important future work.

Acknowledgments

This research was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement no. 815943 and created as part of NCCR Catalysis (grant number 180544), a National Centre of Competence in Research funded by the Swiss National Science Foundation. Ya-Ping Hsieh acknowledges funding through an ETH Foundations of Data Science (ETH-FDS) postdoctoral fellowship.

References

J. Altschuler, S. Chewi, P. R. Gerber, and A. Stromme. Averaging on the Bures-Wasserstein manifold: dimension-free convergence of gradient descent. *Advances in Neural Information Processing Systems (NeurIPS)*, 34, 2021.

L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Springer, 2006.

E. Bernton, J. Heng, A. Doucet, and P. E. Jacob. Schrödinger Bridge Samplers. In *arXiv preprint arXiv:1912.13170*, 2019.

R. Bhatia, T. Jain, and Y. Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2), 2019.

R. Bojilov and A. Galichon. Matching in Closed-Form: Equilibrium, Identification, and Comparative Statics. *Economic Theory*, 61(4), 2016.

N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. Sliced and Radon Wasserstein Barycenters of Measures. *Journal of Mathematical Imaging and Vision*, 51(1), 2015.

C. Bunne, S. G. Stark, G. Gut, J. S. del Castillo, K.-V. Lehmann, L. Pelkmans, A. Krause, and G. Ratsch. Learning Single-Cell Perturbation Responses using Neural Optimal Transport. *bioRxiv*, 2021.

C. Bunne, A. Krause, and M. Cuturi. Supervised Training of Conditional Monge Maps. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

G. Carlier, L. Chizat, and M. Laborde. Lipschitz Continuity of the Schrödinger Map in Entropic Optimal Transport. *arXiv preprint arXiv:2210.00225*, 2022.

T. Chen, G.-H. Liu, and E. A. Theodorou. Likelihood Training of Schrödinger Bridge using Forward-Backward SDEs Theory. In *International Conference on Learning Representations (ICLR)*, 2022.

Y. Chen, T. T. Georgiou, and M. Pavon. Optimal Steering of a Linear Stochastic System to a Final Probability Distribution–Part III. *IEEE Transactions on Automatic Control*, 61(5), 2015.

Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2), 2016.

Y. Chen, Y. Shi, and B. Zhang. Optimal Control Via Neural Networks: A Convex Approach. In *International Conference on Learning Representations (ICLR)*, 2019.

Y. Chen, T. T. Georgiou, and M. Pavon. Stochastic Control Liaisons: Richard Sinkhorn Meets Gaspard Monge on a Schrödinger Bridge. *SIAM Review*, 63(2), 2021.

S. Chewi, T. Maunu, P. Rigollet, and A. J. Stromme. Gradient descent algorithms for Bures-Wasserstein barycenters. In *Conference on Learning Theory (COLT)*. PMLR, 2020.

M. Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 26, 2013.

V. De Bortoli, A. Doucet, J. Heng, and J. Thornton. Simulating Diffusion Bridges with Score Matching. In *arXiv preprint arXiv:2111.07243*, 2021a.

V. De Bortoli, J. Thornton, J. Heng, and A. Doucet. Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 2021b.

E. del Barrio and J.-M. Loubes. The statistical effect of entropic regularization in optimal transportation. *arXiv preprint arXiv:2006.05199*, 2020.

D. Dowson and B. Landau. The Fréchet Distance between Multivariate Normal Distributions. *Journal of Multivariate Analysis*, 12(3), 1982.

C. J. Frangieh, J. C. Melms, P. I. Thakore, K. R. Geiger-Schuller, P. Ho, A. M. Luoma, B. Cleary, L. Jerby-Arnon, S. Malu, M. S. Cuoco, et al. Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. *Nature Genetics*, 53(3), 2021.

I. Gentil, C. Léonard, and L. Ripani. About the analogy between optimal transport and minimal entropy. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 26, 2017.

I. Gentil, C. Léonard, and L. Ripani. Dynamical aspects of the generalized Schrödinger problem via Otto calculus–A heuristic point of view. *Revista Matemática Iberoamericana*, 36(4), 2020.

A. Han, B. Mishra, P. K. Jawanpuria, and J. Gao. On Riemannian optimization over positive definite matrices with the Bures-Wasserstein geometry. *Advances in Neural Information Processing Systems (NeurIPS)*, 34, 2021.

J. Ho, A. Jain, and P. Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

L. Holdijk, Y. Du, F. Hooft, P. Jaini, B. Ensing, and M. Welling. Path Integral Stochastic Optimal Control for Sampling Transition Paths. *arXiv preprint arXiv:2207.02149*, 2022.

C.-W. Huang, J. H. Lim, and A. Courville. A Variational Perspective on Diffusion-Based Generative Models and Score Matching. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021a.

J. Huang, Y. Jiao, L. Kang, X. Liao, J. Liu, and Y. Liu. Schrödinger-Föllmer Sampler: Sampling without Ergodicity. *arXiv preprint arXiv:2106.10880*, 2021b.

H. Janati, B. Muzellec, G. Peyré, and M. Cuturi. Entropic Optimal Transport between Unbalanced Gaussian Measures has a Closed Form. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, 2020.

D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations (ICLR)*, 2014.

A. Kulkarni, A. G. Anderson, D. P. Merullo, and G. Konopka. Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Current Opinion in Biotechnology*, 58:129–136, 2019.

H. Lavenant, S. Zhang, Y.-H. Kim, and G. Schiebinger. Towards a mathematical theory of trajectory inference. *arXiv preprint arXiv:2102.09204*, 2021.

C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.

G.-H. Liu, T. Chen, O. So, and E. A. Theodorou. Deep Generalized Schrödinger Bridge. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

M. D. Luecken and F. J. Theis. Current best practices in single-cell RNA-seqanalysis: a tutorial. *Molecular Systems Biology*, 15(6), 2019.

E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.

A. Mallasto, A. Gerolin, and H. Q. Minh. Entropy-regularized 2-Wasserstein distance between Gaussian measures. *Information Geometry*, pages 1–35, 2021.

R. Mansuy and M. Yor. *Aspects of Brownian motion*. Springer Science & Business Media, 2008.

G. R. Martin and M. J. Evans. Differentiation of Clonal Lines of Teratocarcinoma Cells: Formation of Embryoid Bodies In Vitro. *Proceedings of the National Academy of Sciences*, 72(4), 1975.

R. J. McCann. A convexity principle for interacting gases. *Advances in Mathematics*, 128(1), 1997.

K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. van den Elzen, M. J. Hirn, R. R. Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37(12), 2019.

I. Olkin and F. Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48, 1982.

F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Taylor & Francis*, 2001.

G. Peyré and M. Cuturi. Computational Optimal Transport. *Foundations and Trends in Machine Learning*, 11(5-6), 2019.

E. Platen and N. Bruti-Liberati. *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, volume 64. Springer Science & Business Media, 2010.

P. E. Protter. Stochastic Differential Equations. In *Stochastic Integration and Differential Equations*, pages 249–361. Springer, 2005.

J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein Barycenter and Its Application to Texture Mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2011.

G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, et al. Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell*, 176(4), 2019.

M. J. Shamblott, C. L. Kerr, J. Axelman, J. W. Littlefield, G. O. Clark, E. S. Patterson, R. C. Addis, J. N. Kraszewski, K. C. Kent, and J. D. Gearhart. Derivation and Differentiation of Human Embryonic Germ Cells. In *Essentials of Stem Cell Biology*. Elsevier, 2009.

J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In *International Conference on Machine Learning (ICML)*, 2015.

Y. Song and S. Ermon. Generative Modeling by Estimating Gradients of the Data Distribution. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations (ICLR)*, volume 9, 2021.

A. Takatsu. On Wasserstein geometry of Gaussian measures. In *Probabilistic Approach to Geometry*. Mathematical Society of Japan, 2010.

A. Tong, J. Huang, G. Wolf, D. Van Dijk, and S. Krishnaswamy. TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics. In *International Conference on Machine Learning (ICML)*, 2020.

user26872. Reference for Multidimensional Gaussian Integral. Mathematics Stack Exchange, 2012. URL https://math.stackexchange.com/q/126767.

F. Vargas, P. Thodoroff, N. D. Lawrence, and A. Lamacraft. Solving Schrödinger Bridges via Maximum Likelihood. *Entropy*, 23(9), 2021.

T. Vayer, R. Flamary, R. Tavenard, L. Chapel, and N. Courty. Sliced Gromov-Wasserstein. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

G. Wang, Y. Jiao, Q. Xu, Y. Wang, and C. Yang. Deep Generative Learning via Schrödinger Bridge. In *International Conference on Machine Learning (ICML)*, 2021.

F. A. Wolf, P. Angerer, and F. J. Theis. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1), 2018.

A. Zee. *Quantum Field Theory in a Nutshell*, volume 7. Princeton University Press, 2010.

G. X. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, et al. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1), 2017.

A Proof of Theorem 1

It is known that, for SBs, the optimal solution can be searched within the class of stochastic processes (Léonard, 2013)

$$X_t \sim \mathbb{P}_t : \quad \mathrm{d}X_t = (f_t(X_t) + w_t(X_t))\,\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t. \tag{A.1}$$

The Fokker-Planck equation for the SDE (A.1) is

$$\partial_t \rho_t = -\nabla_x \cdot (\rho_t(f_t + w_t)) + \frac{g_t^2}{2}\Delta\rho_t. \tag{A.2}$$

A simple application of the Girsanov's theorem then shows, up to a constant,

$$D_{\mathrm{KL}}(\mathbb{P}_t \| Y_t) = \mathbb{E}\left[\int_0^1 \frac{\|w_t\|^2}{2g_t^2}\,\mathrm{d}t\right]. \tag{A.3}$$

Using a change of variable $v_t = w_t - \frac{g_t^2}{2}\nabla \log \rho_t$, we see that (A.2) is equivalent to

$$\partial_t \rho_t = -\nabla_x \cdot (\rho_t(f_t + v_t)). \tag{A.4}$$

On the other hand, since $\|w_t\|^2 = \|v_t\|^2 + \frac{g_t^4}{4}\|\nabla \log \rho_t\|^2 + 2\left\langle v_t, \frac{g_t^2}{2}\nabla \log \rho_t \right\rangle$, the integrand in the objective of (A.3) becomes

$$\mathbb{E}\left[\int_0^1 \frac{\|v_t\|^2}{2g_t^2} + \frac{g_t^2}{8}\|\nabla \log \rho_t\|^2 + \frac{1}{2}\langle v_t, \nabla \log \rho_t\rangle\,\mathrm{d}t\right]. \tag{A.5}$$

Letting $H(\rho_t) := \int \rho_t \log \rho_t$ be the entropy, we have

$$\begin{aligned}
H(\rho_1) - H(\rho_0) &= \int_0^1 \partial_t H(\rho_t)\,\mathrm{d}t \\
&= \int_0^1 \int (1 + \log \rho_t)\partial_t \rho_t \mathrm{d}x\,\mathrm{d}t \\
&= \int_0^1 \int (1 + \log \rho_t) \cdot (-\nabla_x \cdot (\rho_t(f_t + v_t)))\mathrm{d}x\,\mathrm{d}t \qquad \text{by (A.2)} \\
&= \int_0^1 \int \rho_t \langle \nabla \log \rho_t, f_t + v_t\rangle \mathrm{d}x\,\mathrm{d}t
\end{aligned}$$

by integration by parts for the divergence operator. Therefore,

$$\mathbb{E}\left[\int_0^1 \langle \nabla \log \rho_t, v_t\rangle\,\mathrm{d}t\right] = H(\rho_1) - H(\rho_0) - \mathbb{E}\left[\int_0^1 \langle \nabla \log \rho_t, f_t\rangle\,\mathrm{d}t\right] \tag{A.6}$$

which concludes the proof. □

B The Bures-Wasserstein Geometry of Gaussian Schrödinger Bridges

B.1 Review of Bures-Wasserstein Geometry

Recall that the *metric tensor* $\langle \cdot, \cdot \rangle_\Sigma$ in the *Bures-Wasserstein geometry* (Takatsu, 2010) is defined in terms of the Lyapunov operator:

$$\forall\, U, V \in \mathcal{T}_\Sigma \mathbb{S}^d_{++}, \quad \langle U, V\rangle_\Sigma := \operatorname{tr} \mathcal{L}_\Sigma[U]\Sigma\mathcal{L}_\Sigma[V] = \frac{1}{2}\operatorname{tr} \mathcal{L}_\Sigma[U]V. \tag{B.1}$$

The corresponding Bures-Wasserstein norm is induced via $\|U\|_\Sigma^2 := \langle U, U\rangle_\Sigma$. Another important operator is the Bures-Wasserstein *gradient*: For any function $F\colon \mathbb{S}^d_{++} \to \mathbb{R}$,

$$\mathcal{T}_\Sigma \mathbb{S}^d_{++} \ni \operatorname{grad} F(\Sigma) := 2\big(\nabla F(\Sigma)\Sigma + \Sigma \nabla F(\Sigma)^\top\big) \tag{B.2}$$

where ∇ is the usual Euclidean gradient of F, viewed as a function from $\mathbb{R}^{d\times d}$ to $\mathbb{R}$. Note that

$$\mathcal{L}_\Sigma[\operatorname{grad} F(\Sigma)] = 2\mathcal{L}_\Sigma[\nabla F(\Sigma)\Sigma + \Sigma\nabla F(\Sigma)] \tag{B.3}$$
$$= 2\nabla F(\Sigma) \tag{B.4}$$

by definition of the Lyapunov operator. In other words,

$$\operatorname{grad} F(\Sigma) = \mathcal{L}_\Sigma^{-1}[2\nabla F]. \tag{B.5}$$

Lastly, we recall the Bures-Wasserstein *acceleration* of a curve $\Sigma_t : [0,1] \to \mathbb{S}^d_{++}$, which we denote by $\nabla_{\dot\Sigma_t}\dot\Sigma_t$:[2]

$$\nabla_{\dot\Sigma_t}\dot\Sigma_t = \ddot\Sigma_t - \Big(\mathcal{L}_{\Sigma_t}[\dot\Sigma_t]\dot\Sigma_t + \dot\Sigma_t\mathcal{L}_{\Sigma_t}[\dot\Sigma_t]\Big) + \Big(\Sigma_t\Big(\mathcal{L}_{\Sigma_t}[\dot\Sigma_t]\Big)^2 + \Big(\mathcal{L}_{\Sigma_t}[\dot\Sigma_t]\Big)^2\Sigma_t\Big). \tag{B.6}$$

B.2 Proof of Theorem 2

For convenience, we restate Theorem 2 in full below:

Theorem 2. *The minimizer of* (16) *(and hence* (4)*) coincides with the solution of the action minimization problem:*

$$\min_{\Sigma_0=\Sigma,\Sigma_1=\Sigma'} \int_0^1 \frac{1}{2}\|\dot\Sigma_t\|^2_{\Sigma_t} - \mathcal{U}_\sigma(\Sigma_t)\,\mathrm{d}t \tag{17}$$

where $\mathcal{U}_\sigma(\Sigma_t) := -\frac{\sigma^4}{8}\operatorname{tr}\Sigma_t^{-1}$ *and the minimum is taken over all piecewise smooth curves in* $\mathbb{S}^d_{++}$*. In particular, the minimizer of* (16) *solves the Euler-Lagrange equation in the Bures-Wasserstein geometry:*

$$\nabla_{\dot\Sigma_t}\dot\Sigma_t = -\operatorname{grad}\mathcal{U}_\sigma(\Sigma_t), \quad \Sigma_0 = \Sigma, \quad \Sigma_1 = \Sigma', \tag{18}$$

where $\nabla_{\dot\Sigma_t}\dot\Sigma_t$ *denotes the Riemannian acceleration and* grad *the Riemannian gradient in the Bures-Wasserstein sense.*

The proof consists of verifying the Euler-Lagrange equation (18) for the curve (19).

B.2.1 Verifying the Euler-Lagrange Equation (18)

We begin by noting that the boundary conditions in (18) hold for the curve in (19).

We now compute the two sides of (18) separately:

The right-hand side of (18): $-\operatorname{grad}\mathcal{U}_\sigma(\Sigma_t)$. Since $\nabla\mathcal{U}_\sigma(\Sigma_t) = -\nabla\Big(\operatorname{tr}\frac{\sigma^4}{8}\Sigma_t^{-1}\Big) = \frac{\sigma^4}{8}\Sigma_t^{-1}\cdot\Sigma_t^{-1}$, we see from (B.2) that the negative Bures-Wasserstein gradient of $\mathcal{U}_\sigma(\Sigma_t)$ is

$$-\operatorname{grad}\mathcal{U}_\sigma(\Sigma_t) = -2\Big(\frac{\sigma^4}{8}\Sigma_t^{-1}\cdot\Sigma_t^{-1}\cdot\Sigma_t + \Sigma_t\cdot\frac{\sigma^4}{8}\Sigma_t^{-1}\cdot\Sigma_t^{-1}\Big)$$
$$= -\frac{\sigma^4}{2}\Sigma_t^{-1}. \tag{B.7}$$

The left-hand side of (18): $\nabla_{\dot\Sigma_t}\dot\Sigma_t$. Computing $\nabla_{\dot\Sigma_t}\dot\Sigma_t$ is significantly trickier than $-\operatorname{grad}\mathcal{U}_\sigma(\Sigma_t)$. The central piece of the proof is the following technical lemma:

Lemma B.1. *Define the matrix* $\tilde S_t$ *to be:*

$$\tilde S_t := t\Sigma' + \bar t C_\sigma - \bar t\Sigma - tC_\sigma^\top + \frac{\sigma^2}{2}(\bar t - t)I. \tag{B.8}$$

Then $\mathcal{L}_{\Sigma_t}[\dot\Sigma_t] = \tilde S_t^\top\Sigma_t^{-1}$*. In other words,* $\tilde S_t^\top\Sigma_t^{-1}$ *is symmetric and solves the Lyapunov equation:*

$$A: \quad A\Sigma_t + \Sigma_t A = \dot\Sigma_t. \tag{B.9}$$

Moreover, $\tilde S_t$ *satisfies the following identity:*

$$\dot{\tilde S}_t - \Sigma_t^{-1}\tilde S_t^2 = -\frac{\sigma^4}{4}\Sigma_t^{-1}. \tag{B.10}$$

[2] More formally, $\nabla_{\dot\Sigma_t}\dot\Sigma_t$ is the Bures-Wasserstein covariant derivative of $\dot\Sigma_t$ in the direction of $\dot\Sigma_t$.

Before commencing the proof of Lemma B.1, let us show how it readily leads us to (18).

Recall the definition of $\nabla_{\dot{\Sigma}_t}\dot{\Sigma}_t$ in (B.6). First, note that, by (19) and (B.8),

$$\frac{1}{2}\ddot{\Sigma}_t = \Sigma + \Sigma' - \left(C_\sigma + C_\sigma^\top + \sigma^2 I\right) \tag{B.11}$$

$$= \dot{\tilde{S}}_t. \tag{B.12}$$

On the other hand, Lemma B.1 entails that

$$\begin{aligned}\Sigma_t\Big(\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\Big)^2 + \Big(\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\Big)^2\Sigma_t &= \Sigma_t\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\cdot\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t] + \mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\cdot\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\Sigma_t \\ &= \Sigma_t\Sigma_t^{-1}\tilde{S}_t\cdot\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t] + \mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\cdot\tilde{S}_t^\top\Sigma_t^{-1}\Sigma_t \\ &= \tilde{S}_t\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t] + \mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\tilde{S}_t^\top. \end{aligned} \tag{B.13}$$

By noting, again from Lemma B.1,

$$\begin{aligned}\dot{\Sigma}_t &= \tilde{S}_t^\top\Sigma_t^{-1}\cdot\Sigma_t + \Sigma_t\cdot\Sigma_t^{-1}\tilde{S}_t \\ &= \tilde{S}_t + \tilde{S}_t^\top, \end{aligned} \tag{B.14}$$

we thus get

$$\begin{aligned}\Sigma_t\Big(\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\Big)^2 + \Big(\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\Big)^2\Sigma_t - \Big(\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\dot{\Sigma}_t + \dot{\Sigma}_t\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\Big) &= \Big(\tilde{S}_t - \dot{\Sigma}_t\Big)\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t] + \mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\Big(\tilde{S}_t^\top - \dot{\Sigma}_t\Big) \\ &= -\Big(\tilde{S}_t^\top\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t] + \mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\tilde{S}_t\Big)\end{aligned}$$

by (B.14). But $\tilde{S}_t^\top\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t] = \tilde{S}_t^\top\cdot\Sigma_t^{-1}\tilde{S}_t = \Sigma_t^{-1}\tilde{S}_t^2$ by symmetry of $\tilde{S}_t^\top\Sigma_t^{-1}$ and, similarly, we have $\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\tilde{S}_t = \Sigma_t^{-1}\tilde{S}_t^2$. As a result, (B.6) reduces to

$$\nabla_{\dot{\Sigma}_t}\dot{\Sigma}_t = 2\dot{\tilde{S}}_t - 2\Sigma_t^{-1}\tilde{S}_t^2. \tag{B.15}$$

In lieu of (18), (B.7), and (B.15), the proof of (17) can thus be reduced to showing

$$2\dot{\tilde{S}}_t - 2\Sigma_t^{-1}\tilde{S}_t^2 = -\frac{\sigma^4}{2}\Sigma_t^{-1} \tag{B.16}$$

which is exactly (B.10).

Proof of Lemma B.1. We now prove Lemma B.1. We begin by proving some useful identities that will inspire our proof for the general GSBs in Section 5.

Useful identities. First, note that the definition of C_σ immediately implies $C_\sigma\Sigma = \Sigma C_\sigma^\top$. In addition, we have

$$\begin{aligned}C_\sigma^{-1}\Sigma &= 2\Big(\Sigma^{\frac{1}{2}}D_\sigma\Sigma^{-\frac{1}{2}} - \sigma^2 I\Big)^{-1}\Sigma \\ &= 2\Big(\Sigma^{-\frac{1}{2}}D_\sigma\Sigma^{-\frac{1}{2}} - \sigma^2\Sigma^{-1}\Big)^{-1} \\ &= \Sigma C_\sigma^{-\top}. \end{aligned} \tag{B.17}$$

Recall from (Janati et al., 2020) that C_σ solves the following matrix equation:

$$C_\sigma^2 + \sigma^2 C_\sigma = \Sigma\Sigma'. \tag{B.18}$$

We therefore have

$$\begin{aligned}C_\sigma &= C_\sigma^{-1}\Sigma\Sigma' - \sigma^2 I, \\ C_\sigma^\top &= \Sigma'\Sigma C_\sigma^{-\top} - \sigma^2 I,\end{aligned}$$

which, together with (B.17), implies

$$\begin{aligned} C_\sigma^\top \Sigma' &= \Sigma' \Sigma C_\sigma^{-\top} \Sigma' - \sigma^2 \Sigma' \\ &= \Sigma' C_\sigma^{-1} \Sigma \Sigma' - \sigma^2 \Sigma' \\ &= \Sigma' C_\sigma. \end{aligned} \tag{B.19}$$

Now, set $\tilde{S}_t = P_t - Q_t^\top + \frac{\sigma^2}{2}(\bar{t} - t)I$ where

$$P_t := t\Sigma' + \bar{t} C_\sigma, \quad Q_t := \bar{t}\Sigma + t C_\sigma. \tag{B.20}$$

Note that, by (B.19),

$$\begin{aligned} \Sigma' P_t^{-1} &= \left(P_t \Sigma'^{-1}\right)^{-1} \\ &= \left(tI + \bar{t} C_\sigma \Sigma'^{-1}\right)^{-1} \\ &= \left(tI + \bar{t} \Sigma'^{-1} C_\sigma^\top\right)^{-1} \\ &= \left(\Sigma'^{-1} P_t^\top\right)^{-1} \\ &= P_t^{-\top} \Sigma'. \end{aligned} \tag{B.21}$$

A similar calculation leading to (B.21) shows

$$Q_t^{-1} \Sigma = \Sigma Q_t^{-\top}. \tag{B.22}$$

Proof of symmetry of $\tilde{S}_t^\top \Sigma_t^{-1}$. We get, by (B.18) and (B.19),

$$\begin{aligned} P_t^2 + \sigma^2 \bar{t} P_t &= t^2 \Sigma'^2 + \bar{t}^2 C_\sigma^2 + t\bar{t}(\Sigma' C_\sigma + C_\sigma \Sigma') + \sigma^2 t \bar{t} \Sigma' + \sigma^2 \bar{t}^2 C_\sigma \\ &= t^2 \Sigma'^2 + \bar{t}^2 \left(C_\sigma^2 + \sigma^2 C_\sigma\right) + t\bar{t}\left(C_\sigma^\top \Sigma' + C_\sigma \Sigma'\right) + \sigma^2 t \bar{t} \Sigma' \\ &= t^2 \Sigma'^2 + \bar{t}^2 \Sigma \Sigma' + t\bar{t}\left(C_\sigma^\top + C_\sigma + \sigma^2 I\right)\Sigma' = \Sigma_t \Sigma'. \end{aligned} \tag{B.23}$$

It then follows from (B.23) that

$$P_t = \Sigma_t \Sigma' P_t^{-1} - \sigma^2 \bar{t} I, \tag{B.24}$$

$$P_t^\top = P_t^{-\top} \Sigma' \Sigma_t - \sigma^2 \bar{t} I. \tag{B.25}$$

As a result, we get, by (B.21) and (B.24)-(B.25),

$$\begin{aligned} \Sigma_t^{-1} P_t &= \Sigma' P_t^{-1} - \sigma^2 \bar{t} \Sigma_t^{-1} \\ &= P_t^{-\top} \Sigma' - \sigma^2 \bar{t} \Sigma_t^{-1} \\ &= P_t^\top \Sigma_t^{-1}. \end{aligned} \tag{B.26}$$

In exactly the same vein, we have

$$Q_t^2 + \sigma^2 t Q_t = \Sigma \Sigma_t \tag{B.27}$$

as well as

$$\Sigma_t^{-1} Q_t^\top = Q_t \Sigma_t^{-1}. \tag{B.28}$$

The symmetry of $\tilde{S}_t^\top \Sigma_t^{-1}$ is then an immediate consequence of (B.26) and (B.28). In addition, we have

$$\begin{aligned} \dot{\Sigma}_t &= 2t\Sigma' - 2\bar{t}\Sigma + (\bar{t} - t)\left(C_\sigma + C_\sigma^\top + \sigma^2 I\right) \\ &= \tilde{S}_t + \tilde{S}_t^\top. \end{aligned} \tag{B.29}$$

Combining the symmetry of $\tilde{S}_t^\top \Sigma_t^{-1}$ and (B.29), we see that

$$\tilde{S}_t^\top \Sigma_t^{-1} \cdot \Sigma_t + \Sigma_t \cdot \Sigma_t^{-1} \tilde{S}_t = \tilde{S}_t + \tilde{S}_t^\top = \dot{\Sigma}_t,$$

i.e., $\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t] = \tilde{S}_t^\top \Sigma_t^{-1}$.

Proof of (B.10). We next compute

$$\begin{aligned} P_t Q_t^\top &= (t\Sigma' + \bar{t}C_\sigma)\big(\bar{t}\Sigma + tC_\sigma^\top\big) \\ &= t\bar{t}\Sigma'\Sigma + t^2\Sigma' C_\sigma^\top + \bar{t}^2 C_\sigma \Sigma + t\bar{t} C_\sigma C_\sigma^\top \\ &= \bar{t}^2 \Sigma C_\sigma^\top + t^2 \Sigma' C_\sigma^\top + t\bar{t}\big(C_\sigma^{\top 2} + \sigma^2 C_\sigma^\top\big) + t\bar{t} C_\sigma C_\sigma^\top \\ &= \Sigma_t C_\sigma^\top \end{aligned} \tag{B.30}$$

where we have used (B.17) in the third equality of (B.30). A similar computation further shows

$$Q_t^\top P_t = \Sigma_t C_\sigma. \tag{B.31}$$

We thus get, by combining (B.23) (B.27)

$$\begin{aligned} \tilde{S}_t^2 &= P_t^2 - P_t Q_t^\top + \frac{\sigma^2}{2}(\bar{t}-t)P_t - Q_t^\top P_t + Q_t^{\top 2} - \frac{\sigma^2}{2}(\bar{t}-t)Q_t^\top + \frac{\sigma^2}{2}(\bar{t}-t)P_t - \frac{\sigma^2}{2}(\bar{t}-t)Q_t^\top + \frac{\sigma^4}{4}(\bar{t}-t)^2 I \\ &= P_t^2 + \sigma^2(\bar{t}-t)P_t + Q_t^{\top 2} - \sigma^2(\bar{t}-t)Q_t^\top - \big(P_t Q_t^\top + Q_t^\top P_t\big) + \frac{\sigma^4}{4}(\bar{t}-t)^2 I \\ &= \Sigma_t\Sigma' - \sigma^2 t P_t + \Sigma_t\Sigma - \sigma^2\bar{t}Q_t^\top - \big(\Sigma_t C_\sigma^\top + \Sigma_t C_\sigma\big) + \frac{\sigma^4}{4}(\bar{t}-t)^2 I - \sigma^2\Sigma_t + \sigma^2\Sigma_t \\ &= \Sigma_t\big(\Sigma + \Sigma' - \big(C_\sigma + C_\sigma^\top + \sigma^2 I\big)\big) + \sigma^2\big(\Sigma_t - tP_t - \bar{t}Q_t^\top\big) + \frac{\sigma^4}{4}(\bar{t}-t)^2 I \\ &= \Sigma_t\dot{\tilde{S}}_t + \sigma^2 \cdot t\bar{t}\sigma^2 I + \frac{\sigma^4}{4}(\bar{t}-t)^2 I \\ &= \Sigma_t\dot{\tilde{S}}_t + \frac{\sigma^4}{4} I \end{aligned} \tag{B.32}$$

where the third equality follows from (B.23), (B.27), and (B.30)-(B.31), and the fifth equality follows from (B.8). Multiplying both sides of (B.32) by Σ_t^{-1} from the right yields the desired (B.10). □

B.2.2 Equivalence between (16) and (17)

We first note that, by (B.1) and Lemma B.1,

$$\begin{aligned} \frac{1}{2}\|\dot{\Sigma}_t\|_{\Sigma_t}^2 &= \frac{1}{2}\operatorname{tr}\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t]\Sigma_t\mathcal{L}_{\Sigma_t}[\dot{\Sigma}_t] \\ &= \frac{1}{2}\operatorname{tr}\tilde{S}_t^\top\Sigma_t^{-1}\cdot\Sigma_t\cdot\Sigma_t^{-1}\tilde{S}_t \\ &= \frac{1}{2}\operatorname{tr}\tilde{S}_t^\top\Sigma_t^{-1}\tilde{S}_t, \end{aligned} \tag{B.33}$$

and therefore the integrand in (17) is equal to

$$\operatorname{tr}\left(\frac{1}{2}\tilde{S}_t^\top\Sigma_t^{-1}\tilde{S}_t + \frac{\sigma^4}{8}\Sigma_t^{-1}\right). \tag{B.34}$$

To proveed, we will need another formulation of (16), which is (Chen et al., 2016; Gentil et al., 2017) specialized to our case:

Lemma B.2. *Let $\mathcal{N}_0 := \mathcal{N}(0,\Sigma)$ and $\mathcal{N}_1 := \mathcal{N}(0,\Sigma')$. Then (16) is equivalent to*

$$\min_{\rho_0=\mathcal{N}_0,\rho_1=\mathcal{N}_1}\int_0^1 \mathbb{E}\left[\frac{1}{2}\|\nabla\Phi_t\|^2 + \frac{\sigma^4}{8}\|\nabla\log\rho_t\|^2\right]\mathrm{d}t \tag{B.35}$$

where the minimization is taken over all pairs $(\rho_t, \nabla\Phi_t)$ such that $\Phi_t : \mathbb{R}^d \to \mathbb{R}$ are differentiable functions and the continuity equation holds:

$$\partial_t\rho_t = -\nabla_x\cdot(\rho_t\nabla\Phi_t). \tag{B.36}$$

We will also need the *Jacobi formula*: Let $A(t)\colon \mathbb{R}^+ \to \mathbb{R}^{d\times d}$ be a differentiable matrix-valued function. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\det A(t) = \det A(t)\cdot \operatorname{tr} A^{-1}(t)\cdot \frac{\mathrm{d}}{\mathrm{d}t}A(t). \tag{B.37}$$

We are now ready to finish the proof of Theorem 2. By Léonard (2013), the optimal curve for (B.35) is Gaussian with zero mean. We denote by Σ_t the covariance of the solution at time t. By (B.37), we have

$$\begin{aligned}
\partial_t \rho_t(x) &= \partial_t\left((2\pi)^{\frac{d}{2}}(\det\Sigma_t)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}x^\top\Sigma_t^{-1}x\right)\right)\\
&= (2\pi)^{\frac{d}{2}}\left(-\frac{1}{2}(\det\Sigma_t)^{-\frac{3}{2}}\right)\cdot\det\Sigma_t\cdot\operatorname{tr}\Sigma_t^{-1}\dot\Sigma_t\exp\left(-\frac{1}{2}x^\top\Sigma_t^{-1}x\right)\\
&\qquad + (2\pi)^{\frac{d}{2}}(\det\Sigma_t)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}x^\top\Sigma_t^{-1}x\right)\cdot\left(\frac{1}{2}x^\top\Sigma_t^{-1}\dot\Sigma_t\Sigma_t^{-1}x\right)\\
&= \rho_t(x)\cdot\left(\frac{1}{2}x^\top\Sigma_t^{-1}\dot\Sigma_t\Sigma_t^{-1}x - \frac{1}{2}\operatorname{tr}\Sigma_t^{-1}\dot\Sigma_t\right).
\end{aligned} \tag{B.38}$$

On the other hand, by the chain rule for the divergence, we have

$$\nabla_x\cdot(\rho_t\nabla\Phi_t) = \langle\nabla\rho_t,\nabla\Phi_t\rangle + \rho_t\Delta\Phi_t. \tag{B.39}$$

Since $\nabla\rho_t = \rho_t\big(-\Sigma_t^{-1}x\big)$, the continuity equation (B.36) together with (B.38)-(B.39) implies that Σ_t must satisfy

$$\Delta\Phi_t = \frac{1}{2}\operatorname{tr}\Sigma_t^{-1}\dot\Sigma_t, \tag{B.40}$$

$$\langle\Sigma_t^{-1}x,\nabla\Phi_t(x)\rangle = \frac{1}{2}\langle\Sigma_t^{-1}x,\dot\Sigma_t\Sigma_t^{-1}x\rangle,\quad \forall x\in\mathbb{R}^d. \tag{B.41}$$

In other words, the optimal vector field is of the form $\nabla\Phi_t(x) = \tilde S_t^\top\Sigma_t^{-1}x$ for some matrix $\tilde S_t$ such that

$$\operatorname{tr}\tilde S_t^\top\Sigma_t^{-1} = \frac{1}{2}\operatorname{tr}\dot\Sigma_t\Sigma_t^{-1}, \tag{B.42}$$

$$\operatorname{tr}\Sigma_t^{-1}\tilde S_t^\top\Sigma_t^{-1}xx^\top = \frac{1}{2}\operatorname{tr}\Sigma_t^{-1}\dot\Sigma_t\Sigma_t^{-1}xx^\top,\quad \forall x\in\mathbb{R}^d. \tag{B.43}$$

Therefore, we see that

$$\begin{aligned}
\mathbb{E}\big[\|\nabla\Phi_t\|^2\big] &= \mathbb{E}\Big[\operatorname{tr}\tilde S_t^\top\Sigma_t^{-1}xx^\top\Sigma_t^{-1}\tilde S_t\Big]\\
&= \operatorname{tr}\tilde S_t^\top\Sigma_t^{-1}\,\mathbb{E}[xx^\top]\Sigma_t^{-1}\tilde S_t\\
&= \operatorname{tr}\tilde S_t^\top\Sigma_t^{-1}\tilde S_t.
\end{aligned} \tag{B.44}$$

Furthermore, we have

$$\begin{aligned}
\mathbb{E}\big[\|\nabla\log\rho_t\|^2\big] &= \mathbb{E}\big[\operatorname{tr}\Sigma_t^{-1}xx^\top\Sigma_t^{-1}\big]\\
&= \operatorname{tr}\Sigma_t^{-1}.
\end{aligned} \tag{B.45}$$

Finally, since the optimal vector field $\nabla\Phi_t$ is a gradient field, we must have $\tilde S_t^\top\Sigma_t^{-1} = \Sigma_t^{-1}\tilde S_t$. Combing all the above, we see that (B.35) is equivalent to

$$\min_{\substack{\Sigma_0=\Sigma,\Sigma_1=\Sigma'\\ \tilde S_t^\top\Sigma_t^{-1}=\Sigma_t^{-1}\tilde S_t}}\int_0^1\operatorname{tr}\left(\frac{1}{2}\tilde S_t^\top\Sigma_t^{-1}\tilde S_t + \frac{\sigma^4}{8}\Sigma_t^{-1}\right)\mathrm{d}t \tag{B.46}$$

which, in view of (B.34), is exactly the same as (17).

B.3 Some Interesting Consequences of Theorem 2

Here, we collect some interesting corollaries of Theorem 2, although they will not be used in the rest of the paper.

B.3.1 Conservation of Hamiltonian

The first result concerns the *Hamiltonian formulation* of the action minimization problem (17).

Corollary 2 (Conservation of Hamiltonian)**.** *Define the* ***Hamiltonian*** *associated with* (17) *to be*

$$\begin{aligned}\mathcal{H}(\Sigma_t) &:= \frac{1}{2}\|\dot{\Sigma}_t\|^2_{\Sigma_t} + \mathcal{U}_\sigma(\Sigma_t) \\ &= \operatorname{tr}\left(\frac{1}{2}\tilde{S}_t^\top \Sigma_t^{-1}\tilde{S}_t - \frac{\sigma^4}{8}\Sigma_t^{-1}\right).\end{aligned} \tag{H}$$

Then the Hamiltonian is conserved along Σ_t:

$$\dot{\mathcal{H}} \equiv 0, \text{ or, equivalently, } \mathcal{H}(\Sigma_t) = \operatorname{tr}(\Sigma + \Sigma' - D_\sigma) \text{ for all } t. \tag{B.47}$$

The fact that the Hamiltonian, commonly interpreted as the *total energy*, is conserved is a well-known fact in physics (Villani, 2009) and directly follows from Theorem 2.

B.3.2 Connection to Fisher Information

The "potential energy" term $\mathcal{U}_\sigma(\Sigma_t)$ in (17) has an interesting origin: It is, up to a constant, the *entropy production rate*, i.e., the *Fisher information*.

Lemma B.3. *Let $\rho \sim \mathcal{N}(0, \Sigma)$, and let $H(\Sigma)$ be the (negative) Shannon entropy of ρ. Then*

$$\mathcal{U}_\sigma(\Sigma_t) = \frac{1}{2}\mathcal{I}_\sigma(\Sigma) \tag{B.48}$$

where

$$\mathcal{I}_\sigma(\Sigma) := \frac{\sigma^4}{4}\|\operatorname{grad} H(\Sigma)\|^2_\Sigma. \tag{B.49}$$

Proof. Recall that $\nabla H(\Sigma) = \nabla\left(-\frac{1}{2}\log\det\Sigma - \frac{d}{2}\log 2\pi e\right) = -\frac{1}{2}\Sigma^{-1}$. Therefore, by (B.1) and (B.5),

$$\begin{aligned}\mathcal{I}_\sigma(\Sigma) &= \frac{\sigma^4}{4}\|\operatorname{grad} H(\Sigma)\|^2_\Sigma \\ &= \frac{\sigma^4}{4}\langle \operatorname{grad} H(\Sigma), \operatorname{grad} H(\Sigma)\rangle_\Sigma \\ &= \frac{\sigma^4}{4}\operatorname{tr}\mathcal{L}_\Sigma[\operatorname{grad} H(\Sigma)]\Sigma\mathcal{L}_\Sigma[\operatorname{grad} H(\Sigma)] \\ &= \frac{\sigma^4}{4}\operatorname{tr}\mathcal{L}_\Sigma[\mathcal{L}_\Sigma^{-1}[2\nabla H(\Sigma)]]\Sigma\mathcal{L}_\Sigma[\mathcal{L}_\Sigma^{-1}[2\nabla H(\Sigma)]] \\ &= \frac{\sigma^4}{4}(-\Sigma^{-1})\Sigma(-\Sigma^{-1}) = \frac{\sigma^4}{4}\Sigma^{-1}.\end{aligned}$$

□

An infinite-dimensional version of Lemma B.3 for non-Gaussian measures is proved in Chen et al. (2016); Gentil et al. (2017); the connection to the Bures-Wasserstein geometry here seems to be new.

The specific form of the potential energy in (B.49) has been shown to be intimately related to the *gradient flow* of entropy:

$$\dot{\Sigma}_t = -\operatorname{grad} H(\Sigma). \tag{B.50}$$

We refer the interested readers to (Gentil et al., 2020) for details.

B.3.3 Solution of the Schrödinger Systems

Another way of solving a system of the form (B.35) is via the so-called forward *Schrödinger system* (Chen et al., 2021; Léonard, 2013):

$$\begin{cases} \partial_t\mu_t + \nabla_x\cdot(\mu_t\nabla\Phi_t) = \frac{\sigma^2}{2}\Delta\mu_t \\ \partial_t\Phi_t + \frac{\|\nabla\Phi_t\|^2}{2} + \frac{\sigma^2}{2}\Delta\Phi_t = 0 \end{cases}. \tag{B.51}$$

By the various identities we prove in Appendix B.2.1, one can easily show that the solution to (B.51) is given by

$$\Phi_t(x) = -\frac{\sigma^2}{4}\log\det\Sigma_t + \frac{\sigma^4}{4}\int_0^t \operatorname{tr}\Sigma_t^{-1}\,\mathrm{d}t + \frac{1}{2}\langle x, \left(\tilde{S}_t^\top - \frac{\sigma^2}{2}I\right)\Sigma_t^{-1}x\rangle + \text{const.} \tag{B.52}$$

This is in fact the same solution of the *fluid mechanical* problem

$$\min_{\substack{\rho_0=\mathcal{N}_0,\rho_1=\mathcal{N}_1\\ \partial_t\rho_t+\nabla_x\cdot(\rho_t\nabla\Phi_t)=\Delta\rho_t}} \int_0^1 \mathbb{E}\left[\frac{1}{2}\|\nabla\Phi_t\|^2\right]\mathrm{d}t \tag{B.53}$$

which is yet another equivalent formulation of (16).

There is also a backward Schrödinger system:

$$\begin{cases} -\partial_t\mu_t + \nabla_x\cdot\left(\mu_t\nabla\hat{\Phi}_t\right) = \frac{\sigma^2}{2}\Delta\mu_t \\ -\partial_t\hat{\Phi}_t + \frac{\|\nabla\hat{\Phi}_t\|^2}{2} + \frac{\sigma^2}{2}\Delta\hat{\Phi}_t = 0 \end{cases}, \tag{B.54}$$

whose solution is given by

$$\hat{\Phi}_t(x) = -\frac{\sigma^2}{4}\log\det\Sigma_t - \frac{\sigma^4}{4}\int_0^t \operatorname{tr}\Sigma_t^{-1}\,\mathrm{d}t - \frac{1}{2}\langle x, \left(\tilde{S}_t^\top + \frac{\sigma^2}{2}I\right)\Sigma_t^{-1}x\rangle + \text{const.} \tag{B.55}$$

Notice that

$$\Phi_t + \hat{\Phi}_t = \sigma^2\log\rho_t \tag{B.56}$$

which is a well-known feature of the solutions to the forward and backward Schrödinger systems (Chen et al., 2021; Léonard, 2013).

C Proof of the Closed-Form Solutions for Gaussian Schrödinger Bridges

C.1 Preliminaries for the Proof of Theorem 3

We need a technical lemma that is intimately related to the "central identity of quantum field theory" (Zee, 2010); the version below is adopted from (user26872, 2012), wherein the readers can find an easy proof.

Lemma C.1 (The central identity of Quantum Field Theory)**.** *The following identity holds for all matrix $M \succ 0$ and all sufficiently regular analytic function v (e.g., polynomials or $v \in \mathcal{C}^\infty(\mathbb{R}^d)$ with compact support):*

$$(2\pi)^{-\frac{d}{2}}(\det M)^{\frac{1}{2}}\int_{\mathbb{R}^d} v(x)\exp\left(-\frac{1}{2}x^\top Mx\right)\mathrm{d}x = \exp\left(\frac{1}{2}\partial_x^\top M^{-1}\partial_x\right)v(x)\Big|_{x=0} \tag{C.1}$$

where $\exp\left(\frac{1}{2}\partial_x^\top M^{-1}\partial_x\right)$ is understood as a power series in the differential operators.

Lastly, we recall the elementary

Lemma C.2 (Conditional Gaussians are Gaussian)**.** *Let $(Y_0, Y_1) \sim \mathcal{N}\left(\begin{bmatrix}\mu_0\\ \mu_1\end{bmatrix}, \begin{bmatrix}\Sigma_{00} & \Sigma_{01}\\ \Sigma_{10} & \Sigma_{11}\end{bmatrix}\right)$. Then $Y_0|Y_1 = y \sim \mathcal{N}(\check{\mu}, \check{\Sigma})$ where*

$$\begin{aligned}\check{\mu} &= \mu_0 + \Sigma_{01}\Sigma_{11}^{-1}(y-\mu_1),\\ \check{\Sigma} &= \Sigma_{00} - \Sigma_{01}\Sigma_{11}^{-1}\Sigma_{10}.\end{aligned} \tag{C.2}$$

C.2 The Proof

We are now ready for the proof. For convenience, we restate Theorem 3 below:

Theorem 3. *Denote by $\mathbb{P}_t$ the solution to Gaussian Schrödinger bridges (GSB). Set*

$$
\begin{aligned}
&r_t := \frac{\kappa(t,1)}{\kappa(1,1)}, \quad \bar{r}_t := \tau_t - r_t\tau_1, \quad \sigma_\star := \sqrt{\tau_1^{-1}\kappa(1,1)},\\
&\zeta(t) := \tau_t \int_0^t \tau_s^{-1}\alpha_s \,\mathrm{d}s, \rho_t := \frac{\int_0^t \tau_s^{-2} g_s^2 \,\mathrm{d}s}{\int_0^1 \tau_s^{-2} g_s^2 \,\mathrm{d}s},\\
&P_t := \dot{r}_t(r_t\Sigma_1 + \bar{r}_t C_{\sigma_\star}), \quad Q_t := -\dot{\bar{r}}_t(\bar{r}_t\Sigma_0 + r_t C_{\sigma_\star}),\\
&S_t := P_t - Q_t^\top + \big[c_t\kappa(t,t)(1-\rho_t) - g_t^2\rho_t\big] I.
\end{aligned} \tag{25}
$$

Then the following holds:

1. *The solution $\mathbb{P}_t$ is a Markov Gaussian process whose marginal variable $X_t \sim \mathcal{N}(\mu_t, \Sigma_t)$, where*

$$\mu_t := \bar{r}_t\mu_0 + r_t\mu_1 + \zeta(t) - r_t\zeta(1), \tag{26}$$

$$\Sigma_t := \bar{r}_t^2\Sigma_0 + r_t^2\Sigma_1 + r_t\bar{r}_t\big(C_{\sigma_\star} + C_{\sigma_\star}^\top\big) + \kappa(t,t)(1-\rho_t)I. \tag{27}$$

2. *X_t admits a closed-form representation as the SDE:*

$$\mathrm{d}X_t = f_{\mathcal{N}}(t, X_t)\,\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t \tag{28}$$

where

$$f_{\mathcal{N}}(t,x) := S_t^\top \Sigma_t^{-1}(x - \mu_t) + \dot{\mu}_t. \tag{29}$$

Moreover, the matrix $S_t^\top\Sigma_t^{-1}$ is symmetric.

As the proof is quite complicated, we first outline the main steps below:

1. Leveraging existing results (Bojilov and Galichon, 2016; del Barrio and Loubes, 2020; Janati et al., 2020; Mallasto et al., 2021), we first solve an appropriately chosen *static* GSB determined by the reference process $\mathbb{Q}_t$.
2. It can be shown from the disintegration formula (Léonard, 2013), the solution of the static GSBs (3), and properties of (22) that $\mathbb{P}_t$ is a Markov Gaussian process with mean (26) and covariance (27).
3. Invoking the *generator theory* (Protter, 2005), to prove (28), it suffices to show that X_t satisfies, for any sufficiently regular test function $u : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$,

$$\lim_{h\to 0} \frac{\mathbb{E}\big[u\big(t+h, X_{t+h}\big) \,\big|\, X_t = x\big]}{h} = \mathcal{L}_t u(t,x), \tag{C.3}$$

   where

$$\mathcal{L}_t u(t,x) := \frac{\partial}{\partial t}u(t,x) + \frac{g_t^2}{2}\Delta u(t,x) + \langle \nabla u(t,x), f_{\mathcal{N}}(t,x)\rangle \tag{C.4}$$

   is the generator for the process (28).
4. Since the marginal/joint/conditional distributions of a Gaussian process are still Gaussian, the expectation in (C.3) requires to express Gaussian integrals as differential operators. To this end, the appropriate tool is the "central identity in quantum field theory" (Zee, 2010).
5. Proof concludes by matching terms in (C.3) and (C.4).

Proof of Theorem 3. **From now on, we will invoke the notations in (25) without explicit mentions.**

The static Gaussian SB. We begin by solving the *static* Gaussian SB

$$\min_{\mathbb{P}_{01}} D_{\mathrm{KL}}(\mathbb{P}_{01} \| \mathbb{Q}_{01}) \tag{C.5}$$

over all $\mathbb{P}_{01}$ having marginals $\mathcal{N}(\mu_0, \Sigma_0)$ and $\mathcal{N}(\mu_1, \Sigma_1)$.

Recall that, conditioned on Y_0, $Y_t \sim \mathbb{Q}_t$ is a Gaussian process with mean (23) and covariance (24). Thus, if we only consider the endpoint marginal distributions (Y_0, Y_1), it is easy to derive the transition probability:

$$\mathbb{Q}(Y_1 = y_1 | Y_0 = y_0) = (2\pi)^{\frac{d}{2}} \det(\kappa(1,1)I)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y_1 - \eta(1))^\top (\kappa(1,1)I)^{-1}(y_1 - \eta(1))\right) \tag{C.6}$$

$$= (2\pi)^{\frac{d}{2}} \det(\kappa(1,1)I)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\kappa(1,1)}\|y_1 - \tau_1 y_0 - \zeta(1)\|^2\right). \tag{C.7}$$

Therefore, abusing the notation by continually writing $\mathbb{P}_{01}$ as the relative density of $\mathbb{P}_{01}$ with respect to the Lebesgue measure, we get

$$D_{\mathrm{KL}}(\mathbb{P}_{01} \| \mathbb{Q}_{01}) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \log \frac{\mathrm{d}\mathbb{P}_{01}}{\mathrm{d}\mathbb{Q}_{01}} \, \mathrm{d}\mathbb{P}_{01} \tag{C.8}$$

$$= \text{const.} + \frac{1}{2\kappa(1,1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y' - \tau_1 y - \tau_1 \zeta(1)\|^2 \, \mathrm{d}\mathbb{P}_{01}(y, y') + \int_{\mathbb{R}^d \times \mathbb{R}^d} \log \mathbb{P}_{01} \, \mathrm{d}\mathbb{P}_{01}. \tag{C.9}$$

If $\mathbb{P}_{01}$ is a joint distribution with marginals $Y \sim \mathcal{N}(\mu_0, \Sigma_0)$ and $Y' \sim \mathcal{N}(\mu_1, \Sigma_1)$, then the change of variable $\tilde{Y} = \tau_1 Y + \zeta(1)$ gives rise to a joint distribution $\tilde{\mathbb{P}}_{01}$ having marginals $\tilde{Y} \sim \mathcal{N}\left(\tilde{\mu}_0, \tilde{\Sigma}_0\right)$ and $Y' \sim \mathcal{N}(\mu_1, \Sigma_1)$, where

$$\tilde{\mu}_0 = \tau_1 \mu_0 + \zeta(1), \tag{C.10}$$

$$\tilde{\Sigma}_0 = \tau_1^2 \Sigma_0. \tag{C.11}$$

Obviously, there is a one-to-one correspondence between $\mathbb{P}_{01}$ and $\tilde{\mathbb{P}}_{01}$.

The first integral in (C.9) is equal to $\mathbb{E}\left[\left\|Y' - \tilde{Y}\right\|^2\right]$. On the other hand, we always have

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \log \tilde{\mathbb{P}}_{01} \, \mathrm{d}\tilde{\mathbb{P}}_{01} = \int_{\mathbb{R}^d \times \mathbb{R}^d} \log \mathbb{P}_{01} \, \mathrm{d}\mathbb{P}_{01} + \text{const.}$$

Therefore, minimizing (C.8) over $\mathbb{P}_{01}$ is equivalent to

$$\min_{\tilde{\mathbb{P}}_{01}} D_{\mathrm{KL}}\left(\tilde{\mathbb{P}}_{01} \| \mathbb{Q}_{01}\right) \equiv \min_{\tilde{\mathbb{P}}_{01}} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\|y - y'\|_2^2}{2} \, \mathrm{d}\tilde{\mathbb{P}}_{01}(y, y') + \kappa(1,1) \int_{\mathbb{R}^d \times \mathbb{R}^d} \log \tilde{\mathbb{P}}_{01} \, \mathrm{d}\tilde{\mathbb{P}}_{01}. \tag{C.12}$$

By (3), the solution to (C.12) is given by the joint Gaussian

$$\tilde{\mathbb{P}}_{01}^\star \sim \mathcal{N}\left(\begin{bmatrix}\tilde{\mu}_0 \\ \mu_1\end{bmatrix}, \begin{bmatrix}\tilde{\Sigma}_0 & \tilde{C}_{\tilde{\sigma}} \\ \tilde{C}_{\tilde{\sigma}}^\top & \Sigma_1\end{bmatrix}\right) \tag{C.13}$$

where $\tilde{\sigma} = \sqrt{\kappa(1,1)}$ and

$$\tilde{C}_{\tilde{\sigma}} = \frac{1}{2}\left(\tilde{\Sigma}_0^{\frac{1}{2}} \tilde{D}_{\tilde{\sigma}} \tilde{\Sigma}_0^{-\frac{1}{2}} - \tilde{\sigma}^2 I\right), \tag{C.14}$$

$$\tilde{D}_{\tilde{\sigma}} = \left(4\tilde{\Sigma}_0^{\frac{1}{2}} \Sigma_1 \tilde{\Sigma}_0^{\frac{1}{2}} + \tilde{\sigma}^4 I\right)^{\frac{1}{2}}. \tag{C.15}$$

The optimal static Gaussian SB $\mathbb{P}_{01}^\star$ is then given by the inverse transform $Y = \tau_1^{-1}\left(\tilde{Y} - \zeta(1)\right)$, i.e.,

$$\mathbb{P}_{01}^\star \sim \mathcal{N}\left(\begin{bmatrix}\mu_0 \\ \mu_1\end{bmatrix}, \begin{bmatrix}\Sigma_0 & \tau_1^{-1}\tilde{C}_{\tilde{\sigma}} \\ \tau_1^{-1}\tilde{C}_{\tilde{\sigma}}^\top & \Sigma_1\end{bmatrix}\right). \tag{C.16}$$

Rearranging terms and using (C.14) and (C.15), we get

$$\tau_1^{-1}\tilde{C}_{\tilde{\sigma}} = C_{\sigma_\star} \tag{C.17}$$

where $\sigma_\star = \frac{\kappa(1,1)}{\tau_1}$.

The $\mathbb{Q}$–bridges. For future use, we will need the distribution of Y_t conditioned on Y_0 and Y_1. When $Y_t \equiv \mathbb{W}_t$, the distribution is called the *Brownian bridge*, which is in itself an important subject in mathematics and financial engineering (Mansuy and Yor, 2008). We thus term the conditional distribution of Y_t the $\mathbb{Q}$–*Bridges*.

From (23) and (24), one can infer that, given Y_0, the joint distribution of (Y_t, Y_1) is

$$Y_t, Y_1|Y_0 \sim \mathcal{N}\left(\begin{bmatrix}\eta(t)\\ \eta(1)\end{bmatrix}, \begin{bmatrix}\kappa(t,t)I & \kappa(t,1)I\\ \kappa(t,1)I & \kappa(1,1)I\end{bmatrix}\right). \tag{C.18}$$

Therefore, Lemma C.2 applied implies that, conditioned on Y_0 and Y_1, Y_t is Gaussian with mean

$$\begin{aligned}\mathbb{E}[Y_t \,|\, Y_0, Y_1] &= \eta(t) + \frac{\kappa(t,1)}{\kappa(1,1)}(Y_1 - \eta(1))\\ &= \tau_t Y_0 + \zeta(t) + \frac{\kappa(t,1)}{\kappa(1,1)}(Y_1 - \tau_1 Y_0 - \zeta(1))\\ &= \left(\tau_t - \frac{\kappa(t,1)}{\kappa(1,1)}\tau_1\right)Y_0 + \frac{\kappa(t,1)}{\kappa(1,1)}Y_1 + \zeta(t) - \frac{\kappa(t,1)}{\kappa(1,1)}\zeta(1)\\ &= \bar{r}_t Y_0 + r_t Y_1 + \zeta(t) - \tau_t\zeta(1)\end{aligned} \tag{C.19}$$

and covariance process (for any $t' \geq t$)

$$\mathbb{E}\Big[(Y_t - \mathbb{E}[Y_t \,|\, Y_0, Y_1])(Y_{t'} - \mathbb{E}[Y_{t'} \,|\, Y_0, Y_1])^\top \,\Big|\, Y_0, Y_1\Big] = \left(\kappa(t,t') - \frac{\kappa(t,1)\kappa(t',1)}{\kappa(1,1)}\right)I. \tag{C.20}$$

Since a Gaussian process is uniquely determined by its mean and covariance processes, we have, for some Gaussian process ξ_t independent of Y_t having zero mean and covariance process (C.20),

$$Y_t|Y_0, Y_1 \overset{\text{law}}{=} \bar{r}_t Y_0 + r_t Y_1 + \zeta(t) - \tau_t\zeta(1) + \xi_t. \tag{C.21}$$

From $\mathbb{Q}$–bridges to μ_t and Σ_t. The disintegration formula of $D_{\mathrm{KL}}(\cdot\|\cdot)$ (Léonard, 2013) implies that the solution to (GSB) is given by first generating $(X_0, X_1) \sim \mathbb{P}^\star_{01}$ for $\mathbb{P}^\star_{01}$ in (C.16), and then connecting X_0 and X_1 using the $\mathbb{Q}$–bridges (C.21). Namely,

$$X_t \overset{\text{law}}{=} \bar{r}_t X_0 + r_t X_1 + \zeta(t) - r_t\zeta(t) + \xi_t \tag{C.22}$$

from which (26) and (27) follow by a straightforward calculation. Furthermore, in view of (C.16) and (C.22), X_t is obviously a Gaussian process. Finally, since $\mathbb{Q}_t$ is a Markov process, (Léonard, 2013, Theorem 2.12) implies that $\mathbb{P}_t$ is also Markov. This concludes the first half of Theorem 3.

The SDE representation of X_t. The main idea of proving (29) is to compute

$$\lim_{h\to 0}\frac{\mathbb{E}\big[u\big(t+h, X_{t+h}\big) \,\big|\, X_t = x\big] - u(t,x)}{h} \tag{C.23}$$

and equate (C.23) with the *generator* of (28), which is (Protter, 2005)

$$\mathcal{L}_t u(t,x) := \frac{\partial}{\partial t}u(t,x) + \frac{g_t^2}{2}\Delta u(t,x) + \langle \nabla u(t,x), f_{\mathcal{N}}(t,x)\rangle. \tag{C.24}$$

Since X_t is a Gaussian process, we may derive the conditional expectation in (C.23) using Lemma C.2. However, since eventually we will divide everything by h and drive $h \to 0$, we can ignore any term that is $o(h)$ during the computation. This simple observation will prove to be extremely useful in the sequel.

We first compute the first-order approximation of Σ_t. In view of (27), and since $r_t\kappa(t,1) = \kappa(t,t)\rho_t$ and $\dot{r}_t\kappa(t,1) = r_t\frac{\partial}{\partial t}\kappa(t,1)$, we have

$$\begin{aligned}
\dot{\Sigma}_t &= 2\dot{\bar{r}}_t\bar{r}_t\Sigma_0 + 2\dot{r}_t r_t\Sigma_1 + (\dot{r}_t\bar{r}_t + r_t\dot{\bar{r}}_t)\big(C_{\sigma_\star} + C_{\sigma_\star}^\top\big) + \left(\frac{\partial}{\partial t}\kappa(t,t) - \dot{r}_t\kappa(t,1) - r_t\frac{\partial}{\partial t}\kappa(t,1)\right)I \\
&= \dot{r}_t\big(r_t\Sigma_1 + \bar{r}_t C_{\sigma_\star} + r_t\Sigma_1 + \bar{r}_t C_{\sigma_\star}^\top\big) + \dot{\bar{r}}_t\big(\bar{r}_t\Sigma_0 + r_t C_{\sigma_\star} + \bar{r}_t\Sigma_0 + r_t C_{\sigma_\star}^\top\big) + \left(\frac{\partial}{\partial t}\kappa(t,t) - 2\dot{r}_t\kappa(t,1)\right)I \\
&= \big(P_t + P_t^\top\big) - \big(Q_t + Q_t^\top\big) + \left(\frac{\partial}{\partial t}\kappa(t,t) - 2\dot{r}_t\kappa(t,1)\right)I.
\end{aligned} \tag{C.25}$$

Next, let $K_{t,t+h}$ denote the covariance process of X_t. We can estimate $K_{t,t+h}$ up to first order by computing:

$$\begin{aligned}
K_{t,t+h} &:= \mathbb{E}\Big[(X_t - \mu_t)\big(X_{t+h} - \mu_{t+h}\big)^\top\Big] \\
&= \bar{r}_t\bar{r}_{t+h}\Sigma_0 + r_t r_{t+h}\Sigma_1 + \bar{r}_t r_{t+h}C_{\sigma_\star} + r_t\bar{r}_{t+h}C_{\sigma_\star}^\top + (\kappa(t,t+h) - r_{t+h}\kappa(t,1))I \\
&= \Sigma_t + \bar{r}_t(\bar{r}_{t+h} - \bar{r}_t)\Sigma_0 + r_t(r_{t+h} - r_t)\Sigma_1 + \bar{r}_t(r_{t+h} - r_t)C_{\sigma_\star} + r_t(\bar{r}_{t+h} - \bar{r}_t)C_{\sigma_\star}^\top \\
&\qquad + (\kappa(t,t+h) - \kappa(t,t) - r_{t+h}\kappa(t,1) + r_t\kappa(t,1))I \\
&= \Sigma_t + \frac{r_{t+h} - r_t}{\dot{r}_t}P_t - \frac{\bar{r}_{t+h} - \bar{r}_t}{\dot{\bar{r}}_t}Q_t^\top + (\kappa(t,t+h) - \kappa(t,t) - r_{t+h}\kappa(t,1) + r_t\kappa(t,1))I \\
&= \Sigma_t + h\left\{P_t - Q_t^\top + \left[\left(\frac{\partial}{\partial t'}\kappa\right)(t,t) - \dot{r}_t\kappa(t,1)\right]I\right\} + o(h),
\end{aligned} \tag{C.26}$$

where $\big(\frac{\partial}{\partial t'}\kappa\big)(t,t') := \lim_{h\to 0}\frac{\kappa\left(t,t'+h\right) - \kappa\left(t,t'\right)}{h}$ denotes the derivative of the function $\kappa(t,\cdot)$. Using (24) and $\dot{\tau}_t = c_t\tau_t$, we have

$$\left(\frac{\partial}{\partial t'}\kappa\right)(t,t) = \frac{\partial}{\partial t'}\left(\tau_t\tau_{t'}\int_0^t \tau_s^{-2}g_s^2\,\mathrm{d}s\right)\bigg|_{t'=t} \tag{C.27}$$

$$\begin{aligned}
&= \dot{\tau}_t\tau_t\int_0^t \tau_s^{-2}g_s^2\,\mathrm{d}s \\
&= c_t\kappa(t,t).
\end{aligned} \tag{C.28}$$

On the other hand, we have

$$\begin{aligned}
\dot{r}_t &= \frac{1}{\kappa(1,1)}\frac{\partial}{\partial t}\left(\tau_t\tau_1\int_0^t \tau_s^{-2}g_s^2\,\mathrm{d}s\right) \\
&= \frac{1}{\kappa(1,1)}\big(c_t\kappa(t,1) + \tau_t^{-1}\tau_1 g_t^2\big) \\
&= c_t r_t + \frac{\tau_1 g_t^2}{\tau_t\kappa(1,1)}.
\end{aligned} \tag{C.29}$$

Combining (C.28) and (C.29), using the fact that $r_t\kappa(t,1) = \kappa(t,t)\rho_t$ and $\frac{\tau_1\kappa(t,1)}{\tau_t\kappa(1,1)} = \rho_t$, we may further write (C.26) as

$$\begin{aligned}
K_{t,t+h} &= \Sigma_t + h\big\{P_t - Q_t^\top + \big[c_t\kappa(t,t)(1-\rho_t) - g_t^2\rho_t\big]I\big\} + o(h) \\
&= \Sigma_t + hS_t + o(h).
\end{aligned} \tag{C.30}$$

We are now ready to derive (28). By Lemma C.2, the random variable X_{t+h} conditioned on $X_t = x$ follows $\mathcal{N}\big(\check{\mu}_{t+h}, \check{\Sigma}_{t+h}\big)$ where, by (C.30),

$$\begin{aligned}\check{\mu}_{t+h} &= \mu_{t+h} + K_{t,t+h}^\top \Sigma_t^{-1}(x - \mu_t)\\ &= \mu_t + h\dot{\mu}_t + \big(I + hS_t^\top \Sigma_t^{-1}\big)(x - \mu_t) + o(h)\\ &= x + h\big(S_t^\top \Sigma_t^{-1}(x - \mu_t) + \dot{\mu}_t\big) + o(h),\end{aligned} \tag{C.31}$$

and, by (C.25) and (C.26),

$$\begin{aligned}\check{\Sigma}_{t+h} &= \Sigma_{t+h} - K_{t,t+h}^\top \Sigma_t^{-1} K_{t,t+h}\\ &= \Sigma_t + h\dot{\Sigma}_t - \big(\Sigma_t + hS_t^\top + hS_t\big) + o(h)\\ &= h\Bigg[P_t + P_t^\top - Q_t - Q_t^\top + \left(\frac{\partial}{\partial t}\kappa(t,t) - 2\dot{r}_t\kappa(t,1)\right)I\\ &\qquad - \left(P_t - Q_t^\top + \left[\left(\frac{\partial}{\partial t'}\kappa\right)(t,t) - \dot{r}_t\kappa(t,1)\right]I\right)^\top - \left(P_t - Q_t^\top + \left[\left(\frac{\partial}{\partial t'}\kappa\right)(t,t) - \dot{r}_t\kappa(t,1)\right]I\right)\Bigg] + o(h)\\ &= h\left(\frac{\partial}{\partial t}\kappa(t,t) - 2\left(\frac{\partial}{\partial t'}\kappa\right)(t,t)\right)I + o(h).\end{aligned} \tag{C.32}$$

However, by (24), we have

$$\begin{aligned}\frac{\partial}{\partial t}\kappa(t,t) &= \frac{\partial}{\partial t}\left(\tau_t^2 \int_0^t \tau_s^{-2} g_s^2 \,\mathrm{d}s\right)\\ &= 2\dot{\tau}_t\tau_t \int_0^t \tau_s^{-2} g_s^2 \,\mathrm{d}s + g_t^2,\end{aligned}$$

$$\left(\frac{\partial}{\partial t'}\kappa\right)(t,t) = \frac{\partial}{\partial t'}\left(\tau_t\tau_{t'} \int_0^t \tau_s^{-2} g_s^2 \,\mathrm{d}s\right)\Bigg|_{t'=t} \tag{C.33}$$

$$= \dot{\tau}_t\tau_t \int_0^t \tau_s^{-2} g_s^2 \,\mathrm{d}s, \tag{C.34}$$

from which (C.32) simplifies to

$$\check{\Sigma}_{t+h} = hg_t^2 I + o(h). \tag{C.35}$$

We can now compute $\mathbb{E}\big[u\big(t+h, X_{t+h}\big) \,\big|\, X_t = x\big]$ as follows:

$$\begin{aligned}\mathbb{E}\big[u\big(t+h, X_{t+h}\big) \,\big|\, X_t = x\big] &= (2\pi)^{\frac{d}{2}}\big(\det \check{\Sigma}_{t+h}\big)^{-\frac{1}{2}} \int_{\mathbb{R}^d} u(t+h, x') \exp\left(-\frac{1}{2}(x' - \check{\mu}_{t+h})^\top \check{\Sigma}_{t+h}^{-1}(x' - \check{\mu}_{t+h})\right) \mathrm{d}x'\\ &= (2\pi)^{\frac{d}{2}}\big(\det \check{\Sigma}_{t+h}\big)^{-\frac{1}{2}} \int_{\mathbb{R}^d} u(t+h, x' + \check{\mu}_{t+h}) \exp\left(-\frac{1}{2}x'^\top \check{\Sigma}_{t+h}^{-1} x'\right) \mathrm{d}x'.\end{aligned} \tag{C.36}$$

Invoking Lemma C.1, we see that (C.36) can be evaluated as

$$\mathbb{E}\big[u\big(t+h, X_{t+h}\big) \,\big|\, X_t = x\big] = \exp\left(\frac{1}{2}\partial_{x'}^\top \check{\Sigma}_{t+h}\partial_{x'}\right) u(t+h, x' + \check{\mu}_{t+h})\Big|_{x'=0}. \tag{C.37}$$

Since $\check{\Sigma}_{t+h} = hg_t^2 I + o(h)$ by (C.35), expanding the power series $\exp\big(\frac{1}{2}\partial_{x'}^\top \check{\Sigma}_{t+h}\partial_{x'}\big)$ and ignoring every $o(h)$ terms, (C.37) becomes

$$\mathbb{E}\big[u\big(t+h, X_{t+h}\big) \,\big|\, X_t = x\big] = \left(u(t+h, x' + \check{\mu}_{t+h}) + \frac{hg_t^2}{2}\Delta u(t+h, x' + \check{\mu}_{t+h})\right)\Big|_{x'=0} + o(h)$$
$$= u(t+h, \check{\mu}_{t+h}) + \frac{hg_t^2}{2}\Delta u(t+h, \check{\mu}_{t+h}) + o(h). \tag{C.38}$$

Recalling from (C.31) that $\check{\mu}_{t+h} = x + h\big(S_t^\top \Sigma_t^{-1}(x - \mu_t) + \dot{\mu}_t\big) + o(h)$, the Taylor expansion in the x variable for $u(t,x)$ shows that

$$\mathbb{E}\big[u\big(t+h, X_{t+h}\big) \,\big|\, X_t = x\big] = u(t+h,x) + h\left(\frac{g_t^2}{2}\Delta u(t+h,x) + \big\langle \nabla u(t+h,x), S_t^\top \Sigma_t^{-1}(x-\mu_t) + \dot{\mu}_t\big\rangle\right) + o(h)$$

whence

$$\lim_{h\to 0} \frac{\mathbb{E}\big[u\big(t+h, X_{t+h}\big) \,\big|\, X_t = x\big] - u(t,x)}{h} = \frac{\partial}{\partial t}u(t,x) + \frac{g_t^2}{2}\Delta u(t,x) + \big\langle \nabla u(t,x), S_t^\top \Sigma_t^{-1}(x-\mu_t) + \dot{\mu}_t\big\rangle.$$

This is exactly (C.24) with $f_{\mathcal{N}}(t,x) \leftarrow S_t^\top \Sigma_t^{-1}(x-\mu_t) + \dot{\mu}_t$, which concludes the proof for (28) and (29).

Finally, by (Léonard, 2013, (4.2)), the optimal drift $f_{\mathcal{N}}(t,x)$ is a *gradient field*:

$$f_{\mathcal{N}}(t,x) = \nabla \psi(t,x) \tag{C.39}$$

for some function $\psi : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$, implying that $S_t^\top \Sigma_t^{-1}$ must be symmetric. □

D Additional Details for Section 6

D.1 Further Empirical Validation of Gaussian Approximation

A central thesis of our paper is that Gaussian approximation provides a reasonable initialization for the objectives in (D.5a) and (D.5b). The purpose of the current section is to empirically validate this Gaussian hypothesis via examining the *marginal distributions* of the two real datasets we considered.

We summarize our finding in Fig. 5, from which we can see that many marginal distributions (e.g., the 8-29th principal components of both datasets) can be reasonably approximated by Gaussians *provided that we take their mean and variances into account*. This is precisely the key feature of GSBFLOW, and thus provides empirical evidence favoring our approach.

D.2 Dynamics Reconstruction via GSBFLOW

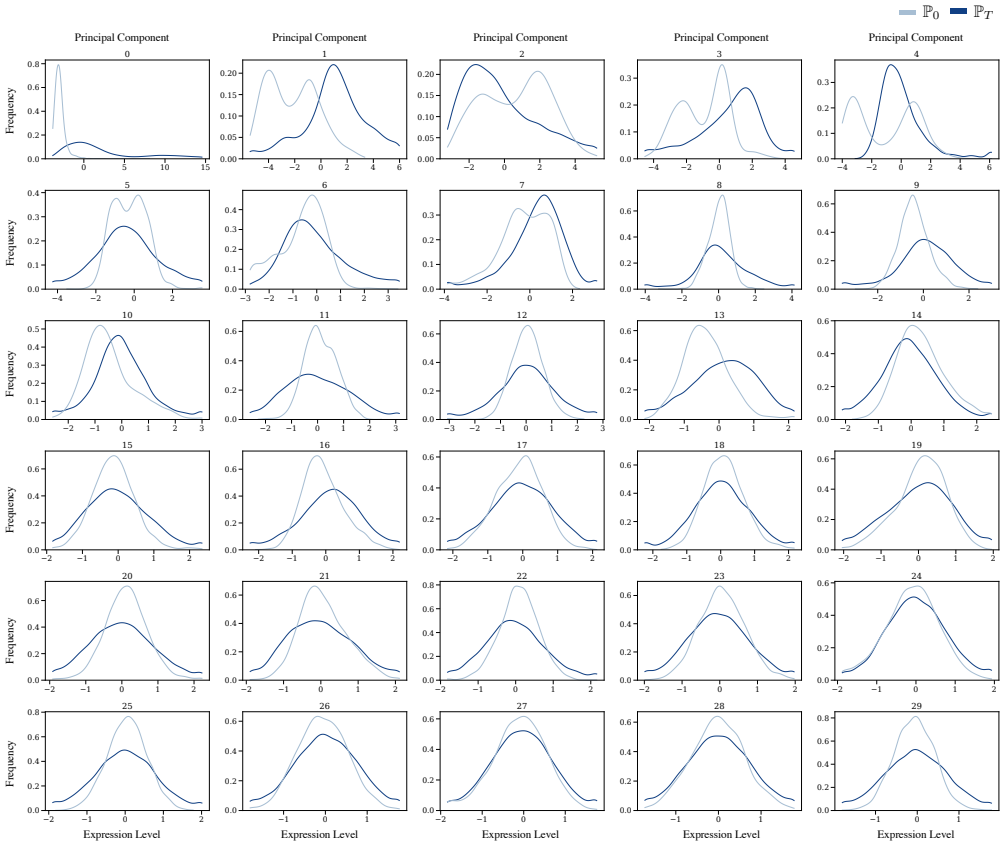
D.2.1 Background on Schrödinger Bridges

Recall the general SB problem (8). It turns out that the solution to (8) is itself given by two coupled SDEs of the form (Léonard, 2013)

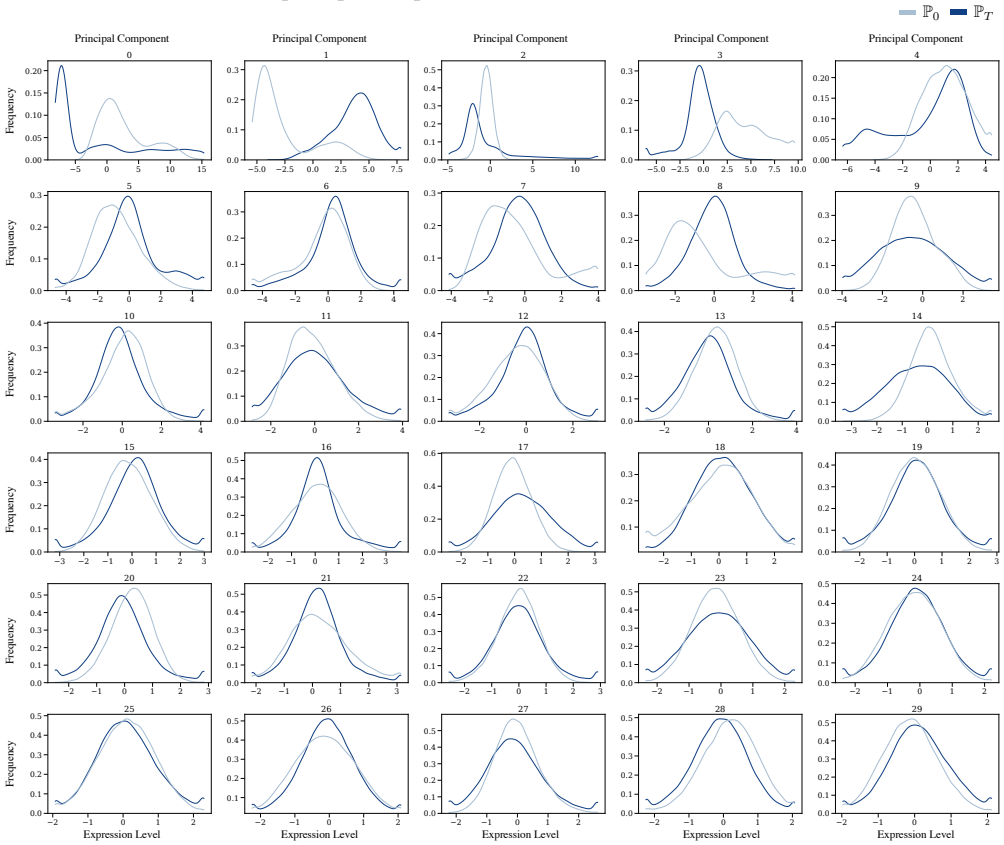
$$\mathrm{d}X_t = (f_t + g_t Z_t)\,\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t,\ X_0 \sim \hat{\mathbb{P}}_0, \tag{D.1a}$$
$$\mathrm{d}X_t = (f_t - g_t \hat{Z}_t)\,\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t,\ X_1 \sim \hat{\mathbb{P}}_1, \tag{D.1b}$$

where $Z_t, \hat{Z}_t \colon \mathbb{R}^d \to \mathbb{R}^d$ are two time-indexed smooth *vector fields* called the optimal forward and backward drift, respectively, and (D.1b) runs backward in time (i.e., from $1 \to 0$). If we parametrize the forward drift by $Z_t^\theta(x)$ and the backward drift by $\hat{Z}_t^\phi(x)$ with some parameters θ, ϕ, then the negative likelihood function for θ and ϕ can be expressed as (Chen et al., 2022)

$$\ell(x_0;\phi) = \int_0^1 \mathbb{E}_{(\text{D.1a})}\left[\frac{1}{2}\|\hat{Z}_t^\phi\|^2 + g\nabla_x \cdot \hat{Z}_t^\phi + \langle Z_t^\theta, \hat{Z}_t^\phi\rangle\,\mathrm{d}t \,\middle|\, X_0 = x_0\right], \tag{D.2a}$$
$$\ell(x_1;\theta) = \int_0^1 \mathbb{E}_{(\text{D.1b})}\left[\frac{1}{2}\|Z_t^\theta\|^2 + g\nabla_x \cdot Z_t^\theta + \langle \hat{Z}_t^\phi, Z_t^\theta\rangle\,\mathrm{d}t \,\middle|\, X_1 = x_1\right]. \tag{D.2b}$$



(a) The first 30 principal components of the dataset from (Moon et al., 2019).



(b) The first 30 principal components of the dataset from (Schiebinger et al., 2019).

Figure 5: Empirical validation of the Gaussian approximation for the considered datasets.

D.2.2 The GSBFLOW Algorithm

Building on the closed-form solutions in Section 5, we present an end-to-end learning paradigm that takes two marginal distributions $\hat{\mathbb{P}}_0, \hat{\mathbb{P}}_1$ to output the reconstruction of the underlying stochastic dynamics $\mathbb{P}_t$. Because our framework relies on GSBs, we call our algorithm the **GSBFLOW**.

Step 1: Moment estimates and GSB initialization. We first compute the means μ_0, μ_1 and covariances Σ_0, Σ_1 of the input distributions, and plug them into (29) and (30)-(33). Note that these computations are done only *once* for every dataset, and can be reused for all subsequent training.

Step 2: Forward and backward pretraining. Denoting by $\mathbb{Q}_t$ the measure of $f_{\mathcal{N}}\,\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t$ in (29), we propose to minimize the objective

$$\min_{\mathbb{P}_0=\hat{\mathbb{P}}_0,\ \mathbb{P}_1=\hat{\mathbb{P}}_1} D_{\mathrm{KL}}(\mathbb{P}_t \| \mathbb{Q}_t). \tag{D.3}$$

Following the framework of Chen et al. (2022), we see that the optimal solution to (D.3) is given by two SDEs of the form:

$$\mathrm{d}X_t = (f_{\mathcal{N}} + g_t Z_t)\,\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t, \quad X_0 \sim \hat{\mathbb{P}}_0, \tag{D.4a}$$

$$\mathrm{d}X_t = \left(f_{\mathcal{N}} - g_t \hat{Z}_t\right)\mathrm{d}t + g_t\,\mathrm{d}\mathbb{W}_t, \quad X_1 \sim \hat{\mathbb{P}}_1, \tag{D.4b}$$

where (D.4b) runs backward in time. After parameterizing Z_t and $\hat{Z}_t$ by two neural networks $Z_t^\theta(x), \hat{Z}_t^\phi(x)$ with parameters θ, ϕ, the corresponding negative likelihood in Appendix D.2.1 becomes

$$\ell(x_0;\phi) = \int_0^1 \mathbb{E}_{(\text{D.4a})}\left[\frac{1}{2}\|\hat{Z}_t^\phi\|^2 + g\nabla_x \cdot \hat{Z}_t^\phi + \langle Z_t^\theta, \hat{Z}_t^\phi\rangle\,\mathrm{d}t \middle| X_0 = x_0\right], \tag{D.5a}$$

$$\ell(x_1;\theta) = \int_0^1 \mathbb{E}_{(\text{D.4b})}\left[\frac{1}{2}\|Z_t^\theta\|^2 + g\nabla_x \cdot Z_t^\theta + \langle \hat{Z}_t^\phi, Z_t^\theta\rangle\,\mathrm{d}t \middle| X_1 = x_1\right]. \tag{D.5b}$$

Following existing work on training SB-based objectives (Chen et al., 2022; De Bortoli et al., 2021b; Vargas et al., 2021), we propose to initialize $\tilde{\theta}_0, \tilde{\phi}_0$ such that $Z_t^{\tilde{\theta}_0}(x), \hat{Z}_t^{\tilde{\phi}_0}(x) \equiv 0$, which can be easily achieved by zeroing out the last layer of the corresponding neural networks. In this case, estimating the conditional expectations in both (D.5a)-(D.5b) reduces to simulating (29) *conditioned* on the given start or end data points. Thanks to our closed-form expressions, this can be easily achieved by drawing Gaussian variables with mean and covariance prescribed in (D.4a)-(D.4b). The pretraining procedure is summarized in Algorithm 1.

Step 3: Alternating minimization. After the pretraining phase, we switch to minimizing (D.5a)-(D.5b) with general drifts in (D.4a)-(D.4b). We carry out this step in an alternating fashion: Since the bottleneck of our framework is to simulate the trajectories of SDEs, we perform several gradient updates for one parameter before drawing another batch of samples. See Algorithm 2 for a summary, and Fig. 1 for an illustration.

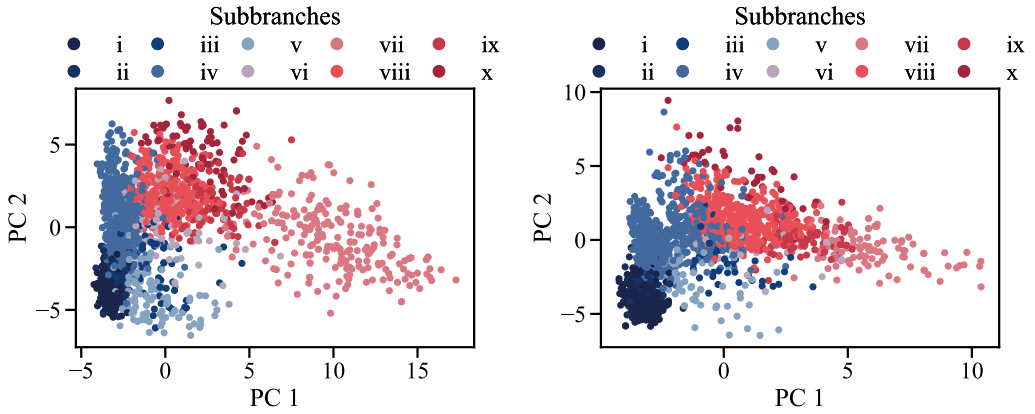
Algorithm 1 Forward and Backward Pretraining

Input: Marginal distributions $\hat{\mathbb{P}}_0, \hat{\mathbb{P}}_1$, initial parameters $\tilde{\theta}_0, \tilde{\phi}_0$ such that $Z_t^{\tilde{\theta}_0}(\cdot\,) = \hat{Z}_t^{\tilde{\phi}_0}(\cdot\,) \equiv 0$, iteration counts K_θ, K_ϕ, learning rates $\gamma_\theta, \gamma_\phi$
Output: Pretrained parameters θ_0, ϕ_0
Initialize $\theta_0 \leftarrow \tilde{\theta}_0$, $\phi_0 \leftarrow \tilde{\phi}_0$.
for $k = 1$ **to** K_ϕ **do**
  Sample X_t from (30)-(31) with $x_0 \sim \hat{\mathbb{P}}_0$
  Compute $\ell(x_0; \phi)$ via (D.5a)
  Update $\phi_0 \leftarrow \phi_0 - \gamma_\phi \nabla \ell(x_0; \phi_0)$
for $k = 1$ **to** K_θ **do**
  Sample X_t from (32)-(33) with $x_1 \sim \hat{\mathbb{P}}_1$
  Compute $\ell(x_1; \theta)$ via (D.5b)
  Update $\theta_0 \leftarrow \theta_0 - \gamma_\theta \nabla \ell(x_1; \theta_0)$

Algorithm 2 GSBFLOW

Input: Marginal distributions $\hat{\mathbb{P}}_0, \hat{\mathbb{P}}_1$, pretrained parameters θ_0, ϕ_0, caching frequency M, iteration counts $K_{\text{in}}, K_{\text{out}}$, learning rates $\gamma_\theta, \gamma_\phi$
Output: Optimal forward and backward drifts $Z_t(\cdot), \hat{Z}_t(\cdot)$ for (D.3)
Initialize $\theta \leftarrow \theta_0$, $\phi \leftarrow \phi_0$.
for $k = 1$ **to** K_{out} **do**
  for $j = 1$ **to** K_{in} **do**
    if $j \mod M = 0$ **then**
      Simulate (D.4a) with $x_0 \sim \hat{\mathbb{P}}_0$
    Compute $\ell(x_0; \phi)$ via (D.5a)
    Update $\phi \leftarrow \phi - \gamma_\phi \nabla \ell(x_0; \phi)$
  for $j = 1$ **to** K_{in} **do**
    if $j \mod M = 0$ **then**
      Simulate (D.4b) with $x_1 \sim \hat{\mathbb{P}}_1$
    Compute $\ell(x_1; \theta)$ via (D.5b)
    Update $\theta \leftarrow \theta - \gamma_\theta \nabla \ell(x_1; \theta)$

D.3 Further Experimental Results



(a) Data (b) GSBFLOW

Figure 6: PCA embedding of the (a) (Moon et al., 2019) data and (b) the GSBFLOW predictions colored by the lineage subbranch class.

Fig. 6 shows the GSBFLOW's prediction of cell lineages on the Moon et al. (2019) dataset but with annotated predictions to cellular sublineages. The result agrees with our observations on the Moon et al. (2019) dataset present in the main text, where GSBFLOW succeeds in learning the cell's differentiation into various cell lineages and capturing biological heterogeneity on a more macroscopic level.

D.4 Single-Cell Datasets

We evaluate GSBFLOW on multiple datasets. This includes synthetic population dynamics, whose results are described in Section 6.1, as well as dynamics of single-cells of a human developmental process, which we cover in Section 6.2.2. In the following, we describe the data collection and data preprocessing steps.

In biology, developmental processes involve complex tasks such as tissue and organ development, body axis formation, cell division, and cell differentiation, i.e., the development of stem cells into functional cell types. One prototype of such a

process is the differentiation of *embryonic stem cells* (ESCs) into distinct lineages such as hematopoietic, cardiac, neural, pancreatic, hepatocytic and germ. It is possible to approximate this development *in vitro* via embryoid bodies (EBs) (Martin and Evans, 1975), three-dimensional aggregates of pluripotent stem cells, including ESCs (Shamblott et al., 2009). Recently, an scRNA-seq analysis is conducted by Moon et al. (2019) in order to unveil the developmental trajectories, as well as cellular and molecular identities through which early lineage precursors emerge from human ESCs. The dataset of Moon et al. (2019) can be found online via Mendeley Data (V6N743H5NG).

In the second task we consider, the natural developmental process is inverted and instead we study the reprogramming of somatic cells into induced pluripotent stem cells (iPSCs). In a recent study, Schiebinger et al. (2019) unveiled this process in depth by measuring 315,000 single-cell RNA sequencing (scRNA-seq) profiles, collected at half-day intervals across 18 days.

In the following, we provide details for the the preprocessing of the raw scRNA-seq data as well as the lineage branch analysis extracting the functional cell types emerging in this developmental process. This dataset can be found via `GEO: GSE122662`.

D.4.1 Data Preprocessing

To preprocess the data, we follow the analysis of Moon et al. (2019) as well as Luecken and Theis (2019). For the analysis, we invoke the Python package `scanpy` (Wolf et al., 2018).

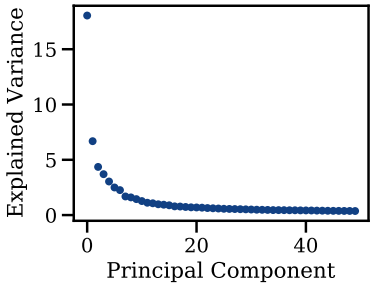


Figure 7: Proportion of explained variance per PC of the embryoid body scRNA-seq data after preprocessing (Moon et al., 2019).

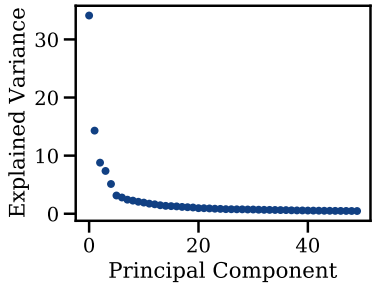
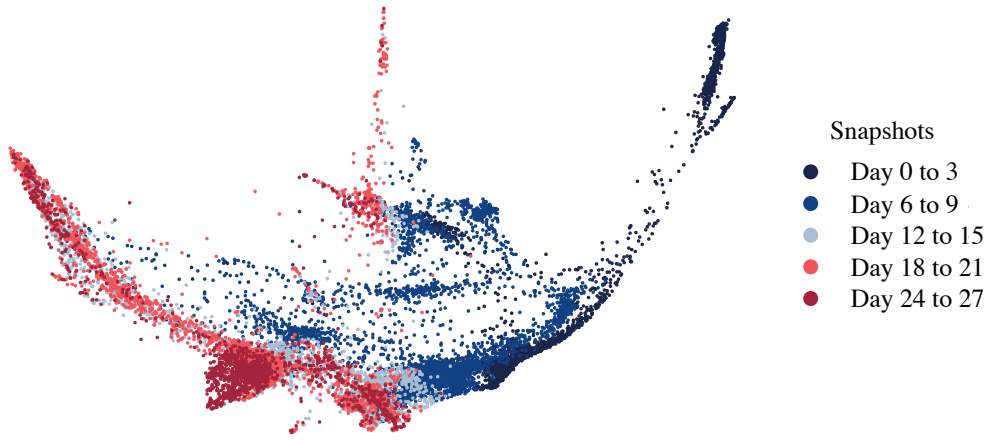


Figure 8: Proportion of explained variance per PC of MEFs reprogramming into iPSCs after preprocessing (Schiebinger et al., 2019).

Moon et al. (2019) take measures of approximately 31,000 cells over a 27-day differentiation time course, comprising gene expression matrices and barcodes, i.e., DNA tags for identifying reads originating from the same cell. Subsequently, the measured cells are filtered in a quality control stage, their gene expression levels normalized and further processed in a feature selection step, where only highly-differentiated genes are selected. The resulting data is then visualized using a standard PCA as well as the dimensionality reduction method of PHATE (Moon et al., 2019) in order to extract biological labels.

The quality control of the data is based on the number of counts per barcode (count depth), the number of genes per barcode, and the fraction of counts from mitochondrial genes per barcode. In our experiments, we only keep cells with at least 4000 and at most 10000 counts, as well as more than 550 expressed genes and less than 20% of mitochondrial counts, as a high fraction is indicative of cells whose cytoplasmic mRNA has leaked out through a broken membrane (Luecken and Theis, 2019). For the subsequent analysis, we further screen out genes which are expressed in less than 10 genes. After quality control procedure described above, the dataset consists of 15150 cells and 17945 genes. We then normalize each cell by total counts over all genes and logarithmize the data matrix. We extract 4000 highly variable genes (HVG) the 10X genomics preprocessing software `Cell Ranger` (Zheng et al., 2017) to further reduce the dimensionality of the dataset and include only the most informative genes. Given the resulting data matrix with 15150 cells and 4000 genes,we compute a corresponding low-dimensional embedding using PCA. Figure 7 shows the proportion of explained variance of each principal component (PC). We use the first 20 or 30 PCs for predicting population dynamics using GSBFLOW.This is in alignment with previous analysis of developmental trajectories which uses 30 PCs (Schiebinger et al., 2019). The same analysis on executed on the mouse embryonic fibroblasts (MEFs) reprogramming dataset by Schiebinger et al. (2019).



(a) PHATE embedding colored by time of snapshot.

(b) PHATE embedding hued by k-Means clustering ($k = 30$).

(c) PHATE embedding hued by predicted lineage branch.

Figure 9: Analysis of embryoid body scRNA-seq data based on PHATE embedding (Moon et al., 2019). Lineage branches are determined based on contiguous k-means clusters.

D.4.2 Lineage Branch Analysis of the Embryoid Body scRNA-Seq Data

In order to annotate the developmental process and detect lineage branches originating from the differentiation of ESCs, we again follow the analysis of Moon et al. (2019). With a 10-dimensional PHATE embedding of the embryoid body scRNA-seq data (see the first two PHATE components in Fig. 9a), we segment the dataset into 30 clusters using k-means. We then assign the resulting cluster to a lineage subbranch (*i* - *x*), using the following assignment of subbranch to cluster identification (see Fig. 9b):

i. 2, 20
ii. 5, 19
iii. 9, 11, 23
iv. 3, 6, 8, 13, 15, 21, 24
v. 0, 7, 14, 25, 28
vi. 16, 18, 27
vii. 4, 10, 12, 17, 22
viii. 1
ix. 26
x. 29.

Then, subbranches are summarized to lineage branches using the assignment in (Moon et al., 2019, Suppl. Note 4):

ESC.	i, ii	Endoderm.	v
Neural Crest.	iii	Mesoderm.	vi, vii
Neuroectoderm.	iv	Other.	viii, ix, x.

The resulting lineage branch annotation of the embryoid body scRNA-seq data can be found in Figure 9c.

D.5 Experimental Details

In the following, we describe model components, as well as provide details on networks architectures and hyperparameters used.

D.5.1 The Reference Processes: VESDEs and sub–VPSDEs

We adopt the VESDE (Song et al., 2021) as the reference process of the GSB

$$\mathrm{d}Y_t = g_t \, \mathrm{d}\mathbb{W}_s, \tag{D.6}$$

where

$$g_t = \sigma_{\min} \left(\frac{\sigma_{\max}}{\sigma_{\min}} \right)^t \sqrt{2 \log \frac{\sigma_{\max}}{\sigma_{\min}}}.$$

Here, $\sigma_{\min}, \sigma_{\max} \in \mathbb{R}^+$ are two hyperparameters that we sweep for each dataset. The corresponding $\mathrm{q}(t)$ in Table 1 is $\sigma_{\min}^2 \left(\frac{\sigma_{\max}}{\sigma_{\min}} \right)^{2t}$, from which one can easily compute all the other functions, and hence (29) as well as (30)-(33).

For existing SB-based methods, we additionally implement the sub–VPSDE (Song et al., 2021), which is determined by a linear function $\beta_t := \beta_{\min} + t(\beta_{\max} - \beta_{\min})$, where as in VESDE, $\beta_{\min}, \beta_{\max} \in \mathbb{R}^+$ are two hyperparameters that we sweep for each dataset. The reference SDE in (7) is given by

$$c_t = -\frac{1}{2}\beta_t,$$
$$g_t = \sqrt{\beta_t \left(1 - e^{-2 \int_0^t \beta_s \, \mathrm{d}s} \right)}.$$

D.5.2 Network Architectures

Forward and backward policies $Z_t^\theta(x), \hat{Z}_t^\phi(x)$ are time-indexed functions parameterizing the optimal forward and backward drift. Throughout the experiments, we parameterize both $Z_t^\theta(x), \hat{Z}_t^\phi(x)$ via multi-layer perceptrons (MLP). The network architectures change with the complexity of the task. We use four hidden layers of size 128 with sigmoid linear units (SiLU, Swish) as activation function for synthetic data and modeling embryoid body development (Moon et al., 2019). In the MEF reprogramming task (Schiebinger et al., 2019) we apply a MLP with five hidden layers.

D.5.3 Hyperparameters and Training

For experiments on synthetic data as well as on the task of (Schiebinger et al., 2019), we train $Z_t^\theta(x)$, $\hat{Z}_t^\phi(x)$ with batch size 1024. For the task of Moon et al. (2019), we set the batch size to 512. For all experiments, we use the Adam optimizer (Kingma and Ba, 2014) with learning rate $\text{lr} = 0.0002$ ($\beta_1 = 0.5$, $\beta_2 = 0.9$). Further, we use exponential moving average (EMA) with the decay rate of 0.99.

The hyperparameters of each SDE class, i.e., VE SDE and sub-VP SDE are chosen based on the underlying dataset. On synthetic data, we set $\sigma_{\max} = 2$ in (D.6), for the single-cell tasks we use $\sigma_{\max} = 10$. For sub-VP SDE we set $\beta_{\max} = 4$ in synthetic data setting, and $\beta_{\max} = 20$ in single-cell experiments. In all experiments, we assume a continuous time variable t.

D.5.4 Additional Evaluation of Single-Cell Experiments

Besides evaluating how well GSBFLOW resembles the spatio-temporal dynamics, we analyze its ability to capture biological heterogeneity. Serving as an *in vitro* model of early embryogenesis, embryoid bodies differentiation captures the development of ESCs into mesoderm, endoderm, neuroectoderm, neural crest and others. Using an initial k-means clustering ($k = 30$) and following Moon et al. (2019, Fig. 6, Suppl. Note 4), we compute lineage branch classes (Fig. 9c) for all cells in a 10-dimensional embedding space using PHATE, a non-linear dimensionality reduction method capturing a denoised representation of both local and global structure of a dataset (Fig. 9b). For details, see Appendix D.4.2.

We then train a k-NN classifier ($k = 5$) to infer the lineage branch class based on a 30-dimensional PCA embedding of a cell (ESC: 0, neural crest: 1, neuroectoderm: 2, endoderm: 3, mesoderm: 4, other: 5). We analyze the captured lineage branch heterogeneity of GSBFLOW's predictions by computing the lineage branch class of each cell using the k-NN classifier. The predicted populations colored by the estimated lineage branch as well as the data with the true lineage branch labels are visualized in Fig. 4c. We study a more fine-grained resolution of lineage branches into subbranches in Appendix D.3 (Fig. 6). This analysis further demonstrates GSBFLOW's ability to learn cells' differentiation into various lineages and to capture biological heterogeneity on a more macroscopic level.

E Reproducibility

An implementation of GSBFLOW is available at `https://github.com/bunnech/gsbflow`.