中国人民大學
RENMIN UNIVERSITY OF CHINA

Recovering Stochastic Dynamics via Gaussian Schrödinger Bridges

Charlotte Bunne*, Ya-Ping Hsieh*, Marco Cuturi, Andreas Krause
https://doi.org/10.48550/arXiv.2202.05722
arXiv preprint arXiv:2202.05722, 2022

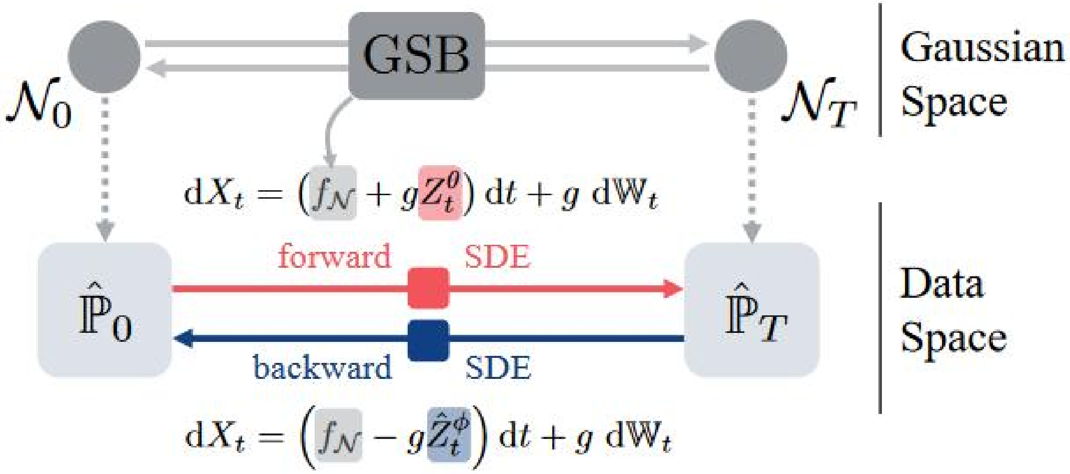
Charlotte Bunne
PhD Candidate in Computer Science, ETH Zurich
Institute for Machine Learning and ETH AI Center
Learning and Adaptive Systems Group
bunnec at inf.ethz.ch

Yajie Zhang
5.26 2020

Recovering Stochastic Dynamics via Gaussian Schrödinger Bridges(GSB)

Main Work: A new framework to reconstruct a stochastic process $\{\mathbb{P}_t : t \in [0,T]\}$ using only samples from its marginal distributions $\widehat{\mathbb{P}}_0$, $\widehat{\mathbb{P}}_T$, observed at start and end times 0 and T .

Applications: Population dynamics inferences, e.g. the time-evolution of cell populations from single-cell sequencing data.



Idea: Relying on Gaussian approximations of the data to provide the reference stochastic process needed to estimate SB.

Results: 1. **Closed-Form Solutions** of Gaussian Schrödinger Bridges(GSB);

2. Dynamics Reconstruction via **GSBFlow**

CONTENTS

Introduction

Problem formulation: Due to the advent of the single-cell RNA sequencing technique, where $\widehat{\mathbb{P}}_0$ and $\widehat{\mathbb{P}}_T$ represent random samples of transcribed genes of a cell at time 0 and time T, such problems have attracted significant attention in biology :

$$\textit{Find a "suitable" } \mathbb{P}_t \textit{ such that } \mathbb{P}_0 \simeq \hat{\mathbb{P}}_0, \mathbb{P}_T \simeq \hat{\mathbb{P}}_T. \quad (\star)$$

Major challenges: the problem ($\star$) is clearly ill-posed because of little prior information about the underlying process $\mathbb{P}_t$.

Introduction

RENMIN UNIVERSITY OF CHINA

- $\widehat{\mathbb{P}}_T \sim \mathcal{N}(0, I)$, $(\star) \rightarrow$ classical *generative modeling* (GM) task, transform a noise distribution $\widehat{\mathbb{P}}_T$ into a complex data distr

Solved by

Solve $(\star)$ means inverting such a "reference process" $\mathbb{Q}_t$ in time(Song et al., 2021)

- ***Score-based Generative Model*** (SGM)
- Define a "reference process" $\mathbb{Q}_t$ that transforms any data distribution $\widehat{\mathbb{P}}_0$ into (almost) $\mathcal{N}(0, I)$. (Sohl-Dickstein et al., 2015; Song and Ermon, 2019; Song et al., 2021)

- SGMs can be seen as approximating an *KL-minimization* problem called the (generalized) *Schrödinger bridge* (SB)

$$\min_{\mathbb{P}_0=\widehat{\mathbb{P}}_0,\ \mathbb{P}_T=\widehat{\mathbb{P}}_T} D_{\mathrm{KL}}(\mathbb{P}_t \| \mathbb{Q}_t), \qquad \text{(SB)}$$

where $\widehat{\mathbb{P}}_0 \equiv p_{data}$, $\widehat{\mathbb{P}}_T = \mathcal{N}(0, I)$.

Background

Schrödinger Bridge:

- A simple case: When $\mathbb{Q}_t = \mathbb{W}_t$, for any marginal $\widehat{\mathbb{P}}_0$, $\mathbb{W}_t$ would predict $\mathbb{P}_T = \widehat{\mathbb{P}}_0 * \mathcal{N}(0, T \cdot I)$. ($\mathbb{W}_t - \mathbb{W}_s \sim \mathcal{N}(0, t-s)$). If $\mathbb{P}_T$ differs from the actual data $\widehat{\mathbb{P}}_T$, then $\mathbb{P}_t$ must also differ from $\mathbb{W}_t$.
- Schrödinger Bridge (SB) problem is identifying $\mathbb{P}_t$ with initial distribution $\widehat{\mathbb{P}}_0$ and terminal distribution $\widehat{\mathbb{P}}_T$ which is the closest to $\mathbb{Q}_t$ in terms of Kullback–Leibler divergence:

$$\min_{\mathbb{P}_0 = \widehat{\mathbb{P}}_0,\ \mathbb{P}_T = \widehat{\mathbb{P}}_T} D_{\mathrm{KL}}(\mathbb{P}_t \| \mathbb{Q}_t), \qquad \text{(SB)}$$

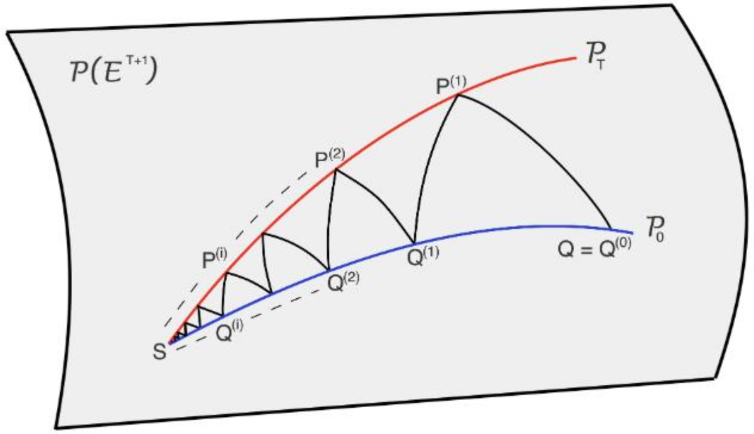
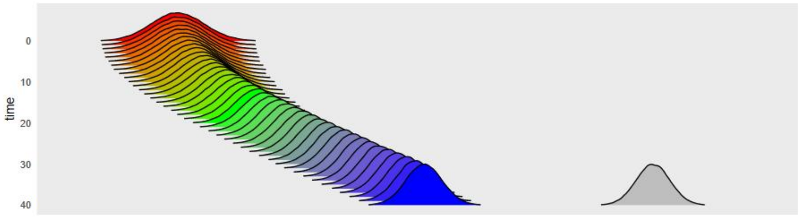
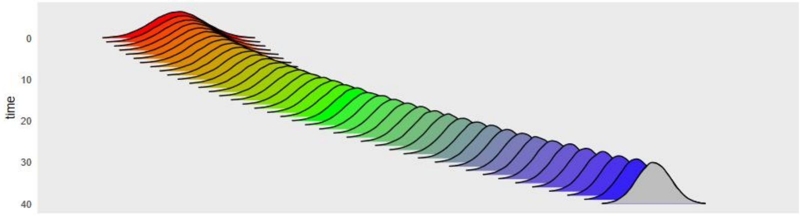


Figure 1: Illustration of the iterative proportional fitting procedure. The blue line represents $\mathcal{P}_0(\pi_0) \subset \mathcal{P}(\mathsf{E}^{T+1})$, denoted $\mathcal{P}_0$ in the figure, while the red line represents $\mathcal{P}_T(\pi_T) \subset \mathcal{P}(\mathsf{E}^{T+1})$, denoted $\mathcal{P}_T$. The black line illustrates that the alternating KL projections $\mathbb{Q}^{(i)} \in \mathcal{P}_0(\pi_0)$ and $\mathbb{P}^{(i)} \in \mathcal{P}_T(\pi_T)$ converge towards the Schrödinger bridge $\mathbb{S}$.



(a) Marginals q_t of a reference process $\mathbb{Q}(\mathrm{d}x_{0:T})$ with $T = 40$ and the target π.



(b) Marginals s_t of the Schrödinger bridge $\mathbb{S}(\mathrm{d}x_{0:T})$ corresponding to the reference process $\mathbb{Q}(\mathrm{d}x_{0:T})$.

Background

Schrödinger Bridges:

In practice, $\mathbb{Q}_t$ is taken to be the measure of an SDE with some initial condition Y_0:

$$\mathrm{d}Y_t = f(t, Y_t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbb{W}_t \equiv f\,\mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t \qquad (2)$$

In this case, it turns out that the solution to (SB) is itself given by two coupled SDEs of the form (Léonard, 2013):

$$\mathrm{d}X_t = (f + gZ_t)\,\mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t,\ X_0 \sim \hat{\mathbb{P}}_0, \qquad (3a)$$

$$\mathrm{d}X_t = (f - g\hat{Z}_t)\,\mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t,\ X_T \sim \hat{\mathbb{P}}_T, \qquad (3b)$$

Parameterize $Z_t, \hat{Z}_t$ by $Z_t^\theta, \hat{Z}_t^\phi$ with θ, ϕ, then the negative likelihood function for θ and ϕ can be expressed as

$$\ell(x_0;\phi) = \int_0^T \mathbb{E}_{(3a)}[\tfrac{1}{2}\|\hat{Z}_t^\phi\|^2 + g\nabla_x \cdot \hat{Z}_t^\phi + \langle Z_t^\theta, \hat{Z}_t^\phi\rangle \mathrm{d}t | X_0 = x_0] \quad (4a)$$

$$\ell(x_T;\theta) = \int_0^T \mathbb{E}_{(3b)}[\tfrac{1}{2}\|Z_t^\theta\|^2 + g\nabla_x \cdot Z_t^\theta + \langle \hat{Z}_t^\phi, Z_t^\theta\rangle \mathrm{d}t | X_T = x_T] \quad (4b)$$

where $\nabla_x \cdot$ denotes the divergence operator w.r.t. the x variable.

Background

SGM seeks to find nonlinear functions that transform simple distributions (typically Gaussian) into complex data distributions. **SB** – as an entropy-regularized optimal transport problem – seeks two optimal policies that transform back-and-forth between two arbitrary distributions in a finite horizon.

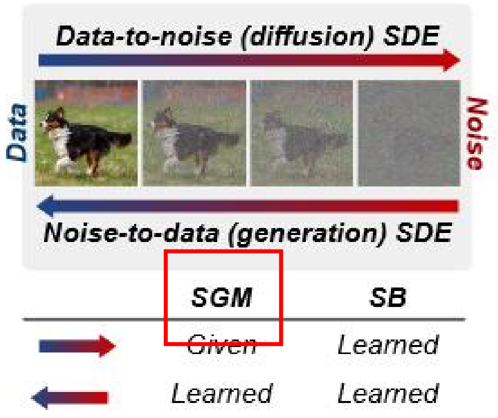


Figure 1: Both Score-based Generative Model (SGM) and Schrödinger Bridge (SB) transform between two distributions. While SGM requires pre-specifying the data-to-noise diffusion, SB instead *learns* the process.

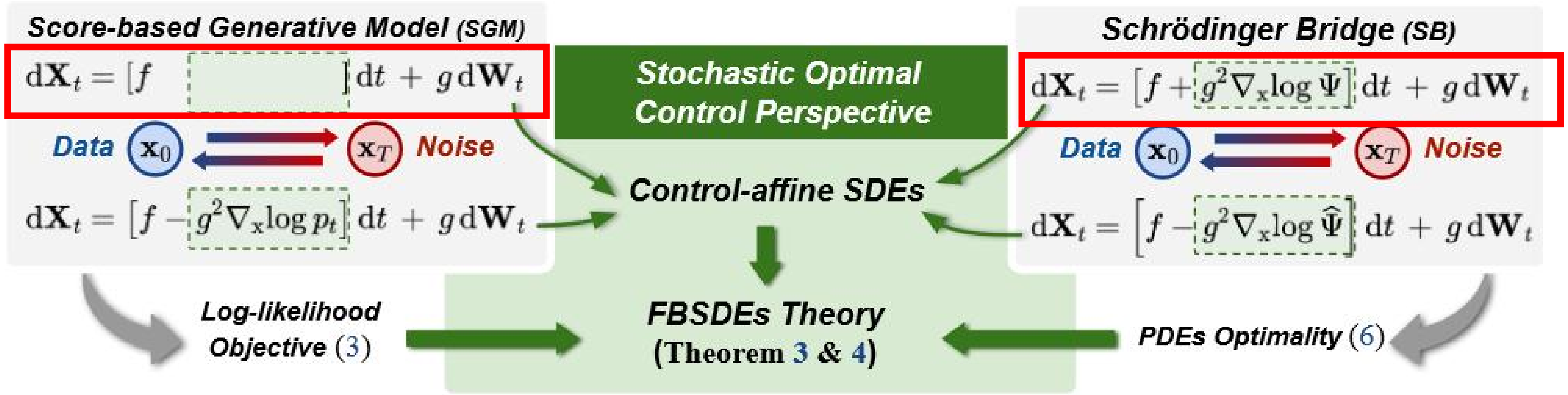


Figure 2: Schematic diagram of the our stochastic optimal control interpretation, and how it connects the objective of SGM (3) and optimality of SB (6) through Forward-Backward SDEs theory.

[Likelihood Training of Schrödinger Bridge using Forward-Backward SDEs Theory]
T. Chen, G.-H. Liu, and E. A. Theodorou. In arXiv Preprint, 2021a.

Background

- The limitations of SGM:

1. The diffusion process (0→T) has to obey a simple form (e.g. linear or degenerate drift) in order to compute the analytic score function for the regression purpose.
2. The diffusion process (0→T) needs to run to sufficiently large time steps so that the end distribution is approximate Gaussian.

In contrast, SB considers a more flexible framework for designing the forward diffusion.

- The limitation of SB: Simulating SDE(3a) with a general drift $Z_t^{\theta_0}(x)$ is expensive.

If $Z_t^{\theta_0} \equiv 0$, it merely simulates the reference SDE Y_t. So we take a pertraining of (SB) in (Chen et al., 2021a):

Pretraining of (SB): initialize θ_0 such that $Z_t^{\theta_0}(\cdot) \equiv 0$ (for instance, $Z_t^{\theta_0}(\cdot)$ being a neural network with zero output weights) and train the backward drift $\hat{Z}_t^{\phi}(x)$ first.

Background

$$\min_{\mathbb{P}_0=\hat{\mathbb{P}}_0,\ \mathbb{P}_T=\hat{\mathbb{P}}_T} D_{\mathrm{KL}}(\mathbb{P}_t \| \mathbb{Q}_t), \qquad \text{(SB)}$$

Strength: (SB) remains a reasonable objective for arbitrary $\hat{\mathbb{P}}_T$ once a “coarse approximation” $\mathbb{Q}_t$ of $\mathbb{P}_t$ is chosen.

Weakness:

1. Choosing an overly simple $\mathbb{Q}_t$ in (SB), such as Brownian motions, presumes unrealistic dynamics of $\mathbb{P}_t$ and leads to solutions with **poor data-fit**.
2. It is **unclear how to design a more sophisticated** $\mathbb{Q}_t$ that takes data into account and retaining the necessary numerical efficiency of the simple dynamics.

Result1. Closed-Form Solutions of Gaussian Schrödinger Bridges

- Consider Gaussian Schrödinger bridges (GSBs) :

$$\min_{\mathbb{P}_0=\mathcal{N}_0,\, \mathbb{P}_T=\mathcal{N}_T} D_{\mathrm{KL}}(\mathbb{P}_t \| \mathbb{Q}_t) \qquad (\mathbf{SB}_{\mathcal{N}})$$

- $\mathbb{Q}_t$ ($\mathrm{SB}_{\mathcal{N}}$): a linear SDE, with Y_0:

$$\mathrm{d}Y_t = (c(t)Y_t + \alpha(t))\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbb{W}_t \qquad (5)$$

$$Y_t = \tau_t\left(Y_0 + \int_0^t \tau_s^{-1}\alpha(s)\,\mathrm{d}s + \int_0^t \tau_s^{-1}g(s)\,\mathrm{d}\mathbb{W}_s\right) \qquad (6)$$

where $\tau_t := \exp\left(\int_0^t c(s)\,\mathrm{d}s\right)$.

- Y_t is a Gaussian process given any Y_0, and is thus characterized by the first two moments：

$$\mathbb{E}[Y_t \,|\, Y_0] = \tau_t\left(Y_0 + \int_0^t \tau_s^{-1}\alpha(s)\,\mathrm{d}s\right) =: \eta(t) \qquad (7)$$

$$\mathbb{E}\left[(Y_t - \eta(t))(Y_{t'} - \eta(t'))^\top \,\middle|\, Y_0\right] = \left(\tau_t\tau_{t'}\int_0^t \tau_s^{-2}g^2(s)\,\mathrm{d}s\right)I =: \kappa(t,t')I. \qquad (8)$$

R1. Closed-Form Solutions of Gaussian Schrödinger Bridges

中国人民大学 RENMIN UNIVERSITY OF CHINA

Theorem 1. *Denote by $\mathbb{P}_t^\star$ the solution to* (SB$_\mathcal{N}$). *Set*

$$r_t := \frac{\kappa(t,T)}{\kappa(T,T)}, \quad \bar{r}_t := \tau_t - r_t\tau_T, \quad \sigma_\star := \sqrt{\tau_T^{-1}\kappa(T,T)},$$

$$\zeta(t) := \tau_t \int_0^t \tau_s^{-1}\alpha(s)\,\mathrm{d}s, \quad \rho_t := \frac{\int_0^t \tau_s^{-2} g^2(s)\,\mathrm{d}s}{\int_0^T \tau_s^{-2} g^2(s)\,\mathrm{d}s},$$

$$P_t := \dot{r}_t(r_t\Sigma_T + \bar{r}_t C_{\sigma_\star}), \quad Q_t := -\dot{\bar{r}}_t(\bar{r}_t\Sigma_0 + r_t C_{\sigma_\star}),$$

$$S_t := P_t - Q_t^\top + \left[c(t)\kappa(t,t)(1-\rho_t) - g^2(t)\rho_t\right] I. \quad (9)$$

Then the following holds:

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

$$\mu_t^\star := \bar{r}_t\mu_0 + r_t\mu_T + \zeta(t) - r_t\zeta(T), \quad (10)$$

$$\Sigma_t^\star := \bar{r}_t^2\Sigma_0 + r_t^2\Sigma_T + r_t\bar{r}_t\left(C_{\sigma_\star} + C_{\sigma_\star}^\top\right) + \kappa(t,t)(1-\rho_t)I. \quad (11)$$

2. *$X_t^\star$ admits a closed-form solution as the SDE:*

$$\mathrm{d}X_t^\star = f_\mathcal{N}(t, X_t^\star)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbb{W}_t \quad (12)$$

where

$$f_\mathcal{N}(t,x) := S_t^\top \Sigma_t^{\star-1}(x - \mu_t^\star) + \dot{\mu}_t^\star. \quad (13)$$

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

Where $\mathcal{N}_0 = \mathcal{N}(\mu_0, \Sigma_0)$, $\mathcal{N}_T = \mathcal{N}(\mu_T, \Sigma_T)$ are two arbitrary Gaussian distributions in (SBN), and

$$D_\sigma := \left(4\Sigma_0^{\frac{1}{2}}\Sigma_T\Sigma_0^{\frac{1}{2}} + \sigma^4 I\right)^{\frac{1}{2}}, \quad C_\sigma := \frac{1}{2}\left(\Sigma_0^{\frac{1}{2}} D_\sigma \Sigma_0^{-\frac{1}{2}} - \sigma^2 I\right)$$

R1. Closed-Form Solutions of Gaussian Schrödinger Bridges

- **Examples**

SDE with $\alpha(t)\equiv 0$	Setting	$\kappa(t,t')$	$\sigma_\star^2$	r_t	$\bar{r}_t$	ρ_t	$\zeta(t)$
BM	$c(t)\equiv 0$ $g(t)\equiv\omega\in\mathbb{R}^+$	$\omega^2 t$	$\omega^2 T$	$\frac{t}{T}$	$1-\frac{t}{T}$	$\frac{t}{T}$	0
VE SDE	$c(t)\equiv 0$ $g(t)=\sqrt{\dot{\mathrm{q}}(t)}$	$\mathrm{q}(t)$	$\mathrm{q}(T)$	$\frac{\mathrm{q}(t)}{\mathrm{q}(T)}$	$1-\frac{\mathrm{q}(t)}{\mathrm{q}(T)}$	$\frac{\mathrm{q}(t)}{\mathrm{q}(T)}$	0
VP SDE	$-2c(t)=g^2(t)$	$\tau_{t'}(\tau_t^{-1}-\tau_t)$	$\tau_T^{-1}-\tau_T$	$\frac{\tau_t^{-1}-\tau_t}{\tau_T^{-1}-\tau_T}$	$\tau_T\left(\frac{\tau_t}{\tau_T}-\frac{\tau_t^{-1}-\tau_t}{\tau_T^{-1}-\tau_T}\right)$	$\frac{\tau_t^{-1}(\tau_t^{-1}-\tau_t)}{\tau_T^{-1}(\tau_T^{-1}-\tau_T)}$	0
sub–VP SDE	$\frac{g^2(t)}{-2c(t)}=1-\tau_t^4$	$\tau_t\tau_{t'}(\tau_t^{-1}-\tau_t)^2$	$\tau_T(\tau_T^{-1}-\tau_T)^2$	$\frac{\tau_t}{\tau_T}\cdot\left(\frac{\tau_t^{-1}-\tau_t}{\tau_T^{-1}-\tau_T}\right)^2$	$\tau_t\left(1-\left(\frac{\tau_t^{-1}-\tau_t}{\tau_T^{-1}-\tau_T}\right)^2\right)$	$\left(\frac{\tau_t^{-1}-\tau_t}{\tau_T^{-1}-\tau_T}\right)^2$	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	$c(t)\equiv-\lambda\in\mathbb{R}$ $\alpha(t)\equiv\mathrm{v}\in\mathbb{R}^d$ $g(t)\equiv\omega\in\mathbb{R}^+$	$\frac{\omega^2 e^{-\lambda t'}\sinh\lambda t}{\lambda}$	$\frac{\omega^2\sinh\lambda T}{\lambda}$	$\frac{\sinh\lambda t}{\sinh\lambda T}$	$\sinh\lambda t\coth\lambda t$ $-\sinh\lambda t\coth\lambda T$	$e^{-\lambda(T-t)}$ $\cdot\frac{\sinh\lambda t}{\sinh\lambda T}$	$\frac{\mathrm{v}}{\lambda}(1-e^{-\lambda t})$
$\alpha(t)$-BDT	$c(t)\equiv 0$ $g(t)\equiv\omega\in\mathbb{R}^+$	$\omega^2 t$	$\omega^2 T$	$\frac{t}{T}$	$1-\frac{t}{T}$	$\frac{t}{T}$	$\int_0^t\alpha(s)\,\mathrm{d}s$

Table 1. Examples of SDEs and the functions relevant to the closed-form solutions of ($\mathrm{SB}_{\mathcal{N}}$).

R1. Closed-Form Solutions of Gaussian Schrödinger Bridges

- **Comparison with Existing Work**

➢ All results in Theorem 1 are novel, except for the case which $c(t) \equiv a(t) \equiv 0$ and $g(t) \equiv const$. (Mallasto et al. (2021))

➢ The solutions in Theorem 1 are explicitly expressed in terms of the inputs μ_0, μ_T, Σ_0, Σ_T which are amenable to efficient implementation. (The key difference between with Chen et al. (2015; 2016))

R2. Dynamics Reconstruction via GSBFLOW

- Resolve the above issues via constructing a data-informed reference process $\mathbb{Q}_t^\star$ based on $\hat{\mathbb{P}}_0$ and $\hat{\mathbb{P}}_T$.

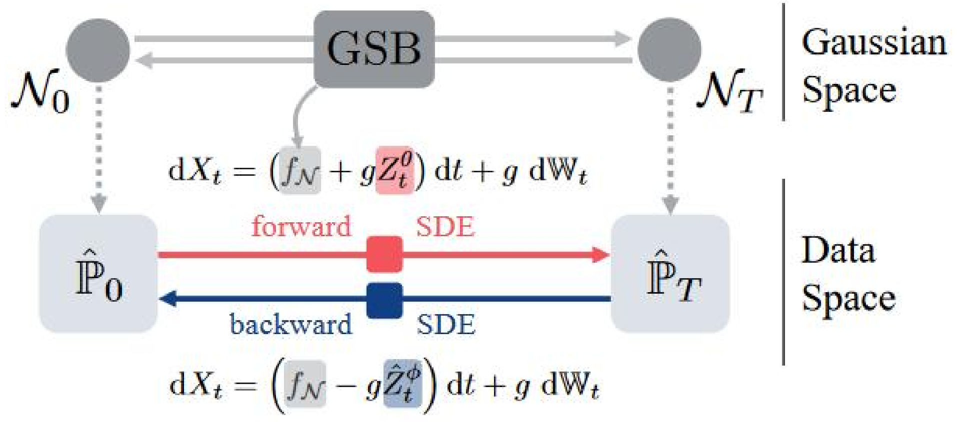
1. Pretaining: derive the closed-form solutions for a simpler variant of (SB) with Gaussian marginal:

$$\mathbb{Q}_t^\star := \underset{\mathbb{P}_0=\mathcal{N}_0,\, \mathbb{P}_T=\mathcal{N}_T}{\operatorname{argmin}} D_{\mathrm{KL}}(\mathbb{P}_t \| Y_t)$$

where $\mathcal{N}_0$, $\mathcal{N}_T$ are two normal distributions and Y_t is a wide class of (SDEs) adopted for reference process for (SB).

2. Learn (SB) with $\mathbb{Q}_t$ replaced by $\mathbb{Q}_t^\star$:

$$\min_{\mathbb{P}_0=\hat{\mathbb{P}}_0,\, \mathbb{P}_T=\hat{\mathbb{P}}_T} D_{\mathrm{KL}}(\mathbb{P}_t \| \mathbb{Q}_t^\star).$$



R2. Dynamics Reconstruction via GSBFLOW

中国人民大学
RENMIN UNIVERSITY OF CHINA

GSBFLOW ALGORITHM:

Step 1: Moment estimates and GSB initialization.

- Compute μ_0, μ_T, Σ_0, Σ_T, and plug them into (13) and (16)~(19)

Step 2: Forward and backward pretraining.

- Denote the measure of $f_{\mathcal{N}}\,dt + g\,d\mathbb{W}_t$ by Q_t^*, then

$$\min_{\mathbb{P}_0=\hat{\mathbb{P}}_0,\ \mathbb{P}_T=\hat{\mathbb{P}}_T} D_{\mathrm{KL}}(\mathbb{P}_t \| \mathbb{Q}_t^\star). \tag{20}$$

- Following (3a)-(3b), the optimal solution to (20) is given by two SDEs of the form:

$$\mathrm{d}X_t = (f_{\mathcal{N}} + gZ_t)\,\mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t, \quad X_0 \sim \hat{\mathbb{P}}_0, \tag{21a}$$

$$\mathrm{d}X_t = \left(f_{\mathcal{N}} - g\hat{Z}_t\right)\mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t, \quad X_T \sim \hat{\mathbb{P}}_T, \tag{21b}$$

- Parameterize Z_t, $\hat{Z}_t$ by $Z_t^\theta, \hat{Z}_t^\phi$ with θ, ϕ, the corresponding negative likelihood becomes

$$\ell(x_0;\phi) = \int_0^T \mathbb{E}_{(21a)}[\tfrac{1}{2}\|\hat{Z}_t^\phi\|^2 + g\nabla_x\cdot\hat{Z}_t^\phi + \langle Z_t^\theta, \hat{Z}_t^\phi\rangle \mathrm{d}t | X_0 = x_0] \tag{22a}$$

$$\ell(x_T;\theta) = \int_0^T \mathbb{E}_{(21b)}[\tfrac{1}{2}\|Z_t^\theta\|^2 + g\nabla_x\cdot Z_t^\theta + \langle \hat{Z}_t^\phi, Z_t^\theta\rangle \mathrm{d}t | X_T = x_T] \tag{22b}$$

Step 3: Alternating minimization.

- Minimize (22a)-(22b) with general drifts in (21a)-(21b) in alternating fashion.

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

$$\begin{aligned}\mu_{t|0}^\star &= \bar{r}_t x_0 + r_t\left(\mu_T + C_{\sigma_\star}^\top \Sigma_0^{-1}(x_0 - \mu_0)\right) + \zeta(t) - r_t\zeta(T)\\ &= \bar{r}_t x_0 + r_t\mu_{T|0}^\star + \zeta(t) - r_t\zeta(T),\end{aligned} \tag{16}$$

$$\begin{aligned}\Sigma_{t|0}^\star &= r_t^2\left(\Sigma_T - C_{\sigma_\star}^\top \Sigma_0^{-1} C_{\sigma_\star}\right) + \kappa(t,t)(1-\rho_t)I\\ &= r_t^2\Sigma_{T|0}^\star + \kappa(t,t)(1-\rho_t)I.\end{aligned} \tag{17}$$

Similarly, $X_t^\star | X_T^\star = x_T \sim \mathcal{N}\left(\mu_{t|T}^\star, \Sigma_{t|T}^\star\right)$, where

$$\begin{aligned}\mu_{t|T}^\star &= r_t x_T + \bar{r}_t\left(\mu_0 + C_{\sigma_\star}\Sigma_T^{-1}(x_T - \mu_T)\right) + \zeta(t) - r_t\zeta(T)\\ &= r_t x_T + \bar{r}_t\mu_{0|T}^\star + \zeta(t) - r_t\zeta(T),\end{aligned} \tag{18}$$

$$\begin{aligned}\Sigma_{t|T}^\star &= \bar{r}_t^2\left(\Sigma_0 - C_{\sigma_\star}\Sigma_T^{-1} C_{\sigma_\star}^\top\right) + \kappa(t,t)(1-\rho_t)I\\ &= \bar{r}_t^2\Sigma_{0|T}^\star + \kappa(t,t)(1-\rho_t)I.\end{aligned} \tag{19}$$

R2. Dynamics Reconstruction via GSBFLOW

中国人民大学
RENMIN UNIVERSITY OF CHINA

Algorithm 1 Forward and Backward Pretraining

Input: Marginal distributions $\hat{\mathbb{P}}_0, \hat{\mathbb{P}}_T$, 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 (16)-(17) with $x_0 \sim \hat{\mathbb{P}}_0$

  Compute $\ell(x_0; \phi)$ via (22a)

  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 (18)-(19) with $x_T \sim \hat{\mathbb{P}}_T$

  Compute $\ell(x_T; \theta)$ via (22b)

  Update $\theta_0 \leftarrow \theta_0 - \gamma_\theta \nabla \ell(x_T; \theta_0)$

Algorithm 2 GSBFLOW

Input: Marginal distributions $\hat{\mathbb{P}}_0, \hat{\mathbb{P}}_T$, 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 (20)

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 (21a) with $x_0 \sim \hat{\mathbb{P}}_0$

    Compute $\ell(x_0; \phi)$ via (22a)

    Update $\phi \leftarrow \phi - \gamma_\phi \nabla \ell(x_0; \phi)$

  for $j = 1$ **to** K_{in} **do**

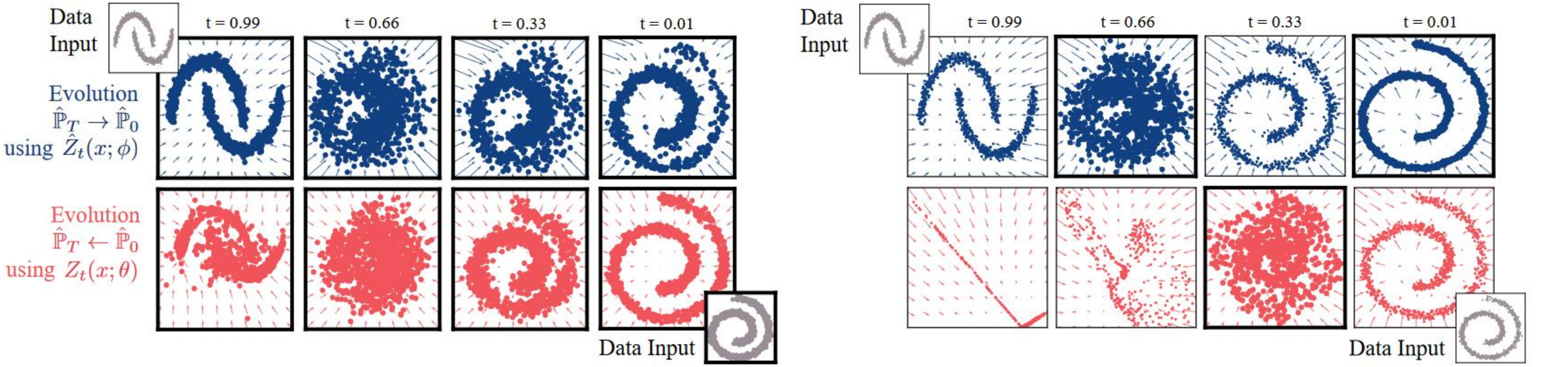
    if $j \mod M = 0$ **then**

      Simulate (21b) with $x_T \sim \hat{\mathbb{P}}_T$

    Compute $\ell(x_T; \theta)$ via (22b)

    Update $\theta \leftarrow \theta - \gamma_\theta \nabla \ell(x_T; \theta)$

Experiments1. Synthetic Dynamics



(a) GSBFLOW with VE SDE.

(b) Chen et al. (2021a) with VE SDE.

Figure 2. Illustration of the time-dependent drifts learned by GSBFLOW compared to those recovered by the method proposed by Chen et al. (2021a) for two toy marginal distributions. *Top.* Evolution of $\hat{\mathbb{P}}_T$ (moons) $\to$ $\hat{\mathbb{P}}_0$ (spiral) via backward policy $\hat{Z}_t^{\phi}(x)$. *Bottom.* Evolution of $\hat{\mathbb{P}}_0$ (spiral) $\to$ $\hat{\mathbb{P}}_T$ (moons) via forward policy $Z_t^{\theta}(x)$.

- As demonstrated in Fig.2a, GSBFLOW is able to successfully learn both policies Z_t^{θ}, $\hat{Z}_t^{\phi}$ and reliably recovers the corresponding targets of the forward and backward evolutionary processes.
- In Fig.2b, (Chen et al., 2021a) fail to recover the forward policy Z_t^{θ}. While this shortcut is feasible in low-dimensional data regimes, it gives rise to numerical instabilities and malfunctions in more complex problems.

Experiments2. Single-Cell Dynamics

中国人民大學
RENMIN UNIVERSITY OF CHINA

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

Method	**Tasks**					
	Generation ($\mathcal{N}_0 \to \hat{\mathbb{P}}_T$) Wasserstein Loss $W_\varepsilon \downarrow$		**Evolution** ($\hat{\mathbb{P}}_0 \to \hat{\mathbb{P}}_T$) Wasserstein Loss $W_\varepsilon \downarrow$			
	Moon et al. 2019	Schiebinger et al. 2019	Moon et al. 2019	Schiebinger et al. 2019	Moon et al. (2019) (scaled)	Schiebinger et al. (2019) (scaled)
Chen et al. (2021a)						
VESDE	20.83 ± 0.18	40.81 ± 0.42	45.01 ± 2.74	62.13 ± 3.81	Num. Inst.[1]	Num. Inst.[1]
sub-VPSDE	$\mathbf{19.96 \pm 0.58}$	48.15 ± 3.38	59.15 ± 44.08	Num. Inst.[1]	Num. Inst.[1]	Num. Inst.[1]
GSBFLOW (*ours*)						
VESDE	25.18 ± 0.10	$\mathbf{27.85 \pm 0.68}$	$\mathbf{43.39 \pm 0.93}$	$\mathbf{59.99 \pm 1.73}$	$\mathbf{40.64 \pm 1.18}$	$\mathbf{58.55 \pm 1.07}$

[1]No predictions available due to numerical instabilities.

Conclusion

Schrödinger bridges promise to provide a powerful tool to infer a wide variety of paths linking particles between two marginals. Estimating such bridges requires determining a backward and forward drift for a pair of coupled SDEs that progressively transform one measure into the other.

Contributions:

1. Derive a closed-form solution for the Schrödinger bridge between two Gaussians (GSB).
2. Use these closed-forms on the Gaussian approximation of the marginals, which provides the prior knowledge necessary to estimate the original Schrödinger bridge between the target measures.
3. Their experiments show that this approach performs significantly better than all existing alternatives.

Limitation:

1. The assumption that the first and second order moments of the marginals are informative.
2. Estimating them in high-dimensions are challenging on its own

References

Some reviews about SBP and OMT:

1. Y. Chen, T. T. Georgiou, and M. Pavon. Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger bridge. SIAM Review, 63(2), 2021b.
2. Y. Chen, T.T. Georgiou and M. Pavon, Optimal Transport in Systems and Control, Annual Review of Control, Robotics, and Autonomous Systems, vol.4, 2021, to appear.
3. Y. Chen, T.T. Georgiou and M. Pavon, On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint, J. Optim. Theory and Applic., 169(2): 671-691, 2016.
4. C. L´eonard, A survey of the Schrödinger problem and some of its connections with optimal transport, Discrete Contin. Dyn. Syst. A, 2014, 34 (4): 1533-1574.

SUMMARY POINTS

1. (OMT) can be cast as a stochastic control problem.
2. (SBP) was conceived as the inference problem of finding the most likely random evolution linking boundary marginal distributions.
3. **(SBP) is seen as an entropy-regularized version of (OMT).**
4. A discrete space counterpart of either, OMT or SBP, relates to control problems for Markov Decision Processes (MDPs) and transport over networks.

FUTURE ISSUES

1. (OMT) and (SBP) represent rapidly developing subjects, with a rich mathematical basis, that impacts a range of scientific disciplines beyond Systems and Control.
2. (OMT) and (SBP) have helped launch a new sub-discipline of Stochastic Control, namely, Control of Uncertainty, where many technical and computations issues remain open.

References

Related works

1. T. Chen, G.-H. Liu, and E. A. Theodorou. Likelihood Training of Schrödinger Bridge using Forward-Backward SDEs Theory. In arXiv Preprint, 2021a.
2. V. De Bortoli, J. Thornton, J. Heng, Arnaud Doucet. Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling. arXiv preprint arXiv:2106.01357, 2021.
3. 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.
4. Chow, SN., Li, W., Mou, C. et al. Dynamical Schrödinger Bridge Problems on Graphs. J Dyn Diff Equat (2021).
5. C. Bunne, L. Meng-Papaxanthos, A. Krause, and M. Cuturi. Proximal Optimal Transport Modeling of Population Dynamics. In International Conference on Artificial Intelligence and Statistics (AISTATS), volume 25, 2022.
6. E. Bernton, J. Heng, A. Doucet, and P. E. Jacob. Schrödinger Bridge Samplers. In arXiv Preprint, 2019.

THANKS !