A variational approach to large deviations in Schrödinger bridges

Viktor Nilsson, KTH Royal Institute of Technology
Dept. of Mathematics
Joint work with Pierre Nyquist

Background

Recently, both optimal transport and Schrödinger bridge problems have found foundational value for many methods within the burgeoning field of generative diffusion models. The relation between the two problem types is a widely researched topic.
Schrödinger Bridges can be viewed as an analog of, or alternative to, entropic regularization for optimal transport problems. Like entropic regularization, dynamic Schrödinger Bridges (using Brownian priors) converge to the dynamic OT-solution as the regularization parameter goes to zero. Detailed information of such convergence can be obtained using the theory of large deviations.
Recent work has derived large deviation principles for both the static and dynamic Schrödinger Bridge problems. In this talk we show how such results can be obtained using the well-established weak convergence approach to large deviations. This systematic approach opens the door for generalizing the results to settings found in the machine learning literature. We provide an overview over current and anticipated results for common Schrödinger bridge setups from machine learning, with varying domains and prior processes.

Schrödinger bridges / OT

For fixed ε consider the diffusion X^ε given as the solution of

$$\begin{aligned} dX^\varepsilon(t) &= f(t, X^\varepsilon(t))dt + \sqrt{\varepsilon} g(t) dW(t),\ t \in [0,1] \\ X^\varepsilon(0) &= \xi \sim \mu_0, \end{aligned} \tag{1}$$

This gives rise to a path measure $\mathbb{Q}$ on $C([0,1] : \mathbb{R}^d)$. The Schrödinger bridge problem is to find the path measure $\mathbb{P}$ minimizing

$$\min_{\mathbb{P}:\mathbb{P}_0=\mu_0,\mathbb{P}_1=\mu_1} R(\mathbb{P}\,||\,\mathbb{Q}) = \mathbb{E}^{\mathbb{P}}\left[\log\left(\frac{d\,\mathbb{P}}{d\,\mathbb{Q}}\right)\right]. \tag{2}$$

When $f = 0$ and $g = 1$, the Schrödinger bridge problem is equivalent to the entropically regularized optimal transport problem:

$$\min_{\mathbb{P}\in\Pi(\mu_0,\mu_1)} \int \frac{1}{2}|x-y|^2\, \mathbb{P}(dx,dy) + \varepsilon R(\mathbb{P}\,||\mu_0 \otimes \mu_1). \tag{3}$$

Computational advances

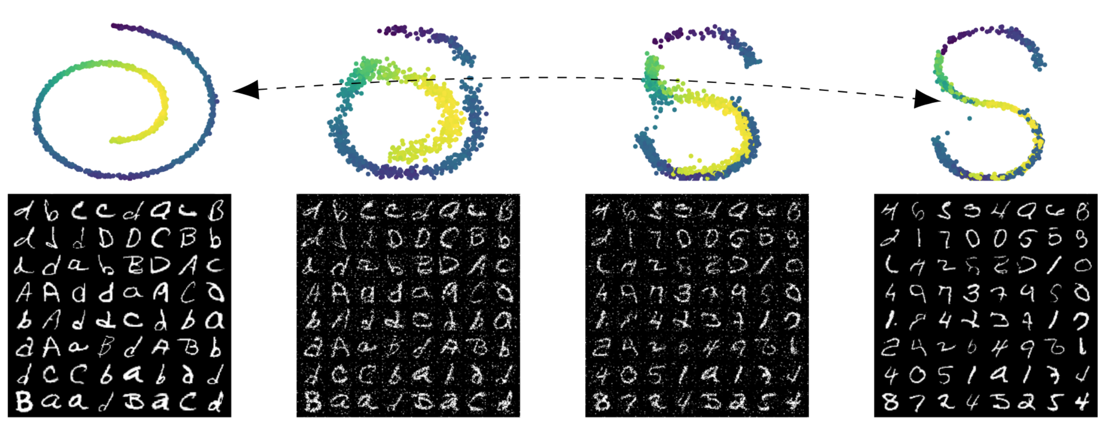

Figure 1: Diffusion Schrödinger bridges on synthetic and real data.

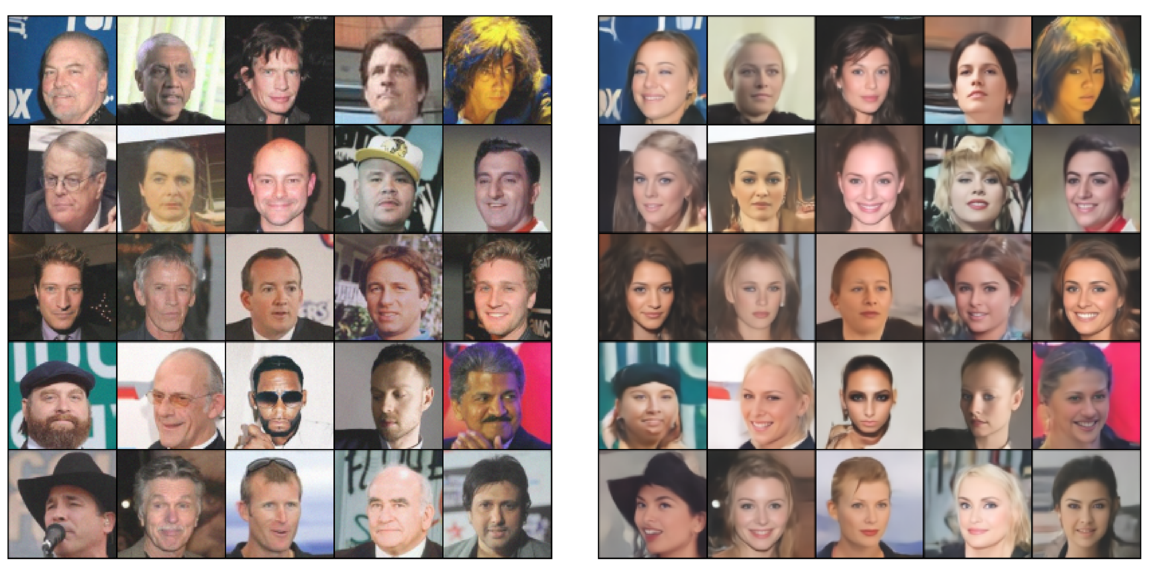

Figure 2: Diffusion Schrödinger bridges on CelebA.

Large deviations

The choice of the Brownian motion reference measure makes the problem equivalent to *entropic optimal transport*, and as $\varepsilon \to 0$, the dynamic Schrödinger bridge converges to the dynamic optimal transport solution.
We are interested in characterizing the speed of this convergence using large deviations theory. I.e. we want a *rate function* $I : C([0,1] : \mathbb{R}^d) \to \mathbb{R}$ such that

$$\begin{aligned} \liminf_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{P}^\varepsilon(F) &\geq -\inf_{\varphi\in F}(\varphi) \quad \text{for all closed } F \subseteq C([0,1]:\mathbb{R}^d), \\ \limsup_{\varepsilon\downarrow 0} \varepsilon \log \mathbb{P}^\varepsilon(G) &\leq -\inf_{\varphi\in G}(\varphi) \quad \text{for all open } G \subseteq C([0,1]:\mathbb{R}^d). \end{aligned} \tag{4}$$

Results

Theoretical

Theorem 0.1 (Bernton, Ghosal, and Nutz, 2022)**.** *The static Schrödinger bridges* $\{\mathbb{P}^\varepsilon_{01}\}_{\varepsilon\in(0,1)}$ *satisfies a large deviation principle, with the rate function*

$$I_{\mathrm{EOT}}(x,y) = \frac{1}{2}|y-x|^2 - \psi^c(y) - \psi(x), \tag{5}$$

where ψ *is a Kantorovich potential of the optimal transport problem, and* ψ^c *its c-transform.* $c(x,y) = \frac{1}{2}|y-x|^2$.

We also know that Brownian bridges satisfy a large deviation principle. This is classical.

Theorem 0.2. *The* ε*-Brownian bridge* $\{B^{\varepsilon,xy}\}$ *satisfies an LDP, as* $\varepsilon \downarrow 0$ *with the rate function*

$$I^{xy}(\varphi) = \frac{1}{2}\int_0^1 |\varphi'(t)|^2 dt - \frac{1}{2}|y-x|^2. \quad \text{(For } \varphi : (\varphi(0),\varphi(1)) = (x,y)\text{, otherwise } \infty\text{).} \tag{6}$$

Can we combine the two results?

Definition 0.3. The LDP (stated as a Laplace principle) is said to be **uniform on compacts** if for any bounded continuous $F : C([0,1] : \mathbb{R}^d) \to \mathbb{R}$, we have

$$\lim_{\varepsilon\downarrow 0} -\varepsilon \log E[\exp -\frac{1}{\varepsilon} F(B^{\varepsilon,(x_\varepsilon,y_\varepsilon)})] = \inf_{\varphi\in C([0,1]:\mathbb{R}^d)} F(\varphi) + I(\varphi), \tag{7}$$

whenever $(x_\varepsilon, y_\varepsilon) \to (x,y)$.
(actually a condition for it).

This has been shown by Kato 2024 and again by us. Restating this result in the variational framework is our main contirubtion so far. This bring the result closer to a more general setting where we consider general SDE.

Theorem 0.4 (Kato, 2024)**.** *The dynamic Schrödinger bridges* $\{\mathbb{P}^\varepsilon\}_{\varepsilon\in(0,1)}$ *satisfies a large deviation principle, with the rate function*

$$I(\varphi) = \frac{1}{2}\int_0^1 |\varphi'(t)|^2 dt - \psi^c(y) - \psi(x), \tag{8}$$

where ψ *is a Kantorovich potential of the optimal transport problem, and* ψ^c *its c-transform.* $c(x,y) = \frac{1}{2}|y-x|^2$.

Experimental

Nothing yet.