4th Symposium on Advances in Approximate Bayesian Inference, 2021 1–5

Shooting Schrödinger's Cat

Anonymous Authors
Anonymous Institution

Abstract

We present a variational inference scheme to learn a model that solves the Schrödinger Bridge Problem (SBP). In contrast to previous work, our approach is solver-agnostic and guarantees solutions that respect the prior beyond the first fitting iteration. Having this solution allows us to generate new samples from one of the distributions by first sampling from the other one and then solving the dynamical system. We show that our model is able to learn the transformation between the Gaussian distribution and arbitrary data, as well as learning dynamics that follow a potential function.

1. Introduction and motivation

The Schrödinger bridge problem (SBP) (Schrödinger, 1931, 1933) seeks to find a transformation between two probability distributions. It has both a static version, where a direct transformation is obtained, and a dynamic version, where the flows between the two distributions are learned. Recent works (De Bortoli et al., 2021; Vargas et al., 2021; Chen et al., 2021) have tackled this dynamic scenario.

Finding these flows can be done using the Iterative Proportional Fitting Procedure (IPFP) (Fortet, 1940; Kullback, 1968; Cramer, 2000), that iteratively optimizes each process so that the Kullback-Leibler divergence (D_{KL}) between the two processes is minimized at each iteration. One way to do this is to update the joint density of the problem using the potentials of the dual representation of the SBP. Approximating these coupled potentials requires estimating an integral involving them, which in higher dimensional problems can be difficult.

The thorough approach of De Bortoli et al. (2021) uses another representation that can more easily be applied where samples of both boundary distributions are available. This tackles the Schrödinger bridge as an iterative mean-matching procedure, where at each iteration, the mean of the reversed forwards process is matched with the mean of the backward process. This requires assuming that at each time-step, the change in the state is small, so that the Euler-Maruyama discretization of the process is valid. Using a small enough time-step ensures this. However, if the time-step is not small enough, it can lead to stability issues. Additionally, while this model is regularized using a prior diffusion process, this regularization is only directly applied in the first iteration. This can lead to the learned dynamics differing wildly from the prior, especially when they are represented by unbounded functions (such as neural networks).

Our model seeks to address these issues by learning the dynamics of the two processes in a way that is independent of the solver used, and so better solvers become available, while also maintaining the regularization in every iteration.

© A. Authors.

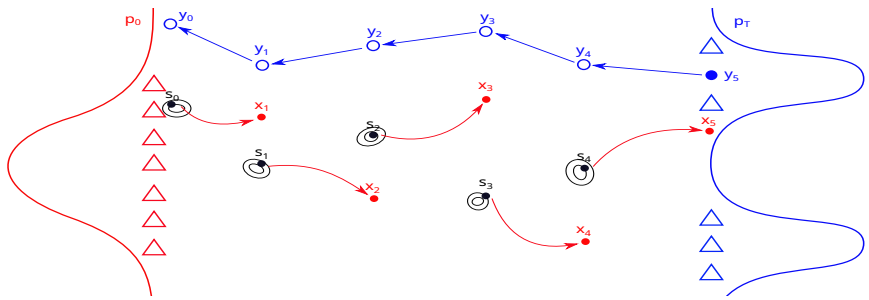


Figure 1: *Diagram of the model, when optimizing the forward process. Firstly, a solution to the backward process is obtained using the solver (blue). Then* $\{s_i\}_{i=0}^{N-1}$ *are sampled from the variational distribution (black) which are then solved forwards one step using the forward process to obtain* $\{x_i\}_{i=1}^{N}$ *(red). Finally, the parameters of the variational distribution and the forward process are optimized so that both* s_i *and* x_i *match their corresponding* y_i*.*

2. Solving the SBP with shooting variables

Writing the two boundary distributions as p_0 and p_T, we define the process that transforms the first into the second as the forward process

$$\mathrm{d}\boldsymbol{y}_t = \boldsymbol{f}^k(\boldsymbol{y}_t)\mathrm{d}t + \gamma_t \mathrm{d}\boldsymbol{B}_t\,, \tag{1}$$

where k is the IPFP iteration, $\boldsymbol{f}^k : \mathbb{R}^d \to \mathbb{R}^d$ is the drift function, $\boldsymbol{B}_t$ is the Brownian process and γ_t is the diffusion scale. The backward process then transforms p_T into p_0, with drift function $\boldsymbol{g}^k : \mathbb{R}^d \to \mathbb{R}^d$ and same diffusion scales for the same values of t.

We consider the time interval $[0, T]$ divided into N sub-intervals Δt_i, for $i \in \{0, \ldots, N-1\}$. For the following derivation, we assume that we have samples $\{\boldsymbol{y}_i\}_{i=0}^{N}$, obtained by solving the backward SDE with an arbitrary solver and then reversing them time-wise (the derivation is equivalent if we have samples from the forward SDE). The probability density of these reversed solutions is then

$$p(\{\boldsymbol{y}_i\}_{i=0}^{N}) = p(\boldsymbol{y}_0)\prod_{i=1}^{N} p(\boldsymbol{y}_i|\boldsymbol{y}_{i-1})\,. \tag{2}$$

Taking inspiration from Hegde et al. (2021), we now introduce the shooting variables $\{\boldsymbol{s}_i\}_{i=0}^{N-1}$ and $\{\boldsymbol{x}_i\}_{i=1}^{N}$with transition probabilities $p(\boldsymbol{x}_i|\boldsymbol{s}_{i-1})$. The probability density in Equation (2) then becomes

$$p(\{\boldsymbol{y}_i\}_{i=0}^{N}) = p(\boldsymbol{y}_0)\int\prod_{i=1}^{N} p(\boldsymbol{y}_i|\boldsymbol{x}_i)p(\boldsymbol{x}_i|\boldsymbol{s}_{i-1})p(\boldsymbol{s}_{i-1}|\boldsymbol{y}_{i-1})\mathrm{d}\boldsymbol{x}_i\mathrm{d}\boldsymbol{s}_{i-1}\,, \tag{3}$$

where we have assumed that $p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{y}_{i-1}) \approx p(\boldsymbol{y}_i|\boldsymbol{x}_i)$.

We now introduce the variational distribution $q(\boldsymbol{s}_{i-1}|y_{i-1})$ and the "variational" transition density $q(\boldsymbol{x}_i|\boldsymbol{s}_{i-1})$, which we define to be the forward process, to obtain the following

lower bound on Equation (3)

$$\mathcal{L} = \sum_{i=1}^{N} p(\boldsymbol{y}_0) \Bigg\{ \int q(\boldsymbol{x}_i|\boldsymbol{s}_{i-1}) q(\boldsymbol{s}_{i-1}|\boldsymbol{y}_{i-1}) \log p(\boldsymbol{y}_i|\boldsymbol{x}_i) \mathrm{d}\boldsymbol{x}_i \mathrm{d}\boldsymbol{s}_{i-1} \\ - \int q(\boldsymbol{s}_{i-1}|\boldsymbol{y}_{i-1}) D_{KL}[q(\boldsymbol{x}_i|\boldsymbol{s}_{i-1}) \| p(\boldsymbol{x}_i|\boldsymbol{s}_{i-1})] \mathrm{d}\boldsymbol{s}_{i-1} - D_{KL}[q(\boldsymbol{s}_{i-1}|\boldsymbol{y}_{i-1}) \| p(\boldsymbol{s}_{i-1}|\boldsymbol{y}_{i-1})] \Bigg\}. \tag{4}$$

The diagram of the model is shown in Figure 1.

The first and last terms can be easily computed once the distributions $p(\boldsymbol{y}_i|\boldsymbol{s}_i)$ and $p(\boldsymbol{s}_{i-1}|\boldsymbol{y}_{i-1})$ are defined. The second term, though, requires us to know the transition probabilities of the forward process, which for more elaborate solvers might be very difficult (or even impossible) to obtain in closed form. To circumvent this, we replace the Kullback-Leibler divergence in this second term by the kernelized Stein discrepancy (KSD, Liu et al. (2016); Chwialkowski et al. (2016)):

$$\mathbb{S}(p, q) = \mathbb{E}_{x,x' \sim p}[\delta_{q,p}(x)^\top k(x, x') \delta_{q,p}(x')] \tag{5}$$

where $\delta_{q,p}(x) = \nabla \log q(x) - \nabla \log p(x)$, (with q and p arbitrary densities) and $k(x, x')$ is the (twice differentiable) strictly positive kernel. It can be shown that there is an estimator for the KSD that only requires the score function of one of the densities, which we can calculate for the prior transition density of the shooting variables, since we define it as the transition density of the reference process used in IPFP.

3. Experiments

In all of these experiments, the SDEs in our model were solved using the weak solver proposed in Rößler (2009).

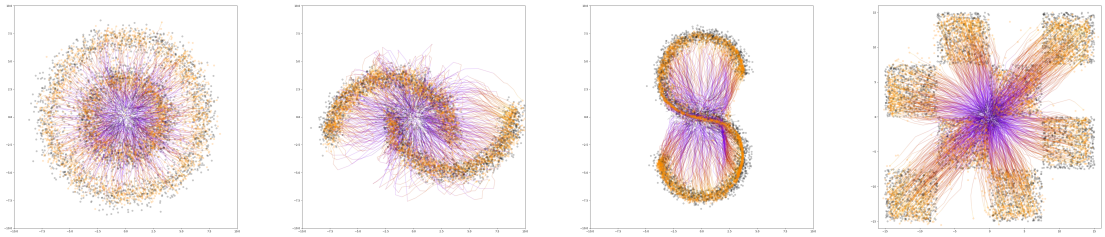
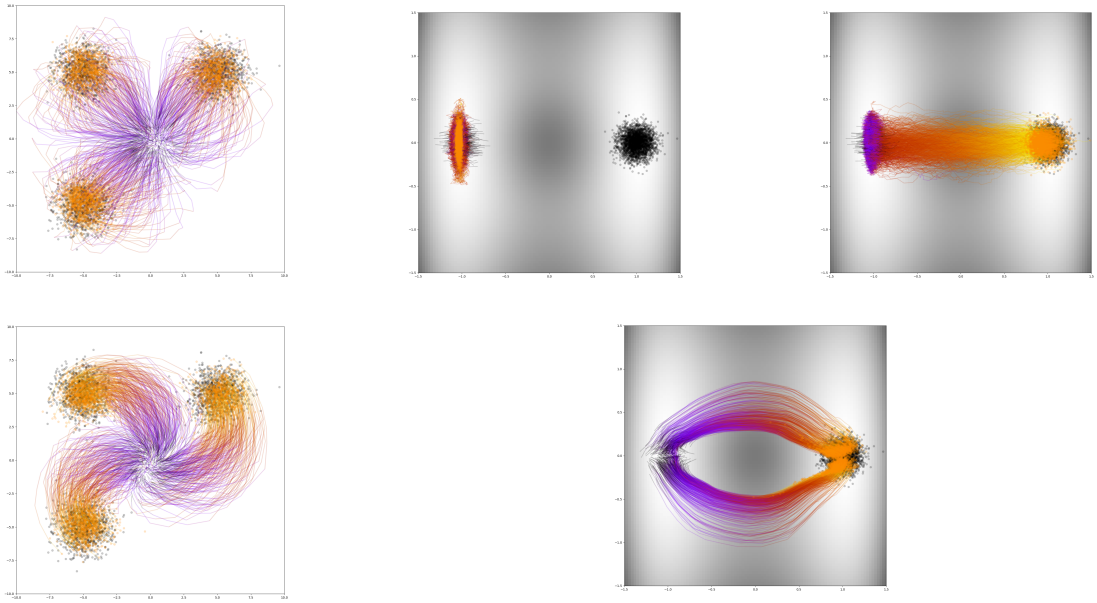


Figure 2: *Learned dynamics, from Gaussian distribution to data set, with Brownian prior. Data points are shown in black and final generated points in orange.*

Brownian prior Firstly, we test whether our model is able to transform a Gaussian distribution into several synthetic data sets and where the prior process is simple Brownian motion. The results are presented in Figure 2.

Whirlpool prior We then verify that the model is able to learned dynamics that are different from a Brownian motion. The first test consists of a "whirlpool" prior on the



(a) Whirlpool prior

(b) Double well prior

Figure 3: *Dynamics under priors. (a) The standard Brownian motion (top) modified by the whirlpool prior (bottom). (b) Double well prior dynamics applied (top-left); learned forwards dynamics with DSB (top-right) and learned forwards dynamics with our model (bottom).*

dynamics: the drift at any particular point is perpendicular to it. So, for a 2D vector $\{x, y\}$ the drift would be $\boldsymbol{f}(x, y) = \{-\frac{7}{2}x, \frac{7}{2}y\}$. The results are shown in Figure 3 (a).

Double well prior We also test our model with a double well potential prior, where the drift follows the negative gradient of the potential

$$U(x, y) = \frac{5}{2}(x^2 - 1)^2 + y^2 + 3\exp\left(-3(x^2 + y^2)\right). \tag{6}$$

We compare our model with De Bortoli et al. (2021) (DSB) in Figure 3(b). While our model is able to avoid the hill, DSB does not. We believe this is because in the first iteration DSB prefers to match the prior paths (see top-left plot of Figure 3(b)) at any cost, and the quickest way to do so is to go over the hill. Since in the following IPFP iterations, DSB only tries to match the learned paths of the other direction, it will never avoid the hill.

4. Conclusion

We have presented a model that is able to obtain solutions to the dynamic Schrödinger Bridge Problem that is both solver-agnostic and maintains the regularization throughout the training iterations, not just the first one. We have shown that it is able to learn transformations from a Gaussian distribution to arbitrary data distributions, both using a simple Brownian prior and a elaborate priors. Finally, we have shown that our model can also be applied to potential priors.

References

Tianrong Chen, Guan-Horng Liu, and Evangelos A. Theodorou. Likelihood training of schrödinger bridge using forward-backward sdes theory, 2021.

Kacper Chwialkowski, Heiko Strathmann, and Arthur Gretton. A kernel test of goodness of fit. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2606–2615, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/chwialkowski16.html.

Erhard Cramer. Probability measures with given marginals and conditionals: I-projections and conditional iterative proportional fitting. *Statistics & Decisions*, 18(3):311–330, 01 2000. doi: 10.1524/strm.2000.18.3.311.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. *arXiv preprint arXiv:2106.01357*, 2021.

R. Fortet. Résolution d'un système d'équations de m. schrödinger. *Journal de Mathématiques Pures et Appliqués*, 1:82–105, 1940.

Pashupati Hegde, Çağatay Yıldız, Harri Lähdesmäki, Samuel Kaski, and Markus Heinonen. Bayesian inference of odes with gaussian processes. *arXiv preprint arXiv:2106.10905*, 2021.

S. Kullback. Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39(4):1236–1243, 1968.

Qiang Liu, Jason Lee, and Michael Jordan. A kernelized stein discrepancy for goodness-of-fit tests. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 276–284, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/liub16.html.

Andreas Rößler. Second order runge–kutta methods for itô stochastic differential equations. *SIAM Journal on Numerical Analysis*, 47(3):1713–1738, 2009. doi: 10.1137/060673308. URL https://doi.org/10.1137/060673308.

E. Schrödinger. Uber die umkehrung der naturgesetze. *Sitzungsberichte der Preuss Akad. Wissen, Berlin, Phys. Math. Klasse*, pages 144–153, 1931.

E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Annales de l'Institut Henri Poincaré*, 2(3):269–310, 1933.

Francisco Vargas, Pierre Thodoroff, Austen Lamacraft, and Neil Lawrence. Solving Schrödinger bridges via maximum likelihood. *Entropy*, 23(9), 2021. ISSN 1099-4300. doi: 10.3390/e23091134. URL https://www.mdpi.com/1099-4300/23/9/1134.