Light and Optimal Schrödinger Bridge Matching

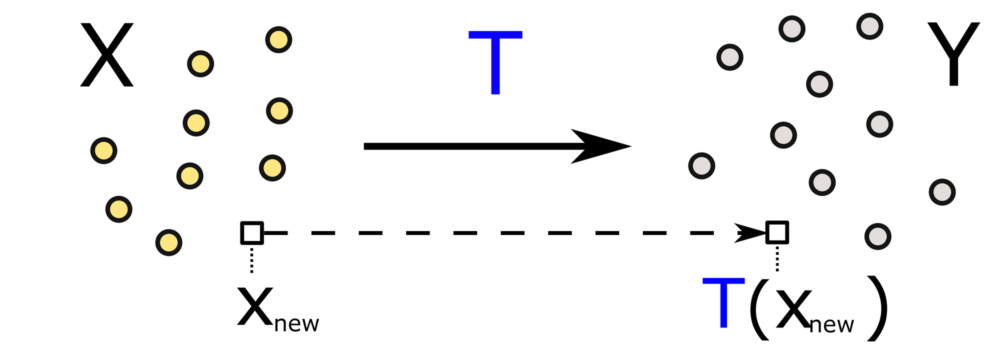
Skoltech[1] Skolkovo Institute of Science and Technology

Nikita Gushchin[*,1] Sergei Kholkin[*,1] Evgeny Burnaev[1,2] Alexander Korotin[1,2]

(* - equal contribution)

AIRI[2]

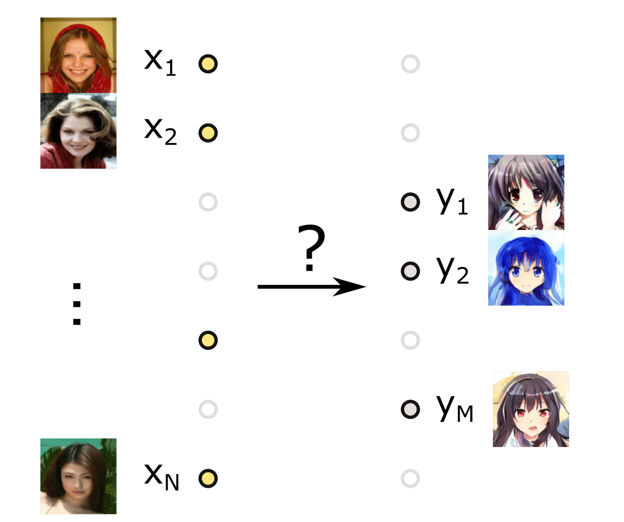
I Motivation: *Unpaired* Domain Translation



The (informal) task: given samples X, Y from two domains, construct a map T which can translate new samples from the input domain to the target domain.

Unpaired setup

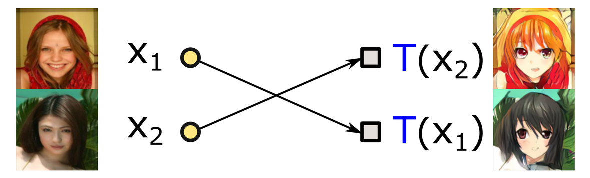
No paired training examples are available.



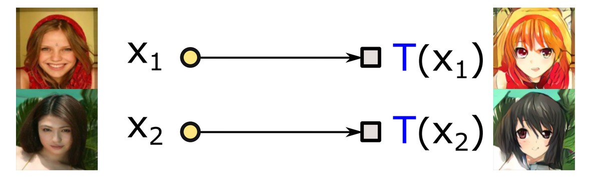
Main problem

Ambiguity in translations

Bad solution (changes the content)



Good solution (keeps the content)



II Optimal Schrödinger Bridge (SB) Matching

Given two probability distributions p_0, p_1, how to transform p_0 to p_1 via a diffusion process and preserve the input-output similarity?

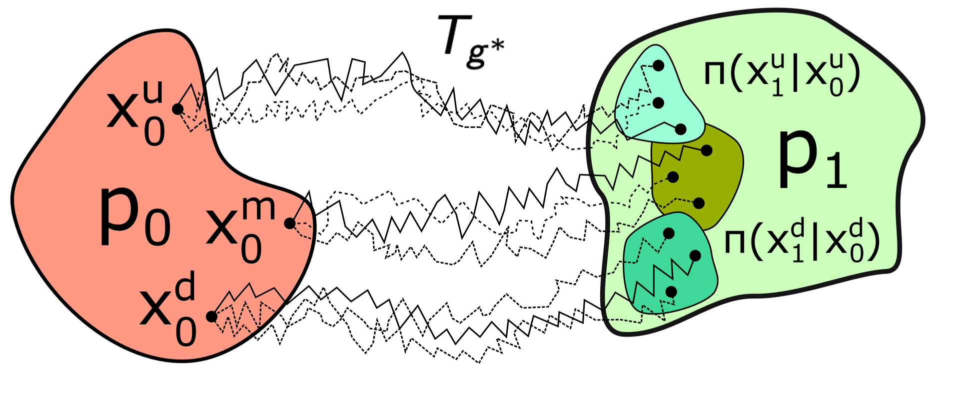
The SB problem.

The problem is about finding a diffusion T_g which transforms a given distribution p_0 to another distribution p_1 and has minimal "kinetic energy."

For two continuous distributions p_0 and p_1 on $\mathbb{R}^D$, the SB problem is:

$$\inf_{T_g \in \mathcal{D}(p_0,p_1)} \frac{1}{2\epsilon} \underbrace{\mathbb{E}_{T_g}\Big[\int_0^1 \|g(x_t,t)\|^2 dt\Big]}_{\text{"average kinetic energy"}}, \quad T_g : dx_t = \underbrace{g(x_t,t)}_{\text{"velocity"}} dt + \sqrt{\epsilon} dW_t,$$

where $\mathcal{D}(p_0,p_1)$ is the set of diffusions with marginals p_0, p_1 at $t=0, t=1$.



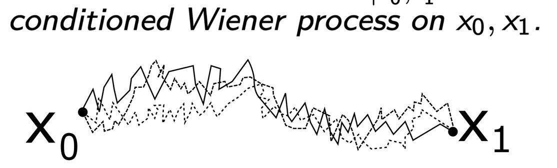
The minimizer T_{g^*} is called the Schrödinger Bridge.

Bridge Matching constructs a diffusion between p_0 and p_1 by combining reciprocal and Markovian projections of stochastic processes.

Reciprocal Projection.

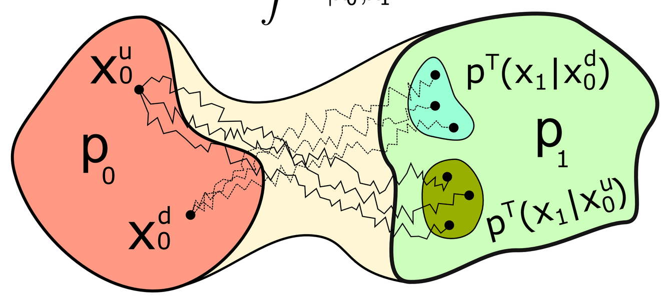
Makes a mixture of Brownian bridges $W^{\epsilon}_{|x_0,x_1}$ with the distribution $p^T(x_0,x_1)$ of stochastic process T at times $t=0$ and $t=1$.

The Brownian Bridge $W^{\epsilon}_{|x_0,x_1}$ is a conditioned Wiener process on x_0, x_1.



$\mathcal{N}(x_t|(1-t)x_0 + tx_1, t(1-t)\epsilon I_D)$

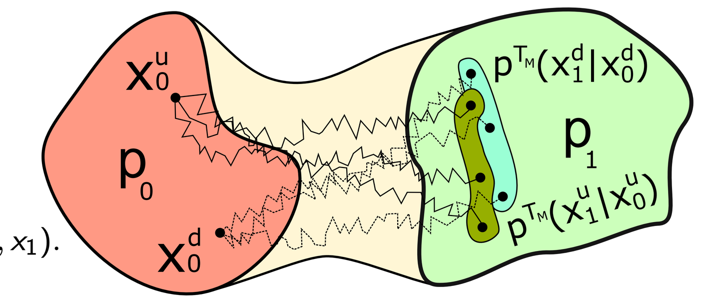
$$\text{proj}_{\mathcal{R}}(T) = \int W^{\epsilon}_{|x_0,x_1} dp^T(x_0,x_1)$$



Markovian Projection.

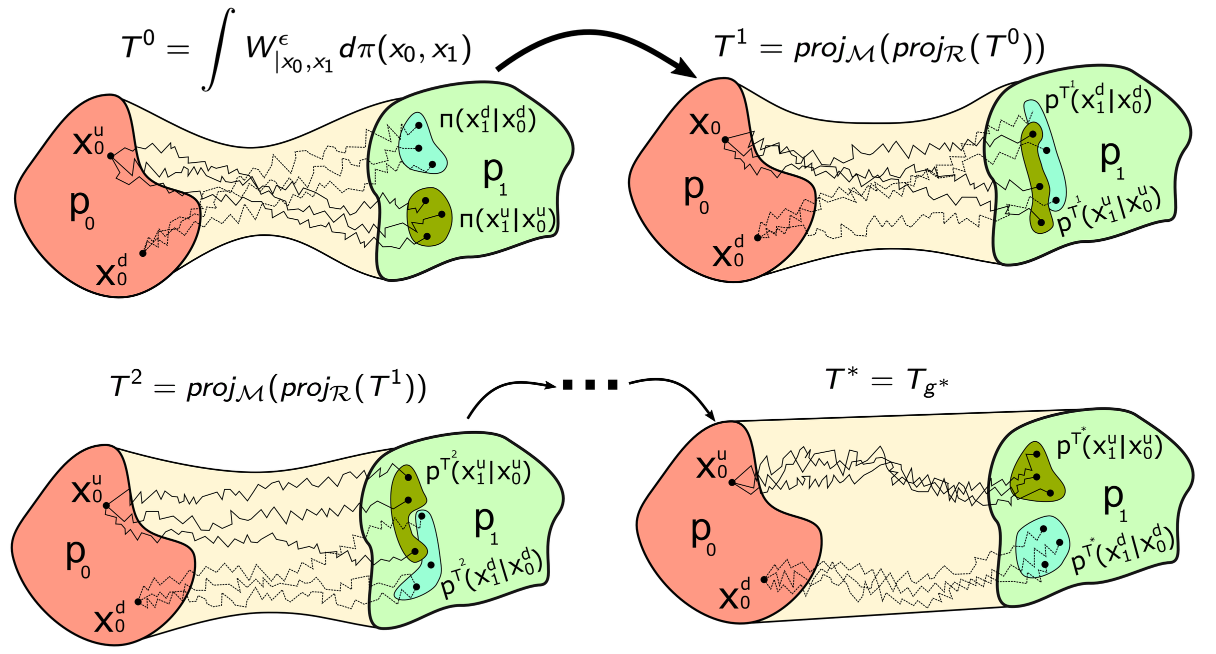
Finds the diffusion $T_{\mathcal{M}}$ which is the most similar to a process T.

$$T_{\mathcal{M}} = \text{proj}_{\mathcal{M}}(T): \quad dx_t = g_{\mathcal{M}}(x_t,t)dt + \sqrt{\epsilon}dW_t,$$
$$g_{\mathcal{M}}(x_t,t) = \underset{g}{\text{argmin}} \int \|g(x_t,t) - \frac{x_1 - x_t}{1-t}\|^2 dp^T(x_t,x_1).$$



Iterative Markovian Fitting (IMF algorithm).

Consider a sequence of alternating Markovian and Reciprocal projections starting from some process T^0, i.e., $T^n = proj_{\mathcal{M}}(proj_{\mathcal{R}}(T^{n-1}))$. This sequence T^n converges to the Schrödinger Bridge $T^* = T_{g^*}$.



Issue: Many iterations of the Bridge Matching.

Optimal Schrödinger Bridge Matching (ours)

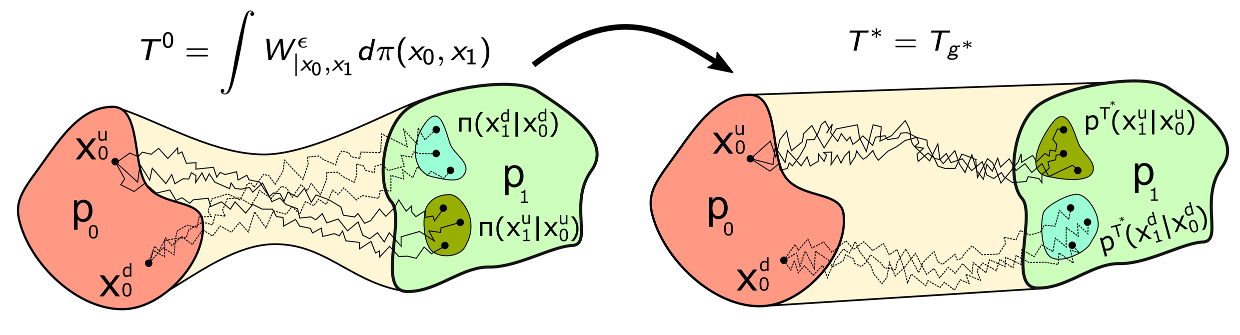
Optimal Projection T_{OP} of a process $T_\pi = \int W^{\epsilon}_{|x_0,x_1} d\pi(x_0,x_1)$ is given by:

$$T_{OP}(T_\pi) : dx_t = g^*(x_t,t)dt + \sqrt{\epsilon}dW_t,$$

$$g^* = \underset{v:\mathbb{R}^d \to \mathbb{R}}{\text{argmin}} \int \|g_v(x_t,t) - \frac{x_1 - x_t}{1-t}\|^2 dp^{T_\pi}(x_t,x_1),$$

where $g_v(x_t,t) = \nabla_{x_t} \log \int \mathcal{N}(x'|x_t,(1-t)\epsilon I_D) \exp(\frac{\|x'\|^2}{2\epsilon}) v(x') dx'$.

For **any** $\pi(x_0,x_1)$ with marginals p_0 and p_1, **our** Optimal Projection of T_π provably recovers the Schrödinger Bridge $T^* = T_{g^*}$.



Issue solved: One iteration of Bridge Matching.

III Proposed Algorithm: Light Schrödinger Bridge Matching (LightSB-M)

The algorithm is based on:

- Our "Optimal Projection" that translates **any** π with marginals p_0 and p_1 to SB
- Our novel Bridge Matching-like optimization objective

$$L_\theta(\pi) = \int_0^1 \int_{\mathbb{R}^D \times \mathbb{R}^D} \|g_\theta(x_t,t) - \frac{x_1 - x_t}{1-t}\|^2 dp_{T_\pi}(x_t,x_1)dt$$

$$g_\theta(x_t,t) = \epsilon \nabla_{x_t} \log \int_{\mathbb{R}^D} \mathcal{N}(x'|x_t,(1-t)\epsilon I_D) \exp(\frac{\|x'\|^2}{2\epsilon}) v_\theta(x') dx'$$

- Parameterization of the SB using mixtures of Gaussians $v_\theta(x) = \sum_{k=1}^K \alpha_k \mathcal{N}(x|r_k, S_k)$. In this case, g_θ admits a closed-form expression (see the paper).

Algorithm 1: Light SB Matching (LightSB-M)

repeat

- Sample a batch of pairs $\{x_0^n, x_1^n\}_{n=0}^N \sim \pi$;
- Sample a batch $\{t_n\}_{n=0}^N \sim U[0,1]$;
- Sample a batch $\{x_t^n\}_{n=0}^N \sim W^{\epsilon}_{|x_0,x_1}$;
- $\mathcal{L}_\theta \leftarrow \frac{1}{N}\sum_{n=1}^N ||g_\theta(x_t^n,t_n) - \frac{1}{1-t_n}(x_1^n - x_t^n)||^2$;
- Update θ using $\frac{\partial \mathcal{L}_\theta}{\partial \theta}$;

until *converged*;

Advantages:

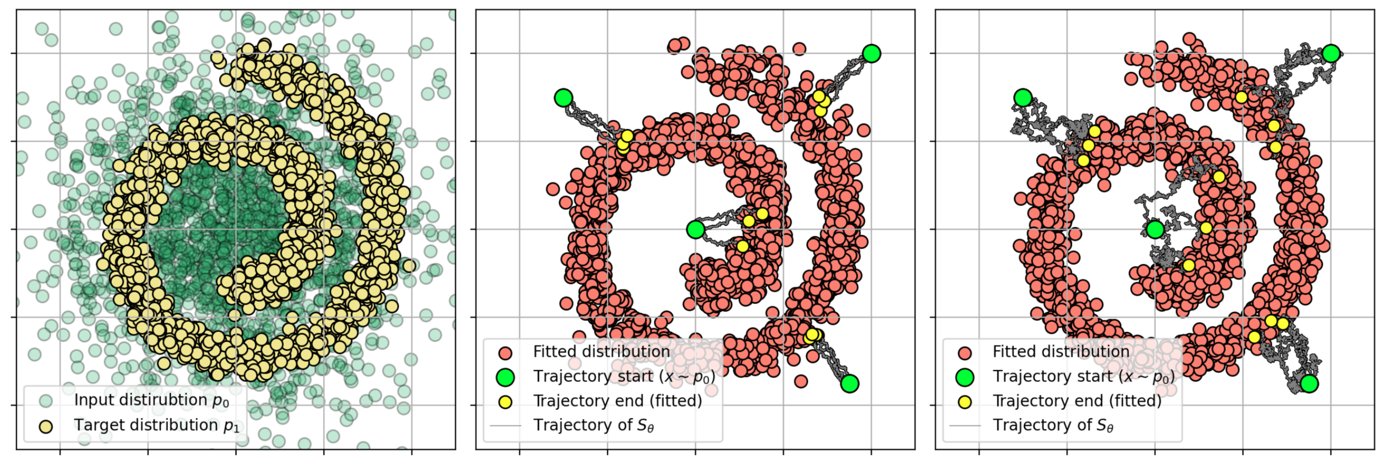
Fast training (< 1 minute on 4 CPU cores, not hours of training on GPU, like others).

No more iterative Bridge Matching Optimal projection guarantee SB solution after one Optimal Bridge Matching step

IV Toy examples

Qualitative results of our solver applied to 2D model distributions ("Gaussian" to "swiss-roll").

The volatility of trajectories increases with ϵ, and the distributions $\pi(x_1|x_0)$ become more disperse.



(a) $x \sim p_0, y \sim p_1$. (b) $\epsilon = 0.01$. (c) $\epsilon = 0.1$.

V Schrödinger Bridge Benchmark

Quantitative results of our solver on the standard benchmark for the Schrödinger bridge problem.

	$\epsilon=0.1$				$\epsilon=1$				$\epsilon=10$			
	$D=2$	$D=16$	$D=64$	$D=128$	$D=2$	$D=16$	$D=64$	$D=128$	$D=2$	$D=16$	$D=64$	$D=128$
DSBM	5.2	16.8	37.3	35	0.3	1.1	9.7	31	3.7	105	3557	15000
SF^2M-Sink	0.54	3.7	9.5	10.9	0.2	1.1	9	23	0.31	4.9	319	819
LightSB-M (ID)	0.04	0.18	0.77	1.66	**0.09**	**0.18**	0.47	**1.2**	**0.12**	0.19	**0.36**	**0.71**
LightSB-M (MB)	**0.02**	**0.1**	**0.56**	**1.32**	**0.09**	**0.18**	**0.46**	**1.2**	0.13	**0.18**	**0.36**	**0.71**

* The cBW-UVP metric is used to compare the built Schrödinger Bridge with the ground truth bridge (lower=better). LighSB-M is trained with transport plans π which are: independent (ID) and mini batch OT (MB).

VI Single cell experiment

Quantitative results in the problem of predicting single-cell trajectories in the feature space (single-cell trajectory inference).

Solver \ DIM	50	100	1000
DSBM (Shi et al., 2023) [1 GPU V100]	2.46 ± 0.1 (6.6 m)	2.35 ± 0.1 (6.6 m)	1.36 ± 0.04 (8.9 m)
SF^2M-Sink (Tong et al., 2023) [1 GPU V100]	2.66 ± 0.18 (8.4 m)	2.52 ± 0.17 (8.4 m)	1.38 ± 0.05 (13.8 m)
LightSB-M (ID) [4 CPU cores]	2.347 ± 0.11 (58 s)	2.174 ± 0.08 (60 s)	1.35 ± 0.05 (147 s)
LightSB-M (MB) [4 CPU cores]	$\mathbf{2.33} \pm 0.09$ (80 s)	$\mathbf{2.172} \pm 0.08$ (80 s)	$\mathbf{1.33} \pm 0.05$ (176 s)

** The Energy distance metric is used to compare the predicted cell position and the observed one (smaller=better). The operating time of the method in question is indicated in parentheses. 50, 100, 1000 - dimension of the feature space. LighSB-M is trained with transport plans π which are: independent (ID) and mini batch OT (MB).

VII Unpaired Image-to-image Translation

Qualitative results of our solver for solving the domain translation problem (in the latent space of the ALAE autoencoder).

Images resolution is 1024x1024.

(a) Man → Woman (b) Woman → Man (c) Adult → Child (d) Child → Adult