Light Schrödinger Bridge

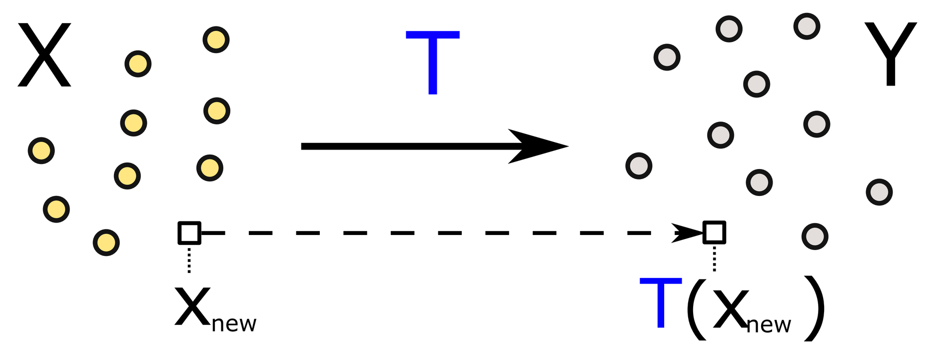
Skoltech[1] Skolkovo Institute of Science and Technology

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

(* - equal contribution)

AIRI[2]

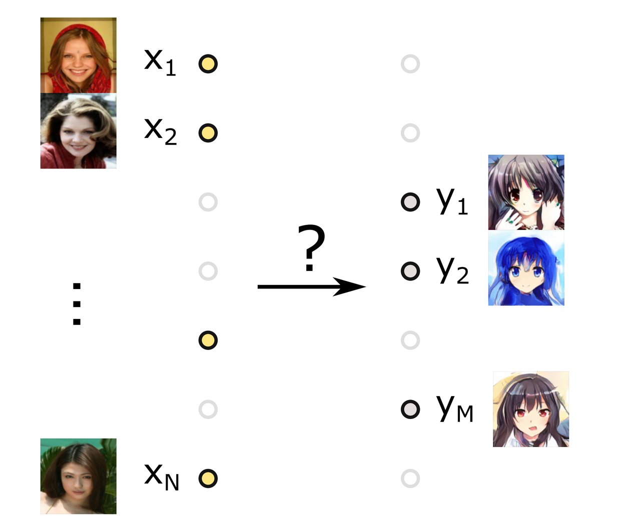
I Motivation: *Unpaired* Domain Translation



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

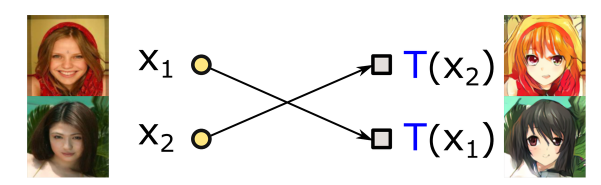
No paired training examples are available.



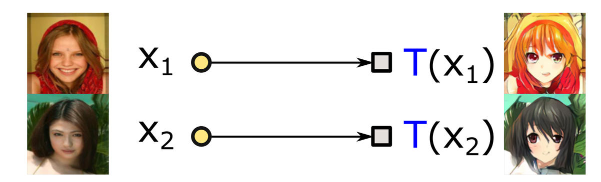
Main problem

Ambiguity in translations

Bad solution (changes the content)

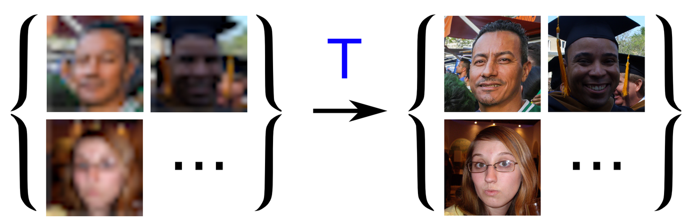


Good solution (keeps the content)

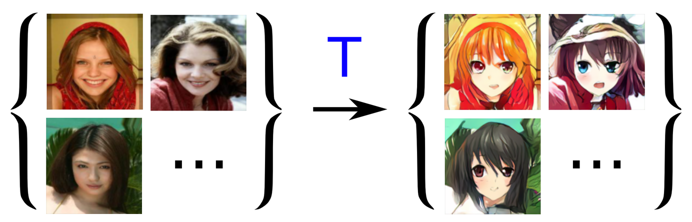


Examples of image unpaired domain translation

Example I: image super-resolution

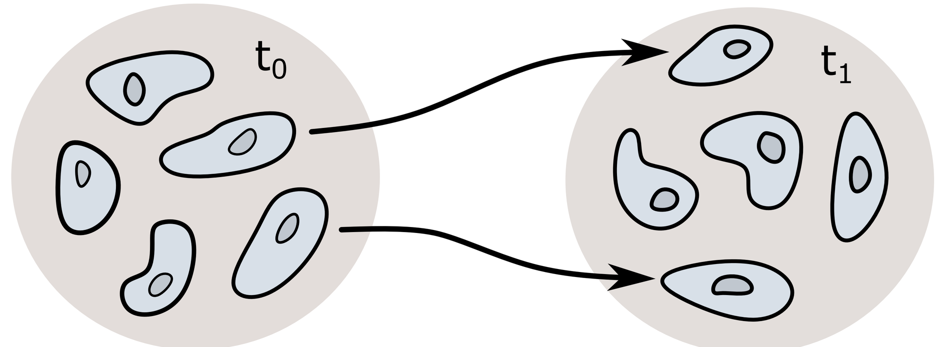


Example II: image style transfer



Single-cell data domain translation

Single-cell (SC) sequencing extracts features for individual cells from a population but destroys them. Therefore, to study individual cell dynamics one needs a method to map cells between different observation times.



II Background: Diffusion Schrödinger Bridge (SB)

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?

1. Formulation of the Schrödinger Bridge Problem:

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

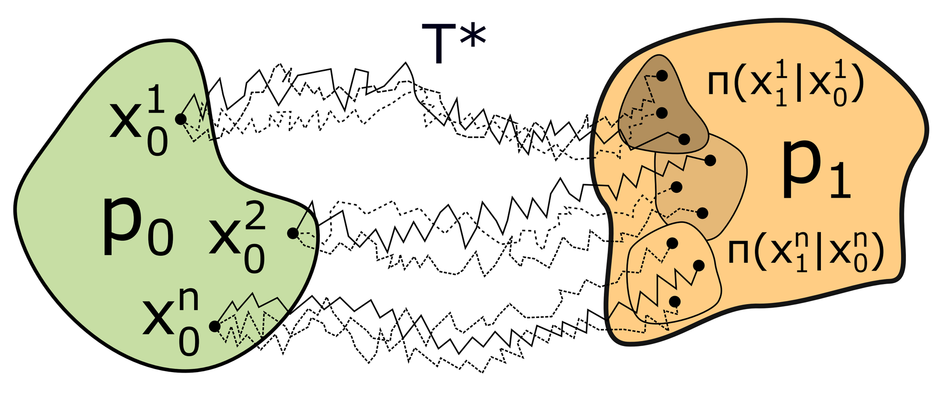
$$\inf_{T\in\mathcal{F}(p_0,p_1)} \mathrm{KL}(T\|W^\epsilon),$$

where $\mathcal{F}(p_0,p_1)$ are stochastic processes with marginals p_0, p_1 at $t=0$ and $t=1$, respectively, while W^ϵ is the Wiener process with variance ϵ, i.e., given by the SDE:

$$W^\epsilon: \quad dX_t = \sqrt{\epsilon}dW_t.$$

This problem has a unique solution, which is a diffusion process T^* described by the SDE:

$$T^*: \quad dX_t = g^*(X_t,t)dt + \sqrt{\epsilon}dW_t.$$



The minimizer T* is called the Schrödinger Bridge.

2. Structure of the Schrödinger Bridge.

The Schrödinger bridge problem can be fully characterized by the initial distribution p_0 and *the Schrödinger potential* $\phi^*(x): \mathbb{R}^D \to \mathbb{R}_+$. The optimal drift can be expressed by the Schrödinger potential as

$$g^*(x,t) = \epsilon\nabla_x \log \int_{\mathbb{R}^D} \mathcal{N}(x'|x,(1-t)\epsilon I_D)\phi^*(x')dx'$$

In our work, we introduce (for convenience) *the adjusted Schrödinger bridge* potential $v^*(x_1) \stackrel{\text{def}}{=} \exp(-\frac{\|x_1\|^2}{2\epsilon})\phi^*(x_1)$.

3. Integration of the Schrödinger Bridge SDE.

The Schrödinger bridge SDE:

$$T^*: \quad dX_t = g^*(X_t,t)dt + \sqrt{\epsilon}dW_t.$$

admits a closed-form solution $\pi^*(x_1|x_0)$ expressed through the adjusted Schrödinger potential v^*:

$$\pi^*(x_1|x_0) \stackrel{\text{def}}{=} \frac{\exp\big(\langle x_0,x_1\rangle/\epsilon\big)v^*(x_1)}{c^*(x_0)},$$

where $c^*(x_0) \stackrel{\text{def}}{=} \int_{\mathbb{R}^D} \exp\big(\langle x_0,x_1\rangle/\epsilon\big)v^*(x_1)dx_1$ is the normalizing constant.

III Proposed Algorithm: Light Schrödinger Bridge (LightSB)

Our solver is based on:

1. Optimal parameterization of the Schrödinger bridge using mixtures of Gaussians:

$$v_\theta(x_1) = \sum_{k=1}^K \alpha_k\mathcal{N}(x_1|r_k,\epsilon S_k), \quad c_\theta(x_0) = \sum_{k=1}^K \alpha_k \exp\Big(\frac{x_0^T S_k x_0 + 2r_k^T x_0}{2\epsilon}\Big).$$

2. New loss function for training the Schrödinger bridge:

$$\mathrm{KL}(T^*\|T_\theta) = \underbrace{\int_{\mathbb{R}^D} \log c_\theta(x_0)p_0(x_0)dx_0 - \int_{\mathbb{R}^D} \log v_\theta(x_1)p_1(x_1)dx_1}_{\mathcal{L}(\theta)} + \text{Const}$$

Algorithm 1: Light Schrödinger Bridge (LightSB)

repeat

Sample batches $\{x_0^1,\dots,x_0^N\}\sim p_0$, $\{x_1^1,\dots,x_1^M\}\sim p_1$;

$\widehat{\mathcal{L}}_\theta \leftarrow \frac{1}{N}\sum_{n=1}^N \log c_\theta(x_0^n) - \frac{1}{M}\sum_{m=1}^M \log v_\theta(x_1^m)$;

Update θ by using $\frac{\partial\widehat{\mathcal{L}}_\theta}{\partial\theta}$;

until *converged*;

Advantages:

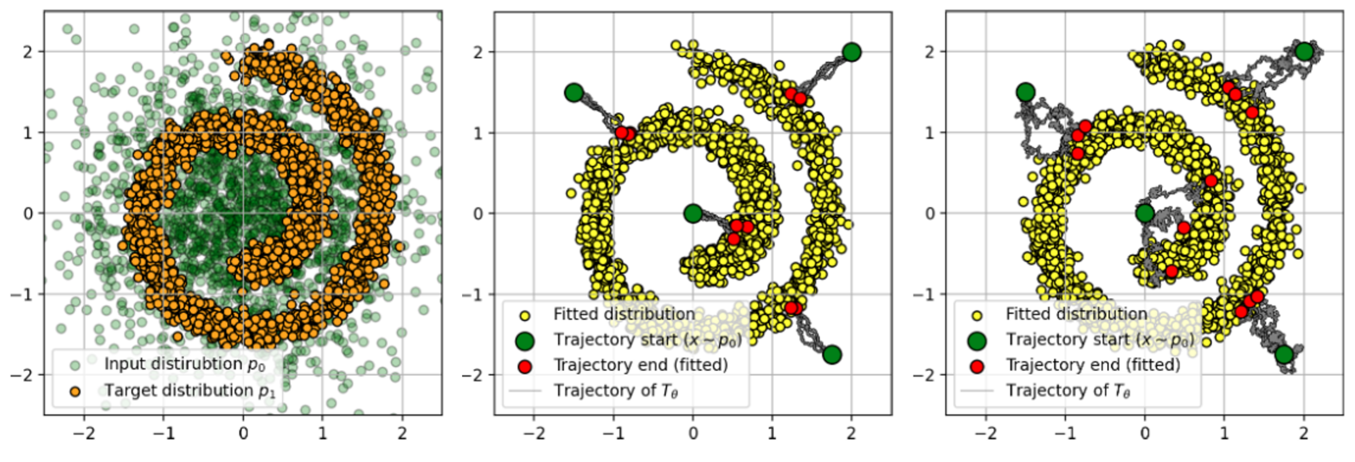
Fast training (< 1 minute on 4 CPU cores, not hours of training on GPU, like others).

Theoretical validity (the guarantees of the method's learning ability from the point of view of statistical learning and approximation theories).

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

V Schrödinger Bridge Benchmark

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

	$\epsilon=0.1$				$\epsilon=1$				$\epsilon=10$			
	$D=2$	$D=16$	$D=64$	$D=128$	$D=2$	$D=16$	$D=64$	$D=128$	$D=2$	$D=16$	$D=64$	$D=128$
Best solver	1.94	13.67	11.74	11.4	1.04	9.08	18.05	15.23	1.40	1.27	2.36	1.31
[LightSB]	**0.03**	**0.08**	**0.28**	**0.60**	**0.05**	**0.09**	**0.24**	**0.62**	**0.07**	**0.11**	**0.21**	**0.37**
$\pm std$	±0.01	±0.04	±0.02	±0.02	±0.003	±0.006	±0.007	±0.007	±0.02	±0.01	±0.01	±0.01

*The metric cBW-UVP is used for comparising build schrödinger bridge with ground-truth bridge (lower=better).

VI Single cell experiment

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

Setup	Solver type	Solver \ DIM	50	100	1000
Discrete EOT	Sinkhorn	(Cuturi, 2013) [1 GPU V100]	2.34 (90 s)	2.24 (2.5 m)	1.864 (9 m)
Continuous EOT	Langevin-based	(Mokrov et al., 2023) [1 GPU V100]	2.39 ± 0.06 (19 m)	2.32 ± 0.15 (19 m)	1.46 ± 0.20 (15 m)
Continuous EOT	Minimax	(Gushchin et al., 2023) [1 GPU V100]	2.44 ± 0.13 (43 m)	2.24 ± 0.13 (45 m)	1.32 ± 0.06 (71 m)
Continuous EOT	IPF	(Vargas et al., 2021) [1 GPU V100]	3.14 ± 0.27 (8 m)	2.86 ± 0.26 (8 m)	2.05 ± 0.19 (11 m)
Continuous EOT	KL minimization	LightSB (**ours**) [4 CPU cores]	2.31 ± 0.27 (65 s)	2.16 ± 0.26 (66 s)	1.27 ± 0.19 (146 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.

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) Male → Female. (b) Female → Male. (c) Adult → Child. (d) Child → Adult.