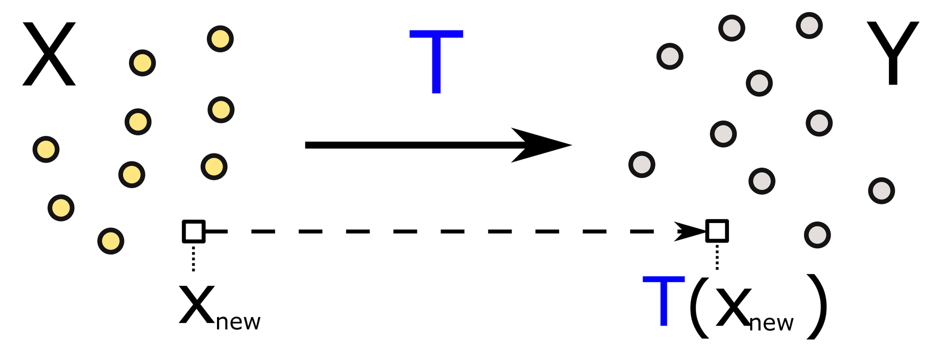
Adversarial Schrödinger Bridge Matching

Nikita Gushchin[*,1] Daniil Selikhanovych[*,1] Sergei Kholkin[*,1]
Evgeny Burnaev[1,2] Alexander Korotin[1,2]
(* - equal contribution)
[1] Skoltech, Skolkovo Institute of Science and Technology
[2] AIRI

NEURAL INFORMATION PROCESSING SYSTEMS

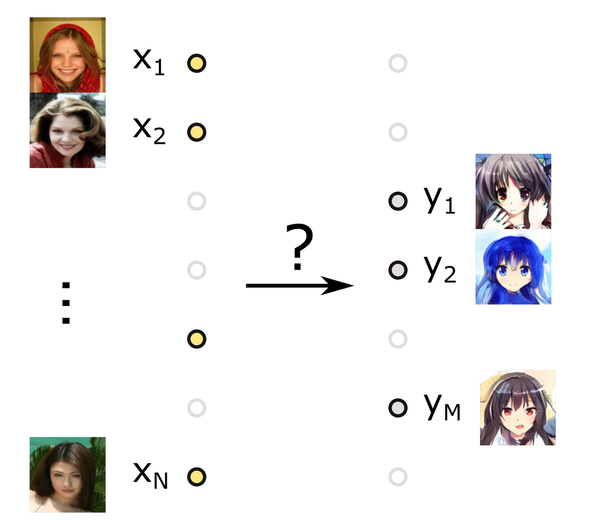
I Motivation: Unpaired Domain Translation



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

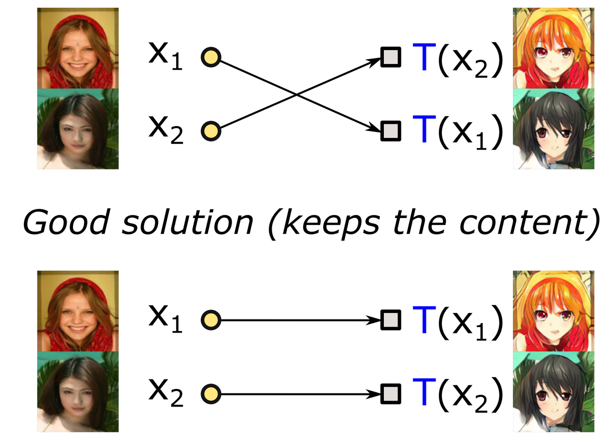
No paired training examples are available.



Main problem

Ambiguity in translations

Bad solution (changes the content)



Good solution (keeps the content)

II Background: Schrödinger Bridges

Static SB problem setup. For two continuous distributions p_0 and p_1 on $\mathbb{R}^D$ the Static Scrödinger Bridge Problem is:

$$\inf_{q\in\Pi(p_0,p_1)} \int \frac{\|x_0 - x_1\|^2}{2} dq(x_0,x_1) - \epsilon \text{Entropy}(q).$$

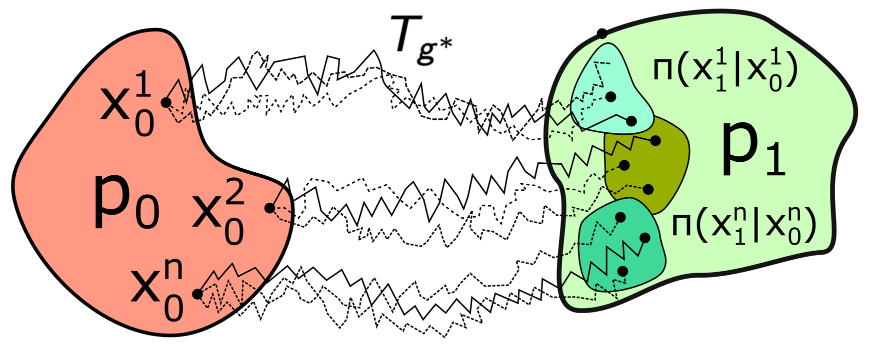
An equivalent dynamic formulation of this problem called the Scrödinger Bridge with Wiener prior (or just Scrödinger Bridge).

SB problem setup with Wiener prior. For two continuous distributions p_0 and p_1 on $\mathbb{R}^D$, the Scrödinger Bridge Problem with Wiener prior is:

$$\inf_{T_g\in\mathcal{D}(p_0,p_1)} \frac{1}{2\epsilon}\mathbb{E}_{T_g}\Big[\int_0^1 \|g(X_t,t)\|^2 dt\Big]$$

$$T_g : dX_t = g(X_t,t)dt + \sqrt{\epsilon}dW_t.$$

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



The minimizer $T^* = T_{g^*}$ is called the **Schrödinger Bridge** and $p^{T^*}(x_0,x_1)$ is the solution to static SB problem.

Bridge Matching is a recent alternative to Diffusion Models. It learns a diffusion between two distributions p_0 and p_1.

Reciprocal Projection

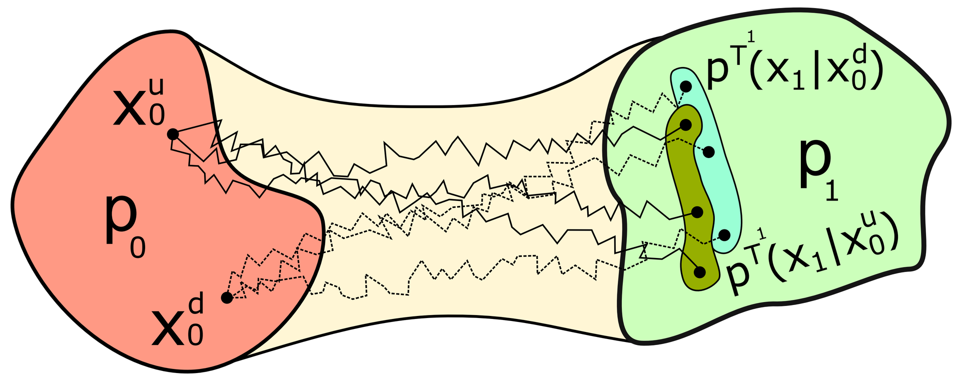
$$\text{proj}_{\mathcal{R}}(T) = \int W^{\epsilon}_{|x_0,x_1} dp^T(x_0,x_1)$$

Markovian Projection

$$\text{proj}_{\mathcal{M}}(T) : dX_t = g_{\mathcal{M}}(X_t,t)dt + \sqrt{\epsilon}dW_t$$

$$g_{\mathcal{M}} = \arg\min_g \int \|g(x_t,t) - \frac{x_1 - x_t}{1-t}\|^2 dp^T(x_t,x_1)$$

Combination of Reciprocal and Markovian projections is called Bridge Matching, i.e., $\text{BM}(T) = \text{proj}_{\mathcal{M}}(\text{proj}_{\mathcal{R}}(T))$.



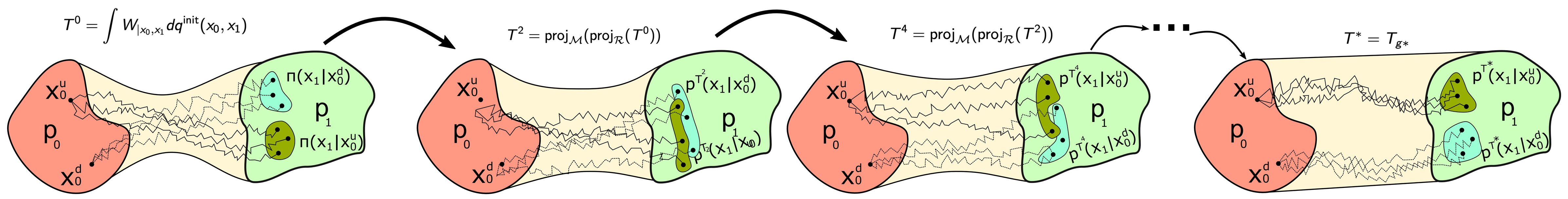
The Schrödinger Bridge is the **only process** in $\mathcal{D}(p_0,p_1)$ that isn't changed by both Reciprocal and Markovian projections, hence to solve SB it is sufficient to find a process $T = T^*$, s.t. $T = \text{proj}_{\mathcal{M}}(\text{proj}_{\mathcal{R}}(T))$.

Background: Schrödinger Bridges via Bridge Matching

Iterative Diffusion Bridge Matching.
Consider a sequence of alternating Markovian and Reciprocal projections, starting from T^0, s.t., $T^0_{t=0,1} = q^{\text{init}}(x_0,x_1)$ with p_0 and p_1 marginals. Let us call it the **IMF** sequence.

$$T^{2l+1} = \text{proj}_{\mathcal{M}}(T^{2l}) \qquad T^{2l+2} = \text{proj}_{\mathcal{R}}(T^{2l+1})$$

IMF sequence converges to a fixed point, i.e., $\lim_{l\to\infty} \text{KL}(T^l \| T^{l+1}) = 0$. In other words it converges to the Scrödinger Bridge between p_0 and p_1.



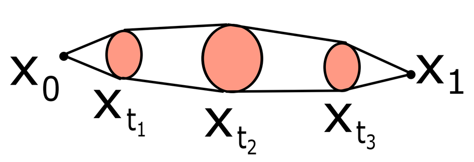
Practical implementation of this algorithm performs sequential Bridge Matching and is called Diffusion Schrödinger Bridge Matching (**DSBM**).

Problem: many NFE is required

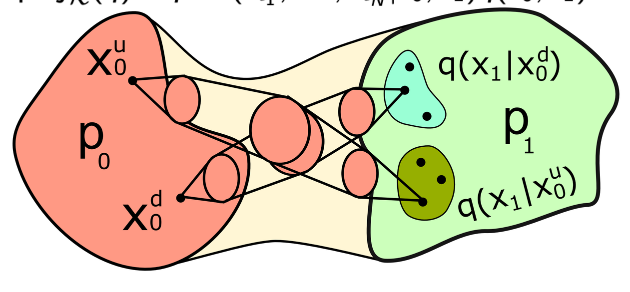
III Theory: Discrete in time Bridge Matching and D-IMF

Discrete Reciprocal Projection.
Makes a mixture of Discrete Brownian bridges $p^{W^\epsilon}(x_{t_1},\dots,x_{t_N}|x_0,x_1)$ with the distribution $p(x_0,x_1)$ of discrete stochastic process q at times $t=0$ and $t=1$.

Discrete Brownian Bridge $p^{W^\epsilon}(x_{t_1},x_{t_2},x_{t_3}|x_0,x_1)$.

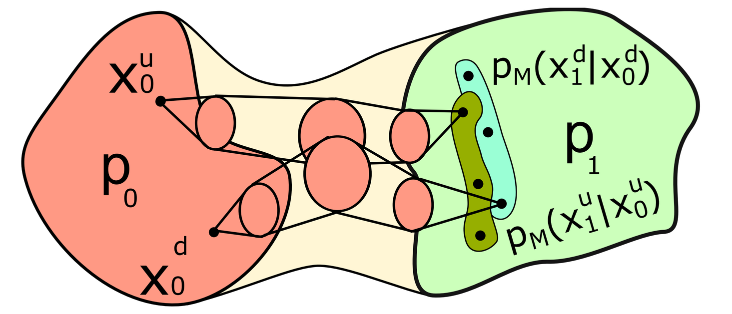


$\text{proj}_{\mathcal{R}}(q) = p^{W^\epsilon}(x_{t_1},\dots,x_{t_N}|x_0,x_1)q(x_0,x_1)$.



Discrite Markovian Projection.
Finds the markovian discrete stochastic process $p_{\mathcal{M}}$ which is the most similar to a process q.

$$p_{\mathcal{M}}(x_0,x_{t_1},\dots,x_{t_N},x_1) = [\text{proj}_{\mathcal{M}}(q)](x_0,x_{t_1},\dots,x_{t_N},x_1) = q(x_0)\prod_{n=1}^{N+1} q(x_{t_n}|x_{t_{n-1}})$$

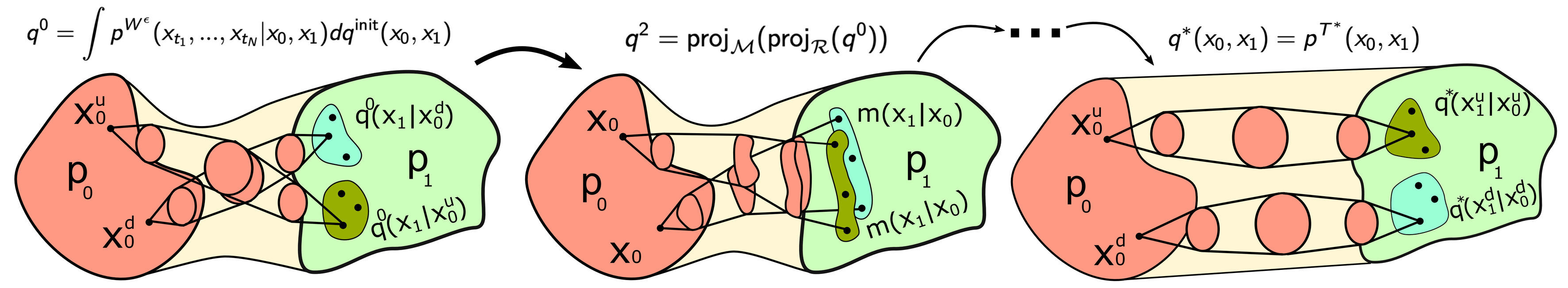


Theorem (Reciprocal and Markovian fixed point)

Consider discrete in time stochastic process q. ***If*** *q is not changed by both Reciprocal and Markovian projections, i.e., $q = \text{proj}_{\mathcal{M}}(q)$ and $q = \text{proj}_{\mathcal{R}}(q)$,* ***then*** *$q(x_0,x_1)$ is the static Schrödinger Bridge.*

Theorem (D-IMF)

Consider a sequence of discrete in time Markovian and Reciprocal projections: $q^{2l+1} = \text{proj}_{\mathcal{M}}(q^{2l})$, $q^{2l+2} = \text{proj}_{\mathcal{R}}(q^{2l+1})$, that starts from q^0, s.t., $q^0(x_0,x_1) = q^{init}(x_0,x_1)$ has marginals p_0 and p_1.
We call it ***D-IMF*** *sequence.*
Under mild assumptions D-IMF sequence converges in KL to p^{T^*}, where $p^{T^*}(x_0,x_1)$ is a static SB solution between p_0 and p_1, i.e., $\lim_{l\to\infty} KL(q^l(x_0,x_1)\|p^{T^*}(x_0,x_1)) = 0$.



IV Algorithm: Adversarial Schrödinger Bridge Matching

Reciprocal projection and sampling from it are available in **closed form**. Result of **markovian** projection is a sequence of transition kernels $q(x_{t_n}|x_{t_{n-1}})$. Each kernel is parametrized by a **conditional GAN** $q_\theta(x_{t_n}|x_{t_{n-1}})$ as in **DD-GAN** model. To battle either p_0 or p_1 distribution shift we implement D-IMF with both **forward** and **backward** in time markovian projections.

Algorithm 1: Adversarial SB matching (ASBM).

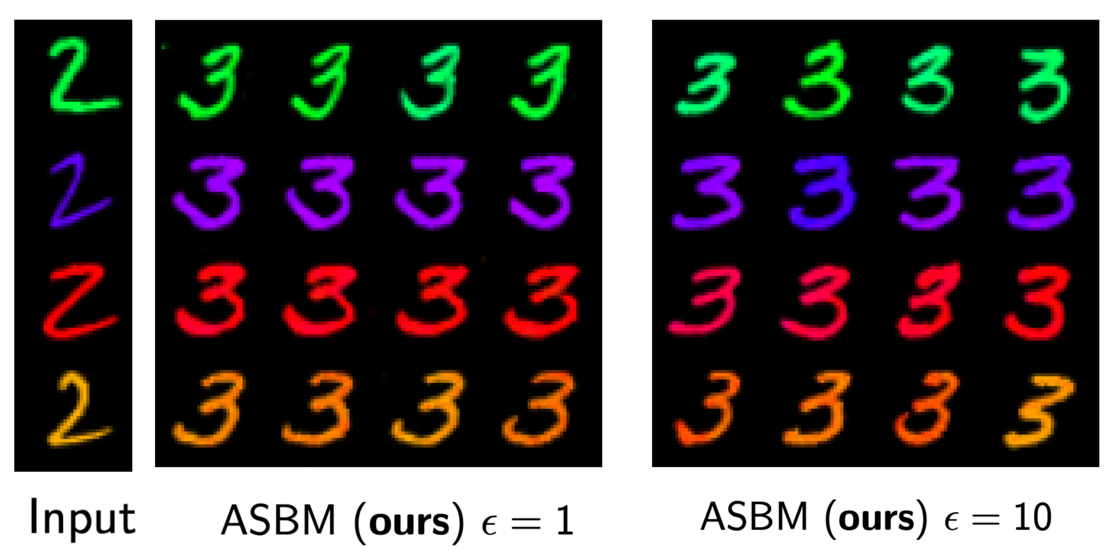
Input : number of intermediate steps N;
initial process $q^0(x_0,x_{t_1},\dots,x_{t_N},x_1)$ accessible by samples;
number of outer iteration $K\in\mathbb{N}$;
forward transitional density network $\{q_\theta(x_{t_n}|x_{t_{n-1}})\}_{n=1}^{N+1}$;
backward transitional density network $\{q_\eta(x_{t_{n-1}}|x_{t_n})\}_{n=1}^{N+1}$;
Output: $p_0(x_0)\prod_{n=1}^{N+1} q_\theta(x_{t_n}|x_{t_{n-1}}) \approx p_1(x_1)\prod_{n=1}^{N+1} q_\eta(x_{t_{n-1}}|x_{t_n}) \approx p^{T^*}(x_0,x_{\text{in}},x_1)$.
for $k=0$ **to** $K-1$ **do**
  $\{q_\theta(x_{t_n}|x_{t_{n-1}})\}_{n=1}^{N+1} := \arg\min_{q_\theta} \sum_{n=1}^{N+1} \mathbb{E}_{q(x_{t_{n-1}})} D_{\text{adv}}(q^{4k}(x_{t_n}|x_{t_{n-1}})\|q_\theta(x_{t_n}|x_{t_{n-1}}))$
  Let q^{4k+1} be given by $p_0(x_0)\prod_{n=1}^{N+1} q_\theta(x_{t_n}|x_{t_{n-1}})$;
  Let q^{4k+2} be given by $p^{W^\epsilon}(x_{t_1},\dots,x_{t_N}|x_0,x_1)q^{4k+1}(x_0,x_1)$;
  $\{q_\eta(x_{t_{n-1}}|x_{t_n})\}_{n=1}^{N+1} = \arg\min_{q_\eta} \sum_{n=1}^{N+1} \mathbb{E}_{q(x_{t_{n-1}})} D_{\text{adv}}(q^{4k+2}(x_{t_{n-1}}|x_{t_n})\|q_\eta(x_{t_{n-1}}|x_{t_n}))$;
  Let q^{4k+3} be given by $p_1(x_1)\prod_{n=1}^{N+1} q_\eta(x_{t_{n-1}}|x_{t_n})$;
  Let q^{4k+4} be given by $p^{W^\epsilon}(x_{t_1},\dots,x_{t_N}|x_0,x_1)q^{4k+3}(x_0,x_1)$;

We call our algorithm Adversarial Schrödinger Bridge Matching (**ASBM**)

Problem solved: NFE=4 is required

V Experiments: Unpaired Image-to-Image Translation

Qualitative results of our solver for solving the domain translation problem
1) between Men and Women on the Celeba dataset (128x128)
2) between digits "2" and digits "3" of the Colored MNIST dataset (32x32)



Model	ASBM (NFE=4)	DSBM (NFE=100)
$\epsilon = 1$	16.08	37.8
$\epsilon = 10$	14.73	89.19

Table 1: FID values *male→female* translation on Celeba dataset with 128 resolution size