Modern Methods for Schrödinger Bridge Problems

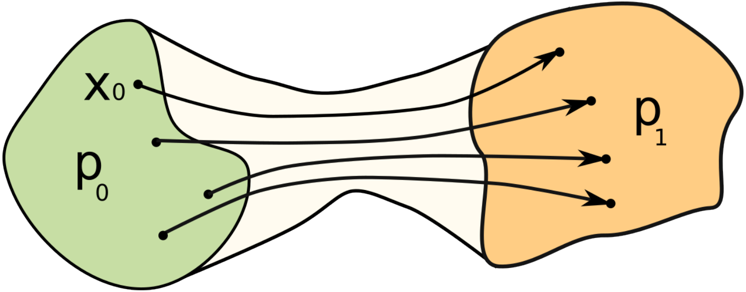
Bair Mikhailov, Ivan Razvorotnev, Nikita Kornilov

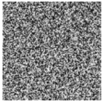
Skoltech
Skolkovo Institute of Science and Technology

Moscow, 25 October 2024

Motivation

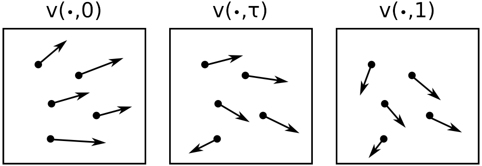
Goal: Map arbitrary distribution p_0 to arbitrary distribution p_1 (maybe with extra conditions).




/


Preliminaries

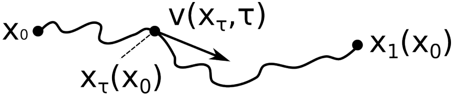
Vector field $v : \mathbb{R}^D \times [0,1] \to \mathbb{R}^D$.



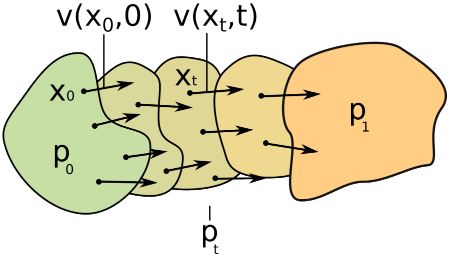
Movement of a point along the field.

Let $x_t(x_0)$ be the solution to $\mathrm{d}x_t = v(x_t, t)\mathrm{d}t$ with initial condition $x_{t=0} = x_0$, i.e.:

$$x_t(x_0) = x_0 + \int_0^t v(x_\tau(x_0), \tau)\mathrm{d}\tau,$$
$$dx_t = v(x_t, t)dt.$$

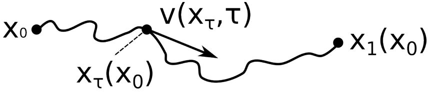


Flow Transport: Key Idea



Find a (time-dependent) vector field $v(x_t, t)$ which transports the probability mass of distribution p_0 to distribution p_1, i.e.:

$$\text{If } x_0 \sim p_0 \text{ then } x_1(x_0) \sim p_1, x_t(x_0) \sim p_t,$$

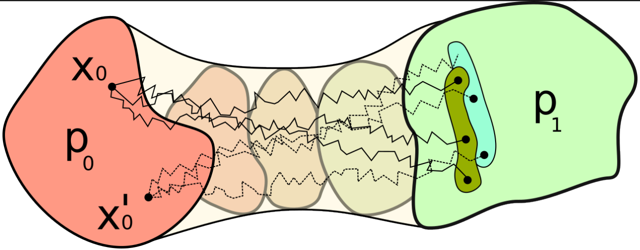


Solution is not unique!

Bridge Transport: Stochasticity

$$\mathrm{d}x_t = g(x_t, t)\mathrm{d}t + \sqrt{\epsilon}\mathrm{d}W_t \quad (\epsilon > 0).$$

We denote the process described by this SDE as T_g and



Bridge Matching: Define a distribution: $p_t^\epsilon \stackrel{\text{def}}{=} \mathcal{N}\left(x_t, \epsilon t(1-t)\right)$

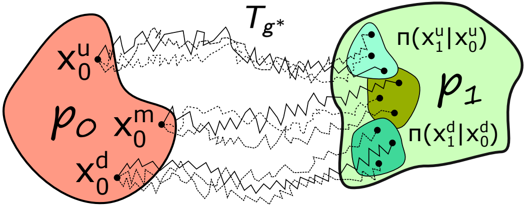
$$\min_{v} \mathbb{E}_{\substack{x_0 \sim p_0 \\ x_1 \sim p_1}} \mathbb{E}_{t \sim [0,1]} \mathbb{E}_{\tilde{x}_t \sim p_t^\epsilon} \left\| v(\tilde{x}_t, t) - \frac{x_1 - \tilde{x}_t}{1 - t} \right\|^2.$$

Schrödinger Bridge

We want to keep input-output similarity during transformation $p_0 \to p_1$ and spend the minimum possible kinetic energy:

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

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

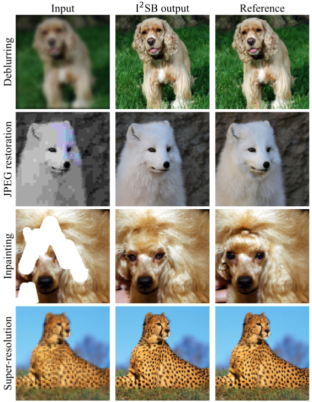


The minimizer T_{g^*}
is called the Schrödinger Bridge.

Examples of Schrödinger Bridge Models for Images

Image-to-image Schrodinger Bridge

for various image restoration problems



Diffusion Schrodinger Bridge Matching

for unpaired image-to-image translation

cat $\longrightarrow$ wild

wild $\longrightarrow$ cat

Key properties

SB-based algorithms typically keep one and enforce another of **two** key properties:

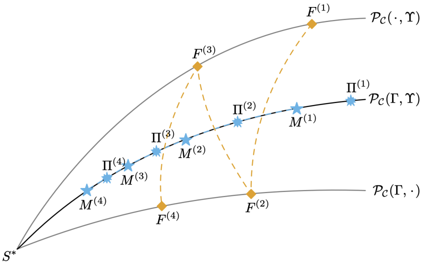
1. **Optimality property**: the similarity between input and translated object,
2. **Marginal matching property**: the input domain is translated to the target domain.

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

Time can be discrete as well with points $0 = t_0 < t_1 < \cdots < t_N < t_{N+1} = 1$:

Previous SB methods

1. **Iterative Bridge Matching** (analogue of Rectifed Flow):
   Keep marginals from initialization and refine optimality.
2. **Iterative Proportional Fitting [4]** (diffusions and drifts):
   Keep optimality from initialization and refine marginals.
   Start with $T_0(x_1, x_0) = p_0(x_0) W^{\epsilon}_{|x_0}(x_1)$ and alternate the final and starting marginals to p_1 and p_0, respectively.

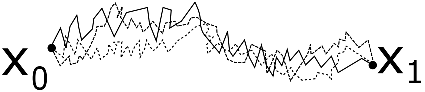


Due to computational errors and many iterations, these methods **forget** the initial condition.

Reciprocal projection

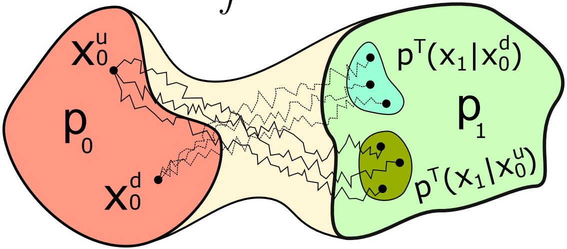
Make a mixture of Brownian bridges $W^{\epsilon}_{|x_0,x_1}$ with distribution $p(x_0, x_1)$ of process T:

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



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

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



Discrete time:

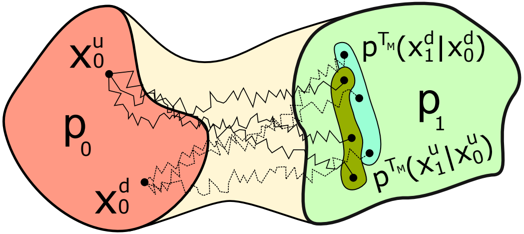
$$\text{proj}_{\mathcal{R}}(T) = p^T(x_0, x_1) W^{\epsilon}_{x_{in}|x_0,x_1}.$$

Marginals do not change.

Markovian projection

Find markovian process that is the most similar to a process T:

$$g_M = \arg\min_{g} \int \|g(x_t, t) - \frac{x_1 - x_t}{1 - t}\| dp^T(x_t, x_1).$$



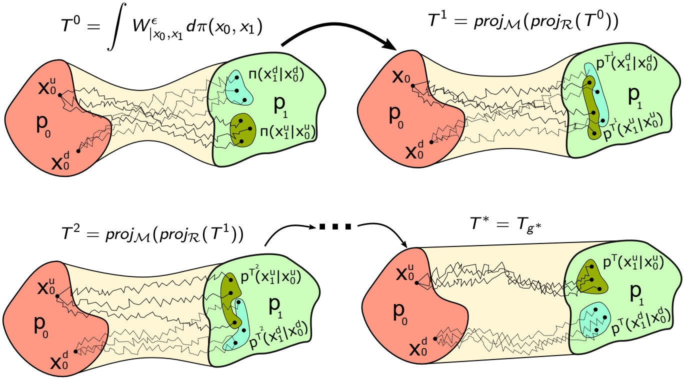
In discrete time, this projection even simplier:

$$\text{proj}_{\mathcal{M}}(T) = p^T(x_0) \cdot \prod_{i=0}^{N} p^T(x_{t_{i+1}}|x_{t_i}) = p^T(x_1) \cdot \prod_{i=0}^{N} p^T(x_{t_i}|x_{t_{i+1}}).$$

Marginals do not change.

Iterative Markovian Fitting (IMF)

Consider a sequence of alternating Markovian and Reciprocal projections from T_0 with plan π with marginals p_0 and p_1. Initial T_0 keeps **marginal matching property**.



With each iteration, **optimality** property refines and converges to SB solution. SB is the one and only process which is both markovian and reciprocal.

IMF in practice: discrete time

$$T^{2k+1} = \mathrm{proj}_{\mathcal{R}}(T^{2k}) \overset{\mathrm{def}}{=} T^{2k}(x_0, x_1) p^{W^\epsilon}(x_{\mathrm{in}} | x_0, x_1),$$

$$T^{2k+2} = \mathrm{proj}_{\mathcal{M}}(T^{2k+1}) \overset{\mathrm{def}}{=} \underbrace{T^{2k+1}(x_0) \prod_{n=1}^{N+1} T^{2k+1}(x_{t_n} | x_{t_{n-1}})}_{\text{forward representation}} = \underbrace{T^{2k+1}(x_1) \prod_{n=0}^{N} T^{2k+1}(x_{t_n} | x_{t_{n+1}})}_{\text{backward representation}}$$

We parametrize via GANs $\{T_\theta(x_{t_n} | x_{t_{n-1}})\}$ (**forward parametrization**) and $T(x_0) = p_0(x_0)$ or $\{T_\phi(x_{t_n} | x_{t_{n+1}})\}$ (**backward parametrization**) and $T(x_1) = p_1(x_1)$.

Vanilla IMF: only forward representation is used, and due to computational errors *marginal matching property is losing*.

Bidirectional IMF (heuristic): alternate forward and backward representations, and *errors do not accumulate*.

New observations [1]

ASBM [2]: Heuristic Bidirectional IMF in discrete time

DSBM [3]: Heuristic. Bidirectional IMF in continuous time (diffusions or drifts).

These methods not only save from error accumulation, but fix the initial errors, allowing *any* initial T_0.

We can even use T_0 which final marginal does not equal to p_1.

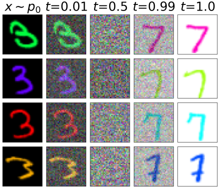
Theoretically prove exponential convergence to the ground truth solution in case of 1D Gaussians

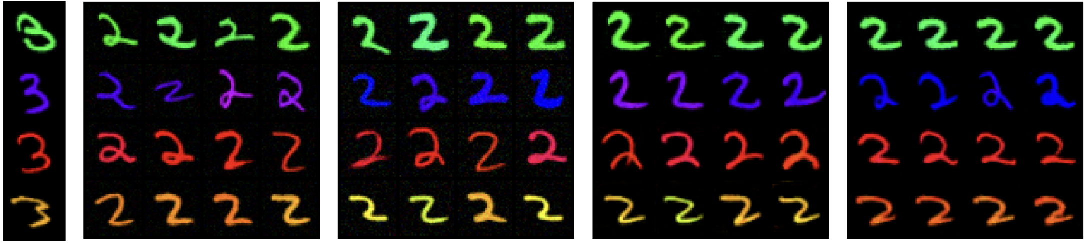
Experiments: Color MNIST 32x32 $3 \to 2$

We fix parameter $\epsilon = 10$.

Initial processes:

- IFM $= p_3(x_0) \times p_2(x_1) \times W^{\epsilon}_{|x_0,x_1}$ – standard
- Inverted 7 $= p_3(x_0) \times p_7^{inv}(x_1) \times W^{\epsilon}_{|x_0,x_1}$ – misleading





(a) $x \sim p_0$ (b) DSBM-IMF (c) DSBM-*Inversed 7* (d) ASBM-IMF (e) ASBM-*Inverted 7*

Experiments: Celeba 64x64 Male→Female

Parameters: $\epsilon = 1$, IMF-OT $= \pi^{OT}(x_0, x_1) \times W^{\epsilon}_{|x_0,x_1}$, Independent $= p_0(x_0) \times p_0(x_1) \times W^{\epsilon}_{|x_0,x_1}$

(d) $x \sim p_0$

(e) ASBM-IMF-OT

(f) ASBM-*Independent* $p_0 \to p_0$

Experiments: Celeba 64x64 Male→Female

Parameters: $\epsilon = 1$, IMF-OT $= \pi^{OT}(x_0, x_1) \times W^{\epsilon}_{|x_0,x_1}$, Ind $= p_{male}(x_0) \times p_{male}(x_1) \times W^{\epsilon}_{|x_0,x_1}$

(a) $x \sim p_0$ (b) DSBM-IMF-OT (c) DSBM-*Independent* $p_0 \to p_0$

Experiments: SB Benchmark

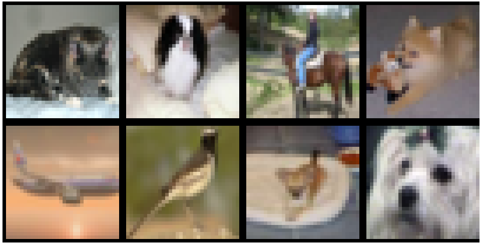
	$\epsilon = 0.1$				$\epsilon = 1$			
	$D=2$	$D=16$	$D=64$	$D=128$	$D=2$	$D=16$	$D=64$	$D=128$
Best algorithm on benchmark†	1.94	13.67	11.74	11.4	1.04	9.08	18.05	15.23
DSBM-IMF	1.21	4.61	9.81	19.8	0.68	**0.63**	5.8	29.5
DSBM-IPF	2.55	17.4	15.85	17.45	0.29	0.76	4.05	29.59
DSBM-$Ind(p_0, p_0)$	2.72	11.7	16.5	17.02	0.41	0.92	**3.7**	29
ASBM-IMF†	0.89	8.2	13.5	53.7	**0.19**	1.6	5.8	**10.5**
ASBM-IPF	3.06	14.37	44.35	32.5	0.18	1.68	9.25	20.47
ASBM-$Ind(p_0, p_0)$	3.99	15.73	39.3	40.32	0.18	1.68	6.16	12.8
Bridge Matching	**0.54**	**3.7**	**9.5**	**10.9**	0.2	1.1	9	23

Comparisons of $\mathrm{cB}\mathbb{W}_2^2$-UVP $\downarrow$ (%) (normalized MSE error) between the ground truth SB solution and the learned solution on the benchmark. The best metric is **bolded**.

Our Experiments: Downsample CIFAR DSBM

Github: https://github.com/MikhailovBair/DASBM

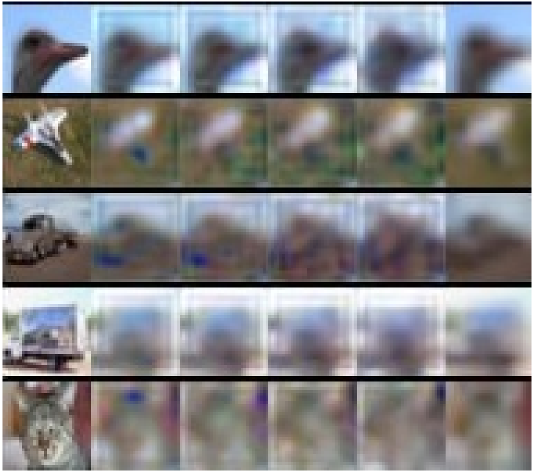
Parameter $\varepsilon = 1$, steps $= 30$, 10 iterations, plan — independent $p_0(x_0) \times p_0(x_1)$



Our Experiments: Downsample CIFAR ASBM

Github: https://github.com/MikhailovBair/DASBM

Parameter $\varepsilon = 1$, 5 points, 10 iterations, plan — independent $p_0(x_0) \times p_0(x_1)$



Conclusions

1. IMF algorithms are much more powerful than previously was thought,
2. New initial plans open new opportunities for generation,
3. Heuristics from SB problems can be used to enhance Rectified Flow.

Team

(a) Bair Mikhailov DSBM Extra Experiments, Repo

(b) Ivan Razvorotnev ASBM Extra Experiments

(c) Nikita Kornilov Idea and Presentation

Thanks for your attention!

Refenences

[1] Sergei Kholkin et al. (2024). **"Diffusion & Adversarial Schr\" odinger Bridges via Iterative Proportional Markovian Fitting".** In: *arXiv preprint arXiv:2410.02601*

[2] Nikita Gushchin et al. (2024). **"Adversarial Schrödinger Bridge Matching".** In: *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. URL: https://openreview.net/forum?id=L3Knnigicu

[3] Yuyang Shi et al. (2024). **"Diffusion Schrödinger bridge matching".** In: *Advances in Neural Information Processing Systems* 36

[4] Stefano Peluchetti (2023). **"Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling".** In: *Journal of Machine Learning Research* 24.374, pp. 1–51