Diffusion and Adversarial Schrodinger Bridges for Image-to-Image problems

Evgeny Burnaev

Prof., Head of AI center, Skoltech
Leading Scientist, AIRI

Progress in Generative AI

Progress in the field

Now (2024)
(results of generation of Kandinsky model by Sber)

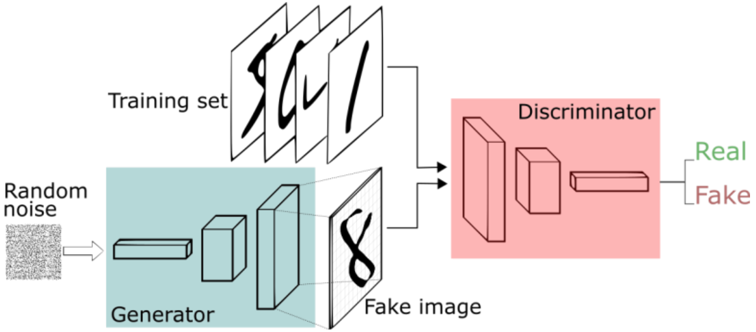
Text prompt: Flamingo, surrealism, nature-inspired, flowerpunk, delicate materials, flamboyant, studio photography.

Text prompt: Cow in the style of floral surrealism in the meadow, nature-inspired camouflage, flowerpunk, delicate materials, flamboyant, studio photography.

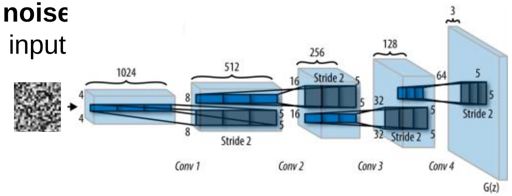
Text Prompt: Tornado of flowers.

Approaches to Generative Modeling

Adversarial models (GANs, 2014)



Random
noise
input

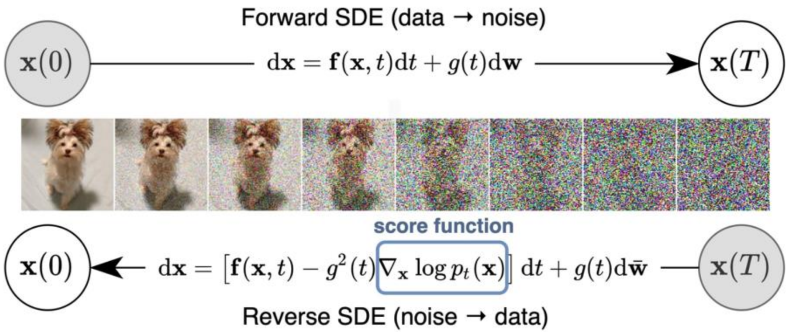


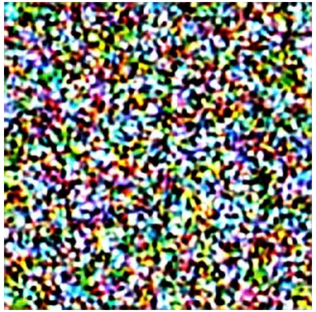
=

Generated images
(these people do not exist!)

Deep **neural network**
(generator)

Diffusion models (DM, 2019)





Limitations of diffusion models

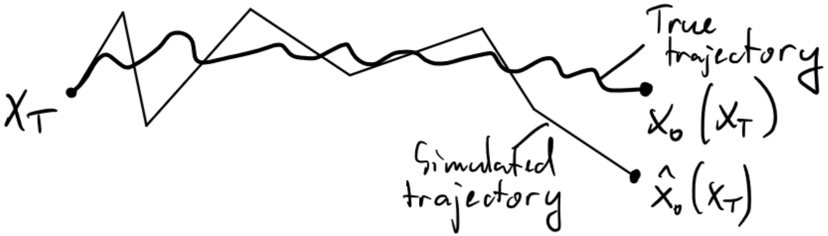
Slow sampling

To simulate the denoising process:

$$x_t = \left[f(x_t, t) - g^2(t)\color{red}{\nabla_x \log p(x_t, t)}\right] t + g(t)\overline{W}_t$$

one uses the discretization (e.g., Euler simulation):

$$x_{t-\Delta t} = x_t - \left[f(x_t, t) - g^2(t)\color{red}{\nabla_x \log p(x_t, t)}\right]\Delta t + \color{blue}{\xi_t},$$
$$\color{blue}{\xi_t} \sim \mathcal{N}(0, g^2(t)\Delta t \cdot I).$$

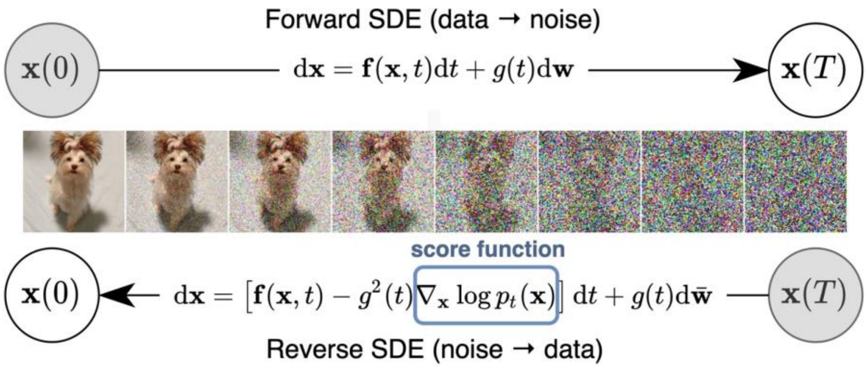


Diffusion trajectories in both SDE and ODE form are not straight and are HARD to simulate.

Only noise to data generation

- The forward process is pre-defined and maps complex data distribution to the normal distribution

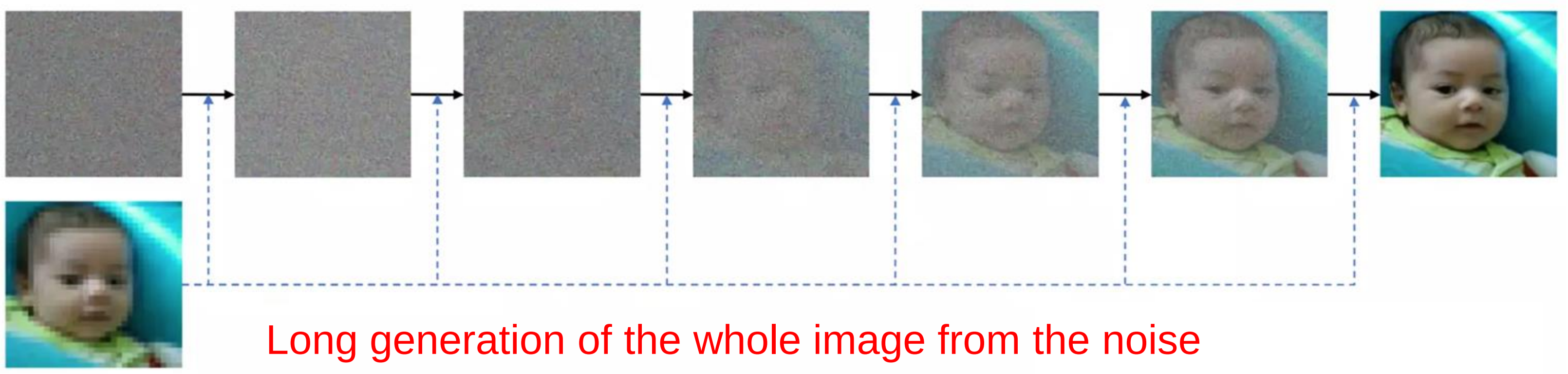
=> The reverse process starts from gaussian noise



Cannot perform direct **data-to-data translation**, e.g image-to-image style transfer or super-resolution.

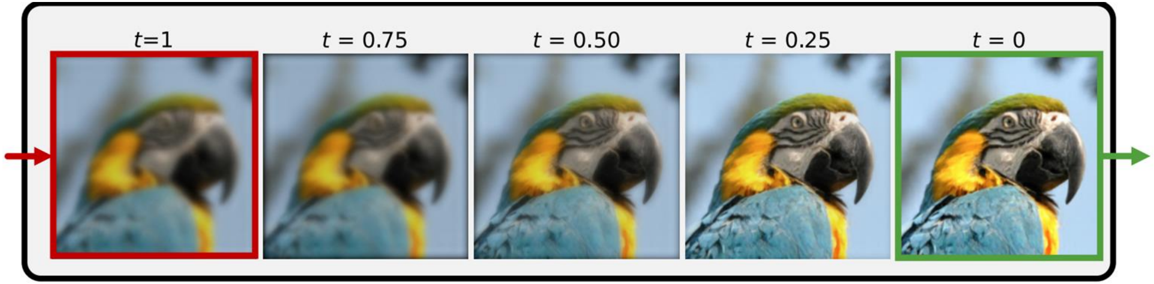
Conditional diffusions vs data-to-data diffusions

Conditional diffusions use additional input as a condition to guide generation from noise.



Long generation of the whole image from the noise

Data-to-data diffusions start from the input image and only add/modify the remaining details.



Faster generation

Better quality

Data-to-data diffusions: Schrödinger Bridge (SB)

Schrödinger Bridge formulation

For two arbitrary distributions $\mathbf{p_0}$ and $\mathbf{p_1}$ it aims to find a diffusion $\mathbf{T_g}$ given by the SDE:

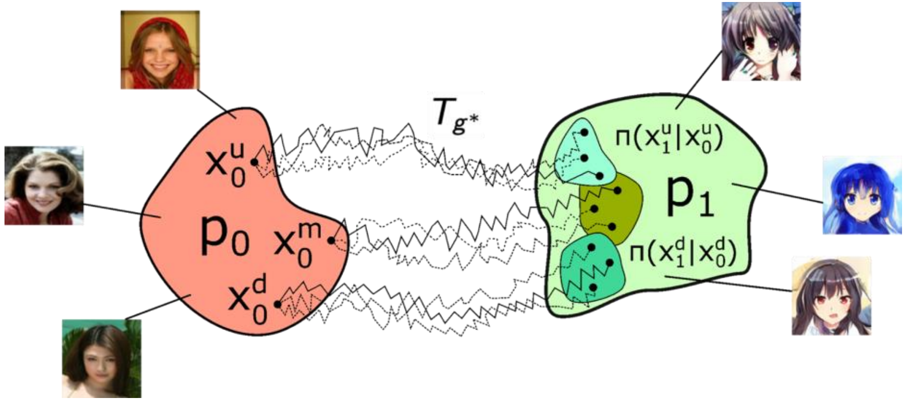
$$T_g : dx_t = \underbrace{g(x_t, t)}_{\text{"velocity"}}\, dt + \sqrt{\epsilon} dW_t$$

which transforms $\mathbf{p_0}$ to $\mathbf{p_1}$ and minimizes the energy:

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

(hyperparameter ε regulates the amount of noise in the trajectories)

Schrödinger Bridge for image-to-image style transfer



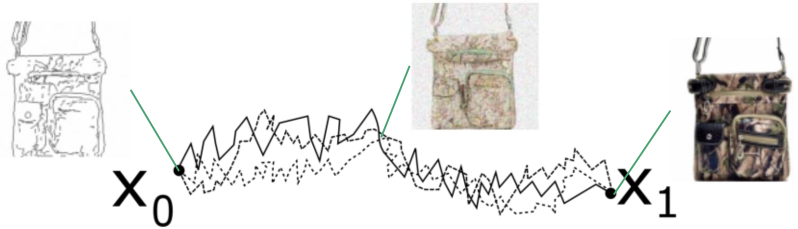
- Data-to-data diffusion
- Can be trained without paired data
- The “most straight” trajectories

Image-to-Image Schrödinger Bridge (I2SB) idea

1. Build a non-markovian forward process **T** using paired data coupling $\mathbf{q_{data}(x_0,x_1)}$ and a stochastic bridge $\mathbf{q(x_t \mid x_0, x_1)}$:

x_0 x_1

$\sim q_{data}(x_0, x_1)$



$$q(X_t|X_0, X_1) = \mathcal{N}(X_t; \mu_t(X_0, X_1), \Sigma_t)$$

$$\mu_t = \frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} X_0 + \frac{\sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} X_1, \quad \Sigma_t = \frac{\sigma_t^2 \bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} \cdot I$$

2. Learn a diffusion that is closest to the forward process **T** using **Bridge Matching** (generalization of *Flow Matching* on diffusions):

$$dx_t = g(x_t, t)dt + \sqrt{\beta_t} dW_t$$

$$g(x_t, t) = \arg\min_{g'} \int \|g'(x_t, t) - \beta_t \frac{x_1 - x_t}{\bar{\sigma}_t}\|^2 dp^T(x_t, x_1)$$

Relation between SDE and bridge parameters

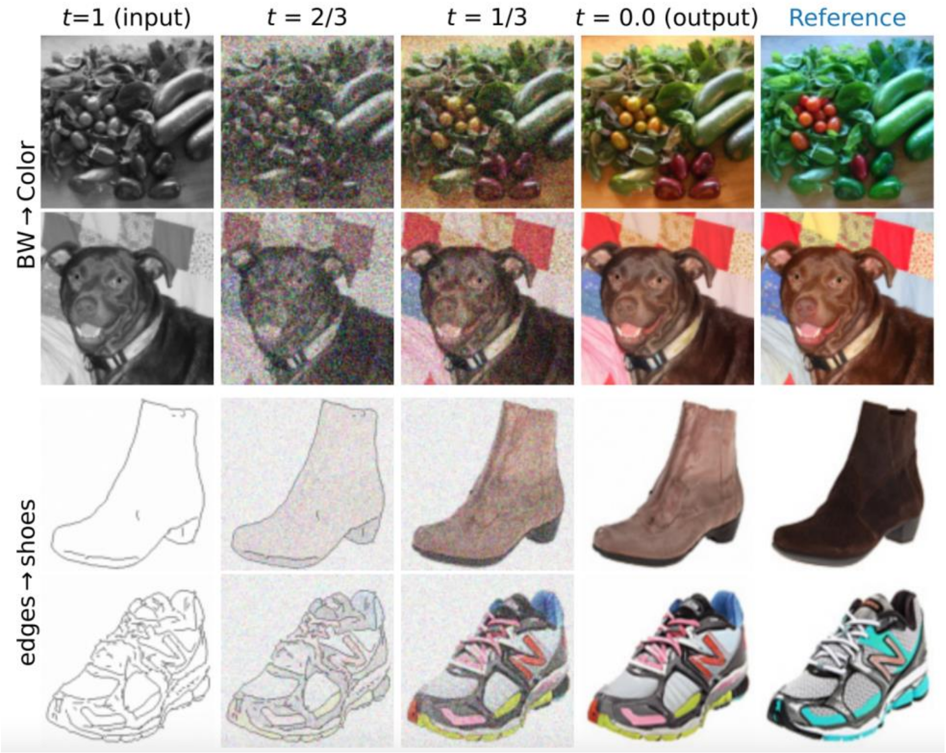
$$\bar{\sigma}_t^2 := \int_t^1 \beta_\tau \mathrm{d}\tau$$

$$\sigma_t^2 := \int_0^t \beta_\tau \mathrm{d}\tau$$

Liu, Guan-Horng, et al. "I2SB: Image-to-Image Schrödinger Bridge." *International Conference on Machine Learning*. PMLR, 2023.

Image-to-Image Schrödinger Bridge (I2SB) examples

I2SB trajectories for colorization and sketch to image translations



I2SB comparison with Palette (conditional diffusion) on inpainting and SR

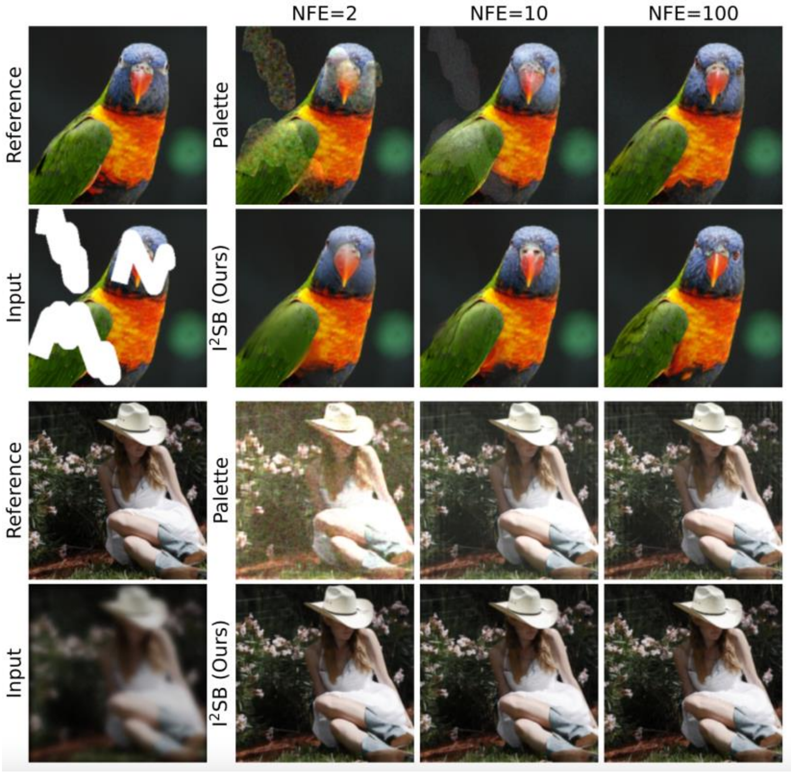


Image-to-Image Schrödinger Bridge (I2SB) more results

Jpeg restoration

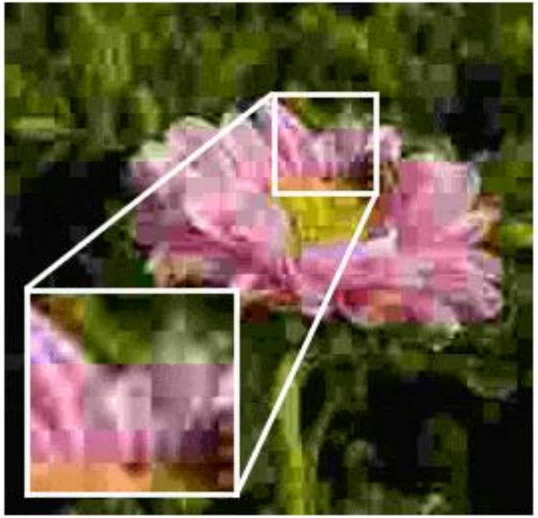
Method	FID-10k ↓
DDRM (Kawar et al., 2022b)	28.2
ΠGDM (Song et al., 2022)	8.6
Palette (Saharia et al., 2022)	8.3
CDSB (Shi et al., 2022)	38.7
I^2SB (Ours)	**4.6**

SR with bicubic degradation

Method	FID-10k ↓
DDRM (Kawar et al., 2022a)	21.3
DDNM (Wang et al., 2022b)	13.6
ΠGDM (Song et al., 2022)	3.6
ADM (Dhariwal & Nichol, 2021)	14.8
CDSB (Shi et al., 2022)	13.6
I^2SB (Ours)	**2.8**

Super-resolution

Degraded Input (JPEG QF5)



Gaussian deblurring

Method	FID-10k ↓
DDRM (Kawar et al., 2022a)	6.1
DDNM (Wang et al., 2022b)	**2.9**
Palette (Saharia et al., 2022)	3.1
CDSB (Shi et al., 2022)	7.7
I^2SB (Ours)	3.0

Freform inpainting

Method	FID-10k ↓
DDRM (Kawar et al., 2022a)	9.7
DDNM (Wang et al., 2022b)	3.2
Palette (Saharia et al., 2022)	4.0
CDSB (Shi et al., 2022)	8.5
I^2SB (Ours)	**2.9**

Limitations:

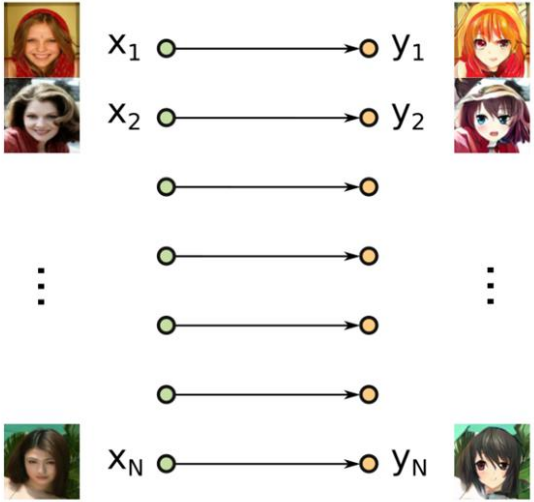
- Requires paired data $\mathbf{q_{data}(x_0, x_1)}$ for training
- Schrödinger Bridge is restored only if "**optimal**" coupling $\mathbf{q^*(x_0,x_1)}$ is used

Paired vs. Unpaired Learning

Supervised

Paired train samples are available:

$$\{(x_1, y_1), \ldots, (x_N, y_N)\}.$$

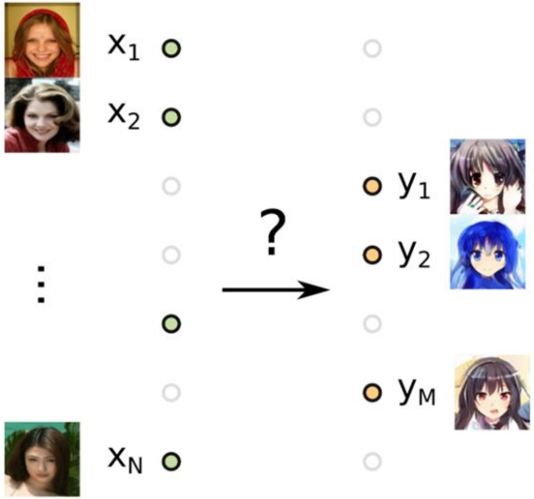


Issue: collecting/constructing pairs may be costly, non-trivial or impossible

Unsupervised

Only *unpaired* train samples are given:

$$\{x_1, \ldots, x_N\}, \{y_1, \ldots, y_M\}.$$



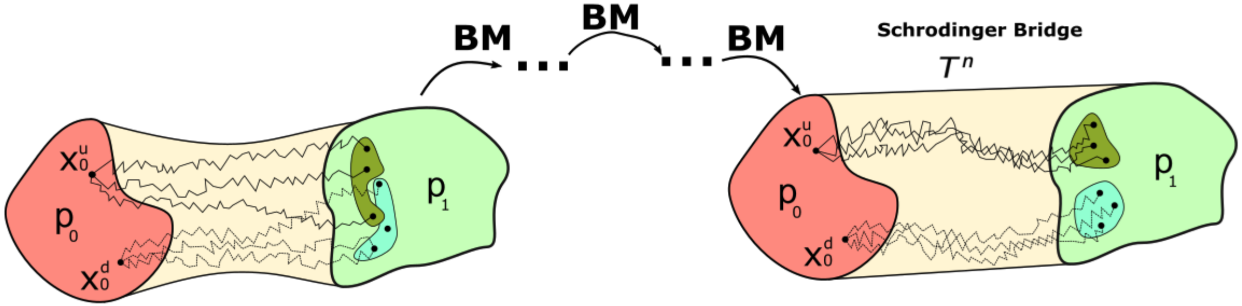
Issue: ill-posed problem, many possible solutions (not all are practically meaningful)

Iterative Bridge Matching

One may learn Bridge Matching diffusion T^{n+1} using another already learned Bridge Matching diffusion inputs and outputs $T^n_{|0,1}$. It is called **Diffusion Schrödinger Bridge Matching** (DSBM) [1] and generalizes the *Rectified Flows*

$$T^{n+1} = BM(T^n_{|0,1})$$

As $n \to \infty$, T^n converge to the Schrödinger Bridge.



Limitations: (1) One has to sample from the previously learned diffusion while learning (2) One has to learn many diffusions iteratively = potential error accumulation, **slow**

[1] Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. In Thirty-seventh Conference on Neural Information Processing Systems, 2023.

DSBM: Examples

Unpaired transfer (*cat→wild*)
on *AFHQ 512x512* dataset

Cat (Input) → **Wild** (Output)

Unpaired transfer (*male→female*)
on *Celeba 128x128* dataset

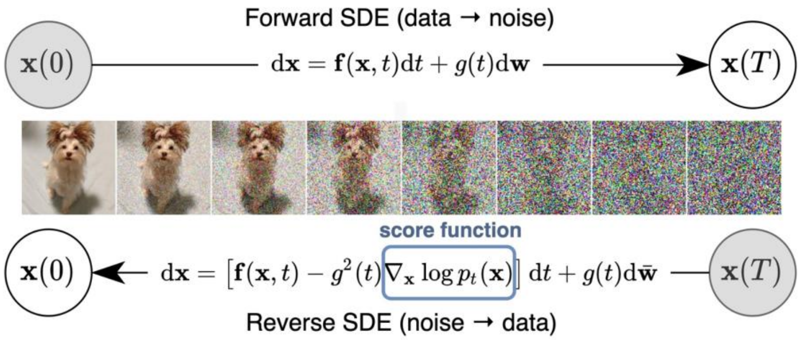
Man (Input) diffuses to **Woman** (Output)

Problem: large NFE = 100 – **long inference**

Denoising Diffusion vs. Denoising Diffusion GAN

Denoising Diffusion

Method: learn continuous in time diffusion via the conditional **score matching**



Requires large NFE (≥100)

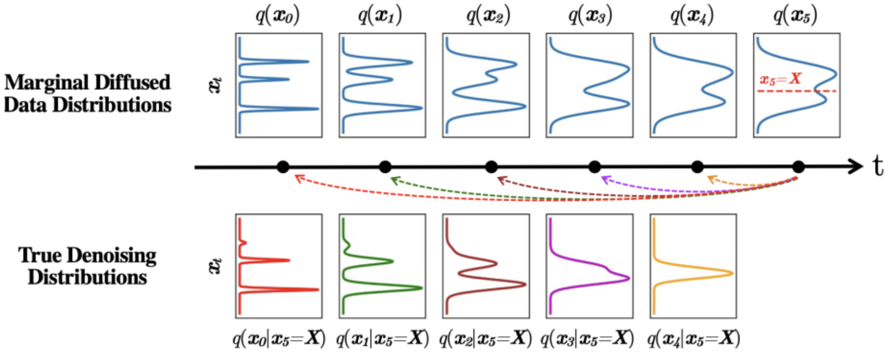
Denoising Diffusion in practice

Method: discrete Markov process

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t\geq 1} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \quad q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Approximation as denoising process

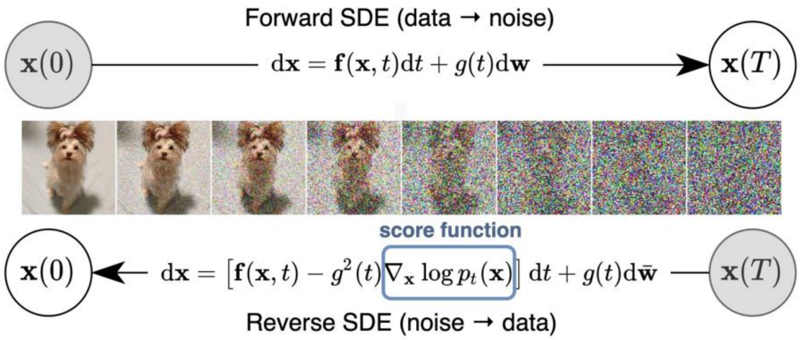
$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T)\prod_{t\geq 1} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \quad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \sigma_t^2\mathbf{I})$$



Denoising Diffusion vs. Denoising Diffusion GAN

Denoising Diffusion

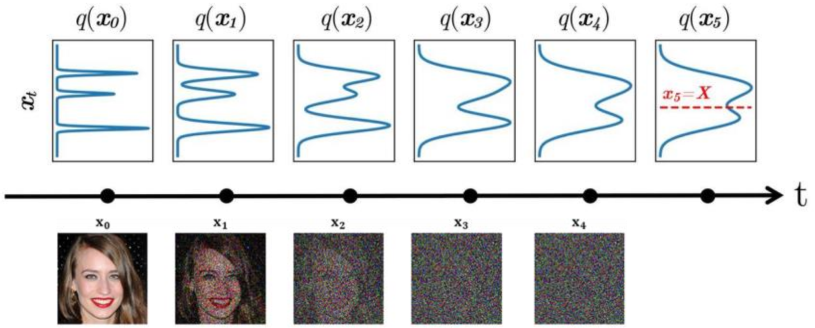
Method: learn continuous in time diffusion via the conditional **score matching**



Requires large NFE (≥100)

Denoising Diffusion GAN [2]

Method: learn markov chain in discrete time via GAN (adversarial) loss



Requires just NFE=4

Problem: No generalization for the Bridge Matching

[2] Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. In *International Conference on Learning Representations*, 2021.

Adversarial Schrodinger Bridge Matching (ASBM, ours)

We proposed an **adversarial bridge matching** technique which can learn diffusion bridges using just several transitions in discrete time instead of hundreds in the diffusion bridge matching

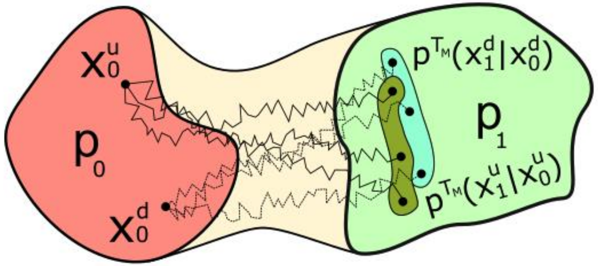
[3] Gushchin, N., Selikhanovych, D., Kholkin, S., Burnaev, E., & Korotin, A. (2024). **Adversarial Schrodinger Bridge Matching**.

Adversarial Schrödinger Bridge Matching

We have generalized DDGAN to Bridge Matching. Our idea is to represent the diffusion in bridge matching diffusion by discrete time Markov chain with learnable transitions [3]

Bridge Matching and DSBM (before us)

NFE=100

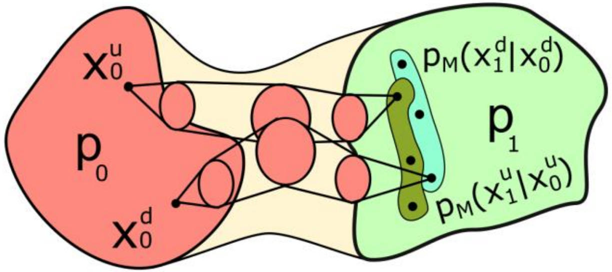


They learn the diffusion drift:

$$L(\theta) = \int \|g_\theta(x_t, t) - \frac{x_1 - x_t}{1 - t}\|^2 dp^T(x_t, x_1)$$

Adversarial Schrödinger Bridge Matching
(ASBM, ours)

NFE=4



We just learn several transition densities:

$$L(\theta) = \sum_{n=1}^{N+1} \mathbb{E}_{q(x_{t_{n-1}})} D_{\text{adv}}\left(q(x_{t_n}|x_{t_{n-1}}) || q_\theta(x_{t_n}|x_{t_{n-1}})\right)$$

Extra Theory

Discrete reciprocal processes

- Finite-time projection of the Brownian bridge $W^{\epsilon}_{|x_0,x_1}$:

$$p^{W^\epsilon}(x_{t_1},\dots,x_{t_N}|x_0,x_1) = \prod_{n=1}^{N} p^{W^\epsilon}(x_{t_n}|x_{t_{n-1}},x_1),$$

$$p^{W^\epsilon}(x_{t_n}|x_{t_{n-1}},x_1) =$$

$$= \mathcal{N}(x_{t_n}|x_{t_{n-1}} + \frac{t_n - t_{n-1}}{1-t_{n-1}}(x_1 - x_{t_{n-1}}), \epsilon\frac{(t_n - t_{n-1})(1-t_n)}{1-t_{n-1}})$$

- Distribution $p^{W^\epsilon}(x_{t_1},\dots,x_{t_N}|x_0,x_1)$ defines a discrete stochastic process, which we call a *discrete Brownian bridge*
- Distribution $q \in \mathcal{P}_{2,ac}(\mathbb{R}^{D\times(N\times2)})$ is a *mixture of discrete Brownian bridges* if

$$\boxed{q(x_0,x_{t_1},\dots,x_{t_N},x_1) = p^{W^\epsilon}(x_{t_1},\dots,x_{t_N}|x_0,x_1)q(x_0,x_1),}$$

where $q(x_0,x_1)$ is a joint marginal distribution
- We denote the set of all such mixtures as $\mathcal{R}(N) \subset \mathcal{P}_{2,ac}(\mathbb{R}^{D\times(N+2)})$ and call them **discrete reciprocal processes**

Discrete Markovian processes

- Discrete process $q \in \mathcal{P}_{2,ac}(\mathbb{R}^{D\times(N+2)})$ is Markovian if

$$\boxed{q(x_0, x_{t_1}, x_{t_2}, \ldots, x_{t_N}, x_1) = q(x_0) \prod_{n=1}^{N+1} q(x_{t_n}|x_{t_{n-1}})}$$

- Let $\mathcal{M}(N) \subset \mathcal{P}_{2,ac}(\mathbb{R}^{D\times(N+2)})$ by a set of all such **discrete Markovian processes**

Solution of static SB in discrete time

- **Theorem**: consider any discrete process $q \in \mathcal{P}_{2,ac}(\mathbb{R}^{D\times(N+2)})$, which is *simultaneously reciprocal and markovian*, i.e. $q \in \mathcal{R}(N)$ and $q \in \mathcal{M}(N)$ and has marginals $q(x_0) = p_0(x_0)$ and $q(x_1) = p_1(x_1)$:

$$\boxed{\begin{aligned} q(x_0, x_{t_1}, \dots, x_{t_N}, x_1) &= p^{W^\epsilon}(x_{t_1}, \dots, x_{t_N} | x_0, x_1) q(x_0, x_1) \\ &= q(x_0) \prod_{n=1}^{N+1} q(x_{t_n} | x_{t_{n-1}}), \end{aligned}}$$

  Then $q(x_0, x_{t_1}, \dots, x_{t_N}, x_1) = p^{\xi^*}(x_0, x_{t_1}, \dots, x_{t_N}, x_1)$, i.e., it is the finite-dimensional projection of the Schrödinger Bridge ξ^* to the considered times. Moreover, its joint marginal $q(x_0, x_1)$ at times $t = 0, 1$ is the solution to the **static SB** problem between p_0 and p_1, i.e., $q(x_0, x_1) = p^{\xi^*}(x_0, x_1)$
- Thus, to solve the static SB problem, it is enough to find a **Markovian mixture of discrete Brownian bridges**
- We propose the Discrete-time Iterative Markovian Fitting (D-IMF) procedure

Discrete Reciprocal Projection

- Let $q \in \mathcal{P}_{2,ac}(\mathbb{R}^{D\times(N+2)})$ be a discrete stochastic process
- The **reciprocal projection** $\text{proj}_{\mathcal{R}}(q)$ is a discrete stochastic process

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

- This projection takes the joint distribution of start and end points $q(x_0, x_1)$ and inserts the Brownian Bridge for intermediate time moments

Discrete Markovian Projection

- Let $q \in \mathcal{P}_{2,ac}(\mathbb{R}^{D\times(N+2)})$ be a discrete stochastic process
- The **Markovian projection** of q is a discrete stochastic process $\text{proj}_{\mathcal{M}}(q) \in \mathcal{P}_{2,ac}(\mathbb{R}^{D\times(N+2)})$ whose joint distribution is

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

D-IMF procedure converges to the Schrödinger Bridge

- Let $p_0 \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$ and $p_1 \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$ be two distributions at $t = 0$ and $t = 1$
- D-IMF starts with any discrete mixture of Brownian bridges $p^{W^\epsilon}(x_{t_1}, \dots, x_{t_N} | x_0, x_1) q(x_0, x_1)$, where $q(x_0, x_1) \in \Pi(p_0, p_1) \cap \mathcal{P}_{2,ac}(\mathbb{R}^{D \times 2})$
- Iterations:

$$\boxed{q^{2l+1} = \text{proj}_{\mathcal{M}}(q^{2l}), \quad q^{2l+2} = \text{proj}_{\mathcal{R}}(q^{2l+1})}$$

- **Theorem**: The sequence q^l converges in KL to p^{ξ^*}. In particular, $q^l(x_0, x_1)$ convergence to the solution $p^{\xi^*}(x_0, x_1)$ of the static SB:

$$\lim_{l \to \infty} \text{KL}\left(q^l \| p^{\xi^*}\right) = 0, \quad \text{and} \quad \lim_{l \to \infty} \text{KL}\left(q^l(x_0, x_1) \| p^{\xi^*}(x_0, x_1)\right) = 0.$$

ASBM vs. DSBM: Examples

Unpaired transfer (*male→female*) on Celeba 128x128 dataset using DSBM and ASBM (ours)

Before:
Long DSBM generation process (NFE=100)

After:
Fast **ASBM** (ours) generation process (NFE=4)

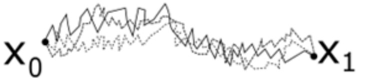
Comparison of approaches

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

Reciprocal Projection.

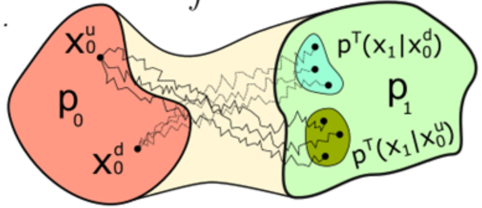
Makes a mixture of Brownian bridges $W^{\epsilon}_{|x_0,x_1}$ with the distribution $p^T(x_0, x_1)$ of stochastic process T at times $t=0$ and $t=1$.

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



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

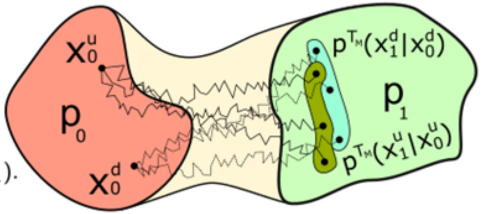
$$\mathrm{proj}_{\mathcal{R}}(T) = \int W^{\epsilon}_{|x_0,x_1} dp^T(x_0, x_1)$$



Markovian Projection.

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

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

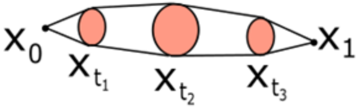


Discrete Bridge Matching constructs a discrete markovian process between p_0 and p_1 by combining discrete reciprocal and discrete Markovian projections of stochastic processes.

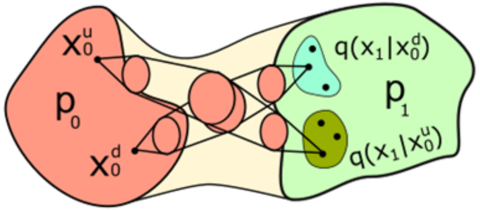
Discrete Reciprocal Projection.

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



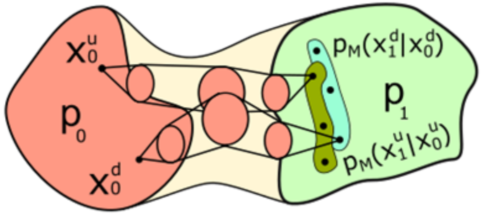
$$\mathrm{proj}_{\mathcal{R}}(q) = p^{W^{\epsilon}}(x_{t_1}, \dots, x_{t_N}|x_0, x_1)q(x_0, x_1).$$



Discrite Markovian Projection.

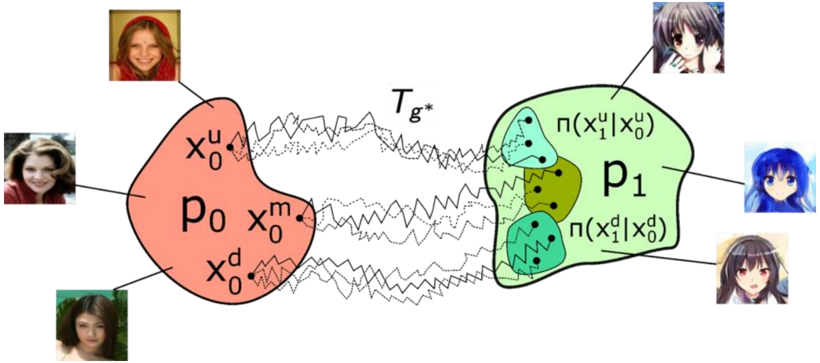
Finds the markovian discrete stochastic process $p_{\mathcal{M}}$ which is the most similar to a process q.

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



Conclusions

1. Schrodinger Bridge (and, more generally, diffusion bridge) framework allows to *construct diffusion processes between arbitrary data distributions* (not just noise to data as in classic score-based diffusion models)
2. The core of the framework is the bridge matching technique which is closely related to the conventional *score matching* from diffusion models
3. Schrodinger Bridges yield *state-of-the-art* results in several image-to-image setups, including image inverse problems (super-resolution, inpainting, etc.)
4. Schrodinger Bridges can be learned via adversarial techniques (**ASBM**, ours) which notably speed up the inference time compared to the diffusion-based approaches



Our related publications (A*)

ICLR International Conference On Learning Representations

NEURAL INFORMATION PROCESSING SYSTEMS

ICML International Conference On Machine Learning

Publications at A* conferences on AI during 2021-2024:

2021-2023

1. Korotin, A., Egiazarian, V., Asadulaev, A., Safin, A., & Burnaev, E. (2020, October). Wasserstein-2 Generative Networks. In1. *International Conference on Learning Representations*.
2. Korotin, A., Li, L., Solomon, J., & Burnaev, E. (2020, October). Continuous Wasserstein-2 Barycenter Estimation without Minimax Optimization. In *International Conference on Learning Representations*.
3. Mokrov, P., Korotin, A., Li, L., Genevay, A., Solomon, J. M., & Burnaev, E. (2021). Large-scale wasserstein gradient flows. *Advances in Neural Information Processing Systems*, *34*, 15243-15256.
4. Korotin, A., Li, L., Genevay, A., Solomon, J. M., Filippov, A., & Burnaev, E. (2021). Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark. *Advances in Neural Information Processing Systems*, *34*, 14593-14605.
5. Barannikov, S., Trofimov, I., Sotnikov, G., Trimbach, E., Korotin, A., Filippov, A., & Burnaev, E. (2021). Manifold Topology Divergence: a Framework for Comparing Data Manifolds. *Advances in Neural Information Processing Systems*, *34*, 7294-7305.
6. Rout, L., Korotin, A., & Burnaev, E. (2021, October). Generative Modeling with Optimal Transport Maps. In *International Conference on Learning Representations*.
7. Korotin, A., Egiazarian, V., Li, L., & Burnaev, E. (2022). Wasserstein iterative networks for barycenter estimation. *Advances in Neural Information Processing Systems*, *35*, 15672-15686.
8. Korotin, A., Kolesov, A., & Burnaev, E. (2022). Kantorovich strikes back! Wasserstein GANs are not optimal transport?. *Advances in Neural Information Processing Systems*, *35*, 13933-13946.
9. **[SPOTLIGHT, TOP 25%]** Korotin, A., Selikhanovych, D., & Burnaev, E. (2023). Neural Optimal Transport. In *The Eleventh International Conference on Learning Representations*.
10. Korotin, A., Selikhanovych, D., & Burnaev, E. (2023). Kernel neural optimal transport. In *The Eleventh International Conference on Learning Representations*.
11. Gazdieva, M., Korotin, A., Selikhanovych, D., & Burnaev, E. (2023). Extremal Domain Translation with Neural Optimal Transport. *Accepted to NeurIPS 2023*
12. Gushchin, N., Kolesov, A., Mokrov, P., Karpikova, P., Spiridonov, A., Burnaev, E., & Korotin, A. (2023). Building the bridge of schrödinger: A continuous entropic optimal transport benchmark. Advances in Neural Information Processing Systems, 36, 18932-18963.
13. **[ORAL, TOP 3%]** Gushchin, N., Kolesov, A., Korotin, A., Vetrov, D. P., & Burnaev, E. (2024). Entropic neural optimal transport via diffusion processes. Advances in Neural Information Processing Systems, 36.

2024

1. Korotin, A., Gushchin, N., & Burnaev, E. (2024). Light Schrödinger Bridge. In *The Twelfth International Conference on Learning Representations*.
2. Mokrov P., Korotin A., Kolesov A., Gushchin N., Burnaev E. (2024) Energy-guided Entropic Neural Optimal Transport. In *The Twelfth International Conference on Learning Representations*.
3. Asadulaev, A., Korotin, A., Egiazarian, V., Mokrov, P., & Burnaev, E. (2024). Neural Optimal Transport with General Cost Functionals. In *The Twelfth International Conference on Learning Representations*.
4. Gushchin, N., Kholkin, S., Burnaev, E., & Korotin, A. (2024, February). Light and Optimal Schrödinger Bridge Matching. In Forty-first International Conference on Machine Learning.
5. Kolesov, A., Mokrov, P., Udovichenko, I., Gazdieva, M., Pammer, G., Burnaev, E., & Korotin, A. Estimating Barycenters of Distributions with Neural Optimal Transport. In Forty-first International Conference on Machine Learning.