Bridge Matching Schrödinger Bridges

MFO Mini-Workshop: Statistical Challenges for Deep Generative Models

Stefano Peluchetti

Sakana AI

The Context

Applications:

- distribution (generative) modeling: $\mathcal{D}^y$
- conditional distribution modeling: $\mathcal{D}^{y|x}$
- alignment of unpaired data: $\mathcal{D}^x$, $\mathcal{D}^y$
- sampling from unnormalized density (not discussed in this talk)

We always assume that samples are available from the target distributions

Stochastic Differential Equations

Diffusion process $X \sim P$ is the d-dimensional Markov process solution to the SDE

$$
\begin{aligned}
dX_t &= \mu(X_t, t)dt + \sigma(t)dW_t, \quad t \in [0, 1] \\
X_0 &\sim P_0,
\end{aligned}
$$

for some drift function $\mu(x, t) : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$ and scalar diffusion $\sigma(t) : [0, 1] \to \mathbb{R}_{\geq 0}$

Differential form is shorthand for

$$X_t = X_0 + \int_0^t \mu(X_s, s)ds + \int_0^t \sigma(s)dW_s$$

Euler-Maruyama discretization is

$$X_{t+\Delta t} \approx X_t + \mu(X_t, t)\Delta t + \sigma(t)\Delta W_t, \quad \Delta W_t := W_{t+\Delta t} - W_t \sim \mathcal{N}(0, \Delta t \mathrm{I}_d),$$

for some step size Δt

Diffusion Time-Reversal

$X \sim P$ diffusion associated to

$$dX_t = \mu(X_t, t)dt + \sigma(t)dW_t, \quad t \in [0, 1],$$
$$X_0 \sim P_0$$

The time reversal (dt is negative, W_t is not the same BM) is still a diffusion with SDE

$$dX_t = -\upsilon(X_t, t)dt + \sigma(t)dW_t, \quad t \in [1, 0],$$
$$-\upsilon(x, t) := \mu(x, t) - \sigma(t)^2 \nabla_x \log p_t(x),$$
$$X_0 \sim P_1,$$

(Anderson, 1982)

Denoising Diffusion (DD) Models

Sohl-Dickstein et al. (2015); Ho et al. (2020); Song et al. (2021)...

Initial target distribution $P_0 = \mathcal{D}^y$

Ergodic / vanishing SNR: $P_{1|0} \approx \mathcal{I}$ for some simple distribution $\mathcal{I}$

Scalable score-matching (Hyvärinen, 2005; Vincent, 2011) to learn $\nabla_x \log p_t(x)$ as

$$\nabla_x \log p_t(x) = \frac{\nabla_x \int p_{t|0}(x|x_0) p_0(x_0) dx_0}{p_t(x)} = \int \nabla_x \log p_{t|0}(x|x_0) \frac{p_{t|0}(x|x_0) p_0(x_0)}{p_t(x)} dx_0$$

$$= \mathop{\mathbb{E}}_P [\nabla_{X_t} \log p_{t|0}(X_t|X_0) | X_t = x]$$

$$= \mathop{\arg\min}_{\alpha(x,t)} \mathop{\mathbb{E}}_P [\| \nabla_{X_t} \log p_{t|0}(X_t|X_0) - \alpha(X_t, t) \|^2]$$

Diffusion Bridges

$X \sim P$ diffusion associated to

$$
\begin{aligned}
&dX_t = \mu(X_t, t)dt + \sigma(t)dW_t, \quad t \in [0, 1],\\
&X_0 \sim P_0
\end{aligned}
$$

Pinning down this diffusion at (x_0, x_1) gives $X \sim P_{|0,1}$

Doob's h-transform: it is still a diffusion associated to

$$
\begin{aligned}
&dX_t = b(X_t, t, x_1)dt + \sigma(t)dW_t, \quad t \in [0, 1],\\
&b(x, t, x_1) := \mu(x, t) + \sigma(t)^2 \nabla_x \log p_{1|t}(x_1|x),\\
&X_0 = x_0
\end{aligned}
$$

Mixtures of Diffusion Bridges

Construct process $X \sim \Pi$ by $\Pi := \Psi_{0,1} P_{|0,1}$

$$\Pi(A) := \int P_{|0,1}(A|x_0, x_1)\Psi_{0,1}(dx_0, dx_1)$$

for some coupling $\Psi_{0,1}$ of two target distributions of interest Ψ_0, Ψ_1

Sample $X \sim \Pi$ by first sampling $(X_0, X_1) \sim \Psi_{0,1}$, then sampling $X_{t\in(0,1)} \sim P_{|0,1}$

Non-Markov diffusion

$$dX_t = [\mu(X_t, t) + \sigma(t)^2 \nabla_x \log p_{1|t}(X_1|X_t)]dt + \sigma(t)dW_t, \quad t \in [0, 1],$$
$$(X_0, X_1) \sim \Psi_{0,1}$$

Markovian Marginal Matching (Gyöngy, 1986)

$X \sim \Pi$ Ito process with differential

$$
\begin{aligned}
dX_t &= \mu(t, \omega)dt + \sigma(t, \omega)dW_t, \quad t \in [0, 1], \\
X_0 &\sim \Pi_0
\end{aligned}
$$

for some nonanticipative processes $\mu(t, \omega), \sigma(t, \omega)$

Consider $X \sim M$ associated to

$$
\begin{aligned}
dX_t &= m(X_t, t)dt + \sigma(X_t, t)dW_t, \quad t \in [0, 1], \\
X_0 &\sim \Pi_0
\end{aligned}
$$

with $m(x, t) := \mathbb{E}_\Pi[\mu(t, \omega)|X_t = x]$, $\sigma(x, t)\sigma(x, t)^\top := \mathbb{E}_\Pi[\sigma(t, \omega)\sigma(t, \omega)^\top|X_t = x]$

Then, $\Pi_t = M_t$, $0 \leq t \leq 1$

Idea

Case $dX_t = \mu^\lambda(X_t, t)dt + \sigma^\lambda(X_t, t)dW_t$, $\Sigma^\lambda(x, t) := \sigma(x, t)\sigma(x, t)^\top$

$$
\begin{aligned}
(\pi_t(x))_t &= \Big(\int_\Lambda p_t^\lambda(x)\mathcal{L}(d\lambda)\Big)_t = \int_\Lambda \Big(p_t^\lambda(x)\Big)_t \mathcal{L}(d\lambda) \\
&= \int_\Lambda \Big(\mu^\lambda(x, t)p_t^\lambda(x)\Big)_x + \frac{1}{2}\Big(\Sigma^\lambda(x, t)p_t^\lambda(x)\Big)_{xx} \mathcal{L}(d\lambda) \\
&= \int_\Lambda \Big(\frac{\mu^\lambda(x, t)p_t^\lambda(x)}{\pi_t(x)}\pi_t(x)\Big)_x + \frac{1}{2}\Big(\frac{\Sigma^\lambda(x, t)p_t^\lambda(x)}{\pi_t(x)}\pi_t(x)\Big)_{xx} \mathcal{L}(d\lambda) \\
&= \Big(\Big[\int_\Lambda \frac{\mu^\lambda(x, t)p_t^\lambda(x)}{\pi_t(x)}\mathcal{L}(d\lambda)\Big]\pi_t(x)\Big)_x + \frac{1}{2}\Big(\Big[\int_\Lambda \frac{\Sigma^\lambda(x, t)p_t^\lambda(x)}{\pi_t(x)}\mathcal{L}(d\lambda)\Big]\pi_t(x)\Big)_{xx}
\end{aligned}
$$

Bridge Matching (BM, Peluchetti (2021))

1. Construct process $\Pi = \Psi_{0,1} P_{|0,1}$
2. Perform Markovian marginal matching of Π

$$dX_t = m(X_t, t)dt + \sigma(t)dW_t, \quad t \in [0, 1],$$
$$X_0 \sim \Psi_0$$

with $m(x, t) := \mathbb{E}_\Pi[b(X_t, t, X_1)|X_t = x] = \arg\min_{\alpha(x,t)} \mathbb{E}_\Pi[\|b(X_t, t, X_1) - \alpha(X_t, t)\|^2]$

Process Π of step 1 is a reciprocal process

Step 2 amounts to a Markovian projection: $M = \arg\min_{Q \in \mathcal{M}} \mathbb{KL}(\Pi \,\|\, Q)$

Example

P given by $X_t = \sigma W_t$: $\mu(x, t) = 0$, $\sigma(t) = \sigma$

$P_{|0,1}$ given by $dX_t = \frac{X_1 - X_t}{1-t} dt + \sigma dW_t$, as $\sigma(t)^2 \nabla_x \log p_{1|t}(x_1|x) = \frac{x_1 - x}{1-t}$

Bridge matching:

$$dX_t = \frac{\mathbb{E}_\Pi[X_1|X_t] - X_t}{1-t} dt + \sigma dW_t, \quad t \in [0, 1],$$
$$X_0 \sim \Psi_0$$

Denoising diffusion: $\mathbb{E}_P[X_1|X_t]$ instead of $\mathbb{E}_\Pi[X_1|X_t]$ (reversed time indexing)

Flow matching (Liu et al., 2022; Liu, 2022) for $\sigma = 0$: $X_t = (1-t)X_0 + tX_1$ gives $P_{|0,1}$, and $m(x, t) = \mathbb{E}_\Pi[X_1 - X_0|X_t]$

Training (Generative Modeling)

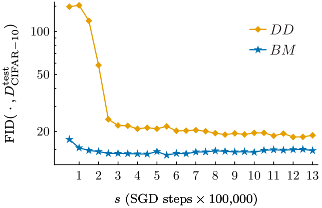
Algorithm Denoising Diffusion

repeat
  $t \sim \mathcal{U}(0, 1)$
  $X_1 \sim \mathcal{D}^y$
  $X_t \sim \mathcal{N}(X_1, (1-t)\sigma^2 \mathrm{I}_d)$
  $Y_t \leftarrow \frac{X_1 - X_t}{1-t}$
  $\mathcal{L} \leftarrow \left\| Y_t - \alpha_\theta(X_t, t) \right\|^2$
  $\theta \leftarrow \texttt{sgdstep}(\theta, \mathcal{L})$
until convergence

Algorithm Bridge Matching

repeat
  $t \sim \mathcal{U}(0, 1)$
  $X_1 \sim \mathcal{D}^y$
  $X_0 \sim \mathcal{N}(0, \mathrm{I}_d)$
  $X_t \sim \mathcal{N}(tX_1 + (1-t)X_0, t(1-t)\sigma^2 \mathrm{I}_d)$
  $Y_t \leftarrow \frac{X_1 - X_t}{1-t}$
  $\mathcal{L} \leftarrow \left\| Y_t - \alpha_\theta(X_t, t) \right\|^2$
  $\theta \leftarrow \texttt{sgdstep}(\theta, \mathcal{L})$
until convergence

Result (Generative Modeling)



Conditional Modeling

BM results updates the input coupling

But sometimes we want to learn a given conditional distribution $\Psi_{1|0} = \mathcal{D}^{y|x}$

Option 1: construct DD and BM conditionally on x (one target distribution simple)

Option 2: construct BM based on $\Pi = \delta_x \otimes \mathcal{D}^{y|x} P_{0,1}$ (one-sided mixing), infer $\mathbb{E}_\Pi[X_1|X_t, X_0]$, see De Bortoli et al. (2023)

Entropic Optimal Transport

We are given two distributions $\Psi_0 = \mathcal{D}^x$ and $\Psi_1 = \mathcal{D}^y$ (samples of), and we want to align the distributions by minimizing a cost function $\kappa(x_0, x_1)$

$$E_{0,1} \coloneqq \underset{C_{0,1} \in \mathcal{C}(\Psi_0, \Psi_1)}{\arg\min} \mathbb{E}[\kappa(X_1, X_0)] - \varepsilon \mathbb{H}(C_{0,1})$$

We learn a coupling $C_{0,1}$, that is a joint (or conditional) distribution with compatible marginal distributions

The scalar $\varepsilon \geq 0$ is the regularization: simpler to solve, better behaved solutions

Schrödinger Bridges

Dynamic Schrödinger bridge for reference diffusion P is

$$S := \operatorname*{arg\,min}_{Q \in \mathcal{M}(\Psi_0, \Psi_1)} \mathbb{KL}(Q \,\|\, P)$$

It holds that $S = S_{0,1} P_{|0,1}$, where $S_{0,1}$ is the static Schrödinger bridge for $P_{0,1}$

$$\begin{aligned} S_{0,1} &= \operatorname*{arg\,min}_{C_{0,1} \in \mathcal{C}(\Psi_0, \Psi_1)} \mathbb{KL}(C_{0,1} \,\|\, P_{0,1}) \\ &= \operatorname*{arg\,min}_{C_{0,1} \in \mathcal{C}(\Psi_0, \Psi_1)} \mathbb{E}_{C_{0,1}} \left[-\log p_{1|0}(X_1 | X_0) \right] - \mathbb{H}(C_{0,1}) \end{aligned}$$

Choosing P s.t. $X_t = \sigma W_t$ yields $S_{0,1} = E_{0,1}$ for $\kappa(x_0, x_1) = \frac{1}{2} \|x_0 - x_1\|^2$ and $\varepsilon = \sigma^2$

S is characterized by: (i) $X_0 \sim \Psi_0$, $X_1 \sim \Psi_1$, (ii) X reciprocal, (iii) X Markov

S is associated to a stochastic optimal control problem

Iterated Proportional Fitting (IPF)

Replace one marginal at a time, keep conditionals fixed

$F_{0,1}^{(0)} \leftarrow \Psi_0 P_{1|0}$

$F_{0,1}^{(1)} \leftarrow \Psi_1 F_{0|1}^{(0)}$

$F_{0,1}^{(2)} \leftarrow \Psi_0 F_{1|0}^{(1)}$

...

Diffusion IPF (Bortoli et al., 2021; Vargas et al., 2021) extends IPF to dynamic setting

- each iteration performs a diffusion time reversal
- inference-simulation mismatch
- forgetting P
- convergence result

Iterated Bridge Matching (IBM, Peluchetti (2023); Shi et al. (2023))

Iteratively initialize BM with the coupling from the previous iteration

$C_{0,1}^{(0)} \leftarrow \Psi_0 \otimes \Psi_1$

$\Pi^{(1)} \leftarrow C_{0,1}^{(0)} P_{|0,1}$

$M^{(1)} \leftarrow \mathrm{BM}(\Pi^{(1)})$

$C_{0,1}^{(1)} \leftarrow M_{0,1}^{(1)}$

$\Pi^{(2)} \leftarrow C_{0,1}^{(1)} P_{|0,1}$

$M^{(2)} \leftarrow \mathrm{BM}(\Pi^{(2)})$

...

Properties

Valid coupling at each step

No simulation-inference mismatch

Sampling $C_{0,1}^{(i)}$ for $i \geq 1$ is expensive, but caching is more efficient

Each step can be solved forward or backward in time, alternating reduces accumulation of errors

Convergence Result (Sketch)

$$\mathbb{KL}(\Pi(C) \,\|\, P) = \mathbb{KL}(\Pi(C) \,\|\, M(C)) + \mathbb{KL}(M(C) \,\|\, P),\, P \in \mathcal{M}$$

$$\Rightarrow \mathbb{KL}(\Pi^{(i)} \,\|\, S) \geq \mathbb{KL}(M^{(i)} \,\|\, S)$$

$$\mathbb{KL}(\Pi^{(i)} \,\|\, S) = \mathbb{KL}(\Pi^{(i)}_{0,1} \,\|\, S_{0,1}) + \mathbb{E}_{\Pi^{(i)}_{0,1}}[\mathbb{KL}(\Pi^{(i)}_{|0,1} \,\|\, S_{|0,1})] = \mathbb{KL}(\Pi^{(i)}_{0,1} \,\|\, S_{0,1})$$

$$\mathbb{KL}(M^{(i)} \,\|\, S) \geq \mathbb{KL}(M^{(i)}_{0,1} \,\|\, S_{0,1}) = \mathbb{KL}(\Pi^{(i+1)}_{0,1} \,\|\, S_{0,1})$$

$$\Rightarrow \mathbb{KL}(M^{(i)} \,\|\, S) \geq \mathbb{KL}(\Pi^{(i+1)} \,\|\, S)$$

$$\mathbb{KL}(\Pi^{(i)} \,\|\, S) \geq \mathbb{KL}(M^{(i)} \,\|\, S) \geq \mathbb{KL}(\Pi^{(i+1)} \,\|\, S)$$

$$\lim_{i\to\infty} \mathbb{KL}(\Pi^{(i)} \,\|\, S) - \mathbb{KL}(M^{(i)} \,\|\, S) = \lim_{i\to\infty} \mathbb{KL}(\Pi^{(i)} \,\|\, M^{(i)}) = 0$$

$$\Pi^{(l)} \xrightarrow{\mathcal{L}} \Pi^{(\infty)},\, M^{(l)} \xrightarrow{\mathcal{L}} M^{(\infty)}$$

$$0 = \liminf_{l\to\infty} \mathbb{KL}(\Pi^{(l)} \,\|\, M^{(l)}) \geq \mathbb{KL}(\Pi^{(\infty)} \,\|\, M^{(\infty)})$$

Gaussian Experiment — Part I

For $d \geq 1$, $\Psi_0 = \mathcal{N}_d(\mu_0, \Sigma_0)$, $\Upsilon_1 = \mathcal{N}_d(\mu_1, \Sigma_1)$, and P with $X_t = \sigma dW_t$

$$S_{0,1} = \mathcal{N}_{2d}\left(\begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}, \begin{bmatrix} \Sigma_0 & \Sigma_S(\sigma) \\ \Sigma_S(\sigma)^\mathsf{T} & \Sigma_1 \end{bmatrix} \right),$$

where $\Sigma_S(\sigma) := (\Sigma_0\Sigma_1 + \frac{\sigma^4}{4}\mathrm{I}_d)^{1/2} - \frac{\sigma^2}{2}\mathrm{I}_d$

Gaussian Experiment — Part II

Consider Gaussian coupling $C_{0,1} \in \mathcal{C}(\Psi_0, \Psi_1)$ parametrized by Σ_C

$\Pi_{0,t,1}$ is Gaussian and

$$\mathbb{E}_{\Pi}[X_1|X_t] = \mu_1 + \Sigma_{\Pi;t,1}^{\top}\Sigma_{\Pi;t,t}^{-1}(X_t - \mu_t),$$

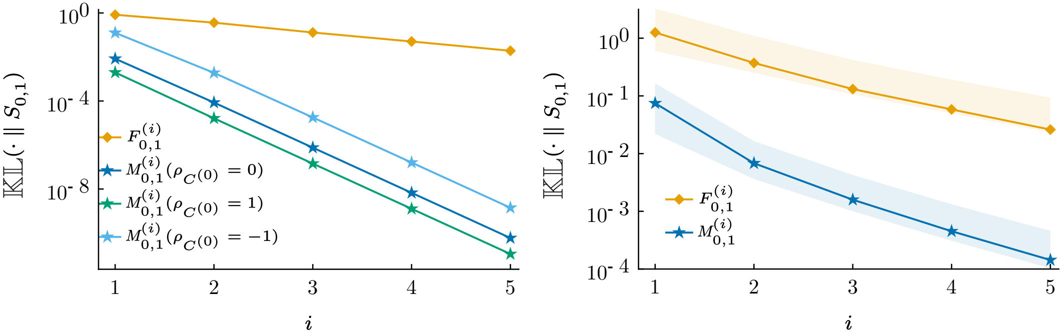
with $\mu_t := (1-t)\mu_0 + t\mu_1$, $\Sigma_{\Pi;t,1} := (1-t)\Sigma_C + t\Sigma_1$, and
$\Sigma_{\Pi;t,t} := (1-t)^2\Sigma_0 + t^2\Sigma_1 + t(1-t)(\Sigma_C + \Sigma_C^{\top} + \sigma^2 \mathrm{I}_d)$

BM SDE of the form $dX_t = (A_t X_t + b_t)dt + \sigma dW_t$: Gaussian transitions

$M_{0,1}$ is Gaussian parametrized by $\Sigma_{C'}$, update $\Sigma_C \to \Sigma_{C'} := \Sigma_0 F_1^{\top}$ where F_t solves the matrix-valued ODE

$$dF_t = A_t F_t, \quad t \in [0,1],$$
$$F_0 = \mathrm{I}_d$$

Gaussian Experiment — Part III



Coupled Bridge Matching (BM^2 Peluchetti (2024); De Bortoli et al. (2024))

Can we do away with iterations?

For simplicity, consider the reference process $X_t = \sigma W_t$

Define forward and backward SDEs

$$X_0 \sim \Psi_0, \quad dX_t = \mu_f(X_t, t, \theta)dt + \sigma dW_t, \quad t \in [0, 1] \qquad (F(\theta))$$

$$X_1 \sim \Psi_1, \quad dX_t = -\upsilon_b(X_t, t, \theta)dt + \sigma dW_t, \quad t \in [1, 0] \qquad (B(\theta))$$

Learn each SDE by performing BM based on the coupling from the other SDE

S is a fixed point of the updates defined by $F' \leftarrow \mathrm{BM}(B_{0,1})$ and $B' \leftarrow \mathrm{BM}(F_{0,1})$

Just matching the two SDEs *does not* result in a Schrödinger bridge

Coupled BM

Define the losses

$$\mathbb{L}_f(\theta;\theta') := \mathbb{E}_{\Pi^{B_{0,1}(\theta')}}\left[\frac{1}{2}\int_0^1 \left\|\frac{X_1 - X_t}{1-t} - \mu_f(X_t, t, \theta)\right\|^2 dt\right]$$
$$\mathbb{L}_b(\theta;\theta') := \mathbb{E}_{\Pi^{F_{0,1}(\theta')}}\left[\frac{1}{2}\int_0^1 \left\|\frac{X_0 - X_t}{t} - v_b(X_t, t, \theta)\right\|^2 dt\right]$$
$$\mathbb{L}(\theta;\theta') := \mathbb{L}_f(\theta;\theta') + \mathbb{L}_b(\theta;\theta')$$

Take a gradient descent step on $\mathbb{L}(\theta;\theta')$ in θ only, with $\theta' = \theta$

Training

Algorithm BM^2 — training loss computation

$f_0 \sim \Psi_0$ ▷ Marginal sampling

$f_{\Delta t}, \ldots, f_1 | f_0 \sim \mathrm{sg}(\mathrm{discretize}(f_0, \Delta t, \mu_f(\cdot, \cdot, \theta)))$ ▷ Discretization

$b_1 \sim \Psi_1$ ▷ Marginal sampling

$b_{1-\Delta t}, \ldots, b_0 | b_1 \sim \mathrm{sg}(\mathrm{discretize}(b_1, \Delta t, \upsilon_b(\cdot, \cdot, \theta)))$ ▷ Discretization

$t \sim \mathcal{U}(0, 1)$ ▷ Time sampling

$pf_t \sim P_{t|0,1}(\cdot | f_0, f_1)$ ▷ Bridge sampling

$pb_t \sim P_{t|0,1}(\cdot | b_0, b_1)$ ▷ Bridge sampling

$l_f(\theta) \leftarrow \frac{1}{2} \| \frac{b_1 - pb_t}{1 - t} - \mu_f(pb_t, t, \theta) \|^2$ ▷ BM based on $B_{0,1}$

$l_b(\theta) \leftarrow \frac{1}{2} \| \frac{f_0 - pf_t}{t} - \upsilon_b(pf_t, t, \theta) \|^2$ ▷ BM based on $F_{0,1}$

$l(\theta) \leftarrow l_f(\theta) + l_b(\theta)$

Properties

Both Diffusion IPF and IBM can be cast as the complete-minimization version of BM^2 with specific drift initializations

First learn BM in both time directions due to simulation-inference mismatch

A single neural network suffices

Simpler and more computationally efficient implementation

Limitations:

- somewhat computationally intensive (still)
- case $\sigma \approx 0$ is problematic
- this (efficient) version limited to simple dynamics / costs

Results

Mixture of Gaussians benchmark

Monte Carlo estimate of $\mathbb{KL}(S \,\|\, P)$ as function of $\boldsymbol{\varepsilon}$ and d, standard deviation in gray:

	$\varepsilon=0.1$				$\varepsilon=1$				$\varepsilon=10$			
Method	$d=2$	$d=16$	$d=64$	$d=128$	$d=2$	$d=16$	$d=64$	$d=128$	$d=2$	$d=16$	$d=64$	$d=128$
BM^2	**0.01** **0.01**	**0.20** **0.02**	**1.03** **0.07**	**3.06** **0.16**	**0.01** **0.00**	**0.11** **0.00**	**1.43** **0.03**	8.29 0.36	**0.11** **0.01**	**2.25** **0.04**	**13.13** **0.13**	**40.46** **0.49**
IBM	0.03 0.01	0.20 0.02	1.24 0.04	5.70 0.42	0.01 0.00	0.16 0.01	1.94 0.04	**7.79** **0.07**	0.16 0.00	4.09 0.03	17.17 0.21	49.17 0.55
DIPF	0.59 0.14	2.39 0.05	7.93 1.23	34.77 0.82	0.23 0.06	1.21 0.18	13.13 0.79	36.51 1.05	0.81 0.06	28.25 2.12	113.8 7.2	345.8 8.1

sakana.ai

Q & A

References i

Anderson, B. D. (1982). Reverse-Time Diffusion Equation Models. *Stochastic Processes and their Applications*, 12(3):313–326.

Bortoli, V. D., Thornton, J., Heng, J., and Doucet, A. (2021). Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling. In *Thirty-Fifth Conference on Neural Information Processing Systems*.

De Bortoli, V., Korshunova, I., Mnih, A., and Doucet, A. (2024). Schr\"odinger Bridge Flow for Unpaired Data Translation.

De Bortoli, V., Liu, G.-H., Chen, T., Theodorou, E. A., and Nie, W. (2023). Augmented Bridge Matching.

References ii

Gyöngy, I. (1986). Mimicking the one-dimensional marginal distributions of processes having an ito differential. *Probability Theory and Related Fields*, 71(4):501–516.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851.

Hyvärinen, A. (2005). Estimation of Non-Normalized Statistical Models by Score Matching. *Journal of Machine Learning Research*, 6(24):695–709.

Liu, Q. (2022). Rectified Flow: A Marginal Preserving Approach to Optimal Transport.

References iii

Liu, X., Gong, C., and Liu, Q. (2022). Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow.

Peluchetti, S. (2021). Non-Denoising Forward-Time Diffusions.

Peluchetti, S. (2023). Diffusion Bridge Mixture Transports, Schrödinger Bridge Problems and Generative Modeling. *Journal of Machine Learning Research*, 24(374):1–51.

Peluchetti, S. (2024). BM^2: Coupled Schrödinger Bridge Matching. *Transactions on Machine Learning Research*.

Shi, Y., De Bortoli, V., Campbell, A., and Doucet, A. (2023). Diffusion Schrödinger Bridge Matching.

References iv

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. (2015). Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265. PMLR.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2021). Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*.

Vargas, F., Thodoroff, P., Lamacraft, A., and Lawrence, N. (2021). Solving Schrödinger Bridges via Maximum Likelihood. *Entropy*, 23(9):1134.

Vincent, P. (2011). A Connection Between Score Matching and Denoising Autoencoders. *Neural Computation*, 23(7):1661–1674.