Diffusion Bridge Mixture Transports, Schrödinger Bridge Problems and Generative Modeling

ICIAM 2023

Theoretical and Computational Advances in Measure Transport

Stefano Peluchetti

August 21, 2023

Cogent Labs

Context

Given Γ, Υ two distributions on $\mathbb{R}^d$, find a transport between them:

- static: realized by any random variable with prescribed marginal distributions;
- dynamic: realized by any stochastic process with prescribed initial and terminal distributions;
- optimal in some sense;
- assume samples from Γ and Υ are available.

Applications:

- generative modeling;
- domain transfer.

Some Notation

$\mathcal{P}_n$: probability measures (PM)s over $\mathbb{R}^{d \times n}$.

$\mathcal{P}_{\mathcal{C}}$: PMs over $\mathcal{C}([0, \tau], \mathbb{R}^d)$.

$\mathcal{P}_2(\Gamma, \Upsilon)$: PMs with marginal distributions Γ, Υ; $\mathcal{P}_2(\Gamma, \Upsilon) \subseteq \mathcal{P}_2$.

$\mathcal{P}_{\mathcal{C}}(\Gamma, \Upsilon)$: PMs with initial-terminal distributions Γ, Υ; $\mathcal{P}_{\mathcal{C}}(\Gamma, \Upsilon) \subseteq \mathcal{P}_{\mathcal{C}}$.

$P_{t_1,\ldots,t_n}$: marginalization of $P \in \mathcal{P}_{\mathcal{C}}$ at $t_1, \ldots, t_n$.

$P_{\bullet|t_1,\ldots,t_n}$: conditioning of $P \in \mathcal{P}_{\mathcal{C}}$ given values at $t_1, \ldots, t_n$.

p instead of P to denote its (Lebesgue) density.

$Q = \Psi P_{\bullet|t_1,\ldots,t_n}$: mixing of $P \in \mathcal{P}_{\mathcal{C}}$ via $\Psi \in \mathcal{P}_n$ at times $t_1, \ldots, t_n$; $Q \in \mathcal{P}_{\mathcal{C}}$.

Schrödinger Bridge Problems

Dynamic Schrödinger bridge problem $(\mathrm{S}_{\mathrm{dyn}})$, reference $R \in \mathcal{P}_{\mathcal{C}}$:

$$S^*(\Gamma, \Upsilon, R, \mathcal{P}_{\mathcal{C}}) := \underset{S \in \mathcal{P}_{\mathcal{C}}(\Gamma, \Upsilon)}{\arg\min} \; D_{KL}(S \,\|\, R).$$

Static Schrödinger bridge problem $(\mathrm{S}_{\mathrm{sta}})$, reference $B \in \mathcal{P}_2$:

$$S^*(\Gamma, \Upsilon, B, \mathcal{P}_2) := \underset{C \in \mathcal{P}_2(\Gamma, \Upsilon)}{\arg\min} \; D_{KL}(C \,\|\, B).$$

Half bridge problems, reference $Q \in \mathcal{P}$:

$$S^*(\Gamma, \cdot, Q, \mathcal{P}) := \underset{H \in \mathcal{P}(\Gamma, \cdot)}{\arg\min} \; D_{KL}(H \,\|\, Q),$$

$$S^*(\cdot, \Upsilon, Q, \mathcal{P}) := \underset{H \in \mathcal{P}(\cdot, \Upsilon)}{\arg\min} \; D_{KL}(H \,\|\, Q).$$

Partial Solutions

$(\mathrm{S}_{\mathrm{sta}}) \to (\mathrm{S}_{\mathrm{dyn}})$: $S^*(\Gamma, \Upsilon, R, \mathcal{P}_{\mathcal{C}}) = S^*(\Gamma, \Upsilon, R_{0,\tau}, \mathcal{P}_2) R_{\bullet|0,\tau}$.

$(\mathrm{S}_{\mathrm{dyn}}) \to (\mathrm{S}_{\mathrm{sta}})$: $S^*(\Gamma, \Upsilon, R, \mathcal{P}_2) = S^*_{0,\tau}(\Gamma, \Upsilon, R, \mathcal{P}_{\mathcal{C}})$.

Half bridge solutions: $S^*(\Gamma, \cdot, Q, \mathcal{P}) = \Gamma Q_{\bullet|0}$ and $S^*(\cdot, \Upsilon, Q, \mathcal{P}) = \Upsilon Q_{\bullet|\tau}$.

Reference SDE

R solution to

$$dX_t = \mu_R(X_t, t)dt + \sigma_R(t)dW_t, \quad t \in [0, \tau],$$
$$X_0 \sim \Gamma.$$

$R_{\bullet|0,\tau}$ is a $(0, x_0, \tau, x_\tau)$ diffusion bridge, solves

$$dX_t = b_R(X_t, t)dt + \sigma_R(t)dW_t, \quad t \in [0, \tau],$$
$$b_R(x, t) := \mu_R(x, t) + \Sigma_R(t)\nabla_x \log r_{\tau|t}(x_\tau|x),$$
$$X_0 = x_0,$$

$\Sigma_R(t) := \sigma_R(t)\sigma_R(t)^\top$.

IPF Algorithm

$$F^{(0)} = R = S^*(\Gamma, \cdot, R, \mathcal{P}_{\mathcal{C}}), F^{(1)} = S^*(\cdot, \Upsilon, F^{(0)}, \mathcal{P}_{\mathcal{C}}), F^{(2)} = S^*(\Gamma, \cdot, F^{(1)}, \mathcal{P}_{\mathcal{C}}), \ldots$$

Algorithm IPF

Input: $\Gamma, \Upsilon, R_{\bullet|0}, n$

Output: $\{F^{(i)}\}_{i=1}^{n}$

$F^{(0)} \leftarrow \Gamma R_{\bullet|0}$

for $i = 1, \ldots, n$ **do**

  if i is even **then**

    $F^{(i)} \leftarrow \Gamma F_{\bullet|0}^{(i-1)}$ ▷ forward IPF

  else

    $F^{(i)} \leftarrow \Upsilon F_{\bullet|\tau}^{(i-1)}$ ▷ backward IPF

  end if

end for

Diffusion Time Reversal

Let $X \sim P$ be the diffusion associated to

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

Under suitable conditions, the time reversed process $\overline{X}_t := X_{\tau-t} \sim \overline{P}$ is still a diffusion process, associated to

$$d\overline{X}_t = \upsilon(\overline{X}_t, t)dt + \sigma_R(\mathfrak{t})dW_t, \quad t \in [0, \tau],$$
$$\upsilon(x, t) := -\mu(x, \mathfrak{t}) + \Sigma_R(\mathfrak{t})\nabla_x \log p_{\mathfrak{t}}(x),$$
$$\overline{X}_0 \sim P_\tau,$$

where $\mathfrak{t} := \tau - t$ (Anderson, 1982).

Diffusion IPF

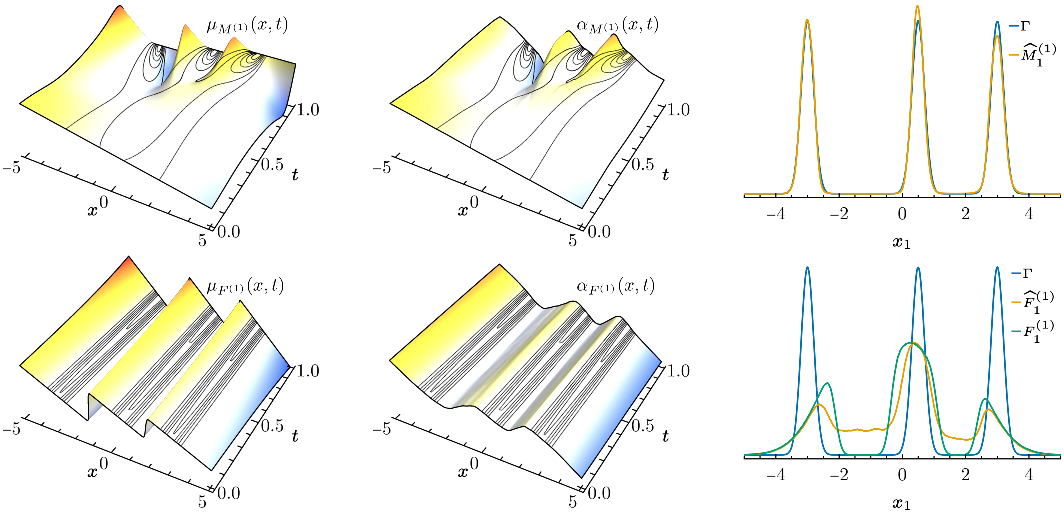
Introduced in Bortoli et al. (2021); Vargas et al. (2021).

Define backward $F^{(i)}$ on reverse timescale $\Rightarrow$ each $F^{(i)} = \{\Gamma, \Upsilon\}\overline{F}^{(i-1)}_{\bullet|0}$.

All $F^{(i)}$ and $\overline{F}^{(i)}$ are diffusions.

For each $i \geq 1$, generate samples from $\overline{F}^{(i-1)}$ (i.e. from $F^{(i-1)}$) to infer the drift of $F^{(i)}$.

Simulation-Inference Mismatch



Score-based Generative Modeling

See Ho et al. (2020); Song et al. (2021), and too many others...

Employs ergodic (or vanishing SNR) $X \sim R = F^{(0)}$.

Scalable score-matching (Hyvärinen, 2005; Vincent, 2011) to infer $\overline{X} \sim \overline{F}^{(0)}$:

$$\nabla_x \log r_t(x) = \frac{\nabla_x \int r_{t|0}(x|x_0) r_0(x_0)(dx_0)}{r_t(x)} = \int \nabla_x \log r_{t|0}(x|x_0) \frac{r_{t|0}(x|x_0) r_0(x_0)}{r_t(x)}(dx_0)$$
$$= \mathop{\mathbb{E}}_R[\nabla_{X_t} \log r_{t|0}(X_t|X_0) | X_t = x].$$

Almost same as $F^{(1)}$ in diffusion IPF, for a trivial SB problem.

Diffusion Mixture Matching

Let $X^\lambda \sim P^\lambda$ be a family of diffusions indexed by $\lambda \in \Lambda$, associated to

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

For a mixing PM Ψ on Λ, let $\Pi \in \mathcal{P}_{\mathcal{C}}$ be obtained by taking the Ψ-mixture of P^λ over $\lambda \in \Lambda$. Under suitable conditions, $X \sim P$, with $X_0 \sim \Pi_0$ associated to

$$dX_t = \mu(X_t, t)dt + \sigma(t)dW_t, \quad t \in [0, \tau],$$
$$\mu(x, t) := \int \mu^\lambda(x, t)\omega(\lambda)d\lambda,$$
$$\sigma(t) := \int \sigma^\lambda(t)\omega(\lambda)d\lambda,$$

satisfies $P_t = \Pi_t$, where $\omega(\lambda) := \frac{p_t^\lambda(x)\psi(\lambda)}{\pi_t(x)}$ and $\pi_t := \int p_t^\lambda(x)\psi(\lambda)d\lambda$.

DBM Transport i

(1) For $C \in \mathcal{P}_2(\Gamma, \Upsilon)$, consider $\Pi := CR_{\bullet|0,\tau} \in \mathcal{P}_{\mathcal{C}}(\Gamma, \Upsilon)$.

(2) Match marginal distributions with a diffusion:

$$\begin{aligned}
\lambda &= (x_0, x_\tau),\\
\Psi(d\lambda) &= C(dx_0, dx_\tau) = \Pi_{0,\tau}(dx_0, dx_\tau),\\
\mu^\lambda(x,t) &= \mu_R(x,t) + \Sigma_R(t)\,\nabla_x \log r_{\tau|t}(x_\tau|x),\\
p_t^\lambda(x) &= r_{t|0,\tau}(x|x_0, x_\tau),
\end{aligned}$$

$$\mu_M(x,t) = \mu_R(x,t) + \Sigma_R(t) \underbrace{\int \nabla_x \log r_{\tau|t}(x_\tau|x_t) \frac{\pi_{t|0,\tau}(x|x_0,x_\tau)\pi_{0,\tau}(x_0,x_\tau)}{\pi_t(x_t)}(dx_0, dx_\tau)}_{\int \nabla_x \log r_{\tau|t}(x_\tau|x)\pi_{\tau|t}(x_\tau|x)(dx_\tau)}$$

DBM Transport ii

$X \sim M$ associated to

$$dX_t = \mu_M(X_t, t)dt + \sigma_R(t)dW_t, \quad t \in [0, \tau],$$
$$\mu_M(x, t) = \mu_R(x, t) + \Sigma_R(t) \mathop{\mathbb{E}}_{\Pi}[\nabla_{X_t} \log r_{\tau|t}(X_\tau|X_t)|X_t = x],$$
$$X_0 \sim \Gamma,$$

satisfies $M_t = \Pi_t$.

Drift of M inferred via MSE regression based on samples from Π:

$$\mathop{\mathbb{E}}_{\Pi}[\Sigma_R(t)\nabla_{X_t} \log r_{\tau|t}(X_\tau|X_t)|X_t = x] = \mathop{\arg\min}_{\alpha(x,t)} \mathbb{L}_{\mathrm{DBM}}(\alpha, \Pi),$$
$$\mathbb{L}_{\mathrm{DBM}}(\alpha, \Pi) := \mathop{\mathbb{E}}_{t \sim \mathcal{U}(0,\tau)}[\varrho_t \mathop{\mathbb{E}}_{\Pi}[\|\alpha(X_t, t) - \Sigma_R(t)\nabla_{X_t} \log r_{\tau|t}(X_\tau|X_t)\|^2]].$$

There is a corresponding backward BDBM transport ($t \leftarrow \mathfrak{r} := \tau - t$).

DBM Transport Example

R associated to $dX_t = \sigma dW_t$ over [0, 1], with $\sigma \geq 0$.

$\Sigma_R(t)\,\nabla_{X_t} \log r_{1|t}(X_1|X_t) = \frac{X_1 - X_t}{1-t}$.

Same form for DBM ($M_0 = \Gamma, \Pi = \Pi(\Gamma, \Upsilon)$) and BDBM ($M_0 = \Upsilon, \Pi = \Pi(\Upsilon, \Gamma)$):

$$dX_t = \frac{\mathbb{E}_\Pi[X_1|X_t] - X_t}{1-t} dt + \sigma dW_t, \quad t \in [0, 1],$$
$$X_0 \sim M_0.$$

$\sigma \to 0$ results in Rectified Flow (Liu et al., 2022; Liu, 2022): $R_{t|0,1}$ given by the deterministic linear interpolant $X_t = (1-t)X_0 + tX_1$, and $\mu_M(x, t) = \mathbb{E}_\Pi[X_1 - X_0|X_t]$.

Practical Implementation

Algorithm SGM training

Input: Γ, $R_{t|0}$, $\nabla_y \log r_{t|0}(y, x)$, $\alpha_\theta(x, t)$
Output: $\alpha_{\mathrm{SGM}}(x, t)$

repeat
  $t \sim \mathcal{U}(0, \tau)$
  $X_0 \sim \Gamma$
  $X_t \sim R_{t|0}(\cdot|X_0)$
  $Y_t \leftarrow \nabla_{X_t} \log r_{t|0}(X_t|X_0)$
  $\mathcal{L} \leftarrow \left\| Y_t - \alpha_\theta(X_t, t) \right\|^2 \lambda_t$
  $\theta \leftarrow \texttt{sgdstep}(\theta, \mathcal{L})$
until convergence

Algorithm BDBM training

Input: Γ, Υ, $R_{t|0,\tau}$, $\nabla_y \log r_{t|0}(y, x)$, $\alpha_\theta(x, t)$
Output: $\alpha_{\mathrm{BDBM}}(x, t)$

repeat
  $t \sim \mathcal{U}(0, \tau)$
  $X_0 \sim \Gamma$
  $X_\tau \sim \Upsilon$
  $X_t \sim R_{t|0,\tau}(\cdot|X_0, X_\tau)$
  $Y_t \leftarrow \nabla_{X_t} \log r_{t|0}(X_t|X_0)$
  $\mathcal{L} \leftarrow \left\| Y_t - \alpha_\theta(X_t, t) \right\|^2 \lambda_t$
  $\theta \leftarrow \texttt{sgdstep}(\theta, \mathcal{L})$
until convergence

A Small Scale Result

Iterated DBM

Algorithm IDBM

Input: $\Gamma, \Upsilon, R_{\bullet|0,\tau}, C^{(0)}, n$

Output: $\{M^{(i)}\}_{i=1}^{n}$

for $i = 1, \ldots, n$ **do**
  $\Pi^{(i)} \leftarrow \Pi(C^{(i-1)}, R_{\bullet|0,\tau})$
  $M^{(i)} \leftarrow M(\Pi^{(i)})$
  $C^{(i)} \leftarrow M_{0,\tau}^{(i)}$
end for

Algorithm IPF

Input: $\Gamma, \Upsilon, R_{\bullet|0}, n$

Output: $\{F^{(i)}\}_{i=1}^{n}$

$F^{(0)} \leftarrow \Gamma R_{\bullet|0}$
for $i = 1, \ldots, n$ **do**
  if i is even **then**
    $F^{(i)} \leftarrow \Gamma F_{\bullet|0}^{(i-1)}$ ▷ forward IPF
  else
    $F^{(i)} \leftarrow \Upsilon F_{\bullet|\tau}^{(i-1)}$ ▷ backward IPF
  end if
end for

Considerations

Valid coupling at each step i.

No simulation-inference mismatch.

Initial coupling $C^{(0)}$ usually product measure.

Sampling $C^{(i)}$ is expensive, but samples can be reused to sample $\Pi^{(i+1)} = C^{(i)} R_{t|0,\tau}$.

Each step i can be solved forward or backward in time, best to alternate.

Many implementation "details".

Gaussian Experiment i

For $d \geq 1$, $\Gamma = \mathcal{N}_d(\mu_0, \Sigma_0)$, $\Upsilon = \mathcal{N}_d(\mu_1, \Sigma_1)$, and R is associated to $dX_t = \sigma dW_t$ over $[0, 1]$, with $\sigma \geq 0$.

Solution to $(\mathrm{S}_{\mathrm{sta}})$, or Euclidean EOT:

$$S^*_{0,1}(\Gamma, \Upsilon, R, \mathcal{P}_\mathcal{C}) = \mathcal{N}_{2d}\left(\begin{bmatrix}\mu_0 \\ \mu_1\end{bmatrix}, \begin{bmatrix}\Sigma_0 & \Sigma_S(\sigma) \\ \Sigma_S(\sigma)^\intercal & \Sigma_1\end{bmatrix}\right),$$

where $\Sigma_S(\sigma) \coloneqq (\Sigma_0\Sigma_1 + \frac{\sigma^4}{4}I)^{1/2} - \frac{\sigma^2}{2}I$.

Gaussian Experiment ii

Consider Gaussian coupling $C \in \mathcal{P}_2(\Gamma, \Upsilon)$ parametrized by Σ_C.

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

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

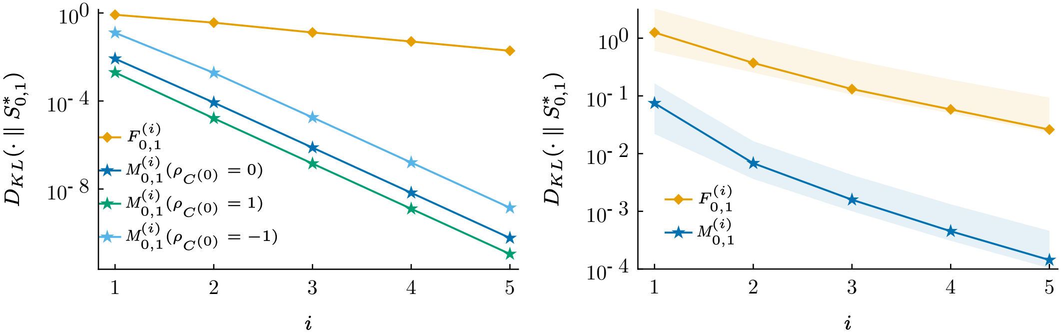
with $\mu_t := (1-t)\mu_0 + t\mu_1$, $\Sigma_{\Pi;t,1} := (1-t)\Sigma_C + t\Sigma_1$, and
$\Sigma_{\Pi;t,t} := (1-t)^2\Sigma_0 + t^2\Sigma_1 + t(1-t)(\Sigma_C + \Sigma_C^{\top} + \sigma^2 I)$.

DBM follows the TI linear SDE $dX_t = (A_tX_t + b_t)dt + \sigma dW_t$ with Gaussian transition probabilities.

Then, $M_{0,1}$ is Gaussian parametrized by $\Sigma_{C'}$, corresponding to the update
$\Sigma_C \to \Sigma_{C'} := \Sigma_0 P_1^{\top}$ where P_t solves the matrix-valued ODE

$$dP_t = A_tP_t, \quad t \in [0,1],$$
$$P_0 = I.$$

Gaussian Experiment iii



An Initial Convergence Result

Assume:

(i) $D_{KL}(C^{(0)} \,\|\, S^*_{0,\tau}) < \infty$;

(ii) for each $i \geq 1$ the Cameron-Martin-Girsanov theorem holds yielding $dM^{(i)}/dS^*$;

(iii) $\sigma_R(t) = I$.

Then:

(i) $\Pi^{(i)} \xrightarrow{\mathcal{L}} S^*$ and $M^{(i)} \xrightarrow{\mathcal{L}} S^*$ as $i \to \infty$;

(ii) $D_{KL}(\Pi^{(i)} \,\|\, S^*) \geq D_{KL}(M^{(i)} \,\|\, S^*) \geq D_{KL}(\Pi^{(i+1)} \,\|\, S^*)$ for $i \geq 1$;

(iii) $D_{KL}(\Pi(C) \,\|\, S^*) = D_{KL}(M(C) \,\|\, S^*)$ if and only if $\Pi(C) = M(C) = S^*$.

Conclusions

Theoretical developments largely incomplete.

Valid coupling at every step.

Ideally suited to simple reference dynamics, can in principle be extended to complex reference dynamics.

BDBM almost plug-in to popular diffusion probabilistic models.

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*.

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.

References ii

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

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

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.