Theoretical guarantees in KL divergence for Diffusion Flow Matching

Gentiloni-Silveri M., Conforti G., Durmus A.

38th Conference on Neural Information Processing Systems

Diffusion Flow Matching (DFM) in Brief

Goal:

- Generate new data $x \sim \nu^\star \in \mathcal{P}(\mathbb{R}^d)$ by learning from existing ones and leveraging a base distribution $\mu \in \mathcal{P}(\mathbb{R}^d)$.

Diffusion Flow Matching (DFM) in Brief

Goal:

- Generate new data $x \sim \nu^\star \in \mathcal{P}(\mathbb{R}^d)$ by learning from existing ones and leveraging a base distribution $\mu \in \mathcal{P}(\mathbb{R}^d)$.

Strategy:

- Build a **Stochastic Interpolant** to interpolate between $\nu^\star$ and μ;
- Build a **Markovian Approximation** to simplify the structure.

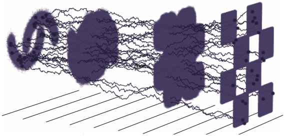

Figure: Figure 1 in (Alberto et al., 2023)

Albergo, Michael S and Boffi, Nicholas M and Vanden-Eijnden, Eric (2023) Stochastic interpolants: A unifying framework for flows and diffusions. In *arXiv preprint arXiv:2303.08797*.

Stochastic Interpolant

Definition:
The stochastic interpolant between μ and $\nu^\star$ is process $(X_t^{\mathrm{I}})_{t\in[0,1]}$ s.t.

$$(X_0^{\mathrm{I}}, X_1^{\mathrm{I}}) \sim \pi \in \Pi(\mu, \nu^\star)\ ,\ (X_t^{\mathrm{I}})_{t\in[0,1]} | (X_0^{\mathrm{I}}, X_1^{\mathrm{I}}) \sim \mathrm{b}\mathbb{B}((X_0^{\mathrm{I}}, X_1^{\mathrm{I}}), \cdot)\ ,$$

with $\Pi(\mu, \nu^\star)$ set of couplings between μ and $\nu^\star$ and $\mathrm{b}\mathbb{B}((x_0, x_1), \cdot)$ Brownian bridge between $x_0, x_1 \in \mathbb{R}^d$.

Remark:
It evolves accordingly to

$$\mathrm{d}X_t^{\mathrm{I}} = 2\nabla \log p_{1-t}(X_1^{\mathrm{I}} | X_t^{\mathrm{I}})\mathrm{d}t + \sqrt{2}\mathrm{d}B_t\ ,\ t \in [0,1]\ ,\ X_0^{\mathrm{I}} \sim \mu\ ,$$

with $(s, x, y) \mapsto p_s(x|y)$ heat kernel.

Markovian Projection

Definition:
The Markovian projection of the stochastic interpolant is the Markovian process $(X_t^{\mathrm{M}})_{t\in[0,1]}$ such that

$$(X_t^{\mathrm{M}})_{t\in[0,1]} \quad : \; X_t^{\mathrm{M}} \overset{\text{dist}}{=} X_t^{\mathrm{I}} \, , \, \forall \, t \in [0,1] \, .$$

Markovian Projection

Definition:
The Markovian projection of the stochastic interpolant is the Markovian process $(X_t^{\mathrm{M}})_{t\in[0,1]}$ such that

$$(X_t^{\mathrm{M}})_{t\in[0,1]} \quad : \; X_t^{\mathrm{M}} \overset{\text{dist}}{=} X_t^{\mathrm{I}} \, , \; \forall \; t \in [0,1] \, .$$

Key point: [Corollary 3.7 in (G. Brunick and S. Shreve, 2013)]
$(X_t^{\mathrm{M}})_{t\in[0,1]}$ is a solution to the Markovian SDE

$$\mathrm{d}X_t^{\mathrm{M}} = \tilde{\beta}_t(X_t^{\mathrm{M}})\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \, , \; t \in [0,1] \, , \; X_0^{\mathrm{M}} \sim \mu \, ,$$

with drift $\tilde{\beta}_t(x) = \mathbb{E}[2\nabla_x \log p_{1-t}(X_1^{\mathrm{I}}|X_t^{\mathrm{I}})|X_t^{\mathrm{I}} = x]$.

G. Brunick and S. Shreve (2013). Mimicking an Itô process by a solution of a stochastic differential equation. In *The Annals of Applied Probability*.

A first draft for DFM

Draft Idea: To match $\nu^\star$, we run the SDE satisfied by the Markovian projection of the stochastic interpolant.

Algorithm 1 Draft algorithm

1: **Input:** μ.
2: **Step 1:** Initialize $X_0^{\mathrm{M}} \sim \mu$.
3: **Step 2:** Compute

$$\mathrm{d}X_t^{\mathrm{M}} = \tilde{\beta}_t(X_t^{\mathrm{M}})\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \quad t \in [0,1].$$

4: **Output:** $\mathrm{Law}(X_1^{\mathrm{M}}) = \nu^\star$.

Key Idea: To approximate $\nu^\star$, we run the **Euler–Maruyama scheme** for the **estimated mimicking drift** (via neural networks).

Algorithm 1

Input: μ, $\{0 = t_0 < t_1 < \cdots < t_N = 1\}$.
Step 1: Initialize $X_0^\star \sim \mu$.
Step 2: **For each** $k = 0, \ldots, N-1$:
  Approximate $\tilde{\beta}_{t_k}(x)$ using $s_{\theta^\star}(t_k, x)$.
  Compute the update:

$$\mathrm{d}X_t^\star = s_{\theta^\star}(t_k, X_{t_k}^\star)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \quad t \in [t_k, t_{k+1}].$$

Output: $\nu_1^{\theta^\star} := \mathrm{Law}(X_1^\star)$.

Non-Asymptotic Guarantees for DFM Models

1. Consider a uniform grid $\{kh\}_k$ for discretizing time and assume the mimicking drift is estimated with precision ε^2.
2. Further assume that μ, $\nu^\star$, and the score functions associated with μ, $\nu^\star$, and π (*i.e.* $\nabla \log(\mathrm{d} \cdot /\mathrm{dLeb})$) have finite 8th-order moments.

Theorem 2 in (Gentiloni-Silveri et al., 2024)

Under these conditions, the Kullback-Leibler (KL) divergence between the output distribution and the target is bounded by:

$$\mathrm{KL}(\text{output}||\nu^\star) \leq \epsilon_{\text{estimation}} + \epsilon_{\text{discretization}}$$

where $\epsilon_{\text{estimation}} = \varepsilon^2$, $\epsilon_{\text{discretization}} = h(h^{1/8} + 1)(d^4 + \text{8th-order-moments})$.

Gentiloni-Silveri M, Conforti G. Durmus A. (2024) Theoretical guarantees in KL for Diffusion Flow Matching. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.