NEURAL INFORMATION PROCESSING SYSTEMS

Formulating Discrete Probability Flow Through Optimal Transport

Pengze Zhang[*], Hubery Yin[*], Chen Li, Xiaohua Xie[†]

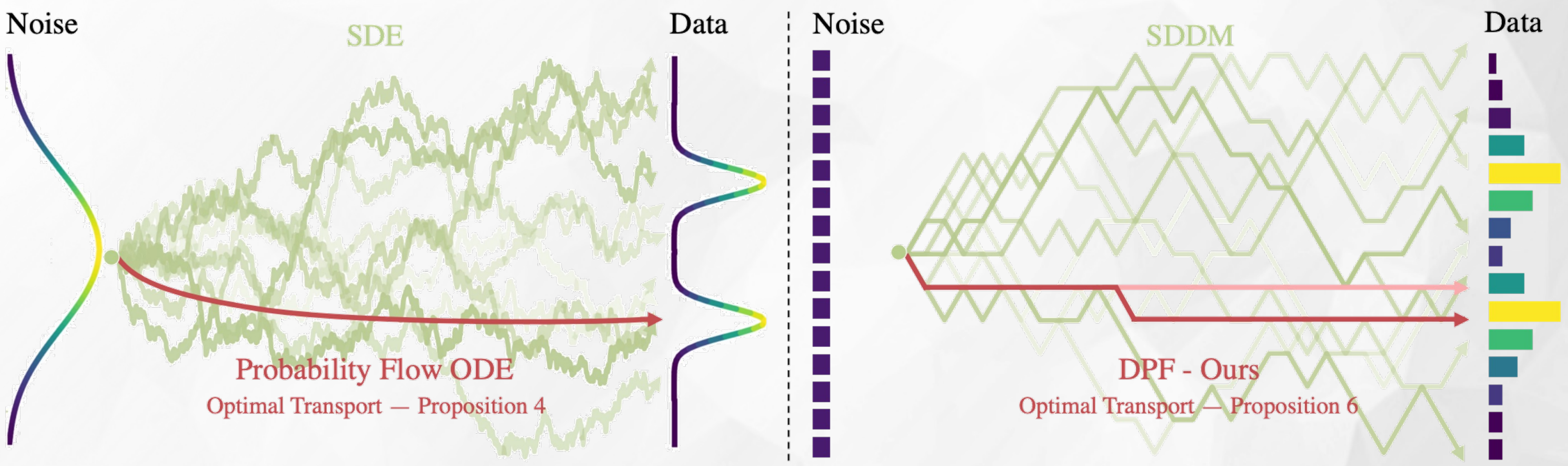
中山大学 SUN YAT-SEN UNIVERSITY

Reporter: Pengze Zhang

WeChat

Contents

Two Types of Diffusion Models



(a) Continuous Diffusion Model

(b) Discrete Diffusion Model

Figure 5: Schematic representation of different diffusion models.

Contributions

- Prove that the **continuous probability flow** is the Monge **optimal transport** map under certain conditions.

- Proposed a **Discrete Probability Flow (DPF)** under the framework of **optimal transport**.

- Proposed a **novel sampling method** based on the DPF, significantly decrease the **uncertainty** of the sampling outcomes.

Motivation

The Kolmogorov forward equation for discrete diffusion model is:

$$\frac{d}{dt}P^i_j(t|s) = \sum_{j'} P^i_{j'}(t|s){Q_D}^{j'}_j(t),$$

where $P^i_{j'}(t|s)$ means $P(x_t = j|x_s = i)$, i, j are states, and Q_D is defined as:

$$ {Q_D}^i_j = \begin{cases} 1, & d_D(i,j) = 1, \\ -\sum_{j' \epsilon \{k: d_D(i,k)=1\}} {Q_D}^i_{j'}, & d_D(i,j) = 0, \\ 0, & \textit{otherwise}, \end{cases} $$

where d is the Manhattan distance. However, the process defined by Q_D is not an **optimal transport map**, as there exist **mutual flows** between the states. (i.e., there exists two states i, j with ${Q_D}^i_j > 0$ and ${Q_D}^j_i > 0$). This process is not an **OT Plan**.

Method

To tackle the mutual flow, we modified the transition rate Q as:

$$Q_j^i(t) = \begin{cases} \dfrac{ReLU(P_{D_i}(t) - P_{D_j}(t))}{P_{D_i}(t)}, & d_D(i,j) = 1, \\ -\sum_{j' \epsilon \{k: d_D(i,k)=1\}} Q_{D}{}_{j'}^{i}(t), & d_D(i,j) = 0, \\ 0, & \textit{otherwise.} \end{cases}$$

The **ReLU** operation ensures that **high probability** states can only **transition** to **lower probability** states in **one direction**, effectively preventing mutual flow.

Important Proposition

Proposition 5. *The processes generated by* Q_D *and* Q *have the same single-time marginal distribution* $\forall t > 0$.

Proposition 6. *Given any* $t > 0$, *there exists a* $\delta_t > 0\ s.t. \forall 0 < s < \delta_t$, *the process generated by* Q *provides an optimal transport map from* $P_D(t)$ *to* $P_D(t+s)$ *under the cost* d_D.

Sampling

With the above proposition, we proposed **a novel sampling method** based on DPF, and the generator of the reverse process is:

$$R_j^i(t) = \begin{cases} ReLU\left(\dfrac{P^{\theta}_{D_{j_l|i\backslash i_l}}(t)}{P^{\theta}_{D_{i_l|i\backslash i_l}}(t)} - 1\right), & d_D(i,j) = 1 \text{ and } i_l \neq j_l, \\ -\sum_{j' \in \{k: d_D(i,k)=1\}} R^i_{j'}(t), & d_D(i,j) = 0 \\ 0, & otherwise. \end{cases}$$

Based on the above reverse transition rate, we can use Euler's method to generate samples.

Quantitative Results on Toy Dataset

Table 2: Comparison of certainty for SDDM and DPF, in terms of CSD on 4,000 initial points, each of which has 10 generated samples. Lower values indicate superior certainty.

	2spirals	8gaussians	checkerboard	circles	moons	pinwheel	swissroll
discrete dimension = 32, state size = 2							
SDDM	14.3053	14.1882	14.7433	14.4327	14.1739	14.0450	14.0548
DPF (ours)	2.1719	1.7945	2.0693	1.7210	2.0573	2.1834	1.8892
discrete dimension = 16, state size = 5							
SDDM	14.4645	14.6143	14.6963	14.4807	14.2397	14.2466	14.2659
DPF (ours)	1.9711	1.9367	1.4172	1.7185	1.7668	1.9633	1.6665
discrete dimension = 12, state size = 10							
SDDM	12.8463	12.7933	13.0158	12.9232	12.6665	12.7634	12.7880
DPF (ours)	1.8123	1.3178	1.1348	1.4625	1.4859	1.8435	1.5227

Visualization Results on Toy Dataset

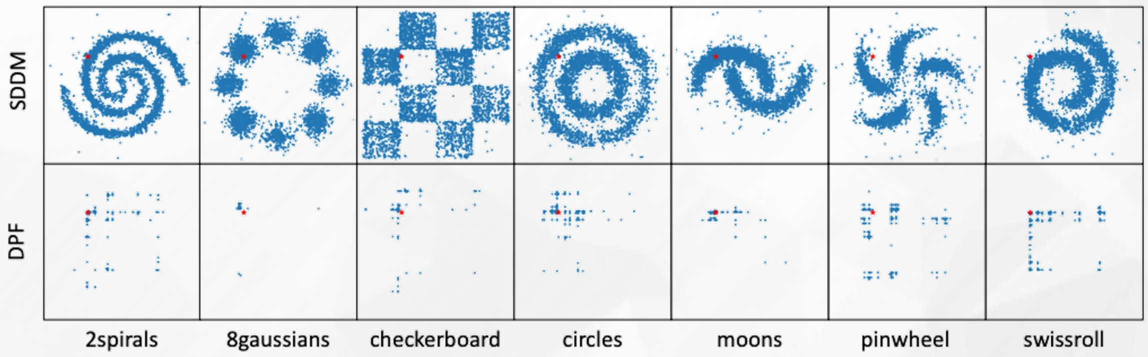


Figure 2: Visualization of the generating certainty on generated binary samples for SDDM and DPF. All the samples (in blue) are randomly generated from the single initial point (in red).

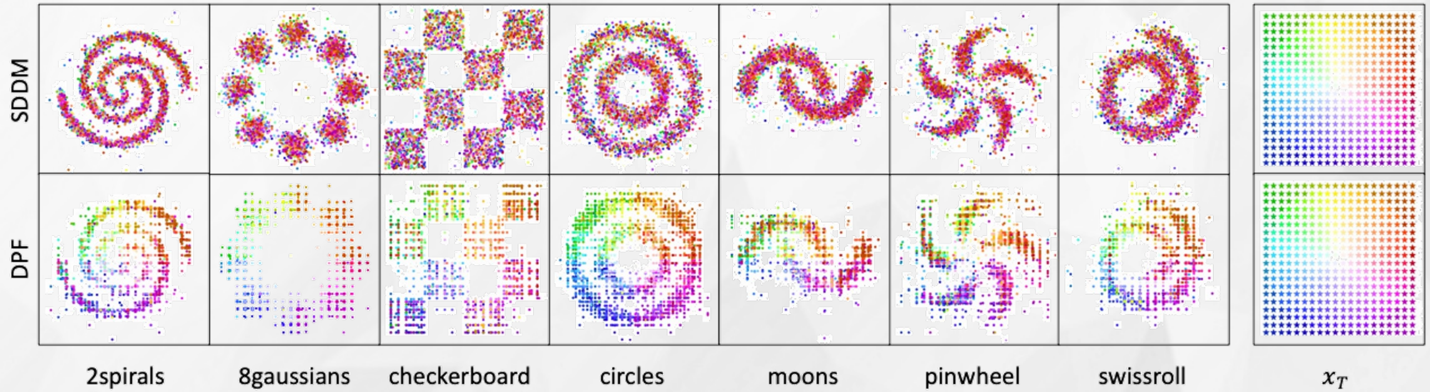


Figure 3: Visualization of the generated binary samples from the given initial points $\boldsymbol{x}_T$. Different colors distinguish the generated samples from different initial points $\boldsymbol{x}_T$.

Quantitative Results on CiFar-10

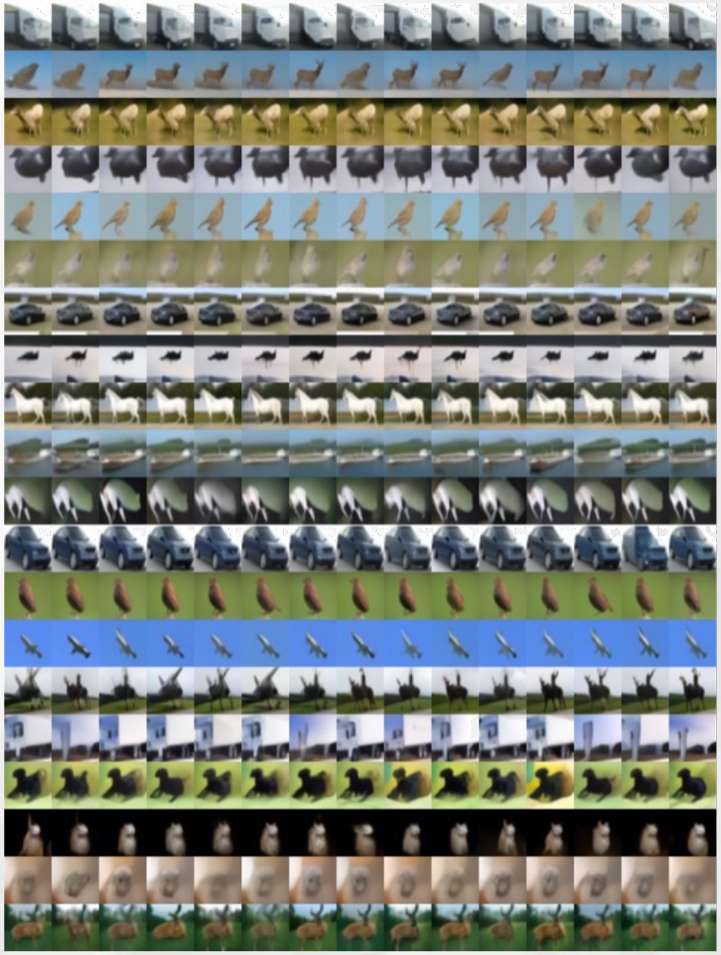
Table 8: Comparison of certainty for τLDR-0 and DPF on the Cifar-10 dataset. Here, CSD , class-std, and class-entropy are calculated on 1,000 initial points, each of which has 10 generated images. Lower values indicate superior certainty.

	CSD	class-std	class-entropy
τLDR-0 [7]	57.6898	2.6628	1.7703
DPF (ours)	9.4420	1.1819	0.5291

Visualization Results on CiFar-10

x_T

τLDR-0



DPF (ours)

THANKS

https://github.com/PangzeCheung/Discrete-Probability-Flow.