Yinuo Ren[1], Haoxuan Chen[1],
Grant M. Rotskoff[3,1], Lexing Ying[2,1]
{yinuoren,haoxuanc,rotskoff,lexing}@stanford.edu

[1]ICME [2]Department of Mathematics [3]Department of Chemistry
Stanford University

April 1, 2025

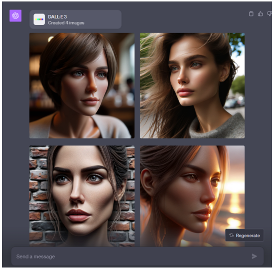
How Discrete and Continuous Diffusion Meet:

Comprehensive Analysis of Discrete Diffusion Models via a Stochastic Integral Framework

Section 1:
Diffusion Models: an Introduction

Diffusion Models

Introduction



(a) DALL·E 3

(b) Stable Diffusion

(c) AI4Science

Figure: Diffusion and flow-based generative models have exerted huge impacts on scientific research in many fields.

Problem Setting: Continuous Diffusion Models

- **Task:** Sample from data distribution p_0 *accurately* and *efficiently*

Problem Setting: Continuous Diffusion Models

- **Task:** Sample from data distribution p_0 *accurately* and *efficiently*
- **Forward SDE:**

$$\mathrm{d}\boldsymbol{x}_s = \boldsymbol{\beta}_s(\boldsymbol{x}_s)\mathrm{d}s + \boldsymbol{\sigma}_s \mathrm{d}\boldsymbol{w}_s, \quad \text{with} \quad \boldsymbol{x}_0 \sim p_0$$

Problem Setting: Continuous Diffusion Models

- **Task:** Sample from data distribution p_0 *accurately* and *efficiently*
- **Forward SDE:**

$$\mathrm{d}\boldsymbol{x}_s = \boldsymbol{\beta}_s(\boldsymbol{x}_s)\mathrm{d}s + \boldsymbol{\sigma}_s \mathrm{d}\boldsymbol{w}_s, \quad \text{with} \quad \boldsymbol{x}_0 \sim p_0$$

- **Backward SDE:**

$$\mathrm{d}\overleftarrow{\boldsymbol{x}}_t = \left[-\overleftarrow{\boldsymbol{\beta}}_t(\overleftarrow{\boldsymbol{x}}_t) + \frac{\overleftarrow{\boldsymbol{\sigma}}_t \overleftarrow{\boldsymbol{\sigma}}_t^\top + \overleftarrow{\boldsymbol{v}}_t \overleftarrow{\boldsymbol{v}}_t^\top}{2} \nabla \log \overleftarrow{p}_t(\overleftarrow{\boldsymbol{x}}_t) \right] \mathrm{d}t + \overleftarrow{\boldsymbol{v}}_t \mathrm{d}\boldsymbol{w}_t$$

with $\overleftarrow{p}_0 = p_T \approx \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and $\overleftarrow{p}_T = p_0$

Problem Setting: Continuous Diffusion Models

- **Task:** Sample from data distribution p_0 *accurately* and *efficiently*
- **Forward SDE:**

$$\mathrm{d}\boldsymbol{x}_s = \boldsymbol{\beta}_s(\boldsymbol{x}_s)\mathrm{d}s + \boldsymbol{\sigma}_s \mathrm{d}\boldsymbol{w}_s, \quad \text{with} \quad \boldsymbol{x}_0 \sim p_0$$

- **Backward SDE:**

$$\mathrm{d}\overleftarrow{\boldsymbol{x}}_t = \left[-\overleftarrow{\boldsymbol{\beta}}_t(\overleftarrow{\boldsymbol{x}}_t) + \frac{\overleftarrow{\boldsymbol{\sigma}}_t \overleftarrow{\boldsymbol{\sigma}}_t^\top + \overleftarrow{\boldsymbol{v}}_t \overleftarrow{\boldsymbol{v}}_t^\top}{2} \nabla \log \overleftarrow{p}_t(\overleftarrow{\boldsymbol{x}}_t) \right] \mathrm{d}t + \overleftarrow{\boldsymbol{v}}_t \mathrm{d}\boldsymbol{w}_t$$

  with $\overleftarrow{p}_0 = p_T \approx \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and $\overleftarrow{p}_T = p_0$
- **Score Function**: $\boldsymbol{s}_t^\theta(\boldsymbol{x}_t) \approx \nabla \log p_t(\boldsymbol{x}_t)$ by optimizing

$$\mathcal{L}(\theta) = \int_0^T \psi_t \mathbb{E}_{\boldsymbol{x}_t \sim p_t} \left[\left\| \nabla \log p_t(\boldsymbol{x}_t) - \boldsymbol{s}_t^\theta(\boldsymbol{x}_t) \right\|^2 \right] \mathrm{d}t$$

Problem Setting: Continuous Diffusion Models

- **Task:** Sample from data distribution p_0 *accurately* and *efficiently*
- **Forward SDE:**

$$\mathrm{d}\boldsymbol{x}_s = \boldsymbol{\beta}_s(\boldsymbol{x}_s)\mathrm{d}s + \boldsymbol{\sigma}_s \mathrm{d}\boldsymbol{w}_s, \quad \text{with} \quad \boldsymbol{x}_0 \sim p_0$$

- **Backward SDE:**

$$\mathrm{d}\overleftarrow{\boldsymbol{x}}_t = \left[-\overleftarrow{\boldsymbol{\beta}}_t(\overleftarrow{\boldsymbol{x}}_t) + \frac{\overleftarrow{\boldsymbol{\sigma}}_t\overleftarrow{\boldsymbol{\sigma}}_t^\top + \overleftarrow{\boldsymbol{v}}_t\overleftarrow{\boldsymbol{v}}_t^\top}{2}\nabla \log \overleftarrow{p}_t(\overleftarrow{\boldsymbol{x}}_t)\right]\mathrm{d}t + \overleftarrow{\boldsymbol{v}}_t \mathrm{d}\boldsymbol{w}_t$$

with $\overleftarrow{p}_0 = p_T \approx \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and $\overleftarrow{p}_T = p_0$

- **Score Function**: $\boldsymbol{s}_t^\theta(\boldsymbol{x}_t) \approx \nabla \log p_t(\boldsymbol{x}_t)$ by optimizing

$$\mathcal{L}(\theta) = \int_0^T \psi_t \mathbb{E}_{\boldsymbol{x}_t \sim p_t}\left[\left\|\nabla \log p_t(\boldsymbol{x}_t) - \boldsymbol{s}_t^\theta(\boldsymbol{x}_t)\right\|^2\right]\mathrm{d}t$$

- **Implementations:** SDE ($\boldsymbol{v}_t = \boldsymbol{\sigma}_t$), Probability Flow ODE (PF-ODE, $\boldsymbol{v}_t \equiv \mathbf{0}$)

Error Analysis: Continuous Diffusion Models

Take $\boldsymbol{\beta}_s(\boldsymbol{x}_s) = -0.5\boldsymbol{x}_s$ and $\boldsymbol{\sigma}_s = \boldsymbol{I}$:

- **Forward SDE:** $\mathrm{d}\boldsymbol{x}_s = -\frac{1}{2}\boldsymbol{x}_s\mathrm{d}s + \mathrm{d}\boldsymbol{w}_s$ with $\boldsymbol{x}_0 \sim p_0$

Error Analysis: Continuous Diffusion Models

Take $\boldsymbol{\beta}_s(\boldsymbol{x}_s) = -0.5\boldsymbol{x}_s$ and $\boldsymbol{\sigma}_s = \boldsymbol{I}$:

- **Forward SDE:** $\mathrm{d}\boldsymbol{x}_s = -\frac{1}{2}\boldsymbol{x}_s \mathrm{d}s + \mathrm{d}\boldsymbol{w}_s$ with $\boldsymbol{x}_0 \sim p_0$
- **Backward SDE:** $\mathrm{d}\overleftarrow{\boldsymbol{x}}_t = \left[\frac{1}{2}\overleftarrow{\boldsymbol{x}}_t + \frac{1+\upsilon^2}{2}\nabla \log \overleftarrow{p}_t(\overleftarrow{\boldsymbol{x}}_t)\right] \mathrm{d}t + \upsilon \mathrm{d}\boldsymbol{w}_t$, with $\overleftarrow{p}_0 = p_T \approx \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\overleftarrow{p}_T = p_0$

Error Analysis: Continuous Diffusion Models

Take $\boldsymbol{\beta}_s(\boldsymbol{x}_s) = -0.5\boldsymbol{x}_s$ and $\boldsymbol{\sigma}_s = \boldsymbol{I}$:

- Forward SDE: $\mathrm{d}\boldsymbol{x}_s = -\frac{1}{2}\boldsymbol{x}_s \mathrm{d}s + \mathrm{d}\boldsymbol{w}_s$ with $\boldsymbol{x}_0 \sim p_0$
- Backward SDE: $\mathrm{d}\overleftarrow{\boldsymbol{x}}_t = \left[\frac{1}{2}\overleftarrow{\boldsymbol{x}}_t + \frac{1+\upsilon^2}{2}\nabla \log \overleftarrow{p}_t(\overleftarrow{\boldsymbol{x}}_t)\right]\mathrm{d}t + \upsilon \mathrm{d}\boldsymbol{w}_t$, with $\overleftarrow{p}_0 = p_T \approx \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and $\overleftarrow{p}_T = p_0$

Theorem (Error Analysis of Continuous Diffusion Models)

Suppose $t_0 = 0 \leq \cdots \leq t_N = T - \delta$ satisfies $t_{k+1} - t_k \leq \kappa(T - t_{k+1})$ and

$$\sum_{k=0}^{N-1} (s_{k+1} - s_k)\mathbb{E}_{\overleftarrow{\boldsymbol{x}}_{s_k} \sim \overleftarrow{p}_{s_k}} \left[\left\|\nabla \log \overleftarrow{p}_{s_k}(\overleftarrow{\boldsymbol{x}}_{s_k}) - \widehat{\overleftarrow{\boldsymbol{s}}}^{\theta}_{s_k}(\boldsymbol{x}_{s_k})\right\|^2\right] \leq \epsilon.$$

Then with

$$T = \mathcal{O}(\log(d\epsilon^{-1})),\ \kappa = \mathcal{O}(d^{-1}\epsilon \log^{-1}(d\epsilon^{-1})),\ N = \mathcal{O}(d\epsilon^{-1}\log^2(d\epsilon^{-1})),$$

we have

$$D_{\mathrm{KL}}(p_\delta \| \widehat{q}_{t_N}) \lesssim de^{-T} + \epsilon + d\kappa T \lesssim \epsilon.$$

Error Analysis: Continuous Diffusion Models

Theorem (Error Analysis of Continuous Diffusion Models)

With

$$T = \mathcal{O}(\log(d\epsilon^{-1})),\ \kappa = \mathcal{O}(d^{-1}\epsilon \log^{-1}(d\epsilon^{-1})),\ N = \mathcal{O}(d\epsilon^{-1}\log^2(d\epsilon^{-1})),$$

we have

$$D_{\mathrm{KL}}(p_\delta \| \widehat{q}_{t_N}) \lesssim de^{-T} + \epsilon + d\kappa T \lesssim \epsilon.$$

Error Analysis: Continuous Diffusion Models

Theorem (Error Analysis of Continuous Diffusion Models)

With

$$T = \mathcal{O}(\log(d\epsilon^{-1})),\ \kappa = \mathcal{O}(d^{-1}\epsilon\log^{-1}(d\epsilon^{-1})),\ N = \mathcal{O}(d\epsilon^{-1}\log^{2}(d\epsilon^{-1})),$$

we have

$$D_{\mathrm{KL}}(p_\delta \| \widehat{q}_{t_N}) \lesssim de^{-T} + \epsilon + d\kappa T \lesssim \epsilon.$$

- **Truncation Error:** Error caused by approximating p_T by p_∞, of the order $\mathcal{O}(d\exp(-T))$;

Error Analysis: Continuous Diffusion Models

Theorem (Error Analysis of Continuous Diffusion Models)

With

$$T = \mathcal{O}(\log(d\epsilon^{-1})),\ \kappa = \mathcal{O}(d^{-1}\epsilon \log^{-1}(d\epsilon^{-1})),\ N = \mathcal{O}(d\epsilon^{-1}\log^2(d\epsilon^{-1})),$$

we have

$$D_{\mathrm{KL}}(p_\delta \| \widehat{q}_{t_N}) \lesssim de^{-T} + \epsilon + d\kappa T \lesssim \epsilon.$$

- **Truncation Error:** Error caused by approximating p_T by p_∞, of the order $\mathcal{O}(d\exp(-T))$;
- **Approximation Error:** Error caused by approximating $\nabla \log p_t(\boldsymbol{x}_t)$ by NN $\widehat{\boldsymbol{s}}_t^\theta(\boldsymbol{x}_t)$, assumed to be of $\mathcal{O}(\epsilon)$;

Error Analysis: Continuous Diffusion Models

Theorem (Error Analysis of Continuous Diffusion Models)

With

$$T = \mathcal{O}(\log(d\epsilon^{-1})),\ \kappa = \mathcal{O}(d^{-1}\epsilon \log^{-1}(d\epsilon^{-1})),\ N = \mathcal{O}(d\epsilon^{-1}\log^2(d\epsilon^{-1})),$$

we have

$$D_{\mathrm{KL}}(p_\delta \| \widehat{q}_{t_N}) \lesssim de^{-T} + \epsilon + d\kappa T \lesssim \epsilon.$$

- **Truncation Error:** Error caused by approximating p_T by p_∞, of the order $\mathcal{O}(d\exp(-T))$;
- **Approximation Error:** Error caused by approximating $\nabla \log p_t(\boldsymbol{x}_t)$ by NN $\widehat{\boldsymbol{s}}_t^\theta(\boldsymbol{x}_t)$, assumed to be of $\mathcal{O}(\epsilon)$;
- **Discretization Error:** Error caused by numerically solving the backward SDE

Problem Setting: Discrete Diffusion Models

- **Task:** Sample from discrete data distribution $p_0 \in \Delta^{|\mathcal{X}|}$

Problem Setting: Discrete Diffusion Models

- **Task:** Sample from discrete data distribution $p_0 \in \Delta^{|\mathcal{X}|}$
- **Forward Continuous Time Markov Chain (CTMC):**

$$\frac{\mathrm{d}\boldsymbol{p}_t}{\mathrm{d}t} = \boldsymbol{Q}_t \boldsymbol{p}_t, \text{ with } Q_t(x,x) = -\sum_{y \neq x} Q_t(y,x) \text{ and } Q_t(x,y) \geq 0 \ (x \neq y)$$

Problem Setting: Discrete Diffusion Models

- **Task:** Sample from discrete data distribution $p_0 \in \Delta^{|\mathcal{X}|}$
- **Forward Continuous Time Markov Chain (CTMC):**

$$\frac{\mathrm{d}\boldsymbol{p}_t}{\mathrm{d}t} = \boldsymbol{Q}_t \boldsymbol{p}_t, \text{ with } Q_t(x,x) = -\sum_{y \neq x} Q_t(y,x) \text{ and } Q_t(x,y) \geq 0 \ (x \neq y)$$

- **Backward CTMC:**

$$\frac{\mathrm{d}\overleftarrow{\boldsymbol{p}}_s}{\mathrm{d}s} = \overline{\boldsymbol{Q}}_s \overleftarrow{\boldsymbol{p}}_s, \text{ with } \overline{Q}_s(y,x) = \begin{cases} \frac{\overleftarrow{p}_s(y)}{\overleftarrow{p}_s(x)} \overleftarrow{Q}_s(x,y), & \forall x \neq y \in \mathbb{X}, \\ -\sum_{y' \neq x} \overline{Q}_s(y',x), & \forall x = y \in \mathbb{X}. \end{cases}$$

with $\boldsymbol{p}_0 = \boldsymbol{p}_T \approx \text{Unif}(\Delta)$ and $\overleftarrow{p}_T = p_0$

Problem Setting: Discrete Diffusion Models

- **Task:** Sample from discrete data distribution $p_0 \in \Delta^{|\mathcal{X}|}$
- **Forward Continuous Time Markov Chain (CTMC):**

$$\frac{\mathrm{d}\boldsymbol{p}_t}{\mathrm{d}t} = \boldsymbol{Q}_t \boldsymbol{p}_t, \text{ with } Q_t(x,x) = -\sum_{y \neq x} Q_t(y,x) \text{ and } Q_t(x,y) \geq 0 \ (x \neq y)$$

- **Backward CTMC:**

$$\frac{\mathrm{d}\overleftarrow{\boldsymbol{p}}_s}{\mathrm{d}s} = \overline{\boldsymbol{Q}}_s \overleftarrow{\boldsymbol{p}}_s, \text{ with } \overline{Q}_s(y,x) = \begin{cases} \frac{\overleftarrow{p}_s(y)}{\overleftarrow{p}_s(x)} \overleftarrow{Q}_s(x,y), & \forall x \neq y \in \mathbb{X}, \\ -\sum_{y' \neq x} \overline{Q}_s(y',x), & \forall x = y \in \mathbb{X}. \end{cases}$$

with $\boldsymbol{p}_0 = \boldsymbol{p}_T \approx \text{Unif}(\Delta)$ and $\overleftarrow{p}_T = p_0$

- **Score Function:** $\boldsymbol{s}_t(x) = (s_t(x,y))_{y \in \mathbb{X}} = \frac{\boldsymbol{p}_t}{p_t(x)}$ by optimizing

$$\int_0^T \psi_t \mathbb{E}_{x_t \sim p_t} \left[\sum_{y \neq x} \left(-\log \frac{\widehat{s}_t^\theta(x,y)}{s_t(x,y)} - 1 + \frac{\widehat{s}_t^\theta(x,y)}{s_t(x,y)} \right) s_t(x,y) Q_t(x,y) \right] \mathrm{d}t$$

Problem Setting: Discrete Diffusion Models

- **Task:** Sample from discrete data distribution $p_0 \in \Delta^{|\mathcal{X}|}$
- **Forward Continuous Time Markov Chain (CTMC):**

$$\frac{\mathrm{d}\boldsymbol{p}_t}{\mathrm{d}t} = \boldsymbol{Q}_t \boldsymbol{p}_t, \text{ with } Q_t(x,x) = -\sum_{y \neq x} Q_t(y,x) \text{ and } Q_t(x,y) \geq 0 \ (x \neq y)$$

- **Backward CTMC:**

$$\frac{\mathrm{d}\overleftarrow{\boldsymbol{p}}_s}{\mathrm{d}s} = \overline{\boldsymbol{Q}}_s \overleftarrow{\boldsymbol{p}}_s, \text{ with } \overline{Q}_s(y,x) = \begin{cases} \frac{\overleftarrow{p}_s(y)}{\overleftarrow{p}_s(x)} \overleftarrow{Q}_s(x,y), & \forall x \neq y \in \mathbb{X}, \\ -\sum_{y' \neq x} \overline{Q}_s(y',x), & \forall x = y \in \mathbb{X}. \end{cases}$$

with $\boldsymbol{p}_0 = \boldsymbol{p}_T \approx$ Unif(Δ) and $\overleftarrow{p}_T = p_0$

- **Score Function**: $\boldsymbol{s}_t(x) = (s_t(x,y))_{y \in \mathbb{X}} = \frac{\boldsymbol{p}_t}{p_t(x)}$ by optimizing

$$\int_0^T \psi_t \mathbb{E}_{x_t \sim p_t} \left[\sum_{y \neq x} \left(-\log \frac{\widehat{s}_t^\theta(x,y)}{s_t(x,y)} - 1 + \frac{\widehat{s}_t^\theta(x,y)}{s_t(x,y)} \right) s_t(x,y) Q_t(x,y) \right] \mathrm{d}t$$

- **Implementations:** τ-leaping scheme ([Gil01, CBDB+22]), uniformization ([VD92, CY24])

Section 2:
A Stochastic Integral Framework for Discrete DMs

Mathematical Preliminaries

Poisson Random Measure with Evolving Intensity

$(\Omega, \mathcal{F}, \mathcal{P})$: probability space
$(\mathbb{X}, \mathcal{B}, \nu)$: measure space
$\lambda_t(y)$: a non-negative predictable process on $\mathbb{R}^+ \times \mathbb{X} \times \Omega$ satisfying

$$\int_0^T \int_{\mathbb{X}} 1 \vee |y| \vee |y|^2 \lambda_t(y) \nu(\mathrm{d}y) \mathrm{d}t < \infty, \text{ a.s..}$$

for any $T > 0$, The random measure $N[\lambda](\mathrm{d}t, \mathrm{d}y)$ on $\mathbb{R}^+ \times \mathbb{X}$ is called a *Poisson random measure with evolving intensity* $\lambda_t(y)$ if

Mathematical Preliminaries

Poisson Random Measure with Evolving Intensity

$(\Omega, \mathcal{F}, \mathcal{P})$: probability space
$(\mathbb{X}, \mathcal{B}, \nu)$: measure space
$\lambda_t(y)$: a non-negative predictable process on $\mathbb{R}^+ \times \mathbb{X} \times \Omega$ satisfying

$$\int_0^T \int_{\mathbb{X}} 1 \vee |y| \vee |y|^2 \lambda_t(y) \nu(\mathrm{d}y) \mathrm{d}t < \infty, \text{ a.s..}$$

for any $T > 0$, The random measure $N[\lambda](\mathrm{d}t, \mathrm{d}y)$ on $\mathbb{R}^+ \times \mathbb{X}$ is called a *Poisson random measure with evolving intensity* $\lambda_t(y)$ if

1. For any $B \in \mathcal{B}$ and $0 \le s < t$, $N[\lambda]((s,t] \times B) \sim \mathcal{P}\left(\int_s^t \int_B \lambda_\tau(y) \nu(\mathrm{d}y) \mathrm{d}\tau\right)$; Interpretation: the number of jumps of magnitude y during the infinitesimal time interval $(t, t + \mathrm{d}t]$ is Poisson distributed with mean $\lambda_t(\nu)\gamma(\mathrm{d}\nu)\mathrm{d}t$.

Mathematical Preliminaries

Poisson Random Measure with Evolving Intensity

$(\Omega, \mathcal{F}, \mathcal{P})$: probability space
$(\mathbb{X}, \mathcal{B}, \nu)$: measure space
$\lambda_t(y)$: a non-negative predictable process on $\mathbb{R}^+ \times \mathbb{X} \times \Omega$ satisfying

$$\int_0^T \int_{\mathbb{X}} 1 \vee |y| \vee |y|^2 \lambda_t(y)\nu(\mathrm{d}y)\mathrm{d}t < \infty, \text{ a.s..}$$

for any $T > 0$, The random measure $N[\lambda](\mathrm{d}t, \mathrm{d}y)$ on $\mathbb{R}^+ \times \mathbb{X}$ is called a *Poisson random measure with evolving intensity* $\lambda_t(y)$ if

1. For any $B \in \mathcal{B}$ and $0 \leq s < t$, $N[\lambda]((s,t] \times B) \sim \mathcal{P}\left(\int_s^t \int_B \lambda_\tau(y)\nu(\mathrm{d}y)\mathrm{d}\tau\right)$;
   Interpretation: the number of jumps of magnitude y during the infinitesimal time interval $(t, t+\mathrm{d}t]$ is Poisson distributed with mean $\lambda_t(\nu)\gamma(\mathrm{d}\nu)\mathrm{d}t$.
2. For any $t \geq 0$ and disjoint sets $\{B_i\}_{i\in[n]} \subset \mathcal{B}$, $\{N_t[\lambda](B_i) := N[\lambda]((0,t] \times B_i)\}_{i\in[n]}$ are independent stochastic processes.

Stochastic Integral Formulation: True Processes

Forward Process:

$$x_t = x_0 + \int_0^t \int_{\mathbb{X}} (y - x_{\tau^-}) N[\lambda](\mathrm{d}\tau, \mathrm{d}y), \text{ with } \lambda_\tau(y) = \widetilde{Q}_\tau(y, x_{\tau^-}),$$

True Backward Process ($\overleftarrow{s}_t$ denotes the true score):

$$\overleftarrow{x}_t = \overleftarrow{x}_0 + \int_0^t \int_{\mathbb{X}} (y - \overleftarrow{x}_{\tau^-}) N[\mu](\mathrm{d}\tau, \mathrm{d}y), \text{ with } \mu_\tau(y) = \overleftarrow{s}_\tau(\overleftarrow{x}_{\tau^-}, y)\overline{Q}_\tau(\overleftarrow{x}_{\tau^-}, y),$$

Stochastic Integral Formulation: τ-Leaping Scheme

Notation: time discretizations $(s_i)_{i\in[0:N]}$ with $s_0 = 0$ and $s_N = T - \delta$, $\lfloor \tau \rfloor = s_n$ for any $\tau \in [s_n, s_{n+1})$

τ-leaping inference scheme:

$$\widehat{y}_{s_{n+1}} \leftarrow \widehat{y}_{s_n} + \sum_{y\in\mathbb{X}} (y - \widehat{y}_{s_n}) \mathcal{P}(\widehat{\mu}^{\theta}_{s_n}(y)(s_{n+1} - s_n))$$

with the evolving intensity $\widehat{\mu}^{\theta}_{\lfloor\tau\rfloor}(y) = \widehat{\overleftarrow{s}}^{\theta}_{\lfloor\tau\rfloor}(\widehat{y}_{\lfloor\tau\rfloor^-}, y)\overline{Q}_{\lfloor\tau\rfloor}(\widehat{y}_{\lfloor\tau\rfloor^-}, y) = \widehat{\mu}^{\theta}_{s_n}(y)$

Approximate Backward Process (equivalent to τ-leaping)

$$\widehat{y}_s = \widehat{y}_0 + \int_0^s \int_{\mathbb{X}} (y - \widehat{y}_{\lfloor\tau\rfloor^-}) N[\widehat{\mu}^{\theta}_{\lfloor\cdot\rfloor}](\mathrm{d}\tau, \mathrm{d}y)$$

Stochastic Integral Formulation: Uniformization Scheme

Notation: time discretization $(s_b)_{b\in[0,N]}$ with $s_0 = 0$ and $s_N = T - \delta$
Uniformization inference scheme:
$M \sim \mathcal{P}(\overline{\lambda}_{s_{b+1}}(s_{b+1} - s_b))$, $\sigma_m \sim \mathrm{Unif}([0,1])$ for $m \in [M]$

$$\widehat{y}_{s_b+\sigma_{(m)}} \leftarrow \begin{cases} y, & \text{w.p. } \widehat{\mu}^\theta_{s_b+\sigma_{(m)}}(y)/\overline{\lambda}_{s_{b+1}}, \text{ for } y \in \mathbb{X}, \\ \widehat{y}_{s_b}, & \text{w.p. } 1 - \sum_{y\in\mathbb{X}} \widehat{\mu}^\theta_{s_b+\sigma_{(m)}}(y)/\overline{\lambda}_{s_{b+1}}; \end{cases}$$

with the evolving intensity $\widehat{\mu}^\theta_s(y) = \overleftarrow{\widehat{s}}^\theta_s(\widehat{y}_{s^-}, y)\overline{Q}_s(\widehat{y}_{s^-}, y)$
Approximate Backward Process (equivalent to uniformization)

$$y_s = y_0 + \int_0^s \int_{\mathbb{X}} \int_{\mathbb{R}} (y - y_{s^-}) \mathbf{1}_{0\le\xi\le\int_{\mathbb{X}} \widehat{\mu}^\theta_s(y)\nu(\mathrm{d}y)} N[\widehat{\mu}^\theta](\mathrm{d}s, \mathrm{d}y, \mathrm{d}\xi)$$

Assumptions for Discrete DMs

- **Regularity of rate matrix $\boldsymbol{Q}$:** (i) $Q(x,y) \le C$ and $\underline{D} \le -Q(x,x) \le \overline{D}$, $\forall x, y \in \mathbb{X}$, where $C, \underline{D}, \overline{D} > 0$; (ii) $\rho(\boldsymbol{Q}) \ge \rho > 0$ for the modified log-Sobolev constant $\rho(\boldsymbol{Q})$ of the rate matrix $\boldsymbol{Q}$.

Assumptions for Discrete DMs

- **Regularity of rate matrix $\boldsymbol{Q}$:** (i) $Q(x,y) \leq C$ and $\underline{D} \leq -Q(x,x) \leq \overline{D}$, $\forall x, y \in \mathbb{X}$, where $C, \underline{D}, \overline{D} > 0$; (ii) $\rho(\boldsymbol{Q}) \geq \rho > 0$ for the modified log-Sobolev constant $\rho(\boldsymbol{Q})$ of the rate matrix $\boldsymbol{Q}$.
- **Boundedness of true score s_t and learned score $\widehat{s}_t$:** $s_t(x,y) \lesssim 1 \vee t^{-1}$ and $\widehat{s}^{\theta}_s(x,y) \in (0, M]$, $\forall x, y \in \mathbb{X}$.

Assumptions for Discrete DMs

- **Regularity of rate matrix $\boldsymbol{Q}$:** (i) $Q(x,y) \leq C$ and $\underline{D} \leq -Q(x,x) \leq \overline{D}$, $\forall x, y \in \mathbb{X}$, where $C, \underline{D}, \overline{D} > 0$; (ii) $\rho(\boldsymbol{Q}) \geq \rho > 0$ for the modified log-Sobolev constant $\rho(\boldsymbol{Q})$ of the rate matrix $\boldsymbol{Q}$.
- **Boundedness of true score s_t and learned score $\widehat{s}_t$:** $s_t(x,y) \lesssim 1 \vee t^{-1}$ and $\widehat{s}^{\theta}_s(x,y) \in (0, M]$, $\forall x, y \in \mathbb{X}$.
- **Continuity of true score** For any $t > 0$ and $y \in \mathbb{X}$ such that $Q(x_{t-}, y) > 0$, we have $\left|\frac{\mu_{t+}(y)}{\mu_t(y)}\right| := \left|\frac{p_t(x_{t-})Q(x_t,y)}{p_t(x_t)Q(x_{t-},y)} - 1\right| \lesssim 1 \vee t^{-\gamma}$ for some exponent $\gamma \in [0,1]$.

Assumptions for Discrete DMs

- **Regularity of rate matrix $\boldsymbol{Q}$:** (i) $Q(x,y) \leq C$ and $\underline{D} \leq -Q(x,x) \leq \overline{D}$, $\forall x, y \in \mathbb{X}$, where $C, \underline{D}, \overline{D} > 0$; (ii) $\rho(\boldsymbol{Q}) \geq \rho > 0$ for the modified log-Sobolev constant $\rho(\boldsymbol{Q})$ of the rate matrix $\boldsymbol{Q}$.
- **Boundedness of true score s_t and learned score $\widehat{s}_t$:** $s_t(x,y) \lesssim 1 \vee t^{-1}$ and $\widehat{s}^\theta_s(x,y) \in (0, M]$, $\forall x, y \in \mathbb{X}$.
- **Continuity of true score** For any $t > 0$ and $y \in \mathbb{X}$ such that $Q(x_{t-}, y) > 0$, we have $\left|\frac{\mu_{t^+}(y)}{\mu_t(y)}\right| := \left|\frac{p_t(x_{t-})Q(x_t,y)}{p_t(x_t)Q(x_{t-},y)} - 1\right| \lesssim 1 \vee t^{-\gamma}$ for some exponent $\gamma \in [0,1]$.
- **δ-accurate score estimation:**

$$\sum_{n=0}^{N-1} (s_{n+1} - s_n)\mathbb{E}\left[\int_{\mathbb{X}} K\left(\frac{\overleftarrow{\widehat{s}}^\theta_{s_n}(\overleftarrow{x}_{s_n^-}, y)}{\overleftarrow{s}_{s_n}(\overleftarrow{x}_{s_n^-}, y)}\right) \overleftarrow{s}_{s_n}(\overleftarrow{x}_{s_n^-}, y)\widetilde{Q}(\overleftarrow{x}_{s_n^-}, y)\nu(\mathrm{d}y)\right] \leq \delta.$$

Assumptions for Continuous DMs (for Comparison)

- **Regularity of data distribution:** p_0 has finite second moment and is normalized, i.e., $\mathrm{cov}_{p_0}(\boldsymbol{x}_0) = \boldsymbol{I}_d$

Assumptions for Continuous DMs (for Comparison)

- **Regularity of data distribution:** p_0 has finite second moment and is normalized, i.e., $\mathrm{cov}_{p_0}(\boldsymbol{x}_0) = \boldsymbol{I}_d$
- **Bounded learned score:** The learned score $\boldsymbol{s}_t^\theta$ has bounded C^1 norm with Lipschitz const $L_{\boldsymbol{s}}$.

Assumptions for Continuous DMs (for Comparison)

- **Regularity of data distribution:** p_0 has finite second moment and is normalized, i.e., $\mathrm{cov}_{p_0}(\boldsymbol{x}_0) = \boldsymbol{I}_d$
- **Bounded learned score:** The learned score $\boldsymbol{s}_t^\theta$ has bounded C^1 norm with Lipschitz const $L_{\boldsymbol{s}}$.
- **δ-accurate score estimation:**

Assumptions for Continuous DMs (for Comparison)

- **Regularity of data distribution:** p_0 has finite second moment and is normalized, i.e., $\mathrm{cov}_{p_0}(\boldsymbol{x}_0) = \boldsymbol{I}_d$
- **Bounded learned score:** The learned score $\boldsymbol{s}_t^\theta$ has bounded C^1 norm with Lipschitz const $L_{\boldsymbol{s}}$.
- **δ-accurate score estimation:**

SDE The learned score $\boldsymbol{s}_t^\theta$ is $L^2([0, t_N])$ δ-accurate:

$$\sum_{j=0}^{N-1} (t_{j+1} - t_j) \mathbb{E}_{\overleftarrow{p}_{t_j}} \left[\left\| \boldsymbol{s}_{t_j}^\theta (\overleftarrow{\boldsymbol{x}}_{t_j}) - \nabla \log \overleftarrow{p}_{t_j} (\overleftarrow{\boldsymbol{x}}_{t_j}) \right\|^2 \right] \leq \delta_2^2.$$

Assumptions for Continuous DMs (for Comparison)

- **Regularity of data distribution:** p_0 has finite second moment and is normalized, i.e., $\mathrm{cov}_{p_0}(\boldsymbol{x}_0) = \boldsymbol{I}_d$
- **Bounded learned score:** The learned score $\boldsymbol{s}_t^\theta$ has bounded C^1 norm with Lipschitz const $L_{\boldsymbol{s}}$.
- **δ-accurate score estimation:**

SDE The learned score $\boldsymbol{s}_t^\theta$ is $L^2([0, t_N])$ δ-accurate:

$$\sum_{j=0}^{N-1} (t_{j+1} - t_j)\mathbb{E}_{\overleftarrow{p}_{t_j}}\left[\left\|\boldsymbol{s}_{t_j}^\theta(\overleftarrow{\boldsymbol{x}}_{t_j}) - \nabla \log \overleftarrow{p}_{t_j}(\overleftarrow{\boldsymbol{x}}_{t_j})\right\|^2\right] \leq \delta_2^2.$$

PF-ODE The learned score $\boldsymbol{s}_t^\theta$ is $L^\infty([0, T-\delta])$ δ-accurate:

$$\mathbb{E}_{\overleftarrow{p}_{t_j}}\left[\left\|\boldsymbol{s}_{t_j}^\theta(\overleftarrow{\boldsymbol{x}}_{t_j}) - \nabla \log \overleftarrow{p}_{t_j}(\overleftarrow{\boldsymbol{x}}_{t_j})\right\|^2\right] \leq \delta_\infty^2.$$

Assumptions for Continuous DMs (for Comparison)

- **Regularity of data distribution:** p_0 has finite second moment and is normalized, i.e., $\mathrm{cov}_{p_0}(\boldsymbol{x}_0) = \boldsymbol{I}_d$
- **Bounded learned score:** The learned score $\boldsymbol{s}_t^\theta$ has bounded C^1 norm with Lipschitz const $L_{\boldsymbol{s}}$.
- **δ-accurate score estimation:**

SDE The learned score $\boldsymbol{s}_t^\theta$ is $L^2([0, t_N])$ δ-accurate:

$$\sum_{j=0}^{N-1} (t_{j+1} - t_j) \mathbb{E}_{\overleftarrow{p}_{t_j}} \left[\left\| \boldsymbol{s}_{t_j}^\theta(\overleftarrow{\boldsymbol{x}}_{t_j}) - \nabla \log \overleftarrow{p}_{t_j}(\overleftarrow{\boldsymbol{x}}_{t_j}) \right\|^2 \right] \leq \delta_2^2.$$

PF-ODE The learned score $\boldsymbol{s}_t^\theta$ is $L^\infty([0, T-\delta])$ δ-accurate:

$$\mathbb{E}_{\overleftarrow{p}_{t_j}} \left[\left\| \boldsymbol{s}_{t_j}^\theta(\overleftarrow{\boldsymbol{x}}_{t_j}) - \nabla \log \overleftarrow{p}_{t_j}(\overleftarrow{\boldsymbol{x}}_{t_j}) \right\|^2 \right] \leq \delta_\infty^2.$$

- **Continuity of true score (PF-ODE):** The true score $\nabla \log p_t$ has bounded C^1 norm with Lipschitz const L_p.

Section 3:

Main Results: Theoretical Guarantees of Inference Algorithms

Convergence Guarantee for τ-leaping

Theorem (Theoretical Guarantees for τ-Leaping)

Take time discretization scheme $(s_i)_{i\in[0,N]}$ *satisfying* $s_0 = 0$, $s_N = T - \delta$ *and* $s_{k+1} - s_k \le \kappa \left(1 \vee (T - s_{k+1})^{1+\gamma-\eta}\right)$ *for* $k \in [0:N-1]$. *Under aforementioned assumptions and the following choices of parameters*

$$T = \mathcal{O}\left(\frac{\log\left(\epsilon^{-1}\log|\mathbb{X}|\right)}{\rho}\right), \ \kappa = \mathcal{O}\left(\frac{\epsilon\rho}{\overline{D}^2 \log(\epsilon^{-1}\log|\mathbb{X}|)}\right), \ \delta = \begin{cases} 0, & \gamma < 1, \\ \Omega(e^{-\sqrt{T}}), & \gamma = 1, \end{cases}$$

we have the following error bound with probability $1 - O(\epsilon)$

$$D_{\mathrm{KL}}(p_\delta \| \widehat{q}_{T-\delta}) \lesssim \exp(-\rho T)\log|\mathbb{X}| + \epsilon + \overline{D}^2 \kappa T \lesssim \epsilon,$$

and the total number of neural network evaluations is

$$N = \kappa^{-1} T = \mathcal{O}\left(\frac{\overline{D}^2 \rho^2 \log^2\left(\epsilon^{-1}\log|\mathbb{X}|\right)}{\epsilon}\right)$$

.

Convergence Guarantee for Uniformization

Theorem (Theoretical Guarantees for Uniformization)

Take block discretization scheme $(s_b)_{b\in[0,N]}$ *satisfying* $s_0 = 0, s_N = T - \delta$ *and* $s_{k+1} - s_k \le \kappa\,(1 \vee (T - s_{k+1}))$ *for* $k \in [0 : N-1]$. *Under aforementioned assumptions and the following choices of parameters*

$$T = \mathcal{O}\left(\frac{\log\left(\epsilon^{-1}\log|\mathbb{X}|\right)}{\rho}\right), \delta = \Omega(e^{-T})$$

we have the following error bound

$$D_{\mathrm{KL}}(p_\delta \| q_{T-\delta}) \lesssim \exp(-\rho T)\log|\mathbb{X}| + \epsilon \lesssim \epsilon,$$

and the total number N *of neural network evaluations satisfies*

$$\mathbb{E}[N] = \mathcal{O}\left(\frac{\overline{D}\log\left(\epsilon^{-1}\log|\mathbb{X}|\right)}{\rho}\right)$$

.

Major technique: Change of Measure for Poisson Random Measure

Theorem (Generalized Girsanov's Theorem for Poisson Random Measure)

Let $N[\lambda](\mathrm{d}t, \mathrm{d}y)$ *be a Poisson random measure with evolving intensity* $\lambda_t(y)$ *in the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *and*

$$\log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{F}_t} = \int_0^t \int_{\mathbb{X}} \log h_t(y) N[\lambda](\mathrm{d}t \times \mathrm{d}y) - \int_0^t \int_{\mathbb{X}} (h_t(y) - 1)\lambda_t(y)\nu(\mathrm{d}y).$$

Then $N[\lambda](\mathrm{d}t, \mathrm{d}y)$ *under* $\mathbb{Q}$ *is a Poisson random measure with evolving intensity* $\lambda_t(y)h_t(y)$.

Thank you for your attention!

References I

Andrew Campbell, Joe Benton, Valentin De Bortoli, Thomas Rainforth, George Deligiannidis, and Arnaud Doucet, *A continuous time framework for discrete denoising models*, Advances in Neural Information Processing Systems **35** (2022), 28266–28279.

Hongrui Chen and Lexing Ying, *Convergence analysis of discrete diffusion model: Exact implementation through uniformization*, arXiv preprint arXiv:2402.08095 (2024).

Daniel T Gillespie, *Approximate accelerated stochastic simulation of chemically reacting systems*, The Journal of chemical physics **115** (2001), no. 4, 1716–1733.

Nico M Van Dijk, *Uniformization for nonhomogeneous markov chains*, Operations research letters **12** (1992), no. 5, 283–291.