Score-Based Generative Models with Lévy Processes

Eunbi Yoon, Keehun Park, Jinhyeok Kim, Sungbin Lim *
Artificial Intelligence Graduate School, UNIST
{eunbiyoon, sungbin}@unist.ac.kr

Abstract

Time reversibility of stochastic processes is a primary cornerstone of the score-based generative models through stochastic differential equations (SDEs). While a broader class of Markov processes is reversible, previous continuous-time approaches restrict the range of noise processes to Brownian motion (BM) since the closed form of the time reversal formula is only known for diffusion processes. In this paper, we propose a class of score-based probabilistic generative models, Lévy-Itō Model (LIM), which utilizes d-dimensional α-stable distribution with independent components for noise injection. To this end, we derive an exact time reversal formula for the SDEs with Lévy processes that can allow discontinuous pure jump motion. Consequently, we advance the score-based generative models with a broad range of non-Gaussian Markov processes. Empirical results on MNIST, CIFAR-10, CelebA, and CelebA-HQ show that our approach is valid.

1 Introduction

The recent successes of score-based generative models [26, 28, 11] and their applications [20, 14, 6] draw huge attention from machine learning communities. Score-based generative models via stochastic differential equations (SDEs) [28] rely on the time reversal theory of diffusion processes, Anderson theorem [1], which shows that the time reversal of the diffusion process belongs to the class of diffusion processes again. One can interpret this result as solving a martingale problem which induces a weak solution to the reverse SDEs [9, 4]. Due to the advances in the SDE theory with jump Markov processes [13, 24, 5], one can desire a positive expectation for applying a class of non-Gaussian noise distribution to score-based generative models. However, since the closed form of the time reversal formula is only known for diffusion processes, whether a score-based method is feasible for a non-Gaussian Markov process other than a Brownian motion has been an open question in this field. To tackle the challenging problem, we propose an exact formula for the time reversal of SDEs with Lévy processes and a novel score-based generative method, Lévy-Itō Model (LIM), which utilizes d-dimensional α-stable Lévy motion with independent components as noise injection. We apply the proposed method to MNIST, CIFAR-10, CelebA, CelebA-HQ. Our approach and empirical results establish the bridge between probability theory and generative models.

2 Score-Based Generative Models with Lévy Processes

2.1 Background

1-Dimensional Symmetric α-stable distribution Let $\alpha \in (0, 2]$ be a characteristic exponent which determines the decay rate at which the tails of the distribution, and γ be a scale parameter. 1-dimensional symmetric α-stable distributions $\mathcal{S}\alpha\mathcal{S}(\gamma)$ have the heavy-tail properties $P(X > x) \sim x^{-\alpha}$ and densities with unknown closed-form expressions, except for $\alpha = 1$ or $\alpha = 2$. [2].

*Corresponding Author
[2]When $\alpha = 2$, it holds $\mathcal{S}\alpha\mathcal{S}(\gamma) = \mathcal{N}(0, \sqrt{2}\gamma)$.

NeurIPS 2022 Workshop on Score-Based Methods.

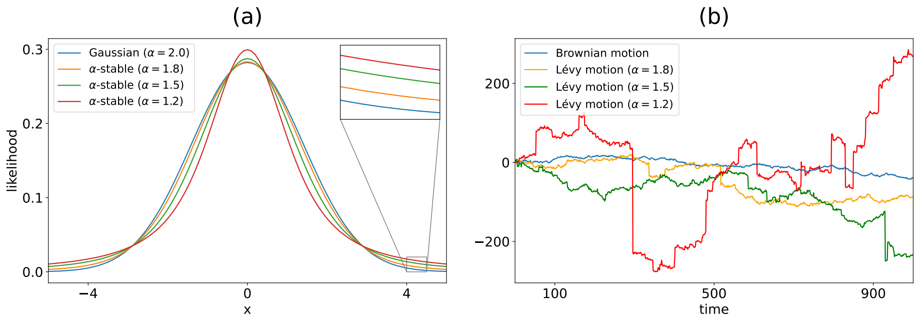


Figure 1: (a) PDF of Gaussian and α-stable distributions. The α-stable distributions have heavier tails as alpha decreases. (b) Trajectories of Brownian motion and the α-stable Lévy processes with different α. Lévy processes can have infinitely many discontinuous jumps, unlike Brownian motion.

Lévy process and α-stable Lévy motion $\mathbb{R}^d$-valued stochastic process $L_t = (L_t)_{t \geq 0}$ with $L_0 = 0$ is called Lévy process if (i) L_t has independent increments, (ii) L_t has stationary increments, (iii) L_t is stochastically continuous. If each components of the difference, $[(L_t - L_s)]_i$ and $[L_{t-s}]_i$ have the same distribution following $\mathcal{S}\alpha\mathcal{S}((t-s)^{1/\alpha})$ for $s < t$, then the Lévy process is called d-dimensional α-stable Lévy motion L_t^α. Due to the stochastic continuity (iii), Lévy processes have a countable number of discontinuous points (i.e. jumps) [33]. Notably, L_t^α is a prototypical pure jump process. The heavy-tail properties of α-stable distribution imply that the frequency of large jumps of L_t^α increases as α gets smaller (see Figure 1).

2.2 Lévy-Itō Model: Time-Reversal of SDEs driven by Lévy Processes

Due to Lévy-Ito decomposition [2, 16], we consider a family of SDEs in $\mathbb{R}^d$ driven by a Lévy process consisting of continuous Brownian motion part B_t and pure jump part L_t^α as follows:

$$d\overrightarrow{X}_t = b(t, \overrightarrow{X}_t)dt + \sigma_B(t)dB_t + \sigma_L(t)dL_t^\alpha, \quad t \in [0, 1]. \tag{1}$$

The following exact time-reversal formula is our main result.

Theorem 2.1 (Time-reversal formula of SDEs with Lévy Processes)**.** The reverse SDE of (1) is

$$d\overleftarrow{X}_t = \left(b(t, \overleftarrow{X}_t) - \sigma_B^2(t)\partial_x \log p_t(\overleftarrow{X}_t) - \alpha \cdot \sigma_L^\alpha(t) \frac{\partial_{|x|}^{\alpha-2} \nabla_x p_t(\overrightarrow{X}_t)}{p_t(\overrightarrow{X}_t)} \right) dt + \sigma_B(t)d\bar{B}_t + \sigma_L(t)d\bar{L}_t^\alpha. \tag{2}$$

where $\partial_{|x|}^{\alpha-2}(f_1(x), \ldots, f_d(x)) = (\partial_{|x_1|}^{\alpha-2} f_1(x), \ldots, \partial_{|x_d|}^{\alpha-2} f_d(x))$ is the partial fractional Riesz potential of order $\alpha - 2$ with $1 < \alpha < 2$ [19] [24] such that $\mathcal{F}[\partial_{|x_i|}^{\alpha} f](k) = |k_i|^\alpha \mathcal{F}[f](k)$ for each $i \in \{1, \ldots, d\}$, $x = (x_1, \ldots, x_d)$, $k = (k_1, \ldots, k_d)$, and $\mathcal{F}$ is the Fourier transformation. $\bar{B}_t$ and $\bar{L}_t$ is a backward Brownian motion and backward d-dimensional α-stable Lévy motion, respectively.

See Theorem A.10 for more details. We also remark that (2) recovers the result of [1] if $\alpha \to 2$. To shed light on the probabilistic approach with the jump Markov process, we propose **L**évy-**It**ō **M**odel (LIM), a novel score-based generative model through SDE driven by d-dimensional α-stable Lévy motion only ($\sigma_B(t) \equiv 0$ in (2)). Considering a beta-scheduling version of LIM, we obtain $d\overrightarrow{X}_t = -\frac{\beta(t)}{\alpha}\overrightarrow{X}_t + (\beta(t))^{1/\alpha} dL_t^\alpha$. Then the solution becomes $\overrightarrow{X}_t \overset{d}{=} a(t)\overrightarrow{X}_0 + \gamma(t)\epsilon$, where $\overset{d}{=}$ means equality in distribution, $[\epsilon]_i \sim \mathcal{S}\alpha\mathcal{S}(1)$ for each $i \in \{1, \ldots, d\}$, $a(t)$ is $\exp(-\int_0^t \frac{\beta(s)}{\alpha} ds)$ and the scale parameter $\gamma(t)$ is $(1 - a(t))^{1/\alpha}$ (see Lemma C.3). Due to the Euler-Maruyama method, we can induce a stochastic sampling of LIM (see Corollary C.4.1).

The Probability ODE We can also derive the probability ODE of (1):

$$d\overrightarrow{X}_t \overset{d}{=} \left[b(t, \overrightarrow{X}_t) - \frac{1}{2}\sigma_B^2(t)\nabla_x \log p_t(\overrightarrow{X}_t) - \sigma_L^\alpha(t) \frac{\partial_{|x|}^{\alpha-2} \nabla_x p_t(\overrightarrow{X}_t)}{p_t(\overrightarrow{X}_t)} \right] dt. \tag{3}$$

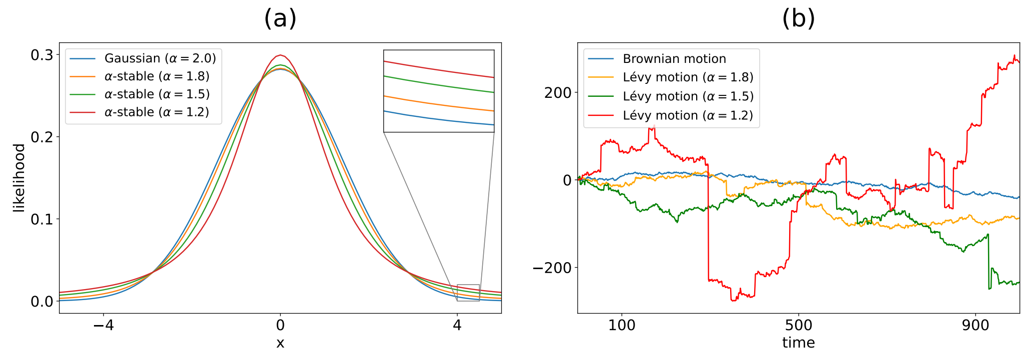


Figure 2: Comparison between (a) Lévy score and (b) ReELS. (c) and (d) the synthetic data sampled by the SDE with (2) ($\alpha = 1.5$) trained with Lévy score and ReELS, respectively. There are divergent points as indicated in the red circle since the value of Lévy score decreases for large noise.

The proof of (3) can be found in Theorem B.4. Deterministic ODE sampling of LIM can be deduced from (3) by using the Euler-Maruyama method (see Theorem C.5).

2.3 Score function for Lévy-Itō Model

Let $q_\alpha(x)$ be the product $q_\alpha(x_1)\cdots q_\alpha(x_n)$ of density functions $q_\alpha(x_i)$ of $\mathcal{S}\alpha\mathcal{S}(1)$ for $x = (x_1, \ldots, x_d)$. Recall the solution of beta-scheduling version of (2) is $\overrightarrow{X}_t \overset{d}{=} a(t)\overrightarrow{X}_0 + \gamma(t)\epsilon$ with the transition density function $p_t(x_t|x_0)$. The score function of $p_t(x_t|x_0)$ satisfies $\nabla_{x_t} \log p_t(x_t|x_0) = \nabla_\epsilon \log q_\alpha(\epsilon)/\gamma(t)$ (See Lemma D.1). We denote $S_\alpha(x) = \nabla_x \log q_\alpha(x)$. Figure 2.(a) shows that the score function of Brownian motion is linearly decreasing, while the Lévy score functions are not monotonic. Hence, if we train the score model to target the Lévy score, it is difficult to denoise the divergent large noise generated at the heavy tail (Figure 2.(b)). These phenomena worsen as α decreases.

Rectified Enhanced Lévy Score (ReELS) To denoise the large noise at the heavy tail without losing the nature of the Lévy score function, we propose **Re**ctified **E**nhanced **L**évy **S**core (ReELS) as follows:

$$\mathrm{ReELS}_\alpha(x) = \begin{cases} S_\alpha(x) & : x \in I_\alpha \\ -\mathrm{sgn}(x)\hat{c}|x|^{\hat{\beta}} & : \text{otherwise} \end{cases}, \quad \hat{\beta}(\alpha) \in (0,1). \tag{4}$$

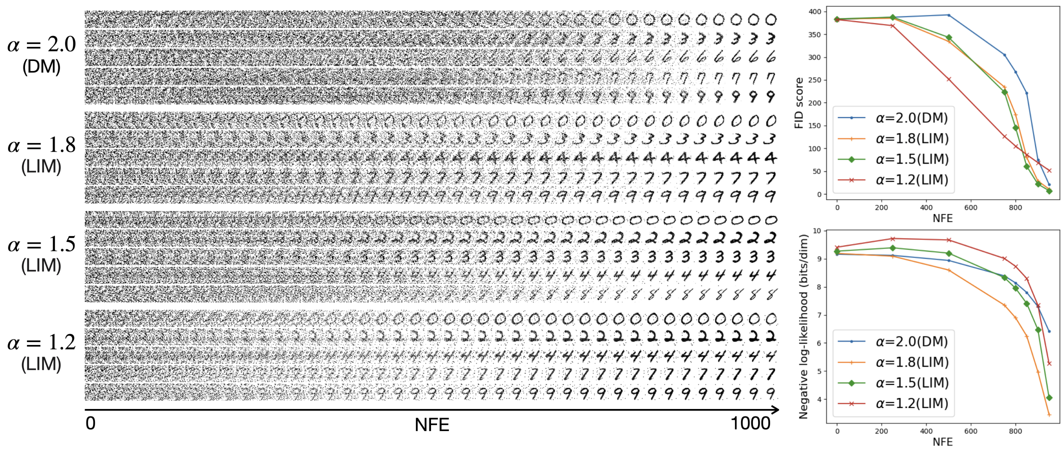
Here we set the range I_α as the interval between two local optimum points of the given Lévy score. We find parameters $\hat{c}, \hat{\beta}$ in ReELS by fitting $-\mathrm{sgn}(x)\hat{c}|x|^{\hat{\beta}}$ to the Lévy score inside I_α (see Figure 2.(b)). This procedure is equivalent to the fitting score function of a generalized Gaussian distribution to the Lévy score [17]. We remark that utilizing the BM score does not outperform ReELS for Lévy-driven SDEs because generalized Gaussian distributions have a score function more similar to the Lévy score function (see Table D.1, D.2, and D.3). The experiments on synthetic data (Mixture of Gaussian, Two-Moon, Swiss-Roll) demonstrate that LIM trained by ReELS converges to the true data distribution and performs better than using BM score for Lévy-driven SDEs.

2.4 Loss function

We use the U-net architecture [21] as in DDPM [11] and apply L_2-loss to train the model $S_\theta(x_t, t)$ using $\mathrm{ReELS}_\alpha(\epsilon)$ as a label through the Denoising Score Matching (DSM) [27] . For $[\epsilon]_i \sim \mathcal{S}\alpha\mathcal{S}(1)$ for each $i \in \{1, \ldots, d\}$ and $x_0 \sim p_{\text{data}}$, we let $x_t = a(t)x_0 + \gamma(t)\epsilon$ where $\beta(t) = \beta_0 + (\beta_1 - \beta_0)t$, $a(t) = \exp(-\frac{(\beta_1-\beta_0)}{2\alpha}t^2 - \frac{\beta_0}{\alpha}t)$, and $\gamma(t) = (1-a(t)^\alpha)^{1/\alpha}$. Let $U(0,1)$ denote a uniform distribution. Then the loss with the relative weight $\gamma(t)$ is defined as

$$L(\theta; \gamma(t)) := \mathbb{E}_{t\sim U(0,1)}\mathbb{E}_{x_0\sim p_{\text{data}}}\mathbb{E}_{\epsilon\sim\mathcal{S}\alpha\mathcal{S}}\|\gamma(t)S_\theta(x_t,t) - \mathrm{ReELS}_\alpha(\epsilon)\|_2^2 \tag{5}$$

3 Experiment



(a) MNIST Stochastic sampling in varied α values (b) Top: FID, Bottom: bits/dim

Figure 3: Generated MNIST images (a) by DM (Brownian motion, [28]) and LIM (α-stable Lévy motion for $\alpha = 1.8, 1.5, 1.2$), and corresponding plots (b) of FID score(↓) and bits per dimension (↓) for different NFEs with various α. LIM shows a faster generation speed than DM.

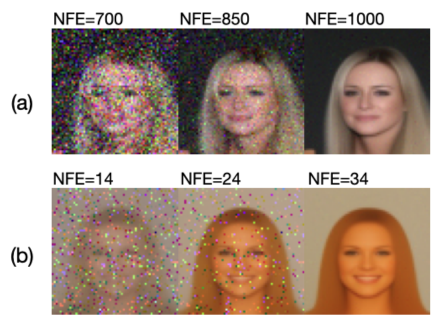


Figure 4: (a) SDE sampling results and (b) deterministic ODE sampling results with different NFEs. ODE shows a faster sampling speed than that of SDE.

We empirically validate the proposed score-based generative model on image data including MNIST (Figure E.1), CIFAR10 (Figure E.2), CelebA (Figure E.4, E.5), and CelebA-HQ (Figure E.6). We adjust the model size of each dataset for training efficiency. For $\alpha \in \{1.8, 1.5, 1.2\}$, we train our model on MNIST for 1000 epochs with $\beta_0 = 0.1$, $\beta_1 = 5.0$ and use the noise clamping to control the large-scale noise to improve the sample quality. See Section E.1 to find the other configurations in different datasets. Figure 3.(a) shows that LIM with ReELS converges faster than DM (Diffusion Models [28], $\alpha = 2.0$). See Figure 3.(b) to compare FID scores [10] and bits per dimension (bits/dim) on the MNIST dataset for each number of function evaluations (NFE) with different α. To evaluate bits/dim, we use a uniformly dequantized test dataset with 5 iterations and compute log-likelihood by using an ODE solver (See Section E.2.2). LIM achieves competitive sample quality compared to DM at $\alpha = 1.5, 1.8$, and tends to converge quickly as lower α. Figure 4 shows ReELS can be adaptive to the probability ODE, which enables fast sampling than stochastic sampling. Although the large jump of the reverse process can be controlled in the reverse SDE sampling, it is challenging to control the jump size in the deterministic ODE sampling. Hence, the image quality may degrade due to the effect of uncontrolled large noise, which leads to the higher bits per dimension in LIM 3.17 ($\alpha = 1.8$) > 1.67 ($\alpha = 2.0$).

4 Conclusion

In this paper, we broaden the range of noise distribution used in score-based generative models by inducing an approximate time reversal formula for SDEs with Lévy processes and by proposing a novel score-based generative model, Lévy-Itō Model (LIM) with Rectified Enhanced Lévy Score (ReELS). Empirical results validate that the proposed approach works well with different ranges of d-dimensional α-stable Lévy motions in synthetic datasets and various image data. Consequently, our study presents a feasible solution and demonstrates the potential for applying a broader class of non-Gaussian Markov processes to score-based generative models.

5 Acknowledgment

This work was also supported by Research Fund (1.220086.01) of UNIST, Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(No. 2022-0-00612, Geometric and Physical Commonsense Reasoning based Behavior Intelligence for Embodied AI), and National Research Foundation of Korea(NRF) funded by the Korea government(MSIT)(2021R1C1C1009256).

References

[1] Brian D.O. Anderson. Reverse-time diffusion equation models. volume 12, pages 313–326, 1982.

[2] David Applebaum. *Lévy processes and stochastic calculus*. Cambridge university press, 2009.

[3] Björn Böttcher. Feller evolution systems: Generators and approximation. *Stochastics and Dynamics*, 14(03):1350025, 2014.

[4] Patrick Cattiaux, Giovanni Conforti, Ivan Gentil, and Christian Léonard. Time reversal of diffusion processes under a finite entropy condition. *arXiv preprint arXiv:2104.07708*, 2021.

[5] Giovanni Conforti and Christian Léonard. Time reversal of markov processes with jumps under a finite entropy condition. 2021.

[6] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *arXiv preprint arXiv:2209.04747*, 2022.

[7] John R Dormand and Peter J Prince. A family of embedded runge-kutta formulae. *Journal of computational and applied mathematics*, 6(1):19–26, 1980.

[8] Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

[9] U. G. Haussmann and E. Pardoux. Time Reversal of Diffusions. *The Annals of Probability*, 14(4):1188 – 1205, 1986.

[10] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Günter Klambauer, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a nash equilibrium. *arXiv preprint arXiv:1706.08500*, 12(1), 2017.

[11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.

[12] Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 18(3):1059–1076, 1989.

[13] Jean Jacod and Philip Protter. Time Reversal on Levy Processes. *The Annals of Probability*, 16(2):620 – 641, 1988.

[14] Jihoon Kim, Jiseob Kim, and Sungjoon Choi. Flame: Free-form language-based motion synthesis & editing. *arXiv preprint arXiv:2209.00349*, 2022.

[15] Thomas G. Kurtz. Equivalence of stochastic equations and martingale problems. 2011.

[16] Paul Lévy. Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 3(3-4):337–366, 1934.

[17] Shengyang Luan, Jiayuan Li, Yinrui Gao, Jinfeng Zhang, and Tianshuang Qiu. Generalized covariance-based esprit-like solution to direction of arrival estimation for strictly non-circular signals under alpha-stable distributed noise. *Digital Signal Processing*, 118:103214, 2021.

[18] M. D. Riesz Ortigueira. Riesz potential operators and inverses via fractional centred derivatives. *International Journal ofMathematics and Mathematical Sciences*, 2006.

[19] Manuel D Ortigueira, Taous-Meriem Laleg-Kirati, and J A Tenreiro Machado. Riesz potential versus fractional laplacian. *Journal of Statistical Mechanics: Theory and Experiment*, 2014.

[20] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[22] Ludger Rueschendorf, Alexander Schnurr, and Viktor Wolf. Comparison of time-inhomogeneous markov processes, 2015.

[23] Tim Salimans and Durk P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. *Advances in neural information processing systems*, 29, 2016.

[24] Umut Şimşekli. Fractional langevin monte carlo: Exploring lévy driven stochastic differential equations for markov chain monte carlo. In *International Conference on Machine Learning*, pages 3200–3209. PMLR, 2017.

[25] John Skilling. *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*, volume 36. Springer Science & Business Media, 2013.

[26] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. 2019.

[27] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[28] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2020.

[29] Padmanabhan Sundar. Time reversal of solutions of equations driven by lévy processes. 1992.

[30] Isabelle Tristani. Fractional fokker-planck equation, 2013.

[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[32] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[33] Umut Şimşekli. Fractional langevin monte carlo: Exploring lévy driven stochastic differential equations for markov chain monte carlo, 2017.

Appendix

First, we will explain the core idea of the proof and then describe the necessary theorems. Detailed definitions are introduced later.

Ω is a probability space and $b(t,x), \sigma_B(t,x), \sigma_L(t,x)$ is a scalar function from Ω to $\mathbb{R}$ a under some smooth condition. If a $\mathbb{R}^d$-valued stochastic process $\overrightarrow{X}_t$ is a solution of a Stochastic Differential Equations(SDE) driven by Lévy process, $d\overrightarrow{X}_t = b(t,\overrightarrow{X}_t)Idt + \sigma_L(t,\overrightarrow{X}_t)IdB_t + \sigma_L(t,\overrightarrow{X}_t)IdL_L^\alpha$, the generator $\mathcal{L}_t$ satisfies

$$\mathcal{L}_t u(x) = b(t,x)\nabla u(x) + \frac{\sigma_B^2(t,x)}{2}\Delta u(x) + \int [u(x+\sigma_L(t,x)y) - u(x) - \nabla u(x)\cdot\sigma_L(t,x)y]\nu(dy) \tag{6}$$

where ν is a symmetric Lévy measure of L_t^α. If for all (t,x), $\sigma_L(t,x) > 0$, then

$$\mathcal{L}_t u(x) = b(t,x)\nabla u(x) + \frac{\sigma_B^2(t,x)}{2}\Delta u(x) \tag{7}$$

$$+ \int [u(x+y) - u(x) - \nabla u(x)\cdot y]\frac{1}{\sigma_L^d(t,s)}\nu\Big(\frac{dy}{\sigma_L(t,s)}\Big) \tag{8}$$

If $\sigma_L(t,s) = 0$, we know the exact time reversal formula [4]. So, our interest is when $\sigma_B(t,s) = 0$ and $\sigma_L(t,s) = \sigma_L(t) > 0$ such that

$$\mathcal{L}_t u(x) = b(t,x)\nabla u(x) + \int [u(x+y - u(x) - \nabla u(x)\cdot y]\frac{1}{\sigma_L^d(t)}\nu\Big(\frac{dy}{\sigma_L(t)}\Big) \tag{9}$$

We know the form of generator $\mathcal{L}_t$ of the given SDE solution $\overrightarrow{X}_t$. Therefore we can get the time-reversal formula of the operator $\mathcal{L}_t$ [5] such that

$$\overleftarrow{\mathcal{L}}_t u(x) = \overleftarrow{b}(t,x)\cdot\nabla u(x) + \int_{\mathbb{R}^n}\int [u(y+x) - u(x) - \nabla u(x)\cdot y]\frac{1}{\sigma_L^d(t)}\frac{p_t(x+y)}{p_t(x)}\nu\Big(\frac{dy}{\sigma_L(t,s)}\Big) \tag{10}$$

where $\mathbf{p}_t(dy)$ is a marginal distribution of $(\overrightarrow{X}_t)_{t\in[0,1]}$ and the backward drift $\overleftarrow{b}(t,x)$ is given by

$$b(t,x) + \overleftarrow{b}(t,x)(t,x) = \int_{\mathbb{R}^n} y\Big(1 + \frac{p_t(x+y)}{p_t(x)}\Big)\frac{1}{\sigma_L^d(t)}\nu\Big(\frac{dy}{\sigma_L(t,s)}\Big) \quad \mathbf{p_t} - \text{a.e.} \tag{11}$$

According to this [29], it can be seen that the reversal of Levy-driven SDE also appears as Levy-driven SDE. Therefore, it can be derived that 10is a generator of a solution of some SDE driven by Lévy process. But exactly what SDE does this generator follow? And $\overleftarrow{b}(t,x)$ appears as an integral equation, can the exact form be calculated? We will answer this question in Appendix A. The proof is divided into two parts. The first is to find the SDE representation of an operator $\mathcal{L}_t$ of the form 10 and the second is to derive the exact form of $\overleftarrow{b}(t,x)$.

A Time-reversal of SDE

In this chapter, given the generator of the general Markov process with jump kernel, we show that the reverse form can be deduced into an exact formula under certain conditions. Finally, we introduce the stochastic sampling and the deterministic ODE sampling of LIM. Let us outline some necessary lemmas before we move on to the proof. A homogeneous Markov process that corresponds to an inhomogeneous Markov process $\overrightarrow{X}_t$ always exists according to Lemma A.1. The existence of an SDE representation of a homogeneous Markov process with a particular generator is given in Lemma A.5. Lemma A.6 introduces the general reverse-time formula. Through the transformation of time-inhomogeneous Markov processes and an SDE representation of given a specific generator, we find the SDE representation that corresponds to the generator of the time-reverse process. From these lemmas, we deduce a reverse SDE representation, on which we also get a stochastic sampling and a deterministic ODE sampling based on probability ODE if a Markov process is provided as a solution to (1).

A.1 Time-Reversal of General Markov process with jump kernel

Let $\overrightarrow{X}_t$ be an $\mathbb{R}^d$-valued continuous time inhomogeneous Markov process on an probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The evolution system is defined as

$$T(s,t)u(x) = \mathbb{E}(u(\overrightarrow{X}_t)|\overrightarrow{X}_s = x) \text{ for } s \leq t, s, t \in [0,1]. \tag{12}$$

and this operator is well-defined on the set of Borel measurable function u on $\mathbb{R}^d$, denoted by $B(\mathbb{R}^d)$. The operator is linear and positive preserving with $T(s,t)1 = 1$ and $T(s,t) = T(s,r)T(r,t)$ for $s \leq r \leq t$. This operator is also strongly continuous such that for each $v, w \in \mathbb{R}, v \leq w$ and $s \leq t$ $\lim_{(s,t)\to(v,w)} ||U(s,t)u - U(v,w)u||_\infty = 0$ where $||\cdot||_\infty$ is the supreme norm. For all $u \in C_\infty(\mathbb{R}^d)$, the set of a continuous function with vanishing at ∞, the generators of the evolution system is given by

$$\mathcal{L}_s u = \lim_{h\to 0} \frac{T(s,s+h)u - u}{h} \text{ for each } s \in \mathbb{R}. \tag{13}$$

A family of linear operators $T(s,t)$ on C_∞ is a Feller evolution system if it is a strongly continuous, positive, contraction semigroup on C_∞.

Definition A.1 (Space-time process)**.** Let $\mathcal{B}$ be a Borel algebra in $\mathbb{R}^d$ and an a state space $(\mathbb{R}_+ \times \mathbb{R}^d, \tilde{\mathcal{B}})$ with $\tilde{x} \in \mathbb{R}_+ \times \mathbb{R}^d$ and σ-algebra $\tilde{\mathcal{B}} = \{B \in \mathbb{R}_+ \times \mathbb{R}^d | B_s \in \mathcal{B}\}$, and a new sample space $(\tilde{\Omega}, \tilde{\mathcal{A}})$ with $\tilde{w} = (s,w) \in \mathbb{R}_+ \times \Omega = \tilde{\Omega}$ and $\tilde{\mathcal{A}} = \{A \subset \mathbb{R}_+ \times \Omega | A_s \in \mathcal{A}, \forall s \in \mathbb{R}_+\}$. A space-time process $(\tilde{X}_t)$ is defined by

$$\tilde{X}_t(\tilde{w}) = (s+t, \overrightarrow{X}_{s+t}(w)). \tag{14}$$

with the probability measure for $A \in \tilde{\mathcal{A}}$ and $\tilde{x} \in \mathbb{R}_+ \times \mathbb{R}^d$ such that $\tilde{P}_{\tilde{x}}(A) = \tilde{P}(A|\tilde{X}_0 = (s,x)) \doteq P(A_s|\overrightarrow{X}_s = x)$ and the transition probabilities are given by $\tilde{P}(\tilde{X}_t \in B|\tilde{X}_0 = \tilde{x}) = \tilde{P}(\tilde{X}_t \in B|\tilde{X}_0 = (s,x)) = P(\overrightarrow{X}_{s+t} \in B_{s+t}|\overrightarrow{X}_s = x)$ where $B \in \tilde{B}, \tilde{x} \in \mathbb{R}_+ \times \mathbb{R}^d$. The transition function is defined by $\tilde{P}(t, \tilde{x}, B) = P(s, x; s+t, B_{s+t})$.

Lemma A.1. *Given a inhomogeneous Markov process (X_t), the space-time process $(\tilde{X}_t)$ on $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ is a homogeneous Markov process.*

Proof. See Transformation 3.1 in [3]. □

Lemma A.2. *Let $(\overrightarrow{X}_t)$ be the stochastic process with Feller evolution system $U(s,t)$ and the generator of $(\overrightarrow{X}_t)$ be $\mathcal{L}_s$. Let $\widetilde{X}_t$ be its space-time process with associated semigroup $T(t)$ by $T_t u(\tilde{x}) = \tilde{\mathbb{E}}(u(\tilde{X}_t)|\tilde{X}_0 = \tilde{x})$ for $\tilde{x} \in \mathbb{R}_+ \times \mathbb{R}^d$ and $u \in B_b(\mathbb{R}_+ \times \mathbb{R}^d)$. Then the extended generator $\tilde{\mathcal{L}}$ of T_t is given for all $u \in C_\infty\left([0,1] \times \mathbb{R}^d\right)$ satisfying some conditions,*

$$\tilde{\mathcal{L}}u(\widetilde{x}) = \frac{\partial}{\partial s}u(s,x) + \mathcal{L}_s u_s(x) \quad \text{where } \widetilde{x} = (s,x) \text{ and } u_s(x) = u(s,x). \tag{15}$$

Proof. See Theorem 3.2 in [3]. □

A Markov process typically has a generator that takes the form

$$\mathcal{L}u(x) = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}u(x) + b(x)\cdot\nabla u(x) \tag{16}$$

$$+ \int_{\mathbb{R}^d} \left(u(x+y) - u(x) - \mathbf{1}_{B_1}(y)y\cdot\nabla u(x)\right)\eta(x,dy). \tag{17}$$

where $b(x)$ is a locally bounded $\mathbb{R}^d$-valued function and (a_{ij}) is a locally bounded and $d \times m$ matrix-valued function, B_1 is the ball with a radius of one and a center of zero and η satisfies

$$\int 1 \wedge \left|y^2\right| \eta(x,dy) < \infty. \tag{18}$$

Suppose there exist $\lambda : \mathbb{R}^d \times S \to [0,1], \hat{\gamma} : \mathbb{R}^d \times S \to \mathbb{R}^d$, and a σ-finite measure v on a measurable space $(S, \mathcal{S})$ such that

$$\eta(x, \Gamma) = \int_S \lambda(x, y) \mathbf{1}_\Gamma(\hat{\gamma}(x, y)) \nu(dy).$$

We decompose S into $S_1 \cup S_2$ such that $1_{S_1} = 1_{B_1}(\hat{\gamma}((s,x),y))$ and $1_{S_2} = 1_{B_1^c}(\hat{\gamma}((s,x),y))$. We can rewrite the form of the generator is

$$\begin{aligned}\mathcal{L}u(x) &= \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} u(x) \\ &\quad + b(x) \cdot \nabla u(x) + \int_S \lambda(x,y) u(x, \hat{\gamma}(x,y)) - u(x) - 1_{S_1}(y) \hat{\gamma}(x, y \cdot \nabla u(x)) \nu(dy).\end{aligned}$$

Lemma A.3. Let the generator $\mathcal{L}$ be the form of (16). Let ξ be a Poisson random measure on $[0,1] \times S \times [0, \infty)$ with mean measure $m \times \nu \times m$, and let $\widetilde{\xi}(A) = \xi(A) - m \times \nu \times m(A)$. Let $(S_0, \mathcal{S}_0)$ be a measurable space, μ a σ_B-finite measure on $(S_0, \mathcal{S}_0)$ where $\sigma_B : \mathbb{R}^d \times S_0 \to \mathbb{R}^d$ satisfies $\int_{S_0} |\sigma_B(x,u)|^2 \mu(du) < \infty$ and

$$a(x) = \int_{S_0} \sigma_B(x,u) \sigma_B^T(x,u) \mu(du). \tag{19}$$

Assume that for each compact $K \subset \mathbb{R}^d$,

$$\sup_{x \in K} \Big(|b(x)| + \int_{S_0} |\sigma_B(x,u)|^2 \mu(du) + \int_{S_1} \lambda(x,u) |\hat{\gamma}(x,u)|^2 \nu(du) \tag{20}$$

$$+ \int_{S_2} \lambda(x,u) |\hat{\gamma}(x,u)| \wedge 1 \nu(du) \Big) < \infty. \tag{21}$$

Then $\overrightarrow{X}$ satisfies a stochastic differential equation of the form

$$\overrightarrow{X}_t = \overrightarrow{X}_0 + \int_0^t \int_{S_0} \sigma_B(\overrightarrow{X}_s, u) W(du \times ds) + \int_0^t b(\overrightarrow{X}_s) ds \tag{22}$$

$$+ \int_{s=0}^{s=t} \int_{u \in S_1} \int_{v=0}^{v=\lambda(\overrightarrow{X}_s, u)} \hat{\gamma}(\overrightarrow{X}_s, u) \widetilde{\xi}(dv \times du \times ds) \tag{23}$$

$$+ \int_{s=0}^{s=t} \int_{u \in S_2} \int_{v=0}^{v=\lambda(\overrightarrow{X}_s, u)} \hat{\gamma}(\overrightarrow{X}_s, u) \xi(dv \times du \times ds), \tag{24}$$

Proof. See Theorem 2.3 in [15]. $\square$

Lemma A.4. Let $\lambda((s,x),y) = \frac{p_s(x+y)}{p_s(x)} \sigma_L^\alpha(s)$ for $\sigma_L(s) \geq 0$ and $\hat{\gamma}((s,x),y)$ be $(0,y)$ and $\nu(dy)$ be a Lévy measure such that it is a Borel measure on $\mathbb{R}^d$ and $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^d} (|x|^2 \wedge 1) \nu(dx) < \infty$. If $(\overrightarrow{X}_t)$ has the corresponding generator $\mathcal{L}_t$

$$\mathcal{L}_t u(x) = b(x) \cdot \nabla u(x) + \int_{\mathbb{R}^d} [u(x+y) - u(x) - y \cdot \nabla u(x) 1_{S_1}(y)] \frac{p_t(x+y)}{p_t(x)} \sigma_L^\alpha(t) \nu(dy). \tag{25}$$

where $u \in B_b(\mathbb{R}^d)$. Then the corresponding generator $\tilde{\mathcal{L}}$ of the space-time process $\tilde{X}_t$ is

$$\tilde{\mathcal{L}} u(s,x) = (1, b(x)) \cdot \nabla u(s,x) \tag{26}$$

$$+ \int_{\mathbb{R}^d} [u((s,x) + \hat{\gamma}((s,x),y)) - u(s,x) - \gamma((s,x),y) \cdot \nabla u(s,x) 1_{S_1}(y)] \lambda((s,x),y) \nu(dy). \tag{27}$$

where $u \in C_\infty([0,1] \times \mathbb{R}^d)$.

Proof.

$$
\begin{aligned}
&\tilde{\mathcal{L}}u(s,x) = \frac{\partial}{\partial s}u(s,x) + \mathcal{L}_s u_s(x) \quad \text{ for } u_s(x) = u(s,x) \\
&= \frac{\partial}{\partial s}u(s,x) + b(x)\cdot\nabla_x u_s(x) + \int [u_s(x+y) - u_s(x) \\
&- y\cdot\nabla_x u_s(x)1_{S_1}(y)]\frac{p_t(x+y)}{p_t(x)}\sigma_L^\alpha(t)\nu(dy) \\
&= (1,b(x))\cdot\nabla u(s,x) + \int [u(s,x+y) - u(s,x) - (0,y)\cdot\nabla u(x)1_{S_1}(y)]\frac{p_t(x+y)}{p_t(x)}\sigma_L^\alpha(t)\nu(dy) \\
&= (1,b(x))\cdot\nabla u(s,x) + \int [u((s,x)+(0,y)) - u(s,x) \\
&- (0,y)\cdot\nabla u(x)1_{S_1}(y)]\frac{p_t(x+y)}{p_t(x)}\sigma_L^\alpha(t)\nu(dy).
\end{aligned}
$$

□

Theorem A.5. A generator $\mathcal{L}_t$ has a jump kernel driven by the 1-dimensional symmetric α-Levy process represented by (25). ξ be a Poisson random measure on $\mathbb{R}_+\times\mathbb{R}^d\times[0,\infty)$ with mean measure $m\times\nu\times m$ such that $\mathbb{E}[\xi(dv\times dy\times ds)] = dm\times\nu(dy)\times dm$ and $\tilde{\xi}(A) = \xi(A) - m\times\nu\times m(A)$. Then the SDE representation of the generator $\tilde{\mathcal{L}}$ satisfies

$$
\begin{aligned}
\overrightarrow{X}_t &= \overrightarrow{X}_0 + \int_0^t b(s,\overrightarrow{X}_s)ds + \int_{s=0}^{s=t}\int_{|y|<1}\int_{v=0}^{v=\frac{p_s(y+\overrightarrow{X}_t)}{p_s(\overrightarrow{X}_s)}\sigma_L^\alpha(s)} y\cdot\tilde{\xi}(dv\times dy\times ds) && (28)\\
&+ \int_{s=0}^{s=t}\int_{|y|>1}\int_{v=0}^{v=\frac{p_s(y+\overrightarrow{X}_s)}{p_s(\overrightarrow{X}_s)}\sigma_L^\alpha(s)} y\cdot\xi(dv\times dy\times ds) && (29)\\
&= \overrightarrow{X}_0 + \int_0^t b(s,\overrightarrow{X}_s)ds + \int_0^t \sigma_L(s)dL_s^\alpha. && (30)
\end{aligned}
$$

Proof. $\lambda((s,x),y)$ is $\frac{p_s(x+y)}{p_s(x)}\sigma_L^\alpha(s)$ for $\sigma_L(s)\geq 0$ and $\hat{\gamma}((s,x),y)$ is $(0,y)$ with $S_1 = \{|y|<1\}$ and $S_2 = \{|y|>1\}$. We know λ satisfies

$$
\begin{aligned}
&\int_{\mathbb{R}} \lambda((s,x),y)1_{S_1}(y)|r((s,x),y))|^2 + 1_{S_2}(y)\nu(dy) && (31)\\
&= \int_{|y|<1}\left[\frac{p_s(x+y)}{p_s(x)}\sigma_L^\alpha(s)|y|^2\nu(dy)\right]dy + \int_{|y|>1}\frac{p_s(x+y)}{p_s(x)}\sigma_L^\alpha(s)\nu(dy) < \infty. && (32)
\end{aligned}
$$

because $\int_{S_1}|y|^2\nu(dy) < \infty$ and $\int_{S_2}\nu(dy) < \infty$ with $\sup_{x,y,s}\frac{p_s(x+y)}{p_s(y)}\sigma_L^\alpha(s) < \infty$. We set $a(s,x) = 0$, so that $\sigma_B((s,x),y)$ is 0. Therefore, for any compact set $K\subset\mathbb{R}^2$,

$$
\begin{aligned}
\sup_{(s,x)\in K}\Big(&|b(x)| + \int_{S_1}\lambda((s,x),y)|r((s,x),y)|^2\nu(du) \\
&+ \int_{S_2}\int \lambda((s,x),y)|r((s,x),y)|\wedge\nu(du)\Big) < \infty.
\end{aligned}
$$

since $\int_{S_1}\lambda((s,x),y)|\hat{\gamma}((s,x),y)|^2\nu(du) + \int_{S_2}\int\lambda((s,x),y)|\hat{\gamma}((s,x),y)|\wedge\nu(du))$ is well-defined and continuous with respect to (s,x) and $b(s,x)$ is locally bounded $\mathbb{R}$-valued function. We can apply Lemma $A.2$ to the transformed homogeneous generator $\tilde{\mathcal{L}}$ of the inhomogeneous generator $\mathcal{L}_t$ from Lemma $A.4$. Now, we define $Y_t = \int_{s=0}^{s=t}\int_{|y|<1}\int_{s=0}^{s=\frac{p_s(y+\overrightarrow{X}_s)}{p_s(\overrightarrow{X}_s)}\sigma_L^\alpha(s)} y\cdot\tilde{\xi}(dv\times dy\times$

$ds) + \int_{s=0}^{s=t}\int_{|y|>1}\int_{v=0}^{v=\frac{p_s(y+\overrightarrow{X}_s)}{p_s(\overrightarrow{X}_s)}\sigma_L^\alpha(s)} y\cdot\xi(dv\times dy\times ds)$ and $Z_t = \int_0^t \sigma_L(s)dL_s^\alpha$. If we show $\mathbb{E}[\exp(i(u, Y_t)) = \mathbb{E}[\exp(i(u, Z_t))$, then we can conclude $dX_t = b(t, X(t))dt + \sigma_L^\alpha(t)dL_t^\alpha$.

$$
\begin{aligned}
\mathbb{E}[\exp(i(u, Y_t))] &= \mathbb{E}\Big[\exp(i(u, \int_{s=0}^{s=t}\int_{|y|<1}\int_{v=0}^{v=\frac{p_s(y+\overrightarrow{X}_s)}{p_s(\overrightarrow{X}_s)}\sigma_L^\alpha(s)} y\cdot\tilde{\xi}(dv\times dy\times ds)\\
&+\int_{s=0}^{s=t}\int_{|y|>1}\int_{v=0}^{v=\frac{p_t(y+\overrightarrow{X}_t)}{p_t(\overrightarrow{X}_t)}\sigma_L^\alpha(t)} y\cdot\xi(dv\times dy\times ds))\Big].
\end{aligned}
$$

Since jumps y occur countably many,

$$
\frac{p_s(y+\overrightarrow{X}_s)}{p_s(\overrightarrow{X}_s)}\sigma_L^\alpha(s) = \frac{p_s(\Delta Y_t+\overrightarrow{X}_s)}{p_s(\overrightarrow{X}_s)}\sigma_L^\alpha(s) = \sigma_L^\alpha(s). \text{ (a.e)} \tag{33}
$$

for each $t\in[0,1]$. Thus,

$$
\begin{aligned}
\mathbb{E}[\exp(i(u, Y_t))] &= \mathbb{E}\Big[\exp(i(u, \int_{s=0}^{s=t}\int_{|y|<1}\int_{v=0}^{v=\sigma_L^\alpha(s)} y\cdot\tilde{\xi}(dv\times dy\times ds)\\
&+\int_{s=0}^{s=t}\int_{|y|>1}\int_{v=0}^{v=\sigma_L^\alpha(s)} y\cdot\xi(dv\times dy\times ds))\Big]\\
&=\exp\Big(\int_0^t\int_{\mathbb{R}}\int_0^{\sigma_L^\alpha(s)}(e^{i\langle u,y\rangle}-1-i\langle u,v\rangle\cdot 1_{|y|<1}(y))dm(v)\times d\nu(y)\times dm(s)\Big)\\
&=\exp\Big(\int_0^t\int_{\mathbb{R}}\sigma_L^\alpha(s)(e^{i\langle u,y\rangle}-1-i\langle u,v\rangle\cdot 1_{|y|<1}(y))dm(s)\times d\nu(y)\Big)\\
&=\exp(-|u|^\alpha\cdot\int_0^t\sigma_L^\alpha(s)ds)\\
&=\mathbb{E}[\exp(i(u, Z_t))].
\end{aligned}
$$

Since the characteristic function uniquely determines the probability distribution, we conclude $Y_t = Z_t$ for almost everywhere (a.e). If L_t^α is a d-dimensional α-stable Lévy motion, then we can apply A.5on the each component $[L_t^\alpha]_i$ of L_t^α for $i\in\{1,\dots,d\}$. □

So far, we have proven that SDE representations can be found for inhomogeneous Markov processes that satisfy certain conditions. Afterward, we will examine how the time reversal of a generator appears when a homogeneous Markov process is given. Then Lemma A.5 is used to obtain the score-based reverse formula of Lévy-driven SDE. We use the reversal formula of Theorem 5.7 in [5] to propose a new class of generative model, LIM.

Lemma A.6. Consider a Markov process $(\overrightarrow{X}_t)_{t\in[0,1]}$ with generator $\mathcal{L}_t$ defined for any function u in set of continuous functions with compact support $C_c^1(\mathbb{R}^d)$ such that $\mathcal{L}_t u(x) = b(t,x)\cdot\nabla u(x) + \int_{\mathbb{R}^n}[u(y)-u(x)-\nabla u(x)\cdot[y-x]^\delta]\overrightarrow{J}_{t,x}(dy)$, $(t,x)\in[0,T]\times\mathbb{R}^n$ for some $\delta>0$, where $b(t,x)$ is a vector field, and the jump kernel is $\overrightarrow{J}_{t,x}(dy)$. Let $[x]^\delta \doteq 1_{|x|\le\delta}x$ Then, under some hypotheses, the Markov generator $\overleftarrow{\mathcal{L}_t}$ of the time-reversed process is

$$
\overleftarrow{\mathcal{L}}_t u(x) = \overleftarrow{b}(t,x)\cdot\nabla u(x) + \int_{\mathbb{R}^n}\int[u(y)-u(x)-\nabla u(x)\cdot[y-x]^\delta\overleftarrow{J}_{t,x}(dy). \tag{34}
$$

where $\mathbf{p}_t(dy)$ is a marginal distribution of $(\overrightarrow{X}_t)_{t\in[0,1]}$ such that it satisfies $\mathbf{p}_t(dy)\overleftarrow{J}_{t,x}(dx) = \mathbf{p}_t(dx)\overrightarrow{J}_{t,x}(dy)$ for almost every t and the backward drift $\overleftarrow{b}(t,x)$ is given by

$$
b(t,x) + \overleftarrow{b}(t,x) = \int_{\mathbb{R}^n}[y-x]^\delta(\overrightarrow{J}_{t,x}+\overleftarrow{J}_{t,x})(dy) \quad \mathbf{p_t}-\text{a.e.} \tag{35}
$$

Proof. See Theorem 5.7 in [5]. $\square$

If we assume the marginal distribution has the density function $p_t(x)$ such that $\mathbf{p}_t(dx) = p_t(x)dx$ and $\overrightarrow{J}_{t,x}(dy)$ is a symmetric kernel with $\overrightarrow{J}_{t,x}(dy) = v_t(y-x)dy$ for some symmetric Lévy measure v_t that is a Borel measure such that $v_t(\{0\}) = 0$ and $\int 1 \wedge |y|^2 v_t(dy) < \infty$ for each t. Then $\overleftarrow{J}_{t,x}(dy) = \frac{p_t(y)}{p_t(x)} v_t(y-x)dy$. Therefore (35) satisfies $b(t,x) + \overleftarrow{b}(t,x)(t,x) = \int_{|y|\leq\delta} y \cdot \frac{p_t(y+x)}{p_t(x)} v_t(y)dy$. Since v_t is symmetric, δ can be ∞ such that

$$
\begin{aligned}
\overleftarrow{\mathcal{L}}_t u(x) &= \overleftarrow{b}(t,x) \cdot \nabla u(x) + \int_{\mathbb{R}^n} \int [u(y+x) - u(x) - \nabla u(x) \cdot [y]^{\delta}] \nu_t(dy) \\
&= \overleftarrow{b}(t,x) \cdot \nabla u(x) + \int_{\mathbb{R}^n} \int [u(y+x) - u(x) - \nabla u(x) \cdot y] \nu_t(dy).
\end{aligned}
$$

Thus, if the jump kernel has the symmetric Lévy measure v_t then

$$
b(t,x) + \overleftarrow{b}(t,x)(t,x) = \int_{\mathbb{R}^n} y \cdot \frac{p_t(y+x)}{p_t(x)} \nu_t(y)dy \quad \mathbf{p_t} - \text{a.e.} \tag{36}
$$

Now, we will deal with 1-dimensional symmetric α-stable Lévy motion for simplicity. 1-dimensional α-stable Lévy motion have the symmetric Lévy measure ν of the symmetric α-stable distribution in $\mathbb{R}$ such that $\nu(dy) = \frac{C}{|y|^{1+\alpha}} dy$ with $C = \frac{\Gamma(\alpha+1)\sin(\alpha\pi/2)}{\pi}$. So, we can use 36 to estimate the reverse drift term $\overleftarrow{b}(t,x)$.

Since d-dimensional Lévy motion consists of independent components of 1-dimensional symmetric α-stable Lévy motions, we can easily extend 1-dimensional results for any d-dimensional cases. Thus we first show our main results for the 1-dimensional cases and then extend the results for d-dimensional cases by applying the results component wisely.

Lemma A.7. If a $\mathbb{R}$-valued stochastic process $(\overrightarrow{X}_t)$ is a solution to $d\overrightarrow{X}_t = -\frac{\beta(t)}{\alpha}\overrightarrow{X}_t + (\beta(t))^{1/\alpha} dL_t^{\alpha}$ then the jump kernel of $\overrightarrow{X}_t$ is

$$
\overrightarrow{J}_t(x,dy) = \frac{\Gamma(\alpha+1)\sin(\alpha\pi/2)}{\pi} \frac{\sigma_L^{\alpha}(t)dy}{|y-x|^{\alpha+1}}. \tag{37}
$$

Proof. See Lemma in [22]. $\square$

By Lemma A.7, the reverse drift term of the $\mathbb{R}$-valued solution $(\overrightarrow{X}_t)$ to $d\overrightarrow{X}_t = -\frac{\beta(t)}{\alpha}\overrightarrow{X}_t + (\beta(t))^{1/\alpha} dL_t^{\alpha}$ satisfies

$$
b(t,x) + \overleftarrow{b}(t,x)(t,x) = \frac{\Gamma(\alpha+1)\sin(\alpha\pi/2)\sigma_L^{\alpha}(t)}{\pi} \int_{\mathbb{R}^n} y \cdot \frac{p_t(y+x)}{p_t(x)} \frac{1}{|y|^{1+\alpha}} dy \quad \mathbf{p_t} - \text{a.e.} \tag{38}
$$

Therefore, the Markov generator $\overleftarrow{\mathcal{L}_t}$ of $(\overleftarrow{X}_t)$ is the form of (25). So, we can use Theorem A.5 to $\overleftarrow{\mathcal{L}_t}$ such that the reverse SDE of $\overleftarrow{X}_t$ is $d\overleftarrow{X}_t = -\overleftarrow{b}(t,x)dt + \sigma_L^{\alpha}(t)dL_t^{\alpha}$. Now, we will calculate the exact form of $\overleftarrow{b}(t,x)$ represented by the integral. For that, we derive a useful equation.

Lemma A.8. $\int_0^{\infty} \frac{\sin x}{x^{\alpha}} dx = \cos(\frac{\pi\alpha}{2}) \cdot \Gamma(1-\alpha)$.

Lemma A.9. $\int_{-\infty}^{\infty} \frac{y}{|y|^{\alpha+1}} e^{-i(u,y)} dy = -2 \cdot iu|u|^{\alpha-2} \cos(\pi\alpha/2)\Gamma(1-\alpha)$.

Proof. Let $uy = k$. If $u > 0$,

$$
\int_{-\infty}^{\infty} \frac{y}{|y|^{\alpha+1}} e^{-i(u,y)} dy = |u|^{\alpha-1} \int_{-\infty}^{\infty} \frac{k}{|k|^{\alpha+1}} e^{ik} dk.
$$

If $u < 0$,

$$
\int_{-\infty}^{\infty} \frac{y}{|y|^{\alpha+1}} e^{-i(u,y)} dy = -|u|^{\alpha-1} \int_{-\infty}^{\infty} \frac{k}{|k|^{\alpha+1}} e^{ik} dk.
$$

Therefore,

$$\int \frac{y}{|y|^{\alpha+1}} e^{-i(u,y)} dy = -\mathrm{sgn}(u)|u|^{\alpha-1} \int_{-\infty}^{\infty} \frac{k}{|k|^{\alpha+1}} e^{ik} dk$$
$$= -2iu|u|^{\alpha-2} \int_0^{\infty} \frac{\sin k}{k^\alpha} dk = -2 \cdot iu|u|^{\alpha-2} \cos(\frac{\pi\alpha}{2})\Gamma(1-\alpha).$$

□

Theorem A.10. If $d\overrightarrow{X}_t = b(t, \overrightarrow{X}_t)dt + \sigma_L(t) dL_t^\alpha$ then the reverse SDE with respect to backward integral is $d\overleftarrow{X}_t = -\overleftarrow{b}(t, \overleftarrow{X}_t)dt^3 + \sigma(t)d\bar{L}_t^\alpha$ with $\overleftarrow{b}$ satisfying

$$b(t,x) + \overleftarrow{b}(t,x) = \sigma_L^\alpha(t) \cdot \alpha \cdot \frac{\partial_{|x|}^{\alpha-2}\partial_x p_t(x)}{p_t(x)}. \tag{39}$$

Proof.

$$\begin{aligned}
\partial_{|x|}^{\alpha-2}\partial_x p_t(x) &= -\int iu|u|^{(\alpha-2)} e^{-i(u,x)} \hat{p}_t(u) du \\
&= \frac{1}{2\cdot\cos(\pi\alpha/2)\Gamma(1-\alpha)} \int\int \frac{y}{|y|^{\alpha-2}} e^{-i(u,y+x)} \hat{p}_t(u) du dy \\
&= \frac{1}{2\cdot\cos(\pi\alpha/2)\Gamma(1-\alpha)} \int p_t(x+y) \frac{y}{|y|^{\alpha+1}} dy \\
&= \frac{\pi}{2\cdot\cos(\pi\alpha/2)\sin(\pi\alpha/2)\Gamma(\alpha+1)\Gamma(1-\alpha)} \int\int \frac{y}{|y|^{\alpha-2}} e^{-i(u,y+x)} \hat{p}_t(u) du dy \\
&= \frac{1}{\alpha} \int C \cdot p_t(x+y) \frac{y}{|y|^{\alpha+1}} dy \text{ for } C = \frac{\sin(\pi\alpha/2)\Gamma(\alpha+1)}{\pi}.
\end{aligned}$$

since $\Gamma(1-\alpha)\Gamma(\alpha) = \frac{\pi}{\sin \pi\alpha}$ and $\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha$. Thus, $b(t,x) + \overleftarrow{b}(t,x)(t,x) = \sigma_L^\alpha(t) \cdot \alpha \cdot \frac{\partial_{|x|}^{\alpha-2}\partial_x p_t(x)}{p_t(x)}$.

□

If a path measure Q on a measure space Ω is given, we denote q_t as a marginal distribution of Q. Then its forward carré du champ is the forward-adapted process defined by

$$\overrightarrow{\Gamma}_t(u,v) = \mathcal{L}_t(uv) - u\mathcal{L}_t v - u\mathcal{L}_t^u. \tag{40}$$

where $\mathrm{dom}\overrightarrow{\Gamma}_t = \{(u,v); u, v, uv \in \text{ dom } L_t\}$. And the IbP of the Carré du champ is that if $u \in \mathrm{dom}\overleftarrow{L}$ and $\overleftarrow{L}u \in L^1(\bar{q})$, then for almose every t

$$\int_{\mathbb{R}^n} \Big\{(\mathcal{L}_t u + \overleftarrow{\mathcal{L}}_t u)v + \overleftarrow{\Gamma}_t(u,v)\Big\} dq_t = 0. \tag{41}$$

By equation 41, the proof of the time reversal formula relies on the integration by parts(IbP) formula for the carré du champ. Thus the reverse formula depends on the form of the Carré du champ.

If the forward generator $\mathcal{L}_t$ can be decomposed into $\mathcal{L}_t = \mathcal{L}_t^1 + \mathcal{L}_t^2$, then its Carré de champ also can be decomposed into $\overrightarrow{\Gamma}_t(u,v) = \overrightarrow{\Gamma}_t^1(u,v) + \overrightarrow{\Gamma}_t^2(u,v)$ such that $\overrightarrow{\Gamma}_t^1(u,v)$ is the Carré du champ of $\mathcal{L}_t^1$ and $\overrightarrow{\Gamma}_t(u,v)$ is the Carré du champ of $\mathcal{L}_t^2$. Since Carreé du champ $\overrightarrow{\Gamma}_t$ is only determined by operator $\mathcal{L}_t$, and if it satisfies $\overrightarrow{\Gamma}_t(u,v) = \overrightarrow{\Gamma}_t^1(u,v) + \overrightarrow{\Gamma}_t^2(u,v)$ then

$$\int_{\mathbb{R}^n} (\overleftarrow{\mathcal{L}}_t u)v = \int_{\mathbb{R}^n} (\mathcal{L}_t u)v + \overleftarrow{\Gamma}_t(u,v) dq_t \tag{42}$$

$$= \int_{\mathbb{R}^n} (\mathcal{L}_t u)v + \int_{\mathbb{R}^n} \overrightarrow{\Gamma}_t^1(u,v) + \int_{\mathbb{R}^n} \overrightarrow{\Gamma}_t^2(u,v) dq_t. \tag{43}$$

[3] If the time flow is in the forward direction, we need to put a minus sign in front of the drift term. The minus sign is not needed if the time flow is in backward direction.

Additional drift term and others are derived from $\int_{\mathbb{R}^n} \overrightarrow{\Gamma}^1_t(u,v)dq_t$ and $\int_{\mathbb{R}^n} \overrightarrow{\Gamma}^2_t(u,v)dq_t$ and if we know each term of $\int_{\mathbb{R}^n} \overrightarrow{\Gamma}^1_t(u,v)dq_t$ and $\int_{\mathbb{R}^n} \overrightarrow{\Gamma}^2_t(u,v)dq_t$ respectively, we can get the time-reversal formula of $\mathcal{L}_t$. From this conclusion, we can induce the time-reversal formula of jump-diffusion processes.

Corollary A.10.1 (The general reversal of SDE)**.** The reverse SDE of $d\overrightarrow{X}_t = b(t,x)dt + \sigma_B(t)dB_t + \sigma_L(t)dL_t^\alpha$ is

$$d\overleftarrow{X}_t = \Big(b(t,\overleftarrow{X}_t) - \sigma_B^2(t)\nabla_x \log p_t(\overleftarrow{X}_t) - \sigma_L^\alpha(t)\cdot\alpha\cdot\frac{\partial_{|x|}^{\alpha-2}\nabla_x p_t(\overleftarrow{X}_t)}{p_t(\overleftarrow{X}_t)}\Big)dt + \sigma_B(t)d\bar{B}_t + \sigma_L(t)d\bar{L}_t^\alpha. \tag{44}$$

where $\bar{B}_t, \bar{L}_t^\alpha$ is a corresponding backward Brownian motion and backward d-dimensional α-stable Lévy motion, respectively.

B Probability ODE

This chapter introduces the fractional Fokker-Planck equation, which is an extended version of the Fokker-Planck equation in diffusion models into a symmetric α-stable distribution and obtains the existence of probability ODEs from the equation. By deriving the probability ODE with the fractional derivative, the computational formula is obtained by using first-order approximation. In order to prove the existence of probability ODE, we first define fractional calculus.

B.1 Fractional Calculus

Fractional calculus is a concept that extends the existing differentiation and has the characteristic that it satisfies (46) when Fourier transformation is performed.

Definition B.1 (Partial fractional Riesz potential)**.** For $\alpha > -1$ and $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, we define the partial fractional Riesz potential $\partial_{|x|}(f_1(x), \ldots, f_d(x))$ as follows [19] [24]:

$$\partial_{|x|}^\alpha(f_1(x), \ldots, f_d(x)) = (\partial_{|x_1|}^\alpha f_1(x), \ldots, \partial_{|x_d|}^\alpha f_d(x)). \tag{45}$$

such that

$$\mathcal{F}[\partial_{|x_i|}^\alpha f](k) = |k_i|^\alpha \mathcal{F}[f](k_1, \ldots, k_d) \text{ for } k = (k_1, \ldots, k_d). \tag{46}$$

where $\mathcal{F}$ denotes the Fourier transform.

Lemma B.1. $\partial_{|x_i|}^\alpha f(x) = -\partial_{x_i}^2 \partial_{|x_i|}^{\alpha-2} f(x) = -\partial_{x_i}\partial_{|x_i|}^{\alpha-2}\partial_{x_i} f(x)$.

Proof.

$$\mathcal{F}[\partial_{|x_i|}^\alpha f](k) = |k_i|^\alpha \mathcal{F}[f](k) = |k_i|^2|k_i|^{\alpha-2}\mathcal{F}[f](k) = \mathcal{F}[-\partial_{x_i}\partial_{|x_i|}^{\alpha-2}\partial_{x_i} f](k). \tag{47}$$

□

B.2 Stochastic Calculus for Lévy-driven Stochastic Differential Equations

Lemma B.2 (Fractional Fokker-Planck equation)**.** Given a Lévy-driven SDE, $d\overrightarrow{X}_t = b(t,\overrightarrow{X}_t)dt + \sigma(t)dL_t^\alpha$ for $dL_t^\alpha = (dL_{t,1}^\alpha, \ldots, dL_{t,d}^\alpha)$ with set of independent symmetric α-stable Lévy motions $(L_{t,i}^\alpha)_{i=1}^d$. Then the marginal distribution $p_t(x)$ satisfies fractional Fokker-Planck equation

$$\frac{\partial p_t(x)}{\partial t} = -\nabla\cdot[b(t,x)p_t(x)] - \sigma_L(t)^\alpha \sum_{i=1}^d \partial_{|x_t|}^\alpha p_t(x). \tag{48}$$

Proof. See Proposition 1 in [30] □

Corollary B.2.1 (General Fractional Fokker-Planck equation)**.** Given a Lévy-driven SDE and, $\overrightarrow{X}_t \in \mathbb{R}^d$ which satisfies

$$d\overrightarrow{X}_t = b(t, \overrightarrow{X}_t)dt + \sigma_B(t)dB_t + \sigma_L(t)dL_t^\alpha. \tag{49}$$

where $dB_t = (dB_{t,1}, \ldots, dB_{t,d})$ with set of independent Brownian motions $(B_{t,i})_{i=1}^d$, and $dL_{t,1}^\alpha, \ldots, dL_{t,d}^\alpha)$ with set of independent symmetric α-stable Lévy motions $(L_{t,i}^\alpha)_{i=1}^d$. Then the marginal distribution $p_t(x)$ satisfies General fractional-Fokker-Planck equation,

$$\frac{\partial p_t(x)}{\partial t} = -\nabla \cdot [b(t,x)p_t(x)] + \frac{\sigma_B^2(t)}{2} \sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial {x_i}^2} - \sigma_L^\alpha(t) \sum_{i=1}^d \partial_{|x_i|}^\alpha p_t(x). \tag{50}$$

Theorem B.3 (Existence of Probability ΨDE)**.** If $p_t(x)$ satisfies Fractional Fokker-Planck equation then $p_t(x)$ satisfies

$$\frac{\partial p_t(x)}{\partial t} = -\nabla \cdot [(b(t,x) - \sigma_L^\alpha(t)F(t,x))p_t(x)]. \tag{51}$$

such that $F_i(t,x) = \frac{\partial_{|x_i|}^{\alpha-2}\partial_{x_i} p_t(x)}{p_t(x)}$. So X_t satisfies the ODE,

$$d\overrightarrow{X}_t = (b(t,x) - \sigma_L^\alpha(t)F(t,x))dt. \tag{52}$$

Proof.

$$\begin{aligned}
\frac{\partial p(t,x)}{\partial t} &= -\sum_{i=1}^d \partial_{x_i}(b_i(t,x)p(t,x)) - \sum_{i=1}^d \sigma_L^\alpha(t)\partial_{|x_i|}^\alpha p(t,x) && (53)\\
&= -\sum_{i=1}^d \left[\partial_{x_i} b_i(t,x)p(t,x) + \sigma_L^\alpha(t)\partial_{|x_i|}^\alpha p(t,x)\right] && (54)\\
&= -\sum_{i=1}^d \left[\partial_{x_i} b_i(t,x)p(t,x) - \sigma_L^\alpha(t)\partial_{x_i}\partial_{|x_i|}^{\alpha-2}\partial_{x_i} p(t,x)\right] && (55)\\
&= -\sum_{i=1}^n \partial_{x_i}\left(\left[b_i(t,x) - \sigma_L^\alpha(t)\frac{\partial_{|x_i|}^{\alpha-2}\partial_{x_i}p(t,x)}{p(t,x)}\right]p(t,x)\right) && (56)\\
&= -\nabla \cdot [(b(t,x) - \sigma_L^\alpha(t)F(t,x))p(t,x)]. && (57)
\end{aligned}$$

□

Theorem B.4 (The general Probability ΨDE)**.** If $p_t(x)$ follows Fractional Fokker-Planck equation, then the transition function $p_t(x)$ satisfies

$$\frac{\partial p_t(x)}{\partial t} = -\partial_x \cdot \left[(b(t,x) - \frac{\sigma_B^2(t)}{2}\partial_x \log p_t(x) - \sigma_L^\alpha(t)F(t,x))p_t(x)\right]. \tag{58}$$

such that $F_i(t,x) = \frac{\partial_{|x_i|}^{\alpha-2}\partial_{x_i} p_t(x)}{p_t(x)}$. Therefore, $\overrightarrow{X}_t$ satisfies the ODE,

$$d\overrightarrow{X}_t \overset{d}{=} \left[b(t,\overrightarrow{X}_t) - \frac{\sigma_B^2(t)}{2}\nabla_{x_t} \log p_t(\overrightarrow{X}_t) - \sigma_L^\alpha(t)F(t,\overrightarrow{X}_t)\right]dt. \tag{59}$$

Proof.

$$\frac{\partial p(t,x)}{\partial t} = -\sum_{i=1}^{d} \partial_{x_i}(b_i(t,x)p(t,x)) + \frac{\sigma_B^2(t)}{2}\sum_{i=1}^{d}\frac{\partial^2 p(t,x)}{\partial_{x_i}^2} - \sum_{i=1}^{d}\sigma_L^\alpha(t)\partial_{|x_i|}^\alpha p(t,x) \tag{60}$$

$$= -\sum_{i=1}^{d}\left[\partial_{x_i} b_i(t,x)p(t,x) - \frac{\sigma_B^2(t)}{2}\partial_{x_i}^2 p(t,x) + \sigma_L^\alpha(t)\partial_{|x_i|}^\alpha p(t,x)\right] \tag{61}$$

$$= -\sum_{i=1}^{d}\left[\partial_{x_i} b_i(t,x)p(t,x) - \frac{\sigma_B^2(t)}{2}\partial_{x_i}^2 p(t,x_t) - \sigma_L^\alpha(t)\partial_{x_i}\partial_{|x_i|}^{\alpha-2}\partial_{x_i} p(t,x)\right] \tag{62}$$

$$= -\sum_{i=1}^{n}\partial_{x_i}\left(\left[b_i(t,x) - \frac{\sigma_B^2(t)}{2}\frac{\partial_{x_i}p(t,x)}{p(t,x)} - \sigma_L^\alpha(t)\frac{\partial_{|x_i|}^{\alpha-2}\nabla_{x_i}p(t,x)}{p(t,x)}\right]p(t,x)\right) \tag{63}$$

$$= -\nabla\cdot[(b(t,x) - \frac{\sigma_B^2(t)}{2}\partial_{x_t}\log p(t,x) - \sigma_L^\alpha(t)F(t,x))p(t,x)]. \quad \square \tag{64}$$

$\square$

C General OU process

Given a SDE driven by a d-dimensional α-stable Lévy motions L_t^α with $[L_t^\alpha]_i \sim \mathcal{S}\alpha\mathcal{S}(t^{1/\alpha})$ for each $i \in \{1,\dots,d\}$ such that

$$d\overrightarrow{X}_t = -\beta\overrightarrow{X}_t dt + (\alpha\cdot\beta)^{1/\alpha}dL_t^\alpha. \tag{65}$$

the solution of the SDE is

$$\overrightarrow{X}_t \overset{d}{=} x_0 e^{-\beta t} + (\alpha\cdot\beta)^{1/\alpha}\int_0^t e^{-\beta(t-s)}dL_s^\alpha. \tag{66}$$

Since the each component of integral $[\int_0^t e^{-\beta(t-s)}dL_s^\alpha]_i$ is also a 1-dimensional symmetric α-stable $\sim \mathcal{S}\alpha\mathcal{S}(\gamma(t))$ for some $\gamma(t)$ as $[L_t^\alpha]_i$ is a 1-dimensional symmetric α-stable Lévy motion for each $i \in \{1,\dots,d\}$. We want to find the scale parameter $\gamma(t)$ of $\int_0^t e^{-\beta(t-s)}dL_s^\alpha$ for each t.

Lemma C.1. Given α with $0 < \alpha < 2$ and f is a measurable function such that $f : [0,T] \to \mathbb{R}$ with $\int_0^T |f(s)|^\alpha ds < \infty$. Let $\mathbb{R}$-valued $\overrightarrow{X}_t = \int_0^t f(s)dL_s^\alpha$ then

$$\overrightarrow{X}_t \sim \mathcal{S}\alpha\mathcal{S}\left(\int_0^t |f(s)|^\alpha ds)^{1/\alpha}\right). \tag{67}$$

Proof. If $f(t) = \sum_{i=1}^N a_i\chi_{(t_{i-1},t_i]}$ with $t_0 = 0, t_N = t$,

$$\overrightarrow{X}_t \overset{d}{=} \int_0^t \sum_{i=1}^N a_i\chi_{(t_{i-1},t_i]}(s)dL_s^\alpha = \sum_{i=1}^n a_i[L_{t_i}^\alpha - L_{t_{i-1}}^\alpha] \overset{d}{=} \sum_{i=1}^N a_i L_{\Delta t_i}^\alpha,\ \Delta t_i = t_i - t_{i-1}. \tag{68}$$

Using the above equation,

$$\mathbb{E}[e^{(iu\overrightarrow{X}_t)}] = \mathbb{E}[e^{iu\sum_{i=1}^N a_i L_{\Delta t_i}^\alpha}] = \prod_{i=1}^N \mathbb{E}[e^{iua_i L_{\Delta i_i}^\alpha}] \tag{69}$$

$$= \prod_{i=1}^N e^{-|u|^\alpha|a_i|^\alpha\Delta t_i} = e^{-\sum_{i=1}^N |a_i|^\alpha\Delta t_i|u|^\alpha} = e^{-(\int_0^t |f(s)|^\alpha ds)|u|^\alpha}. \tag{70}$$

$\Rightarrow \overrightarrow{X}_t \sim \mathcal{S}\alpha\mathcal{S}\left(\int_0^t |f(s)|^\alpha ds)^{1/\alpha}\right)$. Let us prove that f is not a simple function. With the loss of generality, assume $f(t) \geq 0$. If not, we decompose $f(t) = f^+(t) - f^-(t)$ such that f^+, f^- are non-negative functions. Then we can construct an non-decreasing sequence of simple functions f_n such that $\lim_{n\to\infty} f_n(t) = f(t)$ for all $t \in [0,T]$. So, $\sup_n f_n(t) \leq f(t)$ for all t. Define $X_t^n \equiv \int_0^t f_n(s) dL_s^\alpha$. As $\int_0^T |f(s)|^\alpha ds < \infty$, we can use dominated convergence theorem so that $\lim_{n\to\infty} \overrightarrow{X}_t^n(w) = \overrightarrow{X}_t(w)$ for all $w \in \Omega, t \in [0,T]$.

$$\mathbb{E}[e^{(iu\overrightarrow{X}_t)}] = \lim_{n\to\infty} \mathbb{E}[e^{(iuX_t^n)}] = \lim_{n\to\infty} e^{-(\int_0^t |f_n(s)|^\alpha ds)|u|^\alpha} = e^{-(\int_0^t |f(s)|^\alpha ds)|u|^\alpha}. \tag{71}$$

$\therefore\ \overrightarrow{X}_t \sim \mathcal{S}\alpha\mathcal{S}\left(\int_0^t |f(s)|^\alpha ds)^{1/\alpha}\right)$ when f is a measurable function. This theorem can be extended for the solution of the SDE (65) driven by d-dimensional α-stable Lévy motions. □

Theorem C.2. If $a(t) = e^{-\beta t}, \gamma(t) = (1-e^{-\alpha\beta t})^{1/\alpha} = (1-(a(t))^\alpha)^{1/\alpha}$ and $\overrightarrow{X}_t = a(t)x_0 + \gamma(t)\epsilon$ for some $\epsilon \sim \mathcal{S}\alpha\mathcal{S}$ then X_t is a solution to $d\overrightarrow{X}_t = -\beta\overrightarrow{X}_t dt + (\alpha\cdot\beta)^{1/\alpha} dL_t^\alpha$ and

$$\overrightarrow{X}_t \stackrel{d}{=} x_0 e^{-\beta t} + (\alpha\cdot\beta)^{1/\alpha} \int_0^t e^{-\beta(t-s)} dL_s^\alpha. \tag{72}$$

Proof. Use Lemma C.1. □

Lemma C.3. If $\overrightarrow{X}_t$ is a solution to $d\overrightarrow{X}_t = -\frac{\beta(t)}{\alpha}\overrightarrow{X}_t dt + \beta(t)^{1/\alpha} dL_t^\alpha$, then X_t can be represented by

$$\overrightarrow{X}_t \stackrel{d}{=} e^{-\int_0^t \frac{\beta(s)}{\alpha} ds}\overrightarrow{X}_0 + \int_0^t e^{-\int_u^t \frac{\beta(s)}{\alpha} ds} \beta(u)^{1/\alpha} dL_u^\alpha. \tag{73}$$

If we define $a(t) = e^{-\int_0^t \frac{\beta(s)}{\alpha} ds}$, then the scale parameter $\gamma(t)$ of $\int_0^t e^{-\int_u^t \frac{\beta(s)}{\alpha} ds} (\beta(u)^{1/\alpha} dL_t^\alpha$ satisfies $\gamma^\alpha(t) = (1-a^\alpha(t))$. If $\beta(t) = \beta_0 + (\beta_1-\beta_0)t$ then $\mathbb{E}[X_t] = e^{-\frac{(\beta_1-\beta_0)}{2\alpha}t^2 - \frac{\beta_0 t}{\alpha}} x_0 = a(t)x_0$, with $\log a(t) = -\frac{(\beta_1-\beta_0)}{2\alpha}t^2 - \frac{\beta_0}{\alpha}t$.

Proof.

$$\begin{aligned} d(e^{\int_0^t \frac{\beta(s)}{\alpha} ds}) &= e^{\int_0^t \frac{\beta(s)}{\alpha} ds} \cdot \frac{\beta(t)}{\alpha} dt + e^{\int_0^t \frac{\beta(s)}{\alpha} ds}\Big(-\frac{\beta(t)}{\alpha}\overrightarrow{X}_t dt + (\beta(t))^{1/\alpha} dL_t^\alpha\Big) \\ &= e^{\int_0^t \frac{\beta(s)}{\alpha} ds} (\beta(t))^{1/\alpha} dL_t^\alpha. \end{aligned}$$

$\overrightarrow{X}_t = e^{-\int_0^t \frac{\beta(s)}{\alpha} ds} X_0 + \int_0^t e^{-\int_u^t \frac{\beta(s)}{\alpha} ds}\beta(u)^{1/\alpha} dL_t^\alpha$. If we set $a(t) = e^{-\int_0^t \frac{\beta(s)}{\alpha} ds}$ then $\frac{d}{dt}\log a(t) = -\frac{\beta(s)}{\alpha}$. And the scale parameter $\gamma(t)$ satisfies

$$\begin{aligned} \gamma^\alpha(t) &= \int_0^t \frac{a(t)^\alpha}{a(u)^\alpha}(\beta(u))du = \int_0^t \frac{a^\alpha(t)}{a^\alpha(u)}(-\alpha)\frac{d}{dt}\log a(u) du = a^\alpha(t)\int_0^t \frac{-\alpha}{a^\alpha(u)}\frac{a'(u)}{a(u)} du. \\ &= a^\alpha(t)\int_0^t (-\alpha)\frac{a'(u)}{a^{\alpha+1}(u)} du = a^\alpha(t)\int_0^t \frac{d}{du}(a^{-\alpha}(u)) du = a^\alpha(t)[a^{-\alpha}(t) - a^{-\alpha}(0)]. \\ &= (1-a^\alpha(t)). \end{aligned}$$

□

Theorem C.4. The partial fractional Riesz potential can be approximated by

$$\frac{\partial_{|x_i|}^{\alpha-2}\partial_{x_i} p_t(x)}{p_t(x)} \approx \frac{1}{h^{\alpha-2}}\sum_{k\in\mathbb{Z}} \frac{(-1)^k\Gamma(\alpha-1)}{\Gamma\left(\frac{\alpha}{2}-k\right)\Gamma\left(\frac{\alpha}{2}+k\right)}\partial_{x_i}\log p_t(x_1,\ldots,x_i-kh,\ldots,x_d)\left[1-kh\partial_{x_i}\log p_t(x)\right]$$

.

If we only approximate this summation on k=0, then $\frac{\partial_{|x_i|}^{\alpha-2}\partial_{x_i} p_t(x)}{p_t(x)} \approx \frac{1}{h^{\alpha-2}}\frac{\Gamma(\alpha-1)}{\Gamma(\frac{\alpha}{2})^2}\nabla\log p_t(x)$.

See Equation (4.1) in [18].

Corollary C.4.1 (Stochastic sampling of LIM)**.** When $t < s$, $\Delta t = s - t$

$$x(t) = (1 + \frac{\beta(s)}{\alpha}\Delta t)x(s) + \alpha \cdot \Big(\beta(s)\Delta t \frac{1}{h^{\alpha-2}} \frac{\Gamma(\alpha-1)}{\Gamma(\frac{\alpha}{2})^2}\Big)\nabla_x \log p_s(x(s)) + (\beta(s)\Delta t)^{1/\alpha}\epsilon. \quad (74)$$

where $[\epsilon]_i \sim \mathcal{S}\alpha\mathcal{S}(1)$ for each $i \in \{1, \ldots, d\}$.

Proposition C.1.

$$\beta(t) = -\alpha\gamma^{\alpha}(t)\frac{d\lambda(t)}{dt}. \quad (75)$$

where $\lambda(t) = \log \frac{a(t)}{\gamma(t)}$.

Theorem C.5 (Deterministic ODE sampling of LIM)**.**

$$x_t = \frac{a(t)}{a(s)}x_s + \frac{\Gamma(\alpha-1)}{\Gamma^2(\alpha/2)}\frac{\alpha}{h^{\alpha-2}}\gamma^{\alpha-1}(s)\gamma(t)(-1+e^{h_t})s_\theta(x_s, s). \quad (76)$$

Proof. We apply Euler-Maruyama method to $d\overrightarrow{X}_t = (-\frac{\beta(t)}{\alpha}\overrightarrow{X}_t - \frac{\Gamma(\alpha-1)}{\Gamma^2(\alpha/2)}\frac{\beta(t)}{h^{\alpha-2}}\nabla \log p_t(\overrightarrow{X}))dt$. For $s > t$, we can discretize the ODE such as

$$\begin{aligned}
x_t &= \frac{a(t)}{a(s)}x_s + \int_s^t \frac{a(t)}{a(u)}\frac{\Gamma(\alpha-1)}{\Gamma(\alpha/2)^2}\frac{1}{h^{\alpha-2}}\alpha\gamma^{\alpha}(u)\frac{d\lambda(u)}{du}S_\theta(x_u, u)du && (77)\\
&= \frac{a(t)}{a(s)}x_s + \frac{\Gamma(\alpha-1)}{\Gamma^2(\alpha/2)}\frac{\alpha}{h^{\alpha-2}}\int_{\lambda(s)}^{\lambda(t)} e^{-\lambda}\gamma^{\alpha-1}S_\theta(x_\lambda, \lambda)d\lambda && (78)\\
&= \frac{a(t)}{a(s)}x_s + \frac{\Gamma(\alpha-1)}{\Gamma^2(\alpha/2)}\frac{\alpha}{h^{\alpha-2}}\gamma^{\alpha-1}(s)\gamma(t)(-1+e^{h_t})S_\theta(x_s, s). && (79)
\end{aligned}$$

$\square$

D Score Function for Lévy-Itō Models

Lemma D.1. Let q_α be the density function of $\mathcal{S}\alpha\mathcal{S}$ and the value of X_t satisfies $x_t = a(t)x_0 + \gamma(t)\epsilon$ for given x_0 and $[\epsilon]_i \sim \mathcal{S}\alpha\mathcal{S}$ for each $i \in \{1, \dots, d\}$. Then the score function of the transition distribution satisfies $\nabla \log p_t(x_t|x_0) = \nabla \log q_\alpha(\epsilon)/\gamma(t)$.

Proof. Let $\overrightarrow{X}_t$ and Y be defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ where the transition density function of $\overrightarrow{X}_t$ is $p_t(x_t|x_0) = \frac{d\mathbb{P}(\overrightarrow{X}_t \le x_t | \overrightarrow{X}_t = x_0)}{dx_t}$ and the density function of Y is $q_\alpha(y) = \frac{d\mathbb{P}(Y \le y)}{dy}$. Let $\overrightarrow{X}_t = a(t)x_0 + \gamma(t)\epsilon$. Then

$$\mathbb{P}(\overrightarrow{X}_t \le x_t | \overrightarrow{X}_0 = x_0) = \mathbb{P}(a(t)x_0 + \gamma(t)Y \le x_t) \text{ since } \overrightarrow{X}_t = a(t)x_0 + \gamma(t)Y \tag{80}$$

$$= \mathbb{P}(Y \le \frac{x_t - a(t)x_0}{\gamma(t)}) \tag{81}$$

$$= \mathbb{P}(Y \le \epsilon) \tag{82}$$

Then $p_t(x_t|x_0) = \frac{d\mathbb{P}(\overrightarrow{X}_t \le x_t | \overrightarrow{X}_0 = x_0)}{dx_t} = \frac{d\mathbb{P}(Y \le \epsilon)}{d\epsilon}/\gamma(t) = q_\alpha(\epsilon)/\gamma(t)$. If we take logarithm of both sides then $\log p(x_t|x_0) = \log q_\alpha(\epsilon) - \log \gamma(t)$. Therefore, we can get $\nabla_{x_t} \log p(x_t|x_0) = \nabla_\epsilon \log q_\alpha(\epsilon)/\gamma(t)$. □

Definition D.1 (Generalized Gaussian distribution)**.** The generalized Gaussian distribution is two families of parametric probability distributions with a continuous path on $\mathbb{R}$ with a shape parameter $\tilde{\beta}$ and scale parameter $\tilde{\sigma}$ such that

$$G_{\tilde{\sigma},\tilde{\beta}}(x) = \frac{\tilde{\beta}}{2\tilde{\sigma}\Gamma(\tilde{\beta}-1)} \exp\Big(-\frac{|x|^{\tilde{\beta}}}{\tilde{\sigma}^{\tilde{\beta}}}\Big) \tag{83}$$

where $\Gamma(\cdot)$ is the Gamma function.

The score function of the Generalized Gaussian distribution is

$$\nabla_x \log G_{\tilde{\sigma},\tilde{\beta}} = -\frac{\tilde{\beta}}{\tilde{\sigma}^{\tilde{\beta}}} \mathrm{sgn}(x))|x|^{\tilde{\beta}-1} \tag{84}$$

Which is the same form of

$$\mathrm{ReELS}_\alpha(x) = -\mathrm{sgn}(x)\hat{c}|x|^{\hat{\beta}}$$

when $\hat{\beta} = \tilde{\beta} - 1$ and $\hat{c} = \frac{\tilde{\beta}}{\tilde{\sigma}^{\tilde{\beta}}}$.

D.1 ReELS

The principle behind the ReELS approaches to approximate the Lévy score function of the α-stable distribution is similar to that provided [17]. Because computing the Lévy score exactly requires higher computation complexity, score functions of generalized Gaussian distribution are employed [17] as an approximation technique. ReLES is employed with a similar concept to conduct enough denoising at large noise to allow data to converge while maintaining the heavy-tailed features. We empirically observe that $\hat{\beta}$ becomes an approximation with a value less than 1 when α is less than 2. This means that a distribution with a score similar to each Lévy score is a generalized Gaussian distribution where $\tilde{\beta}$ is less than 2.

D.2 Stochastic sampling for synthetic data

To show that the results predicted by the theory are valid, the performance of the model trained with the BM score and ReLES with synthetic data is compared with FID. Additionally, it is tested with synthetic data if the synthetic data converges to the original distribution when using ReELS for LIM. The synthetic data used as the test were two mixtures of Gaussian, Two moons, and swiss-roll. In the case of two mixtures of Gaussian, the simplest MLP with a model depth of 3 was used, and for Two Moon and swiss-roll, MLP with a model depth of 6 was used. Detailed experimental settings are described in (Table D.1, D.2, D.3). When using different score functions for each $\alpha = 1.2, 1.5, 1.8, 2$, the generation ability of LIM was compared with FID. As a result of the experiments, the mean value and variance of FID were low for the overall α in the case of using ReELS. The stochastic sampling according to the time step when different α is given can be seen in (Figure E.3, E.8, E.9).

FID $\downarrow$ \ α	1.2	1.5	1.8
Lévy Score	304192 ± 233	25410 ± 98	1966 ± 31
ReELS	**2.01** ± 0.08	**0.34** ± 0.02	**0.14** ± 0.01
BM Score	2.61 ± 0.09	0.71 ± 0.03	0.20 ± 0.01

Table D.1: FID score (mean $\pm 95\%$ CI) of stochastic sampling on synthetic data (Mixture of Gaussian). The mean values of the data distributions are (5,5), (-5,-5), respectively, and the covariance is $0.2I$. The training data is 5000 pieces and the test data is 5000 pieces. As a score model, an MLP model with a depth of 3 and a channel of [3,32,2] is used. β_0 is set to 0, β_1 is set to 10, and the clamp is set to 20. It can be seen that the FID is low when ReELS is used for all $\alpha = 1.2, 1.5$, and 1.8.

FID $\downarrow$ \ α	1.2	1.5	1.8
Lévy Score	22258 ± 182	3145 ± 62	316 ± 26
ReELS	**0.85** ± 0.013	**0.21** ± 0.0053	**0.11** ± 0.0032
BM Score	0.99 ± 0.023	0.34 ± 0.0083	0.16 ± 0.0038

Table D.2: FID score (mean $\pm 95\%$ CI) of stochastic sampling on synthetic data (Two-moon). The noise of two-moon synthetic data was set to 0.05. The training data is 5000 pieces and the test data is 5000 pieces. As a score model, an MLP model with a depth of 6 and a channel [3,32,64,64,32,2] is used. β_0 is set to 0, β_1 is set to 5, and the clamp is set to 20. It can be seen that the FID is low when ReELS is used for all $\alpha = 1.2, 1.5$, and 1.8.

FID $\downarrow$ \ α	1.2	1.5	1.8
Lévy Score	1486 ± 3.68	197.54 ± 0.73	17 ± 0.13
ReELS	1.16 ± 0.090	**0.210** ± 0.0087	**0.114** ± 0.0019
BM Score	**0.952** ± 0.11	0.44 ± 0.017	0.210 ± 0.0087

Table D.3: FID score (mean $\pm 95\%$ CI) of stochastic sampling on synthetic data (Swiss-roll). The noise of Swiss-roll synthetic data was set to 0.1. The training data is 5000 pieces and the test data is 5000 pieces. As a score model, an MLP model with a depth of 6 and a channel [3,32,64,64,32,2] is used. β_0 is set to 0, β_1 is set to 5, and the clamp is set to 20. It can be seen that the FID is low when ReELS is used for all $\alpha = 1.2, 1.5$, and 1.8.

D.3 Deterministic ODE sampling for synthetic data

In this chapter, we demonstrate the validity of the probability ODE (Theorem 3) from the deterministic ODE sampling of LIM by showing the ability to generate synthetic data. We train score models by using three synthetic data such as Two mixtures of Gaussian, Two moons, and Swiss roll. The deterministic ODE sampling according to the time step when different α is given can be seen in (Figure E.10, E.11, E.12).

FID $\downarrow \setminus \alpha$	1.2	1.5	1.8	2.0
Two mixture	30.36 ± 0.094	2.39 ± 0.0076	2.39 ± 0.0076	0.21 ± 0.0021
Two moon	35.67 ± 0.170	0.99 ± 0.028	0.99 ± 0.028	0.41 ± 0.0026
Swiss roll	30.96 ± 0.584	0.77 ± 0.0097	0.15 ± 0.0021	0.39 ± 0.0039

Table D.4: FID (mean $\pm 95\%$ CI) of deterministic ODE sampling on synthetic data (Two mixtures, Two moons, Swiss-roll). The mean values of the Mixture of Gaussian distributions are (5,5), (-5,-5), respectively, and the covariance is $0.2I$. The training data is 5000 pieces and the test data is 5000 pieces. β_0 is set to 0, β_1 is set to 10, and the clamp is set to 20. It can be seen that the FID is low when ReELS is used for all $\alpha = 1.2, 1.5$, and 1.8.

E Dataset Experiment

E.1 Implementation Detail

Our diffusion model is U-Net[21] following DDPM[11], which replaces weight normalization[23] with group normalization[32] for simple implementation. We set the model size suitable for the dataset, such that MNIST (28×28) is [16, 32, 64], CIFAR10 (32×32) is [128, 256, 256, 256], CelebA (64×64) is [128, 256, 256, 256, 1024], and CelebA-HQ (256×256) is [128, 256, 256, 256, 1024, 1024], but fix the number of residual blocks with 2 in each resolution level, and add self-attention block only in 16×16 resolution level. Continuous diffusion time $t \in [0, 1)$ is injected into the model through Transformer sinusoidal position embedding[31] after adding with 0.0001, and we use swish function as the activation function.

We train our MNIST model used in experiments for 1000 epochs with batch size 128, CIFAR10 model for 250 epochs with batch size 128, CelebA model for 140 epochs with batch size 128, and CelebA-HQ model for 160 epochs with batch size 32. All training and experiments are conducted on NVIDIA A100 GPU and NVIDIA GeForce RTX 3090, and we tune the batch size for sampling adjusted for computation resources. Because the target distribution of our model is α-stable distribution, sample quality is very sensitive to hyperparameter setting according to the α scale. So we improve sample quality by tuning hyperparameters to be optimized for each dataset:

- Though DDPM[11] used linear noise schedule with fixed $\beta_0 = 0.1$, $\beta_1 = 20$, we tuned β_0, β_1 for each α because variance of α-stable distribution depends on α scale. For MNIST dataset, we fixed the β schedule as $\beta_0 = 0.1$, $\beta_1 = 5$ in all α values. In CIFAR10/CelebA/CelebA-HQ, we chose $\beta_1 = 20$ for $\alpha = 1.8$, and $\beta_1 = 15$ for $\alpha = 1.5$ to optimize convergence into sample space, and fix β_0 to 0.1.
- Different from Gaussian distribution, α-stable distribution can have large-scale noise at lower α values, which leads to sample quality degradation. To prevent this problem, we used noise clamping as a heuristic in the training and sampling phase. It consists of 3 clamps, clamp(training, init sample, SDE sample), and we adjusted the scale of clamps suitable for each dataset:
  - 10, 10, 10 at $\alpha = 1.8$, and 100, 50, 100 at $\alpha = 1.5, 1.2$ for MNIST.
  - 30, 30, 30 at $\alpha = 1.8$, and 100, 50, 100 at $\alpha = 1.5$ for CIFAR10.
  - 10, 10, 10 at $\alpha = 1.8$, and 50, 50, 50 at $\alpha = 1.5$ for CelebA and CelebA-HQ.
  - In the case of deterministic ODE sampling, we only used clamp(training, init sample), and set them to 10, 5.

E.2 Evaluation Metric

E.2.1 FID(Fréchet Inception Distance) score

To evaluate generated sample quality, we choose the widely used FID score metric([10]), where a lower score means better sample quality. After computing both mean/variance of distributions in the training dataset and generating 50k samples by using the pre-trained Inception-V3 model, we calculate distances between two distributions as FID score.

E.2.2 Likelihood computation

Our ReELS method is adaptive to probability ODE(Figure 4), so we can compute the exact likelihood on any input data in the same way as [28]. By replacing the score $\nabla_x \log p_t(X_t)$ with score model $S_\theta(X_t, t)$, we can rewrite (3) as

$$d\overrightarrow{X}_t = \underbrace{\left(b(t, \overrightarrow{X}_t) - S_\theta(\overrightarrow{X}_t, t) \frac{\Gamma(\alpha - 1)}{\Gamma^2(\alpha/2)} \frac{\sigma_L^\alpha(t)}{h^{\alpha-2}} \right)}_{=: \tilde{f}_\theta(\overrightarrow{X}_t, t)} dt. \tag{85}$$

Then we can compute the log-likelihood of $p_0(X_0)$ such that

$$\log p_0(\overrightarrow{X}_0) = \log p_T(\overrightarrow{X}_T) + \int_0^T \nabla \cdot \tilde{f}_\theta(\overrightarrow{X}_t, t) dt \tag{86}$$

where X_T is noise mapping to X_0 which can be obtained by solving the probability ODE in (85) with ODE solver. Because of the expensive computation of $\nabla \cdot \tilde{f}_\theta(\overrightarrow{X}_t, t)$, we estimate it by using the Skilling-Hutchinson trace estimator([25], [12]), which is following [8].

To solve the integral term in (85), we choose the RK45 ODE solver[7] which can be used as `solve-ivp` function in `scipy.integrate` library. As same [8], we also set parameters `atol=1e-5`, `rtol=1e-5`. We use a test dataset applied uniform dequantization, and take the average of the bits/dim values over 5 repeats for exact likelihood computation. By changing initial time t_0 of integral $\int_{t_0}^T \nabla \cdot \tilde{f}_\theta(\overrightarrow{X}_t, t) dt$ after adding 0.001, we compute bits/dim with varied number of function evaluations(NFE) like Figure 3(b).

E.3 Additional Samples

Additional sampling results on MNIST 28×28, CIFAR10 32×32, CelebA 64×64, CelebA-HQ 256×256 are reported in below figures.

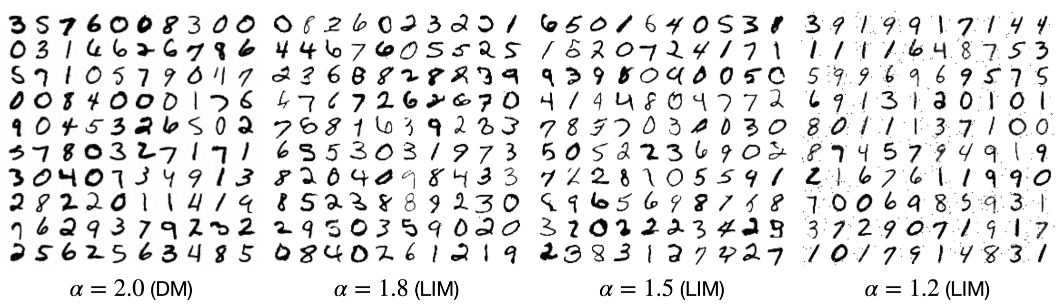


Figure E.1: Samples generated by DM(Brownian motion, [28]), and LIM($\alpha = 1.8, 1.5, 1.2$) with on MNIST(28×28) dataset.

Figure E.2: Samples generated by LIM at $\alpha = 1.8$ on CIFAR10(32×32) dataset.

Figure E.3: Samples generated by LIM at $\alpha = 1.5$ on CIFAR10(32×32) dataset.

Figure E.4: Samples generated by LIM at $\alpha = 1.8$ on CelebA(64×64) dataset.

Figure E.5: Samples generated by LIM at $\alpha = 1.5$ on CelebA(64×64) dataset.

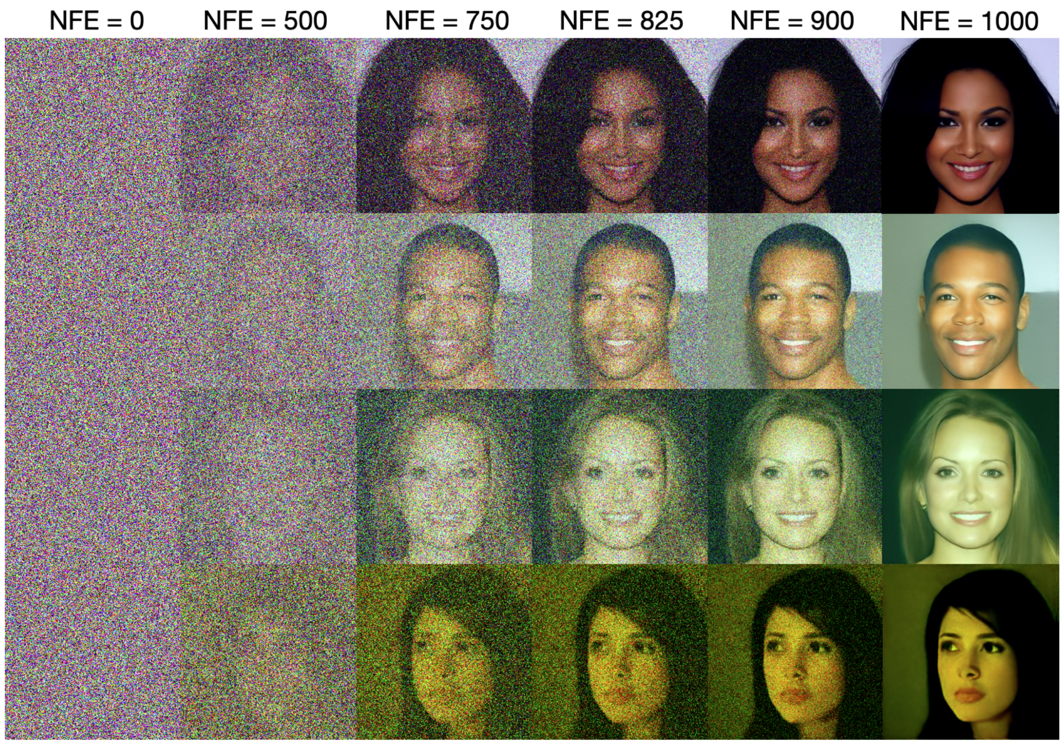


Figure E.6: Samples generated by LIM($\alpha = 1.8$) with 0, 500, 750, 825, 900, 1000 number of function evaluations(NFE) on CelebA-HQ(256×256) dataset.

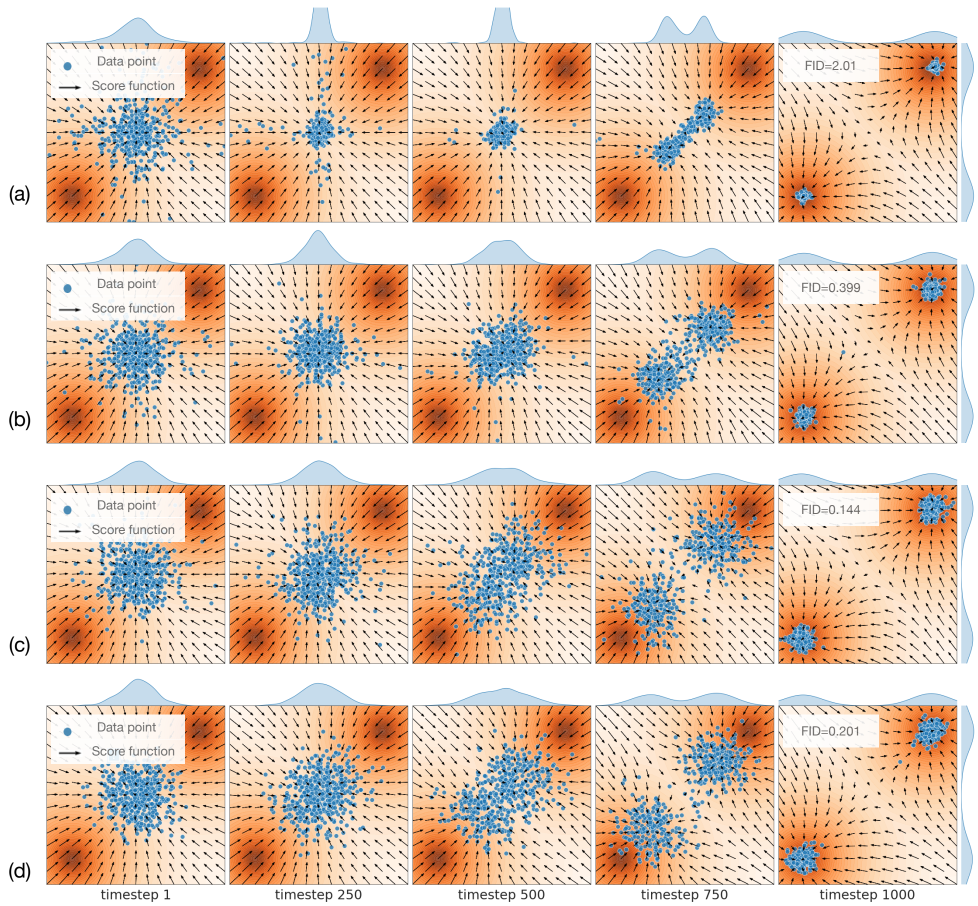


Figure E.7: Stochastic sampling(C.4.1) of two mixtures of Gaussian using ReELS for (a) $\alpha = 1.2$, (b) $\alpha = 1.5$, (c) $\alpha = 1.8$, and (d) BM-driven synthetic image. The orange color represents the original distribution of two moons.

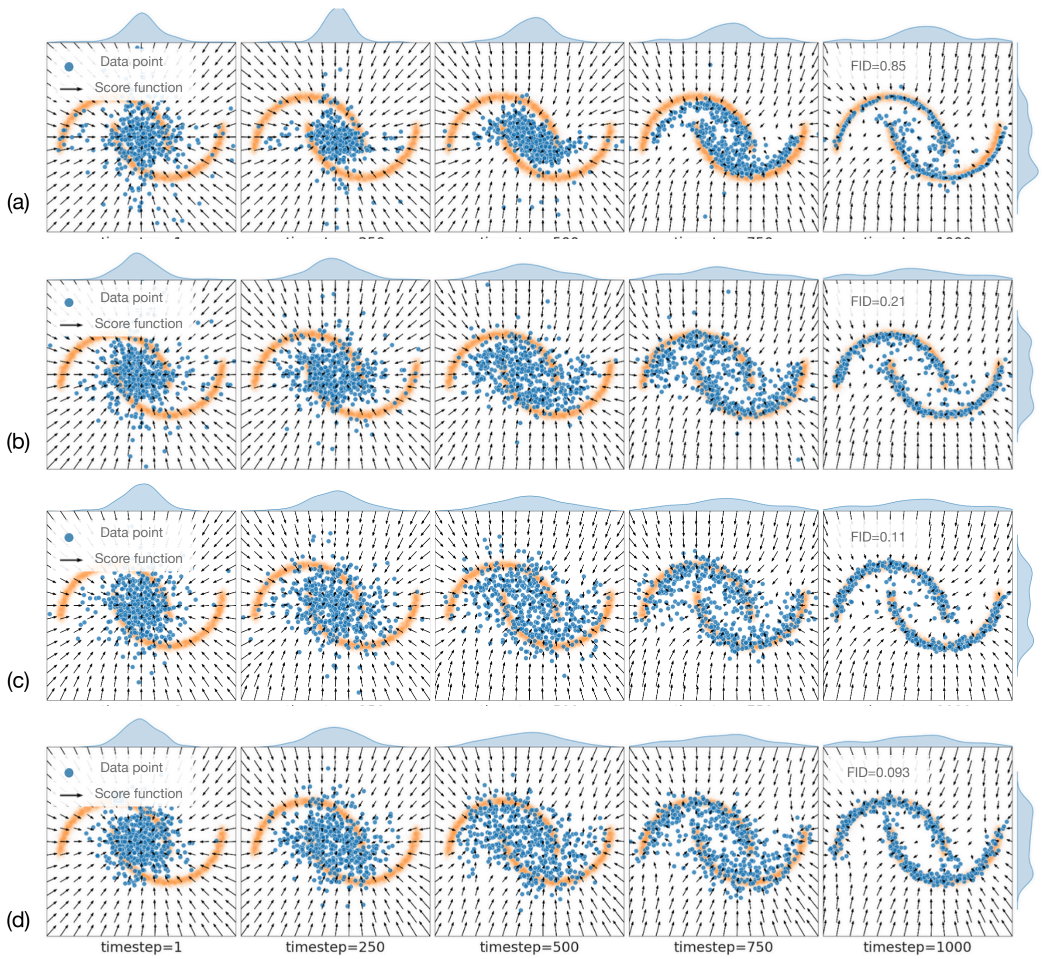


Figure E.8: Stochastic sampling(C.4.1) of two moons using ReELS for (a) $\alpha = 1.2$, (b) $\alpha = 1.5$, (c) $\alpha = 1.8$, and (d) BM-driven synthetic image. The orange color represents the original distribution of two moons.

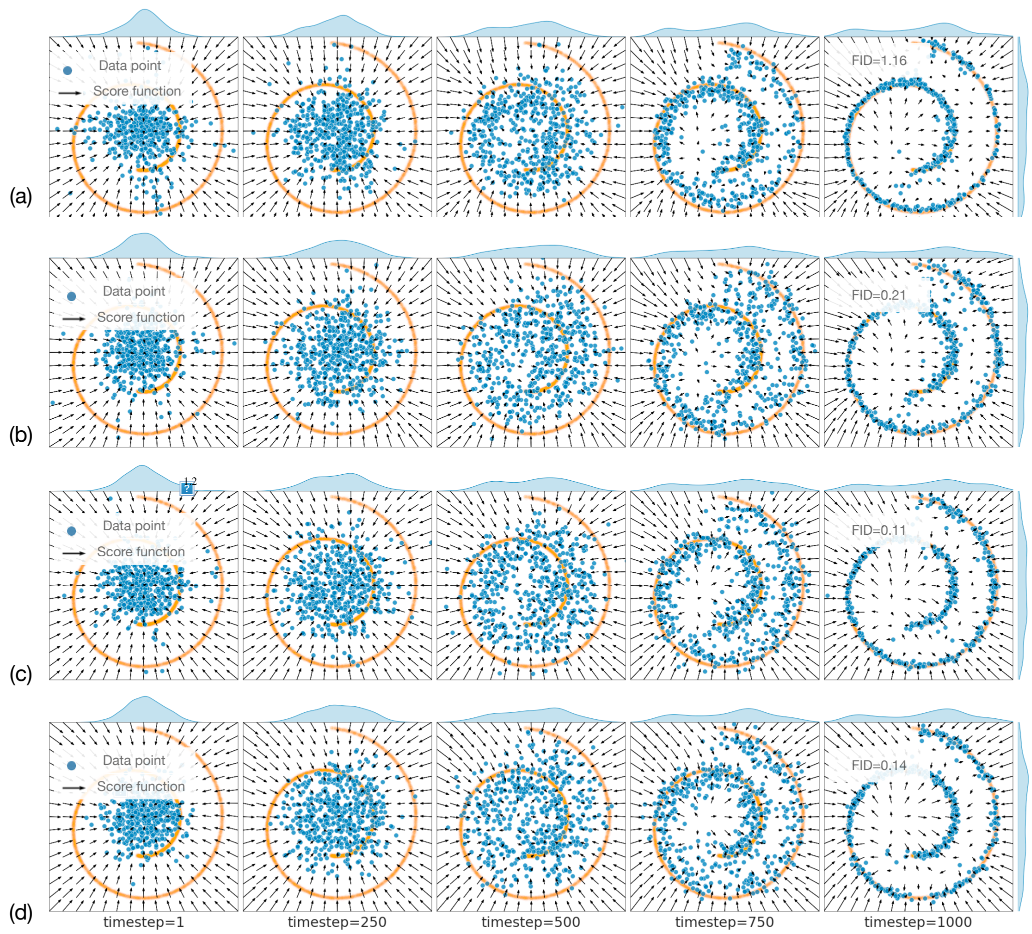


Figure E.9: Stochastic sampling(C.4.1) of swiss roll using ReELS for (a) $\alpha = 1.2$, (b) $\alpha = 1.5$, (c) $\alpha = 1.8$, and (d) BM-driven synthetic image. The orange color represents the original distribution of swiss roll.

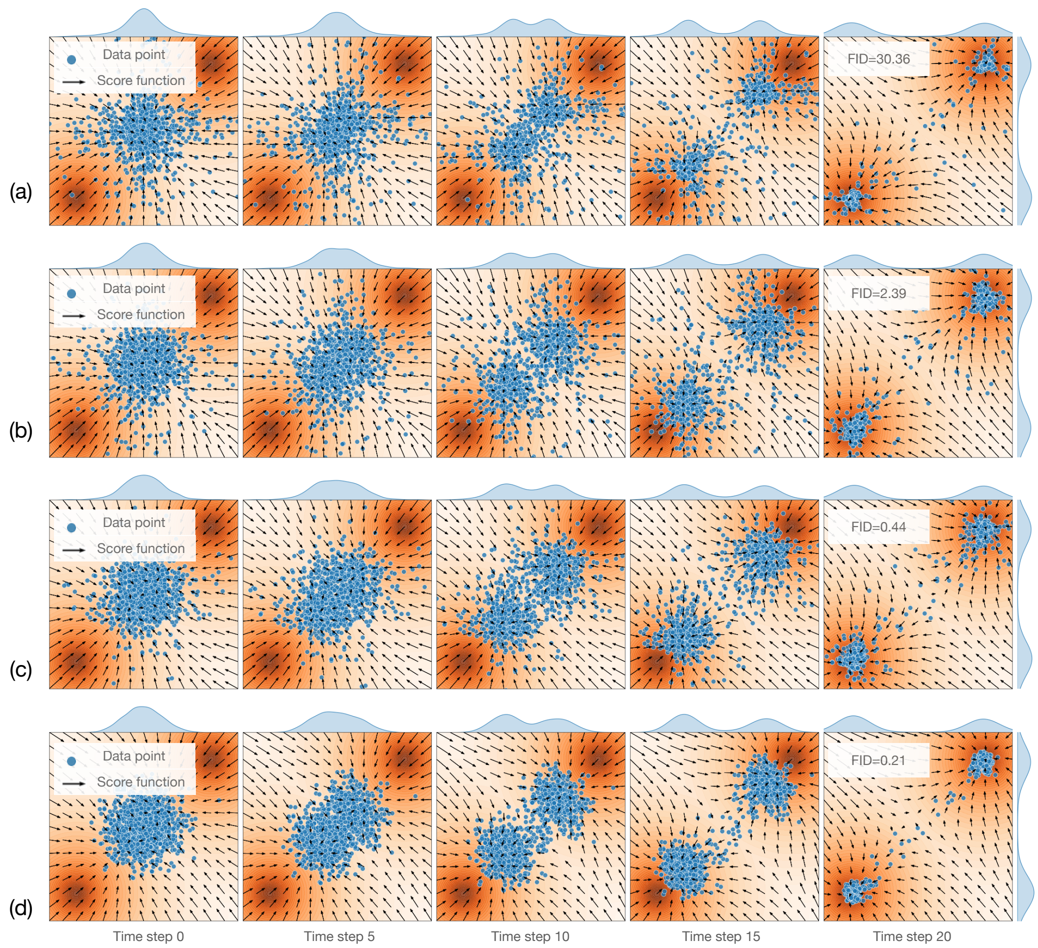


Figure E.10: Deterministic ODE sampling(C.5s) of Mixture of Gaussian synthetic data for (a) $\alpha = 1.2$, (b) $\alpha = 1.5$, (c) $\alpha = 1.8$, and (d) BM-driven synthetic image. Unlike stochastic sampling, there is a small number of points that do not converge to modes. The existence of these points is presumed to have occurred because ODE sampling cannot directly clamp the noise size in the middle of the reverse process.

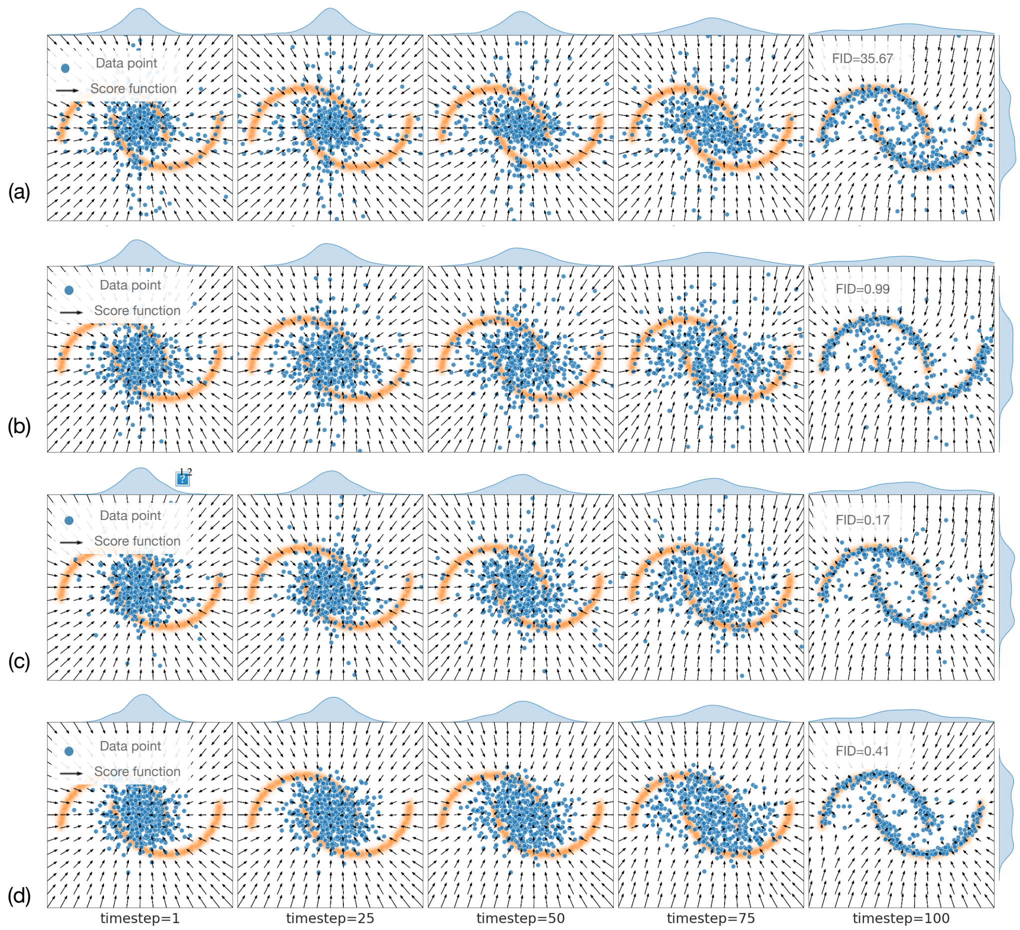


Figure E.11: Deterministic ODE sampling(C.5s) of Two Moons synthetic data for (a) $\alpha = 1.2$, (b) $\alpha = 1.5$, (c) $\alpha = 1.8$, and (d) BM-driven synthetic image. Unlike stochastic sampling, there is a small number of points that do not converge to modes. The existence of these points is presumed to have occurred because ODE sampling cannot directly clamp the noise size in the middle of the reverse process.

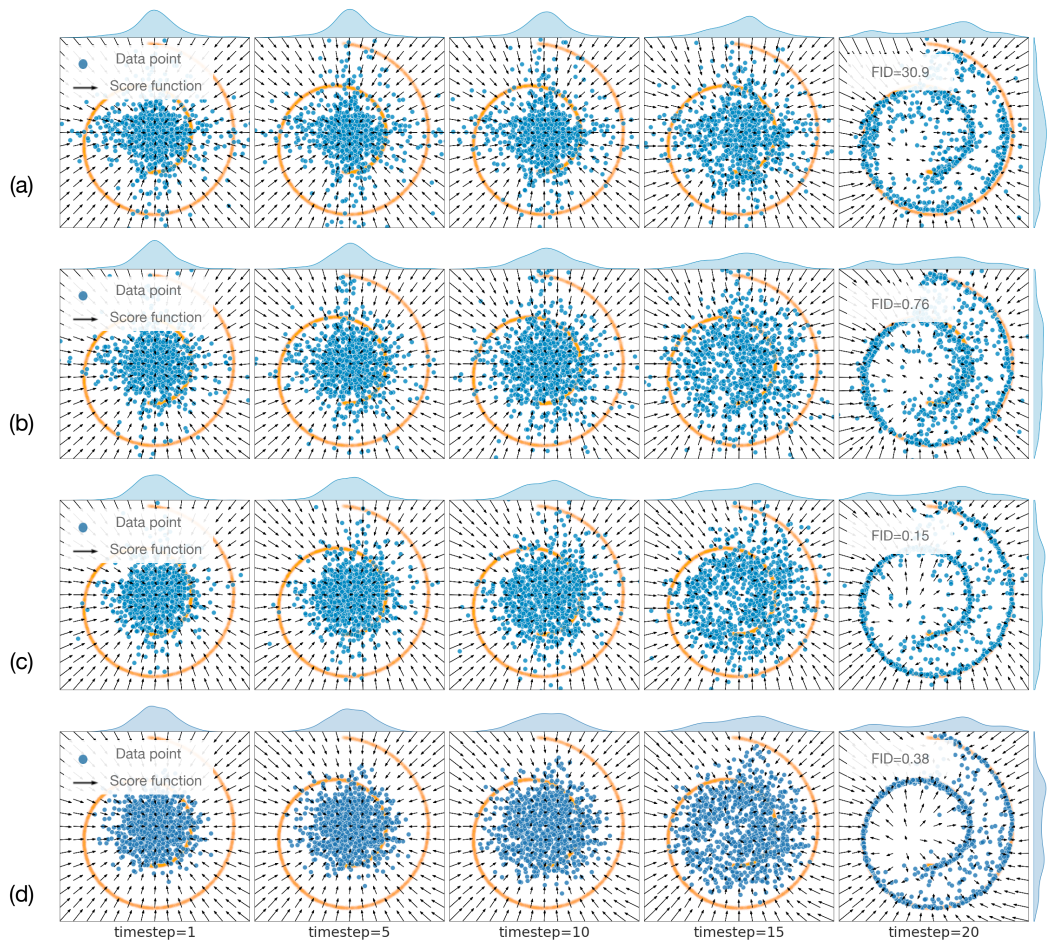


Figure E.12: Deterministic ODE sampling(C.5s) for Swiss Roll synthetic data when (a) $\alpha = 1.2$, (b) $\alpha = 1.5$, (c) $\alpha = 1.8$, and (d) BM-driven synthetic image. Unlike stochastic sampling, there is a small number of points that do not converge to modes. The existence of these points is presumed to have occurred because ODE sampling cannot directly clamp the noise size in the middle of the reverse process.