Diffusion Models for Adapted Sequential Data Generation

Haoyang Cao
Johns Hopkins University
hycao@jhu.edu

Minshuo Chen
Northwestern University
minshuo.chen@northwestern.edu

Yinbin Han
Stanford University
yinbinha@stanford.edu

Renyuan Xu
Stanford University
renyuanxu@stanford.edu

Abstract

Generating realistic synthetic sequential data is critical in real-world applications such as operations research and finance. While diffusion models have achieved remarkable success in generating *static* data, their direct extensions to sequential settings often fail to capture temporal dependencies and information structure. Designing diffusion models that can simulate sequential data in an adaptive, *non-anticipative* manner therefore remains an *open* challenge.

In this work, we propose a sequential forward–backward diffusion framework for adapted time series generation. Our approach progressively injects and removes noise along the sequence, by conditioning on the previously generated history to ensure adaptiveness. We further introduce a novel score-matching objective for efficient parallel training. Finally, we establish a score approximation result using transformer networks as an early step towards a statistical estimation theory.

1 Introduction

Diffusion models [25, 53, 54] have emerged as one of the most powerful frameworks for data generation, achieving state-of-the-art performance across domains such as text-to-image generation [5, 17, 51], molecular design [26, 60] and diffusion language models [3, 52]. By using a forward–backward stochastic system for injecting noise and denoising, diffusion models are capable of generating complex, high-dimensional samples with remarkable fidelity and diversity.

Despite these advances, most existing studies focus on *static* data, where samples are from a fixed distribution. In contrast, the application of diffusion models to sequential data remains relatively underexplored. This gap is particularly significant because in many real-world applications, especially in finance, data naturally evolves over time and is intrinsically adapted to an underlying information flow. In the example of stress testing [1, 2, 15], realistic synthetic data generation requires not only modeling temporal dependencies, but also ensuring that the generated sequences do not "look into the future" and are adaptive to the specified information flow.

A straightforward extension of diffusion models to sequential data is to inject noise into the entire sequence at once and then learn to denoise the full trajectory. Although this approach can approximate the joint distribution, it *fails* to capture the temporal dependency and probabilistic structure intrinsic to sequential data. More importantly, the generated samples are not necessarily adapted to the underlying information flow—a key requirement in sequential settings. This limitation motivates the development of a filtration-adapted sequential diffusion framework, the central focus of our paper.

Our work and contributions. In this work, we introduce a system of forward and time-reversed diffusion processes for adapted time series generation. Specifically, we design noise injection

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making.

and denoising schemes for each entry of the input sequential data, conditioning on the previously generated entries. This construction preserves the conditional distributions of the source data while ensuring that the generated samples are adapted to an enlarged filtration. Additionally, we propose a novel score matching objective that enables parallel training of the score functions across the sequence, facilitating both scalability and efficiency. As a first step towards a statistical estimation theory, we establish a score approximation result using transformer networks. To the best of our knowledge, this is the first work on applying diffusion models to sequential data generation with theoretical guarantees. The analytical framework developed in this paper advances our understanding of diffusion models and has potential to be applied in financial modeling and risk analysis.

Related literature. Our work is related to several emerging research directions.

First, a variety of generative modeling frameworks have been developed for financial time series synthesis. Prior to the emergence of diffusion models, most approaches were based on the GAN framework [15, 20, 30, 35, 39, 44, 57, 58, 62] while others adopted the VAE paradigm [1, 7, 8, 9, 16, 29, 43, 63]. More recently, diffusion models have gained significant attention for data synthesis and have been adapted for time series generation [40, 41, 45, 49, 64]. In parallel, methods based on Schrödinger Bridges have also been explored for generative modeling of time series [2, 6, 23, 59].

Our work is also closely related to the recent advances of diffusion models. A line of research focused on the convergence theory of diffusion models [4, 10, 12, 13, 14, 21, 22, 27, 28, 31, 32, 33, 34, 36, 37, 38, 42, 48, 50, 55]. In parallel, several studies analyzed the statistical properties of score matching, establishing statistical error rates and sample complexity bounds [18, 24, 47, 61, 65]. However, none of the prior work addressed the statistical estimation theory of diffusion models for sequential data, except [19], which investigated diffusion transformers restricted to (discrete) Gaussian process data.

Compared to the existing literature, our work is the first to study diffusion transformers for learning generic sequential data with statistical guarantees. More importantly, the proposed diffusion framework is designed to preserve the temporal dependencies and probabilistic structure inherent in the source data. The adaptiveness guarantee established in this work also paves the way for applications of diffusion models to domains where sequential structure is essential.

Notations. Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_h)_{h=1}^H)$ be a filtered probability space. Denote $(W_t^h)_{h=1}^H$ and $(\bar{W}_t^h)_{h=1}^H$ be $2H$ independent one-dimensional Brownian motions. For a vector $X = (X^1, \dots, X^H) \in \mathbb{R}^H$, we denote $X^{[\ell:h]} := (X^\ell, \dots, X^h)$ and $X^{[\ell:h)} := (X^\ell, \dots, X^{h-1})$. Let $\mathcal{N}(\mu, \sigma^2)$ be Gaussian distribution with mean μ and variance σ^2. Let $\sigma(X)$ be the σ-algebra generated by the random variable X and $\mathcal{A} \vee \mathcal{B}$ the smallest σ-algebra containing $\mathcal{A} \cup \mathcal{B}$. Finally, $\|\cdot\|$ is the Euclidean norm.

2 Problem Setup

We provide here an introduction to diffusion models for adapted time series generation.

Diffusion models. Let $X_0 = (X_0^h)_{h=1}^H \in \mathbb{R}^H$ be a time series of length H, adapted to the information filtration $(\mathcal{F}_h)_{h=1}^H$. Our objective is to generate additional adapted time series following the same distribution as X_0, subject to a meaningful filtration. We consider a progressively conditioned diffusion models based on the Ornstein–Uhlenbeck (OU) dynamics. Specifically, for each coordinate/timestamp h ($1 \leq h \leq H$), we will define a pair of forward and time-reversed SDEs.

For $h = 1$, the forward process follows a standard OU dynamics:

$$\mathrm{d}X_t^1 = -\tfrac{1}{2}X_t^1 \mathrm{d}t + \mathrm{d}W_t^1, \quad X_0^1 \sim \phi^1(0, \cdot), \tag{1}$$

where $\phi^1(0, \cdot)$ is the probability density of X_0^1. To generate new samples, we reverse process (1):

$$\mathrm{d}\bar{X}_t^1 = \left[\tfrac{1}{2}\bar{X}_t^1 + \partial_x \log \phi^1(T-t, \bar{X}_t^1)\right]\mathrm{d}t + \mathrm{d}\bar{W}_t^1, \quad \bar{X}_0^1 \sim \phi^1(T, \cdot), \tag{2}$$

where $\phi^1(t, \cdot)$ is the probability density of X_t^1 and $\partial_x \log \phi^1(t, \cdot)$ denotes its score function. For each $h > 1$, we construct the forward process *iteratively* by conditioning on the history:

$$\mathrm{d}X_t^h = -\tfrac{1}{2}X_t^h \mathrm{d}t + \mathrm{d}W_t^h, \quad X_0^h \sim \phi^h_{\bar{X}_T^{[1:h)}}(0, \cdot), \tag{3}$$

where $\phi_z^h(0, \cdot)$ is the probability density of X_0^h given the history $X_0^{[1:h)} = z$ and $\phi_z^h(t, \cdot)$ is the conditional density of X_t^h given $X_0^{[1:h)} = z$. Compared to (1), the forward process (3) is initialized from a conditional probability distribution, incorporating the temporal dependencies inherent in the time series. By reversing the forward process (3) in time, we obtain the time-reversed dynamics:

$$\mathrm{d}\bar{X}_t^h = \left[\tfrac{1}{2}\bar{X}_t^h + \partial_x \log \phi^h_{\bar{X}_T^{[1:h)}}(T-t, \bar{X}_t^h)\right]\mathrm{d}t + \mathrm{d}\bar{W}_t^h, \quad \bar{X}_0^h \sim \phi^h_{\bar{X}_T^{[1:h)}}(T, \cdot). \tag{4}$$

Here the time-reversed process (4) incorporates previously generated trajectories $\bar{X}_T^{[1:h)}$, ensuring that the synthesized series $\bar{X}_T = (\bar{X}_T^h)_{h=1}^H$ is generated sequentially and matches the joint distribution of original data.

Proposition 2.1 (Property of the time-reversed process) *Let* $(\bar{X}_T^h)_{h=1}^H$ *follow* (1)–(4). *Then the following distributional property holds,*

$$\bar{X}_T^1 \overset{d}{=} X_0^1, \text{ and } (\bar{X}_T^h | \bar{X}_T^{h-1} = x) \overset{d}{=} (X_0^h | X_0^{h-1} = x), \ \forall h > 1, x \in \mathbb{R}.$$

Consequently, we have $(\bar{X}_T^h)_{h=1}^H \overset{d}{=} (X_0^h)_{h=1}^H$.

Proposition 2.1 indicates that the time-reversed process we constructed exhibits the desired distributional property for the sampling step of diffusion models.

In practice, however, the time-reversed SDEs (2) and (4) cannot be directly used to generate samples, as both the score functions $\partial_x \log \phi^h_{\bar{X}_T^{[1:h)}}(t, \cdot)$ and the initial distributions $\phi^h_{\bar{X}_T^{[1:h)}}(T, \cdot)$ are unknown. To address this issue, we follow common practice to replace the initial distribution by the standard Gaussian distribution and replace the ground-truth score $\partial_x \log \phi_z^h(t, x)$ by a score estimator $\hat{s}^h(t, x, z)$. With these modifications, we obtain an approximation of the time-reversed process, which is practically implementable:

$$\mathrm{d}\tilde{X}_t^h = \Big[\tilde{X}_t^h/2 + \hat{s}^h(T-t, \tilde{X}_t^h, \tilde{X}_T^{[1:h)})\Big]\mathrm{d}t + \mathrm{d}\bar{W}_t^h, \quad \tilde{X}_0^h \sim \mathcal{N}(0, 1). \tag{5}$$

Here, by convention, we set $z = \varnothing$ when $h = 1$ and thus $\hat{s}^1(t, x, \varnothing) = \hat{s}^1(t, x)$. Using (5), we propose Algorithm 1 for sequential data sampling.

Algorithm 1 Sampling algorithm for sequential data

1: **Input:** Length of time series H, terminal time T, score networks $\{\hat{s}^h(\cdot)\}_{h=1}^H$.
2: Run the backward process up to T

$$\mathrm{d}\tilde{X}_t^1 = \Big[\tilde{X}_t^1/2 + \hat{s}^1(T-t, \tilde{X}_t^1)\Big]\mathrm{d}t + \mathrm{d}\bar{W}_t^1, \text{ with } \tilde{X}_0^1 \sim \mathcal{N}(0, 1),$$

and obtain a realization x^1.
3: **for** $h = 2, \ldots, H$ **do**
4: Run the backward process up to T

$$\mathrm{d}\tilde{X}_t^h = \Big[\tilde{X}_t^h/2 + \hat{s}^h(T-t, \tilde{X}_t^h, x^{[1:h)})\Big]\mathrm{d}t + \mathrm{d}\bar{W}_t^h, \text{ with } \tilde{X}_0^h \sim \mathcal{N}(0, 1),$$

and obtain a realization x^h.
5: Set $x^{[1:h]} = (x^{[1:h)}, x^h)$.
6: **end for**
7: **return** $x^{[1:H]}$

Score matching. To estimate the score function, a natural choice is to minimize the weighted quadratic loss for all coordinates:

$$\min_{s\in\mathcal{F}} \sum_{h=1}^H \int_{t_0}^T \tfrac{1}{T-T_0} \mathbb{E}_{\bar{X}_T^{[1:h)}} \mathbb{E}_{X_t^h} \Big[\big|s^h(t, X_t^h, \bar{X}_T^{[1:h)}) - \partial_x \log \phi^h_{\bar{X}_T^{[1:h)}}(t, X_t^h)\big|^2\Big]\mathrm{d}t, \tag{6}$$

where $\mathcal{F}$ is a function class (often neural networks). Here, t_0 is an early-stopping time to prevent the blow-up of score functions, which is commonly adopted in practice [11, 46, 54]. As $\bar{X}_T^{[1:h)}$ has the same distribution as $X_0^{[1:h)}$ for each coordinate h, one can equivalently minimize the alternative:

$$\min_{s\in\mathcal{F}} \sum_{h=1}^H \int_{t_0}^T \tfrac{1}{T-T_0} \mathbb{E}_{X_0^{[1:h)}} \mathbb{E}_{X_t^h} \Big[\big|s^h(t, X_t^h, X_0^{[1:h)}) - \partial_x \log \phi^h_{X_0^{[1:h)}}(t, X_t^h)\big|^2\Big]\mathrm{d}t. \tag{7}$$

As shown by [56], rather than minimizing the integral above, we can minimize an equivalent denoising score matching objective:

$$\min_{s\in\mathcal{F}} \sum_{h=1}^H \int_{t_0}^T \tfrac{1}{T-T_0} \mathbb{E}_{X_0^{[1:h)}} \mathbb{E}_{X_0^h} \mathbb{E}_{X_t^h} \Big[\big|s^h(t, X_t^h, X_0^{[1:h)}) - \tfrac{\alpha_t X_0^h - X_t^h}{\sigma_t^2}\big|^2\Big]\mathrm{d}t, \tag{8}$$

with $\alpha_t = e^{-t/2}$ and $\sigma_t^2 = 1 - e^{-t}$. Furthermore, we denote $s = (s^1, \ldots, s^H)$ and use the vector notation to rewrite (8) as

$$\min_{s\in\mathcal{F}} \int_{t_0}^{T} \frac{1}{T-T_0} \mathbb{E}_{X_0}\mathbb{E}_{X_t|X_0}\left[\left\|s(t, X_t, X_0) - \frac{\alpha_t X_0 - X_t}{\sigma_t^2}\right\|^2\right]\mathrm{d}t. \quad (9)$$

We remark that the objective (9) allows for training score functions in parallel in contrast to the sequential sampling in Algorithm 1. In fact, (9) can be implemented in a parallel fashion for each timestamp $h \in [1 : H]$, without knowledge of information at prior timestamps.

In practice, we approximate (9) by its empirical version. Moreover, in many applications, the temporal dependency in the time series diminishes markedly as the horizon distance increases. Therefore, we truncate the historical data to improve the estimation efficiency. Specifically, given n i.i.d. sample trajectories $\{x_i\}_{i=1}^n$ with $x_i \in \mathbb{R}^H$, we sample X_t^h given $X_0^h = x_i^h$ from $\mathcal{N}(\alpha_t x_i^h, \sigma_t^2)$. We focus on the historical window from coordinate k to h and denote the associated loss function by

$$\ell^{[k:h]}(x_i; s) = \int_{t_0}^{T} \frac{1}{T-t_0} \mathbb{E}_{X_t^h \sim \mathcal{N}(\alpha_t x_i^h, \sigma_t^2)}\left[\left|s(t, X_t^h, x_i^{[k:h)}) - \frac{\alpha_t x_i^h - X_t^h}{\sigma_t^2}\right|^2\right]\mathrm{d}t. \quad (10)$$

Here, for simplicity we assume sufficient sampling of $X_t^h|X_0^h = x_i^h$ and t, as they are easy to generate. Let $\hat{s}^h$ be a minimizer to the empirical score matching risk $\widehat{\mathcal{L}}^{[k:h]}(s) := \frac{1}{n}\sum_{i=1}^n \ell^{[k:h]}(x_i; s)$.

3 Main Results

In this section, we establish adaptiveness of the generated time series and a score approximation result.

Adaptiveness. In many applications in operations research and finance, synthetic data is expected to be adapted to a suitably chosen filtration. We make the following assumption on the score estimators.

Assumption 3.1 *For each coordinate h, the score estimator $\hat{s}^h$ is trained such that $\hat{s}^h \in \mathcal{F}_h \vee \sigma(\tilde{U}^h)$, in which $\tilde{U}^h$ is an independent random variable on the same probability space.*

$\tilde{U}^h$ can be viewed as exogenous randomness occurred in training (*e.g.,* adopting a stochastic optimization algorithm). The next proposition states the adaptiveness of the generated time series.

Proposition 3.2 (Adaptiveness) *Suppose Assumption 3.1 holds. In addition, assume the filtered probability space is rich enough to contain H independent Brownian motions $\{\bar{W}_t^h\}_{h=1}^H$, H independent random variables $\tilde{U}^h$, and H independent standard Gaussian random variables $\tilde{X}_0^h \sim \mathcal{N}(0,1)$. Then the output $(\tilde{X}_T^h)_{h=1}^H$ from Algorithm 1 is adapted to the enlarged filtration $\tilde{\mathcal{H}} := (\tilde{\mathcal{H}}_h)_{h=1}^H$ defined as, for $1 \le h \le H$*

$$\tilde{\mathcal{H}}_1 = \sigma(\tilde{X}_0^1) \vee \sigma(\{\bar{W}_t^1\}_{t=0}^T) \vee \sigma(\tilde{U}^1) \vee \mathcal{F}_1, \tilde{\mathcal{H}}_h = \tilde{\mathcal{H}}_{h-1} \vee \sigma(\tilde{X}_0^h) \vee \sigma(\{\bar{W}_t^h\}_{t=0}^T) \vee \sigma(\tilde{U}^h) \vee \mathcal{F}_h.$$

To understand Proposition 3.2, we note that the definition of enlarged filtration consists of four sources of information: 1) data filtration $\mathcal{F}_h$, 2) stochastic algorithm $\sigma(\tilde{U}^h)$, 3) initial distribution $\sigma(\tilde{X}_0^h)$, and 4) Brownian motion $\sigma(\{\bar{W}_t^h\}_{t=0}^T)$. Consequently, $\tilde{\mathcal{H}}$ can be treated as the smallest meaningful filtration that includes all the randomness occurred in training and sampling.

Score approximation. We establish an approximation theory of score functions. In this work, we consider the transformer network class as in [27]. Given input $Z \in \mathbb{R}^{d\times L}$, define self-attention layer $f^{\mathrm{SA}}(Z) := Z + \sum_{i=1}^M W_O^i(W_V^i Z)\mathrm{Softmax}((W_K^i Z)^\top W_Q^i Z)$ and feed-forward layer $f^{\mathrm{FF}}(Z) := Z + W_2\mathrm{ReLU}(W_1 Z + b_1) + b_2$. With these two layers, we define the transformer block with with positional encoding $E \in \mathbb{R}^{d\times L}$ as $f^{\mathrm{trans}}(Z) := f^{\mathrm{FF}}(f^{\mathrm{SA}}(Z+E)) : \mathbb{R}^{d\times L} \to \mathbb{R}^{d\times L}$. Denote transformer class $\mathcal{T}$ as the collection of all functions $f = f_{\mathrm{out}} \circ R^{-1} \circ f_1^{\mathrm{trans}} \circ \cdots \circ f_L^{\mathrm{trans}} \circ R \circ f_{\mathrm{in}}$, where $f_{\mathrm{in}} : \mathbb{R}^{h-k+2} \to \mathbb{R}^{h-k+2}$ and $f_{\mathrm{out}} : \mathbb{R}^{h-k+2} \to \mathbb{R}$ are linear input/output layers, and $R : \mathbb{R}^{h-k+2} \to \mathbb{R}^{d\times L}$ is a reshape layer.

To state the approximation result, we impose the following assumption on the tail probability and the smoothness of the data distribution. We denote $\phi_z^{[k:h]}(0,\cdot)$ as the conditional probability density of X_0^h given $X_0^{[k:h)} = z$.

Assumption 3.3 (Conditional sub-Gaussianity and Hölder smoothness) *For each fixed $k < h$, we assume there exist positive constants $C_1, C_2 > 0$ and a vector $v \in \mathbb{R}^{h-k}$ such that $\phi_{x_0^{[k:h)}}^{[k:h]}(0, x_0^h) \le C_1 \exp\big(-C_2|x_0^h - v^\top x_0^{[k:h)}|^2/2\big)$. Moreover, we assume the conditional density is β-Hölder smooth with its Hölder norm bounded by B.*

The following theorem states the approximation of the score function using transformer networks.

Theorem 3.4 (Score approximation) *Suppose Assumption 3.3 holds. Let $N > 0$ be a sufficiently large integer (neural network hyper-parameter) and we set $t_0 = N^{-C_\sigma}$ and $T = C_\alpha \log N$ for some positive constants C_σ and C_α. There is a transformer network $s^* \in \mathcal{T}$ such that*

$$\mathbb{E}\left[\left|s^\star(t, X_t^h, X_0^{[\ell:h)}) - \partial_x \log \phi^{[\ell:h]}_{X_0^{[\ell:h)}}(t, X_t^h)\right|^2\right] \lesssim \frac{B}{\sigma_t^4} N^{-\beta} (\log N)^{2+\beta/2}, \quad \forall t \in [t_0, T]. \tag{11}$$

Theorem 3.4 provides a transformer network to approximate the score function with a rate depending on the hyperparameter N. This result will serve as a building block for score and distribution estimation theory.

4 Conclusion

In this work, we study diffusion models for adapted time series generation. We introduce a system of forward and backward processes for time series generation and establish the adaptiveness of the resulting synthetic data sequence. In addition, we propose a novel score matching objective for training score functions efficiently in parallel. As a first step towards the statistical estimation theory, we derive a score approximation result using transformer networks. The analytical framework laid out in this paper sheds light on the application of diffusion models for learning time series. Finally, our work leaves several open questions for future investigation, including score estimation and distribution estimation theory as well as connection to financial modeling.

References

[1] Beatrice Acciaio, Stephan Eckstein, and Songyan Hou. Time-causal vae: Robust financial time series generator. *arXiv preprint arXiv:2411.02947*, 2024.

[2] Alexandre Alouadi, Baptiste Barreau, Laurent Carlier, and Huyên Pham. Robust time series generation via Schrödinger Bridge: a comprehensive evaluation. *arXiv preprint arXiv:2503.02943*, 2025.

[3] Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in neural information processing systems*, 34:17981–17993, 2021.

[4] Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *The Twelfth International Conference on Learning Representations*, 2024.

[5] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. Improving image generation with better captions. *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf*, 2(3):8, 2023.

[6] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, 2021.

[7] Hans Buehler, Blanka Horvath, Terry Lyons, Imanol Perez Arribas, and Ben Wood. Generating financial markets with signatures. RISK, 2021.

[8] Hans Bühler, Blanka Horvath, Terry Lyons, Imanol Perez Arribas, and Ben Wood. A data-driven market simulator for small data environments. *ERN: Neural Networks & Related Topics (Topic)*, 2020.

[9] Borui Cai, Shuiqiao Yang, Longxiang Gao, and Yong Xiang. Hybrid variational autoencoder for time series forecasting. *Knowledge-Based Systems*, 281:111079, 2023.

[10] Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *International Conference on Machine Learning*, pages 4735–4763. PMLR, 2023.

[11] Minshuo Chen, Kaixuan Huang, Tuo Zhao, and Mengdi Wang. Score approximation, estimation and distribution recovery of diffusion models on low-dimensional data. In *International Conference on Machine Learning*, pages 4672–4712. PMLR, 2023.

[12] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. *International Conference on Learning Representations*, 2023.

[13] Sitan Chen, Giannis Daras, and Alexandros G Dimakis. Restoration-degradation beyond linear diffusions: A non-asymptotic analysis for DDIM-type samplers. *arXiv preprint arXiv:2303.03384*, 2023.

[14] Xiuyuan Cheng, Jianfeng Lu, Yixin Tan, and Yao Xie. Convergence of flow-based generative models via proximal gradient descent in wasserstein space. *arXiv preprint arXiv:2310.17582*, 2023.

[15] Rama Cont, Mihai Cucuringu, Renyuan Xu, and Chao Zhang. Tail-gan: Learning to simulate tail risk scenarios. *Management Science*, 2025.

[16] Abhyuday Desai, Cynthia Freeman, Zuhui Wang, and Ian Beaver. Timevae: A variational auto-encoder for multivariate time series generation. *arXiv preprint arXiv:2111.08095*, 2021.

[17] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[18] Zehao Dou, Subhodh Kotekal, Zhehao Xu, and Harrison H Zhou. From optimal score matching to optimal sampling. *arXiv preprint arXiv:2409.07032*, 2024.

[19] Hengyu Fu, Zehao Dou, Jiawei Guo, Mengdi Wang, and Minshuo Chen. Diffusion transformer captures spatial-temporal dependencies: A theory for gaussian process data. *arXiv preprint arXiv:2407.16134*, 2024.

[20] Rao Fu, Jie Chen, Shutian Zeng, Yiping Zhuang, and Agus Sudjianto. Time series simulation by conditional generative adversarial net. *International Journal of Neural Networks and Advanced Applications*, 7:25–38, 2020.

[21] Xuefeng Gao and Lingjiong Zhu. Convergence analysis for general probability flow odes of diffusion models in wasserstein distances. *arXiv preprint arXiv:2401.17958*, 2024.

[22] Marta Gentiloni-Silveri and Antonio Ocello. Beyond log-concavity and score regularity: Improved convergence bounds for score-based generative models in w2-distance. *arXiv preprint arXiv:2501.02298*, 2025.

[23] Mohamed Hamdouche, Pierre Henry-Labordere, and Huyên Pham. Generative modeling for time series via Schrödinger bridge, 2023.

[24] Yinbin Han, Meisam Razaviyayn, and Renyuan Xu. Neural network-based score estimation in diffusion models: Optimization and generalization. In *The Twelfth International Conference on Learning Representations*, 2024.

[25] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[26] Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pages 8867–8887. PMLR, 2022.

[27] Jerry Yao-Chieh Hu, Weimin Wu, Yi-Chen Lee, Yu-Chao Huang, Minshuo Chen, and Han Liu. On statistical rates of conditional diffusion transformers: Approximation, estimation and minimax optimality. *arXiv preprint arXiv:2411.17522*, 2024.

[28] Daniel Zhengyu Huang, Jiaoyang Huang, and Zhengjiang Lin. Convergence analysis of probability flow ode for score-based generative models. *arXiv preprint arXiv:2404.09730*, 2024.

[29] Hongbin Huang, Minghua Chen, and Xiao Qiao. Generative learning for financial time series with irregular and scale-invariant patterns. In *The Twelfth International Conference on Learning Representations*, 2024.

[30] Adriano Koshiyama, Nick Firoozye, and Philip Treleaven. Generative adversarial networks for financial trading strategies fine-tuning and combination. *Quantitative Finance*, pages 1–17, 2020.

[31] Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence for score-based generative modeling with polynomial complexity. *Advances in Neural Information Processing Systems*, 35:22870–22882, 2022.

[32] Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In *International Conference on Algorithmic Learning Theory*, pages 946–985. PMLR, 2023.

[33] Gen Li and Yuling Yan. $O(d/T)$ convergence theory for diffusion probabilistic models under minimal assumptions. *arXiv preprint arXiv:2409.18959*, 2024.

[34] Gen Li, Yuting Wei, Yuxin Chen, and Yuejie Chi. Towards faster non-asymptotic convergence for diffusion-based generative models. *arXiv preprint arXiv:2306.09251*, 2023.

[35] Junyi Li, Xintong Wang, Yaoyang Lin, Arunesh Sinha, and Michael Wellman. Generating realistic stock market order streams. *AAAI Conference on Artificial Intelligence*, 34(01):727–734, 2020.

[36] Runjia Li, Qiwei Di, and Quanquan Gu. Unified convergence analysis for score-based diffusion models with deterministic samplers. *arXiv preprint arXiv:2410.14237*, 2024.

[37] Jiadong Liang, Zhihan Huang, and Yuxin Chen. Low-dimensional adaptation of diffusion models: Convergence in total variation. *arXiv preprint arXiv:2501.12982*, 2025.

[38] Yuchen Liang, Peizhong Ju, Yingbin Liang, and Ness Shroff. Broadening target distributions for accelerated diffusion models via a novel analysis approach. *arXiv preprint arXiv:2402.13901*, 2024.

[39] Shujian Liao, Hao Ni, Marc Sabate-Vidales, Lukasz Szpruch, Magnus Wiese, and Baoren Xiao. Sig-Wasserstein GANs for conditional time series generation. *Mathematical Finance*, 34(2): 622–670, 2024. doi: https://doi.org/10.1111/mafi.12423. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12423`.

[40] Haksoo Lim, Minjung Kim, Sewon Park, and Noseong Park. Regular time-series generation using SGM. In *Advancement of Artificial Intelligence (AAAI)*, 2023.

[41] Haksoo Lim, Minjung Kim, Sewon Park, Jaehoon Lee, and Noseong Park. TSGM: Regular and irregular time-series generation using score-based generative models, 2024.

[42] Xingchao Liu, Lemeng Wu, Mao Ye, and Qiang Liu. Let us build bridges: Understanding and extending diffusion generative models. *arXiv preprint arXiv:2208.14699*, 2022.

[43] Yuansan Liu, Sudanthi Wijewickrema, Ang Li, and James Bailey. Time-transformer aae: Connecting temporal convolutional networks and transformer for time series generation. 2022.

[44] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv:1411.1784*, 2014.

[45] Ilan Naiman, Nimrod Berman, Itai Pemper, Idan Arbiv, Gal Fadlon, and Omri Azencot. Utilizing image transforms and diffusion models for generative modeling of short and long time series. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[46] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171, 2021.

[47] Kazusato Oko, Shunta Akiyama, and Taiji Suzuki. Diffusion models are minimax optimal distribution estimators. *arXiv preprint arXiv:2303.01861*, 2023.

[48] Francesco Pedrotti, Jan Maas, and Marco Mondelli. Improved convergence of score-based diffusion models via prediction-correction. *arXiv preprint arXiv:2305.14164*, 2023.

[49] Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International Conference on Machine Learning (ICML)*, 2021.

[50] Yinuo Ren, Haoxuan Chen, Grant M Rotskoff, and Lexing Ying. How discrete and continuous diffusion meet: Comprehensive analysis of discrete diffusion models via a stochastic integral framework. *arXiv preprint arXiv:2410.03601*, 2024.

[51] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

[52] Subham Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. *Advances in Neural Information Processing Systems*, 37:130136–130184, 2024.

[53] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

[54] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2021.

[55] Wenpin Tang and Hanyang Zhao. Contractive diffusion probabilistic models. *arXiv preprint arXiv:2401.13115*, 2024.

[56] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

[57] Milena Vuletić and Rama Cont. VOLGAN: a generative model for arbitrage-free implied volatility surfaces. *Applied Mathematical Finance*, to appear, 2024.

[58] Milena Vuletić, Felix Prenzel, and Mihai Cucuringu. Fin-gan: Forecasting and classifying financial time series via generative adversarial networks. *Available at SSRN 4328302*, 2023.

[59] Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via Schrödinger bridge. In *International Conference on Machine Learning (ICML)*, 2021.

[60] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

[61] Andre Wibisono, Yihong Wu, and Kaylee Yingxi Yang. Optimal score estimation via empirical bayes smoothing. *arXiv preprint arXiv:2402.07747*, 2024.

[62] Magnus Wiese, Robert Knobloch, Ralf Korn, and Peter Kretschmer. Quant GANs: Deep generation of financial time series. *Quantitative Finance*, pages 1–22, 2020.

[63] Magnus Wiese, Ben Wood, Alexandre Pachoud, Ralf Korn, Hans Buehler, Murray Phillip, and Lianjun Bai. Multi-asset spot and option market simulation. *SSRN Electronic Journal*, 2021.

[64] Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, et al. A survey on diffusion models for time series and spatio-temporal data. *arXiv preprint arXiv:2404.18886*, 2024.

[65] Kaihong Zhang, Caitlyn H Yin, Feng Liang, and Jingbo Liu. Minimax optimality of score-based diffusion models: Beyond the density lower bound assumptions. *arXiv preprint arXiv:2402.15602*, 2024.