PROBABILITY-FLOW ODE IN INFINITE-DIMENSIONAL FUNCTION SPACES

Kunwoo Na*
Seoul National University (SNU)
Seoul, Republic of Korea
jamongna@snu.ac.kr

Junghyun Lee, Se-Young Yun
KAIST
Seoul, Republic of Korea
{jh_lee00,yunseyoung}@kaist.ac.kr

Sungbin Lim
Korea University & LG AI Research
Seoul, Republic of Korea
sungbin@korea.ac.kr

ABSTRACT

Recent advances in infinite-dimensional diffusion models have demonstrated their effectiveness and scalability in function generation tasks where the underlying structure is inherently infinite-dimensional. To accelerate inference in such models, we derive, for the first time, an analog of the probability-flow ODE (PF-ODE) in infinite-dimensional function spaces. Leveraging this newly formulated PF-ODE, we reduce the number of function evaluations while maintaining sample quality in function generation tasks, including applications to PDEs.

1 INTRODUCTION

Diffusion model (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021b; Kingma et al., 2021) is a class of generative model that adds noise to real data to train the score network and sequentially approximate the time-reversed process (Föllmer & Wakolbinger, 1986; Anderson, 1982) to generate samples from the true data distribution. This model has shown remarkable empirical success in numerous domains such as image generation (Song et al., 2021b;a), video generation (Luo et al., 2023), medical data processing (Song et al., 2022; Chung & Ye, 2022; Akrout et al., 2023), and audio generation (Kong et al., 2020).

However, "classical" diffusion models formulated on finite-dimensional Euclidean spaces limit their applicability to function generation problems as they can only generate function values realized on a fixed discretization of the function's domain (Li et al., 2020) and cannot capture functional properties of a data such as integrability or smoothness (Kerrigan et al., 2023). Motivated by such a limitation of finite-dimensional models, there has been a line of works extending the finite-dimensional diffusion model to infinite-dimensional Hilbert spaces; for instance, Hagemann et al. (2023); Kerrigan et al. (2023); Lim et al. (2023a;b); Pidstrigach et al. (2023); Phillips et al. (2022); Baldassari et al. (2023). Kerrigan et al. (2023) proposes a discrete-time model that serves as an analog of Ho et al. (2020) in infinite-dimensional space, and Hagemann et al. (2023) introduces a finite-dimensional approximation of an infinite-dimensional SDEs and utilizes the time-reversal formula in finite-dimensional spaces. Lim et al. (2023a); Franzese et al. (2023); Pidstrigach et al. (2023) propose continuous-time models by extending the SDE framework of Song et al. (2021b) to infinite dimensions based on semigroup theory (ref. Da Prato & Zabczyk (2014)); however, their consideration is limited to a relatively simple class of SDEs, such as Langevin type SDE or SDEs with constant-time diffusion coefficients. Later, Lim et al. (2023b) proved a general form of time-reversal formula which encompasses various choices of SDEs such as VPSDE, VESDE, sub-VPSDE (Song et al., 2021b) and variance scheduling (Nichol & Dhariwal, 2021), by exploiting more advanced mathematical machinery, e.g., variational approach and functional derivatives (ref. Krylov & Rozovskii (2007); Bogachev & Mayer-Wolf (1999)).

*This work was done while Kunwoo Na was an intern at KAIST AI

Research works mentioned above are primarily focused on the *training* of diffusion models, i.e., they aim to implement a mathematical framework in which the score-matching objective (Sohl-Dickstein et al., 2015; Vincent, 2011) and time reversal (Föllmer & Wakolbinger, 1986; Millet et al., 1989) of the noising process are possible. Although the "SDE" component and the "score-matching" component of the finite-dimensional diffusion model have been transferred to infinite dimensions, the existence of an infinite dimensional analog of **probability-flow ODE** (PF-ODE; Song et al. (2021b)) is still open. Indeed, PF-ODE has been crucial in the sampling process of diffusion models as it allows for fast sampling (Chen et al., 2023; Lu et al., 2022b) and consistency modeling (Song et al., 2023). In this work, we aim to accelerate the *inference* process of infinite-dimensional diffusion models by extending the probability-flow ODE (Song et al., 2021b) to infinite-dimensional spaces.

Contributions. Our contributions are as follows:

- We derive in a mathematically rigorous manner the notion of probability-flow ODE (Theorem 3.1) associated with a general class of stochastic differential equations (SDEs) in infinite-dimensional spaces, including VPSDE, VESDE, sub-VPSDE (Song et al., 2021b) and variance scheduling (Nichol & Dhariwal, 2021). We note that our infinite-dimensional probability-flow ODE is widely applicable regardless of the specific formulation of the infinite-dimensional diffusion model.
- We empirically demonstrate that sampling with PF-ODE achieves comparable or superior generation quality to the previous SDE-based approach while requiring significantly fewer number of function evaluations (NFEs) in both toy and real-world PDE problems.

2 Preliminaries

2.1 Probability-Flow ODE in $\mathbf{R}^n$

Let us consider the following stochastic differential equation in $\mathbf{R}^n$ ($n < \infty$) over $t \in [0, T]$:

$$dX_t = f(t, X_t)dt + \sigma(t)dB_t, \quad X_0 \sim p_0 = p_{\text{data}}, \tag{1}$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion in $\mathbf{R}^n$, $f : [0, T] \times \mathbf{R}^n \to \mathbf{R}^n$ is the drift term, $\sigma : [0, T] \to \mathrm{Mat}_n(\mathbf{R})$ is the diffusion term, and $p_0 = p_{\text{data}}$ is the probability *density* of the target data distribution. Closely related to this SDE is the so-called **probability-flow ODE** (PF-ODE; Song et al. (2021b)):

$$dY_t = \left[f(t, Y_t) - \frac{1}{2} A(t) \nabla \log p_t(Y_t) \right] dt, \quad Y_0 \sim p_0,$$

where $A(t) = \sigma(t)\sigma(t)^\intercal$, and p_t is the density of X_t. It is well-known that the solution for the PF-ODE has the same density as X_t for each t (Song et al., 2021b, Appendix D.1). The derivation of the PF-ODE heavily relies on the Fokker-Planck equation (ref. Øksendal (2003) for example), a well-studied second-order PDE whose solution is $(t, x) \mapsto p_t(x)$. In infinite-dimensional spaces, however, one cannot utilize the probability density function in the analysis due to the lack of reference measure (ref. Lunardi et al. (2015), Proposition 2.2.1). Hence, a more careful treatment is required for infinite-dimensional cases.

2.2 Infinite-Dimensional Analysis

Let $\mathcal{H}$ denote a real separable Hilbert space, and $(W_t)_{t\geq 0}$ be a Q-Wiener process on $\mathcal{H}$. Denote by $\mathcal{H}_Q$ the Cameron-Martin space (ref. Da Prato & Zabczyk (2014)) of $\mathcal{N}(0, Q)$. Let $\mathscr{L}_2(\mathcal{H})$ be the set of Hilbert-Schmidt operators on $\mathcal{H}$, and let $\{\varphi_i\}$ be an orthonormal basis of $\mathcal{H}$ that consists of eigenvectors of Q corresponding to λ_i. We assume $\mathcal{H}$ is a function space over some set $\Omega \subseteq \mathbf{R}^d$ ($d < \infty$); for example, $\mathcal{H} = L^2(\Omega)$ or $\mathcal{H} = W^{1,2}(\Omega)$.

Due to the lack of reference measure in $\mathcal{H}$, we shall express the time evolution of a family of probability measures in a *weak sense*; that is, we express the evolution of the dual pairings of a probability measure and test functions. Below, we introduce the minimal background required for this work; we refer the readers to Appendix A for a more detailed overview.

Test functions. The class of cylindrical functions $\mathcal{FC}_b^\infty(\mathcal{H})$ is defined as

$$\mathcal{FC}_b^\infty(\mathcal{H}) = \left\{ x \mapsto f(\langle \varphi_1, x\rangle, \cdots, \langle \varphi_m, x\rangle) \;\middle|\; m \in \mathbf{N},\ f \in \mathcal{C}_0^\infty(\mathbf{R}^m) \right\}.$$

We write $f_{\varphi_1,\cdots,\varphi_m}(x) = f(\langle \varphi_1, x\rangle, \cdots, \langle \varphi_m, x\rangle)$ for $x \in \mathcal{H}$. Here, $\mathcal{C}_0^\infty(\mathbf{R}^m)$ is the space of smooth functions on $\mathbf{R}^m$ that vanish at infinity, which serves as a canonical class of test functions in usual finite-dimensional analysis.

Weak formulation. Let $\mathcal{L}$ be an operator such that $\mathcal{L}\psi : \mathcal{H} \to \mathbf{R}$ is in $L^1(\mathcal{H}, \mu)$ for all $\psi \in \mathcal{FC}_b^\infty(\mathcal{H})$. We say $\mathcal{L}^*\mu = 0$ if

$$\int_{\mathcal{H}} \mathcal{L} f_{\varphi_1,\cdots,\varphi_m}(x)\mu(dx) = 0, \quad \forall f_{\varphi_1,\cdots,\varphi_m} \in \mathcal{FC}_b^\infty(\mathcal{H}).$$

In a similar manner, for a family of measures $\{\nu_t\}$, we shall understand the equation $\mathcal{L}^*\mu = \partial_t \nu_t$ in a *weak sense*, i.e., we say $\mathcal{L}^*\mu = \partial_t\, \nu_t$ if

$$\int_{\mathcal{H}} \mathcal{L} f_{\varphi_1,\cdots,\varphi_m}(x)\mu(dx) = \frac{\partial}{\partial t}\int_{\mathcal{H}} f_{\varphi_1,\cdots,\varphi_m}(x)\nu_t(dx), \quad \forall f_{\varphi_1,\cdots,\varphi_m} \in \mathcal{FC}_b^\infty(\mathcal{H}).$$

Logarithmic gradient. We say that a Borel probability measure μ is Fomin differentiable along $h \in \mathcal{H}_Q$ if there exists a function $\rho_h^\mu \in L^1(\mathcal{H}, \mu)$ such that

$$\int_{\mathcal{H}} \partial_h f_{\varphi_1,\cdots,\varphi_m}(x)\mu(dx) = -\int_{\mathcal{H}} f_{\varphi_1,\cdots,\varphi_m}(x)\rho_h^\mu(x)\mu(dx), \quad \forall f_{\varphi_1,\cdots,\varphi_m} \in \mathcal{FC}_b^\infty(\mathcal{H}). \tag{2}$$

Here, $\partial_h f_{\varphi_1,\cdots,\varphi_m}(x)$ denotes the Gâteaux differential of $f_{\varphi_1,\cdots,\varphi_m}$ at x along h. If there exists a function $\rho_{\mathcal{K}}^\mu : \mathcal{H} \to \mathcal{H}$ such that $\langle \rho_{\mathcal{K}}^\mu(x), h\rangle_{\mathcal{K}} = \rho_h^\mu(x)$ for every $x \in \mathcal{H}$ and $h \in \mathcal{K}$, then we call $\rho_{\mathcal{K}}^\mu$ the logarithmic gradient of μ along $\mathcal{K}$.

3 Probability-Flow ODEs in Function Spaces

Let us consider an SDE in $\mathcal{H}$ given by

$$dX_t = B(t, X_t)dt + G(t)dW_t, \quad X_0 \sim \mathbf{P}_0 = \mathbf{P}_{\text{data}}, \tag{3}$$

where $(W_t)_{t\geq 0}$ is a Q-Wiener process on $\mathcal{H}$, $B : [0,T] \times \mathcal{H} \to \mathcal{H}$ and $G : [0,T] \to \mathscr{L}_2(\mathcal{H})$ are progressively measurable, and $\mathbf{P}_0 = \mathbf{P}_{\text{data}}$ is the probability *measure* from which X_0 is sampled. Prior works (Hagemann et al., 2023; Lim et al., 2023b;a; Pidstrigach et al., 2023) on infinite-dimensional diffusion models directly implement Eqn. (3) and its time-reversal. On the other hand, in finite-dimensional models, PF-ODE has played a crucial role in allowing for faster sampling (Lu et al., 2022b) and recently, leading to consistency modeling (Song et al., 2023). Thus, it is only natural to ask whether there is an infinite-dimensional version of the PF-ODE, which, to the best of our knowledge, has not been tackled in the literature yet.

The usual approach of Song et al. (2021b) of deriving the PF-ODE *fails* in infinite-dimensions, as there is no probability density function. Our main question in this section is as follows:

> *Is there an ODE in infinite-dimensional space with a random initial point $Y_0 \sim \mathbf{P}_0$ whose solution evolves like the solution of the original SDE (Eqn. (3))?*

The answer is affirmative.

Consider the following family of operators $\{\mathcal{L}_t\}_{t\in(0,T]}$ defined by

$$\mathcal{L}_t f_{\varphi_1,\cdots,\varphi_m}(u) = \frac{1}{2}\operatorname{Tr}_{\mathcal{H}_Q}\left(A(t)\circ Q \circ D^2 f_{\varphi_1,\cdots,\varphi_m}(u)\right) + \langle Df_{\varphi_1,\cdots,\varphi_m}(u), B(t,u)\rangle_{\mathcal{H}_Q}$$

for $f_{\varphi_1,\cdots,\varphi_m} \in \mathcal{FC}_b^\infty(\mathcal{H})$, where $A(t) = G(t)G(t)^*$ and D stands for the Fréchet derivative. It is known (ref. Belopolskaya & Dalecky (2012), Chapter 5) that for the solution of Eqn. (3) denoted X_t, the law $\mu_t = \operatorname{Law}(X_t)$ satisfies the following *Fokker-Planck-Kolmogorov* equation

$$\begin{cases} \partial_t \mu_t = (\mathcal{L}_t)^*\mu_t, & t \in (0,T], \\ \mu_t\Big|_{t=0} = \mathbf{P}_0. \end{cases}$$

Exploiting the preceding Fokker-Planck-Kolmogorov equation, we explicitly state the PF-ODE in infinite-dimensional spaces as in the following theorem, whose proof is deferred to Appendix B:

Theorem 3.1. *Let X_t be a solution of Eqn. (3) and $\mu_t := \mathrm{Law}(X_t)$. Then, μ_t satisfies the Fokker-Planck-Kolmogorov equation of $(Y_t)_{t\in[0,T]}$, where $(Y_t)_{t\in[0,T]}$ is a solution of the following probability-flow ODE in infinite-dimension:*

$$dY_t = \left[B(t, Y_t) - \frac{1}{2}A(t)\rho^{\mu_t}_{\mathcal{H}_Q}(Y_t)\right]dt, \quad Y_0 \sim \mathbf{P}_0. \tag{4}$$

Here, $A(t) := G(t)G(t)^*$ and $\rho^{\mu_t}_{\mathcal{H}_Q}$ is the logarithmic gradient of μ_t along $\mathcal{H}_Q$.

4 Experiments

In all experiments, we sample synthetic functions via our PF-ODE and the usual time-reversed SDE in infinite-dimensional function spaces, where we employ the Euler's method for the ODE and SDE solving for each NFE. In Appendix C, we provide the missing implementation details.

4.1 1D Function generation

Setting. We use a synthetic dataset `Quadratic` consisting of (noise-corrupted) functions of the form

$$f(x; a) = ax^2 + \varepsilon,$$

where $a \sim \mathrm{Unif}\{-1, 1\}$ and $\varepsilon \sim \mathcal{N}(0, 1)$ are sampled independently. These functions are evaluated at a fixed grid $x =$ `np.linspace(-10, 10, 100)`. We utilize the checkpoint trained on the `Quadratic` dataset by Lim et al. (2023b). For the evaluation, we calculate the power of kernel two-sample test with functional PCA kernel (Wynne & Duncan (2022); lower power is better). We consider the number of function evaluations (NFEs) in the range of $\{10, 20, \cdots, 100\}$.

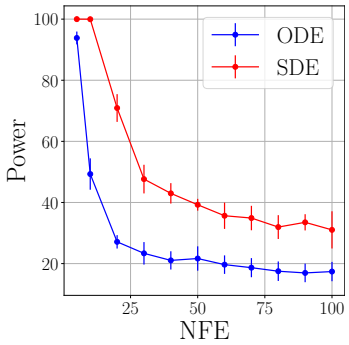
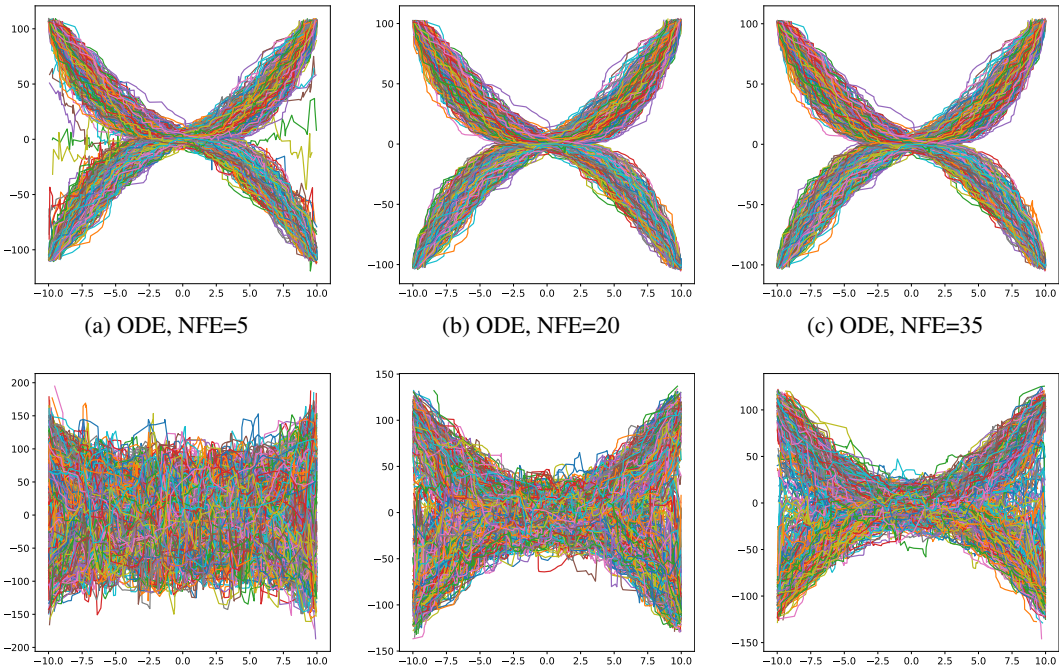


Figure 1: Power vs. NFE.



(a) ODE, NFE=5 (b) ODE, NFE=20 (c) ODE, NFE=35

(d) SDE, NFE=5 (e) SDE, NFE=20 (f) SDE, NFE=35

Figure 2: Qualitative comparison of ODE- and SDE-generated samples in `Quadratic` dataset with NFE$\in\{5, 20, 35\}$. Samples from the (Top) ODE solver and (Bottom) SDE solver.

Discussions. Figure 1 quantitatively compares samples from the ODE and SDE solver, which clearly shows that the ODE solver outperforms the SDE solver at every considered NFE. Remarkably, ODE solving with NFE=20 performs even better than SDE solving with *any* NFE. In Figure 2, we show the samples generated via the ODE and SDE solving with NFE 5, 20, and 35 (with a fixed seed). Qualitatively as well, it is clear that the ODE solver produces much better samples than the SDE solver.

4.2 Synthetic solution of various PDEs

For the PDE tasks, we train an infinite-dimensional diffusion model via the score-matching objective. Then, we sample a synthetic solution for two well-studied PDE problems, namely, the diffusion-reaction and the heat equation, via solving our PF-ODE and the usual SDE, respectively. We use the same checkpoint during the inference via the ODE and SDE solving for a fair comparison.

4.2.1 Diffusion-reaction equation

Setting. We consider the diffusion-reaction equation of the form

$$\partial_t u = D\Delta u + R,$$

where D is a diagonal matrix, and R is a function that accounts for the diffusion of the system and the source term, respectively. We utilize `PDEBench` dataset (Takamoto et al., 2022), which consists of solutions to the preceding diffusion-reaction equation with varying D and R. Figure 3 shows a batch of ground-truth solutions for diffusion-reaction equation from `PDEBench` dataset. We train an infinite-dimensional diffusion model with resolution 64. During inference, we take various NFEs $\in \{10, 20, \cdots, 100\}$. For a quantitative investigation, we compute the sliced Wasserstein (SW) distance (Stein et al., 2024) of synthetic samples (lower SW distance is better) as in Hagemann et al. (2023).

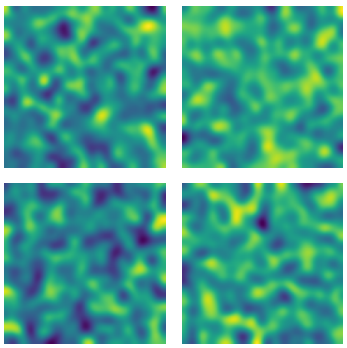


Figure 3: Ground-truth solutions sampled from `PDEBench` dataset.

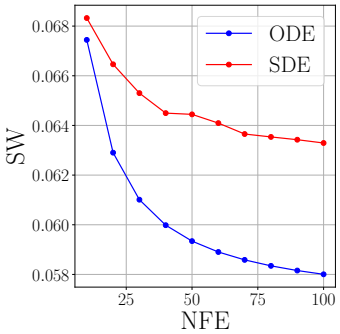


Figure 4: SW vs. NFE.

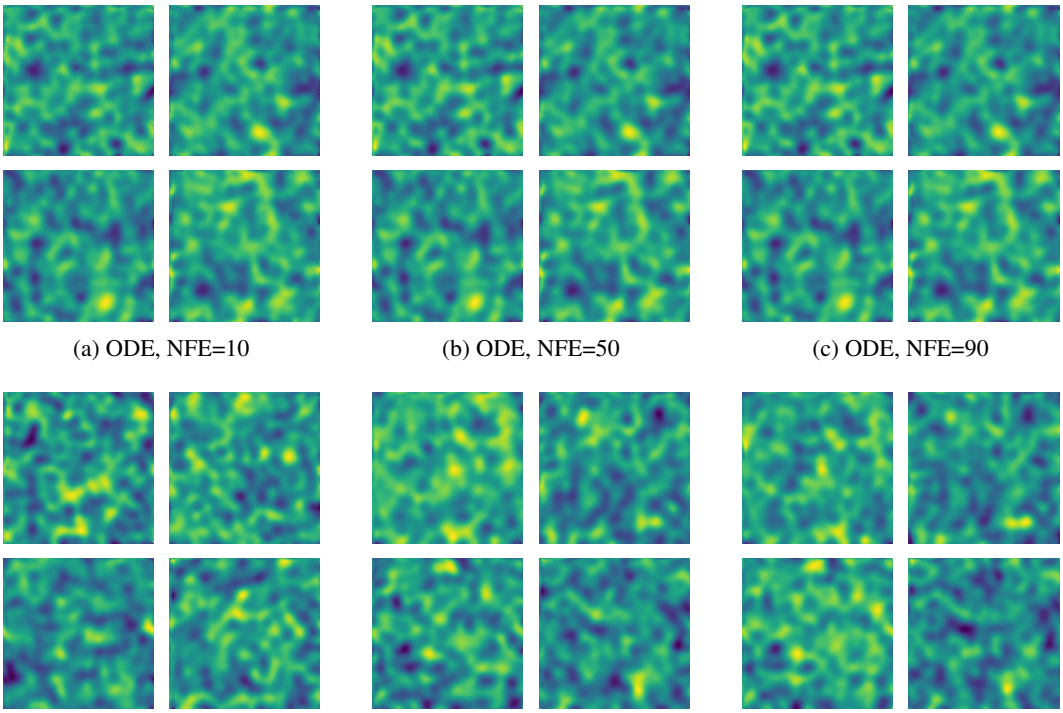
Discussions. Figure 4 shows the SW distance of samples with resolution 256, generated by SDE and ODE solving at various NFEs$\in \{10, 20, \cdots, 100\}$. We note that samples from the ODE solver show a lower SW distance than those from the SDE solver at every NFE. Figure 5 compares samples obtained from the ODE solver and the SDE solver at NFE 10, 50, and 90, where each samples are generated with the same fixed seed. Qualitatively, observe that the ODE samples across NFE $\in \{10, 50, 90\}$ (Figure 5a, 5b, and 5c) are similar to each other, while the SDE samples across the same NFEs (Figure 5d, 5e, and 5f) show severe variations. This suggests that sampling via our PF-ODE is much faster than the SDE; for this specific example, running the ODE solver with NFE=10 is sufficient, while much more is required for the SDE solver.

4.2.2 Heat equation

Setting. We consider the heat equation on $\mathcal{O} = [-1, 1]^2$ with zero Neumann boundary condition:

$$\begin{cases} \partial_t u = \beta \Delta u & \text{on } \mathcal{O}, \\ u = 0 & \text{on } \partial\mathcal{O}. \end{cases}$$

Here, $u : [0, T] \times \mathcal{O} \to \mathbf{R}$ and $\beta \in [2 \times 10^{-3}, 2 \times 10^{-2}]$ is a constant. It is well-known that the preceding heat equation, given with initial condition as an additional datum, has a unique solution



(a) ODE, NFE=10 (b) ODE, NFE=50 (c) ODE, NFE=90

(d) SDE, NFE=10 (e) SDE, NFE=50 (f) SDE, NFE=90

Figure 5: Qualitative comparison of ODE- and SDE-generated solutions for diffusion-reaction equation with NFE$\in\{10, 50, 90\}$. Samples from the (Top) ODE solver and (Bottom) SDE solver.

under mild regularity conditions (ref. Evans (2022)). Furthermore, the unique solution can be easily numerically simulated, using numerical analytic techniques such as finite difference methods (Dawson et al., 1991). We generate a training dataset by randomly generating an initial condition f as a mixture of sine functions (as in Zhou & Farimani (2024)), and then numerically solving the heat equation with $\beta = 0.05$ and initial condition $u(0, \cdot) = f$. We train an (infinite-dimensional) diffusion model with resolution 64. For a systematic comparison, we first generate synthetic solutions u_{Synt} via the ODE or SDE solver. Then, we numerically solve the preceding heat equation via the finite-difference method with the initial condition given as $u_{\text{Synt}}(0, \cdot)$ to obtain a ground truth solution $u_\star$. This allows us to compute the L^p-distance $\|u_{\text{Synt}} - u_\star\|_{L^p([0,T]\times\mathcal{O})}$ between the synthetic solution and corresponding ground truth solution with the same initial condition. In particular, we measure their L^2- and L^∞-distances.

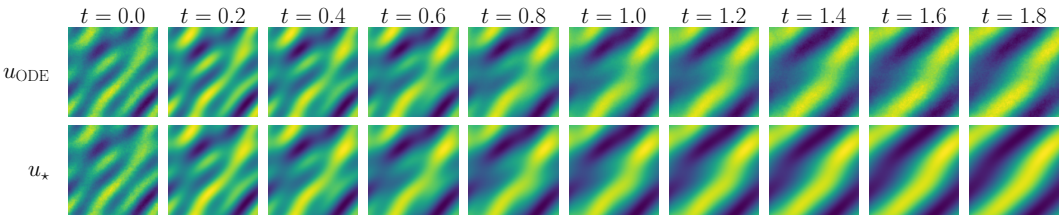


Figure 6: u_{ODE} generated with NFE 10.

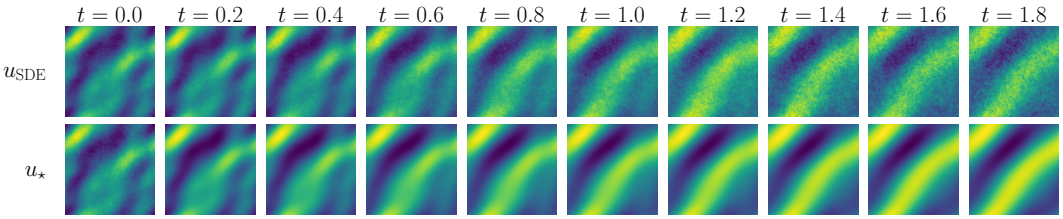


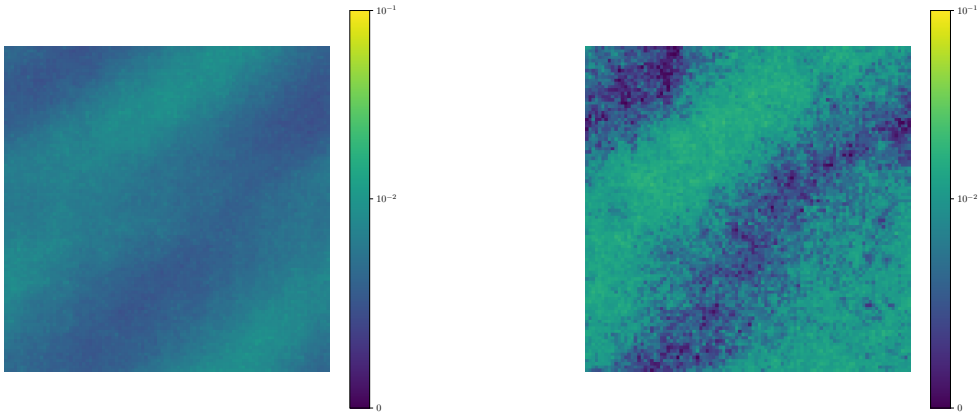
Figure 7: u_{SDE} generated with NFE 10.

Discussions. Figure 6 shows a synthetic solution generated from the ODE solver with NFE 10 (denoted u_{ODE}) and corresponding ground truth solution $u_\star$ with the same initial condition. Similarly, Figure 7 compares a synthetic solution obtained from solving SDE with NFE 10 and the corresponding ground truth solution. Notably, the solution generated by the ODE solver is much less noisy than that generated by the SDE solver. Figure 8 shows the pixel-wise difference between a synthetic solution generated by the ODE solver and the SDE solver with the same NFE 10 and the corresponding ground truth solution with the same initial solution, where samples are generated with resolution 64.

From Figure 8, one can observe that the solution generated from the ODE solver is much more similar to the ground truth than that of the SDE solver. From Table 1, it is notable that samples generated via the ODE solver with NFE 10 have lower L^p-distances to the ground truth solution $u_\star$ than those generated via the SDE solver with the same NFE.

Sampling method	ODE	SDE
L^2-distance (↓)	**12.85**± 1.44	15.58± 1.84
L^∞-distance (↓)	**1.31e-1**± 9.83e-3	1.46e-1± 1.36e-2

Table 1: Comparison of L^2- and L^∞-distance for samples generated via the ODE and SDE solving with NFE 10.



(a) $|u_{\mathrm{ODE}}(2,\cdot) - u_\star(2,\cdot)|$ (b) $|u_{\mathrm{SDE}}(2,\cdot) - u_\star(2,\cdot)|$

Figure 8: Difference between a synthetic solution and the corresponding ground truth solution with same initial condition. Samples are generated with NFE 10.

5 Conclusion and Future Work

In this work, we derive a notion of probability-flow ODE (PF-ODE) in infinite-dimensional function spaces with functional derivatives and measure-valued Fokker-Planck-Kolmogorov equation. By utilizing our infinite-dimensional PF-ODE, we lower the NFEs without affecting the sample quality in various function generation settings. We observe that in some examples, such as time-evolving two-dimensional PDE problems, samples generated via our PF-ODE are of higher quality than those generated via the SDE not only at low NFEs but also for overall NFEs.

Our newly derived infinite-dimensional PF-ODE opens up various avenues for future work in functional diffusion models. First, we leave extending our work to faster solvers (Lu et al., 2022b) and knowledge distillation (Song et al., 2023)) as future work. Also, a rigorous investigation into the discretization error of infinite-dimensional diffusion models, both SDE and our PF-ODE, is another fruitful direction. This direction may shed light on the effectiveness of our PF-ODE over SDE in several function generation tasks, which we believe is because the ODE method only incurs a single discretization error from the initial approximation of infinite-dimensional noise $\xi \sim \mathcal{N}(0, Q)$; in contrast, for the SDE method, repeated discretization error for ξ occurs.

ACKNOWLEDGMENTS

The authors thank Hojung Jung for providing helpful feedback on the initial manuscript. This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No.RS-2019-II190075 Artificial Intelligence Graduate School Program (KAIST), No.2022-0-00612, Geometric and Physical Commonsense Reasoning based Behavior Intelligence for Embodied AI, No.RS-2022-II220311, Development of Goal-Oriented Reinforcement Learning Techniques for Contact-Rich Robotic Manipulation of Everyday Objects), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00410082).

BIBLIOGRAPHY

Mohamed Akrout, Bálint Gyepesi, Péter Holló, Adrienn Poór, Blága Kincső, Stephen Solis, Katrina Cirone, Jeremy Kawahara, Dekker Slade, Latif Abid, et al. Diffusion-based data augmentation for skin disease classification: Impact across original medical datasets to fully synthetic images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 99–109. Springer, 2023.

Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Jean-Pierre Aubin. *Applied functional analysis*. John Wiley & Sons, 2011.

Christopher TH Baker and RL Taylor. The numerical treatment of integral equations. *Journal of Applied Mechanics*, 46(4):969, 1979.

Lorenzo Baldassari, Ali Siahkoohi, Josselin Garnier, Knut Solna, and Maarten V. de Hoop. Conditional score-based diffusion models for bayesian inference in infinite dimensions. In *Advances in Neural Information Processing Systems*, volume 36, pp. 24262–24290. Curran Associates, Inc., 2023. URL https://openreview.net/forum?id=voG6nEW9BV.

Ya I Belopolskaya and Yu L Dalecky. *Stochastic equations and differential geometry*, volume 30. Springer Science & Business Media, 2012.

Vladimir Bogachev and Eduardo Mayer-Wolf. Absolutely continuous flows generated by sobolev class vector fields in finite and infinite dimensions. *Journal of functional analysis*, 167(1):1–68, 1999.

Vladimir I Bogachev, Nicolai V Krylov, Michael Röckner, and Stanislav V Shaposhnikov. *Fokker–Planck–Kolmogorov Equations*, volume 207. American Mathematical Society, 2022.

Vladimir Igorevich Bogachev. *Gaussian Measures*. Number 62 in Mathematical Surveys and Monographs. American Mathematical Society, 1998.

Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=MhK5aXo3gB. Expert Certification.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems*, volume 31, pp. 6572–6583. Curran Associates, Inc., 2018. URL https://arxiv.org/abs/1806.07366.

Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ODE is provably fast. In *Advances in Neural Information Processing Systems*, volume 36, pp. 68552–68575. Curran Associates, Inc., 2023. URL https://openreview.net/forum?id=KD6MFeWSAd.

Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated MRI. *Medical Image Analysis*, 80:102479, 2022. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2022.102479. URL https://www.sciencedirect.com/science/article/pii/S1361841522001268.

John B Conway. *A Course in Functional Analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer New York, NY, 2 edition, 2007.

Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*. Number 152 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 2014.

Clint N Dawson, Qiang Du, and Todd F Dupont. A finite difference domain decomposition algorithm for numerical solution of the heat equation. *Mathematics of computation*, 57(195):63–71, 1991.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *Advances in Neural Information Processing Systems*, 35:2406–2422, 2022.

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

Nathaniel Eldredge. Analysis and Probability on Infinite-Dimensional Spaces. *arXiv preprint arXiv:1607.03591*, 2016. URL https://arxiv.org/abs/1607.03591.

Lawrence C Evans. *Partial differential equations*, volume 19. American Mathematical Society, 2022.

H Föllmer and A Wakolbinger. Time reversal of infinite-dimensional diffusions. *Stochastic processes and their applications*, 22(1):59–77, 1986.

Sergei Vasil'evich Fomin. Differentiable measures in linear spaces. *Uspekhi Matematicheskikh Nauk*, 23(1):221–222, 1968.

Giulio Franzese, Giulio Corallo, Simone Rossi, Markus Heinonen, Maurizio Filippone, and Pietro Michiardi. Continuous-Time Functional Diffusion Processes. In *Advances in Neural Information Processing Systems*, volume 36, pp. 37370–37400. Curran Associates, Inc., 2023. URL https://openreview.net/forum?id=VPrir0p5b6.

Leszek Gawarecki and Vidyadhar Mandrekar. *Stochastic Differential Equations in Infinite Dimensions with Applications to Stochastic Partial Differential Equations*. Probability and Its Applications. Springer Berlin, Heidelberg, 2011.

Paul Hagemann, Sophie Mildenberger, Lars Ruthotto, Gabriele Steidl, and Nicole Tianjiao Yang. Multilevel diffusion: Infinite dimensional score-based diffusion models for image generation. *arXiv preprint arXiv:2303.04772*, 2023.

Tapio Helin and Martin Burger. Maximum a posteriori probability estimates in infinite-dimensional bayesian inverse problems. *Inverse Problems*, 31(8):085009, 2015. URL https://iopscience.iop.org/article/10.1088/0266-5611/31/8/085009.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://arxiv.org/abs/2006.11239.

Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Gavin Kerrigan, Justin Ley, and Padhraic Smyth. Diffusion Generative Models in Infinite Dimensions. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 9538–9563. PMLR, 25–27 Apr 2023. URL https://proceedings.mlr.press/v206/kerrigan23a.html.

Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional flow matching. In *International Conference on Artificial Intelligence and Statistics*, pp. 3934–3942. PMLR, 2024.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*, 2020.

Jean Kossaifi, Nikola Kovachki, Zongyi Li, Davit Pitt, Miguel Liu-Schiaffini, Robert Joseph George, Boris Bonev, Kamyar Azizzadenesheli, Julius Berner, and Anima Anandkumar. A library for learning neural operators, 2024.

Nikola B. Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces. *CoRR*, abs/2108.08481, 2021.

Nicolai V Krylov and Boris L Rozovskii. Stochastic evolution equations. In *Stochastic Differential Equations: Theory And Applications: A Volume in Honor of Professor Boris L Rozovskii*, pp. 1–69. World Scientific, 2007.

Hui-Hsiung Kuo. *Gaussian Measures in Banach Spaces*. Number 463 in Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 1975.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Jae Hyun Lim, Nikola B. Kovachki, Ricardo Baptista, Christopher Beckham, Kamyar Azizzadenesheli, Jean Kossaifi, Vikram Voleti, Jiaming Song, Karsten Kreis, Jan Kautz, Christopher Pal, Arash Vahdat, and Anima Anandkumar. Score-based Diffusion Models in Function Space. *arXiv preprint arXiv:2302.07400*, 2023a. URL https://arxiv.org/abs/2302.07400.

Sungbin Lim, Eunbi Yoon, Taehyun Byun, Taewon Kang, Seungwoo Kim, Kyungjae Lee, and Sungjoon Choi. Score-based Generative Modeling through Stochastic Evolution Equations in Hilbert Spaces. In *Advances in Neural Information Processing Systems*, volume 36, pp. 37799–37812. Curran Associates, Inc., 2023b. URL https://openreview.net/forum?id=GrElRvXnEj.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow Matching for Generative Modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Cheng Lu, Kaiwen Zheng, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Maximum Likelihood Training for Score-based Diffusion ODEs by High Order Denoising Score Matching. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 14429–14460. PMLR, 17–23 Jul 2022a. URL https://proceedings.mlr.press/v162/lu22f.html.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022b.

Alessandra Lunardi, Michele Miranda, and Diego Pallara. Infinite Dimensional Analysis. *19th Internet Seminar*, 2015. URL https://www.mathematik.tu-darmstadt.de/media/analysis/lehrmaterial_anapde/hallerd/Lectures.pdf.

Zhengxiong Luo, Dayou Chen, Yingya Zhang, Yan Huang, Liang Wang, Yujun Shen, Deli Zhao, Jingren Zhou, and Tieniu Tan. VideoFusion: Decomposed Diffusion Models for High-Quality Video Generation. *arXiv preprint arXiv:2303.08320*, 2023. URL https://arxiv.org/abs/2303.08320.

Annie Millet, David Nualart, and Marta Sanz. Time reversal for infinite-dimensional diffusions. *Probability theory and related fields*, 82(3):315–347, 1989.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.

Angus Phillips, Thomas Seror, Michael John Hutchinson, Valentin De Bortoli, Arnaud Doucet, and Emile Mathieu. Spectral Diffusion Processes. *arXiv preprint arXiv:2209.14125*, 2022. URL https://arxiv.org/abs/2209.14125.

Jakiw Pidstrigach, Youssef Marzouk, Sebastian Reich, and Sven Wang. Infinite-Dimensional Diffusion Models. *arXiv preprint arXiv:2302.10130*, 2023. URL https://arxiv.org/abs/2302.10130.

Phil Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, and Tom Goldstein. The Intrinsic Dimension of Images and Its Impact on Learning. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=XJk19XzGq2J.

Claudia Prévôt and Michael Röckner. *A Concise Course on Stochastic Partial Differential Equations*. Number 1905 in Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 2007.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33:12438–12448, 2020.

Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum Likelihood Training of Score-Based Diffusion Models. In *Advances in Neural Information Processing Systems*, volume 34, pp. 1415–1428. Curran Associates, Inc., 2021a. URL https://openreview.net/forum?id=AkltWFnxS9.

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*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving Inverse Problems in Medical Imaging with Score-Based Generative Models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=vaRCHVj0uGI.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency Models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32211–32252. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/song23a.html.

George Stein, Jesse Cresswell, Rasa Hosseinzadeh, Yi Sui, Brendan Ross, Valentin Villecroze, Zhaoyan Liu, Anthony L Caterini, Eric Taylor, and Gabriel Loaiza-Ganem. Exposing flaws of generative model evaluation metrics and their unfair treatment of diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024.

Andrew M Stuart. Inverse problems: a Bayesian perspective. *Acta numerica*, 19:451–559, 2010. URL https://www.cambridge.org/core/journals/acta-numerica/article/abs/inverse-problems-a-bayesian-perspective/587A3A0D480A1A7C2B1B284BCEDF7E23.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Dan MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. PDEBench Datasets, 2022. URL https://doi.org/10.18419/darus-2986.

Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based Generative Modeling in Latent Space. In *Advances in Neural Information Processing Systems*, volume 34, pp. 11287–11302. Curran Associates, Inc., 2021. URL https://openreview.net/forum?id=P9TYG0j-wtG.

Pascal Vincent. A Connection Between Score Matching and Denoising Autoencoders. *Neural Computation*, 23(7):1661–1674, 07 2011. ISSN 0899-7667. doi: 10.1162/NECO_a_00142. URL https://doi.org/10.1162/NECO_a_00142.

George Wynne and Andrew B Duncan. A kernel two-sample test for functional data. *Journal of Machine Learning Research*, 23(73):1–51, 2022.

Anthony Zhou and Amir Barati Farimani. Masked autoencoders are PDE learners. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=rZNuiFwXVs.

Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer Berlin, Heidelberg, 6 edition, 2003.

APPENDIX

Structure of the Appendix.

- In Appendix A, we introduce a minimal mathematical background required in this work. Table 2 lists mathematical symbols and notations used throughout this work.
- In Appendix B, we provide a full proof of our main Theorem 3.1.
- In Appendix C, we provide details regarding the architecture and training of infinite-dimensional diffusion model.

TABLE OF NOTATIONS

Symbol	Description
$T_{\#}\mu$	Pushforward of a measure μ by a map T
X^*	Dual space of X
$\langle x^*, x\rangle$	Dual pairing of $x^* \in X^*$ and $x \in X$
$\hat{\mu}$	Characteristic function of a probability measure μ on X defined by $\hat{\mu}(h) = \int e^{i\langle h,x\rangle}\mu(dx)$, $h \in X^*$.
$\langle \cdot, \cdot\rangle_{\mathcal{H}}$	Inner product on a Hilbert space $\mathcal{H}$
$\mathrm{Tr}_{\mathcal{H}}$	Trace on a Hilbert space $\mathcal{H}$
$\mathscr{L}_2(\mathcal{H})$	The set of Hilbert-Schmidt operators on $\mathcal{H}$
$\mathcal{N}(0, Q)$	Centered Gaussian measure in $\mathcal{H}$ with covariance operator Q
$(W_t)_{t\geq 0}$	A Q-Wiener process in $\mathcal{H}$
$\mathcal{H}_Q$	The Cameron-Martin space of $\mathcal{N}(0, Q)$
$\mathcal{FC}_b^\infty(\mathcal{H})$	The set of all cylindrical functions on $\mathcal{H}$
$\mathcal{M}(\mathcal{H})$	The set of all Borel measures on $\mathcal{H}$
$\mathcal{L}_t$	Kolmogorov operator defined on $\mathcal{FC}_b^\infty(\mathcal{H})$
$\partial_h f(x)$	Gâteaux differential of f at x along h
$Df(x)$	Fréchet derivative of f at x
$\mathrm{Law}(X)$	Distribution (law) of a random variable X
$\rho^{\mu}_{\mathcal{H}_Q}$	Logarithmic gradient of μ along $\mathcal{H}_Q$
Δ	The Laplace operator
∇	The gradient operator

Table 2: Mathematical Symbols and Definitions

A MATHEMATICAL PRELIMINARIES

In this section, we provide a gentle introduction to the theory of Gaussian measures and stochastic processes in infinite dimensional Hilbert spaces. Most of the content of this section can be found in Da Prato & Zabczyk (2014), Bogachev (1998), Bogachev et al. (2022), Prévôt & Röckner (2007), or Kuo (1975).

We introduce several notations and definitions here before introducing precise definitions of mathematical objects we exploit in this research.

Pushforward measure. If $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ are measurable spaces and $T : X \to Y$ is $\mathcal{F}/\mathcal{G}$-measurable, then for any measure μ on $(X, \mathcal{F})$ we define the pushforward measure $T_{\#}\mu$ by

$$(T_{\#}\mu)(A) = \mu(T^{-1}(A)), \quad \forall A \in \mathcal{G}.$$

Duality and pairing. For a locally convex topological vector space X over $\mathbb{k} = \mathbf{R}$ (or $\mathbf{C}$), we denote by X^* the dual space of X, i.e.,

$$X^* = \{\ell : X \to \mathbb{k} \mid \ell \text{ is linear and continuous}\}.$$

For $\ell \in X^*$ and $x \in X$ we denote by $\langle \ell, x\rangle$ the quantity $\ell(x)$.

Characteristic function. If μ is a probability measure on $(X, \mathcal{B}(X))$, we define the characteristic function $\hat{\mu}$ of μ by

$$\hat{\mu}(h) = \int_X e^{i\langle h,x\rangle}\mu(dx), \quad \forall h \in X^*.$$

It is well known that if $\hat{\mu} = \hat{\nu}$, then $\mu = \nu$ (ref. Da Prato & Zabczyk (2014), Proposition 2.5).

A.1 Gaussian measures and Wiener processes

A.1.1 Gaussian measures

Definition A.1. *A Borel probability measure μ on a locally convex space X is called a Gaussian measure if the pushforward measure $h_{\#}\mu$ is Gaussian for every $h \in X^*$. The measure μ is said to be centered if $h_{\#}\mu$ is centered in $\mathbf{R}$ for every $h \in X^*$.*

Theorem A.2 (ref. Bogachev (1998), Theorem 2.2.4)**.** *A measure μ on a locally convex space X is Gaussian if and only if its characteristic function is of the form*

$$\hat{\mu}(h) = \exp\left[iL(h) - \frac{1}{2}B(h,h)\right], \quad \forall h \in X^*,$$

where L is a linear functional on X^* and B is a symmetric, non-negative bilinear form on X^*.

From now on, we stick to the case where $X = \mathcal{H}$ is a separable Hilbert space. In this case, we may identify $\mathcal{H}$ with $\mathcal{H}^*$ via the Riesz representation. If μ is a Gaussian measure on $\mathcal{H}$, we can find some $m \in \mathcal{H}$ and a non-negative symmetric operator $Q : \mathcal{H} \to \mathcal{H}$ such that

$$\hat{\mu}(h) = \exp\left[i\langle m,h\rangle_{\mathcal{H}} - \frac{1}{2}\langle Qh,h\rangle_{\mathcal{H}}\right] \quad \forall h \in \mathcal{H}.$$

It is known that for every $f, g \in \mathcal{H}$,

$$\langle m,f\rangle_{\mathcal{H}} = \int_{\mathcal{H}} \langle f,x\rangle_{\mathcal{H}}\mu(dx)$$
$$\langle Qf,g\rangle_{\mathcal{H}} = \int_{\mathcal{H}} \langle f,x-m\rangle_{\mathcal{H}}\langle g,x-m\rangle_{\mathcal{H}}\mu(dx)$$

In other words, if Z is an $\mathcal{H}$-valued random variable with $\mathrm{Law}(Z) = \mu$, then

$$m = \mathbb{E}[Z], \quad Q = \mathrm{Cov}(Z).$$

We call m the mean vector and Q the covariance operator, and write $\mu = \mathcal{N}(m, Q)$.

We end this subsection by providing a brief notes on Cameron-Martin space of a Gaussian measure. Although there are several equivalent definitions of Cameron-Martin space, we follow that of Da Prato & Zabczyk (2014) as it can be presented without providing additional technical details.

Definition A.3. *Let μ be a centered Gaussian measure on a locally convex space X. A linear space $\mathcal{H}_\mu \subset X$ equipped with an inner product is called a Cameron-Martin space of μ if $\mathcal{H}_\mu$ is continuously embedded in X and for every $h \in X^*$, one has that $\mathrm{Law}(\varphi) = \mathcal{N}(0, |\varphi|_\mu^2)$, where*

$$|\varphi|_\mu = \sup_{h\in\mathcal{H}_\mu, \|h\|_{\mathcal{H}_\mu}\le 1} |\varphi(h)|.$$

For a Gaussian measure $\mu = \mathcal{N}(0, Q)$, it is known that the Cameron-Martin space $\mathcal{H}_\mu = \mathcal{H}_{\mathcal{N}(0,Q)}$ is given by

$$\mathcal{H}_\mu = Q^{1/2}(\mathcal{H}), \quad \langle f,g\rangle_{\mathcal{H}_\mu} = \langle Q^{-1/2}f, Q^{-1/2}g\rangle_{\mathcal{H}}.$$

We shall simply denote $\mathcal{H}_\mu = \mathcal{H}_Q$ in this case. If $\{\varphi_i\}$ is an orthonormal basis of the ambient Hilbert space $\mathcal{H}$, then $\{Q^{1/2}\varphi_i\}$ becomes an orthonormal basis for the Cameron-Martin space $\mathcal{H}_Q$.

A.1.2 WIENER PROCESSES

Definition A.4. *Let Q be a trace class non-negative symmetric operator on $\mathcal{H}$. An $\mathcal{H}$-valued stochastic process $W = (W_t)_{t\in[0,T]}$ on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is called a standard Q-Wiener process, if*

1. *$W(0) = 0$,*
2. *W has continuous trajectories, i.e., W has $\mathbf{P}$-continuous paths,*
3. *W has independent increments, i.e., for any $n \in \mathbf{N}$ and $0 < t_1 < \cdots < t_n < \infty$,*
$$W_{t_1},\ W_{t_2} - W_{t_1}, \cdots,\quad W_{t_n} - W_{t_{n-1}}$$
are independent,
4. *the increments have the following Gaussian laws:*
$$\mathbb{P} \circ (W_t - W_s)^{-1} := \mathrm{Law}(W_t - W_s) = \mathcal{N}(0, (t-s)Q)$$
for all $0 \le s \le t \le T$.

In this work, Q always denote a trace class non-negative symmetric operator on $\mathcal{H}$. Notice that Q is diagonalizable, and in particular, there exists a sequence $\{\varphi_k\}_{k=1}^{\infty}$ consists of eigenvectors of Q and a sequence of non-negative real numbers $\{\lambda_k\}_{k=1}^{\infty}$ such that $Q\varphi_k = \lambda_k\varphi_k$ for all $k = 1, 2, \cdots$ (Conway, 2007, Chapter II). Based on this eigensystem of Q, one can express Q-Wiener process as a series expansion. More precisely, one has the so-called Kosambi–Karhunen–Loève Theorem (see Prévôt & Röckner (2007), for example):

Theorem A.5. *A $\mathcal{H}$-valued stochastic process $W = (W_t)_{t\ge 0}$ is a Q-Wiener process if and only if*
$$W_t = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\beta_t^k \varphi_k$$
where $\beta^k = (\beta_t^k)_{t\in[0,T]}$ are independent real-valued Brownian motions on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The series converges in $L^2(\Omega, \mathcal{F}, \mathbf{P}; C([0,T]; \mathcal{U}))$. (Hence, there exists a $\mathbf{P}$-a.s. continuous version of W.)

A.2 FUNCTIONAL DERIVATIVES AND FOMIN DERIVATIVE

In this subsection, we introduce the notion of Fréchet and Gâteaux derivative (of functions $\mathcal{H} \to \mathbf{R}$) and Fomin derivative (of Borel measures on $\mathcal{H}$). Contents of this section can be found in Helin & Burger (2015) and Bogachev et al. (2022, Chapter 10), for example.

A.2.1 FUNCTIONAL DERIVATIVES

Let X and Y be locally convex spaces, and let $U \subset X$ be open. For a function $F : U \to Y$, the Gâteaux differential of F along $h \in X$ is defined by
$$\partial_h F(u) := \lim_{\varepsilon\to 0} \frac{F(u+\varepsilon h) - F(u)}{\varepsilon} = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} F(u+\varepsilon h),$$
whenever the limit exists. If the limit exists for every $h \in X$, then F is called Gâteaux differentiable at u.

In this paper, we stick to the cases where $X = \mathcal{H}$ and $Y = \mathbf{R}$ or $Y = \mathcal{H}$. In these cases (or more generally whenever X and Y are normed spaces), there is another canonical notion of differentiability called the Fréchet derivative. For a function $F : X \to Y$ (where X and Y are normed spaces), we say that F is Fréchet differentiable at $x \in U$ if there is a bounded linear operator $DF(x) : X \to Y$ such that
$$\lim_{\|h\|\to 0} \frac{\|F(x+h) - F(x) - DF(x)h\|}{\|h\|} = 0.$$

The notion of Fréchet differentiability is stronger than that of Gâteaux differentiability in the sense that whenever F is Fréchet differentiable at $x \in X$, then F is Gâteaux differentiable at x too, and $\partial_h F(x) = DF(x)(h)$. In particular, when $X = \mathcal{H}$ and $Y = \mathbf{R}$, then we can equivalently understand the notion of Fréchet differentiability at $x \in \mathcal{H}$ as an existence of $DF(x) \in \mathcal{H}$ such that

$$\lim_{\|h\| \to 0} \frac{|F(x+h) - F(x) - \langle DF(x), h \rangle_{\mathcal{H}}|}{\|h\|} = 0,$$

via the Riesz isomorphism $\mathcal{H} \cong \mathcal{H}^*$. If $F : \mathcal{H} \to \mathbf{R}$ is Fréchet differentiable at x, then its Gâteaux differential along $h \in \mathcal{H}$ coincides with $\langle DF(x), h \rangle$.

A.2.2 Fomin derivative

In this subsection, we briefly introduce the notion of Fomin differentiability of measures, which is developed by Fomin (1968). Although one can define the notion of Fomin differentiability for any Borel probability measure μ on a locally convex space X, we will stick to the case where $X = \mathcal{H}$ (and $\mu \in \mathcal{M}(\mathcal{H})$).

In an infinite dimensional space $\mathcal{H}$, neither the natural notion of probability density function (p) nor the notion of gradient (∇) exists. Still, a notion of the logarithmic gradient of probability measure μ on $\mathcal{H}$ (that acts as $\nabla \log p$) exists. We provide the formal definition below.

Definition A.6 (ref. Bogachev & Mayer-Wolf (1999))**.** *Let μ be a (Borel) probability measure on $\mathcal{H}$, and let $\mathcal{K} \subset \mathcal{H}$ be a densely embedded Hilbert space. We say μ is Fomin differentiable along $h \in \mathcal{K}$ if there exists a function $\rho_h^\mu \in L^1(\mu)$ such that*

$$\int_{\mathcal{H}} \partial_h f_{\varphi_1, \cdots, \varphi_n}(x) \mu(dx) = -\int_{\mathcal{H}} f_{\varphi_1, \cdots, \varphi_n}(x) \rho_h^\mu(x) \mu(dx). \tag{5}$$

If there exists a function $\rho_{\mathcal{K}}^\mu : \mathcal{H} \to \mathcal{H}$ such that $\langle \rho_{\mathcal{K}}^\mu(x), h \rangle_{\mathcal{K}} = \rho_h^\mu(x)$ for every $x \in \mathcal{H}$ and $h \in \mathcal{K}$, then we call $\rho_{\mathcal{K}}^\mu$ the logarithmic gradient of μ along $\mathcal{K}$.

It is well-known that μ is differentiable along h in the sense of Fomin if and only if the following quantity

$$d_h \mu(A) = \lim_{\varepsilon \to 0} \frac{\mu(A + \varepsilon h) - \mu(A)}{\varepsilon}$$

exists for every Borel set A in $\mathcal{H}$ (ref. Helin & Burger (2015), Proposition 1). Because the zero measure is the only measure on $\mathcal{H}$ which is Fomin differentiable along every vector in $\mathcal{H}$ (Bogachev et al., 2022, p.406), it is a necessary treatment in the above Eqn. (5) to specify the set $\mathcal{K} \subset \mathcal{H}$.

B PROOF OF THEOREM 3.1

This section provides a rigorous proof of our main theoretical result, Theorem 3.1, which we re-state for the sake of convenience.

Theorem (Restatement of Theorem 3.1)**.** *Let X_t be a solution of an SDE in $\mathcal{H}$ of the form*

$$dX_t = B(t, X_t)dt + G(t)dW_t, \quad X_0 \sim \mathbf{P}_0 = \mathbf{P}_{\text{data}}$$

and let $\mu_t := \text{Law}(X_t)$. Then, μ_t satisfies the Fokker-Planck-Kolmogorov equation of $(Y_t)_{t\in[0,T]}$, where $(Y_t)_{t\in[0,T]}$ is a solution of the following (infinite dimensional) probability-flow ODE

$$dY_t = \left[B(t, Y_t) - \frac{1}{2}A(t)\rho^{\mu_t}_{\mathcal{H}_Q}(Y_t)\right] dt, \quad Y_0 \sim \mathbf{P}_0. \tag{6}$$

Here, $A(t) := G(t)G(t)^*$ and $\rho^{\mu_t}_{\mathcal{H}_Q}$ is the logarithmic gradient of μ_t along the Cameron-Martin space $\mathcal{H}_Q$ of $\mathcal{N}(0, Q)$.

In the proof of Theorem 3.1, we slightly abuse notations and write $\varphi(h) = \langle \varphi, h\rangle$ in order to avoid too many brackets in the presentation. Also, we utilize the dual pairing notation of a function and a Borel (probability) measure: We write $\langle f, \nu\rangle$ for $f \in \mathcal{FC}^\infty_b(\mathcal{H})$ and $\nu \in \mathcal{M}(\mathcal{H})$ to denote the quantity $\int_{\mathcal{H}} f(u)\nu(du)$.

Proof of Theorem 3.1. View the probability-flow ODE (Eqn. (6)) as an SDE in $\mathcal{H}$ with no diffusion term, and consider the associated Kolmogorov type operator $\tilde{\mathcal{L}}_t$, $t \in [0, T]$, defined as

$$\tilde{\mathcal{L}}_t f_{\varphi_1,\cdots,\varphi_m}(u) = \Big\langle Df_{\varphi_1,\cdots,\varphi_m}(u), \underbrace{B(t, u) - \frac{1}{2}A(t)\rho^{\mu_t}_{\mathcal{H}_Q}(u)}_{=:\mathcal{G}(t,u)}\Big\rangle_{\mathcal{H}_Q}, \quad \forall f_{\varphi_1,\cdots,\varphi_m} \in \mathcal{FC}^\infty_b(\mathcal{H}),$$

where $Df_{\varphi_1,\cdots,\varphi_m}$ stands for the Fréchet derivative of $f_{\varphi_1,\cdots,\varphi_m}$. One has to check if the time-evolution of $\mu_t = \text{Law}(X_t)$ can be described in terms of the Kolmogorov type operator $\tilde{\mathcal{L}}_t$. That is, one has to check if μ_t satisfies

$$\partial_t\langle f_{\varphi_1,\cdots,\varphi_m}, \mu_t\rangle = \langle \tilde{\mathcal{L}}_t f_{\varphi_1,\cdots,\varphi_m}, \mu_t\rangle, \quad \forall f_{\varphi_1,\cdots,\varphi_m} \in \mathcal{FC}^\infty_b(\mathcal{H}). \tag{7}$$

By the definition of the Kolmogorov operator $\tilde{\mathcal{L}}_t$, one expands the left-hand side of Eqn. (7) as

$$\begin{aligned}
&\left\langle \tilde{\mathcal{L}}_t f_{\varphi_1,\cdots,\varphi_m}, \mu_t\right\rangle \\
&= \int_{\mathcal{H}} \langle Df_{\varphi_1,\cdots,\varphi_m}(u), \mathcal{G}(t, u)\rangle_{\mathcal{H}_Q}\, \mu_t(du) \\
&= \underbrace{\int_{\mathcal{H}} \left[\langle Df_{\varphi_1,\cdots,\varphi_m}(u), B(t, u)\rangle_{\mathcal{H}_Q} - \frac{1}{2}\left\langle Df_{\varphi_1,\cdots,\varphi_m}(u), A(t)\rho^{\mu_t}_{\mathcal{H}_Q}(u)\right\rangle_{\mathcal{H}_Q}\right] \mu_t(du)}_{=:(\mathrm{I})}.
\end{aligned}$$

On the other hand, note that one already knows that $\{\mu_t\}_{t\in[0,T]}$ satisfies the Kolmogorov forward equation for the original stochastic differential equation. That is, if we define

$$\mathcal{L}_t f_{\varphi_1,\cdots,\varphi_m}(u) = \frac{1}{2}\,\text{Tr}_{\mathcal{H}_Q}\left(A(t) \circ Q \circ D^2 f_{\varphi_1,\cdots,\varphi_m}(u)\right) + \langle Df_{\varphi_1,\cdots,\varphi_m}(u), B(t, u)\rangle_{\mathcal{H}_Q}$$

for $f_{\varphi_1,\cdots,\varphi_m} \in \mathcal{FC}^\infty_b(\mathcal{H})$, then the time-evolution of $\{\mu_t\}_{t\in[0,T]}$ can be expressed as the following Cauchy problem in a weak sense (Belopolskaya & Dalecky, 2012):

$$\begin{cases} \partial_t\mu_t = (\mathcal{L}_t)^*\mu_t, & t \in (0, T), \\ \mu_t\Big|_{t=0} = \mathbf{P}_0, \end{cases} \tag{8}$$

where $(\mathcal{L}_t)^*$ denotes the formal adjoint of $\mathcal{L}_t$. From the forward equation (Eqn. (8)), one has that

$$\begin{aligned}
&\partial_t \langle f_{\varphi_1,\cdots,\varphi_m}, \mu_t\rangle \\
&= \underbrace{\int_{\mathcal{H}} \left[\frac{1}{2}\,\text{Tr}_{\mathcal{H}_Q}\left(A(t) \circ Q \circ D^2 f_{\varphi_1,\cdots,\varphi_m}(u)\right) + \langle Df_{\varphi_1,\cdots,\varphi_m}(u), B(t, u)\rangle_{\mathcal{H}_Q}\right] \mu_t(du)}_{=:(\mathrm{II})}
\end{aligned}$$

for every $f_{\varphi_1,\cdots,\varphi_m} \in \mathcal{FC}^\infty_b(\mathcal{H})$. Hence, in order to prove Eqn. (7), one has to check if (I) = (II). To establish this result, it suffices to prove the following

Claim: *It holds that*

$$\left.\begin{aligned} &-\int_{\mathcal{H}} \left\langle Df_{\varphi_1,\cdots,\varphi_m}(u), A(t)\rho^{\mu_t}_{\mathcal{H}_Q}(u)\right\rangle_{\mathcal{H}_Q} \mu_t(du) \\ &\overset{(!)}{=} \int_{\mathcal{H}} \mathrm{Tr}_{\mathcal{H}_Q}\left(A(t)\circ Q\circ D^2 f_{\varphi_1,\cdots,\varphi_m}(u)\right)\mu_t(du) \end{aligned}\right\}. \tag{9}$$

Before proving the preceding claim, we first establish an auxiliary result on the first- and second-order Fréchet derivatives of cylindrical functions.

Lemma B.1. *For $f_{\varphi_1,\cdots,\varphi_m} \in \mathcal{FC}^\infty_b(\mathcal{H})$, $f \in \mathcal{C}^\infty_0(\mathbf{R}^m)$, its first- and second-order Fréchet derivatives are given by*

$$\begin{aligned} Df_{\varphi_1,\cdots,\varphi_m}(u) &= \sum_{i=1}^m (\partial_i f)(\langle\varphi_1,u\rangle,\cdots,\langle\varphi_m,u\rangle)\varphi_i, \\ D^2 f_{\varphi_1,\cdots,\varphi_m}(u)(h) &= \sum_{i,j=1}^m \varphi_i(\partial^2_{ij} f)(\langle\varphi_1,u\rangle,\cdots,\langle\varphi_m,u\rangle)\langle\varphi_j,h\rangle, \quad h\in\mathcal{H}. \end{aligned}$$

We then have

$$\begin{aligned} &\mathrm{Tr}_{\mathcal{H}_Q}\left(A(t)\circ Q\circ D^2 f_{\varphi_1,\cdots,\varphi_m}(u)\right) \\ =&\sum_{\ell=1}^\infty \left\langle A(t)\circ Q\circ D^2 f_{\varphi_1,\cdots,\varphi_m}(u)(Q^{1/2}\varphi_\ell),\, Q^{1/2}\varphi_\ell\right\rangle_{\mathcal{H}_Q} \\ =&\sum_{\ell=1}^\infty \left\langle A(t)\circ Q\circ\left(\sum_{i,j=1}^m (\partial^2_{ij} f)(\varphi_1(u),\cdots,\varphi_m(u))\left\langle\varphi_j, Q^{1/2}\varphi_\ell\right\rangle_{\mathcal{H}}\varphi_i\right),\, Q^{1/2}\varphi_\ell\right\rangle_{\mathcal{H}_Q} \\ &\hfill \text{(Lemma B.1)} \\ =&\sum_{i,j=1}^m (\partial^2_{ij} f)(\varphi_1(u),\cdots,\varphi_m(u))\left[\sum_{\ell=1}^\infty \left\langle A(t)\circ Q(\varphi_i), Q^{1/2}\varphi_\ell\right\rangle_{\mathcal{H}_Q}\left\langle\varphi_j, Q^{1/2}\varphi_\ell\right\rangle_{\mathcal{H}}\right] \\ =&\sum_{i,j=1}^m (\partial^2_{ij} f)(\varphi_1(u),\cdots,\varphi_m(u))\left[\sum_{\ell=1}^\infty \left\langle A(t)\circ Q(\varphi_i), Q^{1/2}\varphi_\ell\right\rangle_{\mathcal{H}_Q}\left\langle Q^{1/2}\varphi_j, Q^{1/2}\circ Q^{1/2}\varphi_\ell\right\rangle_{\mathcal{H}_Q}\right] \\ =&\sum_{i,j=1}^m (\partial^2_{ij} f)(\varphi_1(u),\cdots,\varphi_m(u))\left[\sum_{\ell=1}^\infty \left\langle A(t)\circ Q(\varphi_i), Q^{1/2}\varphi_\ell\right\rangle_{\mathcal{H}_Q}\left\langle Q\varphi_j, Q^{1/2}\varphi_\ell\right\rangle_{\mathcal{H}_Q}\right] \\ &\hfill (Q^{1/2}\text{ is self-adjoint}) \\ =&\sum_{i,j=1}^m (\partial^2_{ij} f)(\varphi_1(u),\cdots,\varphi_m(u))\left\langle A(t)\circ Q(\varphi_i), Q(\varphi_j)\right\rangle_{\mathcal{H}_Q} \\ =&\sum_{i=1}^m\left[\sum_{j=1}^m \partial_j\left(\partial_i f(\varphi_1(u),\cdots,\varphi_m(u))\right)\left\langle A(t)\circ Q(\varphi_i), Q(\varphi_j)\right\rangle_{\mathcal{H}_Q}\right] \\ =&\sum_{i=1}^m \left\langle D(\partial_i f_{\varphi_1,\cdots,m})(u), A(t)\circ Q(\varphi_i)\right\rangle_{\mathcal{H}_Q}. \qquad \text{(Definition of the Fréchet derivative)} \end{aligned}$$

Now, from the preceding chain of equalities, we write

$$\begin{aligned} &\int_{\mathcal{H}} \mathrm{Tr}_{\mathcal{H}_Q}\left(A(t)\circ Q\circ D^2 f_{\varphi_1,\cdots,\varphi_m}(u)\right)\mu_t(du) \\ =&\sum_{i=1}^m \int_{\mathcal{H}} \left\langle D(\partial_i f_{\varphi_1,\cdots,m})(u), A(t)\circ Q(\varphi_i)\right\rangle_{\mathcal{H}_Q}\mu_t(du) \end{aligned}$$

$$
\begin{aligned}
&= -\sum_{i=1}^{m} \int_{\mathcal{H}} \partial_i f_{\varphi_1,\cdots,\varphi_m}(u)\, \rho^{\mu_t}_{A(t)\circ Q(\varphi_i)}(u)\, \mu_t(du) && \text{(Integration-by-parts)} \\
&= -\sum_{i=1}^{m} \int_{\mathcal{H}} \partial_i f_{\varphi_1,\cdots,\varphi_m}(u) \left\langle \rho^{\mu_t}_{\mathcal{H}_Q}(u), A(t)\circ Q(\varphi_i) \right\rangle_{\mathcal{H}_Q} \mu_t(du) \\
& && \text{(Definition of the logarithmic gradient)} \\
&= -\sum_{i=1}^{m} \int_{\mathcal{H}} \partial_i f_{\varphi_1,\cdots,\varphi_m}(u) \left\langle A(t)\rho^{\mu_t}_{\mathcal{H}_Q}(u), Q(\varphi_i) \right\rangle_{\mathcal{H}_Q} \mu_t(du),
\end{aligned}
$$

which eventually proves the claim Eqn. (9), and hence the theorem. □

Proof of Lemma B.1. From the linearity of inner product, note that

$$
\begin{aligned}
f_{\varphi_1,\cdots,\varphi_m}(u+h) &= f(\langle \varphi_1, u+h\rangle, \cdots, \langle \varphi_m, u+h\rangle) \\
&= f(\langle \varphi_1, u\rangle + \varepsilon_1, \cdots, \langle \varphi_m, u\rangle + \varepsilon_m),
\end{aligned}
$$

where $\varepsilon_i = \langle \varphi_i, h\rangle$. From the Cauchy-Schwwarz inequality, it is clear that $\varepsilon_i \to 0$ as $\|h\| \to 0$ for each $i = 1, 2, \cdots, m$. Because $f \in \mathcal{C}_0^\infty(\mathbf{R}^m)$ is smooth, it follows from the Taylor expansion of f applied to the RHS of the preceding equality that

$$
\begin{aligned}
&f_{\varphi_1,\cdots,\varphi_m}(u+h) \\
&\quad = f(\langle \varphi_1, u\rangle, \cdots, \langle \varphi_m, u\rangle) + \sum_{i=1}^{m} (\partial_i f)(\langle \varphi_1, u\rangle, \cdots, \langle \varphi_m, u\rangle)\langle \varphi_i, h\rangle + o(\|h\|).
\end{aligned}
$$

Therefore, it follows that

$$
\partial_h f_{\varphi_1,\cdots,\varphi_m}(u) = \sum_{i=1}^{m} (\partial_i f)(\langle \varphi_1, u\rangle, \cdots, \langle \varphi_m, u\rangle)\langle \varphi_i, h\rangle.
$$

Because $\partial_h f_{\varphi_1,\cdots,\varphi_m}(u) = \langle D f_{\varphi_1,\cdots,\varphi_m}(u), h\rangle$, we conclude that

$$
D f_{\varphi_1,\cdots,\varphi_m}(u) = \sum_{i=1}^{m} (\partial_i f)(\langle \varphi_1, u\rangle, \cdots, \langle \varphi_m, u\rangle)\varphi_i.
$$

Repeating the same method, one observes that

$$
\begin{aligned}
&D f_{\varphi_1,\cdots,\varphi_m}(u+h) \\
&\quad = D f_{\varphi_1,\cdots,\varphi_m}(u) + \sum_{i,j=1}^{m} (\partial^2_{ij} f)(\langle \varphi_1, u\rangle, \cdots, \langle \varphi_m, u\rangle)\varphi_i \langle \varphi_j, h\rangle + o(\|h\|),
\end{aligned}
$$

from which the second statement is deduced. This completes the proof. □

C EXPERIMENTAL DETAILS

C.1 TRAINING AND SAMPLING

Training. We approximate $\rho_{\mathcal{H}_Q}^{\mu_t}(u)$ via a Fourier neural operator $S_\theta(t,u)$ parametrized by θ. The training of S_θ is done via the score-matching objective (Vincent, 2011; Sohl-Dickstein et al., 2015):

$$\underset{\theta}{\text{minimize}} \quad \mathcal{L}(\theta) = \int_0^T \mathop{\mathbb{E}}_{X_0\sim\mathbf{P}_0}\left[\mathop{\mathbb{E}}_{X_t\sim\mu_{t|X_0}}\left[\left\|S_\theta(t,X_t) - \rho_{\mathcal{H}_Q}^{\mu_{t|X_0}}(X_t)\right\|^2\right]\right],$$

where $\mu_{t|X_0}$ is the conditional measure of X_t given X_0. We use the variance-preserving SDE (Song et al., 2021b) in $\mathcal{H}$ of the form

$$dX_t^{\rightarrow} = -\frac{\alpha(t)}{2}X_t^{\rightarrow} + \sqrt{\alpha(t)}X_t^{\rightarrow}\,dW_t, \quad X_0 \sim \mathbf{P}_0 = \mathbf{P}_{\text{data}}$$

where $(W_t)_{t\geq 0}$ is a Q-Wiener process in $\mathcal{H}$.

Sampling. Once S_θ is learned, we sample synthetic data by plugging in $S_\theta(T-t, Y_{T-t}^{\leftarrow})$ in place of $\rho_{\mathcal{H}_Q}^{\mu_{T_t}}(Y_{T-t}^{\leftarrow})$ in our probability-flow ODE (Theorem 3.1) as

$$dY_{T-t}^{\leftarrow} = \frac{\alpha(T-t)}{2}\left[Y_{T-t}^{\leftarrow} + \rho_{\mathcal{H}_Q}^{\mu_{T-t}}(Y_{T-t}^{\leftarrow})\right]dt, \quad Y_0 \sim \mathcal{N}(0,Q), \tag{10}$$

and run the following plug-and-play ODE

$$dY_{T-t}^{\leftarrow} = \frac{\alpha(T-t)}{2}\left[Y_{T-t}^{\leftarrow} + S_\theta(T-t, Y_{T-t}^{\leftarrow})\right]dt, \quad Y_0 \sim \mathcal{N}(0,Q). \tag{11}$$

In every example in this work, we utilize Euler's method to solve the preceding plug-and-play ODE (Eqn. (11)).

C.2 APPROXIMATION OF $\mathcal{N}(0,Q)$

1D (Quadratic) function generation. For one-dimensional function generation task, we let $k(\cdot,\cdot) : [a,b]\times[a,b] \to \mathbf{R}$ be a positive-definite kernel. We let Q be the integral operator on $L^2([a,b])$ corresponding to k. Let $\mathcal{D} = \{x_1, \cdots, x_N\}$ be the fixed grid where functions are evaluated. Define the Gram matrix $K \in \mathrm{Mat}_N(\mathbf{R})$ by $K_{ij} = k(x_i, x_j)$, and let $K = \Phi D \Phi^\intercal$ be the eigen-decomposition of K. As in Baker & Taylor (1979); Phillips et al. (2022); Lim et al. (2023b), we generate a random noise $W \sim \mathcal{N}(0,Q)$ by

$$W = ZD^{1/2}\Phi^\intercal, \quad Z \sim \mathcal{N}(0, id_N).$$

We choose our k to be the Gaussian RBF kernel of the form

$$k(x_1,x_2) = \texttt{gain}\, e^{-|x_1-x_2|^2/\texttt{len}^2}, \quad x_1, x_2 \in [a,b].$$

The hyperparameters `gain` and `len` are called the gain parameter and the length parameter, respectively.

PDE problems. For PDE solution generation tasks (reaction-diffusion equation and time-evolving heat equation), we use the Bessel prior $\mathcal{N}(0,(\gamma-\Delta)^{-s})$, which is introduced in Hagemann et al. (2023), as our noise $\mathcal{N}(0,Q)$ in the ambient Hilbert space $\mathcal{H}$. Sampling from $\mathcal{N}(0,(\gamma-\Delta)^{-s})$ is done by computing

$$W = \mathrm{FFT}^{-1}\left(\lambda(\gamma-\Delta)^{-s/2}) \odot \mathrm{FFT}(Z)\right), \quad Z \sim \mathcal{N}(0, id_{N^2}),$$

where $\lambda((\gamma-\Delta)^{-s/2})$ is the vector whose entries consist of eigenvalues of $(\gamma-\Delta)^{-s/2}$, N is the resolution of samples, $\odot$ denotes the entry-wise product, and FFT and FFT^{-1} denotes the Fast Fourier Transform and its inverse transform, respectively. The hyperparameter $\gamma > 0$ is called the `scale` parameter, and $s > 0$ is called the `power` parameter.

C.3 IMPLEMENTATIONAL DETAILS

In this subsection, we provide details regarding the architecture and training details of the infinite-dimensional diffusion models used in our experiments.

1D (Quadratic) function generation. For the one-dimensional (`Quadratic`) function generation task, we use a modified version of Fourier Neural Operator (Li et al., 2020) that is proposed in Lim et al. (2023b). Here, we use the pre-trained checkpoint by Lim et al. (2023b) without additional training[1]. Table 3 lists up detailed architectural design used in the 1D function generation experiment.

Table 3: Architectural details for 1D (`Quadratic`) function generation

Architecture	Base channels	256
	# of ResBlocks per stage	4
	Lifting channels	256
	Projection channels	256
	# of modes	[100]
	Activation function	Gelu
	Normalization	GroupNorm
Diffusion	Noise schedule	Cosine
	# timesteps	1000
	$\log(\alpha_0^2/\sigma_0^2)$	10
	$\log(\alpha_1^2/\sigma_1^2)$	-10
	Length parameter	0.8
	Gain parameter	1.0

2D Reaction-diffusion equation. For the two-dimensional reaction-diffusion equation problem, we use a two-dimensional Fourier Neural Operator. We follow the architectural detail used in Hagemann et al. (2023)[2], which is based on the official implementation of Neural Operators (Kovachki et al., 2021; Kossaifi et al., 2024)[3]. Table 4 lists up detailed architectural design used in the reaction-diffusion equation experiment.

Table 4: Implementational details for 2D (reaction-diffusion) experiment

Architecture	Base channels	32
	# of ResBlocks per stage	4
	Lifting channels	32
	Projection channels	128
	# of modes	[12, 12]
	Activation function	Gelu
	Normalization	None
Diffusion	Noise schedule	Cosine
	# timesteps	1000
	$\log(\alpha_0^2/\sigma_0^2)$	10
	$\log(\alpha_1^2/\sigma_1^2)$	-10
	Scale parameter	8
	Power parameter	0.55
Training	Optimizer	Adam, $\beta_1 = 0.9$, $\beta_2 = 0.999$
	Learning rate	0.001
	# epochs	200
	Batch size	8

[1]Official code repository: https://github.com/KU-LIM-Lab/hdm-official/
[2]Official code repository: https://github.com/PaulLyonel/multilevelDiff/
[3]Official code repository: https://github.com/neuraloperator/neuraloperator/

2D Heat equation. For the two-dimensional heat equation, we discretize time-evolving solutions of the heat equation on a three-dimensional domain: two dimensions for spacial and one for time. We use a three-dimensional Fourier Neural Operator (Li et al., 2020; Kovachki et al., 2021; Kossaifi et al., 2024). Table 5 lists up implementational details used in the time-evolving heat equation experiment.

Table 5: Implementational details for time-evolving heat equation experiment

Architecture	Base channels	32
	# of ResBlocks per stage	4
	Lifting channels	32
	Projection channels	128
	# of modes	[12, 12, 12]
	Activation function	Gelu
	Normalization	None
Diffusion	Noise schedule	Cosine
	# timesteps	1000
	$\log(\alpha_0^2/\sigma_0^2)$	10
	$\log(\alpha_1^2/\sigma_1^2)$	-10
	Scale parameter	8
	Power parameter	0.55
Training	Optimizer	Adam, $\beta_1 = 0.9$, $\beta_2 = 0.999$
	Learning rate	0.001
	# epochs	200
	Batch size	4