



Review

Cite this article: Franzese G, Michiardi P.
2025 Generative diffusion models in infinite
dimensions: a survey. *Phil. Trans. R. Soc. A* **383**:
20240322.
<https://doi.org/10.1098/rsta.2024.0322>

Received: 21 October 2024

Accepted: 13 March 2025

One contribution of 13 to a theme issue
'Generative modelling meets Bayesian
inference: a new paradigm for inverse
problems'.

Subject Areas:

artificial intelligence

Keywords:

stochastic differential equations, generative
diffusion models, Hilbert spaces

Author for correspondence:

Giulio Franzese

e-mail: giulio.franzese@eurecom.fr

Generative diffusion models in infinite dimensions: a survey

Giulio Franzese¹ and Pietro Michiardi¹

¹Data Science, EURECOM, Biot, France

GF, 0000-0003-4244-2053

Diffusion models have recently emerged as a powerful class of generative models, achieving state-of-the-art performance in various domains such as image and audio synthesis. While most existing work focuses on finite-dimensional data, there is growing interest in extending diffusion models to infinite-dimensional function spaces. This survey provides a comprehensive overview of the theoretical foundations and practical applications of diffusion models in infinite dimensions. We review the necessary background on stochastic differential equations in Hilbert spaces, and then discuss different approaches to define generative models rooted in such formalism. Finally, we survey recent applications of infinite-dimensional diffusion models in areas such as generative modelling for function spaces, conditional generation of functional data and solving inverse problems. Throughout the survey, we highlight the connections between different approaches and discuss open problems and future research directions.

This article is part of the theme issue 'Generative modelling meets Bayesian inference: a new paradigm for inverse problems'.

1. Introduction

Diffusion models have emerged as a prominent class of generative models, achieving notable success in various fields such as image generation [1], audio synthesis [2,3], video [4,5], molecular structures and general three-dimensional shapes [6–9] and conditional generation from open-ended text prompts [10–13]. For applications involving stochastic processes and complex systems where finite-dimensional approximations are insufficient, like weather forecasts or seismology [14,15] or

extremely high-resolution point clouds [16], generative models which operate directly in function space are required. This survey aims to provide a structured overview of the theoretical and practical developments in generative diffusion models operating in infinite dimensions.

A fundamental motivation for considering infinite-dimensional approaches arises from Bayesian inverse problems in function spaces. The work in [17] established a rigorous mathematical framework for inference in such settings, emphasizing that naive finite-dimensional approximations often introduce undesirable pathologies. Rather than discretizing early in the modelling process, a more principled approach is to *avoid discretization until the last possible moment* [17]. Numerous studies have demonstrated that working directly in infinite-dimensional spaces leads to improved stability and accuracy in inference tasks [18,19].

The necessity of such models stems from the intrinsic limitations of finite-dimensional approximations, particularly their inconsistency across different resolutions. In empirical studies on generative diffusion models, it has been widely observed that models trained at one resolution often fail to generalize across different resolutions or scales unless carefully adapted [20–23]. This issue extends beyond neural architecture design (e.g. parameter adjustments) and is fundamentally linked to the role of *trace-class noise* in the perturbation process [24–31]. When the additive noise has a flat spectral density, changing the sampling frequency (i.e. the resolution) typically alters the effective signal-to-noise ratio, affecting consistency in generative models.

In contrast, infinite-dimensional methods allow for the construction of generative models that generalize effectively across different resolutions and are not constrained by regular sampling grids [24–33]. Additionally, infinite-dimensional generative models have been shown to provide enhanced parameter efficiency [26]. These models inherently accommodate function spaces in a way that ensures robustness and consistency in inference and generation tasks, making them particularly well-suited for applications requiring flexibility across multiple scales.

The problem of generative modelling in functional domains can be approached by adapting the techniques valid for the finite-dimensional case to the one where observed data points are considered as finite-resolution realizations of underlying functions [34–36]. These approaches allow for the incorporation of the notion of underlying manifold (whether Euclidean space or not) on which the data is collected, without the need to explicitly consider the machinery of infinite-dimensional stochastic calculus [37]. However, scaling these methods to arbitrary resolution is challenging. A different approach, which has been incorporated by some of the purely functional domain diffusion models, is to consider pseudo-invertible mappings from infinite-dimensional domains onto finite-dimensional ones and consider classical diffusion in those finite spaces [38,39]. Given the limitations of such approaches, a purely function space investigation of diffusion processes has been explored by the community, which we describe in the following.

(a) An overview of existing methods and applications

Although the primary focus of this survey is the exploration of *infinite*-dimensional diffusion models, we first provide an overview of key generative models based on diffusion and related dynamics in finite dimensions, to set the stage for a discussion of their infinite-dimensional counterparts and applications.

On the left column of [table 1](#), we summarize the major classes of finite-dimensional generative models. Among these, *discrete-time* diffusion models, commonly referred to as DDPM [5,40,41], learn to approximate the reversal of a forward Markov chain with a finite number of steps. This forward process perturbs the original data, typically through additive Gaussian noise, which smooths the data distribution. A related approach, annealed Langevin dynamics, replaces single-step transitions with multiple runs of Langevin dynamics, using the perturbed score function as the drift term to iteratively refine the noisy samples [42,43].

As the number of steps tends to infinity, the Markov chain formulation converges to a continuous-time setting, leading to score-based SDEs [1], which model the forward and reverse

Table 1. Finite-dimensional generative models and infinite-dimensional counterparts.

finite-dimensional	infinite-dimensional	position in survey
DDPM [5,40,41]	[28]	§4a
annealed-LD [42,43]	[30]	§3a
score-SDE [1]	[27,31,44,45]	§2a
flow matching [46,47]	[29]	§3b
diffusion bridges and stochastic control [48,49]	[32,33,50]	§5a
Bayesian inverse problems [51,52]	[24,25]	§5b

evolution through stochastic differential equations. Another approach, known as *Flow Matching* [46,47], models the perturbation dynamics deterministically, inducing transport flows that map between the initial noisy distribution and the target data distribution.

Beyond generative modelling, diffusion models have been applied to stochastic control problems and Bayesian inverse problems. In finite dimensions, diffusion bridges provide a framework for modelling stochastic paths conditioned on endpoints, making them useful in molecular dynamics and controlled trajectory planning [48,49]. Similarly, diffusion-based priors have been leveraged in Bayesian inverse problems, where they help reconstruct unknown samples from noisy measurements [51,52]. By providing strong probabilistic priors, these models enhance inference accuracy and enable uncertainty quantification in complex systems.

The transition to infinite-dimensional settings introduces additional mathematical and computational challenges, which we explore in later sections. In particular, we discuss the infinite-dimensional counterpart of score-SDEs [27,31,44,45] in §2a, where they appear as specific cases of equation (2.4). Alternative formulations, including annealed Langevin dynamics [30] and flow matching [28], are examined in §3a,b, in connection with equations (3.1) and (3.4). Applications of diffusion bridges and Bayesian inverse problems in infinite dimensions are covered in §5a,b, particularly in equations (5.2)–(5.4).

(b) Structure of the manuscript

In §2, we delve into the formal definition of diffusion models in infinite dimensions, focusing on Hilbert spaces. We introduce the necessary mathematical foundations, such as SDEs in Hilbert spaces, the properties of infinite-dimensional diffusion processes and key concepts like time-reversal of SDEs and the formulation of probability measures. Special attention is given to generative processes driven by these models, alongside theoretical considerations such as existence and uniqueness of solutions. In §3, we explore alternative methodologies for constructing generative models in infinite-dimensional settings, contrasting them with the diffusion-based approaches discussed earlier. We review methods like annealed Langevin dynamics and functional flow matching and connect their mathematical underpinnings to diffusion models, highlighting both their strengths and limitations. In §4, we outline the loss functions and training objectives associated with these parametric methods, particularly in terms of denoising and likelihood maximization. We also briefly overview neural architectures suitable for functional data and analyse their ability to handle resolution invariance. In §5, we provide an overview of various real-world applications of infinite-dimensional diffusion models. Other than purely generative use cases, we discuss the solution of Bayesian inverse problems, stochastic bridges and optimal control.

Finally, in §6, we summarize the key takeaways from this survey, highlighting the theoretical advancements and practical applications of diffusion models in infinite dimensions. We also outline open problems and future research directions, suggesting ways to further refine these models and explore their broader use in generative modelling and related fields.

2. Diffusion-based generative models in Hilbert spaces

We consider H to be a real, separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$, norm $\|\cdot\|_H$. Let $L(H)$ be the set of bounded linear operators on H , $B(H)$ be its Borel σ -algebra, $B_b(H)$ be the set of bounded $B(H)$ -measurable functions $H \rightarrow \mathbb{R}$ and $P(H)$ be the set of probability measures on $(H, B(H))$.

Our focus is to survey existing *generative models* which leverage the properties of diffusion processes in Hilbert spaces. Informally, given a finite collection of samples drawn from an underlying probability measure $\mu \in P(H)$, we consider different, but strictly related techniques, to *transform* a sample drawn from an *easy* measure into a sample whose measure coincides with the desired μ . Consider the following H -valued SDE whose initial conditions are drawn from μ

$$\begin{cases} dX_t = (\mathcal{A}X_t + f(X_t, t))dt + dW_t, \\ X_0 \sim \mu = \mu \in P(H), \end{cases} \quad (2.1)$$

where $t \in [0, T]$, W_t is a R -Wiener process on H (of *trace* class) defined on the quadruplet $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})$. In our definitions, Ω, \mathcal{F} are the sample space and canonical filtration, respectively: we consider Ω to be $C^1([0, T])$, that is the space of all continuous mappings $[0, T] \rightarrow H$, and $X_t(\omega) = \omega(t), \omega \in \Omega$ to be the *canonical* process. The domain of f is $D(f) \in B([0, T] \times H)$, where $f : D(f) \subset [0, T] \times H \rightarrow H$ is a measurable map. f is assumed to be Lipschitz continuous. The negative, symmetric operator $\mathcal{A} : D(\mathcal{A}) \subset H \rightarrow H$ is the infinitesimal generator of a C_0 -semigroup $\exp(t\mathcal{A})$ in $H(t \geq 0)$, and μ is a probability measure in H . By definition, the measure associated with equation (2.1) is indicated with \mathbb{Q} . The *law* induced at time $\tau \in [0, T]$ by the canonical process on the measure \mathbb{Q} is indicated with $\rho_\tau \in P(H)$, where $\rho_\tau(S) = \mathbb{Q}(\{\omega \in \Omega : X_\tau(\omega) \in S\})$, and S is any element of $B(H)$. We define the law associated with the process defined in equation (2.1), when initial conditions are deterministic ($X_0 = z$) as ρ_τ^z . Notice that, by disintegration properties,

$$d\rho_\tau(x) = \int d\rho_\tau^z(x) d\mu(z). \quad (2.2)$$

Given a countable orthonormal basis $\{e^k\}_{k \in \mathbb{N}}$ for the Hilbert space H , we can always construct an isomorphism with the space of square summable sequences $l_2(\mathbb{N})$. Since R is a trace class operator, there exists a complete orthonormal system in H [37] that *diagonalizes* R , i.e. $Re^k = r^k e^k$, with r^k a positive scalar, where we leverage the fact that R is a symmetric operator being a covariance. The covariance operator can be used to define the Cameron Martin space $H_R \stackrel{\text{def}}{=} R^{\frac{1}{2}}H$ associated with H , where the corresponding induced scalar product is $\langle \cdot, \cdot \rangle_{H_R} = \langle R^{-\frac{1}{2}} \cdot, R^{-\frac{1}{2}} \cdot \rangle$.

Then, it is useful to notice that equation (2.1) can also be expressed as an (infinite) system of stochastic differential equations, in terms of $X_t^k = \langle X_t, e^k \rangle$, as:

$$dX_t^k = a^k(X_t) + f^k(X_t, t)dt + dW_t^k, \quad k \in \mathbb{N}, \quad (2.3)$$

where we introduced the projection $a^k(x) = \sum_{j=1}^{\infty} a^{kj} x^j$, with $a^{kj} \stackrel{\text{def}}{=} \langle \mathcal{A}e^j, e^k \rangle$ and $f^k(x, t) = \langle f(x, t), e^k \rangle$.

Moreover, $dW_t^k = \langle dW_t, e^k \rangle$ with covariance given by $\mathbb{E}_{\mathbb{Q}}[W_t^k W_s^j] = \delta(k-j)r^k \min(s, t)$, δ in the Kronecker sense. Under condition 2.1, the SDE equation (2.1) admits a *strong* solution [37].

Condition 2.1. (i) $R^{\frac{1}{2}}H \subset D(\mathcal{A})$, (ii) $\mathcal{A}R^{\frac{1}{2}}$ is a Hilbert–Schmidt operator, and (iii) $\mu(D(\mathcal{A})) = 1$. Notice that this condition is trivially satisfied whenever \mathcal{A} is a bounded operator.

(a) Time reversal of SDEs

The idea of generative models based on SDEs in the form of equation (2.1) is to consider the *time-reversal* of the stochastic process X_t , namely the process $\hat{X}_t \stackrel{\text{def}}{=} X_{T-t}$. Under certain conditions, the

process \hat{X}_t can be shown to be a solution of a new SDE. When this is the case, simulation of its dynamics allows obtaining a sample from the desired measure ($\hat{X}_T = X_0 \sim \mu$).

More formally, we require that the time reversal of the canonical process, $\hat{X}_t = X_{T-t}$, is again a diffusion process, with distribution given by the *path-reversed* measure $\hat{\mathbb{Q}}(\omega)$, along with the reversed filtration \mathcal{F} . In the finite-dimensional case (which corresponds to the case where the Hilbert space has finite dimensionality), several results exist which prove (under appropriate technical conditions) that the time reversal of SDEs in the form of equation (2.1) (although finite dimensional) has again expression in terms of an SDE, with a new drift term composed as $-a^k(x) - f^k(x, t) + r^k \frac{\partial}{\partial x^k} \log \left(\frac{d\rho_T(x)}{dx} \right)$, where $\frac{d\rho_T(x)}{dx}$ is the *density* of the measure w.r.t. the Lebesgue measure dx [53,54]. However, the time reversal of an infinite-dimensional process is more involved than for the finite-dimensional case. Indeed, in infinite-dimensional spaces, there is no equivalent of the Lebesgue measure to obtain densities from measures (i.e. dx in \mathbb{R}^N does not exist [55]). However, in the infinite-dimensional case, we can consider the single-dimensional conditional density $\frac{d\rho_T(x^i | x^{j \neq i})}{dx^i}$ (provided its existence), dx^i being the single-dimensional Lebesgue measure. At a rather informal level, the reader can understand why considering such objects can alleviate the problem of non-existence of the desired densities in infinite dimensions: in the finite-dimensional case, thanks to disintegration of measures, the second term of the time-reversed drift can be simplified as $\frac{\partial}{\partial x^k} \log \left(\frac{d\rho_T(x^k | x^{j \neq k})}{dx^k} \frac{d\rho_T(x^{j \neq k})}{dx^{j \neq k}} \right) = \frac{\partial}{\partial x^k} \log \left(\frac{d\rho_T(x^k | x^{j \neq k})}{dx^k} \right)$.

It is then possible to derive results for the infinite-dimensional case which are analogous to the finite-dimensional one. There are two major approaches to guarantee the existence of the reverse diffusion process. The first approach [56] relies on a finite local entropy condition. The second approach is based on stochastic calculus of variations [57], which we use to claim what follows.

Theorem 2.1. Consider equation (2.1). Suppose that the conditions listed in Theorem 4.3 of [57] hold. Then \hat{X}_t , corresponding to the path measure $\hat{\mathbb{Q}}(\omega)$, has the following SDE representation:

$$\begin{cases} d\hat{X}_t = (-A\hat{X}_t - f(\hat{X}_t, T-t) + R\nabla \log \rho_{T-t}(\hat{X}_t)) dt + d\hat{W}_t, \\ \hat{X}_0 \sim \rho_T, \end{cases} \quad (2.4)$$

where \hat{W} is a $\hat{\mathbb{Q}}$ -Wiener process and the notation $R\nabla \log \rho_t(x)$ stands for the mapping $H \rightarrow H$ that, when projected, satisfies $\langle R\nabla \log \rho_t(x), e^k \rangle = r^k \frac{\partial}{\partial x^k} \log \left(\frac{d\rho_t(x^k | x^{j \neq k})}{dx^k} \right)$. By projecting onto the eigenbasis, we have an infinite system of SDEs:

$$d\hat{X}_t^k = \left(-a^k(\hat{X}_t) - f^k(\hat{X}_t, T-t) + r^k \frac{\partial}{\partial x^k} \log \left(\frac{d\rho_{T-t}(\hat{X}_t^k | \hat{X}_t^{j \neq k})}{dx^k} \right) \right) dt + d\hat{W}_t^k, \quad k \in \mathbb{N}. \quad (2.5)$$

Notice that strictly speaking, using the symbol ∇ is an abuse of notation, since we cannot identify it with the gradient (in Hilbert space) of a single function.

The result in [57] is remarkably general and elegant, but this generality comes at the cost of increased complexity in verifying that the necessary technical conditions hold. Although not the most general case, a concrete example of conditions that satisfy the coefficient assumptions outlined in [57] are those considered in [26], which we report hereafter, as discussed in detail in appendix A:

Condition 2.2. $f = 0$ and A is diagonal in $\{e^i\}$, i.e. $a^{ij} = \delta_{ij}a^i$.

Condition 2.3. Condition 2.2 is satisfied, and A is bounded: $-a^i \leq M$, with M a finite positive constant.

It is worth mentioning that much effort has been put in by the authors in [31] to generalize the settings considered here for the dynamics in equation (2.1), in particular with the possibility of considering non-constant operators G_t , which precondition the Wiener process via $G_t dW_t$. In

their work, they focus on the time reversal formula at the level of the generator directly: in general, the existence of time reversal of a process and the validity of time reversal formulas do not imply each other. Such an approach is useful for investigating the relationship between different methodologies, as we will explore later.

For the validity of the methodologies described above, it is required that the conditional densities exist. Next, we introduce a different condition, which will be used to prove the existence of such densities and is at the root of the different generative modelling techniques explored in this survey, the *trait d'union* between all the technical requirements of infinite-dimensional diffusion models.

Condition 2.4. For each $t \in (0, T]$, the measure ρ_t is absolutely continuous with respect to the measure ρ_t^0 .

In our introduction, we considered the generic nonlinear ($f \neq 0$) SDE equation (2.1). However, in almost all actual applications in the literature, the simple linear case where $f = 0$ is considered. This assumption allows a simpler investigation of the validity of time reversal equations (see appendix A). Furthermore, this has tremendous practical benefits, as under such conditions, there exists a weak solution to equation (2.1) as

$$X_t = \exp(t\mathcal{A})X_0 + \int_0^t \exp((t-s)\mathcal{A})dW_s, \quad (2.6)$$

which is typically amenable to cheap simulation. In this case, the conditional distribution admits a simple expression $d\rho_t^z(x) = \mathcal{N}_{\exp(t\mathcal{A})z, \mathcal{S}(t)}(x)$, where $\mathcal{S}(t) = \left(\int_0^t \exp(s\mathcal{A})R \exp(s\mathcal{A})ds \right)$.

Condition 2.2 is particularly helpful in practical implementations, simplifying the computation of terms in the form $\mathcal{A}R$, which would generally require working with a different basis and consequently induce extremely high computational cost. For simplicity of exposition in relating different works, we will often consider this assumption in our writing.

Under condition 2.2, $\langle \mathcal{S}(t)e^j, e^i \rangle \stackrel{\text{def}}{=} s^{ij}(t) = \frac{\exp(2\mathcal{A}t)-1}{2\mathcal{A}} r_i \delta_{ij}$. This simplified expression is central in the works [24,27,45] where reverse time dynamics are proved directly with approaches that do not explicitly require the existence of the conditional densities. Furthermore, in this particular case, the measures $d\mu_t^0(x)$ have a simple Gaussian expression $d\mathcal{N}_{0, \mathcal{S}(t)}(x)$. In particular, in the works [24,25,27], it is assumed that $\mathcal{A}x = -\frac{x}{2}$, and consequently, $R\mathcal{S}^{-1}(t) = 1 - \exp(-t)$, $R\nabla \log \rho_{T-t}(x) = -(1 - \exp(-t))^{-1} \left(x - \exp\left(-\frac{t}{2}\right) \right) \mathbb{E}_{\mathbb{Q}} [X_0 | X_t = x]$. The proof strategy in these works revolves around the conditional expectation directly, without the need to explicitly define conditional densities. This is an advantage over the conditions imposed, e.g. in [56,57], which are designed for generally nonlinear processes. In particular, the key idea adopted for the proofs is to define projected, low-dimensional dynamics for the process X_t , via $P^D(X_t)$, where P^D is the projection operator on the first D basis vectors of H . Then, the existence of a backward process with a finite-dimensional expected value is proven using classical results. The core technical contribution is to show that as $D \rightarrow \infty$ the processes remain well-behaved and converge to the original infinite-dimensional dynamics, allowing us to claim that equations in the form of equation (2.4) (with the substitution of equation (2.7)) are valid generative models. Summarizing, the linear case has the following important property

Theorem 2.2. Assume that $\mu(R^{\frac{1}{2}}H) = 1$ and condition 2.2 holds. Then

- (i) Condition 2.4 holds ($\frac{d\rho_t}{d\rho_t^0}$ exists).
- (ii) The conditional densities $\frac{d\rho_t(x^i | x^{j \neq i})}{dx^i}$ exist.

$$(iii) \quad R\nabla \log \rho_t(x) = -R\mathcal{S}(t)^{-1} (x - \exp(t\mathcal{A})\mathbb{E}_{\mathbb{Q}}[X_0 | X_t = x]), \quad (2.7)$$

$$(iv) \quad R\nabla \log \left(\frac{d\rho_t}{d\rho_t^0} \right) (x) = \exp(t\mathcal{A})\mathbb{E}_{\mathbb{Q}}[X_0 | X_t = x]. \quad (2.8)$$

Proof.

(i) Consider temporarily the assumptions $\mu(R^{\frac{1}{2}}H) = 1$ and $\exp(t\mathcal{A})R^{\frac{1}{2}}H \subseteq \mathcal{S}^{\frac{1}{2}}(t)H$. Condition 2.4 is equal to $\rho_t^0(B) = 0 \Rightarrow \rho_t(B) = 0$ for any measurable set B and any $t \in (0, T]$. Since $\rho_t(B) = \int \rho_t^z(B) d\mu(z) = 0$, the implication is true if $\rho_t^z \ll \rho_t^0$ for μ a.e. z . Given that both measures are Gaussian with means $\exp(t\mathcal{A})z$ and 0 , respectively, and covariance $\mathcal{S}(t)$ the absolute continuity holds if $\exp(t\mathcal{A})z \in \mathcal{S}^{\frac{1}{2}}(t)H$ for μ a.e. z . Given that $\mu(R^{\frac{1}{2}}H) = 1$, this is equivalent to $\exp(t\mathcal{A})R^{\frac{1}{2}}y \in \mathcal{S}^{\frac{1}{2}}(t)H$ for all $y \in H$, which is exactly the assumption $\exp(t\mathcal{A})R^{\frac{1}{2}}H \subseteq \mathcal{S}(t)H$. We now prove that $\exp(t\mathcal{A})R^{\frac{1}{2}}H \subseteq \mathcal{S}^{\frac{1}{2}}(t)H$ holds true if condition 2.2 holds. First, let us notice that the assumption is equivalent to assuming that $\forall y \in H, \exists z \in H : \exp(t\mathcal{A})R^{\frac{1}{2}}y = \mathcal{S}^{\frac{1}{2}}(t)z$. In this case, $\mathcal{S}_t^{\frac{1}{2}} = ((1 - \exp(2t\mathcal{A}))(-2\mathcal{A}^{-1}R))^{1/2}$. Given a single $y \in H$, clearly $z = (-2\mathcal{A})^{\frac{1}{2}}(1 - \exp(2t\mathcal{A}))^{-\frac{1}{2}} \exp(t\mathcal{A})y$ satisfies the equality. Furthermore, its norm is equal to $\|z\|^2 = \langle y, -2\mathcal{A}(1 - \exp(2t\mathcal{A}))^{-1} \exp(2t\mathcal{A})y \rangle = \langle y, 2\mathcal{A}(\exp(-2t\mathcal{A}) - 1)^{-1}y \rangle = \sum_i \frac{2a^i}{\exp(-2ta^i) - 1} (y^i)^2 \leq \sum_i (y^i)^2 < \infty$.

(ii) Condition 2.4 implies that the Radon–Nikodym derivative exists and is well defined.

By restriction, $\frac{d\rho_t(x^{j\neq i})}{d\rho_t^0(x^{j\neq i})}$ exists and is uniquely defined as well. Then, by disintegration

$$\frac{d\rho_t(x)}{d\rho_t^0(x)} = \frac{d\rho_t(x^i | x^{j\neq i})}{d\rho_t^0(x^i | x^{j\neq i})} \frac{d\rho_t(x^{j\neq i})}{d\rho_t^0(x^{j\neq i})},$$

from which we conclude the existence of $\frac{d\rho_t(x^i | x^{j\neq i})}{d\rho_t^0(x^i | x^{j\neq i})}$. Based on this, $\rho_t^0(x^i | x^{j\neq i}) = \mathcal{N}_{0, s^i}(x^i)$. Since $d\rho_t(x^i | x^{j\neq i}) \ll d\mathcal{N}_{0, s^i}(x^i) \ll dx^i$, the existence of the densities is proven.

(iii) $\frac{\partial}{\partial x^i} \log \frac{d\rho_t(x^i | x^{j\neq i})}{d\rho_t^0(x^i | x^{j\neq i})} = \left(\frac{d\rho_t(x^i | x^{j\neq i})}{d\rho_t^0(x^i | x^{j\neq i})} \right)^{-1} \frac{\partial}{\partial x^i} \left(\frac{d\rho_t(x^i | x^{j\neq i})}{d\rho_t^0(x^i | x^{j\neq i})} \right)$. Furthermore, $d\rho_t(x^i | x^{j\neq i}) = \int d\rho_t(x^i | x_0, x^{j\neq i}) d\mu(x_0 | x^{j\neq i})$, where $d\rho_t(x^i | x_0, x^{j\neq i}) = d\mathcal{N}_{\exp(ta_i)x_0^i, s^i(t)}(x^i)$ which has density w.t.t dx^i . Then, $\frac{d\rho_t(x^i | x^{j\neq i})}{d\rho_t^0(x^i | x^{j\neq i})} = \int \mathcal{N}_{\exp(ta_i)x_0^i, s^i(t)}(x^i) d\mu(x_0^i | x^{j\neq i})$, and $\frac{\partial}{\partial x^i} \left(\frac{d\rho_t(x^i | x^{j\neq i})}{d\rho_t^0(x^i | x^{j\neq i})} \right) = - \int \frac{x^i - \exp(ta_i)x_0^i}{s^i(t)} \mathcal{N}_{\exp(ta_i)x_0^i, s^i(t)}(x^i) d\mu(x_0^i | x^{j\neq i})$. Upon simple algebraic manipulation, we can get the desired result (the proof is based on the content of [24,26,33]).

(iv) Due to the Feldman–Hajek theorem [37]

$$\frac{d\rho_t^z}{d\rho_t^0}(x) = \exp \left(\langle R^{-1}x, \exp(t\mathcal{A})z \rangle - 1/2 \|R^{-1/2} \exp(t\mathcal{A})z\|^2 \right).$$

Then, $R\nabla \frac{d\rho_t^z}{d\rho_t^0}(x) = \frac{d\rho_t^z}{d\rho_t^0}(x) \exp(t\mathcal{A})z$. Consequently, $R\nabla \frac{d\rho_t}{d\rho_t^0}(x) = R\nabla \int \frac{d\rho_t^z}{d\rho_t^0}(x) d\mu(z) = \int \frac{d\rho_t^z}{d\rho_t^0}(x) \exp(t\mathcal{A})z d\mu(z) = \int \exp(t\mathcal{A})z d\mu(z | X_t = x) \frac{d\rho_t}{d\rho_t^0}(x)$ and equation (2.8) holds (see [30] for further insights). ■

If the linear operator \mathcal{A} satisfies determinate spectral conditions, then $\rho_t^z \sim \rho_\infty$, for all $z \in H$ and $t > 0$, where both measures have known Gaussian form [37,58–60]. Consequently, $\rho_t^z \sim \rho_\infty \sim \rho_t^0$ for all $z \in H$, which implies $\rho_t^0(B) = 0 \rightarrow \rho_t(B) = \int \rho_t^z(B) d\mu(z) = 0$, i.e. §2.4 holds ($\frac{d\rho_t}{d\rho_t^0}$ exists). Unfortunately, this result, which is stronger than condition §2.4, does not hold for the most commonly used operators \mathcal{A} , like $-I$, as it is required $\exp(t\mathcal{A})z \in \mathcal{S}_t^{\frac{1}{2}}H$, for $z \in H$ generic ([60]).

Theorem 2.1 stipulates that, given some conditions, the reverse time dynamics of equation (2.1) exist in the form of SDE. However, for generative modelling purposes, the content of theorem 2.1 is stronger than necessary. What suffices is that there exists a new SDE whose time-varying law

corresponds to the backward in time of the original ρ_t . Clearly, if a time reversal exists, such a process satisfies the weaker measure matching condition.

To expand on this point and connect together the different generative modelling techniques discussed in the literature, it is necessary to introduce the Kolmogorov operators [31,61–64]

$$\mathcal{L}_0 u(x, t) = D_t u(x, t) + \underbrace{\frac{1}{2} \text{Tr} R D_x^2 u(x, t) + \langle \mathcal{A}x + f(x, t), D_x u(x, t) \rangle}_{\mathcal{L}u(x, t)}, \quad x \in H, t \in [0, T], \quad (2.9)$$

where D_t is the time derivative, D_x, D_x^2 are first- and second-order Fréchet derivatives in space (since both the scalar product and the trace are defined on the Hilbert space H , and not its dual, the quantities in the equation above should be interpreted after Riesz mapping $M \circ$). The domain of the operator is assumed to be the set of smooth cylinder functions of finite support [31,61].

Provided appropriate conditions are satisfied, see for example [61,62], the time-varying measure $d\rho_t(x)$ exists, is unique, and solves the Fokker–Planck equation

$$\int_H u(x, 0) d\mu(x) + \int_0^T ds \int_H \mathcal{L}_0 u(x, t) d\rho_s(x) = 0. \quad (2.10)$$

The result of theorem 2.1 can be thus understood at the level of time reversal of measures: by simply considering the negative of the operator \mathcal{L} , we obtain $\frac{1}{2} \text{Tr}\{-R D_x^2 u(x)\} + \langle -\mathcal{A}x, D_x u(x) \rangle$. While $-\mathcal{A}$ can be associated with a new drift term, it is not possible to construct a Brownian motion term with covariance $-R$. However, $-\frac{1}{2} \text{Tr}\{R D_x^2 u(x)\} = -\text{Tr}\{R D_x^2 u(x)\} + \frac{1}{2} \text{Tr}\{R D_x^2 u(x)\}$. As shown in [31, App. B.2],

$$-\int_H \text{Tr}\{R D_x^2 u(x, t)\} d\rho_s(x) = \int_H \langle R \nabla \log \rho_s(x), D_x u(x, s) \rangle d\rho_s(x), \quad (2.11)$$

which allows us, as anticipated, to prove in a simpler way a *weaker* form of the content of theorem 2.1: an SDE in the form of equation (2.4), defined on a proper probability space, induces a time-varying measure which corresponds to the time-reversal of the original ρ_t , i.e. the measure associated with equation (2.1). Furthermore, the equality in equation (2.11) allows us to claim that the measure ρ_t is also a solution of the continuity equation (see [63–65] for definiteness and uniqueness)

$$\int_H u(x, 0) d\mu(x) + \int_0^T ds \int_H \left(\frac{du(x, t)}{dt} + \langle \mathcal{A}x + f(x, t) - \frac{R}{2} \nabla \log \rho_t(x), D_x u(x, t) \rangle \right) d\rho_t(x) = 0, \quad (2.12)$$

which corresponds to an ordinary differential equation (ODE) in the Hilbert space with deterministic drift

$$\begin{cases} dX_t = \mathcal{A}X_t + f(X_t, t) - \frac{R}{2} \nabla \log \rho_t(X_t) dt \\ X_0 \sim \mu \end{cases}. \quad (2.13)$$

These facts allow us to build a connection between the methodology discussed so far and other prominent approaches from the literature, as shown hereafter.

3. Alternative generative approaches

In §2, we introduced a family of generative models rooted in (variants) of the time reversal formula, shown in equation (2.4). As anticipated, inverting the dynamics of a diffusion process is a sufficient, but not necessary, strategy to obtain a generative model in function space. In this section, we briefly overview some different methodologies which appeared in the literature, clearly connecting their working mechanism and underlying assumptions with the ones described previously.

(a) Annealed Langevin dynamics

Consider the $N + 1$ dimensional sequence of measures $\rho_0, \rho_{T/N}, \dots, \rho_T$. Suppose we have access to computational schemes which allow us, starting from a sample drawn from $\rho_{(i+1)T/N}$, to obtain a sample from the target $\rho_{iT/N}$. Then, a valid generative modelling technique is to run sequentially such computational schemes on the reverse order sequence. One particular approach [30] relies on running multiple annealed chains of Langevin sampling schemes, in the following fashion:

$$d\tilde{X}_t = -\tilde{X}_t dt + \mathcal{S}(iT/N) \nabla \log \left(\frac{d\rho_{iT/N}}{d\rho_{iT/N}^0}(\tilde{X}_t) \right) + \sqrt{2\mathcal{S}(iT/N)R^{-1}} d\tilde{W}_t \quad (3.1)$$

where \tilde{W}_t has covariance R . Dynamics in the form of equation (3.1) have the following fundamental property: their time-invariant measure is $\rho_{iT/N}$ [19,66]. Notice that for the scheme to be valid, the Radon–Nikodym derivative should be valid, i.e. condition 2.4 should hold, as required explicitly by the authors of [30]. Consequently, by selecting the sequence of measures as the law of a diffusion process in the form of equation (2.1), with the conditions described in theorem 2.2, the condition is satisfied, allowing comparison of the methods on the same ground.

To connect this scheme with the one described in [30], it is necessary to notice that $M \circ D_x^{(\mathcal{H}\mathcal{S}(iT/N))} \log \left(\frac{d\rho_{iT/N}}{d\rho_{iT/N}^0} \right) = \nabla^{(\mathcal{H}\mathcal{S}(iT/N))} \log \left(\frac{d\rho_{iT/N}}{d\rho_{iT/N}^0} \right) = \mathcal{S}(iT/N) \nabla^{\mathcal{H}} \log \left(\frac{d\rho_{iT/N}}{d\rho_{iT/N}^0} \right)$. The iterative procedure that starts from a sample $\sim \rho_{(i+1)T/N}$, simulates dynamics equation (3.1) for a sufficiently long time (to reach approximately steady state) and obtains a sample from $\rho_{iT/N}$, is consequently a valid generative model for μ .

Our goal hereafter is to discuss more clearly the connections between this approach and the schemes described previously. Considering a simple time-rescaling argument, it can be shown [37] that the dynamics in equation (3.1) are equivalent to

$$d\tilde{X}_t = -\frac{1}{2} \mathcal{S}(iT/N)^{-1} R \tilde{X}_t dt + \frac{R}{2} \nabla \log \left(\frac{d\rho_{iT/N}}{d\rho_{iT/N}^0}(\tilde{X}_t) \right) dt + d\tilde{W}_t, \quad (3.2)$$

where \tilde{W}_t is a Brownian motion with covariance R . Then, under the assumptions of theorem 2.2, equation (3.2) reads

$$d\tilde{X}_t = \frac{R}{2} \nabla \log(\rho_{iT/N}(\tilde{X}_t)) dt + d\tilde{W}_t, \quad (3.3)$$

which is reminiscent of a classical Langevin equation in \mathbb{R}^N . Clearly, equation (3.3) preserves the measure $\rho_{iT/N}$. The reverse time dynamics described by equation (2.5) can then be interpreted as the schemes which anneal through the infinite ($N \rightarrow \infty, [1, T/N, \dots, T] \rightarrow [0, T]$) sequences as follows:

$$d\hat{X}_t = \underbrace{\left(\mathcal{A}\hat{X}_t + \frac{R}{2} \nabla \log(\rho_{T-t}(\hat{X}_t)) dt \right)}_{\text{Equation (2.13). Transports } \rho^{T-t} \text{ into } \rho^{T-dt-t}} + \underbrace{\left(\frac{R}{2} \nabla \log(\rho_{T-t}(\hat{X}_t)) dt + d\hat{W}_t \right)}_{\text{Equation (3.3) with } i/N=T-t. \text{ Preserves } \rho^{T-t}}$$

Furthermore, the approach in [30] can also be directly linked with the denoising approach of [24,25,27] thanks to the result in equation (2.8), which allows us to appreciate the connection among all the methods and a *denoising* interpretation of the generative dynamics.

(b) Flow matching

It is worth mentioning that, in the literature, a purely deterministic approach to the problem (a generalization of the ODE equation (2.13)), built as a generalization of the finite-dimensional *Flow Matching* approach of [46], has been explored in [29]. In this work, the authors avoid the issues associated with time reversal by directly constructing a path of probability measures $(\hat{\rho}_t)_{t \in [0, T]}$ which

anneals from a tractable initial measure $\hat{\rho}_0$ (a Gaussian measure for example) to the target data distribution $\hat{\rho}_T = \mu$. This path is constructed by flowing the initial measure along a time-dependent vector field $v(x, t)$ on the Hilbert space H :

$$d\tilde{X}_t = v(X_t, t)dt, \quad \tilde{X}_0 \sim \hat{\rho}_0 = \mathcal{N}_{0,C}, \quad (3.4)$$

with C being a given trace class covariance. Since it represents a deterministic evolution, equation (3.4) can equivalently be described by a flow mapping $\tilde{X}_t = \phi(\tilde{X}_0, t)$, which is usually the space in which the design of such models is performed. The evolution of $\hat{\rho}_t$ along $v(x, t)$ is governed by the continuity equation (again for u in the set of cylinder test functions)

$$\int_0^T dt \int_H (\partial_t u(x, t) + \langle v(x, t)(x), \nabla_x u(x, t) \rangle) d\hat{\rho}_t(x) = 0. \quad (3.5)$$

Notice that the difference between this methodology and the ones described previously is that the transformation of a Gaussian reference measure into the data measure takes place exactly in *finite time*, whereas for the other methodologies, this does not happen. Indeed, in equation (2.4), the initial conditions are drawn from ρ_{T_i} , which can be close but not equal to a purely Gaussian distribution (unless μ is the steady-state distribution itself of the process, provided it exists, or $T = \infty$). Similarly, the annealed Langevin approach requires running the last step of the chains for an infinite amount of time.

Given a conditional vector field $v(x, t)$, it is possible to compute the induced sequence of measures $\hat{\rho}_t$. The inverse procedure, which is more important from a practical perspective (the annealed sequence is a design parameter and the field is unknown), is much more challenging. *A priori*, it is not even known whether, given a sequence of measures, a vector field which matches this sequence exists. In [29], the strategy for the construction of a sequence for which the field exists is split into steps: (i) selection of a family of vector fields v_t^z which induce a flow of measures $\hat{\rho}_t^z$ whose starting point is $\hat{\rho}_0^z = \hat{\rho}_0$ (thus independent from z) and ending one (at time T) is a measure concentrated around z , (ii) assumption that $\hat{\rho}_t^z \ll \hat{\rho}_t$ for μ a.e. z and almost every $t \in [0, T]$, and (iii) construction of $v(x, t) = \int v_t^z \frac{d\hat{\rho}_t^z}{d\hat{\rho}_t} d\mu(z)$. The biggest technical problem is to ensure that $\hat{\rho}_t^z \ll \hat{\rho}_t$, which is proven to be true if the collection of parametrized measures are μ a.e. mutually absolutely continuous. This holds for a class of conditional flows (induced by the conditional fields) of the form $\phi^z(\tilde{x}_0, t) = \sigma(t)x_0 + m(t)z$, where m, σ are selected appropriately ($\sigma(0) = m(T) = 1, \sigma(1) = m(0) = 0$), which corresponds to conditional vector fields

$$v^z(x, t) = \frac{d \log \sigma(t)}{dt} (x - m(t)z) + \frac{dm(t)}{dt} z. \quad (3.6)$$

In this practical case, the sequence of measures $d\hat{\rho}_t^z(g)$ admits known Gaussian closed form.

The relationship between the technique described in [29] and the other ones presented previously can be understood assuming that $\hat{\rho}_t = \rho_{T-t}$, with $\mathcal{A} = -I/2$. Then $\hat{\rho}_t^z(x) = \mathcal{N}_{\exp(-(T-t)/2)z, (1-\exp(-(T-t)))R}(x)$. Notice that, in this comparison, the flow starts from a measure ρ_T which is only *close* to a purely Gaussian measure. Then, we can explicitly identify the vector field $v(x, t)$ from equation (2.13), after simple manipulation, as $v(x, t) = -\mathcal{A}x + \frac{1}{2}R\nabla \log \rho_{T-t}(x)$, thanks to the content of the second form of the Fokker–Planck equation, as shown in equation (2.12).

This correspondence is useful in that the same conditions of theorem 2.2, which allow us to claim validity for the other class of models, are sufficient to ensure that the models considered in [29] are well defined. Indeed, other than integrability assumptions, the main assumption in Theorem 1 of [29] is that $\hat{\rho}_t^z \ll \hat{\rho}_t$ for μ a.e. z and almost every t . Under the same assumptions of theorem 2.2, this holds.

Theorem 3.1. *As in theorem 2.2, assume that $\mu(R^{\frac{1}{2}}H) = 1$ and condition 2.2 holds. Then, $\hat{\rho}_t^z \ll \hat{\rho}_t$ for μ a.e. z and almost every t .*

Proof. We consider point (ii) of theorem 2.2, $\rho_t \ll \rho_t^0$, which reads $\rho_t^0(B) = 0 \Rightarrow \rho_t(B) = 0$ for any $t \in (0, T]$. Since $\rho_t(B) = \int \rho_t^z(B) d\mu(z)$, we have $\rho_t(B) = 0 \Rightarrow \rho_t^z(B) = 0$, for μ a.e. z . Chaining the results, this implies that $\rho_t^z \ll \rho_t^0$, for μ a.e. z . Since both measures are Gaussian, this implies their equivalence, $\rho_t^z \sim \rho_t^0$ for μ -a.e. z . Then, for any z, z' , $\rho_t^z \sim \rho_t^{z'}$, which is enough to prove that $\rho_t^z \ll \rho_t$ for μ a.e. z and almost every t , since $(0, T]$ has full measure ([29], Theorem 2). ■

4. Parametric approximations

Our discussion so far has been rooted in the underlying assumption of having access to the key vector fields of the dynamics for the generation process: $\nabla \log \rho_t(x)$, $\nabla \log \frac{d\rho_t/N}{d\nu_t/N}(x)$ and $v(x, t)$. Clearly, this is not the case in practice, for realistic (and unknown) μ . In all implementations, the true vector field is replaced by a *parametric* ($\theta \in \mathbb{R}^p$) approximation $s_\theta(x, t)$. Importantly, working in the functional domain calls for architectural choices which are suited for the infinite-dimensional domain. The most popular choice for *resolution invariant* architectures is the family of Neural (Fourier) Operators [67,68]. These architectures are extremely powerful in representational power but suffer from the drawback of requiring the input points to be collected on a regularly spaced grid [68], limiting their usage. Alternatively, transformers architectures [69] have been considered, by interpreting them as mappings between Hilbert spaces [70]. The flexibility of being capable of handling irregularly spaced grids comes at the cost that resolution invariance has to be *learned* during training, and it is not guaranteed by default for Neural Operators. Finally, it is worth mentioning the class of Implicit Neural Representations [71], which have the requested resolution-invariant properties, are known to be naturally good denoisers [72], but require meta-learning techniques for working properly [73]. Another possibility, not yet explored by the community, is to combine the different architectures mentioned, as done in the empirical work [74].

(a) Loss functions and time discretization

While the selection of an appropriate architecture is an important endeavour, a fundamental aspect is the selection of appropriate loss functions for learning the parameters of such architectures. We then proceed by providing an overview of various loss functions employed in different approaches to infinite-dimensional diffusion models, highlighting their mathematical formulations and connections. Interestingly, all variants presented in the literature can be understood under the lens of learning to *denoise* a corrupted version of the input data.

The learning objective associated with equation (2.4) can be formulated as an evidence lower bound (ELBO) on the log-likelihood of the data (see [26] for full details). In particular, such ELBO arises from comparing the path measures of the forward and reverse diffusion processes, leveraging Girsanov's theorem in infinite dimensions [37]. The discrepancy between the measure χ obtained with the approximated dynamics (and with initial conditions drawn from ρ_T^0 in place of ρ_T) and the true one can be expressed in Kullback–Leibler (KL) terms as

$$\text{KL}(\mu || \chi) \leq \frac{1}{2} \mathbb{E}_{\mathbb{Q}} \left[\int_0^T \|R(s_\theta(X_t, t) - \nabla \log \rho_t(X_t | X_0))\|_{R^{1/2}H}^2 dt \right] + \text{KL}(\rho_T || \rho_T^0) + \text{const}(\theta), \quad (4.1)$$

where $\langle \nabla \log \rho_t(x | x_0), e^k \rangle = \frac{\partial}{\partial x^k} \log \frac{d\rho_t(x^k | x_0)}{dx^k}$. In the linear case, the conditional score term has expression [26]

$$R \nabla \log \rho_t(x | x_0) = -R\mathcal{S}(t)^{-1} (x - \exp(t\mathcal{A})x_0). \quad (4.2)$$

Minimizing the loss $\mathcal{L}(\theta) = \mathbb{E}_{\mathbb{Q}} \left[\int_0^T \|R s_\theta(X_t, t) + R\mathcal{S}(t)^{-1} (X_t - \exp(t\mathcal{A})X_0)\|_{R^{1/2}H}^2 dt \right]$ is then equivalent to minimizing the ELBO. Related losses, but with different preconditioning, are

also considered in [24,31,45]. A typical parametrization chosen for $s_\theta(x, t)$ is in the form $S(t)^{-1}(X_t - \exp(t\mathcal{A})n_\theta(x, t))$, from which

$$C(\theta) = \mathbb{E}_Q \left[\int_0^T \left\| RS^{-1}(t) \exp(t\mathcal{A})(s_\theta(X_t, t) - X_0) \right\|_{\frac{1}{R^2 H}}^2 dt \right]$$

and is thus evident that the optimal solution for the parametric network $n_\theta(x, t)$ is the ‘denoiser’ $\mathbb{E}_Q[X_0 | X_t = x]$.

While the continuous-time framework provides a strong theoretical foundation for understanding diffusion processes, practical implementations often rely only on discrete-time approximations. In this setting [28], the first work in chronological order in considering Hilbert space valued generative diffusion models, the continuous path of measures $(\rho_t)_{t \in [0, T]}$ is replaced by the finite sequence of measures, and the forward and reverse diffusion processes are approximated using discrete-time transitions, from which an analogous ELBO to equation (4.1) is obtained (where effectively, integration over t is substituted with a finite sum over the discrete steps of the generation chain). One important aspect of this approach is that it automatically includes in the lower bound the effect of discretization of the SDEs, which is instead not included explicitly in the continuous time methods (when implemented, such methods clearly need to include some form of numerical integration [75–81]). Whenever the number of numerical integration steps is limited, in the finite-dimensional literature of diffusion models, it has been observed that optimizing directly the discrete-time lower bound provides better results [82]. While the same should hold in principle also in the infinite-dimensional settings, to the best of our knowledge, none of the work present in the literature has directly explored the problem.

Other families of approaches [29,30] adopt similar techniques, where a conditional version of the target field is considered for learning the vector fields. In particular, for learning $\nabla \log \frac{d\rho_t/N}{d\nu_t/N}(x)$ through the equality formalized by equation (2.8), the following loss is considered:

$C(\theta) = \sum_{i=1}^N \mathbb{E}_Q \left[\left\| \exp(iT/N\mathcal{A})X_0 - s_\theta(X_{iT/N}, iT/N) \right\|^2 \right]$. Each term in the loss corresponds to the mismatch between the true Langevin drift and the approximated one. Interestingly, in principle, the scheme in [30] could still achieve very good performance while having bad approximations for all the terms but the last one, provided the last chain is run for long enough. A quantitative analysis of such discrepancy is an interesting avenue for future work, but it is worth mentioning that the study of the convergence rate of Langevin dynamics in infinite dimensions is much more challenging than for the finite-dimensional case [19]. Finally, in [29], the considered loss is $C(\theta) = \mathbb{E}_Q \left[\int_0^T \|s_\theta(X_t, T-t) - v^{x_0}(X_t, T-t)\|^2 \right]$ which, upon noticing that $v^{x_0}(X_t, T-t) = \frac{d \log \sigma(t)}{dt}(X_t - m(t)X_0) + \frac{dm(t)}{dt}X_0$ (equation (3.6)), can again be interpreted as a denoising objective.

5. Applications and extensions

One of the primary applications of diffusion models in function spaces is data generation at arbitrary resolutions. This capability is essential in domains where high-resolution or continuous data representations are required. For example, in the fields of image and audio synthesis, where traditional generative models operate in discrete, finite-dimensional spaces, diffusion models in function spaces can inherently capture and generate data that varies continuously across scales. Recent works have demonstrated that these models can be particularly powerful for generating high-resolution images or audio waveforms, with Neural Operators, INRs or Transformer architectures providing a way to model functional data while maintaining resolution invariance [26,27,31,50,83]. Other authors have mainly focused on time series modelling under the functional formalism [28,29], or focused on new methodologies for physical sciences simulations and inverse problems [30]. In general, function space diffusion models can be applied to any modality on which finite-dimensional techniques have been applied, and beyond. We comment that, other than the obvious application scenario of generative modelling, variants of these diffusion models have also been applied for other problems, which we overview in the following.

(a) Diffusion bridges and optimal control

For example, the works presented in [32,33,50] introduce an approach for simulating nonlinear diffusion bridges in infinite-dimensional spaces. The method involves considering a diffusion process (with measure \mathbb{Q})

$$dX_t = (\mathcal{A}X_t + f(X_t))dt + dW_t, X_0 = x_0, \quad (5.1)$$

and proving, through a generalization of Doob's h-transform [84] to infinite dimensions, that the representation of an SDE corresponding to a process with path measure \mathbb{Q}^x , where $\mathbb{Q}^x(\cdot) = \int \mathbb{Q}(\cdot | X_T = y) d\chi(y)$, $\chi \ll \rho_T$, is again an SDE in the form

$$dX_t = (\mathcal{A}X_t + f(X_t) + R\nabla \log h(X_t, t))dt + dW_t, X_0 = x_0, \quad (5.2)$$

where $h(x, t) = \mathbb{E}_{\gamma \sim \chi} \mathbb{E}_{\mathbb{Q}}[1_{d\gamma}(X_T) | X_t = x] = \int d\chi(z) d\mathbb{Q}(X_T = z | X_t = x)$. Unfortunately, excluding simple cases, knowledge of $\nabla \log h(x, t)$ is out of reach. Even worse, *a priori*, it is not even known whether $h(x, t)$ is differentiable and the problem is consequently well defined.

In the linear case, under the assumption $\exp(t\mathcal{A})z \in \mathcal{S}_t^{\frac{1}{2}}H$, the derivative exists and furthermore $\rho_t^x \ll \rho_\infty$, with known differentiable density $q_t(z, x)$ (for the exact form and a proof refer to [59,85,86]).¹ Then, noticing the time homogeneity of the problem $h(x, t) = \int d\chi(z) d\mathbb{Q}(X_{T-t} = z | X_0 = x) = \int d\chi(z) d\rho_{T-t}^x(z) = \int d\chi(z) q_{T-t}(x, z) d\rho_\infty(z)$, that is a form allowing for the computation of $\nabla \log h(x, t)$, which coincides with $\nabla \log \rho_{T-t}(x)$ [50]. In [50], such formalism is investigated in deeper detail and adopted for building bridges between broader modifications of \mathbb{Q} , for example for the measure $\int d\chi(y) d\zeta(x) \mathbb{Q}(\cdot | X_0 = x, X_T = y)$. These extensions allow us, for example, to perform generic bridge matching and implement Bayesian learning in function space [50]. Other than for generative modelling purposes or stochastic control, stochastic bridges are also useful in other contexts, such as phylogenetic shape analysis, which models the stochastic change in animals' morphometry over time [87].

When considering nonlinear SDEs, the situation becomes more complex as, in general, it is harder to prove the existence of the derivative of $h(x, t)$, and it is impossible to sample from $\mathbb{Q}(X_T = \cdot | X_t = x)$ without having first constructed the bridge, which introduces a circular dependency. The authors in [32,33] adopt the strategy of time reversal to solve such problems and simplify the implementation.

For the following passages to be valid, it is important to *assume* the existence of the various Radon–Nikodym derivatives involved, whose proof is in general a far from trivial task. We consider for simplicity of exposition deterministic terminal conditions, i.e. $\mathbb{Q}^x(\cdot) = \mathbb{Q}(\cdot | X_T = y)$, and consequently $h(x, t) = d\mathbb{Q}(X_t = x | X_T = y)$. By writing $d\mathbb{Q}(X_t = x | X_T = y) = \frac{d\mathbb{Q}(X_t = x | X_T = y)}{d\mathbb{Q}(X_t = x)} d\mathbb{Q}(X_T = y)$,

we can explicitly write $\nabla \log h(x, t) = \nabla \log \frac{d\mathbb{Q}(X_t = x | X_T = y)}{d\mathbb{Q}(X_t = x)}$, and in particular then $\langle R\nabla \log h(x, t), e^k \rangle =$

$$r^k \frac{\partial}{\partial x^k} \left(\log \left(\frac{d\mathbb{Q}(X_t^k = x^k | X_T = y, X_t^{j \neq k} = x^{j \neq k})}{d\mathbb{Q}(X_t^k = x^k | X_t^{j \neq k} = x^{j \neq k})} \right) \right).$$

The time reversal of equation (5.2) involves the distribution ρ_t^x associated to \mathbb{Q}^x , and consequently $\langle R\nabla \log \rho_t^x(x), e^k \rangle = r^k \frac{\partial}{\partial x^k} \left(\log \left(\frac{d\rho_t(X_t^k = x^k | X_t^{j \neq k} = x^{j \neq k}, X_T = y)}{dx^k} \right) \right)$. Then

$-R\nabla \log h(x, t) + R\nabla \log \rho_t^x(x) = R\nabla \log \rho_t(x)$. Conceptually, this implies the following: to simulate a bridge in the form of equation (5.2), it is possible to learn $R\nabla \log \rho_t(x)$ from equation (5.1), which has no constraints on the ending value and is thus an easier task, and then simulate the backward dynamics of equation (5.2), where initial conditions are known and the drift term is equal to $-\mathcal{A}x - f(x, T - t) + R\nabla \log \rho_t(x)$. For full details and formal validity of the derivations, we refer the reader to [32,33].

¹Notice that this condition is in general not compatible with the ones guaranteeing the existence of a time reversal as defined by [57].

(b) Bayesian inverse problems

Finite-dimensional diffusion models have been explored as strong priors for Bayesian inverse problems in multiple domains [51,52]. The extension of these techniques to the Hilbert space settings has been explored by the authors in [24,25], in the case of linear and nonlinear observations, respectively. Considering the linear observation problem, it is assumed to have access to $Y = P(X_0) + B$, where $P : H \rightarrow \mathbb{R}^M$ is a linear operator and B is an additive noise random variable with a density ψ w.r.t. the Lebesgue measure. Then, if $\mu \ll \mathcal{N}(0, R)$ (note that the work in [24] considers a more generic \tilde{R}) it is possible to show that the following scheme

$$d\hat{X}_t = \frac{1}{2}\hat{X}_t - (1 - \exp(-t))^{-1} \left(x - \exp\left(-\frac{t}{2}\right) \mathbb{E}[X_0 | X_{T-t} = \hat{X}_t, Y] \right) + d\hat{W}_t, \hat{X}_0 \sim \mathbb{Q}(X_T | Y) \quad (5.3)$$

with \hat{W}_t being a R -Wiener process provides, at time $t = T$, a valid sample from the posterior distribution $\mathbb{Q}(X_0 | Y)$. The result is obtained considering the particular case of equation (2.1) where $f = 0$ and $\mathcal{A} = -\frac{1}{2}I$. The proof requires mimicking the result of [45], extending it for the conditional case. The authors extend their work in two directions: first, considering the case of nonlinear operator P , and second, performing conditional sampling with only knowledge of the unconditional $\mathbb{E}[X_0 | X_T]$ (thus allowing the technique to adopt any pre-trained unconditional model as in [88]). A simplified expression of the scheme proposed in [25] reads

$$d\hat{X}_t = \frac{R}{2}\nabla \log \rho_\tau(\hat{X}_t) + \frac{R}{2}\nabla \log(\psi(Y - P(\hat{X}_t))) + d\hat{W}_t, \hat{X}_0 \sim \mathbb{Q}(X_T | Y), \quad (5.4)$$

where $\tau > 0$ is a parameter which is annealed from large to small values. Given the similarity between equations (5.4) and (3.3), it is natural to interpret the former as an infinite-dimensional generalization of the conditioning trick [89], which in finite dimension reads $\nabla \rho_\tau(x | y) = \nabla \rho_\tau(x) + \nabla \mathbb{Q}(y | x)$. We invite the reader to refer to [25] for generalizations of this scheme and quantitative bounds on the quality of approximated posterior.

6. Conclusions

In this survey, we have explored the extension of diffusion-based generative models from finite-dimensional settings to infinite-dimensional function spaces, focusing on the theoretical underpinnings and practical implementations of such models in Hilbert spaces. By leveraging SDEs and the properties of time reversal in infinite dimensions, we have highlighted how these models can effectively generate function-valued data, offering promising applications in domains that require high-resolution generative capabilities.

While significant progress has been made in adapting diffusion models to infinite-dimensional settings, several challenges remain open. One of the key difficulties is ensuring the efficient computation of generative dynamics in function spaces without sacrificing resolution invariance. Approximations and parametric methods, such as those based on neural Fourier operators and transformer architectures, have shown promise, but further improvements are necessary to fully realize the potential of these models in practical settings.

An interesting direction for future research would be the exploration of perturbation-based scenarios for infinite-dimensional diffusion models, akin to those already developed for finite-dimensional cases. For instance, works such as [90–93] have successfully extended finite-dimensional diffusion models to account for different perturbative settings. Applying these methods to function spaces could open new avenues for developing more robust generative models that operate under complex data perturbations or in noisier, real-world applications. This extension could provide stronger priors for solving inverse problems, denoising functional data or addressing generative tasks in physical sciences. Alternatively, it would be interesting to draw the functional domain equivalent of the alternatives to diffusion in finite dimensions [47,94,95].

Moreover, the theoretical understanding of how these perturbation techniques interact with the infinite-dimensional structure remains an open problem. As with finite-dimensional models,

establishing precise conditions for the stability, convergence and efficiency of these models in infinite dimensions will be critical for both theoretical development and their practical utility.

In conclusion, infinite-dimensional diffusion models represent a powerful and flexible approach to generative modelling in functional domains. Their potential in practical applications is clear but under-explored, and further extensions, especially along the lines of perturbation methods, could significantly advance both the theory and applications of functional models.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. This article has no additional data.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. G.F.: conceptualization, funding acquisition, writing—original draft; P.M.: conceptualization, funding acquisition, writing—original draft.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. The authors were partially funded by Huawei Paris under the Netception Project.

Appendix A. Detailed conditions for theorem 2.1

We provide a complete list of assumptions from [65], where the authors consider a generalization of the infinite system of equations

$$dY_t^k = b^k(Y_t, t)dt + \sigma^k(t)dW_t^k, \quad k \in \mathbb{N}, \quad (\text{A } 1)$$

where W_t^k are Brownian motions, with $\mathbb{E}_{\mathbb{Q}}[W_t^k W_s^j] = \delta(k-j)r^k \min(s, t)$, δ in the Kronecker sense. For clarity, we denote the law of the process by $d\rho_t^Y(y)$ when necessary. The system studied in [65] evolves in the space of weighted square-summable sequences,

$$l_2^r(\mathbb{N}) \stackrel{\text{def}}{=} \{y \in \mathbb{R}^{\mathbb{N}} : \sum r^i (y^i)^2 < \infty\},$$

while the original process X is defined in $l_2(\mathbb{N})$. Since these two spaces are isomorphic, results from one setting can be translated into the other via simple algebraic manipulations. We now list the necessary assumptions for ensuring the existence of the reverse process. We state the assumptions in a simplified form, which leverages the fact that in this work we consider σ to be independent from y .

Assumption A.1. (*Assumption H1 in [65]*). There exists a finite positive constant K such that $\forall y, z \in l_2^r(\mathbb{N})$:

$$\begin{aligned} \sup_t \left\{ \sum_i r^i (b^i(y, t))^2 + \sum_i (r^i)^2 (\sigma^i(t))^2 \right\} &\leq K \left(1 + \sum_i r^i (y^i)^2 \right) \\ \sup_t \left\{ \sum_i r^i (b^i(y, t) - b^i(z, t))^2 \right\} &\leq K \sum_i r^i (y^i - z^i)^2. \end{aligned}$$

Assumption A.2. (*Assumption H2 of [65]*). For each $k \in \mathbb{N}$ and $t \in [0, T]$, the coefficients $b^k(y, t)$ depend on y on at most a number of finite coordinates.

Assumption A.3. (*Assumption H4 of [65]*). There exists a positive constant K such that $\forall y, z \in l_2^r(\mathbb{N})$

$$\sup_t \left\{ \sum_i (r^i)^2 (b^i(y, t) - b^i(z, t))^2 \right\} \leq K \sum_i (r^i)^2 (y^i - z^i)^2.$$

Assumption A.4. (Assumption H5 of [65]). There exists an increasing sequence of finite subsets $I(i), i \geq 0$ which cover \mathbb{N} (i.e. $\cup_i I(i) = \mathbb{N}$) such that $\forall i \in \mathbb{N}, C > 0$, there exists a constant $K(i, C)$ for which

$$\sup_t \sup_{k \in I(i)} \sup\{|b^k(y, t)| + \sum_{j \in \mathbb{N}} r^k |\sigma^{jk}(t)| : \sup_{m \in I(i)} |y^m| \leq C\} \leq K(i, C).$$

Assumption A.5. (parts (iii,iv) of Theorem 4.3. in [65]). Suppose that the initial condition $Y_0 \in l_2^*(\mathbb{N})$.

- Assume that the conditional law of y^i given $y^j, j \neq i$ has density with respect to Lebesgue measure on \mathbb{R} , i.e. the conditional densities $\frac{d\rho_i^Y(y^i | y^{j \neq i})}{dy^i}$ exist.
- Fix a finite subset $J \subset \mathbb{N}$. For every positive t_0 , every $D \in \{(\prod_{i \in J} K_i) \times (\prod_{j \notin J} \mathbb{R}), K_j \text{ compact in } \mathbb{R}\} \cap l_2^*(\mathbb{N})$, and every $i \in \mathbb{N}$ assume that

$$\int_{t_0}^1 \int_D |r^i(\sigma^i(t))^2 \frac{\partial}{\partial y^i} (\frac{d\rho_i^Y(y^i | y^{j \neq i})}{dy^i})| dy^i \rho_i^Y(dy^{j \neq i}) dt < \infty.$$

Given the assumptions outlined above, we can now formally state the main result of Theorem 4.3 in [65]:

Theorem A.1. Consider the system in equation (A 1). Suppose assumptions A.1, A.2, A.3, A.4, and A.5 hold. Then the time reversal of equation (A 1) is a solution of the SDE taking values in $l_2^*(\mathbb{N})$

$$d\hat{Y}_t^k = -b^k(\hat{Y}_t, T - t) + r^k(\sigma^k(T - t))^2 \frac{\partial}{\partial y^k} \log \left(\frac{d\rho_{T-t}^Y(\hat{Y}_t^k | \hat{Y}_t^{j \neq k})}{dy^k} \right) dt + \sigma^k(T - t) d\hat{W}_t^k. \tag{A 2}$$

As a direct corollary, we consider the system of equations in $l_2(\mathbb{N})$,

$$dX_t^k = a^k(X_t) + f^k(X_t, t)dt + dW_t^k, \quad k \in \mathbb{N},$$

which admits strong solutions. We introduce the coordinate transformation $y^k = (r^k)^{-\frac{1}{2}} x^k$, yielding the equivalent system

$$dY_t^k = \underbrace{(r^k)^{-\frac{1}{2}} \left(a^k(R^{\frac{1}{2}} Y_t) + f^k(R^{\frac{1}{2}} Y_t, t) \right)}_{b^k(Y_t, t)} dt + \underbrace{(r^k)^{-\frac{1}{2}}}_{\sigma^k(t)} dW_t^k, \quad k \in \mathbb{N}, \tag{A 3}$$

which evolves in $l_2^*(\mathbb{N})$ since

$$\sum_i (x^i)^2 = \sum_i r^i \left((r^i)^{-\frac{1}{2}} x^i \right)^2 = \sum_i r^i (y^i)^2.$$

For this transformed system, we verify that assumptions A.1, A.2, A.3, A.4 and A.5 hold. If these conditions are satisfied, the time-reversed process of equation (A 3) satisfies equation (A 2). Since the transformation satisfies $d\hat{X}_t^k = (r^k)^{\frac{1}{2}} d\hat{Y}_t^k$, the reversed system takes the form

$$d\hat{X}_t^k = -\left(a^k(\hat{X}_t) + f^k(\hat{X}_t, T - t) \right) + (r^k)^{\frac{1}{2}} \left(\frac{\partial}{\partial y^k} \log \left(\frac{d\rho_{T-t}^Y(y^k | y^{j \neq k})}{dy^k} \right) \right) \Big|_{y=R^{-\frac{1}{2}} \hat{X}_t} dt + d\hat{W}_t^k. \tag{A 4}$$

Applying the change of variable formula for measures, we obtain $\frac{d\rho_{T-t}^Y(y^k | y^{j \neq k})}{dy^k} = (r^k)^{\frac{1}{2}} \frac{d\rho_{T-t}^X(x^k | x^{j \neq k})}{dx^k}$, which leads directly to the form given in equation (2.4).

A.1. The linear case

We now show that condition 2.3 ensures the validity of assumptions A.1, A.2, A.3 and A.4. Under this condition, the forward system satisfies

$$dX_t^k = a^k X_t^k dt + dW_t^k, \quad k \in \mathbb{N}, \quad (\text{A } 5)$$

which admits a strong solution in $L_2(\mathbb{N})$ (see condition 2.1). Transforming via $y^i = (r^i)^{-\frac{1}{2}} x^i$, we obtain

$$dY_t^k = a^k Y_t^k dt + (r^k)^{-\frac{1}{2}} dW_t^k, \quad k \in \mathbb{N}. \quad (\text{A } 6)$$

Under this transformation, assumption A.1 reduces to

$$\begin{aligned} \forall y \in L_2(\mathbb{N}), \quad \sum_i r^i (a^i)^2 (y^i)^2 + \sum_i r^i &\leq K \left(1 + \sum_i r^i (y^i)^2 \right), \\ \forall y, z \in L_2(\mathbb{N}), \quad \sum_i r^i (a^i)^2 (y^i - z^i)^2 &\leq K \sum_i r^i (y^i - z^i)^2. \end{aligned}$$

Since b and σ are independent of t , these assumptions are satisfied, given that $\sum_i r^i = K_1$ and

$$\sum_i r^i (a^i)^2 (y^i)^2 \leq M^2 \sum_i r^i (y^i)^2, \quad \sum_i r^i (a^i)^2 (y^i - z^i)^2 \leq M^2 \sum_i r^i (y^i - z^i)^2,$$

where M is the constant from condition 2.3. Assumption A.3 is verified similarly, while A.2 and A.4 follow from the diagonality of \mathcal{A} . In particular for A.4, consider $I(i) = \{i\}$, from which the assumption is satisfied if $\sup\{|a^i y^i| + (r^i)^{\frac{1}{2}} : |y^i| \leq C\} \leq K(i, C), \forall i \in \mathbb{N}$, which clearly holds since a^i is bounded.

Using again the relation $\frac{d\rho_t^Y(y^k | y^{j \neq k})}{dy^k} = (r^k)^{\frac{1}{2}} \frac{d\rho_t^X(x^k | x^{j \neq k})}{dx^k}$, along with the results in theorem 2.2(iii), we establish that the integrability condition in assumption A.5 holds, since the time interval considered is $(0, 1]$.

References

1. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. 2021 Score-based generative modeling through stochastic differential equations. In *Int. Conf. on Learning Representations*, Vienna, Austria, 4 May 2021. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75296.html>.
2. Kong Z, Ping W, Huang J, Zhao K, Catanzaro B. 2021 Diffwave: a versatile diffusion model for audio synthesis. In *Int. Conf. on Learning Representations*, Vienna, Austria, 4 May 2021. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75296.html>.
3. Liu J, Li C, Ren Y, Chen F, Zhao Z. 2022 Diffsinger: singing voice synthesis via shallow diffusion mechanism. In *Proc. of the AAAI Conf. on Artificial Intelligence*, Online, 22 February – 1 March 2022 vol. 36, pp. 11020–11028. Red Hook, NY: Curran Associates. (doi:10.1609/aaai.v36i10.21350)
4. He Y, Yang T, Zhang Y, Shan Y, Chen Q. 2022 Latent video diffusion models for high-fidelity video generation with arbitrary lengths. *arXiv* (doi:10.48550/arXiv.2302.10130)
5. Ho J *et al.* 2022 Imagen video: high definition video generation with diffusion models. See <https://imagen.research.google/video/paper.pdf>
6. Hoogeboom E, Satorras VG, Vignac C, Welling M. 2022 Equivariant diffusion for molecule generation in 3D. In *Int. Conf. on Machine Learning*, Baltimore, MD, 17–23 July 2022. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/68317.html>.
7. Luo S, Hu W. Diffusion probabilistic models for 3d point cloud generation. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. Online, 19–25 June 2021, pp. 2837–2845. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/60773.html>.

8. Trippe BL, Yim J, Tischer D, Baker D, Broderick T, Barzilay R, Jaakkola T. 2022 Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem *International Conference on Learning Representations (ICLR)*, Kigali, Rwanda, 1-5 May 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75096.html>
9. Zeng X, Vahdat A, Williams F, Gojic Z, Litany O, Fidler S, Kreis K. 2022 Lion: Latent point diffusion models for 3d shape generation. In *Advances in neural information processing systems (neurIPS)*, New Orleans, LA, 28 November - 9 December 2022. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/68431.html>.
10. Esser P, Kulal S, Blattmann A, Entezari R, Müller J, Saini H, Levi Y. 2024 Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first Int. Conf. on Machine Learning*, Vienna, Austria, 21-27 July 2024. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/76048.html>.
11. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. 2022 Hierarchical text-conditional image generation with CLIP latents. *arXiv*. See <https://arxiv.org/abs/2204.06125>
12. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, 18-24 June 2022, pp. 10684–10695. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/65666.html>
13. Saharia C *et al.* 2022 Photorealistic text-to-image diffusion models with deep language understanding. In *Advances in neural information processing systems* (eds AH Oh, A Agarwal, D Belgrave, K Cho), New Orleans, LA, 28 November - 9 December 2022. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/68431.html>.
14. Pathak J, Subramanian S, Harrington P, Raja S, Chattopadhyay A, Mardani M, Kurth T. 2022 Fourcastnet: a global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv*. (doi:10.48550/arXiv.2202.11214)
15. Yang Y, Gao AF, Castellanos JC, Ross ZE, Azizzadenesheli K, Clayton RW. 2021 Seismic wave propagation and inversion with neural operators. *Seism. Rec.* **1**, 126–134. (doi:10.1785/0320210026)
16. Park S, Kim D, Alaa A. 2024 Mean-field chaos diffusion models. *arXiv* (doi:10.48550/arXiv.2406.05396)
17. Stuart AM. 2010 Inverse problems: a Bayesian perspective. *Acta Numer.* **19**, 451–559. (doi:10.1017/s0962492910000061)
18. Cotter SL, Roberts GO, Stuart AM, White D. 2013 MCMC methods for functions: modifying old algorithms to make them faster. *Stat. Sci.* **28**, 424–446. (doi:10.1214/13-STS421)
19. Dashti M, Stuart AM. 2013 The Bayesian approach to inverse problems. *arXiv* (doi:10.48550/arXiv.1302.6989)
20. Chang P, Tang J, Gross M, C.Azevedo V. 2024 How i warped your noise: a temporally correlated noise prior for diffusion models. In *The Twelfth Int. Conf. on Learning Representations*, Vienna, Austria, 7-11 May 2024. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/74958.html>.
21. Chen T. 2023 On the importance of noise scheduling for diffusion models. *arXiv* (doi:10.48550/arXiv.2301.10972)
22. Ge S *et al.* 2023 Preserve your own correlation: a noise prior for video diffusion models. In *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, Vancouver, Canada, 18-22 June 2023, pp. 22930–22941. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/70184.html>
23. Teng J, Zheng W, Ding M, Hong W, Wangni J, Yang Z, Tang J. 2023 Relay diffusion: unifying diffusion process across resolutions for image synthesis. *arXiv* (doi:10.48550/arXiv.2309.03350)
24. Baldassari L, Siahkoohi A, Garnier J, Solna K, de Hoop MV. 2024 Conditional score-based diffusion models for Bayesian inference in infinite dimensions. In *Advances in neural information processing systems*, vol. 36, New Orleans, LA, 10-16 December 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75280.html>.
25. Baldassari L, Siahkoohi A, Garnier J, Solna K, de Hoop MV. 2024 Taming score-based diffusion priors for infinite-dimensional nonlinear inverse problems. *arXiv*. (doi:10.48550/arXiv.2405.15676)

26. Franzese G, Corallo G, Rossi S, Heinonen M, Filippone M, Michiardi P. 2024 Continuous-time functional diffusion processes. In *Advances in neural information processing systems*, vol. 36, New Orleans, LA, 10-16 December 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75280.html>.
27. Hagemann P, Mildenerger S, Ruthotto L, Steidl G, T.Yang N. 2023 Multilevel diffusion: infinite dimensional score-based diffusion models for image generation. *arXiv* (doi:10.48550/arXiv.2303.04772)
28. Kerrigan G, Ley J, Smyth P. 2023 Diffusion generative models in infinite dimensions. In *Int. Conf. on Artificial Intelligence and Statistics*, Valencia, Spain, 25-27 April 2023, pp. 9538–9563. PMLR. <https://proceedings.mlr.press/v206/>
29. Kerrigan G, Migliorini G, Smyth P. 2024 Functional flow matching. In *Int. Conf. on Artificial Intelligence and Statistics*, Valencia, Spain, 25-27 April 2023, pp. 3934–3942. PMLR. <https://proceedings.mlr.press/v238/>
30. Lim JH, Kovachki NB, Baptista R, Beckham C, Azizzadenesheli K, Kossaifi J, Voleti V. 2023 Score-based diffusion models in function space. *arXiv* (doi:10.48550/arXiv.2302.07400)
31. Lim S, Yoon EB, Byun T, Kang T, Kim S, Lee K, Choi S. 2024 Score-based generative modeling through stochastic evolution equations in hilbert spaces. In *Advances in neural information processing systems*, vol. 36, New Orleans, LA, 10-16 December 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75280.html>.
32. Baker EL, Yang G, Severinsen ML, Hipsley CA, Sommer S. 2024 Conditioning non-linear and infinite-dimensional diffusion processes. *arXiv* (doi:10.48550/arXiv.2402.01434)
33. Yang G, Baker EL, Severinsen ML, Hipsley CA, Sommer S. 2024 Simulating infinite-dimensional nonlinear diffusion bridges. *arXiv* (doi:10.48550/arXiv.2405.18353)
34. Dutordoir V, Saul A, Ghahramani Z, Simpson F. 2023 Neural diffusion processes. In *Int. Conf. on Machine Learning*, Honolulu, HI, 23-29 July 2023, pp. 8990–9012. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/73019.html>
35. Mathieu E, Dutordoir V, Hutchinson M, Bortoli V, W.Teh Y, Turner R. 2024 Geometric neural diffusion processes. In *Advances in neural information processing systems*, vol. 36, New Orleans, LA, 10-16 December 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75280.html>.
36. Zhou L, Du Y, Wu J. 2021 3d shape generation and completion through point-voxel diffusion. In *Proc. of the IEEE/CVF Int. Conf. on computer vision*, Online, 19-25 June 2021, pp. 5826-5835. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/60773.html>
37. Da Prato G, Zabczyk J. 2014 *Stochastic equations in infinite dimensions*. Cambridge, UK: Cambridge University Press.
38. Dupont E, Kim H, A.Eslami SM, J.Rezende D, Rosenbaum D. 2022 From data to functa: Your data point is a function and you can treat it like one. In *Int. Conf. on Machine Learning of Proc of Machine Learning Research*, vol. 162, Baltimore, MD, 17-23 July 2022, pp. 5694-5725. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/68317.html>
39. Phillips A, Seror T, Hutchinson M, De Bortoli V, Doucet A, Mathieu E. 2022 Spectral diffusion processes. *arXiv*. (doi:10.48550/arXiv.2209.14125)
40. Ho J, Jain A, Abbeel P. 2020 Denoising diffusion probabilistic models. In *Advances in neural information processing systems* (eds H Larochelle, M Ranzato, R Hadsell, M Balcan, H Lin), vol. 33, Online, 6-12 December 2020, pp. 6840-6851. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/59066.html>
41. Sohl-Dickstein J, Weiss E, Maheswaranathan N, Ganguli S. Deep unsupervised learning using nonequilibrium thermodynamics (eds F Bach, D Blei). In *Proc. of the 32nd Int. Conf. on Machine Learning of Proc. of Machine Learning Research*, vol. 37, Lille, France, 6-11 July 2015, pp. 2256–2265. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/27550.html>
42. Song Y, Ermon S. 2019 Generative modeling by estimating gradients of the data distribution. In *Advances in neural information processing systems* (eds H Wallach, H Larochelle, A Beygelzimer, F d' Alché-Buc, E Fox, R Garnett), vol. 32, Vancouver, Canada, 8-14 December 2019. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/53719.html>

43. Song Y, Ermon S. 2020 Improved techniques for training score-based generative models. In *Advances in neural information processing systems* (eds H Larochelle, M Ranzato, R Hadsell, M Balcan, H Lin), Online, 6–12 December 2020 pp. 12438–12448, vol. 33. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/59066.html>
44. Franzese G, Rossi S, Yang L, Finamore A, Rossi D, Filippone M, Michiardi P. 2022 How much is enough? A study on diffusion times in score-based generative models. *arXiv*. (doi:10.48550/arXiv.2206.05173)
45. Pidstrigach J, Marzouk Y, Reich S, Wang S. 2023 Infinite-dimensional diffusion models for function spaces. *arXiv*. (doi:10.48550/arXiv.2302.10130)
46. Lipman Y, T.Chen R, Ben-Hamu H, Nickel M, Le M. Flow matching for generative modeling. In *The Eleventh Int. Conf. on Learning Representations*, Kigali, Rwanda, 1–5 May 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75096.html>
47. Liu X, Gong C, Qiang liu L. 2023 Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh Int. Conf. on Learning Representations*, Kigali, Rwanda, 1–5 May 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75096.html>
48. Berner J, Richter L, Ullrich K. 2024 An optimal control perspective on diffusion-based generative modeling. *Trans. Mach. Lear. Res.* (doi:10.48550/arXiv.2211.01364)
49. Heng J, Bortoli V, Doucet A, Thornton J. 2021 Simulating diffusion bridges with score matching. *arXiv* (doi:10.48550/arXiv.2111.07243)
50. Park B, Choi J, Lim S, Lee J. 2024 Stochastic optimal control for diffusion bridges in function spaces. *arXiv* (doi:10.48550/arXiv.2405.20630)
51. Cardoso G, idrissi YJ, L.Corff S, Moulines E. 2024 Monte Carlo guided denoising diffusion models for Bayesian linear inverse problems. In *The Twelfth Int. Conf. on Learning Representations*, Vienna, Austria, 7–11 May 2024. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/74958.html>
52. Wu L, Trippe B, Naesseth C, Blei D, Cunningham JP. 2024 Practicaland asymptotically exact conditional sampling in diffusion models. In *Advances in neural information processing systems*, vol. 36, New Orleans, LA, 10–16 December 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75280.html>
53. Anderson BDO. 1982 Reverse-time diffusion equation models. *Stoch. Process. Their Appl.* **12**, 313–326. (doi:10.1016/0304-4149(82)90051-5)
54. Föllmer H. 1985 An entropy approach to the time reversal of diffusion processes. In *Stochastic differential systems filtering and control*, pp. 156–163. Berlin; Heidelberg: Springer.
55. Eldredge N. 2016 Analysis and probability on infinite-dimensional spaces. *arXiv* (doi:10.48550/arXiv.1607.03591)
56. Föllmer H, Wakolbinger A. 1986 Time reversal of infinite-dimensional diffusions. *Stoch. Process. Their Appl.* **22**, 59–77. (doi:10.1016/0304-4149(86)90114-6)
57. Millet A, Nualart D, Sanz M. 1989 Time reversal for infinite-dimensional diffusions. *Probab. Theory Relat. Fields* **82**, 315–347. (doi:10.1007/bf00339991)
58. Da Prato G, Zabczyk J. 1996 *Ergodicity for infinite dimensional systems*. vol. 229. Cambridge, UK: Cambridge University Press.
59. Da Prato G, Zabczyk J. 2002 *Second order partial differential equations in hilbert spaces*. Cambridge, UK: Cambridge University Press.
60. Lunardi A, Pallara D. 2020 Ornstein-Uhlenbeck semigroups in infinite dimension. *Phil. Trans. R. Soc. A* **378**, 20190620. (doi:10.1098/rsta.2019.0620)
61. Bogachev V, Da Prato G, Röckner M. 2009 Existence and uniqueness of solutions for Fokker–Planck equations on Hilbert spaces. *J. Evol. Equ.* **10**, 487–509. (doi:10.1007/s00028-010-0058-y)
62. Bogachev V, Prato GD, Röckner M. 2011 Uniqueness for solutions of Fokker–Planck equations on infinite dimensional spaces. *Commun. Partial Differ. Equ.* **36**, 925–939. (doi:10.1080/03605302.2010.539657)
63. Bogachev VI, Da Prato G, Roeckner M, Shaposhnikov SV. 2015 An analytic approach to infinite-dimensional continuity and Fokker–Planck–Kolmogorov equations. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* **14**, 983–1023. (doi:10.2422/2036-2145.201303_004)
64. Da Prato G, Flandoli F, Röckner M. 2014 Uniqueness for continuity equations in Hilbert spaces with weakly differentiable drift. *Stoch. Partial Differ. Equ.* **2**, 121–145. (doi:10.1007/s40072-014-0031-9)

65. Da Prato G, Flandoli F, Röckner M. 2019 Absolutely continuous solutions for continuity equations in Hilbert spaces. *J. De Mathématiques Pures Et Appliquées* **128**, 42–86. (doi:10.1016/j.matpur.2019.06.010)
66. Ottobre M, S.Pillai N, J.Pinski F, M.Stuart A. 2016 A function space HMC algorithm with second order Langevin diffusion limit. The Hague: Bernoulli Society for Mathematical Statistics and Probability <https://ouci.dntb.gov.ua/en/works/7nMmDWm9/>
67. Kovachki N, Li Z, Liu B, Azizzadenesheli K, Bhattacharya K, Stuart A, Anandkumar A. 2023 Neural operator: learning maps between function spaces with applications to PDEs. *J. Mach. Learn. Res.* **24**, 1–97. <https://www.jmlr.org/papers/v24/21-1524.html>
68. Li Z, Kovachki NB, Azizzadenesheli K, liu B, Bhattacharya K, Stuart A, Anandkumar A. 2021 Fourier neural operator for parametric partial differential equations. In *Int. Conf. on Learning Representations*, Vienna, Austria, 4 May 2021. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75296.html>
69. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017 Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, CA, 4–9 December 2017. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/39083.html>
70. Cao S. 2021 Choose a transformer: Fourier or Galerkin. *Adv. Neural Inf. Process. Syst.* **34**, 24924–24940. (doi:10.5555/3540261.3542170)
71. Sitzmann V, Martel J, Bergman A, Lindell D, Wetzstein G. 2020 Implicit neural representations with periodic activation functions. In *Advances in neural information processing systems* (eds H Larochelle, M Ranzato, R Hadsell, M Balcan, H Lin), Online, 6–12 December 2020, pp. 7462–7473. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/59066.html>
72. Kim C, Lee J, Shin J. 2022 Zero-shot blind image denoising via implicit neural representations. *arXiv*. (doi:10.48550/arXiv.2204.02405)
73. Finn C, Abbeel P, Levine S. 2017 Model-agnostic meta-learning for fast adaptation of deep networks. In *Int. Conf. on Machine Learning*, pp. 1126–1135. PMLR.
74. Jun H, Nichol A. 2023 Shap-e: Generating conditional 3d implicit functions. *arXiv* (doi:10.48550/arXiv.2305.02463)
75. Debussche A. 2010 Weak approximation of stochastic partial differential equations: the nonlinear case. *Math. Comput.* **80**, 89–117. (doi:10.1090/s0025-5718-2010-02395-6)
76. Gyöngy I. 1998 Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise I. *Potential Anal.* **9**, 1–25.
77. Gyöngy I. 1999 Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise II. *Potential Anal.* **1**, 1–37.
78. Hausenblas E. 2003 Approximation for semilinear stochastic evolution equations. *Potential Anal.* **18**, 141–186. (doi:10.1023/A:1020552804087)
79. Hausenblas E. 2003 Weak approximation for semilinear stochastic evolution equations. In *Stochastic analysis and related topics viii*, pp. 111–128. Basel: Springer. (doi:10.1007/978-3-0348-8020-6_5)
80. Shardlow T. 1999 Numerical methods for stochastic parabolic PDEs. *Numer. Funct. Anal. Optim.* **20**, 121–145. (doi:10.1080/01630569908816884)
81. Yoo H. 2000 Semi-discretization of stochastic partial differential equations on R by a finite-difference method. *Math. Comput.* **69**, 653–666. (doi:10.1090/S0025-5718-99-01150-3)
82. Kingma DP, Salimans T, Poole B, Ho J. 2021 Variational diffusion models. In *Advances in neural information processing systems* (eds A Beygelzimer, Y Dauphin, P Liang, JW Vaughan), Online, 6–14 December 2021. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/63069.html>
83. Bond-Taylor S, Willcocks CG. 2023 ∞ -diff: infinite resolution diffusion with subsampled mollified states. In *The Twelfth Int. Conf. on Learning Representations*.
84. Rogers LCG, Williams D. 2000 *Diffusions, markov processes, and martingales: itô calculus*. Cambridge, UK: Cambridge University Press. (doi:10.1017/CBO9780511805141)
85. Chojnowska-Michalik A, Goldys B. 1996 Nonsymmetric Ornstein-Uhlenbeck semigroup as second quantized operator. *J. Math. Kyoto Univ.* **36**, 481–498.
86. Manca L. 2008 Kolmogorov operators in spaces of continuous functions and equations for measures. PhD thesis, Scuola Normale Superiore di Pisa.

87. Sommer SH, Schauer M, Meulen F. 2021 *Stochastic flows and shape bridges*. Statistics of Stochastic Differential Equations on Manifolds and Stratified Spaces (Hybrid Meeting), no.48 in Oberwolfach Reports, Mathematisches Forschungsinstitut Oberwolfach. pp. 18–21.
88. Sun Y, Wu Z, Chen Y, Feng BT, Bouman KL. 2024 Provable probabilistic imaging using score-based generative priors. *IEEE Trans. Comput. Imaging* **10**, 1290–1305. (doi:10.1109/tci.2024.3449114)
89. Ho J, Salimans T. 2021 Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, Online, 6–14 December 2021. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/63069.html>
90. Bansal A, Borgnia E, Chu HM, Li J, Kazemi H, Huang F, Goldblum M, Geiping J, Goldstein T. 2024 Cold diffusion: inverting arbitrary image transforms without noise. In *Advances in neural information processing systems*, New Orleans, LA, 10–16 December 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75280.html>
91. Daras G, Delbracio M, Talebi H, Dimakis A, Milanfar P. 2023 Soft diffusion: score matching with general corruptions. TMLR. <https://jmlr.org/tmlr/>
92. Hooeboom E, Salimans T. 2023 Blurring diffusion models. In *The Eleventh Int. Conf. on Learning Representations*, Kigali, Rwanda, 1–5 May 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/75096.html>
93. Rissanen S, Heinonen M, Solin A. 2022 Generative modelling with inverse heat dissipation. *arXiv*. (doi:10.48550/arXiv.2206.13397)
94. Albergo MS, Boffi NM, Vanden-Eijnden E. 2023 Stochastic interpolants: a unifying framework for flows and diffusions. *arXiv* (doi:10.48550/arXiv.2303.08797)
95. Song Y, Dhariwal P, Chen M, Sutskever I. 2023 Consistency models. In *International Conference on Machine Learning (ICML)*, Honolulu, HI, 23–29 July 2023. Red Hook, NY: Curran Associates, Inc. <https://www.proceedings.com/73019.html>