Stochastic Interpolants in Hilbert Spaces

James Boran Yu

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Master of Philosophy in Machine Learning and Machine Intelligence

Pembroke College

August 2025

To my loving family.

Declaration

I, James Boran Yu of Pembroke College, being a candidate for the MPhil in Machine Learning and Machine Intelligence, hereby declare that this report and the work described in it are my own work, unaided except as may be specified below, and that the report does not contain material that has already been used to any substantial extent for a comparable purpose.

TODO: Signed, Date

TODO: Software declaration

TODO: Word count

James Boran Yu

August 2025

Acknowledgements

TODO Acknowledgements

Abstract

TODO ABSTRACT!

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation and Overview

TODO!

1.2 Contributions

This thesis develops a novel framework for generative modelling on function spaces. Our primary contributions are as follows.

1. We formulate stochastic interpolants directly in infinite-dimensional settings, which forms the core of our proposed framework.

2. Our framework goes beyond bridging marginal distributions and is the first to consider a bridge to the *conditional* distribution of target data, conditional on source data.

3. We provide a rigourous theoretical analysis, establishing sufficient conditions under which the framework is well-posed and satisfies critical theoretical guarantees.

4. We translate these theoretical insights into practical design principles to improve the algorithm's performance.

5. We demonstrate our framework's effectiveness for solving partial differential equation (PDE)-based forward and inverse problems, achieving results competitive with state-of-the-art approaches but with reduced inference time.

6. Finally, we outline areas for further research, such as extending our theoretical guarantees under more relaxed assumptions and developing novel practical designs.

1.3 Outline

This thesis is structured as follows.

Chapter 1 provides the motivation and overview for this thesis.

Chapter 2 presents the necessary groundwork for this thesis: we provide an overview of stochastic interpolants in their original finite-dimensional setting, as proposed by Albergo et al. (2023a), and contrast this with diffusion models for generative modelling (Song et al., 2021). We then give an overview of the key mathematical concepts necessary to generalise stochastic interpolants to infinite-dimensional spaces, and provide a review of related works in generative modelling on function spaces.

Chapter 3 introduces our core framework: a formulation of stochastic interpolants directly in infinite dimensions. We present a Hilbert space-valued SDE and justify its suitability for generative modelling and prove sufficient conditions for the well-posedness of such an SDE. We provide a training objective and relate this to an error bound of the learned generative process. From this theoretical analysis, we describe how our framework is useful for solving both forward and inverse problems and identify key design principles informing the implemention of our method.

Chapter 4 details an application of our framework for solving PDE-based forward and inverse problems. We describe the datasets and methods used, and compare our results with current state-of-the-art stochastic and deterministic solvers.

?? describes the merits of our work, as well as some limitations and potential areas for further work.

TODO: mention optimal transport in future work

TODO: make sure you frame the entire paper from the pov of bayesian inverse problems

TODO: add detail!

Chapter 2

Background and Preliminaries

In this chapter, we establish the conceptual and mathematical preliminaries to lay the necessary groundwork to formally generalise stochastic interpolants to function spaces. To achieve this, we structure our discussion as follows.

1. We begin by presenting diffusion models (DMs; Song et al., 2021) in finite dimensions.
2. Then, we describe key advantages of the stochastic interpolants framework over DMs, and present a form of stochastic interpolants in their original finite-dimensional context, as proposed by Albergo et al. (2023a).
3. We define Hilbert spaces as the underlying setting for our analysis, and present an overview of the key mathematical concepts necessary to describe random variables and stochastic differential equations (SDEs) in Hilbert spaces. Given these concepts, we outline key challenges in extending stochastic interpolants to infinite dimensions in Hilbert spaces.
4. Finally, we provide a review of related works which generalise DDPM and SBDM to function spaces, highlighting the relationship of these methods with their finite-dimensional counterparts.

2.1 Diffusion Models in Finite Dimensions

Diffusion models (DMs; Song et al., 2021) are a family of generative models achieving remarkable empirical success across a broad range of domains. To generate data x distributed according to a target measure μ_{target} on N-dimensional Euclidean space $\mathbb{R}^N$, we define two stochastic processes on a finite time interval $[0,T]$. For a drift coefficient $f : [0,T] \times \mathbb{R}^N \to \mathbb{R}^N$

and diffusion coefficient $g : [0,T] \to \mathbb{R}_{>0}$, the *diffusion process* $\mathbb{X} = \{X_t\}_{t\in[0,T]}$, is the solution to the following *forward SDE*:

$$\mathrm{d}X_t = f(t,X_t)\,\mathrm{d}t + g(t)\,\mathrm{d}W_t\,, \quad X_0 \sim \mu_{\text{target}},$$

where $\mathbb{W} := \{W_t\}_{t\in[0,T]}$ is a standard dimensional Wiener process.

Let μ_t be the law (marginal distribution) of X_t and let $p_t : \mathbb{R}^N \to \mathbb{R}$ be the density of μ_t with respect to the Lebesgue measure. Under some mild regularity conditions (Anderson, 1982) we may define a *time-reversed process* $\overline{\mathbb{X}} = \left\{\overline{X}_t\right\}_{t\in[0,T]}$, which when solved backwards in time from $\overline{X}_T \sim \mu_T$ yields a sample $\overline{X}_0 \sim \mu_{\text{target}}$:

$$\mathrm{d}\overline{X}_t = \left(f(t,\overline{X}_t) - g^2(t)\nabla \log p_t(\overline{X}_t)\right)\mathrm{d}t + g(t)\,\mathrm{d}\overline{W}_t\,, \quad \overline{X}_T \sim \mu_T, \tag{2.1}$$

where $\overline{\mathbb{W}} := \left\{\overline{W}_t\right\}_{t\in[0,T]}$ is a standard Wiener process when time flows backwards from $t = T$ to 0, and $\nabla \log p_t(x)$ is the *score* of the marginal distribution at time t, namely, the spatial derivative of the log-density of X_t.

By learning a time-dependent score network $s_\theta(t,x)$ and plugging this in place of $\nabla \log p_t(x)$ in Equation (2.1), we may generate approximate samples from μ_{target}, provided we have samples from μ_T.

To ensure that μ_T is a simple and tractable distribution, f and g are typically chosen such that the forward process systematically transforms data $X_0 \sim \mu_{\text{target}}$ into a Gaussian $\mathrm{N}(0,\sigma_T^2 I_N)$. However, this transformation is only guaranteed to be perfect asymptotically as $T \to \infty$. In a practical implementation, we must terminate time at a finite time step T. This introduces a bias during sampling, since the final condition for the time-reversed SDE is not a Gaussian at time T.

For example, *score-based diffusion models* (SBDMs) are a special case of DMs in which the forward SDE is an Ornstein-Uhlenbeck process. In this case, the law of X_t converges to a standard Gaussian $\mathrm{N}(0,I_N)$ in the limit $t \to \infty$.

$$\mathrm{d}X_t = -X_t d_t + \sqrt{2}\,\mathrm{d}W_t\,, \quad X_0 \sim \mu_{\text{target}}.$$

While a larger T bridges the data closer to a Gaussian, a smaller T helps improve the learned approximation $s_\theta(t,x)$ of the score and leads to more tractable sampling when solving the reverse process. Hence, a tradeoff must be found when choosing T (see, for example, Franzese et al., 2023).

2.2 Stochastic Interpolants in Finite Dimensions

Stochastic interpolants (SIs) are a class of generative models which provide the following improvements in flexibility over DMs:

1. SIs can bridge between any two arbitrary distributions determined *a priori*, as opposed to between a single target distribution and a fixed noise prior. Moreover, the source and target distributions can be coupled, allowing SIs to model a joint probability law between source and target data. This provides a powerful and flexible framework, where a single trained model can perform unconditional generation in addition to solving both forward and inverse tasks within a Bayesian setting.

2. The interpolation is constructed on a finite time horizon, in contrast to DMs which rely on an asymptotic convergence to the simple noise prior. By design, this has two advantages: it removes approximation bias from the terminal distribution and eliminates the need to tune the time horizon as a hyperparameter.

3. The interpolation path is an explicit design choice, allowing us to construct simple bridges (e.g., linear trajectories) between the two distributions. This contrasts with DMs, where the trajectory is an emergent property determined by the specific SDE. Simple, low-curvature paths are easier for numerical solvers to approximate accurately, which can lead to greater sampling efficiency with fewer function evaluations.

Each of these merits is demonstrated in a function generation setting in Chapter 4: we show that our framework is highly effective for solving PDE-based forward and inverse problems. Notably, this is achieved on a strict finite time interval, and with fewer function evaluations and reduced inference time.

Having stated the key merits of SIs over DMs, we now introduce SIs in their finite-dimensional setting, as proposed by Albergo et al. (2023a,b). To establish the necessary context for our subsequent development in infinite dimensions, the following discussion captures the conceptual essence of SIs in finite dimensions. A formal and detailed presentation of the specific regularity conditions in our infinite-dimensional setting will be provided in Chapter 3.

Let μ be a joint measure on $\mathbb{R}^N \times \mathbb{R}^N$ with marginals μ_0 and μ_1. We draw a (possibly coupled) pair of random variables $\xi = (\xi_0, \xi_1) \sim \mu$, where we refer to μ_0 as the *source* and μ_1 as the *target distribution*.

Let z be a standard N-dimensional Gaussian, distributed independently of ξ. A *stochastic interpolant* is a family of random variables $\{x_t\}_{t\in[0,1]}$ indexed by time $t \in [0,1]$:

$$x_t = \alpha(t)\xi_0 + \beta(t)\xi_1 + \gamma(t)z, \quad t \in [0,1],$$

where $\alpha, \beta, \gamma : [0,1] \to \mathbb{R}_{\geq 0}$ are such that α, β are continuously differentiable on $[0,1]$ and γ is continuous on $[0,1]$ and continuously differentiable on $(0,1)$. They satisfy the boundary conditions $\alpha(0) = \beta(1) = 1$, $\alpha(1) = \beta(0) = 0$, $\gamma(0) = \gamma(1) = 0$ and $\gamma(t) > 0$ for all $t \in (0,1)$. We denote their time derivatives respectively by $\dot{\alpha}, \dot{\beta}, \dot{\gamma}$. Additionally, we denote $\dot{x}_t := \dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1 + \dot{\gamma}(t)z$

Intuitively, the boundary conditions on I and γ ensure that the law of the stochastic interpolant matches the source and target distributions at the endpoints, $x_0 \sim \mu_0$ and $x_1 \sim \mu_1$. For intermmediate times $t \in (0,1)$, the law of x_t is equal to that of a deterministic path between ξ_0 and ξ_1, corrupted by scaled Gaussian noise.

To bridge from μ_0 to μ_1, we choose a positive constant $\varepsilon > 0$ and define a *forward SDE* as follows:

$$\mathrm{d}X_t = \left(\mathbb{E}\left[\dot{x}_t \mid x_t = X_t\right] + \varepsilon \nabla \log p_t(X_t)\right)\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t, \quad X_0 \sim \mu_0, t \in [0,1],$$

where p_t is the density of the law of the interpolant x_t at time t, with respect to the Lebesgue measure. Under suitable regularity conditions, Albergo et al. (2023a) show that the law of X_t at any time $t \in [0,1]$ is equal to the law of x_t. Hence, by solving the forward SDE, we generate a sample from the target distribution μ_1.

Similarly, we define a *time-reversed SDE* which, when solved backwards in time starting from $\overline{X}_1 \sim \mu_1$, gives a sample from the source distribution μ_0:

$$\mathrm{d}\overline{X}_t = \left(\mathbb{E}\left[\dot{x}_t \mid x_t = X_t\right] - \varepsilon \nabla \log p_t(X_t)\right)\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}\overline{W}_t, \quad \overline{X}_1 \sim \mu_1, t \in [0,1].$$

In the special case where $\varepsilon = 0$, the forward and time-reversed SDEs collapse to a *probability flow ODE*, where the source of stochasticity only comes from the initial/final conditions, in contrast to $\varepsilon > 0$ where additional noise is injected by the Wiener process.

2.3 Mathematical Preliminaries

Generalising stochastic interpolants to infinite dimensions requires confronting several theoretical challenges. To understand these challenges and to construct our infinite-dimensional framework in Chapter 3, we review some fundamental mathematical preliminaries.

Hilbert Spaces A *Hilbert space H* is a vector space equipped with a scalar-valued inner product $\langle f,g\rangle_H$, which is *complete* with respect to the norm $\|f\|_H := \sqrt{\langle f,f\rangle_H}$ induced by this inner product, that is, every H-valued Cauchy sequence converges in H-norm to an element in H. The choice of a Hilbert space, as opposed to a more general Banach space, is justified by the fact that the inner product provides essential geometric structure, giving rise to the concept of orthogonality.

Throughout, we let H be an infinite dimensional Hilbert space satisfying the following two properties:

1. H is *real*, meaning that all scalars, including inner products, are real-valued.
2. H is *separable*, which has the implication that there exists a *countable* orthonormal basis for H.

We develop our framework by viewing functions as vectors living in H. Hence, we use the terms *vector* and *function* interchangeably.

Gaussian Measures A random variable x is distributed according to a *Gaussian measure* on a real, separable Hilbert space H if, for all $f \in H$, the inner product $\langle f,x\rangle_H \in \mathbb{R}$ is distributed according to a one-dimensional Gaussian. Such a Gaussian measure is completely determined by its mean $m \in H$ and a *covariance operator*, defined as a bounded, self-adjoint, positive-semidefinite, linear operator $C : H \to H$ which satisfies:

$$\langle Cf,g\rangle_H = \langle f,Cg\rangle_H = \mathrm{Cov}\left[\langle f,x\rangle_H \langle g,x\rangle_H\right] = \mathbb{E}\left[\langle f-m,x\rangle_H \langle g-m,x\rangle_H\right],$$

for all $f,g \in H$. Hence we denote the law of x by $\mathrm{N}(m,C)$.

Let $\{e_n\}_{n=1}^{\infty}$ be an orthonormal basis of eigenvectors of C with corresponding eigenvalues $\{\lambda_n\}_{n=1}^{\infty}$. We call C *trace class*, if

$$\mathrm{Tr}(C) := \sum_{n=1}^{\infty} \langle Ce_n,e_n\rangle_H = \sum_{n=1}^{\infty} \lambda_n < \infty.$$

This condition is critical in infinite dimensions: for a Gaussian to be supported on H, its expected squared norm must be finite, and this value is equal to $\|m\|_H^2 + \mathrm{Tr}(C)$. A Gaussian with non-trace-class noise will have samples which are almost-surely unbounded in norm and hence do not belong to the Hilbert space H. To ensure that samples are well-defined, we focus only on the case of Gaussians with trace-class covariance.

Cameron-Martin Spaces For a covariance operator C, the *Cameron-Martin space*, H_C, is an (infinite-dimensional) subspace of H defined as the image of H under $C^{\frac{1}{2}}$. The Cameron-Martin space is a Hilbert space itself when equipped with the inner product $\langle f,g\rangle_{H_C} := \left\langle C^{-\frac{1}{2}}f, C^{-\frac{1}{2}}g\right\rangle_H$.

If C is trace class its eigenvalues must decay to zero. Hence, the eigenvalues of the operator $C^{-\frac{1}{2}}$ diverge to infinity, making $C^{-\frac{1}{2}}$ an unbounded operator on H. Critically, this implies that the H_C is a strict, dense subspace of H. An element $f \in H$ belongs to the subspace H_C only if its coefficients in the eigenbasis of C decay sufficiently quickly to ensure its Cameron-Martin norm is finite.

Intuitively, since the eigenvalues of C are typically lowest for high-frequency modes, this condition means that elements of H_C are fundamentally smoother than arbitrary elements of H, as they are constrained to have little energy in their high-frequency components.

If $H = L^2(D,\mu_D)$ is the set of all square-integrable functions defined on a domain D with respect to a finite measure μ_D, equipped with the inner product $\langle f,g\rangle_H = \int_D f(x)g(x)\mu_D(\mathrm{d}x)$, then the Cameron-Martin space H_C for a trace-class covariance operator C there exists a unique positive-semidefinite kernel function $k : D \times D \to \mathbb{R}_{\geq 0}$ such that

$$Cf(x) = \int_D k(x,y)f(y)\mu(\mathrm{d}y), \text{ for all } f \in H.$$

Consequently, H_C is a reproducing kernel Hilbert space (RKHS) with k as its reproducing kernel. Intuitively, this provides another reason why H_C is a strict subspace of H: the defining property of the RKHS, that pointwise evaluation of functions is continuous in H_C-norm, imposes a strong regularity condition that functions in H_C are sufficiently smooth.

A fundamental result in the theory of Gaussian measures is that when C is trace-class, samples from $N(0,C)$ are almost surely not in H_C even though they belong to the larger space H.

2.4 Challenges in Extending SIs to Infinite Dimensions

Equipped with these mathematical foundations, we now identify the key challenges which arise when extending SIs to infinite dimensions.

Choice of Gaussian Noise As discussed, samples from a Gaussian $\mathrm{N}(0,C)$ on H almost surely do not belong to H unless C is trace class. Crucially, this rules out allowing the noise z in an interpolant to be isotropic.

To construct a well-defined interpolant, we restrict the noise z to be drawn from a Gaussian with trace-class covariance. We provide design principles for selecting this covariance to achieve desirable properties in the interpolation path.

No Lebesgue measure Typically in finite dimensions, densities are taken with respect to the Lebesgue measure. However, the Lebesgue measure does not exist in infinite dimensions. Crucially, this makes the score $\nabla \log p_t(x)$ ill-defined. One might consider defining the density p_t of the interpolant x_t with respect to some reference Gaussian measure. However due to the time-varying noise schedule $\gamma(t)z$, this approach faces a crucial obstacle stemming from the Feldman-Hajek theorem: Gaussian measures whose covariance operators are different scaled versions of the same operator are mutually singular. This implies the law of x_t is not absolutely continuous with respect to any single reference Gaussian for all t.

To resolve the issue of the ill-defined score, our work extends a key insight from finite-dimensional stochastic interpolants: Albergo et al. (2023a, Theorem 2.8) show that the score $\nabla \log p_t(x)$ can be computed via the conditional expectation $\frac{1}{\gamma(t)} \mathbb{E}[z \mid x_t = x]$. We show a similar principle is true in infinite dimensions. By defining and computing our score operator via a conditional expectation, we avoid the requirement of a global reference measure.

Well-Posedness of SDEs In finite dimensions, the convolution of interpolated data with scaled noise $\gamma(t)z$ has a regularising effect, ensuring the corresponding SDE is well-posed. This guarantee is lost in infinite dimensions, where the regularizing effect of Gaussian noise on arbitrary measures is often insufficient. This can result in a drift term that is unbounded and/or non-Lipschitz, violating the conditions ensuring the uniqueness or even existence of solutions.

To address this challenge, we establish a set of sufficient conditions on the source and target measures which ensure the drift remains well-behaved, thus guaranteeing the existence and uniqueness of the solution to the infinite-dimensional SDE.

We acknowledge that the sufficient conditions required by our formulation to guarantee a well-posed SDE are strong and unlikely to be strictly met in practice. Nevertheless, we contend that the value of our theoretical framework lies in the design principles it provides for constructing models in empirical settings to ensure stable and well-behaved interpolants.

2.5 Related Works

Generalisations of DMs in infinite dimensions

SIs with coupled data

Forward and inverse problems

PDE-based forward and inverse problems

Neural operators

2.6 Summary

Chapter 3

Construction and Well-Posedness

In Chapter 2, we introduced stochastic interpolants (SIs) in their original finite-dimensional setting, noting their advantages over diffusion models (DMs). While DMs have been successfully generalised to achieve state-of-the-art results in function spaces, SIs have not yet been framed in function spaces. Furthermore, existing SI formulations are primarily generative; they do not explicitly guarantee that evolving a process from a point yields a sample from the true conditional target distribution. This conditional sampling capability is essential for the Bayesian inverse problems that are a central motivation for this thesis.

This chapter addresses both of these gaps. We develop a framework for stochastic interpolants on infinite-dimensional Hilbert spaces, explicitly addressing the cases of non-conditional and conditional conditional sampling. We will refer to the former as a *marginal bridge* and the latter as an *conditional bridge*.

For clarity of presentation, our formal analysis will focus on the process that evolves from the source to the target distribution. The corresponding results for the time-reversed evolution are analogous, and we detail this symmetry in **??**.

3.1 Framework

Let H be a real, separable Hilbert space equipped with the inner product $\langle \cdot, \cdot \rangle_H$ and let μ be a Borel probability measure on the product space $H \times H$. The marginals of μ, denoted by μ_0 and μ_1, are the pushforward measures under the canonical projection maps onto the first and second components of the project space, that is, $\mu_0(\mathrm{d}\xi_0) = \mu(\mathrm{d}\xi_0 \times H)$ and $\mu_1(\mathrm{d}\xi_1) = \mu(H \times \mathrm{d}\xi_1)$.

Definition 1. A *stochastic interpolant* (SI) is a family of H-valued random variables $\{x_t\}_{t\in[0,1]}$ indexed by time $t \in [0,1]$ such that

$$x_t = \alpha(t)\xi_0 + \beta(t)\xi_1 + \gamma(t)z,$$

where:

1. $\alpha(t), \beta(t), \gamma(t) : [0,1] \to \mathbb{R}_{\geq 0}$ defined such that α and β are continuously differentiable on $[0,1]$, and γ is continuous on $[0,1]$ and continuously differentiable on $(0,1)$. They satisfy the boundary conditions $\alpha(0) = \beta(1) = 1, \alpha(1) = \beta(0) = 0$, $\gamma(0) = \gamma(1) = 1$, and $\gamma(t) > 0$ for all $t \in (0,1)$.

2. The pair of random variables $\xi = (\xi_0, \xi_1)$ is drawn from the joint probability measure μ.

3. The random variable z distributed independently of ξ and drawn from a Gaussian measure $\mathrm{N}(0,C)$, where $C : H \to H$ is a positive-definite trace-class covariance operator.

Throughout, we denote $\dot{x}_t := \dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1 + \dot{\gamma}(t)z$. We refer to the components of the data pair $\xi = (\xi_0, \xi_1) \sim \mu$ as the *source data* ξ_0 and *target data* ξ_1, with corresponding *source distribution* μ_0 and *target distribution* μ_1. The joint measure μ also induces a conditional distribution of the target given source data: for μ_0-almost every $x \in H$, we write $\mu_{1|0}(\mathrm{d}x, x_0)$ to denote the conditional distribution of ξ_1 on H, conditional on $\xi_0 = x$.

3.1.1 Marginal Bridge

We first construct a stochastic process that bridges the source distribution μ_0, to the target distribution, μ_1. We refer to this process as the *marginal bridge*, which distinguishes it from the *conditional bridge* to be detailed in Section 3.1.2.

Using the same terminology as in Albergo et al. (2023a), we define *velocity* and *denoiser* functions $\zeta, \eta : [0,1] \times H \to H$ to be the following conditional expectations.

$$\zeta(t,x) := \mathbb{E}[\dot{x}_t \mid x_t = x], \tag{3.1}$$

$$\eta(t,x) := \mathbb{E}[z \mid x_t = x]. \tag{3.2}$$

The marginal bridge is a stochastic process X_t governed by the following equation, which we call the MB-SDE:

$$\mathrm{d}X_t := \left(\zeta(t,X_t) - \frac{\varepsilon}{\gamma(t)}\eta(t,X_t)\right)\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t\,, \quad X_0 \sim \mu_0. \tag{3.3}$$

where W_t is a C-Wiener process and $\varepsilon \geq 0$ is a scalar. We use the following to denote the drift coefficient of the MB-SDE (3.3):

$$f(t,x) := \zeta(t,x) - \frac{\varepsilon}{\gamma(t)}\eta(t,x) \tag{3.4}$$

Assuming that the MB-SDE (3.3) has any weak solution on a, possibly strict, subinterval $[0,\bar{t}] \subseteq [0,1]$, standard results (see e.g., Da Prato and Zabczyk, 2014, Chapter 14.2.2) show that for dt-almost every $t \in [0,\bar{t}]$, the marginal distribution ρ_t of this solution at time t satisfies the following *Fokker-Plank* equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_H u(t,x)\rho_t(\mathrm{d}x) = \int_H \mathscr{L}u(t,x)\rho_t(\mathrm{d}x), \tag{3.5}$$

for all test functions $u(t,x)$ in the space E formed by the linear span of the real and imaginary components of functions of the form

$$u_{\phi,h}(t,x) = \phi(t)e^{i\langle x,h(t)\rangle_H}, \text{ for any } \phi \in C^1([0,\bar{t}]), h \in C^1([0,\bar{t}];H), \tag{3.6}$$

and where where $\mathscr{L}$ is a *Kolmogorov operator* given by:

$$\mathscr{L}u(t,x) := \mathrm{Tr}\left(\varepsilon C D_x^2 u(t,x)\right) + D_t u(t,x) + \langle f(t,x), D_x u(t,x)\rangle_H.$$

We use D_t to denote the derivative in time, and D_x, D_x^2 the first and second-order Frechet derivatives in Hilbert space.

The Fokker-Planck equation (3.5) fundamentally describes the evolution of the probability distribution of a stochastic process. In finite dimensions, this is typically stated directly in terms of the density of the law of the solution at each time point, with respect to the Lebesgue measure. In contrast, in infinite dimensions a time-uniform reference measure is not guaranteed to exist and hence we instead state the Fokker-Plank equation in terms of test functions $u(t,x)$.

To show that the MB-SDE (3.3) provides a valid path that correctly transports a source measure μ_0 to a target measure μ_1, we show that the marginal distribution μ_t of our stochatic interpolant also satisfies Equation (3.5) on the entire time interval $t \in [0,1]$. Our main technical contribution is showing this relationship holds in infinite-dimensions via test functions, avoiding the need to express measures via densities.

Lemma 2. *Let μ_t be the marginal distribution of the stochatic interpolant x_t, defined in Definition 1. For every $t \in [0,1]$, the measure μ_t satisfies the Fokker-Plank equation (3.5).*

Proof (sketch). The full proof is presented in Section A.1 in Appendix A. Our strategy is to consider the characteristic function of the real-valued random variable $u(t,x_t)$ to provide an expression for the time derivative of the expected value of $u(t,x_t)$, which is the left-hand side of Equation (3.5). We apply the law of iterated expectations to express this in terms of the drift term $f(t,x_t)$. We then recover the trace term by applying Parseval's theorem and expressing inner products as an infinite sum of projections onto an eigenbasis of the covariance operator C. ■

Having established that both ρ_t and μ_t satisfy the Fokker-Plank equation (3.5), we state our main result justifying the MB-SDE (3.3) as a suitable stochastic process allowing one to bridge μ_0 to μ_1.

Theorem 3. *Let μ_t be the law of the stochastic interpolant x_t at time t.*

1. *Suppose that the MB-SDE (3.3) has solutions which are unique in law on a non-empty time interval $[0,\bar{t}] \subseteq [0,1]$. We denote the law of X_t by ρ_t.*
2. *Suppose that $\mathscr{L}E$ is dense in $L^1([0,\bar{t}] \times H, \nu)$, where ν is the measure on $[0,\bar{t}] \times H$ determined uniquely by*
$$\nu(\mathrm{d}(t,x)) = \nu_t(\mathrm{d}x)\,\mathrm{d}t\,,$$
and $\nu_t := \frac{1}{2}\rho_t + \frac{1}{2}\mu_t$ for each $t \in [0,\bar{t}]$.

Then, for $\mathrm{d}t$*-almost every $t \in [0,\bar{t}]$, we have*

$$\rho_t = \mu_t\,.$$

Proof (sketch). The full proof is presented in Section A.2 in Appendix A. We follow a similar line of reasoning to Bogachev et al. (2010, Theorem 2.1), who study the uniqueness of solutions to Fokker-Plank equations in infinite dimensions. By exploiting the deness of $\mathscr{L}E$ in $L^1([0,1] \times H, \nu)$, we show for $\mathrm{d}t$-almost every t that the signed measure $\rho_t - \mu_t$ is zero, and hence $\rho_t = \mu_t$. ■

Theorem 3 means that the MB-SDE (3.3) successfully bridges from the source to the target distribution: starting with a sample from the source distribution, we can solve the MB-SDE (3.3) forward in time to obtain a samples from the source distribution μ_0 provided we can learn the drift coefficient $f(t,x)$.

The validity of this result rests on two key assumptions. Our subsequent analysis in **??** addresses the first assumption, the existence of a unique weak solution, by proving a stronger result: the existence and uniqueness of a *strong solution*. Strong uniqueness enables us to employ a coupling argument to bound the Wasserstein distance between our generated samples and the true target distribution (see TODO **??**).

Our second assumption adopts the framework of Bogachev et al. (2010, Theorem 2.1). The density condition on the Kolmogorov operator's range guarantees uniqueness for the Fokker-Planck equation. This technical requirement ensures the space of test functions is sufficiently rich to exclude spurious solutions to the Fokker-Planck equation beyond the one generated by the MB-SDE. While essential for our proof, a detailed analysis of the minimal requirements to ensure it holds is a distinct line of inquiry that we leave for future work.

Thus far, we have focused on the marginal bridge SDE, which provides a mechanism to sample from a target distribution μ_1. However, to solve Bayesian forward and inverse problems we are required not to sample from a marginal, but from a conditional distribution. To address this, we now extend our framework to construct a conditional bridge SDE (CB-SDE). We detail this process in the following section.

3.1.2 Conditional Bridge

We now construct a stochastic process called the *conditional bridge* which, conditional on a draw $\xi_0 \sim \mu_0$, forms a bridge to the conditional distribution $\mu_{1|0}(\mathrm{d}\xi_1, \xi_0)$.

We define *conditional velcoity* and *denoiser* functions $\zeta, \eta : [0,1] \times H \times H \to H$ to be the following conditional expectations:

$$\zeta(t, x_0, x) := \mathbb{E}[\dot{x}_t \mid \xi_0 = x_0, x_t = x], \tag{3.7}$$

$$\eta(t, x_0, x) := \mathbb{E}[z \mid \xi_0 = x_0, x_t = x]. \tag{3.8}$$

The conditional bridge is a stochastic process X_t governed by the following equation, which we call the CB-SDE:

$$\mathrm{d}X_t := \left(\zeta(t, \xi_0, X_t) - \frac{\varepsilon}{\gamma(t)} \eta(t, \xi_0, X_t) \right) \mathrm{d}t + \sqrt{2\varepsilon}\, \mathrm{d}W_t, \quad X_0 = \xi_0. \tag{3.9}$$

We use the following to denote the drift coefficient of the CB-SDE:

$$f(t, x_0, x) := \zeta(t, x_0, x) - \frac{\varepsilon}{\gamma(t)} \eta(t, x_0, x).$$

For μ_0-almost every ξ_0, we denote by $\mu_{t|0}(\mathrm{d}x, \xi_0)$ the distribution of the interpolant x_t, conditional on ξ_0. Furthermore, assuming the CB-SDE (3.9) has a unique weak solution on a subinterval $[0,\bar{t}] \subseteq [0,1]$, we let $\rho_{t|0}(\mathrm{d}x\,,\xi_0)$ be the law of X_t at time $t \in [0,\bar{t}]$, conditional on $X_0 = \xi_0$.

We follow an analogous logic to the proof of Lemma 2 to show that $\rho_{t|0}(\mathrm{d}x\,,\xi_0)$ and $\mu_{t|0}(\mathrm{d}x\,,\xi_0)$ are both solutions to a common Fokker-Plank equation with the following Kolmogorov operator indexed by ξ_0:

$$\mathscr{L}_{\xi_0} u(t,x) \coloneqq \operatorname{Tr}\left(\varepsilon C D_x^2 u(t,x)\right) + D_t u(t,x) + \langle f(t,\xi_0,x), D_x u(t,x)\rangle_H\,.$$

Hence, the CB-SDE (3.9) is a suitable stochastic process where, conditional on a starting point $X_0 = \xi_0$, we may bridge to the conditional distribution $\mu_{1|0}(\mathrm{d}\xi_1\,,\xi_0)$. We state this result directly blow, and provide a full proof in **??**.

Theorem 4. *Let $\mu_{t|0}(\mathrm{d}x\,,\xi_0)$ be the law of the stochastic interpolant x_t at time t, conditional on ξ_0.*

1. *Suppose that for μ_0-almost every initial condition $X_0 = \xi_0$, the CB-SDE (3.3) has solutions which are unique in law on a non-empty time interval $[0,\bar{t}] \subseteq [0,1]$. We denote the law of X_t conditional on $X_0 = \xi_0$ by $\rho_{t|0}(\mathrm{d}x\,,\xi_0)$.*

2. *Suppose that for μ_0-almost every ξ_0, the set $\mathscr{L}_{\xi_0}E$ is dense in $L^1([0,\bar{t}] \times H, \nu_{\xi_0})$, where ν_{ξ_0} is the measure on $[0,\bar{t}] \times H$ determined uniquely by*

$$\nu_{\xi_0}(\mathrm{d}(t,x)) = \nu_{\xi_0,t}(\mathrm{d}x\,,\xi_0)\,\mathrm{d}t\,,$$

and $\nu_{\xi_0,t}(\mathrm{d}x\,,\xi_0) \coloneqq \frac{1}{2}\rho_{t|0}(\mathrm{d}x\,,\xi_0) + \frac{1}{2}\mu_{t|0}(\mathrm{d}x\,,\xi_0)$ for each $t \in [0,\bar{t}]$.

Then, for $\mathrm{d}t$-almost every $t \in [0,\bar{t}]$, we have

$$\rho_{t|0}(\mathrm{d}x\,,\xi_0) = \mu_{t|0}(\mathrm{d}x\,,\xi_0).$$

Formally, the drift coefficient $f(t,\xi_0,X_t)$ is a *random function* coupled to the specific initial condition $X_0 = \xi_0$. The uniqueness assumption (1) in Theorem 4 is hence identical to (1) for the marginal bridge (Theorem 3) but restated to emphasise its dependence on this initial condition. In contrast, the dense range condition (2) is necessarily stronger than its marginal counterpart (2) to ensure uniqueness for every conditional path.

The CB-SDE differs from the MB-SDE only in the inclusion of ξ_0 as an additional conditioning variable when defining the conditional velocity and denoiser functions (Equations 3.7 and 3.8), which guarantee a bridge for each conditional path. To the best of our knowledge, this is the first statement of stochastic interpolants explicitly considers conditional paths between the source and target distributions. While Albergo et al. (2023b) consider SIs in which the source and target distributions are coupled, they do so to show that such a coupling provides simpler sampling paths, but without explicitly conditioning on the initial condition, their framework still only provides a marginal bridge. To illustrate this, we note that the CB-SDE and MB-SDE are equivalent when the following mean-independence conditions hold:

$$\mathbb{E}[\dot{x}_t \mid x_t = x] = \mathbb{E}[\dot{x}_t \mid \xi_0 = x_0, x_t = x],$$
$$\mathbb{E}[z \mid x_t = x] = \mathbb{E}[z \mid \xi_0 = x_0, x_t = x],$$

that is, conditioning on ξ_0 provides no further information than already provided by x_t. This is a very strong statistical requirement which we do not assume. For example, these conditions are true when ξ_0 is deterministic.

Since our primary focus is the application of SIs to forward and inverse problems, we henceforth center our analysis on the conditional bridge, with analogous results for the marginal bridge provided in the appendix. This approach is justified since the conditional bridge is a stronger construction: a bridge between marginal distributions can be readily recovered from the conditional bridge by marginalising over the source distribution.

We have established that conditional sample paths between source and target distributions can be obtained by solving the CB-SDE (3.9). To justify approximating (3.9) for conditional sampling, the next section ensures that a solution exists and is unique. This rules out spurious sample paths which result in a distribution other than $\mu_{1|0}(\mathrm{d}x, \xi_0)$.

3.2 Existence and Uniqueness of Strong Solutions

While Theorem 4 only requires the existence and uniqueness of solutions to the CB-SDE in the weak sense, we focus on strong solutions to facilitate later analysis in **??** on the Wasserstein distance between generated samples and the true target distribution. This will allow us to use the same Wiener process to provide a coupling between the true CB-SDE with an SDE based on a drift learned by a neural network.

Our approach is to first show a result on the Lipschitz continuity of the drift coefficient $f(t, x_0, x)$ as a function of x and use this result to show existence and uniqueness of strong solutions. We provide two different settings under which such a Lipschitz condition can be obtained.

In both settings, we assume the source and target data, ξ_0 and ξ_1, are supported on the Cameron-Martin space H_C of the covariance operator C. This is a strong regularity condition which ensures the noise is inherently rougher than the data, allowing for the derivation of the well-defined posterior measures used for conditioning on x_t and ξ_0.

The first setting directly addresses the case of Bayesian forward and inverse problems, in which we assume that the true data distribution μ is supported on the Cameron-Martin space H_C and has a density with respect to a reference Gaussian measure. We state these conditions in the following hypothesis.

Hypothesis 5. Let $H_C := C^{\frac{1}{2}}H$ be the Cameron-Martin space of C. We suppose the following conditions hold.

i The law μ of data ξ is supported on the product space $H_C^2 := H_C \times H_C$ and has zero mean.

ii μ has a density $p : H_C^2 \to \mathbb{R}_{\geq 0}$ with respect to a *prior* Gaussian measure $\mathbb{P} := N(0, Q)$ on H_C^2, where Q is a positive-definite trace-class covariance operator on H_C^2.

iii The negative log-density $\Phi := -\log p$ is twice differentiable and strongly convex, that is, there exists a scalar $k > 0$ where, for every $\lambda \in [0,1]$ and every $u, v \in H_C^2$, we have

$$\Phi(\lambda u + (1-\lambda)v) \leq \lambda \Phi(u) + (1-\lambda)\Phi(v) - \frac{k}{2}\lambda(1-\lambda)\|u-v\|_{H_C^2}^2.$$

Using Hypothesis 5, we establish the following result on the Lipschitz-continuity of the conditional expectation $\mathbb{E}[\xi_1 \mid \xi_0, x_t]$.

Proposition 6. *Suppose Hypothesis 5 holds. Then the map $x \mapsto f(t, x_0, x)$ is Lipschitz continuous with respect to the H_C-norm. Specifically, for each $t \in (0,1)$, $x_0 \in H_C$ and $x \in H$, the following inequality holds:*

$$\|f(t, x_0, x) - f(t, x_0, y)\|_{H_C} \leq L(t)\|x-y\|_{H_C},$$

where the Lipschitz constant $L(t)$ is:

$$L(t) = \max\left\{ \left| \frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)} \right|, \left| \dot{\beta}(t) - \beta\left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)} \right) \right| \frac{\beta(t)}{\beta^2(t) + k\gamma^2(t)} \right\}.$$

Proof (sketch). The full proof is presented in Section A.3 in Appendix A. First, we re-express the drift to isolate its dependence on x into two terms: a linear term and the posterior conditional mean $\mathbb{E}[\xi_1 \mid \xi_0 = x_0, x_t = x]$. The problem thus reduces to proving that this conditional expectation is a Lipschitz-continuous map in x.

Our proof strategy uses a Galerkin-type argument in which we use a sequence of finite-dimensional approximations, combined with Brascamp-Lieb inequality (Brascamp and Lieb, 1976). We introduce a sequence of approximating posterior measures defined on finite-dimensional subspaces $H_N \subset H_C$. For each N, we study the Frechet derivative of the approximate posterior mean on this subspace, with respect to x, which is precisely the corresponding posterior covariance operator C_N.

The core of our contribution is the application of the Brascamp-Lieb inequality to this setting. This inequality provides an upper bound on the operator-norm of C_N, in terms of the expectation of the inverse Hessian of the posterior log-density. By leveraging the strong convexity of the prior potential Φ, we establish a uniform lower bound on this Hessian in the Loewner order. This, in turn, yields a crucial upper bound on the operator norm of C_N, independent of the dimension N.

This uniform bound on the norm of the derivative translates directly into a Lipschitz inequality with Lipschitz constant independent of N. As we let $N \to \infty$, we show that the approximate posterior means converge to the true posterior mean, which therefore inherits this uniform Lipschitz property. ■

Our second setting replaces the density assumption on μ with an assumption that is support is bounded. This approach is particularly useful when the data $\xi = (\xi_0, \xi_1)$ are subject to geometric constraints. For instance, if the data lie on a manifold, it may be natural to assume that their support is bounded.

Proposition 7. *Suppose the law μ_1 of the target data ξ_1 is supported on a bounded subset of H_C, that is, there exists a scalar $R < \infty$ where $\|\xi_1\|_{H_C} < R$, μ_1-almost surely. Then the map $x \mapsto f(t, x_0, x)$ is Lipschitz continuous with respect to the H_C-norm. Specifically, for each $t \in (0,1)$ and $x_0, x \in H$, the following inequality holds:*

$$\|f(t,x_0,x) - f(t,x_0,y)\|_{H_C} \leq L(t)\|x-y\|_{H_C},$$

where the Lipschitz constant $L(t)$ is:

$$L(t) = \max\left\{\left|\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right|, \left|\dot{\beta}(t) - \beta\left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right)\right|\frac{R^2\beta(t)}{\gamma^2(t)}\right\}.$$

Proof. The full proof is presented in Section A.4 in Appendix A. The overarching argument follows that of Lemma 6, except that the argument is substantially simplified by the assumption that ξ_1 has bounded support in H_C, which allows a construction directly in infinite dimensions. ■

Both cases involve the essential assumption that the target data ξ_1 is supported on the Cameron-Martin space H_C. This ensures, via the Cameron-Martin theorem, that the law of the interpolant x_t has a well-defined Radon-Nikodym derivative with respect to a reference measure, which acts as the likelihood function and in turn facilitates an expression for the density of the posterior law of ξ_1 when conditioning on $\xi_0 = x_0$ and $x_t = x$.

Intuitively, the restriction of ξ_1 to H_C a smoothness assumption that confines realisations of ξ_1 to a class of functions that are fundamentally less rough than typical realisations of the noise $\gamma^2(t)z$. This assumption ensures that the laws Gaussian measures corresponding to translations of scaled-noise $\gamma^2(t)$ by different candidates ξ_1', ξ_1'' are always equivalent, allowing for an expression of the posterior measure as a well-defined density with respect to some reference measure. For instance, fix an initial state ξ_0 and two candidates ξ_1', ξ_1''. The likelihood of observing the interpolant x_t is given by the shifted Gaussian measures:

$$\mathrm{N}(\alpha(t)\xi_0 + \beta(t)\xi_1, \gamma^2(t)C), \quad \xi_1 = \xi_1', \xi_1''.$$

The Feldman-Hajek dichotomy states that two Gaussian measures are equivalent if and only if the difference in their means, $\xi_1' - \xi_1''$, is in H_C; otherwise they are mutually singular. Hence, if $\xi_1 \in H_C$ but $\xi_1' \notin H_C$ with positive probability, then the supports of these two measures are disjoint, preventing the construction of a meaningful posterior measure as a density with respect to any reference measure.

Building on the Lipshitz-continuity properties for the drift coefficient, established in Propositions 6 or 7, we can now establish the existence of solutions to the CB-SDE (3.9). We first treat the issue of existence.

Theorem 8. *Suppose that there exists some $\bar{t} \in (0,1]$ such that for each $t \in (0,\bar{t})$ and μ_0-almost every x_0, the mapping $x \mapsto f(t,x_0,x)$ is Lipschitz continuous in H_C norm, satisfying*

$$\|f(t,x_0,x) - f(t,x_0,y)\|_{H_C} \leq L(t)\|x-y\|_{H_C}, \text{ for all } x,y \in H.$$

for some function $L(t)$. If $L(t)$ is continuous on $(0,\bar{t})$ and integrable on $[0,\bar{t}]$, then there exists a strong solution to the CB-SDE (3.9) on the time interval $[0,\bar{t}]$.

Proof. The full proof is presented in Section A.5 in Appendix A. We allow for the possibility that $\lim_{t\to 0} L(t) = +\infty$ or $\lim_{t\to\bar{t}} L(t) = +\infty$ by only requiring an integrability condition for $L(t)$ on $[0,\bar{t}]$ which allows us to consider a deterministic time change $\hat{X}_t = X_{\theta(t)}$ (see Lemma **??** TODO) which leads to a finite Lipschitz constant on the new time domain.

We prove existence for the well-behaved time-changed stochatic process using a piecewise construction: we partition of the new time domain into a finite sequence of small intervals in such a way that the Banach fixed-point theorem yields the existencde of solutions on each subinterval. We then stitch these together to form a single, continuous strong solution for the time-changed process $\hat{X}_t$. The adaptedness of this solution is preserved throughout the iterative construction. Finally, we recover the strong solution X_t to the original CB-SDE by reversing the time-change. ■

Note that Banach's fixed point theorem does not guarantee uniqueness: the arguments we use in the proof only ensure uniqueness among solutions X_t where $X_t - \xi_0 - \sqrt{2\varepsilon}W_t \in H_C$. *A priori*, we cannot rule out other solutions to the CB-SDE (3.9) that do not satisfy this condition.

To help facilitate our proof to the uniqueness of strong solutions to the CB-SDE (3.9), we use an additional decoupling assumption which ensures independence of components of ξ_1 along the eigenvectors of the covariance operator C.

Theorem 9. *Let $\{e_n\}_{n=1}^{\infty}$ be an orthonormal basis of eigenvectors for the covariance operator C, and let H_N be the subspace of H_C spanned by $\{e_1, \dots, e_N\}$. We denote by P_N the orthogonal projection operator from H into H_N.*

Suppose that the distribution μ_1 of target data ξ_1 is such that the projections $\langle \xi_1, e_n \rangle$ are mutually independent random variables for different indices n. Then, under the same Lipschitz continuity conditions as in Theorem 8, the solution to the CB-SDE (3.9) is unique.

Proof. The full proof is presented in Section A.6 in Appendix A. The decoupling assumption on the components of ξ_1 allows us to employ a projection argument coupled with Gronwall's inequality to show that the norm of the difference between any two strong solutions driven by the same Wiener process is zero. ■

3.3 Parameterisation and Training Objective

We now detail our choice of parameterisation in learning an approximation to the drift $f(t, x_0, x)$ of the CB-SDE (3.9). We propose decomposing the drift into two distinct components: a *path velocity* φ and *denoiser* η:

$$f(t, x_0, x) = \varphi(t, x_0, x) + \left(\dot{\gamma}(t) - \frac{\varepsilon}{\gamma(t)} \right) \eta(t, x_0, x), \tag{3.10}$$

where

$$\begin{aligned} \varphi(t, x_0, x) &:= \mathbb{E}\left[\dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1 \,\middle|\, \xi_0 = x_0, x_t = x \right], \\ \eta(t, x_0, x) &:= \mathbb{E}[z \mid \xi_0 = x_0, x_t = x]. \end{aligned}$$

Hence, we decompose training into two learning objectives: one for φ and another for η. This decomposition is natural, as the stochastic interpolant x_t comprises a signal component $\alpha(t)\xi_0 + \beta(t)\xi_1$ and a noise component $\gamma(t)z$. The path velocity φ captures the deterministic path conditioned on the datapoints ξ_0, ξ_1, while the denoiser η controls the injection of stochasticity by the Wiener process.

An alternative approach involves learning the drift via a single objective: estimating the conditional expectation $\mathbb{E}[\xi_1 \mid \xi_0, x_t]$ only. This is justified by the following decomposition of the drift:

$$\begin{aligned} f(t,x_0,x) = & \left(\dot{\alpha}(t) - \alpha(t)\left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right)\right)x_0 \\ & + \left(\dot{\beta}(t) - \beta(t)\left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right)\right)\mathbb{E}[\xi_1 \mid \xi_0 = x_0, x_t = x] \\ & + \left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right)x_t. \end{aligned} \tag{3.11}$$

Although this is mathematically valid and useful for proving the theoretical bounds in Propositions 6 and 7, we find that this parameterisation exhibits far weaker empirical performance compared to the decomposition into path velocity and denoiser. We attribute this underperformance to two primary factors:

1. Inductive bias: our two-component approach provides a more effective inductive bias. The task of directly predicting the target ξ_1 given (ξ_0, x_t) is more complex than the two sub-problems of learning the path velocity and the denoiser. By simplifying the learning task, our decomposition helps the model find a better approximation of the overall drift.

2. Numerical stability: the alternative parameterisation suffers from severe instabilities due to the singularities at times $t = 0, 1$ that amplify approximation errors. This makes sampling unreliable at these critical times.

The crucial difference lies in the severity of the endpoint singularities. In our proposed parameterisation (Equation 3.10), the coefficient $\dot{\gamma}(t) - \frac{\varepsilon}{\gamma(t)}$ on the denoiser $\eta(t, x_0, x)$ is also singular at $t = 0, 1$. However this coefficient is integrable on $[0, 1]$ as long as $\frac{1}{\gamma(t)}$ is integrable. This property allows us to mitigate sampling instabilities by introducing a change of time (see Lemma **??** TODO).

In contrast, the coefficient in the alternative parameterisation (Equation 3.11) includes an additional $\frac{1}{\gamma(t)}$ factor which results in a stronger singularity that is non-integrable on $[0, 1]$ for any choice of $\gamma(t)$. This argument is formalised in Section A.7 in Appendix chapter A, and means that the alternative parameterisation is fundamentally less stable and cannot be resolved by a similar time-change technique.

Having established our choice in parameterising the drift $f(t,x_0,x)$ as a decomposition into a path velocity term $\varphi(t,x_0,x)$ and denoiser term $\eta(t,x_0,x)$, we introduce our loss functions. We will consider losses with respect to both the H-norm and the H_C-norm, and hence introduce the variable U as a Hilbert space representing either H or U. For approximations $\widetilde{\varphi}$ and $\widetilde{\eta}$, we define the *true path matching* (TPM) and *true denoiser matching* (TDM) objectives below:

$$\mathrm{TPM}_t(\widetilde{\varphi}) := \mathbb{E}\left[\|\widetilde{\varphi}(t,\xi_0,x_t) - \varphi(t,\xi_0,x_t)\|_U^2\right] \tag{3.12}$$

$$\mathrm{TDM}_t(\widetilde{\eta}) := \mathbb{E}\left[\|\widetilde{\eta}(t,\xi_0,x_t) - \eta(t,\xi_0,x_t)\|_U^2\right] \tag{3.13}$$

In practical terms, we do not have access to the ground-truth conditional expectations needed to calculate the TPM and TDM losses. Hence, we introduce two auxiliary losses, the *practical path matching* (PPM) and *practical denoiser matching* (PDM) objectives below:

$$\mathrm{PPM}_t(\widetilde{\varphi}) := \mathbb{E}\left[\left\|\widetilde{\varphi}(t,\xi_0,x_t) - (\dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1)\right\|_U^2\right] \tag{3.14}$$

$$\mathrm{PDM}_t(\widetilde{\eta}) := \mathbb{E}\left[\|\widetilde{\eta}(t,\xi_0,x_t) - z\|_U^2\right] \tag{3.15}$$

These losses are analogous to technique employed when training stochastic interpolants in finite dimensions (see Albergo et al., 2023a, Theorems 2.7–2.8), which makes the loss functions tractable by replacing the target conditional expectations with a sample of the underlying random variable to form a practical loss objective. However, in infinitte dimensions, the true matching objectives could finite while the practical objectives are not: special care must be taken to ensure that both sets of losses are finite. This is established in the next result.

Proposition 10. *Let U be the Hilbert space H in the definitions of the true objectives (Equations 3.12 and 3.13) and practical objectives (Equations 3.14 and 3.15). Given candidate approximations $\widetilde{\varphi}$ and $\widetilde{\eta}$ for which the TPM and TDM objectives are finite, the practical objectives* $\mathrm{PPM}_t(\widetilde{\varphi})$ *and* $\mathrm{PDM}_t(\widetilde{\varphi})$ *differ from* $\mathrm{TPM}_t(\widetilde{\varphi})$ *and* $\mathrm{TDM}_t(\widetilde{\varphi})$ *only by a finite constant for any $t \in (0,1)$.*

Furthermore, if U is instead the subspace H_C, the same result is true if the target data ξ_1 is supported on H_C and has finite second moment, that is, $\mathbb{E}\left[\|\mathbb{E}[\xi_1 \mid \xi_0, x_t] - \xi_1\|_{H_C}^2\right] < \infty$.

Proof. The full proof is given in Section A.8 in Appendix A. ■

Note that under both settings of Proposition 6 or Proposition 7, the target data ξ_1 is supported on H_C and has finite second moment. Hence, Proposition 10 shows that both the H-norm and H_C-norm are valid choices for our training objectives.

Having established conditions under which our training objectives are well-defined in infinite dimensions, we now quantify the quality of samples generated by simulating the CB-SDE (3.9) when the drift coefficient f is replaced by a learned drift $\widetilde{f}$, compared with the true conditional law $\mu_{t|0}(\mathrm{d}x_t\,,\xi_0)$.

Theorem 11. *TODO: result regarding Wasserstein-2 distance. This will be stated in U-norm $U = H$ or H_C This requires a result on Lipschitz continuity in U-norm. Note that all the proofs we had so far are in H_C-norm.*

We acknowledge a subtle but important distinction between our theoretical analysis and practical implementation. The Lipschitz continuity results established in Propositions 6 and 7 are derived with respect to the H_C-norm. An ideal training procedure would therefore employ loss functions measured in the H_C-norm, that is, apply Equations (3.14) and (3.15) with $U = H_C$, to align directly with these guarantees.

However, implementing such a loss is computationally demanding, as it requires access to the inverse covariance operator, C^{-1}. For tractability, we instead adopt the standard H-norm for our training objectives, which translates to standard mean-squared-error loss in implementation. A crucial consequence of this choice is that minimizing the loss in H-norm does not guarantee control on the corresponding loss in H_C-norm: one could have an arbitrarily small but positive H-norm loss that has an unbounded loss in H_C norm, since high-frequency components of the learned functions, by which we mean components in directions of eigenvectors of C^{-1} for which the corresponding eigenvalue is large, are strongly penalised in H_C-norm but not in H-norm. This observation suggests that a promising direction for future work is the inclusion of an explicit regularization term that penalizes high-frequency outputs when training with an H-norm loss.

TODO: note that the Lipschitz constant is finite on $[0,1)$ when $\gamma(t) = \sqrt{bt(1-t)}$.

3.4 Change of time

TODO. Integrability of $\frac{1}{\gamma(t)}$ implies $\exists$ of time change that makes coefficient on $\eta(t,\xi_0,x_t)$ not blow up on $[0,1]$ and hence

3.5 Algorithms

TODO. Describe the use of predictor-corrector.

Also consider the ODE and argue why the SDE is better ($\varepsilon > 0$ helps to control the injection of noise by the Wiener process and "correct" learning errors; can refer to the EDM paper for diffusion)

3.6 Backwards SDE

TODO: note that CB-SDE involves training two separate models for the forwrad/inverse tasks. In constrast, the MB-SDE allows for the same model to be used for both tasks.

Empirically, we find that there is only a marginal loss in performance, even though the MB-SDE does not guarantee a bridge to the *conditional* distribution. Suggests most information captured by x_t.

Chapter 4

Methodology and Results

Outline: use 1D darcy to experiment with different configurations. This uses an FNO. Results are bad in the 1D case but I'll try to argue that's okay because the point was to give informative design choices.

4.1 1D Dataset

Narrative for 1D goes as follows:

1. to choose roughness of noise, look at graph of relative L2 error when projecting 1D dataset onto the RKHS of C for RBF kernel of different length scales
2. compare this to empirical results when training on differnet length scales
3. show experiment with time re-weightings and argue that the time change $\theta(t) = t^2$ is best – it outperforms even the ease-in-out schedule which we attribute to *starvation* of the intermediate time steps
4. having establisehd $\theta(t) = t^2$ as the best time change, show experiment where t is sampled according to u^2 where $u \sim \mathscr{U}[0,1]$. This does not perform as well which we attribute to *model starvation* – not enough emphasis on crucial earlier time steps. hence we argue that the change-in-time is there primarily to mitigate the explosion in the coefficient on η and thus preventing amplification of training errors
5. compare with the "alternative parameterisation" where we only learn $\mathbb{E}[\xi_1 \mid \xi_0, x_t]$ and show this does not work well
6. hence for the expensive 2D datasets we go ahead with uniform time sampling during training, and learn φ, η

4.2 2D Dataset

Narrative for 2D datasets goes as follows

1. present results for different length scales
2. compare performance with FunDPS and vanilla stochatic interpolants baselines and show performance is very competitive with the former and (hopefully) far exceeds that of the latter
3. ablation: training the marginal model. performance not much worse, so it's useful in circumstances where we want to train less and are willing to give up some performance. likely due to diffusion paths being mainly driven by the conditioning on x_t not ξ_0
4. ablation: ODE. note that we can predict ξ_1 given ξ_0 using only one step. haven't yet done this experiment

References

Albergo, M. S., Boffi, N. M., and Vanden-Eijnden, E. (2023a). Stochastic interpolants: A unifying framework for flows and diffusions.

Albergo, M. S., Goldstein, M., Boffi, N. M., Ranganath, R., and Vanden-Eijnden, E. (2023b). Stochastic interpolants with data-dependent couplings. *arXiv preprint arXiv:2310.03725*.

Ames, W. F. and Pachpatte, B. (1997). *Inequalities for differential and integral equations*, volume 197. Elsevier.

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

Berger, M. S. (1977). *Nonlinearity and functional analysis: lectures on nonlinear problems in mathematical analysis*, volume 74. Academic press.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Bogachev, V., Prato, G. D., and Röckner, M. (2010). Uniqueness for solutions of fokker-planck equations on infinite dimensional spaces.

Bogachev, V. I. (1998). *Gaussian measures*. Number 62. American Mathematical Soc.

Brascamp, H. J. and Lieb, E. H. (1976). On extensions of the brunn-minkowski and prékopa-leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of functional analysis*, 22(4):366–389.

Da Prato, G. and Zabczyk, J. (2014). *Stochastic equations in infinite dimensions*, volume 152. Cambridge university press.

Dubins, L. E. and Schwarz, G. (1965). On continuous martingales. *Proceedings of the National Academy of Sciences*, 53(5):913–916.

Franzese, G., Rossi, S., Yang, L., Finamore, A., Rossi, D., Filippone, M., and Michiardi, P. (2023). How much is enough? a study on diffusion times in score-based generative models. *Entropy*, 25(4):633.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2021). Score-based generative modeling through stochastic differential equations.

Walnut, D. (2011). Vitali's convergence theorems. online.

Appendix A

Mathematical Proofs

A.1 Proof of Lemma 2

Lemma 2. *Let μ_t be the marginal distribution of the stochatic interpolant x_t, defined in Definition 1. For every $t \in [0,1]$, the measure μ_t satisfies the Fokker-Plank equation (3.5).*

Proof. It is sufficient to restrict our attention to any real-valued test function of the form $u(t,x) = \mathrm{Re}\left[\phi(t)e^{i\langle x,h(t)\rangle_H}\right]$ or $\mathrm{Im}\left[\phi(t)e^{i\langle x,h(t)\rangle_H}\right]$, where ϕ and h satisfy the properties given in Equation (3.6).

Fix $t \in [0,1]$ and consider the characteristic function of the real-valued random variable $u(t,x_t)$. For any $k \in \mathbb{R}$, we define

$$\chi(t,k) := \mathbb{E}\left[e^{iku(t,x_t)}\right] \tag{A.1}$$

Taking derivatives with respect to t and k and evaluating at $k=0$ allows us to compute the time derivative of the expected value of $u(t,x_t)$:

$$\frac{1}{i}\frac{\partial^2}{\partial t \partial k}\chi(t,k)\bigg|_{k=0} = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[u(t,x_t)\right] = \mathbb{E}\left[D_t u(t,x_t) + \langle \dot{x}_t, D_x u(t,x_t)\rangle_H\right]. \tag{A.2}$$

Since the inner product $\langle \dot{x}_t, D_x u(t,x_t)\rangle_H$ is linear in its first argument, we may apply the law of iterated expectations and replace $\dot{x}_t$ with $\zeta(t,x_t) = \mathbb{E}\left[\dot{x}_t \mid x_t\right]$ as defined in Equation (3.1):

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[u(t,x_t)\right] = \mathbb{E}\left[D_t u(t,x_t) + \langle \zeta(t,x_t), D_x u(t,x_t)\rangle_H\right]$$

Adding and subtracting $\frac{\varepsilon}{\gamma(t)}\eta(t,x_t)$, where $\eta(t,x_t) = \mathbb{E}[z \mid x_t]$ as defined in Equation (3.2), we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[u(t,x_t)] &= \mathbb{E}\left[D_t u(t,x_t) + \left\langle \frac{\varepsilon}{\gamma(t)}\eta(t,x_t) + \zeta(t,x_t) - \frac{\varepsilon}{\gamma(t)}\eta(t,x_t), D_x u(t,x_t)\right\rangle_H\right] \\ &= \frac{\varepsilon}{\gamma(t)}\mathbb{E}[\langle z, D_x u(t,x_t)\rangle_H] + \mathbb{E}[D_t u(t,x_t) + \langle f(t,x_t), D_x u(t,x_t)\rangle_H], \qquad \text{(A.3)}\end{aligned}$$

where we simplified the first term using the law of iterated expectations to simplify the first term, and substituted the definition $f(t,x) = \zeta(t,x) - \frac{\varepsilon}{\gamma(t)}\eta(t,x)$ given in Equation (3.4) for the second term.

For the following, we assume that $u(t,x) = \mathrm{Re}[\phi(t)e^{i\langle x,h(t)\rangle_H}]$, but an identical line of reasoning applies if $u(t,x) = \mathrm{Im}\left[\phi(t)e^{i\langle x,h(t)\rangle_H}\right]$.

Let us focus on the first term in Equation (A.3). We have:

$$\begin{aligned}\frac{\varepsilon}{\gamma(t)}\mathbb{E}[\langle z, D_x u(t,x_t)\rangle_H] &= \mathrm{Re}\left[i\frac{\varepsilon}{\gamma(t)}\mathbb{E}\left[\phi(t)e^{i\langle x_t,h(t)\rangle_H}\langle z,h(t)\rangle_H\right]\right] \\ &= \mathrm{Re}\left[i\frac{\varepsilon}{\gamma^2(t)}\mathbb{E}\left[\phi(t)e^{i\langle \alpha(t)\xi_0+\beta(t)\xi_1,h(t)\rangle_H}\right]\mathbb{E}\left[e^{i\langle\gamma(t)z,h(t)\rangle_H}\langle\gamma(t)z,h(t)\rangle_H\right]\right], \qquad \text{(A.4)}\end{aligned}$$

where the second line follows since $z \perp (\xi_0, \xi_1)$.

Let $\{\lambda_n, e_n\}_{n=1}^{\infty}$ be an orthonormal system for C (i.e. $Ce_n = \lambda e_n$ for each n) and define the scalar-valued functions $h_n(t) := \langle h(t), e_n\rangle_H$. The projections $z_n = \langle z, e_n\rangle$ for each n are mutually independent 1-dimensional Gaussians with zero mean and variances equal to λ_n. By Parseval's theorem, we have the identity $\langle\gamma(t)z,h(t)\rangle = \sum_{n=1}^{\infty}\gamma(t)h_n(t)z_n$. We may therefore write

$$\mathbb{E}\left[\langle\gamma(t)z,h(t)\rangle_H e^{i\langle\gamma(t)z,h(t)\rangle_H}\right] = \sum_{n=1}^{\infty}\mathbb{E}\left[\gamma(t)h_n(t)z_n e^{i\gamma(t)h_n(t)z_n}\right]\prod_{m\neq n}\mathbb{E}\left[e^{i\gamma(t)h_m(t)z_m}\right]$$

Using the identity $\mathbb{E}\left[qe^{iq}\right] = iv\mathbb{E}\left[e^{iq}\right]$ for a 1-dimensional Gaussian $v \sim \mathrm{N}(0,q)$, we have

$$\mathbb{E}\left[\langle\gamma(t)z,h(t)\rangle_H e^{i\langle\gamma(t)z,h(t)\rangle_H}\right] = \sum_{n=1}^{\infty} i\gamma^2(t)h_n^2(t)\lambda_n\mathbb{E}\left[e^{i\langle\gamma(t)z,h(t)\rangle_H}\right]$$

Substituting into Equation (A.4), we have

$$\frac{\varepsilon}{\gamma(t)}\mathbb{E}[\langle z, D_x u(t,x_t)\rangle_H] = \mathbb{E}\left[\sum_{n=1}^{\infty} -\varepsilon\lambda_n h_n^2(t)u(t,x_t)\right] = \mathbb{E}\left[\mathrm{Tr}\left(\varepsilon C D_x^2 u(t,x_t)\right)\right].$$

Finally, substituting this expression into Equation (A.3) and re-writing expectations via integrals, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_H u(t,x)\mu_t(\mathrm{d}x) = \int_H \mathrm{Tr}\left(\varepsilon C D_x^2 u(t,x)\right) + D_t u(t,x) + \langle f(t,x), D_x u(t,x)\rangle_H \,\mu_t(\mathrm{d}x).$$

Since the choice of t was arbitrary, it follows that μ_t satisfies the Fokker-Plank equation (3.5) for any $t \in [0,t]$. This concludes the proof. ■

A.2 Proof of Theorem 3

Theorem 3. *Let μ_t be the law of the stochastic interpolant x_t at time t.*

1. *Suppose that the MB-SDE (3.3) has solutions which are unique in law on a non-empty time interval $[0,\bar{t}] \subseteq [0,1]$. We denote the law of X_t by ρ_t.*
2. *Suppose that $\mathscr{L}E$ is dense in $L^1([0,\bar{t}] \times H, \nu)$, where ν is the measure on $[0,\bar{t}] \times H$ determined uniquely by*
$$\nu(\mathrm{d}(t,x)) = \nu_t(\mathrm{d}x)\,\mathrm{d}t\,,$$
and $\nu_t := \frac{1}{2}\rho_t + \frac{1}{2}\mu_t$ for each $t \in [0,\bar{t}]$.

Then, for $\mathrm{d}t$*-almost every $t \in [0,\bar{t}]$, we have*

$$\rho_t = \mu_t.$$

Proof. In addition to ν, we define the measures ρ and μ on the product space $[0,\bar{t}] \times H$ determined uniquely by $\rho(\mathrm{d}(t,x)) = \rho_t(\mathrm{d}x)\,\mathrm{d}t$ and $\mu(\mathrm{d}(t,x)) = \mu_t(\mathrm{d}x)\,\mathrm{d}t$. Hence, it follows by construction that $\nu = \frac{1}{2}\rho + \frac{1}{2}\nu$ and both $\rho$and μ are absolutely continuous with respect to ν. We define their densities p,q with respect to ν:

$$p(t,x) := \frac{\mathrm{d}\rho}{\mathrm{d}\nu} \quad \text{and} \quad q(t,x) = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}.$$

From Lemma 2 we know that both ρ_t and μ_t solve the Fokker-Plank equation (3.5). Hence,

$$0 = \int_{[0,\bar{t}]\times H} \mathscr{L}u(t,x)(p(t,x) - q(t,x))\nu(\mathrm{d}(t,x)) \tag{A.5}$$

for every test function $u \in E$. Note that for ν-almost every (t,x), we have $0 \le p(t,x), q(t,x) \le 2$, so their difference is bounded almost everywhere. Since Equation (A.5) holds for every

$u \in E$ and by assumption, $\mathscr{L}E$ is dense in $L^1([0,\bar{t}] \times H, \nu)$, it follows that

$$p(t,x) = q(t,x)$$

for ν-almost every (t,x). Hence, the signed measure $\rho - \mu = 0$ and $\rho_t = \mu_t$ for dt-almost every t. This concludes the proof. ■

A.3 Proof of Proposition 6

Proposition 6. *Suppose Hypothesis 5 holds. Then the map $x \mapsto f(t,x_0,x)$ is Lipschitz continuous with respect to the H_C-norm. Specifically, for each $t \in (0,1)$, $x_0 \in H_C$ and $x \in H$, the following inequality holds:*

$$\|f(t,x_0,x) - f(t,x_0,y)\|_{H_C} \leq L(t)\|x-y\|_{H_C},$$

where the Lipschitz constant $L(t)$ is:

$$L(t) = \max\left\{\left|\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right|, \left|\dot{\beta}(t) - \beta\left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right)\right|\frac{\beta(t)}{\beta^2(t)+k\gamma^2(t)}\right\}.$$

Proof. The proof proceeds in steps TODO

Step 0 First, we notice that the drift term can be re-written:

$$\begin{aligned} f(t,x_0,x) &= \mathbb{E}\left[\dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1 + \left(\dot{\gamma}(t) - \frac{\varepsilon}{\gamma(t)}\right)z \,\middle|\, \xi_0 = x_0, x_t = x\right] \\ &= \left(\dot{\alpha}(t) - \alpha(t)\left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right)\right)x_0 \\ &\quad + \left(\dot{\beta}(t) - \beta(t)\left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right)\right)\mathbb{E}[\xi_1 \mid \xi_0 = x_0, x_t = x] \\ &\quad + \left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right)x_t. \end{aligned} \tag{A.6}$$

Hence, if we can show that the mapping $x \mapsto \mathbb{E}[\xi_1 \mid \xi_0 = x_0, x_t = x]$ is Lipschitz continuous in H_C-norm, this translates to Lipschtiz continuity in the overall mapping $x \mapsto f(t,x_0,x)$.

Step 1 Let $\mu_{1|0,t}(\mathrm{d}\xi_1, x_0, x)$ denote the posterior law of ξ_1, conditional on $\xi_0 = x_0$ and $x_t = x$. Furthermore, let $\mathbb{P}_{1|0}(\mathrm{d}\xi_1, x_0)$ be the corresponding conditional prior, which is a well-defined Gaussian measure on H_C (see, e.g., Bogachev, 1998, Chapter 3.10). We use

$m_{1|0}(x_0)$ and $Q_{1|0}$ respectively to denote the mean and covariance operator of this Gaussian on H_C. Note that the prior conditional mean $m_{1|0}(x_0)$ is a linear function of x_0. Then for μ_0-almost every $x_0 \in H_C$, the law $\mu_{1|0,t}(\mathrm{d}\xi_1, x_0, x)$ is absolutely continuous with respect to the reference measure $\mathbb{P}_{1|0}(\mathrm{d}\xi_1, x_0)$ with the following density:

$$\frac{\mathrm{d}\mu_{1|0,t}(\cdot, x_0, x)}{\mathrm{d}\mathbb{P}_{1|0}(\cdot, x_0)}(\xi_1) = \frac{1}{Z_{1|0,t}(x_0,x)} \exp\big(-V_{1|0,t}(\xi_1, x_0, x)\big),$$
$$\text{where } V_{1|0,t}(\xi_1, x_0, x) := \frac{1}{2\gamma^2(t)} \|\alpha(t)x_0 + \beta(t)\xi_1 - x\|_{H_C}^2 + \Phi(x_0, \xi_1),$$

and $Z_{1|0,t}(x_0,x) := \int_{H_C} \exp\big(-V_{1|0,t}(\xi_1, x_0, x)\big)\mathbb{P}_{1|0}(\mathrm{d}\xi_1, x_0)$ is a normalising constant.

Step 2 Let $\{e_n\}_{n=1}^{\infty}$ be an orthonormal basis for H_C and for each $N \geq 1$, let H_N be the linear span of $\{e_1, \dots, e_N\}$. We define $\Pi_N : H_C \to H_N$ as the self-adjoint orthogonal projection operator onto the finite-dimensional subspace H_N of H_C and let $\xi_{1,N} := \Pi_N \xi_1$. Furthermore, we define a reference measure by projecting $\mathbb{P}_{1|0}$ onto this subspace:

$$\mathbb{P}_{1|0,N}(\mathrm{d}\xi_{1,N}, x_0) := \mathrm{N}(m_{1|0,N}(x_0), Q_N),$$
$$\text{where } m_{1|0,N}(x_0) := \Pi_N m_{1|0}(x_0),$$
$$\text{and } Q_N := \Pi_N Q_{1|0} \Pi_N.$$

Using this, we create a sequence of approximating posterior measures $\mu_{1|0,t,N}$ by restricting the potential to H_N: for each $\xi_{1,N} \in H_N$.

$$\frac{\mathrm{d}\mu_{1|0,t,N}(\cdot, x_0, x)}{\mathrm{d}\mathbb{P}_{1|0,N}(\cdot, x_0)}(\xi_{1,N}) := \frac{1}{Z_{1|0,t,N}(x_0,x)} \exp\big(-V_{1|0,t,N}(\xi_{1,N}, x_0, x)\big),$$
$$\text{where } V_{1|0,t,N}(\xi_{1,N}, x_0, x) := \frac{1}{2\gamma^2(t)} \big\|\alpha(t)\Pi_N x_0 + \beta(t)\xi_{1,N} - x\big\|_{H_C}^2 + \Phi(x_0, \xi_{1,N}),$$

where $Z_{1|0,t,N}(x_0,x) := \int_{H_N} \exp\big(-V_{1|0,t,N}\big)(\xi_{1,N}, x_0, x)\mathbb{P}_{1|0,N}(\mathrm{d}\xi_{1,N}, x_0)$ is a normalising constant.

Given these definitions, we study the following approximation of the posterior mean:

$$m_{1|0,t,N}(x_0,x) := \mathbb{E}_{\mu_{1|0,t,N}(\cdot, x_0, x)}[\xi_{1,N}] = \int_{H_N} \xi_{1,N} \mu_{1|0,t,N}(\mathrm{d}\xi_{1,N}, x_0, x). \tag{A.7}$$

We aim to find a Lipschitz constant for the map $x \mapsto m_{1|0,t,N}(x_0,x)$ that is independent of N and x_0. To do so, we consider the Frechet derivative of $m_{1|0,t,N}(x_0,x)$ with respect to x, applied in a direction $h \in H_C$. This is a covariance (see Lemma 12):

$$
\begin{aligned}
D_x m_{1|0,t,N}(x_0,x)[h] &= \frac{\beta(t)}{\gamma^2(t)}\,\mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\Big[(\xi_{1,N}-m_{1|0,t,N}(x_0,x))\,\big\langle \xi_{1,N}-m_{1|0,t,N}(x_0,x),h\big\rangle_{H_C}\Big]\\
&= \frac{\beta(t)}{\gamma^2(t)}\,\mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\Big[(\xi_{1,N}-m_{1|0,t,N}(x_0,x))\,\big\langle \xi_{1,N}-m_{1|0,t,N}(x_0,x),\Pi_N h\big\rangle_{H_N}\Big],
\end{aligned}
\tag{A.8}
$$

where the second equality follows from the first since the components of $\xi_{1,N}-m_{1|0,t,N}(x_0,x)$ along the basis vectors $\{e_n\}_{n=N+1}^{\infty}$ are all zero.

By the Riesz representation theorem, the N-dimensional subspace H_N is isomorphic with $\mathbb{R}^N$, so all vectors on H_N can be identified with an N-dimensional column vector in $\mathbb{R}^N$. We may therefore re-write the derivative using an N-dimensional covariance matrix C_N acting on the vector $\Pi_N h$:

$$
D_x m_{1|0,t,N}(x_0,x)[h] = \frac{\beta(t)}{\gamma^2(t)} C_N \Pi_N h, \tag{A.9}
$$
$$
\text{where } C_N = \mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\Big((\xi_{1,N}-m_{1|0,t,N}(x_0,x))(\xi_{1,N}-m_{1|0,t,N}(x_0,x))^{\mathsf{T}}\Big).
$$

For the rest of the proof, we identify C_N with a self-adjoint covariance operator on H_N.

Step 3 We now use the Brascamp-Lieb inequality (Brascamp and Lieb, 1976) to place a bound on the operator norm of C_N. We proceed by expressing the approximate posterior measure $\mu_{1|0,t,N}(\mathrm{d}\xi_{1,N},x_0,x)$ via a density relative to the Lebesgue measure on H_N (identified with $\mathbb{R}^N$). The density of the reference measure $\mathbb{P}_{1|0,N}(\mathrm{d}\xi_{1,N},x_0)$ with respect to the Lebesgue measure, evaluated at $\xi_{1,N}\in H_N$, is proportional to

$$
\exp\Big(-\frac{1}{2}\big\langle Q_N^{-1}(\xi_{1,N}-m_{1|0,N}(x_0)),\xi_{1,N}-m_{1|0,N}(x_0)\big\rangle_{H_N}\Big),
$$

where the inverse Q_N^{-1} is well-defined because $Q_N : H_N \to H_N$ is positive-definite and bounded. Hence,

$$
\begin{aligned}
&\mu_{1|0,t,N}(\mathrm{d}\xi_{1,N},x_0,x)\\
&\quad\propto \exp\Big(-V_{1|0,t,N}(\xi_{1,N},x_0,x)-\frac{1}{2}\big\langle Q_N^{-1}(\xi_{1,N}-m_{1|0,N}(x_0)),\xi_{1,N}-m_{1|0,N}(x_0)\big\rangle_{H_N}\Big)\,\mathrm{d}\xi_{1,N}.
\end{aligned}
$$

Let $W_{1|0,t,N}(\xi_{1,N},x_0,x) := V_{1|0,t,N}(\xi_{1,N},x_0,x)+\frac{1}{2}\big\langle Q_N^{-1}(\xi_{1,N}-m_{1|0,N}(x_0)),\xi_{1,N}-m_{1|0,N}(x_0)\big\rangle_{H_N}$ be the total potential with respect to the Lebesgue measure on H_N. Since this is twice-differentiable and strictly convex, the conditions for the Brascamp-Lieb inequality are satisfied (see (Brascamp and Lieb, 1976, Theorem 4.1)): for any continuously differentiable function $f : H_N \to \mathbb{R}$, we have

$$\begin{aligned}
&\mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\left[\left(f(\xi_{1,N})-\bar{f}\right)^2\right]\\
&\quad\le \mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\left[\left\langle \left(D^2_{\xi_{1,N}}W_{1|0,t,N}(\xi_{1,N},x_0,x)\right)^{-1}Df(\xi_{1,N}),Df(\xi_{1,N})\right\rangle_{H_N}\right],
\end{aligned}$$

where $\overline{f}$ is the expectation of $f(\xi_{1,N})$ under the measure $\mu_{1|0,t,N}(\mathrm{d}\xi_{1,N},x_0,x)$ and $D^2_{\xi_{1,N}}W_{1|0,t,N}(\xi_{1,N},x_0,x)$ is the inverse Hessian of $W_{1|0,t,N}(\xi_{1,N},x_0,x)$ with respect to $\xi_{1,N}$ on H_N. In the case where $f(\xi_{1,N}) = \left\langle \xi_{1,N},u\right\rangle_{H_N}$ for any $u\in H_N$, we have $Df(\xi_{1,N}) = u$, and

$$\begin{aligned}
&\mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\left[\left(f(\xi_{1,N})-\bar{f}\right)^2\right] = \langle C_N u,u\rangle\\
&\quad\le \mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\left[\left\langle \left(D^2_{\xi_{1,N}}W_{1|0,t,N}(\xi_{1,N},x_0,x)\right)^{-1}u,u\right\rangle_{H_N}\right]. \qquad \text{(A.10)}
\end{aligned}$$

Step 4 We now aim to place a Loewner order on the inverse Hessian $\left(D^2_{\xi_{1,N}}W_{1|0,t,N}(\xi_{1,N},x_0,x)\right)^{-1}$ irrespective of $\xi_{1,N}$, which will in turn allow us to form a Loewner order on C_N.

Taking the second-order Frechet derivatives of $W_{1|0,t,N}(\xi_{1,N},x_0,x)$ with respect to $\xi_{1,N}$ in the directions $u,v\in H_N$, we have

$$D^2_{\xi_{1,N}}W_{1|0,t,N}(\xi_N,x_0,x)[u,v] = \left\langle \left(\frac{\beta^2(t)}{\gamma^2(t)}I_N + \Pi_N\nabla^2_{\xi_1}\Phi(x_0,\xi_{1,N})\Pi_N + Q_N^{-1}\right)u,v\right\rangle_{H_N},$$

where $\nabla^2_{\xi_1}\Phi(\xi_0,\xi_1)$ is the partial Hessian of the potential Φ with respect to the second coordinate. This allows us to identify the Hessian with a self-adjoint Hessian operator from H_N to H_N:

$$D^2_{\xi_{1,N}}W_{1|0,t,N}(\xi_N,x_0,x)[u,v] = \frac{\beta^2(t)}{\gamma^2(t)}I_N + \Pi_N\nabla^2_{\xi_1}\Phi(x_0,\xi_{1,N})\Pi_N + Q_N^{-1} \qquad \text{(A.11)}$$

Since Φ is k-strongly convex, it is also k-strongly convex in the second coordinate and hence the projection of its partial Hessian satisfies the following Loewner order:

$$\Pi_N\nabla^2_{\xi_1}\Phi(x_0,\xi_{1,N}) \succcurlyeq kI_N,$$

which allows us to place a Loewner order on Equation (A.11):

$$D^2_{\xi_{1,N}} W_{1|0,t,N}(\xi_N, x_0, x)[u,v] \succcurlyeq \left(\frac{\beta^2(t)}{\gamma^2(t)} + k\right) I_N + Q_N^{-1}$$

Since the right-hand side of this quantity is positive-definite, this Loewner order is reversed when taking inverses:

$$\left(D^2_{\xi_{1,N}} W_{1|0,t,N}(\xi_N, x_0, x)[u,v]\right)^{-1} \preccurlyeq \left(\left(\frac{\beta^2(t)}{\gamma^2(t)} + k\right) I_N + Q_N^{-1}\right)^{-1}.$$

This relationship holds uniformly for all $\xi_{1,N} \in H_N$. Substituting into Equation (A.10), we have

$$\begin{aligned} \langle C_N u, u\rangle &\leq \left\langle \left(\left(\frac{\beta^2(t)}{\gamma^2(t)} + k\right) I_N + Q_N^{-1}\right)^{-1} u, u \right\rangle_{H_N}, \text{ for all } u \in H_N \\ \iff C_N &\preccurlyeq \left(\left(\frac{\beta^2(t)}{\gamma^2(t)} + k\right) I_N + Q_N^{-1}\right)^{-1}. \end{aligned}$$

Step 5 Having established a Loewner order on C_N, we now use this to place a bound on the operator norm of C_N. Since C_N is positive semi-definite, the Loewner order translates directly into an ordering on operator norms:

$$|||C_N||| \leq \left|\left|\left| \left(\left(\frac{\beta^2(t)}{\gamma^2(t)} + k\right) I_N + Q_N^{-1}\right)^{-1} \right|\right|\right|.$$

The spectrum of the operator $\left(\left(\frac{\beta^2(t)}{\gamma^2(t)} + k\right) I_N + Q_N^{-1}\right)^{-1}$ is given by the function $\sigma(\lambda) = \frac{\lambda\gamma^2(t)}{\lambda(\beta^2(t)+k\gamma^2(t))+\gamma^2(t)}$ evaluated over the spectrum of Q_N. This function is monotone and increasing for $\lambda \geq 0$, attaining its supremum at $\frac{\gamma^2(t)}{\beta^2(t)+k\gamma^2(t)}$. Hence, we have

$$|||C_N||| \leq \frac{\gamma^2(t)}{\beta^2(t) + k\gamma^2(t)}.$$

Substituting this relationship in Equation (A.9),

$$\left\|D_x m_{1|0,t,N}(x_0, x)[h]\right\|_{H_C} \leq \frac{\beta(t)}{\gamma^2(t)} |||C_N||| \, |||\Pi_N||| \, \|h\|_{H_C} \leq \frac{\beta(t)}{\beta^2(t) + k\gamma^2(t)} \|h\|_{H_C}.$$

It follows from the mean-value inequality (Berger, 1977, Theorem 2.1.19), that for any $x, y \in H$,

$$\begin{aligned}\left\|m_{1|0,t,N}(x_0,x) - m_{1|0,t,N}(x_0,y)\right\|_{H_C} &= \left\|m_{1|0,t,N}(x_0,x) - m_{1|0,t,N}(x_0,y)\right\|_{H_N} \\ &\leq \frac{\beta(t)}{\beta^2(t) + k\gamma^2(t)} \|x-y\|_{H_C}. \end{aligned} \tag{A.12}$$

Passing $N \to \infty$, the sequence of approximate posterior means $m_{1|0,t,N}(x_0,x)$ converges to the true posterior mean $m_{1|0,N}(x_0,x)$ (see Lemma 13). Since each approximation satisfies the inequality (A.12) that is uniform in N and the norm is a continuous mapping, the true posterior mean $m_{1|0,t}(x_0,x)$ also inherits the inequality.

$$\left\|m_{1|0,t}(x_0,x) - m_{1|0,t}(x_0,y)\right\|_{H_C} \leq \frac{\beta(t)}{\beta^2(t) + k\gamma^2(t)} \|x-y\|_{H_C}.$$

Step 7 We now substitute this relationship into the expression for the drift coefficient in Equation (A.6): a Lipschitz constant for the overall drift is the maximum of the Lipschitz constants for each term involving x_t:

$$\|f(t,x_0,x) - f(t,x_0,y)\|_{H_C} \leq L(t)\|x-y\|_{H_C},$$

where

$$L(t) = \max\left\{\left|\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right|, \left|\dot{\beta}(t) - \beta\left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}\right)\right| \frac{\beta(t)}{\beta^2(t) + k\gamma^2(t)}\right\}.$$

This concludes the proof. ∎

Lemma 12. *Let $m_{1|0,t,N}(x_0,x)$ be an approximate posterior mean as defined in Equation* (A.7), *with $t \in (0,1)$ and $N \geq 0$. Then the Frechet derivative of the mapping $x \mapsto m_{1\|0,t,N}(x_0,x)$ in H_C-norm, in a direction $h \in H_C$ is given by*

$$D_x m_{1|0,t,N}(x_0,x)[h] = \frac{\beta(t)}{\gamma^2(t)} \mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\left[(\xi_{1,N} - m_{1|0,t,N}(x_0,x))\left\langle \xi_{1,N} - m_{1|0,t,N}(x_0,x), h\right\rangle_{H_C}\right]$$

Proof. We begin by taking the Frechet derivative of $m_{1|0,t,N}(x_0,x)$ at x in a direction $h \in H_C$. Applying the quotient rule (see Berger, 1977, Chapter 2.1) and simplifying, we have

$$\begin{aligned} D_x m_{1|0,t,N}(x_0,x)[h] &= D_x \left\{ \frac{\int_{H_N} \xi_{1,N} \exp\big(-V_{1|0,t,N}(\xi_{1,N},x_0,x)\big)\mathbb{P}_{1|0,N}(\mathrm{d}\xi_{1,N},x_0)}{Z_{1|0,t,N}(x_0,x)} \right\}[h] \\ &= \frac{1}{Z_{1|0,t,N}(x_0,x)} D_x U_{1|0,t,N}(x_0,x)[h] - m_{1|0,t,N}(x_0,x)\frac{D_x Z_{1|0,t,N}(x_0,x)[h]}{Z_{1|0,t,N}(x_0,x)}, \end{aligned} \tag{A.13}$$

where we define $U_{1|0,t,N}(x_0,x) := \int_{H_N} \xi_{1,N} \exp\big(-V_{1|0,t,N}(\xi_{1,N},x_0,x)\big)\mathbb{P}_{1|0,N}(\mathrm{d}\xi_{1,N},x_0)$ to simplify notation. Evaluating the Frechet derivatives, we have

$$\begin{aligned} D_x U_{1|0,t,N}(x_0,x)[h] &= \frac{1}{\gamma^2(t)} \int_{H_N} \xi_{1,N} \left\langle \alpha(t)\Pi_N x_0 + \beta(t)\xi_{1,N} - x, h \right\rangle_{H_C} \\ &\qquad \cdot \exp\big(-V_{1|0,t,N}(\xi_{1,N},x_0,x)\big)\mathbb{P}_{1|0,N}(\mathrm{d}\xi_{1,N},x_0), \\ D_x Z_{1|0,t,N}(x_0,x)[h] &= \frac{1}{\gamma^2(t)} \int_{H_N} \left\langle \alpha(t)\Pi_N x_0 + \beta(t)\xi_{1,N} - x, h \right\rangle_{H_C} \\ &\qquad \cdot \exp\big(-V_{1|0,t,N}(\xi_{1,N},x_0,x)\big)\mathbb{P}_{1|0,N}(\mathrm{d}\xi_{1,N},x_0). \end{aligned}$$

Substituting these into Equation (A.13) and recognising that the fractions come together to form the approximate posterior density, we have:

$$D_x m_{1|0,t,N}(x_0,x)[h] = \frac{1}{\gamma^2(t)} \mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)} \Big[\big(\xi_{1,N} - m_{1|0,t,N}(x_0,x)\big) \left\langle \alpha(t)\Pi_N x_0 + \beta(t)\xi_{1,N} - x, h \right\rangle_{H_C} \Big].$$

Adding and subtracting zero,

$$0 = \frac{1}{\gamma^2(t)} \mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)} \Big[\big(\xi_{1,N} - m_{1|0,t,N}(x_0,x)\big) \left\langle -\alpha(t)\Pi_N x_0 + \beta(t) m_{1|0,t,N}(x_0,x) + x, h \right\rangle_{H_C} \Big],$$

we arrive at the expression

$$D_x m_{1|0,t,N}(x_0,x)[h] = \frac{\beta(t)}{\gamma^2(t)} \mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)} \Big[\big(\xi_{1,N} - m_{1|0,t,N}(x_0,x)\big) \left\langle \xi_{1,N} - m_{1|0,t,N}(x_0,x), h \right\rangle_{H_C} \Big].$$

This concludes the proof. ■

Lemma 13. *For every $x_0, x \in H$ and $t \in (0,1)$, the sequence of approximate posterior means $\{m_{1|0,t,N}(x_0,x)\}_{N=1}^{\infty}$ as defined in Equation* (A.7) *converges to the true posterior mean $m_{1|0,t}(x_0,x)$.*

Proof. First, let us re-express the definition of $m_{1|0,t,N}(x_0,x)$ by lifting the integrals into a common infinite-dimensional space:

$$m_{1|0,t,N}(x_0,x) = \int_{H_C} \Pi_N \xi_1 \frac{1}{Z_{1|0,t,N}(x_0,x)} \exp\bigl(-V_{1|0,t}(\Pi_N \xi_1, \Pi_N x_0, x)\bigr)\mathbb{P}_{1|0}(\mathrm{d}\xi_1\,, x_0), \tag{A.14}$$

where $Z_{1|0,t,N}(x_0,x) = \int_{H_C} V_{1|0,t}(\Pi_N \xi_1, \Pi_N x_0, x)\mathbb{P}_{1|0}(\mathrm{d}\xi_1\,, x_0)$.

We define the sequence of functions

$$f_N(\xi_1) := \Pi_N \xi_1 \frac{1}{Z_{1|0,t,N}(x_0,x)} \exp\bigl(-V_{1|0,t}(\Pi_N \xi_1, \Pi_N x_0, x)\bigr),$$

and

$$f(\xi_1) := \xi \frac{1}{Z_{1|0,t}(x_0,x)} \exp\bigl(-V_{1|0,t}(\xi_1, x_0, x)\bigr),$$

for fixed x_0 and x. To show convergence, we appeal to the Vitali convergence theorem (Walnut, 2011), which is a generalisation of the dominated convergence theorem and states that if the sequence of functions f_N is pointwise-convergent to f and uniformly integrable, then the integral of the functions also converges to the integral of f. We proceed in two steps: we first show pointwise convergence, and then show uniform integrability.

Step 1: Pointwise Convergence The numerator $\Pi_N \xi_1 \exp\bigl(-V_{1|0,t}(\Pi_N \xi_1, \Pi_N x_0, x)\bigr)$ is clearly pointwise convergent to $\xi_1 \exp\bigl(-V_{1|0,t}(\xi_1, x_0, x)\bigr)$ since for any fixed $\xi_1 \in H_C$, the projection $\Pi_N \xi_1$ converges to ξ_1 in H_C-norm, and $V_{1|0,t,x}$ is continuous in all of its inputs. Hence, it remains to show convergence of the sequence of normalising constants $Z_{1|0,t,N}(x_0,x)$.

To this end, we apply the dominated convergence theorem to show that

$$\lim_{N\to\infty} \int_{H_C} \exp\bigl(-V_{1|0}(\Pi_N \xi_1, \Pi_N x_0, x)\bigr)\mathbb{P}_{1|0}(x_0) = \lim_{N\to\infty} \int_{H_C} \exp\bigl(-V_{1|0}(\xi_1, x_0, x)\bigr)\mathbb{P}_{1|0}(\mathrm{d}\xi_1\,, x_0).$$

Since Φ is strongly convex, it has a unique global minimum. This implies that the integrand on both sides are bounded by a constant $M_1 < \infty$ that does not depend on N. Since the constant function is integrable on any probability space, it follows from the dominated convergence theorem that $\lim_{N\to\infty} Z_{1|0,t,N}(x_0,x) = Z_{1|0,t}(x_0,x)$.

Finally, since the normalising constant is nonzero for any N and converges to a non-zero value, the functions $f_N(\xi_1)$ are pointwise convergent to $f(\xi_1)$.

Step 2: Uniform Integrability A sufficient condition for uniform integrability is that there exists a uniform bound on the expected squared norm of sequence of the functions f_N (Billingsley, 2013, Theorem 3.5):

$$\int_{H_C} \|\Pi_N \xi_1\|_{H_C}^2 \frac{1}{Z_{1|0,t,N}^2(x_0,x)} \exp\big(-2V_{1|0,t}(\Pi_N \xi_1, \Pi_N x_0, x)\big)\mathbb{P}_{1|0}(\mathrm{d}\xi_1, x_0). \tag{A.15}$$

We will again employ the dominated convergence theorem to show that this sequence converges, and hence is bounded. First, pointwise convergence holds trivially since both the numerator and denominators converge, and the squared normalising factors $Z_{1|0,t,N}^2(x_0,x)$ are positive for all N and converge to a positive value. Furthermore, the integrand is uniformly bounded by a constant $\overline{M}$, since the strong convexity of Φ ensures that the potential grows at least quadratically as $\|\Pi_N \xi_1\|_{H_C} \to \infty$ and hence overwhelms the quadratic growth of the $\|\Pi_N \xi_1\|_{H_C}^2$ pre-factor.

The dominated convergence theroem therefore applies and it follows that the sequence of integrals in Equation (A.15) is convergent and therefore bounded. Hence, the sequence of functions f_N is uniformly integrable.

Since we have shown that the sequence of functions f_N is pointwise convergent and uniformly integrable, it follows that their integrals, which are equal to the approximate posterior means $m_{1|0,t,N}(x_0,x)$, are convergent and converge to the true posterior mean $m_{1|0,t}(x_0,x)$. ■

A.4 Proof of Proposition 7

Proposition 7. *Suppose the law μ_1 of the target data ξ_1 is supported on a bounded subset of H_C, that is, there exists a scalar $R < \infty$ where $\|\xi_1\|_{H_C} < R$, μ_1-almost surely. Then the map $x \mapsto f(t,x_0,x)$ is Lipschitz continuous with respect to the H_C-norm. Specifically, for each $t \in (0,1)$ and $x_0, x \in H$, the following inequality holds:*

$$\|f(t,x_0,x) - f(t,x_0,y)\|_{H_C} \leq L(t)\|x-y\|_{H_C},$$

where the Lipschitz constant $L(t)$ is:

$$L(t) = \max\left\{ \left| \frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)} \right|, \left| \dot{\beta}(t) - \beta\left(\frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)} \right) \right| \frac{R^2 \beta(t)}{\gamma^2(t)} \right\}.$$

Proof. Our proof follows a similar overarching argument to to proof of Proposition 6 in Section A.3: again, the expression Equation (A.6) means it is sufficient to consider Lipschitz continuity of the mapping $x \mapsto \mathbb{E}[\xi_1 \mid \xi_0 = x_0, x_t = x]$. As before, we find a bound for the expression for the Frechet derivative of the posterior mean, expressed as a covariance. The assumption of bounded support in H_C-norm allows us to greatly simplify our arguments, meaning that we no longer require a Galerkin-type projection argument and directly provide our proof in infinite dimensions.

As in Section A.3, we let $\mu_{1|0,t}(\mathrm{d}\xi_1, x_0, x)$ denote the posterior law of ξ_1, conditional on $\xi_0 = x_0$ and $x_t = x$. This time however, for each $t \in (0,1)$ we let the reference measure be $\mathbb{P}_t := \mathrm{N}(\alpha(t)\xi_0, \gamma^2(t)C)$. Note that the Cameron-Martin space of $\gamma^2(t)C$ is identical to that of C, equipped with an inner product scaled by $\frac{1}{\gamma^2(t)}$. Since $\beta(t)\xi_1$ is almost-surely in H_C, and hence also the Cameron-Martin space of $\gamma^2(t)C$, $H_{\gamma^2(t)C}$, the measure $\mu_{1|0,t}(\mathrm{d}\xi_1, x_0, x)$ is absolutely continuous with respect to $\mathbb{P}_t$:

$$\frac{\mathrm{d}\mu_{1|0,t}(\cdot, x_0, x)}{\mathrm{d}\mathbb{P}_t}(\xi_1) = \frac{1}{Z_{1|0,t}(x_0,x)} \exp\big(-V_{1|0,t}(\xi_1, x_0, x)\big),$$
$$\text{where } V_{1|0,t}(\xi_1, x_0, x) = \frac{1}{\gamma^2(t)} \|\alpha(t)x_0 + \beta(t)\xi_1 - x\|_{H_C}^2,$$

and $Z_{1|0,t}(x_0,x) := \int_{H_C} \exp\big(-V_{1|0,t}(\xi_1, x_0, x)\big)\mathbb{P}_t(\mathrm{d}\xi_1)$ is a normalising constant. We define $m_t(x_0,x)$ as the posterior mean:

$$m_t(x_0,x) := \mathbb{E}_{\mu_{1|0,t}(\cdot,x_0,x)}[\xi_1] = \int_{H_C} \xi_1 \mu_{1|0,t}(\mathrm{d}\xi_1, x_0, x).$$

Following an approach analogous to that given in the proof to Lemma 12, we take the Frechet derivative in the direction $h \in H_C$ and again arrive at a covariance:

$$D_x m_t(x_0,x)[h] = \frac{\beta(t)}{\gamma^2(t)} \mathbb{E}_{\mu_{1|0,t}(\cdot,x_0,x)}\Big[(\xi_1 - m_t(x_0,x))\,\langle \xi_1 - m_t(x_0,x), h\rangle_{H_C}\Big]$$

Taking the norm in H_C and applying the Cauchy-Schwarz inequality, we have

$$\|D_x m_t(x_0,x)[h]\|_{H_C} \le \frac{\beta(t)}{\gamma^2(t)} \mathbb{E}_{\mu_{1|0,t}(\cdot,x_0,x)}\Big[\|\xi_1 - m_t(x_0,x)\|_{H_C}^2\Big] \|h\|_{H_C}$$

Using the fact that $0 \le \mathbb{E}\Big[\|\xi_1 - m_t(x_0,x)\|_{H_C}^2\Big] = \mathbb{E}\Big[\|\xi_1\|_{H_C}^2\Big] - \|m_t(x_0,x)\|^2$ and $\big\|\xi_1^2\big\|_{H_C} \le R^2$ almost surely, we conclude

$$\|D_x m_t(x_0,x)[h]\|_{H_C} \le \frac{R^2\beta(t)}{\gamma^2(t)} \|h\|_{H_C}.$$

Finally, we apply the mean-value inequality (Berger, 1977, Theorem 2.1.19) and conclude that $m_t(x_0,x)$ is Lipschitz in H_C-norm with Lipschitz constant at most $\frac{R^2\beta(t)}{\gamma^2(t)}$:

$$\|m_t(x_0,x) - m_t(x_0,y)\|_{H_C} \leq \frac{R^2\beta(t)}{\gamma^2(t)}\|x-y\|_{H_C}.$$

Substituting this into Equation (A.6) gives the Lipschitz constant for the overall mapping $x \mapsto f(t,x_0,x)$. This concludes the proof. ■

A.5 Proof of Theorem 8

Theorem 8. *Suppose that there exists some $\bar{t} \in (0,1]$ such that for each $t \in (0,\bar{t})$ and μ_0-almost every x_0, the mapping $x \mapsto f(t,x_0,x)$ is Lipschitz continuous in H_C norm, satisfying*

$$\|f(t,x_0,x) - f(t,x_0,y)\|_{H_C} \leq L(t)\|x-y\|_{H_C}, \textit{ for all } x,y \in H.$$

for some function $L(t)$. If $L(t)$ is continuous on $(0,\bar{t})$ and integrable on $[0,\bar{t}]$, then there exists a strong solution to the CB-SDE (3.9) on the time interval $[0,\bar{t}]$.

Proof. We allow for the possibility that $\lim_{t\to 0} L(t) = +\infty$ or $\lim_{t\to\bar{t}} L(t) = +\infty$ by only requiring an integrability condition for $L(t)$ on $[0,\bar{t}]$. Without loss of generality, we may assume that $L(t) > 0$ on $[0,\bar{t}]$ and hence by **??** there exists some monotonic function $\theta \in C^1([0,\bar{t}])$ for which there exists $\hat{\bar{t}} \in (0,1]$ satisfing $\theta(s) \geq 0, \dot{\theta}(s) \geq 0$ for all $s \in [0,\hat{\bar{t}}]$ and $\theta(\hat{\bar{t}}) = \bar{t}$ such that the function $L(\theta(t))\dot{\theta}(t)$ is continuous on $[0,\hat{\bar{t}}]$.

The existence of such a function $\theta(t)$ allows us to consider a time-changed stochastic process $\hat{X}_t$ on $[0,\hat{\bar{t}}]$, defined by

$$\hat{X}_t := X_{\theta(t)}, \quad \hat{X}_0 = \xi_0,$$

such that any possible singularities in $L(t)$ at $t=0$ are eliminated: $\lim_{t\to 0} L(\theta(t))\dot{\theta}(t) < \infty$.

Throughout, fix any C-Wiener process W_t on $[0,\bar{t}]$ and let $\mathbb{F} = \{\mathscr{F}_t\}_{t\in[0,\bar{t}]}$ be the smallest normal filtration generated by (ξ_0, W_t).

Step 1: Time-Changed Stochastic Process From the Dambis-Dubins-Scharz theorem (Dubins and Schwarz, 1965), there exists a stochastic process $\hat{W}_t$ on $[0,\hat{\bar{t}}]$ such that $\hat{W}_t = W_{\theta(t)}$ and $\hat{W}_t$ is a C-Wiener process with respect to the filtration $\hat{\mathbb{F}} = \left\{\mathscr{F}_{\theta(t)}\right\}_{t\in[0,\hat{\bar{t}}]}$. Furthermore, $\hat{\mathbb{F}}$ is the smallest normal filtration generated by $(\xi_0, \hat{W}_t)$. The time-changed stochastic process

$\hat{X}_t$ is then governed by the following SDE:

$$\mathrm{d}\hat{X}_t = f(\theta(t), \xi_0, \hat{X}_t)\dot{\theta}(t)\,\mathrm{d}t + \sqrt{2\varepsilon\dot{\theta}(t)}\,\mathrm{d}\hat{W}_t\,. \tag{A.16}$$

Let $\hat{f}(t,x_0,x) := f(\theta(t),x_0,x)\dot{\theta}(t)$ denote the drift for this time-changed stochastic process. Note that at time $t = 0$, the drift $\hat{f}(t,x_0,x)$ is well-defined only when $x_0 = x$, but this is guaranteed by the definition $\hat{X}_0 = \xi_0$. By construction, the mapping $x \mapsto \hat{f}(\theta(t),x_0,x)$ is Lipschitz-continuous and has a Lipschitz constant at most

$$\hat{L}(t) = L(\theta(t))\dot{\theta}(t), \tag{A.17}$$

which is continuous and finite for each $t \in [0,\hat{\hat{t}}]$.

Hence, it is possible to create a finite partition $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_k < \cdots < \tau_K = \hat{\hat{t}}$ of $[0,\hat{\hat{t}}]$ with $K < \infty$ such that

$$q_k := (\tau_k - \tau_{k-1}) \sup_{t\in[\tau_{k-1},\tau_k]} \hat{L}(t) < 1, \qquad \text{for all } k = 1,\dots,K.$$

Step 2: Existence of Strong Solutions For each $k = 1,\dots,K$, consider the Banach space B_k of all continuous, H_C-valued functions on $[\tau_{k-1},\tau_k]$ equipped with the following norm:

$$\|Y\|_{B_k} := \sup_{t\in[\tau_{k-1},\tau_k]} \|Y(t)\|_{H_C}.$$

To argue existence of a strong solution to the CB-SDE on $[0,\hat{\hat{t}}]$, we will apply Banach's fixed point theorem inductively and piecewise on the intervals $[\tau_{k-1},\tau_k]$ and pathwise for all events ω in the sample space Ω, to build a solution $\hat{X}_t$ to Equation (A.16).

Fix any event $\omega \in \Omega$, so that $\xi_0(\omega)$ and $\hat{W}_t(\omega)$ are respectively the outcomes of the random variable ξ_0 and the Wiener process at time t, and define $\hat{X}_0(\omega) := \xi_0(\omega)$. Furthermore, let

$$\widetilde{W}_{k,t} := \int_{\tau_{k-1}}^{t} \sqrt{2\varepsilon\dot{\theta}(s)}\,\mathrm{d}\hat{W}_s\,.$$

We proceed by induction: for each $k = 1,\dots,K$, having defined $\hat{X}_{\tau_{k-1}}(\omega)$, we define the mapping $\Psi_{k,\omega} : B_k \to B_k$ as follows. For any $Y \in B_k$,

$$(\Psi_{k,\omega}Y)(t) = \int_{\tau_{k-1}}^{t} \hat{f}\Big(s, \xi_0(\omega), \hat{X}_{\tau_{k-1}}(\omega) + \widetilde{W}_{k,s}(\omega) + Y(s)\Big)\,\mathrm{d}s\,, \qquad \text{for all } t \in [\tau_{k-1},\tau_k]. \tag{A.18}$$

For any $Y, Y' \in B_k$, we have

$$\begin{aligned}
\left\|\Psi_{k,\omega}Y - \Psi_{k,\omega}Y'\right\|_{B_k} &= \sup_{t\in[\tau_{k-1},\tau]} \left\|(\Psi_{k,\omega}Y - \Psi_{k,\omega}Y')(t)\right\|_{H_C} \\
&\le \int_{\tau_{k-1}}^{\tau_k} \left\|\hat{f}\Big(s, \xi_0(\omega), \hat{X}_{\tau_{k-1}}(\omega) + \widetilde{W}_{k,s}(\omega) + Y(s)\Big) - \hat{f}\Big(s, \xi_0(\omega), \hat{X}_{\tau_{k-1}}(\omega) + \widetilde{W}_{k,s}(\omega) + Y'(s)\Big)\right\|_{H_C} \mathrm{d}s \\
&\le (\tau_k - \tau_{k-1}) \sup_{t\in[\tau_{k-1},\tau]} \left[\hat{L}(t)\left\|Y(t) - Y'(t)\right\|_{H_C}\right] \\
&\le (\tau_k - \tau_{k-1}) \sup_{t\in[\tau_{k-1},\tau]} \hat{L}(t) \sup_{t\in[\tau_{k-1},\tau]} \left\|Y(t) - Y'(t)\right\|_{H_C} \\
&= q_k \left\|Y - Y'\right\|_{B_k},
\end{aligned}$$

where $q_k < 1$ by construction of the interval. By Banach's fixed point theorem, it follows that there exists a unique $Y^* \in B_k$ such that $\Psi_{k,\omega}Y^* = Y^*$.

For every $t \in [\tau_{k-1}, \tau_k]$, we let $\hat{X}_t(\omega) := \hat{X}_{\tau_{k-1}}(\omega) + \widetilde{W}_{k,t}(\omega) + Y^*(t)$ for all $t \in [\tau_{k-1}, \tau_k]$. Substituting this definition into the fixed point identity $\Psi_{k,\omega}Y^* = Y^*$, we have

$$\begin{aligned}
\hat{X}_t(\omega) - \hat{X}_{\tau_{k-1}}(\omega) - \widetilde{W}_{k,t}(\omega) &= \int_{\tau_{k-1}}^{t} \hat{f}(s, \xi_0(\omega), \hat{X}_s(\omega))\,\mathrm{d}s \\
\implies \hat{X}_t(\omega) &= \hat{X}_{\tau_{k-1}}(\omega) + \int_{\tau_{k-1}}^{\tau_k} \hat{f}(s, \xi_0(\omega)\hat{X}_s(\omega))\,\mathrm{d}s + \int_{\tau_{k-1}}^{t} \sqrt{2\varepsilon\dot{\theta}(t)}\,\mathrm{d}\hat{W}(\omega),
\end{aligned}$$

which is the integral form of the time-changed CB-SDE (A.16), expressed pathwise with the chosen probability event $\omega \in \Omega$ and defined on the interval $t \in [\tau_{k-1}, \tau_k]$.

Since ω was chosen arbitrarily, we may repeat this process for every $\omega \in \Omega$ to build a stochatic process $\hat{X}_t$ on the interval $t \in [\tau_{k-1}, \tau_k]$. Now that we have a definition of $\hat{X}_{\tau_k}(\omega)$, we may repeat the inductive step for $k \leftarrow k+1$. This builds a stochatic process $\hat{X}_t$ on the entire desired interval $t \in [0, \hat{\bar{t}}]$.

It remains to check that $\hat{X}_t$ is $\hat{\mathbb{F}}$-adapted on $[0, \hat{\bar{t}}]$. Again, employing induction, we may observe that $X_0 = \xi_0$ is by definition $\mathscr{F}_0$-measurable. Then, for each $k = 1, \ldots, K$, we are given that $X_{\tau_{k-1}}$ is $\mathscr{F}_{\theta(\tau_{k-1})}$-measurable. We can view every contraction-mapping iteration as if it were applied for all $\omega \in \Omega$ simultaneously. Suppose the initial guesses $Y_\omega \in B_k$ are such that $Y_\omega(t)$ is $\mathscr{F}_{\theta(t)}$-measurable as a function of ω, for all $t \in [\tau_{k-1}, \tau_k]$. Each application of the contraction mapping, $(\Psi_{k,\omega}(Y_\omega))(t)$, is also $\mathscr{F}_{\theta(t)}$-measurable as a function of ω, since the integrand in Equation (A.18) is the composition of a continuous function with a $\mathscr{F}_{\theta(t)}$-measurable function. Hence, every time we perform a Banach iteration, the outcome at time $t \in [\tau_{k-1}, \tau]$ is $\mathscr{F}_{\theta(t)}$-measurable. Since σ-fields are closed under countable pointwise limits, it follows that $Y^*_\omega(t)$ and thus $\hat{X}_t(\omega)$ are $\mathscr{F}_{\theta(t)}$-measurable for all $t \in [\tau_{k-1}, \tau_k]$. Repeating the induction for all steps up to $k = K$ ensures that $\hat{X}_t(\omega)$ is $\mathscr{F}^*_{\theta(t)}$-measurable for all $t \in [0,\hat{\bar{t}}]$ and hence $\hat{X}_t$ is $\hat{\mathbb{F}}$-adapted on $[0,\hat{\bar{t}}]$.

Since the time change $\theta(t)$ is a bijection between $[0,\bar{t}]$ and $[0,\hat{\bar{t}}]$, we recover the stochastic process X_t on $[0,\bar{t}]$ by reversing the time change: $X_t := \hat{X}_{\theta^{-1}(t)}$. This process is by construction $\mathbb{F}$-adapted. We have therefore shown the existence of a strong solution to the CB-SDE (3.9) on $[0,\bar{t}]$. This concludes the proof. ■

A.6 Proof of Theorem 9

Theorem 9. *Let $\{e_n\}_{n=1}^{\infty}$ be an orthonormal basis of eigenvectors for the covariance operator C, and let H_N be the subspace of H_C spanned by $\{e_1,\dots,e_N\}$. We denote by P_N the orthogonal projection operator from H into H_N.*

Suppose that the distribution μ_1 of target data ξ_1 is such that the projections $\langle \xi_1, e_n \rangle$ are mutually independent random variables for different indices n. Then, under the same Lipschitz continuity conditions as in Theorem 8, the solution to the CB-SDE (3.9) is unique.

Proof. As in the proof in Section A.5 for the existence of strong solutions (Theorem 8), it helps to first consider uniqueness of strong solutions for the time-changed CB-SDE (A.16). Let $\hat{X}_t$ and $\hat{X}'_t$ be two strong solutions for the same initial condition, $\hat{X}_0 = \hat{X}'_0 = \xi_0$ and driven by the same Wiener process $\hat{W}_t$ on $[0,\hat{\bar{t}}]$. *A priori*, it is not guaranteed that $\|\hat{X}_t - \hat{X}'_t\|_{H_C} < \infty$ since $\hat{X}_t - \hat{X}'_t$ may not be in H_C. However, for each $N \geq 1$, it is guaranteed that the projected diference $P_N(\hat{X}_t - \hat{X}'_t) \in H_C$ since the range of P_N is by definition a subspace of H_C due to C being a positive-definite operator. It therefore holds that

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} P_N\big(\hat{X}_t - \hat{X}'_t\big) &= P_N\big(\hat{f}(t,\xi_0,\hat{X}_t) - \hat{f}(t,\xi_0,\hat{X}'_t)\big) \\
\implies \frac{\mathrm{d}}{\mathrm{d}t} \big\|P_N\big(\hat{X}_t - \hat{X}'_t\big)\big\|_{H_C} &\leq \big\|P_N\big(\hat{f}(t,\xi_0,\hat{X}_t) - \hat{f}(t,\xi_0,\hat{X}'_t)\big)\big\|_{H_C} \\
&\leq \hat{L}(t)\big\|P_N\big(\hat{X}_t - \hat{X}'_t\big)\big\|_{H_C},
\end{aligned}
$$

where $\hat{L}(t)$ is as defined in Equation (A.17).

We now apply Groenwall's inequality (Ames and Pachpatte, 1997, Theorem 1.2.2) to the quantity $\big\|P_N\big(\hat{X}_t - \hat{X}'_t\big)\big\|_{H_C}$ as a function of t: since $\hat{L}(t)$ is real-valued and continuous on $[0,\hat{\bar{t}}]$, we have:

$$
\big\|P_N\big(\hat{X}_t - \hat{X}'_t\big)\big\|_{H_C} \leq \big\|P_N\big(\hat{X}_0 - \hat{X}'_0\big)\big\|_{H_C} \exp\left(\int_0^{\hat{\bar{t}}} \hat{L}(t)\big\|P_N\big(\hat{X}_t - \hat{X}'_t\big)\big\|_{H_C}\right).
$$

Since by definition $\hat{X}_0 = \hat{X}'_0 = \xi_0$, so $\hat{X}_0 - \hat{X}'_0 = 0$, it follows that

$$\left\| P_N \left(\hat{X}_t - \hat{X}'_t \right) \right\|_{H_C} = 0,$$

for all $t \in [0, \hat{\bar{t}}]$. Since this equality is true for every $N \geq 1$, we pass $N \to \infty$. It follows that $\left\| \hat{X}_t - \hat{X}'_t \right\|_{H_C} = 0$ and therefore

$$\hat{X}_t = \hat{X}'_t.$$

Hence, there is a unique solution to the time-changed CB-SDE (A.16). By inverting the time change, $X_t := \hat{X}_{\theta^{-1}(t)}$, it follows that the CB-SDE (3.9) has a unique strong solution on $[0, \bar{t}]$. This concludes the proof. ■

A.7 Non-integrability of coefficients

In this section, we argue formally why the coefficients in the alternative parameterisation (Equation 3.11) are not integrable on $[0,1]$. We make use of the following lemma.

Lemma 14. *For any $\varepsilon \geq 0$, there exists no function $\gamma : [0,1] : \mathbb{R}_{\geq 0}$ that is continuous on $[0,1]$, continuously differentiable on $(0,1)$, and satisfies the boundary conditions $\gamma(0) = \gamma(1) = 0$ and $\gamma(t) > 0$ for all $t \in (0,1)$, for which the function*

$$c(t) := \frac{\dot{\gamma}(t)}{\gamma(t)} - \frac{\varepsilon}{\gamma^2(t)}$$

is integrable on $[0,1]$.

Proof. We consider the function $y(t) := \gamma^2(t)$, which satisfies $y(0) = y(1) = 0$ and $\dot{y}(t) = 2\dot{\gamma}(t)\gamma(t)$. The function $c(t)$ can be re-written in terms of $y(t)$ as

$$c(t) = \frac{\dot{y}(t) - 2\varepsilon}{2y(t)}.$$

Consider the improper integral

$$\begin{aligned} I_- &= \lim_{a \to 0^+} \int_a^{\frac{1}{2}} c(t)\, \mathrm{d}t \\ &= \frac{1}{2} \log y\left(\frac{1}{2}\right) - \lim_{a \to 0^+} \left[\frac{1}{2} \log y(a) + \varepsilon \int_a^{\frac{1}{2}} \frac{1}{y(t)}\, \mathrm{d}t \right] \end{aligned}$$

A necessary condition for I_- to converge to a finite number is that $\varepsilon > 0$. In this case, it is necessary that

$$-1 = \lim_{a\to 0^-} \frac{\log y(a)}{2\varepsilon \int_a^{\frac{1}{2}} \frac{1}{y(t)}\,\mathrm{d}t} = \lim_{a\to 0^-} -\frac{\dot{y}(a)}{2\varepsilon} \iff \dot{y}(0) = 2\varepsilon.$$

A similar analysis for the integral $I_+ := \lim\limits_{a\to 1^-} \int_{\frac{1}{2}}^{a} c(t)$ shows that it is necessary that $\dot{y}(1) = -2\varepsilon$.

Taken together, these conditions imply that there exists some function $h(t)$ differentiable on $(0,1)$ satisfying the boundary conditions $h(0) = 2\varepsilon$ and $h(1) = -2\varepsilon$, such that

$$y(t) = t(1-t)h(t).$$

Substituting this into the definition of $c(t)$, we have

$$I_+ = \int_{\frac{1}{2}}^{1} c(t)\,\mathrm{d}t = \int_{\frac{1}{2}}^{1} \frac{(1-2t)h(t) - 2\varepsilon}{2t(1-t)h(t)}\,\mathrm{d}t\,.$$

In the limit as $t \to 1^-$, the integrand converges to $-\infty$ and satisfies the following limit:

$$\lim_{t\to 1^-} \frac{-\left[\frac{(1-2t)h(t)-2\varepsilon}{2t(1-t)h(t)}\right]}{\frac{1}{1-t}} = 1.$$

Hence, the integral I_+ converges if and only if the integral $\int_{\frac{1}{2}}^{1} \frac{1}{1-t}\,\mathrm{d}t$ converges. Since this integral does not converge, we conclude that I_+ does not converge and hence $c(t)$ is not integrable for any permissible choice of γ. ■

We have established that the function $c(t)$ is not integrable on $[0,1]$. We can now argue that the coefficient $\dot{\beta}(t) - \beta(t)c(t)$ on the expectation $\mathbb{E}[\xi_1 \mid \xi_0, x_t]$ in the alternative parameterisation (Equation 3.11) is not integrable on $[0,1]$. The singularity as $t \to 1^-$, which we have shown is of the order $\frac{1}{1-t}$, is not avoided in the coefficient $\dot{\beta}(t) - \beta(t)c(t)$. This is because $\beta(1) = 1$ by definition, and $\dot{\beta}$ is bounded on $[0,1]$ due to the continuous differentiability of β. Hence, the coefficient $\dot{\beta}(t) - \beta(t)c(t)$ has a singularity of order $\frac{1}{1-t}$ as $t \to 1^-$ and is not integrable on $[0,1]$.

A.8 Proof of Proposition 10

Proposition 10. *Let U be the Hilbert space H in the definitions of the true objectives (Equations 3.12 and 3.13) and practical objectives (Equations 3.14 and 3.15). Given candidate approximations $\widetilde{\varphi}$ and $\widetilde{\eta}$ for which the TPM and TDM objectives are finite, the practical objectives* $\mathrm{PPM}_t(\widetilde{\varphi})$ *and* $\mathrm{PDM}_t(\widetilde{\varphi})$ *differ from* $\mathrm{TPM}_t(\widetilde{\varphi})$ *and* $\mathrm{TDM}_t(\widetilde{\varphi})$ *only by a finite constant for any $t \in (0,1)$.*

Furthermore, if U is instead the subspace H_C, the same result is true if the target data ξ_1 is supported on H_C and has finite second moment, that is, $\mathbb{E}\left[\|\mathbb{E}[\xi_1 \mid \xi_0, x_t] - \xi_1\|_{H_C}^2\right] < \infty$.

Proof. We first consider the difference between $\mathrm{TDM}_t(\widetilde{\eta})$ and $\mathrm{PDM}_t(\widetilde{\eta})$ when $\mathrm{TDM}_t(\widetilde{\eta})$ is finite. The result for the path velocity matching objectives follows via analogous arguments.

Step 1: $\mathrm{PDM}_t(\widetilde{\eta}) - \mathrm{TDM}_t(\widetilde{\eta})$ Let $\{e_n\}_{n=1}^{\infty}$ be an eigenbasis of U, let U_N be the linear span of $\{e_1, \ldots, e_N\}$, and denote by Π_N the orthogonal projection operator from U into U_N. We perform these projections to ensure that all terms we work with are finite when manipulating the expectations. For any $N \geq 1$, we have:

$$
\begin{aligned}
&\mathbb{E}\left[\|\Pi_N(\widetilde{\eta}(t,\xi_0,x_t) - z)\|_U^2\right] && \text{(A.19)}\\
&\quad = \mathbb{E}\left[\left\|\Pi_N\big((\widetilde{\eta}(t,\xi_0,x_t) - \eta(t,\xi_0,x_t)) - (\eta(t,\xi_0,x_t) - z)\big)\right\|_U^2\right]\\
&\quad = \mathbb{E}\left[\|\Pi_N(\widetilde{\eta}(t,\xi_0,x_t) - \eta(t,\xi_0,x_t))\|_U^2\right] + \mathbb{E}\left[\left\|\Pi_N\big(\eta(t,\xi_0,x_t) - z\big)\right\|_U^2\right]\\
&\qquad +2\,\mathbb{E}\left[\left\langle \Pi_N\big(\widetilde{\eta}(t,\xi_0,x_t) - \eta(t,\xi_0,x)\big), \Pi_N\big(\eta(t,\xi_0,x_t) - z\big)\right\rangle\right]\\
&\quad = \mathbb{E}\left[\|\Pi_N(\widetilde{\eta}(t,\xi_0,x_t) - \eta(t,\xi_0,x_t))\|_U^2\right] + \mathbb{E}\left[\left\|\Pi_N\big(\eta(t,\xi_0,x_t) - z\big)\right\|_U^2\right], && \text{(A.20)}
\end{aligned}
$$

where the final equality is due to an application of the law of iterated expectations which holds from the linearity of the inner product and projection operator.

We now take the limit as $N \to \infty$. The first term in Equation (A.20) converges to $\mathrm{TDM}_t(\widetilde{\eta})$ which by assumption is finite. To analyse the second term, we consider the cases $U = H$ and $U = H_C$.

In the case where $U = H$, we have

$$
\lim_{n\to\infty} \mathbb{E}\left[\|\Pi_N \eta(t,\xi_0,x_t) - z\|_H^2\right] = \mathbb{E}\left[\|\mathbb{E}[z \mid \xi_0, x_t] - z\|_H^2\right] < \infty,
$$

since the Gaussian $z \sim \mathrm{N}(0,C)$ has finite second moment in H-norm due to the covariance operator C being trace-class. Therefore in the limit as $N \to \infty$, the left-hand side (Equation A.19) is finite and converges to $\mathrm{PDM}_t(\widetilde{\eta})$.

In the case where $U = H_C$, we use the fact that $\eta(t,\xi_0,x_t) = \frac{\beta(t)}{\gamma(t)}(\mathbb{E}[\xi_1 \mid \xi_0,x_t] - \xi_1)$. If ξ_1 is supported on the subspace H_C and has finite second moment, then

$$\lim_{n\to\infty} \mathbb{E}\left[\|\Pi_N \eta(t,\xi_0,x_t) - z\|_H^2\right] = \frac{\beta^2(t)}{\gamma^2(t)}\mathbb{E}\left[\|\mathbb{E}[\xi_1 \mid \xi_0,x_t] - \xi_1\|_{H_C}^2\right] < \infty.$$

Hence in both cases, $\mathrm{TDM}_t(\widetilde{\eta})$ and $\mathrm{PDM}_t(\widetilde{\eta})$ differ only by a finite constant.

Step 2: $\mathrm{PPM}_t(\widetilde{\eta}) - \mathrm{TPM}_t(\widetilde{\eta})$ We now consider the difference between $\mathrm{TPM}_t(\widetilde{\varphi})$ and $\mathrm{PPM}_t(\widetilde{\varphi})$ when $\mathrm{TPM}_t(\widetilde{\varphi})$ is finite. Following similar analysis to the above, we have

$$\begin{aligned}\mathbb{E}\left[\left\|\Pi_N(\widetilde{\varphi}(t,\xi_0,x_t) - (\dot{\alpha}(t)\xi_0 + \dot{\beta}_t\xi_1))\right\|_U^2\right] &= \dot{\beta}^2(t)\,\mathbb{E}\left[\left\|\Pi_N\big(\mathbb{E}[\xi_1 \mid \xi_0,x_t] - \xi_1\big)\right\|_U^2\right] \\ &\quad + \mathbb{E}\left[\|\Pi_N(\widetilde{\varphi}(t,\xi_0,x_t) - \varphi(t,\xi_0,x_t))\|_U^2\right].\end{aligned}$$

In the limit as $N \to \infty$, the second term converges to $\mathrm{TPM}_t(\widetilde{\varphi})$. In the case where $U = H_C$, if ξ_1 is supported on H_C with finite second moment, then the first term also converges, and hence the left-hand side converges to $\mathrm{PPM}_t(\widetilde{\varphi})$, is finite, and differs from $\mathrm{TPM}_t(\widetilde{\varphi})$ only by a constant.

In the case where $U = H_C$, we use the fact that $\mathbb{E}[\xi_1 \mid \xi_0,x_t] - \xi_1 = \frac{\gamma(t)}{\beta(t)}(\mathbb{E}[z \mid \xi_0,x_t] - z)$ and hence the first term converges to

$$\frac{\dot{\beta}^2(t)\gamma^2(t)}{\beta^2(t)}\mathbb{E}\left[\|\mathbb{E}[z \mid \xi_0,x_t] - z\|_H^2\right],$$

which is finite since C is trace-class. Again, the left-hand side converges to $\mathrm{PPM}_t(\widetilde{\varphi})$, is finite, and differs from $\mathrm{TPM}_t(\widetilde{\varphi})$ only by a constant. This concludes the proof. ■

Appendix B

Installing the CUED class file

LaTeX.cls files can be accessed system-wide when they are placed in the <texmf>/tex/latex directory, where <texmf> is the root directory of the user's TeXinstallation. On systems that have a local texmf tree (<texmflocal>), which may be named "texmf-local" or "localtexmf", it may be advisable to install packages in <texmflocal>, rather than <texmf> as the contents of the former, unlike that of the latter, are preserved after the LaTeXsystem is reinstalled and/or upgraded.

It is recommended that the user create a subdirectory <texmf>/tex/latex/CUED for all CUED related LaTeXclass and package files. On some LaTeXsystems, the directory look-up tables will need to be refreshed after making additions or deletions to the system files. For TeXLive systems this is accomplished via executing "texhash" as root. MIKTeXusers can run "initexmf -u" to accomplish the same thing.

Users not willing or able to install the files system-wide can install them in their personal directories, but will then have to provide the path (full or relative) in addition to the filename when referring to them in LaTeX.