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 xiii

List of tables xv

1 Introduction 1

1.1 Motivation and Overview . . . 1

1.2 Contributions . . . 1

1.3 Outline . . . 2

2 Background and Preliminaries 3

2.1 Diffusion Models in Finite Dimensions . . . 3

2.2 Stochastic Interpolants in Finite Dimensions . . . 5

2.3 Mathematical Preliminaries . . . 6

2.4 Challenges in Extending SIs to Infinite Dimensions . . . 8

2.5 Related Works . . . 9

2.6 Summary . . . 10

3 Construction and Well-Posedness 11

3.1 Framework . . . 11

3.1.1 Marginal Bridge . . . 12

3.1.2 Conditional Bridge . . . 15

3.2 Existence and Uniqueness of Strong Solutions . . . 17

References 19

Appendix A Mathematical Proofs 21

A.1 Proof of Lemma 2 . . . 21

A.2 Proof of Theorem 3 . . . 23

A.3 Proof of Lemma 6 . . . 24

A.4 Proof of **??** 32

Appendix B Installing the CUED class file **35**

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

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

4. 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.

5. 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.

?? 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 **??**: 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 $\alpha, \beta, \gamma \colon [0,1] \to \mathbb{R}_{\geq 0}$ are continuously differentiable and satisfy $\alpha(0) = \beta(1) = 1$, $\alpha(1) = \beta(0) = 0$, $\gamma(0) = \gamma(1) = 0$ 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:

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

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:

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

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-definite 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. $\alpha(t), \beta(t), \gamma(t) : [0,1] \to \mathbb{R}_{\geq 0}$ are continuously differentiable on $(0,1)$ and satisfy $\alpha(0) = \beta(1) = 1, \alpha(1) = \beta(0) = 0$, $\gamma(0) = \gamma(1) = 1$, 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 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:

$$f(t, x_0, x) := \zeta(t, \xi_0, x) - \frac{\varepsilon}{\gamma(t)} \eta(t, x_0, x).$$

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:

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

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 $\nu_{\xi_0,t}(\mathrm{d}x\,,\xi_0) := \frac{1}{2}\rho_{t|0}(\mathrm{d}x\,,\xi_0) + \frac{1}{2}\mu_{t|0}(\mathrm{d}x\,,\xi_0)$ 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:

$$\begin{aligned}\mathbb{E}\left[\dot{x}_t \mid x_t = x\right] &= \mathbb{E}\left[\dot{x}_t \mid \xi_0 = x_0, x_t = x\right],\\ \mathbb{E}\left[z \mid x_t = x\right] &= \mathbb{E}\left[z \mid \xi_0 = x_0, x_t = x\right],\end{aligned}$$

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 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 $\mathbb{E}[\xi_1 \mid \xi_0, x_t]$.

Lemma 6. *Suppose Hypothesis 5 holds and let $m_{1|0,t}(x_0, x)$ be the posterior mean of ξ_1 when conditioning on $\xi_0 = x_0$ and $x_t = x$:*

$$m_{1|0,t}(x_0, x) := \mathbb{E}[\xi_1 \mid \xi_0 = x_0, x_t = x].$$

For any $t \in (0, 1)$ and μ_0-almost every $x_0 \in H_C$, the map $x \mapsto m_{1|0,t}(x_0, x)$ is Lipschitz continuous in H_C-norm, with a Lipschitz constant L_t at most $\frac{1}{\beta(t)}$. That is,

$$\left\|m_{1|0,t}(x_0, x) - m_{1|0,t}(x_0, y)\right\|_{H_C} \leq \frac{1}{\beta(t)}\|x-y\|_{H_C} \textit{ for all } x, y \in H.$$

Proof (sketch). The full proof is presented in Section A.3 in Appendix A. TODO more details ■

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*.

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.

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

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

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:

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

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 dt*-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 Lemma 6

Lemma 6. *Suppose Hypothesis 5 holds and let $m_{1|0,t}(x_0,x)$ be the posterior mean of ξ_1 when conditioning on $\xi_0 = x_0$ and $x_t = x$:*

$$m_{1|0,t}(x_0,x) := \mathbb{E}\,[\xi_1 \mid \xi_0 = x_0, x_t = x].$$

For any $t \in (0,1)$ and μ_0-almost every $x_0 \in H_C$, the map $x \mapsto m_{1|0,t}(x_0,x)$ is Lipschitz continuous in H_C-norm, with a Lipschitz constant L_t at most $\frac{1}{\beta(t)}$. That is,

$$\left\| m_{1|0,t}(x_0,x) - m_{1|0,t}(x_0,y) \right\|_{H_C} \leq \frac{1}{\beta(t)} \|x-y\|_{H_C} \textit{ for all } x,y \in H.$$

Proof. The proof proceeds in steps TODO

Step 0 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:

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

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

Step 1 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:

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

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

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

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:

$$m_{1|0,t,N}(x_0, x) := \mathbb{E}_{\mu_{1|0,t,N}(\cdot, x_0, x)}[\xi_{1,N}] = \int_{H_N} \xi_{1,N} \mu_{1|0,t,N}(\mathrm{d}\xi_{1,N}, x_0, x). \tag{A.6}$$

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 7):

$$\begin{aligned}D_x m_{1|0,t,N}(x_0, x)[h] &= \frac{\beta(t)}{\gamma^2(t)} \mathbb{E}_{\mu_{1|0,t,N}(\cdot, x_0, x)}\Big[(\xi_{1,N} - m_{1|0,t,N}(x_0, x))\big\langle \xi_{1,N} - m_{1|0,t,N}(x_0, x), h\big\rangle_{H_C}\Big]\\ &= \frac{\beta(t)}{\gamma^2(t)} \mathbb{E}_{\mu_{1|0,t,N}(\cdot, x_0, x)}\Big[(\xi_{1,N} - m_{1|0,t,N}(x_0, x))\big\langle \xi_{1,N} - m_{1|0,t,N}(x_0, x), \Pi_N h\big\rangle_{H_N}\Big],\end{aligned} \tag{A.7}$$

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 $\Pi_N h$:

$$D_x m_{1|0,t,N}(x_0,x)[h] = \frac{\beta(t)}{\gamma^2(t)} C_N \Pi_N h, \tag{A.8}$$
$$\text{where } C_N = \mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\left((\xi_{1,N} - m_{1|0,t,N}(x_0,x))(\xi_{1,N} - m_{1|0,t,N}(x_0,x))^\top\right).$$

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

Step 2 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

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

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

$$\mu_{1|0,t,N}(\mathrm{d}\xi_{1,N}, x_0, x)$$
$$\propto \exp\left(-V_{1|0,t,N}(\xi_{1,N}, x_0, x) - \frac{1}{2}\left\langle Q_N^{-1}(\xi_{1,N} - m_{1|0,N}(x_0)), \xi_{1,N} - m_{1|0,N}(x_0)\right\rangle_{H_N}\right) \mathrm{d}\xi_{1,N}.$$

Let $W_{1|0,t,N}(\xi_{1,N}, x_0, x) := V_{1|0,t,N}(\xi_{1,N}, x_0, x) + \frac{1}{2}\left\langle Q_N^{-1}(\xi_{1,N} - m_{1|0,N}(x_0)), \xi_{1,N} - m_{1|0,N}(x_0)\right\rangle_{H_N}$ 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

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

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

$f(\xi_{1,N}) = \langle \xi_{1,N}, u \rangle_{H_N}$ for any $u \in H_N$, we have $Df(\xi_{1,N}) = u$, and

$$\begin{aligned}\mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\left[\left(f(\xi_{1,N}) - \bar{f}\right)^2\right] &= \langle C_N u, u\rangle \\ &\leq \mathbb{E}_{\mu_{1|0,t,N}(\cdot,x_0,x)}\left[\left\langle \left(D^2_{\xi_{1,N}} W_{1|0,t,N}(\xi_{1,N}, x_0, x)\right)^{-1} u, u\right\rangle_{H_N}\right]. \end{aligned} \tag{A.9}$$

Step 3 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:

$$D^2_{\xi_{1,N}} W_{1|0,t,N}(\xi_N, x_0, x)[u,v] = \frac{\beta^2(t)}{\gamma^2(t)} I_N + \Pi_N \nabla^2_{\xi_1} \Phi(x_0, \xi_{1,N}) \Pi_N + Q_N^{-1} \tag{A.10}$$

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:

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

which allows us to place a Loewner order on Equation (A.10):

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

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.9), 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 4 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

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

Substituting this relationship in Equation (A.8),

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

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

$$\left\| m_{1|0,t,N}(x_0, x) - m_{1|0,t,N}(x_0, y) \right\|_{H_C} = \left\| m_{1|0,t,N}(x_0, x) - m_{1|0,t,N}(x_0, y) \right\|_{H_N} \leq \frac{1}{\beta(t)} \|x - y\|_{H_C}. \tag{A.11}$$

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 8). Since each approximation satisfies the inequality (A.11) 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 inequality.

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

This concludes the proof.

■

Lemma 7. *Let $m_{1|0,t,N}(x_0,x)$ be an approximate posterior mean as defined in Equation* (A.6)*, 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

$$\begin{aligned} D_x m_{1|0,t,N}(x_0,x)[h] &= D_x\left\{\frac{\int_{H_N}\xi_{1,N}\exp\left(-V_{1|0,t,N}(\xi_{1,N},x_0,x)\right)\mathbb{P}_{1|0,N}(\mathrm{d}\xi_{1,N},x_0)}{Z_{1|0,t,N}(x_0,x)}\right\}[h] \\ &= \frac{1}{Z_{1|0,t,N}(x_0,x)}D_x U_{1|0,t,N}(x_0,x)[h] - m_{1|0,t,N}(x_0,x)\frac{D_x Z_{1|0,t,N}(x_0,x)[h]}{Z_{1|0,t,N}(x_0,x)}, \end{aligned} \tag{A.12}$$

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

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

Substituting these into Equation (A.12) and recognising that the fractions come together to form the approximate posterior density, we have:

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

Adding and subtracting zero,

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

we arrive at the expression

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

This concludes the proof. ∎

Lemma 8. *Let* $\{m_{1|0,t,N}(x_0,x)\}_{N=1}^{\infty}$ *be a sequence of approximate posterior means as defined in Equation* (A.6)*, with* $t \in (0,1)$*. Then for almost every* x_0, x*, the sequence 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:

$$m_{1|0,t,N}(x_0,x) = \int_{H_C} \Pi_N \xi_1 \frac{1}{Z_{1|0,t,N}(x_0,x)} \exp\left(-V_{1|0,t}(\Pi_N \xi_1, \Pi_N x_0, x)\right) \mathbb{P}_{1|0}(\mathrm{d}\xi_1, x_0), \tag{A.13}$$

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

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

and

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

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 $\Pi_N \xi_1 \exp\left(-V_{1|0,t}(\Pi_N \xi_1, \Pi_N x_0, x)\right)$ is clearly pointwise convergent to $\xi_1 \exp\left(-V_{1|0,t}(\xi_1, x_0, x)\right)$ 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

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

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 $\lim\limits_{N\to\infty} Z_{1|0,t,N}(x_0,x)=Z_{1|0,t}(x_0,x)$.

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):

$$\int_{H_C}\|\Pi_N\xi_1\|^2_{H_C}\frac{1}{Z^2_{1|0,t,N}(x_0,x)}\exp\big(-2V_{1|0,t}(\Pi_N\xi_1,\Pi_N x_0,x)\big)\mathbb{P}_{1|0}(\mathrm{d}\xi_1\,,x_0). \tag{A.14}$$

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 $Z^2_{1|0,t,N}(x_0,x)$ 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 $\|\Pi_N\xi_1\|^2_{H_C}$ pre-factor.

The dominated convergence theroem therefore applies and it follows that the sequence of integrals in Equation (A.14) 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 ??

Our proof follows a similar overarching argument to to proof of Lemma 6 in Section A.3: 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 which we apply directly 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 $\mathbb{P}_t := \mathrm{N}(0, \gamma^2(t)C)$. Note that the Cameron-Martin space of $\gamma^2(t)C$ is identical to of C, equipped with an inner product scaled by $\frac{1}{\gamma^2(t)}$. Since $\alpha(t)\xi_0 + \beta(t)\xi_1$ is supported 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 $\mathbb{P}_t$:

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

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. Defining $m_t(x_0, x)$ as the posterior mean:

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

we take the Frechet derivative in the direction $h \in H_C$, we again arrive at a covariance:

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

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

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

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

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

Finally, since the coefficient on $\|h\|_{H_C}$ is uniform in h, we apply the mean-value inequality and conclude that $m_t(x_0, x)$ is Lipschitz in H_C-norm with Lipschitz inequality at most $\frac{R^2\beta(t)}{\gamma^2(t)}$:

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

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.