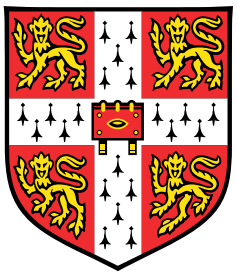
Stochastic Interpolants in Hilbert Spaces



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 (Albergo et al., 2023a) to function spaces. To achieve this, we structure our discussion as follows.

1. We begin by presenting diffusion models (DMs; Ho et al., 2020; Hyvärinen and Dayan, 2005; 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 score-based diffusion models (Song et al., 2021) to function spaces, highlighting the relationship of these methods with their finite-dimensional counterparts.

2.1 Diffusion Models in Finite Dimensions

Diffusion models (DMs; Ho et al., 2020; Hyvärinen and Dayan, 2005; 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 $g : [0,T] \times \mathbb{R}^N \to \mathbb{R}_{>0}$, the *diffusion process* X_t, is the solution to the following *forward SDE*:

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

where W_t is a standard N-dimensional Wiener process on $[0,T]$.

Let μ_t be the law (marginal distribution) of X_t and let $p_t : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$ 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* $\overline{X}_t$, which when solved backwards in time from $\overline{X}_T \sim \mu_T$ yields a sample $\overline{X}_0 \sim \mu_{\text{target}}$:

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

where $\overline{W}_t$ 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 $\widetilde{s}(t,x)$ 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. The score $\log p_t(x)$ is generally intractable, so the learned approximation $\widetilde{s}(t,x)$ can be obtained by minimising a *conditional* score-matching objective:

$$\mathbb{E}_{t\sim\nu, X_0\sim\mu_{\text{target}}, X_t\sim\mu_{t|0}}\Big[\big\|\widetilde{s}(t,X_t) - \log p_{t|0}(X_t \mid X_0)\big\|^2\Big], \tag{2.2}$$

where ν is a distribution over $[0,T]$ and the *noising kernel* $\mu_{t|0}$ is the conditional distribution of X_t, conditional on X_0 with corresponding density $p_{t|0}$.

In the case of a *linear SDE*, where $f(t,X_t) = b(t)X_t$ for some $b : [0,T] \to \mathbb{R}_{\geq 0}$ and $g(t,X_t) = g(t)$, the noising kernel is a Gaussian with closed form expression for its mean and variance:

$$\mu_{1|0} = \mathrm{N}(a(t)X_0, \sigma^2(t)I_N),$$

where

$$a(t) = \exp\left(\int_0^t b(s)\,\mathrm{d}s\right) \text{ and } \sigma^2(t) = \int_0^t \exp\left(2\int_s^t b(u)\,\mathrm{d}u\, g^2(s)\,\mathrm{d}s\right).$$

Hence, the conditional score-matching objective simpifies to the following closed form:

$$\mathbb{E}_{t\sim\nu,X_0\sim\mu_{\text{target}},X_t\sim\mu_{t|0}}\left[\left\|\widetilde{s}(t,X_t)-\frac{1}{\sigma^2(t)}(a(t)X_0-X_t)\right\|^2\right]. \tag{2.3}$$

This objective is a generalisation of *denoising score matching* (DSM) (Song and Ermon, 2019; Vincent, 2011). To learn the score, the network implicitly learns to infer the clean signal X_0 from the corrupted observation X_t. Learning the score is equivalent to learning a denoising function that can reverse the corruption introduced by the forward SDE. This insight allows for reparameterizing the model to directly predict the clean data X or the noise itself, which is often more stable to train (Karras et al., 2022). Indeed, by Tweedie's formula (Efron, 2011), the true score can be written as a conditional expectation:

$$\nabla\log p_t(X_t)=\frac{1}{\sigma^2(t)}(a(t)\mathbb{E}[X_0\mid X_t]-X_t), \tag{2.4}$$

and hence the DSM objective can be seen as a mean-squared error loss in which the intractable conditional expectation $\mathbb{E}[X_0\mid X_t]$ is substituted with a sample from the underlying random variable.

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.

Score-based diffusion models (SBDMs; Song et al., 2021, Equation 11) are a special case of DMs in which the forward SDE is a *variance-preserving* SDE which defines a Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein, 1930). For a diffusion rate $b:[0,T]\to\mathbb{R}_{>0}$, the forward SDE is given by

$$\mathrm{d}X_t=-\frac{1}{2}b(t)X_t\,\mathrm{d}t+\sqrt{b(t)}\,\mathrm{d}W_t\,,\quad X_0\sim\mu_{\text{target}}.$$

The law of X_t converges at an exponential rate to a standard Gaussian $\mathrm{N}(0,I_N)$ in the limit $t\to\infty$. Hence, in a practical implementation for sampling, we truncate T at a sufficiently large value and simulate the time-reversed SDE using the learned score $\widetilde{s}$ starting from a sample $\overline{X}_T\sim\mathrm{N}(0,I_N)$ using any SDE solver (see Karras et al., 2022).

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. While the *rate* of convergence for the forward process in DMs is exponential for variance-preserving SDEs, a large T is still required in practice to bridge sufficiently close to a Gaussian: this makes score estimation more challenging and imposes higher costs for training and sampling (Franzese et al., 2023).

3. The interpolation path is an explicit design choice, allowing us to construct simple bridges (e.g., linear trajectories) between the two distributions. Simple, low-curvature paths are easier for numerical solvers to approximate accurately, which can lead to greater sampling efficiency with fewer function evaluations. While the path in DMs is determined by the chosen drift and diffusion schedules, recent work (Karras et al., 2022; Williams et al., 2024) has shown that the most effective sampling trajectories are an important design choice with significant effects on the quality of generated samples. Stochastic interpolants codifies this principle, by making the path itself a primary object of design, rather than having it emerge from a specific SDE formulation.

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 random variable, 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 $\dot{x}_t \coloneqq \dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1 + \dot{\gamma}(t)z$

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:

$$\mathrm{d}X_t = (\mathbb{E}[\dot{x}_t \mid x_t = X_t] + \varepsilon \nabla \log p_t(X_t))\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t, \quad X_0 \sim \mu_0, t \in [0,1], \tag{2.5}$$

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:

$$\mathrm{d}\overline{X}_t = (\mathbb{E}[\dot{x}_t \mid x_t = X_t] - \varepsilon \nabla \log p_t(X_t))\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}\overline{W}_t, \quad \overline{X}_1 \sim \mu_1, t \in [0,1]. \tag{2.6}$$

Albergo et al. (2023a, Theorem 2.8) show that in finite dimensions, the following relationships holds between the score $\nabla \log p_t(x)$ and a conditional expectation $\mathbb{E}[z \mid x_t = x]$ called the *denoiser*:

$$\nabla \log p_t(x) = \frac{1}{\gamma(t)}\,\mathbb{E}[z \mid x_t = x].$$

This can be seen as an analogue of Tweedie's formula in DMs (Equation 2.4) stated for stochastic interpolants. Hence, to learn the drift coefficient, one can define two networks: a *velocity network* $\widetilde{\varphi}(t,x)$ which predicts $\mathbb{E}\left[\dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1 \mid x_t = x\right]$ and a *denoiser network* $\widetilde{\eta}(t,x)$ which predicts $\mathbb{E}[z \mid x_t = x]$. Since these conditional expectations are intractable, the networks are trained by minimising the following losses in which the conditional expectations are replaced by the sample of the underlying random variables:

$$\mathbb{E}_{t\sim\mathscr{U}[0,1],x_t\sim p_t}\left[\left\|\widetilde{\varphi}(t,x_t) - (\dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1)\right\|^2\right] \text{ and } \mathbb{E}_{t\sim\mathscr{U}[0,1],x_t\sim p_t}\left[\|\widetilde{\eta}(t,x_t) - z\|^2\right].$$

The denoiser network here is in contrast to DMs in which the "denoiser" often refers to a re-parameterisation of the score network designed to predict the clean data given noisy data in the context of DSM (Equation 2.3).

During training, samples x_t are obtained by taking (possibly) paired data $(\xi_0, \xi_1) \sim \mu$ and noise $z \sim \mathrm{N}(0, I_N)$, calculating the interpolant x_t, and using (t, x_t) as inputs to the respective neural networks to predict $(\dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1)$ and z respectively.

The learned approximations $\widetilde{\varphi}$ and $\widetilde{\eta}$ can then be substituted in place of their true counterparts in Equations (2.5) and (2.6) to define a stochastic process to approximately bridge from source to target distribution.

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 Preliminaries for Function Spaces

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 in Hilbert Spaces For a real, separable Hilbert space H, a random variable $x \in H$ is distributed according to a *Gaussian measure* 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

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

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 $\langle f, g \rangle_{H_C} \coloneqq \left\langle C^{-\frac{1}{2}} f, C^{-\frac{1}{2}} g \right\rangle_H$.

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.

A fundamental result in the theory of Gaussian measures is that when C is trace-class, samples from $N(0,C)$ are almost surely[1] not in H_C. Intuitively, samples from $N(0,C)$ are too "rough" to count as part of the smaller subspace of smoother functions H_C.

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 $\nabla \log p_t(x)$ and hence the forward and time-reverse SDEs 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 the key insight from finite-dimensional stochastic interpolants that the score can be computed via the conditional expectation, i.e. $\nabla \log p_t(x) = \frac{1}{\gamma(t)} \mathbb{E}[z \mid x_t = x]$ (Albergo et al., 2023a, Theorem 2.8). We show that 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.

[1] By *almost surely* we mean with probability one, so if a statement P is true μ-almost surely, we mean that P is true with probability one according to a measure μ.

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 regularising 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, a central goal of this thesis is to solve Bayesian inverse problems, which demands the ability to set up a true conditional bridge. Existing SI formulations in finite dimensions do not explicitly guarantee that evolving a process from a point yields a sample from the target distribution, conditional on the source point. Instead, treatment has been given to *additional* conditioning variables, such as class labels, rather than a focus on the source point itself.

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 sampling. We will refer to the former as a *marginal bridge* and the latter as an *conditional bridge*. While we develop the conditional bridge directly in infinite dimensions, our results naturally also apply to the finite-dimensional case.

For clarity and to avoid repetition, our formal analysis will focus on the forward process which evolves from the source to the target distribution. The corresponding results for the reversed evolution follow from a direct symmetry, which we establish in Section 3.4. We will therefore present our main theorems for the forward process only, with the understanding that analogous statements for the reverse evolution hold.

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 $\dot{x}_t \coloneqq \dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1 + \dot{\gamma}(t)z$. 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 $\mu_{1|0}(\mathrm{d}\xi_1, x_0)$ to denote the conditional distribution of ξ_1 on H, conditional on $\xi_0 = x_0$. Analogously, we write $\mu_{0|1}(\mathrm{d}\xi_0, x_1)$ to denote the condition distribution of ξ_0, conditional on $\xi_1 = x_1$.

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.

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

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

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

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

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):

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

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 $\mathrm{d}t$-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:

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

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 $\nu_t \coloneqq \frac{1}{2}\rho_t + \frac{1}{2}\mu_t$ 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). By exploiting the denseness 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 Section 3.2 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 in Theorem 13.

Our second assumption is adapted from Bogachev et al. (2010), who propose that a denseness condition on the range of the Kolmogorov operator helps guarantee the uniqueness of solutions to Fokker-Plank equations in infinite dimensions. 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 the *conditional velcoity* and *denoiser* functions $\zeta, \eta : [0,1] \times H \times H \to H$ to be the following conditional expectations:

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

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

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

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

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

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

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:

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

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 **??** TODO!.

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 finite dimensions) 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, or if the stochastic interpolant x_t is Markovian which is not true in general. This stands in contrast to techniques such as rectified flow (Liu et al., 2022), where paths from source to target are Markovized by constructing a new coupling between source and target. In our setting of Bayesian inverse problems, however, the coupling between the source and target is fixed by the problem's underlying structure and cannot be modified.

Remark 4.1. The theoretical distinction we draw between the marginal and conditional bridge can be connected to techniques such as *classifier-free guidance* (CFG; Ho and Salimans, 2022) for DMs. In CFG, the score network for a DM is trained by randomly dropping the conditioning information during training. At sampling time, guidance is achieved by combining outputs with and without conditioning information to guide the sampling process. An interesting line of future work is to investigate CFG-style techniques for SIs, particularly in infinite dimensions.

The primary focus of this thesis is the application of stochastic interpolants to forward and inverse problems, which are inherently conditional tasks. Consequently, our subsequent theoretical development will concentrate exclusively on the conditional bridge and the CB-SDE. We justify this approach by the fact that the conditional bridge is a more powerful

construction: indeed, the marginal bridge can be recovered by marginalizing the conditional bridge over the source distribution μ_0. Hence, all subsequent results, including the proofs presented in Appendix A, will be presented for the conditional bridge. The corresponding results for the marginal bridge follow from analogous arguments and are omitted for conciseness.

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 on the Wasserstein distance between generated samples and the true target distribution (see Theorem 13). 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.

To show existence and uniqueness of strong solutions, our approach begins by presenting a result on the Lipschitz continuity of the drift coefficient $f(t, x_0, x)$ as a function of x. 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. We provide a more detailed discussion in Remark 7.1.

Our 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:

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

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. ■

Remark 6.1. We acknowledge the primary limitation of this first setting is the strong assumption of strong convexity on the potential Φ. This restricts the density p to having a single maximum, which excludes multi-modal distributions such as Gaussian mixtures. Nevertheless, this requirement is central for the arguments of our proof and $k > 0$ ensures that it is possible to define an SI for which the $\lim_{t\to 0^+} L(t)$ is finite, a requirement for our proof of uniqueness (see Theorem 9). Relaxing this condition is an important direction for future work.

Our second setting replaces the density assumption on μ with an assumption that its 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:

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

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. ■

Remark 7.1. 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.

Remark 7.2. A key feature of the stochatic interpolants framework is the boundary condition $\gamma(1) = 0$ and positivity condition $\gamma(t) > 0$ on $(0,1)$. These have the direct consequence that the Lipschitz constants $L(t)$ derived in Proposition 6 and Proposition 7 suffer from a singularity at the endpoint $t = 1$, that is, $\lim_{t \to 1^-} L(t) = +\infty$ for any choice of γ. This behaviour is characteristic of bridge processes and presents a well-known challenge for establishing existence and uniqueness guarantees on the entire time domain (see, e.g., Li, 2016). Similar singularities are present in related infinite dimensional frameworks, such as in studying the revese-time SDEs in score-based diffusion models (Pidstrigach et al., 2023, Theorem 12).

While extending these guarantees to the entire time domain is an important direction for future work, likely requiring tools from the theory of singular SDEs (see, e.g., Cherny et al., 2005; Flandoli et al., 2010; Hairer, 2014), our analysis provides the necessary

foundation for the practical implementation of our samplers and the derivation of the Wasserstein error bounds.

In contrast to the unavoidable singularity at the terminal time $t = 1$, our framework allows for choices of $\gamma(t)$ for which $\lim_{t\to 0} L(t)$ is finite, which we detailed in our methods section. We therefore establish existence and uniqueness for strong solutions on any compact sub-interval $[0,\bar{t}] \subset [0,1)$. This methodology is consistent with the treatment of similar endpoint issues in related literature, such as Pidstrigach et al. (2023).

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 $\lim_{t\to 0^+} L(t)$ is finite, 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 prove existence using a piecewise construction: we partition the time domain into a finite sequence of small intervals in such a way that the Banach fixed-point theorem yields the existence of solutions on each subinterval. We then stitch these together to form a single, continuous strong solution for the process. The adaptedness of this solution is preserved throughout the iterative construction. ■

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 Groenwall'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). Similarly to in finite dimensions (see Albergo et al., 2023a, Section 2.4), we propose decomposing the drift into two distinct components: a *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

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

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 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:

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

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 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 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, which we detail in Section 3.3.2.

In contrast, the coefficient in the alternative parameterisation (Equation 3.11) has singularities of the order $\frac{1}{\gamma^2(t)}$ rather than $\frac{1}{\gamma(t)}$, which results in a stronger singularity that is non-integrable on $[0,1]$ for any choice of $\gamma(t)$. This argument is formalised below, and means that the alternative parameterisation is fundamentally less stable and cannot be resolved by a similar time-change technique.

Lemma 10. *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. The full proof is presented in Section A.7 in Appendix A. ■

3.3.1 Losses

Having established our choice in parameterising the drift $f(t,x_0,x)$ as a decomposition into a 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 H_C. For approximations $\widetilde{\varphi}$ and $\widetilde{\eta}$, we define the *true velocity matching* (TVM) and *true denoiser matching* (TDM) objectives below:

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

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

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

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

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

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 11. *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 TVM and TDM objectives are finite, the practical objectives* $\mathrm{PVM}_t(\widetilde{\varphi})$ *and* $\mathrm{PDM}_t(\widetilde{\varphi})$ *differ from* $\mathrm{TVM}_t(\widetilde{\varphi})$ *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 11 shows that both the H-norm and H_C-norm are valid choices for our training objectives.

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 minimising 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.

Having established conditions under which our training objectives are well-defined in infinite dimensions, we now turn to quantifying the quality of samples generated from our learned model. The Wasserstein-2 distance provides a natural metric for this purpose, measuring the discrepancy between the law of the generated process and the true conditional law $\mu_{t|0}(\mathrm{d}x_t, \xi_0)$.

However, a direct analysis of the CB-SDE (3.9) is complicated by the singular behavior of the drift coefficient at the endpoints $t = 0$ and $t = 1$. As discussed, the coefficient $\dot{\gamma}(t) - \frac{\varepsilon}{\gamma(t)}$

on the denoiser term becomes at these endpoints. We address this in the next section, where we introduce our technique of time reparameterisation to regularise the CB-SDE (3.9), producing an equivalent SDE, which we call the time-changed CB-SDE (TC-SB-SDE).

3.3.2 Regularising Time Change

The integrability of $\frac{1}{\gamma(t)}$ on $[0,1]$ is a crucial condition which allows us to create a time-changed stochastic process which cancels out the singularity introduced by the coefficient $\dot{\gamma}(t) - \frac{\varepsilon}{\gamma(t)}$ on the denoiser. We state this in the following result.

Lemma 12. *Let the coefficient $c(t) \coloneqq \dot{\gamma}(t) - \frac{\varepsilon}{\gamma(t)}$. Suppose the improper integral $\int_0^1 \frac{1}{\gamma(t)}\,\mathrm{d}t$ is finite and the product $\dot{\gamma}(t)\gamma(t)$ has a (unique) continuous extension on $[0,1]$. Then, there exists a strictly increasing, bijective, continuously differentiable time change $\theta(t) : [0,1] \leftrightarrow [0,1]$ such that the time-transformed coefficient*

$$\hat{c}(t) \coloneqq c(\theta(t))\dot{\theta}(t) = \left(\dot{\gamma}(\theta(t)) - \frac{\varepsilon}{\gamma(\theta(t))}\right)\dot{\theta}(t), \tag{3.16}$$

defined for $t \in (0,1)$, has a continuous extension on the compact interval $[0,1]$.

Proof. The full proof is presented in Section A.9 in Appendix A. We exploit the integrability of $\frac{1}{\gamma(t)}$ to construct the time change $\theta(t)$. ■

Remark 12.1. Note that since $\theta(t)$ is a strictly increasing bijection on $[0,1]$, it necessarily satisfies the boundary conditions $\theta(0) = 0$ and $\theta(1) = 1$, that is, the endpoints of the time domain are unchanged.

The time-changed stochatic process $\hat{X}_t \coloneqq X_{\theta(t)}$ satisfies the following SDE, which we call the time-changed conditional bridge SDE (TC-CB-SDE):

$$\mathrm{d}Y_t \coloneqq f(\theta(t), \xi_0, Y_t)\dot{\theta}(t) + \sqrt{2\varepsilon\dot{\theta}(t)}\,\mathrm{d}\hat{W}_t\,, \quad Y_0 = \xi_0, \tag{3.17}$$

where $\hat{W}_t$ is a C-Wiener process. Since θ is strictly increasing and bijective on $[0,1]$, the TC-CB-SDE has a unique strong solution on $[0, \theta^{-1}(\overline{\theta})]$ as long as X_t has a unique solution on $[0,\overline{t}]$. Intuitively, the reparameterisation slows down time near the original singularities, causing the time-changed process to spend more "computational time" at the endpoints and hence regularising the drift. The benefits of the TC-CB-SDE are two-fold:.

First, we have improved numerical stability: without time reparameterisation, the burden of a well-behaved numerical integration implicitly lies with the training process: the

denoiser's output must decay sufficiently rapdily to zero as $t \to 1^-$ in order to counteract the singularity introduced by the coefficient $c(t)$. The time-change decouples the learning objective from this implicit regularisation, ensuring that training errors are not amplified by the singular coefficient during simulation.

Second, analysis of the time-changed process helps us establish theoretical guarantees: since $\hat{c}(t)$ is continuous and hence bounded on the compact interval $[0,1]$, we are able to derive a meaningful and finite bound on the Wasserstein-2 distance based on the training loss. We turn to this in the next section.

3.3.3 Wasserstein-2 Distance

We now present a result bounding the squared Wasserstein-2 distance, stated in terms of the TVM_t and PDM_t losses (Equations 3.12 and 3.13). Intuitively, the Wasserstein-2 distance between two measures π_1 and π_2 on H lifts distance induced by the H-norm into the space of measures on H, and is defined by:

$$\left(\inf_{\pi_\times} \int_{H^2} \|x-y\|_H^2 \pi_\times(\mathrm{d}(x,y))\right)^{\frac{1}{2}},$$

where the infimum is taken over the space of measures $\pi_\times$ on H^2 which marginalise on π_1 and π_2.

Theorem 13. *Let $\widetilde{\varphi}$ and $\widetilde{\eta}$ be the approximations of φ and η respectively, and let $\gamma(t)$ and $c(t)$ satisfy the conditions in Lemma 12.*

Suppose that for all $t \in [0,1], x_0 \in H$, the mappings $x \mapsto \widetilde{\varphi}(t,x_0,x)$ and $x \mapsto \widetilde{\eta}(t,x_0,x)$ are Lipschitz continuous in H-norm, that is, there exists a constant $\widetilde{L} < \infty$ where for all $x, y \in H$,

$$\|\widetilde{\varphi}(t,x_0,x) - \widetilde{\varphi}(t,x_0,y)\|_H \leq \widetilde{L}\|x-y\|_H \text{ and } \|\widetilde{\eta}(t,x_0,x) - \widetilde{\eta}(t,x_0,y)\|_H \leq \widetilde{L}\|x-y\|_H.$$

Furthermore, suppose the CB-SDE has a unique strong solution X_t on $[0,\bar{t}] \subseteq [0,1]$ (see Propositions 6 or 7 for sufficient conditions) and let $\widetilde{X}_t$ be the unique strong solution to the CB-SDE when replacing the velocity φ and denoiser η with their approximations, solved with $\widetilde{X}_0 = X_0 = \xi_0$.

Then, the expected squared Wasserstein distance $\mathscr{W}_2^2(\bar{t})$ between the law of the approximate path $\widetilde{X}_t$ and the law of the conditional interpolant $\mu_{t|0}(\mathrm{d}x_t, \xi_0)$ at time $t = \bar{t}$

is bounded by:

$$\mathcal{W}_2^2(\bar{t}) \leq 2\bar{c}^2 e^{2\bar{c}\widetilde{L}+1} \int_0^{\theta^{-1}(\bar{t})} \mathrm{TVM}_{\theta(t)}(\widetilde{\varphi}) + \mathrm{TDM}_{\theta(t)}(\widetilde{\eta})\,\mathrm{d}t\,, \tag{3.18}$$

where

$$\bar{c} := \max_{t\in[0,\theta^{-1}(\bar{t})]} \left(\dot{\theta}(t) + |\hat{c}(t)|\right) < \infty.$$

Proof. The full proof is presented in Section A.10 in Appendix A. From Theorem 4, the law of X_t is equal to $\mu_{t|0}(\mathrm{d}x_t\,, \xi_0)$ and hence we couple the solution X_t to the CB-SDE with the solution $\widetilde{X}_t$ to the CB-SDE when replacing φ and η with their approximations, by driving both stochastic processes with a common C-Wiener process W_t. We make use of Lemma 12 to obtain a tractable bound by considering the time-changed counterparts $\hat{X}_t$ and $\hat{\widetilde{X}}_t$. ■

Theorem 13 provides two key insights into the behavior and design of our learned sampler. First, the primary theoretical requirement of the theorem is the uniform Lipschitz continuity of the learned *approximations* $\widetilde{\varphi}$ and $\widetilde{\eta}$, which stands in contrast to Theorems 8 and 9 which consider the *true* drift. The uniform Lipschitz continuity of the approximations $\widetilde{\varphi}$ and $\widetilde{\eta}$ in H-norm is satisfied by the neural operator architectures we employ. This condition is standard for the analysis of infinite-dimensional frameworks, although its treatment varies. For example in diffusion models, Hagemann et al. (2023) make an equivalent, explicit assumption on their learned score, while Pidstrigach et al. (2023, Theorem 14) state the assumption for the true score and implicitly apply it to the learned approximation to bound the error. A key contribution of our work is to state this assumption and its role in the final error bound transparently, providing a discussion of design implications in Section 4.1

A crucial implication of this regularity is that the CB-SDE driven by the *approximate* drift has a unique strong solution on the full interval $[0,1]$. This result can be seen by making use of the fact that $x \mapsto \widetilde{f}(\theta(t), x_0, x)\dot{\theta}(t)$ is uniformly Lipschitz in H-norm, with Lipschitz constant $2\bar{c}\widetilde{L}$. Hence, we may applying the arguments in the proof of Theorem 8 to the TC-CB-SDE, simplified by the new uniform Lipschitz constant $\bar{c}\widetilde{L}$. This directly obtains a unique solution on $[0,1]$ since the Lipschitz continuity in H-norm now means that Banach's fixed point theorem gives us uniqueness for free, without requiring the additional arguments in Theorem 9. We then obtain a unique solution to the approximate CB-SDE by reversing the time change.

Hence, we run our inference by implementing an SDE integrator for our learned drift $\widetilde{f}$ on the entire time domain $[0,1]$, even though the existence and uniqueness results of Theorems 8

and 9 for the true CB-SDE were restricted to strict sub-intervals $[0,\bar{t}] \subset [0,1]$ due to the singularity of the true drift at $t = 1$.

Second, our time-change regularization is essential to guarantee a tractable error bound. Theorem 13 guarantees that the law of the generated process can be made arbitrarily close to the true conditional law if $\mathrm{TVM}_t(\widetilde{\varphi})$ and $\mathrm{TDM}_t(\widetilde{\varphi})$ are bounded and sufficiently small for a dense subset of the time domain.

Since the time domain is re-weighted by the time-change $\theta(t)$ in Equation (3.18), an intuitive training strategy is to sample training times non-uniformly according to this change of time. However, our preliminary experiments demonstrate this to be suboptimal. We therefore maintain a uniform time sampling strategy in our implementation. We hypothesize this is due to "model starvation", where the network receives too few training samples in certain regions of the original time interval. We explore this in more detail in Section 4.3.

3.4 Bridging from Target to Source

Our preceeding analysis has focused on establishing a stochastic bridge from the source $\xi_0 \sim \mu_0$ to the conditional target distribution $\mu_{1|0}(\mathrm{d}\xi_1, \xi_0)$. To bridge from a target point $\xi_1 \sim \mu_1$ to the conditional source distribution $\mu_{0|1}(\mathrm{d}\xi_0, \xi_1)$, we may consider a *reverse interpolant*

$$x_t^{\mathrm{rev}} \coloneqq \alpha(1-t)\xi_0 + \beta(1-t)\xi_1 + \gamma(1-t).$$

Following our parameterisation of the CB-SDE in Equation (3.10), it is easy to see that the analogous SDE which bridges from ξ_1 to $\mu_{0|1}(\mathrm{d}\xi_0, \xi_1)$, which we call the *reverse conditional bridge SDE* (RCB-SDE), is

$$\mathrm{d}X_t^{\mathrm{rev}} = f^{\mathrm{rev}}(t, \xi_1, X_t^{\mathrm{rev}})\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t, \quad X_0^{\mathrm{rev}} = \xi_1,$$

where

$$\begin{aligned}
f^{\mathrm{rev}}(t, x_1, x) &\coloneqq \varphi^{\mathrm{rev}}(t, x_0, x) - \left(\dot{\gamma}(1-t) + \frac{\varepsilon}{\gamma(1-t)}\right)\eta^{\mathrm{rev}}(t, x_0, x) \\
\varphi^{\mathrm{rev}}(t, x_1, x) &\coloneqq -\,\mathbb{E}\left[\dot{\alpha}(1-t)\xi_0 + \dot{\beta}(1-t)\xi_1 \mid \xi_1 = x_1, x_t^{\mathrm{rev}} = x\right], \\
\eta^{\mathrm{rev}}(t, x_1, x) &\coloneqq \mathbb{E}[z \mid \xi_1 = x_1, x_t = x].
\end{aligned}$$

Unlike in Albergo et al. (2023a), we state this as an SDE to be solved *forward* in time, starting from the initial condition $X_0^{\mathrm{rev}} = \xi_1$. This gives a consistent indexing of time and assuming $\gamma(t) = \gamma(1-t)$, the same time change $\theta(t)$ can be applied to regularise the RCB-SDE. Our

preceeding results apply analogously, by re-stating the conditions on ξ_1 in Proposition 7 and Theorem 9 as conditions on ξ_0. Similarly, the *reverse marginal bridge SDE* (RMB-SDE) can be recovered by dropping the conditioning on ξ_1 in the definitions above to form a stochastic bridge from the marginal target distribution μ_1 to the source μ_0.

Since the conditional expectations are taken conditional on $(\xi_1, x_t^{\text{rev}})$, we must in general train additional networks in order to learn both the CB-SDE and the RCB-SDE. In contrast, in applications where only the forward and reverse *marginal* bridges matter, we no longer require conditioning on ξ_0 and ξ_1 for the forward and reverse bridge respectively, and hence the same trained networks can be used for both forward and reverse tasks by the deterministic relationship $x_t = x_{1-t}^{\text{rev}}$. In Section 4.4, we show that simulating the MB-SDE for conditional tasks causes only a modest loss in performance, and hence a MB-SDE/RMB-SDE may be still be useful in conditional settings.

3.5 Summary

In this chapter, we have presented our main theoretical framework of stochastic interpolants in infinite dimensions, justifying the CB-SDE as the primary mechanism by which we can establish a conditional bridge from $\xi_0 \sim \mu_0$ to $\xi_{1|0}(\mathrm{d}\xi_1 \mid \xi_0)$. We provided sufficient conditions under which the CB-SDE is well-posed, and justified a reparameterisation of its drift coefficient and a change in time as key techniques using which learning and inference are well-behaved. Finally, we quantified the quallity of generated samples by providing a bound on the Wasserstein-2 distance between generated and true samples.

Chapter 4

Methodology and Results

Building on the theory from Chapter 2, this chapter details the practical application and validation of our framework.

1. We first establish practical design guidelines from our theory and provide a specific instantiation of our framework for solving PDE-based forward and inverse problems.
2. Next, we conduct preliminary tests on 1D Darcy flow to verify our method and gain practical insights.
3. Finally, we present our main results, applying the framework to challenging 2D Darcy and Navier-Stokes equations.

4.1 Design Choices

Our theory provides three key design choices which we discuss below. These are the choice of noise, model capacity, and noise scale factor $\gamma(t)$.

4.1.1 Tradeoff between noise regularity and learnability

The choice of noise and model capacity are closely tied, so we discuss them together. Our results on well-posedness require the noise to be sufficiently rough such that the target data (and also the source data in the case of inverse problems) to be supported on the Cameron-Martin space H_C (see Remark 7.1). Consequently, the networks $\widetilde{\varphi}, \widetilde{\eta}$ must have enough capacity to process the less regular interpolant x_t, and predict the rough training targets $\dot{\alpha}(t)\xi_0 + \dot{\beta}(t)\xi_1$ and z.

However, excessively rough noise is detrimental for two reasons. First, it creates a more difficult learning problem, as the input x_t is also rougher and less informative, and the noise

target z is harder for the denoiser $\widetilde{\varphi}$ to predict. Second, it can harm sample quality: the Wasserstein-2 error bound in Equation (3.18) grows exponentially with the Lipschitz constant $\widetilde{L}$ of the *learned* approximations $\widetilde{\varphi}, \widetilde{\eta}$. A rougher input x_t and noise target can lead to a larger $\widetilde{L}$, weakening the performance guarantee.

This tension creates a practical "sweet spot" where noise must be just rough enough to satisfy the theoretical condition that ξ_0, ξ_1 are supported in H_C, but smooth enough to ensure a tractable learning problem and well-behaved Lipschitz constant. In this work, we employ implicit regularisation through our choice of network architecture and leave explicit control over network smoothness to future work.

4.1.2 Choice of $\gamma(t)$

As detailed in Section 3.3.2, the noise scaling factor γ must be chosen such that $\frac{1}{\gamma(t)}$ is integrable on $[0,1]$. In practice, this means

$$\lim_{t\to 0} \frac{t}{\gamma(t)} = \lim_{t\to 1} \frac{1-t}{\gamma(t)} = 0, \tag{4.1}$$

that is, $\gamma(t)$ must approach zero more slowly than t near the endpoint $t = 0$ (and similarly for $1-t$ near $t = 1$). Then, during sampling we solve the time-changed CB-SDE (3.17) which re-parameterises the integration for numerical stability.

4.2 Instantiation of Framework

We now provide a concrete setup of the framework and algorithms that we use to solve PDE-based forward and inverse problems.

Hilbert spaces Throughout, we work with data that lie in the Hilbert space of square-integrable functions on a compact Euclidean subset: this will be $L^2 = L^2(D)$ where $D = [0,1]$ for functions defined on a unit interval $D = [0,1]^2$ for functions on the unit square. We equip this with the canonical L^2-inner product:

$$\langle f, g \rangle_{L^2} \coloneqq \int_D f(x) g(x) \, \mathrm{d}x.$$

We work with two distinct settings for source and target data. In the first, we address homogeneous data where ξ_0 and ξ_1 represent similar physical quantities and can be naturally modelled on the same function space. We therefore define the interpolant space as $H \coloneqq L^2$.

This approach is suitable for problems like predicting a future fluid pressure field ξ_1 from a past one ξ_0.

In the second setting, we address heterogeneous data, where ξ_0 and ξ_1 are different physical quantities (e.g. a permeability field and a pressure field). Interpolating directly between these in a single channel would be unnatural and impose a difficult disentanglement task during learning. To provide a stronger inductive bias, we define the interpolant space as the *product space* $H := L^2 \times L^2$. Here, for a pair of source and target data functions, we simply define the data as $\xi_0' = (\xi_0, 0)$ and $\xi_1' = (0, \xi_1)$, where 0 represents the zero function on D. Under this construction, there is no signal bleed in the interpolant x_t as heterogeneous data channels are kept separate. Since the product space H is still a Hilbert space, all results of our theory hold true for the new source and target random variables ξ_0' and ξ_1'.

Noise When $H = L^2$, we define noise z as samples from a Gaussian process (Rasmussen and Williams, 2006) with zero mean and radial basis kernel k:

$$z \sim \mathrm{GP}(0, k), \text{ where } k(x,y) := \exp\left(-\frac{1}{2\ell}\|x-y\|_D^2\right).$$

equal to a radial basis function of gain 1 and varying length scales ℓ to investigate the impact of noise roughness on evaluation performance. This is equivalent to sampling z from a Gaussian measure $N(0, C)$ on H where the covariance operator is given by

$$Cf(x) := \int_D f(y)k(x,y)\,\mathrm{d}y\,.$$

In the product space setting where $H = L^2 \times L^2$, we define the noise z as a pair of independent samples from this process, i.e. $z = (z_0, z_1)$ where each component $z_0, z_1 \overset{\text{i.i.d.}}{\sim} \mathrm{GP}(0, k)$. Formally, this is equivalent to sampling from a GP with matrix-valued kernel K:

$$z \sim \mathrm{GP}(0, K), \text{ where } K(x,y) := \begin{bmatrix} k(x,y) & 0 \\ 0 & k(x,y) \end{bmatrix}.$$

Choice of $\gamma(t)$ Following (Albergo et al., 2023a), we define

$$\gamma(t) := \sqrt{bt(1-t)}.$$

This satisfies Equation (4.1), so $\frac{1}{\gamma(t)}$ is integrable on $[0,1]$. The following result provides necessary and sufficient conditions on the time change function θ such that the time-changed

coefficient $\hat{c}(t) = c(\theta(t))\dot{\theta}(t)$ (Equation 3.16) on the denoiser is finite on $[0,1]$. We provide the proof in Section A.11 in Appendix A.

Lemma 14. *A strictly increasing, bijective, continuously differentiable time change function $\theta(t)$ on $[0,1]$ is a valid change-of-time ensuring that $\hat{c}(t)$ is finite on $[0,1]$ if and only if $\theta(t)$ satisfies the following conditions.*

1. $\lim_{t\to 1^-} \frac{\dot{\theta}(t)}{2(1-t)} < \infty$*; and*
2. $\lim_{t\to 0^+} \frac{\dot{\theta}(t)}{2t} < \infty$ *if* $\varepsilon \neq \frac{b}{2}$.

Remark 14.1. Intuitively, these conditions (1) and (2) mean that the *rate of time change* decays to zero at the endpoints. This controlled deceleration is precisely what resolves the singularity.

For SDE inference, we will choose $\varepsilon = \frac{b}{2}$, which simplifies the original coefficient to $c(t) = -\sqrt{\frac{bt}{1-t}}$. This resolves the singularity at $t = 0$ leaving only the singularity at $t = 1$ to be managed by the time change. Therefore, only condition (1) is required to ensure the time-changed coefficient $\hat{c}(t)$ is finite on $[0,1]$.

Choice of $\alpha(t)$ and $\beta(t)$ Following work on rectified flow (Liu et al., 2022), we choose $\alpha(t) \coloneqq 1-t$ and $\beta(t) \coloneqq t$, which makes the signal $\alpha(t)\xi_0 + \beta(t)\xi_1$ a linear interpolation between source and target data. This straight line path has two advantages:

1. The instantaneous velocity is $\varphi(t, x_t) = \mathbb{E}[\xi_1 - \xi_0 \mid \xi_0, x_t]$, which simplifies the training task as the network $\widetilde{\varphi}$ targets the constant vector $\xi_1 - \xi_0$.
2. The lack of curvature in the trajectory means that the ODE component is easier to solve during inference. In fact, in the deterministic case where $\varepsilon = 0$, the probability flow ODE can be solved in just one step.

4.3 1D Dataset

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.

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

4.4 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.

Cherny, A. S. et al. (2005). *Singular stochastic differential equations*. Springer Science & Business Media.

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

Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614.

Flandoli, F., Gubinelli, M., and Priola, E. (2010). Well-posedness of the transport equation by stochastic perturbation. *Inventiones mathematicae*, 180(1):1–53.

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.

Hagemann, P., Mildenberger, S., Ruthotto, L., Steidl, G., and Yang, N. T. (2023). Multilevel diffusion: Infinite dimensional score-based diffusion models for image generation. *arXiv preprint arXiv:2303.04772*.

Hairer, M. (2014). A theory of regularity structures. *Inventiones mathematicae*, 198(2):269–504.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models.

Ho, J. and Salimans, T. (2022). Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*.

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

Karras, T., Aittala, M., Aila, T., and Laine, S. (2022). Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577.

Li, X.-M. (2016). Generalised brownian bridges: examples. *arXiv preprint arXiv:1612.08716*.

Liu, X., Gong, C., and Liu, Q. (2022). Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*.

Pidstrigach, J., Marzouk, Y., Reich, S., and Wang, S. (2023). Infinite-dimensional diffusion models. *arXiv preprint arXiv:2302.10130*.

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. MIT press Cambridge, MA.

Song, Y. and Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32.

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

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical review*, 36(5):823.

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

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

Williams, C., Campbell, A., Doucet, A., and Syed, S. (2024). Score-optimal diffusion schedules. *Advances in Neural Information Processing Systems*, 37:107960–107983.

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

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

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 $\mathrm{d}t$-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:

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

Proof. The proof proceeds in steps TODO

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

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

Hence, if we can show that the mapping $x \mapsto \mathbb{E}\left[\xi_1 \mid \xi_0 = x_0, x_t = x\right]$ 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:

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

and $Z_{1|0,t}(x_0, x) \coloneqq \int_{H_C} \exp\big(-V_{1|0,t}(\xi_1, x_0, x)\big)\mathbb{P}_{1|0}(\mathrm{d}\xi_1, x_0)$ 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 $\{e_1, \ldots, e_N\}$. 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 $\xi_{1,N} \coloneqq \Pi_N \xi_1$. Furthermore, we define a reference measure by projecting $\mathbb{P}_{1|0}$ onto this subspace:

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

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

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

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

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

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

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 15):

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

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 $\Pi_N h$:

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

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

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

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

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

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

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

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:

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

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 16). 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 15. *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*

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

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 $U_{1|0,t,N}(x_0,x) \coloneqq \int_{H_N} \xi_{1,N} \exp\big(-V_{1|0,t,N}(\xi_{1,N},x_0,x)\big)\mathbb{P}_{1|0,N}(\mathrm{d}\xi_{1,N},x_0)$ 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:

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

Adding and subtracting zero,

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

we arrive at the expression

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

This concludes the proof. ■

Lemma 16. *For every $x_0, x \in H$ and $t \in (0,1)$, the sequence of approximate posterior means $\big\{m_{1|0,t,N}(x_0,x)\big\}_{N=1}^{\infty}$ 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:

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

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

We define the sequence of functions

$$f_N(\xi_1) \coloneqq \Pi_N \xi_1 \frac{1}{Z_{1|0,t,N}(x_0,x)} \exp\big(-V_{1|0,t}(\Pi_N\xi_1, \Pi_N x_0, x)\big),$$

and

$$f(\xi_1) \coloneqq \xi \frac{1}{Z_{1|0,t}(x_0,x)} \exp\big(-V_{1|0,t}(\xi_1, x_0, x)\big),$$

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 $\Pi_N\xi_1 \exp\big(-V_{1|0,t}(\Pi_N\xi_1, \Pi_N x_0, x)\big)$ is clearly pointwise convergent to $\xi_1 \exp\big(-V_{1|0,t}(\xi_1, x_0, x)\big)$ 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 $\lim_{N\to\infty} Z_{1|0,t,N}(x_0,x) = Z_{1|0,t}(x_0,x)$.

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:

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

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 $\mathbb{P}_t$:

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

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. 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 15, 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

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

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, 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*

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

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

Proof. We begin by addressing the behavior of the drift and its associated Lipschitz constant, $L(t)$, at the initial time $t = 0$. The drift coefficient $f(t,x_0,x)$ is defined via conditional expectations of the stochastic interpolant x_t, conditioned on the events $\xi_0 = x_0$ and $x_t = x$.

At the specific instant $t = 0$, the conditioning event $x_t = x$ becomes $\xi_0 = x$ by definition of the stochastic interpolant. For this event to be consistent with the condition $\xi_0 = x_0$, we must have $x = x_0$. Consequently at time 0, the drift $f(0,x_0,x)$ is only well-defined where $x_0 = x$. This is satisfied by the initial condition $X_0 = \xi_0$ for the CB-SDE (3.9).

However, the Lipschitz condition is a statement about the behavior of the drift under perturbations, i.e., comparing $f(t,x_0,x)$ and $f(t,x_0,y)$ for $x \neq y$. Since the drift is not defined for such perturbations at $t = 0$, the Lipschitz condition is only meaningful for $t > 0$.

Therefore, for the purpose the arguments below, we extend the function $L(t)$ to be continuous on the entire closed interval $[0,\bar{t}]$. Without loss of generality, we define $L(0) := \lim_{t\to 0^+} L(t)$. This is justified because the value of the drift at a single point in time does not affect the SDE's solution.

Step 1: Partitioning of time domain With this definition, we can proceed with the proof assuming that $L(t)$ is continuous and therefore bounded on the compact interval $[0,\bar{t}]$. Hence, it is possible to create a finite partition $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_k < \cdots < \tau_K = \bar{t}$ of $[0,\bar{t}]$

with $K < \infty$ such that

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

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:

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

To argue existence of a strong solution to the CB-SDE on $[0,\bar{t}]$, 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 X_t on $[0,\bar{t}]$.

Fix any event $\omega \in \Omega$, so that $\xi_0(\omega)$ and $W_t(\omega)$ are respectively the outcomes of the random variable ξ_0 and the Wiener process at time t, and define $X_0(\omega) \coloneqq \xi_0(\omega)$. Furthermore, let

$$\widetilde{W}_{k,t} \coloneqq \int_{\tau_{k-1}}^{t} \sqrt{2\varepsilon}\,\mathrm{d}W_s.$$

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

$$(\Psi_{k,\omega}Y)(t) = \int_{\tau_{k-1}}^{t} f\Big(s,\xi_0(\omega),X_{\tau_{k-1}}(\omega)+\widetilde{W}_{k,s}(\omega)+Y(s)\Big)\,\mathrm{d}s, \quad \text{for all } t \in [\tau_{k-1},\tau_k]. \tag{A.16}$$

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

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

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 $X_t(\omega) \coloneqq X_{\tau_{k-1}}(\omega) + \widetilde{W}_{k,t}(\omega) + Y^*(t)$ for all $t \in [\tau_{k-1}, \tau_k]$. Substituting this definition into the fixed point identity $\Psi_{k,\omega} Y^* = Y^*$, we have

$$X_t(\omega) - X_{\tau_{k-1}}(\omega) - \widetilde{W}_{k,t}(\omega) = \int_{\tau_{k-1}}^{t} f(s, \xi_0(\omega), X_s(\omega))\,\mathrm{d}s$$
$$\implies X_t(\omega) = X_{\tau_{k-1}}(\omega) + \int_{\tau_{k-1}}^{\tau_k} f(s, \xi_0(\omega) X_s(\omega))\,\mathrm{d}s + \int_{\tau_{k-1}}^{t} \sqrt{2\varepsilon}\,\mathrm{d}W(\omega)\,,$$

which is the integral form of the CB-SDE (3.9), 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 X_t on the interval $t \in [\tau_{k-1}, \tau_k]$. Now that we have a definition of $X_{\tau_k}(\omega)$, we may repeat the inductive step for $k \leftarrow k+1$. This builds a stochatic process X_t on the entire desired interval $t \in [0, \bar{t}]$.

It remains to check that X_t is $\mathbb{F}$-adapted on $[0, \bar{t}]$. 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 $\mathscr{F}_{\tau_{k-1}}$-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 $\mathscr{F}_t$-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 $\mathscr{F}_t$-measurable as a function of ω, since the integrand in Equation (A.16) is the composition of a continuous function with a $\mathscr{F}_t$-measurable function. Hence, every time we perform a Banach iteration, the outcome at time $t \in [\tau_{k-1}, \tau]$ is $\mathscr{F}_t$-measurable. Since σ-fields are closed under countable pointwise limits, it follows that $Y^*_\omega(t)$ and thus $X_t(\omega)$ are $\mathscr{F}_t$-measurable for all $t \in [\tau_{k-1}, \tau_k]$. Repeating the induction for all steps up to $k = K$ ensures that $X_t(\omega)$ is $\mathscr{F}^*_t$-measurable for all $t \in [0, \bar{t}]$ and hence X_t is $\mathbb{F}$-adapted 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 $\{e_1, \ldots, e_N\}$. 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, we assume without loss of generality that $L(t)$ is continuous on $[0, \bar{t}]$. Let X_t and X'_t be two strong solutions

for the same initial condition, $X_0 = X_0' = \xi_0$ and driven by the same Wiener process W_t on $[0,\bar{t}]$. *A priori*, it is not guaranteed that $\|X_t - X_t'\|_{H_C} < \infty$ since $X_t - X_t'$ may not be in H_C. However, for each $N \geq 1$, it is guaranteed that the projected diference $P_N(X_t - X_t') \in H_C$ 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

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

where $L(t)$ is as defined in **??**.

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

$$\big\|P_N\big(X_t - X_t'\big)\big\|_{H_C} \leq \big\|P_N\big(X_0 - X_0'\big)\big\|_{H_C} \exp\left(\int_0^{\bar{t}} L(t)\big\|P_N\big(X_t - X_t'\big)\big\|_{H_C}\right).$$

Since by definition $X_0 = X_0' = \xi_0$, so $X_0 - X_0' = 0$, it follows that

$$\big\|P_N\big(X_t - X_t'\big)\big\|_{H_C} = 0,$$

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

$$X_t = X_t', \text{ for all } t \in [0,\bar{t}].$$

This concludes the proof. ■

A.7 Proof Lemma 10

Lemma 10. *For any $\varepsilon \geq 0$, there exists no function $\gamma\colon [0,1] : \mathbb{R}_{\geq 0}$ 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*

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

is integrable on $[0,1]$.

Proof. We consider the function $y(t) \coloneqq \gamma^2(t)$, 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 $I_+ \coloneqq \lim_{a\to 1^-}\int_{\frac{1}{2}}^a c(t)$ 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 11

Proposition 11. *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 TVM and TDM objectives are finite, the practical objectives* $\mathrm{PVM}_t(\widetilde{\varphi})$ *and* $\mathrm{PDM}_t(\widetilde{\varphi})$ *differ from* $\mathrm{TVM}_t(\widetilde{\varphi})$ *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 the denoising matching objectives $\mathrm{TDM}_t(\widetilde{\eta})$ and $\mathrm{PDM}_t(\widetilde{\eta})$ when $\mathrm{TDM}_t(\widetilde{\eta})$ is finite. The result for the 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:

$$
\begin{aligned}
&\mathbb{E}\left[\|\Pi_N(\widetilde{\eta}(t,\xi_0,x_t) - z)\|_U^2\right] \qquad (A.17)\\
&\quad = \mathbb{E}\left[\left\|\Pi_N\big((\widetilde{\eta}(t,\xi_0,x_t) - \eta(t,\xi_0,x_t)) - (\eta(t,\xi_0,x_t) - z)\big)\right\|_U^2\right]\\
&\quad = \mathbb{E}\left[\|\Pi_N(\widetilde{\eta}(t,\xi_0,x_t) - \eta(t,\xi_0,x_t))\|_U^2\right] + \mathbb{E}\left[\left\|\Pi_N\big(\eta(t,\xi_0,x_t) - z\big)\right\|_U^2\right]\\
&\qquad +2\,\mathbb{E}\left[\left\langle \Pi_N\big(\widetilde{\eta}(t,\xi_0,x_t) - \eta(t,\xi_0,x)\big), \Pi_N\big(\eta(t,\xi_0,x_t) - z\big)\right\rangle\right]\\
&\quad = \mathbb{E}\left[\|\Pi_N(\widetilde{\eta}(t,\xi_0,x_t) - \eta(t,\xi_0,x_t))\|_U^2\right] + \mathbb{E}\left[\left\|\Pi_N\big(\eta(t,\xi_0,x_t) - z\big)\right\|_U^2\right], \qquad (A.18)
\end{aligned}
$$

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.18) 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.17) 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: $\mathrm{PVM}_t(\widetilde{\eta}) - \mathrm{TVM}_t(\widetilde{\eta})$ We now consider the difference between $\mathrm{TVM}_t(\widetilde{\varphi})$ and $\mathrm{PVM}_t(\widetilde{\varphi})$ when $\mathrm{TVM}_t(\widetilde{\varphi})$ 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 $\mathrm{TVM}_t(\widetilde{\varphi})$. 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 $\mathrm{PVM}_t(\widetilde{\varphi})$, is finite, and differs from $\mathrm{TVM}_t(\widetilde{\varphi})$ 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 $\mathrm{PVM}_t(\widetilde{\varphi})$, is finite, and differs from $\mathrm{TVM}_t(\widetilde{\varphi})$ only by a constant. This concludes the proof. ■

A.9 Proof of Lemma 12

Lemma 12. *Let the coefficient $c(t) \coloneqq \dot{\gamma}(t) - \frac{\varepsilon}{\gamma(t)}$. Suppose the improper integral $\int_0^1 \frac{1}{\gamma(t)}\,\mathrm{d}t$ is finite and the product $\dot{\gamma}(t)\gamma(t)$ has a (unique) continuous extension on $[0,1]$. Then, there exists a strictly increasing, bijective, continuously differentiable time change $\theta(t) : [0,1] \leftrightarrow [0,1]$ such that the time-transformed coefficient*

$$\hat{c}(t) \coloneqq c(\theta(t))\dot{\theta}(t) = \left(\dot{\gamma}(\theta(t)) - \frac{\varepsilon}{\gamma(\theta(t))}\right)\dot{\theta}(t), \tag{3.16}$$

defined for $t \in (0,1)$, has a continuous extension on the compact interval $[0,1]$.

Proof. Let $h(t) \coloneqq \varepsilon/\gamma(t)$ and $C \coloneqq \int_0^1 h(\sigma)d\sigma$, which is finite and positive by assumption. We define a new time variable $s(t)$ by

$$s(t) \coloneqq \frac{1}{C}\int_0^t h(\sigma)\,\mathrm{d}\sigma\,.$$

Since $h(t) > 0$ for $t \in (0,1)$, $s(t)$ is a strictly increasing, continuously differentiable bijection from $[0,1]$ to itself. We define the time-change $\theta(t)$ as its inverse, $\theta(t) \coloneqq s^{-1}(t)$. This is also strictly increasing and continuously differentiable on $(0,1)$, with derivative $\dot{\theta}(t) = \frac{C}{\varepsilon}\gamma(\theta(t))$. Substituting this into the definition of $\hat{c}(t)$, we have

$$\hat{c}(t) = \frac{C}{\varepsilon}(\dot{\gamma}(\theta(t))\gamma(\theta(t))) - C.$$

By assumption, the function $\dot{\gamma}(t)\gamma(t)$ has a continuous extension to $[0,1]$. Since $\theta(t)$ is also continuous on $[0,1]$, their composition $\dot{\gamma}(\theta(t))\gamma(\theta(t))$ is continuous on $[0,1]$. Therefore, the final expression for $\hat{c}(t)$ is continuous on $[0,1]$. This implies that $\hat{c}(t)$, initially defined only on $(0,1)$, has well-defined finite limits as $t \to 0^+$ and $t \to 1^-$, and thus admits a continuous extension to the compact interval $[0,1]$. ■

A.10 Proof of Theorem 13

Theorem 13. *Let $\widetilde{\varphi}$ and $\widetilde{\eta}$ be the approximations of φ and η respectively, and let $\gamma(t)$ and $c(t)$ satisfy the conditions in Lemma 12.*

Suppose that for all $t \in [0,1], x_0 \in H$, the mappings $x \mapsto \widetilde{\varphi}(t,x_0,x)$ and $x \mapsto \widetilde{\eta}(t,x_0,x)$ are Lipschitz continuous in H-norm, that is, there exists a constant $\widetilde{L} < \infty$ where for all $x, y \in H$,

$$\|\widetilde{\varphi}(t,x_0,x) - \widetilde{\varphi}(t,x_0,y)\|_H \leq \widetilde{L}\|x-y\|_H \text{ and } \|\widetilde{\eta}(t,x_0,x) - \widetilde{\eta}(t,x_0,y)\|_H \leq \widetilde{L}\|x-y\|_H.$$

Furthermore, suppose the CB-SDE has a unique strong solution X_t on $[0,\bar{t}] \subseteq [0,1]$ (see Propositions 6 or 7 for sufficient conditions) and let $\widetilde{X}_t$ be the unique strong solution to the CB-SDE when replacing the velocity φ and denoiser η with their approximations, solved with $\widetilde{X}_0 = X_0 = \xi_0$.

Then, the expected squared Wasserstein distance $\mathscr{W}_2^2(\bar{t})$ between the law of the approximate path $\widetilde{X}_t$ and the law of the conditional interpolant $\mu_{t|0}(\mathrm{d}x_t\,, \xi_0)$ at time $t = \bar{t}$ is bounded by:

$$\mathscr{W}_2^2(\bar{t}) \leq 2\bar{c}^2 e^{2\bar{c}\widetilde{L}+1} \int_0^{\theta^{-1}(\bar{t})} \mathrm{TVM}_{\theta(t)}(\widetilde{\varphi}) + \mathrm{TDM}_{\theta(t)}(\widetilde{\eta})\,\mathrm{d}t\,, \tag{3.18}$$

where

$$\bar{c} := \max_{t\in[0,\theta^{-1}(\bar{t})]} \left(\dot{\theta}(t) + |\hat{c}(t)|\right) < \infty.$$

Proof. Let $\widetilde{X}_t$ be the solution to the CB-SDE when using the approximate velocity $\widetilde{\varphi}$ and denoiser $\widetilde{\eta}$ to form the approximate drift $\widetilde{f}(t,x_0,x) := \widetilde{\varphi}(t,x_0,x) + c(t)\widetilde{\eta}(t,x_0,x)$. From Theorem 4, we know that the law of X_t is equal to $\mu_{t|0}(\mathrm{d}x_t\,, \xi_0)$. Hence, we couple $\widetilde{X}_t$ with X_t via the same C-Wiener process W_t and analyse the expected squared distance between these processes.

We consider the TC-CB-SDE, which has a unique solution $\hat{X}_t = X_{\theta(t)}$ on the interval $[0, \theta^{-1}(\bar{t})]$ when driven by the Wiener process $\hat{W}_t = W_{\theta(t)}$. Let $\hat{\widetilde{X}}_t = \widetilde{X}_{\theta(t)}$ be the time-changed approximate counterpart. Applying Ito's lemma (Da Prato and Zabczyk, 2014, Theorem 4.2) to $E(t) := \mathbb{E}\left[\left\|\hat{\widetilde{X}}_t - \hat{X}_t\right\|_H^2\right]$, we have

$$\frac{\mathrm{d}E(t)}{\mathrm{d}t} = 2\,\mathbb{E}\left[\left\langle \hat{\widetilde{X}}_t - \hat{X}_t, \widetilde{f}(\theta(t),\xi_0,\hat{\widetilde{X}}_t) - f(\theta(t),\xi_0,\hat{X}_t)\right\rangle_H \dot{\theta}(t)\right].$$

We add $0 = -\widetilde{f}(\theta(t),\xi_0,\hat{X}_t) + \widetilde{f}(\theta(t),\xi_0,\hat{X}_t)$ to the second argument of the inner product to split this into two terms:

$$\begin{aligned}\frac{\mathrm{d}E(t)}{\mathrm{d}t} = &\overbrace{2\,\mathbb{E}\left[\left\langle \hat{\widetilde{X}}_t - \hat{X}_t, \widetilde{f}(\theta(t),\xi_0,\hat{\widetilde{X}}_t) - \widetilde{f}(\theta(t),\xi_0,\hat{X}_t)\right\rangle_H \dot{\theta}(t)\right]}^{\text{propagation error term}} \\ &\underbrace{+\,2\,\mathbb{E}\left[\left\langle \hat{\widetilde{X}}_t - \hat{X}_t, \widetilde{f}(\theta(t),\xi_0,\hat{X}_t) - f(\theta(t),\xi_0,\hat{X}_t)\right\rangle_H \dot{\theta}(t)\right]}_{\text{training error term}}.\end{aligned} \tag{A.19}$$

We place a bound on each term using the Cauchy-Schwarz inequality. For the propagation error term, we make use of the fact that $\widetilde{\varphi}$ and $\widetilde{\eta}$ are Lipschitz continuous, so that for each

$t \in [0, \theta^{-1}(\bar{t})]$ and $x_0 \in H$, the mapping

$$x \mapsto \widetilde{f}(\theta(t), x_0, x)\dot{\theta}(t) = \widetilde{\varphi}(\theta(t), x_0, x)\dot{\theta}(t) + \hat{c}(t)\widetilde{\eta}(\theta(t), x_0, x)$$

is Lipschitz continuous in H-norm with Lipschitz constant $L(t) = \max\left\{\dot{\theta}(t), \hat{c}(t)\right\}\widetilde{L}$. Since by Lemma 12, both $\hat{c}$ and $\dot{\theta}$ are continuous on the compact interval $[0, \theta^{-1}(\bar{t})]$, the uniform bound $\bar{c} \coloneqq \max_{t \in [0, \theta^{-1}(\bar{t})]} \left(\dot{\theta}(t) + |\hat{c}(t)|\right)$ is finite. It follows that

$$2\,\mathbb{E}\left[\left\langle \hat{\widetilde{X}}_t - \hat{X}_t, \widetilde{f}(\theta(t), \xi_0, \hat{\widetilde{X}}_t) - \widetilde{f}(\theta(t), \xi_0, \hat{X}_t)\right\rangle_H \dot{\theta}(t)\right] \leq 2\bar{c}\widetilde{L}E(t). \tag{A.20}$$

For the training error term, we have

$$\begin{aligned}
&2\,\mathbb{E}\left[\left\langle \hat{\widetilde{X}}_t - \hat{X}_t, \widetilde{f}(\theta(t), \xi_0, \hat{X}_t) - f(\theta(t), \xi_0, \hat{X}_t)\right\rangle_H \dot{\theta}(t)\right] \\
&\quad \leq 2\,\mathbb{E}\left[\left\|\hat{\widetilde{X}}_t - \hat{X}_t\right\|_H \left\|\left(\widetilde{f}(\theta(t), \xi_0, \hat{X}_t) - f(\theta(t), \xi_0, \hat{X}_t)\right)\dot{\theta}(t)\right\|_H\right] \\
&\quad \leq E(t) + \mathbb{E}\left[\left\|\widetilde{f}(\theta(t), \xi_0, \hat{X}_t) - f(\theta(t), \xi_0, \hat{X}_t)\dot{\theta}(t)\right\|_H^2\right] \\
&\quad \leq E(t) + 2\dot{\theta}^2(t)\mathrm{TVM}_{\theta(t)}(\widetilde{\varphi}) + 2\hat{c}^2(t)\mathrm{TDM}_{\theta(t)}(\widetilde{\eta}) \\
&\quad \leq E(t) + 2\bar{c}^2\left(\mathrm{TVM}_{\theta(t)}(\widetilde{\varphi}) + \mathrm{TDM}_{\theta(t)}(\widetilde{\eta})\right)
\end{aligned} \tag{A.21}$$

Substituting our bounds on the propagation error term (Equation A.20) and training error term (Equation A.21) into Equation (A.19), we have

$$\frac{\mathrm{d}E(t)}{\mathrm{d}t} \leq (2\bar{c}\widetilde{L} + 1)E(t) + 2\bar{c}^2\left(\mathrm{TVM}_{\theta(t)}(\widetilde{\varphi}) + \mathrm{TDM}_{\theta(t)}(\widetilde{\eta})\right).$$

Applying Groenwall's inequality to $E(t)$ on the interval $[0, \theta^{-1}(\bar{t})]$, we have

$$\begin{aligned}
\mathscr{W}_2^2(\bar{t}) &\leq \mathbb{E}\left[\left\|\overline{\widetilde{X}}_t - \overline{X}_t\right\|_H^2\right] \\
&= E(\theta^{-1}(\bar{t})) \leq 2\bar{c}^2 \int_0^{\theta^{-1}(\bar{t})} \left(\mathrm{TVM}_{\theta(t)}(\widetilde{\varphi}) + \mathrm{TDM}_{\theta(t)}(\widetilde{\eta})\right) e^{(2\bar{c}\widetilde{L}+1)(\theta^{-1}(\bar{t})-t)}\,\mathrm{d}t \\
&\leq 2\bar{c}^2 e^{2\bar{c}\widetilde{L}+1} \int_0^{\theta^{-1}(\bar{t})} \mathrm{TVM}_{\theta(t)}(\widetilde{\varphi}) + \mathrm{TDM}_{\theta(t)}(\widetilde{\eta})\,\mathrm{d}t
\end{aligned}$$

This concludes the proof. ■

A.11 Proof of Lemma 14

Lemma 14. *A strictly increasing, bijective, continuously differentiable time change function $\theta(t)$ on $[0,1]$ is a valid change-of-time ensuring that $\hat{c}(t)$ is finite on $[0,1]$ if and only if $\theta(t)$ satisfies the following conditions.*

1. $\lim_{t\to 1^-} \frac{\dot{\theta}(t)}{2(1-t)} < \infty$*; and*
2. $\lim_{t\to 0^+} \frac{\dot{\theta}(t)}{2t} < \infty$ *if* $\varepsilon \neq \frac{b}{2}$.

Proof. We have:

$$\hat{c}(t) = \frac{b-2\varepsilon}{2\sqrt{b\theta(t)(1-\theta(t))}}\dot{\theta}(t) - \sqrt{\frac{b\theta(t)}{1-\theta(t)}}\dot{\theta}(t) \tag{A.22}$$

By inspection, $\hat{c}(t)$ is finite on all $(0,1)$ since $\theta(t) \in (0,1)$ and $\dot{\theta}(t)$ is continuous. We analyse the limits at the endpoints by considering each case in turn.

First, when $\varepsilon = \frac{b}{2}$, the first term of Equation (A.22) vanishes, reducing analysis to the second term. This is finite on $[0,1)$, so we need only consider the limit $t \to 1^-$. This is finite if and only if $\lim_{t\to 1^-} q(t)$ is finite, where $q(t) \coloneqq \frac{\dot{\theta}(t)}{\sqrt{1-\theta(t)}}$. Using a substitution $y(t) = \sqrt{1-\theta(t)}$, we have $q(t) = -2\dot{y}(t)$ and hence

$$\begin{aligned}\lim_{t\to 1^-} q(t) < \infty &\iff \lim_{t\to 1^-} -\dot{y}(t) = \lim_{t\to 1^-} \frac{y(t)}{1-t} < \infty \\ &\iff \lim_{t\to 1^-} \frac{y^2(t)}{(1-t)^2} = \lim_{t\to 1^-} \frac{1-\theta(t)}{(1-t)^2} = \lim_{t\to 1^-} \frac{\dot{\theta}(t)}{2(1-t)} < \infty.\end{aligned}$$

Therefore, when $\varepsilon = \frac{b}{2}$, $\hat{c}(t)$ has a finite continuous extension on $[0,1]$ if and only if condition (1) in Lemma 14 holds.

Now we consider the case $\varepsilon \neq \frac{b}{2}$. We now additionally require the first term in Equation (A.22) to have finite limits at the endpoints. In the limit $t \to 1^-$, the first term is finite if and only if $\lim_{t\to 1^-} \frac{\dot{\theta}(t)}{\sqrt{1-\theta(t)}}$ is finite, which is the same condition as discussed above. It remains to check the limit $t \to 0^+$. Here, the first term is finite if and only if $\lim_{t\to 0^+} r(t) < \infty$, where $r(t) \coloneqq \frac{\dot{\theta}(t)}{\sqrt{\theta(t)}}$. Considering a substitution $u(t) \coloneqq \sqrt{\theta(t)}$, we have $r(t) = 2\dot{u}(t)$ and hence

$$\begin{aligned}\lim_{t\to 0^+} r(t) < \infty &\iff \lim_{t\to 0^+} \dot{u}(t) = \lim_{t\to 0^+} \frac{u(t)}{t} < \infty \\ &\iff \lim_{t\to 0^+} \frac{u^2(t)}{t^2} = \lim_{t\to 0^+} \frac{\theta(t)}{t^2} = \lim_{t\to 0^+} \frac{\dot{\theta}(t)}{2t} < \infty.\end{aligned}$$

Hence, when $\varepsilon \neq \frac{b}{2}$, $\hat{c}(t)$ has a finite continuous extension on $[0,1]$ if and only if both conditions (1) and (2) in Lemma 14 hold. 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.