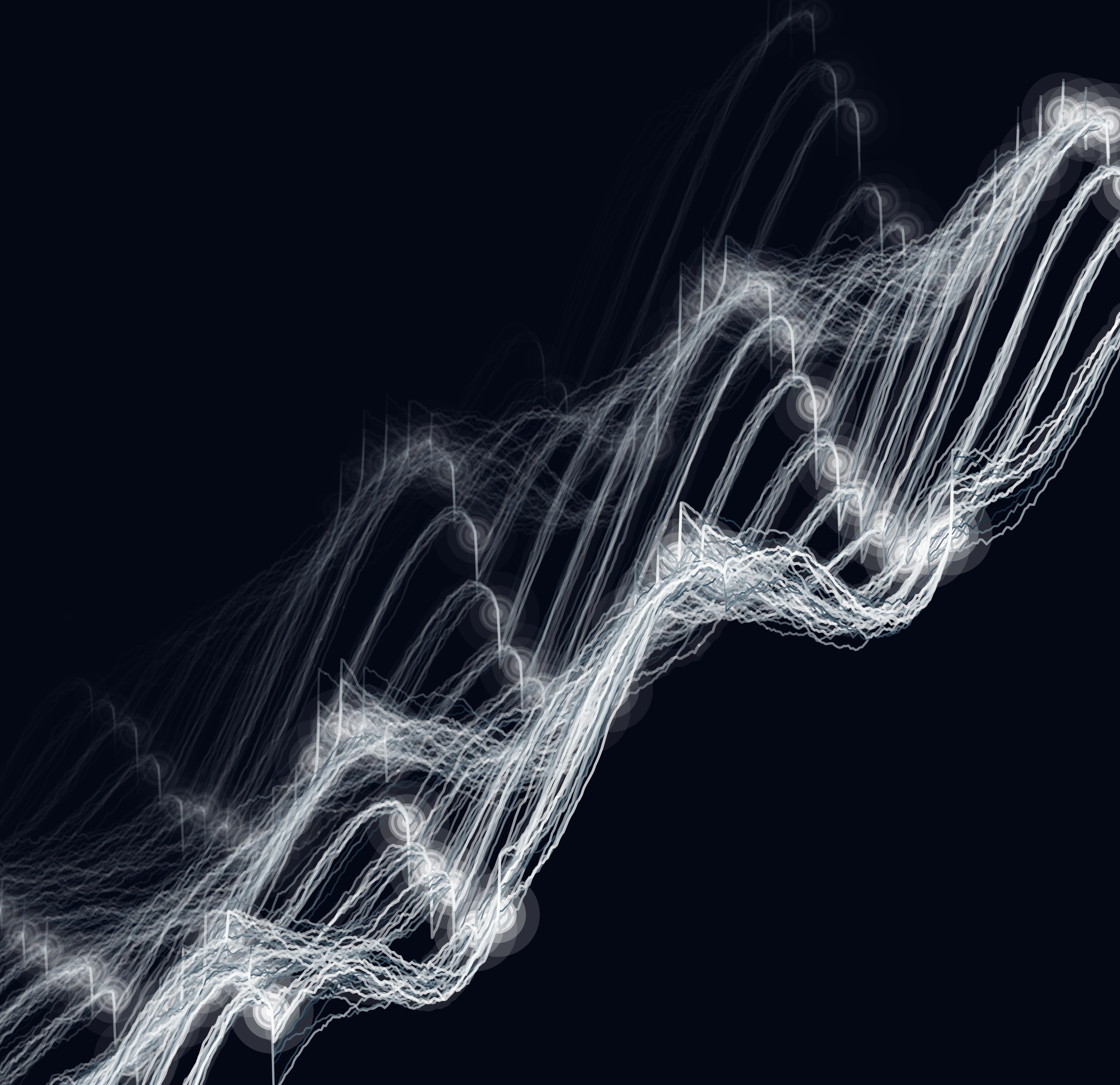
Bayesian Computation for Stochastic Partial Differential Equations

Thorben Pieper-Sethmacher



Bayesian Computation for Stochastic Partial Differential Equations

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus
Prof.dr.ir. H. Bijl;
Chair of the Board for Doctorates

to be defended publicly on
Monday, 11 May 2026 at 12:30

by

Thorben PIEPER-SETHMACHER

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr. A.W. van der Vaart	Delft University of Technology, *promotor*
Prof.dr.ir. F.H. van der Meulen	Vrije Universiteit Amsterdam, *promotor*

Independent members:

Prof.dr.ir. G. Jongbloed	Delft University of Technology
Prof.dr. G.J. Lord	Radboud University
Prof.dr. M. Reiß	Humboldt University, Germany
Dr. M.V. Gnann	Delft University of Technology
Dr. D. Paulin	Nanyang Technological University, Singapore
Prof.dr.ir. M.C. Veraar	Delft University of Technology, *reserve member*

The research in this dissertation was supported by the NWO Spinoza prize awarded to A.W. van der Vaart by the Netherlands Organisation for Scientific Research (NWO).

Contents

Chapter 1

Introduction

1.1 A motivating example

Roughly five years ago, in early 2020, I began to work on my Master's thesis in collaboration between TU Delft and the University of Sydney. Just a few months prior to that, the state of New South Wales (NSW) had seen bushfires "that were some of the worst in the world and in recorded history" (Owens and O'Kane [88]).
A key challenge in fighting an active bushfire is the real-time estimation and prediction of its spread. The NSW government's bushfire inquiry report states in their executive summary the "need to push available technologies harder, especially fire science, remote sensing, data science and artificial intelligence to equip us better to understand what happens during a bush fire and respond more quickly." (Owens and O'Kane [88, page v])
Although recent advances in remote sensing capabilities have improved online monitoring of a moving firefront, the fire is typically only observed partially and noisily as well as sparsely in both time and space. As a reference example, the data gathered by NASA's Landsat and MODIS satellite programs correspond respectively to a temporal resolution of approximately 16 days at a spatial resolution of 30m (Landsat 8, Earth Resources Observation and Science (EROS) Center [48]) and a temporal resolution of 24 hours at a spatial resolution of 1km (MODIS, NASA VIIRS Land Science Team [86]). Moreover, smoke coverage produced by severe fires may cause a significant amount of missing data points.
From a mathematical viewpoint, the estimation and prediction of an evolving firefront based on sparse measurements defines a statistical problem. Specifically, it requires estimating a spatio-temporal stochastic process $\{X(t,\xi) : t \geq 0,\, \xi \in D\}$ that represents the 'fire intensity' $X(t,\xi)$ at time $t \geq 0$ and location $\xi \in D$ in a spatial domain $D \subset \mathbb{R}^2$. Developing solutions to this estimation task was the original aim of the proposed Master's project.

At the heart of the statistical problem lies a stochastic model of X. A promising approach is the assumption that X satisfies a *stochastic partial differential equation (SPDE)*

$$\mathcal{F}(X)(t,\xi) = \dot{W}(t,\xi)$$

where $\mathcal{F}$ is a (nonlinear) differential operator and $\dot{W}(t,\xi)$ represents a stochastic forcing term, typically Gaussian, that incorporates intrinsic noise into the dynamics of the model. The motivation behind such an SPDE model is to marry the physical dynamics of a spreading fire, described by the differential operator $\mathcal{F}$, with the treatment of X as a spatio-temporal stochastic process. The stochastic noise $\dot{W}$ may represent the system's

inherent stochasticity, model uncertainty or extrinsic stochastic factors acting on the process.
In simplified models (see e.g. Asensio and Ferragut [7], Mandel et al. [83] and references within), PDEs describing the propagation of wildfires are based on reaction-convection-diffusion operators $\mathcal{F}$ of the general form

$$\mathcal{F}(X)(t,\xi) = \partial_t X(t,\xi) - [\Delta X(t,\xi) - v(t,\xi)\cdot\nabla X(t,\xi) + F(X)(t,\xi)],$$

with Laplacian $\Delta = \partial_{\xi_1}^2 + \partial_{\xi_2}^2$, Nabla operator $\nabla = (\partial_{\xi_1}, \partial_{\xi_2})$ and wind vector field $v(t,\xi)$. The function $F(X)$ is a reaction term that models combustion and typically depends on a coupled process that represents available fuel load.
The first challenge of this approach lies in the fact that, to make any attempt at modelling the complex behaviour of a spreading wildfire, the reaction term F is highly irregular and nonlinear. This irregularity severely complicates the application of classical SPDE theory, rendering the existence of a solution - whether through a weak, mild or rough path formulation - a non-trivial problem in itself.

This 'solution question' put aside, the bigger challenge during the proposed Master's project was the apparent lack of statistical methodology for estimation problems concerning stochastic partial differential equations. A few exceptions were to be found for inference of model parameters in the frequentist literature, as will be reviewed in more detail further below.
Additionally, in the literature for spatio-temporal statistics, there has been a growing interest in so-called 'physically inspired spatio-temporal models'. In this context, formulating spatio-temporal processes as solutions to linear SPDEs has proven a computationally efficient alternative to traditional approaches based on Gaussian process regression. However, the methodology developed in that direction does not generalise to nonlinear and non-Gaussian models, rendering it unsuitable for a highly nonlinear system such as an evolving firefront.

Eventually, we judged an SPDE-based approach to estimate the spread of a bushfire infeasible for the scope of a Master's thesis. Instead, I turned my attention to statistical inference for linear SPDEs. I remember joking with a colleague at the time that I would "never touch the nonlinear stuff, because that seems way too difficult". This PhD thesis is the outcome of an attempt at doing exactly that.

1.2 Framework and problem statement

The process X of interest throughout this thesis is assumed to be the unique solution to a semilinear stochastic partial differential equation. There are various approaches on how such a solution is to be understood. A common approach, and the one followed in this work, is the so-called *mild solution framework*, in which an SPDE is phrased as an infinite-dimensional stochastic differential equation

$$\begin{cases} \mathrm{d}X_t &= [AX_t + F(t,X_t)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t, \\ X_0 &= x_0 \in H. \end{cases} \tag{1.1}$$

It is infinite-dimensional in the sense that a solution X_t to Equation (1.1) takes values in an infinite-dimensional Hilbert space H, typically a function space over some spatial domain. The operator A is assumed to be the generator of a strongly continuous semigroup $(S_t)_{t\geq 0}$ on H, whereas F is a 'sufficiently regular' nonlinear function. The process W is a cylindrical Wiener process on H, defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and the operator Q is a bounded linear operator on H that represents the spatial covariance structure of the noise. A mild solution X_t to Equation (1.1) is a stochastic process that satisfies the integral equation

$$X_t = S_t x_0 + \int_0^t S_{t-s} F(s, X_s)\,\mathrm{d}s + \int_0^t S_{t-s} Q^{\frac{1}{2}}\,\mathrm{d}W_s. \tag{1.2}$$

Conditions for the existence and uniqueness of mild solutions are well established in the literature, see for example Chapter 7 in the classical reference Da Prato and Zabczyk [41]. It is assumed throughout that the necessary conditions for the existence and uniqueness of the mild solution X are fulfilled.

If Equation (1.1) represents an SPDE, the operator A is a linear differential operator, typically defined only on a dense domain $\mathrm{dom}(A) \subset H$, although other situations in which A is a bounded linear operator are also possible. In both cases, borrowing from the terminology for stochastic differential equations in Euclidean state spaces, the mild solution to Equation (1.1) might be referred to as an *infinite-dimensional diffusion process*.

Besides the previously given example, semilinear SPDEs of the form (1.1) arise naturally in a wide range of applied disciplines such as in finance (Carmona and Tehranchi [19], Cont [34], Benth and Paraschiv [11]), biology (Spill et al. [106], Infante et al. [69], Altmeyer et al. [2]), fluid dynamics (Da Prato and Debussche [40], Barbu et al. [8], Llopis et al. [78]) and neuroscience (Walsh [111], Coombes et al. [36], Faugeras and Inglis [53]).

In many real-world applications, a system that is modelled by a process X satisfying (1.2) is observed indirectly at observation times $0 < t_1 < \ldots < t_n = T$ through random variables Y_i defined by

$$Y_i \mid X_{t_i} \sim k_i(X_{t_i}, \cdot), \quad i = 1, \ldots, n, \tag{1.3}$$

where k_i denotes a Markov kernel from H to $\mathbb{R}^{m_i}$. This includes the special case that one observes the Gaussian variables

$$Y_i \mid X_{t_i} \sim \mathcal{N}(L_i X_{t_i}, \Sigma_i),$$

where $L_i : H \to \mathbb{R}^{m_i}$ are bounded linear *observation operators* and Σ_i, $i = 1, \ldots, n$ are positive-definite matrices. If the noise degenerates, one might also observe a partial representation $L_i X_{t_i}$ of X_{t_i} directly, in which case $k_i(X_{t_i}, \cdot) = \delta_{L_i X_{t_i}}$ is the Dirac measure at $L_i X_{t_i}$. Common observation schemes include observing localised measurements of X, represented by the observation operator

$$(Lx)_j = \int \rho_j(\xi) x(\xi)\,\mathrm{d}\xi, \quad x \in H,\ j = 1, \ldots, m_i, \tag{1.4}$$

with weight functions ρ_j, or observing a projection onto a finite dimensional subspace

$$Lx = \sum_{j=1}^{m_i} \langle x, e_j \rangle e_j, \quad x \in H, \tag{1.5}$$

for an orthonormal basis $(e_j)_{j=1}^{\infty}$ of H. This is typical, for example, in low pass filters.

Jointly, the SPDE model of X in Equation (1.1) and the observation scheme (1.3) define multiple statistical problems. Firstly, there is the problem of state estimation of the latent continuous-time, continuous-space signal X based on the discrete observations $Y = (Y_1, \ldots, Y_n)$. This problem is usually considered in two settings. In the *online* setting, the distribution of the current state of X_{t_i} is inferred based on all previous observations, i.e. one estimates

$$\mathbb{P}(X_{t_i} \in B \mid Y_1, \ldots, Y_i), \quad i = 1, \ldots, n, \tag{1.6}$$

where $B \subset H$ is a Borel set. This is known as the *filtering problem* and the distribution defined by (1.6) as the *filtering distribution*. The motivating example of estimating an evolving firefront based on satellite imagery of the fire's intensity falls within this setting. On the other hand, in the *offline* setting, the *smoothing problem* aims to estimate the complete path measure of X based on the full set of observations, i.e. one infers the *smoothing distribution*

$$\mathbb{P}(X \in \mathcal{B} \mid Y_1, \ldots, Y_n) \tag{1.7}$$

for any Borel set $\mathcal{B} \subset C([0,T]; H)$. Here, $C([0,T]; H)$ denotes the space of continuous functions from $[0,T]$ to H.

On top of state estimation, there is the question of parameter estimation. Typically, any SPDE model (1.1) involves a vector of parameters $\theta \in \mathbb{R}^p$ that parametrise the operators $A = A_\theta$, $F = F_\theta$ and $Q = Q_\theta$. These might, for example, relate to physical constants that determine the dynamics of the system. If the parameters θ are unknown, they are to be estimated, possibly jointly with X, based on the observations Y. This allows for efficient model calibration of SPDEs based on real-world data, hence improving their predictive power, consistency with the underlying physical systems and applicability in the sciences and engineering.

1.3 State of the literature

Broadly speaking, the related literature for this thesis can be categorised into three fields. Firstly, there has been a recent surge of interest in the frequentist literature on statistical inference for SPDEs. In fact, a survey by Cialenco [28] reveals that the majority of research in statistics for SPDEs falls within this category. Given an SPDE model (1.1) and an observation scheme thereof, a point estimator for a parameter θ that parametrises either A_θ, F_θ or Q_θ is constructed and its asymptotic properties studied as the given observations tend to a fully observed path of X. Typically, due to the difficulty of the problem, parameters are treated individually and not estimated jointly. Common observation schemes are the *spectral approach* of observing a finite dimensional approximation of X via basis functions of A_θ as in Eq. (1.5), the *localised approach* in which one observes a continuous path of X, integrated in the spatial domain with respect to a localising kernel akin to (1.4) and the *discrete* scheme of observing $X(t_i, \xi_j)$ at discrete points in space and time.

Of high importance is the fact that, if θ parametrises the drift A_θ or volatility Q_θ, the path measures $\mathcal{L}_\theta(X)$ of X on $C([0,T]; H)$ are singular for different values of θ. This turns the estimation of θ into a singular problem, rendering it particularly attractive from the frequentist perspective. Indeed, early works in Huebner et al. [67], Huebner and Rozovskii [68] and Lototsky [80] focus on parametrisations of the drift A_θ for linear SPDEs in the spectral approach. In the localised observation scheme, non-parametric estimation of a linear drift has been considered in Altmeyer and Reiß [4] and parametric estimation in the case of multiplicative noise in Janák and Reiß [70]. Moreover, Cialenco and Huang [30], Hildebrandt and Trabs [64] and Kaino and Uchida [74] estimate drift parametrisations for linear SPDEs based on discrete observations. Parameter estimation for volatility parameters based on discrete observations of a linear SPDE has been considered in Bibinger and Trabs [15] and Chong [25].

For semilinear SPDEs, estimation of a parameter in the drift has been considered in the spectral approach for the two-dimensional stochastic Navier-Stokes equation in Cialenco and Glatt-Holtz [29] and for reaction-diffusion equations in Pasemann and Stannat [93]. Additionally, the latter has been studied in the localised observation scheme in Altmeyer et al. [3] and in the discretised observation scheme in Cialenco et al. [31].

On the other hand, there are few results concerning estimation for the nonlinearity F_θ. Notable exceptions are given for localised observations in Gaudlitz and Reiß [56] and Gaudlitz [55] and for discrete observations in Hildebrandt and Trabs [65].
Despite differences in their observation schemes and estimation target, most approaches study the asymptotic behaviour of point estimators for θ as the given observations tend to a fully observed path of X, often exploiting the fact that θ parametrises the model (1.1) in a 'nice', tractable way. In contrast, if observations are sparse and possibly noisy or the parametrisation of the SPDE is more complex, the performance of such estimators is uncertain. In that case, the flexibility and uncertainty quantification inherent to the Bayesian paradigm makes it an attractive choice.

The Bayesian approach is commonly found in the literature for spatio-temporal statistics. Inherently, the focus here lies on state estimation of a spatio-temporal signal X, given sparse observations thereof. Additionally, parameter estimation might be included in a Bayesian hierarchical model. In this context, SPDEs are used as a building block for so-called *dynamic spatio-temporal models*, see Wikle et al. [114], Chapter 5 for a recent introduction.
There are two key aspects that motivate this approach. Firstly, it is a natural way to build spatio-temporal statistical models that incorporate physical dynamics of the system at hand. Secondly, the computational complexity of the widely used *geostatistical* (or *kriging*) approach, an approach based on Gaussian process (GP) regression, renders it advantageous to reformulate spatio-temporal GPs as solutions to SPDEs. Making use of the inherent Markovian property of SPDEs, computational costs can then be drastically reduced, leading to methods that are particularly attractive for large spatio-temporal datasets.
Early work in this direction has been carried out by Jones and Zhang [73] for convection-diffusion models. A link between the mesh discretisation of a linear SPDE and Gaussian Markov random fields has been established in Lindgren et al. [76]. The interpretation of GP regression as a filtering and smoothing problem for linear SPDEs has been studied in Sarkka et al. [100]. In Sigrist et al. [103], spatio-temporal covariance kernels are constructed based on the spectral approximation of convection-diffusion SPDEs. This approach has been extended in Liu et al. [77] for non-stationary kernels and in Clarotto et al. [33] based on finite element and finite difference approximations.
However, to stay within the framework of Gaussian models, the methodology developed in this direction concerns linear SPDEs only and does not generalise to nonlinear and non-Gaussian models. Moreover, the general aim is not to infer the infinite-dimensional exact solution of an SPDE. Instead, after a spatio-temporal model based on an SPDE is formulated, a model approximation using, for example, a basis function decomposition or discretised spatial mesh is utilised and statistical estimation is pursued within this approximation only.
Notable exceptions that address state estimation for nonlinear SPDEs in the online setting can be found in Llopis et al. [78] and Lang et al. [75]. Here, particle filtering schemes - designed to handle the high dimensionality of the problem - have been proposed to solve the filtering problem for the stochastic Navier-Stokes equation and stochastic rotating shallow water model, respectively.

The literature most relevant to this thesis concerns Bayesian inference for stochastic ordinary differential equations (SODEs). The mild solution framework of formulating SPDEs as infinite-dimensional SDEs makes this a natural starting point and much attention will be paid to the extension of methods that have been developed for finite-dimensional diffusion processes to the infinite-dimensional case. The literature in this

field has grown extensive over the past two decades. In the following, a few key contributions that are relevant to this work are highlighted. A comprehensive overview can be found in the recent survey by Craigmile et al. [38]. Additionally, subsequent chapters include short summaries of relevant literature for the chapter at hand.
Early work on sampling diffusion bridges, the special case of the smoothing problem with one full and noiseless observation, can be found in the context of parameter estimation for discretely observed diffusion processes in Elerian et al. [49], Eraker [50] and Roberts and Stramer [97]. Other works in this direction include Clark [32], Delyon and Hu [44], Beskos et al. [14], Beskos et al. [13], Papaspiliopoulos and Roberts [90], Cotter et al. [37] and, more recently, Schauer et al. [101], Whitaker et al. [113] and Heng et al. [63]. For hypoelliptic diffusions, bridge sampling has been considered in Hairer et al. [62] and Bierkens et al. [16]. More generally, the joint smoothing and parameter estimation problem has been studied in van der Meulen and Schauer [110], Mider et al. [85] and Graham et al. [60]. On the other hand, results concerning online state estimation for SDEs based on discrete observations can be found in the particle filtering literature, for example in Sottinen and Särkkä [105], Fearnhead et al. [54], Beskos et al. [12], Ruzayqat et al. [99], Jasra et al. [71] and Chopin et al. [26].

1.4 Approach and contribution

This thesis contributes to the emerging literature on statistical inference for infinite-dimensional diffusion processes and stochastic PDEs. Both state and parameter estimation are considered in a fully Bayesian approach and methodology to estimate the joint posterior of $(\theta, X) \mid Y$ is developed, whereby the SPDE (1.1) is treated as defining a prior on the sample paths of X and a prior on θ is specified as needed.

A crucial aspect of this thesis is that all results concerning the conditioned process $X^\star = X \mid Y$ are derived in the fully infinite-dimensional setting. This leads to methodology that infers the law $\mathcal{L}(X^\star)$ of $X^\star$ on the path space $C([0,T];H)$.
Of course, in practical terms, any implementation of an algorithm that estimates $\mathcal{L}(X^\star)$ relies on the discretisation of $X^\star$ in both space and time. However, as has been shown in the context of sampling finite-dimensional diffusion bridges, methods that are developed in the function space setting prove to be more robust to the curse of dimensionality. As a result, algorithms are obtained that outperform traditional 'discretisation-first' algorithms as the temporal discretisation mesh size tends to zero (cf. Beskos et al. [13], Cotter et al. [37]).
Since the SPDE case concerns discretisation in both space and time, the need to develop methodology in the infinite-dimensional setting is even more critical.

Following this, the first question arises in exactly how a mild solution X, conditioned on the observations Y, is to be understood. This is one of the central questions addressed in Chapter 2. It has received surprisingly little attention in the literature so far, with exceptions in Simão [104] and Goldys and Maslowski [59] for the case of linear SPDEs. Consider instead the well-studied problem of conditioning a solution $(x_t)_{t\geq 0}$ to an SODE

$$\mathrm{d}x = b(t, x_t)\,\mathrm{d}t + \sqrt{q}\,\mathrm{d}w_t.$$

Assume one observes the full state $x_T = y$ at some time $T > 0$, in which case the conditioned process $(x_t^\star)_{t\geq 0}$ defines a diffusion bridge. It is a well-known result that the bridge process is characterised by yet another SODE given by

$$\mathrm{d}x_t^\star = b(t, x_t^\star)\,\mathrm{d}t + q\,\mathrm{D}_x \log h(t, x_t^\star)\,\mathrm{d}t + \sqrt{q}\,\mathrm{d}w_t, \quad t < T,$$

where the function h is defined by $h(t,x) = p(t,x;T,y)$ with p denoting the transition density of $(x_t)_{t\geq 0}$ (see, for example, Rogers and Williams [98]). This diffusion bridge equation can be derived by applying *Doob's h-transform* (cf. Doob [46]), which defines a unique change of measure $\mathbb{P}^h$ on $\mathcal{F}_T$ via

$$\frac{\mathrm{d}\mathbb{P}^h}{\mathrm{d}\mathbb{P}}_{|\mathcal{F}_t} = \frac{h(t,x_t)}{h(0,x_0)}, \quad t < T.$$

It turns out that Doob's h-transform is a special case of a general class of exponential changes of measures $\mathbb{P}^h$, defined with respect to $\mathbb{P}$ on $\mathcal{F}_T$ by the Radon–Nikodym density process

$$E_t^h = \frac{h(t,x_t)}{h(0,x_0)} \exp\left(- \int_0^t \frac{Lh}{h}(s,x_s)\,\mathrm{d}s \right), \quad t < T. \tag{1.8}$$

The operator L denotes the infinitesimal space-time generator of the process $(x_t)_{t\geq 0}$, acting on suitably smooth functions φ as the Kolmogorov differential operator

$$(L_0\varphi)(t,x) = \partial_t\varphi(t,x) + \langle b(t,x), \mathrm{D}_x\varphi(t,x)\rangle + \frac{1}{2}\operatorname{tr}[q\,\mathrm{D}_x^2\varphi(t,x)].$$

For Markov processes in Euclidean state spaces, this change of measure has been studied in detail in Palmowski and Rolski [89], where suitable conditions are established under which E^h is a martingale and hence the measure $\mathbb{P}^h$ well defined.

In Chapter 2, the measure transformations defined by martingales E^h of the form (1.8) are generalised to the setting of SPDEs. This generalisation poses three key challenges. Firstly, the transition semigroup of an SPDE is not strongly continuous in the norm topology of any suitable function space (cf. Goldys and Kocan [58]). This raises the question in what sense the infinitesimal generator L of X is to be defined. One way to address this issue is to consider the transition semigroup in a weaker topology on a suitable function space. The approach presented here adopts the topology of *bounded pointwise convergence* on the space of continuous functions $\varphi(t,x)$ of at most polynomial growth in x. In this setting, transition semigroups and their generators for SPDEs have been studied previously in Priola [96] and Manca [82].

Secondly, since X is assumed to be a mild solution, essential tools from classical Itô calculus, such as Itô's lemma, are unavailable. One therefore needs to approximate the mild solution X by a sequence of strong solutions to SPDEs, defined through the *Yosida approximations* of the operator A in (1.1).

Lastly, the assumption that $h(t,x)$ is regular enough to be in the domain of the Kolmogorov operator L_0 plays a critical role in deriving the diffusion bridge SODE from the change of measure $\mathbb{P}^h$. For SPDEs, the equivalent differential operator

$$(L_0\varphi)(t,x) = \partial_t\varphi(t,x) + \langle Ax + F(t,x), \mathrm{D}_x\varphi(t,x)\rangle + \frac{1}{2}\operatorname{tr}[Q\,\mathrm{D}_x^2\varphi(t,x)]$$

includes the unbounded differential operator A. Hence, one requires that there exists a continuous extension of the mapping

$$\mathbb{R}_+ \times \operatorname{dom}(A) \to \mathbb{R},\ (t,x) \mapsto \langle Ax, \mathrm{D}_x\varphi(t,x)\rangle.$$

This severely limits the class of functions for which $L_0\varphi$ is well-defined. In particular, for the applications of our interest, one can generally not assume that h is regular enough for $L_0 h$ to exist. Instead, approximations of h by a suitable class of test functions are required.

Ultimately, the main result of Chapter 2, given in Theorem 2.16, states sufficient assumptions on any function h in the domain of L under which the change of measure $\mathbb{P}^h$ is well-defined. Moreover, the theorem provides sufficient conditions for $\mathbb{P}^h$ to define a Girsanov-type change of measure, under which X is a mild solution to

$$\mathrm{d}X_t^h = \left[AX_t^h + F(t,X_t^h) + Q\,\mathrm{D}_x \log h(t,X_t^h)\right]\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t^h, \quad t < T,$$

for some $\mathbb{P}^h$-Wiener process W^h. An application of Theorem 2.16 in Section 2.4.1 derives the *infinite-dimensional diffusion bridge* (or *SPDE bridge*) equation

$$\mathrm{d}X_t^\star = [AX_t^\star + F(t, X_t^\star) + Q\,\mathrm{D}_x \log p_X(t, X_t^\star; T, y)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t^\star \tag{1.9}$$

of X given $X_T = y$. Crucially, since the Lebesgue reference measure is unavailable in infinite-dimensional state spaces, sufficient conditions are given for the transition density p_X of X to be well-defined with respect to a Gaussian reference measure. The results of Chapter 2 led to the following publication;

> Pieper-Sethmacher, T., F. van der Meulen and A. van der Vaart (2025). On a class of exponential changes of measure for stochastic pdes. *Stochastic Processes and their Applications 185*, 104630.

Chapter 3 continues the development of Chapter 2 and concerns sampling (or simulation) of the infinite-dimensional diffusion bridge $X^\star$, conditioned on one partial observation $LX_T = y$. Sampling from the diffusion bridge distribution may be considered a fundamental building block in addressing each of the three estimation problems studied in this thesis. Not only does it represent the 'simplest' smoothing case, but it also plays an essential role in parameter inference for S(P)DEs and the construction of proposal distributions in sequential importance sampling schemes within the filtering context.
It is a well-known issue in the literature for SODEs that sampling of diffusion bridges by simulating Equation (1.9) is generally infeasible due to the intractability of the transition density p_X. This has motivated several works in which p_X is replaced by a tractable function g in Equation (1.9), defining a so-called guided process X° (cf. Clark [32], Delyon and Hu [44], Schauer et al. [101], Mider et al. [85]).
The changes of measure introduced in Chapter 2 provide a natural framework to generalise this approach to the SPDE setting. If, for example, g is induced by the infinite-dimensional Ornstein-Uhlenbeck process Z, it defines a change of measure $\mathbb{P}^g$

$$\frac{\mathrm{d}\mathbb{P}^g}{\mathrm{d}\mathbb{P}}_{|\mathcal{F}_t} = \frac{g(t, X_t)}{g(0, x_0)} \exp\left(-\int_0^t \langle F(s, X_s), \mathrm{D}_x \log g(s, X_s)\rangle\,\mathrm{d}s\right)\mathrm{d}\mathbb{P}_t, \quad t < T,$$

under which X satisfies Equation (1.9) with p_X replaced by g and $\mathrm{D}_x \log g(s, x)$ being a tractable functional. This is shown in Proposition 3.6 and proves the existence of the infinite-dimensional guided process X°. As a consequence of the construction of $\mathbb{P}^h$ and $\mathbb{P}^g$, the laws of the bridge process $X^\star$ and the guided process X° are absolutely continuous with respect to each other on the path space $C([0, t]; H)$ for any $t < T$.
The main result of chapter 3, stated in Theorem 3.11, provides conditions on the SPDE parameters A, F and Q under which this absolute continuity extends onto the complete path space $C([0, T]; H)$. Moreover, it provides a Radon-Nikodym derivative $\Phi(X)$, tractable up to a proportionality constant, between the laws of $X^\star$ and X°. This enables sampling of $X^\star$ based on proposals drawn from X°, for example by means of the Metropolis-Hastings (MH) sampler given in Algorithm 1. A critical step in proving Theorem 3.11 is to show that the guided process X° converges to the conditioning point y as $t \to T$ and that it does so at the appropriate rate. This is shown in Theorem 3.9.
The findings of Chapter 3 led to the following article;

> Pieper-Sethmacher, T., F. van der Meulen and A. van der Vaart (2025). Simulation of infinite-dimensional diffusion bridges. *arXiv preprint arXiv:2503.13177.* (To appear in *Annals of Applied Probability.*)

Ultimately, Chapter 4 extends the methodology developed in Chapter 3 to address the three previously introduced statistical problems: filtering, smoothing and parameter estimation. Multiple observations $Y_1, \dots, Y_n$, as defined in (1.3) for non-degenerate Markov kernels with densities $k(X_{t_i}, y)\,\mathrm{d}y$, are assumed. In a similar setup, the filtering problem has been studied for the two-dimensional stochastic Navier-Stokes equation in Llopis et al. [78]. On the other hand, smoothing and Bayesian parameter inference in this context have not been considered before in the literature.
The approach to the filtering problem presented here closely follows the work in Llopis et al. [78], where a particle filter scheme based on likelihood-informed proposals was developed. Additionally, to overcome the high dimensionality of the problem, tempering and resample-move steps were incorporated.
The key contribution to the filtering problem made in this thesis is to show how the infinite-dimensional guided process can be incorporated as a proposal distribution into the particle filter. This generalises the proposals used in Llopis et al. [78]. The steps of the resulting guided particle filter are given in Algorithms 3 and 4. In experiments on the stochastic Amari equation, an equation from mathematical neuroscience that models the dynamics of a stochastic neural field, it is shown that the quality of the proposals based on the guiding distribution leads to improved performance compared to established methods.
The smoothing problem is addressed by extending the change of measure approach of Chapter 2 to include multiple observations. Specifically, denoting by μ the Markov transition kernel of X and defining $h(t,x)$ on $(t_{i-1}, t_i] \times H$ by

$$h(t,x) = \int k(x_{t_i}, y_i) \left(\prod_{j=i}^{n-1} k(x_{t_{j+1}}, y_{j+1})\, \mu_{t_j, t_{j+1}}(x_{t_j}, \mathrm{d}x_{t_{j+1}}) \right) \mu_{t,t_i}(x, \mathrm{d}x_{t_i})$$

leads to a change of measure $\mathbb{P}^h$ that conditions X on the entire vector of observations $Y = (Y_1, \dots, Y_n)$. This defines the distribution (1.7) of the smoothed process $X^\star = X \mid Y$. The function $h(t,x)$ may at first seem abstract but has a natural interpretation as the likelihood of $X_t = x$ given all non-past observations at time t under the dynamics governed by Equation (1.1).
A tractable approximation g of h is obtained by substituting the transition kernel μ with the kernel ν of an SPDE with linearised dynamics. This defines a change of measure $\mathbb{P}^g$ and the guided process X°, steered into areas of high posterior probability given the observations Y. Crucially, as shown in Theorem 4.9, the function g and R.N. derivative Φ of $\mathbb{P}^h$ with respect to $\mathbb{P}^g$ can be derived by solving an infinite-dimensional backwards Ricatti equation. Hence, weighted samples of $X^\star$ can be drawn by sampling the guided process X°. Algorithm 5 presents an MH sampler targeting the smoothing distribution of X based on this proposal scheme.
If the initial state x_0 and model parameters θ are unknown, Algorithm 5 can be extended to sample from the joint posterior of $(\theta, x_0, X) \mid Y$. However, due to the singularity of the path measures $\mathcal{L}_\theta(X^\star)$ for different parameters θ, this extension needs to be carried out with care. A naive implementation of iterating between the updates $X \mid \theta, Y$ and $\theta \mid X, Y$ will lead to a degenerate Markov chain.
This issue has first been considered in Roberts and Stramer [97] within the context of finite-dimensional diffusion bridges and the singularity of their path measures for different volatility parameters. In the SPDE setting, this issue worsens as it affects both the drift operator A_θ and the volatility operator Q_θ. The solution, as suggested in Roberts and Stramer [97], is to define a reparametrisation of the conditioning that decouples the close dependency between $\mathcal{L}_\theta(X^\star)$ and the parameter θ.
In Section 4.4, a measurable mapping

$$\Gamma_\theta(x_0, \cdot) : C([0,T]; H') \to C([0,T]; H)$$

and a measure $\mathbb{V}_\theta^{x_0}$ on $C([0,T];H')$, absolutely continuous with respect to a Gaussian measure $\mathbb{W}$ on $C([0,T];H')$, are derived that jointly accomplish this decoupling. Here, H' denotes an enlarged Hilbert space such that $H \hookrightarrow H'$ is embedded in a Hilbert-Schmidt way. If V is then a process with $V \sim \mathbb{V}_\theta^{x_0}$, the law of $\Gamma_\theta(x_0, V)$ equals $\mathcal{L}_\theta(X^\star)$. A Gibbs sampler based on this reparametrisation that targets the joint posterior of $(\theta, x_0, V) \mid Y$ is presented in Algorithm 7. Performance of the algorithm is tested in a case study for the Amari equation, where it successfully estimates model parameters that were previously assumed to be known in the literature due to challenges in their estimation.
The findings of Chapter 4 resulted in the following article;

> Pieper-Sethmacher, T., D. Avitabile, F. van der Meulen (2025). Guided filtering and smoothing for infinite-dimensional diffusions. *arXiv preprint arXiv:2507.06786.* (Under Review).

Chapter 2

On a class of exponential changes of measure for stochastic PDEs

In the following chapter, we consider a class of exponential changes of measure for a mild solution X to a stochastic partial differential equation. The changed measures $\mathbb{P}^h$ depend on the choice of a function h in the domain of the infinitesimal generator L of X, defined in the topology of bounded pointwise convergence on a suitable function space. In our main result, we derive conditions on h for which the change of measure is of Girsanov-type. The process X under $\mathbb{P}^h$ is then shown to be a mild solution to another SPDE with an additional drift-term. We illustrate how different choices of h impact the law of X under $\mathbb{P}^h$ in selected applications. These include the derivation of an infinite-dimensional diffusion bridge as well as the introduction of guided processes for SPDEs, generalising results known for finite-dimensional diffusion processes to the infinite-dimensional case.

2.1 Introduction

Consider a semilinear *stochastic partial differential equation (SPDE)* of the form

$$\begin{cases} \mathrm{d}X(t) &= [AX(t) + F(t, X(t))]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W(t), \quad t \geq s, \\ X(s) &= x. \end{cases} \tag{2.1}$$

The operator A denotes the generator of a strongly continuous semigroup $(S_t)_{t\geq 0}$ on a Hilbert space H, whereas F denotes a nonlinear operator and Q is a symmetric, positive operator on H. The process W is a cylindrical Wiener process on H, defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. We assume that the operators A, F and Q satisfy suitable conditions such that Equation (2.1) admits a unique mild solution $X = (X(t,s,x))_{t\geq s}$ for any $s \geq 0$ and $x \in H$. Throughout the chapter we fix some arbitrary $x_0 \in H$ and simply write $X(t)$ if the SPDE in (2.1) is assumed to be initialised at $X(0) = x_0$.
For any $m \in \mathbb{N}$, let $C_m(\mathbb{R}_+ \times H)$ be the Banach space of continuous functions $\varphi : \mathbb{R}_+ \times H \to \mathbb{R}$ such that $\|\varphi\|_m = \sup_{t,x}(1 + \|x\|^m)^{-1}|\varphi(t,x)| < \infty$. The process X is Markovian and defines a transition semigroup

$$(T_t\varphi)(s,x) = \mathbb{E}[\varphi(s+t, X(t+s,s,x))], \quad s,t \geq 0,\ x \in H,$$

on $C_m(\mathbb{R}_+ \times H)$. It is well-known that the semigroup $(T_t)_{t\geq 0}$ is not strongly continuous with respect to the norm topology on $C_m(\mathbb{R}_+ \times H)$, see e.g. Cerrai [20] and Da Prato [39]. However, it does possess the properties of a strongly continuous semigroup in

several weaker 'modes of convergence'. This has been studied in the framework of $\mathcal{K}$-*convergence* in Cerrai [20], Cerrai [21] and Cerrai and Gozzi [22], the mixed topology in Goldys and Kocan [58] and of *bp- (bounded pointwise) or π- convergence* in Priola [96]. See also Fabbri et al. [52], Appendix B for a recent survey. In the respective convergence of choice, one can then define an infinitesimal generator $(L, \mathrm{dom}_m(L))$ of the semigroup $(T_t)_{t\geq 0}$ in the usual way. In this chapter, we will work within the framework of π-convergence as introduced in Priola [96].
Crucially, the operator $(L, \mathrm{dom}_m(L))$ exhibits the common properties that are characteristic for infinitesimal generators of strongly continuous semigroups. Of particular importance for us is the fact that *Dynkin's formula* holds, i.e. for any $h \in (L, \mathrm{dom}_m(L))$ the process

$$D^h(t) = h(t, X(t)) - \int_0^t Lh(s, X(s))\,\mathrm{d}s$$

is a $\mathbb{P}$-martingale. In other words, X solves the *martingale problem* of $(L, \mathrm{dom}_m(L))$ as introduced in Stroock and Varadhan [109]. Furthermore, one can show that for any positive $h \in \mathrm{dom}_m(L)$, the process

$$E^h(t) = \frac{h(t, X(t))}{h(0, x_0)} \exp\left(-\int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s\right), \quad t \geq 0,$$

whenever existent, is a positive, continuous local $\mathbb{P}$-martingale with $\mathbb{E}[E^h(0)] = 1$. If E^h is a true $\mathbb{P}$-martingale, it defines an *exponential change of measure* $\mathbb{P}^h$ on $\mathcal{F}$ such that for any $t \geq 0$

$$\mathrm{d}\mathbb{P}^h_{|\mathcal{F}_t} = E^h(t)\,\mathrm{d}\mathbb{P}_{|\mathcal{F}_t}. \tag{2.2}$$

The change of measure $\mathbb{P}^h$ is well-known in the literature for Markov processes, see Palmowski and Rolski [89] and references within. If the function h is harmonic, i.e. $Lh = 0$, it is known as *Doob's h-transform*, following its introduction in Doob [46]. In Palmowski and Rolski [89] it was shown that X remains Markovian under $\mathbb{P}^h$ and solves the martingale problem corresponding to a perturbation of L.
In this chapter we aim to establish conditions on the h-function under which X is not only Markovian under the changed measure, but again the mild solution of another SPDE, differing from Equation (2.1) by an additional drift-term dependent on h. This can be viewed as a special case in which $\mathbb{P}^h$ is a *Girsanov*-type change of measure. In this spirit, we show the following as the main result of this chapter.

Theorem 2.1 (Informal). *Under suitable assumptions on $h \in \mathrm{dom}_m(L)$, there exists a unique measure $\mathbb{P}^h$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ that satisfies (2.2). Furthermore, the process*

$$W^h(t) = W(t) - \int_0^t Q^{\frac{1}{2}}\,\mathrm{D}_x \log h(s, X(s))\,\mathrm{d}s, \quad t \in [0, T],$$

is a cylindrical Wiener process with respect to $\mathbb{P}^h$. In particular, X under $\mathbb{P}^h$ solves the SPDE

$$\mathrm{d}X(t) = [AX(t) + F(t, X(t)) + Q\,\mathrm{D}_x \log h(t, X(t))]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W^h(t), \quad t \in [0, T].$$

2.1.1 Approach and challenges

Let us demonstrate our approach on how to derive Theorem 2.1 in the case that $H = \mathbb{R}^d$. Equation (2.1) then describes a *stochastic ordinary differential equation (SODE)*

$$\mathrm{d}x(t) = b(t, x(t))\,\mathrm{d}t + \sqrt{q}\,\mathrm{d}w(t), \tag{2.3}$$

where b is some Lipschitz continuous function, q is a symmetric positive definite matrix and w is an $\mathbb{R}^d$-valued Wiener process. Let $h \in C_b^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$ be differentiable with bounded derivatives. An application of *Itô's formula* then gives that $h(t, x(t))$ is the semimartingale given by

$$h(t, x(t)) = h(0, x_0) + \int_0^t L_0 h(s, x(s))\, \mathrm{d}s + \int_0^t \langle \sqrt{q}\, \mathrm{D}_x h(s, x(s)), \mathrm{d}w(s) \rangle,$$

where L_0 is the *Kolmogorov operator* associated with equation (2.3), defined via

$$(L_0 h)(t, x) = \partial_t h(t, x) + \langle b(t, x), \mathrm{D}_x h(t, x) \rangle + \frac{1}{2} \operatorname{tr}[q\, \mathrm{D}_x^2 h(t, x)].$$

From this, one can conclude that $h \in \mathrm{dom}_m(L)$ with $Lh = L_0 h$ and Dynkin martingale D^h given by

$$D^h(t) = h(0, x_0) + \int_0^t \langle \sqrt{q}\, \mathrm{D}_x h(s, x(s)), \mathrm{d}w(s) \rangle.$$

An application of the integration by parts formula for semimartingales then shows that the local martingale E^h equals the *stochastic exponential*

$$E^h(t) = \exp\left(M^h(t) - \frac{1}{2} \left[M^h \right]_t \right)$$

of the Itô process $M^h(t) = \int_0^t \langle \sqrt{q}\, \mathrm{D}_x \log h(s, x(s)), \mathrm{d}w(s) \rangle$. Therefore, if E^h is a true martingale, the *Girsanov theorem* implies that the process

$$w^h(t) = w(t) - \int_0^t \sqrt{q}\, \mathrm{D}_x \log h(s, x(s))\, \mathrm{d}s$$

is a Wiener process under $\mathbb{P}^h$ and that x solves the SODE

$$\mathrm{d}x(t) = b(t, x(t))\, \mathrm{d}t + q\, \mathrm{D}_x \log h(t, x(t))\, \mathrm{d}t + \sqrt{q}\, \mathrm{d}w^h(t).$$

We face two key challenges when generalising this approach to the setting of an infinite-dimensional Hilbert space H. Firstly, since A is an unbounded operator on H, we generally cannot expect the SPDE (2.1) to admit a strong solution. This renders any direct application of Itô's formula infeasible and even for smooth functions h, the process $h(t, X(t))$ is in general not a semimartingale.
We circumvent this issue by approximating X by a sequence of strong solutions X_n that satisfy Equation (2.1) when substituting A by its *Yosida approximations* $(A_n)_n$. Under suitable assumptions on h, one can then approximate the process $h(t, X(t))$ by the sequence of semimartingales $h(t, X_n(t))$ for which Itô's formula is applicable.
Secondly, consider the Kolmogorov operator

$$(L_0 \varphi)(t, x) = \partial_t \varphi(t, x) + \langle Ax + F(t, x), \mathrm{D}_x \varphi(t, x) \rangle + \frac{1}{2} \operatorname{tr}[Q\, \mathrm{D}_x^2 \varphi(t, x)]$$

associated with the SPDE (2.1). In order for $L_0 \varphi$ to be a well-defined and continuous function on $\mathbb{R}_+ \times H$, besides the usual smoothness properties of φ, one requires that there exists a continuous extension of the mapping

$$\mathbb{R}_+ \times \mathrm{dom}(A) \to \mathbb{R},\ (t, x) \mapsto \langle Ax, \mathrm{D}_x \varphi(t, x) \rangle.$$

This severely limits the class of functions for which $L_0 \varphi$ is a well-defined differential operator and substantial work has been done to construct suitable spaces of test functions for Kolmogorov operators in infinite dimensions, see Da Prato [39] and references within.
In particular, for most of our applications, the h-functions of interest cannot be expected to be in the domain of L_0. However, in Manca [81] and Manca [82] it is shown that the space of *exponential test functions* $\mathcal{E}_A(\mathbb{R}_+ \times H)$, defined as the real and imaginary parts of the functions

$$\mathbb{R}_+ \times H \to \mathbb{R},\ (s, x) \mapsto \exp(i(\langle x, a \rangle + cs)), \quad a \in \mathrm{dom}(A^*), c \in \mathbb{R},$$

acts as a core for the infinitesimal generator $(L, \mathrm{dom}_m(L))$ with respect to π-convergence. Therefore, under the weaker assumption that $h \in \mathrm{dom}_m(L)$, we may approximate h with a sequence of suitable test functions $(h_n)_n \subset \mathcal{E}_A(\mathbb{R}_+ \times H)$ for which $L_0 h_n$ remains well-defined.

2.1.2 Related work and applications

In applications of the exponential change of measure E^h, one chooses a suitable h-function such that X under $\mathbb{P}^h$ exhibits certain desired properties. A well-known application from the finite-dimensional setting is the derivation of *diffusion bridges* that describe the process x conditioned on hitting an endpoint $x(T) = y \in \mathbb{R}^d$. An application of our results lifts this to the infinite-dimensional case, thereby allowing us to derive an equation for the *infinite-dimensional diffusion bridge* (or *SPDE bridge*).
To the best of our knowledge, the existing literature on infinite-dimensional bridges is limited to the case where $F \equiv 0$. This ensures mild solutions are Gaussian processes, which leads to an explicit expression for the SPDE bridge, also called an *Ornstein-Uhlenbeck (OU)* bridge in this case. Simão [104] shows that an infinite system of one-dimensional OU bridges defines an OU bridge on a Hilbert space. In a more general, non-diagonal setting, Goldys and Maslowski [59] derive an equation for the OU bridge and apply it to study basic properties of transition semigroups for semilinear SPDEs. More recently, Di Nunno et al. [45] consider a linear stochastic reaction–diffusion equation on a bounded domain where the process is conditioned on a noisy observation at time T. A general framework for the spatial discretization of these bridge processes is developed.
Our approach via the change of measure E^h is more general. In the specific case that $h(t, x) = p_X(t, x; T, y)$, with p_X denoting the transition density of the process X with respect to an appropriately chosen reference measure, it gives rise to the infinite-dimensional diffusion bridge that conditions the process to hit y at time T. Our results allow for other choices of h, for example $h(t, x) = p_Z(t, x; T, y)$, where p_Z is the transition density of the Ornstein-Uhlenbeck process Z. The resulting process is called a *guided process*, analogous to the finite-dimensional setting introduced in Schauer et al. [101]. The guided process is different from the conditioned process, but it mimics properties of that process, though the extra term in the drift ignores the nonlinearity. Contrary to p_X, the transition density p_Z is explicitly known. Therefore, the SPDE for the guided process can be numerically approximated. Taking into account the likelihood ratio of the law of the true conditioned process with respect to the guided process, *weighted samples* of the conditioned process are obtained. It is exactly this approach which has been proposed in earlier works in the simpler setting of stochastic ordinary differential equations (see, for instance, Schauer et al. [101], Delyon and Hu [44], Bierkens et al. [16] and applications in Mider et al. [85]). The results in this chapter prove the existence of the guided process as the mild solution to a particular SPDE.
Another application that we consider is that of *forcing* the process so that its marginal distribution at time T is fixed to a specified distribution. This extends the results for the finite-dimensional case considered in Baudoxin [9].

2.1.3 Outline

We provide an overview of the needed preliminaries in Section 2.2. Particular attention will be given to semigroups that are strongly continuous with respect to π-convergence as well as their infinitesimal generators. In Section 2.3, we present, in a detailed manner, the main result of this chapter and its proof. Additionally, a modified version is given for the special case that the change of measure is limited to a finite time horizon. We showcase some applications of the main result in Section 2.4. These include the derivations of the infinite-dimensional diffusion bridge and the forced process as well as the guided process.

2.2 Preliminaries

2.2.1 On the stochastic basis

We assume to be working on the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ defined as follows. Let H be some Hilbert space, embedded in another Hilbert space H' such that the embedding $J : H \hookrightarrow H'$ is a Hilbert-Schmidt operator. Note that in particular, JJ^* is a positive definite trace-class operator on H'. Define $\Omega = C(\mathbb{R}_+; H')$ as the space of all continuous functions from $\mathbb{R}_+$ to H' equipped with the metric

$$d(\omega, \tilde{\omega}) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\|\omega - \tilde{\omega}\|_n}{1 + \|\omega - \tilde{\omega}\|_n},$$

where $\|\omega - \tilde{\omega}\|_n = \sup_{t\in[0,n]} |\omega(t) - \tilde{\omega}(t)|$. Then (Ω, d) is a Polish space and we denote by $\mathcal{F}$ the Borel σ-algebra of (Ω, d).
On $(\Omega, \mathcal{F})$ define the canonical process $\eta : \mathbb{R}_+ \times \Omega \to H'$, $\eta_t(\omega) = \omega(t)$ and let $(\mathcal{F}_t)_{t\geq 0}$ be the right-continuous extension of the natural filtration of η_t, i.e. $\mathcal{F}_t = \bigcap_{\varepsilon>0} \sigma(\eta_s : s \leq t + \varepsilon)$ for any $t \geq 0$. By the Kolmogorov extension theorem, there exists a Gaussian measure $\mathbb{P}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ such that, under $\mathbb{P}$, the canonical process η is a Wiener process on H' with covariance operator JJ^*. In particular, η is a cylindrical Wiener process on H.

Denote by $\mathbb{P}_t$ the restriction of $\mathbb{P}$ onto $\mathcal{F}_t$ for any $t \geq 0$. We will need the following result (see Stroock [108], Lemma 4.2.).

Lemma 2.2. *Let $(E(t))_{t\geq 0}$ be a non-negative martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with $\mathbb{E}[E(0)] = 1$. Then there exists a unique measure $\mathbb{Q}$ on $\mathcal{F}$ such that $\mathrm{d}\mathbb{Q}_t = E(t)\,\mathrm{d}\mathbb{P}_t$ for all $t \geq 0$.*

Remark 2.3. Note that Lemma 2.2 does not give absolute continuity of $\mathbb{Q}$ with respect to $\mathbb{P}$. However, under the stronger assumption that E is a uniformly integrable martingale, it can be shown that $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$ on $\mathcal{F}$ with $\mathrm{d}\mathbb{Q} = E_\infty\,\mathrm{d}\mathbb{P}$, where E_∞ is the unique random variable such that $E(t) = \mathbb{E}[E_\infty \mid \mathcal{F}_t]$ for all $t \geq 0$.

In many applications the martingale E is only defined on some half-open interval $t \in [0, T)$. In that circumstance, the following variation of Lemma 2.2 will be useful. See Appendix 2.A for the proof.

Lemma 2.4. *Let $(E(t))_{t\in[0,T)}$ be a non-negative martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T)}, \mathbb{P})$ with $\mathbb{E}[E(0)] = 1$. Then there exists a unique measure $\mathbb{Q}$ on $\mathcal{F}_T$ such that $\mathrm{d}\mathbb{Q}_t = E(t)\,\mathrm{d}\mathbb{P}_t$ for all $0 \leq t < T$.*

2.2.2 On stochastic evolution equations

The following is a standing assumption on the components involved in Equation (2.1).
Assumption 2.5.

(i) *A is the generator of a strongly continuous semigroup $(S_t)_{t\geq 0}$ on H. In particular there exists a $C_S > 0, \omega_S \in \mathbb{R}$ such that $\|S_t\|_{L(H)} \leq C_S \exp(\omega_S t)$ for all $t \geq 0$.*

(ii) *W is a cylindrical Wiener process on H.*

(iii) *Q is a symmetric, positive and bounded operator on H. Furthermore, the family of operators $(Q_t)_{t\geq 0}$ defined by*

$$Q_t = \int_0^t S_s Q S_s^* \, \mathrm{d}s \tag{2.4}$$

is such that $\sup_t \operatorname{tr}(Q_t) < \infty$.

(iv) *F is such that there exists a constant $C_F > 0$ with*

$$\|F(t,x) - F(t,y)\| \leq C_F \|x - y\| \text{ and } \|F(t,x)\| \leq C_F(1 + \|x\|)$$

for all $t \geq 0$ and $x, y \in H$.

Under Assumption 2.5, Equation (2.1) admits a unique mild solution $X = (X(t,s,x))_{t\geq s}$ for any initial value $x \in H$, i.e. X is an H-valued, $\mathcal{F}_t$-adapted process that satisfies

$$X(t,s,x) = S_{t-s}x + \int_s^t S_{t-u}F(u, X(u,s,x)) \, \mathrm{d}u + \int_s^t S_{t-u}Q^{\frac{1}{2}} \, \mathrm{d}W(u), \quad t \geq s. \tag{2.5}$$

The process X is Markovian and has a $\mathbb{P}$-almost surely continuous modification. Furthermore, for any $m \geq 1$, there exist some constants $C_m > 0, \gamma_m \geq 0$, also depending on ω_S, C_S, C_F and Q, such that

$$\mathbb{E}[\|X(t,s,x)\|^m] \leq C_m \exp(\gamma_m(t-s))(1 + \|x\|^m). \tag{2.6}$$

Though bounds similar to (2.6) are well known results in the literature, this particular bound follows from Goldys and Kocan [58], Proposition 3.1. To abbreviate notation, we fix some arbitrary $x_0 \in H$ and simply write $X(t)$ whenever (2.1) is assumed to be initialised at $X(0) = x_0$.

Due to the unbounded nature of A, one generally cannot assume Equation (2.1) to admit a strong solution. However, one can approximate its mild solution with strong solutions to a sequence of substitute equations in the following way. Let $(A_n)_{n>\omega_S}$ be the Yosida approximation of A, i.e. $(A_n)_{n>\omega_S}$ is the sequence of bounded, linear operators on H defined via

$$A_n = nAR(n, A),$$

where $R(n,A) = (n - A)^{-1}$ is the resolvent of A. It then holds that $\lim_n A_n x = Ax$ for any $x \in \operatorname{dom}(A)$ and that A_n defines a semigroup $S_t^{(n)}$ such that $\lim_n S_t^{(n)}x = S_t x$ for all $x \in H$ and

$$\|S_t^{(n)}\|_{L(H)} \leq C_S \exp(\omega_n t) \tag{2.7}$$

with $\omega_n = \dfrac{\omega_S n}{n - \omega_S}$. Now let X_n be the strong solution to the equation

$$\mathrm{d}X_n(t) = [A_n X_n(t) + F(t, X_n(t))] \, \mathrm{d}t + Q^{\frac{1}{2}} \, \mathrm{d}W(t), \quad X_n(s) = x. \tag{2.8}$$

It is well-known (see e.g. Da Prato and Zabczyk [41], Proposition 7.4) that

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\in[s,T]} |X_n(t) - X(t)|^p\right] = 0 \tag{2.9}$$

for any $T > s, p > 1$ and in particular $X_n \to X$ in probability as $C([s,T];H)$-valued random variables.

In the special case that $F = 0$, we denote the time homogeneous mild solution to Equation (2.1) by Z and refer to it as the *Ornstein-Uhlenbeck (OU) process*. The random variables $Z(t,s,x)$ are Gaussian with mean $S_{t-s}x$ and covariance operator Q_{t-s}. Under Assumption 2.5 (iii), $Z(t,s,x)$ converges in distribution to its invariant distribution $\nu \sim \mathcal{N}(0, Q_\infty)$ with

$$Q_\infty = \int_0^\infty S_u Q S_u^* \, \mathrm{d}u. \tag{2.10}$$

2.2.3 On transition semigroups and their generators on spaces of polynomial growth

Let $(E, |\cdot|_E)$ be a Banach space. For any $m \geq 0$, we let $C_m(\mathbb{R}_+ \times H; E)$ be the space of all continuous functions $\varphi : \mathbb{R}_+ \times H \to E$ that are bounded in t and of at most polynomial growth of order m in x. The space $C_m(\mathbb{R}_+ \times H; E)$ is a Banach space with norm

$$\|\varphi\|_m = \sup_{(t,x)\in\mathbb{R}_+\times H} \frac{|\varphi(t,x)|_E}{1 + \|x\|^m}.$$

If $E = \mathbb{R}^d$, equipped with the Euclidean norm, we simply write $C_m(\mathbb{R}_+ \times H)$.
It follows from the inequality (2.6) that the time-homogeneous space-time process of X, given by $Y(t,(s,x)) = (t+s, X(t+s,s,x))$, defines a transition semigroup $(T_t)_{t\geq 0}$ on $C_m(\mathbb{R}_+ \times H)$ via

$$(T_t\varphi)(s,x) = \mathbb{E}[\varphi(s+t, X(t+s,s,x))], \quad \varphi \in C_m(\mathbb{R}_+ \times H), t \geq 0, x \in H. \tag{2.11}$$

As noted in the introduction, $(T_t)_{t\geq 0}$ is not strongly continuous with respect to the norm topology on $C_m(\mathbb{R}_+ \times H)$. Instead, one has to turn to the weaker mode of *π-convergence*. We say a sequence $(\varphi_n)_n \subset C_m(\mathbb{R}_+ \times H)$ is *π-convergent* to $\varphi \in C_m(\mathbb{R}_+ \times H)$ and write $\pi\text{-}\lim_n \varphi_n = \varphi$ if

$$\sup_n \|\varphi_n\|_m < \infty \text{ and } \lim_n \varphi_n(t,x) = \varphi(t,x) \text{ for all } (t,x) \in \mathbb{R}_+ \times H.$$

The *π-closure* of a subset B in $C_m(\mathbb{R}_+ \times H)$ is defined as

$$\overline{B}^\pi = \{\varphi \in C_m(\mathbb{R}_+ \times H) : \exists (\varphi_n)_n \subset B \text{ s.t. } \pi\text{-}\lim_n \varphi_n = \varphi\}.$$

The set B is said to be *π-closed* if $\overline{B}^\pi = B$ and *π-dense* if $\overline{B}^\pi = C_m(\mathbb{R}_+ \times H)$. A linear operator $L : \mathrm{dom}(L) \subset C_m(\mathbb{R}_+ \times H) \to C_m(\mathbb{R}_+ \times H)$ is a *π-closed operator* if the graph $\{(\varphi, L\varphi) : \varphi \in \mathrm{dom}(L)\}$ is π-closed in $C_m(\mathbb{R}_+ \times H) \times C_m(\mathbb{R}_+ \times H)$. If a subdomain $D \subset \mathrm{dom}(L)$ is such that for any $\varphi \in \mathrm{dom}(L)$ there exists a sequence $(\varphi_n)_n$ in D with

$$\pi\text{-}\lim_n \varphi_n = \varphi \text{ and } \pi\text{-}\lim_n L\varphi_n = L\varphi \tag{2.12}$$

we call D a π-core for $(L, \mathrm{dom}(L))$.
The following definition of π-semigroups is based on Priola [96]. They are exactly those semigroups that are 'strongly continuous' with respect to π-convergence.

Definition 2.6 (π-semigroup). Let $(T_t)_{t\geq 0}$ be a semigroup of bounded, linear operators on $C_m(\mathbb{R}_+ \times H), m \geq 0$. We say $(T_t)_{t\geq 0}$ is a *π-semigroup* if the following conditions hold:

(i) There exist some $M \geq 1$ and $\omega \in \mathbb{R}$ such that for all $t \geq 0$
$$\|T_t\| \leq M \exp(\omega t). \tag{2.13}$$

(ii) For any $t \geq 0$ and any $(\varphi_n)_n \subset C_m(\mathbb{R}_+ \times H)$ such that $\pi\text{-}\lim_n \varphi_n = \varphi \in C_m(\mathbb{R}_+ \times H)$ we have
$$\pi\text{-}\lim_n T_t\varphi_n = T_t\varphi \in C_m(\mathbb{R}_+ \times H). \tag{2.14}$$

(iii) For any $\varphi \in C_m(\mathbb{R}_+ \times H)$ and $(s,x) \in \mathbb{R}_+ \times H$ fixed, the mapping
$$[0,\infty) \longrightarrow \mathbb{R}, \ t \mapsto T_t\varphi(s,x) \tag{2.15}$$
is continuous.

If condition (i) holds with $M = 1$ and $\omega = 0$ we call $(T_t)_{t\geq 0}$ a *contraction π-semigroup.*

Remark 2.7. Note that contrary to the case of semigroups that are strongly continuous with respect to the norm topology on $C_m(\mathbb{R}_+ \times H)$, condition (i) needs to be assumed as it does not follow from the other conditions.

Remark 2.8. Conditions (i) and (iii) imply that any π-semigroup $(T_t)_{t\geq 0}$ is 'strongly continuous' with respect to π-convergence, i.e. if $(t_n)_n$ is a sequence such that $t_n \downarrow 0$ then

$$\pi\text{-}\lim_n T_{t_n}\varphi = \varphi, \quad \varphi \in C_m(\mathbb{R}_+ \times H).$$

We then write $\pi\text{-}\lim_{t\downarrow 0} T_t\varphi = \varphi$.

Lemma 2.9. *For any $m \geq 0$, the semigroup $(T_t)_{t\geq 0}$ defined in (2.11) is a π-semigroup on $C_m(\mathbb{R}_+ \times H)$.*

Proof. We show that $(T_t)_{t\geq 0}$ satisfies properties (i) to (iii) in Definition 2.6. Let $\varphi \in C_m(\mathbb{R}_+ \times H)$. Then it holds for any (s,x) that

$$\begin{aligned} |(T_t\varphi)(s,x)| &\leq \mathbb{E}[|\varphi(s+t, X(s+t,s,x))|] \\ &\leq \|\varphi\|_m \mathbb{E}[1 + \|X(s+t,s,x)\|^m] \\ &\leq \|\varphi\|_m (1 + C_m \exp(\gamma_m t)(1 + \|x\|^m)), \end{aligned} \tag{2.16}$$

where we used the definition of $\|\varphi\|_m$ in the second and the inequality (2.6) in the third step. It follows that

$$\begin{aligned} \|T_t\varphi\|_m &= \sup_{(s,x)\in\mathbb{R}_+\times H} \frac{|T_t\varphi(s,x)|}{1+\|x\|^m} \\ &\leq \|\varphi\|_m \sup_{(s,x)\in\mathbb{R}_+\times H} \frac{(1 + C_m \exp(\gamma_m t)(1+\|x\|^m))}{1+\|x\|^m} \\ &\leq \|\varphi\|_m (1 + C_m) \exp(\gamma_m t) \end{aligned}$$

and thus (2.13) holds with $M = (1 + C_m)$ and $\omega = \gamma_m$.
To show (ii), let $(\varphi_n)_n$ be a sequence in $C_m(\mathbb{R}_+ \times H)$ such that $\pi\text{-}\lim_n \varphi_n = \varphi$. It then follows by the dominated convergence theorem that

$$\begin{aligned} \lim_n (T_t\varphi_n)(s,x) &= \lim_n \mathbb{E}[\varphi_n(t+s, X(t+s,s,x))] \\ &= \mathbb{E}[\varphi(t+s, X(t+s,s,x))] = (T_t\varphi)(s,x). \end{aligned}$$

Furthermore, from (2.16) and $\sup_n \|\varphi\|_m < \infty$ it follows that $\sup_n \|T_t\varphi_n\|_m < \infty$ and therefore that $\pi\text{-}\lim_n T_t\varphi_n = T_t\varphi$. To show (iii), it suffices to note that continuity of $t \mapsto T_t\varphi(s,x)$ follows from the almost sure continuity of $t \mapsto X(\cdot,s,x)$ and dominated convergence. □

The *infinitesimal generator* of a π-semigroup is defined in a similar manner as for a strongly continuous semigroup.

Definition 2.10. Let $(T_t)_{t\geq 0}$ be a π-semigroup on $C_m(\mathbb{R}_+ \times H)$. The *infinitesimal generator* L of $(T_t)_{t\geq 0}$ is the operator defined via

$$\begin{cases} \mathrm{dom}_m(L) &= \left\{\varphi \in C_m(\mathbb{R}_+ \times H) : \exists \psi \in C_m(\mathbb{R}_+ \times H) \text{ s.t. } \pi\text{-}\lim_{t\downarrow 0} \dfrac{T_t\varphi - \varphi}{t} = \psi\right\} \\ (L\varphi)(s,x) &= \lim_{t\downarrow 0} \dfrac{(T_t\varphi)(s,x) - \varphi(s,x)}{t}, \quad \varphi \in \mathrm{dom}_m(L), (s,x) \in \mathbb{R}_+ \times H. \end{cases} \tag{2.17}$$

As the upcoming result shows, the infinitesimal generator of a π-semigroup satisfies the common properties characteristic for generators of strongly continuous semigroups.

Theorem 2.11. *Let L be the infinitesimal generator of a π-semigroup $(T_t)_{t\geq 0}$ on $C_m(\mathbb{R}_+ \times H)$. Then:*

(i) *The domain $\mathrm{dom}_m(L)$ is π-dense in $C_m(\mathbb{R}_+ \times H)$.*

(ii) *The operator L is the unique π-closed operator such that for all $\lambda > \omega$ the resolvent $R(\lambda, L) = (\lambda - L)^{-1}$ is a bounded, linear operator on $C_m(\mathbb{R}_+ \times H)$ with*

$$R(\lambda, L)\varphi(s,x) = \int_0^\infty \exp(-\lambda t) T_t \varphi(s,x) \, \mathrm{d}t. \tag{2.18}$$

(iii) *It holds that $T_t(\mathrm{dom}_m(L)) \subset \mathrm{dom}_m(L)$ and for all $\varphi \in \mathrm{dom}_m(L)$ and fixed $(s,x) \in \mathbb{R}_+ \times H$ the function $t \mapsto T_t\varphi(s,x)$ is differentiable with*

$$\frac{\mathrm{d}}{\mathrm{d}t} T_t \varphi(s,x) = L T_t \varphi(s,x) = T_t L \varphi(s,x). \tag{2.19}$$

Proof. See Priola [96], Propositions 3.2 to 3.6., where these properties have been shown in the context of π-semigroups on $C_b(H)$. □

The following is essentially a version of *Dynkin's formula* in the context of π-semigroups.

Lemma 2.12. *Let $Y(t) = (t, X(t))$ be the space-time process of $X(t)$. Then for any $\varphi \in \mathrm{dom}(L)$ the process*

$$D^\varphi(t) = \varphi(Y(t)) - \int_0^t L\varphi(Y(s)) \, \mathrm{d}s \tag{2.20}$$

is an $\mathcal{F}_t$-adapted $\mathbb{P}$-martingale. We call $D^\varphi(t)$ the Dynkin martingale (of $(\varphi, L\varphi)$).

Proof. Clearly $Y(t) = (t, X(t,x))$ is $\mathcal{F}_t$-adapted with initial condition $Y(0) = (0, x_0)$. It holds for any $r \le t$ that

$$\begin{aligned} \mathbb{E}[D^\varphi(t) \mid \mathcal{F}_r] &= \mathbb{E}\left[\varphi(Y(t)) - \int_0^t L\varphi(Y(u)) \, \mathrm{d}u \mid \mathcal{F}_r\right] \\ &= \mathbb{E}[\varphi(Y(t)) \mid \mathcal{F}_r] - \mathbb{E}\left[\int_0^r L\varphi(Y(u)) \, \mathrm{d}u \mid \mathcal{F}_r\right] - \mathbb{E}\left[\int_r^t L\varphi(Y(u)) \, \mathrm{d}u \mid \mathcal{F}_r\right] \\ &= T_{t-r}\varphi(Y(r)) - \int_0^r L\varphi(Y(u)) \, \mathrm{d}u - \int_r^t T_{u-r} L\varphi(Y(r)) \, \mathrm{d}u, \end{aligned} \tag{2.21}$$

where we used the Markov property and $\mathcal{F}_t$-adaptedness of $Y(t)$ in the last step. By Theorem 2.11 (iii) and a substitution we have

$$\begin{aligned} \int_r^t T_{u-r} L\varphi(Y(r)) \, \mathrm{d}u &= \int_0^{t-r} T_u L\varphi(Y(r)) \, \mathrm{d}u \\ &= T_{t-r}\varphi(Y(r)) - \varphi(Y(r)) \quad \mathbb{P}\text{-a.s.} \end{aligned}$$

Plugging this into (2.21) gives $\mathbb{E}[D^\varphi(t) \mid \mathcal{F}_r] = \varphi(Y(r)) - \int_0^r L\varphi(Y(u)) \, \mathrm{d}u$ which gives the claim. □

In general, the abstract definition of the infinitesimal generator given in (2.17) does not lend itself easily to a closed form expression of $L\varphi$ for arbitrary $\varphi \in \mathrm{dom}_m(L)$. However, in many cases one can instead construct a suitable π-core for $(L, \mathrm{dom}_m(L))$ on which L acts in a more 'descriptive' manner. It turns out that in this case, L acts as the Kolmogorov operator associated with (2.1) on the space of *exponential test functions* $\mathcal{E}_A(\mathbb{R}_+ \times H)$, defined as the span of the real and imaginary parts of the functions

$$\mathbb{R}_+ \times H \to \mathbb{R}, \ (s,x) \mapsto \exp(i(\langle x, a\rangle + cs)), \quad a \in \mathrm{dom}(A^*), c \in \mathbb{R}, \tag{2.22}$$

and $\mathcal{E}_A(\mathbb{R}_+ \times H)$ defines a π-core of $(L, \mathrm{dom}_m(L))$ as the following lemma shows. The π-core property of $\mathcal{E}_A(\mathbb{R}_+ \times H)$ will play a crucial part in the proof our main result.

Lemma 2.13. *For any $m \geq 1$, the space $\mathcal{E}_A(\mathbb{R}_+ \times H)$ is a subset of* $\mathrm{dom}_m(L)$ *with* $L\varphi = L_0\varphi$, *where*

$$(L_0\varphi)(t,x) = \partial_t\varphi(t,x) + \langle x, A^* \,\mathrm{D}_x\varphi(t,x)\rangle + \langle F(t,x), \mathrm{D}_x\varphi(t,x)\rangle + \frac{1}{2}\,\mathrm{tr}\left(Q\,\mathrm{D}_x^2\varphi(t,x)\right) \tag{2.23}$$

for any $\varphi \in \mathcal{E}_A(\mathbb{R}_+ \times H)$. Moreover, $\mathcal{E}_A(\mathbb{R}_+ \times H)$ is a π-core for $(L, \mathrm{dom}_m(L))$, i.e. for any $\varphi \in \mathrm{dom}_m(L)$ there exists a sequence $(\varphi_n)_n \subset \mathcal{E}_A(\mathbb{R}_+ \times H)$ [1] *such that*

$$\pi\text{-}\lim_n \varphi_n = \varphi \text{ and } \pi\text{-}\lim_n L_0\varphi_n = L\varphi. \tag{2.24}$$

Furthermore, if $\varphi \in (L, \mathrm{dom}_m(L))$ is such that $\mathrm{D}_x\varphi \in C_m(\mathbb{R}_+ \times H; H)$, the approximating sequence $(\varphi_n)_n$ in (2.24) can be chosen such that

$$\pi\text{-}\lim_n \mathrm{D}_x\varphi_n = \mathrm{D}_x\varphi. \tag{2.25}$$

Proof. This follows from some slight generalisations of the results in Manca [82], where the π-core property of exponential test functions for generators of π-semigroups was studied in the context of time homogeneous SPDEs. For more details, see Appendix 2.B. □

Remark 2.14. Note that for $\varphi \in \mathcal{E}_A(\mathbb{R}_+ \times H)$, $L\varphi$ is in general not a bounded function anymore. Let for example $\varphi(t,x) = \sin(\langle x, h\rangle), h \in \mathrm{dom}(A^*)$. Then $\mathrm{D}_x\varphi = h\cos(\langle x, h\rangle)$ and $\mathrm{D}_x^2\varphi(x) = -h \otimes h \sin(\langle x, h\rangle)$ and thus

$$L\varphi = \cos(\langle x, h\rangle)(\langle x, A^*h\rangle + \langle F(t,x), h\rangle) - \frac{1}{2}\sin(\langle x, h\rangle)\langle Q^{\frac{1}{2}}h, Q^{\frac{1}{2}}h\rangle,$$

which is not bounded but of linear growth under the Lipschitz assumption on F. In particular, $\mathcal{E}_A(\mathbb{R}_+ \times H)$ is not a π-core for $(L, \mathrm{dom}_m(L))$ if $m = 0$.

2.3 Main result

In this section we present the proof of our main result. It establishes conditions on the h-function for which, under the exponential change of measure $\mathbb{P}^h$, the mild solution X to the SPDE (2.1) is a mild solution to yet another SPDE with an additional drift term dependent on the h-function of choice.
Our point of departure is that, following Lemma 2.12, for any $h \in \mathrm{dom}_m(L)$, the process

$$D^h(t) = h(t, X(t)) - \int_0^t Lh(s, X(s))\,\mathrm{d}s$$

is a $\mathbb{P}$-martingale. Additionally, it can be shown that for any positive function $h \in \mathrm{dom}_m(L)$, the process

$$E^h(t) = \frac{h(t, X(t))}{h(0, x_0)} \exp\left(-\int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s\right), \quad t \geq 0, \tag{2.26}$$

is a continuous $\mathbb{P}$-local martingale whenever it exists, see for example Palmowski and Rolski [89], Lemma 3.1. If E^h is a true $\mathbb{P}$-martingale, it follows from Lemma 2.2 that it defines a unique measure $\mathbb{P}^h$ on $\mathcal{F}$ such that

$$\frac{\mathrm{d}\mathbb{P}_t^h}{\mathrm{d}\mathbb{P}_t} = E^h(t), \quad t > 0. \tag{2.27}$$

Throughout this section we fix some arbitrary $m \geq 1$. We will need the following assumptions.

[1] This approximation relies on multi-indexed sequences, see Appendix 2.B. However, for ease of notation, we may assume that the sequence only has one index.

Assumption 2.15. *The function* $h : \mathbb{R}_+ \times H \longrightarrow \mathbb{R}_{>0}$ *satisfies:*

(i) $h \in \mathrm{dom}_m(L)$ *such that* $h^{-1}Lh \in C_m(\mathbb{R}_+ \times H)$.
(ii) h *is Fréchet differentiable in* x *such that* $\mathrm{D}_x h \in C_m(\mathbb{R}_+ \times H; H)$.
(iii) h *is such that* E^h *is a* $\mathbb{P}$*-martingale.*

The following is this sections main result.

Theorem 2.16. *Let* h *satisfy Assumption 2.15 and let* $\mathbb{P}^h$ *be the measure defined by (2.27). Then, for any* $T > 0$, *the process*

$$W^h(t) = W(t) - \int_0^t Q^{\frac{1}{2}} \mathrm{D}_x \log h(s, X(s))\, \mathrm{d}s, \quad t \in [0, T], \tag{2.28}$$

is a cylindrical Wiener process with respect to $\mathbb{P}^h$. *In particular,* X *under* $\mathbb{P}^h$ *is a mild solution to the SPDE*

$$\mathrm{d}X(t) = [AX(t) + F(t, X(t)) + Q\, \mathrm{D}_x \log h(t, X(t))]\, \mathrm{d}t + Q^{\frac{1}{2}}\, \mathrm{d}W^h(t), \quad t \in [0, T]. \tag{2.29}$$

The main part of the proof of Theorem 2.16 lies in the following lemma.

Lemma 2.17. *Let* h *satisfy Assumption 2.15 (i) and (ii) and let* D^h *be the corresponding Dynkin martingale. Then*

$$D^h(t) = h(0, x_0) + \int_0^t \langle Q^{\frac{1}{2}} \mathrm{D}_x h(s, X(s)), \mathrm{d}W(s)\rangle, \quad t \geq 0. \tag{2.30}$$

Furthermore, let E^h *be the process as defined in (2.26). Then* E^h *is the stochastic exponential given by*

$$E^h(t) = \exp\left(M^h(t) - \frac{1}{2}[M^h]_t\right), \quad t \geq 0, \tag{2.31}$$

where $M^h(t) = \int_0^t \langle Q^{\frac{1}{2}} \mathrm{D}_x \log h(s, X(s)), \mathrm{d}W(s)\rangle$.

Proof. *Step One.* We begin by showing the claim in Equation (2.30) for any exponential test function $h \in \mathcal{E}_A(\mathbb{R}_+ \times H)$.
Let $(A_n)_{n > \omega_S}$ be the Yosida approximation of A such that $\lim_n A_n x = Ax$ for all $x \in \mathrm{dom}(A)$ and let X_n be the sequence of strong solutions to equation (2.8). Define the sequence of processes $(D_n^h)_n$ via

$$D_n^h(t) = h(t, X_n(t)) - \int_0^t Lh(s, X_n(s))\, \mathrm{d}s. \tag{2.32}$$

We first show that $D_n^h(t) \xrightarrow{\mathbb{P}} D^h(t)$ for any $t \geq 0$. It follows from (2.9) that $h(t, X_n(t)) \xrightarrow{\mathbb{P}} h(t, X(t))$ and $Lh(t, X_n(t)) \xrightarrow{\mathbb{P}} Lh(t, X(t))$. Furthermore, it holds that

$$\begin{aligned} \sup_n \mathbb{E}\left[\left|\int_0^t Lh(s, X_n(s))\, \mathrm{d}s\right|\right] &\leq \sup_n \mathbb{E}\left[\int_0^t \frac{|Lh(s, X_n(s))|}{1 + \|X_n(s)\|^m}(1 + \|X_n(s)\|^m)\, \mathrm{d}s\right] \\ &\leq \|Lh\|_m \sup_n \int_0^t (1 + \mathbb{E}[\|X_n(s)\|^m])\, \mathrm{d}s \\ &< \infty. \end{aligned} \tag{2.33}$$

Here, we used in the last step that, by the bounds in (2.6) and (2.7), there exists for any $m \geq 1$ some $\tilde{C}_m > 0$ and $\tilde{\gamma}_m \geq 0$ such that

$$\sup_n \mathbb{E}[\|X_n(s)\|^m] \leq \tilde{C}_m \exp(\tilde{\gamma}_m s). \tag{2.34}$$

By the dominated convergence theorem it follows from (2.33) that

$$\int_0^t Lh(s, X_n(s))\,\mathrm{d}s \xrightarrow{L^1(\mathbb{P})} \int_0^t Lh(s, X(s))\,\mathrm{d}s$$

and therefore in total we get that $D_n^h(t) \xrightarrow{\mathbb{P}} D^h(t)$.
On the other hand, for any $n > \omega_S$, an application of Itô's lemma gives that

$$\mathrm{d}h(t, X_n(t)) = L_n h(t, X_n(t))\,\mathrm{d}t + \langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(t, X_n(t)), \mathrm{d}W(t)\rangle,$$

where $L_n h(t,x) = \partial_t h(t,x) + \langle x, A_n^*\,\mathrm{D}_x h(t,x)\rangle + \langle F(x), \mathrm{D}_x h(t,x)\rangle + \frac{1}{2}\operatorname{tr}[Q\,\mathrm{D}_x^2 h(t,x)]$ is the Kolmogorov operator associated with the approximating SPDE in (2.8). Plugging this into (2.32) gives that

$$\begin{aligned} D_n^h(t) &= h(0, x_0) + \int_0^t \big\langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(s, X_n(s)), \mathrm{d}W(s)\big\rangle \\ &\quad + \int_0^t [L_n h(s, X_n(s)) - Lh(s, X_n(s))]\,\mathrm{d}s \\ &= h(0, x_0) + \int_0^t \big\langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(s, X_n(s)), \mathrm{d}W(s)\big\rangle \\ &\quad + \int_0^t \big\langle X_n(s), (A_n^* - A^*)\,\mathrm{D}_x h(s, X_n(s))\big\rangle\,\mathrm{d}s. \end{aligned} \tag{2.35}$$

It remains to show that

$$\lim_n \int_0^t \big\langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(s, X_n(s)), \mathrm{d}W(s)\big\rangle = \int_0^t \big\langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(s, X(s)), \mathrm{d}W(s)\big\rangle \tag{2.36}$$

and

$$\lim_n \int_0^t \big\langle X_n(s), (A_n^* - A^*)\,\mathrm{D}_x h(s, X_n(s))\big\rangle\,\mathrm{d}s = 0 \tag{2.37}$$

in probability. Then (2.30) follows from $D_n^h(t) \xrightarrow{\mathbb{P}} D^h(t)$ and (2.35)-(2.37).
To show (2.36), note that by the Itô isometry and (2.34) we have

$$\begin{aligned} &\sup_n \mathbb{E}\left[\left\|\int_0^t \big\langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(s, X_n(s)), \mathrm{d}W(s)\big\rangle\right\|^2\right] \\ &\quad = \sup_n \int_0^t \mathbb{E}\left[\|Q^{\frac{1}{2}}\,\mathrm{D}_x h(s, X_n(s))\|^2\,\mathrm{d}s\right] \\ &\quad \le \sup_n \|Q\|\|\,\mathrm{D}_x h\|_m^2 \int_0^t \mathbb{E}[(1 + \|X_n(s)\|^m)^2]\,\mathrm{d}s \\ &\quad < \infty. \end{aligned}$$

Since $\mathrm{D}_x h(s, X_n(s)) \xrightarrow{\mathbb{P}} \mathrm{D}_x h(s, X(s))$, we get (2.36) by an application of the dominated convergence theorem.
Lastly, for any $h \in \mathcal{E}_A(\mathbb{R}_+ \times H)$, it holds that $\mathrm{D}_x h(s,x) = \sum_{i=1}^k z_i g_i(s,x)$ for some $z_i \in \operatorname{dom}(A^*)$ and $g_i \in \mathcal{E}_A(\mathbb{R}_+ \times H)$, $i = 1, \ldots k$. We show (2.37) for $k = 1$. The case that $k > 1$ follows from linearity. We have

$$\int_0^t \big\langle X_n(s), (A_n^* - A^*)\,\mathrm{D}_x h(s, X_n(s))\big\rangle\,\mathrm{d}s = \int_0^t \big\langle X_n(s), (A_n^* - A^*)\,z_1\big\rangle g_1(s, X_n(s))\,\mathrm{d}s.$$

By (2.9) it follows that $X_n \xrightarrow{\mathbb{P}} X$ in $C([0,t]; H)$ and in particular $\sup_n \sup_{s\in[0,t]} \|X_n(s)\| < \infty$ a.s. Thus (2.37) follows from $\lim_n (A_n^* - A^*)\,z_1 = 0$. In total this concludes that (2.30) holds for any $h \in \mathcal{E}_A(\mathbb{R}_+ \times H)$.

Step Two. We proceed to show (2.30) for any $h \in \mathrm{dom}_m(L)$ that satisfies Assumptions 2.15 (i) and (ii). By Lemma 2.13 there exists a sequence $(h_n)_n \in \mathcal{E}_A(\mathbb{R}_+ \times H)$ such that

$$\pi\text{-}\lim_n h_n = h, \;\; \pi\text{-}\lim_n \mathrm{D}_x h_n = \mathrm{D}_x h \text{ and } \pi\text{-}\lim_n Lh_n = Lh. \tag{2.38}$$

Denote by $D^{h_n}(t)$ the Dynkin martingale of (h_n, Lh_n). Firstly, it follows by dominated convergence, the bound in (2.6) and the π-convergence in (2.38) that

$$\begin{aligned} D^{h_n}(t) &= h_n(t, X(t)) - \int_0^t Lh_n(s, X(s))\,\mathrm{d}s \\ &\xrightarrow{\mathbb{P}} h(t, X(t)) - \int_0^t Lh(s, X(s))\,\mathrm{d}s \\ &= D^h(t). \end{aligned} \tag{2.39}$$

On the other hand, since $(h_n)_n \subset \mathcal{E}_A(\mathbb{R}_+ \times H)$, it follows from Step One that

$$D^{h_n}(t) = h_n(0, x_0) + \int_0^t \big\langle Q^{\frac{1}{2}}\,\mathrm{D}_x h_n(s, X(s)), \mathrm{d}W(s)\big\rangle, \quad n \in \mathbb{N}. \tag{2.40}$$

We already have that $\lim_n h_n(0, x_0) = h(0, x_0)$. It thus remains to show that

$$\int_0^t \big\langle Q^{\frac{1}{2}}\,\mathrm{D}_x h_n(s, X(s)), \mathrm{d}W(s)\big\rangle \xrightarrow{\mathbb{P}} \int_0^t \langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(s, X(s)), \mathrm{d}W(s)\rangle. \tag{2.41}$$

By Itô isometry it suffices to establish that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^t \|Q^{\frac{1}{2}}\,\mathrm{D}_x h_n(s, X(s))\|^2\,\mathrm{d}s\right] = \mathbb{E}\left[\int_0^t \|Q^{\frac{1}{2}}\,\mathrm{D}_x h(s, X(s))\|^2\,\mathrm{d}s\right],$$

but this again follows from an application of the dominated convergence theorem, the π-convergence of $\pi\text{-}\lim_n \mathrm{D}_x h_n = \mathrm{D}_x h$ and the bound in (2.6). In total, the claim (2.30) thus follows from (2.39) - (2.41).

Step Three. We finish the proof by showing the claim in (2.31). First note that by Assumption 2.15 (i) and the bound in (2.6), the process

$$t \mapsto \int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s \tag{2.42}$$

is finite $\mathbb{P}$-a.s. and the process E^h in (2.26) is therefore well-defined with $\mathbb{E}\left[E^h(0)\right] = 1$. Furthermore, the mapping in (2.42) is continuous and of finite variation and thus (see e.g. Palmowski and Rolski [89], Lemma 3.1)

$$\left[h(t, X(t)), \exp\left(-\int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s\right)\right]_t = 0. \tag{2.43}$$

By Lemma 2.12, the process $h(t, X(t))$ is a semimartingale and with the integration by parts formula for semimartingales it follows that

$$\begin{aligned} \mathrm{d}E^h(t) &= \frac{1}{h(0, x_0)}\left[h(t, X(t))\,\mathrm{d}\left(\exp\left(-\int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s\right)\right)\right. \\ &\quad \left. + \exp\left(-\int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s\right)\mathrm{d}h(t, X(t))\right] \\ &= \frac{1}{h(0, x_0)}\left[-\exp\left(-\int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s\right) Lh(t, X(t))\,\mathrm{d}t\right. \\ &\quad \left. + \exp\left(-\int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s\right)\mathrm{d}h(t, X(t))\right] \\ &= \frac{1}{h(0, x_0)}\exp\left(-\int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s\right)\mathrm{d}D^h(t), \end{aligned} \tag{2.44}$$

where $D^h(t) = h(t, X(t)) - \int_0^t Lh(s, X(s))\,\mathrm{d}s$ is the Dynkin martingale of (h, Lh). From step two it follows that $\mathrm{d}D^h(t) = \langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(t, X(t)), \mathrm{d}W(t)\rangle$ and plugging this into (2.44) gives

$$\begin{aligned}\mathrm{d}E^h(t) &= \frac{1}{h(0,x_0)} \exp\left(-\int_0^t \frac{Lh}{h}(s, X(s))\,\mathrm{d}s\right) \langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(t, X(t)), \mathrm{d}W(t)\rangle \\ &= E^h(t)\langle Q^{\frac{1}{2}}\,\mathrm{D}_x \log h(t, X(t)), \mathrm{d}W(t)\rangle.\end{aligned}$$

In particular, the process $E^h(t)$ is the stochastic exponential

$$E^h(t) = \exp\left(M^h(t) - \frac{1}{2}[M^h]_t\right)$$

of the Itô process

$$M^h(t) = \int_0^t \big\langle Q^{\frac{1}{2}}\,\mathrm{D}_x \log h(s, X(s)), \mathrm{d}W(s)\big\rangle.$$

□

We are now ready to give the remainder of the proof of Theorem 2.16, which is essentially just an application of Girsanov's theorem.

Proof of Theorem 2.16. Let E^h be the process as defined in (2.26). By Assumption 2.15 (iii), E^h is a $\mathbb{P}$-martingale and thus, following Lemma 2.2, defines a measure $\mathbb{P}^h$ on $\mathcal{F}$ such that $\mathrm{d}\mathbb{P}^h_T = E^h(T)\,\mathrm{d}\mathbb{P}_T$ on $\mathcal{F}_T$ for any $T > 0$. On the other hand, Lemma 2.17 shows that E^h is the stochastic exponential of the Itô process

$$M^h(t) = \int_0^t \big\langle Q^{\frac{1}{2}}\,\mathrm{D}_x \log h(s, X(s)), \mathrm{d}W(s)\big\rangle. \tag{2.45}$$

It therefore follows from Girsanov's theorem that

$$W^h(t) = W(t) - \int_0^t Q^{\frac{1}{2}}\,\mathrm{D}_x \log h(s, X(s))\,\mathrm{d}s, \quad t \in [0,T], \tag{2.46}$$

is a cylindrical Wiener process with respect to the measure $\mathbb{P}^h$. In particular, under $\mathbb{P}^h$, the process X given in (2.5) is a mild solution to the SPDE

$$\mathrm{d}X(t) = [AX(t) + F(t, X(t)) + Q\,\mathrm{D}_x \log h(t, X(t))]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W^h(t), \quad t \in [0,T].$$

□

2.3.1 The change of measure on a finite time interval

In certain applications we work with h-functions that are only defined on the half open interval $[0,T)$ for some $T > 0$. We then replace Assumption 2.15 with the following.

Assumption 2.18. *The function $h : [0,T) \times H \longrightarrow \mathbb{R}_+$ satisfies for any $S < T$:*

(i) *$h \in C_m([0,S] \times H)$ such that $Lh(t,x)$ exists for any $[0,S] \times H$ and $h^{-1}Lh \in C_m([0,S] \times H)$.*
(ii) *h is Fréchet differentiable in x such that $\mathrm{D}_x h \in C_m([0,S] \times H; H)$.*
(iii) *h is such that $(E^h(t))_{t\in[0,S]}$ is a $\mathbb{P}$-martingale.*

Under Assumption 2.18 (iii), $(E^h(t))_{t\in[0,T)}$ is a non-negative martingale with $\mathbb{E}[E(0)] = 1$. Following Lemma 2.4, it thus defines a measure $\mathbb{P}^h$ on $\mathcal{F}_T$ such that

$$\frac{\mathrm{d}\mathbb{P}^h_t}{\mathrm{d}\mathbb{P}_t} = E^h(t), \quad t \in [0,T). \tag{2.47}$$

We get the following version of Theorem 2.16.

Theorem 2.19. *Let h satisfy Assumption 2.18 and let $\mathbb{P}^h$ be the measure defined on $\mathcal{F}_T$ by (2.47). Then for any $S < T$, the process*

$$W^h(t) = W(t) - \int_0^t Q^{\frac{1}{2}} \,\mathrm{D}_x \log h(s, X(s))\, \mathrm{d}s, \quad t \in [0, S], \tag{2.48}$$

is a cylindrical Wiener process with respect to $\mathbb{P}^h$. In particular, X under $\mathbb{P}^h$ is a mild solution to the SPDE

$$\mathrm{d}X(t) = [AX(t) + F(t, X(t)) + Q\,\mathrm{D}_x \log h(t, X(t))]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W^h(t), \quad t \in [0, T). \tag{2.49}$$

Proof. Let $S < T$ be arbitrary but fixed. Let $\bar{h}$ be an extension of h onto $\mathbb{R}_+ \times H$ defined via

$$\bar{h}(t, x) = \begin{cases} h(t, x), & t \le S, \\ h(S, x), & t > S. \end{cases}$$

Then $\bar{h}$ satisfies Assumption 2.15. Following the proof of Lemma 2.17, we get that

$$E^h(t) = \exp\left(M^h(t) - \frac{1}{2}[M^h]_t\right), \quad t \in [0, S],$$

where $M^h(t) = \int_0^t \langle Q^{\frac{1}{2}} \,\mathrm{D}_x \log h(s, X(s)), \mathrm{d}W(s)\rangle$. The claim then follows from Assumption 2.18 (iii) and an application of Girsanov's theorem. □

Typically, the martingale property of E^h in Assumption 2.18 (iii) is the most difficult of the three to verify. The following lemma summarises conditions under which it is satisfied.

Lemma 2.20. *Either of the following conditions are sufficient for $(E^h(t))_{t\in[0,S]}$ to be a martingale:*

1. *In addition to Assumption 2.18 (i), $h^{-1}Lh \in C_0([0, S] \times H)$ is bounded.*
2. *In addition to Assumption 2.18 (i) and (ii), it holds that*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^S \|Q^{\frac{1}{2}}\,\mathrm{D}_x \log h(t, X(t))\|^2\,\mathrm{d}t\right)\right] < \infty.$$

3. *In addition to Assumption 2.18 (i) and (ii), the mapping*

$$(t, x) \mapsto Q^{\frac{1}{2}}\,\mathrm{D}_x \log h(t, x)$$

is Lipschitz in x, uniformly in $t \in [0, S]$.

Proof. The first condition has been shown in Palmowski and Rolski [89], Proposition 3.2. The second condition is the well-known Novikov condition, see for example Da Prato and Zabczyk [41], Proposition 10.17. A proof for the last condition can be found in Appendix 2.C. □

2.4 Applications

2.4.1 Conditioned SPDEs

In this subsection we introduce a class of h-functions for which the change of measure $\mathbb{P}^h$ corresponds to conditioning the process $(X(t))_{t\in[0,T)}$ on $X(T)$. Applications of this type of conditioning include, for example, the infinite-dimensional diffusion bridge.
For the construction of the h-functions in this section, we rely on the transition density of X with respect to some suitable reference measure. Since the state space of X is not Euclidean, the typical choice of the Lebesgue measure is not available to us. However, one can construct a suitable Gaussian reference measure on $(H, \mathcal{B}(H))$ as follows.

In addition to Assumption 2.5, let the following hold.

ASSUMPTION 2.21 (Strong Feller assumption). *The semigroup* $(S_t)_{t\geq 0}$ *and covariance operators* $(Q_t)_{t\geq 0}$ *defined in* (2.4) *are such that*

$$\operatorname{im}(S_t) \subset \operatorname{im}(Q_t^{1/2}), \quad t \geq 0.$$

Under Assumption 2.21, the Ornstein-Uhlenbeck process Z is strongly Feller and absolutely continuous with respect to its invariant measure $\nu \sim \mathcal{N}(0, Q_\infty)$ with covariance operator Q_∞ as defined in (2.10).

Furthermore, it follows from the Girsanov theorem that $\mathcal{L}(X(t,s,x)) \sim \mathcal{L}(Z(t,s,x))$ for all $t \geq s$. In particular, we have that $\mathcal{L}(X(t,s,x)) \sim \nu$ and we define by $p_X(s,x;t,y)$ the transition density

$$p_X(s,x;t,y) = \frac{\mathrm{d}\mathcal{L}(X(t,s,x))}{\mathrm{d}\nu}(y), \quad \nu\text{-a.e. } y \in H, \tag{2.50}$$

of X with respect to ν. From the Markov property of X it follows that $p_X(s,x;t,y)$ satisfies the *Chapman-Kolmogorov* equation

$$p_X(s,x;t,y) = \int_H p_X(s,x;r,z) p_X(r,z;t,y)\, \nu(\mathrm{d}z) \tag{2.51}$$

for all $s < r < t$ and for ν-a.e. every $y \in H$. Let us now define the function $h : [0,T) \times H \to \mathbb{R}_{>0}$ via

$$h(t,x) = \int_H p_X(t,x;T,y)\, \mu(\mathrm{d}y), \tag{2.52}$$

for some measure μ on $(H, \mathcal{B}(H))$ such that $\int_H p_X(0,x_0;T,y)\, \mu(\mathrm{d}y) < \infty$. The h-function constructed in (2.52) satisfies the following.

LEMMA 2.22. *The function* h *is space-time harmonic, i.e. for any* $(s,x) \in [0,T) \times H$ *it holds that*

$$(T_t h)(s,x) = h(s,x), \quad t < T - s.$$

In particular, $Lh = 0$ *and* $h(t, X(t))$ *is a* $\mathbb{P}$*-martingale.*

PROOF. Let $(s,x) \in [0,T) \times H$ be fixed. Then for any $t < T - s$ we have

$$\begin{aligned}
(T_t h)(s,x) &= \mathbb{E}\left[h(t+s, X(t+s,s,x))\right] \\
&= \int_H h(t+s,z) p_X(s,x;t+s,z)\, \nu(\mathrm{d}z) \\
&= \int_H \int_H p_X(t+s,z;T,y) p_X(s,x;t+s,z)\, \nu(\mathrm{d}z) \mu(\mathrm{d}y) \\
&= \int_H p_X(s,x;t,y)\, \mu(\mathrm{d}y) \\
&= h(s,x),
\end{aligned}$$

where we use the definition of h in the third and the Chapman-Kolmogorov equation the fourth line. By definition of L it follows that $Lh = 0$. Furthermore, using the Markov property of $(t, X(t))$ we have for any $s < t < T$ that

$$\mathbb{E}[h(t, X(t)) \mid \mathcal{F}_s] = (T_{t-s} h)(s, X(s)) = h(s, X(s)).$$

□

From Lemma 2.22 it follows that h satisfies Assumption 2.18 (i) and (iii). In particular, $h(t,X(t))/h(0,x_0)$ is a non-negative martingale with mean one and thus defines a unique measure $\mathbb{P}^h$ on $\mathcal{F}_T$ via

$$\frac{\mathrm{d}\mathbb{P}^h_t}{\mathrm{d}\mathbb{P}_t} = \frac{h(t,X(t))}{h(0,x_0)}, \quad t < T. \tag{2.53}$$

On the other hand, Assumption 2.18 (ii) is hard to verify in general, as the Fréchet differentiability of h depends on $p_X(s,x;t,y)$ as well as the choice of μ. We thus keep it as an assumption.

Assumption 2.23. *The function h defined in (2.52) is Fréchet differentiable in x such that $\mathrm{D}_x h \in C_m([0,S]\times H; H)$ for any $S < T$.*

The following result shows how the measure $\mathbb{P}^h$ changes the law of X.

Proposition 2.24. *Let $\mathbb{P}^h$ be the measure defined by (2.53). It then holds for any bounded and measurable function f and $0 \le t_1 \le \ldots \le t_n < T$ that*

$$\mathbb{E}^h[f(X(t_1),\ldots,X(t_n))] = \int_H \mathbb{E}[f(X(t_1),\ldots,X(t_n)) \mid X(T)=y]\,\xi(\mathrm{d}y), \tag{2.54}$$

where ξ is the measure defined on $(H,\mathcal{B}(H))$ via

$$\xi(\mathrm{d}y) = \frac{p_X(0,x_0;T,y)\mu(\mathrm{d}y)}{\int_H p_X(0,x_0;T,y)\,\mu(\mathrm{d}y)}. \tag{2.55}$$

Additionally, if Assumption 2.23 is satisfied, X under $\mathbb{P}^h$ satisfies the SPDE

$$\mathrm{d}X(t) = [AX(t) + F(t,X(t)) + Q\,\mathrm{D}_x \log h(t,X(t))]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W^h(t), \quad t\in[0,T),$$

where W^h is the $\mathbb{P}^h$-cylindrical Wiener process as defined in (2.48).

Proof. To show (2.54) it suffices to show that

$$\mathbb{E}^h[f(X(t))] = \int_H \mathbb{E}[f(X(t)) \mid X(T)=y]\,\xi(\mathrm{d}y)$$

for any $t < T$ and continuous and bounded $g : H \to \mathbb{R}$. The claim then follows by a standard cylindrical argument, see e.g. Ethier and Kurtz [51], Proposition 4.1.6. Indeed it holds that

$$\begin{aligned}
\mathbb{E}^h[f(X(t))] &= \mathbb{E}\left[f(X(t))\frac{h(t,X(t))}{h(0,x_0)}\right]\\
&= \int_H f(x)\frac{h(t,x)}{h(0,x_0)}p_X(0,x_0;t,x)\,\nu(\mathrm{d}x)\\
&= \int_H f(x)\frac{\int_H p_X(t,x;T,y)\,\mu(\mathrm{d}y)}{\int_H p_X(0,x_0;T,y)\,\mu(\mathrm{d}y)}p_X(0,x_0;t,x)\,\nu(\mathrm{d}x)\\
&= \int_H\int_H f(x)\frac{p_X(0,x_0;t,x)p_X(t,x;T,y)}{p_X(0,x_0;T,y)}\,\nu(\mathrm{d}x)\frac{p_X(0,x_0;T,y)}{\int_H p_X(0,x_0;T,y)\,\mu(\mathrm{d}y)}\,\mu(\mathrm{d}y)\\
&= \int_H \mathbb{E}[f(X(t)) \mid X(T)=y]\,\xi(\mathrm{d}y)
\end{aligned}$$

with $\xi(\mathrm{d}y) = \frac{p_X(0,x_0;T,y)\mu(\mathrm{d}y)}{\int_H p_X(0,x_0;T,y)\,\mu(\mathrm{d}y)}$. Here we used the definition of h in the third, Fubini's theorem in the fourth and Bayes' theorem in the last step.
The second claim of the proposition is a direct consequence of Theorem 2.19 under Assumption 2.23. ∎

The formula

$$\mathbb{E}^h[f(X(t_1),...,X(t_n))] = \int_H \mathbb{E}[f(X(t_1),...,X(t_n)) \mid X(T) = y]\, \xi(\mathrm{d}y)$$

provides a disintegration of the conditioned process: to draw from it one first generates an endpoint $X(T) = y$ from $\xi(\mathrm{d}y)$, followed by drawing the path conditioned on this value of y, see Example 2.25 below. Different choices of the measure μ in (2.52) correspond to different kinds of conditioning. We illustrate this in the following examples.

Example 2.25 (The infinite-dimensional diffusion bridge). Let $y \in H$ be such that $p_X(t, x; T, y)$ is well-defined. [2] Set $\mu = \delta_y$ as the Dirac measure

$$\delta_y(A) = \begin{cases} 1, & \text{if } y \in A \\ 0, & \text{else.} \end{cases}$$

It follows that $\xi = \delta_y$ and thus the formula (2.54) reduces to

$$\mathbb{E}^h[f(X(t_1),...,X(t_n))] = \mathbb{E}[f(X(t_1),...,X(t_n)) \mid X(T) = y].$$

In other words, X under $\mathbb{P}^h$ is the process $(X(t))_{t\in[0,T)}$ conditioned on hitting the endpoint $X(T) = y$. We refer to this process as the *infinite-dimensional diffusion bridge*. If the transition density $p_X(t, x; T, y)$ satisfies Assumption 2.23, the infinite-dimensional diffusion bridge is characterised by the bridge equation

$$\mathrm{d}X^\star(t) = [AX^\star(t) + F(t, X^\star(t)) + Q\,\mathrm{D}_x \log p_X(t, X^\star(t); T, y)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W(t). \quad (2.56)$$

This equation corresponds to the well known diffusion bridge equation for the case of a finite-dimensional state space.

Example 2.26 (Conditioning on a noisy observation). Suppose we do not observe the endpoint $X(T) = y$ directly, but instead we observe a sample v from a distribution $q(v \mid y)\nu(\mathrm{d}v)$ where $q(\cdot \mid y)$ is some probability density function with respect to ν. This corresponds to observing $X(T)$ under noise q. Furthermore, assume q is such that

$$\mu(\mathrm{d}y) = q(v \mid y)\nu(\mathrm{d}y)$$

is a finite measure. It then follows that

$$\xi(\mathrm{d}y) = \frac{p_X(0, x_0; T, y)q(v \mid y)\nu(\mathrm{d}y)}{\int p_X(0, x_0; T, y)q(v \mid y)\,\nu(\mathrm{d}y)}.$$

This has a nice Bayesian interpretation where the endpoint y gets assigned prior density $\pi(y) = p_X(0, x_0; T, y)$ and the observation is given by v. The likelihood for this observation is $\ell(y \mid v) = q(v \mid y)$ and hence

$$\xi(\mathrm{d}y) = \frac{\pi(y)\ell(y \mid v)\nu(\mathrm{d}y)}{\int \pi(y)\ell(y \mid v)\,\nu(\mathrm{d}y)}.$$

This shows that $\xi(\mathrm{d}y)$ gives the posterior measure of y, conditional upon observing v. Therefore, using this h-transform, the conditioned process is constructed by first sampling the endpoint y conditional on the observation, followed by sampling the bridge to y.

Example 2.27 (The forced/tilted process). Let q be some arbitrary density function with respect to ν such that

$$\mu(\mathrm{d}y) = \frac{q(y)}{p_X(0, x_0; T, y)}\nu(\mathrm{d}y)$$

[2] This holds for ν-a.e. $y \in H$.

defines a finite measure on $(H, \mathcal{B}(H))$. By straightforward substitution we get that

$$h(t,x) = \int_H \frac{p_X(t,x;T,y)}{p_X(0,x_0;T,y)} q(y)\,\nu(\mathrm{d}y)$$

and $\xi(\mathrm{d}y) = q(y)\nu(\mathrm{d}y)$. Hence this corresponds to *forcing/tilting* $X(T)$ to have the distribution $q(y)\nu(\mathrm{d}y)$.

2.4.2 Guided processes

In Example 2.25, setting $h(t,x) = p_X(t,x;T,y)$, we have derived the infinite-dimensional diffusion bridge equation

$$\mathrm{d}X^\star(t) = [AX^\star(t) + F(t, X^\star(t)) + Q\,\mathrm{D}_x \log p_X(t, X^\star(t); T, y)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W(t)$$

that characterises the law of the conditioned process $X(t) \mid X(T) = y$ for ν-a.e. $y \in H$. In many practical applications, one seeks to draw samples from $X^\star$. In general, however, the transition density $p_X(t,x;T,y)$ is not tractable, rendering a direct simulation of $X^\star$ infeasible.

This motivates the following construction. Let $p_Z(t,x;T,y)$ be a tractable density of the mild solution Z of another auxiliary SPDE. Then, setting $g(t,x) = p_Z(t,x;T,y)$, one aims to define a change of measure $\mathbb{P}^g$ such that

(i) X under $\mathbb{P}^g$ equals in law the mild solution X° to the SPDE

$$\mathrm{d}X^\circ(t) = [AX^\circ(t) + F(t, X^\circ(t)) + Q\,\mathrm{D}_x \log g(t, X^\circ(t))]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W(t), \quad t \in [0,T).$$

(ii) $\mathbb{P}^h$ and $\mathbb{P}^g$ - and therefore, the laws of $X^\star$ and X° on $C([0,T];H)$ - are absolutely continuous with some likelihood ratio Φ.

Samples of the bridge process $X^\star$ can then be obtained by drawing proposal samples from the law of X° and accepting or rejecting the proposals based on the likelihood ratio Φ. We now showcase the idea of this construction in more detail.

For this, let p_Z denote the transition density of the OU process Z with respect to ν, i.e.

$$p_Z(s,x;t,y) = \frac{\mathrm{d}\mathcal{L}(Z(t,s,x))}{\mathrm{d}\nu}(y), \quad \nu\text{-a.e. } y \in H, s < t.$$

Since Z is a Gaussian process and ν a Gaussian measure on $(H, \mathcal{B}(H))$, the densities $p_Z(s,x;t,y)$ can be obtained by the Cameron-Martin formula. We shall place an additional assumption on the covariance operators of Z.

Assumption 2.28. *The covariance operators $(Q_t)_{t\geq 0}$ as defined in (2.4) commute.*

The following proposition shows that p_Z defines a function g that satisfies Assumption 2.18 and that the SPDE induced by the changed measure $\mathbb{P}^g$ remains tractable.

Proposition 2.29. *For ν-a.e. $y \in H$, the function $g(t,x) = p_Z(t,x;T,y)$ satisfies Assumption 2.18. Moreover,*

$$\mathrm{D}_x \log g(t,x) = \Gamma^*_{T-t} Q^{-\frac{1}{2}}_{T-t}[y - S(T-t)x], \tag{2.57}$$

where $\Gamma_{T-t} = Q^{-\frac{1}{2}}_{T-t} S_{T-t}$ is a bounded operator on H by Assumption 2.21. In particular, there exists a unique change of measure $\mathbb{P}^g$ such that X under $\mathbb{P}^g$ is a mild solution to the SPDE

$$\mathrm{d}X^\circ(t) = \Big[AX^\circ(t) + F(t, X^\circ(t)) + Q\Gamma^*_{T-t} Q^{-\frac{1}{2}}_{T-t}[y - S(T-t)X^\circ(t)]\Big]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W^g(t). \tag{2.58}$$

Proof. Let $S < T$ be arbitrary but fixed. We start by showing that Assumption 2.18 (ii) is satisfied, i.e. that $g(t,x)$ is Fréchet differentiable in x on $[0,S] \times H$ with derivative $\mathrm{D}_x g(t,x)$ of at most polynomial growth in x. For this, write

$$p_Z(t,x;T,y) = \frac{\mathrm{d}\mathcal{L}(Z(T,t,x))}{\mathrm{d}\mathcal{L}(Z(T,t,0))}(y)\frac{\mathrm{d}\mathcal{L}(Z(T,t,0))}{\mathrm{d}\nu}(y), \quad \nu\text{-a.e. } y \in H.$$

Since Z is a Gaussian process, the Cameron-Martin formula gives that

$$\begin{aligned}\frac{\mathrm{d}\mathcal{L}(Z(T,t,x))}{\mathrm{d}\mathcal{L}(Z(T,t,0))}(y) &= \exp\left(\langle Q_{T-t}^{-\frac{1}{2}}y, Q_{T-t}^{-\frac{1}{2}}S_{T-t}x\rangle - \frac{1}{2}\|Q_{T-t}^{-\frac{1}{2}}S_{T-t}x\|^2\right)\\ &= \exp\left(\langle \Gamma_{T-t}^* Q_{T-t}^{-\frac{1}{2}}y, x\rangle - \frac{1}{2}\|\Gamma_{T-t}x\|^2\right), \quad \nu\text{-a.e. } y \in H,\end{aligned}$$

where the mapping $t \mapsto \Gamma_{T-t}^* Q_{T-t}^{-\frac{1}{2}}y$ is continuous for ν-a.e. $y \in H$, see Appendix 2.D for details. It follows that, for ν-a.e. $y \in H$ fixed, the function $g(t,x) = p_Z(t,x;T,y)$ is well-defined and bounded on $[0,S] \times H$. Moreover, it is Fréchet differentiable in x with derivative

$$\mathrm{D}_x g(t,x) = g(t,x)\Gamma_{T-t}^*\left(Q_{T-t}^{-\frac{1}{2}}y - \Gamma_{T-t}x\right).$$

It holds that $(S_t)_t$ and $(Q_t)_t$ are strongly continuous, from which it follows that $t \mapsto \Gamma_{T-t}x$ is continuous for any fixed $x \in H$. In particular, by the uniform boundedness principle and the continuity of $t \mapsto \Gamma_{T-t}^* Q_{T-t}^{-\frac{1}{2}}y$, we have that

$$\sup_{t\in[0,S]}\left(\|\Gamma_{T-t}\| + \|\Gamma_{T-t}^* Q_{T-t}^{-\frac{1}{2}}y\|\right) < \infty.$$

In total we get that $\mathrm{D}_x g \in C_1([0,S]\times H; H)$ since

$$\sup_{t\in[0,S],x\in H}\frac{\|\mathrm{D}_x g(t,x)\|}{1+\|x\|} \le \sup_{t\in[0,S],x\in H}|g(t,x)|\left(\frac{\|\Gamma_{T-t}^* Q_{T-t}^{-\frac{1}{2}}y\|}{1+\|x\|} + \frac{\|\Gamma_{T-t}^*\Gamma_{T-t}x\|}{1+\|x\|}\right) < \infty.$$

We proceed to show that g satisfies Assumption 2.18 (i). By the Fréchet differentiability of g, it follows from Manca [82], Theorem 4.1, that

$$(Lg)(t,x) = \tilde{L}g(t,x) + \langle F(t,x), \mathrm{D}_x g(t,x)\rangle,$$

where $\tilde{L}$ is the infinitesimal generator of the Ornstein-Uhlenbeck π-semigroup. By the same arguments as in Lemma 2.22, one shows that g is harmonic with respect to $\tilde{L}$, i.e. $\tilde{L}g(t,x) = 0$ for any $t < T$. With the Lipschitz continuity of F, it thus follows that $(Lg)(t,x) \in C_m([0,S]\times H)$ for any $m \ge 2$ with

$$(Lg)(t,x) = \langle F(t,x), \mathrm{D}_x g(t,x)\rangle.$$

Likewise, we have that $g^{-1}Lg \in C_m([0,S]\times H)$ for any $m \ge 2$ since

$$\begin{aligned}(g^{-1}Lg)(t,x) &= \langle F(t,x), \mathrm{D}_x \log g(t,x)\rangle\\ &= \left\langle F(t,x), \Gamma_{T-t}^*\left(Q_{T-t}^{-\frac{1}{2}}y - \Gamma_{T-t}x\right)\right\rangle.\end{aligned}$$

Therefore, g satisfies Assumption 2.18 (i) and E^g is a well-defined continuous $\mathbb{P}$-local martingale given by

$$E^g(t) = \frac{g(t,X(t))}{g(0,x_0)}\exp\left(-\int_0^t \left\langle F(s,X(s)), \Gamma_{T-s}^*\left(Q_{T-s}^{-\frac{1}{2}}y - \Gamma_{T-s}X(s)\right)\right\rangle \mathrm{d}s\right). \quad (2.59)$$

It remains to show that E^g is a true $\mathbb{P}$-martingale. However, this follows directly from the uniform Lipschitz continuity of $(t,x) \mapsto Q^{\frac{1}{2}}\mathrm{D}_x \log g(t,x)$ and Lemma 2.20. □

By construction we immediately get the following.

COROLLARY 2.30. *The measures $\mathbb{P}^h$ and $\mathbb{P}^g$ are absolutely continuous on $\mathcal{F}_t, t < T$, with likelihood ratio*

$$\frac{\mathrm{d}\mathbb{P}^h_t}{\mathrm{d}\mathbb{P}^g_t}(X) = \frac{h(t,X(t))}{g(t,X(t))}\frac{g(0,x_0)}{h(0,x_0)} \exp\left(\int_0^t \left\langle F(s,X(s)), \Gamma^*_{T-s}\left(Q_{T-s}^{-\frac{1}{2}}y - \Gamma_{T-s}X(s)\right)\right\rangle \mathrm{d}s\right).$$

PROOF. This follows from plugging in the likelihood functions $\mathrm{d}\mathbb{P}^h_t = E^h(t)\,\mathrm{d}\mathbb{P}_t$ and $\mathrm{d}\mathbb{P}^g_t = E^g(t)\,\mathrm{d}\mathbb{P}_t$ as given in (2.59) into $\mathrm{d}\mathbb{P}^h_t/\,\mathrm{d}\mathbb{P}^g_t = \mathrm{d}\mathbb{P}^h_t/\,\mathrm{d}\mathbb{P}_t\ \ \mathrm{d}\mathbb{P}_t/\,\mathrm{d}\mathbb{P}^g_t$. □

REMARK 2.31. Corollary 2.30 shows the absolute continuity of the measures $\mathbb{P}^h$ and $\mathbb{P}^g$ on any $\mathcal{F}_t, t < T$. In other words, the measures are equivalent as long as we 'stay away from the conditioning time T'. However, in order to draw samples from the bridge process $X^\star$, we require absolute continuity of $\mathbb{P}^h$ and $\mathbb{P}^g$ on $\mathcal{F}_T$, i.e. on the complete interval $[0,T]$.
To demonstrate this, note that $\mathbb{P}^h_t \sim \mathbb{P}_t$ for any $t < T$, but under $\mathbb{P}$ the event $\{X(T) = y\}$ has measure zero, implying that samples of X under $\mathbb{P}$ will almost surely not hit the endpoint $X(T) = y$.
In contrast to this, the additional drift term $Q\,\mathrm{D}_x \log p_Z(t,x;T,y)$ forces the process X° to hit the conditioning point $X^\circ(T) = y$, and we hypothesise that this results in absolute continuity of the laws of the processes $X^\star$ and X° on the complete interval.
In the case of observing a finite-dimensional representation of X_T, the proof of this result can be found in the next chapter. For a fully observed state, proving absolute continuity on the complete time interval remains an open problem.

Acknowledgements

We thank professor B. Goldys (University of Sydney) for providing the arguments in Lemma 2.37 and his comments on the regularity of the OU transition density.

2.A

LEMMA 2.32 (Lemma 2.4 above). *Let $(E(t))_{t\in[0,T)}$ be a non-negative martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T)}, \mathbb{P})$ with $\mathbb{E}[E(0)] = 1$. Then there exists a unique measure $\mathbb{Q}$ on $\mathcal{F}_T$ such that $\mathrm{d}\mathbb{Q}_t = E(t)\,\mathrm{d}\mathbb{P}_t$ for all $0 \le t < T$.*

PROOF. *Uniqueness.* By left-continuity of $(\mathcal{F}_t)_t$ we have that $\mathcal{F}_T = \sigma(\bigcup_{t<T}\mathcal{F}_t)$. Thus, if $\mathbb{Q}$ and $\tilde{\mathbb{Q}}$ are two measures on $\mathcal{F}_T$ such that $\mathbb{Q}_t = \tilde{\mathbb{Q}}_t$ for all $t < T$ it immediately follows that $\mathbb{Q} = \tilde{\mathbb{Q}}$.
Existence. For any $n \in \mathbb{N}$, let $t_n = T - \frac{1}{n}$ and let $\eta_n(\omega)(t) = \omega(t \wedge t_n)$ be the stopped canonical process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T)}, \mathbb{P})$. Furthermore, let $\{\mathbb{Q}_n\}$ be the measures on $\mathcal{F}_T$ defined by

$$\mathbb{Q}_n(A) = \mathbb{E}\left[E_{t_n}\mathbb{1}_{\eta_n^{-1}(A)}\right], \quad A \in \mathcal{F}_T. \tag{2.60}$$

Noting that $\eta_{n+1}^{-1}(A) = \eta_n^{-1}(A)$ for any $A \in \mathcal{F}_{t_n}$ it holds that

$$\begin{aligned}\mathbb{Q}_{n+1}(A) &= \mathbb{E}\left[\mathbb{E}\left[E_{t_{n+1}}\mathbb{1}_{\eta_n^{-1}(A)} \mid \mathcal{F}_{t_n}\right]\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\eta_n^{-1}(A)}\mathbb{E}\left[E_{t_{n+1}} \mid \mathcal{F}_{t_n}\right]\right] \\ &= \mathbb{Q}_n(A),\end{aligned} \tag{2.61}$$

i.e. $\mathbb{Q}_{n+1|\mathcal{F}_{t_n}} = \mathbb{Q}_{n|\mathcal{F}_{t_n}}$ for all $n \in \mathbb{N}$.

Denote by $\mathcal{S}$ the algebra $\mathcal{S} = \bigcup_{n\geq 1} \mathcal{F}_{t_n}$. By left-continuity of $(\mathcal{F}_t)_t$ it holds that $\mathcal{F}_T = \sigma(\mathcal{S})$ and from (2.61) it follows that the mapping

$$\mathbb{Q}(A) = \lim_n \mathbb{Q}_n(A), \quad A \in \mathcal{S},$$

is well-defined and a pre-measure on $\mathcal{S}$. It therefore follows from the Carathéodory extension theorem that $\mathbb{Q}$ extends uniquely to a measure on $\mathcal{F}_T$ and it is straightforward to check that $\mathbb{Q}$ satisfies $\mathbb{Q}(A) = \mathbb{E}[E_t \mathbb{1}_A]$ for any $A \in \mathcal{F}_t, t < T$. □

2.B

We give more details on the approximation properties of the exponential test functions $\mathcal{E}_A(\mathbb{R}_+ \times H)$ with respect to π-convergence in $C_m(\mathbb{R}_+ \times H)$. For this, it does not suffice to consider single-indexed sequences. Hence, the results stated in this section rely on using k-indexed sequences, i.e. sequences $(\varphi_{n_1,\dots,n_k})_{n_1,\dots,n_k}$ that depend on k indices for some $k \in \mathbb{N}$.
A k-indexed sequence $(\varphi_{n_1,\dots,n_k})_{n_1,\dots,n_k}$ in $C_m(\mathbb{R}_+ \times H)$ is said to be π-convergent to some φ in $C_m(\mathbb{R}_+ \times H)$ if for any $i \in \{2, \dots, k\}$ there exists an $(i-1)$-indexed sequence $(\varphi_{n_1,\dots,n_{i-1}})_{n_1,\dots,n_{i-1}}$ in $C_m(\mathbb{R}_+ \times H)$ such that

$$\varphi_{n_1,\dots,n_{i-1}} = \pi\text{-}\lim_{n_i} \varphi_{n_1,\dots,n_i}$$

and $\varphi = \pi\text{-}\lim_{n_1} \varphi_{n_1}$. We then write $\pi\text{-}\lim_{n_1,\dots,n_k} \varphi_{n_1,\dots,n_k} = \varphi$.
The following is a slight generalisation of Manca [81], Proposition 2.7.

Lemma 2.33. *Let $\varphi \in C_0(\mathbb{R}_+ \times H)$ such that $\mathrm{D}_x\varphi \in C_0(\mathbb{R}_+ \times H; H)$. Then there exists a three-indexed sequence $(\varphi_{n_1,n_2,n_3})_{n_1,n_2,n_3} \subset \mathcal{E}_A(\mathbb{R}_+ \times H)$ such that*

$$\begin{aligned} \pi\text{-}\lim_{n_1,n_2,n_3} \varphi_{n_1,n_2,n_3} &= \varphi, \\ \pi\text{-}\lim_{n_1,n_2,n_3} \mathrm{D}_x\varphi_{n_1,n_2,n_3} &= \mathrm{D}_x\varphi. \end{aligned} \tag{2.62}$$

Proof. We give a rough sketch only.
Step One. We first show (2.62) for the case that $H = \mathbb{R}^d$ and the case that $Ax = x$. We denote by $\mathcal{E}(\mathbb{R}_+ \times H)$ the space of exponential test functions as defined in (2.22) for the identity operator $Ax = x$.
For any $n \in \mathbb{N}$, let $\psi_n \in C_0(\mathbb{R}_+ \times \mathbb{R}^d)$ be such that

(i) $\psi_n = \varphi$ and $\mathrm{D}_x\psi_n = \mathrm{D}_x\varphi$ on $(0, 2n) \times (-n, n)^d$,
(ii) ψ_n and $\mathrm{D}_x\psi_n$ are $2n$-periodic in any direction,
(iii) $\|\psi_n\|_0 \leq \|\varphi\|_0$ and $\|\mathrm{D}_x\psi_n\|_0 \leq \|\mathrm{D}_x\varphi\|_0$.

Then, for any fixed $n \in \mathbb{N}$, let $(\psi_{n,m})_{n,m}$ be the sequence given by

$$\begin{aligned} \psi_{n,m}(t,x) &= \left(\frac{1}{2n}\right)^{d+1} \int_{[0,2n]\times[-n,n]^d} \psi_n(t-s, x-y) F_{n,m}(y)\, \mathrm{d}s\, \mathrm{d}y \\ &= (\psi_n * F_{n,m})(t,x), \end{aligned}$$

where $F_{n,m}$ is the m-th order Fejér kernel of period $2n$. It can be shown that $\psi_{n,m}$ equals the Cesàro mean of the m-th partial Fourier sum of ψ_n and hence $(\psi_{n,m})_{n,m} \subset \mathcal{E}(\mathbb{R}_+ \times \mathbb{R}^d)$.
Note that $\|\psi_{n,m}\|_0 \leq \|\psi_n\|_0$ for all $n, m \in \mathbb{N}$. Furthermore, since $\mathrm{D}_x\psi_n$ is continuous and bounded, it holds by standard properties of the convolution operator that

$$\mathrm{D}_x\psi_{n,m}(t,x) = (\mathrm{D}_x\psi_n * F_{n,m})(t,x)$$

with $\| \mathrm{D}_x\psi_{n,m}\|_0 \leq \| \mathrm{D}_x\psi_n\|_0$. It follows from Fejér's theorem that

$$\lim_m \|\psi_{n,m} - \psi_n\|_0 = 0,$$
$$\lim_m \| \mathrm{D}_x\psi_{n,m} - \mathrm{D}_x\psi_n\|_0 = 0.$$

By diagonalisation of $(\psi_{n,m})_{n,m}$[3] we find a sequence $(\varphi_n)_n \subset \mathcal{E}(\mathbb{R}_+ \times \mathbb{R}^d)$ that satisfies (2.62).

Step Two. We keep the assumption that $Ax = x$ but now assume that H is infinite-dimensional with some orthonormal basis $(e_j)_j$. Let $P_{n_1}x = \sum_{j=1}^{n_1}\langle x, e_j\rangle e_j$ be the orthogonal projection onto $H^{n_1} = \mathrm{span}\{e_j : j = 1, ..., n_1\}$. Then, for any n_1 fixed and $\varphi \in C_0(\mathbb{R}_+ \times H)$ consider the function

$$\varphi_{n_1}(t,x) = \varphi(t, P_{n_1}x).$$

By the first step above, there exists a two-indexed sequence $(\varphi_{n_1,n_2})_{n_1,n_2}$ such that $\pi\text{-}\lim_{n_2}\varphi_{n_1,n_2} = \varphi_{n_1}$ and $\pi\text{-}\lim_{n_2}\mathrm{D}_x\varphi_{n_1,n_2} = \mathrm{D}_x\varphi_{n_1}$. In particular, from $\lim_{n_1}\varphi_{n_1}(t,x) = \varphi(t,x)$ and

$$\begin{aligned}\lim_{n_1}(\mathrm{D}_x\varphi_{n_1})(t,x) &= \lim_{n_1} P_{n_1}(\mathrm{D}_x\varphi)(t, P_{n_1}x)\\ &= \mathrm{D}_x\varphi(t,x)\end{aligned}$$

as well as $\|\varphi_{n_1}\|_0 < \|\varphi\|_0$ and $\| \mathrm{D}\varphi_{n_1}\|_0 \leq \| \mathrm{D}\varphi\|_0$ it follows that (2.62) holds.

Step Three. Now let H be infinite-dimensional and A be the generator of a C_0-semigroup on H as in Assumption 2.5. For any $n > \omega_S$ let $R(n,A) = (n-A)^{-1}$ be the resolvent of A such that $\lim_n nR(n,A)x = x$ and $R(n,A)x \in \mathrm{dom}(A)$ for any $x \in H$. Let φ be as assumed in the lemma. By step two there exists a sequence $(\varphi_{n_1,n_2})_{n_1,n_2} \subset \mathcal{E}(\mathbb{R}_+ \times H)$ such that $\pi\text{-}\lim_{n_1,n_2}\varphi_{n_1,n_2} = \varphi$ and $\pi\text{-}\lim_{n_1,n_2}\mathrm{D}_x\varphi_{n_1,n_2} = \mathrm{D}_x\varphi$. Then, setting

$$\varphi_{n_1,n_2,n_3}(t,x) = \varphi_{n_1,n_2}(t, n_3R(n_3,A)x)$$

it holds that $(\varphi_{n_1,n_2,n_3})_{n_1,n_2,n_3} \subset \mathcal{E}_A(\mathbb{R}_+ \times H)$ and it is straightforward to show that (2.62) holds. □

LEMMA 2.34. *Let $\varphi \in C_m(\mathbb{R}_+ \times H)$ be such that $\mathrm{D}_x\varphi \in C_m(\mathbb{R}_+ \times H; H)$. Then there exists a four-indexed sequence $(\varphi_{n_1,n_2,n_3,n_4})_{n_1,n_2,n_3,n_4} \subset \mathcal{E}_A(\mathbb{R}_+ \times H)$ such that*

$$\begin{aligned}\pi\text{-}\lim_{n_1,n_2,n_3,n_4}\varphi_{n_1,n_2,n_3,n_4} &= \varphi,\\ \pi\text{-}\lim_{n_1,n_2,n_3,n_4}\mathrm{D}_x\varphi_{n_1,n_2,n_3,n_4} &= \mathrm{D}_x\varphi.\end{aligned} \tag{2.63}$$

PROOF. Let $\varphi \in C_m(\mathbb{R}_+ \times H)$ be such that $\mathrm{D}_x\varphi \in C_m(\mathbb{R}_+ \times H; H)$. For any $n_1 \in \mathbb{N}$, define

$$\varphi_{n_1}(t,x) = \frac{n_1\varphi(t,x)}{n_1 + \|x\|^{2m}}.$$

Then $\lim_{n_1}\varphi_{n_1}(t,x) = \varphi(t,x)$ pointwise and $\sup_{n_1}\|\varphi_{n_1}\|_m \leq \|\varphi\|_m$. Furthermore, one can show that $\lim_{n_1}\mathrm{D}_x\varphi_{n_1}(t,x) = \varphi(t,x)$ pointwise with $\sup_{n_1}\| \mathrm{D}_x\varphi_{n_1}\|_m \leq (\|\varphi\|_m + \| \mathrm{D}_x\varphi\|_{0,m})$ and in particular it holds that

$$\pi\text{-}\lim_{n_1}\varphi_{n_1} = \varphi, \pi\text{-}\lim_{n_1}\mathrm{D}_x\varphi_{n_1} = \mathrm{D}_x\varphi.$$

[3]For example, for any $n \in \mathbb{N}$ let $\varphi_n = \psi_{n,m(n)}$ where $m(n)$ is such that $\|\psi_{n,m} - \psi_n\|_0 < \frac{1}{n}$ and $\| \mathrm{D}_x\psi_{n,m} - \mathrm{D}_x\psi_n\|_0 < \frac{1}{n}$.

Noting that $\varphi_{n_1} \in C_0(\mathbb{R}_+ \times H)$ with bounded derivative $\mathrm{D}_x\varphi_{n_1} \in C_0(\mathbb{R}_+ \times H; H)$, the claim follows from Lemma 2.33 by approximating φ_{n_1} with a suitable sequence $(\varphi_{n_1,n_2,n_3,n_4})_{n_1,n_2,n_3,n_4} \subset \mathcal{E}_A(\mathbb{R}_+ \times H)$ such that

$$\pi\text{-}\lim_{n_1,n_2,n_3,n_4} \varphi_{n_1,n_2,n_3,n_4} = \varphi_{n_1},$$
$$\pi\text{-}\lim_{n_1,n_2,n_3,n_4} \mathrm{D}_x\varphi_{n_1,n_2,n_3,n_4} = \mathrm{D}_x\varphi_{n_1}.$$

□

LEMMA 2.35 (Lemma 2.13 above). *For any $m \geq 1$, the space $\mathcal{E}_A(\mathbb{R}_+ \times H)$ is a subset of $\mathrm{dom}_m(L)$ with $L\varphi = L_0\varphi$, where*

$$(L_0\varphi)(t,x) = \partial_t\varphi(t,x) + \langle x, A^*\,\mathrm{D}_x\varphi(t,x)\rangle + \langle F(t,x), \mathrm{D}_x\varphi(t,x)\rangle + \frac{1}{2}\operatorname{tr}\left(Q\,\mathrm{D}_x^2\varphi(t,x)\right) \tag{2.64}$$

for any $\varphi \in \mathcal{E}_A(\mathbb{R}_+ \times H)$. Moreover, $\mathcal{E}_A(\mathbb{R}_+ \times H)$ is a π-core for $(L, \mathrm{dom}_m(L))$, i.e. for any $\varphi \in \mathrm{dom}_m(L)$ there exists a sequence $(\varphi_n)_n \subset \mathcal{E}_A(\mathbb{R}_+ \times H)$ such that

$$\pi\text{-}\lim_n \varphi_n = \varphi \text{ and } \pi\text{-}\lim_n L_0\varphi_n = L\varphi. \tag{2.65}$$

Furthermore, if $\varphi \in (L, \mathrm{dom}_m(L))$ is such that $\mathrm{D}_x\varphi \in C_m(\mathbb{R}_+ \times H; H)$, the approximating sequence $(\varphi_n)_n$ in (2.65) can be chosen such that

$$\pi\text{-}\lim_n \mathrm{D}_x\varphi_n = \mathrm{D}_x\varphi. \tag{2.66}$$

PROOF. Using Lemma 2.34, the proof goes just as in Manca [82], Theorem 1.3 after noting that $Y(t,(s,x))$ is the mild solution to the noise-degenerate SPDE

$$\begin{cases} \mathrm{d}Y(t) &= \left[\tilde{A}Y(t) + \tilde{F}(Y(t))\right]\mathrm{d}t + \sqrt{\tilde{Q}}\,\mathrm{d}\tilde{W}(t), \quad t \geq 0, \\ Y(0) &= (s,x). \end{cases} \tag{2.67}$$

Here $\tilde{A}(s,x) = (0, Ax)$ and $\tilde{F}(s,x) = (1, F(s,x))$ and $\tilde{Q} = (0, Q)$ are defined on the product space $\mathbb{R} \times H$ and it is straightforward to show that $\tilde{A}, \tilde{F}$ and $\tilde{Q}$ satisfy Assumption 2.5. □

2.C

LEMMA 2.36. *Let $G(s,x)$ be a Lipschitz continuous function in x, uniformly on $[0,T]$, i.e. there exists some constant $C > 0$ such that*

$$\|G(s,x) - G(s,y)\| \leq C\|x - y\|$$

for all $s \in [0,T]$ and $x, y \in H$. Let X be the mild solution to (2.1). Then the process $E(t)$ defined by

$$E(t) = \exp\left(\int_0^t \langle G(s,X(s)), \mathrm{d}W(s)\rangle - \frac{1}{2}\int_0^t \|G(s,X(s))\|^2\,\mathrm{d}s\right), \quad t \in [0,T],$$

is a $\mathbb{P}$-martingale.

PROOF. Since $(E(t))_{t\in[0,T]}$ is a supermartingale, it suffices to show that $\mathbb{E}[E(T)] = 1$. From the almost sure continuity of X and the Lipschitz continuity of G it follows that

$$\int_0^T \|G(s,X(s))\|^2\,\mathrm{d}s < \infty \quad \mathbb{P}\text{-a.s.}$$

Thus, defining the stopping times

$$\tau_n(X) = \inf\left\{t \in [0,T] : \int_0^t \|G(s,X(s))\|^2 \, \mathrm{d}s \geq n\right\} \wedge T$$

we have that $\mathbb{P}\left(\lim_n \tau_n = T\right) = 1$. Since

$$\int_0^{T\wedge\tau_n} \|G(s,X(s))\|^2 \, \mathrm{d}s < n \quad \mathbb{P}\text{-a.s.}$$

it follows from the Novikov condition (see e.g. Da Prato and Zabczyk [41], Proposition 10.17) that $E_n(t) = E(t \wedge \tau_n), t \in [0,T]$, is a $\mathbb{P}$-martingale for each $n \in \mathbb{N}$. In particular, E_n defines a measure $\mathbb{P}_n$ on $\mathcal{F}_T$ such that $\mathrm{d}\mathbb{P}_n = E_n(T)\,\mathrm{d}\mathbb{P}$. Define $\chi_n(s) = \mathbb{1}_{s\leq\tau_n}$. Noting that

$$\begin{aligned} E_n(t) &= \exp\left(\int_0^{t\wedge\tau_n} \langle G(s,X(s)), \mathrm{d}W(s)\rangle - \frac{1}{2}\int_0^{t\wedge\tau_n} \|G(s,X(s))\|^2 \, \mathrm{d}s\right) \\ &= \exp\left(\int_0^t \langle \chi_n(s)G(s,X(s)), \mathrm{d}W(s)\rangle - \frac{1}{2}\int_0^t \|\chi_n(s)G(s,X(s))\|^2 \, \mathrm{d}s\right), \end{aligned}$$

it follows from the Girsanov theorem that

$$W_n(t) = W(t) - \int_0^t \chi_n(s)G(s,X(s)) \, \mathrm{d}s, \quad t \in [0,T],$$

is a $\mathbb{P}_n$-cylindrical Wiener process. It follows that for any $n \in \mathbb{N}$, X under $\mathbb{P}_n$ is a mild solution to the equation

$$\mathrm{d}X_n = \left[AX_n + F(t,X_n(t)) + Q^{\frac{1}{2}}\chi_n(t)G(t,X_n(t))\right]\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_n(t).$$

In particular, by the Lipschitz continuity of G, there exists a $\mathbb{P}_{n_0}$-a.s. continuous version of X for any $n_0 \in \mathbb{N}$ fixed, from which we conclude that

$$\int_0^T \|G(s,X(s))\|^2 \, \mathrm{d}s < \infty \quad \mathbb{P}_{n_0}\text{-a.s.}$$

It follows, using the monotonicity of τ_n in n in the second line and monotone convergence in the last step, that

$$\begin{aligned} 1 = \lim_n \mathbb{P}_{n_0}(\tau_n = T) &= \lim_n \int_{\{\tau_n=T\}} E(T\wedge\tau_{n_0}) \, \mathrm{d}\mathbb{P} \\ &= \lim_{n\geq n_0} \int_{\{\tau_n=T\}} E(T) \, \mathrm{d}\mathbb{P} \\ &= \mathbb{E}[E(T)]. \end{aligned}$$

□

2.D

Lemma 2.37. *Under Assumption 2.21, for any $S < T$, the operators Γ_{T-t} are uniformly Hilbert-Schmidt on $[0,S]$, i.e.*

$$\sup_{t\in[0,S]} \|\Gamma_{T-t}\|_{HS} < \infty.$$

Proof. From the Strong Feller assumption 2.21 it follows that $\Gamma_r = Q_r^{-\frac{1}{2}} S_r$ is a bounded linear operator and thus

$$S_r = Q_r^{\frac{1}{2}} \Gamma_r$$

is a Hilbert-Schmidt operator for any $r > 0$. Moreover, it follows from Assumption 2.21 that

$$\mathrm{im}(Q_\infty^{\frac{1}{2}}) = \mathrm{im}(Q_r^{\frac{1}{2}}), \quad r > 0,$$

see Proposition 2 in Chojnowska-Michalik and Goldys [24]. From this one concludes that $Q_r^{-\frac{1}{2}} Q_\infty^{\frac{1}{2}}$ and $Q_\infty^{-\frac{1}{2}} S_r$ are bounded linear operators for all $r > 0$. Now, fix some arbitrary $S < T$. Then for any $t \in [0, S]$ it holds that

$$\begin{aligned}\Gamma_{T-t} &= Q_{T-t}^{-\frac{1}{2}} S_{T-t} \\ &= (Q_{T-t}^{-\frac{1}{2}} Q_\infty^{\frac{1}{2}})(Q_\infty^{-\frac{1}{2}} S_{T-t}) \\ &= (Q_{T-t}^{-\frac{1}{2}} Q_\infty^{\frac{1}{2}})(Q_\infty^{-\frac{1}{2}} S_{T+S-t}) S_{T-S}.\end{aligned}$$

Then, noting that $(Q_{T-t}^{-\frac{1}{2}} Q_\infty^{\frac{1}{2}})$ and $(Q_\infty^{-\frac{1}{2}} S_{T+S-t})$ are strongly continuous in t, it follows from the uniform boundedness principle that

$$\sup_{t\in[0,S]} \|\Gamma_{T-t}\|_{HS} \leq \sup_{t\in[0,S]} \|(Q_{T-t}^{-\frac{1}{2}} Q_\infty^{\frac{1}{2}})(Q_\infty^{-\frac{1}{2}} S_{T+S-t})\| \|S_{T-S}\|_{HS} < \infty.$$

□

Lemma 2.38. *For any $S < T$, the random process*

$$\Gamma_{T-t}^* Q_{T-t}^{-\frac{1}{2}} y = \sum_{j=1}^{\infty} q_{j,T-t}^{-\frac{1}{2}} \langle y, e_j \rangle \Gamma_{T-t}^* e_j, \quad t \in [0, S], \tag{2.68}$$

is well-defined as a limit in $L^2(H, \nu; C([0, S]; H))$. Moreover, there exists a measurable space H_S with $\nu(H_S) = 1$ such that the limit exists pointwise for ν-a.e. $y \in H_S$.

Proof. For any $n \in \mathbb{N}$ define the process

$$\Upsilon_n(t) = \sum_{j=1}^{n} q_{j,T-t}^{-\frac{1}{2}} \langle y, e_j \rangle \Gamma_{T-t}^* e_j$$

where $(q_{j,T-t}, e_j)_j$ is the eigenbasis of Q_{T-t}. From the strong continuity of $(Q_t)_t$ and $(\Gamma_t)_t$ it follows that $\Upsilon_n \in C([0, S]; H)$ for any $n \in \mathbb{N}$. Furthermore, it holds that

$$\begin{aligned}\int_H \|\Upsilon_n\|_0^2 \, \nu(\mathrm{d}y) &= \int_H \sup_{t\in[0,S]} \|\sum_{j=1}^{n} q_{j,T-t}^{-\frac{1}{2}} \langle y, e_j \rangle \Gamma_{T-t}^* e_j\|^2 \, \nu(\mathrm{d}y) \\ &= \sup_{t\in[0,S]} \sum_{j=1}^{n} q_{j,T-t}^{-1} \left(\int_H |\langle y, e_j \rangle|^2 \, \nu(\mathrm{d}y) \right) \|\Gamma_{T-t}^* e_j\|^2 \\ &= \sup_{t\in[0,S]} \sum_{j=1}^{n} q_{j,T-t}^{-1} \langle Q_\infty e_j, e_j \rangle \, \|\Gamma_{T-t}^* e_j\|^2 \\ &= \sup_{t\in[0,S]} \sum_{j=1}^{n} \langle Q_{T-t}^{-\frac{1}{2}} Q_\infty^{\frac{1}{2}} e_j, Q_{T-t}^{-\frac{1}{2}} Q_\infty^{\frac{1}{2}} e_j \rangle \langle \Gamma_{T-t}^* e_j, \Gamma_{T-t}^* e_j \rangle \\ &\leq \sup_{t\in[0,S]} \left(\|Q_{T-t}^{-\frac{1}{2}} Q_\infty^{\frac{1}{2}}\| \, \|\Gamma_{T-t}^*\|_{HS}^2 \right) < \infty\end{aligned}$$

from which we conclude the convergence of $\Upsilon_n \to \Gamma_{T-t}^* Q_{T-t}^{-\frac{1}{2}} y$ in $L^2(H, \nu; C([0, S]; H))$. Now, the second claim follows by an application of the Itô–Nisio theorem. □

Corollary 2.39. *There exists a measurable space $H_0 \subset H$ with $\nu(H_0) = 1$ such that*

$$[0, T) \to H, \; t \mapsto \Gamma_{T-t}^* Q_{T-t}^{-\frac{1}{2}} y \tag{2.69}$$

is well-defined and continuous for any $y \in H_0$.

Proof. Set $H_0 = \bigcap_n H_{T-1/n}$ where $H_{T-1/n}$ is the measurable space given by Lemma 2.38. □

Chapter 3

Simulation of infinite-dimensional diffusion bridges

Let X be the mild solution to a semilinear stochastic partial differential equation. In this chapter, we develop methodology to sample from the infinite-dimensional diffusion bridge that arises from conditioning X on a linear transformation LX_T of the final state X_T at some time $T > 0$. This solves a problem that has so far not been attended to in the literature. Our main contribution is the derivation of a path measure that is absolutely continuous with respect to the path measure of the infinite-dimensional diffusion bridge. This lifts previously known results for stochastic ordinary differential equations to the setting of infinite-dimensional diffusions and stochastic partial differential equations. We demonstrate our findings through numerical experiments on stochastic reaction-diffusion equations.

3.1 Introduction

Consider a semilinear *stochastic partial differential equation (SPDE)* of the form

$$\begin{cases} \mathrm{d}X_t &= [AX_t + F(t, X_t)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t, \quad t \geq 0, \\ X_0 &= x_0. \end{cases} \tag{3.1}$$

The operator A denotes the generator of a strongly continuous semigroup $(S_t)_{t\geq 0}$ on a Hilbert space H, whereas F denotes a nonlinear operator and Q is a symmetric, positive operator on H. The process W is a cylindrical Wiener process on H, defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. We assume that the operators A, F and Q satisfy suitable conditions such that Equation (3.1) admits a unique mild solution $X = (X_t)_{t\geq 0}$ for any $x_0 \in H$.

For some $k \geq 1$, let $L : H \to \mathbb{R}^k$ denote a bounded, linear operator, termed the observation operator. Suppose we observe the linear transformation LX_T of X_T at some time $T > 0$. This includes the highly relevant cases that we observe an orthogonal projection of X_T onto a finite-dimensional subspace or that, in the case of $H = L^2(D)$ for some $D \subset \mathbb{R}^d$, we observe weighted integrals of X_T.

We are interested in sampling the process X on the interval $[0, T]$, conditioned on the event $\{LX_T = y\}$ for some fixed $y \in \mathbb{R}^k$. This problems arises, for example, in parameter estimation for partially observed SPDEs. Following the literature for diffusion processes with finite-dimensional state spaces, we refer to the conditioned process as an *infinite-dimensional diffusion bridge* and denote it by $X^\star$. To the best of our knowledge, the problem of drawing samples of $X^\star$ has not been considered before in the generality of our setup.

This stands in contrast to the corresponding problem for a stochastic *ordinary* differential equation, which has attracted considerable attention in the past two decades (see for instance Delyon and Hu [44], Beskos et al. [14], Papaspiliopoulos and Roberts [90], Schauer et al. [101], Bierkens et al. [16], Heng et al. [63]). By Doob's h-transform (Rogers and Williams [98], Section IV.39), it can be shown that a *diffusion bridge* $x^\star$ associated with the strong solution x to the SODE

$$\mathrm{d}x_t = f(x_t)\,\mathrm{d}t + \sigma\,\mathrm{d}w_t$$

is characterised by yet another SODE given by

$$\mathrm{d}x_t^\star = [f(x_t^\star) + \sigma\sigma^*\,\mathrm{D}_x \log h(t, x_t^\star)]\,\mathrm{d}t + \sigma\,\mathrm{d}w_t.$$

Here, h is a mapping that depends on both the transition densities p and the matrix representation of the map L. In particular, the conditioned process $x^\star$ solves an SODE itself, where the additional term $\sigma\sigma^*\,\mathrm{D}_x \log h(t, x_t^\star)$ is added to the drift coefficient to steer the process towards the point of conditioning.

Unfortunately, the transition density p, and consequently the function h, is not known in closed form except for a few special cases. This has motivated several works in which the map h is replaced by a substitute g in the steering term (Clark [32], Delyon and Hu [44], Schauer et al. [101], Mider et al. [85]). If g is chosen carefully, then indeed the laws of the conditioned process and the process that substitutes g for h are absolutely continuous on path space. In particular, *weighted samples* of the conditioned process can be obtained. This has important applications for example in Bayesian parameter estimation for discretely observed diffusion processes (see for instance Roberts and Stramer [97] and Papaspiliopoulos et al. [91]).

The literature on the corresponding problem for infinite-dimensional diffusions is scarce. Notable exceptions are Di Nunno et al. [45] and Yang et al. [115]. The former considers the situation where $F \equiv 0$, in which case the diffusion bridge is given by a conditioned Gaussian process on H. Moreover, convergence rates are derived for the spatial discretisation of $X^\star$. In Yang et al. [115], the process X is conditioned on a set of positive measure and projected onto a finite-dimensional subspace of H after which the problem is considered in finite dimensions. Additionally, Equation (3.1) is assumed to admit a strong solution, an assumption that is generally not satisfied when working with SPDEs.

3.1.1 Contribution

In this chapter, we lift earlier results obtained for SODEs in Schauer et al. [101] and Bierkens et al. [16] to SPDEs, resulting in computational methods for sampling infinite-dimensional diffusion bridges. This extension is nontrivial as we need to deal with complications arising from working in infinite-dimensional state spaces. We strongly depend on our earlier work in Pieper-Sethmacher et al. [94] on exponential change of measure for SPDEs. Here, it was shown that the infinite-dimensional diffusion bridge $X^\star$ can be obtained by an *exponential change of measure* $\mathbb{P}^\star$ on the underlying filtered probability space. Under additional assumptions, it was shown that $X^\star$ is equivalent in law to the mild solution of the *bridge SPDE*

$$\mathrm{d}X^\star(t) = [AX^\star(t) + F(t, X^\star(t)) + Q\,\mathrm{D}_x \log h(t, X^\star(t))]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W(t), \quad t < T,$$

where $h(t,z) := \rho_X(t, z; T, y)$ denotes the density of $LX_T \mid X_t = z$, evaluated at y. As in the finite-dimensional case, ρ_X is in general intractable, rendering a direct simulation of $X^\star$ infeasible. As a further complication, the formulation of the bridge SPDE relies on the existence of the gradient $\mathrm{D}_x \log h$ in the first place. However, to the best of our knowledge, no such regularity result for transition densities of SPDEs of the form (3.1) has been obtained in the literature so far. On the other hand, the *Ornstein-Uhlenbeck* process Z, which satisfies the linearised version of (3.1) with $F \equiv 0$, is a Gaussian process for which the the density of $LZ_T \mid Z_t = z$ is tractable. Moreover, for this density, which we denote by ρ_Z, the required regularity assumptions can be validated. This motivates substituting ρ_X by ρ_Z, by which we obtain the *guided process* X° as the mild solution to the SPDE

$$\mathrm{d}X^\circ(t) = [AX^\circ(t) + F(t, X^\circ(t)) + Q\,\mathrm{D}_x \log g(t, X^\circ(t))]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W(t),$$

where $g(t,x) = \rho_Z(t,x;T,y)$. In Pieper-Sethmacher et al. [94], it was shown that X° equals in law the process X under yet another exponential change of measure $\mathbb{P}^\circ$. Moreover, it was shown that the laws $\mathcal{L}^\star$ and $\mathcal{L}^\circ$ of $X^\star$ and X° are absolutely continuous on the path space $C([0,t];H)$ for any $t < T$. Our main result in this work, presented in Theorem 3.11, shows that the absolute continuity persists in the limit $t \uparrow T$. This facilitates $\mathcal{L}^\circ$ as a suitable proposal distribution for $\mathcal{L}^\star$, hence enabling a wide range of computational methods to sample from infinite-dimensional diffusion bridges. These include, for example, Markov chain Monte Carlo algorithms, importance sampling or variational methods. Algorithm 1 shows how Theorem 3.11 can be used to define a Metropolis-Hastings algorithm for sampling diffusion bridges for SPDEs. We show numerical performance in stochastic reaction diffusion equations, including the Allen-Cahn equation. A crucial step towards the main result concerns the convergence of the guided process $X^\circ(t)$ towards the conditioning point y at an appropriate rate (Cf. Theorem 3.9). While the main steps in the proof are similar to the proof of absolute continuity in Bierkens et al. [16], complications arise from dealing with the infinite-dimensional setting. As it is unclear under which conditions transition densities exist, contrary to Bierkens et al. [16], the proof of Theorem 3.11 avoids their existence. Since we restrict Equation (3.1) to the case of additive noise, we also obtain much "cleaner" conditions for absolute continuity to hold. Our examples show that Assumption 3.13 can easily be verified in the diagonalisable case (Cf. Section 3.5). The other assumption, Assumption 3.10, is shown to hold under the easily verified conditions of Lemma 3.13.

3.1.2 Outline

The necessary preliminaries are outlined in Section 3.2. This contains a brief description on the derivation of the infinite-dimensional diffusion bridge and the guided process. In Section 3.3 we state the main results of the chapter and in Section 3.4 we present a Metropolis-Hastings algorithm for sampling infinite-dimensional diffusion bridges. We showcase the performance of this algorithm in numerical experiments in Section 3.5. The proofs of our main results, Theorem 3.9 and Theorem 3.11, are given in Section 3.6 and Section 3.7 respectively.

3.1.3 Frequently used notation

For the readers convenience we summarise here the general notation as well as processes and operators that are frequently used throughout this work.

General We denote by H a Hilbert space equipped with inner product $\langle x, y\rangle$ and norm $|x| = \sqrt{\langle x, x\rangle}$. For any linear, bounded operator $B : H \to H$ we let $\|B\|$ be the operator norm of B. If B is a trace-class operator, we denote by $\mathrm{tr}[B]$ its trace and if B is a Hilbert-Schmidt operator we write $\|B\|_{\mathrm{HS}}$ for its Hilbert-Schmidt norm. Moreover, B^* denotes the adjoint of B, and if B is positive, $B^{\frac{1}{2}}$ denotes the square root operator of B.

The space $C([0,T];H)$ denotes the space of continuous, H-valued function, endowed with the supremum norm $\|\varphi\| = \sup_{t\in[0,T]} |\varphi(t)|$. Without further mention, we assume that all normed spaces are equipped with their corresponding Borel σ-Algebra.

For any arbitrary set $\mathcal{X}$ and two functions $f, g : \mathcal{X} \to \mathbb{R}$, we write $f(x) \lesssim g(x)$ if there exists some constant $C > 0$, independent of x, such that $f(x) \le Cg(x)$ for all x. If $\mathcal{X}$

equals $\mathbb{R}$ or $\mathbb{N}$ and f and g are asymptotically equivalent, we write $f \sim g$. The tuple $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ denotes a stochastic basis and we let $\mathbb{P}_t$ be the restriction of $\mathbb{P}$ onto $\mathcal{F}_t$.

Operators and processes For any $k \geq 1$, the operator $L : H \to \mathbb{R}^k$ is a bounded, linear operator that we refer to as the observation operator. The operator $(A, \mathrm{dom}(A))$ is the generator of a strongly continuous semigroup $(S_t)_{t\geq 0}$ on H. Moreover, Q is a positive symmetric operator and F a continuous nonlinearity, both acting on H as well. The process X denotes the mild solution to the SPDE (3.1), whereas Z is the Ornstein-Uhlenbeck process, i.e. the mild solution to (3.1) for the case that $F \equiv 0$. By Q_t we denote the covariance operator of Z_t, whereas $R_t = LQ_tL^*$ and $L_t = LS_t$. The process $X^\star$ is the conditioned process X given $LX_T = y$ and X° the guided process, respectively derived from X under the changes of measure $\mathbb{P}^\star$ and $\mathbb{P}^\circ$ on $\mathcal{F}_T$.
For the process $(X_t)_{t\in[0,T]}$ we denote by $\mathcal{L}(X)$ the law of X on the path space $C([0,T]; H)$ and by $\mathcal{L}_t(X)$ the restriction of its law on $C([0,t]; H)$, whereas $\mathcal{L}(X_t)$ is the law of X_t on H. Finally, for the measures $\mathcal{L}(X^\star)$ and $\mathcal{L}(X^\circ)$ we shorten notation and simply write $\mathcal{L}^\star$ and $\mathcal{L}^\circ$ instead.

3.2 Preliminaries

3.2.1 Mild solutions

The following is a standing assumption on the components involved in Equation (3.1).
ASSUMPTION 3.1.

(i) *A is the generator of a strongly continuous semigroup $(S_t)_{t\geq 0}$ on H. In particular there exists a $C_S > 0, \omega_S \in \mathbb{R}$ such that $\|S_t\| \leq C_S \exp(\omega_S t)$ for all $t \geq 0$.*

(ii) *W is a cylindrical Wiener process on H.*

(iii) *Q is a symmetric, positive and bounded operator on H. Moreover, the operators $(Q_t)_{t\geq 0}$ given by*

$$Q_t = \int_0^t S_u Q S_u^* \,\mathrm{d}u \tag{3.2}$$

are well-defined and trace-class.

(iv) *F is such that there exists a constant $C_F > 0$ with*

$$|F(t,x) - F(t,y)| \leq C_F|x-y| \text{ and } |F(t,x)| \leq C_F$$

for all $t \in [0,T]$ and $x, y \in H$.

Under Assumption 3.1, Equation (3.1) admits, for any initial value $x_0 \in H$, a unique mild solution X that satisfies

$$X_t = S_t x_0 + \int_0^t S_{t-s}F(s, X_s)\,\mathrm{d}s + \int_0^t S_{t-s}Q^{\frac{1}{2}}\,\mathrm{d}W_s, \quad t \geq 0. \tag{3.3}$$

The process X is Markovian and has an almost surely continuous modification. In the special case that $F = 0$ we denote the time homogeneous mild solution to equation (3.1) by Z and refer to it as the *Ornstein-Uhlenbeck (OU) process*. The Ornstein-Uhlenbeck process is a Gaussian process and, in particular, for any $s < t$ and $x \in H$, the distribution of $Z_t \mid Z_s = x$ is Gaussian with mean $S_{t-s}x$ and covariance operator Q_{t-s}. Moreover, this implies that the distribution of $LZ_t \mid Z_s = x$ is Gaussian in $\mathbb{R}^k$ with mean $L_{t-s}x$ and covariance matrix R_{t-s}, with operators L_t and R_t defined by

$$\begin{aligned} L_t &= LS_t, \\ R_t &= LQ_tL^*, \quad t \in [0,T]. \end{aligned} \tag{3.4}$$

3.2.2 Derivation of the diffusion bridge and guided process

For the reader's convenience, we give an outline of the derivation of the infinite-dimensional diffusion bridge $X^\star$ and the guided process X° based on an exponential change of measure as introduced in Pieper-Sethmacher et al. [94]. The details of this subsection's proofs can be found in Appendix 3.A.

For this, let $(T_t)_{t\geq 0}$ denote the transition semigroup

$$(T_t\varphi)(s,x) = \mathbb{E}[\varphi(s+t, X_{t+s}) \mid X_s = x]$$

defined on the Banach space $C_m(\mathbb{R}_+ \times H)$ of continuous functions $\varphi : \mathbb{R}_+ \times H \to \mathbb{R}$ such that $\|\varphi\|_m = \sup_{t,x}(1+\|x\|^m)^{-1}|\varphi(t,x)| < \infty$. A sequence $(\varphi_n)_n \subset C_m(\mathbb{R}_+ \times H)$ is said to converge to $\varphi \in C_m(\mathbb{R}_+ \times H)$ in the topology of *bounded-pointwise* convergence if $\lim_n \varphi_n(t,x) = \varphi(t,x)$ for any $t \geq 0, x \in H$ and $\sup_n \|\varphi_n\|_m < \infty$. We then write $\pi\text{-}\lim_n \varphi_n = \varphi$.

In the topology of bounded-pointwise convergence, define the infinitesimal generator $(K, \mathrm{dom}_m(K))$ of $(T_t)_t$ via

$$\begin{cases} \mathrm{dom}_m(K) & = \left\{ \varphi \in C_m(\mathbb{R}_+ \times H) : \exists \psi \in C_m(\mathbb{R}_+ \times H) \text{ s.t. } \pi\text{-}\lim_{t\downarrow 0} \dfrac{T_t\varphi - \varphi}{t} = \psi \right\} \\ (K\varphi)(s,x) & = \lim_{t\downarrow 0} \dfrac{(T_t\varphi)(s,x) - \varphi(s,x)}{t}, \quad \varphi \in \mathrm{dom}_m(K), (s,x) \in \mathbb{R}_+ \times H. \end{cases} \tag{3.5}$$

For any positive function $h \in \mathrm{dom}_m(K)$, define the process

$$E_t^h = \frac{h(t,X_t)}{h(0,x_0)} \exp\left(-\int_0^t \frac{Kh}{h}(s,X_s)\,\mathrm{d}s \right). \tag{3.6}$$

It can be shown that E^h is a continuous local $\mathbb{P}$-martingale whenever it exists, see for example Lemma 3.1 in Palmowski and Rolski [89]. If E^h is a true $\mathbb{P}$-martingale, it defines an *exponential change of measure* $\mathbb{P}^h$ on $\mathcal{F}_T$. Moreover, under additional assumptions on h, the change of measure $\mathbb{P}^h$ is of Girsanov-type, as was shown in the following theorem.

THEOREM 3.2 (Theorem 3.5, Pieper-Sethmacher et al. [94]). *Let $h : [0,T) \times H \longrightarrow \mathbb{R}_{>0}$ satisfy the following for any $S < T$:*

- (i) *$h \in C_m([0,S] \times H)$ such that $Kh(t,x)$ exists for any $[0,S] \times H$ and $h^{-1}Kh \in C_m([0,S] \times H)$.*
- (ii) *h is such that the process $(E_t^h)_{t\in[0,S]}$ is a $\mathbb{P}$-martingale.*

Then h defines a unique change of measure $\mathbb{P}^h$ on $(\Omega, \mathcal{F}_T)$ via

$$\mathrm{d}\mathbb{P}_t^h = \mathbb{E}_t^h\,\mathrm{d}\mathbb{P}_t, \quad t < T.$$

Additionally, let h satisfy the following:

- (iii) *h is Fréchet differentiable in x such that $\mathrm{D}_x h \in C_m([0,S] \times H; H)$ for any $S < T$.*

Then X under $\mathbb{P}^h$ is a mild solution to the SPDE

$$\mathrm{d}X_t = [AX_t + F(t,X_t) + Q\,\mathrm{D}_x \log h(t,X_t)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t^h, \quad t \in [0,T),$$

where W^h is a $\mathbb{P}^h$-cylindrical Wiener process.

The diffusion bridge $X^\star$ and corresponding guided process X° are defined through an application of Theorem 3.2 for different choices of the function h as follows. Let $y \in \mathbb{R}^k$ and $T > 0$ be fixed. For any $x \in H$ and $t \in [0,T)$, denote by $\rho_Z(t,x;T,y)$ the density of $LZ_T \mid Z_t = x$ with respect to the Lebesgue measure on $\mathbb{R}^k$, evaluated at y. As stated in Section 3.2.1, $\rho_Z(t,x;T,y)$ is Gaussian in y with mean $L_{T-t}x$ and covariance matrix R_{T-t} defined in (3.4).

Moreover, let $\rho_X(t,x;T,y)$ denote the density of $LX_T \mid X_t = x$, evaluated at y. The existence of $\rho_X(t,x;T,y)$ is a consequence of the Girsanov theorem, from which the absolute continuity of $\mathcal{L}(X_T)$ with respect to $\mathcal{L}(Z_T)$, and hence of $\mathcal{L}(LX_T)$ with respect to $\mathcal{L}(LZ_T)$ follows.
On $[0,T) \times H$, define the functions h and g by

$$\begin{aligned} h(t,x) &= \rho_X(t,x;T,y), \\ g(t,x) &= \rho_Z(t,x;T,y). \end{aligned} \tag{3.7}$$

Remark 3.3. The function $h(t,x) = \rho_X(t,x;T,y)$ is only defined $\mathcal{L}(LX_T)$-almost surely in y, i.e. there exists a Borel set $\mathcal{A}_0$ such that $\mathbb{P}(LX_T \in \mathcal{A}_0) = 1$ and $(t,x,y) \mapsto \rho_X(t,x;T,y)$ is uniquely well-defined for all $t \in [0,T), x \in H$ and $y \in \mathcal{A}_0$. The existence of such a set, independently of (t,x), follows from the Girsanov theorem and the continuity of $(t,x) \mapsto LZ_T \mid Z_t = x$. Throughout the remainder of this work, we assume without further notice that $y \in \mathcal{A}_0$.

The following two propositions verify that h and g satisfy the assumptions of Theorem 3.2, and consequently, prove the existence of $X^\star$ and X°.

Proposition 3.4 (Existence of the diffusion bridge). *The mapping h defined in (3.7) satisfies Assumptions (i) and (ii) of Theorem 3.2 with $Kh = 0$. Moreover, the measure $\mathbb{P}^\star$ defined on $\mathcal{F}_T$ by*

$$\mathrm{d}\mathbb{P}^\star_t = \frac{h(t,X_t)}{h(0,x_0)} \,\mathrm{d}\mathbb{P}_t, \quad t < T,$$

is such that, for any bounded and measurable function φ and $0 \le t_1 \le ... \le t_n < T$, it holds

$$\mathbb{E}^\star[\varphi(X_{t_1},...,X_{t_n})] = \mathbb{E}[\varphi(X_{t_1},...,X_{t_n}) \mid LX_T = y], \quad \mathcal{L}(LX_T) - a.s.$$

We call the process X under $\mathbb{P}^\star$ the infinite-dimensional diffusion bridge (of X given $LX_T = y$).

Remark 3.5. If, in addition to Assumption (i) and (ii) of Theorem 3.2, h satisfies Assumption (iii), X under $\mathbb{P}^\star$ is a mild solution to the diffusion bridge equation

$$\mathrm{d}X^\star_t = [AX^\star_t + F(t,X^\star_t) + Q\,\mathrm{D}_x \log h(t,X^\star_t)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W^\star_t, \quad t \in [0,T), \tag{3.8}$$

for some $\mathbb{P}^\star$-cylindrical Wiener process $W^\star$. This generalises the diffusion bridge equation known for conditioned diffusions in Euclidean spaces.
However, we do not impose this assumption on h for two reasons. Firstly, verifying the Fréchet differentiablity of h is a difficult problem in this infinite-dimensional setting. If the state space is Euclidean, this assumption can be verified based on regularity results for transition densities of diffusion processes. To the best of our knowledge, no such results are known for SPDEs.
Secondly, except for a few special cases, the term $\mathrm{D}_x \log h$ is intractable. This renders a direct simulation of the bridge process infeasible, even if it satisfies Equation (3.8). This motivates the construction of the guided process as a tractable substitute process in the next proposition.

Proposition 3.6 (Existence of the guided process). *The mapping g defined in (3.7) satisfies Assumptions (i) to (iii) of Theorem 3.2 with $Kg(t,x) = \langle F(t,x), \mathrm{D}_x g(t,x)\rangle$. In particular, there exists a unique measure $\mathbb{P}^\circ$ on $\mathcal{F}_T$, defined by*

$$\mathrm{d}\mathbb{P}^\circ_t = \frac{g(t,X_t)}{g(0,x_0)} \exp\left(- \int_0^t \langle F(s,X_s), \mathrm{D}_x \log g(s,X_s)\rangle \,\mathrm{d}s \right) \mathrm{d}\mathbb{P}_t, \quad t < T,$$

such that X under $\mathbb{P}^\circ$ is the unique mild solution to the SPDE

$$\mathrm{d}X_t^\circ = [AX_t^\circ + F(t, X_t^\circ) + Q\,\mathrm{D}_x \log g(t, X_t^\circ)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t^\circ, \quad t \in [0, T), \tag{3.9}$$

where $(W_t^\circ)_{t\in[0,T)}$ is a $\mathbb{P}^\circ$-cylindrical Wiener process. The process X under $\mathbb{P}^\circ$ is referred to as the guided process.

The proofs of Proposition 3.4 and Proposition 3.6 can be found in Appendix 3.A.

Remark 3.7. To highlight the fact that the process X under $\mathbb{P}^\circ$ solves the SPDE (3.9), we write from now on, with slight abuse of notation, X° instead of X whenever we assume it to be defined on the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P}^\circ)$. Likewise, we write $X^\star$ for the process X defined under the measure $\mathbb{P}^\star$. Furthermore, we denote by $\mathcal{L}^\star$ and $\mathcal{L}^\circ$ the law of X on $C([0,T];H)$ under $\mathbb{P}^\star$ and $\mathbb{P}^\circ$ respectively and by $\mathcal{L}_t^\star$ and $\mathcal{L}_t^\circ$ their restrictions onto $C([0,t];H)$ for any $t < T$.

Contrary to the function h, the function g is available in closed form with

$$g(t,x) = \rho_Z(t,x;T,y) = \frac{1}{\sqrt{(2\pi)^k \det(R_{T-t})}} \exp\left(-\frac{1}{2}|R_{T-t}^{-\frac{1}{2}}(y - L_{T-t}x)|^2\right).$$

Hence, the additional drift term $G(t,x) = \mathrm{D}_x \log g(t,x)$ appearing in Equation (3.9) is given by

$$G(t,x) = L_{T-t}^* R_{T-t}^{-1}(y - L_{T-t}x). \tag{3.10}$$

In particular, this enables us to draw samples directly from X° by forward simulating the SPDE (3.9).

3.3 Main results

Throughout this section we fix some $T > 0$ and $y \in \mathbb{R}^k$ and let $X^\star$ be the infinite-dimensional diffusion bridge and X° be the guided process, respectively defined as the process X under the changes of measure $\mathbb{P}^\star$ and $\mathbb{P}^\circ$ as introduced in Propositions 3.4 and 3.6.

By construction, the measures $\mathbb{P}^\star$ and $\mathbb{P}^\circ$ are absolutely continuous on $\mathcal{F}_t, t < T$, with

$$\frac{\mathrm{d}\mathbb{P}_t^\star}{\mathrm{d}\mathbb{P}_t^\circ}(X^\circ) = \frac{h(t,X_t^\circ)}{g(t,X_t^\circ)}\frac{g(0,x_0)}{h(0,x_0)}\Psi_t(X^\circ) \quad \mathbb{P}^\circ\text{-a.s.} \tag{3.11}$$

where the process $\Psi_t(X^\circ)$ is given by

$$\Psi_t(X^\circ) = \exp\left(\int_0^t \Big\langle F(s, X_s^\circ), G(s, X_s^\circ)\Big\rangle\,\mathrm{d}s\right). \tag{3.12}$$

Our main results come in two parts. Firstly, we show that the guided process $X^\circ(t)$ converges to the conditioning set $\{x \in H : Lx = y\}$. Besides the convergence in itself, it is crucial that this convergence occurs at an appropriate rate.

Secondly, we show that the absolute continuity in (3.11) persists as we take the limit $t \uparrow T$, i.e. that the measures $\mathbb{P}^\star$ and $\mathbb{P}^\circ$ are equivalent on $\mathcal{F}_T$. As a consequence thereof, the laws $\mathcal{L}^\star$ and $\mathcal{L}^\circ$ are equivalent on the path space $C([0,T];H)$. This is critical for the guided process to serve as a valid proposal distribution for the infinite-dimensional diffusion bridge.

3.3.1 Limit behavior of the Guided Process towards the conditioning time

We first study the limit behavior of $X^\circ(t)$ as $t \uparrow T$. Let the following hold.

Assumption 3.8. *There exist positive constants* $\underline{c}, \bar{c}$ *such that for all* $t \in (0, T]$

$$\underline{c}\, t^{-1} \leq \|R_t^{-1}\| \leq \bar{c}\, t^{-1}. \tag{3.13}$$

We get the following convergence result for the guided process X°.

Theorem 3.9. *Under Assumptions 3.1 and 3.8, there exists a constant* $M > 0$ *such that*

$$\limsup_{t \uparrow T} \frac{|y - L_{T-t} X_t^\circ|}{\sqrt{(T-t)\ln(1/(T-t))}} \leq M \quad \mathbb{P}^\circ\text{-a.s.} \tag{3.14}$$

3.3.2 Absolute continuity of $\mathbb{P}^\star$ and $\mathbb{P}^\circ$

In this section we establish sufficient conditions for the absolute continuity of $\mathbb{P}^\star$ with respect to $\mathbb{P}^\circ$ on $\mathcal{F}_T$. We require the following assumption on the functions h and g.

Assumption 3.10. *There exist continuous, positive functions* $\lambda_1(t) \leq \lambda_2(t)$ *satisfying* $\lim_{t \to 0} \lambda_1(t) = \lim_{t \to 0} \lambda_2(t) = 1$ *such that*

$$\lambda_1(T-t) g(t,x) \leq h(t,x) \leq \lambda_2(T-t) g(t,x), \quad t < T, x \in H. \tag{3.15}$$

Theorem 3.11. *Under Assumptions 3.1, 3.8 and 3.10, it holds that* $\mathbb{P}^\star$ *and* $\mathbb{P}^\circ$ *are absolutely continuous on* $\mathcal{F}_T$ *with*

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ}(X^\circ) = \Phi_T(X^\circ) \quad \mathbb{P}^\circ\text{-a.s.}, \tag{3.16}$$

where $\Phi_T(X^\circ) = \frac{g(0,x_0)}{h(0,x_0)} \Psi_T(X^\circ)$ *with* $\Psi_t(X^\circ)$ *defined as in Equation (3.12).*

We give the proof of Theorem 3.11 in Section 3.7.

Remark 3.12. As an immediate consequence of Theorem 3.11 and a change of variables, it follows that the laws $\mathcal{L}^\star$ and $\mathcal{L}^\circ$ are absolutely continuous on the path space $C([0,T]; H)$. This implies that we can obtain weighted samples of the intractable diffusion bridge $X^\star$ by sampling from the tractable guided process X° and evaluating the weights given by the Radon–Nikodym derivative (3.16). The weights contain an intractable multiplicative term $h(0, x_0)$. This term acts as a proportionality constant that cancels out in a renormalisation step of the weights.

3.3.3 Remarks on assumptions

Let us briefly comment on the assumptions underlying our main results. The assumptions in 3.1 are standard, with the exception of the boundedness of F in Assumption 3.1(iv). It is needed for our proof of Theorem 3.11. However, our numerical illustrations in Example 3.21 indicate that this assumption is stronger than required for the absolute continuity to hold.
Assumption 3.8 essentially states that the covariance matrix R_{T-t}^{-1} present in the guiding term $G(t,x)$ explodes as the guided process approaches the conditioning time $t \uparrow T$. Moreover, it is crucial that this divergence occurs at a linear rate. This aligns with the intuition gained from the simplest conditioned diffusion, the one-dimensional Brownian bridge, where the equivalent term in the Brownian bridge equation is given by $(T-t)^{-1}$. In this infinite-dimensional setting, Assumption 3.8 depends through the operators Q_t on the interplay between the drift A and covariance operator Q of Equation (3.1), as well as on the observation operator L. As illustrated in Section 3.5, this assumption can typically be validated directly in applications of interest. It may fail though for hypoelliptic diffusions on $\mathbb{R}^n$, such as an integrated diffusion. In that case, the temporal regularity of the sample paths is not the same for all components of the diffusion and

therefore the conclusion of Theorem 3.9 won't hold. We refer to Bierkens et al. [16] for a more detailed discussion.
Assumption 3.10 dictates the functions h and g to be asymptotically equivalent in the limit $t \uparrow T$. For finite dimensional state spaces, this assumption can be traced back to bounds on the transition densities of X and Z, using, for example, Aronson's inequality (cf. Aronson and Oser [6]). In our setting, such transition densities are not guaranteed to exist, making it necessary to work with h and g directly instead.
The following lemma gives a sufficient condition for Assumption 3.10 to be satisfied. It is shown in Section 3.7.2.

Lemma 3.13. *Let the nonlinearity F and covariance operator Q be such that $\tilde{F}(t,x) = Q^{-\frac{1}{2}}F(t,x)$ is well defined with $|\tilde{F}(t,x)| \leq C_{\tilde{F}}$ for all $t \in [0,T], x \in H$. Then there exists a continuous function $\lambda(t) \geq 1$ satisfying* $\lim_{t\to 0}\lambda(t) = 1$ *such that*

$$\frac{g(t,x)}{\lambda(T-t)} \leq h(t,x) \leq \lambda(T-t)g(t,x), \quad t < T,\ x \in H. \tag{3.17}$$

3.4 A Metropolis-Hastings algorithm for sampling infinite-dimensional diffusion bridges

We present a short description on how the theoretical result of Theorem 3.11 can be used to simulate the infinite-dimensional diffusion bridge $X^\star$. This can, for example, be achieved through a simple Metropolis-Hastings sampler (MH sampler), using the law $\mathcal{L}^\circ$ of the guided process as a proposal distribution to draw samples $X^\star$ of $\mathcal{L}^\star$ by iterating between the following steps:

1. Draw a sample from the proposal distribution $X^\circ \sim \mathcal{L}^\circ$.
2. Accept/reject X° with acceptance probability based on the Radon–Nikodym derivative $\Phi_T(X^\circ)$.

In practical applications, one has to numerically approximate X and $X^\star$ on some gridded domain $D \times [0,T], D \subset \mathbb{R}^d$. For such computational implementations, we write $\mathrm{solve}(A,F,Q,W)$ for any *numerical SPDE solver* that approximates the mild solution X of an SPDE of the form (3.1) on the given grid. We refer to Lord et al. [79], Chapter 10, for an introduction to the solvers used in the field. Since we are dealing with diagonalisable operators A and Q, a natural choice for the SPDE solver is the *spectral Galerkin method*.
The proposals of X° can be constructed to depend on the current value of the chain based on the *preconditioned Crank-Nicolson (pCN)* scheme (see Neal [87], Beskos et al. [14], Cotter et al. [37]) as follows. In step i, let V denote the process such that for the current value of the chain $X_i^\star = \mathrm{solve}(A, F + QG^\circ, Q, V)$. The proposal in the i-th sampling step then takes on the form:

(i) Draw a Wiener process W, independent of V;
(ii) Set $V^\circ = \sqrt{1-\beta^2}V + \beta W$;
(iii) Compute $X^\circ = \mathrm{solve}(A, F + QG^\circ, Q, V^\circ)$.

Here, $\beta \in (0,1]$ denotes a hyperparameter that determines the size of the pCN step. For $\beta = 1$, independent proposals of X° are drawn.
We summarise the complete algorithm in the following.

Algorithm 1: MH Sampler of $X^\star$

Input: SPDE Parameters A, F, Q and x_0, conditioning point $y \in H$, gridded domain $D \times [0,T]$, iterations N, step size β

Output: Samples $(X_i^\star)_{i=0}^N$ of $X^\star$ on $D \times [0,T]$

Initialise: Draw a Wiener process V and set $X_0^\star = \text{solve}(A, F + QG^\circ, Q, V)$;

for $i = 0 \dots N-1$ **do**

- **Proposal**
  - (i) Draw a Wiener process W;
  - (ii) Set $V^\circ = \sqrt{1-\beta^2} V + \beta W$;
  - (iii) Compute $X^\circ = \text{solve}(A, F + QG^\circ, Q, V^\circ)$;
- **Update**
  - (i) Compute $M = \min\left(1, \dfrac{\Phi_T(X^\circ)}{\Phi_T(X_i^\star)}\right)$;
  - (ii) Draw $U \sim \text{Unif}(0,1)$;
  - **if** $U < M$ **then**
    - Set $X_{i+1}^\star = X^\circ$ and $V = V^\circ$;
  - **else**
    - Set $X_{i+1}^\star = X_i^\star$.

Remark 3.14. Since the results in Theorem 3.11 are infinite-dimensional, the validity of Algorithm 1 depends neither on the choice of the numerical solver nor on the discretisation of the domain $D \times [0,T]$. In particular, for a practical implementation, both the solver and the mesh size can be freely chosen and the implementation will remain valid if the mesh size tends to zero. With a trade-off of greater computational costs, a finer mesh size will not only improve the approximation quality of the numerical SPDE solver, but also the approximation of the likelihood ratios $\Phi_T(X^\circ)$.

Remark 3.15. A substantial benefit of Algorithm 1 is that it samples from the diffusion bridge law $\mathcal{L}^\star$ independently of the existence of the SPDE bridge equation (3.8). Instead, $\mathcal{L}^\star$ is defined purely as the push-forward measure of X under $\mathbb{P}^\star$. The pCN steps introduced in Algorithm 1 can accordingly be interpreted as follows.

By the well-posedness of Equation (3.9), there exists a measurable map $\Gamma : C([0,T]; H') \to C([0,T]; H)$ such that $X^\circ = \Gamma(W)$. Here, W is a $\mathbb{P}^\circ$-cylindrical Wiener process on H, taking values in a larger Hilbert H' such that $H \hookrightarrow H'$ is embedded in a Hilbert-Schmidt way. Let $\mathbb{W}$ denote the corresponding Wiener measure on $C([0,T]; H')$ and define a measure $\mathbb{V}$ via

$$\frac{\mathrm{d}\mathbb{V}}{\mathrm{d}\mathbb{W}}(V) = \Phi_T(\Gamma(V)).$$

If V is then an H'-valued process with $V \sim \mathbb{V}$, it follows that

$$\mathbb{E}_{V\sim\mathbb{V}}[f(\Gamma(V))] = \mathbb{E}_{V\sim\mathbb{W}}[f(\Gamma(V))\Phi_T(\Gamma(V))] = \mathbb{E}^\circ[f(X^\circ)\Phi_T(X^\circ)] = \mathbb{E}^\star[f(X^\star)]$$

for any bounded and measurable functional f. Hence, the law of $\Gamma(V)$ equals the diffusion bridge law $\mathcal{L}^\star$. In this way, the MH sampler in Algorithm 1 can be interpreted as a pCN scheme in Wiener space that targets the measure $\mathbb{V}$. In the literature, any such V with law $\mathbb{V}$ is commonly referred to as the innovation process of $X^\star$.

3.5 Examples

3.5.1 Diagonalisable equations

Consider the case that the linear operators A and Q share an eigenbasis of H. Specifically, assume the following.

Assumption 3.16. *The operators A and Q are diagonalisable, i.e. there exists an orthonormal basis $(e_j)_{j\in\mathbb{N}}$ of H and positive sequences $(a_j)_{j\in\mathbb{N}}$ and $(q_j)_{j\in\mathbb{N}}$ such that $a_j \to \infty$, $\sup q_j < \infty$ and*

$$\begin{aligned} Ae_j &= -a_j e_j \\ Qe_j &= q_j e_j. \end{aligned} \tag{3.18}$$

Given Assumption 3.16, Equation (3.1) is commonly referred to as a *diagonalisable SPDE*. If F is a linear operator that commutes with A and Q, such an equation is referred to as *fully diagonalisable*. In that case, the mild solution X is an Ornstein-Uhlenbeck process that can be represented as an infinite series of decoupled SODEs in its eigenmodes. Note, however, that such a decomposition is no longer valid when F is a nonlinear operator.
In the diagonalisable setting, A generates a strongly continuous semigroup $(S_t)_{t\geq 0}$ with $S_t e_j = \exp(-a_j t)e_j$ and, in particular, satisfies Assumption 3.1(i). Moreover, Assumption 3.1(iii) is equivalent to

$$\sum_{j=1}^{\infty} \frac{q_j}{a_j} < \infty, \tag{3.19}$$

under which the operators $(Q_t)_{t\geq 0}$ are diagonalisable with eigenvalues

$$q_j(t) = \frac{q_j}{2a_j}\left[1 - \exp(-2a_j t)\right]. \tag{3.20}$$

In the following proposition, we show that, given a diagonalisable SPDE, Assumption 3.8 is satisfied for two common choices of the observation operator L.

Proposition 3.17. *Let Assumption 3.16 and Equation (3.19) be satisfied. Then, Assumption 3.8 holds in either of the following cases:*

- *$L = P_k$ is the projection onto the first k coordinates with respect to the eigenbasis $(e_j)_{j\in\mathbb{N}}$, i.e.*

$$P_k x = \left(\langle x, e_j\rangle\right)_{j=1}^{k}, \quad x \in H. \tag{3.21}$$

- *$L = P_w$ is the projection onto the coordinate of the subspace span$\{w\}$ for some fixed $w \in H$, i.e.*

$$P_w x = \langle x, w\rangle, \quad x \in H. \tag{3.22}$$

Proof. Since $R_t = LQ_tL^*$ is symmetric positive definite, to show (3.13) it suffices to show that

$$\begin{aligned} &\lambda_{\min}(R_t) \le t\, \underline{c}^{-1}, \\ &\lambda_{\min}(R_t) \ge t\, \bar{c}^{-1}, \quad t \in (0, T], \end{aligned} \tag{3.23}$$

for some constants $\bar{c} > 0$ and $\underline{c} > 0$, with $\lambda_{\min}(R_t)$ and $\lambda_{\max}(R_t)$ denoting the smallest and largest eigenvalue of R_t. We will use that, for any $j \geq 1$,

$$\begin{aligned} &\sup_{t\in(0,T]} t^{-1}(1 - \exp(-2a_j t)) = 2a_j \\ &\inf_{t\in(0,T]} t^{-1}(1 - \exp(-2a_j t)) = T^{-1}(1 - \exp(-2a_j T)), \end{aligned} \tag{3.24}$$

which follows from a Taylor expansion at $t = 0$.

Consider the first case of $L = P_k$. A simple computation shows that R_t is given by the diagonal $k \times k$-matrix with entries $q_j(t), j = 1, ..., k$. Hence, plugging in $q_j(t)$ as in (3.20) and (3.24) we have

$$\max_{1 \le j \le k} \sup_{t \in (0,T]} t^{-1} q_j(t) = \max_{1 \le j \le k} \sup_{t \in (0,T]} \frac{q_j}{2a_j} t^{-1}(1 - \exp(-2a_j t)) = \max_{1 \le j \le k} q_j =: \underline{c}^{-1}$$

as well as

$$\min_{1 \le j \le k} \inf_{t \in (0,T]} t^{-1} q_j(t) = \min_{1 \le j \le k} \inf_{t \in (0,T]} \frac{q_j}{2a_j} t^{-1}(1 - \exp(-2a_j t)) = \min_{1 \le j \le k} \frac{q_j}{2a_j} T^{-1}(1 - \exp(-2a_j T)) =: \bar{c}^{-1}.$$

This shows (3.23) in the first case. For the second case, note that P_w^* is the mapping $y \mapsto yw \in H$ for any $y \in \mathbb{R}$. From this, it follows that

$$R_t y = (P_w Q_t P_w^*) y = (LQ_t)(yw) = \langle Q_t w, w \rangle y$$

and hence $R_t = \langle Q_t w, w \rangle$. Plugging in $q_j(t)$ as given in (3.20), it holds

$$\begin{aligned} \sup_{t \in (0,T]} t^{-1} R_t &= \sup_{t \in (0,T]} t^{-1} \langle Q_t w, w \rangle \\ &\le \sum_{j=1}^{\infty} \frac{q_j}{2a_j} \left(\sup_{t \in (0,T]} t^{-1} (1 - \exp(-2a_j t)) \right) |\langle w, e_j \rangle|^2 \\ &= \sum_{j=1}^{\infty} q_j |\langle w, e_j \rangle|^2 = \langle Q^{\frac{1}{2}} w, Q^{\frac{1}{2}} w \rangle =: \underline{c}^{-1}. \end{aligned}$$

Here, we used (3.24) in the third step and the fact that Q is positive in the last step. This shows the first inequality in (3.23). The second inequality follows from the fact that

$$\begin{aligned} \inf_{t \in (0,T]} t^{-1} R_t &\ge \sum_{j=1}^{\infty} \frac{q_j}{2a_j} \left(\inf_{t \in (0,T]} t^{-1} (1 - \exp(-2a_j t)) \right) |\langle w, e_j \rangle|^2 \\ &= \sum_{j=1}^{\infty} \frac{q_j}{2a_j} T^{-1} (1 - \exp(-2a_j T)) |\langle w, e_j \rangle|^2 =: \bar{c}^{-1}, \end{aligned}$$

where we again used (3.24) in the second step and (3.19) in the last step to justify the convergence of the sum. This shows (3.23) and thus finishes the proof.

□

REMARK 3.18. Consider the special case that $H = L^2(D)$ for some bounded domain $D \subset \mathbb{R}^d$. In that case, the observation operator

$$P_w x = \langle x, w \rangle = \int_D x(\xi) w(\xi) \, \mathrm{d}\xi$$

of Proposition 3.17 can be understood as observing a weighted average $P_w X_T$ of X_T. Depending on the choice of $w \in L^2(D)$, this can, for example, correspond to observing a global or local average of X_T. Moreover, it is clear that the second case in Proposition 3.17 extends to the case that one observes a set of weighted averages $\int_D x(\xi) w_j(\xi) \, \mathrm{d}\xi$ for functions $w_1, ..., w_k \in L^2(D)$.

EXAMPLE 3.19 (Stochastic reaction-diffusion equations). Consider a stochastic reaction-diffusion equation, formally written as

$$\partial_t X(t, \xi) = \partial_\xi^2 X(t, \xi) + f(X(t, \xi)) + \dot{w}(t, \xi), \quad t, \xi \in [0, T] \times [0, \pi], \tag{3.25}$$

with initial condition $X(0,\xi) = x_0(\xi)$ and homogeneous Dirichlet boundary conditions $X(t,0) = X(t,\pi) = 0$ for all $t,\xi \in [0,T] \times [0,\pi]$. Here, f is a continuous, scalar-valued function of at most linear growth and $\dot{w}$ is a Gaussian noise term that is white in time and possibly spatially dependent.
It is well known that (3.25) can be stated as a semilinear SPDE of the form (3.1), where A denotes the Dirichlet Laplace operator $A = \partial_\xi^2$ on $H = L^2([0,\pi])$ with domain $D(A) = H^2([0,\pi]) \cap H_0^1([0,\pi])$. Here, $H^2([0,\pi])$ is the Sobolev space $W^{2,2}([0,\pi])$ and H_0^1 denotes the closure of the smooth and compactly supported test functions in $W^{1,2}([0,\pi])$. Furthermore, F is the nonlinearity given by the Nemytskii operator $F(X_t)(\xi) = f(X(t,\xi))$ generated by f and Q is a symmetric, positive definite operator on H.
The Dirichlet Laplace operator admits an eigenbasis $e_j(\xi) = \sin(j\xi)$ with eigenvalues $-a_j = -j^2, j \geq 1$ on H. Hence, Assumption 3.16 and Equation (3.19) are satisfied for any operator Q on H that is diagonalisable with respect to $(e_j)_j$ with eigenvalues $q_j \sim j^{-r}$ for some $r \geq 0$.

3.5.2 Stochastic Amari equation

Assume the stochastic Amari equation, formally given by

$$\partial_t X(t,\xi) = -X(t,\xi) + \int_{[0,\pi]} f(\xi,\xi')s(X(t,\xi') - \theta)\,\mathrm{d}\xi' + \dot{w}(t,\xi), \quad t,\xi \in [0,T] \times [0,\pi], \tag{3.26}$$

with initial condition $X(0,\xi) = x_0(\xi)$. Here, $X(t,\xi)$ models the activity of a neural field at time t and location ξ. The mapping f represents a spatial connectivity function on $[0,\pi]$, whereas s is an activation function with threshold $\theta > 0$. The term $-X(t,\xi)$ is a linear damping term that models relaxation of neural activity.
For more details on the Amari equation and neural field models we refer to the recent survey in Cook et al. [35] and to Faugeras and Inglis [53] for a mathematical view on the field.
The Amari equation can be formulated as a semilinear SPDE of the form (3.1) on $H = L^2([0,\pi])$ with $AX = -X$ and $F(X)(\xi) = \int_{[0,\pi]} f(\xi,\xi')s(X(\xi') - \theta)\,\mathrm{d}\xi'$. Typically, f is chosen to be continuous and s to be Lipschitz continuous, in which case F is itself a bounded and Lipschitz continuous mapping on H. Since the semigroup generated by A is no longer smoothing, one requires the noise operator Q to be trace-class. In that case, Assumption 3.1 is met. Moreover, it follows, just as in the proof of Proposition 3.17, that Assumption 3.8 is satisfied for the observation operators $L = P_k$ and $L = P_w$.
We apply Lemma 3.13 to show that Assumption 3.10 is satisfied for common choices of the spatial connectivity function f. One such typical choice is given by a difference of Gaussian kernels

$$f(\xi,\xi') = A_1 \exp\left(-\frac{|\xi - \xi'|^2}{2\sigma_1^2}\right) - A_2 \exp\left(-\frac{|\xi - \xi'|^2}{2\sigma_2^2}\right)$$

for some positive parameters $A_1, A_2, \sigma_1, \sigma_2$. Consider the associated Hilbert-Schmidt integral operator T_f defined by

$$(T_f u)(\xi) = \int_{[0,\pi]} f(\xi,\xi')u(\xi')\,\mathrm{d}\xi', \quad u \in H.$$

By Mercer's theorem, it follows that T_f admits an eigenbasis $(e_j)_{j\geq 1}$ of H and it is a well-known result that the eigenvalues λ_j of T_f are of exponential decay $\lambda_j \sim \exp(-j^2)$. Hence, representing the nonlinearity F as $F(X) = T_f s_\theta(X)$, where we write with slight abuse of notation s_θ for the Nemytskii operator $s_\theta(X)(\xi) = s(X(\xi) - \theta)$, it follows

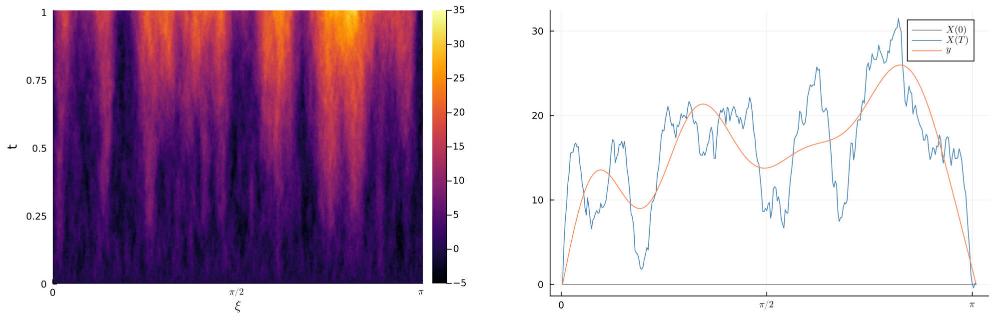


Figure 3.5.1. Forward simulation of X. Left: Heatmap of a sample path $X(t, \xi)$ of Equation (3.27) with parameters as specified in (3.29). Right: The states X_0 and X_T and the observation $y = P_k X_T$, $k = 10$.

that for any trace-class operator Q with eigenvalues $q_j \sim j^{-r}, r > 1$, we have that $Q^{-1/2}F = Q^{-1/2}T_f s_\theta$ is well-defined and bounded.

3.5.3 Numerical illustrations

We illustrate the performance of Algorithm 1 with numerical examples for two distinct stochastic reaction-diffusion equations.

Example 3.20. Assume the SPDE given by

$$\mathrm{d}X_t = \left[\eta A X_t + \frac{\zeta_1 X_t^2}{1 + \zeta_2 X_t^2}\right] \mathrm{d}t + \sigma \,\mathrm{d}W_t, \;\; X(0) = x_0, \tag{3.27}$$

with Dirichlet Laplacian A, white noise covariance operator $Q = \sigma^2 \mathrm{Id}$ and nonlinearity F generated by

$$f(x) = \frac{\zeta_1 x^2}{1 + \zeta_2 x^2}. \tag{3.28}$$

Reactions terms of the form (3.28) are present for example in predator-prey models (Holling [66], Dawes and Souza [42]), activator-inhibitor models (Young [116], Pasemann et al. [92]) or Michaelis-Menten kinetics (Johnson and Goody [72]). Typically, the given examples model multiple interacting populations with a system of coupled SPDEs and nonlinearity (3.28) depending linearly on another population. However, for the sake of simplicity, we restrict our attention to the case where such additional populations are assumed to be constant, leading to a reaction-diffusion equation of the form (3.27).
Following Example 3.19, Equation (3.27) is diagonalisable such that the condition (3.19) is met. Moreover, since F is bounded and Q is boundedly invertible, it follows that both Assumption 3.1 and Assumption 3.10 are satisfied. We assume one partial observation $y \in \mathbb{R}^k$ given at time T by the spectral projection $y = P_k X_T$ as defined in (3.21). By Proposition 3.17, Assumption 3.8 is then also satisfied.
For our numerical implementation, we set the parameters of Equation (3.27) as

$$[\eta, \zeta_1, \zeta_2, \sigma] = [3 \times 10^{-3}, 3, 0.1, 1]. \tag{3.29}$$

Figure 3.5.1 shows the heatmap of one sample path of X on the time interval $[0, T] = [0, 1]$ with initial value $x_0 \equiv 0$. The path is sampled based on a spectral Galerkin approximation, using the first 100 eigenfunctions, and a semi-implicit Euler-Maruyama scheme to approximate the resulting SODE with time steps $\Delta t = 0.005$. For the observation $y = P_k X_T$ we set $k = 10$. The initial state x_0, true state X_T and observed state y are shown on the right-hand side in Figure 3.5.1.

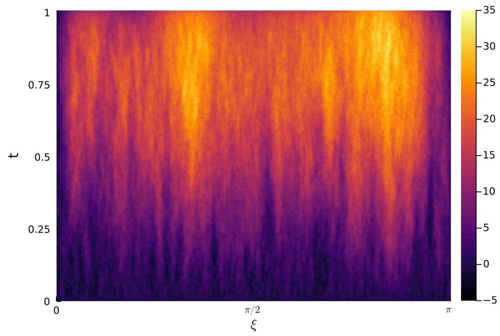


Figure 3.5.2. Heatmap of a sample path $X^\circ(t, \xi)$ of the guided process corresponding to (3.27) with conditioning state $y = P_k X_T, k = 10$.

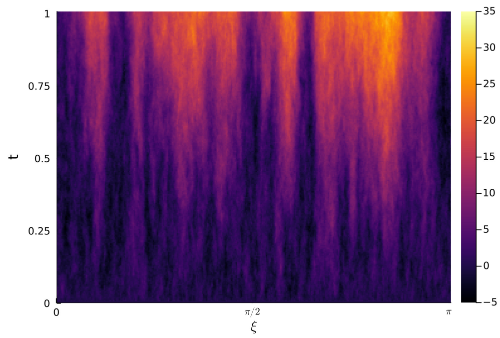


Figure 3.5.3. Heatmap of the mean sample path of 100 samples of the estimated diffusion bridge $X^\star(t, \xi)$ of Equation (3.27) returned by Algorithm 1.

We employ Algorithm 1 to sample from the infinite-dimensional diffusion bridge $X^\star$ of Equation (3.27), conditioned on the observed state y. In Figure 3.5.2, we display a random sample path of the guided process X°, simulated with the same SPDE solver and mesh size as the path of Figure 3.5.1. It serves as the proposal distribution in the MH sampler. While indeed the process is forced to satisfy $LX_t^\circ = y$, one can visually assess that the guided process looks different from the forward simulated path in Figure 3.5.1, motivating the application of Algorithm 1.
The MH sampler is run with step size $\beta = 0.08$ and 50 000 iterations. The resulting acceptance rate of proposals equals 26%. Figure 3.5.3 shows the mean sample path of the last 100 samples of the estimated diffusion bridge distribution. A visual comparison with the sample paths of the data generating process X and the guided process X° strongly suggests that Algorithm 1 correctly samples from the distribution of the diffusion bridge $X^\star$.
This is further supported by the display of the sampler's performance in Figures 3.5.4 and 3.5.5. For this, we let

$$j_1 = \arg\max_j x_j(T) = \arg\max_j \langle X_T, e_j \rangle, \tag{3.30}$$

$$j_2 = \arg\min_j x_j(T) = \arg\min_j \langle X_T, e_j \rangle \tag{3.31}$$

be the indices of the largest and smallest spectral mode of X_T. In our example, this corresponds to $j_1 = 1$ and $j_2 = 2$. Out of the 50 000 sampled paths, Figure 3.5.4 shows

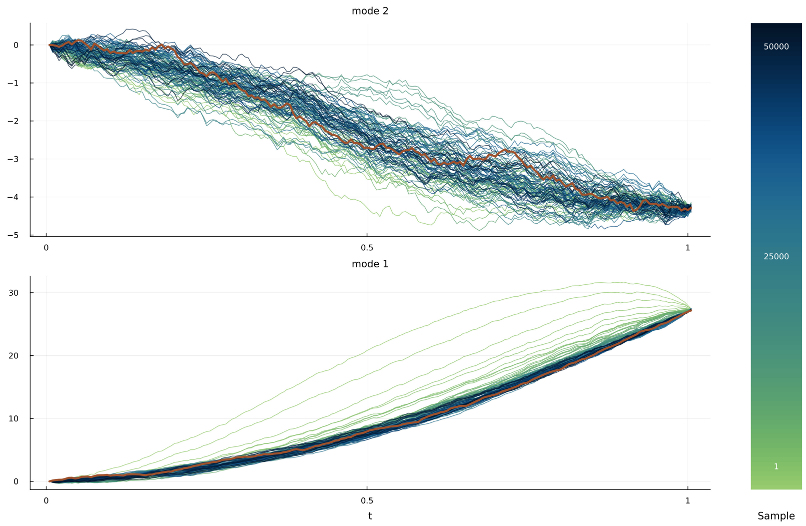


FIGURE 3.5.4. Paths of two spectral modes of the diffusion bridge samples of Equation (3.27) returned by Algorithm 1. Every 500-th sample is shown. Green paths represent 'earlier' samples in the Markov chain, whereas blue paths represent 'later' samples. The orange path is the spectral mode of the data generating process X.

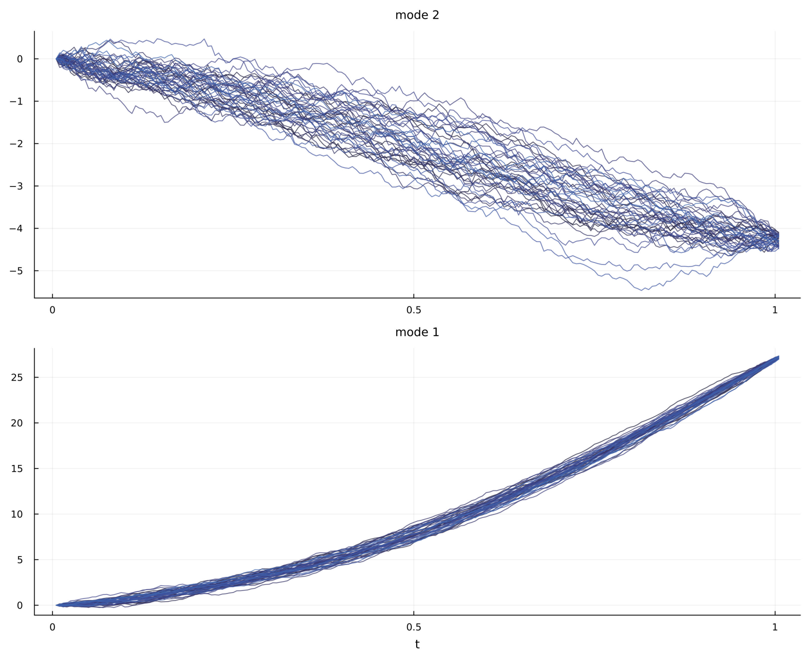


FIGURE 3.5.5. Sample paths in the spectral modes 1 and 2 obtained by forward sampling the SPDE (3.27) with parameters as specified in (3.29). Only paths that satisfy (3.32) are kept.

every 500-th path of the spectral modes $\{x^{\star}_{j_k}(t) : t \in [0, T]\}, k = 1, 2$, as well as the 'true' path of the data generating process $x_j(t)$.

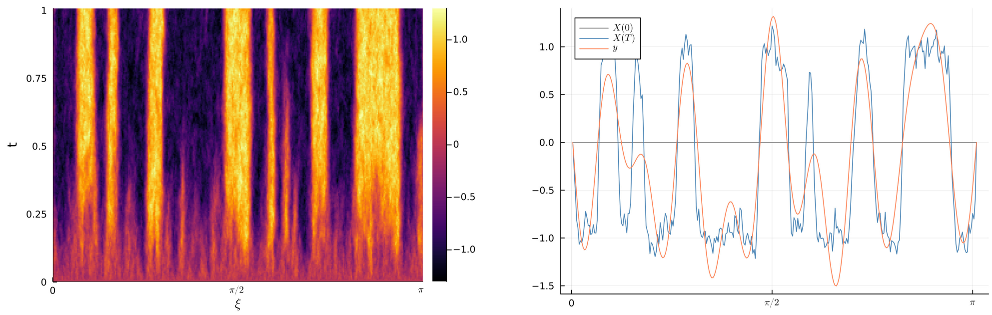


Figure 3.5.6. Forward simulation of X. Left: Heatmap of a sample path $X(t, \xi)$ of Equation (3.33) with parameters as specified in (3.34). Right: The states X_0 and X_T and the observation $y = P_k X_T$, $k = 20$.

For a comparison, we plot in Figure 3.5.5 the 'typical' sample path of the diffusion 'bridge' conditioned on hitting $x_{j_k}(T), k = 1, 2$. This is achieved by forward sampling 100 000 paths Y of Equation (3.27) and only keeping those sample paths that satisfy

$$|y_{j_k}(T) - x_{j_k}(T)| < \varepsilon, \quad k = 1, 2. \tag{3.32}$$

Here we choose $\varepsilon = 0.2$. Note that the other spectral modes are disregarded. Even still, due to the high dimensionality of the problem, this is a rare event - from the 100 000 forward samples all but 47 samples are discarded. A comparison between Figures 3.5.4 and 3.5.5 shows that the sample paths resemble each other well.

Example 3.21 (Stochastic Allen-Cahn Equation). Consider the stochastic Allen-Cahn equation

$$\mathrm{d}X_t = \left[\eta A X_t + \zeta(X_t - X_t^3)\right] \mathrm{d}t + Q_\sigma^{\frac{1}{2}} \,\mathrm{d}W_t, \;\; X(0) = x_0, \tag{3.33}$$

with Dirichlet Laplacian A, nonlinearity F generated by $f(x) = x - x^3$, diffusion parameter $\eta > 0$ and reaction rate $\zeta > 0$. Here, we choose a diagonalisable trace-class operator Q_σ with eigenvalues

$$q_j = \sigma_0^2(\rho^{-2} + (2\pi j)^2)^{-(1/2+\nu)}$$

parametrized by $\sigma = (\sigma_0, \rho, \nu)$. The operator Q_σ then corresponds to the Hilbert-Schmidt integral operator

$$(Q_\sigma x)(\xi) = \int_{[0,\pi]} q(|\xi - \xi'|) x(\xi') \,\mathrm{d}\xi', \quad x \in L^2([0, \pi]),$$

where $q(r)$ is the Matérn covariance function $q(r) = \sigma_0^2 \rho^{-1} r K_\nu(r/\rho)$, with K_ν denoting the second kind modified Bessel function of order ν. This follows from the fact that a Gaussian variable on H with covariance operator Q_σ is itself the weak solution to a stochastic fractional Laplace equation, see for example Borovitskiy et al. [17], Theorem 5. The parameter σ_0 corresponds to the marginal variance of $q(r)$, whereas ρ is a precision parameter. The parameter ν controls the 'smoothness' of Q_σ. Specifically, it can be shown that a Gaussian variable on H with covariance operator Q_σ is ν-times mean differentiable.

Just as in Example 3.20, A and Q_σ satisfy the Assumptions 3.1 and 3.8. However, since F is not bounded, Assumption 3.1(iv) is not met. Moreover, the unbounded nature of F and the fact that Q_σ is trace-class renders Assumption 3.10 difficult to verify as Lemma 3.13 is no longer applicable. Nonetheless, the results of our numerical experiments

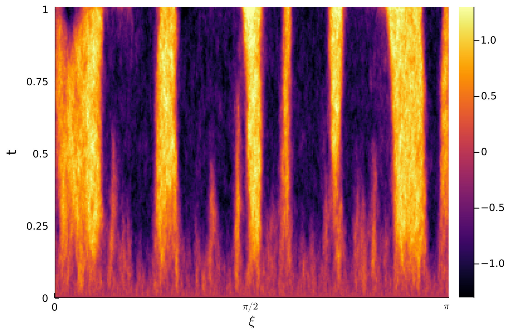


FIGURE 3.5.7. Heatmap of a sample path $X^\circ(t,\xi)$ of the guided process corresponding to (3.33) with conditioning state $y = P_k X_T, k = 20$.

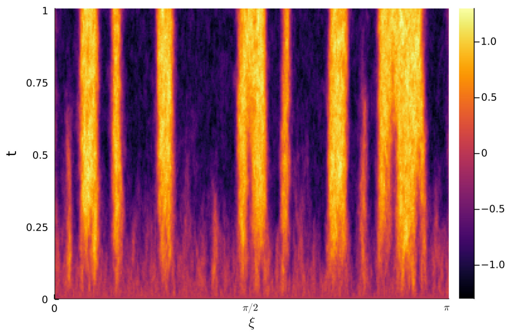


FIGURE 3.5.8. Heatmap of the mean sample path of 100 samples of the estimated diffusion bridge $X^\star(t,\xi)$ of Equation (3.33) returned by Algorithm 1.

indicate that these assumptions are stricter than necessary for the results of Theorem 3.11 to hold.

We follow the experimental setup of Example 3.20, using the observation scheme $y = P_k X_T$ with $k = 20$ and the same numerical solver on an identical space-time grid. For the parametrisation of (3.33) we set

$$[\eta, \zeta, \sigma_0, \rho, \nu] = [2 \times 10^{-3}, 10, 10^7, 5 \times 10^{-6}, 1]. \tag{3.34}$$

Figure 3.5.6 shows the heatmap of a corresponding sample path X. It displays the typical pattern formation of a noisy Allen-Cahn equation with steady states at ± 1. The initial state x_0, true state X_T and observed state y are shown on the right in Figure 3.5.6.

In Figure 3.5.7 we display a random sample path of the guided process X°, whereas Figure 3.5.8 shows the mean sample path of 100 samples of the estimated diffusion bridge distribution after running Algorithm 1 with step size $\beta = 0.07$, 50 000 iterations and a resulting 27% acceptance rate.

A close inspection shows that the guided process fails to capture steady states that are not represented in the data y, most notably around $\xi \approx \pi/8$. In contrast, the diffusion bridge mean captures all steady states displayed by the data generating path X, with a separation between the states that is characteristic for the Allen-Cahn equation.

In Figure 3.5.9 we display the paths of the spectral modes $\{x^\star_{j_k}(t) : t \in [0,T], k = 1,2\}$, with j_1, j_2 as defined in (3.30). In this case, we have $j_1 = 5$ and $j_2 = 11$. For comparison, Figure 3.5.10 shows the 'typical' spectral sample paths of the diffusion 'bridge', obtained

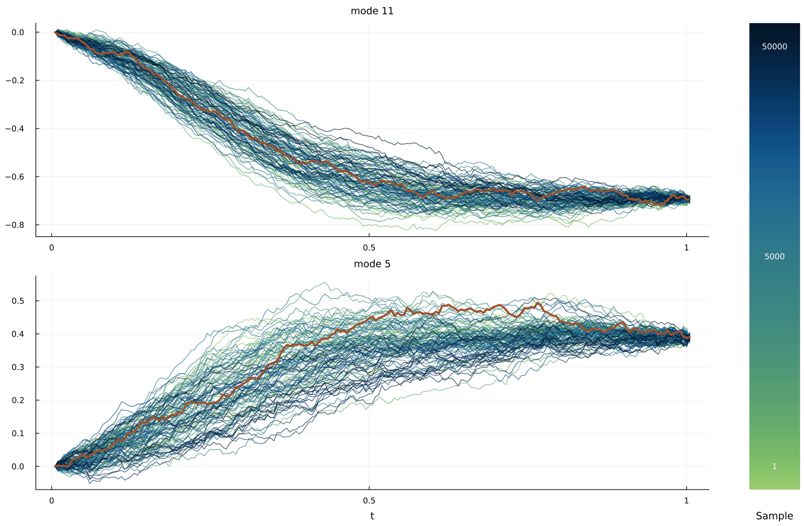


FIGURE 3.5.9. Paths of two spectral modes of the diffusion bridge samples of Equation (3.33) returned by Algorithm 1. Every 500-th sample is shown. Green paths represent 'earlier' samples in the Markov chain, whereas blue paths represent 'later' samples. The orange path is the spectral mode of the data generating process X.

by sampling 100 000 paths Y of Equation (3.33) and only keeping those paths that satisfy (3.32) with $\varepsilon = 0.05$. This results in 52 out of the 100 000 samples being kept. The comparison of Figures 3.5.9 and 3.5.10 again shows that the 'typical' spectral path obtained through this heuristic sampling approach matches the output of Algorithm 1 quite well.

We remark that a precise assessment of the accuracy of the bridge sampling procedure is hard to obtain in this infinite-dimensional setting. Therefore, we provide in Appendix 3.C additional numerical validation that the result in Theorem 3.11 remains true when Assumptions 3.1(iii) and 3.10 are not necessarily met.

3.6 Proof of Theorem 3.9

To ease readability of the proof of Theorem 3.9, we write $\Delta_t = T - t$ for any $t < T$. The structure of the proof follows that of the proof of Proposition 2.2 in Bierkens et al. [16]. It can be outlined as follows:

1. Given $Y_t = y - L_{\Delta_t} X_t^\circ$, define the Lyapunov function $V(t, Y) = \frac{1}{2}\langle R_{\Delta_t}^{-1} Y_t, Y_t\rangle$.
2. Apply Itô's lemma on $V(t, Y)$.
3. Bound the terms appearing in $\mathrm{d}V(t, Y)$.
4. Apply a Gronwall-type inequality.

Steps 1 and 2 are covered in the following lemma.

LEMMA 3.22. *Let $Y_t = y - L_{\Delta_t} X_t^\circ$ and let $V(t, Y_t) = \frac{1}{2}\langle R_{\Delta_t}^{-1} Y_t, Y_t\rangle$. Then*

$$\begin{aligned} \mathrm{d}V(t, Y_t) = &-\langle L_{\Delta_t}^* R_{\Delta_t}^{-1} Y_t, F(t, X_t^\circ)\rangle\,\mathrm{d}t \\ &+ \langle Q^{\frac{1}{2}} L_{\Delta_t}^* R_{\Delta_t}^{-1} Y_t, \mathrm{d}\tilde{W}_t\rangle - \frac{1}{2}\langle Q^{\frac{1}{2}} L_{\Delta_t}^* R_{\Delta_t}^{-1} Y_t, Q^{\frac{1}{2}} L_{\Delta_t}^* R_{\Delta_t}^{-1} Y_t\rangle\,\mathrm{d}t \\ &+ \frac{1}{4}\operatorname{tr}\left[R_{\Delta_t}^{-1} L_{\Delta_t} Q L_{\Delta_t}^*\right]\mathrm{d}t, \quad \mathbb{P}^\circ\text{-}a.s., \end{aligned} \tag{3.35}$$

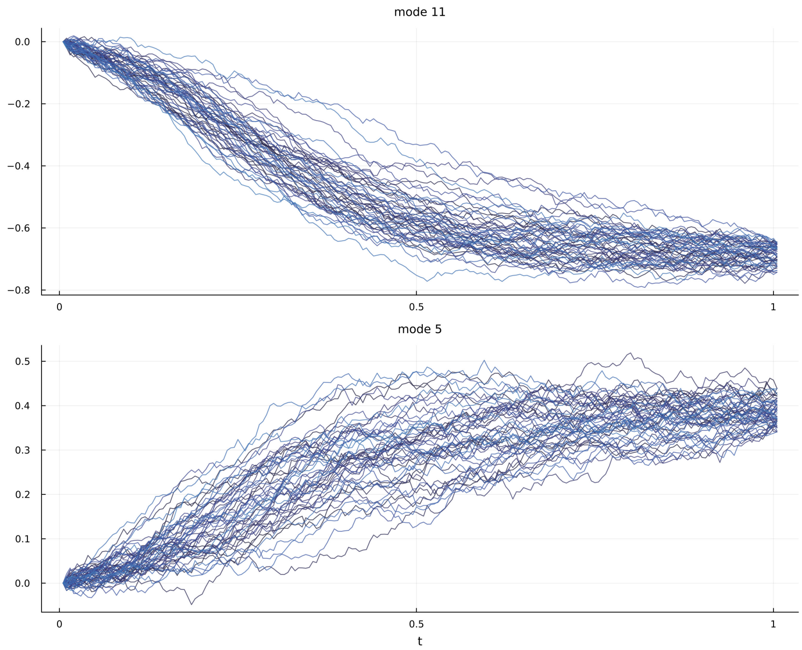


FIGURE 3.5.10. Sample paths in the spectral modes 5 and 11 obtained by forward sampling the SPDE (3.33) with parameters as specified in (3.34). Only paths that satisfy (3.32) are kept.

where $\tilde{W}$ is the cylindrical Wiener process $\tilde{W}_t = -W_t^\circ$.

PROOF. Since X° is a mild solution to the SPDE (3.9), it satisfies

$$X_t^\circ = S_t x_0 + \int_0^t S_{t-s}\left[F(s, X_s^\circ) + QG(s, X_s^\circ)\right] \mathrm{d}s + \int_0^t S_{t-s} Q^{\frac{1}{2}} \,\mathrm{d}W_s^\circ, \quad \mathbb{P}^\circ\text{-a.s.}$$

Hence, using the semigroup property $S_{T-t}S_{t-s} = S_{T-s}$, it follows that

$$\begin{aligned} L_{\Delta_t} X_t^\circ &= L S_{T-t} X_t^\circ \\ &= L\left(S_T x_0 + \int_0^t S_{T-s}\left[F(s, X_s^\circ) + QG(s, X_s^\circ)\right] \mathrm{d}s + \int_0^t S_{T-s} Q^{\frac{1}{2}} \,\mathrm{d}W_s^\circ\right). \end{aligned}$$

Setting $\tilde{W}_t := -W_t^\circ$, it thus holds

$$\mathrm{d}Y_t = -L_{\Delta_t}\left[F(t, X_t^\circ) + QG(t, X_t^\circ)\right] \mathrm{d}t + L_{\Delta_t} Q^{\frac{1}{2}} \,\mathrm{d}\tilde{W}_t. \tag{3.36}$$

Next, by definition, Q_t satisfies $\mathrm{d}Q_t = S_t Q S_t^* \,\mathrm{d}t$. Hence, for $R_t = LQ_tL^*$, it follows that

$$\mathrm{d}R_t = L\,\mathrm{d}Q_t L^* = L S_t Q S_t^* L^* \,\mathrm{d}t = L_t Q L_t^* \,\mathrm{d}t$$

which gives that

$$\frac{\mathrm{d}}{\mathrm{d}t} R_{\Delta_t}^{-1} = -R_{\Delta_t}^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t} R_{\Delta_t}\right) R_{\Delta_t}^{-1} = R_{\Delta_t}^{-1} L_{\Delta_t} Q L_{\Delta_t}^* R_{\Delta_t}^{-1}. \tag{3.37}$$

Plugging in (3.36) and (3.37), we get

$$\begin{aligned}\mathrm{d}\left(R_{\Delta_t}^{-1}Y_t\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}R_{\Delta_t}^{-1}\right)Y_t\,\mathrm{d}t + R_{\Delta_t}^{-1}\,\mathrm{d}Y_t \\ &= R_{\Delta_t}^{-1}L_{\Delta_t}QL_{\Delta_t}^{*}R_{\Delta_t}^{-1}Y_t\,\mathrm{d}t \\ &\quad - R_{\Delta_t}^{-1}L_{\Delta_t}\left[F(t,X_t^{\circ}) + QG(t,X_t^{\circ})\right]\mathrm{d}t + R_{\Delta_t}^{-1}L_{\Delta_t}Q^{\frac{1}{2}}\,\mathrm{d}\tilde{W}_t \\ &= -R_{\Delta_t}^{-1}L_{\Delta_t}F(t,X_t^{\circ})\,\mathrm{d}t + R_{\Delta_t}^{-1}L_{\Delta_t}Q^{\frac{1}{2}}\,\mathrm{d}\tilde{W}_t,\end{aligned} \tag{3.38}$$

where in the last step we use that $QG(t,X_t^{\circ}) = QL_{\Delta_t}^{*}R_{\Delta_t}^{-1}Y_t$.
Finally, it follows from Itô's lemma in the second line, plugging in (3.36) and (3.38) in the third line and again using $QG(t,X_t^{\circ}) = QL_{\Delta_t}^{*}R_{\Delta_t}^{-1}Y_t$ in the last equality that

$$\begin{aligned}2\,\mathrm{d}V(t,Y_t) &= \mathrm{d}\langle R_{\Delta_t}^{-1}Y_t, Y_t\rangle \\ &= \langle R_{\Delta_t}^{-1}Y_t, \mathrm{d}Y_t\rangle + \langle \mathrm{d}\left(R_{\Delta_t}^{-1}Y_t\right), Y_t\rangle + \frac{1}{2}\operatorname{tr}\left[R_{\Delta_t}^{-1}L_{\Delta_t}QL_{\Delta_t}^{*}\right]\mathrm{d}t \\ &= -\langle R_{\Delta_t}^{-1}Y_t, L_{\Delta_t}\left[F(t,X_t^{\circ}) + QG(t,X_t^{\circ})\right]\rangle\,\mathrm{d}t + \langle R_{\Delta_t}^{-1}Y_t, L_{\Delta_t}Q^{\frac{1}{2}}\,\mathrm{d}\tilde{W}_t\rangle \\ &\quad - \langle R_{\Delta_t}^{-1}L_{\Delta_t}F(t,X_t^{\circ}), Y_t\rangle\,\mathrm{d}t + \langle R_{\Delta_t}^{-1}L_{\Delta_t}Q^{\frac{1}{2}}\,\mathrm{d}\tilde{W}_t, Y_t\rangle \\ &\quad + \frac{1}{2}\operatorname{tr}\left[R_{\Delta_t}^{-1}L_{\Delta_t}QL_{\Delta_t}^{*}\right]\mathrm{d}t \\ &= -2\langle L_{\Delta_t}^{*}R_{\Delta_t}^{-1}Y_t, F(t,X_t^{\circ})\rangle\,\mathrm{d}t \\ &\quad + 2\langle Q^{\frac{1}{2}}L_{\Delta_t}^{*}R_{\Delta_t}^{-1}Y_t, \mathrm{d}\tilde{W}_t\rangle - \langle Q^{\frac{1}{2}}L_{\Delta_t}^{*}R_{\Delta_t}^{-1}Y_t, Q^{\frac{1}{2}}L_{\Delta_t}^{*}R_{\Delta_t}^{-1}Y_t\rangle\,\mathrm{d}t \\ &\quad + \frac{1}{2}\operatorname{tr}\left[R_{\Delta_t}^{-1}L_{\Delta_t}QL_{\Delta_t}^{*}\right]\mathrm{d}t.\end{aligned} \tag{3.39}$$

□

We proceed with the more technical steps 3 and 4.

Proof. (of Theorem 3.9) We aim to bound all terms appearing on the right hand side of (3.35), starting with the stochastic integral term. For this, fix some $t_0 \in [0,T)$ and define

$$M_t = \int_{t_0}^{t} \langle Q^{\frac{1}{2}}L_{\Delta_s}^{*}R_{\Delta_s}^{-1}Y_s, \mathrm{d}\tilde{W}_s\rangle, \quad t \in [t_0, T).$$

Note that M_t is a square-integrable martingale with quadratic variation

$$[M]_t = \int_{t_0}^{t} \langle Q^{\frac{1}{2}}L_{\Delta_s}^{*}R_{\Delta_s}^{-1}Y_s, Q^{\frac{1}{2}}L_{\Delta_s}^{*}R_{\Delta_s}^{-1}Y_s\rangle\,\mathrm{d}s$$

and in particular $\exp\left(M_t - \frac{1}{2}[M]_t\right)$ is a martingale which lets us apply the exponential martingale inequality as follows;
For any sequence $(\gamma_k)_k$ of positive numbers and $t_k = T - 1/k,\ k \in \mathbb{N}$, define the events

$$E_k = \left\{ \sup_{t_0 \le t \le t_{k+1}} \left(M_t - \frac{1}{2}[M]_t \right) > \gamma_k \right\}.$$

An application of Mao [84], Theorem 1.7.4, then shows that $\mathbb{P}(E_k) \le \exp(-\gamma_k)$.
Now, let $(\gamma_k)_k$ be an arbitrary but fixed sequence that satisfies

$$\sum_{k=1}^{\infty} \exp(-\gamma_k) < \infty. \tag{3.40}$$

It then follows that $\sum_{k=1}^{\infty} \mathbb{P}(E_k) \leq \sum_{k=1}^{\infty} \exp(-\gamma_k) < \infty$, and thus, by the Borel-Cantelli lemma, that $\mathbb{P}(\limsup_k E_k) = 0$. Hence, for $\mathbb{P}$-a.e. ω there exists some $k_0(\omega)$ such that for all $k \geq k_0(\omega)$

$$\sup_{t_0 \leq t \leq t_{k+1}} \left(M_t - \frac{1}{2} \left[M\right]_t \right) \leq \gamma_k. \tag{3.41}$$

The remaining terms in (3.35) are easily bounded as follows. From Assumption 3.1, it follows that there exists some $c > 0$ such that $\sup_{x \in H, t \in [0,T]} |L_{\Delta_t} F(t,x)| \leq c$. Hence, with $\|R_{\Delta_t}^{-1}\| \leq \bar{c} \Delta_t^{-1}$ as given in Assumption 3.8, it holds on $[0,T)$ that

$$\left| \langle L_{\Delta_t}^* R_{\Delta_t}^{-1} Y_t, F(t, X_t^\circ) \rangle \right| \leq c\ \bar{c}\ \Delta_t^{-1} |Y_t|.$$

Moreover, by Lemma 3.23, we have $\operatorname{tr}[L_t Q L_t^*] \leq \bar{c}'$ for any $t \in [0,T]$ and thus

$$\left| \operatorname{tr}\left[R_{\Delta_t}^{-1} L_{\Delta_t} Q L_{\Delta_t}^* \right] \right| \leq \bar{c}\bar{c}' \Delta_t^{-1}.$$

In total, we get for any $t \in [t_0, t_{k+1}]$ that

$$\begin{aligned} V(t, Y_t) &= V(t_0, Y_{t_0}) + \int_{t_0}^t \mathrm{d}V(s, Y_s) \\ &\leq V(t_0, Y_{t_0}) + \gamma_k + c\ \bar{c} \int_{t_0}^t \Delta_s^{-1}\ |Y_s|\ \mathrm{d}s \\ &\quad + \frac{\bar{c}\bar{c}'}{4} \int_{t_0}^t \Delta_s^{-1}\ \mathrm{d}s. \end{aligned} \tag{3.42}$$

Now define $\xi_t = \Delta_t^{-1} |Y_t|^2$. With $\|R_{\Delta_t}^{-1}\| \geq \underline{c}\ \Delta_t^{-1}$ as given in Assumption 3.8 and the symmetry of $R_{\Delta_t}^{-1}$, it follows that

$$V(t, Y_t) = \frac{1}{2} \langle R_{\Delta_t}^{-1} Y_t, Y_t \rangle \geq \frac{1}{2} \underline{c}\ \Delta_t^{-1} |Y_t|^2 = \frac{1}{2} \underline{c}\ \xi_t.$$

Hence, the inequality in (3.42) can be rewritten as

$$\begin{aligned} \xi_t &\leq M_0 + \frac{2}{\underline{c}} \gamma_k + M_1 \int_{t_0}^t \sqrt{\xi_s} \Delta_s^{-\frac{1}{2}}\ \mathrm{d}s + M_2 \int_{t_0}^t \Delta_s^{-1}\ \mathrm{d}s \\ &= M_0 + \frac{2}{\underline{c}} \gamma_k + M_1 \int_{t_0}^t \sqrt{\xi_s} \Delta_s^{-\frac{1}{2}}\ \mathrm{d}s + M_2 \ln\left(\frac{T - t_0}{T - t} \right), \end{aligned}$$

where $M_0 = \frac{2V(t_0, Y_{t_0})}{\underline{c}}$, $M_1 = \frac{2c\bar{c}}{\underline{c}}$ and $M_2 = \frac{\bar{c}\bar{c}'}{2\underline{c}}$.
Thus, applying the Gronwall-type inequality in Lemma 3.28, we get

$$\begin{aligned} \sqrt{\xi_t} &\leq \sqrt{M_0 + \frac{2}{\underline{c}} \gamma_k + M_2 \int_{t_0}^t \Delta_s^{-1}\ \mathrm{d}s} + \frac{M_1}{2} \int_{t_0}^t \Delta_s^{-\frac{1}{2}}\ \mathrm{d}s \\ &= \sqrt{M_0 + \frac{2}{\underline{c}} \gamma_k + M_2 \ln\left((T - t_0)/(T - t) \right)} + M_1 \left[\sqrt{\Delta_{t_0}} - \sqrt{\Delta_t} \right]. \end{aligned} \tag{3.43}$$

Now, for any fixed $\varepsilon > 0$, consider the sequence $\gamma_k = \ln(k^{1+\varepsilon})$. Then $(\gamma_k)_k$ satisfies (3.40). Moreover, for any $t \in [t_k, t_{k+1}]$, it holds

$$\frac{\gamma_k}{\ln\left(1/(T - t)\right)} \leq \frac{\gamma_k}{\ln\left(1/(T - t_k)\right)} = \frac{\ln(k^{1+\varepsilon})}{\ln(k)} = (1 + \varepsilon).$$

Hence, dividing both sides in (3.43) by $\sqrt{\ln\left(1/(T-t)\right)}$, considering $t \in [t_k, t_{k+1}]$ and taking the limit in k gives that

$$\limsup_{t \uparrow T} \frac{|y - L_{\Delta_t} X_t^\circ|}{\sqrt{(T - t) \ln\left(1/(T - t)\right)}} \leq \sqrt{\frac{2}{\underline{c}}(1 + \varepsilon) + \frac{\bar{c}\bar{c}'}{2\underline{c}}}.$$

The claim follows from noting that $\varepsilon > 0$ was chosen arbitrarily. □

The following lemma was used in the proof of Theorem 3.9.

Lemma 3.23. *Under Assumption 3.1, there exists some constant* $\bar{c}'$ *such that*

$$\operatorname{tr}[L_t Q L_t^*] \leq \bar{c}', \quad t \in [0,T]. \tag{3.44}$$

Proof. Since H is reflexive, it is well known that the family of operators $(S_t^*)_{t\geq 0}$ is a strongly continuous semigroup on H. In particular, $(L_t^*)_{t\geq 0}$ is strongly continuous. Hence, denoting by $(b_j)_j$ the standard basis in $\mathbb{R}^k$, the mapping

$$t \mapsto \operatorname{tr}[L_t Q L_t^*] = \sum_{j=1}^k \langle L_t Q L_t^* b_j, b_j\rangle = \sum_{j=1}^k \|Q^{\frac{1}{2}} L_t^* b_j\|^2$$

is continuous and thus (3.44) follows. □

3.7 Proof of Theorem 3.11

Proof. (of Theorem 3.11)
Let $(\Phi_t(X^\circ))_{t<T}$ be the density process defined by

$$\Phi_t(X^\circ) = \frac{\mathrm{d}\mathbb{P}_t^\star}{\mathrm{d}\mathbb{P}_t^\circ}(X^\circ) = \frac{h(t,X_t^\circ)}{g(t,X_t^\circ)}\frac{g(0,x_0)}{h(0,x_0)}\Psi_t(X^\circ).$$

Following Theorem 3.9, the random variable $\Psi_T(X^\circ)$ is $\mathbb{P}^\circ$-a.s. finite and hence $\Psi_t(X^\circ) \to \Psi_T(X^\circ)$ $\mathbb{P}^\circ$-a.s. Moreover, from Assumption 3.10, it follows that $\lim_{t\to T} h(t,X_t^\circ)/g(t,X_t^\circ) = 1$ $\mathbb{P}^\circ$-a.s. and thus, in total,

$$\lim_{t\uparrow T}\Phi_t(X^\circ) = \Phi_T(X^\circ) \quad \mathbb{P}^\circ\text{-a.s.} \tag{3.45}$$

We will show that this convergence holds in $L^1(\mathbb{P}^\circ)$. Since $(\Phi_t(X^\circ))_{t<T}$ is a martingale under $\mathbb{P}^\circ$, the convergence in $L^1(\mathbb{P}^\circ)$ then implies that

$$\Phi_t(X^\circ) = \mathbb{E}^\circ[\Phi_T(X^\circ) \mid \mathcal{F}_t] \quad \mathbb{P}^\circ\text{-a.s.}$$

and hence that $\mathbb{P}^\star$ is absolutely continuous with respect to $\mathbb{P}^\circ$ on $\mathcal{F}_T$ with $\mathrm{d}\mathbb{P}^\star = \Phi_T(X^\circ)\,\mathrm{d}\mathbb{P}^\circ$.
For this, let $r(t) := \sqrt{(T-t)\ln(1/(T-t))}$ and define for any $m\geq 1$ the stopping time

$$\sigma_m = T\wedge\inf\{t\in[0,T] : \|y - L_{T-t}X_t\| \geq m\, r(t)\}. \tag{3.46}$$

Following Equation (3.11), it holds for any $\mathcal{F}_t$-measurable random variable f_t that

$$\frac{g(0,x_0)}{h(0,x_0)}\mathbb{E}^\circ\left[\Psi_t(X^\circ)\ f_t\right] = \mathbb{E}^\star\left[\frac{g(t,X_t^\star)}{h(t,X_t^\star)}\ f_t\right].$$

Hence, setting $f_t = \mathbb{1}_{\{t\leq\sigma_m\}}$, we get

$$\frac{g(0,x_0)}{h(0,x_0)}\mathbb{E}^\circ\left[\Psi_t(X^\circ)\mathbb{1}_{\{t\leq\sigma_m\}}\right] = \mathbb{E}^\star\left[\frac{g(t,X_t^\star)}{h(t,X_t^\star)}\mathbb{1}_{\{t\leq\sigma_m\}}\right]. \tag{3.47}$$

We proceed by taking $\lim_{m\to\infty}\lim_{t\uparrow T}$ on both sides of (3.47). Starting with the left-hand side, it follows from the dominated convergence theorem that

$$\lim_{t\uparrow T}\mathbb{E}^\circ\left[\Psi_t(X^\circ)\mathbb{1}_{\{t\leq\sigma_m\}}\right] = \mathbb{E}^\circ\left[\Psi_T(X^\circ)\mathbb{1}_{\{T\leq\sigma_m\}}\right]$$

Here, we use that, following Lemma 3.24, $\sup_t \Psi_t(X^\circ)\mathbb{1}_{\{t\leq\sigma_m\}}$ is bounded for any fixed $m\in\mathbb{N}$. Furthermore, by the definition of σ_m and Theorem 3.9, it holds that $\lim_m \mathbb{1}_{\{T\leq\sigma_m\}} = \lim_m \mathbb{1}_{\{T=\sigma_m\}} \uparrow 1$ $\mathbb{P}^\circ$-a.e. Hence, by the previous display and monotone convergence, it holds that

$$\lim_{m\to\infty}\lim_{t\uparrow T}\mathbb{E}^\circ\left[\Psi_t(X^\circ)\mathbb{1}_{\{t\leq\sigma_m\}}\right] = \mathbb{E}^\circ\left[\Psi_T(X^\circ)\right]. \tag{3.48}$$

It remains to take the limit on the right-hand side of (3.47). Applying Proposition 3.4 in the first line, we have

$$\begin{aligned}\mathbb{E}^\star\left[\frac{g(t,X_t^\star)}{h(t,X_t^\star)}\mathbb{1}_{\{t\le\sigma_m\}}\right] &= \mathbb{E}\left[\frac{g(t,X_t)}{h(0,x_0)}\mathbb{1}_{\{t\le\sigma_m\}}\right]\\ &= \mathbb{E}\left[\frac{g(t,X_t)}{h(0,x_0)}\right]-\mathbb{E}\left[\frac{g(t,X_t)}{h(0,x_0)}\mathbb{1}_{\{t>\sigma_m\}}\right].\end{aligned}$$

From Lemma 3.25, it follows that $\lim_{m\to\infty}\lim_{t\uparrow T}\mathbb{E}\left[g(t,X_t)/h(0,x_0)\mathbb{1}_{\{t>\sigma_m\}}\right] = 0$. Moreover, following Assumption 3.10, it holds that

$$\begin{aligned}\lambda_2^{-1}(T-t)=\lambda_2^{-1}(T-t)\ \mathbb{E}\left[\frac{h(t,X_t)}{h(0,x_0)}\right] &\le \mathbb{E}\left[\frac{g(t,X_t)}{h(0,x_0)}\right]\\ &\le \lambda_1^{-1}(T-t)\ \mathbb{E}\left[\frac{h(t,X_t)}{h(0,x_0)}\right]=\lambda_1^{-1}(T-t),\end{aligned}$$

which shows that $\lim_{t\uparrow T}\mathbb{E}\left[\frac{g(t,X_t)}{h(0,x_0)}\right]=1$. We conclude that

$$\lim_{m\to\infty}\lim_{t\uparrow T}\mathbb{E}^\star\left[\frac{g(t,X_t^\star)}{h(t,X_t^\star)}\mathbb{1}_{\{t\le\sigma_m\}}\right]=1.$$

Jointly with (3.47) and (3.48), this shows that $\mathbb{E}^\circ[\Phi_T(X^\circ)]=1$.
To conclude the proof, note that $\mathbb{E}^\circ[\Phi_t(X^\circ)] = 1$ for all $t<T$. Hence, by Scheffé's lemma and (3.45), this implies that $\lim_{t\uparrow}\Phi_t(X^\circ)=\Phi_T(X^\circ)$ in $L^1(\mathbb{P}^\circ)$. □

3.7.1 Supplementary lemmas in the proof of Theorem 3.11

Throughout this section we work under the Assumptions of Theorem 3.11.

Lemma 3.24. *Let σ_m be defined as in (3.46). Then there exists a constant $K>0$ such that*

$$\Psi_t(X^\circ)\mathbb{1}_{\{t\le\sigma_m\}}\le\exp(Km)\quad \mathbb{P}^\circ\text{-}a.s.$$

Proof. Per definition of σ_m, it holds on the event $\{t\le\sigma_m\}$ that

$$|y-L_{T-t}X_t^\circ|\le m\, r(t),$$

with $r(t)=\sqrt{(T-t)\ln(1/(T-t))}$. Hence, plugging in $G(s,x)$ as given in (3.10), it follows on $\{t\le\sigma_m\}$ for any $s\le t$ that

$$|G(s,X_s^\circ)|=|L_{T-s}^*R_{T-s}^{-1}(y-L_{T-s}X_s^\circ)|\le\tilde c\,\bar c\,m(T-s)^{-1/2}\sqrt{\ln(1/(T-s))},\qquad(3.49)$$

where we use the bound $\|R_{T-s}^{-1}\|\le\bar c(T-s)^{-1}$ of Assumption 3.8 and the fact that $\|L_{T-s}^*\|\le\tilde c$ for some $\tilde c>0$. In total, using the boundedness of $|F(s,x)|\le C_F$ and (3.49), we get

$$\begin{aligned}\Psi_t(X^\circ)\mathbb{1}_{\{t\le\sigma_m\}} &= \exp\left(\int_0^t\left\langle F(s,X_s^\circ),G(s,X_s^\circ)\right\rangle\mathrm{d}s\right)\mathbb{1}_{\{t\le\sigma_m\}}\\ &\le\exp\left(\int_0^t\left|\left\langle F(s,X_s^\circ),G(s,X_s^\circ)\right\rangle\right|\mathrm{d}s\right)\mathbb{1}_{\{t\le\sigma_m\}}\\ &\le\exp\left(\int_0^t C_F\tilde c\,\bar c\,m(T-s)^{-1/2}\sqrt{\ln(1/(T-s))}\,\mathrm{d}s\right)\mathbb{1}_{\{t\le\sigma_m\}}\quad\mathbb{P}^\circ\text{-a.s.}\end{aligned}$$

Noting that the term in the exponential is integrable over $[0,T]$, the claim follows with $K=C_F\tilde c\,\bar c\int_0^T(T-s)^{-1/2}\sqrt{\ln(1/(T-s))}\,\mathrm{d}s$. □

Lemma 3.25. *Let σ_m be defined as in (3.46). Then*

$$\lim_{m\to\infty}\lim_{t\uparrow T}\mathbb{E}[g(t,X_t)\mathbb{1}_{\{t>\sigma_m\}}]=0.$$

Proof. Just as in the proof of Bierkens et al. [16], Lemma 6.4, it can be shown that

$$\lim_{m\to\infty}\lim_{t\uparrow T}\mathbb{E}\left[g(\sigma_m,X_{\sigma_m})\mathbb{1}_{\{t>\sigma_m\}}\right]=0. \tag{3.50}$$

The claim then follows from the fact that

$$\begin{aligned}\mathbb{E}[g(t,X_t)\mathbb{1}_{\{t>\sigma_m\}}] &\le C\ \mathbb{E}[h(t,X_t)\mathbb{1}_{\{t>\sigma_m\}}]\\ &= C\,\mathbb{E}[\,\mathbb{E}[h(t,X_t)\mathbb{1}_{\{t>\sigma_m\}}\mid\mathcal{F}_{\sigma_m}]]\\ &= C\,\mathbb{E}[h(\sigma_m,X_{\sigma_m})\mathbb{1}_{\{t>\sigma_m\}}]\\ &\le C^2\mathbb{E}[g(\sigma_m,X_{\sigma_m})\mathbb{1}_{\{t>\sigma_m\}}].\end{aligned}$$

Here, we used that, following Assumption 3.10, $g(t,x)\le C_1h(t,x)$ and $h(t,x)\le C_2g(t,x)$ for all $t<T, x\in H$ with $C_1:=\sup_{t\in[0,T]}\lambda_1^{-1}(t)$ and $C_2:=\sup_{t\in[0,T]}\lambda_2(t)$. □

3.7.2 Proof of Lemma 3.13

Proof. Recall that $h(t,x)=\rho_X(t,x;T,y)$ and $g(t,x)=\rho_Z(t,x;T,y)$ are the transition densities of $LX_T\mid X_t=x$ and $LZ_T\mid Z_t=x$. Denote by $X^{t,x}$ and $Z^{t,x}$ the processes X and Z initiated at $X_t=x$ and $Z_t=x$ respectively.
We start by showing the second inequality in (3.17). By the Girsanov theorem and the abstract Bayes formula we have, for any bounded and measurable functional f, that

$$\mathbb{E}[f(X^{t,x})\mid LX_T^{t,x}=y]=\frac{\rho_Z(t,x;T,y)}{\rho_X(t,x;T,y)}\mathbb{E}\left[f(Z^{t,x})\frac{\mathrm{d}\mathcal{L}(X^{t,x})}{\mathrm{d}\mathcal{L}(Z^{t,x})}(Z^{t,x})\mid LZ_T^{t,x}=y\right],$$

where

$$\frac{\mathrm{d}\mathcal{L}(X^{t,x})}{\mathrm{d}\mathcal{L}(Z^{t,x})}(Z)=\exp\left(\int_t^T\langle\tilde{F}(s,Z_s^{t,x}),\mathrm{d}W_s\rangle-\frac{1}{2}\int_t^T\|\tilde{F}(s,Z_s^{t,x})\|^2\ \mathrm{d}s\right)\quad\mathbb{P}\text{-a.s.}$$

is the Girsanov likelihood between $X^{t,x}$ and $Z^{t,x}$ on $C([t,T];H)$.
Denote by $(M(u))_{u\ge t}$ the continuous local martingale $M(u)=\int_t^u\langle\tilde{F}(s,Z_s^{t,x}),\mathrm{d}W_s\rangle$ with quadratic variation $[M]_u=\int_t^u\|\tilde{F}(s,Z_s^{t,x})\|^2\ \mathrm{d}s$. Setting $f\equiv 1$ we get that

$$\begin{aligned}\rho_X(t,x;T,y)&=\rho_Z(t,x;T,y)\mathbb{E}\left[\exp\left(M(T)-\frac{1}{2}[M]_T\right)\mid LZ_T^{t,x}=y\right]\\ &\le\rho_Z(t,x;T,y)\mathbb{E}\left[\exp\left(M(T)\right)\mid LZ_T^{t,x}=y\right].\end{aligned}\tag{3.51}$$

From the Dambis-Dubins-Schwarz theorem it follows that there exists a real-valued Wiener process $\bar{W}$ such that

$$M(u)=\bar{W}_{[M]_u}\quad\mathbb{P}\text{-a.s.}, u\ge t.$$

Furthermore, by the boundedness of $\tilde{F}$, we have that

$$[M]_T=\int_t^T\|\tilde{F}(s,Z_s^{t,x})\|^2\ \mathrm{d}s\le C_{\tilde{F}}^2(T-t).$$

Hence it holds that

$$\begin{aligned}\mathbb{E}\left[\exp\left(M(T)\right) \mid LZ_T^{t,x}=y\right] &= \mathbb{E}\left[\exp\left(\bar{W}_{[M]_T}\right) \mid LZ_T^{t,x}=y\right] \\ &\leq \mathbb{E}\left[\exp\left(\sup_{0\leq s\leq C_{\tilde{F}}^2(T-t)} \bar{W}_s\right) \mid LZ_T^{t,x}=y\right] \\ &= \mathbb{E}\left[\exp\left(\sup_{0\leq s\leq C_{\tilde{F}}^2(T-t)} \bar{W}_s\right)\right].\end{aligned} \tag{3.52}$$

Noting that, for any $\xi \geq 0$, $\tilde{W}_\xi := \sup_{0\leq s\leq \xi} \bar{W}_s$ has density

$$f_{\tilde{W}_\xi}(z) = \sqrt{2/(\pi\xi)}\exp(-z^2/(2\xi))\mathbb{1}_{[0,\infty)}(z),$$

it follows with $\xi_t := C_{\tilde{F}}^2(T-t)$ that the term on the right hand side of (3.52) equals

$$\begin{aligned}\sqrt{2/(\pi\xi_t)}\int_0^\infty \exp(z)\exp(-z^2/(2\xi_t))\,\mathrm{d}z &= \lim_{a\to\infty}\left[-\exp(\xi_t/2)\mathrm{erf}\left(\frac{\xi_t - z}{\sqrt{2\xi_t}}\right)\right]_{z=0}^{z=a} \\ &= \exp(\xi_t/2)\left[1+\mathrm{erf}(\sqrt{\xi_t/2})\right].\end{aligned}$$

Here, $\mathrm{erf}(z) = 2/\sqrt{\pi}\int_0^z \exp(-v^2)\,\mathrm{d}v$ denotes the Gaussian error function. This, jointly with (3.51) and (3.52), shows the second inequality in the lemma with

$$\lambda(T-t) = \exp\left(\frac{C_{\tilde{F}}^2(T-t)}{2}\right)\left[1+\mathrm{erf}\left(\sqrt{\frac{C_{\tilde{F}}^2(T-t)}{2}}\right)\right]. \tag{3.53}$$

The first inequality follows by interchanging the roles of ρ_X and ρ_Z in (3.51). By the same arguments, we then get

$$\rho_Z(t,x;T,y) \leq \rho_X(t,x;T,y)\lambda(T-t),$$

which shows (3.17).

□

3.8 Discussion and future work

In this chapter, we introduced novel methodology to sample from the infinite-dimensional diffusion bridge $X^\star$, defined as the mild solution to a stochastic partial differential equation conditioned on a finite-dimensional, linear transformation $y = LX_T$ of the state X_T for some $T > 0$.

For this, we introduced the guided process X° that mimics the dynamics of the original process while being forced to the conditioning state y. The key aspect of our results is the absolute continuity of its law $\mathcal{L}^\circ$ with respect to the intractable law $\mathcal{L}^\star$ of the diffusion bridge. This enabled the definition of Algorithm 1, a Metropolis-Hastings sampler targeting $\mathcal{L}^\star$ whose performance was illustrated in numerical examples for stochastic reaction-diffusion equations.

Fully observed states One problem that we have not addressed is the case that L equals the identity operator, i.e. when one 'fully' observes the state $y = X_T$ instead of a finite-dimensional transformation thereof. Several challenges arise in this scenario. Firstly, the natural candidate for the h-function is given by $h(t,x) = p(t,x;T,y)$, where p is the transition density of X with respect to some suitable reference measure ν on $(H, \mathcal{B}(H))$. The existence of such transition densities is not necessarily given when H is infinite-dimensional. In some cases, a suitable Gaussian reference measure ν can be constructed as the invariant Ornstein-Uhlenbeck measure. However, this imposes additional assumptions on A and Q.

The more difficult challenge is to show that the guided process X° converges to the conditioning state as $t \uparrow T$, similar to our result of Theorem 3.9. Heuristically speaking, this proves difficult to show as the convergence must occur at the same rate, simultaneously in all infinite coordinates of X°. In Theorem 3.9, it is the observation operator L that maps y and X° to a finite dimensional subspace, therefore decreasing the degrees of freedom in the components of LX°, that allows us to prove convergence at a suitable rate.
While the case of observing a full state X_T is certainly of mathematical interest, let us remark that the assumption of observing finite dimensional transformations thereof certainly appears to be of more practical relevance.

Multiple observations A natural extension of our work is to consider partial observations $L_i X_{t_i}$ of X at multiple observation times t_i, possibly corrupted by some observation noise η_i. The task of state estimation then branches into the two well-known problems of *filtering* and *smoothing*. For finite-dimensional state space models, many state-of-the-art solutions to the filtering and smoothing problems are *particle-based* solutions, in which states of the latent process X are estimated by a set of weighted samples. Without going into too much detail, we refer to Chopin et al. [27] for an introduction into the rich literature in this field.
A common challenge in particle-based inference is to construct importance sampling distributions for the unobserved latent path that are tractable, mimic the behaviour of the dynamical system and are informed by the data. We hypothesise that the path measure of the guided process derived in this chapter serves as a good starting point into the construction of such distributions for partially observed stochastic PDEs. We will investigate this further in future research.

3.A

Proposition 3.26 (Proposition 3.4 above.). *The mapping h defined in (3.7) satisfies Assumptions (i) and (ii) of Theorem 3.2 with $Kh = 0$. Moreover, the measure $\mathbb{P}^\star$ defined on $\mathcal{F}_T$ by*

$$\mathrm{d}\mathbb{P}^\star_t = \frac{h(t, X_t)}{h(0, x_0)} \,\mathrm{d}\mathbb{P}_t, \quad t < T, \tag{3.54}$$

is such that, for any bounded and measurable function φ and $0 \le t_1 \le ... \le t_n < T$, it holds

$$\mathbb{E}^\star[\varphi(X_{t_1}, ..., X_{t_n})] = \mathbb{E}[\varphi(X_{t_1}, ..., X_{t_n}) \mid LX_T = y]. \tag{3.55}$$

We call the process X under $\mathbb{P}^\star$ the infinite-dimensional diffusion bridge (of X given $LX_T = y$).

Proof. We first show that $Kh = 0$. For this, it suffices to show that h satisfies

$$h(s, x) = \mathbb{E}[h(t + s, X_{t+s}) \mid X_s = x] \tag{3.56}$$

for all $s, t \ge 0$ such that $s + t < T$ and $x \in H$. The claim then follows immediately from the definition of K.
Recall that $h(s, x) = \rho_X(s, x; T, y)$ is defined as the density of $LX_T \mid X_s = x$, evaluated at some fixed $y \in \mathbb{R}^k$. Hence, to show (3.56), let $\mathcal{A} \in \mathcal{B}(\mathbb{R}^d)$ be arbitrary and denote by $\mu_{s,t}(x, \mathcal{A})$ the Markov transition kernel of X. Then indeed for any $s, t \ge 0$ such that

$t+s<T$ it holds that

$$\begin{aligned}\mathbb{P}(LX_T \in \mathcal{A} \mid X_s = x) &= \int_H \mathbb{P}(LX_T \in \mathcal{A} \mid X_s = x, X_{s+t} = z)\, \mu_{s,s+t}(x, \mathrm{d}z) \\ &= \int_H \mathbb{P}(LX_T \in \mathcal{A} \mid X_{s+t} = z)\, \mu_{s,s+t}(x, \mathrm{d}z) \\ &= \int_H \int_{\mathcal{A}} \rho_X(s+t, z; T, y)\ \mathrm{d}y\ \mu_{s,s+t}(x, \mathrm{d}z) \\ &= \int_{\mathcal{A}} \mathbb{E}\left[\rho_X(s+t, X_{s+t}; T, y) \mid X_s = x\right] \mathrm{d}y.\end{aligned}$$

Since $\mathcal{A}$ was arbitrary, this shows (3.56) for almost every $y \in \mathbb{R}^k$.
To show the second claim (3.55), we write, for any $y \in \mathbb{R}^k$ with slight abuse of notation, $\mathbb{E}^y\left[\varphi(X_t)\right]$ for the change of measure $\mathbb{P}^\star$ as defined in (3.54) with choice $h(t,x) = \rho_X(t, x; T, y)$. We show that $\mathbb{E}^y\left[\varphi(X_t)\right]$ is a version of $\mathbb{E}\left[\varphi(X_t) \mid LX_T = y\right]$. For this, it suffices to show that

$$\mathbb{E}\left[\mathbb{E}^{LX_T}\left[\varphi(X_t)\right] \mathbb{1}_{\{LX_T \in \mathcal{A}\}}\right] = \mathbb{E}\left[\varphi(X_t)\mathbb{1}_{\{LX_T \in \mathcal{A}\}}\right] \tag{3.57}$$

for any $\mathcal{A} \in \mathcal{B}(\mathbb{R}^d)$. Indeed we have for any $t \in [0,T)$ that

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}^{LX_T}\left[\varphi(X_t)\right] \mathbb{1}_{\{LX_T \in \mathcal{A}\}}\right] &= \int_{\mathcal{A}} \mathbb{E}^y\left[\varphi(X_t)\right] \rho_X(0, x_0; T, y)\, \mathrm{d}y \\ &= \int_{\mathcal{A}} \mathbb{E}\left[\varphi(X_t) \frac{\rho_X(t, X_t; T, y)}{\rho_X(0, x_0; T, y)}\right] \rho_X(0, x_0; T, y)\, \mathrm{d}y \\ &= \mathbb{E}\left[\varphi(X_t) \int_{\mathcal{A}} \rho_X(t, X_t; T, y)\, \mathrm{d}y\right] \\ &= \mathbb{E}\left[\varphi(X_t) \mathbb{E}\left[\mathbb{1}_{\{LX_T \in \mathcal{A}\}} \mid X_t\right]\right] \\ &= \mathbb{E}\left[\varphi(X_t)\mathbb{1}_{\{LX_T \in \mathcal{A}\}}\right].\end{aligned}$$

Here we use the definition of $\rho_X(0, x_0; T, y)$ in the first line, the definition of $\mathbb{E}^y\left[\varphi(X_t)\right]$ in the second line, Fubini's theorem in the third step and the law of total expectation in the last step. The claim now follows from a standard cylindrical argument, see e.g. Ethier and Kurtz [51], Proposition 4.1.6. □

PROPOSITION 3.27 (Proposition 3.6 above.). *The mapping g defined in* (3.7) *satisfies Assumptions (i) to (iii) of Theorem 3.2 with $Kg(t,x) = \langle F(t,x), \mathrm{D}_x g(t,x)\rangle$. In particular, there exists a unique measure $\mathbb{P}^\circ$ on $\mathcal{F}_T$, defined by*

$$\mathrm{d}\mathbb{P}^\circ_t = \frac{g(t, X_t)}{g(0, x_0)} \exp\left(-\int_0^t \langle F(s, X_s), \mathrm{D}_x \log g(s, X_s)\rangle\ \mathrm{d}s\right) \mathrm{d}\mathbb{P}_t, \quad t < T,$$

such that X under $\mathbb{P}^\circ$ is the unique mild solution to the SPDE

$$\mathrm{d}X^\circ_t = \left[AX^\circ_t + F(t, X^\circ_t) + Q\,\mathrm{D}_x \log g(t, X^\circ_t)\right] \mathrm{d}t + Q^{\frac{1}{2}}\, \mathrm{d}W^\circ_t, \quad t \in [0,T), \tag{3.58}$$

where $(W^\circ_t)_{t\in[0,T)}$ is a $\mathbb{P}^\circ$-cylindrical Wiener process.

PROOF. We begin by showing that g satisfies Assumption (iii) of Theorem 3.2. By definition of g we have

$$g(t,x) = \frac{1}{\sqrt{(2\pi)^k \det(R_{T-t})}} \exp\left(-\frac{1}{2}|R_{T-t}^{-\frac{1}{2}}(y - L_{T-t}x)|^2\right),$$

from which it follows that g is continuous and bounded on $[0,S]$ for any $S<T$. Moreover, g is Fréchet differentiable in x with bounded derivative

$$\mathrm{D}_x g(t,x) = g(t,x)\left(L^*_{T-t} R^{-1}_{T-t}(y - L_{T-t}x)\right).$$

This proves Assumption (iii). To show the first assumption, denote by $\tilde{K}$ the infinitesimal generator of the Ornstein-Uhlenbeck process, defined as in (3.5) by substituting T with the transition semigroup of Z. By the Fréchet differentiability of g, it follows from Manca [82], Theorem 4.1, that

$$(Kg)(t,x) = \tilde{K}g(t,x) + \langle F(t,x), \mathrm{D}_x g(t,x)\rangle.$$

Reiterating the arguments in the proof of Proposition 3.4 gives that $\tilde{K}g = 0$. Hence, it follows that

$$(g^{-1}Kg)(t,x) = \langle F(t,x), \mathrm{D}_x \log g(t,x)\rangle,$$

which shows Assumption (i) and the existence of the continuous local martingale $(E^g_t)_{t<T}$. Lastly, to show that E^g is a martingale, it suffices to note that

$$\mathrm{D}_x \log g(t,x) = L^*_{T-t} R^{-1}_{T-t}(y - L_{T-t}x)$$

is Lipschitz in x, uniformly in $t \in [0,S]$ for any $S < T$. The martingale property then follows from Pieper-Sethmacher et al. [94], Lemma 3.6. □

3.B

Lemma 3.28. *Let $t \mapsto \zeta(t)$ be nonnegative and continuously differentiable on $[t_0, t_1)$ and let $t \mapsto f(t)$ be nonnegative and continuous on $[t_0, t_1)$. Assume that $t \mapsto u(t)$ is a nonnegative and continuous function on $[t_0, t_1)$ such that*

$$u(t) \leq \zeta(t) + \int_{t_0}^{t_1} f(s)\sqrt{u(s)}\,\mathrm{d}s.$$

Then $u(t)$ satisfies

$$u(t) \leq \left(\sqrt{\zeta(t_0) + \int_{t_0}^{t} |\zeta'(s)|\,\mathrm{d}s} + \frac{1}{2}\int_{t_0}^{t} f(s)\,\mathrm{d}s\right)^2, \quad t \in [t_0, t_1).$$

Proof. This follows from Agarwal et al. [1], Theorem 2.1. In their notation we have $n = 1, w_1(x) = \sqrt{x}, W_1(x) = 2\sqrt{x}$. □

3.C

We provide additional numerical evidence that the results in Theorem 3.11 hold true in the setting of Example 3.21, even when Assumption 3.1(iv) and the condition of Lemma 3.13 are not met. Assume that Equation (3.16) holds true. It then follows that

$$\rho_X(0, x_0; T, y) = \rho_Z(0, x_0; T, y)\mathbb{E}^\circ[\Psi_T(X^{\circ,y})]. \tag{3.59}$$

Here, to underline the dependence on y, we add a superscript y to the guided process X° with end state $LX^\circ_T = y$. Furthermore, we denote the corresponding guiding term G° of $X^{\circ,y}$ by G^y.
Fix x_0, T and denote by $\pi(y) = \rho_X(0, x_0; T, y)$ the density of $LX_T \mid X_0 = x_0$. It then follows from (3.59) that, for samples $\mathbb{X}^{\circ,y} = \{X^{\circ,y}_i, i = 1, ..., n\}$, the estimator

$$\hat{\pi}(y, \mathbb{X}^{\circ,y}) = \rho_Z(0, x_0; T, y)\frac{1}{n}\sum_{i=1}^{n}\Psi_T(X^{\circ,y}_i) \tag{3.60}$$

is an unbiased estimator of the otherwise intractable density $\pi(y)$. This allows us to numerically sample from the distribution of LX_T in two ways:

1. By approximating the solution to the Allen-Cahn equation (3.33) by a numerical SPDE solver and evaluating LX_T.

2. By sampling from $\mathcal{L}(LX_T)$ via an MCMC scheme based on the unbiased estimator $\hat{\pi}(y, \mathbb{X}^\circ)$.

For the numerical solver we use a spectral Galerkin approximation and a semi-implicit Euler-Maruyama scheme to approximate the resulting SODE. For the second option we implement the correlated pseudomarginal (CPM) sampler as introduced in Deligiannidis et al. [43]. The proposals in y are simple random walk proposals. For the proposals in the latent variable X° we use the same pCN scheme as introduced in Algorithm 1. The complete algorithm is outlined in Algorithm 2.

Algorithm 2: CPM Sampler of LX_T

Input: SPDE Parameters A, F, Q and x_0, observation operator L, gridded domain $D \times [0, T]$, iterations N, step size β, step size $\rho \in [0, 1)$
Output: Samples $(y_i)_{i=0}^N$ of LX_T.
Initialize: Draw $y_0 \sim \mathcal{N}(0, \beta^2 I)$, draw a Wiener process W and compute $X^{\circ,y_0} = \text{solve}(A, F + QG^{y_0}, Q, W)$;
for $i = 0... N-1$ **do**
- **Proposal**
  - (i) Draw $v \sim \mathcal{N}(0, I)$ and set $y^\circ = y_i + \beta v$;
  - (ii) Draw a Wiener process V and set $W^\circ = \sqrt{1-\rho^2}W + \rho V$;
  - (iii) Compute $X^{\circ,y^\circ} = \text{solve}(A, F + QG^{y^\circ}, Q, W^\circ)$;
- **Update**
  - (i) Compute $M = \min\left(1, \dfrac{\hat{\pi}(y^\circ, X^{\circ,y^\circ})}{\hat{\pi}(y_i, X^{\circ,y_i})}\right)$;
  - (ii) Draw $U \sim \text{Unif}(0, 1)$;
  - **if** $U < M$ **then**
    - Set $y_{i+1} = y^\circ$, $W = W^\circ$ and $X^{\circ,y} = X^{\circ,y^\circ}$;
  - **else**
    - Set $y_{i+1} = y_i$.

We employ Algorithm 2 for the Allen-Cahn equation (3.33) with parametrisation

$$[\eta, \zeta, \sigma_0, \rho, \nu] = [1 \times 10^{-2}, 10, 10^7, 3 \times 10^{-6}, 1] \tag{3.61}$$

and initial value $x_0(\xi) = 0.5\sin(4\xi), \xi \in [0, \pi]$. As observation operator we take $L = P_4$, in which case the algorithm returns samples of the first four spectral modes of the Allen-Cahn equation. The chain is run with step size $\beta = 0.03$ and $\rho = 0.1$ for $N = 100\ 000$ iterations of which we discard the first 50 000 as burn-in. The resulting acceptance percentage equals 24%.

To compare the chain output with the empirical distribution of LX_T, we numerically solve the SPDE (3.33) and evaluate LX_T for a total of 10 000 samples. Figure 3.C.1 shows a quantile-quantile plot of the two sample sets for the first and fourth spectral modes of X_T. It indicates that Algorithm 2 correctly samples from the target distribution of LX_T, providing further evidence that Equation (3.60) and hence the results of Theorem 3.11 remain valid, even if Assumption 3.1(iv), and possibly Assumption 3.10, are not satisfied.

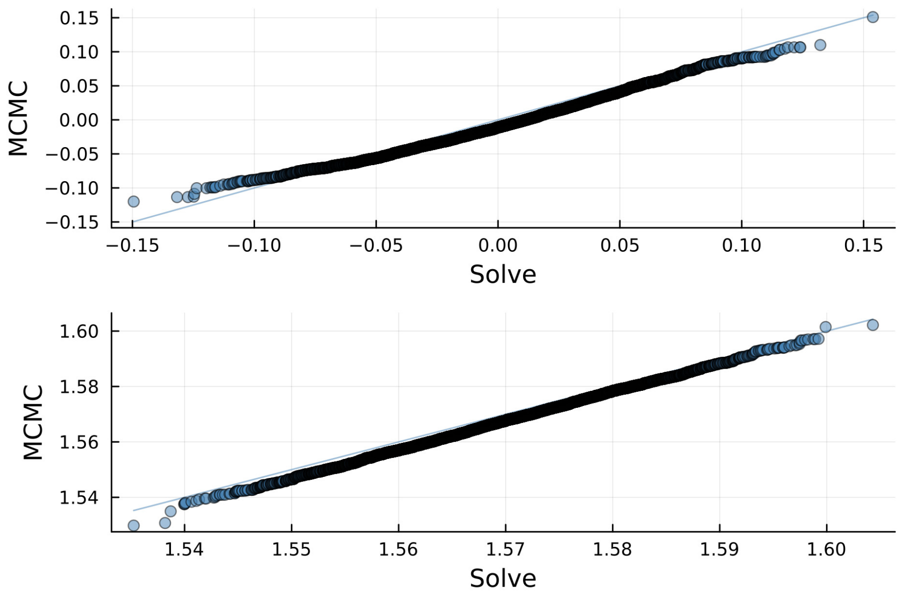


Figure 3.C.1. Quantile-quantile plot of samples of the first and fourth spectral modes of LX_T. Solve: samples obtained by approximating the solution to the Allen-Cahn equation (3.33) by a numerical SPDE solver and evaluating LX_T. MCMC: samples obtained by Algorithm 2, with the first 50 000 out of 100 000 samples discarded as burn-in. Top: first spectral mode. Bottom: fourth spectral mode.

Chapter 4

Guided filtering and smoothing for infinite-dimensional diffusions

This chapter presents novel methodology for the filtering and smoothing problems of an infinite-dimensional diffusion process X, observed through a finite-dimensional representation at discrete points in time. At the heart of our proposed methodology lies the construction of a path measure, termed the guided distribution of X, that is absolutely continuous with respect to the law of X, conditioned on the observations. We show that this distribution can be incorporated as a proposal measure for both sequential Monte Carlo as well as Markov Chain Monte Carlo schemes to tackle the filtering and smoothing problems respectively. In the offline setting, we extend our approach to include parameter estimation of unknown model parameters. The proposed methodology is numerically illustrated in a case study for the stochastic Amari equation.

4.1 Introduction

Consider an *infinite-dimensional diffusion* equation of the form

$$\begin{cases} \mathrm{d}X_t & = [AX_t + F(t, X_t)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t, \quad t \geq 0, \\ X_0 & \sim \mu_0. \end{cases} \tag{4.1}$$

Here, A denotes the generator of a strongly continuous semigroup $(S_t)_{t\geq 0}$ on a Hilbert space H, whereas F is a Lipschitz continuous nonlinearity and Q a symmetric, positive operator of trace class on H. The process $(W_t)_{t\geq 0}$ is a cylindrical Wiener process on H, defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and μ_0 denotes a Borel measure on H. Infinite-dimensional diffusions like that of Equation (4.1) arise, for example, in the context of semilinear stochastic partial differential equations or stochastic field equations.

Throughout this chapter, we assume that Equation (4.1) admits a unique mild solution X satisfying

$$X_t = S_t X_0 + \int_0^t S_{t-s} F(s, X_s)\,\mathrm{d}s + \int_0^t S_{t-s} Q^{\frac{1}{2}}\,\mathrm{d}W_s. \tag{4.2}$$

The process X is an H-valued, predictable Markov process with transition kernel $(\mu_{s,t})_{s\leq t}$ defined by $\mu_{s,t}(x, B) := \mathbb{P}(X_t \in B \mid X_s = x)$ for all Borel sets $B \subset H$ and $x \in H$.

Suppose at discrete observation times $0 < t_1 < t_2 < \ldots < t_n = T$, we are given *finite-dimensional observations* y_i through realisations $Y_i = y_i$ of

$$Y_i \mid X_{t_i} \sim k_i(X_{t_i}, \cdot), \quad i = 1, \ldots, n. \tag{4.3}$$

Here, $k_i(X_{t_i}, \cdot)$ denotes a conditional density on $\mathbb{R}^{m_i}$, where m_i is the dimension of the observation at time t_i. For the sake of convenience, we assume from now on that

$$Y_i \mid X_{t_i} \sim f(\cdot; LX_{t_i}, \Sigma), \quad i = 1, \ldots, n, \tag{4.4}$$

where $L : H \to \mathbb{R}^m$ is a bounded linear operator, termed the *observation operator*, Σ is a positive-definite matrix and $f(\cdot; Lx, \Sigma)$ is the Gaussian density with mean Lx and covariance Σ. However, our methods generalise to non-Gaussian observation densities and we will discuss how to adapt to such cases towards the end of the chapter.
Based on the described setup, we address three estimation problems. Firstly, we are concerned with *state estimation* of the unobserved, latent path X. Here, we consider both the *online* problem of estimating the *filtering* distribution

$$\mathbb{P}(X_{t_i} \in B \mid Y_1, \ldots, Y_i), \quad i = 1, \ldots, n, \tag{4.5}$$

and as well as the *offline* problem of estimating the *smoothing* distribution

$$\mathbb{P}(X_t \in B \mid Y_1, \ldots, Y_n), \quad t \in [0, T], \tag{4.6}$$

for any Borel set $B \subset H$. Moreover, if Equation (4.1) is parametrised by some unknown model parameters $\theta \in \mathbb{R}^p$, we are interested in *parameter estimation* of θ in the offline setting.

4.1.1 Related Work

Although interest in statistical inference for SPDEs has grown in the past decade, it remains an emerging field, in particular concerning nonlinear equations and Bayesian methodology.
The majority of recent advances concern parameter estimation from the frequentist perspective. Inference of a drift parameter for the two-dimensional stochastic Navier-Stokes equation has been considered in Cialenco and Glatt-Holtz [29] and for semi-linear SPDEs in Pasemann and Stannat [93], Altmeyer et al. [3], and Cialenco et al. [31]. Moreover, estimation concerning the nonlinearity F has been treated in the parametric case in Gaudlitz and Reiß [56] and in the non-parametric case in Gaudlitz [55], and Hildebrandt and Trabs [65].
On the other hand, the literature regarding state estimation for partially observed SPDEs and parameter estimation within the Bayesian framework remains limited. Notable exceptions address the filtering problem for SPDEs. In Llopis et al. [78], a guided particle filter, including tempering and resample-move steps, has been introduced to estimate the filtering distribution of a two-dimensional stochastic Navier-Stokes signal. Moreover, a similar approach was used in Lang et al. [75] in the context of filtering for the stochastic rotating shallow water (SRSW) model.
Smoothing and Bayesian parameter inference for infinite-dimensional diffusions in the generality of our setup has, to the best of our knowledge, not been considered previously in the literature. However, in the case of a single noiseless observation, sampling of the infinite-dimensional diffusion bridge has been investigated for strong solutions of Equation (4.1) in Yang et al. [115] and for mild solutions in Pieper-Sethmacher et al. [95].
In contrast, in the setting of finite-dimensional state spaces, smoothing for discretely observed diffusion processes has been studied in Mider et al. [85], Graham et al. [60], and in the more recent work by Stanton and Beskos [107]. In this work, we closely follow the idea of *guiding*, as was first introduced in Schauer et al. [101] for sampling of diffusion bridges and extended to the smoothing problem in Mider et al. [85].

4.1.2 Approach

Our approach relies heavily on the exponential changes of measure studied in Pieper-Sethmacher et al. [94]. Consider for now the case that $n = 1$, i.e. we observe one realisation y of the random variable $Y \mid X_T \sim k(X_T, \cdot)$. In this simplified case, estimating the conditional distribution of X given $Y = y$ solves both the filtering as well as the smoothing problem.
It is well-known that this conditional distribution can be derived by a change of measure known as *Doob's h-transform* as follows. Defining for any fixed $y \in \mathbb{R}^m$ the likelihood function

$$h(t,x) = \int_H k(z,y)\mu_{t,T}(x, \mathrm{d}z), \quad t \in [0,T], x \in H,$$

one shows that the process $(h(t,X_t))_{t\in[0,T]}$ is a non-negative $\mathbb{P}$-martingale and hence defines a unique change of measure $\mathbb{P}^h$ on $\mathcal{F}_T$ by

$$\mathrm{d}\mathbb{P}^h_{|\mathcal{F}_T} = \frac{h(T,X_T)}{C^h}\,\mathrm{d}\mathbb{P}_{|\mathcal{F}_T},$$

with $C^h := \mathbb{E}[h(0,X_0)]$ acting as a normalising constant. Under the new measure $\mathbb{P}^h$, the law $\mathcal{L}^h(X)$ of X is exactly the conditional distribution of X given $Y = y$ in the sense that

$$\mathbb{E}^h[\varphi(X)] = \mathbb{E}[\varphi(X) \mid Y = y]$$

for any bounded and measurable function φ. Moreover, under certain regularity conditions on h, it was shown in Pieper-Sethmacher et al. [94] that the change of measure $\mathbb{P}^h$ is of Girsanov type. The process X under $\mathbb{P}^h$ is then a mild solution to the SPDE

$$\mathrm{d}X_t^h = \left[AX_t^h + F(t,X_t^h) + Q\,\mathrm{D}_x \log h(t,X_t^h)\right]\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t^h, \tag{4.7}$$

where W^h is a $\mathbb{P}^h$-cylindrical Wiener process. Equation (4.7) differs from the original dynamics in Equation (4.1) only by an additional drift term involving the *score* $\mathrm{D}_x \log h(t,x)$ of the likelihood function $x \mapsto h(t,x)$ that steers the process X^h into regions of high probability given the observation $Y = y$. With slight abuse of notation, we write X^h for the process X defined on the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}^h)$ and call it the *conditioned process*. Note that the existence of $\mathbb{P}^h$ and X^h does not rely on the existence of a mild solution to Equation (4.7).

In most cases, the transition kernel μ of X, and hence the function h, is intractable. This renders directly sampling the conditioned process infeasible, even if conditions for the well-posedness of Equation (4.7) are satisfied. A natural approach to overcome this problem is to substitute the intractable function h by a tractable function g that 'approximates' h in some sense. Besides being tractable, the function g should

(i) be informed by the observed data in a way that resembles the original function h,

(ii) be such that there exists a mild solution X^g to the equation

$$\mathrm{d}X_t^g = [AX_t^g + F(t,X_t^g) + Q\,\mathrm{D}_x \log g(t,X_t^g)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t, \tag{4.8}$$

(iii) be such that the laws of X^h and X^g are absolutely continuous in path space with a tractable Radon-Nikodym (R.N.) derivative Φ.

One can then obtain *weighted samples* of the conditioned process X^h by sampling the process X^g and associating with each sample the weight $\Phi(X^g)$. Since X^g is 'guided' by the observation $Y = y$, we refer to it as the *guided process* and its law as the *guided distribution*.

In our approach, we define g by substituting the transition kernel μ in the definition of h with a tractable transition kernel ν of an auxiliary process Z. If Z is an Ornstein-Uhlenbeck process, we show that the function g defines a change of measure on $\mathcal{F}_T$ by

$$\mathrm{d}\mathbb{P}^g_{|\mathcal{F}_T} = \frac{g(T, X_T)}{C^g} \exp\left(-\int_0^T \frac{\mathcal{K}g}{g}(s, X_s)\,\mathrm{d}s\right) \mathrm{d}\mathbb{P}_{|\mathcal{F}_T}$$

such that X under $\mathbb{P}^g$ satisfies Equation (4.8) with respect to a $\mathbb{P}^g$-cylindrical Wiener process W^g. Here, $\mathcal{K}$ denotes the infinitesimal generator of the original, unconditioned process X and C^g is another normalising constant.
The function $x \mapsto g(t, x)$ is to be interpreted as the likelihood function of $Z_t = x$ under the simplified dynamics of Z and therefore satisfies condition (i). Moreover, condition (ii) is ensured by the Gaussian nature of $g(t, x)$ implying a global Lipschitz condition of the score function $\mathrm{D}_x \log g(t, x)$. Lastly, by construction of $\mathbb{P}^h$ and $\mathbb{P}^g$, the laws of X^g and X^h are absolutely continuous in path space, hence satisfying the third condition imposed on g. In Pieper-Sethmacher et al. [95] it was shown that this absolute continuity persists if the observation kernel degenerates to a Dirac measure in Lx, i.e. if $Y \mid X_T \sim \delta_{Lx}(\cdot)$. In other words, the guided distribution remains a valid importance sampling distribution for X^h even under highly informative observations.

Our aim in this chapter is to apply this basic idea of constructing the conditioned process X^h as well as the guided process X^g by a change of measure to derive solutions to both the filtering as well as the smoothing problem.
For the filtering problem, we propose the law of X^g as a natural candidate for the proposal distribution in a *particle filter* (or *sequential Monte Carlo*) approach. The particle filter is a well-established algorithm that approximates the filtering distribution in (4.5) by a set of samples ('particles') $X^{(j)}_{t_i}$ with associated weights $w^{(j)}_i$. At each iteration i, the particles are evolved forward following a proposal distribution $X^{(j)}_{[t_i,t_{i+1}]} \sim \mathbb{Q}(\cdot \mid X^{(j)}_{t_i})$ with the associated weights following the recursive update

$$w^{(j)}_{i+1} \propto w^{(j)}_i\, k(X^{(j)}_{t_{i+1}}, y_{i+1}) \frac{\mathrm{d}\mathbb{P}\left(X^{(j)}_{[t_i,t_{i+1}]} \in \cdot \mid X^{(j)}_{t_i}\right)}{\mathrm{d}\mathbb{Q}(\cdot \mid X^{(j)}_{t_i})} \left(X^{(j)}_{[t_i,t_{i+1}]}\right).$$

It is well-known that the locally optimal proposal distribution $\mathbb{Q}$ is given by the conditional distribution of $X^{(j)}_{[t_i,t_{i+1}]} \mid X^{(j)}_{t_i}, Y_{i+1} = y_{i+1}$ (cf. 10.3, Chopin et al. [27]). This corresponds to the h-transform introduced above with the intractable function

$$h_i(t, x) = \int_H k(z, y_{i+1}) \mu_{t,t_{i+1}}(x, \mathrm{d}z), \quad t \in (t_i, t_{i+1}], x \in H.$$

A natural candidate for a substitute proposal distribution is therefore the guided distribution corresponding to the observation y_{i+1}.

On the other hand, the smoothing problem can be solved after extending the change of measure $\mathbb{P}^h$ to include the complete set of observations $(Y_1, \ldots, Y_n) = (y_1, \ldots, y_n)$ by defining the function

$$h(t, x) := \int k(x_{t_i}, y_i) \left(\prod_{j=i}^{n-1} k(x_{t_{j+1}}, y_{j+1})\, \mu_{t_j,t_{j+1}}(x_{t_j}, \mathrm{d}x_{t_{j+1}})\right) \mu_{t,t_i}(x, \mathrm{d}x_{t_i})$$

for any $t \in (t_{i-1}, t_i]$. In Section 4.2 we show that h defines a change measure $\mathbb{P}^h$ under which the law of X^h is exactly the smoothing distribution (4.6) of X. Substituting again the Markov kernel μ with the kernel ν of an auxiliary process Z leads to a tractable

change of measure $\mathbb{P}^g$ and a corresponding guided process X^g that satisfies yet another SPDE, while maintaining absolute continuity of the laws of X^h and X^g with a tractable Radon-Nikodym derivative Φ.
For certain choices of Z, we show that g can be directly computed on the basis of the recursion known as the *backwards information filter*

$$\begin{aligned} g(t,x) &= \int g(t_i, x_{t_i})\, \nu_{t,t_i}(x, \mathrm{d}x_{t_i}), \quad t \in (t_{i-1}, t_i], \\ g(t_i, x) &= k(x, y_j)\, g^+(t_i, x), \quad i \in \{1, \dots, n\}, \end{aligned} \tag{4.9}$$

with $g^+(t_i, x) := \lim_{\substack{t \downarrow t_i \\ t \neq t_i}} g(t,x)$. This enables drawing weighted samples of X^h which can then be passed onto general Markov chain Monte Carlo or importance sampling schemes to sample from the smoothing distribution of X.

4.1.3 Outline

In Section 4.2 we derive the two changes of measure that define the smoothing and guided distribution of X. In particular, we present two ways of efficiently computing the score $\mathrm{D}_x \log g$ needed to sample from the guiding distribution and evaluate the corresponding weights. The implementation of the guided distribution into a particle filtering scheme is discussed in Section 4.3. Section 4.4 derives a Markov Chain Monte Carlo algorithm that targets the smoothing distribution of X and posterior of model parameters θ. Performance of the proposed methodology is illustrated in Section 4.5 through an example based on the stochastic Amari equation.

4.1.4 Frequently used notation

We denote by H a Hilbert space with inner product $\langle x, y\rangle$ and norm $|x| = \sqrt{\langle x, x\rangle}$. The process X denotes the mild solution to Equation (4.1), defined on the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, with Markov transition kernel $\mu = (\mu_{s,t})_{s\leq t}$. If F is independent of X and instead $F(t, X_t) = a_t$ for some predictable, H-valued process $(a_t)_t$, we write Z in place of X for the Ornstein-Uhlenbeck process that satisfies (4.1) and $\nu = (\nu_{s,t})_{s\leq t}$ for its transition kernel.

The kernel $k_i(x, \cdot)$ defines a conditional density on $\mathbb{R}^{m_i}$ for any $x \in H$. For most of this chapter, we assume $k_i(x, \cdot) = k(x, \cdot)$ to be a Gaussian density on $\mathbb{R}^m$ with mean Lx and covariance matrix Σ, where $L : H \to \mathbb{R}^m$ is a bounded linear observation operator. By $Y_i \in \mathbb{R}^{m_i}$ we denote the observations of X_{t_i} with density $k_i(X_{t_i}, \cdot)$ at observation time t_i, with $Y = (Y_1, \dots Y_n)$ defined as the joint vector of all observations. We refer to individual components of one observation vector $Y_i \in \mathbb{R}^{m_i}$ as measurements. For realisations of Y_i, we write $y_i \in \mathbb{R}^{m_i}$. We shall also write $y_i^+ = (y_i, \dots, y_n)$ for the vector of all non-past observations at time t_i. Moreover, m_i^+ denotes the dimension of y_i^+, i.e. $m_i^+ = \sum_{j=i}^n m_j$. For a function f, $f^+(t) := \lim_{s\downarrow t} f(s)$ is defined as the right-sided limit of f at t whenever existent.

By $C([0,T]; H)$ we denote the space of continuous, H-valued functions endowed with the supremum norm and its Borel algebra. For the laws of X and Z on the Borel algebra of $C([0,T]; H)$ we write $\mathcal{L}(X)$ and $\mathcal{L}(Z)$ respectively. With slight abuse of notation, we shall also write $X_{[t_{i-1}, t_i]}$ for the process $(X_t)_{t\in[t_{i-1}, t_i]}$ and $\mathcal{L}(X_{[t_{i-1}, t_i]})$ for its law on $C([t_{i-1}, t_i]; H)$.
The realised observations y together with the transition kernels μ and ν define the functions h and g. These define the changes of measure $\mathbb{P}^h$ and $\mathbb{P}^g$ under which the law

of X is, respectively, the smoothing and guided distribution given the observations y. We denote these by $\mathcal{L}^h(X)$ and $\mathcal{L}^g(X)$.

4.2 Changes of measure

The aim of this section is to derive the two changes of measure $\mathbb{P}^h$ and $\mathbb{P}^g$ that respectively define the intractable smoothing distribution $\mathcal{L}^h(X)$ and the tractable guided distribution $\mathcal{L}^g(X)$ of X. In particular, the construction will show that

(i) $\mathcal{L}^h(X)$ and $\mathcal{L}^g(X)$ are absolutely continuous with Radon-Nikodym derivative

$$\Phi(X) := \frac{\mathrm{d}\mathcal{L}^h(X)}{\mathrm{d}\mathcal{L}^g(X)}(X) = \frac{C^g}{C^h} \exp\left(\int_0^T \langle \tilde{F}(s, X_s), G(s, X_s)\rangle \,\mathrm{d}s\right),$$

where C^g and C^h are constants and $\tilde{F}(t,x)$ and $G(t,x) = \mathrm{D}_x \log g(t,x)$ are tractable functions,

(ii) the distribution $\mathcal{L}^g(X)$ can be sampled by numerically solving the SPDE

$$\mathrm{d}X_t^g = [AX_t^g + F(t, X_t^g) + QG(t, X_t^g)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t^g.$$

Moreover, we derive two ways of efficiently computing G needed to sample from X^g and evaluate the corresponding weights. The proofs of the results in this section can be found in Appendix 4.A.

4.2.1 The smoothing distribution

We start by constructing the smoothing distribution of X given $Y_1 = y_1, \dots, Y_n = y_n$. We do so by defining a measure $\mathbb{P}^h$ on $\mathcal{F}_T$ such that

$$\mathbb{E}^h\left[\varphi(X)\right] = \mathbb{E}[\varphi(X) \mid Y_1 = y_1, \dots, Y_n = y_n] \tag{4.10}$$

for all bounded and measurable functionals $\varphi : C([0,T], H) \to \mathbb{R}$. Note that for $n = 1$, this includes the locally optimal proposal distribution in the guided particle filter as introduced in Section 4.1.2.

To this end, define the function $h : [0,T] \times H \to \mathbb{R}_+$ for any $t \in (t_{i-1}, t_i]$ by

$$h(t,x) := \int k(x_{t_i}, y_i)\left(\prod_{j=i}^{n-1} k(x_{t_{j+1}}, y_{j+1})\,\mu_{t_j,t_{j+1}}(x_{t_j}, \mathrm{d}x_{t_{j+1}})\right)\mu_{t,t_i}(x, \mathrm{d}x_{t_i}) \tag{4.11}$$

with $h(0,x) := h^+(0,x)$. For the sake of notational clarity, we drop the dependence of h on the observations $(y_1, \dots, y_n)$. However, it is worth pointing out that on any time interval $(t_{i-1}, t_i]$, $h(t,x)$ has a natural interpretation as the likelihood function of $X_t = x$ given all non-past observations $y_i^+ = (y_i, \dots, y_n)$, i.e. in Bayesian notation

$$h(t, x_t) = p(y_i^+ \mid x_t), \quad t \in (t_{i-1}, t_i].$$

A direct consequence of the definition of h is that it satisfies the following recursive relation known as the *backwards information filter.*

PROPOSITION 4.1. *The function $h(t,x)$ defined in (4.11) satisfies*

$$h(t,x) = \int h(t_i, x_{t_i})\,\mu_{t,t_i}(x, \mathrm{d}x_{t_i}), \quad t \in (t_{i-1}, t_i].$$

In particular, h is continuous on any interval (t_{i-1}, t_i) and left continuous at any $t_i, i = 1, \dots, n$. Moreover, at the observation times t_i, it holds that

$$h(t_i, x) = k(x, y_j) \int h(t_{i+1}, x_{t_{i+1}})\,\mu_{t_i,t_{i+1}}(x, \mathrm{d}x_{t_i}) = k(x, y_j)\,h^+(t_i, x).$$

With the choice of h established, define now the process $(E_t^h)_{t\in[0,T]}$ by

$$E_t^h := \frac{1}{C^h}\left(\prod_{j=1}^{i-1} k(X_{t_j}, y_j)\right) h(t, X_t), \quad t \in (t_{i-1}, t_i], \tag{4.12}$$

with $E_0^h := h(0, X_0)/C^h$ and $C^h := \mathbb{E}[h(0, X_0)]$ acting as a normalising constant.

Remark 4.2. At any time $t \in (t_{i-1}, t_i]$, the product $\prod_{j=1}^{i-1} k(x_{t_j}, y_j)$ appearing on the right-hand side of (4.12) is the likelihood of previous states of X at observation times given past observations, whereas $h(t, x_t)$ is the likelihood of $X_t = x_t$ given all non-past observations. Hence, in total, the term

$$\left(\prod_{j=1}^{i-1} k(x_{t_j}, y_j)\right) h(t, x_t, y), \quad t \in (t_{i-1}, t_i],$$

is the likelihood of $(X_{t_1} = x_{t_1}, \dots, X_{t_{i-1}} = x_{t_{i-1}}, X_t = x_t)$ given all observations y_j, $j = 1, \dots, n$.

Proposition 4.3. *The process $(E_t^h)_{t\in[0,T]}$ defined in (4.12) is a non-negative $\mathbb{P}$-martingale with respect to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$.*

Given its martingale property and noting that $E_0^h = 1$, the process E^h defines a measure $\mathbb{P}^h$ on $\mathcal{F}_T$ by

$$\mathrm{d}\mathbb{P}^h_{|\mathcal{F}_T} = E_T^h \,\mathrm{d}\mathbb{P}_{|\mathcal{F}_T}. \tag{4.13}$$

The following theorem establishes that this is indeed the change of measure under which the law of X is the smoothing distribution.

Theorem 4.4.

(i) *The measure $\mathbb{P}^h$ satisfies (4.10).*

(ii) *In addition, if h is Fréchet differentiable in x such that $\mathrm{D}_x h \in C_m((t_{i-1}, t_i); H)$ for all $i = 1 \dots n$, then the process X under $\mathbb{P}^h$ is a mild solution to the SPDE*

$$\begin{cases} \mathrm{d}X_t^h &= \left[AX_t^h + F(t, X_t^h) + Q\,\mathrm{D}_x \log h(t, X_t^h)\right]\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t^h, \\ X_0^h &\sim \mu_0^h, \end{cases} \tag{4.14}$$

where W^h is a $\mathbb{P}^h$-cylindrical Wiener process and $\mu_0^h(B) := \mathbb{P}^h(X_0 \in B)$.

Remark 4.5. The assumption in Theorem 4.4 (ii) is in general hard to verify. However, to sample from the law of X under $\mathbb{P}^h$, the differential form of X^h in Equation (4.14) is not necessary. Instead, weighted samples of the smoothing distribution are obtained by sampling X under the substitute measure $\mathbb{P}^g$ defined in the next subsection.

4.2.2 The guided distribution

To construct the guided distribution, we follow similar steps as in the previous section, replacing μ in the definition of h with a tractable transition kernel ν of an auxiliary process Z.

For this, let Z be the Ornstein-Uhlenbeck process, defined as the unique mild solution to the SPDE

$$\begin{cases} \mathrm{d}Z_t &= [AZ_t + a_t]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t, \quad t \geq 0, \\ Z_0 &\sim \nu_0. \end{cases} \tag{4.15}$$

The H-valued process $(a_t)_{t\geq 0}$ is assumed to be predictable with integrable trajectories and ν_0 denotes a Gaussian measure on H. Let $(\nu_{s,t})_{s\leq t}$ be the transition kernel of Z defined by $\nu_{s,t}(x, B) := \mathbb{P}(Z_t \in B \mid Z_s = x)$ for any $0 \leq s \leq t$, $x \in H$ and Borel set $B \subset H$. It is well-known that $\nu_{s,t}(x,\cdot)$ is a Gaussian measure on H with mean $S_{t-s}x + \int_s^t S_{t-u}a_u \,\mathrm{d}u$ and covariance operator

$$Q_{t-s} := \int_0^{t-s} S_u Q S_u^* \,\mathrm{d}u. \tag{4.16}$$

Following the definition of h in Equation (4.11), we define the mapping g by

$$g(t,x) := \int k(x_{t_i}, y_i) \left(\prod_{j=i}^{n-1} k(x_{t_{j+1}}, y_{j+1})\, \nu_{t_j,t_{j+1}}(x_{t_j}, \mathrm{d}x_{t_{j+1}}) \right) \nu_{t,t_i}(x, \mathrm{d}x_{t_i}) \tag{4.17}$$

for any $t \in (t_{i-1}, t_i]$ with the convention that $g(0,x) := g^+(0,x)$.
The function $g(t,x)$ on $(t_{i-1}, t_i]$ has a natural interpretation as the likelihood of $Z_t = x$ given the vector $(y_i, \dots, y_n)$ of non-past observations $y_i \sim k(Z_{t_i}, \cdot)$ under the simplified dynamics of Equation (4.15). Moreover, just like the h-function, g satisfies the recursion of the backwards information filter

$$\begin{aligned} g(t,x) &= \int g(t_i, x_{t_i})\, \nu_{t,t_i}(x, \mathrm{d}x_{t_i}), \quad t \in (t_{i-1}, t_i], \\ g(t_i, x) &= k(x, y_i)\, g^+(t_i, x), \quad i \in \{1, \dots, n\}, \end{aligned} \tag{4.18}$$

with $g^+(t_i, x) = \int g(t_{i+1}, x_{t_{i+1}})\, \nu_{t_i,t_{i+1}}(x, \mathrm{d}x_{t_i})$.
Contrary to the generally intractable h, the Gaussian nature of $\nu_{s,t}(x,\cdot)$ and $k(x,\cdot)$ enables us to compute the function g using the recursion in (4.18) as shown in the next theorem.

Theorem 4.6. *On any interval $(t_{i-1}, t_i]$, $i = 1, \dots, n$, let $L_t \in L(H, \mathbb{R}^{m_i^+})$ be given by*

$$L_t := \begin{bmatrix} LS_{t_i - t} \\ LS_{t_{i+1}-t} \\ \vdots \\ LS_{t_n - t} \end{bmatrix} \tag{4.19}$$

and let $R_t \in L(\mathbb{R}^{m_i^+})$ and $\alpha_t \in \mathbb{R}^{m_i^+}$ be defined by the backwards recursions

$$\begin{aligned} R_t &:= \begin{cases} \Sigma + LQ_{t_n - t}L^*, & i = n, \\ \begin{bmatrix} \Sigma & 0 \\ 0 & R_{t_i}^+ \end{bmatrix} + L_{t_i} Q_{t_i - t} L_{t_i}^*, & \text{else}, \end{cases} \\ \alpha_t &:= \begin{cases} L\left(\int_t^{t_n} S_{t_n - s} a_s \,\mathrm{d}s \right), & i = n, \\ \begin{bmatrix} 0 \\ \alpha_{t_i}^+ \end{bmatrix} + L_{t_i} \left(\int_t^{t_i} S_{t_i - s} a_s \,\mathrm{d}s \right), & \text{else}. \end{cases} \end{aligned} \tag{4.20}$$

Then, the function $g(t,x)$ is given by

$$g(t,x) = f(y_i^+; L_t x + \alpha_t, R_t), \quad t \in (t_{i-1}, t_i]. \tag{4.21}$$

Following Theorem 4.6, the function g can be expressed on any interval $(t_{i-1}, t_i]$ as the Gaussian density in Equation (4.21). This lets us define a tractable change of measure $\mathbb{P}^g$ on $\mathcal{F}_T$ as follows. Denote by G the score function of the likelihood $x \mapsto g(t,x)$, i.e.

$$G(t,x) := \mathrm{D}_x \log g(t,x) = L_t^* R_t^{-1}(y_i^+ - L_t x - \alpha_t), \quad t \in (t_{i-1}, t_i], x \in H. \tag{4.22}$$

Moreover, define the process $(E_t^g)_{t\in[0,T]}$ by setting $\tilde{F}(t,x) := F(t,x) - a_t$ and

$$E_t^g := \frac{1}{C^g}\left(\prod_{j=1}^{i-1} k(X_{t_j}, y_j)\right) g(t, X_t) \exp\left(-\int_0^t \langle \tilde{F}(s, X_s), G(s, X_s)\rangle \, \mathrm{d}s\right), \quad t \in (t_{i-1}, t_i], \tag{4.23}$$

with $E_0^g := g(0, X_0)/C^g$ and $C^g := \mathbb{E}[g(0, X_0)]$ acting as a normalising constant.
The theorem presented below shows that E^g is a martingale that defines a measure $\mathbb{P}^g$ on $\mathcal{F}_T$ under which the law of X is exactly the guided distribution as introduced in Section 4.1.2.

THEOREM 4.7.

(i) *The process E^g defined in (4.23) is a non-negative $\mathbb{P}$-martingale with respect to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$.*

(ii) *Under the measure $\mathbb{P}^g$ defined on $\mathcal{F}_T$ by $\mathrm{d}\mathbb{P}^g_{|\mathcal{F}_T} = E_T^g \, \mathrm{d}\mathbb{P}_{|\mathcal{F}_T}$, the process X is the unique mild solution to the SPDE*

$$\begin{cases} \mathrm{d}X_t^g &= [AX_t^g + F(t, X_t^g) + QG(t, X_t^g)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t^g, \quad t \in [0,T], \\ X_0^g &\sim \mu_0^g, \end{cases} \tag{4.24}$$

where W^g is a cylindrical Wiener process under $\mathbb{P}^g$.

From the construction of $\mathbb{P}^h$ and $\mathbb{P}^g$ it follows that the smoothing and guided distribution of X are absolutely continuous with Radon-Nikodym derivative given by

$$\Phi(X) = \frac{C^g}{C^h}\Psi(X) \tag{4.25}$$

where $\Psi(X)$ is defined as

$$\Psi(X) := \exp\left(\int_0^T \langle \tilde{F}(s, X_s), G(s, X_s)\rangle \, \mathrm{d}s\right). \tag{4.26}$$

Hence, weighted samples of the smoothing distribution $\mathcal{L}^h(X)$ can be obtained by drawing samples from $\mathcal{L}^g(X)$ and evaluating the weights $\Phi(X)$. As a consequence of Theorem 4.7, the sampling step corresponds to numerically solving Equation (4.24).

REMARK 4.8. Drawing samples from $\mathcal{L}^g(X)$ as well as evaluating the weights $\Phi(X)$ requires computation of the function G defined in Equation (4.22). In practical applications, this is carried out on a gridded domain in both space and time.
In that case, the infinite-dimensional operators A, S_t and Q are typically approximated by $M\times M$-dimensional matrices on the given spatial - or spectral - grid. Likewise, the observation operator L is represented by a $m\times M$-dimensional matrix. Evaluating G based on Theorem 4.6 on a temporal grid of $(t_{i-1}, t_i]$ with N grid points then requires the following operations:

1. The operators L_t are given in closed form and are a simple concatenation of linear compositions of L and the semigroup S. For each observation time t_i, the matrix multiplication $L_t = L_{t_i} S_{t_i - t}$ needs to be carried out for N grid points of $(t_{i-1}, t_i]$. This results in $O(Nm_i^+ M^2)$ operations.
2. To compute R_t, the matrix approximations of $Q_{t_i - t} = \int_0^{t_i - t} S_s Q S_s^* \,\mathrm{d}s$ need to evaluated. If this integral cannot be solved directly, it needs to be numerically integrated with computational complexity of $O(NM^3)$. Additionally, the matrix multiplications $L_{t_i} Q_{t_i - t} L_{t_i}^*$ will need to be carried out, leading to an additional total cost of $O(Nm_i^+ M^2)$. Moreover, evaluation of G will require the matrix inversion of R_t. Due to the block matrix structure of R_t, this is of cost $O((n-i+1)m^3) = O(m_i^+ m^2)$ for each inversion.

3. Lastly, to compute α_t, we need to numerically backwards integrate the term $\int_t^{t_i} S_{t_i-s}a_s \, \mathrm{d}s$. This is of complexity $O(NM^2)$ after discretisation, on top of the cost $O(Nm_i^+ M)$ of evaluating $L_{t_i} \int_t^{t_i} S_{t_i-s}a_s \, \mathrm{d}s$.

Hence, in total, the cost of the operations required to compute G using Theorem 4.6 on a spatio-temporal grid with (N, M) grid points is of complexity

$$O(N \max\{M^3, m_i^+ M^2, m_i^+ m^2\}).$$

In principle, the spatial grid size M chosen by the user is fixed but can be arbitrarily large. However, with growing number of observations n, the dimension $m_i^+ = (n-i+1)m$ quickly grows larger than M, leading to computational costs of complexity $O(Nm_i^+ M^2)$. In that case, it is beneficial to parametrise g in a different manner as we will explore in the next section.

4.2.3 Computationally efficient guiding

We derive an alternative to Theorem 4.6 of computing G to sample from the guided process X^g. Define the functions $(V_t)_{t\in[0,T]}$ and $(U_t)_{t\in[0,T]}$ by

$$\begin{aligned} U_t &:= L_t^* R_t^{-1} L_t \\ V_t &:= L_t^* R_t^{-1}(y_i^+ - \alpha_t), \quad t \in (t_{i-1}, t_i] \end{aligned} \tag{4.27}$$

for each $i = 1, \dots, n$ with $U_0 := U_0^+$ and $V_0 := V_0^+$. Following Theorem 4.6, it holds that

$$G(t, x) = V_t - U_t x. \tag{4.28}$$

As the upcoming theorem shows, U_t and V_t can be obtained by solving a set of infinite-dimensional backwards differential equations.

Theorem 4.9. *On each interval*[1] $(t_{i-1}, t_i]$, $i = 1, \dots, n$, *U is the unique mild solution to the backwards Ricatti equation*

$$\begin{cases} \mathrm{d}U_t &= [-A^*U_t - U_t A + U_t Q U_t]\,\mathrm{d}t \\ U_{t_i} &= L^*\Sigma^{-1}L + U_{t_i}^+, \end{cases} \tag{4.29}$$

whereas V is the unique mild solution to the backwards equation

$$\begin{cases} \mathrm{d}V_t &= [-A^*V_t + U_t Q V_t + U_t a_t]\,\mathrm{d}t, \\ V_{t_i} &= L^*\Sigma^{-1}y_i + V_{t_i}^+. \end{cases} \tag{4.30}$$

Remark 4.10. Without going into too much detail, it is worth pointing out that there exists extensive literature on the approximation of the infinite-dimensional Ricatti equation. We refer to Gibson [57], Burns and Rautenberg [18], Cheung [23] and references within for an introduction into the relevant literature.

For numerical purposes, the infinite-dimensional operators A and Q are typically approximated by $M \times M$-dimensional matrices on a spatial - or spectral - grid. Given such approximations of A and Q, the backwards Ricatti equation (4.29) can be numerically solved using standard solvers for ODEs with total costs of $O(NM^3)$ on a temporal grid of N points. In a similar fashion, the backwards evolution equation (4.30) can be solved using standard ODE solvers with total costs of $O(NM^3)$. Note that, contrary to the computation of G based on Theorem 4.6, this cost is independent of the number of observations. Moreover, in many cases, structural properties of the matrix approximations of A and Q, such as sparsity, can be taken advantage of to further reduce the computational costs.

[1] If $i = n$, U needs to be initialised with $U_{t_n} = L^*\Sigma L$ and V with $V_{t_n} = L^*\Sigma^{-1}y_n$.

Remark 4.11. In certain cases, the backwards Ricatti equation (4.29) can be solved in closed form. This includes the highly relevant case that A, Q and L are diagonalisable, i.e. if there exists an orthonormal basis $(e_j)_{j=1}^{\infty}$ with

$$\begin{aligned} Ae_j &= -a_j e_j, \\ Qe_j &= q_j e_j \end{aligned}$$

such that $0 < a_j \to \infty$ and $0 < q_j < \sup_j q_j < \infty$ and $\operatorname{im}(L^*) \subset \operatorname{span}\{e_j : j = 1, \dots m\}$. Equation (4.29) then decouples into a sequence $(u^j)_{j=1}^m$ of scalar-valued Ricatti differential equations

$$du_t^j = 2a_j u_t^j + q_j (u_t^j)^2 \,\mathrm{d}t, \quad t \in (t_{i-1}, t_i),$$

which are solved in closed form for all $j = 1 \dots m$ by (cf. Zaitsev and Polyanin [117], Chapter 1.1)

$$u_t^j = \frac{\exp(-2a_j(t_i - t))}{(u_{t_i}^j)^{-1} + \frac{q_j}{2a_j}\left[1 - \exp(-2a_j(t_i - t))\right]}, \qquad t \in (t_{i-1}, t_i).$$

When computing G based on Theorem 4.9, the following proposition is useful to derive the full expression of g. This is required, for example, when inferring not only unobserved states of X but also unknown model parameters θ.

Proposition 4.12. *In addition to U and V as defined in Theorem 4.9, let c be defined by*

$$c_t = -\frac{1}{2}\left[\log((2\pi)^{m_i^+}) + \log(\det(R_t)) + \langle y_i^+ - \alpha_t, R_t^{-1}(y_i^+ - \alpha_t)\rangle\right], \quad t \in (t_{i-1}, t_i].$$

Then $\log g(t,x) = c_t + \langle V_t, x\rangle - \frac{1}{2}\langle x, U_t x\rangle$. Moreover, on any interval $(t_{i-1}, t_i]$, $i = 1, \dots, n$, c_t solves the backwards differential equation

$$\begin{cases} \mathrm{d}c_t &= \left[\frac{1}{2}\operatorname{tr}\left[U_t Q\right] - \langle a_t, V_t\rangle - \frac{1}{2}\langle V_t, QV_t\rangle\right]\mathrm{d}t, \\ c_{t_i} &= \log f(y_i; 0, \Sigma) - c_{t_i}^+. \end{cases} \tag{4.31}$$

4.3 Guided particle filtering

In this section, we show how a special case of the guided distribution can be used as proposal distribution in a particle filtering scheme. For the convenience of the reader, we give a brief outline of the particle filter (PF). We refer to Doucet et al. [47] or Chopin et al. [27] for book-length treatments of this field.
The particle filter uses a set of particles $\{X_{t_i}^{(j)} : j = 1, \dots, J\}$ with associated weights $\{w_i^{(j)} : j = 1, \dots, J\}$ to approximate the filtering distribution (4.5) at any observation time t_i by a Monte Carlo approximation

$$\mathbb{E}[\varphi(X_{t_i}) \mid Y_1 = y_1, \dots, Y_n = y_i] \approx \sum_j w_{t_i}^{(j)} \varphi(X_{t_i}^{(j)})$$

for any bounded and measurable function $\varphi : H \to \mathbb{R}$.
The particles and their weights are sequentially updated with incoming observations as per the following recursion. Given an estimate $\{(X_{t_{i-1}}^{(j)}, w_{i-1}^{(j)}) : j = 1, \dots, J\}$ of the filtering distribution at time t_{i-1} and given a Markov kernel $\mathbb{Q}(\cdot \mid x)$ from H to $C([t_{i-1}, t_i]; H)$, the particles evolve forward to the next observation time t_i according to

$$X_{[t_{i-1}, t_i]}^{(j)} \sim \mathbb{Q}(\cdot \mid X_{t_{i-1}}^{(j)}), \quad j = 1, \dots J.$$

The weights are then updated (and normalised) via

$$w_i^{(j)} \propto w_{i-1}^{(j)} k(X_{t_i}^{(j)}, y_i) \frac{\mathrm{d}\mathcal{L}(X_{[t_{i-1},t_i]} \mid X_{t_{i-1}}^{(j)})}{\mathrm{d}\mathbb{Q}(\cdot \mid X_{t_{i-1}}^{(j)})} \left(X_{[t_{i-1},t_i]}^{(j)}\right), \quad \text{s.t.} \sum_{j=1}^{J} w_i^{(j)} = 1.$$

Here, $\mathcal{L}(X_{[t_{i-1},t_i]} \mid X_{t_{i-1}}^{(j)})$ denotes the law of $X_{[t_{i-1},t_i]}$ satisfying (4.1) with initial state $X_{t_{i-1}} = X_{t_{i-1}}^{(j)}$.

Typically, to prevent *weight degeneracy*, i.e. one or few weights dominating the others, a resampling step of the particles based on the normalised weights is introduced. This can be done every fixed number of steps or in an adaptive number based on maintaining the *effective sample size* $ESS_i := 1/\left(\sum_{j=1}^{J}(w_i^{(j)})^2\right)$ above a predefined threshold.

The proposal kernel $\mathbb{Q}$ is a parameter of the particle filter that is to be chosen by the user. If $\mathbb{Q}(\cdot \mid X_{t_{i-1}}^{(j)}) = \mathcal{L}(X_{[t_{i-1},t_i]} \mid X_{t_{i-1}}^{(j)})$, the algorithm is known as the *bootstrap particle filter (BPF)*. Since proposals in the BPF are 'uninformed' by the observations, it is known to typically struggle with highly degenerate weights.

On the other hand, the locally optimal proposal kernel is given by the conditional distribution

$$\mathbb{Q}(\cdot \mid X_{t_{i-1}}^{(j)}) = \mathbb{P}(X_{[t_{i-1},t_i]} \in \cdot \mid X_{t_{i-1}} = X_{t_{i-1}}^{(j)}, Y_i = y_i).$$

The proposal $\mathbb{Q}$ minimises the variance of the weights $w_i^{(j)}$, see for example Chopin et al. [27], Theorem 10.1 for details. However, as noted in the introduction and previous section, this proposal kernel is typically intractable.

Algorithm 3: Guided Particle Filter

Input: Observations $\{y_i\}_{i=1}^n$, number of particles J, threshold J_0

Output: Particles and weights $\left\{\left(X_{t_i}^{(j)}, w_i^{(j)}\right) : j = 1 \dots J\right\}$ for each $i = 1, \dots, n$

Initialise $\{X_0^{(j)}\}_{j=1}^J \sim \mu_0$ and set $w_0^{(j)} = \frac{1}{J}$;

for $i = 1$ **to** n **do**

1. Evolve particles:

for $j = 1$ **to** J **do**

(i) Sample guided proposal:

$$X_{[t_{i-1},t_i]}^{(j)} \sim \mathcal{L}^{g_i}(X_{[t_{i-1},t_i]} \mid X_{t_{i-1}}^{(j)});$$

(ii) Compute unnormalised weight:

$$\tilde{w}_i^{(j)} = w_{i-1}^{(j)} g_i(t_{i-1}, X_{t_{i-1}}^{(j)}) \exp\left(\int_{t_{i-1}}^{t_i} \langle \tilde{F}(s, X_s^{(j)}), G_i(s, X_s^{(j)}) \rangle \,\mathrm{d}s\right);$$

2. Normalize Weights:

$$w_i^{(j)} = \frac{\tilde{w}_i^{(j)}}{\sum_{k=1}^J \tilde{w}_i^{(k)}};$$

Compute effective sample size: $\mathrm{ESS} = 1/\sum_{j=1}^J \left(w_i^{(j)}\right)^2$;

3. Resample: if ESS $< J_0$ **then**

Resample $\{X_{t_i}^{(j)}\}_{j=1}^J$ according to weights $w_i^{(j)}$;

Set $w_i^{(j)} = \frac{1}{J}$ for all j;

We propose instead to steer the particles on the interval $[t_{i-1}, t_i]$ into regions of high probability given $Y_i = y_i$ by the guided distribution introduced in Section 4.2.2. For this, define

$$g_i(t,x) := \int k(x_{t_i}, y_i)\nu_{t,t_i}(x, \mathrm{d}x_{t_i}), \quad t \in [t_{i-1}, t_i],$$

and let $\mathbb{P}^{g_i}$ be the change of measure on $\mathcal{F}_{t_i}$ as defined in Theorem 4.7. We adopt as proposal kernel the guided distribution

$$\mathbb{Q}(\cdot \mid X^{(j)}_{t_{i-1}}) = \mathcal{L}^{g_i}(X_{[t_{i-1},t_i]} \mid X^{(j)}_{t_{i-1}}), \tag{4.32}$$

i.e. the law of $X_{[t_{i-1},t_i]}$ under $\mathbb{P}^{g_i}$ initiated at $X_{t_{i-1}} = X^{(j)}_{t_{i-1}}$. Particles are then evolved forward in time by simulating the guided process

$$\mathrm{d}X^{g_i}_t = [AX^{g_i}_t + F(t, X^{g_i}_t) + QG_i(t, X^{g_i}_t)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W^{g_i}_t, \quad t \in [t_{i-1}, t_i], \tag{4.33}$$

with respective initial conditions $X^{(j)}_{t_{i-1}}, j = 1, \dots J$ and $G_i(t,x) = \mathrm{D}_x \log g_i(t,x)$. Moreover, Theorem 4.7 implies that the update step in the corresponding weights is given by

$$\begin{aligned} w^{(j)}_i &\propto w^{(j)}_{i-1}\, k(X^{(j)}_{t_i}, y_i) \frac{\mathrm{d}\mathcal{L}(X_{[t_{i-1},t_i]} \mid X^{(j)}_{t_{i-1}})}{\mathrm{d}\mathcal{L}^{g_i}(X_{[t_{i-1},t_i]} \mid X^{(j)}_{t_{i-1}})} \left(X^{(j)}_{[t_{i-1},t_i]}\right) \\ &= w^{(j)}_{i-1}\, k(X^{(j)}_{t_i}, y_i) \frac{g_i(t_{i-1}, X^{(j)}_{t_{i-1}})}{g_i(t_i, X^{(j)}_{t_i})} \exp\left(\int_{t_{i-1}}^{t_i} \langle \tilde{F}(s, X^{(j)}_s), G_i(s, X^{(j)}_s)\rangle\,\mathrm{d}s\right) \\ &= w^{(j)}_{i-1}\, g_i(t_{i-1}, X^{(j)}_{t_{i-1}}) \exp\left(\int_{t_{i-1}}^{t_i} \langle \tilde{F}(s, X^{(j)}_s), G_i(s, X^{(j)}_s)\rangle\,\mathrm{d}s\right). \end{aligned} \tag{4.34}$$

The resulting *guided particle filter* is summarised in Algorithm 3.

Remark 4.13. The measure $\mathcal{L}^{g_i}(X_{[t_{i-1},t_i]})$ is the 'one-step-ahead' guided distribution on $[t_{i-1}, t_i]$ corresponding to the next observation y_i. Hence, the functions g_i and G_i only need to be computed as functions of y_i and the parametrisation of Theorem 4.6 remains a computationally valid choice. Specifically, on any $[t_{i-1}, t_i]$, $g_i(t,x)$ is the Gaussian likelihood $g_i(t,x) = f(y_i; L_t x + \alpha_t, R_t)$ with $L_t = LS_{t_i - t}$, $\alpha_t = L\left(\int_t^{t_i} S_{t_i - s} a_s\,\mathrm{d}s\right)$ and $R_t = \Sigma + LQ_{t_i - t}L^*$, whereas G_i is the score

$$G_i(t,x) = L^*_t R^{-1}_t (y_i - L_t x - \alpha_t), \quad t \in [t_{i-1}, t_i].$$

4.3.1 Tempering and moving steps

Even though the proposal kernel introduced in Equation (4.32) defines a suitable, likelihood-informed importance sampling distribution, the guided particle filter might still face weight degeneracy due to the high dimensionality of the state space. Since this is a common issue in particle-based inference, many solutions to overcome it have been proposed and studied in the literature. For the purpose of our numerical experiments in Section 4.5, we introduce two such solutions - *tempering* and *moving* (or *particle rejuvenation*) steps - into the guided particle filter. This follows the suggestions made in Llopis et al. [78], where the filtering problem was considered in a similar setup to ours. The following brief exposition is based on that work.

Denoting by Π_i the conditioned law $\Pi_i := \mathcal{L}(X_{[0,t_i]} \mid Y_1 \dots Y_i)$, an application of the Bayes theorem and the change of measure (4.32) gives the recursion

$$\begin{aligned} \frac{\mathrm{d}\Pi_i}{\mathrm{d}(\Pi_{i-1} \otimes \mathcal{L}^{g_i})}(X_{[0,t_i]}) &\propto g_i(t_{i-1}, X_{t_{i-1}}) \exp\left(\int_{t_{i-1}}^{t_i} \langle \tilde{F}(s, X_s), G_i(s, X_s) \rangle \, \mathrm{d}s \right) \\ &=: \Lambda_i(X_{[t_{i-1},t_i]}). \end{aligned} \tag{4.35}$$

Suppose that, at observation time t_i, we have an evenly weighted set of samples of the distribution Π_{i-1}. The filtering step at t_i then corresponds to targeting the measure Π_i by an importance sampling scheme with proposal distribution $\mathbb{Q}_i := \Pi_{i-1} \otimes \mathcal{L}^{g_i}$. Samples from this proposal are drawn by forward propagating the given particles of Π_{i-1} following the dynamics in (4.33) with IS weights given by Λ_i in (4.35).

Now fix the observation time t_i and drop the subscript i to simplify notation. In state spaces of high dimension and/or the situation of highly informative observations, the proposal distribution $\mathbb{Q}$ might not be a sufficiently good approximation for the target Π.
Tempering aims to overcome the distance between Π and $\mathbb{Q}$ by introducing a sequence of intermediate distributions Π^l between Π and $\mathbb{Q}$ defined by

$$\frac{\mathrm{d}\Pi^l}{\mathrm{d}\mathbb{Q}}(X) \propto \Lambda(X)^{\psi_l} \tag{4.36}$$

for a set of inverse temperatures $0 = \psi_0 < \ldots < \psi_L = 1$. Notably, we have absolute continuity of Π^{l+1} with respect to Π^l with R.N. derivative proportional to $\Lambda(X)^{\psi_{l+1}-\psi_l}$. Instead of targeting Π directly by $\mathbb{Q}$, one progressively steps through this sequence of artificial distributions, starting from $\mathbb{Q}$ to target the flattened Π^1 and ending up at targeting Π with proposals drawn from Π^{L-1}.
Crucially, the parameters $\{\psi_l\}_l$ can be chosen adaptively as follows. Given a set of evenly weighted particles of Π^l at current inverse temperature ψ_l, the parameter ψ_{l+1} is set by

$$\psi_{l+1} := \inf\{\psi \in (\psi_l, 1] : \mathrm{ESS}_{l,l+1} \leq \alpha J\}$$

with the definition of $\inf \emptyset := 1$. The hyperparameter $\alpha \in (0,1)$ is pre-specified by the user and $\mathrm{ESS}_{l,l+1}$ denotes the essential sample size of the weights $\Lambda(X)^{\psi_{l+1}-\psi_l}$ between Λ^{l+1} and Λ^l.
After Π^{l+1} has been determined, particles are resampled based on the weights $\Lambda(X)^{\psi_{l+1}-\psi_l}$. Subsequently, to increase particle diversity, each particle is moved by a few steps of an MCMC scheme that targets Π^{l+1}. This can be done by proposing independent samples from $\mathbb{Q}$ and accepting/rejecting the proposals in a Metropolis-Hastings step based on the likelihood in Equation (4.36). Alternatively, proposals can be localised based on the preconditioned Crank-Nicolson scheme, see Section 4.4 for details. The complete *tempering and move* procedure is summarised in Algorithm 4. It is to be carried out at each observation time t_i in place of Steps 2. and 3. in Algorithm 3.

Remark 4.14. For ease of exposition, the MCMC move step in Algorithm 4 is presented using independent proposals drawn from $X' \sim \mathbb{Q}$. These proposals can be localised in Wiener space by using the preconditioned Crank-Nicolson scheme as is further explained in Section 4.4. For practical purposes, this implies having to store at the observation time t_i for each particle $X^{(j)}$ the sample of the Wiener process W that corresponds to simulating the sample path $X^{(j)}_{[t_{i-1},t_i]}$ following the dynamics in (4.33). Note that this is still carried out in an online setting since the stored Wiener paths relate to the evolution on $[t_{i-1}, t_i]$ only and may hence be discarded after each complete filtering step.

4.4 Guided smoothing via Metropolis-Hastings sampling

The following section introduces a solution to the smoothing problem via a Metropolis-Hastings (MH) algorithm that targets the smoothing distribution $\mathcal{L}^h(X)$ based on proposals drawn from the guided distribution $\mathcal{L}^g(X)$.

4.4.1 The case of known X_0.

Assume first that the initial distribution μ_0 is given by a Dirac measure δ_{x_0} for some $x_0 \in H$. In that case, $X_0 = x_0$ remains unchanged under the changes of measure $\mathbb{P}^h$ and $\mathbb{P}^g$ and the Radon-Nikodym derivative in (4.25) reduces to

$$\Phi(X) = \frac{g(0, x_0)}{h(0, x_0)} \Psi(X)$$

with $\Psi(X)$ as defined in Equation (4.26). A basic MH sampler that targets $\mathcal{L}^h(X)$ can then be constructed by proposing samples from $\mathcal{L}^g(X)$ and accepting or rejecting the proposals with an acceptance probability based on $\Phi(X)$. Note that $\Phi(X)$ includes the intractable term $h(0, x_0)$ which acts as a proportionality constant and cancels out in an MH acceptance step for a known x_0.
Sampling from $\mathcal{L}^g(X)$ and evaluating $\Psi(X)$ relies on the function G. This function needs to be computed *once* as a function of the complete set of observations on the basis of the backwards information filter in Equation (4.18). Either of the parametrisations in

Algorithm 4: Tempering and move steps at fixed observation time t_i

Input: Samples $X^{(j)}$ of $\mathbb{Q} = \Pi_{i-1} \otimes \mathcal{L}^{g_i}$ with weights $\Lambda(X^{(j)})$, hyperparameters α and N
Output: Evenly weighted samples $X^{(j)}$ of $\Pi = \Pi_i$
Initialise $l = 0$ and $\psi_l = 0$;
while $\psi < 1$ **do**
  1. Compute next ψ: set $l \leftarrow l + 1$ and

$$\psi_l = \inf\{\psi \in (\psi_{l-1}, 1] : \mathrm{ESS}_{l-1,l} \leq \alpha J\};$$

  2. Specify Π^l and resample:
   (i) Compute weights and normalise

$$w_l^{(j)} \propto \Lambda(X^{(j)})^{\psi_l - \psi_{l-1}} \quad \text{s.t.} \sum_{j=1}^{J} w_l^{(j)} = 1$$

   (ii) Resample $\{X^{(j)}\}_{j=1}^{J}$ according to weights $w_l^{(j)}$;
  3. Move particles:
  for $j = 1$ **to** J **do**
   for $n = 1$ **to** N **do**
    (i) Sample guided proposal: $X' \sim \mathbb{Q}$;
    (ii) Compute $M = \min\left(1, \left(\frac{\Lambda(X')}{\Lambda(X^{(j)})}\right)^{\psi_l}\right)$;
    (iii) Draw $U \sim \mathrm{Unif}(0, 1)$;
     if $U < M$ **then**
      Set $X^{(j)} = X'$;

Algorithm 5: MH Sampler of $\mathcal{L}^h(X)$

Input: Parameters A, F, Q and x_0, observations y, iterations N, N_0, step size β

Output: Samples $\{X_i\}$ of $\mathcal{L}^h(X)$ (after burn-in period N_0)

Initialize:

Precompute G by solving the backwards ODEs $(U_t)_{t\in[0,T]}$ and $(V_t)_{t\in[0,T]}$ in (4.29) and (4.30);

Draw a Wiener process V and set $X = \Gamma(x_0, V)$;

for $i = 0 \dots N-1$ **do**

  1. MH step in X

  (i) Draw a Wiener process W and set $V' = \sqrt{1-\beta^2}V + \beta W$;

  (ii) Compute $X' = \Gamma(x_0, V')$ and $M = \min\left(1, \dfrac{\Psi(X')}{\Psi(X)}\right)$;

  (iii) Draw $U \sim \mathrm{Unif}(0,1)$;

  if $U < M$ **then**

    Set $X = X'$ and $V = V'$;

  2. Save current state X

  if $i > N_0$ **then**

    Set $X_i = X$;

Theorem 4.6 or Theorem 4.9 can be used for this, though in practice the latter is usually the computationally favourable one.

To localise the proposals drawn from $\mathcal{L}^g(X)$ we introduce random-walk-type proposals in the Wiener space by adapting a *preconditioned Crank-Nicolson (pCN)* scheme (see Neal [87], Beskos et al. [14], Cotter et al. [37]) as follows.
By Theorem 4.7, drawing samples of $\mathcal{L}^g(X)$ is equivalent to sampling the mild solution X^g to Equation (4.24). The well-posedness of Equation (4.24) implies the existence of a measurable map Γ such that

$$X^g = \Gamma(x_0, W), \quad \mathbb{P}^g\text{-a.s.}, \tag{4.37}$$

where W is a $\mathbb{P}^g$-cylindrical Wiener process. Assume now that the current value X of the MH sampler is such that $X = \Gamma(x_0, V)$ for some process V. The pCN proposal is then given by:

(i) Draw a Wiener process W, independent of V;
(ii) Set $V' = \sqrt{1-\beta^2}V + \beta W$;
(iii) Propose $X' = \Gamma(x_0, V')$.

The parameter $\beta \in (0,1]$ determines the size of the pCN step. For $\beta = 1$, independent proposals of X^g are drawn. The resulting MH sampler is given in Algorithm 5.

Remark 4.15. Since W is a cylindrical Wiener process on H, it cannot be treated directly as an H-valued process. However, W can be defined as a Wiener process taking values in a larger Hilbert space H' such that $H \hookrightarrow H'$ is embedded in a Hilbert-Schmidt way. Note that such an embedding always exists, see Hairer [61], Section 4.4 for details. The solution map Γ in (4.37) is then to be understood as a measurable mapping $\Gamma : H \times C([0,T]; H') \to C([0,T]; H)$.

Remark 4.16. In the spirit of Remark 4.15, the MH sampler in Algorithm 5 can be rephrased as targeting a measure $\mathbb{V}^{x_0}$ that is absolutely continuous with respect to a Wiener measure $\mathbb{W}$ on $C([0,T]; H')$. For this, let V be a $\mathbb{P}^g$-cylindrical Wiener process on H, taking values in the enlargement H' of H. Note that the measures $\mathbb{P}^g$ and $\mathbb{P}^h$

depend on the fixed initial state x_0. Define $\mathbb{V}^{x_0} = \mathcal{L}^h(V)$ as the law of V under $\mathbb{P}^h$ on $C([0,T];H')$. It immediately follows that

$$\frac{\mathrm{d}\mathbb{V}^{x_0}}{\mathrm{d}\mathbb{W}}(V) = \frac{\mathrm{d}\mathcal{L}^h(V)}{\mathrm{d}\mathcal{L}^g(V)}(V) = \Phi(\Gamma(x_0,V)) = \frac{g(0,x_0)}{h(0,x_0)}\Psi(\Gamma(x_0,V)). \tag{4.38}$$

On the other hand, it holds for any bounded and measurable functional f that

$$\mathbb{E}^h\left[f(\Gamma(x_0,V))\right] = \mathbb{E}^g[f(\Gamma(x_0,V))\Phi(\Gamma(x_0,V))] = \mathbb{E}^g[f(X)\Phi(X)] = \mathbb{E}^h[f(X)].$$

Hence, $\mathbb{V}^{x_0}$ is the unique measure on $C([0,T];H')$ that is absolutely continuous with respect to $\mathbb{W}$ such that, if $V \sim \mathbb{V}^{x_0}$, then $\Gamma(x_0,V)$ equals in law the smoothing distribution of X given x_0.

4.4.2 The case of unknown X_0.

Let now X_0 be unknown with prior distribution μ_0 and suppose the following assumption is satisfied.

ASSUMPTION 4.17. *There exists a Gaussian measure ν_0 on $(H,\mathcal{B}(H))$ such that $\mu_0 \ll \nu_0$ with density $\rho(x_0) := \frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(x_0)$.*

A simple calculation shows that the conditional law $\mu_0^h = \mathcal{L}^h(X_0)$ of X_0 given $Y = y$ remains absolutely continuous with respect to ν_0 with density

$$\frac{\mathrm{d}\mu_0^h}{\mathrm{d}\nu_0}(x_0) = \frac{h(0,x_0)\rho(x_0)}{C^h}. \tag{4.39}$$

Alternatively, using Bayesian notation and recalling that $h(0,x_0) = p(y \mid x_0)$ and $C^h = \mathbb{E}[h(0,X_0)] = p(y)$, Equation (4.39) is simply a restatement of the Bayes theorem.
The decomposition $X = \Gamma(X_0,V)$ introduced in the previous section suggests to sample from the full smoothing distribution of X by targeting the joint measure $\mathbb{V}\otimes\mu_0^h$ defined through the disintegration $\mathbb{V}\otimes\mu_0^h := \mathbb{V}^{x_0}(\mathrm{d}V)\mu_0^h(\mathrm{d}x_0)$ on $C([0,T];H')\otimes H$. Sampling from the disintegrated measure corresponds to first sampling an initial state $X_0 = x_0$ given the observations y and subsequently drawing a sample $V \sim \mathbb{V}^{x_0}$ of the measure $\mathbb{V}^{x_0}$ introduced in Remark 4.16. The sample $\Gamma(x_0,V)$ then represents a realisation from the smoothing distribution of X.
Plugging in the densities in (4.38) and (4.39) shows that $\mathbb{V}\otimes\mu_0^h$ is absolutely continuous with respect to $\mathbb{W}\otimes\nu_0$ with Radon–Nikodym derivative

$$\frac{\mathrm{d}(\mathbb{V}\otimes\mu_0^h)}{\mathrm{d}(\mathbb{W}\otimes\nu_0)}(V,x_0) = \frac{g(0,x_0)\rho(x_0)}{C^h}\Psi(\Gamma(x_0,V)). \tag{4.40}$$

This density can be used to construct a Gibbs sampler targeting $\mathbb{V}\otimes\mu_0^h$ as summarised in Algorithm 6.

4.4.3 Parameter estimation

Suppose the parameters A, F, Q and μ_0 of the dynamics in Equation (4.1) depend on some unknown parameter $\theta \in \mathbb{R}^p$. This dependence carries over to the majority of the processes and functions that we have discussed so far. However, for notational clarity, we only make this dependence explicit in certain cases by adding a subscript θ onto the dependent variables. In particular, let us highlight the dependence of the functions h_θ, g_θ and Γ_θ as well as the measures $\mathbb{V}_\theta^{x_0}$ and $\mu_{0,\theta}$ on θ.
The unknown parameter θ is to be inferred jointly with the unobserved states of X, based on the observations y. Our methodology makes it natural to do so in a fully Bayesian approach. For this, assume that θ is the realisation of some $\mathcal{F}_0$-measurable random variable Θ with Lebesgue density $\pi(\theta)$. With slight abuse of notation, denote by π^y the

conditional distribution of θ given y. We then aim to sample from the joint posterior measure defined by $\mathbb{V}\otimes\mu_0^h\otimes\pi^y := \mathbb{V}_\theta^{x_0}(\mathrm{d}V)\mu_{0,\theta}^{h_\theta}(\mathrm{d}x_0)\pi^y(\mathrm{d}\theta)$. Following Equation (4.40) and the Bayes theorem, $\mathbb{V}\otimes\mu_0^h\otimes\pi^y$ has density with respect to $\mathbb{W}\otimes\nu_0\otimes\mathrm{Leb}(\mathbb{R}^p)$ given by

$$\begin{aligned}\frac{\mathrm{d}(\mathbb{V}\otimes\mu_0^h\otimes\pi^y)}{\mathrm{d}(\mathbb{W}\otimes\nu_0\otimes\mathrm{Leb}(\mathbb{R}^p))}(V,x_0,\theta) &= \frac{g_\theta(0,x_0)\rho_\theta(x_0)}{C^{h_\theta}}\Psi_\theta(\Gamma_\theta(x_0,V))\frac{p(y\mid\theta)\pi(\theta)}{\int p(y\mid\theta)\pi(\theta)\,\mathrm{d}\theta}\\ &\propto g_\theta(0,x_0)\rho_\theta(x_0)\pi(\theta)\Psi_\theta(\Gamma_\theta(x_0,V)),\end{aligned} \tag{4.41}$$

where we used that $C^{h_\theta}=\mathbb{E}[h_\theta(0,X_0)]=p(y\mid\theta)$. An MCMC algorithm sampler that targets $\mathbb{V}\otimes\mu_0^h\otimes\pi^y$ can hence be defined by adding a Gibbs sampling step $\theta\mid V,x_0$ to Algorithm 6. The details are summarised in Algorithm 7.

REMARK 4.18. If A and Q are independent of the unknown parameter θ, one only needs to solve the backwards ODEs $(U_t)_{t\in[0,T]}$ and $(V_t)_{t\in[0,T]}$ in (4.29) and (4.30) once. This severely eases the computational burden of Algorithm 7 for the highly relevant case that the parameter θ of interest parametrises only the nonlinearity $F=F_\theta$.

4.4.4 Non-Gaussian observations

So far we have assumed the observations Y_i given X_{t_i} to be Gaussian with conditional density $k(X_{t_i},\cdot)=f(\cdot;LX_{t_i},\Sigma)$. It is straightforward to let the kernel k depend on the

---

Algorithm 6: Gibbs Sampler of $\mathbb{V}\otimes\mu_0^h$

Input: Parameters A, F, Q and ρ, observations y, iterations N,N_0, step sizes β,β_0

Output: Samples $\{(V,X_0)_i\}$ of $\mathbb{V}\otimes\mu_0^h$ (after burn-in period N_0)

Initialise:

Precompute G by solving the backwards ODEs $(U_t)_{t\in[0,T]}$ and $(V_t)_{t\in[0,T]}$ in (4.29) and (4.30);

Draw a Wiener process V and $X_0\sim\nu_0$;

for $i=0\ldots\, N-1$ **do**

- **1. Update** $V\mid X_0$
  - (i) Draw a Wiener process W and set $V'=\sqrt{1-\beta^2}V+\beta W$;
  - (ii) Compute $M=\min\left(1,\dfrac{\Psi(\Gamma(X_0,V'))}{\Psi(\Gamma(X_0,V))}\right)$;
  - (iii) Draw $U\sim\mathrm{Unif}(0,1)$;
  - **if** $U<M$ **then**
    - Set $V=V'$;
- **2. Update** $X_0\mid V$
  - (i) Draw $z\sim\nu_0$ and set $X_0'=\sqrt{1-\beta_0^2}X_0+\beta_0 z$;
  - (ii) Compute $M=\min\left(1,\dfrac{\rho(X_0')g(0,X_0')\Psi(\Gamma(X_0',V))}{\rho(X_0)g(0,X_0)\Psi(\Gamma(X_0,V))}\right)$;
  - (iii) Draw $U\sim\mathrm{Unif}(0,1)$;
  - **if** $U<M$ **then**
    - Set $X_0=X_0'$;
- **3. Save current state** (V,X_0)
  - **if** $i>N_0$ **then**
    - Set $(V,X_0)_i=(V,X_0)$;

Algorithm 7: Gibbs Sampler of $\mathbb{V} \otimes \mu_0^h \otimes \pi^y$

Input: Parameters A_θ, F_θ, Q_θ and ρ_θ, observations y, prior π, proposal kernel q,
iterations N, N_0, step sizes β, β_0
Output: Samples $\{(V, X_0, \theta)_i\}$ of $\mathbb{V} \otimes \mu_0^h \otimes \pi^y$ (after burn-in period N_0)
Initialize:
Draw $\theta \sim \pi$, a Wiener process V and $X_0 \sim \rho_\theta \, \mathrm{d}\nu_0$;
Compute G_θ by solving the backwards ODEs $(U_t)_{t\in[0,T]}$ and $(V_t)_{t\in[0,T]}$ in (4.29) and (4.30);
for $i = 0 ... N-1$ **do**
  1. Update $V \mid X_0, \theta$
  See Algorithm 6, Step 1 with $(\Psi, \Gamma) = (\Psi_\theta, \Gamma_\theta)$;
  2. Update $X_0 \mid V, \theta$
  See Algorithm 6, Step 2 with $(\rho_\theta, g_\theta, \Psi, \Gamma) = (\rho_\theta, g_\theta, \Psi_\theta, \Gamma_\theta)$;
  3. Update $\theta \mid X_0, V$
   (i) Draw $\theta' \sim q(\cdot \mid \theta)$;
   (ii) If A, Q depend on θ: Compute $G_{\theta'}$ by solving Eq. (4.29) and Eq. (4.30);
   (iii) Compute
$$M = \min\left(1, \frac{q(\theta \mid \theta')\pi(\theta')\rho_{\theta'}(X_0)g_{\theta'}(0, X_0)\Psi_{\theta'}(\Gamma_{\theta'}(X_0, V))}{q(\theta' \mid \theta)\pi(\theta)\rho_\theta(X_0)g_\theta(0, X_0)\Psi_\theta(\Gamma_\theta(X_0, V))}\right);$$
   (iii) Draw $U \sim \mathrm{Unif}(0, 1)$;
   if $U < M$ **then**
    Set $\theta = \theta'$ and $G_\theta = G_{\theta'}$;
  4. Save current state (V, X_0, θ)
   if $i > N_0$ **then**
    Set $(V, X_0, \theta)_i = (V, X_0, \theta)$;

observation time t_i by setting $k_i(X_{t_i}, \cdot) = f(\cdot; L_i X_{t_i}, \Sigma_i)$ for time varying observation operators $L_i : H \to \mathbb{R}^{m_i}$ and extrinsic noise matrices $\Sigma_i \in \mathbb{R}^{m_i \times m_i}$.
More crucially, our proposed methodology can be extended to the case of non-Gaussian observations in the following way. Let l_i be some Markov kernel from H to $\mathbb{R}^{m_i}$ and assume observations Y_i are distributed following

$$Y_i \mid X_{t_i} \sim l_i(X_{t_i}, \cdot). \tag{4.42}$$

Substituting $l_i(x_{t_i}, y_i)$ for $k(x_{t_i}, y_i)$ in the definition of the h-function in Equation (4.11) preserves the properties of $h(t, x)$ established in Propositions 4.1 and 4.3 as well as Theorem 4.4. Hence, the smoothing distribution of X under the observation scheme in (4.42) is the law of X under the change of measure

$$\frac{\mathrm{d}\mathbb{P}^h}{\mathrm{d}\mathbb{P}}(X) = E_T^h = \frac{1}{C^h}\left(\prod_{i=1}^n l_i(X_{t_i}, y_i)\right)$$

on $\mathcal{F}_T$. For the guided distribution of X, the Markov kernel k_i needs to remain Gaussian in order to preserve the tractable expressions for $g(t, x)$ obtained in Section 4.2.2. It is therefore defined under the change of measure

$$\frac{\mathrm{d}\mathbb{P}^g}{\mathrm{d}\mathbb{P}}(X) = E_T^g = \frac{1}{C^g}\left(\prod_{i=1}^n k_i(X_{t_i}, y_i)\right)\Psi(X).$$

In total, the R.N. derivative of the smoothing distribution with respect to the guided distribution then takes on the form

$$\Phi(X) = \frac{C^g}{C^h} \left(\prod_{i=1}^{n} \frac{l_i(X_{t_i}, y_i)}{k_i(X_{t_i}, y_i)} \right) \Psi(X).$$

4.5 Case study: stochastic Amari equation

4.5.1 Stochastic Amari equation

In the following section, we evaluate our proposed methodology for filtering, smoothing, and parameter estimation in the context of a partially observed stochastic Amari model. For this, let $H = L^2(D)$ for some bounded spatial domain $D \subset \mathbb{R}^d$. For ease of exposition, we assume the domain D to be rectangular and impose periodic boundary conditions on it. Consider then the mild solution X to the stochastic Amari equation

$$\begin{aligned} \mathrm{d}X_t &= [-X_t + F(X_t)]\,\mathrm{d}t + Q^{\frac{1}{2}}\,\mathrm{d}W_t, \quad t \geq 0, \\ X_0 &= x_0, \end{aligned} \tag{4.43}$$

with nonlinearity $F : L^2(D) \to L^2(D)$ defined by

$$F(X_t) = \int_D k_F(\xi, \xi') f(X_t(\xi'))\,\mathrm{d}\xi'. \tag{4.44}$$

The process X_t models a stochastic neural field on D, with $X_t(\xi)$ representing the average neural activity at location $\xi \in D$ and time $t \geq 0$.
The kernel $k_F(\xi, \xi')$ captures neural connectivity between locations ξ and ξ', whereas f is an activation function, typically a sigmoid-like function or similar. In our experiments, we consider

$$\begin{aligned} k_F(\xi, \xi') &= \frac{A}{\sqrt{\pi}} \exp\left(-(|\xi - \xi'| - \delta)^2\right) - \frac{A}{\sqrt{\pi}B} \exp\left(-\left(\frac{|\xi - \xi'| - \delta}{B}\right)^2\right) \\ f(x) &= \frac{1}{1 + \exp\left(-\eta\, x + \zeta\right)} - \frac{1}{1 + \exp(\zeta)}, \end{aligned} \tag{4.45}$$

for a set of parameters $A > 0, B > 0, \eta > 0, \zeta > 0$ and $\delta \in \mathbb{R}$.
For the mild solution X to exist, one may consider any trace-class operator Q on $L^2(D)$. We let Q be a diagonal operator defined by its eigenvalues

$$q_l = \sigma_0^2(\rho_0^{-2} + (2\pi l)^2)^{-(d/2+\eta_0)} \tag{4.46}$$

with respect to the Fourier basis functions $(e_l)_{l=1}^{\infty}$ on $L^2(D)$. The operator Q is then the Hilbert-Schmidt integral operator defined by the Matérn kernel on D, see for example Borovitskiy et al. [17], Theorem 5.

Observation and guiding scheme We consider observations $Y_i \in \mathbb{R}^m$ of the latent process X at observation times $(t_i)_{i=1}^n$ defined as in Eq. (4.4). The observation operator $L : L^2(D) \to \mathbb{R}^m$ we set as

$$Lx = \left(\int_D x(\xi)\omega_j(\xi)\,\mathrm{d}\xi \right)_{j=1}^{m} = (\langle x, \omega_j \rangle)_{j=1}^{m} \in \mathbb{R}^m \tag{4.47}$$

for a set of positive, square-integrable functions $\omega_j \in L^2(D), j = 1, \ldots m$. The adjoint operator L^* then takes the form

$$L^* y = \sum_{j=1}^{m} y_j \omega_j \in L^2(D), \quad y \in \mathbb{R}^m.$$

Specifically, we let $\omega_j = \frac{1}{|D_j|}\mathbb{1}_{D_j}$ for a series of mutually disjoint subsets $D_j \subset D$, $j = 1 \dots m$. The operator L then represents observing localised spatial averages of $X_{t_i}(\xi)$ over the domains D_j at observation time t_i.

To construct the guiding distribution, we set $(a_t)_{t \geq 0}$ in (4.15) as $a_t \equiv 0$. For the one-step ahead guiding distribution we receive the following closed-term expression. The computations can be found in Appendix 4.B.

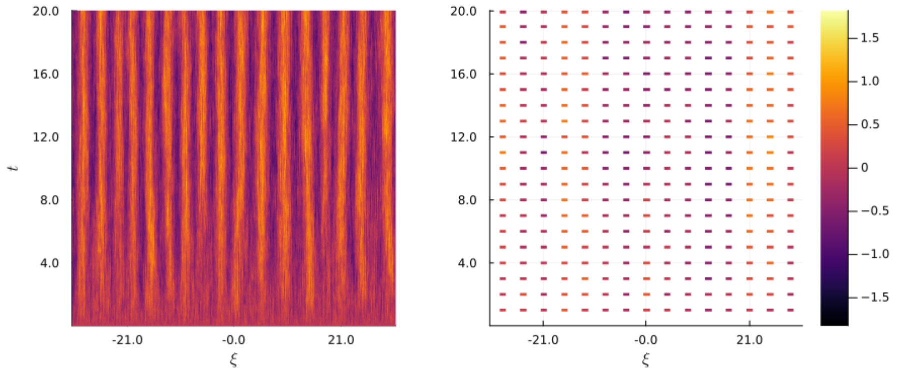


Figure 4.5.1. Data generating process of Experiment 1. Colour intensity represents the value of the process and observations, whereas time t is plotted along the y-axis and spatial coordinate ξ along the x-axis. Left: Heatmap of the sample path $X(t, \xi)$ of Eq. (4.43) with parameters as in (4.48) and $\delta_0 = 0$. Right: Observations Y_i at $n = 20$ observation times with $m = 15$ local averages of $X(t_i, \xi)$ per observation.

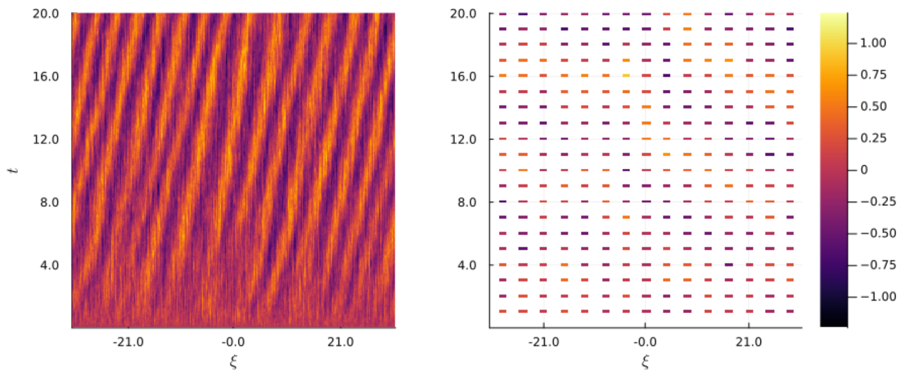


Figure 4.5.2. Data generating process of Experiment 2. Colour intensity represents the value of the process and observations, whereas time t is plotted along the y-axis and spatial coordinate ξ along the x-axis. Left: Heatmap of the sample path $X(t, \xi)$ of Eq. (4.43) with parameters as in (4.48) and $\delta_0 = 0.5$. Right: Observations Y_i at $n = 20$ observation times with $m = 15$ local averages of $X(t_i, \xi)$ per observation.

Proposition 4.19. *In the setup above, the guiding term* G_i *of Eq.* (4.33) *at observation time* t_i *is given by*

$$G_i(t,x) = \exp(-(t-t_i))L^* \left(\Sigma + \frac{1-\exp(-2(t-t_i))}{2} LQL^*\right)^{-1} (y_i - \exp(-(t-t_i))Lx)$$

where LQL^* *is the symmetric, positive definite* $m \times m$*-matrix with entries*

$$(LQL^*)[i,j] = \sum_{l=1}^{\infty} q_l \langle \omega_i, e_l \rangle \langle \omega_j, e_l \rangle.$$

4.5.2 Filtering

Model setup For the filtering problem, we run two experiments with different choices $\delta_0 = 0$ and $\delta_1 = 0.5$ of the model parameter δ. The remaining model parameters in (4.45) and (4.46) are set in both experiments as

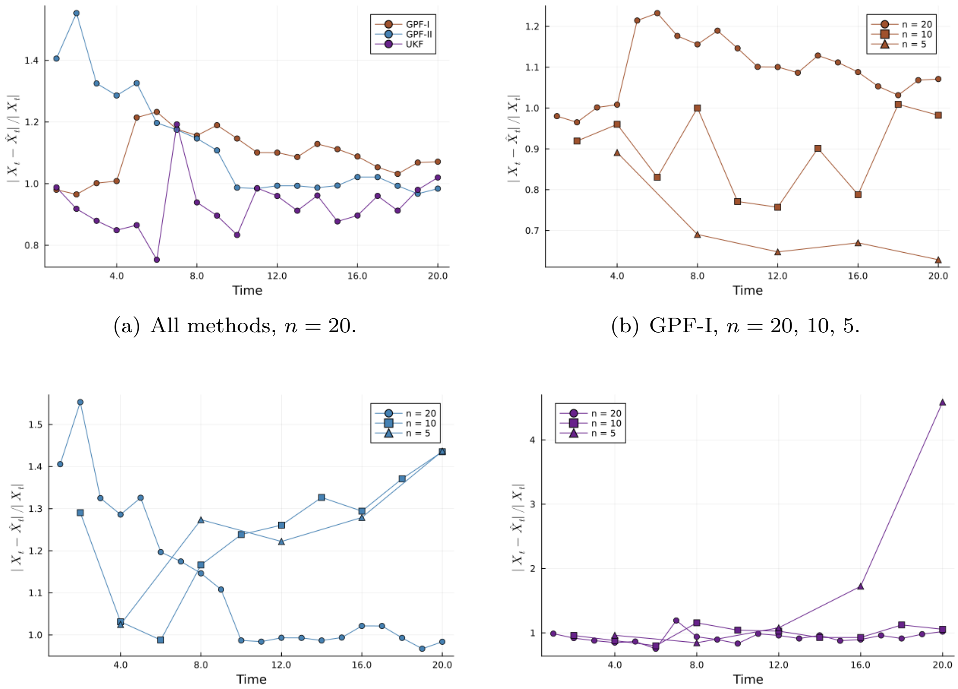
$$[A, B, \eta, \zeta, \sigma_0, \rho_0, \eta_0] = [4.0, 1.5, 10.0, 0.5, 3 \cdot 10^5, 5 \cdot 10^{-5}, 1.0] \tag{4.48}$$

and the initial state is set to be $X_0 \equiv 0$. Under this parametrisation, the stochastic Amari equation exhibits either steady state Turing patterns ($\delta_0 = 0$) or travelling waves ($\delta_1 = 0.5$) emerging from the initial state X_0.
Experiments are carried out on data generated by forward simulating the solution X to Equation (4.43) on the spatio-temporal domain $[0,T] \times D = [0,20] \times [-10\pi, 10\pi]$ with step sizes $\Delta_t = 0.02$ and $\Delta_x = 20\pi/2^8$.

Data A total number of $n = 20$ observations are taken at observation times $t_1 = 1.0$, $t_2 = 2.0, \ldots, t_{20} = 20.0$. Each observation $Y_i, i = 1, \ldots, n$, is defined as in Equation (4.47) with $\omega_j = \frac{1}{|D_j|}\mathbb{1}_{D_j}$ and measurement domains D_j defined at $m = 15$ equal spatial intervals with length $|D_j| = 1$. As measurement noise we set $\Sigma_i = 0.01 \cdot I$.
The data generating processes X and resulting observations are shown in Figure 4.5.1 and 4.5.2 for the two choices of δ respectively.
Additionally, for each experiment, we investigate performance of the proposed methods when downsampling the data from $n = 20$ to $n = 10$ and $n = 5$ observations.

Methods For each of the two experiments and the three (downsampled) observation schemes, we run the proposed guided particle filter in Algorithm 3, including the tempering and moving steps of Algorithm 4 with $J = 100$ particles, $N = 30$ MCMC move steps with pCN step size $\beta = 0.1$ and ESS threshold parameter $\alpha = 0.75$.
Additionally, we implement the following two established methods as a baseline for comparison:

1. The guided particle filter as proposed in Llopis et al. [78]. This method differs to ours in the proposal distribution $\mathbb{Q}$ of Eq. (4.32). The proposals chosen in Llopis et al. [78] are defined by solving the guided process (4.33) with guiding term
$$\tilde{G}_i(t,x) = L^* \left(\Sigma + (t_i - t)LQL^*\right)^{-1} (y - Lx), \quad t \in [t_{i-1}, t_i]. \tag{4.49}$$
In the context of our work, this guiding term can be derived by taking $A = 0$ in the auxiliary process Z. To differentiate between the two guided particle filters in the experiments we denote by GPF-I our version and GPF-II the one based on the guiding term $\tilde{G}$. For each experiment, both algorithms are implemented with the same hyperparameter choices of J, N, β and α.
2. The unscented Kalman filter (UKF) introduced in Wan and Van Der Merwe [112]. The UKF has been proposed for the filtering problem of an Amari model



(a) All methods, $n = 20$. (b) GPF-I, $n = 20$, 10, 5.

(c) GPF-II, $n = 20$, 10, 5. (d) UKF, $n = 20$, 10, 5.

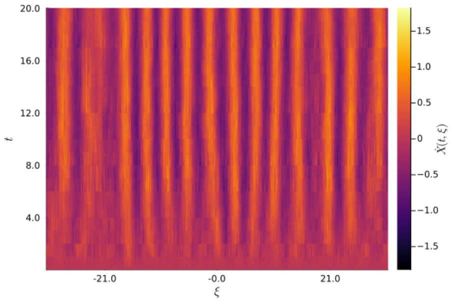
Figure 4.5.3. Relative errors $|\hat{X}_{t_i} - X_{t_i}|/|X_{t_i}|$ at observation times t_i for the filtering estimates $\hat{X}_{t_i}$ based on the data set of Figure 4.5.1 and the three methods GPF-I (our proposed method), GPF-II (method with guiding term in (4.49)) and UKF. Subfigure (a): a comparison of estimation errors for all three methods in the dense observation case. Subfigures (b) - (d): estimation errors of individual methods across subsampled datasets.

in Schiff and Sauer [102], where it has been shown to perform well in the setting of a discretisation of Equation (4.43).

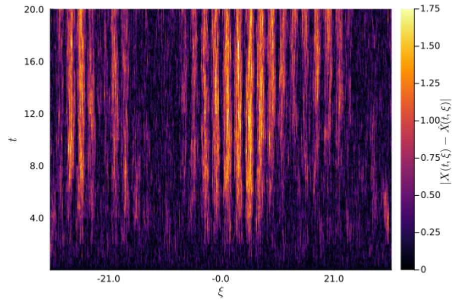
Results Experiment 1 Figure 4.5.3 shows relative errors $|\hat{X}_{t_i} - X_{t_i}|/|X_{t_i}|$ of the filtering estimates $\hat{X}_{t_i}$ returned by the three respective methods at observation times $t_1, \dots t_n$. In Figure 4.5.3(a), all three methods are compared in the dense observation scheme with $n = 20$ observations. In this setting, the UKF outperforms both particle filtering methods. We hypothesise that this is due to two reasons. Firstly, the UKF relies on the temporal discretisation of the Amari equation between observation points. In this dense observation setting, the discretisation proves to be sufficient for the UKF to perform well on the filtering task. Secondly, in the chosen parametrisation with $\delta = 0$, the Amari equation converges to an equilibrium state at which the dynamics behave roughly as

$$\mathrm{d}X_t \approx \sqrt{Q} dW_t. \tag{4.50}$$

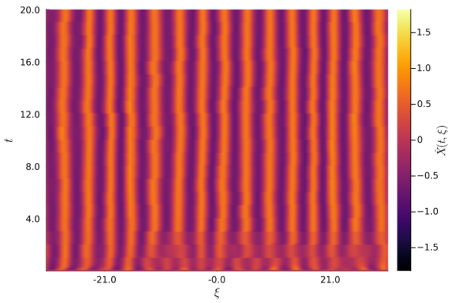
Hence, once equilibrium is reached, the nonlinear part of the dynamical system vanishes, rendering a temporal discretisation more forgiving. As can be observed in Figure 4.5.3(d), with increasing time lag between observations, the performance of the UKF worsens drastically to the point of diverging in the case of $n = 5$.



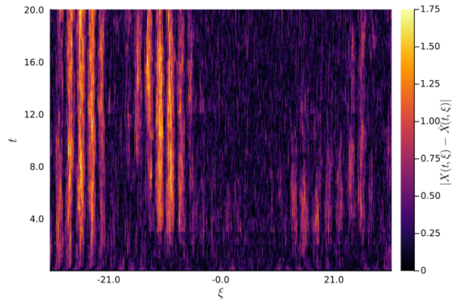
(a) GPF-I, $n = 20$.



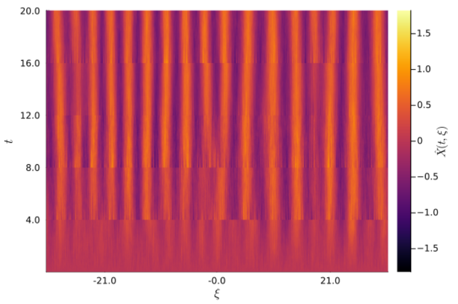
(b) GPF-I, $n = 20$.



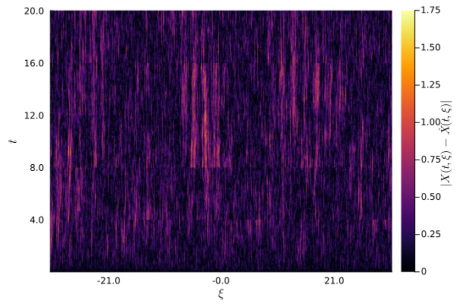
(c) GPF-II, $n = 20$.



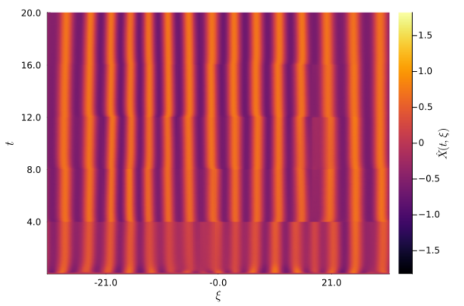
(d) GPF-II, $n = 20$.



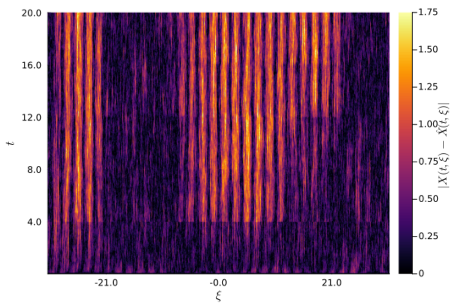
(e) GPF-I, $n = 5$.



(f) GPF-I, $n = 5$.

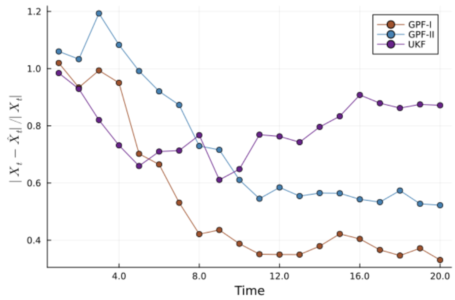


(g) GPF-II, $n = 5$.

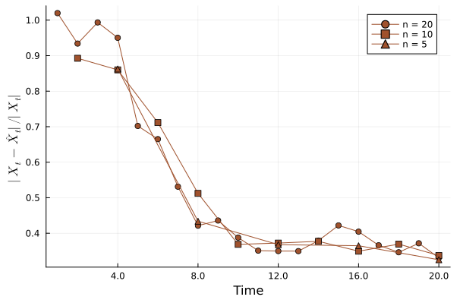


(h) GPF-II, $n = 5$.

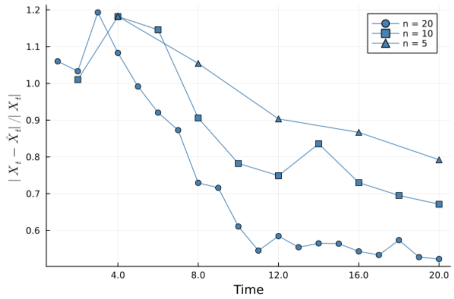
Figure 4.5.4. Filtering estimates for the two guided particle filter methods based on the dataset in 4.5.1. Left column: Heatmaps of the filtered path estimate. Right column: Heatmaps of the absolute estimation error. The first and third rows show GPF-I (our proposed method), while the second and fourth rows show GPF-II (method with guiding term in (4.49)) under the dense and sparse observation schemes, respectively.



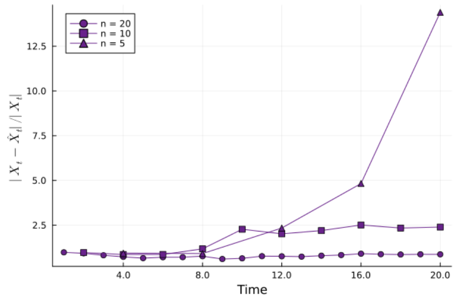
(a) All methods, $n = 20$.



(b) GPF-I, $n = 20$, 10, 5.



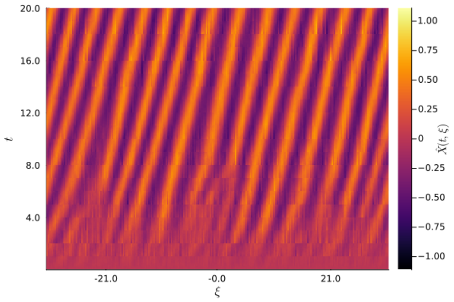
(c) GPF-II, $n = 20$, 10, 5.



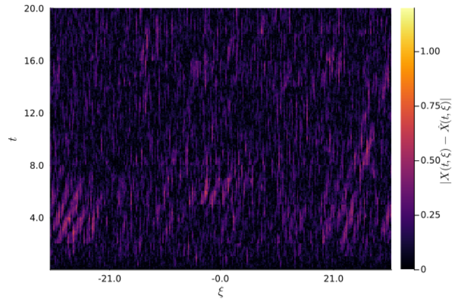
(d) UKF, $n = 20$, 10, 5.

Figure 4.5.5. Relative errors $|\hat{X}_{t_i} - X_{t_i}|/|X_{t_i}|$ at observation times t_i for the filtering estimates $\hat{X}_{t_i}$ based on the data set of Figure 4.5.2 and the three methods GPF-I (our proposed method), GPF-II (method with guiding term in (4.49)) and UKF. Subfigure (a): a comparison of estimation errors for all three methods in the dense observation case. Subfigures (b) - (d): estimation errors of individual methods across subsampled datasets.

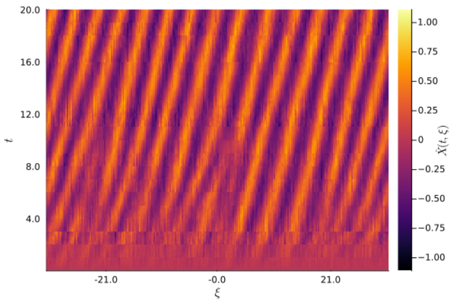
A comparison of the two particle filter methods in Figure 4.5.3(a) shows that GPF-II with proposals based on the guiding function in (4.49) slightly outperforms our proposed GPF-I in the dense observation setting. This may be explained yet again by the models convergence to an equilibrium state. Since the guiding function defined in (4.49) is derived based on the choice $\mathrm{d}Z_t = \sqrt{Q}\,\mathrm{d}W_t$ for the auxiliary process Z, once equilibrium is reached and X behaves following (4.50), such proposals are almost locally optimal. Hence, GPF-II starts outperforming GPF-I around the observation times just before $t = 8.0$ when an equilibrium state is 'fully' reached the first time. On the other hand, when observations are sparse, the quality of the proposals for the GPF-I at states away from equilibrium lead to a better performance in the cases of $n = 10$ and $n = 5$. Interestingly, as can be seen in Figure 4.5.3(b), the performance of GPF-I in this experiment increases as observations become more sparse. This coincides with findings in Llopis et al. [78] for GPF-II in the context of a different dynamical system. In our experiment, this increase in performance is drastic enough for GPF-I in the $n = 5$ case to be the best performing method over all observation schemes.
Figure 4.5.4 shows heatmaps of the filtered path reconstructions returned by the guided particle filters as well as the corresponding error heatmaps. Note that we are only interested in the filtering distribution and hence do not resample the whole path history



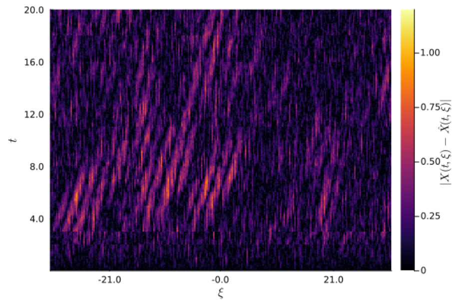
(a) GPF-I, $n = 20$.



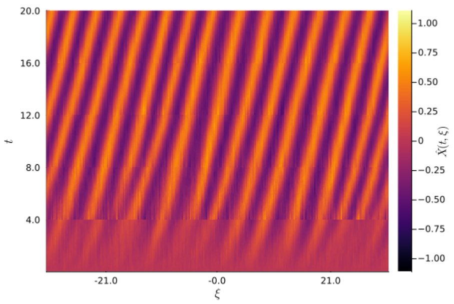
(b) GPF-I, $n = 20$.



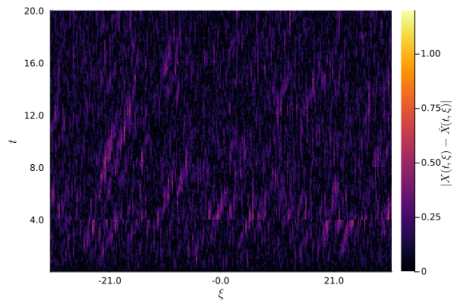
(c) GPF-II, $n = 20$.



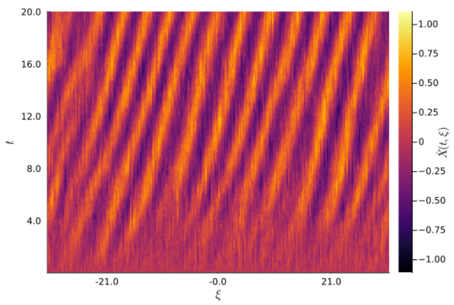
(d) GPF-II, $n = 20$.



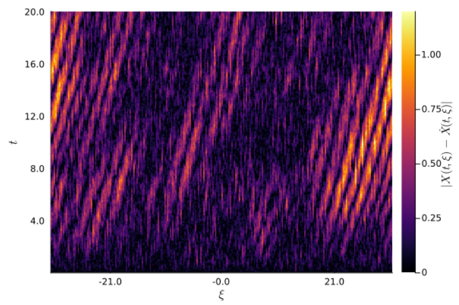
(e) GPF-I, $n = 5$.



(f) GPF-I, $n = 5$.



(g) GPF-II, $n = 5$.



(h) GPF-II, $n = 5$.

Figure 4.5.6. Filtering estimates for the two guided particle filter methods based on the dataset in 4.5.1. Left column: Heatmaps of the filtered path estimate. Right column: Heatmaps of the absolute estimation error. The first and third rows show GPF-I (our proposed method), while the second and fourth rows show GPF-II (method with guiding term in (4.49)) under the dense and sparse observation schemes, respectively.

at each filtering step. This results in the discontinuity of the paths at observation times seen in Figure 4.5.4.

It can clearly be seen that the filtering task at hand reduces to estimating the right equilibrium state early on. Once such a steady state is estimated, the particle filters struggle to correct their estimates. Only GPF-I with $n = 5$ seems to have sufficient time for the guided proposals to evolve into the correct steady state.

Results Experiment 2 For the second filtering experiment, the comparison between relative filtering errors of the three methods can be found in Figure 4.5.5. In this parametrisation with $\delta = 0.5$, the system does not converge to an equilibrium; instead, it exhibits traveling wave behaviour. Expectedly, the proposed GPF-I outperforms GPF-II over all three observation schemes. We also see the UKF performing worse than the particle filter methods, even in the dense observation scheme. Moreover, the large time steps between observations plus the nonlinear dynamics of the traveling waves states result in significantly more divergent behaviour in the case of $n = 5$ observations. On the other hand, GPF-II performs only slightly worse as observation numbers decrease, whereas GPF-I remains robust to the sparsity of observations.

This is confirmed by the heatmaps of the reconstructed paths in 4.5.6. In the dense observation scheme, performance of the filtering methods is similar. However, for $n = 5$, only the better informed proposals of GPF-I estimate the correct traveling waves at the first observation time $t_1 = 4$. Still, in contrast to the first experiment, both particle filters manage to overcome poor estimates instead of getting stuck in misestimated equilibria.

4.5.3 Smoothing and parameter estimation

Model setup and data The smoothing and parameter estimation experiments are run on the same data generating process and observation scheme as depicted in Figure 4.5.2. We assume the parameters $x_0, B, \sigma_0, \rho_0, \eta_0$ as defined in (4.48) to be known. On the other hand, the parameters η, ζ, A and δ are treated as unknown and therefore to be estimated.

Method We apply Algorithm 7 to the given dataset with $N = 15\,000$ MCMC iterations and pCN step size $\beta = 0.1$. As priors we choose

$$\eta \sim \mathrm{Unif}([0,15]),\ \zeta \sim \mathrm{Unif}([0,3]),\ A \sim \mathrm{Unif}([0,8]),\ \delta \sim \mathrm{Unif}([0,1]).$$

The parameters $\{\eta, \zeta, A, \delta\}$ are updated individually in a Gibbs sampling manner. Proposals for each parameter are chosen via an adaptive scaling Metropolis scheme (see e.g. Andrieu and Thoms [5]) as follows. At iteration j, given a step size S_j^η, say we propose η_{j+1} via

$$\eta_{j+1} = \eta_j + \exp(S_j^\eta)V$$

with $V \sim \mathcal{N}(0,1)$. Setting $\theta_{j+1} = [\eta_{j+1}, \zeta_j, A_j, \delta_j]$, we accept/reject θ_{j+1} with Metropolis acceptance probability

$$M_{j+1} = \min\left(1, \frac{\pi(\theta_{j+1})g_{\theta_{j+1}}(0,x_0)\Psi_{\theta_{j+1}}(\Gamma_{\theta_{j+1}}(x_0,V))}{\pi(\theta_j)g_{\theta_j}(0,x_0)\Psi_{\theta_j}(\Gamma_{\theta_j}(x_0,V))}\right).$$

Subsequently, the step size S_j^η is updated via

$$S_{j+1}^\eta = S_j^\eta + r_j(M_{j+1} - \alpha^*)$$

where r_j is a scaling function and α^* a target acceptance rate. The remaining parameters ζ_j, A_j, δ_j are updated equivalently in the standard Gibbs sampling fashion, each with their own adaptive step sizes $S_j^\zeta, S_j^A, S_j^\eta$. In our implementation, we set $r_j = j^{-2/3}$ and $\alpha^* = 0.234$. The step sizes are initiated at $S_0 = 1$ for all parameters. To stress-test Algorithm 7, we run the experiment twice, initialising each parameter at either end of its prior interval.

	True Value	Exp. 1 - Mean (Std)	Exp. 1 - Acceptance Ratio	Exp. 2 - Mean (Std)	Exp. 2 - Acceptance Ratio
η	**10.0**	**10.95** (1.72)	0.239	**11.26** (2.2)	0.237
ζ	**0.5**	**0.31** (0.23)	0.236	**0.33** (0.23)	0.234
A	**4.0**	**3.89** (0.17)	0.223	**3.89** (0.24)	0.226
δ	**0.5**	**0.5** (0.006)	0.211	**0.5** (0.007)	0.208

Table 4.5.1. Posterior means and standard deviations as well as acceptance ratios returned for the model parameters by Algorithm 7.

Results Figure 4.5.7 shows the trace-plots of the MCMC outputs returned by Algorithm 7 for the two different initialisations. Given the initialisations on either end of the prior intervals, the chains appear convergent remarkably fast. The true parameters fall within the range of one standard deviation of the posterior mean for each individual parameter. The exact posterior mean and standard deviation, computed after a burn-in period of 5000 iterations, are given in Table 4.5.1.
An estimate of the smoothed path, given by the mean sample path of the final 1000 MCMC iterations, is shown in Figure 4.5.8. It is noteworthy that the quality of the path reconstruction matches that of the filtering experiments, even in this setting, in which half of the model parameters are assumed to be unknown. Errors in the reconstruction also appear to be mostly due to local noise and the slight misestimation of parameters, rather than a mismatch in the structural patterns of the Amari equation. Additionally, we plot a spatially localised path of every 100th of the first 5000 sample paths of the MCMC output. Each path is localised in space by applying the observation operator L and taking the first and eighth component of the resulting vector. These correspond to the first and eighth measurement domains with centres at $\xi_1 \approx -28.9$ and $\xi_8 = 0.0$ plotted on the right-hand side of Figure 4.5.2. The localised paths as well as the true localised signal and measurements thereof are depicted in Figure 4.5.9. The figures show a clear estimation quality of the posterior samples, with deviations to the true signal being stronger in areas where measurements deviate as well.

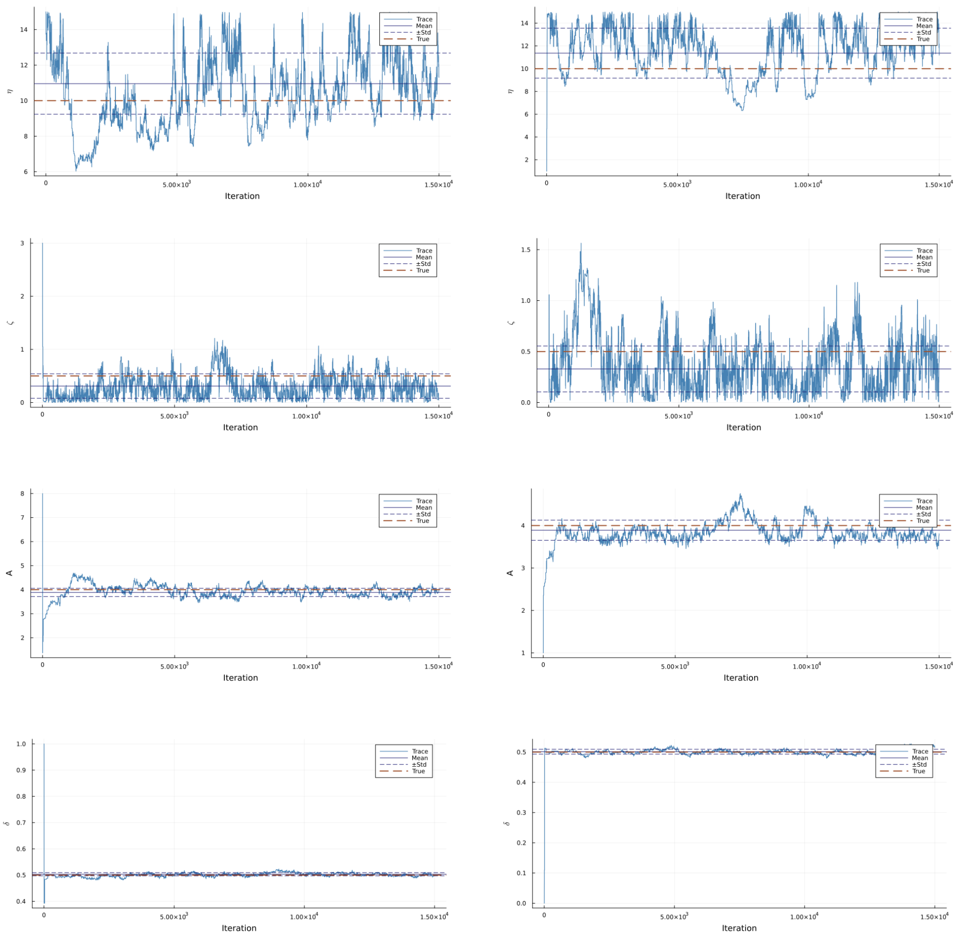


FIGURE 4.5.7. Trace-plots of the MCMC outputs returned by Algorithm 7 for various model parameters. The rows show, in order, the trace-plots of η, ζ, A and δ. The dark blue solid line represents the sample mean after a burn-in period of 5000 samples, whereas the dashed lines represent the interval spanning one empirical standard deviation above and below the sample mean. The orange lines represent the true value. Different columns relate to a different initialisation of the algorithm.

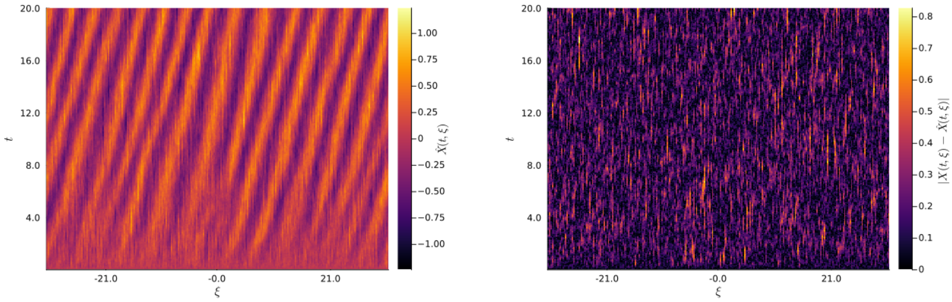


FIGURE 4.5.8. Heatmap of the smoothing estimate of X. Left: Path estimate $\hat{X}$ given by the mean of the final 1000 MCMC iterations returned by Algorithm 7. Right: Absolute error with respect to the true signal X.

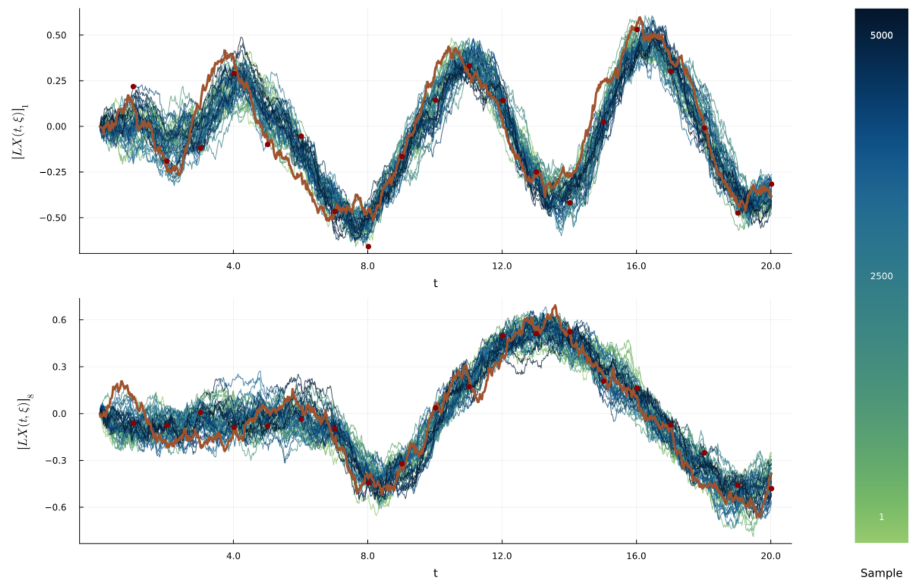


FIGURE 4.5.9. Spatially localised sample paths of the smoothing distribution of X. Spatial localisations are computed by applying the observation operator L. Green represents early samples in the Markov chain, whereas blue represents later samples. The orange line shows the true localised signal, with red dots symbolising the measurements thereof. Top: Local average over the first measurement domain around $\xi_1 \approx -28.9$. Bottom: Local average over the eighth measurement domain around $\xi_8 = 0.0$.

4.A

4.A.1 Proofs of Section 4.2.1

PROOF OF PROPOSITION 4.1. Both equalities follow from rewriting the definition of h in (4.11), whereas the continuity claims are a consequence of the dominated convergence theorem and almost sure continuity of X. □

PROOF OF PROPOSITION 4.3. We first show that the function h defined in (4.11) is space-time harmonic on any time interval $(t_{i-1}, t_i]$, i.e.

$$\mathbb{E}[h(t, X_t) \mid X_s = x] = h(s, x) \tag{4.51}$$

for any $x \in H$ and $s, t \in (t_{i-1}, t_i]$ such that $s < t$.
Indeed, from plugging in (4.11) and using the Chapman-Kolmogorov equation, it follows that

$$\begin{aligned}
\mathbb{E}[h(t, X_t) \mid X_s = x] &= \int h(t, x_t)\mu_{s,t}(x, \mathrm{d}x_t) \\
&= \int k(x_{t_i}, y_i)\kappa(x_{t_i})\mu_{s,t}(x, \mathrm{d}x_t)\mu_{t,t_i}(x_t, \mathrm{d}x_{t_i}) \\
&= \int k(x_{t_i}, y_i)\kappa(x_{t_i})\mu_{s,t_i}(x, \mathrm{d}x_{t_i}) \\
&= h(s, x),
\end{aligned}$$

where $\kappa(x_{t_i}) := \int \prod_{j=i}^{n-1} k(x_{t_{j+1}}, y_{j+1})\, \mu_{t_j,t_{j+1}}(x_{t_j}, \mathrm{d}x_{t_{j+1}})$. This gives (4.51). To show that E^h is a $\mathbb{P}$-martingale, first consider the case that $t_{i-1} < s < t \leq t_i$. Then, by the Markov property of X, the definition of E^h and the preceding display we have

$$\begin{aligned}
\mathbb{E}[E_t^h \mid \mathcal{F}_s] &= \frac{1}{C^h}\mathbb{E}\left[\left(\prod_{j=1}^{i-1} k(X_{t_j}, y_j)\right) h(t, X_t) \mid \mathcal{F}_s\right] \\
&= \frac{1}{C^h}\prod_{j=1}^{i-1} k(X_{t_j}, y_j)\mathbb{E}\left[h(t, X_t) \mid X_s\right] \\
&= \frac{1}{C^h}\prod_{j=1}^{i-1} k(X_{t_j}, y_j)h(s, X_s) = E_s^h.
\end{aligned}$$

On the other hand, if $t_{i-2} < s \leq t_{i-1} < t \leq t_i$, then it holds that

$$\begin{aligned}
\mathbb{E}[E_t^h \mid \mathcal{F}_s] &= \frac{1}{C^h}\mathbb{E}\left[\left(\prod_{j=1}^{i-1} k(X_{t_j}, y_j)\right) h(t, X_t) \mid \mathcal{F}_s\right] \\
&= \frac{1}{C^h}\left(\prod_{j=1}^{i-2} k(X_{t_j}, y_j)\right)\mathbb{E}\left[k(X_{t_{i-1}}, y_{i-1})h(t, X_t) \mid X_s\right] \\
&= \frac{1}{C^h}\left(\prod_{j=1}^{i-2} k(X_{t_j}, y_j)\right) \\
&\quad \cdot \int k(x_{t_{i-1}}, y_{i-1})h(t, x_t)\, \mu_{s,t_{i-1}}(X_s, \mathrm{d}x_{t_{i-1}})\mu_{t_{i-1},t}(x_{t_{i-1}}, \mathrm{d}x_t) \\
&= \frac{1}{C^h}\left(\prod_{j=1}^{i-2} k(X_{t_j}, y_j)\right) h(s, X_s) = E_s^h.
\end{aligned}$$

Here, we use the Markov property of X as well as the definition of E_t^h in the first step and the $\mathcal{F}_s$-measurability of $X_{t_j}, j \leq i-2$ in the second step. The last step follows by plugging in the definition of $h(t, x_t)$ and integrating out the latent variable X_t. The general case $s < t$ follows by the same arguments.

□

PROOF OF THEOREM 4.4. We begin by showing that $\mathbb{P}^h$ satisfies (4.10). For the proof of this first claim, let us write $\mathbb{P}^{h,y}$ and $E^{h,y}$ to emphasise the dependence of the measure $\mathbb{P}^h$ and process E^h on the observation $Y = y$. To prove (4.10) it suffices to show that

$$\mathbb{E}\left[\mathbb{E}^{h,Y}\left[\varphi(X_t)\right]\mathbb{1}_{\{Y\in A\}}\right] = \mathbb{E}[\varphi(X_t)\mathbb{1}_{\{Y\in A\}}] \tag{4.52}$$

for any $A \in (\mathcal{B}(\mathbb{R}^k))^n$ and bounded, measurable function φ and $t \in [0, T]$. Without loss of generality let $t \in (t_{i-1}, t_i]$ for some $i \in \{1, \ldots, n\}$. Then indeed we have that

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}^{h,Y}\left[\varphi(X_t)\right]\mathbb{1}_{\{Y\in A\}}\right] &= \int_A \mathbb{E}^{h,y}\left[\varphi(X_t)\right] p(y)\,\mathrm{d}y \\
&= \int_A \mathbb{E}\left[\varphi(X_t)\frac{1}{C^{h,y}}\left(\prod_{j=1}^{i-1} k(X_{t_j}, y_j)\right) h(t, X_t)\right] p(y)\,\mathrm{d}y \\
&= \int_A \mathbb{E}\left[\varphi(X_t)\left(\prod_{j=1}^{i-1} k(X_{t_j}, y_j)\right) h(t, X_t)\right] \mathrm{d}y \\
&= \mathbb{E}\left[\varphi(X_t)\int_A \left(\prod_{j=1}^{i-1} k(X_{t_j}, y_j)\right) h(t, X_t)\,\mathrm{d}y\right] \\
&= \mathbb{E}\left[\varphi(X_t)\mathbb{E}\left[\mathbb{1}_{\{Y\in A\}} \mid X_{t_1}, \ldots, X_{t_{i-1}}, X_t\right]\right] \\
&= \mathbb{E}\left[\varphi(X_t)\mathbb{1}_{\{Y\in A\}}\right].
\end{aligned}$$

Here, we used in the second step the definition of $\mathbb{P}^{h,y}$ and the martingale property of $E_t^{h,y}$ and in the third step the equality $C^{h,y} = p(y)$. Moreover, we used Fubini's theorem in the fourth equality as well as the identity of $p(y \mid x_{t_1}, \ldots, x_t) = \left(\prod_{j=1}^{i-1} k(x_{t_j}, y_j)\right) h(t, x_t)$ as derived in Remark 4.2 in the second to last step.
We continue to prove the second claim of Theorem 4.4. To do so, we show that the martingale E_t^h defined in (4.12) is given by

$$E_t^h = \frac{h(0, X_0)}{C^h}\mathcal{E}(M^h)_t, \quad t \in [0, T], \tag{4.53}$$

where $\mathcal{E}(M^h)$ denotes the Doléans-Dade exponential of

$$M_t^h := \int_0^t \langle Q^{\frac{1}{2}}\,\mathrm{D}_x \log h(s, X_s), \mathrm{d}W_s\rangle, \quad t \in [0, T].$$

The claim then follows by an application of the Girsanov theorem.
Let first $t \in [0, t_1]$ and denote by K the infinitesimal space-time generator of the Markov process X. Following (4.51), it holds that $Kh = 0$. From this and the Fréchet differentiability of h in x it follows with Lemma 3.3 in Pieper-Sethmacher et al. [94] that

$$h(t, X_t) = h(0, X_0) + \int_0^t \langle Q^{\frac{1}{2}}\,\mathrm{D}_x h(s, X_s), \mathrm{d}W_s\rangle, \quad t \in [0, t_1).$$

Hence the process $\bar{h}(t, X_t) := h(t, X_t)/h(0, X_0)$ is the unique solution to

$$\begin{cases} \mathrm{d}\bar{h}(t, X_t) &= \bar{h}(t, X_t)\,\mathrm{d}M_t^h \\ \bar{h}(0, X_0) &= 1. \end{cases}$$

From this and the left-continuity of h it follows that

$$h(t, X_t) = h(0, X_0)\mathcal{E}(M^h)_t, \quad t \in [0, t_1].$$

This, jointly with $E_t^h = h(t, X_t)/C^h$, $t \in [0, t_1]$, shows (4.53) on $[0, t_1]$. Now, let $t \in (t_{i-1}, t_i], i \geq 2$, and assume (4.53) holds for all $[0, t_{i-1}]$. Again, by the definition of E^h, Equation (4.51) and Lemma 3.3 in Pieper-Sethmacher et al. [94] we have that

$$\begin{aligned}
\mathrm{d}E_t^h &= \frac{1}{C^h}\left(\prod_{j=1}^{i-1} k(X_{t_j}, y_j)\right) \mathrm{d}h(t, X_t) \\
&= \frac{1}{C^h}\left(\prod_{j=1}^{i-1} k(X_{t_j}, y_i)\right) \langle Q^{\frac{1}{2}}\, \mathrm{D}_x h(t, X_t), \mathrm{d}W_t\rangle \\
&= E_t^h\, \mathrm{d}M_t^h, \quad t \in (t_{i-1}, t_i).
\end{aligned}$$

From this it is apparent that $\bar{E}_t^h := E_t^h / E_{t_{i-1}}^h$ solves

$$\begin{cases} \mathrm{d}\bar{E}_t^h &= \bar{E}_t^h\, \mathrm{d}M_t^h, \quad t \in (t_{i-1}, t_i), \\ \mathrm{d}\bar{E}_{t_{i-1}}^h &= 1. \end{cases} \tag{4.54}$$

It follows that

$$\bar{E}_t^h = \exp\left(\int_{t_{i-1}}^t \langle Q^{\frac{1}{2}}\, \mathrm{D}_x \log h(s, X_s), \mathrm{d}W_s\rangle - \frac{1}{2}\int_{t_{i-1}}^t |\langle Q^{\frac{1}{2}}\, \mathrm{D}_x \log h(s, X_s)|^2\, \mathrm{d}s\right)$$

for all $t \in (t_{i-1}, t_i]$ and hence

$$\begin{aligned}
E_t^h &= E_{t_{i-1}}^h \exp\left(\int_{t_{i-1}}^t \langle Q^{\frac{1}{2}}\, \mathrm{D}_x \log h(s, X_s), \mathrm{d}W_s\rangle - \frac{1}{2}\int_{t_{i-1}}^t |\langle Q^{\frac{1}{2}}\, \mathrm{D}_x \log h(s, X_s)|^2\, \mathrm{d}s\right) \\
&= \frac{h(0, X_0)}{C^h}\mathcal{E}(M^h)_{t_{i-1}} \\
&\quad \cdot \exp\left(\int_{t_{i-1}}^t \langle Q^{\frac{1}{2}}\, \mathrm{D}_x \log h(s, X_s), \mathrm{d}W_s\rangle - \frac{1}{2}\int_{t_{i-1}}^t |\langle Q^{\frac{1}{2}}\, \mathrm{D}_x \log h(s, X_s)|^2\, \mathrm{d}s\right) \\
&= \frac{h(0, X_0)}{C^h}\mathcal{E}(M^h)_t.
\end{aligned}$$

The claim in (4.53) then follows inductively, thereby finishing the proof. □

4.A.2 Proofs of Section 4.2.2

Proof of Theorem 4.6. *Step One.* We begin deriving L_t, R_t and α_t on the interval $(t_{n-1}, t_n]$. By definition of g in (4.17), we have $g(t_n, x) = k(x_n, y_n) = f(y_n; Lx, \Sigma)$. Moreover, from the backwards recursion (4.18) it follows that

$$\begin{aligned}
g(t, x) &= \int f(y_n; Lx_{t_n}, \Sigma)\, \nu_{t,t_n}(x, \mathrm{d}x_{t_n}) \\
&= \int f(y_n; \xi, \Sigma) f\left(\xi; LS_{t_n - t}x + L\int_t^{t_n} S_{t_n - s}a_s\, \mathrm{d}s, LQ_{t_n - t}L^*\right) \mathrm{d}\xi \\
&= \int f(y_n; \xi, \Sigma) f\left(\xi; L_t x + \alpha_t, LQ_{t_n - t}L^*\right) \mathrm{d}\xi.
\end{aligned}$$

Here we use in the second equality that $\xi = Lx_{t_n}$ under $\nu_{t,t_n}(x, \mathrm{d}x_{t_n})$ is a Gaussian measure on $\mathbb{R}^m$ with mean

$$LS_{t_n-t}x + L\int_t^{t_n} S_{t_n-s}a_s \,\mathrm{d}s = L_t x + \alpha_t$$

and covariance matrix $LQ_{t_n-t}L^*$. Now, integrating out the latent Gaussian variable ξ gives

$$g(t,x) = f(y_n; L_t x + \alpha_t, \Sigma + LQ_{t_n-t}L^*) = f(y_n; L_t x + \alpha_t, R_t),$$

with L_t, α_t and R_t as defined in (4.19) and (4.20).
Step Two. Consider $g(t_{n-1}, x)$ evaluated at the observation time t_{n-1}. From the second equality in the backwards recursion (4.18) it follows that

$$g(t_{n-1},x) = k(y_{n-1},x)\, g^+(t_{n-1},x) = f(y_{n-1}; Lx, \Sigma)\, f(y_n; L^+_{t_{n-1}} + \alpha^+_{t_{n-1}}, R^+_{t_{n-1}}).$$

Expanding the Gaussian densities on the right-hand side and simple algebra gives that

$$g(t_{n-1},x) = f(y^+_{n-1}; L_{t_{n-1}}x + \alpha_{t_{n-1}}, R_{t_{n-1}})$$

with

$$L_{t_{n-1}} = \begin{bmatrix} L \\ L^+_{t_{n-1}} \end{bmatrix}, R_{t_{n-1}} = \begin{bmatrix} \Sigma & 0 \\ 0 & R^+_{t_{n-1}} \end{bmatrix}, \alpha_{t_{n-1}} = \begin{bmatrix} 0 \\ \alpha^+_{t_{n-1}} \end{bmatrix}.$$

Step Three. The general claim follows by repeating the backwards recursion of the first two steps to obtain $g(t,x)$ on the complete interval $[0, t_n]$. □

Proof of Theorem 4.7. The proof of Theorem 4.7 follows roughly the same structure as the second part of the proof of Theorem 4.4. We show that E^g as defined in Eq. (4.23) equals the Doléans-Dade exponential

$$E^g_t = \frac{g(0,X_0)}{C^g}\mathcal{E}(M^g)_t, \quad t \in [0,T], \tag{4.55}$$

of the martingale $M^g_t := \int_0^t \langle Q^{\frac{1}{2}} G(s,X_s), \mathrm{d}W_s\rangle$. If $\mathcal{E}(M^g)_t$ is then a $\mathbb{P}$-martingale, Theorem 4.7 follows from an application of the Girsanov theorem.
First note that, following Theorem 4.6, $g(t,x)$ is bounded, continuous and Fréchet differentiable in x on any interval (t_{i-1}, t_i) with

$$\mathrm{D}_x g(t,x) = g(t,x)\,\mathrm{D}_x \log g(t,x) = g(t,x)G(t,x).$$

Hence, denoting by K the infinitesimal space-time generator of X, it follows from Manca [82], Theorem 4.1 that $Kg(t,x) = \langle \tilde{F}(t,x), \mathrm{D}_x g(t,x)\rangle$ with $\tilde{F}(t,x) := F(t,x) - a_t$. Lemma 3.3 of Pieper-Sethmacher et al. [94] therefore gives for any $t \in [0,t_1)$ that

$$g(t,X_t) = g(0,X_0) + \int_0^t \langle \tilde{F}(s,X_s), \mathrm{D}_x g(s,X_s)\rangle \,\mathrm{d}s + \int_0^t \langle Q^{\frac{1}{2}}\,\mathrm{D}_x g(s,X_s), \mathrm{d}W_s\rangle. \tag{4.56}$$

Consequently, plugging in E^g as in (4.23) and applying the integration by parts for semimartingales we have on $[0, t_1)$ that

$$\begin{aligned}
\mathrm{d}E_t^g &= \frac{1}{C^g}\,\mathrm{d}\left(g(t,X_t)\exp\left(-\int_0^t \langle \tilde{F}(s,X_s), G(s,X_s)\rangle\;\mathrm{d}s\right)\right)\\
&= \frac{1}{C^g}\exp\left(-\int_0^t \langle \tilde{F}(s,X_s), G(s,X_s)\rangle\;\mathrm{d}s\right)\\
&\quad\cdot\left[\langle \tilde{F}(t,X_t), \mathrm{D}_x g(t,X_t)\rangle\,\mathrm{d}t + \langle Q^{\frac{1}{2}}\,\mathrm{D}_x g(t,X_t), \mathrm{d}W_t\rangle\right]\\
&\quad - \frac{1}{C^g} g(t,X_t)\exp\left(-\int_0^t \langle \tilde{F}(s,X_s), G(s,X_s)\rangle\;\mathrm{d}s\right)\langle \tilde{F}(t,X_t), G(t,X_t)\rangle\;\mathrm{d}t\\
&= \frac{1}{C^g}\exp\left(-\int_0^t \langle \tilde{F}(s,X_s), G(s,X_s)\rangle\;\mathrm{d}s\right)\langle Q^{\frac{1}{2}}\,\mathrm{D}_x g(t,X_t), \mathrm{d}W_t\rangle\\
&= \frac{g(t,X_t)}{C^g}\exp\left(-\int_0^t \langle \tilde{F}(s,X_s), G(s,X_s)\rangle\;\mathrm{d}s\right)\langle Q^{\frac{1}{2}} G(t,X_t), \mathrm{d}W_t\rangle\\
&= E_t^g\,\mathrm{d}M_t^g.
\end{aligned}$$

This shows that $\bar{E}_t^g := E_t^g/E_0^g = \mathcal{E}(M^g)_t$ and hence (4.55) for all $t \in [0, t_1]$. The extension onto the complete time interval $[0, T]$ follows by the same inductive argument as in the proof of Theorem 4.4. Lastly, to show that $\mathcal{E}(M^g)$ is a proper martingale, it suffices to note that $x \mapsto G(t, x)$ is Lipschitz continuous in x, uniformly on $[0, T]$. The martingale property of $\mathcal{E}(M^g)$ then follows from Pieper-Sethmacher et al. [94], Lemma C.1. □

4.A.3 Proofs of Section 4.2.3

Proof of Theorem 4.9. We start by showing that U_t is a solution to (4.29) in a mild sense, i.e. that U_t satisfies

$$U_t = S_{t_i-t}^* U_{t_i} S_{t_i-t} - \int_t^{t_i} S_{s-t}^* U_s Q U_s S_{s-t}\;\mathrm{d}s, \quad t \in (t_{i-1}, t_i], \tag{4.57}$$

and $U_{t_i} = L^*\Sigma L + U_{t_i}^+$. The latter follows from plugging in L_{t_i} and R_{t_i} as given in Theorem 4.6 into the definition of U_t. To show (4.57), note that R_t as defined in Theorem 4.6 satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} R_t = \frac{\mathrm{d}}{\mathrm{d}t}\left(L_{t_i} Q_{t_i-t} L_{t_i}^*\right) = -L_{t_i} S_{t_i-t} Q S_{t_i-t}^* L_{t_i}^* = -L_t Q L_t^*. \tag{4.58}$$

Here, we used that $\mathrm{d}Q_t = S_t Q S_t^*\,\mathrm{d}t$ following the definition of Q_t in (4.16). It follows

$$\frac{\mathrm{d}}{\mathrm{d}t} R_t^{-1} = -R_t^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t} R_t\right) R_t^{-1} = R_t^{-1} L_t Q L_t^* R_t^{-1} \tag{4.59}$$

and thus

$$R_t^{-1} = R_{t_i}^{-1} - \int_t^{t_i} R_s^{-1} L_s Q L_s^* R_s^{-1}\;\mathrm{d}s. \tag{4.60}$$

Hence, plugging (4.60) into (4.27) and using that $L_t = L_s S_{s-t}$ for all $s \in [t, t_i]$, this gives

$$\begin{aligned} L_t^* R_t^{-1} L_t &= L_t^* R_{t_i}^{-1} L_t - L_t^* \left(\int_t^{t_i} R_s^{-1} L_s Q L_s^* R_s^{-1} \, \mathrm{d}s \right) L_t \\ &= S_{t_i-t}^* L_{t_i}^* R_{t_i}^{-1} L_{t_i} S_{t_i-t} - \int_t^{t_i} S_{s-t}^* L_s^* R_s^{-1} L_s Q L_s^* R_s^{-1} L_s S_{s-t} \, \mathrm{d}s \\ &= S_{t_i-t}^* U_{t_i} S_{t_i-t} - \int_t^{t_i} S_{s-t}^* U_s Q U_s S_{s-t} \, \mathrm{d}s. \end{aligned}$$

This shows (4.57) and uniqueness of the mild solution follows from uniqueness of the infinite-dimensional forward Ricatti equation, see for example Burns and Rautenberg [18], Theorem 3.6 as well as Bensoussan et al. [10], Chapter 2.2 and references within. To show that V_t satisfies (4.30) in a mild sense, first note that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \left(R_t^{-1} (y_i^+ - \alpha_t) \right) &= \left(\frac{\mathrm{d}}{\mathrm{d}t} R_t^{-1} \right) (y_i^+ - \alpha_t) - R_t^{-1} \left(\frac{\mathrm{d}}{\mathrm{d}t} \alpha_t \right) \\ &= R_t^{-1} L_t Q L_t^* R_t^{-1} (y_i^+ - \alpha_t) + R_t^{-1} L_t a_t \\ &= R_t^{-1} L_t Q V_t + R_t^{-1} L_t a_t, \quad t \in (t_{i-1}, t_i). \end{aligned} \tag{4.61}$$

Here we used (4.59) and the definition of α_t in the second equality and the definition of V_t in the last. This gives

$$R_t^{-1} (y_i^+ - \alpha_t) = R_{t_i}^{-1} (y_i^+ - \alpha_{t_i}) - \int_t^{t_i} R_s^{-1} L_s Q V_s + R_s^{-1} L_s a_s \, \mathrm{d}s, \quad t \in (t_{i-1}, t_i),$$

and consequently, using that $L_t^* = S_{s-t}^* L_s^*$ for all $s \in [t, t_i]$ and the definition of U_s,

$$\begin{aligned} V_t &= L_t^* R_t^{-1} (y_i^+ - \alpha_t) \\ &= L_t^* R_{t_i}^{-1} (y_i^+ - \alpha_{t_i}) - L_t^* \left(\int_t^{t_i} R_s^{-1} L_s Q V_s + R_s^{-1} L_s a_s \, \mathrm{d}s \right) \\ &= S_{t_i-t}^* L_{t_i}^* R_{t_i}^{-1} (y_i^+ - \alpha_{t_i}) - \left(\int_t^{t_i} S_{s-t}^* L_s^* R_s^{-1} L_s Q V_s + S_{s-t}^* L_s^* R_s^{-1} L_s a_s \, \mathrm{d}s \right) \\ &= S_{t_i-t}^* V_{t_i} - \left(\int_t^{t_i} S_{s-t}^* U_s Q V_s + S_{s-t}^* U_s a_s \, \mathrm{d}s \right), \quad t \in (t_{i-1}, t_i). \end{aligned}$$

This, together with the fact that $V_{t_i} = L^* \Sigma^{-1} y_i + V_{t_i}^+$ follows from plugging in L_{t_i}, R_{t_i} and α_{t_i} as given by Theorem 4.6 into the definition of V_t, shows that V_t is a mild solution to (4.30). Moreover, uniqueness of V is a direct consequence of classical results concerning the uniqueness of mild solutions to the abstract Cauchy problem. □

Proof of Proposition 4.12. The first claim follows from taking the log of $g(t, x)$ in (4.21). To derive the ODE (4.31), first note that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \det(R_t) &= \det(R_t) \operatorname{tr} \left[R_t^{-1} \left(\frac{\mathrm{d}}{\mathrm{d}t} R_t \right) \right] \\ &= -\det(R_t) \operatorname{tr} \left[R_t^{-1} L_t Q L_t^* \right] \\ &= -\det(R_t) \operatorname{tr} \left[U_t Q \right]. \end{aligned}$$

on any (t_{i-1}, t_i). Here, we use (4.58) in the second step and the definition of U_t in the last. It follows that

$$\frac{\mathrm{d}}{\mathrm{d}t} \log(\det(R_t)) = -\operatorname{tr} \left[U_t Q \right]. \tag{4.62}$$

Furthermore, plugging in $\mathrm{d}\alpha_t = -L_t a_t \,\mathrm{d}t$ and (4.59) it holds that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle y_i^+ - \alpha_t, R_t^{-1}(y_i^+ - \alpha_t)\right\rangle &= \left\langle \frac{\mathrm{d}}{\mathrm{d}t}(y_i^+ - \alpha_t), R_t^{-1}(y_i^+ - \alpha_t)\right\rangle \\ &\quad + \left\langle y_i^+ - \alpha_t, \frac{\mathrm{d}}{\mathrm{d}t}\left(R_t^{-1}(y_i^+ - \alpha_t)\right)\right\rangle \\ &= \left\langle L_t a_t, R_t^{-1}(y_i^+ - \alpha_t)\right\rangle \\ &\quad + \left\langle y_i^+ - \alpha_t, R_t^{-1} L_t Q L_t^* R_t^{-1}(y_i^+ - \alpha_t) + R_t^{-1} L_t a_t\right\rangle \\ &= 2\left\langle a_t, L_t^* R_t^{-1}(y_i^+ - \alpha_t)\right\rangle \\ &\quad + \left\langle y_i^+ - \alpha_t, R_t^{-1} L_t Q L_t^* R_t^{-1}(y_i^+ - \alpha_t)\right\rangle \\ &= 2\left\langle a_t, V_t\right\rangle + \left\langle V_t, Q V_t\right\rangle. \end{aligned} \tag{4.63}$$

Hence, combining (4.62) and (4.63) gives

$$\frac{\mathrm{d}}{\mathrm{d}t} c_t = \frac{1}{2}\operatorname{tr}\left[U_t Q\right] - \langle a_t, V_t\rangle - \frac{1}{2}\langle V_t, Q V_t\rangle, \quad t \in (t_{i-1}, t_i).$$

Lastly, the expression for the terminal condition c_{t_i} follows from plugging R_{t_i} and α_{t_i} as given in Theorem 4.6 into the definition of c_t. □

4.B

4.B.1 Proofs of Section 4.5

PROOF OF PROPOSITION 4.19. As noted in Remark 4.13, the function G_i for the one-step-ahead guiding distribution is given by

$$G_i(t, x) = L_t^* R_t^{-1}\left(y_i - L_t x\right), \quad t \in [t_{i-1}, t_i],$$

with $L_t = L S_{t_i - t}$ and $R_t = \Sigma + L Q_{t_i - t} L^*$. Let us define, with slight abuse of notation, $\Delta_i := t_i - t$.
Given $Ax = -x$, it holds that $(S_t)_t$ is the semigroup given by $S_{\Delta_i} = \exp(-\Delta_i)$. Hence, following the definition of Q_t given in (4.16), we have

$$Q_{\Delta_i} = \int_0^{\Delta_i} \exp(-2s) Q \,\mathrm{d}s = \frac{1 - \exp(-2\Delta_i)}{2} Q.$$

Plugging this into the expression of G_i above gives the first claim of Proposition 4.19. The matrix representation of LQL^* in the second claim can be deduced from the fact that, for any $y \in \mathbb{R}^m$,

$$\begin{aligned}(LQL^*)y &= LQ\left(\sum_{i=1}^m y_i \omega_i\right) \\ &= L\left(\sum_{i=1}^m y_i \sum_{l=1}^\infty q_l \langle \omega_i, e_l\rangle e_l\right) \\ &= \left(\sum_{i=1}^m y_i \sum_{l=1}^\infty q_l \langle \omega_i, e_l\rangle \langle \omega_j, e_l\rangle\right)_{j=1}^m \in \mathbb{R}^m.\end{aligned}$$

□

Chapter 5

Conclusion

This thesis developed Bayesian approaches to statistical inference for stochastic partial differential equations. A unifying framework for conditioning and guiding of SPDEs has been introduced, based on a class of exponential measure transformations. This enabled the design of Bayesian computational methodology for both state and parameter estimation in the setting of discretely and partially observed semilinear SPDEs.
The results obtained in this thesis therefore fill a gap in the emerging literature on statistical inference for SPDEs, which has so far been concerned mostly with parameter estimation from a frequentist perspective and state estimation for linear SPDEs in the context of spatio-temporal statistics. By enabling non-linear dynamics in physically inspired spatio-temporal models and calibration of SPDE parameters based on incomplete and noisy data, the developed methodology broadens the applicability of SPDEs in a wide range of fields such as neuroscience, fluid dynamics, and finance.

In the main contribution of Chapter 2, a class of exponential measure transformations for the mild solution X to a semilinear SPDE was derived. The changed measure $\mathbb{P}^h$ depends on a function h in the domain of the infinitesimal generator L of X, defined in the topology of bounded pointwise convergence on the space of continuous function of at most polynomial growth. Conditions on h were established under which $\mathbb{P}^h$ is a Girsanov-type change of measure and X under $\mathbb{P}^h$ the mild solution to an SPDE with an additional drift term.
Different choices of h and how they impact the process X under the changed measure $\mathbb{P}^h$ were studied. In particular, an SPDE for the process X, conditioned on hitting a terminal state $X_T = y$ - the so-called infinite-dimensional diffusion bridge - was derived. This generalised results that were previously only obtained for linear SPDEs. Moreover, the guided process, a tractable process steered towards a terminal state, was lifted from the finite-dimensional to the infinite-dimensional setting. These transformations constituted the theoretical foundation for the statistical methodology developed throughout the remainder of this thesis.

In Chapter 3, methodology to sample from the law of the infinite-dimensional diffusion bridge was developed. Sampling of diffusion bridges serves as a fundamental building block for computational Bayesian approaches to statistical inference for discretely observed S(P)DEs. From the finite-dimensional setting, this is known to be a challenging problem due to an intractable drift term in the diffusion bridge equation.
The main result of Chapter 3 provided conditions for the absolute continuity between the laws of the guided process and the diffusion bridge. Moreover, a tractable Radon–Nikodym derivative was derived. This established the law of the guided process as a valid proposal distribution for the infinite-dimensional diffusion bridge law. As a consequence,

weighted samples of the diffusion bridge can be generated and subsequently used within an importance sampling or Metropolis-Hastings scheme that target the bridge distribution. A Metropolis-Hastings sampler of this form was formulated, making use of localised random walk proposals in Wiener space based on the preconditioned Crank–Nicolson scheme. The performance of the sampler in simulating the infinite-dimensional diffusion bridge was validated in numerical experiments on a class of stochastic reaction-diffusion equations.

Chapter 4 introduced novel methodology to address state estimation, both in the online (filtering) and the offline (smoothing) setting, as well as parameter estimation for discretely observed infinite-dimensional diffusion processes. The main contribution concerning the filtering problem was to show how the infinite-dimensional guided process can be incorporated as a proposal distribution in a sequential Monte Carlo scheme. This generalised a class of proposal distributions previously used in the literature. The resulting guided particle filter was shown to outperform established methods in a case study for the stochastic Amari equation.
To address the smoothing problem and parameter estimation, the change of measure approach of Chapter 2 was extended to the setting of multiple observations. This led to a unifying framework for conditioning and guiding of infinite-dimensional diffusion processes, based on complete sets of observations over finite time intervals. A computationally efficient way to simulate the guided process by solving an infinite-dimensional backwards Ricatti equation was derived. To sample from the joint posterior of the smoothed process and unknown model parameters, a reparametrisation of the conditioning was introduced, leading to the formulation of a latent path measure that is absolutely continuous with respect to the Wiener measure. A Gibbs sampler based on this reparametrisation and the preconditioned Crank-Nicolson scheme was presented. In numerical experiments on the Amari equation, the algorithm successfully estimated model parameters that were previously assumed to be known in the literature due to challenges in their estimation.

Bibliography

[1] Agarwal, R. P., S. Deng, and W. Zhang (2005). Generalization of a retarded gronwall-like inequality and its applications. *Applied Mathematics and Computation 165*(3), 599–612.

[2] Altmeyer, R., T. Bretschneider, J. Janák, and M. Reiß (2022). Parameter estimation in an spde model for cell repolarization. *SIAM/ASA Journal on Uncertainty Quantification 10*(1), 179–199.

[3] Altmeyer, R., I. Cialenco, and G. Pasemann (2023). Parameter estimation for semi-linear spdes from local measurements. *Bernoulli 29*(3), 2035–2061.

[4] Altmeyer, R. and M. Reiß (2020). Nonparametric estimation for linear spdes from local measurements. *Annals of Applied Probability, to appear*.

[5] Andrieu, C. and J. Thoms (2008). A tutorial on adaptive mcmc. *Statistics and computing 18*, 343–373.

[6] Aronson, B. D. G. and J. Oser (1967). Bounds for the fundamental solution of a parabolic equation. *Bulletin of the American Mathematical Society 73*, 890–896.

[7] Asensio, M. and L. Ferragut (2002). On a wildland fire model with radiation. *International Journal for Numerical Methods in Engineering 54*(1), 137–157.

[8] Barbu, V., G. Da Prato, and A. Debussche (2004). The kolmogorov equation associated to the stochastic navier–stokes equations in 2d. *Infinite Dimensional Analysis, Quantum Probability and Related Topics 7*(02), 163–182.

[9] Baudoxin, F. (2002). Conditioned stochastic differential equations: theory, examples and application to finance. *Stochastic Processes and their Applications 100*(1), 109–145.

[10] Bensoussan, A., G. Da Prato, M. C. Delfour, and S. K. Mitter (2007). *Representation and control of infinite dimensional systems*, Volume 1. Springer.

[11] Benth, F. E. and F. Paraschiv (2018). A space-time random field model for electricity forward prices. *Journal of Banking & Finance 95*, 203–216.

[12] Beskos, A., D. Crisan, A. Jasra, K. Kamatani, and Y. Zhou (2017). A stable particle filter for a class of high-dimensional state-space models. *Advances in Applied Probability 49*(1), 24–48.

[13] Beskos, A., F. Pinski, J. Sanz-Serna, and A. Stuart (2011). Hybrid monte carlo on hilbert spaces. *Stochastic Processes and their Applications 121*(10), 2201–2230.

[14] Beskos, A., G. Roberts, A. Stuart, and J. Voss (2008). Mcmc methods for diffusion bridges. *Stochastics and Dynamics 08*(03), 319–350.

[15] Bibinger, M. and M. Trabs (2020). Volatility estimation for stochastic pdes using high-frequency observations. *Stochastic Processes and their Applications 130*(5), 3005–3052.

[16] Bierkens, J., F. van der Meulen, and M. Schauer (2020). Simulation of elliptic and hypo-elliptic conditional diffusions. *Advances in Applied Probability 52*(1), 173–212.

[17] Borovitskiy, V., A. Terenin, P. Mostowsky, et al. (2020). Matérn gaussian processes on riemannian manifolds. *Advances in Neural Information Processing Systems 33*, 12426–12437.

[18] Burns, J. A. and C. N. Rautenberg (2015). Solutions and approximations to the riccati integral equation with values in a space of compact operators. *SIAM Journal on Control and Optimization 53*(5), 2846–2877.

[19] Carmona, R. and M. R. Tehranchi (2007). *Interest rate models: an infinite dimensional stochastic analysis perspective*. Springer Science & Business Media.

[20] Cerrai, S. (1994). A hille-yosida theorem for weakly continuous semigroups. *Semigroup Forum 49*(1), 349–367.

[21] Cerrai, S. (1995). Weakly continuous semigroups in the space of functions with polynomial growth. *Dynam. Syst. Appl 4*, 351–372.

[22] Cerrai, S. and F. Gozzi (1995). Strong solutions of cauchy problems associated to weakly continuous semigroups. *Differential and Integral Equations 8*, 465 – 486.

[23] Cheung, J. (2025). On the approximation of operator-valued riccati equations in hilbert spaces. *Journal of Mathematical Analysis and Applications 547*(1), 129250.

[24] Chojnowska-Michalik, A. and B. Goldys (1996). On regularity properties of nonsymmetric ornstein-uhlenbeck semigroup in lp spaces. *Stochastics: An International Journal of Probability and Stochastic Processes 59*(3-4), 183–209.

[25] Chong, C. (2020). High-frequency analysis of parabolic stochastic PDEs. *The Annals of Statistics 48*(2), 1143 – 1167.

[26] Chopin, N., A. Fulop, J. Heng, and A. H. Thiery (2023). Computational doob h-transforms for online filtering of discretely observed diffusions. In *International Conference on Machine Learning*, pp. 5904–5923. PMLR.

[27] Chopin, N., O. Papaspiliopoulos, et al. (2020). *An introduction to sequential Monte Carlo*, Volume 4. Springer.

[28] Cialenco, I. (2018). Statistical inference for spdes: an overview. *Statistical Inference for Stochastic Processes*.

[29] Cialenco, I. and N. Glatt-Holtz (2011). Parameter estimation for the stochastically perturbed navier–stokes equations. *Stochastic Processes and their Applications 121*, 701–724.

[30] Cialenco, I. and Y. Huang (2020). A note on parameter estimation for discretely sampled spdes. *Stochastics and Dynamics 20*, 1 – 13.

[31] Cialenco, I., H.-J. Kim, and G. Pasemann (2024). Statistical analysis of discretely sampled semilinear spdes: a power variation approach. *Stochastics and Partial Differential Equations: Analysis and Computations 12*(1), 326–351.

[32] Clark, J. (1990). The simulation of pinned diffusions. In *29th IEEE conference on decision and control*, pp. 1418–1420. IEEE.

[33] Clarotto, L., D. Allard, T. Romary, and N. Desassis (2024). The spde approach for spatio-temporal datasets with advection and diffusion. *Spatial Statistics*, 100847.

[34] Cont, R. (2005). Modeling term structure dynamics: an infinite dimensional approach. *International Journal of theoretical and applied finance 8*(03), 357–380.

[35] Cook, B. J., A. D. Peterson, W. Woldman, and J. R. Terry (2022). Neural field models: A mathematical overview and unifying framework. *Mathematical Neuroscience and Applications 2*.

[36] Coombes, S., P. beim Graben, R. Potthast, and J. Wright (2014). *Neural fields: theory and applications*. Springer.

[37] Cotter, S. L., G. O. Roberts, A. M. Stuart, and D. White (2013). MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science 28*(3), 424 – 446.

[38] Craigmile, P., R. Herbei, G. Liu, and G. Schneider (2023). Statistical inference for stochastic differential equations. *Wiley Interdisciplinary Reviews: Computational Statistics 15*(2), e1585.

[39] Da Prato, G. (2004). *Kolmogorov equations for stochastic PDEs*. Springer Science & Business Media.

[40] Da Prato, G. and A. Debussche (2002). Two-dimensional navier–stokes equations driven by a space–time white noise. *Journal of Functional Analysis 196*(1), 180–210.

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

[42] Dawes, J. and M. Souza (2013). A derivation of holling's type i, ii and iii functional responses in predator–prey systems. *Journal of theoretical biology 327*, 11–22.

[43] Deligiannidis, G., A. Doucet, and M. K. Pitt (2018, 07). The correlated pseudo-marginal method. *Journal of the Royal Statistical Society Series B: Statistical Methodology 80*(5), 839–870.

[44] Delyon, B. and Y. Hu (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Applications 116*(11), 1660–1675.

[45] Di Nunno, G., S. Ortiz–Latorre, and A. Petersson (2023). Spde bridges with observation noise and their spatial approximation. *Stochastic Processes and their Applications 158*, 170–207.

[46] Doob, J. L. (1984). *Classical potential theory and its probabilistic counterpart*, Volume 262. Springer.

[47] Doucet, A., N. De Freitas, N. J. Gordon, et al. (2001). *Sequential Monte Carlo methods in practice*, Volume 1. Springer.

[48] Earth Resources Observation and Science (EROS) Center (2020). Landsat 8-9 operational land imager / thermal infrared sensor level-2, collection 2. Date Accessed: 2025-07-03.

[49] Elerian, O., S. Chib, and N. Shephard (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica 69*(4), 959–993.

[50] Eraker, B. (2001). Mcmc analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics 19*(2), 177–191.

[51] Ethier, S. N. and T. G. Kurtz (2009). *Markov processes: characterization and convergence*. John Wiley & Sons.

[52] Fabbri, G., F. Gozzi, and A. Swiech (2017). Stochastic optimal control in infinite dimension. *Probability and Stochastic Modelling. Springer*.

[53] Faugeras, O. and J. Inglis (2015). Stochastic neural field equations: a rigorous footing. *Journal of mathematical biology 71*(2), 259–300.

[54] Fearnhead, P., O. Papaspiliopoulos, and G. O. Roberts (2008). Particle filters for partially observed diffusions. *Journal of the Royal Statistical Society Series B: Statistical Methodology 70*(4), 755–777.

[55] Gaudlitz, S. (2023). Non-parametric estimation of the reaction term in semi-linear spdes with spatial ergodicity. *arXiv preprint arXiv:2307.05457*.

[56] Gaudlitz, S. and M. Reiß (2023). Estimation for the reaction term in semi-linear spdes under small diffusivity. *Bernoulli 29*(4), 3033–3058.

[57] Gibson, J. S. (1983). Linear-quadratic optimal control of hereditary differential systems: Infinite dimensional riccati equations and numerical approximations. *SIAM Journal on Control and Optimization 21*(1), 95–139.

[58] Goldys, B. and M. Kocan (2001). Diffusion semigroups in spaces of continuous functions with mixed topology. *Journal of Differential Equations 173*(1), 17–39.

[59] Goldys, B. and B. Maslowski (2008). The ornstein–uhlenbeck bridge and applications to markov semigroups. *Stochastic Processes and their Applications 118*(10), 1738–1767.

[60] Graham, M. M., A. H. Thiery, and A. Beskos (2022). Manifold markov chain monte carlo methods for bayesian inference in diffusion models. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(4), 1229–1256.

[61] Hairer, M. (2009). An introduction to stochastic pdes. *arXiv preprint arXiv:0907.4178*.

[62] Hairer, M., A. M. Stuart, and J. Voss (2011). Sampling conditioned hypoelliptic diffusions. *The Annals of Applied Probability 21*(2), 669 – 698.

[63] Heng, J., V. De Bortoli, A. Doucet, and J. Thornton (2021). Simulating diffusion bridges with score matching. *arXiv preprint arXiv:2111.07243*.

[64] Hildebrandt, F. and M. Trabs (2021). Parameter estimation for SPDEs based on discrete observations in time and space. *Electronic Journal of Statistics 15*(1), 2716 – 2776.

[65] Hildebrandt, F. and M. Trabs (2023). Nonparametric calibration for stochastic reaction–diffusion equations based on discrete observations. *Stochastic Processes and their Applications 162*, 171–217.

[66] Holling, C. S. (1959). The components of predation as revealed by a study of small-mammal predation of the european pine sawfly. *The Canadian Entomologist 91*(5), 293–320.

[67] Huebner, M., R. Khasminskii, and B. Rozovskii (1993). Two examples of parameter estimation for stochastic partial differential equations. *Stochastic processes: A Festschrift in Honour of G. Kallianpur*, 149 – 160.

[68] Huebner, M. and B. Rozovskii (1995). On asymptotic properties of maximum likelihood estimators for parabolic stochastic pde's. *Probability Theory and Related Fields volume 103*, 143 – 163.

[69] Infante, S., A. Hernández, et al. (2019). Stochastic models to estimate population dynamics. *Statistics, Optimization & Information Computing 7*(2), 311–328.

[70] Janák, J. and M. Reiß (2024). Parameter estimation for the stochastic heat equation with multiplicative noise from local measurements. *Stochastic Processes and their Applications 175*, 104385.

[71] Jasra, A., K. J. Law, and F. Yu (2022). Unbiased filtering of a class of partially observed diffusions. *Advances in Applied Probability 54*(3), 661–687.

[72] Johnson, K. A. and R. S. Goody (2011). The original michaelis constant: translation of the 1913 michaelis–menten paper. *Biochemistry 50*(39), 8264–8269.

[73] Jones, R. H. and Y. Zhang (1997). Models for continuous stationary space-time processes. In *Modelling longitudinal and spatially correlated data*, pp. 289–298. Springer.

[74] Kaino, Y. and M. Uchida (2021). Parametric estimation for a parabolic linear spde model based on discrete observations. *Journal of Statistical Planning and Inference 211*, 190–220.

[75] Lang, O., P. J. Van Leeuwen, D. Crisan, and R. Potthast (2022). Bayesian inference for fluid dynamics: a case study for the stochastic rotating shallow water model. *Frontiers in Applied Mathematics and Statistics 8*, 949354.

[76] Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology 73*(4), 423–498.

[77] Liu, X., K. Yeo, and S. Lu (2022). Statistical modeling for spatio-temporal data from stochastic convection-diffusion processes. *Journal of the American Statistical Association 117*(539), 1482–1499.

[78] Llopis, F. P., N. Kantas, A. Beskos, and A. Jasra (2018). Particle filtering for stochastic navier–stokes signal observed with linear additive noise. *SIAM Journal on Scientific Computing 40*(3), A1544–A1565.

[79] Lord, G. J., C. E. Powell, and T. Shardlow (2014). *An introduction to computational stochastic PDEs*, Volume 50. Cambridge University Press.

[80] Lototsky, S. V. (2009). Statistical inference for stochastic parabolic equations: a spectral approach. *Publicacions Matemàtiques 53*(1), 3 – 45.

[81] Manca, L. (2008). Kolmogorov equations for measures. *Journal of Evolution Equations 8*, 231–262.

[82] Manca, L. (2009). Fokker–planck equation for kolmogorov operators with unbounded coefficients. *Stochastic analysis and applications 27*(4), 747–769.

[83] Mandel, J., L. S. Bennethum, J. D. Beezley, J. L. Coen, C. C. Douglas, M. Kim, and A. Vodacek (2008). A wildland fire model with data assimilation. *Mathematics and Computers in Simulation 79*(3), 584–606.

[84] Mao, X. (2007). *Stochastic Differential Equations and Applications* (2nd ed.). Woodhead Publishing.

[85] Mider, M., M. Schauer, and F. van der Meulen (2021). Continuous-discrete smoothing of diffusions. *Electronic Journal of Statistics 15*(2), 4295 – 4342.

[86] NASA VIIRS Land Science Team (2021). *VIIRS (NOAA-21/JPSS-2) I Band 375 m Active Fire Product NRT (Vector Data)*. NASA LANCE MODIS at the MODAPS. Date Accessed: 2025-07-03.

[87] Neal, R. M. (1999, 08). Regression and Classification Using Gaussian Process Priors. In *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting June 6-10, 1998*. Oxford University Press.

[88] Owens, D. and M. O'Kane. Final report of the nsw bushfire inquiry.

[89] Palmowski, Z. and T. Rolski (2002). A technique for exponential change of measure for markov processes. *Bernoulli*, 767–785.

[90] Papaspiliopoulos, O. and G. Roberts (2012). Importance sampling techniques for estimation of diffusion models. *Statistical methods for stochastic differential equations 124*, 311–340.

[91] Papaspiliopoulos, O., G. O. Roberts, and O. Stramer (2013). Data augmentation for diffusions. *Journal of Computational and Graphical Statistics 22*(3), 665–688.

[92] Pasemann, G., S. Flemming, S. Alonso, C. Beta, and W. Stannat (2021). Diffusivity estimation for activator–inhibitor models: Theory and application to intracellular dynamics of the actin cytoskeleton. *Journal of nonlinear science 31*(3), 59.

[93] Pasemann, G. and W. Stannat (2020). Drift estimation for stochastic reaction-diffusion systems. *Electronic Journal of Statistics 14*, 547 –579.

[94] Pieper-Sethmacher, T., F. van der Meulen, and A. van der Vaart (2025). On a class of exponential changes of measure for stochastic pdes. *Stochastic Processes and their Applications 185*, 104630.

[95] Pieper-Sethmacher, T., F. van der Meulen, and A. van der Vaart (2025). Simulation of infinite-dimensional diffusion bridges. *arXiv preprint arXiv:2503.13177*.

[96] Priola, E. (1999). On a class of markov type semigroups in spaces of uniformly continuous and bounded functions. *Studia Mathematica 136*, 3.

[97] Roberts, G. O. and O. Stramer (2001). On inference for partially observed nonlinear diffusion models using the metropolis–hastings algorithm. *Biometrika 88*(3), 603–621.

[98] Rogers, L. C. G. and D. Williams (2000). *Diffusions, Markov processes, and martingales: Itô calculus*, Volume 2. Cambridge university press.

[99] Ruzayqat, H., A. Er-Raiy, A. Beskos, D. Crisan, A. Jasra, and N. Kantas (2022). A lagged particle filter for stable filtering of certain high-dimensional state-space models. *SIAM/ASA Journal on Uncertainty Quantification 10*(3), 1130–1161.

[100] Sarkka, S., A. Solin, and J. Hartikainen (2013). Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering. *IEEE Signal Processing Magazine 30*(4), 51–61.

[101] Schauer, M., F. van der Meulen, and H. van Zanten (2017). Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli 23*(4A), 2917 – 2950.

[102] Schiff, S. J. and T. Sauer (2007). Kalman filter control of a model of spatiotemporal cortical dynamics. *Journal of neural engineering 5*(1), 1.

[103] Sigrist, F., H. R. Künsch, and W. A. Stahel (2015). Stochastic partial differential equation based modelling of large space–time data sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology 77*(1), 3–33.

[104] Simão, I. (1991). A conditioned ornstein–uhlenbeck process on a hilbert space. *Stochastic Analysis and Applications 9*(1), 85–98.

[105] Sottinen, T. and S. Särkkä (2008). Application of Girsanov theorem to particle filtering of discretely observed continuous-time non-linear systems. *Bayesian Analysis 3*(3), 555 – 584.

[106] Spill, F., P. Guerrero, T. Alarcon, P. K. Maini, and H. Byrne (2015). Hybrid approaches for multiple-species stochastic reaction–diffusion models. *Journal of Computational Physics 299*, 429–445.

[107] Stanton, C. and A. Beskos (2024). Particle based inference for continuous-discrete state space models. *arXiv preprint arXiv:2407.15666*.

[108] Stroock, D. W. (1987). *Lectures on stochastic analysis: diffusion theory.* Cambridge University Press.

[109] Stroock, D. W. and S. S. Varadhan (1997). *Multidimensional diffusion processes*, Volume 233. Springer Science & Business Media.

[110] van der Meulen, F. and M. Schauer (2017). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electronic Journal of Statistics 11*(1), 2358 – 2396.

[111] Walsh, J. B. (1986). An introduction to stochastic partial differential equations. *Lecture notes in mathematics*, 265–439.

[112] Wan, E. A. and R. Van Der Merwe (2000). The unscented kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 adaptive systems for signal processing, communications, and control symposium (Cat. No. 00EX373)*, pp. 153–158. Ieee.

[113] Whitaker, G. A., A. Golightly, R. J. Boys, and C. Sherlock (2017). Improved bridge constructs for stochastic differential equations. *Statistics and Computing 27*, 885–900.

[114] Wikle, C. K., A. Zammit-Mangion, and N. Cressie (2019). *Spatio-temporal statistics with R.* Chapman and Hall/CRC.

[115] Yang, G., E. L. Baker, M. L. Severinsen, C. A. Hipsley, and S. Sommer (2024). Simulating infinite-dimensional nonlinear diffusion bridges.

[116] Young, D. A. (1984). A local activator-inhibitor model of vertebrate skin patterns. *Mathematical Biosciences 72*(1), 51–58.

[117] Zaitsev, V. F. and A. D. Polyanin (2002). *Handbook of exact solutions for ordinary differential equations.* Chapman and Hall/CRC.

Summary

Stochastic partial differential equations (SPDEs) are mathematical models that describe the evolution of dynamical systems in space and time under the influence of random noise. The noise may represent the system's inherent stochasticity, model uncertainties, or extrinsic stochastic forces acting on the system. By marrying the deterministic dynamics of classical partial differential equations with the influence of random noise, SPDEs allow for modeling of uncertainty in complex spatio-temporal phenomena in a wide range of fields, including geophysics, neuroscience, and finance.

In many real-world applications, a process modelled by an SPDE is observed only discretely and partially. As an example, consider an evolving firefront, whose fire intensity may be observed through satellite sensors at discrete points in time at a fixed spatial resolution. The SPDE model, jointly with the observed data, defines two statistical problems. Firstly, there is the estimation of the latent state of the system, given the observations. Typically, this is separated into two subproblems: the filtering problem, dealing with online state estimation as new observations become available, and the smoothing problem, dealing with offline state estimation given a full set of observations over a fixed time interval. Secondly, there is the question of parameter estimation. Typically, any SPDE model is governed by a set of parameters. These might, for example, represent physical constants that determine the dynamics of the system. In applications where such parameters are unknown, they are to be estimated jointly with the unobserved signal, based on the observed data.

This thesis develops Bayesian computational approaches to statistical inference for SPDEs. Both state estimation, in the online and offline setting, as well as parameter estimation for discretely and partially observed semilinear SPDEs are addressed. It therefore fills a gap in the current literature on statistics for SPDEs, which has so far focused primarily on frequentist parameter estimation or state estimation of linear SPDEs. By introducing methodology that enables inference of the latent state, model calibration and uncertainty quantification for semilinear SPDEs, based on incomplete and noisy data, this work broadens the applicability of SPDEs in real-world settings.

In Chapter 2, a class of exponential measure transformations for SPDEs is introduced. Conditions are derived under which the transformed measure is of Girsanov-type such that the mild solution to the SPDE evolves according to yet another SPDE with an additional drift term. An application this result gives rise to the infinite-dimensional diffusion bridge - a mild solution to an SPDE conditioned on hitting a predefined terminal state. This generalises results previously known only for linear systems. Moreover, the guided process, the mild solution to a tractable SPDE that steers a process towards a terminal state, is introduced as an approximation to the diffusion bridge.

Chapter 3 develops sampling methodology for the intractable infinite-dimensional diffusion bridge. This serves as a fundamental building block for computational Bayesian approaches to inference for discretely observed SPDEs. In the main result of Chapter 3, conditions are derived under which absolute continuity holds between the laws of the guided process and the diffusion bridge. This legitimises the guided process as a proposal distribution for importance sampling or Metropolis-Hastings schemes that target the law of the infinite-dimensional diffusion bridge.

In Chapter 4, these ideas are extended to build methodology for all of the aforementioned statistical tasks. To address the filtering problem, a sequential Monte Carlo scheme is introduced, built upon the law of the guided process between observation times as a proposal distribution. The smoothing and parameter inference problems are solved by generalising the measure transformations of Chapter 2 to include conditioning and guiding based on multiple observations. Building on these transformations and a reparametrisation of the conditioned process, a Gibbs sampler is derived that samples from the joint posterior of the smoothed process and unknown model parameters.

Samenvatting

Stochastische partiële differentiaalvergelijkingen (SPDV's) zijn wiskundige modellen die de evolutie van dynamische systemen in ruimte en tijd onder invloed van stochastische ruis beschrijven. De ruis kan de inherente stochasticiteit van het systeem, modelonzekerheden of extrinsieke stochastische krachten die op het systeem inwerken, vertegenwoordigen. Door de deterministische dynamica van klassieke partiële differentiaalvergelijkingen te combineren met de invloed van stochastische ruis, maken SPDV's het mogelijk om onzekerheid in complexe spatiotemporele fenomenen te modelleren in een breed scala aan vakgebieden, waaronder geofysica, neurowetenschappen en financiën.

In veel praktijktoepassingen wordt een proces dat door een SPDV wordt gemodelleerd slechts discreet en gedeeltelijk waargenomen. Neem als voorbeeld een zich ontwikkelend vuurfront, waarvan de vuurintensiteit kan worden waargenomen door satellietsensoren op discrete tijdstippen met een vaste ruimtelijke resolutie. Het SPDV-model definieert, samen met de geobserveerde data, twee statistische problemen. Ten eerste is er de schatting van de latente toestand van het systeem, gegeven de waarnemingen. Meestal wordt dit opgesplitst in twee subproblemen: het filterprobleem, dat betrekking heeft op online toestandsschatting naarmate nieuwe waarnemingen beschikbaar komen, en het smoothing-probleem (afvlakking), dat betrekking heeft op offline toestandsschatting op basis van een volledige set waarnemingen over een vast tijdsinterval. Ten tweede is er het vraagstuk van parameterschatting. Kenmerkend is dat elk SPDV-model wordt bepaald door een set parameters. Deze kunnen bijvoorbeeld fysische constanten vertegenwoordigen die de dynamica van het systeem bepalen. In toepassingen waar dergelijke parameters onbekend zijn, moeten deze gezamenlijk met het niet-waargenomen signaal worden geschat op basis van de geobserveerde data.

Dit proefschrift ontwikkelt computationele Bayesiaanse benaderingen voor statistische inferentie voor SPDV's. Zowel toestandsschatting, in de online en offline setting, als parameterschatting voor discreet en gedeeltelijk waargenomen semilineaire SPDV's worden behandeld. Het vult daarmee een leemte in de huidige literatuur over statistiek voor SPDV's, die zich tot nu toe voornamelijk heeft gericht op frequentistische parameterschatting of toestandsschatting van lineaire SPDV's. Door methodologie te introduceren die inferentie van de latente toestand, modelkalibratie en onzekerheidskwantificatie voor semilineaire SPDV's mogelijk maakt, gebaseerd op onvolledige en ruisbehepte data, verbreedt dit werk de toepasbaarheid van SPDV's in de praktijk.

In Hoofdstuk 2 wordt een klasse van exponentiële maattransformaties voor SPDV's geïntroduceerd. Er worden voorwaarden afgeleid waaronder de getransformeerde maat van het Girsanov-type is, zodanig dat de milde oplossing van de SPDV evolueert volgens weer een andere SPDV met een extra driftterm. Een toepassing van dit resultaat leidt tot de oneindig-dimensionale diffusiebrug - een milde oplossing van een SPDV die is

geconditioneerd op het bereiken van een vooraf gedefinieerde eindtoestand. Dit generaliseert resultaten die voorheen alleen bekend waren voor lineaire systemen. Bovendien wordt het gestuurde proces, de milde oplossing van een hanteerbare SPDV die een proces naar een eindtoestand stuurt, geïntroduceerd als een benadering van de diffusiebrug.

Hoofdstuk 3 ontwikkelt samplingmethodologie voor de rekenkundig onhanteerbare oneindig-dimensionale diffusiebrug. Dit dient als een fundamentele bouwsteen voor computationele Bayesiaanse benaderingen van inferentie voor discreet geobserveerde SPDV's. In het belangrijkste resultaat van Hoofdstuk 3 worden voorwaarden afgeleid waaronder absolute continuïteit geldt tussen de verdelingen van het gestuurde proces en de diffusiebrug. Dit legitimeert het gestuurde proces als een voorstelverdeling voor importance sampling of Metropolis-Hastings-schema's die gericht zijn op de verdeling van de oneindig-dimensionale diffusiebrug.

In Hoofdstuk 4 worden deze ideeën uitgebreid om methodologie op te bouwen voor alle bovengenoemde statistische taken. Om het filterprobleem aan te pakken, wordt een sequentieel Monte Carlo-schema geïntroduceerd, dat is gebouwd op de verdeling van het gestuurde proces tussen observatietijdstippen als voorstelverdeling. De smoothing- en parameterinferentieproblemen worden opgelost door de maattransformaties van Hoofdstuk 2 te generaliseren, zodat deze conditionering en sturing op basis van meerdere waarnemingen omvatten. Voortbouwend op deze transformaties en een reparametrisatie van het geconditioneerde proces, wordt een Gibbs-sampler afgeleid die een steekproef trekt uit de gezamenlijke a posteriori-verdeling van het afgevlakte proces en de onbekende modelparameters.

Acknowledgments

I would first like to extend my gratitude to Frank and Aad for their guidance throughout the years. I am grateful for all your advice, patience, positive reinforcement and your level-headed approach to supervision. Thank you, Frank, for initiating this research with me, all the way back in 2019 when we were merely discussing ideas for a Master's thesis. And thank you, Aad, for your support without which this project would not have existed in this shape.

Many thanks to Geurt Jongbloed, Gabriel Lord, Markus Reiß, Manuel Gnann, Daniel Paulin and Mark Veraar for their service as committee members. I appreciate the time you took to read this thesis and your valuable feedback.

I would also like to thank everyone from the Statistics group of the Delft Institute of Applied Mathematics. In particular, my thanks go to my fellow PhD students with whom I have shared parts of the academic journey: Marc, Lasse, Sebastiano, Koen, Wieger, Thomas, Chris, Andreas, Hidde, Serena, Dan, Vicente, Jonathan, Francesco and Matthias. The PhD journey can feel long and challenging at times, and being able to share it with great peers plays a fundamental role in finishing it with success. Most importantly, I would like to express my gratitude to Fra and Ardjen. I deeply appreciate the friendship that has grown out of our time spent together in Delft.

The end of my PhD also marks the end of my stay in the Netherlands, at least for the foreseeable future. With this in mind, I would like to thank all the wonderful people who have contributed to turning the Netherlands into my home away from home over the past few years.
Thank you, Betty, for being there for me from the first day that I moved to Rotterdam. Your friendship has been a great pillar throughout.
To Manolis and Nadia: you already know how much I appreciated and relied on your warmth and welcoming nature. Let me thank you again for it here. Thank you also, Andrei, for all the shared laughs and banter and for being a good housemate and an even better friend. I would also like to express my deep gratitude to Shiyu for her support and the time spent together.
I am also grateful to Alba and Bennet for the parties, Saturday morning market hangouts and DnD sessions. And to Anna, for all the fun times we had together, and Rohit and Bartek, whom I wish I had met sooner!
Thank you to everyone whom I have climbed with over the years at Beest for the shared sessions. Shoutout to the Monday night training group - Franz, Samuel, Sil, Nino, Andrés and all the others. Climbing together was much appreciated to clear the head after long days of thesis writing. In particular, I would like to thank Iñaki, whose passion not only for the sport but also for its community is infectious. Thanks also to Domen, whose advice I deeply valued and relied on. Lastly, my gratitude goes out to the baristas at Harvest Coffee Brewers, without whom this thesis might still be in production.

There are too many to thank without turning this section into its own thesis, so to everyone else with whom I have crossed paths over the past years - thank you!

To my friends abroad, especially Vale, Jared and Chloe, I want to express my appreciation for the on-going efforts to visit each other and the great memories made during those visits. Spring break never ends.
I would like to extend my deepest gratitude to Marisa, in whose company the larger part of this thesis was written. Thank you for your support and the patience you have shown while listening to my complaints whenever I got stuck.
Special thanks also go to Gerrit and Lukas for visiting so frequently and bringing a piece of home whenever doing so.
But most importantly, I thank my parents for everything. This thesis would not have been possible without your continuous support and unconditional love.

Written by a human.

Curriculum Vitae

Thorben Pieper-Sethmacher
Born: 3rd of August, 1995, Peine, Germany

Education

Delft University of Technology 2021 – 2026
PhD Mathematics & Statistics
Promotors: Prof. Frank van der Meulen, Prof. Aad van der Vaart

Delft University of Technology 2018 – 2021
MSc Applied Mathematics, *cum laude*
Supervisors: Prof. Frank van der Meulen, Prof. Sally Cripps

Westfälische Wilhelms-Universität Münster 2015 – 2018
BSc Mathematics, *cum laude*
Supervisor: Prof. Siegfried Echterhoff

Experience

Postdoctoral Researcher Nov. 2025 – ...
Nanyang Technological University
(Supervisor: A/Prof. Daniel Paulin)

PhD Candidate Sep. 2021 – Sep. 2025
Delft University of Technology

Guest Researcher Mar. 2020 – Aug. 2021
Centre for Translational Data Science, University of Sydney

Machine Learning Intern Dec. 2019 – Feb. 2020
Central Bank of Germany

Teaching Assistant Oct. 2016 – Jul. 2018
Westfälische Wilhelms-Universität Münster

Publications

Published

Pieper-Sethmacher, T., F. van der Meulen and A. van der Vaart (2025). On a class of exponential changes of measure for stochastic PDEs. *Stochastic Processes and their Applications*, 185, 104630.

Accepted for publication

Pieper-Sethmacher, T., F. van der Meulen and A. van der Vaart (2025). Simulation of infinite-dimensional diffusion bridges. To appear in *Annals of Applied Probability.* arXiv:2503.13177.

Submitted

Pieper-Sethmacher, T., D. Avitabile and F. van der Meulen (2025). Guided filtering and smoothing for infinite-dimensional diffusions. *arXiv preprint arXiv:2507.06786.* (Under review).