Asymptotic theory for Bayesian nonparametric inference in statistical models arising from partial differential equations

Matteo Giordano

Supervisor: Prof. Richard Nickl

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Clare Hall

July 2021

Declaration

I hereby declare that my dissertation entitled *Asymptotic theory for Bayesian nonparametric inference in statistical models arising from partial differential equations* is not substantially the same as any that I have submitted, or is being concurrently submitted, for a degree or diploma or other qualification at the University of Cambridge or any other university or similar institution. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other university or similar institution. It does not exceed the prescribed word limit for the relevant Degree Committee.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared below and specified in the text. Chapter 2 consists of original research, which was conducted in collaboration with Hanne Kekkonen and published as [105]. Chapter 3 consists of original research, which was conducted in collaboration with Richard Nickl and published as [106]. Chapter 4 consists of original research, which was conducted in collaboration with Kolyan Ray and submitted to the arXiv [107]. Chapter 1 contains introductory and background material, from sources cited throughout, and the results of numerical simulations that I have personally implemented.

Matteo Giordano
July 2021

Abstract

Partial differential equations (PDEs) are primary mathematical tools to model the behaviour of complex real-world systems. PDEs generally include a collection of parameters in their formulation, which are often unknown in applications and need to be estimated from the data. In the present thesis, we investigate the theoretical performance of nonparametric Bayesian procedures in such parameter identification problems in PDEs. In particular, inverse regression models for elliptic equations and stochastic diffusion models are considered.

In Chapter 2, we study the statistical inverse problem of recovering an unknown function from a linear indirect measurement corrupted by additive Gaussian white noise. We employ a nonparametric Bayesian approach with standard Gaussian priors, for which the posterior-based reconstruction corresponds to a Tikhonov regulariser with a reproducing kernel Hilbert space norm penalty. We prove a semiparametric Bernstein–von Mises theorem for a large collection of linear functionals of the unknown, implying that semiparametric posterior estimation and uncertainty quantification are valid and optimal from a frequentist point of view. The general result is applied to three concrete examples that cover both the mildly and severely ill-posed cases: specifically, elliptic inverse problems, an elliptic boundary value problem, and the recovery of the initial condition of the heat equation. For the elliptic boundary value problem, we also obtain a nonparametric version of the theorem that entails the convergence of the posterior distribution to a prior-independent infinite-dimensional Gaussian probability measure with minimal covariance. As a consequence, it follows that the Tikhonov regulariser is an efficient estimator, and we derive frequentist guarantees for certain credible balls centred around it.

Chapter 3 is concerned with statistical nonlinear inverse problems. We focus on the prototypical example of recovering the unknown conductivity function in an elliptic PDE in divergence form from discrete noisy point evaluations of the PDE solution. We study the statistical performance of Bayesian nonparametric procedures based on a flexible class of Gaussian (or hierarchical Gaussian) process priors, whose implementation is feasible by MCMC methods. We show that, as the number of measurements increases, the resulting

posterior distributions concentrate around the true parameter generating the data, and derive a convergence rate, algebraic in inverse sample size, for the estimation error of the associated posterior means.

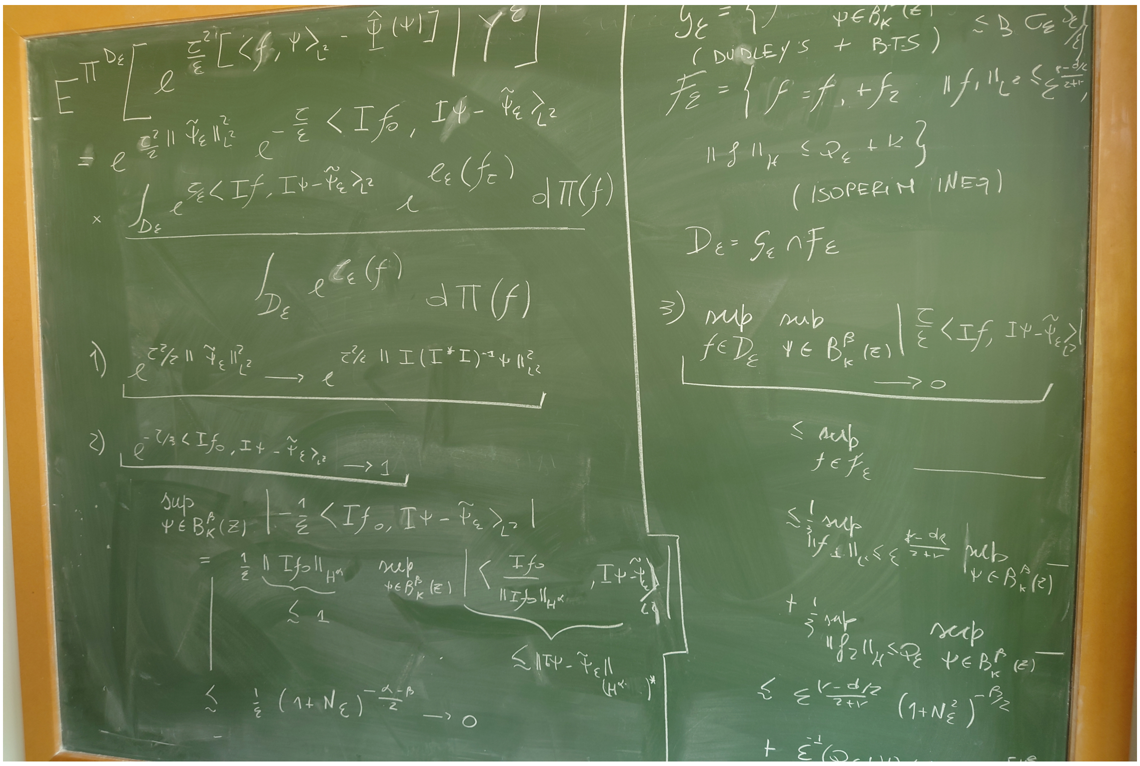
Finally, in Chapter 4 we extend the posterior consistency analysis to dynamical models based on stochastic differential equations. We study nonparametric Bayesian models for reversible multi-dimensional diffusions with periodic drift. For continuous observation paths, reversibility is exploited to prove a general posterior contraction rate theorem for the drift gradient vector field under approximation-theoretic conditions on the induced prior for the invariant measure. The general theorem is applied to Gaussian priors and p-exponential priors, which are shown to converge to the truth at the minimax optimal rate over Sobolev smoothness classes in any dimension.

Chapter 1 is dedicated to introducing the statistical models considered in Chapters 2 - 4, and to providing an overview of the theoretical results derived therein. The main theorems of Chapter 2 and Chapter 3 are illustrated via the results of simulations, and detailed comments are provided on the implementation.

To my parents, for their love and support.
And to my grandfather Ugo: I wish you were still here to share in this.

Acknowledgements

First and foremost, I wish to express my deepest and most sincere gratitude to my supervisor, Richard Nickl. Thanks, Richard, for believing in me and giving me the opportunity to join your amazing research group; thanks for your guidance and support throughout my Ph.D.; and thanks for helping me to develop as a mathematician and as an independent researcher. I have been inspired by your mathematical insight and intuition, and found our collaboration a profoundly enriching and rewarding experience. During one of our very first meetings, I took some pictures to keep track of our calculations on the blackboard, and you (jokingly) suggested that I should keep them and put them somewhere in my thesis. It seemed to me a very long way back then; but almost four years have passed, and here is one of those pictures:



I then wish to extend my thanks to my coauthors, Hanne and Kolyan, and to the whole research group: to Alberto, Chris, Gabriel, Jan, Kweku, Matthias, Neil, Randolf, Simon, Sven and Thomas. It has been a privilege to be part of this community, and I am grateful that I got to know, learn from and exchange ideas with all of you. That we managed to keep on regularly meeting (albeit virtually) during the months of lockdown has been of great help. Thanks to Daniel and Hamed for being such excellent (and humorous!) office mates. Thanks to the whole 2017/2018 CCA cohort for sharing this journey; and thanks to Tessa for all her support.

Thanks to my *Alma mater*, the University of Torino, where I studied as an undergraduate and for my master's degree. Thanks to the Collegio Carlo Alberto community in Torino for pushing me towards pursuing doctoral studies. And thanks to my former professors Antonio, Igor, Matteo and Pierpaolo for their advice.

I am very grateful to my College, Clare Hall, for providing such a welcoming home at my arrival in Cambridge. Throughout these four years, I have greatly enjoyed and shared in the egalitarian, friendly and intimate atmosphere of the Clare Hall community. And, of course, thanks to everyone involved, in the past and present, with the Clare Hall Boat Club: you have made those countless winter mornings on the river Cam worth every moment, and I am incredibly proud to having contributed to your legacy. Finally, thanks to the City of Cambridge for being such a special place to live in, and for making myself and my partner feel so warmly welcome.

Throughout this journey I have had the good fortune of meeting so many great people and new friends. Thanks to Alex, Collin, Jou, Leonard, Lukas, Marno, Mans, Matthias, Nino, Patrick, Sam, Tobias, and Timo for your dedication to the Clare Hall men's rowing crew; thanks to Andy, Giovanna, Josh, Neil, Rhea, Vishal and the other housemates at Michael Stoker Building; thanks to Emily, Mariana, Simone and the whole Clare Hall GSB for organising some spectacular events; thanks to Arran and Dr. Forster for some brilliant conversation at the mathematical lunches; thanks to Francisco and the entire Cambridge '99 rowing community. Thanks to everyone I am not mentioning here explicitly for enriching these fours years in a unique way. And, above all, thanks to my cousin Simone and to my lifetime friends Lorenzo and Luca: I look forward to meeting you after such a long time apart.

I wish to conclude these acknowledgements expressing my most affectionate gratitude to my parents, to whom this thesis is dedicated, to my sisters Camilla and Margherita, to Bilbo and to my wider family for their love and unceasing support towards my scholastic pursuits. My thought goes to my grandmother Anna after the recent passing of my grandfather Ugo: to his memory also I wish to dedicate this thesis. And, finally, my most heartfelt and loving gratitude is for Silvia, for all the happiness she has brought to me ever since we met, and for sharing with me this English adventure, of which I cannot wait to write with her the next chapter.

Matteo Giordano
July 2021

Formal acknowledgements

I gratefully acknowledge the financial support, that has made this thesis possible, of Richard Nickl's ERC grant No. 647812 and of the EPSRC grant EP/L016516/1 for the Cambridge Centre for Analysis. I also would like to thank two anonymous referees for their comments on the article [105] on which Chapter 2 is based, and three anonymous referees for comments on the article [106] forming the basis of Chapter 3.

Table of contents

List of figures

Così tra questa
immensità s'annega il pensier mio:
e il naufragar m'è dolce in questo mare.

Giacomo Leopardi, L'infinito (13-15)

Chapter 1

Introduction

Partial differential equations (PDEs) are primary mathematical tools to model the behaviour of complex real-world systems, such as - to name but a few major examples - transport of mass, diffusing heat and travelling waves, with ubiquitous applications in physics, engineering and throughout the applied sciences. PDEs generally include a collection of parameters in their formulation (for example, a conductivity functional coefficient modelling spatially-varying diffusion of heat throughout an inhomogeneous medium), which are often unknown in applications. When data are collected from a system modelled by a PDE, it is then of central interest - for the purpose of scientific investigation or to make predictions about its future behaviour - to recover the unknown parameters in the governing equation [88, 220, 122].

The design of algorithms to solve such *parameter identification problems* in PDEs must formally incorporate the underlying PDE structure into the analysis, and also coherently account for (potential) sources of noise within the data generating process. A realistic description of such measurement errors often requires a probabilisitic model; furthermore, in many applications the unknown parameter of interest cannot be modelled by a finite set of numbers (i.e., a finite-dimensional vector), but rather has to be treated as a genuine function, giving rise to PDE-based *infinite-dimensional* (also commonly termed *nonparametric*) statistical inference problems. In the present thesis, two major, and intimately related, classes of nonparametric statistical models arising from PDEs of elliptic and parabolic type are considered:

1. *inverse regression models*, wherein the parameter of interest is observed indirectly through noisy point evaluations of the solution of a PDE (e.g., [88, 127, 223, 220, 122]);

2. *stochastic diffusion models*, in which the data-generating mechanism is described by a stochastic differential equation (SDE) (e.g., [188, 145, 96]).

The (nonparametric) Bayesian paradigm of statistical inference provides a principled and methodologically attractive approach to such statistical models, and has gained great popularity in the last decade since influential work by Stuart [220, 73]. In the Bayesian framework, the underlying PDE structure and the statistical properties of the noise are naturally incorporated into the analysis via the *likelihood function*, which describes the data sampling distribution for each given element in the parameter space. Following the Bayesian paradigm, the unknown parameter is regarded as a random variable (with values in a function space or a subset thereof) and assigned a *prior probability distribution*. The likelihood and the prior then jointly induce, through *Bayes' formula*, a *posterior distribution*, which represents the conditional distribution of the parameter given the observations, and is used to draw inferential conclusions about the unknown parameter of interest.

The posterior distribution can be formally expressed in terms of evaluations only of the likelihood function and the prior distribution, thereby conceptually sidestepping the need of an inversion formula for the complex PDE-based relationship between the parameter and the observations. At the same time, the *posterior mean* provides a broadly applicable approach to point estimation that constitutes an attractive alternative to traditional penalised least squares procedures, which may give rise, in nonlinear likelihood structures, to non-convex optimisation problems. A further fundamental feature of the Bayesian methodology is that, alongside point estimates, it also automatically delivers a *quantification of the uncertainty* in the reconstruction, measured by the spread of the posterior distribution around its centre, that may be used in applications to construct *credible sets* (i.e., regions of the parameter space of high posterior probability) and hypothesis tests.

In practice, in the statistical models considered in the present thesis, outside of certain *conjugate* settings where explicit formulae for the posterior distribution can be derived, nonparametric Bayesian procedures are implemented using suitable numerical methods. In particular, recent advances in high- and infinite-dimensional Markov chain Monte Carlo (MCMC) algorithms (e.g., [201, 63, 66, 26, 210]) allow to reliably and efficiently obtain approximate posterior samples that can be used to construct posterior mean estimates (via MCMC averages) and credible sets (through the empirical quantiles of the MCMC samples). The Bayesian approach thus provides concrete algorithms for estimation *and* uncertainty quantification that can be successfully employed in a great variety of complex real-world applications.

It is therefore of great importance to provide theoretical performance guarantees for nonparametric Bayesian procedures in statistical models arising from PDEs, in order to certify the validity of the inferential conclusions that such methodology delivers, and also to allow an objective comparison with potential competing statistical procedures. In particular, the performance of Bayesian methods depends on a suitable choice of the prior, which in the nonparametric framework primarily serves as a regularisation tool for the likelihood (rather than representing subjective beliefs about the unknown), and whose specification is a complex task in its own merit (cfr. Section 1.2 in [101]). Thus, a natural question arises as to whether Bayesian procedures may provide valid and prior-independent inferential conclusions - at least in the presence of informative data.

The established paradigm under which such investigation can be carried out consists in performing a *frequentist analysis* of Bayesian methods (see [101], and also Section 7.3 in [104]), assuming that the observations are generated by a fixed *ground truth* in the parameter space (as opposed to the parameter being randomly drawn from the prior distribution), and studying the asymptotic behaviour of the posterior distribution in the large sample size (or equivalently, small noise variance) limit. More specifically, two fundamental questions arise in this framework in regards to the theoretical performance of nonparametric Bayesian procedures:

1. *posterior consistency*, i.e., the concentration of the posterior distribution around the assumed true parameter. Related questions are the convergence of relevant Bayesian estimators, such as the posterior mean, as well as a quantitative characterisation of the rate at which the contraction occurs;

2. validity of Bayesian uncertainty quantification, that is, whether credible sets are asymptotically proper *confidence set*, achieving in the limit frequentist coverage probability at least equal to their credible level.

While related, the two above questions entail distinct analyses. Indeed, while posterior consistency pertain to the global behaviour of the posterior distribution, understanding the frequentist coverage of Bayesian credible sets can be approached by studying the local microscopic fluctuations of the posterior distribution over set where it concentrates, which, as we shall elaborate in Chapter 2, may depend subtly on the geometries involved. Both of these issues have been intensely investigated in the literature in the past two decades in classical nonparametric statistical models such as regression and density estimation [104, 101]. More recently, significant advances have been achieved in extending such results to PDE-based statistical models, including linear [138, 6, 195, 133, 166] and nonlinear inverse problems [241, 2, 175, 171, 167, 168], and stochastic diffusion models [232, 110, 174, 173].

The purpose of the present thesis is to further advance our understanding in this emerging area. In particular, elliptic equations and stochastic diffusions originating from advection-diffusion models are considered. Chapter 2 investigates the validity of Bayesian uncertainty quantification in a class of linear inverse problems. For Bayesian procedures based on standard Gaussian priors, a (semiparametric) Bernstein–von Mises theorem is proved, providing a detailed description of the limiting shape of the posterior distributions of certain functionals of the parameter, and implying that credible intervals are asymptotically valid confidence intervals with radius shrinking at the optimal rate. For a concrete example based on an elliptic boundary value problem, an infinite-dimensional extension of the result is further derived. Chapter 3 is concerned with nonlinear inverse problems. Focusing on the prototypical example of recovering the conductivity function in an elliptic PDE in divergence form, guarantees for posterior consistency and convergence of the posterior mean (with upper bounds, algebraic in inverse sample size, on the rates) are provided for Bayesian nonparametric procedures based on Gaussian (or hierarchically Gaussian) priors. The results presented stem from a general posterior contraction rates theory developed for a class of nonlinear inverse problems. Finally, Chapter 4 extends the posterior consistency analysis to dynamical models based on SDEs. It is shown that the posterior distributions on the drift vector field arising from Gaussian and p-exponential priors contracts towards the ground truth at (minimax) optimal rate in any dimension. These chapters are aimed to be self-contained and can also be read independently.

The remainder of this introductory chapter is mostly dedicated to presenting the statistical models considered in Chapters 2 - 4, and to discussing the theoretical results derived therein, with an overview on the main mathematical challenges and the proofs strategies. The main theorems of Chapter 2 and Chapter 3 are illustrated via the results of simulations, and detailed comments are provided on the implementation. Section 1.3 contains additional discussion, with references to relevant related literature and core background material.

1.1 Advection-diffusion equations and related statistical models

Advection-diffusion equations are a major class of second-order PDEs that describe a large variety of phenomena in physics and applied mathematics, involving the transfer of a physical quantity (such as particles, energy, or individuals in a population) inside a system due to the superposition of two processes:

1. *advection*, i.e., the transport by the bulk motion of a fluid containing or holding the transferred physical quantity;

2. *diffusion*, that is, the transport from regions of higher concentration to regions of lower concentration.

See, e.g., [12, Section 3.5]. Formally, let $\mathcal{O} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be an open and bounded domain with boundary $\partial\mathcal{O}$, and let $u(t,x)$ denote the density at time $t \geq 0$ and location $x \in \mathcal{O}$ of a physical quantity being transferred within $\mathcal{O}$. Then, the advection-diffusion model postulates that the evolution of u is governed by a linear parabolic PDE of the form

$$\frac{\partial u}{\partial t} = \nabla \cdot (A\nabla u) - b.\nabla u + cu + s, \qquad t \geq 0, \quad x \in \mathcal{O}, \tag{1.1}$$

where $\nabla \cdot = \sum_{i=1}^d \partial_{x_i}$ and ∇ denote the divergence and gradient operators respectively. Above, $A = A(x) = [A_{ij}(x)]_{i,j=1}^d$ is a (matrix-valued) *diffusion* (or *conductivity*) *coefficient* modelling anisotropic and spatially-varying diffusion throughout the domain $\mathcal{O}$, assumed to satisfy $A_{ij} = A_{ji},\ i \neq j$, and the uniform ellipticity condition

$$\sum_{i,j=1}^d A_{ij}(x)\xi_i\xi_j \geq a|\xi|^2 \tag{1.2}$$

for some constant $a > 0$, all $x \in \mathcal{O}$ and $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$. The *drift* vector field $b = b(x) = [b_1(x), \dots, b_d(x)]$ represents the velocity of the moving fluid, the scalar field $c = c(x)$ is an *attenuation coefficient* modelling local absorption (or release) of the physical quantity, and the term $s = s(x)$ describes the spatial distribution of local *sources* $(s(x) > 0)$ or *sinks* $(s(x) < 0)$. For the purpose of this thesis, the coefficients A, b, c and s are assumed to be independent of time. Equation (1.1) is typically complemented with some *initial condition*

$$u(0,x) = h(x), \qquad x \in \mathcal{O},$$

and some *boundary conditions*, which we take to be of Dirichlet type, prescribing for some function g defined on the boundary $\partial\mathcal{O}$,

$$u(t,x) = g(x), \qquad t \geq 0, \quad x \in \partial\mathcal{O}.$$

The advection-diffusion equation (1.1) can be derived from a *continuity equation* (e.g., [89, Chapter 11]) for the physical quantity being transferred: in particular, the processes of advection and diffusion jointly induce a *flux*

$$F\left(t, x, u(t, x)\right) = b(x)u(t, x) - A(x)\nabla u(t, x), \tag{1.3}$$

where the first term represents the advective flux associated to the fluid moving with velocity b, while the second term is the diffusive flux, proportional, according to *Fick's first law*, to the concentration gradient. Note that, in view of the ellipticity condition (1.2),

$$-A(x)\nabla u(t, x).\nabla u(t, x) \leq 0,$$

i.e., the diffusive flux is directed towards regions of lower concentration. Then, the continuity equation states that for any open subregion $\mathcal{O}' \subset \mathcal{O}$ with smooth boundary $\partial\mathcal{O}'$, the rate of change in the total amount of the physical quantity contained in $\mathcal{O}'$ equals the sum of the negative net flux through $\partial\mathcal{O}'$ and the net rate at which the quantity is created or depleted within $\mathcal{O}'$ (due to attenuation and the presence of sources or sinks), in integral form:

$$\frac{d}{dt}\int_{\mathcal{O}'} u(t, x)dx = -\int_{\partial\mathcal{O}'} F.\nu dS + \int_{\mathcal{O}'} \left[c(x)u(t, x) + s(x)\right] dx, \tag{1.4}$$

where ν is the unit outer normal field to the surface $\partial\mathcal{O}'$. By the divergence theorem, $\int_{\partial\mathcal{O}'} F.\nu dS = \int_{\mathcal{O}'} \nabla \cdot F(t, x, u(t, x))dx$, and therefore Equation (1.4) can be rewritten as

$$\int_{\mathcal{O}'} \frac{\partial}{\partial t}u(t, x)dx = \int_{\mathcal{O}'} \left[-\nabla \cdot F(t, x, u(t, x)) + c(x)u(t, x) + s(x)\right] dx.$$

By the arbitrariness of the subdomain $\mathcal{O}'$, it can be concluded that

$$\frac{\partial}{\partial t}u = -\nabla \cdot F + cu + s,$$

and recalling the specific form of the flux given in (1.3),

$$\begin{aligned}\frac{\partial}{\partial t}u &= \nabla \cdot (A\nabla u) - \nabla \cdot (bu) + cu + s \\ &= \nabla \cdot (A\nabla u) - b.\nabla u - u\nabla \cdot b + cu + s,\end{aligned}$$

whence Equation (1.1) follows as a particular case for *incompressible flows*, characterised by velocity with zero divergence, $\nabla \cdot b = 0$.

Assuming that observations from a system governed by an advection-diffusion equation of the form (1.1) are available, several parameter identification problems can be formulated depending on the application and the specific measurement model at hand, including the recovery of the coefficients A, b, c and s, and of the initial condition h. For example, among the numerous application areas are tomography and medical imaging (e.g., [25, 11]), fluid dynamics (e.g., [64]), finance (e.g., [191, 130]), photoacustic and scattering problems (e.g., [18, 17]) and reservoir hydrology (e.g., [247]); see [16] for an overview.

In the next subsections, we introduce the two major classes of statistical models originating from advection-diffusion equations that will be investigated in Chapters 2 - 4.

1.1.1 Inverse regression models for elliptic PDEs

Measurement models based on elliptic PDEs naturally arise from the advection-diffusion setting for systems in equilibrium, for which the density u in Equation (1.1) satisfies the steady-state condition $\partial_t u = 0$. Then, denoting by $u(x) = u(t, x)$, Equation (1.1) simplifies to the elliptic PDE with Dirichlet boundary conditions

$$\begin{cases} \nabla \cdot (A\nabla u) - b.\nabla u + cu + s = 0, & x \in \mathcal{O}, \\ u(x) = g(x), & x \in \partial\mathcal{O}. \end{cases} \tag{1.5}$$

Recall that above $A = [A_{ij}]_{i,j=1}^d$, with $A_{ij} = A_{ji}$, $i \neq j$, is a diffusion coefficient satisfying the uniform ellipticity condition (1.2), $b = [b_1, \dots, b_d]$ is the velocity vector field, c is the attenuation coefficient and s is the source term.

Under appropriate regularity conditions on the coefficients A, b, c, s and g, and on the boundary $\partial\mathcal{O}$, a unique classical twice continuously-differentiable solution $u \in C^2(\mathcal{O})$ to (1.5) exists (e.g., [89, Chapter 6]). Assuming that measurements throughout the domain $\mathcal{O}$ are available, a prototypical statistical model for applications based on a system governed by equation (1.5) then postulates data given by discrete noisy observations of u over a grid of design points $\{X_1, \dots, X_N\} \subset \mathcal{O}$,

$$Y_i = u(X_i) + \sigma W_i, \qquad i = 1, \dots, N, \tag{1.6}$$

corrupted by statistical measurement errors $W_1, \dots, W_N$ scaled by the noise level $\sigma > 0$. Since observational noise typically arise as a superposition of many small independent random effects, a Gaussian model $W_1, \dots, W_N \overset{\text{iid}}{\sim} N(0, 1)$ is often realistic in view of the central limit theorem. The design points can either be derministic (in which case we speak of *fixed design*), or themselves be random variables drawn from some probability distributions on the domain $\mathcal{O}$ (*random design* case). The task at hand is then to infer

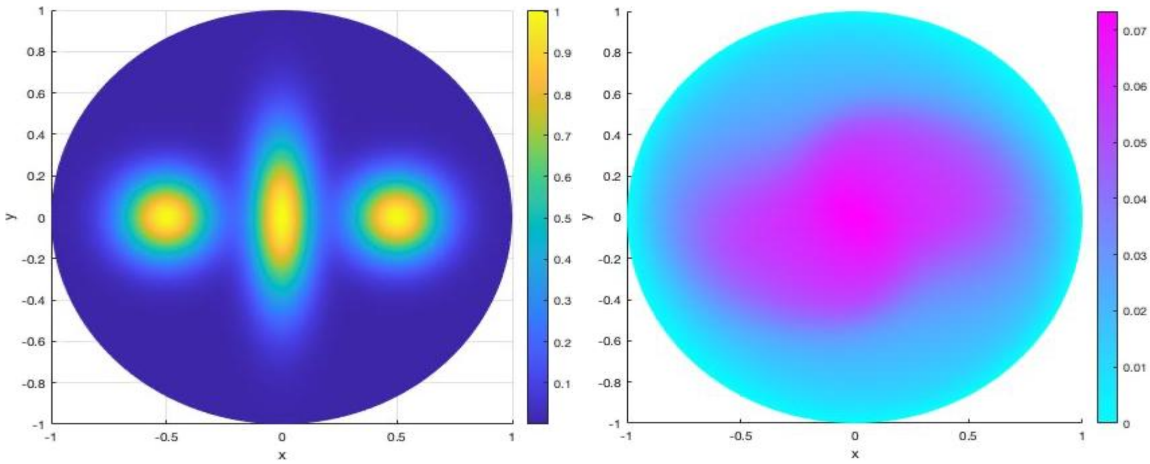


Fig. 1.1 Left: an example of source function s. Right: the corresponding PDE solution u_s (for a given diffusion coefficient A).

one of the coefficients appearing in equation (1.5) from data $(Y_i, X_i)_{i=1}^N$, giving rise, in view of the assumed randomness of the noise, to a *statistical inverse (regression) problem*.

Concrete instances of statistical inverse problems of the form (1.6) (or a continuous version thereof) are considered in Chapters 2 and 3. In particular, Chapter 2 investigates *linear* inverse problems, wherein the relationship between the unknown parameter of interest and the observed object (e.g., the PDE solution u in (1.6)) is given by a *linear forward map*. The main example considered entails the recovery of the unknown source function s in the steady-state diffusion equation with zero Dirichlet boundary conditions

$$\begin{cases} \nabla \cdot (A\nabla u) = s, & x \in \mathcal{O}, \\ u(x) = 0, & x \in \partial\mathcal{O}, \end{cases} \tag{1.7}$$

from noisy observations of its solution $u = u_s$, for smooth $\partial\mathcal{O}$ and known diffusion coefficient A. By standard elliptic theory (e.g., [89, Chapter 6]), the associated forward map

$$G : L^2(\mathcal{O}) \to L^2(\mathcal{O}), \qquad s \mapsto u_s,$$

where $L^2(\mathcal{O})$ denotes the usual space of measurable and square integrable functions on $\mathcal{O}$, defines an injective compact linear operator with unbounded inverse, giving rise to a linear *ill-posed* inverse problem. Figure 1.1 provides an illustration of the problem on the unit disk for a specific choice of the source s (and fixed diffusion coefficient A); see Section 1.2.1 for details on the implementation and numerical results.

In Chapter 3 the scope of the investigation is broadened to *nonlinear* inverse problems. In the representative example considered here, the goal is to recover the unknown positive

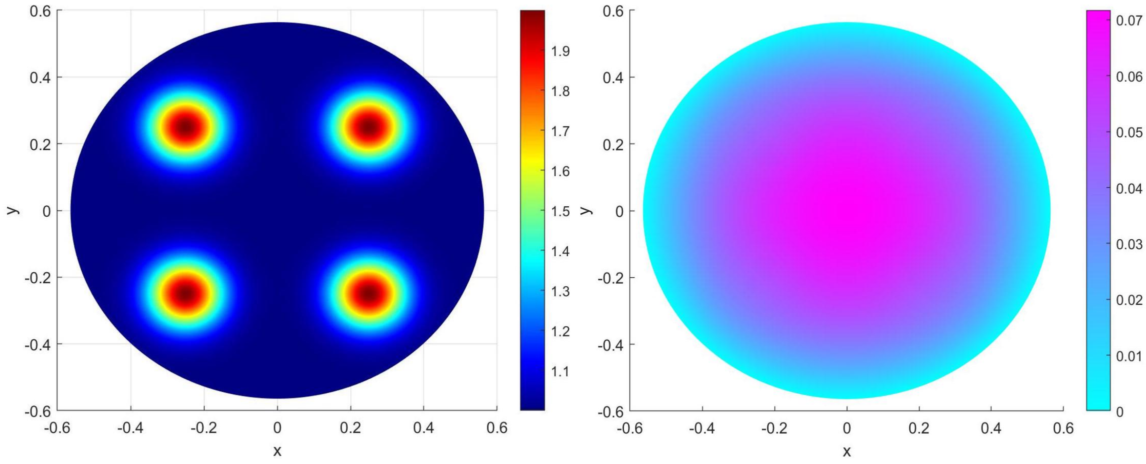


Fig. 1.2 Left: an example of scalar diffusion coefficient a. Right: the corresponding PDE solution u_a (for a given source function s)

scalar diffusion coefficient $a : \mathcal{O} \to [K_{min}, \infty),\ K_{min} > 0$, from noisy discrete observations of the solution $u = u_a$ of the elliptic boundary value problem

$$\begin{cases} \nabla \cdot (a \nabla u) = s, & x \in \mathcal{O}, \\ u(x) = 0, & x \in \partial\mathcal{O}, \end{cases} \tag{1.8}$$

for smooth $\partial\mathcal{O}$ and known positive source term s satisfying $\inf_{x\in\mathcal{O}} s(x) > 0$. While the PDE (1.8) is linear in u, its solution depends nonlinearly on the diffusion coefficient a, as can be observed already in the simplest one-dimensional setting with $\mathcal{O} = (0, 1)$ and constant $a(x) = a > 0$ and $s(x) = 1$:

$$a\frac{d^2u}{dx^2} = 1, \qquad x \in (0, 1), \quad u(0) = u(1) = 0,$$

with unique solution

$$u_a(x) = \frac{x(x-1)}{2a}, \qquad x \in [0, 1].$$

For any continuously differentiable $a \in C^1(\mathcal{O})$, the Schauder theory for elliptic PDEs (e.g., Theorem 6.14 in [89]) implies that the boundary value problem (1.8) has a unique classical solution $u_a \in C^2(\mathcal{O})$, resulting in the nonlinear forward map

$$G : \left\{ a \in C^1(\mathcal{O}) : \inf_{x\in\mathcal{O}} a(x) \geq K_{min} \right\} \to C^2(\mathcal{O}), \qquad a \mapsto u_a.$$

The positivity constraint on the source term s is a natural assumption that guarantees injectivity of the above forward map (cfr. [199, 176]), a necessary condition for the identification of a from observations of the solution u_a. The problem is illustrated in Figure 1.2; see Section 1.2.2 for numerical results.

1.1.2 Stochastic diffusion models

If it is of interest to explicitly take into account the time evolution nature of the advection-diffusion equation (1.1), an observation model of the form (1.6) may still be relevant, for example considering measurements of the solution $u(T,x)$, $x \in \mathcal{O}$, at some fixed time instant $T > 0$ (e.g, as in Example 2.2 in Chapter 2 for the recovery of the initial condition of the heat equation), or alternatively discrete observations over a collection of several time instants and spatial locations (e.g., as in [131]). A second class of statistical models arises from considering the microscopic components of the system. In particular, the process of advection causes the particles in the system to drift following the bulk motion of the fluid. Such particles are then further subject to random collisions with each other, with intensity depending on the characteristic of the medium, that perturb their trajectories and cause the phenomenon of diffusion. Formally, letting X_t denote the state of a particle at time $t \geq 0$, the trajectory $(X_t \ : \ t \geq 0)$ is modelled probabilistically as a random process satisfying the multi-dimensional SDE

$$dX_t = b(X_t)dt + A(X_t)dW_t, \qquad t \geq 0, \tag{1.9}$$

where $(W_t \ : \ t \geq 0)$ is a standard d-dimensional Brownian motion, and where we recall that $b = [b_1, \dots, b_d] : \mathbb{R}^d \to \mathbb{R}^d$ is the fluid velocity vector field and $A = [A_{ij}]_{i,j=1}^d : \mathbb{R}^d \to \mathbb{R}^{d,d}$ is the diffusion coefficient. Advection-diffusion equations of the form (1.1) then provide a description of the macroscopic behaviour of a system of particles evolving according to the SDE (1.9): specifically, for any sufficiently regular function $h : \mathbb{R}^d \to \mathbb{R}$ and any initial state $X_0 = x \in \mathbb{R}^d$, the conditional expectation $u(t,x) = E[h(X_t)|X_0 = x]$ satisfies the parabolic PDE (e.g., [15, Section 10.1])

$$\begin{cases} \partial_t u = \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^d a_{ik}a_{kj}\partial^2_{x_i x_j}u + \sum_{i=1}^d b_i \partial_{x_i} u, & t \geq 0, \quad x \in \mathbb{R}^d, \\ u(0,x) = h(x), & x \in \mathbb{R}^d. \end{cases}$$

The collection of all such expectations characterises the conditional law, given the initial state, of X_t for any time instant $t \geq 0$, thereby providing a description of the time evolution of the spatial distribution of a system composed by a large number of particles evolving according to the SDE (1.9).

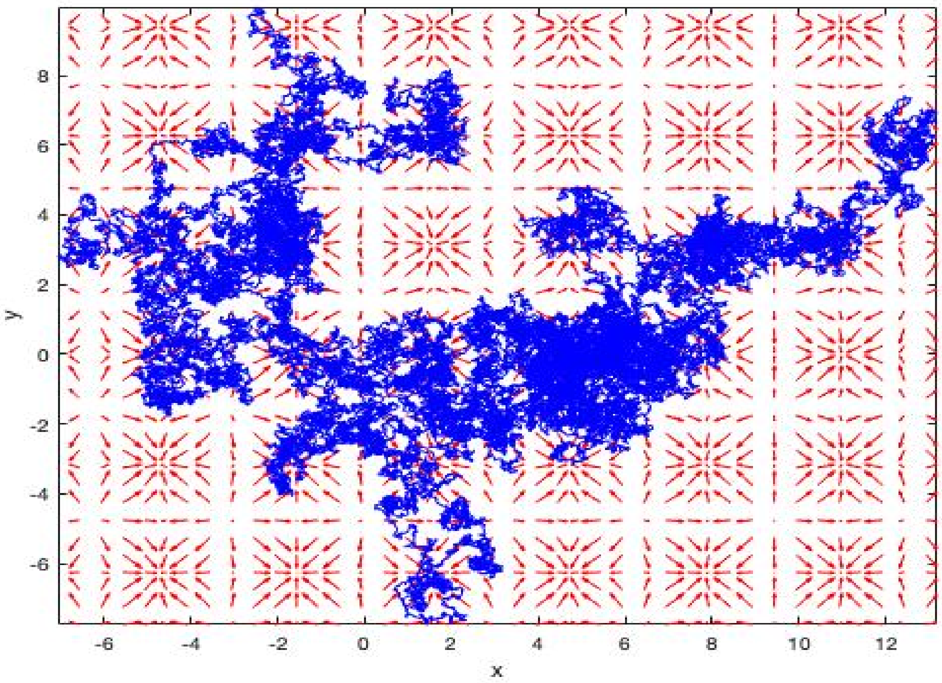


Fig. 1.3 In red, an example of gradient drift vector field, arising from the potential $B(x, y) = \cos^2 x + \sin^2 y$. In blue, a diffusion path.

Under suitable regularity conditions on A and b, there exists a unique strong pathwise solution to the SDE (1.9), that defines a d-dimensional Markov diffusion process (e.g., Chapters 24 and 39 in [19]). Assuming that observations on the state of a particle are available through time, a coherent measurement model then postulates data given by the continuous trajectory $X^T = (X_t \ : \ 0 \leq t \leq T)$ up to time $T > 0$, or of discrete samples thereof. The task at hand then consists in inferring either the diffusion coefficient A or the drift b. Such inferential problems naturally arise in a large variety of applications areas that involve dynamical models based on SDEs, including finance (e.g., [36, 121]), life sciences (e.g., [96]), physics (e.g., [58]) and biology (e.g., [245]).

In Chapter 4, a specific instance of the multi-dimensional SDE model (1.9) is considered, where the diffusion coefficient A is assumed to be known and set equal to the identity matrix of $\mathbb{R}^d$ (cfr. Section 4.3.4 for models with non-constant diffusivity), and the drift arises as a gradient vector field $b = \nabla B$ of some unknown potential $B : \mathbb{R}^d \to \mathbb{R}$,

$$dX_t = \nabla B(X_t)dt + W_t, \qquad t \geq 0.$$

Such assumption on the drift reflects a typical physical situation in which a particle moves within a potential energy field, modelled by the map B, which exerts a force on the particle directed towards its local extrema (e.g., [77]). Figure 1.3 shows a realisation of the diffusion path for a specific (periodic) choice of the potential.

1.2 Overview of results

The purpose of the present thesis is to investigate the theoretical properties of Bayesian nonparametric procedures in the PDE-based statistical models introduced above. Chapters 2 - 3 consider the inverse regression setting described in Section 1.1.1, while Chapter 4 studies the stochastic diffusion models introduced in Section 1.1.2. The next subsections provide an overview of the results derived in the later chapters, discussing the main mathematical challenges and the proofs strategies. The theory developed in Chapter 2 and Chapter 3 is illustrated via the results of simulations.

1.2.1 Statistical guarantees for Bayesian uncertainty quantification in linear inverse problems

Bernstein–von Mises theorems for linear inverse problems

In Chapter 2, statistical guarantees on the performance of Bayesian nonparametric procedures based on standard Gaussian priors are provided for a general class of linear inverse problems, in which the forward map defines an injective linear and continuous operator between separable Hilbert spaces

$$G : \mathcal{W}_1 \to \mathcal{W}_2.$$

For conciseness, we here present the results in the context of the source identification problem described in Section 1.1.1, entailing the recovery of the unknown source function $s \in L^2(\mathcal{O}) = \mathcal{W}_1$ in the elliptic boundary value problem (1.7) from noisy observations of its solution $Gs = u_s \in L^2(\mathcal{O}) = \mathcal{W}_2$. See Section 2.2 for the general theory and Section 2.3 for further examples.

The results are developed in a continuous analogue (in the sense of *Le Cam equivalence* [48, 197]) of the discrete inverse regression model (1.6), assuming observations of the functional equation

$$Y^\varepsilon = Gs + \varepsilon \mathbb{W}, \qquad \varepsilon > 0, \tag{1.10}$$

where $\mathbb{W}$ is a Gaussian white noise process indexed by $L^2(\mathcal{O})$; see Section 2.2.1 for details. In particular, observing data Y^ε in (1.10) is understood as observing a realisation of the Gaussian process $(Y^\varepsilon(\varphi) : \varphi \in \mathcal{W}_2)$, with marginal distributions $Y^\varepsilon(\varphi) \sim N(\langle Gs, \varphi\rangle_{L^2}, \varepsilon^2\|\varphi\|^2_{L^2})$. Providing that $Gs \in L^2(\mathcal{O})$, the (cylindrically defined) law $P_s^{Y^\varepsilon}$ of Y^ε in (1.10) is absolutely continuous with respect to the law $P_0^{Y^\varepsilon}$ of $\varepsilon\mathbb{W}$, with log-likelihood

$$\ell_\varepsilon(s) = \frac{1}{\varepsilon^2}\langle Y^\varepsilon, Gs\rangle_{L^2} - \frac{1}{2\varepsilon^2}\|Gs\|^2_{L^2}.$$

Assigning to s a prior distribution Π on $L^2(\mathcal{O})$ then induces, via Bayes' formula, the posterior distribution

$$\Pi(B|Y^\varepsilon) = \frac{\int_B e^{\ell_\varepsilon(s)} d\Pi(s)}{\int_{L^2} e^{\ell_\varepsilon(s')} d\Pi(s')}, \qquad B \subseteq L^2(\mathcal{O}) \text{ measurable.}$$

In particular, Gaussian priors represent a natural choice for Bayesian nonparametric inference in model (1.10), since, in view of the linearity of the forward map G, they are conjugate and thus allow for an explicit characterisation of the posterior (cfr. (1.16) and (1.17) below). Specifically, it is of interest to consider standard Gaussian priors whose specification does not require additional (and often unavailable) information about the forward map G, such as its singular value decomposition. In Chapter 2, a large class of Gaussian priors is considered with reproducing kernel Hilbert spaces (RKHSs) of Sobolev type: these include the frequently used Matérn processes, as well as Gaussian series on standard basis functions; see Section 2.3. For such priors, posterior inference can be concretely implemented exploiting the appropriate conjugate formulae.

As mentioned in the introduction (and as further elaborated in Section 2.1), the properties of Bayesian credible sets depend on the fine interplay between the local asymptotic behaviour of the posterior distribution and the geometries involved, which in the nonparametric setting may give rise to substantial obstructions even in the simplest direct sequence space model [95]. To address these issues, in Chapter 2 we build on the *semi-parametric* approach developed by Castillo and Nickl [52, 53] (and refined in [166] and [171] in the inverse problems setting), and study the induced posterior distributions on one-dimensional functionals of the form $\langle s, \psi\rangle_{L^2}$, for test functions $\psi \in L^2(\mathcal{O})$ satisfying a minimal (Sobolev) regularity condition. Our main result is a *semiparametric Bernstein–von Mises theorem* (cfr. Theorem 2.2), which, under mild assumptions on the true source s_0 and on the prior regularity, identifies a precise Gaussian limit for the (centred and scaled) one-dimensional posterior distributions:

$$\mathcal{L}\left(\frac{1}{\varepsilon}\langle s-\bar{s}_\varepsilon,\psi\rangle_{L^2}\Big|Y^\varepsilon\right)\xrightarrow{\mathcal{L}}N(0,\|\nabla\cdot(A\nabla\psi)\|^2_{L^2}),\qquad \varepsilon\to 0, \tag{1.11}$$

in $P^Y_{s_0}$-probability. The asymptotic variance $\|\nabla\cdot(A\nabla\psi)\|^2_{L^2}$ is minimal, as it can be shown to coincide with the Cramér-Rao lower bound for estimating $\langle s_0,\psi\rangle_{L^2}$ from data (1.10) (cfr. Appendix 2.A). Above, $\bar{s}_\varepsilon = E^\Pi[s|Y^\varepsilon]$ is the posterior mean, which in view of the Gaussianity of the posterior distribution coincides with the posterior mode, and hence can be characterised as a *Tikhonov regulariser* found by minimising the penalised least squares functional

$$Q(s)=-\ell_\varepsilon(s)+\frac{1}{2}\|s\|^2_{H^\alpha}=-\frac{1}{\varepsilon^2}\langle Y^\varepsilon,Gs\rangle_{L^2}+\frac{1}{2\varepsilon^2}\|Gs\|^2_{L^2}+\frac{1}{2}\|s\|^2_{H^\alpha}, \tag{1.12}$$

with squared Sobolev norm penalty $\|\cdot\|^2_{H^\alpha}$ induced by the prior; see Section 2.2.2 for details. From the convergence of moments in (1.11), a central limit theorem for the plug-in Tikhonov regularisers can then be deduced as well (cfr. Remark 2.1):

$$\frac{1}{\varepsilon}\left(\langle\bar{s}_\varepsilon,\psi\rangle_{L^2}-\langle s_0,\psi\rangle_{L^2}\right)\xrightarrow{d}N(0,\|\nabla\cdot(A\nabla\psi)\|^2_{L^2}),\qquad \varepsilon\to 0,$$

showing, in view of the minimality of the asymptotic variance $\|\nabla\cdot(A\nabla\psi)\|^2_{L^2}$, that $\langle\bar{s}_\varepsilon,\psi\rangle_{L^2}$ is an asymptotically efficient estimator of $\langle s_0,\psi\rangle_{L^2}$.

Finally, alongside efficiency of the plug-in Tikhonov regularisers, the Bernstein–von Mises result (1.11) also implies that Bayesian credible intervals built around such estimators are asymptotically valid and optimal confidence intervals. In particular, for any $\alpha\in(0,1)$, the $(1-\alpha)\%$-credible interval

$$C_\varepsilon=\{x\in\mathbb{R}\ :\ |\langle\bar{s}_\varepsilon,\psi\rangle_{L^2}-x|\le R_\varepsilon\},\qquad \Pi\left(s:\langle s,\psi\rangle_{L^2}\in C_\varepsilon|Y^\varepsilon\right)=1-\alpha,$$

for $R_\varepsilon>0$ the appropriate posterior quantile, has asymptotically $(1-\alpha)\%$ frequentist coverage,

$$P^{Y^\varepsilon}_{s_0}\left(\langle s_0,\psi\rangle_{L^2}\in C_\varepsilon\right)\to 1-\alpha,\qquad \varepsilon\to 0,$$

and radius shrinking at optimal (parametric) rate $R_\varepsilon = O_{P^{Y^\varepsilon}_{s_0}}(\varepsilon^{-1})$; see Corollary 2.2.

The results presented in the chapter thus shows that semi-parametric posterior estimation and uncertainty quantification based on standard Gaussian priors, which can be straightforwardly implemented in practice thanks to conjugacy, are valid and optimal from a frequentist point of view. In fact, exploiting the uniformity in the limit (1.11) for ψ in a Sobolev ball, a nonparametric version of the result is further derived (cfr. Theorem

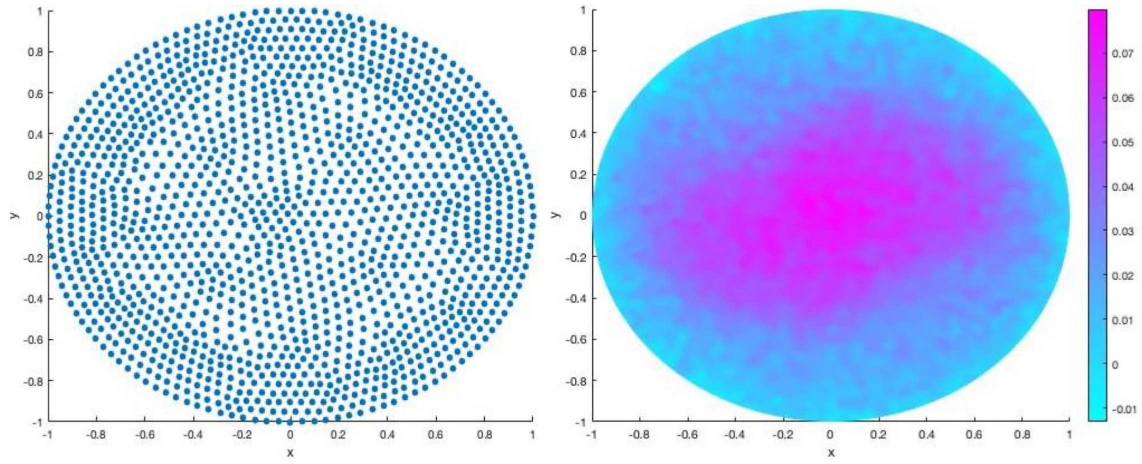


Fig. 1.4 Left: $N = 1453$ nodes in a deterministic triangular mesh. Right: $N = 1453$ noisy discrete observation of the PDE solution Gs_0 over the design points shown on the left (cfr. Figure 1.1 for the noiseless PDE solution).

2.3), that entails the weak convergence, in a dual Sobolev space topology, of the (scaled and centred) posterior distribution to a fixed Gaussian measure whose one-dimensional marginals are identified by the right-hand side of (1.11). As a consequence, a central limit theorem for the posterior mean $\bar{s}_\varepsilon$ is derived, as well as frequentist coverage properties for certain credible balls centred around $\bar{s}_\varepsilon$ in the dual space where the convergence is attained.

Implementation of the algorithm

For the inverse problem of recovering the unknown source function s in the elliptic boundary value problem (1.7), we take the unit disk $\mathcal{O} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ as the working domain, and we assume in practice that we are given N noisy point evaluations of the solution $Gs = u$ over a deterministic grid $\{x_1, \dots, x_N\} \subset \mathcal{O}$,

$$Y_i = Gs(x_i) + \sigma W_i, \qquad W_i \overset{\text{iid}}{\sim} N(0, 1), \quad \sigma > 0. \tag{1.13}$$

Specifically, we take the design point as the nodes in a triangular mesh in $\mathcal{O}$ (see Figure 1.4, left).

We discretise the parameter space by modelling the unknown source function s as a finite sum

$$s = \sum_{m=1}^{M} s_m \phi_m, \qquad s_m \in \mathbb{R}, \tag{1.14}$$

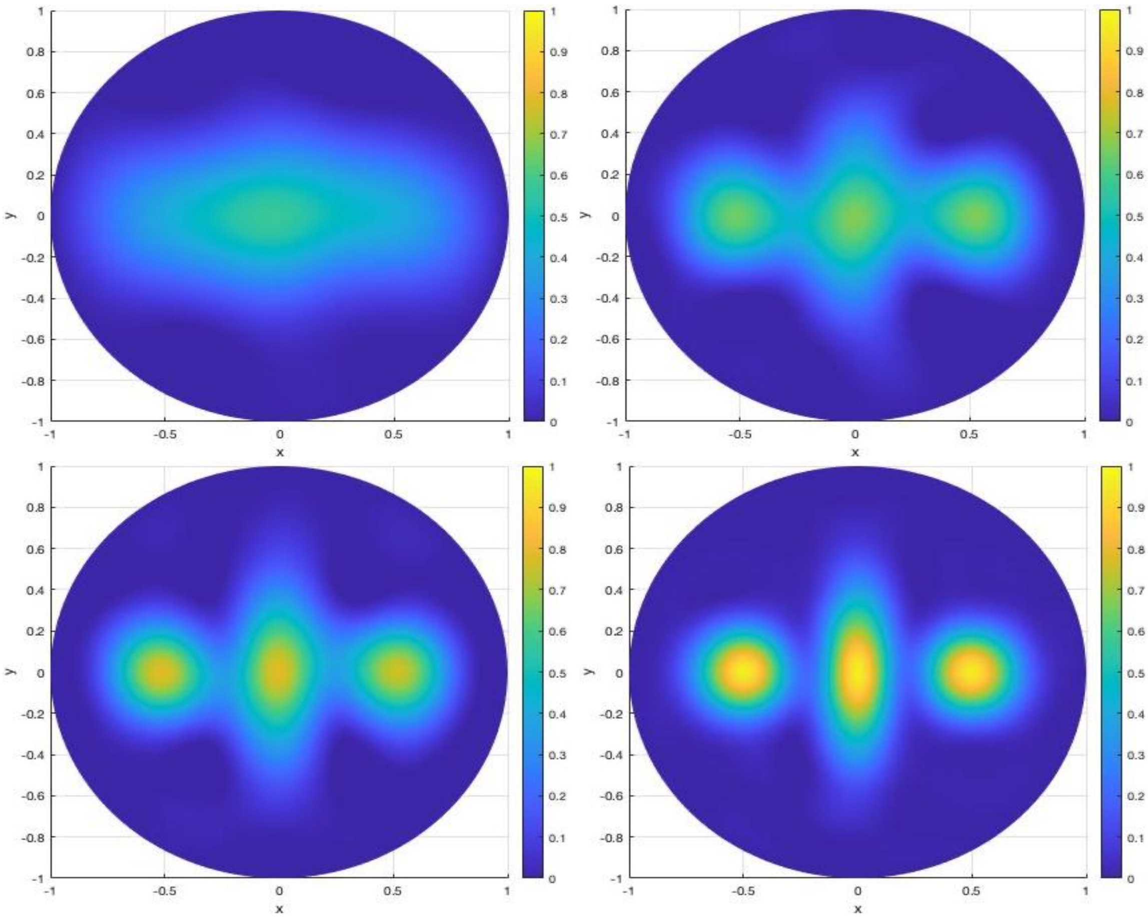


Fig. 1.5 Left to right, top to bottom: posterior mean estimates $\bar{\mathbf{s}}_\sigma$ of the source function s for decreasing values of the noise standard deviation $\sigma = .005, .001, .0005, .0001$.

where $\{\phi_m : m \geq 1\}$ is the family of eigenfunctions of the Laplacian on the disk [61] (satisfying zero Dirichlet boundary conditions), with corresponding eigenvalues $\{\lambda_m : m \geq 1\}$. Such *disk harmonics* form an orthonormal system in $L^2(\mathcal{O})$ and span the regularity scale of Sobolev functions (with zero-trace). Identifying s in (1.14) with the coefficient vector $\mathbf{s} = [s_1, \dots, s_M]$, the observation model (1.13) then becomes, in vectorial notation,

$$\mathbf{Y} = \mathrm{G}\mathbf{s} + \sigma\mathbf{W}, \tag{1.15}$$

where $\mathbf{Y} = [Y_1, \dots, Y_n]$, $\mathbf{W} = [W_1, \dots, W_n] \sim N(0, I_N)$ for I_N the identity matrix of $R^{N,N}$, and where G is the discretisation of the forward operator G given by the $N \times M$ matrix with entries $\mathrm{G}_{im} = G\phi_m(x_i)$. The computation of the discretised forward operator is performed using Matlab PDE Toolbox (which also provides the code to generate the triangular mesh) and the implementation of the disk harmonics within the Chebfun package [244].

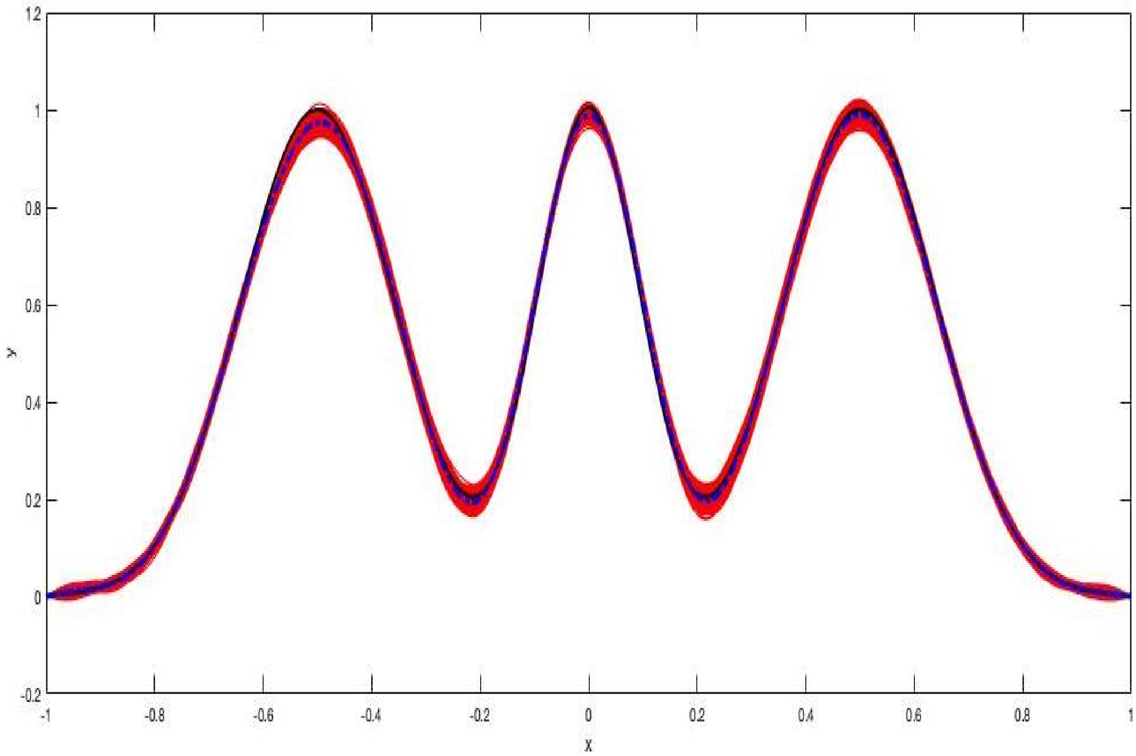


Fig. 1.6 In red, cross-section along the x-axis of 2500 posterior samples. In black and dotted blue respectively, the cross-section of the true source function s_0 and of the posterior mean estimate $\bar{\mathbf{s}}_\sigma$.

For given $\mathbf{s} \in \mathbb{R}^M$, the observation $\mathbf{Y}$ in (1.15) has conditional distribution $\mathbf{Y}|\mathbf{s} \sim N(\mathrm{G}\mathbf{s}, \sigma^2 I_N)$. Specifying a prior distribution $\mathbf{s} \sim N(0, \Gamma)$, for some known covariance matrix $\Gamma \in \mathbb{R}^{M,M}$, then gives, by a standard conjugate computation, the posterior

$$\mathbf{s}|\mathbf{Y} \sim N(\bar{\mathbf{s}}_\sigma, \Gamma_\sigma), \tag{1.16}$$

where

$$\Gamma_\sigma = \left(\sigma^{-2}\mathrm{G}^T\mathrm{G} + \Gamma^{-1}\right)^{-1}; \qquad \bar{\mathbf{s}}_\sigma = \sigma^{-2}\Gamma_\sigma \mathrm{G}^T\mathbf{Y}. \tag{1.17}$$

For concrete instances of the prior covariance matrix Γ, the posterior mean $\bar{\mathbf{s}}_\sigma$ and covariance matrix Γ_σ can readily be computed from the data, allowing direct posterior sampling according to (1.16). In particular, natural choices are diagonal prior covariance matrices modelling Sobolev regularities, of the form $\Gamma = \mathrm{diag}(\lambda_1^{-\alpha}, \dots, \lambda_M^{-\alpha})$ for $\alpha > 0$, and where $\{\lambda_m : m \geq 1\}$ are the Laplacian eigenvalues.

Figure 1.4 (right) shows noisy discrete observations of the PDE solution corresponding to the true source function $s_0(x, y) = e^{-(5x-2.5)^2-(5y)^2} + e^{-(7.5x)^2-(2.5y)^2} + e^{-(5x-2.5)^2-(5y)^2}$ (shown above in Figure 1.1, left). The diffusion coefficient was taken to be $A = aI_2$, for I_2 the identity matrix of $\mathbb{R}^{2,2}$ and for, in polar coordinates, $a(r,\theta) = 40.5 - 40\left\{\cos\left[(\sin(\pi r)\cos(\theta) + \sin(2\pi r)\sin(\theta))/4\right]\right\}$, $r \in [0,1]$, $\theta \in [0,2\pi)$. The triangular mesh used has $N = 1453$ nodes, and the noise standard deviation was set $\sigma = .005$. The posterior mean estimates obtained for decreasing values of the noise standard deviation $\sigma = .005, .001, .0005, .0001.$ are shown in Figure 1.5, to be compared to the true source function pictured in 1.1, left. For all four estimates, an expansion (1.14) with $M = 210$ basis function was used, and the prior covariance matrix was taken to be $\Gamma = \mathrm{diag}(\lambda_1^{-2}, \dots, \lambda_{210}^{-2})$. Finally, Figure 1.6 provides a visualisation of the Bayesian uncertainty quantification, showing the cross section along the x-axis of 2500 posterior samples (obtained with $\sigma = .0005$).

1.2.2 Consistent Bayesian inference in nonlinear inverse problems

Posterior contraction rates and convergence of the posterior mean

In Chapter 3, we turn to the study of nonlinear inverse problems. We focus on the representative elliptic example introduced in Section 1.1.1, consisting in the recovery of the unknown scalar positive diffusion coefficient a from noisy discrete observations of the solution $G(a) = u$ of the elliptic PDE (1.8). In particular, we assume observations

$$Y_i = G(a)(X_i) + \sigma W_i, \qquad W_i \overset{\text{iid}}{\sim} N(0,1), \quad \sigma > 0, \tag{1.18}$$

for design points $X_i \overset{\text{iid}}{\sim} \mathrm{U}(\mathcal{O})$ uniformly drawn at random on $\mathcal{O}$. For any fixed $a \in C^1(\mathcal{O})$ satisfying $\inf_{x\in\mathcal{O}} a(x) \geq K_{min}$, $K_{min} > 0$, the law P_a^N of $(Y^{(N)}, X^{(N)}) = ((Y_i)_{i=1}^N, (X_i)_{i=1}^N)$ on $\mathbb{R}^N \times \mathcal{O}^N$ is then absolutely continuous with respect to the (product) Lebesgue measure, with negative log-likelihood equal to, up to an additive constant, the least-squares functional

$$-\ell_N(a) = \frac{1}{2\sigma^2}\sum_{i=1}^N \left[Y_i - G(a)(X_i)\right]^2. \tag{1.19}$$

As observed in Section 1.1.1, the PDE solution $G(a)$ depends nonlinearly on the diffusion coefficient a, implying that the above functional is non-convex. As a consequence, commonly used recovery methods (such as maximum likelihood estimators or Tikhonov regularisers) defined as optimisers of likelihood-based objective functionals cannot be reliably computed by standard convex optimisation techniques. Bayesian procedures represent an attractive alternative non-optimisation based approach to estimation and uncertainty quantification in such non-convex statistical models, that can reliably be implemented via modern MCMC methodology; see Section 3.1 for further discussion. In particular, Gaussian priors are widely used in Bayesian inverse problems [220, 73], due to advantageous methodological and computational features, including a natural connection with classical regularisation methods (cfr. (1.12)) and the existence of *ad hoc* sampling methods (such as the pCN algorithm used below), as well as for their rich theory on sample paths properties.

In Chapter 3, we study posterior consistency and the convergence of the posterior mean for Bayesian nonparametric procedures based on a large class of Gaussian (and hierarchical Gaussian) priors with RKHSs of Sobolev type, suitably parametrised to incorporate the positivity constraint on a. Our main results (cfr. Theorems 3.2 and 3.5) show that the posterior distributions arising from such priors concentrate in L^2-distance around the true diffusion coefficient a_0 generating the data, with an upper bound (algebraic in inverse sample size) on the contraction rates:

$$\Pi\left(a : \|a - a_0\|_{L^2} > LN^{-\lambda} | Y^{(N)}, X^{(N)}\right) \xrightarrow{P_{a_0}} 0, \tag{1.20}$$

as $N \to \infty$ for large enough $L > 0$ and some $\lambda > 0$ (explicitly characterised in the proofs). As a corollary, we derive the same convergence rates for the corresponding posterior mean estimators $\bar{a}_N = E^{\Pi}[a|Y^{(N)}, X^{(N)}]$ (cfr. Theorems 3.3 and 3.6),

$$\|\bar{a}_N - a_0\|_{L^2} = O_{P_{a_0}^N}\left(N^{-\lambda}\right), \qquad N \to \infty. \tag{1.21}$$

We briefly outline the proofs. Applying ideas developed in [167] for an inverse problem with (non-Abelian) X-ray transforms, we first consider the related PDE-constrained regression problem of estimating the true PDE solution $G(a_0)$ from data (1.18). Theorems 3.1 and 3.4 show that, if a_0 is an α-regular Sobolev function satisfying some minimal smoothness assumptions, the induced (non-Gaussian) posterior distributions on the PDE solution $G(a)$ concentrates around $G(a_0)$ at the (minimax) optimal posterior contraction rate,

$$\Pi\left(a : \|G(a) - G(a_0)\|_{L^2} > LN^{-\frac{\alpha+2}{2\alpha+4+d}} | Y^{(N)}, X^{(N)}\right) \xrightarrow{P_{a_0}} 0, \tag{1.22}$$

as $N \to \infty$ for sufficiently large $L > 0$. Theorems 3.1 and 3.4 are based on general theory for posterior contraction rates in forward risk developed in Appendix 3.A for a class of nonlinear inverse problems with forward map satisfying local Lipschitz regression estimates and boundedness conditions (see (3.32) and (3.33) below). Under these assumptions, the Hellinger contraction rates theory for Gaussian priors (for direct models) [235] can be successfully employed in the inverse problems setting exploiting a standard inequality for the Hellinger distance due to Birgé [30]. For the elliptic example considered here, the specific estimates used were established in [176].

The posterior contraction rates (1.20) for the diffusion coefficient then follow combining (1.22) with a suitable *stability estimate*, that is, a 'local' Lipschitz estimate (uniform on sets of bounded Hölder norm) for the inverse of the forward map G, proved in [176],

$$\|a - a_0\|_{L^2} \lesssim \|a\|_{C^1} \|G(a) - G(a_0)\|_{H^2},$$

holding for all sufficiently regular a, a_0, with multiplicative constant independent of a. Lastly, we deduce the convergence rates (1.21) for the posterior mean using uniform integrability arguments applied to the contracting posterior distributions.

Apart from the elliptic inverse problem considered here, the general framework assumed in Appendix 3.A further encompasses several other nonlinear inverse problems, including the Schrödinger equation [171, 176, 177] and the parabolic Schrödinger equation [131], as well as many linear inverse problems such as the classical Radon transform [176], for which the general theory developed in Chapter 3 provides a template to obtain posterior contraction rates via regression and stability estimates.

The local nature of the Lipschitz condition on the forward map G and of the above stability estimate implies the necessity of additional *a priori* regularisation in order to quantitatively control the norm of the posterior draws (in the relevant function spaces). In Chapter 3, we first address such issues introducing a sample size dependent scaling of a base standard Gaussian prior (cfr. (3.11)), which implies asymptotic concentration of the posterior distribution over sets of bounded Hölder norms. Secondly, we show that sufficient regularisation can alternatively be achieved by (fully Bayesian) hierarchical prior modelling, introducing a random truncation point with a suitable hyper-prior in the series expansion of the base Gaussian prior (cfr. (3.19)). For such hierarchical Gaussian priors, the forward contraction rates (1.22) for the PDE solution are also shown to be adaptive to the smoothness α.

Implementation of the algorithm

We take the unit disk $\mathcal{O} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ as the working domain, and assume we are given N noisy discrete evaluations of the PDE solution as in (1.18) over a grid of design points uniformly drawn at random on $\mathcal{O}$, shown in Figure 1.7.

Following the approach in Chapter 3, the positivity constraint on the diffusion coefficient a is incorporated by modelling a as the composition $a = \Phi \circ a'$, for $a' : \mathcal{O} \to \mathbb{R}$ and where $\Phi : \mathbb{R} \to [K_{min}, \infty)$, $K_{min} > 0$, is a fixed bijective smooth link function (see

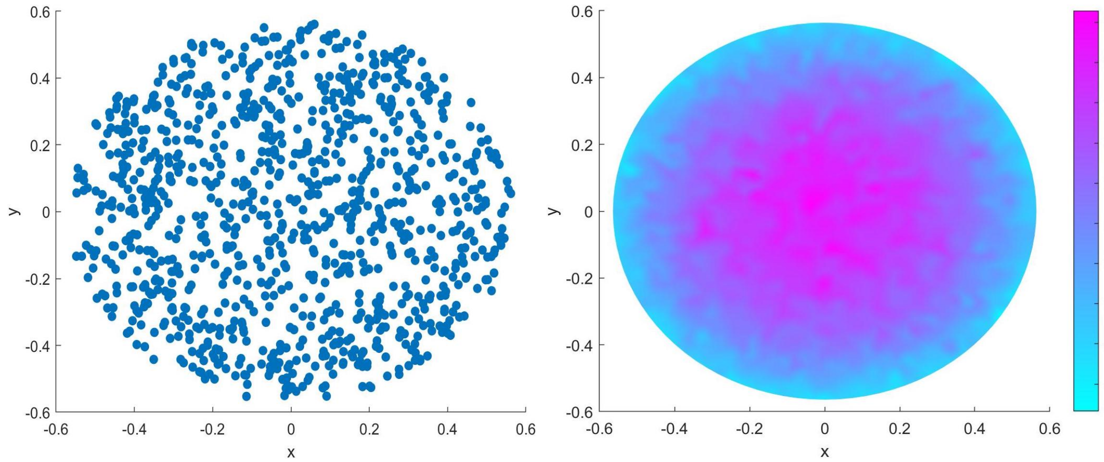


Fig. 1.7 Left: mesh with $N = 1000$ random design points. Right: $N = 1000$ noisy discrete observation of the PDE solution $G(a_0)$ over the design points shown on the left (cfr. Figure 1.2 for the noiseless PDE solution).

Section 3.2.2 for details). We then discretise the parameter space by assuming that a' is given by the finite sum

$$a' = \sum_{m=1}^{M} a'_m \psi_m, \qquad a_m \in \mathbb{R}, \tag{1.23}$$

where $\{\psi_1, \dots, \psi_M\}$ are piecewise linear functions on a deterministic triangular mesh with nodes $\{z_1, \dots, z_M\} \subset \mathcal{O}$ (displayed in Figure 1.8), uniquely characterised by the relation $\psi_m(z_{m'}) = 1_{\{m=m'\}}$. Accordingly, a' in (1.23) satisfies $a'(z_m) = a'_m$, and for any $x \in \mathcal{O}$ the value $a'(x)$ is obtained by linear interpolation over the pairs $\{(z_m, a_m) \ : \ m = 1, \dots, M\}$. We note that a discretisation approach based on an orthonormal system of basis functions as described in Section 1.2.1 could be used as well.

We assign to a' a Gaussian process prior Π, which can readily be implemented under the discretisation described above. In particular, if Π arises as the law of a centred Gaussian process on $\mathcal{O}$ with covariance function $C(x, x')$, $x, x' \in \mathcal{O}$, a sample $a' \sim \Pi$ is drawn in practice by sampling the vector of coefficients $\mathbf{a}' = [a'_1, \dots, a'_M]$ in (1.23) from the multivariate Gaussian distribution on $\mathbb{R}^M$

$$\mathbf{a}' \sim N(0, \mathrm{C}), \qquad \mathrm{C} = [\mathrm{C}_{mm'}]_{m,m'=1}^{M}, \quad \mathrm{C}_{mm'} = C(z_m, z'_m).$$

Specifically, we consider covariance functions of Matérn type, given by

$$C(x, x') = c(|x - x'|), \qquad c(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r\sqrt{2\nu}}{\ell} \right)^{\nu} K_\nu \left(\frac{r\sqrt{2\nu}}{\ell} \right), \quad r > 0,$$

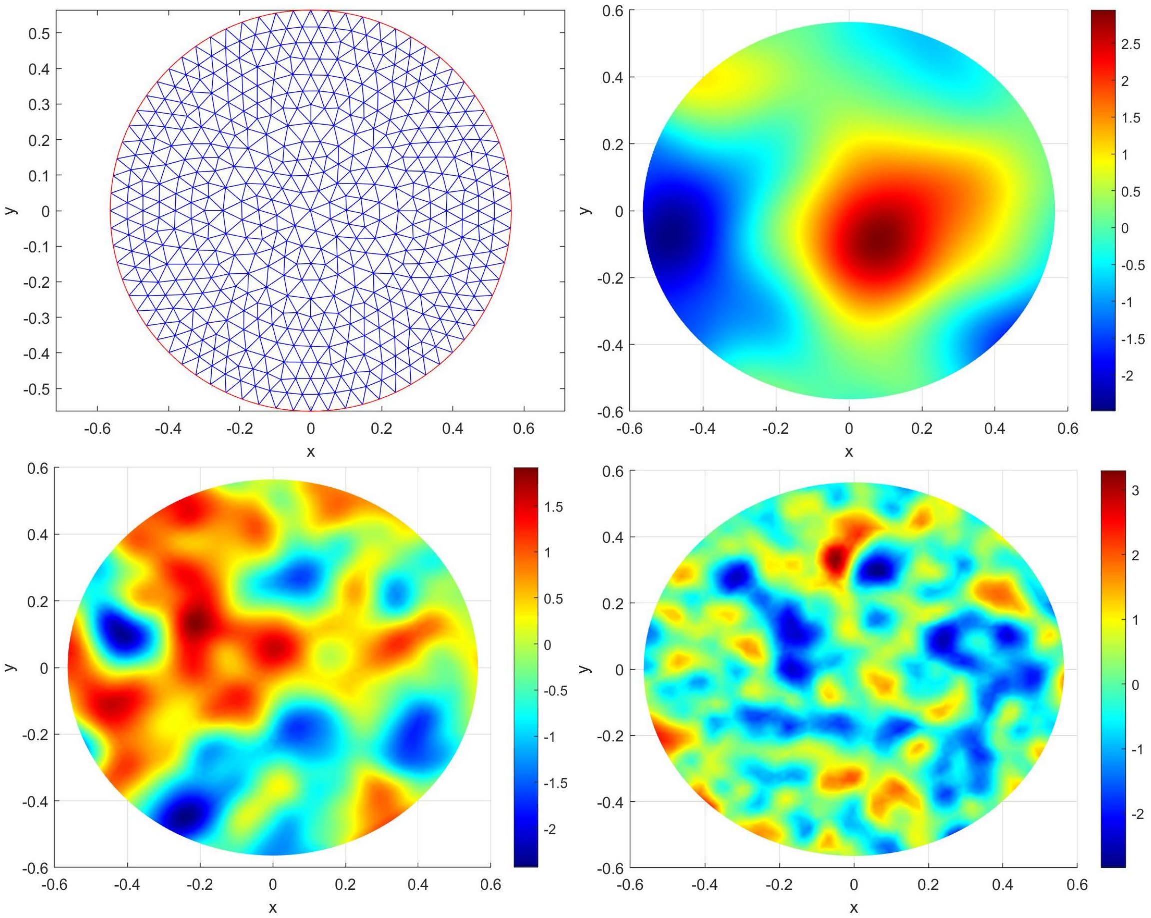


Fig. 1.8 Left to right, top to bottom: deterministic triangular mesh with $M = 1969$ nodes; samples from a Matérn process prior with smoothness parameter $\nu = 5$ and decreasing length parameter $\ell = .25, .1, .05$ respectively.

where Γ denotes the gamma function and K_ν is the modified Bessel function of the second kind. The hyperparameter $\nu > 0$ controls the Sobolev regularity of the sample paths, while $\ell > 0$ determines the characteristic length-scale (cfr. Figure 1.8).

Unlike the conjugate situation considered in Chapter 2, in the nonlinear setting the posterior distribution is in general not available in closed form. For Gaussian priors, specific MCMC algorithms have been developed to (approximately) sample from the posterior distributions, including the preconditioned Crank-Nicholson (pCN) method [63]. In the present setting, the pCN algorithm generates a Markov chain $(\vartheta_k)_{k\geq 1}$ with invariant measure equal to the posterior distribution of a', starting from an initialisation point ϑ_0 and then, for $k \geq 0$, repeating the following steps:

1. draw a prior sample $\xi \sim \Pi$ and for $\delta > 0$ define the proposal $p = \sqrt{1-2\delta}\vartheta_k + \sqrt{2\delta}\xi$;

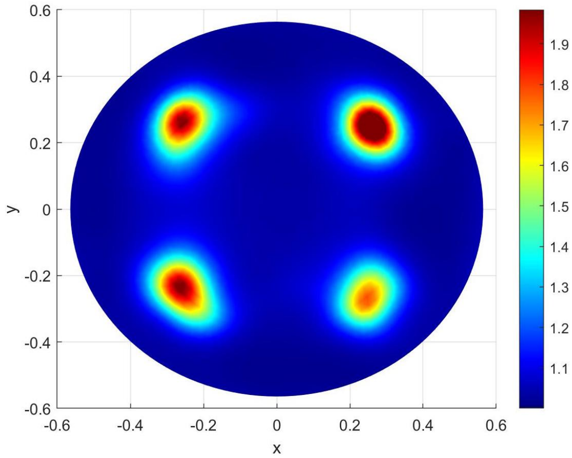


Fig. 1.9 Posterior mean estimate $\bar{a}_N$ of the diffusion coefficient a, computed via the pCN algorithm.

2. set
$$\vartheta_{k+1} = \begin{cases} p, & \text{with probability } 1 \wedge e^{\ell_N(\Phi\circ p)-\ell_N(\Phi\circ\vartheta_k)}, \\ \vartheta_k, & \text{otherwise,} \end{cases}$$
where ℓ_N is the log-likelihood function in (1.19).

For each iteration, step 2. requires the evaluation of the log-likelihood $\ell_N(\Phi\circ p)$, which in turn entails the numerical evaluation of the PDE solution $G(\Phi\circ p)$ at the design points $\{X_1,\dots,X_N\}$ (cfr. (1.19)), performed in practice using Matlab PDE Toolbox (which also provide an implementation of the triangular mesh for the discretisation of the parameter space). Prior samples $\xi \sim \Pi$ required in step 1. are drawn as described above.

The algorithm is terminated after K steps, returning approximate samples $(\vartheta_k : k = 0,\dots,K)$ from the posterior distribution of a'. Under certain assumptions on the forward map that are compatible with the present setting, Hairer, Stuart and Vollmer [113] derived dimension-free spectral gaps which imply rapid convergence towards the invariant measure. As a consequence, the posterior mean can be reliably computed numerically by the MCMC average
$$\bar{\vartheta} = \frac{1}{K+1}\sum_{k=0}^{K}\vartheta_k,$$
with non-asymptotic bounds for the numerical approximation error. Posterior credible sets can likewise be reliably computed by considering the empirical quantiles of the pCN samples.

Figure 1.7 (right) shows $N = 1000$ noisy discrete observations of the PDE solution corresponding to the true diffusion coefficient $a_0(x,y) = e^{-(10x-2.5)^2-(10y-2.5)^2} +$

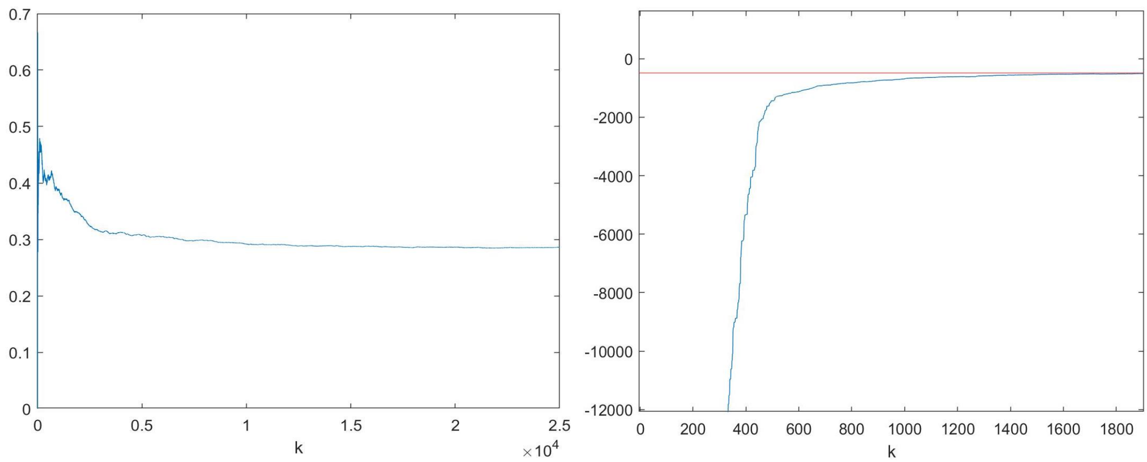


Fig. 1.10 Left: acceptance rate over $K = 25000$ pCN samples. The rate stabilises around 30% after the initial burn-in time (first 2000 iterates). Right: in blue, the log-likelihood $\ell_N(\Phi \circ \vartheta_k)$ of the first 2000 iterates (burn-in time); in red, the loglikelihood $\ell_N(a_0)$ of the true diffusion coefficient a_0.

$e^{-(10x-2.5)^2-(10y*2.5)^2} + e^{-(10x+2.5)^2-(10y+2.5)^2} + e^{-(10x+2.5)^2-(10y-2.5)^2}$ (shown above in Figure 1.2, left). The source function was taken to be constant $s(x, y) = 1$. The noise standard deviation was set $\sigma = .005$. The posterior mean estimate, computed via the pCN algorithm (with $K = 25000$ iterations), is shown in Figure 1.9, to be compared to the true diffusion coefficient pictured in Figure 1.2, left. The parameter space was discretised using a triangular mesh with $M = 1969$ nodes, and the hyperparameter for the Matérn process prior were set $\nu = 5$ and $\ell = .125$. The step-size δ for the pCN algorithm was set $\delta = .00125$, tuned so that after the initial burn-in time (here seen to roughly correspond to the first 2000 iterates), the acceptance rate of new proposals stabilises around 30%; see Figure 1.10 (left). A non-informative initialisation point $\vartheta_0 = 0$ for the pCN algorithm was chosen. Figure 1.10 (right) shows the log-likelihood $\ell_N(\Phi \circ \vartheta_k)$ over the burn-in time, seen to rapidly increase towards, and then stabilise around, the log-likelihood $\ell_N(a_0)$ attained by the true diffusion coefficient a_0.

1.2.3 Consistent Bayesian inference in multi-dimensional stochastic diffusion models

In the fourth and final chapter of the present thesis we investigate the posterior consistency of Bayesian nonparametric procedures in the stochastic diffusion setting introduced in Section 1.1.2. In particular, we study the problem of estimating a twice-continuously differentiable potential function $B : \mathbb{R}^d \to \mathbb{R}$ from continuous observations $X^T = (X_t : 0 \le t \le T)$, $T > 0$, of the trajectory of the multi-dimensional diffusion process governed by the SDE

$$dX_t = \nabla B(X_t) + dW_t, \qquad t \geq 0, \tag{1.24}$$

for some initial condition $X_0 = x_0 \in \mathbb{R}^d$, where $(W_t : t \geq 0)$ is a standard d-dimensional Brownian motion. Drift vector fields in gradient form naturally arise in physical applications in the presence of potential energy fields; see Section 4.1 for more discussion.

In order for consistent estimation of B from X^T to be possible, additional assumptions on the model are typically needed to ensure suitable *recurrence* properties on the diffusion sample paths, so that growing information about the value $B(x)$ at each $x \in \mathbb{T}^d$ is accumulated as the time horizon T increases. As in [181, 187, 1, 173] among others, we here assume that the unknown potential B is periodic, and regard it as a function defined on the d-dimensional torus $\mathbb{T}^d$. In analogy to Kolmogorov's characterisation of time-reversible diffusions on $\mathbb{R}^d$ (e.g., [15]), it then follows that the periodised diffusion (modulo $\mathbb{Z}^d$) is reversible on $\mathbb{T}^d$, with unique invariant probability density function

$$\mu_B(x) = \frac{e^{2B(x)}}{\int_{\mathbb{T}^d} e^{2B(x')}dx'}, \qquad x \in \mathbb{T}^d, \tag{1.25}$$

for which it holds that, for any continuous and one-periodic $\varphi \in C(\mathbb{T}^d)$,

$$\frac{1}{T}\int_0^T \varphi(X_t) \to \int_{\mathbb{T}^d} \varphi(x)\mu_B(x)dx$$

in probability under the law of the data as $T \to \infty$. See Section 4.2.2 for details. In such periodic setting, for any $B \in C^2(\mathbb{T}^d)$, Girsanov's theorem (e.g., [19]) implies that the law P_B^T of X^T in (1.24) is absolutely continuous with respect to the law P_0^T of $(W_t : 0 \leq t \leq T)$, with log-likelihood

$$\ell_T(B) = -\frac{1}{2}\int_0^T \|\nabla B(X_t)\|^2 dt + \int_0^T \nabla B(X_t).dX_t. \tag{1.26}$$

In Chapter 4, two major classes of priors for B on $C^2(\mathbb{T}^d)$ are considered, modelling the unknown potential according to either a Gaussian or a p-exponential prior. Gaussian priors are widely used in drift estimation problems for diffusions (e.g., [187, 110, 238, 20, 173]), and, as we show in Section 4.3.1, in the present setting represent a natural choice in view of conjugacy, which stems from the quadratic dependence on B of the log-likelihood in (1.26). The class of p-exponential priors [5] (which include Besov priors, popular in the inverse problem community, e.g., [239, 209, 73, 4]) generalise the series construction of Gaussian priors allowing for heavier-tailed random coefficients. Such added flexibility has been observed to imply a number of desirable properties, including sparsity-promoting and edge-preserving reconstruction, suited to the recovery of spatially-irregular, blocky structures such as images. While in general non-conjugate, p-exponential prior maintain a log-concave structure which, in view of the quadratic form of the log-likelihood in (1.26), implies that posterior inference in the reversible diffusion model (1.24) can be approached with convex optimisation and sampling techniques; see Remark 4.1.

Our main results (cfr. Theorems 4.1 and 4.2) show that, for suitable choices of the prior hyperparameters, the posterior distributions arising from both classes of priors concentrates around the true potential B_0 generating the data at the (minimax) optimal posterior contraction rates. In particular, if B_0 is an $(\alpha+1)$-regular Sobolev function, then

$$\Pi\left(B : \|\nabla B - \nabla B_0\|_{L^p} > LT^{-\frac{\alpha}{2\alpha+d}} | X^T\right) \xrightarrow{P_{B_0}} 0 \tag{1.27}$$

as $T \to \infty$ for large enough $L > 0$. Above, $p = 2$ in the case of Gaussian priors. The result is formulated in terms of the distance $\|\nabla B - \nabla B_0\|_{L^p}$, since B is identifiable only up to additive constants in view of the dependence of $\ell_T(B)$ on ∇B in (1.26). In fact, after a standard identifiability assumption, the same contraction rates in (1.27) are attained in the Sobolev norm $\|B - B_0\|_{W^{1,p}}$; see the discussion after Theorem 4.1.

While the asymptotic theory for Bayesian nonparametric procedures for drift estimation is well-developed in the one-dimensional case $d = 1$ (e.g., [232, 238, 174, 1]), the multi-dimensional setting $d \geq 2$ is considerably more challenging and essentially unexplored, with the exception of the recent work by Nickl and Ray [173] for *non-reversible* diffusions. Specifically, the key challenges in studying the asymptotic behaviour of posterior distributions in the multi-dimensional reversible diffusion model (1.24) are:

1. the non-availability in the multi-dimensional setting of the powerful diffusion local times techniques;

2. the more involved dependence in the reversible setting of the log-likelihood $\ell_T(B)$ on B, resulting in substantial difficulty in employing the Hellinger testing theory for contraction rates in diffusion models [232, 173].

More discussion can be found in Section 4.1. In Chapter 4, we address these issues using ideas developed in [103] for density estimation, and directly construct plug-in tests with exponentially decaying type-I and type-II error probabilities based on preliminary estimators with suitable concentration inequalities (see Lemma 4.2). In particular, we exploit the relationship (1.25) between the invariant measure μ_B and the potential B to construct, starting from an estimator $\hat{\mu}_T$ of μ_B (of wavelet projection type, cfr. (4.21)), the estimator $\frac{1}{2}\nabla \log \hat{\mu}_T$ of the drift ∇B, for which we derive a stability estimate of the form (cfr. (4.23))

$$\left\| \frac{1}{2}\nabla \log \hat{\mu}_T - \nabla B \right\|_{L^p} \lesssim \|\hat{\mu}_T - \mu_B\|_{W^{1,p}}, \tag{1.28}$$

holding with high probability under the law of the data $X^T \sim P_B^T$, uniformly for B in sets of bounded Hölder norm. The required concentration inequality for $\|\frac{1}{2}\nabla \log \hat{\mu}_T - \nabla B\|_{L^p}$ then follows from a key concentration inequality for $\|\hat{\mu}_T - \mu_B\|_{W^{1,p}}$, which we derive by a duality argument (cfr. the proof of Lemma 4.3), and a refinement of the bound proved in [173] for the suprema of the empirical process (see Proposition 4.1).

In following the program outlined above, the nonlinearity in the relationship (1.25) between the invariant measure μ_B and the potential B gives rise to certain nonlinear effects in the proofs similar to those encountered in the inverse problem setting described in Section 1.2.2, including the locality of the stability estimate (1.28). In Chapter 4 we address such issues using tools from the nonlinear inverse problems literature [167, 2, 177, 168], introducing a T-dependent scaling of the Gaussian and p-exponential priors analogous the scaling used in Chapter 3, which enforces a bound on the Hölder norm of the posterior draws (cfr. Lemma 4.5).

1.3 Further discussion and related literature

1.3.1 Computational guarantees for MCMC algorithms

The focus of the present thesis is on deriving statistical guarantees for Bayesian nonparametric procedures in statistical models arising from PDEs. As mentioned in the introduction, outside of certain conjugate settings such as those outlined in Sections 1.2.1 and 1.2.3, the implementation of such procedures rely in practice on numerical algorithms to extract information from the non-explicitly available posterior distributions. A fundamental question closely related to the investigation carried out in this thesis then arises as to whether, and by what algorithms, the numerical implementation of Bayesian methods can be achieved with a feasible computational cost in complex and high-dimensional statistical models.

Among the most prominent computational methods for Bayesian inference is (approximate) posterior sampling via MCMC algorithms; see, e.g., [200] for an introduction. In the last two decades, the design of sampling methods suited to the high- and infinite-dimensional framework has enjoyed considerable progresses, leading to the development of a wide range of methodologies; see [165, 203, 224, 29, 63, 66, 62, 26] and the many references therein. In this setting, two major classes of MCMC algorithms are Metropolis-

Hastings methods [28, 29, 120] (including the pCN algorithm [63] implemented for the simulations in Section 1.2.2), and gradient-based Langevin Monte Carlo methods [202, 9].

Recently, a line of theoretical work has emerged to provide computational guarantees on the performance and computational cost of such algorithms, which generally aims to quantify the number of required MCMC iterations in order to numerically evaluate quantities of interest up to a desired precision level. For the (unadjusted) Langevin algorithm, the first results in this direction were derived by Dalalyan [69], and by Durmus and Moulines [81, 82] for posterior distributions with (strongly) log-concave densities, providing non-asymptotic upper bounds on the mixing times that scale polynomially with respect to sample size and dimension. Related results are in [44, 85, 84] for Metropolis-adjusted Langevin algorithms and in [154, 43, 59] for Hamiltonian Monte Carlo. Extensions under geometric conditions close to log-concavity are studied in [60, 158, 240].

While the statistical linear inverse problems studied in Chapter 2 and the stochastic diffusion model considered in Chapter 4 are characterised by log-concave likelihoods, the posterior densities arising in nonlinear inverse problems are in general non log-concave (cfr. Sections 1.2.2 and 3.1). In particular, the nonlinearity of the forward map may cause the presence of multiple posterior modes, potentially leading to mixing times that depend exponentially on the sample size in the statistically relevant large sample scenario, in which, due to posterior contraction, the posterior distributions tend to become increasingly spiked around their modes (cfr. [86], and the discussion in [177]). In such non log-concave framework, the aforementioned work by Hairer, Stuart and Vollmer [113] first derived dimension-independent spectral gaps for the pCN algorithm, under uniformly boundedness and local Lipschitz assumptions compatible with nonlinear inverse problems in PDEs (including the elliptic example considered in Chapter 3), which imply dimension-free computational guarantees (however, with an implicit dependence on sample size, potentially exponential for spiked multimodal posterior densities, cfr. [177]). The computational cost of MCMC methods in the large sample scenario for high-dimensional and non log-concave models is investigated in [22] under a Bernstein–von Mises assumption. Finally, Nickl and Wang [177] very recently obtained non-asymptotic computational guarantees scaling polynomially in sample size and dimension for a Langevin-type algorithm in an elliptic nonlinear inverse problem for the Schrödinger equation, based on a general theory for nonlinear inverse regression models satisfying local regularity and curvature assumptions. Such results have been extended in [40] to a general class of PDEs (in particular, to transport equations for non-Abelian X-ray transforms).

1.3.2 Classical and statistical approaches to inverse problems

The study of inverse problems arising from PDEs has a long history in applied mathematics and numerical analysis. The typical general mathematical formulation of an inverse problem involves a *forward operator* $G : \mathcal{W}_1 \to \mathcal{W}_2$ between two Banach spaces $\mathcal{W}_1$, $\mathcal{W}_2$, and the recovery of an unknown element $f \in \mathcal{W}_1$ from data

$$Y = G(f) + W, \tag{1.29}$$

where W represents the observational noise. An intrinsic characteristic of a wide class of inverse problems is that they are *ill-posed* (in the sense of Hadamard [112]): the forward map G is in general not continuously invertible on its range, so that even in the absence of noise (i.e., $W = 0$), equation (1.29) might not have a solution, or it might have multiple solutions or one that depend sensitively on the data Y.

In the classical (deterministic) approach to inverse problems, the observational noise W is assumed to be a deterministic unknown element of $\mathcal{W}_2$ with bounded norm, and the convergence of recovery algorithms is then evaluated according to a worst-case error analysis. In particular, in the *regularisation* approach to inverse problems, originated in the work of Tikhonov [225, 226] and Philips [184], ill-posed problems are addressed by constructing a regularised version thereof, typically defined as a penalised minimisation problem of the form

$$\min_{f \in \mathcal{W}_1} \left\{ \|Y - G(f)\|^2_{\mathcal{W}_2} + \lambda R(f) \right\}, \qquad \lambda > 0, \tag{1.30}$$

where $R : \mathcal{W}_1 \to [0, \infty]$ is a suitable penalty functional that enforces well-posedness and imposes particular properties on the solution such as smoothness (e.g., [163]). The literature on the subject is vast, and for an extensive treatment we refer to the monographs [88, 135], where many additional references can be found. We further mention that iterative methods (such as *Landweber iteration* or *ν-method*), are often used in practice to solve the regularised minimisation problem (1.30), in particular in the case of nonlinear forward maps G, where the objective functional is in general non-convex; see [114, 192, 129] and the monograph [128].

Explicit statistical models for the noise, on the other hand, have been systematically studied only more recently [94, 179, 150, 178, 92, 90, 127]. Concretely, given data Y from model (1.6), the primary goal is then to construct an estimator of the unknown parameter of interest f. Furthermore, as opposed to the deterministic framework, the statistical setting naturally allows to approach quantification of the uncertainty in the reconstruction and the testing of hypotheses.

The nonparametric estimation theory is relatively well developed for *linear* inverse problems, wherein the forward map G is assumed to be a linear and bounded operator. Broadly speaking, the (minimax) optimal convergence rates achievable by estimators in linear inverse problems are determined by the degree of ill-posedness, in turn connected to the smoothing properties of the forward map. A smoothing operator effectively shrinks the distance between different parameters, rendering more difficult to distinguish between them based on observations of their image under the forward map: thus, the greater the smoothing effect of G, the harder the estimation problem. A precise measure of the degree of ill-posedness can be formulated in terms of the action of G upon different smoothness scales (e.g., the Sobolev or Besov scale) of functions [159, 161, 111] or via the decay of the singular values of G (e.g., [55, 195]). According to the latter principle, two classes of linear inverse problems are identified as:

1. *mildly* ill-posed problems, characterised by an algebraic decay of the singular values of the forward map and by algebraic minimax rates of convergence. Concrete examples of mildly ill-posed problems include the Radon transform in computerised tomography [169, 126, 195], elliptic pseudo-differential operators (e.g., [6, 133]), and the recovery of the source function in boundary value problems (e.g., [111]). The latter two examples are studied in Chapter 2;

2. *severely* ill-posed problems, characterised by an exponential decay of the singular values and by logarithmic minimax rates of convergence over smoothness classes. Examples include the inverse heat equation considered in Chapter 2, Chaucy problems for the Laplace and Helmotz equations (e.g., [248]) and certain deconvolution problems with smooth kernels (e.g., [195]).

A number of statistical procedures have been shown to attain such minimax rates of estimation, including Tikhonov regularised estimators based on penalised least-squares formulations analogous to (1.30); see, e.g., [178, 65, 162]. A large body of work is dedicated to develop and study methods based on knowledge of the singular value decomposition of G, such as spectral cut-off estimators [126, 159, 87, 134, 56]. Further important classes of methods are wavelet-based estimators [78, 3, 125] and iterative methods (e.g., [34, 37]). The paper [55] provides an overview on statistical estimation in linear inverse problems.

Recently, statistical inference methodologies have also been studied in the linear setting, including the construction of confidence bands and tests in density deconvolution problems [32, 35, 156, 212], in inverse regression models with convolution-type forward operators [31, 189], and for the inverse heat equation [35]. Further results are in [74, 246, 14], where more references can be found.

The statistical literature on *nonlinear* inverse problem is, on the other hand, far less developed. Extensions of the convergence rates theory of Tikhonov-type regularisers are provided in certain nonlinear setting in [180, 33, 117, 155, 243]. The results in the recent paper [176] are related to the investigation in Chapter 3, and provide (using variational arguments) convergence rates for the maximum a posteriori estimators associated to the scaled Gaussian priors employed in Section 3.3.1. The convergence of iterative methods in statistical nonlinear inverse problems is studied in [21, 118, 242].

The Bayesian approach to statistical inverse problems was described in the introduction to this chapter. During the last decade, the frequentist analysis of Bayesian nonparametric procedures in inverse problems has received a great amount of interest. In particular, in the linear setting, minimax optimal posterior contraction rates have been derived in a number of papers under various assumptions on the forward map and the priors employed (e.g., [138, 139, 195, 93, 137] and references therein). Statistical guarantees for Bayesian uncertainty quantification are provided in a diagonal SVD-based setting in [138], and more recently in [166] for X-ray transforms using standard Gaussian process priors. Further results are derived in Chapter 2 in a general setting and applied to three concrete examples covering both the mildly and severely ill-posed case.

Regarding nonlinear inverse problems, an early result on posterior consistency was proved in [241] for the elliptic inverse problem studied in Chapter 3, assuming uniformly bounded priors. More recently, employing similar bounded random wavelet series priors, [175] derived posterior contraction rates for estimating the drift and diffusion coefficients from discretely observed one-dimensional diffusions, while posterior contraction rates and Bernstein–von Mises theorems, implying coverage properties of posterior credible sets, were proved in [171] for an inverse problem for the Schrödinger equation and in [175] for compound Poisson processes.

A very significant recent advance has been the extension of the theory to *unbounded* priors, and in particular to Gaussian priors for which *ad hoc* sampling algorithms are available. Posterior consistency results for Gaussian priors have been established in nonlinear inverse problems arising from PDEs of elliptic [2, 177], parabolic [131], and transport type [167, 168, 39]. In Chapters 3 and 4, further results are derived for the problem of estimating the scalar diffusion coefficient in an elliptic PDE, and for drift estimation in reversible diffusion models. Finally, (semi-parametric) Bernstein–von Mises theorems and statistical guarantees for Bayesian uncertainty quantification have been proved in [168] in a class of nonlinear inverse problems that includes the Schrödinger equation as well as (non-Abelian) X-ray transforms. Some negative results about the possibility of a Bernstein–von Mises theorem in the elliptic inverse problem considered in Chapter 3 are discussed in [172].

1.3.3 Background reading

We refer to [101] and to Section 7.3 in [104] for basics on the nonparametric Bayesian approach to statistical inference, as well as for the general theory of the frequentist analysis of posterior distributions. Chapter 11 in [101] and Chapter 2 in [104] also cover fundamental results on Gaussian processes used throughout the thesis. The properties of p-exponential priors used in Chapter 4 are largely due to [5].

For the general minimax theory of estimation, see [230], while for the semiparametric theory of efficiency relevant to Chapter 2, see [234].

We refer to [122] for an overview on the field of inverse problems arising from PDEs. The nonparametric Bayesian approach to inverse problems is described in [220, 73]. For an introduction to PDEs, see [89, 153]. Finally, we refer to [19, 15] for the theory of diffusion processes and SDEs.

Chapter 2

Bernstein–von Mises theorems and uncertainty quantification for linear inverse problems

We consider the statistical inverse problem of recovering an unknown function f from a linear indirect measurement corrupted by additive Gaussian white noise. We employ a nonparametric Bayesian approach with standard Gaussian priors, for which the posterior-based reconstruction of f corresponds to a Tikhonov regulariser $\bar{f}$ with a reproducing kernel Hilbert space norm penalty. We prove a semiparametric Bernstein–von Mises theorem for a large collection of linear functionals of f, implying that semiparametric posterior estimation and uncertainty quantification are valid and optimal from a frequentist point of view. The result is applied to study three concrete examples that cover both the mildly and severely ill-posed cases: specifically, an elliptic inverse problem, an elliptic boundary value problem and the recovery of the initial condition of the heat equation. For the elliptic boundary value problem, we also obtain a nonparametric version of the theorem that entails the convergence of the posterior distribution to a prior-independent infinite-dimensional Gaussian probability measure with minimal covariance. As a consequence, it follows that the Tikhonov regulariser $\bar{f}$ is an efficient estimator of f, and we derive frequentist guarantees for certain credible balls centred at $\bar{f}$.

2.1 Introduction

Inverse problems arise in a variety of scientific disciplines, where the relationship between the quantity of interest and the data collected in an experiment is determined by the physics of the underlying system and can be mathematically modelled. Real world measurements are always discrete and carry statistical noise, which is often most naturally modelled by independent Gaussian random variables. The observation scheme then gives rise to an inverse regression model of the form

$$Y_i = (Gf)_i + W_i, \qquad i = 1, \dots, N, \quad W_i \overset{\text{iid}}{\sim} N(0,1),$$

where G describes the forward process and $(Gf)_i$ is a discrete observation of the transformed signal.

The formulation and analysis of the inverse problem is often best done by working with an analogous continuous model. This guarantees, among other things, discretisation invariance that allows to switch consistently between different discretisations [73, 146, 147, 220]. In this chapter we consider the case where the forward operator $G : \mathcal{W}_1 \to \mathcal{W}_2$ is linear between separable Hilbert spaces $\mathcal{W}_1$ and $\mathcal{W}_2$, and assume the continuous equivalent model (in the sense of [48, 197])

$$Y^\varepsilon = Gf + \varepsilon \mathbb{W}, \qquad \varepsilon > 0, \tag{2.1}$$

where $\mathbb{W}$ is a Gaussian white noise process indexed by $\mathcal{W}_2$. Note that while $\mathbb{W}$ can be defined by its actions on $\mathcal{W}_2$, it almost surely does not take values on it, making the noise in (2.1) 'rougher' than the forward signal Gf.

We adopt the Bayesian approach to inverse problems [220, 73] and study the performance of nonparametric procedures based on centred Gaussian priors Π for f. We are interested in what kind of objective guarantees can be achieved for Bayesian inference based on standard Gaussian priors used in practice. The specification of these priors does not require additional - and often unavailable - information on the forward map G, such as its singular value decomposition (SVD). The solution to the statistical inverse problem is the conditional distribution of f given Y^ε, whose mean or mode can be used as point estimators. The main appeal of the method is, however, that it automatically delivers quantification of uncertainty in the reconstruction, obtained through credible sets, i.e., regions of the parameter space with specified high posterior probability. In many applications this method can be efficiently implemented using modern (possibly infinite-dimensional) MCMC algorithms that allow fast sampling from the posterior distribution [26, 127].

Our goal is to investigate whether the methodology delivers correct, prior-independent and possibly optimal inference on the unknown parameter in the small noise limit. These questions can be addressed under the frequentist assumption that Y^ε is in reality generated through model (2.1) from a fixed true signal f_0 (instead of f being randomly drawn from Π). We then study the asymptotic concentration of the posterior distribution around f_0 as $\varepsilon \to 0$. The frequentist analysis of nonparametric Bayesian procedures for inverse problems has received increasing interest in the last decade, and several contributions in the linear setting have established consistency results and derived posterior contraction rates; see [6, 7, 133, 138, 139, 137, 136, 195, 241] among others. We also mention [175, 176, 171] for results for nonlinear inverse problems.

However, determining whether the resulting uncertainty quantification is objectively valid requires finer analysis of the posterior distribution. The central question is: do credible sets have the correct frequentist coverage in the small noise limit? That is, do we have, for some set $C_\varepsilon = C_\varepsilon(Y^\varepsilon)$,

$$\Pi\left(f \in C_\varepsilon | Y^\varepsilon\right) \approx 1 - \alpha \qquad \Leftrightarrow \qquad P_{f_0}^{Y^\varepsilon}\left(f_0 \in C_\varepsilon\right) \approx 1 - \alpha, \tag{2.2}$$

with small $\alpha \in (0, 1)$ as $\varepsilon \to 0$? The importance of the above questions is not restricted just to the Bayesian paradigm. In linear Bayesian inverse problems with Gaussian priors the conditional mean estimator can be shown to coincide with a Tikhonov regulariser $\bar{f}_\varepsilon$ arising from a reproducing kernel Hilbert space norm penalty, see [71, 115]. Thus, if (2.2) holds for a credible set C_ε centred at the posterior mean, we can use C_ε as an (asymptotic) frequentist confidence region based on the Tikhonov regulariser $\bar{f}_\varepsilon$.

Obtaining optimal contraction rates is not enough to answer the above question even in the parametric case. For classical finite-dimensional models the Bernstein–von Mises (BvM) theorem states that, under mild conditions, the posterior distribution is approximated in total variation distance by a normal distribution, centred at the maximum likelihood estimator and with minimal asymptotic variance. This implies that credible sets are asymptotically valid and optimal confidence regions; see, e.g., [234, Chapter 10]. Understanding the frequentist properties of nonparametric credible sets presents a more delicate matter. It was observed by [65], and later in [95], that the theorem may fail to hold even in a simple nonparametric regression model, for which credible balls in L^2 can be shown to have null asymptotic coverage.

One way of tackling the problem is to start by examining the limit behaviour of the one-dimensional marginals $\langle f, \psi\rangle_{\mathcal{W}_1} | Y^\varepsilon$ instead of the full posterior. This semiparametric approach was introduced for a direct problem where $G = I$ in [52, 53], where it is shown that (approximately) in the small noise limit

$$\langle f, \psi\rangle_{\mathcal{W}_1} | Y^\varepsilon \sim N\left(\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}, \varepsilon^2 I^{-1}(\psi)\right), \tag{2.3}$$

for a large collection of test functions ψ. Above $I^{-1}(\psi)$ is the asymptotic minimal variance. Note that nonparametric BvM theorems cannot hold in total variation distance like the classical BvM theorem. Instead one has to employ some metric for weak convergence of probability measures. Utilising a Wasserstein-type metric [52, 53] achieve weak convergence of the posterior distribution to a prior-independent infinite-dimensional Gaussian distribution on a large enough function space. More recently similar techniques were used in the inverse setting [166], for the linear X-ray transform problem, obtaining a semiparametric BvM theorem relative to *smooth* functionals of the unknown, while [171] proved a nonparametric result for a nonlinear problem arising in partial differential equations. See also [175, 173] for further related results. Positive results have also been obtained in [138, 149, 221], for priors defined on the SVD basis of the forward operator.

The first contribution of the present chapter is to extend the semiparametric BvM theorem in [166] for linear inverse problems of the form (2.1), formulating a general framework that translates the C^∞ smoothness assumption on the test functions ψ into a general 'source-type condition' that depends on the properties of the forward map and of the chosen prior (cfr. Theorem 2.1). As a consequence, we then deduce that the plug-in Tikhonov regularisers $\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}$ are consistent and efficient estimators for $\langle f_0, \psi\rangle_{\mathcal{W}_1}$, and that credible intervals centred at such estimators have asymptotically correct coverage and optimal width.

We subsequently employ the general theory to study three concrete examples of interest, where properties of the forward map can be exploited to check the condition for the semiparametric BvM theorem to hold. Specifically, we consider elliptic inverse problems on closed manifolds (Example 2.1), an inverse problem arising from an elliptic boundary value problem (Example 2.1), and the severely ill-posed problem of finding the initial source of the heat equation (Example 2.2). Similar examples have been considered, e.g, respectively in [6, 133], in [111] and in [7, 139, 195].

Our second contribution is a refinement of the result obtained for the elliptic boundary value problem, for which we further relax the assumption on the test functions to a minimal smoothness requirement that only depends on the degree of ill-posedness (cfr. Theorem 2.2). Adapting the program laid out in [171] to the problem at hand, we show that the asymptotic approximation of the marginal distributions holds uniformly across a suitable collection of test functions, leading to the formulation of a nonparametric BvM theorem. This entails the convergence of the posterior distribution to a limiting Gaussian probability measure with minimal covariance in suitable function spaces (cfr. Theorem 2.3), and implies frequentist guarantees for the reconstruction and uncertainty quantification relative to the entire function f.

This chapter is organised as follows: we introduce the general setting in Section 2.2, and state the semiparametric BvM theorem for linear functionals of the unknown in Section 2.2.1. In Section 2.2.2 we derive the asymptotic normality of $\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}$ and the coverage properties of credible intervals. Section 2.3 is dedicated to the examples. In Section 2.4 we refine the general theorem to achieve optimal semiparametric result for the elliptic boundary value problem, and obtain the nonparametric BvM theorem. The proofs are postponed to Section 3.5 and Appendix 2.C. Finally, Appendix 2.A and 2.B provide some of the background facts used throughout the chapter.

Regarding the notation, we will write $\lesssim$ and $\gtrsim$ for inequalities holding (possibly asymptotically) up to a universal constant. Also, for two real sequences (a_n) and (b_n), we say that $a_n \simeq b_n$ if both $a_n \lesssim b_n$ and $b_n \lesssim a_n$ for all n (large enough). Below, we will denote by $\xrightarrow{d}$ the usual convergence in distribution of a sequence of random variables. The notation $\mu_\varepsilon \xrightarrow{\mathcal{L}} \mu$ will be used for the weak convergence of random laws in probability, meaning that for any metric d for weak convergence of probability measures the real random variables $d(\mu_\varepsilon, \mu)$ converge to zero in probability (see [79] for definitions).

2.2 Results for general inverse problems

2.2.1 A semiparametric Bernstein–von Mises theorem

We start by considering general linear inverse problems with minimal assumptions on the forward operator. We are interested in the nonparametric statistical inverse problem of recovering an unknown function f from a noisy measurement of the form

$$Y^\varepsilon = Gf + \varepsilon \mathbb{W}, \qquad \varepsilon > 0. \tag{2.4}$$

The forward operator $G : \mathcal{W}_1 \to \mathcal{W}_2$ is assumed to be linear, bounded and injective between separable Hilbert spaces $\mathcal{W}_1$ and $\mathcal{W}_2$ of real valued functions (that can be defined on different sets). The operator G has a well defined adjoint $G^* : \mathcal{W}_2 \to \mathcal{W}_1$ for which $\langle Gf, g\rangle_{\mathcal{W}_2} = \langle f, G^* g\rangle_{\mathcal{W}_1}$, for all $f \in \mathcal{W}_1$ and $g \in \mathcal{W}_2$. In order to deal with possibly non-smooth unknowns, we define a third space $\overline{\mathcal{W}}$ as a separable Hilbert space for which $G : \overline{\mathcal{W}} \to \mathcal{W}_2$ is continuous and $\mathcal{W}_1 \subset \overline{\mathcal{W}}$ is dense in the norm of $\overline{\mathcal{W}}$. In particular, there exists $c > 0$ such that

$$\|Gf\|_{\mathcal{W}_2} \le c\|f\|_{\overline{\mathcal{W}}}, \qquad \forall f \in \overline{\mathcal{W}}. \tag{2.5}$$

The above can be thought of as a smoothing property of G, in that the more smoothing the forward operator is, the larger the space $\overline{\mathcal{W}}$ can be chosen. For example, if we assume that $G : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is an elliptic (pseudo-)differential operator smoothing of order t, we may choose $\overline{\mathcal{W}} = H^{-t}(\mathbb{R}^d)$, see Section 2.3.1. Since our general semiparametric result only requires that $f \in \overline{\mathcal{W}}$, this allows dealing with possibly non-smooth unknowns $f \notin L^2(\mathbb{R}^d)$ as long as $t > 0$ (cfr. Example 2.1 and the following discussion). Note that we can always make the trivial choice $\overline{\mathcal{W}} = \mathcal{W}_1$.

The measurement noise $\mathbb{W}$ is taken to be a centred Gaussian white noise process $(\mathbb{W}(\varphi) : \varphi \in \mathcal{W}_2)$ defined on some probability space $(\Omega, \Sigma, \Pr)$, with covariance $E\left(\mathbb{W}(\varphi)\mathbb{W}(\psi)\right) = \langle \varphi, \psi\rangle_{\mathcal{W}_2}$. Below we often write $\langle \mathbb{W}, \varphi\rangle_{\mathcal{W}_2}$ for the random variable $\mathbb{W}(\varphi)$. The noise amplitude is modelled by $\varepsilon > 0$. Observing data Y^ε then means that we observe a realisation of the Gaussian process $(Y^\varepsilon(\varphi) = \langle Y^\varepsilon, \varphi\rangle_{\mathcal{W}_2} : \varphi \in \mathcal{W}_2)$ with marginal distributions $\langle Y^\varepsilon, \varphi\rangle_{\mathcal{W}_2} \sim N(\langle Gf, \varphi\rangle_{\mathcal{W}_2}, \varepsilon^2\|\varphi\|^2_{\mathcal{W}_2})$.

For a fixed $f \in \overline{\mathcal{W}}$, let $P_f^{Y^\varepsilon} = \mathcal{L}(Y^\varepsilon)$ be the (cylindrically defined) law of Y^ε. Arguing as in Section 7.4 in [171] (see also [67, Theorem 2.23]), we can use the law $P_0^{Y^\varepsilon}$ of $\varepsilon\mathbb{W}$ as a common dominating measure, and apply the Cameron–Martin theorem [38, Corollary 2.4.3.] to define the log-likelihood function as

$$\ell_\varepsilon(f) = \log p_f(Y^\varepsilon) = \log \frac{dP_f^{Y^\varepsilon}}{dP_0^{Y^\varepsilon}}(Y^\varepsilon) = \frac{1}{\varepsilon^2}\langle Y^\varepsilon, Gf\rangle_{\mathcal{W}_2} - \frac{1}{2\varepsilon^2}\|Gf\|^2_{\mathcal{W}_2}. \tag{2.6}$$

We consider a Bayesian approach to the problem, assigning f a centred Gaussian prior Π on $\overline{\mathcal{W}}$. The reproducing kernel Hilbert space (RKHS) or Cameron-Martin space of Π is denoted by $\mathcal{H}$. Provided that $\ell_\varepsilon(f)$ can be taken to be jointly measurable, we can then use Bayes' theorem to deduce that the posterior distribution of $f|Y^\varepsilon$ arising from observation (2.4) can be written as

$$\Pi(B|Y^\varepsilon) = \frac{\int_B p_f(Y^\varepsilon)d\Pi(f)}{\int_{\overline{\mathcal{W}}} p'_f(Y^\varepsilon)d\Pi(f')}, \qquad B \in \mathcal{B}_{\overline{\mathcal{W}}} \text{ a Borel set in } \overline{\mathcal{W}}. \tag{2.7}$$

In the following we will study the asymptotic behaviour of $\Pi(\cdot|Y^\varepsilon)$ in the small noise limit $\varepsilon \to 0$, under the assumption that the measurement is generated from a fixed true unknown $f_0 \in \overline{\mathcal{W}}$. In order to do so, we assume that the prior satisfies a standard concentration function condition.

Condition 2.1. *Let* Π *be a centred Gaussian Borel probability measure on the separable Hilbert space* $\overline{\mathcal{W}}$ *for which* (2.5) *holds, and let* $\mathcal{H}$ *be the RKHS of* Π. *Define the* *concentration function of* Π *for a fixed* $f_0 \in \overline{\mathcal{W}}$ *as*

$$\phi_{\Pi,f_0}(\delta) = \inf_{g\in\mathcal{H}\,:\,\|g-f_0\|_{\overline{\mathcal{W}}}\leq\delta} \frac{\|g\|_{\mathcal{H}}^2}{2} - \log\Pi(f : \|f\|_{\overline{\mathcal{W}}} \leq \delta), \qquad \delta > 0. \tag{2.8}$$

Given Π *and* $f_0 \in \overline{\mathcal{W}}$*, assume that there exists a sequence* $\delta_\varepsilon \to 0$*, with* $\delta_\varepsilon/\varepsilon \to \infty$ *as* $\varepsilon \to 0$*, such that*

$$\phi_{\Pi,f_0}\left(\frac{\delta_\varepsilon}{2c}\right) \leq \left(\frac{\delta_\varepsilon}{\varepsilon}\right)^2. \tag{2.9}$$

The above condition characterises the asymptotics of the small ball probabilities, and guarantees that the prior puts sufficient mass around the truth: in particular $\Pi(f : \|f - f_0\|_{\overline{\mathcal{W}}} \leq \delta_\varepsilon) > e^{-\frac{1}{2}(\delta_\varepsilon/\varepsilon)^2}$ as $\varepsilon \to 0$ (cfr. the proof of Lemma 2.C.3). Analogous conditions underpin many results in Bayesian asymptotics, and play a fundamental role in the theory of posterior contraction rates, see e.g. [101, 104, 235]. The concentration functions of Gaussian priors are generally well understood, and explicit forms for the sequences δ_ε can readily be computed for many standard choices of practical interest, such as the commonly used Matérn process priors (see Section 2.3).

Next we formulate a semiparametric Bernstein–von Mises theorem in the above general linear inverse problems setting.

Theorem 2.1. *Let* $P_{f_0}^{Y^\varepsilon}$ *be the law of* Y^ε *generated by* (2.4) *with* $f = f_0 \in \overline{\mathcal{W}}$ *, where* $\overline{\mathcal{W}}$ *is a separable Hilbert space for which* (2.5) *holds. We assume a centred Gaussian prior* Π *that satisfies Condition 2.1 for a fixed* $f_0 \in \overline{\mathcal{W}}$ *and denote its RKHS by* $\mathcal{H}$*. Consider a test function* $\psi \in \mathcal{W}_1$ *such that* $|\langle\psi, \varphi\rangle_{\mathcal{W}_1}| \lesssim \|\varphi\|_{\overline{\mathcal{W}}}$*, for all* $\varphi \in \mathcal{W}_1$*, and suppose that* $\psi = -G^*G\tilde{\psi}$ *for some* $\tilde{\psi} \in \mathcal{H}$*. Then,*

$$\mathcal{L}\left(\varepsilon^{-1}\left(\langle f, \psi\rangle_{\mathcal{W}_1} - \widehat{\Psi}\right)\Big|Y^\varepsilon\right) \xrightarrow{\mathcal{L}} N(0, \|G\tilde{\psi}\|_{\mathcal{W}_2}^2) \tag{2.10}$$

in $P_{f_0}^{Y}$*-probability as* $\varepsilon \to 0$*, where*

$$\widehat{\Psi} = \langle f_0, \psi\rangle_{\mathcal{W}_1} - \varepsilon\langle G\tilde{\psi}, \mathbb{W}\rangle_{\mathcal{W}_2}. \tag{2.11}$$

The next corollary states that we can replace the centring $\widehat{\Psi}$ by a linear functional of the conditional mean. This implies that the posterior distribution of the functionals are asymptotically approximated by a normal distribution centred at the conditional mean and with asymptotic minimal variance (see Remark 2.1 below). The proof of Corollary

2.1 can be adapted from the proof of Theorem 2.7 in [166] and is therefore omitted (see also Step V in Appendix 2.C at the end of the chapter).

Corollary 2.1. *Let $\bar{f}_\varepsilon = E^\Pi[f|Y^\varepsilon]$ be the mean of the posterior $\Pi(\cdot|Y^\varepsilon)$. Then, for every $\psi \in \mathcal{W}_1$ satisfying the conditions in Theorem 2.1, we have*

$$\frac{1}{\varepsilon}\left(\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1} - \hat{\Psi}\right) \to 0, \tag{2.12}$$

in $P^Y_{f_0}$-probability as $\varepsilon \to 0$. As a consequence, we can replace $\hat{\Psi}$ with $\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}$ in Theorem 2.1.

Note that, since $\mathcal{W}_1 \subset \overline{\mathcal{W}}$ is dense and $L_\psi(\cdot) = \langle \cdot, \psi\rangle_{\mathcal{W}_1}$ is assumed to be a bounded linear operator (and hence uniformly continuous), we can extend L_ψ continuously to $\overline{\mathcal{W}}$. The condition on the test functions requires that ψ is in the range of the 'Fisher information operator' G^*G acting upon the RKHS of Π. This can normally be translated into suitable smoothness assumptions on ψ, see Section 2.3 for examples. The requirement resembles certain source conditions often used in inverse problems [88, 157, 214]. The main conceptual difference is that instead of requiring extra smoothness for the unknown f_0 to attain convergence in a predefined space, we allow f_0 to be non-smooth and impose constraints on the test functions in order to achieve convergence.

2.2.2 Efficiency and uncertainty quantification for Tikhonov regularisers

Since the forward operator G is assumed to be linear, the posterior distribution $\Pi(\cdot|Y^\varepsilon)$ is Gaussian. It follows that the conditional mean $\bar{f}_\varepsilon = \bar{f}_\varepsilon(Y^\varepsilon) = E^\Pi[f|Y^\varepsilon]$ coincides with the maximum a posterior (MAP) estimator, and using Corollary 3.10 in [71] (under appropriate conditions on G) the latter can be seen to be a Tikhonov-type regulariser found by minimising the following Onsager-Machlup functional:

$$Q(f) = -\frac{1}{\varepsilon^2}\langle Y^\varepsilon, Gf\rangle_{\mathcal{W}_2} + \frac{1}{2\varepsilon^2}\|Gf\|^2_{\mathcal{W}_2} + \frac{1}{2}\|f\|^2_{\mathcal{H}}.$$

Using Theorem 2.1 and Corollary 2.1 we can derive the asymptotic distribution of the plug-in estimators $\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}$.

Remark 2.1 (Minimax optimality of the plug-in Tikhonov regulariser)**.** Corollary 2.1 implies that

$$\frac{1}{\varepsilon}\langle \bar{f}_\varepsilon - f_0, \psi\rangle_{\mathcal{W}_1} \xrightarrow{d} Z \sim N(0, \|G\tilde{\psi}\|^2_{\mathcal{W}_2}) \tag{2.13}$$

in $P^Y_{f_0}$-probability as $\varepsilon \to 0$. The above random variable Z identifies the asymptotic minimal variance (in the minimax sense) in estimating $\langle f_0, \psi\rangle_{\mathcal{W}_1}$ from model (2.4), in that

$$\liminf_{\varepsilon\to 0} \inf_T \sup_{f\in B_\varepsilon} \varepsilon^{-2} E_f^{Y^\varepsilon}(\langle f, \psi\rangle_{\mathcal{W}_1} - T)^2 \geq \|G\tilde{\psi}\|^2_{\mathcal{W}_2}, \tag{2.14}$$

the infimum being over all estimators $T = T(Y^\varepsilon, \psi)$ of $\langle f_0, \psi\rangle_{\mathcal{W}_1}$ based on observing Y^ε in (2.4) with $f = f_0$, and the supremum is taken over balls B_ε in $\overline{\mathcal{W}}$ centred at f_0 and with radius $\varepsilon > 0$; see Appendix 2.A.

We note that (2.13) implies the convergence of all moments (see Step V in Appendix 2.C). Consequently, for all $\psi \in \mathcal{W}_1$ fulfilling the conditions of Theorem 2.1, the plug-in Tikhonov regulariser $\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}$ attains the lower bound in (2.14), and hence is an asymptotic minimax estimator of $\langle f_0, \psi\rangle_{\mathcal{W}_1}$.

Besides the question of efficiency, the most relevant consequence of Theorem 2.1 is that credible intervals built around the estimators $\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}$ are asymptotically valid frequentist confidence intervals with optimal diameter. Specifically, for ψ as above, consider a credible interval for $\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}$ of the form

$$C_\varepsilon = \left\{ x \in \mathbb{R} \ : \ |\langle \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1} - x| \leq R_\varepsilon \right\}, \tag{2.15}$$

with $R_\varepsilon = R_\varepsilon(\alpha, Y^\varepsilon)$ chosen so that

$$\Pi\left(f : \langle f, \psi\rangle_{\mathcal{W}_1} \in C_\varepsilon | Y^\varepsilon\right) = 1 - \alpha, \qquad \alpha \in (0,1).$$

Then it follows that C_ε has the correct asymptotic coverage and that its diameter shrinks at the optimal rate ε. The proof of the following corollary can be found in Appendix 2.C.

Corollary 2.2. *Let $\psi \in \mathcal{W}_1$ satisfy the conditions in Theorem 2.1, and let C_ε be as in (2.15). Then, as $\varepsilon \to 0$,*

$$P^{Y^\varepsilon}_{f_0}(\langle f_0, \psi\rangle_{\mathcal{W}_1} \in C_\varepsilon) \to 1 - \alpha$$

and

$$\varepsilon^{-1} R_\varepsilon \xrightarrow{P^Y_{f_0}} \Phi^{-1}(1-\alpha).$$

Here $\Phi(t) = \Pr(|Z| \leq t)$ with $Z \sim N(0, \|G\tilde{\psi}\|^2_{\mathcal{W}_2})$.

Note that although an explicit formulation of C_ε would require the computation of the quantiles of the posterior distribution of $\langle f, \psi\rangle_{\mathcal{W}_1} | Y^\varepsilon$, these type of credible intervals can often in practice be implemented by numerically approximating the radius R_ε with a posterior sampling method. See, e.g., [127], or Section 2.2 in [166]. The implementation of the algorithm is further discussed in 1.2.1, where some simulation results are shown.

For the inferential problem for elliptic partial differential equations studied in Section 2.3.2, Remark 2.4 below will extend the conclusions of Corollary 2.2 to entire credible balls in suitable function spaces centred at $\bar{f}_\varepsilon$.

2.3 Examples

In this section we consider examples of linear inverse problems fitting in the framework of Section 2.2, studying the conditions under which the semiparametric Bernstein–von Mises phenomenon occurs in such instances. We first need to introduce some notation on Sobolev spaces (see [153, 164] for background).

The Sobolev space on $\mathbb{R}^d$ of order $s \in \mathbb{R}$ is defined as

$$H^s(\mathbb{R}^d) = \left\{ u \in \mathcal{S}'(\mathbb{R}^d) \ : \ (1 + |\cdot|^2)^{s/2} \mathcal{F} u \in L^2(\mathbb{R}^d) \right\}, \tag{2.16}$$

where $\mathcal{S}'(\mathbb{R}^d)$ is the space of tempered distributions on $\mathbb{R}^d$ and $\mathcal{F}$ is the Fourier transform. For $\mathcal{O} \subset \mathbb{R}^d$ a non-empty, open and bounded set with smooth boundary $\partial\mathcal{O}$ (a smooth domain), Sobolev spaces on $\mathcal{O}$ can be defined via the restriction operator $|_{\mathcal{O}}$ as

$$H^s(\mathcal{O}) = \left\{ u = U|_{\mathcal{O}}, \ U \in H^s(\mathbb{R}^d) \right\}, \qquad \|u\|_{H^s(\mathcal{O})} = \inf_{U \in H^s(\mathbb{R}^d) \ : \ U|_{\mathcal{O}} = u} \|U\|_{H^s(\mathbb{R}^d)}. \tag{2.17}$$

To correctly address issues relative to the behaviour of functions near $\partial\mathcal{O}$, we will need to consider certain subspaces of $H^s(\mathcal{O})$. We denote the set of functions in $H^s(\mathcal{O})$ that are compactly supported in $\mathcal{O}$ by $H^s_c(\mathcal{O})$, and for any fixed compact subset $K \subset \mathcal{O}$, we write $H^s_K(\mathcal{O}) := \{u \in H^s(\mathcal{O}) \ : \ \mathrm{supp}(u) \subseteq K\}$. Finally, for all $s > 1/2$, let $H^s_0(\mathcal{O})$ be the usual subspace of $H^s(\mathcal{O})$ of functions with null trace on $\partial\mathcal{O}$. Below we will often suppress the dependence on the underlying domain denoting $H^s = H^s(\mathcal{O})$.

2.3.1 Elliptic inverse problems

We start with a basic example to demonstrate how Theorem 2.1 can be applied when G is assumed to be a smoothing elliptic pseudo-differential operator and $\mathcal{O}$ a closed manifold (see [119, 216] for general theory on pseudo-differential operators). The previous definitions of Sobolev spaces can straightforwardly be adapted to this setting, see, e.g., [216, Chapter I.7]. The absence of a boundary and the properties of the forward map allow for a clean exposition of the results. In the Section 2.3.2 we will instead assume that $\mathcal{O}$ is a smooth domain in $\mathbb{R}^d$, and take G to be the solution operator associated with an elliptic boundary value problem. We then have to refine the results to take into account some subtleties of the behaviour of functions near the boundary.

Example 2.1. Let $\mathcal{O}$ be a closed d-dimensional manifold and $G : L^2(\mathcal{O}) \to L^2(\mathcal{O})$ an injective and elliptic pseudo-differential operator smoothing of order t, that is, $G : H^s(\mathcal{O}) \to H^{s+t}(\mathcal{O})$ with all $s \in \mathbb{R}$ [216, Section I.5.]. We can then choose $\overline{\mathcal{W}} = H^{-t}(\mathcal{O})$.

Let $P_{f_0}^{Y^\varepsilon}$ be the law of Y^ε generated by (2.4) with $f = f_0 \in H^\alpha(\mathcal{O})$, $\alpha > -t$. We assume a centred Gaussian prior Π with RKHS $\mathcal{H} = H^r(\mathcal{O})$, where $r \geq \max\{0, d_0 - t\}$ and $d_0 > d/2$. This guarantees that $f \in H^{r-d_0}(\mathcal{O}) \subset H^{-t}(\mathcal{O}) = \overline{\mathcal{W}}$ almost surely. For example, we can take $\Pi = N(0, C)$, where C is a self-adjoint, injective and elliptic covariance operator smoothing of order $2r$ [6, 133]. Another example is to assume Π to be the law of the Matérn process of smoothness $r - d/2$ (see Example 11.8 in [101] for details), namely the centred Gaussian process $\{M(x),\ x \in \mathcal{O}\}$ with covariance kernel

$$C(x, y) = \int_{\mathbb{R}^d} e^{-i\langle x-y,\xi\rangle_{\mathbb{R}^d}} \mu(d\xi), \qquad \mu(d\xi) = (1 + |\xi|^2)^{-r} d\xi.$$

Since G is elliptic and $\mathcal{O}$ is a closed manifold, G^*G has a well defined inverse $(G^*G)^{-1} : H^s(\mathcal{O}) \to H^{s-2t}(\mathcal{O})$, $s \in \mathbb{R}$, see e.g. [132]. We can then take $\psi \in H^{r+2t}(\mathcal{O})$, which guarantees $\tilde{\psi} = -(G^*G)^{-1}\psi \in H^r(\mathcal{O}) = \mathcal{H}$ and $|\langle \psi, \varphi\rangle_{L^2}| \leq C\|\varphi\|_{H^{-t}}$, for all $\varphi \in L^2(\mathcal{O})$.

Denote by $\bar{f}_\varepsilon = E^\Pi[f|Y^\varepsilon]$ the mean of the posterior distribution $\Pi(\cdot|Y^\varepsilon)$ arising from observing (2.4). Then, for all test functions $\psi \in H^{r+2t}(\mathcal{O})$, the following convergence occurs in $P_{f_0}^Y$-probability as $\varepsilon \to 0$

$$\mathcal{L}\left(\varepsilon^{-1}\langle f - \bar{f}_\varepsilon, \psi\rangle_{L^2} \middle| Y^\varepsilon\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \|G(G^*G)^{-1}\psi\|_{L^2}^2\right).$$

Note that if $t > \frac{d}{2} - 1$ we can allow unknowns of bounded variation $f_0 \in BV(\mathcal{O})$, since $BV(\mathcal{O}) \subset H^\alpha(\mathcal{O})$ when $\alpha \leq 1 - \frac{d}{2}$. Functions of bounded variation are widely used, e.g., in image analysis due to their ability to deal with discontinuities. One standard example is total variation denoising [51, 207].

Remark 2.2. Let Π and f_0 be as above. Then, as $\delta \to 0$

$$\phi_{\Pi, f_0}(\delta) \lesssim \delta^{-\frac{2\max\{0, r-\alpha\}}{t+\alpha}} + \delta^{-\frac{d}{r+t-d/2}}, \tag{2.18}$$

so that the concentration condition $\phi_{\Pi,f_0}(\delta_\varepsilon) \lesssim (\delta_\varepsilon/\varepsilon)^2$ is satisfied by taking

$$\delta_\varepsilon \simeq \max\left\{\varepsilon^{\frac{t+\alpha}{t+r}}, \varepsilon^{\frac{t+r-d/2}{t+r}}\right\}.$$

The proof of Remark 2.2 is omitted since it is a simplified version of the proof of Remark 2.3 where, $\mathcal{O}$ being a closed manifold, one does not need to address the technicalities arising at the boundary.

2.3.2 An elliptic boundary value problem

Let $\mathcal{O} \subset \mathbb{R}^d$ be a non-empty, open and bounded set with smooth boundary $\partial\mathcal{O}$. We consider the problem of recovering the unknown source $f \in L^2 = L^2(\mathcal{O})$ in the elliptic boundary value problem

$$\begin{cases} Lu = f, & x \in \mathcal{O}, \\ u(x) = 0, & x \in \partial\mathcal{O}, \end{cases} \tag{2.19}$$

from noisy observations of the solution u corrupted by additive Gaussian white noise in L^2. We take L to be the following partial differential operator in divergence form:

$$Lu = -\sum_{i,j=1}^{d} \frac{\partial}{\partial x_j}\left(a_{ij}\frac{\partial u}{\partial x_i}\right), \tag{2.20}$$

for known $a_{ij} \in C^\infty(\overline{\mathcal{O}})$, with $a_{ij} = a_{ji}$. The problem represents an 'elliptic counterpart' of the transport PDE arising in [166].

Assuming that L is uniformly elliptic (see Appendix 2.B), it follows that for each $f \in H^s$, $s \geq 0$, there exists a unique weak solution $L^{-1}f = u \in H_0^{s+2}$ to (2.19). In particular, $L^{-1} : H^s \to H_0^{s+2}$ defines a bounded isomorphism, self-adjoint with respect to $\langle\cdot,\cdot\rangle_{L^2}$, and for all $s \geq 0$ we also have the dual estimates

$$\|L^{-1}f\|_{(H^s)^*} = \sup_{u \in H^s \,:\, \|u\|_{H^s} \leq 1} |\langle L^{-1}f, u\rangle_{L^2}| \leq c_s \|f\|_{(H_0^{s+2})^*}, \qquad c_s > 0. \tag{2.21}$$

Rephrasing in the notation of Section 2.2 with $G = L^{-1}$, we consider the observation

$$Y^\varepsilon = L^{-1}f + \varepsilon\mathbb{W}, \qquad \varepsilon > 0, \tag{2.22}$$

where $\mathbb{W}$ is Gaussian white noise in L^2. For $\mathcal{W}_1 = \mathcal{W}_2 = L^2$, the dual estimate (2.21) implies that we can take $\overline{\mathcal{W}} = (H_0^2)^*$.

We assume that $f \sim \Pi$, where Π is a centred Gaussian Borel probability measure on L^2 with RKHS $\mathcal{H} = H^r$, for some $r > d/2$. For example, we can take Π to be the law of the Matérn process of smoothness $r - d/2$ introduced in the previous example.

For $f_0 \in H_c^\alpha$, with some $\alpha \geq 0$, we show that the semiparametric BvM phenomenon occurs under appropriate smoothness conditions on the test functions ψ. In particular, assuming that $\psi \in H_c^{r+4}$ automatically verifies the requirements of Theorem 2.1, since taking $\tilde{\psi} = -LL\psi$ implies $\tilde{\psi} \in \mathcal{H} = H^r$ and $\psi = -L^{-1}L^{-1}\tilde{\psi}$, as $\text{supp}(L\psi) \subseteq \text{supp}(\psi) \subsetneq \mathcal{O}$. The proof of the following proposition can be found in Section 2.5.2.

Proposition 2.1. *Let Π be a Gaussian Borel probability measure on $L^2(\mathcal{O})$ with RKHS $\mathcal{H} = H^r(\mathcal{O})$, $r > d/2$. Assume that $f_0 \in H_c^\alpha(\mathcal{O})$, $\alpha \geq 0$, and let $P_{f_0}^{Y^\varepsilon}$ be the law of Y^ε generated by (2.22) with $f = f_0$. Let $\bar{f}_\varepsilon = E^\Pi[f|Y^\varepsilon]$ be the mean of the posterior distribution $\Pi(\cdot|Y^\varepsilon)$ arising from observing (2.22). Then, for all $\psi \in H_c^{r+4}(\mathcal{O})$, we have*

$$\mathcal{L}(\varepsilon^{-1}\langle f - \bar{f}_\varepsilon, \psi\rangle_{L^2}|Y^\varepsilon) \xrightarrow{\mathcal{L}} N(0, \|L\psi\|_{L^2}^2) \tag{2.23}$$

in $P_{f_0}^Y$-probability as $\varepsilon \to 0$.

Remark 2.3. Let Π and f_0 be as above. In the proof of Proposition 2.1 we show that as $\delta \to 0$

$$\phi_{\Pi,f_0}(\delta) \lesssim \delta^{-\frac{2\max\{0,r-\alpha\}}{2+\alpha}} + \delta^{-\frac{d}{r+2-d/2}}, \tag{2.24}$$

so that the concentration condition $\phi_{\Pi,f_0}(\delta_\varepsilon) \lesssim (\delta_\varepsilon/\varepsilon)^2$ is satisfied by taking

$$\delta_\varepsilon \simeq \max\left\{\varepsilon^{\frac{2+\alpha}{2+r}}, \varepsilon^{\frac{2+r-d/2}{2+r}}\right\}. \tag{2.25}$$

2.3.3 Recovery of the initial condition of the heat equation

We will conclude this section by applying the general framework studied in Section 2.2 to the severely ill-posed problem of finding the initial source of the heat equation. Contraction rates for similar inverse problems have been studied in [7, 139, 195].

Example 2.2. Let $\mathcal{O} \subset \mathbb{R}^d$ be an open bounded set with C^∞ boundary $\partial\mathcal{O}$. We consider the boundary value problem for the heat equation

$$\begin{cases} \partial_t u - \Delta u = 0, & t \geq 0, \quad x \in \mathcal{O}, \\ u(x) = 0, & t \geq 0, \quad x \in \partial\mathcal{O}, \\ u(0, x) = f(x), & x \in \mathcal{O}. \end{cases}$$

The inverse problem is to recover the initial heat source $f \in L^2$ from a noisy observation of the solution u at time T, corrupted by additive Gaussian white noise on L^2. The solution to the boundary value problem is given by

$$u(T,x) = Gf(x) = \sum_{j=1}^{\infty} \langle f, \varphi_j \rangle_{L^2} e^{-\lambda_j T} \varphi_j(x), \qquad x \in \mathcal{O},$$

where $-\Delta \varphi_j = \lambda_j \varphi_j$, and $\{\varphi_j\}_{j=1}^{\infty}$ forms an orthonormal basis of L^2. If we order the eigenvalues to be increasing, that is, $\lambda_1 \leq \lambda_2 \leq \dots$, then Weyl's law yields that $\lambda_j \simeq j^{2/d}$ (e.g., [204, Theorem 8.16]). Thus, the singular values of the compact forward operator G decay exponentially to zero, meaning that the recovery of the initial condition of the heat equation is a severely ill-posed inverse problem.

Assume that $f \sim \Pi$, where Π is a centred Gaussian Borel probability measure on L^2 with RKHS $\mathcal{H} = H^r$, $r > d/2$. Let $\psi \in L^2$ be of the form

$$\psi = -G^* G \tilde{\psi} = -\sum_{j=1}^{\infty} \langle \tilde{\psi}, \varphi_j \rangle_{L^2} e^{-2\lambda_j T} \varphi_j, \tag{2.26}$$

for some $\tilde{\psi} \in H^r$. Then

$$\begin{aligned} |\langle \psi, \phi \rangle_{L^2}| &= \left| \left\langle \psi, \sum_{j=1}^{\infty} \langle \phi, \varphi_j \rangle_{L^2} \varphi_j \right\rangle_{L^2} \right| \\ &\leq \sum_{j=1}^{\infty} |\langle \phi, \varphi_j \rangle_{L^2}| \left| \left\langle \sum_{i=1}^{\infty} \langle \tilde{\psi}, \varphi_i \rangle_{L^2} e^{-2\lambda_i T} \varphi_i, \varphi_j \right\rangle_{L^2} \right| \\ &\leq \sum_{j=1}^{\infty} |\langle \phi, \varphi_j \rangle_{L^2}| |\langle \tilde{\psi}, \varphi_j \rangle_{L^2}| e^{-2\lambda_j T} \\ &\leq C \|\phi\|_{H^{-t}}, \end{aligned}$$

for all $t \geq 0$, verifying the condition of Theorem 2.1. Hence, for $f_0 \in L^2$ and ψ as above, we get the following convergence in $P_{f_0}^Y$-probability as $\varepsilon \to 0$

$$\mathcal{L}\left(\varepsilon^{-1} \langle f - \bar{f}_\varepsilon, \psi \rangle_{L^2} \middle| Y^\varepsilon \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \|G\tilde{\psi}\|_{L^2}^2\right).$$

Note that the contraction rate ε entailed by the semiparametric BvM theorem is a very strong requirement for severely ill-posed inverse problems, usually characterised by logarithmic rates even for smooth functionals [139]. To achieve the rate ε, we then need to assume the analytic-type condition (2.26) on the test function ψ, which reflects the natural condition of ψ being in the range of G^* which is necessary for efficient semiparametric estimation [234, Theorem 25.32].

2.4 A nonparametric Bernstein–von Mises theorem for elliptic boundary value problems

In this section we continue the investigation of the BvM phenomenon in the setting of the elliptic boundary value problem studied in Section 2.3.2. We develop Proposition 2.1 along two related directions: first, we extend the class of test functions ψ for which the convergence (2.23) occurs, identifying a natural lower limit for the smoothness of ψ that only depends on the level of ill-posedness of the inverse problem. Secondly, combining the result with the program laid out in [171], we derive a nonparametric BvM theorem that entails the weak convergence, in a suitable function space, of the centred and scaled posterior to a prior-independent infinite-dimensional Gaussian probability measure whose covariance function attains the information lower bound. From the latter result we then obtain frequentist guarantees for uncertainty quantification in the reconstruction of the entire function f.

We briefly recall that, for unknown $f \in L^2 = L^2(\mathcal{O})$, we consider observations $Y^\varepsilon = L^{-1}f + \varepsilon\mathbb{W}$, $\varepsilon > 0$, where L^{-1} is the solution map associated with the boundary value problem (2.19) (see Section 2.3.2 for details) and $\mathbb{W}$ is a Gaussian white noise in L^2. We assign f a centred Gaussian prior in L^2 with RKHS H^r, $r > d/2$, and assume that the observation Y^ε is generated from a fixed $f_0 \in H^\alpha_c$ with some $\alpha > 0$. For the results in this section we assume an undersmoothing prior. That is, we consider the case $r - d/2 \leq \alpha$. The proofs can be found in Section 3.5.

Theorem 2.2. *Let Π be a Gaussian Borel probability measure on $L^2(\mathcal{O})$ with RKHS $\mathcal{H} = H^r(\mathcal{O})$, $r > d/2$. Assume that $f_0 \in H^\alpha_c(\mathcal{O})$, $\alpha \geq r - d/2$, and let $P^{Y^\varepsilon}_{f_0}$ be the law of Y^ε generated by (2.22) with $f = f_0$. Let $\bar{f}_\varepsilon = E^\Pi[f|Y^\varepsilon]$ be the mean of the posterior distribution $\Pi(\cdot|Y^\varepsilon)$ arising from observing (2.22). Then, for all $\beta > 2 + d/2$, and any $\psi \in H^\beta_c(\mathcal{O})$, we have*

$$\mathcal{L}(\varepsilon^{-1}\langle f - \bar{f}_\varepsilon, \psi\rangle_{L^2}|Y^\varepsilon) \xrightarrow{\mathcal{L}} N(0, \|L\psi\|^2_{L^2}) \tag{2.27}$$

in $P^Y_{f_0}$-probability as $\varepsilon \to 0$.

Assuming that $\beta > 2 + d$, we will strengthen the above result to a nonparametric Bernstein–von Mises theorem in the dual spaces $(H^\beta_K)^*$, for any compact set $K \subset \mathcal{O}$. In particular, we note that the Gaussian laws in the right hand side of (2.27) identify the one-dimensional marginal distributions of a nonparametric Gaussian probability measure μ, induced via Kolmogorov's extension (see, e.g., [79, Section 12.1]) on the cylindrical σ-field of $\mathbb{R}^{H^\beta_K}$ by the centred Gaussian process

$$X = (X(\psi) : \psi \in H^\beta_K), \qquad E[X(\psi)X(\psi')] = \langle L\psi, L\psi' \rangle_{L^2}. \tag{2.28}$$

In fact, in the view of the efficiency considerations in Appendix 2.A, μ represents the 'canonical' asymptotic distribution for the problem of inferring f in model (2.22), as its covariance function is minimal in the information theoretic sense of Remark 2.1. In the following lemma we derive the values of β for which μ is a tight Borel probability measure on $(H^\beta_K)^*$, a necessary condition for any sequence of laws on such spaces to weakly converge to μ. The proof adapts the argument in the proof of Proposition 6 in [171], and is included in Appendix 2.C.

Lemma 2.1. *Fix any compact set $K \subset \mathcal{O}$. Let X be as in (2.28), and let μ be the law of X on the cylindrical σ-field of $\mathbb{R}^{H^\beta_K}$. Then,*

1. *for all $\beta > 2 + d/2$, μ is a tight Gaussian Borel probability measure on $(H^\beta_K)^*$;*
2. *for $\beta < 2 + d/2$, we have*

$$\mu\left(x : \|x\|_{(H^\beta_K)^*} < \infty\right) = 0;$$

3. *for $\beta = 2 + d/2$, μ is not tight on $(H^\beta_K)^*$.*

Similarly, the stochastic process obtained by collecting the random variables in the left hand side of (2.27),

$$X_\varepsilon = \left(\varepsilon^{-1}\langle f - \bar{f}_\varepsilon, \psi \rangle_{L^2} | Y^\varepsilon \ : \ \psi \in H^\beta_K\right), \qquad \varepsilon > 0, \tag{2.29}$$

can also be shown to induce a tight Borel probability measure on $(H^\beta_K)^*$ when $\beta > 2 + d/2$ (see Step IV in Section 2.5.3). We will interpret the law of X_ε as the nonparametric centred and scaled posterior distribution arising from observing (2.22), denoted by

$$\mathcal{L}(\varepsilon^{-1}(f - \bar{f}_\varepsilon)|Y^\varepsilon) = \mathcal{L}(X_\varepsilon). \tag{2.30}$$

Theorem 2.2 implies the convergence of the finite-dimensional distributions of the stochastic process X_ε to those of X (cfr. Lemma 2.7), and by showing that (2.27) holds *uniformly* across the set of test functions, we then deduce the weak convergence of the respective induced laws on $(H^\beta_K)^*$. As mentioned in the introduction, nonparametric BvM theorems cannot hold in total variation distance like the classical BvM theorem in finite dimensional Euclidean spaces. Instead we use a Wasserstein-type metric for weak convergence of probability measures. Recall that on a given complete separable metric space (S,ρ), the notion of weak convergence of sequences of Borel probability measures can be metrised by the bounded Lipschitz (BL) metric

$$d_S(\nu_1,\nu_2) = \sup_{F:S\to\mathbb{R},\ \|F\|_{Lip}\le 1} \left| \int_S F d(\nu_1-\nu_2) \right|, \tag{2.31}$$

where

$$\|F\|_{Lip} = \sup_{x\in S}|F(x)| + \sup_{x,y\in S,\ x\neq y} \frac{|F(x)-F(y)|}{\rho(x,y)};$$

see, e.g., [80, Theorem 3.28].

Theorem 2.3. *Let* Π *be a Gaussian Borel probability measure on* $L^2(\mathcal{O})$ *with RKHS* $\mathcal{H}=H^r(\mathcal{O})$, $r>d/2$. *Assume that* $f_0\in H^\alpha_c(\mathcal{O})$, $\alpha\ge r-d/2$, *and let* $P^{Y^\varepsilon}_{f_0}$ *be the law of* Y^ε *generated by (2.22) with* $f=f_0$. *Let* $\bar f_\varepsilon = E^\Pi[f|Y^\varepsilon]$ *be the mean of the posterior distribution* $\Pi(\cdot|Y^\varepsilon)$ *arising from observing* (2.22). *Then, for all* $\beta>2+d$ *and any compact* $K\subset\mathcal{O}$, *denoting* $d_{(H^\beta_K)^*}$ *the BL-metric for weak convergence on* $(H^\beta_K(\mathcal{O}))^*$,

$$d_{(H^\beta_K)^*}\left(\mathcal{L}(\varepsilon^{-1}(f-\bar f_\varepsilon)|Y^\varepsilon),\mu\right)\to 0 \tag{2.32}$$

in $P^Y_{f_0}$*-probability as* $\varepsilon\to 0$. *Above* $\mathcal{L}(\varepsilon^{-1}(f-\bar f_\varepsilon)|Y^\varepsilon)$ *is the centred and scaled posterior* (2.30), *and* μ *is the Gaussian distribution induced by* X *in (2.28).*

Similar results as Theorem 2.2 and Theorem 2.3 could be formulated for Example 2.1, exploiting the fact that the 'Fisher information operator' G^*G has a well defined inverse $(G^*G)^{-1}: H^s(\mathcal{O})\to H^{s-2t}(\mathcal{O})$, for all $s\in\mathbb{R}$. In particular, since $\mathcal{O}$ was assumed to be a closed manifold, the weak convergence will be achieved in $H^{-\beta}(\mathcal{O})$ for all $\beta>t+d$.

Remark 2.4 (Applications to uncertainty quantification)**.** With similar reasoning as in Section 2.2.2, Theorem 2.2 implies that for all $\psi\in H^\beta_c$, $\beta>2+d/2$, the credible intervals C_ε in (2.15) centred at the plug-in Tikhonov regulariser $\langle \bar f_\varepsilon,\psi\rangle_{L^2}$ have asymptotically correct frequentist coverage and optimal diameter.

On the other hand, the full strength of Theorem 2.3 can be employed to show that the posterior distribution delivers valid uncertainty quantification also for the entire unknown f, by considering credible sets in the weak topology where the limit is attained.

The weak convergence in the dual space $(H_K^\beta)^*$ is indeed enough to deduce frequentist guarantees for a sufficiently rich class of credible sets (see the related discussion in Section 7.3.4 in [104]). In particular, choosing posterior quantiles $\tilde{R}_\varepsilon = \tilde{R}(\alpha, Y^\varepsilon)$ so that

$$\tilde{C}_\varepsilon = \left\{ f \in L^2 : \|f - \bar{f}_\varepsilon\|_{(H_K^\beta)^*} \leq \tilde{R}_\varepsilon \right\}, \qquad \Pi(\tilde{C}_\varepsilon | Y^\varepsilon) = 1 - \alpha, \quad \alpha \in (0,1),$$

we have for all $\beta > 2 + d$

$$P_{f_0}^{Y^\varepsilon}(f_0 \in \tilde{C}_\varepsilon) \to 1 - \alpha$$

as $\varepsilon \to 0$, with asymptotically vanishing diameter $\tilde{R}_\varepsilon = O_{P_{f_0}^{Y^\varepsilon}}(\varepsilon)$.

Finally, while the optimal rate ε is obtained for the relatively weak norm of $(H_K^\beta)^*$, arguing as in Section 2 in [52] (see also Section 5.1 in [171]), we can intersect $\tilde{C}_\varepsilon$ with additional prior smoothness information (cfr. Step I in Section 2.5.3) to show that the diameter of $\tilde{C}_\varepsilon$ decays at polynomial rate ε^γ, for any $\gamma < \alpha/(\alpha + 2 + d)$, also with respect to the stronger norm of interest $\|\cdot\|_{L^2(K')}$, for any compact $K' \subsetneq K$.

Remark 2.5 (Smoothness requirement)**.** Regarding the weak convergence to μ on $(H_K^\beta)^*$, the requirement that $\beta > 2 + d$ under which (2.32) is obtained is stronger than the necessary tightness condition $\beta > 2 + d/2$ of Lemma 2.1. While the proof of Theorem 2.3 does imply the convergence of the finite-dimensional distributions of $\mathcal{L}(\varepsilon^{-1}(f - \bar{f}_\varepsilon)|Y^\varepsilon)$ to those of μ in the full range $\beta > 2 + d/2$ (see Lemma 2.7), the stronger condition $\beta > 2 + d$ is used crucially in order to control the arising semiparametric bias term *uniformly* in the collection $\{\psi \in H_K^\beta,\ \|\psi\|_{H^\beta} \leq 1\}$. This in turn implies that the L^2-diameter of $\tilde{C}_\varepsilon$ does not attain the minimax rate $\varepsilon^{\alpha/(\alpha+2+d/2)}$, and hence can potentially deliver polynomially sub-optimal results.

To the best of our knowledge, examples of Gaussian priors that attain a nonparametric BvM theorem in the optimal function space are known in literature only in the SVD-based framework considered in [52, 53, 196], or in the 'nearly-diagonal' problem studied very recently by [173]. Applying our proof to a Gaussian prior defined via SVD would here recover the result of [196]. However, the main interest of this chapter is in the performance of standard Gaussian priors that are not defined on the SVD basis of the forward operator - such as the Matérn priors considered in the examples - since this information is rarely available in inverse problems encountered in practice. Our results show that for the inverse problem (2.22) standard Gaussian priors indeed yield optimal semiparametric inference for the maximal class of functionals, and provide a validation of the associated nonparametric credible sets.

2.5 Proofs

2.5.1 Proof of Theorem 2.1

The proof of Theorem 2.1 follows ideas developed in [166] for the special case of G being the X-ray transform and $\psi \in C^\infty$. We will here outline the proof and comment on the main steps. We start by noting that the posterior concentrates on events that have high enough prior probability. As a result, one can confine the analysis to an approximate posterior arising from restricting the prior over such sets. This observation allows to conveniently incorporate concentration properties of the prior into the analysis.

Lemma 2.2. *Let $\Pi(\cdot|Y^\varepsilon)$ be the posterior distribution arising from observation Y^ε in (2.4) and prior Π satisfying Condition 2.1 for a fixed $f_0 \in \overline{\mathcal{W}}$ and some sequence $\delta_\varepsilon \to 0$, such that $\delta_\varepsilon/\varepsilon \to \infty$. Then, for any Borel set $\mathcal{D}_\varepsilon \subset \overline{\mathcal{W}}$ for which*

$$\Pi(\mathcal{D}_\varepsilon^c) \lesssim e^{-D(\delta_\varepsilon/\varepsilon)^2}, \qquad \text{for some } D > 3, \tag{2.33}$$

and all $\varepsilon > 0$ small enough, we have

$$\Pi(\mathcal{D}_\varepsilon^c|Y^\varepsilon) \to 0 \qquad \text{and} \qquad \|\Pi(\cdot|Y^\varepsilon) - \Pi^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon)\|_{TV} \to 0 \tag{2.34}$$

in $P^Y_{f_0}$-probability as $\varepsilon \to 0$. Above $\Pi^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon)$ is the posterior arising from the prior $\Pi(\cdot \cap \mathcal{D}_\varepsilon)/\Pi(\mathcal{D}_\varepsilon)$ restricted to $\mathcal{D}_\varepsilon$ and renormalised.

The proof of Lemma 2.2 (and Lemma 2.3 below) can be adapted from the corresponding results in [166]. They are included for completeness in Appendix 2.C.

Next we need to find a suitable set $\mathcal{D}_\varepsilon$. If $f \sim \Pi$, we have $\langle f, \tilde{\psi}\rangle_{\mathcal{H}} \sim N(0, \|\tilde{\psi}\|^2_{\mathcal{H}})$ for all $\tilde{\psi} \in \mathcal{H}$, and the standard Gaussian tail bound guarantees for all $t \geq 0$ that

$$\Pi\left(f : \frac{|\langle f, \tilde{\psi}\rangle_{\mathcal{H}}|}{\|\tilde{\psi}\|_{\mathcal{H}}} > \frac{t\delta_\varepsilon}{\varepsilon}\right) \leq e^{-\frac{t^2}{2}(\delta_\varepsilon/\varepsilon)^2}.$$

Hence we can choose

$$\mathcal{D}_\varepsilon = \left\{f : \frac{|\langle f, \tilde{\psi}\rangle_{\mathcal{H}}|}{\|\tilde{\psi}\|_{\mathcal{H}}} \leq \frac{T\delta_\varepsilon}{\varepsilon}\right\}, \qquad T > \sqrt{6}. \tag{2.35}$$

We assume that the test function $\psi \in \mathcal{W}_1$ fulfils $|\langle \psi, \varphi\rangle_{\mathcal{W}_1}| \leq \|\varphi\|_{\overline{\mathcal{W}}}$, for all $\varphi \in \mathcal{W}_1$, in order to extend $L_\psi(\cdot) = \langle \cdot, \psi\rangle_{\mathcal{W}_1}$ continuously to $\overline{\mathcal{W}}$. If we assume furthermore that $\psi = -G^*G\tilde{\psi}$, with some $\tilde{\psi} \in \mathcal{H}$, we can proceed to study the moment generating function of $\varepsilon^{-1}(\langle f, \psi\rangle_{\mathcal{W}_1} - \widehat{\Psi})$ under the posterior $\Pi^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon)$, and conclude that it converges to the moment generating function of the limiting Gaussian law.

Lemma 2.3. *Under the conditions of Lemma 2.2, consider a test function $\psi \in \mathcal{W}_1$ such that $|\langle \psi, \varphi\rangle_{\mathcal{W}_1}| \lesssim \|\varphi\|_{\overline{\mathcal{W}}}$, for all $\varphi \in \mathcal{W}_1$, and suppose that $\psi = -G^*G\tilde{\psi}$, for some $\tilde{\psi} \in \mathcal{H}$. Define the random variable*

$$\widehat{\Psi} = \langle f_0, \psi\rangle_{\mathcal{W}_1} - \varepsilon\langle G\tilde{\psi}, \mathbb{W}\rangle_{\mathcal{W}_2}.$$

Then, for all $\tau \in \mathbb{R}$ we have as $\varepsilon \to 0$

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau}{\varepsilon}(\langle f,\psi\rangle_{\mathcal{W}_1} - \widehat{\Psi})}\middle| Y^\varepsilon\right] = e^{\frac{\tau^2}{2}\|G\tilde{\psi}\|^2_{\mathcal{W}_2}}\left(1 + o_{P^{Y^\varepsilon}_{f_0}}(1)\right). \tag{2.36}$$

To conclude, we note that the exponential in the right hand side of (2.36) coincides with the moment generating function of the $N(0, \|G\tilde{\psi}\|^2_{\mathcal{W}_2})$ distribution. Since the convergence of the Laplace transforms implies weak convergence (see, e.g., Proposition 1 in the supplement of [54]), we obtain from Lemma 2.3 that the conclusion of Theorem 2.1 holds for the approximate posterior $\Pi^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon)$. Furthermore, convergence in total variation distance implies convergence in any metric for weak convergence and hence Theorem 2.1 follows from Lemma 2.2.

2.5.2 Proof of Proposition 2.1

We now apply Theorem 2.1 to show the semiparametric result in the elliptic boundary value problem setting of Section 2.3.2. As already noted before Proposition 2.1, any test function $\psi \in H^{r+4}_c = H^{r+4}_c(\mathcal{O})$ verifies the requirements of Theorem 2.1. Hence, we only need to derive Condition 2.1 for the chosen prior. In particular, for Π a Gaussian prior on L^2 with RKHS H^r, $r > d/2$, and the true unknown $f_0 \in H^\alpha_c$, $\alpha \geq 0$, we find suitable sequences δ_ε that satisfy the estimate (2.18) for the concentration function

$$\phi_{\Pi,f_0}(\delta) = \inf_{g\in H^r,\ \|g-f_0\|_{(H^2_0)^*}\leq\delta} \frac{\|g\|^2_{H^r}}{2} - \log \Pi(f : \|f\|_{(H^2_0)^*} \leq \delta), \qquad \delta > 0. \tag{2.37}$$

We proceed by calculating suitable upper bounds for the two terms. For the first term, we need to find approximations for the unknown $f_0 \in H^\alpha_c$ in the RKHS H^r of Π, for which we can both control the approximation error and the norm in the latter space. We employ the approximations used in Section 4.3.3 of [170]. In particular, we fix a compact set F such that $\mathrm{supp}(f_0) \subsetneq F \subsetneq \mathcal{O}$, and a cut-off function $\zeta \in C^\infty_c(\mathbb{R}^d)$ such that $\zeta = 1$

on $\text{supp}(f_0)$, $0 \leq \zeta \leq 1$ and $\text{supp}(\zeta) \subseteq F$. Noting that we can (isometrically) extend f_0 to zero outside F to form an element in $H^\alpha(\mathbb{R}^d)$, we then define

$$f_{0,\varepsilon} = (\zeta \mathcal{F}^{-1} 1_{\{|\cdot| \leq N_\varepsilon\}} \mathcal{F} f_0)|_{\mathcal{O}}, \tag{2.38}$$

for a sequence $N_\varepsilon \to \infty$ as $\varepsilon \to 0$ that will be chosen below.

Lemma 2.4. *Let $f_0 \in H^\alpha_c(\mathcal{O})$ for some $\alpha > 0$, and fix a compact set F such that $\text{supp}(f_0) \subsetneq F \subsetneq \mathcal{O}$. Then we have, for $f_{0,\varepsilon}$ as in (2.38) and for any sequence $N_\varepsilon \to \infty$ as $\varepsilon \to 0$,*

1. *$f_{0,\varepsilon} \in H^t_F(\mathcal{O})$ for all $t \geq 0$ and*

$$\|f_{0,\varepsilon}\|^2_{H^t} \leq (1 + N_\varepsilon^2)^{\max\{0, t-\alpha\}} \|f\|^2_{H^\alpha}; \tag{2.39}$$

2. *for all $0 \leq s < \alpha$*

$$\|f_{0,\varepsilon} - f_0\|^2_{H^s} \leq (1 + N_\varepsilon^2)^{s-\alpha} \|f_0\|^2_{H^\alpha}; \tag{2.40}$$

and for all $s \geq 0$,

$$\|f_{0,\varepsilon} - f_0\|^2_{(H^s)^*} \leq (1 + N_\varepsilon^2)^{-s-\alpha} \|f_0\|^2_{H^\alpha}. \tag{2.41}$$

Proof. Let $t \geq 0$ be fixed. Clearly $\text{supp}(f_{0,\varepsilon}) \subseteq \text{supp}(\zeta) \subseteq F$, and we can compute directly

$$\begin{aligned}
\|f_{0,\varepsilon}\|^2_{H^t(\mathcal{O})} &\leq \|\zeta \mathcal{F}^{-1} 1_{\{|\cdot| \leq N_\varepsilon\}} \mathcal{F} f_0\|^2_{H^t(\mathbb{R}^d)} \\
&\lesssim \int_{\mathbb{R}^d} (1 + |\xi|^2)^t (1_{\{|\xi| \leq N_\varepsilon\}} \mathcal{F} f_0(\xi))^2 d\xi \\
&= \int_{|\xi| \leq N_\varepsilon} (1 + |\xi|^2)^{t-\alpha} (1 + |\xi|^2)^\alpha (\mathcal{F} f_0(\xi))^2 d\xi \\
&\leq (1 + N_\varepsilon^2)^{\max\{0, t-\alpha\}} \|f_0\|^2_{H^\alpha(\mathcal{O})}.
\end{aligned}$$

For $0 \leq s < \alpha$ we proceed similarly, observing that $f_0 = \zeta f_0$ since $\zeta = 1$ on $\text{supp}(f_0)$. Then

$$\begin{aligned}
\|f_{0,\varepsilon} - f_0\|^2_{H^s(\mathcal{O})} &\leq \|\zeta \mathcal{F}^{-1} 1_{\{|\cdot| \leq N_\varepsilon\}} \mathcal{F} f_0 - \zeta f_0\|^2_{H^s(\mathbb{R}^d)} \\
&\lesssim \int_{\mathbb{R}^d} (1 + |\xi|^2)^s (1_{\{|\xi| \leq N_\varepsilon\}} \mathcal{F} f_0(\xi) - \mathcal{F} f_0(\xi))^2 d\xi \\
&\leq (1 + N_\varepsilon^2)^{s-\alpha} \|f_0\|^2_{H^\alpha(\mathcal{O})}.
\end{aligned}$$

Finally, recalling that both f_0 and $f_{0,\varepsilon}$ are supported in $F \subsetneq \mathcal{O}$, we get for all $s \geq 0$

$$\begin{aligned}
\|f_{0,\varepsilon} - f_0\|^2_{(H^s(\mathcal{O}))^*} &= \sup_{u \in H^s(\mathcal{O}),\ \|u\|_{H^s(\mathcal{O})} \leq 1} \left| \int_F (f_{0,\varepsilon} - f_0) u dx \right| \\
&\leq \sup_{U \in H^s(\mathbb{R}^d),\ \|U\|_{H^s(\mathbb{R}^d)} \leq 1} \left| \int_F (f_{0,\varepsilon} - f_0) U|_{\mathcal{O}} dx \right| \\
&= \sup_{U \in H^s(\mathbb{R}^d),\ \|U\|_{H^s(\mathbb{R}^d)} \leq 1} \left| \int_F (\zeta \mathcal{F}^{-1} 1_{\{|\cdot| \leq N_\varepsilon\}} \mathcal{F} f_0 - \zeta f_0) U dx \right| \\
&\lesssim \|\mathcal{F}^{-1} 1_{\{|\cdot| \leq N_\varepsilon\}} \mathcal{F} f_0 - f_0\|_{H^{-s}(\mathbb{R}^d)} \\
&\leq (1 + N_\varepsilon^2)^{-s-\alpha} \|f\|^2_{H^\alpha(\mathcal{O})},
\end{aligned}$$

where the last line follows arguing just as above for the case $0 \leq s < \alpha$.

□

We next derive an upper bound for the second term in (2.37). The proof adapts to the inverse problem (2.22) standard computations in the theory of small balls probabilities of Gaussian priors (e.g., [104, Section 7.3]).

Lemma 2.5. *Let Π be a Gaussian Borel probability measure on $L^2(\mathcal{O})$ with RKHS $\mathcal{H} = H^r(\mathcal{O})$, $r > d/2$. Then, as $\delta \to 0$,*

$$-\log \Pi(f : \|f\|_{(H_0^2(\mathcal{O}))^*} \leq \delta) \lesssim \delta^{-\frac{d}{r+2-d/2}}.$$

Proof. Since for any $f \in L^2$ we have $f = L(L^{-1}f)$, we can write

$$\Pi(f : \|f\|_{(H_0^2)^*} \leq \delta) = \Pi(f : \|L(L^{-1}f)\|_{(H_0^2)^*} \leq \delta).$$

Recalling that L is self-adjoint when acting on H_0^2, we have for some $c > 0$ that

$$\|L(L^{-1}f)\|_{(H_0^2)^*} = \sup_{v \in H_0^2,\ \|v\|_{H^2} \leq 1} |\langle L(L^{-1}f), v \rangle_{L^2}| \leq c \|L^{-1}f\|_{L^2},$$

having used the boundedness of L. Thus,

$$\begin{aligned}
-\log \Pi(f : \|f\|_{(H_0^2)^*} \leq \delta) &\leq -\log \Pi\left(f : \|L^{-1}f\|_{L^2} \leq \delta/c\right) \\
&= -\log \tilde{\Pi}\left(h : \|h\|_{L^2} \leq \delta/c\right)
\end{aligned}$$

where $h = L^{-1}f \sim \tilde{\Pi}$ for $f \sim \Pi$. From Exercise 2.6.5 in [104] and the linearity of L^{-1}, we see that $\tilde{\Pi}$ is a Gaussian probability measure with RKHS $\tilde{\mathcal{H}} = L^{-1}(H^r) = H_0^{r+2}$, with unit ball $B_{\tilde{\mathcal{H}}}$ included in the unit ball B^{r+2} of H^{r+2}. We thus get the following upper bound for the minimal number $N(\delta; B_{\tilde{\mathcal{H}}}, \|\cdot\|_{L^2})$ of L^2-balls or radius δ to cover $B_{\tilde{\mathcal{H}}}$:

$$N(\delta; B_{\tilde{\mathcal{H}}}, \|\cdot\|_{L^2}) \leq N(\delta; B^{r+2}, \|\cdot\|_{L^2}).$$

Theorem 4.3.36 in [104] now implies that

$$\log N(\delta; B^{r+2}, \|\cdot\|_{L^2}) \lesssim \delta^{-\frac{d}{r+2}},$$

and by applying the small ball estimates in Theorem 1.2 of [152], we obtain that as $\delta \to 0$

$$-\log \tilde{\Pi}\left(h : \|h\|_{L^2} \leq \delta/c\right) \lesssim \delta^{-\frac{d}{r+2-d/2}},$$

concluding the proof.

□

Thus, applying Lemma 2.4, the first term in the estimate (2.24) follows by choosing, for any fixed $\delta \geq 0$, N_ε in (2.38) in such a way that

$$(1 + N_\varepsilon^2)^{-2-\alpha} \leq \delta^2,$$

so that, in view of (2.40) and (2.39) respectively, $\|f_{0,\varepsilon} - f_0\|_{(H_0^2)^*} \lesssim \delta$ and $\|f_{0,\varepsilon}\|_{H^r}^2 \lesssim \delta^{-\frac{2\max\{0,r-\alpha\}}{2+\alpha}}$. It can then be readily checked from (2.24) that the sequence δ_ε in (2.25) satisfies the required inequality $\phi_{\Pi,f_0}(\delta_\varepsilon) \lesssim (\delta_\varepsilon/\varepsilon)^2$, concluding the proof of Proposition 2.1.

2.5.3 Proofs of Theorem 2.2 and Theorem 2.3

The key steps of the proof consist in a refinement of the strategy developed to prove Theorem 2.1 and Proposition 2.1. Following [52, 53, 171], we first aim at obtaining the Laplace transform convergence (2.36) uniformly with respect to the test functions $\psi \in H_c^\beta$, $\beta > 2 + d/2$ (cfr. Steps I-II). We subsequently exploit the result to show Theorem 2.2, and to derive the convergence of the finite dimensional distributions of the centred and scaled posterior $\mathcal{L}(\varepsilon^{-1}(f - \bar{f}_\varepsilon)|Y^\varepsilon)$ to those of the limiting Gaussian measure μ (Step III). Finally, combining this observation with a suitable bound on the covariance of the process X_ε in (2.29), we show for each $\beta > 2 + d$ that the distance between $\mathcal{L}(\varepsilon^{-1}(f - \bar{f}_\varepsilon)|Y^\varepsilon)$ and μ, measured in the BL metric on $(H_K^\beta)^*$, vanishes with $P_{f_0}^Y$-probability converging to one (Step IV-V).

Step I: construction of the approximating sets

Let Π be a centred Gaussian prior on $L^2 = L^2(\mathcal{O})$ with RKHS H^r, $r > d/2$, and let $f_0 \in H^\alpha_c$ be fixed. Recall that we assume the prior to undersmooth f_0, namely that $\alpha \geq r - d/2$. Then Remark 2.3 implies that Condition 2.1 is satisfied by taking

$$\delta_\varepsilon \simeq \varepsilon^{\frac{2+r-d/2}{2+r}}. \tag{2.42}$$

In the first step we need to construct appropriate approximating sets $\mathcal{D}_\varepsilon$, by adapting the events introduced in (2.35) for the proof of Theorem 2.1. First, to extend the semiparametric result in Proposition 2.1 to the range $2 + d/2 < \beta < r + 4$, we replace the element $\tilde{\psi} = -LL\psi$ (here not in the RKHS $\mathcal{H} = H^r$) with a suitable approximation. To deal with the possibly diverging norm of such approximations, we will then impose additional constraints to control the size of $f \in \mathcal{D}_\varepsilon$. Finally, to achieve the required uniformity in the Laplace transform convergence (2.36), we will further intersect the resulting events across all test functions, in such a way as to maintain the exponential decay (2.33) for $\Pi(\mathcal{D}^c_\varepsilon)$.

To proceed, let $\beta > 2 + d/2$, let $K \subset \mathcal{O}$ be compact and fix a compact set F such that $K \subsetneq F \subsetneq \mathcal{O}$. Then, for each ψ in a ball

$$B^\beta_K(z) := \left\{ v \in H^\beta_K,\ \|v\|_{H^\beta} \leq z \right\} \tag{2.43}$$

of fixed radius $z > 0$, consider the approximation of $L\psi$ given by Lemma 2.4, of the form

$$\tilde{\psi}_\varepsilon = \left(\zeta \mathcal{F}^{-1} 1_{\{|\cdot| \leq N_\varepsilon\}} \mathcal{F}[L\psi] \right)|_{\mathcal{O}}, \qquad N_\varepsilon \simeq \varepsilon^{-\frac{1}{2+r}}. \tag{2.44}$$

By point 1. in Lemma 2.4, we can uniformly control the Sobolev norms of the resulting collection of approximations. Indeed, by the continuity of L, for all $\psi \in B^\beta_K(z)$ we have $\|L\psi\|_{H^{\beta-2}} \leq z'$ for some constant $z' > 0$, so that in view (2.39), for all $t \geq 0$,

$$\left\{ \tilde{\psi}_\varepsilon,\ \psi \in B^\beta_K(z) \right\} \subseteq B^t_F(b^t_\varepsilon),$$

where

$$b^t_\varepsilon := \sup_{\psi \in B^\beta_K(z)} \|\tilde{\psi}_\varepsilon\|_{H^t} \leq z'(1 + N^2_\varepsilon)^{\max\{0, t-\beta+2\}/2}. \tag{2.45}$$

Then, for all $\psi \in B^\beta_K(z)$, it follows in particular $\tilde{\psi}_\varepsilon \in B^{r+2}_F(b^{r+2}_\varepsilon)$, from which we deduce that $L\tilde{\psi}_\varepsilon \in \mathcal{H} = H^r$. Thus, if $f \sim \Pi$, then $\langle f, L\tilde{\psi}_\varepsilon \rangle_\mathcal{H} \sim N(0, \|L\tilde{\psi}_\varepsilon\|^2_\mathcal{H})$, with variance uniformly bounded, in view of the isomorphism property of L, by

$$\sigma_\varepsilon^2 := \sup_{\psi \in B_K^\beta(z)} E^\Pi |\langle f, L\tilde{\psi}_\varepsilon\rangle_{\mathcal{H}}|^2 \simeq (b_\varepsilon^{r+2})^2. \tag{2.46}$$

Define, for each $\varepsilon > 0$, and $D > 0$ to be chosen below, the approximating set

$$\mathcal{G}_\varepsilon = \left\{ f : \sup_{\psi \in B_K^\beta(z)} |\langle f, L\tilde{\psi}_\varepsilon\rangle_{\mathcal{H}}| \le D\sigma_\varepsilon \delta_\varepsilon/\varepsilon \right\}. \tag{2.47}$$

Here $\mathcal{G}_\varepsilon$ serves as the counterpart of the events (2.35), with the constraint holding simultaneously for all ψ.

We derive the exponential decay (2.33) for $\Pi(\mathcal{G}_\varepsilon^c)$. First, denoting E^Π the expectation under the prior, we have by the Borell-Sudakov-Tirelson inequality [104, Theorem 2.5.8] that for all $\tilde{D} > 0$

$$\Pi\left(f : \sup_{\psi \in B_K^\beta(z)} |\langle f, L\tilde{\psi}_\varepsilon\rangle_{\mathcal{H}}| > E^\Pi \sup_{\psi \in B_K^\beta(z)} |\langle f, L\tilde{\psi}_\varepsilon\rangle_{\mathcal{H}}| + \tilde{D}\sigma_\varepsilon \delta_\varepsilon/\varepsilon \right) \le e^{-\frac{\tilde{D}^2}{2}(\delta_\varepsilon/\varepsilon)^2}. \tag{2.48}$$

Thus, the condition (2.33) will follow if we show that $E^\Pi \sup_{\psi \in B_K^\beta(z)} |\langle f, L\tilde{\psi}_\varepsilon\rangle_{\mathcal{H}}| \lesssim \sigma_\varepsilon \delta_\varepsilon/\varepsilon$. Indeed, in view of (2.45), denoting $B^s(z)$ a ball in H^s of radius z, for general $s \ge 0$ and $z > 0$,

$$E^\Pi \sup_{\psi \in B_K^\beta(z)} |\langle f, L\tilde{\psi}_\varepsilon\rangle_{\mathcal{H}}| \le E^\Pi \sup_{v \in B^{t+2}(z' b_\varepsilon^{t+2})} |\langle f, Lv\rangle_{\mathcal{H}}| \lesssim E^\Pi \sup_{w \in B^t(z'' b_\varepsilon^{t+2})} |\langle f, w\rangle_{\mathcal{H}}|,$$

and Dudley's bound for the expectation of suprema of Gaussian processes [104, Theorem 2.3.8] yields, for σ_ε the constant in (2.46),

$$\begin{aligned} E^\Pi \sup_{w \in B^t(z'' b_\varepsilon^{t+2})} |\langle f, w\rangle_{\mathcal{H}}| &\lesssim \int_0^{\sigma_\varepsilon} \sqrt{\log N(\eta; B^t(z'' b_\varepsilon^{t+2}), \|\cdot\|_{\mathcal{H}})}\, d\eta \\ &= \int_0^{\sigma_\varepsilon} \sqrt{\log N\left(\frac{\eta}{z'' b_\varepsilon^{t+2}}, B^t(1), \|\cdot\|_{\mathcal{H}} \right)}\, d\eta. \end{aligned}$$

Fixing $t > \max\{r + d/2, \beta - 4\}$, recalling $\mathcal{H} = H^r$ and using the known metric entropy estimates for Sobolev balls (see, e.g., [227]), we then obtain

$$E^\Pi \sup_{w \in B^t(z'' b_\varepsilon^{t+2})} |\langle f, w\rangle_{\mathcal{H}}| \lesssim \int_0^{\sigma_\varepsilon} \left(b_\varepsilon^{t+2}/\eta \right)^{\frac{d}{2(t-r)}} d\eta \lesssim (b_\varepsilon^{t+2})^{\frac{d}{2(t-r)}} \sigma_\varepsilon^{\frac{2t-2r-d}{2(t-r)}}.$$

Using (2.46), it follows that

$$\begin{aligned}\frac{1}{\sigma_\varepsilon} E^\Pi \sup_{w\in B^t(z'' b_\varepsilon^{t+2})} |\langle f, w\rangle_{\mathcal{H}}| &\lesssim (b_\varepsilon^{t+2})^{\frac{d}{2(t-r)}} (b_\varepsilon^r)^{-\frac{d}{2(t-r)}} \\ &= (1+N_\varepsilon^2)^{\frac{d(t-\beta+4)}{4(t-r)}} (1+N_\varepsilon^2)^{-\frac{d\max\{0,r-\beta+4\}}{4(t-r)}} \\ &\leq (1+N_\varepsilon^2)^{d/4}.\end{aligned}$$

Recalling that $\delta_\varepsilon \simeq \varepsilon^{\frac{2+r-d/2}{2+r}}$ the choice $N_\varepsilon \simeq \varepsilon^{-\frac{1}{2+r}}$, finally yields

$$\frac{1}{\sigma_\varepsilon} E^\Pi \sup_{\psi\in B_K^\beta(z)} |\langle f, L\tilde{\psi}_\varepsilon\rangle_{\mathcal{H}}| \lesssim \varepsilon^{-\frac{d}{2(2+r)}} \simeq \delta_\varepsilon/\varepsilon.$$

Taking $\tilde{D} > \sqrt{6}$ in (2.48), and sufficiently large $D > \tilde{D}$ in the definition (2.47) of $\mathcal{G}_\varepsilon$, yields the exponential decay (2.33) for $\Pi(\mathcal{G}_\varepsilon^c)$.

Next, we proceed by suitably controlling the size of the elements in the approximating sets. To do so, let, for Φ is the standard normal cumulative distribution function,

$$Q_\varepsilon = -2\Phi^{-1}\left(e^{-\frac{\tilde{D}^2}{2}(\delta_\varepsilon/\varepsilon)^2}\right) \simeq \delta_\varepsilon/\varepsilon.$$

For $\rho > 0$ to be chose below and arbitrary $\kappa > 0$, Consider the event

$$\mathcal{F}_\varepsilon = \left\{ f = f_1 + f_2 : \|f_1\|_{L^2} \leq \rho\varepsilon^{\frac{r-d/2}{2+r}},\ \|f_2\|_{\mathcal{H}} \leq Q_\varepsilon + \kappa \right\} \tag{2.49}$$

in which we constraint the prior draws $f \sim \Pi$ to belong to (a slight enlargement of) a ball of the RKHS $\mathcal{H}$ of growing radius. By the isoperimetric inequality for Gaussian processes [104, Theorem 2.6.12] we can lower bound the prior probability of $\mathcal{F}_\varepsilon$ by

$$\Pi(\mathcal{F}_\varepsilon) \geq \Phi\left(\Phi^{-1}\left[\Pi\left(f : \|f\|_{L^2} \leq \rho\varepsilon^{\frac{r-d/2}{2+r}}\right)\right] + Q_\varepsilon\right). \tag{2.50}$$

Applying again the small ball estimate for Π in Theorem 1.2 in [152] as in the proof of Lemma 2.5, we see that for some $b > 0$

$$-\log\Pi\left(f : \|f\|_{L^2} \leq \rho\varepsilon^{\frac{r-d/2}{2+r}}\right) \leq b\rho^{-\frac{d}{r-d/2}}\varepsilon^{-\frac{d}{2+r}}$$

and recalling that $\delta_\varepsilon/\varepsilon \simeq \varepsilon^{-\frac{d/2}{2+r}}$, we can choose $\rho > 0$ so that

$$-\log\Pi\left(f : \|f\|_{L^2} \leq \rho\varepsilon^{\frac{r-d/2}{2+r}}\right) \leq (\delta_\varepsilon/\varepsilon)^2.$$

Combining the above with (2.50) yields

$$\Pi(\mathcal{F}_\varepsilon) \geq \Phi\left(\Phi^{-1}\left(e^{-(\delta_\varepsilon/\varepsilon)^2}\right) + Q_\varepsilon\right) \geq \Phi\left(\Phi^{-1}\left(e^{-\frac{\tilde{D}^2}{2}(\delta_\varepsilon/\varepsilon)^2}\right) + Q_\varepsilon\right) = \Phi(Q_\varepsilon/2),$$

and finally $\Pi(\mathcal{F}_\varepsilon^c) \leq e^{-\frac{\tilde{D}^2}{2}(\delta_\varepsilon/\varepsilon)^2}$.

We conclude by taking

$$\mathcal{D}_\varepsilon = \mathcal{G}_\varepsilon \cap \mathcal{F}_\varepsilon, \tag{2.51}$$

for which the bounds on $\Pi(\mathcal{G}_\varepsilon^c)$ and $\Pi(\mathcal{F}_\varepsilon^c)$ imply $\Pi(\mathcal{D}_\varepsilon^c) \leq 2e^{-\frac{\tilde{D}^2}{2}(\delta_\varepsilon/\varepsilon)^2}$, $\tilde{D}^2/2 > 3$.

Step II: Laplace transform expansion

We proceed deriving an asymptotic expression, analogous to the one obtained in Lemma 2.3, for the Laplace transform of the linear functionals $\langle f, \psi\rangle_{L^2}$. In view of the simultaneous constraint imposed in (2.47), the result holds uniformly with respect to test functions ψ.

Lemma 2.6. *Let Π be a Gaussian Borel probability measure on $L^2(\mathcal{O})$ with RKHS $\mathcal{H} = H^r(\mathcal{O})$, $r > d/2$, and assume that $f_0 \in H_c^\alpha(\mathcal{O})$, $\alpha \geq r - d/2$. For all $\beta > 2 + d/2$, and any $\psi \in B_K^\beta(z)$ (defined as in (2.43)), $z > 0$, let $\tilde{\psi}_\varepsilon$ be the approximation in (2.44), and define*

$$\hat{\Psi}(\psi) = \langle f_0, \psi\rangle_{L^2} + \varepsilon\langle\tilde{\psi}_\varepsilon, \mathbb{W}\rangle_{L^2}. \tag{2.52}$$

Then, for all fixed $\tau \in \mathbb{R}$

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau^2}{\varepsilon}[\langle f,\psi\rangle_{L^2} - \hat{\Psi}(\psi)]}\middle| Y^\varepsilon\right] = e^{R_\varepsilon} e^{\frac{\tau^2}{2}\|L\psi\|_{L^2}^2} \frac{\Pi(\mathcal{D}_{\varepsilon,\tau}|Y^\varepsilon)}{\Pi(\mathcal{D}_\varepsilon|Y^\varepsilon)}, \tag{2.53}$$

where $\mathcal{D}_{\varepsilon,\tau} = \{f - \tau\varepsilon L\tilde{\psi}_\varepsilon,\ f \in \mathcal{D}_\varepsilon\}$ and $R_\varepsilon \to 0$ uniformly in $B_K^\beta(z)$ for any $z > 0$ as $\varepsilon \to 0$.

Proof. We have

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau^2}{\varepsilon}[\langle f,\psi\rangle_{L^2} - \hat{\Psi}(\psi)]}\middle| Y^\varepsilon\right] = e^{-\tau\langle\tilde{\psi}_\varepsilon, \mathbb{W}\rangle_{L^2}} E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau}{\varepsilon}\langle f - f_0, \psi\rangle_{L^2}}\middle| Y^\varepsilon\right],$$

and letting $f_\tau = f - \tau\varepsilon L\tilde{\psi}_\varepsilon$, the expectation in the right hand side becomes (cfr. (2.7))

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau}{\varepsilon}\langle f-f_0,\psi\rangle_{L^2}}\middle|Y^\varepsilon\right] = \frac{\int_{\mathcal{D}_\varepsilon} e^{\frac{\tau}{\varepsilon}\langle f-f_0,\psi\rangle_{L^2}} e^{\ell_\varepsilon(f)-\ell_\varepsilon(f_\tau)} e^{\ell_\varepsilon(f_\tau)} d\Pi(f)}{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f)} d\Pi(f)}.$$

From the expression of the log-likelihood (2.6) we readily obtain

$$\ell_\varepsilon(f) - \ell_\varepsilon(f_\tau) = \frac{\tau^2}{2}\|\tilde{\psi}_\varepsilon\|_{L^2}^2 + \frac{\tau}{\varepsilon}\langle L^{-1}(f-f_0), \tilde{\psi}_\varepsilon\rangle_{L^2} + \tau\langle\tilde{\psi}_\varepsilon, \mathbb{W}\rangle_{L^2},$$

which substituted into the previous expression yields, using the self-adjointness of L^{-1},

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau^2}{\varepsilon}[\langle f,\psi\rangle_{L^2}-\hat{\Psi}(\psi)]}\middle|Y^\varepsilon\right] \tag{2.54}$$

$$= e^{\frac{\tau^2}{2}\|\tilde{\psi}_\varepsilon\|_{L^2}^2} e^{-\frac{\tau}{\varepsilon}\langle L^{-1}f_0, L\psi-\tilde{\psi}_\varepsilon\rangle_{L^2}} \frac{\int_{\mathcal{D}_\varepsilon} e^{\frac{\tau}{\varepsilon}\langle L^{-1}f, L\psi-\tilde{\psi}_\varepsilon\rangle_{L^2}} e^{\ell_\varepsilon(f_\tau)} d\Pi(f)}{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f)} d\Pi(f)}. \tag{2.55}$$

In view of (2.40), we have that $\|\tilde{\psi}_\varepsilon - L\psi\|_{L^2} \to 0$ as $\varepsilon \to 0$ uniformly in $B_K^\beta(z)$ for all $z > 0$, and hence

$$e^{\frac{\tau^2}{2}\|\tilde{\psi}_\varepsilon\|_{L^2}^2} = (1+o(1))e^{\frac{\tau^2}{2}\|L\psi\|_{L^2}^2}. \tag{2.56}$$

Next we prove that, uniformly in $B_K^\beta(z)$,

$$e^{-\frac{\tau}{\varepsilon}\langle L^{-1}f_0, L\psi-\tilde{\psi}_\varepsilon\rangle_{L^2}} = 1 + o(1). \tag{2.57}$$

To do so, note

$$\begin{aligned}
&\sup_{\psi\in B_K^\beta(z)}\left| -\frac{1}{\varepsilon}\langle L^{-1}f_0, L\psi-\tilde{\psi}_\varepsilon\rangle_{L^2}\right| \\
&\quad= \frac{1}{\varepsilon}\|L^{-1}f_0\|_{H^{\alpha+2}} \sup_{\psi\in B_K^\beta(z)}\left|\left\langle \frac{L^{-1}f_0}{\|L^{-1}f_0\|_{H^{\alpha+2}}}, L\psi-\tilde{\psi}_\varepsilon\right\rangle_{L^2}\right| \\
&\quad\lesssim \frac{1}{\varepsilon}\sup_{\psi\in B_K^\beta(z)}\|L\psi-\tilde{\psi}_\varepsilon\|_{(H_0^{\alpha+2})^*} \\
&\quad\lesssim \varepsilon^{-1}(1+N_\varepsilon^2)^{\frac{-\alpha-\beta}{2}}
\end{aligned}$$

where the last line follows by (2.41). Recalling that $N_\varepsilon \simeq \varepsilon^{-\frac{1}{2+r}}$, $\alpha \geq r - d/2$ and $\beta > 2 + d/2$,

$$\sup_{\psi \in B_K^\beta(z)} \left| -\frac{\tau}{\varepsilon} \langle L^{-1} f_0, L\psi - \tilde{\psi}_\varepsilon \rangle_{L^2} \right| \lesssim \frac{1}{\varepsilon}(1 + N_\varepsilon^2)^{\frac{-\alpha-\beta}{2}} \simeq \varepsilon^{\frac{\alpha+\beta-2-r}{2+r}} \to 0.$$

The following step consists in showing that, uniformly in $B_K^\beta(z)$,

$$\int_{\mathcal{D}_\varepsilon} e^{\frac{\tau}{\varepsilon}\langle L^{-1} f, L\psi - \tilde{\psi}_\varepsilon \rangle_{L^2}} e^{\ell_\varepsilon(f_\tau)} d\Pi(f) = (1 + o(1)) \int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f_\tau)} d\Pi(f). \tag{2.58}$$

The result will follow from the dominated convergence theorem upon showing that

$$\sup_{f \in \mathcal{D}_\varepsilon} \sup_{\psi \in B_K^\beta(z)} \left| \frac{\tau}{\varepsilon} \langle L^{-1} f, L\psi - \tilde{\psi}_\varepsilon \rangle_{L^2} \right| \to 0.$$

Recalling the definition (2.51) of $\mathcal{D}_\varepsilon$, we bound the left hand side by

$$\begin{aligned} &\sup_{f \in \mathcal{F}_\varepsilon} \sup_{\psi \in B_K^\beta(z)} \left| \frac{\tau}{\varepsilon} \langle L^{-1} f, L\psi - \tilde{\psi}_\varepsilon \rangle_{L^2} \right| \\ &\quad \lesssim \frac{1}{\varepsilon} \sup_{\|f_1\|_{L^2} \leq \rho \varepsilon^{\frac{r-d/2}{2+r}}} \sup_{\psi \in B_K^\beta(z)} \left| \langle L^{-1} f_1, L\psi - \tilde{\psi}_\varepsilon \rangle_{L^2} \right| \\ &\qquad + \frac{1}{\varepsilon} \sup_{\|f_2\|_{\mathcal{H}} \leq Q_\varepsilon + \kappa} \sup_{\psi \in B_K^\beta(z)} \left| \langle L^{-1} f_2, L\psi - \tilde{\psi}_\varepsilon \rangle_{L^2} \right|. \end{aligned}$$

Accordingly, it is enough to show the joint convergence of the two terms above, which can be done similarly as in the derivation of (2.57). In particular

$$\frac{1}{\varepsilon} \sup_{\|f_1\|_{L^2} \leq \rho \varepsilon^{\frac{r-d/2}{2+r}}} \sup_{\psi \in B_K^\beta(z)} \left| \langle L^{-1} f, L\psi - \tilde{\psi}_\varepsilon \rangle_{L^2} \right| \lesssim \varepsilon^{\frac{r-d/2}{2+r} - 1}(1 + N_\varepsilon^2)^{-\frac{\beta}{2}} \to 0.$$

On the other hand, recalling that $\mathcal{H} = H^r$,

$$\frac{1}{\varepsilon} \sup_{\|f_2\|_{\mathcal{H}} \leq Q_\varepsilon + \kappa} \sup_{\psi \in B_K^\beta(z)} \left| \langle L^{-1} f_2, L\psi - \tilde{\psi}_\varepsilon \rangle_{L^2} \right| \lesssim \varepsilon^{-1}(Q_\varepsilon + \kappa)(1 + N_\varepsilon^2)^{-\frac{r+\beta}{2}} \to 0,$$

since $\delta_\varepsilon \simeq \varepsilon^{\frac{2+r-d/2}{2+r}}$ and $Q_\varepsilon = -2\Phi^{-1}\left(c' e^{-\frac{\bar{D}^2}{2}(\delta_\varepsilon/\varepsilon)^2}\right) \simeq \delta_\varepsilon/\varepsilon$. Replacing (2.56), (2.57) and (2.58) into (2.54) we obtain, uniformly in $B_K^\beta(z)$,

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau}{\varepsilon}[\langle f,\psi\rangle_{L^2}-\hat{\Psi}(\psi)]}\Big|Y^\varepsilon\right] = (1+o(1))e^{\frac{\tau^2}{2}\|L\psi\|_{L^2}^2}\frac{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f_\tau)}d\Pi(f)}{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f)}d\Pi(f)}. \tag{2.59}$$

We conclude by further simplifying the ratio in the right hand side of (2.59) in the same way as in the conclusion of the proof of Proposition 3.2 in [166]. Let Π_τ be the law of the shifted parameter $f_\tau = f - \tau\varepsilon L\tilde{\psi}_\varepsilon$, and $\mathcal{D}_{\varepsilon,\tau} = \{f - \tau\varepsilon L\tilde{\psi}_\varepsilon,\ f \in \mathcal{D}_\varepsilon\}$. Then, the Cameron-Martin theorem (e.g., Theorem 2.6.13 in [104]) yields

$$\frac{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f_\tau)}d\Pi(f)}{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f)}d\Pi(f)} = e^{-\frac{(\tau\varepsilon)^2}{2}\|L\tilde{\psi}_\varepsilon\|_{\mathcal{H}}^2}\frac{\int_{\mathcal{D}_{\varepsilon,\tau}} e^{\ell_\varepsilon(g)}e^{-\tau\varepsilon\langle L\tilde{\psi}_\varepsilon,g\rangle_{\mathcal{H}}}d\Pi(g)}{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(g)}d\Pi(g)}.$$

First note that by (2.39),

$$\sup_{\psi\in B_K^\beta(z)} \varepsilon^2\|L\tilde{\psi}_\varepsilon\|_{\mathcal{H}}^2 \lesssim \sup_{\psi\in B_K^\beta(z)} \varepsilon^2\|\tilde{\psi}_\varepsilon\|_{H^{r+2}}^2 \lesssim \varepsilon^{2-\frac{2\max\{0,r-\beta+4\}}{2+r}} \to 0. \tag{2.60}$$

Next, recalling the definitions (2.47) and (2.51) of $\mathcal{G}_\varepsilon$ and $\mathcal{D}_\varepsilon$ respectively, we have

$$\begin{aligned}
\sup_{g\in\mathcal{D}_{\varepsilon,\tau}}\sup_{\psi\in B_K^\beta(z)} |\varepsilon\langle L\tilde{\psi}_\varepsilon, g\rangle_{\mathcal{H}}| &= \varepsilon \sup_{f\in\mathcal{D}_\varepsilon}\sup_{\psi\in B_K^\beta(z)} |\langle L\tilde{\psi}_\varepsilon, f - \tau\varepsilon L\tilde{\psi}_\varepsilon\rangle_{\mathcal{H}}| \\
&\le \varepsilon\left(\sup_{f\in\mathcal{G}_\varepsilon}\sup_{\psi\in B_K^\beta(z)} |\langle L\tilde{\psi}_\varepsilon, f\rangle_{\mathcal{H}}| + |\tau|\varepsilon \sup_{\psi\in B_K^\beta(z)}\|L\tilde{\psi}_\varepsilon\|_{\mathcal{H}}^2\right) \\
&\lesssim \sigma_\varepsilon\delta_\varepsilon + o(1)
\end{aligned}$$

and, since $\beta > 2 + d/2$ and $r > d/2$, then $\sigma_\varepsilon\delta_\varepsilon \lesssim \varepsilon^{\frac{\min\{r+2-d/2,\beta-2-d/2\}}{2+r}} \to 0$. Thus, we conclude that uniformly in $B_K^\beta(z)$ as $\varepsilon \to 0$,

$$\frac{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f_\tau)}d\Pi(f)}{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f)}d\Pi(f)} = (1+o(1))\frac{\int_{\mathcal{D}_{\varepsilon,\tau}} e^{\ell_\varepsilon(g)}d\Pi(g)}{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(g)}d\Pi(g)} = (1+o(1))\frac{\Pi(\mathcal{D}_{\varepsilon,\tau}|Y^\varepsilon)}{\Pi(\mathcal{D}_\varepsilon|Y^\varepsilon)}.$$

□

Step III: convergence of the finite dimensional distributions

We now exploit the previous lemma to show the convergence of the finite dimensional distributions of the centred and scaled posterior to those of the Gaussian measure μ

induced by the process X in (2.28). The result is obtained by showing that the ratio in the right hand side of (2.53) converges to one, yielding, as in Lemma 2.3, the desired asymptotic expression for the Laplace transform. This will in turn conclude the proof of Theorem 2.2.

To proceed, consider the centring Gaussian process obtained by collecting the random variables introduced in (2.52),

$$\hat{\Psi}_\varepsilon = \left(\hat{\Psi}_\varepsilon(\psi) : \psi \in H^\beta_c\right), \qquad \hat{\Psi}_\varepsilon(\psi) = \langle f_0, \psi\rangle_{L^2} + \varepsilon\langle\tilde{\psi}_\varepsilon, \mathbb{W}\rangle_{L^2}, \qquad \varepsilon > 0. \tag{2.61}$$

In view of Lemma 2.1, and by the continuity of the linear map $\psi \in H^\beta_c \mapsto \tilde{\psi}_\varepsilon \in H^t$, $t \geq 0$, $\hat{\Psi}_\varepsilon$ defines a Borel measurable map on $(H^\beta_c)^*$, for $\beta > 2 + d/2$. Then, denote by

$$\mathcal{L}\left(\varepsilon^{-1}\left(f - \hat{\Psi}_\varepsilon\right)\Big|Y^\varepsilon\right) = \mathcal{L}(\hat{X}_\varepsilon), \qquad f \sim \Pi \tag{2.62}$$

the tight conditional law on $(H^\beta_c)^*$ of

$$\hat{X}_\varepsilon = \left(\hat{X}_\varepsilon(\psi) : \psi \in H^\beta_c\right), \qquad \hat{X}_\varepsilon(\psi) = \varepsilon^{-1}\left(\langle f, \psi\rangle_{L^2} - \hat{\Psi}_\varepsilon(\psi)\right)\Big|Y^\varepsilon. \tag{2.63}$$

Lemma 2.7. *For any fixed $\psi_1, \dots, \psi_n \in H^\beta_c$, consider the following Borel probability measures on $\mathbb{R}^n$:*

$$\mathcal{L}\left(\varepsilon^{-1}(f - \hat{\Psi}_\varepsilon)\Big|Y^\varepsilon\right)_n := \mathcal{L}\left(\varepsilon^{-1}[\langle f, \psi_1\rangle_{L^2} - \hat{\Psi}_\varepsilon(\psi_1), \dots, \langle f, \psi_n\rangle_{L^2} - \hat{\Psi}_\varepsilon(\psi_n)]\Big|Y^\varepsilon\right),$$

where $f \sim \Pi$, and

$$\mu_n := \mathcal{L}(X(\psi_1), \dots, X(\psi_n)),$$

where X is as in (2.28). Then, denoting $d_{\mathbb{R}^n}$ the BL-metric for weak convergence on $\mathbb{R}^n$, we have in $P^Y_{f_0}$-probability as $\varepsilon \to 0$.

$$d_{\mathbb{R}^n}\left(\mathcal{L}\left(\varepsilon^{-1}\left(f - \hat{\Psi}_\varepsilon\right)\Big|Y^\varepsilon\right)_n, \mu_n\right) \to 0. \tag{2.64}$$

Proof. By Lemma 2.2, it is enough to show (2.64) for $f \sim \Pi^{\mathcal{D}_\varepsilon}$, with $\mathcal{D}_\varepsilon$ as in (2.51). Let $\psi \in H^\beta_c$ be fixed. Then, by taking $K = \mathrm{supp}(\psi)$, Lemma 2.6 implies

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau}{\varepsilon}[\langle f,\psi\rangle_{L^2} - \hat{\Psi}_\varepsilon(\psi)]}\Big|Y^\varepsilon\right] = (1 + o(1))e^{\frac{\tau^2}{2}\|L\psi\|^2_{L^2}}\frac{\Pi(\mathcal{D}_{\varepsilon,\tau}|Y^\varepsilon)}{\Pi(\mathcal{D}_\varepsilon|Y^\varepsilon)},$$

and the proof is concluded by showing that the ratio on the right hand side converges to 1 in $P^Y_{f_0}$-probability as $\varepsilon \to 0$. Indeed, if this is the case, the convergence of the Laplace transform will imply that for any fixed $\psi \in H_c^\beta$,

$$d_{\mathbb{R}}\left(\mathcal{L}\left(\varepsilon^{-1}\left(\langle f,\psi\rangle_{L^2} - \hat{\Psi}_\varepsilon(\psi)\right)\Big|Y^\varepsilon\right), \mathcal{L}(X(\psi))\right) \to 0, \tag{2.65}$$

in $P_{f_0}^Y$-probability as $\varepsilon \to 0$; and, by the Cramer-Wold device, also (2.64) will follow by replacing ψ with any finite linear combination $\sum_{i=1}^n a_i\psi_i \in H_c^\beta$.

To proceed, first recall that Lemma 2.2 implies that $\Pi(\mathcal{D}_\varepsilon|Y^\varepsilon) \to 1$ in $P_{f_0}^Y$-probability as $\varepsilon \to 0$. To apply the same result to the numerator, we show that the prior probability of $D_{\varepsilon,\tau}^c$ decays exponentially as required by (2.33). Note

$$\mathcal{D}_{\varepsilon,\tau} = \{f - \tau\varepsilon L\tilde{\psi}_\varepsilon,\ f \in \mathcal{G}_\varepsilon \cap \mathcal{F}_\varepsilon\} = \mathcal{G}_{\varepsilon,\tau} \cap \mathcal{F}_{\varepsilon,\tau},$$

where $\mathcal{G}_{\varepsilon,\tau},\ \mathcal{F}_{\varepsilon,\tau}$ are defined analogously to the set $\mathcal{D}_{\varepsilon,\tau}$ introduced in the previous lemma. It is hence enough to deduce (2.33) for $\mathcal{G}_{\varepsilon,\tau}$ and $\mathcal{F}_{\varepsilon,\tau}$ separately.

First, from the definition of $\mathcal{G}_\varepsilon$ in (2.47), we see that

$$\mathcal{G}_{\varepsilon,\tau} \supseteq \left\{g : \sup_{\phi\in B_K^\beta(z)} |\langle g, L\tilde{\phi}_\varepsilon\rangle_{\mathcal{H}}| \le D\sigma_\varepsilon\delta_\varepsilon/\varepsilon - |\tau\varepsilon|\|L\tilde{\psi}_\varepsilon\|_{\mathcal{H}} \sup_{\phi\in B_K^\beta(z)} \|L\tilde{\phi}_\varepsilon\|_{\mathcal{H}}\right\}.$$

Now, using (2.46) and recalling $\delta_\varepsilon \simeq \varepsilon^{\frac{2+r-d/2}{2+r}}$,

$$\varepsilon\|L\tilde{\psi}_\varepsilon\|_{\mathcal{H}} \sup_{\phi\in B_K^\beta(z)} \|L\tilde{\phi}_\varepsilon\|_{\mathcal{H}} \lesssim \varepsilon^{1-\frac{\max\{0,r-\beta+4\}}{2+r}}\sigma_\varepsilon = o\left(\sigma_\varepsilon\delta_\varepsilon/\varepsilon\right).$$

Then, for all $\varepsilon > 0$ small enough

$$\mathcal{G}_{\varepsilon,\tau} \supseteq \left\{g : \sup_{\phi\in B_K^\beta(z)} |\langle g, L\tilde{\phi}_\varepsilon\rangle_{\mathcal{H}}| \le D\sigma_\varepsilon\delta_\varepsilon/\varepsilon\right\},$$

and by our particular choices of $D > \tilde{D} > \sqrt{6}$ we obtain (via the Borel-Sudakov-Tirelson inequality) that $\Pi(G_{\varepsilon,\tau}^c) \le e^{-\frac{\tilde{D}^2}{2}(\delta_\varepsilon/\varepsilon)^2}$. On the other hand, for $\mathcal{F}_\varepsilon$ defined in (2.49),

$$\mathcal{F}_{\varepsilon,\tau} \supseteq \left\{f_1 + f_2' : \|f_1\|_{L^2} \le \rho\varepsilon^{\frac{r-d/2}{2+r}},\ \|f_2'\|_{\mathcal{H}} \le Q_\varepsilon + \kappa - |\tau|\varepsilon\|L\tilde{\psi}_\varepsilon\|_{\mathcal{H}}\right\}$$

and since, by (2.60), $\varepsilon\|L\tilde{\psi}_\varepsilon\|_{\mathcal{H}} \to 0$ as $\varepsilon \to 0$, we have $Q_\varepsilon + \kappa - |\tau|\varepsilon\|L\tilde{\psi}_\varepsilon\|_{\mathcal{H}} > Q_\varepsilon$ for $\varepsilon > 0$ small enough. Then, for all such $\varepsilon > 0$,

$$\mathcal{F}_{\varepsilon,\tau} \supseteq \left\{f_1 + f_2' : \|f_1\|_{L^2} \le \rho\varepsilon^{\frac{r-d/2}{2+r}},\ \|f_2'\|_{\mathcal{H}} \le Q_\varepsilon\right\},$$

and by the isoperimetric inequality for Gaussian processes we can conclude that $\Pi(F^c_{\varepsilon,\tau}) \lesssim e^{-\frac{\tilde D^2}{2}(\delta_\varepsilon/\varepsilon)^2}$, $\tilde D > \sqrt{6}$, as required.

□

Using the same argument as in the conclusion of the proof of Theorem 2.1, we deduce from the above lemma that the semiparametric BvM phenomenon displayed in (2.27) does occur for all $\beta > 2 + d/2$, concluding the proof of Theorem 2.2.

Step IV: weak convergence in $(H^\beta_K)^*$

Assume now that $\beta > 2 + d$. Combining the convergence of the finite dimensional distribution established in the previous step with a uniform bound on the covariance of the process $\hat X_\varepsilon$ in (2.63) implied by Lemma 2.6, we show that $\mathcal{L}(\varepsilon^{-1}(f - \hat\Psi_\varepsilon)|Y^\varepsilon)$ converges weakly to μ. That $\hat\Psi_\varepsilon$ can be replaced by the posterior mean $\bar f_\varepsilon$ can then be shown analogously as in the proof of Theorem 2.7 in [166]; see Step V in Appendix 2.C.

It is again enough to consider the restricted prior $\Pi^{\mathcal{D}_\varepsilon}$. Thus, for any fixed compact set $K \subset \mathcal{O}$, let $\tilde\Pi^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon)$ be the tight Gaussian law on $(H^\beta_K)^*$ induced by $\hat X_{\varepsilon,K} := (\hat X_\varepsilon(\psi) : \psi \in H^\beta_K)$, where $\hat X_\varepsilon(\psi)$ is as in (2.63) with $f \sim \Pi^{\mathcal{D}_\varepsilon}$. To exploit the convergence of the finite dimensional distributions, we further consider 'projections' of $\tilde\Pi^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon)$ onto suitable subspaces. In particular, let $\{\Phi_{lr},\ l \geq -1,\ r = 1,\dots N_l\}$, $N_l \lesssim 2^{ld}$, be an orthonormal basis of $L^2(\mathcal{O})$ of sufficiently regular boundary corrected Daubechies wavelets. We will exploit the fact that such basis conveniently characterises the Sobolev regularity of the test functions in terms of the decay of the wavelets coefficients (see [228] or also Chapter 4 of [104] for details).

For any $\lambda \in \mathbb{N}$ and all $\psi \in H^\beta_K$, let $P_\lambda\psi$ denote the projection of ψ onto the finite dimensional subspace spanned by $\{\Phi_{lr},\ l \leq \lambda,\ r \leq N_l\}$. Next, define the projected posterior $\tilde\Pi^{\mathcal{D}_\varepsilon}_\lambda(\cdot|Y^\varepsilon)$ as the law of the process $P_\lambda \hat X_{\varepsilon,K} := (\hat X_\varepsilon(P_\lambda\psi) : \psi \in H^\beta_K)$; define analogously the projected limiting law μ_λ. For $d = d_{(H^\beta_K)^*}$ the BL-metric for weak convergence of probability measures defined on $(H^\beta_K)^*$, the triangular inequality then yields

$$d(\tilde\Pi^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon), \mu) \leq d\left(\tilde\Pi^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon), \tilde\Pi^{\mathcal{D}_\varepsilon}_\lambda(\cdot|Y^\varepsilon)\right) + d\left(\tilde\Pi^{\mathcal{D}_\varepsilon}_\lambda(\cdot|Y^\varepsilon), \mu_\lambda\right) + d\left(\mu_\lambda, \mu\right). \qquad (2.66)$$

We show that the three term in the right hand side vanish. For the first, recalling the definition of the BL-metric (2.31), we have

$$d\left(\tilde{\Pi}^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon), \tilde{\Pi}^{\mathcal{D}_\varepsilon}_\lambda(\cdot|Y^\varepsilon)\right) = \sup_{F:(H^\beta_K)^*\to\mathbb{R},\ \|F\|_{Lip}\le 1} \left|E^{\Pi^{\mathcal{D}_\varepsilon}}[F(\hat{X}_{\varepsilon,K}) - F(P_\lambda \hat{X}_{\varepsilon,K})]\right|$$
$$\le E^{\Pi^{\mathcal{D}_\varepsilon}} \left\|\hat{X}_{\varepsilon,K} - P_\lambda \hat{X}_{\varepsilon,K}\right\|_{(H^\beta_K)^*},$$

which, by definition of the norm in $(H^\beta_K)^*$, equals

$$E^{\Pi^{\mathcal{D}_\varepsilon}} \sup_{\psi\in B^\beta_K(1)} |\hat{X}_\varepsilon(\psi - P_\lambda\psi)| \le E^{\Pi^{\mathcal{D}_\varepsilon}} \sup_{\psi\in B^\beta_K(1)} \sum_{l>\lambda}\sum_{r=1}^{N_l} |\langle\psi, \Phi_{lr}\rangle_{L^2}||\hat{X}_\varepsilon(\Phi_{lr})|,$$

with $B^\beta_K(1)$ defined as in (2.43). Note that, as $\mathrm{supp}(\psi) \subset K$, for λ large enough (only depending on K) the above sum involves only wavelets that are compactly supported within $\mathcal{O}$. We now apply Hölder's inequality and the wavelet characterisation of Sobolev norms to upper bound the right hand side by

$$E^{\Pi^{\mathcal{D}_\varepsilon}} \sup_{\psi\in B^\beta_K(1)} \sum_{l>\lambda} \sqrt{\sum_{r=1}^{N_l}\langle\psi, \Phi_{lr}\rangle^2_{L^2}} \sqrt{\sum_{r=1}^{N_l} |\hat{X}_\varepsilon(\Phi_{lr})|^2} \lesssim \sum_{l>\lambda} 2^{-l\beta} E^{\Pi^{\mathcal{D}_\varepsilon}} \sqrt{\sum_{r=1}^{N_l} |\hat{X}_\varepsilon(\Phi_{lr})|^2}.$$

Jensen's inequality implies the further upper bound

$$\sum_{l>\lambda} 2^{-(\beta-\beta')l} \sqrt{\sum_{r=1}^{N_l} E^{\Pi^{\mathcal{D}_\varepsilon}} |\hat{X}_\varepsilon(2^{-\beta' l}\Phi_{lr})|^2}$$

having scaled the wavelets by a factor $2^{-\beta' l}$ for some $2 + d/2 < \beta' < \beta - d/2$ (possible since we here assume $\beta > 2 + d$). In particular, since $\|\Phi_{lr}\|^2_{H^{\beta'}} \simeq 2^{2\beta' l}$, we have that $2^{-\beta' l}\Phi_{lr} \in B^{\beta'}_c(1)$. Using Lemma 2.6 and the fact that $\|L\Phi_{lr}\|^2_{L^2} \lesssim \|\Phi_{lr}\|_{H^2} \simeq 2^{4l}$, we obtain

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\hat{X}_\varepsilon(2^{-\beta' l}\Phi_{lr})}\Big|Y^\varepsilon\right] = e^{R_\varepsilon} e^{\frac{1}{2}\|2^{-\beta' l} L\Phi_{lr}\|^2_{L^2}} \frac{\Pi(\mathcal{D}_{\varepsilon,\tau}|Y^\varepsilon)}{\Pi(\mathcal{D}_\varepsilon|Y^\varepsilon)}$$
$$\le \frac{e^{R_\varepsilon}}{\Pi(\mathcal{D}_\varepsilon|Y^\varepsilon)} e^{2^{-2\beta' l-1}\|L\Phi_{lr}\|^2_{L^2}}$$
$$\lesssim r_\varepsilon$$
$$= O_{P^{Y^\varepsilon}_{f_0}}(1)$$

having lower bounded the probability on the denominator by Lemma 2.2. Then, since $x^2 \le e^x + e^{-x}$ for all $x \in \mathbb{R}$,

$$E^{\Pi^{\mathcal{D}_\varepsilon}}|\hat{X}_\varepsilon(2^{-\beta' l}\Phi_{lr})|^2 \le E^{\Pi^{\mathcal{D}_\varepsilon}} e^{\hat{X}_\varepsilon(2^{-\beta' l}\Phi_{lr})} + E^{\Pi^{\mathcal{D}_\varepsilon}} e^{-\hat{X}_\varepsilon(2^{-\beta' l}\Phi_{lr})} \lesssim r_\varepsilon.$$

We thus obtain, recalling $N_l \simeq 2^{ld}$,

$$d\left(\tilde{\Pi}^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon), \tilde{\Pi}^{\mathcal{D}_\varepsilon}_\lambda(\cdot|Y^\varepsilon)\right) \lesssim \sum_{l>\lambda} 2^{-(\beta-\beta')l} \sqrt{\sum_{r=1}^{N_l} r_\varepsilon} \lesssim r'_\varepsilon \sum_{l>\lambda} 2^{-(\beta-\beta'-d/2)l}.$$

Since we have chosen $\beta' < \beta - d/2$, the latter series is convergent, implying that the right hand side vanishes if $\lambda \to \infty$.

For the second term in (2.66), we can deduce directly from Lemma 2.7 that for any fixed λ we have $d(\tilde{\Pi}^{\mathcal{D}_\varepsilon}_\lambda(\cdot|Y^\varepsilon), \mu_\lambda) \to 0$ in $P^Y_{f_0}$-probability as $\varepsilon \to 0$. For the third term we proceed similarly to the first, obtaining

$$d(\mu, \mu_\lambda) = E \sup_{\psi \in B^\beta_K(1)} |X(\psi - P_\lambda \psi)| \lesssim \sum_{l>\lambda} 2^{-\beta l} \sqrt{\sum_{r=1}^{N_l} E|X(\Phi_{lr})|^2}.$$

Recall $X(\Phi_{lr}) \sim N(0, \|L\Phi_{lr}\|^2_{L^2})$. Hence

$$d(\mu, \mu_\lambda) \lesssim \sum_{l>\lambda} 2^{-\beta l} 2^{2l} \sqrt{N_l} \le \sum_{l>\lambda} 2^{-(\beta-2-d/2)l}$$

which again converges since $\beta > 2 + d$. To conclude, we can fix arbitrarily $\varepsilon' > 0$, and then find $\lambda = \lambda(\varepsilon')$ sufficiently large so that the first and third term in (2.66) are smaller than ε'. For such value of λ, the second term can be made smaller that ε' by choosing ε small enough with $P^Y_{f_0}$-probability approaching one, implying that $d(\tilde{\Pi}^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon), \mu) \to 0$ in $P^Y_{f_0}$-probability as $\varepsilon \to 0$.

Appendix 2.A Information lower bound for linear inverse problems

Let Y^ε be given by (2.4) with $f = f_0$, and let $\ell_\varepsilon(f) = \log p_f(Y^\varepsilon)$, $f \in \overline{\mathcal{W}}$, be the log-likelihood in (2.6). For any $h \in \overline{\mathcal{W}}$, $\varepsilon > 0$, we have

$$\log \frac{p_{f_0+\varepsilon h}(Y^\varepsilon)}{p_{f_0}(Y^\varepsilon)} = \ell_\varepsilon(f_0 + \varepsilon h) - \ell_\varepsilon(f_0) = \langle Gh, \mathbb{W}\rangle_{\mathcal{W}_2} - \frac{1}{2}\|Gh\|^2_{\mathcal{W}_2}.$$

Recalling $\langle Gh, \mathbb{W}\rangle_{\mathcal{W}_2} \sim N(0, \|Gh\|^2_{\mathcal{W}_2})$, the model is seen to be locally (asymptotically) normal (LAN), with LAN-inner product and norm respectively given by

$$\langle \cdot, \cdot \rangle_{LAN} = \langle G\cdot, G\cdot \rangle_{\mathcal{W}_2}, \qquad \| \cdot \|_{LAN} = \|G \cdot \|_{\mathcal{W}_2}.$$

Let $\psi \in \mathcal{W}_1$ satisfy the assumptions of Theorem 2.1, and consider the continuous linear map

$$L_\psi : \overline{\mathcal{W}} \to \mathbb{R}, \qquad L_\psi(h) = \langle h, \psi \rangle_{\mathcal{W}_1},$$

defined by extension using the fact that $\mathcal{W}_1 \subseteq \overline{\mathcal{W}}$ is dense. As by assumption $\psi = -G^*G\tilde{\psi}$ for some $\tilde{\psi} \in \mathcal{H}$, then for all $h \in \overline{\mathcal{W}}$

$$L_\psi(h) = \langle h, \psi \rangle_{\mathcal{W}_1} = \langle h, -G^*G\tilde{\psi} \rangle_{\mathcal{W}_1} = \langle h, -\tilde{\psi} \rangle_{LAN},$$

so that the Riesz representer with respect to the LAN-inner product of the linear functional L_ψ is $-\tilde{\psi}$. We then deduce from the semiparametric theory of efficiency (see Chapter 25 in [234], or Section 7.5 in [171]) that the information lower bound for estimating $L_\psi(f_0) = \langle f_0, \psi \rangle_{\mathcal{W}_1}$ from model (2.4) is identified by the random variable

$$Z \sim N(0, \|\tilde{\psi}\|_{LAN}) = \mathcal{N}0, \|G\tilde{\psi}\|_{\mathcal{W}_1}),$$

and we have the lower bound (2.14) for the asymptotic minimal variance . Note that when G^*G has a well defined inverse we can write $\|G\tilde{\psi}\|^2_{\mathcal{W}_2} = \|(G^*G)^{-1}\psi\|^2_{LAN}$. In analogy with the finite-dimensional case, we then sometimes call G^*G the Fisher information operator.

Appendix 2.B Properties of elliptic boundary value problems

We list here some key facts relative to the boundary value problem (2.19) following from the general elliptic theory (see, e.g., the monographs [153, 205]). We start noting that the operator L defines a bounded linear operator from $H^s = H^s(\mathcal{O})$ into H^{s-2} for all $s \geq 2$, and, in view of the symmetry of the coefficients a_{ij}, it is also self-adjoint with respect to $\langle \cdot, \cdot \rangle_{L^2}$ when acting upon H^2_0.

If, in addition, we assume the uniform ellipticity condition:

$$\sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geq a|\xi|^2, \qquad \forall x \in \mathcal{O}, \quad \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d,$$

for some constant $a > 0$, then for all $s \geq 0$ and any $f \in H^s$, there exists a unique weak solution $u_f \in H_0^{s+2}$ of (2.19) satisfying the variational formulation of the problem:

$$\int_{\mathcal{O}} \sum_{i,j=1}^{d} a_{ij} \frac{\partial u_f}{\partial x_i} \frac{\partial v}{\partial x_j} = \int_{\mathcal{O}} fv, \qquad \forall v \in H_0^1. \tag{2.67}$$

Furthermore, we have the elliptic estimates

$$\|u_f\|_{H^{s+2}} \leq c_s \|f\|_{H^s},$$

for constants $c_s > 0$ depending only on s. These results follow directly from Theorem 5.4 in [153, Chapter 2] (see also remark 7.2 in the same reference) by noting that $u_f = 0$ is the unique smooth solution of (2.19) with $f = 0$ (e.g., in view of Theorem 3 and Theorem 4 in [89, Section 6.3]). Finally, as pointed out in Remark (ii), page 310 in [89], it follows that $Lu_f = f$ almost everywhere on $\mathcal{O}$.

With a slight abuse of notation, let L^{-1} denote the solution map, so that $L^{-1}f = u_f$ is the unique element in H_0^{s+2} that satisfies (2.67). Also, in view of the uniqueness of weak solutions, $L^{-1}Lu = u$ for all $u \in H_0^2$. From the above results we have that $L^{-1} : H^s \to H_0^{s+2}$, $s \geq 0$, defines a linear and bounded isomorphism which is self-adjoint on L^2 (following from the self-adjointness of L).

Appendix 2.C Remaining proofs

2.C.1 Proof of Corollary 2.2

The proof follows the argument in Section 2 of [52]. We start by noting that the function $\Phi : [0, \infty) \to [0, 1]$ is uniformly continuous and strictly increasing, with continuous inverse $\Phi^{-1} : [0, 1] \to [0, \infty)$. Thus, for every $\gamma > 0$ we can find $\varepsilon > 0$ such that $|\Phi(t+\varepsilon) - \Phi(t)| \leq \gamma$, $\forall t \geq 0$. For such ε and for all $t \geq 0$, denoting $B(0, t) = \{x \in \mathbb{R}, \ |x| \leq t\}$, we have

$$\Pr(t - \varepsilon < |Z| \leq t + \varepsilon) = \Phi(t + \varepsilon) - \Phi(t - \varepsilon) \leq 2\gamma.$$

Thus, applying Corollary 7.3.22 in [104] to $\mathcal{L}(\varepsilon^{-1}\langle f - \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}|Y^\varepsilon)$ converging weakly to $\mathcal{L}(Z)$ in $P^Y_{f_0}$-probability as $\varepsilon \to 0$, we deduce that

$$\sup_{0\le t<\infty}\left|\Pi(\varepsilon^{-1}\langle f - \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1} \in B(0,t)|Y^\varepsilon) - \Pr(Z \in B(0,t))\right| = o_{P^{Y^\varepsilon}_{f_0}}(1)$$

as $\varepsilon \to 0$. Thus, recalling the definition of R_ε after (2.15)

$$\begin{aligned}\Phi(\varepsilon^{-1}R_\varepsilon) &= \Pr(|Z| \le \varepsilon^{-1}R_\varepsilon) - \Pi(\varepsilon^{-1}|\langle f - \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}| \le \varepsilon^{-1}R_\varepsilon|Y^\varepsilon) + 1 - \alpha \\ &= 1 - \alpha + o_{P^{Y^\varepsilon}_{f_0}}(1)\end{aligned}$$

as $\varepsilon \to 0$ by the above with $t = \varepsilon^{-1}R_\varepsilon$. Since Φ^{-1} is continuous, the continuous mapping theorem yields

$$\varepsilon^{-1}R_\varepsilon = \Phi^{-1}[\Phi(\varepsilon^{-1}R_\varepsilon)] \xrightarrow{P^Y_{f_0}} \Phi^{-1}(1-\alpha).$$

Then the first claim follows using Theorem 2.1, as

$$\begin{aligned}P^{Y^\varepsilon}_{f_0}(\langle f_0, \psi\rangle_{\mathcal{W}_1} \in C_\varepsilon) &= P^{Y^\varepsilon}_{f_0}(\varepsilon^{-1}|\langle f_0 - \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}| \le \varepsilon^{-1}R_\varepsilon) \\ &= P^{Y^\varepsilon}_{f_0}\left(\varepsilon^{-1}|\langle f_0 - \bar{f}_\varepsilon, \psi\rangle_{\mathcal{W}_1}| \le \Phi^{-1}(1-\alpha)\right) + o(1) \\ &= \Pr\left(|Z| \le \Phi^{-1}(1-\alpha)\right) + o(1) \\ &= 1 - \alpha + o(1).\end{aligned}$$

2.C.2 Proof of Lemma 2.1

1. First assume that $\beta > 2 + d/2$, and let $B^\beta_c(1) := \{h \in H^\beta_c,\ \|h\|_{H^\beta} \le 1\}$. According to (2.28), $X_{|B^\beta_c(1)} := (X(\psi) : \psi \in B^\beta_c(1))$ is a Gaussian process with intrinsic distance

$$d^2_X(\psi, \psi') := E[X(\psi) - X(\psi')]^2 \lesssim \|\psi - \psi'\|^2_{H^2}.$$

Next, from Edmund and Triebel's upper bound for the entropy numbers in general Besov spaces (see [227]) we deduce that, for positive reals $s_1 < s_2$, letting $B^s(r) := \{h \in H^s,\ \|h\|_{H^s} \le r\},\ r > 0$,

$$\log N(\eta; B^{s_2}(1), \|\cdot\|_{H^{s_1}}) \lesssim \eta^{-\frac{d}{s_2 - s_1}}, \qquad \eta > 0. \tag{2.68}$$

Then it follows from Dudley's metric entropy inequality [104, Theorem 2.3.7] that for all $z > 0$

$$E \sup_{\psi \in B_c^\beta(1),\ \|\psi\|_{H^2} \le z} |X(\psi)| \lesssim \int_0^z \sqrt{2 \log N(\eta; B^\beta(1), \|\cdot\|_{H^2})} d\eta \lesssim \int_0^z \eta^{-\frac{d}{2(\beta-2)}} d\eta, \quad (2.69)$$

which is indeed convergent for all $\beta > 2 + d/2$. Thus, letting $z \to 0$ in (2.69) implies that $X_{|B_c^\beta(1)}$ has a version taking values in the separable Banach space

$$(\mathbb{B}, \|\cdot\|_{\mathbb{B}}),\ \mathbb{B} = UC(B_c^\beta(1), d_X), \qquad \|x\|_{\mathbb{B}} = \sup_{\psi \in B_c^\beta(1)} |x(\psi)|, \quad (2.70)$$

of bounded and uniformly continuous (with respect to the metric d_X on $B_c^\beta(1)$) pre-linear functionals on $B_c^\beta(1)$, the separability following from Corollary 11.2.5 in [79] since, in view of (2.68), $B_c^\beta(1)$ is totally bounded for the metric d_X if $\beta > 2$. Finally, as according to (2.70) $\mathbb{B}$ is an isometrically imbedded closed subspace of $(H_c^\beta)^*$, we deduce from Oxtoby-Ulam theorem (Proposition 2.1.4 in [104]) that $X_{|B_c^\beta(1)}$ induces a tight Borel Gaussian probability measure on $\mathbb{B}$, which has a unique extension to $(H_c^\beta)^*$.

2. For $\beta < 2 + d/2$, as $H_c^\beta \subset H_c^{\beta'}$ with continuous embedding if $\beta' < \beta$, it is enough to show that $\Pr\left(\sup_{\psi \in B_c^\beta(1)} |X(\psi)| < \infty\right) = 0$ for $2 < \beta < 2 + d/2$. We proceed by contradiction, assuming on the contrary that

$$\Pr\left(\sup_{\psi \in B_c^\beta(1)} |X(\psi)| < \infty\right) > 0. \quad (2.71)$$

In view of (2.68), $B_c^\beta(1)$ is separable with respect to the intrinsic metric d_X for any $\beta > 2$. Hence, Proposition 2.1.12 in [104] and (2.71) jointly imply, by Proposition 2.1.20 in [104], that $E \sup_{\psi \in B_c^\beta(1)} |X(\psi)| < \infty$, which we will show to yield a contradiction. To do so, note that $(X(L^{-1}\psi) : \psi \in H_c^\beta)$ has the same law on $\mathbb{R}^{H_c^\beta}$ as the standard Gaussian white noise $\mathbb{W}$. Thus,

$$E \sup_{\psi \in B_c^\beta(1)} |X(\psi)| = E \sup_{\psi \in B_c^\beta(1)} |\langle L\psi, \mathbb{W}\rangle_{L^2}|,$$

and the proof is completed by finding a suitable lower bound to show that the right hand side diverges.

Considering the orthonormal Daubechies wavelet basis of L^2 introduced in Step IV in Section 2.5.3, select for each $j \ge 1$, $n_j = c'2^{jd}$, $c' > 0$, wavelets $\{\Phi_{jr},\ r = 1, \dots, n_j\}$ with disjoint compact support within $\mathcal{O}$. Next, for each $m = 1, \dots, 2^{n_j}$ and $b_{m\cdot} = (b_{mr},\ r =$

$1, \dots, n_j) \in \{-1, 1\}^{n_j}$, define

$$h_{jm}(x) = k_j \sum_{r=1}^{n_j} b_{mr} 2^{-j\beta} \Phi_{jr}(x), \qquad x \in \mathcal{O}, \tag{2.72}$$

where $k_j > 0$ is to be fixed. Recall that it is enough to consider $2 < \beta < 2 + d/2$. We have $h_{jm} \in H_c^\beta$, and by the usual wavelet characterisation of the Sobolev norms

$$\|h_{jm}\|_{H^\beta}^2 \simeq \sum_{l \ge -1} \sum_{s=1}^{n_j} 2^{2l\beta} \langle h_{jm}, \Phi_{ls} \rangle_{L^2}^2 = k_j^2 n_j.$$

Hence, choosing $k_j < n_j^{-1/2}$ guarantees that $\{h_{jm}, \ m = 1, \dots, 2^{n_j}\} \subset B_c^\beta(1)$, yielding the lower bound

$$E \sup_{\psi \in B_c^\beta(1)} |\langle L\psi, \mathbb{W} \rangle_{L^2}| \ge E \max_{m=1,\dots,2^{n_j}} |\langle L h_{mj}, \mathbb{W} \rangle_{L^2}|, \qquad j \ge 1,$$

which we can further bound from below by restricting the maximum to a suitable smaller subset. In particular, the Gaussian vector $(\mathbb{W}(L h_{jm}), \ m = 1, \dots, 2^{n_j})$ has intrinsic metric

$$d_j^2(h_{jm}, h_{jm'}) = \|L(h_{jm} - h_{jm'})\|_{L^2}^2 = k_j^2 2^{-2j\beta} \left\| \sum_{r=1}^{n_j} (b_{mr} - b_{m'r}) L \Phi_{jr} \right\|_{L^2}^2;$$

and arguing as in the proof of Proposition 6 in [171] we can select, for sufficiently large j, a subset $\{h_{j1}, \dots, h_{jm_j}\} \subseteq \{h_{jm}, \ m = 1, \dots, 2^{n_j}\}$ of cardinality $m_j \ge 3^{n_j/4}$, such that

$$d_j^2(h_{jh}, h_{jk}) \gtrsim 2^{2j(2-\beta)}, \qquad h \ne k.$$

Thus, by applying Sudakov's lower bound [104, Theorem 2.4.12], we deduce that for all such j

$$\begin{aligned} E \max_{m=1,\dots,2^{n_j}} |\langle L h_{jm}, \mathbb{W} \rangle_{L^2}| &\ge E \max_{h=1,\dots,m_j} \{|\langle L h_{jh}, \mathbb{W} \rangle_{L^2}|\} \\ &\ge c 2^{j(2-\beta)} \sqrt{\log N(2^{j(2-\beta)}; \{h_{j1}, \dots, h_{jm_j}\}, d_j)} \\ &\ge c' 2^{j(2-\beta)} \sqrt{\log m_j} \\ &\ge c'' 2^{j(2+d/2-\beta)}. \end{aligned}$$

The last line diverges as $j \to \infty$ for all $\beta < 2 + d/2$, yielding the contradiction.

3. Assuming tightness on $(H_c^\beta)^*$ for $\beta = 2 + d/2$ would imply (exactly as above) that X were sample bounded and, in view of Proposition 2.1.7 in [104], also sample continuous with respect to d_X. Then, Proposition 2.4.14 in [104] would yield

$$\lim_{\eta \to 0} \eta \sqrt{\log N(\eta; B_c^\beta(1), d)} = 0$$

which, taking the sequence $\eta_j = 2^{j(2-\beta)} = 2^{-jd/2}$, is in contrast with the fact that

$$2^{j(2-\beta)} \sqrt{\log N(2^{j(2-\beta)}; B_c^\beta(1), d)} \geq 2^{j(2-\beta)} \sqrt{\log N(2^{j(2-\beta)}; \{h_{j1}, \dots, h_{jm_j}\}, d_j)},$$

and that the right hand side is bounded below by a positive constant for $\beta = 2 + d/2$, as seen above. □

2.C.3 Proof of supporting lemmas for Theorem 2.1

Proof of Lemma 2.2

We start by noting that $\Pi(B) = \Pi(B \cap \mathcal{D}_\varepsilon) + \Pi(B \cap \mathcal{D}_\varepsilon^c)$ and

$$\Pi(B \cap \mathcal{D}_\varepsilon) - \Pi^{\mathcal{D}_\varepsilon}(B) = \frac{\Pi(B \cap \mathcal{D}_\varepsilon)}{\Pi(\overline{\mathcal{W}})} - \frac{\Pi(B \cap \mathcal{D}_\varepsilon)}{\Pi(\mathcal{D}_\varepsilon)} = -\Pi(\mathcal{D}_\varepsilon^c)\Pi^{\mathcal{D}_\varepsilon}(B)$$

which implies $\|\Pi(\cdot|Y^\varepsilon) - \Pi^{\mathcal{D}_\varepsilon}(\cdot|Y^\varepsilon)\|_{TV} \leq 2\Pi(\mathcal{D}_\varepsilon^c|Y^\varepsilon)$. Hence it suffices to prove the first limit in (2.34). This will be done using Markov's inequality and showing that $E_{f_0}^{Y^\varepsilon}(\Pi(\mathcal{D}_\varepsilon^c|Y^\varepsilon)) \to 0$. In particular, we split the expectation as

$$E_{f_0}^{Y^\varepsilon}(\Pi(\mathcal{D}_\varepsilon^c|Y^\varepsilon)) = E_{f_0}^{Y^\varepsilon}(\Pi(\mathcal{D}_\varepsilon^c|Y^\varepsilon)1_{\mathcal{F}_\varepsilon}) + E_{f_0}^{Y^\varepsilon}(\Pi(\mathcal{D}_\varepsilon^c|Y^\varepsilon)1_{\mathcal{F}_\varepsilon^c}) \tag{2.73}$$

where $\mathcal{F}_\varepsilon$ is a suitable event to be specified for which $P_{f_0}^{Y^\varepsilon}(\mathcal{F}_\varepsilon) \to 0$, yielding the cancellation of the first term, at a sufficiently slow rate so that also the second vanishes due to the assumption on $\Pi(\mathcal{D}_\varepsilon^c)$.

We proceed constructing $\mathcal{F}_\varepsilon$. For $\ell_\varepsilon(f)$ the log-likelihood defined in (2.6), we can rewrite the posterior (2.7) as

$$\Pi(B|Y^\varepsilon) = \frac{\int_B e^{\ell_\varepsilon(f) - \ell_\varepsilon(f_0)} d\Pi(f)}{\int_{\overline{\mathcal{W}}} e^{\ell_\varepsilon(f) - \ell_\varepsilon(f_0)} d\Pi(f)}, \qquad B \in \mathcal{B}_{\overline{\mathcal{W}}}. \tag{2.74}$$

It follows from (2.6) that under $P_{f_0}^{Y^\varepsilon}$ we have

$$\ell_\varepsilon(f) - \ell_\varepsilon(f_0) = \frac{1}{\varepsilon}\langle G(f-f_0), \mathbb{W}\rangle_{\mathcal{W}_2} - \frac{1}{2\varepsilon^2}\|G(f-f_0)\|_{\mathcal{W}_2}^2.$$

Let ν be any probability measure on the set $B = \{f : \|G(f-f_0)\| \leq \delta_\varepsilon\}$. Applying Jensen's inequality to the exponential function we get for any $\tilde{C} > -1/2$

$$\begin{aligned} &P_{f_0}^{Y^\varepsilon}\left(\int_B e^{\ell_\varepsilon(f)-\ell_\varepsilon(f_0)} d\nu(f) \leq e^{-(1+\widetilde{C})(\delta_\varepsilon/\varepsilon)^2}\right) \\ &\quad \leq \Pr\left(E^\nu\left(\frac{1}{\varepsilon}\langle G(f-f_0), \mathbb{W}\rangle_{\mathcal{W}_2} - \frac{1}{2\varepsilon^2}\|G(f-f_0)\|_{\mathcal{W}_2}^2\right) \leq -(1+\widetilde{C})\,(\delta_\varepsilon/\varepsilon)^2\right). \end{aligned}$$

Denote $Z = \frac{1}{\varepsilon}\int_B \langle G(f-f_0), \mathbb{W}\rangle_{\mathcal{W}_2} d\nu(f) \sim N(0, C_Z)$ where, using again Jensen's inequality,

$$\begin{aligned} C_Z &= \frac{1}{\varepsilon^2} E\left(E^\nu \langle G(f-f_0), \mathbb{W}\rangle_{\mathcal{W}_2}\right)^2 \\ &\leq \frac{1}{\varepsilon^2} E^\nu\left(E\langle G(f-f_0), \mathbb{W}\rangle_{\mathcal{W}_2}^2\right) \\ &= \frac{1}{\varepsilon^2}\int_B \|G(f-f_0)\|_{\mathcal{W}_2}^2 d\nu(f) \\ &\leq (\delta_\varepsilon/\varepsilon)^2. \end{aligned}$$

We can then conclude

$$\begin{aligned} P_{f_0}^{Y^\varepsilon}\left(\int_B e^{\ell_\varepsilon(f)-\ell_\varepsilon(f_0)} d\nu(f) \leq e^{-(1+\widetilde{C})(\delta_\varepsilon/\varepsilon)^2}\right) &= \Pr\left(|Z - EZ| \geq \left(\frac{1}{2} + \widetilde{C}\right)(\delta_\varepsilon/\varepsilon)^2\right) \\ &\leq e^{-\frac{(1/2+\tilde{C})^2}{2}(\delta_\varepsilon/\varepsilon)^2} \end{aligned}$$

the last inequality following from the standard Gaussian tail bound $\Pr(|Z-EZ| \geq c) \leq e^{-c^2/(2Var(Z))}$. We can now choose $\nu = \Pi(\cdot \cap B)/\Pi(B)$ and let

$$\mathcal{F}_\varepsilon = \left\{f : \int_B e^{\ell_\varepsilon(f)-\ell_\varepsilon(f_0)} d\nu(f) \leq e^{-\frac{3}{2}(\delta_\varepsilon/\varepsilon)^2}\right\}.$$

Using the above with $\tilde{C} = 1/2$ we see that $P_{f_0}^{Y^\varepsilon}(\mathcal{F}_\varepsilon) \leq e^{-\frac{1}{2}(\delta_\varepsilon/\varepsilon)^2} \to 0$, which implies that the first term in (2.73) tends to zero since $\Pi(\cdot|Y^\varepsilon) \leq 1$.

For the second term study the small ball probabilities $\Pi(B) = \Pi(f : \|G(f-f_0)\|_{\mathcal{W}_2} \leq \delta_\varepsilon)$ using the condition (2.9) on the concentration function. We see from (2.74) that

$$E_{f_0}^{Y^\varepsilon}(\Pi(\mathcal{D}_\varepsilon^c|Y^\varepsilon)1_{\mathcal{F}_\varepsilon^c}) \le E_{f_0}^{Y^\varepsilon}\left(\frac{\int_{\mathcal{D}_\varepsilon^c} e^{\ell_\varepsilon(f)-\ell_\varepsilon(f_0)}d\Pi(f)}{\int_B e^{\ell_\varepsilon(f)-\ell_\varepsilon(f_0)}\Pi(B)d\nu(f)}\,1_{\mathcal{F}_\varepsilon^c}\right)$$
$$\le \frac{e^{2(\delta_\varepsilon/\varepsilon)^2}}{\Pi(f:\|G(f-f_0)\|_{\mathcal{W}_2}^2 \le \delta_\varepsilon^2)}\int_{\mathcal{D}_\varepsilon^c} E_{f_0}^{Y^\varepsilon}\left(e^{\ell_\varepsilon(f)-\ell_\varepsilon(f_0)}\right)d\Pi(f).$$

For $f\sim\Pi$, denote the concentration function of the (image) Gaussian measure $\tilde{\Pi}=\mathcal{L}(Gf)$ as

$$\widetilde{\phi}_{\Pi,f_0}(\delta) = \inf_{Gg\in\tilde{\mathcal{H}},\ \|G(g-f_0)\|_{\mathcal{W}_2}\le\delta} \frac{\|Gg\|_{\tilde{\mathcal{H}}}^2}{2} - \log\Pi(f:\|Gf\|_{\mathcal{W}_2}\le\delta).$$

Following Proposition 2.6.19 in [104] we next show that

$$\Pi(f:\|G(f-f_0)\|_{\mathcal{W}_2}^2\le\delta^2) \ge e^{-\widetilde{\phi}_{\Pi,f_0}(\delta/2)}.$$

Let $g\in\mathcal{H}$ be such that $\|G(g-f_0)\|_{\mathcal{W}_2}\le\delta/2$. Then $\|G(f-f_0)\|_{\mathcal{W}_2}\le\|G(f-g)\|_{\mathcal{W}_2}+\delta/2$. We denote $\Pi_g(B)=\Pi(B-g)=\Pi(f:f+g\in B)$. Using the Cameron-Martin theorem [38, Corollary 2.4.3.] and the fact that f is a centred Gaussian random variable we can write

$$\begin{aligned}
&\Pi(f:\|G(f-f_0)\|_{\mathcal{W}_2}\le\delta)\\
&\quad\ge \Pi(f:\|G(f-g)\|_{\mathcal{W}_2}\le\delta/2)\\
&\quad= \frac{1}{2}\left(\Pi_{-Gg}\left(f:\|Gf\|_{\mathcal{W}_2}\le\frac{\delta}{2}\right)+\Pi_{Gg}\left(f:\|Gf\|_{\mathcal{W}_2}\le\frac{\delta}{2}\right)\right)\\
&\quad= \frac{1}{2}\left(\int_{\{\|\widetilde{f}\|_{\mathcal{W}_2}\le\frac{\delta}{2}\}}\frac{d\widetilde{\Pi}_{-\widetilde{g}}(\widetilde{f})}{d\widetilde{\Pi}(\widetilde{f})}d\widetilde{\Pi}(\widetilde{f})+\int_{\{\|\widetilde{f}\|_{\mathcal{W}_2}\le\frac{\delta}{2}\}}\frac{d\widetilde{\Pi}_{\widetilde{g}}(\widetilde{f})}{d\widetilde{\Pi}(\widetilde{f})}d\widetilde{\Pi}(\widetilde{f})\right)\\
&\quad= \frac{1}{2}\left(\int_{\{\|\widetilde{f}\|_{\mathcal{W}_2}\le\frac{\delta}{2}\}}\left(e^{-\langle\widetilde{g},\widetilde{f}\rangle_{\tilde{\mathcal{H}}}}+e^{\langle\widetilde{g},\widetilde{f}\rangle_{\tilde{\mathcal{H}}}}\right)e^{-\frac{\|\widetilde{g}\|_{\tilde{\mathcal{H}}}^2}{2}}d\widetilde{\Pi}(\widetilde{f})\right)\\
&\quad\ge e^{-\frac{\|Gg\|_{\tilde{\mathcal{H}}}^2}{2}}\Pi\left(f:\|Gf\|_{\mathcal{W}_2}\le\frac{\delta}{2}\right)
\end{aligned}$$

where $Gg=\widetilde{g}$ and $Gf=\widetilde{f}$. The last inequality follows from the fact $e^{-x}+e^{x}\ge 2$ for all $x\in\mathbb{R}$. We can then conclude

$$E_{f_0}^{Y^\varepsilon}(\Pi(\mathcal{D}_\varepsilon^c|Y^\varepsilon)1_{\mathcal{F}_\varepsilon^c}) \le e^{2(\delta_\varepsilon/\varepsilon)^2}e^{\widetilde{\phi}_{\Pi,f_0}(\delta_\varepsilon/2)}\Pi(\mathcal{D}_\varepsilon^c)$$

since $E_{f_0}^{Y^\varepsilon}\left(e^{\ell_\varepsilon(f)-\ell_\varepsilon(f_0)}\right)=1$.

Note that G is assumed to be linear and injective and hence the RKHS $\tilde{\mathcal{H}}=G(\mathcal{H})$ of Gf is isometric to $\mathcal{H}$ (see Exercise 2.6.5 in [104]). By assumption, $\|Gf\|_{\mathcal{W}_2}\le c\|f\|_{\overline{\mathcal{W}}}$ for all $f\in\overline{\mathcal{W}}$, which implies

$$-\log\Pi(f:\|Gf\|_{\mathcal{W}_2}\le\delta)\le-\log\Pi\left(f:\|f\|_{\overline{\mathcal{W}}}\le\delta/c\right).$$

We also have $\|G(g-f_0)\|_{\mathcal{W}_2}\le c\|g-f_0\|_{\overline{\mathcal{W}}}$ and hence $\widetilde{\phi}_{\Pi,f_0}(\delta)\le\phi_{\Pi,f_0}(\delta/c)$ for all δ. Thus by (2.9) and assumption (2.33) we can conclude

$$E_{f_0}^{Y^\varepsilon}(\Pi(\mathcal{D}_\varepsilon^c|Y^\varepsilon)1_{\mathcal{F}_\varepsilon^c})\le e^{2(\delta_\varepsilon/\varepsilon)^2}e^{\phi_{\Pi,f_0}(\delta_\varepsilon/2c)}\Pi(\mathcal{D}_\varepsilon^c)\le e^{(3-D)(\delta_\varepsilon/\varepsilon)^2}\to 0.$$

□

Proof of Lemma 2.3

Denote $f_\tau=f+\tau\varepsilon\tilde{\psi}$. Then the left hand side of (2.36) can be written as

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau}{\varepsilon}\langle f-f_0,\psi\rangle_{\mathcal{W}_1}+\tau\langle G\tilde{\psi},\mathbb{W}\rangle_{\mathcal{W}_2}}\Big|Y^\varepsilon\right]$$
$$=\frac{\int_{\overline{\mathcal{W}}}e^{\frac{\tau}{\varepsilon}\langle f-f_0,\psi\rangle_{\mathcal{W}_1}+\tau\langle G\tilde{\psi},\mathbb{W}\rangle_{\mathcal{W}_2}+\ell_\varepsilon(f_\tau)-\ell_\varepsilon(f_\tau)+\ell_\varepsilon(f)}d\Pi^{\mathcal{D}_\varepsilon}(f)}{\int_{\overline{\mathcal{W}}}e^{\ell_\varepsilon(f)}d\Pi^{\mathcal{D}_\varepsilon}(f)}.$$

Using (2.6) we see that under $P_{f_0}^{Y^\varepsilon}$

$$\ell_\varepsilon(f)-\ell_\varepsilon(f_\tau)=\frac{\tau}{\varepsilon}\langle G(f-f_0),G\tilde{\psi}\rangle_{\mathcal{W}_2}+\frac{\tau^2}{2}\|G\tilde{\psi}\|_{\mathcal{W}_2}^2-\tau\langle G\tilde{\psi},\mathbb{W}\rangle_{\mathcal{W}_2}$$

and hence

$$E^{\Pi^{\mathcal{D}_\varepsilon}}\left[e^{\frac{\tau}{\varepsilon}(\langle f,\psi\rangle_{\mathcal{W}_1}-\widehat{\Psi})}\Big|Y^\varepsilon\right]=e^{\frac{\tau^2}{2}\|G\tilde{\psi}\|_{\mathcal{W}_2}^2}\frac{\int_{\mathcal{D}_\varepsilon}e^{\ell_\varepsilon(f_\tau)}d\Pi(f)}{\int_{\mathcal{D}_\varepsilon}e^{\ell_\varepsilon(f)}d\Pi(f)}.\tag{2.75}$$

Let Π_τ be the shifted law of f_τ, $f\sim\Pi$. Then by the Cameron-Martin theorem [38, Corollary 2.4.3.] we get, denoting $\mathcal{D}_{\varepsilon,\tau}=\{g=f_\tau:f\in\mathcal{D}_\varepsilon\}$,

$$\frac{\int_{\mathcal{D}_{\varepsilon,\tau}}e^{\ell_\varepsilon(g)}\frac{d\Pi_\tau(g)}{d\Pi(g)}d\Pi(g)}{\int_{\mathcal{D}_\varepsilon}e^{\ell_\varepsilon(g)}d\Pi(g)}=\frac{\int_{\mathcal{D}_{\varepsilon,\tau}}e^{\ell_\varepsilon(g)}e^{\tau\varepsilon\langle\tilde{\psi},g\rangle_{\mathcal{H}}-\frac{(\tau\varepsilon)^2}{2}\|\tilde{\psi}\|_{\mathcal{H}}^2}d\Pi(g)}{\int_{\mathcal{D}_\varepsilon}e^{\ell_\varepsilon(g)}d\Pi(g)}.\tag{2.76}$$

Since $\tilde{\psi}$ is a fixed element in $\mathcal{H}$, we see that $\varepsilon^2\|\tilde{\psi}\|_{\mathcal{H}}^2 \to 0$ as $\varepsilon \to 0$. Using the definition of $\mathcal{D}_\varepsilon$ in (2.35) we see, as $\varepsilon \to 0$,

$$\varepsilon \sup_{g\in\mathcal{D}_{\varepsilon,\tau}} |\langle\tilde{\psi}, g\rangle_{\mathcal{H}}| = \varepsilon \sup_{f\in\mathcal{D}_\varepsilon} |\langle\tilde{\psi}, f + \tau\varepsilon\tilde{\psi}\rangle_{\mathcal{H}}| \leq T\delta_\varepsilon\|\tilde{\psi}\|_{\mathcal{H}} + |\tau|\varepsilon^2\|\tilde{\psi}\|_{\mathcal{H}}^2 \to 0.$$

We have thus shown that a small shift of f along $\mathcal{H}$ in (2.75) correspond asymptotically to a shift in $\mathcal{D}_\varepsilon$:

$$\frac{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f_\tau)} d\Pi(f)}{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(f)} d\Pi(f)} = \frac{\int_{\mathcal{D}_{\varepsilon,\tau}} e^{\ell_\varepsilon(g)} d\Pi(g)}{\int_{\mathcal{D}_\varepsilon} e^{\ell_\varepsilon(g)} d\Pi(g)}(1+o(1)) = \frac{\Pi(\mathcal{D}_{\varepsilon,\tau}|Y^\varepsilon)}{\Pi(\mathcal{D}_\varepsilon|Y^\varepsilon)}(1+o(1)).$$

Using Lemma 2.2 we see that $\Pi(\mathcal{D}_\varepsilon|Y^\varepsilon) \to 1$ in $P_{f_0}^Y$-probability. We also note that

$$\begin{aligned}\Pi(D_{\varepsilon,\tau}^c) &= \Pi\left(g : \frac{|\langle\tilde{\psi}, g - \tau\varepsilon\tilde{\psi}\rangle_{\mathcal{H}}|}{\|\tilde{\psi}\|_{\mathcal{H}}} > \frac{T\delta_\varepsilon}{\varepsilon}\right) \\ &\leq \Pi\left(g : \frac{|\langle\tilde{\psi}, g\rangle_{\mathcal{H}}|}{\|\tilde{\psi}\|_{\mathcal{H}}} > \frac{T\delta_\varepsilon}{\varepsilon} - |\tau|\varepsilon\|\tilde{\psi}\|_{\mathcal{H}}^2\right) \\ &\leq e^{-\frac{t^2}{2}(\delta_\varepsilon/\varepsilon)^2},\end{aligned}$$

for any $\sqrt{6} < t < T$. Using Lemma 2.2 again we then conclude that $\Pi(\mathcal{D}_{\varepsilon,\tau}|Y^\varepsilon) \to 1$ in $P_{f_0}^Y$-probability.

2.C.4 Step V in the proof of Theorem 2.3: convergence of the moments

The last step consists in replacing, in the result derived in the previous step, the centring $\hat{\Psi}_\varepsilon$ (defined in (2.61)) with the posterior mean $\bar{f}_\varepsilon = E^\Pi[f|Y^\varepsilon]$. The proof only requires minor adjustments from the proof of Theorem 2.7 in [166]. In particular, we show that as $\varepsilon \to 0$,

$$\|\varepsilon^{-1}(\bar{f}_\varepsilon - \hat{\Psi}_\varepsilon)\|_{(H_K^\beta)^*} = \|E^\Pi[\varepsilon^{-1}(f - \hat{\Psi}_\varepsilon)|Y^\varepsilon]\|_{(H_K^\beta)^*} = o_{P_{f_0}^{Y^\varepsilon}}(1). \tag{2.77}$$

W argue by contradiction: let $(\Omega, \Sigma, \Pr)$ be the probability space on which $Y^\varepsilon = L^{-1}f_0 + \varepsilon\mathbb{W}$ is defined, and assume that for some $\Omega' \in \Sigma$, $\Pr(\Omega') > 0$, and $\xi > 0$, we have along a certain vanishing sequence $(\varepsilon_n)_{n\geq 1}$

$$\|E^\Pi[\varepsilon_n^{-1}(f - \hat{\Psi}_{\varepsilon_n})|Y^{\varepsilon_n}(\omega)]\|_{(H_K^\beta)^*} \geq \xi, \qquad \forall\omega \in \Omega'. \tag{2.78}$$

In view of the convergence established in Step IV and since convergence in probability implies almost sure convergence for a subsequence, we can find $\Omega_0 \in \Sigma,\ \Pr(\Omega_0) = 1$, such that along a further subsequence (denoted again as $(\varepsilon_n)_{n\geq 1}$ for convenience)

$$\beta_{(H_K^\beta)^*}(\mathcal{L}(\varepsilon_n^{-1}(f - \hat{\Psi}_{\varepsilon_n})|Y^{\varepsilon_n}(\omega)), \mu) \to 0, \qquad \forall \omega \in \Omega_0,$$

as $n \to \infty$. Thus, for each $\omega \in \Omega_0$, recalling the definition (2.63) of the process $\hat{X}_\varepsilon$ with law $\mathcal{L}(\varepsilon^{-1}(f - \hat{\Psi}_\varepsilon)|Y^\varepsilon)$ on $(H_K^\beta)^*$, the sequence $\{\hat{X}_{\varepsilon_n}(\omega),\ n \geq 1\}$ of Borel random elements in $(H_K^\beta)^*$ will convergence in distribution to the process X in (2.28). By Skorohod's embedding theorem [79, Theorem 11.7.2] we can find a probability space and random elements with values in $(H_K^\beta)^*$, $\tilde{X}_{\varepsilon_n}(\omega) \overset{d}{=} \hat{X}_{\varepsilon_n}(\omega)$, $\tilde{X} \overset{d}{=} X$, defined on it such that $\tilde{X}_{\varepsilon_n}(\omega) \xrightarrow{\text{a.s.}} \tilde{X}$, or, equivalently,

$$\|\tilde{X}_{\varepsilon_n}(\omega) - \tilde{X}\|_{(H_K^\beta)^*} \xrightarrow{\text{a.s.}} 0. \tag{2.79}$$

From the standard conjugacy property of Gaussian priors with respect to linear inverse problems with Gaussian noise, $\hat{X}_{\varepsilon_n}(\omega)$ is a Gaussian random element in $(H_K^\beta)^*$ for each $\omega \in \Omega_0,\ n \geq 1$. Then, also $\hat{X}_{\varepsilon_n}(\omega) - \tilde{X}$ is Gaussian, and by the Paley-Zygmund argument in Exercise 2.1.4 in [104], (2.79) in fact implies the convergence of all norm-moments; in particular:

$$E^\Pi \|\tilde{X}_{\varepsilon_n}(\omega) - \tilde{X}\|_{(H_K^\beta)^*} \to 0.$$

Thus, since X is a centred process, we obtain that for each $\omega \in \Omega_0$

$$\|E^\Pi[\varepsilon_n^{-1}(f - \hat{\Psi}_{\varepsilon_n})|Y^{\varepsilon_n}(\omega)]\|_{(H_K^\beta)^*} = \|E\tilde{X}_{\varepsilon_n}(\omega) - E\tilde{X}\|_{(H_K^\beta)^*} \to 0,$$

contradicting (2.78) since $\Pr(\Omega_0) = 1$.

Chapter 3

Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem

For $\mathcal{O}$ a bounded domain in $\mathbb{R}^d$ and a given smooth function $g : \mathcal{O} \to \mathbb{R}$, we consider the statistical nonlinear inverse problem of recovering the conductivity $f > 0$ in the divergence form equation

$$\begin{cases} \nabla \cdot (f \nabla u) = g, & x \in \mathcal{O}, \\ u(x) = 0, & x \in \partial\mathcal{O}, \end{cases}$$

from N discrete noisy point evaluations of the solution $u = u_f$ on $\mathcal{O}$. We study the statistical performance of Bayesian nonparametric procedures based on a flexible class of Gaussian (or hierarchical Gaussian) process priors, whose implementation is feasible by MCMC methods. We show that, as the number N of measurements increases, the resulting posterior distributions concentrate around the true parameter generating the data, and derive a convergence rate $N^{-\lambda}$, $\lambda > 0$, for the reconstruction error of the associated posterior means, in $L^2(\mathcal{O})$-distance.

3.1 Introduction

Statistical inverse problems arise naturally in many applications in physics, imaging, tomography, and generally in engineering and throughout the sciences. A prototypical example involves a domain $\mathcal{O} \subset \mathbb{R}^d$, some function $f : \mathcal{O} \to \mathbb{R}$ of interest, and indirect measurements $G(f)$ of f, where G is a given solution (or 'forward') operator of some partial differential equation (PDE) governed by the unknown coefficient f. A natural statistical observational model postulates data

$$Y_i = G(f)(X_i) + \sigma W_i, \qquad i = 1, \dots, N, \tag{3.1}$$

where the X_i's are design points at which the PDE solution $G(f)$ is measured, and where the W_i's are standard Gaussian noise variables scaled by a noise level $\sigma > 0$. The aim is then to infer f from the data $(Y_i, X_i)_{i=1}^N$. The study of problems of this type has a long history in applied mathematics, see the monographs [88, 128], although explicit *statistical* noise models have been considered only more recently [127, 33, 34, 117]. Recent survey articles on the subject are [23, 10] where many more references can be found.

For many of the most natural PDEs – such as the divergence form elliptic equation (3.2) considered below – the resulting maps G are *nonlinear* in f, and this poses various challenges: among other things, the negative log-likelihood function associated to the model (3.1), which equals the least squares criterion (see (3.10) below for details), is then possibly *non-convex*, and commonly used statistical algorithms (such as maximum likelihood estimators, Tikhonov regularisers or MAP estimates) defined as optimisers in f of likelihood-based objective functions can not reliably be computed by standard convex optimisation techniques. While iterative optimisation methods (such as Landweber iteration) may overcome such challenges [114, 192, 128, 129], an attractive alternative methodology arises from the Bayesian approach to inverse problems advocated in an influential paper by Stuart [220]: one starts from a *Gaussian process prior* Π for the parameter f or in fact, as is often necessary, for a suitable vector-space valued re-parameterisation F of f. One then uses Bayes' theorem to infer the best posterior guess for f given data $(Y_i, X_i)_{i=1}^N$. Posterior distributions and their expected values can be approximately computed via Markov Chain Monte Carlo (MCMC) methods (see, e.g., [63, 62, 26] and references therein) as soon as the forward map $G(\cdot)$ can be evaluated numerically, avoiding optimisation algorithms as well as the use of (potentially tedious, or non-existent) inversion formulas for G^{-1}; see Section 3.4.1 below for more discussion. The Bayesian approach has been particularly popular in application areas as it does not only deliver an estimator for the unknown parameter f but simultaneously provides uncertainty quantification methodology for the recovery algorithm via the probability distribution of $f|(Y_i, X_i)_{i=1}^N$ (see, e.g., [73]). Conceptually related is the area of 'probabilistic numerics' [47] in the noise-less case $\sigma = 0$, with key ideas dating back to work by Diaconis [75].

As successful as this approach may have proved to be in algorithmic practice, for the case when the forward map G is nonlinear we currently only have a limited understanding of the statistical validity of such Bayesian inversion methods. By validity we mean here *statistical guarantees* for convergence of natural Bayesian estimators such as the posterior mean $\bar{f}_N = E^{\Pi}[f|(Y_i, X_i)_{i=1}^N]$ towards the ground truth f_0 generating the data. Without such guarantees, the interpretation of posterior based inferences remains vague: the randomness of the prior may have propagated into the posterior in a way that does not 'wash out' even when very informative data is available (e.g., small noise variance and/or large sample size N), rendering Bayesian methods potentially ambiguous for the purposes of valid statistical inference and uncertainty quantification.

In the present chapter we attempt to advance our understanding of this problem area in the context of the following basic but representative example for a nonlinear inverse problem: let g be a given smooth 'source' function, and let $f : \mathcal{O} \to \mathbb{R}$ be a an unknown conductivity parameter determining solutions $u = u_f$ of the PDE

$$\begin{cases} \nabla \cdot (f \nabla u) = g, & x \in \mathcal{O}, \\ u(x) = 0, & x \in \partial\mathcal{O}, \end{cases} \tag{3.2}$$

where we denote by $\nabla \cdot$ the divergence and by ∇ the gradient operator, respectively. Under mild regularity conditions on f, and assuming that $f \geq K_{min} > 0$ on $\mathcal{O}$, standard elliptic theory implies that (3.2) has a unique classical C^2-solution $G(f) = u_f$. Identification of f from an observed solution u_f of this PDE has been considered in a large number of articles both in the applied mathematics and statistics communities – we mention here [199, 91, 116, 144, 8, 141, 123, 140, 220, 72, 215, 241, 73, 42, 26, 176, 47] and the many references therein.

The main contributions of this chapter are as follows: we show that posterior means arising from a large class of Gaussian (or conditionally Gaussian) process priors for f provide statistically consistent recovery (with explicit polynomial convergence rates as the number N of measurements increases) of the unknown parameter f in (3.2) from data in (3.1). While we employ the theory of posterior contraction from Bayesian non-parametric statistics [235, 236, 101], the nonlinear nature of the problem at hand leads to substantial additional challenges arising from the fact that a) the Hellinger distance induced by the statistical experiment is not naturally compatible with relevant distances on the actual parameter f and that b) the 'push-forward' prior induced on the information-theoretically relevant regression functions $G(f)$ is non-explicit (in particular, non-Gaussian) due to the nonlinearity of the map G. Our proofs apply recent ideas from [167] to the present elliptic situation. In the first step we show that the posterior distributions arising from the priors considered (optimally) solve the PDE-constrained regression problem of inferring $G(f)$ from data (3.1). Such results can then be combined with a suitable 'stability estimate' for the inverse map G^{-1} to show that, for large sample size N, the posterior distributions concentrate around the true parameter generating the data at a convergence rate $N^{-\lambda}$ for some $\lambda > 0$. We ultimately deduce the same rate of consistency for the posterior mean from quantitative uniform integrability arguments.

The first results we obtain apply to a large class of 're-scaled' Gaussian process priors similar to those considered in [167], addressing the need for additional a-priori regularisation of the posterior distribution in order to tame nonlinear effects of the 'forward map'. This rescaling of the Gaussian process depends on sample size N. From a non-asymptotic point of view this just reflects an adjustment of the covariance operator of the prior, but following [75] one may wonder whether a 'fully Bayesian' solution of this nonlinear inverse problem, based on a prior that does *not* depend on N, is also possible. We show indeed that a hierarchical prior that randomises a finite truncation point in the Karhunen-Loève-type series expansion of the Gaussian base prior will also result in consistent recovery of the conductivity parameter f in eq. (3.2) from data (3.1), at least if f is smooth enough.

Let us finally discuss some related literature on statistical guarantees for Bayesian inversion: to the best of our knowledge, the only previous paper concerned with (frequentist) consistency of Bayesian inversion in the elliptic PDE (3.2) is by Vollmer [241]. The proofs in [241] share a similar general idea in that they rely on a preliminary treatment of the associated regression problem for $G(f)$, which is then combined with a suitable stability estimate for G^{-1}. However, the convergence rates obtained in [241] are only implicitly given and sub-optimal, also (unlike ours) for 'prediction risk' in the PDE-constrained regression problem. Moreover, when specialised to the concrete nonlinear elliptic problem (3.2) considered here, the results in Section 4 in [241] only hold for priors with bounded C^β-norms, such as 'uniform wavelet type priors', similar to the ones used in [174, 171, 175] for different nonlinear inverse problems. In contrast, our results hold for the more practical Gaussian process priors which are commonly used in applications, and which permit the use of tailor-made MCMC methodology – such as the pCN algorithm discussed in Section 3.4.1 – for computation.

The results obtained in [176] for the maximum a posteriori (MAP) estimates associated to the priors studied here are closely related to our findings in several ways. Ultimately the proof methods in [176] are, however, based on variational methods and hence entirely different from the Bayesian ideas underlying our results. Moreover, the MAP estimates in [176] are difficult to compute due to the lack of convexity of the forward map, whereas posterior means arising from Gaussian process priors admit explicit computational guarantees, see [113] and also Section 3.4.1 for more details.

It is further of interest to compare our results to those recently obtained in [2], where the statistical version of the *Calderón problem* is studied. There the 'Dirichlet-to-Neumann map' of solutions to the PDE (3.2) is observed, corrupted by appropriate Gaussian matrix noise. In this case, as only boundary measurements of u_f at $\partial\mathcal{O}$ are available, the statistical convergence rates are only of order $\log^{-\gamma}(N)$ for some $\gamma > 0$ (as $N \to \infty$), whereas our results show that when interior measurements of u_f are available throughout $\mathcal{O}$, the recovery rates improve to $N^{-\lambda}$ for some $\lambda > 0$.

There is of course a large literature on consistency of Bayesian *linear* inverse problems with Gaussian priors, we only mention [138, 195, 6, 133, 166] and references therein. The nonlinear case considered here is fundamentally more challenging and cannot be treated by the techniques from these papers – however, some of the general theory we develop in the appendix provides novel proof methods also for the linear setting.

This chapter is structured as follows. Section 3.2 formally introduces the problem and the Bayesian approach. Section 3.3 contains all the main results for the inverse problem arising from the PDE model (3.2). The proofs, which also include some theory for general nonlinear inverse problems that is of independent interest, are given in Section 3.5 and Appendix 3.A. Finally, Appendix 3.B provides additional details on some facts used throughout the chapter.

3.2 A statistical inverse problem with elliptic PDEs

3.2.1 Main notation

Throughout the chapter, $\mathcal{O} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, is a given nonempty open and bounded set with smooth boundary $\partial\mathcal{O}$ and closure $\bar{\mathcal{O}}$.

The spaces of continuous functions defined on $\mathcal{O}$ and $\bar{\mathcal{O}}$ are respectively denoted $C(\mathcal{O})$ and $C(\bar{\mathcal{O}})$, and endowed with the supremum norm $\|\cdot\|_\infty$. For positive integers $\beta \in \mathbb{N}$, $C^\beta(\mathcal{O})$ is the space of β-times differentiable functions with uniformly continuous derivatives; for non-integer $\beta > 0$, $C^\beta(\mathcal{O})$ is defined as

$$C^\beta(\mathcal{O}) = \left\{ f \in C^{\lfloor\beta\rfloor}(\mathcal{O}) : \forall |i| = \lfloor\beta\rfloor, \sup_{x,y\in\mathcal{O}\,:\,x\neq y} \frac{|D^i f(x) - D^i f(y)|}{|x-y|^{\beta-\lfloor\beta\rfloor}} < \infty \right\},$$

where $\lfloor \beta \rfloor$ denotes the largest integer less than or equal to β, and for any multi-index $i = (i_1, \ldots, i_d)$, D^i is the i-th partial differential operator. $C^\beta(\mathcal{O})$ is normed by

$$\|f\|_{C^\beta(\mathcal{O})} = \sum_{|i| \leq \lfloor \beta \rfloor} \sup_{x \in \mathcal{O}} |D^i f(x)| + \sum_{|i| = \lfloor \beta \rfloor} \sup_{x,y \in \mathcal{O} \,:\, x \neq y} \frac{|D^i f(x) - D^i f(y)|}{|x-y|^{\beta - \lfloor \beta \rfloor}},$$

where the second summand is removed for integer β. We denote by $C^\infty(\mathcal{O}) = \cap_\beta C^\beta(\mathcal{O})$ the set of smooth functions, and by $C_c^\infty(\mathcal{O})$ the subspace of elements in $C^\infty(\mathcal{O})$ with compact support contained in $\mathcal{O}$.

Denote by $L^2(\mathcal{O})$ the Hilbert space of square integrable functions on $\mathcal{O}$, equipped with its usual inner product $\langle \cdot, \cdot \rangle_{L^2(\mathcal{O})}$. For integer $\alpha \geq 0$, the order-α Sobolev space on $\mathcal{O}$ is the separable Hilbert space

$$H^\alpha(\mathcal{O}) = \{f \in L^2(\mathcal{O}) : \forall |i| \leq \alpha,\ \exists\ D^i f \in L^2(\mathcal{O})\}, \quad \langle f, g \rangle_{H^\alpha(\mathcal{O})} = \sum_{|i| \leq \alpha} \langle D^i f, D^i g \rangle_{L^2(\mathcal{O})}.$$

For non-integer $\alpha \geq 0$, $H^\alpha(\mathcal{O})$ can be defined by interpolation, see, e.g., [153]. For any $\alpha \geq 0$, $H_0^\alpha(\mathcal{O})$ will denote the completion of $C_c^\infty(\mathcal{O})$ with respect to the norm $\|\cdot\|_{H^\alpha(\mathcal{O})}$. Finally, if K is a nonempty compact subset of $\mathcal{O}$, we denote by $H_K^\alpha(\mathcal{O})$ the closed subspace of functions in $H^\alpha(\mathcal{O})$ with support contained in K. Whenever there is no risk of confusion, we will omit the reference to the underlying domain $\mathcal{O}$.

Throughout, we use the symbols $\lesssim$ and $\gtrsim$ for inequalities holding up to a universal constant. Also, for two real sequences (a_N) and (b_N), we say that $a_N \simeq b_N$ if both $a_N \lesssim b_N$ and $b_N \lesssim a_N$ for all N large enough. For a sequence of random variables Z_N we write $Z_N = O_{\Pr}(a_N)$ if for all $\varepsilon > 0$ there exists $M_\varepsilon < \infty$ such that for all N large enough, $\Pr(|Z_N| \geq M_\varepsilon a_N) < \varepsilon$. Finally, we will denote by $\mathcal{L}(Z)$ the law of a random variable Z.

3.2.2 Parameter spaces and link functions

Let $g \in C^\infty(\mathcal{O})$ be an arbitrary source function, which will be regarded as fixed throughout. For $f \in C^\beta(\mathcal{O})$, $\beta > 1$, consider the boundary value problem

$$\begin{cases} \nabla \cdot (f \nabla u) = g, & x \in \mathcal{O}, \\ u(x) = 0, & x \in \partial \mathcal{O}. \end{cases} \tag{3.3}$$

If we assume that $f \geq K_{min} > 0$ on $\mathcal{O}$, then standard elliptic theory (e.g., [102]) implies that (3.3) has a classical solution $G(f) = u_f \in C(\bar{\mathcal{O}}) \cap C^{1+\beta}(\mathcal{O})$.

We consider the following parameter space for f: for integer $\alpha > 1+d/2$, $K_{min} \in (0,1)$, and denoting by $n = n(x)$ the outward pointing normal at $x \in \partial\mathcal{O}$, let

$$\mathcal{F}_{\alpha,K_{min}} = \left\{ f \in H^\alpha(\mathcal{O}) : \inf_{x\in\mathcal{O}} f(x) > K_{min},\ f_{|\partial\mathcal{O}} = 1,\ \frac{\partial^j f}{\partial n^j}_{|\partial\mathcal{O}} = 0 \text{ for } 1 \le j \le \alpha - 1 \right\}. \tag{3.4}$$

Our approach will be to place a prior probability measure on the unknown conductivity f and base our inference on the posterior distribution of f given noisy observations of $G(f)$, via Bayes' theorem. It is of interest to use *Gaussian process priors*. Such probability measures are naturally supported in linear spaces (in our case $H_0^\alpha(\mathcal{O})$) and we now introduce a bijective re-parametrisation so that the prior for f is supported in the relevant parameter space $\mathcal{F}_{\alpha,K_{min}}$. We follow the approach of using regular link functions Φ as in [176].

Condition 3.1. *For given $K_{min} > 0$, let $\Phi : \mathbb{R} \to (K_{min}, \infty)$ be a smooth, strictly increasing bijective function such that $\Phi(0) = 1$, $\Phi'(t) > 0$, $t \in \mathbb{R}$, and assume that all derivatives of Φ are bounded on $\mathbb{R}$.*

For some of the results to follow it will prove convenient to slightly strengthen the previous condition.

Condition 3.2. *Let Φ be as in Condition 3.1, and assume furthermore that Φ' is nondecreasing and that $\liminf_{t\to-\infty} \Phi'(t)t^a > 0$ for some $a > 0$.*

For $a = 2$, an example of such a link function is given in Example 3.1 below. Note however that the choice of $\Phi = \exp$ is not permitted in either condition.

Given any link function Φ satisfying Condition 3.1, one can show (cfr. [176], Section 3.1) that the set $\mathcal{F}_{\alpha,K_{min}}$ in (3.4) can be realised as the family of composition maps

$$\mathcal{F}_{\alpha,K_{min}} = \{\Phi \circ F : F \in H_0^\alpha(\mathcal{O})\}, \qquad \alpha \in \mathbb{N}.$$

We then regard the solution map associated to (3.3) as one defined on H_0^α via

$$\mathscr{G} : H_0^\alpha(\mathcal{O}) \to L^2(\mathcal{O}), \qquad F \mapsto \mathscr{G}(F) := G(\Phi \circ F), \tag{3.5}$$

where $G(\Phi \circ F)$ is the solution to (3.3) now with $f = \Phi \circ F \in \mathcal{F}_{\alpha,K_{min}}$. In the results to follow, we will implicitly assume a link function Φ to be given and fixed, and understand the re-parametrised solution map $\mathscr{G}$ as being defined as in (3.5) for such choice of Φ.

3.2.3 Measurement model

Define the uniform distribution on $\mathcal{O}$ by $\mu = dx/\mathrm{vol}(\mathcal{O})$, where dx is the Lebesgue measure and $\mathrm{vol}(\mathcal{O}) = \int_{\mathcal{O}} dx$, and consider random design variables

$$(X_i)_{i=1}^N \overset{\text{iid}}{\sim} \mu, \qquad N \in \mathbb{N}. \tag{3.6}$$

For unknown $f \in \mathcal{F}_{\alpha,K_{min}}$, we model the statistical errors under which we observe the corresponding measurements $\{G(f)(X_i)\}_{i=1}^N$ by i.i.d. Gaussian random variables $W_i \sim N(0,1)$, all independent of the X_i's. Using the re-parameterisation $f = \Phi \circ F$ via a given link function from the previous subsection, the observation scheme is then

$$Y_i = \mathscr{G}(F)(X_i) + \sigma W_i, \qquad i = 1, \dots, N, \tag{3.7}$$

where $\sigma > 0$ is the noise amplitude. We will often use the shorthand notation $Y^{(N)} = (Y_i)_{i=1}^N$, with analogous definitions for $X^{(N)}$ and $W^{(N)}$. The random vectors (Y_i, X_i) on $\mathbb{R} \times \mathcal{O}$ are then i.i.d with laws denoted as P_F^i. Writing dy for the Lebesgue measure on $\mathbb{R}$, it follows that P_F^i has Radon-Nikodym density

$$p_F(y,x) := \frac{dP_F^i}{dy \times d\mu}(y,x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[y-\mathscr{G}(F)(x)]^2/(2\sigma^2)}, \qquad y \in \mathbb{R},\ x \in \mathcal{O}. \tag{3.8}$$

We will write $P_F^N = \otimes_{i=1}^N P_F^i$ for the joint law of $(Y^{(N)}, X^{(N)})$ on $\mathbb{R}^N \times \mathcal{O}^N$, with E_F^i, E_F^N the expectation operators corresponding to the laws P_F^i, P_F^N respectively. In the sequel we sometimes use the notation P_f^N instead of P_F^N when convenient.

3.2.4 The Bayesian approach

In the Bayesian approach one models the parameter $F \in H_0^\alpha(\mathcal{O})$ by a Borel probability measure Π supported in the Banach space $C(\mathcal{O})$. Since the map $(F,(y,x)) \mapsto p_F(y,x)$ can be shown to be jointly measurable, the posterior distribution $\Pi(\cdot|Y^{(N)}, X^{(N)})$ of $F|(Y^{(N)}, X^{(N)})$ arising from data in model (3.7) equals, by Bayes' formula (p.7, [101]),

$$\Pi(B|Y^{(N)}, X^{(N)}) = \frac{\int_B e^{\ell^{(N)}(F)} d\Pi(F)}{\int_{C(\mathcal{O})} e^{\ell^{(N)}(F')} d\Pi(F')} \qquad \text{any Borel set } B \subseteq C(\mathcal{O}), \tag{3.9}$$

where

$$\ell^{(N)}(F) = -\frac{1}{2\sigma^2} \sum_{i=1}^N [Y_i - \mathscr{G}(F)(X_i)]^2 \tag{3.10}$$

is (up to an additive constant) the joint log-likelihood function.

3.3 Main results

In this section we will show that the posterior distribution arising from certain priors concentrates near any sufficiently regular ground truth F_0 (or, equivalently, f_0), and provide a bound on the rate of this contraction, assuming the observation $(Y^{(N)}, X^{(N)})$ to be generated through model (3.7) of law $P^N_{F_0}$. We will regard $\sigma > 0$ as a fixed and known constant; in practice it may be replaced by the estimated sample variance of the Y_i's.

The priors we will consider are built around a Gaussian process base prior Π', but to deal with the nonlinearity of the inverse problem, some additional regularisation will be required. We first show how this can be done by a N-dependent 'rescaling' step as suggested in [167]. We then further show that a randomised truncation of a Karhunen-Loève-type series expansion of the base prior also leads to a consistent, 'fully Bayesian' solution of this inverse problem.

3.3.1 Statistical convergence rates with re-scaled Gaussian priors

We will freely use terminology from the basic theory of Gaussian processes and measures, see, e.g., [104], Chapter 2 for details.

Condition 3.3. *Let $\alpha > 1 + d/2$, $\beta \geq 1$, and let $\mathcal{H}$ be a Hilbert space continuously imbedded into $H^\alpha_0(\mathcal{O})$. Let Π' be a centred Gaussian Borel probability measure on the Banach space $C(\mathcal{O})$ that is supported on a separable measurable linear subspace of $C^\beta(\mathcal{O})$, and assume that the reproducing-kernel Hilbert space (RKHS) of Π' equals $\mathcal{H}$.*

As a basic example of a Gaussian base prior Π' satisfying Condition 3.3, consider a Matérn process $M = \{M(x),\ x \in \mathcal{O}\}$ indexed by $\mathcal{O}$ and of regularity $\alpha - d/2$ (cfr. Example 3.2 below for full details). We will assume that it is known that $F_0 \in H^\alpha(\mathcal{O})$ is supported inside a given compact subset K of the domain $\mathcal{O}$, and fix any smooth cut-off function $\chi \in C^\infty_c(\mathcal{O})$ such that $\chi = 1$ on K. Then, $\Pi' = \mathcal{L}(\chi M)$ is supported on the separable linear subspace $C^{\beta'}(\mathcal{O})$ of $C^\beta(\mathcal{O})$ for any $\beta < \beta' < \alpha - d/2$, and its RKHS $\mathcal{H} = \{\chi F,\ F \in H^\alpha(\mathcal{O})\}$ is continuously imbedded into $H^\alpha_0(\mathcal{O})$ (and contains $H^\alpha_K(\mathcal{O})$). The condition $F_0 \in \mathcal{H}$ that is employed in the following theorems then amounts to the standard assumption that $F_0 \in H^\alpha(\mathcal{O})$ be supported in a strict subset K of $\mathcal{O}$.

To proceed, if Π' is as above and $F' \sim \Pi'$, we consider the 're-scaled' prior

$$\Pi_N = \mathcal{L}(F_N), \qquad F_N = \frac{1}{N^{d/(4\alpha+4+2d)}} F', \tag{3.11}$$

Then Π_N again defines a centred Gaussian prior on $C(\mathcal{O})$, and a basic calculation (e.g., Exercise 2.6.5 in [104]) shows that its RKHS $\mathcal{H}_N$ is still given by $\mathcal{H}$ but now with norm

$$\|F\|_{\mathcal{H}_N} = N^{d/(4\alpha+4+2d)} \|F\|_{\mathcal{H}} \qquad \forall F \in \mathcal{H}. \tag{3.12}$$

Our first result shows that the posterior contracts towards F_0 in 'prediction'-risk at rate $N^{-(\alpha+1)/(2\alpha+2+d)}$ and that, moreover, the posterior draws possess a bound on their C^β-norm with overwhelming frequentist probability.

Theorem 3.1. *For fixed integer $\alpha > \beta + d/2$, $\beta \geq 1$, consider the Gaussian prior Π_N in (3.11) with base prior $F' \sim \Pi'$ satisfying Condition 3.3 for RKHS $\mathcal{H}$. Let $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)}, X^{(N)})$ in (3.7), set $\delta_N = N^{-(\alpha+1)/(2\alpha+2+d)}$, and assume $F_0 \in \mathcal{H}$.*

Then for any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, F_0, D, \alpha, \beta$, as well as on $\mathcal{O}, d, g$) such that, as $N \to \infty$,

$$\Pi_N(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2} > L\delta_N | Y^{(N)}, X^{(N)}) = O_{P^N_{F_0}}(e^{-DN\delta_N^2}), \tag{3.13}$$

and for sufficiently large $M > 0$ (depending on σ, D, α, β)

$$\Pi_N(F : \|F\|_{C^\beta} > M | Y^{(N)}, X^{(N)}) = O_{P^N_{F_0}}(e^{-DN\delta_N^2}). \tag{3.14}$$

Following ideas in [167], we can combine (3.13) with the regularisation property (3.14) and a suitable stability estimate for G^{-1} to show that the posterior contracts about f_0 also in L^2-risk. We shall employ the stability estimate proved in [176, Lemma 24] which requires the source function g in the base PDE (3.3) to be strictly positive, a natural condition ensuring injectivity of the map $f \mapsto G(f)$, see [199]. Denote the push-forward posterior on the conductivities f by

$$\tilde{\Pi}_N(\cdot|Y^{(N)}, X^{(N)}) := \mathcal{L}(f), \qquad f = \Phi \circ F \ : \ F \sim \Pi_N(\cdot|Y^{(N)}, X^{(N)}). \tag{3.15}$$

Theorem 3.2. *Let $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$, δ_N and F_0 be as in Theorem 3.1 for integer $\beta \geq 2$. Let $f_0 = \Phi \circ F_0$ and assume in addition that $\inf_{x \in \mathcal{O}} g(x) \geq g_{min} > 0$. Then for any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, f_0, D, \alpha, \beta, \mathcal{O}, g_{min}, d$) such that, as*

$N \to \infty$,

$$\tilde{\Pi}_N(f : \|f - f_0\|_{L^2} > LN^{-\lambda}|Y^{(N)}, X^{(N)}) = O_{P_{f_0}^N}(e^{-DN\delta_N^2}), \qquad \lambda = \frac{(\alpha+1)(\beta-1)}{(2\alpha+2+d)(\beta+1)}.$$

We note that as the smoothness α of f_0 increases, we can employ priors of higher regularity α, β. In particular, if $F_0 \in C^\infty = \cap_{\alpha>0} H^\alpha$, we can let the above rate $N^{-\lambda}$ be as closed as desired to the 'parametric' rate $N^{-1/2}$.

We conclude this section showing that the posterior mean $E^\Pi[F|Y^{(N)}, X^{(N)}]$ of $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$ converges to F_0 at the rate $N^{-\lambda}$ from Theorem 3.2. We formulate this result at the level of the vector space valued parameter F (instead of for conductivities f), as the most commonly used MCMC algorithms (such as pCN, see Section 3.4.1) target the posterior distribution of F.

Theorem 3.3. *Under the hypotheses of Theorem 3.2, let $\bar{F}_N = E^\Pi[F|Y^{(N)}, X^{(N)}]$ be the (Bochner-) mean of $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$. Then, as $N \to \infty$,*

$$P_{F_0}^N\Big(\|\bar{F}_N - F_0\|_{L^2} > N^{-\lambda}\Big) \to 0. \tag{3.16}$$

The same result holds for the implied conductivities, that is, for $\|\Phi \circ \bar{F}_N - f_0\|_{L^2}$ replacing $\|\bar{F}_N - F_0\|_{L^2}$, since composition with Φ is Lipschitz.

3.3.2 Extension to high-dimensional Gaussian sieve priors

It is often convenient, for instance for computational reasons as discussed in Section 3.4.1, to employ 'sieve'-priors that are concentrated on a finite-dimensional approximation of the parameter space supporting the prior. For example a truncated Karhunen-Loève-type series expansion (or some other discretisation) of the Gaussian base prior Π' is frequently used [72, 113]. The theorems of the previous subsection remain valid if the approximation spaces are appropriately chosen.

Let us illustrate this by considering a Gaussian series prior based on an orthonormal basis $\{\Phi_{lr},\ l \geq -1,\ r \in \mathbb{Z}^d\}$ of $L^2(\mathbb{R}^d)$ composed of sufficiently regular, compactly supported Daubechies wavelets (see Chapter 4 in [104] for details). We assume that $F_0 \in H_K^\alpha(\mathcal{O})$ for some $K \subset \mathcal{O}$, and denote by $\mathcal{R}_l$ the set of indices r for which the support of Φ_{lr} intersects K. Fix any compact $K' \subset \mathcal{O}$ such that $K \subsetneq K'$, and a cut-off function $\chi \in C_c^\infty(\mathcal{O})$ such that $\chi = 1$ on K'. For any real $\alpha > 1 + d/2$, consider the prior

Π'_J arising as the law of the Gaussian random sum

$$\Pi'_J = \mathcal{L}(\chi F), \qquad F = \sum_{l \le J, r \in \mathcal{R}_l} 2^{-l\alpha} F_{lr} \Phi_{lr}, \quad F_{lr} \overset{\text{iid}}{\sim} N(0,1), \tag{3.17}$$

where $J = J_N \to \infty$ is a (deterministic) truncation point to be chosen. Then Π'_J defines a centred Gaussian prior that is supported on the finite-dimensional space

$$\mathcal{H}_J = \text{span}\{\chi \Phi_{lr},\ l \le J,\ r \in \mathcal{R}_l\} \subset C(\mathcal{O}). \tag{3.18}$$

Proposition 3.1. *Consider a prior Π_N as in (3.11) where now $F' \sim \Pi'_J$ and $J = J_N \in \mathbb{N}$ is such that $2^J \simeq N^{1/(2\alpha+2+d)}$. Let $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)}, X^{(N)})$ in (3.7), and assume $F_0 \in H^\alpha_K(\mathcal{O})$. Then the conclusions of Theorems 3.1-3.3 remain valid (under the respective hypotheses on α, β, g).*

A similar result could be proved for more general Gaussian priors (not of wavelet type), but we refrain from giving these extensions here.

3.3.3 Statistical convergence rates with randomly truncated Gaussian series priors

In this section we show that instead of rescaling Gaussian base priors Π', Π'_J in a $N-$dependent way to attain extra regularisation, one may also randomise the dimensionality parameter J in (3.17) by a hyper-prior with suitable tail behaviour. While this is computationally somewhat more expensive (by necessitating a hierarchical sampling method, see Section 3.4.1), it gives a possibly more principled approach to ('fully') Bayesian regularisation in our inverse problem. The theorem below will show that such a procedure is consistent in the frequentist sense, at least for smooth enough F_0.

For the wavelet basis and cut-off function χ introduced before (3.17), we consider again a random (conditionally Gaussian) sum

$$\Pi = \mathcal{L}(\chi F), \qquad F = \sum_{l \le J, r \in \mathcal{R}_l} 2^{-l\alpha} F_{lr} \Phi_{lr}, \quad F_{lr} \overset{\text{iid}}{\sim} N(0,1) \tag{3.19}$$

where now J is a random truncation level, independent of the random coefficients F_{lr}, satisfying the following inequalities

$$\Pr(J > j) = e^{-2^{jd} \log 2^{jd}} \quad \forall j \ge 1; \qquad \Pr(J = j) \gtrsim e^{-2^{jd} \log 2^{jd}}, \quad j \to \infty. \tag{3.20}$$

When $d = 1$, a (log-) Poisson random variable satisfies these tail conditions, and for $d > 1$ such a random variable J can be easily constructed too – see Example 3.3 below.

Our first result in this section shows that the posterior arising from the truncated series prior in (3.19) achieves (up to a log-factor) the same contraction rate in L^2-prediction risk as the one obtained in Theorem 3.1. Moreover, as is expected in light of the results in [236, 195], the posterior adapts to the unknown regularity α_0 of F_0 when it exceeds the base smoothness level α.

Theorem 3.4. *For any $\alpha > 1 + d/2$, let Π be the random series prior in (3.19), and let $\Pi(\cdot|Y^{(N)}, X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)}, X^{(N)})$ in (3.7). Then, for each $\alpha_0 \geq \alpha$ and any $F_0 \in H_K^{\alpha_0}(\mathcal{O})$, we have that for any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, F_0, D, \alpha, \mathcal{O}, d, g$) such that, as $N \to \infty$,*

$$\Pi(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2} > L\xi_N|Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\xi_N^2}),$$

where $\xi_N = N^{-(\alpha_0+1)/(2\alpha_0+2+d)} \log N$. Moreover, for $\mathcal{H}_J$ the finite-dimensional subspaces in (3.18) and $J_N \in \mathbb{N}$ such that $2^{J_N} \simeq N^{1/(2\alpha_0+2+d)}$, we also have that for sufficiently large $M > 0$ (depending on D, α)

$$\Pi(F : F \in \mathcal{H}_{J_N},\ \|F\|_{H^\alpha} \leq M2^{J_N\alpha}N\xi_N^2|Y^{(N)}, X^{(N)}) = 1 - O_{P_{F_0}^N}(e^{-DN\xi_N^2}). \tag{3.21}$$

We can now exploit the previous result along with the finite-dimensional support of the posterior and again the stability estimate from [176] to obtain the following consistency theorem for $F_0 \in H^{\alpha_0}$ if α_0 is large enough (with a precise bound $\alpha_0 \geq \alpha^*$ given in the proof of Lemma 3.2).

Theorem 3.5. *Let the link function Φ in the definition (3.5) of $\mathscr{G}$ satisfy Condition 3.2. Let $\Pi(\cdot|Y^{(N)}, X^{(N)})$, ξ_N be as in Theorem 3.4, assume in addition $g \geq g_{min} > 0$ on $\mathcal{O}$, and let $\tilde{\Pi}(\cdot|Y^{(N)}, X^{(N)})$ be the posterior distribution of f as in (3.15). Then for $f_0 = \Phi \circ F_0$ with $F_0 \in H_K^{\alpha_0}(\mathcal{O})$ for α_0 large enough (depending on α, d, a) and for any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, f_0, D, \alpha, \mathcal{O}, g_{min}, d$) such that, as $N \to \infty$,*

$$\tilde{\Pi}(f : \|f - f_0\|_{L^2} > LN^{-\rho}|Y^{(N)}, X^{(N)}) = O_{P_{f_0}^N}(e^{-DN\xi_N^2}), \qquad \rho = \frac{(\alpha_0+1)(\alpha-1)}{(2\alpha_0+2+d)(\alpha+1)}.$$

Just as before, for $f_0 \in C^\infty$ the above rate can be made as close as desired to $N^{-1/2}$ by choosing α large enough. Moreover, the last contraction theorem also translates into a convergence result for the posterior mean of F.

Theorem 3.6. *Under the hypotheses of Theorem 3.5, let $\bar{F}_N = E^\Pi[F|Y^{(N)}, X^{(N)}]$ be the mean of $\Pi(\cdot|Y^{(N)}, X^{(N)})$. Then, as $N \to \infty$,*

$$P^N_{F_0}\Big(\|\bar{F}_N - F_0\|_{L^2} > N^{-\rho}\Big) \to 0. \tag{3.22}$$

We note that the proof of the last two theorems crucially takes advantage of the 'non-symmetric' and 'non-exponential' nature of the stability estimate from [176], and may not hold in other nonlinear inverse problems where such an estimate may not be available (e.g., as in [167, 2] or also in the Schrödinger equation setting studied in [171, 176]).

Let us conclude this section by noting that hierarchical priors such as the one studied here are usually devised to 'adapt to unknown' smoothness α_0 of F_0, see [236, 195]. Note that while our posterior distribution is adaptive to α_0 in the 'prediction risk' setting of Theorem 3.4, the rate $N^{-\rho}$ obtained in Theorems 3.5 and 3.6 for the inverse problem *does* depend on the minimal smoothness α, and is therefore *not adaptive*. Nevertheless, this hierarchical prior gives an example of a fully Bayesian, consistent solution of our inverse problem.

3.4 Concluding discussion

3.4.1 Posterior computation

As mentioned in the introduction, in the context of the elliptic inverse problem considered in the present chapter, posterior distributions arising from Gaussian process priors such as those above can be computed by MCMC algorithms, see [63, 62, 26], and computational guarantees can be obtained as well: for Gaussian priors, [113] establish non-asymptotic sampling bounds for the 'preconditioned Crank-Nicholson (pCN)' algorithm, which hold even in the absence of log-concavity of the likelihood function, and which imply bounds on the approximation error for the computation of the posterior mean. The algorithm can be implemented as long as it is possible to evaluate the forward map $F \mapsto \mathscr{G}(F)(x)$ at $x \in \mathcal{O}$, which in our context can be done by using standard numerical methods to solve the elliptic PDE (3.3). In practice, these algorithms often employ a finite-dimensional approximation of the parameter space (cfr. Section 3.3.2). See Section 1.2.2 for further discussion on the implementation of the algorithm and for some numerical results.

In order to sample from the posterior distribution arising from the more complex hierarchical prior (3.19), MCMC methods based on fixed Gaussian priors (such as the pCN algorithm) can be employed within a suitable Gibbs-sampling scheme that exploits the conditionally Gaussian structure of the prior. The algorithm would then alternate, for given J, an MCMC step targeting the marginal posterior distribution of $F|(Y^{(N)}, X^{(N)}, J)$, followed by, given the actual sample of F, a second MCMC run with objective the marginal posterior of $J|(Y^{(N)}, X^{(N)}, F)$. A related approach to hierarchical inversion is empirical Bayesian estimation. In the present setting this would entail first estimating the truncation level J from the data, via an estimator $\hat{J} = \hat{J}(Y^{(N)}, X^{(N)})$ (e.g., the marginal maximum likelihood estimator), and then performing inference based on the fixed finite-dimensional prior $\Pi_{\hat{J}}$ (defined as in (3.19) with J replaced by $\hat{J}$). See [137] where this is studied in a diagonal linear inverse problem.

3.4.2 Open problems: towards optimal convergence rates

The convergence rates obtained in this chapter demonstrate the frequentist consistency of a Bayesian (Gaussian process) inversion method in the elliptic inverse problem (3.2) with data (3.1) in the large sample limit $N \to \infty$. While the rates approach the optimal rate $N^{-1/2}$ for very smooth models ($\alpha \to \infty$), the question of optimality for fixed α remains an interesting avenue for future research. We note that for the 'PDE-constrained regression' problem of recovering $\mathscr{G}(F_0)$ in 'prediction' loss, the rate $\delta_N = N^{-(\alpha+1)/(2\alpha+2+d)}$ obtained in Theorems 3.1 and 3.4 can be shown to be minimax optimal (as in [176, Theorem 10]). But for the recovery rates for f obtained in Theorems 3.3 and 3.6, no matching lower bounds are currently known. Related to this issue, in [176] faster (but still possibly suboptimal) rates are obtained for the modes of our posterior distributions (MAP estimates, which are not obviously computable in polynomial time), and one may loosely speculate here about computational hardness barriers in our nonlinear inverse problem. These issues pose formidable challenges for future research and are beyond the scope of the present investigation.

3.5 Proofs

We assume without loss of generality that $\mathrm{vol}(\mathcal{O}) = 1$. In the proof, we will repeatedly exploit properties of the (re-parametrised) solution map $\mathscr{G}$ defined in (3.5), which was studied in detail in [176]. Specifically, in the proof of Theorem 9 in [176] it is shown that, for all $\alpha > 1 + d/2$ and any $F_1, F_2 \in H_0^\alpha(\mathcal{O})$,

$$\|\mathscr{G}(F_1) - \mathscr{G}(F_2)\|_{L^2(\mathcal{O})} \lesssim (1 + \|F_1\|^4_{C^1(\mathcal{O})} \vee \|F_2\|^4_{C^1(\mathcal{O})})\|F_1 - F_2\|_{(H^1(\mathcal{O}))^*}, \tag{3.23}$$

where we denote by X^* the topological dual Banach space of a normed linear space X. Secondly, we have (Lemma 20 in [176]) for some constant $c > 0$ (only depending on d, $\mathcal{O}$ and K_{min}),

$$\sup_{F \in H_0^\alpha} \|\mathscr{G}(F)\|_\infty \le c\|g\|_\infty < \infty. \tag{3.24}$$

Therefore the inverse problem (3.7) falls in the general framework considered in Appendix 3.A below (with $\beta = \kappa = 1$, $\gamma = 4$ in (3.32) and $S = c\|g\|_\infty$ in (3.33)) ; in particular Theorems 3.1 and 3.4 then follow as particular cases of the general contraction rate results derived in Theorem 3.8 and Theorem 3.9, respectively. It thus remains to derive Theorems 3.2 and 3.3 from Theorem 3.1, and Theorems 3.5 and 3.6 from Theorem 3.4, respectively.

To do so we recall here another key result from [176], namely their stability estimate Lemma 24: for $\alpha > 2 + d/2$, if $G(f)$ denotes the solution of the PDE (3.3) with g satisfying $\inf_{x \in \mathcal{O}} g(x) \ge g_{min} > 0$, then for fixed $f_0 \in \mathcal{F}_{\alpha, K_{min}}$ and all $f \in \mathcal{F}_{\alpha, K_{min}}$

$$\|f - f_0\|_{L^2(\mathcal{O})} \lesssim \|f\|_{C^1(\mathcal{O})}\|G(f) - G(f_0)\|_{H^2(\mathcal{O})}, \tag{3.25}$$

with multiplicative constant independent of f.

3.5.1 Proofs for Section 3.3.1

Proof of Theorem 3.2

The conclusions of Theorem 3.1 can readily be translated for the push-forward posterior $\tilde{\Pi}_N(\cdot|Y^{(N)}, X^{(N)})$ from (3.15). In particular, (3.13) implies, for $f_0 = \Phi \circ F_0$, as $N \to \infty$,

$$\tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{L^2} > L\delta_N|Y^{(N)}, X^{(N)}) = O_{P^N_{f_0}}(e^{-DN\delta_N^2}); \tag{3.26}$$

and using Lemma 29 in [176] and (3.14) we obtain for sufficiently large $M' > 0$

$$\tilde{\Pi}_N(f : \|f\|_{C^\beta} > M'|Y^{(N)}, X^{(N)}) \le \Pi_N(F : \|F\|_{C^\beta} > M|Y^{(N)}, X^{(N)}) = O_{P^N_{f_0}}(e^{-DN\delta_N^2}). \tag{3.27}$$

From the previous bounds we now obtain the following result.

Lemma 3.1. *For $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$, δ_N and F_0 as in Theorem 3.1, let $\tilde{\Pi}_N(\cdot|Y^{(N)}, X^{(N)})$ be the push-forward posterior distribution from (3.15). Then, for $f_0 = \Phi \circ F_0$ and any $D > 0$ there exists $L > 0$ large enough such that, as $N \to \infty$,*

$$\tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{H^2} > L\delta_N^{(\beta-1)/(\beta+1)}|Y^{(N)}, X^{(N)}) = O_{P^N_{F_0}}(e^{-DN\delta_N^2}).$$

Proof. Using the continuous imbedding of $C^\beta \subset H^\beta, \beta \in \mathbb{N}$, and (3.27), for some $M' > 0$ as $N \to \infty$,

$$\tilde{\Pi}_N(f : \|f\|_{H^\beta} > M'|Y^{(N)}, X^{(N)}) = O_{P^N_{F_0}}(e^{-DN\delta_N^2}).$$

Now if $f \in H^\beta$ with $\|f\|_{H^\beta} \leq M'$, Lemma 23 in [176] implies $G(f), G(f_0) \in H^{\beta+1}$, with

$$\|G(f_0)\|_{H^{\beta+1}} \lesssim 1 + \|f_0\|_{H^\beta}^{\beta(\beta+1)} < \infty, \qquad \|G(f)\|_{H^{\beta+1}} \lesssim 1 + \|f\|_{H^\beta}^{\beta(\beta+1)} < M'' < \infty;$$

and by the usual interpolation inequality for Sobolev spaces [153],

$$\begin{aligned}\|G(f) - G(f_0)\|_{H^2} &\lesssim \|G(f) - G(f_0)\|_{L^2}^{(\beta-1)/(\beta+1)}\|G(f) - G(f_0)\|_{H^{\beta+1}}^{2/(\beta+1)} \\ &\lesssim \|G(f) - G(f_0)\|_{L^2}^{(\beta-1)/(\beta+1)}.\end{aligned}$$

Thus, by what precedes and (3.26), for sufficiently large $L > 0$

$$\begin{aligned}&\tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{H^2} > L\delta_N^{(\beta-1)/(\beta+1)}|Y^{(N)}, X^{(N)}) \\ &\leq \tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{L^2} > L'\delta_N|Y^{(N)}, X^{(N)}) + \tilde{\Pi}_N(f : \|f\|_{H^\beta} > M''|Y^{(N)}, X^{(N)}) \\ &= O_{P^N_{F_0}}(e^{-DN\delta_N^2}),\end{aligned}$$

as $N \to \infty$.

□

To prove Theorem 3.2 we use (3.25), (3.27) and Lemma 3.1 to the effect that for any $D > 0$ we can find $L, M > 0$ large enough such that, as $N \to \infty$,

$$\begin{aligned}&\tilde{\Pi}_N\left(f : \|f - f_0\|_{L^2} > L\delta_N^{\frac{\beta-1}{\beta+1}}\middle|Y^{(N)}, X^{(N)}\right) \\ &\leq \tilde{\Pi}_N\left(f : \|G(f) - G(f_0)\|_{H^2} > L'\delta_N^{\frac{\beta-1}{\beta+1}}\middle|Y^{(N)}, X^{(N)}\right) + \tilde{\Pi}_N\left(f : \|f\|_{C^\beta} > M|Y^{(N)}, X^{(N)}\right) \\ &= O_{P^N_{F_0}}(e^{-DN\delta_N^2}).\end{aligned}$$

Proof of Theorem 3.3

The proof largely follows ideas of [167] but requires a slightly more involved, iterative uniform integrability argument to also control the probability of events $\{F : \|F\|_{C^\beta} > M\}$ on whose complements we can subsequently exploit regularity properties of the inverse link function Φ^{-1}.

Using Jensen's inequality, it is enough to show, as $N \to \infty$,

$$P_{F_0}^N\Big(E^\Pi\left[\|F-F_0\|_{L^2}^2\Big|Y^{(N)},X^{(N)}\right] > N^{-\lambda}\Big) \to 0.$$

For $M>0$ sufficiently large to be chosen, we decompose

$$\begin{aligned} E^\Pi\left[\|F-F_0\|_{L^2}^2\Big|Y^{(N)},X^{(N)}\right] &= E^\Pi\left[\|F-F_0\|_{L^2}1_{\{\|F\|_{C^\beta}\le M\}}\Big|Y^{(N)},X^{(N)}\right] \\ &\quad + E^\Pi\left[\|F-F_0\|_{L^2}1_{\{\|F\|_{C^\beta}> M\}}\Big|Y^{(N)},X^{(N)}\right]. \end{aligned} \tag{3.28}$$

Using the Cauchy-Schwarz inequality we can upper bound the expectation in the second summand by

$$\sqrt{E^\Pi\left[\|F-F_0\|_{L^2}^2\Big|Y^{(N)},X^{(N)}\right]}\sqrt{\Pi_N(F:\|F\|_{C^\beta}>M|Y^{(N)},X^{(N)})}.$$

In view of (3.14), for all $D>0$ we can choose $M>0$ large enough to obtain

$$\begin{aligned} &P_{F_0}^N\Big(E^\Pi\left[\|F-F_0\|_{L^2}^2\Big|Y^{(N)},X^{(N)}\right]\Pi_N(F:\|F\|_{C^\beta}>M|Y^{(N)},X^{(N)}) > N^{-2\lambda}\Big) \\ &\quad\le P_{F_0}^N\Big(E^\Pi\left[\|F-F_0\|_{L^2}^2\Big|Y^{(N)},X^{(N)}\right]e^{-DN\delta_N^2} > N^{-2\lambda}\Big) + o(1). \end{aligned}$$

To bound the probability in the last line, let $\mathcal{B}_N$ be the sets defined in (3.34) below, note that Lemma 3.3 and Lemma 3.9 below jointly imply that $\Pi_N(\mathcal{B}_N)\ge ae^{-AN\delta_N^2}$ for some $a,A>0$. Also, let $\nu(\cdot)=\Pi_N(\cdot\cap\mathcal{B}_N)/\Pi_N(\mathcal{B}_N)$, and let $\mathcal{C}_N$ be the event from (3.40), for which Lemma 7.3.2 in [104] implies that $P_{F_0}^N(\mathcal{C}_N)\to 1$ as $N\to\infty$. Then

$$\begin{aligned} &P_{F_0}^N\Big(E^\Pi\left[\|F-F_0\|_{L^2}^2\Big|Y^{(N)},X^{(N)}\right]e^{-DN\delta_N^2} > N^{-2\lambda}\Big) \\ &\le P_{F_0}^N\left(\frac{\int_{C(\mathcal{O})}\|F-F_0\|_{L^2}^2\prod_{i=1}^N p_F/p_{F_0}(Y_i,X_i)d\Pi_N(F)}{\Pi(\mathcal{B}_N)\int_{\mathcal{B}_N}\prod_{i=1}^N p_F/p_{F_0}(Y_i,X_i)d\nu(F)}e^{-DN\delta_N^2} > N^{-2\lambda},\mathcal{C}_N\right) + o(1) \\ &\le P_{F_0}^N\Big(\int_{C(\mathcal{O})}\|F-F_0\|_{L^2}^2\prod_{i=1}^N p_F/p_{F_0}(Y_i,X_i)d\Pi_N(F) > N^{-2\lambda}ae^{(D-A-2)N\delta_N^2}\Big) + o(1) \end{aligned}$$

which is upper bounded, using Markov's inequality and Fubini's theorem, by

$$\frac{1}{a}e^{-(D-A-2)N\delta_N^2}N^{2\lambda}\int_{C(\mathcal{O})}\|F-F_0\|_{L^2}^2 E_{F_0}^N\left(\prod_{i=1}^N \frac{p_F}{p_{F_0}}(Y_i,X_i)\right)d\Pi_N(F).$$

Taking $D > A+2$ (and M large enough in (3.28)), using the fact that $E_{F_0}^N\big(\prod_{i=1}^N p_F/p_{F_0}(Y_i,X_i)\big) = 1$, and that $E^{\Pi_N}\|F\|_{L^2} < \infty$ (by Fernique's theorem, e.g., [104, Exercise 2.1.5]), we then conclude

$$P_{F_0}^N\left(E^{\Pi}\left[\|F-F_0\|_{L^2}^2 1_{\{\|F\|_{C^\beta}>M\}}\Big|Y^{(N)},X^{(N)}\right] > N^{-\lambda}\right) \to 0, \qquad N\to\infty. \tag{3.29}$$

To handle the first term in (3.28), let $f=\Phi\circ F$ and $f_0=\Phi\circ F_0$. Then for all $x\in\mathcal{O}$, by the mean value and inverse function theorems,

$$|F(x)-F_0(x)| = |\Phi^{-1}\circ f(x)-\Phi^{-1}\circ f_0(x)| = \frac{1}{|\Phi'(\Phi^{-1}(\eta))|}|f(x)-f_0(x)|$$

for some η lying between $f(x)$ and $f_0(x)$. If $\|F\|_{C^\beta}\le M$ then, as Φ is strictly increasing, necessarily $f(x)=\Phi(F(x))\in[\Phi(-M),\Phi(M)]$ for all $x\in\mathcal{O}$. Similarly, the range of f_0 is contained in the compact interval $[\Phi(-M),\Phi(M)]$ for $M\ge\|F_0\|_\infty$, so that

$$|\Phi^{-1}\circ f(x)-\Phi^{-1}\circ f_0(x)| \le \frac{1}{\min_{z\in[-M,M]}\Phi'(z)}|f(x)-f_0(x)| \lesssim |f(x)-f_0(x)|$$

for a multiplicative constant not depending on $x\in\mathcal{O}$. It follows

$$\|F-F_0\|_{L^2}1_{\{\|F\|_{C^\beta}\le M\}} \lesssim \|f-f_0\|_{L^2}1_{\{\|F\|_{C^\beta}\le M\}},$$

and

$$E^{\Pi}\left[\|F-F_0\|_{L^2}1_{\{\|F\|_{C^\beta}\le M\}}\Big|Y^{(N)},X^{(N)}\right] \lesssim E^{\tilde\Pi}\left[\|f-f_0\|_{L^2}\Big|Y^{(N)},X^{(N)}\right].$$

Noting that for each $L>0$ the last expectation is upper bounded by

$$\begin{aligned}
&LN^{-\lambda}+E^{\tilde\Pi}\Big[\|f-f_0\|_{L^2}1_{\{\|f-f_0\|_{L^2}>LN^{-\lambda}\}}|Y^{(N)},X^{(N)}]\\
&\quad\le LN^{-\lambda}+\sqrt{E^{\tilde\Pi}[\|f-f_0\|_{L^2}^2|Y^{(N)},X^{(N)}]}\sqrt{\tilde\Pi_N(f:\|f-f_0\|_{L^2}>LN^{-\lambda}|Y^{(N)},X^{(N)})},
\end{aligned}$$

we can repeat the above argument, with the event $\{F:\|F\|_{C^\beta}>M\}$ replaced by the event $\{f:\|f-f_0\|_{L^2}>LN^{-\lambda}\}$, to deduce from Theorem 3.2 that for $D>A+2$ there exists $L > 0$ large enough such that

$$P_{F_0}^N\Big(E^{\tilde{\Pi}}[\|f - f_0\|_{L^2}^2 | Y^{(N)}, X^{(N)}]\tilde{\Pi}_N(f : \|f - f_0\|_{L^2}^2 > LN^{-\lambda} | Y^{(N)}, X^{(N)}) > N^{-\lambda}\Big)$$
$$\lesssim e^{-(D-A-2)N\delta_N^2} N^{2\lambda}$$

which combined with (3.29) and the definition of δ_N concludes the proof.

3.5.2 Sieve prior proofs

The proof only requires minor modification from the proofs of Section 3.3.1. We only discuss here the main points: one first applies the L^2-prediction risk Theorem 3.8 with a sieve prior. In the proof of the small ball Lemma 3.3 one uses the following observations: the projection $P_J F_0 \in \mathcal{H}_J$ of $F_0 \in H_K^\alpha$ defined in (3.61) satisfies by (3.63)

$$\|F_0 - P_J F_0\|_{(H^1(\mathcal{O}))^*} \lesssim 2^{-J(\alpha+1)};$$

hence choosing J such that $2^J \simeq N^{1/(2\alpha+2+d)}$, and noting also that $\|P_J F_0\|_{C^1} \leq \|F_0\|_{C^1} < \infty$ for all J by standard properties of wavelet bases, it follows from (3.23) that

$$\|\mathscr{G}(F_0) - \mathscr{G}(P_J F_0)\|_{L^2} \lesssim \|F_0 - P_J F_0\|_{(H^1)^*} \lesssim N^{-(\alpha+1)/(2\alpha+2+d)} = \delta_N.$$

Therefore, by the triangle inequality,

$$\Pi_N(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2} \geq \delta_N/q) \geq \Pi_N(F : \|\mathscr{G}(F) - \mathscr{G}(P_J F_0)\|_{L^2} \geq q'\delta_N).$$

The rest of the proof of Lemma 3.3 then carries over (with $P_J F_0$ replacing F_0), upon noting that (3.60) and a Sobolev imbedding imply

$$\sup_{J\in\mathbb{N}} E^{\Pi'_J}\|F\|_{C^1}^2 < \infty, \quad \text{as well as} \quad \|F\|_{H^\alpha} \leq c\|F\|_{\mathcal{H}_J} \text{ for all } F \in \mathcal{H}_J$$

for some constant $c > 0$ independent of J. Moreover, the last two properties are sufficient to prove an analogue of Lemma 3.4 as well, so that Theorem 3.8 indeed applies to the sieve prior. The proof from here onwards is identical to the ones of Theorems 3.1-3.3 for the unsieved case, using also that what precedes implies that $\sup_J E^{\Pi'_J}\|F\|_{L^2}^2 < \infty$, relevant in the proof of convergence of the posterior mean.

3.5.3 Proofs for Section 3.3.3

Inspection of the proofs for re-scaled priors implies that Theorems 3.5 and 3.6 can be deduced from Theorem 3.4 if we can show that posterior draws lie in a α-Sobolev ball of fixed radius with sufficiently high frequentist probability. This is the content of the next result.

Lemma 3.2. *Under the hypotheses of Theorem 3.5, there exists $\alpha^* > 0$ (depending on α, d and a) such that for each $F_0 \in H_K^{\alpha_0}(\mathcal{O})$, $\alpha_0 > \alpha^*$, and any $D > 0$ we can find $M > 0$ large enough such that, as $N \to \infty$,*

$$\Pi(F : \|F\|_{H^\alpha} \le M | Y^{(N)}, X^{(N)}) = 1 - O_{P_{F_0}^N}(e^{-DN\xi_N^2}).$$

Proof. Theorem 3.4 implies that for all $D > 0$ and sufficiently large $L, M > 0$, if $J_N \in \mathbb{N} : 2^{J_N} \simeq N^{1/(2\alpha_0+2+d)}$ and denoting by

$$\mathcal{A}_N = \left\{ F \in \mathcal{H}_{J_N} : \|F\|_{H^\alpha} \le M 2^{J_N\alpha}\sqrt{N}\xi_N,\ \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2} \le L\xi_N \right\},$$

then as $N \to \infty$

$$\Pi(\mathcal{A}_N | Y^{(N)}, X^{(N)}) = 1 - O_{P_{F_0}^N}(e^{-DN\xi_N^2}). \tag{3.30}$$

Next, note that if $F \in \mathcal{H}_{J_N}$, then by standard properties of wavelet bases (cfr. (3.64)), $\|F\|_{H^\alpha} \lesssim 2^{J_N\alpha}\|F\|_{L^2}$ for all N large enough. Thus, for $P_{J_N}F_0$ the projection of F_0 onto $\mathcal{H}_{J_N}$ defined according to (3.61),

$$\|F\|_{H^\alpha} \le \|F - P_{J_N}F_0\|_{H^\alpha} + \|P_{J_N}F_0\|_{H^\alpha} \lesssim 2^{J_N\alpha}\|F - F_0\|_{L^2} + \|F_0\|_{H^\alpha},$$

and a Sobolev imbedding further gives $\|F\|_{L^\infty} \le M' 2^{J_N\alpha}\sqrt{N}\xi_N$, for some $M' > 0$. Now letting $f = \Phi \circ F$ and $f_0 = \Phi \circ F_0$, by similar argument as in the proof of Theorem 3.3 combined with monotonicity of Φ', we see that for all N large enough

$$\|F - F_0\|_{L^2} \le \frac{1}{\Phi'(-M' 2^{J_N\alpha}\sqrt{N}\xi_N)}\|f - f_0\|_{L^2}.$$

Then, using the assumption on the left tail of Φ in Condition 3.2, and the stability estimate (3.25),

$$\|F - F_0\|_{L^2} \lesssim (2^{J_N\alpha}\sqrt{N}\xi_N)^a \|f\|_{H^\alpha}\|G(f) - G(f_0)\|_{H^2}.$$

Finally, by the interpolation inequality for Sobolev spaces [153] and Lemma 23 in [176],

$$\begin{aligned}\|G(f)-G(f_0)\|_{H^2} &\lesssim \|G(f)-G(f_0)\|_{L^2}^{(\alpha-1)/(\alpha+1)}\|G(f)-G(f_0)\|_{H^{\alpha+1}}^{2/(\alpha+1)}\\ &\lesssim \xi_N^{(\alpha-1)/(\alpha+1)}(\|G(f)\|_{H^{\alpha+1}}+\|G(f_0)\|_{H^{\alpha+1}})^{2/(\alpha+1)}\\ &\lesssim \xi_N^{(\alpha-1)/(\alpha+1)}(1+\|f\|_{H^\alpha}^{\alpha^2+\alpha})^{2/(\alpha+1)},\end{aligned}$$

so that, in conclusion, for each $F\in\mathcal{A}_N$ and sufficiently large N,

$$\|F\|_{H^\alpha}\lesssim 1+2^{J_N\alpha}(2^{J_N\alpha}\sqrt{N}\xi_N)^a\|f\|_{H^\alpha}\xi_N^{\frac{\alpha-1}{\alpha+1}}(1+\|f\|_{H^\alpha}^{\alpha^2+\alpha})^{\frac{2}{\alpha+1}}.$$

The last term is bounded, using Lemma 29 in [176], by a multiple of

$$\xi_N^{\frac{\alpha-1}{\alpha+1}}2^{J_N\alpha}(2^{J_N\alpha}\sqrt{N}\xi_N)^{2\alpha^2+2\alpha+a}=N^{-\frac{(\alpha_0+1)(\alpha-1)}{(2\alpha_0+2+d)(\alpha+1)}}N^{\frac{2\alpha^3+(2+d)\alpha^2+(1+a+d)\alpha+ad/2}{2\alpha_0+2+d}}$$

the last identity holding up to a log factor. Hence, if

$$\alpha_0>\alpha^*:=\frac{[2\alpha^3+(2+d)\alpha^2+(1+a+d)\alpha+ad/2](\alpha+1)}{(\alpha-1)}$$

then we conclude overall that $\|F\|_{H^\alpha}\lesssim 1+o(1)$ as $N\to\infty$ for all $F\in\mathcal{A}_N$, proving the claim in view of (3.30).

□

Replacing β by α in the conclusion of Lemma 3.1, the proof of Theorem 3.5 now proceeds as in the proof of Theorem 3.2 without further modification. Likewise, Theorem 3.6 can be shown following the same argument as in the proof of Theorem 3.3, noting that for Π the random series prior in (3.19), it also holds that $E^\Pi\|F\|_{L^2}^2<\infty$.

Appendix 3.A Results for general inverse problems

Let $\mathcal{O}\subset\mathbb{R}^d$, $d\in\mathbb{N}$, be a nonempty open and bounded set with smooth boundary, and assume that $\mathcal{D}$ is a nonempty and bounded measurable subset of $\mathbb{R}^p$, $p\geq 1$. Let $\mathcal{F}\subseteq L^2(\mathcal{O})$ be endowed with the trace Borel σ-field of $L^2(\mathcal{O})$, and consider a Borel-measurable 'forward mapping'

$$\mathscr{G}:\mathcal{F}\to L^2(\mathcal{D}),\qquad F\mapsto\mathscr{G}(F).$$

For $F \in \mathcal{F}$, we are given noisy discrete measurements of $\mathscr{G}(F)$ over a grid of points drawn uniformly at random on $\mathcal{D}$,

$$Y_i = \mathscr{G}(F)(X_i) + \sigma W_i, \qquad i = 1, \dots, N, \quad X_i \overset{\text{iid}}{\sim} \mu, \quad W_i \overset{\text{iid}}{\sim} N(0,1), \tag{3.31}$$

for some $\sigma > 0$. Above μ denotes the uniform (probability) distribution on $\mathcal{D}$ and the design variables $(X_i)_{i=1}^N$ are independent of the noise vector $(W_i)_{i=1}^N$. We assume without loss of generality that $\mathrm{vol}(\mathcal{D}) = 1$, so that $\mu = dx$, the Lebesgue measure on $\mathcal{D}$.

We take the noise amplitude $\sigma > 0$ in (3.31) to be fixed and known, and work under the assumption that the forward map $\mathscr{G}$ satisfies the following local Lipschitz condition: for given $\beta, \gamma, \kappa \geq 0$, and all $F_1, F_2 \in C^\beta(\mathcal{O}) \cap \mathcal{F}$,

$$\|\mathscr{G}(F_1) - \mathscr{G}(F_2)\|_{L^2(\mathcal{D})} \lesssim (1 + \|F_1\|^\gamma_{C^\beta(\mathcal{O})} \vee \|F_2\|^\gamma_{C^\beta(\mathcal{O})}) \|F_1 - F_2\|_{(H^\kappa(\mathcal{O}))^*} \tag{3.32}$$

where we recall that X^* denotes the topological dual Banach space of a normed linear space X. Additionally, we will require $\mathscr{G}$ to be uniformly bounded on its domain,

$$S := \sup_{F \in \mathcal{F}} \|\mathscr{G}(F)\|_{L^\infty(\mathcal{D})} < \infty. \tag{3.33}$$

As observed in (3.23), the elliptic inverse problem considered in this chapter falls in this general framework, which also encompasses other examples of nonlinear inverse problems such as those involving the Schrödinger equation considered in [171, 176], for which the results in this section would apply as well. It also includes many linear inverse problems such as the classical Radon transform, see [176].

3.A.1 General contraction rates in Hellinger distance

Using the same notation as in Section 3.2.2, and given a sequence of Borel prior probability measures Π_N on $\mathcal{F}$, we write $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$ for the posterior distribution of $F|(Y^{(N)}, X^{(N)})$ (arising as after (3.9) and (3.10)). We also continue to use the notation p_F for the densities from (3.8) now in the general observation model (3.31) (and implicitly assume that the map $(F, (y, x)) \mapsto p_F(y, x)$ is jointly measurable to ensure existence of the posterior distribution). Below we formulate a general contraction theorem in the Hellinger distance that forms the basis of the proofs of the main results. It closely follows the general theory in [101] and its adaptation to the inverse problem setting in [167] – we include a proof for conciseness and convenience of the reader.

Define the Hellinger distance $h(\cdot,\cdot)$ on the set of probabilities density functions on $\mathbb{R}\times\mathcal{D}$ (with respect to the product measure $dy \times dx$) by

$$h^2(p_1,p_2) := \int_{\mathbb{R}\times\mathcal{D}} \left[\sqrt{p_1(y,x)} - \sqrt{p_2(y,x)}\right]^2 dydx.$$

For any set $\mathcal{A}$ of such densities, let $N(\eta;\mathcal{A},h),\ \eta>0$, be the minimal number of Hellinger balls of radius η needed to cover $\mathcal{A}$.

Theorem 3.7. *Let Π_N be a sequence of prior Borel probability measures on $\mathcal{F}$, and let $\Pi_N(\cdot|Y^{(N)},X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)},X^{(N)})$ in model (3.31). Assume that for some fixed $F_0\in\mathcal{F}$, and a sequence $\delta_N>0$ such that $\delta_N\to 0$ and $\sqrt{N}\delta_N\to\infty$ as $N\to\infty$, the sets*

$$\mathcal{B}_N := \left\{F:\ E^1_{F_0}\left[\log\frac{p_{F_0}(Y_1,X_1)}{p_F(Y_1,X_1)}\right]\le\delta_N^2,\ E^1_{F_0}\left[\log\frac{p_{F_0}(Y_1,X_1)}{p_F(Y_1,X_1)}\right]^2\le\delta_N^2\right\}, \tag{3.34}$$

satisfy for all N large enough

$$\Pi_N(\mathcal{B}_N)\ge a e^{-AN\delta_N^2}, \qquad \text{for some } a,A>0. \tag{3.35}$$

Further assume that there exists a sequence of Borel sets $\mathcal{A}_N\subset\mathcal{F}$ for which

$$\Pi_N(\mathcal{A}_N^c)\lesssim e^{-BN\delta_N^2}, \qquad \text{for some } B>A+2 \tag{3.36}$$

for all N large enough, as well as

$$\log N(\delta_N;\mathcal{A}_N,h)\le CN\delta_N^2, \qquad \text{for some } C>0. \tag{3.37}$$

Then, for sufficiently large $L=L(B,C)>4$ such that $L^2>12(B\vee C)$, and all $0<D<B-A-2$, as $N\to\infty$,

$$\Pi_N\left(F\in\mathcal{A}_N: h(p_F,p_{F_0})\le L\delta_N|Y^{(N)},X^{(N)}\right) = 1-O_{P^N_{F_0}}(e^{-DN\delta_N^2}). \tag{3.38}$$

Proof. We start noting that by Theorem 7.1.4 in [104], for each $L>4$ satisfying $L^2>12(B\vee C)$ we can find tests (random indicator functions) $\Psi_N=\Psi_N(Y^{(N)},X^{(N)})$ such that as $N\to\infty$

$$E^N_{F_0}\Psi_N\to 0, \qquad \sup_{F\in\mathcal{A}_N\,:\,h(p_F,p_{F_0})\ge L\delta_N} E^N_F(1-\Psi_N)\le e^{-BN\delta_N^2}. \tag{3.39}$$

Next, denote the set whose posterior probability we want to lower bound by

$$\tilde{\mathcal{A}}_N = \{F \in \mathcal{A}_N : h(p_F, p_{F_0}) \leq L\delta_N\};$$

and, using the first display in (3.39), decompose the probability of interest as

$$\begin{aligned}
&P^N_{F_0}\Big(\Pi_N(\tilde{\mathcal{A}}^c_N|Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}\Big)\\
&\quad= P^N_{F_0}\Big(\Pi_N(\tilde{\mathcal{A}}^c_N|Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0\Big)\\
&\qquad+ P^N_{F_0}\Big(\Pi_N(\tilde{\mathcal{A}}^c_N|Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 1\Big)\\
&\quad= P^N_{F_0}\Big(\Pi_N(\tilde{\mathcal{A}}^c_N|Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0\Big) + o(1).
\end{aligned}$$

Next, let $\nu(\cdot) = \Pi_N(\cdot \cap \mathcal{B}_N)/\Pi_N(\mathcal{B}_N)$ be the restricted normalised prior on $\mathcal{B}_N$, and define the event

$$\mathcal{C}_N = \Big\{\int_{\mathcal{B}_N} \prod_{i=1}^N \frac{p_F}{p_{F_0}}(Y_i, X_i)d\nu(F) \geq e^{-2N\delta_N^2}\Big\}, \tag{3.40}$$

for which Lemma 7.3.2 in [104] implies that $P^N_{F_0}(\mathcal{C}_N) \to 1$ as $N \to \infty$. We then further decompose

$$\begin{aligned}
&P^N_{F_0}\Big(\Pi_N(\tilde{\mathcal{A}}^c_N|Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0\Big)\\
&\qquad\qquad= P^N_{F_0}\Big(\Pi_N(\tilde{\mathcal{A}}^c_N|Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0, \mathcal{C}_N\Big) + o(1)
\end{aligned}$$

and in view of condition (3.35) and the above definition of $\mathcal{C}_N$, we see that

$$\begin{aligned}
&P^N_{F_0}\Big(\Pi_N(\tilde{\mathcal{A}}^c_N|Y^{(N)}, X^{(N)}) \geq e^{-DN\delta_N^2}, \Psi_N = 0, \mathcal{C}_N\Big)\\
&\quad= P^N_{F_0}\left(\frac{\int_{\tilde{\mathcal{A}}^c_N}\prod_{i=1}^N p_F/p_{F_0}(Y_i, X_i)d\Pi_N(F)}{\int_{\mathcal{F}}\prod_{i=1}^N p_F/p_{F_0}(Y_i, X_i)d\Pi_N(F)} \geq e^{-DN\delta_N^2}, \Psi_N = 0, \mathcal{C}_N\right)\\
&\quad\leq P^N_{F_0}\left(\frac{\int_{\tilde{\mathcal{A}}^c_N}(1-\Psi_N)\prod_{i=1}^N p_F/p_{F_0}(Y_i, X_i)d\Pi_N(F)}{\int_{\mathcal{B}_N}\prod_{i=1}^N p_F/p_{F_0}(Y_i, X_i)d\nu(F)} \geq \Pi_N(\mathcal{B}_N)e^{-DN\delta_N^2}, \mathcal{C}_N\right)\\
&\quad\leq P^N_{F_0}\left(\int_{\tilde{\mathcal{A}}^c_N}(1-\Psi_N)\prod_{i=1}^N \frac{p_F}{p_{F_0}}(Y_i, X_i)d\Pi_N(F) \geq ae^{-(A+D+2)N\delta_N^2}\right).
\end{aligned}$$

We conclude applying Markov's inequality and Fubini's theorem, jointly with the fact that for all $F \in \mathcal{F}$

$$E_{F_0}^N \Big[(1-\Psi_N)\prod_{i=1}^N \frac{p_F}{p_{F_0}}(Y_i, X_i)\Big] = E_{F_0}^N \Big[(1-\Psi_N)\prod_{i=1}^N \frac{dP_F^1}{dP_{F_0}^1}(Y_i, X_i)\Big] = E_F^N[1-\Psi_N],$$

to upper bound the last probability by

$$\begin{aligned}
&\frac{1}{a}e^{(A+D+2)N\delta_N^2}\Big(\int_{\mathcal{A}_N^c} E_F^N[1-\Psi_N]d\Pi_N(F) + \int_{\{F\in\mathcal{A}_N\,:\,h(p_{F_0},p_F)>L\delta_N\}} E_F^N[1-\Psi_N]d\Pi_N(F) \\
&\quad + \int_{\{F\in\mathcal{A}_N^c\,:\,h(p_{F_0},p_F)>L\delta_N\}} E_F^N[1-\Psi_N]d\Pi_N(F)\Big) \\
&\quad \le \frac{1}{a}e^{(A+D+2)N\delta_N^2}\Big(2\Pi_N(\mathcal{A}_N^c) + \int_{\{F\in\mathcal{A}_N\,:\,h(p_{F_0},p_F)>L\delta_N\}} E_F^N[1-\Psi_N]d\Pi_N(F)\Big) \\
&\quad \lesssim e^{-(B-A-D-2)N\delta_N^2} = o(1)
\end{aligned}$$

as $N \to \infty$ since $B > A + D + 2$, having used the excess mass condition (3.36) and the second display in (3.39).

□

3.A.2 Contraction rates for re-scaled Gaussian priors

While the previous result assumed a general sequence of priors, we now derive explicit contraction rates in L^2-prediction risk for the specific choices of priors considered in Section 3.3. We start with the 're-scaled' priors of Section 3.3.1.

Theorem 3.8. *Let the forward map $\mathscr{G}$ satisfy (3.32) and (3.33) for given $\beta, \gamma, \kappa, \ge 0$ and $S > 0$. For integer $\alpha > \beta + d/2$, consider a Gaussian prior Π_N constructed as in (3.11) with scaling $N^{d/(4\alpha+4\kappa+2d)}$ and with base prior $F' \sim \Pi'$ satisfying Condition 3.3 with RKHS $\mathcal{H}$. Let $\Pi_N(\cdot|Y^{(N)}, X^{(N)})$ be the resulting posterior arising from observations $(Y^{(N)}, X^{(N)})$ in (3.31), assume $F_0 \in \mathcal{H}$ and set $\delta_N = N^{-(\alpha+\kappa)/(2\alpha+2\kappa+d)}$.*

Then for any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, F_0, D, \alpha,$ and $\beta, \gamma, \kappa, S, d$) such that, as $N \to \infty$,

$$\Pi_N(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2(\mathcal{D})} > L\delta_N | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\delta_N^2}), \tag{3.41}$$

and for sufficiently large $M > 0$ (depending on $\sigma, D, \alpha, \beta, \gamma, \kappa, d$)

$$\Pi_N(F : \|F\|_{C^\beta} > M | Y^{(N)}, X^{(N)}) = O_{P_{F_0}^N}(e^{-DN\delta_N^2}). \tag{3.42}$$

Remark 3.1. Inspection of the proof (cfr. after (3.45)) shows that if $\kappa = 0$ in (3.32), then the RKHS $\mathcal{H}$ in Condition 3.3 can be assumed to be continuously imbedded in $H^\alpha(\mathcal{O})$ instead of $H_0^\alpha(\mathcal{O})$. The same remark in fact applies for $\kappa < 1/2$.

Proof. In view of the boundedness assumption (3.33) on $\mathscr{G}$, we have by Lemma 3.9 below that for some $q > 0$ (depending on σ, S)

$$E_{F_0}^1\left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)}\right] \vee E_{F_0}^1\left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)}\right]^2 \le q\|\mathscr{G}(F_0) - \mathscr{G}(F)\|_{L^2(\mathcal{D})}^2.$$

Hence, for $\mathcal{B}_N$ the sets from (3.34) we have $\{F : \|\mathscr{G}(F_0) - \mathscr{G}(F)\|_{L^2(\mathcal{D})} \le \delta_N/q\} \subseteq \mathcal{B}_N$, which in turn implies the small ball condition (3.35) since by Lemma 3.3 below (premultiplying, if needed, δ_N by a sufficiently large but fixed constant)

$$\Pi_N(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2(\mathcal{D})} \le \delta_N/q) \gtrsim e^{-AN\delta_N^2}$$

for some $A > 0$ and all N large enough. Next, for all $D > 0$ and any $B > A + D + 2$, we can choose sets $\mathcal{A}_N$ as in Lemmas 3.4 and 3.5 and verify the excess mass condition (3.36) as well as the complexity bound (3.37). Note that $\|F\|_{C^\beta} \le M$ for all $F \in \mathcal{A}_N$. We then conclude by Theorem 3.7 that for some $L' > 0$ large enough

$$\Pi_N(F \in \mathcal{A}_N : h(p_F, p_{F_0}) \le L'\delta_N | Y^{(N)}, X^{(N)}) = 1 - O_{P_{F_0}^N}(e^{-DN\delta_N^2})$$

yielding the claim for some appropriate $L > 0$ using the first inequality in (3.57).

□

The following key lemma shows that the (non-Gaussian) prior induced on the regression functions $\mathscr{G}(F)$ assigns sufficient mass to a L^2-neighbourhood of $\mathscr{G}(F_0)$.

Lemma 3.3. *Let Π_N, F_0 and δ_N be as in Theorem 3.8. Then, for sufficiently large $c > 0$ there exists $A > 0$ (depending on $c, F_0, \alpha, \beta, \gamma, \kappa, d$) such that*

$$\Pi_N(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2(\mathcal{D})} \le c\delta_N) \gtrsim e^{-AN\delta_N^2} \tag{3.43}$$

for all N large enough.

Proof. Since $F_0 \in \mathcal{H}$, $\|F_0\|_{C^\beta} < \infty$ by a Sobolev imbedding. Let $M > \|F_0\|_{C^\beta} \vee 1$ be a fixed constant. Using (3.32), we obtain for some $k > 0$

$$\begin{aligned}\Pi_N(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2(\mathcal{D})} \leq c\delta_N) \\ &\geq \Pi_N(F : \|F - F_0\|_{(H^\kappa)^*} \leq ckM^{-\gamma}\delta_N, \|F - F_0\|_{C^\beta} \leq M) \\ &= \Pi_N(F : F - F_0 \in C_1 \cap C_2),\end{aligned}$$

where

$$C_1 := \{F : \|F\|_{(H^\kappa)^*} \leq ckM^{-\gamma}\delta_N\}, \qquad C_2 := \{F : \|F\|_{C^\beta} \leq M\}.$$

Then, recalling that the RKHS $\mathcal{H}_N$ of Π_N coincides with $\mathcal{H}$ with RKHS norm $\|\cdot\|_{\mathcal{H}_N}$ given in (3.12), now with scaling $N^{d/(4\alpha+4\kappa+2d)} = \sqrt{N}\delta_N$, we can use Corollary 2.6.18 in [104] to lower bound the last probability by

$$\begin{aligned}e^{-\|F_0\|^2_{\mathcal{H}_N}/2}\Pi_N(C_1 \cap C_2) &= e^{-\frac{1}{2}N\delta_N^2\|F_0\|^2_{\mathcal{H}}}\Pi_N(C_1 \cap C_2) \\ &\geq e^{-\frac{1}{2}N\delta_N^2\|F_0\|^2_{\mathcal{H}}}\left(\Pi_N(C_1) - \Pi_N(C_2^c)\right)\end{aligned}$$

To upper bound $\Pi_N(C_2^c)$, note that by construction of Π_N

$$\Pi_N(C_2^c) = \Pr(\|F'\|_{C^\beta} > MN\delta_N^2), \qquad F' \sim \Pi'.$$

By Condition 3.3, F' defines a centred Gaussian Borel random element in a separable measurable subspace $\mathcal{C}$ of C^β, and by the Hahn-Banach theorem and the separability of $\mathcal{C}$, $\|F'\|_{C^\beta}$ can then be represented as a countable supremum

$$\|F'\|_{C^\beta} = \sup_{T \in \mathcal{T}} |T(F')|$$

of actions of bounded linear functionals $\mathcal{T} = (T_m)_{m\in\mathbb{N}} \subset (C^\beta)^*$. It follows that the collection $\{T_m(F')\}_{m\in\mathbb{N}}$ is a centred Gaussian process with almost surely finite supremum, so that by Fernique's theorem [104, Theorem 2.1.20]

$$E\|F'\|_{C^\beta} = E\sup_{m\in\mathbb{N}} |T_m(F')| < \infty; \qquad \tau^2 := \sup_{m\in\mathbb{N}} E|T_m(F')|^2 < \infty.$$

We then apply the Borell-Sudakov-Tirelson inequality [104, Theorem 2.5.8] to obtain for all N large enough,

$$\Pr\left(\|F'\|_{C^\beta} \geq M\sqrt{N}\delta_N\right) \leq \Pr\left(\|F'\|_{C^\beta} \geq E\|F'\|_{C^\beta} + M\sqrt{N}\delta_N/2\right) \leq e^{-\frac{1}{8}(M/\tau)^2N\delta_N^2}. \tag{3.44}$$

We proceed finding a lower bound for the prior probability of C_1, which, by construction of Π_N, satisfies

$$\Pi_N(F \in C_1) = \Pi'\left(F' : \|F'\|_{(H^\kappa)^*} \le ckM^{-\gamma}\sqrt{N}\delta_N^2\right).$$

For any integer $\alpha > 0$ and any $\kappa \ge 0$, letting $B_0^\alpha(r) := \{F \in H_0^\alpha,\ \|F\|_{H^\alpha} \le r\},\ r > 0$, we have the metric entropy estimate:

$$\log N(\eta; B_0^\alpha(r), \|\cdot\|_{(H^\kappa)^*}) \lesssim (r/\eta)^{d/(\alpha+\kappa)} \qquad \forall \eta > 0; \tag{3.45}$$

see the proof of Lemma 19 in [176] for the case $\kappa \ge 1/2$, and Theorem 4.10.3 in [227] for $\kappa < 1/2$ (in the latter case, we note in fact that the estimate holds true also for balls in the whole space H^α). Hence, since $\mathcal{H}$ is continuously imbedded into H_0^α, letting $B_{\mathcal{H}}(1)$ be the unit ball of $\mathcal{H}$, we have $B_{\mathcal{H}}(1) \subseteq B_0^\alpha(r)$ for some $r > 0$, implying that for all $\eta > 0$

$$\log N(\eta; B_{\mathcal{H}}(1), \|\cdot\|_{(H^\kappa)^*}) \le \log N(\eta; B_0^\alpha(r), \|\cdot\|_{(H^\kappa)^*}) \lesssim \eta^{-d/(\alpha+\kappa)}. \tag{3.46}$$

Then, for all N large enough, the small ball estimate in Theorem 1.2 in [152] yields

$$\begin{aligned} -\log \Pi'(F' : \|F'\|_{(H^\kappa)^*} &\le ckM^{-\gamma}\sqrt{N}\delta_N^2) \\ &\lesssim (cM^{-\gamma}\sqrt{N}\delta_N^2)^{-2\frac{d}{\alpha+\kappa}(2-d/(\alpha+\kappa))^{-1}} \\ &= (M^\gamma/c)^{\frac{2d}{2\alpha+2\kappa-d}}[N^{-(\alpha+\kappa-d/2)/(2\alpha+2\kappa+d)}]^{-\frac{2d}{2\alpha+2\kappa-d}} \\ &= (M^\gamma/c)^{\frac{2d}{2\alpha+2\kappa-d}} N\delta_N^2. \end{aligned}$$

Thus, for $k' > 0$ a fixed constant, we obtain the lower bound

$$\begin{aligned} \Pi_N(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2(\mathcal{D})} &\le c\delta_N) \\ &\ge e^{-\frac{1}{2}N\delta_N^2\|F_0\|_{\mathcal{H}}^2}\left(e^{-k'(M^\gamma/c)^{\frac{2d}{2\alpha+2\kappa-d}}N\delta_N^2} - e^{-\frac{1}{8}(M/\tau)^2 N\delta_N^2}\right) \\ &\gtrsim e^{-AN\delta_N^2} \end{aligned}$$

having taken $c > 0$ large enough (satisfying $k'(M^\gamma/c)^{\frac{2d}{2\alpha+2\kappa-d}} < \frac{1}{8}(M/\tau)^2)$, and where $A = \frac{1}{2}\|F_0\|_{\mathcal{H}}^2 + k'(M^\gamma/c)^{\frac{2d}{2\alpha+2\kappa-d}}$.

□

We now construct suitable approximating sets for which we check the excess mass condition (3.36).

Lemma 3.4. *Let Π_N and δ_N be as in Theorem 3.8. Define for any $M, Q > 0$*

$$\mathcal{A}_N = \left\{F : F = F_1 + F_2 : \|F_1\|_{(H^\kappa)^*} \le Q\delta_N,\ \|F_2\|_{\mathcal{H}} \le M,\ \|F\|_{C^\beta} \le M\right\}. \tag{3.47}$$

Then for any $B > 0$ and for sufficiently large M, Q (both depending on $B, \alpha, \beta, \gamma, \kappa, d$), for all N large enough,

$$\Pi_N(\mathcal{A}_N^c) \le 2e^{-BN\delta_N^2}. \tag{3.48}$$

Proof. By (3.44), taking $M \gtrsim \sqrt{B}$, we obtain for all N large enough that $\Pi_N(F : \|F\|_{C^\beta} \le M) \ge 1 - e^{-BN\delta_N^2}$. Thus, the claim will follow if we can derive a similar lower bound for

$$\begin{aligned}\Pi_N\Big(F : F = F_1 + F_2 : \|F_1\|_{(H^\kappa)^*} \le Q\delta_N, \|F_2\|_{\mathcal{H}} \le M\Big)\\ = \Pi'\left(F' : F' = F_1' + F_2',\ \|F_1'\|_{(H^\kappa)^*} \le Q\sqrt{N}\delta_N^2, \|F_2'\|_{\mathcal{H}} \le M\sqrt{N}\delta_N\right),\end{aligned}$$

having used that $N^{d/(4\alpha+4\kappa+d)} = \sqrt{N}\delta_N$. Using Theorem 1.2 in [152] as after (3.46), we deduce that for some $q > 0$

$$-\log \Pi'\left(F' : \|F'\|_{(H^\kappa)^*} \le Q\sqrt{N}\delta_N^2\right) \le q(Q\sqrt{N}\delta_N^2)^{-\frac{2d}{2\alpha+2\kappa-d}}$$

so that for any $Q > (B/q)^{-(2\alpha+2\kappa-d)/(2d)}$

$$-\log \Pi'\left(F' : \|F'\|_{(H^\kappa)^*} \le Q\sqrt{N}\delta_N^2\right) \le B(\sqrt{N}\delta_N^2)^{-\frac{2d}{2\alpha+2\kappa-d}} = BN\delta_N^2. \tag{3.49}$$

Next, denote by

$$M_N = -2\Phi^{-1}(e^{-BN\delta_N^2})$$

where Φ is the standard normal cumulative distribution function. Then by standard inequalities for Φ^{-1} we have $M_N \simeq \sqrt{BN}\delta_N$ as $N \to \infty$, so that taking $M \gtrsim \sqrt{B}$ implies

$$\begin{aligned}\Pi'\Big(F' : F' = F_1' + F_2',\ \|F_1'\|_{(H^\kappa)^*} \le Q\sqrt{N}\delta_N^2, \|F_2'\|_{\mathcal{H}} \le M\sqrt{N}\delta_N\Big)\\ \ge \Pi'\left(F' : F' = F_1' + F_2',\ \|F_1'\|_{(H^\kappa)^*} \le Q\sqrt{N}\delta_N^2, \|F_2'\|_{\mathcal{H}} \le M_N\right).\end{aligned}$$

By the isoperimetric inequality for Gaussian processes [104, Theorem 2.6.12], the last probability is then lower bounded, using (3.49), by

$$\Phi\left(\Phi^{-1}\left[\Pi'\left(F' : \|F'\|_{(H^\kappa)^*} \le Q\sqrt{N}\delta_N^2\right)\right] + M_N\right) \ge \Phi\left(\Phi^{-1}[e^{-BN\delta_N^2}] + M_N\right) = 1 - e^{-BN\delta_N^2},$$

concluding the proof.

□

We conclude with the verification of the complexity bound (3.37) for the sets $\mathcal{A}_N$.

Lemma 3.5. *Let $\mathcal{A}_N$ be as in Lemma 3.4 for some fixed $M, Q > 0$. Then,*

$$\log N(\delta_N; \mathcal{A}_N, h) \leq CN\delta_N^2,$$

for some constant $C > 0$ (depending on $\sigma, M, Q, \alpha, \beta, \gamma, \kappa, d, S$) and all N large enough.

Proof. If $F \in \mathcal{A}_N$, then $F = F_1 + F_2$ with $\|F_1\|_{(H^\kappa)^*} \leq Q\delta_N$ and $\|F_2\|_{H^\alpha} \leq M'$, the latter inequality following from the continuous imbedding of $\mathcal{H}$ into H_0^α. Thus, recalling the metric entropy estimate (3.45), if

$$\{H_1, \dots, H_P\} \subset B_0^\alpha(M'), \qquad P \leq e^{-q\delta_N^{-d/(\alpha+\kappa)}} = e^{-qN\delta_N^2}, \quad q > 0,$$

is a δ_N-net with respect to $\|\cdot\|_{(H^\kappa)^*}$, we can find H_i such that $\|F_2 - H_i\|_{(H^\kappa)^*} \leq \delta_N$. Then, using the second inequality in (3.57) below and the local Lipschitz estimate (3.32),

$$\begin{aligned} h(p_F, H_i) &\lesssim \|\mathscr{G}(F) - \mathscr{G}(H_i)\|_{L^2(\mathcal{D})} \\ &\lesssim (1 + \|F\|_{C^\beta}^\gamma \vee \|H_i\|_{C^\beta}^\gamma)\|F - H_i\|_{(H^\kappa)^*}. \end{aligned}$$

Recalling that if $F \in \mathcal{A}_N$ then also $\|F\|_{C^\beta} \leq M$, and using the Sobolev imbedding of H^α into C^β to bound $\|H_i\|_{C^\beta}$, we then obtain

$$h(p_F, H_i) \lesssim \|F - H_i\|_{(H^\kappa)^*} \lesssim \|F - F_2\|_{(H^\kappa)^*} + \|F_2 - H_i\|_{(H^\kappa)^*} \lesssim \delta_N.$$

It follows that $\{H_1, \dots, H_P\}$ also forms a $q'\delta_N$-net for $\mathcal{A}_N$ in the Hellinger distance for some $q' > 0$, so that

$$\log N(\delta_N; \mathcal{A}_N, h) \leq \log N(\delta_N/q'; B_0^\alpha(M), \|\cdot\|_{(H^\kappa)^*}) \lesssim N\delta_N^2.$$

□

3.A.3 Contraction rates for hierarchical Gaussian series priors

We now derive contraction rates in L^2-prediction risk in the inverse problem (3.31), for the truncated Gaussian random series priors introduced in Section 3.3.3. The proof again proceeds by an application of Theorem 3.7.

Theorem 3.9. *Let the forward map $\mathscr{G}$ satisfy (3.32) and (3.33) for given $\beta, \gamma, \kappa \geq 0$ and $S > 0$. For any $\alpha > \beta + d/2$, let Π be the random series prior in (3.19), and let $\Pi(\cdot|Y^{(N)}, X^{(N)})$ be the resulting posterior distribution arising from observations $(Y^{(N)}, X^{(N)})$ in (3.31). Then, for each $\alpha_0 \geq \alpha$, any $F_0 \in H_K^{\alpha_0}(\mathcal{O})$ and any $D > 0$ there exists $L > 0$ large enough (depending on $\sigma, F_0, D, \alpha, \beta, \gamma, \kappa, S, d$) such that, as $N \to \infty$,*

$$\Pi(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2(\mathcal{D})} > L\xi_N | Y^{(N)}, X^{(N)}) = O_{P^N_{F_0}}(e^{-DN\xi_N^2}), \tag{3.50}$$

where $\xi_N = N^{-(\alpha_0+\kappa)/(2\alpha_0+2\kappa+d)} \log N$. Moreover, for $\mathcal{H}_J$ the finite-dimensional subspaces from (3.18) and $J_N \in \mathbb{N}$ such that $2^{J_N} \simeq N^{1/(2\alpha_0+2\kappa+d)}$, we also have that for sufficiently large $M > 0$ (depending on D, α, β, d)

$$\Pi\left(F : F \in \mathcal{H}_{J_N},\ \|F\|_{H^\alpha} \leq M 2^{J_N \alpha} N\xi_N^2 | Y^{(N)}, X^{(N)}\right) = O_{P^N_{F_0}}(e^{-DN\xi_N^2}). \tag{3.51}$$

We begin deriving a suitable small ball estimate in the L^2-prediction risk.

Lemma 3.6. *Let Π, F_0 and ξ_N be as in Theorem 3.9. Then, for sufficiently large $q > 0$ there exists $A > 0$ (depending on $q, F_0, \alpha, \beta, \gamma, \kappa, d$) such that*

$$\Pi(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2(\mathcal{D})} \leq q\xi_N) \gtrsim e^{-AN\xi_N^2} \tag{3.52}$$

for all N large enough.

Proof. For each $j \in \mathbb{N}$, denote by Π_j the Gaussian probability measure on the finite dimensional subspace $\mathcal{H}_j$ in (3.18) defined as after (3.19) with the series truncated at j. For $J_N \in \mathbb{N} : 2^{J_N} \simeq N^{1/(2\alpha_0+2\kappa+d)}$, note

$$2^{J_N d} \log 2^{J_N d} \simeq N^{d/(2\alpha_0+2\kappa+d)} \log N = N\xi_N^2, \tag{3.53}$$

so that, recalling the properties (3.20) of the random truncation level J, for some $s > 0$,

$$\Pr(K = J_N) \gtrsim e^{-2^{J_N d} \log 2^{J_N d}} \geq e^{-sN\xi_N^2}$$

for all N large enough. It follows

$$\begin{aligned}\Pi(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2} \leq q\xi_N) &\geq \Pi_{J_N}(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2} \leq q\xi_N) \Pr(K = J_N)\\ &\gtrsim \Pi_{J_N}(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2} \leq q\xi_N) e^{-sN\xi_N^2}.\end{aligned}$$

Next, let

$$P_{J_N}F_0 = \chi \sum_{l\le J_N, r\in\mathcal{R}_l} \langle F_0, \Phi_{lr}\rangle_{L^2}\Phi_{lr}$$

be the ‘projection’ of F_0 onto $\mathcal{H}_{J_N}$. Since $F_0 \in H_K^{\alpha_0} \subset C^\beta$ by a Sobolev imbedding, it follows using (3.32) and standard approximation properties of wavelets (cfr. (3.63)),

$$\|\mathscr{G}(F_0) - \mathscr{G}(P_{J_N}F_0)\|_{L^2(\mathcal{D})} \lesssim \|F_0 - P_{J_N}F_0\|_{(H^\kappa)^*} \lesssim 2^{-J_N(\alpha_0+\kappa)} = N^{-\frac{\alpha_0+\kappa}{2\alpha_0+2\kappa+d}},$$

which implies by the triangle inequality that

$$\begin{aligned}
&\Pi_{J_N}(F : \|\mathscr{G}(F) - \mathscr{G}(F_0)\|_{L^2} \le q\xi_N) \\
&\quad \ge \Pi_{J_N}(F : \|\mathscr{G}(F) - \mathscr{G}(P_{J_N}F_0)\|_{L^2} \le q\xi_N - \|\mathscr{G}(F_0) - \mathscr{G}(P_{J_N}F_0)\|_{L^2}) \\
&\quad \ge \Pi_{J_N}(F : \|\mathscr{G}(F) - \mathscr{G}(P_{J_N}F_0)\|_{L^2} \le q'\xi_N).
\end{aligned}$$

Using again that H^α imbeds continuously into C^β as well as (3.32) and (4.30), we can lower bound the last probability by

$$\begin{aligned}
&\Pi_{J_N}(F : \|\mathscr{G}(F) - \mathscr{G}(P_{J_N}F_0)\|_{L^2(\mathcal{D})} \le q'\xi_N, \|F - P_{J_N}F_0\|_{H^\alpha(\mathcal{O})} \le \xi_N) \\
&\quad \ge \Pi_{J_N}(F : \|F - P_{J_N}F_0\|_{(H^\kappa(\mathcal{O}))^*} \le q''\xi_N, \|F - P_{J_N}F_0\|_{H^\alpha(\mathcal{O})} \le \xi_N) \\
&\quad \ge \Pi_{J_N}(F : \|F - P_{J_N}F_0\|_{H^\alpha(\mathcal{O})} \le q'''\xi_N),
\end{aligned}$$

which, by Corollary 2.6.18 in [104] and in view of (3.59) is further lower bounded by

$$e^{-\frac{1}{2}\|P_{J_N}F_0\|^2_{\mathcal{H}_{J_N}}}\Pi_{J_N}(F : \|F\|_{H^\alpha} \le q'''\xi_N) \ge e^{-s'\|F_0\|^2_{H^{\alpha_0}}}\Pi_{J_N}(F : \|F\|_{H^\alpha} \le q'''\xi_N).$$

Now since $f \mapsto \chi f,\ \chi \in C^\infty(\mathcal{O})$, is continuous on $H^\alpha(\mathcal{O})$,

$$\begin{aligned}
\Pi_{J_N}(F : \|F\|_{H^\alpha} \le q'''\xi_N) &= \Pr\left(\Big\|\chi \sum_{l\le J_N, r\in\mathcal{R}_l} 2^{-l\alpha}F_{lr}\Phi_{lr}\Big\|_{H^\alpha} \le q'''\xi_N\right) \\
&\ge \Pr\left(\sum_{m=1}^{\dim(\mathcal{H}_{J_N})} Z_m^2 \le t\xi_N^2\right)
\end{aligned}$$

for some $t > 0$, where $Z_m \overset{\text{iid}}{\sim} N(0,1)$, and where we have used the wavelet characterisation of the $H^\alpha(\mathbb{R}^d)$ norm. To conclude, note that the last probability is greater than

$$\begin{aligned}
&\Pr\Big(\sqrt{\dim(\mathcal{H}_{J_N})}\max_{m\le\dim(\mathcal{H}_{J_N})}|Z_m| \le \sqrt{t}\xi_N\Big)\\
&\quad\ge \Pr\Big(\max_{m\le\dim(\mathcal{H}_{J_N})}|Z_m| \le t' N^{-\frac{\alpha_0+\kappa}{2\alpha_0+2\kappa+d}} N^{-\frac{d/2}{2\alpha_0+2\kappa+d}}\Big)\\
&\quad= \prod_{m\le\dim(\mathcal{H}_{J_N})}\Pr\Big(|Z_m| \le t' N^{-\frac{\alpha_0+\kappa+d/2}{2\alpha_0+2\kappa+d}}\Big).
\end{aligned}$$

Finally, a standard calculation shows that $\Pr(|Z_1|\le t)\gtrsim t$ if $t\to 0$, and hence the last product is lower bounded, for large N, by

$$\begin{aligned}
\Big(t' N^{-\frac{\alpha_0+\kappa+d/2}{2\alpha_0+2\kappa+d}}\Big)^{\dim(\mathcal{H}_{J_N})} &= e^{\dim(\mathcal{H}_{J_N})\log\Big(t' N^{-\frac{\alpha_0+\kappa+d/2}{2\alpha_0+2+d}}\Big)}\\
&\ge e^{-t'' 2^{J_N d}\log N}\\
&= e^{-t''' N\xi_N^2}.
\end{aligned}$$

□

In the following lemma we construct suitable approximating sets, for which we check the excess mass condition (3.36) and the complexity bound (3.37) required in Theorem 3.7.

Lemma 3.7. *Let Π, ξ_N and J_N be as in Theorem 3.4, and let $\mathcal{H}_{J_N}$ be the finite dimensional subspace defined in (3.18) with $J = J_N$. Define for each $M > 0$*

$$\mathcal{A}_N = \Big\{F \in \mathcal{H}_{J_N},\ \|F\|_{H^\alpha} \le M 2^{J_N\alpha} N\xi_N^2\Big\}. \tag{3.54}$$

Then, for any $B > 0$ there exists $M > 0$ large enough (depending on B, α, β, d) such that, for sufficiently large N

$$\Pi(\mathcal{A}_N^c) \le 2e^{-BN\xi_N^2}. \tag{3.55}$$

Moreover, for each fixed $M > 0$ and all N large enough

$$\log N(\xi_N; \mathcal{A}_N, h) \le CN\xi_N^2 \tag{3.56}$$

for some $C > 0$ (depending on $\sigma, \alpha, \beta, \gamma, \kappa, S, d$).

Proof. Letting $Z_m \overset{\text{iid}}{\sim} N(0,1)$, noting $\|F\|^2_{H^\alpha} \le 2^{2J_N\alpha}\sum_{l\le J_N, r\in\mathcal{R}_l} F^2_{lr}$ for all $F\in\mathcal{H}_{J_N}$ (cfr. (3.59)) and using (3.53) and (3.20), we have for sufficiently large N

$$\begin{aligned}\Pi(\mathcal{A}_N^c) &\le \Pr(J > J_N) + \Pr\left(\sum_{l\le J\wedge J_N, r\in\mathcal{R}_l} F^2_{lr} \le MN\xi_N^2\right)\\ &\le e^{-2^{J_N d}\log 2^{J_N d}} + \Pr\left(\sum_{m\le \dim(\mathcal{H}_{J_N})} Z^2_m > MN\xi_N^2\right)\\ &\le e^{-BN\xi_N^2} + \Pr\left(\sum_{m\le \dim(\mathcal{H}_{J_N})} (Z^2_m - 1) > \bar{M}N\xi_N^2\right)\end{aligned}$$

for any constant $0<\bar{M}<M^2-1$, since $\dim(\mathcal{H}_{J_N}) \lesssim 2^{J_N d} \simeq N^{d/(2\alpha+2+d)} = o(N\xi_N^2)$. The bound (3.55) then follows applying Theorem 3.1.9 in [104] to upper bound the last probability, for any $B>$ and for sufficiently large M and $\bar{M}$, by

$$e^{-\frac{\bar{M}^2(N\xi_N^2)^2}{4\dim(\mathcal{H}_{J_N})+\bar{M}N\xi_N^2}} \le e^{-BN\xi_N^2}.$$

We proceed with the derivation of (3.56). By choice of J_N, if $F\in\mathcal{A}_N$ then $\|F\|^2_{H^\alpha} \lesssim N^{(2\alpha)/(2\alpha+2\kappa+d)}N\xi_N^2$. Hence, by the second inequality in (3.57), using (3.32) and the Sobolev imbedding of H^α into C^β, if $F_1, F_2\in\mathcal{A}_N$ then

$$\begin{aligned}h(p_{F_1}, p_{F_2}) &\lesssim \|\mathscr{G}(F_1) - \mathscr{G}(F_2)\|_{L^2(\mathcal{D})}\\ &\lesssim (1 + (N^{\frac{\alpha}{2\alpha+2\kappa+d}}\sqrt{N}\xi_N)^\gamma)\|F_1 - F_2\|_{(H^\kappa)^*}\\ &\lesssim N^{\frac{\alpha\gamma}{2\alpha+2\kappa+d}}(\sqrt{N}\xi_N)^\gamma\sqrt{\sum_{l\le J_N, r\in\mathcal{R}_l}(F_{1,lr} - F_{2,lr})^2}.\end{aligned}$$

Therefore, using the standard metric entropy estimate for balls $B_{\mathbb{R}^p}(r)$, $r>0$, in Euclidean spaces [104, Proposition 4.3.34], we see that for N large enough

$$\begin{aligned}\log N(\xi_N;\mathcal{A}_N, h) &\lesssim \log N\left(\xi_N N^{\frac{-\alpha\gamma}{2\alpha+2\kappa+d}}(\sqrt{N}\xi_N)^{-\gamma}; B_{\mathbb{R}^{\dim(\mathcal{H}_{J_N})}}(M\sqrt{N}\xi_N), \|\cdot\|_{\mathbb{R}^{\dim(\mathcal{H}_{J_N})}}\right)\\ &\le \dim(\mathcal{H}_{J_N})\log\frac{3M\sqrt{N}\xi_N}{\xi_N N^{\frac{\alpha\gamma}{2\alpha+2\kappa+d}}(\sqrt{N}\xi_N)^{-\gamma}}\\ &\lesssim N\xi_N^2.\end{aligned}$$

□

3.A.4 Information theoretic inequalities

In the following lemma (due to [30]) we exploit the boundedness assumption (3.33) on $\mathscr{G}$ to show the equivalence between the Hellinger distance appearing in the conclusion of Theorem 3.7 and the L^2-distance on the 'regression functions' $\mathscr{G}(F)$.

Lemma 3.8. *Let the forward map* $\mathscr{G}$ *satisfy* (3.33) *for some* $S > 0$*. Then, for all* $F_1, F_2 \in \mathcal{F}$

$$\frac{1 - e^{-S^2/(2\sigma^2)}}{4S^2} \|\mathscr{G}(F_1) - \mathscr{G}(F_2)\|^2_{L^2(\mathcal{D})} \le h^2(p_{F_1}, p_{F_2}) \le \frac{1}{4\sigma^2} \|\mathscr{G}(F_1) - \mathscr{G}(F_2)\|^2_{L^2(\mathcal{D})}. \tag{3.57}$$

Proof. Note $h^2(p_{F_1}, p_{F_2}) = 2 - 2\rho(p_{F_1}, p_{F_2})$, where

$$\rho(p_{F_1}, p_{F_2}) := \int_{\mathbb{R}\times\mathcal{D}} \sqrt{p_{F_1}(y,x) p_{F_2}(y,x)} dy dx$$

is the Hellinger affinity. Using the expression of the likelihood in (3.8) (with $\mathcal{D}$ instead of $\mathcal{O}$), the right hand side is seen to be equal to

$$\begin{aligned}
&\int_{\mathbb{R}\times\mathcal{D}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\{[y-\mathscr{G}(F_1)(x)]^2 - [y-\mathscr{G}(F_2)(x)]^2\}/(4\sigma^2)} dy dx \\
&\quad = \int_{\mathcal{D}} e^{-\{[\mathscr{G}(F_1)(x)]^2 + [\mathscr{G}(F_2)(x)]^2\}/(4\sigma^2)} \Big[\int_{\mathbb{R}} \frac{e^{-y^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} e^{y[\mathscr{G}(F_1)(x) + \mathscr{G}(F_2)(x)]/(2\sigma^2)} dy \Big] dx \\
&\quad = \int_{\mathcal{D}} e^{-\{[\mathscr{G}(F_1)(x)]^2 + [\mathscr{G}(F_2)(x)]^2\}/(4\sigma^2)} e^{[\mathscr{G}(F_1)(x) + \mathscr{G}(F_2)(x)]^2/(8\sigma^2)} dx
\end{aligned}$$

having used that the moment generating function of $Z \sim N(0, \sigma^2)$ satisfies $Ee^{tZ} = e^{\sigma^2 t^2/2}$, $t \in \mathbb{R}$. Thus, the latter integral equals

$$\int_{\mathcal{D}} e^{-\{[\mathscr{G}(F_1)(x)]^2 + [\mathscr{G}(F_2)(x)]^2 - 2\mathscr{G}(F_2)(x)\mathscr{G}(F_2)(x)\}/(8\sigma^2)} dx = E^\mu e^{-\{\mathscr{G}(F_1)(X) - \mathscr{G}(F_2)(X)\}^2/(8\sigma^2)}.$$

To derive the second inequality in (3.57), we use Jensen's inequality to lower bound the expectation in the last line by

$$e^{-E^\mu\{\mathscr{G}(F_1)(X) - \mathscr{G}(F_2)(X)\}^2/(8\sigma^2)} = e^{-\|\mathscr{G}(F_1) - \mathscr{G}(F_2)\|^2_{L^2(\mathcal{D})}/(8\sigma^2)}.$$

Hence

$$h^2(p_{F_1}, p_{F_2}) \le 2\Big[1 - e^{-\|\mathscr{G}(F_1) - \mathscr{G}(F_2)\|^2_{L^2(\mathcal{D})}/(8\sigma^2)}\Big],$$

whereby the claim follows using the basic inequality $1 - e^{-z/c} \le z/c$, for all $c, z > 0$.

To deduce the first inequality we follow the proof of Proposition 1 in [30]: note that for all $0 \leq z_1 < z_2$

$$e^{-z_1} \leq \frac{z_1}{z_2} e^{-z_2} + \left(1 - \frac{z_1}{z_2}\right) = \frac{e^{-z_2} - 1}{z_2} z_1 + 1.$$

Then taking $z_1 = \{\mathscr{G}(F_1)(X) - \mathscr{G}(F_2)(X)\}^2/(8\sigma^2)$ and $z_2 = S^2/(2\sigma^2)$,

$$E^\mu e^{-\{\mathscr{G}(F_1)(X) - \mathscr{G}(F_2)(X)\}^2/(8\sigma^2)} \leq \frac{e^{-S^2/(2\sigma^2)} - 1}{4S^2} \|\mathscr{G}(F_1) - \mathscr{G}(F_2)\|^2_{L^2(\mathcal{D})} + 1$$

which in turn yields the result.

□

The next lemma bounds the Kullback-Leibler divergences appearing in (3.34) in terms of the L^2-prediction risk.

Lemma 3.9. *Let the observation Y_i in (3.31) be generated by some fixed $F_0 \in \mathcal{F}$. Then, for each $F \in \mathcal{F}$,*

$$E^1_{F_0}\left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)}\right] = \frac{1}{\sigma^2} \|\mathscr{G}(F_0) - \mathscr{G}(F)\|^2_{L^2(\mathcal{D})},$$

and

$$E^1_{F_0}\left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)}\right]^2 \leq \frac{2(S^2 + \sigma^2)}{\sigma^4} \|\mathscr{G}(F_0) - \mathscr{G}(F)\|^2_{L^2(\mathcal{D})}.$$

Proof. If $Y_1 = \mathscr{G}(F_0)(X_1) + \sigma W_1$, then

$$\begin{aligned}
&\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)} \\
&\quad = -\frac{1}{2\sigma^2}\{[\mathscr{G}(F_0)(X_1) + \sigma W_1 - \mathscr{G}(F_0)(X_1)]^2 - [\mathscr{G}(F_0)(X_1) + \sigma W_1 - \mathscr{G}(F)(X_1)]^2\} \\
&\quad = \frac{1}{2\sigma^2}\{\mathscr{G}(F_0)(X_1) - \mathscr{G}(F)(X_1)\}^2 + \frac{1}{\sigma} W_1 \{\mathscr{G}(F_0)(X_1) - \mathscr{G}(F)(X_1)\}.
\end{aligned}$$

Hence, since $EW_1 = 0$ and $X_1 \sim \mu$,

$$\begin{aligned}
E^1_{F_0}\left[\log \frac{p_{F_0}(Y_1, X_1)}{p_F(Y_1, X_1)}\right] &= E^\mu\left[\frac{1}{2\sigma^2}\{\mathscr{G}(F_0)(X_1) - \mathscr{G}(F)(X_1)\}^2\right] \\
&= \frac{1}{2\sigma^2} \|\mathscr{G}(F_0) - \mathscr{G}(F)\|^2_{L^2(\mathcal{D})}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\Big[\log\frac{p_{F_0}(Y_1,X_1)}{p_F(Y_1,X_1)}\Big]^2\\
&\quad=\Big[\frac{1}{2\sigma^2}\{\mathscr{G}(F_0)(X_1)-\mathscr{G}(F)(X_1)\}^2+\frac{1}{\sigma}W_1\{\mathscr{G}(F_0)(X_1)-\mathscr{G}(F)(X_1)\}\Big]^2\\
&\quad\le 2\Big[\frac{1}{2\sigma^2}\{\mathscr{G}(F_0)(X_1)-\mathscr{G}(F)(X_1)\}^2\Big]^2+2\Big[\frac{1}{\sigma}W_1\{\mathscr{G}(F_0)(X_1)-\mathscr{G}(F)(X_1)\}\Big]^2\\
&\quad=\frac{2S^2}{\sigma^4}\{\mathscr{G}(F_0)(X_1)-\mathscr{G}(F)(X_1)\}^2+\frac{2}{\sigma^2}W_1^2\{\mathscr{G}(F_0)(X_1)-\mathscr{G}(F)(X_1)\}^2,
\end{aligned}$$

whence the second claim follows since $EW_1^2=1$.

□

Appendix 3.B Additional background material

In this final appendix we collect some standard materials used in the proofs for convenience of the reader.

Example 3.1. Take

$$\phi:\mathbb{R}\to(0,\infty),\qquad \phi(t)=\frac{1}{1-t}1_{\{t<0\}}+(1+t)1_{\{t\ge 0\}},$$

and let $\psi:\mathbb{R}\to[0,\infty)$ be a smooth compactly supported function such that $\int_{\mathbb{R}}\psi(t)dt=1$. Define for any $K_{min}\in(0,1)$

$$\Phi(t)=K_{min}+\frac{1-K_{min}}{\psi*\phi(0)}\psi*\phi(t),\qquad t\in\mathbb{R}. \tag{3.58}$$

Then it is elementary to check that Φ is a regular link function that satisfies Condition 3.2 (with $a=2$).

Example 3.2. For any real $\alpha>d/2$, the Matérn process with index set $\mathcal{O}$ and regularity $\alpha-d/2>0$ (cfr. Example 11.8 in [101]) is the stationary centred Gaussian process $M=\{M(x),\ x\in\mathcal{O}\}$ with covariance kernel

$$C(x,y)=\int_{\mathbb{R}^d}e^{-i\langle x-y,\xi\rangle_{\mathbb{R}^d}}\mu(d\xi),\qquad \mu(d\xi)=(1+\|\xi\|^2_{\mathbb{R}^d})^{-\alpha}d\xi,\quad x,y\in\mathcal{O}.$$

From the results in Chapter 11 in [101] we see that the RKHS of $(M(x):x\in\mathcal{O})$ equals the set of restrictions to $\mathcal{O}$ of elements in the Sobolev space $H^\alpha(\mathbb{R}^d)$, which equals, with equivalent norms, the space $H^\alpha(\mathcal{O})$ (since $\mathcal{O}$ has a smooth boundary). Moreover, Lemma I.4 in [101] shows that M has a version with paths belonging almost surely to $C^{\beta'}$ for all $\beta' < \alpha - d/2$. Let now $K \subset \mathcal{O}$ be a nonempty compact set, and let M be a $C^{\beta'}$-smooth version of a Matérn process on $\mathcal{O}$ with RKHS $H^\alpha(\mathcal{O})$. Taking $F' = \chi M$ implies (cfr. Exercise 2.6.5 in [104]) that $\Pi' = \mathcal{L}(F')$ defines a centred Gaussian probability measure supported on $C^{\beta'}$, whose RKHS is given by

$$\mathcal{H} = \{\chi F,\ F \in H^\alpha(\mathcal{O})\},$$

and the RKHS norm satisfies that for all $F \in H^\alpha(\mathcal{O})$ there exists $F^* \in H^\alpha(\mathcal{O})$ such that $\chi F = \chi F^*$ and

$$\|\chi F\|_{\mathcal{H}} = \|F^*\|_{H^\alpha(\mathcal{O})}.$$

Thus if $F' = \chi F$ is an arbitrary element of $\mathcal{H}$, then

$$\|F'\|_{H^\alpha} = \|\chi F^*\|_{H^\alpha} \lesssim \|F^*\|_{H^\alpha} = \|F'\|_{\mathcal{H}},$$

which shows that $\mathcal{H}$ is continuously embedded into $H^\alpha_0(\mathcal{O})$.

Remark 3.2. Let $\{\Phi_{lr},\ l \geq -1, r \in \mathbb{Z}^d\}$ be an orthonormal basis of $L^2(\mathbb{R}^d)$ composed of S-regular and compactly supported Daubechies wavelets (see Chapter 4 in [104] for construction and properties). For each $0 \leq \alpha \leq S$, we have

$$H^\alpha(\mathbb{R}^d) = \Big\{F \in L^2(\mathbb{R}^d) : \sum_{l,r} 2^{2l\alpha}\langle F, \Phi_{lr}\rangle^2_{L^2(\mathbb{R}^d)} < \infty\Big\},$$

and the square root of the latter series defines an equivalent norm to $\|\cdot\|_{H^\alpha(\mathbb{R}^d)}$. Note that $S > 0$ can be taken arbitrarily large.

For any $\alpha \geq 0$ the Gaussian random series

$$\bar{F}_j = \sum_{l\leq j, r\in\mathcal{R}_l} F_{lr} 2^{-l\alpha}\Phi_{lr}, \qquad F_{lr} \overset{\text{iid}}{\sim} N(0,1)$$

defines a centred Gaussian probability measure supported on the finite-dimensional space $\bar{\mathcal{H}}_j$ spanned by the $\{\Phi_{lr}, l \leq j, r \in \mathcal{R}_l\}$, and its RKHS equals $\bar{\mathcal{H}}_j$ endowed with norm

$$\|\bar{H}_j\|^2_{\bar{\mathcal{H}}_j} = \sum_{l\leq j, r\in\mathcal{R}_l} 2^{2l\alpha} H^2_{lr} = \|\bar{H}_j\|^2_{H^\alpha(\mathbb{R}^d)} \qquad \forall \bar{H}_j \in \bar{\mathcal{H}}_j$$

(cfr. Example 2.6.15 in [104]). Basic wavelet theory implies $\dim(\bar{\mathcal{H}}_j) \lesssim 2^{jd}$.

If we now fix compact $K' \subset \mathcal{O}$ such that $K \subsetneq K'$, and consider a cut-off function $\chi \in C^\infty_c(\mathcal{O})$ such that $\chi = 1$ on K', then multiplication by χ is a bounded linear operator

$\chi : H^s(\mathbb{R}^d) \to H_0^s(\mathcal{O})$. It follows that the random function

$$F_j = \chi(\bar{F}_j) = \sum_{l\leq j, r\in\mathcal{R}_l} F_{lr} 2^{-l\alpha} \chi\Phi_{lr}, \qquad F_{lr} \overset{\text{iid}}{\sim} N(0,1)$$

defines, according to Exercise 2.6.5 in [104], a centred Gaussian probability measure $\Pi_j = \mathcal{L}(F_j)$ supported on the finite dimensional subspace $\mathcal{H}_j$ from (3.18), with RKHS norm satisfying

$$\Big\|\chi\Big(\sum_{l\leq j, r\in\mathcal{R}_l} H_{lr}\Phi_{lr}\Big)\Big\|_{\mathcal{H}_j} \leq \Big\|\sum_{l\leq j, r\in\mathcal{R}_l} H_{lr}\Phi_{lr}\|_{\bar{\mathcal{H}}_j} = \sqrt{\sum_{l\leq j, r\in\mathcal{R}_l} 2^{2l\alpha} H_{lr}^2}. \tag{3.59}$$

Arguing as in the previous remark one shows further that for some constant $c > 0$,

$$\|H_j\|_{H^\alpha(\mathcal{O})} \leq c\|H_j\|_{\mathcal{H}_j} \qquad \forall H_j \in \mathcal{H}_j. \tag{3.60}$$

Remark 3.3. Using the notation of the previous remark, for fixed $F_0 \in H_K^\alpha(\mathcal{O})$, consider the finite-dimensional approximations

$$P_J F_0 = \sum_{l\leq j, r\in\mathcal{R}_l} \langle F_0, \Phi_{lr}\rangle_{L^2} \chi\Phi_{lr} \in \mathcal{H}_j, \qquad j \in \mathbb{N}. \tag{3.61}$$

Then in view of (3.59), we readily check that for each $j \geq 1$

$$\|P_J F_0\|_{\mathcal{H}_j} \leq \sqrt{\sum_{l\leq j, r\in\mathcal{R}_l} 2^{2l\alpha} \langle F_0, \Phi_{lr}\rangle_{L^2}^2} \leq \|F_0\|_{H^\alpha(\mathcal{O})} < \infty. \tag{3.62}$$

Also, for each $\kappa \geq 0$, and any $G \in H^\kappa(\mathcal{O})$, we see that (implicitly extending to 0 on $\mathbb{R}^d \backslash \mathcal{O}$ functions that are compactly supported inside $\mathcal{O}$)

$$\langle F_0 - P_J F_0, G\rangle_{L^2(\mathcal{O})} = \langle F_0 - P_J F_0, \chi' G\rangle_{L^2(\mathbb{R}^d)}$$

where $\chi' \in C_c^\infty(\mathcal{O})$, with $\chi' = 1$ on $\text{supp}(\chi)$. We also note that, in view of the localisation properties of Daubechies wavelets, for some $J_{min} \in \mathbb{N}$ large enough, if $l \geq J_{min}$ and the support of Φ_{lr} intersects K, then necessarily $\text{supp}(\Phi_{lr}) \subseteq K'$, so that

$$\chi\Phi_{lr} = \Phi_{lr} \qquad \forall l \geq J_{min},\ r \in \mathcal{R}_l.$$

Therefore, for $j \geq J_{min}$, by Parseval's identity and the Cauchy-Schwarz inequality

$$\begin{aligned}
&\langle F_0 - P_J F_0, \chi' G\rangle_{L^2(\mathbb{R}^d)} \\
&\quad = \sum_{l'>j, r' \in \mathcal{R}_l} 2^{l\alpha} \langle F_0, \Phi_{l'r'}\rangle_{L^2(\mathbb{R}^d)} 2^{l'\kappa} \langle \chi' G, \Phi_{l'r'}\rangle_{L^2(\mathbb{R}^d)} 2^{-l'(\alpha+\kappa)} \\
&\quad \leq 2^{-j(\alpha+\kappa)} \sqrt{\sum_{l'>j, r' \in \mathcal{R}_l} 2^{2l\alpha} \langle F_0, \Phi_{l'r'}\rangle^2_{L^2(\mathbb{R}^d)}} \sqrt{\sum_{l'>j, r' \in \mathcal{R}_l} 2^{2l\kappa} \langle \chi' G, \Phi_{l'r'}\rangle^2_{L^2(\mathbb{R}^d)}} \\
&\quad \leq 2^{-j(\alpha+\kappa)} \|F_0\|_{H^\alpha(\mathcal{O})} \|\chi' G\|_{H^\kappa(\mathbb{R}^d)}.
\end{aligned}$$

It follows by duality that for all j large enough

$$\|F_0 - P_j F_0\|_{(H^\kappa(\mathcal{O}))^*} \lesssim 2^{-j(\alpha+\kappa)} \|F_0\|_{H^\alpha(\mathcal{O})}. \tag{3.63}$$

We conclude remarking that

$$\|F\|_{H^\alpha(\mathcal{O})} \lesssim 2^{j\alpha} \|F\|_{L^2(\mathcal{O})}, \qquad \forall F \in \mathcal{H}_j,\ j \geq J_{min}. \tag{3.64}$$

Indeed, let $j \geq J_{min}$, and fix $F \in \mathcal{H}_j$; then

$$F = P_{J_{min}} F + (F - P_{J_{min}} F) = \sum_{l \leq J_{min}, r \in \mathcal{R}_l} F_{lr} \chi \Phi_{lr} + \sum_{J_{min} < l \leq j, r \in \mathcal{R}_l} F_{lr} \Phi_{lr}.$$

But as $\mathcal{H}_{J_{min}}$ is a fixed finite dimensional subspace, then we have $\|P_{J_{min}} F\|_{H^s(\mathcal{O})} \lesssim \|P_{J_{min}} F\|_{L^2(\mathcal{O})} \leq \|F\|_{L^2(\mathcal{O})}$ for some fixed multiplicative constant only depending on J_{min}. On the other hand, we also have

$$\begin{aligned}
\|F - P_{J_{min}} F\|^2_{H^\alpha(\mathcal{O})} &= \sum_{J_{min} < l \leq j, r \in \mathcal{R}_l} 2^{2l\alpha} F^2_{lr} \\
&\leq 2^{2j\alpha} \|F - P_{J_{min}} F\|^2_{L^2(\mathcal{O})} \\
&\leq 2^{2j\alpha} \|F\|^2_{L^2(\mathcal{O})},
\end{aligned}$$

yielding (3.64).

Example 3.3. Consider the integer-valued random variable

$$J = \lfloor \log_2(\phi^{-1}(T)^{1/d}) \rfloor + 1, \qquad T \sim \mathrm{Exp}(1),$$

where $\phi(x) = x \log x,\ x \geq 1$. Then for any $j \geq 1$

$$\Pr(J > j) = \Pr(\phi^{-1}(T) \geq 2^{jd}) = \Pr(T \geq 2^{jd} \log 2^{jd}) = e^{-2^{jd} \log 2^{jd}}.$$

On the other hand, since $e^{-2^{jd}(1-2^{-d})\log 2^{(j-1)d}} \to 0$ as $j \to \infty$,

$$\begin{aligned}
\Pr(J = j) &= \Pr\left(2^{(j-1)d} \leq \phi^{-1}(T) < 2^{jd}\right) \\
&= e^{-2^{(j-1)d}\log 2^{(j-1)d}} + 1 - e^{-2^{jd}\log 2^{jd}} - 1 \\
&\geq e^{-2^{(j-1)d}\log 2^{(j-1)d}}(1 - e^{-2^{jd}(1-2^{-d})\log 2^{(j-1)d}}) \\
&\gtrsim e^{-2^{jd}\log 2^{jd}}.
\end{aligned}$$

Chapter 4

Nonparametric Bayesian inference for reversible multi-dimensional diffusions

We study nonparametric Bayesian models for reversible multi-dimensional diffusions with periodic drift. For continuous observation paths, reversibility is exploited to prove a general posterior contraction rate theorem for the drift gradient vector field under approximation-theoretic conditions on the induced prior for the invariant measure. The general theorem is applied to Gaussian priors and p-exponential priors, which are shown to converge to the truth at the minimax optimal rate over Sobolev smoothness classes in any dimension.

4.1 Introduction

Consider observing a continuous trajectory $X^T = (X_t = (X_t^1, \dots, X_t^d) : 0 \le t \le T)$ of the multi-dimensional Markov diffusion process given by the solution to the stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + dW_t, \qquad X_0 = x_0 \in \mathbb{R}^d, \quad t \ge 0, \tag{4.1}$$

where $(W_t = (W_t^1, \dots, W_t^d) : t \ge 0)$ is a standard Brownian motion on $\mathbb{R}^d$ and $b = (b_1, \dots, b_d)$ is a Lipschitz vector field. We are interested in nonparametric Bayesian inference on the drift term b in the practically important case when the diffusion process is *time-reversible.* By a result of Kolmogorov, this is equivalent to the drift b equalling the gradient vector field ∇B of a potential function $B : \mathbb{R}^d \to \mathbb{R}$ (e.g., [15], p. 46).

In many applications, reversibility arises due to physical considerations and is a key property which one wants to incorporate into the model. Indeed, one is often interested in inference on the potential B, which carries important physical information. For example, the SDE (4.1) with $b = \nabla B$ arises as the Smoluchowski-Kramers approximation to the Langevin equation for the motion of a chemically bounded particle [143, 58, 213], in which case B describes the chemical bonding forces. Other applications of the reversible diffusion model (4.1) in physics and chemistry include vacancy diffusion and Lennard-Jones clusters [185, 186] and chemical reaction equations [41].

A Bayesian who wants to model reversible diffusion dynamics must do so explicitly via the prior, namely by constructing one that draws gradient vector fields for b. The natural approach is then to directly place a prior on the potential B rather than on b, which is the approach we pursue here. Moreover, B typically has a strong physical meaning and estimating it is often the primary inferential goal, in which case explicitly modelling the potential provides interpretable inference. Our goal is to provide theoretical guarantees for this Bayesian approach, which arises directly from physical considerations in the modelling step.

The theoretical performance of nonparametric Bayesian procedures for drift estimation is well-studied in the one-dimensional case $d = 1$ [232, 187, 110, 238, 174, 1, 237]. However, much less is known in the general multi-dimensional setting $d \geq 2$, even for nonparametric frequentist methods. In the continuous observation model, Dalalyan and Reiß [68] established pointwise convergence rates and Strauch [217–219] obtained adaptive rates, both using multivariate kernel-type estimators. In the Bayesian setting, Nickl and Ray [173] obtained L^2 and L^∞ posterior contraction rates, as well as Bernstein–von Mises results, for certain *non-reversible* drift vector fields - more discussion can be found below.

In this chapter, we obtain contraction rates for the posterior distribution of B about the true potential B_0 in the model (4.1). We prove a general theorem for diffusions governed by gradient vector fields based on the classical testing approach of Bayesian nonparametrics [99, 235], and firstly apply it to Gaussian priors for B, obtaining minimax optimal rates over Sobolev smoothness classes in any dimension. Gaussian priors are widely used in diffusion models [187, 208, 110, 238, 237, 20, 173] and are a canonical choice, not least for computational reasons [181, 208, 231]. Apart from the non-reversible results in [173], these are the first multi-dimensional Bayesian contraction results for diffusions, and the first for potential-modelling priors.

In applications, the potential B is often spatially inhomogeneous for physical reasons, for instance being spiky in some regions and flat or smooth in others (e.g., molecular communication [98]), see Figure 4.1 for an illustrative example. Gaussian priors are

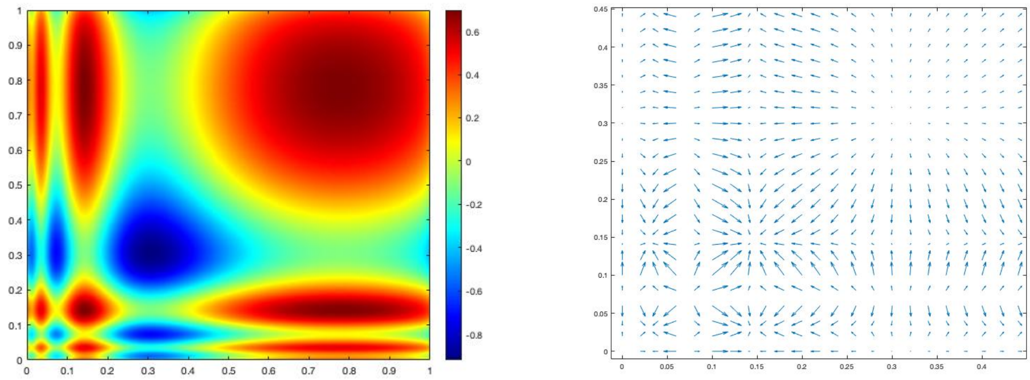


Fig. 4.1 Left: an example of a spatially heterogeneous potential B. Right: the corresponding gradient vector field ∇B. Note the axis scales are different for clarity.

known to be unsuited to modelling such inhomogeneities (see, e.g., [5]), which has motivated the use of heavier tailed priors, especially Besov space priors in the inverse problems [195, 4, 73, 5] and medical imaging [209, 239] communities. To address such physically motivated situations, we also consider modelling the potential using heavier-tailed p-exponential priors [5], for which we establish minimax optimal rates. Such priors have been employed to recover spatially inhomogeneous functions, such as those with a 'blocky' structure with sudden changes from one block to another. They have a number of attractive properties including edge preservation, discretization invariance [147, 146], promoting sparse solution representations, while also maintaining a log-concave structure that aids posterior sampling. Contraction rates for such priors have recently been obtained in direct linear models [5], and we present here a first extension to a nonlinear diffusion setting.

The multi-dimensional case $d \geq 2$ is intrinsically more challenging than the one-dimensional case. The testing approach has been used for model (4.1) first by van der Meulen et al. [232] to obtain contraction rates in the 'natural distance' induced by the statistical experiment, which is a 'random Hellinger semimetric' depending on the observation path $(X_t : 0 \leq t \leq T)$. In dimension $d = 1$, the theory of diffusion local times can then be used to relate this random path-dependent metric to the L^2-distance [232, 187, 238], but when $d > 1$ such local time arguments are not available. In the multi-dimensional setting, Nickl and Ray [173] relate this random Hellinger metric to the L^2-distance for specific truncated Gaussian series product priors on b. They exploit concentration properties of the high-dimensional random matrices induced by the Hellinger semimetric on finite-dimensional projection spaces to relate this problem to a random design type regression problem. However, this approach crucially uses that the priors for each coordinate of $b = (b_1, \ldots, b_d)$ are supported on the same finite-dimensional projection spaces, which is typically only the case if $b_1, \ldots, b_d$ have *independent* priors. Since product priors for b draw gradient vector fields $b = \nabla B$ with probability zero, they are inherently unable to model reversibility.

Aside from requiring different priors, modelling the potential B introduces fundamentally new features to the inference problem at hand. Whereas one can relate the non-gradient vector field case to a direct linear regression type model [173], modelling B is equivalent to modelling the invariant measure (see (4.6)) and leads to a genuinely nonlinear regression problem. We must thus employ a completely different approach to [173] here, using instead tools from nonlinear inverse problems [167, 2, 106, 177, 168] - our work can thus also be viewed as a contribution to the Bayesian inverse problem literature.

Instead of Hellinger testing theory, we develop concentration inequalities for preliminary estimators to directly construct suitable plug-in tests following ideas in [103], see also [195, 174, 1, 2, 160]. In the present setting, the invariant measure μ_b of the diffusion describes the probabilities

$$\mu_b(A) \overset{\text{a.s.}}{=} \lim_{T\to\infty} \frac{1}{T} \int_0^T 1_A(X_t)dt, \tag{4.2}$$

corresponding to the average asymptotic time spent by the process (X_t) in a given measurable subset A of the state space. In the reversible case, we can exploit the one-to-one correspondence $\mu_b \propto e^{2B}$ (see (4.6) below) between the potential B and the invariant measure μ_b, to construct estimators for B based on preliminary estimators for μ_b. We then combine elliptic PDE and martingale techniques with concentration of measure arguments to obtain exponential inequalities for such estimators, and hence bounds for the type-II errors of suitable tests.

As in [181, 187, 238, 1, 97, 237, 173], we restrict to the *periodic* setting and thus periodic potentials B. Under this simplification, a potential B still implies the corresponding (periodised) diffusion is reversible ([97], Proposition 2) and so our results maintain the key modelling link between reversibility and potential functions. As well as ensuring mixing of the diffusion, periodicity simplifies the elliptic PDE arguments involved in studying the mapping properties of the generator of the underlying semigroup, see, e.g., [24]. The underlying PDE techniques in principle extend to the non-periodic setting under certain conditions [182], albeit at the expense of additional technicalities that are beyond the scope of the present chapter.

As well as deriving theoretical results, we also discuss numerical implementation in the continuous observation model considered here. Posterior sampling using simulation techniques is well-studied for 'real-world' discrete data, including for some priors we study here, which we highlight when discussing concrete prior choices. For recent work on this active research topic, see for example [27, 181, 208, 231, 210, 20, 97] and references therein.

4.2 Inference for reversible multi-dimensional diffusions

4.2.1 Basic notation and definitions

Let $\mathbb{T}^d$ be the d-dimensional torus, isomorphic to $[0,1]^d$ with the opposite points on the cube identified. We denote by $L^p(\mathbb{T}^d)$ the usual Lebesgue spaces on $\mathbb{T}^d$ equipped with norm $\|\cdot\|_p$, and by $\langle\cdot,\cdot\rangle_2$ the inner product on $L^2(\mathbb{T}^d)$. We further define the subspaces

$$\dot{L}^2(\mathbb{T}^d) := \left\{ f \in L^2(\mathbb{T}^d) : \int_{\mathbb{T}^d} f dx = 0 \right\}, \qquad L^2_\mu(\mathbb{T}^d) := \left\{ f \in L^2(\mathbb{T}^d) : \int_{\mathbb{T}^d} f d\mu = 0 \right\},$$

where μ is a probability measure on $\mathbb{T}^d$.

Let $C(\mathbb{T}^d)$ be the space of continuous functions on $\mathbb{T}^d$, equipped with the supremum norm $\|\cdot\|_\infty$. For $\beta > 0$, denote by $C^\beta(\mathbb{T}^d)$ the usual Hölder space of $\lfloor\beta\rfloor$-times continuously differentiable functions on $\mathbb{T}^d$ whose $\lfloor\beta\rfloor^{\text{th}}$-derivative is $(\beta - \lfloor\beta\rfloor)$-Hölder continuous. We let $H^\alpha(\mathbb{T}^d)$, $\alpha \in \mathbb{R}$, denote the usual L^2-Sobolev spaces on $\mathbb{T}^d$, defined by duality when $\alpha < 0$. We further define the Sobolev norms $\|f\|_{W^{1,q}} = \|f\|_q + \sum_{i=1}^d \|\partial_{x_i} f\|_q$, and note that $\|\cdot\|_{W^{1,2}}$ is equivalent to $\|\cdot\|_{H^1}$.

Let $\{\Phi_{lr} \;:\; l \in \{-1,0\} \cup \mathbb{N},\; r = 0,\dots,\max(2^{ld}-1,0)\}$ be an orthonormal tensor product wavelet basis of $L^2(\mathbb{T}^d)$, obtained from a periodised Daubechies wavelet basis of $L^2(\mathbb{T})$, which we take to be S-regular for $S \in \mathbb{N}$ large enough; see Section 4.3. in [104] for details. For $J \in \mathbb{N}$, define the finite-dimensional approximation space

$$V_J := \operatorname{span}\left\{ \Phi_{lr} : \; l \le J,\; r = 0,\dots,\max(2^{ld}-1,0) \right\} \tag{4.3}$$

and let $P_J : L^2(\mathbb{T}^d) \to V_J$ be the associated L^2-projection operator. Note that V_J has dimension $v_J := \dim(V_J) = O(2^{Jd})$ as $J \to \infty$. For $0 \le t \le S$, $1 \le p,q \le \infty$, define the

Besov spaces via their wavelet characterisation:

$$B^t_{pq}(\mathbb{T}^d) = \left\{ f \in L^p(\mathbb{T}^d) : \ \|f\|^q_{B^t_{pq}} := \sum_l 2^{ql\left(t+\frac{d}{2}-\frac{d}{p}\right)} \left(\sum_r |\langle f, \Phi_{lr}\rangle_2|^p \right)^{\frac{q}{p}} < \infty \right\},$$

replacing the ℓ^p or ℓ^q-norm above with ℓ^∞ if $p = \infty$ or $q = \infty$, respectively. Recall that $H^t(\mathbb{T}^d) = B^t_{22}(\mathbb{T}^d)$ and the continuous embedding $C^t(\mathbb{T}^d) \subseteq B^t_{\infty\infty}(\mathbb{T}^d)$ for $t \geq 0$, see Chapter 3 in [211].

When no confusion may arise, we suppress the dependence of the function spaces on the underlying domain, writing for example B^t_{pq} instead of $B^t_{pq}(\mathbb{T}^d)$. We also employ the same function space notation for vector fields $f = (f_1, \dots, f_d)$. For instance, $f \in H^\alpha = (H^\alpha)^{\otimes d}$ will mean each $f_i \in H^\alpha$ and the norm on H^α is $\|f\|_{H^\alpha} = \sum_{i=1}^d \|f_i\|_{H^\alpha}$. Similarly, $\|\nabla g\|_p = \sum_{i=1}^d \|\partial_{x_i} g\|_p$.

We write $\lesssim$, $\gtrsim$ and $\simeq$ to denote one- or two-sided inequalities up to multiplicative constants that may either be universal or 'fixed' in the context where the symbols appear. We also write $a_+ = \max(a, 0)$ and $a \vee b = \max(a, b)$ for real numbers a, b. The ε-covering number of a set Θ for a semimetric d, denoted $N(\varepsilon; \Theta, d)$, is the minimal number of d-balls of radius ε needed to cover Θ.

4.2.2 Diffusions with periodic drift and Bayesian inference

Consider the SDE (4.1) with drift $b = \nabla B$, for a twice-continuously differentiable and one-periodic potential $B : \mathbb{R}^d \to \mathbb{R}$, that is $B(x + m) = B(x)$ for all $m \in \mathbb{Z}^d$,

$$dX_t = \nabla B(X_t)dt + dW_t, \qquad X_0 = x_0 \in \mathbb{R}^d, \quad t \geq 0. \tag{4.4}$$

There exists a d-dimensional strong pathwise solution $X = (X_t = (X^1_t, \dots, X^d_t) \ : \ t \geq 0)$ with cylindrically defined law on the path space $C([0, \infty); \mathbb{R}^d)$; see, e.g., Chapters 24 and 39 in [19]. For $T > 0$, let $X^T := (X_t \ : \ 0 \leq t \leq T)$ and denote by P^T_B the law of X^T on $C([0, T], \mathbb{R}^d)$. We omit the dependence on the initial condition $X_0 = x_0$ since it plays no role in our results.

By periodicity, we can consider B as a function on $\mathbb{T}^d$. In model (4.4), the law P^T_B depends on B only through $b = \nabla B$ (see (4.7) below), which is thus only identifiable up to an additive constant. We therefore without loss of generality assume that $\int_{\mathbb{T}^d} B(x)dx = 0$, i.e., $B \in \dot{L}^2(\mathbb{T}^d)$. Our goal is to estimate the drift $b = \nabla B : \mathbb{R}^d \to \mathbb{R}^d$ from an observed trajectory $X_T \sim P^T_B$. We will sometimes write P^T_b when a technical result also applies to possibly non-gradient vector field drifts, but this will be clarified in each instance.

The periodic model effectively restricts the diffusion to the bounded state space $\mathbb{T}^d$. More precisely, while the diffusion defined in (4.4) takes values on all of $\mathbb{R}^d$, its values (X_t) modulo $\mathbb{Z}^d$ contain all the relevant statistical information about ∇B (note that (X_t) will not be globally recurrent on $\mathbb{R}^d$). This allows us to define an *invariant measure* on $\mathbb{T}^d$, since it holds that (arguing as in the proof of Lemma 6 in [173])

$$\frac{1}{T}\int_0^T \varphi(X_s)ds \xrightarrow{P_B} \int_{\mathbb{T}^d} \varphi d\mu_b, \qquad T\to\infty, \quad \forall\varphi\in C(\mathbb{T}^d),$$

where μ_b is a uniquely defined probability measure on $\mathbb{T}^d$ and we identify φ with its periodic extension to $\mathbb{R}^d$ on the left-hand side. The measure μ_b thus inherits the usual probabilistic interpretation as the limiting ergodic average in (4.2).

Recall that the periodic generator $L_b : H^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ of the possibly non-reversible diffusion from (4.1) is

$$L_b = \frac{1}{2}\Delta + b.\nabla = \frac{1}{2}\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} + \sum_{i=1}^d b_i(\cdot)\frac{\partial}{\partial x_i}. \tag{4.5}$$

The corresponding invariant measure μ_b is known to be the weak solution of the PDE $L_b^*\mu_b = 0$ on $\mathbb{T}^d$, where $L_b^* = \frac{1}{2}\Delta - b.\nabla - \nabla\cdot b$ is the L^2-adjoint operator of L_b, see, e.g., p. 45 in [15] (in a slight abuse of notation, we use μ_b for both the probability measure and its density function). If $b = \nabla B$ for some potential $B \in C^2(\mathbb{T}^d)$, one can check directly that

$$\mu_B(x) = \frac{e^{2B(x)}}{\int_{\mathbb{T}^d} e^{2B(x')}dx'}, \qquad x\in\mathbb{T}^d, \tag{4.6}$$

is a classical C^2 solution to $L^*_{\nabla B}\mu_B = 0$ and hence is the unique invariant probability density function of the diffusion (again in a slight abuse of notation, we write μ_B instead of $\mu_{\nabla B}$ when $b = \nabla B$). In particular, if we have a potential B, then we can recover b from μ_b via $b = \nabla B = \frac{1}{2}\nabla \log \mu_b$, a connection we exploit in our proofs.

The log-likelihood for $B \in C^2(\mathbb{T}^d)$ for our observation model is given by Girsanov's theorem (e.g., Section 17.7 in [19]):

$$\ell_T(B) := \log \frac{dP_B^T}{dP_0^T}(X^T) = -\frac{1}{2}\int_0^T \|\nabla B(X_t)\|^2 dt + \int_0^T \nabla B(X_t).dX_t, \tag{4.7}$$

where P_0^T is the law of a d-dimensional Brownian motion $(W_t \ : \ 0 \le t \le T)$. We consider a Bayesian approach to the problem, assigning a (possibly T-dependent) prior $\Pi = \Pi_T$ to B, which for identifiability we assume is supported on $\dot{L}^2(\mathbb{T}^d)\cap C^2(\mathbb{T}^d)$. The posterior

$\Pi(\cdot|X^T)$ then takes the form

$$\Pi(A|X^T) = \frac{\int_A e^{\ell_T(B)} d\Pi(B)}{\int_{C^2(\mathbb{T}^d)} e^{\ell_T(B')} d\Pi(B')}, \qquad \text{any Borel } A \subseteq C^2(\mathbb{T}^d). \tag{4.8}$$

Note that this induces a prior and posterior for both $b = \nabla B$ and μ_B. In the following, we study the concentration of the posterior about the 'ground truth' gradient vector field $b_0 = \nabla B_0$, assuming that the observation $X^T \sim P^T_{B_0}$ is generated according to the SDE (4.4) with $B = B_0$.

4.3 Main results

4.3.1 Gaussian process priors

Posterior contraction rates

Gaussian priors are widely employed for diffusion models [181, 187, 208, 110, 238, 231, 237, 20, 173] and we provide here theoretical guarantees for such priors. We consider a class of Gaussian process priors constructed from a base Gaussian probability measure Π_W, which we assume satisfies the following condition. We refer, e.g., to Chapter 2 in [104] for definitions and terminology regarding the theory of Gaussian processes and measures.

Condition 4.1. *For $\alpha > d/2 + (d/2 - 1)_+$, let Π_W be a centred Gaussian Borel probability measure on the Banach space $C(\mathbb{T}^d)$ that is supported on a separable (measurable) linear subspace of $C^{(d/2+\kappa)\vee 2}(\mathbb{T}^d)$ for some $\kappa > 0$, and assume its reproducing kernel Hilbert space (RKHS) $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ is continuously embedded into the Sobolev space $H^{\alpha+1}(\mathbb{T}^d)$.*

Examples of Gaussian processes priors satisfying Condition 4.1 include the periodic Matérn process and high-dimensional Gaussian series expansions, see Examples 4.1-4.2 below.

To account for the nonlinearity of the problem, we rescale the base Gaussian processes following ideas from the Bayesian inverse problem literature [168, 2, 106, 177, 168]. Given a random draw $W \sim \Pi_W$ satisfying Condition 4.1, consider the following rescaled function:

$$B(x) := \frac{W(x)}{T^{d/(4\alpha+2d)}}, \qquad x \in \mathbb{T}^d, \tag{4.9}$$

whose law Π we take as the prior for the potential B. It follows that Π is a centred Gaussian probability measure on $C(\mathbb{T}^d)$, with the same support and RKHS as Π_W, but with rescaled RKHS norm $\|h\|_{\mathcal{H}_B} = T^{d/(4\alpha+2d)}\|h\|_{\mathcal{H}}$.

The rescaling enforces additional regularisation in the induced posterior distribution for the invariant measure μ_B, implying in particular a bound for $\|\mu_B\|_{C^{(d/2+\kappa)\vee 2}}$, needed to control the nonlinear map $B \mapsto \mu_B$ given in (4.6) under the posterior. Such issues are commonly encountered in nonlinear inverse problems, where one often requires the posterior to place most of its mass on sets of bounded higher-order smoothness in order to use stability estimates. Note that such priors are special cases of the rescaled Gaussian process priors considered in several benchmark statistical settings in [233].

Theorem 4.1. *Let* Π *be the rescaled Gaussian process prior for* B *in* (4.9) *with* $W \sim \Pi_W$ *satisfying Condition 4.1 for some* $\alpha > d/2 + (d/2-1)_+$, $\kappa > 0$, *and RKHS* $\mathcal{H}$. *Suppose that* $B_0 \in H^{\alpha+1}(\mathbb{T}^d)$ *and that there exists a sequence* $B_{0,T} \in \mathcal{H}$ *such that* $\|B_0 - B_{0,T}\|_{C^1} = O(T^{-\alpha/(2\alpha+d)})$ *and* $\|B_{0,T}\|_{\mathcal{H}} = O(1)$ *as* $T \to \infty$. *Then for* $M > 0$ *large enough, as* $T \to \infty$,

$$P_{B_0}^T \Pi(B : \|\nabla B - \nabla B_0\|_2 \geq M T^{-\alpha/(2\alpha+d)} | X^T) \to 0.$$

The resulting posterior thus contracts about the truth at the minimax rate (cfr. [68]) for any dimension d. We recall that the law P_B^T depends on the potential B only through ∇B in (4.4), and thus it is natural to study recovery of the gradient vector field ∇B_0. In fact, after making the identifiability assumption $B, B_0 \in \dot{L}^2(\mathbb{T}^d)$, the norm $\|\nabla B - \nabla B_0\|_2$ is equivalent to the usual Sobolev norm $\|B - B_0\|_{H^1}$.

Theorem 4.1 requires that the true B_0 be approximable by elements $B_{0,T}$ of the RKHS of W at a suitable rate, which reflects the notion of smoothness being modelled by the Gaussian process prior. One can thus think of this condition as 'B_0 is $(\alpha+1)$-smooth' in both the Sobolev and prior sense. For instance, if the Gaussian prior already models Sobolev smooth functions (e.g., a Matérn process prior - see Example 4.1), then this poses no additional conditions.

Examples of Gaussian priors

We now provide concrete examples of Gaussian priors to which Theorem 4.1 applies. As discussed in Section 4.2.2, the potential B is only identified up to an additive constant, which we without loss of generality select via $\int_{\mathbb{T}^d} B(x)dx = 0$. For series expansions, one can enforce this by setting the coefficient of $e_0 \equiv 1$ (Fourier basis) or $\Phi_{-10} \equiv 1$ (wavelet basis) equal to zero. For more general Gaussian processes, one can simply recenter the prior draws by $B \mapsto B - \int_{\mathbb{T}^d} B(x)dx$.

A common choice for this problem is a mean-zero Gaussian process with covariance operator equal to an inverse power of the Laplacian [187, 238, 237], for which posterior inference based on discrete data can be computed efficiently using a finite element method

[181] (note that in the continuous model considered here, Gaussian priors for B are conjugate, see Lemma 4.1 below). Such priors can be defined via a Karhunen-Loève expansion in the Fourier basis and are equivalent to periodic Matérn processes, see Appendix 4.A.1 for details

Example 4.1 (Periodic Matérn process)**.** *For $\alpha+1>d/2+(d/2)\vee 2$, consider the base Gaussian prior*

$$W(x)=(2\pi)^{d/2}\sum_{k\in\mathbb{Z}^d}\frac{1}{(1+4\pi^2\|k\|^2)^{(\alpha+1)/2}}g_k e_k(x),\qquad g_k\overset{\text{iid}}{\sim}N(0,1),\quad x\in\mathbb{T}^d,\quad (4.10)$$

corresponding to the series expansion of a periodic Matérn process with smoothness parameter $\alpha+1-d/2$ (cfr. Appendix 4.A.1 for details). By the Fourier series characterisation of Sobolev spaces, its RKHS $\mathcal{H}$ equals $H^{\alpha+1}(\mathbb{T}^d)$ with equivalent RKHS norm $\|\cdot\|_{\mathcal{H}}\simeq\|\cdot\|_{H^{\alpha+1}}$. Furthermore, W defines a Borel random element in $C^{\alpha+1-\frac{d}{2}-\eta}(\mathbb{T}^d)$ for all $\eta>0$, which is a separable linear subspace of $C^{(d/2+\kappa)\vee 2}(\mathbb{T}^d)$ for some $\kappa,\eta>0$ if $\alpha+1>d/2+(d/2)\vee 2$. Condition 4.1 therefore holds for periodic Matérn processes.

We may thus apply Theorem 4.1 to the periodic Matérn base prior in (4.10) and any $B_0\in H^{\alpha+1}(\mathbb{T}^d)=\mathcal{H}$ with $\alpha+1>d/2+(d/2)\vee 2$ by taking the trivial sequence $B_{0,T}=B_0\in H^{\alpha+1}(\mathbb{T}^d)$.

Another common approach to prior modelling is to obtain a high-dimensional discretisation by a truncated Gaussian series expansion. We illustrate this considering a wavelet expansion for concreteness, but analogous results can be derived for any basis compatible with the Sobolev smoothness scales, such as the Fourier basis.

Example 4.2 (Truncated Gaussian series)**.** *Let $\{\Phi_{lr},\ l\geq -1, r=0,\dots,\max(2^{ld}-1,0)\}$ be a periodised Daubechies wavelet basis of $L^2(\mathbb{T}^d)$ as described in Section 4.2.1, and consider the base prior*

$$W(x)=\sum_{l\leq J}\sum_r 2^{-l(\alpha+1)}g_{lr}\Phi_{lr}(x),\qquad g_{lr}\overset{\text{iid}}{\sim}N(0,1),\quad x\in\mathbb{T}^d,$$

for some $\alpha>d/2+(d/2-1)_+$, where $2^J\simeq T^{1/(2\alpha+d)}$, which tends to infinity as $T\to\infty$, is the optimal dimension of a finite-dimensional model for α-smooth functions. The support of Π_W equals the finite-dimensional approximation space V_J, which is a separable linear subspace of $C^{(d/2+\kappa)\vee 2}(\mathbb{T}^d)$. Its RKHS $\mathcal{H}$ equals V_J with norm

$$\|h\|_{\mathcal{H}}^2=\sum_{l\leq J}\sum_r 2^{2l(\alpha+1)}|\langle h,\Phi_{lr}\rangle_2|^2=\|h\|_{H^{\alpha+1}}^2,\qquad h\in V_J,$$

so that Π_W *satisfies Condition 4.1.*

For $B_0 \in H^{\alpha+1}(\mathbb{T}^d) \cap C^{\alpha+1}(\mathbb{T}^d)$*, the wavelet projections* $B_{0,T} = P_J B_0 \in V_J = \mathcal{H}$ *satisfy* $\|P_J B_0\|_{\mathcal{H}} \leq \|B_0\|_{H^{\alpha+1}} < \infty$ *and* $\|B_0 - P_J B_0\|_{C^1} \lesssim 2^{-J\alpha} \simeq T^{-\alpha/(2\alpha+d)}$*. Theorem 4.1 therefore applies with* Π_W *a Gaussian wavelet series and all* $B_0 \in H^{\alpha+1}(\mathbb{T}^d) \cap C^{\alpha+1}(\mathbb{T}^d)$ *with* $\alpha > d/2 + (d/2-1)_+$*.*

In Section 4.3.2, we extend the last result to truncated p-exponential series priors. For $p = 2$, Theorem 4.2 below shows that the above additional smoothness requirement $B_0 \in C^{\alpha+1}(\mathbb{T}^d)$ can be removed under the slightly stronger minimal smoothness assumption $\alpha + 1 > (d/2) \vee 2 + d/2$.

Conjugacy of Gaussian priors

For the continuous observation model $X^T = (X_t : 0 \leq t \leq T)$, the likelihood (4.7) is of quadratic form in the potential B, and hence Gaussian priors are conjugate as we now show. This parallels the known conjugacy property of Gaussian priors for the drift vector field b [187], which in our setting corresponds to Gaussianity of the posterior on ∇B.

Lemma 4.1. *Let* Π *be a centred Gaussian Borel probability measure on* $L^2(\mathbb{T}^d)$ *that is supported on* $C^2(\mathbb{T}^d) \cap \dot{L}^2(\mathbb{T}^d)$*. Then the posterior distribution (4.8) is almost surely (under the law of the data* X^T*) Gaussian on* $L^2(\mathbb{T}^d)$*.*

The $C^2(\mathbb{T}^d)$ condition is a standard assumption on B to ensure the existence of a strong pathwise solution to the SDE (4.4), see, e.g., [19], and thus is natural in our setting. Conjugacy implies that posterior sampling is straightforward to implement in this model. Consider a discretisation step by a Karhunen-Loève (KL) truncation, taking as prior the random function

$$B(x) = \sum_{k=1}^{K} \upsilon_k g_k h_k(x), \qquad g_k \overset{\text{iid}}{\sim} N(0,1), \quad x \in \mathbb{T}^d, \tag{4.11}$$

for some fixed $K \in \mathbb{N}$, scaling coefficients $\upsilon_k > 0$, and some ‘basis’ functions $(h_k,\ k \in \mathbb{N}) \subset C^2(\mathbb{T}^d) \cap \dot{L}^2(\mathbb{T}^d)$ (e.g., the Fourier or wavelet bases). Identifying a function $B = \sum_{k=1}^K B_k h_k$ with its coefficient vector $\mathbf{B} = (B_1, \dots, B_K)^T \in \mathbb{R}^K$, a standard conjugate computation yields

$$\mathbf{B}|X^T \sim N\left((\Sigma + \Upsilon^{-1})^{-1}\mathbf{H}, (\Sigma + \Upsilon^{-1})^{-1}\right), \tag{4.12}$$

where $\Upsilon = \mathrm{diag}(\upsilon_1^2, \dots, \upsilon_K^2)$ is a $K \times K$ diagonal matrix and

$$\Sigma = \left[\int_0^T \nabla h_k(X_t).\nabla h_{k'}(X_t)dt\right]_{k,k'} \in \mathbb{R}^{K,K}, \quad \mathbf{H} = \left[\left(\int_0^T \nabla h_k(X_t).dX_t\right)_{k=1}^K\right]^T \in \mathbb{R}^K.$$

Additional details and the proof of Lemma 4.1 can be found in Appendix 4.A.1. For concrete basis choices for the KL-expansion (4.11), Σ and $\mathbf{H}$ can be computed from the data, allowing direct posterior sampling according to (4.12). Gaussian conjugacy no longer holds for the more realistic discrete data setting, where more advanced sampling techniques must be employed, see for instance [181, 208, 210, 20].

4.3.2 p-exponential priors

We next consider modelling the potential function B using the class of heavier-tailed p-exponential priors [5], known in the inverse problems literature as Besov priors [147]. These priors are constructed via random basis expansions, assigning i.i.d. random coefficients distributed according to the probability density function

$$f_p(x) \propto e^{-\frac{|x|^p}{p}}, \qquad x \in \mathbb{R}, \quad p \in [1, 2].$$

This generalises the series construction of Gaussian priors ($p = 2$), allowing heavier-tailed random coefficients for $p < 2$, while preserving a log-concave structure favourable to computation, see Remark 4.1. The class includes products of Laplace distributions ($p = 1$), which have recently received significant interest in the Bayesian inverse problem community [142, 70, 195, 50, 73, 4] due to their edge-preserving and sparsity-promoting properties. For Laplace priors, these advantages stem from the ℓ^1-type penalty induced by the prior, which promotes sparse reconstructions that have been observed to perform better in practice for the recovery of spatially-irregular, blocky structures such as images, see, e.g., [193, 151, 142, 49, 124] and references therein. It is therefore of interest to provide theoretical guarantees for these methods which are employed in practice.

We consider p-exponential priors defined via a truncated wavelet expansion. Let $\{\Phi_{lr},\ l \geq -1, r = 0, \dots, \max(2^{ld} - 1, 0)\}$ be a periodised Daubechies wavelet basis of $L^2(\mathbb{T}^d)$ as described in Section 4.2.1. For $p \in [1, 2]$ and $s \geq 0$, consider the *p-exponential measure* [5] Π_W arising as the law of the random function

$$W(x) = \sum_{l=0}^{J} \sum_{r} 2^{-l\left(\alpha+1+\frac{d}{2}-\frac{d}{p}\right)} \xi_{lr} \Phi_{lr}(x), \qquad x \in \mathbb{T}^d, \tag{4.13}$$

where the truncation level satisfies $2^J \simeq T^{1/(2\alpha+d)}$ as $J \to \infty$. Note that this class includes both the product Laplace prior ($p = 1$) and Gaussian series prior ($p = 2$). For identifiability, the wavelet coefficient corresponding to $\Phi_{-10} \equiv 1$ is again set to zero to enforce the zero-integral condition, so that $W \in \dot{L}^2(\mathbb{T}^d)$ almost surely. Similar to the Gaussian priors considered in the previous section, we introduce a suitable scaling of

$W \sim \Pi_W$, taking as prior Π for B the law of

$$B(x) = \frac{W(x)}{\left(T^{\frac{d}{2\alpha+d}}\right)^{\frac{1}{p}}}, \qquad x \in \mathbb{T}^d. \tag{4.14}$$

This is the correct scaling, since scaling at a different rate yields suboptimal contraction rates even in the simpler Gaussian white noise model [5, Proposition 5.8]. The next theorem shows that the resulting posterior contracts about the truth at the minimax optimal rate in any dimension d, extending the Gaussian white noise results of [5] to the present nonlinear diffusion setting.

Theorem 4.2. *Let Π be the scaled p-exponential truncated series prior (4.14), where W is as in (4.13) for some $\alpha > [(d/2+\kappa) \vee 2] + d/p - 1$ and $p \in [1,2]$. Suppose that $B_0 \in H^{\alpha+1}(\mathbb{T}^d)$. Then, for $M > 0$ large enough, as $T \to \infty$,*

$$P^T_{B_0} \Pi\Big(B : \|\nabla B - \nabla B_0\|_p \geq M T^{-\alpha/(2\alpha+d)} \Big| X^T\Big) \to 0.$$

For $p = 2$, Theorem 4.2 implies the result in Example 4.2 for truncated Gaussian wavelet series priors, but removes the additional assumption that $B_0 \in C^{\alpha+1}(\mathbb{T}^d)$ under a slightly stronger minimal smoothness assumption on α.

Remark 4.1 (MAP estimation and sampling)**.** For the p-exponential prior (4.14), the posterior density for any $B = \sum_{l \leq J,r} B_{lr}\Phi_{lr} \in V_J$ takes the form $d\Pi(B|X^T) \propto e^{-\Psi_T(B)}$, where

$$\begin{aligned}
\Psi_T(B) &= -\ell_T(B) + \frac{(T^{\frac{d}{2\alpha+d}})^{\frac{1}{p}}}{p} \|B\|^p_{B^{\alpha+1}_{pp}} \\
&= \frac{1}{2} \sum_{l \leq J,r} \sum_{l' \leq J,r'} B_{lr} B_{l'r'} \left[\int_0^T \nabla\Phi_{lr}(X_t).\nabla\Phi_{l'r'}(X_t)dt\right] \\
&\quad - \sum_{l \leq J,r} B_{lr} \left[\int_0^T \nabla\Phi_{lr}(X_t).dX_t\right] + \frac{(T^{\frac{d}{2\alpha+d}})^{\frac{1}{p}}}{p} \sum_{l \leq J,r} 2^{pl\left(\alpha+1+\frac{d}{2}-\frac{d}{p}\right)} |B_{lr}|^p.
\end{aligned}$$

Since $p \geq 1$, $\Psi_T(B)$ is, given the data X^T, a convex functional of B, implying that the posterior is log-concave. Computation of the Bayesian MAP (maximum-a-posteriori) estimator, defined as the minimiser over V_J of $\Psi_T(\cdot)$, can then be performed using efficient high-dimensional convex optimisation algorithms [46, 109, 57]. Furthermore, (approximate) sampling from the posterior distribution is also feasible by employing MCMC algorithms for log-concave distributions. In particular, non-asymptotic convergence guarantees suitable for high-dimensional settings have been derived for gradient-based

Langevin Monte Carlo methods [69, 81, 82] (also for the relevant non-smooth case $p = 1$, using suitable proximal regularisation of the log-posterior density [183, 83]).

4.3.3 A general contraction theorem for multi-dimensional diffusions with gradient drift vector field

The results for Gaussian and p-exponential priors presented in the preceding sections are based on the following general contraction rate theorem for the drift $b = \nabla B$. We employ the testing approach of [99], which requires the construction of suitable tests with exponentially decaying type-II errors. This has been done in [232] for the 'natural distance' for this model, which is an observation-dependent 'random Hellinger semimetric'. In dimension $d = 1$, this can be related to the L^2 distance using the theory of diffusion local times, something which is unavailable in dimension $d > 1$. We instead directly construct plug-in tests based on the concentration properties of preliminary estimators following ideas from the i.i.d. density estimation model [103]. In the present multi-dimensional diffusion setting, suitable estimators can be obtained by exploiting the one-to-one connection between the potential B and the invariant measure μ_B given by (4.6).

Theorem 4.3. *Let $q \in [1,2]$, $J \in \mathbb{N}$, $\varepsilon_T \to 0$ and $\xi_T \to 0$ satisfy $2^J \to \infty$, $T\varepsilon_T^2 \to \infty$ and $T^{-1/2}2^{Jd/2} + \varepsilon_T = O(\xi_T)$, and let $\Pi = \Pi_T$ be priors for B supported on the Banach space $C^2(\mathbb{T}^d)$. Assume further that*

$$2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}\varepsilon_T = O(1) \qquad \text{and} \qquad T^{-1/2}2^{J[d+\kappa+(d/2+\kappa-1)_+]} = O(1) \tag{4.15}$$

for some $\kappa > 0$. Consider sets

$$\Lambda_T \subseteq \left\{\mu : \int_{\mathbb{T}^d} \mu(x)dx = 1,\ \mu(x) \geq \delta,\ \|\mu\|_{C^{(d/2+\kappa)\vee 2}} \leq m,\ \|\mu - P_J\mu\|_{W^{1,q}} \leq C_\Lambda \xi_T\right\} \tag{4.16}$$

for some $\delta, C_\Lambda, m > 0$ and define $\Theta_T = \{B : \nabla B = \frac{1}{2}\nabla \log \mu \text{ for some } \mu \in \Lambda_T\}$. Let B_0 be the true potential and suppose that $\mu_0 = \frac{e^{2B_0}}{\int_{\mathbb{T}^d} e^{2B_0(x)}dx}$ satisfies $\|\mu_0 - P_J\mu_0\|_{W^{1,q}} \leq D_0\xi_T$. Suppose further that

(i) $\Pi(\Theta_T^c) \leq e^{-(C+4)T\varepsilon_T^2}$,

(ii) there exist deterministic sets $\mathcal{SB}_T$ *for* B *with* $\Pi(\mathcal{SB}_T) \geq e^{-CT\varepsilon_T^2}$ *and*

$$P_{B_0}^T \left(\sup_{B \in \mathcal{SB}_T} \sum_{i=1}^d \int_0^T \|\partial_{x_i} B(X_s) - \partial_{x_i} B_0(X_s)\|^2 ds \leq T\varepsilon_T^2 \right) \to 1. \tag{4.17}$$

Then for $M > 0$ *large enough, as* $T \to \infty$*,*

$$P_{B_0}^T \Pi \left(B : \|\nabla B - \nabla B_0\|_q \geq M\xi_T | X^T \right) \to 0.$$

Remark 4.2 (Small ball probability)**.** One can always take $\mathcal{SB}_T = \{B : \sum_i \|\partial_{x_i} B - \partial_{x_i} B_0\|_\infty^2 \leq \varepsilon_T^2\}$, in which case (4.17) holds automatically. However, for truncated wavelet series priors, this leads to unnecessary smoothness conditions on the true underlying function B_0, which can be avoided by instead taking $\mathcal{SB}_T = \{B : \|\nabla B - \nabla B_0\|_{L^2(\mu_0)}^2 \leq \varepsilon_T^2\}$ and verifying (4.17) directly (cfr. Lemma 4.4). Note that $\mathcal{SB}_T$ is required to be deterministic in order to apply certain martingale arguments in the proof of Theorem 4.3.

Beyond the 'usual' conditions arising from the testing approach [99], the main additional assumption in the last theorem is that the prior puts most of its mass on a set of potentials Θ_T, where the corresponding invariant measures $\mu \in \Lambda_T$ can be well approximated by their wavelet projections $P_J\mu$ (cfr. the last inequality in (4.16)). If this is the case, it suffices to study the deviations of the wavelet projection estimator $\hat{\mu}_T$ about its expectation $P_J\mu$. We then use results from empirical process theory, martingale theory and PDEs in order to obtain suitable concentration inequalities, and hence exponential probability bounds for the type-II errors of suitable tests.

A significant additional difficulty in carrying out this program is the nonlinearity of the map $B \mapsto \mu_B$ given by (4.6). One can relate the error of $\nabla B - \nabla B_0$ to that of $\mu_B - \mu_{B_0}$ by a type of stability estimate, see the proof of Lemma 4.3 below. However, without controlling the norm $\|\mu\|_{C^{(d/2+\kappa)\vee 2}}$, the constant for this stability estimate grows rapidly, rendering it unusable in the proofs. This reinforces the connection of recovery of the gradient vector field ∇B_0 in the diffusion model (4.4) with nonlinear inverse problems, where similar phenomena are often encountered, and indeed motivated the use of rescaled priors to overcome these considerable technical challenges [168, 2, 106, 177, 168].

4.3.4 On models with non-constant diffusivity

We comment on some implications for likelihood-based estimation in generalizations of the models (4.1) and (4.4). Consider observing a continuous trajectory $X^T = (X_t : 0 \leq t \leq T)$

from a diffusion dynamics of the form

$$dX_t = b(X_t)dt + \Sigma^{1/2}(X_t)dW_t, \qquad X_0 = x_0 \in \mathbb{R}^d, \quad t \geq 0,$$

with non-constant local variance (diffusion or volatility matrix) $\Sigma : \mathbb{R}^d \to \mathbb{R}^{d,d}$. For general positive-definite $\Sigma(\cdot)$, this model is not identifiable and so we restrict to the case of scalar local variance $\Sigma(X_t) = \sigma(X_t)I_d$, where $\sigma : \mathbb{R}^d \to [0,\infty)$ and I_d is the $d \times d$ identity matrix.

In this model, one can exactly recover the quadratic variation process of any component of X, namely $([X^i]_t = \int_0^t \sigma(X_s)^2 ds : 0 \leq t \leq T)$, $i = 1,\dots,d$, and hence $\{\sigma(X_t) : 0 \leq t \leq T\}$ is perfectly identified from the data X^T. In the well-studied scalar case $d = 1$, this corresponds to knowledge of $\{\sigma(x) : \inf_{t\in[0,T]} X_t \leq x \leq \sup_{t\in[0,T]} X_t\}$ and yields the conventional wisdom in the statistical diffusion literature with continuous data that one can treat $\sigma(\cdot)$ as known, usually taking $\sigma(x) \equiv 1$ for simplicity as we do here. For dimension $d \geq 2$, while we still perfectly identify σ *along the trajectory* of the diffusion, this trajectory now has zero Lebesgue measure in $\mathbb{R}^d$ (note that for $d \geq 2$, the diffusion process will not be recurrent). Thus the diffusivity function $\sigma(\cdot)$ is a non-trivial parameter and there may still be statistical interest in modelling it.

Consider placing a prior on $\sigma(\cdot)$ and let $P_{b,\sigma} = P_{b,\sigma}^T$ be the law of the above process. Girsanov's theorem (e.g., Section 17.7 in [19]) implies that the two measures $P_{b,\sigma} = P_{b',\sigma'}$ are singular unless $\sigma = \sigma'$. The likelihood $e^{\ell_T(b,\sigma)}$ is thus zero unless σ *exactly* matches the observed diffusivity along the trajectory. Therefore, for a Bayesian posterior to be well-defined, the prior must assign positive probability to $\{\sigma(\cdot) : \sigma(X_t) = (\frac{d}{dt}[X^i]_t)^{1/2},\ 0 \leq t \leq T\}$, i.e., it must be conditioned to match these values along the trajectory. Since the random trajectory X^T typically has fractal-like behaviour, this is a highly non-standard and non-trivial prior construction. It is a somewhat unusual feature of this model that any prior for $\sigma(\cdot)$ must heavily rely on the observed data X^T. Similar considerations apply to other likelihood-based procedures, such as maximum likelihood estimation.

The above features stem from the continuous observation model and do not occur in the more realistic low frequency discrete observation model, where estimation of (b,σ) is an ill-posed inverse problem, see [108] for the scalar case $d = 1$. While both cases are interesting mathematically, the continuous and discrete models are fundamentally different problems with regards to estimating the diffusivity $\sigma(\cdot)$. We are unaware of any results concerning minimax rates in dimension $d \geq 2$ in the low frequency setting.

4.4 Proof of Theorem 4.3

We employ the general testing approach for non-i.i.d. sampling models [100] combined with tools from the diffusion setting [232]. In order to construct suitable plug-in tests, we extend ideas from the i.i.d. density model [103] to the multi-dimensional diffusion setting with drift arising as a gradient vector field $b = \nabla B$.

We start with the following contraction rate theorem, which applies also to non-reversible diffusions, based on the existence of abstract tests. In a slight abuse of notation, denote by P_b^T the law of $(X_t : 0 \leq t \leq T)$ from model (4.1), i.e., we do not assume $b = \nabla B$ in the next result.

Theorem 4.4. *Let d_T be a semimetric on the parameter space $\mathcal{D} \subseteq C^1(\mathbb{T}^d)$ for the drift b and let $\Pi = \Pi_T$ be priors for b. Let $\varepsilon_T \to 0$ satisfy $\sqrt{T}\varepsilon_T \to \infty$, let $\xi_T \to 0$, $\mathcal{D}_T \subseteq \mathcal{D}$ and let ϕ_T be a sequence of test functions satisfying*

$$P_{b_0}^T \phi_T \to 0, \qquad \sup_{b \in \mathcal{D}_T \,:\, d_T(b,b_0) \geq D\xi_T} P_b^T(1-\phi_T) \leq L e^{-(C+4)T\varepsilon_T^2}$$

for some $C, D, L > 0$ with $\Pi(\mathcal{D}_T^c) \leq e^{-(C+4)T\varepsilon_T^2}$. Suppose further that there exist deterministic sets $\mathcal{SB}_T \subseteq \mathcal{D}$ with $\Pi(\mathcal{SB}_T) \geq e^{-CT\varepsilon_T^2}$ and

$$P_{b_0}^T \left(\sup_{b \in \mathcal{SB}_T} \int_0^T \|b(X_s) - b_0(X_s)\|^2 ds \leq T\varepsilon_T^2 \right) \to 1,$$

where b_0 is the true drift function. Then for $M > 0$ large enough, as $T \to \infty$,

$$P_{b_0}^T \Pi \left(b : d_T(b, b_0) \geq M\xi_T \middle| X^T \right) \to 0.$$

The proof of Theorem 4.4 follows similarly to results in [99, 232] and is deferred to Appendix 4.A.2. The required tests are contained in the next lemma, whose proof can be found in Section 4.4.2 below.

Lemma 4.2. *Let $q \in [1,2]$, $J \in \mathbb{N}$, $\varepsilon_T \to 0$ and $\xi_T \to 0$ satisfy $2^J \to \infty$, $T\varepsilon_T^2 \to \infty$ and $T^{-1/2}2^{Jd/2} + \varepsilon_T = O(\xi_T)$. Assume further that*

$$2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}\varepsilon_T = O(1) \qquad \text{and} \qquad T^{-1/2}2^{J[d+\kappa+(d/2+\kappa-1)_+]} = O(1)$$

for some $\kappa > 0$. Consider sets

$$\Lambda_T \subseteq \left\{ \mu : \int_{\mathbb{T}^d} \mu(x)dx = 1,\ \mu(x) \geq \delta,\ \|\mu\|_{C^{(d/2+\kappa)\vee 2}} \leq m,\ \|\mu - P_J\mu\|_{W^{1,q}} \leq C_\Lambda \xi_T \right\}$$

for some $\delta, C_\Lambda, m > 0$ *and define* $\Theta_T = \{B : \nabla B = \frac{1}{2}\nabla \log \mu$ *for some* $\mu \in \Lambda_T\}$*. Let* B_0 *and* μ_0 *be the true potential and invariant measure, respectively, and assume that* $\|\mu_0 - P_J\mu_0\|_{W^{1,q}} \leq D_0\xi_T$ *for some* $D_0 > 0$*. Then for any* $M > 0$*, there exist tests* ϕ_T *such that for* $D = D(q, \delta, m, C_\Lambda, D_0, M) > 0$ *large enough,*

$$P^T_{B_0}\phi_T \to 0, \qquad \sup_{B \in \Theta_T \,:\, \|\nabla B - \nabla B_0\|_q \geq D\xi_T} P^T_B(1-\phi_T) \leq 4e^{-MT\varepsilon_T^2}.$$

Proof of Theorem 4.3. The conclusion of Theorem 4.3 follows by applying Theorem 4.4 with the set $\mathcal{D}_T = \{b = \nabla B : B \in \Theta_T\}$, the distance $d_T(f,g) = \|f-g\|_q$ and the tests ϕ_T constructed in Lemma 4.2.

□

4.4.1 A concentration of measure result for empirical processes

The following concentration inequality is a key technical tool in the proof of Theorem 4.3, providing uniform stochastic control of functionals of the (possibly non-reversible) diffusion process (4.1). It is based on a chaining argument for stochastic processes with mixed tails (cfr. Theorem 2.2.28 in Talagrand [222] and Theorem 3.5 in Dirksen [76]). We again write P^T_b for the law of X^T to emphasise that we do not assume $b = \nabla B$ in the next result. Recall the notation $\dot{L}^2 = \{f \in L^2 : \int_{\mathbb{T}^d} f dx = 0\}$ and $L^2_\mu = \{f \in L^2 : \int_{\mathbb{T}^d} f d\mu = 0\}$.

Proposition 4.1. *Suppose* $b \in C^{(d/2+\kappa)\vee 1}(\mathbb{T}^d)$ *for some* $\kappa > 0$*, and let* $\mathcal{F}_T \subset V_J \cap L_{\mu_b}(\mathbb{T}^d)$ *for* J *satisfying* $2^{J[d/2+\kappa+(d/2+\kappa-1)_+]} \lesssim \sqrt{T}$*. Define the empirical process*

$$\mathbb{G}_T(f) := \frac{1}{\sqrt{T}}\int_0^T f(X_s)ds, \qquad f \in \mathcal{F}_T,$$

and let $D_{\mathcal{F}_T} := \dim(\mathcal{F}_T)$ *and* $|\mathcal{F}_T|_{H^{-1}} := \sup_{f\in\mathcal{F}_T}\|f\|_{H^{-1}}$*. Then for all* $T \geq \eta > 0$ *and* $x \geq 0$*,*

$$P^T_b\left(\sup_{f\in\mathcal{F}_T}|\mathbb{G}_T(f)| \geq C|\mathcal{F}_T|_{H^{-1}}\left\{D^{1/2}_{\mathcal{F}_T} + \sqrt{x} + T^{-\frac{1}{2}}2^{J[\frac{d}{2}+\kappa+(\frac{d}{2}+\kappa-1)_+]}(D_{\mathcal{F}_T}+x)\right\}\right) \leq 2e^{-x},$$

where C *depends on* d, κ, η *and upper bounds for* $\|b\|_{B^{|d/2+\kappa-1|\vee 1}_{\infty\infty}}$ *and* $\|\mu_b\|_\infty$*.*

Proof. We first note that since $b \in C^1$, a corresponding unique invariant probability measure $\mu = \mu_b$ indeed exists by Proposition 1 of [173]. For $f \in L^2_\mu \cap V_J \subset L^2_\mu \cap H^{d/2+\kappa}$ and $b \in C^{d/2+\kappa}$, by Lemma 4.8 and the Sobolev embedding theorem, the Poisson equation $L_b u = f$ has a unique solution $L_b^{-1}[f] \in \dot{L}^2 \cap H^{d/2+\kappa+2} \subset C^2$ satisfying $L_b L_b^{-1}[f] = f$.

For such f, we may thus define

$$\begin{aligned} Z_T(f) &:= \int_0^T \nabla L_b^{-1}[f](X_s).dW_s \\ &= L_b^{-1}[f](X_T) - L_b^{-1}[f](X_0) - \int_0^T L_b L_b^{-1}[f](X_s)ds \\ &= L_b^{-1}[f](X_T) - L_b^{-1}[f](X_0) - \sqrt{T}\mathbb{G}_T[f], \end{aligned}$$

where we have used Itô's lemma (Theorem 39.3 of [19]). Thus for $\mathcal{F}_T \subset L^2_\mu \cap H^{d/2+\kappa}$,

$$\sup_{f\in\mathcal{F}_T} |\mathbb{G}_T[f]| \leq \frac{1}{\sqrt{T}} \sup_{f\in\mathcal{F}_T} |Z_T(f)| + \frac{2}{\sqrt{T}} \sup_{f\in\mathcal{F}_T} \|L_b^{-1}[f]\|_\infty. \tag{4.18}$$

We derive a concentration inequality for $\sup_f |Z_T(f)|$ and hence for $\sup_f |\mathbb{G}_T[f]|$.

Recall Bernstein's inequality for continuous local martingales (p. 153 of [198]): if M is a continuous local martingale vanishing at 0 with quadratic variation $[M]$, then for any stopping time T and any $y, K > 0$,

$$\Pr\left(\sup_{0\leq t\leq T} |M_t| \geq y, [M]_T \leq K \right) \leq 2e^{-\frac{y^2}{2K}}. \tag{4.19}$$

For fixed f, $(Z_T(f) : T \geq 0)$ is a continuous square integrable local martingale with quadratic variation $[Z_\cdot(f)]_T = \int_0^T \|\nabla L_b^{-1}[f](X_s)\|^2 ds$. Applying Bernstein's inequality,

$$\begin{aligned} P_b^T\left(|Z_T(f)| \geq x\right) &\leq P_b^T\left(|Z_T(f)| \geq x, [Z_\cdot(f)]_T \leq K_T(f)\right) + P_b^T\left([Z_\cdot(f)]_T > K_T(f)\right) \\ &\leq 2\exp\left(-\tfrac{x^2}{2K_T(f)}\right) + P_b^T([Z_\cdot(f)]_T > K_T(f)) \end{aligned} \tag{4.20}$$

for any $x > 0$ and $K_T(f) > 0$. We now upper bound the right-hand side.

Since $x \mapsto \|x\|^2$ is a smooth map, the function

$$\gamma_f(x) = \|\nabla L_b^{-1}[f](x)\|^2 - \int_{\mathbb{T}^d} \|\nabla L_b^{-1}[f](y)\|^2 d\mu(y)$$

is in $L^2_\mu \cap H^{d/2+\kappa}$ for all $f \in \mathcal{F}_T$. Recall the distance $d_L^2(f,g) := \sum_{i=1}^d \|\partial_{x_i} L_b^{-1}[f-g]\|^2_\infty$ defined in Lemma 1 of [173]. Using the Sobolev embedding theorem, Lemma 4.8 and the

Runst-Sickel lemma ([173], Lemma 2),

$$\begin{aligned} d_L(\gamma_f,0) &\lesssim \|\gamma_f\|_{H^{d/2+\kappa-1}} \\ &\lesssim \sum_{i=1}^d \|(\partial_{x_i}L_b^{-1}[f])^2 - \|\partial_{x_i}L_b^{-1}[f]\|_{L^2(\mu)}^2\|_{H^{d/2+\kappa-1}} \\ &\lesssim \sum_{i=1}^d \|\partial_{x_i}L_b^{-1}[f]\|_\infty \|\partial_{x_i}L_b^{-1}[f]\|_{H^{(d/2+\kappa-1)_+}} + \|\nabla L_b^{-1}[f]\|_{L^2(\mu)}^2 \|1\|_{H^{d/2+\kappa-1}} \\ &\lesssim \|f\|_{H^{d/2+\kappa-1}}\|f\|_{H^{(d/2+\kappa-1)_+-1}} + \|\mu\|_\infty \|f\|_{H^{-1}}^2 \\ &\lesssim \|f\|_{H^{d/2+\kappa-1}}\|f\|_{H^{(d/2+\kappa-1)_+-1}}, \end{aligned}$$

where the constants depend only on d,κ and upper bounds for $\|b\|_{B^{|d/2+\kappa-1|\vee 1}_{\infty\infty}}$ and $\|\mu\|_\infty$. Applying the exponential inequality in Lemma 1 of [173] to the class $\{\gamma_f, 0\}$ gives for $y \geq 0$,

$$P_b^T\left(|\mathbb{G}_T[\gamma_f]| \geq 2T^{-1/2}\|L_b^{-1}[\gamma_f]\|_\infty + Cd_L(\gamma_f,0)(1+y)\right) \leq e^{-y^2/2}.$$

One can identically prove that $\|L_b^{-1}[\gamma_f]\|_\infty \lesssim \|f\|_{H^{d/2+\kappa-1}}\|f\|_{H^{(d/2+\kappa-1)_+-1}}$ with the constant depending on the same quantities as above. Since $\sqrt{T}\mathbb{G}_T[\gamma_f] = [Z.(f)]_T - T\|\nabla L_b^{-1}[f]\|_{L^2(\mu)}^2$, this and the last two displays yield

$$P_b^T\left(\left|[Z.(f)]_T - T\|\nabla L_b^{-1}[f]\|_{L^2(\mu)}^2\right| \geq C\sqrt{T}\|f\|_{H^{d/2+\kappa-1}}\|f\|_{H^{(d/2+\kappa-1)_+-1}}(1+y)\right) \leq e^{-y^2/2}$$

for all $T \geq \eta > 0$. Since by Lemma 4.8, $\|\nabla L_b^{-1}[f]\|_{L^2(\mu)}^2 \lesssim \|\mu\|_\infty \|f\|_{H^{-1}}^2$, substituting the last display into (4.20) with $K_T(f) = CT\|f\|_{H^{-1}}^2 + C\sqrt{T}\|f\|_{H^{d/2+\kappa-1}}\|f\|_{H^{(d/2+\kappa-1)_+-1}}(1+y)$ gives for all $x,y \geq 0$

$$\begin{aligned} &P_b^T\left(|Z_T(f)| \geq x\right) \\ &\quad \leq 2\exp\left(-\frac{x^2}{2CT\|f\|_{H^{-1}}^2 + 2C\sqrt{T}\|f\|_{H^{d/2+\kappa-1}}\|f\|_{H^{(d/2+\kappa-1)_+-1}}(1+y)}\right) + e^{-y^2/2}. \end{aligned}$$

For $f \in V_J$ and $u \geq -1$, $\|f\|_{H^u}^2 = \sum_{l\leq J}\sum_r 2^{2ul}|\langle f, \Phi_{lr}\rangle_2|^2 \leq 2^{2J(u+1)}\|f\|_{H^{-1}}^2$. Using this bound, that $2^{J[d/2+\kappa+(d/2+\kappa-1)_+]} \lesssim \sqrt{T}$ and setting $y = x/(CT\|f\|_{H^{-1}}^2)^{1/2}$, the last display gives

$$P_b^T\left(|Z_T(f)| \geq x\right) \leq 3\exp\left(-\frac{x^2}{2CT\|f\|_{H^{-1}}^2 + 2C2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}\|f\|_{H^{-1}}x}\right).$$

Using the linearity of $f \mapsto Z_T(f)$ and rearranging, we get the following Bernstein inequality:

$$P_b^T\left(|Z_T(f-g)| \geq C\|f-g\|_{H^{-1}}(\sqrt{Tz} + 2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}z)\right) \leq 3e^{-z},$$

where again C depends on d, κ, η and upper bounds for $\|b\|_{B_{\infty\infty}^{|d/2+\kappa-1|\vee 1}}$ and $\|\mu\|_\infty$.

We now apply Theorem 3.5 in Dirksen [76], which is a refinement of Theorem 2.2.28 in Talagrand [222], to bound the supremum of the random process $(Z_T(f))_{f\in\mathcal{F}_T}$. In particular, the last display shows that the process has 'mixed tails' (cfr. (3.8) in [76]) with respect to the metrics $d_1(f,g) = C2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}\|f-g\|_{H^{-1}}$ and $d_2(f,g) = C\sqrt{T}\|f-g\|_{H^{-1}}$. The diameters $\Delta_{d_1}(\mathcal{F}_T)$ and $\Delta_{d_2}(\mathcal{F}_T)$ appearing in the second display of Theorem 3.5 in [76] can be bounded by

$$\Delta_{d_1}(\mathcal{F}_T) := \sup_{f,g\in\mathcal{F}_T} d_1(f,g) \leq 2C2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}|\mathcal{F}_T|_{H^{-1}},$$

$$\Delta_{d_2}(\mathcal{F}_T) := \sup_{f,g\in\mathcal{F}_T} d_2(f,g) \leq 2C\sqrt{T}|\mathcal{F}_T|_{H^{-1}},$$

so that Theorem 3.5 of [76] yields that for all $x \geq 0$,

$$P_b^T\left(\sup_{f\in\mathcal{F}_T}|Z_T(f)| \geq C\left(\gamma_2(\mathcal{F}_T,d_2) + \gamma_1(\mathcal{F}_T,d_1) + |\mathcal{F}_T|_{H^{-1}}\left\{\sqrt{Tx} + 2^{J[\frac{d}{2}+\kappa+(\frac{d}{2}+\kappa-1)_+]}x\right\}\right)\right)$$
$$\leq e^{-x},$$

where γ_1, γ_2 are 'generic chaining functionals'. Using the estimate (2.3) in [76], and recalling that $\mathcal{F}_T \subset V_J$ has dimension $D_{\mathcal{F}_T}$, we can bound $\gamma_2(\mathcal{F}_T, d_2)$ by a multiple of

$$\begin{aligned}\int_0^\infty \sqrt{\log N(\eta;\mathcal{F}_T,d_2)}d\eta &\leq \int_0^{\Delta_{d_2}(\mathcal{F}_T)}\sqrt{D_{\mathcal{F}_T}\log\left(\frac{3\Delta_{d_2}(\mathcal{F}_T)}{\eta}\right)}d\eta \\ &= D_{\mathcal{F}_T}^{1/2}\Delta_{d_2}(\mathcal{F}_T)\int_0^1\sqrt{\log(3/u)}du \\ &\leq C_1 D_{\mathcal{F}_T}^{1/2}\sqrt{T}|\mathcal{F}_T|_{H^{-1}}\end{aligned}$$

for $C_1 > 0$, where the first inequality follows from the usual metric entropy estimate for balls in finite-dimensional Euclidean spaces [104, Proposition 4.3.34]. Similarly, $\gamma_1(\mathcal{F}_T, d_1)$

is bounded by a multiple of

$$\int_0^\infty \log N(\eta; \mathcal{F}_T, d_1) d\eta \le D_{\mathcal{F}_T} \int_0^{\Delta_{d_1}(\mathcal{F}_T)} \log\left(\frac{3\Delta_{d_1}(\mathcal{F}_T)}{\eta}\right) d\eta$$
$$\le C_2 D_{\mathcal{F}_T} \Delta_{d_1}(\mathcal{F}_T) \le C_3 D_{\mathcal{F}_T} 2^{J[d/2+\kappa+(d/2+\kappa-1)_+]} |\mathcal{F}_T|_{H^{-1}}.$$

In summary, Theorem 3.5 of [76] implies that for all $x \ge 0$

$$P_b^T\left(\sup_{f\in\mathcal{F}_T} |Z_T(f)| \ge C|\mathcal{F}_T|_{H^{-1}}\left\{\sqrt{T}(D_{\mathcal{F}_T}^{1/2} + \sqrt{x}) + 2^{J[\frac{d}{2}+\kappa+(\frac{d}{2}+\kappa-1)_+]} D_{\mathcal{F}_T} + x)\right\}\right) \le e^{-x}.$$

This provides the concentration inequality for the first term in (4.18). For the second term in (4.18), using the Sobolev embedding theorem and Lemma 4.8,

$$\|L_b^{-1}[f]\|_\infty \lesssim \|L_b^{-1}[f]\|_{H^{d/2+\kappa}} \lesssim \|f\|_{H^{d/2+\kappa-1}}$$
$$\le 2^{J[\frac{d}{2}+\kappa]}\|f\|_{H^{-1}} \lesssim 2^{J[\frac{d}{2}+\kappa+(\frac{d}{2}+\kappa-1)_+]}|\mathcal{F}_T|_{H^{-1}}$$

for all $f \in \mathcal{F}_T$, where the constant depends on d, κ, η and an upper bound for $\|b\|_{B_{\infty\infty}^{|d/2+\kappa-1|\vee 1}}$. The result then follows from the last two displays and (4.18).

□

4.4.2 Construction of tests and proof of Lemma 4.2

Using Proposition 4.1, we now derive concentration inequalities for a preliminary estimator $\hat{\mu}_T$ of the invariant measure μ, uniformly over certain sets $\Gamma(\delta, m)$ below. We can then exploit the correspondence $\mu \propto e^{2B}$ to obtain an estimator for the gradient vector field $\nabla B = \frac{1}{2}\nabla \log \mu = \frac{1}{2}(\nabla\mu)/\mu$ based on $\hat{\mu}_T$. Consider the wavelet estimator for the invariant measure

$$\hat{\mu}_T(x) = \sum_{l\le J}\sum_r \hat{\beta}_{lr}\Phi_{lr}(x), \qquad x \in \mathbb{T}^d, \tag{4.21}$$

where $\hat{\beta}_{lr} = \frac{1}{T}\int_0^T \Phi_{lr}(X_t)dt$ and $J \to \infty$ as $T \to \infty$. We now prove the required concentration inequality for the induced estimator $\frac{1}{2}\nabla \log \hat{\mu}_T$ of ∇B under the conditions of Theorem 4.3. Recall that P_J denotes the L^2-projection onto the wavelet approximation space V_J given in (4.3), and that we write $f \in L^2_\mu$ if $\int_{\mathbb{T}^d} f d\mu = 0$.

Lemma 4.3. *Let $q \in [1, 2]$, $2^J \to \infty$ and $\varepsilon_T \to 0$ satisfy $T\varepsilon_T^2 \to \infty$. Assume further that*

$$2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}\varepsilon_T = O(1) \qquad \text{and} \qquad T^{-1/2}2^{J[d+\kappa+(d/2+\kappa-1)_+]} = O(1)$$

for some $\kappa > 0$. Define

$$\Gamma = \Gamma(\delta, m) = \left\{\mu : \int_{\mathbb{T}^d} \mu(x)dx = 1,\ \mu(x) \geq \delta \text{ for all } x \in \mathbb{T}^d,\ \|\mu\|_{C^{(d/2+\kappa)\vee 2}} \leq m\right\}$$

for some $\delta, m > 0$. Then there exists $C = C(\delta, m, q) > 0$ such that for all $T > 0$ large enough and all $M > 0$,

$$\sup_{\substack{B:\nabla B = \frac{1}{2}\nabla\log\mu,\\ \mu\in\Gamma}} P_B^T\left(\|\nabla\log\hat{\mu}_T - \nabla\log\mu\|_q \geq C\|\mu - P_J\mu\|_{W^{1,q}} + C(1+M)\left(T^{-\frac{1}{2}}2^{J\frac{d}{2}} + \varepsilon_T\right)\right)$$
$$\leq 4e^{-MT\varepsilon_T^2}.$$

Proof. All constants in this proof are taken uniform over $\delta, m > 0$ in the definition of the set Γ above and we write $b_\mu = \frac{1}{2}\nabla\log\mu$. We first derive a preliminary L^∞-convergence rate ζ_T for $\hat{\mu}_T$, for which we establish an exponential inequality. Write

$$\|\hat{\mu}_T - P_J\mu\|_\infty = \sup_{x\in\mathbb{T}^d}\left|\frac{1}{T}\int_0^T \sum_{l\leq J}\sum_r [\Phi_{lr}(X_s) - \langle\mu, \Phi_{lr}\rangle_2]\Phi_{lr}(x)ds\right| = \frac{1}{\sqrt{T}}\sup_{x\in\mathbb{T}^d}|\mathbb{G}_T[h_x]|$$

with $h_x(u) = \sum_{l\leq J}\sum_r[\Phi_{lr}(u) - \langle\mu, \Phi_{lr}\rangle_2]\Phi_{lr}(x) \in V_J \cap L^2_\mu$ and $\mathbb{G}_T[f] = \frac{1}{\sqrt{T}}\int_0^T f(X_s)ds$ the empirical process for μ-centered functions $f \in L^2_\mu$. Recalling that the periodised father wavelet $\Phi_{-10} \equiv 1$ ([104], p. 354) and that $\sup_x \sum_r \Phi_{lr}(x)^2 \lesssim 2^{Jd}$,

$$\|h_x\|_{H^{-1}}^2 = \sum_{l=0}^{J}\sum_r 2^{-2l}|\Phi_{lr}(x)|^2 + \left|\sum_{l\leq J}\sum_r\langle\mu, \Phi_{lr}\rangle_2\Phi_{lr}(x)\right|^2$$
$$\lesssim \sum_{l=0}^{J} 2^{J(d-2)} + \|P_J\mu\|_\infty^2 \lesssim J_d 2^{J(d-2)_+},$$

where $J_d = J1_{\{d=2\}}$ and we have used $\|P_J\mu\|_\infty \leq \|\mu\|_\infty + \|\mu - P_J\mu\|_\infty \lesssim (1+2^{-J})\|\mu\|_{B^1_{\infty\infty}} \lesssim m$. Applying Proposition 4.1 with $D_{\mathcal{F}_T} = \dim(\mathbb{T}^d) = d$, $|\mathcal{F}_T|_{H^{-1}} = \sup_{x\in\mathbb{T}^d}\|h_x\|_{H^{-1}} = O(J_d^{1/2}\,2^{J(d/2-1)_+})$ and $x = MT\varepsilon_T^2 \to \infty$ then gives

$$P_b^T\left(\|\hat{\mu}_T - P_J\mu\|_\infty \geq C(1+M)J_d^{1/2}2^{J(d/2-1)_+}\varepsilon_T\right) \leq 2e^{-MT\varepsilon_T^2}$$

for $T > 0$ large enough, having used that $2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}\varepsilon_T \lesssim 1$. Note that the constant in the last display depends on d, κ and upper bounds for $\|\mu\|_\infty \lesssim \|\mu\|_{B^1_{\infty\infty}} \lesssim m$ and $\|b_\mu\|_{B^{|d/2+\kappa-1|\vee 1}_{\infty\infty}}$. For this last quantity, by the chain rule, for $\mu \geq \delta$ bounded away from zero, $\|b_\mu\|_{B^\alpha_{\infty\infty}} \lesssim \|\nabla\log\mu\|_{B^\alpha_{\infty\infty}} \lesssim \|\log\mu\|_{B^{\alpha+1}_{\infty\infty}} \lesssim \|\log\mu\|_{C^{\alpha+1}} \lesssim 1 + \|\mu\|_{C^{\alpha+1}}$ for all

$\alpha > 0$. Thus in particular, $\|b_\mu\|_{B^{|d/2+\kappa-1|\vee 1}_{\infty\infty}} \lesssim 1 + \|\mu\|_{C^{(d/2+\kappa)\vee 2}} \lesssim 1 + m$ is also uniformly bounded over Γ. Combined with the bias bound $\|\mu - P_J\mu\|_\infty \lesssim 2^{-J}\|\mu\|_{B^1_{\infty\infty}} \lesssim 2^{-J}m$, this yields

$$P_b^T\left(\|\hat\mu_T - \mu\|_\infty \geq C(1+M)J_d^{1/2}2^{J(d/2-1)_+}\varepsilon_T + C2^{-J}m\right) \leq 2e^{-MT\varepsilon_T^2} \tag{4.22}$$

with C a uniform constant over Γ. Set $\zeta_T = C(1+M)J_d^{1/2}2^{J(d/2-1)_+}\varepsilon_T + C2^{-J}m \lesssim 2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}\varepsilon_T + 2^{-J}m \to 0$. On the event $\{\|\hat\mu_T - \mu\|_\infty \leq \zeta_T\}$, for $x \in \mathbb{T}^d$, $\mu \in \Gamma$ and $i = 1, \dots, d$,

$$\begin{aligned}
|\partial_{x_i}\log\hat\mu_T(x) - \partial_{x_i}\log\mu(x)| &= \left|\frac{\partial_{x_i}\hat\mu_T(x)}{\hat\mu_T(x)} - \frac{\partial_{x_i}\mu(x)}{\mu(x)}\right| \\
&\leq \left|\frac{\partial_{x_i}\hat\mu_T(x) - \partial_{x_i}\mu(x)}{\hat\mu_T(x)}\right| + |\partial_{x_i}\mu(x)|\left|\frac{\mu(x) - \hat\mu_T(x)}{\hat\mu_T(x)\mu(x)}\right| \\
&\leq \frac{|\partial_{x_i}\hat\mu_T(x) - \partial_{x_i}\mu(x)|}{\delta - \zeta_T} + |\partial_{x_i}\mu(x)|\frac{|\hat\mu_T(x) - \mu(x)|}{\delta(\delta - \zeta_T)}.
\end{aligned}$$

Taking the q^{th} power, integrating, using that $\|\partial_{x_i}\mu\|_\infty \lesssim \|\mu\|_{C^2} \leq m$ and (4.22) gives for large enough $T > 0$,

$$\inf_{\substack{B\,:\,\nabla B = \frac{1}{2}\nabla\log\mu,\\ \mu\in\Gamma}} P_B^T\left(\|\nabla\log\hat\mu_T - \nabla\log\mu\|_q \leq C(\delta, m)\|\hat\mu_T - \mu\|_{W^{1,q}}\right) \geq 1 - 2e^{-MT\varepsilon_T^2} \tag{4.23}$$

since $\zeta_T \to 0$. It thus suffices to prove an exponential inequality for $\|\hat\mu_T - \mu\|_{W^{1,q}}$.

For $1 \leq q \leq 2$, we have the continuous embedding $H^1(\mathbb{T}^d) = W^{1,2}(\mathbb{T}^d) \subset W^{1,q}(\mathbb{T}^d)$. For the variance term, by Hilbert space duality,

$$\begin{aligned}
\|\hat\mu_T - P_J\mu\|_{H^1} &= \sup_{\varphi\in C^\infty\,:\,\|\varphi\|_{H^{-1}}\leq 1}\left|\int_{\mathbb{T}^d}[\hat\mu_T(x) - P_J\mu(x)]\varphi(x)dx\right| \\
&= \sup_{\varphi\in V_J\,:\,\|\varphi\|_{H^{-1}}\leq 1}\left|\sum_{l\leq J}\sum_r\langle\hat\mu_T - \mu, \Phi_{lr}\rangle_2\langle\varphi, \Phi_{lr}\rangle_2\right| \\
&= \sup_{\varphi\in V_J\,:\,\|\varphi\|_{H^{-1}}\leq 1}\left|\frac{1}{T}\int_0^T\sum_{l\leq J}\sum_r[\Phi_{lr}(X_s) - \langle\mu, \Phi_{lr}\rangle_2]\langle\varphi, \Phi_{lr}\rangle_2 ds\right| \\
&= \frac{1}{\sqrt{T}}\sup_{\varphi\in V_J\,:\,\|\varphi\|_{H^{-1}}\leq 1}|\mathbb{G}_T[g_\varphi]|,
\end{aligned}$$

where $g_\varphi(u) = \sum_{l\leq J}\sum_r [\Phi_{lr}(u) - \langle \mu, \Phi_{lr}\rangle_2]\langle \varphi, \Phi_{lr}\rangle_2 \in V_J \cap L^2_\mu$. Using that $\Phi_{-10} \equiv 1$, for $\|\varphi\|_{H^{-1}} \leq 1$,

$$\begin{aligned}
\|g_\varphi\|^2_{H^{-1}} &= \sum_{l=0}^{J}\sum_r 2^{-2l}|\langle \varphi, \Phi_{lr}\rangle_2|^2 + \left|\sum_{l\leq J}\sum_r \langle \mu, \Phi_{lr}\rangle_2 \langle \varphi, \Phi_{lr}\rangle_2\right|^2 \\
&\leq \|\varphi\|^2_{H^{-1}} + |\langle \mu, \varphi\rangle_2|^2 \\
&\leq (1 + \|\mu\|^2_{H^1})\|\varphi\|^2_{H^{-1}} \\
&\lesssim 1 + \|\mu\|^2_{C^2} \leq 1 + m^2.
\end{aligned}$$

Applying now Proposition 4.1 with $D_{\mathcal{F}_T} = \dim(V_J) = O(2^{Jd})$, $|\mathcal{F}_T|_{H^{-1}} = \sup_\varphi \|g_\varphi\|_{H^{-1}} \lesssim 1 + m$ and $x = MT\varepsilon_T^2 \to \infty$ gives

$$P_B^T\left(\|\hat{\mu}_T - P_J\mu\|_{H^1} \geq C(1+M)\left(T^{-1/2}2^{Jd/2} + \varepsilon_T\right)\right) \leq 2e^{-MT\varepsilon_T^2},$$

where we have used $2^{J[d/2+\kappa+(d/2+\kappa-1)_+]}\varepsilon_T \lesssim 1$ and $T^{-1/2}2^{J[d+\kappa+(d/2+\kappa-1)_+]} \lesssim 1$ and where the constant in the last display again depends on d, κ and m. Using the embedding $H^1(\mathbb{T}^d) \subset W^{1,q}(\mathbb{T}^d)$, $1 \leq q \leq 2$, this yields

$$P_B^T\left(\|\hat{\mu}_T - \mu\|_{W^{1,q}} \geq \|\mu - P_J\mu\|_{W^{1,q}} + C(1+M)\left(T^{-1/2}2^{Jd/2} + \varepsilon_T\right)\right) \leq 2e^{-MT\varepsilon_T^2}.$$

Combining the last inequality with (4.23) proves the result.

□

Proof of Lemma 4.2. Consider the test $\phi_T = 1\{\|\nabla \log \hat{\mu}_T - \nabla \log \mu_0\|_q \geq M_0\xi_T\}$, where $\hat{\mu}_T$ is the wavelet estimator in (4.21) and M_0 is to be selected below. Since $T^{-1/2}2^{Jd/2} + \varepsilon_T \lesssim \xi_T$ and $\|P_J\mu_0 - \mu_0\|_{W^{1,q}} \leq D_0\xi_T$ by assumption, Lemma 4.3 gives that for any $M > 0$ and large enough $T > 0$,

$$P_{B_0}^T\left(\|\nabla \log \hat{\mu}_T - \nabla \log \mu_0\|_q \geq C(D_0 + 1 + M)\xi_T\right) \leq 4e^{-MT\varepsilon_T^2}.$$

Taking $M_0 > C(2 + D_0)$, the type-I error then satisfies $P_{B_0}^T\phi_T \leq 4e^{-T\varepsilon_T^2} \to 0$.

Turning to the type-II error, since $\|\mu - P_J\mu\|_{W^{1,q}} \leq C_\Lambda\xi_T$ for all $\mu \in \Lambda_T$, and $\Lambda_T \subset \Gamma(\delta, m)$ for $\Gamma(\delta, m)$ the set in Lemma 4.3, applying that lemma yields that for all $M > 0$ and large enough $T > 0$,

$$\sup_{\substack{B\,:\,\nabla B = \frac{1}{2}\nabla \log \mu,\\ \mu \in \Gamma}} P_B^T\left(\|\nabla \log \hat{\mu}_T - \nabla \log \mu\|_q \geq C(C_\Lambda + 1 + M)\xi_T\right) \leq 4e^{-MT\varepsilon_T^2}.$$

Now consider $B \in \Theta_T$ (implying $\nabla B = \frac{1}{2}\nabla \log \mu$ for some $\mu \in \Lambda_T$) such that $\|\nabla B - \nabla B_0\|_q = \frac{1}{2}\|\nabla \log \mu - \nabla \log \mu_0\|_q \geq D\xi_T$. Applying the triangle inequality and the last display,

$$\begin{aligned} P_B^T(1-\phi_T) &= P_B^T(\|\nabla \log \hat{\mu}_T - \nabla \log \mu_0\|_q \leq M_0\xi_T) \\ &\leq P_B^T(\|\nabla \log \mu_0 - \nabla \log \mu\|_q - \|\nabla \log \mu - \nabla \log \hat{\mu}_T\|_q \leq M_0\xi_T) \\ &\leq P_B^T((2D - M_0)\xi_T \leq \|\nabla \log \mu - \nabla \log \hat{\mu}_T\|_q) \leq 4e^{-MT\varepsilon_T^2}, \end{aligned}$$

for $D > M_0/2 + C(C_\Lambda + 1 + M)/2$. This completes the proof.

□

4.5 Proofs for Gaussian and p-exponential priors

4.5.1 Proof of Theorem 4.1

We verify the assumptions of Theorem 4.3 with $q = 2$, $\varepsilon_T = \xi_T \simeq T^{-\alpha/(2\alpha+d)}$ and $2^J \simeq T^{1/(2\alpha+d)}$. In particular, the quantitative conditions (4.15) in Theorem 4.3 are then satisfied for all $\alpha > d/2 + (d/2-1)_+$.

As pointed out in Remark 4.2, the 'small ball condition' (assumption (ii) in Theorem 4.3) follows if we show that for some $C > 0$,

$$\Pi\left(\sum_{i=1}^{d} \|\partial x_i B - \partial x_i B_0\|_\infty \leq \varepsilon_T^2\right) \geq e^{-CT\varepsilon_T^2}. \tag{4.24}$$

Note that

$$\left(\sum_{i=1}^{d} \|\partial_{x_i} B - \partial_{x_i} B_0\|_\infty^2\right)^{1/2} \leq \sum_{i=1}^{d} \|\partial_{x_i} B - \partial_{x_i} B_0\|_\infty \lesssim \|B - B_0\|_{C^1},$$

and since $\|B_0 - B_{0,T}\|_{C^1} = O(T^{-\alpha/(2\alpha+d)}) = O(\varepsilon_T)$ by assumption, it thus suffices to lower bound $\Pi(\|B - B_{0,T}\|_{C^1} \leq \varepsilon_T/2)$, upon replacing ε_T with a multiple of itself if necessary. Recall that the RKHS $\mathcal{H}_B$ of the scaled Gaussian process $B = W/T^{d/(4\alpha+2d)} = W/(\sqrt{T}\varepsilon_T)$ equals the RKHS $\mathcal{H}$ of W, with scaled norm $\|h\|_{\mathcal{H}_B} = \sqrt{T}\varepsilon_T\|h\|_{\mathcal{H}}$. Since $\|B_{0,T}\|_{\mathcal{H}} = O(1)$, using Corollary 2.6.18 of [104] we lower bound the probability of interest by

$$e^{-\frac{1}{2}\|B_{0,T}\|_{\mathcal{H}_B}^2}\Pi(\|B\|_{C^1} \leq \varepsilon_T/2) \geq e^{-c_1T\varepsilon_T^2}\Pi_W(\|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2/2)$$

for some $c_1 > 0$. Since $\sqrt{T}\varepsilon_T^2 \to 0$ for $\alpha > d/2$, the small ball estimate (4.29) below gives

$$\Pi_W(\|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2/2) \geq e^{-c_2(\sqrt{T}\varepsilon_T^2)^{-2d/(2\alpha-d)}} = e^{-c_3 T\varepsilon_T^2},$$

for some $c_2, c_3 > 0$. Taking $C = c_1 + c_3 < \infty$, the last two displays yields (4.24) as required.

For $M > 0$ and $p = 2$, let $\mathcal{B}_T$ be the set in (4.27) and define $\Lambda_T = \{\mu_B = e^{2B}/\int_{\mathbb{T}^d} e^{2B}dx : B \in \mathcal{B}_T\}$. Taking $M > 0$ large enough, Lemma 4.5 (i) implies that $\Pi(\mathcal{B}_T^c) \leq e^{-(C+4)T\varepsilon_T^2}$ as required by assumption (i) in Theorem 4.3. We now show that Λ_T satisfies the inclusion condition (4.16). Since $\|B\|_\infty \leq \|B\|_{C^{(d/2+\kappa)\vee 2}} \leq m$ for every $B \in \mathcal{B}_T$, it holds that $\mu_B \geq e^{-4m} > 0$ for all $\mu_B \in \Lambda_T$. Moreover, using again the boundedness of $B \in \mathcal{B}_T$,

$$\|\mu_B\|_{C^{(d/2+\kappa)\vee 2}} \leq e^{2m}\|e^{2B}\|_{C^{(d/2+\kappa)\vee 2}} \lesssim 1 + \|B\|_{C^{(d/2+\kappa)\vee 2}} + \|B\|^{(d/2+\kappa)\vee 2}_{C^{(d/2+\kappa)\vee 2}} \leq c(m),$$

where for $d \leq 3$ the second inequality (with $C^{(d/2+\kappa)\vee 2} = C^2$) follows readily by differentiation, while for $d \geq 4$ it is implied by Lemma 4.6 upon noting that $C^{(d/2+\kappa)\vee 2} = C^{d/2+\kappa} = B^{d/2+\kappa}_{\infty\infty}$, since $d/2 + \kappa \notin \mathbb{N}$ (cfr. Chapter 3 in [211]). Finally, the bias bound follows from Lemma 4.7 with $p = 2$. This shows that Λ_T satisfies the required assumptions in Theorem 4.3.

It remains only to consider the true potential B_0. However, since $B_0 \in H^{\alpha+1}$, Lemma 4.6 similarly implies that $\|\mu_0 - P_J\mu_0\|_{W^{1,2}} \lesssim 2^{-J\alpha}\|\mu_0\|_{H^{\alpha+1}} \simeq \varepsilon_T$. The result thus follows from Theorem 4.3.

□

4.5.2 Proof of Theorem 4.2

We verify the assumptions of Theorem 4.3 with $q = p$, $\varepsilon_T = \xi_T \simeq T^{-\alpha/(2\alpha+d)}$ and $2^J \simeq T^{1/(2\alpha+d)}$. The quantitative conditions (4.15) are satisfied since we have assumed $\alpha > [(d/2+\kappa)\vee 2] + d/p - 1 > d/2 + (d/2-1)_+$. The remaining assumptions are verified using tools for p-exponential random elements mainly due to [5].

We first consider assumption (ii) in Theorem 4.3. Since $\alpha > [(d/2+\kappa)\vee 2] + d/p - 1 > d/2 + (d/2+\kappa)\vee 1$, and since the prior Π arising as the law of B in (4.14) is supported on $V_J \cap \dot{L}^2$, Lemma 4.4 below implies that it is enough to show that for some $M, C > 0$

$$\Pi(\|\nabla B - \nabla B_0\|_{L^2(\mu_0)} \leq M\varepsilon_T) \geq e^{-CT\varepsilon_T^2},$$

where $\mu_0 \propto e^{2B_0}$ is the invariant density and $\|f\|^2_{L^2(\mu_0)} = \int_{\mathbb{T}^d} |f(x)|^2 \mu_0(x)dx$. Since μ_0 is bounded, the probability on the left hand side is greater than

$$\Pi(\|\nabla B - \nabla B_0\|_2 \le m_1 \varepsilon_T) \ge \Pi(\|B - B_0\|_{H^1} \le m_2 \varepsilon_T)$$

for some $m_1, m_2 > 0$. Let $P_J B_0$ the wavelet projection of $B_0 \in H^{\alpha+1} \cap \dot{L}^2$ onto V_J. Then $\|B_0 - P_J B_0\|_{H^1} \lesssim 2^{-J\alpha} \simeq \varepsilon_T$, so that by the triangle inequality the latter probability is lower bounded by $\Pi(\|B - P_J B_0\|_{H^1} \le m_3 \varepsilon_T)$ for some $m_3 > 0$. In the language of [5], the $\mathcal{Z}$-space (Definition 2.8 in [5]) associated to the p-exponential random element $B = W/(T^{d/(2\alpha+d)})^{1/p} = W/(T\varepsilon_T^2)^{1/p}$ is equal to $V_J \cap \dot{L}^2$, with norm

$$\|h\|_{\mathcal{Z}} = \left(T\varepsilon_T^2 \sum_{l=0}^{J} \sum_r 2^{pl\left(\alpha+1+\frac{d}{2}-\frac{d}{p}\right)} |\langle h, \Phi_{lr}\rangle_2|^p \right)^{\frac{1}{p}} = \left(T\varepsilon_T^2\right)^{\frac{1}{p}} \|h\|_{B^{\alpha+1}_{pp}}, \qquad h \in \mathcal{Z}. \quad (4.25)$$

Since $P_J B_0 \in \mathcal{Z}$, by Proposition 2.11 in [5] the probability of interest is thus greater than

$$e^{-\frac{1}{p}\|P_J B_0\|^p_{\mathcal{Z}}} \Pi\Big(\|B\|_{H^1} \le m_3 \varepsilon_T\Big) \ge e^{-\frac{1}{p} T\varepsilon_T^2 \|B_0\|^p_{B^{\alpha+1}_{pp}}} \Pi\Big(\|B\|_{H^1} \le m_3 \varepsilon_T\Big),$$

where $\|B_0\|_{B^{\alpha+1}_{pp}} < \infty$ in view of the continuous embedding $H^{\alpha+1}(\mathbb{T}^d) \subseteq B^{\alpha+1}_{pp}(\mathbb{T}^d)$ holding for all $p \le 2$ (e.g., p. 360 in [104]). We conclude estimating the above centred small ball probability. Using Theorem 4.2 in [13] (whose conclusion can readily be adapted for double index sums), we have as $T \to \infty$,

$$\begin{aligned} -\log \Pi\Big(\|B\|_{H^1} \le m_3 \varepsilon_T\Big) &= -\log \Pi_W\Big(\|W\|_{H^1} \le m_3 (T\varepsilon_T^2)^{\frac{1}{p}} \varepsilon_T\Big) \\ &\simeq \Big[(T\varepsilon_T^2)^{\frac{1}{p}} \varepsilon_T\Big]^{-\frac{d}{\alpha - d/p}} \\ &= T\varepsilon_T^2. \end{aligned}$$

Thus, for $c_1 > 0$ a large enough constant and $C = c_1 + \|B_0\|^p_{B^{\alpha+1}_{pp}}/p < \infty$, we obtain as required that $\Pi\left(\|\nabla B - \nabla B_0\|_{L^2(\mu_0)} \le M\varepsilon_T^2\right) \ge e^{-CT\varepsilon_T^2}$.

The remaining conditions in Theorem 4.3 are verified arguing as in the proof of Theorem 4.1, using the sets $\mathcal{B}_T$ in (4.27) with $M > 0$ large enough and $p \in [1, 2]$, taking $\Lambda_T = \{\mu_B = e^{2B}/\int_{\mathbb{T}^d} e^{2B}dx : B \in \mathcal{B}_T\}$, and noting that since $B_0 \in B^{\alpha+1}_{pp}$, we have by Lemma 4.6 that $\|\mu_0 - P_J\mu_0\|_{W^{1,p}} \lesssim 2^{-J\alpha}\|\mu_0\|_{B^{\alpha+1}_{pp}} \lesssim \varepsilon_T$.

□

Lemma 4.4. *Suppose $B_0 \in H^{\alpha+1}(\mathbb{T}^d) \cap \dot{L}^2(\mathbb{T}^d)$ for $\alpha > d/2 + (d/2+\kappa) \vee 1$ and some $\kappa > 0$. If $\varepsilon_T = T^{-\alpha/(2\alpha+d)}$, $J \in \mathbb{N}$ satisfies $2^J \simeq T^{1/(2\alpha+d)}$ and $M > 0$, then as $T \to \infty$,*

$$P_{B_0}^T \left(\sup_{B \in V_J \cap \dot{L}^2 \,:\, \|\nabla B - \nabla B_0\|_{L^2(\mu_0)} \le M\varepsilon_T} \frac{1}{T} \int_0^T \|\nabla B(X_s) - \nabla B_0(X_s)\|^2 ds \le M^2 \varepsilon_T^2 + o(\varepsilon_T^2) \right) \to 1.$$

The restriction to $\dot{L}^2$ is for identifiability and simplifies certain norms. A similar result holds under other identifiability constraints.

Proof. We first show that $\frac{1}{T}\int_0^T \|\nabla B(X_s) - \nabla B_0(X_s)\|^2 ds$ is close to its ergodic average $\|\nabla B - \nabla B_0\|_{L^2(\mu_0)}^2$, uniformly over $\mathcal{SB}_T = \{B \in V_J \cap \dot{L}^2 : \|\nabla B - \nabla B_0\|_{L^2(\mu_0)} \le M\varepsilon_T\}$. Define

$$\mathcal{F}_T = \left\{ f_B(x) := \|\nabla B(x) - \nabla B_0(x)\|^2 - \|\nabla B - \nabla B_0\|_{L^2(\mu_0)}^2 : B \in \mathcal{SB}_T \cup \{B_0\} \right\}.$$

Since $x \mapsto \|x\|^2$ is a smooth function, $\mathcal{SB}_T \subset V_J$ and $B_0 \in H^{\alpha+1}$, it holds that $\mathcal{F}_T \subset L^2_{\mu_0} \cap H^s$. By Lemma 4.8, the Poisson equation $L_{B_0} u = f_B$ (writing L_{B_0} for $L_{\nabla B_0}$) has a unique solution $L_{B_0}^{-1}[f_B] \in \dot{L}^2$ for any $B \in \mathcal{SB}_T$. Applying Lemma 1 of [173], since $\alpha > d/2 + (d/2+\kappa) \vee 1$, gives that for any $x \ge 0$,

$$P_{B_0}^T \left(\sup_{f_B \in \mathcal{F}_T} |\mathbb{G}_T[f_B]| \ge \sup_{f \in \mathcal{F}_T} \frac{2\|L_{B_0}^{-1}[f_B]\|_\infty}{\sqrt{T}} + J_{\mathcal{F}_T}(4\sqrt{2} + 192x) \right) \le e^{-x^2/2}, \tag{4.26}$$

where $J_{\mathcal{F}_T} = \int_0^{D_{\mathcal{F}_T}} \sqrt{\log 2N(\tau; \mathcal{F}_T, 6d_L)} d\tau$, $d_L^2(f,g) = \sum_{i=1}^d \|\partial_{x_i} L_{B_0}^{-1}[f-g]\|_\infty^2$ and $D_{\mathcal{F}_T}$ is the d_L-diameter of $\mathcal{F}_T$. We now proceed to bound $J_{\mathcal{F}_T}$.

For $B \in \mathcal{SB}_T$, write $h_i = \partial_{x_i}(B - B_0)$ so that $f_B = \sum_{i=1}^d h_i^2 - \|h_i\|_{L^2(\mu_0)}^2$. Using the Sobolev embedding theorem, Lemma 4.8 and the Runst-Sickel lemma ([173], Lemma 2), for $\kappa > 0$,

$$\begin{aligned} d_L(f_B, f_{\bar{B}}) &\lesssim \|f_B - f_{\bar{B}}\|_{H^{d/2+\kappa-1}} \\ &\lesssim \sum_{i=1}^d \left\| h_i^2 - \bar{h}_i^2 - \int_{\mathbb{T}^d} (h_i^2 - \bar{h}_i^2) d\mu_0 \right\|_{H^{d/2+\kappa-1}} \\ &\lesssim \sum_{i=1}^d \|h_i - \bar{h}_i\|_\infty \|h_i + \bar{h}_i\|_{H^{(d/2+\kappa-1)_+}} + \|h_i - \bar{h}_i\|_{H^{(d/2+\kappa-1)_+}} \|h_i + \bar{h}_i\|_\infty \\ &\qquad + \|\mu_0\|_\infty \|h_i - \bar{h}_i\|_2 \|h_i + \bar{h}_i\|_2 \|1\|_{H^{d/2+\kappa-1)}}, \end{aligned}$$

where the constants depend only on d, κ and $\|\nabla B_0\|_{B^{|d/2+\kappa-1|}_{\infty\infty}} \lesssim \|B_0\|_{H^{|d/2+\kappa-1|+d/2+1}} \lesssim \|B_0\|_{H^{\alpha+1}}$, and we note that $\|\mu_0\|_\infty \leq e^{4\|B_0\|_\infty} < \infty$ using (4.6). Since $B, \bar{B} \in V_J$, we have $\|h_i - \bar{h}_i\|_{H^u} \lesssim \|B - \bar{B}\|_{H^{u+1}} \leq 2^{Ju}\|B - \bar{B}\|_{H^1}$ for $u \geq 0$ and $\|h_i - \bar{h}_i\|_\infty \lesssim \|B - \bar{B}\|_{B^1_{1\infty}} \lesssim 2^{Jd/2}\|B - \bar{B}\|_{H^1}$. Furthermore,

$$\begin{aligned}\|h_i + \bar{h}_i\|_{H^u} \lesssim \sup_{B \in \mathcal{SB}_T} \|B - B_0\|_{H^{u+1}} &\lesssim \sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{H^{u+1}} + \|B_0 - P_J B_0\|_{H^{u+1}} \\ &\lesssim 2^{Ju} \sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{H^1} + 2^{-J(\alpha-u)}\|B_0\|_{H^{\alpha+1}}.\end{aligned}$$

Note that $\|\nabla B - \nabla B_0\|_{L^2(\mu_0)} \simeq \|B - B_0\|_{H^1}$ since μ_0 is both bounded and bounded away from zero and $B, B_0 \in \dot{L}^2$. Therefore, $\sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{H^1} + 2^{-J\alpha}\|B_0\|_{H^{\alpha+1}} \leq C(M+1)\varepsilon_T$, so that $\|h_i + \bar{h}_i\|_{H^u} \leq C(M+1)2^{Ju}\varepsilon_T$. Similarly, using the Sobolev embedding theorem,

$$\begin{aligned}\|h_i + \bar{h}_i\|_\infty \lesssim \sup_{B \in \mathcal{SB}_T} \|B - B_0\|_{B^1_{1\infty}} &\lesssim \sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{B^1_{1\infty}} + \|B_0 - P_J B_0\|_{B^1_{1\infty}} \\ &\lesssim 2^{Jd/2} \sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{H^1} + 2^{-J(\alpha-d/2)}\|B_0\|_{H^{\alpha+1}} \\ &\leq C(M+1)2^{Jd/2}\varepsilon_T.\end{aligned}$$

Combining these bounds yields

$$d_L(f_B, f_{\bar{B}}) \leq C(M+1)2^{J[d/2+(d/2+\kappa-1)_+]}\varepsilon_T\|B - \bar{B}\|_{H^1}.$$

Since $\mathcal{SB}_T \subseteq (V_J, \|\cdot\|_{H^1})$ is finite dimensional, using the last display and the usual covering argument for balls in finite dimensional spaces (e.g., [104], Proposition 4.3.34),

$$\begin{aligned}\log N(\tau; \mathcal{F}_T, 6d_L) &\leq \log N\left(\tau; \mathcal{SB}_T, C(M+1)2^{J[d/2+(d/2+\kappa-1)_+]}\varepsilon_T\|\cdot\|_{H^1}\right) \\ &\lesssim \dim(V_J)\log\left(C(M+1)2^{J[d/2+(d/2+\kappa-1)_+]}\varepsilon_T \sup_{B,\bar{B} \in \mathcal{SB}_T} \|B - \bar{B}\|_{H^1}/\tau\right) \\ &\lesssim \dim(V_J)\log\left(CR_T/\tau\right),\end{aligned}$$

where $C > 0$ and

$$R_T := (M+1)^2 2^{J[d/2+(d/2+\kappa-1)_+]}\varepsilon_T^2 \to 0$$

under the current assumption on α. Recall the inequality $\int_0^a \sqrt{\log(A/x)}dx \leq 4a\sqrt{\log(A/a)}$ for any $A \geq 2$ and $0 < a \leq 1$ ([104], p. 190). Using this inequality, the last display with

$\dim(V_J) = O(2^{Jd})$, and that $\mathcal{F}_T$ has d_L-diameter $D_{\mathcal{F}_T} \lesssim R_T \to 0$, we obtain

$$J_{\mathcal{F}_T} \lesssim \dim(V_J) \int_0^{D_{\mathcal{F}_T}} \sqrt{\log([CR_T] \vee 2/\tau)} d\tau \lesssim 2^{Jd/2} D_{\mathcal{F}_T} \sqrt{\log([CR_T] \vee 2/D_{\mathcal{F}_T})}.$$

Taking $D_{\mathcal{F}_T} \simeq R_T$ in the last display gives $J_{\mathcal{F}_T} \lesssim 2^{Jd/2} R_T \left(1 + \sqrt{\log(1/R_T)}\right)$. Arguing as for the bound for $d_L(f_B, f_{\bar{B}})$ above, one has for all $B \in \mathcal{SB}_T$,

$$\begin{aligned} \|L_{B_0}^{-1}[f_B]\|_\infty &\lesssim \|f_B\|_{H^{d/2+\kappa-1}} \\ &\lesssim \sum_{i=1}^d \left\| h_i^2 - \int_{\mathbb{T}^d} h_i^2 d\mu_0 \right\|_{H^{d/2+\kappa-1}} \\ &\lesssim \sum_{i=1}^d \|h_i\|_\infty \|h_i\|_{H^{(d/2+\kappa-1)_+}} \\ &\lesssim 2^{J[d/2+(d/2+\kappa-1)_+]} (M+1)^2 \varepsilon_T^2 \end{aligned}$$

where the constants depend only on d, κ, $\|\mu_0\|_\infty$ and $\|\nabla B_0\|_{B_{\infty\infty}^{|d/2+\kappa-1|}} \lesssim \|B_0\|_{H^{|d/2+\kappa-1|+d/2+1}} \lesssim \|B_0\|_{H^{\alpha+1}}$. Substituting this bound and $J_{\mathcal{F}_T} \lesssim 2^{Jd/2} R_T \left(1 + \sqrt{\log(1/R_T)}\right)$ into (4.26),

$$P_{B_0}^T \left(\sup_{f_B \in \mathcal{F}_T} |\mathbb{G}_T[f_B]| \geq C 2^{Jd/2} R_T \left(1 + \sqrt{\log(1/R_T)}\right) (1+x) \right) \leq e^{-x^2/2}$$

for any $x \geq 0$. Set

$$\zeta_T = M_T T^{-1/2} 2^{Jd/2} R_T \left(1 + \sqrt{\log(1/R_T)}\right) = O\left(M_T \sqrt{\log T} T^{-\frac{\alpha - d/2 - (d/2+\kappa-1)_+}{2\alpha+d}} \varepsilon_T^2 \right),$$

which satisfies $\zeta_T = o(\varepsilon_T^2)$ for $M_T \to \infty$ growing slow enough, since $\alpha > d/2 + |d/2+\kappa-1|$. Then using the definition of the empirical process $\mathbb{G}_T[f_B]$, for any $M_T \to \infty$,

$$P_{B_0}^T \left(\sup_{B \in \mathcal{SB}_T} \left| \frac{1}{T} \int_0^T \|\nabla B(X_s) - \nabla B_0(X_s)\|^2 ds - \|\nabla B - \nabla B_0\|_{L^2(\mu_0)}^2 \right| \geq \zeta_T \right) \to 0$$

as $T \to \infty$. The result then follows because on the complement of the event in the last display,

$$\begin{aligned} \sup_{B \in \mathcal{SB}_T} \frac{1}{T} \int_0^T \|\nabla B(X_s) - \nabla B_0(X_s)\|^2 ds &\leq \sup_{B \in \mathcal{SB}_T} \|\nabla B - \nabla B_0\|_{L^2(\mu_0)}^2 + \zeta_T \\ &\leq M^2 \varepsilon_T^2 + o(\varepsilon_T^2). \end{aligned}$$

□

4.5.3 Concentration inequalities and bias bounds

In this section, we build on techniques for Gaussian process priors [235], p-exponential priors [5] and rescaled priors [167] to obtain the prior bias bounds needed to apply our general contraction theorem for multi-dimensional diffusions.

Lemma 4.5. *For $s, M, \kappa > 0$, $p \in [1, 2]$ and sequences $\varepsilon_T = T^{-\alpha/(2\alpha+d)}$, $\bar{\varepsilon}_T = T^{-(\alpha+1)/(2\alpha+d)}$, define the sets*

$$\mathcal{B}_T = \{B = B_1 + B_2 : \|B_1\|_\infty \leq \bar{\varepsilon}_T, \|B_1\|_{C^1} \leq \varepsilon_T, \|B_2\|_{B^{\alpha+1}_{pp}} \leq M, \|B\|_{C^{(d/2+\kappa)\vee 2}} \leq M\}. \tag{4.27}$$

Assume either:

(i) $p = 2$ and $B = W/(\sqrt{T}\varepsilon_T)$ for $W \sim \Pi_W$ a Gaussian process satisfying Condition 4.1;

(ii) $B = W/(T\varepsilon_T^2)^{\frac{1}{p}}$ for $W \sim \Pi_W$ a p-exponential random element as in (4.13) with $\alpha > (d/2 + \kappa) \vee 2 + d/p - 1$.

Let Π be the law of B. Then, for every $K > 0$, there exists $M > 0$ large enough such that $\Pi(\mathcal{B}_T^c) \leq e^{-KT\varepsilon_T^2}$.

Proof. (i) In the Gaussian case, define the sets

$$\mathcal{B}_{T,1} := \{B = B_1 + B_2 : \|B_1\|_\infty \leq \bar{\varepsilon}_T, \|B_1\|_{C^1} \leq \varepsilon_T, \|B_2\|_{B^{\alpha+1}_{22}} \leq m\},$$
$$\mathcal{B}_{T,2} := \{B : \|B\|_{C^{(d/2+\kappa)\vee 2}} \leq m\}.$$

To upper bound $\Pi(\mathcal{B}_T^c)$ it thus suffices to upper bound $\Pi(\mathcal{B}_{T,1}^c)$ and $\Pi(\mathcal{B}_{T,2}^c)$. Since $\|g\|_{B^{\alpha+1}_{22}} = \|g\|_{H^{\alpha+1}} \leq c_0\|g\|_{\mathcal{H}}$ for all $g \in \mathcal{H}$ under Condition 4.1, Borell's isoperimetric inequality ([104], Theorem 2.6.12) gives

$$\begin{aligned}
&\Pi(\mathcal{B}_{T,1}) \\
&\quad = \Pi_W(W = W_1 + W_2 : \|W_1\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T, \|W_1\|_{C^1} \leq \sqrt{T}\varepsilon_T^2, \|W_2\|_{B^{\alpha+1}_{22}} \leq m\sqrt{T}\varepsilon_T) \\
&\quad \geq \Phi\left(\Phi^{-1}\left(\Pi_W\left(W : \|W\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T, \|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2\right)\right) + m\sqrt{T}\varepsilon_T/c_0\right),
\end{aligned} \tag{4.28}$$

where Φ is the standard normal cumulative distribution function. Now for $\mathcal{H}_1$ and $H_1^{\alpha+1}$ the unit balls of $\mathcal{H}$ and $H^{\alpha+1}$ respectively, we have under Condition 4.1 that

$$\log N(\tau; \mathcal{H}_1, \|\cdot\|_{C^1}) \leq \log N(\tau; c_0 H_1^{\alpha+1}, \|\cdot\|_{C^1}) \lesssim \tau^{-d/\alpha},$$

where the last inequality follows by arguing as in Theorem 4.3.36 of [104]. By Theorem 1.2 of Li and Linde [152], this yields

$$\Pi_W(\|W\|_{C^1} \leq \eta) \geq e^{-c_1^2 \eta^{-2d/(2\alpha-d)}} \qquad \text{as } \eta \to 0, \tag{4.29}$$

for any $d/2 < s < \infty$ and some $c_1 = c_1(d, \alpha, c_0) > 0$, which implies, since $\sqrt{T}\varepsilon_T^2 \to 0$,

$$\Pi_W\left(\|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2\right) \geq e^{-c_1^2(\sqrt{T}\varepsilon_T^2)^{-2d/(2\alpha-d)}} = e^{-c_1^2 T\varepsilon_T^2}.$$

Using the same argument, now with the bound

$$\log N(\tau; \mathcal{H}_1, \|\cdot\|_\infty) \leq \log N(\tau; c_0 H_1^{\alpha+1}, \|\cdot\|_\infty) \lesssim \tau^{-d/(\alpha+1)},$$

it follows for some $c_2 > 0$ that $\Pi_W(\|W\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T) \geq e^{-c_2^2 T\varepsilon_T^2}$. The Gaussian correlation inequality (which holds for Gaussian measures in separable Banach spaces, see Lemma 4.9 below) then gives for $c_3 = c_1^2 + c_2^2$

$$\begin{aligned}\Pi_W\Big(\|W\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T, &\|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2\Big) \\ &\geq \Pi_W\left(\|W\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T\right)\Pi_W\left(\|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2\right) \\ &\geq e^{-c_3 T\varepsilon_T^2}.\end{aligned}$$

Using the standard inequality $\Phi^{-1}(y) \geq -\sqrt{2\log(1/y)}$ for $0 < y < 1$, the right hand side of (4.28) is thus lower bounded by

$$\Phi\left((m/c_0 - \sqrt{2}c_3)\sqrt{T}\varepsilon_T\right).$$

Defining $m_T = -\Phi^{-1}(e^{-KT\varepsilon_T^2}/3)$, this further gives $m_T \leq \sqrt{2\log 2} + \sqrt{2KT}\varepsilon_T$, which can be made smaller than $(m/c_0 - \sqrt{2}c_3)\sqrt{T}\varepsilon_T$ by taking $M = M(K, c_0, c_3)$ large enough. For such M, the last display is lower bounded by $\Phi(m_T) = 1 - \Phi(\Phi^{-1}(e^{-KT\varepsilon_T^2}/2)) = 1 - \frac{1}{3}e^{-KT\varepsilon_T^2}$.

To bound $\Pi(\mathcal{B}_{T,2}^c)$, recall that by Condition 4.1 W defines a Gaussian Borel random element in a separable linear subspace $\mathcal{S}$ of $C^{(d/2+\kappa)\vee 2}$. Using the Hahn-Banach theorem, we may thus represent its norm as

$$\|W\|_{C^{(d/2+\kappa)\vee 2}} = \sup_{L\in\mathcal{L}} |L(W)|,$$

where $\mathcal{L}$ is a countable set of bounded linear functionals on $(\mathcal{S}, \|\cdot\|_{C^{(d/2+\kappa)\vee 2}})$. Applying Fernique's theorem [104, Theorem 2.1.20] to the centred Gaussian process $(X(L) = L(W) : L \in \mathcal{L})$, we have $E\|W\|_{C^{(d/2+\kappa)\vee 2}} = E\sup_{L\in\mathcal{L}}|X(L)| \leq D < \infty$, and for $m = m(D) > 0$ large enough and since $\sqrt{T}\varepsilon_T \to \infty$,

$$\Pi(\mathcal{B}^c_{T,2}) \leq \Pi_W\left(W : \|W\|_{C^{(d/2+\kappa)\vee 2}} - E\|W\|_{C^{(d/2+\kappa)\vee 2}} \geq M\sqrt{T}\varepsilon_T\right) \leq 2e^{-D'M^2T\varepsilon_T^2}$$

for some fixed constant $D' > 0$. Taking $M > 0$ large enough, the right-hand side can be made less than $\frac{1}{3}e^{-KT\varepsilon_T^2}$, concluding the proof.

(ii). Turning to p-exponential priors, define the set

$$\mathcal{B}'_T = \Big\{B = B'_1 + B'_2 + B'_3 \ : B'_i \in V_J \cap \dot{L}^2,\ \|B'_1\|_\infty \leq \bar{\varepsilon}_T, \\ \|B'_2\|_{H^{\alpha+1+d/2-d/p}} \leq M^{\frac{p}{2}}(T\varepsilon_T^2)^{\frac{1}{2}-\frac{1}{p}},\ \ \|B'_3\|_{B^{\alpha+1}_{pp}} \leq M\Big\}.$$

We lower bound the prior probability of $\mathcal{B}'_T$ using the generalization of Borell's inequality to p-exponential measures. The space of admissible shifts (cfr. Proposition 2.7 in [5]) of the scaled p-exponential random element $B = W/(T\varepsilon_T^2)^{\frac{1}{p}}$ is $\mathcal{Q} = V_J \cap \dot{L}^2$, with norm

$$\|h\|_{\mathcal{Q}} = (T\varepsilon_T^2)^{\frac{1}{p}}\left(\sum_{l=0}^{J}\sum_r 2^{2l\left(\alpha+1+\frac{d}{2}-\frac{d}{p}\right)}|\langle h, \Phi_{lr}\rangle_2|^2\right)^{\frac{1}{2}} = \left(T\varepsilon_T^2\right)^{\frac{1}{p}}\|h\|_{H^{\alpha+1+d/2-d/p}}, \quad h \in \mathcal{Q}.$$

Then, recalling the $\mathcal{Z}$-norm defined in (4.25), Proposition 2.15 in [5] implies

$$\Pi(\mathcal{B}'_T) = \Pi\Big(B = B'_1 + B'_2 + B'_3 : \|B'_1\|_\infty \leq \bar{\varepsilon}_T, \|B'_2\|_{\mathcal{Q}} \leq (M^pT\varepsilon_T^2)^{\frac{1}{2}}, \|B'_3\|_{\mathcal{Z}} \leq M(T\varepsilon_T^2)^{\frac{1}{p}}\Big) \\ \geq 1 - \frac{1}{\Pi(\|B\|_\infty \leq \bar{\varepsilon}_T\big)}\exp\left(-(M^p/k)T\varepsilon_T^2\right)$$

for some $k = k(p) > 0$. By an analogous small ball computation as in the proof of Proposition 6.3 in [5], it follows that as $T \to \infty$

$$-\log\Pi(\|B\|_\infty \leq \bar{\varepsilon}_T) = -\log\Pi_W(\|W\|_\infty \leq \bar{\varepsilon}_T(T\varepsilon_T^2)^{\frac{1}{p}}) \\ \simeq [\bar{\varepsilon}_T(T\varepsilon_T^2)^{\frac{1}{p}}]^{\frac{d}{\alpha+1-d/p}} = T\varepsilon_T^2.$$

Thus, for some constant $c_1 = c_1(\alpha, p, d) > 0$, we have $\Pi(\mathcal{B}'_T) \geq 1 - e^{-[(M^p/k)-c_1]T\varepsilon_T^2}$, so that for any $K > 0$ we can choose $M = M(K, c_1, k) = M(K, \alpha, p, d) > 0$ large enough to obtain $\Pi(\mathcal{B}'_T) \geq 1 - e^{-KT\varepsilon_T^2}$.

We conclude the proof by showing that $\mathcal{B}'_T \subset \mathcal{B}_T$. First, since $B'_1 \in V_J$ we have $\|B'_1\|_{C^1} \leq 2^J \|B'_1\|_\infty \lesssim T^{1/(2\alpha+d)}\bar{\varepsilon}_T = \varepsilon_T$, and therefore both norm bounds on $B_1 = B'_1$ in (4.27) are satisfied. Also note that

$$\|B'_1\|_{C^{(d/2+\kappa)\vee 2}} \leq 2^{J(d/2+\kappa)\vee 2}\|B'_1\|_\infty \lesssim T^{\frac{(d/2+\kappa)\vee 2}{2\alpha+d}}\bar{\varepsilon}_T = o(1) \tag{4.30}$$

since by assumption $\alpha+1 > (d/2+\kappa)\vee 2 + d/p$. Next, since also $B'_2 \in V_J$, using the continuous embedding $H^{\alpha+1}(\mathbb{T}^d) \subseteq B^{\alpha+1}_{pp}(\mathbb{T}^d)$, $p \leq 2$,

$$\begin{aligned}\|B'_2\|_{B^{\alpha+1}_{pp}} \lesssim \|B'_2\|_{H^{\alpha+1}} &\leq 2^{Jd\left(\frac{1}{p}-\frac{1}{2}\right)}\|B'_2\|_{H^{\alpha+1+d/2-d/p}} \\ &\leq 2^{Jd\left(\frac{1}{p}-\frac{1}{2}\right)} M^{\frac{p}{2}} (T\varepsilon_T^2)^{\frac{1}{2}-\frac{1}{p}} \simeq M^{\frac{p}{2}}.\end{aligned}$$

Thus, taking $B_2 = B'_2 + B'_3$ implies $\|B_2\|_{B^{\alpha+1}_{pp}} \lesssim M^{p/2} + M \lesssim M$ for $M \geq 1$ as required. Finally, using (4.30) and the continuous embedding $B^{\alpha+1}_{pp} \subset C^{(d/2+\kappa)\vee 2}$, holding for all $\alpha+1 > (d/2+\kappa)\vee 2 + d/p$,

$$\|B\|_{C^{(d/2+\kappa)\vee 2}} \leq \|B'_1\|_{C^{(d/2+\kappa)\vee 2}} + \|B_2\|_{B^{\alpha+1}_{pp}} \lesssim M,$$

concluding the proof.

□

Lemma 4.6. *Let $1 \leq p, q \leq \infty$ and $t > d/p$. If $\|B\|_\infty \leq m$, then*

$$\|e^B\|_{B^t_{pq}} \leq C(1 + \|B\|_{B^t_{pq}} + \|B\|^t_{B^t_{pq}})$$

for some constant $C = C(m,t,p,q) > 0$.

Proof. Consider a function $f \in C^\infty(\mathbb{R})$ such that $f(x) = e^x - 1$ for $|x| \leq m$ and $\|f\|_{L^\infty(\mathbb{R})} \leq 2m$. Since $\|B\|_\infty \leq m$, we have $f \circ B(x) = e^{B(x)} - 1$ for all $x \in \mathbb{T}^d$. By Theorem 11 of Bourdaud and Sickel [45],

$$\|e^B - 1\|_{B^t_{pq}} \leq c\|f\|_{C_b^{\lfloor t \rfloor+1}(\mathbb{R})}(\|B\|_{B^t_{pq}} + \|B\|^t_{B^t_{pq}})$$

for some $c > 0$. Since $\|1\|_{B^t_{pq}} < \infty$ for periodic Besov spaces, the result follows.

□

Lemma 4.7. *For $\alpha, M, \kappa > 0$ and $p \in [1,2]$, let $\mathcal{B}_T$ be the set in (4.27), with $\varepsilon_T, \bar{\varepsilon}_T$ as in Lemma 4.5. If $2^J \simeq T^{1/(2\alpha+d)}$, then there exists a finite constant C depending on α, d, m*

and the wavelet basis $\{\Phi_{lr}\}$ *such that*

$$\left\{\mu_B = \frac{e^{2B}}{\int_{\mathbb{T}^d} e^{2B(x)}dx} : B \in \mathcal{B}_T\right\} \subset \{\mu : \|\mu - P_J\mu\|_{W^{1,p}} \leq C\varepsilon_T\}.$$

Proof. Since $\|B\|_\infty \leq \|B\|_{C^2} \leq M$ for every $B \in \mathcal{B}_T$, this implies $e^{-2M} \leq \int_{\mathbb{T}^d} e^{2B}dx \leq e^{2M}$ and hence $\|\mu_B - P_J\mu_B\|_{W^{1,p}} \leq e^{2M}\|e^{2B} - P_Je^{2B}\|_{W^{1,p}}$, so it suffices to bound the last quantity. For a function f on the torus $\mathbb{T}^d$, denote by $\bar{f}$ its periodic extension to $\mathbb{R}^d$. Recall that the periodic projection satisfies $P_Jf(x) = \int_{\mathbb{R}^d} K_J(x,y)\bar{f}(y)dy$ for all $x \in (0,1]^d$, where $K_J(x,y) = 2^{Jd}\sum_{k\in\mathbb{Z}^d}\phi(2^Jx-k)\phi(2^Jy-k)$ is the unperiodised wavelet kernel and ϕ is the unperiodised father wavelet used in the construction of the periodised wavelet basis (see (4.127) in [104]). Using that $\int_{\mathbb{R}^d} K_J(x,y)dy = 1$ for all $x \in (0,1]^d$ and writing $B = B_1 + B_2$ as in (4.27),

$$\begin{aligned}
|\partial_{x_i}(e^{2B} - P_Je^{2B})(x)| &= \left|\partial_{x_i}\int_{\mathbb{R}^d} K_J(x,y)(e^{2B_1(x)+2B_2(x)} - e^{2\bar{B}_1(y)+2\bar{B}_2(y)})dy\right| \\
&\leq \left|\partial_{x_i}\left(e^{2B_1(x)}\int_{\mathbb{R}^d} K_J(x,y)(e^{2B_2(x)} - e^{2\bar{B}_2(y)})dy\right)\right| \\
&\quad + \left|\partial_{x_i}\left(\int_{\mathbb{R}^d} K_J(x,y)(e^{2B_1(x)} - e^{2\bar{B}_1(y)})e^{2\bar{B}_2(y)}dy\right)\right| \\
&\leq e^{2B_1(x)}\left|2\partial_{x_i}B_1(x)[e^{2B_2(x)} - P_Je^{2B_2}(x)] + \partial_{x_i}[e^{2B_2(x)} - P_Je^{2B_2}(x)]\right| \\
&\quad + \left|\int_{\mathbb{R}^d}\partial_{x_i}K_J(x,y)(e^{2B_1(x)} - e^{2\bar{B}_1(y)})e^{2\bar{B}_2(y)}dy\right| \\
&\quad + e^{2B_1(x)}\left|2\partial_{x_i}B_1(x)\int_{\mathbb{R}^d} K_J(x,y)e^{2\bar{B}_2(y)}dy\right|.
\end{aligned}$$

Using that $\int_{\mathbb{R}^d}\partial_{x_i}K_J(x,y)dy \lesssim 2^J$ by the localization property of wavelets and that $|e^x - 1| \lesssim x$ for small $|x|$, the last display is bounded by a multiple of

$$\begin{aligned}
&e^{2\|B_1\|_\infty}\|B_1\|_{C_1}\left|(e^{2B_2} - P_Je^{2B_2})(x)\right| + e^{2\|B_1\|_\infty}\left|\partial_{x_i}(e^{2B_2} - P_Je^{2B_2})(x)\right| \\
&\quad + 2^Je^{2\|B_2\|_\infty}\|B_1\|_\infty + e^{2\|B_1\|_\infty}\|B_1\|_{C^1}|P_Je^{2B_2}(x)|.
\end{aligned}$$

Taking p^{th} powers and integrating, and using the embedding $B^0_{p1}(\mathbb{T}^d) \subset L^p(\mathbb{T}^d)$ [104, Section 4.3.2], then yields

$$\begin{aligned}
\|\partial_{x_i}(e^{2B} - P_Je^{2B})\|_p \lesssim\ & e^{2\|B_1\|_\infty}\|B_1\|_{C_1}\|e^{2B_2} - P_Je^{2B_2}\|_{B^0_{p1}} + e^{2\|B_1\|_\infty}\|e^{2B_2} - P_Je^{2B_2}\|_{B^1_{p1}} \\
&+ 2^Je^{2\|B_2\|_\infty}\|B_1\|_\infty + e^{2\|B_1\|_\infty}\|B_1\|_{C_1}\|P_Je^{2B_2}\|_{B^0_{p1}}.
\end{aligned}$$

By Lemma 4.6,

$$\|e^{2B_2}\|_{B^{\alpha+1}_{pp}} \lesssim 1 + \|B_2\|_{B^{\alpha+1}_{pp}} + \|B_2\|^{\alpha+1}_{B^{\alpha+1}_{pp}} \leq c(m, \alpha, p),$$

which implies

$$\begin{aligned}
\|e^{2B_2} - P_J e^{2B_2}\|_{B^1_{p1}} &= \sum_{l>J} 2^{l(1+d/2-d/p)} \Big(\sum_r |\langle e^{2B_2}, \Phi_{lr}\rangle_2|^p \Big)^{1/p} \\
&= \sum_{l>J} 2^{-l\alpha} \Big(2^{pl(\alpha+1+d/2-d/p)} \sum_r |\langle e^{2B_2}, \Phi_{lr}\rangle_2|^p \Big)^{1/p} \\
&\leq \|e^{2B_2}\|_{B^{\alpha+1}_{pp}} \sum_{l>J} 2^{-l\alpha} \\
&\lesssim 2^{-J\alpha}.
\end{aligned}$$

By a similar computation, $\|e^{2B_2} - P_J e^{2B_2}\|_{B^0_{p1}} \lesssim 2^{-J(\alpha+1)}$, while by Hölder's inequality with exponents $(p/(p-1), p)$,

$$\begin{aligned}
\|P_J e^{2B_2}\|_{B^0_{p1}} &= \sum_{l\leq J} 2^{l(d/2-d/p)} \Big(\sum_r |\langle e^{2B_2}, \Phi_{lr}\rangle_2|^p \Big)^{1/p} \\
&\leq \Big(\sum_{l\leq J} [2^{-l(\alpha+1)}]^{\frac{p}{p-1}} \Big)^{\frac{p-1}{p}} \Big(\sum_{l\leq J} 2^{pl(\alpha+1+d/2-d/p)} \sum_r |\langle e^{2B_2}, \Phi_{lr}\rangle_2|^p \Big)^{1/p} \\
&\lesssim \|e^{2B_2}\|_{B^{\alpha+1}_{pp}} \\
&\lesssim 1.
\end{aligned}$$

Combining the above bounds, using the definition of $\varepsilon_T, \bar{\varepsilon}_T$ in Lemma 4.5 and that $2^J \simeq T^{1/(2\alpha+d)}$, we thus obtain that for any $B \in \mathcal{B}_T$,

$$\begin{aligned}
\|\partial_{x_i}(e^{2B} - P_J e^{2B})\|_p &\lesssim e^{2\bar{\varepsilon}_T} \varepsilon_T 2^{-J(\alpha+1)} + e^{2\bar{\varepsilon}_T} 2^{-J\alpha} + 2^J e^{2M} \bar{\varepsilon}_T + e^{2\bar{\varepsilon}_T} \varepsilon_T \\
&\lesssim \varepsilon_T.
\end{aligned}$$

By a similar, in fact easier, computation, we also obtain

$$\begin{aligned}
\|e^{2B} - P_J e^{2B}\|_p &\lesssim e^{2\bar{\varepsilon}_T} 2^{-J(\alpha+1)} + e^{2M} \bar{\varepsilon}_T \\
&= o(\varepsilon_T).
\end{aligned}$$

The required bias bound then follows from the last two displays.

□

Appendix 4.A Additional material and technical results

4.A.1 Some properties of Gaussian priors

Periodic Matérn processes

The Matérn process on $\mathbb{R}^d$ with smoothness parameter $\alpha + 1 - d/2 > 0$ is a stationary Gaussian process with covariance kernel (Example 11.8 in [101])

$$C(x, y) \equiv C(x - y) = \int_{\mathbb{R}^d} e^{-i(x-y).\xi}(1 + \|\xi\|^2)^{-\alpha-1} d\xi, \qquad x, y \in \mathbb{R}^d. \tag{4.31}$$

Using a standard approach to periodization [194], one can construct a corresponding stationary periodic kernel via the Poisson summation formula:

$$C_{\text{per}}(x, y) \equiv C_{\text{per}}(x - y) = \sum_{m \in \mathbb{Z}^d} C(x - y + m), \qquad x, y \in \mathbb{R}^d.$$

Since the Poisson summation formula preserves the Fourier transform (Theorem 8.31 of [95]), C_{per} has Fourier coefficients

$$\int_{\mathbb{T}^d} e^{-2\pi i k.u} C_{\text{per}}(u) du = \int_{\mathbb{R}^d} e^{-2\pi i k.u} C(u) du = \frac{(2\pi)^d}{(1 + 4\pi^2\|k\|^2)^{\alpha+1}}, \qquad k \in \mathbb{Z}^d,$$

the last equality following from (4.31) and the Fourier inversion formula (e.g., Section 8.3 in [95]). Thus, using the Fourier series of $K_{\text{per}}(u)$, the periodised Matérn kernel has series representation

$$C_{\text{per}}(x, y) = (2\pi)^d \sum_{k \in \mathbb{Z}^d} \frac{1}{(1 + 4\pi^2\|k\|^2)^{\alpha+1}} e_k(x)\overline{e_k(y)},$$

where $\{e_k,\ k \in \mathbb{Z}^d\}$ is the Fourier basis of $L^2(\mathbb{T}^d)$, i.e. $e_k(x) = e^{2\pi i k.x}$. Theorem I.21 in [101] then implies that the centred Gaussian process $M = \{M(x),\ x \in \mathbb{T}^d\}$ with covariance kernel C_{per} has RKHS

$$\mathcal{H} = \left\{ h = \sum_{k \in \mathbb{Z}^d} h_k e_k :\ \|h\|_{\mathcal{H}}^2 = (2\pi)^d \sum_{k \in \mathbb{Z}^d} h_k^2 (1 + 4\pi^2\|k\|^2)^{\alpha+1} < \infty \right\}.$$

By the Fourier series characterisation of Sobolev spaces, we thus have $\mathcal{H} = H^{\alpha+1}(\mathbb{T}^d)$ and $\|\cdot\|_{\mathcal{H}}$ is an equivalent norm to $\|\cdot\|_{H^{\alpha+1}}$. The above computation also implies that the periodic Matérn process has Karhunen-Loève series expansion given by (4.10), equivalent (up to constants) to the mean-zero Gaussian processes with inverse Laplacian covariance appearing in [181, 187, 238, 237].

For the regularity of the sample paths, by Proposition I.4 in [101], which applies also to the periodised kernel C_{per}, the periodic Matérn process has a version M whose sample paths are in $C^{\alpha+1-\frac{d}{2}-\eta}(\mathbb{T}^d)$ for any $\eta > 0$. By Lemma I.7 in [101], this version then defines a Borel random element in $C^r(\mathbb{T}^d)$ for all $r < \alpha + 1 - d/2$. In particular, $C^r(\mathbb{T}^d)$ is a separable linear subspace of $C^{(d/2+\kappa)\vee 2}(\mathbb{T}^d)$ for some $\kappa > 0$ provided that $\alpha + 1 > d/2 + (d/2) \vee 2$ and $r > (d/2 + \kappa) \vee 2$. In conclusion, the periodic Matérn process satisfies Condition 4.1 for $\alpha + 1 > d/2 + (d/2) \vee 2$.

Gaussian conjugacy formulae and proof of Lemma 4.1

We first provide details for the derivation of the conjugate formula (4.12). For basis functions $(h_k,\ k \in \mathbb{N}) \subset C^2 \cap \dot{L}^2$, and fixed $K \in \mathbb{N}$, identifying a function $B = \sum_{k=1}^K B_k h_k$ with its coefficient vector $\mathbf{B} = (B_1, \dots, B_K)^T \in \mathbb{R}^K$, we can write the log-likelihood (4.7) in quadratic form as

$$\begin{aligned}
\ell_T(\mathbf{B}) &= -\frac{1}{2} \int_0^T \sum_{j=1}^d \sum_{k,k'} B_k B_{k'} \partial_{x_j} h_k(X_t) \partial_{x_j} h_{k'}(X_t) dt + \sum_{j=1}^d \int_0^T \sum_{k=1}^K B_k \partial_{x_j} h_k(X_t) dX_t^j \\
&= -\frac{1}{2} \sum_{k,k'} B_k B_{k'} \left[\int_0^T \nabla h_k(X_t).\nabla h_{k'}(X_t) dt \right] + \sum_{k=1}^K B_k \left[\int_0^T \nabla h_k(X_t).dX_t \right] \\
&= -\frac{1}{2} \mathbf{B}^T \Sigma \mathbf{B} + \mathbf{B}^T \mathbf{H},
\end{aligned}$$

for Σ and $\mathbf{H}$ defined as after (4.12). Under the same identification, we may regard the prior (4.11) as a multivariate normal distribution on $\mathbb{R}^K$ with diagonal covariance matrix $\Upsilon = \mathrm{diag}(\upsilon_1^2, \dots, \upsilon_K^2)$. Then, by Bayes' formula, the posterior density is given by

$$d\Pi(\mathbf{B}|X^T) \propto e^{\ell_T(\mathbf{B})} e^{-\frac{1}{2}\mathbf{B}^T \Upsilon^{-1} \mathbf{B}} = e^{-\frac{1}{2}\mathbf{B}^T \Sigma \mathbf{B} + \mathbf{B}^T \mathbf{H} - \frac{1}{2}\mathbf{B}^T \Upsilon^{-1} \mathbf{B}}.$$

Completing the squares then gives

$$d\Pi(\mathbf{B}|X^T) \propto e^{-\frac{1}{2}[\mathbf{B} - (\Sigma + \Upsilon^{-1})^{-1}\mathbf{H}]^T (\Sigma + \Upsilon^{-1}) [\mathbf{B} - (\Sigma + \Upsilon^{-1})^{-1}\mathbf{H}]},$$

which completes the derivation of (4.12).

Proof of Lemma 4.1. We follow ideas in [187]. For $K \in \mathbb{N}$, let $P_K : L^2 \to L^2$ be the projection onto the Fourier approximation space $E_K = \mathrm{span}\{e_k,\ |k| \le K\}$, $e_k(x) = e^{2\pi i k.x}$.

Consider the approximate posterior

$$\Pi_K(A|X^T) = \frac{\int_A e^{\ell_T(P_K B)} d\Pi(B)}{\int_{C^2} e^{\ell_T(P_K B')} d\Pi(B')}, \qquad A \subset C^2 \cap \dot{L}^2 \text{ measurable.}$$

Since $\ell_T(P_K B)$ only depends on $P_K B$, $\Pi_K(\cdot|X^T)$ can be decomposed as the product of a Gaussian measure on E_K (satisfying the conjugate formula (4.12)) and a Gaussian measure obtained as the push-forward of Π under the projection operator onto the orthogonal complement of E_K in L^2. It follows that $\Pi_K(\cdot|X^T)$ is Gaussian in L^2. We now show that, a.s. under the law of the data X^T, $\Pi_K(\cdot|X^T) \to \Pi(\cdot|X^T)$ weakly in L^2 as $K \to \infty$, which in turn will conclude the proof since the class of Gaussian measures in closed with respect to weak convergence.

For $B \in C^2$, we have $\|B - P_K B\|_{H^2} \to 0$ as $K \to \infty$. It follows that $e^{\ell_T(P_K B)} \to e^{\ell_T(B)}$ a.s., since, given the data X^T, the function $B \mapsto \ell_T(B)$ in (4.7) is continuous with respect to the H^2-norm. We next show that the limiting function $e^{\ell_T(B)}$, $B \in C^2$, is dominated by a Π-integrable function. Under $P^T_{B_0}$, using Itô's lemma (Theorem 39.3 of [19]),

$$\begin{aligned}
\ell_T(B) &\leq \int_0^T \nabla B(X_t).dX_t \\
&= B(X_T) - B(X_0) - \frac{1}{2}\int_0^T \Delta B(X_t)dt \\
&\leq 2\|B\|_\infty + \frac{1}{2}\sum_{j=1}^d \int_0^T \left|\frac{\partial^2}{\partial x_j^2} B(X_t)\right| dt,
\end{aligned}$$

and using Young's inequality with ε and $p = q = 2$, we obtain that for all $\eta > 0$,

$$\begin{aligned}
\ell_T(B) &\leq \frac{2}{\eta} + \frac{\eta}{2}\|B\|_\infty^2 + \frac{1}{2}\sum_{j=1}^d \int_0^T \left[\frac{1}{2\eta} + \frac{\eta}{2}\left|\frac{\partial^2}{\partial x_j^2} B(X_t)\right|^2\right] dt \\
&\leq \frac{2}{\eta} + \frac{\eta\|B\|_\infty^2}{2} + \frac{dT}{4\eta} + \frac{dT\eta}{4}\|B\|_{C^2}^2 = \frac{dT+8}{4\eta} + \frac{\eta(dT+2)}{4}\|B\|_{C^2}^2.
\end{aligned}$$

In conclusion, for all $\eta > 0$,

$$e^{\ell_T(B)} \leq e^{(dT+8)/(4\eta)} e^{\eta(dT+2)\|B\|_{C^2}^2/4}.$$

Taking $\eta > 0$ small enough, since Π can be regarded as a Gaussian measure on C^2, Fernique's theorem [38, Theorem 2.8.5] implies that the right hand side in the last display is Π-integrable. By the dominated convergence theorem, we then conclude that a.s. under the law of the data X^T, as $K \to \infty$,

$$\int_{C^2} e^{\ell_T(P_K B')} d\Pi(B') \to \int_{C^2} e^{\ell_T(B')} d\Pi(B'),$$

and likewise, for all measurable $A \subset C^2 \cap \dot{L}^2$, by boundedness of the indicator function $1_A(\cdot)$,

$$\int_A e^{\ell_T(P_K B)} d\Pi(B) = \int_{C^2} e^{\ell_T(P_K B)} 1_A(B) d\Pi(B) \to \int_A e^{\ell_T(B)} d\Pi(B),$$

showing as required that $\Pi_K(\cdot|X^T) \to \Pi(\cdot|X^T)$ weakly in C^2 (and also in L^2).

□

4.A.2 Proof of Theorem 4.4

The proof follows by the standard arguments for test-based contraction rates, see, e.g., Theorem 2.1 of [99], with the only difference being how we control the denominator in the Bayes formula. We now detail this argument, which is a modification of Lemma 4.2 of van der Meulen et al. [232] adapted to our setting. By (4.7), the log-likelihood ratio equals (replacing ∇B with b), under $P_{b_0}^T$,

$$\log \frac{dP_b^T}{dP_{b_0}^T}(X^T) = \int_0^T [b(X_s) - b_0(X_s)].dW_s - \tfrac{1}{2} \int_0^T \|b(X_s) - b_0(X_s)\|^2 ds =: M_T^b - \tfrac{1}{2}[M^b]_T.$$

Set $\overline{\Pi} = \Pi(\cdot \cap \mathcal{SB}_T)/\Pi(\mathcal{SB}_T)$ to be the normalised prior restricted to the set $\mathcal{SB}_T$. Since $P_{b_0}^T(\sup_{b \in \mathcal{SB}_T}[M^b]_T \le T\varepsilon_T^2) \to 1$ by assumption, we henceforth work on this event. By Jensen's inequality,

$$\log \int \frac{dP_b^T}{dP_{b_0}^T}(X_T) \frac{d\Pi(b)}{\Pi(\mathcal{SB}_T)} \ge \int_{\mathcal{SB}_T} \log \frac{dP_b^T}{dP_{b_0}^T}(X_T) d\overline{\Pi}(b) = \int_{\mathcal{SB}_T} M_T^b - \tfrac{1}{2}[M^b]_T d\overline{\Pi}(b).$$

For fixed $T > 0$, define

$$Z_t^T = \int_{\mathcal{SB}_T} M_t^b d\overline{\Pi}(b) = \sum_{i=1}^d \int_0^t \int_{\mathcal{SB}_T} b_i(X_s) - b_{0,i}(X_s) d\overline{\Pi}(b) dW_s^i,$$

where the last equality follows from stochastic Fubini's theorem (Theorem 64 of Chapter IV in [190]) since $\mathcal{SB}_T$ is deterministic. The process $(Z_t^T : t \ge 0)$ is a continuous local martingale under $P_{b_0}^T$ with quadratic variation

$$[Z^T]_t = \int_0^t \sum_{i=1}^d \left(\int_{\mathcal{SB}_T} b_i(X_s) - b_{0,i}(X_s) d\overline{\Pi}(b) \right)^2 ds.$$

Using Jensen's inequality, Fubini's theorem and the definition of $\mathcal{SB}_T$,

$$[Z^T]_T \leq \int_0^T \sum_{i=1}^d \int_{\mathcal{SB}_T} (b_i(X_s) - b_{0,i}(X_s))^2 d\overline{\Pi}(b)\, ds$$
$$= \int_{\mathcal{SB}_T} \int_0^T \|b(X_s) - b_0(X_s)\|^2 ds\, d\overline{\Pi}(b) = \int_{\mathcal{SB}_T} [M^b]_T d\overline{\Pi}(b) \leq T\varepsilon_T^2.$$

By Bernstein's inequality (4.19), for any $x > 0$,

$$P_{b_0}^T \left(|Z_T^T| \geq x \right) = P_{b_0}^T \left(|Z_T^T| \geq x, [Z^T]_T \leq T\varepsilon_T^2 \right) \leq 2 \exp\left(-\frac{x^2}{2T\varepsilon_T^2} \right).$$

Setting $x = LT\varepsilon_T^2$ gives $P_{b_0}^T(|Z_T^T| \geq LT\varepsilon_T^2) \to 0$ for any $L > 0$. On the event $\{|Z_T^T| \leq LT\varepsilon_T^2\}$, which has $P_{b_0}^T$-probability tending to one, and using the second to last display,

$$\int_{\mathcal{SB}_T} M_T^b - \tfrac{1}{2}[M^b]_T d\overline{\Pi}(b) = Z_T^T - \tfrac{1}{2} \int_{\mathcal{SB}_T} [M^b]_T d\overline{\Pi}(b) \geq -(L + 1/2)T\varepsilon_T^2.$$

In conclusion, we have shown $P_{b_0}^T(\int p_b^T/p_{b_0}^T d\Pi(b) \geq e^{-(L+1/2)T\varepsilon_T^2}\Pi(\mathcal{SB}_T)) \to 1$ for any $L > 0$. Substituting in $\Pi(\mathcal{SB}_T) \geq e^{-CT\varepsilon_T^2}$ gives $P_{b_0}^T(\int p_b/p_{b_0} d\Pi(b) \geq e^{-(C+L+1/2)T\varepsilon_T^2}) \to 1$ for any $L > 0$.

□

4.A.3 A PDE estimate

In the proofs we used a stability estimate for solutions to the Poisson equation $L_b u = f$, where the generator L_b of the diffusion is the strongly elliptic second order partial differential operator given in (4.5). For some basic facts about this elliptic PDE, see Section 6 in [173], while for more general theory for periodic elliptic PDEs, see Chapter II.3 in [24]. The following stability estimate is only proved for $t \geq 2$ in Lemma 11 of [173], but the proof can be extended to general $t \in \mathbb{R}$.

Lemma 4.8. *Let $t \in \mathbb{R}$ and assume $b \in C^{|t-2|}(\mathbb{T}^d)$. For any $f \in L^2_{\mu_b}(\mathbb{T}^d)$, there exists a unique solution $L_b^{-1}[f] \in L^2(\mathbb{T}^d)$ of the equation*

$$L_b u = f, \qquad f \in L^2(\mathbb{T}^d),$$

satisfying $L_b L_b^{-1}[f] = f$ almost everywhere and $\int_{\mathbb{T}^d} L_b^{-1}[f](x)dx = 0$. Moreover,

$$\|L_b^{-1}[f]\|_{H^t} \lesssim \|f\|_{H^{t-2}},$$

with constants depending on t, d and on an upper bound for $\|b\|_{B^{|t-2|}_{\infty\infty}}$.

Proof. Recall the multiplication inequality for Besov-Sobolev norms with $t \geq 0$ ([229], p. 143),

$$\|fg\|_{B^t_{pq}} \leq c(t,p,q,d)\|f\|_{B^t_{pq}}\|g\|_{B^t_{\infty\infty}} \leq c'(t,p,q,d)\|f\|_{B^t_{pq}}\|g\|_{C^t},$$

which by duality implies that for $t \geq 0$,

$$\|fg\|_{B^{-t}_{pq}} = \sup_{\|\varphi\|_{B^t_{p'q'}} \leq 1} \left| \int_{\mathbb{T}^d} f \frac{g\varphi}{\|g\varphi\|_{B^t_{p'q'}}} dx \right| \|g\varphi\|_{B^t_{p'q'}} \leq c(t,p,q,d)\|f\|_{B^{-t}_{pq}}\|g\|_{B^t_{\infty\infty}},$$

where $1/p + 1/p' = 1/q + 1/q' = 1$. Combining the last displays yields the multiplier inequality

$$\|fg\|_{B^t_{pq}} \leq c(|t|,p,q,d)\|f\|_{B^t_{pq}}\|g\|_{B^{|t|}_{\infty\infty}}, \qquad t \in \mathbb{R}. \tag{4.32}$$

The proof of the result then follows as in Lemma 11 of [173] upon using the multiplier inequality (4.32) instead of (3) when establishing (69) in that paper. We omit the details here.

□

4.A.4 The Gaussian correlation inequality

The Gaussian correlation inequality states that for any closed, convex and symmetric around the origin sets K, L in $\mathbb{R}^d$ and any centred Gaussian measure μ in $\mathbb{R}^d$, we have

$$\mu(K \cap L) \geq \mu(K)\mu(L).$$

The inequality was proved by Royen [206]; see also [148] for a self-contained presentation and proof. In the proof of Lemma 4.5 we have used the following extension to Gaussian measures on separable Banach spaces, which was referred to, without full details, in Remark 3 (i) in [148]. We include a proof for completeness.

Lemma 4.9. *Let μ be a centred Borel Gaussian measure on a separable Banach space $\mathbb{B}$. Let K, L be closed (with respect to the topology induced by $\|\cdot\|_{\mathbb{B}}$), convex and symmetric around the origin subsets of $\mathbb{B}$. Then,*

$$\mu(K \cap L) \geq \mu(K)\mu(L).$$

Proof. For $X \sim \mu$, the Karhunen-Loève expansion ([104], Theorem 2.6.10) of X implies that there exists a complete orthonormal system $\{x_n,\ n \geq 1\}$ of the RKHS of X and i.i.d. standard normal random variables $(\xi_n)_{n\geq 1}$, such that $X_n = \sum_{i=1}^n \xi_i x_i$ converges almost surely to X in the norm of $\mathbb{B}$. We first show that for any closed, convex and symmetric around the origin set $K \subseteq \mathbb{B}$,

$$\mu(K) = \lim_{n\to\infty} \Pr\left(X_n \in K\right). \tag{4.33}$$

First, since K is closed and $X_n \to X$ almost surely, and hence also in law, the Portmanteau lemma ([234], Lemma 2.2) implies

$$\limsup_{n\to\infty} \Pr(X_n \in K) \leq \Pr(X \in K).$$

On the other hand, since K is convex and symmetric around the origin, by independence of the random variables $(\xi_n)_{n\geq 1}$ and Anderson's inequality [104, Theorem 2.4.5], denoting by ν_n the law on $\mathbb{B}$ of the tail of the series $\sum_{i=n+1}^\infty \xi_i x_i$,

$$\begin{aligned}\Pr(X \in K) &= \Pr\left(X_n + \sum_{i=n+1}^\infty \xi_i x_i \in K\right)\\ &= \int_{\mathbb{B}} \Pr\left(X_n + x \in K\right) \nu_n(dx)\\ &\leq \Pr(X_n \in K).\end{aligned}$$

The two above inequalities thus jointly establish (4.33). Now let $K, L \subseteq \mathbb{B}$ be arbitrary closed, convex and symmetric around the origin sets and define

$$K_n = \left\{ z = (z_1, \dots, z_n) \in \mathbb{R}^n : \sum_{i=1}^n z_i x_i \in K \right\} \subseteq \mathbb{R}^n$$

and analogously L_n. It is straightforward to check that K_n is a convex and symmetric subset of $\mathbb{R}^n$. Furthermore, if $\{z^{(k)}\}_{k\geq 1} \subset K_n$ converges to some $z = (z_1, \dots, z_n) \in \mathbb{R}^n$, then

$$\left\| \sum_{i=1}^n z_i^{(k)} x_i - \sum_{i=1}^n z_i x_i \right\|_{\mathbb{B}} \leq \max_{i=1,\dots,n} \|x_i\|_{\mathbb{B}} \sum_{i=1}^n |z_i^{(k)} - z_i| \to 0, \qquad k \to \infty,$$

so that, K being closed in $\mathbb{B}$, $\sum_{i=1}^n z_i x_i \in K$ and $z \in K_n$. This shows that K_n is closed, convex and symmetric around the origin in $\mathbb{R}^n$. Denoting by $\gamma_n = \mathcal{L}(\xi_1, \dots, \xi_n)$ the standard Gaussian measure on $\mathbb{R}^n$, (4.33) then implies $\mu(K) = \lim_{n\to\infty} \gamma_n(K_n)$. Note that identical considerations apply to L and L_n and, since $K \cap L$ is also closed, convex and symmetric around the origin in $\mathbb{B}$, also to $K \cap L$ and $K_n \cap L_n$.

Applying this and the Gaussian correlation inequality for the finite-dimensional Gaussian measure γ_n,

$$\mu(K \cap L) = \lim_{n\to\infty} \gamma_n(K_n \cap L_n) \geq \liminf_{n\to\infty} \gamma_n(K_n)\gamma_n(L_n) = \mu(K)\mu(L).$$

□

References

[1] Abraham, K. (2019). Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. *Bernoulli*, 25(4A):2696–2728.

[2] Abraham, K. and Nickl, R. (2019). On statistical Calderón problems. *Math. Stat. Learn.*, 2(2):165–216.

[3] Abramovich, F. and Silverman, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85(1):115–129.

[4] Agapiou, S., Burger, M., Dashti, M., and Helin, T. (2018). Sparsity-promoting and edge-preserving maximum *a posteriori* estimators in non-parametric Bayesian inverse problems. *Inverse Problems*, 34(4):045002, 37.

[5] Agapiou, S., Dashti, M., and Helin, T. (2021). Rates of contraction of posterior distributions based on p-exponential priors. *Bernoulli*, 27(3):1616 – 1642.

[6] Agapiou, S., Larsson, S., and Stuart, A. M. (2013). Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Process. Appl.*, 123(10):3828–3860.

[7] Agapiou, S., Stuart, A. M., and Zhang, Y.-X. (2014). Bayesian posterior contraction rates for linear severely ill-posed inverse problems. *J. Inverse Ill-Posed Probl.*, 22(3):297–321.

[8] Alessandrini, G. (1986). An identification problem for an elliptic equation in two variables. *Annali di Matematica Pura ed Applicata*, 145(1):265–295.

[9] Apte, A., Hairer, M., Stuart, A. M., and Voss, J. (2007). Sampling the posterior: an approach to non-Gaussian data assimilation. *Phys. D*, 230(1-2):50–64.

[10] Arridge, S., Maass, P., Öktem, O., and Schönlieb, C.-B. (2019). Solving inverse problems using data-driven models. *Acta Numer.*, 28:1–174.

[11] Arridge, S. R. (1999). Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41–R93.

[12] Atangana, A. (2018). *Fractional Operators with Constant and Variable Order with Application to Geo-Hydrology*. Academic Press.

[13] Aurzada, F. (2007). On the lower tail probabilities of some random sequences in l_p. *J. Theoret. Probab.*, 20(4):843–858.

[14] Babii, A. (2020). Honest confidence sets in nonparametric IV regression and other ill-posed models. *Econometric Theory*, 36(4):658–706.

[15] Bakry, D., Gentil, I., and Ledoux, M. (2014). *Analysis and geometry of Markov diffusion operators*, volume 348. Springer, Cham.

[16] Bal, G. (2009). Inverse transport theory and applications. *Inverse Problems*, 25(5):053001, 48.

[17] Bal, G. and Uhlmann, G. (2010). Inverse diffusion theory of photoacoustics. *Inverse Problems*, 26(8):085010, 20.

[18] Bao, G. and Li, P. (2005). Inverse medium scattering problems for electromagnetic waves. *SIAM J. Appl. Math.*, 65(6):2049–2066.

[19] Bass, R. F. (2011). *Stochastic processes*, volume 33 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

[20] Batz, P., Ruttor, A., and Opper, M. (2018). Approximate Bayes learning of stochastic differential equations. *Phys. Rev. E*, 98:022109.

[21] Bauer, F., Hohage, T., and Munk, A. (2009). Iteratively regularized Gauss-Newton method for nonlinear inverse problems with random noise. *SIAM J. Numer. Anal.*, 47(3):1827–1846.

[22] Belloni, A. and Chernozhukov, V. (2009). On the computational complexity of MCMC-based estimators in large samples. *Ann. Statist.*, 37(4):2011–2055.

[23] Benning, M. and Burger, M. (2018). Modern regularization methods for inverse problems. *Acta Numerica*, 27:1–111.

[24] Bers, L., John, F., and Schechter, M. (1964). *Partial differential equations*. Lectures in Applied Mathematics, Vol. III. Wiley & Sons, Inc. New York-London-Sydney.

[25] Bertero, M. and Boccacci, P. (1998). *Introduction to inverse problems in imaging*. Institute of Physics Publishing, Bristol.

[26] Beskos, A., Girolami, M., Lan, S., Farrell, P., and Stuart, A. (2017). Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, 335:327–351.

[27] Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(3):333–382. With discussions and a reply by the authors.

[28] Beskos, A., Roberts, G., and Stuart, A. (2009). Optimal scalings for local Metropolis-Hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.*, 19(3):863–898.

[29] Beskos, A. and Stuart, A. (2009). MCMC methods for sampling function space. In *ICIAM 07—6th International Congress on Industrial and Applied Mathematics*, pages 337–364. Eur. Math. Soc., Zürich.

[30] Birgé, L. (2004). Model selection for Gaussian regression with random design. *Bernoulli*, 10:1039–1051.

[31] Bissantz, N. and Birke, M. (2009). Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators. *J. Multivariate Anal.*, 100(10):2364–2375.

[32] Bissantz, N., Dümbgen, L., Holzmann, H., and Munk, A. (2007a). Non-parametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(3):483–506.

[33] Bissantz, N., Hohage, T., and Munk, A. (2004). Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise. *Inverse Problems*, 20(6):1773–1789.

[34] Bissantz, N., Hohage, T., Munk, A., and Ruymgaart, F. (2007b). Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM Journal on Numerical Analysis*, 45(6):2610–2636.

[35] Bissantz, N. and Holzmann, H. (2008). Statistical inference for inverse problems. *Inverse Problems*, 24(3):034009, 17.

[36] Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *J. Polit. Econ.*, 81(3):637–654.

[37] Blanchard, G., Hoffmann, M., and Reiß, M. (2018). Optimal adaptation for early stopping in statistical inverse problems. *SIAM/ASA J. Uncertain. Quantif.*, 6(3):1043–1075.

[38] Bogachev, V. (1998). *Gaussian Measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society.

[39] Bohr, J. (2021). Stability of the non-abelian x-ray transform in dimension $d \geq 3$. *arXiv preprint arXiv:2008.07951*.

[40] Bohr, J. and Nickl, R. (2021). On log-concave approximations of high-dimensional posterior measures and stability properties in non-linear inverse problems. *arXiv preprint arXiv:2105.07835*.

[41] Bolhuis, P. G., Chandler, D., Dellago, C., and Geissler, P. L. (2002). Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical Chemistry*, 53(1):291–318. PMID: 11972010.

[42] Bonito, A., Cohen, A., DeVore, R., Petrova, G., and Welper, G. (2017). Diffusion coefficients estimation for elliptic partial differential equations. *SIAM J. Math. Anal.*, 49(2):1570–1592.

[43] Bou-Rabee, N., Eberle, A., and Zimmer, R. (2020). Coupling and convergence for Hamiltonian Monte Carlo. *Ann. Appl. Probab.*, 30(3):1209–1250.

[44] Bou-Rabee, N. and Hairer, M. (2013). Nonasymptotic mixing of the MALA algorithm. *IMA J. Numer. Anal.*, 33(1):80–110.

[45] Bourdaud, G. and Sickel, W. (2011). Composition operators on function spaces with fractional order of smoothness. In *Harmonic analysis and nonlinear partial differential equations*, RIMS Kôkyûroku Bessatsu, B26, pages 93–132. Res. Inst. Math. Sci. (RIMS), Kyoto.

[46] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

[47] Briol, F.-X., Oates, C. J., Girolami, M., Osborne, M. A., and Sejdinovic, D. (2019). Probabilistic integration: a role in statistical computation? *Statist. Sci.*, 34(1):1–22.

[48] Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24(6):2384–2398.

[49] Bui-Thanh, T. and Ghattas, O. (2015). A scalable algorithm for MAP estimators in Bayesian inverse problems with Besov priors. *Inverse Probl. Imaging*, 9(1):27–53.

[50] Burger, M. and Helin, T. (2015). Maximum *a posteriori* probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8):085009, 22.

[51] Burger, M. and Osher, S. (2004). Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421.

[52] Castillo, I. and Nickl, R. (2013). Nonparametric Bernstein–von Mises Theorems in Gaussian white noise. *Ann. Statist.*, 41(4):1999–2028.

[53] Castillo, I. and Nickl, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.*, 42(5):1941–1969.

[54] Castillo, I. and Rousseau, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.*, 43(6):2353–2383.

[55] Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems*, (24).

[56] Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields*, 123(3):323–354.

[57] Chambolle, A. and Pock, T. (2016). An introduction to continuous optimization for imaging. *Acta Numer.*, 25:161–319.

[58] Chandrasekhar, S. (1943). Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15:1–89.

[59] Chen, Y., Dwivedi, R., Wainwright, M. J., and Yu, B. (2020). Fast mixing of metropolized Hamiltonian Monte Carlo: benefits of multi-step gradients. *J. Mach. Learn. Res.*, 21:Paper No. 92, 71.

[60] Cheng, X. and Bartlett, P. (2018). Convergence of Langevin MCMC in KL-divergence. In *Algorithmic learning theory 2018*, volume 83 of *Proc. Mach. Learn. Res. (PMLR)*, page 26. Proceedings of Machine Learning Research PMLR, [place of publication not identified].

[61] Churchill, R. V. and Brown, J. W. (1978). *Fourier series and boundary value problems*. McGraw-Hill Book Co., New York, third edition.

[62] Conrad, P. R., Marzouk, Y. M., Pillai, N. S., and Smith, A. (2016). Accelerating asymptotically exact MCMC for computationally intensive models via local approximations. *Journal of the American Statistical Association*, 111(516):1591–1607.

[63] Cotter, S., Roberts, G., Stuart, A., and White, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446.

[64] Cotter, S. L., Dashti, M., Robinson, J. C., and Stuart, A. M. (2009). Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, 25(11):115008, 43.

[65] Cox, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.*, 21(2):903–923.

[66] Cui, T., Law, K. J. H., and Marzouk, Y. M. (2016). Dimension-independent likelihood-informed MCMC. *J. Comput. Phys.*, 304:109–137.

[67] Da Prato, G. and Zabczyk, J. (2014). *Stochastic equations in infinite dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition.

[68] Dalalyan, A. and Reiß, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probab. Theory Related Fields*, 137(1-2):25–47.

[69] Dalalyan, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676.

[70] Dashti, M., Harris, S., and Stuart, A. M. (2012). Besov priors for Bayesian inverse problems. *Inverse Probl. Imaging*, 6(2):183–200.

[71] Dashti, M., Law, K. J. H., Stuart, A. M., and Voss, J. (2013). MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 27.

[72] Dashti, M. and Stuart, A. M. (2011). Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM J. Numer. Anal.*, 49(6):2524–2542.

[73] Dashti, M. and Stuart, A. M. (2017). The Bayesian approach to inverse problems. In *Handbook of uncertainty quantification. Vol. 1, 2, 3*, pages 311–428. Springer, Cham.

[74] Delaigle, A., Hall, P., and Jamshidi, F. (2015). Confidence bands in non-parametric errors-in-variables regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 77(1):149–169.

[75] Diaconis, P. (1988). Bayesian Numerical Analysis. In Berger, J. and Gupta, S., editors, *Statistical Decision Theory and Related Topics IV*, pages 163–175. Springer, New York.

[76] Dirksen, S. (2015). Tail bounds via generic chaining. *Electron. J. Probab.*, 20:no. 53, 29.

[77] Doi, M., Edwards, S., and Edwards, S. (1988). *The Theory of Polymer Dynamics*. International series of monographs on physics. Clarendon Press.

[78] Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2(2):101–126.

[79] Dudley, R. M. (2002). *Real analysis and probability*. Cambridge University Press, Cambridge.

[80] Dudley, R. M. (2014). *Uniform central limit theorems*. Cambridge University Press, New York, second edition.

[81] Durmus, A. and Moulines, E. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587.

[82] Durmus, A. and Moulines, E. (2019). High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882.

[83] Durmus, A., Moulines, E., and Pereyra, M. (2018). Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau. *SIAM J. Imaging Sci.*, 11(1):473–506.

[84] Dwivedi, R., Chen, Y., Wainwright, M. J., and Yu, B. (2019). Log-concave sampling: Metropolis-Hastings algorithms are fast. *J. Mach. Learn. Res.*, 20:Paper No. 183, 42.

[85] Eberle, A. (2014). Error bounds for Metropolis-Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. *Ann. Appl. Probab.*, 24(1):337–377.

[86] Eberle, A. (2015). Reflection couplings and contraction rates for diffusions. *Probability Theory and Related Fields*, 166:851–886.

[87] Efromovich, S. (1997). Robust and efficient recovery of a signal passed through a filter and then contaminated by non-Gaussian noise. *IEEE Trans. Inform. Theory*, 43(4):1184–1191.

[88] Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht.

[89] Evans, L. C. (2010). *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition.

[90] Evans, S. N. and Stark, P. B. (2002). Inverse problems as statistics. *Inverse Problems*, 18(4):R55–R97.

[91] Falk, R. S. (1983). Error estimates for the numerical identification of a variable coefficient. *Mathematics of Computation*, 40(162):537–546.

[92] Fitzpatrick, B. G. (1991). Bayesian analysis in inverse problems. *Inverse Problems*, 7(5):675–702.

[93] Florens, J.-P. and Simoni, A. (2016). Regularizing priors for linear inverse problems. *Econometric Theory*, 32(1):71–121.

[94] Franklin, J. N. (1970). Well-posed stochastic extensions of ill-posed linear problems. *J. Math. Anal. Appl.*, 31:682–716.

[95] Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Ann. Statist.*, 27(4):1119–1140.

[96] Fuchs, C. (2013). *Inference for diffusion processes.* Springer, Heidelberg. With applications in life sciences, With a foreword by Ludwig Fahrmeir.

[97] García-Portugués, E., Sørensen, M., Mardia, K. V., and Hamelryck, T. (2019). Langevin diffusions on the torus: estimation and applications. *Stat. Comput.*, 29(1):1–22.

[98] Garralda, N., Llatser, I., Cabellos-Aparicio, A., Alarcón, E., and Pierobon, M. (2011). Diffusion-based physical channel identification in molecular nanonetworks. *Nano Communication Networks*, 2(4):196 – 204.

[99] Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531.

[100] Ghosal, S. and van der Vaart, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.*, 35(1):192–223.

[101] Ghosal, S. and van der Vaart, A. W. (2017). *Fundamentals of Nonparametric Bayesian Inference.* Cambridge University Press, New York.

[102] Gilbarg, D. and Trudinger, N. S. (1998). *Elliptic partial differential equations of second order.* Springer-Verlag, Berlin-New York.

[103] Giné, E. and Nickl, R. (2011). Rates of contraction for posterior distributions in L^r-metrics, $1 \le r \le \infty$. *Ann. Statist.*, 39(6):2883–2911.

[104] Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models.* Cambridge University Press, New York.

[105] Giordano, M. and Kekkonen, H. (2020). Bernstein–von Mises theorems and uncertainty quantification for linear inverse problems. *SIAM/ASA J. Uncertain. Quantif.*, 8(1):342–373.

[106] Giordano, M. and Nickl, R. (2020). Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Problems*, 36(8):085001–85036.

[107] Giordano, M. and Ray, K. (2020). Nonparametric bayesian inference for reversible multi-dimensional diffusions. *arXiv preprint arXiv:2012.12083.*

[108] Gobet, E., Hoffmann, M., and Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.*, 32(5):2223–2253.

[109] Green, P. J., Latuszyński, K., Pereyra, M., and Robert, C. P. (2015). Bayesian computation: a summary of the current state, and samples backwards and forwards. *Stat. Comput.*, 25(4):835–862.

[110] Gugushvili, S. and Spreij, P. (2014). Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. *Lith. Math. J.*, 54(2):127–141.

[111] Gugushvili, S., van der Vaart, A., and Yan, D. (2020). Bayesian linear inverse problems in regularity scales. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):2081 – 2107.

[112] Hadamard, J. (1923). *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Yale University Press.

[113] Hairer, M., Stuart, A., and Vollmer, S. (2014). Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *The Annals of Applied Probability*, 24(6):2455–2490.

[114] Hanke, M., Neubauer, A., and Scherzer, O. (1995). A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.*, 72(1):21–37.

[115] Helin, T. and Burger, M. (2015). Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8):085009, 22.

[116] Hoffmann, K. H. and Sprekels, J. (1985). On the identification of coefficients of elliptic problems by asymptotic regularization. *Numerical Functional Analysis and Optimization*, 7(2-3):157–177.

[117] Hohage, T. and Pricop, M. (2008). Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise. *Inverse Probl. Imaging*, 2(2):271–290.

[118] Hohage, T. and Werner, F. (2013). Iteratively regularized Newton-type methods for general data misfit functionals and applications to Poisson data. *Numer. Math.*, 123(4):745–779.

[119] Hörmander, L. (1994). *The analysis of linear partial differential operators. III*, volume 274 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin. Pseudo-differential operators, Corrected reprint of the 1985 original.

[120] Hosseini, B. (2019). Two Metropolis-Hastings algorithms for posterior measures with non-Gaussian priors in infinite dimensions. *SIAM/ASA J. Uncertain. Quantif.*, 7(4):1185–1223.

[121] Hull, J. (2009). *Options, Futures and Other Derivatives*. Prentice Hall finance series. Pearson/Prentice Hall.

[122] Isakov, V. (2017). *Inverse problems for partial differential equations*, volume 127 of *Applied Mathematical Sciences*. Springer, Cham, third edition.

[123] Ito, K. and Kunisch, K. (1994). On the injectivity and linearization of the coefficient-to-solution mapping for elliptic boundary value problems. *Journal of Mathematical Analysis and Applications*, 188(3):1040–1066.

[124] Jia, J., Peng, J., and Gao, J. (2016). Bayesian approach to inverse problems for functions with a variable-index Besov prior. *Inverse Problems*, 32(8):085006, 32.

[125] Johnstone, I. M., Kerkyacharian, G., Picard, D., and Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(3):547–573.

[126] Johnstone, I. M. and Silverman, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.*, 18(1):251–280.

[127] Kaipio, J. and Somersalo, E. (2004). *Statistical and Computational Inverse Problems*. Number 160 in Applied Mathematical Sciences. Springer-Verlag New York.

[128] Kaltenbacher, B., Neubauer, A., and Scherzer, O. (2008). *Iterative regularization methods for nonlinear ill-posed problems*, volume 6 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin.

[129] Kaltenbacher, B., Schöpfer, F., and Schuster, T. (2009). Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems. *Inverse Problems*, 25(6):065003, 19.

[130] Kamynin, V. L. and Bukharova, T. I. (2019). On inverse problem of determination of the coefficient in the Black-Scholes type equation. In *Finite difference methods*, volume 11386 of *Lecture Notes in Comput. Sci.*, pages 313–320. Springer, Cham.

[131] Kekkonen, H. (2021). Consistency of bayesian inference with Gaussian process priors for a parabolic inverse problem. *arXiv preprint arXiv:2103.13213*.

[132] Kekkonen, H., Lassas, M., and Siltanen, S. (2014). Analysis of regularized inversion of data corrupted by white Gaussian noise. *Inverse Problems*, 30(4):045009, 18.

[133] Kekkonen, H., Lassas, M., and Siltanen, S. (2016). Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators. *Inverse Problems*, 32(8):085005, 31.

[134] Kim, P. T. and Koo, J.-Y. (2002). Optimal spherical deconvolution. *J. Multivariate Anal.*, 80(1):21–42.

[135] Kirsch, A. (2011). *An introduction to the mathematical theory of inverse problems*, volume 120 of *Applied Mathematical Sciences*. Springer, New York, second edition.

[136] Knapik, B. and Salomond, J.-B. (2018). A general approach to posterior contraction in nonparametric inverse problems. *Bernoulli*, 24(3):2091–2121.

[137] Knapik, B., Szabò, B., van der Vaart, A. W., and van Zanten, H. (2015). Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probab. Theory Relat. Fields*, (164):771–813.

[138] Knapik, B., van der Vaart, A. W., and van Zanten, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.*, 39(5):2626–2657.

[139] Knapik, B. T., van der Vaart, A. W., and van Zanten, J. H. (2013). Bayesian recovery of the initial condition for the heat equation. *Comm. Statist. Theory Methods*, 42(7):1294–1313.

[140] Knowles, I. (2001). Parameter identification for elliptic problems. *Journal of Computational and Applied Mathematics*, 131(1):175 – 194.

[141] Kohn, R. V. and Lowe, B. D. (1988). A variational method for parameter identification. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 22(1):119–158.

[142] Kolehmainen, V., Lassas, M., Niinimäki, K., and Siltanen, S. (2012). Sparsity-promoting Bayesian inversion. *Inverse Problems*, 28(2):025005, 28.

[143] Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304.

[144] Kravaris, C. and Seinfeld, J. (1985). Identification of parameters in distributed parameter systems by regularization. *SIAM Journal on Control and Optimization*, 23.

[145] Kutoyants, Y. A. (2004). *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London, Ltd., London.

[146] Lasanen, S. (2012). Non-Gaussian statistical inverse problems. Part I: Posterior distributions. *Inverse Probl. Imaging*, 6(2):215–266.

[147] Lassas, M., Saksman, E., and Siltanen, S. (2009). Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl. Imaging*, 3(1):87–122.

[148] Latala, R. and Matlak, D. (2017). Royen's proof of the Gaussian correlation inequality. In *Geometric aspects of functional analysis*, volume 2169 of *Lecture Notes in Math.*, pages 265–275. Springer, Cham.

[149] Leahu, H. (2011). On the Bernstein-von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.*, 5:373–404.

[150] Lehtinen, M. S., Päivärinta, L., and Somersalo, E. (1989). Linear inverse problems for generalised random variables. *Inverse Problems*, 5(4):599–612.

[151] Leporini, D. and Pesquet, J.-C. (2001). Bayesian wavelet denoising: Besov priors and non-Gaussian noises. *Signal Processing*, 81(1):55–67. Special section on Markov Chain Monte Carlo (MCMC) Methods for Signal Processing.

[152] Li, W. V. and Linde, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.*, 27(3):1556–1578.

[153] Lions, J. L. and Magenes, E. (1972). *Non-Homogeneous Boundary Value Problems and Applications*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag Berlin Heidelberg, 1 edition.

[154] Livingstone, S., Betancourt, M., Byrne, S., and Girolami, M. (2019). On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli*, 25(4A):3109–3138.

[155] Loubes, J.-M. and Ludeña, C. (2010). Penalized estimators for non linear inverse problems. *ESAIM: Probability and Statistics*, 14:173–191.

[156] Lounici, K. and Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.*, 39(1):201–231.

[157] Lu, S. and Pereverzev, S. V. (2013). *Regularization theory for ill-posed problems*, volume 58 of *Inverse and Ill-posed Problems Series*. De Gruyter, Berlin. Selected topics.

[158] Ma, Y.-A., Chen, Y., Jin, C., Flammarion, N., and Jordan, M. (2018). Sampling can be faster than optimization. *arxiv preprint*.

[159] Mair, B. A. and Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5):1424–1444.

[160] Mariucci, E., Ray, K., and Szabó, B. (2020). A Bayesian nonparametric approach to log-concave density estimation. *Bernoulli*, 26(2):1070–1097.

[161] Mathé, P. and Pereverzev, S. V. (2001). Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods. *SIAM J. Numer. Anal.*, 38(6):1999–2021.

[162] Mathé, P. and Pereverzev, S. V. (2006). Regularization of some linear ill-posed problems with discretized random noisy data. *Math. Comp.*, 75(256):1913–1929.

[163] Mazzieri, G., Spies, R., and Temperini, K. (2012). Existence, uniqueness and stability of minimizers of generalized tikhonov–phillips functionals. *Journal of Mathematical Analysis and Applications*, 396(1):396–411.

[164] McLean, W. C. H. (2000). *Strongly Elliptic Systems and Boundary Integral Equations*. Cambridge University Press.

[165] Meyn, S. P. and Tweedie, R. L. (1993). *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London, Ltd., London.

[166] Monard, F., Nickl, R., and Paternain, G. P. (2019). Efficient nonparametric Bayesian inference for X-ray transforms. *Ann. Statist.*, 47(2):1113–1147.

[167] Monard, F., Nickl, R., and Paternain, G. P. (2021a). Consistent inversion of noisy non-Abelian X-ray transforms. *Comm. Pure Appl. Math.*, 74(5):1045–1099.

[168] Monard, F., Nickl, R., and Paternain, G. P. (2021b). Statistical guarantees for bayesian uncertainty quantification in non-linear inverse problems with Gaussian process priors. *Ann. Statist. (to appear)*.

[169] Natterer, F. (1986). *The mathematics of computerized tomography*. B. G. Teubner, Stuttgart; John Wiley & Sons, Ltd., Chichester.

[170] Nickl, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields*, 138(3-4):411–449.

[171] Nickl, R. (2020). Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc. (JEMS)*, 22(8):2697–2750.

[172] Nickl, R. and Paternain, G. (2021). On some information-theoretic aspects of non-linear statistical inverse problems. *arXiv preprint arXiv:2107.09488*.

[173] Nickl, R. and Ray, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist.*, 48(3):1383–1408.

[174] Nickl, R. and Söhl, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.*, 45(4):1664–1693.

[175] Nickl, R. and Söhl, J. (2019). Bernstein–von Mises theorems for statistical inverse problems II: compound Poisson processes. *Electron. J. Stat.*, 13(2):3513–3571.

[176] Nickl, R., van de Geer, S., and Wang, S. (2020). Convergence rates for penalized least squares estimators in PDE constrained regression problems. *SIAM/ASA J. Uncertain. Quantif.*, 8(1):374–413.

[177] Nickl, R. and Wang, S. (2020). On polynomial-time computation of high-dimensional posterior measures by langevin-type algorithms.

[178] Nychka, D. W. and Cox, D. D. (1989). Convergence rates for regularized solutions of integral equations from discrete noisy data. *Ann. Statist.*, 17(2):556–572.

[179] O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statist. Sci.*, 1(4):502–527. With comments and a rejoinder by the author.

[180] O'Sullivan, F. (1990). Convergence characteristics of methods of regularization estimators for nonlinear operator equations. *SIAM J. Numer. Anal.*, 27(6):1635–1649.

[181] Papaspiliopoulos, O., Pokern, Y., Roberts, G. O., and Stuart, A. M. (2012). Nonparametric estimation of diffusions: a differential equations approach. *Biometrika*, 99(3):511–531.

[182] Pardoux, E. and Veretennikov, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085.

[183] Pereyra, M. (2016). Proximal Markov chain Monte Carlo algorithms. *Stat. Comput.*, 26(4):745–760.

[184] Phillips, D. L. (1962). A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9:84–97.

[185] Pinski, F. and Stuart, A. (2010). Transition paths in molecules: gradient descent in pathspace. *Journal of Chemical Physics*, 132:184104.

[186] Pinski, F. J., Stuart, A. M., and Theil, F. (2012). Γ-limit for transition paths of maximal probability. *J. Stat. Phys.*, 146(5):955–974.

[187] Pokern, Y., Stuart, A. M., and van Zanten, J. H. (2013). Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.*, 123(2):603–628.

[188] Prakasa Rao, B. L. S. (1999). *Statistical inference for diffusion type processes*, volume 8 of *Kendall's Library of Statistics*. Edward Arnold, London; Oxford University Press, New York.

[189] Proksch, K., Bissantz, N., and Dette, H. (2015). Confidence bands for multivariate and time dependent inverse regression models. *Bernoulli*, 21(1):144–175.

[190] Protter, P. E. (2004). *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition. Stochastic Modelling and Applied Probability.

[191] Pyatkov, S. G. (2007). Inverse problems for the Black-Scholes equation and related problems. *J. Inverse Ill-Posed Probl.*, 15(9):955–974.

[192] Qi-nian, J. (2000). Error estimates of some Newton-type methods for solving nonlinear inverse problems in Hilbert scales. *Inverse Problems*, 16(1):187–197.

[193] Rantala, M., Vänskä, S., Järvenpää, S., Kalke, M., Lassas, M., Moberg, J., and Siltanen, S. (2006). Wavelet-based reconstruction for limited-angle x-ray tomography. *IEEE Transactions on Medical Imaging*, 25(2):210–217.

[194] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA.

[195] Ray, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.*, 7:2516–2549.

[196] Ray, K. (2014). *Asymptotic theory for Bayesian nonparametric procedures in inverse problems*. PhD thesis, University of Cambridge.

[197] Reiß, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.*, 36(4):1957–1982.

[198] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, volume 293. Springer-Verlag, Berlin, third edition.

[199] Richter, G. R. (1981). An inverse problem for the steady state diffusion equation. *SIAM J. Appl. Math.*, 41(2):210–221.

[200] Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition.

[201] Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, 88(3):603–621.

[202] Roberts, G. O. and Tweedie, R. L. (1996a). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363.

[203] Roberts, G. O. and Tweedie, R. L. (1996b). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110.

[204] Roe, J. (1998). *Elliptic operators, topology and asymptotic methods*, volume 395 of *Pitman Research Notes in Mathematics Series*. Longman, Harlow, second edition.

[205] Roitberg, Y. (1996). *Elliptic Boundary Value Problems in the Spaces of Distributions*. Mathematics and Its Applications. Springer Netherlands, 1 edition.

[206] Royen, T. (2014). A simple proof of the Gaussian correlation conjecture extended to some multivariate gamma distributions. *Far East J. Theor. Stat.*, 48(2):139–145.

[207] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. volume 60, pages 259–268. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).

[208] Ruttor, A., Batz, P., and Opper, M. (2013). Approximate Gaussian process inference for the drift function in stochastic differential equations. In *Advances in Neural Information Processing Systems*, volume 26, pages 2040–2048.

[209] Sakhaee, E. and Entezari, A. (2015). Spline-based sparse tomographic reconstruction with Besov priors. In Ourselin, S. and Styner, M. A., editors, *Medical Imaging 2015: Image Processing*, volume 9413, pages 101 – 108. International Society for Optics and Photonics, SPIE.

[210] Schauer, M., van der Meulen, F., and van Zanten, H. (2017). Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli*, 23(4A):2917–2950.

[211] Schmeisser, H.-J. and Triebel, H. (1987). *Topics in Fourier analysis and function spaces*. A Wiley-Interscience Publication. John Wiley & Sons, Ltd., Chichester.

[212] Schmidt-Hieber, J., Munk, A., and Dümbgen, L. (2013). Multiscale methods for shape constraints in deconvolution: confidence statements for qualitative features. *Ann. Statist.*, 41(3):1299–1328.

[213] Schuss, Z. (1980). Singular perturbation methods in stochastic differential equations of mathematical physics. *SIAM Rev.*, 22(2):119–155.

[214] Schuster, T., Kaltenbacher, B., Hofmann, B., and Kazimierski, K. S. (2012). *Regularization methods in Banach spaces*, volume 10 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin.

[215] Schwab, C. and Stuart, A. M. (2012). Sparse deterministic approximation of Bayesian inverse problems. *Inverse Problems*, 28(4):045003, 32.

[216] Shubin, M. A. (1987). *Pseudodifferential operators and spectral theory*. Springer Series in Soviet Mathematics. Springer-Verlag, Berlin. Translated from the Russian by Stig I. Andersson.

[217] Strauch, C. (2015). Sharp adaptive drift estimation for ergodic diffusions: the multivariate case. *Stochastic Process. Appl.*, 125(7):2562–2602.

[218] Strauch, C. (2016). Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions. *Probab. Theory Related Fields*, 164(1-2):361–400.

[219] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist.*, 46(6B):3451–3480.

[220] Stuart, A. M. (2010). Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559.

[221] Szabó, B., van der Vaart, A. W., and van Zanten, H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist., with discussion*, 43(4):1391–1428.

[222] Talagrand, M. (2014). *Upper and lower bounds for stochastic processes.* Springer, Heidelberg.

[223] Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[224] Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9.

[225] Tikhonov, A. N. (1943). On the stability of inverse problems. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 39:176–179.

[226] Tikhonov, A. N. (1963). On the solution of incorrectly put problems and the regularisation method. In *Outlines Joint Sympos. Partial Differential Equations (Novosibirsk, 1963)*, pages 261–265. Acad. Sci. USSR Siberian Branch, Moscow.

[227] Triebel, H. (1978). *Interpolation theory, function spaces, differential operators*, volume 18 of *North-Holland Mathematical Library.* North-Holland Publishing Co., Amsterdam-New York.

[228] Triebel, H. (2008). *Function spaces and wavelets on domains*, volume 7 of *EMS Tracts in Mathematics.* European Mathematical Society (EMS), Zürich.

[229] Triebel, H. (2010). *Theory of function spaces.* Modern Birkhäuser Classics. Birkhäuser/Springer Basel AG, Basel. Reprint of 1983 edition [MR0730762], Also published in 1983 by Birkhäuser Verlag [MR0781540].

[230] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[231] van der Meulen, F. and Schauer, M. (2017). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.*, 11(1):2358–2396.

[232] van der Meulen, F. H., van der Vaart, A. W., and van Zanten, J. H. (2006). Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli*, 12(5):863–888.

[233] van der Vaart, A. and van Zanten, H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electron. J. Stat.*, 1:433–448.

[234] van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge University Press, Cambridge.

[235] van der Vaart, A. W. and van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.*, 36(3):1435–1463.

[236] van der Vaart, A. W. and van Zanten, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist.*, 37(5B):2655–2675.

[237] van Waaij, J. (2019). Adaptive posterior contraction rates for empirical Bayesian drift estimation of a diffusion. *arXiv e-prints*, page arXiv:1909.12710.

[238] van Waaij, J. and van Zanten, H. (2016). Gaussian process methods for one-dimensional diffusions: Optimal rates and adaptation. *Electron. J. Stat.*, 10(1):628–645.

[239] Vänskä, S., Lassas, M., and Siltanen, S. (2009). Statistical X-ray tomography using empirical Besov priors. *Int. J. Tomogr. Stat.*, 11(S09):3–32.

[240] Vempala, S. S. and Wibisono, A. (2019). Rapid convergence of the unadjusted langevin algorithm: Isoperimetry suffices. *arXiv preprint.*

[241] Vollmer, S. J. (2013). Posterior consistency for Bayesian inverse problems through stability and regression results. *Inverse Problems*, 29(12):125011, 32.

[242] Werner, F. (2015). On convergence rates for iteratively regularized Newton-type methods under a Lipschitz-type nonlinearity condition. *J. Inverse Ill-Posed Probl.*, 23(1):75–84.

[243] Werner, F. and Hohage, T. (2012). Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data. *Inverse Problems*, 28(10):104004, 15.

[244] Wilber, H., Townsend, A., and Wright, G. B. (2017). Computing with functions in spherical and polar geometries II. The disk. *SIAM J. Sci. Comput.*, 39(3):C238–C262.

[245] Wilkinson, D. J. (2006). *Stochastic modelling for systems biology.* Chapman & Hall/CRC Mathematical and Computational Biology Series. Chapman & Hall/CRC, Boca Raton, FL.

[246] Wojdył a, J. and Szkutnik, Z. (2018). Nonparametric confidence bands in Wicksell's problem. *Statist. Sinica*, 28(1):93–113.

[247] Yeh, W. W.-G. (1986). Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research*, 22(2):95–108.

[248] Zhang, Y.-X., Fu, C.-L., and Deng, Z.-L. (2013). An *a posteriori* truncation method for some Cauchy problems associated with Helmholtz-type equations. *Inverse Probl. Sci. Eng.*, 21(7):1151–1168.