Statistical inference and computation in elliptic PDE models

Sixuan Sven Wang

Supervisor: Prof. Richard Nickl

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge
Trinity College

This dissertation is submitted for the degree of
Doctor of Philosophy

September 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It 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 except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis 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 except as declared in the Preface and specified in the text.

Chapter 1 reviews known results, from sources cited throughout. Chapter 2 consists of original research, which was conducted in collaboration with Richard Nickl and submitted to the arXiv [137]. Chapter 3 consists of original research, conducted in collaboration with Richard Nickl and Sara van de Geer and published as [136]. Chapter 4 consists of original research and was published in [178].

Sven Wang
September 2020

Abstract

Partial differential equations (PDE) are ubiquitous in describing real-world phenomena. In many statistical models, PDE are used to encode complex relationships between unknown quantities and the observed data. We investigate statistical and computational questions arising in such models, adopting an infinite-dimensional 'nonparametric' framework and assuming the observed data are subject to random noise. The main PDE examples are of elliptic or parabolic type.

In Chapter 2, we investigate the problem of sampling from high-dimensional Bayesian posterior distributions. The main results consist of non-asymptotic computational guarantees for Langevin-type Markov chain Monte Carlo (MCMC) algorithms which scale polynomially in key quantities such as the dimension of the model, the desired precision level, and the number of available statistical measurements. The bounds hold with high probability under the distribution of the data, assuming that certain 'local geometric' assumptions are fulfilled and that a good initialiser of the algorithm is available. We study a representative non-linear PDE example where the unknown is a coefficient function in a steady-state Schrödinger equation, and the solution to a corresponding boundary value problem is observed.

Chapter 3 investigates statistical convergence rates for nonparametric Tikhonov-type estimators, which can be interpreted also as Bayesian maximum a posteriori (MAP) estimators arising from certain Gaussian process priors. The theory is derived in a general setting for non-linear inverse problems and then applied to two examples, the steady-state Schrödinger equation studied in Chapter 2 and a model for the steady-state heat equation. It is shown that the rates obtained are minimax-optimal in prediction loss.

The final Chapter 4 considers a model for scalar diffusion processes $(X_t : t \geq 0)$ with an unknown drift function which is modelled nonparametrically. It is shown that in the low frequency sampling case, when the sample consists of $(X_0, X_\Delta, ..., X_{n\Delta})$ for some fixed sampling distance $\Delta > 0$, under mild regularity assumptions, the model satisfies the local asymptotic normality (LAN) property. The key tools used are regularity estimates and spectral properties for certain parabolic and elliptic PDE related to $(X_t : t \geq 0)$.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor, Richard Nickl. At the start of my PhD, I looked forward to an exciting intellectual journey during which I would have your outstanding guidance. Not only did this fully turn into reality, but I had no idea that I would also get to play football on Queens' splendid pitches and receive introductions to new literary and musical spheres. Thank you so much for being available to meet whenever needed, for believing in me, for your continuous encouragement, for enabling me to attend conferences, and for sharing your deep mathematical insights and creativity. Likely I still have not completed the list of required non-mathematical readings to qualify for this degree. Our collaborative work, though it happened virtually towards the end, was a really inspiring and enriching experience.

My thanks extend to the whole research group, to my collaborators Matthias and Sergios and to the CCA cohort. It has been a joy to learn, exchange and discuss ideas in this great research community. Thanks to Sam and Ferdia, for being brilliant office mates and friends. I am very grateful to Sara van de Geer for our collaboration and for sharing her enthusiasm for science and music, thereby inspiring me to learn the wonderful 'story' of D959, and to Andrew Stuart for being such a welcoming host during my visit at Caltech this year; I truly enjoyed the great atmosphere and interactions with the research group.

Outside of academia, I feel tremendously lucky to have met great people over the last years, many through the Trinity community and the BBC crew; I look up to all of you. Marius and Olimpia, I always think back happily to our time living together in Portugal street No. 10. Tobi, your endless energy and adventures inspire me. Daniel, thank you for all our good times in the Frazer room, UWA and infinite joys with Bach. Thank you, Heidi, for our countless good conversations; Drasko, for enduring my singing and for celebrating Nikšić; Lydia, for the most 'productive' study sessions in the UL; Jenny, for all the great conversations and travels; Arun, for your ever-refreshing unconventional wisdoms; and finally Nick S., for keeping our friendship amazing throughout and for visiting me in the UK regularly. A big thank you to my dear friends Charlotte, Christoph, Eesha, Elisa, Esma, Fabi, Nick G., Konrad, Leah, Pooneh and Peggy, who have made Cambridge feel home.

I am grateful to all those people whom I had the privilege of sharing the love for music with, many of whom I got to meet through the Cambridge choirs, the Pembroke Lieder Scheme, the Brixen Musikakademie and the UWA meetings (where I learned what true 'binge'

playing of chamber music meant). Daniel, Fanny, Louisa, Manon, Mateo and Max, you have been the greatest musical companions and friends I could imagine. I thank J.S. Bach, W.A. Mozart and F. Schubert for having enriched the world with their incredible musical genius, from which I continuously draw energy and inspiration.

Finally, I would like to express my deepest gratitude to my family. It is hard to imagine how this could have been possible without your support and care, ever since I can remember. That I was able to spend extended time with all of you during the last months was certainly the best side effect of the Corona lockdown. Jiejie and Nikki, thank you so much for our time together at Weigandufer 18 where you introduced me to many great new life perspectives, and I am so excited for a new family member to join soon. Jiejie, thank you for being the most supportive sister I could wish for, and for always having been a role model to me. Baba and Mama, you have always believed in me and only through your love I have made it this far – Danke für alles!

Sven Wang
September 2020

Formal Acknowledgements

I would like to thank the Cantab Capital Institute for the Mathematics of Information and ERC grant No. 647812 for funding my PhD studies. I am grateful to Trinity College, Cambridge and the Cambridge Center for Analysis for further financial support.

Contents

Sind es nicht die kühnen, phantasievollen Hypothesen, zu denen nur der phantastische Geist findet – und die dann vom logischen Denker bewiesen werden können?

Nikolaus Harnoncourt, Musik als Klangrede

Chapter 1

Introduction

Partial differential equations (PDE) are powerful tools to describe real-world phenomena. Hence, there is an abundance of applications in the sciences and engineering in which the collected measurement data follows the pattern of some PDE, and statistical procedures that draw inferences from these data have to account for this PDE structure. Two prominent types of such models are *inverse regression models*, where the relationship between the unknown parameter of interest and the observed data are described by some PDE-governed 'forward map' [91, 155, 10, 94, 159] and *diffusion* or *data assimilation* models in which the data-generating mechanism is governed by stochastic differential equations (SDE) [103, 147, 119, 16]; these models will be introduced in further detail below.

For the inference tasks arising in these settings, a range of methodologies have emerged in the last decades, including classical regularisation methods [63, 17], Bayesian methods [155, 147] and various recent machine-learning based schemes, see, e.g., [8]. Moreover, often it is desired not just to provide point reconstructions, but to quantify the uncertainty in the reconstruction. The Bayesian paradigm offers an elegant way of achieving the latter – via regions of high posterior probability, so-called 'credible sets' – and has been adopted widely in the PDE context at least since influential work by A. Stuart [155, 53].

To understand whether such methodologies can be *trusted* in their conclusions, it is key to understand their theoretical convergence properties. In many PDE models, it is natural to assume that the unknown model parameter is a *function* and hence infinite-dimensional. Thus in practice, the above-mentioned statistical procedures typically rely on algorithms being able to solve challenging computational tasks on high- or infinite-dimensional spaces, usually optimisation problems [28] or sampling problems [150]. Therefore, in giving theoretical performance guarantees, one would ideally like to understand both (1) whether a given procedure is in principle able to guarantee satisfactory recovery, assuming it can be numerically evaluated, and (2) whether, and by what algorithm, the (approximate) numerical computation of the procedure can be guaranteed with a feasible computational cost.

Both of these problem areas have been investigated intensely in the literature for the past decades. The first of them has been studied in the nonparametric statistics literature, see [72, 167] for an overview, and has notably led to the minimax theory of estimation [167, 72, 169] and the frequentist analysis of Bayes procedures [69]. Many results were first derived for the *nonparametric regression* model and variants thereof, where the function of interest is observed directly, corrupted by additive noise. Only recently, significant advances have also been made in extending such results to PDE models, in particular a number of non-linear inverse problems [177, 131, 136, 126, 135, 132, 2] and diffusion models [134, 1].

The theoretical study of high-dimensional computation likewise is a vast field – in this thesis, we shall mainly be concerned with the computation of Bayesian posterior distributions using *Markov chain Monte Carlo* (MCMC) [150]. The methodological design of MCMC schemes in high- and infinite-dimensional spaces has recently attracted a lot of attention, and it would not be possible to name all relevant articles here; we refer to [44, 18, 47, 43] where many more references can be found. There has also been a surge in activity in computational statistics, probability and machine learning to theoretically examine the convergence behaviour of such MCMC schemes, see for instance [48, 57, 58, 80, 26]. Such computational guarantees typically aim to quantify the number of required MCMC iterations (and thus the computational complexity) required in order to numerically evaluate the quantity of interest up to a desired precision level. However, there are very few results of this kind which apply to even basic PDE models – the few notable exceptions will be discussed below. A major hurdle is that many quantitative 'non-asymptotic' computational guarantees in the literature depend on strong geometric assumptions like log-concavity of the target measure, which can be hard or impossible to verify in non-linear settings. Without such assumptions, the computational cost may scale *exponentially* as the model dimension and sample size increase, see Sections 1.3 and 2.1 for more discussion.

In this thesis, we will study three theoretical problems which directly emanate from the above discussion. Chapter 2 aims to develop mathematical theory which allows to assert the 'polynomial-time' feasibility of Bayesian posterior computation even in certain non-linear settings, Chapter 3 studies the convergence rates of penalised least squares estimators in non-linear statistical inverse problems and finally, Chapter 4 investigates the so-called *local asymptotic normality* property for a nonparametric diffusion model. In the rest of this introduction, we review the relevant mathematical context for those topics and thereafter give a brief summary of the Chapters 2–4 to follow. These chapters may also be read independently.

1.1 Inverse regression models with PDE

Suppose that $\mathcal{O} \subseteq \mathbb{R}^d,\ d \in \mathbb{N}$ is some open, bounded subset with smooth boundary $\partial\mathcal{O}$ and that the parameter space $\mathcal{F}$ is a collection of functions $f : \mathcal{O} \to \mathbb{R}$. Then, a prototypical statistical model[1] is the *nonparametric regression model*, where point evaluations of f are observed across the domain $\mathcal{O}$, corrupted by Gaussian noise. Concretely, our data consist of N independent and identically (i.i.d.) distributed pairs $(Y_1, X_1), \dots, (Y_N, X_N)$ given by

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1, ..., N, \tag{1.1}$$

where the $X_i \in \mathcal{O}$ are called the *design points* and are $\varepsilon_i \sim^{i.i.d.} N(0, 1)$ are i.i.d. noise variables. The unknown is the regression function $f \in \mathcal{F}$ itself. This model is ubiquitous in nonparametric statistics, see for instance [167, 72, 69]. For much of this thesis, we shall assume that the X_i are themselves i.i.d. random variables which are uniformly distributed across $\mathcal{O}$ (independently of the ε_i's) – this is known as the *random design* regression model and provides a natural randomised way of modelling 'equally spaced' measurements on $\mathcal{O}$.

However, often the model parameter f cannot be directly observed and it is customary to instead consider an *inverse regression model.* Here we have a collection of functions $\{u_f : f \in \mathcal{F}\} \subseteq L^2(\mathcal{O})$ indexed by some parameter space $\mathcal{F}$, where $L^2(\mathcal{O})$ denotes the usual Lebesgue space of square integrable functions. The relationship between f and the corresponding regression function u_f may also be expressed by a 'forward map'

$$G : \mathcal{F} \mapsto L^2(\mathcal{O}), \quad f \mapsto u_f, \tag{1.2}$$

and in analogy to (1.1) the data are then given by

$$Y_i = G(f)(X_i) + \varepsilon_i, \quad i = 1, ..., N, \tag{1.3}$$

where again $\varepsilon_i \sim^{i.i.d.} N(0, 1)$ and $X_i \sim^{i.i.d.} \mathrm{Uniform}(\mathcal{O})$, independently of the ε_i's. In what follows, we will also write shorthand $Z^{(N)} = ((Y_i, X_i) : i = 1, ...N) \in (\mathbb{R} \times \mathcal{O})^N$ for the full data vector. The law of $Z^{(N)}$ will be denoted by P_f, with associated expectation operator E_f. Since the measurement noise is assumed to be random, the recovery of f from the indirect measurements (1.3) of $G(f)$ constitutes a *statistical inverse problem.*

As elaborated above, in a variety of applications the map G is implicitly given through the solution of some PDE or a system of PDE [155, 91, 45]. Concrete examples are, for instance, the famous Calderón problem [31, 157, 128], linear and non-linear X-ray transforms

[1] When we speak of a statistical model, we generally mean a family $\{P_f : f \in \mathcal{F}\}$ of probability distributions indexed by some set $\mathcal{F}$ (called the *parameter space*), each of which is a candidate for having generated the observed data. When $\mathcal{F}$ is a subset of finite-dimensional Euclidean space $\mathbb{R}^p$ for some $p \geq 1$, we speak of a *parametric* model, and when $\mathcal{F}$ is infinite-dimensional, we will use the standard terminology of saying that $\{P_f : f \in \mathcal{F}\}$ is 'nonparametric'.

[143, 125, 126], time-evolution PDE from fluid dynamics [45, 155] or coefficient-to-solution maps for various elliptic PDE [155, 131, 149], just to name a few. In most of the above cases, the forward map is *non-linear*.

In Chapters 2 and 3 of this thesis, we will study two representative example problems arising from elliptic PDE. In both cases f is a non-negative coefficient function of a partial differential operator and G is a non-linear coefficient-to-solution map for a corresponding boundary value problem. In the first example, suppose $g : \partial O \to (0,\infty)$ is some smooth function representing *known* 'boundary temperatures' on ∂O. Then, for every sufficiently regular $f : \mathcal{O} \to [0,\infty)$, $G(f)$ is given as the unique solution $u = u_f$ of the time-independent *Schrödinger equation*

$$\begin{cases} \Delta u - 2fu = 0 \quad \text{on } \mathcal{O}, \\ u = g \quad \text{on } \partial O, \end{cases} \tag{1.4}$$

where $\Delta = \sum_{i=1}^{d} \partial^2/\partial x_i^2$ denotes the standard Laplace operator. Equations of this kind appear for instance in photoacoustics [10] and scattering problems [11, 88], where the function $f > 0$ may be interpreted as an unknown 'attenuation potential' which one would like to infer from the data, see also [131].

In the second example, for some known 'source function' $g : \mathcal{O} \to (0,\infty)$ and any sufficiently smooth $f : \mathcal{O} \to (0,\infty)$, $G(f)$ is given by the unique solution to the following *divergence form* Dirichlet boundary value problem (for ∇ and $\nabla\cdot$ respectively denoting gradient and divergence)

$$\begin{cases} \nabla \cdot (f\nabla u) = g \quad \text{on } \mathcal{O}, \\ u = 0 \quad \text{on } \partial O, \end{cases} \tag{1.5}$$

The equation (1.5) may be viewed as a steady-state heat equation, where f is a spatially varying heat *diffusivity coefficient*, and u_f describes the temperatures of an equilibrium state. This equation has applications for instance in groundwater flow [182] and has been studied frequently in the mathematical literature [155, 149, 53, 102, 24].

In both models, the constraint $f > 0$ has a clear physical interpretation, and whenever f lies in some Sobolev space $H^\alpha(\mathcal{O})$ for α large enough, unique solutions $G(f)$ exist by means of classical elliptic PDE theory (see, e.g., Chapter 6 of [71]), see Chapter 3 for more details. Therefore, to be concrete, in both problems we may for instance choose

$$\mathcal{F} := \{f : \mathcal{O} \to \mathbb{R} | \inf_{x\in\mathcal{O}} f(x) \geq K_{min} \text{ and } \|f\|_{H^\alpha(\mathcal{O})} \leq R\}$$

as the parameter space, for some $0 < K_{min} < R$ and large enough $\alpha \in \mathbb{N}$.

Given a forward map G, a natural initial question which one may pose is whether G is *injective* – this is a necessary condition to permit the identification of f even if one were to observe $G(f)$ in a 'noiseless' manner. Such injectivity properties for non-linear PDE inverse problems can be very challenging to prove and require a case-by-case analysis, see, e.g., the seminal papers [157, 128] where this is established for the Calderón problem [31]. More quantitative statements about the injectivity of G are often referred to as 'stability estimates', where one studies, very broadly speaking, the continuity properties of the inverse operator $u_f \mapsto f$, between suitable function spaces. For instance, in the divergence form problem (1.5), such stability properties were first examined by Richter [149] by considering the hyperbolic problem

$$\nabla f \cdot \nabla u + f \Delta u = g \quad \text{on } \mathcal{O}$$

subject to appropriate boundary conditions (indeed the above equation is just (1.5) re-written as a PDE for f with known u), where the author also demonstrates that injectivity can be guaranteed so long as the 'heat source function' g is positive, $\inf_{x \in \mathcal{O}} g(x) > 0$. Later these results were refined and extended by various authors, see, e.g., [24, 102], and such stability estimates will play an important role in Chapter 3.

Let us now return to the situation with noisy and discretised measurements $Z^{(N)}$ from (1.3): Here an exact reconstruction of f is typically infeasible, and instead one seeks reconstruction rules which can grant an approximate answer $\hat{f}_N \approx f$. In statistical terminology, we seek *estimators*, i.e., measurable functions $\hat{f}_N = \hat{f}_N(Z^{(N)})$ of the data $Z^{(N)}$ taking values in the parameter space $\mathcal{F}$. A principled way to study the quality of estimators are *convergence rates* of $\hat{f}_N$ towards f in a frequentist statistical sense. Here, for each possible hypothetical 'ground truth' value f_0 generating the data, one evaluates the performance of $\hat{f}_N$ in recovering f_0 in the large sample limit $N \to \infty$. Specifically, if $d : \mathcal{F} \times \mathcal{F} \to [0, \infty)$ is some metric (called the 'loss function') on $\mathcal{F}$, we wish to derive upper bounds $\delta_N \xrightarrow{N\to\infty} 0$ for the expected loss

$$E_{f_0}[d(\hat{f}_N, f_0)] \leq \delta_N, \quad f_0 \in \mathcal{F}. \tag{1.6}$$

In the minimax paradigm (cf. [72, 167]), one further considers the *worst-case risk* over $f \in \mathcal{F}$, given by $\sup_{f \in \mathcal{F}} E_f[d(\hat{f}_N, f)]$. The best such attainable risk for any estimator is called the *minimax risk* r_N, and we say that $\hat{f}_N$ converges at the *minimax rate* if for some constant $C > 0$ and all $N \geq 1$,

$$\sup_{f \in \mathcal{F}} E_f[d(\hat{f}_N, f)] \leq C r_N.$$

For statistical inverse problems with *linear* forward operator G, the minimax theory of estimation is fairly well-understood – roughly speaking, the minimax convergence rate depends on the degree of *ill-posedness* of the forward map [which can for instance be encapsulated by the action of G on different (e.g., Sobolev or Besov) regularity scales of functions [56, 77], or by the rate of decay of the singular values of G [37]]. In the linear case, a variety of methods have been proven to achieve the minimax rate, notably spectral/SVD-based methods and wavelet shrinkage procedures, see [92, 56, 118, 37, 36] and references therein.

In the non-linear setting, however, such methods are unavailable since they rely crucially on the linearity of G. Hence, in the present thesis we will focus our study on *likelihood-based* methods. Recalling (1.3), we see that up to an additive constant the log-likelihood function of the data $Z^{(N)}$ equals the least squares criterion

$$\ell_N(f) = \ell_N(f, Z^{(N)}) = -\frac{1}{2}\sum_{i=1}^{N}[Y_i - G(f)(X_i)]^2. \tag{1.7}$$

For instance, the Bayesian approach to be discussed in the subsequent section is likelihood-based. Another classical example is *Tikhonov regularisation* [164, 63], where one minimises a *penalised least squares* functional $J : \mathcal{F} \to \mathbb{R} \cup \{+\infty\}$ of the form

$$J(f) = \frac{1}{2}\sum_{i=1}^{N}\big[Y_i - G(f)(X_i)\big]^2 + \lambda^2\mathcal{S}(f), \qquad \hat{f}_N \in \arg\min_{f\in\mathcal{F}} J(f), \tag{1.8}$$

where $\lambda > 0$ is a regularisation parameter and $\mathcal{S} : \mathcal{F} \to \mathbb{R} \cup \{+\infty\}$ is a suitable penalty functional. Heuristically, the first term in (1.8) quantifies the 'data misfit' and the second term penalises the 'complexity' of any candidate solution f. Convergence rates of the form (1.6) for Tikhonov regularised estimators have been studied extensively both in linear [46, 120, 36] and non-linear statistical inverse problems [141, 20, 87, 112, 180]. In Chapter 3 we will derive further convergence rate results for the elliptic PDE examples (1.5) and (1.4) in $\|\cdot\|_{L^2(\mathcal{O})}$-loss, where the relation to those existing references will also be discussed.

Lastly, we mention that *iterative regularisation* methods [81, 95, 96, 22, 23, 179] are also commonly used in non-linear inverse problems.

1.2 The Bayesian approach

The Bayesian approach to inverse problems for PDE was advocated for in a number of important papers in the last decade [155, 53]. Since then, significant advances were also made in investigating the performance of Bayes methods in a statistical setting with random measurement noise, see, e.g., [131, 125, 126, 177] and the references below. In this section we briefly review basic definitions and some of those results.

Let us consider again the inverse regression setting of the previous section with forward map $G : \mathcal{F} \to L^2(\mathcal{O})$ and data $Z^{(N)}$ from (1.3), we also recall the log-likelihood function ℓ_N from (1.7).[2] The principle underlying the Bayesian approach is to model the unknown f itself as a *random variable*, distributed according to some prior probability distribution Π on $\mathcal{F}$ – together with the model (1.3) which specifies the law of $Z^{(N)}$ given f, this yields a joint probability law for the pair $(f, Z^{(N)})$. In the Bayesian paradigm, rather than 'explicitly'

[2] We tacitly assume that $\mathcal{F}$ is endowed with a σ-algebra, making it a measurable space, and that G is measurable.

devising reconstruction methods $\hat{f} = \hat{f}(Z^{(N)}) \in \mathcal{F}$, statistical inference is based on the *posterior distribution* on the parameter space $\mathcal{F}$, which is the conditional distribution of f given the observed data $Z^{(N)}$.

Under very mild regularity assumptions, the posterior distribution is given by the well-known *Bayes' formula* – see for instance p.7 in [69]. In our case, it suffices to note that the family of laws P_f of $Z^{(N)}$ has a common dominating measure (e.g., the Lebesgue measure on $(\mathbb{R}\times\mathcal{O})^N$) and to assume that the log-likelihood $(f, Z^{(N)}) \to \ell_N(f, Z^{(N)})$ is jointly measurable. Then, Bayes' formula states that the posterior distribution $\Pi(\cdot|Z^{(N)})$ is given by

$$\Pi(B|Z^{(N)}) = \frac{\int_B e^{\ell_N(f)} d\Pi(f)}{\int_{\mathcal{F}} e^{\ell_N(f)} d\Pi(f)}, \quad B \subseteq \mathcal{F} \text{ measurable}. \tag{1.9}$$

Colloquially, one may paraphrase the above equation as

$$\text{posterior} \quad \propto \quad \text{prior} \ \times \ \text{likelihood}.$$

Of course, an appropriate choice of the prior distribution is key to ensure that the Bayesian approach can succeed. For instance, if the prior distribution puts no mass on a neighbourhood of the true parameter, then the posterior will neither, and in this case posterior-based inference will not be able to recover the true parameter. In the (for this thesis) prototypical case where $\mathcal{F} \subseteq L^2(\mathcal{O})$ is some function space on $\mathcal{O}$, e.g. of Sobolev or Hölder type, one would like to devise infinite-dimensional prior distributions which reflect the regularity properties of the statistical parameter f appropriately. Various possible ways of constructing such priors are discussed, e.g., in the monograph [69] and in [105, 155, 49]. Frequently, priors for functions $f : \mathcal{O} \to \mathbb{R}$ arise as the laws of a *random sequence* or 'basis expansion'

$$f = \sum_{k=1}^{\infty} g_k e_k, \tag{1.10}$$

where $(e_k : k \geq 1) \subseteq L^2(\mathcal{O})$ are the 'basis functions' and $g_k \in \mathbb{R}$ are random coefficients. A plethora of choices for e_k are possible, including polynomials, wavelets, trigonometric series and more. For g_k there are likewise many possibilities, for instance Gaussian random series (see Chapter 11 of [69]), uniform wavelet priors [131] or Besov-type wavelet priors where the g_k are exponentially distributed [105, 49, 6] – the regularity of typical prior draws is then reflected by decay properties of the coefficients g_k.

In fact, by means of the *Karhunen-Loéve expansion*, any Gaussian random field taking values in $L^2(\mathcal{O})$ may be written as a sequence (1.10) with normally distributed coefficients g_k, see, e.g., Example 11.16 in [69]. Later in Chapter 2, we will employ such priors of 'Whittle-Matérn' type, where the basis functions $(e_k : k \geq 1)$ are given as the $L^2(\mathcal{O})$-orthonormal system of eigenfunctions of the Dirichlet-Laplacian Δ on $\mathcal{O}$, see Chapter 2. In numerical practice, priors of the form (1.10) are often discretised by choosing a suitable high-dimensional truncation of the sequence.

1.2.1 Bayesian posterior contraction rates

When the data stem from a ground truth parameter f_0, it is a natural question whether infinite-dimensional Bayes procedures are capable of recovering f_0 as the number of samples N increases, in analogy to convergence rates for point estimates from (1.6). When this 'frequentist' perspective is taken, rather than reflecting any 'subjective beliefs' of the scientist, the Bayes approach is treated as yet another statistical inference procedure whose performance we can evaluate within a framework which does not depend on the prior at all.

For the posterior distribution $\Pi(\cdot|Z^{(N)})$, a central notion of convergence towards f_0 are *posterior contraction rates*, which quantify the size of neighbourhoods of f_0 on which most of the posterior mass concentrates. As before, let $d : \mathcal{F} \times \mathcal{F} \to [0, \infty)$ denote some metric on $\mathcal{F}$.

Definition 1.2.1 (Contraction rate)**.** *Let $\Pi = \Pi_N$ be a sequence of prior distributions on $\mathcal{F}$ and recall the definition of $\Pi_N(\cdot|Z^{(N)})$ from (1.9). We say that a sequence $(\delta_N : N \in \mathbb{N})$, $\delta_n > 0$, is a posterior contraction rate at the parameter f_0 with respect to d if for all $M_N \to \infty$,*

$$\Pi_N\big(\{f : d(f, f_0) \geq M_N\delta_N\}|Z^{(N)}\big) \xrightarrow{N\to\infty} 0, \tag{1.11}$$

in P_{f_0}-probability.

In general nonparametric i.i.d. sampling models, the seminal paper [67] first devised general conditions under which contraction rates for nonparametric Bayes procedures can be established. Since then, the theory has been extended and further developed for a variety of settings, including for Gaussian process priors [173], non-i.i.d. sampling models (such as the Markov chain models considered in Section 1.4 below) [68] and recently exponential Besov-type priors [6]. The study of posterior contraction rates in various statistical models is an active field of research and we refer to Chapters 8-9 of [69] for an excellent overview.

In inverse problems, the investigation of posterior contraction rates began relatively recently and with the linear case. A *conjugate* setting with the Gaussian white noise model, where both prior and posterior are Gaussian, is considered in [99, 7]. These results were then extended by [144] to the non-conjugate setting and in [100] to obtain adaptive contraction rates; we further mention the recent references [98, 77]. The above references all require some 'compatibility' assumption between the forward map and prior distribution in one form or another: The papers [99, 100] achieve this by assuming that one may *simultaneously diagonalise* the prior covariance and forward map, while [7, 77] achieve this, very roughly speaking, by assuming that smoothness properties of the prior and of the forward map can be measured on the same regularity scale of functions.

Recently, significant progress was also made in obtaining contraction rates for a number of non-linear inverse problems of the form (1.3) (or closely related measurement models), including the Schrödinger model (1.4) [131], statistical Calderón problems [2] and non-Abelian X-ray transforms [126]. The divergence form PDE problem (1.5) was treated in [177, 132], where the latter paper achieves minimax optimal convergence rates in prediction loss. On a high level, a unifying feature in the above references is that the proofs proceed by addressing the following two main challenges. First, one proves a contraction rate on the 'forward level'

$$\Pi_N(f : \|G(f) - G(f_0)\|_{L^2(\mathcal{O})} \geq \tilde{\delta}_N | Z^{(N)}) \xrightarrow{N\to\infty} 0 \quad \text{in} \quad P_{f_0}\text{-probability},$$

for some sequence $\tilde{\delta}_N > 0$, and second, one needs to establish a suitable (or use an existing) *stability estimate* for the inverse map $G(f) \mapsto f$. [As discussed earlier in Section 1.1, the latter needs to be studied case-by-case for most non-linear inverse problems.] In our proofs for the Schrödinger problem (1.4) and the divergence form problem (1.5) in Chapters 2 and 3 to follow, we will employ a similar proof principle.

1.2.2 MAP and mean estimates

Bayes methods can also be used to achieve *point reconstruction* via estimators $\hat{f}$ for f_0 which are based on the posterior distribution (1.9). A natural candidate is the *posterior mean* $\bar{f} = \bar{f}(Z^{(N)})$, which is defined by the Bochner integral

$$\bar{f} \equiv \int_{\mathcal{F}} f d\Pi(f|Z^{(N)}), \tag{1.12}$$

whenever the latter is well-defined. Widely used are also maximum a posteriori (MAP) estimators $\hat{f}_{MAP}$, which one may think of as elements $f \in \mathcal{F}$ that are 'most likely' relative to the posterior distribution (1.9). In infinite-dimensional models, this concept requires a careful definition since one cannot simply define $\hat{f}_{MAP}$ as a maximiser of a Lebesgue density on $\mathbb{R}^D$ (as is possible in finite-dimensional models). Instead, it has been proposed to resort to a more abstract notion where MAP estimators are defined via a property that small balls centred around $\hat{f}_{MAP}$ carry maximal probability in an asymptotic sense, see [51, 84, 5].

If $\mathcal{F}$ is a separable Hilbert space and Π a Gaussian process prior on $\mathcal{F}$, then under mild regularity conditions, such 'generalised' MAP estimators $\hat{f}_{MAP}$ can in fact be shown to coincide with minimisers of Tikhonov-type penalised least squares functionals from (1.8) with appropriate penalty norm. Specifically, if $\|\cdot\|_{\mathcal{H}}$ denotes the reproducing kernel Hilbert space (RKHS) norm of Π (see, e.g., Chapter 11 of [69]), the relevant functional is given by

$$J : \mathcal{F} \to \mathbb{R} \cup \{+\infty\}, \quad J(f) = \begin{cases} -\ell_N(f) + \frac{1}{2}\|f\|_{\mathcal{H}}^2 & \text{if } f \in \mathcal{H}, \text{ and} \\ +\infty, & \text{else,} \end{cases} \tag{1.13}$$

see Corollary 3.10 in [51] for a precise statement.

This variational characterisation is not just of theoretical appeal; it also asserts that MAP estimates may be *computed* by means of optimisation methods. In contrast, the evaluation of the posterior mean from (1.12) constitutes a completely different computational task – namely that of high-dimensional numerical integration – and is typically achieved by *averaging* over sufficiently many (approximate) draws $(\vartheta_k : 1 \leq k \leq J)$ from the posterior distribution,

$$\bar{f} \approx \frac{1}{J} \sum_{k=1}^{J} \vartheta_k.$$

Both the optimisation problem of minimising $\mathcal{J}$ as well as the sampling problem of generating (approximate) draws from $\Pi(\cdot|Z^{(N)})$ can be challenging. The latter will be further discussed in Section 1.3 and forms the primary subject of study in Chapter 2 below.

For MAP estimators with Gaussian process priors, by virtue of (1.13) the convergence behaviour can be understood in the same manner as for the Tikhonov-estimators discussed in Section 1.1, see, e.g., [141, 20, 87, 112, 180], and by using our results derived in Chapter 3 below. Convergence rates for $\bar{f}$ were only recently asserted in some non-linear inverse problems [132, 126] by very different proof techniques from Bayesian nonparametrics, namely by combining posterior contraction rates of the form (1.11) with a uniform control over higher moments of the posterior distribution, see for instance Section 3.1 in [132].

1.2.3 Credible and confidence sets

In many applications, one of the main attractions of the Bayesian approach is that it not only yields point reconstructions, but also provides a natural way of quantifying the uncertainty in the reconstruction. Heuristically, for the Bayesian, the more the posterior distribution from (1.9) is 'concentrated' on some region (for instance centred around the mean or the MAP), the more precise the estimate is. This is formalized in the following notion: for any $\gamma \in (0, 1)$, we say that a set $A = A(Z^{(N)}) \subseteq \mathcal{F}$ is a level $1 - \gamma$ *credible set* if

$$\Pi_N(A|Z^{(N)}) \geq 1 - \gamma. \tag{1.14}$$

Of course, a trivial way to devise a credible set would be to choose $A = \mathcal{F}$, but this choice is entirely uninformative. Hence, in practice, we should be preferably interested in constructing credible sets which are 'small' by a suitable criterion (such as $L^2(\mathcal{O})$-diameter) and whose posterior probability is in fact close to $1 - \gamma$.

From the frequentist perspective, one of the central goals of statistical inference is to devise *confidence sets*, i.e., sets $A = A(Z^{(N)}) \subseteq \mathcal{F}$ which satisfy the asymptotic property

$$P_{f_0}(f_0 \in A(Z^{(N)})) \xrightarrow{N \to \infty} 1 - \gamma, \quad f_0 \in \mathcal{F}, \tag{1.15}$$

again for some prescribed level $\gamma \in (0,1)$.[3] One says that A has *asymptotic coverage* $1-\gamma$.

Of course, for the practicioner, it is tempting to utilise a credible set arising from the posterior distribution as an 'off-the-shelf' candidate for a confidence set. However, whether this approach is theoretically justified is an intricate question, and the previously discussed results do not yet provide a satisfactory answer. One way to rigorously assert that Bayesian credible sets indeed constitute confidence sets in the sense of (1.15) is to study the precise limiting shape of the posterior distribution via so-called Bernstein-von-Mises (or, in short, BvM) theorems. In the parametric case where f is a finite-dimensional parameter, the BvM theorem asserts that if the data $Z^{(N)}$ arise from some true parameter f_0, then the law of the rescaled parameter $\sqrt{N}(f-\hat{f})$ under the posterior distribution tends in total variation distance to the *normal* distribution $N(0, I_{f_0}^{-1})$, where $\hat{f}$ is suitable centring estimator and I_{f_0} denotes the Fisher information matrix, see for instance, Chapter 10 in [172].

While the parametric BvM theorem holds under mild conditions which are satisfied by most 'regular' statistical models and prior distributions, the situation is far more subtle in infinite dimensions, and a full discussion of this topic goes beyond the scope of this thesis. Nonparametric BvM theorems which assert the convergence of the rescaled posterior $\sqrt{N}(f-\hat{f})|Z^{(N)}$ to a limiting infinite-dimensional Gaussian process have been established for the standard nonparametric regression model [33, 34, 145] and thereafter also in some linear [125, 73] and non-linear inverse problems [131, 135], see also Chapter 12 of [69] for an overview. In numerous PDE problems including (1.5), it remains a challenging open problem to establish such a BvM theorem.

1.3 Bayesian computation by MCMC

The computational tasks arising from the posterior-based methods discussed in the previous section can be challenging to solve numerically. One of the key methodologies for Bayesian computation is Markov chain Monte Carlo (MCMC) [150], and in this section we shall discuss some basic examples as well as some relevant theoretical *convergence guarantees.*

While the statistical models introduced in the previous sections were generally indexed by an infinite-dimensional parameter $f \in \mathcal{F}$, in practice it is common to employ a D-dimensional Euclidean parameter ($D \geq 1$) which we shall denote by $\theta \in \mathbb{R}^D$. Usually $\mathbb{R}^D$ can be viewed as an 'approximation subspace' of the parameter space $\mathcal{F}$ via some suitable one-to-one parametrisation $\theta \to f_\theta \in \mathcal{F}$. Since D may grow as the sample size N increases, we speak of θ as a 'high-dimensional' parameter. In this section, suppose that Π is some prior probability distribution on $\mathbb{R}^D$ with Lebesgue density π and that $\theta \mapsto \ell_N(\theta)$ is a likelihood function arising from some statistical model (not necessarily of inverse regression type). Then, the posterior probability distribution $\Pi(\cdot|Z^{(N)})$ from which we wish to sample also possesses a

[3]Alternative definitions, for instance where uniformity $\inf_{f\in\mathcal{F}} P_f(f \in A(Z^{(N)}))$ is required, are also common. However, we shall not discuss this further.

Lebesgue density, of the form

$$\pi(\theta|Z^{(N)}) \propto e^{\ell_N(\theta)}\pi(\theta),\ \theta \in \mathbb{R}^D. \tag{1.16}$$

Sampling from a measure of the form (1.16) is relevant not just in the statistical inverse problem setting discussed above, but also a variety of other contexts, for instance conditioned diffusions or data assimilation, as illustrated for instance in [44].

1.3.1 Examples of MCMC algorithms

The literature on MCMC in high-dimensional and function spaces has grown rapidly in recent years and it would not be possible here to recall all the schemes which have been proposed; see for instance [44, 18, 43, 47] for a range of methodologies. Let us discuss two representative examples of MCMC algorithms which are used regularly in the context of high-dimensional statistical models. The first example is a *gradient-based* algorithm which arises as the Euler-Maruyama discretisation of the Langevin SDE

$$dL_t = [\nabla \log \pi(\cdot|Z^{(N)})](L_t)dt + \sqrt{2}dW_t, \quad L_t \in \mathbb{R}^D,\ t \geq 0, \tag{1.17}$$

where $(W_t : t \geq 0)$ is a D-dimensional Brownian motion. It is well-known that under standard regularity assumptions, (1.17) possesses a unique strong solution and that $\Pi(\cdot|Z^{(N)})$ is the unique invariant measure for $(L_t : t \geq 0)$, see, e.g., [48]. The MCMC scheme requires choice of a step size $\gamma > 0$ as well as an initialisation point θ_{init} and is stated precisely in Algorithm 1 below. Due to the discretisation step, the invariant distribution of the resulting Markov chain is biased, which is not being corrected for – hence the algorithm is often called *Unadjusted Langevin algorithm* [152].

Algorithm 1 Unadjusted Langevin algorithm

Input: Initialiser $\theta_{init} \in \mathbb{R}^D$, step size $\gamma > 0$, *i.i.d.* sequence $\xi_k \sim N(0, I_{D\times D})$.

initialise $\vartheta_0 = \theta_{init}$
for $k = 0, 1, \ldots$ **do**
    $\vartheta_{k+1} = \vartheta_k + \gamma\nabla \log \pi(\vartheta_k|Z^{(N)}) + \sqrt{2\gamma}\xi_{k+1}$
return $(\vartheta_k : k \geq 1)$

The second example is the *preconditioned Crank-Nicolson* (pCN) algorithm, which was first proposed in [44]. The pCN algorithm is an instance of a *Metropolis-Hastings* Markov chain [123, 83], where each iteration of the Markov chain consists of two steps: (1) a proposal step and (2) an accept-reject step. The advantage of such methods is that it allows to choose from a broad class of proposal distributions. It can be shown under very general conditions that the accept-reject mechanism ensures that the resulting Markov chain has the correct stationary distribution [163]. For a general overview of Metropolis-Hastings methods, see [150].

The pCN algorithm is designed specifically to take advantage of *Gaussian prior* distributions. Suppose Π is an $N(0,\mathcal{C})$ distribution with some $D\times D$ positive definite covariance matrix $\mathcal{C}$, such that the posterior density is given by

$$\pi(\theta|Z^{(N)}) \propto \exp\Big(\ell_N(\theta) - \frac{1}{2}\theta^T\mathcal{C}^{-1}\theta\Big),\ \theta\in\mathbb{R}^D.$$

The pCN proposal distribution is then based on a semi-explicit Euler discretisation of the Ornstein-Uhlenbeck type SDE

$$dL_t = -L_t dt + \sqrt{2\mathcal{C}}dW_t, \quad L_t\in\mathbb{R}^D,\ t\geq 0, \tag{1.18}$$

where $(W_t : t\geq 0)$ again is a D-dimensional Brownian motion, see [44] for the construction. The dynamics (1.18) are chosen to leave the prior distribution $N(0,\mathcal{C})$ invariant; as a result, the acceptance probabilities for each proposal step only depend on *likelihood ratios*, see Algorithm 2. One of the premier advantages of the pCN algorithm is that it can be formulated in an infinite dimensional setting with essentially no modifications, see, e.g., [44, 80]. Therefore, its ergodicity properties may be analysed in a manner that is dimension-independent – this done in the paper [80] which will be discussed further below.

Algorithm 2 preconditioned Crank-Nicolson algorithm

Input: Initialiser $\theta_{init}\in\mathbb{R}^D$, step size $\beta\in(0,1)$, $i.i.d.$ sequence $\xi_k\sim N(0,\mathcal{C})$.

1: **initialise** $\vartheta_0 = \theta_{init}$
2: **for** $k=0,1,\dots$ **do**
3: **set** $z_{k+1} = \sqrt{1-\beta}\vartheta_k + \sqrt{\beta}\xi_{k+1}$.
4: **set** $\vartheta_{k+1} = \begin{cases} z_{k+1} & \text{with probability} \quad \min\{1, e^{\ell_N(z_{k+1})-\ell_N(\vartheta_k)}\} \\ \vartheta_k & \text{else}\end{cases}$
5: **return** $(\vartheta_k : k\geq 1)$

As mentioned above, the ULA and pCN algorithms are just two representatives of a larger class of MCMC schemes based on time-discretisations of SDEs. Other commonly used algorithms of this type include the Metropolis-adjusted Langevin algorithm (MALA) [151, 44] and Hamiltonian Monte Carlo (HMC) [18].

1.3.2 Computational cost of MCMC

A main topic of this thesis is to study the *computational cost* of MCMC in high-dimensional models. In the context here, we shall use this term synonymously with the number of *MCMC iterations* required to perform some given computational task up to a prescribed precision level.[4] Specifically, we will consider the following two central aims of posterior computation:

(1) To synthesize a random variable whose law on $\mathbb{R}^D$ (approximately) equals $\Pi(\cdot|Z^{(N)})$.

(2) To (approximately) compute *integrals* against the posterior distribution

$$E^\Pi[H|Z^{(N)}] := \int_{\mathbb{R}^D} H(\theta) d\Pi(\cdot|Z^{(N)})(\theta), \tag{1.19}$$

for suitable test functions $H : \mathbb{R}^D \to \mathbb{R}$.

For a given MCMC scheme $(\vartheta_k : k \geq 1)$, studying the computational cost to address these tasks typically boils down to a detailed understanding of the ergodicity properties of $(\vartheta_k : k \geq 1)$. Firstly, given a prescribed precision level $\varepsilon > 0$ and some metric d on the space of probability distributions on $\mathbb{R}^D$, how long does the Markov chain $(\vartheta_k : k \geq 1)$ take to mix up to precision ε? In other words, one seeks to bound the *mixing time*

$$k_{mix}(\varepsilon) := \inf \Big\{ k \in \mathbb{N} : d(\mathcal{L}(\vartheta_k), \mu) \leq \varepsilon \Big\}. \tag{1.20}$$

Computation of the integral (1.19) is most commonly addressed by taking an ergodic average over $J \geq 1$ elements of the Markov chain after some 'burn-in time' $J_{in} \geq 1$,

$$\hat{\pi}^J_{J_{in}}(H) := \sum_{k=J_{in}+1}^{J_{in}+J} H(\vartheta_k). \tag{1.21}$$

Thus the second relevant question is as follows. Given $\varepsilon > 0$, how large do $J, J_{in} \geq 1$ need to be chosen in (1.21) to ensure

$$|\hat{\pi}^J_{J_{in}}(H) - E^\Pi[H|Z^{(N)}]| \leq \varepsilon, \tag{1.22}$$

with high probability under the law of $(\vartheta_k : k \geq 1)$?

Somewhat 'non-explicit' convergence properties for various MCMC schemes including Langevin-type algorithms were studied in a line of important early work, see, e.g., [151, 152, 124]. However, in our setting, it is particularly important to understand the explicit scaling of the mixing properties of $(\vartheta_k : k \geq 1)$ both in dimension D and sample size N, since this is the asymptotic regime in which consistent statistical recovery can be achieved. Recently,

[4] We note that in practice, computational challenges of different nature may also arise, for instance when evaluation of the likelihood function ℓ_N requires numerical solution of complicated PDE forward maps. However this will not be discussed further in the present thesis.

there have been significant research efforts to obtain such explicit, non-asymptotic guarantees suited for the high-dimensional setting, which we shall discuss in the rest of this section.

Mixing of Langevin-type algorithms under log-concavity assumptions

For ULA from Algorithm 1 targeting *strongly log-concave* distributions, the first explicit convergence guarantees suited for the high-dimensional setting were proved by Dalalyan [48] as well as Durmus and Moulines [57, 58]. The papers [48, 57] considered mixing times in total variation whereas [58] extends these results to the Wasserstein distance.

To illustrate, and since those results will play an important role in Chapter 2, let us briefly state a simplified version of Theorem 5 in [58]. For any Borel probability measures μ_1, μ_2 on $\mathbb{R}^D$ with finite second moments we define the Wasserstein distance

$$W_2^2(\mu_1, \mu_2) = \inf_{\nu \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^D \times \mathbb{R}^D} \|\theta - \vartheta\|_{\mathbb{R}^D}^2 d\nu(\theta, \vartheta), \tag{1.23}$$

where $\Gamma(\mu_1, \mu_2)$ is the set of all 'couplings' of μ_1 and μ_2 (see, e.g., [176]).

Proposition 1.3.1 (Wasserstein mixing of ULA)**.** *Suppose that $\pi(\theta|Z^{(N)})$ from (1.16) is positive throughout $\mathbb{R}^D$ and write $U(\theta) := -\log \pi(\theta|Z^{(N)})$. Assume $U : \mathbb{R}^D \to \mathbb{R}$ is continuously differentiable and that there exist constants $0 < m \leq \Lambda < \infty$ such that for all $\theta, \bar{\theta} \in \mathbb{R}^D$,*

$$\begin{aligned} \|\nabla U(\theta) - \nabla U(\bar{\theta})\|_{\mathbb{R}^D} &\leq \Lambda \|\theta - \bar{\theta}\|_{\mathbb{R}^D}, \qquad (\Lambda\text{-gradient Lipschitz}), \\ U(\bar{\theta}) - U(\theta) - \langle \nabla U(\theta), \bar{\theta} - \theta \rangle_{\mathbb{R}^D} &\geq \frac{m}{2} \|\theta - \bar{\theta}\|_{\mathbb{R}^D}^2, \qquad (m\text{-strongly convex}). \end{aligned}$$

Then, if θ_U denotes the unique minimiser of U and $(\vartheta_k : k \geq 1)$ is given by Algorithm 1 for some $\gamma \leq 1/\Lambda$ and $\theta_{init} \in \mathbb{R}^D$, then the laws $\mathcal{L}(\vartheta_k)$ of ϑ_k on $\mathbb{R}^D$ satisfy

$$W_2^2(\mathcal{L}(\vartheta_k), \Pi(\theta|Z^{(N)})) \leq C\Big((1 - m\gamma/2)^k \Big[\|\theta_{init} - \theta_U\|_{\mathbb{R}^D}^2 + \frac{D}{m} \Big] + \frac{\gamma D \Lambda^2}{m^2} + \frac{\gamma^2 D \Lambda^4}{m^3} \Big),$$

for some universal constant $C > 0$ and all $k \geq 0$.

The derivation of this result can be found in Section 2.5.1. An important implication of the preceding proposition is that in the strongly log-concave setting, the mixing times (1.20) of ULA in Wasserstein distance depends *polynomially* on the dimension D and the desired precision level. In fact, the paper [58] also asserts the polynomial-time feasibility of Monte Carlo integration by ergodic averages $\hat{\pi}^J_{J_{in}}(H)$ from (1.21) by proving non-asymptotic concentration inequalities akin to (1.22), see also Section 2.5.1. The proofs in [58] are based on a rather 'explicit' analysis of both the continuous-time and discretised Langevin dynamics. For both cases one can construct explicit (synchronous) couplings, where Wasserstein contractivity follows relatively directly from strong log-concavity (see the proofs of Proposition 1 and 3 in [58]). Similarly, explicit arguments are used to prove concentration bounds for $\hat{\pi}^J_{J_{in}}(H)$ – here the authors exploit the fact that the transition kernel for the ULA chain, conditional on the current state ϑ_k, is a Gaussian distribution $\vartheta_{k+1}|\vartheta_k \sim N(\vartheta_k - \gamma\nabla U(\vartheta_k), 2\gamma I_{D\times D})$. The quantitative bounds are then obtained using martingale arguments and concentration inequalities for Gaussian distributions, see the proof of Theorem 15 in [58]. Such concrete arguments seem harder to conduct for MCMC schemes like pCN or MALA which involve a Metropolis-Hastings correction, see, e.g., [80, 59].

In the past years, the results from [48, 57, 58] for the ULA were extended to various other settings, asserting non-asymptotic mixing bounds under closely related 'geometric' conditions such as log-Sobolev inequalities, or log-concavity outside of some Euclidean ball [40, 175, 117]. We conclude by mentioning some results for related Langevin-type algorithms. For continuous-time diffusions, mixing times are obtained in [61, 62] by means of so-called reflection couplings. Again in the strongly log-concave setting, the papers [27, 60, 59] derive convergence results for MALA; HMC algorithms are considered in [26, 110, 39, 38].

Beyond log-concavity

When the posterior distribution $\Pi(\cdot|Z^{(N)})$ from (1.16) is non-log-concave, the situation is far less clear and and polynomial-time mixing bounds seem more challenging to obtain. The elliptic PDE inverse problems (1.4) and (1.5) fall into this setting. Here, the nonlinearity of the forward map G causes the negative log-likelihood from (1.7) to be non-convex in general. To illustrate the difficulties which can arise, note that non-log-concave target densities may in principle be multi-modal. That this leads to a deterioration of the mixing time can already be shown in the one-dimensional $D = 1$ case – as elaborated in Example 1 in [61], one can see from a basic double-well example that the *exit time* of the continuous time Langevin-dynamics (1.17) from local optima of the log-posterior density $\log \pi(\cdot|Z^{(N)})$ can scale exponentially in the height of the local optimum.

However, there do exist a number of papers which analyse the computational complexity of MCMC in high dimensions under different structural assumptions than the above-mentioned ones. The first such article is [80] by Hairer, Stuart and Vollmer, where the pCN scheme from Algorithm 2 is studied in a setting with Gaussian process priors. There, under general local Lipschitz assumptions on the likelihood function as well as a lower bound assumption on the acceptance probabilities $1 \wedge \exp(\ell_N(\theta) - \ell_N(\bar{\theta}))$ (for suitable $\theta, \bar{\theta} \in \mathbb{R}^D$) from Algorithm 2, the authors first obtain a contractivity property in a Wasserstein-type distance (using probabilistic results from [79]), which is then shown to yield an $L^2(\mu)$-spectral gap. An important aspect is that the spectral gap is shown to be dimension-independent, thus demonstrating that it is possible in principle to devise MCMC schemes whose mixing properties are invariant with respect to the dimension D at which an infinite-dimensional model is discretized. However, the results derived in [80] remain implicit in other important manners, which makes it non-obvious how to obtain non-asymptotic mixing guarantees for our setting in Chapter 2 below – see Section 2.1.2 for more discussion.

The paper [14] discusses the computational cost of MCMC in a frequentist setting with high-dimensional statistical models where the model dimension D and sample size N tend to infinity simultaneously. The starting point of their analysis is a 'Bernstein-von-Mises' type assumption that the posterior distribution, when suitably rescaled and recentred, converges to a Gaussian measure in a strong enough sense (cf. Section 1.2.3). Lastly, we mention [181] where the computational complexity of Bayesian *variable selection* in a high-dimensional linear regression model is analysed. Here the non-log-concavity of the sampling task arises from sparsity assumptions on the parameter.

1.4 Nonparametric models for scalar SDE

We now discuss another class of statistical models in which the data is governed by a differential equation and which will be studied in Chapter 4 of this thesis, namely models for stochastic *diffusion processes*. Specifically, for some *drift function* $b : \mathbb{R} \to \mathbb{R}$ and *diffusion coefficient* $\sigma : \mathbb{R} \to (0, \infty)$, consider the one-dimensional time-homogeneous Itô stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + \sqrt{2}\sigma(X_t)dW_t, \qquad t \geq 0, \tag{1.24}$$

where $(W_t : t \geq 0)$ is a standard Brownian motion. Whenever b and σ are Lipschitz continuous, there exists a unique pathwise solution to (1.24), see, e.g., Theorem 24.3 in [13].

In the statistical setting, at least one of the two coefficients b and σ are unknown and the goal is to infer b and/or σ from successive observations from one realisation of the stochastic process $(X_t : t \geq 0)$. Hence, in order for consistent statistical recovery to be possible in the large sample limit, one typically needs additional assumptions ensuring that $(X_t : t \geq 0)$ has sufficiently strong 'recurrence' properties. For instance, [75, 134] consider a modification of (1.24) where the diffusion is constrained to $[0, 1] \subseteq \mathbb{R}$ and reflected at the boundary points $\{0, 1\}$. Alternatively one may assume a periodic setting [133, 111, 1] or that b satisfies a drift condition at $\pm\infty$ [4]. For the sake of this introduction, we shall ignore this important technical point and refer to [75, 134, 1] and Chapter 4 below for further discussion.

In contrast to the i.i.d. inverse regression setting from Section 1.1, the randomness in the observations stems from the inherent stochasticity of the dynamics (1.24). The literature on statistical inference for diffusions has considered various different observation schemes. For instance, the references [103, 174, 4, 3, 133] assume the *continuous observation* scheme in which the data are given by a continuous path $X^{(T)} = (X_t : 0 \leq t \leq T)$ with asymptotics $T \to \infty$. In the *high frequency observation* scheme, one instead observes discrete states

$X^{(n)} = (X_0, X_{\Delta_n}, \dots, X_{n\Delta_n})$ where the time steps Δ_n between observations tend to 0 as $n \to \infty$, see, e.g., [1, 74].

In the present thesis, we will be concerned with the *low frequency observation* scheme, where the data are given by

$$X^{(n)} = (X_0, X_\Delta, \dots, X_{n\Delta}), \tag{1.25}$$

for some *fixed* time difference $\Delta > 0$ between measurements. In this setting, both upper and lower bounds for the minimax rate of estimation were first obtained by Gobet, Hoffmann and Reiss in the seminal paper [75]. More specifically, they considered reflected diffusions $(X_t : t \geq 0)$ taking values in $[0, 1]$, and assumed the pair (σ, b) to belong to the Sobolev regularity class

$$\Theta_s := \Big\{(\sigma, b) \in H^s((0,1)) \times H^{s-1}((0,1)) : \ \|\sigma\|_{H^s} \leq C, \ \|b\|_{H^{s-1}} \leq C, \ \inf_{x\in(0,1)} \sigma(x) \geq c\Big\} \tag{1.26}$$

for some integer $s > 1$ and constants $C \geq c > 0$. In this case the Markov chain $(X_{i\Delta} : i \geq 0)$ possesses a unique invariant measure μ. The estimation method proposed in [75] then proceeds in two steps. First, one estimates μ by a standard wavelet projection estimator $\hat{\mu}$ as well as the *transition operator* $P_{\Delta,\sigma,b}$, which is defined via the action

$$P_{\Delta,\sigma,b}h(x) = E_{\sigma,b}[h(X_\Delta)|X_0 = x]$$

on appropriate test functions h, by a high-dimensional matrix estimator $\hat{\mathbf{P}}_\Delta$. [Here we have denoted expectation w.r.t. the law of (1.25) under (σ, b) by $E_{\sigma,b}$.] Second, by exploiting explicit reconstruction formulas for one-dimensional diffusions (see Section 3 of [75]) one obtains simple 'plug-in' estimates $\hat{\sigma}_n, \hat{b}_n$ from $\hat{\mathbf{P}}_\Delta, \hat{\mu}$. For any $0 < a_0 < a_1 < 1$, and , the spectral method is then shown to achieve the convergence rates

$$\begin{aligned} E_{\sigma,b}\big[\|\hat{\sigma}_n^2 - \sigma^2\|_{L^2([a_0,a_1])}\big] &\lesssim n^{-s/(2s+3)}, \\ E_{\sigma,b}\big[\|\hat{b}_n - b\|_{L^2([a_0,a_1])}\big] &\lesssim n^{-(s-1)/(2s+3)}, \end{aligned} \tag{1.27}$$

which are proved to be *minimax-optimal* – see Theorem 2.4 in [75] for the precise statement.

Interestingly, these spectral methods circumvent the explicit use of the *likelihood structure* of the low frequency diffusion model. Only recently, it was proved by Nickl and Söhl [134] that (likelihood-based) nonparametric Bayes methods are able to achieve recovery at the same minimax convergence rate (1.27) over Θ_s from (1.26), up to log-factors.

1.5 Contributions

In this thesis, we study three problems related to the PDE/SDE models discussed above.

1.5.1 On polynomial-time posterior computation

In Chapter 2, we consider sampling from high-dimensional posterior distributions $\Pi(\cdot|Z^{(N)})$ by means of MCMC algorithms. As in Section 1.3, $\Pi(\cdot|Z^{(N)})$ is assumed to have a density of the form

$$\pi(\cdot|Z^{(N)}) = e^{\ell_N(\theta)}\pi(\theta),\ \ \theta \in \mathbb{R}^D, \tag{1.28}$$

where $\ell_N(\theta)$ is some log-likelihood function, N denotes sample size, and π is the density of the prior distribution. Specifically, we will study (1) the computation of posterior characteristics (such as its mean) by ergodic averages as well as (2) global mixing times in Wasserstein distance, as defined by (1.20). As discussed already in Section 1.3.2, without further geometric assumptions on the target density (such as log-concavity) it can be challenging to give polynomial-time convergence guarantees, and the aim of Chapter 2 is to develop mathematical techniques which allow to overcome such hardness barriers in certain non-linear and high-dimensional settings.

The particular MCMC scheme considered will be a Langevin-type Markov chain $(\vartheta_k : k \geq 1)$ which is closely related to the Unadjusted Langevin algorithm from Algorithm 1 and which is described in detail in Section 2.2. Our theory is developed in general non-linear inverse regression models (1.3) with random design and a 'high-dimensional' parameter space $\mathbb{R}^D$. The main PDE example studied is the inverse problem for the Schrödinger equation described in (1.4) – in the terminology of Section 1.1, the forward map $G : \mathbb{R}^D \to L^2(\mathcal{O})$ then arises via solutions $u \equiv G(\theta)$ of the boundary value problem

$$\begin{cases} \Delta u - 2f_\theta u = 0 \ \text{ on } \ \mathcal{O}, \\ u = g \ \text{ on } \ \partial\mathcal{O}, \end{cases}$$

where $\theta \mapsto f_\theta > 0$ is a suitable parameterisation. We will assume the prior distributions to be Gaussian – thus the main difficulty arises from the non-log-concavity of the log-likelihood function ℓ_N.

Our main results in the Schrödinger model can be summarised as follows. First, in Proposition 2.2.4 we obtain non-asymptotic concentration inequalities for ergodic averages of the Langevin-type algorithm. As as consequence, with high probability under the distribution of the data, the posterior mean vector can be computed up to prescribed precision level $\varepsilon > 0$ using polynomially many (in D, N and ε) MCMC iterations, see Theorem 2.2.5. In Theorem 2.2.7, we obtain similar polynomial-time bounds for the mixing time of $(\vartheta_k : k \geq 1)$ in the Wasserstein distance from (1.23) – specifically, we have

$$W_2(\mathcal{L}(\vartheta_k), \Pi(\cdot|Z^{(N)})) \leq \varepsilon \ \text{ for all } \ k \geq k_{mix}(\varepsilon), \quad k_{mix}(\varepsilon) = O(D^{a_1}N^{a_2}\varepsilon^{-a_3})$$

for some exponents $a_1, a_2, a_3 > 0$. A further consequence of the proofs is that maximum a posteriori (MAP) estimators are computable by a gradient descent method within $O(D^{b_1} N^{b_2} \varepsilon^{-b_3})$, $b_1, b_2, b_3 > 0$, iterations (cf. Theorem 2.2.8) and that the 'ground truth' parameter θ_0 can be computed within the statistical estimation error in polynomial time.

Our approach is to study in detail the local geometry of the posterior distribution $\Pi(\cdot|Z^{(N)})$. The first key step lies in showing that even when $\Pi(\cdot|Z^{(N)})$ is not globally log-concave, the posterior density satisfies a *local* log-concavity property which is satisfied on some region around the ground truth parameter, with high probability under the law of the data. This can be proven using concentration of measure arguments and assuming that the derivative vector field $\nabla_\theta G(\theta) : \mathcal{O} \to \mathbb{R}^D$ satisfies a suitable *stability estimate* for all θ which are sufficiently close to θ_0 (see Lemma 2.3.4 below). We verify such stability estimates in the Schrödinger model using techniques from elliptic PDE (see Lemma 2.4.7).

The second main step consists in proving that most of the mass of the posterior distribution concentrates on the region of log-concavity. To show this, we use techniques in Bayesian nonparametric inference which were recently developed to tackle non-linear inverse problems with Gaussian priors [126, 131, 173]; see also the discussion in Section 1.2 above. The importance of this step is that it allows to deduce that the posterior distribution can be approximated in Wasserstein distance by a globally log-concave measure (this is stated in Theorem 2.4.14 for the Schrödinger model). As auxiliary results, convergence rates for MAP estimators and contraction rate results for the posterior akin to (1.11) are proved.

The previous two steps can then be used to conclude that when sufficiently good initialisation (into the region of log-concavity) is feasible, polynomial-time mixing bounds can be derived from the recent non-asymptotic convergence results for Langevin algorithms [48, 58] discussed above in Section 1.3.2.

1.5.2 Convergence rates of Tikhonov regularised estimators

In Chapter 3, we study convergence rates for Tikhonov-regularised estimators introduced in (1.8) where the penalty norms considered are of squared Sobolev-norm type $\|\cdot\|_{H^\alpha(\mathcal{O})}$, $\alpha \in \mathbb{N}$. As discussed in Section 1.1, these estimators can also be interpreted as Bayesian maximum a posteriori estimates for Gaussian process priors with reproducing kernel Hilbert space $\mathcal{H} = H^\alpha(\mathcal{O})$.

The two primary examples are the non-linear PDE models (1.5) for the divergence form equation and (1.4) for the Schrödinger equation. The measurement model considered is the *Gaussian white noise* which is intimately related to the random design regression model (1.3). Here one observes a realisation of the 'continuum' stochastic process

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

where $\varepsilon > 0$ is the noise level and $\mathbb{W} = (\mathbb{W}(\psi) : \psi \in L^2(\mathcal{O}))$ denotes a Gaussian white noise process in the Hilbert space $L^2(\mathcal{O})$ – see Chapter 3 for details. Indeed, under standard smoothness assumptions on the regression functions $G(f)$ and with scaling $\varepsilon \simeq N^{-1/2}$ of the noise level it can be shown that (1.29) and (1.3) are asymptotically equivalent in the sense of *Le Cam equivalence* [148, 30] and in the literature it is standard to consider the white noise model in place of (1.3), see, e.g., [72, 167, 99, 144, 131, 125, 136].

The final objective function is a version of (1.8) adapted to the white noise setting and to incorporate the constraint $f > 0$ required from the PDE models. Thus, for some smooth one-to-one link function $\Phi : \mathbb{R} \to (0, \infty)$, the criterion takes the form

$$J(f) = -2\langle Y^{(\varepsilon)}, G(f)\rangle_{L^2(\mathcal{O})} + \|G(f)\|^2_{L^2(\mathcal{O})} + \lambda^2 \|\Phi^{-1} \circ f\|^2_{H^\alpha(\mathcal{O})}, \quad \lambda > 0,$$

for a suitably chosen scalar $\lambda > 0$, see Section 3.3.

Our main results can be summarised as follows. First, we prove that minimisers $\hat{f}$ of J exist almost surely under the law of $Y^{(\varepsilon)}$ from (1.29). We then obtain the following convergence rates for $\hat{f}$ towards f and for the 'plug-in' estimator $G(\hat{f})$ towards $G(f)$:

$$E_{f_0}\|G(\hat{f}) - G(f_0)\|_{L^2(\mathcal{O})} \le C\varepsilon^{\frac{2\alpha+2}{2\alpha+2+d}}, \quad E_{f_0}\|\hat{f} - f_0\|_{L^2(\mathcal{O})} \le C\varepsilon^{\frac{2\alpha-2}{2\alpha+2+d}} \quad \text{(Div. form)}, \tag{1.30}$$

$$E_{f_0}\|G(\hat{f}) - G(f_0)\|_{L^2(\mathcal{O})} \le C\varepsilon^{\frac{2\alpha+4}{2\alpha+4+d}}, \quad E_{f_0}\|\hat{f} - f_0\|_{L^2(\mathcal{O})} \le C\varepsilon^{\frac{2\alpha}{2\alpha+4+d}} \quad \text{(Schrödinger)}. \tag{1.31}$$

Here, $C > 0$ is a constant which can be chosen over suitable positivity-constrained α-regular Sobolev classes, for the precise statements we refer to the main Theorems 3.3.3-3.3.7. Note that the respective convergence rates for $\|G(\hat{f}) - G(f_0)\|_{L^2}$ correspond to the minimax rate (with direct observations) for estimating an $(\alpha + 1)$-smooth function in the divergence form problem and an $(\alpha + 2)$-smooth function in the Schrödinger problem. By proving a lower bound for the convergence rate in divergence form problem which matches the upper bound (1.30), both convergence rates for $\|G(\hat{f}) - G(f_0)\|_{L^2}$ are shown to be minimax optimal – in this sense, one may think of the forward maps as 1- and 2- smoothing respectively.

Let us briefly discuss the proof ideas. First, we establish a convergence rate result (Theorem 3.2.2 below) on the level of the regression functions $\|G(\hat{f}) - G(f_0)\|_{L^2}$ in general non-linear inverse problems – this result is shown under a *local Lipschitz condition* on the forward map (see (3.8) below) between $L^2(\mathcal{O})$ and $(H^\kappa)^*$, $\kappa \in \mathbb{N}$, where $(H^\kappa)^*$ denotes the topological dual space of $H^\kappa(\mathcal{O})$. One may think of κ as the 'level of smoothing' of G. In the proofs, we crucially use metric entropy and Gaussian process arguments from M-estimation, which were previously used to obtain convergence rates for 'direct' nonparametric regression [170, 169]. Thereafter, the convergence rates for $\|\hat{f} - f_0\|_{L^2}$ are proved by use of suitable *stability estimates* of the inverse problem, that is, local Lipschitz estimates for the inverse map G^{-1}. In the divergence form model, such stability estimates had previously been considered in the literature, e.g., [149, 24, 102], and in Lemma 3.5.5 we prove a new quantitative $L^2(\mathcal{O})$-$H^2(\mathcal{O})$ stability estimate building on ideas from [102].

Note that our approach circumvents source conditions which existing results for general non-linear inverse problems frequently rely on see, e.g., [64, 20, 21] and also the discussion in Section 3.1. For papers which have employed a similar high-level strategy of first solving the 'regression problem' and thereafter employing stability estimates, see, e.g., [177, 131].

We conclude by mentioning an interesting unresolved question, namely to determine whether the convergence rate $E_{f_0}\|\hat{f} - f_0\|_{L^2} \lesssim \varepsilon^{\frac{2\alpha-2}{2\alpha+2+d}}$ in (1.30) is minimax-optimal. In fact, the latter is the minimax rate for a 2-ill posed problem with $(\alpha+1)$-regular 'forward' regression functions. In our case this rate arises due to the use of a '1-smoothing' L^2-$(H^1)^*$ Lipschitz estimate for G and a '2-degenerate' L^2-H^2 stability estimate for G^{-1}. Whether an improved L^2-H^1 stability estimate can be obtained is a question for future research.

1.5.3 Local asymptotic normality of low frequency diffusion models

In the fourth and final chapter of this thesis, we consider nonparametric models for diffusion processes $(X_t : t \geq 0)$ which were introduced in Section 1.4, where we will make the simplifying assumption that $\sigma \equiv 1$ is constant and known. Hence the goal is to recover the *unknown drift function* b from *low-frequency* samples $X^{(n)} = (X_0, X_\Delta, \dots, X_{n\Delta})$ of the diffusion process

$$dX_t = b(X_t)dt + \sqrt{2}dW_t, \quad t \geq 0, \tag{1.32}$$

where $(W_t : t \geq 0)$ is a standard Brownian motion and $\Delta > 0$ is the *fixed* distance between measurement times. Specifically, we will follow [75, 134] in studying a version of the model with reflection at the boundary points $\{0, 1\}$, see Section 4.2.1 for the precise definition.

The main aim of the chapter is to establish a *local asymptotic normality* (in short, LAN) property for the low frequency diffusion model. The primary motivation to study the LAN property is that it constitutes a first step towards understanding the 'efficiency' of inference procedures in diffusion models, beyond the minimax estimation [75] and Bayesian contraction rate results [134] discussed in Section 1.4. Specifically, the potential statistical applications in mind are to study *semiparametric efficiency bounds* (see, e.g., Chapter 25 in [172] for a general exposition) as well as infinite-dimensional *Bernstein-von Mises theorems* akin to [33, 34, 133], which we briefly discussed in Section 1.2.3. However, in order to prove the latter, there remain highly non-trivial challenges which have not yet been overcome, see also Section 4.2.4.

Suppose that $d\mu_{init}$ denotes the Lebesgue density of some initial distribution such that $X_0 \sim \mu_{init}$ (independently of $(W_t : t \geq 0)$), and let $p_{\Delta,b}(\cdot, \cdot)$ denote the transition densities of

$(X_t : t \geq 0)$ at time Δ. Then, the log-likelihood function of the sample $X^{(n)}$ is given by

$$\ell_n(b) = \log d\mu_{init}(X_0) + \sum_{i=1}^{n} \log p_{\Delta,b}(X_{(i-1)\Delta}, X_{i\Delta}).$$

The LAN property asserts that the log-likelihood ratios satisfy a quadratic expansion of a specific form – in our setting this will mean that for all sufficiently regular functions $b, h : [0,1] \to \mathbb{R}$,

$$\ell_n(b+h) - \ell_n(b) = \Delta_{n,h} - \frac{1}{2}\|h\|_{LAN}^2 + o_{P_b}(1), \tag{1.33}$$

where P_b denotes the law of $X^{(n)}$ under some drift function b, $\Delta_{n,h}$ are random vectors satisfying $\Delta_{n,h} \xrightarrow{n\to\infty} N(0, \|h\|_{LAN}^2)$ in distribution (under P_b) and $\|\cdot\|_{LAN}$ is a suitable norm, see Section 4.2.

In parametric models with i.i.d. data, one can establish a LAN expansion of the form (1.33) under simple differentiability assumptions, see Chapter 7 of [172]. In the diffusion setting, the LAN expansion has to be proven by different means, due to the non-i.i.d. nature of the Markov chain data. In parametric models for diffusion processes observed at *high frequency* (cf. Section 1.4) the LAN property was shown by Gobet [74] by use of Malliavin calculus. For the low-frequency setting here, our two main proof ingredients are as follows. Firstly, we require differentiability properties of the transition densities $b \mapsto p_{\Delta,b}(x,y)$ allowing to form a second-order Taylor expansion of $\ell_n(b)$ with suitable control over remainder terms. Secondly, we require limit theorems for Markov chains which ensure that the first and second order terms in the Taylor expansion respectively converge to the desired expressions in (1.33). Such limit theorems have been established previously, see, e.g., [29] – the main work of Chapter 4 thus lies in establishing the differentiability properties. Since $p_{\Delta,b}(\cdot,\cdot)$ does not admit a closed-form expression in terms of b, we resort to an implicit characterisation of $p_{\Delta,b}(\cdot,\cdot)$ as fundamental solutions for a natural parabolic PDE corresponding to $(X_t : t \geq 0)$. The regularity properties of $p_{\Delta,b}(\cdot,\cdot)$ are then obtained by means of perturbation arguments and using regularity theory for parabolic equations [116].

Chapter 2

On polynomial-time computation of high-dimensional posterior measures

This chapter studies the problem of generating random samples from high-dimensional posterior distributions. The main results consist of non-asymptotic computational guarantees for Langevin-type MCMC algorithms which scale polynomially in key quantities such as the dimension of the model, the desired precision level, and the number of available statistical measurements. As a direct consequence, it is shown that posterior mean vectors as well as optimisation based maximum a posteriori (MAP) estimates are computable in polynomial time, with high probability under the distribution of the data. These results are complemented by statistical guarantees for recovery of the ground truth parameter generating the data.

The theory is first developed in a general high-dimensional non-linear regression setting (with Gaussian process priors) where posterior measures are not necessarily log-concave, employing a set of local 'geometric' assumptions on the parameter space, and assuming that a good initialiser of the algorithm is available. Thereafter we derive mixing bounds for the non-linear PDE model with the steady-state Schrödinger equation which we encountered already in Section 1.1.

2.1 Introduction

Markov chain Monte Carlo (MCMC) type algorithms are a key methodology in computational mathematics and statistics. The main idea is to generate a Markov chain $(\vartheta_k : k \in \mathbb{N})$ whose laws $\mathcal{L}(\vartheta_k)$ on $\mathbb{R}^D$ approximate its invariant measure. In Bayesian inference the relevant invariant measure has a probability density of the form

$$\pi(\theta|Z^{(N)}) \propto e^{\ell_N(\theta)}\pi(\theta), \ \ \theta \in \mathbb{R}^D. \tag{2.1}$$

Here π is a *prior density function* for a parameter $\theta \in \mathbb{R}^D$ and the map $\ell_N : \mathbb{R}^D \to \mathbb{R}$ is the 'data-log-likelihood' based on N observations $Z^{(N)}$ from some statistical model, so that $\pi(\cdot|Z^{(N)})$ is the density of the Bayesian *posterior probability distribution* on $\mathbb{R}^D$ arising from the observations.

It can be challenging to give performance guarantees for MCMC algorithms in the increasingly complex and high-dimensional statistical models relevant in contemporary data science. By 'high-dimensional' we mean that the model dimension D may be large (e.g., proportional to a power of N). Without any further assumptions accurate sampling from $\pi(\cdot|Z^{(N)})$ in high dimensions can then be expected to be intractable (see below for more discussion). For MCMC methods the computational hardness typically manifests itself in an *exponential* dependence in D or N of the 'mixing time' of the Markov chain $(\vartheta_k : k \in \mathbb{N})$ towards its equilibrium measure (2.1).

In this work we develop mathematical techniques which allow to overcome such computational hardness barriers. We consider diffusion-based MCMC algorithms targeting the Gibbs-type measure with density $\pi(\cdot|Z^{(N)})$ from (2.1) in a non-linear and high-dimensional setting. The prior π will be assumed to be Gaussian – the main challenge thus arises from the non-convexity of $-\ell_N$. We will show how local geometric properties of the statistical model can be combined with recent developments in Bayesian nonparametric statistics [131, 126] and the non-asymptotic theory of Langevin algorithms [48, 57, 58] to justify the '*polynomial time*' feasibility of such sampling methods.

While the approach is general, it crucially takes advantage of the particular geometric structure of the statistical model at hand. In a large class of high-dimensional non-linear inference problems arising throughout applied mathematics, such structure is described by *partial differential equations* (PDEs). Examples that come to mind are inverse and data assimilation problems, and in particular since influential work by A. Stuart [155], MCMC-based Bayesian methodology is frequently used in such settings, especially for the task of uncertainty quantification. We refer the reader to [93], [94], [78], [32], [105], [45], [155], [119], [44], [147], [53], [8] and the references therein. A main contribution of this work is to demonstrate the feasibility of our proof strategy in a (for such PDE problems) prototypical non-linear example where the parameter θ models the potential in a steady-state Schrödinger equation. This PDE arises in various applications such as photo-acoustics, e.g., [10], and provides a suitable framework to lay out the main mathematical ideas underpinning our proofs.

2.1.1 Basic setting and contributions

To summarise our key results we now introduce a more concrete setting. For $\mathcal{O}$ a bounded subset of $\mathbb{R}^d, d \in \mathbb{N}$, and Θ some parameter space, consider a family of 'regression' functions $\{\mathcal{G}(\theta) : \theta \in \Theta\} \subset L^2(\mathcal{O})$, where $L^2(\mathcal{O})$ denotes the usual Lebesgue space $L^2(\mathcal{O})$ of square-integrable functions. This induces a 'forward map'

$$\mathcal{G} : \Theta \to L^2(\mathcal{O}), \tag{2.2}$$

and we suppose that N observations $Z^{(N)} = (Y_i, X_i : i = 1, \dots N)$ arising via

$$Y_i = \mathcal{G}(\theta)(X_i) + \varepsilon_i, \qquad i = 1, ..., N, \tag{2.3}$$

are given, where $\varepsilon_i \sim N(0,1)$ are independent noise variables, and design variables X_i are drawn uniformly at random from the domain $\mathcal{O}$ (independently of ε_i). While natural parameter spaces Θ can be infinite-dimensional, in numerical practice a D-dimensional discretisation of Θ is employed, where D can possibly be large. The log-likelihood function of the data (Y_i, X_i) then equals, up to additive constants, the usual least squares criterion

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \big[Y_i - \mathcal{G}(\theta)(X_i)\big]^2, \quad \theta \in \mathbb{R}^D. \tag{2.4}$$

The aim is to recover θ from $Z^{(N)}$. A wide-spread practice in statistical science is to employ Gaussian (process) priors Π with multivariate normal probability densities π on $\mathbb{R}^D$; from a numerical point of view the Bayesian approach to inference in such problems is then precisely concerned with (approximate) evaluation of the posterior measure (2.1).

As discussed above, in important physical applications the forward map $\mathcal{G}$ is described implicitly by a *partial differential equation*. For example suppose that $\mathcal{G}(\theta) = u_{f_\theta}$ arises as the solution $u = u_{f_\theta}$ to the following elliptic boundary value problem for a Schrödinger equation

$$\begin{cases} \frac{1}{2}\Delta u - f_\theta u = 0 \text{ on } \mathcal{O}, \\ u = g \text{ on } \partial O, \end{cases} \tag{2.5}$$

with a suitable parameterisation $\theta \mapsto f_\theta > 0$, $\theta \in \mathbb{R}^D$ (see (2.17) below for details). In such cases, the map $\mathcal{G}$ is non-linear and $-\ell_N(\theta)$ is not convex. The probability measure with density $\pi(\cdot|Z^{(N)})$ given in (2.1) may then be highly complex to evaluate in a high-dimensional setting, with computational cost scaling exponentially as $D \to \infty$. For instance, complexity theory for high-dimensional numerical integration (see [139, 140] for general references) implies that computing the integral of a D-dimensional real-valued Lipschitz function – such as the normalising factor implicit in (2.1) – by a deterministic algorithm has worst case cost scaling as $D^{D/5}$ [156, 85]. Relaxing a worst case analysis, Monte Carlo methods can in principle obtain dimension-free guarantees (with high probability under the randomisation scheme). However, a curse of dimensionality may persist as one typically is only able to sample *approximately* from the target measure, and since the approximation error incurred, e.g., by the mixing time of a Markov chain, could scale exponentially in dimension. The references [15, 19], [14], [146], [181, 117] discuss this issue in a variety of contexts. In addition, since the distribution becomes increasingly 'spiked' as the statistical information increases (i.e., $N \to \infty$), commonly used iterative algorithms can take an exponential in N time to exit neighbourhoods of local optima of the posterior surface $\pi(\cdot|Z^{(N)})$ (e.g., [61], Example 4).

In light of the preceding discussion one may ask whether the approximate calculation of basic aspects of $\pi(\cdot|Z^{(N)})$ – such as its mean vector (expected value), real-valued functionals $\int_{\mathbb{R}^D} H(\theta)\pi(\theta|Z^{(N)})d\theta$, or mode – is feasible at a computational cost which grows at most *polynomially in* D, N and the desired (inverse) precision level. Very few rigorous results providing even just partial such guarantees appear to be available. The notable exception Hairer, Stuart and Vollmer [80] along with some other important references will be discussed below.

Let us describe the scope of the methods to be developed in this article in the problem of approximate computation of the high-dimensional *posterior mean vector* in the PDE model (2.5) with the Schrödinger equation. We will require mild regularity assumptions on D, Π and on the ground truth θ_0 generating the data (2.3) – full details can be found in Section 2.2. If Π is a D-dimensional Gaussian process prior with covariance equal to a rescaled inverse Laplacian raised to some large enough power $\alpha \in \mathbb{N}$, if the model dimension grows at most as $D \lesssim N^{d/(2\alpha+d)}$, and if θ_0 is sufficiently well-approximated by its 'discretisation' in $\mathbb{R}^D$ (see (2.28)), we obtain the following main result.

Theorem 2.1.1. *Suppose that data $Z^{(N)} = (Y_i, X_i : i = 1, ..., N)$ arise through (2.3) in the Schrödinger model (2.5) and let $P > 0$. Then, for any precision level $\varepsilon \geq N^{-P}$ there exists a (randomised) algorithm whose output $\hat{\theta}_\varepsilon \in \mathbb{R}^D$ can be computed with computational cost*

$$O(N^{b_1} D^{b_2} \varepsilon^{-b_3}) \quad (b_1, b_2, b_3 > 0), \tag{2.6}$$

and such that with high probability (under the joint law of $Z^{(N)}$ and the randomisation mechanism),

$$\|\hat{\theta}_\varepsilon - E^\Pi[\theta|Z^{(N)}]\|_{\mathbb{R}^D} \leq \varepsilon,$$

where $E^\Pi[\theta|Z^{(N)}] = \int_{\mathbb{R}^D} \theta\pi(\theta|Z^{(N)})d\theta$ denotes the mean vector of the posterior distribution $\Pi(\cdot|Z^{(N)})$ with density (2.1).

We further show in Theorem 2.2.6 that $\hat{\theta}_\varepsilon$ also recovers the ground truth θ_0, within precision ε. The method underlying Theorem 2.1.1 consists of an initialisation step which requires solving a standard convex optimisation problem, followed by iterations (ϑ_k) of a discretised gradient based Langevin-type MCMC algorithm, at each step requiring a single evaluation of $\nabla \ell_N$ (which itself amounts to solving a standard linear elliptic boundary value problem). In particular our results will imply that the posterior mean can be computed by ergodic averages $(1/J)\sum_{k \leq J} \vartheta_k$ along the MCMC chain (after some burn-in time), see Theorem 2.2.5 (which implies Theorem 2.1.1). The laws $\mathcal{L}(\vartheta_k)$ of the iterates (ϑ_k) in fact provide a *global* approximation

$$W_2(\mathcal{L}(\vartheta_k), \Pi(\cdot|Z^{(N)})) \leq \varepsilon, \quad k \geq k_{mix},$$

of the high-dimensional posterior measure on $\mathbb{R}^D$, in Wasserstein-distance W_2. Our explicit convergence guarantees will ensure that both the 'mixing time' k_{mix} and the number of required iterations J to reach precision level ε scales polynomially in D, N, ε^{-1}. Similar statements hold true for the computation of real-valued functionals $\int_{\mathbb{R}^D} H(\theta)\pi(\theta|Z^N)d\theta$ for Lipschitz maps $H : \mathbb{R}^D \to \mathbb{R}$ and of maximum a posteriori (MAP) estimates. See Theorems 2.2.7, 2.2.8 as well as Proposition 2.2.4 for precise statements.

The key idea underlying our proofs is to demonstrate first that, with high probability under the law generating the data $Z^{(N)}$, the target measure $\Pi(\cdot|Z^{(N)})$ from (2.1) is locally log-concave on a region in $\mathbb{R}^D$ where most of its mass concentrates. Then we show that a 'localised' Langevin-type algorithm, when initialised into the region of log-concavity, possesses polynomial time convergence guarantees in 'moderately' high-dimensional models. That sufficiently precise initialisation is possible has to be shown in each problem individually (for the Schrödinger model, see Section 2.5.4). Our proofs provide a template (outlined in Section 2.3) that can be used in principle also in general settings as long as the linearisation $\nabla_\theta G(\theta_0)$ of $\mathcal{G}$ at the ground truth parameter θ_0 satisfies a suitable stability estimate (i.e., a quantitative injectivity property related to the 'Fisher information' operator of the statistical model). We verify this stability property for the Schrödinger equation using elliptic PDE techniques (see Lemma 2.4.7) but our approach may succeed in a variety of other non-linear forward models arising in inverse problems [97, 168, 155, 127], integral X-ray geometry [143, 126, 90], and also in the context of data assimilation and filtering [119, 147]. Further advancing our understanding of the computational complexity of such PDE-constrained high-dimensional inference problems poses a formidable challenge for future research.

2.1.2 Discussion of related literature

Both the statistical and computational aspects of high-dimensional Bayes procedures have been subject of great interest in recent years. Frequentist convergence properties of high- and infinite-dimensional Bayes procedures were intensely studied in the last two decades. For 'direct' statistical models we refer to the recent monograph [69] (and references therein), and in the non-linear (PDE) setting relevant here to [134, 126, 132, 2, 131, 135, 136].

We now discuss some representative papers studying mixing properties of MCMC algorithms in high-dimensional settings, and refer to the references cited in these articles for various further important results.

2.1.2.1 Mixing times for pCN-type algorithms

The important contribution [80] by Hairer, Stuart and Vollmer derives dimension-independent convergence guarantees for the preconditioned Crank-Nicolson (pCN) algorithm, using ergodicity results for infinite-dimensional Markov chains from Hairer, Mattingly and Scheutzow [79]. The task of sampling from a general measure arising from a Gaussian process prior and a general likelihood function $\exp(-\Phi(\theta))$ is considered there. Their results are hence naturally compatible with the setting considered in this thesis, where Φ is given by (2.4), i.e. $\Phi = \Phi_N = \ell_N$ and it is natural to ask (a) whether the bounds from [80] apply to this class of problems and (b) if they apply, how they quantitatively depend on N and model dimension.

The key Assumptions 2.10, 2.11, and 2.13 made in [80] can be summarised as (A) a global lower bound on the acceptance probability of the pCN as well as (B) a (local) Lipschitz continuity requirement on Φ. In non-linear PDE problems, part (B) can usually be verified (e.g., [136]), while part (A) is more challenging: due to the global nature of the assumption, it seems that verification of (A) will typically require bounds for likelihood ratios $\exp(\Phi(\theta) - \Phi(\bar{\theta}))$ with $\theta, \bar{\theta}$ arbitrarily far apart. Of course, in some specific problems an initial bound may be obtained by invoking inequalities like (2.18). However the resulting lower bounds on the acceptance probabilities in the pCN scheme will decrease exponentially in N. We also note that though dimension-independent, the main Theorems 2.12 and 2.14 from [80] remain implicit (non-quantitative) in the relevant quantities from Assumptions (A) and (B); this seems to stem both from the utilised proof techniques, such as considerations regarding level sets of Lyapunov functions (cf. [80], p.2474), as well as the qualitative nature of the key underlying probabilistic weak Harris theorem proved by [79]. Summarising, while it would be very exciting to see the results [80] be extended to yield quantitative bounds which are polynomial in both N, D, serious technical and conceptual innovations seem to be required. In the present context, when exploiting local average curvature of the likelihood surface arising from PDE structure, it appears more promising to investigate gradient based MCMC schemes.

2.1.2.2 Computational guarantees for Langevin-type algorithms

For the important gradient-based class of Langevin Monte Carlo (LMC) algorithms, the first nonasymptotic convergence guarantees which are suited for high-dimensional settings were obtained by Dalalyan [48] for log-concave densities, shortly after to be extended by Durmus and Moulines [57, 58] to closely related cases. Our proofs rely substantially on these convergence results for the strongly log-concave case (see Section 2.5.1 for a review).

Very recently further extensions have emerged, notably [117] and [175], which estabish convergence guarantees assuming that either the density to be sampled from is convex outside of some region, or that the target measure satisfies functional inequalities of log-Sobolev and Poincaré type. However, it appears that both of these results, when applied to (2.4) without any further substantial work, yield bounds that scale exponentially in N. Indeed, the bound in Theorem 1 of [117] evidently depends exponentially on the Lipschitz constant of the gradient $\nabla \ell_N$; and ad hoc verification of assumptions from [175] would utilise the Holley-Stroock perturbation principle [89] (and (2.18)), exhibiting the same exponential dependence. Alternative, more elaborate ways of verifying functional inequalities in this context would be highly interesting, but this is not the approach we take in this thesis.

2.1.2.3 Relationship to Bernstein-von Mises theorems

A key idea in our proofs is to use approximate curvature of $\ell_N(\theta)$ 'near' the ground truth θ_0. On a deeper level this idea is related to the possibility of a Bernstein-von Mises theorem which would establish precise Gaussian ('Laplace-type') approximations to posterior distributions, see [104, 106, 172] for the classical versions of such results in 'low-dimensional' statistical models, and [66, 33–35] for high- or infinite-dimensional versions thereof.

Such an approach is taken by [14] who attempt to exploit the asymptotic 'normality' of the posterior measure to establish bounds on the computation time of MCMC-based posterior sampling, building on seminal work by Lovasz, Simonovits and Vempala [113, 114] on the complexity of general Metropolis-Hastings schemes. While [14] allow potentially for moderately high-dimensional situations (by appealing to high-dimensional Bernstein-von Mises theorems from [66]), their sampling guarantees hold for rescaled posterior measures arising as laws of $\sqrt{N}(\theta-\tilde{\theta})|Z^{(N)}$ where $\tilde{\theta}=\tilde{\theta}(Z^{(N)})$ is an initial 'semi-parametrically efficient centring' of the posterior draws $\theta|Z^{(N)}$ (cf. also Remark 2.2.10 below). In our setting such a centring is not generally available (in fact that one can compute such centrings, such as the posterior mode or mean, in polynomial time, is a main aim of our analysis). The setting in [14] thus appears somewhat unnatural for the problems studied here, also because the conditions there do not appear to permit Gaussian priors.

For the Schrödinger equation example considered in this thesis, Bernstein-von Mises theorems were obtained in the recent paper [131] (in a slightly different but closely related measurement setting). While we follow [131] in using elliptic PDE theory to quantify the amount of curvature expressed in the 'limiting information operator' arising from the Schrödinger model, our proofs are in fact not based on an asymptotic Gaussian approximation of the posterior distribution. Rather we use tools from high-dimensional probability to deduce local curvature bounds directly for the likelihood surface, and then show that the posterior measure is approximated, in Wasserstein distance, by a globally log-concave measure that concentrates around the posterior mode (see Theorem 2.4.14). While one can think of this as a 'non-asymptotic' version of a Bernstein-von Mises theorem, the underlying techniques do not require the full inversion of the information operator (as in [131] or also in [125, 135]), but solely rely on a 'stability estimate' for the local linearisation of the forward map, and hence are likely to apply to a larger class of PDEs. A further key advantage of our approach is that we do not require the initialiser for the algorithm to be a 'semi-parametrically efficient' estimator (as [14] do), instead only a sufficiently fast 'nonparametric' convergence rate is required, which substantially increases the class of admissible initialisation strategies.

2.1.2.4 Regularisation/optimisation literature

Regularisation-driven optimisation methods have been studied for a long time in applied mathematics, see for instance the monographs [63, 95]. In the setting of non-linear operator equations in Hilbert spaces and with deterministic noise, 'local' convergence guarantees for iterative (gradient or 'Landweber') methods have been obtained in [81, 95], assuming that optimisation is performed over a (sufficiently small) neighbourhood of a maximum. The proof techniques underlying our main results allow as well to derive guarantees for gradient descent algorithms targeting, for instance, maximum a posteriori (MAP) estimates, see Section 2.2.2.5. Specifically, in Theorem 2.2.8, global convergence guarantees for the computation of MAP estimates over a high-dimensional discretisation space are given, in our genuine statistical framework, paralleling our main results for Langevin sampling methods, which can be regarded as randomised versions of classical gradient methods. A main attraction of studying such randomised algorithms, and more generally of solving the problem of Bayesian computation, is of course that one can access *entire* posterior distributions, which is required for quantifying the statistical uncertainty in the reconstruction provided by point estimates such as posterior mean or mode.

2.1.3 Notations and conventions

Throughout, N will denote the number of observations in (2.3) and D will denote the dimension of the model from (2.4). For a real-valued function $f : \mathbb{R}^D \to \mathbb{R}$, its gradient and Hessian are denoted by ∇f and $\nabla^2 f$, respectively, while $\Delta = \nabla^T \nabla$ denotes the Laplace operator. For any matrix $A \in \mathbb{R}^{D\times D}$, we denote the operator norm by

$$\|A\|_{op} := \sup_{\psi: \|\psi\|_{\mathbb{R}^D} \leq 1} \|A\psi\|_{\mathbb{R}^D}.$$

If A is positive definite and symmetric, then we denote the minimal and maximal eigenvalues of A by $\lambda_{min}(A)$ and $\lambda_{max}(A)$ respectively, with condition number $\kappa(A) := \lambda_{max}(A)/\lambda_{min}(A)$. The Euclidean norm on $\mathbb{R}^D$ will be denoted by $\|\cdot\|_{\mathbb{R}^D}$. The space $\ell_2(\mathbb{N})$ denotes the usual sequence space of square-summable sequence $(a_n : n \in \mathbb{N})$, normed by $\|\cdot\|_{\ell_2}$. For any $a \in \mathbb{R}$, we write $a_+ = \min\{a, 0\}$. Throughout, $\lesssim, \gtrsim, \simeq$ will denote (in-)equalities up to multiplicative constants.

For a Borel subset $\mathcal{O} \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, let $L^p = L^p(\mathcal{O})$ be the usual spaces of functions endowed with the norm $\|\cdot\|_{L^p}^p = \int_{\mathcal{O}} |h(x)|^p dx$, where dx is Lebesgue measure. The usual $L^2(\mathcal{O})$ inner product is denoted by $\langle \cdot, \cdot \rangle_{L^2(\mathcal{O})}$. If $\mathcal{O}$ is a smooth domain in $\mathbb{R}^d$, then $C(\mathcal{O})$

denotes the space of bounded continuous functions $h : \mathcal{O} \to \mathbb{R}$ equipped with the supremum norm $\|\cdot\|_\infty$ and $C^\alpha(\mathcal{O}), \alpha \in \mathbb{N}$, denote the usual spaces of α-times continuously differentiable functions on $\mathcal{O}$ with bounded derivatives. Likewise we denote by $H^\alpha(\mathcal{O})$ the usual order-α Sobolev spaces of weakly differentiable functions with square integrable partial derivatives up to order $\alpha \in \mathbb{N}$, and this definition extends to positive $\alpha \notin \mathbb{N}$ by interpolation [161]. We also define $(H_0^2(\mathcal{O}))^*$ as the topological dual space of

$$(H_0^2(\mathcal{O}) = \{h \in H^2(\mathcal{O}) : tr(h) = 0\}, \|\cdot\|_{H^2(\mathcal{O})}),$$

where $tr(\cdot)$ denotes the usual trace operator on $\mathcal{O}$. We will repeatedly use the inequalities

$$\|gh\|_{H^\alpha} \leq c(\alpha, \mathcal{O})\|g\|_{H^\alpha}\|h\|_{H^\alpha}, \qquad \alpha > d/2, \tag{2.7}$$

$$\|h\|_{H^\beta} \leq c(\beta, \alpha, \mathcal{O})\|h\|_{L^2}^{(\alpha-\beta)/\alpha}\|h\|_{H^\alpha}^{\beta/\alpha}, \quad 0 \leq \beta \leq \alpha \tag{2.8}$$

for $g, h \in H^\alpha$, see, e.g., [109]. For Borel probability measures μ_1, μ_2 on $\mathbb{R}^D$ with finite second moments we define the Wasserstein distance

$$W_2^2(\mu_1, \mu_2) = \inf_{\nu \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^D \times \mathbb{R}^D} \|\theta - \vartheta\|_{\mathbb{R}^D}^2 d\nu(\theta, \vartheta), \tag{2.9}$$

where $\Gamma(\mu_1, \mu_2)$ is the set of all 'couplings' of μ_1 and μ_2, i.e., probability measures ν on $\mathbb{R}^D \times \mathbb{R}^D$ satisfying $\nu(A \times \mathbb{R}^D) = \mu_1(\mathbb{R}^D)$ and $\nu(\mathbb{R}^D \times A) = \mu_2(\mathbb{R}^D)$ for all Borel sets $A \subseteq \mathbb{R}^D$ (cf. [176]). Finally we say that a map $H : \mathbb{R}^D \to \mathbb{R}$ is Lipschitz if it has finite Lipschitz norm

$$\|H\|_{Lip} := \sup_{x \neq y, x, y \in \mathbb{R}^D} \frac{|H(x) - H(y)|}{\|x - y\|_{\mathbb{R}^D}}. \tag{2.10}$$

2.2 Main results for the Schrödinger model

Our object of study in this section is a nonlinear forward model arising with a (steady state) Schrödinger equation. Throughout, let $\mathcal{O} \subset \mathbb{R}^d$ be a bounded domain with smooth boundary $\partial\mathcal{O}$. For convenience we will restrict to $d \leq 3$, dimensions $d \geq 4$ could be considered as well at the expense of further technicalities. Moreover, without loss of generality we assume $vol(\mathcal{O}) = 1$.

Suppose that $g \in C^\infty(\partial\mathcal{O})$ is a given function prescribing boundary values $g \geq g_{min} > 0$ on ∂O. For an 'attenuation potential' $f \in H^\alpha(\mathcal{O})$, consider solutions $u = u_f$ of the PDE

$$\begin{cases} \frac{1}{2}\Delta u - fu = 0 \quad \text{on } \mathcal{O}, \\ u = g \quad \text{on } \partial\mathcal{O}. \end{cases} \tag{2.11}$$

If $\alpha > d/2$ and $f \geq 0$ then standard theory for elliptic PDEs (see Chapter 6 of [71] or Chapter 4 in [42]) implies that a unique classical solution $u_f \in C^2(\mathcal{O}) \cap C(\bar{\mathcal{O}})$ to the Schrödinger equation (2.11) exists. The non-linearity of the map $f \mapsto u_f$ becomes apparent from the classical Feynman-Kac formula (e.g., Theorem 4.7 in [42])

$$u_f(x) = u_{f,g}(x) = E^x \left[g(X_{\tau_\mathcal{O}}) e^{-\int_0^{\tau_\mathcal{O}} f(X_s)ds} \right], \ x \in \mathcal{O}, \tag{2.12}$$

where $(X_s : s \geq 0)$ is a d-dimensional Brownian motion started at x with exit time $\tau_\mathcal{O}$ from $\mathcal{O}$. This PDE appears in various settings in applied mathematics; for example an application to photo-acoustics is discussed in Section 3 in [10].

2.2.1 Bayesian inference with Gaussian process priors

2.2.1.1 The Dirichlet-Laplacian and Gaussian random fields

In Bayesian statistics popular choices of prior probability measures arise from Gaussian random fields whose covariance kernels are related to the Laplace operator Δ, see, e.g., Section 2.4 in [155] and also Example 11.8 in [69] (where the closely related 'Whittle-Matérn' processes are considered).

Let $g_\mathcal{O}$ denote the symmetric Green kernel of the Dirichlet Laplacian on $\mathcal{O}$, which for $\psi \in L^2(\mathcal{O})$ describes the unique solution $v = \mathbb{V}[\psi] = \int_\mathcal{O} g_\mathcal{O}(\cdot, y)\psi(y)dy \in H_0^2(\mathcal{O})$ of the Poisson equation $\Delta v/2 = \psi$ on $\mathcal{O}$. By standard results (Section 5.A in [161]) the compact $\langle \cdot, \cdot \rangle_{L^2(\mathcal{O})}$-self-adjoint operator $\mathbb{V}$ has eigenfunctions $(e_k : k \in \mathbb{N})$ forming an orthonormal basis of $L^2(\mathcal{O})$ such that $\mathbb{V}[\psi] = \sum_{k=1}^\infty \mu_k \langle e_k, \psi \rangle_{L^2(\mathcal{O})} e_k$, with (negative) eigenvalues μ_k satisfying the Weyl asymptotics (e.g., Corollary 8.3.5 in [162])

$$\lambda_k = \frac{1}{|\mu_k|} \simeq k^{2/d} \ \text{ as } k \to \infty, \ \ 0 < \lambda_k < \lambda_{k+1}, \ \ k \in \mathbb{N}. \tag{2.13}$$

The 'spectrally defined' Sobolev-type spaces $\mathcal{H}_\alpha = \{F \in L^2(\mathcal{O}) : \sum_{k=1}^\infty \lambda_k^\alpha \langle F, e_k \rangle_{L^2(\mathcal{O})}^2 < \infty\}$ are isomorphic to corresponding Hilbert sequence spaces

$$h^\alpha := \{\theta \in \ell_2(\mathbb{N}) : \|\theta\|_{h^\alpha}^2 = \sum_{k=1}^\infty \lambda_k^\alpha \theta_k^2 < \infty\}, \quad h^0 =: \ell_2(\mathbb{N}).$$

One shows that $\mathcal{H}_\alpha$ is a closed subspace of $H^\alpha(\mathcal{O})$ and that the sequence norm $\|\cdot\|_{h^\alpha}$ is equivalent to $\|\cdot\|_{H^\alpha(\mathcal{O})}$ on $\mathcal{H}_\alpha$. For α even, this follows from the usual isomorphism theorems for the $\alpha/2$-fold application of the inverse Dirichlet-Laplacian, and extends to general α by interpolation, see Section 5.A in [161]. One also shows that any $F \in H^\alpha(\mathcal{O})$ supported strictly inside of $\mathcal{O}$ belongs to $\mathcal{H}_\alpha$.

A centred Gaussian random field $\mathcal{M}_\alpha$ on $\mathcal{O}$ can be defined by the infinite random series

$$\mathcal{M}_\alpha(x) = \sum_{k=1}^{\infty} \lambda_k^{-\alpha/2} g_k e_k(x), \; x \in \mathcal{O}, \quad g_k \sim^{i.i.d.} N(0,1). \tag{2.14}$$

For $\alpha > d/2$ one shows that $\mathcal{M}_\alpha$ defines a Gaussian Borel random variable in $C(\mathcal{O}) \cap \{h \text{ uniformly continuous} : h = 0 \text{ on } \partial\mathcal{O}\}$ with reproducing kernel Hilbert space equal to $\mathcal{H}_\alpha$ (see Example 2.6.15 in [72]), thus providing natural priors for α-regular functions vanishing at $\partial\mathcal{O}$. Such Dirichlet boundary conditions could be replaced by Neumann conditions at the expense of minor changes (see p.473 in [161]). We finally note that our techniques in principle may extend to other classes of priors such as exponential Besov-type priors considered in [105], but we focus our development here on the most commonly used class of α-regular Gaussian process priors.

2.2.1.2 Re-parameterisation, regular link functions, and forward map

To use Gaussian random fields such as $\mathcal{M}_\alpha$ to model a potential $f \geq 0$ featuring in the Schrödinger equation (2.11), we need to enforce positivity by use of a 'link function' Φ. While $\Phi = \exp$ would appear natural, it will be convenient (following [136]) to choose a function that does not grow exponentially towards ∞.

Definition 2.2.1 (Regular link function)**.** *Let $K_{min} \in [0, \infty)$. We say that $\Phi : \mathbb{R} \to (K_{min}, \infty)$ is a regular link function if it is bijective, smooth, strictly increasing (i.e. $\Phi' > 0$ on $\mathbb{R}$) and if for any $k \geq 1$, the k-th derivative of Φ satisfies* $\sup_{x\in\mathbb{R}} |\Phi^{(k)}(x)| < \infty$.

For a simple example of a regular link function Φ, see e.g. Example 3.2 of [136]. To ease notation, we denote the composition operator associated to Φ by

$$\Phi^* : L^2(\mathcal{O}) \to L^2(\mathcal{O}), \qquad F \mapsto \Phi \circ F = \Phi^*(F). \tag{2.15}$$

Now to describe a natural parameter space for f, we will first expand functions $F \in L^2(\mathcal{O})$ in the orthonormal basis from Section 2.2.1.1,

$$F = F_\theta = \sum_{k=1}^{\infty} \theta_k e_k, \quad (\theta_k : k = 1, 2, \dots) \in \ell_2(\mathbb{N}), \tag{2.16}$$

and denote by $\Psi(\theta) = F_\theta$ the map $\Psi : \ell_2(\mathbb{N}) \to L^2(\mathcal{O})$ that associates to the vector θ the 'Fourier' series of F_θ. We then apply a regular link function Φ to F_θ and set $f_\theta := \Phi \circ F_\theta$. For $\alpha > d/2$, one shows (see (2.178) below) that $F_\theta \in H^\alpha(\mathcal{O})$ implies $f_\theta \in H^\alpha(\mathcal{O})$ and hence solutions of the Schrödinger equation (2.11) exist for such f. If we denote the solution map $f \mapsto u_f$ from (2.11) by G, then the overall forward map describing our parametrisation is given by

$$\mathcal{G} : h^\alpha \to L^2(\mathcal{O}), \; \mathcal{G}(\theta) = u_{f_\theta} = [G \circ \Phi^* \circ \Psi](\theta). \tag{2.17}$$

We shall frequently regard $\mathcal{G}$ as a map on the closed linear subspace $\mathbb{R}^D$ of h^α consisting of the first D coefficients $(\theta_1, \ldots, \theta_D)$ of $\theta \in h^\alpha$. Moreover it will be tacitly assumed that a regular link function $\Phi : \mathbb{R} \to (K_{min}, \infty)$, $K_{min} \geq 0$, has been chosen. We also note that the solutions of (2.11) are uniformly bounded by a constant independent of $\theta \in h^\alpha$, specifically

$$\|\mathcal{G}(\theta)\|_\infty = \|u_{f_\theta}\|_\infty \leq \|g\|_\infty, \tag{2.18}$$

as follows from (2.12) and $f_\theta \geq 0$. This 'bounded range' property of $\mathcal{G}$ is relative to the norm employed; for instance the $\|u_{f_\theta}\|_{H^\alpha}$-norms are *not* uniformly bounded in $\theta \in h^\alpha$ for general α.

2.2.1.3 Measurement model, prior, likelihood and posterior

For the forward map $\mathcal{G}$ from (2.17), we now consider the measurement model

$$Y_i = \mathcal{G}(\theta)(X_i) + \varepsilon_i, \quad i = 1, \ldots, N, \quad \varepsilon_i \sim^{i.i.d.} N(0,1), \quad X_i \sim^{i.i.d.} \text{Uniform}(\mathcal{O}). \tag{2.19}$$

The i.i.d. random vectors

$$Z^{(N)} = (Z_i)_{i=1}^N = (Y_i, X_i)_{i=1}^N \tag{2.20}$$

are drawn from a product measure on $(\mathbb{R} \times \mathcal{O})^N$ that we denote by $P_\theta^N = \otimes_{i=1}^N P_\theta$. The coordinate (Lebesgue) densities p_θ of the joint probability density $p_\theta^N = \prod_{i=1}^N p_\theta$ of P_θ^N are of the form

$$p_\theta(y, x) := \frac{1}{\sqrt{2\pi}} \exp\Big\{ -\frac{1}{2}[y - \mathcal{G}(\theta)(x)]^2 \Big\}, \quad y \in \mathbb{R}, x \in \mathcal{O}, \tag{2.21}$$

(recalling $vol(\mathcal{O}) = 1$) and we can define the *log-likelihood function* as

$$\ell_N(\theta) \equiv \log p_\theta^N + N \log \sqrt{2\pi} = -\frac{1}{2} \sum_{i=1}^N \big(Y_i - \mathcal{G}(\theta)(X_i)\big)^2. \tag{2.22}$$

When using Gaussian process prior models in Bayesian statistics, a common discretisation approach is to truncate the ('Karhunen-Loéve' type) expansion of the prior in a suitable basis, cf. [105, 155, 80, 52]. In our context this will mean that we truncate the series defining the random field $\mathcal{M}_\alpha$ in (2.14) at some finite dimension D to be specified. For integer α to be chosen, and recalling the eigenvalues $(\lambda_k : k \in \mathbb{N})$ of the Dirichlet Laplacian from (2.13), we thus consider priors

$$\theta \sim \Pi = \Pi_N \sim N(0, N^{-d/(2\alpha+d)} \Lambda_\alpha^{-1}), \quad \Lambda_\alpha = diag(\lambda_1^\alpha, \ldots, \lambda_D^\alpha), \tag{2.23}$$

supported in the subspace $\mathbb{R}^D$ of h^α consisting of its first D coordinates. The Lebesgue density $d\Pi$ of Π on $\mathbb{R}^D$ will be denoted by π. The posterior measure $\Pi(\cdot|Z^{(N)})$ on $\mathbb{R}^D$ then arises from data $Z^{(N)}$ in (2.19) via Bayes' formula. Writing $\|\theta\|_{h^\alpha} = \|F_\theta\|_{h^\alpha}$, its probability density function of $\Pi(\cdot|Z^{(N)})$ is given by

$$\pi(\theta|Z^{(N)}) \propto e^{\ell_N(\theta)}\pi(\theta) \tag{2.24}$$
$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(Y_i - \mathcal{G}(\theta)(X_i)\right)^2 - \frac{N^{d/(2\alpha+d)}}{2}\|\theta\|_{h^\alpha}^2\right\}, \ \theta \in \mathbb{R}^D.$$

2.2.2 Polynomial time guarantees for Bayesian posterior computation

2.2.2.1 Description of the algorithm

We now describe the Langevin-type algorithm targeting the posterior measure $\Pi(\cdot|Z^{(N)})$. It requires the choice of an initialiser θ_{init} and of constants ϵ, K, γ. Throughout, we use the initialiser $\theta_{init} = \theta_{init}(Z^{(N)}) \in \mathbb{R}^D$ constructed in Theorem 2.5.10 in Section 2.5.4 (computable in $O(N^{b_0})$ polynomially many steps, for some $b_0 > 0$). For $\epsilon > 0$ to be chosen we define the high-dimensional region

$$\hat{\mathcal{B}} = \{\theta \in \mathbb{R}^D : \|\theta - \theta_{init}\|_{\mathbb{R}^D} \le \epsilon D^{-4/d}/2\}. \tag{2.25}$$

We then construct a proxy function $\tilde{\ell}_N : \mathbb{R}^D \to \mathbb{R}$ which agrees on $\hat{\mathcal{B}}$ with the log-likelihood function ℓ_N from (2.22). Specifically, take the cut-off function $\alpha = \alpha_\eta$ from (2.53) and the convex function $g = g_\eta$ from (2.52) with choice $\eta = \epsilon D^{-4/d}$ and $|\cdot|_1 = \|\cdot\|_{\mathbb{R}^D}$. Note that α is compactly supported and identically one on $\hat{\mathcal{B}}$ and that g vanishes on $\hat{\mathcal{B}}$. Then for K to be chosen, $\tilde{\ell}_N$ takes the form

$$\tilde{\ell}_N(\theta) := \alpha(\theta)\ell_N(\theta) - Kg(\theta), \quad \theta \in \mathbb{R}^D. \tag{2.26}$$

This induces a proxy probability measure, correspondingly denoted by $\tilde{\Pi}(\cdot|Z^{(N)})$, with log-density

$$\log \tilde{\pi}(\theta|Z^{(N)}) = \tilde{\ell}_N(\theta) - N^{\frac{d}{2\alpha+d}}\|\theta\|_{h^\alpha}^2/2 + const., \ \theta \in \mathbb{R}^D. \tag{2.27}$$

Note that $\tilde{\pi}(\cdot|Z^{(N)})$ coincides with the posterior density $\pi(\cdot|Z^{(N)})$ on the set $\hat{\mathcal{B}}$ up to a (random) normalising constant. The MCMC scheme we consider is then given in Algorithm 3 and the law of the resulting Markov chain $(\vartheta_k) \in \mathbb{R}^D$ will be denoted by $\mathbf{P}_{\theta_{init}}$.

While the algorithm is related to stochastic optimisation methods based on gradient descent, the diffusivity term is of constant order in k, allowing (ϑ_k) to explore the entire support of the target measure. It coincides with the *unadjusted Langevin algorithm* (see Section 2.5.1) targeting $\pi(\cdot|Z^{(N)})$ as long as the iterates (ϑ_k) stay within the region $\hat{\mathcal{B}} \subset \mathbb{R}^D$ we have initialised to. When (ϑ_k) exits $\hat{\mathcal{B}}$, the Markov chain is forced by the 'proxy' function $\tilde{\ell}_N$ to eventually return to $\hat{\mathcal{B}}$. This procedure is justified since most of the posterior mass will be shown to concentrate on $\hat{\mathcal{B}}$ with high probability under the law of $Z^{(N)}$. [In fact a key step of our proofs is to control the Wasserstein-distance between the measures induced by the

Algorithm 3

Input: Initialiser $\theta_{init} \in \mathbb{R}^D$, convexification parameters $\epsilon, K > 0$, step size $\gamma > 0$, *i.i.d.* sequence $\xi_k \sim N(0, I_{D\times D})$.

Output: Markov chain $\vartheta_1, \dots, \vartheta_k, \dots \in \mathbb{R}^D$.

initialise $\vartheta_0 = \theta_{init}$
for $k = 0, 1, \dots$ **do**
  $\vartheta_{k+1} = \vartheta_k + \gamma \nabla \log \tilde{\pi}(\vartheta_k | Z^{(N)}) + \sqrt{2\gamma}\xi_{k+1}$
return $(\vartheta_k : k = 1, \dots)$

densities $\pi(\cdot|Z^{(N)}), \tilde{\pi}(\cdot|Z^{(N)})$, cf. Theorem 2.4.14.] Note that while the ball in (2.25) shrinks as dimension $D \to \infty$, relative to the step-sizes γ permitted below, $\hat{\mathcal{B}}$ has asymptotically *growing* diameter. The results that follow show that the Markov chain (ϑ_k) nevertheless mixes sufficiently fast to reconstruct the posterior surface on $\hat{\mathcal{B}}$ with arbitrary precision after a polynomial runtime.

To demonstrate the performance of Algorithm 3 in a large N, D scenario, we now make the following specific choices of the key algorithm parameters ϵ, K, γ.

Condition 2.2.2. *Let θ_{init} be the initialiser from Theorem 2.5.10 and suppose that*

$$\epsilon := \frac{1}{\log N}, \quad K := ND^{8/d}(\log N)^3, \quad \gamma \le \frac{1}{ND^{8/d}(\log N)^4}.$$

2.2.2.2 Conditions involving θ_0

The convergence guarantees obtained below hold for moderately high-dimensional models where D is permitted to grow polynomially in N, and under the frequentist assumption that the data $Z^{(N)}$ from (2.19) is generated from a fixed ground truth θ_0 inducing the law $P^N_{\theta_0}$. Note that we do *not* assume that $\theta_0 \in \mathbb{R}^D$, but rather that $\theta_0 \in h^\alpha$ is sufficiently well approximated by its $\ell_2(\mathbb{N})$-projection $\theta_{0,D}$ onto $\mathbb{R}^D$. The precise condition, which is discussed in more detail in Remark 2.2.9 below, reads as follows.

Condition 2.2.3. *For integers $d \le 3$ and $\alpha > 6$, suppose data $Z^{(N)}$ from (2.20) arise in the Schrödinger model (2.19) for some fixed $\theta_0 \in h^\alpha$. Moreover, suppose that $D \in \mathbb{N}$ is such that for some constants $c_0 > 0, 0 < c_0' < 1/2$, and $\theta_{0,D} = ((\theta_0)_1, ..., (\theta_0)_D)$*

$$D \le c_0 N^{d/(2\alpha+d)}, \quad \|\mathcal{G}(\theta_{0,D}) - \mathcal{G}(\theta_0)\|_{L^2(\mathcal{O})} \le c_0' N^{-\alpha/(2\alpha+d)}. \tag{2.28}$$

Though it will be left implicit, the results we obtain in this section depend on θ_0 only through c_0' and an upper bound $S \ge \|\theta_0\|_{h^\alpha}$.

2.2.2.3 Computational guarantees for ergodic MCMC averages

We first present a concentration inequality for ergodic averages along the Markov chain (ϑ_k). Proposition 2.2.4 is non-asymptotic in nature; hence its statement necessarily involves various constants whose dependence on D and N is tracked. Theorems 2.2.5 and 2.2.6 then demonstrate how the desired polynomial time computation guarantees, including Theorem 2.1.1, can be deduced from it.

For 'burn-in' time $J_{in} \in \mathbb{N}$ and MCMC samples $(\vartheta_k : k = J_{in}+1, ..., J_{in}+J)$ from Algorithm 3, define

$$\hat{\pi}^J_{J_{in}}(H) = \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} H(\vartheta_k), \quad H : \mathbb{R}^D \to \mathbb{R}.$$

We also set, for $c_1 > 0$ to be chosen,

$$B(\gamma) := c_1\Big[\gamma D^{(d+24)/d}(\log N)^6 + \gamma^2 N D^{(d+44)/d}(\log N)^{12}\Big] + \exp(-N^{-\frac{d}{2\alpha+d}}). \tag{2.29}$$

The quantity $B(\gamma)$ is an upper bound for the error incurred by the Euler discretisation of the Langevin dynamics (see (2.163) below) and by the 'proxy' construction (2.27).

Proposition 2.2.4. *Assume Condition 2.2.3 is satisfied and consider iterates ϑ_k of the Markov chain from Algorithm 3 with $\theta_{init}, \epsilon, K, \gamma$ satisfying Condition 2.2.2. Then there exist constants $c_1, c_2, ..., c_5 > 0$ such that for all $N \in \mathbb{N}$, any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$, any burn-in period*

$$J_{in} \geq \frac{\log N}{\gamma N D^{-4/d}} \times \log\big(D + B(\gamma)^{-1}\big), \tag{2.30}$$

any $J \in \mathbb{N}$, any $t \geq 2\|H\|_{Lip}\sqrt{B(\gamma)}$ and on events $\mathcal{E}_N$ (measurable subsets of $(\mathbb{R}\times\mathcal{O})^N$) of probability $P^N_{\theta_0}(\mathcal{E}_N) \geq 1 - c_2\exp(-c_3 N^{d/(2\alpha+d)})$,

$$\mathbf{P}_{\theta_{init}}\big(|\hat{\pi}^J_{J_{in}} - E^\Pi(H|Z^{(N)})| \geq t\big) \leq c_5 \exp\Big(-c_4\frac{t^2N^2J\gamma}{D^{8/d}\|H\|^2_{Lip}(1+D^{4/d}/(NJ\gamma))}\Big).$$

The next result concerns computation of the posterior mean vector

$$E^\Pi[\theta|Z^{(N)}] = \int_{\mathbb{R}^D}\theta\pi(\theta|Z^{(N)})d\theta$$

by ergodic averages

$$\bar{\theta}^J_{J_{in}} := \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J}\vartheta_k, \quad J_{in}, J \in \mathbb{N},$$

within prescribed precision level ε. For convenience we assume $\varepsilon \geq N^{-P}$, which is natural in view of the statistical error to be considered in Theorem 2.2.6 below. To this end, we make an explicit choice for the step size parameter

$$\gamma = \gamma_\varepsilon = \min\Big(\frac{\varepsilon^2}{D^{(d+24)/d}}, \frac{\varepsilon}{\sqrt{N} D^{(22+d/2)/d}}, \frac{1}{ND^{8/d}}\Big) \times (\log N)^{-7}. \tag{2.31}$$

Theorem 2.2.5. *Assume Condition 2.2.3 is satisfied. Fix $P > 0$ and let $\varepsilon \geq N^{-P}$. Consider iterates ϑ_k of the Markov chain from Algorithm 3 with $\theta_{init}, \epsilon, K$ satisfying Condition 2.2.2 and with $\gamma = \gamma_\varepsilon$ as in (2.31). Then there exist $c_6, c_7, c_8 > 0$ and at most polynomially growing constants*

$$g_{D,N,\varepsilon} = O(D^{\bar{b}_1} N^{\bar{b}_2} \varepsilon^{-\bar{b}_3}), \quad \bar{b}_1, \bar{b}_2, \bar{b}_3 > 0, \tag{2.32}$$

such that for all $N \in \mathbb{N}$, $J_{in} \geq g_{D,N,\varepsilon}$, $J \in \mathbb{N}$ and on events $\mathcal{E}_N$ of probability $P_{\theta_0}^N(\mathcal{E}_N) \geq 1 - c_7 \exp(-c_8 N^{d/(2\alpha+d)})$,

$$\mathbf{P}_{\theta_{init}}\Big(\|\bar{\theta}^J_{J_{in}} - E^\Pi[\theta|Z^{(N)}]\|_{\mathbb{R}^D} \geq \varepsilon\Big) \leq c_6 D \exp\Big(-\frac{J}{g_{D,N,\varepsilon}}\Big). \tag{2.33}$$

Theorem 2.2.5 implies that for $J_{in} \wedge J \gg g_{D,N,\varepsilon} \times \log D$, one can compute the posterior mean vector within precision $\varepsilon > 0$ with probability as close to one as desired. Using this and Theorem 2.5.10 (whose hypotheses are implied by those of Theorem 2.2.5), we have in particular also proven Theorem 2.1.1. Similar bounds for computation of $E^\Pi(H|Z^{(N)})$ can be obtained as long as $\|H\|_{Lip}$ grows at most polynomially in D.

We conclude this subsection with a result concerning recovery of the actual target of statistical inference, that is, the ground truth θ_0. It combines Theorem 2.2.5 with a statistical rate of convergence of $E^\Pi[\theta|Z^N]$ to θ_0, obtained by adapting recent results from [126] to the present situation.

Theorem 2.2.6. *Consider the setting of Theorem 2.2.5 with $P = \alpha^2/((2\alpha+d)(\alpha+2))$. There exist further constants $c_9, c_{10}, c_{11}, c_{12} > 0$ such that for all $N \in \mathbb{N}$, all $\varepsilon \geq c_{11} N^{-\frac{\alpha}{2\alpha+d}\frac{\alpha}{\alpha+2}}$, with $g_{D,N,\varepsilon}$ from (2.32) and on events $\mathcal{E}_N$ of probability $P_{\theta_0}^N(\mathcal{E}_N) \geq 1 - c_9 \exp(-c_{10} N^{d/(2\alpha+d)})$,*

$$\mathbf{P}_{\theta_{init}}\Big(\|\bar{\theta}^J_{J_{in}} - \theta_0\|_{\ell_2} \geq \varepsilon\Big) \leq c_{12} \exp\Big(-\frac{J}{4 g_{D,N,\varepsilon}}\Big). \tag{2.34}$$

While the statistical minimax-optimal rate towards $\theta_0 \in h^\alpha$ in this problem can be expected to be faster than N^{-P} (see [131]), it appears unclear how to obtain this rate when F_θ is discretised by means of the (for the purposes of the theory developed in this chapter essential) spectral decomposition of the Dirichlet-Laplacian from Section 2.2.1.1. The difficulty arises with the approximation theory of the space $H_c^\alpha(\mathcal{O})$ (equal to the completion of $C_c^\infty(\mathcal{O})$ in $H^\alpha(\mathcal{O})$) and is not discussed further here.

2.2.2.4 Global bounds for posterior approximation in Wasserstein distance

The previous theorems concern the computation of specific posterior characteristics; one may also be interested in *global* mixing properties of the laws $\mathcal{L}(\vartheta_k)$ induced by the Markov chain $(\vartheta_k : k \in \mathbb{N})$ towards the target $\Pi(\cdot|Z^{(N)})$, for instance in the Wasserstein distance from (2.9).

Theorem 2.2.7. *Assume Condition 2.2.3 is satisfied, let $\mathcal{L}(\vartheta_k)$ denote the law of the k-th iterate ϑ_k of the Markov chain from Algorithm 3 with $\theta_{init}, \epsilon, K, \gamma$ satisfying Condition 2.2.2, and let $B(\gamma), c_1$ be as in (2.29). For any $P > 0$ there exist constants $c_1, c_{13}, c_{14}, c_{15}, c_{16} > 0$ such that on events $\mathcal{E}_N$ of probability $P_{\theta_0}^N(\mathcal{E}_N) \geq 1 - c_{13}\exp(-c_{14}N^{d/(2\alpha+d)})$ and for all $N \in \mathbb{N}$, the following holds.*

i) *For any $k \geq 1$,*

$$W_2^2(\mathcal{L}(\vartheta_k), \Pi[\cdot|Z^{(N)}]) \leq c_{15}D^{2\alpha/d}(1 - c_{16}ND^{-4/d}\gamma)_+^k + B(\gamma). \tag{2.35}$$

ii) *For any 'precision level' $\varepsilon \geq N^{-P}$ and for $\gamma = \gamma_\varepsilon$ from (2.31), there exists*

$$k_{mix} = O(N^{\tilde{b}_1}D^{\tilde{b}_2}\varepsilon^{-\tilde{b}_3}), \quad \tilde{b}_1, \tilde{b}_2, \tilde{b}_3 > 0, \tag{2.36}$$

such that for any $k \geq k_{mix}$,

$$W_2(\mathcal{L}(\vartheta_k), \Pi[\cdot|Z^{(N)}]) \leq \varepsilon.$$

The first term on the right hand side of (2.35) characterises the rate of geometric convergence towards equilibrium of (ϑ_k); the factor $ND^{-4/d}\gamma$ can be thought of as a spectral gap of the Markov chain (related to the 'average local curvature' of $\ell_N(\cdot)$ near θ_0 in the Schrödinger model). Choosing $\gamma = \gamma_\varepsilon$ as in (2.31), part ii) further establishes 'polynomial-time' mixing of the MCMC scheme towards the posterior measure.

2.2.2.5 Computation of the MAP estimate

Our techniques also imply the following guarantees for the computation of *maximum a posteriori* (MAP) estimates

$$\hat{\theta}_{MAP} \in \arg\max_{\theta\in\mathbb{R}^D} \pi(\theta|Z^{(N)})$$

by a classical gradient descent method applied to the 'proxy' posterior surface (2.27).

Theorem 2.2.8. *Assume Condition 2.2.3 is satisfied and let θ_{init} denote the initialiser from Theorem 2.5.10. For $k = 0, 1, 2, \dots,$ consider the gradient descent algorithm*

$$\vartheta_0 = \theta_{init}, \quad \vartheta_{k+1} = \vartheta_k + \gamma\nabla\log\tilde{\pi}(\vartheta_k|Z^{(N)}), \quad \gamma = \frac{1}{ND^{8/d}(\log N)^4}.$$

There exist constants $c_{17}, c_{18}, c_{19}, c_{20}, c_{21} > 0$ *such that for all* $N \in \mathbb{N}$ *and on events* $\mathcal{E}_N$ *of probability at least* $P_{\theta_0}^N(\mathcal{E}_N) \geq 1 - c_{17}\exp(-c_{18}N^{d/(2\alpha+d)})$ *we have the following:*

i) *A unique maximiser* $\hat{\theta}_{MAP}$ *of* $\pi(\theta|Z^{(N)})$ *over* $\mathbb{R}^D$ *exists.*

ii) *For all* $k \geq 1$*, we have the geometric convergence*

$$\|\vartheta_k - \hat{\theta}_{MAP}\|_{\mathbb{R}^D}^2 \leq c_{19}D^{4/d}\Big(1 - \frac{c_{20}}{D^{12/d}(\log N)^4}\Big)_+^k.$$

iii) *Finally, we can choose* $k = O(D^{12/d}(\log N)^5)$ *such that*

$$\|\vartheta_k - \theta_0\|_{\ell_2} \leq c_{21}N^{-\frac{\alpha}{2\alpha+d}\frac{\alpha}{\alpha+2}}.$$

2.2.2.6 Remarks

Remark 2.2.9 (About Condition 2.2.3)**.** In principle the upper bound for D required in Condition 2.2.3 could be replaced by general conditions on D (alike those from Lemma 2.3.4) which do not become more stringent as α increases. From a statistical point of view, however, a choice $D \leq c_0 N^{d/(2\alpha+d)}$ is natural as it corresponds to the optimal 'bias-variance' tradeoff underpinning the convergence rate towards $\theta_0 \in h^\alpha$ from Theorem 2.2.6. [In fact, the second requirement in (2.28) can be checked for $\theta_0 \in h^\alpha$ and $D \simeq N^{d/(2\alpha+d)}$, since $\mathcal{G}$ is $\ell_2(\mathbb{N}) - L^2(\mathcal{O})$ Lipschitz.] Moreover, combined with $\alpha > 6$, such a choice of D provides a convenient sufficient condition throughout our proofs: It is used critically when showing (in Theorem 2.4.14) that the proxy posterior measure $\tilde{\Pi}(\cdot|Z^{(N)})$ contracts about a $\|\cdot\|_{\mathbb{R}^D}$-neighbourhood of θ_0 of radius $D^{-4/d}$ on which the Fisher information in the Schrödinger model has a stable behaviour (see (2.116)). It is also required for our initialiser θ_{init} to lie in this neighbourhood (Theorem 2.5.10). While it is conceivable that the condition on α could be weakened (as discussed, e.g., in the next remark), it would come at the expense of considerable further technicalities that we wish to avoid here.

Remark 2.2.10 (Preconditioning and rescaling)**.** Given the 'local' nature of Algorithm 3, one may be interested in sampling from the distribution of $\theta|Z^{(N)}$ by first running an appropriate modification of Algorithm 3 generating samples $(\psi_k : k \geq 0)$ of the rescaled and recentred law ν_N of $A_N^{-1}(\theta - \theta_{init})$ with probability density propotional $d\nu_N(\psi) \propto \pi(\theta_{init} + A_N\psi|Z^{(N)})$, where $A_N \in \mathbb{R}^{D\times D}$ is a sequence of 'preconditioning' matrices, and then setting $\vartheta_k = \theta_{init} + A_N\psi_k$, see, e.g., Section 4.2 in [48]. The techniques underlying our proofs also apply to such preconditioned algorithms by obvious modifications of the surrogate construction (using also that $W_2(\mathcal{L}(\vartheta_k), \Pi(\cdot|Z^{(N)})) \lesssim W_2(\mathcal{L}(\psi_k), \nu_N)$, where the constant in $\lesssim$ depends only polynomially on the eigenvalues of A_N). This may speed up the algorithm (e.g., in terms of explicit constants $b_i, \bar{b}_i, \tilde{b}_i$ in Theorems 2.1.1, 2.2.5 and 2.2.7 respectively), for instance, in the Schrödinger equation it would be natural to choose for A_N the action of the Laplace operator

Δ on $\mathbb{R}^D$ to 'stablise' the curvature bounds in Lemma 2.4.7. However, when investigating the question of existence of polynomial time sampling algorithms, such preconditioning arguments appear less relevant. For instance, for the pCN algorithm discussed in Section 2.1.2, the global likelihood ratios determining the mixing time of the Markov chain obtained in [80] still grow exponentially in N after rescaling. Likewise, for rescaled Langevin algorithms, the 'qualitative' picture of computational hardness (in the context of this thesis) remains unchanged.

2.3 General theory for random design regression

In proving the results from Section 2.2, we will first develop some theory which applies to general nonlinear regression models. We thus consider in this section the measurement model (2.3) for a general forward model $\mathcal{G}$ that satisfies a set of analytic conditions to be detailed below. Let Θ be a (measurable) linear subspace of $\ell_2(\mathbb{N})$ which itself admits a subspace $\mathbb{R}^D \subseteq \Theta$ for some $D \in \mathbb{N}$. Let $\mathcal{O}$ be a Borel subset of $\mathbb{R}^d, d \geq 1$, and consider a model of regression functions $\{\mathcal{G}(\theta) : \theta \in \Theta\}$ via a Borel-measurable forward map $\mathcal{G} : \Theta \to C(\mathcal{O})$. While we regard each $\mathcal{G}(\theta)$ as a continuous real-valued function, the results of this section readily extend to vector or matrix fields over manifolds $\mathcal{O}$, see Remark 2.3.11. Our data is given by $Z_i = (Y_i, X_i)$ arising from

$$Y_i = \mathcal{G}(\theta)(X_i) + \varepsilon_i, \qquad i = 1, ..., N, \tag{2.37}$$

where $X_i \sim^{i.i.d.} P^X$, P^X a Borel probability measure on $\mathcal{O}$, and where $\varepsilon_i \sim^{i.i.d.} N(0,1)$, independently of the X_i's. We write $Z^{(N)} = (Z_1, ..., Z_N)$ for the full data vector with joint distribution $P_\theta^N = \otimes_{i=1}^N P_\theta$ on $(\mathbb{R} \times \mathcal{O})^N$, with expectation operator $E_\theta^N = \otimes_{i=1}^N E_\theta$. Then the log-likelihood functions of the data $Z^{(N)}$ and of a single observation $Z = (Y, X) \sim P_\theta$ are given by

$$\ell_N(\theta) \equiv \ell_N(\theta, Z^{(N)}) = -\frac{1}{2}\sum_{i=1}^N [Y_i - \mathcal{G}(\theta)(X_i)]^2, \quad \ell(\theta) \equiv \ell(\theta, Z) = -\frac{1}{2}[Y - \mathcal{G}(\theta)(X)]^2, \tag{2.38}$$

respectively. If we regard these maps as being defined on $\mathbb{R}^D \subseteq \Theta$, and if Π is a Gaussian prior Π supported in $\mathbb{R}^D$, then we obtain the posterior measure $\Pi(\cdot|Z^{(N)})$ with probability density $\pi(\cdot|Z^{(N)})$ on $\mathbb{R}^D$ as in (2.24).

The main results of this section are Theorems 2.3.7 and 2.3.8, providing convergence guarantees for a Langevin sampling method for the posterior distribution that depend polynomially on model dimension D and number N of measurements, and which hold on an *event* (i.e., a measurable subset $\mathcal{E}$ of the sample space $(\mathbb{R} \times \mathcal{O})^N$ supporting the data $Z^{(N)}$) of the form

$$\mathcal{E} := \mathcal{E}_{conv} \cap \mathcal{E}_{init} \cap \mathcal{E}_{wass}.$$

On $\mathcal{E}_{conv}$ the negative log-likelihood $-\ell_N(\theta)$ will be strongly convex in some region $\mathcal{B} \subseteq \mathbb{R}^D$, while $\mathcal{E}_{init}$ is the event that allows one to initialise the method at some (data-driven) $\theta_{init} = \theta_{init}(Z^{(N)})$ in that set $\mathcal{B}$. Finally, intersection with $\mathcal{E}_{wass}$ further guarantees that the posterior measure $\Pi(\cdot|Z^{(N)})$ is close in Wasserstein distance to a *globally* log-concave surrogate probability measure $\tilde{\Pi}(\cdot|Z^{(N)})$ which locally coincides with $\Pi(\cdot|Z^{(N)})$ up to proportionality factors. In applying the results of this section to a concrete sampling problem, one needs to show that all the events $\mathcal{E}_{conv}, \mathcal{E}_{init}, \mathcal{E}_{wass}$ have sufficiently high frequentist $P_{\theta_0}^N$-probability, where θ_0 is the ground truth parameter generating data (2.37). For the event $\mathcal{E}_{conv}$ we provide a generic method in Lemma 2.3.4, based on a stability estimate for the linearisation of the map $\mathcal{G}$ combined with high-dimensional concentration of measure techniques. The events $\mathcal{E}_{init}$ and $\mathcal{E}_{wass}$ are somewhat more specific to a given problem, see Remark 2.3.10 for more discussion.

We will assume the set $\mathcal{B} \subseteq \mathbb{R}^D$ of local convexity to be of *ellipsoidal* form.

Definition 2.3.1. *A norm $|\cdot|$ on $\mathbb{R}^D$ is called ellipsoidal if there exists a positive definite, symmetric matrix $M \in \mathbb{R}^{D\times D}$ such that $|\theta|^2 = \theta^T M\theta$ for any $\theta \in \mathbb{R}^D$.*

Throughout this section, for some centring $\theta^* \in \mathbb{R}^D$, scalar $\eta > 0$ and ellipsoidal norm $|\cdot|_1$ with associated matrix M, let $\mathcal{B}$ denote the open subset of $\mathbb{R}^D$ given by

$$\mathcal{B} := \{\theta \in \mathbb{R}^D : |\theta - \theta^*|_1 < \eta\}. \tag{2.39}$$

One may think of θ^* as the projection of θ_0 onto $\mathbb{R}^D$, but at this stage this is not necessary. While for the Schrödinger model with $d \leq 3$ we can choose $|\cdot|_1 = \|\cdot\|_{\mathbb{R}^D}$, in general (e.g., when $d \geq 4$ or in other non-linear problems) it may be convenient to consider other (ellipsoidal) localisation regions.

2.3.1 Local curvature bounds for the likelihood function

In what follows, $\theta_0 \in \Theta$ is an arbitrary 'ground truth' and the gradient operator $\nabla = \nabla_\theta$ will always act on $\mathcal{G}$ viewed as a map on the subspace $\mathbb{R}^D \subseteq \Theta$. Specifically we shall write $\nabla\mathcal{G}(\theta)$ and $\nabla^2\mathcal{G}(\theta)$ for the following vector and matrix fields

$$\nabla\mathcal{G}(\theta) : \mathcal{O} \to \mathbb{R}^D, \quad \nabla^2\mathcal{G}(\theta) : \mathcal{O} \to \mathbb{R}^{D\times D},$$

respectively. The following condition summarises some quantitative regularity conditions on the map $\mathcal{G}$. These have to hold locally on the set $\mathcal{B}$ (and are satisfied, for instance, for any smooth $\mathcal{G}$). To formulate them we equip $\mathbb{R}^D$ and $\mathbb{R}^{D\times D}$ with the Euclidean norm $\|\cdot\|_{\mathbb{R}^D}$ and the operator norm $\|\cdot\|_{op} = \|\cdot\|_{\mathbb{R}^D\to\mathbb{R}^D}$ (for linear maps from $\mathbb{R}^D \to \mathbb{R}^D$) respectively, and the functional norms of $\mathbb{R}^D$- or $\mathbb{R}^{D\times D}$-valued fields are understood relative to these norms.

[So for instance, in (2.40), one requires a bound k_2 for $\sup_{x\in\mathcal{O}} \|\nabla^2\mathcal{G}(\theta)(x)\|_{\mathbb{R}^D\to\mathbb{R}^D}$ that is uniform in $\theta \in \mathcal{B}$.]

Assumption 2.3.2 (Local regularity)**.** *Let $\mathcal{B}$ be given in (2.39).*

i) *For any $x \in \mathcal{O}$, the map $\theta \mapsto \mathcal{G}(\theta)(x)$ is twice continuously differentiable on $\mathcal{B}$.*

ii) *For some $k_0, k_1, k_2 > 0$,*

$$\begin{aligned} \sup_{\theta\in\mathcal{B}} \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_\infty &\leq k_0, \\ \sup_{\theta\in\mathcal{B}} \|\nabla\mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} &\leq k_1, \\ \sup_{\theta\in\mathcal{B}} \|\nabla^2\mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} &\leq k_2. \end{aligned} \tag{2.40}$$

iii) *For some $m_0, m_1, m_2 > 0$ and any $\theta, \bar\theta \in \mathcal{B}$, we have*

$$\begin{aligned} \|\mathcal{G}(\theta) - \mathcal{G}(\bar\theta)\|_\infty &\leq m_0|\theta - \bar\theta|_1, \\ \|\nabla\mathcal{G}(\theta) - \nabla\mathcal{G}(\bar\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} &\leq m_1|\theta - \bar\theta|_1, \\ \|\nabla^2\mathcal{G}(\theta) - \nabla^2\mathcal{G}(\bar\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} &\leq m_2|\theta - \bar\theta|_1. \end{aligned}$$

We now turn to the central condition underlying the results in this section in terms of a local curvature bound on $E_{\theta_0}[-\nabla^2\ell(\theta, Z)]$, with $\ell(\theta) : \mathbb{R}^D \to \mathbb{R}$ from (2.38). To motivate it, notice that

$$-\nabla^2\ell(\theta, Z) = [\nabla\mathcal{G}(\theta)(X)][\nabla\mathcal{G}(\theta)(X)]^T + [\mathcal{G}(\theta)(X) - Y]\nabla^2[\mathcal{G}(\theta)(X)]. \tag{2.41}$$

If the design distribution P^X is uniform on a bounded domain $\mathcal{O}$ then at $\theta = \theta_0$, the $E^N_{\theta_0}$-expectation of the last expression can be represented as

$$v^T E_{\theta_0}[-\nabla^2\ell(\theta_0, Z)]v = \|\nabla\mathcal{G}(\theta_0)^T v\|^2_{L^2(\mathcal{O})}, \quad v \in \mathbb{R}^D. \tag{2.42}$$

Therefore, if a suitable '$L^2(\mathcal{O})$-stability estimate' for the linearisation $\nabla\mathcal{G}$ of $\mathcal{G}$ at θ_0 is available, the key condition (2.43) below holds at θ_0; by regularity of $\mathcal{G}$ this should extend to θ sufficiently close to θ_0. In the example with the Schrödinger equation studied in Section 2.2, such a stability estimate indeed follows from elliptic PDE theory, see Lemma 2.4.7.

Note that the Hessian $E_{\theta_0}[-\nabla^2\ell(\theta, Z)]$ is symmetric (by (2.41) and Assumption 2.3.2i)), and recall that $\lambda_{min}(A)$ denotes the smallest eigenvalue of a symmetric matrix A.

Assumption 2.3.3 (Local curvature)**.** *Let $\mathcal{B}$ be given in (2.39) and let $\ell : \mathbb{R}^D \to \mathbb{R}$ be as in (2.38).*

i) *For some $c_{min} > 0$, we have*

$$\inf_{\theta\in\mathcal{B}} \lambda_{min}\Big(E_{\theta_0}[-\nabla^2\ell(\theta, Z)]\Big) \geq c_{min}. \tag{2.43}$$

ii) *For some $c_{max} \geq c_{min} > 0$, we have*

$$\sup_{\theta\in\mathcal{B}} \Big[|E_{\theta_0}\ell(\theta, Z)| + \|E_{\theta_0}[\nabla\ell(\theta, Z)]\|_{\mathbb{R}^D} + \|E_{\theta_0}[\nabla^2\ell(\theta, Z)]\|_{op}\Big] \leq c_{max}. \tag{2.44}$$

The following lemma, which is based on concentration of measure arguments, shows that the local 'average' curvature bound in (2.43) carries over to the 'observed' log-likelihood function, with high frequentist $P_{\theta_0}^N$-probability, and whenever $D \leq \mathcal{R}_N$, where the dimension constraint is explicitly quantified in terms of the constants featuring in the previous hypotheses. The expression for $\mathcal{R}_N$ substantially simplifies in concrete settings but, in this general form, reflects the various non-asymptotic stochastic regimes of the log-likelihood function and its derivatives.

Lemma 2.3.4. *Suppose that data arises from (2.37) with $\ell_N : \mathbb{R}^D \to \mathbb{R}$ given by (2.38). Suppose Assumptions 2.3.2, 2.3.3 are satisfied. There exists a universal constant $C > 0$ such that if*

$$\mathcal{R}_N := CN \min\left\{\frac{c_{min}^2}{C_{\mathcal{G}}^2\eta^2}, \frac{c_{min}}{C_{\mathcal{G}}\eta}, \frac{c_{min}^2}{C_{\mathcal{G}}'^2}, \frac{c_{min}}{k_2}, \frac{c_{max}^2}{C_{\mathcal{G}}''^2\eta^2}, \frac{c_{max}}{C_{\mathcal{G}}''\eta}, \frac{c_{max}^2}{C_{\mathcal{G}}'''^2}, \frac{c_{max}}{k_0+k_1}\right\}, \tag{2.45}$$

where

$$\begin{aligned} C_{\mathcal{G}} &:= k_0m_2 + k_1m_1 + k_2m_0 + m_2, \qquad & C_{\mathcal{G}}' &:= k_1^2 + k_0k_2 + k_2,\\ C_{\mathcal{G}}'' &:= k_0m_1 + k_1m_0 + m_1 + k_0m_0 + m_0, \qquad & C_{\mathcal{G}}''' &= k_0k_1 + k_1 + k_0^2 + k_0, \end{aligned} \tag{2.46}$$

then for any $D, N \geq 1$ satisfying $D \leq \mathcal{R}_N$, we have

$$P_{\theta_0}^N\Big(\inf_{\theta\in\mathcal{B}} \lambda_{min}\big[-\nabla^2\ell_N(\theta, Z^{(N)})\big] < \frac{1}{2}Nc_{min}\Big) \leq 8e^{-\mathcal{R}_N}, \tag{2.47}$$

as well as

$$\begin{aligned} P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \Big[|\ell_N(\theta, Z^{(N)})| + \|\nabla\ell_N(\theta, Z^{(N)})\|_{\mathbb{R}^D} + \|\nabla^2\ell_N(\theta, Z^{(N)})\|_{op}\Big] > N(5c_{max}+1)\Big)&\\ \leq 24e^{-\mathcal{R}_N} + e^{-N/8}.& \end{aligned} \tag{2.48}$$

Inspection of the proof shows that for the first inequality (2.47), the terms involving c_{max} can be removed from the definition of $\mathcal{R}_N$. In the sequel we will restrict considerations to the event

$$\mathcal{E}_{conv} := \Big\{ \inf_{\theta \in \mathcal{B}} \lambda_{min}\big[-\nabla^2 \ell_N(\theta)\big] \geq N c_{min}/2 \Big\} \\ \cap \Big\{ \sup_{\theta \in \mathcal{B}} \Big[|\ell_N(\theta)| + \|\nabla \ell_N(\theta)\|_{\mathbb{R}^D} + \|\nabla^2 \ell_N(\theta)\|_{op} \Big] \leq N(5c_{max}+1) \Big\}, \tag{2.49}$$

whose $P_{\theta_0}^N$-probability is controlled by Lemma 2.3.4.

2.3.2 Construction of the likelihood surrogate function

For Bayesian computation via Langevin-type algorithms one needs to ensure recurrence of the underlying diffusion process, a sufficient condition for which is *global log-concavity* (on $\mathbb{R}^D$) of the target measure to be sampled from, see Section 2.5.1. To this end we now construct a 'surrogate log-likelihood function' $\tilde{\ell}_N : \mathbb{R}^D \to \mathbb{R}$ for the log-likelihood ℓ_N such that $\tilde{\ell}_N = \ell_N$ identically on the subset $\{\theta \in \mathbb{R}^D : |\theta - \theta^*|_1 \leq 3\eta/8\}$ of $\mathcal{B}$ from (2.39), and which will be shown to be globally log-concave on the event $\mathcal{E}$ from (2.60) below.

In order to perform the convexification of $-\ell_N$, one needs to identify the region $\mathcal{B}$ up to sufficient precision. In what follows, we denote by $\theta_{init} = \theta_{init}(Z^{(N)}) \in \mathbb{R}^D$ a (data-driven) point estimator where the sampling algorithm is initialised; and we define the event $\mathcal{E}_{init}$ (measurable subset of $(\mathbb{R} \times \mathcal{O})^N$) by

$$\mathcal{E}_{init} := \{|\theta_{init} - \theta^*|_1 \leq \eta/8\}, \tag{2.50}$$

where θ_{init} belongs to the region $\mathcal{B}$. That such initialisation is possible (i.e., that $\mathcal{E}_{init}$ has sufficiently high $P_{\theta_0}^N$-probability for appropriate $\eta > 0$) is proved for the Schrödinger model in Theorem 2.5.10.

We require two auxiliary functions, g_η (globally convex) and α_η (cut-off function): For some smooth and symmetric (about 0) function $\varphi : \mathbb{R} \to [0, \infty)$ satisfying $\mathrm{supp}(\varphi) \subseteq [-1, 1]$ and $\int_{\mathbb{R}} \varphi(x)dx = 1$, let us define the mollifiers $\varphi_h(x) := h^{-1}\varphi(x/h), h > 0$. Then, we define the functions $\tilde{\gamma}_\eta, \gamma_\eta : \mathbb{R} \to \mathbb{R}$ by

$$\begin{aligned} \tilde{\gamma}_\eta(t) &:= \begin{cases} 0 & \text{if} \quad t < 5\eta/8, \\ (t - 5\eta/8)^2 & \text{if} \quad t \geq 5\eta/8, \end{cases} \\ \gamma_\eta(t) &:= \big[\varphi_{\eta/8} * \tilde{\gamma}_\eta\big](t), \end{aligned} \tag{2.51}$$

where $*$ denotes convolution, and

$$g_\eta : \mathbb{R}^D \to [0, \infty), \qquad g_\eta(\theta) := \gamma_\eta(|\theta - \theta_{init}|_1). \tag{2.52}$$

Finally, for some smooth $\alpha : [0, \infty) \to [0, 1]$ which satisfies $\alpha(t) = 1$ for $t \in [0, 3/4]$ and $\alpha(t) = 0$ for $t \in [7/8, \infty)$, we define the 'cut-off' function

$$\alpha_\eta : \mathbb{R}^D \to [0, 1], \qquad \alpha_\eta(\theta) = \alpha(|\theta - \theta_{init}|_1/\eta). \tag{2.53}$$

Definition 2.3.5. *For the auxiliary functions g_η, α_η from (2.52), (2.53) and $K > 0$, we define the surrogate likelihood function $\tilde{\ell}_N$ by*

$$\tilde{\ell}_N : \mathbb{R}^D \to \mathbb{R}, \quad \tilde{\ell}_N(\theta) := \alpha_\eta(\theta)\ell_N(\theta) - K g_\eta(\theta). \tag{2.54}$$

When the choice of the constant $K > 0$ is large enough relative to c_{max} from Assumption 2.3.2, the following global convexity property can be proved for $\tilde{\ell}_N$ (see Section 2.5 for a proof).

Proposition 2.3.6. *On the event $\mathcal{E}_{conv} \cap \mathcal{E}_{init}$ (cf. (2.49), (2.50)), when $\tilde{\ell}_N$ from (2.54) is defined with any constant K satisfying*

$$K \geq CN(c_{max} + 1) \cdot \frac{1 + \lambda_{max}(M)/\eta^2}{\lambda_{min}(M)}, \tag{2.55}$$

($C > 1$ depending only on the function α above), we have

$$\ell_N(\theta) = \tilde{\ell}_N(\theta) \quad \text{for all } \theta \in \mathbb{R}^D \text{ s.t. } |\theta - \theta^*|_1 \leq 3\eta/8,$$

and

$$\inf_{\theta \in \mathbb{R}^D} \lambda_{min}\big(- \nabla^2 \tilde{\ell}_N(\theta)\big) \geq N c_{min}/2, \tag{2.56}$$

as well as

$$\|\nabla \tilde{\ell}_N(\theta) - \nabla \tilde{\ell}_N(\bar{\theta})\|_{\mathbb{R}^D} \leq 7K\lambda_{max}(M)\|\theta - \bar{\theta}\|_{\mathbb{R}^D}, \quad \theta, \bar{\theta} \in \mathbb{R}^D. \tag{2.57}$$

2.3.3 Non-asymptotic bounds for Bayesian posterior computation

We now consider the problem of generating random samples from the posterior measure

$$\Pi[B|Z^{(N)}] = \frac{\int_B e^{\ell_N(\theta, Z^{(N)})} d\Pi(\theta)}{\int_{\mathbb{R}^D} e^{\ell_N(\theta, Z^{(N)})} d\Pi(\theta)}, \qquad B \subseteq \mathbb{R}^D \text{ measurable},$$

arising from data (2.37) with log-likelihood (2.38) and Gaussian $N(0, \Sigma)$ prior Π of density π on $\mathbb{R}^D$, with positive definite covariance matrix $\Sigma \in \mathbb{R}^{D \times D}$.

We use the stochastic gradient method obtained from an Euler discretisation of the D-dimensional Langevin diffusion (see Section 2.5.1) with drift vector field $\nabla(\tilde{\ell}_N + \log \pi)$

based on the surrogate likelihood function. More precisely, for *stepsize* $\gamma > 0$ and auxiliary variables $\xi_k \sim^{i.i.d.} N(0, I_{D\times D})$, define a Markov chain as

$$\begin{aligned} \vartheta_0 &= \theta_{init}, \\ \vartheta_{k+1} &= \vartheta_k + \gamma[\nabla\tilde{\ell}_N(\vartheta_k) - \Sigma^{-1}\vartheta_k] + \sqrt{2\gamma}\xi_{k+1}, \quad k = 0, 1, \ldots \end{aligned} \tag{2.58}$$

Probabilities and expectations with respect to the law of this Markov chain (random only through the ξ_k, conditional on the data $Z^{(N)}$) will be denoted by $\mathbf{P}_{\theta_{init}}, \mathbf{E}_{\theta_{init}}$ respectively. The invariant measure of the underlying continuous time Langevin diffusion equals the *surrogate posterior distribution* given by

$$\tilde{\Pi}[B|Z^{(N)}] := \frac{\int_B e^{\tilde{\ell}_N(\theta, Z^{(N)})} d\Pi(\theta)}{\int_{\mathbb{R}^D} e^{\tilde{\ell}_N(\theta, Z^{(N)})} d\Pi(\theta)}, \qquad B \subseteq \mathbb{R}^D \text{ measurable.}$$

In the following results we assume that the Wasserstein distance W_2 between $\tilde{\Pi}(\cdot|Z^{(N)})$ and $\Pi(\cdot|Z^{(N)})$ can be controlled, specifically, for any $\rho > 0$, let us define the event

$$\mathcal{E}_{wass}(\rho) := \{W_2^2(\Pi[\,\cdot\,|Z^{(N)}], \tilde{\Pi}[\,\cdot\,|Z^{(N)}]) \leq \rho/2\}. \tag{2.59}$$

For the Schrödinger model this is achieved in Theorem 2.4.14, for ρ decaying exponentially in N, using that most of the posterior mass (and its mode) concentrate on the set $\mathcal{B}$ from (2.39).

Our first result consists of a global Wasserstein-approximation of $\Pi(\cdot|Z^{(N)})$ by the law $\mathcal{L}(\vartheta_k)$ on $\mathbb{R}^D$ of the k-th iterate ϑ_k arising from (2.58).

Theorem 2.3.7 (Non-asymptotic Wasserstein mixing)**.** *Suppose that the model given by (2.37)-(2.38) fulfills the Assumptions 2.3.2, 2.3.3 for some $0 < \eta \leq 1$, that $D, N \in \mathbb{N}$ are such that $D \leq \mathcal{R}_N$ with $\mathcal{R}_N$ from (2.45) and let K be as in (2.55). Further define the constants*

$$m := Nc_{min}/2 + \lambda_{min}(\Sigma^{-1}), \qquad \Lambda := 7K\lambda_{max}(M) + \lambda_{max}(\Sigma^{-1}).$$

Then for any $0 < \gamma \leq 1/\Lambda$ and any $\rho > 0$ the algorithm $(\vartheta_k : k \geq 0)$ from (2.58) satisfies, on the event (i.e., measurable subset of $(\mathbb{R} \times \mathcal{O})^N$)

$$\mathcal{E} := \mathcal{E}_{conv} \cap \mathcal{E}_{init} \cap \mathcal{E}_{wass}(\rho), \tag{2.60}$$

(with $\mathcal{E}_{conv}, \mathcal{E}_{init}, \mathcal{E}_{wass}(\rho)$ defined in (2.49), (2.50), (2.59), respectively), and all $k \geq 0$,

$$W_2^2(\mathcal{L}(\vartheta_k), \Pi[\cdot|Z^{(N)}]) \leq \rho + b(\gamma) + 4\big(\tau(\Sigma, M, R) + \frac{D}{m}\big)\Big(1 - \frac{\gamma m}{2}\Big)^k, \tag{2.61}$$

where, for some universal constants $c_1, c_2 > 0$, any $R \geq \|\theta^*\|_{\mathbb{R}^D}$ and $\kappa(\Sigma) = \lambda_{max}(\Sigma)/\lambda_{min}(\Sigma)$,

$$b(\gamma) = c_1\Big[\frac{\gamma D\Lambda^2}{m^2} + \frac{\gamma^2 D\Lambda^4}{m^3}\Big], \quad \tau(\Sigma, M, R) = c_2\kappa(\Sigma)\Big[1 + \frac{\eta^2}{\lambda_{min}(M)} + R^2\Big]. \tag{2.62}$$

From the previous theorem we can obtain the following bound on the computation of posterior functionals by ergodic averages of ϑ_k collected after some burn-in time $J_{in} \in \mathbb{N}$. Specifically, if we define, for any $H : \mathbb{R}^D \to \mathbb{R}$ integrable with respect to $\Pi(\cdot|Z^{(N)})$, the random variable

$$\hat{\pi}^J_{J_{in}}(H) = \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} H(\vartheta_k), \tag{2.63}$$

we obtain the following non-asymptotic concentration bound.

Theorem 2.3.8 (Lipschitz functionals)**.** *In the setting of the previous theorem, there exist further constants $c_3, c_4 > 0$ such that for any $\rho > 0$, any burn-in period*

$$J_{in} \geq \frac{c_3}{m\gamma} \times \log\Big(1 + \frac{1}{\rho + b(\gamma)} + \tau(\Sigma, M, R) + \frac{D}{m}\Big), \tag{2.64}$$

any $J \in \mathbb{N}$, any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$, any

$$t \geq \sqrt{8}\|H\|_{Lip}\sqrt{\rho + b(\gamma)} \tag{2.65}$$

and on the event $\mathcal{E}$ from (2.60), we have

$$\mathbf{P}_{\theta_{init}}\Big(|\hat{\pi}^J_{J_{in}}(H) - E^\Pi[H|Z^{(N)}]| \geq t\Big) \leq 2\exp\Big(-c_4\frac{t^2m^2J\gamma}{\|H\|^2_{Lip}(1 + 1/(mJ\gamma))}\Big). \tag{2.66}$$

From the last theorem one can obtain as a direct consequence the following guarantee for computation of the posterior mean $E^\Pi[\theta|Z^{(N)}]$ by the ergodic average accrued along the Markov chain.

Corollary 2.3.9. *In the setting of Theorem 2.3.8, if we define*

$$\bar{\theta}^J_{J_{in}} = \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} \vartheta_k,$$

then on the event $\mathcal{E}$ and for $t \geq \sqrt{8}\sqrt{\rho + b(\gamma)}$, we have for some constant $c_5 > 0$ that

$$\mathbf{P}_{\theta_{init}}\Big(\|\bar{\theta}^J_{J_{in}} - E^\Pi[\theta|Z^{(N)}]\|_{\mathbb{R}^D} \geq t\Big) \leq 2D\exp\Big(-c_5\frac{t^2m^2J\gamma}{D(1 + 1/(mJ\gamma))}\Big). \tag{2.67}$$

The two previous results imply that one can compute the posterior mean (or $E^\Pi[H|Z^{(N)}]$ with $\|H\|_{Lip} \leq 1$) within precision $\varepsilon > 0$ as long as $\epsilon \gtrsim \sqrt{\rho}$: For instance if γ is chosen as

$$\gamma \simeq \min\Big\{\frac{\varepsilon^2 m^2}{D\Lambda^2}, \frac{\varepsilon m^{3/2}}{D^{1/2}\Lambda^2}\Big\},$$

then the overall number of required MCMC iterations $J_{in} + J$ depends polynomially on the quantities $N, D, m^{-1}, \Lambda, \varepsilon^{-1}$. When the latter three constants exhibit at most polynomial growth in N, D (as is the case for the Schrödinger equation treated in Section 2.2), we can deduce that polynomial-time computation of such posterior characteristics is feasible, on the event $\mathcal{E}$ from (2.60) at computational cost $J_{in} + J = O(N^{b_1} D^{b_2} \varepsilon^{-b_3}), b_1, b_2, b_3 > 0$, with $\mathbf{P}_{\theta_{init}}$-probability as close to 1 as desired.

Remark 2.3.10 (About the events $\mathcal{E}_{init}, \mathcal{E}_{wass}$)**.** Controlling the probability of the events $\mathcal{E}_{init}, \mathcal{E}_{wass}$ (featuring in the definition of $\mathcal{E}$ in (2.60)) on which the preceding bounds hold may pose a formidable challenge in its own right when considering a concrete 'forward map' $\mathcal{G}$. For our prototypical example of the Schrödinger equation from Section 2.2, this is achieved in Sections 2.4.2 and 2.5.4. The proofs there give some guidance for how to proceed in other settings, too. In essence one can expect that in bounding the $P_{\theta_0}^N$-probability of the events $\mathcal{E}_{init}, \mathcal{E}_{wass}$, *global* 'stability' and 'range' properties of the map $\mathcal{G}$ will play a role, whereas the Assumptions 2.3.2, 2.3.3 employed in this section are 'local' in the sense that they concern properties of $\mathcal{G}$ on $\mathcal{B}$ from (2.39) only. Discerning local from global requirements on $\mathcal{G}$ in this way appears helpful both in the proofs and in the exposition of the main ideas of this chapter.

Remark 2.3.11 (Extensions to vector-valued data)**.** The key results of this section apply to other settings where the 'forward' map $\mathcal{G}(\theta)$ defines an element of the space of continuous maps $C(\mathcal{M} \to V)$ from a d-dimensional compact Riemannian manifold $\mathcal{M}$ (possibly with boundary) into a finite-dimensional vector space V of fixed finite dimension $dim(V) < \infty$. If we assume that the statistical errors $(\varepsilon_i : i = 1, \dots, N)$ in equation (2.37) are i.i.d. $N(0, Id_V)$ in V, then the log-likelihood function of the model is not given by (2.38) but instead of the form

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^N \|Y_i - \mathcal{G}(\theta)(X_i)\|_V^2, \quad \ell(\theta) = -\frac{1}{2}\|Y - \mathcal{G}(\theta)(X)\|_V^2,$$

where the X_i, X are drawn i.i.d. from a Borel measure P^X on $\mathcal{M}$. Imposing Assumption 2.3.2 with the obvious modification of the norms there for V-valued maps, and if Assumption 2.3.3 holds for the preceding definition of $\ell(\theta)$, then the conclusion of Lemma 2.3.4 remains valid as stated, after basic notational adjustments in its proof.

2.3.4 Proof of Lemma 2.3.4

It suffices to prove the assertion for $\mathcal{R}_N \geq 1$. We first need some more notation: For any $x \in \mathcal{O}$, we denote the point evaluation map by

$$\mathcal{G}^x: \quad \Theta \to \mathbb{R}, \qquad \theta \mapsto \mathcal{G}(\theta)(x).$$

For $Z = (Y, X) \sim P_{\theta_0}$, we will frequently use the following identities in the proofs below (where we recall that ∇ and ∇^2 act on the θ-variable).

$$\begin{aligned} -\ell(\theta, Z) &= \frac{1}{2}\big[Y - \mathcal{G}^X(\theta)\big]^2 = \frac{1}{2}\big[\mathcal{G}^X(\theta_0) + \varepsilon - \mathcal{G}^X(\theta)\big]^2, \\ -\nabla\ell(\theta, Z) &= \big[\mathcal{G}^X(\theta) - \mathcal{G}(\theta_0) - \varepsilon\big]\nabla\mathcal{G}^X(\theta), \\ -\nabla^2\ell(\theta, Z) &= \nabla\mathcal{G}^X(\theta)\nabla\mathcal{G}^X(\theta)^T + \big[\mathcal{G}^X(\theta) - \mathcal{G}(\theta_0) - \varepsilon\big]\nabla^2\mathcal{G}^X(\theta), \\ -E_{\theta_0}[\ell(\theta, Z)] &= \frac{1}{2} + \frac{1}{2}E^X[\mathcal{G}^X(\theta_0) - \mathcal{G}^X(\theta)]^2, \end{aligned} \tag{2.68}$$

where we note that by Assumption 2.3.2, the Hessian $\nabla^2\ell(\theta, Z)$ is a symmetric $D \times D$ matrix field. When no confusion can arise, we will suppress the second argument Z and write $\ell(\theta)$ for $\ell(\theta, Z)$.

Throughout, $P_N := N^{-1}\sum_{i=1}^N \delta_{Z_i}$ denotes the empirical measure induced by $Z^{(N)}$, which acts on measurable functions $h : \mathbb{R} \times \mathcal{O} \to \mathbb{R}$ via

$$P_N(h) = \int_{\mathbb{R}\times\mathcal{O}} h dP_N = \frac{1}{N}\sum_{i=1}^N h(Z_i).$$

2.3.4.1 Proof of (2.47)

Let us write $\bar{\ell}_N := \ell_N/N$. Then, by a standard inequality due to Weyl as well as Assumption 2.3.3, we have for any $\theta \in \mathcal{B}$ that

$$\begin{aligned} \lambda_{min}\big[-\nabla^2\bar{\ell}_N(\theta)\big] &\geq \lambda_{min}\big(E_{\theta_0}\big[-\nabla^2\ell(\theta)\big]\big) - \big\|\nabla^2\bar{\ell}_N(\theta) - E_{\theta_0}\big[\nabla^2\ell(\theta)\big]\big\|_{op} \\ &\geq c_{min} - \big\|\nabla^2\bar{\ell}_N(\theta) - E_{\theta_0}\big[\nabla^2\ell(\theta)\big]\big\|_{op}. \end{aligned} \tag{2.69}$$

Hence we deduce

$$\begin{aligned} &P_{\theta_0}^N\Big(\inf_{\theta\in\mathcal{B}} \lambda_{min}[\nabla^2\ell_N(\theta, Z)] < Nc_{min}/2\Big) \\ &\qquad \leq P_{\theta_0}^N\Big(\big\|\nabla^2\bar{\ell}_N(\theta) - E_{\theta_0}[\nabla^2\ell(\theta)]\big\|_{op} \geq c_{min}/2 \text{ for some } \theta \in \mathcal{B}\Big) \\ &\qquad \leq P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\ \sup_{v:\|v\|_{\mathbb{R}^D}\leq 1}\Big|v^T\Big(\nabla^2\bar{\ell}_N(\theta) - E_{\theta_0}[\nabla^2\ell(\theta)]\Big)v\Big| \geq c_{min}/2\Big) \\ &\qquad = P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\ \sup_{v:\|v\|_{\mathbb{R}^D}\leq 1}\big|P_N(g_{v,\theta})\big| \geq c_{min}/2\Big), \end{aligned} \tag{2.70}$$

where

$$g_{v,\theta}(\cdot) := v^T\Big(\nabla^2\ell(\theta,\cdot) - E_{\theta_0}[\nabla^2\ell(\theta)]\Big)v, \qquad v\in\mathbb{R}^D.$$

The next step is to reduce the supremum over $\{v : \|v\|_{\mathbb{R}^D} \le 1\}$ to a suitable finite maximum over grid points v_i by a contraction argument (commonly used in high-dimensional probability). For $\rho > 0$, let $N(\rho)$ denote the minimal number of balls of $\|\cdot\|_{\mathbb{R}^D}-$radius ρ required to cover $\{v : \|v\|_{\mathbb{R}^D} \le 1\}$, and let v_i, $\|v_i\|_{\mathbb{R}^D} \le 1$, be the centre points of a minimal covering. Thus for any $v \in \mathbb{R}^D$ there exists an index i such that $\|v - v_i\|_{\mathbb{R}^D} \le \rho$. Hence, writing shorthand

$$M_\theta = \nabla^2\bar{\ell}_N(\theta) - E_{\theta_0}[\nabla^2\ell(\theta)], \quad \theta\in\mathcal{B},$$

we have by the Cauchy-Schwarz inequality and the symmetry of the matrix M_θ,

$$\begin{aligned} v^T M_\theta v &= v_i^T M_\theta v_i + (v-v_i)^T M_\theta v + v_i^T M_\theta (v - v_i) \\ &= v_i^T M_\theta v_i + \|v-v_i\|_{\mathbb{R}^D}\|M_\theta v\|_{\mathbb{R}^D} + \|v-v_i\|_{\mathbb{R}^D}\|M_\theta v_i\|_{\mathbb{R}^D} \\ &\le v_i^T M_\theta v_i + 2\rho \sup_{v:\|v\|_{\mathbb{R}^D}\le 1} v^T M_\theta v. \end{aligned}$$

Choosing $\rho = \frac{1}{4}$ and taking suprema it follows that for any $\theta\in\mathcal{B}$,

$$\sup_{v:\|v\|_{\mathbb{R}^D}\le 1} v^T M_\theta v \le 2 \max_{i=1,\dots,N(1/4)} v_i^T M_\theta v_i. \tag{2.71}$$

Since the covering (v_i) is independent of θ, we can further estimate the right hand side of (2.70) by a union bound to the effect that

$$\begin{aligned} &P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\sup_{v:\|v\|_{\mathbb{R}^D}\le 1}\Big|v^T M_\theta v\Big| \ge c_{min}/2\Big) \\ &\le N(1/4)\cdot \sup_{v:\|v\|_{\mathbb{R}^D}\le 1} P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\Big|v^T M_\theta v\Big| \ge c_{min}/4\Big) \\ &\le N(1/4)\cdot \sup_{v:\|v\|_{\mathbb{R}^D}\le 1}\Big[P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\Big|P_N(g_{v,\theta} - g_{v,\theta^*})\Big| \ge c_{min}/8\Big) + P_{\theta_0}^N\big(\big|P_N(g_{v,\theta^*})\big| \ge c_{min}/8\big)\Big], \end{aligned} \tag{2.72}$$

where we recall that θ^* is the centrepoint of the set $\mathcal{B}$ from (2.39). For the rest of the proof, we fix any $v\in\mathbb{R}^D$ with $\|v\|_{\mathbb{R}^D} = 1$. Next, we use (2.68) to decompose the 'uncentred' part of $g_{v,\theta}$ as

$$\begin{aligned} -v^T\nabla^2\ell(\theta, Z)v &= v^T\Big[\mathcal{G}^X(\theta)\nabla\mathcal{G}^X(\theta)^T + [\mathcal{G}^X(\theta) - \mathcal{G}^X(\theta_0)]\nabla^2\mathcal{G}^X(\theta)\Big]v - \varepsilon v^T\nabla^2\mathcal{G}^X(\theta)v \\ &=: \tilde{g}_{v,\theta}^I(X) + \varepsilon g_{v,\theta}^{II}(X), \end{aligned}$$

such that

$$g_{v,\theta}(z) = g_{v,\theta}^I(x) + \varepsilon g_{v,\theta}^{II}(x),$$

where we have defined the centred version of $\tilde{g}^I_{v,\theta}$ as

$$g^I_{v,\theta}(x) = \tilde{g}^I_{v,\theta}(x) - E_{\theta_0}[\tilde{g}^I_{v,\theta}(X)], \quad x \in \mathcal{O}.$$

We can therefore bound the right hand side of (2.72) by

$$\begin{aligned}
N\Big(\frac{1}{4}\Big)\cdot \sup_{v:\|v\|_{\mathbb{R}^D}\leq 1} \Big[& P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} \Big|\frac{1}{N}\sum_{i=1}^N (g^I_{v,\theta} - g^I_{v,\theta^*})(X_i)\Big| \geq \frac{c_{min}}{16}\Big) + P^N_{\theta_0}\Big(\Big|\frac{1}{N}\sum_{i=1}^N g^I_{v,\theta^*}(X_i)\Big| \geq \frac{c_{min}}{16}\Big) \\
&+ P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} \Big|\frac{1}{N}\sum_{i=1}^N \varepsilon_i (g^{II}_{v,\theta} - g^{II}_{v,\theta^*})(X_i)\Big| \geq \frac{c_{min}}{16}\Big) + P^N_{\theta_0}\Big(\Big|\frac{1}{N}\sum_{i=1}^N \varepsilon_i g^{II}_{v,\theta^*}(X_i)\Big| \geq \frac{c_{min}}{16}\Big)\Big] \\
&=: N(1/4)\cdot (i + ii + iii + iv).
\end{aligned}$$

We now use empirical process techniques (Lemma 2.3.12 and also Hoeffding's inequality) to bound the preceding probabilities.

Terms i and ii. In order to apply Lemma 2.3.12 to term i, we require some preparations. By the definition of $\tilde{g}^I_{v,\theta}$ and of the operator norm $\|\cdot\|_{op}$, using the elementary identity $v^T(aa^T - bb^T)v = v^T(a+b)(a-b)^T v$ for any $v, a, b \in \mathbb{R}^D$ and Assumption 2.3.2, we have that for any $\theta, \bar{\theta} \in \mathcal{B}$,

$$\begin{aligned}
\|\tilde{g}^I_{v,\theta} - \tilde{g}^I_{v,\bar{\theta}}\|_\infty &\leq \Big\| \big[\nabla\mathcal{G}(\theta)\nabla\mathcal{G}(\theta)^T + [\mathcal{G}(\theta) - \mathcal{G}(\theta_0)]\nabla^2\mathcal{G}(\theta)\big] \\
&\qquad\qquad - \big[\nabla\mathcal{G}(\bar{\theta})\nabla\mathcal{G}(\bar{\theta})^T + [\mathcal{G}(\bar{\theta}) - \mathcal{G}(\theta_0)]\nabla^2\mathcal{G}(\bar{\theta})\big]\Big\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \\
&\leq \Big\| [\nabla\mathcal{G}(\theta) - \nabla\mathcal{G}(\bar{\theta})]\,[\nabla\mathcal{G}(\theta) + \nabla\mathcal{G}(\bar{\theta})]^T \Big\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \\
&\qquad + \Big\| [\mathcal{G}(\theta) - \mathcal{G}(\bar{\theta})]\nabla^2\mathcal{G}(\theta) \Big\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \\
&\qquad + \Big\| [\mathcal{G}(\bar{\theta}) - \mathcal{G}(\theta_0)]\,[\nabla^2\mathcal{G}(\theta) - \nabla^2\mathcal{G}(\bar{\theta})] \Big\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \\
&\leq 2m_1k_1|\theta - \bar{\theta}|_1 + m_0k_2|\theta - \bar{\theta}|_1 + m_2k_0|\theta - \bar{\theta}|_1 \\
&\leq 2C_{\mathcal{G}}|\theta - \bar{\theta}|_1.
\end{aligned} \tag{2.73}$$

In particular, by (2.39) we obtain the uniform bound

$$\sup_{\theta\in\mathcal{B}} \|g^I_{v,\theta} - g^I_{v,\theta^*}\|_\infty \leq 2\sup_{\theta\in\mathcal{B}} \|\tilde{g}^I_{v,\theta}(X) - \tilde{g}^I_{v,\theta^*}\|_\infty \leq 4C_{\mathcal{G}}|\theta - \theta^*|_1 \leq 4C_{\mathcal{G}}\eta. \tag{2.74}$$

We introduce the rescaled function class

$$h^I_\theta := \frac{g^I_{v,\theta} - g^I_{v,\theta^*}}{16C_{\mathcal{G}}\eta}, \qquad \mathcal{H}^I = \{h^I_\theta : \theta \in \mathcal{B}\},$$

which has envelope and variance proxy bounded as

$$\sup_{\theta\in\mathcal{B}} \|h^I_\theta\|_\infty \le 1/4 \equiv U, \quad \sup_{\theta\in\mathcal{B}} \big(E_{\theta_0}[h^I_\theta(X)^2]\big)^{\frac{1}{2}} \le 1/4 \equiv \sigma. \tag{2.75}$$

Next, if

$$d_2^2(\theta,\bar\theta) = E_{\theta_0}\big[(h^I_\theta(X) - h^I_{\bar\theta}(X))^2\big], \quad d_\infty(\theta,\bar\theta) = \|h^I_\theta - h^I_{\bar\theta}\|_\infty, \qquad \theta,\bar\theta\in\mathcal{B},$$

then using (2.73) we have that

$$d_2(\theta,\bar\theta) \le d_\infty(\theta,\bar\theta) \le |\theta-\bar\theta|_1/\eta, \qquad \theta,\bar\theta\in\mathcal{B}.$$

Thus for any $\rho\in(0,1)$, using Proposition 4.3.34 in [72], we obtain that

$$N(\mathcal{H}^I, d_2, \rho) \le N(\mathcal{H}^I, d_\infty, \rho) \le N(\mathcal{B}, |\cdot|_1/\eta, \rho) \le (3/\rho)^D. \tag{2.76}$$

For any $A\ge 2$ we have

$$\int_0^1 \log(A/x)dx = \log(A) + 1, \qquad \int_0^1 \sqrt{\log(A/x)}dx \le \frac{2\log A}{2\log A - 1}\sqrt{\log(A)},$$

[see p.190 of [72] for the latter inequality], and hence, using this for $A=3$, we can respectively bound the L^∞ and L^2 metric entropy integrals of $\mathcal{H}^I$ by

$$\begin{aligned} \mathcal{J}_\infty(\mathcal{H}^I) &= \int_0^{4U} \log N(\mathcal{H}^I, d_\infty, \rho)d\rho \lesssim D, \\ J_2(\mathcal{H}^I) &\le \int_0^{4\sigma} \sqrt{\log N(\mathcal{H}^I, d_2, \rho)}d\rho \lesssim \sqrt{D}. \end{aligned}$$

Now, an application of Lemma 2.3.12 below implies that for any $x\ge 1$ and some universal constant $L'>0$, we have that

$$P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N h^I_\theta(X_i)\Big| \ge L'\Big[\sqrt{D} + \sqrt{x} + (D+x)/\sqrt{N}\Big]\Big) \le 2e^{-x}. \tag{2.77}$$

We also have by the definition of g^I_{v,θ^*} that

$$\|g^I_{v,\theta^*}\|_\infty \le 2\|\tilde g^I_{v,\theta^*}\|_\infty \le 2(k_1^2 + k_0k_2),$$

and hence by Hoeffding's inequality (Theorem 3.1.2 in [72]) that

$$ii \le 2\exp\Big(-\frac{2Nc^2_{min}}{256\cdot 4(k_1^2+k_0k_2)^2}\Big) \le 2\exp\Big(-\frac{Nc^2_{min}}{512C'^2_{\mathcal{G}}}\Big). \tag{2.78}$$

Now if we define

$$\mathcal{R}_N^{2,I} := CN \min\Big\{\frac{c_{min}^2}{C_{\mathcal{G}}^2\eta^2}, \frac{c_{min}}{C_{\mathcal{G}}\eta}, \frac{c_{min}^2}{C_{\mathcal{G}}'^2}\Big\}, \tag{2.79}$$

then for any $D \le \mathcal{R}_N^{2,I}$ and choosing $x = 4\mathcal{R}_N^{2,I}$ we have

$$L\Big[\sqrt{D} + \sqrt{x} + (D+x)/\sqrt{N}\Big] \le \frac{c_{min}\sqrt{N}}{256C_{\mathcal{G}}\eta}, \qquad 4\mathcal{R}_N^{2,I} \le \frac{Nc_{min}^2}{512C_{\mathcal{G}}'^2},$$

whenever $C > 0$ is small enough. Therefore, combining (2.77) and (2.78), and using the definitions of the term i and of h_θ^I, we obtain

$$ii + i \le 2e^{-4\mathcal{R}_N^{2,I}} + P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N h_\theta^I(X_i)\Big| \ge \frac{c_{min}\sqrt{N}}{256C_{\mathcal{G}}\eta}\Big) \le 4e^{-4\mathcal{R}_N^{2,I}}. \tag{2.80}$$

Terms iii and iv. Let us now treat the empirical process indexed by the functions $\{g_{v,\theta}^{II} : \theta \in \mathcal{B}\}$. Since $\|v\|_{\mathbb{R}^D} \le 1$, we have for any $\theta, \bar\theta \in \mathcal{B}$,

$$\|g_{v,\theta}^{II} - g_{v,\bar\theta}^{II}\|_\infty \le \|\nabla^2\mathcal{G}(\theta) - \nabla^2\mathcal{G}(\bar\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \le m_2|\theta-\bar\theta|_1,$$

which also yields the envelope bound

$$\sup_{\theta\in\mathcal{B}} \big\|g_{v,\theta}^{II} - g_{v,\theta^*}^{II}\big\|_\infty \le m_2 \sup_{\theta\in\mathcal{B}} |\theta-\theta^*|_1 \le m_2\eta.$$

Now the rescaled function class

$$h_\theta^{II} := \frac{g_{v,\theta}^{II} - g_{v,\theta^*}^{II}}{4m_2\eta}, \qquad \mathcal{H}^{II} = \{h_\theta^{II} : \theta\in\mathcal{B}\},$$

admits envelopes

$$\sup_{\theta\in\mathcal{B}} \|h_{v,\theta}^{II}\|_\infty \le 1/4 \equiv U, \qquad \sup_{\theta\in\mathcal{B}} \big(E_{\theta_0}\big[h_{v,\theta}^{II}(X)^2\big]\big)^{\frac12} \le 1/4 \equiv \sigma.$$

Thus defining

$$d_2^2(\theta,\bar\theta) := E_{\theta_0}\big[(h_{v,\theta}^{II}(X) - h_{v,\bar\theta}^{II}(X))^2\big], \qquad d_\infty(\theta,\bar\theta) = \|h_{v,\theta}^{II} - h_{v,\bar\theta}^{II}\|_\infty, \qquad \theta,\bar\theta\in\mathcal{B}$$

we have

$$d_2(\theta,\bar\theta) \le d_\infty(\theta,\bar\theta) \le |\theta-\bar\theta|_1/\eta, \qquad \theta,\bar\theta\in\mathcal{B}.$$

Therefore, just as with the bounds obtained for term i, we have $N(\mathcal{H}^{II}, d_2, \rho) \leq (3/\rho)^D$ and thus, by Lemma 2.3.12 below,

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N \varepsilon_i h_\theta^{II}(X_i)\Big| \geq L'\Big[\sqrt{D}+\sqrt{x}+(D+x)/\sqrt{N}\Big]\Big) \leq 2e^{-x}, \quad x\geq 1. \tag{2.81}$$

Moreover, by the hypotheses, $\|g_{v,\theta^*}^{II}\|_\infty \leq k_2$, and hence, invoking the Bernstein inequality (2.96) with $U=\sigma\equiv k_2$, we obtain that

$$P_{\theta_0}^N\Big(\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^N \varepsilon_i g_{v,\theta^*}^{II}(X_i)\Big| \geq k_2\sqrt{2x}+\frac{k_2 x}{3\sqrt{N}}\Big) \leq 2e^{-x}, \quad x>0. \tag{2.82}$$

We can now set

$$\mathcal{R}_N^{2,II} := CN\min\Big\{\frac{c_{min}^2}{m_2^2\eta^2}, \frac{c_{min}}{m_2\eta}, \frac{c_{min}^2}{k_2^2}, \frac{c_{min}}{k_2}\Big\},$$

and choosing $x=4\mathcal{R}_N^{2,II}$ in the preceding displays, we obtain that for $C>0$ small enough and any $D\leq \mathcal{R}_N^{2,II}$,

$$\begin{aligned} iii+iv &\leq P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N \varepsilon_i h_\theta^{II}(X_i)\Big| \geq \frac{c_{min}\sqrt{N}}{96 m_2\eta}\Big) + P_{\theta_0}^N\Big(\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^N \varepsilon_i g_{v,\theta^*}^{II}(X_i)\Big| \geq \frac{c_{min}\sqrt{N}}{16}\Big) \\ &\leq 4e^{-4\mathcal{R}_N^{2,II}}. \end{aligned} \tag{2.83}$$

Combining the terms. By combining the bounds (2.70), (2.72), (2.80), (2.83) and using that $N(1/4)\leq 9^D\leq e^{3D}$ (cf. Proposition 4.3.34 in [72]) we obtain that since $D\leq \mathcal{R}_N\leq \min(\mathcal{R}_N^{2,I}, \mathcal{R}_N^{2,II})$ from (2.45),

$$\begin{aligned} P_{\theta_0}^N\Big(\inf_{\theta\in\mathcal{B}} \lambda_{min}\big(-\nabla^2\ell_N(\theta,Z)\big) < Nc_{min}/2\Big) &\leq N(1/4)\cdot(i+ii+iii+iv) \\ &\leq 4e^{3D-4\mathcal{R}_N^{2,I}} + 4e^{3D-4\mathcal{R}_N^{2,II}} \leq 8e^{-\mathcal{R}_N}, \end{aligned}$$

completing the proof of (2.47). □

2.3.4.2 Proof of (2.48)

We derive probability bounds for each of the three terms in (2.48) separately. The general scheme of proof for each of the three bounds is similar to the proof of (2.47), and we condense some of the steps to follow.

Second order term. Using that $c_{max}\geq c_{min}$, we can replace (2.70) by

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \lambda_{max}\big[-\nabla^2\ell_N(\theta,Z)\big] \geq 3Nc_{max}/2\Big) \leq P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\sup_{v:\|v\|_{\mathbb{R}^D}\leq 1}\big|P_N(g_{v,\theta})\big| \geq c_{min}/2\Big).$$

From here onwards, this term can be treated exactly as in the proof of (2.47) and thus, for $D \leq \mathcal{R}_n$ from (2.45), we deduce

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \lambda_{max}\big[-\nabla^2\ell_N(\theta,Z)\big] \geq 3Nc_{max}/2\Big) \leq 8e^{-\mathcal{R}_N}. \tag{2.84}$$

First order term. First, let us denote

$$f_{v,\theta}(z) := v^T\Big(\nabla\ell(\theta,z) - E_{\theta_0}[\nabla\ell(\theta,Z)]\Big),\ \ \|v\|_{\mathbb{R}^D} \leq 1, \theta\in\mathcal{B},$$

and let $(v_i : i = 1, ..., N(1/2))$ be the centre points of a $\|\cdot\|_{\mathbb{R}^D}$-covering with balls of radius $1/2$, of the unit ball $\{\theta : \|\theta\|_{\mathbb{R}^D} \leq 1\}$. Then for any v there exists v_i such that $\|v - v_i\|_{\mathbb{R}^D} \leq 1/2$ so that by the Cauchy-Schwarz inequality,

$$\begin{aligned}|P_N(f_{v,\theta})| &\leq |P_N(f_{v,\theta} - f_{v_i,\theta})| + |P_N(f_{v_i,\theta})|\\ &\leq \|v - v_i\|_{\mathbb{R}^D}\big\|\nabla\bar{\ell}_N(\theta) - E_{\theta_0}\big[\nabla\ell(\theta)\big]\big\|_{\mathbb{R}^D} + |P_N(f_{v_i,\theta})|\\ &\leq \frac{1}{2}\big\|\nabla\bar{\ell}_N(\theta) - E_{\theta_0}\big[\nabla\ell(\theta)\big]\big\|_{\mathbb{R}^D} + |P_N(f_{v_i,\theta})|.\end{aligned}$$

Therefore, since $\|u\|_{\mathbb{R}^D} = \sup_{v:\|v\|_{\mathbb{R}^D}\leq 1}|v^T u|$ for any $u \in \mathbb{R}^D$, we deduce for any $\theta \in \mathcal{B}$,

$$\sup_{v:\|v\|_{\mathbb{R}^D}\leq 1}|P_N(f_{v,\theta})| \leq 2\max_{1\leq i\leq N(1/2)}|P_N(f_{v_i,\theta})|. \tag{2.85}$$

We can hence estimate

$$\begin{aligned}P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\|\nabla\bar{\ell}_N(\theta)\|_{\mathbb{R}^D} \geq 3c_{max}/2\Big) &\leq P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\sup_{v:\|v\|_{\mathbb{R}^D}\leq 1}\big|v^T\big[\nabla\bar{\ell}_N(\theta) - E_{\theta_0}[\nabla\ell(\theta)]\big]\big| \geq c_{max}/2\Big)\\ &\leq N(1/2)\cdot\sup_{v:\|v\|_{\mathbb{R}^D}\leq 1} P_{\theta_0}^N\big(\sup_{\theta\in\mathcal{B}}\big|P_N\big(f_{v,\theta}\big)\big| \geq c_{max}/4\big).\end{aligned} \tag{2.86}$$

We fix $v \in \mathbb{R}^D$ with $\|v\|_{\mathbb{R}^D} \leq 1$. Using (2.68), by decomposing the 'uncentred' part of $f_{v,\theta}$ into

$$v^T\nabla\ell(\theta,Z) = v^T\nabla\mathcal{G}^X(\theta)\big[\mathcal{G}^X(\theta) - \mathcal{G}(\theta_0)\big] - \varepsilon v^T\nabla\mathcal{G}^X(\theta) =: \tilde{f}^I_{v,\theta}(X) - \varepsilon f^{II}_{v,\theta}(X),$$

we can then write

$$f_{v,\theta}(z) = f^I_{v,\theta}(x) + \varepsilon f^{II}_{v,\theta}(x),$$

where we have further defined $f^I_{v,\theta}(x) := \tilde f^I_{v,\theta}(x) - E_{\theta_0}[\tilde f^I_{v,\theta}(X)]$. We then estimate the probability on the right hand side of (2.86) as follows,

$$\begin{aligned} P^N_{\theta_0}\big(\sup_{\theta\in\mathcal{B}} |P_N(f_{v,\theta})| \geq c_{max}/4\big) \\ &\leq P^N_{\theta_0}\big(\sup_{\theta\in\mathcal{B}} |P_N(f^I_{v,\theta} - f^I_{v,\theta^*})| \geq c_{max}/16\big) + P^N_{\theta_0}\big(|P_N(f^I_{v,\theta^*})| \geq c_{max}/16\big) \\ &\qquad + P^N_{\theta_0}\big(\sup_{\theta\in\mathcal{B}} |P_N(f^{II}_{v,\theta} - f^{II}_{v,\theta^*})| \geq c_{max}/16\big) + P^N_{\theta_0}\big(|P_N(f^{II}_{v,\theta^*})| \geq c_{max}/16\big) \\ &=: i + ii + iii + iv. \end{aligned} \tag{2.87}$$

We first treat the terms i and ii. By the definition of $\tilde f^I_{v,\theta}$ and Assumption 2.3.2, we have that for any $\theta, \bar\theta \in \mathcal{B}$,

$$\begin{aligned} \|\tilde f^I_{v,\theta} - \tilde f^I_{v,\bar\theta}\|_\infty &\leq \big\|[\nabla\mathcal{G}(\theta) - \nabla\mathcal{G}(\bar\theta)][\mathcal{G}(\theta) - \mathcal{G}(\theta_0)] + \nabla\mathcal{G}(\bar\theta)[\mathcal{G}(\theta) - \mathcal{G}(\bar\theta)]\big\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} \\ &\leq (k_0 m_1 + k_1 m_0)|\theta - \bar\theta|_1. \end{aligned}$$

Again using Assumption 2.3.2, we have likewise

$$\sup_{\theta\in\mathcal{B}} \|\tilde f^I_{v,\theta} - \tilde f^I_{v,\theta^*}\|_\infty \leq (k_0 m_1 + k_1 m_0)\eta.$$

Moreover, using that $\|f^I_{v,\theta^*}\|_\infty \leq 2k_0k_1$, Hoeffding's inequality yields that

$$ii \leq 2\exp\Big(-\frac{Nc^2_{max}}{512k_0^2k_1^2}\Big).$$

Therefore, by using Lemma 2.3.12 in the same manner as in (2.77), we obtain that the rescaled process

$$h^I_{v,\theta} := \frac{\tilde f^I_{v,\theta} - \tilde f^I_{v,\theta^*}}{8(k_0m_1 + k_1m_0)\eta}$$

satisfies

$$P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N h^I_\theta(X_i)\Big| \geq L'\Big[\sqrt{D} + \sqrt{x} + (D+x)/\sqrt{N}\Big]\Big) \leq 2e^{-x}, \quad x \geq 1. \tag{2.88}$$

Thus, setting

$$\mathcal{R}^{1,I}_N =: CN\min\Big\{\frac{c^2_{max}}{(k_0m_1 + k_1m_0)^2\eta^2}, \frac{c_{max}}{(k_0m_1 + k_1m_0)\eta}, \frac{c^2_{max}}{k_0^2k_1^2}\Big\},$$

and choosing $x = 3\mathcal{R}_N^{1,I}$ in (2.88), we obtain that for $C > 0$ small enough and any $D \le \mathcal{R}_N^{1,I}$,

$$ii + i \le 2e^{-3\mathcal{R}_N^{1,I}} + P_{\theta_0}^N\Big(\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^N h_{v,\theta}^I(X_i)\Big| \ge \frac{c_{max}\sqrt{N}}{128(k_0m_1 + k_1m_0)\eta}\Big) \le 4e^{-3\mathcal{R}_N^{1,I}}. \tag{2.89}$$

We now treat the terms iii and iv. As $\|v\|_{\mathbb{R}^D} \le 1$, we have that for any $\theta, \bar\theta \in \mathcal{B}$,

$$\|f_{v,\theta}^{II} - f_{v,\bar\theta}^{II}\|_\infty \le m_1|\theta - \bar\theta|_1, \quad \|f_{v,\theta}^{II} - f_{v,\theta^*}^{II}\|_\infty \le m_1\eta, \quad \|f_{v,\theta^*}^{II}\|_\infty \le k_1.$$

Therefore, by utilising the Lemma 2.3.12 below as well as Bernstein's inequality (2.96) in precisely the same manner as in the derivations of (2.81) and (2.82) respectively, we obtain the two inequalities

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N \varepsilon_i\frac{f_{v,\theta}^{II}(X_i) - f_{v,\theta^*}^{II}(X_i)}{4m_1\eta}\Big| \ge L'\Big[\sqrt{D} + \sqrt{x} + (D+x)/\sqrt{N}\Big]\Big) \le 2e^{-x}, \quad x \ge 1,$$

and

$$P_{\theta_0}^N\Big(\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^N \varepsilon_i f_{v,\theta^*}^{II}(X_i)\Big| \ge k_1\sqrt{2x} + \frac{k_1x}{3\sqrt{N}}\Big) \le 2e^{-x}, \quad x > 0.$$

Thus, if we set

$$\mathcal{R}_N^{1,II} := CN\min\Big\{\frac{c_{max}^2}{m_1^2\eta^2}, \frac{c_{max}}{m_1\eta}, \frac{c_{max}^2}{k_1^2}, \frac{c_{max}}{k_1}\Big\},$$

then for $C > 0$ small enough, for any $D \le 3\mathcal{R}_N^{1,II}$ and choosing $x = 3\mathcal{R}_N^{1,II}$ in the preceding displays, we obtain

$$iii + iv \le 4e^{-3\mathcal{R}_N^{1,II}}. \tag{2.90}$$

By combining (2.86), (2.87), (2.89), (2.90), using that $N(1/2) \le e^{2D}$ (cf. Proposition 4.3.34 in [72]) and since $D \le \mathcal{R}_N \le \min(\mathcal{R}_N^{1,I}, \mathcal{R}_N^{1,II})$, we conclude that

$$\begin{aligned} P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\|\nabla\bar\ell_N(\theta)\|_{\mathbb{R}^D} \ge 3c_{max}/2\Big) &\le N(1/2)\cdot(i + ii + iii + iv) \\ &\le 4e^{2D-3\mathcal{R}_N^{1,I}} + 4e^{2D-3\mathcal{R}_N^{1,II}} \le 8e^{-\mathcal{R}_N}. \end{aligned} \tag{2.91}$$

Order zero term. As with the previous terms, we introduce a decomposition

$$\begin{aligned} -\ell(\theta, Z) &= \frac{1}{2}\big[\mathcal{G}^X(\theta_0) - \mathcal{G}^X(\theta)\big]^2 - \varepsilon\big[\mathcal{G}^X(\theta_0) - \mathcal{G}^X(\theta)\big] + \frac{\varepsilon^2}{2} \\ &=: \tilde l_\theta^I(X) + \varepsilon l_\theta^{II}(X) + \frac{\varepsilon^2}{2}, \end{aligned}$$

and therefore, defining

$$l_\theta^I(x) =: \tilde l_\theta^I(x) - E_{\theta_0}[\tilde l_\theta^I(X)], \quad x \in \mathcal{O},$$

we have that

$$-\ell(\theta, Z) + E_{\theta_0}[\ell(\theta)] = l^I_\theta(X) + \varepsilon l^{II}_\theta(X) + \frac{\varepsilon^2}{2}.$$

Then, using Assumption 2.3.3, we can estimate

$$\begin{aligned}
&P^N_{\theta_0}\big(\sup_{\theta\in\mathcal{B}} |\bar{\ell}_N(\theta, Z)| \geq 2c_{max} + 1\big)\\
&\quad \leq P^N_{\theta_0}\big(\sup_{\theta\in\mathcal{B}} |\bar{\ell}_N(\theta, Z) - E_{\theta_0}[\ell(\theta, Z)]| \geq c_{max} + 1\big)\\
&\quad \leq P^N_{\theta_0}\big(\sup_{\theta\in\mathcal{B}} |P_N(l^I_\theta - l^I_{\theta^*})| \geq \frac{c_{max}}{4}\big) + P^N_{\theta_0}\big(\sup_{\theta\in\mathcal{B}} |P_N(l^I_{\theta^*})| \geq \frac{c_{max}}{4}\big)\\
&\qquad + P^N_{\theta_0}\big(\sup_{\theta\in\mathcal{B}} |P_N(l^{II}_\theta - l^{II}_{\theta^*})| \geq \frac{c_{max}}{4}\big) + P^N_{\theta_0}\big(\sup_{\theta\in\mathcal{B}} |P_N(l^{II}_{\theta^*})| \geq \frac{c_{max}}{4}\big)\\
&\qquad + P^N_{\theta_0}\Big(\frac{1}{2N}\sum_{i=1}^N \varepsilon_i^2 \geq 1\Big) =: i + ii + iii + iv + v.
\end{aligned}$$

To bound the preceding terms, we use Assumption 2.3.2 to deduce that for all $\theta, \bar{\theta} \in \mathcal{B}$,

$$\begin{aligned}
\|l^I_\theta - l^I_{\bar{\theta}}\|_\infty \leq 2\|\tilde{l}^I_\theta - \tilde{l}^I_{\bar{\theta}}\|_\infty &= \big\| - 2\mathcal{G}(\theta_0)\big[\mathcal{G}(\theta) - \mathcal{G}(\bar{\theta})\big] + \mathcal{G}(\theta)^2 - \mathcal{G}(\bar{\theta})^2\big\|_\infty\\
&= \big\|\big[(\mathcal{G}(\theta) - \mathcal{G}(\theta_0)) + (\mathcal{G}(\bar{\theta}) - \mathcal{G}(\theta_0))\big]\big[\mathcal{G}(\theta) - \mathcal{G}(\bar{\theta})\big]\big\|_\infty\\
&\leq 2k_0 m_0 |\theta - \bar{\theta}|_1,
\end{aligned}$$

as well as

$$\sup_{\theta\in\mathcal{B}} \|l^I_\theta - l^I_{\theta^*}\|_\infty \leq 2k_0 m_0 \eta, \quad \|l^I_{\theta^*}\|_\infty \leq k_0^2.$$

Moreover, again by Assumption 2.3.2 we have that for all $\theta, \bar{\theta} \in \mathcal{B}$,

$$\|l^{II}_\theta - l^{II}_{\bar{\theta}}\|_\infty \leq 2m_0|\theta - \bar{\theta}|_1, \quad \sup_{\theta\in\mathcal{B}} \|l^{II}_\theta - l^{II}_{\theta^*}\|_\infty \leq 2m_0\eta, \quad \|l^{II}_{\theta^*}\|_\infty \leq 2k_0.$$

Next, similarly as for the second and first order terms, in order to control the terms i and iii we now apply Lemma 2.3.12 to the empirical processes indexed by the rescaled empirical processes

$$h^I_\theta := \frac{l^I_\theta - l^I_{\theta^*}}{8k_0 m_0 \eta}, \quad h^{II}_\theta := \frac{l^{II}_\theta - l^{II}_{\theta^*}}{8m_0\eta},$$

and in order to control the terms ii and iv, we respectively apply Hoeffding's inequality and Bernstein's inequality (2.96) in the same manner as before. Overall, if we set

$$\begin{aligned}
\mathcal{R}^{0,I}_N &:= CN \min\Big\{\frac{c^2_{max}}{k_0^2 m_0^2 \eta^2}, \frac{c_{max}}{k_0 m_0 \eta}, \frac{c^2_{max}}{k_0^4}\Big\},\\
\mathcal{R}^{0,II}_N &:= CN \min\Big\{\frac{c^2_{max}}{m_0^2 \eta^2}, \frac{c_{max}}{m_0 \eta}, \frac{c^2_{max}}{k_0^2}, \frac{c_{max}}{k_0}\Big\},
\end{aligned} \tag{2.92}$$

then for $C > 0$ small enough, we obtain that for any $D \le \mathcal{R}_N \le \min(\mathcal{R}_N^{0,I}, \mathcal{R}_N^{0,II})$,

$$\begin{aligned} i + ii + iii + iv &\le P_{\theta_0}^N\big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\big|\sum_{i=1}^N h_\theta^I(X_i)\big| \ge \frac{c_{max}\sqrt{N}}{32k_0m_0\eta}\big) + 2\exp\Big(-\frac{Nc_{max}^2}{8k_0^4}\Big) \\ &\quad + P_{\theta_0}^N\big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\big|\sum_{i=1}^N h_\theta^{II}(X_i)\big| \ge \frac{c_{max}\sqrt{N}}{32m_0\eta}\big) + 2e^{-\mathcal{R}_N^{0,II}} \\ &\le 4e^{-\mathcal{R}_N^{0,I}} + 4e^{-\mathcal{R}_N^{0,II}} \le 8e^{-\mathcal{R}_N}. \end{aligned}$$

Finally, we estimate the term v by a standard tail inequality (see Theorem 3.1.9 in [72]),

$$v = P_{\theta_0}^N\Big(\sum_{i=1}^N(\varepsilon_i^2 - 1) \ge N\Big) \le e^{-N/8},$$

and thus obtain

$$P_{\theta_0}^N\big(\sup_{\theta\in\mathcal{B}} |\bar{\ell}_N(\theta, Z)| \ge 2c_{max} + 1\big) \le i + ii + iii + iv + v \le 8e^{-\mathcal{R}_N} + e^{-N/8}. \tag{2.93}$$

Conclusion. By combining (2.84), (2.91) and (2.93), the proof of (2.48) is completed. □

2.3.5 A chaining lemma for empirical processes

The following key technical lemma is based on a chaining argument for stochastic processes with a mixed tail (cf. Theorem 2.2.28 in Talagrand [158] and Theorem 3.5 in Dirksen [55]). For us it will be sufficient to control the 'generic chaining' functionals employed in these references by suitable metric entropy integrals. For any (semi-)metric d on a metric space T, we denote by $N = N(T, d, \rho)$ the minimal cardinality of a covering of T by balls with centres $(t_i : i = 1, \dots, N) \subset T$ such that for all $t \in T$ there exists i such that $d(t, t_i) < \rho$. Below we require the index set Θ to be countable (to avoid measurability issues). Whenever we apply Lemma 2.3.12 in this article with an uncountable set Θ, one can show that the supremum can be realised as one over a countable subset of it.

Lemma 2.3.12. *Let Θ be a countable set. Suppose a class of real-valued measurable functions*

$$\mathcal{H} = \{h_\theta : \mathcal{X} \to \mathbb{R}, \theta \in \Theta\}$$

defined on a probability space $(\mathcal{X}, \mathcal{A}, P^X)$ is uniformly bounded by $U \ge \sup_\theta \|h_\theta\|_\infty$ and has variance envelope $\sigma^2 \ge \sup_\theta E^X h_\theta^2(X)$ where $X \sim P^X$. Define metric entropy integrals

$$J_2(\mathcal{H}) = \int_0^{4\sigma} \sqrt{\log N(\mathcal{H}, d_2, \rho)}d\rho, \quad d_2(\theta, \theta') := \sqrt{E^X[h_\theta(X) - h_{\theta'}(X)]^2},$$

$$J_\infty(\mathcal{H}) = \int_0^{4U} \log N(\mathcal{H}, d_\infty, \rho)d\rho, \quad d_\infty(\theta, \theta') := \|h_\theta - h_{\theta'}\|_\infty.$$

For $X_1, \dots, X_N$ drawn i.i.d. from P^X and $\varepsilon_i \sim^{iid} N(0,1)$ independent of all the X_i's, consider empirical processes arising either as

$$Z_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^N h_\theta(X_i)\varepsilon_i, \quad \theta \in \Theta,$$

or as

$$Z_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (h_\theta(X_i) - Eh_\theta(X)), \quad \theta \in \Theta.$$

We then have for some universal constant $L > 0$ and all $x \geq 1$,

$$\Pr\left(\sup_{\theta\in\Theta} |Z_N(\theta)| \geq L\Big[J_2(\mathcal{H}) + \sigma\sqrt{x} + (J_\infty(\mathcal{H}) + Ux)/\sqrt{N}\Big]\right) \leq 2e^{-x}.$$

Proof. We only prove the case where $Z_N(\theta) = \sum_i h_\theta(X_i)\varepsilon_i/\sqrt{N}$, the simpler case without Gaussian multipliers is proved in the same way. We will apply Theorem 3.5 in [55], whose condition (3.8) we need to verify. First notice that for $|\lambda| < 1/\|h_\theta - h_{\theta'}\|_\infty$, and E^ε denoting the expectation with respect to ε,

$$\begin{aligned} E\exp\{\lambda\varepsilon(h_\theta - h_{\theta'})(X)\} &\leq 1 + \sum_{k=2}^\infty \frac{|\lambda|^k E^\varepsilon|\varepsilon|^k E^X |h_\theta - h_{\theta'}|^k(X)}{k!} \\ &\leq 1 + \lambda^2 E^X[h_\theta(X) - h_{\theta'}(X)]^2 \sum_{k=2}^\infty \frac{E^\varepsilon|\varepsilon|^k}{k!}\left(|\lambda|\|h_\theta - h_{\theta'}\|_\infty\right)^{k-2} \\ &\leq \exp\left\{\frac{\lambda^2 d_2^2(\theta,\theta')}{1 - |\lambda| d_\infty(\theta,\theta')}\right\} \qquad (2.94) \end{aligned}$$

where we have used the basic fact $E^\varepsilon|\varepsilon|^k/k! \leq 1$. By the i.i.d. hypothesis we then also have

$$E\exp\Big\{\lambda(Z_N(\theta) - Z_N(\theta'))\Big\} \leq \exp\left\{\frac{\lambda^2 d_2^2(\theta,\theta')}{1 - |\lambda| d_\infty(\theta,\theta')/\sqrt{N}}\right\}.$$

An application of the exponential Chebyshev inequality (and optimisation in λ, as in the proof of Proposition 3.1.8 in [72]) then implies that condition (3.8) in [55] holds for the stochastic process $Z_N(\theta)$ with metrics $\bar{d}_2 = 2d_2$ and $\bar{d}_1 = d_\infty/\sqrt{N}$. In particular, the $\bar{d}_2$-diameter $\Delta_2(\mathcal{H})$ of $\mathcal{H}$ is at most 4σ and the $\bar{d}_1$-diameter $\Delta_1(\mathcal{H})$ of $\mathcal{H}$ is bounded by $4U/\sqrt{N}$. [These bounds are chosen so that they remain valid for the process without Gaussian multipliers as well.] Theorem 3.5 in [55] now gives, for some universal constant M, and any $\theta_\dagger \in \Theta$ that

$$\Pr\left(\sup_{\theta\in\Theta} |Z_N(\theta) - Z_N(\theta_\dagger)| \geq M(\gamma_2(\mathcal{H}) + \gamma_1(\mathcal{H}) + \sigma\sqrt{x} + (U/\sqrt{N})x)\right) \leq e^{-x}$$

where the 'generic chaining' functionals γ_1, γ_2 are upper bounded by the respective metric entropy integrals of the metric spaces $(\mathcal{H}, \bar{d}_i), i = 1, 2$, up to universal constants (see (2.3) in [55]). For γ_1 also notice that a simple substitution $\rho' = \rho\sqrt{N}$ implies that

$$\int_0^{4U/\sqrt{N}} \log N(\mathcal{H}, \bar{d}_1, \rho) d\rho = \frac{1}{\sqrt{N}} \int_0^{4U} \log N(\mathcal{H}, d_\infty, \rho') d\rho',$$

and we hence deduce that

$$\Pr\left(\sup_{\theta\in\Theta} |Z_N(\theta) - Z_N(\theta_\dagger)| \geq \bar{L}\Big[J_2(\mathcal{H}) + \sigma\sqrt{x} + (J_\infty(\mathcal{H}) + Ux)/\sqrt{N}\Big]\right) \leq e^{-x} \tag{2.95}$$

for some universal constant $\bar{L}$.

Now what precedes also implies the classical Bernstein-inequality

$$\Pr\left(|Z_N(\theta)| \geq \sigma\sqrt{2x} + \frac{Ux}{3\sqrt{N}}\right) \leq 2e^{-x}, \ x > 0, \tag{2.96}$$

for any fixed $\theta \in \Theta, U \geq \|h_\theta\|_\infty$ and $\sigma^2 \geq E^X h_\theta^2(X)$, proved as (3.24) in [72], using (2.94). Applying this with $\theta_\dagger$ and using (2.95), the final result follows now from

$$\Pr\big(\sup_{\theta\in\Theta} |Z_N(\theta)| > 2\tau(x)\big) \leq \Pr\big(\sup_{\theta\in\Theta} |Z_N(\theta) - Z_N(\theta_\dagger)| > \tau(x)\big) + \Pr\big(|Z_N(\theta_\dagger)| > \tau(x)\big)\big) \leq 2e^{-x},$$

for any $x \geq 1$, where $\tau(x) = \bar{L}\big[J_2(\mathcal{H}) + \sigma\sqrt{x} + (J_\infty(\mathcal{H}) + Ux)/\sqrt{N}\big]$ and $L \geq 2\bar{L} > 0$ is large enough. □

2.3.6 Proofs for Section 2.3.3

We apply the results from Section 2.5.1 to $\mu = \tilde{\Pi}(\cdot|Z^{(N)})$.

Proof of Theorem 2.3.7. For any $\theta, \bar{\theta} \in \mathbb{R}^D$, we have for the log-prior density that

$$\begin{aligned} \|\nabla \log \pi(\theta) - \nabla \log \pi(\bar{\theta})\|_{\mathbb{R}^D} = \|\Sigma^{-1}(\theta - \bar{\theta})\|_{\mathbb{R}^D} \leq \lambda_{max}(\Sigma^{-1})\|\theta - \bar{\theta}\|_{\mathbb{R}^D}, \\ \lambda_{min}(-\nabla^2 \log \pi(\theta)) \geq \lambda_{min}(\Sigma^{-1}), \end{aligned}$$

and for the likelihood surrogate $\tilde{\ell}_N$, by Proposition 2.3.6 and on the event $\mathcal{E}$, that

$$\begin{aligned} \|\nabla \tilde{\ell}_N(\theta) - \nabla \tilde{\ell}_N(\bar{\theta})\|_{\mathbb{R}^D} \leq 7K\lambda_{max}(M)\|\theta - \bar{\theta}\|_{\mathbb{R}^D}, \\ \lambda_{min}(-\nabla^2 \tilde{\ell}_N(\theta)) \geq Nc_{min}/2. \end{aligned}$$

Combining the last two displays, and on the event $\mathcal{E}$, we can verify Assumption 2.5.1 below for $-\log d\tilde{\Pi}(\cdot|Z^{(N)})$ with constants

$$m = Nc_{min}/2 + \lambda_{min}(\Sigma^{-1}), \qquad \Lambda = 7K\lambda_{max}(M) + \lambda_{max}(\Sigma^{-1}).$$

We may thus apply Proposition 2.5.4 to obtain,

$$
\begin{aligned}
W_2^2(\mathcal{L}(\vartheta_k), \Pi(\cdot|Z^{(N)})) &\leq 2W_2^2(\Pi(\cdot|Z^{(N)}), \tilde{\Pi}(\cdot|Z^{(N)})) + 2W_2^2(\mathcal{L}(\vartheta_k), \tilde{\Pi}(\cdot|Z^{(N)})) \\
&\leq \rho + b(\gamma) + 4(1 - m\gamma/2)^k \Big[\|\theta_{init} - \theta_{max}\|_{\mathbb{R}^D}^2 + \frac{D}{m}\Big],
\end{aligned}
$$

where θ_{max} denotes the unique maximiser of $\log d\tilde{\Pi}(\cdot|Z^{(N)})$ over $\mathbb{R}^D$ (which exists on the event $\mathcal{E}_{conv}$, by virtue of strong concavity).

We conclude by an estimate for $\|\theta_{init} - \theta_{max}\|_{\mathbb{R}^D}$. To start, notice that for any $\theta \in \mathbb{R}^D$ we have

$$
|\theta - \theta_{init}|_1^2 = (\theta - \theta_{init})^T M(\theta - \theta_{init}) \geq \lambda_{min}(M)\|\theta - \theta_{init}\|_{\mathbb{R}^D}^2. \tag{2.97}
$$

Thus, for any $\theta \in \mathbb{R}^D$ with $\|\theta - \theta_{init}\|_{\mathbb{R}^D}^2 \geq 4\eta^2/\lambda_{min}(M)$, we have that $|\theta - \theta_{init}|_1 \geq 2\eta$, and therefore also that $g_\eta(\theta) \geq \big(|\theta - \theta_{init}|_1 - \eta\big)^2 \geq \frac{1}{4}|\theta - \theta_{init}|_1^2$. Thus, for C from (2.55) and any $\theta \in \mathbb{R}^D$ satisfying

$$
\|\theta - \theta_{init}\|_{\mathbb{R}^D}^2 \geq \frac{20}{C} + \frac{4\eta^2}{\lambda_{min}(M)},
$$

using (2.97), (2.55) as well as the upper bound for $|\ell_N(\theta)|$ in the definition of $\mathcal{E}_{conv}$, we obtain

$$
\begin{aligned}
-\tilde{\ell}_N(\theta) = Kg_\eta(\theta) &\geq CN(c_{max} + 1)\frac{1 + \lambda_{max}(M)/\eta^2}{\lambda_{min}(M)} \cdot \frac{|\theta - \theta_{init}|_1^2}{4} \\
&\geq \frac{C}{4}N(c_{max} + 1)\|\theta - \theta_{init}\|_{\mathbb{R}^D}^2 \\
&\geq 5N(c_{max} + 1) \geq -\tilde{\ell}_N(\theta_{init}).
\end{aligned}
$$

This implies that necessarily the unique maximiser $\theta_{\tilde{\ell}}$ of the (on $\mathcal{E}_{conv}$) strongly concave map $\tilde{\ell}_N$ over $\mathbb{R}^D$ satisfies $\|\theta_{\tilde{\ell}} - \theta_{init}\|_{\mathbb{R}^D}^2 \leq 20/C + 4\eta^2/\lambda_{min}(M)$. Moreover, in view of the definition of $\mathcal{B}$ and the hypotheses on θ^* we have that

$$
\|\theta_{init}\|_{\mathbb{R}^D} \leq \|\theta_{init} - \theta^*\|_{\mathbb{R}^D} + \|\theta^*\|_{\mathbb{R}^D} \leq \frac{|\theta_{init} - \theta^*|_1}{\sqrt{\lambda_{min}(M)}} + R \leq \frac{\eta}{\sqrt{\lambda_{min}(M)}} + R,
$$

which also allows us to deduce

$$
\|\theta_{\tilde{\ell}}\|_{\mathbb{R}^D} \leq \|\theta_{\tilde{\ell}} - \theta_{init}\|_{\mathbb{R}^D} + \|\theta_{init}\|_{\mathbb{R}^D} \leq \sqrt{20/C} + \frac{3\eta}{\sqrt{\lambda_{min}(M)}} + R.
$$

We further have that $\theta_{max}^T\Sigma^{-1}\theta_{max} \leq \theta_{\tilde{\ell}}^T\Sigma^{-1}\theta_{\tilde{\ell}}$ (otherwise θ_{max} would not be a maximiser of $\log d\tilde{\Pi}(\cdot|Z^{(N)})$) and thus, for $\kappa(\Sigma)$ the condition number of Σ,

$$
\|\theta_{max}\|_{\mathbb{R}^D}^2 \leq \frac{1}{\lambda_{min}(\Sigma^{-1})}\theta_{max}^T\Sigma^{-1}\theta_{max} \leq \frac{1}{\lambda_{min}(\Sigma^{-1})}\theta_{\tilde{\ell}}^T\Sigma^{-1}\theta_{\tilde{\ell}} \leq \kappa(\Sigma)\|\theta_{\tilde{\ell}}\|_{\mathbb{R}^D}^2.
$$

Combining the preceding displays, the proof is now completed as follows:

$$\begin{aligned}\|\theta_{max}-\theta_{init}\|^2_{\mathbb{R}^D} &\lesssim \|\theta_{max}\|^2_{\mathbb{R}^D}+\|\theta_{init}\|^2_{\mathbb{R}^D}\\ &\lesssim \kappa(\Sigma)\|\theta_{\tilde{\ell}}\|^2_{\mathbb{R}^D}+\frac{\eta^2}{\lambda_{min}(M)}+R^2\\ &\lesssim \kappa(\Sigma)\Big[1+\frac{\eta^2}{\lambda_{min}(M)}+R^2\Big].\end{aligned}$$

Proof of Theorem 2.3.8. For any $t \geq 0$ and any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$ we have

$$\begin{aligned}&\mathbf{P}_{\theta_{init}}\Big(|\hat{\pi}^J_{J_{in}}(H)-E^\Pi[H|Z^{(N)}]|\geq t\Big)\\ &\quad\leq \mathbf{P}_{\theta_{init}}\Big(|\hat{\pi}^J_{J_{in}}(H)-\mathbf{E}_{\theta_{init}}[\hat{\pi}^J_{J_{in}}(H)]|\geq t-|\mathbf{E}_{\theta_{init}}[\hat{\pi}^J_{J_{in}}(H)]-E^\Pi[H|Z^{(N)}]|\Big).\end{aligned} \tag{2.98}$$

To further estimate the right side, note that for any $k \geq J_{in}$, by (2.64) and Theorem 2.3.7, we have

$$W_2^2(\mathcal{L}(\vartheta_k),\Pi(\cdot|Z^{(N)}))\leq 2(\rho+b(\gamma)).$$

Noting that (2.167) below in fact holds for any probability measure μ and thus in particular for $\mu=\Pi(\cdot|Z^{(N)})$, it follows that for any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$,

$$\big(\mathbf{E}_{\theta_{init}}[\hat{\pi}^J_{J_{in}}(H)]-E^\Pi[H|Z^{(N)}]\big)^2\leq 2\|H\|^2_{Lip}(\rho+b(\gamma)).$$

Thus if $t \geq 0$ satisfies (2.65), then applying Proposition 2.5.3 to both H and $-H$ yields that the r.h.s. in (2.98) is further bounded by

$$\mathbf{P}_{\theta_{init}}\Big(|\hat{\pi}^J_{J_{in}}(H)-\mathbf{E}_{\theta_{init}}[\hat{\pi}^J_{J_{in}}(H)]|\geq t/2\Big)\leq 2\exp\Big(-c\frac{t^2m^2J\gamma}{\|H\|^2_{Lip}(1+1/(mJ\gamma))}\Big).$$

Proof of Corollary 2.3.9. We first estimate the probability to be bounded by

$$\mathbf{P}_{\theta_{init}}\Big(\|\bar{\theta}^J_{J_{in}}-\mathbf{E}_{\theta_{init}}[\bar{\theta}^J_{J_{in}}]\|_{\mathbb{R}^D}\geq t-\|\mathbf{E}_{\theta_{init}}[\bar{\theta}^J_{J_{in}}]-E^\Pi[\theta|Z^{(N)}]\|_{\mathbb{R}^D}\Big).$$

Next, for any $k \geq 1$, let ν_k denote an optimal coupling between $\mathcal{L}(\vartheta_k)$ and $\Pi[\cdot|Z^{(N)}]$ (cf. Theorem 4.1 in [176]). Then by Jensen's inequality and the definition of W_2 from (2.9),

$$\begin{aligned}
\|\mathbf{E}_{\theta_{init}}[\bar{\theta}^J_{J_{in}}] - E^\Pi[\theta|Z^{(N)}]\|^2_{\mathbb{R}^D} &= \Big\| \frac{1}{J} \sum_{k=J_{in}+1}^{J_{in}+J} \int_{\mathbb{R}^D \times \mathbb{R}^D} (\theta - \theta') d\nu_k(\theta, \theta') \Big\|^2_{\mathbb{R}^D} \\
&= \sum_{j=1}^{D} \Big(\frac{1}{J} \sum_{k=J_{in}+1}^{J_{in}+J} \int_{\mathbb{R}^D \times \mathbb{R}^D} (\theta_j - \theta'_j) d\nu_k(\theta, \theta') \Big)^2 \\
&\leq \frac{1}{J} \sum_{k=J_{in}+1}^{J_{in}+J} \int_{\mathbb{R}^D \times \mathbb{R}^D} \sum_{j=1}^{D} (\theta_j - \theta'_j)^2 d\nu_k(\theta, \theta') \\
&= \frac{1}{J} \sum_{k=J_{in}+1}^{J_{in}+J} W_2^2(\mathcal{L}(\vartheta_k), \Pi[\cdot|Z^{(N)}]).
\end{aligned}$$

Thus we obtain from (2.61), (2.64) (as after (2.98)) that

$$\|E_{\theta_{init}}[\bar{\theta}^J_{J_{in}}] - E^\Pi[\theta|Z^{(N)}]\|_{\mathbb{R}^D} \leq \sqrt{2}\sqrt{\rho + b(\gamma)}.$$

Now for any $j = 1, ..., d$, let us write $H_j : \mathbb{R}^D \to \mathbb{R},\ \theta \mapsto \theta_j$, for the j-the coordinate projection map, of Lipschitz constant 1. Then in the notation (2.63) we can write

$$[\bar{\theta}^J_{J_{in}}]_j = \hat{\pi}^J_{J_{in}}(H_j), \qquad j = 1, ..., D.$$

For $t \geq \sqrt{8(\rho + b(\gamma))}$ and applying Proposition 2.5.3 as in the proof of Theorem 2.3.8 as well as a union bound gives

$$\begin{aligned}
\mathbf{P}_{\theta_{init}}\Big(\|\bar{\theta}^J_{J_{in}} - E^\Pi[\theta|Z^{(N)}]\|_{\mathbb{R}^D} \geq t\Big) &\leq \mathbf{P}_{\theta_{init}}\Big(\|\bar{\theta}^J_{J_{in}} - \mathbf{E}_{\theta_{init}}[\bar{\theta}^J_{J_{in}}]\|_{\mathbb{R}^D} \geq t/2\Big) \\
&= \mathbf{P}_{\theta_{init}}\Big(\sum_{j=1}^{D} \Big[\hat{\pi}^J_{J_{in}}(H_j) - \mathbf{E}_{\theta_{init}}[\hat{\pi}^J_{J_{in}}(H_j)]\Big]^2 \geq \frac{t^2}{4}\Big) \\
&\leq \sum_{j=1}^{D} \mathbf{P}_{\theta_{init}}\Big(\Big[\hat{\pi}^J_{J_{in}}(H_j) - \mathbf{E}_{\theta_{init}}[\hat{\pi}^J_{J_{in}}(H_j)]\Big]^2 \geq \frac{t^2}{4D}\Big) \\
&\leq 2D \exp\Big(- c \frac{t^2 m^2 J\gamma}{D\big[1 + 1/(mJ\gamma)\big]}\Big),
\end{aligned}$$

completing the proof of the corollary.

2.4 Proofs for the Schrödinger model

In this section, we will show how the results from Section 2.3 can be applied to the nonlinear problem for the Schrödinger equation (2.17). Recalling the notation of Sections 2.2 and 2.3, we will set $\theta^* = \theta_{0,D}$, the norm $|\cdot|_1 := \|\cdot\|_{\mathbb{R}^D}$ as well as $\eta := \epsilon D^{-4/d}$ (for ϵ to be chosen), such that the region $\mathcal{B}$ from (2.39) equals the Euclidean ball

$$\mathcal{B}_\epsilon := \Big\{\theta \in \mathbb{R}^D : \|\theta - \theta_{0,D}\|_{\mathbb{R}^D} < \epsilon D^{-4/d}\Big\}. \tag{2.99}$$

The first key observation is the following result on the local log-concavity of the likelihood function on $\mathcal{B}_\epsilon$, which will be proved by a combination of the concentration result Lemma 2.3.4 with the PDE estimates below, notably the 'average curvature' bound from Lemma 2.4.7.

Proposition 2.4.1. *Let $\theta_0 \in h^2$ satisfy $\|\theta_0\|_{h^2} \leq S$ for some $S > 0$ and consider ℓ_N from (2.22) with forward map $\mathcal{G} : \mathbb{R}^D \to \mathbb{R}$ from (2.17). Then there exist constants $0 < \epsilon_S = \epsilon_S(\mathcal{O}, g, \Phi) \leq 1$ and $c_1, c_2, c_3, c_4 > 0$ such that for any $\epsilon \leq \epsilon_S$ and all D, N satisfying $D \leq c_2 N^{\frac{d}{d+12}}$ as well as $\|\mathcal{G}(\theta_0) - \mathcal{G}(\theta_{0,D})\|_{L^2(\mathcal{O})} \leq c_1 D^{-4/d}$, the event*

$$\begin{aligned}\mathcal{E}_{conv}(\epsilon) = \Big\{ &\inf_{\theta\in\mathcal{B}_\epsilon} \lambda_{min}\big(-\nabla^2 \ell_N(\theta)\big) > c_3 N D^{-4/d}, \\ &\sup_{\theta\in\mathcal{B}_\epsilon} \Big[|\ell_N(\theta)| + \|\nabla \ell_N(\theta)\|_{\mathbb{R}^D} + \|\nabla^2 \ell_N(\theta)\|_{op}\Big] < c_4 N\Big\}\end{aligned}$$

satisfies

$$P^N_{\theta_0}\big(\mathcal{E}_{conv}(\epsilon)\big) \geq 1 - 33 e^{-c_2 N^{\frac{d}{d+12}}}. \tag{2.100}$$

Proof. For any $\theta \in \mathbb{R}^D$, F_θ as in (2.16), by a Sobolev embedding and (2.13), we have $\|F_\theta\|_\infty \lesssim \|\theta\|_{h^2} \lesssim D^{2/d}\|\theta\|_{\mathbb{R}^D}$. This and the Lemmas 2.4.4, 2.4.5, 2.4.6 verify Assumption 2.3.2 in the present setting, with constants

$$k_0 \simeq k_1 \simeq \text{const.}, \quad k_2 \simeq m_0 \simeq m_1 \simeq D^{2/d}, \quad m_2 \simeq D^{4/d},$$

whence the constants from (2.46) satisfy

$$C_\mathcal{G} \simeq D^{4/d}, \quad C'_\mathcal{G} \simeq D^{2/d}, \quad C''_\mathcal{G} \simeq D^{2/d}, \quad C'''_\mathcal{G} \simeq \text{const.}.$$

Moreover, Lemmas 2.4.7 and 2.4.8 verify Assumption 2.3.3 for our choice of η with

$$c_{min} \simeq D^{-4/d}, \quad c_{max} \simeq \text{const.} \tag{2.101}$$

Then the minimum (2.45) is dominated by the third term, yielding that

$$\mathcal{R}_N = \mathcal{R}_{N,D} \simeq c^2_{min}/C'^2_\mathcal{G} \simeq N D^{-12/d}.$$

Therefore, we can choose $c > 0$ small enough such that for any $D, N \in \mathbb{N}$ satisfying $D \leq c N^{d/(d+12)}$, we also have $D \leq \mathcal{R}_{N,D}$. Lemma 2.3.4 then implies that for all such D, N, we have

$$P_{\theta_0}^N(\mathcal{E}_{conv}^c) \leq 32e^{-\mathcal{R}_N} + e^{-N/8} \leq 33e^{-cN^{\frac{d}{d+12}}}. \tag{2.102}$$

□

Next, if θ_{init} is the estimator from Theorem 2.5.10, then in the present setting with $\epsilon = 1/\log N$, the event (2.50) equals

$$\mathcal{E}_{init} = \Big\{\|\theta_{init} - \theta_{0,D}\|_{\mathbb{R}^D} \leq \frac{1}{8(\log N)D^{4/d}}\Big\}.$$

Proposition 2.4.2. *Assuming Condition 2.2.3, there exist constants* $c_5, c_6 > 0$ *such that for all* $N \in \mathbb{N}$*,*

$$P_{\theta_0}^N(\mathcal{E}_{init}) \geq 1 - c_5 e^{-c_6 N^{d/(2\alpha+d)}}.$$

Proof. Using Theorem 2.5.10 and $\alpha > 6$, we obtain that with sufficiently high probability,

$$\|\theta_{init} - \theta_{0,D}\|_{\mathbb{R}^D} \lesssim N^{-(\alpha-2)/(2\alpha+d)} = o\big((\log N)^{-1} D^{-4/d}\big).$$

□

Next, denoting by $\tilde{\Pi}(\cdot|Z^{(N)})$ the 'surrogate' posterior measure with density (2.27), and if

$$\mathcal{E}_{wass} = \Big\{W_2^2(\tilde{\Pi}(\cdot|Z^{(N)}), \Pi(\cdot|Z^{(N)})) \leq \exp(-N^{d/(2\alpha+d)})/2\Big\},$$

is given by (2.59) with $\rho = \exp(-N^{d/(2\alpha+d)})$, then Theorem 2.4.14 implies the following approximation result in Wasserstein distance.

Proposition 2.4.3. *Assume Conditions 2.2.2 and 2.2.3. Then there exist constants* $c_7, c_8 > 0$ *such that for all* $N \in \mathbb{N}$*,*

$$P_{\theta_0}^N(\mathcal{E}_{wass}) \geq 1 - c_7 e^{-c_8 N^{d/(2\alpha+d)}}.$$

The preceding propositions imply that the events

$$\mathcal{E}_N := \mathcal{E}_{conv} \cap \mathcal{E}_{init} \cap \mathcal{E}_{wass} \tag{2.103}$$

satisfy the probability bound $P_{\theta_0}^N(\mathcal{E}_N) \geq 1 - c' e^{-c'' N^{d/(2\alpha+d)}}$. In what follows, the events $\mathcal{E}_N$ will be tacitly further intersected with events which have probability 1 for all N large enough, ensuring that the non-asymptotic conditions required in the results of Section 2.3 are eventually verified.

Proof of Theorem 2.2.7. We will prove Theorem 2.2.7 by applying Theorem 2.3.7 with the choices $\mathcal{B} = \mathcal{B}_\epsilon$ from (2.99), $\epsilon = 1/\log N$ and K from Condition 2.2.2, $\rho = \exp(-N^{d/(2\alpha+d)})$ and $M = I_{D\times D}$ generating the ellipsoidal norm $\|\cdot\|_{\mathbb{R}^D}$. Using (2.13), the prior covariance Σ

from (2.23) satisfies

$$\lambda_{min}(\Sigma^{-1}) \simeq N^{\frac{d}{2\alpha+d}}, \qquad \lambda_{max}(\Sigma^{-1}) \simeq N^{\frac{d}{2\alpha+d}} D^{2\alpha/d}.$$

Then using Condition 2.2.2, we first have that

$$K \gtrsim ND^{8/d}(\log N)^2 \simeq Nc_{max} \cdot (1+\eta^{-2}),$$

verifying the lower bound (2.55), and then also that $m, \Lambda > 0$ from Theorem 2.3.7 satisfy

$$m \simeq ND^{-4/d} + N^{\frac{d}{2\alpha+d}}, \quad \Lambda \simeq ND^{8/d}(\log N)^3 + N^{\frac{d}{2\alpha+d}} D^{\frac{2\alpha}{d}}.$$

The dimension condition (2.28) and the condition on α further imply

$$ND^{-4/d} \gtrsim N^{\frac{d}{2\alpha+d}}, \qquad N^{\frac{d}{2\alpha+d}} D^{\frac{2\alpha}{d}} \lesssim N,$$

whence we further obtain

$$m \simeq ND^{-4/d}, \qquad \Lambda \simeq ND^{8/d}(\log N)^3. \tag{2.104}$$

Noting that also $\gamma = o(\Lambda^{-1})$ with our choices, Theorem 2.3.7 yields that on the event $\mathcal{E}_N$ from (2.103), the Markov chain (ϑ_k) satisfies the Wasserstein bound (2.61) with

$$b(\gamma) \lesssim \frac{\gamma D\Lambda^2}{m^2} + \frac{\gamma^2 D\Lambda^4}{m^3} \lesssim \gamma D^{(d+24)/d}(\log N)^6 + \gamma^2 ND^{(d+44)/d}(\log N)^{12}, \tag{2.105}$$

as well as

$$\tau(\Sigma, M, \|\theta_{0,D}\|_{\mathbb{R}^D}) \lesssim \kappa(\Sigma) \simeq D^{2\alpha/d}.$$

Using also that $D/m \lesssim$ const., the first part of Theorem 2.2.7 follows.

For the choice of $\gamma = \gamma_\varepsilon$ from (2.31), straightforward calculation yields that (for N large enough)

$$B(\gamma_\varepsilon) = o(\varepsilon^2 + N^{-2P}), \tag{2.106}$$

which proves the second part of Theorem 2.2.7.

Proof of Proposition 2.2.4 and of Theorems 2.2.5, 2.2.6. The proof of Proposition 2.2.4 now follows directly from Theorem 2.3.8 and the preceding computations. Noting that for all N large enough we have $B(\gamma) \leq N^{-P}$, Theorem 2.2.5 follows from Corollary 2.3.9, (2.106) as well as (2.67), for $J_{in} \geq (\log N)^3/(\gamma_\varepsilon ND^{-4/d})$. Finally, intersecting further with the event

$$\mathcal{E}_{mean} := \{\|E^\Pi[\theta|Z^{(N)}] - \theta_0\|_{\ell_2} \leq LN^{-\frac{\alpha}{2\alpha+d}\frac{\alpha}{\alpha+2}}\}, \quad L > 0,$$

Theorem 2.2.6 now follows from the triangle inequality and (2.153).

Proof of Theorem 2.2.8. In the proof we intersect $\mathcal{E}_N$ from (2.103) further with the event on which the conclusion of Theorem 3.14 holds. Part iii) then follows from part ii) and straightforward calculations. Part i) follows from the arguments following (2.159) below, where it is proved in particular that $\hat{\theta}_{MAP}$ is the unique maximiser of the proxy posterior density $\tilde{\pi}(\cdot|Z^{(N)})$ over $\mathbb{R}^D$. We can now apply Proposition 2.5.2 with m, Λ from (2.104), using also that

$$\begin{aligned}
&|\log \tilde{\pi}(\theta_{init}|Z^{(N)}) - \log \tilde{\pi}(\hat{\theta}_{MAP}|Z^{(N)})| \\
&\lesssim \sup_{\theta \in \mathcal{B}_{1/8\log N}} |\ell_N(\theta)| + N^{d/(2\alpha+d)} \|\hat{\theta}_{MAP}\|^2_{h^\alpha} + N^{d/(2\alpha+d)} \|\theta_{init}\|^2_{h^\alpha} \\
&\lesssim N + N^{d/(2\alpha+d)}(1 + D^{2\alpha/d}) \lesssim N,
\end{aligned}$$

in view of $\ell_N = \tilde{\ell}_N$ on $\mathcal{B}_{1/8\log N}$, the definition of $\mathcal{E}_{init}$, (2.13) and since $\theta_0 \in h^\alpha$.

2.4.1 Analytical properties of the Schrödinger forward map

This section is devoted to proving the four auxiliary Lemmas 2.4.5-2.4.8 used in the proof of Proposition 2.4.1. Throughout we consider forward map $\mathcal{G} : \mathbb{R}^D \to L^2(\mathcal{O})$, $\mathcal{G} = G \circ \Phi^* \circ \Psi$ given by (2.17) and assume the hypotheses of Proposition 2.4.1, where the set $\mathcal{B}_\epsilon$ was defined in (2.99).

For any $f \in C(\mathcal{O})$ with $f \geq 0$, by standard theory for elliptic PDEs (see e.g. Chapter 6.3 of [65]) there exists a linear, continuous operator $V_f : L^2(\mathcal{O}) \to H_0^2(\mathcal{O})$ describing (weak) solutions $V_f[\psi] = w \in H_0^2$ of the (inhomogeneous) Schrödinger equation

$$\begin{cases} \frac{\Delta}{2} w - f w = \psi & \text{on } \mathcal{O}, \\ w = 0 & \text{on } \partial\mathcal{O}. \end{cases} \tag{2.107}$$

Lemma 2.4.4. *For any $x \in \mathcal{O}$, the map $\theta \mapsto \mathcal{G}(\theta)(x)$ is twice continuously differentiable on $\mathbb{R}^D$. The vector field $\nabla\mathcal{G}_\theta : \mathcal{O} \to \mathbb{R}^D$ is given by*

$$v^T \nabla \mathcal{G}_\theta(x) = V_{f_\theta}\big[u_{f_\theta}(\Phi' \circ F_\theta)\Psi(v)\big](x), \quad x \in \mathcal{O},\ v \in \mathbb{R}^D.$$

Moreover, for any $v_1, v_2 \in \mathbb{R}^D$ and $x \in \mathcal{O}$, the matrix field $\nabla^2\mathcal{G}_\theta : \mathcal{O} \to \mathbb{R}^{D\times D}$ is given by

$$\begin{aligned}
v_1^T \nabla^2 \mathcal{G}_\theta(x) v_2 =& V_{f_\theta}\big[u_{f_\theta}\Psi(v_1)\Psi(v_2)(\Phi'' \circ F_\theta)\big](x) \\
&+ V_{f_\theta}\big[(\Phi' \circ F_\theta)\Psi(v_1) V_{f_\theta}\big[u_{f_\theta}(\Phi' \circ F_\theta)\Psi(v_2)\big]\big](x) \\
&+ V_{f_\theta}\big[(\Phi' \circ F_\theta)\Psi(v_2) V_{f_\theta}\big[u_{f_\theta}(\Phi' \circ F_\theta)\Psi(v_1)\big]\big](x).
\end{aligned}$$

Proof. In the notation from (2.17), the map $\theta \mapsto \mathcal{G}(\theta)(x)$ can be represented as the composition $\delta_x \circ G \circ \Phi^* \circ \Psi$, where $\delta_x : w \mapsto w(x)$ denotes point evaluation. We first show that each of these four operators is twice differentiable. The continuous linear maps $\Psi : \mathbb{R}^D \to C(\mathcal{O})$

and $\delta_x : C(\mathcal{O}) \to \mathbb{R}$ are infinitely differentiable (in the Frechét sense). Moreover, the maps $G : C(\mathcal{O}) \cap \{f > 0\} \to C(\mathcal{O})$ and $\Phi^* : C(\mathcal{O}) \to C(\mathcal{O}) \cap \{f > 0\}$ are twice Frećhet differentiable with derivatives DG, DG^2 and $D\Phi^*, D^2\Phi^*$ given by Lemma 2.5.6 and (2.177) respectively. We deduce overall by the chain rule for Fréchet derivates (cf. Lemma 2.5.7), that $x \mapsto \mathcal{G}(\theta)(x)$ is twice differentiable, with the desired expressions for the vector and matrix fields. The continuity of the second partial derivatives follows from inspection of the expression for the matrix field, and by applying the regularity results for V_f, G and Φ^* from Section 2.5. $\square$

Now since $\|\theta_0\|_{h^2} \leq S$ and by the definition (2.99) of the set $\mathcal{B}_1$, we have from (2.13) that

$$\sup_{\theta \in \mathcal{B}_1} \|\theta\|_{h^2} \leq \|\theta_{0,D}\|_{h^2} + \sup_{\theta \in \mathcal{B}_1} \|\theta - \theta_{0,D}\|_{h^2} \lesssim S + D^{\frac{2}{d}} \sup_{\theta \in \mathcal{B}_1} \|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \lesssim S + 1.$$

It follows further from the Sobolev embedding and regularity of the link function Φ (Section 2.5.2.1) that there exists a constant $B = B(S, \Phi, \mathcal{O}) < \infty$, such that

$$\sup_{\theta \in \mathcal{B}_1} \Big[\|F_\theta\|_\infty + \|F_\theta\|_{H^2} + \|f_\theta\|_{H^2} + \|f_\theta\|_\infty \Big] \leq B. \tag{2.108}$$

In particular, this estimate implies that the constants appearing in the inequalities from Lemma 2.5.5 can be chosen independently of $\theta \in \mathcal{B}$, which we use frequently below.

For notational convenience we also introduce spaces

$$E_D := \mathrm{span}(e_1, ..., e_D) \subseteq L^2(\mathcal{O}), \quad D \in \mathbb{N}, \tag{2.109}$$

spanned by the first D eigenfunctions of Δ on $\mathcal{O}$ (cf. Section 2.2.1.1).

We first verify the boundedness property required in Assumption 2.3.2 ii).

Lemma 2.4.5. *There exists a constant $C > 0$ such that*

$$\sup_{\theta \in \mathcal{B}_1} \|\mathcal{G}(\theta)\|_{L^\infty} \leq C, \qquad \sup_{\theta \in \mathcal{B}_1} \|\nabla \mathcal{G}(\theta)\|_{L^\infty(\mathcal{O}, \mathbb{R}^D)} \leq C, \qquad \sup_{\theta \in \mathcal{B}_1} \|\nabla^2 \mathcal{G}(\theta)\|_{L^\infty(\mathcal{O}, \mathbb{R}^{D \times D})} \leq C D^{2/d}.$$

Proof. The estimate for $\|\mathcal{G}(\theta)\|_\infty$ follows immediately from (2.18). To estimate $\|\nabla \mathcal{G}(\theta)\|_{L^\infty(\mathcal{O}, \mathbb{R}^D)}$, we first note that by Lemma 2.4.4,

$$\|\nabla \mathcal{G}(\theta)\|_{L^\infty(\mathcal{O}, \mathbb{R}^D)} = \sup_{v : \|v\|_{\mathbb{R}^D} \leq 1} \|v^T \nabla \mathcal{G}(\theta)\|_{L^\infty} \leq \sup_{H \in E_D : \|H\|_{L^2} \leq 1} \big\|V_{f_\theta}\big[u_{f_\theta}(\Phi' \circ F_\theta) H\big]\big\|_\infty.$$

Thus by the Sobolev embedding $\|\cdot\|_\infty \lesssim \|\cdot\|_{H^2}$, Lemma 2.5.5 and boundedness of Φ', we have that for any $\theta \in \mathcal{B}_1$ and any $H \in E_D$,

$$\begin{aligned} \big\|V_{f_\theta}[u_{f_\theta}(\Phi' \circ F_\theta) H]\big\|_\infty &\lesssim \big\|V_{f_\theta}[u_{f_\theta}(\Phi' \circ F_\theta) H]\big\|_{H^2} \\ &\lesssim \big\|u_{f_\theta}(\Phi' \circ F_\theta) H\big\|_{L^2} \\ &\lesssim \|u_{f_\theta}\|_\infty \|\Phi' \circ F_\theta\|_\infty \|H\|_{L^2} \lesssim \|H\|_{L^2}. \end{aligned}$$

Again using Lemma 2.4.4, we can similarly estimate $\|\nabla^2\mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)}$ by

$$\begin{aligned}\|\nabla^2\mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} &\leq \sup_{v:\|v\|_{\mathbb{R}^D}\leq 1} \|v^T\nabla^2\mathcal{G}(\theta)v\|_{L^\infty} \\ &\leq \sup_{H\in E_D:\|H\|_{L^2}\leq 1} 2\big\|V_{f_\theta}\big[H(\Phi'\circ F_\theta)V_{f_\theta}\big[H(\Phi'\circ F_\theta)u_{f_\theta}\big]\big]\big\|_\infty + \big\|V_{f_\theta}\big[H^2(\Phi''\circ F_\theta)u_{f_\theta}\big]\big\|_\infty \\ &=: \sup_{H\in E_D:\|H\|_{L^2}\leq 1} I + II.\end{aligned} \tag{2.110}$$

Arguing as in the estimate for $\|\nabla\mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)}$, we have that for any $\theta\in\mathcal{B}_1$ and $H\in E_D$,

$$\begin{aligned} I &\lesssim \|H(\Phi'\circ F_\theta)V_{f_\theta}\big[H(\Phi'\circ F_\theta)u_{f_\theta}\big]\|_{L^2} \\ &\lesssim \|H\|_{L^2}\|\Phi'\circ F\|_\infty\|V_f[H(\Phi'\circ F)u_f]\|_\infty \\ &\lesssim \|H\|_{L^2}\|H(\Phi'\circ F)u_f\|_{L^2} \lesssim \|H\|_{L^2}^2,\end{aligned}$$

as well as

$$II \lesssim \|H^2(\Phi''\circ F_\theta)u_{f_\theta}\|_{L^2} \lesssim \|u_{f_\theta}\|_\infty\|\Phi''\circ F_\theta\|_\infty\|H\|_{L^2}\|H\|_\infty \lesssim \|H\|_{L^2}\|H\|_{H^2} \lesssim D^{2/d}\|H\|_{L^2}^2,$$

where we used the basic norm estimate on $E_D\subseteq L^2(\mathcal{O})$ from Lemma 2.4.9. By combining the last three displays, the proof is completed. □

Next, we verify the increment bound needed in Assumption 2.3.2 iii).

Lemma 2.4.6. *There exists a constant $C>0$ such that for any $D\in\mathbb{N}$ and any $\theta,\theta'\in\mathbb{R}^D$,*

$$\|\mathcal{G}(\theta)-\mathcal{G}(\bar\theta)\|_\infty \leq C\|F_\theta - F_{\bar\theta}\|_\infty, \quad \|\mathcal{G}(\theta)-\mathcal{G}(\bar\theta)\|_{L^2} \leq C\|F_\theta - F_{\bar\theta}\|_{L^2}, \tag{2.111}$$

as well as, for any $\theta,\theta'\in\mathcal{B}_1$,

$$\|\nabla\mathcal{G}(\theta)-\nabla\mathcal{G}(\bar\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} \leq C\|F_\theta - F_{\bar\theta}\|_\infty, \tag{2.112}$$

$$\|\nabla^2\mathcal{G}(\theta)-\nabla^2\mathcal{G}(\bar\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \leq CD^{2/d}\|F_\theta - F_{\bar\theta}\|_\infty. \tag{2.113}$$

Proof. The estimate (2.111) follows immediately from (2.173) and (2.179). Now fix any $\theta,\bar\theta\in\mathcal{B}_1$. To ease notation, in what follows we write $F=\Psi(\theta), \bar F = \Psi(\bar\theta)$, $f=\Phi\circ F$ and

$\bar{f} = \Phi \circ \bar{F}$. For (2.112), arguing as in the proof of Lemma 2.4.5, we first have

$$\begin{aligned}
&\|\nabla\mathcal{G}(\theta) - \nabla\mathcal{G}(\bar{\theta})\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} \\
&\quad\le \sup_{v:\|v\|_{\mathbb{R}^D}\le 1} \|v^T(\nabla\mathcal{G}(\theta) - \nabla\mathcal{G}(\bar{\theta}))\|_\infty \\
&\quad\le \sup_{H\in E_D:\|H\|_{L^2}\le 1} \|V_f[H(\Phi'\circ F)u_f] - V_{\bar{f}}[H(\Phi'\circ\bar{F})u_{\bar{f}}]\|_\infty \\
&\quad= \sup_{H\in E_D:\|H\|_{L^2}\le 1} \|(V_f - V_{\bar{f}})[H(\Phi'\circ F)u_f]\|_\infty + \|V_{\bar{f}}[H(\Phi'\circ F - \Phi'\circ\bar{F})u_{\bar{f}}]\|_\infty \\
&\qquad\qquad + \|V_{\bar{f}}[H(\Phi'\circ F)(u_f - u_{\bar{f}})]\|_\infty \\
&\quad=: \sup_{H\in E_D:\|H\|_{L^2}\le 1} I_a + I_b + I_c.
\end{aligned}$$

Now, we fix $H \in E_D$ for the rest of the proof. The term I_a can further be estimated by repeatedly using the Sobolev embedding $\|\cdot\|_\infty \lesssim \|\cdot\|_{H^2}$, Lemma 2.5.5 as well as (2.108) and (2.179):

$$\begin{aligned}
I_a &= \|V_f[(f-\bar{f})V_{\bar{f}}[u_{\bar{f}}(\Phi'\circ F)H]]\|_\infty \\
&\lesssim \|V_f[(f-\bar{f})V_{\bar{f}}[u_{\bar{f}}(\Phi'\circ F)H]]\|_{H^2} \\
&\lesssim \|(f-\bar{f})V_{\bar{f}}[u_{\bar{f}}(\Phi'\circ F)H]\|_{L^2} \\
&\lesssim \|f-\bar{f}\|_\infty \|u_{\bar{f}}(\Phi'\circ\bar{F})H\|_{L^2} \\
&\lesssim \|F-\bar{F}\|_\infty \|H\|_{L^2}.
\end{aligned} \tag{2.114}$$

Similarly, I_b is estimated as follows:

$$I_b \lesssim \|H(\Phi'\circ F - \Phi'\circ\bar{F})u_{\bar{f}}\|_{L^2} \lesssim \|\Phi'\circ F - \Phi'\circ\bar{F}\|_\infty \|u_{\bar{f}}\|_\infty \|H\|_{L^2} \lesssim \|F-\bar{F}\|_\infty \|H\|_{L^2}.$$

Finally, we can similarly estimate

$$I_c \lesssim \|(u_f - u_{\bar{f}})(\Phi'\circ F)H\|_{L^2} \lesssim \|u_f - u_{\bar{f}}\|_\infty \|\Phi'\circ F\|_\infty \|H\|_{L^2} \lesssim \|F-\bar{F}\|_\infty \|H\|_{L^2},$$

where we have also used (2.111). By combining the estimates for I_a, I_b and I_c, we have completed the proof of (2.112).

It remains to prove (2.113). In analogy to (2.110), we may fix any $v \in \mathbb{R}^D$, and it suffices to derive a bound for $v^T(\nabla^2\mathcal{G}(\theta) - \nabla^2\mathcal{G}(\bar{\theta}))v$. To ease notation, let us write $H = \Psi v \in E_D \cong \mathbb{R}^D$, as well as $h = H(\Phi'\circ F)$ and $\bar{h} = H(\Phi'\circ\bar{F})$. Then by Lemma 2.4.4, we have the following decomposition into eight terms:

$$
\begin{aligned}
&v^T(\nabla^2\mathcal{G}(\theta) - \nabla^2\mathcal{G}(\bar\theta))v \\
&= 2V_{\bar f}[\bar h V_{\bar f}[\bar h u_{\bar f}]] - 2V_f[hV_f[hu_f]] + V_{\bar f}[u_{\bar f}H^2(\Phi''\circ\bar F)] - V_f[u_f H^2(\Phi''\circ F)] \\
&= 2(V_{\bar f} - V_f)[\bar h V_{\bar f}[\bar h u_{\bar f}]] + 2V_f[(\bar h - h)V_{\bar f}[\bar h u_{\bar f}]] \\
&\quad + 2V_f[h(V_{\bar f} - V_f)[\bar h u_{\bar f}]] + 2V_f[hV_f[(\bar h - h)u_{\bar f}]] + 2V_f[hV_f[h(u_{\bar f} - u_f)]] \\
&\quad + (V_{\bar f} - V_f)[u_{\bar f}H^2(\Phi''\circ\bar F)] + V_f[(u_{\bar f} - u_f)H^2(\Phi''\circ\bar F)] + V_f[u_f H^2(\Phi''\circ\bar F - \Phi''\circ F)] \\
&=: II_a + II_b + II_c + II_d + II_e + II_f + II_g + II_h.
\end{aligned}
\tag{2.115}
$$

To estimate these terms, we will again repeatedly use (2.108), the regularity estimates from Lemmas 2.5.5- 2.5.6 below, the estimates $\|h\|_{L^2}, \|\bar h\|_{L^2} \lesssim \|H\|_{L^2}$ as well as $\|f-\bar f\|_\infty \lesssim \|F-\bar F\|_\infty$, which all hold uniformly in $\theta\in\mathcal{B}_1$.

Using Lemma 2.5.5, including the estimate (2.171) with $\psi = \bar h V_{\bar f}[\bar h u_{\bar f}]$, we obtain

$$
\begin{aligned}
\|II_a\|_\infty &\lesssim \|f-\bar f\|_\infty \|\bar h V_{\bar f}[\bar h u_{\bar f}]\|_{L^2} \lesssim \|f-\bar f\|_\infty \|\bar h\|_{L^2}\|V_{\bar f}[\bar h u_{\bar f}]\|_\infty \\
&\lesssim \|f-\bar f\|_\infty \|H\|_{L^2}\|\bar h u_{\bar f}\|_{L^2} \lesssim \|f-\bar f\|_\infty \|H\|_{L^2}^2\|u_{\bar f}\|_\infty \\
&\lesssim \|F-\bar F\|_\infty \|H\|_{L^2}^2.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\|II_b\|_\infty &\lesssim \|(\bar h - h)V_{\bar f}[\bar h u_{\bar f}]\|_{L^2} \lesssim \|H(\Phi'\circ\bar F - \Phi'\circ F)\|_{L^2}\|V_{\bar f}[\bar h u_{\bar f}]\|_\infty \\
&\lesssim \|u_f\|_\infty \|H\|_{L^2}\|\bar F - F\|_\infty \|\bar h u_{\bar f}\|_{L^2} \\
&\lesssim \|H\|_{L^2}^2 \|\bar F - F\|_\infty,
\end{aligned}
$$

and, again using (2.171),

$$
\begin{aligned}
\|II_c\|_\infty &\lesssim \|h(V_{\bar f} - V_f)[\bar h u_{\bar f}]\|_{L^2} \lesssim \|h\|_{L^2}\|(V_{\bar f} - V_f)[\bar h u_{\bar f}]\|_\infty \lesssim \|H\|_{L^2}\|\bar f - f\|_\infty \|\bar h u_{\bar f}\|_{L^2} \\
&\lesssim \|H\|_{L^2}^2\|\bar F - F\|_\infty.
\end{aligned}
$$

For II_d, by following similar steps as for II_b, we see that

$$
\|II_d\|_\infty \lesssim \|H\|_{L^2}\|V_f[(\bar h - h)u_{\bar f}]\|_\infty \lesssim \|H\|_{L^2}^2\|\bar F - F\|_\infty,
$$

and similarly, using also (2.111), we obtain

$$
\|II_e\|_\infty \lesssim \|H\|_{L^2}\|V_f[h(u_{\bar f} - u_f)]\|_\infty \lesssim \|H\|_{L^2}^2\|u_{\bar f} - u_f\|_\infty \lesssim \|H\|_{L^2}^2\|\bar F - F\|_\infty.
$$

For the term II_f, we note that by the Sobolev embedding,

$$\|w\|_{(H_0^2)^*} \leq \sup_{\psi:\|\psi\|_{H^2}\leq 1} |\int_{\mathcal{O}} w\psi| \lesssim \|w\|_{L^1} \sup_{\psi:\|\psi\|_{H^2}\leq 1} \|\psi\|_\infty \lesssim \|w\|_{L^1}, \quad w \in L^1(\mathcal{O}),$$

and consequently by Lemma 2.5.5,

$$\begin{aligned}
\|II_f\|_\infty &= \|V_f[(\bar{f}-f)V_{\bar{f}}[u_{\bar{f}}H^2(\Phi''\circ F)]]\|_\infty \\
&\lesssim \|\bar{f}-f\|_\infty \|V_{\bar{f}}[u_{\bar{f}}H^2(\Phi''\circ F)]\|_{L^2} \\
&\lesssim \|\bar{f}-f\|_\infty \|u_{\bar{f}}H^2(\Phi''\circ F)\|_{(H_0^2)^*} \\
&\lesssim \|\bar{f}-f\|_\infty \|u_{\bar{f}}H^2(\Phi''\circ F)\|_{L^1} \\
&\lesssim \|\bar{F}-F\|_\infty \|H\|_{L^2}^2.
\end{aligned}$$

For terms II_g and II_h, by similar steps and additionally using that by Lemma 2.4.9, $\|H\|_\infty \lesssim \|H\|_{H^2} \lesssim D^{2/d}\|H\|_{L^2}$ for any $H \in E_D$, we obtain

$$\|II_g\|_\infty \lesssim \|u_{\bar{f}} - u_f\|_\infty \|H^2\|_{L^2} \|\Phi''\circ\bar{F}\|_\infty \lesssim \|\bar{f}-f\|_\infty \|H\|_{L^2}\|H\|_\infty \lesssim D^{2/d}\|\bar{F}-F\|_\infty \|H\|_{L^2}^2,$$

as well as

$$\|II_h\|_\infty \leq \|u_f H^2(\Phi''\circ\bar{F} - \Phi''\circ F)\|_{L^2} \lesssim \|H\|_{L^2}\|H\|_\infty \|\bar{F}-F\|_\infty \lesssim D^{2/d}\|\bar{F}-F\|_\infty \|H\|_{L^2}^2.$$

By combining (2.115) with the estimates for the terms $II_a - II_h$, the proof of (2.113) is complete. □

We now turn to the key 'geometric' bound from the first part of Assumption 2.3.3, which quantifies the average curvature of the likelihood function ℓ_N near $\theta_{0,D}$ in a high-dimensional setting (when P^X is uniform on $\mathcal{O}$). The curvature deteriorates with rate $D^{-4/d}$ as $D \to \infty$, which is in line with the (local) ill-posedness of the Schrödinger model, and the related fact that the associated 'Fisher information operator' is of the form I^2, with I being the inverse of a second order (elliptic Schrödinger-type) operator (cf. also Section 4 in [131]).

Lemma 2.4.7. *Let $\ell(\theta)$ be as in (2.38) with $\mathcal{G}: \mathbb{R}^D \to \mathbb{R}$ from (2.17), and let $\mathcal{B}_\epsilon$ be as in (2.99). Let $\theta_0 \in h^2$ satisfy $\|\theta_0\|_{h^2} \leq S$ for some $S > 0$. Then there exist constants $0 < \epsilon_S \leq 1, c_1, c_2 > 0$ such that if also $\|\mathcal{G}(\theta_0) - \mathcal{G}(\theta_{0,D})\|_{L^2(\mathcal{O})} \leq c_1 D^{-4/d}$, then for all $D \in \mathbb{N}$ and all $\epsilon \leq \epsilon_S$,*

$$\inf_{\theta\in\mathcal{B}_\epsilon} \lambda_{min}(E_{\theta_0}[-\nabla^2\ell(\theta)]) \geq c_2 D^{-4/d}. \tag{2.116}$$

Proof. We begin by noting that for any $Z = (Y, X) \in \mathbb{R}\times\mathcal{O}$, we have

$$-\nabla^2\ell(\theta, Z) = \nabla\mathcal{G}^X(\theta)\nabla\mathcal{G}^X(\theta)^T - (Y - \mathcal{G}^X(\theta))\nabla^2\mathcal{G}^X(\theta).$$

Using this and Lemma 2.4.4, we obtain that for any $v \in \mathbb{R}^D$, with the previous notation $H = \Psi(v)$ and $h = (\Phi' \circ F_\theta)H$,

$$\begin{aligned} v^T E_{\theta_0}[-\nabla^2 \ell(\theta, Z)]v = \|V_{f_\theta}[u_{f_\theta}(\Phi' \circ F_\theta)H]\|^2_{L^2(\mathcal{O})} &- \langle u_{f_{\theta_0}} - u_{f_\theta}, 2V_{f_\theta}[hV_{f_\theta}[hu_{f_\theta}]]\rangle_{L^2(\mathcal{O})} \\ &- \langle u_{f_{\theta_0}} - u_{f_\theta}, V_{f_\theta}[u_{f_\theta} H^2(\Phi'' \circ F_\theta)]\rangle_{L^2(\mathcal{O})} \\ =: I + II + III. \end{aligned} \tag{2.117}$$

We next derive a lower bound on the term I and upper bounds for the terms II and III, for any fixed $v \in \mathbb{R}^D$.

Lower bound for I. Writing $a_\theta := u_{f_\theta}(\Phi' \circ F_\theta)$, using the elliptic L^2-$(H_0^2)^*$ coercivity estimate (2.170) from Lemma 2.5.5 below as well as (2.108), we have

$$\sqrt{I} = \|V_{f_\theta}[a_\theta H]\|_{L^2(\mathcal{O})} \gtrsim \frac{\|a_\theta H\|_{(H_0^2)^*}}{1 + \|f_\theta\|_\infty} \gtrsim \|a_\theta H\|_{(H_0^2)^*}, \quad \theta \in \mathcal{B}_1. \tag{2.118}$$

The next step is to lower bound a_θ. By Theorem 1.17 in [42], the expected exit time $\tau_{\mathcal{O}}$ featuring in the Feynman-Kac formula (2.12) satisfies the uniform estimate $\sup_{x\in\mathcal{O}} E^x \tau_{\mathcal{O}} \le K(vol(\mathcal{O}), d) < \infty$. Therefore, using also Jensen's inequality and $g \ge g_{min} > 0$, we have that, with B from (2.108),

$$\inf_{\theta\in\mathcal{B}_1} \inf_{x\in\mathcal{O}} u_{f_\theta}(x) \ge g_{min} e^{-BK(vol(\mathcal{O}),d)} =: u_{min} > 0. \tag{2.119}$$

Also, since Φ is a regular link function, for some $k = k(B) > 0$ we have

$$\inf_{\theta\in\mathcal{B}_1} \inf_{x\in\mathcal{O}} [\Phi' \circ F_\theta](x) \ge \inf_{t\in[-k,k]} \Phi'(t) > 0,$$

and therefore for some $a_{min} = a_{min}(\Phi, B, \mathcal{O}, g_{min}) > 0$,

$$\inf_{\theta\in\mathcal{B}_1} \inf_{x\in\mathcal{O}} a_\theta(x) \ge a_{min} > 0. \tag{2.120}$$

We thus obtain, by definition of $(H_0^2)^*$ and the multiplication inequality (2.7) that for some $c = c(a_{min}) > 0$,

$$\|H\|_{(H_0^2)^*} = \|a_\theta a_\theta^{-1} H\|_{(H_0^2)^*} \le \|a_\theta^{-1}\|_{H^2} \|a_\theta H\|_{(H_0^2)^*} \le c(1 + \|a_\theta\|^2_{H^2})\|a_\theta H\|_{(H_0^2)^*}, \tag{2.121}$$

where in the last inequality we used (2.178) for the function $x \mapsto 1/x$. Using again (2.108), regularity of Φ', the chain rule as well as the elliptic regularity estimate (2.175), we obtain that

$$\sup_{\theta\in\mathcal{B}_1} \|a_\theta\|_{H^2} \le \sup_{\theta\in\mathcal{B}_1} \|u_{f_\theta}\|_{H^2} \sup_{\theta\in\mathcal{B}_1} \|\Phi' \circ F_\theta\|_{H^2} \le C(g, S, \mathcal{O}, \Phi) < \infty. \tag{2.122}$$

Therefore, combining the displays (2.118), (2.121), (2.122), we have proved that, uniformly in $\theta \in \mathcal{B}_1$,

$$I \gtrsim \|a_\theta H\|^2_{(H_0^2)^*} \gtrsim \frac{\|H\|^2_{(H_0^2)^*}}{c^2 \sup_{\theta\in\mathcal{B}_1}(1+\|a_\theta\|^2_{H^2})^2} \gtrsim D^{-4/d}\|H\|^2_{L^2}, \tag{2.123}$$

where we have used Lemma 2.4.9 below in the last inequality.

Upper bound for II and III. Using the self-adjointness of V_{f_θ} on $L^2(\mathcal{O})$, a Sobolev embedding, Lemma 2.5.5, (2.108), the Lipschitz estimate (2.173) as well as (2.18), we have uniformly in $\theta \in \mathcal{B}_1$,

$$\begin{aligned}
|II| &\lesssim \Big|\int_{\mathcal{O}} (u_{f_{\theta_0}} - u_{f_\theta}) V_{f_\theta}[hV_{f_\theta}[hu_{f_\theta}]]\Big| = \Big|\int_{\mathcal{O}} V_{f_\theta}[u_{f_{\theta_0}} - u_{f_\theta}][hV_{f_\theta}[hu_{f_\theta}]]\Big| \\
&\lesssim \|V_{f_\theta}[u_{f_{\theta_0}} - u_{f_\theta}]\|_\infty \|hV_{f_\theta}[hu_{f_\theta}]\|_{L^1} \\
&\lesssim \|u_{f_{\theta_0}} - u_{f_\theta}\|_{L^2}\|h\|_{L^2}\|V_{f_\theta}[hu_{f_\theta}]\|_{L^2} \\
&\lesssim \|u_{f_{\theta_0}} - u_{f_\theta}\|_{L^2}\|H\|^2_{L^2}.
\end{aligned} \tag{2.124}$$

Similarly, for the term III, using also $\|\Phi''\|_\infty < \infty$, we estimate

$$\begin{aligned}
|III| &= \big|\langle u_{f_{\theta_0}} - u_{f_\theta}, V_{f_\theta}[u_{f_\theta}H^2(\Phi''\circ F_\theta)]\rangle_{L^2(\mathcal{O})}\big| \\
&= \big|\langle V_{f_\theta}[u_{f_{\theta_0}} - u_{f_\theta}], u_{f_\theta}H^2(\Phi''\circ F_\theta)\rangle_{L^2(\mathcal{O})}\big| \\
&\le \|V_{f_\theta}[u_{f_{\theta_0}} - u_{f_\theta}]\|_\infty \|u_{f_\theta}\|_\infty \|\Phi''\circ F_\theta\|_\infty \|H^2\|_{L^1} \\
&\lesssim \|u_{f_{\theta_0}} - u_{f_\theta}\|_{L^2}\|H\|^2_{L^2}
\end{aligned} \tag{2.125}$$

Combining the displays (2.117), (2.123), (2.124) and (2.125), we have proved that for any $\theta \in \mathcal{B}_1$, any $v \in \mathbb{R}^D$ and some constants $c', c'' > 0$,

$$v^T E_{\theta_0}[-\nabla^2 \ell(\theta, Z)]v \ge \big[c' D^{-4/d} - c''\|u_{f_{\theta_0}} - u_{f_\theta}\|_{L^2}\big]\|H\|^2_{L^2}.$$

Using (2.111) and the hypotheses, we obtain that for some $c_g > 0$,

$$\|u_{f_{\theta_0}} - u_{f_\theta}\|_{L^2} \le \|\mathcal{G}(\theta_0) - \mathcal{G}(\theta_{0,D})\|_{L^2} + c_g\|\theta_{0,D} - \theta\|_{\mathbb{R}^D} \le (c_1 + c_g\epsilon_S)D^{-4/d}.$$

Thus for all $c_1, \varepsilon_S > 0$ small enough and taking the infimum over $v \in \mathbb{R}^D$ with $\|v\|_{\mathbb{R}^D} = \|\Psi(v)\|_{L^2} = \|H\|_{L^2} = 1$, we obtain that for any $\theta \in \mathcal{B}_{\epsilon_S}$ and some $c''' > 0$,

$$\lambda_{min}(E_{\theta_0}[-\nabla^2\ell(\theta, Z)]) \ge c''' D^{-4/d},$$

which completes the proof. □

Finally, we prove the upper bound required for Assumption 2.3.3 ii).

Lemma 2.4.8 (Upper bound)**.** *For every $S > 0$, there exists a constant $c_{max} > 0$ such that for $\|\theta_0\|_{h^2} \le S$ and all $D \in \mathbb{N}$, we have*

$$\sup_{\theta\in\mathcal{B}_1}\Big[|E_{\theta_0}[\ell(\theta,Z)]| + \|E_{\theta_0}[\nabla\ell(\theta,Z)]\|_{\mathbb{R}^D} + \|E_{\theta_0}[\nabla^2\ell(\theta,Z)]\|_{op}\Big] \le c_{max}.$$

Proof. For the zeroeth order term, using Lemma 2.4.5, we have that for some $K_0 > 0$ and any $\theta \in \mathcal{B}_1$,

$$|E_{\theta_0}[\ell(\theta)]| = 1/2 + 1/2\|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2}^2 \lesssim 1 + \|\mathcal{G}(\theta)\|_\infty^2 + \|u_{f_0}\|_\infty^2 \le K_0.$$

For the first order term, similarly by Lemma 2.4.5 there exists some $K_1 > 0$ such that for any $\theta \in \mathcal{B}_1$,

$$\begin{aligned}\big\|E_{\theta_0}\big[-\nabla\ell(\theta)\big]\big\|_{\mathbb{R}^D} &\lesssim \big\|\langle \mathcal{G}(\theta_0) - \mathcal{G}(\theta), \nabla\mathcal{G}(\theta)\rangle_{L^2(\mathcal{O})}\big\|_{\mathbb{R}^D}\\ &\lesssim \big\|G(\theta_0) - \mathcal{G}(\theta)\big\|_\infty \big\|\nabla\mathcal{G}(\theta)\big\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} \le K_1.\end{aligned}$$

For the second order term, we recall the decomposition

$$\lambda_{max}(E_{\theta_0}\big[-\nabla^2\ell(\theta)\big]) = \sup_{v:\|v\|_{\mathbb{R}^D}\le 1} v^T E_{\theta_0}\big[-\nabla^2\ell(\theta)\big]v = \sup_{v:\|v\|_{\mathbb{R}^D}\le 1}\big[I + II + III\big],$$

where the terms $I - III$ were defined in (2.117). Suitable uniform upper bounds for the terms II and III have already been shown in (2.124) and (2.125) respectively, whence it suffices to upper bound the term I. We do this by using (2.108) and Lemma 2.5.5: for any $\theta \in \mathcal{B}_1$ and any $H = \Psi(v)$, $v \in \mathbb{R}^D$,

$$\sqrt{I} = \|V_{f_\theta}[u_{f_\theta}(\Phi' \circ F_\theta)H]\|_{L^2} \lesssim \|u_{f_\theta}(\Phi' \circ F_\theta)H\|_{L^2} \lesssim \|u_{f_\theta}\|_\infty \|\Phi' \circ F_\theta\|_\infty \|H\|_{L^2} \lesssim \|v\|_{\mathbb{R}^D}.$$

□

We conclude with the following basic comparison lemma for Sobolev norms on the subspaces $E_D \subseteq L^2(\mathcal{O})$ from (2.109).

Lemma 2.4.9. *There exists $C > 0$ such that for any $D \in \mathbb{N}$ and any $H \in E_D$,*

$$\|H\|_{H^2} \le CD^{2/d}\|H\|_{L^2}, \qquad \|H\|_{L^2} \le CD^{2/d}\|H\|_{(H_0^2)^*}. \tag{2.126}$$

Proof. Fix $D \in \mathbb{N}$. By the isomorphism property of Δ between the spaces H_0^2 and L^2 (see e.g. Theorem II.5.4 in [109]), we first have the norm equivalence

$$\|\Delta H\|_{L^2} \lesssim \|H\|_{H_0^2} \lesssim \|\Delta H\|_{L^2}, \quad H \in E_D.$$

It follows by Weyl's law (2.13) that

$$\|H\|_{H_0^2}^2 \lesssim \sum_{k=1}^{D} |\langle H, e_k\rangle_{L^2}|^2 \lambda_k^2 \lesssim D^{4/d} \|H\|_{L^2}^2.$$

Thus, combining the above display with the following duality argument completes the proof:

$$\|H\|_{L^2} = \sup_{\psi \in E_D : \|\psi\|_{L^2} \le 1} |\langle H, \psi\rangle_{L^2}| \lesssim D^{2/d} \sup_{\psi \in E_D : \|\psi\|_{H_0^2} \le 1} |\langle H, \psi\rangle_{L^2}| \le D^{2/d} \|H\|_{(H_0^2)^*}.$$

□

2.4.2 Wasserstein approximation of the posterior measure

The main purpose of this section is to prove Theorem 2.4.14, which provides a bound on the Wasserstein distance between the posterior measure $\Pi(\cdot|Z^{(N)})$ from (2.24) and the surrogate posterior $\tilde{\Pi}(\cdot|Z^{(N)})$ from (2.27) in the Schrödinger model. The idea behind the proof of this theorem is to show that both $\Pi(\cdot|Z^{(N)})$ and $\tilde{\Pi}(\cdot|Z^{(N)})$ concentrate most of their mass on the region (2.99) where the log-likelihood function ℓ_N is strongly concave (with high $P_{\theta_0}^N$-probability, cf. Proposition 2.4.1). This involves initially a careful study of the mode (maximiser) of the posterior density, given in Theorem 3.14.

2.4.2.1 Convergence rate of MAP estimates

For $(Y_i, X_i)_{i=1}^N$ arising from (2.19) with $\mathcal{G} : \mathbb{R}^D \to \mathbb{R}$ from (2.17), we now study maximisers

$$\hat{\theta}_{MAP} \in \arg\max_{\theta \in \mathbb{R}^D} \left[-\frac{1}{2N} \sum_{i=1}^{N} \big(Y_i - \mathcal{G}(\theta)(X_i)\big)^2 - \frac{\delta_N^2}{2} \|\theta\|_{h^\alpha}^2 \right], \quad \delta_N = N^{-\frac{\alpha}{2\alpha+d}}, \tag{2.127}$$

of the posterior density (2.24). For Λ_α from (2.23) we will write $I(\theta) := \frac{1}{2}\|\theta\|_{h^\alpha}^2 = \frac{1}{2}\theta^T \Lambda_\alpha \theta$ for $\theta \in \mathbb{R}^D$. We denote the empirical measure on $\mathbb{R} \times \mathcal{O}$ induced by the $Z_i = (X_i, Y_i)$'s as

$$P_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{(Y_i, X_i)}, \quad \text{so that} \quad \int h dP_N = \frac{1}{N} \sum_{i=1}^{N} h(Y_i, X_i) \tag{2.128}$$

for any measurable map $h : \mathbb{R} \times \mathcal{O} \to \mathbb{R}$. Recall also that $p_\theta : \mathbb{R} \times \mathcal{O} \to [0, \infty)$ denotes the marginal probability densities of P_θ^N defined in (2.21).

Lemma 2.4.10. *Let $\hat\theta_{MAP}$ be any maximiser in (2.127), and denote by $\theta_{0,D}$ the projection of θ_0 onto $\mathbb{R}^D$. We have ($P_{\theta_0}^N$-a.s.)*

$$\begin{aligned}&\frac{1}{2}\|\mathcal{G}(\hat\theta_{MAP})-\mathcal{G}(\theta_0)\|_{L^2}^2+\delta_N^2 I(\hat\theta_{MAP})\\&\qquad\le\int\log\frac{p_{\hat\theta_{MAP}}}{p_{\theta_{0,D}}}d(P_N-P_{\theta_0})+\delta_N^2 I(\theta_{0,D})+\frac{1}{2}\|\mathcal{G}(\theta_{0,D})-\mathcal{G}(\theta_0)\|_{L^2}^2.\end{aligned}$$

Proof. By the definitions

$$\ell_N(\hat\theta_{MAP})-\ell_N(\theta_{0,D})-N\delta_N^2 I(\hat\theta_{MAP})\ge -N\delta_N^2 I(\theta_{0,D})$$

which is the same as

$$N\int\log\frac{p_{\hat\theta_{MAP}}}{p_{\theta_{0,D}}}d(P_N-P_{\theta_0})+N\delta_N^2 I(\theta_{0,D})\ge N\delta_N^2 I(\hat\theta_{MAP})-N\int\log\frac{p_{\hat\theta_{MAP}}}{p_{\theta_{0,D}}}dP_{\theta_0}. \quad (2.129)$$

The last term can be decomposed as

$$\begin{aligned}-\int\log\frac{p_{\hat\theta_{MAP}}}{p_{\theta_{0,D}}}dP_{\theta_0}&=-\int\log\frac{p_{\hat\theta_{MAP}}}{p_{\theta_0}}dP_{\theta_0}+\int\log\frac{p_{\theta_{0,D}}}{p_{\theta_0}}dP_{\theta_0}\\&=\frac{1}{2}\|\mathcal{G}(\hat\theta_{MAP})-\mathcal{G}(\theta_0)\|_{L^2(\mathcal{O})}^2-\frac{1}{2}\|\mathcal{G}(\theta_{0,D})-\mathcal{G}(\theta_0)\|_{L^2(\mathcal{O})}^2\end{aligned}$$

where we have used a standard computation of likelihood ratios (see also Lemma 23 in [132]). The result follows from the last two displays after dividing by N. □

The following result can be proved by adapting techniques from M-estimation [170] (see also [169], [136]) to the present situation. We will make crucial use of the concentration Lemma 2.3.12.

Proposition 2.4.11. *Let $\alpha>d$. Suppose $\|\theta_0\|_{h^\alpha}\le c_0$ and that D is such that $\|\mathcal{G}(\theta_0)-\mathcal{G}(\theta_{0,D})\|_{L^2}\le c_1\delta_N$ for some $c_0,c_1>0$. Then, for any $c\ge 1$ we can choose $C=C(c,c_0,c_1)$ large enough so that every $\hat\theta_{MAP}$ maximising (2.127) satisfies*

$$P_{\theta_0}^N\left(\frac{1}{2}\|\mathcal{G}(\hat\theta_{MAP})-\mathcal{G}(\theta_0)\|_{L^2}^2+\delta_N^2 I(\hat\theta_{MAP})>C\delta_N^2\right)\lesssim e^{-c^2N\delta_N^2}. \quad (2.130)$$

Proof. We define functionals

$$\tau(\theta,\theta')=\frac{1}{2}\|\mathcal{G}(\theta)-\mathcal{G}(\theta')\|_{L^2}^2+\delta_N^2 I(\theta),\quad \theta\in\mathbb{R}^D,\theta'\in h^\alpha,$$

and empirical processes

$$W_N(\theta)=\int\log\frac{p_\theta}{p_{\theta_{0,D}}}d(P_N-P_{\theta_0}),\ W_{N,0}(\theta)=\int\log\frac{p_\theta}{p_{\theta_0}}d(P_N-P_{\theta_0}),\ \theta\in\mathbb{R}^D,$$

so that

$$W_N(\theta) = W_{N,0}(\theta) - W_{N,0}(\theta_{0,D}), \quad \theta \in \mathbb{R}^D.$$

Using the previous lemma it suffices to bound

$$P_{\theta_0}^N \Big(\tau(\hat\theta_{MAP}, \theta_0) > C\delta_N^2, W_N(\hat\theta_{MAP}) \geq \tau(\hat\theta_{MAP}, \theta_0) - \delta_N^2 I(\theta_{0,D}) - \|\mathcal{G}(\theta_{0,D}) - \mathcal{G}(\theta_0)\|_{L^2}^2/2\Big)$$

Since

$$I(\theta_{0,D}) = \|\theta_{0,D}\|_{h^\alpha}^2/2 \leq \|\theta_0\|_{h^\alpha}^2/2 \leq c_0^2/2 \text{ and } \|\mathcal{G}(\theta_{0,D}) - \mathcal{G}(\theta_0)\|_{L^2}^2 \leq c_1^2\delta_N^2$$

by hypothesis, we can choose C large enough so that the last probability is bounded by

$$\begin{aligned}
&P_{\theta_0}^N \Big(\tau(\hat\theta_{MAP}, \theta_0) > C\delta_N^2, |W_N(\hat\theta_{MAP})| \geq \tau(\hat\theta_{MAP}, \theta_0)/2\Big) \\
&\leq \sum_{s=1}^\infty P_{\theta_0}^N \left(\sup_{\theta \in \mathbb{R}^D : 2^{s-1}C\delta_N^2 \leq \tau(\theta,\theta_0) \leq 2^s C\delta_N^2} |W_{N,0}(\theta)| \geq 2^s C\delta_N^2/8 \right) + P_{\theta_0}^N(|W_{N,0}(\theta_{0,D})| \geq C\delta_N^2/8) \\
&\leq 2\sum_{s=1}^\infty P_{\theta_0}^N \left(\sup_{\theta\in\Theta_s} |W_{N,0}(\theta)| \geq 2^s C\delta_N^2/8 \right), \qquad (2.131)
\end{aligned}$$

where, for $s \in \mathbb{N}$,

$$\Theta_s := \{\theta \in \mathbb{R}^D : \tau(\theta,\theta_0) \leq 2^s C\delta_N^2\} = \{\theta \in \mathbb{R}^D : \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2}^2 + \delta_N^2\|\theta\|_{h^\alpha}^2 \leq 2^{s+1}C\delta_N^2\}, \qquad (2.132)$$

and where we have used that $\theta_{0,D} \in \Theta_1$ for C large enough by the hypotheses. To proceed, notice that

$$NW_{N,0}(\theta) = \ell_N(\theta) - \ell_N(\theta_0) - E_{\theta_0}[\ell_N(\theta) - \ell_N(\theta_0)]$$

and that, for $(Y_i, X_i) \sim^{i.i.d.} P_{\theta_0}$,

$$\begin{aligned}
\ell_N(\theta) - \ell_N(\theta_0) &= -\frac{1}{2}\sum_{i=1}^N \left[(\mathcal{G}(\theta_0)(X_i) - \mathcal{G}(\theta)(X_i) + \varepsilon_i)^2 - \varepsilon_i^2\right] \\
&= -\sum_{i=1}^N (\mathcal{G}(\theta_0)(X_i) - \mathcal{G}(\theta)(X_i))\varepsilon_i - \frac{1}{2}\sum_{i=1}^N (\mathcal{G}(\theta_0)(X_i) - \mathcal{G}(\theta)(X_i))^2, \quad (2.133)
\end{aligned}$$

so that we have to deal with two empirical processes separately. We first bound

$$\sum_{s=1}^\infty P_{\theta_0}^N \left(\sup_{\theta\in\Theta_s} |Z_N(\theta)| \geq \sqrt{N} 2^s C\delta_N^2/16 \right) \qquad (2.134)$$

where

$$Z_N = \frac{1}{\sqrt{N}}\sum_{i=1}^N h_\theta(X_i)\varepsilon_i, \quad h_\theta = \mathcal{G}(\theta_0) - \mathcal{G}(\theta), \quad \theta \in \Theta = \Theta_s, s \in \mathbb{N},$$

is as in Lemma 2.3.12. We will apply that lemma with bounds (recalling $vol(\mathcal{O}) = 1$)

$$E^X h_\theta^2(X) = \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2}^2 \leq 2^{s+1} C\delta_N^2 =: \sigma_s^2, \quad \|h_\theta\|_\infty \leq 2 \sup_\theta \|\mathcal{G}(\theta)\|_\infty \leq U < \infty \tag{2.135}$$

uniformly in all $\theta \in \Theta_s$, for some fixed constant $U = U(g, \mathcal{O})$ (cf. (2.18)). For the entropy bounds, we use that on each slice $\sup_{\theta \in \Theta_s} \|F_\theta\|_{H^\alpha} \leq C' 2^{s/2}$, which for $\alpha > d$ implies (using (4.184) in [72] and standard extension properties of Sobolev norms)

$$\log N\big(\{F_\theta : \theta \in \Theta_s\}, \|\cdot\|_\infty, \rho\big) \leq K\Big(\frac{2^{s/2}}{\rho}\Big)^{d/\alpha}, \quad \rho > 0,$$

for some constant $K = K(\alpha, d, C')$. Since the map $F_\theta \mapsto \mathcal{G}(\theta)$ is Lipschitz for the $\|\cdot\|_\infty$-norm (Lemma 2.4.6) we deduce that also

$$\log N\big(\{h_\theta = \mathcal{G}(\theta) - \mathcal{G}(\theta_0) : \theta \in \Theta_s\}, \|\cdot\|_\infty, \rho\big) \leq K'\Big(\frac{2^{s/2}}{\rho}\Big)^{d/\alpha}, \quad \rho > 0, \tag{2.136}$$

and as a consequence, for $\alpha > d$ and $J_2(\mathcal{H}), J_\infty(\mathcal{H})$ defined in Lemma 2.3.12,

$$J_2(\mathcal{H}) \lesssim \int_0^{4\sigma_s} \Big(\frac{2^{s/2}}{\rho}\Big)^{d/2\alpha} d\rho \lesssim 2^{sd/4\alpha} \sigma_s^{1-\frac{d}{2\alpha}}, \quad J_\infty(\mathcal{H}) \lesssim \int_0^{4U} \Big(\frac{2^{s/2}}{\rho}\Big)^{d/\alpha} d\rho \lesssim 2^{sd/2\alpha} U^{1-\frac{d}{\alpha}}. \tag{2.137}$$

The sum in (2.134) can now be bounded by Lemma 2.3.12 with $x = c^2 N 2^s \delta_N^2$ and the choices of σ_s, U in (2.135) for $C = C(c) > 0$ large enough,

$$\sum_{s \in \mathbb{N}} P_{\theta_0}^N \left(\sup_{\theta \in \Theta_s} |Z_N(\theta)| \geq \sqrt{N} \sigma_s^2 / 16 \right) \leq 2 \sum_{s \in \mathbb{N}} e^{-c^2 2^s N \delta_N^2} \lesssim e^{-c^2 N \delta_N^2} \tag{2.138}$$

since then, by definition of δ_N, for $\alpha > d$ and C large enough, the quantities

$$\mathcal{J}_2(\mathcal{H}) \lesssim 2^{sd/4\alpha} (2^{s/2} \sqrt{C} \delta_N)^{1-\frac{d}{2\alpha}} \lesssim \frac{1}{C^{(4\alpha+d)/4\alpha}} \sqrt{N} \sigma_s^2, \quad \sigma_s \sqrt{x} \leq \frac{c}{\sqrt{C}} \sqrt{N} \sigma_s^2, \tag{2.139}$$

and

$$\frac{1}{\sqrt{N}} \mathcal{J}_\infty(\mathcal{H}) \lesssim \frac{2^{sd/2\alpha}}{\sqrt{N}} \lesssim \frac{1}{C} \sqrt{N} \sigma_s^2, \quad \frac{x}{\sqrt{N}} = \frac{c^2}{C} \sqrt{N} \sigma_s^2 \tag{2.140}$$

are all of the correct order of magnitude compared to $\sqrt{N}\sigma_s^2$.

We now turn to the process corresponding to the second term in (2.133), which is bounded by

$$\sum_{s \in \mathbb{N}} P_{\theta_0}^N \left(\sup_{\theta \in \Theta_s} |Z_N'(\theta)| \geq \sqrt{N} 2^s C \delta_N^2 / 16 \right) \tag{2.141}$$

where Z'_N is now the centred empirical process

$$Z'_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (h_\theta - E^X h_\theta(X)), \quad \text{with } \mathcal{H} = \{h_\theta = (\mathcal{G}(\theta) - \mathcal{G}(\theta_0))^2 : \theta \in \Theta_s\}$$

to which we will again apply Lemma 2.3.12. Just as in (2.135) the envelopes of this process are uniformly bounded by a fixed constant, again denoted by U, which implies in particular that the bounds (2.137) also apply to $\mathcal{H}$ as then, for some constant $c_U > 0$,

$$\|h_\theta - h_{\theta'}\|_\infty \leq c_U \|\mathcal{G}(\theta) - \mathcal{G}(\theta')\|_\infty.$$

Moreover on each slice Θ_s the weak variances are bounded by

$$E^X h_\theta^2(X) \leq c'_U \|h_\theta\|_{L^2}^2 \leq \sigma_s^2$$

with σ_s as in (2.135) and some $c'_U > 0$. We see that all bounds required to obtain (2.134) apply to the process Z'_N as well, and hence the series in (2.131) is indeed bounded as required in the proposition, completing the proof. $\square$

From a stability estimate for $\theta \mapsto \mathcal{G}(\theta)$ we now obtain the following convergence rate for $\|\hat{\theta}_{MAP} - \theta_0\|_{\ell_2}$ which in turn also bounds $\|\hat{\theta}_{MAP} - \theta_{0,D}\|_{\mathbb{R}^D}$.

Theorem 2.4.12. *Let $Z^{(N)} \sim P_{\theta_0}^N$ be as in (2.20) where $\theta_0 \in h^\alpha, \alpha > d, d \leq 3$. Define*

$$\bar{\delta}_N := N^{-r(\alpha)} \quad \textit{where } r(\alpha) = \frac{\alpha}{2\alpha + d} \frac{\alpha}{\alpha + 2}.$$

Suppose $\|\theta_0\|_{h^\alpha} \leq c_0$ and that D is such that $\|\mathcal{G}(\theta_0) - \mathcal{G}(\theta_{0,D})\|_{L^2} \leq c_1 \delta_N$, for some constants $c_0, c_1 > 0$. Then given $c > 0$ we can choose $\bar{C}, \bar{c}$ large enough (depending on $c, c_0, c_1, \alpha, \mathcal{O}$) so that for all N and any maximiser $\hat{\theta}_{MAP}$ satisfying (2.127), one has

$$P_{\theta_0}^N \left(\|\hat{\theta}_{MAP} - \theta_0\|_{\ell_2} \leq \bar{C} \bar{\delta}_N, \quad \|\hat{\theta}_{MAP}\|_{h^\alpha} \leq \bar{C} \right) \geq 1 - \bar{c} e^{-c^2 N \delta_N^2}. \tag{2.142}$$

Proof. By Proposition 2.4.11 we can restrict to events

$$T_N := \{\|\mathcal{G}(\hat{\theta}_{MAP}) - \mathcal{G}(\theta_0)\|_{L^2}^2 \leq 2C\delta_N^2, \|F_{\hat{\theta}_{MAP}}\|_{H^\alpha} = \|\hat{\theta}_{MAP}\|_{h^\alpha} \leq \sqrt{2C}\} \tag{2.143}$$

of sufficiently high $P_{\theta_0}^N$-probability. If we write $\hat{f} = \Phi \circ F_{\hat{\theta}_{MAP}}$ for Φ from (2.17) then by (2.178), on the events T_N we also have $\|\hat{f}\|_{H^\alpha} \leq C'$ and $\|\hat{f}\|_\infty \leq C'$, for some $C' > 0$. We write $u_{\hat{f}} = \mathcal{G}(\hat{\theta}_{MAP})$ for the unique solution of the Schrödinger equation (2.11) corresponding to $\hat{f}$. We then necessarily have $f = \Delta u_f / (2u_f)$ both for $f = \hat{f}$ and $f = f_0$, where we also use that denominator u_f is bounded away from zero by a constant depending only on $C' \geq \|f\|_\infty, \mathcal{O}, g$, see (2.119). Then using the multiplication and interpolation inequalities

(2.7), (2.8), the regularity estimate from (2.176) and (2.178), we have for $t = \alpha/(\alpha+2)$,

$$\begin{aligned}\|\hat{f} - f_0\|_{L^2} &\lesssim \|u_{\hat{f}} - u_{f_0}\|_{H^2} \\ &\lesssim \|\mathcal{G}(\hat{\theta}_{MAP}) - \mathcal{G}(\theta_0)\|_{L^2}^t \|u_{\hat{f}} - u_{f_0}\|_{H^{\alpha+2}}^{1-t} \\ &\lesssim \delta_N^t (\|\hat{f}\|_{H^\alpha} + \|f_0\|_{H^\alpha}) \lesssim \delta_N^t \end{aligned} \tag{2.144}$$

on the event T_N. From a Sobolev imbedding (some $\kappa > 0$) and applying (2.8) again we further deduce $\|\hat{f} - f_0\|_\infty \lesssim \delta_N^{(\alpha-d/2-\kappa)/(\alpha+2)} \to 0$ as $N \to \infty$, hence using $\inf_x f_0(x) > K_{min}$ we also have $\inf_x \hat{f}(x) \geq K_{min} + k$ for some $k > 0$ (on T_N, for all N large enough). We deduce

$$\|\hat{\theta}_{MAP} - \theta_0\|_{\ell_2} \leq \|F_{\hat{\theta}_{MAP}} - F_{\theta_0}\|_{L^2} = \|\Phi^{-1} \circ \hat{f} - \Phi^{-1} \circ f_0\|_{L^2} \lesssim \|\hat{f} - f_0\|_{L^2} \lesssim \delta_N^t$$

on the events T_N, where in the last inequality we have used regularity of the inverse link function $\Phi^{-1} : [K_{min} + k, \infty)$ and (2.179). This completes the proof. □

2.4.2.2 Posterior contraction rates

We now study the full posterior distribution (2.24) arising from the Gaussian prior Π for θ from (2.23). The result we shall prove parallels Theorem 3.14 but holds for most of the 'mass' of the posterior measure instead of just for its 'mode' $\hat{\theta}_{MAP}$. This requires very different techniques and we rely on ideas from Bayesian nonparametrics [173, 69], specifically recent progress [126] that allows one to deal with non-linear settings (see also [132]).

In the proof of Theorem 2.4.14 to follow we will require control of the posterior 'normalising factors', expressed via sets

$$\mathcal{C}_N = \mathcal{C}_{N,K} = \left\{ \int_{\mathbb{R}^D} e^{\ell_N(\theta) - \ell_N(\theta_0)} d\Pi(\theta) \geq \Pi(B(\delta_N)) \exp\{-(1+K)N\delta_N^2\} \right\}, \tag{2.145}$$

for some $K > 0$, where $\delta_N = N^{-\alpha/(2\alpha+d)}$ and

$$B(\delta_N) = \{\theta \in \mathbb{R}^D : \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2(\mathcal{O})} \leq \delta_N\}.$$

This is achieved in the course of the proof of our next result. We denote by c_g the global Lipschitz constant of the map $\theta \mapsto \mathcal{G}(\theta)$ from $\ell_2(\mathbb{N}) \to L^2(\mathcal{O})$, see (2.111).

Theorem 2.4.13. *Let $Z^{(N)}, \theta_0, \alpha, d, \bar{\delta}_N$ be as in Theorem 3.14 and let $\Pi(\cdot|Z^{(N)})$ denote the posterior distribution from (2.24). Suppose $\|\theta_0\|_{h^\alpha} \leq c_0$ and that $D \leq c_2 N\delta_N^2$ is such that*

$$\|\mathcal{G}(\theta_0) - \mathcal{G}(\theta_{0,D})\|_{L^2(\mathcal{O})} \leq c_1 \delta_N \tag{2.146}$$

for some finite constants $c_0, c_2 > 0, 0 < c_1 < 1/2$. Then for any $a > 0$ there exist c', c'' such that for $K, L = L(a, c_0, c_2, c_g, \alpha, \mathcal{O})$ large enough,

$$P_{\theta_0}^N(\{\Pi(\theta : \|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \leq L\bar{\delta}_N,\ \|\theta\|_{h^\alpha} \leq L|Z^{(N)}) \geq 1 - e^{-aN\delta_N^2}\}, \mathcal{C}_{N,K}) \geq 1 - c'e^{-c''N\delta_N^2}. \tag{2.147}$$

Proof. We initially establish some auxiliary results that will allow us to apply a standard contraction theorem from Bayesian non-parametrics, specifically in a form given in Theorem 13 in [132]. By Lemma 23 in [132] and (2.18) we can lower bound $\Pi_N(\mathcal{B}_N)$ in (35) in [132] by our $\Pi_N(B(\delta_N))$ (after adjusting the choice of δ_N in [132] by a multiplicative constant). Then using (2.146), Corollary 2.6.18 in [72], and ultimately Theorem 1.2 in [107] combined with (4.184) in [72], we have for $\theta' \sim N(0, \Lambda_\alpha^{-1})$,

$$\begin{aligned}\Pi_N(\|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2(\mathcal{O})} < \delta_N) &\geq \Pi_N(\|\mathcal{G}(\theta) - \mathcal{G}(\theta_{0,D})\|_{L^2(\mathcal{O})} < \delta_N/2)\\ &\geq \Pi_N(\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} < \delta_N/2c_g)\\ &\geq e^{-N\delta_N^2\|\theta_{0,D}\|_{h^\alpha}^2/2}\,\mathrm{Pr}(\|\theta'\|_{\mathbb{R}^D} < \sqrt{N}\delta_N^2/2c_g) \geq e^{-\bar{d}N\delta_N^2}\end{aligned} \tag{2.148}$$

for some $\bar{d} > 0$. From this we deduce further from Borell's Gaussian iso-perimetric inequality [25] (in the form of Theorem 2.6.12 in [72]), arguing just as in Lemma 17 in [132] (and invoking the remark after that lemma with $\kappa = 0$ there), that given $B > 0$ we can find M large enough (depending on $\bar{d}, B$) such that

$$\Pi_N(\theta = \theta_1 + \theta_2 \in \mathbb{R}^D : \|\theta_1\|_{\mathbb{R}^D} \leq M\delta_N, \|\theta_2\|_{h^\alpha} \leq M) \geq 1 - 2e^{-BN\delta_N^2}.$$

Next the eigenvalue growth $\lambda_k^\alpha \lesssim k^{2\alpha/d}$ from (2.13) and the hypothesis on D imply that for $\bar{L}$ large enough we have

$$\|\theta_1\|_{h^\alpha} \lesssim D^{\alpha/d}\|\theta_1\|_{\mathbb{R}^D} \leq (c_2N\delta_N^2)^{\alpha/d}M\delta_N \leq \bar{L}/2 \tag{2.149}$$

and then also

$$\Pi_N(\mathcal{A}_N^c) \leq 2e^{-BN\delta_N^2} \text{ where } \mathcal{A}_N = \{\theta \in \mathbb{R}^D : \|\theta\|_{h^\alpha} \leq \bar{L}\}. \tag{2.150}$$

The $\|\cdot\|_\infty$-covering numbers of the implied set of regression functions $\mathcal{G}(\theta)$ satisfy the bounds

$$\begin{aligned}\log N(\{\mathcal{G}(\theta) : \theta \in \mathcal{A}_N\}, \|\cdot\|_\infty, \delta_N) &\lesssim \log N(\{F_\theta : \theta \in \mathcal{A}_N\}, \|\cdot\|_\infty, \tilde{c}\delta_N)\\ &\lesssim \log N(\{F : \|F\|_{H^\alpha(\mathcal{O})} \leq c\bar{L}\}, \|\cdot\|_\infty, \tilde{c}\delta_N) \lesssim N\delta_N^2,\end{aligned}$$

for some $\tilde{c}, c > 0$, using that the map $F_\theta \mapsto \mathcal{G}(\theta)$ is globally Lipschitz for the $\|\cdot\|_\infty$-norm (Lemma 2.4.6) and also the bound (4.184) in [72]. By (2.18) and Lemma 22 in [132] the previous metric entropy inequality also holds for the Hellinger distance replacing $\|\cdot\|_\infty$-distance on the l.h.s. in the last display. Theorem 13 and again Lemma 22 in [132] now imply that for any $a > 0$ and L large enough,

$$P_{\theta_0}^N\Big(\Pi(\{\theta : \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2} > L\delta_N\} \cup \mathcal{A}_N^c | Z^{(N)}) \le e^{-aN\delta_N^2}\Big) \to 0 \qquad (2.151)$$

as $N \to \infty$. The convergence in probability to zero obtained in the proof of Theorem 13 in [132] is in fact exponentially fast, as required in (2.147): This is true by virtue of the bound to follow in the next display (which forms part of the proof in [132] as well), and since the type-one testing errors in (39) in [132] are controlled at the required exponential rate (via Theorem 7.1.4 in [72]). The inequality

$$P_{\theta_0}^N\Big(\int_{B(\delta_N)} e^{\ell_N(\theta)-\ell_N(\theta_0)} d\Pi(\theta) \ge \Pi(B(\delta_N)) \exp\{-(1+K)N\delta_N^2\}\Big) \le c' e^{-c''N\delta_N^2},$$

bounding $P_{\theta_0}^N(\mathcal{C}_{N,K}^c)$ as required in the theorem follows from Lemma 2.4.15 below for large enough K and $\bar{C} = 1/2$.

Now to conclude, we can define subsets of $\mathbb{R}^D$ as

$$\Theta_N := \{\theta : \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2} \le L\delta_N\} \cap \mathcal{A}_N = \{\theta : \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2} \le L\delta_N, \|F_\theta\|_{H^\alpha} = \|\theta\|_{h^\alpha} \le \bar{L}\}$$

paralleling the events T_N from (2.143) above. Then arguing as in and after (2.144), one shows that

$$\Theta_N \subset \tilde{\Theta}_N = \{\theta : \|\theta - \theta_0\|_{\mathbb{R}^D} \le LN^{-r(\alpha)}, \|\theta\|_{h^\alpha} \le L\}, \qquad (2.152)$$

increasing also the constant L if necessary, and hence the posterior probability of this event is also lower bounded by $\Pi(\tilde{\Theta}_N | Z^{(N)}) \ge 1 - e^{-aN\delta_N^2}$, with the desired $P_{\theta_0}^N$-probability, proving the theorem. □

Moreover, a quantitative uniform integrability argument from Section 5.4.5 in [126] (see the proof of Theorem 2.4.14, term III, below) then also gives a convergence rate for the posterior mean $E^\Pi[\theta | Z^{(N)}]$ towards θ_0, namely that for L large enough there exist $\bar{c}', \bar{c}'' > 0$ such that

$$P_{\theta_0}^N(\|E^\Pi[\theta|Z^{(N)}] - \theta_0\|_{\ell_2} > L\bar{\delta}_N) \le \bar{c}' e^{-\bar{c}''N\delta_N^2}. \qquad (2.153)$$

2.4.2.3 Globally log-concave approximation of the posterior in Wasserstein distance

Recall the surrogate posterior measure $\tilde{\Pi}(\cdot | Z^{(N)})$ from (2.27) with log-density

$$\log \tilde{\pi}_N(\theta) = const. + \tilde{\ell}_N(\theta) - \frac{N\delta_N^2}{2}\|\theta\|_{h^\alpha}^2, \quad \theta \in \mathbb{R}^D \qquad (2.154)$$

with θ_{init} and parameters ϵ, K chosen as in Condition 2.2.2, and with $\delta_N = N^{-\alpha/(2\alpha+d)}$. We now prove the main result of this section.

Theorem 2.4.14. *Assume Condition 2.2.3 and let $\tilde{\Pi}(\cdot|Z^{(N)})$ be the probability measure of density given in (2.27) with $K, \varepsilon > 0$ chosen as in Condition 2.2.2. Then for some $a_1, a_2 > 0$ and all $N \in \mathbb{N}$,*

$$P_{\theta_0}^N\big(W_2^2(\tilde{\Pi}(\cdot|Z^{(N)}), \Pi(\cdot|Z^{(N)})) > e^{-N\delta_N^2/2}\big) \le a_1 e^{-a_2 N\delta_N^2}.$$

Proof. In the proof we will require a new sequence

$$\tilde{\delta}_N = N^{(-\alpha+2)/(2\alpha+d)}\sqrt{\log N} \tag{2.155}$$

describing the 'rate of contraction' of the surrogate posterior obtained below. We first notice that the definitions of $\bar{\delta}_N$ (from Theorem 3.14) and of δ_N imply by straightforward calculations and using $D \lesssim N\delta_N^2, \alpha > 6$, the asymptotic relations as $N \to \infty$,

$$\delta_N D^{2/d}\sqrt{\log N} = O(\tilde{\delta}_N), \quad \delta_N \ll \bar{\delta}_N \ll \tilde{\delta}_N \ll \frac{1}{\log N} D^{-\frac{4}{d}}, \tag{2.156}$$

which we shall use in the proof. We will prove the bound for all N large enough, which is sufficient to prove the desired inequality after adjusting the constant in $\lesssim$ (since probabilities are always bounded by one).

Geometry of the surrogate posterior. To set things up, consider MAP estimates $\hat{\theta}_{MAP}$ from (2.127). In view of (2.18), the function q_N to be maximised over $\mathbb{R}^D$ in (2.127) satisfies $q_N(\theta) < q_N(0)$ for all θ such that $\|\theta\|_{h^\alpha}$ exceeds some positive constant k. Then on the compact set $M = \{\theta \in \mathbb{R}^D : \|\theta\|_{h^\alpha} \le k\}$ the function q_N is continuous (as $\mathcal{G}$ is continuous from $\mathbb{R}^D \to L^\infty(\mathcal{O})$, Lemma 2.4.6), and hence attains its maximum at some $\hat{\theta}_M \in M$, which must be a global maximiser of q_N since $q_N(\hat{\theta}_M) \ge q_N(0) > \inf_{\theta \in M^c} q_N(\theta)$. Conclude that a maximiser $\hat{\theta}_{MAP}$ exists (one shows that it can be taken to be measurable, Exercise 7.2.3 in [72]).

In view of Proposition 2.4.1, Theorem 3.14, Theorem 2.5.10 (and the remark before it), $\alpha > 6$ as well as Proposition 2.3.6, we may restrict the rest of the proof to the following event

$$\mathcal{S}_N := \Big\{\|\theta_{init} - \theta_{0,D}\|_{\mathbb{R}^D} \le \frac{1}{8D^{4/d}\log N}, \inf_{\theta \in \mathcal{B}_{1/\log N}} \lambda_{min}(-\nabla^2 \tilde{\ell}_N(\theta)) \ge \underline{c} N D^{-4/d}\Big\}$$
$$\cap \Big\{\text{any } \hat{\theta}_{MAP} \text{ satisfies } \|\hat{\theta}_{MAP} - \theta_{0,D}\|_{\mathbb{R}^D} \le \min\big\{\frac{1}{8D^{4/d}\log N}, \bar{C}\bar{\delta}_N\big\}\Big\},$$

where $\mathcal{B}_\epsilon$ was defined in (2.99), where $\bar{C}$ is from (2.142) and where $\underline{c} = c_2$ from Proposition 2.4.1. On $\mathcal{S}_N$ we have the following properties of $\tilde{\ell}_N$. First, from (2.26),

$$\tilde{\ell}_N(\theta) = \ell_N(\theta) \text{ for any } \theta \ s.t. \ \|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \le \frac{3}{8D^{4/d}\log N}. \tag{2.157}$$

Moreover, by Proposition 2.3.6, $\log \tilde{\pi}(\cdot|Z^{(N)})$ is strongly concave in view of

$$\sup_{\theta,\vartheta\in\mathbb{R}^D, \|\vartheta\|_{\mathbb{R}^D}=1} \vartheta^T[\nabla^2(\log \tilde{\pi}_N(\theta))]\vartheta \le \sup_{\theta,\vartheta\in\mathbb{R}^D, \|\vartheta\|_{\mathbb{R}^D}=1} \vartheta^T[\nabla^2\tilde{\ell}_N(\theta)]\vartheta \le -\underline{c}ND^{-4/d}. \tag{2.158}$$

Finally, any $\hat{\theta}_{MAP}$ necessarily satisfies

$$0 = \nabla \log \pi(\hat{\theta}_{MAP}|Z^{(N)}) = \nabla \log \tilde{\pi}(\hat{\theta}_{MAP}), \tag{2.159}$$

from which we conclude that $\hat{\theta}_{MAP}$ necessarily equals the *unique* global maximiser of the strongly concave function $\log \tilde{\pi}(\cdot|Z^{(N)})$ over $\mathbb{R}^D$.

Decomposition of the Wasserstein distance. Now let us write

$$\hat{\mathcal{B}}(r) = \{\theta \in \mathbb{R}^D : \|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D} \le r\},$$

for the Euclidean ball of radius $r > 0$ centred at $\hat{\theta}_{MAP}$. Then using Theorem 6.15 in [176] with $x_0 = \hat{\theta}_{MAP}$, we obtain for any $m > 0$ that

$$\begin{aligned}
W_2^2(\tilde{\Pi}(\cdot|Z^{(N)}), \Pi(\cdot|Z^{(N)})) &\le 2\int_{\mathbb{R}^D} \|\theta - \hat{\theta}_{MAP}\|^2_{\mathbb{R}^D} d|\tilde{\Pi}(\cdot|Z^{(N)}) - \Pi(\cdot|Z^{(N)})|(\theta) \\
&\le 2\int_{\hat{\mathcal{B}}(m\tilde{\delta}_N)} \|\theta - \hat{\theta}_{MAP}\|^2_{\mathbb{R}^D} d|\tilde{\Pi}(\cdot|Z^{(N)}) - \Pi(\cdot|Z^{(N)})|(\theta) \\
&\quad + 2\int_{\mathbb{R}^D\setminus\hat{\mathcal{B}}(m\tilde{\delta}_N)} \|\theta - \hat{\theta}_{MAP}\|^2_{\mathbb{R}^D} d|\tilde{\Pi}(\cdot|Z^{(N)}) - \Pi(\cdot|Z^{(N)})|(\theta) \\
&\le 2m^2\tilde{\delta}_N^2 \int_{\hat{\mathcal{B}}(m\tilde{\delta}_N)} d|\Pi(\cdot|Z^{(N)}) - \tilde{\Pi}(\cdot|Z^{(N)})|d\theta \\
&\quad + 2\int_{\|\theta-\hat{\theta}_{MAP}\|_{\mathbb{R}^D}>m\tilde{\delta}_N} \|\theta - \hat{\theta}_{MAP}\|^2_{\mathbb{R}^D} d\tilde{\Pi}(\cdot|Z^{(N)}) \\
&\quad + 2\int_{\|\theta-\hat{\theta}_{MAP}\|_{\mathbb{R}^D}>m\tilde{\delta}_N} \|\theta - \hat{\theta}_{MAP}\|^2_{\mathbb{R}^D} d\Pi(\cdot|Z^{(N)}) \\
&\equiv I + II + III,
\end{aligned}$$

and we now bound I, II, III in separate steps.

Term II. We can write the surrogate posterior density as

$$\tilde{\pi}(\theta|Z^{(N)}) = \frac{e^{\tilde{\ell}_N(\theta)-\tilde{\ell}_N(\hat{\theta}_{MAP})}\pi(\theta)}{\int_{\mathbb{R}^D} e^{\tilde{\ell}_N(\theta)-\tilde{\ell}_N(\hat{\theta}_{MAP})}\pi(\theta)d\theta}, \quad \theta \in \mathbb{R}^D,$$

and will first lower bound the normalising factor. From (2.156) we have for any $c > 0$ the set inclusion

$$B_N \equiv \{\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \leq c\delta_N\} \subset \Big\{\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \leq \frac{3}{8D^{4/d}\log N}\Big\}$$

whenever N is large enough. Since $\ell_N(\theta) = \tilde{\ell}_N(\theta)$ on the last set we have on an event of large enough $P_{\theta_0}^N$-probability,

$$\begin{aligned}\int_{\mathbb{R}^D} e^{\tilde{\ell}_N(\theta)-\tilde{\ell}_N(\hat{\theta}_{MAP})} d\Pi(\theta) &\geq \int_{B_N} e^{\tilde{\ell}_N(\theta)-\tilde{\ell}_N(\hat{\theta}_{MAP})} d\Pi(\theta) \\ &= \int_{B_N} e^{\ell_N(\theta)-\ell_N(\hat{\theta}_{MAP})} d\nu(\theta) \times \Pi(B_N) \geq e^{-\bar{c}N\delta_N^2}\end{aligned}$$

for some $\bar{c} = \bar{c}(\bar{d}, c)$, where we have used Lemma 2.4.15 for our choice of B_N (permitted for appropriate choice of $c > 0$ by (2.28) and since $\mathcal{G} : \mathbb{R}^D \to L^2$ is Lipschitz, see Section 2.5) with $\nu = \Pi(\cdot)/\Pi(B_N), \bar{C} = 1/2$; as well as the small ball estimate for Π in (2.148).

Now recall the prior (2.23) and define scaling constants

$$V_N = (2\pi)^{-D/2}\sqrt{\det(N\delta_N^2\Lambda_\alpha)} \times e^{\bar{c}N\delta_N^2}.$$

Then on the preceding events the term II can be bounded, using a second order Taylor expansion of $\log\tilde{\pi}(\cdot|Z^{(N)})$ around its maximum $\hat{\theta}_{MAP}$ combined with (2.158), (2.159), as

$$\begin{aligned}&\int_{\|\theta-\hat{\theta}_{MAP}\|_{\mathbb{R}^D} > m\tilde{\delta}_N} \|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D}^2 \tilde{\pi}(\theta|Z^{(N)}) d\theta \\ &\leq e^{\bar{c}N\delta_N^2} \int_{\|\theta-\hat{\theta}_{MAP}\|_{\mathbb{R}^D} > m\tilde{\delta}_N} \|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D}^2 e^{\tilde{\ell}_N(\theta)-\tilde{\ell}_N(\hat{\theta}_{MAP})}\pi(\theta) d\theta \\ &\leq V_N \times \int_{\|\theta-\hat{\theta}_{MAP}\|_{\mathbb{R}^D} > m\tilde{\delta}_N} \|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D}^2 e^{\tilde{\ell}_N(\theta)-\frac{N\delta_N^2}{2}\|\theta\|_{h^\alpha}^2 - \tilde{\ell}_N(\hat{\theta}_{MAP}) + \frac{N\delta_N^2}{2}\|\hat{\theta}_{MAP}\|_{h^\alpha}^2} d\theta \\ &= V_N \times \int_{\|\theta-\hat{\theta}_{MAP}\|_{\mathbb{R}^D} > m\tilde{\delta}_N} \|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D}^2 e^{\log\tilde{\pi}_N(\theta) - \log\tilde{\pi}_N(\hat{\theta}_{MAP})} d\theta \\ &\leq V_N \times \int_{\|\theta-\hat{\theta}_{MAP}\|_{\mathbb{R}^D} > m\tilde{\delta}_N} \|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D}^2 e^{-\underline{c}ND^{-4/d}\|\theta-\hat{\theta}_{MAP}\|_{\mathbb{R}^D}^2/2} d\theta \\ &\leq 2V_N \times \big(\frac{4\pi}{\underline{c}ND^{-4/d}}\big)^{D/2} \Pr\big(\|Z\|_{\mathbb{R}^D} > m\tilde{\delta}_N\big)\end{aligned}$$

where we have used $x^2 e^{-cx^2} \leq 2e^{-cx^2/2}$ for all $x \in \mathbb{R}$, $c \geq 1$ (and N such that $\underline{c}ND^{-4/d} \geq 1$) and where

$$Z \sim N\Big(0, \frac{2}{\underline{c}D^{-4/d}N} I_{D\times D}\Big).$$

Now by $D \leq c_0 N\delta_N^2$ and (2.156),

$$E\|Z\|_{\mathbb{R}^D} \leq \sqrt{E\|Z\|_{\mathbb{R}^D}^2} \leq \sqrt{2D/(\underline{c}D^{-4/d}N)} \leq (2c_0/\underline{c})^{1/2}\delta_N D^{2/d} \leq (m/2)\tilde{\delta}_N$$

for m large enough, so that

$$\Pr\left(\|Z\|_{\mathbb{R}^D} > m\tilde{\delta}_N\right) \leq \Pr\left(\|Z\|_{\mathbb{R}^D} - E\|Z\|_{\mathbb{R}^D} > (m/2)\tilde{\delta}_N\right) \leq e^{-m^2\underline{c}ND^{-4/d}\tilde{\delta}_N^2/16}$$

by a concentration inequality for Lipschitz-functionals of D-dimensional Gaussian random vectors (e.g., Theorem 2.5.7 in [72] applied to $(\underline{c}ND^{-4/d}/2)^{1/2}Z \sim N(0, I_{D\times D})$ and $F = \|\cdot\|_{\mathbb{R}^D}$). By (2.13) and since $D \lesssim N\delta_N^2$ we have for some $c' > 0$

$$V_N \leq e^{c'N\delta_N^2 \log N}$$

so that for m large enough and using (2.156), the last term in the displayed array above, and hence II, is bounded by

$$2V_N \times (\frac{4\pi}{\underline{c}ND^{-4/d}})^{D/2} \times e^{-m^2\underline{c}D^{-4/d}N\tilde{\delta}_N^2/16} \leq e^{-m^2D^{-4/d}N\tilde{\delta}_N^2/32} \leq \frac{1}{8}e^{-N\delta_N^2}.$$

Term III: We first note that Theorem 2.4.13 and (2.156) imply that for every $a > 0$ we can find m large enough such that

$$\begin{aligned}\Pi(\|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D} > m\tilde{\delta}_N|Z^{(N)}) &\leq \Pi(\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} > m\bar{\delta}_N - \|\hat{\theta}_{MAP} - \theta_{0,D}\|_{\mathbb{R}^D}|Z^{(N)}) \\ &\leq \Pi(\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} > m\bar{\delta}_N/2|Z^{(N)}) \leq e^{-aN\delta_N^2}\end{aligned}$$

on events $\mathcal{S}'_N \subset \mathcal{S}_N$ of sufficiently high probability. Moreover, again by Theorem 2.4.13, we can further restrict the argument that follows to the event $\mathcal{C}_{N,K}$ from (2.145) for some $K > 0$. Now using the Cauchy-Schwarz and Markov inequalities as well as $E_{\theta_0}^N e^{\ell_N(\theta)-\ell_N(\theta_0)} = 1$ and the small ball estimate for Π in (2.148), we have

$$\begin{aligned}&P_{\theta_0}^N\Big(\mathcal{C}_{N,K} \cap \mathcal{S}'_N, \int_{\|\theta-\hat{\theta}_{MAP}\|_{\mathbb{R}^D} > m\tilde{\delta}_N} \|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D}^2 d\Pi(\cdot|Z^{(N)}) > e^{-N\delta_N^2}/8\Big) \\ &\leq P_{\theta_0}^N\Big(\mathcal{C}_{N,K} \cap \mathcal{S}'_N, \Pi(\|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D} > m\tilde{\delta}_N|Z^{(N)})E^{\Pi}[\|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D}^4|Z^{(N)}] > e^{-2N\delta_N^2}/8\Big) \\ &\leq P_{\theta_0}^N\Big(\mathcal{S}'_N, e^{(1+K+\bar{d}+2-a)N\delta_N^2}\int_{\mathbb{R}^D} \|\theta - \hat{\theta}_{MAP}\|_{\mathbb{R}^D}^4 e^{\ell_N(\theta)-\ell_N(\theta_0)} d\Pi(\theta) > 1/8\Big) \\ &\lesssim e^{(1+K+\bar{d}+2-a)N\delta_N^2}\int_{\mathbb{R}^D}(1 + \|\theta\|_{\mathbb{R}^D}^4) d\Pi(\theta) \leq e^{-a_2N\delta_N^2}\end{aligned}$$

whenever m and then a are large enough, since Π has uniformly bounded fourth moments and since $\|\hat{\theta}_{MAP}\|_{\mathbb{R}^D}$ is uniformly bounded by a constant depending only on $\|\theta_0\|_{\ell_2}$ on the events $\mathcal{S}_N$.

Term I: On the events $\mathcal{S}_N$ we have from (2.156) that for fixed $m > 0$ and all N large enough

$$\hat{\mathcal{B}}(m\tilde{\delta}_N) \subseteq \{\theta : \|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \leq 3/(8D^{4/d}\log N)\}.$$

On the latter set, by (2.157), the probability measures $\tilde{\Pi}(\cdot|Z^{(N)})$ and $\Pi(\cdot|Z^{(N)})$ coincide up to a normalising factor, and thus we can represent their Lebesgue densities as

$$\tilde{\pi}(\theta|Z^{(N)}) = p_N \pi(\theta|Z^{(N)}), \quad \theta \in \hat{\mathcal{B}}(m\tilde{\delta}_N),$$

for some $0 < p_N < \infty$. Moreover, by the preceding estimates for terms II and III (which hold just as well without the integrating factors $\|\theta - \hat{\theta}_{MAP}\|^2_{\mathbb{R}^D}$), we have both

$$p_N \Pi(\hat{\mathcal{B}}(m\tilde{\delta}_N)|Z^{(N)}) = \tilde{\Pi}(\hat{\mathcal{B}}(m\tilde{\delta}_N)|Z^{(N)}) \geq 1 - e^{-N\delta_N^2}/8 \ \Rightarrow \ 1 - e^{-N\delta_N^2}/8 \leq p_N,$$

$$p_N^{-1} \tilde{\Pi}(\hat{\mathcal{B}}(m\tilde{\delta}_N)|Z^{(N)}) = \Pi(\hat{\mathcal{B}}(m\tilde{\delta}_N)|Z^{(N)}) \geq 1 - e^{-N\delta_N^2}/8 \ \Rightarrow \ 1 - e^{-N\delta_N^2}/8 \leq \frac{1}{p_N}$$

on events of sufficiently high $P^N_{\theta_0}$-probability. On these events necessarily

$$p_N \in \Big[1 - \frac{e^{-N\delta_N^2}}{8}, \frac{1}{1 - \frac{e^{-N\delta_N^2}}{8}}\Big]$$

and so for N large enough

$$\int_{\hat{\mathcal{B}}(m\tilde{\delta}_N)} d|\Pi(\cdot|Z^{(N)}) - \tilde{\Pi}(\cdot|Z^{(N)})|(\theta) = |1 - p_N| \int_{\hat{\mathcal{B}}(m\tilde{\delta}_N)} \pi(\theta|Z^{(N)}) d\theta \leq |1 - p_N| \leq e^{-N\delta_N^2}/4,$$

which is obvious for $p_N \leq 1$ and follows from the mean value theorem applied to $f(x) = (1-x)^{-1}$ near $x = 0$ also for $p_N > 1$. Collecting the bounds for I, II, III completes the proof. □

2.4.2.4 An 'exponential' small ball lemma

Lemma 2.4.15. *Let $\mathcal{G}$ be as in (2.17) and let ν be a probability measure on some ($\ell_2(\mathbb{N})$-measurable) set*

$$B_N \subseteq \{\theta \in \ell_2(\mathbb{N}) : \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|^2_{L^2} \leq 2\bar{C}\delta_N^2\}, \ \textit{for some } \bar{C} > 0. \tag{2.160}$$

Then for ℓ_N from (2.22) we have for every $K = K(\bar{C}) > 0$ large enough and some fixed constant $b > 0$ that

$$P^N_{\theta_0}\left(\int_{B_N} e^{\ell_N(\theta) - \ell_N(\hat{\theta}_{MAP})} d\nu(\theta) \leq e^{-(1+K)\bar{C}^2 N\delta_N^2}\right) \lesssim e^{-bN\delta_N^2}. \tag{2.161}$$

The same conclusion holds true with $\ell_N(\hat{\theta}_{MAP})$ replaced by $\ell_N(\theta_0)$.

Proof. We proceed as in Lemma 7.3.2 in [72] to deduce from Jensen's inequality (applied to log and $\int(\cdot)d\nu$) that, for P_N the empirical measure from (2.128), the probability in question is bounded by

$$P_{\theta_0}^N\left(\int\int_{B_N}\log\frac{p_\theta}{p_{\hat\theta_{MAP}}}d\nu(\theta)d(P_N-P_{\theta_0})\le -(1+K)\bar C^2\delta_N^2-\int\int_{B_N}\log\frac{p_\theta}{p_{\hat\theta_{MAP}}}d\nu(\theta)dP_{\theta_0}\right).$$

Now just as in the proof of Lemma 2.4.10 and using Theorem 3.14 we see that

$$\begin{aligned}-\int\log\frac{p_\theta}{p_{\hat\theta_{MAP}}}dP_{\theta_0}&=-\int\log\frac{p_\theta}{p_{\theta_0}}dP_{\theta_0}+\int\log\frac{p_{\theta_0}}{p_{\hat\theta_{MAP}}}dP_{\theta_0}\\&=\frac{1}{2}\|\mathcal{G}(\theta)-\mathcal{G}(\theta_0)\|_{L^2}^2-\frac{1}{2}\|\mathcal{G}(\hat\theta_{MAP})-\mathcal{G}(\theta_0)\|_{L_2}^2\le\bar C^2\delta_N^2\end{aligned}$$

so that using also Fubini's theorem the last probability can be bounded by

$$\begin{aligned}&P_{\theta_0}^N\left(\sqrt{N}\int\int_{B_N}\log\frac{p_{\theta_0}}{p_\theta}d\nu(\theta)d(P_N-P_{\theta_0})\ge K\bar C^2\sqrt{N}\delta_N^2/2\right)\\&+P_{\theta_0}^N\left(\sqrt{N}\int\log\frac{p_{\hat\theta_{MAP}}}{p_{\theta_0}}d(P_N-P_{\theta_0})\ge K\bar C^2\sqrt{N}\delta_N^2/2\right).\end{aligned}$$

For the first probability we decompose as in (2.133) and consider Z_N as in Lemma 2.3.12 for fixed h_θ equal to either h_1 or h_2, where

$$h_1(x)=\int_{B_N}(\mathcal{G}(\theta)(x)-\mathcal{G}(\theta_0)(x))d\nu(\theta),\quad\text{and } h_2(x)=\int_{B_N}(\mathcal{G}(\theta)(x)-\mathcal{G}(\theta_0)(x))^2d\nu(\theta).$$

To each of these we apply Bernstein's inequality (2.96) with $x=N\sigma^2$ and K large enough to obtain the desired exponential bound, using uniform boundedness $\|\mathcal{G}(\theta)-\mathcal{G}(\theta_0)\|_\infty\le 2U$ from (2.18) and Jensen's inequality in the variance estimates $E^Xh_1^2(X)\le 2\bar C^2\delta_N^2\equiv\sigma^2$ in the first case and

$$E^Xh_2^2(X)\le 4U^2\int_{B_N}\|\mathcal{G}(\theta)-\mathcal{G}(\theta_0)\|_{L^2}^2d\nu(\theta)\le 8U^2\bar C\delta_N^2\equiv\sigma^2$$

for the second case. [This already proves the case where $\hat\theta_{MAP}$ is replaced by θ_0.]

For the second probability, restricting to the event in the supremum below, which has sufficiently high $P_{\theta_0}^N$-probability in view of Proposition 2.4.11, it suffices to bound

$$P_{\theta_0}^N\left(\sup_{\|\theta\|_{h^\alpha}\le 2C,\|\mathcal{G}(\theta)-\mathcal{G}(\theta_0)\|_{L_2}^2\le 2C\delta_N^2}\sqrt{N}\left|\int\log\frac{p_\theta}{p_{\theta_0}}d(P_N-P_{\theta_0})\right|\ge K\bar C^2\sqrt{N}\delta_N^2/2\right).$$

This term corresponds to the empirical process bounded in and after (2.131) for $s=1$. Choosing K large enough the proof there now applies directly, giving the desired exponential bound. □

2.5 Auxiliary results

2.5.1 Review of convergence guarantees for ULA

In this section we collect some key results (that were used in our proofs) about convergence guarantees for an Unadjusted Langevin Algorithm (ULA) for sampling from *strongly log-concave target measures* [48, 57, 58]. Our presentation follows the recent article [58].

Suppose that μ is a Borel probability measure on $\mathbb{R}^D$ which has a Lebesgue density proportional to e^{-U} for some potential $U : \mathbb{R}^D \to \mathbb{R}$, specifically

$$\mu(B) = \frac{\int_B e^{-U(\theta)} d\theta}{\int_{\mathbb{R}^D} e^{-U(\theta)} d\theta}, \quad B \subseteq \mathbb{R}^D \text{ measurable.} \tag{2.162}$$

Following [58] (cf. H1 and H2 there) we will assume that the potential U has a Λ-Lipschitz gradient and is m-strongly convex.

Assumption 2.5.1. *1. The function $U : \mathbb{R}^D \to \mathbb{R}$ is continuously differentiable and there exists a constant $\Lambda \geq 0$ such that for all $\theta, \bar{\theta} \in \mathbb{R}^D$,*

$$\|\nabla U(\theta) - \nabla U(\bar{\theta})\|_{\mathbb{R}^D} \leq \Lambda \|\theta - \bar{\theta}\|_{\mathbb{R}^D}.$$

2. There exists a constant $0 < m \leq \Lambda$ such that for all $\theta, \bar{\theta} \in \mathbb{R}^D$, we have

$$U(\bar{\theta}) \geq U(\theta) + \langle \nabla U(\theta), \bar{\theta} - \theta \rangle_{\mathbb{R}^D} + \frac{m}{2} \|\theta - \bar{\theta}\|^2_{\mathbb{R}^D}.$$

Under Assumption 2.5.1, the potential U has a unique minimiser over $\mathbb{R}^D$, which we shall denote by θ_U. For the computation of θ_U via gradient descent methods, we have the following standard result from convex optimisation (see Theorem 1 in [48] and (9.18) in [28]).

Proposition 2.5.2. *Suppose $U : \mathbb{R}^D \to \mathbb{R}$ satisfies Assumption 2.5.1. Then the gradient descent algorithm given by*

$$\vartheta_{k+1} = \vartheta_k - \frac{1}{2\Lambda} \nabla U(\vartheta_k), \quad k = 0, 1, 2, \ldots,$$

satisfies that

$$\|\vartheta_k - \theta_U\|^2_{\mathbb{R}^D} \leq \frac{2(U(\vartheta_0) - U(\theta_U))}{m} \big(1 - \frac{m}{2\Lambda}\big)^k, \quad k = 0, 1, 2, \ldots$$

The results presented below establish corresponding geometric convergence bounds for *stochastic* gradient methods which target the entire probability measure μ (instead of just its mode θ_U). Define the continuous time Langevin diffusion process as the unique strong solution $(L_t : t \geq 0)$ of the stochastic differential equation

$$dL_t = -\nabla U(L_t)dt + \sqrt{2}dW_t, \quad t \geq 0,\ L_t \in \mathbb{R}^D, \tag{2.163}$$

where $(W_t : t \geq 0)$ is a D-dimensional standard Brownian motion. It is well known that the Markov process $(L_t : t \geq 0)$ has μ from (2.162) as its invariant measure. The Euler-Maruyama discretisation of the dynamics (2.163) gives rise to the discrete-time Markov chain $(\vartheta_k : k \geq 0)$,

$$\vartheta_{k+1} = \vartheta_k - \gamma\nabla U(\vartheta_k) + \sqrt{2\gamma}\xi_{k+1}, \quad k \geq 0, \tag{2.164}$$

where $(\xi_k : k \geq 1)$ form an i.i.d. sequence of D-dimensional standard Gaussian $N(0, I_{D\times D})$ vectors, and $\gamma > 0$ is some fixed *step size*. We will refer to (ϑ_k) as the unadjusted Langevin algorithm (ULA) in what follows. We denote by $\mathbf{P}_{\theta_{init}}, \mathbf{E}_{\theta_{init}}$ the law and expectation operator, respectively, of the Markov chain $(\vartheta_k : k \geq 1)$ when started at a deterministic point $\vartheta_0 = \theta_{init}$. We also write $\mathcal{L}(\vartheta_k)$ for the (marginal) distribution of the k-th iterate ϑ_k.

For any measurable function $H : \mathbb{R}^D \to \mathbb{R}$ and any $J_{in}, J \geq 0$, let us define the average of H along an ULA trajectory after 'burn-in' period J_{in} by

$$\hat{\mu}^J_{J_{in}}(H) = \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} H(\vartheta_k).$$

Proposition 2.5.3. *Suppose that U satisfies Assumption 2.5.1 and suppose $\gamma \leq 2/(m+\Lambda)$. Then for all $J, J_{in} \geq 1, x > 0$ and any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$, we have the concentration inequality*

$$\mathbf{P}_{\theta_{init}}\Big(\hat{\mu}^J_{J_{in}}(H) - \mathbf{E}_{\theta_{init}}[\hat{\mu}^J_{J_{in}}(H)] \geq x\Big) \leq \exp\Big(-\frac{J\gamma x^2 m^2}{16\|H\|^2_{Lip}(1+2/(mJ\gamma))}\Big).$$

Proof. The statement follows directly from Theorem 17 of [58], noting that $\kappa = 2m\Lambda/(m+\Lambda) \in [m, 2m]$ and that the constant $v_{N,n}(\gamma)$ from (28) of [58] can be upper bounded by

$$1 + \frac{m^{-1} + 2/(m+\Lambda)}{\gamma J} \leq 1 + 2/(m\gamma J).$$

□

Proposition 2.5.4. *Suppose that U satisfies Assumption 2.5.1 and let γ, J_{in}, J and H be as in Proposition 2.5.3. Then we have for μ as in (2.162) that*

$$W_2^2(\mathcal{L}(\vartheta_k), \mu) \leq 2(1 - m\gamma/2)^k\Big[\|\theta_{init} - \theta_U\|^2_{\mathbb{R}^D} + \frac{D}{m}\Big] + b(\gamma)/2, \qquad k \geq 0, \tag{2.165}$$

where

$$b(\gamma) = 36\frac{\gamma D\Lambda^2}{m^2} + 12\frac{\gamma^2 D\Lambda^4}{m^3}, \tag{2.166}$$

as well as

$$\Big(\mathbf{E}_{\theta_{init}}[\hat{\mu}^J_{J_{in}}(H)] - E_\mu H\Big)^2 \leq \|H\|^2_{Lip}\ \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} W_2^2(\mathcal{L}(\vartheta_k), \mu). \tag{2.167}$$

Proof. The display (2.167) is derived in (27) of [58]. The bound (2.165) follows from an application of Theorem 5 in [58] with fixed step size $\gamma > 0$, where in our case, noting again that $\kappa \in [m, 2m]$, the expression $u_n^{(1)}(\gamma)$ there is upper bounded by $2(1-m\gamma/2)^k$ and the expression $u_n^{(2)}(\gamma)$ there is upper bounded by (using that $\gamma \leq \min\{2/\Lambda, 1/m\} \leq \min\{2/\Lambda, 2/\kappa\}$)

$$\begin{aligned}
&\Lambda^2 D\gamma^2(\kappa^{-1}+\gamma)\Big(2+\frac{\Lambda^2\gamma}{m}+\frac{\Lambda^2\gamma^2}{6}\Big)\sum_{i=1}^k(1-\kappa\gamma/2)^{k-i}\\
&\leq \Lambda^2 D\gamma^2(\kappa^{-1}+\gamma)\Big(2+\frac{\Lambda^2\gamma}{m}+\frac{\Lambda^2\gamma^2}{6}\Big)\frac{2}{\kappa\gamma}\\
&\leq \Lambda^2 D\gamma\Big(\kappa^{-2}+\frac{\gamma}{\kappa}\Big)\Big(6+\frac{2\Lambda^2\gamma}{m}\Big)\\
&\leq \Lambda^2 D\gamma m^{-2}\Big(18+\frac{6\Lambda^2\gamma}{m}\Big),
\end{aligned}$$

which equals (2.166). □

2.5.2 Analytical properties of Schrödinger operators and link functions

Recall the inverse Schrödinger operators V_f from (2.107).

Lemma 2.5.5. *There exists a constant $C > 0$ such that for any $f \in C(\mathcal{O})$ with $f \geq 0$, the following holds.*

i) *We have the estimates*

$$\begin{aligned}
&\|V_f[\psi]\|_{L^2} \leq C\|\psi\|_{L^2},\ \ \psi \in L^2(\mathcal{O}),\\
&\|V_f[\psi]\|_\infty \leq C\|\psi\|_\infty,\ \ \psi \in C(\mathcal{O}).
\end{aligned} \tag{2.168}$$

ii) *For any $\psi \in L^2(\mathcal{O})$, we have that*

$$\|V_f[\psi]\|_{H^2} \leq C(1+\|f\|_\infty)\|\psi\|_{L^2}, \tag{2.169}$$

as well as

$$\frac{1}{C(1+\|f\|_\infty)}\|\psi\|_{(H_0^2)^*} \leq \|V_f[\psi]\|_{L^2} \leq C(1+\|f\|_\infty)\|\psi\|_{(H_0^2)^*}. \tag{2.170}$$

iii) *If also $d \leq 3$, then for any $\psi \in L^2(\mathcal{O})$ and any $f, \bar{f} \in C(\mathcal{O})$ with $f, \bar{f} \geq 0$, we have that*

$$\|V_f[\psi] - V_{\bar{f}}[\psi]\|_\infty \lesssim (1 + \|f\|_\infty)\|\psi\|_{L^2}\|f - \bar{f}\|_\infty. \tag{2.171}$$

Proof. Part i) is a direct consequence of the Feynman-Kac formula for $V_f[\psi]$ from [42] (see also Lemma 25 in [136]). The upper bounds in part ii) likewise are proved by standard arguments for elliptic PDEs (see, e.g., Lemma 26 in [136]). In order to prove the lower bound in (2.170), let us denote the Schrödinger operator by $S_f[w] = \frac{1}{2}\Delta w - fw$. Since $S_f : H_0^2 \to L^2$ satisfies $S_f V_f[\psi] = \psi$, it suffices to show that

$$\|S_f w\|_{(H_0^2)^*} \lesssim (1 + \|f\|_\infty)\|w\|_{L^2}, \ w \in H_0^2.$$

Using the divergence theorem we have that for such w,

$$\begin{aligned}\|S_f w\|_{(H_0^2)^*} &= \sup_{\psi \in H_0^2 : \|\psi\|_{H_0^2} \leq 1} \Big| \int_{\mathcal{O}} \psi S_f w \Big| \\ &= \sup_{\psi \in H_0^2 : \|\psi\|_{H_0^2} \leq 1} \Big| \int_{\mathcal{O}} w S_f \psi \Big| \leq \|w\|_{L^2} \sup_{\psi \in H_0^2 : \|\psi\|_{H_0^2} \leq 1} \|S_f \psi\|_{L^2},\end{aligned}$$

and the term on the right hand side is further estimated by

$$\|S_f \psi\|_{L^2} \lesssim \|\Delta\psi\|_{L^2} + \|f\psi\|_{L^2} \lesssim 1 + \|f\|_\infty \|\psi\|_{L^2} \leq 1 + \|f\|_\infty,$$

which proves (2.170). Finally, (2.171) is proved by using a Sobolev embedding as well as (2.168), (2.169):

$$\begin{aligned}\|V_f[\psi] - V_{\bar{f}}[\psi]\|_\infty &\lesssim \|V_f[(f - \bar{f})V_{\bar{f}}[\psi]]\|_{H^2} \lesssim (1 + \|f\|_\infty)\|(f - \bar{f})V_f[\psi]\|_{L^2} \\ &\lesssim (1 + \|f\|_\infty)\|f - \bar{f}\|_\infty \|\psi\|_{L^2}.\end{aligned}$$

□

For any normed vector spaces $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ let $L(V, W)$, denote the space of bounded linear operators $V \to W$, equipped with the operator norm. For $g \in C^\infty(\partial\mathcal{O})$ and any $f \in C(\mathcal{O})$ with $f > 0$, there exists a unique (weak) solution $G(f) \in C(\mathcal{O})$ of (2.11), see Theorem 4.7 in [42]. We define the operators $DG_f \in L(C(\mathcal{O}), C(\mathcal{O}))$ and $D^2G_f \in L(C(\mathcal{O}), L(C(\mathcal{O}), C(\mathcal{O})))$ as

$$DG_f[h_1] = V_f[h_1 u_f], \quad (D^2G_f[h_1])[h_2] = V_f[h_1 DG_f[h_2]] + V_f[h_2 DG_f[h_1]], \quad h_1, h_2 \in C(\mathcal{O}). \tag{2.172}$$

The next lemma establishes that these operators are suitable Fréchet derivatives of G on the open subset $\{f \in C(\mathcal{O}), f > 0\}$ of $C(\mathcal{O})$.

Lemma 2.5.6. ***i)*** *For any $f \in C(\mathcal{O})$ with $f > 0$, we have $G(f) \in C(\mathcal{O})$. Moreover there exists $C > 0$ such that for any $f, \bar{f} \in C(\mathcal{O})$ with $f, \bar{f} > 0$,*

$$\|G(\bar{f}) - G(f)\|_\infty \le C\|\bar{f} - f\|_\infty, \tag{2.173}$$

as well as

$$\begin{aligned} \|G(\bar{f}) - G(f) - DG_f[\bar{f} - f]\|_\infty &\le C\|\bar{f} - f\|_\infty^2, \\ \|DG_{\bar{f}} - DG_f - D^2G_f[\bar{f} - f]\|_{L(C(\mathcal{O}),C(\mathcal{O}))} &\le C\|\bar{f} - f\|_\infty^2. \end{aligned} \tag{2.174}$$

ii) *For any integer $\alpha > d/2$ there exists a constant $C > 0$ such that for all $f \in H^\alpha$ with $\inf_{x\in\mathcal{O}} f(x) > 0$, we have*

$$\|G(f)\|_{H^2} \le C(\|f\|_{L^2} + \|g\|_{C^2(\partial\mathcal{O})}), \tag{2.175}$$

$$\|G(f)\|_{H^{\alpha+2}} \le C(1 + \|f\|_{H^\alpha}^{\alpha/2+1})\|g\|_{C^{\alpha+2}(\partial\mathcal{O})}. \tag{2.176}$$

Proof. The estimate (2.173) follows from the identity $G(\bar{f}) - G(f) = V_f[(\bar{f} - f)G(\bar{f})]$, (2.168) and (2.18). Arguing similarly and using (2.173), we further obtain

$$\begin{aligned} \|G(\bar{f}) - G(f) - DG_f[\bar{f} - f]\|_\infty &= \|V_f[(\bar{f} - f)(G(\bar{f}) - G(f))]\|_\infty \\ &\lesssim \|(\bar{f} - f)(G(\bar{f}) - G(f))\|_\infty \lesssim \|\bar{f} - f\|_\infty^2, \end{aligned}$$

which proves the first part of (2.174). For the second part of (2.174), we have for any $h \in C(\mathcal{O})$ that

$$\begin{aligned} DG_{\bar{f}}[h] - DG_f[h] &= V_{\bar{f}}[hu_{\bar{f}}] - V_f[hu_f] \\ &= V_{\bar{f}}[h(u_{\bar{f}} - u_f)] + (V_{\bar{f}} - V_f)[hu_f] \\ &= V_f[hDG_f[\bar{f} - f]] + R_1 + V_f[(\bar{f} - f)V_f[hu_f]] + R_2 \\ &= (D^2G_f[\bar{f} - f])[h] + R_1 + R_2, \end{aligned}$$

with remainder terms R_1, R_2 given by

$$\begin{aligned} R_1 &= [V_{\bar{f}} - V_f][h(u_{\bar{f}} - u_f)] + V_f[h(u_{\bar{f}} - u_f - DG[h])], \\ R_2 &= [V_{\bar{f}} - V_f](hu_f) - V_f[(\bar{f} - f)V_f[hu_f]]. \end{aligned}$$

Using the identity $(V_{\bar{f}} - V_f)\psi = V_f[(\bar{f} - f)V_{\bar{f}}[\psi]]$ with $\psi = h(u_{\bar{f}} - u_f)$, Lemma 2.5.5 as well as the first part of (2.174), we have

$$\|R_1\|_\infty \lesssim \|\bar{f} - f\|_\infty\|h(u_{\bar{f}} - u_f)\|_\infty + \|h\|_\infty\|u_{f+h} - u_f - D\bar{G}[h]\|_\infty \lesssim \|\bar{f} - f\|_\infty^2\|h\|_\infty,$$

and arguing similarly,

$$\|R_2\|_\infty = \|V_f[(\bar f - f)(V_{\bar f} - V_f)[hu_f]]\|_\infty \lesssim \|\bar f - f\|_\infty \|(V_{\bar f} - V_f)[hu_f]\|_\infty \lesssim \|\bar f - f\|_\infty^2 \|h\|_\infty.$$

This completes the proof of (2.174).

To prove (2.175), we use that $(\Delta, \mathrm{tr}) : H^2(\mathcal{O}) \to L^2 \times H^{3/2}(\partial\mathcal{O})$ [where tr denotes the boundary trace operator for the domain $\mathcal{O}$] is a topological isomorphism, see Theorem II.5.4 in [109], such that in particular

$$\|G(f)\|_{H^2} \lesssim \|fu_f\|_{L^2} + \|g\|_{C^2(\partial\mathcal{O})} \leq \|f\|_{L^2} + \|g\|_{C^2(\partial\mathcal{O})}.$$

where we also used (2.18). Finally, (2.176) is proved in Lemma 27 in [136]. □

2.5.2.1 Properties of the map Φ^*

We summarise some properties of 'regular' link functions from Definition 2.2.1. We recall the notation Φ^* for the associated composition operator from (2.15). For any $F \in C(\mathcal{O})$, define the operators $D\Phi^*_F \in L(C(\mathcal{O}), C(\mathcal{O}))$, $D^2\Phi^*_F \in L(C(\mathcal{O}), L(C(\mathcal{O}), C(\mathcal{O})))$ by

$$D\Phi^*_F[H] = H\Phi' \circ F, \qquad (D^2\Phi^*_F[H])[J] = HJ\Phi'' \circ F, \quad H, J \in C(\mathcal{O}). \tag{2.177}$$

Then for any $F, H, J \in C(\mathcal{O})$ and $x \in \mathcal{O}$, a Taylor expansion immediately implies that, with $\zeta_x, \bar\zeta_x$ denoting intermediate points between $F(x)$ and $(F+H)(x)$,

$$|(\Phi^*(F+H) - \Phi^*(F) - D\Phi^*_F[H])(x)| = |H^2(x)\Phi''(\zeta_x)/2| \leq \|H\|_\infty^2 \sup_{t\in\mathbb{R}} |\Phi''(t)|,$$
$$\big|\big(D\Phi^*_{F+H} - D\Phi^*_F - D^2\Phi^*_F[H]\big)[J](x)\big| = \big|J(x)H^2(x)\Phi'''(\bar\zeta_x)/2\big| \leq \|J\|_\infty \|H\|_\infty^2 \sup_{t\in\mathbb{R}} |\Phi'''(t)|,$$

whence $D\Phi^*, D^2\Phi^*$ are the Fréchet derivatives of $\Phi^* : C(\mathcal{O}) \to C(\mathcal{O})$.

We also need the basic fact that for any integer $\alpha > d/2$ there exists $C > 0$ such that for all $F \in H^\alpha(\mathcal{O})$,

$$\|\Phi \circ F\|_{H^\alpha} \leq C(1 + \|\Phi \circ F\|_{H^\alpha}^\alpha), \tag{2.178}$$

see Lemma 29 in [136]. Finally, note that by the definition of Φ, there exists $C' > 0$ such that for any $\bar F, F \in C(\mathcal{O})$,

$$\|\Phi \circ \bar F - \Phi \circ F\|_\infty \leq C\|\bar F - F\|_\infty, \quad \|\Phi \circ \bar F - \Phi \circ F\|_{L^2} \leq C\|\bar F - F\|_{L^2}. \tag{2.179}$$

2.5.2.2 Chain rule for Fréchet derivatives

Let U, V be normed vector spaces and $\mathcal{D} \subseteq U$ an open subset. For a map $T : \mathcal{D} \to V$ we denote by $DT_\theta \in L(U, V)$ and $D^2T_\theta \in L(U, L(U, V))$ the first and second order Fréchet derivatives at $\theta \in \mathcal{D}$, respectively, whenever they exist. The following basic lemma then follows directly from the chain rule.

Lemma 2.5.7. *Suppose U, V, W are (open subsets of) normed vector spaces, and suppose that $A : U \to V$ and $B : V \to W$ are both twice differentiable in the Fréchet sense. Then for any $\theta \in U$ and $H_1, H_2 \in U$, we have that $D(B \circ A)_\theta = DB_{A(\theta)} \circ DA_\theta$ and*

$$(D^2(B \circ A)_\theta[H_1])[H_2] = (D^2 B_{A(\theta)}[DA_\theta[H_1]])[DA_\theta[H_2]] + DB_{A(\theta)}[(D^2 A_\theta[H_1])[H_2]]. \tag{2.180}$$

2.5.3 Proof of Proposition 2.3.6

We first record the following basic lemma without proof.

Lemma 2.5.8. *Let $|\cdot|$ be an ellipsoidal norm on $\mathbb{R}^D$ with associated matrix M, $|\theta|^2 = \theta^T M \theta$ and define the function $n : \theta \to |\theta|$. Then for any $\theta \neq 0$, we have*

$$\nabla n(\theta) = \frac{M\theta}{|\theta|}, \qquad \nabla^2 n(\theta) = \frac{M}{|\theta|} - \frac{M\theta(M\theta)^T}{|\theta|^3}, \tag{2.181}$$

as well as the norm estimates

$$\|\nabla n(\theta)\|_{\mathbb{R}^D} \leq \sqrt{\lambda_{max}(M)}, \tag{2.182}$$

$$\|\nabla^2 n(\theta)\|_{op} \leq 2\lambda_{max}(M)/|\theta|_1. \tag{2.183}$$

Using Lemma 2.5.8, we prove the following bounds on the cut-off function α_η.

Lemma 2.5.9. *If $|\cdot|_1$ is an ellipsoidal norm with associated matrix M, $|\theta|_1^2 = \theta^T M \theta$, then the function α_η from (2.53) satisfies that for all $\theta \in \mathbb{R}^D$,*

$$\|\nabla \alpha_\eta(\theta)\|_{\mathbb{R}^D} \leq \frac{\|\alpha\|_{C^1}\sqrt{\lambda_{max}(M)}}{\eta}, \qquad \|\nabla^2 \alpha_\eta(\theta)\|_{op} \leq \frac{4\|\alpha\|_{C^2}\lambda_{max}(M)}{\eta^2}.$$

Proof. We may assume w.l.o.g. that $\theta_{init} = 0$ and we write $n(\theta) = |\theta|_1$. The gradient bound is obtained by the chain rule and (2.182):

$$\|\nabla \alpha_\eta(\theta)\|_{\mathbb{R}^D} = \left\|\eta^{-1}\alpha'(|\theta|_1/\eta)\nabla n(\theta)\right\|_{\mathbb{R}^D} \leq \eta^{-1}\|\alpha\|_{C_1}\sqrt{\lambda_{max}(M)}.$$

For the Hessian, we similarly employ the chain rule, (2.182), (2.183) as well as the fact that $\alpha'(t) = 0$ when $t \in (0, 3/4)$:

$$\begin{aligned}
\|\nabla^2 \alpha_\eta(\theta)\|_{op} &\leq \eta^{-2} \big\|\alpha''(|\theta|_1/\eta)\nabla n(\theta)\nabla n(\theta)^T\big\|_{op} + \eta^{-1}\big\|\alpha'(|\theta|_1/\eta)\nabla^2 n(\theta)\big\|_{op} \\
&\leq \eta^{-2}\|\alpha\|_{C^2}\|\nabla n(\theta)\|^2_{\mathbb{R}^D} + \eta^{-1}\|\alpha\|_{C_1}\mathbb{1}_{\{|\theta|\geq 3\eta/4\}} \cdot \frac{2\lambda_{max}(M)}{|\theta|_1} \\
&\leq 4\eta^{-2}\|\alpha\|_{C^2}\lambda_{max}(M).
\end{aligned}$$

□

We now turn to the proof of Proposition 2.3.6. Throughout, we work on the event $\mathcal{E}_{conv} \cap \mathcal{E}_{init}$ defined by (2.49),(2.50); moreover we assume without loss of generality that $\theta_{init} = 0$.

Proof of Proposition 2.3.6. We divide the proof into five steps.

1. Local lower bound for $\alpha_\eta \ell_N$. For the set

$$V := \{\theta : |\theta|_1 \leq 3\eta/4\},$$

by definition of $\mathcal{E}_{init}$, we have that $V \subseteq \mathcal{B}$. Thus using the definitions of $\mathcal{E}_{conv}$ and of α_η, we obtain

$$\inf_{\theta \in V} \lambda_{min}\big(-\nabla^2[\alpha_\eta \ell_N](\theta)\big) \geq N c_{min}/2. \tag{2.184}$$

2. Upper bound for $\alpha_\eta \ell_N$. By the chain rule, Lemma 2.5.9, the definition of $\mathcal{E}_{conv}$ and using that $\|\alpha\|_{C^2} \geq 1$, we obtain that for any $\theta \in \mathbb{R}^D$ and some $c = c(\alpha)$,

$$\begin{aligned}
\|\nabla^2[\alpha_\eta \ell_N](\theta)\|_{op} &\leq |\ell_N(\theta)|\|\nabla^2\alpha_\eta(\theta)\|_{op} + 2\|\nabla\alpha_\eta(\theta)\|_{\mathbb{R}^D}\|\nabla\ell_N(\theta)\|_{\mathbb{R}^D} + |\alpha_\eta(\theta)|\|\nabla^2\ell_N(\theta)\|_{op} \\
&\leq 2\sup_{\theta\in\mathcal{B}}\Big(\big[|\alpha_\eta(\theta)| + \|\nabla\alpha_\eta(\theta)\|_{\mathbb{R}^D} + \|\nabla^2\alpha_\eta(\theta)\|_{op}\big]\big[|\ell_N(\theta)| + \|\nabla\ell_N(\theta)\|_{\mathbb{R}^D} + \|\nabla^2\ell_N(\theta)\|_{op}\big]\Big) \\
&\leq c\big(1 + \lambda_{max}(M)/\eta^2\big) \cdot N(c_{max} + 1).
\end{aligned} \tag{2.185}$$

3. Global lower bound for $\nabla^2 g_\eta$. First we note that g_η is convex on all of $\mathbb{R}^D$: Indeed, this follows from the the identity $\gamma_\eta = \tilde{\gamma}_\eta * \varphi_{\eta/8}$, the convexity of the functions $n : \theta \mapsto |\theta|_1$, $\tilde{\gamma}_\eta$ and the fact that convolution with the positive function $\varphi_{\eta/8}$ preserves convexity. As g_η has C^2 regularity, it follows that $\nabla^2 g_\eta \succeq 0$ on all of $\mathbb{R}^D$.

We next prove a quantitative lower bound for $\nabla^2 g_\eta$ on the set V^c. By the chain rule and Lemma 2.5.8, we have that for any $\theta \in \mathbb{R}^D$, writing $v = \nabla n(\theta)$,

$$\begin{aligned}
\nabla^2 g_\eta(\theta) &= \gamma_\eta''(|\theta|_1)\nabla n(\theta)\nabla n(\theta)^T + \gamma_\eta'(|\theta|_1)\nabla^2 n(\theta) \\
&= \gamma_\eta''(|\theta|_1)vv^T + \frac{\gamma_\eta'(|\theta|_1)}{|\theta|_1}(M - vv^T) \\
&= \Big(\gamma_\eta''(|\theta|_1) - \frac{\gamma_\eta'(|\theta|_1)}{|\theta|_1}\Big)vv^T + \frac{\gamma_\eta'(|\theta|_1)}{|\theta|_1}M \\
&=: A(|\theta|_1)vv^T + B(|\theta|_1)M.
\end{aligned} \tag{2.186}$$

To derive lower bounds for the functions $B(\cdot)$ and $A(\cdot)$, we first observe that by the symmetry of $\varphi_{\eta/8}$ around 0, it holds for any $t \geq 3\eta/4$ that

$$\gamma_\eta'(t) = \int_{[-\eta/8,\eta/8]} \varphi_{\eta/8}(y) \cdot 2(t - y - 5\eta/8) = 2(t - 5\eta/8). \tag{2.187}$$

Thus the function $B(t) = \gamma_\eta'(t)/t$ strictly increases on $(3\eta/4, \infty)$, and for any $t \geq 3\eta/4$, we obtain

$$B(t) \geq B(3\eta/4) = \frac{\gamma_\eta'(3\eta/4)}{3\eta/4} = 2\frac{3\eta/4 - 5\eta/8}{3\eta/4} = \frac{1}{3}. \tag{2.188}$$

For the term $A(\cdot)$, we note that for any $t \geq 3\eta/4$, using that $\gamma_\eta''(t) = 2$ as well as (2.187), we have

$$A(t) = 2 - \frac{2(t - 5\eta/8)}{t} \geq 0. \tag{2.189}$$

Combining the displays (2.186), (2.188), (2.189), we have proved the lower bound

$$\inf_{\theta \in V^c} \lambda_{min}(\nabla^2 g_\eta(\theta)) \geq \lambda_{min}(M)/3, \quad . \tag{2.190}$$

4. Global upper bound for $\nabla^2 g_\eta$. We note that the functions $A(\cdot)$, $B(\cdot)$ from (2.186) satisfy

$$\sup_{t\in(0,\infty)} |A(t)| \leq \sup_{t\in(0,\infty)} |\gamma_\eta'(t)/t| + |\gamma_\eta''(t)| \leq 4, \qquad \sup_{t\in(0,\infty)} |B(t)| \leq \sup_{t\in(0,\infty)} |\gamma_\eta'(t)/t| \leq 2.$$

Hence, by (2.186) and Lemma 2.5.8, we obtain that

$$\|\nabla^2 g_\eta(\theta)\|_{op} \leq 4\|vv^T\|_{op} + 2\|M\|_{op} \leq 6\lambda_{max}(M), \quad \theta \in \mathbb{R}^D. \tag{2.191}$$

5. Combining the bounds. Combining the estimates (2.184), (2.185) and (2.190), we obtain that

$$\begin{aligned}\inf_{\theta\in V}\lambda_{min}\big(-\nabla^2\tilde{\ell}_N(\theta)\big) &\geq \frac{Nc_{min}}{2},\\ \inf_{\theta\in V^c}\lambda_{min}\big(-\nabla^2\tilde{\ell}_N(\theta)\big) &\geq \frac{K\lambda_{min}(M)}{3} - c(1+\lambda_{max}(M)/\eta^2)N(c_{max}+1).\end{aligned} \tag{2.192}$$

In particular, there exists $C \geq 3$ such that for any K satisfying (2.55), we have

$$\inf_{\theta\in\mathbb{R}^D}\lambda_{min}\big(-\nabla^2\tilde{\ell}_N(\theta)\big) \geq \min\Big\{\frac{Nc_{min}}{2},\frac{K\lambda_{min}(M)}{6}\Big\} = Nc_{min}/2,$$

which completes the proof of (2.56). To prove (2.57), we use (2.185), (2.191) and (2.55) to obtain that for all $\theta \neq \bar{\theta} \in \mathbb{R}^D$,

$$\begin{aligned}\frac{\|\nabla\tilde{\ell}_N(\theta)-\nabla\tilde{\ell}_N(\bar{\theta})\|_{\mathbb{R}^D}}{\|\theta-\bar{\theta}\|_{\mathbb{R}^D}} &\leq \sup_{\theta\in\mathbb{R}^D}\|\nabla^2\tilde{\ell}_N(\theta)\|_{op}\\ &\leq c\|\alpha\|_{C^2}(1+\lambda_{max}(M)/\eta^2)N(c_{max}+1)+6K\lambda_{max}(M)\\ &\leq 7K\lambda_{max}(M).\end{aligned}$$

□

2.5.4 Initialisation

In this section we prove the existence of polynomial time 'initialiser' $\theta_{init} = \theta_{init}(Z^{(N)}) \in \mathbb{R}^D$ (that lies in the region $\mathcal{B}_{1/\log N}$ from (2.99) of strong log-concavity of the posterior measure with high $P^N_{\theta_0}$-probability, when $\alpha > 6$), in the Schrödinger model.

Theorem 2.5.10. *Suppose $\theta_0 \in h^\alpha(\mathcal{O})$ for some $\alpha > 2 + d/2, d \leq 3$. Then there exists a measurable function $\theta_{init} \in \mathbb{R}^D$ of the data $Z^{(N)}$ from (2.20) and large enough $M' > 0$ such that for all $N, D \in \mathbb{N}$ and some $\bar{c} > 0$,*

$$P^N_{\theta_0}\big(\|\theta_{init}-\theta_{0,D}\|_{\mathbb{R}^D} > M'N^{-(\alpha-2)/(2\alpha+d)}\big) \lesssim e^{-\bar{c}N^{d/(2\alpha+d)}}.$$

Moreover θ_{init} is the output of a polynomial time algorithm involving $O(N^{b_0}), b_0 > 0$, iterations of gradient descent (each requiring a multiplication with a fixed $D' \times D'$ matrix, $D' \lesssim N^{d/(2\alpha+d)}$).

Proof. **Step I.** To start, consider the wavelet frame

$$\{\phi_{l,r}, 1 \leq r \leq N_l, l \in \mathbb{N}\}, N_l \lesssim 2^{ld},$$

of $L^2(\mathcal{O})$ constructed in Theorem 5.51 in [167]. Then for data arising from (2.19), choosing

$$2^J \simeq N^{1/(2\alpha+d)} = (N\delta_N^2)^{1/d}, \ \delta_N = N^{-\alpha/(2\alpha+d)}, \ n_J \equiv \sum_{l\le J} N_l \lesssim 2^{Jd},$$

and for multiscale vectors $(\lambda_{l,r}) \in \mathbb{R}^{n_J}$, define

$$\hat\lambda = \arg\min_{\lambda\in\mathbb{R}^{n_J}} \left[\frac{1}{N}\sum_{i=1}^N \big(Y_i - \sum_{l\le J,r} \lambda_{l,r}\phi_{l,r}(X_i)\big)^2 + \delta_N^2\|\lambda\|_{h^\alpha}^2\right], \quad \|\lambda\|_{h^\alpha}^2 = \sum_{l,r} 2^{2l\alpha}\lambda_{l,r}^2. \quad (2.193)$$

Next we set

$$\hat u = \hat u(Z^{(N)}) = \sum_{l\le J,r} \hat\lambda_{l,r}\phi_{l,r}, \quad u_{f_0,J} = \sum_{l\le J,r} \lambda_{0,l,r}\phi_{l,r},$$

where the $\lambda_{0,l,r} \in h^{\alpha+2}$ are frame coefficients of $u_{f_0} = \mathcal{G}(\theta_0) \in H^{\alpha+2}$ furnished by Theorem 5.51 in [167] and the elliptic regularity estimate (2.176). In particular by the Sobolev embedding $h^{\alpha+2} \subset b^\alpha_{\infty\infty}$ $(d<4)$ and again Theorem 5.51 in [167] we can prove

$$\|u_{f_0} - u_{f_0,J}\|_{L^2} \lesssim \|u_{f_0} - u_{f_0,J}\|_\infty \lesssim 2^{-J\alpha} \lesssim \delta_N. \quad (2.194)$$

We now apply a standard result from M estimation [170, 169], with empirical norms

$$\|u\|_{(N)}^2 = \frac{1}{N}\sum_{i=1}^N u^2(X_i),$$

conditional on the design $X_1, \dots, X_n$, to obtain the following bound.

Proposition 2.5.11. *We have for $\alpha > d/2$, all N and some constant $c > 0$,*

$$P_{\theta_0}^N\big(\|\hat u - u_{f_0}\|_{(N)}^2 + \delta_N^2\|\hat\lambda\|_{h^\alpha}^2 > \|u_{f_0} - u_{f_0,J}\|_{(N)}^2 + \delta_N^2\|\lambda_{0,l,r}\|_{h^\alpha}^2 |(X_i)_{i=1}^N\big) \le e^{-cN\delta_N^2}. \quad (2.195)$$

Proof. We apply Theorem 2.1 in [169]. We can bound the $\|\cdot\|_\infty$ and then also $\|\cdot\|_{(N)}$-metric entropy of the class of functions

$$\Big\{u : u = \sum_{l\le J,r} \lambda_{l,r}\phi_{l,r}; \|\lambda\|_{h^\alpha}^2 \le m\Big\}, \quad m > 0,$$

by the metric entropy of a ball of radius m in a H^α-Sobolev space, which by (4.184) in [72] is of order $H(\tau) \lesssim (m/\tau)^{d/\alpha}$ for every $m > 0$. Then arguing as in Section 3.1.1 in [169] (the only notational difference being that here $d > 1$), the result follows. □

This implies in particular, using $\|u\|_{(N)} \le \|u\|_\infty$, (2.194), $\lambda_{0,l,r} \in h^{\alpha+2}$ and Theorem 5.51 in [167], that for some $C, C' > 0$,

$$P^N_{\theta_0}(\|\hat{u}\|^2_{H^\alpha} > C) \le P^N_{\theta_0}(\|\hat{\lambda}\|^2_{h^\alpha} > C') \le \exp\{-cN\delta_N^2\}. \tag{2.196}$$

as well as

$$P^N_{\theta_0}(\|\hat{u} - u_{f_0,J}\|^2_{(N)} > C\delta_N^2) \le \exp\{-cN\delta_N^2\}. \tag{2.197}$$

In Step IV below we establish the following restricted isometry type bound

$$P^N_{\theta_0}\left(\Big|\frac{\|\hat{u} - u_{f_0,J}\|^2_{(N)}}{\|\hat{u} - u_{f_0,J}\|^2_{L^2}} - 1\Big| \le \frac{1}{2}\right) \ge 1 - c''e^{-c'N\delta_N^2} \tag{2.198}$$

for some constants $c', c'' > 0$ so that in particular

$$P^N_{\theta_0}\left(\frac{1}{2} \le \frac{\|\hat{u} - u_{f_0,J}\|^2_{(N)}}{\|\hat{u} - u_{f_0,J}\|^2_{L^2}} \le \frac{3}{2}\right) \ge 1 - c''e^{-c'N\delta_N^2}.$$

On the event $\mathcal{A}_N$ in the last probability we can write, using again (2.194) and (2.197), for M large enough,

$$\begin{aligned}
&P^N_{\theta_0}(\|\hat{u} - u_{f_0}\|^2_{L^2} > M\delta_N^2) \le P^N_{\theta_0}(\|\hat{u} - u_{f_0,J}\|^2_{L^2} > (M/2)\delta_N^2) \\
&\le P^N_{\theta_0}\left(\frac{\|\hat{u} - u_{f_0,J}\|^2_{L^2}}{\|\hat{u} - u_{f_0,J}\|^2_{(N)}}\|\hat{u} - u_{f_0,J}\|^2_{(N)} > (M/2)\delta_N^2, \mathcal{A}_N\right) + c''e^{-c'N\delta_N^2} \\
&\le P^N_{\theta_0}\left(\|\hat{u} - u_{f_0,J}\|^2_{(N)} > (M/4)\delta_N^2\right) + c''e^{-c'N\delta_N^2} \lesssim e^{-cN\delta_N^2} + e^{-c'N\delta_N^2}.
\end{aligned}$$

Overall what precedes implies that we can find M large enough such that for some constants $\bar{c}, \bar{c}' > 0$,

$$P^N_{\theta_0}(\|\hat{u} - u_{f_0}\|^2_{L^2} \le M\delta_N^2 \text{ and } \|\hat{u}\|^2_{H^\alpha} \le M) \ge 1 - \bar{c}'e^{-\bar{c}N\delta_N^2}. \tag{2.199}$$

Step II. By definition of the $\|\cdot\|_{h^\alpha}$-norm, the objective function minimised in (2.193) over $\mathbb{R}^{n_J}$ is m-strongly convex with convexity bound $m \ge \delta_N^2$. Moreover, noting that the sum-of-squares term Q_N appearing in (2.193) satisfies

$$\frac{\partial Q_N}{\partial \lambda_{l',r'}}(\lambda) = -\frac{2}{N}\sum_{i=1}^N [Y_i - \sum_{l\le J,r} \lambda_{l,r}\phi_{l,r}(X_i)]\phi_{l',r'}(X_i), \quad l' \le J,\ 1 \le r' \le N_{l'},$$

we can deduce that the gradient of the objective function is globally Lipschitz with constant at most of order $O(2^{Jd}) = O(N\delta_N^2)$, using standard properties of the wavelet frame from Definition 5.25 in [167]. Using (2.18), (2.96) and a standard tail inequality for χ^2-random variables (Theorem 3.1.9 in [72]), one shows further that for some $\bar{C} > 0$ and on events of sufficiently high $P_{\theta_0}^N$-probability,

$$Q_N(0) = \frac{1}{N}\sum_{i=1}^{N}\left(\varepsilon_i^2 + 2\varepsilon_i u_{f_0}(X_i) + u_{f_0}^2(X_i)\right) \leq \bar{C}.$$

By Proposition 2.5.2 and using the standard sequence norm inequality

$$\|v\|_{h^\beta} \leq 2^{J\beta}\|v\|_{\ell_2} \lesssim N^{\frac{\beta}{2\alpha+d}}\|v\|_{\ell_2}, \quad v \in \mathbb{R}^{n_J},\ \beta \geq 0,$$

we deduce that on preceding events and for any fixed $p > 0$ there exists $b_0 > 0$ such that the output $\lambda_{init} \in \mathbb{R}^{n_J}$ from $O(N^{b_0})$ iterations of gradient descent satisfies $\|\lambda_{init} - \hat{\lambda}\|_{h^\alpha} \leq N^{-p}$. In particular we can choose p such that, denoting

$$u_{init} := \sum_{l \leq J, r}\lambda_{init,l,r}\phi_{l,r},$$

we have that $\|\hat{u} - u_{init}\|_{H^\alpha} \lesssim \|\hat{\lambda} - \lambda_{init}\|_{h^\alpha} = o(\delta_N)$; hence by virtue of (2.199), we may restrict the rest of the proof to an event of sufficiently large probability where u_{init} satisfies

$$\|u_{init} - u_{f_0}\|_{L^2}^2 + \delta_N^2\|u_{init}\|_{H^\alpha}^2 \leq (2M+1)\delta_N^2. \tag{2.200}$$

Step III. From the interpolation inequality for Sobolev norms from Section 2.1.3 and (2.200) we now obtain, with sufficiently high $P_{\theta_0}^N$-probability,

$$\|u_{init} - u_{f_0}\|_{H^2} \leq \bar{M}N^{-(\alpha-2)/(2\alpha+d)} \tag{2.201}$$

and the Sobolev imbedding ($d < 4$) further implies $\|u_{init} - u_{f_0}\|_\infty \to 0$ as $N \to \infty$ so that we deduce from (2.119) $\hat{u} \geq u_{f_0}/2 \geq c > 0$ with sufficiently high $P_{\theta_0}^N$-probability. So on these events we can define a new estimator

$$f_{init} = \frac{\Delta u_{init}}{2u_{init}}, \quad \text{noting that } f_0 = \frac{\Delta u_{f_0}}{2u_{f_0}}. \tag{2.202}$$

For $F_{init} = \Phi^{-1} \circ f_{init}$, using also the regularity of the inverse link function (2.179), we then see

$$\|F_{init} - F_{\theta_0}\|_{L^2} \lesssim \|f_{init} - f_0\|_{L^2} \lesssim \|u_{init} - u_{f_0}\|_{H^2},$$

and hence for some $M' > 0$,

$$P_{\theta_0}^N\left(\|F_{init} - F_{\theta_0}\|_{L^2} \leq M'N^{-(\alpha-2)/(2\alpha+d)}\right) \geq 1 - \bar{c}'e^{-\bar{c}N\delta_N^2}.$$

We finally define θ_{init} as

$$\theta_{init} = (\langle F_{init}, e_k \rangle_{L^2} : k \leq D) \in \mathbb{R}^D, \quad D \in \mathbb{N},$$

the vector of the first D 'Fourier coefficients' of F_{init}. Then we obtain from Parseval's identity that $\|\theta_{init} - \theta_{0,D}\|_{\mathbb{R}^D} \leq \|F_{init} - F_{\theta_0}\|_{L^2}$, which combined with the last probability inequality establishes convergence rate desired in Theorem 2.5.10.

Step IV. Proof of (2.198). Let us introduce the symmetric $n_J \times n_J$, $n_J \lesssim 2^{Jd}$, matrices

$$\hat{\Gamma}_{(l,r),(l',r')} = \frac{1}{N}\sum_{i=1}^{N} \phi_{l,r}(X_i)\phi_{l',r'}(X_i), \quad \Gamma_{(l,r),(l',r')} = \int_{\mathcal{O}} \phi_{l,r}(x)\phi_{l',r'}(x)dP^X(x),$$

and vectors $(\hat{\lambda} = \hat{\lambda}_{l,r}), (\lambda_0 = \lambda_{0,l,r}) \in \mathbb{R}^{n_J}$. Then we can write

$$\|\hat{u} - u_{f_0,J}\|^2_{(N)} - \|\hat{u} - u_{f_0,J}\|^2_{L^2(\mathcal{O})} = (\hat{\lambda} - \lambda_0)^T(\hat{\Gamma} - \Gamma)(\hat{\lambda} - \lambda_0)$$

and hence (one minus the) probability relevant in (2.198) can be bounded as

$$\Pr\left(\Big|\frac{(\hat{\lambda} - \lambda_0)^T(\hat{\Gamma} - \Gamma)(\hat{\lambda} - \lambda_0)}{(\hat{\lambda} - \lambda_0)^T\Gamma(\hat{\lambda} - \lambda_0)}\Big| > 1/2\right) \leq \Pr\left(\sup_{v\in\mathbb{R}^{n_J}: v^T\Gamma v\leq 1} |v^T(\hat{\Gamma} - \Gamma)v| > 1/2\right).$$

We also note that by the frame property of the $\{\phi_{l,r}\}$, specifically from (5.252) in [167] with $s = 0, p = q = 2$, for any $u_v = \sum_{l\leq J,r} v_{l,r}\phi_{l,r}$ we have the norm equivalence

$$\|v\|^2_{\mathbb{R}^{n_J}} \simeq \|u_v\|^2_{L^2} = \sum_{l,l'\leq J,r,r'} v_{l,r}v_{l',r'}\Gamma_{(l,r),(l',r')} = v^T\Gamma v =: \|v\|^2_{\Gamma}, \tag{2.203}$$

with the constants implied by $\simeq$ independent of J. Next for any $\kappa > 0$ let

$$\{v_m, m = 1, \dots, M_{J,\kappa}\}, \ M_{J,\kappa} \lesssim (3/\kappa)^{n_J}$$

denote the centres of balls of $\|\cdot\|_\Gamma$-radius κ covering the unit ball V_Γ of $(\mathbb{R}^{n_J}, \|\cdot\|_\Gamma)$ (e.g., as in Prop. 4.3.34 in [72] and using (2.203)). Then using the Cauchy-Schwarz inequality

$$\begin{aligned}
|v^T(\hat{\Gamma} - \Gamma)v| &= |(v - v_m + v_m)^T(\hat{\Gamma} - \Gamma)(v - v_m + v_m)| \\
&\leq \|v - v_m\|^2_\Gamma \sup_{v\in V_\Gamma} |v^T(\hat{\Gamma} - \Gamma)v| + 2\|v - v_m\|_\Gamma \|(\hat{\Gamma} - \Gamma)v\|_\Gamma + |v_m^T(\hat{\Gamma} - \Gamma)v_m| \\
&\leq (\kappa^2 + 2\kappa) \sup_{v\in V_\Gamma} |v^T(\hat{\Gamma} - \Gamma)v| + |v_m^T(\hat{\Gamma} - \Gamma)v_m|
\end{aligned}$$

so choosing κ small enough so that $\kappa^2 + 2\kappa < 1/4$ we obtain

$$\sup_{v\in V_\Gamma} |v^T(\hat{\Gamma} - \Gamma)v| \leq (4/3) \max_{m=1,\dots,M_J} |v_m^T(\hat{\Gamma} - \Gamma)v_m|, \quad M_J \equiv M_{J,\kappa}. \tag{2.204}$$

In particular, using also that $M_J \lesssim e^{c_0 2^{Jd}} \le e^{c_1 N\delta_N^2}$, the last probability is thus bounded by

$$\Pr\Big(\max_{m=1,\dots,M_J} |v_m^T(\hat\Gamma - \Gamma)v_m| > 1/4\Big) \le e^{c_1 N\delta_N^2} \max_m \Pr\Big(|v_m^T(\hat\Gamma - \Gamma)v_m| > 1/4\Big). \tag{2.205}$$

Each of the last probabilities can be bounded by Bernstein's inequality (Prop. 3.1.7 in [72]) applied to

$$v_m^T(\hat\Gamma - \Gamma)v_m = \frac{1}{N}\sum_{i=1}^N Z_i - EZ_i,$$

with i.i.d. variables $Z_i = Z_{i,m}$ given by

$$Z_i = \sum_{l,l'\le J,r,r'} v_{m,l,r}v_{m,l',r'}\phi_{l,r}(X_i)\phi_{l',r'}(X_i) = \sum_{l\le J,r} v_{m,l,r}\phi_{l,r}(X_i) \sum_{l'\le J,r'} v_{m,l',r'}\phi_{l',r'}(X_i), \tag{2.206}$$

wit vectors v_m all satisfying $\|v_m\|_\Gamma \le 1$. For these variables we have from the Cauchy-Schwarz inequality

$$|Z_i| \le \Big|\sum_{l\le J,r} v_{m,l,r}\phi_{l,r}(\cdot)\Big|^2 \le \|v_m\|^2_{\mathbb{R}^{n_J}} \sum_{l\le J,r} (\phi_{l,r}(\cdot))^2 \le c2^{Jd} \equiv U$$

where the constant c depends only on the wavelet frame (cf. (2.203) and also Definition 5.25 in [167]). Similarly, using the previous estimate, we can bound

$$EZ_i^2 = E\Big[\sum_{l\le J,r} v_{m,l,r}\phi_{l,r}(X_i)\Big]^4 \le U\int_{\mathcal{O}}\Big[\sum_{l\le J,r} v_{m,l,r}\phi_{l,r}(x)\Big]^2 dx = U\|v_m\|_\Gamma^2 \le U.$$

Now Proposition 3.1.7 in [72] implies for some constant $c_0 > 0$

$$\Pr\Big(N|v_m(\hat\Gamma - \Gamma)v_m| > N/4\Big) \le 2\exp\Big\{-\frac{N^2/16}{2NU + (2/12)NU}\Big\} \le 2e^{-c_0/\delta_N^2}$$

since $U = c2^{Jd} \simeq N\delta_N^2$. Now since $\alpha > d/2$ we have $\delta_N^2 = o(1/\sqrt{N})$ and thus $(1/\delta_N^2) \gg N\delta_N^2$ which means that the r.h.s in (2.205) is bounded by a constant multiple of $e^{-c'N\delta_N^2}$ for some $c' > 0$, completing the proof. $\square$

Chapter 3

Convergence rates for Penalised Least Squares estimators

In this chapter we study convergence rates for Tikhonov-type penalised least squares estimators in inverse regression problems. The main examples are two non-linear models, respectively arising with an elliptic divergence form PDE and with a steady-state Schrödinger equation, both of which we previously encountered in Section 1.1. In both cases the parameter f is an unknown coefficient function of a partial differential operator L_f and the unique solution u_f of an elliptic boundary value problem corresponding to L_f is observed, corrupted by additive Gaussian white noise.

The penalty terms we consider are of squared Sobolev-norm type, and thus the estimators $\hat{f}$ can also be interpreted as Bayesian MAP-estimators corresponding to some Gaussian process prior. We derive rates of convergence of $\hat{f}$ and of $u_{\hat{f}}$, to f, u_f, respectively. We prove that the rates obtained are minimax-optimal in prediction loss. Our bounds are derived from a general convergence rate result for non-linear inverse problems whose forward map satisfies a modulus of continuity condition, a result of independent interest that is applicable also to linear inverse problems, illustrated in an example with the Radon transform.

3.1 Introduction

Observations obeying certain physical laws can often be described by a partial differential equation (PDE). Real world measurements carry statistical noise and thus do not generally exactly exhibit the idealised pattern of the PDE, but it is desirable that recovery of parameters from data is consistent with the PDE structure. In the mathematical literature on inverse problems several algorithms that incorporate such constraints have been proposed, notably optimisation based methods such as Tikhonov regularisation [63, 17] and maximum a posteriori (MAP) estimates related to Bayesian inversion techniques [155, 53]. In statistical terminology these methods can be viewed as penalised least squares estimators over parameter spaces of regression functions that are restricted to lie in the range of some 'forward operator' $\mathscr{G}$ describing the solution map of the PDE. The case where $\mathscr{G}$ is linear is reasonably well studied in the inverse problems literature, but already in basic elliptic PDE examples, the map $\mathscr{G}$ is *non-linear* and the analysis is more involved. The observation scheme considered here will be a natural continuous analogue of the standard Gaussian regression model

$$Y_i = u_f(x_i) + \varepsilon_i,\ i = 1, \dots, N;\ \{\varepsilon_i\} \sim^{i.i.d.} N(0,1), \tag{3.1}$$

where $(x_i)_{i=1}^N$ are 'equally spaced' design points on a bounded domain $\mathcal{O} \subset \mathbb{R}^d$ with smooth boundary $\partial\mathcal{O}$. The function $u_f : \mathcal{O} \to \mathbb{R}$ is, in our first example, the solution $u = u_f$ of the elliptic PDE (with ∇ denoting the gradient and $\nabla\cdot$ the divergence operator)

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

where $g > 0$ is a given source function defined on $\mathcal{O}$ and $f : \mathcal{O} \to (0, \infty)$ is an unknown *conductivity (or diffusion) coefficient.* The second model example arises with solutions $u = u_f$ of the time-independent Schrödinger equation (with Δ equal to the standard Laplacian operator)

$$\begin{cases} \Delta u - 2fu = 0 & \text{on } \mathcal{O}, \\ u = g & \text{on } \partial O, \end{cases} \tag{3.3}$$

corresponding to the unknown *attenuation potential (or reaction coefficient)* $f : \mathcal{O} \to (0, \infty)$, and given positive 'boundary temperatures' $g > 0$. Both PDEs have a fundamental physical interpretation and feature in many application areas, see, e.g., [63, 20, 87, 10, 155, 24, 53], and references therein.

When $f > 0$ belongs to some Sobolev space $H^\alpha(\mathcal{O})$ for appropriate $\alpha > 0$, unique solutions u_f of the PDEs (3.2), (3.3) exist, and the 'forward' map $f \mapsto u_f$ is non-linear. [In fact, in (3.3) only $f \geq 0$ is required.] A natural method to estimate f is by a penalised least squares approach: one minimises over $f \in H^\alpha(\mathcal{O})$ with $f > 0$ the squared Euclidean distance

$$Q_N(f) = \|Y - u_f\|^2$$

of the observation vector $(Y_i : i = 1, \dots, N)$ to the fitted values $(u_f(x_i) : i = 1, \dots, N)$, and penalises too complex solutions f by, for instance, an additive Sobolev norm $\|\cdot\|_{H^\alpha}$ -type penalty. The (from a PDE perspective) natural constraint $f > 0$ can be incorporated by a smooth one-to-one transformation Φ of the penalty function, and a final estimator $\hat{f}$ minimises a criterion function of the form

$$Q_N(f) + \lambda^2 \|\Phi^{-1}[f]\|_{H^\alpha}^2,$$

over $f \in H^\alpha(\mathcal{O})$ with $f > 0$, where λ is a scalar regularisation parameter to be chosen. Both Tikhonov regularisers as well as Bayesian maximum a posteriori (MAP) estimates arising from suitable Gaussian priors fall into this class of estimators. We show in this chapter that suitable choices of λ, α, Φ give rise to statistically optimal solutions of the above PDE constrained regression problems from data (3.1), in prediction loss. The convergence rates obtained can be combined with 'stability estimates' to obtain bounds also for the recovery of the parameter f itself.

Our main results are based on a general convergence rate theorem for minimisers over H^α of functionals of the form

$$F \mapsto \|Y - \mathscr{G}(F)\|^2 + \lambda^2 \|F\|_{H^\alpha}^2$$

in possibly non-linear inverse problems whose forward map $F \mapsto \mathscr{G}(F)$ satisfies a certain modulus of continuity assumption between Hilbert spaces. This result, which adapts M-estimation techniques [169, 170] to the inverse problems setting, is of independent interest, and provides novel results also for linear forward maps, see Remark 3.2.5 for an application to Radon transforms.

For sake of conciseness, our theory is given in the Gaussian white noise model introduced in (3.4) below – it serves as an asymptotically equivalent (see [30, 148]) continuous analogue of the discrete model (3.1), and facilitates the application of PDE techniques in our proofs. Transferring our results to discrete regression models is possible, but the additional difficulties are mostly of a technical nature and will note be pursued here.

Recovery for non-linear inverse problems such as those mentioned above has been studied initially in the deterministic regularisation literature [64, 130, 154, 63, 160], and the convergence rate theory developed there has been adapted to the statistical regression model (3.1) in [20, 21, 87, 112]. These results all assume that a suitable Fréchet derivative $D\mathscr{G}$ of the non-linear forward map $\mathscr{G}$ exists at the 'true' parameter F, and moreover require that F lies in the range of the adjoint operator of $D\mathscr{G}$ – the so called 'source condition'. Particularly for the PDE (3.2), such conditions are problematic and do not hold in general for rich enough classes of F's (such as Sobolev balls) unless one makes very stringent additional model assumptions. Our results circumvent such source conditions. Further remarks, including a discussion of related convergence analysis of estimators obtained from Bayesian inversion techniques [177, 50, 131] can be found in Section 3.3.4.

The article is organised as follows. The main results are stated in Sections 3.2 and 3.3; their proofs are contained in Sections 3.4 and 3.7. Some key auxiliary results about the elliptic PDE (3.2)-(3.3) and the 'link functions' Φ used below are proved in Section 3.5 and 3.6 respectively.

3.1.1 Some preliminaries and basic notation

Throughout, $\mathcal{O} \subseteq \mathbb{R}^d$, $d \geq 1$, denotes a bounded non-empty C^∞-domain (an open bounded set with smooth boundary) with closure $\bar{\mathcal{O}}$. The usual space $L^2(\mathcal{O})$ of square integrable functions carries a norm $\|\cdot\|_{L^2(\mathcal{O})}$ induced by the inner product

$$\langle h_1, h_2 \rangle_{L^2(\mathcal{O})} = \int_{\mathcal{O}} h_1(x) h_2(x) dx, \quad h_1, h_2 \in L^2(\mathcal{O}),$$

where dx denotes Lebesgue measure. For any multi-index $i = (i_1, ..., i_d)$ of 'order' $|i|$, let D^i denote the i-th (weak) partial derivative operator of order $|i|$. Then for integer $\alpha \geq 0$, the usual Sobolev spaces are defined as

$$H^\alpha(\mathcal{O}) := \left\{ f \in L^2(\mathcal{O}) \;\middle|\; \text{for all } |i| \leq \alpha,\ D^i f \text{ exists and } D^i f \in L^2(\mathcal{O}) \right\},$$

normed by $\|f\|_{H^\alpha(\mathcal{O})} = \sum_{|i| \leq \alpha} \|D^i f\|_{L^2(\mathcal{O})}$. For non-integer real values $\alpha \geq 0$, we define $H^\alpha(\mathcal{O})$ by interpolation, see, e.g., [108] or [165].

The spaces of bounded and continuous functions on $\mathcal{O}$ and $\bar{\mathcal{O}}$ are denoted by $C(\mathcal{O})$ and $C(\bar{\mathcal{O}})$, respectively, equipped with the supremum norm $\|\cdot\|_\infty$. For $\eta \in \mathbb{N}$, the space of η-times differentiable functions on $\mathcal{O}$ with (bounded) uniformly continuous derivatives is denoted by $C^\eta(\mathcal{O})$. For $\eta > 0, \eta \notin \mathbb{N}$, we say $f \in C^\eta(\mathcal{O})$ if for all multi-indices β with $|\beta| \leq \lfloor \eta \rfloor$ (the integer part of η), $D^\beta f$ exists and is $\eta - \lfloor \eta \rfloor$-Hölder continuous. The norm on $C^\eta(\mathcal{O})$ is

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

We also define the set of smooth functions as $C^\infty(\mathcal{O}) = \cap_{\eta > 0} C^\eta(\mathcal{O})$ and its subspace $C_c^\infty(\mathcal{O})$ of functions compactly supported in $\mathcal{O}$.

The previous definitions will be used also for $\mathcal{O}$ replaced by $\partial\mathcal{O}$ or $\mathbb{R}^d$. When there is no ambiguity, we omit $\mathcal{O}$ from the notation.

For any normed linear space $(X, \|\cdot\|_X)$ its topological dual space is

$$X^* := \{ L : X \to \mathbb{R} \text{ linear s.t. } \exists C > 0\ \forall x \in X :\ |L(x)| \leq C \|x\|_X \},$$

which is a Banach space for the norm $\|L\|_{X^*} = \sup_{x \in X} |L(x)| / \|x\|_X$.

We need further Sobolev-type spaces to address routine subtleties of the behaviour of functions near $\partial\mathcal{O}$: denote by $H_c^\alpha(\mathcal{O})$ the completion of $C_c^\infty(\mathcal{O})$ for the $H^\alpha(\mathcal{O})$-norm, and let $\tilde{H}^\alpha(\mathcal{O})$ denote the closed subspace of $H^\alpha(\mathbb{R}^d)$ consisting of functions supported in $\bar{\mathcal{O}}$. We have $H_c^\alpha(\mathcal{O}) = \tilde{H}^\alpha(\mathcal{O})$ unless $\alpha = k + 1/2, k \in \mathbb{N}$ (Section 4.3.2 in [165]), and one defines negative order Sobolev spaces $H^{-\kappa}(\mathcal{O}) = (\tilde{H}^\kappa(\mathcal{O}))^*, \kappa > 0$, cf. also Theorem 3.30 in [121].

We use the symbols "$\lesssim, \gtrsim$" for inequalities that hold up to multiplicative constants that are universal, or whose dependence on other constants will be clear from the context. We also use the standard notation $\mathbb{R}_+ := \{x | x \geq 0\}$ and $a \vee b := \max\{a, b\}$ for $a, b \in \mathbb{R}$.

3.2 A convergence rate result for general inverse problems

3.2.1 Forward map and white noise model

Let $\mathbb{H}$ be a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$. Suppose that $\tilde{\mathcal{V}} \subseteq L^2(\mathcal{O})$ and that

$$\mathscr{G} : \tilde{\mathcal{V}} \to \mathbb{H}, \qquad F \mapsto \mathscr{G}(F),$$

is a given 'forward' map. For some $F \in \tilde{\mathcal{V}}$, and for scalar 'noise level' $\varepsilon > 0$, we observe a realisation of the equation

$$Y^{(\varepsilon)} = \mathscr{G}(F) + \varepsilon \mathbb{W}, \tag{3.4}$$

where $(\mathbb{W}(\psi) : \psi \in \mathbb{H})$ is a centred Gaussian white noise process indexed by the Hilbert space $\mathbb{H}$ (see p.19-20 in [72]). Let $\mathbb{E}_F^\varepsilon, F \in \tilde{\mathcal{V}}$, denote the expectation operator under the law $\mathbb{P}_F^\varepsilon$ of $Y^{(\varepsilon)}$ from (3.4). Observing (3.4) means to observe a realisation of the Gaussian process $(\langle Y^{(\varepsilon)}, \psi \rangle_{\mathbb{H}} : \psi \in \mathbb{H})$ with marginal distributions

$$\langle Y^{(\varepsilon)}, \psi \rangle_{\mathbb{H}} \sim N(\langle \mathscr{G}(F), \psi \rangle_{\mathbb{H}}, \varepsilon^2 \|\psi\|_{\mathbb{H}}^2).$$

In the case $\mathbb{H} = L^2(\mathcal{O})$ relevant in Section 3.3 below, (3.4) can be interpreted as a Gaussian shift experiment in the Sobolev space $H^{-\kappa}(\mathcal{O}), \kappa > d/2$ (see, e.g., [33, 131]), and also serves as a theoretically convenient (and, for $\varepsilon = 1/\sqrt{N}$, as $N \to \infty$ asymptotically equivalent) continuous surrogate model for observing $(Y_i, x_i)_{i=1}^N$ in the standard fixed design Gaussian regression model

$$Y_i = \mathscr{G}(F)(x_i) + \varepsilon_i, \ i = 1, \dots, N, \ \{\varepsilon_i\} \sim^{i.i.d.} N(0, 1), \tag{3.5}$$

where the x_i are 'equally spaced' design points in the domain $\mathcal{O}$ (see [30, 148]).

In the discrete model (3.5) the least squares criterion can be decomposed as $\|Y - \mathscr{G}(F)\|_{\mathbb{R}^N}^2 = \|Y\|_{\mathbb{R}^N}^2 - 2\langle Y, \mathscr{G}(F) \rangle_{\mathbb{R}^N} + \|\mathscr{G}(F)\|_{\mathbb{R}^N}^2$. The first term $\|Y\|_{\mathbb{R}^N}^2$ is independent of F and can be neglected when optimising in F. In the continuous model (3.4) we have $\|Y\|_{\mathbb{H}} = \infty$ a.s. (unless $\dim(\mathbb{H}) < \infty$), which motivates to define a 'Tikhonov-regularised' functional

$$\mathscr{J}_{\lambda,\varepsilon} : \tilde{\mathcal{V}} \to \mathbb{R}, \quad \mathscr{J}_{\lambda,\varepsilon}(F) := 2\langle Y^{(\varepsilon)}, \mathscr{G}(F) \rangle_{\mathbb{H}} - \|\mathscr{G}(F)\|_{\mathbb{H}}^2 - \lambda^2 \|F\|_{H^\alpha}^2, \tag{3.6}$$

where $\lambda > 0$ is a regularisation parameter to be chosen, and where we set $\mathscr{J}_{\lambda,\varepsilon}(F) = -\infty$ for $F \notin H^\alpha$. Maximising $\mathscr{J}_{\lambda,\varepsilon}$ thus amounts to minimising the natural least squares fit with a $H^\alpha(\mathcal{O})$-penalty for F, and we note that it also corresponds to maximising the penalised log-likehood function arising from (3.4), see, e.g., [131], Section 7.4. In all that follows $\|\cdot\|_{H^\alpha}$ could be replaced by any equivalent norm on $H^\alpha(\mathcal{O})$.

We note that when $\mathscr{G}$ is non-linear, computation of a global maximiser of the (then non-convex) functional $\mathscr{J}_{\lambda,\varepsilon}$ may pose an infeasible computational task in practice. Nevertheless, the convergence rates we obtain below provide a first rigorous understanding of the statistical complexity of the PDE inference problems at hand. It is an interesting open question whether algorithms that are computable in 'polynomial time' can attain the same performance guarantees. This is subject of ongoing research (see, e.g., [126]) and beyond the scope of the present thesis.

3.2.2 Results

For $F_1 \in \tilde{\mathcal{V}} \cap H^\alpha$, $F_2 \in \tilde{\mathcal{V}}$ and $\lambda > 0$, define the functional

$$\tau_\lambda^2(F_1, F_2) := \|\mathscr{G}(F_1) - \mathscr{G}(F_2)\|_{\mathbb{H}}^2 + \lambda^2 \|F_1\|_{H^\alpha}^2. \tag{3.7}$$

The main result of this section, Theorem 3.2.2, proves the existence of maximisers $\hat{F}$ for $\mathscr{J}_{\lambda,\varepsilon}$ over suitable subsets $\mathcal{V} \subseteq \tilde{\mathcal{V}} \cap H^\alpha$ and concentration properties for $\tau_\lambda(\hat{F}, F_0)$, where F_0 is the 'true' function generating the law $\mathbb{P}_{F_0}^\varepsilon$ from equation (3.4). Note that bounds for $\tau_\lambda(\hat{F}, F_0)$ simultaneously control the 'prediction error' $\|\mathscr{G}(\hat{F}) - \mathscr{G}(F_0)\|_{\mathbb{H}}$ as well as the regularity $\|\hat{F}\|_{H^\alpha}$ of the estimated output $\hat{F}$.

Theorem 3.2.2 is proved under a general 'modulus of continuity' condition on the map $\mathscr{G}$ which reads as follows.

Definition 3.2.1. *Let $\alpha, \gamma, \kappa \in \mathbb{R}_+$ be non-negative real numbers and $\tilde{\mathcal{V}} \subseteq L^2(\mathcal{O})$. Set $\mathcal{H} := H^\alpha(\mathcal{O})$ if $\kappa < 1/2$, and $\mathcal{H} := H_c^\alpha(\mathcal{O})$ if $\kappa \geq 1/2$. A map $\mathscr{G} : \tilde{\mathcal{V}} \to \mathbb{H}$ is called (κ, γ, α)-regular if there exists a constant $C > 0$ such that for all $F, H \in \tilde{\mathcal{V}} \cap \mathcal{H}$, we have*

$$\|\mathscr{G}(F) - \mathscr{G}(H)\|_{\mathbb{H}} \leq C\big(1 + \|F\|_{H^\alpha(\mathcal{O})}^\gamma \vee \|H\|_{H^\alpha(\mathcal{O})}^\gamma\big)\|F - H\|_{(H^\kappa(\mathcal{O}))^*}, \tag{3.8}$$

This condition is easily checked for 'κ-smoothing' *linear* maps $\mathscr{G}$ with $\gamma = 0$, see Remark 3.2.5 for an example. But (3.8) also allows for certain non-linearities of $\mathscr{G}$ on unbounded parameter spaces $\tilde{\mathcal{V}}$ that will be seen later on to accommodate the forward maps induced by the PDEs (3.2), (3.3). See also Remarks 3.2.6, 3.3.10 below.

Theorem 3.2.2. *Suppose that $\mathscr{G} : \tilde{\mathcal{V}} \to \mathbb{H}$ is a (κ, γ, α)-regular map for some integer $\alpha > (d/2 - \kappa) \vee (\gamma d/2 - \kappa)$. Let $Y^{(\varepsilon)} \sim \mathbb{P}_{F_0}^\varepsilon$ from (3.4) for some fixed $F_0 \in \tilde{\mathcal{V}}$. Then the following holds.*

1. Let $\mathcal{V} \subseteq \tilde{\mathcal{V}} \cap \mathcal{H}$ be closed for the weak topology of the Hilbert space $\mathcal{H}$. Then for all $\lambda, \varepsilon > 0$, almost surely under $\mathbb{P}_{F_0}^\varepsilon$, there exists a maximiser $\hat{F} = \hat{F}_{\lambda,\varepsilon} \in \mathcal{V}$ of $\mathscr{J}_{\lambda,\varepsilon}$ from (3.6) *over $\mathcal{V}$, satisfying*

$$\sup_{F\in\mathcal{V}} \mathscr{J}_{\lambda,\varepsilon}(F) = \mathscr{J}_{\lambda,\varepsilon}(\hat{F}). \tag{3.9}$$

2. Let $\mathcal{V} \subseteq \tilde{\mathcal{V}} \cap \mathcal{H}$. There exist constants $c_1, c_2, c_3 > 0$ such that for all $\varepsilon, \lambda, \delta > 0$ satisfying

$$\varepsilon^{-1}\delta \geq c_1\big(1 + \lambda^{-\frac{1}{2s}}\big(1 + (\delta/\lambda)^{\frac{\gamma}{2s}}\big)\big), \;\; s := (\alpha+\kappa)/d, \tag{3.10}$$

all $R \geq \delta$, any maximiser $\hat{F} = \hat{F}_{\lambda,\varepsilon} \in \mathcal{V}$ of $\mathscr{J}_{\lambda,\varepsilon}$ over $\mathcal{V}$ and any $F_* \in \mathcal{V}$, we have

$$\mathbb{P}^{\varepsilon}_{F_0}\big(\tau^2_\lambda(\hat{F}, F_0) \geq 2(\tau^2_\lambda(F_*, F_0) + R^2)\big) \leq c_2 \exp\Big(-\frac{R^2}{c_2^2\varepsilon^2}\Big), \tag{3.11}$$

and also

$$\mathbb{E}^{\varepsilon}_{F_0}\left[\tau^2_\lambda(\hat{F}, F_0)\right] \leq c_3\left(\tau^2_\lambda(F_*, F_0) + \delta^2 + \varepsilon^2\right). \tag{3.12}$$

Various applications of Theorem 3.2.2 for specific choices of κ, γ, $\mathcal{V}$ and $\tilde{\mathcal{V}}$ will be illustrated in the following - besides the main PDE applications from Section 3.3, see Remarks 3.2.4, 3.3.10 and 3.3.11 as well as Example 3.2.5 below.

Theorem 3.2.2 does not necessarily require $F_0 \in \mathcal{V}$ as long as F_0 can be suitably approximated by some $F_* \in \mathcal{V}$, see Remark 3.2.4 for an instance of when this is relevant. If $F_0 \in \mathcal{V}$ then we can set $F_* = F_0$ in the above theorem and obtain the following convergence rates, which are well known to be optimal for κ-smoothing linear forward maps $\mathscr{G}$, and which will be seen to be optimal also for the non-linear inverse problems arising from the PDE models (3.2) and (3.3).

Corollary 3.2.3. *Under the conditions of Part 2 of Theorem 3.2.2, for all $R > 0$ there exists $c < \infty$ such that for all $\varepsilon > 0$ small enough, $\lambda = \varepsilon^{2(\alpha+\kappa)/(2(\alpha+\kappa)+d)}$ and any maximizer $\hat{F}_{\lambda,\varepsilon}$ of $\mathscr{J}_{\lambda,\varepsilon}$ over $\mathcal{V}$,*

$$\sup_{F_0\in\mathcal{V}:\|F_0\|_{H^\alpha}\leq R} \mathbb{E}^{\varepsilon}_{F_0}\left\|\mathscr{G}(\hat{F}_{\lambda,\varepsilon}) - \mathscr{G}(F_0)\right\|_{\mathbb{H}} \leq c\varepsilon^{\frac{2(\alpha+\kappa)}{2(\alpha+\kappa)+d}}. \tag{3.13}$$

When images of $\|\cdot\|_{H^\alpha}$-bounded subsets of $\mathcal{V}$ under a forward map $\mathscr{G} : L^2(\mathcal{O}) \to L^2(\mathcal{O})$ are bounded in $H^\beta(\mathcal{O})$ for some $\beta > 0$, then the L^2-bound (3.13) extends (via interpolation and bounds for $\|\hat{F}\|_{H^\alpha}$ implied by Theorem 3.2.2) to H^η-norms, $\eta \in [0, \beta]$, which in turn can be used to obtain convergence rates also for $\hat{F} - F_0$ by using stability estimates. See the results in Section 3.3 and also Example 3.2.5 below for examples.

Remark 3.2.4 (MAP estimates)**.** Let Π be a Gaussian process prior measure for F with reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ and RKHS-norm $\bar{\lambda}\|\cdot\|_{H^\alpha}, \bar{\lambda} > 0$. Taking note of the form of the likelihood function in the model (3.4) (see, e.g., Section 7.4 in [131]), maximisers $\hat{F}$ of $\mathscr{J}_{\lambda,\varepsilon}$ over $\mathcal{V} = \mathcal{H}$ with $\lambda = \varepsilon\bar{\lambda}$ have a formal interpretation as maximum a posteriori (MAP) estimators for the resulting posterior distributions $\Pi(\cdot|Y^{(\varepsilon)})$, see also

[50, 84]. For instance, let $\alpha > d/2, \kappa \geq 0$, and consider a *linear* inverse problem where for $\beta = \alpha - d/2$ and $\tilde{\mathcal{V}} = H^\beta(\mathcal{O})$, $\mathscr{G} : H^\beta(\mathcal{O}) \to \mathbb{H}$ is a linear map satisfying (3.8) with $\gamma = 0$ for all $F, H \in H^\beta(\mathcal{O})$. Then, applying Theorem 3.2.2 with $\lambda = \varepsilon$ (so that $\bar{\lambda} = 1$) and $\delta \approx \varepsilon^{(2\beta+2\kappa)/(2\beta+2\kappa+d)}$ yields

$$\sup_{F_0 \in \tilde{H}^\beta(\mathcal{O}_0): \|F_0\|_{\tilde{H}^\beta} \leq R} \mathbb{E}^\varepsilon_{F_0} \left\| \mathscr{G}(\hat{F}) - \mathscr{G}(F_0) \right\|_{\mathbb{H}} \lesssim \delta, \quad R > 0, \tag{3.14}$$

for any fixed sub-domain $\mathcal{O}_0$ such that $\bar{\mathcal{O}}_0 \subsetneq \mathcal{O}$. Indeed, one easily checks (3.10), and given $F_0 \in \tilde{H}^\beta(\mathcal{O}_0)$ set $F_* = \zeta \mathscr{F}^{-1}[(1_{[|\cdot| \leq (\delta/\lambda)^{2/d}]} \mathscr{F}[F_0]] \in H^\alpha_c(\mathcal{O})$, where $\zeta \in C^\infty_c(\mathcal{O})$ is such that $\zeta = 1$ on $\mathcal{O}_0$ and $\mathscr{F}$ is the Fourier transform. Then $\|F_*\|_{H^\alpha(\mathcal{O})} \lesssim \delta/\lambda$ and $\|F^* - F_0\|_{(H^\kappa(\mathcal{O}))^*} \lesssim \|F^* - F_0\|_{H^{-\kappa}(\mathcal{O})} \lesssim \delta$ in (3.12) yield (3.14). Similar comments apply to non-linear $\mathscr{G}$, with appropriate choice of $\bar{\lambda}$, see Remark 3.3.10.

Example 3.2.5 (Rates for the Radon transform)**.** Let $\mathscr{R} : \tilde{\mathcal{V}} \equiv L^2(\mathcal{O}) \to \mathbb{H}$ be the Radon transform, where $\mathcal{O} = \{x \in \mathbb{R}^2 : \|x\| < 1\}$ and $\mathbb{H} = L^2(\Sigma), \Sigma := (0, 2\pi] \times \mathbb{R}$, equipped with Lebesgue measure, see p.9 in [129] for definitions. Then $\mathscr{G} = \mathscr{R}$ satisfies (3.8) with $\kappa = 1/2, \gamma = 0$ and any $\alpha \in \mathbb{N}$ – see p.42 in [129] and note that our $\|\cdot\|_{(H^{1/2}(\mathcal{O}))^*}$-norm is the $\|\cdot\|_{H_0^{-1/2}(\mathcal{O})}$-norm used in [129] (cf. Theorem 3.30 in [121]). Applying Corollary 3.2.3 with $\alpha \geq 1$, $\mathcal{V} = H^\alpha_c(\mathcal{O})$ and $\lambda = \varepsilon^{(2\alpha+1)/(2\alpha+3)}$ implies that for any $F_0 \in H^\alpha_c(\mathcal{O})$,

$$\mathbb{E}^\varepsilon_{F_0}[\|\mathscr{R}(\hat{F}_{\lambda,\varepsilon}) - \mathscr{R}(F_0)\|^2_{L^2(\Sigma)} + \lambda^2 \|\hat{F}_{\lambda,\varepsilon}\|^2_{H^\alpha_c(\mathcal{O})}] \lesssim \varepsilon^{(4\alpha+2)/(2\alpha+3)}. \tag{3.15}$$

Using again the estimates on p.42 in [129] and that Hölder's inequality implies

$$\|g\|_{H^{1/2}(\Sigma)} \leq \|g\|^{2\alpha/(2\alpha+1)}_{L^2(\Sigma)} \|g\|^{1/(2\alpha+1)}_{H^{\alpha+1/2}(\Sigma)}$$

for $H^\alpha(\Sigma)$ defined as in [129], we deduce from (3.15) and Markov's inequality that as $\varepsilon \to 0$,

$$\|\hat{F}_{\lambda,\varepsilon} - F_0\|_{L^2(\mathcal{O})} \lesssim \|\mathscr{R}(\hat{F}) - \mathscr{R}(F_0)\|_{H^{1/2}(\Sigma)} = O_{\mathbb{P}^\varepsilon_{f_0}}(\varepsilon^{\frac{2\alpha}{2\alpha+3}})$$

in probability [recall that random variables $(Z_n : n \in \mathbb{N})$ are $O_{\Pr}(r_n)$ if $\forall \delta > 0\ \exists M = M_\delta$ s.t. $\Pr(|Z_n| > M r_n) < \delta$ for all $n \in \mathbb{N}$], with constants uniform in $\|F_0\|_{H^\alpha_c(\mathcal{O})} \leq R$ for any $R > 0$. Similarly, if one chooses $\lambda = \varepsilon$ instead, then the MAP estimate from Remark 3.2.4 satisfies

$$\left\|\hat{F}_{\lambda,\varepsilon} - F_0\right\|_{L^2(\mathcal{O})} = O_{\mathbb{P}^\varepsilon_{f_0}}(\varepsilon^{\frac{2\beta}{2\beta+3}}), \quad \text{where } \beta := \alpha - 1 > 0,$$

uniformly over $\|F_0\|_{H^\beta_c(\mathcal{O}_0)} \leq R$ for $R > 0$.

Remark 3.2.6 (The effect of nonlinearity)**.** In the proof of Theorem 3.2.2 we follow ideas for M-estimation from [170, 169], and condition (3.8) is needed to bound the entropy numbers of images $\{\mathscr{G}(F) \mid \|F\|_{H^\alpha} \leq R\}, 0 < R < \infty$, of Sobolev balls under $\mathscr{G}$, which in turn control the modulus of continuity of the Gaussian process that determines the convergence rate of $\hat{F}$ to F_0. The at most polynomial growth in $\|F\|_{H^\alpha}$ of the Lipschitz constants

$$(1 + \|F\|_{H^\alpha}^\gamma \vee \|H\|_{H^\alpha}^\gamma), \quad \gamma \geq 0, \tag{3.16}$$

in (3.8) turns out to be essential in the proof of Theorem 3.2.2. But even when only a 'polynomial nonlinearity' is present ($\gamma > 0$), the last term in the condition (3.10) can become dominant if the penalisation parameter λ is too small. The intuition is that, for non-linear problems, too little penalisation can mean that the maximisers $\hat{F}$ over unbounded parameter spaces behave erratically, yielding sub-optimal convergence rates.

3.3 Results for elliptic PDE models

In this section, we apply Theorem 3.2.2 to the inverse problems induced by the PDEs (3.2) and (3.3). We also discuss the implied convergence rates for the parameter f.

3.3.1 Basic setup and link functions

For any integer $\alpha > d/2$ and any constant $K_{min} \in [0,1)$, and denoting the outward pointing normal vector at $x \in \partial\mathcal{O}$ by $n = n(x)$, define the parameter space (boundary derivatives are understood in the trace sense)

$$\begin{aligned}\mathcal{F} := \mathcal{F}_{\alpha,K_{min}} = \big\{ f \in H^\alpha(\mathcal{O}) : f > K_{min} \text{ on } \mathcal{O},\ f = 1 \text{ on } \partial\mathcal{O}, \\ \frac{\partial^j f}{\partial n^j} = 0 \text{ on } \partial\mathcal{O} \text{ for } j = 1, ..., \alpha - 1 \big\},\end{aligned} \tag{3.17}$$

and its subclasses

$$\mathcal{F}_{\alpha,r}(R) := \big\{ f \in \mathcal{F} : f > r \text{ on } \mathcal{O},\ \|f\|_{H^\alpha} \leq R \big\},\ r \geq K_{min}, R > 0.$$

We note that the restrictions $K_{min} < 1$ and $f = 1$ on $\partial\mathcal{O}$ in (3.17) are made only for convenience, and could be replaced by any $K_{min} > 0$ and $f = \tilde{g}$ for <u>fixed</u> $\tilde{g} \in C^\infty(\partial\mathcal{O})$ satisfying $\tilde{g} > K_{min}$. Moreover, for estimation over parameter spaces without prescribed boundary values for f, see Remark 3.3.11.

We will assume that the coefficient f of the second order linear elliptic partial differential operators featuring in the boundary value problems (3.2) and (3.3), respectively, belong to $\mathcal{F}_{\alpha,K_{min}}$ for large enough α, and denote by

$$G : \mathcal{F} \to L^2(\mathcal{O}), \qquad f \mapsto G(f) := u_f, \tag{3.18}$$

the corresponding solution maps. Following (3.4) with $\mathbb{H} = L^2(\mathcal{O})$, we then observe

$$Y^{(\varepsilon)} = G(f) + \varepsilon\mathbb{W}, \ \varepsilon > 0, \tag{3.19}$$

whose law will now be denoted by $\mathbb{P}_f^\varepsilon$ for $f \in \mathcal{F}$.

We will apply Theorem 3.2.2 to a suitable bijective re-parameterisation of $\mathcal{F}$ for which the set $\mathcal{V}$ one optimises over is a linear space. This is natural for implementation purposes but also necessary to retain the Bayesian interpretation of our estimators from Remark 3.2.4. To this end, we introduce 'link functions' Φ – the lowercase and uppercase notation for corresponding functions $f \in \mathcal{F}$ and $F = \Phi^{-1} \circ f$ will be used throughout.

Definition 3.3.1. *1. A function Φ is called a link function if Φ is a smooth, strictly increasing bijective map $\Phi : \mathbb{R} \to (K_{min}, \infty)$ satisfying $\Phi(0) = 1$ and $\Phi' > 0$ on $\mathbb{R}$.*

2. A function $\Phi : (a, b) \to \mathbb{R}$, $-\infty \leq a < b \leq \infty$, is called regular if all derivatives of Φ of order $k \geq 1$ are bounded, i.e.

$$\forall k \geq 1 : \quad \sup_{x \in (a,b)} \left|\Phi^{(k)}(x)\right| < \infty. \tag{3.20}$$

In the notation of Theorem 3.2.2, throughout this section we set $\mathbb{H} = L^2(\mathcal{O})$, $\tilde{\mathcal{V}} = \mathcal{V} := \{\Phi^{-1} \circ f : f \in \mathcal{F}\}$ to be the 'pulled-back' parameter space, and

$$\mathscr{G} : \mathcal{V} \to L^2(\mathcal{O}), \quad \mathscr{G}(F) := G(\Phi \circ F), \tag{3.21}$$

For $\mathcal{F}$ as in (3.17), one easily verifies that

$$\mathcal{V} = \left\{ F \in H^\alpha : \frac{\partial^j F}{\partial n^j} = 0 \text{ on } \partial\mathcal{O} \text{ for } j = 0, ..., \alpha - 1 \right\} = H_c^\alpha(\mathcal{O}),$$

where the second equality follows from the characterization of $H_c^\alpha(\mathcal{O})$ in Theorem 11.5 of [108]. Given a realisation of (3.19) and a regular link function Φ, we define the generalised Tikhonov regularised functional $J_{\lambda,\varepsilon} : \mathcal{F} \to \mathbb{R}$,

$$J_{\lambda,\varepsilon}(f) := 2\langle Y^{(\varepsilon)}, G(f)\rangle_{L^2} - \|G(f)\|_{L^2}^2 - \lambda^2\|\Phi^{-1} \circ f\|_{H^\alpha}^2, \quad \lambda > 0. \tag{3.22}$$

Then for all $f \in \mathcal{F}$, we have $\mathscr{J}_{\lambda,\varepsilon}(F) = J_{\lambda,\varepsilon}(f)$ in the notation (3.6), and maximising $J_{\lambda,\varepsilon}$ over $\mathcal{F}$ is equivalent to maximising $\mathscr{J}_{\lambda,\varepsilon}$ over $H_c^\alpha = \mathcal{V}$. Any pair of maximisers will be denoted by

$$\hat{f} \in \arg\max_{f \in \mathcal{F}} J_{\lambda,\varepsilon}(f), \quad \hat{F} = \Phi^{-1} \circ \hat{f} \in \arg\max_{F \in H_c^\alpha} \mathscr{J}_{\lambda,\varepsilon}(F), \quad G(\hat{f}) = \mathscr{G}(\hat{F}).$$

The proofs of the theorems which follow are based on an application of Theorem 3.2.2, after verifying that the map (3.21) satisfies (3.8) with $\mathcal{V} = H_c^\alpha$ and suitable values of κ, γ, α. The verification of (3.8) is based on PDE estimates that control the modulus of continuity of the solution map (3.18), and on certain analytic properties of the link function Φ. In practice often the choice $\Phi = \exp$ is made (cf. [155]), but our results suggest that the use of a regular link function might be preferable. Indeed, the polynomial growth requirement (3.16) discussed above is not met if one chooses for Φ the exponential function. Before we proceed, let us give an example of a regular link function.

Example 3.3.2. Define the function $\phi : \mathbb{R} \to (0, \infty)$ by $\phi(x) = e^x 1_{x<0} + (1 + x)1_{x\geq 0}$, let $\psi : \mathbb{R} \to [0, \infty)$ be a smooth, compactly supported function with $\int_{\mathbb{R}} \psi = 1$, and write $\phi * \psi = \int_{\mathbb{R}} \phi(\cdot - y)\psi(y)dy$ for their convolution. It follows from elementary calculations that, for any $K_{min} \in \mathbb{R}$,

$$\Phi : \mathbb{R} \to (K_{min}, \infty),\ \Phi := K_{min} + \frac{1 - K_{min}}{\psi * \phi(0)} \psi * \phi,$$

is a regular link function with range (K_{min}, ∞).

3.3.2 Divergence form equation

For a given source function $g \in C^\infty(\mathcal{O})$, we consider the Dirichlet boundary value problem

$$\begin{cases} \nabla \cdot (f \nabla u) = g & \text{on } \mathcal{O}, \\ u = 0 & \text{on } \partial O, \end{cases} \tag{3.23}$$

where $f \in \mathcal{F}_{\alpha, K_{min}}$ (see (3.17)) for some $\alpha > d/2 + 1, K_{min} > 0$. Then (3.59) implies $f \in C^{1+\eta}(\mathcal{O})$ for some $\eta > 0$, and the Schauder theory for elliptic PDEs (Theorem 6.14 in [71]) then gives that (3.23) has a unique classical solution in $C(\bar{\mathcal{O}}) \cap C^{2+\eta}(\mathcal{O})$ which we shall denote by $G(f) = u_f$.

Upper bounds For a link function Φ and $f_1, f_2 \in \mathcal{F}$, define (cf. (3.7))

$$\mu_\lambda(f_1, f_2) := \|G(f_1) - G(f_2)\|_{L^2}^2 + \lambda^2 \|\Phi^{-1} \circ f_1\|_{H^\alpha}^2 = \tau_\lambda(F_1, F_2).$$

Theorem 3.3.3 (Prediction error)**.** *Let $\mathcal{F}$ be given by (3.17) for some integer $\alpha > (d/2 + 2) \vee (2d - 1)$ and $K_{min} \in (0, 1)$. Let $G(f) = u_f$ denote the unique solution of (3.23) and let $Y^{(\varepsilon)} \sim \mathbb{P}_{f_0}^\varepsilon$ from (3.19) for some $f_0 \in \mathcal{F}$. Moreover, suppose that $\Phi : \mathbb{R} \to (K_{min}, \infty)$ is a regular link function and that $J_{\lambda_\varepsilon, \varepsilon}$ is given by (3.22), where*

$$\lambda_\varepsilon := \varepsilon^{\frac{2(\alpha+1)}{2(\alpha+1)+d}}.$$

Then the following holds.

1. *For each* $f_0 \in \mathcal{F}$ *and* $\varepsilon > 0$, *almost surely under* $\mathbb{P}^{\varepsilon}_{f_0}$, *there exists a maximiser* $\hat{f}_\varepsilon \in \mathcal{F}$ *of* $J_{\lambda_\varepsilon,\varepsilon}$ *over* $\mathcal{F}$.

2. *For each* $R > 0$, $r > K_{min}$, *there exist finite constants* $c_1, c_2 > 0$ *such that for any maximiser* $\hat{f}_\varepsilon \in \mathcal{F}$ *of* $J_{\lambda_\varepsilon,\varepsilon}$, *all* $0 < \varepsilon < 1$ *and all* $M \geq c_1$,
$$\sup_{f_0 \in \mathcal{F}_{\alpha,r}(R)} \mathbb{P}^{\varepsilon}_{f_0}\Big(\mu^2_{\lambda_\varepsilon}(\hat{f}_\varepsilon, f_0) \geq M^2 \varepsilon^{\frac{4(\alpha+1)}{2(\alpha+1)+d}}\Big) \leq \exp\Big(-\frac{M^2\lambda_\varepsilon^2}{c_2\varepsilon^2}\Big). \tag{3.24}$$

3. *For each* $R > 0$, $r > K_{min}$ *and* $\beta \in [0, \alpha+1]$, *there exists a constant* c_3 *such that for any maximiser* $\hat{f}_\varepsilon \in \mathcal{F}$ *of* $J_{\lambda_\varepsilon,\varepsilon}$ *with corresponding* $u_{\hat{f}_\varepsilon}$, *for all* $0 < \varepsilon < 1$,
$$\sup_{f_0 \in \mathcal{F}_{\alpha,r}(R)} \mathbb{E}^{\varepsilon}_{f_0}\Big\|u_{\hat{f}_\varepsilon} - u_{f_0}\Big\|_{H^\beta} \leq c_3 \varepsilon^{\frac{2(\alpha+1-\beta)}{2(\alpha+1)+d}}. \tag{3.25}$$

Lower bounds We now give a minimax lower bound on the rate of estimation for u_f which matches the bound in (3.25). To facilitate the exposition we only consider the unit ball $\mathcal{O} = D := \{x \in \mathbb{R}^d : \|x\| < 1\}$, set $g = 1$ identically on $\mathcal{O}$, and fix H^β-loss with $\beta = 2$.

Theorem 3.3.4. *For* $K_{min} \in (0,1), \alpha > d/2+1$, $\mathcal{O} = D$ *and* $g = 1$ *on* $\mathcal{O}$, *consider solutions* $u_f, f \in \mathcal{F}$, *to (3.23). Then there exists* $C < \infty$ *such that for all* $\varepsilon > 0$ *small enough,*
$$\inf_{\hat{u}_\varepsilon} \sup_{f_0 \in \mathcal{F}_{\alpha,r}(R)} \mathbb{E}^{\varepsilon}_{f_0}\|\hat{u}_\varepsilon - u_{f_0}\|_{H^2} \geq C \varepsilon^{\frac{2(\alpha-1)}{2(\alpha+1)+d}}, \ r > K_{min}, R > 0, \tag{3.26}$$
where the infimum ranges over all measurable functions $\hat{u}_\varepsilon = \hat{u}(Y^{(\varepsilon)})$ *of* $Y^{(\varepsilon)}$ *from (3.19) that take values in* H^2.

Observe that (3.26) coincides with the lower bound for estimating u_{f_0} as a regression function without PDE-constraint in $H^{\alpha+1}$ under H^2-loss. Note however that unconstrained 'off the shelf' regression function estimators $\tilde{u}_\varepsilon$ for u_f will not satisfy the non-linear PDE constraint $\tilde{u} = G(\tilde{f})$ for some $\tilde{f} \in \mathcal{F}$, thus providing no recovery of the PDE coefficient f_0 itself.

Rates for f via stability estimates For estimators $u_{\hat{f}_\varepsilon}$ that lie in the range of the forward map G, we can resort to 'stability estimates' which allow to control the convergence rate of $\hat{f}_\varepsilon$ to f_0 by the rate of $G(\hat{f}_\varepsilon) = u_{\hat{f}_\varepsilon}$ towards $G(f_0) = u_{f_0}$, in appropriate norms. Injectivity and global stability estimates for this problem have been studied in several papers since Richter [149], see the recent contribution [24] and the discussion therein. They require additional assumptions, a very common choice being that $g > 0$ throughout $\bar{\mathcal{O}}$. The usefulness of these estimates depends in possibly subtle ways on the class of f's one constrains the problem to.

The original stability estimate given in [149] controls $\|f_1 - f_2\|_\infty$ in terms of $\|u_{f_1} - u_{f_2}\|_{C^2}$ which does not combine well with the H^β- convergence rates obtained in Theorem 3.3.3. The results proved in [24] are designed for 'low regularity' cases where $\alpha \in (0, 1)$: they give at best

$$\|f_1 - f_2\|_{L^2} \leq C(f_1, f_2)\|u_{f_1} - u_{f_2}\|_{H^1}^{1/2}, \qquad f_1, f_2 \in \mathcal{F},\ d \geq 2, \tag{3.27}$$

which via Theorem 3.3.3 would imply a convergence rate of $\varepsilon^{\frac{\alpha}{2(\alpha+1)+d}}$ for $\|\hat{f}_\varepsilon - f_0\|_{L^2}$. For higher regularity $\alpha \geq 2$ relevant here, this can be improved. We prove in Lemma 3.5.5 below a Lipschitz stability estimate for the map $u_f \mapsto f$ between the spaces H^2 and L^2, and combined with Theorem 3.3.3 this gives the following rate bound for $\hat{f}_\varepsilon - f_0$.

Theorem 3.3.5. *Suppose that α, K_{min}, $\mathcal{F}$, G, Φ, λ_ε are as in Theorem 3.3.3 and that in addition,* $\inf_{x \in \mathcal{O}} g(x) \geq g_{min}$ *for some $g_{min} > 0$. Let $\hat{f}_\varepsilon \in \mathcal{F}$ be any maximiser of $J_{\lambda_\varepsilon, \varepsilon}$. Then, for each $r > K_{min}$ and $R < \infty$, there exists a constant $C > 0$ such that we have for all $0 < \varepsilon < 1$,*

$$\sup_{f_0 \in \mathcal{F}_{\alpha,r}(R)} \mathbb{E}^\varepsilon_{f_0} \|\hat{f}_\varepsilon - f_0\|_{L^2} \leq C\varepsilon^{\frac{2(\alpha-1)}{2(\alpha+1)+d}}. \tag{3.28}$$

The rate in Theorem 3.3.5 is strictly better than what can be obtained from (3.27), or by estimating $\|u_{\hat{f}_\varepsilon} - u_{f_0}\|_{C^2}$ by $\|u_{\hat{f}_\varepsilon} - u_{f_0}\|_{H^{2+d/2+\eta}}, \eta > 0$, and using Richter's stability estimate. A more detailed study of the stability problem, and of the related question of optimal rates for estimating f, is beyond the scope of the present thesis and will be pursued elsewhere.

3.3.3 Schrödinger equation

We now turn to the Schrödinger equation

$$\begin{cases} \Delta u - 2fu = 0 \quad \text{on } \mathcal{O}, \\ u = g \quad \text{on } \partial O, \end{cases} \tag{3.29}$$

with given $g \in C^\infty(\partial\mathcal{O})$. By standard results for elliptic PDEs (Theorem 6.14 in [71]), for $f \in \mathcal{F}_{\alpha, K_{min}}$ from (3.17) with $K_{min} \geq 0, \alpha > d/2$, a unique classical solution $u_f = G(f)$ to (3.3) exists which lies in $C^{2+\eta}(\mathcal{O}) \cap C^0(\bar{\mathcal{O}})$ for some $\eta > 0$.

The results for this PDE are similar to the previous section, although the convergence rates are quantitatively different due to the fact that the forward operator is now 2-smoothing.

Theorem 3.3.6 (Prediction error)**.** *Let $\mathcal{F}$ be given by (3.17) for some integer $\alpha > (d/2 + 2) \vee (2d - 2)$ and $K_{min} \in [0, 1)$. Let $G(f) = u_f$ denote the unique solution to (3.29) and let $Y^{(\varepsilon)} \sim \mathbb{P}^\varepsilon_{f_0}$ from (3.19) for some $f_0 \in \mathcal{F}$. Moreover, suppose that $\Phi : \mathbb{R} \to (K_{min}, \infty)$ is a regular link function and that $J_{\lambda_\varepsilon, \varepsilon}$ is given by (3.22), where*

$$\lambda_\varepsilon = \varepsilon^{\frac{2(\alpha+2)}{2(\alpha+2)+d}}.$$

Then the following holds.

1. *For each $f_0 \in \mathcal{F}$ and $\varepsilon > 0$, almost surely under $\mathbb{P}^\varepsilon_{f_0}$, there exists a maximiser $\hat{f}_\varepsilon \in \mathcal{F}$ of $J_{\lambda_\varepsilon,\varepsilon}$ over $\mathcal{F}$.*

2. *For each $R > 0$, $r > K_{min}$, there exist finite constants $c_1, c_2 > 0$ such that for any maximiser $\hat{f}_\varepsilon \in \mathcal{F}$ of $J_{\lambda_\varepsilon,\varepsilon}$, all $0 < \varepsilon < 1$ and all $M \geq c_1$, we have*

$$\sup_{f_0 \in \mathcal{F}_{\alpha,r}(R)} \mathbb{P}^\varepsilon_{f_0}\Big(\mu^2_{\lambda_\varepsilon}(\hat{f}_\varepsilon, f_0) \geq M^2 \varepsilon^{\frac{4(\alpha+2)}{2(\alpha+2)+d}}\Big) \leq \exp\Big(-\frac{M^2\lambda^2_\varepsilon}{c_2\varepsilon^2}\Big). \tag{3.30}$$

3. *For each $R > 0$, $r > K_{min}$ and $\beta \in [0, \alpha+2]$, there exists a constant $c_3 > 0$ such that for any maximiser $\hat{f}_\varepsilon \in \mathcal{F}$ of $J_{\lambda_\varepsilon,\varepsilon}$ and all $0 < \varepsilon < 1$,*

$$\sup_{f_0 \in \mathcal{F}_{\alpha,r}(R)} \mathbb{E}^\varepsilon_{f_0} \left\| u_{\hat{f}_\varepsilon} - u_{f_0} \right\|_{H^\beta} \leq C \varepsilon^{\frac{2(\alpha+2-\beta)}{2(\alpha+2)+d}}. \tag{3.31}$$

For the PDE (3.29) the stability estimate is easier to obtain than the one required in Theorem 3.3.5, and here is the convergence rate for estimation of $f \in \mathcal{F}$. We note that the rates obtained in Theorems 3.3.6 and 3.3.7 are minimax-optimal in view of Proposition 2 in [131] (and its proof).

Theorem 3.3.7. *Assume that $\alpha, K_{min}, \mathcal{F}, G, \Phi, \lambda_\varepsilon$ are as in Theorem 3.3.6 and that in addition, $\inf_{x\in\partial O} g(x) \geq g_{min}$ for some $g_{\min} > 0$. Let $\hat{f}_\varepsilon \in \mathcal{F}$ be any maximiser of $J_{\lambda_\varepsilon,\varepsilon}$. Then for all $r > K_{min}$ and $R > 0$, there exists a constant $C > 0$ such that for all $\varepsilon > 0$ small enough,*

$$\sup_{f_0 \in \mathcal{F}_{\alpha,r}(R)} \mathbb{E}^\varepsilon_{f_0} \left\| \hat{f}_\varepsilon - f_0 \right\|_{L^2} \leq C \varepsilon^{\frac{2\alpha}{2(\alpha+2)+d}}.$$

3.3.4 Concluding remarks and discussion

Remark 3.3.8. The classical literature on 'deterministic inverse problems' deals with convergence rate questions of Tikhonov and related regularisers, see the monograph [63], [64, 130, 154, 63, 160, 95] and also [17]. The convergence analysis conducted there is typically for observations $y_\delta = \mathscr{G}(F) + \delta$ where δ is a fixed perturbation vector in data space, equal to $L^2(\mathcal{O})$ in the present setting. For non-linear problems, rates are obtained as $\|\delta\| \to 0$ under some invertibility assumptions on a suitable adjoint $D\mathscr{G}^*_F$ of the linearisation $D\mathscr{G}_F[\cdot]$ of the forward operator at the 'true' parameter F ('source conditions'), see, e.g., Section 10 in [63]. These results are not directly comparable since our noise $\mathbb{W}$ models genuine statistical error and hence is random and, in particular, almost surely *not* an element of data space $L^2(\mathcal{O})$. As shown in [20, 21, 87, 112], the 'deterministic' analysis extends to the Gaussian regression model (3.4) to a certain degree, but the results obtained there still rely, among other things, on invertibility properties of $D\mathscr{G}^*_F$. For the PDE (3.2) such 'source conditions' can be problematic as $D\mathscr{G}_F^*$ may not be invertible in general (due to the fact ∇u_f can vanish on $\mathcal{O}$ unless some further assumptions are made, see, e.g., [102]). Our techniques circumvent source conditions by first optimally solving the 'forward problem', and then feeding this solution into a suitable stability estimate for $\mathscr{G}^{-1}$.

Remark 3.3.9 (Bayesian inversion)**.** The Bayesian approach [155, 99, 53] to inverse problems has been very popular recently, but only few theoretical guarantees for such algorithms are available in non-linear settings: In [177], convergence rates for the PDE (3.23) are obtained for certain Bayes procedures that arise from priors for f that concentrate on specific bounded subsets of H^α. The main idea to combine regression results with stability estimates is related to our approach, but the rates obtained in [177] are suboptimal, and for the elliptic PDE models considered here do not apply to Gaussian priors. Bayesian inference for the PDE (3.29) has been studied in [131], where it is shown that procedures based on a uniform wavelet prior *do* attain minimax optimal convergence rates for f and u_f (up to log-factors). The paper [131] also addresses the question of uncertainty quantification via the posterior distribution, by proving nonparametric Bernstein-von Mises theorems, whereas our results only concern the convergence rate of the MAP estimate for certain Gaussian priors (see Remarks 3.2.4, 3.3.10). A related recent reference is [125] where the asymptotics for linear functionals of Gaussian MAP estimates are obtained in linear inverse problems involving Radon and more general X-ray transforms – see also [99, 144] for earlier results for diagonalisable linear inverse problems. Finally, convergence rates for posterior distributions of PDE coefficients in certain non-linear parabolic (diffusion) settings have been studied in [134, 1].

Remark 3.3.10 (MAP estimates, non-linear $\mathscr{G}$)**.** As explained in Remark 3.2.6, Theorem 3.2.2 does not necessarily produce optimal rates for the choice $\lambda = \varepsilon$ in the non-linear settings from this section where $\gamma > 0$. For MAP estimates as discussed in Remark 3.2.4 our results then imply optimal convergence rates for $G(f)$ only if the Gaussian prior is re-scaled in an ε-dependent way, more specifically if its RKHS norm is $\bar{\lambda}\|\cdot\|_{H^\alpha}$ with $\bar{\lambda} = \varepsilon^{-d/(2\alpha+2\kappa+d)}$. Moreover, positivity of f is enforced by a 'regular link function' Φ, excluding the exponential map. Whether these restrictions on admissible priors are artefacts of our proofs remains a challenging open question, however, to the best of our knowledge, these are the first convergence rate results for proper MAP-estimators in the non-linear PDE constrained inverse problems studied here, improving in particular upon the ('sub-sequential') consistency results [50].

Remark 3.3.11. For both PDEs (3.23) and (3.29), one can also consider estimation over the parameter space

$$\tilde{\mathcal{F}} := \{f \in H^\alpha(\mathcal{O}) : \inf_{x\in\mathcal{O}} f(x) > K_{min} \text{ on } \mathcal{O}\}, \tag{3.32}$$

without the boundary restrictions on f from (3.17). Note that $\mathcal{F} \subset \tilde{\mathcal{F}}$. Then, with $\kappa = 1/2 - \eta, \eta \in (0, 1/2)$, $\tilde{\mathcal{V}} = \mathcal{V} = H^\alpha(\mathcal{O})$ in (3.21), $\alpha > (d/2+2) \vee 2d - \kappa$ and K_{min} as before, Theorem 3.2.2 applies as in Theorems 3.3.3 and 3.3.6, and

$$\sup_{f_0 \in \tilde{\mathcal{F}}_{\alpha,r}(R)} \mathbb{E}^\varepsilon_{f_0} \|u_{\hat{f}} - u_{f_0}\|_{H^\beta(\mathcal{O})} \lesssim \varepsilon^{\frac{2(\alpha+1/2-\eta-\beta)}{2(\alpha+1/2-\eta)+d}}, \quad r > K_{min}, R > 0,$$

where, respectively, $\beta \in [0, \alpha+1]$ (divergence form eq.) and $\beta \in [0, \alpha+2]$ (Schrödinger eq.), and $\tilde{\mathcal{F}}_{\alpha,r}(R) := \{f \in \tilde{\mathcal{F}} : f > r \text{ on } \mathcal{O},\ \|f\|_{H^\alpha} \le R\}$. By the stability Lemmas 3.5.5 and 3.5.9 (which apply to $\tilde{\mathcal{F}}$ as well) and arguing as in the proofs of Theorems 3.3.5 and 3.3.7, this yields the respective convergence rates

$$\varepsilon^{\frac{2(\alpha-3/2-\eta)}{2(\alpha+1/2-\eta)+d} \cdot \frac{\alpha-1}{\alpha+1}} \text{ (div. form eq.)}, \quad \varepsilon^{\frac{2(\alpha-3/2-\eta)}{2(\alpha+1/2-\eta)+d} \cdot \frac{\alpha}{\alpha+2}} \text{ (Schrödinger eq.)},$$

for $\mathbb{E}^\varepsilon_{f_0} \|\hat{f} - f_0\|_{L^2}$, uniform over $\tilde{\mathcal{F}}_{\alpha,r}(R)$.

3.4 Proofs of the main results

3.4.1 Convergence rates in M-estimation

For the convenience of the reader we recall here some classical techniques for proving convergence rates for M-estimators (see [169, 170]) – these will form the basis for the proof of Theorem 3.2.2. In the following $\tilde{\mathcal{V}} \subseteq L^2(\mathcal{O})$, $\mathscr{G} : \tilde{\mathcal{V}} \to \mathbb{H}$ is a Borel-measurable map, and the functionals $\mathscr{J}_{\lambda,\varepsilon}, \tau^2_\lambda(\cdot,\cdot)$ are given by (3.6), (3.7), respectively. Let $\mathcal{V} \subseteq \tilde{\mathcal{V}} \cap H^\alpha(\mathcal{O})$ be a subset over which we aim to maximise $\mathscr{J}_{\lambda,\varepsilon}$. For any $F_* \in \mathcal{V}$ and $\lambda, R \ge 0$, define sets

$$\mathcal{V}_*(\lambda, R) := \{F \in \mathcal{V} : \tau^2_\lambda(F, F_*) \le R^2\}, \tag{3.33}$$

their images under $\mathscr{G}$,

$$\mathcal{D}_*(\lambda, R) = \{\mathscr{G}(F) : F \in \mathcal{V} \text{ with } \tau^2_\lambda(F, F_*) \le R^2\}, \tag{3.34}$$

and also

$$J_*(\lambda, R) := R + \int_0^{2R} H^{1/2}(\rho, \mathcal{D}_*(\lambda, R), \|\cdot\|_{\mathbb{H}})\, d\rho, \tag{3.35}$$

where the usual metric entropy of $A \subset \mathbb{H}$ is denoted by $H(\rho, A, \|\cdot\|_{\mathbb{H}})$ $(\rho > 0)$. The following theorem is, up to some modifications which adapt it to the continuum sampling scheme (3.4) and the inverse problem setting considered here, a version of Theorem 2.1 in [169].

Theorem 3.4.1. *Let $F_* \in \mathcal{V}, \lambda > 0$, and let $\mathbb{P}^\varepsilon_{F_0}$ be the law of $Y^{(\varepsilon)}$ from (3.4) for some fixed $F_0 \in \tilde{\mathcal{V}}$. Suppose $\Psi_*(\lambda, R) \ge J_*(\lambda, R)$ is some upper bound such that $R \mapsto \Psi_*(\lambda, R)/R^2$ is non-increasing. Then there exist universal constants c_1, c_2, c_3 such that for all $\varepsilon, \lambda, \delta > 0$*

satisfying

$$\delta^2 \geq c_1 \varepsilon \Psi_*(\lambda, \delta) \tag{3.36}$$

and any $R \geq \delta$, *we have that*

$$\begin{aligned}\mathbb{P}^{\varepsilon}_{F_0}\Big(&\mathscr{J}_{\lambda,\varepsilon} \text{ has a maximizer } \hat{F} \text{ over } \mathcal{V} \text{ s.t. } \tau_\lambda^2(\hat{F}, F_0) \geq 2(\tau_\lambda^2(F_*, F_0) + R^2)\Big)\\ &\leq c_2 \exp\Big(-\frac{R^2}{c_1^2 \varepsilon^2}\Big).\end{aligned} \tag{3.37}$$

Moreover, for any maximiser $\hat{F}$ *of* $\mathscr{J}_{\lambda,\varepsilon}$ *over* $\mathcal{V}$ *we have for some universal constant* c_3

$$\mathbb{E}^{\varepsilon}_{F_0}[\tau_\lambda^2(\hat{F}, F_0)] \leq c_3(\tau_\lambda^2(F_*, F_0) + \delta^2 + \varepsilon^2). \tag{3.38}$$

Proof. 1. Let $\hat{F}$ denote any maximiser of $\mathscr{J}_{\lambda,\varepsilon}$. By completing the square, we see that $\hat{F}$ also maximises

$$Q_{\lambda,\varepsilon}(F) := 2\langle \varepsilon \mathbb{W}, \mathscr{G}(F)\rangle_{\mathbb{H}} - \|\mathscr{G}(F) - \mathscr{G}(F_0)\|^2_{\mathbb{H}} - \lambda^2 \|F\|^2_{H^\alpha}.$$

Rewriting the inequality $Q_{\lambda,\varepsilon}(\hat{F}) \geq Q_{\lambda,\varepsilon}(F_*)$, we obtain

$$2\langle \varepsilon \mathbb{W}, \mathscr{G}(\hat{F}) - \mathscr{G}(F_*)\rangle_{\mathbb{H}} \geq \tau_\lambda^2(\hat{F}, F_0) - \tau_\lambda^2(F_*, F_0).$$

Elementary calculations as in [169], p.3-4, give that for all $R > 0$, if

$$\tau_\lambda^2(\hat{F}, F_0) \geq 2\left(\tau_\lambda^2(F_*, F_0) + R^2\right)$$

holds then we also have the inequalities

$$\begin{aligned}\tau_\lambda^2(\hat{F}, F_*) &\geq R^2 \qquad \text{and}\\ \tau_\lambda^2(\hat{F}, F_0) - \tau_\lambda^2(F_*, F_0) &\geq \frac{1}{6}\tau_\lambda^2(\hat{F}, F_*).\end{aligned}$$

It follows that for any $R > 0$ and for $\mathbb{P}$ the law of the centred Gaussian process $(\mathbb{W}(\psi) = \langle \mathbb{W}, \psi\rangle_{\mathbb{H}} : \psi \in \mathbb{H})$,

$$\begin{aligned}&\mathbb{P}^{\varepsilon}_{F_0}\left(\tau_\lambda^2(\hat{F}, F_*) \geq 2\left(\tau_\lambda^2(F_*, F_0) + R^2\right)\right)\\ &\leq \mathbb{P}^{\varepsilon}_{F_0}\Big(\tau_\lambda^2(\hat{F}, F_*) \geq R^2,\ 2\langle \varepsilon \mathbb{W}, \mathscr{G}(\hat{F}) - \mathscr{G}(F_*)\rangle_{\mathbb{H}} \geq \frac{1}{6}\tau_\lambda^2(\hat{F}, F_*)\Big)\\ &\leq \sum_{l=1}^{\infty} \mathbb{P}\left(\sup_{\psi \in \mathcal{D}_*(\lambda, 2^l R)} \langle \varepsilon \mathbb{W}, \psi - \mathscr{G}(F_*)\rangle_{\mathbb{H}} \geq \frac{1}{48} 2^{2l} R^2\right) =: \sum_{l=1}^{\infty} P_l.\end{aligned}$$

2. For all $\lambda, R \geq 0$, we have that $\sup_{\psi,\varphi\in\mathcal{D}_*(\lambda,R)} \|\psi - \varphi\|_{\mathbb{H}} \leq 2R$, so that by Dudley's theorem (see [72], p.43),

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\psi\in\mathcal{D}_*(\lambda,R)} |\langle \mathbb{W}, \psi - \mathscr{G}(F_*)\rangle_{\mathbb{H}}|\right] \\
&\quad\lesssim \inf_{\psi\in\mathcal{D}_*(\lambda,R)} \mathbb{E}|\langle \mathbb{W}, \psi - \mathscr{G}(F_*)\rangle_{\mathbb{H}}| + \int_0^{2R} H^{1/2}\left(\rho, \mathcal{D}_*(\lambda,R), \|\cdot\|_{\mathbb{H}}\right) d\rho \\
&\quad\lesssim R + \int_0^{2R} H^{1/2}\left(\rho, \mathcal{D}_*(\lambda,R), \|\cdot\|_{\mathbb{H}}\right) d\rho = J_*(\lambda,R) \leq \Psi_*(\lambda,R).
\end{aligned}$$

3. Let us write $S_*(\lambda,R) := \sup_{\psi\in\mathcal{D}_*(\lambda,R)} |\langle \mathbb{W}, \psi - \mathscr{G}(F_*)\rangle|$. By choosing c large enough and δ such that (3.36) holds, we have that for all $R \geq \delta$, $\frac{1}{48}R^2 - \varepsilon\Psi_*(\lambda,R) \geq \frac{1}{96}R^2$. Thus by the preceding display, the Borell-Sudakov-Tsirelson inequality (see Theorem 2.5.8 in [72]), and possibly making $c > 0$ larger, we obtain for all $R \geq \delta$ and $l = 1, 2, ...$

$$\begin{aligned}
P_l &\leq \mathbb{P}\left(\varepsilon S_*(\lambda, 2^l R) - \varepsilon\mathbb{E}[S_*(\lambda, 2^l R)] \geq \frac{1}{48} 2^{2l} R^2 - \varepsilon\Psi_*(\lambda, 2^l R)\right) \\
&\leq \mathbb{P}\left(S_*(\lambda, 2^l R) - \mathbb{E}[S_*(\lambda, 2^l R)] \geq \frac{2^{2l}R^2}{96\varepsilon}\right) \\
&\leq \exp\left(-\frac{1}{2}\left(\frac{2^{2l}R^2}{96\varepsilon}\right)^2 2^{-2l}R^{-2}\right) \leq \exp\left(-\frac{2^{2l}R^2}{c\varepsilon^2}\right),
\end{aligned} \tag{3.39}$$

where in the penultimate inequality, we have used

$$\sup_{\psi\in\mathcal{D}_*(\lambda,2^l R)} \mathbb{E}[|\langle \mathbb{W}, \psi - \mathscr{G}(F_*)\rangle_{\mathbb{H}}|^2] \leq 2^{2l}R^2.$$

The inequality (3.37) now follows from summing (3.39), and (3.38) follows from arguing as in the proof of Lemma 2.2 in [169]. □

3.4.2 Proof of 3.2.2, Part 2

We will apply Theorem 3.4.1 and need the following lemma. For $F_* \in \mathcal{V}$, define $\mathcal{V}_*(\lambda,R), \mathcal{D}_*(\lambda,R)$ and $J_*(\lambda,R)$ by (3.33), (3.34) and (3.35) respectively. We also use the notation $H^\alpha(\mathcal{O}, r) := \{F \in H^\alpha(\mathcal{O}) \mid \|F\|_{H^\alpha} \leq r\}$ and $H^\alpha_c(\mathcal{O}, r) := \{F \in H^\alpha_c(\mathcal{O}) \mid \|F\|_{H^\alpha} \leq r\}, r > 0$ and recall $s = (\alpha + \kappa)/d$.

Lemma 3.4.2. *Suppose that $\mathcal{V}$ and $\mathscr{G}$ are as in Part 2 of Theorem 3.2.2. Then there exists a positive constant c such that for all $\lambda, R > 0$ and $F_* \in \mathcal{V}$,*

$$\Psi_*(\lambda,R) := R + c\big(R\lambda^{-\frac{1}{2s}}\left(1 + (R/\lambda)^{\gamma/2s}\right)\big)$$

is an upper bound for $J_*(\lambda,R)$.

Proof. Let us first assume that $\kappa \geq 1/2$. We estimate the metric entropy in $J_*(\lambda, R)$. Let $\rho, \lambda, R > 0$ and define

$$m := C\left(1 + R^\gamma \lambda^{-\gamma}\right),$$

where C is the constant from (3.8). By definition of τ_λ, we have $\mathcal{V}_*(\lambda, R) \subseteq H_c^\alpha(\mathcal{O}, R/\lambda)$. Fix some larger, bounded C^∞-domain $\tilde{\mathcal{O}} \supset \bar{\mathcal{O}}$ and some function $\zeta \in C_c^\infty(\mathbb{R}^d)$ such that $0 \leq \zeta \leq 1$, $\zeta = 1$ on $\mathcal{O}$ and $\mathrm{supp}(\zeta) \subset \tilde{\mathcal{O}}$. By the main theorem of Section 4.2.2 in [165], there exists a bounded, linear extension operator $\mathcal{E} : H^\kappa(\mathcal{O}) \to H^\kappa(\mathbb{R}^d)$. Define the map $e : \phi \mapsto \zeta\mathcal{E}(\phi)$ which maps $H^\kappa(\mathcal{O})$ continuously into $\tilde{H}^\kappa(\tilde{\mathcal{O}})$, and for $\phi \in L^2(\mathcal{O})$, let $\tilde{\phi} : \mathbb{R}^d \to \mathbb{R}$ denote its extension by 0 on $\mathbb{R}^d \setminus \mathcal{O}$. We then have, for some $c_1 > 0$,

$$\|\phi\|_{(H^\kappa(\mathcal{O}))^*} = \sup_{\varphi \in H^\kappa(\mathcal{O},1)} \left| \int_{\mathcal{O}} \phi\varphi \right| = \sup_{\varphi \in H^\kappa(\mathcal{O},1)} \left| \int_{\tilde{\mathcal{O}}} \tilde{\phi} e(\varphi) \right| \leq c_1 \|\tilde{\phi}\|_{H^{-\kappa}(\tilde{\mathcal{O}})}. \tag{3.40}$$

By Theorem 11.4 in [108] and its proof, the zero extension $\phi \mapsto \tilde{\phi}$ is continuous from $H_c^\alpha(\mathcal{O})$ to $H^\alpha(\tilde{\mathcal{O}})$ with norm 1, so that

$$\mathcal{W} := \left\{\tilde{F} : F \in \mathcal{V}_*(\lambda, R)\right\} \subseteq H_c^\alpha(\tilde{\mathcal{O}}, R/\lambda).$$

By Theorem 4.10.3 of [165], we can pick $\tilde{F}_1, ..., \tilde{F}_N \in \mathcal{W}$ with

$$N \leq \exp\left(c_2 \left(\frac{Rmc_1}{\lambda\rho}\right)^{\frac{1}{s}}\right)$$

for some universal constant c_2, such that the balls

$$\tilde{B}_i := \left\{\psi \in \mathcal{W} \;:\; \|\psi - \tilde{F}_i\|_{H^{-\kappa}(\tilde{\mathcal{O}})} \leq \frac{\rho}{mc_1}\right\}, \qquad i = 1, ..., N,$$

form a covering of $\mathcal{W}$. Then it follows from (3.8) and (3.40) that for all $i = 1, ..., N$ and F with $\tilde{F} \in \tilde{B}_i$,

$$\|\mathscr{G}(F) - \mathscr{G}(F_i)\|_{\mathbb{H}} \leq m\|F - F_i\|_{(H^\kappa(\mathcal{O}))^*} \leq mc_1\|\tilde{F} - \tilde{F}_i\|_{H^{-\kappa}(\tilde{\mathcal{O}})},$$

whence the balls

$$B_i' := \{\psi \in \mathcal{D}_*(\lambda, R) : \|\psi - \mathscr{G}(F_i)\|_{\mathbb{H}} \leq \rho\}, \quad i = 1, ..., N$$

form a covering of $\mathcal{D}_*(\lambda, R)$. Hence we obtain the bound

$$H\left(\rho, \mathcal{D}_*(\lambda, R), \|\cdot\|_{\mathbb{H}}\right) \lesssim \left(\frac{Rm}{\lambda\rho}\right)^{\frac{1}{s}}, \tag{3.41}$$

and hence also

$$\int_0^{2R} H^{1/2}\left(\rho, \mathcal{D}_*(\lambda, R), \|\cdot\|_{\mathbb{H}}\right) d\rho \lesssim \int_0^{2R} \left(\frac{Rm}{\lambda\rho}\right)^{\frac{1}{2s}} d\rho \lesssim R\lambda^{-\frac{1}{2s}}(1 + (R/\lambda)^{\frac{\gamma}{2s}}),$$

which proves that $\Psi_* \geq J_*$ for the case $\kappa \geq 1/2$.

For $\kappa < 1/2$, by Theorem 11.1 in [108], we have $\tilde{H}^\kappa(\mathcal{O}) = H^\kappa_c(\mathcal{O}) = H^\kappa(\mathcal{O})$ and hence $\|\cdot\|_{(H^\kappa(\mathcal{O}))^*} = \|\cdot\|_{H^{-\kappa}(\mathcal{O})}$, whence we can use Theorem 4.10.3 of [165] directly to cover $\mathcal{V}_*(R, \lambda) \subseteq H^\alpha(\mathcal{O}, R/\lambda)$ by $H^{-\kappa}(\mathcal{O})$-balls, and using (3.8) as above yields the entropy bound (3.41). $\square$

By assumption on α we have $1 + \frac{\gamma}{2s} < 2$ and hence the map $R \mapsto \Psi_*(\lambda, R)/R^2$, for $\Psi_*(\lambda, R)$ as defined in Lemma 3.4.2, is decreasing. The bounds (3.11) and (3.12) then follow from Theorem 3.4.1. The proof of existence of maximisers is given in Section 3.7. Finally, we obtain Corollary 3.2.3 by taking $F_* = F_0$ and $\delta := c\varepsilon^{2(\alpha+\kappa)/(2\alpha+2\kappa+d)}$ for which (3.10) is easily verified for λ chosen as in the corollary, so that Theorem 3.2.2 applies.

3.4.3 Proof of Theorems 3.3.3, 3.3.4 and 3.3.5

Proof of Theorem 3.3.3. We verify that $\mathscr{G}$ given by (3.21) with G the solution map of (3.23), satisfies (3.8) for $\mathcal{V} = H^\alpha_c, \mathbb{H} = L^2(\mathcal{O})$, $\gamma = 4, \kappa = 1$, in order to apply Theorem 3.2.2. Let $F, H \in H^\alpha$, and let us write $f := \Phi \circ F$, $h := \Phi \circ H$. With L_f, V_f introduced in Section 3.5.2 we have by (3.21) and (3.23)

$$\begin{aligned} L_f[\mathscr{G}(F) - \mathscr{G}(H)] &= L_f[u_f - u_h] \\ &= L_f[u_f] - L_h[u_h] + (L_h - L_f)[u_h] = \nabla \cdot ((h - f)\nabla u_h), \end{aligned} \tag{3.42}$$

and then, by Lemma 3.5.2 with H^2_0 defined in (3.62), the estimate

$$\begin{aligned} \|\mathscr{G}(F) - \mathscr{G}(H)\|_{L^2} &= \|V_f\left[\nabla \cdot ((h - f)\nabla u_h)\right]\|_{L^2} \\ &\leq C(1 + \|f\|_{C^1})\, \|\nabla \cdot ((h - f)\nabla u_h)\|_{(H^2_0)^*}\,. \end{aligned} \tag{3.43}$$

By applying the divergence theorem to the vector field $\varphi(h - f)\nabla u_h$, where $\varphi \in C^2_0$ is any C^2-function that vanishes at the boundary, we have

$$\begin{aligned} \|\nabla \cdot ((h - f)\nabla u_h)\|_{(H^2_0)^*} &= \sup_{\varphi \in C^2_0,\, \|\varphi\|_{H^2} \leq 1} \left|\int_{\mathcal{O}} \varphi \nabla \cdot ((h - f)\nabla u_h)\right| \\ &= \sup_{\varphi \in C^2_0,\, \|\varphi\|_{H^2} \leq 1} \left|\int_{\mathcal{O}} (h - f)\nabla\varphi \cdot \nabla u_h\right| \\ &\leq \|h - f\|_{(H^1)^*} \sup_{\varphi \in C^2_0,\, \|\varphi\|_{H^2} \leq 1} \|\nabla\varphi \cdot \nabla u_h\|_{H^1} \\ &\lesssim \|h - f\|_{(H^1)^*} \|u_h\|_{\mathcal{C}^2}, \end{aligned}$$

where we used the multiplicative inequality (3.57) in the last step. Combining this with (3.43) and Lemma 3.5.3 yields that

$$\|\mathscr{G}(F) - \mathscr{G}(H)\|_{L^2} \lesssim (1 + \|f\|_{C^1})(1 + \|h\|_{C^1}^2)\|h - f\|_{(H^1)^*}.$$

Hence, by (3.94), (3.96) and the Sobolev embedding (3.59), we obtain

$$\|\mathscr{G}(F) - \mathscr{G}(H)\|_{L^2} \lesssim (1 + \|F\|_{H^\alpha}^4 \vee \|H\|_{H^\alpha}^4)\|F - H\|_{(H^1)^*},$$

so $\mathscr{G}$ indeed fulfills (3.8) for $\gamma = 4$ and $\kappa = 1$.

The existence of maximisers $\hat{f}_\varepsilon$ now follows from the first part of Theorem 3.2.2, and we prove (3.24) by applying Theorem 3.2.2 with $F_* = F_0$. First, we note that for all $\hat{f}_\varepsilon$ and f_0,

$$\mu_\lambda(\hat{f}_\varepsilon, f_0) = \tau_\lambda(\hat{F}_\varepsilon, F_0). \tag{3.44}$$

For the choice $\delta_\varepsilon = c\varepsilon^{\frac{2(\alpha+1)}{2(\alpha+1)+d}}$ and c large enough, the triple $(\varepsilon, \lambda_\varepsilon, \delta_\varepsilon)$ satisfies (3.10) and Theorem 3.2.2 and (3.44) yield that for some $c' > 0$ and any $m \geq \delta_\varepsilon$,

$$\mathbb{P}_{f_0}^\varepsilon\left(\mu_{\lambda_\varepsilon}^2(\hat{f}_\varepsilon, f_0) \geq 2(\delta_\varepsilon^2 + m^2)\right) \leq \exp\left(-\frac{m^2}{c'\varepsilon^2}\right),$$

which proves (3.24).

To show (3.25), let now $\beta \in [0, \alpha+1]$, $R > 0$ and $r > K_{min}$. By Lemma 3.5.4, we have that

$$M := \sup_{f \in \mathcal{F}: \|f\|_{H^\alpha} \leq R} \|u_f\|_{H^{\alpha+1}} < \infty.$$

Now for any $f_0 \in \mathcal{F}_{\alpha,r}(R)$, we can use (3.64) to estimate

$$\begin{aligned} \|u_{\hat{f}} - u_{f_0}\|_{H^\beta} &\lesssim \|u_{\hat{f}} - u_{f_0}\|_{L^2}^{\frac{\alpha+1-\beta}{\alpha+1}} \|u_{\hat{f}} - u_{f_0}\|_{H^{\alpha+1}}^{\frac{\beta}{\alpha+1}} \\ &\lesssim \|u_{\hat{f}} - u_{f_0}\|_{L^2}^{\frac{\alpha+1-\beta}{\alpha+1}} \left(M^{\frac{\beta}{\alpha+1}} + \|u_{\hat{f}}\|_{H^{\alpha+1}}^{\frac{\beta}{\alpha+1}}\right). \end{aligned} \tag{3.45}$$

Further, Lemma 3.5.4 and (3.95) yield that

$$\|u_{\hat{f}}\|_{H^{\alpha+1}}^{\frac{\beta}{\alpha+1}} \lesssim 1 + \|\hat{f}\|_{H^\alpha}^{\alpha\beta} \lesssim 1 + \|\hat{F}\|_{H^\alpha}^{\alpha^2\beta} \lesssim 1 + \left(\lambda_\varepsilon^{-1}\mu_{\lambda_\varepsilon}(\hat{f}, f_0)\right)^{\alpha^2\beta}. \tag{3.46}$$

Now set $\delta_\varepsilon := c_1\varepsilon^{\frac{2(\alpha+1)}{2(\alpha+1)+d}}$ for c_1 from the second part of the theorem. We define the events

$$\begin{cases} A_0 := \{\mu_{\lambda_\varepsilon}(\hat{f}_\varepsilon, f_0) < \delta_\varepsilon\} \\ A_j := \{\mu_{\lambda_\varepsilon}(\hat{f}_\varepsilon, f_0) \in (2^{j-1}\delta_\varepsilon, 2^j\delta_\varepsilon]\}, \quad j \geq 1. \end{cases} \tag{3.47}$$

By (3.24) and (3.45)-(3.46), and writing $\hat{\mu}_{\lambda_\varepsilon} := \mu_{\lambda_\varepsilon}(\hat{f}_\varepsilon, f_0)$, we then obtain

$$\begin{aligned}
\mathbb{E}^{\varepsilon}_{F_0}\left[\|u_{\hat{f}} - u_{f_0}\|_{H^\beta}\right] &\lesssim \sum_{j=0}^{\infty} \mathbb{E}^{\varepsilon}_{F_0}\left[1_{A_j}\|u_{\hat{f}} - u_{f_0}\|_{L^2}^{\frac{\alpha+1-\beta}{\alpha+1}}\left(1+\lambda_\varepsilon^{-\alpha^2\beta}\hat{\mu}_{\lambda_\varepsilon}^{\alpha^2\beta}\right)\right] \\
&\lesssim \delta_\varepsilon^{\frac{\alpha+1-\beta}{\alpha+1}} + \sum_{j=1}^{\infty}(2^j\delta_\varepsilon)^{\frac{\alpha+1-\beta}{\alpha+1}}\left(1+\lambda_\varepsilon^{-\alpha^2\beta}(2^j\delta_\varepsilon)^{\alpha^2\beta}\right)\mathbb{P}^{\varepsilon}_{f_0}(A_j) \\
&\lesssim \delta_\varepsilon^{\frac{\alpha+1-\beta}{\alpha+1}}\left(1+\sum_{j=1}^{\infty}2^{\frac{j(\alpha+1-\beta)}{\alpha+1}}(1+(c2^j)^{\alpha^2\beta})\exp\big(-\frac{2^{2j}\delta_\varepsilon^2}{c_2^2\varepsilon^2}\big)\right) \\
&\lesssim \delta_\varepsilon^{\frac{\alpha+1-\beta}{\alpha+1}}(1+o(\varepsilon)),
\end{aligned} \tag{3.48}$$

where c_2 is the constant from (3.24). The theorem is proved. □

Proof of Theorem 3.3.5. We apply Lemma 3.5.5 with $f_2 = \hat{f}$ and $f_1 = f_0 \in \mathcal{F}_{\alpha,r}(R)$, so that $\|u_{f_1}\|_{C^1} \vee \|f_1\|_{C^1}$ is bounded by some fixed $B = B(R)$ (cf. (3.59) and Lemma 3.5.3). Thus, writing $\hat{F}_\varepsilon := \Phi^{-1} \circ \hat{f}_\varepsilon$ and using (3.94),

$$\begin{aligned}
\mathbb{E}^{\varepsilon}_{f_0}\|\hat{f}_\varepsilon - f_0\|_{L^2} &\lesssim \mathbb{E}^{\varepsilon}_{f_0}\left[\|u_{\hat{f}_\varepsilon} - u_{f_0}\|_{H^2}\|\hat{f}_\varepsilon\|_{C^1}\right] \\
&\lesssim \mathbb{E}^{\varepsilon}_{f_0}\left[\|u_{\hat{f}_\varepsilon} - u_{f_0}\|_{L^2}^{\frac{(\alpha-1)}{\alpha+1}}\|u_{\hat{f}_\varepsilon} - u_{f_0}\|_{H^{\alpha+1}}^{\frac{2}{\alpha+1}}(1+\|\hat{F}_\varepsilon\|_{C^1})\right].
\end{aligned}$$

We now choose $\delta_\varepsilon := c_1\varepsilon^{\frac{2(\alpha+1)}{2(\alpha+1)+d}}$ where c_1 is the constant from the second part of Theorem 3.3.3. Bounding $\|u_{\hat{f}_\varepsilon} - u_{f_0}\|_{H^{\alpha+1}}$ as in (3.45)-(3.46), splitting the expectation into A_j, $j \geq 0$ as defined in (3.47) and using the concentration inequality (3.24), we obtain as in (3.48) the desired inequality

$$\mathbb{E}^{\varepsilon}_{f_0}\|\hat{f}_\varepsilon - f_0\|_{L^2} \lesssim \delta_\varepsilon^{\frac{\alpha-1}{\alpha+1}}(1+o(\varepsilon)).$$

□

Proof of Theorem 3.3.4. We only prove the more difficult case $d \geq 2$.

1. Let $f_0 = 1$. By direct computation, one verifies that the unique classical solution to (3.23) with $g = 1, \mathcal{O} = D$ is

$$u_{f_0}(x) = \frac{1}{2d}\left(\|x\|^2 - 1\right), \qquad \nabla u_{f_0}(x) = \frac{x}{d}.$$

Thus we have that for some $1/2 < a < b < 1$,

$$[a,b]^d \subset D, \qquad \frac{1}{2d} \leq \partial_{x_i}u_{f_0}(x) \leq \frac{1}{d} \quad \text{for all } i = 1, ..., d \text{ and } x \in [a,b]^d.$$

2. Now let $\Psi : \mathbb{R} \to \mathbb{R}$ be a 1-dimensional, compactly supported, at least $(\alpha+1)$-regular Daubechies wavelet (see [72], Theorem 4.2.10). Then, for all integers $j \geq 1$, for suitable constants n_j, $c > 1$ and shift vectors $v^{j,r} = (v_1^{j,r}, ..., v_d^{j,r})$ to be chosen later, we define the tensor wavelets $\Psi_{j,r}$, $r = 1, ..., n_j$ by

$$\Psi_{j,r}(x) = 2^{\frac{jd}{2}} c^{-\frac{d-1}{2}} \Psi(2^j x_1 + v_1^{j,r}) \prod_{i=2}^{d} \Psi\left(\frac{2^j}{c} x_i + v_i^{j,r}\right).$$

Note that the $\Psi_{j,r}$ are 'steeper' by a fixed constant c in x_1-direction than in any other direction. Due to the compact support of Ψ, there exists a constant c_0 which depends only on c and Ψ such that for all $j \geq j_0$ large enough, we can set $n_j = c_0 2^{jd}$ and find suitable vectors $v^{j,r}$ such that all $\Psi_{j,r}$ are supported in the interior $[a,b]^d$ with disjoint support. For some sufficiently small constant $\kappa > 0$, we define

$$f_m := f_0 + \kappa 2^{-j(\alpha+d/2)} \sum_{r=1}^{n_j} \beta_{r,m} \Psi_{j,r}, \qquad m = 1, ..., M, \tag{3.49}$$

where $\beta_{r,m},\ m = 1, ..., M$ will be chosen later as a suitably separated elements of the hypercube $\beta_r \in \{-1, 1\}^{n_j}$.

3. We choose κ small enough (independently of $c > 1$), as follows. By the wavelet characterisation of Sobolev norms, all f_m of the form (3.49) lie in a fixed H^α-ball of radius $C\kappa$, for some universal constant $C > 0$, in particular $\|f_m - f_0\|_\infty$ can be made as small as desired for κ small enough, so that all the $f_m > K_{min}$. Arguing as in (3.42), using $L_{f_0} = \Delta$ (the standard Laplacian), (3.79), the multiplicative inequality (3.58), Lemma 3.5.3 and the Sobolev embedding $H^\alpha \subseteq C^{1+\eta}$ (for some small $\eta > 0$), we have (uniformly for all f_m)

$$\begin{aligned} \|u_{f_m} - u_{f_0}\|_{\mathcal{C}^2} &= \|V_{f_0}\left[\nabla \cdot ((f_m - f_0)\nabla u_{f_m})\right]\|_{\mathcal{C}^2} \\ &\lesssim \|\nabla \cdot ((f_m - f_0)\nabla u_{f_m})\|_{\mathcal{C}^0} \\ &\lesssim \|(f_m - f_0)\nabla u_{f_m}\|_{\mathcal{C}^1} \\ &\lesssim \|f_m - f_0\|_{\mathcal{C}^1}\|u_{f_m}\|_{\mathcal{C}^2} \\ &\lesssim \|f_m - f_0\|_{H^\alpha}(1 + \|f_m\|_{\mathcal{C}^1}^2). \end{aligned}$$

Therefore, $\sup_m \|u_{f_m}\|_{\mathcal{C}^2} < \infty$ and we can pick κ so small that for all f_m of the form (3.49),

$$\frac{1}{4d} \leq \partial_{x_i} u_{f_m}(x) \leq \frac{2}{d} \quad \text{for all } \ i = 1, ..., d \ \text{ and } \ x \in [a,b]^d. \tag{3.50}$$

4. Next, we want to apply Theorem 6.3.2 from [72], for which two steps are needed: an appropriate lower bound on the H^2-distance between the u_{f_m}'s and a suitable upper bound on the KL-divergence of the laws $\mathbb{P}^\varepsilon_{f_m}, \mathbb{P}^\varepsilon_{f_0}$.

5. We begin with the lower bound. By the isomorphism (3.75), for all $u \in H_0^2$ and $f \in \mathcal{F}$, we have that

$$\|u\|_{H^2} \gtrsim \|\Delta u\|_{L^2} = \|f^{-1}(L_f u - \nabla u \cdot \nabla f)\|_{L^2} \geq \|f\|_\infty^{-1} \|L_f u - \nabla u \cdot \nabla f\|_{L^2}.$$

For all $m, m' = 1, ..., M$, using this inequality with $f = f_m$,

$$u = u_{f_m} - u_{f_{m'}} = V_{f_m}[\nabla \cdot (f_{m'} - f_m) \nabla u_{f_{m'}})] \tag{3.51}$$

in view of (3.42), and $\sup_m \|f_m\|_{C^1} < \infty$,

$$\begin{aligned} \|u_{f_m} - u_{f_{m'}}\|_{H^2} &\gtrsim \left\| \nabla \cdot \Big((f_m - f_{m'}) \nabla u_{f_{m'}} \Big) \right\|_{L^2} - \|\nabla(u_{f_m} - u_{f_{m'}}) \cdot \nabla f_m\|_{L^2} \\ &\geq \|\nabla(f_m - f_{m'}) \cdot \nabla u_{f_{m'}}\|_{L^2} - \|(f_m - f_{m'}) \Delta u_{f_{m'}}\|_{L^2} \\ &\quad - \|u_{f_m} - u_{f_{m'}}\|_{H^1} \|f_m\|_{C^1} =: I - II - III. \end{aligned} \tag{3.52}$$

We will later show that the second and third terms are of smaller order than the first term. Using (3.50), we see

$$\begin{aligned} I &= \left\| \sum_{i=1}^d \partial_{x_i}(f_m - f_{m'}) \partial_{x_i} u_{f_{m'}} \right\|_{L^2} \\ &\geq \|\partial_{x_1}(f_m - f_{m'}) \partial_{x_1} u_{f_{m'}}\|_{L^2} - \sum_{i=2}^d \|\partial_{x_i}(f_m - f_{m'}) \partial_{x_i} u_{f_{m'}}\|_{L^2} \\ &\geq \frac{1}{4d} \|\partial_{x_1}(f_m - f_{m'})\|_{L^2} - \frac{2}{d} \sum_{i=2}^d \|\partial_{x_i}(f_m - f_{m'})\|_{L^2}. \end{aligned} \tag{3.53}$$

To estimate this further, we calculate that for any $i = 2, ..., d$,

$$\begin{aligned} \partial_{x_i} \Psi_{j,r}(x) &= 2^{\frac{jd}{2}} c^{-\frac{d-1}{2}} \Psi(2^j x_1 + v_1^{j,r}) \\ &\qquad \times \Big(\prod_{k=2,\, k \neq i}^d \Psi\Big(\frac{2^j}{c} x_k + v_k^{j,r}\Big)\Big) \frac{2^j}{c} \Psi' \left(\frac{2^j}{c} x_i + v_i^{j,r} \right). \end{aligned}$$

Similarly calculating $\partial_{x_1} \Psi_{j,r}$ and summing over $r = 1, ..., n_j$, we obtain

$$\|\partial_{x_i}(f_m - f_{m'})\|_{L^2} = \frac{1}{c} \|\partial_{x_1}(f_m - f_{m'})\|_{L^2}, \quad i = 2, ..., d.$$

Thus, choosing c large enough and combining this with (3.53), we can ensure that

$$\begin{aligned} I &\gtrsim \frac{1}{4d} \|\partial_{x_1}(f_m - f_{m'})\|_{L^2} - \frac{2(d-1)}{cd} \|\partial_{x_1}(f_m - f_{m'})\|_{L^2} \\ &\geq \frac{1}{8d} \|\partial_{x_1}(f_m - f_{m'})\|_{L^2}. \end{aligned}$$

Moreover, as the first partial derivatives of the $\Psi_{j,r}$ still have disjoint support, they are orthonormal in L^2 and by Parseval's identity we have

$$\begin{aligned}\|\partial_{x_1}(f_m - f_{m'})\|_{L^2}^2 &= \kappa^2 2^{-2j(\alpha+d/2)} \sum_{j=1}^{n_j} |\beta_{r,m} - \beta_{r,m'}|^2 \|\partial_{x_1}\Psi_{j,r}\|_{L^2}^2 \\ &= \|\partial_{x_1}\Psi_{0,1}\|_{L^2}^2 \kappa^2 2^{-2j(\alpha-1+d/2)} \sum_{j=1}^{n_j} |\beta_{r,m} - \beta_{r,m'}|^2.\end{aligned} \tag{3.54}$$

By the Varshamov-Gilbert-bound (Example 3.1.4 in [72]), for constants $c_1, c_2 > 0$ independent of j, we can find a subset $\mathcal{M}_j \subset \{-1,1\}^{c_0 2^{jd}}$ of cardinality $M_j = 2^{c_1 2^{jd}}$ such that

$$\sum_{j=1}^{n_j} |\beta_{r,m} - \beta_{r,m'}|^2 \geq c_2 2^{jd}$$

whenever $m \neq m'$. For such a subset $\mathcal{M}_j$, by (3.54) we have

$$I \gtrsim \|\partial_{x_1}(f_m - f_{m'})\|_{L^2} \gtrsim 2^{-j(\alpha-1)}. \tag{3.55}$$

6. We next show that II and III in (3.52) are of smaller order as $j \to \infty$. With the above choice of f_m's, we have from Parseval's identity and (3.57)

$$\begin{aligned}II^2 &\leq \|f_m - f_{m'}\|_{L^2}^2 \|u_{f_m}\|_{\mathcal{C}^2}^2 = \kappa^2 2^{-2j(\alpha+d/2)} \sum_{r=1}^{n_j} |\beta_{r,m} - \beta_{r,m'}|^2 \|u_{f_m}\|_{\mathcal{C}^2}^2 \\ &\lesssim 2^{-2j\alpha} = o(2^{-2j(\alpha-1)}),\end{aligned}$$

and for term III we have, by (3.51), (3.64), Lemma 3.5.2 and arguing as in the first display of Step 7 to follow, that

$$\begin{aligned}\|u_{f_m} - u_{f_{m'}}\|_{H^1} &\lesssim \|u_{f_m} - u_{f'_m}\|_{H^2}^{1/2} \|u_{f_m} - u_{f_{m'}}\|_{L^2}^{1/2} \\ &\lesssim \|\nabla \cdot ((f_m - f_{m'})\nabla u_{f_{m'}})]\|_{L^2}^{1/2} \|[\nabla \cdot ((f_m - f_{m'})\nabla u_{f_{m'}})]\|_{(H_0^2)^*}^{1/2} \\ &\lesssim \|f_m - f_{m'}\|_{H^1}^{1/2} \|f_m - f_{m'}\|_{H^{-1}}^{1/2} \lesssim 2^{-j\alpha} = o(2^{-j(\alpha-1)}),\end{aligned}$$

where the first factor in the last line is bounded by $2^{-j(\alpha/2-1/2)}$ by similar arguments as in (3.54). Combining the last two displayed estimates with (3.52) and (3.55) gives the overall lower bound

$$\|u_{f_m} - u_{f_{m'}}\|_{H^2} \gtrsim 2^{-j(\alpha-1)} \approx \varepsilon^{\frac{2(\alpha-1)}{2(\alpha+1)+d}}$$

with choice $j = j_\varepsilon$ such that $2^j \simeq \varepsilon^{-2/(2\alpha+2+d)}$.

7. Now we show the upper bound. Arguing as in (3.42), using Lemma 3.5.2, integrating by parts and using the wavelet characterisation of the $H^{-1}(\mathbb{R}^d)$-norm (e.g., Section 4.3 in

[72] with $B^s_{2,2} = H^s, s \in \mathbb{R}$) as well as the interior support of the $\Psi_{j,r}$, we estimate

$$\begin{aligned}
\|u_{f_m} - u_{f_0}\|^2_{L^2} &\lesssim \|\nabla \cdot ((f_m - f_0)\nabla u_{f_0})\|^2_{(H^2_0)^*} \\
&= \left(\sup_{\|\psi\|_{H^2_0} \le 1} \left| \int_{\mathbb{R}^d} \nabla\psi \cdot \nabla u_{f_0}(f_m - f_0) \right| \right)^2 \\
&\lesssim \|f_m - f_0\|^2_{H^{-1}(\mathbb{R}^d)} \|u_{f_0}\|_{\mathcal{C}^1} \\
&\simeq \kappa^2 2^{-2j(\alpha + d/2 + 1)} \sum_{r=1}^{n_j} 1 \lesssim 2^{-2j(\alpha+1)}.
\end{aligned}$$

By definition of M_j, using the results in Section 7.4 in [131] and arguing as in (6.16) in [72] we thus bound the information distances as

$$\mathrm{KL}(\mathbb{P}^\varepsilon_{u_{f_m}}, \mathbb{P}^\varepsilon_{u_{f_0}}) \lesssim \varepsilon^{-2} \|u_{f_m} - u_{f_0}\|^2_{L^2} \lesssim \varepsilon^{-2} 2^{-2j(\alpha+1)} = 2^{jd} \lesssim \log M_j,$$

so that the overall result now follows from Theorem 6.3.2 in [72]. □

3.4.4 Proof of Theorems 3.3.6 and 3.3.7

The proof of Theorem 3.3.6 follows the same principle as the proof of Theorem 3.3.3. By arguing exactly as in the first two steps of the proof of Theorem 3.3.3, in order to be able to apply Theorem 3.2.2, we now verify that the map

$$\mathscr{G} : H^\alpha_c \to L^2, \qquad \mathscr{G}(F) := G(\Phi \circ F),$$

satisfies (3.8) with $\mathbb{H} = L^2, \gamma = 4, \kappa = 2$. Let $F, H \in H^\alpha$ and $f = \Phi \circ F$, $h = \Phi \circ H \in \mathcal{F}$. By (3.29), $u_f - u_h$ satisfies

$$(u_f - u_h)|_{\partial\mathcal{O}} = 0, \qquad L_f[u_f - u_h] = (L_h - L_f)[u_h] = (f - h)u_h$$

where L_f is defined in Section 3.5.3.1 below. Using this, the norm estimate (3.87), Lemma 3.5.8, the embedding $H^\alpha \subseteq C^2(\mathcal{O})$ as well as (3.96), we can then estimate

$$\begin{aligned}
\|\mathscr{G}(F) - \mathscr{G}(H)\|_{L^2} &= \|u_f - u_h\|_{L^2} \\
&\lesssim (1 + \|f\|_\infty) \|(f-h)u_h\|_{(H^2_0)^*} \\
&\le (1 + \|f\|_\infty) \|u_h\|_{\mathcal{C}^2} \|f - h\|_{(H^2_0)^*} \\
&\lesssim (1 + \|f\|_\infty)(1 + \|h\|_\infty) \|f - h\|_{(H^2)^*} \\
&\lesssim \left(1 + \|F\|^2_\infty \vee \|H\|^2_\infty\right) \|F - H\|_{(H^2)^*} \left(1 + \|F\|^2_{C^2} \vee \|H\|^2_{C^2}\right) \\
&\lesssim \left(1 + \|F\|^4_{H^\alpha} \vee \|H\|^4_{H^\alpha}\right) \|F - H\|_{(H^2)^*}.
\end{aligned}$$

Thus (3.8) is fulfilled for $\gamma = 4$ and $\kappa = 2$. The existence of maximizers now follows from the first part of Theorem 3.2.2. The proof of the concentration inequality (3.30) is completely analogous to the proof of (3.24), and the convergence rate (3.31) follows from the same argument as in the proof of Theorem 3.3.3, utilizing Lemma 3.5.8 in place of Lemma 3.5.4.

Finally, the proof Theorem 3.3.7 is analogous to that of Theorem 3.3.5, but using Lemma 3.5.9 instead of Lemma 3.5.5, and is left to the reader.

3.5 Some PDE facts

In this section, we collect some key PDE facts which are needed to prove the results in Section 3.3.

3.5.1 Preliminaries

Besides the classical Hölder spaces $C^\alpha(\mathcal{O})$, we will also need the Hölder-Zygmund spaces $\mathcal{C}^\alpha(\mathcal{O})$, see Section 3.4.2 in [166] for definitions. For $\alpha \geq 0$, $\alpha \notin \mathbb{N}$, we have that $C^\alpha = \mathcal{C}^\alpha$ with equivalent norms, and we have the continuous embeddings $\mathcal{C}^{\alpha'} \subseteq C^\alpha \subseteq \mathcal{C}^\alpha$ for all $\alpha' > \alpha \geq 0$.

We will repeatedly use the multiplicative inequalities

$$\|fg\|_{H^\alpha} \lesssim \|f\|_{H^\alpha}\|g\|_{H^\alpha}, \qquad \alpha > d/2, \tag{3.56}$$

$$\|fg\|_{H^\alpha} \lesssim \|f\|_{\mathcal{C}^\alpha}\|g\|_{H^\alpha}, \qquad \alpha \geq 0, \tag{3.57}$$

$$\|fg\|_{\mathcal{C}^\alpha} \lesssim \|f\|_{\mathcal{C}^\alpha}\|g\|_{\mathcal{C}^\alpha}, \qquad \alpha \geq 0 \tag{3.58}$$

for all f, g in the appropriate function spaces, which follow from Remark 1 on p.143 and Theorem 2.8.3 in [166]. For any $\alpha > d/2$ and $0 \leq \eta < \alpha - d/2$, we also need the continuous embedding $H^\alpha \subseteq C^\eta$, with the norm estimate

$$\forall f \in H^\alpha, \ \ \|f\|_{C^\eta} \lesssim \|f\|_{H^\alpha}. \tag{3.59}$$

Let $\mathrm{tr}[\cdot]$ denote the usual trace operator for functions defined on $\mathcal{O}$ (for the definition on Sobolev spaces, see, e.g., Chapter 5.5 in [65]). In this and the next section, we will repeatedly use the fact that the standard Laplacian Δ and $\mathrm{tr}[\cdot]$ establish topological isomorphisms between appropriate Sobolev and Hölder-Zygmund spaces. That is, for each $\alpha \geq 0$, we have the topological isomorphisms

$$(\Delta, \mathrm{tr}) : H^{\alpha+2}(\mathcal{O}) \to H^\alpha(\mathcal{O}) \times H^{\alpha+3/2}(\partial\mathcal{O}), \quad u \mapsto (\Delta u, \mathrm{tr}[u]), \tag{3.60}$$

$$(\Delta, \mathrm{tr}) : \mathcal{C}^{\alpha+2}(\mathcal{O}) \to \mathcal{C}^\alpha(\mathcal{O}) \times \mathcal{C}^{\alpha+2}(\partial\mathcal{O}), \quad u \mapsto (\Delta u, \mathrm{tr}[u]), \tag{3.61}$$

which follow from Theorem II.5.4 in [108] and Theorem 4.3.4 in [166] respectively. Moreover, for any $\alpha \geq 1$, we will use the notation

$$H_0^\alpha(\mathcal{O}) := \{f \in H^\alpha(\mathcal{O}) \mid \mathrm{tr}[f] = 0\}, \quad \mathcal{C}_0^\alpha(\mathcal{O}) := \{f \in \mathcal{C}^\alpha(\mathcal{O}) \mid \mathrm{tr}[f] = 0\}. \tag{3.62}$$

We also need the following interpolation inequalities. For all $\beta_1, \beta_2 \geq 0$ and $\theta \in [0,1]$, there exists a constant $C < \infty$ such that

$$\forall u \in \mathcal{C}^{\beta_1} \cap \mathcal{C}^{\beta_2}: \quad \|u\|_{\mathcal{C}^{\theta\beta_1+(1-\theta)\beta_2}} \leq C\|u\|_{\mathcal{C}^{\beta_1}}^\theta \|u\|_{\mathcal{C}^{\beta_2}}^{1-\theta}, \tag{3.63}$$

$$\forall u \in H^{\beta_1} \cap H^{\beta_2}: \quad \|u\|_{H^{\theta\beta_1+(1-\theta)\beta_2}} \leq C\|u\|_{H^{\beta_1}}^\theta \|u\|_{H^{\beta_2}}^{1-\theta}, \tag{3.64}$$

see Theorems 1.3.3 and 4.3.1 in [165] (and note $\mathcal{C}^\beta = B_{\infty,\infty}^\beta, H^\beta = B_{2,2}^\beta$).

3.5.2 Divergence form equation

3.5.2.1 Estimates for V_f

For each $f \in C^1(\bar{\mathcal{O}})$ with $f \geq K_{min} > 0$, we define the differential operator

$$L_f : H_0^2(\mathcal{O}) \to L^2(\mathcal{O}), \qquad L_f[u] = \nabla \cdot (f \nabla u).$$

By standard theory for elliptic PDEs, L_f has a linear, continuous inverse operator, which we denote by

$$V_f : L^2(\mathcal{O}) \to H_0^2(\mathcal{O}), \qquad \psi \mapsto V_f[\psi],$$

see [65], Theorem 4 in Chapter 6.3. In other words, for each right hand side $\psi \in L^2$, there exists a unique function $w_{f,\psi} := V_f[\psi] \in H_0^2$ solving the Dirichlet problem

$$\begin{cases} L_f[w_{f,\psi}] = \psi & \text{on } \mathcal{O}, \\ w_{f,\psi} = 0 & \text{on } \partial\mathcal{O} \end{cases} \tag{3.65}$$

weakly, i.e. in the sense that the identity

$$-\int_{\mathcal{O}} \sum_{i=1}^d f D_i w_{f,\psi} D_i v = \int_{\mathcal{O}} \psi v \tag{3.66}$$

holds for all test functions $v \in H_0^1(\mathcal{O})$ (cf. [65], Chapter 6). By the zero boundary conditions of (3.23) and the divergence theorem, any classical solution (i.e. C^2 solution) must be equal to the unique weak solution when interpreted as an H_0^2 function.

Theorem 4 in Chapter 6.3 of [65] implies that there exists a constant $C = C_f$ (allowed to depend on f) such that for all $\psi \in L^2$, we have the norm estimate $\|V_f[\psi]\|_{H^2} \leq C_f \|\psi\|_{L^2}$, and we need a result that tracks the dependence of C_f on f in a quantitative way. We first establish that when we only seek an $L^p \to L^p$-estimate, $p \in \{2, \infty\}$, rather than an $L^2 \to H^2$-estimate, the constant merely depends on the lower bound K_{min} for f.

Lemma 3.5.1. *Let $K_{min} > 0$. Then there exists $C = C(d, \mathcal{O}, K_{min})$ such that for all $f \in C^2(\mathcal{O})$ with $f \geq K_{min} > 0$ and $\psi \in L^2$, we have*

$$\|V_f[\psi]\|_{L^2} \leq C\|\psi\|_{L^2} \tag{3.67}$$

and for all $\psi \in C^\eta(\mathcal{O}), \eta > 0$,

$$\|V_f[\psi]\|_\infty \leq C\|\psi\|_\infty. \tag{3.68}$$

Proof. Assume first that $\psi \in C^\eta(\mathcal{O})$ so that $V_f[\psi] \in C(\bar{\mathcal{O}}) \cap C^2(\mathcal{O})$ (see after (3.23)). Then we have the Feynman-Kac formula

$$V_f[\psi](x) = -\frac{1}{2}\mathbb{E}^x\left[\int_0^{\tau_\mathcal{O}} \psi(X_s^f)ds\right], \qquad x \in \mathcal{O}, \tag{3.69}$$

where $(X_s^f : s \geq 0)$ is a diffusion Markov process started at $x \in \mathcal{O}$ with infinitesimal generator $L_f/2$ and expectation operator $\mathbb{E}^x$, and where $\tau_\mathcal{O}$ is the exit time of X_s^f from $\mathcal{O}$, see, e.g., Theorem 1.2 in Section II of [12]. We also record that, by Theorem 4.3 in Section VII of [12] and inspection of its proof, there exists a constant c_1 only depending on the lower bound $K_{min} < f$ and on d, such that the transition densities of $(X_s^f : s \geq 0)$ exist and satisfy the estimate

$$p_f(t, x, y) \leq c_1 t^{-d/2}, \qquad t > 0, \quad x, y \in \mathbb{R}^d. \tag{3.70}$$

Then, arguing as in the proof of Theorem 1.17 in [42], with (3.70) replacing the standard heat kernel estimate for Brownian motion, we obtain that $\sup_{x\in\mathcal{O}} \mathbb{E}^x\tau_\mathcal{O} \leq c$, with $c = c(\mathcal{O}, d, c_1)$, and hence (3.68) follows from

$$\|V_f[1]\|_\infty \leq \sup_{x\in\mathcal{O}} \mathbb{E}^x\tau_\mathcal{O} \leq c. \tag{3.71}$$

Using what precedes one further shows that V_f has a representation via a non-negative and symmetric integral kernel $G_f(\cdot,\cdot)$, such that

$$V_f[\psi](x) = -\int_\mathcal{O} G_f(x,y)\psi(y)dy, \ x \in \mathcal{O}, \quad \forall\ \psi \in C^\eta(\mathcal{O}). \tag{3.72}$$

Then using (3.71), the Cauchy-Schwarz inequality and the positivity of G we have for all $\psi \in C^\eta(\mathcal{O})$,

$$\|V_f[\psi]\|_{L^2}^2 \leq \int_\mathcal{O}\int_\mathcal{O} G_f(x,y)dy \int_\mathcal{O} G_f(x,y)\psi^2(y)dydx \leq \|V_f[1]\|_\infty^2\|\psi\|_{L^2}^2,$$

whence (3.67) follows for $\psi \in C^\eta(\mathcal{O})$, and extends to $\psi \in L^2$ by approximation since V_f is a continuous operator on $L^2(\mathcal{O})$ (as established above). □

Lemma 3.5.1 will be used in the proof of the following stronger elliptic regularity estimate.

Lemma 3.5.2. *Let $K_{min} > 0$. Then there exists a universal constant $C > 0$ such that for all $f \in C^2(\mathcal{O})$ with $f \geq K_{min}$ and $\psi \in L^2(\mathcal{O})$, the unique weak solution $w_{f,\psi} = V_f[\psi]$ to (3.65) satisfies*

$$\|V_f[\psi]\|_{H^2} \leq C(1 + \|f\|_{C^1})\|\psi\|_{L^2}, \tag{3.73}$$

$$\|V_f[\psi]\|_{L^2} \leq C(1 + \|f\|_{C^1})\|\psi\|_{(H_0^2)^*}, \tag{3.74}$$

where C only depends on K_{min} and $\mathcal{O}, d$.

Proof. Let $f \in C^1$ and $\psi \in L^2$. By (3.60), there exists a constant $C > 0$ depending only on $\mathcal{O}, d$ such that for all $u \in H_0^2$,

$$C^{-1}\|\Delta u\|_{L^2} \leq \|u\|_{H^2} \leq C\|\Delta u\|_{L^2}. \tag{3.75}$$

Moreover we have by the definition of L_f that

$$\Delta u = f^{-1}(L_f u - \nabla f \cdot \nabla u). \tag{3.76}$$

Writing $w = w_{f,\psi}$ and utilising (3.75) and (3.76), we can estimate

$$\begin{aligned} \|w\|_{H^2} &\leq C\|\Delta w\|_{L^2} = C\left\|f^{-1}(\psi - \nabla w \cdot \nabla f)\right\|_{L^2} \\ &\leq CK_{min}^{-1}(\|\psi\|_{L^2} + \|f\|_{C^1}\|w\|_{H^1}.) \end{aligned} \tag{3.77}$$

By choosing the test function $-w \in H_0^1$ in the weak formulation (3.66), we have that

$$K_{min}\int_{\mathcal{O}}|Dw|^2 \leq \int_{\mathcal{O}}\sum_{i=1}^{d} f(D_i w)^2 = \int_{\mathcal{O}} -\psi w \ \leq \frac{1}{2}\int_{\mathcal{O}}(\psi^2 + w^2).$$

Combining this with (3.77) and Lemma 3.5.1, we finally obtain that for constants C', C'', C''' only depending on K_{min} and $\mathcal{O}$, we have

$$\begin{aligned} \|w\|_{H^2} &\leq C'K_{min}^{-1}(\|\psi\|_{L^2} + \|f\|_{C^1}C''(\|\psi\|_{L^2} + \|w\|_{L^2})) \\ &= C'''(1 + \|f\|_{C^1})\|\psi\|_{L^2}, \end{aligned}$$

which proves (3.73).

Next, using the divergence theorem and (3.73), we obtain (3.74) from

$$\begin{aligned}
\|V_f[\psi]\|_{L^2} &= \sup_{\varphi \in C_c^\infty,\ \|\varphi\|_{L^2} \le 1} \left| \int_{\mathcal{O}} V_f[\psi]\varphi \right| \\
&= \sup_{\varphi \in C_c^\infty(\mathcal{O}),\ \|\varphi\|_{L^2} \le 1} \left| \int_{\mathcal{O}} V_f[\psi] L_f V_f[\varphi] \right| \\
&= \sup_{\varphi \in C_c^\infty(\mathcal{O}),\ \|\varphi\|_{L^2} \le 1} \left| \int_{\mathcal{O}} \psi V_f[\varphi] \right| \\
&\le C(1 + \|f\|_{C^1}) \sup_{\varphi \in H_0^2,\ \|\varphi\|_{H^2} \le 1} \left| \int_{\mathcal{O}} \psi\varphi \right| = C(1 + \|f\|_{C^1}) \|\psi\|_{(H_0^2)^*}.
\end{aligned}$$

□

3.5.2.2 Estimates for G

Now we turn to the forward map G representing the solutions of the PDE (3.23). The following norm estimate for the $\mathcal{C}^2$-Hölder-Zygmund norm of $G(f) = u_f$ is needed.

Lemma 3.5.3. *Suppose that for some $K_{min} > 0$, $\alpha > d/2 + 2$ and $g \in \mathcal{C}^\eta(\mathcal{O}), \eta > 0$, $\tilde{\mathcal{F}}$ is as in (3.32) and u_f denotes the unique solution of (3.23). Then there exists $C = C(d, \mathcal{O}, K_{min}, \|g\|_\infty)$ such that for all $f \in \tilde{\mathcal{F}}$,*

$$\|u_f\|_{\mathcal{C}^2} \le C\left(1 + \|f\|_{\mathcal{C}^1}^2\right). \tag{3.78}$$

Proof. The proof is similar to that of Lemma 3.5.2. By (3.61), there exists a constant $C > 0$ depending only on $\mathcal{O}, d$ such that for all functions $u \in \mathcal{C}_0^2(\mathcal{O})$, we have

$$C^{-1}\|\Delta u\|_{\mathcal{C}^0} \le \|u\|_{\mathcal{C}^2} \le C\|\Delta u\|_{\mathcal{C}^0}. \tag{3.79}$$

Using this, the PDE (3.23), the multiplicative inequality (3.58) and the interpolation inequality (3.63), we can estimate as in (3.77)

$$\begin{aligned}
\|u_f\|_{\mathcal{C}^2} &\lesssim \|f^{-1}(g - \nabla f \cdot \nabla u_f)\|_{\mathcal{C}^0} \lesssim \|f^{-1}\|_{\mathcal{C}^0} \left(\|g\|_{\mathcal{C}^0} + \|f\|_{\mathcal{C}^1}\|u_f\|_{\mathcal{C}^1}\right) \\
&\lesssim K_{min}^{-1}\left(\|g\|_{\mathcal{C}^0} + \|f\|_{\mathcal{C}^1}\|u_f\|_{\mathcal{C}^2}^{1/2}\|u_f\|_{\mathcal{C}^0}^{1/2}\right).
\end{aligned}$$

Dividing this inequality by $\|u_f\|_{\mathcal{C}^2}^{1/2}$ whenever $\|u_f\|_{\mathcal{C}^2}^{1/2} \ge 1$ and otherwise estimating it by 1, we obtain that

$$\|u_f\|_{\mathcal{C}^2} \lesssim 1 + K_{min}^{-2}\left(\|g\|_{\mathcal{C}^0}^2 + \|f\|_{\mathcal{C}^1}^2\|u_f\|_{\mathcal{C}_0}\right) \lesssim 1 + K_{min}^{-2}\left(\|g\|_\infty^2 + \|f\|_{\mathcal{C}^1}^2\|g\|_\infty\right)$$

where in last step we used $\|\cdot\|_{\mathcal{C}^0} \lesssim \|\cdot\|_\infty$ and Lemma 3.5.1. □

We also need that the forward map G maps bounded sets in H^α onto bounded sets in $H^{\alpha+1}$.

Lemma 3.5.4. *Suppose that $\alpha, \tilde{\mathcal{F}}$ are as in Lemma 3.5.3 and for some $g \in H^{\alpha-1}(\mathcal{O})$, let $u_f = w_{f,g}, f \in \tilde{\mathcal{F}}$, be the unique solution of (3.23). Then $u_f \in H^{\alpha+1}(\mathcal{O})$ and there exists a constant $C = C(\alpha, d, \mathcal{O}, K_{min}) > 0$ such that*

$$\|u_f\|_{H^{\alpha+1}} \leq C\big(1 + \|f\|_{H^\alpha}^{\alpha^2+\alpha}\big)\big(\|g\|_{H^{\alpha-1}}^{\alpha+1} \vee \|g\|_{H^{\alpha-1}}^{1/(\alpha+1)}\big). \tag{3.80}$$

Proof. First, suppose $f \in C^\infty \cap \tilde{\mathcal{F}}$. By (3.60), the standard Laplacian Δ establishes an isomorphism between $H_0^{\alpha+1}$ and $H^{\alpha-1}$, and by Theorem 8.13 in [71], $u_f \in H_0^{\alpha+1}$. Then (3.76) and the multiplicative inequality (4.18) give

$$\begin{aligned}\|u_f\|_{H^{\alpha+1}} &\lesssim \|f^{-1}(g - \nabla f \cdot \nabla u_f)\|_{H^{\alpha-1}} \\ &\lesssim \|f^{-1}\|_{H^{\alpha-1}}(\|g\|_{H^{\alpha-1}} + \|f\|_{H^\alpha}\|u_f\|_{H^\alpha}).\end{aligned}$$

Noting that the map $\Psi : (K_{min}, \infty) \to \mathbb{R},\ x \mapsto x^{-1}$ satisfies (3.20), (3.95) implies that there exists $c > 0$ such that for all $f \in \mathcal{F}$,

$$\|f^{-1}\|_{H^{\alpha-1}} \leq c(1 + \|f\|_{H^{\alpha-1}}^{\alpha-1}).$$

Using this and (3.64), we obtain

$$\begin{aligned}\|u_f\|_{H^{\alpha+1}} &\lesssim (1 + \|f\|_{H^{\alpha-1}}^{\alpha-1})(\|g\|_{H^{\alpha-1}} + \|f\|_{H^\alpha}\|u_f\|_{H^\alpha}) \\ &\lesssim (1 + \|f\|_{H^\alpha}^{\alpha})\big(\|g\|_{H^{\alpha-1}} + \|u_f\|_{H^{\alpha+1}}^{\frac{\alpha}{\alpha+1}}\|u_f\|_{L^2}^{\frac{1}{\alpha+1}}\big)\end{aligned}$$

When $\|u_f\|_{H^{\alpha+1}} \leq 1$ we use (3.67) to deduce

$$\|u_f\|_{H^{\alpha+1}} \lesssim (1 + \|f\|_{H^\alpha}^{\alpha})\big(\|g\|_{H^{\alpha-1}} + \|g\|_{L^2}^{\frac{1}{\alpha+1}}\big),$$

and when $\|u_f\|_{H^{\alpha+1}} \geq 1$, then dividing both sides by $\|u_f\|_{H^{\alpha+1}}^{\frac{\alpha}{\alpha+1}}$ and using again (3.67) yields

$$\|u_f\|_{H^{\alpha+1}}^{1/(\alpha+1)} \lesssim (1 + \|f\|_{H^\alpha}^{\alpha})\big(\|g\|_{H^{\alpha-1}} + \|g\|_{L^2}^{\frac{1}{\alpha+1}}\big).$$

Combining the preceding bounds and using $\|\cdot\|_{L^2} \lesssim \|\cdot\|_{H^{\alpha-1}}$ implies (3.80) for smooth $f \in \tilde{\mathcal{F}}$. Now for any $f \in \tilde{\mathcal{F}}$, take $f_n \in C^\infty(\mathcal{O}), f_n > K_{min}/2$, such that $f_n \to f$ in H^α as $n \to \infty$, and hence by (3.80) the sequence u_{f_n} is bounded in $H^{\alpha+1}$. Then applying (3.42) to $u_{f_n} - u_{f_m}, m, n \in \mathbb{N}$, and applying (3.80) with $g = \nabla \cdot ((f_m - f_n)\nabla u_{f_m})$, one shows that u_{f_n} is a Cauchy sequence in $H^{\alpha+1}$ converging to u_f, and taking limits extends the inequality (3.80) to the general case $f \in \mathcal{F}$. $\square$

3.5.2.3 Stability Estimates for G^{-1}

The following estimate for the inverse map $u_f \mapsto f$ allows to obtain convergence rates for $\|\hat f - f_0\|_{L^2}$ via rates for $\|u_{\hat f} - u_{f_0}\|_{H^2}$, with choices $f_0 = f_1$ and $\hat f = f_2$. As $\hat f$ is random we explicitly track the dependence of the constants on f_2.

Lemma 3.5.5. *Let $\alpha > d/2+2, g_{min}, K_{min}, B, \eta$ be given, positive constants and let $\tilde{\mathcal{F}}$ be given by (3.32). For $g \in C^\eta(\mathcal{O})$ with $\inf_{x\in\mathcal{O}} g(x) \geq g_{min}$, denote by u_f the unique solution of (3.23). Then there exists $C = C(g_{min}, K_{min}, B, \mathcal{O}, d) < \infty$ such that for all $f_1, f_2 \in \tilde{\mathcal{F}}$ with $\|f_1\|_{C^1} \vee \|u_{f_1}\|_{C^2} \leq B$, we have*

$$\|f_1 - f_2\|_{L^2} \leq C\|f_2\|_{C^1}\|u_{f_1} - u_{f_2}\|_{H^2}.$$

Proof. For $f_1, f_2 \in \tilde{\mathcal{F}}$ write $h = f_1 - f_2$. By (3.23), we have

$$\begin{aligned}\nabla\cdot(h\nabla u_{f_1}) &= \nabla\cdot(f_1\nabla u_{f_1}) - \nabla\cdot(f_2\nabla u_{f_2}) - \nabla\cdot(f_2\nabla(u_{f_1}-u_{f_2}))\\ &= \nabla\cdot(f_2\nabla(u_{f_2}-u_{f_1})).\end{aligned} \tag{3.81}$$

We can upper bound the $\|\cdot\|_{L^2}$-norm of the right hand side by

$$\begin{aligned}\|\nabla\cdot(f_2\nabla(u_{f_2}-u_{f_1}))\|_{L^2} &\leq \|\nabla f_2\|_\infty\|u_{f_2}-u_{f_1}\|_{H^1} + \|f_2\|_\infty\|u_{f_2}-u_{f_1}\|_{H^2}\\ &\leq 2\|f_2\|_{C^1}\|u_{f_2}-u_{f_1}\|_{H^2}.\end{aligned} \tag{3.82}$$

Next, we lower bound the $\|\cdot\|_{L^2}$-norm of the left side of (3.81). For regular enough v we see from Green's identity (p.17 in [71]) that

$$\langle\Delta u_{f_1}, v^2\rangle_{L^2} + \frac{1}{2}\langle\nabla u_{f_1}, \nabla(v^2)\rangle_{L^2} = \frac{1}{2}\langle\Delta u_{f_1}, v^2\rangle_{L^2} + \frac{1}{2}\int_{\partial\mathcal{O}}\frac{\partial u_{f_1}}{\partial n}v^2.$$

Moreover for $v = e^{-\lambda u_{f_1}}h$ with $\lambda > 0$ to be chosen we have

$$\frac{1}{2}\int_{\mathcal{O}}\nabla(v^2)\cdot\nabla u_{f_1} = -\int_{\mathcal{O}}\lambda\|\nabla u_{f_1}\|^2v^2 + \int_{\mathcal{O}}ve^{-\lambda u_{f_1}}\nabla h\cdot\nabla u_{f_1},$$

so that by the Cauchy-Schwarz inequality

$$\begin{aligned}&\left|\int_{\mathcal{O}}\Big(\frac{1}{2}\Delta u_{f_1} + \lambda\|\nabla u_{f_1}\|^2\Big)v^2 + \int_{\partial\mathcal{O}}\frac{1}{2}\frac{\partial u_{f_1}}{\partial n}v^2\right|\\ &= \left|\langle(\Delta u_{f_1} + \lambda\|\nabla u_{f_1}\|^2), v^2\rangle_{L^2} + \frac{1}{2}\langle\nabla u_{f_1}, \nabla(v^2)\rangle_{L^2}\right|\\ &= \left|\langle h\Delta u_{f_1} + \nabla h\cdot\nabla u_{f_1}, he^{-2\lambda u_{f_1}}\rangle_{L^2}\right| \leq \mu\|\nabla\cdot(h\nabla u_{f_1})\|_{L^2}\|h\|_{L^2}\end{aligned} \tag{3.83}$$

for $\mu = \exp(2\lambda\|u_{f_1}\|_\infty)$. [The preceding argument is adapted from the proof of Theorem 4.1 in [102].] We next lower bound the multipliers of v^2 in the integrands in the first line of the last display. First we have

$$0 < g_{min} \leq g = L_{f_1} u_{f_1} = f_1(x)\Delta u_{f_1} + \nabla f_1 \cdot \nabla u_{f_1}, \quad \text{on } \mathcal{O},$$

so that either $\Delta u_{f_1}(x) \geq g_{min}/2\|f_1\|_\infty$ or $\|\nabla u_{f_1}(x)\|^2 \geq (g_{min}/2\|f_1\|_{C^1})^2$ on $\mathcal{O}$. Since $\|\Delta u_{f_1}\|_\infty \leq c(B)$ this implies for $\lambda = \lambda(g_{min}, B)$ large enough that

$$\frac{1}{2}\Delta u_{f_1}(x) + \lambda\|\nabla u_{f_1}(x)\|^2 \geq c_0 > 0, \quad x \in \mathcal{O}, \tag{3.84}$$

for some $c_0 = c_0(g_{min}, B)$. Next, for the integral over $\partial\mathcal{O}$, we use again $L_{f_1}u_{f_1} = g > 0$ and apply the Hopf boundary point Lemma 6.4.2 in [65]: We have $u_{f_1}(x_0) = 0$ for any $x_0 \in \partial\mathcal{O}$ but $u_{f_1}(x) < 0$ for all $x \in \mathcal{O}$: Indeed, by $g \geq g_{min} > 0$ and the Feynman-Kac formula (3.69) (with $g = \psi$), it suffices to lower bound $\mathbb{E}^x\tau_{\mathcal{O}}$ which satisfies, by Markov's inequality

$$\mathbb{E}^x\tau_{\mathcal{O}} \geq \mathbb{P}^x(\tau_{\mathcal{O}} > 1) \geq \mathbb{P}^x\big(\sup_{0<s\leq 1}\|X_s - x\| < \|x - \partial\mathcal{O}\|\big) > 0$$

in view Theorem V.2.5 in [12] with $\psi(s) = x$ identically for all s. Lemma 6.4.2 in [65] now gives $\partial u_{f_1}/\partial n \geq 0$ for all $x \in \partial\mathcal{O}$. Combining this with (3.83) and (3.84) we deduce

$$\|\nabla \cdot (h\nabla u_{f_1})\|_{L^2}\|h\|_{L^2} \geq c'(g_{min}, K_{min}, B, \mathcal{O}, d)\|v\|_{L^2}^2 \gtrsim \|h\|_{L^2}^2,$$

which together with (3.82) yields the desired estimate. $\square$

3.5.3 Schrödinger equation

3.5.3.1 Estimates for V_f and G

In this section, for each $f \in C(\mathcal{O})$ with $f \geq 0$, let L_f denote the Schrödinger differential operator

$$L_f : H_0^2(\mathcal{O}) \to L^2(\mathcal{O}), \qquad L_f[u] = \Delta u - 2fu,$$

where H_0^2 is given by (3.62). As in the divergence form case, L_f is a bijection with a linear, continuous inverse operator which we again denote by

$$V_f : L^2(\mathcal{O}) \to H_0^2(\mathcal{O}), \qquad \psi \mapsto V_f\,[\psi]\,.$$

In other words, for any $f \in C(\mathcal{O})$ and $\psi \in L^2$ the inhomogeneous equation

$$\begin{cases} \Delta u - 2fu = \psi \quad \text{on} \quad \mathcal{O}, \\ u = 0 \quad \text{on} \quad \partial O \end{cases} \tag{3.85}$$

has a unique weak solution which we shall denote by $\omega_{f,\psi} := V_f[\psi] \in H_0^2(\mathcal{O})$, see Theorem 4 in Chapter 6.3 of [65] for this standard result for elliptic PDEs.

As in the divergence form case, we first observe that for $p \in \{2, \infty\}$, the $L^p \to L^p$ operator norm of V_f can be upper bounded uniformly in f.

Lemma 3.5.6. *There exists a constant $C > 0$ such that for all $f \in C(\mathcal{O})$ with $f \geq 0$ and $\psi \in L^2(\mathcal{O})$, $w_{f,\psi} = V_f[\psi]$ satisfies*

$$\|V_f\,[\psi]\,\|_{L^2} \leq C\|\psi\|_{L^2}$$

and if $\psi \in C(\mathcal{O})$, then also

$$\|V_f\,[\psi]\,\|_{\infty} \leq C\|\psi\|_{\infty}.$$

Proof. We have the Feynman-Kac representation

$$w_{f,\psi}(x) = -\frac{1}{2}\mathbb{E}^x\Big[\int_0^{\tau_{\mathcal{O}}} \psi(X_s)e^{-\int_0^s f(X_r)dr}ds\Big],\ x \in \mathcal{O},\ \psi \in C(\mathcal{O}),$$

where $(X_s : s \geq 0)$ is a standard d-dimensional Brownian motion started at x, with exit time $\tau_{\mathcal{O}}$ from $\mathcal{O}$, see p.84 and Theorem 3.22 of [42] . [These results are applicable as $C(\mathcal{O}) \subseteq J$ with J defined on p.62 of [42], and $C(\mathcal{O}) \subseteq \mathbb{F}(D, q)$ with $\mathbb{F}(D, q)$ defined on p.80 of [42].] The proof is now similar to that of Lemma 3.5.1, using $f \geq 0$ and that $\sup_{x\in\mathcal{O}} \mathbb{E}^x[\tau_{\mathcal{O}}] \leq K(vol(\mathcal{O}), d) < \infty$ by Theorem 1.17 in [42]. □

Using the above lemma, we now show the following regularity estimate.

Lemma 3.5.7. *There exists a constant C such that for all $f \in C^1(\mathcal{O})$ with $f \geq 0$ and $\psi \in L^2(\mathcal{O})$, we have*

$$\|V_f\,[\psi]\,\|_{H^2} \leq C(1 + \|f\|_\infty)\|\psi\|_{L^2}, \tag{3.86}$$

$$\|V_f\,[\psi]\,\|_{L^2} \leq C(1 + \|f\|_\infty)\|\psi\|_{(H_0^2)^*}. \tag{3.87}$$

Proof. By the norm equivalence (3.75) and (3.85), we have that

$$\begin{aligned}\|V_f\,[\psi]\,\|_{H^2} &\lesssim \|\Delta V_f\,[\psi]\|_{L^2} \leq \|L_f V_f\,[\psi]\,\|_{L^2} + \|f V_f\,[\psi]\,\|_{L^2}\\ &\leq \|\psi\|_{L^2} + \|f\|_\infty\|V_f\,[\psi]\,\|_{L^2} \lesssim (1 + \|f\|_\infty)\|\psi\|_{L^2},\end{aligned}$$

which proves (3.73). The second estimate (3.87) now follows from the same duality argument as in the proof of (3.74). □

Next, we prove some basic boundedness properties of the forward map $G : f \mapsto u_f$.

Lemma 3.5.8. *Suppose that for some $g \in C^\infty(\partial\mathcal{O})$, $\alpha > d/2$ and $K_{min} \geq 0$, $\tilde{\mathcal{F}}$ is as in (3.32), and let u_f be the unique solution of (3.29).*

1. *There exists* $C > 0$ *(independent of* g*) such that for all* $f \in \tilde{\mathcal{F}}$*, we have*

$$\|u_f\|_{\mathcal{C}^2(\mathcal{O})} \leq C(1 + \|f\|_\infty)\|g\|_{\mathcal{C}^2(\mathcal{O})}.$$

2. *There exists* $C > 0$ *(possibly depending on* g*) such that for all* $f \in \tilde{\mathcal{F}}$*,*

$$\|u_f\|_{H^{\alpha+2}(\mathcal{O})} \leq C(1 + \|f\|_{H^\alpha}^{\alpha/2+1}).$$

Proof. By (3.61), $(\Delta, \mathrm{tr}[\cdot])$ is a topological isomorphism between the spaces $\mathcal{C}^2(\mathcal{O})$ and $\mathcal{C}^0(\mathcal{O}) \times \mathcal{C}^2(\partial O)$, whence we deduce that for all $u \in \mathcal{C}^2(\mathcal{O})$, we have the norm estimate

$$\|u\|_{\mathcal{C}^2(\mathcal{O})} \leq C\left(\|\Delta u\|_{\mathcal{C}^0(\mathcal{O})} + \|\mathrm{tr}[u]\|_{\mathcal{C}^2(\partial\mathcal{O})}\right).$$

Using this, the PDE (3.29) and the triangle inequality, we have for $f \in \mathcal{F}$,

$$\begin{aligned}\|u_f\|_{\mathcal{C}^2(\mathcal{O})} &\lesssim \|L_f u_f\|_{\mathcal{C}^0(\mathcal{O})} + \|f u_f\|_{\mathcal{C}^0(\mathcal{O})} + \|\mathrm{tr}[u_f]\|_{\mathcal{C}^2(\partial\mathcal{O})} \\ &\leq \|f\|_\infty \|u_f\|_\infty + \|g\|_{\mathcal{C}^2(\partial\mathcal{O})}.\end{aligned} \tag{3.88}$$

Next, we claim that there exists a constant $C > 0$ such that for all f, g as in the hypotheses, we have

$$\|u_f\|_\infty \leq C\|g\|_\infty. \tag{3.89}$$

Indeed, this can be seen immediately from the fact that $f \geq 0$ and the Feynman-Kac representation (see [42], Theorem 4.7)

$$u_f(x) = \frac{1}{2}\mathbb{E}^x\left[g(X_{\tau_\mathcal{O}})e^{-\int_0^{\tau_\mathcal{O}} f(X_s)ds}\right], \quad x \in \mathcal{O}, \tag{3.90}$$

where $(X_s : s \geq 0), \tau_\mathcal{O}$ are as in the proof of Lemma 3.5.6. Hence, combining (3.89) with (3.88) yields the desired estimate

$$\|u_f\|_{\mathcal{C}^2(\mathcal{O})} \lesssim \|f\|_\infty \|g\|_{L^\infty(\mathcal{O})} + \|g\|_{\mathcal{C}^2(\partial\mathcal{O})} \leq (1 + \|f\|_\infty)\|g\|_{\mathcal{C}^2(\partial\mathcal{O})}.$$

For the second part, we initially assume $f \in C^\infty(\mathcal{O})$ so that $u_f \in C^\infty(\mathcal{O})$ too (see Corollary 8.11 in [71]), and then use the topological isomorphism (Δ, tr) between $H^{\alpha+2}(\mathcal{O})$ and $H^\alpha(\mathcal{O}) \times H^{\alpha+3/2}(\partial\mathcal{O})$, which yields

$$\begin{aligned}\|u_f\|_{H^{\alpha+2}(\mathcal{O})} &\lesssim \|\Delta u_f\|_{H^\alpha(\mathcal{O})} + \|\mathrm{tr}[u_f]\|_{H^{\alpha+3/2}(\partial\mathcal{O})} \lesssim \|f u_f\|_{H^\alpha(\mathcal{O})} + \|g\|_{C^{\alpha+2}(\partial\mathcal{O})} \\ &\lesssim 1 + \|f\|_{H^\alpha}\|u_f\|_{H^\alpha} \lesssim 1 + \|u_f\|_{H^{\alpha+2}}^{\frac{\alpha}{\alpha+2}} \|u_f\|_{L^2}^{\frac{2}{\alpha+2}} \|f\|_{H^\alpha}.\end{aligned}$$

Dividing this by $\|u_f\|_{H^{\alpha+2}}^{\frac{\alpha}{\alpha+2}}$ when $\|u_f\|_{H^{\alpha+2}} \geq 1$ and otherwise estimating it by 1, and using (3.89), we have that

$$\|u_f\|_{H^{\alpha+2}} \lesssim 1 + \|u_f\|_{L^2}\|f\|_{H^\alpha}^{\frac{\alpha+2}{2}} \lesssim 1 + \|g\|_\infty \|f\|_{H^\alpha}^{\frac{\alpha+2}{2}} \lesssim 1 + \|f\|_{H^\alpha}^{\frac{\alpha+2}{2}}.$$

The case of general $f \in \tilde{\mathcal{F}}$ now follows from taking smooth $f_n > K_{min}/2, f_n \to f$ in H^α, showing that u_{f_n} is Cauchy in $H^{\alpha+2}$ (by using (3.75), (3.64), Lemma 3.5.6), and taking limits in the last inequality. Details are left to the reader. □

3.5.3.2 Estimates for G^{-1}

Lemma 3.5.9. *Suppose that for some $\alpha > d/2, K_{min} \geq 0$, $g_{min} > 0$ and $g \in C^\infty(\partial\mathcal{O})$ with $\inf_{x\in\partial\mathcal{O}} g(x) \geq g_{min}$, $\tilde{\mathcal{F}}$ is given by (3.32), and let u_f denote the unique solution of (3.29). Then there exist constants $c_1, c_2 > 0$ such that for all $f_1, f_2 \in \tilde{\mathcal{F}}$, we have*

$$\begin{aligned}\|f_1 - f_2\|_{L^2} \leq c_1\big(&e^{c_2\|f_1\|_\infty} \|u_{f_1} - u_{f_2}\|_{H^2} \\ &+ \|u_{f_2}\|_{C^2}\, e^{c_2\|f_1\vee f_2\|_\infty}\|u_{f_1} - u_{f_2}\|_{L^2}\big).\end{aligned}$$

Proof. Applying Jensen's inequality to the Feynman-Kac representation (3.90), and since $\sup_x \mathbb{E}^x\tau_{\mathcal{O}} \leq c < \infty$ (see the proof of Lemma 3.5.6) yields

$$\inf_{x\in\mathcal{O}} u_f(x) \geq g_{min} \inf_{x\in\mathcal{O}} e^{-\|f\|_\infty \mathbb{E}^x\tau_{\mathcal{O}}} \geq g_{min}e^{-c\|f\|_\infty} > 0. \tag{3.91}$$

Moreover, (3.29) yields that we have $f = \frac{\Delta u_f}{2u_f}$ on $\mathcal{O}$, for all $f \in \tilde{\mathcal{F}}$. Thus, for any $f_1, f_2 \in \tilde{\mathcal{F}}$, we can estimate

$$\begin{aligned}\|f_1 - f_2\|_{L^2} &= \frac{1}{2}\|\frac{\Delta u_{f_1}}{u_{f_1}} - \frac{\Delta u_{f_2}}{u_{f_2}}\|_{L^2} \\ &\lesssim \left\|(\Delta u_{f_1} - \Delta u_{f_2})\, u_{f_1}^{-1}\right\|_{L^2} + \left\|\Delta u_{f_2}\left(u_{f_1}^{-1} - u_{f_2}^{-1}\right)\right\|_{L^2} \\ &\lesssim \big(\inf_{x\in\mathcal{O}} |u_{f_1}(x)|\big)^{-1} \|u_{f_1} - u_{f_2}\|_{H^2} + \|\Delta u_{f_2}\|_{C^2}\|u_{f_1}^{-1} - u_{f_2}^{-1}\|_{L^2}.\end{aligned} \tag{3.92}$$

Further, using the mean value theorem and (3.91), we have that

$$\left|u_{f_1}^{-1} - u_{f_2}^{-1}\right| \leq \max\{u_{f_1}^{-2}, u_{f_2}^{-2}\}\, |u_{f_1} - u_{f_2}| \leq g_{min}^{-2} e^{2c\|f_1\vee f_2\|_\infty} |u_{f_1} - u_{f_2}|.$$

Combining this with (3.92) and using (3.91) once more, we obtain that

$$\|f_1 - f_2\|_{L^2} \lesssim e^{c\|f_1\|_\infty} \|u_{f_1} - u_{f_2}\|_{H^2} + e^{2c\|f_1\vee f_2\|_\infty} \|u_{f_1} - u_{f_2}\|_{L^2},$$

which concludes the proof. □

3.6 Some properties of regular link functions

We define L^p-norms, $0 < p \leq \infty$, in the usual way. By obvious modifications, the following lemma holds also for regular functions $\Phi : (a, b) \to \mathbb{R}$ with arbitrary $-\infty \leq a < b \leq \infty$ and suitable $F, J : \mathcal{O} \to (a, b)$, we restrict to the case $(a, b) = \mathbb{R}$ here.

Lemma 3.6.1. *Suppose $\Phi : \mathbb{R} \to \mathbb{R}$ is a smooth and regular function in the sense of (3.20).*

1. There exists $C < \infty$ such that for all $p \in [1, \infty]$,

$$\forall F \in L^p(\mathcal{O}), \quad \|\Phi \circ F\|_{L^p} \leq C(1 + \|F\|_{L^p}). \tag{3.93}$$

2. For each integer $m \geq 0$, there exists $C < \infty$ such that

$$\forall F \in C^m(\mathcal{O}), \quad \|\Phi \circ F\|_{C^m} \leq C\left(1 + \|F\|_{C^m}^m\right). \tag{3.94}$$

3. For each integer $m \geq d/2$, there exists $C < \infty$ such that for all $F \in H^m(\mathcal{O})$, we have $\Phi \circ F \in H^m(\mathcal{O})$ and

$$\|\Phi \circ F\|_{H^m} \leq C(1 + \|F\|_{H^m}^m). \tag{3.95}$$

4. There exists $C < \infty$ such that for $\kappa \in \{1, 2\}$ and all $F, J \in C^\kappa(\mathcal{O})$,

$$\|\Phi \circ F - \Phi \circ J\|_{(H^\kappa)^*} \leq C \|F - J\|_{(H^\kappa)^*} \left(1 + \|F\|_{C^\kappa}^\kappa \vee \|J\|_{C^\kappa}^\kappa\right). \tag{3.96}$$

The rest of this section is devoted to proving Lemma 3.6.1. To prove (3.94)-(3.95), we need Faá di Bruno's formula (a generalization of the chain rule), which classically holds for C^m functions, and by the chain rule for Sobolev functions (see e.g. [183], Thm. 2.1.11) also holds for H^m functions.

Lemma 3.6.2. *Let $m \in \mathbb{N}$ and suppose that $F : \mathcal{O} \to \mathbb{R}$ and $\Phi : \mathbb{R} \to \mathbb{R}$ are of class $H^m(\mathcal{O})$ and $C^m(\mathbb{R})$ respectively. Then for any $\alpha \in \{1, ..., d\}^m$, the m-th order partial derivative of $f := \Phi \circ F$ in direction $x_{\alpha_1}...x_{\alpha_m}$ is given by*

$$\frac{\partial^m f}{\partial x_{\alpha_1}...\partial x_{\alpha_m}}(x) = \sum_{\pi \in \Pi} \Phi^{(|\pi|)}(F(x)) \prod_{B \in \pi} \frac{\partial^{|B|} F}{\prod_{j \in B} \partial x_{\alpha_j}}(x), \tag{3.97}$$

where π runs through the set Π of all partitions of $\{1, ..., m\}$, and the $B \in \pi$ runs over all 'blocks' B of each partition π.

Proof of (3.93)-(3.94). By (3.20), there exists a constant $c > 0$ only depending on the values of $\Phi(0)$ and $\|\Phi'\|_\infty$ such that for all $x \in \mathbb{R}$, $|\Phi(x)| \leq c(1 + |x|)$, which yields (3.93). For (3.94), let $\alpha \in \{1, ..., d\}^{|\alpha|}$, $1 \leq |\alpha| \leq m$ and let π be a partition of $\{1, ..., |\alpha|\}$. Then the corresponding summand on the right side of (3.97) can be estimated by

$$\left\| \prod_{B\in\pi} \frac{\partial^{|B|}F}{\prod_{j\in B}\partial x_{\alpha_j}} \right\|_\infty \le \|F\|_{C^m}^{|\pi|} \lesssim (1+\|F\|_{C^m}^m)\,.$$

By summing the above display over all such α, π and using (3.93) with $p=\infty$, we obtain (3.94). □

To prove (3.95), we also need the Gagliardo-Nirenberg interpolation inequality (see [138], p.125) in the special case $r=q=2$.

Lemma 3.6.3. *Suppose that $\mathcal{O}\subseteq\mathbb{R}^d$ is a bounded C^∞ domain and that $i=1,...,m$, $a\in[i/m,1]$ and $p\in[1,\infty)$ satisfy*

$$\frac{1}{p} = \frac{1}{2} + \frac{i}{d} - \frac{m}{d}a. \tag{3.98}$$

Then for any $s>0$, there exist constants C_1, C_2 depending only on $m, d, i, a, \mathcal{O}$ and s such that for all $F\in H^m$, we have that $D^iF\in L^p$, and

$$\|D^iF\|_{L^p} \le C_1\|D^mF\|_{L^2}^a\|F\|_{L^2}^{1-a} + C_2\|F\|_{L^s}.$$

Proof of (3.95). Let us write $f=\Phi\circ F$. By (3.93), we have that $\|f\|_{L^2}\le C(1+\|F\|_{L^2})$ whence we only need to estimate $\|D^mf\|_{L^2}$. For any $\alpha\in\{1,...,d\}^m$ we have by (3.97) that

$$\left|\frac{\partial^m f}{\partial x_{\alpha_1}...\partial x_{\alpha_m}}(x)\right|^2 \lesssim \sum_{\pi\in\Pi}\left|\prod_{B\in\pi}\frac{\partial^{|B|}F}{\prod_{j\in B}\partial x_{\alpha_j}}(x)\right|^2$$

Similarly to the proof of (3.94), it thus suffices to prove that for all $\alpha\in\{1,...,d\}^m$ and partition π of $\{1,...,m\}$,

$$\left\| \prod_{B\in\pi} \frac{\partial^{|B|}F}{\prod_{j\in B}\partial x_{\alpha_j}} \right\|_{L^2} \lesssim (1+\|F\|_{H^m}^m). \tag{3.99}$$

Fix some π for the rest of the proof. For $i=1,...,m$, define

$$\pi_i := \{B\in\pi \mid |B|=i\}, \qquad p_i := \frac{2m}{i}.$$

Then we have $\sum_{i=1}^m i|\pi_i| = m$, and hence by Hölder's inequality

$$\begin{aligned}\left\| \prod_{B\in\pi} \frac{\partial^{|B|}F}{\prod_{j\in B}\partial x_{\alpha_j}} \right\|_{L^2} &\le \left\|\prod_{i=1}^m |D^iF|^{|\pi_i|}\right\|_{L^2} \le \prod_{i=1}^m \left\||D^iF|^{|\pi_i|}\right\|_{L^{p_i/|\pi_i|}} \\ &= \prod_{i=1}^m \left\|D^iF\right\|_{L^{p_i}}^{|\pi_i|}.\end{aligned} \tag{3.100}$$

Next, define

$$a_i := \left(\frac{i}{d} + \frac{1}{2} - \frac{i}{2m}\right)\frac{d}{m} \qquad \text{for } i = 1, ..., m. \tag{3.101}$$

To apply Lemma 3.6.3, we verify that for each $i = 1, ..., m$, (i, a_i, p_i) satisfies the conditions of Lemma 3.6.3. By definition, (3.98) is satisfied. Moreover, as $i \le m$, it follows that

$$ma_i = i + \left(\frac{d}{2} - \frac{di}{2m}\right) \ge i,$$

whence we have $\frac{i}{m} \le a_i$. Finally, we need to verify $a_i \le 1$. For this, we note that for $i = 1, ..., m$, choosing $m = d/2$ in (3.101) yields $a_i = a_i(m) = 1$. Moreover, for $m \ge d/2$, we have

$$\frac{\partial a_i(m)}{\partial m} = \frac{2di - 2mi - dm}{2m^3} \le \frac{di - dm}{2m^3} \le 0, \tag{3.102}$$

so that $\alpha_i \le 1$.

Applying Lemma 3.6.3 with $s = 2$ to (3.100) and using that $\sum_{i=1}^m |\pi_i| \in [1, m]$ yields that

$$\begin{aligned}\left\| \prod_{B\in\pi} \frac{\partial^{|B|} F}{\prod_{j\in B} \partial x_{\alpha_j}} \right\|_{L^2} &\lesssim \prod_{i=1}^m \left(\|D^m F\|_{L^2}^{a_i} \|F\|_{L^2}^{1-a_i} + \|F\|_{L^2} \right)^{|\pi_i|} \\ &\lesssim \prod_{i=1}^m \|F\|_{H^m}^{|\pi_i|} \lesssim 1 + \|F\|_{H^m}^m.\end{aligned}$$

□

Proof of (3.96). 1. Let $\kappa \in \{1, 2\}$ and fix $F, J \in C^\kappa(\mathcal{O})$. Define the function

$$\omega : \mathcal{O} \to \mathbb{R}, \qquad \omega(x) := \begin{cases} \frac{\Phi(F(x)) - \Phi(J(x))}{F(x) - J(x)} & \text{if } x \in \{F \ne J\} \\ \Phi'(F(x)) & \text{if } x \in \{F = J\}. \end{cases}$$

Then we have, using also (3.57), that

$$\begin{aligned}\|\Phi \circ F - \Phi \circ J\|_{(H^\kappa)^*} &= \sup_{\varphi \in C^\infty(\mathcal{O}),\, \|\varphi\|_{H^\kappa} \le 1} \left| \int_{\mathcal{O}} \varphi (\Phi \circ F - \Phi \circ J) \mathbb{1}_{\{F \ne J\}} \right| \\ &= \sup_{\varphi \in C^\infty(\mathcal{O}),\, \|\varphi\|_{H^\kappa} \le 1} \left| \int_{\mathcal{O}} (F - J) \varphi \omega \right| \\ &\le \|F - J\|_{(H^\kappa)^*} \sup_{\varphi \in C^\infty(\mathcal{O}),\, \|\varphi\|_{H^\kappa} \le 1} \|\varphi \omega\|_{H^k} \\ &\lesssim \|F - J\|_{(H^\kappa)^*} \|\omega\|_{C^\kappa}.\end{aligned}$$

2. Thus it suffices to prove that $\|\omega\|_{C^\kappa} \leq C(1 + \|F\|_{C^\kappa}^\kappa \vee \|J\|_{C^\kappa}^\kappa)$ for some $C > 0$ independent of F and J. Writing $\omega = \psi \circ \phi$, where

$$\phi : \mathcal{O} \to \mathbb{R}^2, \qquad \phi(z) = (F(z), J(z)),$$

$$\psi : \mathbb{R}^2 \to (0,\infty), \qquad \psi(x,y) = \begin{cases} \frac{\Phi(x)-\Phi(y)}{x-y} & \text{if } x \neq y \\ \Phi'(x) & \text{if } x = y, \end{cases}$$

we see by the multivariate chain rule that it suffices to show that ψ is κ-times continuously differentiable with bounded derivatives, and we achieve this by showing that the partial derivatives of ψ of order κ exist and are continuous throughout $\mathbb{R}^2$.

3. We will repeatedly use the following basic fact: Let $h : \mathbb{R} \to \mathbb{R}$ be continuous and continuously differentiable on $\mathbb{R} \setminus \{0\}$. If h' has a continuous extension g to $\mathbb{R}$ with some value $g(0) = \xi$, then $h \in C^1(\mathbb{R})$ with $h'(0) = \xi$.

4. Clearly, ψ is smooth on $\mathbb{R}^2 \setminus \{x = y\}$. For $k \geq 0$ and $x, y \in \mathbb{R}$, we denote the remainder of the k-th order Taylor expansion by

$$R_{k,x}(y) := \Phi(y) - \sum_{j=0}^{k} \frac{\Phi^{(j)}(x)}{j!}(y-x)^j.$$

For $x \neq y$, we have $\psi(x,y) = \frac{R_{0,x}(y)}{y-x}$ and also, by induction

$$\partial_1^k \psi(x,y) = \frac{k! R_{k,x}(y)}{(y-x)^{k+1}}, \quad k \geq 0, \tag{3.103}$$

where ∂_1 denotes the partial derivative with respect to x. By the mean value form of the remainder, we know that $R_{k,x}(y) = \frac{\Phi^{(k+1)}(\xi)}{(k+1)!}(y-x)^{k+1}$ for some ξ between x and y. Thus we can continuously extend $\partial_1^k \psi$ to $\{x = y\}$ by

$$\partial_1^k \psi(x,x) = \frac{\Phi^{(k+1)}(x)}{k+1}.$$

It follows that the partial derivatives with respect to x of all orders exist and are continuous on $\mathbb{R}^2$. The same holds for the partial derivatives with respect to y, by symmetry, concluding the proof of the case $\kappa = 1$. The case $\kappa = 2$ follows by adapting the previous arguments for mixed partial derivative $\partial_1 \partial_2$ and is left to the reader. $\square$

3.7 Proof of Theorem 3.2.2, Part 1

Let $\lambda, \varepsilon > 0$ be fixed throughout and let us write $\mathscr{J} = \mathscr{J}_{\lambda,\varepsilon}$. We denote by $\mathcal{T}_w = \mathcal{T}_{w,\alpha}$ the weak topology on $\mathcal{H}$ (recall $\mathcal{H} = H^\alpha(\mathcal{O})$ if $\kappa < 1/2$ and $\mathcal{H} = H_c^\alpha(\mathcal{O})$ if $\kappa \geq 1/2$), i.e. the coarsest topology with respect to which all bounded linear functionals $L : \mathcal{H} \to \mathbb{R}$ are continuous. We also denote the subspace topology on subsets of $\mathcal{H}$ by $\mathcal{T}_w$. On any closed ball $\mathcal{H}(R) := \{F \in \mathcal{H} : \|F\|_{H^\alpha} \leq R\}$, this topology is metrisable by some metric d, see e.g. Theorem 2.6.23 in [122].

Step 1: Localisation In Lemma 3.4.2, by assumption on α, we have that $\Psi_*(\lambda, R)/R^2 \xrightarrow{R\to\infty} 0$ and so there exists $\delta > 0$ such that for all $R \geq \delta$, we have that $R^2 \geq c_1 \varepsilon \Psi_*(\lambda, R)$, where c_1 is the constant from (3.36). Thus, applying Theorem 3.4.1, we have that the events

$$A_j := \left\{ \mathscr{J} \text{ has a maximizer } \hat{F} \notin \mathcal{V} \cap \mathcal{H}(2^j) \right\}$$

satisfy $\mathbb{P}(A_j) \xrightarrow{j\to\infty} 0$, whence choosing $j \in \mathbb{N}$ large enough ensures that

$$\sup_{F \in \mathcal{V} \cap H^\alpha(2^j)} \mathscr{J}(F) = \sup_{F \in \mathcal{V}} \mathscr{J}(F)$$

holds with probability as close to one as desired.

Step 2: Local existence via direct method By the previous step, it suffices to show that for any $j \in \mathbb{N}$, $\mathscr{J}$ almost surely has a maximizer over $\mathcal{V} \cap \mathcal{H}(2^j)$. We fix some $j \in \mathbb{N}$. As $\mathcal{V}$ is weakly closed and $\mathcal{H}(2^j)$ is weakly sequentially compact by the Banach-Alaoglu Theorem, it follows that any sequence $F_n \in \mathcal{V} \cap \mathcal{H}(2^j)$ has a weakly convergent subsequence $F_n \to F$ with weak limit $F \in \mathcal{V} \cap \mathcal{H}(2^j)$. Moreover, we claim that $-\mathscr{J} : \mathcal{V} \cap \mathcal{H}(2^j) \to \mathbb{R}$ is lower semicontinuous with respect to $\mathcal{T}_w$. To see this, we decompose $-\mathscr{J}$ as

$$-\mathscr{J}(F) = -2\langle Y, \mathscr{G}(F)\rangle_{\mathbb{H}} + \|\mathscr{G}(F)\|_{\mathbb{H}}^2 + \lambda^2 \|F\|_{H^\alpha}^2 =: I + II + III.$$

The term I is, almost surely under $\mathbb{P}_{F_0}^{\varepsilon}$, continuous w.r.t. $\mathcal{T}_w$ by Lemma 3.7.3, II is continuous w.r.t. $\mathcal{T}_w$ by Lemma 3.7.2 and III is lower semicontinuous by a standard fact from functional analysis. Thus the existence of minimisers follows from the direct method of the calculus of variations.

The next three lemmas are needed to prove lower semicontinuity of $-\mathscr{J}$.

Lemma 3.7.1. *Let $\alpha > 0$ and let $(F_n : n \in \mathbb{N}) \subseteq \mathcal{H}$, for $\mathcal{H} = H^\alpha$ or H_c^α, be a sequence such that $F_n \to F$ for $\mathcal{T}_w$. Then also $F_n \to F$ in L^2.*

Proof. It suffices to show that for any subsequence $(F_{n_j} : j \in \mathbb{N})$, there exists a further subsequence $(F_{n_{j'}} : j' \in \mathbb{N})$ such that $F_{n_{j'}} \to F$ in L^2. By the uniform boundedness principle, there exists $R > 0$ such that for all $n \in \mathbb{N}$, $\|F_n\|_{H^\alpha} \leq R$. By the Rellich-Kondrashov compactness theorem, the closed ball $\mathcal{H}(R)$ is pre-compact with respect to L^2 topology, hence for any subsequence (F_{n_j}) of (F_n), there exists a further convergent subsequence $(F_{n_{j'}})$ with limit $\tilde{F}$ in L^2. In particular, we have $F_n \to F$ weakly in L^2 and $F_{n_{j'}} \to \tilde{F}$ in L^2, so that by the uniqueness of weak limits, we have $\tilde{F} = F$ as elements in L^2, and therefore $\tilde{F} = F$ a.e. in $\mathcal{O}$ and $F_n \to F$ in L^2. □

Lemma 3.7.2. *Let $\alpha > 0$, $\kappa, \gamma \in \mathbb{R}_+$ and $\mathcal{V}_0 \subseteq \mathcal{V}$ be a bounded subset of $\mathcal{H} = H^\alpha$ or H^α_c. If a map $\mathscr{G} : \mathcal{V} \to \mathbb{H}$ is (κ, γ, α)-regular, then it is continuous as a mapping from $(\mathcal{V}_0, d)$ to $\mathbb{H}$.*

Proof. Take any $F_n, F \in \mathcal{V}_0$ such that $F_n \to F$ for $\mathcal{T}_w$ and note that $\|F_n\|_{H^\alpha} \leq R$ for some $R > 0$. By Lemma 3.7.1 we have $\|F_n - F\|_{L^2} \to 0$ and by (3.8) and the continuous imbedding $L^2 \subseteq (H^\kappa)^*, \kappa \geq 0$, we obtain

$$\|\mathscr{G}(F_n) - \mathscr{G}(F)\|_{\mathbb{H}} \leq C\left(1 + R^\gamma\right)\|F_n - F\|_{L^2} \xrightarrow{n\to\infty} 0. \tag{3.104}$$

□

We finally establish a continuity result for the Gaussian process $Y^{(\varepsilon)}$.

Lemma 3.7.3. *Suppose that $Y^{(\varepsilon)}$ and $\mathscr{G}$ are as in Theorem 3.2.2. Then there exists a version of the Gaussian white noise process $\mathbb{W}$ in $\mathbb{H}$ such that for all $R > 0$, the map (between metric spaces)*

$$\Psi : (\mathcal{V} \cap \mathcal{H}(R), d) \to \mathbb{R}, \qquad F \mapsto \langle Y^{(\varepsilon)}, \mathscr{G}(F)\rangle_{\mathbb{H}}$$

is almost surely uniformly continuous.

Proof. For any $\delta > 0$, define the modulus of continuity

$$M_\delta := \sup_{F,H \in \mathcal{V}\cap\mathcal{H}(R),\ d(F,H)\leq\delta} \left|\langle Y^{(\varepsilon)}, \mathscr{G}(F) - \mathscr{G}(H)\rangle_{\mathbb{H}}\right|,$$

a random variable. Moreover, we define the set

$$A := \left\{\omega \in \Omega \;\middle|\; M_\delta \xrightarrow{\delta\to 0} 0\right\},$$

where Ω is a probability space supporting the law $\mathbb{P}$ of $\mathbb{W}$. It is sufficient to show that $\mathbb{P}(A) = 1$, and noting that M_δ is decreasing in δ, it hence suffices to prove $\mathbb{E}\left[M_\delta\right] \xrightarrow{\delta\to 0} 0$. To see this, similarly to the proof of Lemma 3.4.1, we apply Dudley's theorem (see [72], Theorem 2.3.7) to the Gaussian process

$$\left(\mathbb{W}(\psi) :\ \psi \in \mathcal{D}_R\right), \qquad \mathcal{D}_R := \left\{\mathscr{G}(F) \mid F \in \mathcal{V} \cap \mathcal{H}(R)\right\}.$$

For any $\delta > 0$, define

$$R_\delta := \sup_{F,H \in \mathcal{V}\cap\mathcal{H}(R),\ d(F,H)\leq\delta} \|\mathscr{G}(F) - \mathscr{G}(H)\|_{\mathbb{H}}.$$

By Lemma 3.7.2, we know that $\mathscr{G}$ is continuous as a mapping from $(\mathcal{V} \cap \mathcal{H}(R), d)$ to $\mathbb{H}$. As $(\mathcal{V} \cap \mathcal{H}(R), d)$ is a compact metric space, $\mathscr{G}$ is in fact uniformly continuous, so we have that $R_\delta \xrightarrow{\delta\to 0} 0$. By the same argument as in the proof of Lemma 3.4.2 (but choosing here $m := (1 + R^\gamma))$ we can use (3.8) to obtain

$$H(\rho, \mathcal{D}_R, \|\cdot\|_{\mathbb{H}}) \lesssim \left(\frac{Rm}{\rho}\right)^{\frac{d}{(\alpha+\kappa)}}, \quad \rho > 0,$$

whence by Dudley's theorem, the modulus of continuity is controlled by

$$\mathbb{E}\left[M_\delta\right] \le \mathbb{E}\left[\sup_{\psi,\varphi\in\mathcal{D}_R,\, \|\psi-\varphi\|_{\mathbb{H}}\le R_\delta} |\langle \mathbb{W}, \psi - \varphi\rangle_{\mathbb{H}}|\right] \lesssim \int_0^{R_\delta} \left(\frac{Rm}{\rho}\right)^{\frac{d}{2(\alpha+\kappa)}} d\rho,$$

which converges to zero as $\delta \to 0$ since $\alpha > d/2 - \kappa$. $\square$

Chapter 4

The nonparametric LAN expansion for discretely observed diffusions

This chapter considers scalar reflected diffusion processes $(X_t : t \geq 0)$, where the unknown drift function b is modelled nonparametrically. We show that in the low frequency sampling case, when the sample consists of $(X_0, X_\Delta, ..., X_{n\Delta})$ for some fixed sampling distance $\Delta > 0$, the model satisfies the local asymptotic normality (LAN) property, assuming that b satisfies some mild regularity assumptions. This is established by using the connections of diffusion processes to elliptic and parabolic PDEs. The key tools used are regularity estimates for certain parabolic PDEs as well as a detailed analysis of the spectral properties of the elliptic differential operator related to $(X_t : t \geq 0)$.

4.1 Introduction

Consider a scalar diffusion, described by a stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + \sqrt{2}dW_t, \qquad t \geq 0,$$

where $(W_t : t \geq 0)$ is a standard Brownian motion and b is the unknown *drift function* that is to be estimated. We investigate the so-called *low frequency* observation scheme, where the data consists of states

$$X^{(n)} = (X_0, X_\Delta ..., X_{n\Delta}) \tag{4.1}$$

of one sample path of $(X_t : t \geq 0)$, where $\Delta > 0$ is the *fixed* time difference between measurements. To ensure ergodicity and to limit technical difficulties, we follow [75] and [134] and consider a version of the model where the diffusion takes values on $[0, 1]$ with reflection at the boundary points $\{0, 1\}$, see Section 4.2.1 for the precise definition.

The nonparametric estimation of the coefficients of a diffusion process has attracted a great deal of attention in the past. For the low-frequency sampling scheme (4.1), Gobet, Hoffmann and Reiss [75] determined the minimax rate of estimation for both the drift and diffusion coefficient and also devised a spectral estimation method which achieves this rate. Thereafter, Nickl and Söhl [134] proved that the Bayesian posterior distribution contracts at the minimax rate, giving a frequentist justification for the use of Bayesian methods. In other sampling schemes, various methods have been studied, see e.g. [86] for a frequentist approach, [76, 101, 174, 1, 133] for recent posterior consistency and contraction rate results for Bayesian methods as well as [142, 171] for MCMC methodology for the computation of the Bayesian posterior.

However, often one desires a more detailed understanding of the performance of both frequentist and Bayesian methods, e.g. by establishing semi-parametric efficiency bounds or by proving a nonparametric Bernstein-von Mises theorem (BvM), which would give a frequentist justification for the use of Bayesian credible sets as confidence sets (see [72], Chapter 7.3). Nonparametric BvMs have been explored in the papers [33, 34] and have recently been proven for a number of statistical inverse problems [131, 135, 125], by Nickl and co-authors. In a diffusion model with continuous observations $(X_t : t \leq T)$, Nickl and Ray [133] recently proved a nonparametric BvM for estimating the drift b.

To order to achieve such a detailed understanding, a key step lies in studying the local information geometry of the parameter space, which in terms of semiparametric efficiency theory (see e.g. [172], Chapter 25) involves finding the LAN expansion and the corresponding (Fisher) information operator. This in turn determines the Cramér-Rao lower bound for estimating a certain class of functionals of the parameter of interest. While in the Gaussian white noise model with direct observations, the LAN expansion of the log-likelihood ratio is exact and given by the Cameron-Martin theorem, in inverse problems proving the LAN property is often not straightforward.

In a finite-dimensional (parametric) model for multidimensional diffusions which are sampled at high frequency, where the sample consists of states

$$X^{(n)} = (X_0, X_{\Delta_n} ..., X_{n\Delta_n})$$

with asymptotics such that $\Delta_n \to 0$ and $n\Delta_n \to \infty$, the LAN property was shown by Gobet [74] by use of Malliavin calculus.

The main contribution of this paper is to prove that also with low frequency observations, the reflected diffusion model satisfies the LAN property, under mild regularity assumptions on the drift b. If the transition densities of the Markov chain $(X_{i\Delta} : i \in \mathbb{N})$ are denoted by $p_{\Delta,b}$, then the log-likelihood of the sample (4.1) is approximately equal to

$$\ell_b(X^{(n)}) \approx \sum_{i=1}^{n} \log p_{\Delta,b}(X_{(i-1)\Delta}, X_{i\Delta}),$$

from which one can see the necessity of two ingredients to show the LAN expansion (see also [153]):

- The first is a result on the differentiability of the transition densities $b \mapsto p_{\Delta,b}(x,y)$, which guarantees that we can form the second-order Taylor expansion of the log-likelihood in certain 'directions' $h/\sqrt{n}$ with sufficiently good control over the remainder. See Theorem 4.2.1 for the precise statement, where we importantly also obtain an explicit form for the first derivative A_b, the 'score operator'.

- The second main ingredient consists of two well known limit theorems, the central limit theorem for martingale difference sequences [29] and the ergodic theorem, which ensure the right limits for the first and second order terms in the Taylor expansion respectively.

In view of this, the main work done in this chapter lies in establishing the regularity needed for $p_{\Delta,b}(x,y)$, see Theorem 4.2.1 below. As there is no explicit formula for $p_{\Delta,b}(x,y)$ in terms of b, our approach relies on techniques from the theory of parabolic PDE and spectral theory. We use a PDE perturbation argument, based on the fact that the transition densities of a diffusion process can naturally be viewed as the fundamental solution to a related parabolic PDE.

The main difficulty in the proofs lies in the singular behaviour of $p_{t,b}(x,y)$ as (t,x) approaches $(0,y)$, which is why standard PDE results cannot be applied directly, but only in a regularised setting. Thus the arguments will first be carried out for any fixed regularisation parameter $\delta > 0$, where the analysis needs to be done carefully in order to ensure that the estimates obtained are uniform in $\delta > 0$ and hence still valid in the limit $\delta \to 0$.

In the context of a statistical inverse problem for the (elliptic) Schrödinger equation [131, 115], where the above singular behaviour is not present, PDE perturbation arguments have previously been used to linearize the log-likelihood.

We also remark that the use of more probabilistic proof techniques like in [74] would have been conceivable, too. However, we found the PDE approach employed here to be more naturally suited to dealing with boundary conditions, and it avoids dealing with pathwise properties of the diffusions by working with the transitions densities directly, which are ultimately the objects of interest for analyzing the likelihood.

Potential applications of the LAN expansion presented in Theorem 4.2.2 include the study of semiparametric efficiency for a certain class of functionals of b which is implicitly defined by the range of the 'information operator' $A_b^*A_b$ (where A_b is the score operator (4.9)), as well as an infinite-dimensional Bernstein-von-Mises theorem similar to [125, 131, 133, 135]. However, studying the properties of $A_b^*A_b$ needed for this poses a highly non-trivial challenge which still has to be overcome, see Section 4.2.4 for a more detailed discussion.

In Section 4.2, we state and prove the LAN expansion. Section 4.3 is devoted to proving Theorem 4.2.1. Finally, in Section 4.4, we derive the spectral properties of the differential operator $\mathcal{L}_b$ and the transition semigroup $(P_{t,b} : t \geq 0)$ needed throughout the proofs.

4.2 Main results

4.2.1 A reflected diffusion model

We shall work with boundary reflected diffusions on the interval $[0,1]$, following [75, 134]. Consider the stochastic process $(X_t : t \geq 0)$, whose evolution is described by the stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + \sqrt{2}dW_t \ + \ dK_t(X), \qquad X_t \in [0,1], \quad t \geq 0. \tag{4.2}$$

Here $(W_t : t \geq 0)$ is a standard Brownian motion, $(K_t(X) : t \geq 0)$ is a non-anticipative finite variation process that only changes when $X_t \in \{0,1\}$ and

$$b : [0,1] \to \mathbb{R}$$

is the unknown drift function. We note that $K(X)$, which accounts for the reflecting boundary behaviour, is part of a solution to (4.2) and is in fact given by the difference of the local times of X at 0 and 1.

For any integer $s \geq 0$, let $C^s = C^s((0,1))$ and $H^s = H^s((0,1))$ denote the spaces of s-times continuously differentiable functions and s-times weakly differentiable functions with L^2-derivatives, respectively, endowed with the usual norms. We also define the subspace

$$C_0^1 := \{f \in C^1 : f(0) = f(1) = 0\}.$$

We assume throughout that for some $B < \infty$, b lies in the C_0^1-ball

$$\Theta := \{f \in C_0^1 : \|f\|_{C^1} := \|f\|_\infty + \|f'\|_\infty \leq B\}. \tag{4.3}$$

This ensures the existence of a pathwise solution $(X_t : t \geq 0)$ to (4.2) which can be constructed by a reflection argument, see e.g. Section I.§23 in [70] or [134]. For some $\Delta > 0$, which we assume to be *fixed* throughout the paper, our sample consists of measurements $X^{(n)} = (X_0, X_\Delta ..., X_{n\Delta})$ of one sample path, with asymptotics $n \to \infty$.

The process $(X_t : t \geq 0)$ forms an ergodic Markov process with invariant distribution μ_b, whose Lebesgue density (which we also denote by μ_b) is identified by b via

$$\mu_b(x) = \frac{e^{\int_0^x b(y)dy}}{\int_0^1 e^{\int_0^u b(y)dy} du}, \quad x \in [0,1], \tag{4.4}$$

see e.g. Chapter 4 in [12]. Moreover, we denote the Lebesgue transition densities and the semigroup associated to $(X_t : t \geq 0)$ by $p_{t,b}$ and $P_{t,b}$ respectively:

$$p_{t,b} : [0,1]^2 \to \mathbb{R}, \quad p_{t,b}(x,y) = \mathbb{P}_x(X_t \in dy), \quad t > 0, \tag{4.5}$$

$$P_{t,b}f(x) = \mathbb{E}_x[f(X_t)] = \int_0^1 p_{t,b}(x,z)f(z)dz, \quad t > 0, \quad f \in L^2. \tag{4.6}$$

Here, by Proposition 9 in [134], the transition densities are well-defined as well as bounded above and below for each $t > 0$, so that (4.6) is well-defined, too.

Let $\mathbb{P}_b$ denote the law of $(X_{i\Delta} : i \geq 0)$ on $[0,1]^{\mathbb{N}}$. For ease of exposition, we assume throughout that $X_0 \sim \mu_b$ under $\mathbb{P}_b$, a common assumption (cf. [75, 134]) which we make due to the uniform spectral gap over $b \in \Theta$ guaranteed by Lemma 4.4.1 below, which yields exponentially fast convergence of X_t to μ_b. Then under any $\mathbb{P}_b$, $b \in \Theta$, the law of $X^{(n)}$ from (4.1) on $[0,1]^{n+1}$ is absolutely continuous with respect to the $n+1$ -dimensional Lebesgue measure, and the log-density, which also constitutes the *log-likelihood* (when viewed as a function of b), is given by

$$\log d\mathbb{P}_b(X^{(n)}) = \log \mu_b(X_0) + \sum_{i=1}^{n} \log p_{\Delta,b}(X_{(i-1)\Delta}, X_{i\Delta}). \tag{4.7}$$

We note that some of the above assumptions can be relaxed at the expense of further technicalities in the proofs: Firstly, the assumption $X_0 \sim \mu_b$ could be replaced by $X_0 \sim \pi_b$ (under $\mathbb{P}_b$), for any measures π_b with Lebesgue densities such that for all $b \in \Theta$, $\log d\pi_{\tilde{b}}(X_0) - \log d\pi_b(X_0) = o_{\mathbb{P}_b}(1)$ as $\|\tilde{b} - b\|_{H^1} \to 0$. Secondly, it is conceivable that the main Theorems 4.2.1 and 4.2.2 below can be generalized to all $b \in H^1$ and $h \in \{f \in H^1 : f(0) = f(1) = 0\}$, which we shall not pursue further here, however.

4.2.2 Differentiability of the transition densities

In order to prove the LAN property, we need to differentiate the log-likelihood (4.7) at any drift parameter $b \in \Theta$, and the following theorem shows that for any $x, y \in [0,1]$, maps of the form $b \mapsto p_{\Delta,b}(x,y)$ are infinitely differentiable in 'directions' $h \in C_0^1$ (and in fact, Fréchet differentiable). For $b, h \in C_0^1$, $\eta \in \mathbb{R}$ and $x, y \in [0,1]$, for convenience we introduce the notation

$$\Phi_{b,h,x,y} = \Phi : \mathbb{R} \to \mathbb{R}, \qquad \Phi(\eta) = p_{\Delta,b+\eta h}(x,y).$$

Theorem 4.2.1. *For all $b, h \in C_0^1$ and $x, y \in [0,1]$, $\Phi = \Phi_{b,h,x,y}$ is a smooth (in fact, real analytic) function on $\mathbb{R}$, and we have*

$$\Phi'(0) = \int_0^{\Delta} P_{\Delta-s,b}[h\partial_1 p_{s,b}(\cdot, y)](x)ds. \tag{4.8}$$

Moreover, for each integer $k \geq 1$, we have the following bound on the k-th derivative of Φ at 0:

$$\sup_{b\in\Theta} \sup_{h\in C_0^1, \|h\|_{H^1}\leq 1} \sup_{x,y\in[0,1]} \left|\Phi^{(k)}(0)\right| < \infty.$$

Section 4.3 is devoted to the proof of Theorem 4.2.1.

Heuristically speaking, the right hand side of (4.8) has the form of a solution to an inhomogeneous parabolic PDE (cf. Proposition 4.3.1), and this PDE perspective will be key in the proofs. However, one has to be careful with such an interpretation, as the singular 'source term' $h\partial_1 p_{b,t}(\cdot, y)$ does not fall within the scope of classical PDE theory. Therefore, the above intuition needs to be made rigorous via a regularisation argument, see Section 4.3.

4.2.3 LAN expansion

By Lemma 4.4.4, for each $b \in \Theta$, $p_{\Delta,b}(\cdot,\cdot)$ is bounded above and below. Hence by Theorem 4.2.1 and the chain rule, the score operator is given by

$$A_b: \; C_0^1([0,1]) \;\to\; L^2([0,1]\times[0,1]), \qquad A_b h(x,y) = \frac{[\Phi_{b,h,x,y}]'(0)}{p_{\Delta,b}(x,y)}. \tag{4.9}$$

For any $f, g \in L^2([0,1]\times[0,1])$, we also define the corresponding 'LAN inner product' and 'LAN norm' as follows:

$$\begin{aligned} \langle f,g\rangle_{L^2(p_{\Delta,b}\mu_b)} &:= \int_0^1\int_0^1 f(x,y)g(x,y)\mu_b(x)p_{\Delta,b}(x,y)dxdy, \\ \langle f,g\rangle_{LAN} &:= \langle A_b f, A_b g\rangle_{L^2(p_{\Delta,b}\mu_b)}, \qquad \|f\|_{LAN}^2 := \langle f,f\rangle_{LAN}. \end{aligned} \tag{4.10}$$

Here is our main result, the proof can be found in Section 4.2.5.

Theorem 4.2.2 (LAN expansion)**.** *For any $b, h \in C_0^1$, we have that*

$$\log \frac{d\mathbb{P}_{b+h/\sqrt{n}}}{d\mathbb{P}_b}(X^{(n)}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n A_b h(X_{(i-1)\Delta}, X_{i\Delta}) - \frac{1}{2}\|h\|_{LAN}^2 + o_{\mathbb{P}_b}(1) \tag{4.11}$$

as $n\to\infty$ and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n A_b h(X_{(i-1)\Delta}, X_{i\Delta}) \xrightarrow{n\to\infty}{}^{d} N(0, \|h\|_{LAN}^2). \tag{4.12}$$

Note that due to the nature of the non-i.i.d. Markov chain data at hand, A_b necessarily needs to map into a function space of two variables, as the overall log-likelihood cannot be formed as a sum of functions of single states of the chain, but only of increments of the chain.

4.2.4 Potential statistical applications of Theorem 4.2.2

The LAN expansion can be used to obtain semiparametric lower bounds for the estimation of certain linear functionals $L(b)$ for which there exists a Riesz representer $\Psi \in C_0^1$ such that $L(\cdot) = \langle \Psi, \cdot \rangle_{LAN}$, and can potentially further be used to prove a non-parametric Bernstein-von-Mises theorem.

To make this more precise, we define the 'information operator' (which generalizes the Fisher information) by $I_b := A_b^* A_b : C_0^1 \to L^2$, where A_b from (4.9) is viewed as a densely defined operator on L^2 with domain C_0^1 and A_b^* is the adjoint of A_b with respect to the inner products $\langle \cdot, \cdot \rangle_{L^2}$ and $\langle \cdot, \cdot \rangle_{L^2(p_{\Delta,b}\mu_b)}$. Then, for example, to study semiparametric Cramér-Rao lower bounds for functionals of the form $L(b) = \langle \psi, b \rangle_{L^2}$, $\psi \in L^2$, one needs that there is some $\Psi \in C_0^1$ such that

$$\forall w \in C_0^1 : \quad \langle \psi, w \rangle_{L^2} = \langle \Psi, w \rangle_{LAN} = \langle I_b \Psi, w \rangle_{L^2}.$$

Hence one needs that ψ lies in the range $R(I_b)$ of I_b (or at least of $R(A_b^*)$), see p.372-373 in [172] for a detailed discussion. Assuming the injectivity of I_b, the 'optimal asymptotic variance' for estimators of $L(b)$ is then given by

$$\|\Psi\|_{LAN}^2 = \langle A_b \Psi, A_b \Psi \rangle_{L^2(p_{\Delta,b}\mu_b)},$$

which may intuitively be understood as an 'inverse Fisher information $\langle \psi, I_b^{-1} \psi \rangle_{L^2}$', in analogy to the parametric setting.

When $R(I_b)$ is known to contain at least a 'nice' subspace of functions, e.g. C_c^∞, I_b can be inverted on that subspace, and if key mapping properties of I_b^{-1} are known, then along the lines of [131, 135, 133, 125], one can further try to prove a nonparametric BvM. This would assert the convergence of infinite-dimensional posterior distributions to a Gaussian limit measure $\mathcal{G}$ whose covariance is given by the LAN inner product via $\mathrm{Cov}[G(\psi_1), G(\psi_2)] = \langle \Psi_1, \Psi_2 \rangle_{LAN}$, cf. (28) in [131].

The identification of $R(I_b)$ in the present case of diffusions sampled at low frequency, as well as the study of mapping properties of I_b, remain challenging open problems.

4.2.5 Proof of the LAN expansion

We now give the proof of Theorem 4.2.2, assuming the validity of Theorem 4.2.1 which is proven in Section 4.3 below. Besides Theorem 4.2.1, the other key ingredient for Theorem 4.2.2 is the following CLT for martingale difference sequences. It is due to Brown (building on ideas of Billingsley and Lévy) and follows immediately from the special case $t = 1$ in Theorem 2 of [29].

Proposition 4.2.3 (cf. [29])**.** *Suppose* $(\Omega, \mathcal{F}, (\mathcal{F}_n : n \geq 0), \mathbb{P})$ *is a filtered probability space and let* $(M_n : n \in \mathbb{N})$ *be a* $\mathcal{F}_n$*-martingale with* $M_0 = 0$*. For* $n \geq 1$*, define the increments*

$Y_n := M_n - M_{n-1}$ and let

$$\sigma_n^2 := \mathbb{E}\left[Y_n^2 \middle| \mathcal{F}_{n-1}\right], \qquad V_n^2 := \sum_{i=1}^n \sigma_i^2, \qquad s_n^2 := \mathbb{E}[V_n^2].$$

Suppose that $V_n^2 s_n^{-2} \xrightarrow{n\to\infty} 1$ in probability and that for all $\epsilon > 0$,

$$\frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}\left[Y_i^2 \mathbb{1}\{|Y_i| \geq \epsilon s_n\}\right] \xrightarrow{n\to\infty} 0 \tag{4.13}$$

in probability. Then, as $n \to \infty$, we have

$$M_n/s_n \xrightarrow{d} \mathcal{N}(0,1).$$

Proof of Theorem 4.2.2. Fix $b, h \in C_0^1$. Due to the spectral gap of the generator $\mathcal{L}_b$ (see Lemma 4.4.1), the Markov chain $(X_{n\Delta} : n \in \mathbb{N})$ originating from the diffusion (4.2) with initial distribution $X_0 \sim \mu_b$, is stationary and geometrically ergodic – we will use this fact repeatedly.

For notational convenience, we write

$$f(\eta, x, y) = \log p_{\Delta, b+\eta h}(x, y), \qquad g(\eta, x, y) = p_{\Delta, b+\eta h}(x, y).$$

By Theorem 4.2.1, f is smooth in η on a neighbourhood of 0, and for some $C < \infty$, the second order Taylor remainder satisfies

$$R_f(\eta) := \sup_{x,y\in[0,1]} |f(\eta,x,y) - f(0,x,y) - \eta\partial_\eta f(0,x,y) - \frac{\eta^2}{2}\partial_\eta^2 f(0,x,y)| \leq C|\eta|^3. \tag{4.14}$$

Thus, Taylor-expanding the log-likelihood (4.7) in direction $h/\sqrt{n}$ yields that

$$\begin{aligned}
&\log \frac{d\mathbb{P}^n_{b+h/\sqrt{n}}}{d\mathbb{P}^n_b}(X_0, ..., X_{n_\Delta}) \\
&= \left(\log \mu_{b+h/\sqrt{n}}(X_0) - \log \mu_b(X_0)\right) + \frac{1}{\sqrt{n}}\sum_{i=1}^n \partial_\eta f(0, X_{(i-1)\Delta}, X_{i\Delta}) \\
&\quad + \frac{1}{2n}\sum_{i=1}^n \partial_\eta^2 f(0, X_{(i-1)\Delta}, X_{i\Delta}) + D_n \\
&=: A_n + B_n + C_n + D_n.
\end{aligned} \tag{4.15}$$

For the remainder term D_n, we immediately see from (4.14) that $|D_n| \leq nR_f(n^{-1/2})$, whence $D_n = o_{\mathbb{P}_b}(1)$ as $n \to \infty$.

For C_n, observe that the function $\partial_\eta^2 f(0,\cdot,\cdot)$ is bounded by Theorem 4.2.1, such that the almost sure ergodic theorem yields that

$$C_n \xrightarrow{n\to\infty} \frac{1}{2}\mathbb{E}_b[\partial_\eta^2 f(0,X_0,X_\Delta)] \quad \text{a.s.},$$

where $\mathbb{E}_b$ denotes the expectation with respect to $\mathbb{P}_b$. Moreover, we have

$$\partial_\eta^2 f(0,X_0,X_\Delta) = \frac{\partial_\eta^2 g(0,X_0,X_\Delta)}{g(0,X_0,X_\Delta)} - \left(\partial_\eta f(0,X_0,X_\Delta)\right)^2 =: I + II,$$

and by interchanging differentiation and integration (which is possible by Theorem 4.2.1), we see that

$$\mathbb{E}[I] = \int_0^1\int_0^1 \partial_\eta^2 g(0,x,y)\mu_b(x)dxdy = 0,$$

and hence $\mathbb{E}_b[\partial_\eta^2 f(0,X_0,X_\Delta)] = -\langle A_b h, A_b h\rangle_{L^2(p_{\Delta,b}\mu_b)} = -\|h\|_{LAN}^2$.

We next treat B_n. Let $(\mathcal{F}_n : n \geq 0)$ denote the natural filtration of $(X_{\Delta n} : n \geq 0)$. In view of Proposition 4.2.3, let us write

$$Y_n = \partial_\eta f(0,X_{(n-1)\Delta},X_{n\Delta}), \qquad M_n := \sqrt{n}B_n = \sum_{i=1}^n Y_n,$$

$$\sigma_n^2 = \mathbb{E}\left[Y_n^2 \middle| X_{(n-1)\Delta}\right], \qquad V_n^2 = \sum_{i=1}^n \sigma_i^2, \qquad s_n^2 = \mathbb{E}[V_n^2].$$

Then, using dominated convergence and the Markov property, we see that $M_0 = 0$ and that $(M_n : n \geq 0)$ is a martingale:

$$\begin{aligned}
\mathbb{E}[Y_n|\mathcal{F}_{n-1}] &= \int_0^1 \partial_\eta f(0,X_{(n-1)\Delta},y)p_{\Delta,b}(X_{(n-1)\Delta},y)dy \\
&= \int_0^1 \partial_\eta g(0,X_{(n-1)\Delta},y)dy \\
&= \partial_\eta \int_0^1 p_{\Delta,b+\eta h}(X_{(n-1)\Delta},y)dy \,\Big|_{\eta=0} = 0.
\end{aligned}$$

Moreover, we have that $\sigma_n^2 = \tilde{\sigma}^2(X_{(n-1)\Delta})$ for some bounded measurable function $\tilde{\sigma}^2 : [0,1] \to [0,\infty)$ and by the stationarity of $(X_{i\Delta} : i \geq 0)$, we have $s_n^2 = n\mathbb{E}_b[\tilde{\sigma}^2(X_0)] = n\|h\|_{LAN}^2$, whence the ergodic theorem yields that $\mathbb{P}_b$– a.s.,

$$V_n^2 s_n^{-2} = \frac{1}{n\|h\|_{LAN}^2}\sum_{i=1}^n \tilde{\sigma}^2(X_{(n-1)\Delta}) \xrightarrow{n\to\infty} \|h\|_{LAN}^{-2}\mathbb{E}_b[(\partial_\eta f(0,X_0,X_1))^2] = 1.$$

Lastly, as the Y_i's are bounded random variables, the condition (4.13) is fulfilled. Hence Proposition 4.2.3 yields that $B_n \to^d \mathcal{N}(0,\|h\|_{LAN}^2)$.

Finally, we observe that the term A_n in (4.15) from the invariant measure is of order $o_{\mathbb{P}_b}(1)$, as it can be bounded uniformly over b, h using (4.4):

$$
\begin{aligned}
&|\log \mu_{b+h/\sqrt{n}}(X_0) - \log \mu_b(X_0)| \lesssim \|\mu_{b+h/\sqrt{n}} - \mu_b\|_\infty \\
&\lesssim \Big\| \frac{e^{\int_0^\cdot (b+h/\sqrt{n})(y)dy}}{\int_0^1 e^{\int_0^x (b+h/\sqrt{n})(y)dy} dx} - \frac{e^{\int_0^\cdot b(y)dy}}{\int_0^1 e^{\int_0^x b(y)dy} dx} \Big\|_\infty \xrightarrow{n\to\infty} 0.
\end{aligned}
$$

□

4.3 Local approximation of transition densities

In this section, we study the differentiability properties of $p_{t,b}(x, y)$ as a function of the drift b, and the main goal is to prove Theorem 4.2.1. For technical reasons, we first prove a regularized version of it (Lemma 4.3.5 in Section 4.3.2) and then let the regularization parameter $\delta > 0$ tend to 0 to obtain Theorem 4.2.1 (Section 4.3.3).

4.3.1 Preliminaries and notation

We begin by introducing some notation and important classical results.

4.3.1.1 Some function spaces

For any integer $s \geq 0$, we equip the Sobolev space $H^s = H^s((0, 1))$ with the inner product

$$
\langle g_1, g_2 \rangle_{H^s} = \langle g_1, g_2 \rangle_{L^2} + \langle g_1^{(s)}, g_2^{(s)} \rangle_{L^2}, \tag{4.16}
$$

where L^2 is the usual space of square integrable functions with respect to Lebesgue measure. Occasionally it will be convenient to replace the L^2–inner product above by the $L^2(\mu_b)$–inner product, where μ_b is the invariant measure of $(X_t : t \geq 0)$, which by (4.24) induces a norm which is equivalent to the norm induced by (4.16).

We will also use the fractional Sobolev spaces H^s for real $s \geq 0$, which are obtained by interpolation, see [108]. For $s > \frac{1}{2}$, the Sobolev embedding (4.19) implies that any function $f \in H^s$ extends uniquely to a continuous function on $[0, 1]$. The following standard interpolation equalities and embeddings (see [108], p.44-45) will be used throughout. For all $s_1, s_2 \geq 0$ and $\theta \in (0, 1)$, we have

$$
\forall f \in H^{s_1} \cap H^{s_2} : \quad \|f\|_{H^{\theta s_1 + (1-\theta)s_2}} \leq C(\theta, s_1, s_2) \|f\|_{H^{s_1}}^\theta \|f\|_{H^{s_2}}^{1-\theta}, \tag{4.17}
$$

and for each $s > 1/2$, we have the multiplicative inequality

$$
\forall f, g \in H^s : \quad \|fg\|_{H^s} \lesssim C(s) \|f\|_{H^s} \|g\|_{H^s} \tag{4.18}
$$

as well as the continuous embedding

$$H^s \subseteq C([0,1]), \quad \|f\|_\infty \leq C(s)\|f\|_{H^s}, \tag{4.19}$$

where $C([0,1])$ denotes the space of continuous functions on $[0,1]$. Moreover, for any $s > 0$, we define the negative order Sobolev space H^{-s} as the topological dual space of H^s, where for any $f \in L^2$, the norm can be written as

$$\|f\|_{H^{-s}} = \sup_{\psi \in H^s, \|\psi\|_{H^s} \leq 1} |\int_0^1 f\psi|.$$

For any $T > 0$, any Banach space $(X, \|\cdot\|)$ and any integer $k \geq 0$, we denote by $C^k([0,T], X)$ the k-times continuously differentiable functions from $[0,T]$ to X, equipped with the norm

$$\|f\|_{C^k([0,T],X)} = \sum_{i=0}^{k} \sup_{t \in [0,T]} \|\frac{d^i}{dt^i} f(t)\|.$$

For $\alpha > 0$ with $\alpha \notin \mathbb{N}$, we denote the space of α-Hölder continuous functions $f : [0,T] \to X$ by $C^\alpha([0,T], X)$ and equip it with the usual norm

$$\|f\|_{C^\alpha([0,T],X)} = \|f\|_{C^{\lfloor \alpha \rfloor}([0,T],X)} + \sup_{s,t \in [0,T], s \neq t} \frac{\|\frac{d^{\lfloor \alpha \rfloor}}{dt^{\lfloor \alpha \rfloor}} f(t) - \frac{d^{\lfloor \alpha \rfloor}}{dt^{\lfloor \alpha \rfloor}} f(s)\|}{\left|t - s\right|^{\alpha - \lfloor \alpha \rfloor}}.$$

We will frequently, without further comment, interpret functions $f : [0,T] \times [0,1] \to \mathbb{R}$ as L^2-valued maps $f : [0,T] \to L^2, f(t) = f(t, \cdot)$, and vice versa.

4.3.1.2 The differential operator $\mathcal{L}_b$

For any drift function $b \in H^1$, we define the differential operator

$$\mathcal{L}_b f(x) := f''(x) + b(x)f'(x) \qquad \text{for } f \in \mathcal{D},$$
$$\mathcal{D} := \left\{ f \in H^2 : f'(0) = f'(1) = 0 \right\}.$$

It is well-known that $\mathcal{L}_b$ is the infinitesimal generator of the semigroup $(P_{t,b} : t \geq 0)$ defined in (4.6), so that we get by the usual functional calculus that $P_{t,b} = e^{t\mathcal{L}_b}$ for all $t \geq 0$ (with the convention $e^0 = Id$). The fact that the domain $\mathcal{D}$ of $\mathcal{L}_b$ is equipped with Neumann boundary conditions corresponds to the diffusion being reflected at the boundary, see [82] for a detailed discussion. We equip $\mathcal{D}$ with the graph norm

$$\|f\|_{\mathcal{L}_b} := \left(\|\mathcal{L}_b f\|^2_{L^2(\mu_b)} + \|f\|^2_{L^2(\mu_b)}\right)^{1/2},$$

which by Lemma 4.4.2 is equivalent to the H^2-norm on $\mathcal{D}$. Moreover, for $h \in H^1$, we define the first order differential operator

$$L_h f(x) = h(x) f'(x), \qquad f \in H^1. \tag{4.20}$$

The operator $\mathcal{L}_b$ has a purely discrete spectrum $\mathrm{Spec}(\mathcal{L}_b) \subseteq (-\infty, 0]$ (see [54], Theorem 7.2.2). We will denote by $(u_{j,b})_{j\geq 0}$ the $L^2(\mu_b)$-normalized orthogonal basis of $L^2(\mu_b)$ consisting of the eigenfunctions $u_{j,b} \in \mathcal{D}$ of $\mathcal{L}_b$, ordered such that the corresponding eigenvalues $(\lambda_{j,b})_{j\geq 0}$ are non-increasing. When there is no ambiguity, we will often simply write λ_j and u_j. We will use throughout the spectral decomposition

$$p_{t,b}(x,y) = \sum_{j\geq 0} e^{\lambda_j t} u_j(x) u_j(y) \mu_b(y), \quad x, y \in [0,1],\ t > 0, \tag{4.21}$$

see e.g. p. 101 in [9], and the spectral representations

$$\mathcal{L}_b f = \sum_{j\geq 0} \lambda_j \langle f, u_j \rangle_{L^2(\mu_b)} u_j, \quad f \in \mathcal{D}, \tag{4.22}$$

$$P_{t,b} f = \sum_{j\geq 0} e^{t\lambda_j} \langle f, u_j \rangle_{L^2(\mu_b)} u_j, \quad f \in L^2,\ t > 0. \tag{4.23}$$

We also note that (4.4) immediately yields that there exist constants $0 < C < C' < \infty$ such that for all $b \in \Theta$ and all $x \in [0,1]$,

$$C \leq \mu_b(x) \leq C'. \tag{4.24}$$

4.3.1.3 A key PDE result

For any $f \in C([0,T], L^2)$ and $u_0 \in \mathcal{D}$, consider the inhomogeneous parabolic equation

$$\begin{cases} \frac{d}{dt} u(t) = \mathcal{L}_b u(t) + f(t) & \text{for all } t \in [0,T], \\ u(0) = u_0. \end{cases} \tag{4.25}$$

We say that a function $u : [0,T] \to L^2$ is a solution to (4.25) if $u \in C^1([0,T], L^2) \cap C([0,T], \mathcal{D})$ and (4.25) holds. The next proposition regarding the existence, uniqueness and regularity properties of solutions to (4.25) will play a key role for the proofs in the rest of Section 4.3. To state the result, we need the following interpolation spaces $\mathcal{D}(\alpha), 0 \leq \alpha \leq 1$, between L^2 and $\mathcal{D}$:

$$\begin{aligned} \mathcal{D}(\alpha) &:= \{ f \in L^2 : \omega(t) := t^{-\alpha} \| P_{t,b} f - f \|_{L^2(\mu_b)} \text{ is bounded on } t \in [0,1] \}, \\ \|f\|_{\mathcal{D}(\alpha)} &:= \|f\|_{L^2(\mu_b)} + \sup_{t\in[0,1]} \omega(t). \end{aligned} \tag{4.26}$$

Proposition 4.3.1. *Suppose* $0 < \alpha < 1$*,* $f \in C^\alpha([0,T], L^2)$ *and* $u_0 \in \mathcal{D}$*. Then there exists a unique solution* u *to (4.25), given by the Bochner integral*

$$u(t) = P_{t,b}u_0 + \int_0^t P_{t-s,b}f(s)ds, \quad t \in [0,T]. \tag{4.27}$$

If also $f(0) + \mathcal{L}_b u_0 \in \mathcal{D}(\alpha)$*, then we have* $u \in C^{1+\alpha}([0,T], L^2) \cap C^\alpha([0,T], \mathcal{D})$ *and there exists* $C < \infty$ *so that for all such* f *and* u_0*,*

$$\begin{aligned} &\|u\|_{C^{1+\alpha}([0,T],L^2)} + \|u\|_{C^\alpha([0,T],\mathcal{D})} \\ &\qquad \le C\left(\|f\|_{C^\alpha([0,T],L^2)} + \|u_0\|_{\mathcal{L}_b} + \|f(0) + \mathcal{L}_b u_0\|_{\mathcal{D}(\alpha)}\right). \end{aligned}$$

Proof. This is a special case of Theorem 4.3.1 (iii) in [116] with $X = L^2(\mu_b)$ and $A = \mathcal{L}_b$, where we note that the integral formula (4.27) is given by Proposition 4.1.2 in the same reference. We also note that $\mathcal{D}(\alpha)$ coincides with the space $D_A(\alpha, \infty)$ from [116] with equivalent norms, see Proposition 2.2.4 in [116]. It therefore suffices to verify that the general theory for parabolic PDEs developed in [116] applies to our particular case. For that, we need to check that $(P_{t,b} : t \ge 0)$ is an analytic semigroup of operators on L^2 in the sense of [116], p.34, which requires the following.

1. For some $\theta \in (\pi/2, \pi)$ and $\omega \in \mathbb{R}$, the resolvent set $\rho(\mathcal{L}_b)$ of $\mathcal{L}_b$ contains the sector $S_{\theta,\omega} \subseteq \mathbb{C}$, where $S_{\theta,\omega}$ is defined by

$$S_{\theta,\omega} := \{\lambda \in \mathbb{C} : \lambda \ne \omega, |\arg(\lambda - \omega)| < \theta\}.$$

2. There exists some $M < \infty$ such that we have the resolvent estimate

$$\|R(\lambda, \mathcal{L}_b)\|_{L^2 \to L^2} \le M|\lambda - \omega|^{-1} \quad \forall \lambda \in S_{\theta,\omega}.$$

As $\mathcal{L}_b$ has a discrete spectrum contained in the non-positive half line, both of the above properties are easily checked with $\omega = 0$ and any $\theta \in (\frac{\pi}{2}, \pi)$. □

Definition 4.3.2 (Solution operator)**.** *In what follows, we denote by* $\mathcal{S} = \left(\frac{d}{dt} - \mathcal{L}_b\right)^{-1}$ *the linear solution operator which maps any* $f \in C^\alpha([0,T], L^2)$*,* $0 < \alpha < 1$*, to the unique solution* $u = \mathcal{S}(f)$ *of the parabolic problem*

$$\begin{cases} \left(\frac{d}{dt} - \mathcal{L}_b\right) u(t) = f(t), & t \in [0,T], \\ u(0) = 0. \end{cases} \tag{4.28}$$

4.3.2 Approximation of regularized transition densities

The main result of this section is Lemma 4.3.5, which can be viewed as a regularized version of Theorem 4.2.1. The main tools used to prove it are Proposition 4.3.1 as well as the spectral analysis of $\mathcal{L}_b$ and $P_{t,b}$ from Section 4.4.

In order to apply Proposition 4.3.1, we view the transition densities $p_{t,b}(x,y)$ as functions of the two variables $(t,x) \in [0,T] \times [0,1]$ with $y \in [0,1]$ fixed, where T is an arbitrary constant $T > \Delta > 0$ (with the convention that $p_{0,b}(\cdot, y)$ is the point mass at y). Due to the singular behaviour of $p_{t,b}(x,y)$ for $(t,x) \to (0,y)$, a regularisation argument is needed. For any $\delta > 0$ and $d \in C_0^1$, define the δ- regularized transition densities by

$$u_d^\delta : [0,\infty) \times [0,1] \to \mathbb{R}, \quad u_d^\delta(t,x) := P_{\delta,0}(p_{t,d}(x,\cdot))(y),$$

where $(P_{t,0} : t \geq 0)$ denotes the transition semigroup for $b = 0$, which corresponds to reflected Brownian motion.

4.3.2.1 Recursive definition of approximations

We now implicitly define the ‘candidate’ local approximations to u_d^δ as solutions to certain parabolic PDEs. To that end, we note that using (4.6), one easily checks that for all $t \geq 0$,

$$u_d^\delta(t) = P_{t,d}\varphi_\delta, \quad \text{where} \quad \varphi_\delta(x) := p_{\delta,0}(y,x). \tag{4.29}$$

Hence we can give the following crucial PDE interpretation to u_d^δ.

Lemma 4.3.3. *For any $d \in C_0^1$, we have that $u_d^\delta \in C^{3/2}([0,T], L^2) \cap C^{1/2}([0,T], \mathcal{D})$, and u_d^δ is the unique solution to the initial value problem*

$$\begin{cases} (\frac{d}{dt} - \mathcal{L}_d)u(t) = 0 \;\; \textit{for all } t \in [0,T], \\ u(0) = \varphi_\delta. \end{cases} \tag{4.30}$$

Proof. We check that Proposition 4.3.1 applies with $\alpha = 1/2$. For this, we need that $\varphi_\delta \in \mathcal{D}$ and that $\mathcal{L}_d\varphi_\delta \in \mathcal{D}(1/2)$. Using the spectral decomposition (4.21) and the fact that $\mu_b = \mathrm{Leb}([0,1])$ for $b = 0$, we see by differentiating under the sum that $\varphi_\delta \in \mathcal{D}$. This is possible by Lemma 4.4.1 and the dominated convergence theorem. The same argument yields that $\varphi_\delta \in H^3$. Thus, we have that $\mathcal{L}_d\varphi_\delta \in H^1$, which is a subset of $\mathcal{D}(1/2)$ by the second part of Lemma 4.4.5. □

We now recursively define the functions $R_k^\delta[h]$ and $v_k^\delta[h]$, $k \geq 0$. The norm estimates in Section 4.3.2.2 justify that they are the correct remainder and approximating terms, respectively, in the k-th order Taylor expansion of $\eta \mapsto u_{b+\eta h}^\delta$.

Definition 4.3.4. *Let $b, h \in C_0^1$ and $\delta > 0$.*

1. For $k = 0$, we define the '0-th order local approximation' of $\eta \mapsto u_{b+\eta h}^\delta$ at 0, and the remainder of this approximation, by

$$v_0^\delta[h] = v_0^\delta := u_b^\delta, \qquad R_0^\delta[h] = R_0^\delta := u_{b+h}^\delta - u_b^\delta.$$

2. For $k \geq 1$, we recursively define the functions $R_k^\delta[h] = R_k^\delta, v_k^\delta[h] = v_k^\delta \in C^{3/2}([0,T], L^2) \cap C^{1/2}([0,T], \mathcal{D})$ by

$$R_k^\delta[h] := \mathcal{S}(L_h R_{k-1}^\delta[h]), \quad v_k^\delta[h] := R_{k-1}^\delta[h] - R_k^\delta[h], \tag{4.31}$$

where $\mathcal{S}$ is the solution operator defined in (4.28) and L_h was defined in (4.20).

We should justify why the definition (4.31) is admissible, and we do so by induction. By Lemma 4.3.3, we have $R_0^\delta[h] \in C^{3/2}([0,T], L^2) \cap C^{1/2}([0,T], \mathcal{D})$. Hence, using the definition of $R_k^\delta[h]$ and Proposition 4.3.1 inductively, we obtain that for all $k \geq 1$, $L_h R_{k-1}^\delta[h] \in C^{1/2}([0,T], H^1)$ as well as $L_h R_{k-1}^\delta[h](0) = 0$, so that R_k^δ, v_k^δ have the stated regularity. Thus, (4.31) is well-defined.

By definition of $\mathcal{L}_b$ and (4.30), we see that $R_0^\delta[h]$ is the unique solution to

$$(\frac{d}{dt} - \mathcal{L}_b) R_0^\delta(t) = L_h u_{b+h}^\delta(t) \ \ \forall t \in [0,T] \quad \text{and} \quad R_0^\delta(0) = 0, \tag{4.32}$$

and (4.31) yields that

$$u_{b+h}^\delta = \sum_{i=0}^k v_i^\delta[h] + R_k^\delta[h] \qquad \forall b, h \in C^1, \quad k \geq 0. \tag{4.33}$$

The regularity estimates for $R_k^\delta[h]$ in the next section will justify that (4.33) is in fact the Taylor approximation for $\eta \mapsto u_{b+\eta h}^\delta$. Before proceeding to this, we need to check that the $v_k^\delta[h]$ are homogeneous of degree k in h, i.e. that

$$\forall h \in C_0^1 \ \forall \eta \in \mathbb{R}: \quad v_k^\delta[\eta h] = \eta^k v_k^\delta[h]. \tag{4.34}$$

This is again seen by induction. For $k = 0$, we have that $v_0^\delta[\eta h] = u_b^\delta = v_0^\delta[h]$, and if (4.34) holds for some $k \geq 0$, then we have that

$$v_{k+1}^\delta[\eta h] = \mathcal{S}(L_{\eta h} v_k^\delta[\eta h]) = \eta^{k+1} \mathcal{S}(L_h v_k^\delta[h]),$$

where we have used that for each $k \in \mathbb{N} \cup \{0\}$,

$$(\frac{d}{dt} - \mathcal{L}_b) v_{k+1}^\delta = L_h(R_{k-1}^\delta - R_k^\delta) = L_h v_k^\delta.$$

4.3.2.2 Regularity estimates

We now derive norm estimates for the remainders $R_k^\delta[h]$ from (4.31) and (4.33), using Propsition 4.3.1 and the results from Section 4.4.

The following Lemma is the main result of Section 4.3. It can be viewed as a regularised version of Theorem 4.2.1. Crucially, the estimate below is uniform in $\delta > 0$ such that it can be preserved in the limit $\delta \to 0$.

Lemma 4.3.5. *For each $\epsilon > 0$, there exists $C > 0$ such that for all $b \in \Theta$ from (4.3), $h \in C_0^1$ with $\|h\|_{H^1} \leq 1$, $y \in [0,1]$, $k \in \mathbb{N} \cup \{0\}$ and $\delta > 0$,*

$$\|R_k^\delta[h](\Delta)\|_\infty \leq C^k \|h\|_{H^1}^{k+1/2-\epsilon}.$$

The rest of this section is concerned with proving Lemma 4.3.5. In what follows, when we write that an inequality is 'uniform' without further comment, or when we use the symbols $\lesssim, \gtrsim, \simeq$, we mean that the constants involved can be chosen uniformly over b, h, y, k and δ as in the statement of Lemma 4.3.5.

The proof of Lemma 4.3.5 consists of two separate lemmas, which establish an L^2-estimate (4.38) and an H^1-estimate (4.41) for $R_k^\delta[h](\Delta)$ respectively. Given these two estimates, Lemma 4.3.5 then immediately follows from interpolating, and taking C to be the larger of the two constants from (4.38) and (4.41):

$$\|R_k[h](\Delta)\|_\infty \lesssim \|R_k[h](\Delta)\|_{H^{\frac{1}{2}+\varepsilon}} \lesssim \|R_k(\Delta)\|_{L^2}^{\frac{1}{2}-\varepsilon} \|R_k(\Delta)\|_{H^1}^{\frac{1}{2}+\varepsilon} \leq C^k \|h\|_{H^1}^{k+\frac{1}{2}-\epsilon}.$$

The L^2-estimate To obtain estimates which are uniform in $\delta > 0$, we 'regularise' R_k^δ further by integrating in time. For $k \geq 0$, define

$$Q_k^\delta[h] : [0,T] \to L^2, \quad Q_k^\delta[h](t) := \int_0^t R_k^\delta[h](s) ds.$$

Here is the L^2-estimate.

Lemma 4.3.6. *1. Let $b, h \in C_0^1$, $\delta > 0$ and recall the definition (4.28) of $\mathcal{S}$. Then we have that*

$$Q_0^\delta[h] = \mathcal{S}\big(L_h \int_0^\cdot u_{b+h}^\delta(s) ds\big), \tag{4.35}$$

and for $k \geq 1$, we have that

$$Q_k^\delta[h] = \mathcal{S}\big(L_h Q_{k-1}^\delta[h]\big). \tag{4.36}$$

2. For all $\alpha < 1/4$, there exists $C < \infty$ such that for all b, h, y, k, δ as in Lemma 4.3.5,

$$\|Q_k^\delta[h]\|_{C^{1+\alpha}([0,T],L^2)} + \|Q_k^\delta[h]\|_{C^\alpha([0,T],\mathcal{D})} \leq C^k \|h\|_\infty^{k+1}. \tag{4.37}$$

In particular, we have that

$$\|R_k^\delta[h]\|_{C^\alpha([0,T],L^2)} \leq C^k \|h\|_\infty^{k+1}. \tag{4.38}$$

Proof. We first show (4.35). Using Riemann sums to approximate the integrals below, the closedness of the operators $\mathcal{L}_b$ and L_h as well as (4.32), we obtain that

$$\begin{aligned}(\frac{d}{dt} - \mathcal{L}_b)Q_0^\delta = R_0^\delta(t) - \int_0^t \mathcal{L}_b R_0^\delta(s)ds = R_0^\delta(t) - R_0^\delta(0) - \int_0^t \mathcal{L}_b R_0^\delta(s)ds \\ = \int_0^t (\frac{d}{ds} - \mathcal{L}_b)R_0^\delta(s)ds = L_h \int_0^t u_{b+h}^\delta(s)ds.\end{aligned} \tag{4.39}$$

Moreover, we have $Q_0^\delta(0) = 0$ and $Q_0^\delta \in C^{3/2}([0,T],L^2) \cap C^{1/2}([0,T],\mathcal{D})$, so that (4.35) follows from Proposition 4.3.1. For $k \geq 1$, (4.36) is proved in the same manner.

Next, we prove (4.37) for $k = 0$. Let $\alpha < 1/4$, $\delta > 0$, $b \in \Theta$, $\|h\| \in C_0^1$ with $\|h\|_{H^1} \leq 1$, and let us write

$$f(t) = \partial_x\Big(\int_0^t u_{b+h}^\delta(s)ds\Big).$$

In view of (4.35) and Proposition 4.3.1, and noting that $hf(0) = 0$, it suffices to show that $\|f\|_{C^\alpha([0,T],L^2)} \leq C$ for some uniform constant C. For all $t < t' \in [0,T]$, we have by the definition of u_{b+h}^δ and Fubini's theorem that

$$\begin{aligned}\left[f(t') - f(t)\right](x) =& \partial_x \int_t^{t'} \int_0^1 p_{s,b+h}(x,z)\varphi_\delta(z)dzds \\ =& \partial_x \int_0^1 \Big(\int_t^{t'} p_{s,b+h}(x,z)ds\Big)\varphi_\delta(z)dz.\end{aligned}$$

For convenience, let us for now write μ for μ_{b+h} and $(\lambda_j, u_j)_{j\geq 0}$ for the eigenpairs of $\mathcal{L}_{b+h}$. Using the spectral decomposition (4.21) with $b+h$ in place of b and Fubini's theorem, integrating each summand separately yields that

$$\begin{aligned}\left[f(t') - f(t)\right](x) = (t'-t)\partial_x \int_0^1 \varphi_\delta(z)\mu(z)dz \\ + \partial_x \int_0^1 \sum_{j\geq 1} \frac{1}{\lambda_j}(e^{t'\lambda_j} - e^{t\lambda_j})u_j(x)u_j(z)\varphi_\delta(z)\mu(z)dz \\ = \partial_x \sum_{j\geq 1} \frac{1}{\lambda_j}(e^{t'\lambda_j} - e^{t\lambda_j})u_j(x)\langle u_j, \varphi_\delta\rangle_{L^2(\mu)},\end{aligned} \tag{4.40}$$

where Fubini's theorem is applicable due to Lemma 4.4.1:

$$\sum_{j\geq 1} \Big|\frac{1}{\lambda_j}(e^{t'\lambda_j} - e^{t\lambda_j})u_j(x)\langle u_j, \varphi_\delta\rangle_{L^2(\mu)}\Big| \lesssim \|\mu\varphi_\delta\|_{L^2} \sum_{j\geq 1} j^{-2}\|u_j\|_\infty \lesssim \sum_{j\geq 1} j^{-3/2+\varepsilon}.$$

From Lemma 4.4.3 and (4.40), it follows that

$$f(t') - f(t) = \partial_x \left(\mathcal{L}_{b+h}^{-1} \left(P_{t',b+h} - P_{t,b+h} \right) \varphi_\delta \right).$$

Using this, (4.61), the self-adjointness of $P_{t,b+h}$ with respect to $\langle \cdot, \cdot \rangle_{L^2(\mu)}$ and (4.65), we obtain that

$$\begin{aligned}
\|f(t') - f(t)\|_{L^2} &\leq \|\mathcal{L}_{b+h}^{-1} \left(P_{t',b+h} - P_{t,b+h} \right) \varphi_\delta\|_{H^1} \\
&\lesssim \|P_{t,b+h}(P_{t'-t,b+h} - Id)\varphi_\delta\|_{H^{-1}} \\
&\lesssim \sup_{\phi \in H^1, \|\phi\|_{H^1} \leq 1} \left| \langle (P_{t'-t,b+h} - Id) P_{t,b+h}\varphi_\delta, \phi \rangle_{L^2(\mu)} \right| \\
&= \sup_{\phi \in H^1, \|\phi\|_{H^1} \leq 1} \left| \langle P_{t,b+h}\varphi_\delta, P_{t'-t,b+h}\phi - \phi \rangle_{L^2(\mu)} \right| \\
&\lesssim \sup_{t>0} \|P_{t,b+h}\varphi_\delta\|_{L^1} \sup_{\phi \in H^1, \|\phi\|_{H^1} \leq 1} \|P_{t'-t,b+h}\phi - \phi\|_\infty \\
&\lesssim \sup_{\phi \in H^1, \|\phi\|_{H^1} \leq 1} \|P_{t'-t,b+h}\phi - \phi\|_\infty \\
&\lesssim (t'-t)^\alpha.
\end{aligned}$$

Hence, Proposition 4.3.1 and (4.35) imply (4.37) for $k = 0$. Choosing C large enough and inductively using Proposition 4.3.1 and (4.36), we also obtain (4.37) for $k \geq 1$:

$$\begin{aligned}
\|Q_k^\delta\|_{C^\alpha([0,T],\mathcal{D})} + \|Q_k^\delta\|_{C^{1+\alpha}([0,T],L^2)} &\leq C \|L_h Q_{k-1}^\delta\|_{C^\alpha([0,T],L^2)} \\
&\leq \|h\|_\infty \|Q_{k-1}^\delta\|_{C^\alpha([0,T],\mathcal{D})} \leq C^k \|h\|_\infty^{k+1}.
\end{aligned}$$

Finally, (4.38) follows upon differentiating (4.37) in t. □

The H^1 estimate The H^1-estimate reads as follows.

Lemma 4.3.7. *Let $k \geq 0$ be an integer and $\Delta > 0$. Then there exists $C < \infty$ such that for all b, h, y, k, δ as in Lemma 4.3.5,*

$$\|R_k^\delta(\Delta)\|_{H^1} \leq C^k \|h\|_{H^1}^k. \tag{4.41}$$

To prove Lemma 4.3.7, we express $R_k^\delta[h]$ using (4.27) and decompose the integral into times close to 0 and times bounded away from 0. The following Lemma allows us to control the respective integrals.

Lemma 4.3.8. *Let $T > 0$ and $0 < \eta < T$. Then there exists $C < \infty$ such that for all b, h, y, k, δ as in Lemma 4.3.5, the following estimates hold.*

1. For all $\tilde{T} \in [0,T)$, we have

$$\Big\| \int_0^{\tilde{T}} P_{T-s} L_h R_k^\delta(s) ds \Big\|_{H^1} \le \frac{C}{(T-\tilde{T})^{5/4}} \|h\|_{H^1} \sup_{s\in[0,\tilde{T}]} \|R_k^\delta(s)\|_{L^2}. \tag{4.42}$$

2. For all $\tilde{T} \in (0,T]$, we have

$$\Big\| \int_{\tilde{T}}^{T} P_{T-s} L_h R_k^\delta(s) ds \Big\|_{H^1} \le C\|h\|_\infty \sup_{s\in[\tilde{T},T]} \|R_k^\delta(s)\|_{H^1}. \tag{4.43}$$

Proof. We first show (4.43). By Lemma 4.4.2, we can estimate the $(-\mathcal{L}_b)^{1/2}$-graph norm instead of the H^1 norm. Using Lemma 4.4.1, we have

$$\begin{aligned}
&\Big\| (-\mathcal{L}_b)^{1/2} \int_{\tilde{T}}^{T} P_{T-s} L_h R_k^\delta(s) ds \Big\|^2_{L^2(\mu_b)} \\
&= \sum_{j=1}^{\infty} \Big(\int_{\tilde{T}}^{T} |\lambda_j|^{\frac12} e^{\lambda_j (T-s)} \langle u_j, h R_k^\delta(s)' \rangle_{L^2(\mu_b)} ds \Big)^2 \\
&\lesssim \sum_{j=1}^{\infty} \Big(\int_{\tilde{T}}^{T} j e^{-cj^2(T-s)} \|h R_k^\delta(s)'\|_{L^2} ds \Big)^2 \\
&\lesssim \|h\|_\infty^2 \sup_{s\in[\tilde{T},T]} \|R_k^\delta(s)\|_{H^1}^2 \sum_{j=1}^{\infty} \Big(j \int_{\tilde{T}}^{T} e^{-cj^2(T-s)} ds \Big)^2 \\
&\lesssim \|h\|_\infty^2 \sup_{s\in[\tilde{T},T]} \|R_k^\delta(s)\|_{H^1}^2 \sum_{j=1}^{\infty} \frac{1}{j^2}.
\end{aligned}$$

A similar calculation yields that

$$\Big\| \int_{\tilde{T}}^{T} P_{T-s} L_h R_k^\delta(s) ds \Big\|^2_{L^2(\mu_b)} \lesssim \|h\|_\infty^2 \sup_{s\in[\tilde{T},T]} \|R_k^\delta(s)\|_{H^1}^2 \Big(T^2 + \sum_{j=1}^{\infty} \frac{1}{j^4} \Big).$$

Combining the last two displays completes the proof of (4.43).

Next, we prove (4.42). Using (4.66) with $\alpha = 1$, the boundary condition $h(0) = h(1) = 0$ to integrate by parts and (4.17), we obtain

$$\begin{aligned}
\Big\| \int_0^{\tilde{T}} P_{T-s} L_h R_k^\delta(s) ds \Big\|_{H^1} &\lesssim \int_0^{\tilde{T}} (T-s)^{-\frac54} \Big\| L_h R_k^\delta(s) \Big\|_{H^{-1}} ds \\
&\le (T-\tilde{T})^{-\frac54} \int_0^{\tilde{T}} \sup_{\psi\in C^\infty, \|\psi\|_{H^1}\le 1} \Big| \int_0^1 (\psi h)' R_k^\delta(s) \Big| ds \\
&\lesssim (T-\tilde{T})^{-\frac54} \|h\|_{H^1} \sup_{s\in[0,\tilde{T}]} \Big\| R_k^\delta(s) \Big\|_{L^2}.
\end{aligned}$$

□

Proof of Lemma 4.3.7. The case $k=0$ follows from Lemma 4.4.4. For $k\geq 1$, we iteratively apply the estimates (4.42) and (4.43). We first define the points Δ_j at which we will split the integrals involved below:

$$\Delta_j := \Delta\frac{1+j/k}{2}, \quad j=0,...,k, \quad \text{and} \quad \eta_k := \frac{\Delta}{2k} = \Delta_k - \Delta_{k-1}.$$

Then, using (4.31) and (4.27), we can estimate

$$\begin{aligned}\|R_k^\delta(\Delta)\|_{H^1} &\leq \Big\| \int_0^{\Delta_{k-1}} P_{\Delta-s} L_h R_{k-1}^\delta(s)ds\Big\|_{H^1} + \Big\| \int_{\Delta_{k-1}}^{\Delta} P_{\Delta-s} L_h R_{k-1}^\delta(s)ds\Big\|_{H^1}\\ &=: I + II.\end{aligned}$$

Now let C be the largest of the constants from (4.38), (4.42) and (4.43). From (4.42) with $\tilde{T}=\Delta_{k-1}$ and (4.38), we obtain

$$I \leq C\eta_k^{-\frac{5}{4}} \|h\|_{H^1} \sup_{s\in[0,\Delta_{k-1}]} \Big\|R_{k-1}^\delta(s)\Big\|_{L^2} \leq C^k \eta_k^{-\frac{5}{4}} \|h\|_{H^1}^{k+1}.$$

For the second term, we apply (4.43) to obtain

$$II \leq C\|h\|_\infty \sup_{s\in[\Delta_{k-1},\Delta]} \|R_{k-1}^\delta(s)\|_{H^1}.$$

To further estimate the right hand side, we can repeat the argument for any $s\in[\Delta_{k-1},\Delta]$:

$$\begin{aligned}\|R_{k-1}^\delta(s)\|_{H^1} &\leq \Big\| \int_0^{\Delta_{k-2}} P_{\Delta-s} L_h R_{k-2}^\delta(s)ds\Big\|_{H^1} + \Big\| \int_{\Delta_{k-2}}^{s} P_{\Delta-u} L_h R_{k-2}^\delta(u)du\Big\|_{H^1}\\ &\leq C^k \eta_k^{-\frac{5}{4}} \|h\|_{H^1}^{k} + C\|h\|_\infty \sup_{s\in[\Delta_{k-2},\Delta]} \|R_{k-2}^\delta(s)\|_{H^1}.\end{aligned}$$

By iterating this argument k times, we obtain that for some larger constant $\tilde{C}$ independent of k,

$$\|R_k^\delta(\Delta)\|_{H^1} \leq kC^k(\frac{2k}{\Delta})^{\frac{5}{4}} \|h\|_{H^1}^{k+1} + C^k\|h\|_\infty^k \sup_{s\in[\Delta/2,\Delta]} \|R_0^\delta(s)\|_{H^1} \leq \tilde{C}^k \|h\|_{H^1}^k,$$

where we used (4.64) in the last step. This completes the proof. □

4.3.3 Proof of Theorem 4.2.1

We now prove Theorem 4.2.1 by letting $\delta>0$ in Lemma 4.3.5 tend to 0. Let us fix $b\in\Theta$, $h\in C_0^1$ with $\|h\|_{H^1}\leq 1$ and $x,y\in[0,1]$, and recall the notation $\Phi(\eta) := p_{\Delta,b+\eta h}(x,y)$ for

$\eta \in \mathbb{R}$. For notational convenience, for any $\delta > 0, \eta \in \mathbb{R}$ and integer $k \geq 0$, define

$$\Phi^\delta(\eta) := u^\delta_{b+\eta h}(\Delta, x), \quad a^\delta_k := v^\delta_k[h](\Delta, x), \quad p^\delta_k(\eta) := \sum_{i=0}^{k} a^\delta_i \eta^i.$$

Then by Lemma 4.3.5 and (4.34), there exists $C < \infty$ such that for all $\delta > 0$, $k \geq 0$ and $\eta \in [-1, 1]$,

$$\left|\Phi^\delta(\eta) - p^\delta_k(\eta)\right| = \left|R^\delta_k[\eta h](\Delta, x)\right| \leq \|R^\delta_k[\eta h](\Delta)\|_\infty \leq C^k |\eta|^{k+1/4}. \tag{4.44}$$

Hence for all $\delta > 0$, on the interval $\eta \in [-\frac{1}{2C}, \frac{1}{2C}] \cap [-1, 1]$, Φ^δ is given by the power series $\Phi^\delta(\eta) = \sum_{i=0}^\infty a^\delta_i \eta^i$. We divide the rest of the proof into three steps. The first two steps imply an analogous power series for Φ, and the third proves the integral formula (4.8).

1. Convergence of $\Phi^\delta(\eta)$. Note that by the definition of $u^\delta_{b+\eta h}$, we have that

$$\forall \eta \in \mathbb{R}: \quad \left|\Phi^\delta(\eta) - \Phi(\eta)\right| = \left|P_{\delta,0}\big(p_{\Delta,b+\eta h}(x, \cdot)\big)(y) - p_{\Delta,b+\eta h}(x, y)\right|.$$

Moreover, by (4.64) we have for any $R > 0$ that

$$\sup_{x \in [0,1], \|d\|_{H^1} \leq R} \|p_{\Delta,d}(x, \cdot)\|_{H^1} < \infty. \tag{4.45}$$

Thus, using (4.65), it follows that for any $\alpha < 1/4$, there is $c < \infty$ such that for all $b \in \Theta, h \in C^1_0$ with $\|h\|_{H^1} \leq 1$ and $|\eta| \leq 1$,

$$\begin{aligned} &\left|P_{\delta,0}\big(p_{\Delta,b+\eta h}(x, \cdot)\big)(y) - p_{\Delta,b+\eta h}(x, y)\right| \\ &\quad \leq \sup_{x \in [0,1], \|d\|_{H^1} \leq B+1} \|P_{\delta,0} p_{\Delta,d}(x, \cdot) - p_{\Delta,d}(x, \cdot)\|_\infty \leq c\delta^\alpha \xrightarrow{\delta \to 0} 0. \end{aligned} \tag{4.46}$$

2. Convergence of a^δ_k. Fix some $\eta \neq 0$ and some sequence $\delta_n > 0$ tending to 0 as $n \to \infty$. Using (4.44), it is easily seen inductively that for all $k \geq 0$, the sequence $(a^{\delta_n}_k : n \in \mathbb{N})$ is bounded. Hence, by a diagonal argument there exists a subsequence $(\delta_{n_l} : l \in \mathbb{N})$ and some sequence $a_k \in \mathbb{R}$ such that for all k, $a^{\delta_{n_l}}_k \xrightarrow{l \to \infty} a_k$. Defining the polynomials

$$p_k(\eta) := \sum_{i=0}^{k} a_i \eta^i, \qquad \eta \in \mathbb{R}, \quad k = 0, 1, 2, ..., \tag{4.47}$$

we see that (4.44) still holds with Φ and p_k in place of Φ^δ and p^δ_k. Hence, Φ is analytic and $\Phi(\eta) = \sum_{k=0}^\infty a_i \eta^i$ holds for $\eta \in \left[-\frac{1}{2C}, \frac{1}{2C}\right] \cap [-1, 1]$.

3. Proof of (4.8). It remains to show the integral formula (4.8) for $\Phi'(0)$. By what precedes, we know that the constants a_0, a_1 from (4.47) satisfy

$$\forall \eta \in [-1,1] : |\Phi(\eta) - a_0 - \eta a_1| \leq C|\eta|^{5/4}, \quad \Phi(0) = a_0, \quad \Phi'(0) = a_1 = \lim_{\delta \to 0} a_1^\delta.$$

Moreover, by definition of $v_1^\delta[h]$, we have for all $\delta > 0$ that

$$a_1^\delta = v_1^\delta[h](\Delta, x) = \mathcal{S}(L_h u_b^\delta)(\Delta, x) = \int_0^\Delta \left[P_{\Delta-s,b} L_h u_b^\delta(s)\right](x) ds.$$

Therefore, (4.8) is proven if we can show that the following expression converges to 0 as $\delta \to 0$ (recall that φ_δ was defined in (4.29)):

$$\begin{aligned}
&\int_0^\Delta [P_{\Delta-s,b} L_h P_{s,b}\varphi_\delta](x) ds - \int_0^\Delta [P_{\Delta-s,b} L_h p_{s,b}(\cdot, y)](x) ds \\
&= \int_0^\Delta \int_0^1 p_{\Delta-s,b}(x,z) h(z) \partial_z \Big(\int_0^1 p_{s,b}(z,u)\varphi_\delta(u) du - p_{s,b}(z,y) \Big) dz ds \\
&= - \int_0^{\Delta/2} \int_0^1 \partial_z [p_{\Delta-s,b}(x,z) h(z)] \Big(\int_0^1 p_{s,b}(z,u)\varphi_\delta(u) du - p_{s,b}(z,y) \Big) dz ds \\
&\quad + \int_{\Delta/2}^\Delta \int_0^1 p_{\Delta-s,b}(x,z) h(z) \partial_z \Big(\int_0^1 p_{s,b}(z,u)\varphi_\delta(u) du - p_{s,b}(z,y) \Big) dz ds \\
&=: I + II.
\end{aligned}$$

Here we have integrated by parts and used that the boundary terms vanish due to $h(0) = h(1) = 0$. For the term I, by arguing as in (4.45)-(4.46) (with s and z in place of Δ and x), we have that

$$\forall s \in (0, \Delta/2] : \quad \sup_{z \in [0,1]} \Big| \int_0^1 p_{s,b}(z,u)\varphi_\delta(u) du - p_{s,b}(z,y) \Big| \xrightarrow{\delta \to 0} 0,$$

showing that the ds-integrand in I tends to 0 pointwise. By the heat kernel estimate (4.63) and (4.64), we can also bound the ds-integrand uniformly in δ by

$$\frac{2C}{\sqrt{s}} \|p_{\Delta-s,b}(x,\cdot) h\|_{H^1} \leq \frac{2C\|h\|_{H^1}}{\sqrt{s}} \sup_{x \in [0,1], s \in [\Delta/2, \Delta]} \|p_{s,b}(x,\cdot)\|_{H^1} < \infty,$$

where C is the constant from (4.63). Hence, we have by the dominated convergence theorem that $|I| \xrightarrow{\delta \to 0} 0$.

For II, we argue similarly. By Lemma 4.4.4, we have that

$$\sup_{s \in [\Delta/2, \Delta], z \in [0,1]} \|\partial_z p_{s,b}(z,\cdot)\|_{H^2} < \infty,$$

whence (4.65) yields that

$$\begin{aligned}
&\Big|\partial_z\Big(\int_0^1 p_{s,b}(z,u)\varphi_\delta(u)du - p_{s,b}(z,y)\Big)\Big| \\
&\quad = \Big|\int_0^1 \partial_z p_{s,b}(z,u)\varphi_\delta(u)du - \partial_z p_{s,b}(z,y)\Big| \\
&\quad \leq \big\|P_{\delta,0}(\partial_z p_{s,b}(z,\cdot)) - \partial_z p_{s,b}(z,\cdot)\big\|_\infty \xrightarrow{\delta\to 0} 0.
\end{aligned}$$

Moreover, the ds-integrand is bounded by (cf. Lemma 4.4.4)

$$\frac{2C\|h\|_\infty}{\sqrt{\Delta - s}} \sup_{s\in[\Delta/2,\Delta],z\in[0,1]} \|p_{s,b}(z,\cdot)\|_{H^1},$$

such that by dominated convergence, we have $|II| \xrightarrow{\delta\to 0} 0$.

4.4 Spectral analysis of $\mathcal{L}_b$ and $(P_{t,b} : t \geq 0)$

In this section, we collect some properties of the generator $\mathcal{L}_b$, the differential equation related to $\mathcal{L}_b$ and the transition semigroup $(P_{t,b} : t \geq 0)$ which are needed for the proofs of Section 4.3. Although some results can be obtained using well-known, more general theory, our proofs are based on more or less elementary arguments, using the spectral analysis of $\mathcal{L}_b$ in Section 4.4.1.

4.4.1 Bounds on eigenvalues and eigenfunctions of $\mathcal{L}_b$

The following lemma summarizes some key properties of the eigenparis (u_j, λ_j) of $\mathcal{L}_b$. Note that the estimate (4.49) is an improvement on the bound in Lemma 6.6 of [75], and that (4.49) moreover coincides with the intuition from the eigenvalue equation $\mathcal{L}_b u_j = \lambda_j u_j$ that "two derivatives of u_j correspond to one order of growth in λ_j".

Lemma 4.4.1. *Let $s \geq 1$ be an integer and $B > 0$.*

1. *Suppose $b \in H^s \cap C_0^1$. Then for all $j \geq 0$, we have $u_j \in H^{s+2}$.*

2. *There exist $0 < C' < C < \infty$ such that for all $b \in C_0^1$ with $\|b\|_\infty \leq B$,*

$$\forall j \geq 0, \qquad \lambda_j \in [-Cj^2, -C'j^2]. \tag{4.48}$$

   Moreover, we have $u_0 = 1, \lambda_0 = 0$.

3. *There exists $C < \infty$ such that for all $0 \leq \alpha \leq s+2$,*

$$\forall j \geq 1 : \sup_{b\in H^s\cap C_0^1:\|b\|_{H^s}\leq B} \|u_j\|_{H^\alpha} \leq C|\lambda_j|^{\frac{\alpha}{2}}. \tag{4.49}$$

In particular, we have $\|u_j\|_\infty \lesssim |\lambda_j|^{1/4+\epsilon}$ *for all* $\epsilon > 0$.

Proof. Using that $u_j \in \mathcal{D} \subseteq H^2$ and (4.18), we obtain that for all $j \geq 0$, $u_j'' = \lambda u_j - b u_j' \in H^1$. Differentiating this equation $s-1$ times and bootstrapping this argument yields that $u_j^{(s+1)} \in H^1$.

Next, we prove (4.48) by adapting arguments from Chapter 4 of [54]. The standard Laplacian $\mathcal{L}_0 = \Delta$ with domain $\mathcal{D}$ is a nonpositive operator, self-adjoint with respect to the L^2-inner product, with spectrum $\{-j^2\pi^2 : j = 0, 1, 2, ...\}$ and associated quadratic form

$$Q_0(f) = \langle f', f' \rangle_{L^2} \quad \text{for all } f \in Dom((-\mathcal{L}_0)^{1/2}) = H^1,$$

where the fact that $Dom((-\mathcal{L}_0)^{1/2}) = H^1$ is shown in Chapter 7 of [54]. Similarly, using (4.4) and integrating by parts using $f'(0) = f'(1) = 0$, we have that $\mathcal{L}_b$, with domain $\mathcal{D}$, is self-adjoint with respect to the $L^2(\mu_b)$-inner product, and that for any $f \in \mathcal{D}$, the associated quadratic form is given by

$$\begin{aligned} Q_b(f) = \langle -\mathcal{L}_b f, f \rangle_{L^2(\mu_b)} &= \int_0^1 f'^2 \mu_b dx + \int_0^1 f' f \mu_b' dx - \int_0^1 f' f b \mu_b dx \\ &= \langle f', f' \rangle_{L^2(\mu_b)}. \end{aligned} \tag{4.50}$$

For any finite-dimensional subspace $L \subseteq \mathcal{D}$, define

$$\lambda^{(0)}(L) := \inf_{f \in L, \|f\|_{L^2} \leq 1} -Q_0(f), \qquad \lambda^{(b)}(L) := \inf_{f \in L, \|f\|_{L^2(\mu_b)} \leq 1} -Q_b(f). \tag{4.51}$$

Then by Theorem 4.5.3 of [54], the eigenvalues of $\mathcal{L}_0$ and $\mathcal{L}_b$ are given by

$$\lambda_j^{(0)} = \sup_{L \subseteq \mathcal{D}, \dim L \leq j} \lambda^{(0)}(L) = -j^2\pi^2, \qquad \lambda_j^{(b)} = \sup_{L \subseteq \mathcal{D}, \dim L \leq j} \lambda^{(b)}(L) \tag{4.52}$$

respectively. This, combined with (4.50) and (4.24), yields (4.48).

We now prove (4.49). Iterating the equation $\mathcal{L}_b u_j = \lambda_j u_j$, we have

$$\begin{aligned} \lambda_j^2 u_j = \mathcal{L}_b^2 u_j &= (u_j'' + b u_j')'' + b(u_j'' + b u_j')' \\ &= u_j^{(4)} + b'' u_j' + 2b' u_j'' + b u_j''' + b u_j''' + b b' u_j' + b^2 u_j''. \end{aligned}$$

Note that in each summand above, except for the first one, the sum of the orders of all derivatives is at most 3. This generalizes to $n \geq 3$, in that there exist polynomials $P_{n,m}$ such that

$$\lambda_j^n u_j = \mathcal{L}_b^n u_j = u_j^{(2n)} + \sum_{m=1}^{2n-1} P_{n,m}(b, b', ..., b^{(2n-2)}) u_j^{(m)}, \tag{4.53}$$

for which one can check the following properties by induction:

1. For all $n \geq 1$ and $m \leq 2n-1$, $P_{n,m}$ has degree at most n.

2. The only summand in (4.53) with factor $b^{(2n-2)}$ is $u_j' b^{(2n-2)}$.

For the odd order derivatives of u_j, there similarly exist polynomials $\tilde{P}_{n,m}$ of degree at most n such that

$$\begin{aligned} u_j^{(2n+1)} &= \Big(\mathcal{L}_b^n u_j - \sum_{m=1}^{2n-1} P_{n,m}(b, b', ..., b^{(2n-2)}) u_j^{(m)}\Big)' \\ &= \lambda_j^n u_j' - \sum_{m=1}^{2n} \tilde{P}_{n,m}(b, b', ..., b^{(2n-1)}) u_j^{(m)}, \end{aligned} \tag{4.54}$$

where the only summand containing the factor $b^{(2n-1)}$ is $u_j' b^{(2n-1)}$.

We now use these facts to show (4.49) by an induction argument, consisting of the base case and two induction steps.

Base Case $\alpha \leq 2$: To show (4.49) for all $\alpha \leq 2$, it suffices to prove the case $\alpha = 2$, as the case $\alpha \in (0, 2)$ then follows from $\|u_j\|_{L^2(\mu_b)} = 1$ and (4.17). We also note that the estimate for $\|u_j\|_\infty$ then follows by the Sobolev embedding (4.19). The case $\alpha = 2$ follows immediately from (4.57) and (4.48):

$$\|u_j\|_{H^2}^2 \simeq \|\mathcal{L}_b u_j\|_{L^2(\mu_b)}^2 + \|u_j\|_{L^2(\mu_b)}^2 = (\lambda_j^2 + 1)\|u_j\|_{L^2(\mu_b)}^2 = \lambda_j^2 + 1 \lesssim \lambda_j^2, \quad j \geq 1.$$

Induction step $2n \to 2n+1$: Assume that for some integer n, (4.49) holds for all $\alpha \leq 2n < s + 2$. Then, using (4.54), the Sobolev embedding $C^{2n-2} \subseteq H^s$ (note that $s \geq 2n - 1$) and the induction hypothesis, we obtain

$$\|u_j^{(2n+1)}\|_{L^2} \lesssim |\lambda_j|^n \|u_j'\|_{L^2} + \|b^{(2n-1)}\|_{L^2}\|u_j'\|_\infty + \|b\|_{C^{2n-2}}^n \|u_j\|_{H^{2n}} \lesssim |\lambda_j|^{n+\frac{1}{2}}.$$

The non-integer case $\alpha \in (2n, 2n+1)$ follows by interpolation.

Induction step $2n - 1 \to 2n$: Similarly, using (4.53), the embedding $C^{2n-3} \subseteq H^s$ (note that $s \geq 2n - 2$) and the induction hypothesis, we have

$$\|u_j^{(2n)}\|_{L^2} \lesssim |\lambda_j|^n + \|b^{(2n-2)}\|_{L^2}\|u_j'\|_\infty + \|b\|_{C^{2n-3}}^n \|u_j\|_{H^{2n-1}} \lesssim |\lambda_j|^n,$$

and the non-integer case $\alpha \in (2n - 1, 2n)$ again follows by interpolation. □

4.4.2 Characterisation of Sobolev norms in terms of (λ_j, u_j)

Using Lemma 4.4.1, we now prove that the graph norms of the non-negative self-adjoint operators $(-\mathcal{L}_b)^\theta$, $\theta \in \{0, \frac{1}{2}, 1\}$, on their respective domains, are equivalent to standard Sobolev norms. Let $\ell^2 = \ell^2(\mathbb{N} \cup \{0\})$ denote the usual space of square-summable sequences. For any Banach space $(X, \|\cdot\|_X)$ and linear operator $T : D \to X$ with domain $D \subseteq X$, we denote the graph norm of T by

$$\|x\|_T := (\|x\|_X^2 + \|Tx\|_X^2)^{1/2}, \qquad x \in D.$$

Lemma 4.4.2. *1. Let $\theta \in [0,1]$. Then for any $f \in L^2$, we have*

$$f \in Dom((-\mathcal{L}_b)^\theta) \iff \sum_{j=0}^{\infty}(1+|\lambda_j|^{2\theta})|\langle f, u_j\rangle_{L^2(\mu_b)}|^2 \ < \infty \tag{4.55}$$

and for any $f \in Dom((-\mathcal{L}_b)^\theta)$, we have

$$(-\mathcal{L}_b)^\theta f = \sum_{j=1}^{\infty}(-\lambda_j)^\theta \langle f, u_j\rangle_{L^2(\mu_b)} u_j. \tag{4.56}$$

2. There exists $0 < C < \infty$ such that for any $\theta \in \{0, \frac{1}{2}, 1\}$, we have

$$C^{-1}\|f\|_{H^{2\theta}} \leq \|f\|_{(-\mathcal{L}_b)^\theta} \leq C\|f\|_{H^{2\theta}}, \quad f \in Dom((-\mathcal{L}_b)^{\theta/2}). \tag{4.57}$$

3. There exists $0 < C < \infty$ such that for all $f \in L^2$,

$$C^{-1}\|f\|_{H^{-1}} \leq \Big\|\Big(\frac{\langle f, u_j\rangle_{L^2(\mu_b)}}{\sqrt{1+|\lambda_j|}} : j \geq 0\Big)\Big\|_{\ell^2} \leq C\|f\|_{H^{-1}} \tag{4.58}$$

Proof. 1. We first prove (4.55) for $\theta = 1$. Define the dense linear subspace

$$D := \bigcup_{n=0}^{\infty} \operatorname{span}\{u_j : j = 0, ..., n\} \subseteq L^2(\mu_b).$$

Then by Lemma 1.2.2 in [54], we know that the restriction of $\mathcal{L}_b$ to D, which we shall denote by $\mathcal{L}_b^D$, is an essentially self-adjoint operator on $L^2(\mu_b)$. Moreover, under the unitary operator

$$U : L^2(\mu_b) \to \ell^2, \quad f \mapsto (\langle f, u_j\rangle_{L^2(\mu_b)} : j \geq 0),$$

$\mathcal{L}_b^D$ is unitarily equivalent to the essentially self-adjoint multiplication operator $M^D : (a_j : j \geq 0) \mapsto (\lambda_j a_j : j \geq 0)$ on ℓ^2 with domain

$$U(D) = \{a \in \ell^2 : a_j = 0 \text{ for all } j \text{ large enough}\}.$$

Thus, the unique self-adjoint extentions of both operators (cf. [54], Theorem 1.2.7), which we denote by $\mathcal{L}_b$ and M, are also unitarily equivalent. Hence, for all $f \in L^2(\mu_b)$,

$$f \in \mathcal{D} \iff \sum_{j=0}^{\infty}(1+\lambda_j^2)|\langle f, u_j\rangle_{L^2(\mu_b)}|^2 < \infty$$

(The above condition defines the domain of the self-adjoint extension of M^D, see [54], Lemma 1.3.1), which proves (4.55) for $\theta = 1$. To see (4.55) for $\theta \in [0, 1)$, we note that the fractional power $(-\mathcal{L}_b)^\theta$ is unitarily equivalent to multiplication with $\big(|\lambda_j|^\theta : j \geq 0\big)$, and that $f \in \mathrm{Dom}\left((-\mathcal{L}_b)^\theta\right)$ iff

$$Uf \in \mathrm{Dom}(M^\theta) = \{f \in L^2 : \sum_{j=0}^{\infty}(1 + |\lambda_j|^{2\theta})|\langle f, u_j\rangle_{L^2(\mu_b)}|^2 \ < \infty\}.$$

2. We now show (4.57). For $\theta = 0$, there is nothing to prove. For $\theta = 1/2$, note that by Theorem 7.2.1 in [54] and (4.50), we have $\mathrm{Dom}\left((-\mathcal{L}_b)^{1/2}\right) = H^1$ and

$$\forall f \in H^1 : \|f\|^2_{\mathcal{L}_b^{1/2}} = \|f\|^2_{L^2(\mu_b)} + \langle \mathcal{L}_b^{1/2} f, \mathcal{L}_b^{1/2} f\rangle_{L^2(\mu_b)} = \|f\|^2_{H^1(\mu_b)}.$$

The case $\theta = 1/2$ now follows from (4.24). Finally, let $\theta = 1$. It is clear that $\|f\|^2_{\mathcal{L}_b} \lesssim \|f\|^2_{H^2}$, so that it remains to show $\|f\|^2_{H^2} \lesssim \|f\|^2_{\mathcal{L}_b}$. For this, we use Cauchy's inequality with ϵ to obtain that for some c_1,

$$\|\mathcal{L}_b f\|^2_{L^2} = \|f''\|^2_{L^2} + 2\langle f'', bf'\rangle_{L^2} + \|bf'\|^2_{L^2} \geq \frac{1}{2}\|f''\|^2_{L^2} - c_1\|f'\|^2_{L^2}.$$

Hence, integrating by parts and using Cauchy's inequality with ϵ again yields that for some c_2,

$$\begin{aligned}\|f''\|^2_{L^2} &\leq 2\|\mathcal{L}_b f\|^2_{L^2} + 2c_1\|f'\|^2_{L^2} \leq 2\|\mathcal{L}_b f\|^2_{L^2} + 2c_1\|f\|_{L^2}\|f''\|_{L^2} \\ &\leq 2\|\mathcal{L}_b f\|^2_{L^2} + c_2\|f\|^2_{L^2} + \frac{1}{2}\|f''\|^2_{L^2},\end{aligned}$$

proving that $\|f\|^2_{H^2} \lesssim \|f\|^2_{\mathcal{L}_b}$.

3. For any $f \in L^2$ and any test function $\psi \in H^1$, let us write $f_j = \langle f, u_j\rangle_{L^2(\mu_b)}$ and $\psi_j = \langle \psi, u_j\rangle_{L^2(\mu_b)}$, $j \geq 0$ respectively. Then by (4.56)-(4.57), we have

$$\begin{aligned}\|f\|_{H^{-1}} &\simeq \sup_{\psi\in H^1, \|\psi\|_{H^1}\leq 1} \left|\langle f, \psi\rangle_{L^2(\mu_b)}\right| = \sup_{\psi\in H^1, \|\psi\|_{H^1}\leq 1} \Big|\sum_{j=0}^{\infty} f_j\psi_j\Big| \\ &\simeq \sup_{\psi\in L^2, \|\psi\|_{L^2}\leq 1} \Big|\sum_{j=0}^{\infty} f_j\,(1 + |\lambda_j|)^{-1/2}\,\psi_j\Big| \\ &\simeq \left\|\left(f_j\,(1 + |\lambda_j|)^{-1/2} : j \geq 0\right)\right\|_{\ell^2}.\end{aligned} \tag{4.59}$$

□

4.4.3 Basic norm estimates for the one-dimensional Neumann problem

From the preceding Lemma, we can immediately derive some basic properties of the (elliptic) boundary value problem

$$\mathcal{L}_b u = f \quad \text{on } (0,1), \qquad u'(0) = u'(1) = 0 \tag{4.60}$$

needed in the proof of Lemma 4.3.6. Let us denote the orthogonal complement of the first eigenfunction $u_0 \equiv 1$ of $\mathcal{L}_b$ in $L^2(\mu_b)$ by

$$u_0^\perp = \Big\{ f \in L^2 : \int f d\mu_b = 0 \Big\}.$$

Lemma 4.4.3. *For every $f \in u_0^\perp$, there exists a unique function $u \in \mathcal{D} \cap u_0^\perp$ such that $\mathcal{L}_b u = f$, for which we use the notation $u = \mathcal{L}_b^{-1} f$. Moreover, for every $B > 0$ there exists $C < \infty$ such that for all $b \in C_0^1$ with $\|b\|_\infty \le B$ and $f \in u_0^\perp$,*

$$\|u\|_{H^s} \le C \|f\|_{H^{s-2}} \qquad \text{for } s \in \{0,1,2\}. \tag{4.61}$$

Proof. It follows immediately from the domain characterisation (4.55) and the spectral representation (4.22) that $\mathcal{L}_b$ is a one-to-one map from $\mathcal{D} \cap u_0^\perp$ to $L^2 \cap u_0^\perp$, and that $\mathcal{L}_b^{-1}$ is unitarily equivalent to multiplication by $(\lambda_j^{-1} \mathbb{1}_{j \ge 1} : j \ge 0)$ in the spectral domain, so that the $L^2 \to L^2$ norm of $\mathcal{L}_b^{-1}$ is finite. Hence, for $s = 2$, the estimate (4.61) follows from (4.57):

$$\Big\| \mathcal{L}_b^{-1} f \Big\|_{H^2}^2 \simeq \Big\| \mathcal{L}_b \mathcal{L}_b^{-1} f \Big\|_{L_2}^2 + \|\mathcal{L}_b^{-1} f\|_{L^2}^2 \simeq \|f\|_{L_2}^2.$$

The case $s = 0$ is obtained by duality. Using that $\mathcal{L}_b^{-1}$ is self-adjoint on $u_0^\perp$ and the previous case $s = 2$, we have that

$$\begin{aligned} \|\mathcal{L}_b^{-1} f\|_{L^2(\mu_b)} &= \sup_{\phi \in u_0^\perp, \|\phi\|_{L^2} \le 1} \Big| \int_0^1 \mathcal{L}_b^{-1} f \phi d\mu_b \Big| = \sup_{\phi \in u_0^\perp, \|\phi\|_{L^2} \le 1} \Big| \int_0^1 f \mathcal{L}_b^{-1} \phi d\mu_b \Big| \\ &\lesssim \|f\|_{H^{-2}}. \end{aligned}$$

Finally, for $s = 1$, Lemma 4.4.2 implies that

$$\Big\| \mathcal{L}_b^{-1} f \Big\|_{H^1}^2 \simeq \sum_{j=1}^\infty (1 + |\lambda_j|) \Big| \frac{\langle f, u_j \rangle_{L^2(\mu_b)}}{\lambda_j} \Big|^2 \lesssim \sum_{j=1}^\infty \frac{|\langle f, u_j \rangle_{L^2(\mu_b)}|^2}{1 + |\lambda_j|} \lesssim \|f\|_{H^{-1}}^2.$$

□

4.4.4 Estimates on $p_{t,b}(\cdot, \cdot)$ and $P_{t,b}$

Using Lemmata 4.4.1 and 4.4.2, we now collect some basic (partially well-known) results about the Lebesgue transition densities $p_{t,b}(\cdot, \cdot)$ (Lemma 4.4.4) and the semigroup $P_{t,b}$ (Lemma 4.4.5). Recall that they were defined in (4.5) and (4.6).

Lemma 4.4.4. *Let $s \geq 1$ be an integer, $t_0 > 0$ and $B > 0$. Then we have the following.*

1. *There exist constants $0 < C < C' < \infty$ such that for all $t \geq t_0$, $b \in C_0^1$ with $\|b\|_{C^1} \leq B$ and $x, y \in [0, 1]$,*
$$C \leq p_{t,b}(x, y) \leq C'. \tag{4.62}$$

2. *There exists $C < \infty$ such that for all $t \in (0, 1]$ and $b \in C_0^1$ with $\|b\|_\infty \leq B$,*
$$\|p_{t,b}(x, y)\|_\infty \leq Ct^{-\frac{1}{2}}, \quad x, y \in [0, 1]. \tag{4.63}$$

3. *For each $n \leq s + 2$, $m \leq s$ and $n', m' \leq s + 1$,*
$$\begin{aligned} &\sup_{t \geq t_0} \sup_{y \in [0,1]} \sup_{b \in C_0^1 \cap H^s : \|b\|_{H^s} \leq B} \|\partial_x^n \partial_y^m p_{t,b}(\cdot, y)\|_{L^2} < \infty \\ &\sup_{t \geq t_0} \sup_{x \in [0,1]} \sup_{b \in C_0^1 \cap H^s : \|b\|_{H^s} \leq B} \|\partial_x^{n'} \partial_y^{m'} p_{t,b}(x, \cdot)\|_{L^2} < \infty. \end{aligned} \tag{4.64}$$

Proof. For a proof of (4.62), we refer to Proposition 9 in [134] and for a proof of (4.63), we refer to Theorem 2.12 in [41]. Let us now prove the first part of (4.64); the second is obtained analogously. Let $n \leq s + 2$, $m \leq s$. Then (4.4) yields that

$$\sup_{\|b\|_{H^s} \leq B} \|\mu_b\|_{H^{s+1}} < \infty.$$

Using the multiplicative inequality (4.18), the spectral decomposition (4.21) and Lemma 4.4.1, we have

$$\begin{aligned} \|\partial_x^n \partial_y^m p_{t,b}(\cdot, y)\|_{L^2} &\leq \sum_{j=0}^\infty e^{t\lambda_j} \|u_j^{(n)}\|_{L^2} |\left(u_j \mu_b\right)^{(m)}(y)| \\ &\leq \sum_{j=0}^\infty e^{t_0\lambda_j} \|u_j^{(n)}\|_{L^2} \|(u_j \mu_b)^{(m)}\|_\infty \lesssim \sum_{j=0}^\infty e^{t_0\lambda_j} \|u_j\|_{H^{s+2}} \|u_j\|_{H^{s+1}} \|\mu_b\|_{H^{s+1}} \\ &\lesssim \sum_{j=0}^\infty e^{-cj^2} |\lambda_j|^{\frac{s+2}{2} + \frac{s+1}{2}} \lesssim \sum_{j=0}^\infty e^{-cj^2} j^{2s+3} < \infty, \end{aligned}$$

where Lemma 4.4.1 implies that the constants above are uniform in $\|b\|_{H^s} \leq B$. $\square$

Finally, we collect some properties of $(P_{t,b} : t \geq 0)$.

Lemma 4.4.5. *Let $B > 0$. The following holds.*

1. *For all $b \in C_0^1$, $p \in [1, \infty]$ and $f \in L^p$, we have $\|P_{t,b} f\|_{L^p(\mu_b)} \leq \|f\|_{L^p(\mu_b)}$.*

2. *For every $\epsilon > 0$, there exists $C < \infty$ such that for all $b \in C_0^1$ with $\|b\|_\infty \leq B$, $f \in H^1$ and $t > 0$,*

$$\|P_{t,b} f - f\|_{L^2} \leq C t^{1/2} \|f\|_{H^1} \quad \textit{and} \quad \|P_{t,b} f - f\|_\infty \leq C t^{1/4-\epsilon} \|f\|_{H^1}. \tag{4.65}$$

   In particular, we have that $H^1 \subseteq \mathcal{D}(1/2)$, with $D(1/2)$ defined by (4.26).

3. *Let $s \geq 1$ be an integer. Then for all $t > 0$, $b \in H^s \cap C_0^1$ with $\|b\|_{H^s} \leq B$ and $f \in L^2$, we have $P_{t,b} f \in H^{s+2}$. Moreover, there exists $C < \infty$ such that for all such t, b, f and all $\alpha \leq s + 2$,*

$$\|P_{t,b} f\|_{H^\alpha} \leq C(1 + t^{-\frac{\alpha}{2} - \frac{3}{4}}) \|f\|_{H^{-1}}. \tag{4.66}$$

Proof. 1. For the case $p = 1$, we have by Fubini's theorem that

$$\int_0^1 \Big| \int_0^1 p_{t,b}(x, z) f(z) dz \Big| d\mu(x) \leq \int_0^1 \int_0^1 p_{t,b}(x, z) d\mu(x) |f(z)| dz = \int_0^1 |f(z)| d\mu(z).$$

For the case $p = \infty$, we observe that for all $x \in [0, 1]$

$$|P_{t,b} f(x)| \leq \|f\|_\infty \int p_{t,b}(x, z) dz = \|f\|_\infty.$$

The case $p \in (1, \infty)$ follows by the Riesz-Thorin interpolation theorem.

2. To prove the first part of (4.65), let $f \in H^1 = \mathrm{Dom}\left((-\mathcal{L}_b)^{1/2}\right)$. By the 1/2-Hölder continuity of $x \mapsto e^x$ on $(-\infty, 0]$ and Lemma 4.4.2, we have that for all $t \geq 0$,

$$\begin{aligned} \|P_{t,b} f - f\|_{L^2(\mu_b)}^2 &= \sum_{j=1}^\infty \left(e^{\lambda_j t} - 1\right)^2 |\langle f, u_j \rangle_{L^2(\mu_b)}|^2 \\ &\lesssim t \sum_{j=1}^\infty |\lambda_j| |\langle f, u_j \rangle_{L^2(\mu_b)}|^2 \lesssim t \|f\|_{H^1}^2. \end{aligned}$$

The second estimate in (4.65) now follows from the $H^1 \to H^1$ boundedness of $P_{t,b}$, the embedding (4.19), the interpolation inequality (4.17) and the first part of (4.65). Indeed, we have for any $\varepsilon > 0$ that

$$\|P_{t,b} f - f\|_\infty \lesssim \|P_{t,b} f - f\|_{L^2}^{(1-4\varepsilon)/2} \|P_{t,b} f - f\|_{H^1}^{(1+4\varepsilon)/2} \lesssim t^{1/4-\varepsilon} \|f\|_{H^1}.$$

3. By Lemma 4.4.1, we have that $u_j \in H^{s+2}$ for all $j \geq 0$. Using the spectral representation (4.23), Lemma 4.4.1, Lemma 4.4.2 and Cauchy-Schwarz, we have

$$
\begin{aligned}
\|P_{t,b}f\|_{H^\alpha} &\lesssim \sum_{j=0}^{\infty} e^{\lambda_j t}\|u_j\|_{H^\alpha}(1+|\lambda_j|)^{1/2}\frac{\langle f, u_j\rangle_{L^2(\mu_b)}}{(1+|\lambda_j|)^{1/2}} \\
&\lesssim \Big(\sum_{j=0}^{\infty} e^{2\lambda_j t}(1+|\lambda_j|)^{\alpha+1}\Big)^{1/2}\|f\|_{H^{-1}} \\
&\lesssim \Big(1+\int_0^\infty e^{-2cx^2t}x^{2(\alpha+1)}dx\Big)^{1/2}\|f\|_{H^{-1}} \\
&\lesssim \Big(1+t^{-\alpha-1-\frac{1}{2}}\Big)^{1/2}\|f\|_{H^{-1}} \\
&\lesssim (1+t^{-\frac{\alpha}{2}-\frac{3}{4}})\|f\|_{H^{-1}}.
\end{aligned}
$$

□

Bibliography

[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 Caldéron problems. Mathematical Statistics and Learning, 2:165–216.

[3] Aeckerle-Willems, C. and Strauch, C. (2018a). Concentration of scalar ergodic diffusions and some statistical implications. arXiv preprint.

[4] Aeckerle-Willems, C. and Strauch, C. (2018b). Sup-norm adaptive simultaneous drift estimation for ergodic diffusions. arXiv preprint.

[5] 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.

[6] Agapiou, S., Dashti, M., and Helin, T. (2020). Rates of contraction of posterior distributions based on p-exponential priors. arXiv preprint.

[7] 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.

[8] Arridge, S., Maass, P., Öktem, O., and Schönlieb, C.-B. (2019). Solving inverse problems using data-driven models. Acta Numer., 28:1–174.

[9] Bakry, D., Gentil, I., and Ledoux, M. (2013). Analysis and Geometry of Markov Diffusion Operators. Springer.

[10] Bal, G. and Uhlmann, G. (2010). Inverse diffusion theory of photoacoustics. Inverse Problems, 26(8):085010, 20.

[11] Bao, G. and Li, P. (2005). Inverse medium scattering problems for electromagnetic waves. SIAM J. Appl. Math., 65(6):2049–2066.

[12] Bass, R. F. (1998). Diffusions and elliptic operators. Probability and its Applications (New York). Springer-Verlag, New York.

[13] Bass, R. F. (2011). Stochastic processes, volume 33 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

[14] Belloni, A. and Chernozhukov, V. (2009). On the computational complexity of MCMC-based estimators in large samples. Annals of Statistics, 37.

[15] Bengtsson, T., Bickel, P., and Li, B. (2008). Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems. In Probability and statistics: essays in honor of David A. Freedman, volume 2 of Inst. Math. Stat. (IMS) Collect., pages 316–334.

[16] Bennett, A. F. (2002). Inverse modeling of the ocean and atmosphere. Cambridge University Press, Cambridge.

[17] Benning, M. and Burger, M. (2018). Modern regularization methods for inverse problems. Acta Numer., 27:1–111.

[18] 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.

[19] Bickel, P., Li, B., and Bengtsson, T. (2008). Sharp failure rates for the bootstrap particle filter in high dimensions. In Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh, volume 3 of Inst. Math. Stat. (IMS) Collect., pages 318–329.

[20] 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.

[21] Bissantz, N., Hohage, T., Munk, A., and Ruymgaart, F. (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. SIAM J. Numer. Anal., 45(6):2610–2636.

[22] Blanchard, G., Hoffmann, M., and Reiß, M. (2018a). Early stopping for statistical inverse problems via truncated SVD estimation. Electron. J. Stat., 12(2):3204–3231.

[23] Blanchard, G., Hoffmann, M., and Reiß, M. (2018b). Optimal adaptation for early stopping in statistical inverse problems. SIAM/ASA J. Uncertain. Quantif., 6(3):1043–1075.

[24] Bonito, A., Cohen, A., DeVore, R., Petrova, G., and Welper, G. (2017). Diffusion Coefficients Estimation for Elliptic Partial Differential Equations. SIAM Journal on Mathematical Analysis, 49(2):1570–1592.

[25] Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. Invent. Math., 30(2):207–216.

[26] Bou-Rabee, N., Eberle, A., and Zimmer, R. (2020). Coupling and convergence for hamiltonian monte carlo. Ann. Appl. Probab., 30(3):1209–1250.

[27] Bou-Rabee, N. and Hairer, M. (2013). Nonasymptotic mixing of the MALA algorithm. IMA J. Numer. Anal., 33(1):80–110.

[28] Boyd, S. and Vandenberge, L. (2004). Convex optimization. Cambridge University Press.

[29] Brown, B. (1971). Martingale Central Limit Theorems. Ann. Math. Statist., 42:59–66.

[30] Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. Ann. Statist., 24(6):2384–2398.

[31] Calderón, A.-P. (1980). On an inverse boundary value problem. In Seminar on Numerical Analysis and its Applications to Continuum Physics (Rio de Janeiro, 1980), pages 65–73. Soc. Brasil. Mat., Rio de Janeiro.

[32] Cappé, O., Moulines, E., and Rydén, T. (2005). Inference in hidden Markov models. Springer Series in Statistics. Springer, New York.

[33] Castillo, I. and Nickl, R. (2013). Nonparametric Bernstein–von Mises Theorems in Gaussian white noise. Ann. Statist., 41(4):1999–2028.

[34] Castillo, I. and Nickl, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. Ann. Statist., 42(5):1941–1969.

[35] Castillo, I. and Rousseau, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. Ann. Statist., 43(6):2353–2383.

[36] Cavalier, L. (2008). Nonparametric statistical inverse problems. Inverse Problems, 24(3):034004, 19.

[37] Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. Probab. Theory Related Fields, 123(3):323–354.

[38] 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. 92–70.

[39] Chen, Z. and Vempala, S. S. (2019). Optimal convergence rate of Hamiltonian Monte Carlo for strongly logconcave distributions. In Approximation, randomization, and combinatorial optimization. Algorithms and techniques, volume 145 of LIPIcs. Leibniz Int. Proc. Inform., pages Art. No. 64, 12. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern.

[40] 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.

[41] Chorowski, J. (2016). Statistics for diffusion processes with low and high-frequency observations. PhD Thesis.

[42] Chung, K. L. and Zhao, Z. X. (1995). From Brownian motion to Schrödinger's equation. Springer-Verlag, Berlin.

[43] Conrad, P. R., Marzouk, Y. M., Pillai, N. S., and Smith, A. (2016). Accelerating asymptotically exact MCMC for computationally intensive models via local approximations. J. Amer. Statist. Assoc., 111(516):1591–1607.

[44] 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.

[45] 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.

[46] Cox, D. D. (1988). Approximation of method of regularization estimators. Ann. Statist., 16(2):694–712.

[47] Cui, T., Law, K. J. H., and Marzouk, Y. M. (2016). Dimension-independent likelihood-informed MCMC. J. Comput. Phys., 304:109–137.

[48] 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.

[49] Dashti, M., Harris, S., and Stuart, A. M. (2012). Besov priors for Bayesian inverse problems. Inverse Probl. Imaging, 6(2):183–200.

[50] Dashti, M., Law, K., Stuart, A. M., and Voss, J. (2013a). MAP estimators an their consistency in Bayesian nonparametric inverse problems. Inverse Problems, 29(9).

[51] Dashti, M., Law, K. J. H., Stuart, A. M., and Voss, J. (2013b). MAP estimators and their consistency in Bayesian nonparametric inverse problems. Inverse Problems, 29(9):095017, 27.

[52] 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.

[53] Dashti, M. and Stuart, A. M. (2016). The Bayesian approach to inverse problems. In: Handbook of Uncertainty Quantification, Eds R. Ghanem et al, Springer.

[54] Davies, E. B. (1995). Spectral Theory and Differential Operators. Cambridge University Press.

[55] Dirksen, S. (2015). Tail bounds via generic chaining. Electron. J. Probab., 20:no. 53, 29.

[56] Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Appl. Comput. Harmon. Anal., 2(2):101–126.

[57] Durmus, A. and Moulines, E. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. Ann. Appl. Probab., 27(3):1551–1587.

[58] Durmus, A. and Moulines, E. (2019). High-dimensional Bayesian inference via the unadjusted Langevin algorithm. Bernoulli, 25.

[59] 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.

[60] 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.

[61] Eberle, A. (2015). Reflection couplings and contraction rates for diffusions. Probability Theory and Related Fields, 166:851–886.

[62] Eberle, A., Guillin, A., and Zimmer, R. (2019). Couplings and quantitative contraction rates for Langevin dynamics. Ann. Probab., 47(4):1982–2010.

[63] Engl, H. W., Hanke, M., and Neubauer, A. (1996). Regularization of inverse problems, volume 375 of Mathematics and its Applications. Kluwer, Dordrecht.

[64] Engl, H. W., Kunisch, K., and Neubauer, A. (1989). Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. Inverse Problems, 5(4):523–540.

[65] Evans, L. C. (2010). Partial differential equations. American Math. Soc., Second edition.

[66] Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. Journal of Multivariate Analysis, 74(1):49 – 68.

[67] Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. Ann. Statist., 28(2):500–531.

[68] Ghosal, S. and van der Vaart, A. W. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. Ann. Statist., 35(1):192–223.

[69] Ghosal, S. and van der Vaart, A. W. (2017). Fundamentals of Nonparametric Bayesian Inference. Cambridge University Press, New York.

[70] Gihman, I. and Skorohod, A. (1972). Stochastic differential equations. Springer.

[71] Gilbarg, D. and Trudinger, N. S. (1998). Elliptic partial differential equations of second order. Springer-Verlag, Berlin-New York.

[72] Giné, E. and Nickl, R. (2016). Mathematical foundations of infinite-dimensional statistical models. Cambridge University Press, New York.

[73] Giordano, M. and Kekkonen, H. (2018). Bernstein-von mises theorems and uncertainty quantification for linear inverse problems. SIAM /ASA Journal on Uncertainty Quantification, to appear.

[74] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. Annales de l'Institut Henri Poincaré B, 38:711–737.

[75] Gobet, E., Hoffmann, M., and Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. Ann. Statist., 32(5):2223–2253.

[76] Gugushvili, S. and Spreij, P. (2014). Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. Lith. Math. J., 54(2):127–141.

[77] Gugushvili, S., van der Vaart, A., and Yan, D. (2020). Bayesian linear inverse problems in regularity scales. Ann. Inst. Henri Poincaré Probab. Stat., 56(3):2081–2107.

[78] Haario, H., Laine, M., Lehtinen, M., Saksman, E., and Tamminen, J. (2004). Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. J. R. Stat. Soc. Ser. B Stat. Methodol., 66(3):591–607.

[79] Hairer, M., Mattingly, J., and Scheutzow, M. (2011). Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. Probab. Theory Relat. Fields, 149:223–259.

[80] 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.

[81] Hanke, M., Neubauer, A., and Scherzer, O. (1995). A convergence analysis of the landweber iteration for nonlinear ill-posed problems. Numerische Mathematik.

[82] Hansen, L. P., Scheinkman, J. A., and Touzi, N. (1998). Spectral methods for identifying scalar diffusions. Journal of Econometrics, 86:1–32.

[83] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57(1):97–109.

[84] Helin, T. and Burger, M. (2015). Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. Inverse Problems, 31(8):085009, 22.

[85] Hinrichs, A., Novak, E., Ullrich, M., and Woźniakowski, H. (2014). The curse of dimensionality for numerical integration of smooth functions. Math. Comp., 83(290):2853–2863.

[86] Hoffmann, M., Munk, A., and Schmidt-Hieber, J. (2012). Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. Annals of the Institute H. Poincaré, 48:1186–1216.

[87] 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.

[88] Hohage, T. and Weidling, F. (2015). Verification of a variational source condition for acoustic inverse medium scattering problems. Inverse Problems, 31(7):075006, 14.

[89] Holley, R. and Stroock, D. (1987). Logarithmic Sobolev inequalities and stochastic Ising models. Journal of Statistical Physics, 46:1159–1194.

[90] Ilmavirta, J. and Monard, F. (2020, to appear). Integral geometry on manifolds with boundary and applications. Inverse problems.

[91] Isakov, V. (2006). Inverse problems for partial differential equations, volume 127 of Applied Mathematical Sciences. Springer, New York, second edition.

[92] 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.

[93] Kaipio, J., Kolehmainen, V., Somersalo, E., and Vauhkonen, M. (2000). Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. Inverse Problems, 16(5):1487–1522.

[94] Kaipio, J. and Somersalo, E. (2005). Statistical and computational inverse problems, volume 160 of Applied Mathematical Sciences. Springer-Verlag, New York.

[95] Kaltenbacher, B., Neubauer, A., and Scherzer, O. (2008). Iterative regularization methods for nonlinear ill-posed problems. de Gruyter, Berlin.

[96] Kaltenbacher, B., Nguyen, T. T. N., and Scherzer, O. (2019). The tangential cone condition for some coefficient identification model problems in parabolic pdes. arXiv preprint.

[97] Katchalov, A., Kurylev, Y., and Lassas, M. (2001). Inverse boundary spectral problems. Chapman & Hall/CRC, Boca Raton, FL.

[98] 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.

[99] 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.

[100] Knapik, B. T., Szabó, B. T., van der Vaart, A. W., and van Zanten, J. H. (2016). Bayes procedures for adaptive inference in inverse problems for the white noise model. Probab. Theory Related Fields, 164(3-4):771–813.

[101] Koskela, J., Spano, D., and Jenkins, P. A. (2015). Consistency of Bayesian nonparametric inference for discretely observed jump diffusions. arXiv:1506.04709.

[102] Kunisch, K. and Ito, 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.

[103] Kutoyants, Y. A. (2004). Statistical inference for ergodic diffusion processes. Springer Series in Statistics. Springer-Verlag London, Ltd., London.

[104] Laplace, P.-S. M. d. (1812). Theorie analytiques des probabilités. Courcier, Paris.

[105] Lassas, M., Saksman, E., and Siltanen, S. (2009). Discretization-invariant Bayesian inversion and Besov space priors. Inverse Probl. Imaging, 3(1):87–122.

[106] Le Cam, L. (1986). Asymptotic methods in statistical decision theory. Springer-Verlag, New York.

[107] Li, W. V. and Linde, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. Ann. Probab., 27(3):1556–1578.

[108] Lions, J. and Magenes, E. (1972a). Non-Homogeneous Boundary Value Problems and Applications, Vol. 1. Springer.

[109] Lions, J.-L. and Magenes, E. (1972b). Non-homogeneous boundary value problems and applications. Vol. I. Springer-Verlag, New York-Heidelberg.

[110] Livingstone, S., Betancourt, M., Byrne, S., and Girolami, M. (2019). On the geometric ergodicity of Hamiltonian Monte Carlo. Bernoulli, 25(4A):3109–3138.

[111] Löffler, M. and Picard, A. (2018). Spectral thresholding for the estimation of markov chain transition operators. arXiv preprint.

[112] Loubes, J.-M. and Ludeña, C. (2010). Penalized estimators for non linear inverse problems. ESAIM Probab. Stat., 14:173–191.

[113] Lovász, L. and Simonovits, M. (1993). Random walks in a convex body and an improved volume algorithm. Random Structures Algorithms, 4(4):359–412.

[114] Lovász, L. and Vempala, S. (2007). The geometry of logconcave functions and sampling algorithms. Random Structures Algorithms, 30(3):307–358.

[115] Lu, Y. (2017). On the bernstein-von mises theorem for high dimensional nonlinear bayesian inverse problems. arXiv:1706.00289.

[116] Lunardi, A. (1995). Analytic Semigroups and Optimal Regularity for Parabolic Problems. Birkhäuser.

[117] Ma, Y.-A., Chen, Y., Jin, C., Flammarion, N., and Jordan, M. (2019). Sampling can be faster than optimization. Proc. Nat. Acad. Sciences.

[118] Mair, B. A. and Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales. SIAM J. Appl. Math., 56(5):1424–1444.

[119] Majda, A. J. and Harlim, J. (2012). Filtering complex turbulent systems. Cambridge University Press, Cambridge.

[120] 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.

[121] McLean, W. (2000). Strongly elliptic systems and boundary integral equations. Cambridge University Press, Cambridge.

[122] Megginson, R. E. (1998). An Introduction to Banach Space Theory, volume 183 of Graduate Texts in Mathematics. Springer.

[123] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. Journal of Chemical Physics, 21(6):1087–1092.

[124] Meyn, S. and Tweedie, R. L. (2009). Markov chains and stochastic stability. Cambridge University Press, Cambridge, second edition. With a prologue by Peter W. Glynn.

[125] Monard, F., Nickl, R., and Paternain, G. P. (2019). Efficient nonparametric Bayesian inference for X-ray transforms. Ann. Statist., 47(2):1113–1147.

[126] Monard, F., Nickl, R., and Paternain, G. P. (2020, to appear). Consistent inversion of noisy non-Abelian X-ray transforms. Comm. Pure Appl. Math.

[127] Mueller, J. L. and Siltanen, S. (2012). Linear and nonlinear inverse problems with practical applications, volume 10. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[128] Nachman, A. I. (1996). Global uniqueness for a two-dimensional inverse boundary value problem. Ann. of Math. (2), 143(1):71–96.

[129] Natterer, F. (1986). The mathematics of computerized tomography. B. G. Teubner, Stuttgart; John Wiley & Sons, Ltd., Chichester.

[130] Neubauer, A. (1992). Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales. Appl. Anal., 46(1-2):59–72.

[131] Nickl, R. (2020). Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation. J. Eur. Math. Soc., 22: 2697–2750.

[132] Nickl, R. and Giordano, M. (2020). Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. Inverse Problems, to appear.

[133] Nickl, R. and Ray, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. Ann. Statist., 48(3):1383–1408.

[134] Nickl, R. and Söhl, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. Ann. Statist., 45(4):1664–1693.

[135] 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.

[136] Nickl, R., van de Geer, S., and Wang, S. (2020). Convergence rates for penalised least squares estimators in PDE-constrained regression problems. SIAM J. Uncert. Quant., 8.

[137] Nickl, R. and Wang, S. (2020). On polynomial-time computation of high-dimensional posterior measures by langevin-type algorithms. arXiv preprint.

[138] Nirenberg, L. (1959). On elliptic partial differential equations. Ann. Sc. Norm. Sup. Pisa, 13(2):115–162.

[139] Novak, E. and Woźniakowski, H. (2008). Tractability of multivariate problems. Vol. 1: Linear information, volume 6 of EMS Tracts in Mathematics.

[140] Novak, E. and Woźniakowski, H. (2010). Tractability of multivariate problems. Volume II: Standard information for functionals, volume 12 of EMS Tracts in Mathematics.

[141] O'Sullivan, F. (1990). Convergence characteristics of methods of regularization estimators for nonlinear operator equations. SIAM J. Numer. Anal., 27(6):1635–1649.

[142] Papaspiliopoulos, O., Pokern, Y., Roberts, G. O., and Stuart, A. M. (2012). Nonparametric estimation of diffusions: a differential equations approach. Biometrika, 99:511–531.

[143] Paternain, G. P., Salo, M., and Uhlmann, G. (2012). The attenuated ray transform for connections and Higgs fields. Geom. Funct. Anal., 22(5):1460–1489.

[144] Ray, K. (2013). Bayesian inverse problems with non-conjugate priors. Electron. J. Stat., 7:2516–2549.

[145] Ray, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. Ann. Statist., 45(6):2511–2536.

[146] Rebeschini, P. and van Handel, R. (2015). Can local particle filters beat the curse of dimensionality? Ann. Appl. Probab., 25(5):2809–2866.

[147] Reich, S. and Cotter, C. (2015). Probabilistic forecasting and Bayesian data assimilation. Cambridge University Press, New York.

[148] Reiß, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. Ann. Statist., 36(4):1957–1982.

[149] Richter, G. R. (1981). An inverse problem for the steady state diffusion equation. SIAM J. Appl. Math., 41(2):210–221.

[150] Robert, C. P. and Casella, G. (2004). Monte Carlo statistical methods. Springer Texts in Statistics. Springer-Verlag, New York, second edition.

[151] Roberts, G. O. and Tweedie, R. L. (1996a). Exponential convergence of Langevin distributions and their discrete approximations. Bernoulli, 2(4):341–363.

[152] 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.

[153] Roussas, G. G. (1972). Contiguity of probability measures: some applications in statistics. Cambridge University Press, London-New York. Cambridge Tracts in Mathematics and Mathematical Physics, No. 63.

[154] Scherzer, O., Engl, H. W., and Kunisch, K. (1993). Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. SIAM J. Numer. Anal., 30(6):1796–1838.

[155] Stuart, A. M. (2010). Inverse problems: a Bayesian perspective. Acta Numer., 19:451–559.

[156] Suharev, A. G. (1979). Optimal formulas of numerical integration for some classes of functions of several variables. Dokl. Akad. Nauk SSSR, 246(2):282–285.

[157] Sylvester, J. and Uhlmann, G. (1987). A global uniqueness theorem for an inverse boundary value problem. Ann. of Math. (2), 125(1):153–169.

[158] Talagrand, M. (2014). Upper and lower bounds for stochastic processes. Springer, Heidelberg.

[159] Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[160] Tautenhahn, U. and Jin, Q.-n. (2003). Tikhonov regularization and a posteriori rules for solving nonlinear ill posed problems. Inverse Problems, 19(1):1–21.

[161] Taylor, M. E. (2011a). Partial differential equations I. Basic theory. Springer, New York.

[162] Taylor, M. E. (2011b). Partial differential equations II. Qualitative studies of linear equations. Springer, New York.

[163] Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. Ann. Appl. Probab., 8(1):1–9.

[164] Tikhonov, A. N. and Arsenin, V. Y. (1977). Solutions of ill-posed problems. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York-Toronto, Ont.-London. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.

[165] Triebel, H. (1978). Interpolation theory, function spaces, differential operators, volume 18 of North-Holland Mathematical Library. North-Holland, New York.

[166] Triebel, H. (1983). Theory of function spaces, volume 78 of Monographs in Mathematics. Birkhäuser Verlag, Basel.

[167] Triebel, H. (2008). Function spaces and wavelets on domains. Eur. Math. Soc., Zürich.

[168] Uhlmann, G. (2009). Electrical impedance tomography and Calderón's problem. Inverse Problems, 25(12):123011, 39.

[169] van de Geer, S. (2001). Least squares estimation with complexity penalties. volume 10, pages 355–374. Meeting on Mathematical Statistics (Marseille, 2000).

[170] van de Geer, S. A. (2000). Applications of empirical process theory, volume 6 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

[171] van der Meulen, F., Schauer, M., and van Zanten, H. (2014). Reversible jump MCMC for nonparametric drift estimation for diffusion processes. Comput. Statist. Data Anal., 71:615–632.

[172] van der Vaart, A. (1998). Asymptotic statistics. Cambridge University Press, Cambridge.

[173] van der Vaart, A. and van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. Ann. Statist., 36(3):1435–1463.

[174] 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.

[175] Vempala, S. and Wibisono, A. (2019). Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. Advances in Neural Information Processing Systems 32.

[176] Villani, C. (2009). Optimal transport. Old and new. Springer-Verlag, Berlin.

[177] Vollmer, S. J. (2013). Posterior consistency for Bayesian inverse problems through stability and regression results. Inverse Problems, 29(12):125011, 32.

[178] Wang, S. (2019). The nonparametric LAN expansion for discretely observed diffusions. Electron. J. Stat., 13(1):1329–1358.

[179] Wei, Y., Yang, F., and Wainwright, M. J. (2019). Early stopping for kernel boosting algorithms: a general analysis with localized complexities. IEEE Trans. Inform. Theory, 65(10):6685–6703.

[180] 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.

[181] Yang, Y., Wainwright, M., and Jordan, M. I. (2016). On the computational complexity of high-dimensional Bayesian variable selection. Annals of Statistics.

[182] Yeh, W. W.-G. (1986). Review of parameter identification procedures in groundwater hydrology: The inverse problem. Water Resources Research, 22(2):95–108.

[183] Ziemer, W. (1990). Weakly Differentiable Functions. Springer.