J. Eur. Math. Soc. 26, 1031–1112 (2024) DOI 10.4171/JEMS/1304

Richard Nickl · Sven Wang

On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms

Received October 5, 2020; revised February 18, 2022

Abstract. The problem of generating random samples of high-dimensional posterior distributions is considered. 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.

Our results are derived 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. The theory is applied to a representative non-linear example from PDEs involving a steady-state Schrödinger equation.

Keywords. Bayesian inference, Markov Chain Monte Carlo, non-linear inverse problems, Bayesian computation, Markov chain mixing times

Contents

Richard Nickl: Centre for Mathematical Sciences, University of Cambridge, Cambridge, CB3 0WB, UK; nickl@maths.cam.ac.uk

Sven Wang: IDSS, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; svenwang@mit.edu

Mathematics Subject Classification (2020): Primary 35J10; Secondary 60J22, 62F15

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

$$\pi(\theta \mid Z^{(N)}) \propto e^{\ell_N(\theta)}\pi(\theta), \quad \theta \in \mathbb{R}^D. \tag{1}$$

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 $\pi(\cdot \mid Z^{(N)})$ 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 $\pi(\cdot \mid Z^{(N)})$ 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* (or worse) dependence on D or N of the 'mixing time' of the Markov chain $(\vartheta_k : k \in \mathbb{N})$ towards its equilibrium measure (1).

In this work we develop mathematical techniques which allow one to overcome such computational hardness barriers. We consider diffusion-based MCMC algorithms targeting the Gibbs-type measure with density $\pi(\cdot \mid Z^{(N)})$ from (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 non-parametric statistics [74, 77] and the non-asymptotic theory of Langevin algorithms [32, 36, 37] 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 [92], MCMC-based Bayesian methodology has been frequently used in such settings, especially for the task of uncertainty quantification. We refer the reader to [3, 8, 13, 22, 23, 29–31, 34, 47, 48, 56, 57, 63, 72, 88, 90, 92] and the references therein. A main contribution of this paper is to demonstrate the feasibility of our proof strategy in a 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., [6, 7]), and provides a suitable framework to lay out the main mathematical ideas underpinning our proofs.

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 $\{\mathcal{G}(\theta) : \theta \in \Theta\}$ of real-valued bounded 'regression' functions defined on $\mathcal{O}$. If $L^2(\mathcal{O})$ denotes the usual space of square Lebesgue-integrable functions, this induces a 'forward map'

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

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

$$Y_i = \mathcal{G}(\theta)(X_i) + \varepsilon_i, \quad i = 1, \dots, N, \tag{3}$$

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

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N}[Y_i - \mathcal{G}(\theta)(X_i)]^2, \quad \theta \in \mathbb{R}^D. \tag{4}$$

The aim is to recover the unknown θ from $Z^{(N)}$. A widespread 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 (1).

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

$$\begin{cases} \frac{1}{2}\Delta u - f_\theta u = 0 & \text{on } \mathcal{O}, \\ u = g & \text{on } \partial\mathcal{O}, \end{cases} \tag{5}$$

with a suitable parameterisation $\theta \mapsto f_\theta > 0$, $\theta \in \mathbb{R}^D$ (see (17) below for details). In such cases, the map $\mathscr{G}$ is non-linear and $-\ell_N(\theta)$ is not convex. The probability measure with density $\pi(\cdot \mid Z^{(N)})$ given in (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 [82, 83] for general references) implies that computing the integral of a D-dimensional real-valued Lipschitz function—such as the normalising factor implicit in (1)—by a deterministic algorithm has worst case cost scaling as $D^{D/5}$ [52, 93]. 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 [9–12, 14, 71, 87, 105] 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 $\pi(\cdot \mid Z^{(N)})$ (e.g., [38], Example 4).

In light of the preceding discussion one may ask whether the approximate calculation of basic aspects of $\pi(\cdot \mid Z^{(N)})$—such as its mean vector (expected value), real-valued functionals $\int_{\mathbb{R}^D} H(\theta)\pi(\theta \mid Z^{(N)})\, d\theta$, or mode—is feasible at a computational cost which grows at most *polynomially in* D, N and the desired (inverse) precision level. While answering this question in the affirmative may not directly identify a practical algorithm, it clarifies a fundamental aspect of the computational complexity of the problem at hand. Very few rigorous results providing even just partial such guarantees appear to be available in PDE settings. The notable exception of Hairer, Stuart and Vollmer [50] 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 (5) with the Schrödinger equation. We will require mild regularity assumptions on D, Π and on the ground truth θ_0 generating the data (3)—full details can be found in Section 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 (28)), we obtain the following main result.

Theorem 1.1. *Suppose that data $Z^{(N)} = (Y_i, X_i : i = 1, \dots, N)$ arise through* (3) *in the Schrödinger model* (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*

$$O(N^{b_1} D^{b_2} \varepsilon^{-b_3}) \quad (b_1, b_2, b_3 > 0), \tag{6}$$

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

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

where

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

denotes the mean vector of the posterior distribution $\Pi(\cdot \mid Z^{(N)})$ *with density* (1).

We further show in Theorem 2.6 that $\hat{\theta}_\varepsilon$ also recovers the ground truth θ_0, within precision ε. The method underlying Theorem 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 $(1/J)\sum_{k\le J} \vartheta_k$ along the MCMC chain (after some burn-in time); see Theorem 2.5 (which implies Theorem 1.1). The laws $\mathcal{L}(\vartheta_k)$ of the iterates (ϑ_k) in fact provide a *global* approximation

$$W_2(\mathcal{L}(\vartheta_k), \Pi(\cdot \mid Z^{(N)})) \le \varepsilon, \quad k \ge k_{\text{mix}},$$

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 J of required iterations to reach precision level ε scales polynomially in D, N, ε^{-1}. Similar statements hold true for the computation of real-valued functionals $\int_{\mathbb{R}^D} H(\theta)\pi(\theta \mid Z^N)\, d\theta$ for Lipschitz maps $H : \mathbb{R}^D \to \mathbb{R}$ and of maximum a posteriori (MAP) estimates. See Theorems 2.7, 2.8 as well as Proposition 2.4 for precise statements.

The main ideas of this article can be summarised as follows. We first demonstrate that, with high probability under the law generating the data $Z^{(N)}$, the target measure $\Pi(\cdot \mid Z^{(N)})$ from (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 B.3). Our proofs provide a template (outlined in Section 3) that can be used in principle also in general settings as long as the linearisation $\nabla_\theta \mathscr{G}(\theta_0)$ of $\mathscr{G}$ at the ground truth parameter θ_0 satisfies a suitable stability estimate (i.e., a quantitative injectivity property related to the 'information' operator of the statistical model). We note that this 'gradient stability' hypothesis remains entirely 'local' and is hence weaker than the 'Polyak–Łojasiewicz' gradient condition used in non-convex optimisation [67,86] (see also [59]). We verify our local stability property for the Schrödinger equation using elliptic PDE techniques (see Lemma 4.7) but our approach may succeed in a variety of other non-linear forward models arising in inverse problems [60,76,92,98], integral X-ray geometry [54,74,84,85], and also in the context of data assimilation and filtering [29,72,88]. In fact, the very recent reference [16] achieves this for the non-linear inverse problem considered in [74,84]. Further advancing our understanding of the computational complexity of such PDE-constrained high-dimensional inference problems poses a formidable challenge for future research.

1.2. *Discussion of related literature*

Both the statistical and computational aspects of high-dimensional Bayes procedures have been a subject of great interest in recent years. Frequentist convergence properties of high- and infinite-dimensional Bayes procedures were intensively studied in the last two decades. For 'direct' statistical models we refer to the recent monograph [42] (and references therein), and in the non-linear (PDE) setting relevant here to [1, 2, 15, 16, 18, 45, 61, 74, 75, 77, 79–81, 91].

We now discuss a variety of mixing time results of MCMC algorithms in high-dimensional settings, and refer to the references cited in these articles for further important results.

1.2.1. Mixing times for pCN-type algorithms. The important contribution [50] 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 [49]. 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 paper, where Φ is given by (4), i.e. $\Phi = \Phi_N = -\ell_N$ and it is natural to ask (a) whether the bounds from [50] 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 [50] can be summarised as

(A) a global lower bound on the acceptance probability of the pCN, and

(B) a (local) Lipschitz continuity requirement on Φ.

In PDE models, part (B) can usually be verified (e.g., [81]), 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 (18) below. 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 [50] 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. [50, p. 2474]), and from the qualitative nature of the key underlying probabilistic weak Harris theorem proved in [49].

Summarising, while it would be very exciting to see the results of [50] extended to yield quantitative bounds which are polynomial in both N, D, serious technical and conceptual innovations seem to be required. These remarks apply as well to recent dimension-free mixing time bounds on Hamiltonian Monte Carlo (HMC) methods in [19, 20, 46], which scale exponentially in N via the Lipschitz constant of ℓ_N. In our context, when exploiting local average curvature of the likelihood surface arising from PDE structure, it is initially more promising to use diffusion-based methods.

1.2.2. Computational guarantees for Langevin-type algorithms. For the important gradient-based class of Langevin Monte Carlo (LMC) algorithms, non-asymptotic convergence guarantees which are suited for high-dimensional settings were obtained by Dalalyan [32] for log-concave densities, and extended shortly after by Durmus and Moulines [36, 37] to closely related cases. Our proofs rely substantially on these convergence results for the strongly log-concave case (see Appendix A for a review). We emphasise that the fundamental ideas underpinning the fast mixing of 'hypercontractive' Langevin diffusions in high dimensions go back to earlier seminal work [4, 55]; see also the monograph [5].

Very recently further extensions have emerged, notably [27, 70, 71, 103], which establish convergence guarantees assuming that either the density to be sampled from is convex outside of some region, or the target measure satisfies functional inequalities of log-Sobolev and Poincaré type. However, it appears that both of these results, when applied to (4) without any further substantial work, yield bounds that scale exponentially in N. Indeed, the bound in [71, Theorem 1] evidently depends exponentially on the Lipschitz constant of the gradient $\nabla \ell_N$; and ad hoc verification of assumptions from [103] would utilise the Holley–Stroock perturbation principle [53] (and (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 here.

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') approximations to posterior distributions; see [62, 64, 101] for the classical versions of such results in 'low-dimensional' statistical models, and [24–26, 41] for high- or infinite-dimensional versions.

Such an approach is taken by [9] 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 Lovász, Simonovits and Vempala [68, 69] on the complexity of general Metropolis–Hastings schemes. While [9] potentially allows for moderately high-dimensional situations (by appealing to high-dimensional Bernstein–von Mises theorems from [41]), their sampling guarantees hold for rescaled posterior measures arising as laws of $\sqrt{N}\,(\theta - \tilde{\theta}) \mid Z^{(N)}$ where $\tilde{\theta} = \tilde{\theta}(Z^{(N)})$ is an initial 'semi-parametrically efficient centring' of the posterior draws $\theta \mid Z^{(N)}$. In our setting such a centring is not generally available (in fact, to show that one can compute such centrings, such as the posterior mode or mean, in polynomial time, is one of the main aims of our analysis). The setting in [9] 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 the present paper, Bernstein–von Mises theorems were obtained in [77]—see also the more recent paper [75]. While we follow [77] 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 (via Le Cam theory, as in [75, 77]). 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 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 [77] or in [73, 75, 80]), 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 PDE model where this difference matters is discussed in [78]). 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 [9] does), instead only a sufficiently fast 'non-parametric' convergence rate is required, which substantially increases the class of admissible initialisation strategies.

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 [39, 58]. 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 [51, 58], assuming that optimisation is performed over a (sufficiently small) neighbourhood of a maximum. The proof techniques underlying our main results allow one as well to derive guarantees for gradient descent algorithms targeting, for instance, maximum a posteriori (MAP) estimates; see Section 2.2.5. Specifically, in Theorem 2.8, global convergence guarantees for the computation of MAP estimates over a high-dimensional discretisation space are given, in our 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.

1.3. Notations and conventions

Throughout, N will denote the number of observations in (3) and D will denote the dimension of the model from (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

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

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

$\ell^2(\mathbb{N})$ is the usual space of square-summable sequences $(a_n : n \in \mathbb{N})$, 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 $L^2(\mathcal{O})$ 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 $C^\alpha(\mathcal{O}), \alpha \in \mathbb{N}$, are 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 [95]. We also define $(H_0^2(\mathcal{O}))^*$ as the topological dual space of

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

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

$$\|gh\|_{H^\alpha} \le c(\alpha, \mathcal{O}) \|g\|_{H^\alpha} \|h\|_{H^\alpha}, \qquad \alpha > d/2, \tag{7}$$

$$\|h\|_{H^\beta} \le c(\beta, \alpha, \mathcal{O}) \|h\|_{L^2}^{(\alpha-\beta)/\alpha} \|h\|_{H^\alpha}^{\beta/\alpha}, \quad 0 \le \beta \le \alpha, \tag{8}$$

for $g, h \in H^\alpha$; see, e.g., [66]. For Borel probability measures μ_1, μ_2 on $\mathbb{R}^D$ with finite second moments we define the *Wasserstein distance*

$$W_2^2(\mu_1, \mu_2) = \inf_{\nu \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^D \times \mathbb{R}^D} \|\theta - \vartheta\|_{\mathbb{R}^D}^2 \, d\nu(\theta, \vartheta), \tag{9}$$

where $\Gamma(\mu_1, \mu_2)$ is the set of all 'couplings' of μ_1 and μ_2 (see, e.g., [104]). Finally, we say that a map $H : \mathbb{R}^D \to \mathbb{R}$ is Lipschitz if it has finite Lipschitz norm

$$\|H\|_{\mathrm{Lip}} := \sup_{x \neq y, x, y \in \mathbb{R}^D} \frac{|H(x) - H(y)|}{\|x - y\|_{\mathbb{R}^D}}. \tag{10}$$

2. Main results for the Schrödinger model

Our object of study in this section is a non-linear forward model arising from 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 restrict ourselves throughout to $d \le 3$; dimensions $d \ge 4$ could be considered as well at the expense of further technicalities. Moreover, without loss of generality we assume $\mathrm{vol}(\mathcal{O}) = 1$.

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

$$\begin{cases} \frac{1}{2}\Delta u - fu = 0 & \text{on } \mathcal{O}, \\ u = g & \text{on } \partial\mathcal{O}. \end{cases} \tag{11}$$

If $\alpha > d/2$ and $f \geq 0$ then standard theory for elliptic PDEs (see [43, Chapter 6] or [28, Chapter 4]) implies that a unique classical solution $u_f \in C^2(\mathcal{O})$ to the Schrödinger equation (11) exists. The non-linearity of the map $f \mapsto u_f$ becomes apparent from the classical Feynman–Kac formula (e.g., [28, Theorem 4.7])

$$u_f(x) = u_{f,g}(x) = E^x\big[g(X_{\tau_{\mathcal{O}}})e^{-\int_0^{\tau_{\mathcal{O}}} f(X_s)\,ds}\big], \quad x \in \mathcal{O}, \tag{12}$$

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 applications, for example in photo-acoustics [7, Section 3].

2.1. *Bayesian inference with Gaussian process priors*

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., [92, Section 2.4] and also [42, Example 11.8] (where the closely related 'Whittle–Matérn' processes are considered).

For $\psi \in L^2(\mathcal{O})$, let $v \equiv \mathbb{V}[\psi]$ denote the (unique) solution in H_0^2 to the Poisson equation $\Delta v/2 = \psi$ on $\mathcal{O}$. By standard results [95, Section 5.A] 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., [96, Corollary 8.3.5])

$$\lambda_k = \frac{1}{|\mu_k|} \simeq k^{2/d} \quad \text{as } k \to \infty, \quad 0 < \lambda_k < \lambda_{k+1}, \quad k \in \mathbb{N}. \tag{13}$$

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

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

One shows that $\mathcal{H}^\alpha$ is a closed subspace of $H^\alpha(\mathcal{O})$ and 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 [95, Section 5.A]. 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

$$\mathcal{M}_\alpha(x) = \sum_{k=1}^\infty \lambda_k^{-\alpha/2} g_k e_k(x), \quad x \in \mathcal{O}, \quad g_k \overset{\text{i.i.d.}}{\sim} N(0,1). \tag{14}$$

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 [44, Example 2.6.15]), 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 [95, p. 473]). Our techniques in principle may extend to other classes of priors such as exponential Besov-type priors considered in [2, 63], but we focus our development here on the most commonly used class of α-regular Gaussian process priors.

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 (11), we need to enforce positivity by use of a 'link function' Φ. While $\Phi = \exp$ is common, for technical convenience (following [81]) we choose a function that is globally Lipschitz.

Definition 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 [81, Example 3.2]. We denote the composition operator associated to Φ by

$$\Phi^* : L^2(\mathcal{O}) \to L^2(\mathcal{O}), \quad F \mapsto \Phi \circ F = \Phi^*(F). \tag{15}$$

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 2.1.1,

$$F = F_\theta = \sum_{k=1}^{\infty} \theta_k e_k, \quad (\theta_k : k = 1, 2, \dots) \in \ell^2(\mathbb{N}), \tag{16}$$

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 (176) below) that $F_\theta \in H^\alpha(\mathcal{O})$ implies $f_\theta \in H^\alpha(\mathcal{O})$ and hence solutions of the Schrödinger equation (11) exist for such f. If we denote the solution map $f \mapsto u_f$ from (11) by G, then the overall forward map describing our parametrisation is given by

$$\mathcal{G} : h^\alpha \to L^2(\mathcal{O}), \quad \mathcal{G}(\theta) = u_{f_\theta} = [G \circ \Phi^* \circ \Psi](\theta). \tag{17}$$

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 $(\theta_1, \dots, \theta_D)$ of $\theta \in h^\alpha$, and suppress the dependence of $\mathcal{G}$ on Φ in the notation. We also note that the solutions of (11) are uniformly bounded by a constant independent of $\theta \in h^\alpha$, specifically

$$\|\mathcal{G}(\theta)\|_\infty = \|u_{f_\theta}\|_\infty \leq \|g\|_\infty, \tag{18}$$

as follows from (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.

2.1.3. Measurement model, prior, likelihood and posterior. For the forward map $\mathscr{G}$ from (17), we now consider the measurement model

$$Y_i = \mathscr{G}(\theta)(X_i) + \varepsilon_i, \quad i = 1, \dots, N, \\ \varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0,1), \quad X_i \overset{\text{i.i.d.}}{\sim} \text{Uniform}(\mathcal{O}). \tag{19}$$

The i.i.d. random vectors

$$Z^{(N)} = (Z_i)_{i=1}^N = (Y_i, X_i)_{i=1}^N \tag{20}$$

are drawn from a product measure on $(\mathbb{R} \times \mathcal{O})^N$ that we denote by $P_\theta^N = \bigotimes_{i=1}^N P_\theta$. 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

$$p_\theta(y,x) := \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}[y - \mathscr{G}(\theta)(x)]^2\}, \quad y \in \mathbb{R}, x \in \mathcal{O}, \tag{21}$$

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

$$\ell_N(\theta) \equiv \log p_\theta^N + N \log \sqrt{2\pi} = -\frac{1}{2} \sum_{i=1}^N (Y_i - \mathscr{G}(\theta)(X_i))^2. \tag{22}$$

When using Gaussian process prior models in Bayesian statistics, a common discretisation approach is to truncate the ('Karhunen–Loève' type) expansion of the prior in a suitable basis [33, 50, 63, 92]. In our context this will mean that we truncate the series defining the random field $\mathcal{M}_\alpha$ in (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 (13), we thus consider priors

$$\theta \sim \Pi = \Pi_N \sim N(0, N^{-d/(2\alpha+d)} \Lambda_\alpha^{-1}), \quad \Lambda_\alpha = \text{diag}(\lambda_1^\alpha, \dots, \lambda_D^\alpha), \tag{23}$$

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 $\Pi(\cdot \mid Z^{(N)})$ on $\mathbb{R}^D$ then arises from data $Z^{(N)}$ in (19) via Bayes' formula, with probability density function

$$\pi(\theta \mid Z^{(N)}) \propto e^{\ell_N(\theta)} \pi(\theta) \\ \propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^N (Y_i - \mathscr{G}(\theta)(X_i))^2 - \frac{N^{d/(2\alpha+d)}}{2} \|\theta\|_{h^\alpha}^2 \right\}, \quad \theta \in \mathbb{R}^D. \tag{24}$$

2.2. Polynomial time guarantees for Bayesian posterior computation

2.2.1. Description of the algorithm. We now describe the Langevin-type algorithm targeting the posterior measure $\Pi(\cdot \mid Z^{(N)})$. It requires the choice of an initialiser θ_{init} and of constants ϵ, K, γ. Our goal is merely to exhibit its polynomial runtime and we do not attempt to optimize the constants involved.

Throughout, we use the initialiser $\theta_{\text{init}} = \theta_{\text{init}}(Z^{(N)}) \in \mathbb{R}^D$ constructed in Theorem B.6 in Section B.3 (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

$$\hat{\mathcal{B}} = \{\theta \in \mathbb{R}^D : \|\theta - \theta_{\text{init}}\|_{\mathbb{R}^D} \leq \epsilon D^{-4/d}/2\}. \tag{25}$$

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 (22). Specifically, take the cut-off function $\alpha = \alpha_\eta$ from (53) and the convex function $g = g_\eta$ from (52) with choice $\eta = \epsilon D^{-4/d}$ and $|\cdot|_1 = \|\cdot\|_{\mathbb{R}^D}$. Note that α is compactly supported and identically 1 on $\hat{\mathcal{B}}$ and that g vanishes on $\hat{\mathcal{B}}$. Then for K to be chosen, $\tilde{\ell}_N$ takes the form

$$\tilde{\ell}_N(\theta) := \alpha(\theta)\ell_N(\theta) - Kg(\theta), \quad \theta \in \mathbb{R}^D. \tag{26}$$

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

$$\log \tilde{\pi}(\theta \mid Z^{(N)}) = \tilde{\ell}_N(\theta) - N^{d/(2\alpha+d)}\|\theta\|_{h^\alpha}^2/2 + \text{const}, \quad \theta \in \mathbb{R}^D. \tag{27}$$

Note that $\tilde{\pi}(\cdot \mid Z^{(N)})$ coincides with the posterior density $\pi(\cdot \mid Z^{(N)})$ on the set $\hat{\mathcal{B}}$ up to a (random) normalising constant. The MCMC scheme we consider is then given in Algorithm 1 and the law of the resulting Markov chain $(\vartheta_k) \in \mathbb{R}^D$ will be denoted by $\mathbf{P}_{\theta_{\text{init}}}$.

Algorithm 1

Input: Initialiser $\theta_{\text{init}} \in \mathbb{R}^D$, 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 $\vartheta_1, \ldots, \vartheta_k, \ldots \in \mathbb{R}^D$.

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

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 Appendix A) targeting $\pi(\cdot \mid Z^{(N)})$ 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 densities $\pi(\cdot \mid Z^{(N)}), \tilde{\pi}(\cdot \mid Z^{(N)})$; cf. Theorem 4.14.] Note that while the ball in (25) shrinks as $D \to \infty$, 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) 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 1 in a large N, D scenario, we now make the following specific choices of the key algorithm parameters ϵ, K, γ.

Condition 2.2. *Let θ_{init} be the initialiser from Theorem B.6 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. Conditions involving θ_0. The convergence guarantees obtained below hold for high-dimensional models where D is permitted to grow polynomially in N, and under the frequentist assumption that the data $Z^{(N)}$ from (19) is generated from a fixed ground truth θ_0 inducing the law $P_{\theta_0}^N$. 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 2.9 below, reads as follows.

Condition 2.3. *For integers $d \le 3$ and $\alpha > 6$, suppose data $Z^{(N)}$ from (20) arise in the Schrödinger model (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 $\theta_{0,D} = ((\theta_0)_1, \dots, (\theta_0)_D)$,*

$$D \le c_0 N^{d/(2\alpha+d)}, \quad \|\mathcal{G}(\theta_{0,D}) - \mathcal{G}(\theta_0)\|_{L^2(\mathcal{O})} \le c_0' N^{-\alpha/(2\alpha+d)}. \tag{28}$$

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 $S \ge \|\theta_0\|_{h^\alpha}$.

2.2.3. Computational guarantees for ergodic MCMC averages. We first present a concentration inequality for ergodic averages along the Markov chain (ϑ_k). Proposition 2.4 is non-asymptotic in nature; hence its statement necessarily involves various constants whose dependence on D and N is tracked. Theorems 2.5 and 2.6 then demonstrate how the desired polynomial time computation guarantees, including Theorem 1.1, can be deduced from it.

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

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

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

$$B(\gamma) := c_1\big[\gamma D^{(d+24)/d}(\log N)^6 + \gamma^2 N D^{(d+44)/d}(\log N)^{12}\big] + 2\exp(-N^{-(d)/2\alpha+d}). \tag{29}$$

The quantity $B(\gamma)$ is an upper bound for the error incurred by the discretisation of the Langevin dynamics (see (162) below) and by the 'proxy' construction (27).

Proposition 2.4. *Assume Condition 2.3 is satisfied and consider iterates ϑ_k of the Markov chain from Algorithm 1 with $\theta_{\text{init}}, \epsilon, K, \gamma$ satisfying Condition 2.2. Then there exist constants $c_1, \dots, c_5 > 0$ such that for all $N \in \mathbb{N}$, any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$, any burn-in period*

$$J_{\text{in}} \geq \frac{\log N}{\gamma N D^{-4/d}} \times \log(D + B(\gamma)^{-1}), \tag{30}$$

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

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

The next result concerns computation of the posterior mean vector

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

by ergodic averages

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

within prescribed precision level ε. For convenience we assume $\varepsilon \geq N^{-P}$ for some $P > 0$, which is natural in view of the statistical error to be considered in Theorem 2.6 below. To this end, we make an explicit choice for the step size parameter

$$\gamma = \gamma_\varepsilon = \min\left(\frac{\varepsilon^2}{D^{(d+24)/d}}, \frac{\varepsilon}{\sqrt{N} \, D^{(22+d/2)/d}}, \frac{1}{N D^{8/d}}\right) \times (\log N)^{-7}. \tag{31}$$

Theorem 2.5. *Assume Condition 2.3 is satisfied. Fix $P > 0$ and let $\varepsilon \geq N^{-P}$. Consider iterates ϑ_k of the Markov chain from Algorithm 1 with $\theta_{\text{init}}, \epsilon, K$ satisfying Condition 2.2 and with $\gamma = \gamma_\varepsilon$ as in (31). Then there exist $c_6, c_7, c_8 > 0$ and at most polynomially growing constants*

$$g_{D,N,\varepsilon} = O(D^{\bar{b}_1} N^{\bar{b}_2} \varepsilon^{-\bar{b}_3}), \quad \bar{b}_1, \bar{b}_2, \bar{b}_3 > 0, \tag{32}$$

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

$$\mathbf{P}_{\theta_{\text{init}}}(\|\bar{\theta}_{J_{\text{in}}}^J - E^\Pi[\theta \mid Z^{(N)}]\|_{\mathbb{R}^D} \geq \varepsilon) \leq c_6 D \exp\left(-\frac{J}{g_{D,N,\varepsilon}}\right). \tag{33}$$

Theorem 2.5 implies that for $J_{\text{in}} \wedge J \gg g_{D,N,\varepsilon} \times \log D$, one can compute the posterior mean vector within precision $\varepsilon > 0$ with probability as close to 1 as desired. Using this and Theorem B.6 (whose hypotheses are implied by those of Theorem 2.5), we have in particular also proven Theorem 1.1. Similar bounds for computation of $E^\Pi[H \mid Z^{(N)}]$ can be obtained as long as $\|H\|_{\text{Lip}}$ 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 2.5 with a statistical rate of convergence of $E^\Pi[\theta \mid Z^{(N)}]$ to θ_0, obtained by adapting recent results from [74] to the present situation.

Theorem 2.6. *Consider the setting of Theorem 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 (32) and on events $\mathcal{E}_N$ of probability $P^N_{\theta_0}(\mathcal{E}_N) \geq 1 - c_9 \exp(-c_{10} N^{d/(2\alpha+d)})$,*

$$\mathbf{P}_{\theta_{\mathrm{init}}}(\|\bar{\theta}^J_{J_{\mathrm{in}}} - \theta_0\|_{\ell^2} \geq \varepsilon) \leq c_{12} \exp\left(-\frac{J}{4 g_{D,N,\varepsilon}}\right). \tag{34}$$

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 [77]), it appears unclear how to obtain this rate when F_θ is discretised by means of the (for the purposes of the present paper essential) spectral decomposition of the Dirichlet-Laplacian from Section 2.1.1. The difficulty arises with the approximation theory of the space $H^\alpha_c(\mathcal{O})$ (equal to the completion of $C^\infty_c(\mathcal{O})$ in $H^\alpha(\mathcal{O})$) and is not discussed further here.

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 $\Pi(\cdot \mid Z^{(N)})$, for instance in the Wasserstein distance from (9).

Theorem 2.7. *Assume Condition 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 1 with $\theta_{\mathrm{init}}, \epsilon, K, \gamma$ satisfying Condition 2.2, and let $B(\gamma), c_1$ be as in (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 $P^N_{\theta_0}(\mathcal{E}_N) \geq 1 - c_{13}\exp(-c_{14} N^{d/(2\alpha+d)})$ and for all $N \in \mathbb{N}$, the following holds:*

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

$$W_2^2(\mathcal{L}(\vartheta_k), \Pi(\cdot \mid Z^{(N)})) \leq c_{15} D^{2\alpha/d}(1 - c_{16} N D^{-4/d}\gamma)^k_+ + B(\gamma). \tag{35}$$

(ii) *For any 'precision level' $\varepsilon \geq N^{-P}$ and for $\gamma = \gamma_\varepsilon$ from (31), there exists*

$$k_{\mathrm{mix}} = O(N^{\tilde{b}_1} D^{\tilde{b}_2} \varepsilon^{-\tilde{b}_3}), \quad \tilde{b}_1, \tilde{b}_2, \tilde{b}_3 > 0, \tag{36}$$

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

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

The first term on the right hand side of (35) characterises the rate of geometric convergence towards equilibrium of (ϑ_k); the factor $N D^{-4/d}\gamma$ 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 (31), part (ii) further establishes 'polynomial-time' mixing of the MCMC scheme towards the posterior measure.

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

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

by a classical gradient (ascent) method applied to the 'proxy' posterior surface (27).

Theorem 2.8. *Assume Condition 2.3 is satisfied and let θ_{init} denote the initialiser from Theorem B.6. For $k = 0, 1, 2, \ldots,$ consider the gradient algorithm*

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

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) *There exists a unique maximiser $\hat{\theta}_{\mathrm{MAP}}$ of $\pi(\theta \mid Z^{(N)})$ over $\mathbb{R}^D$.*

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

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

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

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

Remark 2.9 (about Condition 2.3). In principle the upper bound for D required in Condition 2.3 could be replaced by general conditions on D (like those from Lemma 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 2.6. [In fact, the second requirement in (28) can be checked for $\theta_0 \in h^\alpha$ and $D \simeq N^{d/(2\alpha+d)}$, since $\mathscr{G}$ 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 4.14) that the proxy posterior measure $\tilde{\Pi}(\cdot \mid Z^{(N)})$ contracts about a $\|\cdot\|_{\mathbb{R}^D}$-neighbourhood of θ_0 of radius $D^{-4/d}$ on which the information in the Schrödinger model has a stable behaviour (see (115)). It is also required for our initialiser θ_{init} to lie in this neighbourhood (Theorem B.6). 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.

3. General theory for random design regression

In proving the results from Section 2, we will first develop some theory which applies to general non-linear regression models. We thus consider in this section the measurement model (3) for a general forward model $\mathscr{G}$ 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 has a subspace $\mathbb{R}^D \subseteq \Theta$, $D \in \mathbb{N}$. Let $\mathcal{O}$ be a Borel subset of $\mathbb{R}^d$, $d \geq 1$, and consider a model of regression functions $\{\mathscr{G}(\theta) : \theta \in \Theta\}$ via a Borel-measurable forward map $\mathscr{G} : \Theta \to C(\mathcal{O})$. While we regard each $\mathscr{G}(\theta)$ as a continuous *real-valued* function, the results of this section readily extend to vector or matrix fields over manifolds $\mathcal{O}$; see Remark 3.11. Our data is given by $Z_i = (Y_i, X_i)$ arising from

$$Y_i = \mathscr{G}(\theta)(X_i) + \varepsilon_i, \quad i = 1, \dots, N, \tag{37}$$

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

$$\begin{aligned} \ell_N(\theta) &\equiv \ell_N(\theta, Z^{(N)}) = -\frac{1}{2}\sum_{i=1}^N [Y_i - \mathscr{G}(\theta)(X_i)]^2, \\ \ell(\theta) &\equiv \ell(\theta, Z) = -\frac{1}{2}[Y - \mathscr{G}(\theta)(X)]^2, \end{aligned} \tag{38}$$

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 $\Pi(\cdot \mid Z^{(N)})$ with probability density $\pi(\cdot \mid Z^{(N)})$ on $\mathbb{R}^D$ as in (24).

The main results of this section are Theorems 3.7 and 3.8, providing convergence guarantees for a Langevin sampling method for the posterior distribution that depend polynomially on the model dimension D and the 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

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

On $\mathcal{E}_{\text{conv}}$ the negative log-likelihood $-\ell_N(\theta)$ will be strongly convex in some region $\mathcal{B} \subseteq \mathbb{R}^D$, while $\mathcal{E}_{\text{init}}$ is the event that allows one to initialise the method at some (data-driven) $\theta_{\text{init}} = \theta_{\text{init}}(Z^{(N)})$ in that set $\mathcal{B}$. Finally, intersection with $\mathcal{E}_{\text{wass}}$ further guarantees that the posterior measure $\Pi(\cdot \mid Z^{(N)})$ is close in Wasserstein distance to a *globally* log-concave surrogate probability measure $\tilde{\Pi}(\cdot \mid Z^{(N)})$ which locally coincides with $\Pi(\cdot \mid Z^{(N)})$ up to proportionality factors. In applying the results of this section to a concrete sampling problem, one needs to show that all the events $\mathcal{E}_{\text{conv}}$, $\mathcal{E}_{\text{init}}$, $\mathcal{E}_{\text{wass}}$ have sufficiently high frequentist $P_{\theta_0}^N$-probability, where θ_0 is the ground truth parameter generating data (37). For the event $\mathcal{E}_{\text{conv}}$ we provide a generic method in Lemma 3.4, based on a stability estimate for the linearisation of the map $\mathscr{G}$ combined with high-dimensional concentration of measure techniques. Techniques for controlling the respective probabilities of $\mathcal{E}_{\text{init}}$ and $\mathcal{E}_{\text{wass}}$ are discussed in Remark 3.10.

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

Definition 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

$$\mathcal{B} := \{\theta \in \mathbb{R}^D : |\theta - \theta^*|_1 < \eta\}. \tag{39}$$

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.

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 $\mathcal{G}, \ell, \ell_N$ viewed as maps on the subspace $\mathbb{R}^D \subseteq \Theta$. Specifically we shall write $(\nabla\mathcal{G}(\theta)(x) : x \in \mathcal{O})$ and $(\nabla^2\mathcal{G}(\theta)(x) : x \in \mathcal{O})$ for the 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 assumption 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 (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 3.2 (Local regularity). *Let $\mathcal{B}$ be given in (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 $k_0, k_1, k_2 > 0$,*

$$\begin{aligned} \sup_{\theta\in\mathcal{B}} \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_\infty &\leq k_0, \\ \sup_{\theta\in\mathcal{B}} \|\nabla\mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} &\leq k_1, \\ \sup_{\theta\in\mathcal{B}} \|\nabla^2\mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} &\leq k_2. \end{aligned} \tag{40}$$

(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 (38). To motivate it, notice that

$$-\nabla^2 \ell(\theta, Z) = [\nabla \mathcal{G}(\theta)(X)][\nabla \mathcal{G}(\theta)(X)]^T + [\mathcal{G}(\theta)(X) - Y]\nabla^2[\mathcal{G}(\theta)(X)]. \quad (41)$$

If the design distribution P^X is uniform on a bounded domain $\mathcal{O}$ (say, of unit volume) then at $\theta = \theta_0$, the $E_{\theta_0}^N$-expectation of the last expression can be represented as

$$v^T E_{\theta_0}[-\nabla^2 \ell(\theta_0, Z)]v = \|\nabla \mathcal{G}(\theta_0)^T v\|_{L^2(\mathcal{O})}^2, \quad v \in \mathbb{R}^D. \quad (42)$$

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 (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, such a stability estimate indeed follows from elliptic PDE theory (see Lemma 4.7), and the recent reference [16] verifies this condition for the non-Abelian X-ray transform considered in [74].

Note that the Hessian $E_{\theta_0}[-\nabla^2 \ell(\theta, Z)]$ is symmetric (by (41) and Assumption 3.2 (i)), and recall that $\lambda_{\min}(A)$ denotes the smallest eigenvalue of a symmetric matrix A.

Assumption 3.3 (Local curvature). *Let $\mathcal{B}$ be given in (39) and let $\ell : \mathbb{R}^D \to \mathbb{R}$ be as in (38).*

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

$$\inf_{\theta \in \mathcal{B}} \lambda_{\min}(E_{\theta_0}[-\nabla^2 \ell(\theta, Z)]) \ge c_{\min}. \quad (43)$$

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

$$\sup_{\theta \in \mathcal{B}} \big[|E_{\theta_0}[\ell(\theta, Z)]| + \|E_{\theta_0}[\nabla \ell(\theta, Z)]\|_{\mathbb{R}^D} + \|E_{\theta_0}[\nabla^2 \ell(\theta, Z)]\|_{\mathrm{op}}\big] \le c_{\max}. \quad (44)$$

The following lemma, which is based on concentration of measure arguments, shows that the local 'average' curvature bound in (43) carries over to the 'observed' log-likelihood function, with high frequentist $P_{\theta_0}^N$-probability, and whenever $D \le \mathcal{R}_N$, 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 3.4. *Suppose that the data arises from (37) with $\ell_N : \mathbb{R}^D \to \mathbb{R}$ given by (38). Suppose Assumptions 3.2 and 3.3 are satisfied. There exists a universal constant $C > 0$ such that if*

$$\mathcal{R}_N := CN \min\left\{\frac{c_{\min}^2}{C_{\mathcal{G}}^2 \eta^2}, \frac{c_{\min}}{C_{\mathcal{G}} \eta}, \frac{c_{\min}^2}{C_{\mathcal{G}}'^2}, \frac{c_{\min}}{k_2}, \frac{c_{\max}^2}{C_{\mathcal{G}}''^2 \eta^2}, \frac{c_{\max}}{C_{\mathcal{G}}'' \eta}, \frac{c_{\max}^2}{C_{\mathcal{G}}'''^2}, \frac{c_{\max}}{k_0 + k_1}\right\}, \quad (45)$$

where

$$C_{\mathscr{G}} := k_0 m_2 + k_1 m_1 + k_2 m_0 + m_2, \quad C'_{\mathscr{G}} := k_1^2 + k_0 k_2 + k_2,$$
$$C''_{\mathscr{G}} := k_0 m_1 + k_1 m_0 + m_1 + k_0 m_0 + m_0, \quad C'''_{\mathscr{G}} = k_0 k_1 + k_1 + k_0^2 + k_0, \tag{46}$$

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

$$P_{\theta_0}^N \Big(\inf_{\theta \in \mathcal{B}} \lambda_{\min}[-\nabla^2 \ell_N(\theta, Z^{(N)})] < \tfrac{1}{2} N c_{\min} \Big) \leq 8 e^{-\mathcal{R}_N}, \tag{47}$$

as well as

$$P_{\theta_0}^N \Big(\sup_{\theta \in \mathcal{B}} [|\ell_N(\theta, Z^{(N)})| + \|\nabla \ell_N(\theta, Z^{(N)})\|_{\mathbb{R}^D} + \|\nabla^2 \ell_N(\theta, Z^{(N)})\|_{\mathrm{op}}] > N(5c_{\max} + 1) \Big) \leq 24 e^{-\mathcal{R}_N} + e^{-N/8}. \tag{48}$$

Inspection of the proof (given in Section 3.4) shows that for the first inequality (47), the terms involving $c_{\max}$ can be removed from the definition of $\mathcal{R}_N$. In what follows, we will restrict considerations to the event

$$\mathcal{E}_{\mathrm{conv}} := \Big\{ \inf_{\theta \in \mathcal{B}} \lambda_{\min}[-\nabla^2 \ell_N(\theta)] \geq N c_{\min}/2 \Big\} \cap \Big\{ \sup_{\theta \in \mathcal{B}} [|\ell_N(\theta)| + \|\nabla \ell_N(\theta)\|_{\mathbb{R}^D} + \|\nabla^2 \ell_N(\theta)\|_{\mathrm{op}}] \leq N(5c_{\max} + 1) \Big\}, \tag{49}$$

whose $P_{\theta_0}^N$-probability is controlled by Lemma 3.4.

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* (*strong*) *log-concavity* (on $\mathbb{R}^D$) of the target measure to be sampled from; see Appendix A. 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 (39), and which will be shown to be globally log-concave on the event $\mathcal{E}$ from (60) below.

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

$$\mathcal{E}_{\mathrm{init}} := \{ |\theta_{\mathrm{init}} - \theta^*|_1 \leq \eta/8 \}, \tag{50}$$

where θ_{init} belongs to the region $\mathcal{B}$. That such initialisation is possible (i.e., $\mathcal{E}_{\mathrm{init}}$ has sufficiently high $P_{\theta_0}^N$-probability for appropriate $\eta > 0$) is proved for the Schrödinger model in Theorem B.6.

We require two auxiliary functions, g_η (globally convex) and α_η (a 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

$$\tilde{\gamma}_\eta(t) := \begin{cases} 0 & \text{if } t < 5\eta/8, \\ (t - 5\eta/8)^2 & \text{if } t \geq 5\eta/8, \end{cases} \qquad \gamma_\eta(t) := [\varphi_{\eta/8} * \tilde{\gamma}_\eta](t), \tag{51}$$

where $*$ denotes convolution, and

$$g_\eta : \mathbb{R}^D \to [0, \infty), \quad g_\eta(\theta) := \gamma_\eta(|\theta - \theta_{\text{init}}|_1). \tag{52}$$

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

$$\alpha_\eta : \mathbb{R}^D \to [0, 1], \quad \alpha_\eta(\theta) = \alpha(|\theta - \theta_{\text{init}}|_1/\eta). \tag{53}$$

Definition 3.5. For the auxiliary functions g_η, α_η from (52), (53) and $K > 0$, we define the surrogate likelihood function $\tilde{\ell}_N$ by

$$\tilde{\ell}_N : \mathbb{R}^D \to \mathbb{R}, \quad \tilde{\ell}_N(\theta) := \alpha_\eta(\theta)\ell_N(\theta) - K g_\eta(\theta). \tag{54}$$

When the choice of the constant $K > 0$ is large enough relative to $c_{\max}$ from Assumption 3.2, the following global convexity property can be proved for $\tilde{\ell}_N$ (see Appendix B for a proof).

Proposition 3.6. *On the event $\mathcal{E}_{\text{conv}} \cap \mathcal{E}_{\text{init}}$ (cf. (49), (50)), when $\tilde{\ell}_N$ from (54) is defined with any constant K satisfying*

$$K \geq CN(c_{\max} + 1) \cdot \frac{1 + \lambda_{\max}(M)/\eta^2}{\lambda_{\min}(M)} \tag{55}$$

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

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

Moreover, $\tilde{\ell}_N \in C^2(\mathbb{R}^D)$ and

$$\inf_{\theta \in \mathbb{R}^D} \lambda_{\min}(-\nabla^2 \tilde{\ell}_N(\theta)) \geq N c_{\min}/2, \tag{56}$$

as well as

$$\|\nabla \tilde{\ell}_N(\theta) - \nabla \tilde{\ell}_N(\bar{\theta})\|_{\mathbb{R}^D} \leq 7K\lambda_{\max}(M)\|\theta - \bar{\theta}\|_{\mathbb{R}^D}, \quad \theta, \bar{\theta} \in \mathbb{R}^D. \tag{57}$$

3.3. Non-asymptotic bounds for Bayesian posterior computation

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

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

arising from data (37) with log-likelihood (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 Appendix A) 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 $\xi_k \overset{\text{i.i.d.}}{\sim} N(0, I_{D\times D})$, define a Markov chain as

$$\begin{aligned} \vartheta_0 &= \theta_{\text{init}}, \\ \vartheta_{k+1} &= \vartheta_k + \gamma[\nabla\tilde{\ell}_N(\vartheta_k) - \Sigma^{-1}\vartheta_k] + \sqrt{2\gamma}\,\xi_{k+1}, \quad k = 0, 1, \ldots. \end{aligned} \tag{58}$$

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 $\mathbf{P}_{\theta_{\text{init}}}, \mathbf{E}_{\theta_{\text{init}}}$ respectively. The invariant measure of the underlying continuous time Langevin diffusion equals the *surrogate posterior distribution* given by

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

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

$$\mathcal{E}_{\text{wass}}(\rho) := \{W_2^2(\Pi(\cdot \mid Z^{(N)}), \tilde{\Pi}(\cdot \mid Z^{(N)})) \le \rho/2\}. \tag{59}$$

For the Schrödinger model this is achieved in Theorem 4.14, for ρ decaying exponentially in N, using that most of the posterior mass (and its mode) concentrates on the set $\mathcal{B}$ from (39), and the ideas underlying this proof extend to general settings; see Remark 3.10.

Our first result consists of a global Wasserstein approximation of $\Pi(\cdot \mid Z^{(N)})$ by the law $\mathcal{L}(\vartheta_k)$ on $\mathbb{R}^D$ of the k-th iterate ϑ_k arising from (58).

Theorem 3.7 (Non-asymptotic Wasserstein mixing). *Suppose that the model given by (37)–(38) fulfills Assumptions 3.2 and 3.3 for some $0 < \eta \le 1$, let $D, N \in \mathbb{N}$ be such that $D \le \mathcal{R}_N$ with $\mathcal{R}_N$ from (45), and let K be as in (55). Further define the constants*

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

Then for any $0 < \gamma \le 1/\Lambda$ and any $\rho > 0$ the algorithm $(\vartheta_k : k \ge 0)$ from (58) satisfies, on the event (i.e., measurable subset of $(\mathbb{R} \times \mathcal{O})^N$)

$$\mathcal{E} := \mathcal{E}_{\text{conv}} \cap \mathcal{E}_{\text{init}} \cap \mathcal{E}_{\text{wass}}(\rho), \tag{60}$$

(with $\mathcal{E}_{\text{conv}}, \mathcal{E}_{\text{init}}, \mathcal{E}_{\text{wass}}(\rho)$ defined in (49), (50), (59), respectively), and all $k \ge 0$,

$$W_2^2(\mathcal{L}(\vartheta_k), \Pi(\cdot|Z^{(N)})) \le \rho + b(\gamma) + 4(\tau(\Sigma, M, R) + D/m)(1 - \gamma m/2)^k, \tag{61}$$

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

$$b(\gamma) = c_1\left[\frac{\gamma D\Lambda^2}{m^2} + \frac{\gamma^2 D\Lambda^4}{m^3}\right], \quad \tau(\Sigma, M, R) = c_2\kappa(\Sigma)\left[1 + \frac{\eta^2}{\lambda_{\min}(M)} + R^2\right]. \tag{62}$$

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 $J_{\text{in}} \in \mathbb{N}$. Specifically, if we define, for any $H : \mathbb{R}^D \to \mathbb{R}$ integrable with respect to $\Pi(\cdot \mid Z^{(N)})$, the random variable

$$\hat{\pi}^J_{J_{\text{in}}}(H) = \frac{1}{J} \sum_{k=J_{\text{in}}+1}^{J_{\text{in}}+J} H(\vartheta_k), \tag{63}$$

we obtain the following non-asymptotic concentration bound.

Theorem 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*

$$J_{\text{in}} \geq \frac{c_3}{m\gamma} \times \log\left(1 + \frac{1}{\rho + b(\gamma)} + \tau(\Sigma, M, R) + \frac{D}{m}\right), \tag{64}$$

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

$$t \geq \sqrt{8}\,\|H\|_{\text{Lip}} \sqrt{\rho + b(\gamma)} \tag{65}$$

and on the event $\mathcal{E}$ from (60), we have

$$\mathbf{P}_{\theta_{\text{init}}}\big(|\hat{\pi}^J_{J_{\text{in}}}(H) - E^\Pi[H \mid Z^{(N)}]| \geq t\big) \leq 2\exp\left(-c_4 \frac{t^2 m^2 J\gamma}{\|H\|^2_{\text{Lip}}(1 + 1/(mJ\gamma))}\right). \tag{66}$$

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

Corollary 3.9. *In the setting of Theorem 3.8, if we define*

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

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

$$\mathbf{P}_{\theta_{\text{init}}}\big(\|\bar{\theta}^J_{J_{\text{in}}} - E^\Pi[\theta \mid Z^{(N)}]\|_{\mathbb{R}^D} \geq t\big) \leq 2D\exp\left(-c_5 \frac{t^2 m^2 J\gamma}{D(1 + 1/(mJ\gamma))}\right). \tag{67}$$

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

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

then the overall number of required MCMC iterations $J_{\text{in}} + J$ depends polynomially on the quantities $N, D, m^{-1}, \Lambda, \varepsilon^{-1}$. When the last three constants exhibit at most polynomial growth in N, D (as is the case for the Schrödinger equation treated in Section 2), we can deduce that polynomial-time computation of such posterior characteristics is feasible, on the event $\mathcal{E}$ from (60) at computational cost $J_{\text{in}} + J = O(N^{b_1} D^{b_2} \varepsilon^{-b_3}), b_1, b_2, b_3 > 0$, with $\mathbf{P}_{\theta_{\text{init}}}$-probability as close to 1 as desired.

Remark 3.10 (about the events $\mathcal{E}_{\mathrm{init}}, \mathcal{E}_{\mathrm{wass}}$). Controlling the probability of the events $\mathcal{E}_{\mathrm{init}}, \mathcal{E}_{\mathrm{wass}}$ (featuring in the definition of $\mathcal{E}$ in (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, this is achieved in Sections B.3 and 4.2, respectively. 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 $\mathcal{E}_{\mathrm{init}}, \mathcal{E}_{\mathrm{wass}}$, *global* 'stability' and 'range' properties of the map $\mathcal{G}$ will play a role. In contrast, Assumptions 3.2, 3.3 employed in this section are 'local' in the sense that they concern properties of $\mathcal{G}$ on $\mathcal{B}$ from (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 paper. Following the ideas in Section 4.2 below, the recent contribution [16] provides a set of conditions on $\mathcal{G}$ under which a log-concave approximation of the posterior measure similar to Theorem 4.14 holds true.

Remark 3.11 (Extensions to vector-valued data). The key results of this section apply to other settings (e.g. in [74, 85]) 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 manifold $\mathcal{M}$ (possibly with boundary) into a finite-dimensional inner product space V of fixed finite dimension $\dim(V) < \infty$. If we assume that the statistical errors $(\varepsilon_i : i = 1, \dots, N)$ in equation (37) are i.i.d. $N(0, \mathrm{Id}_V)$ in V, then the log-likelihood function of the model is not given by (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 3.2 with the obvious modification of the norms there for V-valued maps, and if Assumption 3.3 holds for the preceding definition of $\ell(\theta)$, then the conclusion of Lemma 3.4 remains valid as stated (see also [16, proof of Lemma 5.6]).

3.4. *Proof of Lemma* 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

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

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):

$$\begin{aligned}
-\ell(\theta, Z) &= \tfrac{1}{2}[Y - \mathcal{G}^X(\theta)]^2 = \tfrac{1}{2}[\mathcal{G}^X(\theta_0) + \varepsilon - \mathcal{G}^X(\theta)]^2, \\
-\nabla\ell(\theta, Z) &= [\mathcal{G}^X(\theta) - \mathcal{G}^X(\theta_0) - \varepsilon]\nabla\mathcal{G}^X(\theta), \\
-\nabla^2\ell(\theta, Z) &= \nabla\mathcal{G}^X(\theta)\nabla\mathcal{G}^X(\theta)^T + [\mathcal{G}^X(\theta) - \mathcal{G}^X(\theta_0) - \varepsilon]\nabla^2\mathcal{G}^X(\theta), \\
-E_{\theta_0}[\ell(\theta, Z)] &= \tfrac{1}{2} + \tfrac{1}{2}E^X[(\mathcal{G}^X(\theta_0) - \mathcal{G}^X(\theta))^2],
\end{aligned} \tag{68}$$

where we note that by Assumption 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

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

3.4.1. Proof of (47). Let us write $\bar\ell_N := \ell_N/N$. Then, by a standard inequality due to Weyl as well as Assumption 3.3, we see that for any $\theta \in \mathcal{B}$,

$$\begin{aligned}\lambda_{\min}[-\nabla^2\bar\ell_N(\theta)] &\geq \lambda_{\min}(E_{\theta_0}[-\nabla^2\ell(\theta)]) - \|\nabla^2\bar\ell_N(\theta) - E_{\theta_0}[\nabla^2\ell(\theta)]\|_{\mathrm{op}}\\ &\geq c_{\min} - \|\nabla^2\bar\ell_N(\theta) - E_{\theta_0}[\nabla^2\ell(\theta)]\|_{\mathrm{op}}. \end{aligned}\tag{69}$$

Hence we deduce

$$\begin{aligned}&P_{\theta_0}^N\Big(\inf_{\theta\in\mathcal{B}} \lambda_{\min}[\nabla^2\ell_N(\theta, Z)] < Nc_{\min}/2\Big)\\ &\quad\leq P_{\theta_0}^N\big(\|\nabla^2\bar\ell_N(\theta) - E_{\theta_0}[\nabla^2\ell(\theta)]\|_{\mathrm{op}} \geq c_{\min}/2 \text{ for some } \theta\in\mathcal{B}\big)\\ &\quad\leq P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\sup_{v:\|v\|_{\mathbb{R}^D}\leq 1}\big|v^T\big(\nabla^2\bar\ell_N(\theta) - E_{\theta_0}[\nabla^2\ell(\theta)]\big)v\big| \geq c_{\min}/2\Big)\\ &\quad= P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\sup_{v:\|v\|_{\mathbb{R}^D}\leq 1}|P_N(g_{v,\theta})| \geq c_{\min}/2\Big),\end{aligned}\tag{70}$$

where

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

The next step is to reduce the supremum over $\{v : \|v\|_{\mathbb{R}^D} \leq 1\}$ 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 $\{v : \|v\|_{\mathbb{R}^D} \leq 1\}$, and let v_i with $\|v_i\|_{\mathbb{R}^D} \leq 1$ be the centre points of a minimal covering. Thus for any such $v \in \mathbb{R}^D$ there exists an index i such that

$$\|v - v_i\|_{\mathbb{R}^D} \leq \rho.$$

Hence, with shorthand

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

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

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

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

$$\sup_{v:\, \|v\|_{\mathbb{R}^D} \le 1} v^T M_\theta v \le 2 \max_{i=1,\dots,N(1/4)} v_i^T M_\theta v_i. \tag{71}$$

Since the covering (v_i) is independent of θ, we can further estimate the right hand side of (70) by a union bound to the effect that

$$\begin{aligned} &P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \sup_{v:\, \|v\|_{\mathbb{R}^D} \le 1} |v^T M_\theta v| \ge c_{\min}/2\Big) \\ &\quad\le N(1/4)\cdot \sup_{v:\, \|v\|_{\mathbb{R}^D} \le 1} P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} |v^T M_\theta v| \ge c_{\min}/4\Big) \\ &\quad\le N(1/4) \\ &\qquad\cdot \sup_{v:\, \|v\|_{\mathbb{R}^D} \le 1} \Big[P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} |P_N(g_{v,\theta}-g_{v,\theta^*})| \ge c_{\min}/8\Big) + P_{\theta_0}^N(|P_N(g_{v,\theta^*})| \ge c_{\min}/8)\Big], \end{aligned} \tag{72}$$

where we recall that θ^* is the centre point of the set $\mathcal{B}$ from (39). For the rest of the proof, we fix any $v \in \mathbb{R}^D$ with $\|v\|_{\mathbb{R}^D} \le 1$. Next, we use (68) to decompose the 'uncentred' part of $g_{v,\theta}$ as

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

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 $\tilde g_{v,\theta}^I$ as

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

We can therefore bound the right hand side of (72) by

$$\begin{aligned} &N\Big(\frac14\Big)\cdot \sup_{v:\, \|v\|_{\mathbb{R}^D} \le 1} \Bigg[P_{\theta_0}^N\Bigg(\sup_{\theta\in\mathcal{B}}\Bigg|\frac1N\sum_{i=1}^N (g_{v,\theta}^I - g_{v,\theta^*}^I)(X_i)\Bigg| \ge \frac{c_{\min}}{16}\Bigg) \\ &\qquad\qquad + P_{\theta_0}^N\Bigg(\Bigg|\frac1N\sum_{i=1}^N g_{v,\theta^*}^I(X_i)\Bigg| \ge \frac{c_{\min}}{16}\Bigg) \\ &+P_{\theta_0}^N\Bigg(\sup_{\theta\in\mathcal{B}}\Bigg|\frac1N\sum_{i=1}^N \varepsilon_i(g_{v,\theta}^{II}-g_{v,\theta^*}^{II})(X_i)\Bigg| \ge \frac{c_{\min}}{16}\Bigg)+P_{\theta_0}^N\Bigg(\Bigg|\frac1N\sum_{i=1}^N \varepsilon_i g_{v,\theta^*}^{II}(X_i)\Bigg| \ge \frac{c_{\min}}{16}\Bigg)\Bigg] \\ &\qquad\qquad =: N\big(\tfrac14\big)\cdot(i + ii + iii + iv). \end{aligned}$$

We now use empirical process techniques (Lemma 3.12 and also Hoeffding's inequality) to bound the preceding probabilities.

Terms i and ii. In order to apply Lemma 3.12 to term i, we require some preparations. By the definition of $\tilde{g}^I_{v,\theta}$ and of the operator norm $\|\cdot\|_{\mathrm{op}}$, 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 3.2, we find that for any $\theta, \bar\theta \in \mathcal{B}$,

$$\begin{aligned}
\|\tilde{g}^I_{v,\theta} - \tilde{g}^I_{v,\bar\theta}\|_\infty &\le \big\| \big[\nabla\mathcal{G}(\theta)\nabla\mathcal{G}(\theta)^T + [\mathcal{G}(\theta) - \mathcal{G}(\theta_0)]\nabla^2\mathcal{G}(\theta)\big] \\
&\qquad - \big[\nabla\mathcal{G}(\bar\theta)\nabla\mathcal{G}(\bar\theta)^T + [\mathcal{G}(\bar\theta) - \mathcal{G}(\theta_0)]\nabla^2\mathcal{G}(\bar\theta)\big]\big\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \\
&\le \|[\nabla\mathcal{G}(\theta) - \nabla\mathcal{G}(\bar\theta)][\nabla\mathcal{G}(\theta) + \nabla\mathcal{G}(\bar\theta)]^T\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \\
&\quad + \|[\mathcal{G}(\theta) - \mathcal{G}(\bar\theta)]\nabla^2\mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \\
&\quad + \|[\mathcal{G}(\bar\theta) - \mathcal{G}(\theta_0)][\nabla^2\mathcal{G}(\theta) - \nabla^2\mathcal{G}(\bar\theta)]\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \\
&\le 2m_1k_1|\theta - \bar\theta|_1 + m_0k_2|\theta - \bar\theta|_1 + m_2k_0|\theta - \bar\theta|_1 \\
&\le 2C_{\mathcal{G}}|\theta - \bar\theta|_1. \qquad (73)
\end{aligned}$$

In particular, by (39) we obtain the uniform bound

$$\sup_{\theta\in\mathcal{B}} \|g^I_{v,\theta} - g^I_{v,\theta^*}\|_\infty \le 2\sup_{\theta\in\mathcal{B}} \|\tilde{g}^I_{v,\theta}(X) - \tilde{g}^I_{v,\theta^*}\|_\infty \le 4C_{\mathcal{G}}|\theta - \theta^*|_1 \le 4C_{\mathcal{G}}\eta. \qquad (74)$$

We introduce the rescaled function class

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

which has envelope and variance proxy bounded as

$$\sup_{\theta\in\mathcal{B}} \|h^I_\theta\|_\infty \le 1/4 \equiv U, \quad \sup_{\theta\in\mathcal{B}} (E_{\theta_0}[h^I_\theta(X)^2])^{1/2} \le 1/4 \equiv \sigma. \qquad (75)$$

Next, if

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

then using (73) we have

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

Thus for any $\rho \in (0, 1)$, using [44, Proposition 4.3.34], we obtain

$$N(\mathcal{H}^I, d_2, \rho) \le N(\mathcal{H}^I, d_\infty, \rho) \le N(\mathcal{B}, |\cdot|_1/\eta, \rho) \le (3/\rho)^D. \qquad (76)$$

For any $A \ge 2$ we have

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

(see [44, p. 190] 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

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

Now, an application of Lemma 3.12 below implies that for any $x \geq 1$ and some universal constant $L' > 0$, we have

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N h_\theta^I(X_i)\Big| \geq L'[\sqrt{D} + \sqrt{x} + (D+x)/\sqrt{N}]\Big) \leq 2e^{-x}. \tag{77}$$

By the definition of g_{v,θ^*}^I we also have

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

and hence by Hoeffding's inequality [44, Theorem 3.1.2],

$$ii \leq 2\exp\Big(-\frac{2Nc_{\min}^2}{256\cdot 4(k_1^2+k_0k_2)^2}\Big) \leq 2\exp\Big(-\frac{Nc_{\min}^2}{512C_{\mathcal{G}}'^2}\Big). \tag{78}$$

Now if we define

$$\mathcal{R}_N^{2,I} := CN\min\Big\{\frac{c_{\min}^2}{C_{\mathcal{G}}^2\eta^2}, \frac{c_{\min}}{C_{\mathcal{G}}\eta}, \frac{c_{\min}^2}{C_{\mathcal{G}}'^2}\Big\}, \tag{79}$$

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

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

whenever $C > 0$ is small enough. Therefore, combining (77) and (78), and using the definitions of the term i and of h_θ^I, we obtain

$$ii + i \leq 2e^{-4\mathcal{R}_N^{2,I}} + P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N h_\theta^I(X_i)\Big| \geq \frac{c_{\min}\sqrt{N}}{256C_{\mathcal{G}}\eta}\Big) \leq 4e^{-4\mathcal{R}_N^{2,I}}. \tag{80}$$

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

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

which also yields the envelope bound

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

Now the rescaled function class

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

admits envelopes

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

Thus defining

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

we have

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

Therefore, just as with the bounds obtained for term i, we have $N(\mathcal{H}^{II}, d_2, \rho) \le N(\mathcal{H}^{II}, d_\infty, \rho) \le (3/\rho)^D$ and thus, by Lemma 3.12 below,

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt N}\Big|\sum_{i=1}^N \varepsilon_i h_\theta^{II}(X_i)\Big| \ge L'[\sqrt D + \sqrt x + (D+x)/\sqrt N]\Big) \le 2e^{-x}, \quad x \ge 1. \tag{81}$$

Moreover, by the hypotheses, $\|g_{v,\theta*}^{II}\|_\infty \le k_2$, and hence, invoking the Bernstein inequality (96) with $U = \sigma \equiv k_2$, we deduce

$$P_{\theta_0}^N\Big(\Big|\frac{1}{\sqrt N}\sum_{i=1}^N \varepsilon_i g_{v,\theta*}^{II}(X_i)\Big| \ge k_2\sqrt{2x} + \frac{k_2 x}{3\sqrt N}\Big) \le 2e^{-x}, \quad x > 0. \tag{82}$$

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 deduce that for $C > 0$ small enough and any $D \le \mathcal{R}_N^{2,II}$,

$$\begin{aligned} iii + iv &\le P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt N}\Big|\sum_{i=1}^N \varepsilon_i h_\theta^{II}(X_i)\Big| \ge \frac{c_{\min}\sqrt N}{64 m_2\eta}\Big) \\ &\quad + P_{\theta_0}^N\Big(\Big|\frac{1}{\sqrt N}\sum_{i=1}^N \varepsilon_i g_{v,\theta*}^{II}(X_i)\Big| \ge \frac{c_{\min}\sqrt N}{16}\Big) \\ &\le 4e^{-4\mathcal{R}_N^{2,II}}. \end{aligned} \tag{83}$$

Combining the terms. By combining the bounds (70), (72), (80), (83) and using $N(1/4) \le 9^D \le e^{3D}$ (cf. [44, Proposition 4.3.34]) we find that since $D \le \mathcal{R}_N \le \min\{\mathcal{R}_N^{2,I}, \mathcal{R}_N^{2,II}\}$ from (45),

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

completing the proof of (47).

3.4.2. Proof of (48). We derive probability bounds for each of the three terms in (48) separately. The general scheme of proof for each of the three bounds is similar to the proof of (47), and we condense some of the steps to follow.

Second order term. Using $c_{\max} \geq c_{\min}$, we can replace (70) by

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

From here onwards, this term can be treated exactly as in the proof of (47) and thus, for $D \leq \mathcal{R}_n$ from (45), we deduce

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} \lambda_{\max}[-\nabla^2\ell_N(\theta,Z)] \geq 3Nc_{\max}/2\Big) \leq 8e^{-\mathcal{R}_N}. \tag{84}$$

First order term. First, let us denote

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

and let $(v_i : i = 1,\dots,N(1/2))$ be the centre points of a $\|\cdot\|_{\mathbb{R}^D}$-covering of the unit ball $\{\theta : \|\theta\|_{\mathbb{R}^D} \leq 1\}$ with balls of radius $1/2$. 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,

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

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

$$\sup_{v:\,\|v\|_{\mathbb{R}^D}\leq 1}|P_N(f_{v,\theta})| \leq 2\max_{1\leq i\leq N(1/2)}|P_N(f_{v_i,\theta})|. \tag{85}$$

We can hence estimate

$$\begin{aligned}&P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\|\nabla\bar{\ell}_N(\theta)\|_{\mathbb{R}^D} \geq 3c_{\max}/2\Big) \\ &\qquad\leq P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\sup_{v:\,\|v\|_{\mathbb{R}^D}\leq 1}\big|v^T\big[\nabla\bar{\ell}_N(\theta) - E_{\theta_0}[\nabla\ell(\theta)]\big]\big| \geq c_{\max}/2\Big) \\ &\qquad\leq N(1/2)\cdot\sup_{v:\,\|v\|_{\mathbb{R}^D}\leq 1}P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}|P_N(f_{v,\theta})| \geq c_{\max}/4\Big).\end{aligned} \tag{86}$$

We fix $v\in\mathbb{R}^D$ with $\|v\|_{\mathbb{R}^D} \leq 1$. Using (68), by decomposing the 'uncentred' part of $f_{v,\theta}$ into

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

we can then write

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

where we have further defined $f^I_{v,\theta}(x) := \tilde f^I_{v,\theta}(x) - E_{\theta_0}[\tilde f^I_{v,\theta}(X)]$ (and still write $P_N(f^I_{v,\theta}) = N^{-1}\sum_{i=1}^N f^I_{v,\theta}(X_i)$ to simplify notation). We then estimate the probability on the right hand side of (86) as follows:

$$
\begin{aligned}
&P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} |P_N(f_{v,\theta})| \geq c_{\max}/4\Big)\\
&\leq P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} |P_N(f^I_{v,\theta} - f^I_{v,\theta^*})| \geq c_{\max}/16\Big) + P^N_{\theta_0}\big(|P_N(f^I_{v,\theta^*})| \geq c_{\max}/16\big)\\
&\quad + P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} |P_N(f^{II}_{v,\theta} - f^{II}_{v,\theta^*})| \geq c_{\max}/16\Big) + P^N_{\theta_0}\big(|P_N(f^{II}_{v,\theta^*})| \geq c_{\max}/16\big)\\
&=: i + ii + iii + iv.
\end{aligned}
\tag{87}
$$

We first treat the terms i and ii. By the definition of $\tilde f^I_{v,\theta}$ and Assumption 3.2, for any $\theta, \bar\theta \in \mathcal{B}$ we have

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

Again using Assumption 3.2, we also have

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

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

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

Therefore, by using Lemma 3.12 in the same manner as in (77), we find 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

$$P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt N}\Big|\sum_{i=1}^N h^I_{v,\theta}(X_i)\Big| \geq L'[\sqrt D + \sqrt x + (D+x)/\sqrt N]\Big) \leq 2e^{-x}, \quad x\geq 1. \tag{88}$$

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 $x = 3\mathcal{R}^{1,I}_N$ in (88), we find that for $C > 0$ small enough and any $D \leq \mathcal{R}^{1,I}_N$,

$$ii + i \leq 2e^{-3\mathcal{R}^{1,I}_N} + P^N_{\theta_0}\Big(\Big|\frac{1}{\sqrt N}\sum_{i=1}^N h^I_{v,\theta}(X_i)\Big| \geq \frac{c_{\max}\sqrt N}{128(k_0m_1 + k_1m_0)\eta}\Big) \leq 4e^{-3\mathcal{R}^{1,I}_N}. \tag{89}$$

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

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

Therefore, by utilising Lemma 3.12 below as well as Bernstein's inequality (96) in precisely the same manner as in the derivations of (81) and (82) respectively, we obtain the two inequalities

$$P^N_{\theta_0}\left(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\left|\sum_{i=1}^N \varepsilon_i \frac{f^{II}_{v,\theta}(X_i) - f^{II}_{v,\theta^*}(X_i)}{4m_1\eta}\right| \geq L'[\sqrt{D} + \sqrt{x} + (D+x)/\sqrt{N}]\right) \\ \leq 2e^{-x}, \quad x \geq 1,$$

and

$$P^N_{\theta_0}\left(\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N \varepsilon_i f^{II}_{v,\theta^*}(X_i)\right| \geq k_1\sqrt{2x} + \frac{k_1 x}{3\sqrt{N}}\right) \leq 2e^{-x}, \quad x > 0.$$

Thus, if we set

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

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

$$iii + iv \leq 4e^{-3\mathcal{R}^{1,II}_N}. \tag{90}$$

By combining (86), (87), (89), (90), using $N(1/2) \leq e^{2D}$ (cf. [44, Proposition 4.3.34]) and since $D \leq \mathcal{R}_N \leq \min\{\mathcal{R}^{1,I}_N, \mathcal{R}^{1,II}_N\}$, we conclude that

$$P^N_{\theta_0}\left(\sup_{\theta\in\mathcal{B}} \|\nabla\bar{\ell}_N(\theta)\|_{\mathbb{R}^D} \geq 3c_{\max}/2\right) \leq N(1/2)\cdot(i + ii + iii + iv) \\ \leq 4e^{2D-3\mathcal{R}^{1,I}_N} + 4e^{2D-3\mathcal{R}^{1,II}_N} \leq 8e^{-\mathcal{R}_N}. \tag{91}$$

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

$$\begin{aligned} -\ell(\theta, Z) &= \tfrac{1}{2}[\mathcal{G}^X(\theta_0) - \mathcal{G}^X(\theta)]^2 - \varepsilon[\mathcal{G}^X(\theta_0) - \mathcal{G}^X(\theta)] + \varepsilon^2/2 \\ &=: \tilde{l}^I_\theta(X) + \varepsilon l^{II}_\theta(X) + \varepsilon^2/2, \end{aligned}$$

and therefore, defining

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

we have

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

Then, using Assumption 3.3, we can estimate

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

To bound the preceding terms, we use Assumption 3.2 to deduce that for all $\theta,\bar{\theta}\in\mathcal{B}$,

$$
\begin{aligned}
\|l_\theta^I-l_{\bar{\theta}}^I\|_\infty\leq 2\|\tilde{l}_\theta^I-\tilde{l}_{\bar{\theta}}^I\|_\infty &= \|-2\mathcal{G}(\theta_0)[\mathcal{G}(\theta)-\mathcal{G}(\bar{\theta})]+\mathcal{G}(\theta)^2-\mathcal{G}(\bar{\theta})^2\|_\infty\\
&=\|[(\mathcal{G}(\theta)-\mathcal{G}(\theta_0))+(\mathcal{G}(\bar{\theta})-\mathcal{G}(\theta_0))][\mathcal{G}(\theta)-\mathcal{G}(\bar{\theta})]\|_\infty\\
&\leq 2k_0m_0|\theta-\bar{\theta}|_1,
\end{aligned}
$$

as well as

$$
\sup_{\theta\in\mathcal{B}}\|l_\theta^I-l_{\theta*}^I\|_\infty\leq 2k_0m_0\eta,\quad \|l_{\theta*}^I\|_\infty\leq k_0^2.
$$

Moreover, again by Assumption 3.2 we have, for all $\theta,\bar{\theta}\in\mathcal{B}$,

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

Next, similarly to the second and first order terms, in order to control terms i and iii we now apply Lemma 3.12 to the rescaled empirical processes

$$
h_\theta^I:=\frac{l_\theta^I-l_{\theta*}^I}{8k_0m_0\eta},\quad h_\theta^{II}:=\frac{l_\theta^{II}-l_{\theta*}^{II}}{8m_0\eta},
$$

and in order to control terms ii and iv, we respectively apply Hoeffding's inequality and Bernstein's inequality (96) in the same manner as before. Overall, if we set

$$
\begin{aligned}
\mathcal{R}_N^{0,I}&:=CN\min\Big\{\frac{c_{\max}^2}{k_0^2m_0^2\eta^2},\frac{c_{\max}}{k_0m_0\eta},\frac{c_{\max}^2}{k_0^4}\Big\},\\
\mathcal{R}_N^{0,II}&:=CN\min\Big\{\frac{c_{\max}^2}{m_0^2\eta^2},\frac{c_{\max}}{m_0\eta},\frac{c_{\max}^2}{k_0^2},\frac{c_{\max}}{k_0}\Big\},
\end{aligned}
\tag{92}
$$

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

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

Finally, we estimate the term v by a standard tail inequality (see [44, Theorem 3.1.9]),

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

and thus obtain

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}} |\bar{\ell}_N(\theta, Z)| \geq 2c_{\max} + 1\Big) \leq i + ii + iii + iv + v \leq 8e^{-\mathcal{R}_N} + e^{-N/8}. \tag{93}$$

Conclusion. By combining (84), (91) and (93), the proof of (48) is completed.

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. [94, Theorem 2.2.28] and [35, Theorem 3.5]). 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 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 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 $U \geq \sup_\theta \|h_\theta\|_\infty$ and has variance envelope $\sigma^2 \geq \sup_\theta E^X[h_\theta^2(X)]$ 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 $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, 1)$ 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

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

Then, for some universal constant $L > 0$ *and all* $x \geq 1$,

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

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 of [35], 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,

$$\begin{aligned}
E[\exp\{\lambda\varepsilon(h_\theta - h_{\theta'})(X)\}] &\leq 1 + \sum_{k=2}^{\infty}\frac{|\lambda|^k E^\varepsilon[|\varepsilon|^k]E^X[|h_\theta - h_{\theta'}|^k(X)]}{k!} \\
&\leq 1 + \lambda^2 E^X[(h_\theta(X) - h_{\theta'}(X))^2]\sum_{k=2}^{\infty}\frac{E^\varepsilon[|\varepsilon|^k]}{k!}(|\lambda|\,\|h_\theta - h_{\theta'}\|_\infty)^{k-2} \\
&\leq \exp\left\{\frac{\lambda^2 d_2^2(\theta,\theta')}{1 - |\lambda|d_\infty(\theta,\theta')}\right\} \qquad (94)
\end{aligned}$$

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

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

An application of the exponential Chebyshev inequality (and optimisation in λ, as in [44, proof of Proposition 3.1.8]) then implies that condition (3.8) in [35] 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 [35] now implies, for some universal constant M and any $\theta_\dagger \in \Theta$, that

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

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 [35, (2.3)]). 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

$$\Pr\Big(\sup_{\theta\in\Theta}|Z_N(\theta) - Z_N(\theta_\dagger)| \geq \bar{L}[J_2(\mathcal{H}) + \sigma\sqrt{x} + (J_\infty(\mathcal{H}) + Ux)/\sqrt{N}]\Big) \leq e^{-x} \qquad (95)$$

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

The preceding argument also implies the classical Bernstein inequality

$$\Pr\Big(|Z_N(\theta)| \geq \sigma\sqrt{2x} + \frac{Ux}{3\sqrt{N}}\Big) \leq 2e^{-x}, \quad x > 0, \tag{96}$$

for any fixed $\theta \in \Theta$, $U \geq \|h_\theta\|_\infty$ and $\sigma^2 \geq E^X[h_\theta^2(X)]$, proved [44, (3.24)], using (94). Applying this with $\theta_\dagger$ and using (95), the final result follows now from

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

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

3.6. *Proofs for Section* 3.3

We apply the results from Appendix A to $\mu = \tilde{\Pi}(\cdot \mid Z^{(N)})$.

Proof of Theorem 3.7. For any $\theta, \bar{\theta} \in \mathbb{R}^D$, for the log-prior density we have

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

and for the likelihood surrogate $\tilde{\ell}_N$, by Proposition 3.6 and on the event $\mathcal{E}$ from (60),

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

Combining the last two displays, and on the event $\mathcal{E}$, we can verify Assumption A.1 below for $-\log d\tilde{\Pi}(\cdot \mid Z^{(N)})$ with constants

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

We may thus apply Proposition A.4 to obtain

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

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

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

$$|\theta - \theta_{\text{init}}|_1^2 = (\theta - \theta_{\text{init}})^T M(\theta - \theta_{\text{init}}) \geq \lambda_{\min}(M)\|\theta - \theta_{\text{init}}\|_{\mathbb{R}^D}^2. \tag{97}$$

Thus, for any $\theta \in \mathbb{R}^D$ with $\|\theta - \theta_{\text{init}}\|_{\mathbb{R}^D}^2 \geq 4\eta^2/\lambda_{\min}(M)$, we have $|\theta - \theta_{\text{init}}|_1 \geq 2\eta$, and therefore also $g_\eta(\theta) \geq (|\theta - \theta_{\text{init}}|_1 - \eta)^2 \geq \frac{1}{4}|\theta - \theta_{\text{init}}|_1^2$. Thus, for C from (55) and any $\theta \in \mathbb{R}^D$ satisfying

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

using (97), (55) as well as the upper bound for $|\ell_N(\theta)|$ in the definition of $\mathcal{E}_{\text{conv}}$, we obtain

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

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

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

which also allows us to deduce

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

We further have $\theta_{\max}^T \Sigma^{-1} \theta_{\max} \leq \theta_{\tilde{\ell}}^T \Sigma^{-1} \theta_{\tilde{\ell}}$ (otherwise $\theta_{\max}$ would not maximise $\log d\tilde{\Pi}(\cdot \mid Z^{(N)})$), 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:

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

Proof of Theorem 3.8. For any $t \geq 0$ and any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$ we have

$$\begin{aligned} &\mathbf{P}_{\theta_{\text{init}}}\big(\big|\hat{\pi}^J_{J_{\text{in}}}(H) - E^\Pi[H \mid Z^{(N)}]\big| \geq t\big) \\ &\quad\leq \mathbf{P}_{\theta_{\text{init}}}\big(\big|\hat{\pi}^J_{J_{\text{in}}}(H) - \mathbf{E}_{\theta_{\text{init}}}[\hat{\pi}^J_{J_{\text{in}}}(H)]\big| \geq t - \big|\mathbf{E}_{\theta_{\text{init}}}[\hat{\pi}^J_{J_{\text{in}}}(H)] - E^\Pi[H \mid Z^{(N)}]\big|\big). \qquad (98) \end{aligned}$$

To further estimate the right side, note that for c_3 large enough and any $k \geq J_{\mathrm{in}}$, by (64) and Theorem 3.7, we have

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

Noting that (165) below in fact holds for any probability measure μ and thus in particular for $\mu = \Pi(\cdot \mid Z^{(N)})$, it follows that for any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$,

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

Thus if $t \geq 0$ satisfies (65), then applying Proposition A.3 to both H and $-H$ shows that the r.h.s. in (98) is further bounded by

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

Proof of Corollary 3.9. We first estimate the probability to be bounded by

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

Next, for any $k \geq 1$, let ν_k denote an optimal coupling between $\mathcal{L}(\vartheta_k)$ and $\Pi(\cdot \mid Z^{(N)})$ (cf. [104, Theorem 4.1]). Then by Jensen's inequality and the definition of W_2 from (9),

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

Thus from (61), (64) (as after (98)) we obtain

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

Now for any $j = 1, \dots, d$, let us write $H_j : \mathbb{R}^D \to \mathbb{R}$, $\theta \mapsto \theta_j$, for the j-th coordinate projection map, of Lipschitz constant 1. Then in the notation (63) we can write

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

For $t \geq \sqrt{8(\rho + b(\gamma))}$ and applying Proposition A.3 as in the proof of Theorem 3.8, as well as a union bound, gives

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

4. Proofs for the Schrödinger model

In this section, we will show how the results from Section 3 can be applied to the non-linear problem for the Schrödinger equation (17). Recalling the notation of Sections 2 and 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 (39) equals the Euclidean ball

$$\mathcal{B}_\epsilon:=\{\theta\in\mathbb{R}^D:\|\theta-\theta_{0,D}\|_{\mathbb{R}^D}<\epsilon D^{-4/d}\}. \tag{99}$$

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 of Lemma 3.4 with the PDE estimates below, notably the ‘average curvature’ bound from Lemma 4.7.

Proposition 4.1. *Let $\theta_0\in h^2$ satisfy $\|\theta_0\|_{h^2}\leq S$ for some $S>0$ and consider ℓ_N from (22) with forward map $\mathcal{G}:\mathbb{R}^D\to\mathbb{R}$ from (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 $D\leq c_2N^{\frac{d}{d+12}}$ as well as $\|\mathcal{G}(\theta_0)-\mathcal{G}(\theta_{0,D})\|_{L^2(\mathcal{O})}\leq c_1D^{-4/d}$, the event*

$$\mathcal{E}_{\mathrm{conv}}(\epsilon)=\Big\{\inf_{\theta\in\mathcal{B}_\epsilon}\lambda_{\min}(-\nabla^2\ell_N(\theta))>c_3ND^{-4/d},\\ \sup_{\theta\in\mathcal{B}_\epsilon}[|\ell_N(\theta)|+\|\nabla\ell_N(\theta)\|_{\mathbb{R}^D}+\|\nabla^2\ell_N(\theta)\|_{\mathrm{op}}]<c_4N\Big\}$$

satisfies

$$P^N_{\theta_0}(\mathcal{E}_{\mathrm{conv}}(\epsilon))\geq 1-33e^{-c_2N^{d/(d+12)}}. \tag{100}$$

Proof. For any $\theta\in\mathbb{R}^D$, F_θ as in (16), by a Sobolev embedding and (13), we have $\|F_\theta\|_\infty\lesssim\|\theta\|_{h^2}\lesssim D^{2/d}\|\theta\|_{\mathbb{R}^D}$. This and Lemmas 4.4–4.6 verify Assumption 3.2 in the present setting, with constants

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

whence the constants from (46) satisfy

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

Moreover, using (28) and (107), Lemmas 4.7 and 4.8 verify Assumption 3.3 for our choice of η with

$$c_{\min} \simeq D^{-4/d}, \quad c_{\max} \simeq \text{const.} \tag{101}$$

Then the minimum (45) is dominated by the third term, yielding

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

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 3.4 then implies that for all such D, N, we have

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

Next, if θ_{init} is the estimator from Theorem B.6, then in the present setting with $\epsilon = 1/\log N$, the event (50) equals

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

Proposition 4.2. *Assuming Condition 2.3, there exist constants $c_5, c_6 > 0$ such that for all $N \in \mathbb{N}$,*

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

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

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

Next, denoting by $\tilde{\Pi}(\cdot \mid Z^{(N)})$ the 'surrogate' posterior measure with density (27), and if

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

is given by (59) with $\rho = 2\exp(-N^{d/(2\alpha+d)})$, then Theorem 4.14 implies the following approximation result in Wasserstein distance.

Proposition 4.3. *Assume Conditions 2.2 and 2.3. Then there exist constants $c_7, c_8 > 0$ such that for all $N \in \mathbb{N}$,*

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

The preceding propositions imply that the events

$$\mathcal{E}_N := \mathcal{E}_{\text{conv}} \cap \mathcal{E}_{\text{init}} \cap \mathcal{E}_{\text{wass}} \tag{102}$$

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 3 are eventually verified.

Proof of Theorem 2.7. We will prove Theorem 2.7 by applying Theorem 3.7 with the choices $\mathcal{B} = \mathcal{B}_\epsilon$ from (99), $\epsilon = 1/\log N$ and K from Condition 2.2, $\rho = 2\exp(-N^{d/(2\alpha+d)})$ and $M = I_{D\times D}$ generating the ellipsoidal norm $\|\cdot\|_{\mathbb{R}^D}$. Using (13), the prior covariance Σ from (23) satisfies

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

Then using Condition 2.2, we first have

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

verifying the lower bound (55), and then also $m, \Lambda > 0$ from Theorem 3.7 satisfy

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

The dimension condition (28) and the condition on α further imply

$$ND^{-4/d} \gtrsim N^{d/(2\alpha+d)}, \quad N^{d/(2\alpha+d)}D^{2\alpha/d} \lesssim N,$$

whence we further obtain

$$m \simeq ND^{-4/d}, \quad \Lambda \simeq ND^{8/d}(\log N)^3. \tag{103}$$

Noting that also $\gamma = o(\Lambda^{-1})$ with our choices, Theorem 3.7 implies that on the event $\mathcal{E}_N$ from (102), the Markov chain (ϑ_k) satisfies the Wasserstein bound (61) with

$$b(\gamma) \lesssim \frac{\gamma D\Lambda^2}{m^2} + \frac{\gamma^2 D\Lambda^4}{m^3} \lesssim \gamma D^{(d+24)/d}(\log N)^6 + \gamma^2 ND^{(d+44)/d}(\log N)^{12}, \tag{104}$$

as well as

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

Using also $D/m \lesssim$ const, the first part of Theorem 2.7 follows.

For the choice of $\gamma = \gamma_\varepsilon$ from (31), straightforward calculation yields (for N large enough)

$$B(\gamma_\varepsilon) = o(\varepsilon^2 + N^{-2P}), \tag{105}$$

which proves the second part of Theorem 2.7. ∎

Proof of Proposition 2.4 *and of Theorems* 2.5, 2.6. Proposition 2.4 now follows directly from Theorem 3.8 and the preceding computations. Noting that for all N large enough we have $B(\gamma) \le N^{-P}$, Theorem 2.5 follows from Corollary 3.9, (105) as well as (67), for $J_{\text{in}} \ge (\log N)^3/(\gamma_\varepsilon ND^{-4/d})$. Finally, intersecting further with the event

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

Theorem 2.6 follows from the triangle inequality and (151). ∎

Proof of Theorem 2.8. In the proof we intersect $\mathcal{E}_N$ from (102) further with the event on which the conclusion of Theorem 4.12 holds. Part (iii) then follows from part (ii) and straightforward calculations. Part (i) follows from the arguments following (157) below, where it is proved in particular that $\hat{\theta}_{\mathrm{MAP}}$ is the unique maximiser of the proxy posterior density $\tilde{\pi}(\cdot \mid Z^{(N)})$ over $\mathbb{R}^D$. We can now apply Proposition A.2 with m, Λ from (103), using also

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

in view of $\ell_N = \tilde{\ell}_N$ on $\mathcal{B}_{1/(8\log N)}$, the definition of $\mathcal{E}_{\mathrm{init}}$, (13) and since $\theta_0 \in h^\alpha$. ■

4.1. Analytical properties of the Schrödinger forward map

This section is devoted to proving the four auxiliary Lemmas 4.5–4.8 used in the proof of Proposition 4.1. Throughout we consider the forward map $\mathscr{G} : \mathbb{R}^D \to L^2(\mathcal{O})$, $\mathscr{G} = G \circ \Phi^* \circ \Psi$, given by (17) and assume the hypotheses of Proposition 4.1, where the set $\mathcal{B}_\epsilon$ was defined in (99).

For any $f \in C(\mathcal{O})$ with $f \geq 0$, by standard theory for elliptic PDEs (see e.g. [40, Chapter 6.3]) 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

$$\begin{cases} \frac{\Delta}{2} w - f w = \psi & \text{on } \mathcal{O}, \\ w = 0 & \text{on } \partial\mathcal{O}. \end{cases} \tag{106}$$

Lemma 4.4. *For any $x \in \mathcal{O}$, the map $\theta \mapsto \mathscr{G}(\theta)(x)$ is twice continuously differentiable on $\mathbb{R}^D$. The vector field $\nabla \mathscr{G}_\theta : \mathcal{O} \to \mathbb{R}^D$ is given by*

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

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

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

Proof. In the notation from (17), the map $\theta \mapsto \mathscr{G}(\theta)(x)$ 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 Fréchet 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 Fréchet differentiable with derivatives DG, D^2G and $D\Phi^*$, $D^2\Phi^*$ given by Lemma B.2 and (175) respectively. By the chain rule for Fréchet derivatives (see Lemma B.3), we deduce that $x \mapsto \mathscr{G}(\theta)(x)$ 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 Appendix B. ∎

Now since $\|\theta_0\|_{h^2} \leq S$ and by the definition (99) of the set $\mathcal{B}_1$, we see from (13) that

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

It follows further from the Sobolev embedding and regularity of the link function Φ (Appendix B.1.1) that there exists a constant $B = B(S, \Phi, \mathcal{O}) < \infty$ such that

$$\sup_{\theta\in\mathcal{B}_1} [\|F_\theta\|_\infty + \|F_\theta\|_{H^2} + \|f_\theta\|_{H^2} + \|f_\theta\|_\infty] \leq B. \tag{107}$$

In particular, this estimate implies that the constants appearing in the inequalities from Lemma B.1 can be chosen independently of $\theta \in \mathcal{B}$, which we use frequently below.

For notational convenience we also introduce the spaces

$$E_D := \mathrm{span}(e_1, \dots, e_D) \subseteq L^2(\mathcal{O}), \quad D \in \mathbb{N}, \tag{108}$$

spanned by the first D eigenfunctions of Δ on $\mathcal{O}$ (cf. Section 2.1.1).

We first verify the boundedness property required in Assumption 3.2 (ii).

Lemma 4.5. *There exists a constant $C > 0$ such that*

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

Proof. The estimate for $\|\mathscr{G}(\theta)\|_\infty$ follows immediately from (18). To estimate $\|\nabla\mathscr{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)}$, we first note that by Lemma 4.4,

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

Thus by the Sobolev embedding $\|\cdot\|_\infty \lesssim \|\cdot\|_{H^2}$, Lemma B.1 and boundedness of Φ', we deduce that for any $\theta \in \mathcal{B}_1$ and any $H \in E_D$,

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

Again using Lemma 4.4, we can similarly estimate $\|\nabla^2 \mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)}$ by

$$\begin{aligned}
\|\nabla^2 \mathcal{G}(\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} &\leq \sup_{v:\|v\|_{\mathbb{R}^D}\leq 1} \|v^T \nabla^2 \mathcal{G}(\theta) v\|_{L^\infty} \\
&\leq \sup_{H\in E_D:\|H\|_{L^2}\leq 1} \big(2\|V_{f_\theta}[H(\Phi'\circ F_\theta)V_{f_\theta}[H(\Phi'\circ F_\theta)u_{f_\theta}]]\|_\infty \\
&\qquad\qquad + \|V_{f_\theta}[H^2(\Phi''\circ F_\theta)u_{f_\theta}]\|_\infty\big) \\
&=: \sup_{H\in E_D:\|H\|_{L^2}\leq 1} (I+II). \qquad (109)
\end{aligned}$$

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

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

as well as

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

where we have used the basic norm estimate on $E_D \subseteq L^2(\mathcal{O})$ from Lemma 4.9. By combining the last three displays, the proof is completed. ■

Next, we verify the increment bound needed in Assumption 3.2 (iii).

Lemma 4.6. *There exists a constant $C > 0$ such that for any $D \in \mathbb{N}$ and any $\theta, \theta' \in \mathbb{R}^D$,*

$$\|\mathcal{G}(\theta) - \mathcal{G}(\bar\theta)\|_\infty \leq C\|F_\theta - F_{\bar\theta}\|_\infty, \quad \|\mathcal{G}(\theta) - \mathcal{G}(\bar\theta)\|_{L^2} \leq C\|F_\theta - F_{\bar\theta}\|_{L^2}, \qquad (110)$$

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

$$\|\nabla\mathcal{G}(\theta) - \nabla\mathcal{G}(\bar\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^D)} \leq C\|F_\theta - F_{\bar\theta}\|_\infty, \qquad (111)$$

$$\|\nabla^2\mathcal{G}(\theta) - \nabla^2\mathcal{G}(\bar\theta)\|_{L^\infty(\mathcal{O},\mathbb{R}^{D\times D})} \leq CD^{2/d}\|F_\theta - F_{\bar\theta}\|_\infty. \qquad (112)$$

Proof. The estimate (110) follows immediately from (171) and (177) in Appendix B. Now fix any $\theta, \bar\theta \in \mathcal{B}_1$. To simplify 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 (111), arguing as in the proof of Lemma 4.5, we first have

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

Now, we fix $H \in E_D$ for the rest of the proof. The term I_a can be estimated by repeatedly using the Sobolev embedding $\|\cdot\|_\infty \lesssim \|\cdot\|_{H^2}$, Lemma B.1 as well as (107) and (177):

$$\begin{aligned} I_a &= \|V_f[(f-\bar{f})V_{\bar{f}}[u_{\bar{f}}(\Phi' \circ F)H]]\|_\infty \lesssim \|V_f[(f-\bar{f})V_{\bar{f}}[u_{\bar{f}}(\Phi' \circ F)H]]\|_{H^2} \\ &\lesssim \|(f-\bar{f})V_{\bar{f}}[u_{\bar{f}}(\Phi' \circ F)H]]\|_{L^2} \lesssim \|f-\bar{f}\|_\infty \|u_{\bar{f}}(\Phi' \circ \bar{F})H\|_{L^2} \\ &\lesssim \|F-\bar{F}\|_\infty \|H\|_{L^2}. \end{aligned} \tag{113}$$

Similarly, I_b is estimated as follows:

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

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 (110). By combining the estimates for I_a, I_b and I_c, we have completed the proof of (111).

It remains to prove (112). In analogy to (109), we may fix any $v \in \mathbb{R}^D$, and it suffices to derive a bound for $v^T(\nabla^2\mathscr{G}(\theta) - \nabla^2\mathscr{G}(\bar{\theta}))v$. To simplify 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 4.4, we have the following decomposition into eight terms:

$$\begin{aligned} &v^T(\nabla^2\mathscr{G}(\theta) - \nabla^2\mathscr{G}(\bar{\theta}))v \\ &= 2V_{\bar{f}}\big[\bar{h}V_{\bar{f}}[\bar{h}u_{\bar{f}}]\big] - 2V_f\big[hV_f[hu_f]\big] + V_{\bar{f}}[u_{\bar{f}}H^2(\Phi'' \circ \bar{F})] - V_f[u_f H^2(\Phi'' \circ F)] \\ &= 2(V_{\bar{f}} - V_f)\big[\bar{h}V_{\bar{f}}[\bar{h}u_{\bar{f}}]\big] + 2V_f\big[(\bar{h}-h)V_{\bar{f}}[\bar{h}u_{\bar{f}}]\big] \\ &\quad + 2V_f\big[h(V_{\bar{f}} - V_f)[\bar{h}u_{\bar{f}}]\big] + 2V_f\big[hV_f[(\bar{h}-h)u_{\bar{f}}]\big] + 2V_f\big[hV_f[h(u_{\bar{f}} - u_f)]\big] \\ &\quad + (V_{\bar{f}} - V_f)[u_{\bar{f}}H^2(\Phi'' \circ \bar{F})] + V_f[(u_{\bar{f}} - u_f)H^2(\Phi'' \circ \bar{F})] \\ &\quad + V_f[u_f H^2(\Phi'' \circ \bar{F} - \Phi'' \circ F)] \\ &=: II_a + II_b + II_c + II_d + II_e + II_f + II_g + II_h. \end{aligned} \tag{114}$$

To estimate these terms, we will again repeatedly use (107), the regularity estimates from Lemmas B.1–B.2 below, the estimates $\|h\|_{L^2}, \|\bar{h}\|_{L^2} \lesssim \|H\|_{L^2}$ together with $\|f-\bar{f}\|_\infty \lesssim \|F-\bar{F}\|_\infty$, which all hold uniformly in $\theta \in \mathcal{B}_1$.

Using Lemma B.1, including the estimate (169) 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 (169),

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

For II_d, by following similar steps to those 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 (110), 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 II_f, we note that by the Sobolev embedding,

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

and consequently by Lemma B.1,

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

For II_g and II_h, by similar steps and additionally using the fact that by Lemma 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 (114) with the estimates for II_a–II_h, the proof of (112) is complete. ∎

We now turn to the key 'geometric' bound from the first part of Assumption 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 'information operator' is of the form I^2, with I being the inverse of a second order (elliptic Schrödinger-type) operator (cf. also [77, Section 4]).

Lemma 4.7. *Let $\ell(\theta)$ be as in (38) with $\mathcal{G}:\mathbb{R}^D\to\mathbb{R}$ from (17), and let $\mathcal{B}_\epsilon$ be as in (99). Let $\theta_0\in h^2$ satisfy $\|\theta_0\|_{h^2}\le S$ for some $S>0$. Then there exist constants $0<\epsilon_S\le 1$ and $c_1,c_2>0$ such that if also $\|\mathcal{G}(\theta_0)-\mathcal{G}(\theta_{0,D})\|_{L^2(\mathcal{O})}\le c_1D^{-4/d}$, then for all $D\in\mathbb{N}$ and all $\epsilon\le\epsilon_S$,*

$$\inf_{\theta\in\mathcal{B}_\epsilon}\lambda_{\min}(E_{\theta_0}[-\nabla^2\ell(\theta)])\ge c_2D^{-4/d}. \tag{115}$$

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

$$\begin{aligned}v^TE_{\theta_0}[-\nabla^2\ell(\theta,Z)]v&=\|V_{f_\theta}[u_{f_\theta}(\Phi'\circ F_\theta)H]\|^2_{L^2(\mathcal{O})}\\&\quad-\big\langle u_{f_{\theta_0}}-u_{f_\theta},2V_{f_\theta}\big[hV_{f_\theta}[hu_{f_\theta}]\big]\big\rangle_{L^2(\mathcal{O})}\\&\quad-\langle u_{f_{\theta_0}}-u_{f_\theta},V_{f_\theta}[u_{f_\theta}H^2(\Phi''\circ F_\theta)]\rangle_{L^2(\mathcal{O})}\\&=:I+II+III.\end{aligned}\tag{116}$$

We next derive a lower bound on I and upper bounds for 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-$(H^2_0)^*$ coercivity estimate (168) from Lemma B.1 below as well as (107), we have

$$\sqrt{I}=\|V_{f_\theta}[a_\theta H]\|_{L^2(\mathcal{O})}\gtrsim\frac{\|a_\theta H\|_{(H^2_0)^*}}{1+\|f_\theta\|_\infty}\gtrsim\|a_\theta H\|_{(H^2_0)^*},\qquad\theta\in\mathcal{B}_1. \tag{117}$$

The next step is to lower bound a_θ. By [28, Theorem 1.17], the expected exit time $\tau_\mathcal{O}$ featuring in the Feynman–Kac formula (12) satisfies the uniform estimate $\sup_{x\in\mathcal{O}}E^x\tau_\mathcal{O}\le K(\mathrm{vol}(\mathcal{O}),d)<\infty$. Therefore, using also Jensen's inequality and $g\ge g_{\min}>0$, we obtain, with B from (107),

$$\inf_{\theta\in\mathcal{B}_1}\inf_{x\in\mathcal{O}}u_{f_\theta}(x)\ge g_{\min}e^{-BK(\mathrm{vol}(\mathcal{O}),d)}=:u_{\min}>0. \tag{118}$$

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

$$\inf_{\theta\in\mathcal{B}_1}\inf_{x\in\mathcal{O}}a_\theta(x)\ge a_{\min}>0. \tag{119}$$

We thus find, by definition of $(H^2_0)^*$ and the multiplication inequality (7), that for some $c=c(a_{\min})>0$,

$$\begin{aligned}\|H\|_{(H^2_0)^*}&=\|a_\theta a_\theta^{-1}H\|_{(H^2_0)^*}\le\|a_\theta^{-1}\|_{H^2}\|a_\theta H\|_{(H^2_0)^*}\\&\le c(1+\|a_\theta\|^2_{H^2})\|a_\theta H\|_{(H^2_0)^*},\end{aligned}\tag{120}$$

where in the last inequality we have used (176) for the function $x \mapsto 1/x$. Using again (107), regularity of Φ', the chain rule as well as the elliptic regularity estimate (173), we obtain

$$\sup_{\theta \in \mathcal{B}_1} \|a_\theta\|_{H^2} \le \sup_{\theta \in \mathcal{B}_1} \|u_{f_\theta}\|_{H^2} \sup_{\theta \in \mathcal{B}_1} \|\Phi' \circ F_\theta\|_{H^2} \le C(g, S, \mathcal{O}, \Phi) < \infty. \tag{121}$$

Therefore, combining the displays (117), (120), (121), we have proved that, uniformly in $\theta \in \mathcal{B}_1$,

$$I \gtrsim \|a_\theta H\|^2_{(H_0^2)^*} \gtrsim \frac{\|H\|^2_{(H_0^2)^*}}{c^2 \sup_{\theta \in \mathcal{B}_1} (1 + \|a_\theta\|^2_{H^2})^2} \gtrsim D^{-4/d} \|H\|^2_{L^2}, \tag{122}$$

where we have used Lemma 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 B.1, (107), the Lipschitz estimate (171) as well as (18), we have, uniformly in $\theta \in \mathcal{B}_1$,

$$\begin{aligned} |II| &\lesssim \left| \int_{\mathcal{O}} (u_{f_{\theta_0}} - u_{f_\theta}) V_{f_\theta} \big[h V_{f_\theta} [h u_{f_\theta}] \big] \right| = \left| \int_{\mathcal{O}} V_{f_\theta} [u_{f_{\theta_0}} - u_{f_\theta}] \big[h V_{f_\theta} [h u_{f_\theta}] \big] \right| \\ &\lesssim \|V_{f_\theta}[u_{f_{\theta_0}} - u_{f_\theta}]\|_\infty \|h V_{f_\theta}[h u_{f_\theta}]\|_{L^1} \lesssim \|u_{f_{\theta_0}} - u_{f_\theta}\|_{L^2} \|h\|_{L^2} \|V_{f_\theta}[h u_{f_\theta}]\|_{L^2} \\ &\lesssim \|u_{f_{\theta_0}} - u_{f_\theta}\|_{L^2} \|H\|^2_{L^2}. \end{aligned} \tag{123}$$

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

$$\begin{aligned} |III| &= |\langle u_{f_{\theta_0}} - u_{f_\theta}, V_{f_\theta}[u_{f_\theta} H^2 (\Phi'' \circ F_\theta)] \rangle_{L^2(\mathcal{O})}| \\ &= |\langle V_{f_\theta}[u_{f_{\theta_0}} - u_{f_\theta}], u_{f_\theta} H^2 (\Phi'' \circ F_\theta) \rangle_{L^2(\mathcal{O})}| \\ &\le \|V_{f_\theta}[u_{f_{\theta_0}} - u_{f_\theta}]\|_\infty \|u_{f_\theta}\|_\infty \|\Phi'' \circ F_\theta\|_\infty \|H^2\|_{L^1} \\ &\lesssim \|u_{f_{\theta_0}} - u_{f_\theta}\|_{L^2} \|H\|^2_{L^2}. \end{aligned} \tag{124}$$

Combining the displays (116), (122), (123) and (124), we have proved that for any $\theta \in \mathcal{B}_1$, any $v \in \mathbb{R}^D$ and some constants $c', c'' > 0$,

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

Using (110) and the hypotheses, we see 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 conclude 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 3.3 (ii).

Lemma 4.8 (Upper bound). *For every $S>0$, there exists a constant $c>0$ such that for $\|\theta_0\|_{h^2}\le S$ and all $D\in\mathbb{N}$, we have*

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

Proof. For the first term, using Lemma 4.5, we see 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 derivative, similarly by Lemma 4.5 there exists some $K_1>0$ such that for any $\theta\in\mathcal{B}_1$,

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

For the second derivative, we recall the decomposition

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

where the terms I–III were defined in (116). Suitable uniform upper bounds for II and III have already been shown in (123) and (124) respectively, whence it suffices to upper bound the term I. We do this by using (107) and Lemma B.1: for any $\theta\in\mathcal{B}_1$ and any $H=\Psi(v)$, $v\in\mathbb{R}^D$,

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

We conclude with the following basic comparison lemma for Sobolev norms on the subspaces $E_D\subseteq L^2(\mathcal{O})$ from (108).

Lemma 4.9. *There exists $C>0$ such that for any $D\in\mathbb{N}$ and any $H\in E_D$,*

$$\|H\|_{H^2}\le CD^{2/d}\|H\|_{L^2},\quad \|H\|_{L^2}\le CD^{2/d}\|H\|_{(H_0^2)^*}.\tag{125}$$

Proof. Fix $D\in\mathbb{N}$. By the isomorphism property of Δ between the spaces H_0^2 and L^2 (see e.g. [66, Theorem II.5.4]), 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 (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:

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

4.2. Wasserstein approximation of the posterior measure

The main purpose of this section is to prove Theorem 4.14, which provides a bound on the Wasserstein distance between the posterior measure $\Pi(\cdot \mid Z^{(N)})$ from (24) and the surrogate posterior $\tilde{\Pi}(\cdot \mid Z^{(N)})$ from (27) in the Schrödinger model. The idea behind the proof is to show that both $\Pi(\cdot \mid Z^{(N)})$ and $\tilde{\Pi}(\cdot \mid Z^{(N)})$ concentrate most of their mass on the region (99) where the log-likelihood function ℓ_N is strongly concave (with high $P_{\theta_0}^N$-probability, cf. Proposition 4.1). We achieve this by a careful study of the mode (maximiser) of the posterior density, given in Theorem 4.12. Our derivations reveal that both $\pi(\cdot \mid Z^{(N)})$ and $\tilde{\pi}(\cdot \mid Z^{(N)})$ possess a *unique* mode, with high frequentist probability (see after (157)).

4.2.1. Convergence rate of MAP estimates. For $(Y_i, X_i)_{i=1}^N$ arising from (19) with $\mathscr{G} : \mathbb{R}^D \to \mathbb{R}$ from (17), we now study maximisers

$$\hat{\theta}_{\text{MAP}} \in \underset{\theta \in \mathbb{R}^D}{\arg\max} \left[-\frac{1}{2N} \sum_{i=1}^N (Y_i - \mathscr{G}(\theta)(X_i))^2 - (\delta_N^2)/2 \|\theta\|_{h^\alpha}^2 \right], \qquad \delta_N = N^{-\frac{\alpha}{2\alpha+d}}, \tag{126}$$

of the posterior density (24). For Λ_α from (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 $Z_i = (Y_i, X_i)$'s as

$$P_N = \frac{1}{N} \sum_{i=1}^N \delta_{(Y_i, X_i)}, \qquad \text{so that} \qquad \int h \, dP_N = \frac{1}{N} \sum_{i=1}^N h(Y_i, X_i) \tag{127}$$

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 (21).

Lemma 4.10. *Let $\hat{\theta}_{\text{MAP}}$ be any maximiser in (126), and denote by $\theta_{0,D}$ the projection of θ_0 onto $\mathbb{R}^D$. We have ($P_{\theta_0}^N$-a.s.)*

$$\frac{1}{2}\|\mathscr{G}(\hat{\theta}_{\text{MAP}}) - \mathscr{G}(\theta_0)\|_{L^2}^2 + \delta_N^2 I(\hat{\theta}_{\text{MAP}})$$
$$\le \int \log \frac{p_{\hat{\theta}_{\text{MAP}}}}{p_{\theta_{0,D}}} \, d(P_N - P_{\theta_0}) + \delta_N^2 I(\theta_{0,D}) + \frac{1}{2}\|\mathscr{G}(\theta_{0,D}) - \mathscr{G}(\theta_0)\|_{L^2}^2.$$

Proof. By the definitions,

$$\ell_N(\hat{\theta}_{\text{MAP}}) - \ell_N(\theta_{0,D}) - N\delta_N^2 I(\hat{\theta}_{\text{MAP}}) \ge -N\delta_N^2 I(\theta_{0,D}),$$

which is the same as

$$N \int \log \frac{p_{\hat\theta_{\mathrm{MAP}}}}{p_{\theta_{0,D}}} \, d(P_N - P_{\theta_0}) + N\delta_N^2 I(\theta_{0,D}) \\ \geq N\delta_N^2 I(\hat\theta_{\mathrm{MAP}}) - N \int \log \frac{p_{\hat\theta_{\mathrm{MAP}}}}{p_{\theta_{0,D}}} \, dP_{\theta_0}. \tag{128}$$

The last term can be decomposed as

$$- \int \log \frac{p_{\hat\theta_{\mathrm{MAP}}}}{p_{\theta_{0,D}}} \, dP_{\theta_0} = - \int \log \frac{p_{\hat\theta_{\mathrm{MAP}}}}{p_{\theta_0}} \, dP_{\theta_0} + \int \log \frac{p_{\theta_{0,D}}}{p_{\theta_0}} \, dP_{\theta_0} \\ = \tfrac{1}{2}\|\mathscr{G}(\hat\theta_{\mathrm{MAP}}) - \mathscr{G}(\theta_0)\|^2_{L^2(\mathcal{O})} - \tfrac{1}{2}\|\mathscr{G}(\theta_{0,D}) - \mathscr{G}(\theta_0)\|^2_{L^2(\mathcal{O})},$$

where we have used a standard computation of likelihood ratios (see also [45, Lemma 23]). The result follows from the last two displays after dividing by N. ■

The following result can be proved by adapting techniques from M-estimation [100] (see also [81, 99]) to the present situation. We will make crucial use of the concentration Lemma 3.12.

Proposition 4.11. *Let $\alpha > d$. Suppose that $\|\theta_0\|_{h^\alpha} \leq c_0$ and D is such that $\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_{0,D})\|_{L^2} \leq c_1\delta_N$ for some $c_0, c_1 > 0$. Then for any $c \geq 1$ we can choose $C = C(c, c_0, c_1)$ large enough so that every $\hat\theta_{\mathrm{MAP}}$ maximising (126) satisfies*

$$P_{\theta_0}^N\big(\tfrac{1}{2}\|\mathscr{G}(\hat\theta_{\mathrm{MAP}}) - \mathscr{G}(\theta_0)\|^2_{L^2} + \delta_N^2 I(\hat\theta_{\mathrm{MAP}}) > C\delta_N^2\big) \lesssim e^{-c^2 N\delta_N^2}. \tag{129}$$

Proof. We define functionals

$$\tau(\theta, \theta') = \tfrac{1}{2}\|\mathscr{G}(\theta) - \mathscr{G}(\theta')\|^2_{L^2} + \delta_N^2 I(\theta), \quad \theta \in \mathbb{R}^D, \ \theta' \in h^\alpha,$$

and empirical processes

$$W_N(\theta) = \int \log \frac{p_\theta}{p_{\theta_{0,D}}} \, d(P_N - P_{\theta_0}), \quad W_{N,0}(\theta) = \int \log \frac{p_\theta}{p_{\theta_0}} \, d(P_N - P_{\theta_0}), \quad \theta \in \mathbb{R}^D,$$

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

$$P_{\theta_0}^N\big(\tau(\hat\theta_{\mathrm{MAP}}, \theta_0) > C\delta_N^2, \\ W_N(\hat\theta_{\mathrm{MAP}}) \geq \tau(\hat\theta_{\mathrm{MAP}}, \theta_0) - \delta_N^2 I(\theta_{0,D}) - \|\mathscr{G}(\theta_{0,D}) - \mathscr{G}(\theta_0)\|^2_{L^2}/2\big).$$

Since

$$I(\theta_{0,D}) = \|\theta_{0,D}\|^2_{h^\alpha}/2 \leq \|\theta_0\|^2_{h^\alpha}/2 \leq c_0^2/2$$

and

$$\|\mathscr{G}(\theta_{0,D}) - \mathscr{G}(\theta_0)\|^2_{L^2} \leq c_1^2\delta_N^2$$

by hypothesis, we can choose C large enough so that the last probability is bounded by

$$
\begin{aligned}
P_{\theta_0}^N\big(\tau(\hat\theta_{\mathrm{MAP}},\theta_0) > C\delta_N^2,\ |W_N(\hat\theta_{\mathrm{MAP}})| \ge \tau(\hat\theta_{\mathrm{MAP}},\theta_0)/2\big)& \\
\le \sum_{s=1}^{\infty} P_{\theta_0}^N\Big(\sup_{\theta\in\mathbb{R}^D:\,2^{s-1}C\delta_N^2\le\tau(\theta,\theta_0)\le 2^sC\delta_N^2} |W_{N,0}(\theta)| \ge 2^sC\delta_N^2/8\Big)& \\
+ P_{\theta_0}^N\big(|W_{N,0}(\theta_{0,D})| \ge C\delta_N^2/4\big)& \\
\le 2\sum_{s=1}^{\infty} P_{\theta_0}^N\Big(\sup_{\theta\in\Theta_s} |W_{N,0}(\theta)| \ge 2^sC\delta_N^2/8\Big),& \qquad (130)
\end{aligned}
$$

where, for $s \in \mathbb{N}$,

$$
\begin{aligned}
\Theta_s &:= \{\theta\in\mathbb{R}^D : \tau(\theta,\theta_0) \le 2^sC\delta_N^2\} \\
&= \{\theta\in\mathbb{R}^D : \|\mathscr{G}(\theta)-\mathscr{G}(\theta_0)\|_{L^2}^2 + \delta_N^2\|\theta\|_{h^\alpha}^2 \le 2^{s+1}C\delta_N^2\}, \qquad (131)
\end{aligned}
$$

and where we have used the fact 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 $(Y_i, X_i) \overset{\text{i.i.d.}}{\sim} P_{\theta_0}$,

$$
\begin{aligned}
\ell_N(\theta) - \ell_N(\theta_0) &= -\frac{1}{2}\sum_{i=1}^{N}[(\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta)(X_i) + \varepsilon_i)^2 - \varepsilon_i^2] \\
&= -\sum_{i=1}^{N}(\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta)(X_i))\varepsilon_i - \frac{1}{2}\sum_{i=1}^{N}(\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta)(X_i))^2, \qquad (132)
\end{aligned}
$$

so that we have to deal with two empirical processes separately. We first bound

$$
\sum_{s=1}^{\infty} P_{\theta_0}^N\Big(\sup_{\theta\in\Theta_s} |Z_N(\theta)| \ge \sqrt{N}\,2^sC\delta_N^2/16\Big) \qquad (133)
$$

where

$$
Z_N = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} h_\theta(X_i)\varepsilon_i, \quad h_\theta = \mathscr{G}(\theta_0) - \mathscr{G}(\theta), \quad \theta\in\Theta=\Theta_s,\ s\in\mathbb{N},
$$

is as in Lemma 3.12. We will apply that lemma with bounds (recalling $\mathrm{vol}(\mathcal{O}) = 1$)

$$
\begin{aligned}
&E^X[h_\theta^2(X)] = \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2}^2 \le 2^{s+1}C\delta_N^2 =: \sigma_s^2, \\
&\|h_\theta\|_\infty \le 2\sup_\theta \|\mathscr{G}(\theta)\|_\infty \le U < \infty
\end{aligned} \qquad (134)
$$

uniformly in all $\theta \in \Theta_s$, for some fixed constant $U = U(g, \mathcal{O})$ (cf. (18)). For the entropy bounds, we use the fact that on each slice $\sup_{\theta\in\Theta_s} \|F_\theta\|_{H^\alpha} \le \sqrt{2C}\, 2^{s/2}$, which for $\alpha > d$ implies (using [44, (4.184)] and standard extension properties of Sobolev norms)

$$\log N(\{F_\theta : \theta \in \Theta_s\}, \|\cdot\|_\infty, \rho) \le K(\sqrt{C}\, 2^{s/2}/\rho)^{d/\alpha}, \quad \rho > 0,$$

for some constant $K = K(\alpha, d)$. Since the map $F_\theta \mapsto \mathcal{G}(\theta)$ is Lipschitz for the $\|\cdot\|_\infty$-norm (Lemma 4.6) we deduce that also

$$\log N\big(\{h_\theta = \mathcal{G}(\theta) - \mathcal{G}(\theta_0) : \theta \in \Theta_s\}, \|\cdot\|_\infty, \rho\big) \le K'(\sqrt{C}\, 2^{s/2}/\rho)^{d/\alpha}, \quad \rho > 0, \tag{135}$$

and as a consequence, for $\alpha > d$ and $J_2(\mathcal{H}), J_\infty(\mathcal{H})$ defined in Lemma 3.12,

$$\begin{aligned} J_2(\mathcal{H}) &\lesssim \int_0^{4\sigma_s} \left(\frac{\sqrt{C}\, 2^{s/2}}{\rho}\right)^{d/2\alpha} d\rho \lesssim C^{d/(4\alpha)} 2^{sd/(4\alpha)} \sigma_s^{1-d/(2\alpha)}, \\ J_\infty(\mathcal{H}) &\lesssim \int_0^{4U} \left(\frac{\sqrt{C}\, 2^{s/2}}{\rho}\right)^{d/\alpha} d\rho \lesssim C^{d/(2\alpha)} 2^{sd/(2\alpha)} U^{1-d/\alpha}. \end{aligned} \tag{136}$$

The sum in (133) can now be bounded by Lemma 3.12 with $x = c^2 N 2^s \delta_N^2$ and the choices of σ_s, U in (134) for $C > 0$ large enough:

$$\sum_{s=1}^\infty P_{\theta_0}^N \Big(\sup_{\theta\in\Theta_s} |Z_N(\theta)| \ge \sqrt{N}\, \sigma_s^2/32\Big) \le 2 \sum_{s\in\mathbb{N}} e^{-c^2 2^s N \delta_N^2} \lesssim e^{-c^2 N \delta_N^2} \tag{137}$$

since then, by definition of δ_N, for $\alpha > d$ and C large enough, the quantities

$$\mathcal{J}_2(\mathcal{H}) \lesssim C^{d/(4\alpha)} 2^{sd/(4\alpha)} (2^{s/2}\sqrt{C}\, \delta_N)^{1-d/(2\alpha)} \lesssim \frac{1}{\sqrt{C}} \sqrt{N}\, \sigma_s^2, \quad \sigma_s \sqrt{x} \le \frac{c}{\sqrt{2C}} \sqrt{N}\, \sigma_s^2, \tag{138}$$

and

$$\frac{1}{\sqrt{N}} \mathcal{J}_\infty(\mathcal{H}) \lesssim \frac{C^{d/(2\alpha)} 2^{sd/(2\alpha)}}{\sqrt{N}} \lesssim \frac{1}{C^{d/(2\alpha)-1}} \sqrt{N}\, \sigma_s^2, \quad \frac{x}{\sqrt{N}} = \frac{c^2}{2C} \sqrt{N}\, \sigma_s^2 \tag{139}$$

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 (132), which is bounded by

$$\sum_{s\in\mathbb{N}} P_{\theta_0}^N \Big(\sup_{\theta\in\Theta_s} |Z'_N(\theta)| \ge \sqrt{N}\, 2^s C \delta_N^2/16\Big) \tag{140}$$

where Z'_N is now the centred empirical process

$$Z'_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (h_\theta - E^X[h_\theta(X)])$$

with

$$\mathcal{H} = \{h_\theta = (\mathcal{G}(\theta) - \mathcal{G}(\theta_0))^2 : \theta \in \Theta_s\},$$

to which we will again apply Lemma 3.12. Just as in (134) the envelopes of this process are uniformly bounded by a fixed constant, again denoted by U, which implies in particular that the bounds (136) also apply to $\mathcal{H}$ as then, for some constant $c_U > 0$,

$$\|h_\theta - h_{\theta'}\|_\infty \le c_U \|\mathscr{G}(\theta) - \mathscr{G}(\theta')\|_\infty.$$

Moreover, on each slice Θ_s the weak variances are bounded by

$$E^X[h_\theta^2(X)] \le c'_U \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2}^2 \le c'_U \sigma_s^2$$

with σ_s as in (134) and some $c'_U > 0$. We see that all bounds required to obtain (133) apply to the process Z'_N as well, and hence the series in (130) is indeed bounded as required in the proposition, completing the proof. ∎

From a stability estimate for $\theta \mapsto \mathscr{G}(\theta)$ we now obtain the following convergence rate for $\|\hat{\theta}_{\mathrm{MAP}} - \theta_0\|_{\ell^2}$ which in turn also bounds $\|\hat{\theta}_{\mathrm{MAP}} - \theta_{0,D}\|_{\mathbb{R}^D}$.

Theorem 4.12. *Let $Z^{(N)} \sim P_{\theta_0}^N$ be as in (20) where $\theta_0 \in h^\alpha$, $\alpha > d$, $d \le 3$. Define*

$$\bar{\delta}_N := N^{-r(\alpha)} \quad \textit{where} \quad r(\alpha) = \frac{\alpha}{2\alpha + d}\,\frac{\alpha}{\alpha + 2}.$$

Suppose $\|\theta_0\|_{h^\alpha} \le c_0$ and that D is such that $\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_{0,D})\|_{L^2} \le c_1\delta_N$ for some constants $c_0, c_1 > 0$. Then given $c \ge 1$ 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 $\hat{\theta}_{\mathrm{MAP}}$ satisfying (126), one has

$$P_{\theta_0}^N(\|\hat{\theta}_{\mathrm{MAP}} - \theta_0\|_{\ell^2} \le \bar{C}\bar{\delta}_N,\ \|\hat{\theta}_{\mathrm{MAP}}\|_{h^\alpha} \le \bar{C}) \ge 1 - \bar{c}e^{-c^2 N\delta_N^2}. \tag{141}$$

Proof. By Proposition 4.11 we can restrict to events

$$T_N := \{\|\mathscr{G}(\hat{\theta}_{\mathrm{MAP}}) - \mathscr{G}(\theta_0)\|_{L^2}^2 \le 2C\delta_N^2,\ \|F_{\hat{\theta}_{\mathrm{MAP}}}\|_{H^\alpha} = \|\hat{\theta}_{\mathrm{MAP}}\|_{h^\alpha} \le \sqrt{2C}\} \tag{142}$$

of sufficiently high $P_{\theta_0}^N$-probability. If we write $\hat{f} = \Phi \circ F_{\hat{\theta}_{\mathrm{MAP}}}$ for Φ from (17) then by (176), on the events T_N we also have $\|\hat{f}\|_{H^\alpha} \le C'$ and $\|\hat{f}\|_\infty \le C'$, for some $C' > 0$. We write $u_{\hat{f}} = \mathscr{G}(\hat{\theta}_{\mathrm{MAP}})$ for the unique solution of the Schrödinger equation (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 the fact that the denominator u_f is bounded away from zero by a constant $C'' > 0$ depending only on $\|f\|_\infty, \mathcal{O}, g$ (see (118)). Then using the multiplication and interpolation inequalities (7), (8), and the regularity estimate from (174) and (176), we have, for $t = \alpha/(\alpha + 2)$,

$$\begin{aligned}\|\hat{f} - f_0\|_{L^2} &\lesssim \|u_{\hat{f}} - u_{f_0}\|_{H^2} \lesssim \|\mathscr{G}(\hat{\theta}_{\mathrm{MAP}}) - \mathscr{G}(\theta_0)\|_{L^2}^t \|u_{\hat{f}} - u_{f_0}\|_{H^{\alpha+2}}^{1-t} \\ &\lesssim \delta_N^t(\|\hat{f}\|_{H^\alpha} + \|f_0\|_{H^\alpha}) \lesssim \delta_N^t \end{aligned} \tag{143}$$

on the event T_N. From a Sobolev imbedding (for some $\kappa > 0$) and applying (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 the fact that

$\inf_x f_0(x) > K_{\min}$ 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

$$\|\hat{\theta}_{\mathrm{MAP}} - \theta_0\|_{\ell^2} \leq \|F_{\hat{\theta}_{\mathrm{MAP}}} - F_{\theta_0}\|_{L^2} = \|\Phi^{-1} \circ \hat{f} - \Phi^{-1} \circ f_0\|_{L^2}$$
$$\lesssim \|\hat{f} - f_0\|_{L^2} \lesssim \delta_N^t$$

on the events T_N, where in the last inequality we have used regularity of the inverse link function $\Phi^{-1} : [K_{\min} + k, \infty) \to \mathbb{R}$ and (177). This completes the proof. ∎

4.2.2. Posterior contraction rates. We now study the full posterior distribution (24) arising from the Gaussian prior Π for θ from (23). The result we shall prove parallels Theorem 4.12 but holds for most of the 'mass' of the posterior measure instead of just for its 'mode' $\hat{\theta}_{\mathrm{MAP}}$. This requires very different techniques and we rely on ideas from Bayesian nonparametrics [42, 102], specifically recent progress [74] that allows one to deal with non-linear settings (see also [45]).

In the proof of Theorem 4.14 to follow we will require control of the posterior 'normalising factors', expressed via sets

$$\mathcal{C}_N = \mathcal{C}_{N,K} = \left\{ \int_{\mathbb{R}^D} e^{\ell_N(\theta) - \ell_N(\theta_0)} \, d\Pi(\theta) \geq \Pi(B(\delta_N)) \exp\{-(1+K)N\delta_N^2\} \right\}, \tag{144}$$

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 (110)).

Theorem 4.13. *Let $Z^{(N)}, \theta_0, \alpha, d, \bar{\delta}_N$ be as in Theorem 4.12 and let $\Pi(\cdot \mid Z^{(N)})$ denote the posterior distribution from (24). Suppose that $\|\theta_0\|_{h^\alpha} \leq c_0$ and $D \leq c_2 N \delta_N^2$ is such that*

$$\|\mathcal{G}(\theta_0) - \mathcal{G}(\theta_{0,D})\|_{L^2(\mathcal{O})} \leq c_1 \delta_N \tag{145}$$

for some finite constants $c_0, c_2 > 0$ and $0 < c_1 < 1/2$. 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,

$$P_{\theta_0}^N \Big(\{\Pi(\theta : \|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \leq L\bar{\delta}_N, \ \|\theta\|_{h^\alpha} \leq L \mid Z^{(N)}) \geq 1 - e^{-aN\delta_N^2}\}, \mathcal{C}_{N,K} \Big)$$
$$\geq 1 - c' e^{-c'' N\delta_N^2}. \tag{146}$$

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 [45, Theorem 13]. By [45, Lemma 23] and (18) we can lower bound $\Pi_N(\mathcal{B}_N)$ in (A5) in [45] by our $\Pi_N(B(\delta_N))$ (after adjusting the choice of δ_N in [45] by a multiplicative constant).

Then using (145), [44, Corollary 2.6.18], and ultimately [65, Theorem 1.2] combined with [44, (4.184)] we have, for $\theta' \sim N(0, \Lambda_\alpha^{-1})$,

$$\begin{aligned}\Pi_N(\|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2(\mathcal{O})} < \delta_N) &\geq \Pi_N(\|\mathscr{G}(\theta) - \mathscr{G}(\theta_{0,D})\|_{L^2(\mathcal{O})} < \delta_N/2) \\ &\geq \Pi_N(\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} < \delta_N/(2c_g)) \\ &\geq e^{-N\delta_N^2\|\theta_{0,D}\|_{h^\alpha}^2/2} \Pr(\|\theta'\|_{\mathbb{R}^D} < \sqrt{N}\,\delta_N^2/(2c_g)) \geq e^{-\bar{d}N\delta_N^2} \end{aligned} \tag{147}$$

for some $\bar{d} > 0$. From this we deduce further from Borell's Gaussian isoperimetric inequality [17] (in the form of [44, Theorem 2.6.12]), arguing just as in [45, Lemma 17] (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 (13) and the hypothesis on D imply that for $\bar{L}$ large enough we have

$$\|\theta_1\|_{h^\alpha} \lesssim D^{\alpha/d}\|\theta_1\|_{\mathbb{R}^D} \leq (c_2 N\delta_N^2)^{\alpha/d} M\delta_N \leq \bar{L}/2 \tag{148}$$

and then also

$$\Pi_N(\mathcal{A}_N^c) \leq 2e^{-BN\delta_N^2} \quad \text{where} \quad \mathcal{A}_N = \{\theta \in \mathbb{R}^D : \|\theta\|_{h^\alpha} \leq \bar{L}\}. \tag{149}$$

The $\|\cdot\|_\infty$-covering numbers of the implied set of regression functions $\mathscr{G}(\theta)$ satisfy the bounds

$$\begin{aligned}\log N(\{\mathscr{G}(\theta) : \theta \in \mathcal{A}_N\}, \|\cdot\|_\infty, \delta_N) &\lesssim \log N(\{F_\theta : \theta \in \mathcal{A}_N\}, \|\cdot\|_\infty, c\delta_N) \\ &\lesssim \log N(\{F : \|F\|_{H^\alpha(\mathcal{O})} \leq \bar{L}\}, \|\cdot\|_\infty, c\delta_N) \lesssim N\delta_N^2,\end{aligned}$$

for some $c > 0$, using the fact that the map $F_\theta \mapsto \mathscr{G}(\theta)$ is globally Lipschitz for the $\|\cdot\|_\infty$-norm (Lemma 4.6) and also the bound in [44, (4.184)]. By (18) and [45, Lemma 22] 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 [45] now imply that for any $a > 0$ there exists L large enough,

$$P_{\theta_0}^N\left(\Pi(\{\theta : \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2} > L\delta_N\} \cup \mathcal{A}_N^c \mid Z^{(N)}) \leq e^{-aN\delta_N^2}\right) \to 0 \tag{150}$$

as $N \to \infty$. The convergence in probability to zero obtained in [45, proof of Theorem 13] is in fact exponentially fast, as required in (146): This is true by virtue of the bound to follow in the next display (which forms part of the proof in [45] as well), and since the type-one testing errors in [45, (39)] are controlled at the required exponential rate (via [44, Theorem 7.1.4]). The inequality

$$P_{\theta_0}^N\left(\int_{B(\delta_N)} e^{\ell_N(\theta) - \ell_N(\theta_0)}\, d\Pi(\theta) \geq \Pi(B(\delta_N))\exp\{-(1+K)N\delta_N^2\}\right) \leq c'e^{-c''N\delta_N^2},$$

bounding $P_{\theta_0}^N(\mathcal{C}_{N,K}^c)$ as required in the theorem follows from Lemma 4.15 below for large enough K and $\bar{C} = 1/2$.

Now to conclude, we can define subsets of $\mathbb{R}^D$ as

$$\begin{aligned}\Theta_N &:= \{\theta : \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2} \leq L\delta_N\} \cap \mathcal{A}_N \\ &= \{\theta : \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2} \leq L\delta_N, \ \|F_\theta\|_{H^\alpha} = \|\theta\|_{h^\alpha} \leq \bar{L}\}\end{aligned}$$

parallelling the events T_N from (142) above. Then arguing as in and after (143), one shows that

$$\Theta_N \subset \tilde{\Theta}_N = \{\theta : \|\theta - \theta_0\|_{\mathbb{R}^D} \leq LN^{-r(\alpha)}, \ \|\theta\|_{h^\alpha} \leq L\},$$

increasing also the constant L if necessary, and hence the posterior probability of this event is also lower bounded by $\Pi(\tilde{\Theta}_N \mid Z^{(N)}) \geq 1 - e^{-aN\delta_N^2}$, with the desired $P_{\theta_0}^N$-probability, proving the theorem, since $\|\theta - \theta_0\|_{\ell^2} \leq \|\theta - \theta_{0,D}\|_{\mathbb{R}^D}$. ■

Moreover, a quantitative uniform integrability argument from [74, Section 5.4.5] (see the proof of Theorem 4.14, term III, below) then also gives a convergence rate for the posterior mean $E^\Pi[\theta \mid Z^{(N)}]$ towards θ_0, namely that for L large enough there exist $\bar{c}', \bar{c}'' > 0$ such that

$$P_{\theta_0}^N(\|E^\Pi[\theta \mid Z^{(N)}] - \theta_0\|_{\ell^2} > L\bar{\delta}_N) \leq \bar{c}' e^{-\bar{c}'' N\delta_N^2}. \tag{151}$$

4.2.3. Globally log-concave approximation of the posterior in Wasserstein distance. Recall the surrogate posterior measure $\tilde{\Pi}(\cdot \mid Z^{(N)})$ from (27) with log-density

$$\log \tilde{\pi}_N(\theta) = \text{const} + \tilde{\ell}_N(\theta) - \frac{N\delta_N^2}{2}\|\theta\|_{h^\alpha}^2, \quad \theta \in \mathbb{R}^D, \tag{152}$$

with θ_{init} and parameters ϵ, K chosen as in Condition 2.2, and with $\delta_N = N^{-\alpha/(2\alpha+d)}$. We now prove the main result of this section.

Theorem 4.14. *Assume Condition* 2.3 *and let* $\tilde{\Pi}(\cdot \mid Z^{(N)})$ *be the probability measure of density given in* (27) *with* $K, \varepsilon > 0$ *chosen as in Condition* 2.2. *Then for some* $a_1, a_2 > 0$ *and all* $N \in \mathbb{N}$,

$$P_{\theta_0}^N\left(W_2^2(\tilde{\Pi}(\cdot \mid Z^{(N)}), \Pi(\cdot \mid Z^{(N)})) > e^{-N\delta_N^2}\right) \leq a_1 e^{-a_2 N\delta_N^2}.$$

Proof. In the proof we will require a new sequence

$$\tilde{\delta}_N = N^{(-\alpha+2)/(2\alpha+d)}\sqrt{\log N} \tag{153}$$

describing the ‘rate of contraction’ of the surrogate posterior obtained below. We first notice that the definitions of $\bar{\delta}_N$ (from Theorem 4.12) and of δ_N imply by straightforward calculations and using $D \lesssim N\delta_N^2, \alpha > 6$, the asymptotic relations as $N \to \infty$,

$$\delta_N D^{2/d}\sqrt{\log N} = O(\tilde{\delta}_N), \quad \delta_N \ll \bar{\delta}_N \ll \tilde{\delta}_N \ll \frac{1}{\log N} D^{-4/d}, \tag{154}$$

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 1).

Geometry of the surrogate posterior. To set things up, consider MAP estimates $\hat{\theta}_{\mathrm{MAP}}$ from (126). In view of (18), the function q_N to be maximised over $\mathbb{R}^D$ in (126) 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 $\mathscr{G}$ is continuous from $\mathbb{R}^D$ to $L^\infty(\mathcal{O})$, by Lemma 4.6), and hence attains its maximum at some $\hat{\theta}_M \in M$, which must be a global maximiser of q_N since $q_N(\hat{\theta}_M) \ge q_N(0) > \inf_{\theta\in M^c} q_N(\theta)$. We conclude that a maximiser $\hat{\theta}_{\mathrm{MAP}}$ exists (one shows that it can be taken to be measurable, [44, Exercise 7.2.3]).

In view of Proposition 4.1, Theorem 4.12, Theorem B.6 (and the remark before it) and $\alpha > 6$, we may restrict ourselves in the rest of the proof to the event

$$\begin{aligned}\mathcal{S}_N := &\left\{\|\theta_{\mathrm{init}} - \theta_{0,D}\|_{\mathbb{R}^D} \le \frac{1}{8\log ND^{4/d}}\right\} \cap \left\{\inf_{\theta\in\mathcal{B}_{1/\log N}} \lambda_{\min}(-\nabla^2\ell_N(\theta)) \ge \underline{c}ND^{-4/d}\right\}\\ &\cap \left\{\sup_{\theta\in\mathcal{B}_{1/\log N}} [|\ell_N(\theta)| + \|\nabla\ell_N(\theta)\|_{\mathbb{R}^D} + \|\nabla^2\ell_N(\theta)\|_{\mathrm{op}}] < \underline{c}'N\right\}\\ &\cap \left\{\text{any } \hat{\theta}_{\mathrm{MAP}} \text{ satisfies } \|\hat{\theta}_{\mathrm{MAP}} - \theta_{0,D}\|_{\mathbb{R}^D} \le \min\left\{\frac{1}{8\log ND^{4/d}}, \bar{C}\bar{\delta}_N\right\}\right\},\end{aligned}$$

where $\mathcal{B}_\epsilon$ was defined in (99), where $\bar{C}$ is from (141) and where $\underline{c} = c_3$, $\underline{c}' = c_4$ from Proposition 4.1. On $\mathcal{S}_N$ we have the following properties of $\tilde{\ell}_N$. First, from (26),

$$\tilde{\ell}_N(\theta) = \ell_N(\theta) \quad \text{for any } \theta \text{ with } \|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \le \frac{3}{8D^{4/d}\log N}. \tag{155}$$

Moreover, by Proposition 3.6, $\log\tilde{\pi}(\cdot\,|\,Z^{(N)})$ is strongly concave in view of

$$\begin{aligned}&\sup_{\theta\in\mathcal{B}_{1/\log N},\, \vartheta\in\mathbb{R}^D,\, \|\vartheta\|_{\mathbb{R}^D}=1} \vartheta^T[\nabla^2\log\tilde{\pi}_N(\theta)]\vartheta\\ &\qquad \le \sup_{\theta\in\mathcal{B}_{1/\log N},\, \vartheta\in\mathbb{R}^D,\, \|\vartheta\|_{\mathbb{R}^D}=1} \vartheta^T[\nabla^2\tilde{\ell}_N(\theta)]\vartheta \le -\underline{c}ND^{-4/d}.\end{aligned} \tag{156}$$

Finally, any $\hat{\theta}_{\mathrm{MAP}}$ satisfies

$$0 = \nabla\log\pi(\hat{\theta}_{\mathrm{MAP}}\,|\,Z^{(N)}) = \nabla\log\tilde{\pi}(\hat{\theta}_{\mathrm{MAP}}), \tag{157}$$

from which we conclude that $\hat{\theta}_{\mathrm{MAP}}$ 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

$$\hat{\mathcal{B}}(r) = \{\theta \in \mathbb{R}^D : \|\theta - \hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D} \le r\}$$

for the Euclidean ball of radius $r > 0$ centred at $\hat\theta_{\mathrm{MAP}}$. Then using [104, Theorem 6.15] with $x_0 = \hat\theta_{\mathrm{MAP}}$ we obtain, for any $m > 0$

$$\begin{aligned}
&W_2^2(\tilde\Pi(\cdot \,|\, Z^{(N)}), \Pi(\cdot \,|\, Z^{(N)})) \\
&\quad \le 2 \int_{\mathbb{R}^D} \|\theta - \hat\theta_{\mathrm{MAP}}\|^2_{\mathbb{R}^D}\, d\big|\tilde\Pi(\cdot \,|\, Z^{(N)}) - \Pi(\cdot \,|\, Z^{(N)})\big|(\theta) \\
&\quad \le 2 \int_{\hat{\mathcal{B}}(m\tilde\delta_N)} \|\theta - \hat\theta_{\mathrm{MAP}}\|^2_{\mathbb{R}^D}\, d\big|\tilde\Pi(\cdot \,|\, Z^{(N)}) - \Pi(\cdot \,|\, Z^{(N)})\big|(\theta) \\
&\qquad + 2 \int_{\mathbb{R}^D \setminus \hat{\mathcal{B}}(m\tilde\delta_N)} \|\theta - \hat\theta_{\mathrm{MAP}}\|^2_{\mathbb{R}^D}\, d\big|\tilde\Pi(\cdot \,|\, Z^{(N)}) - \Pi(\cdot \,|\, Z^{(N)})\big|(\theta) \\
&\quad \le 2m^2\tilde\delta_N^2 \int_{\hat{\mathcal{B}}(m\tilde\delta_N)} d\big|\Pi(\cdot \,|\, Z^{(N)}) - \tilde\Pi(\cdot \,|\, Z^{(N)})\big|(\theta) \\
&\qquad + 2 \int_{\|\theta - \hat\theta_{\mathrm{MAP}}\|_{\mathbb{R}^D} > m\tilde\delta_N} \|\theta - \hat\theta_{\mathrm{MAP}}\|^2_{\mathbb{R}^D}\, d\tilde\Pi(\theta \,|\, Z^{(N)}) \\
&\qquad + 2 \int_{\|\theta - \hat\theta_{\mathrm{MAP}}\|_{\mathbb{R}^D} > m\tilde\delta_N} \|\theta - \hat\theta_{\mathrm{MAP}}\|^2_{\mathbb{R}^D}\, d\Pi(\theta \,|\, Z^{(N)}) \\
&\quad \equiv I + II + III,
\end{aligned}$$

and we now bound I, II, III in separate steps.

Term II: We can write the surrogate posterior density as

$$\tilde\pi(\theta \,|\, Z^{(N)}) = \frac{e^{\tilde\ell_N(\theta) - \tilde\ell_N(\hat\theta_{\mathrm{MAP}})}\pi(\theta)}{\int_{\mathbb{R}^D} e^{\tilde\ell_N(\theta) - \tilde\ell_N(\hat\theta_{\mathrm{MAP}})}\pi(\theta)\, d\theta}, \quad \theta \in \mathbb{R}^D,$$

and will first lower bound the normalising factor. From (154) we have for any $c > 0$ the set inclusion

$$B_N \equiv \{\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \le c\delta_N\} \subset \left\{\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \le \frac{3}{8D^{4/d}\log N}\right\}$$

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 $P^N_{\theta_0}$-probability,

$$\begin{aligned}
\int_{\mathbb{R}^D} e^{\tilde\ell_N(\theta) - \tilde\ell_N(\hat\theta_{\mathrm{MAP}})}\, d\Pi(\theta) &\ge \int_{B_N} e^{\tilde\ell_N(\theta) - \tilde\ell_N(\hat\theta_{\mathrm{MAP}})}\, d\Pi(\theta) \\
&= \int_{B_N} e^{\ell_N(\theta) - \ell_N(\hat\theta_{\mathrm{MAP}})}\, d\nu(\theta) \times \Pi(B_N) \\
&\ge e^{-\bar c N\delta_N^2}
\end{aligned}$$

for some $\bar c = \bar c(\bar d, c)$, where we have used Lemma 4.15 for our choice of B_N (permitted for an appropriate choice of $c > 0$ by (28) and since $\mathcal{G} : \mathbb{R}^D \to L^2$ is Lipschitz, see Appendix B) with $\nu = \Pi(\cdot)/\Pi(B_N)$, $\bar C = 1/2$; as well as the small ball estimate for Π in (147).

Now recall the prior (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 $\log \tilde{\pi}(\cdot \mid Z^{(N)})$ around its maximum $\hat{\theta}_{\mathrm{MAP}}$ combined with (156), (157), as

$$\begin{aligned}
&\int_{\|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}>m\tilde{\delta}_N} \|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^2 \tilde{\pi}(\theta \mid Z^{(N)})\,d\theta \\
&\le e^{\bar{c}N\delta_N^2}\int_{\|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}>m\tilde{\delta}_N} \|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^2 e^{\tilde{\ell}_N(\theta)-\tilde{\ell}_N(\hat{\theta}_{\mathrm{MAP}})}\pi(\theta)\,d\theta \\
&\le V_N\int_{\|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}>m\tilde{\delta}_N} \|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^2 e^{\tilde{\ell}_N(\theta)-\frac{N\delta_N^2}{2}\|\theta\|_{h^\alpha}^2-\tilde{\ell}_N(\hat{\theta}_{\mathrm{MAP}})+\frac{N\delta_N^2}{2}\|\hat{\theta}_{\mathrm{MAP}}\|_{h^\alpha}^2}\,d\theta \\
&= V_N\int_{\|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}>m\tilde{\delta}_N} \|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^2 e^{\log\tilde{\pi}_N(\theta)-\log\tilde{\pi}_N(\hat{\theta}_{\mathrm{MAP}})}\,d\theta \\
&\le V_N\int_{\|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}>m\tilde{\delta}_N} \|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^2 e^{-\underline{c}ND^{-4/d}\|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^2/2}\,d\theta \\
&\le 2V_N\left(\frac{4\pi}{\underline{c}ND^{-4/d}}\right)^{D/2}\Pr(\|Z\|_{\mathbb{R}^D}>m\tilde{\delta}_N),
\end{aligned}$$

where we have used $x^2e^{-cx^2} \le 2e^{-cx^2/2}$ for all $x \in \mathbb{R}$, $c \ge 1$ (and N such that $\underline{c}ND^{-4/d} \ge 1$) and where

$$Z \sim N\left(0, \frac{2}{\underline{c}D^{-4/d}N}I_{D\times D}\right).$$

Now by $D \le c_0N\delta_N^2$ and (154),

$$E\|Z\|_{\mathbb{R}^D} \le \sqrt{E\|Z\|_{\mathbb{R}^D}^2} \le \sqrt{2D/(\underline{c}D^{-4/d}N)} \le (2c_0/\underline{c})^{1/2}\delta_N D^{2/d} \le (m/2)\tilde{\delta}_N$$

for m large enough, so that

$$\Pr(\|Z\|_{\mathbb{R}^D}>m\tilde{\delta}_N) \le \Pr\big(\|Z\|_{\mathbb{R}^D}-E\|Z\|_{\mathbb{R}^D}>(m/2)\tilde{\delta}_N\big) \le e^{-m^2\underline{c}ND^{-4/d}\tilde{\delta}_N^2/16}$$

by a concentration inequality for Lipschitz functionals of D-dimensional Gaussian random vectors (e.g., [44, Theorem 2.5.7] 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 (13) and since $D \lesssim N\delta_N^2$ we have, for some $c' > 0$,

$$V_N \le e^{c'N\delta_N^2\log N},$$

so that for m large enough and using (154), the last term in the displayed array above, and hence $II/2$ is bounded by

$$2V_N\left(\frac{4\pi}{\underline{c}ND^{-4/d}}\right)^{D/2} e^{-m^2\underline{c}D^{-4/d}N\tilde{\delta}_N^2/16} \le e^{-m^2D^{-4/d}N\tilde{\delta}_N^2/32} \le \tfrac{1}{8}e^{-N\delta_N^2}.$$

Term III: We first note that Theorem 4.13 and (154) imply that for every $a > 0$ we can find m large enough such that

$$\Pi(\|\theta - \hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D} > m\tilde{\delta}_N \mid Z^{(N)}) \le \Pi(\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} > m\bar{\delta}_N - \|\hat{\theta}_{\mathrm{MAP}} - \theta_{0,D}\|_{\mathbb{R}^D} \mid Z^{(N)}) \le \Pi(\|\theta - \theta_{0,D}\|_{\mathbb{R}^D} > m\bar{\delta}_N/2 \mid Z^{(N)}) \le e^{-aN\delta_N^2}$$

on events $\mathcal{S}_N' \subset \mathcal{S}_N$ of sufficiently high probability. Moreover, again by Theorem 4.13, we can further restrict the argument that follows to the event $\mathcal{C}_{N,K}$ from (144) 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 (147), we have

$$\begin{aligned}
&P_{\theta_0}^N\Big(\mathcal{C}_{N,K} \cap \mathcal{S}_N', \int_{\|\theta-\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D} > m\tilde{\delta}_N} \|\theta - \hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^2 \, d\Pi(\theta \mid Z^{(N)}) > e^{-N\delta_N^2}/8\Big) \\
&\le P_{\theta_0}^N(\mathcal{C}_{N,K} \cap \mathcal{S}_N', \\
&\qquad \Pi(\|\theta - \hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D} > m\tilde{\delta}_N \mid Z^{(N)}) E^{\Pi}[\|\theta - \hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^4 \mid Z^{(N)}] > e^{-2N\delta_N^2}/64) \\
&\le P_{\theta_0}^N\Big(\mathcal{S}_N', e^{(1+K+\bar{d}+2-a)N\delta_N^2} \int_{\mathbb{R}^D} \|\theta - \hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^4 e^{\ell_N(\theta)-\ell_N(\theta_0)} \, d\Pi(\theta) > 1/64\Big) \\
&\lesssim e^{(1+K+\bar{d}+2-a)N\delta_N^2} \int_{\mathbb{R}^D} (1 + \|\theta\|_{\mathbb{R}^D}^4) \, d\Pi(\theta) \le e^{-a_2 N\delta_N^2}
\end{aligned}$$

whenever m and then a are large enough, since Π has uniformly bounded fourth moments and since $\|\hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}$ 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 (154) that for fixed $m > 0$ and all N large enough

$$\hat{\mathcal{B}}(m\tilde{\delta}_N) \subseteq \{\theta : \|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \le 3/(8D^{4/d} \log N)\}.$$

On the latter set, by (155), the probability measures $\tilde{\Pi}(\cdot \mid Z^{(N)})$ and $\Pi(\cdot \mid Z^{(N)})$ coincide up to a normalising factor, and thus we can represent their Lebesgue densities as

$$\tilde{\pi}(\theta \mid Z^{(N)}) = p_N \pi(\theta \mid Z^{(N)}), \quad \theta \in \hat{\mathcal{B}}(m\tilde{\delta}_N),$$

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 $\|\theta - \hat{\theta}_{\mathrm{MAP}}\|_{\mathbb{R}^D}^2$), we have both

$$\begin{aligned}
&p_N \Pi(\hat{\mathcal{B}}(m\tilde{\delta}_N) \mid Z^{(N)}) = \tilde{\Pi}(\hat{\mathcal{B}}(m\tilde{\delta}_N) \mid Z^{(N)}) \ge 1 - e^{-N\delta_N^2}/8, \ \text{so } 1 - e^{-N\delta_N^2}/8 \le p_N, \\
&p_N^{-1} \tilde{\Pi}(\hat{\mathcal{B}}(m\tilde{\delta}_N) \mid Z^{(N)}) = \Pi(\hat{\mathcal{B}}(m\tilde{\delta}_N) \mid Z^{(N)}) \ge 1 - e^{-N\delta_N^2}/8, \ \text{so } 1 - e^{-N\delta_N^2}/8 \le 1/p_N
\end{aligned}$$

on events of sufficiently high $P_{\theta_0}^N$-probability. On these events necessarily

$$p_N \in \left[1 - e^{-N\delta_N^2}/8, \frac{1}{1 - e^{-N\delta_N^2}/8}\right]$$

and so for N large enough

$$\int_{\hat{\mathcal{B}}(m\tilde{\delta}_N)} d\big|\Pi(\cdot\,|\,Z^{(N)}) - \tilde{\Pi}(\cdot\,|\,Z^{(N)})\big|(\theta) \\ = |1-p_N| \int_{\hat{\mathcal{B}}(m\tilde{\delta}_N)} \pi(\theta\,|\,Z^{(N)})\,d\theta \le |1-p_N| \le e^{-N\delta_N^2}/4,$$

which is obvious for $p_N \le 1$ 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. ■

4.2.4. An 'exponential' small ball lemma.

Lemma 4.15. *Let $\mathcal{G}$ be as in (17) and let ν be a probability measure on some ($\ell^2(\mathbb{N})$-measurable) set*

$$B_N \subseteq \{\theta \in h^\alpha : \|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2}^2 \le 2\bar{C}\delta_N^2\} \quad \text{for some } \bar{C} > 0. \tag{158}$$

Then for ℓ_N from (22) there exists $b > 0$ such that for every $K > 0$ large enough,

$$P_{\theta_0}^N\left(\int_{B_N} e^{\ell_N(\theta)-\ell_N(\hat{\theta}_{\mathrm{MAP}})}\,d\nu(\theta) \le e^{-(1+K)\bar{C}^2N\delta_N^2}\right) \lesssim e^{-bN\delta_N^2}. \tag{159}$$

The same conclusion holds true with $\ell_N(\hat{\theta}_{\mathrm{MAP}})$ replaced by $\ell_N(\theta_0)$.

Proof. We proceed as in [44, Lemma 7.3.2] to deduce from Jensen's inequality (applied to log and $\int(\cdot)\,d\nu$) that, for P_N the empirical measure from (127), the probability in question is bounded by

$$P_{\theta_0}^N\bigg(\int\int_{B_N} \log\frac{p_\theta}{p_{\hat{\theta}_{\mathrm{MAP}}}}\,d\nu(\theta)\,d(P_N - P_{\theta_0}) \\ \le -(1+K)\bar{C}^2\delta_N^2 - \int\int_{B_N}\log\frac{p_\theta}{p_{\hat{\theta}_{\mathrm{MAP}}}}\,d\nu(\theta)\,dP_{\theta_0}\bigg).$$

Now just as in the proof of Lemma 4.10 we see that for all $\theta \in B_N$,

$$-\int \log\frac{p_\theta}{p_{\hat{\theta}_{\mathrm{MAP}}}}\,dP_{\theta_0} = -\int\log\frac{p_\theta}{p_{\theta_0}}\,dP_{\theta_0} - \int\log\frac{p_{\theta_0}}{p_{\hat{\theta}_{\mathrm{MAP}}}}\,dP_{\theta_0} \\ = \tfrac{1}{2}\|\mathcal{G}(\theta) - \mathcal{G}(\theta_0)\|_{L^2}^2 - \tfrac{1}{2}\|\mathcal{G}(\hat{\theta}_{\mathrm{MAP}}) - \mathcal{G}(\theta_0)\|_{L_2}^2 \le \bar{C}^2\delta_N^2$$

so that using also Fubini's theorem the last probability can be bounded by

$$P_{\theta_0}^N\left(\sqrt{N}\int\int_{B_N}\log\frac{p_{\theta_0}}{p_\theta}\,d\nu(\theta)\,d(P_N - P_{\theta_0}) \ge K\bar{C}^2\sqrt{N}\,\delta_N^2/2\right) \\ + P_{\theta_0}^N\left(\sqrt{N}\int\log\frac{p_{\hat{\theta}_{\mathrm{MAP}}}}{p_{\theta_0}}\,d(P_N - P_{\theta_0}) \ge K\bar{C}^2\sqrt{N}\,\delta_N^2/2\right).$$

For the first probability we decompose as in (132) and consider Z_N as in Lemma 3.12 for fixed h_θ equal to either h_1 or h_2, where

$$h_1(x) = \int_{B_N} (\mathscr{G}(\theta)(x) - \mathscr{G}(\theta_0)(x))\, d\nu(\theta),$$
$$h_2(x) = \int_{B_N} (\mathscr{G}(\theta)(x) - \mathscr{G}(\theta_0)(x))^2\, d\nu(\theta).$$

To each of these we apply Bernstein's inequality (96) with $x = N\sigma^2$ and K large enough to obtain the desired exponential bound, using uniform boundedness $\|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_\infty \leq 2U$ from (18) and Jensen's inequality in the variance estimates

$$E^X[h_1^2(X)] \leq 2\bar{C}^2\delta_N^2 \equiv \sigma^2$$

in the first case and

$$E^X[h_2^2(X)] \leq 4U^2 \int_{B_N} \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2}^2\, d\nu(\theta) \leq 8U^2\bar{C}\delta_N^2 \equiv \sigma^2$$

for the second case. [This already proves the case where $\hat{\theta}_{\mathrm{MAP}}$ 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 4.11, it suffices to bound for some $C > 0$,

$$P_{\theta_0}^N\left(\sup_{\|\theta\|_{h^\alpha}\leq 2C,\, \|\mathscr{G}(\theta)-\mathscr{G}(\theta_0)\|_{L_2}^2\leq 2C\delta_N^2} \sqrt{N}\left|\int \log\frac{p_\theta}{p_{\theta_0}}\, d(P_N - P_{\theta_0})\right| \geq K\bar{C}^2\sqrt{N}\,\delta_N^2/2\right).$$

This term corresponds to the empirical process bounded in and after (130) for $s = 1$. Choosing K large enough the proof there now applies directly, giving the desired exponential bound. ■

Appendix A. 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*; see [32, 36, 37] and also the classical reference [89]. Our presentation follows the recent article [37].

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

$$\mu(B) = \frac{\int_B e^{-U(\theta)}\, d\theta}{\int_{\mathbb{R}^D} e^{-U(\theta)}\, d\theta}, \quad B \subseteq \mathbb{R}^D \text{ measurable.} \tag{160}$$

Following [37, H1, H2] we will assume that the potential U has a Λ-Lipschitz gradient and is m-strongly convex.

Assumption A.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 A.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 [32, Theorem 1] and [21, (9.18)]).

Proposition A.2. *Suppose* $U : \mathbb{R}^D \to \mathbb{R}$ *satisfies Assumption* A.1. *Then the gradient descent algorithm given by*

$$\vartheta_{k+1} = \vartheta_k - \frac{1}{2\Lambda}\nabla U(\vartheta_k), \quad k = 0, 1, 2, \dots,$$

satisfies

$$\|\vartheta_k - \theta_U\|^2_{\mathbb{R}^D} \leq \frac{2(U(\vartheta_0) - U(\theta_U))}{m}\left(1 - \frac{m}{2\Lambda}\right)^k, \quad k = 0, 1, 2, \dots.$$

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

$$dL_t = -\nabla U(L_t)\,dt + \sqrt{2}\,dW_t, \quad t \geq 0,\ L_t \in \mathbb{R}^D, \tag{161}$$

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 (160) as its invariant measure. The Euler–Maruyama discretisation of the dynamics (161) gives rise to the discrete-time Markov chain $(\vartheta_k : k \geq 0)$,

$$\vartheta_{k+1} = \vartheta_k - \gamma \nabla U(\vartheta_k) + \sqrt{2\gamma}\,\xi_{k+1}, \quad k \geq 0, \tag{162}$$

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 $\mathbf{P}_{\theta_{\text{init}}}, \mathbf{E}_{\theta_{\text{init}}}$ the law and expectation operator, respectively, of the Markov chain $(\vartheta_k : k \geq 1)$ when started at a deterministic point $\vartheta_0 = \theta_{\text{init}}$. 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 $J_{\text{in}}, J \geq 0$, let us define the average of H along an ULA trajectory after 'burn-in' period J_{in} by

$$\hat{\mu}^J_{J_{\text{in}}}(H) = \frac{1}{J}\sum_{k=J_{\text{in}}+1}^{J_{\text{in}}+J} H(\vartheta_k).$$

Proposition A.3. *Suppose that U satisfies Assumption A.1 and suppose $\gamma \leq 2/(m+\Lambda)$. Then for all $J, J_{\mathrm{in}} \geq 1$, $x > 0$ and any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$, we have the concentration inequality*

$$\mathbf{P}_{\theta_{\mathrm{init}}}\big(\hat{\mu}^J_{J_{\mathrm{in}}}(H) - \mathbf{E}_{\theta_{\mathrm{init}}}[\hat{\mu}^J_{J_{\mathrm{in}}}(H)] \geq x\big) \leq \exp\bigg(-\frac{J\gamma x^2 m^2}{16\|H\|^2_{\mathrm{Lip}}(1+2/(mJ\gamma))}\bigg).$$

Proof. The statement follows directly from [37, Theorem 17], noting that $\kappa = 2m\Lambda/(m+\Lambda) \in [m, 2m]$ and that the constant $v_{N,n}(\gamma)$ from [37, (28)] can be upper bounded by

$$1 + \frac{m^{-1} + 2/(m+\Lambda)}{\gamma J} \leq 1 + 2/(m\gamma J). \qquad \blacksquare$$

Proposition A.4. *Suppose that U satisfies Assumption A.1 and let γ, J_{in}, J and H be as in Proposition A.3. Then for μ as in (160) we have*

$$W_2^2(\mathcal{L}(\vartheta_k), \mu) \leq 2(1 - m\gamma/2)^k[\|\theta_{\mathrm{init}} - \theta_U\|^2_{\mathbb{R}^D} + D/m] + b(\gamma)/2, \quad k \geq 0, \tag{163}$$

where

$$b(\gamma) = 36\frac{\gamma D\Lambda^2}{m^2} + 12\frac{\gamma^2 D\Lambda^4}{m^3}, \tag{164}$$

as well as

$$\big(\mathbf{E}_{\theta_{\mathrm{init}}}[\hat{\mu}^J_{J_{\mathrm{in}}}(H)] - E_\mu[H]\big)^2 \leq \|H\|^2_{\mathrm{Lip}} \frac{1}{J} \sum_{k=J_{\mathrm{in}}+1}^{J_{\mathrm{in}}+J} W_2^2(\mathcal{L}(\vartheta_k), \mu). \tag{165}$$

Proof. The display (165) is derived in [37, (27)]. The bound (163) follows from an application of [37, Theorem 5] 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 $\gamma \leq \min\{2/\Lambda, 1/m\} \leq \min\{2/\Lambda, 2/\kappa\}$)

$$\begin{aligned}
&\Lambda^2 D\gamma^2(\kappa^{-1} + \gamma)\bigg(2 + \frac{\Lambda^2\gamma}{m} + \frac{\Lambda^2\gamma^2}{6}\bigg)\sum_{i=1}^k (1 - \kappa\gamma/2)^{k-i} \\
&\quad \leq \Lambda^2 D\gamma^2(\kappa^{-1} + \gamma)\bigg(2 + \frac{\Lambda^2\gamma}{m} + \frac{\Lambda^2\gamma^2}{6}\bigg)\frac{2}{\kappa\gamma} \leq \Lambda^2 D\gamma\bigg(\kappa^{-2} + \frac{\gamma}{\kappa}\bigg)\bigg(6 + \frac{2\Lambda^2\gamma}{m}\bigg) \\
&\quad \leq \Lambda^2 D\gamma m^{-2}\bigg(18 + \frac{6\Lambda^2\gamma}{m}\bigg),
\end{aligned}$$

which equals (164). ■

Appendix B. Auxiliary results

B.1. Analytical properties of Schrödinger operators and link functions

Recall the inverse Schrödinger operators V_f from (106).

Lemma B.1. *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*

$$\|V_f[\psi]\|_{L^2} \leq C\|\psi\|_{L^2}, \quad \psi \in L^2(\mathcal{O}),$$
$$\|V_f[\psi]\|_\infty \leq C\|\psi\|_\infty, \quad \psi \in C(\mathcal{O}). \tag{166}$$

(ii) *For any $\psi \in L^2(\mathcal{O})$, we have*

$$\|V_f[\psi]\|_{H^2} \leq C(1 + \|f\|_\infty)\|\psi\|_{L^2}, \tag{167}$$

as well as

$$\frac{1}{C(1 + \|f\|_\infty)}\|\psi\|_{(H_0^2)^*} \leq \|V_f[\psi]\|_{L^2} \leq C(1 + \|f\|_\infty)\|\psi\|_{(H_0^2)^*}. \tag{168}$$

(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*

$$\|V_f[\psi] - V_{\bar{f}}[\psi]\|_\infty \lesssim (1 + \|f\|_\infty)\|\psi\|_{L^2}\|f - \bar{f}\|_\infty. \tag{169}$$

Proof. Part (i) is a direct consequence of the Feynman–Kac formula for $V_f[\psi]$ from [28] (see also [81, Lemma 25]). The upper bounds in (ii) likewise are proved by standard arguments for elliptic PDEs (see, e.g., [81, Lemma 26]). In order to prove the lower bound in (168), 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

$$\|S_f w\|_{(H_0^2)^*} \lesssim (1 + \|f\|_\infty)\|w\|_{L^2}, \quad w \in H_0^2.$$

Using the divergence theorem we have, for such w,

$$\begin{aligned}\|S_f w\|_{(H_0^2)^*} &= \sup_{\psi \in H_0^2 : \|\psi\|_{H_0^2} \leq 1} \left|\int_{\mathcal{O}} \psi S_f w\right| \\ &= \sup_{\psi \in H_0^2 : \|\psi\|_{H_0^2} \leq 1} \left|\int_{\mathcal{O}} w S_f \psi\right| \leq \|w\|_{L^2} \sup_{\psi \in H_0^2 : \|\psi\|_{H_0^2} \leq 1} \|S_f \psi\|_{L^2},\end{aligned}$$

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 (168). Finally, (169) is proved by using a Sobolev embedding as well as (166) and (167):

$$\begin{aligned}\|V_f[\psi] - V_{\bar{f}}[\psi]\|_\infty &\lesssim \big\|V_f\big[(f - \bar{f})V_{\bar{f}}[\psi]\big]\big\|_{H^2} \lesssim (1 + \|f\|_\infty)\|(f - \bar{f})V_f[\psi]\|_{L^2} \\ &\lesssim (1 + \|f\|_\infty)\|f - \bar{f}\|_\infty\|\psi\|_{L^2}.\end{aligned}$$

∎

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 (11) (see [28, Theorem 4.7]. 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

$$\begin{aligned} &DG_f[h_1] = V_f[h_1 u_f], \\ &(D^2G_f[h_1])[h_2] = V_f\big[h_1 DG_f[h_2]\big] + V_f\big[h_2 DG_f[h_1]\big], \quad h_1, h_2 \in C(\mathcal{O}). \end{aligned} \tag{170}$$

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 B.2. (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 $f, \bar{f} > 0$,*

$$\|G(\bar{f}) - G(f)\|_\infty \le C\|\bar{f} - f\|_\infty, \tag{171}$$

as well as

$$\begin{aligned} \|G(\bar{f}) - G(f) - DG_f[\bar{f} - f]\|_\infty &\le C\|\bar{f} - f\|_\infty^2, \\ \|DG_{\bar{f}} - DG_f - D^2G_f[\bar{f} - f]\|_{L(C(\mathcal{O}),C(\mathcal{O}))} &\le C\|\bar{f} - f\|_\infty^2. \end{aligned} \tag{172}$$

(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*

$$\|G(f)\|_{H^2} \le C(\|f\|_{L^2} + \|g\|_{C^2(\partial\mathcal{O})}), \tag{173}$$

$$\|G(f)\|_{H^{\alpha+2}} \le C(1 + \|f\|_{H^\alpha}^{\alpha/2+1})\|g\|_{C^{\alpha+2}(\partial\mathcal{O})}. \tag{174}$$

Proof. The estimate (171) follows from the identity $G(\bar{f}) - G(f) = V_f[(\bar{f} - f)G(\bar{f})]$, (166) and (18). Arguing similarly and using (171), 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 (172). For the second part of (172) we have, for any $h \in C(\mathcal{O})$,

$$\begin{aligned} DG_{\bar{f}}[h] - DG_f[h] &= V_{\bar{f}}[hu_{\bar{f}}] - V_f[hu_f] \\ &= V_{\bar{f}}[h(u_{\bar{f}} - u_f)] + (V_{\bar{f}} - V_f)[hu_f] \\ &= V_f\big[hDG_f[\bar{f} - f]\big] + R_1 + V_f\big[(\bar{f} - f)V_f[hu_f]\big] + R_2 \\ &= (D^2G_f[\bar{f} - f])[h] + R_1 + R_2, \end{aligned}$$

with remainder terms R_1, R_2 given by

$$\begin{aligned} R_1 &= [V_{\bar{f}} - V_f][h(u_{\bar{f}} - u_f)] + V_f\big[h(u_{\bar{f}} - u_f - DG[h])\big], \\ R_2 &= [V_{\bar{f}} - V_f](hu_f) - V_f\big[(\bar{f} - f)V_f[hu_f]\big]. \end{aligned}$$

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 B.1 as well as the first part of (172), we have

$$\begin{aligned}\|R_1\|_\infty &\lesssim \|\bar{f} - f\|_\infty \|h(u_{\bar{f}} - u_f)\|_\infty + \|h\|_\infty \|u_{f+h} - u_f - D\bar{G}[h]\|_\infty \\ &\lesssim \|\bar{f} - f\|_\infty^2 \|h\|_\infty,\end{aligned}$$

and arguing similarly,

$$\begin{aligned}\|R_2\|_\infty &= \big\|V_f\big[(\bar{f} - f)(V_{\bar{f}} - V_f)[hu_f]\big]\big\|_\infty \lesssim \|\bar{f} - f\|_\infty \|(V_{\bar{f}} - V_f)[hu_f]\|_\infty \\ &\lesssim \|\bar{f} - f\|_\infty^2 \|h\|_\infty.\end{aligned}$$

This completes the proof of (172).

To prove (173), we use the fact 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 [66, Theorem II.5.4]) such that in particular

$$\|G(f)\|_{H^2} \lesssim \|fu_f\|_{L^2} + \|g\|_{C^2(\partial\mathcal{O})} \le \|f\|_{L^2} + \|g\|_{C^2(\partial\mathcal{O})}.$$

where we have also used (18). Finally, (174) is proved in [81, Lemma 27]. ■

B.1.1. Properties of the map Φ^*. We summarise some properties of ‘regular’ link functions from Definition 2.1. We recall the notation Φ^* for the associated composition operator from (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

$$D\Phi_F^*[H] = H\Phi' \circ F, \quad (D^2\Phi_F^*[H])[J] = HJ\Phi'' \circ F, \quad H, J \in C(\mathcal{O}). \tag{175}$$

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 $(F + H)(x)$,

$$\begin{aligned}|(\Phi^*(F + H) - \Phi^*(F) - D\Phi_F^*[H])(x)| &= |H^2(x)\Phi''(\zeta_x)/2| \\ &\le \|H\|_\infty^2 \sup_{t\in\mathbb{R}} |\Phi''(t)|, \\ |(D\Phi_{F+H}^* - D\Phi_F^* - D^2\Phi_F^*[H])[J](x)| &= |J(x)H^2(x)\Phi'''(\bar{\zeta}_x)/2| \\ &\le \|J\|_\infty \|H\|_\infty^2 \sup_{t\in\mathbb{R}} |\Phi'''(t)|,\end{aligned}$$

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 $F \in H^\alpha(\mathcal{O})$,

$$\|\Phi \circ F\|_{H^\alpha} \le C(1 + \|\Phi \circ F\|_{H^\alpha}^\alpha) \tag{176}$$

(see [81, Lemma 29]). Finally, note that by the definition of Φ, there exists $C' > 0$ such that for any $\bar{F}, F \in C(\mathcal{O})$,

$$\|\Phi \circ \bar{F} - \Phi \circ F\|_\infty \le C\|\bar{F} - F\|_\infty, \quad \|\Phi \circ \bar{F} - \Phi \circ F\|_{L^2} \le C\|\bar{F} - F\|_{L^2}. \tag{177}$$

B.1.2. 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 $D^2 T_\theta \in L(U, L(U, V))$ 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 B.3. *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 $D(B \circ A)_\theta = DB_{A(\theta)} \circ DA_\theta$ and*

$$(D^2(B \circ A)_\theta[H_1])[H_2] = \big(D^2 B_{A(\theta)}\big[DA_\theta[H_1]\big]\big)\big[DA_\theta[H_2]\big] + DB_{A(\theta)}\big[(D^2 A_\theta[H_1])[H_2]\big]. \tag{178}$$

B.2. Proof of Proposition 3.6

We first record the following basic lemma without proof.

Lemma B.4. *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*

$$\nabla n(\theta) = \frac{M\theta}{|\theta|}, \quad \nabla^2 n(\theta) = \frac{M}{|\theta|} - \frac{M\theta(M\theta)^T}{|\theta|^3}, \tag{179}$$

as well as the norm estimates

$$\|\nabla n(\theta)\|_{\mathbb{R}^D} \leq \sqrt{\lambda_{\max}(M)}, \tag{180}$$

$$\|\nabla^2 n(\theta)\|_{\mathrm{op}} \leq 2\lambda_{\max}(M)/|\theta|_1. \tag{181}$$

Using Lemma B.4, we prove the following bounds on the cut-off function α_η.

Lemma B.5. *If $|\cdot|_1$ is an ellipsoidal norm with associated matrix M, $|\theta|_1^2 = \theta^T M \theta$, then the function α_η from (53) satisfies, for all $\theta \in \mathbb{R}^D$,*

$$\|\nabla \alpha_\eta(\theta)\|_{\mathbb{R}^D} \leq \frac{\|\alpha\|_{C^1} \sqrt{\lambda_{\max}(M)}}{\eta}, \quad \|\nabla^2 \alpha_\eta(\theta)\|_{\mathrm{op}} \leq \frac{4\|\alpha\|_{C^2} \lambda_{\max}(M)}{\eta^2}.$$

Proof. We may assume without loss of generality that $\theta_{\mathrm{init}} = 0$ and we write $n(\theta) = |\theta|_1$. The gradient bound is obtained by the chain rule and (180):

$$\|\nabla \alpha_\eta(\theta)\|_{\mathbb{R}^D} = \|\eta^{-1} \alpha'(|\theta|_1/\eta) \nabla n(\theta)\|_{\mathbb{R}^D} \leq \eta^{-1} \|\alpha\|_{C_1} \sqrt{\lambda_{\max}(M)}.$$

For the Hessian, we similarly employ the chain rule, (180), (181) as well as the fact that $\alpha'(t) = 0$ when $t \in (0, 3/4)$:

$$\begin{aligned}\|\nabla^2 \alpha_\eta(\theta)\|_{\mathrm{op}} &\leq \eta^{-2} \|\alpha''(|\theta|_1/\eta) \nabla n(\theta) \nabla n(\theta)^T\|_{\mathrm{op}} + \eta^{-1} \|\alpha'(|\theta|_1/\eta) \nabla^2 n(\theta)\|_{\mathrm{op}} \\ &\leq \eta^{-2} \|\alpha\|_{C^2} \|\nabla n(\theta)\|_{\mathbb{R}^D}^2 + \eta^{-1} \|\alpha\|_{C_1} \mathbb{1}_{\{|\theta| \geq 3\eta/4\}} \cdot \frac{2\lambda_{\max}(M)}{|\theta|_1} \\ &\leq 4\eta^{-2} \|\alpha\|_{C^2} \lambda_{\max}(M).\end{aligned}$$

■

We now turn to the proof of Proposition 3.6. Throughout, we work on the event $\mathcal{E}_{\text{conv}} \cap \mathcal{E}_{\text{init}}$ defined by (49), (50); moreover, we assume without loss of generality that $\theta_{\text{init}} = 0$.

Proof of Proposition 3.6. We divide the proof into five steps.

1. Local lower bound for $\alpha_\eta \ell_N$. For the set

$$V := \{\theta : |\theta|_1 \le 3\eta/4\},$$

by definition of $\mathcal{E}_{\text{init}}$, we have $V \subseteq \mathcal{B}$. Thus using the definitions of $\mathcal{E}_{\text{conv}}$ and of α_η, we obtain

$$\inf_{\theta \in V} \lambda_{\min}(-\nabla^2[\alpha_\eta \ell_N](\theta)) \ge N c_{\min}/2. \tag{182}$$

2. Upper bound for $\alpha_\eta \ell_N$. By the chain rule, Lemma B.5, the definition of $\mathcal{E}_{\text{conv}}$ and using $\|\alpha\|_{C^2} \ge 1$, we deduce that for any $\theta \in \mathbb{R}^D$ and some $c = c(\alpha)$,

$$\begin{aligned}
&\|\nabla^2[\alpha_\eta \ell_N](\theta)\|_{\text{op}} \\
&\le |\ell_N(\theta)|\, \|\nabla^2 \alpha_\eta(\theta)\|_{\text{op}} + 2\|\nabla \alpha_\eta(\theta)\|_{\mathbb{R}^D} \|\nabla \ell_N(\theta)\|_{\mathbb{R}^D} + |\alpha_\eta(\theta)|\, \|\nabla^2 \ell_N(\theta)\|_{\text{op}} \\
&\le 2 \sup_{\theta \in \mathcal{B}} \big([|\alpha_\eta(\theta)| + \|\nabla \alpha_\eta(\theta)\|_{\mathbb{R}^D} + \|\nabla^2 \alpha_\eta(\theta)\|_{\text{op}}][|\ell_N(\theta)| + \|\nabla \ell_N(\theta)\|_{\mathbb{R}^D} \\
&\qquad\qquad\qquad + \|\nabla^2 \ell_N(\theta)\|_{\text{op}}]\big) \\
&\le c(1 + \lambda_{\max}(M)/\eta^2) \cdot N(c_{\max} + 1).
\end{aligned} \tag{183}$$

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 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 B.4, for any $\theta \in \mathbb{R}^D$, writing $v = \nabla n(\theta)$, we have

$$\begin{aligned}
\nabla^2 g_\eta(\theta) &= \gamma_\eta''(|\theta|_1) \nabla n(\theta) \nabla n(\theta)^T + \gamma_\eta'(|\theta|_1) \nabla^2 n(\theta) \\
&= \gamma_\eta''(|\theta|_1) v v^T + \frac{\gamma_\eta'(|\theta|_1)}{|\theta|_1}(M - v v^T) \\
&= \left(\gamma_\eta''(|\theta|_1) - \frac{\gamma_\eta'(|\theta|_1)}{|\theta|_1}\right) v v^T + \frac{\gamma_\eta'(|\theta|_1)}{|\theta|_1} M \\
&=: A(|\theta|_1) v v^T + B(|\theta|_1) M.
\end{aligned} \tag{184}$$

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, for any $t \ge 3\eta/4$ we have

$$\gamma_\eta'(t) = \int_{[-\eta/8, \eta/8]} \varphi_{\eta/8}(y) \cdot 2(t - y - 5\eta/8)\, dy = 2(t - 5\eta/8). \tag{185}$$

Thus the function $B(t) = \gamma_\eta'(t)/t$ strictly increases on $(3\eta/4, \infty)$, and for any $t \ge 3\eta/4$, we obtain

$$B(t) \ge B(3\eta/4) = \frac{\gamma_\eta'(3\eta/4)}{3\eta/4} = 2\frac{3\eta/4 - 5\eta/8}{3\eta/4} = \frac{1}{3}. \tag{186}$$

For the term $A(\cdot)$, we note that for any $t \geq 3\eta/4$, using $\gamma_\eta''(t) = 2$ as well as (185), we have

$$A(t) = 2 - \frac{2(t - 5\eta/8)}{t} \geq 0. \tag{187}$$

Combining the displays (184), (186), (187), we have proved the lower bound

$$\inf_{\theta \in V^c} \lambda_{\min}(\nabla^2 g_\eta(\theta)) \geq \lambda_{\min}(M)/3. \tag{188}$$

4. Global upper bound for $\nabla^2 g_\eta$. We note that the functions $A(\cdot)$, $B(\cdot)$ from (184) satisfy

$$\sup_{t\in(0,\infty)} |A(t)| \leq \sup_{t\in(0,\infty)} \big(|\gamma_\eta'(t)/t| + |\gamma_\eta''(t)|\big) \leq 4, \qquad \sup_{t\in(0,\infty)} |B(t)| \leq \sup_{t\in(0,\infty)} |\gamma_\eta'(t)/t| \leq 2.$$

Hence, by (184) and Lemma B.4, we obtain

$$\|\nabla^2 g_\eta(\theta)\|_{\mathrm{op}} \leq 4\|vv^T\|_{\mathrm{op}} + 2\|M\|_{\mathrm{op}} \leq 6\lambda_{\max}(M), \quad \theta \in \mathbb{R}^D. \tag{189}$$

5. Combining the bounds. Combining the estimates (182), (183) and (188), we obtain that

$$\begin{aligned} &\inf_{\theta\in V} \lambda_{\min}(-\nabla^2 \tilde{\ell}_N(\theta)) \geq Nc_{\min}/2, \\ &\inf_{\theta\in V^c} \lambda_{\min}(-\nabla^2 \tilde{\ell}_N(\theta)) \geq K\lambda_{\min}(M)/3 - c(1 + \lambda_{\max}(M)/\eta^2)N(c_{\max} + 1). \end{aligned} \tag{190}$$

In particular, there exists $C \geq 3$ such that for any K satisfying (55), we have

$$\inf_{\theta\in\mathbb{R}^D} \lambda_{\min}(-\nabla^2 \tilde{\ell}_N(\theta)) \geq \min\{Nc_{\min}/2, K\lambda_{\min}(M)/6\} = Nc_{\min}/2,$$

which completes the proof of (56). To prove (57), we use (183), (189) and (55) to find that for all $\theta \neq \bar{\theta} \in \mathbb{R}^D$,

$$\begin{aligned} \frac{\|\nabla\tilde{\ell}_N(\theta) - \nabla\tilde{\ell}_N(\bar{\theta})\|_{\mathbb{R}^D}}{\|\theta - \bar{\theta}\|_{\mathbb{R}^D}} &\leq \sup_{\theta\in\mathbb{R}^D} \|\nabla^2 \tilde{\ell}_N(\theta)\|_{\mathrm{op}} \\ &\leq c\|\alpha\|_{C^2}(1 + \lambda_{\max}(M)/\eta^2)N(c_{\max} + 1) + 6K\lambda_{\max}(M) \\ &\leq 7K\lambda_{\max}(M). \end{aligned}$$

∎

B.3. Initialisation

In this section we prove the existence of a polynomial time 'initialiser' $\theta_{\mathrm{init}} = \theta_{\mathrm{init}}(Z^{(N)}) \in \mathbb{R}^D$ (that lies in the region $\mathcal{B}_{1/\log N}$ from (99) of strong log-concavity of the posterior measure with high $P_{\theta_0}^N$-probability, when $\alpha > 6$) in the Schrödinger model.

Theorem B.6. *Suppose $\theta_0 \in h^\alpha(\mathcal{O})$ for some $\alpha > 2 + d/2$ with $d \leq 3$. Then there exists a measurable function $\theta_{\mathrm{init}} \in \mathbb{R}^D$ of the data $Z^{(N)}$ from (20) and large enough $M' > 0$ such that for all $N, D \in \mathbb{N}$ and some $\bar{c} > 0$,*

$$P_{\theta_0}^N(\|\theta_{\mathrm{init}} - \theta_{0,D}\|_{\mathbb{R}^D} > M'N^{-(\alpha-2)/(2\alpha+d)}) \lesssim e^{-\bar{c}N^{d/(2\alpha+d)}}.$$

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

$$\{\phi_{l,r} : 1 \le r \le N_l,\, l \in \mathbb{N}\}, \quad N_l \lesssim 2^{ld},$$

of $L^2(\mathcal{O})$ constructed in [97, Theorem 5.51]. Then for data arising from (19), choosing

$$2^J \simeq N^{1/(2\alpha+d)} = (N\delta_N^2)^{1/d}, \quad \delta_N = N^{-\alpha/(2\alpha+d)}, \quad n_J \equiv \sum_{l \le J} N_l \lesssim 2^{Jd},$$

and for multiscale vectors $(\lambda_{l,r}) \in \mathbb{R}^{n_J}$, define

$$\hat{\lambda} = \arg\min_{\lambda \in \mathbb{R}^{n_J}} \left[\frac{1}{N} \sum_{i=1}^{N} \Big(Y_i - \sum_{l \le J, r} \lambda_{l,r} \phi_{l,r}(X_i) \Big)^2 + \delta_N^2 \|\lambda\|_{h^\alpha}^2 \right], \quad \|\lambda\|_{h^\alpha}^2 = \sum_{l,r} 2^{2l\alpha} \lambda_{l,r}^2, \tag{191}$$

noting that the arg min set is a singleton due to strong convexity. 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 $u_{f_0} = \mathscr{G}(\theta_0) \in H^{\alpha+2}$ furnished by [97, Theorem 5.51] and the elliptic regularity estimate (174). In particular, by the Sobolev embedding $h^{\alpha+2} \subset b^\alpha_{\infty\infty}$ $(d < 4)$ and again [97, Theorem 5.51] we can prove

$$\|u_{f_0} - u_{f_0,J}\|_{L^2} \lesssim \|u_{f_0} - u_{f_0,J}\|_\infty \lesssim 2^{-J\alpha} \lesssim \delta_N. \tag{192}$$

We now apply a standard result from M estimation [99, 100], 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 B.7. *For $\alpha > d/2$, all N and some constant $c > 0$, we have*

$$P_{\theta_0}^N \big(\|\hat{u} - u_{f_0}\|_{(N)}^2 + \delta_N^2 \|\hat{\lambda}\|_{h^\alpha}^2 > \|u_{f_0} - u_{f_0,J}\|_{(N)}^2 + \delta_N^2 \|\lambda_{0,l,r}\|_{h^\alpha}^2 | (X_i)_{i=1}^N \big) \le e^{-cN\delta_N^2}. \tag{193}$$

Proof. We apply [99, Theorem 2.1]. 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 an H^α-Sobolev space, which by [44, (4.184)] is of order $H(\tau) \lesssim (m/\tau)^{d/\alpha}$ for every $m > 0$. Then arguing as in [99, Section 3.1.1] (the only notational difference being that here $d > 1$), the result follows. ∎

This implies in particular, using $\|u\|_{(N)} \le \|u\|_\infty$, (192), $\lambda_{0,l,r} \in h^{\alpha+2}$ and [97, Theorem 5.51], that for some $C, C' > 0$,

$$P^N_{\theta_0}(\|\hat{u}\|^2_{H^\alpha} > C) \le P^N_{\theta_0}(\|\hat{\lambda}\|^2_{h^\alpha} > C') \le \exp\{-cN\delta_N^2\}. \tag{194}$$

as well as

$$P^N_{\theta_0}(\|\hat{u} - u_{f_0,J}\|^2_{(N)} > C\delta_N^2) \le \exp\{-cN\delta_N^2\}. \tag{195}$$

In Step IV below we establish the following restricted isometry type bound:

$$P^N_{\theta_0}\left(\left|\frac{\|\hat{u} - u_{f_0,J}\|^2_{(N)}}{\|\hat{u} - u_{f_0,J}\|^2_{L^2}} - 1\right| \le \frac{1}{2}\right) \ge 1 - c''e^{-c'N\delta_N^2} \tag{196}$$

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 (192) and (195), for M large enough,

$$\begin{aligned}
P^N_{\theta_0}(\|\hat{u} - u_{f_0}\|^2_{L^2} > M\delta_N^2) &\le P^N_{\theta_0}(\|\hat{u} - u_{f_0,J}\|^2_{L^2} > (M/2)\delta_N^2) \\
&\le P^N_{\theta_0}\left(\frac{\|\hat{u} - u_{f_0,J}\|^2_{L^2}}{\|\hat{u} - u_{f_0,J}\|^2_{(N)}}\|\hat{u} - u_{f_0,J}\|^2_{(N)} > (M/2)\delta_N^2, \mathcal{A}_N\right) + c''e^{-c'N\delta_N^2} \\
&\le P^N_{\theta_0}(\|\hat{u} - u_{f_0,J}\|^2_{(N)} > (M/4)\delta_N^2) + c''e^{-c'N\delta_N^2} \lesssim e^{-cN\delta_N^2} + e^{-c'N\delta_N^2}.
\end{aligned}$$

Overall what precedes implies that we can find M large enough such that for some constants $\bar{c}, \bar{c}' > 0$,

$$P^N_{\theta_0}(\|\hat{u} - u_{f_0}\|^2_{L^2} \le M\delta_N^2 \text{ and } \|\hat{u}\|^2_{H^\alpha} \le M) \ge 1 - \bar{c}'e^{-\bar{c}N\delta_N^2}. \tag{197}$$

Step II. By definition of the $\|\cdot\|_{h^\alpha}$-norm, the objective function minimised in (191) 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 (191) satisfies

$$\frac{\partial Q_N}{\partial \lambda_{l',r'}}(\lambda) = -\frac{2}{N}\sum_{i=1}^N\Big[Y_i - \sum_{l\le J,r}\lambda_{l,r}\phi_{l,r}(X_i)\Big]\phi_{l',r'}(X_i), \quad l' \le J,\ 1 \le r' \le N_{l'},$$

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 [97, Definition 5.25]. Using (18), (96) and a standard tail inequality for χ^2-random variables [44, Theorem 3.1.9], one shows further that for some $\bar{C} > 0$ and on events of sufficiently high $P^N_{\theta_0}$-probability,

$$Q_N(0) = \frac{1}{N}\sum_{i=1}^N(\varepsilon_i^2 + 2\varepsilon_i u_{f_0}(X_i) + u^2_{f_0}(X_i)) \le \bar{C}.$$

By Proposition A.2 and using the standard sequence norm inequality

$$\|v\|_{h^\beta} \le 2^{J\beta}\|v\|_{\ell^2} \lesssim N^{\beta/(2\alpha+d)}\|v\|_{\ell^2}, \quad v \in \mathbb{R}^{n_J},\ \beta \ge 0,$$

we deduce that on preceding events and for any fixed $p > 0$ there exists $b_0 > 0$ such that the output $\lambda_{\mathrm{init}} \in \mathbb{R}^{n_J}$ from $O(N^{b_0})$ iterations of gradient descent satisfies $\|\lambda_{\mathrm{init}} - \hat{\lambda}\|_{h^\alpha} \le N^{-p}$. In particular, we can choose p such that, denoting

$$u_{\mathrm{init}} := \sum_{l\le J,r} \lambda_{\mathrm{init},l,r}\phi_{l,r},$$

we have $\|\hat{u} - u_{\mathrm{init}}\|_{H^\alpha} \lesssim \|\hat{\lambda} - \lambda_{\mathrm{init}}\|_{h^\alpha} = o(\delta_N)$; hence by virtue of (197), we may restrict the rest of the proof to an event of sufficiently high probability where u_{init} satisfies

$$\|u_{\mathrm{init}} - u_{f_0}\|_{L^2}^2 + \delta_N^2\|u_{\mathrm{init}}\|_{H^\alpha}^2 \le (2M+1)\delta_N^2. \tag{198}$$

Step III. From the interpolation inequality for Sobolev norms from Section 1.3 and (198) we now obtain, with sufficiently high $P_{\theta_0}^N$-probability,

$$\|u_{\mathrm{init}} - u_{f_0}\|_{H^2} \le \bar{M}N^{-(\alpha-2)/(2\alpha+d)} \tag{199}$$

and the Sobolev imbedding ($d < 4$) further implies $\|u_{\mathrm{init}} - u_{f_0}\|_\infty \to 0$ as $N \to \infty$ so that we deduce from (118) that $\hat{u} \ge u_{f_0}/2 \ge c > 0$ with sufficiently high $P_{\theta_0}^N$-probability. So on these events we can define a new estimator

$$f_{\mathrm{init}} = \frac{\Delta u_{\mathrm{init}}}{2u_{\mathrm{init}}}, \quad \text{noting that} \quad f_0 = \frac{\Delta u_{f_0}}{2u_{f_0}}. \tag{200}$$

For $F_{\mathrm{init}} = \Phi^{-1} \circ f_{\mathrm{init}}$, using also the regularity of the inverse link function (177), we then see

$$\|F_{\mathrm{init}} - F_{\theta_0}\|_{L^2} \lesssim \|f_{\mathrm{init}} - f_0\|_{L^2} \lesssim \|u_{\mathrm{init}} - u_{f_0}\|_{H^2},$$

and hence for some $M' > 0$,

$$P_{\theta_0}^N\big(\|F_{\mathrm{init}} - F_{\theta_0}\|_{L^2} \le M'N^{-(\alpha-2)/(2\alpha+d)}\big) \ge 1 - \bar{c}'e^{-\bar{c}N\delta_N^2}.$$

We finally define θ_{init} as

$$\theta_{\mathrm{init}} = (\langle F_{\mathrm{init}}, e_k\rangle_{L^2} : k \le D) \in \mathbb{R}^D, \quad D \in \mathbb{N},$$

the vector of the first D 'Fourier coefficients' of F_{init}. Then we deduce from Parseval's identity that $\|\theta_{\mathrm{init}} - \theta_{0,D}\|_{\mathbb{R}^D} \le \|F_{\mathrm{init}} - F_{\theta_0}\|_{L^2}$, which combined with the last probability inequality establishes convergence rate desired in Theorem B.6.

Step IV. Proof of (196). 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

$$\|\hat{u} - u_{f_0,J}\|_{(N)}^2 - \|\hat{u} - u_{f_0,J}\|_{L^2(\mathcal{O})}^2 = (\hat{\lambda} - \lambda_0)^T (\hat{\Gamma} - \Gamma)(\hat{\lambda} - \lambda_0)$$

and hence (one minus the) probability relevant in (196) can be bounded as

$$\Pr\left(\left|\frac{(\hat{\lambda} - \lambda_0)^T (\hat{\Gamma} - \Gamma)(\hat{\lambda} - \lambda_0)}{(\hat{\lambda} - \lambda_0)^T \Gamma (\hat{\lambda} - \lambda_0)}\right| > 1/2\right) \le \Pr\Big(\sup_{v \in \mathbb{R}^{n_J}: v^T \Gamma v \le 1} |v^T (\hat{\Gamma} - \Gamma) v| > 1/2\Big).$$

We also note that by the frame property of the $\{\phi_{l,r}\}$, specifically from [97, (5.252)] with $s = 0$, $p = q = 2$, for any $u_v = \sum_{l \le J, r} v_{l,r} \phi_{l,r}$ we have the norm equivalence

$$\|v\|_{\mathbb{R}^{n_J}}^2 \simeq \|u_v\|_{L^2}^2 = \sum_{l,l' \le J, r, r'} v_{l,r} v_{l',r'} \Gamma_{(l,r),(l',r')} = v^T \Gamma v =: \|v\|_\Gamma^2, \tag{201}$$

with the constants implied by $\simeq$ independent of J. Next for any $\kappa > 0$ let

$$\{v_m : m = 1, \dots, M_{J,\kappa}\}, \quad M_{J,\kappa} \lesssim (3/\kappa)^{n_J},$$

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 [44, Prop. 4.3.34] and using (201)). Then by the Cauchy–Schwarz inequality,

$$\begin{aligned}
|v^T (\hat{\Gamma} - \Gamma) v| &= |(v - v_m + v_m)^T (\hat{\Gamma} - \Gamma)(v - v_m + v_m)| \\
&\le \|v - v_m\|_\Gamma^2 \sup_{v \in V_\Gamma} |v^T (\hat{\Gamma} - \Gamma) v| + 2\|v - v_m\|_\Gamma \|(\hat{\Gamma} - \Gamma) v\|_\Gamma + |v_m^T (\hat{\Gamma} - \Gamma) v_m| \\
&\le (\kappa^2 + 2\kappa) \sup_{v \in V_\Gamma} |v^T (\hat{\Gamma} - \Gamma) v| + |v_m^T (\hat{\Gamma} - \Gamma) v_m|,
\end{aligned}$$

so choosing κ small enough so that $\kappa^2 + 2\kappa < 1/4$ we obtain

$$\sup_{v \in V_\Gamma} |v^T (\hat{\Gamma} - \Gamma) v| \le (4/3) \max_{m=1,\dots,M_J} |v_m^T (\hat{\Gamma} - \Gamma) v_m|, \quad M_J \equiv M_{J,\kappa}. \tag{202}$$

In particular, using also $M_J \lesssim e^{c_0 2^{Jd}} \le e^{c_1 N \delta_N^2}$, the last probability is thus bounded by

$$\Pr\Big(\max_{m=1,\dots,M_J} |v_m^T (\hat{\Gamma} - \Gamma) v_m| > 1/4\Big) \le e^{c_1 N \delta_N^2} \max_m \Pr\big(|v_m^T (\hat{\Gamma} - \Gamma) v_m| > 1/4\big). \tag{203}$$

Each of the last probabilities can be bounded by Bernstein's inequality [44, Prop. 3.1.7] applied to

$$v_m^T (\hat{\Gamma} - \Gamma) v_m = \frac{1}{N} \sum_{i=1}^N (Z_i - E Z_i),$$

with i.i.d. variables $Z_i = Z_{i,m}$ given by

$$\begin{aligned}
Z_i &= \sum_{l,l' \le J, r, r'} v_{m,l,r} v_{m,l',r'} \phi_{l,r}(X_i) \phi_{l',r'}(X_i) \\
&= \sum_{l \le J, r} v_{m,l,r} \phi_{l,r}(X_i) \sum_{l' \le J, r'} v_{m,l',r'} \phi_{l',r'}(X_i),
\end{aligned} \tag{204}$$

with vectors v_m all satisfying $\|v_m\|_\Gamma \leq 1$. For these variables we have, by the Cauchy–Schwarz inequality,

$$|Z_i| \leq \Big|\sum_{l\leq J,r} v_{m,l,r}\phi_{l,r}(\cdot)\Big|^2 \leq \|v_m\|^2_{\mathbb{R}^{n_J}} \sum_{l\leq J,r} (\phi_{l,r}(\cdot))^2 \leq c2^{Jd} \equiv U$$

where the constant c depends only on the wavelet frame (cf. (201) and also [97, Definition 5.25]). Similarly, using the previous estimate, we can bound

$$E[Z_i^2] = E\Big[\Big(\sum_{l\leq J,r} v_{m,l,r}\phi_{l,r}(X_i)\Big)^4\Big] \leq U\int_{\mathcal{O}}\Big[\sum_{l\leq J,r} v_{m,l,r}\phi_{l,r}(x)\Big]^2 dx = U\|v_m\|^2_\Gamma \leq U.$$

Now Proposition 3.1.7 in [44] implies, for some constant $c_0 > 0$,

$$\Pr\big(N|v_m(\hat\Gamma - \Gamma)v_m| > N/4\big) \leq 2\exp\Big\{-\frac{N^2/16}{2NU + (2/12)NU}\Big\} \leq 2e^{-c_0/\delta_N^2}$$

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 (203) is bounded by a constant multiple of $e^{-c' N\delta_N^2}$ for some $c' > 0$, completing the proof. ∎

Acknowledgements. We gratefully acknowledge support by the *European Research Council*, ERC grant agreement 647812 (UQMSI). This work is part of the second author's PhD thesis written at the University of Cambridge.

References

[1] Abraham, K., Nickl, R.: On statistical Calderón problems. Math. Stat. Learn. **2**, 165–216 (2019) Zbl 1445.35144 MR 4130599

[2] Agapiou, S., Wang, S.: Laplace priors and spatial inhomogeneity in Bayesian inverse problems. arXiv:2112.05679v2 (2022)

[3] Arridge, S., Maass, P., Öktem, O., Schönlieb, C.-B.: Solving inverse problems using data-driven models. Acta Numer. **28**, 1–174 (2019) Zbl 1429.65116 MR 3963505

[4] Bakry, D., Émery, M.: Diffusions hypercontractives. In: Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math. 1123, Springer, Berlin, 177–206 (1985) Zbl 0561.60080 MR 889476

[5] Bakry, D., Gentil, I., Ledoux, M.: Analysis and Geometry of Markov Diffusion Operators. Grundlehren Math. Wiss. 348, Springer, Cham (2014) Zbl 1376.60002 MR 3155209

[6] Bal, G., Ren, K.: Multi-source quantitative photoacoustic tomography in a diffusive regime. Inverse Problems **27**, art. 075003, 20 pp. (2011) Zbl 1225.92024 MR 2806367

[7] Bal, G., Uhlmann, G.: Inverse diffusion theory of photoacoustics. Inverse Problems **26**, art. 085010, 20 pp. (2010) Zbl 1197.35311 MR 2658827

[8] Bardsley, J. M., Cui, T., Marzouk, Y. M., Wang, Z.: Scalable optimization-based sampling on function space. SIAM J. Sci. Comput. **42**, A1317–A1347 (2020) Zbl 1471.62023 MR 4091174

[9] Belloni, A., Chernozhukov, V.: On the computational complexity of MCMC-based estimators in large samples. Ann. Statist. **37**, 2011–2055 (2009) Zbl 1175.65015 MR 2533478

[10] Ben Arous, G., Gheissari, R., Jagannath, A.: Algorithmic thresholds for tensor PCA. Ann. Probab. **48**, 2052–2087 (2020) Zbl 1444.62080 MR 4124533

[11] Ben Arous, G., Wein, A., Zadik, I. Free energy wells and overlap gap property in sparse PCA. In: Conference on Learning Theory 2020, 63 pp.

[12] Bengtsson, T., Bickel, P., Li, B.: 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, Inst. Math. Statist. Collect. 2, Inst. Math. Statist., Beachwood, OH, 316–334 (2008) Zbl 1166.93376 MR 2459957

[13] Beskos, A., Girolami, M., Lan, S., Farrell, P. E., Stuart, A. M.: Geometric MCMC for infinite-dimensional inverse problems. J. Comput. Phys. **335**, 327–351 (2017) Zbl 1375.35627 MR 3612501

[14] Bickel, P., Li, B., Bengtsson, T.: 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, Inst. Math. Statist. Collect. 3, Inst. Math. Statist., Beachwood, OH, 318–329 (2008) MR 2459233

[15] Bohr, J.: Stability of the non-abelian X-ray transform in dimension ≥ 3. J. Geom. Anal. **31**, 11226–11269 (2021) Zbl 1486.53087 MR 4310169

[16] Bohr, J., Nickl, R.: On log-concave approximations of high-dimensional posterior measures and stability properties in non-linear inverse problems. arXiv:2105.07835 (2021)

[17] Borell, C.: The Brunn–Minkowski inequality in Gauss space. Invent. Math. **30**, 207–216 (1975) Zbl 0292.60004 MR 399402

[18] Borggaard, J., Glatt-Holtz, N., Krometis, J.: On Bayesian consistency for flows observed through a passive scalar. Ann. Appl. Probab. **30**, 1762–1783 (2020) Zbl 1461.62215 MR 4133382

[19] Bou-Rabee, N., Eberle, A.: Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions. Stoch. Partial Differ. Equ. Anal. Comput. **9**, 207–242 (2021) Zbl 1470.60202 MR 4218791

[20] Bou-Rabee, N., Eberle, A., Zimmer, R.: Coupling and convergence for Hamiltonian Monte Carlo. Ann. Appl. Probab. **30**, 1209–1250 (2020) Zbl 07325608 MR 4133372

[21] Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge Univ. Press, Cambridge (2004) Zbl 1058.90049 MR 2061575

[22] Briol, F.-X., Oates, C. J., Girolami, M., Osborne, M. A., Sejdinovic, D.: Rejoinder: Probabilistic integration: a role in statistical computation? Statist. Sci. **34**, 38–42 (2019) Zbl 1420.62136 MR 3938962

[23] Cappé, O., Moulines, E., Rydén, T.: Inference in Hidden Markov Models. Springer Ser. Statist., Springer, New York (2005) Zbl 1080.62065 MR 2159833

[24] Castillo, I., Nickl, R.: Nonparametric Bernstein–von Mises theorems in Gaussian white noise. Ann. Statist. **41**, 1999–2028 (2013) Zbl 1285.62052 MR 3127856

[25] Castillo, I., Nickl, R.: On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. Ann. Statist. **42**, 1941–1969 (2014) Zbl 1305.62190 MR 3262473

[26] Castillo, I., Rousseau, J.: A Bernstein–von Mises theorem for smooth functionals in semiparametric models. Ann. Statist. **43**, 2353–2383 (2015) Zbl 1327.62302 MR 3405597

[27] Chewi, S., Erdogdu, M. A.. Li, M. B., Shen, R., Zhang, M.: Analysis of Langevin Monte Carlo from Poincaré to log-Sobolev. arXiv:2112.12662 (2021)

[28] Chung, K. L., Zhao, Z. X.: From Brownian Motion to Schrödinger's Equation. Grundlehren Math. Wiss. 312, Springer, Berlin (1995) Zbl 0819.60068 MR 1329992

[29] Cotter, S. L., Dashti, M., Robinson, J. C., Stuart, A. M.: Bayesian inverse problems for functions and applications to fluid mechanics. Inverse Problems **25**, art. 115008, 43 pp. (2009) Zbl 1228.35269 MR 2558668

[30] Cotter, S. L., Roberts, G. O., Stuart, A. M., White, D.: MCMC methods for functions: modifying old algorithms to make them faster. Statist. Sci. **28**, 424–446 (2013) Zbl 1331.62132 MR 3135540

[31] Cui, T., Law, K. J. H., Marzouk, Y. M.: Dimension-independent likelihood-informed MCMC. J. Comput. Phys. **304**, 109–137 (2016) Zbl 1349.65009 MR 3422405

[32] Dalalyan, A. S.: Theoretical guarantees for approximate sampling from smooth and log-concave densities. J. R. Statist. Soc. Ser. B. Statist. Methodol. **79**, 651–676 (2017) Zbl 1411.62030 MR 3641401

[33] Dashti, M., Stuart, A. M.: Uncertainty quantification and weak approximation of an elliptic inverse problem. SIAM J. Numer. Anal. **49**, 2524–2542 (2011) Zbl 1234.35309 MR 2873245

[34] Dashti, M., Stuart, A. M.: The Bayesian approach to inverse problems. In: Handbook of Uncertainty Quantification, Vol. 1, Springer, Cham, 311–428 (2017) MR 3839555

[35] Dirksen, S.: Tail bounds via generic chaining. Electron. J. Probab. **20**, no. 53, 29 pp. (2015) Zbl 1327.60048 MR 3354613

[36] Durmus, A., Moulines, É.: Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. Ann. Appl. Probab. **27**, 1551–1587 (2017) Zbl 1377.65007 MR 3678479

[37] Durmus, A., Moulines, É.: High-dimensional Bayesian inference via the unadjusted Langevin algorithm. Bernoulli **25**, 2854–2882 (2019) Zbl 1428.62111 MR 4003567

[38] Eberle, A.: Reflection couplings and contraction rates for diffusions. Probab. Theory Related Fields **166**, 851–886 (2016) Zbl 1367.60099 MR 3568041

[39] Engl, H. W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Math. Appl. 375, Kluwer, Dordrecht (1996) Zbl 0859.65054 MR 1408680

[40] Evans, L. C.: Partial Differential Equations. 2nd ed., Grad. Stud. Math. 19, Amer. Math. Soc., Providence, RI (2010) Zbl 1194.35001 MR 2597943

[41] Ghosal, S.: Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. J. Multivariate Anal. **74**, 49–68 (2000) Zbl 1118.62309 MR 1790613

[42] Ghosal, S., van der Vaart, A.: Fundamentals of Nonparametric Bayesian Inference. Cambridge Ser. Statist. Probab. Math. 44, Cambridge Univ. Press, Cambridge (2017) Zbl 1376.62004 MR 3587782

[43] Gilbarg, D., Trudinger, N. S.: Elliptic Partial Differential Equations of Second Order. Grundlehren Math. Wiss. 224, Springer, Berlin (1977) Zbl 1042.35002 MR 0473443

[44] Giné, E., Nickl, R.: Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Ser. Statist. Probab. Math., Cambridge Univ. Press, New York (2016) Zbl 1358.62014 MR 3588285

[45] Giordano, M., Nickl, R.: Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. Inverse Problems **36**, art. 085001, 35 pp. (2020) Zbl 1445.35330 MR 4151406

[46] Glatt-Holtz, N., Mondaini, C.: Mixing rates for Hamiltonian Monte Carlo algorithms in finite and infinite dimensions. Stoch. PDE Anal. Comput. **10**, 1318–1391 (2022) Zbl 07613003 MR 4503169

[47] Graham, I. G., Kuo, F. Y., Nichols, J. A., Scheichl, R., Schwab, C., Sloan, I. H.: Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients. Numer. Math. **131**, 329–368 (2015) Zbl 1341.65003 MR 3385149

[48] Haario, H., Laine, M., Lehtinen, M., Saksman, E., Tamminen, J.: Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. J. R. Statist. Soc. Ser. B Statist. Methodol. **66**, 591–607 (2004) Zbl 1098.86503 MR 2088292

[49] Hairer, M., Mattingly, J. C., Scheutzow, M.: Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. Probab. Theory Related Fields **149**, 223–259 (2011) Zbl 1238.60082 MR 2773030

[50] Hairer, M., Stuart, A. M., Vollmer, S. J.: Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. Ann. Appl. Probab. **24**, 2455–2490 (2014) Zbl 1307.65002 MR 3262508

[51] Hanke, M., Neubauer, A., Scherzer, O.: A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. Numer. Math. **72**, 21–37 (1995) Zbl 0840.65049 MR 1359706

[52] Hinrichs, A., Novak, E., Ullrich, M., Woźniakowski, H.: The curse of dimensionality for numerical integration of smooth functions. Math. Comput. **83**, 2853–2863 (2014) Zbl 1345.65014 MR 3246812

[53] Holley, R., Stroock, D.: Logarithmic Sobolev inequalities and stochastic Ising models. J. Statist. Phys. **46**, 1159–1194 (1987) Zbl 0682.60109 MR 893137

[54] Ilmavirta, J., Monard, F.: Integral geometry on manifolds with boundary and applications. In: The Radon Transform—the First 100 Years and Beyond, Radon Ser. Comput. Appl. Math. 22, de Gruyter, Berlin, 43–113 (2019) Zbl 1454.44002 MR 4059256

[55] Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker–Planck equation. SIAM J. Math. Anal. **29**, 1–17 (1998) Zbl 0915.35120 MR 1617171

[56] Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Appl. Math. Sci. 160, Springer, New York (2005) Zbl 1068.65022 MR 2102218

[57] Kaipio, J. P., Kolehmainen, V., Somersalo, E., Vauhkonen, M.: Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. Inverse Problems **16**, 1487–1522 (2000) Zbl 1044.78513 MR 1800606

[58] Kaltenbacher, B., Neubauer, A., Scherzer, O.: Iterative Regularization Methods for Nonlinear Ill-Posed Problems. Radon Ser. Comput. Appl. Math. 6, de Gruyter, Berlin (2008) Zbl 1145.65037 MR 2459012

[59] Karimi, H., Nutini, J., Schmidt, M.: Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition. arXiv:1608.0436v4 (2020)

[60] Katchalov, A., Kurylev, Y., Lassas, M.: Inverse Boundary Spectral Problems. Chapman & Hall/CRC Monogr. Surveys in Pure Appl. Math. 123, Chapman & Hall/CRC, Boca Raton, FL (2001) Zbl 1037.35098 MR 1889089

[61] Kekkonen, H.: Consistency of Bayesian inference with Gaussian process priors for a parabolic inverse problem. Inverse Problems **38**, art. 035002, 29 pp. (2022) Zbl 1487.80018 MR 4385425

[62] Laplace, P.-S.: Théorie analytique des probabilités. Courcier, Paris (1812)

[63] Lassas, M., Saksman, E., Siltanen, S.: Discretization-invariant Bayesian inversion and Besov space priors. Inverse Probl. Imaging **3**, 87–122 (2009) Zbl 1191.62046 MR 2558305

[64] Le Cam, L.: Asymptotic Methods in Statistical Decision Theory. Springer Ser. Statist., Springer, New York (1986) Zbl 0605.62002 MR 856411

[65] Li, W. V., Linde, W.: Approximation, metric entropy and small ball estimates for Gaussian measures. Ann. Probab. **27**, 1556–1578 (1999) Zbl 0983.60026 MR 1733160

[66] Lions, J.-L., Magenes, E.: Non-homogeneous Boundary Value Problems and Applications. Vol. I. Grundlehren Math. Wiss. 181, Springer, New York (1972) Zbl 0227.35001 MR 0350177

[67] Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels. In: Les Équations aux Dérivées Partielles (Paris, 1962), Éditions du CNRS, Paris, 87–89 (1963) Zbl 0234.57007 MR 0160856

[68] Lovász, L., Simonovits, M.: Random walks in a convex body and an improved volume algorithm. Random Structures Algorithms **4**, 359–412 (1993) Zbl 0788.60087 MR 1238906

[69] Lovász, L., Vempala, S.: The geometry of logconcave functions and sampling algorithms. Random Structures Algorithms **30**, 307–358 (2007) Zbl 1122.65012 MR 2309621

[70] Ma, Y.-A., Chatterji, N. S., Cheng, X., Flammarion, N., Bartlett, P. L., Jordan, M. I.: Is there an analog of Nesterov acceleration for gradient-based MCMC? Bernoulli **27**, 1942–1992 (2021) Zbl 1475.62123 MR 4278799

[71] Ma, Y.-A., Chen, Y., Jin, C., Flammarion, N., Jordan, M. I.: Sampling can be faster than optimization. Proc. Nat. Acad. Sci. USA **116**, 20881–20885 (2019) Zbl 1433.68397 MR 4025861

[72] Majda, A. J., Harlim, J.: Filtering Complex Turbulent Systems. Cambridge Univ. Press, Cambridge (2012) Zbl 1250.93002 MR 2934167

[73] Monard, F., Nickl, R., Paternain, G. P.: Efficient nonparametric Bayesian inference for X-ray transforms. Ann. Statist. **47**, 1113–1147 (2019) Zbl 1417.62060 MR 3909962

[74] Monard, F., Nickl, R., Paternain, G. P.: Consistent inversion of noisy non-Abelian X-ray transforms. Comm. Pure Appl. Math. **74**, 1045–1099 (2021) Zbl 07363259 MR 4230066

[75] Monard, F., Nickl, R., Paternain, G. P.: Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. Ann. Statist. **49**, 3255–3298 (2021) Zbl 1486.62068 MR 4352530

[76] Mueller, J. L., Siltanen, S.: Linear and Nonlinear Inverse Problems with Practical Applications. Comput. Sci. Engrg. 10, SIAM, Philadelphia, PA (2012) Zbl 1262.65124 MR 2986262

[77] Nickl, R.: Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. J. Eur. Math. Soc. **22**, 2697–2750 (2020) Zbl 1445.62099 MR 4118619

[78] Nickl, R., Paternain, G. P.: On some information-theoretic aspects of non-linear statistical inverse problems. arXiv:2107.09488 (2021)

[79] Nickl, R., Söhl, J.: Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. Ann. Statist. **45**, 1664–1693 (2017) Zbl 1411.62087 MR 3670192

[80] Nickl, R., Söhl, J.: Bernstein–von Mises theorems for statistical inverse problems II: compound Poisson processes. Electron. J. Statist. **13**, 3513–3571 (2019) Zbl 1429.62168 MR 4013745

[81] Nickl, R., van de Geer, S., Wang, S.: Convergence rates for penalized least squares estimators in PDE constrained regression problems. SIAM/ASA J. Uncertain. Quantif. **8**, 374–413 (2020) Zbl 1436.62163 MR 4074017

[82] Novak, E., Woźniakowski, H.: Tractability of Multivariate Problems. Vol. 1: Linear Information. EMS Tracts in Math. 6, Eur. Math. Soc. EMS, Zürich (2008) Zbl 1156.65001 MR 2455266

[83] Novak, E., Woźniakowski, H.: Tractability of Multivariate Problems. Vol. II: Standard Information for Functionals. EMS Tracts in Math. 12, Eur. Math. Soc., Zürich (2010) Zbl 1241.65025 MR 2676032

[84] Paternain, G. P., Salo, M., Uhlmann, G.: The attenuated ray transform for connections and Higgs fields. Geom. Funct. Anal. **22**, 1460–1489 (2012) Zbl 1256.53021 MR 2989440

[85] Paternain, G. P., Salo, M., Uhlmann, G.: Geometric Inverse Problems, with Emphasis on two Dimensions. Cambridge Univ. Press, to appear

[86] Polyak, B. T.: Gradient methods for minimizing functionals. Zh. Vychisl. Mat i Mat. Fiz. **3**, 643–653 (1963) (in Russian) Zbl 0196.47701 MR 158568

[87] Rebeschini, P., van Handel, R.: Can local particle filters beat the curse of dimensionality? Ann. Appl. Probab. **25**, 2809–2866 (2015) Zbl 1325.60058 MR 3375889

[88] Reich, S., Cotter, C.: Probabilistic Forecasting and Bayesian Data Assimilation. Cambridge Univ. Press, New York (2015) Zbl 1314.62005 MR 3242790

[89] Roberts, G. O., Tweedie, R. L.: Exponential convergence of Langevin distributions and their discrete approximations. Bernoulli **2**, 341–363 (1996) Zbl 0870.60027 MR 1440273

[90] Schwab, C., Stuart, A. M.: Sparse deterministic approximation of Bayesian inverse problems. Inverse Problems **28**, 045003, 32 (2012) Zbl 1236.62014 MR 2903278

[91] St-Amant, S.: Stability estimate for the broken non-abelian x-ray transform in Minkowski space. Inverse Problems **38**, art. 105007, 36 pp. (2022) Zbl 07596737 MR 4482473

[92] Stuart, A. M.: Inverse problems: a Bayesian perspective. Acta Numer. **19**, 451–559 (2010) Zbl 1242.65142 MR 2652785

[93] Sukharev, A. G.: Optimal formulas of numerical integration for some classes of functions of several variables. Dokl. Akad. Nauk SSSR **246**, 282–285 (1979) (in Russian) Zbl 0433.65009 MR 533623

[94] Talagrand, M.: Upper and Lower Bounds for Stochastic Processes. Ergeb. Math. Grenzgeb. (3) 60, Springer, Heidelberg (2014) Zbl 1293.60001 MR 3184689

[95] Taylor, M. E.: Partial Differential Equations I. Basic Theory. 2nd ed., Appl. Math. Sci. 115, Springer, New York (2011) Zbl 1206.35002 MR 2744150

[96] Taylor, M. E.: Partial Differential Equations II. Qualitative Studies of Linear Equations. 2nd ed., Appl. Math. Sci. 116, Springer, New York (2011) Zbl 1206.35003 MR 2743652

[97] Triebel, H.: Function Spaces and Wavelets on Domains. EMS Tracts in Math. 7, Eur. Math. Soc., Zürich (2008) Zbl 1158.46002 MR 2455724

[98] Uhlmann, G.: Electrical impedance tomography and Calderón's problem. Inverse Problems **25**, art. 123011, 39 pp. (2009) Zbl 1181.35339 MR 3460047

[99] van de Geer, S.: Least squares estimation with complexity penalties. Math. Methods Statist. **10**, 355–374 (2001) Zbl 1005.62043 MR 1867165

[100] van de Geer, S. A.: Applications of Empirical Process Theory. Cambridge Ser. Statist. Probab. Math. 6, Cambridge Univ. Press, Cambridge (2000) Zbl 0953.62049 MR 1739079

[101] van der Vaart, A. W.: Asymptotic Statistics. Cambridge Ser. Statist. Probab. Math. 3, Cambridge Univ. Press, Cambridge (1998) Zbl 0910.62001 MR 1652247

[102] van der Vaart, A. W., van Zanten, J. H.: Rates of contraction of posterior distributions based on Gaussian process priors. Ann. Statist. **36**, 1435–1463 (2008) Zbl 1141.60018 MR 2418663

[103] Vempala, S., Wibisono, A.: Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. In: Adv. Neural Information Processing Systems **32**, 13 pp. (2019)

[104] Villani, C.: Optimal Transport. Grundlehren Math. Wiss. 338, Springer, Berlin (2009) Zbl 1156.53003 MR 2459454

[105] Yang, Y., Wainwright, M. J., Jordan, M. I.: On the computational complexity of high-dimensional Bayesian variable selection. Ann. Statist. **44**, 2497–2532 (2016) Zbl 1359.62088 MR 3576552