NONPARAMETRIC STATISTICAL INFERENCE FOR DRIFT VECTOR FIELDS OF MULTI-DIMENSIONAL DIFFUSIONS

By Richard Nickl and Kolyan Ray

University of Cambridge and King's College London

The problem of determining a periodic Lipschitz vector field $b = (b_1, \ldots, b_d)$ from an observed trajectory of the solution $(X_t : 0 \leq t \leq T)$ of the multi-dimensional stochastic differential equation

$$dX_t = b(X_t)dt + dW_t, \quad t \geq 0,$$

where W_t is a standard d-dimensional Brownian motion, is considered. Convergence rates of a penalised least squares estimator, which equals the maximum a posteriori (MAP) estimate corresponding to a high-dimensional Gaussian product prior, are derived. These results are deduced from corresponding contraction rates for the associated posterior distributions. The rates obtained are optimal up to log-factors in L^2-loss in any dimension, and also for supremum norm loss when $d \leq 4$. Further, when $d \leq 3$, nonparametric Bernstein-von Mises theorems are proved for the posterior distributions of b. From this we deduce functional central limit theorems for the implied estimators of the invariant measure μ_b. The limiting Gaussian process distributions have a covariance structure that is asymptotically optimal from an information-theoretic point of view.

CONTENTS

1

1. Introduction. For $W_t = (W_t^1, \ldots, W_t^d)$ a d-dimensional Brownian motion and $b = (b_1, \ldots, b_d)$ a Lipschitz vector field, consider the multi-dimensional Markov diffusion process $(X_t = (X_t^1, \ldots, X_t^d) : t \geq 0)$ describing the solution to the stochastic differential equation (SDE)

$$(1) \qquad dX_t = b(X_t)dt + dW_t, \qquad X_0 = x_0 \in \mathbb{R}^d, \ t \geq 0.$$

The random process $(X_t : t \geq 0)$ describes a Brownian motion whose trajectories are subject to spatially variable displacements enforced by the drift vector field b. We are interested in recovering the parameter b based on observing the process up to time T. A closely related problem is that of estimating the invariant measure μ_b of the diffusion, which in the *ergodic* case describes the probabilities

$$(2) \qquad \mu_b(A) =^{a.s.} \lim_{T \to \infty} \frac{1}{T} \int_0^T 1_A(X_t)dt$$

corresponding to the average asymptotic time spent by the process (X_t) in a given measurable subset A of the state space.

While the one-dimensional case $d = 1$ is well studied (e.g., [32], [1–3, 14, 15, 33, 42, 57, 60]), comparably little is known about the important multi-dimensional setting, particularly when b is modelled in a nonparametric or high-dimensional way. In the measurement model we consider here, Dalalyan and Reiß [16] first obtained convergence rates of multivariate nonparametric kernel-type estimators. Schmisser [51] established adaptive L^2-convergence rates of certain model selection based projection estimators and Strauch [52–54] obtained adaptive convergence rate results for b in pointwise and L^2-loss, and for μ_b in $\| \cdot \|_\infty$-loss – more discussion can be found below.

For observations $(X_t : 0 \le t \le T)$, the likelihood function is directly available from Girsanov's theorem and has a convenient 'Gaussian' form in the parameter b. This motivates the use of likelihood based inference procedures: the estimators $\hat{b}_T$ for b we study in the present paper are minimisers of a penalised likelihood (or least squares) criterion over a high-dimensional approximation space. In fact, since the penalties we use are squared Hilbert norms, $\hat{b}_T$ equals a Bayesian 'maximum a posteriori' (MAP) estimate arising from a truncated Gaussian series prior. The Bayesian interpretation of $\hat{b}_T$ is exploited in our proofs and has further appeal since it directly suggests uncertainty quantification methodology ('posterior credible sets'). In particular, posterior sampling is feasible even for 'real-world' discrete data by simulation techniques, see [7, 8, 40, 49, 55] and references therein.

Let us briefly describe our contributions: we obtain convergence rates of $\hat{b}_T$ to the 'true' vector field b_0 generating equation (1) and also frequentist contraction rates about b_0 for the corresponding posterior distributions, both in L^2- and $\|\cdot\|_\infty$-distances. For L^2-loss the rates obtained are minimax optimal (up to log-factors) over Hölder classes in any dimension, and this remains true for $\|\cdot\|_\infty$-loss whenever dimension $d \le 4$. When $d \le 3$, we further prove nonparametric Bernstein-von Mises theorems that establish asymptotic normality of the re-centred and scaled posterior distributions $\sqrt{T}(b - \hat{b}_T)|(X_t : 0 \le t \le T)$ in a (large enough) function space. From this we deduce central limit theorems for the implied plug-in estimators for the invariant density μ_b. The proofs imply that the limiting covariances obtained coincide with the semiparametric information lower bounds for these estimation problems. We exploit that the non-linear identification map $b \mapsto \mu_b$ can be shown to be 'one-smoothing' – as inference on b is approximately a nonparametric regression problem [16], this offers an analytical explanation for why the invariant density μ_b of the process can be estimated at $1/\sqrt{T}$ rate in stronger norms than is the case in i.i.d. density estimation.

The multi-dimensional case $d \ge 2$ is fundamentally more challenging than the one-dimensional one for various reasons. First, when $d = 1$ properties of diffusion local times can be used to take advantage of regularity properties of the sample paths of (X_t) as in [2, 3, 15, 42, 57, 60], whereas for $d > 1$ these local times are no longer appropriately defined. Second, Markovian concentration properties can be derived using martingale techniques combined with mapping properties of the generator of the underlying semigroup (via the Poisson equation and Itô's formula, see Lemma 1). In dimension one this involves the study of an explicitly solvable ordinary differential equation (ODE), whereas for $d \ge 2$ the theory of elliptic partial differential equations (PDEs) is required. PDE techniques are an effective alternative to the func-

tional inequalities used in [16, 53, 54], in particular the requirement that b be a gradient vector field ∇B for some $B : \mathbb{R}^d \to \mathbb{R}$, and thus of *reversibility* of (X_t) – used in the references [16, 51, 53, 54] – can be avoided this way; neither does $X_0 \sim \mu_b$ have to be started in equilibrium as in [51]. To simplify the PDE arguments in our proofs we restrict to periodic vector fields b. In our setting, periodicity ensures the required mixing properties of (X_t), replacing spectral gap assumptions in [16, 53, 54]. The techniques of the present paper extend in principle, albeit at the expense of considerable technicalities, to the non-periodic case if b is known outside of a compact subset of $\mathbb{R}^d$ and upon employing assumptions on b as in [41]. Finally, that b is not required to be a gradient field is crucial in the multi-dimensional setting for the use of Bayesian (or penalisation) methods as standard Gaussian priors for b will draw gradient vector fields with probability zero. Moreover, for $d > 1$ the potential absence of reversibility of (X_t) introduces some fundamentally new features to the inference problem at hand, since the invariant measure μ_b no longer identifies the law P_b of the process (X_t) – see after Proposition 1 below. Unlike in the one-dimensional case (e.g., [37], [1]), Bayesian inference thus cannot be based on a prior assigned directly to the invariant measure μ_b. In contrast we show how Gaussian priors for b give valid Bayesian models for the data and allow one to make optimal inference on b and μ_b.

Our proofs employ techniques from Bayesian non- and semi-parametric statistics, specifically [9, 11–13]. In this regard our results are related to recent investigations of Bayesian inverse problems [4, 17, 30, 31, 38, 43], Bernstein-von Mises theorems [34–36], [44, 45, 47], [27] and diffusion models [1, 28, 37, 42, 56, 57, 60].

2. Main results.

2.1. *Basic notation and definitions.* Let $\mathbb{T}^d$ denote the d-dimensional torus, isomorphic to $(0,1]^d$ if opposite points on the cube are identified. By $L^2(\mathbb{T}^d)$ we denote the usual L^2-spaces with respect to Lebesgue measure dx on $\mathbb{T}^d$ equipped with inner product $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{L^2}$. Let μ be a probability measure on $\mathbb{T}^d$. If its Lebesgue density, also denoted by μ, exists and is bounded and bounded away from zero, then an equivalent norm $\| \cdot \|_\mu$ on $L^2(\mathbb{T}^d)$ arises from the inner product $\langle f, g \rangle_\mu = \int f g d\mu$ for $f, g \in L^2(\mathbb{T}^d)$. The symbol $L_0^2(\mathbb{T}^d)$ denotes the subspace of functions f for which $\int_{\mathbb{T}^d} f(x)dx = 0$, and $L_\mu^2(\mathbb{T}^d)$ denotes the subspace for which $\int_{\mathbb{T}^d} f d\mu = 0$.

We define the space $C(\mathbb{T}^d) = C^0(\mathbb{T}^d)$ of continuous functions on $\mathbb{T}^d$ normed by the usual supremum norm $\| \cdot \|_\infty$. For $s > 0$, we denote by $C^s(\mathbb{T}^d)$ the usual Hölder spaces of $[s]$-times continuously differentiable functions on $\mathbb{T}^d$, where $[s]$ is the integer part of s. For $s \in \mathbb{R}$, let $H^s(\mathbb{T}^d)$ denote

the usual Sobolev space of functions from $\mathbb{T}^d$ to $\mathbb{R}$ (defined by duality when $s < 0$). They form the special case $p = q = 2$ in the scale of Besov spaces $B^s_{pq}(\mathbb{T}^d), 1 \leq p, q \leq \infty$, see Chapter 3 of [50] for definitions, where it is also shown that $C^s(\mathbb{T}^d)$ embeds continuously into $B^s_{\infty\infty}(\mathbb{T}^d), s \geq 0$. When no confusion may arise, we employ the same function space notation for vector fields $f = (f_1, \ldots, f_d)$. For instance $f \in H^s \equiv (H^s)^{\otimes d}$ will then mean that each $f_j \in H^s(\mathbb{T}^d)$ and the norm on H^s is given by $\|f\|^2_{H^s} = \sum_{j=1}^d \|f_j\|^2_{H^s}$. We shall repeatedly use multiplication inequalities for Besov-Sobolev norms,

$$(3) \quad \|fg\|_{B^s_{pq}} \leq c(s, p, q, d)\|f\|_{B^s_{pq}}\|g\|_{B^s_{\infty\infty}} \leq c'(s, p, q, d)\|f\|_{B^s_{pq}}\|g\|_{C^s}, s \geq 0.$$

Starting with a periodised Daubechies' wavelet basis of $L^2(\mathbb{T})$, we consider a tensor product wavelet basis of $L^2(\mathbb{T}^d)$ given by

$$(4) \quad \begin{aligned} \{\Phi_{l,r} : r &= 0,.., \max(0, 2^{ld} - 1), \ l = \{-1, 0\} \cup \mathbb{N}\}, \\ V_J &\equiv \operatorname{span}(\Phi_{l,r} : r, l \leq J), \end{aligned}$$

see Section 4.3 of [26], where the base Daubechies wavelets are taken 'S-regular', $S \in \mathbb{N}$. The dimension of V_J is $O(2^{Jd})$ as $J \to \infty$, and the decay of wavelet coefficients in this basis, or equivalently the scaling of approximation errors from L^2-projections P_{V_J} onto V_J, characterise the norms of the Besov spaces $B^s_{pq}(\mathbb{T}^d)$ and Sobolev spaces $H^s(\mathbb{T}^d)$ (p.370f. in [26]).

If μ is a probability measure on some metric space, then $Z \sim \mu$ means that Z is a random variable in that space drawn from the distribution μ, also called the law $\mathcal{L}(Z) = \mu$ of Z. We write $Z_T \to^d Z$, or $Z_T \to^d \mathcal{L}(Z)$ when no confusion can arise, to denote the usual notion of weak convergence of the laws $\mathcal{L}(Z_T) \to \mathcal{L}(Z)$ as $T \to \infty$, see, e.g., Chapter 11 in [18].

For a normed linear space $(X, \|\cdot\|_X)$, the topological dual space is

$$X^* = (X, \|\cdot\|_X)^* := \{L : X \to \mathbb{R} \text{ linear s.t. } |L(x)| \leq C\|x\|_X \ \forall x \in X, C > 0\},$$

which is a Banach space for the norm $\|L\|_{X^*} \equiv \sup_{x \in X, \|x\|_X \leq 1} |L(x)|$. We will sometimes use the symbols $\lesssim, \gtrsim, \simeq$ to denote one- or two-sided inequalities up to multiplicative constants that may either be universal or 'fixed' in the context where the symbols appear. We also write $(\cdot)_+ = \max(\cdot, 0)$ to denote the non-negative part of a real number, and $a \vee b, a \wedge b$ to denote maximum and minimum of real numbers a, b, respectively.

2.2. *Diffusions with periodic drift; likelihood, prior and posterior.* Consider the SDE (1) where the vector field $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ is Lipschitz continuous and one-periodic, that is $b(\cdot + m) = b(\cdot)$ for every $m \in \mathbb{Z}^d$. Then a

strong pathwise solution of this SDE exists which is a d-dimensional diffusion Markov process $X_t = (X_t^1, \ldots, X_t^d)$. We denote by $P_b = P_b^x$ the cylindrical probability measure describing the law of (X_t) in path space $C([0, \infty) \to \mathbb{R}^d)$ when $X_0 = x$; its restriction $P_b^T = P_b^{T,x}$ to the separable space $C([0, T] \to \mathbb{R}^d)$ describes the law of the process $X^T \equiv (X_t : t \in [0, T])$ until time T, see, e.g., Sections 24 and 39 in [5]. We suppress the dependence on the starting value x as our results do not depend on it.

We seek to recover the drift function $b : \mathbb{T}^d \to \mathbb{R}^d$ from an observed trajectory X^T. The periodic model (which has also been used in [42, 60] when $d = 1$) is convenient in our context as it effectively confines the diffusion process (X_t) to a bounded state space $\mathbb{T}^d$. To be precise, while our diffusion takes values in the whole of $\mathbb{R}^d$ (in particular (X_t) will not be globally recurrent), the values of the process (X_t) modulo $\mathbb{Z}^d$ contain all relevant statistical information. In particular, we have (arguing as in the proof of Lemma 6 below),

$$\frac{1}{T} \int_0^T \varphi(X_t)dt \to^{P_b} \int_{\mathbb{T}^d} \varphi d\mu_b \text{ as } T \to \infty, \quad \forall \varphi \in C(\mathbb{T}^d),$$

where μ_b is a uniquely defined probability measure on $\mathbb{T}^d$ and where we identify φ with its periodic extension to $\mathbb{R}^d$ on the left-hand side. The measure μ_b has the usual probabilistic interpretation as an invariant measure appearing in the limit of ergodic averages, but for our purposes it is more convenient to define it in terms of a partial differential equation involving the generator of the diffusion Markov process. Heuristically, if $(P_t = e^{tL} : t \geq 0)$ is the transition operator of a diffusion process with invariant measure μ and generator L, then we can differentiate the invariant identity $\int P_t[\varphi]d\mu = \int \varphi d\mu \ \forall t$ at $t = 0$, so that $\int L\varphi d\mu = 0$ for all smooth φ. If L^* is the adjoint operator for the standard L^2-inner product, then it must satisfy $\int \varphi L^* \mu = 0$ for all smooth φ, and hence necessarily $L^*\mu = 0$ (in the weak sense), which can be used to identify μ via the adjoint generator L^*.

More precisely, in our setting the generator $L : H^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ is

$$(5) \qquad L = L_b = \frac{1}{2}\Delta + b.\nabla = \frac{1}{2}\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} + \sum_{i=1}^d b_i(\cdot)\frac{\partial}{\partial x_i},$$

and from integration by parts the adjoint operator for $\langle \cdot, \cdot \rangle_{L^2}$ equals

$$(6) \qquad L^* = L_b^* = \frac{1}{2}\Delta - b.\nabla - div(b), \quad div(b) = \sum_{j=1}^d \frac{\partial b_j}{\partial x_j},$$

so that μ_b can be identified as the (weak) solution of the PDE

$$(7) \qquad L_b^* \mu_b \equiv \frac{1}{2}\Delta\mu_b - b.\nabla\mu_b - div(b)\mu_b = 0.$$

One can prove the following result (see after (67) in Section 6 below).

PROPOSITION 1. *Let* $b \in C^1(\mathbb{T}^d)$. *A unique periodic solution* μ_b *to* (7) *satisfying* $\int_{\mathbb{T}^d} d\mu_b = 1$ *exists. Moreover,* μ_b *is Lipschitz continuous and bounded away from zero on* $\mathbb{T}^d$, *with* $\|1/\mu_b\|_\infty$ *and the Lipschitz constant* $\|\mu_b\|_{Lip}$ *depending on* b *only through a bound for* $\|b\|_\infty$.

One may show (e.g., as after (63) below) that for smoother vector fields b the resulting invariant measure actually equals a classical C^2-solution of (7), but for existence of μ_b a weak solution suffices.

If b arises as a gradient vector field ∇B for some $B \in C^2(\mathbb{T}^d)$, one can check directly that $\mu_b \propto e^{2B}$ is a classical solution of (7), and we can then recover b from μ_b via $b = (1/2)\nabla\log\mu_b$. But the invariant measure μ_b does *not* identify b or the law P_b of $(X_t : t \geq 0)$ for general vector fields b (unless $d = 1$). To see this, start with $b = \nabla B$ and invariant measure $\mu_b \propto e^{2B}$. For any smooth *divergence free* vector field $\bar{v}$ and $v = \bar{v}/\mu$ (so that $div(v\mu) = 0$) one checks by integration by parts that $\int \phi L_{b+v}^* \mu_b = \int \mu_b L_{b+v}\phi = 0$ for all smooth ϕ, and as a consequence μ_b is also the invariant measure for L_{b+v}. Thus any statistical approach to recover b via first estimating μ_b is bound to fail in our general setting.

We instead propose likelihood-based inference methods. The log-likelihood function $\ell_T(b)$ of our measurement model can be obtained from Girsanov's theorem (Section 17.7 in [5]): for any periodic and Lipschitz $b : \mathbb{T}^d \to \mathbb{R}^d$,

$$(8) \qquad e^{\ell_T(b)} = \frac{dP_b^T}{dP_0^T}(X^T) = \exp\left(-\frac{1}{2}\int_0^T \|b(X_t)\|^2 dt + \int_0^T b(X_t).dX_t\right),$$

where P_0^T is the law of a d-dimensional Brownian motion $(W_t : t \in [0,T])$.

Our approach to inference on b amounts to computing a penalised maximum likelihood estimator over a high-dimensional wavelet approximation space. More precisely, set

$$(9) \qquad \hat{b}_T = \hat{b}(X^T) = \operatorname{argmin}_{b \in V_J^{\otimes d}}\left[-\ell_T(b) + \frac{1}{2}\|b\|_\mathbb{H}^2\right],$$

where $V_J^{\otimes d} = \otimes_{j=1}^d V_J$ (cf. (4)) and $\|\cdot\|_\mathbb{H}$ is a Hilbert tensor norm on $V_J^{\otimes d}$. The estimator $\hat{b}_T$ has a natural Bayesian interpretation as the maximum a posteriori (MAP) estimate arising from a mean zero Gaussian prior $\Pi = \otimes_{j=1}^d \Pi_j$

on $V_J^{\otimes d}$ with reproducing kernel Hilbert space $\mathbb{H}$. Indeed, the posterior distribution $\Pi(\cdot|X^T)$ arising from observing $X^T \sim P_b^T$ is of the form

$$(10) \qquad d\Pi(b|X^T) = \frac{e^{\ell_T(b)}d\Pi(b)}{\int e^{\ell_T(b)}d\Pi(b)} \propto e^{\ell_T(b)-\frac{1}{2}\|b\|_{\mathbb{H}}^2}, \quad b \in V_J^{\otimes d}.$$

Our proofs imply that the denominator in the last expression is finite and non-zero with probability approaching one under the law of X^T as $T \to \infty$. The map $(b,c) \mapsto \int_0^T b(X_t)c(X_t)dt + \langle b,c\rangle_{\mathbb{H}}$ induces an inverse covariance D_H^{-1} on some linear subspace $H \subset V_J^{\otimes d}$. [Since $1 \in V_J^{\otimes d}$, $dimH \neq 0$, and our proofs imply in fact that $H = V_J^{\otimes d}$ with probability approaching one as $T \to \infty$.] By characterisations of Gaussian laws (e.g., Theorem 9.5.7 in [18]) and linearity of $b \mapsto \int_0^T b(X_t).dX_t$, the distribution $\Pi(\cdot|X^T)$ is thus Gaussian on $V_J^{\otimes d}$ and the MAP estimate (9) equals the posterior mean $E^\Pi[b|X^T]$.

The Gaussian process priors $\Pi = \Pi_T$ we will use here are constructed from high-dimensional wavelet expansions for $b = (b_1, \ldots, b_d)$ of the form:

$$(11) \qquad b_j = \sum_{l \leq J} \sum_{r=0}^{2^{ld}-1} \sigma_l g_{l,r,j}\Phi_{l,r}, \quad g_{l,r,j} \sim^{iid} \mathcal{N}(0,1), \quad j = 1, \ldots, d,$$

where the $\Phi_{l,r}$ form a S-regular periodised wavelet basis of $L^2(\mathbb{T}^d)$ (cf. (4)), where $J = J_T \to \infty$ as $T \to \infty$ in a way to be chosen below, and where the weights σ_l govern the regularisation prescribed by the penalty functional. We will tacitly assume throughout that S is large enough (depending on parameters s,a,α to be specified). We choose wavelets for convenience and B-spline bases, which give rise to the same MAP estimates, could have been used as well. Recall (p.75 in [26]) that the Gaussian process (11) has reproducing kernel Hilbert space (RKHS) inner product of tensor form

$$(12) \quad \langle g_1, g_2\rangle_{\mathbb{H}} = \sum_{j=1}^d \sum_{l \leq J} \sum_{r=0}^{2^{ld}-1} \sigma_l^{-2}\langle g_{1,j}, \Phi_{l,r}\rangle_{L^2}\langle g_{2,j}, \Phi_{l,r}\rangle_{L^2}, \quad g_1, g_2 \in V_J^{\otimes d}.$$

2.3. *Contraction rates for the posterior distribution and MAP estimate.* We now give results concerning the concentration of the posterior measure $\Pi(\cdot|X^T)$ around the 'ground truth' vector field b_0 that generated X^T according to the diffusion equation (1). This implies convergence rates of the same order of magnitude for the MAP estimate $\hat{b}_T$ (see Corollary 1). We denote the 'true' invariant measure from Proposition 1 by $\mu_0 = \mu_{b_0}$.

Our first theorem gives a contraction rate in the 'natural distance' induced by the statistical experiment, following the general theory [22, 57]. Initially

this distance is a 'random Hellinger semimetric' (see Theorem 7 below). In dimension $d = 1$, the theory of diffusion local times can then be used to compare this metric to the standard $\|\cdot\|_{\mu_0}, \|\cdot\|_{L^2}$-distances [42, 57, 60], but when $d > 1$ such local time arguments are not available. We instead exploit concentration properties of the high-dimensional random matrices induced by the Hellinger semimetric on $V_J^{\otimes d}$ (Lemma 4).

THEOREM 1. *Let* $s > \max(d/2, 1), d \in \mathbb{N}$. *Suppose* $b_0 \in C^s(\mathbb{T}^d) \cap H^s(\mathbb{T}^d)$. *Consider the Gaussian prior* Π_T *from* (11) *with* $2^J \approx T^{\frac{1}{2a+d}}$ *and* $\sigma_l = 2^{-l(\alpha+d/2)}$ *for* $a > \max(d - 1, 1/2)$ *and* $0 \leq \alpha \leq a$. *Then for* $\varepsilon_T = T^{-\frac{a \wedge s}{2a+d}}(\log T)$ *and every* $M_T \to \infty$, *as* $T \to \infty$,

$$\Pi_T \left(b : \|b - b_0\|_{\mu_0} \geq M_T \varepsilon_T | X^T \right) \to^{P_{b_0}} 0.$$

In particular, if $a = s$ *then* $\varepsilon_T = T^{-\frac{s}{2s+d}}(\log T)$.

Since we wish to perform the primary regularization via the truncation level J rather than the variance scaling α we have taken $0 \leq \alpha \leq a$.

REMARK 1 (Adaptation). The previous theorem extends to adaptive priors, where J is randomised according to a hyperprior on $\mathbb{N}$ of the form $\Pi(J = j) \sim \exp\{-C2^{jd}\}$, without requiring knowledge of the smoothness s. Given the techniques underlying Theorem 1, the proof of such a result follows standard patterns (e.g., [43], [1]) and is left to the reader.

From the previous theorem, and imposing slightly stronger conditions on b_0 and Π_T, one can obtain perturbation approximations of the Laplace transform of $\Pi(\cdot|X^T)$ by the Laplace transform of a certain Gaussian distribution (see Proposition 2), which makes more precise 'semiparametric' tools available for the analysis of the posterior distribution. Following ideas in [9] (see also [10, 12, 13, 36]) we obtain contraction results in the $\|\cdot\|_\infty$-norm.

THEOREM 2. *Let* $a \wedge s > \max(3d/2 - 1, 1), d \in \mathbb{N}$. *Suppose* $b_0 \in C^s(\mathbb{T}^d) \cap H^s(\mathbb{T}^d)$. *Consider the Gaussian prior* Π_T *from* (11) *with* $2^J \approx T^{\frac{1}{2a+d}}$ *and* $\sigma_l = 2^{-l(\alpha+d/2)}$ *for* $0 \leq \alpha < a \wedge s - d/2$. *Assume further that* $a \leq s + 1$ *if* $d \leq 4$ *or* $a \leq s + d/2 - 1$ *if* $d \geq 5$. *Then for every* $\delta > 5/2$,

$$\Pi_T \left(b : \sum_{j=1}^d \|b_j - b_{0,j}\|_\infty \geq (\log T)^\delta T^{-\frac{s \wedge [a-(d/2-2)_+]}{2a+d}} \Big| X^T \right) \to^{P_{b_0}} 0 \quad as \ T \to \infty.$$

In particular, if $a = s, 0 \leq \alpha \leq s - d/2$ *and* $d \leq 4$, *then the convergence rate is* $(\log T)^\delta T^{-\frac{s}{2s+d}}$.

By Gaussianity of the posterior distribution, the previous theorems translate into convergence rates of the MAP estimates from (9).

COROLLARY 1. *Let* $\hat{b}_T = E^{\Pi_T}[b|X^T]$. *Under the conditions of Theorem 1, for every* $M_T \to \infty$,

$$\|\hat{b}_T - b_0\|_{\mu_0} = O_{P_{b_0}}(M_T T^{-\frac{a \wedge s}{2a+d}} \log T) \quad as \ T \to \infty,$$

while under the conditions of Theorem 2, for every $\delta > 5/2$,

$$\|\hat{b}_T - b_0\|_\infty = O_{P_{b_0}}(T^{-\frac{s \wedge [a - (d/2-2)_+]}{2a+d}}(\log T)^\delta) \quad as \ T \to \infty.$$

PROOF. Consider the function

$$H(b') = \Pi_T(b : \|b - b'\|_{\mu_0} \le M_T \varepsilon_T | X^T), \ b' \in V_J^{\otimes d}.$$

The posterior is a Gaussian measure on the finite-dimensional space $V_J^{\otimes d}$, centered at $\hat{b}_T$. Since $\| \cdot \|_{\mu_0}$-norm balls centred at the origin are convex symmetric sets, Anderson's Lemma (Theorem 2.4.5 of [26]) yields that $\hat{b}_T$ is a maximizer of H. Using Theorem 2.5 in [21] with the contraction rate from Theorem 1, we deduce that $\|\hat{b}_T - b_0\|_{\mu_0} = O_{P_{b_0}}(M_T \varepsilon_T)$ as $T \to \infty$. The $\| \cdot \|_\infty$-rate follows similarly using the contraction rate from Theorem 2. $\square$

Up to log-factors, the $\| \cdot \|_{L^2}$-rates obtained are minimax optimal for any dimension d (the lower bounds follow, e.g., via the asymptotic equivalence results in [16], see also [52, 53]). The $\| \cdot \|_\infty$-rates are then also optimal whenever $d \le 4$, up to log-factors. The sub-optimality of our rate for $d > 5$ is related to the presence of common semiparametric 'bias terms' in the approximation-theoretic Lemma 9 below.

2.4. *Bernstein-von Mises theorems for b.* We now adopt the framework of nonparametric Bernstein-von Mises theorems from [11, 12], see also the recent contributions [10, 34–36, 45]. The idea is to obtain a Gaussian approximation for the posterior distribution in a function space in which $1/\sqrt{T}$-convergence rates can be obtained. We will view the re-centred and re-scaled posterior draws $\sqrt{T}(b - \hat{b}_T)|X^T$ as (conditionally on X^T) random vector fields acting linearly on test functions $\phi = (\phi_1, \ldots, \phi_d)$ by integration

$$\left(\phi \mapsto \sqrt{T} \int_{\mathbb{T}^d} (b - \hat{b}_T).\phi : \phi \in B_{1\infty}^\rho \big| X^T\right),$$

and show that a Bernstein-von Mises theorem holds true uniformly in ϕ belonging to any bounded subset of the Besov space $B_{1\infty}^\rho, \rho > d/2, d \le 3$.

Equivalently, the limit theorem holds for the probability laws induced by these stochastic processes in the 'dual' Banach space $(B_{1\infty}^\rho)^*$. The limit will be the tight Gaussian probability measure $\mathcal{N}_{b_0}$ on $(B_{1\infty}^\rho)^*$ induced by the centred Gaussian white noise process $(\mathbb{W}_0(\phi) : \phi \in B_{1\infty}^\rho)$ with covariance

$$E\mathbb{W}_0(\phi)\mathbb{W}_0(\phi') = \langle \phi, \phi' \rangle_{1/\mu_0} = \sum_{j=1}^d \int_{\mathbb{T}^d} \phi_j(x)\phi_j'(x)\mu_0^{-1}(x), \quad \phi, \phi' \in B_{1\infty}^\rho;$$

its existence is established in the proof of the following theorem.

By embedding other spaces into $B_{1\infty}^\rho$ one may deduce various further limit theorems from the results below, for example in negative Sobolev spaces $H^{-\rho} = (H^\rho)^*, \rho > d/2$. For the applications to estimation of μ_b in the next subsection, this particular choice of Besov space is, however, crucial, and restriction to the simpler scale of Sobolev spaces would be insufficient to obtain the results in Section 2.5 below.

For two probability measures τ, τ' on a metric space (S, e), define the bounded Lipschitz (BL) metric for weak convergence (p.157 in [19]) by

$$\beta_S(\tau, \tau') = \sup_{F:S\to\mathbb{R}, \|F\|_{Lip}\le 1} \left| \int_S F d(\tau - \tau') \right|,$$

$$\|F\|_{Lip} \equiv \sup_{x\in S} |F(x)| + \sup_{x\ne y, x,y\in S} \frac{|F(x) - F(y)|}{e(x,y)}.$$

THEOREM 3. *Let $1 \le d \le 3$, $\rho > d/2$, $a > \max(3d/2 - 1, 1)$ and let $s \ge a$ be such that $s > a - 1 + d/2$. Suppose $b_0 \in C^s(\mathbb{T}^d) \cap H^s(\mathbb{T}^d)$. Let Π_T be the Gaussian prior from (11) with $\sigma_l = 2^{-l(\alpha+d/2)}, 0 \le \alpha < a \wedge s - d/2$ and J chosen such that $2^J \approx T^{1/(2a+d)}$. Let $\tilde{\Pi}_T(\cdot|X^T)$ be the conditional law $\mathcal{L}(\sqrt{T}(b - \hat{b}_T)|X^T)$, where $b \sim \Pi_T(\cdot|X^T)$ and $\hat{b}_T = E^{\Pi_T}[b|X^T]$ is the posterior mean, and let $\mathcal{N}_{b_0}$ denote the law in $(B_{1\infty}^\rho)^*$ of a centred Gaussian white noise process for $\langle \cdot, \cdot \rangle_{1/\mu_0}$. Then, as $T \to \infty$,*

$$(13) \qquad \beta_{(B_{1\infty}^\rho)^*}(\tilde{\Pi}(\cdot|X^T), \mathcal{N}_{b_0}) \to^{P_{b_0}} 0.$$

As in related situations in [11, 35], the condition $\rho > d/2$ cannot be relaxed as otherwise the limiting process does not exist as a tight probability measure in $(B_{1\infty}^\rho)^*$. Also the choices $p = 1, q = \infty$ are maximal for Besov spaces. From convergence of moments in (13) we deduce the following.

THEOREM 4. *Under the conditions of the previous theorem, the MAP estimate $\hat{b}_T = E^\Pi[b|X^T]$ satisfies, as $T \to \infty$,*

$$\sqrt{T}(\hat{b}_T - b_0) \to^d \mathcal{N}_{b_0} \text{ in } (B_{1\infty}^\rho)^*.$$

A confidence set for b can now be constructed by using the posterior quantiles to create a multiscale ball around $\hat{b}_T$, which can be further intersected with smoothness information as in [11, 12] to obtain confidence bands that are valid and near-optimal also in $\|\cdot\|_\infty$-diameter.

As remarked at the end of Section 2.3, the presence of semi-parametric bias terms prevents our proof from giving a Bernstein-von Mises theorem when $d \geq 4$, and also necessitates $s > a - 1 + d/2$ in Theorem 3. Unlike in Theorem 2, the case $d = 4$ is excluded as we need to suppress $\log T$-factors to obtain precise limit distributions. Similar phenomena occur in nonparametric smoothing (e.g. Section 3.6 in [25]).

2.5. *Bayesian inference on the invariant measure.* We now turn to the problem of making inference on the invariant measure μ_b. Frequentist estimators of μ_b can be suggested directly based on (2), e.g., [54]. For the Bayesian statistician, modelling μ_b directly by a prior is not coherent since μ_b does not identify the law P_b^T generating the likelihood (8) (cf. the discussion after Proposition 1). Instead, given the MAP estimate $\hat{b}_T$, we can (numerically) solve (7) to obtain a point estimate $\mu_{\hat{b}_T}$. For uncertainty quantification we can generate posterior samples $\mu_b | X^T$ from $b \sim \Pi_T(\cdot | X^T)$. Although numerical solvers for elliptic PDEs such as (7) are available, this algorithm may be computationally expensive. Nonetheless, it gives a principled Bayesian approach to inference on μ_b that, as the results in this section show, is optimal from an information theoretic point of view.

For the formulation of the following general result, we define spaces

$$\mathbb{B}_r = B_{1\infty}^r(\mathbb{T}^d) \cap L^2(\mathbb{T}^d), \quad r > 0, \ d \leq 3,$$

normed by $\|\cdot\|_{L^2} + \|\cdot\|_{B_{1\infty}^r}$; as before the conditional laws $\mathcal{L}(\sqrt{T}(\mu_b - \mu_{\hat{b}_T})|X^T)$ induce stochastic processes in the normed dual space $\mathbb{B}_r^*$ via actions

$$g \mapsto \sqrt{T} \int_{\mathbb{T}^d} (\mu_b - \mu_{\hat{b}_T})g, \quad g \in \mathbb{B}_r,$$

and weak convergence occurs in $\mathbb{B}_r^*$. We note that the inverse $L_{b_0}^{-1}$ of the generator L_{b_0} from (5) exists as a well-defined mapping from $L_{\mu_0}^2(\mathbb{T}^d)$ into $H^2(\mathbb{T}^d) \cap L_0^2(\mathbb{T}^d)$, see Lemma 11 in Section 6. We postpone the special case $d = 1$ to Theorem 6 below.

THEOREM 5. *Let $d = 2, 3$ and $r > d/2 - 1$. Under the conditions of Theorem 3, if $\mu_b, \mu_{\hat{b}_T}$ are the solutions of (7) (invariant measures) associated*

with a posterior draw $b \sim \Pi_T(\cdot|X^T)$ and $\hat{b}_T = E^{\Pi_T}[b|X^T]$, respectively, then for $\tau(\cdot|X^T)$ the conditional law $\mathcal{L}(\sqrt{T}(\mu_b - \mu_{\hat{b}_T})|X^T)$ in $\mathbb{B}_r^*$ we have

$$\beta_{\mathbb{B}_r^*}(\tau(\cdot|X^T), \mathcal{N}_{\mu_0}) \to^{P_{b_0}} 0, \quad \text{and} \quad \sqrt{T}(\mu_{\hat{b}_T} - \mu_0) \to^d \mathcal{N}_{\mu_0} \text{ in } \mathbb{B}_r^*$$

as $T \to \infty$, where $\mathcal{N}_{\mu_0}$ is the tight Borel probability measure on $\mathbb{B}_r^*$ induced by the centred Gaussian process $\mathbb{M}$ with covariance metric

$$E\mathbb{M}(g)\mathbb{M}(g') = \langle \nabla L_{b_0}^{-1}[\bar{g}], \nabla L_{b_0}^{-1}[\bar{g}'] \rangle_{\mu_0}, \quad \bar{g} = g - \int_{\mathbb{T}^d} g d\mu_0, \quad g, g' \in \mathbb{B}_r.$$

This theorem has various corollaries, upon using the richness of the spaces $\mathbb{B}_r, r > d/2 - 1$. For instance, since H^r embeds continuously into $\mathbb{B}^r$ on the bounded domain $\mathbb{T}^d$, one deduces weak convergence in P_{b_0}-probability of the conditional laws in the negative Sobolev spaces $H^{-r}(\mathbb{T}^d) = (H^r(\mathbb{T}^d))^*$:

$$\beta_{H^{-r}}\left(\mathcal{L}(\sqrt{T}(\mu_b - \mu_{\hat{b}_T})|X^T), \mathcal{N}_{\mu_0}\right) \to^{P_{b_0}}_{T \to \infty} 0, \quad r > d/2 - 1, \ d = 2, 3.$$

2.5.1. *Bayesian inference on invariant probabilities.* Indicator functions of measurable subsets C of $\mathbb{T}^d$ of finite perimeter define elements of $B_{1\infty}^1(\mathbb{T}^d)$ (proved, e.g., as in Lemma 8b, [24]) and we can thus make inference on invariant probabilities $\mu_b(C) = \int_{\mathbb{T}^d} 1_C d\mu$ for $d = 2, 3$. Let $\mathcal{C} = \mathcal{C}_K$ be a class of Borel subsets of $(0, 1]^d$ that have perimeter bounded by a fixed constant K. This includes, in particular, all convex subsets of $\mathbb{T}^d$ (e.g., Remark 5 in [24]). Then the collection of functions $\{1_C : C \in \mathcal{C}\}$ is bounded in $B_{1\infty}^1(\mathbb{T}^d) \cap L^2(\mathbb{T}^d)$, and for the resulting set-indexed process of posterior invariant probabilities $(\mu_b(C) : C \in \mathcal{C}), b \sim \Pi_T(\cdot|X^T)$, we deduce from Theorem 5 and the continuous mapping theorem

$$\beta\left(\mathcal{L}(\sqrt{T}(\mu_b(\cdot) - \mu_{\hat{b}_T}(\cdot))|X^T), \mathcal{N}_{\mu_0}\right) \to^{P_{b_0}} 0, \ \sqrt{T}(\mu_{\hat{b}_T} - \mu_0) \to^d \mathcal{N}_{\mu_0} \text{ in } \ell^\infty(\mathcal{C}),$$

as $T \to \infty$, where $\beta = \beta_{\ell^\infty(\mathcal{C})}$, $\ell^\infty(\mathcal{C}) \supset \mathbb{B}_r^*$ is the Banach space of bounded functions on $\mathcal{C}$ (see Proposition 3.7.24 in [26] for a precise definition of β_S for non-separable S). One further deduces that the estimated invariant probabilities induced by the MAP estimate $\hat{b}_T$ obey the limit law

$$\sqrt{T} \sup_{C \in \mathcal{C}} |\mu_{\hat{b}_T}(C) - \mu_0(C)| \to^d \sup_{C \in \mathcal{C}} |\mathbb{M}(1_C)| < \infty \text{ a.s., } T \to \infty.$$

2.5.2. *The one-dimensional case.* We finally turn to the special case $d = 1$, where the proof of Theorem 5 needs adaptations as then $r > d/2 - 1$ can be negative. We obtain a central limit theorem for the invariant probability densities $(\mu_b(x), x \in \mathbb{T})$ viewed as random functions in $C(\mathbb{T})$.

For $d = 1$ the solution map L_b^{-1} from before Theorem 5 has a representation $L_b^{-1}[g] = \int_{\mathbb{T}} G_b(\cdot, y)g(y)dy, g \in L_{\mu_b}^2(\mathbb{T})$, with periodic Green kernel $G_b : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ such that $G_b(\cdot, x) \in H^1(\mathbb{T})$ for all $x \in \mathbb{T}$. This follows, e.g., from directly deriving explicit expressions for the solution v of the ODE $bv' + v''/2 = (2\mu_b)^{-1}(\mu_b v')' = g$, where $\mu_b \propto e^{2B}$ and $B' = b$.

THEOREM 6.   *Under the conditions of Theorem 3 with $d = 1, a > 3/2$, if $\mu_b(x), \mu_{\hat{b}_T}(x), x \in \mathbb{T}$, are the invariant probability density functions associated to $b \sim \Pi(\cdot|X^T), \hat{b}_T = E^{\Pi_T}[b|X^T]$, respectively, then, as $T \to \infty$,*

$$\beta_{C(\mathbb{T})}\big(\mathcal{L}(\sqrt{T}(\mu_b - \mu_{\hat{b}_T})|X^T), \bar{\mathcal{N}}_{b_0}\big) \to^{P_{b_0}} 0, \ \text{and} \ \sqrt{T}(\mu_{\hat{b}_T} - \mu_0) \to^d \bar{\mathcal{N}}_{b_0} \ \text{in} \ C(\mathbb{T}),$$

where $\bar{\mathcal{N}}_{b_0}$ is the Borel probability law in $C(\mathbb{T})$ induced by the centred Gaussian random function $(\bar{\mathbb{M}}(x) : x \in \mathbb{T})$ with covariance

$$E\bar{\mathbb{M}}(x)\bar{\mathbb{M}}(x') = \int_{\mathbb{T}} \frac{d}{dy}G_{b_0}(y, x)\frac{d}{dy}G_{b_0}(y, x')d\mu_0(y), \quad x, x' \in \mathbb{T}.$$

In [3] an analogue of the second limit in the above theorem was obtained for an estimator based on smoothing the empirical measure $\hat{\mu}_T$ from (2). Their proof is very different from ours and based on first establishing that their estimator is asymptotically close to the local time of the diffusion process, in conceptual analogy to the i.i.d. setting [25].

2.5.3. *Information lower bounds.* The LAN expansion of our measurement model under P_{b_0} is obtained in Lemma 6 below, with LAN-inner product $\langle \cdot, \cdot \rangle_\mu$. Standard arguments from asymptotic semiparametric statistics ([58], Chapter 25) then imply that the asymptotic variance occurring in Theorems 3 and 4 is optimal in an information-theoretic sense. This is also true in the case of Theorem 5, where inference on a non-linear functional $\Phi_g(b) = \int_{\mathbb{T}^d} g d\mu_b, g \in L^2(\mathbb{T}^d)$, of b is considered. Indeed, the expansion

$$\Phi_g(b + h) - \Phi_g(b) = \langle \nabla L_b^{-1}[\bar{g}], h \rangle_{\mu_b} + o(\|h\|_\infty), \quad \bar{g} = g - \int_{\mathbb{T}^d} g d\mu_b,$$

follows from the proof of Theorem 5. Thus arguing as in Section 7.5 in [35], the information lower bound for estimating $\Phi_g(b)$ from our observations is

$$\|\nabla L_b^{-1}[\bar{g}]\|_{\mu_b}^2 = \int_{\mathbb{T}^d} \left\| \nabla L_b^{-1}\left[g - \int_{\mathbb{T}^d} g d\mu_b\right](x) \right\|^2 d\mu_b(x), \quad \text{any } d \geq 1.$$

Examining the proof of Theorem 6, a similar remark applies to the covariance appearing in that theorem. See also [3] for the case $d = 1$.

3. Proofs of main results.

3.1. *Concentration of measure tools for multi-dimensional diffusions.* The following results provide uniform stochastic control of functionals of the diffusion process (1) with periodic drift b_0 in terms of metric entropy bounds via a metric d_L involving the inverse generator $L_{b_0}^{-1}$ from Lemma 11.

LEMMA 1. *Suppose $b_0 \in C^{(d/2+\kappa)\vee 1}(\mathbb{T}^d)$ and let $\mathcal{F}_T \subset L_{\mu_0}^2(\mathbb{T}^d) \cap H^{d/2+\kappa}(\mathbb{T}^d)$ for some $\kappa > 0$ be such that $0 \in \mathcal{F}_T$. Define the empirical process*

$$\mathbb{G}_T[f] = \frac{1}{\sqrt{T}} \int_0^T f(X_s) ds, \quad f \in \mathcal{F}_T,$$

the pseudo-distance d_L on $\mathcal{F}_T$ by

$$(14) \qquad d_L^2(f,g) = \sum_{i=1}^d \left\| \partial_{x_i} L_{b_0}^{-1}[f-g] \right\|_\infty^2,$$

and let $D_{\mathcal{F}_T}$ be the d_L-diameter of $\mathcal{F}_T$. Further set

$$J_{\mathcal{F}_T} = J(\mathcal{F}_T, 6d_L, D_{\mathcal{F}_T}) = \int_0^{D_{\mathcal{F}_T}} \sqrt{\log 2N(\mathcal{F}_T, 6d_L, \tau)} d\tau,$$

where $N(\mathcal{F}_T, 6d_L, \tau)$ denotes the covering number of the set $\mathcal{F}_T$ by d_L-balls of radius $\tau/6$. Then

$$E_{b_0} \sup_{f \in \mathcal{F}_T} |\mathbb{G}_T(f)| \le \frac{2}{\sqrt{T}} \sup_{f \in \mathcal{F}_T} \|L_{b_0}^{-1}[f]\|_\infty + 4\sqrt{2} J_{\mathcal{F}_T},$$

and for any $x > 0$,

$$P_{b_0}\left(\sup_{f \in \mathcal{F}_T} |\mathbb{G}_T(f)| \ge \sup_{f \in \mathcal{F}_T} \frac{2\|L_{b_0}^{-1}[f]\|_\infty}{\sqrt{T}} + J_{\mathcal{F}_T}(4\sqrt{2} + 192x) \right) \le e^{-x^2/2}.$$

PROOF. By Lemma 11 and the Sobolev embedding theorem, the Poisson equation $Lu = L_{b_0}u = f$ has a unique solution $L^{-1}[f] \in H^{d/2+\kappa+2} \cap L_0^2 \subset C^2$ satisfying $LL^{-1}[f] = f$ for any $f \in \mathcal{F}_T$. We may therefore define for $f \in \mathcal{F}_T$,

$$\begin{aligned} Z_T(f) &= \int_0^T \nabla L^{-1}[f](X_s).dW_s \\ &= L^{-1}[f](X_T) - L^{-1}[f](X_0) - \int_0^T LL^{-1}[f](X_s)ds, \end{aligned}$$

where we have used Itô's lemma (Theorem 39.3 in [5]). Since

$$\sup_{f \in \mathcal{F}_T} \left| \int_0^T f(X_s)ds \right| - 2 \sup_{f \in \mathcal{F}_T} \|L^{-1}[f]\|_\infty \leq \sup_{f \in \mathcal{F}_T} |Z_T(f)|,$$

it suffices to control $\sup_{f \in \mathcal{F}_T} |Z_T(f)|$. For fixed $f \in \mathcal{F}_T$, $Z_T(f)$ is a continuous square integrable local martingale with quadratic variation

$$[Z.(f)]_T = \int_0^T \|\nabla L^{-1}[f](X_s)\|^2 ds \leq T \sum_{i=1}^d \|\partial_{x_i} L^{-1}[f]\|_\infty^2 = T d_L^2(f, 0).$$

Recall Bernstein's inequality for continuous local martingales (p. 153 of [46]): if M is a continuous local martingale vanishing at 0 with quadratic variation $[M]$, then for any stopping time T and any $y, K > 0$,

$$(15) \qquad P\big(\sup_{0 \leq t \leq T} |M_t| \geq y, [M]_T \leq K \big) \leq 2e^{-\frac{y^2}{2K}}.$$

Applying this to $Z_T(f)$ gives for any $f \in \mathcal{F}_T$ and $x > 0$,

$$P_{b_0}\big(|Z_T(f)| \geq \sqrt{T}x\big) = P_{b_0}\big(|Z_T(f)| \geq \sqrt{T}x, [Z.(f)]_T \leq T d_L^2(f, 0)\big) \leq 2e^{-\frac{x^2}{2d_L^2(f,0)}}.$$

Since L^{-1} is linear, so is $f \mapsto Z_T(f)$, and consequently

$$P_{b_0}\big(|Z_T(f) - Z_T(g)| \geq \sqrt{T}x\big) \leq 2\exp\left(-\frac{x^2}{2d_L^2(f, g)}\right),$$

a non-asymptotic inequality. The process $(T^{-1/2} Z_T(f) : f \in \mathcal{F}_T)$ is thus mean-zero and subgaussian with respect to d_L. From this we deduce that $E_{b_0} \sup_{f \in \mathcal{F}_T} T^{-1/2} |Z_T(f)| \leq 4\sqrt{2} J_{\mathcal{F}_T}$ by the usual chaining bound for subgaussian processes (e.g., Theorem 2.3.7 of [26] - the factor 6 scales the subgaussian constant, see after Definition 2.3.5 of [26]). This chaining bound extends to exponential ψ_2-Orlicz norms $\|\cdot\|_{\psi_2}$ (see Exercise 2.3.1 of [26]), so one further has $\big\|\sup_{f \in \mathcal{F}_T} T^{-1/2} |Z_T(f)|\big\|_{\psi_2} \leq 16\sqrt{6} J_{\mathcal{F}_T}$. Using Lemma 2.3.1 of [26] and that for any random variable X, $\|X - \mathbb{E}X\|_{\psi_2} \leq 2\|X\|_{\psi_2}$, we obtain for any $x > 0$,

$$P_{b_0}\left(\sup_{f \in \mathcal{F}_T} \frac{|Z_T(f)|}{\sqrt{T}} \geq E_{b_0} \sup_{f \in \mathcal{F}_T} \frac{|Z_T(f)|}{\sqrt{T}} + x \right) \leq \exp\left(-\frac{x^2}{2(196 J_{\mathcal{F}_T})^2}\right).$$

Using the expectation bound just derived, the above inequality yields

$$P_{b_0}\left(\sup_{f \in \mathcal{F}_T} T^{-1/2} |Z_T(f)| \geq 4\sqrt{2} J_{\mathcal{F}_T} + 196 J_{\mathcal{F}_T} x \right) \leq e^{-\frac{x^2}{2}}.$$

Combining the above gives the required subgaussian inequality. $\qquad\square$

We now establish usable bounds for the metric d_L. The following is a special case of the Runst-Sickel lemma.

LEMMA 2 ([48], p. 345). *For $t > 0$ and any bounded $f, g \in H^t(\mathbb{T}^d)$,*

$$\|fg\|_{H^t} \leq C(t, d) \left(\|f\|_{H^t} \|g\|_\infty + \|g\|_{H^t} \|f\|_\infty \right).$$

LEMMA 3. *Suppose $b_0 \in C^s(\mathbb{T}^d)$ for $s > \max(d/2 - 1, 0)$. Then for any $0 < \kappa < s - d/2 + 1$ (or $\kappa = 0$ if $d = 1$) and $f, g \in L^2_{\mu_0}(\mathbb{T}^d)$, the pseudo-distance d_L in (14) satisfies*

$$d_L(f, g) \leq C(d, \kappa, b_0) \|f - g\|_{H^{(d/2+\kappa-1)_+}},$$

where $H^0 = L^2$. Moreover, let V_J denote the span of all wavelets up to resolution level J of an S-regular wavelet basis of $L^2(\mathbb{T}^d)$. If $\gamma, \rho \in V_J$ and $0 \leq p < S$, then for $C = C(p, d, \Phi, \|\mu_0\|_\infty)$,

$$\|\gamma\rho - \langle\gamma, \rho\rangle_{\mu_0}\|_{H^p} \leq C 2^{J(p+d/2)} \|\gamma\|_{L^2} \|\rho\|_{L^2}.$$

PROOF. If $d \geq 2$, then for any $0 < \kappa < s - d/2 + 1$, by the Sobolev embedding theorem and Lemma 11 there exists $C = C(d, \kappa, b_0)$ such that

$$d_L^2(f, g) = \sum_{i=1}^d \|\partial_{x_i} L^{-1}[f-g]\|_\infty^2 \leq C\|L^{-1}[f-g]\|_{H^{d/2+\kappa+1}}^2 \leq C\|f-g\|_{H^{d/2+\kappa-1}}^2.$$

If $d = 1$, one similarly has $d_L^2(f, g) \leq C\|L^{-1}[f - g]\|_{H^2}^2 \leq C\|f - g\|_{L^2}^2$. For the second statement, if $p > 0$, then the triangle inequality, Lemma 2 and the Cauchy-Schwarz inequality bound the quantity in question by

$$C(p, d)(\|\gamma\|_{H^p}\|\rho\|_\infty + \|\rho\|_{H^p}\|\gamma\|_\infty) + \|\mu_0\|_\infty\|\gamma\|_{L^2}\|\rho\|_{L^2}\|1\|_{H^p}.$$

If $p = 0$, one instead uses the simpler bound $\|\gamma\rho\|_{L^2} \leq \|\gamma\|_{L^2}\|\rho\|_\infty$. By the wavelet characterisation of the Sobolev norm,

$$(16) \qquad \|\gamma\|_{H^p}^2 = \sum_{l \leq J} \sum_r 2^{2lp} |\langle\gamma, \Phi_{l,r}\rangle|^2 \leq 2^{2Jp}\|\gamma\|_{L^2}^2.$$

Using Cauchy-Schwarz and that for all $l \geq 0$, $\|\sum_r \Phi_{l,r}^2\|_\infty \leq C(\Phi)2^{ld}$,

$$(17) \qquad \|\gamma\|_\infty \leq \sup_x \sum_{l \leq J} \sum_r |\langle\gamma, \Phi_{l,r}\rangle||\Phi_{l,r}(x)| \leq C(\Phi)2^{Jd/2}\|\gamma\|_{L^2}.$$

Applying these bounds to $\gamma, \rho \in V_J$ gives the result. $\qquad\square$

3.2. *A restricted isometry inequality for* $h_T^2(\cdot, \cdot)$ *on* $V_J^{\otimes d}$. We next con-
sider the action on the spaces $V_J^{\otimes d}$ of the random distance h_T defined by

$$Th_T^2(b_1, b_2) \equiv \int_0^T \|b_1(X_s) - b_2(X_s)\|^2 ds = \sum_{j=1}^d \int_0^T |b_{1,j}(X_s) - b_{2,j}(X_s)|^2 ds.$$

From the preceding concentration inequalities, and using a commonly used
contraction principle to bound minimal eigenvalues of random matrices (e.g.,
Section 5.6 in [61]), we establish the following key inequality.

LEMMA 4.   *Suppose* $b_0 \in C^s(\mathbb{T}^d)$ *for* $s > \max(d/2, 1)$ *and let* $J \in \mathbb{N}$. *For*
V_J *as in (4) set* $v_J := dim(V_J) = O(2^{Jd})$. *Then for any* $0 < \kappa < s - d/2 + 1$
(or $\kappa = 0$ *if* $d = 1$*), there exist positive constants* $c_0 = c_0(b_0)$ *and* $C = C(d, b_0, \kappa, \Phi)$ *such that for any* $x > 0$,

$$P_{b_0}\left[\sup_{b,\bar b \in V_J^{\otimes d}: b \neq \bar b} \left| \frac{h_T^2(b, \bar b) - \|b - \bar b\|_{\mu_0}^2}{\|b - \bar b\|_{\mu_0}^2} \right| \geq \frac{2^{J[\frac{d}{2} + (\frac{d}{2} + \kappa - 1)_+]}}{\sqrt{T}/C}(1 + x) \right] \leq d e^{c_0 v_J - \frac{x^2}{2}}.$$

PROOF.   Let $b_j, \bar b_j \in V_J$ and write $b = (b_1, \ldots, b_d)$, $\bar b = (\bar b_1, \ldots, \bar b_d)$ with
$b_j = \sum_{l \leq J, r} \theta_{l,r,j} \Phi_{l,r}$ and $\bar b_j = \sum_{l \leq J, r} \bar\theta_{l,r,j} \Phi_{l,r}$. Then

$$h_T^2(b, \bar b) = \frac{1}{T} \sum_{j=1}^d \int_0^T \left(\sum_{l,r} (\theta_{l,r,j} - \bar\theta_{l,r,j}) \Phi_{l,r}(X_s) \right)^2 ds$$

$$= \sum_{j=1}^d \sum_{l,r} \sum_{l',r'} (\theta_{l,r,j} - \bar\theta_{l,r,j})(\theta_{l',r',j} - \bar\theta_{l',r',j}) \frac{1}{T} \int_0^T \Phi_{l,r}(X_s) \Phi_{l',r'}(X_s) ds$$

$$= \sum_{j=1}^d (\theta_{j,\cdot} - \bar\theta_{j,\cdot})^T \hat\Gamma (\theta_{j,\cdot} - \bar\theta_{j,\cdot}),$$

where $\hat\Gamma_{(l,r)(l',r')} = \frac{1}{T} \int_0^T \Phi_{l,r}(X_s) \Phi_{l',r'}(X_s) ds$, so that $\hat\Gamma$ is a $v_J \times v_J$ symmet-
ric matrix. Similarly,

$$\|b - \bar b\|_{\mu_0}^2 = \sum_{j=1}^d (\theta_{j,\cdot} - \bar\theta_{j,\cdot})^T \Gamma (\theta_{j,\cdot} - \bar\theta_{j,\cdot}),$$

where $\Gamma_{(l,r)(l',r')} = \int_{\mathbb{T}^d} \Phi_{l,r}(x) \Phi_{l',r'}(x) d\mu_0(x)$. Denote the quantity on the
r.h.s. in the inequality in Lemma 4 by $\zeta_T = CT^{-1/2} 2^{J[d/2 + (d/2 + \kappa - 1)_+]}(1 + x)$.

Since $(\theta_{j,\cdot} - \bar\theta_{j,\cdot})^T \Gamma (\theta_{j,\cdot} - \bar\theta_{j,\cdot}) = \|b_j - \bar b_j\|_{\mu_0}^2 \geq 0$ for all j, applying a union bound to the probability in Lemma 4 gives

$$\sum_{j=1}^d P_{b_0} \left(\sup_{\theta_{j,\cdot}, \bar\theta_{j,\cdot} \in \mathbb{R}^{v_J} : (\theta_{j,\cdot} - \bar\theta_{j,\cdot})^T \Gamma(\theta_{j,\cdot} - \bar\theta_{j,\cdot}) \neq 0} \left| \frac{(\theta_{j,\cdot} - \bar\theta_{j,\cdot})^T (\hat\Gamma - \Gamma)(\theta_{j,\cdot} - \bar\theta_{j,\cdot})}{(\theta_{j,\cdot} - \bar\theta_{j,\cdot})^T \Gamma(\theta_{j,\cdot} - \bar\theta_{j,\cdot})} \right| \geq \frac{\zeta_T}{d} \right).$$

(Note that at least one $(\theta_{j,\cdot} - \bar\theta_{j,\cdot})^T \Gamma(\theta_{j,\cdot} - \bar\theta_{j,\cdot}) \neq 0$ by assumption and the above supremum is maximized when $\theta_{j,\cdot} \neq \bar\theta_{j,\cdot}$, so the denominator is well-defined for all j). Setting $u = (\theta_{j,\cdot} - \bar\theta_{j,\cdot}) \in \mathbb{R}^{v_J}$ and using the bilinearity of the above quadratic form, each of the previous probabilities, which are all equal, are bounded by

$$(18) \qquad P_{b_0} \left(\sup_{u \in \Theta} |u^T \Lambda u| \geq \zeta_T/d \right),$$

where $\Theta = \{u \in \mathbb{R}^{v_J} : u^T \Gamma u \leq 1\}$ and $\Lambda = \hat\Gamma - \Gamma$. Let $\|u\|_\Gamma^2 := u^T \Gamma u$, $u \in \mathbb{R}^{v_J}$, and for $0 < \delta < 1$, let $(u^l)_{l=1}^{N(\delta)}$ be a minimal δ-covering of Θ in $\|\cdot\|_\Gamma$-distance. For every $u \in \Theta$, let $u^l = u^l(u)$ denote the closest point in this δ-covering, so that $\|u - u^l\|_\Gamma \leq \delta$. By bilinearity, for any $u \in \Theta$,

$$|(u - u^l)^T \Lambda(u - u^l)| \leq \delta^2 \sup_{w \in \Theta} |w^T \Lambda w|.$$

For any $u \in \Theta$, set $g_u = \sum_{l \leq J, r} u_{l,r} \Phi_{l,r}$. By Proposition 1, $\|u\|_{\mathbb{R}^{v_J}} = \|g_u\|_{L^2} \leq \|1/\mu_0\|_\infty^{1/2} \|g_u\|_{\mu_0} = \|1/\mu_0\|_\infty^{1/2} \|u\|_\Gamma$. For $(\lambda_i)_{i=1}^{v_J}$ the eigenvalues of the symmetric matrix Λ and $\lambda_{max} = \max_i |\lambda_i|$, applying the Cauchy-Schwarz inequality gives

$$|(u - u^l)^T \Lambda u^l| \leq \|u - u^l\|_{\mathbb{R}^{v_J}} \|\Lambda u^l\|_{\mathbb{R}^{v_J}} \leq \delta \|1/\mu_0\|_\infty^{1/2} \lambda_{max} \|u^l\|_{\mathbb{R}^{v_J}}$$
$$\leq \delta \|1/\mu_0\|_\infty \sup_{v : \|v\|_{\mathbb{R}^{v_J}} \leq 1} |v^T \Lambda v|,$$

where the last inequality follows from p.234 of [29]. Since $\sup_{v : \|v\|_{\mathbb{R}^{v_J}} \leq 1} |v^T \Lambda v| \leq \|\mu_0\|_\infty \sup_{w \in \Theta} |w^T \Lambda w|$, then $|(u - u^l)^T \Lambda u^l| \leq \delta \|1/\mu_0\|_\infty \|\mu_0\|_\infty \sup_{w \in \Theta} |w^T \Lambda w|$ for all $u \in \Theta$. Combining the above yields for $0 < \delta < 1$,

$$\sup_{u \in \Theta} |u^T \Lambda u| \leq (\delta^2 + 2\delta \|1/\mu_0\|_\infty \|\mu_0\|_\infty) \sup_{w \in \Theta} |w^T \Lambda w| + \max_{1 \leq l \leq N(\delta)} |(u^l)^T \Lambda u^l|,$$

and for $\delta_0 = \delta_0(\mu_0)$ small enough that $\delta_0^2 + 2\|1/\mu_0\|_\infty \|\mu_0\|_\infty \delta_0 \leq 1/2$,

$$(19) \qquad \sup_{u \in \Theta} |u^T \Lambda u| \leq 2 \max_{1 \leq l \leq N(\delta_0)} |(u^l)^T \Lambda u^l|.$$

A union bound now yields that (18) is bounded by $N(\delta_0) \sup_{u \in \Theta} P_{b_0}(|u^T \Lambda u| \geq \zeta_T/(2d))$. The covering number of the unit ball in a v_J-dimensional space is bounded by $N(\delta_0) \leq (C/\delta_0)^{v_J} = e^{c_0 v_J}$ (Proposition 4.3.34 of [26]).

For $u \in \Theta$, set $f_u(x) = g_u(x)^2 - \langle g_u, g_u \rangle_{\mu_0} \in L^2_{\mu_0}(\mathbb{T}^d) \cap H^S(\mathbb{T}^d)$, where $S > d/2$ is the regularity of the wavelet basis. Since also $b_0 \in C^s$ with $s > d/2$, applying Lemma 1 to the class $\mathcal{F} = \{f_u, 0\}$ and noting that $u^T \Lambda u = T^{-1} \int_0^T f_u(X_t) dt$ yields

(20)
$$P_{b_0} \left(|u^T \Lambda u| \geq CT^{-1} \|L^{-1}[f_u]\|_\infty + CT^{-1/2} d_L(f_u, 0)(1 + x) \right) \leq e^{-x^2/2}.$$

For $0 < \kappa < s - d/2 + 1$ (or $\kappa = 0$ if $d = 1$), applying Lemma 3 with $\gamma = \rho = g_u \in V_J$ and $p = (d/2 + \kappa - 1)_+$ gives

$$d_L(f_u, 0) \leq C \|g_u^2 - \langle g_u, g_u \rangle_{\mu_0}\|_{H^{(d/2+\kappa-1)_+}} \leq C 2^{J[d/2+(d/2+\kappa-1)_+]} \|g_u\|_{L^2}^2.$$

By Proposition 1, $\|g_u\|_{L^2}^2 \leq \|1/\mu_0\|_\infty \|g_u\|_{\mu_0}^2 = \|1/\mu_0\|_\infty u^T \Gamma u \leq \|1/\mu_0\|_\infty$, so that $d_L(f_u, 0) \leq C 2^{J[d/2+(d/2+\kappa-1)_+]}$ for any $u \in \Theta$. Applying the Sobolev embedding theorem, Lemma 11 and Lemma 3 as above, $\|L^{-1}[f_u]\|_\infty \lesssim \|f_u\|_{H^{(d/2+\kappa-2)_+}} \lesssim 2^{J[d/2+(d/2+\kappa-1)_+]}$ for κ as above and any $u \in \Theta$. Substituting this into (20) gives

$$\sup_{u \in \Theta} P_{b_0} \left(|u^T \Lambda u| \geq CT^{-1/2} 2^{J[d/2+(d/2+\kappa-1)_+]}(1 + x) \right) \leq e^{-x^2/2},$$

where the right-hand side equals ζ_T up to constants. Combining the last inequality with (19) and the remarks after it completes the proof. □

3.3. *Proof of Theorem 1.* As a first step we obtain a convergence rate in the 'random Hellinger distance h_T' defined before Lemma 4 corresponding to the regression problem posed by equation (1). This random semimetric arises naturally in the classical testing approach (see [22], and formulated in the Brownian semi-martingale setting relevant here by van der Meulen et al. [57]), since the log-likelihood with respect to $P_{b_0}^T$ can be expressed as $M - \frac{1}{2}[M]$, where M is a continuous local martingale with quadratic variation $[M]_T = Th_T^2(b, b_0)$. The next result is a combination of Theorem 2.1 and Lemma 2.2 of [57], restated in the present context. The proof relies on martingale arguments which generalize to the multidimensional setting without difficulty, hence the proof is left to the reader.

Consider the statistical experiments $(P_b^T : b \in \mathfrak{B}_T)$, where the parameter spaces $\mathfrak{B}_T$, which are allowed to vary with T, are arbitrary sets equipped with σ-algebras satisfying mild measurability conditions, see Section 2 of

[57]. In particular, these are satisfied by the finite-dimensional spaces considered in Theorem 1.

THEOREM 7. *Let $\varepsilon_T \to 0$ be such that $T\varepsilon_T^2 \to \infty$. Suppose that for any $C_1 > 0$, there exist measurable sets $\mathcal{B}_{\mathcal{T}}$ and $C_2, C_3 > 0$ such that*

$$\Pi_T(\mathcal{B}_T^c) \le e^{-C_1 T\varepsilon_T^2}, \tag{21}$$

$$\log N(\mathcal{B}_T, \|\cdot\|_{\mu_0}, \varepsilon_T) \le C_2 T\varepsilon_T^2, \tag{22}$$

$$\Pi_T(b : \|b - b_0\|_{\mu_0} \le \varepsilon_T) \ge e^{-C_3 T\varepsilon_T^2}. \tag{23}$$

Assume further that for every $\gamma > 0$ there exist $c_\gamma, C_\gamma > 0, D_\gamma \ge 0$ such that

(24)
$$\liminf_{T\to\infty} P_{b_0}\Big(c_\gamma\|b - b_0\|_{\mu_0} \le h_T(b, b_0), \forall b \in \mathfrak{B}_T \text{ s.t. } h_T(b, b_0) \ge D_\gamma\varepsilon_T, \text{ and}$$

$$h_T(b_1, b_2) \le C_\gamma\|b_1 - b_2\|_{\mu_0}, \forall b_1, b_2 \in \mathfrak{B}_T \text{ s.t. } h_T(b_1, b_2) \ge D_\gamma\varepsilon_T\Big) \ge 1 - \gamma.$$

Then for every $M_T \to \infty$, $\Pi_T(b \in \mathfrak{B}_T : h_T(b, b_0) \ge M_T\varepsilon_T | X^T) \xrightarrow[T\to\infty]{P_{b_0}} 0$.

The proof of the theorem implies in particular that the denominator in (10) is non-zero on events of P_{b_0}-probability approaching one. We now turn to the proof of Theorem 1 and verify the conditions (21)-(24) of Theorem 7. By Proposition 1, $\|\cdot\|_{L^2}$ and $\|\cdot\|_{\mu_0}$ are equivalent norms. Applying Theorem 4.5 of [59] (see also Sections 11.3 and 11.4.5 in [22]), there exist measurable sets $B_T \subset V_J$ such that for $\varepsilon_T = T^{-\frac{a \wedge s}{2a+d}}(\log T)$,

(i) $\log N(B_T, \|\cdot\|_\infty, 3\varepsilon_T) \le 6CT\varepsilon_T^2$,
(ii) $\Pi_T(b_j \notin B_T) \le e^{-CT\varepsilon_T^2}$ and
(iii) $\Pi_T(\|b_j - b_{0,j}\|_\infty < 4\varepsilon_T) \ge e^{-T\varepsilon_T^2}$.

The set $\mathcal{B}_T = B_T^{\otimes d}$ satisfies $\log N(\mathcal{B}_T, \|\cdot\|_{\mu_0}, 3d^{1/2}\|\mu_0\|_\infty^{1/2}\varepsilon_T) \le 6dCT\varepsilon_T^2$ and $\Pi_T(\mathcal{B}_T^c) \le de^{-CT\varepsilon_T^2}$, which verifies (21) and (22) for (a constant multiple of) ε_T. Further,

$$\Pi_T\left(\|b - b_0\|_{\mu_0} \le 4\sqrt{d}\|\mu_0\|_\infty^{1/2}\varepsilon_T\right) \ge \Pi_T\left(\sup_{j=1,\ldots,d}\|b_j - b_{0,j}\|_{L^2} \le 4\varepsilon_T\right)$$

$$\ge \prod_{j=1}^d \Pi_T\left(\|b_j - b_{0,j}\|_\infty \le 4\varepsilon_T\right) \ge e^{-dT\varepsilon_T^2},$$

thereby verifying (23) for (a constant multiple of) ε_T.

We now verify (24). Since $\Pi_T(V_J^{\otimes d}) = 1$, we may take as parameter space $\mathfrak{B}_T = V_J^{\otimes d} \cup \{b_0\}$. Let $b_{0,j,J}$ denote the orthogonal projection of $b_{0,j}$ onto V_J and set $b_{0,J} = (b_{0,1,J}, \ldots, b_{0,d,J})$. Since $b_0 \in C^s$, $\|b_{0,j} - b_{0,j,J}\|_\infty \leq C(b_0)2^{-Js} \leq C(b_0)\varepsilon_T$, so that for our choice of J this yields $h_T(b_0, b_{0,J}) \leq r\varepsilon_T$ and $\|b_0 - b_{0,J}\|_{\mu_0} \leq \|\mu_0\|_\infty^{1/2}\|b_0 - b_{0,J}\|_{L^2} \leq r\varepsilon_T$ for some $r = r(d, b_0)$. By considering the cases $b_1 \in V_J^{\otimes d}$ and $b_1 = b_0$ separately, the event in (24) therefore contains the event

$$\{c_\gamma\|b - b_{0,J}\|_{\mu_0} + c_\gamma r\varepsilon_T \leq h_T(b, b_{0,J}) - r\varepsilon_T$$
$$\text{and } h_T(b, b_{0,J}) + r\varepsilon_T \leq C_\gamma\|b - b_{0,J}\|_{\mu_0} - C_\gamma r\varepsilon_T,$$
$$\forall b \in V_J^{\otimes d} \text{ with } h_T(b, b_{0,J}) \geq (D_\gamma - r)\varepsilon_T\}$$
$$\cap \{h_T(b_1, b_2) \leq C_\gamma\|b_1 - b_2\|_{\mu_0}, \forall b_1, b_2 \in V_J^{\otimes d} \text{ with } h_T(b_1, b_2) \geq D_\gamma\varepsilon_T\}.$$

For D_γ large enough that $(D_\gamma - r) \geq \max\{(C_\gamma + 1)r, 2(c_\gamma + 1)r\}$, the last event contains

$$\{2c_\gamma\|b - b_{0,J}\|_{\mu_0} \leq h_T(b, b_{0,J}) \leq \tfrac{1}{2}C_\gamma\|b - b_{0,J}\|_{\mu_0},$$
$$\forall b \in V_J^{\otimes d} \text{ with } h_T(b, b_{0,J}) \geq (D_\gamma - r)\varepsilon_T\}$$
$$\cap \{h_T(b_1, b_2) \leq C_\gamma\|b_1 - b_2\|_{\mu_0}, \forall b_1, b_2 \in V_J^{\otimes d} \text{ with } h_T(b_1, b_2) \geq D_\gamma\varepsilon_T\}$$
$$\supset \{2c_\gamma\|b_1 - b_2\|_{\mu_0} \leq h_T(b_1, b_2) \leq \tfrac{1}{2}C_\gamma\|b_1 - b_2\|_{\mu_0}, \forall b_1, b_2 \in V_J^{\otimes d}\}$$

since $b_{0,J} \in V_J^{\otimes d}$. It thus suffices to lower bound the P_{b_0}-probability of the last event. For $C_\gamma > 2$ and $0 < c_\gamma < 1/2$, this probability equals

$$(25) \quad \begin{aligned} &P_{b_0}\left(4c_\gamma^2 - 1 \leq \frac{h_T^2(b_1, b_2)}{\|b_1 - b_2\|_{\mu_0}^2} - 1 \leq \frac{1}{4}C_\gamma^2 - 1, \ \forall b_1, b_2 \in V_J^{\otimes d}, b_1 \neq b_2\right) \\ &\geq 1 - P_{b_0}\left(\sup_{b_1, b_2 \in V_J^{\otimes d}: b_1 \neq b_2} \frac{|h_T^2(b_1, b_2) - \|b_1 - b_2\|_{\mu_0}^2|}{\|b_1 - b_2\|_{\mu_0}^2} > K_\gamma\right), \end{aligned}$$

where $K_\gamma = \min\{1 - 4c_\gamma^2, C_\gamma^2/4 - 1\} > 0$. Since $b_0 \in C^s, s > \max(d/2, 1)$, Lemma 4 with $x = \sqrt{2}M_0 2^{Jd/2} \to \infty$ and $M_0 > 1$ large enough yields

$$\begin{aligned} &P_{b_0}\left(\sup_{b_1, b_2 \in V_J^{\otimes d}: b_1 \neq b_2} \left|\frac{h_T^2(b_1, b_2) - \|b_1 - b_2\|_{\mu_0}^2}{\|b_1 - b_2\|_{\mu_0}^2}\right| \geq \frac{CM_0}{\sqrt{T}}2^{J[d+(d/2+\kappa-1)_+]}\right) \\ &\leq de^{(c_0' - M_0^2)2^{Jd}} \to 0, \end{aligned}$$

where $0 < \kappa < s - d/2 + 1$ (or $\kappa = 0$ if $d = 1$). Since $T^{-1/2}2^{J[d+(d/2+\kappa-1)_+]} \to 0$ as $T \to \infty$ for $a > \max(d-1, 1/2)$ and $\kappa > 0$ small enough, the right-hand side of (25) equals $1 - o_{P_{b_0}}(1)$ as $T \to \infty$. This verifies (24) for $C_\gamma > 2$, $0 < c_\gamma < 1/2$ and $D_\gamma > 0$ large enough, so that applying Theorem 7 yields posterior contraction rate ε_T in the Hellinger distance h_T.

We have shown above that for $C_\gamma > 2$ and $0 < c_\gamma < 1/2$,

$$P_{b_0}(2c_\gamma \|b_1 - b_2\|_{\mu_0} \le h_T(b_1, b_2) \le \tfrac{1}{2}C_\gamma \|b_1 - b_2\|_{\mu_0}, \forall b_1, b_2 \in V_J^{\otimes d}) \to 1.$$

Using again the bias bounds $h_T(b_0, b_{0,J}) \le r\varepsilon_T$ and $\|b_0 - b_{0,J}\|_{\mu_0} \le r\varepsilon_T$, with P_{b_0}-probability tending to one it holds that

$$\{b \in V_J^{\otimes d} : \|b - b_0\|_{\mu_0} \ge \tilde{M}_T \varepsilon_T\} \subseteq \{b \in V_J^{\otimes d} : h_T(b, b_0) \ge M_T \varepsilon_T\}$$

for $\tilde{M}_T = (M_T + r)/(2c_\gamma) + r$. Since the posterior probability of the last set tends to zero in P_{b_0}-probability, this completes the proof of Theorem 1.

3.4. *Proofs of Theorems 2-4.* The proof of Theorem 2 is based on combination of Theorem 1 – which allows an initial localisation of the posterior distribution in a neighbourhood contracting about b_0 in L^2-norm via (26) below – with the key Lemma 5(i) (which itself follows from 'quantitiative' semiparametric techniques developed in Section 4). Once Theorem 2 is proved one can refine Lemma 5 (see its Part (ii)) and apply it to the $\|\cdot\|_\infty$-localised posterior, from which one can derive Theorems 3 and 4.

3.4.1. *Localisation and a key lemma.* We will repeatedly use the following basic fact that allows to 'localise' the posterior distribution to sets $\mathcal{D}_T$ of high frequentist posterior probability: let $\mathcal{D}_T$ be any measurable set in the support of the prior satisfying $\Pi(\mathcal{D}_T | X^T) = 1 - o_{P_{b_0}}(1)$ as $T \to \infty$, let

$$\Pi^{\mathcal{D}_T}(\cdot) = \Pi(\cdot \cap \mathcal{D}_T)/\Pi(\mathcal{D}_T)$$

denote the prior conditioned to $\mathcal{D}_T$ and let $\Pi^{\mathcal{D}_T}(\cdot | X^T)$ denote the posterior distribution arising from prior $\Pi^{\mathcal{D}_T}$. By a standard inequality ([58], p. 142),

$$(26) \qquad \sup_{A \text{ measurable}} |\Pi(A | X^T) - \Pi^{\mathcal{D}_T}(A | X^T)| \le 2\Pi(\mathcal{D}_T^c | X^T) \to_{T\to\infty}^{P_{b_0}} 0.$$

The relevant choices D_T and $\bar{D}_T$ for $\mathcal{D}_T$, to be defined below, depend on a further choice $\Gamma_T \subset V_J^{\otimes d}$ of vector fields γ admitting envelopes

$$(27) \qquad |\Gamma_T|_2 \ge \sup_{\gamma \in \Gamma_T} \|\gamma\|_{L^2}, \qquad \sigma_{\Gamma_T} \ge \sup_{\gamma \in \Gamma_T} \|\gamma\|_{\mathbb{H}},$$

where the RKHS norm $\| \cdot \|_{\mathbb{H}}$ arises from (12) with $\sigma_l = 2^{-l(\alpha+d/2)}$. For any $M > 0$ and $\varepsilon_T = T^{-\frac{a \wedge s}{2a+d}} (\log T)$ as in Theorem 1, define

$$(28) \quad D_T = \left\{ b \in V_J^{\otimes d} : \|b - b_0\|_{L^2} \leq M_T \varepsilon_T, \ \sup_{\gamma \in \Gamma_T} |\langle b, \gamma \rangle_{\mathbb{H}}| \leq M\sqrt{T} \varepsilon_T \sigma_{\Gamma_T} \right\},$$

where $M_T \to \infty$ arbitrarily slowly; and for $\bar{M}_T = (\log T)^{\delta-1}, \delta > 5/2$, define

$$(29) \quad \bar{D}_T = \left\{ b \in V_J^{\otimes d} : \|b - b_0\|_{\infty} \leq \bar{M}_T \varepsilon_T, \ \sup_{\gamma \in \Gamma_T} |\langle b, \gamma \rangle_{\mathbb{H}}| \leq M\sqrt{T} \varepsilon_T \sigma_{\Gamma_T} \right\}.$$

For $\lambda \leq J, 1 \leq j \leq d$ and $a_\lambda > 0$ to be chosen, define the vector fields $\tilde{\Phi}_{\lambda,k,j} = (\tilde{\Phi}_{\lambda,k,j,1}, \ldots, \tilde{\Phi}_{\lambda,k,j,d}) : \mathbb{T}^d \to \mathbb{R}^d$ with

$$\tilde{\Phi}_{\lambda,k,j,i} = \begin{cases} a_\lambda P_{V_J}[\Phi_{\lambda,k}/\mu_0] & i = j, \\ 0 & i \neq j. \end{cases}$$

Thus $\tilde{\Phi}_{\lambda,k,j}$ is the vector field which projects $a_\lambda \Phi_{\lambda,k}/\mu_0$ onto V_J in the j-th coordinate and is uniformly zero on all other coordinates. Denote the collection of all such functions by

$$(30) \quad \Gamma_T = \{\tilde{\Phi}_{\lambda,k,j} : \lambda \leq J, k, 1 \leq j \leq d\} \subset V_J^{\otimes d}.$$

By Lemma 10 and for Γ_T as above we can take the envelopes from (27) as

$$(31) \quad |\Gamma_T|_2 = C(d, \mu_0) \max_{\lambda \leq J} a_\lambda, \quad \sigma_{\Gamma_T} = C(d, \mu_0, \Phi) 2^{J(\alpha+d/2)} \max_{\lambda \leq J} a_\lambda.$$

The following central lemma will be proved in Section 4.4 below.

LEMMA 5. *(i) Assume the conditions of Theorem 2 and let D_T be the set from (28) with Γ_T as in (30), envelope σ_{Γ_T} as in (31) and with*

$$a_\lambda = \begin{cases} 2^{\lambda d/2} 2^{-Jd/2} (\log T)^{-\eta} & \text{if } d \leq 4, \\ 2^{\lambda d/2} 2^{-J(d-2)} (\log T)^{-\eta} & \text{if } d \geq 5, \end{cases}$$

for any $\eta > 1$. Then $\Pi(D_T|X^T) = 1 - o_{P_{b_0}}(1)$ as $T \to \infty$. Moreover if $b \sim \Pi^{D_T}(\cdot|X^T)$ then for all $\lambda \leq J$, as $T \to \infty$,

$$(32) \quad E^{\Pi^{D_T}} \left[\max_{k,j} \sqrt{T} |\langle b_j - b_{0,j}, a_\lambda \Phi_{\lambda,k} \rangle_{L^2}| | X^T \right] = O_{P_{b_0}}(\sqrt{\lambda}),$$

$$(33) \quad E^{\Pi^{D_T}} \left[\max_{\lambda \leq J, k, j} \sqrt{T} |\langle b_j - b_{0,j}, a_\lambda \Phi_{\lambda,k} \rangle_{L^2}| | X^T \right] = O_{P_{b_0}}(\sqrt{J}).$$

(ii) Assume the conditions of Theorem 3 and let $\bar{D}_T$ be the set from (29) with Γ_T as in (30) with $a_\lambda = 1$ for all λ and envelope σ_{Γ_T} as in (31). Then $\Pi(\bar{D}_T|X^T) = 1 - o_{P_{b_0}}(1)$ as $T \to \infty$ and if $b \sim \Pi^{\bar{D}_T}(\cdot|X^T)$, then for all $\lambda \leq J$,

$$E^{\Pi^{\bar{D}_T}} \left[\max_{k,j} \sqrt{T} |\langle b_j - b_{0,j}, \Phi_{\lambda,k} \rangle_{L^2}| | X^T \right] = O_{P_{b_0}}(\sqrt{\lambda}).$$

3.4.2. *Proof of Theorem 2.* Take the set D_T from (28) with Γ_T, σ_{Γ_T} as in Lemma 5(i). Then by that Lemma and (26) with $\mathcal{D}_T = D_T$, it suffices to prove Theorem 2 for b drawn from the localised posterior distribution $\Pi^{D_T}(\cdot|X^T)$. Denote by P_{V_J}, $P_{V_J^{\otimes d}}$ the projection operators onto V_J, $V_J^{\otimes d}$, respectively. Setting

$$\tilde{\varepsilon}_T = (\log T)^\delta T^{-\frac{s\wedge[a-(d/2-2)_+]}{2a+d}}$$

and applying Markov's inequality,

$$\Pi^{D_T}\big(b : \sum_{j=1}^d \|b_j - b_{0,j}\|_\infty \geq \tilde{\varepsilon}_T\big|X^T\big) \leq \tilde{\varepsilon}_T^{-1} \sum_{j=1}^d E^{\Pi^{D_T}}[\|b_j - b_{0,j}\|_\infty|X^T]$$

(34)

$$\leq \tilde{\varepsilon}_T^{-1} \sum_{j=1}^d E^{\Pi^{D_T}}[\|b_j - P_{V_J}[b_{0,j}]\|_\infty|X^T] + \tilde{\varepsilon}_T^{-1} \sum_{j=1}^d \|P_{V_J}[b_{0,j}] - b_{0,j}\|_\infty.$$

Since $b_0 \in C^s$, the second sum is of order $O(2^{-Js}) = O(T^{-\frac{s}{2a+d}}) = o(\tilde{\varepsilon}_T)$ by standard results in approximation theory (cf. after (4) above). Suppose first that $d \leq 4$ and let $a_\lambda = 2^{\lambda d/2} 2^{-Jd/2} (\log T)^{-\eta}$ for some $\eta > 1$. The standard inequality $\sup_x \sum_k |\Phi_{\lambda,k}(x)| \lesssim 2^{\lambda d/2}$ (Section 4.3 in [26]) now gives

$$\begin{aligned}
\sum_{j=1}^d \|b_j - P_{V_J}[b_{0,j}]\|_\infty &= \sum_{j=1}^d \sup_x \Big|\sum_{\lambda \leq J}\sum_k \langle b_j - b_{0,j}, \Phi_{\lambda,k}\rangle_{L^2}\Phi_{\lambda,k}(x)\Big| \\
&\lesssim \sum_j \sum_{\lambda \leq J} \frac{2^{\lambda d/2}}{\sqrt{T}} \max_k \sqrt{T}|\langle b_j - b_{0,j}, \Phi_{\lambda,k}\rangle_{L^2}| \\
&= \frac{1}{\sqrt{T}} \sum_j \sum_{\lambda \leq J} 2^{\lambda d/2} a_\lambda^{-1} \max_k \sqrt{T}|\langle b_j - b_{0,j}, a_\lambda\Phi_{\lambda,k}\rangle_{L^2}| \\
&\lesssim \frac{J2^{Jd/2}(\log T)^\eta}{\sqrt{T}} \max_{\lambda \leq J,k,j} \sqrt{T}|\langle b_j - b_{0,j}, a_\lambda\Phi_{\lambda,k}\rangle_{L^2}|.
\end{aligned}$$

Taking posterior expectations in the last inequality, Lemma 5(i) implies that the first term in (34) is bounded by

$$\begin{aligned}
&\tilde{\varepsilon}_T^{-1} \frac{J2^{Jd/2}(\log T)^\eta}{\sqrt{T}} E^{\Pi^{D_T}}\Big[\max_{\lambda \leq J,k,j} \sqrt{T}|\langle b_j - b_{0,j}, a_\lambda\Phi_{\lambda,k}\rangle_{L^2}|\big|X^T\Big] \\
&= O_{P_{b_0}}\Big(\tilde{\varepsilon}_T^{-1} \frac{J^{3/2}2^{Jd/2}(\log T)^\eta}{\sqrt{T}}\Big) = O_{P_{b_0}}\Big((\log T)^{3/2+\eta-\delta}\Big).
\end{aligned}$$

Taking $\delta > 3/2 + \eta$ completes the proof when $d \leq 4$ since $\eta > 1$ was arbitrary. If $d > 4$, we set $a_\lambda = 2^{\lambda d/2} 2^{-J(d-2)} (\log T)^{-\eta}$ for $\eta > 1$ and use again Lemma 5(i) to obtain, as $T \to \infty$, the convergence to zero of

$$\tilde{\varepsilon}_T^{-1} \sum_j E^{\Pi^{D_T}}[\|b_j - P_{V_J} b_{0,j}\|_\infty | X^T] = O_{P_{b_0}}\Big(\frac{J^{\frac{3}{2}} 2^{J(d-2)} (\log T)^\eta}{\tilde{\varepsilon}_T \sqrt{T}}\Big) = o_{P_{b_0}}(1).$$

3.4.3. *Proof of Theorems 3 and 4.* Let $b \sim \Pi^{\bar{D}_T}(\cdot | X^T)$ conditionally on X^T, where $\bar{D}_T$ is the event from (29) with $\Gamma_T, \sigma_{\Gamma_T}$ chosen as in Lemma 5(ii). Then by that lemma and (26) with $\mathcal{D}_T = \bar{D}_T$, it suffices to prove Theorem 3 for $\Pi^{\bar{D}_T}(\cdot | X^T)$ in place of $\Pi(\cdot | X^T)$.

Denote the centred ball of radius r in $B_{1\infty}^\rho = B_{1\infty}^{\rho, \otimes d}$ by $\mathcal{B}^\rho(r)$ and let $\eta = (\eta_1, \ldots, \eta_d) \in \mathcal{B}^\rho(1)$. For projections

$$P_{V_J^{\otimes d}}[\eta/\mu_0] = (P_{V_J}[\eta_j/\mu_0] : j = 1, \ldots, d),$$

define the centring process $(\hat{G}_J(\eta) \equiv \langle \hat{G}_J, \eta \rangle_{L^2} : \eta \in \mathcal{B}^\rho(1))$ by

$$\langle \hat{G}_J, \eta \rangle_{L^2} = \sum_{j=1}^d \langle \hat{G}_{J,j}, \eta_j \rangle_{L^2} = \langle b_0, \eta \rangle_{L^2} + \frac{1}{T} \int_0^T P_{V_J^{\otimes d}}[\eta/\mu_0](X_t).dW_t,$$

where the notation $\langle \hat{G}_J, \cdot \rangle_{L^2}$ is justified by linearity of the stochastic integral. Next define stochastic processes

$$(Z_1(\eta) = \sqrt{T}(\langle b, \eta \rangle_{L^2} - \hat{G}_J(\eta)) : \eta \in \mathcal{B}^\rho(1)), \quad (Z_2(\eta) : \eta \in \mathcal{B}^\rho(1)),$$

where Z_2 has (cylindrical) law $\mathcal{N}_{b_0}$, and denote the (conditional) law of Z_1 by $\bar{\Pi}^{\bar{D}_T} = \bar{\Pi}^{\bar{D}_T}(\cdot | X^T)$. Both processes prescribe linear actions on $\mathcal{B}^\rho(1)$ – this is clear for Z_1 and follows also for Z_2 as explained before (39) below. The estimates that follow imply moreover that the Z_i define proper random variables in $(B_{1\infty}^\rho)^*$. For $\kappa \in \mathbb{N}$ to be chosen, define probability measures $\bar{\Pi}_\kappa^{\bar{D}_T}, \mathcal{N}_{b_0,\kappa}$ as the laws of the stochastic processes

$$(35) \qquad P_{(\kappa)}(Z_i) \equiv (Z_i(P_{V_\kappa^{\otimes d}}[\eta]) : \eta \in \mathcal{B}^\rho(1)), \quad i = 1, 2,$$

which, as projections, are defined on the same probability space as the Z_i's. Using the triangle inequality for the metric $\beta = \beta_{(B_{1\infty}^\rho)^*}$ we obtain

$$\beta(\bar{\Pi}^{\bar{D}_T}, \mathcal{N}_{b_0}) \leq \beta(\bar{\Pi}_\kappa^{\bar{D}_T}, \mathcal{N}_{b_0,\kappa}) + \beta(\bar{\Pi}^{\bar{D}_T}, \bar{\Pi}_\kappa^{\bar{D}_T}) + \beta(\mathcal{N}_{b_0}, \mathcal{N}_{b_0,\kappa})$$

$$= \beta_{V_\kappa^{\otimes d}}(\bar{\Pi}_\kappa^{\bar{D}_T}, \mathcal{N}_{b_0,\kappa}) + \sum_{i=1}^2 \sup_{\|F\|_{Lip} \leq 1} |E[F(Z_i) - F(P_{(\kappa)}(Z_i))]|$$

$$(36) \qquad \leq \beta_{V_\kappa^{\otimes d}}(\bar{\Pi}_\kappa^{\bar{D}_T}, \mathcal{N}_{b_0,\kappa}) + \sum_{i=1}^2 E\|Z_i - P_{(\kappa)}(Z_i)\|_{(B_{1\infty}^\rho)^*} = A + B + C.$$

For term B, we use Parseval's identity and the fact that (cf. after (4))

$$(37) \qquad \|\eta\|_{B_{1\infty}^{\rho}} \leq 1 \Rightarrow \sum_{j} \sum_{r} |\langle \eta_j, \Phi_{l,r} \rangle_{L^2}| \lesssim 2^{-l(\rho-d/2)} \quad \forall l$$

to obtain, with $E = E^{\Pi^{\bar{D}_T}}[\cdot|X^T]$,

$$E\|Z_1 - P_{(\kappa)}(Z_1)\|_{(B_{1\infty}^{\rho})^*} = E \sup_{\|\eta\|_{B_{1\infty}^{\rho}} \leq 1} \sqrt{T} \left| \langle b - \hat{G}_J, \eta - P_{V_\kappa^{\otimes d}}[\eta] \rangle_{L^2} \right|$$

$$\lesssim \sum_{\kappa<\lambda} 2^{-\lambda(\rho-d/2)} E \max_{k,j} \sqrt{T} \left| \langle b_j - \hat{G}_{J,j}, \Phi_{\lambda,k} \rangle_{L^2} \right|$$

$$(38) \quad \lesssim \sum_{\kappa<\lambda} 2^{-\lambda(\rho-d/2)} E \max_{k,j} \sqrt{T} |\langle b_j - b_{0,j}, \Phi_{\lambda,k} \rangle_{L^2}|$$

$$+ \sum_{\kappa<\lambda} 2^{-\lambda(\rho-d/2)} \max_{k,j} \sqrt{T} \left| \langle \hat{G}_{J,j} - b_{0,j}, \Phi_{\lambda,k} \rangle_{L^2} \right|.$$

By Lemma 5(ii) and the usual decay bound for wavelet coefficients of $b_0 \in C^s$, the first sum is bounded in P_{b_0}-probability by

$$\sum_{\kappa<\lambda\leq J,j} 2^{-\lambda(\rho-d/2)} \sqrt{\lambda} + \sqrt{T} \sum_{\lambda>J} 2^{-\lambda(\rho+s)} = o(1)$$

as $T \to \infty$ and $\kappa \to \infty$, since $\rho > d/2$ and $s \geq a$. To deal with the second sum, note that by definition

$$\sqrt{T}\langle \hat{G}_{J,j} - b_{0,j}, \Phi_{\lambda,k} \rangle_{L^2} = \frac{1}{\sqrt{T}} \int_0^T P_{V_J}[\Phi_{\lambda,k}/\mu_0](X_t)dW_t^j.$$

Arguing as after (57) below, Bernstein's inequality (15) implies that these variables are subgaussian under P_{b_0} with variance proxy bounded by

$$\|P_{V_J}[\Phi_{\lambda,k}/\mu_0]\|_{\mu_0}^2 + \frac{1}{T} \int_0^T |P_{V_J}[\Phi_{\lambda,k}/\mu_0](X_t)|^2 dt - \|P_{V_J}[\Phi_{\lambda,k}/\mu_0]\|_{\mu_0}^2.$$

The first quantity is bounded by $\|\Phi_{\lambda,k}\|_{L^2}\|1/\mu_0\|_\infty \lesssim 1$ whereas Proposition 2(ii) implies that the second quantity is $O_{P_{b_0}}(\tilde{R}_T) = O_{P_{b_0}}(1)$ uniformly over λ, k for our choice of J, s. Thus by the usual subgaussian maximal inequality (Lemma 2.3.4 in [26]), the last term in (38) is $O_{P_{b_0}}(\sum_{\lambda>J} 2^{-\lambda(\rho-d/2)}\sqrt{\lambda}) = o_{P_{b_0}}(1)$ for $\rho > d/2$, so that the last sum in (38) is $o_{P_{b_0}}(1)$ as $\kappa, T \to \infty$.

For term C, we first note that $\mathcal{N}_{b_0}$ defines a tight Gaussian probability measure in the space of bounded functions on $\mathcal{B}^\rho(1)$ (using Theorem 2.3.7, Proposition 2.1.5 and (4.184) in [26]), and arguing as in Theorem 3.7.28 in

[26] one shows further that $\mathcal{N}_{b_0}$ extends to a Gaussian probability measure on $(B_{1\infty}^\rho)^*$. In particular, a version of Z_2 exists that acts linearly on $\mathcal{B}^\rho(1)$. Define $\Phi_{\lambda,k,j} = (0,\ldots,0,\Phi_{\lambda,k},0,\ldots,0) : \mathbb{T}^d \to \mathbb{R}^d$, where the non-zero coordinate occurs in the jth entry. Then, recalling the definition (35) of $P_{(\kappa)}(Z_2)$ and using again (37) and the standard subgaussian maximal inequality, now for the variables $(Z_2(\Phi_{\lambda,k,j}) \sim N(0, \|\Phi_{\lambda,k}\|_{1/\mu_0}^2))$,

$$E\|Z_2 - P_{(\kappa)}(Z_2)\|_{(B_{1\infty}^\rho)^*} = E \sup_{\|\eta\|_{B_{1\infty}^\rho} \leq 1} \left| Z_2(\eta - P_{V_\kappa^{\otimes d}}[\eta]) \right|$$

$$(39) \quad \lesssim \sum_{\kappa < \lambda} 2^{-\lambda(\rho-d/2)} E \max_{k,j} |Z_2(\Phi_{\lambda,k,j})| \lesssim \sum_{\kappa < \lambda} 2^{-\lambda(\rho-d/2)}\sqrt{\lambda} =_{\kappa \to \infty} o(1).$$

For term A we show convergence of the finite-dimensional distributions by the semi-parametric techniques from Section 4: consider the basis $(\Phi_{\lambda,k,j} : k,j; \lambda \leq \kappa)$ of $V_\kappa^{\otimes d}$ for κ fixed. We apply Proposition 2(iii) with $\gamma = P_{V_J^{\otimes d}}[\Phi_{\lambda,k,j}/\mu_0]$, then Lemma 8(ii) and the third part of Lemma 9 to obtain

$$E^{\Pi^{\bar{D}_T}} \left[e^{u\sqrt{T}(\langle b-b_0,\Phi_{\lambda,k,j}/\mu_0\rangle_{\mu_0})-(u/\sqrt{T}) \int_0^T P_{V_J^{\otimes d}}[\Phi_{\lambda,k,j}/\mu_0](X_t).dW_t} \Big| X^T \right]$$

$$= C_T \exp\left\{ \frac{u^2}{2} \int_{\mathbb{T}^d} \|P_{V_J^{\otimes d}}[\Phi_{\lambda,k,j}/\mu_0]\|^2 d\mu_0 \right\},$$

where we can take the envelopes in Lemma 8(ii) to satisfy $|\Gamma_T|_2 \lesssim 1, \varepsilon_T \sigma_{\Gamma_T} = o(1)$ as in the proof of Lemma 5, and where $C_T = 1 + o_{P_{b_0}}(1)$ as $T \to \infty$ for fixed $u \in \mathbb{R}$. We also have $\|P_{V_J^{\otimes d}}[\Phi_{\lambda,k,j}/\mu_0]\|_{\mu_0} \to \|\Phi_{\lambda,k}/\mu_0\|_{\mu_0} = \|\Phi_{\lambda,k}\|_{1/\mu_0}$ as $J \to \infty$ since P_{V_J} are L^2-projections. The same is true if $\Phi_{\lambda,k,j}$ is replaced by arbitrary finite linear combinations $\sum_j \sum_{\lambda \leq \kappa,k} a_{\lambda,k,j}\Phi_{\lambda,k,j}$, κ fixed, and thus by Proposition 29 in [35] we deduce joint weak convergence of the finite-dimensional distributions. In particular, for every fixed $\kappa \in \mathbb{N}$,

$$(40) \qquad\qquad \beta_{V_\kappa^{\otimes d}}(\bar{\Pi}_\kappa^{\bar{D}_T}, \mathcal{N}_{b_0,\kappa}) \to^{P_{b_0}} 0 \quad \text{as } T \to \infty.$$

Combining the above bounds, given $\epsilon' > 0$ we can choose $\kappa = \kappa(\epsilon')$ large enough so that by virtue of the bounds following (38) and (39), the terms B, C in (36) are each less than $\epsilon'/3$ (for B on an event of P_{b_0}-probability as close to one as desired). Then applying (40) for this choice of κ, we can also make the term A less than $\epsilon'/3$ for T large enough and with P_{b_0}-probability as close to one as desired, completing the proof of Theorem 3 with $\bar{\Pi}_T$ replacing $\tilde{\Pi}_T$, that is, with centring equal to $\hat{G}_J$.

That $\hat{G}_J$ can be replaced by the posterior mean in Theorem 3 is the last step: since the laws $\bar{\Pi}_T$ form a sequence of (conditionally on X^T) Gaussian

distributions on $(B_{1\infty}^\rho)^*$ that converge weakly (in probability), we also have convergence of moments of that sequence (in probability) in $(B_{1\infty}^\rho)^*$, using Exercise 2.1.4 in [26] and arguing as in the proof of Theorem 2.7 in [34]. Since $\mathcal{N}_{b_0}$ has Bochner-mean zero we deduce that

$$(41) \qquad \sqrt{T}(E^{\Pi_T}[b|X^T] - \hat{G}_J) = o_{P_{b_0}}(1) \text{ in } (B_{1\infty}^\rho)^*.$$

This concludes the proof of Theorem 3. Theorem 4 now follows from (41) and asymptotic normality of the $\sqrt{T}(\hat{G}_J(\eta) - \langle b_0, \eta\rangle)$ variables in the space $(B_{1\infty}^\rho)^*$, proved as follows: if we denote by ν_T the law of the latter variables, then arguing just as in (36) we have

$$(42) \quad \beta_{(B_{1\infty}^\rho)^*}(\nu_T, \mathcal{N}_{b_0}) \le \beta_{V_\kappa^{\otimes d}}(\nu_{T,\kappa}, \mathcal{N}_{b_0,\kappa}) + \sum_{i=1}^{2} E\|\tilde{Z}_i - P_{(\kappa)}(\tilde{Z}_i)\|_{(B_{1\infty}^\rho)^*}$$

where $\tilde{Z}_2 =^{\mathcal{L}} Z_2$ from above, $\tilde{Z}_1$ has law ν_T and $P_{(\kappa)}(\tilde{Z}_i)$ refers to the projected processes as in (35). The first term on the right hand side converges to zero, for every fixed κ, by applying the martingale central limit theorem as in (43) to $(1/\sqrt{T})\int_0^T (\Phi_{\lambda,r,j}/\mu_0)(X_t).dW_t$, $\lambda \le \kappa$ fixed, and using (15) to show that the term

$$\frac{1}{\sqrt{T}}\int_0^T [P_{V_J^{\otimes d}}[\Phi_{\lambda,r,j}/\mu_0] - \Phi_{l,r,j}/\mu_0](X_t).dW_t = o_{P_{b_0}}(1)$$

in view of $\|P_{V_J^{\otimes d}}[\Phi_{\lambda,r,j}/\mu_0] - \Phi_{\lambda,r,j}/\mu_0\|_\infty \to 0$ as $J \to \infty$ for fixed $\lambda \le \kappa$. The third term in (42) was bounded as $o(1)$ for $\kappa \to \infty$ in (39), and the second term also converges to zero as $\kappa \to \infty$ by the arguments below (38). Thus choosing κ large enough but fixed, and letting $T \to \infty$, Theorem 4 follows since $\beta_{(B_{1\infty}^\rho)^*}$ metrises weak convergence.

4. Bayesian semi-parametric techniques for diffusions. We show here how techniques developed in [9, 12, 13] for Gaussian white noise and i.i.d. density models extend to the multi-dimensional diffusion case.

4.1. *Asymptotic expansion of the posterior Laplace transform.* We start with the following basic 'LAN expansion' for ℓ_T from (8).

LEMMA 6. *Suppose* $b_0 \in C^{(d/2+\kappa)\vee 1}(\mathbb{T}^d), h \in H^{d/2+\kappa}(\mathbb{T}^d), \kappa > 0$. *Then*

$$\ell_T(b_0 + h/\sqrt{T}) - \ell_T(b_0) = W_T(h) - \frac{1}{2T}\int_0^T \|h(X_t)\|^2 dt,$$

where, as $T \to \infty$ *and under* P_{b_0},

$$W_T(h) \equiv \frac{1}{\sqrt{T}}\int_0^T h(X_t).dW_t \to^d N(0, \|h\|_{\mu_0}^2), \ \frac{1}{T}\int_0^T \|h(X_t)\|^2 dt \to^P \|h\|_{\mu_0}^2.$$

PROOF. Using (1) with $b = b_0$ and (8),

$$\ell_T(b_0 + h/\sqrt{T}) - \ell_T(b_0) = \frac{1}{\sqrt{T}} \int_0^T h(X_t).dX_t - \frac{1}{\sqrt{T}} \int_0^T b_0(X_t).h(X_t)dt$$
$$- \frac{1}{2T} \int_0^T \|h(X_t)\|^2 dt$$
$$= \frac{1}{\sqrt{T}} \int_0^T h(X_t).dW_t - \frac{1}{2T} \int_0^T \|h(X_t)\|^2 dt.$$

Since $x \mapsto \|x\|^2$ is a smooth map, the function $f_h(x) = \|h(x)\|^2 - \|h\|^2_{\mu_0} \in L^2_{\mu_0}(\mathbb{T}^d) \cap H^{d/2+\kappa}(\mathbb{T}^d) \subset C(\mathbb{T}^d)$. In particular, $LL^{-1}[f_h] = f_h$ where $L^{-1} = L^{-1}_{b_0}$ is the inverse of the generator L constructed in Lemma 11 below. Moreover, by that Lemma and the Sobolev embedding theorem, $L^{-1}[f_h] \in H^{d/2+\kappa+2} \subset C^2$. By Itô's formula (Theorem 39.3 in [5]),

$$\int_0^T f_h(X_t)dt = \int_0^T LL^{-1}[f_h](X_t)dt$$
$$= L^{-1}[f_h](X_T) - L^{-1}[f_h](X_0) - \int_0^T \nabla L^{-1}[f_h](X_t).dW_t.$$

Since $L^{-1}[f_h] \in C^2$, the first term on the right-hand side is $O(1)$, while the second term satisfies

$$E_{b_0}\Big[\int_0^T \nabla L^{-1}[f_h](X_t).dW_t\Big]^2 = E_{b_0} \int_0^T \|\nabla L^{-1}[f_h](X_t)\|^2 dt \lesssim T\|L^{-1}[f_h]\|^2_{C^1}$$

so that $T^{-1} \int_0^T f_h(X_t)dt \to 0$ in $L^2(P_{b_0})$, and hence also in P_{b_0}-probability. Next, set $M_T^h = \int_0^T h(X_t).dW_t$, so that $(M_T^h)_{T \geq 0}$ is a continuous local L^2-martingale with quadratic variation $[M^h]_T = \int_0^T \|h(X_t)\|^2 dt$. By what precedes $T^{-1}[M^h]_T - \|h\|^2_{\mu_0} = T^{-1} \int_0^T f_h(X_t)dt \to 0$ in $L^2(P_{b_0})$ as $T \to \infty$. Applying the martingale central limit theorem (p.338f. in [20]),

$$(43) \qquad\qquad\qquad T^{-1/2} M_T^h \to^d N(0, \|h\|^2_{\mu_0})$$

as $T \to \infty$, completing the proof. $\square$

A key result for our proofs is the following expansion of the Laplace transform of the posterior distribution $\Pi^{\mathcal{D}_T}(\cdot|X^T)$ arising from a 'localised' prior $\Pi^{\mathcal{D}_T}$ for the choices of $\mathcal{D}_T$ from Lemma 5.

PROPOSITION 2.   *Suppose $b_0 \in C^s(\mathbb{T}^d) \cap H^s(\mathbb{T}^d)$, $s > \max(d/2, 1)$, and consider the Gaussian prior Π from (11) with $2^J \approx T^{\frac{1}{2a+d}}$ and $\sigma_l = 2^{-l(\alpha+d/2)}$ for $a > \max(d-1, 1/2)$ and $0 \le \alpha \le a$. Let $\Gamma_T \subset V_J^{\otimes d}$ be a set of functions admitting envelopes as in (27) and let $D_T \subset V_J^{\otimes d}$ denote the set (28) for this choice of Γ_T and arbitrary $M > 0$. For $u \in \mathbb{R}$, $b \in V_J^{\otimes d}$ and fixed $\gamma \in \Gamma_T$, define the perturbations*

$$(44) \qquad\qquad b_u = b_u(T, \gamma) = b - \frac{u}{\sqrt{T}}\gamma \in V_J^{\otimes d}.$$

For any measurable function $G : L^2(\mathbb{T}^d) \to \mathbb{R}$, write

$$(45) \qquad E^{\Pi^{D_T}}[e^{u\sqrt{T}G(b)}|X^T] = e^{\Lambda_T(u)} \frac{\int_{D_T} e^{S_T(b)+\ell_T(b_u)}d\Pi(b)}{\int_{D_T} e^{\ell_T(b)}d\Pi(b)},$$

for some Λ_T to be determined and where

$$S_T(b) = u\sqrt{T}\left(G(b) - \langle b - b_0, \gamma\rangle_{\mu_0}\right).$$

(i) If for some $\kappa > 0$ (or $\kappa = 0$ if $d = 1$),

$$R_T := 2^{J[d+(d/2+\kappa-1)_+]}M_T\varepsilon_T|\Gamma_T|_2\left(1 + \sqrt{\log(1/(M_T\varepsilon_T))} + \sqrt{\log(1/|\Gamma_T|_2)}\right)$$

satisfies $R_T \to 0$ as $T \to \infty$, then we can take

$$\Lambda_T(u) = \frac{u}{\sqrt{T}}\int_0^T \gamma(X_t).dW_t + \frac{u^2}{2T}\int_0^T \|\gamma(X_t)\|^2 dt + ur_T, \ u \in \mathbb{R},$$

in (45) with $r_T = O_{P_{b_0}}(R_T) = o_{P_{b_0}}(1)$ uniformly over $\gamma \in \Gamma_T$.
(ii) Furthermore,

$$E_{b_0} \sup_{\gamma \in \Gamma_T}\left|\frac{1}{T}\int_0^T \|\gamma(X_t)\|^2 dt - \|\gamma\|_{\mu_0}^2\right| \lesssim \tilde{R}_T,$$

where

$$\tilde{R}_T := T^{-1/2}2^{J[d+(d/2+\kappa-1)_+]}|\Gamma_T|_2^2\left(1 + \sqrt{\log(1/|\Gamma_T|_2)}\right)$$

for any $\kappa > 0$ ($\kappa = 0$ if $d = 1$). In particular, if both $R_T, \tilde{R}_T \to 0$, then we can take

$$\Lambda_T(u) = \frac{u}{\sqrt{T}}\int_0^T \gamma(X_t).dW_t + \frac{u^2}{2}\|\gamma\|_{\mu_0}^2 + ur_T + u^2\tilde{r}_T, \ u \in \mathbb{R},$$

in (45) with $r_T = O_{P_{b_0}}(R_T) = o_{P_{b_0}}(1)$ and $\tilde{r}_T = O_{P_{b_0}}(\tilde{R}_T) = o_{P_{b_0}}(1)$ uniformly over $\gamma \in \Gamma_T$.
(iii) Parts (i) and (ii) remain true if D_T is replaced by $\bar{D}_T$ from (29) and if M_T is replaced by $\bar{M}_T$ in the definition of R_T.

PROOF. (i) For $\gamma = (\gamma_1, \ldots, \gamma_d) \in \Gamma_T$,

$$E^{\Pi^{D_T}}[e^{u\sqrt{T}G(b)}|X^T] = E^{\Pi^{D_T}}[e^{S_T(b)+u\sqrt{T}\langle b-b_0,\gamma\rangle_{\mu_0}}|X^T]$$
$$= Z_T^{-1} \int_{D_T} e^{S_T(b)+u\sqrt{T}\langle b-b_0,\gamma\rangle_{\mu_0}+\ell_T(b_u)+\ell_T(b)-\ell_T(b_u)} d\Pi(b),$$

with $Z_T = \int_{D_T} e^{\ell_T(b)} d\Pi(b)$. Define the empirical process

$$\mathbb{G}_T[h] = \sqrt{T} \left(\frac{1}{T} \int_0^T h(X_t)dt - \int_{\mathbb{T}^d} h d\mu_0 \right), \quad h \in L^2(\mathbb{T}^d).$$

Using the LAN expansion from Lemma 6,

$$\ell_T(b) - \ell_T(b_u) = \frac{u}{\sqrt{T}} \int_0^T \gamma(X_t).dW_t - \frac{u}{\sqrt{T}} \int_0^T [b(X_t) - b_0(X_t)].\gamma(X_t)dt$$
$$+ \frac{u^2}{2T} \int_0^T \|\gamma(X_t)\|^2 dt$$
$$= \frac{u}{\sqrt{T}} \int_0^T \gamma(X_t).dW_t - u\mathbb{G}_T[(b-b_0).\gamma] - u\sqrt{T}\langle b-b_0,\gamma\rangle_{\mu_0}$$
$$+ \frac{u^2}{2T} \int_0^T \|\gamma(X_t)\|^2 dt.$$

The Laplace transform from the first equation of this proof therefore equals

$$e^{\frac{u}{\sqrt{T}} \int_0^T \gamma(X_t).dW_t + \frac{u^2}{2T} \int_0^T \|\gamma(X_t)\|^2 dt} Z_T^{-1} \int_{D_T} e^{-u\mathbb{G}_T[(b-b_0).\gamma]} e^{S_T(b)+\ell_T(b_u)} d\Pi(b).$$

We use Lemma 1 to control the empirical process term uniformly over $b \in D_T$, $\gamma \in \Gamma_T$. Set

$$\mathcal{F}_T = \left\{ f_{b,\gamma} := (b-b_0).\gamma - \int_{\mathbb{T}^d} (b-b_0).\gamma d\mu_0 : b \in D_T, \gamma \in \Gamma_T \right\},$$

which is a subset of $L^2_{\mu_0}(\mathbb{T}^d) \cap H^{d/2+\kappa}(\mathbb{T}^d)$ for $0 < \kappa < s - d/2$ since $\Gamma_T \subset V_J^{\otimes d} \subset H^p$ for any $p \leq S$. Suppose $d \geq 2$. Lemma 3 with $p = d/2 + \kappa - 1$

gives that for any $0 < \kappa < s - d/2 + 1$, $b, \bar{b} \in D_T$ and $\gamma, \bar{\gamma} \in \Gamma_T$,

$$d_L(f_{b,\gamma}, f_{\bar{b},\bar{\gamma}}) \lesssim \|f_{b,\gamma} - f_{\bar{b},\bar{\gamma}}\|_{H^{d/2+\kappa-1}}$$

$$\leq \sum_{j=1}^{d} \left\| (b_j - \bar{b}_j)\gamma_j + (\bar{b}_j - b_{0,j})(\gamma_j - \bar{\gamma}_j) \right.$$

$$\left. - \langle b_j - \bar{b}_j, \gamma_j \rangle_{\mu_0} - \langle \bar{b}_j - b_{0,j}, \gamma_j - \bar{\gamma}_j \rangle_{\mu_0} \right\|_{H^{d/2+\kappa-1}}$$

$$\lesssim \sum_{j=1}^{d} 2^{J(d+\kappa-1)} \|b_j - \bar{b}_j\|_{L^2} \|\gamma_j\|_{L^2}$$

$$+ \sum_{j=1}^{d} \|(\bar{b}_j - P_{V_J} b_{0,j})(\gamma_j - \bar{\gamma}_j) - \langle \bar{b}_j - P_{V_J} b_{0,j}, \gamma_j - \bar{\gamma}_j \rangle_{\mu_0}\|_{H^{d/2+\kappa-1}}$$

$$+ \sum_{j=1}^{d} \|(P_{V_J} b_{0,j} - b_{0,j})(\gamma_j - \bar{\gamma}_j) - \langle P_{V_J} b_{0,j} - b_{0,j}, \gamma_j - \bar{\gamma}_j \rangle_{\mu_0}\|_{H^{d/2+\kappa-1}}.$$

The first sum above is bounded by $C 2^{J(d+\kappa-1)} |\Gamma_T|_2 \|b - \bar{b}\|_{L^2}$, while by Lemma 3 the second sum is bounded by $C \sum_{j=1}^{d} 2^{J(d+\kappa-1)} M_T \varepsilon_T \|\gamma - \bar{\gamma}\|_{L^2}$. Using Lemma 2, that $b_0 \in C^s \cap H^s$ for $s > d/2$ and (16)-(17), the third sum is bounded by

$$\sum_{j=1}^{d} \|P_{V_J} b_{0,j} - b_{0,j}\|_{L^\infty} \|\gamma_j - \bar{\gamma}_j\|_{H^{d/2+\kappa-1}} + \|P_{V_J} b_{0,j} - b_{0,j}\|_{H^{d/2+\kappa-1}} \|\gamma_j - \bar{\gamma}_j\|_{L^\infty}$$

$$\lesssim \sum_{j=1}^{d} \left(2^{-Js} 2^{J(d/2+\kappa-1)} \|\gamma_j - \bar{\gamma}_j\|_{L^2} + 2^{J(d+\kappa-1-s)} \|\gamma_j - \bar{\gamma}_j\|_{L^2} \right)$$

$$\lesssim 2^{J(d+\kappa-1)} M_T \varepsilon_T \|\gamma - \bar{\gamma}\|_{L^2},$$

using again that $\Gamma_T \subset V_J^{\otimes d}$. Summarizing,

$$d_L(f_{b,\gamma}, f_{\bar{b},\bar{\gamma}}) \lesssim 2^{J(d+\kappa-1)} (|\Gamma_T|_2 \|b - \bar{b}\|_{L^2} + M_T \varepsilon_T \|\gamma - \bar{\gamma}\|_{L^2}).$$

In particular, $\mathcal{F}_T$ has d_L-diameter $D_{\mathcal{F}_T} \lesssim 2^{J(d+\kappa-1)} M_T \varepsilon_T |\Gamma_T|_2 = o(R_T) = o(1)$. Since $D_T, \Gamma_T \subset (V_J^{\otimes d}, \|\cdot\|_{L^2})$ where $v_J = dim(V_J) = O(2^{Jd})$, applying Proposition 4.3.34 of [26] yields

$$N(\mathcal{F}_T, d_L, \tau)$$

$$\leq N(D_T, c2^{J(d+\kappa-1)} |\Gamma_T|_2 \|\cdot\|_{L^2}, \tau/2) N(\Gamma_T, c2^{J(d+\kappa-1)} M_T \varepsilon_T \|\cdot\|_{L^2}, \tau/2)$$

$$\leq (C 2^{J(d+\kappa-1)} |\Gamma_T|_2 / \tau)^{dv_J} (C 2^{J(d+\kappa-1)} M_T \varepsilon_T / \tau)^{dv_J}$$

for some $c, C > 0$. Recall the inequality

$$\int_0^a \sqrt{\log(A/x)} dx \leq \frac{2 \log A}{2 \log A - 1} a \sqrt{\log(A/a)} \leq 4a\sqrt{\log(A/a)}$$

for any $A \geq 2$ and $0 < a \leq 1$ (p. 190 of [26]). Using the last two displays and that $D_{\mathcal{F}_T} \to 0$, $\int_0^{D_{\mathcal{F}_T}} \sqrt{\log 2N(\mathcal{F}_T, d_L, \tau)} d\tau$ is bounded by a multiple of

$$v_J^{1/2} \int_0^{D_{\mathcal{F}_T}} \left[\sqrt{\frac{\log([C2^{J(d+\kappa-1)}|\Gamma_T|_2] \vee 2)}{\tau}} + \sqrt{\frac{\log([C2^{J(d+\kappa-1)}M_T\varepsilon_T] \vee 2)}{\tau}} \right] d\tau$$

$$\lesssim 2^{Jd/2} D_{\mathcal{F}_T} \sqrt{\log([C2^{J(d+\kappa-1)}|\Gamma_T|_2] \vee 2/D_{\mathcal{F}_T})}$$

$$+ 2^{Jd/2} D_{\mathcal{F}_T} \sqrt{\log([C2^{J(d+\kappa-1)}M_T\varepsilon_T] \vee 2/D_{\mathcal{F}_T})}.$$

Taking $D_{\mathcal{F}_T} \approx 2^{J(d+\kappa-1)}M_T\varepsilon_T|\Gamma_T|_2$ for $\kappa > 0$ arbitrarily small, one can therefore bound the quantity $J(\mathcal{F}_T, d_L, D_{\mathcal{F}_T})$ in Lemma 1 by

$$2^{J(3d/2+\kappa-1)}M_T\varepsilon_T|\Gamma_T|_2(1 + \sqrt{\log(1/(M_T\varepsilon_T))} + \sqrt{\log(1/|\Gamma_T|_2)}) = R_T.$$

Using the Sobolev embedding theorem, Lemma 11, Lemma 3 and similar computations to the above, we see that

$$\sup_{f_{b,\gamma} \in \mathcal{F}_T} \|L^{-1}[f_{b,\gamma}]\|_\infty \lesssim \sup_{f_{b,\gamma} \in \mathcal{F}_T} \|f_{b,\gamma}\|_{H^{(d/2+\kappa-2)_+}} \lesssim 2^{J[d/2+(d/2+\kappa-2)_+]}M_T\varepsilon_T|\Gamma_T|_2$$

is also $o(R_T)$. Substituting these bounds into Lemma 1 yields

$$E_{b_0} \sup_{b \in D_T, \gamma \in \Gamma_T} |\mathbb{G}_T[(b-b_0).\gamma]| \lesssim R_T \to 0,$$

proving the first statement. The case $d = 1$ is proved similarly, using instead the simpler bound

$$d_L(f_{b,\gamma}, f_{\bar{b},\bar{\gamma}}) \lesssim 2^{J/2}|\Gamma_T|_2\|b - \bar{b}\|_{L^2} + 2^{J/2}M_T\varepsilon_T\|\gamma - \bar{\gamma}\|_{L^2}.$$

(ii) Since $x \mapsto \|x\|^2$ is a smooth map, the function $g_\gamma(x) = \|\gamma(x)\|^2 - \int_{\mathbb{T}^d} \|\gamma(y)\|^2 d\mu_0(y) \in L^2_{\mu_0}(\mathbb{T}^d) \cap H^{d/2+\kappa}$ for $\kappa > 0$. Since $\gamma \in V_J^{\otimes d}$, Lemma 3 with $p = (d/2+\kappa-1)_+$ gives that for any $\kappa > 0$ small enough and $\gamma, \bar{\gamma} \in \Gamma_T$,

$$d_L(g_\gamma, g_{\bar{\gamma}}) \lesssim \|g_\gamma - g_{\bar{\gamma}}\|_{H^{(d/2+\kappa-1)_+}}$$

$$\lesssim \sum_{j=1}^d \|\gamma_j^2 - \bar{\gamma}_j^2 - \int_{\mathbb{T}^d}(\gamma_j^2 - \bar{\gamma}_j^2)d\mu_0\|_{H^{(d/2+\kappa-1)_+}}$$

$$\lesssim \sum_{j=1}^d 2^{J[d/2+(d/2+\kappa-1)_+]}\|\gamma_j - \bar{\gamma}_j\|_{L^2}\|\gamma_j + \bar{\gamma}_j\|_{L^2}$$

$$\lesssim 2^{J[d/2+(d/2+\kappa-1)_+]}|\Gamma_T|_2\|\gamma - \bar{\gamma}\|_{L^2}.$$

In particular, $\mathcal{G}_T = \{g_\gamma : \gamma \in \Gamma_T\} \cup \{0\}$ has d_L-diameter

$$D_{\mathcal{G}_T} \lesssim 2^{J[d/2+(d/2+\kappa-1)_+]} |\Gamma_T|_2^2.$$

Using the same arguments as above, one deduces

$$N(\mathcal{G}_T, d_L, \tau) \leq N(\Gamma_T, 2^{J[d/2+(d/2+\kappa-1)_+]} |\Gamma_T|_2 \| \cdot \|_2, \tau)$$
$$\leq (C 2^{J[d/2+(d/2+\kappa-1)_+]} |\Gamma_T|_2 / \tau)^{dv_J}$$

and hence

$$J(\mathcal{G}_T, d_L, D_{\mathcal{G}_T}) \lesssim 2^{J[d+(d/2+\kappa-1)_+]} |\Gamma_T|_2^2 \left(1 + \sqrt{\log(1/|\Gamma_T|_2)}\right) = \sqrt{T} \tilde{R}_T.$$

In exactly the same way,

$$\sup_{g_\gamma \in \mathcal{G}_T} \|L^{-1}[g_\gamma]\|_\infty \lesssim 2^{J[d/2+(d/2+\kappa-2)_+]} \|\gamma\|_{L^2}^2 \leq 2^{J[d/2+(d/2+\kappa-2)_+]} |\Gamma_T|_2^2.$$

Applying Lemma 1 thus gives

$$E_{b_0} \sup_{\gamma \in \Gamma_T} \left| \frac{1}{T} \int_0^T \|\gamma(X_t)\|^2 dt - \int_{\mathbb{T}^d} \|\gamma(x)\|^2 d\mu_0(x) \right|$$
$$= \frac{1}{\sqrt{T}} E_{b_0} \sup_{g_\gamma \in \mathcal{G}_T} |\mathbb{G}_T(g_\gamma)| \lesssim \frac{1}{\sqrt{T}} 2^{J[d+(d/2+\kappa-1)_+]} |\Gamma_T|_2^2 \left(1 + \sqrt{\log(1/|\Gamma_T|_2)}\right)$$

which equals $\tilde{R}_T$. Finally, the proof of Part (iii) follows in the same way, using that $\| \cdot \|_{L^2} \leq \| \cdot \|_\infty$ and replacing M_T by $\bar{M}_T$. $\square$

4.2. *Change of measure.* In this section, let $\Pi = \Pi_T$ be the prior from (11). Using the lower bound for the small-ball probability (23) in the proof of Theorem 1, the proof of the following lemma is similar to the one of Theorem 8.20 in [22], and hence omitted.

LEMMA 7. *Suppose $b_0 \in C^s$ for $s > 0$. Then there exists a finite constant $C = C(b_0) > 0$ such that if B_T are measurable sets satisfying $\Pi_T(B_T) = o(e^{-CT\varepsilon_T^2})$ for $\varepsilon_T = T^{-\frac{a \wedge s}{2a+d}}(\log T)$, then $E_{b_0} \Pi_T(B_T | X^T) \to 0$.*

We now bound the ratio of Gaussian integrals from (45) in Proposition 2.

LEMMA 8. *(i) Suppose $b_0 \in C^s(\mathbb{T}^d) \cap H^s(\mathbb{T}^d)$ for some $s > \max(d/2, 1)$. Let $2^J \approx T^{\frac{1}{2a+d}}$ for $a > \max(d-1, 1/2)$ and $\varepsilon_T = T^{-\frac{a \wedge s}{2a+d}}(\log T)$. Let D_T be as in (28) for a choice of $\Gamma_T \subset V_J^{\otimes d}$ whose envelopes from (27) satisfy*

$|\Gamma_T|_2 = O(\sqrt{T}\varepsilon_T)$ *and* $\varepsilon_T \sigma_{\Gamma_T} \to 0$ *as* $T \to \infty$. *Then for all* $M > 0$ *large enough,* $\Pi(D_T | X^T) = 1 - o_{P_{b_0}}(1)$. *Moreover for* b_u *as in ([44]) and all* $u \in \mathbb{R}$,

$$(46) \qquad \frac{\int_{D_T} e^{\ell_T(b_u)} d\Pi(b)}{\int_{D_T} e^{\ell_T(b)} d\Pi(b)} = 1 + \zeta_T(u) \le C_T e^{r_T u^2},$$

where $\zeta_T(u) = o_{P_{b_0}}(1)$ *for every fixed* u, *where both* $C_T = O_{P_{b_0}}(1)$ *and nonrandom* $r_T = o(1)$ *are independent of* u, *and all terms are uniform over* $\gamma \in \Gamma_T$.

(ii) The conclusion of Part (i) remains true for $d \le 4$ *and under the conditions of Theorem [2] if* D_T *is replaced by the set* $\bar{D}_T$ *from ([29]) with* $\bar{M}_T = (\log T)^{\delta-1}, \delta > 5/2$, *and if in addition* $|\Gamma_T|_2 = O(1)$ *as* $T \to \infty$.

PROOF. (i) The set of b's satisfying the L^2-constraint in the definition ([28]) of D_T has posterior probability tending to one by Theorem [1]. Recall that by definition of the RKHS, $\langle b, \gamma \rangle_{\mathbb{H}} \sim N(0, \|\gamma\|_{\mathbb{H}}^2)$ for $b \sim \Pi$ and $\gamma \in \mathbb{H} = V_J^{\otimes d}$. By Dudley's metric entropy inequality (Section 2.3 in [26]) applied to the Gaussian process $(\langle b, \gamma \rangle_{\mathbb{H}} : \gamma \in \Gamma_T)$ indexed by bounded subsets of the finite-dimensional space $V_J^{\otimes d}$ (with covering numbers bounded in Proposition 4.3.34 in [26]), we have

$$(47) \qquad E^{\Pi} \sup_{\gamma \in \Gamma_T} |\langle b, \gamma \rangle_{\mathbb{H}}| \lesssim 2^{Jd/2} \sigma_{\Gamma_T} \sqrt{\log(1/\sigma_{\Gamma_T})} \le M_0 \sqrt{T} \varepsilon_T \sigma_{\Gamma_T}$$

for some $M_0 > 0$, since we may always take $\sigma_{\Gamma_T} \ge 1$. By the Borell-Sudakov-Tsirelson inequality (Theorem 2.5.8 of [26]), for $M > M_0$,

$$\Pi\Big(\sup_{\gamma \in \Gamma_T} |\langle b, \gamma \rangle_{\mathbb{H}}| > M\sqrt{T}\varepsilon_T \sigma_{\Gamma_T} \Big)$$
$$\le \Pi\Big(\sup_{\gamma \in \Gamma_T} |\langle b, \gamma \rangle_{\mathbb{H}}| > E^{\Pi} \sup_{\gamma \in \Gamma_T} |\langle b, \gamma \rangle_{\mathbb{H}}| + (M - M_0)\sqrt{T}\varepsilon_T \sigma_{\Gamma_T} \Big)$$
$$\le e^{-\frac{1}{2}(M-M_0)^2 T \varepsilon_T^2}.$$

Taking $M > 0$ large enough, the posterior probability of the set in the last display is then $o_{P_{b_0}}(1)$ by Lemma [7]. In summary this establishes that $\Pi(D_T^c | X_T) = o_{P_{b_0}}(1)$.

We now establish ([46]). Letting Π_u denote the law of b_u under the prior and applying the Cameron-Martin theorem (Theorem 2.6.13 of [26]), the desired ratio equals

$$(48) \qquad \frac{\int_{D_{T,u}} e^{\ell_T(g)} \frac{d\Pi_u}{d\Pi}(g) d\Pi(g)}{\int_{D_T} e^{\ell_T(g)} d\Pi(g)} = \frac{\int_{D_{T,u}} e^{\ell_T(g)} e^{-\frac{u}{\sqrt{T}}\langle \gamma, g \rangle_{\mathbb{H}} - \frac{u^2}{2T}\|\gamma\|_{\mathbb{H}}^2} d\Pi(g)}{\int_{D_T} e^{\ell_T(g)} d\Pi(g)},$$

where $D_{T,u} = \{g = b_u : b \in D_T\}$. By the definition of D_T,

$$\sup_{g \in D_{T,u}, \gamma \in \Gamma_T} \left| \frac{u}{\sqrt{T}} \langle \gamma, g \rangle_{\mathbb{H}} + \frac{u^2}{2T} \|\gamma\|_{\mathbb{H}}^2 \right| \leq \frac{|u|}{\sqrt{T}} \sup_{b \in D_T, \gamma \in \Gamma_T} |\langle \gamma, b - \frac{u\gamma}{\sqrt{T}} \rangle_{\mathbb{H}}| + \frac{u^2 \sigma_{\Gamma_T}^2}{2T}$$

$$\leq |u| M \varepsilon_T \sigma_{\Gamma_T} + \frac{3u^2 \sigma_{\Gamma_T}^2}{2T}.$$

We thus upper bound (48) by

$$(49) \qquad e^{\tilde{r}_T u^2 + \tilde{r}_T' |u|} \frac{\int_{D_{T,u}} e^{\ell_T(g)} d\Pi(g)}{\int_{D_T} e^{\ell_T(g)} d\Pi(g)} = e^{\tilde{r}_T u^2 + \tilde{r}_T' |u|} \frac{\Pi(D_{T,u}|X^T)}{\Pi(D_T|X^T)},$$

where $\tilde{r}_T, \tilde{r}_T' \to 0$ are non-random and uniform over $\gamma \in \Gamma_T$. Since $\alpha |u| \leq \alpha^2 u^2 + 1$ for all $\alpha \geq 0$ and $u \in \mathbb{R}$, the exponential in the last display is bounded by $e^{r_T u^2 + 1}$ for all $u \in \mathbb{R}$, where $r_T = \tilde{r}_T + (\tilde{r}_T')^2 = 3\sigma_{\Gamma_T}^2/(2T) + M^2 \varepsilon_T^2 \sigma_{\Gamma_T}^2 \to 0$. Since we have already shown that $\Pi(D_T|X^T) = 1 - o_{P_{b_0}}(1)$ and the posterior probability $\Pi(D_{T,u}|X^T)$ is bounded by one, the inequality in (46) follows.

Turning to the exact asymptotics for fixed $u \in \mathbb{R}$, the right hand side in (49) equals $\Pi(D_{T,u}|X^T)(1 + o_{P_{b_0}}(1))$, and (48) can be lower bounded by (49) with $e^{\tilde{r}_T u^2 + \tilde{r}_T' |u|}$ replaced by $e^{-\tilde{r}_T u^2 - \tilde{r}_T' |u|}$. It consequently suffices to prove $\Pi(D_{T,u}|X^T) = 1 - o_{P_{b_0}}(1)$. Now

$$\Pi(D_{T,u}^c|X^T) \leq \Pi(g \in V_J^{\otimes d} : \|g + \tfrac{u}{\sqrt{T}}\gamma - b_0\|_{\mu_0} > M_T \varepsilon_T | X^T)$$

$$+ \Pi\Big(g \in V_J^{\otimes d} : \sup_{\gamma \in \Gamma_T} |\langle g + \tfrac{u}{\sqrt{T}}\gamma, \gamma \rangle_{\mathbb{H}}| > M\sqrt{T}\varepsilon_T \sigma_{\Gamma_T} \Big| X^T \Big).$$

By Proposition 1, $\|\tfrac{u}{\sqrt{T}}\gamma\|_{\mu_0} \lesssim \tfrac{|u|}{\sqrt{T}} |\Gamma_T|_2 = O(\varepsilon_T) = o(M_T \varepsilon_T)$, so that the first posterior probability tends to zero by Theorem 1. Using (47), that $\sigma_{\Gamma_T}^2/\sqrt{T} = o(\sqrt{T}\varepsilon_T \sigma_{\Gamma_T})$ and the Borell-Sudakov-Tsirelson inequality (Theorem 2.5.8 of [26]), the prior probability of the second event is bounded by

$$\Pi\Big(g : \sup_{\gamma \in \Gamma_T} |\langle g, \gamma \rangle_{\mathbb{H}}| + \frac{|u|\sigma_{\Gamma_T}^2}{\sqrt{T}} > E^{\Pi} \sup_{\gamma \in \Gamma_T} |\langle b, \gamma \rangle_{\mathbb{H}}| + (M - M_0)\sqrt{T}\varepsilon_T \sigma_{\Gamma_T} \Big)$$

$$\leq e^{-\frac{1}{4}(M - M_0)^2 T \varepsilon_T^2}$$

for T large enough depending on u. For $M > 0$ large enough, Lemma 7 then yields that the posterior probability of this last set is $o_{P_{b_0}}(1)$, which shows $\Pi(D_{T,u}|X^T) = 1 - o_{P_{b_0}}(1)$ as desired.

Part (ii) is proved in the same way using Theorem 2 (whose proof only relies on Part (i) of the present lemma) to ensure that $\Pi(\bar{D}_T | X^T) \to^{P_{b_0}} 1$ as $T \to \infty$, and upon noting that

$$\|\tfrac{u}{\sqrt{T}}\gamma\|_\infty \lesssim \tfrac{2^{Jd/2}|u|}{\sqrt{T}}|\Gamma_T|_2 = O(\varepsilon_T) = o(\bar{M}_T \varepsilon_T)$$

since $|\Gamma_T|_2 = O(1)$ as $T \to \infty$.   □

4.3. *An approximation lemma.*

LEMMA 9.   *Suppose $b_0 \in C^s(\mathbb{T}^d)$ for some $s \geq 1$ and let $\lambda \leq J$, $1 \leq j \leq d$ and $b \in V_J^{\otimes d}$. If $a_\lambda = 2^{\lambda d/2} 2^{-Jd/2}(\log T)^{-\eta}$ for some $\eta \geq 0$, then*

$$a_\lambda |\langle \mu_0(b_j - b_{0,j}), \Phi_{\lambda,k}/\mu_0 - P_{V_J}[\Phi_{\lambda,k}/\mu_0] \rangle_{L^2}|$$
$$\leq C(2^{-2J}\|b_j - b_{0,j}\|_{L^2} + 2^{-J(s+d/2+1)})(\log T)^{-\eta}.$$

If instead $a_\lambda = 2^{\lambda d/2} 2^{-J(d-2)}(\log T)^{-\eta}$ for some $\eta \geq 0$, then

$$a_\lambda |\langle \mu_0(b_j - b_{0,j}), \Phi_{\lambda,k}/\mu_0 - P_{V_J}[\Phi_{\lambda,k}/\mu_0] \rangle_{L^2}|$$
$$\leq C(2^{-Jd/2}\|b_j - b_{0,j}\|_{L^2} + 2^{-J(s+d-1)})(\log T)^{-\eta}.$$

Finally,

$$\left|\langle \mu_0(b_j - b_{0,j}), \tfrac{\Phi_{\lambda,k}}{\mu_0} - P_{V_J}\big[\tfrac{\Phi_{\lambda,k}}{\mu_0}\big]\rangle_{L^2}\right| \leq C 2^{-\lambda d/2}\big(2^{-2J}\|b_j - b_{0,j}\|_\infty + 2^{-J(s+1)}\big).$$

In all cases, the constant C depends only on b_0, Φ and d.

PROOF.   By the triangle inequality, the desired quantity is bounded by

$$a_\lambda |\langle \mu_0(b_j - P_{V_J} b_{0,j}), \Phi_{\lambda,k}/\mu_0 - P_{V_J}[\Phi_{\lambda,k}/\mu_0] \rangle_{L^2}|$$
$$+ a_\lambda |\langle \mu_0(b_{0,j} - P_{V_J} b_{0,j}), \Phi_{\lambda,k}/\mu_0 - P_{V_J}[\Phi_{\lambda,k}/\mu_0] \rangle_{L^2}| =: (I) + (II).$$

By Parseval's identity,

$$(I) = a_\lambda \left| \sum_{l>J} \sum_r \langle \mu_0(b_j - P_{V_J} b_{0,j}), \Phi_{l,r} \rangle_{L^2} \langle \Phi_{\lambda,k}/\mu_0, \Phi_{l,r} \rangle_{L^2} \right|$$

(50)

$$\leq a_\lambda \sum_{l>J} \max_r |\langle \mu_0(b_j - P_{V_J} b_{0,j}), \Phi_{l,r} \rangle_{L^2}| \sum_r |\langle \Phi_{\lambda,k}/\mu_0, \Phi_{l,r} \rangle_{L^2}|.$$

By Proposition 1 we know that μ_0 has finite Lipschitz norm $\|\mu_0\|_{\mathrm{Lip}}$. Let $x_{l,r} \in I_{l,r} := \mathrm{supp}(\Phi_{l,r})$ and note that $\mathrm{diam}(I_{l,r}) = O(2^{-l})$ by construction

of the wavelets. Using that $b_j - P_{V_J} b_{0,j} \in V_J$ is orthogonal to $\Phi_{l,r}$ for any $l > J$, $\|\Phi_{l,r}\|_{L^1} \lesssim 2^{-ld/2}$ and (17),

$$
|\langle \mu_0(b_j - P_{V_J} b_{0,j}), \Phi_{l,r} \rangle_{L^2}|
$$

$$
= \left| \int_{\mathbb{T}^d} (\mu_0(x) - \mu_0(x_{l,r}))(b_j(x) - P_{V_J} b_{0,j}(x))\Phi_{l,r}(x) dx \right|
$$

$$
\leq \|\mu_0\|_{\mathrm{Lip}} \mathrm{diam}(I_{l,r}) \int_{\mathbb{T}^d} |b_j(x) - P_{V_J} b_{0,j}(x)||\Phi_{l,r}(x)| dx
$$

$$
(51) \qquad \leq C(b_0, \Phi) 2^{-l} \|b_j - P_{V_J} b_{0,j}\|_\infty \|\Phi_{l,r}\|_{L^1}
$$

$$
\leq C(b_0, \Phi) 2^{Jd/2} \|b_j - P_{V_J} b_{0,j}\|_{L^2} 2^{-l(d/2+1)}.
$$

Moreover, using the following standard properties of wavelet bases,

$$
\sup_x \sum_r |\Phi_{l,r}(x)| \lesssim 2^{ld/2}, l \geq 0, \quad \langle \Phi_{\lambda,k}, \Phi_{l,r} \rangle_{L^2} = 0, \quad \lambda \leq J < l,
$$

we deduce

$$
\sum_r |\langle \Phi_{\lambda,k}/\mu_0, \Phi_{l,r} \rangle_{L^2}| = \sum_r \left| \int_{\mathbb{T}^d} \left(\frac{1}{\mu_0(x)} - \frac{1}{\mu_0(x_{l,r})} \right) \Phi_{\lambda,k}(x)\Phi_{l,r}(x) dx \right|
$$

$$
\leq \|1/\mu_0\|_{\mathrm{Lip}} \mathrm{diam}(I_{l,r}) \int_{\mathbb{T}^d} |\Phi_{\lambda,k}(x)| \sum_r |\Phi_{l,r}(x)| dx
$$

$$
(52) \qquad \leq C(b_0, \Phi) 2^{l(d/2-1)} 2^{-\lambda d/2}.
$$

Substituting (51) and (52) into (50) yields

$$
(I) \lesssim a_\lambda 2^{Jd/2} 2^{-\lambda d/2} \|b_j - P_{V_J} b_{0,j}\|_{L^2} \sum_{l>J} 2^{-2l} \lesssim 2^{-2J} (\log T)^{-\eta} \|b_j - b_{0,j}\|_{L^2}
$$

as desired. Next expanding (II) as in (50) and using (52),

$$
(II) \leq C(b_0, \Phi) a_\lambda \sum_{l>J} \max_r |\langle \mu_0(b_{0,j} - P_{V_J}[b_{0,j}]), \Phi_{l,r} \rangle_{L^2}| 2^{l(d/2-1)} 2^{-\lambda d/2}
$$

$$
\lesssim a_\lambda 2^{-Js} 2^{-\lambda d/2} \sum_{l>J} 2^{-l} \lesssim 2^{-J(s+1+d/2)} (\log T)^{-\eta},
$$

where we have used Hölder's inequality and $\|b_{0,j} - P_{V_J}[b_0, j]\|_\infty \lesssim 2^{-Js}$. The last two displays imply the first inequality in the lemma. If instead $a_\lambda = 2^{\lambda d/2} 2^{-J(d-2)} (\log T)^{-\eta}$, then substituting this value into the final bounds for (I) and (II) gives the required result. The final inequality of the lemma is proved in the same way, but by using the penultimate $\|\cdot\|_\infty$ bound in (51) rather than the $\|\cdot\|_{L^2}$ bound in the last line. Taking $a_\lambda = 1$ in the rest of the argument gives the result. $\square$

4.4. *Proof of Lemma 5.*

LEMMA 10. *Suppose $b_0 \in C^s(\mathbb{T}^d)$ for some $s > d/2$. Then for Γ_T as in (30) and the RKHS norm $\| \cdot \|_{\mathbb{H}}$ defined in (12) with $\sigma_l = 2^{-l(\alpha+d/2)}$ for $\alpha \geq 0$, we can take the envelopes from (27) as*

$$|\Gamma_T|_2 = C(d,\mu_0) \max_{\lambda \leq J} a_\lambda, \quad \sigma_{\Gamma_T} = C(d,\mu_0,\Phi) 2^{J(\alpha+d/2)} \max_{\lambda \leq J} a_\lambda.$$

PROOF. Using Proposition 1,

$$\|\tilde{\Phi}_{\lambda,k,j}\|_{L^2}^2 = \|P_{V_J}[a_\lambda \Phi_{\lambda,k}/\mu_0]\|_{L^2}^2 \leq a_\lambda^2 \|1/\mu_0\|_\infty^2 \|\Phi_{\lambda,k}\|_{L^2}^2 \lesssim a_\lambda^2.$$

To prove the second bound, note that Proposition 1 implies that $1/\mu_0$ has finite Lipschitz norm $\|1/\mu_0\|_{\text{Lip}}$ on $\mathbb{T}^d$. Let $x_{l,r} \in \text{supp}(\Phi_{l,r})$ and note that $\text{diam}(\text{supp}(\Phi_{l,r})) = O(2^{-l})$ by construction of the wavelets. Using the orthogonality of the wavelets and Hölder's inequality, for $(l,r) \neq (\lambda,k)$ with $\lambda \leq J$,

$$|\langle P_{V_J}[\Phi_{\lambda,k}/\mu_0], \Phi_{l,r}\rangle_{L^2}| = \left| \int_{\mathbb{T}^d} \Phi_{\lambda,k}(x)\Phi_{l,r}(x) \left(\frac{1}{\mu_0(x)} - \frac{1}{\mu_0(x_{l,r})} \right) dx \right|$$

$$\lesssim \|1/\mu_0\|_{\text{Lip}} 2^{-\max(l,\lambda)} \int_{\mathbb{T}^d} |\Phi_{\lambda,k}(x)||\Phi_{l,r}(x)|dx$$

$$\lesssim 2^{-\max(l,\lambda)} 2^{-|l-\lambda|d/2},$$

while $|\langle P_{V_J}[\Phi_{\lambda,k}/\mu_0], \Phi_{\lambda,k}\rangle_{L^2}| \leq \|1/\mu_0\|_\infty \|\Phi_{\lambda,k}\|_{L^2}^2 \lesssim 1$. Note that for $l \leq \lambda$, there are a constant number of wavelets $\Phi_{l,r}$ intersecting $\text{supp}(P_{V_J}[\Phi_{\lambda,k}/\mu_0])$, while for $l \geq \lambda$, there are $O(2^{(l-\lambda)d})$ such wavelets. Splitting the following sum into these two cases, while separately keeping track of the term $(l,r) = (\lambda,k)$, and using the above bounds gives that for $\lambda \leq J$,

$$\|\tilde{\Phi}_{\lambda,k,j}\|_{\mathbb{H}}^2 = \sum_{l \leq J} \sum_r \sigma_l^{-2} |\langle P_{V_J}[a_\lambda \Phi_{\lambda,k}/\mu_0], \Phi_{l,r}\rangle_{L^2}|^2$$

$$\lesssim a_\lambda^2 \left(\sum_{l=0}^{\lambda} \sigma_l^{-2} 2^{-2\lambda-(\lambda-l)d} + \sum_{l=\lambda+1}^{J} \sigma_l^{-2} 2^{(l-\lambda)d} 2^{-2l-(l-\lambda)d} + \sigma_\lambda^{-2} \right)$$

$$\lesssim a_\lambda^2 \left(2^{-(d+2)\lambda} \sum_{l=0}^{\lambda} 2^{2l(\alpha+d)} + \sum_{l=\lambda+1}^{J} 2^{(2\alpha+d-2)l} + 2^{(2\alpha+d)\lambda} \right)$$

$$\lesssim a_\lambda^2 2^{(2\alpha+d)J}.$$

This yields $\sigma_{\Gamma_T}^2 \lesssim a_\lambda^2 2^{(2\alpha+d)J}$.   □

We now prove Lemma 5 from above. The first assertion in Part (i) follows from Lemmas 8 and 10 since the envelopes satisfy $|\Gamma_T|_2 = O(a_J) = O((\log T)^{-\eta})$, $\sqrt{T}\varepsilon_T \to \infty$ and $\varepsilon_T \sigma_{\Gamma_T} = o(1)$ for $0 \leq \alpha < a \wedge s - d/2$ and the specified choice $2^J \approx T^{1/(2a+d)}$. To prove the first two maximal inequalities, we start by verifying the conditions of Proposition 2 for the case $d \leq 4$. Using the envelopes from Lemma 10, we can bound R_T and $\tilde{R}_T$ in Proposition 2 by

$$R_T \lesssim M_T T^{\frac{d - a \wedge s + (d/2 + \kappa - 1)_+}{2a+d}} (\log T)^{3/2 - \eta} \to 0$$

for $a \wedge s > \max(3d/2 - 1, 1)$, $0 < \kappa < a \wedge s - 3d/2 + 1$ ($\kappa = 0$ if $d = 1$) and $M_T \to \infty$ slowly enough, while

$$(53) \qquad \tilde{R}_T \lesssim T^{-\frac{1}{2} + \frac{d + (d/2 + \kappa - 1)_+}{2a+d}} (\log T)^{-2\eta} \sqrt{\log \log T} \to 0$$

for $a > \max(d - 1, 1/2)$ and $0 < \kappa < s - d + 1$ ($\kappa = 0$ if $d = 1$). We may thus apply Proposition 2 to $G(b) = \langle b_j - b_{0,j}, a_\lambda \Phi_{\lambda,k} \rangle_{L^2}$ with $\gamma = \tilde{\Phi}_{\lambda,k,j} \in V_J^{\otimes d}$, so that for $b_u = b - \frac{u}{\sqrt{T}} \tilde{\Phi}_{\lambda,k,j}$, $u \in \mathbb{R}$,

$$E^{\Pi^{D_T}}[e^{u\sqrt{T}\langle b_j - b_{0,j}, a_\lambda \Phi_{\lambda,k}\rangle_{L^2}} | X^T] = e^{\frac{u}{\sqrt{T}} \int_0^T \tilde{\Phi}_{\lambda,k,j}(X_t).dW_t + \frac{u^2}{2}\|\tilde{\Phi}_{\lambda,k,j}\|_{\mu_0}^2 + ur_T + u^2 \tilde{r}_T}$$

$$\times \frac{\int_{D_T} e^{S_T(b) + \ell_T(b_u)} d\Pi(b)}{\int_{D_T} e^{\ell_T(b)} d\Pi(b)},$$

where $r_T = O_{P_{b_0}}(R_T)$, $\tilde{r}_T = O_{P_{b_0}}(\tilde{R}_T)$ uniformly over $\lambda \leq J$ and k, j, and

$$S_T(b) = u\sqrt{T}\langle b_j - b_{0,j}, a_\lambda \Phi_{\lambda,k} \rangle_{L^2} - u\sqrt{T}\langle b_j - b_{0,j}, a_\lambda P_{V_J}[\Phi_{\lambda,k}/\mu_0] \rangle_{\mu_0}$$

$$= u\sqrt{T}a_\lambda \langle \mu_0(b_j - b_{0,j}), \Phi_{\lambda,k}/\mu_0 - P_{V_J}[\Phi_{\lambda,k}/\mu_0] \rangle_{L^2}.$$

Applying the first bound from Lemma 9 with $d \leq 4$, $\eta > 1$,

$$(54)$$
$$\sup_{b \in D_T} |S_T(b)| \lesssim |u|\sqrt{T}(\log T)^{-\eta}(2^{-2J}M_T\varepsilon_T + 2^{-J(s+1+d/2)}) = |u| \times o(1),$$

for $M_T \to \infty$ slowly enough and $s \geq a - 1$. Applying $\alpha|u| \leq \alpha^2 u^2 + 1$ for all $\alpha \geq 0$ to (54), and using Lemmas 8 and 10 gives for all $u \in \mathbb{R}$,

$$\frac{\int_{D_T} e^{S_T(b) + \ell_T(b_u)} d\Pi(b)}{\int_{D_T} e^{\ell_T(b)} d\Pi(b)} \leq C_T e^{c_T u^2},$$

where $C_T = O_{P_{b_0}}(1)$ and non-random $c_T = o(1)$ are independent of u and uniform over $\lambda \leq J$ and k, j.

Setting $Z_{\lambda,k,j} = \langle b_j - b_{0,j}, a_\lambda \Phi_{\lambda,k} \rangle_{L^2} - \frac{1}{T} \int_0^T \tilde{\Phi}_{\lambda,k,.}(X_t).dW_t$ and using again that $\alpha|u| \leq \alpha^2 u^2 + 1$ for all $\alpha \geq 0$, the Laplace transform satisfies the conditional subgaussian bound

$$E^{\Pi^{D_T}}[e^{u\sqrt{T}Z_{\lambda,k,j}}|X^T] \leq C'_T e^{\frac{u^2}{2}(\|\tilde{\Phi}_{\lambda,k,j}\|^2_{\mu_0}+c'_T)}$$

for sequences $C'_T = O_{P_{b_0}}(1)$ and $c'_T = 2(\tilde{r}_T + r_T^2 + c_T) = o_{P_{b_0}}(1)$, which are independent of u and uniform over $\lambda \leq J$ and k, j. Since $\|\tilde{\Phi}_{\lambda,k,j}\|^2_{\mu_0} = a_\lambda^2 \|P_{V_J}[\Phi_{\lambda,k}/\mu_0]\|^2_{\mu_0} \leq a_\lambda^2 \|\mu_0\|_\infty \|1/\mu_0\|^2_\infty \leq C(b_0)$, standard subgaussian inequalities (Lemmas 2.3.2 and 2.3.4 of [26]) give

$$(55) \quad \sqrt{T}E^{\Pi^{D_T}}[\max_{\lambda \leq J,k,j} |Z_{\lambda,k,j}||X^T]$$

$$\lesssim \sqrt{2\log 2\dim(V_J^{\otimes d})}(C'_T + 1) \max_{\lambda \leq J,k,j}(\|\tilde{\Phi}_{\lambda,k,j}\|^2_{\mu_0} + c'_T) = O_{P_{b_0}}(\sqrt{J})$$

since $\dim(V_J^{\otimes d}) = O(d2^{Jd})$. Similarly, for $\lambda \leq J$,

$$(56) \quad \sqrt{T}E^{\Pi^{D_T}}[\max_{k,j} |Z_{\lambda,k,j}||X^T]$$

$$\lesssim \sqrt{2\log 2\dim(V_\lambda^{\otimes d})}(C'_T + 1) \max_{k,j}(\|\tilde{\Phi}_{\lambda,k,j}\|^2_{\mu_0} + c'_T) = O_{P_{b_0}}(\sqrt{\lambda}).$$

We now deduce (33) from (55), the same arguments then also show that (32) follows from (56). Decompose

$$E^{\Pi^{D_T}}\left[\max_{\lambda \leq J,k,j} \sqrt{T}|\langle b_j - b_{0,j}, a_\lambda \Phi_{\lambda,k} \rangle_{L^2}|\,\Big|\, X^T\right]$$

$$\leq E^{\Pi^{D_T}}\left[\max_{\lambda \leq J,k,j} \sqrt{T}|Z_{\lambda,k,j}|\,\Big|\, X^T\right] + \max_{\lambda \leq J,k,j}\left|\frac{1}{\sqrt{T}}\int_0^T \tilde{\Phi}_{\lambda,k,j}(X_t).dW_t\right|$$

and we have shown the first term is $O_{P_{b_0}}(\sqrt{J})$. We now control the P_{b_0}-expectation of the second term by showing that $M_T^{\lambda,k,j} = \int_0^T \tilde{\Phi}_{\lambda,k,j}(X_t).dW_t$ are subgaussian with uniform constants on a suitable event A_T. For $\epsilon > 0$ fixed, set

$$(57) \quad A_T = \left\{\max_{\lambda,k,j}\left|\frac{1}{T}\int_0^T \|\tilde{\Phi}_{\lambda,k,j}(X_t)\|^2 dt - \|\tilde{\Phi}_{\lambda,k,j}\|^2_{\mu_0}\right| \leq \epsilon\right\}.$$

Applying Markov's inequality, Proposition 2(ii) and (53), $P_{b_0}(A_T^c) \lesssim \epsilon^{-1}\tilde{R}_T \to 0$. On A_T we have

$$T^{-1}[M^{\lambda,k,j}]_T = T^{-1}\int_0^T \|\tilde{\Phi}_{\lambda,k,j}(X_T)\|^2 dt \le \|\tilde{\Phi}_{\lambda,k,j}\|_{\mu_0}^2 + \epsilon \le C_0(b_0) + \epsilon,$$

so that applying Bernstein's inequality (15), for any $x > 0$,

$$P_{b_0}(T^{-1/2}|M_T^{\lambda,k,j}|1_{A_T} \ge x) \le P_{b_0}(|M_T^{\lambda,k,j}| \ge x\sqrt{T}, [M^{\lambda,k,j}]_T \le (C_0 + \epsilon)T)$$

$$\le 2e^{-\frac{x^2}{2(C_0+\epsilon)}}.$$

Consequently, $(T^{-1/2}M^{\lambda,k,j}1_{A_T} : \lambda, k)$ are subgaussian random variables with uniformly bounded constants, so that $E_{b_0}\max_{\lambda \le J,k,j} T^{-1/2}|M_T^{\lambda,k,j}|1_{A_T} = O(\sqrt{J})$ by Lemma 2.3.4 of [26]. When $d \ge 5$, one proceeds exactly as above with the only difference to the case $d \le 4$ being that we use the second bound in Lemma 9 with $a \le s + d/2 - 1$ rather than the first bound, which is needed to ensure $\sup_{b \in D_T}|S_T(b)| = o(|u|)$.

For Part (ii), we can invoke Lemma 8(ii) to obtain $\Pi(\bar{D}_T|X^T) \to 1$ in P_{b_0}-probability. The maximal inequality then follows from the same proof as in (i), using Proposition 2(iii), that the $\|\cdot\|_\infty$-contraction rate implies the same rate in L^2-norm, and replacing the bias bound (54) by the third inequality of Lemma 9 so that scaling by a_λ is not necessary. The conditions $s > a-1+d/2$ and $d < 4$ ensure that the terms $\sqrt{T}2^{-J(s+1)}$ and $\sqrt{T}2^{-2J}\|b-b_0\|_\infty$ are both $o(1)$ and hence asymptotically negligible.

5. Proofs for Section 2.5. The proofs follow from Theorems 3 and 4 and a version of the 'Delta'-method for weak convergence applied to the map $b \mapsto \mu_b$. We will represent $\mu_b - \mu_{b+h}$ by a linear transformation of the vector field h plus a remainder term that will be seen to be quadratic in (suitable norms of) h. The identity (60) below is the key to these proofs and can be derived from perturbation arguments for the PDE (7) as follows: let μ_b and μ_{b+h} correspond to vector fields $b, b + h \in C^1(\mathbb{T}^d)$ (cf. Proposition 1). Then necessarily $L_b^*\mu_b = L_{b+h}^*\mu_{b+h}$ or in other words

$$\frac{\Delta}{2}\mu_b - b.\nabla\mu_b - div(b)\mu_b = \frac{\Delta}{2}\mu_{b+h} - (b+h).\nabla\mu_{b+h} - div(b+h)\mu_{b+h},$$

which is the same as

$$\frac{\Delta}{2}(\mu_b - \mu_{b+h}) - b.\nabla(\mu_b - \mu_{b+h}) - div(b)(\mu_b - \mu_{b+h}) = -h.\nabla\mu_{b+h} - div(h)\mu_{b+h}.$$

Thus $u = \mu_b - \mu_{b+h}$ solves the equation

$$(58) \qquad\qquad L_b^*u = -h.\nabla\mu_{b+h} - div(h)\mu_{b+h}.$$

Next denote by $v_h = v_{b,h}$ the unique periodic solution of the PDE

$$(59) \qquad L_b^* v_h = -h.\nabla \mu_b - div(h)\mu_b = -\sum_{j=1}^d \frac{\partial}{\partial x_j}(h_j \mu_b) \equiv f_h$$

satisfying $\int v_h = 0$. In view of the results in Section 6 and since, with $dx^{(j)} = \prod_{i \neq j} dx_i$,

$$\int_{\mathbb{T}^d} f_h(x)dx = \sum_{j=1}^d \int_0^1 \cdots \int_0^1 \frac{\partial}{\partial x_j}(h_j(x)\mu_b(x))dx$$

$$= \sum_j \int_{\mathbb{T}^{d-1}} \Big[(h_j \mu_b)(x_1, \ldots, x_{j-1}, 1, x_{j+1}, \ldots, x_d)$$

$$- (h_j \mu_b)(x_1, \ldots, x_{j-1}, 0, x_{j+1} \ldots, x_d)\Big] dx^{(j)} = 0,$$

such a solution exists and can be represented as $v_h = (L_b^*)^{-1}[f_h]$, a map that is linear in h. Now since $\int \mu_{b+h} - \int \mu_b = 1 - 1 = 0$, we can use (58) and (59) to see that the differences $w_{b,h} = \mu_b - \mu_{b+h} - v_h$ are the unique (periodic) integral-zero solutions of

$$L_{b+h}^* w_{b,h} = L_b^* w_{b,h} - h.\nabla w_{b,h} - div(h)w_{b,h} = \sum_{j=1}^d \frac{\partial}{\partial x_j}[h_j v_h] \equiv \bar{f}_h,$$

where again $\int \bar{f}_h = 0$ as in the penultimate display, so that we can write $w_{b,h} = (L_{b+h}^*)^{-1}[\bar{f}_h]$. Thus we have, for any $h \in C^1(\mathbb{T}^d)$, the decomposition

$$(60) \qquad \mu_b - \mu_{b+h} = v_{b,h} + w_{b,h} = (L_b^*)^{-1}[f_h] + (L_{b+h}^*)^{-1}[\bar{f}_h],$$

which for sufficiently smooth b, h (such that also $\mu_b, \mu_{b+h} \in C^r, r > 2$, see after (63)) holds classically (pointwise on $\mathbb{T}^d$).

5.1. *Proof of Theorem 5.* It suffices to prove the theorem for

$$\sqrt{T}(\mu_b - \mu_{\hat{b}_T})|X^T, \quad b \sim \Pi^{\bar{D}_T}(\cdot|X^T),$$

where $\Pi^{\bar{D}_T}(\cdot|X^T)$ was introduced at the beginning of the proof of Theorem 3 and $\bar{D}_T$ is given by (29). On the set $\bar{D}_T$ we have the estimate

$$\|b\|_{B_{\infty\infty}^1} \lesssim 2^J \|b - b_0\|_\infty + \max_j \|b_{0,j} - P_{V_J}(b_{0,j})\|_{B_{\infty\infty}^1} = O(1)$$

as $T \to \infty$, and the same argument shows $\|\hat{b}_T\|_{B^1_{\infty\infty}} = O_{P_{b_0}}(1)$ by virtue of Corollary 1. Proposition 1 then further implies that $\|\mu_b\|_{Lip}, \|\mu_{\hat{b}_T}\|_{Lip}$ are also $O(1)$ and $O_{P_{b_0}}(1)$, respectively – these bounds will be used repeatedly in the proof without further mention. Recall that $v_h = (L_b^*)^{-1}[f_h]$ and $w_{b,h} = (L_{b+h}^*)^{-1}[\bar{f}_h]$ for $f_h = -\sum_{j=1}^d \frac{\partial}{\partial x_j}(h_j\mu_b)$ and $\bar{f}_h = \sum_{j=1}^d \frac{\partial}{\partial x_j}[h_j v_h]$. We will use the decomposition (60) with $h = \hat{b}_T - b$. First, for the 'remainder' term, we can use (73) below and (3) to deduce that, uniformly in $\|g\|_{\mathbb{B}_r} \leq 1$,

$$\left| \int_{\mathbb{T}^d} w_{b,h} g \right| \leq \|g\|_{L^2} \|(L_{b+h}^*)^{-1}[\bar{f}_h]\|_{L^2}$$

$$\lesssim \|\bar{f}_h\|_{H^{-2}} = \sup_{\|\phi\|_{H^2}\leq 1} \left| \sum_{j=1}^d \int_{\mathbb{T}^d} \phi \frac{\partial}{\partial x_j}[h_j v_h] \right|$$

$$(61) \qquad \leq \sup_{\|\phi\|_{H^2}\leq 1} \left| \sum_{j=1}^d \int_{\mathbb{T}^d} h_j v_h \frac{\partial}{\partial x_j}\phi \right| \lesssim \|h\|_\infty \|(L_b^*)^{-1}[f_h]\|_{L^2}$$

$$\lesssim \|h\|_\infty \|f_h\|_{H^{-2}} \lesssim \|h\|_\infty \|h\mu_b\|_{L^2} \lesssim \|h\|_\infty^2$$

is $O_{P_{b_0}}((\|\hat{b}_T - b_0\|_\infty + \|b - b_0\|_\infty)^2) = o_{P_{b_0}}(1/\sqrt{T})$ on $\bar{D}_T$ and by Corollary 1. For the 'linear' term we may write, noting the dependence $f_h = f_{h,b}$ on b,

$$\int v_{b,h} g = \int (L_{b_0}^*)^{-1}[f_{h,b_0}]g + \int (L_{b_0}^*)^{-1}[f_{h,b} - f_{h,b_0}]g + \int [(L_b^*)^{-1} - (L_{b_0}^*)^{-1}][f_{h,b}]g$$

and we denote the right hand side as $A_0 + A_1 + A_2$. The last term A_2 is $o_{P_{b_0}}(1/\sqrt{T})$ in $\mathbb{B}_r^*$ since $[(L_b^*)^{-1} - (L_{b_0}^*)^{-1}][f_{h,b}]$ can be written as $-(L_b^*)^{-1}[(b - b_0).\nabla\omega + div(b - b_0)\omega]$ for $\omega = (L_{b_0}^*)^{-1}[f_{h,b}]$ (arguing just as in (58)), so that using (73) gives (as in (61)) the inequality
(62)
$$\|[(L_b^*)^{-1} - (L_{b_0}^*)^{-1}][f_h]\|_{L^2} \lesssim \|b - b_0\|_\infty(\|\hat{b}_T - b_0\|_\infty + \|b - b_0\|_\infty) = o_{P_{b_0}}(1/\sqrt{T}).$$

Similarly the term A_1 can be bounded in $\mathbb{B}_r^*$ by

$$\|(L_{b_0}^*)^{-1}[f_{h,b} - f_{h,b_0}]\|_{L^2} \lesssim \|f_{h,b} - f_{h,b_0}\|_{H^{-2}} \lesssim \|h\|_\infty \|b - b_0\|_\infty = o_{P_{b_0}}(1/\sqrt{T}).$$

Finally, for the term A_0, we show that the linear operator

$$h \mapsto v_{b_0,h} = (L_{b_0}^*)^{-1}[f_{h,b_0}]$$

is Lipschitz on $C^1(\mathbb{T}^d)$ for the norms $(B_{1\infty}^{r+1})^*$ and $\mathbb{B}_r^*$ for any $d/2-1 < r < 1$. Using that $v_{b_0,h} \in L_0^2(\mathbb{T}^d)$ and writing $\bar{g} = g - \int g d\mu_0$,

$$\|v_{b_0,h}\|_{\mathbb{B}_r^*}$$

$$= \sup_{\|g\|_{L^2}+\|g\|_{B_{1\infty}^r} \leq 1} \left| \int_{\mathbb{T}^d} g v_{b_0,h} \right| = \sup_{\|g\|_{L^2}+\|g\|_{B_{1\infty}^r} \leq 1} \left| \int_{\mathbb{T}^d} L_{b_0} L_{b_0}^{-1}[\bar{g}] v_{b_0,h} \right|$$

$$= \sup_{\|g\|_{L^2}+\|g\|_{B_{1\infty}^r} \leq 1} \left| \int_{\mathbb{T}^d} L_{b_0}^* v_{b_0,h} L_{b_0}^{-1}[\bar{g}] \right|$$

$$= \sup_{\|g\|_{L^2}+\|g\|_{B_{1\infty}^r} \leq 1} \left| \sum_{j=1}^d \int_{\mathbb{T}^d} h_j \left(\frac{\partial}{\partial x_j} L_{b_0}^{-1}[\bar{g}] \right) \mu_0 \right|$$

(63)

$$= \sup_{\|g\|_{L^2}+\|g\|_{B_{1\infty}^r} \leq 1} \left| \langle h, \mu_0 \nabla L_{b_0}^{-1}[\bar{g}] \rangle_{L^2} \right|$$

$$\lesssim \|\mu_0\|_{C^{r+1}} \sup_{\|g\|_{L^2}+\|g\|_{B_{1\infty}^r} \leq 1} \|\nabla L_{b_0}^{-1}[\bar{g}]\|_{B_{1\infty}^{r+1}} \sup_{\tilde{g}: \|\tilde{g}\|_{B_{1\infty}^{r+1}} \leq 1} |\langle h, \tilde{g} \rangle_{L^2}| \lesssim \|h\|_{(B_{1\infty}^{r+1})^*},$$

where we have used (3), (74) below and that ∇ maps $B_{1\infty}^{r+2}$ continuously into $B_{1\infty}^{r+1,\otimes d}$. We also used that $\mu_0 \in C^t(\mathbb{T}^d), t \geq 1$, whenever $b_0 \in C^s \cap H^s, s > t + d/2 - 1$. Indeed, $\|\mu_0\|_{H^1} \lesssim \|\mu_0\|_{Lip} \leq C(d, \|b_0\|_\infty)$ (by Proposition 1 and Rademacher's theorem) allows an iterated application of the inequality (72) below with $u = \mu_0$ to bound $\|\mu_0\|_{H^{s+1}}$ by a constant $C(d, s, \|b_0\|_{C^s})$, which in turn bounds $\|\mu_0\|_{C^t}$ by the Sobolev embedding theorem.

Summarizing, with $h = b - \hat{b}_T$ we have proved uniformly in $\|g\|_{\mathbb{B}_r} \leq 1$,

$$\sqrt{T} \int_{\mathbb{T}^d} (\mu_b - \mu_{\hat{b}_T}) g = \int_{\mathbb{T}^d} v_{b_0, \sqrt{T}(b-\hat{b}_T)} g + o_{P_{b_0}}(1)$$

and that the linear operator $h \mapsto v_{b_0,h}$ is continuous from $(C^1(\mathbb{T}^d), \|\cdot\|_{(B_{1\infty}^{r+1})^*})$ to $\mathbb{B}_r^*$. Theorem 5 now follows from Theorem 3 with $r + 1 = \rho$ and the continuous mapping theorem for weak convergence applied to $\sqrt{T}(b - \hat{b}_T)$. We note that the calculation leading to (63) shows that the co-variance of the limiting Gaussian process is the one of the Gaussian process $g \mapsto \mathbb{W}_0(\mu_0 \nabla L_{b_0}^{-1}[\bar{g}])$, $\mathbb{W}_0 \sim \mathcal{N}_{b_0}$, of the required form. In particular, $\mathcal{N}_{\mu_0}$ exists as a tight Gaussian probability measure in $\mathbb{B}_r^*$ as the image of $\mathcal{N}_{b_0}$ under the continuous map $v_{b_0,\cdot}$. The limit of the MAP-estimate follows from similar (in fact simpler) arguments and Theorem 4, and is left to the reader.

5.2. *Proof of Theorem 6.* We finally prove Theorem 6 and explain the necessary modifications to the arguments from the proof of Theorem 5. As

in the proof of Theorem 5, one shows that $\hat{b}_T, b$ are (in the former case, stochastically) bounded in $B^1_{\infty\infty}$ on the set $\bar{D}_T$, and so are then $\mu_{\hat{b}_T}, \mu_b$ by Proposition 1. Using the Sobolev-embedding $H^1(\mathbb{T}) \subset C(\mathbb{T})$ and then repeatedly Lemma 12, (73) and the basic interpolation inequality $\|g\|_{H^1} \lesssim \|g\|_{H^2}^{1/2}\|g\|_{L^2}^{1/2}$, the second term in the decomposition (60) can be bounded by

$$\|(L^*_{b+h})^{-1}[\bar{f}_h]\|_{H^1} \lesssim \|\bar{f}_h\|_{L^2}^{1/2}\|\bar{f}_h\|_{H^{-2}}^{1/2} \lesssim \|h\|_{H^1}\|h\|_{L^2},$$

which for $a > 3/2$ and $h = b - \hat{b}_T = b - b_0 - (\hat{b}_T - b_0), b \sim \Pi^{\bar{D}_T}(\cdot|X^T)$, is of order $\|h\|_{H^1}\|h\|_{L^2} = o_P(1/\sqrt{T})$ since $\|h\|_{H^1} \lesssim 2^J\|h\|_{L^2}$ for $h \in V_J$. The linear term in (60) can be decomposed as

$$(L^*_{b_0})^{-1}[f_h](x) - [(L^*_{b_0})^{-1} - (L^*_b)^{-1}][f_h](x).$$

Then arguing as before (62) and using the Sobolev embedding $H^1 \subset C(\mathbb{T})$ as well as Lemma 12, the second term is bounded, for $a > 3/2$, by

$$\|[(L^*_b)^{-1} - (L^*_{b_0})^{-1}][f_h]\|_\infty \lesssim \|b - b_0\|_\infty(\|\hat{b}_T - b_0\|_{H^1} + \|b - b_0\|_{H^1}) = o_{P_{b_0}}(1/\sqrt{T}).$$

Similarly, noting the dependence $f_h = f_{h,b}$ on b, the term $(L^*_{b_0})^{-1}[f_{h,b}] - (L^*_{b_0})^{-1}[f_{h,b_0}]$ can be shown to be $o_{P_{b_0}}(1/\sqrt{T})$ in $C(\mathbb{T})$.

We next establish continuity of the linear operator

$$h \mapsto v_{b_0,h} = (L^*_{b_0})^{-1}[f_{h,b_0}]$$

on $C^1(\mathbb{T})$ for the norms of $(B^1_{1\infty}(\mathbb{T}))^*$ and $C(\mathbb{T})$, so that the theorem follows from Theorem 3 and the continuous mapping theorem for weak convergence, just as in the proof of Theorem 5. We use a dual representation for the weighted wavelet sequence norms characterising Besov spaces – more precisely, that the classical identities $(c_0)^* = \ell_1, (\ell_1)^* = \ell_\infty$, where $c_0 = \{(a_k) : \lim_{k\to\infty} a_k = 0\}$ is equipped with the supremum-norm on sequences, imply

$$\|g\|_{B^0_{\infty 1}} \lesssim \sup_{\phi \in C(\mathbb{T}): \|\phi\|_{B^0_{1\infty}} \leq 1} |\langle g, \phi \rangle_{L^2}|, \quad g \in C(\mathbb{T}).$$

Moreover (3) and (74) imply

$$\sup_{\|\phi\|_{B^0_{1\infty}} \leq 1} \|\mu_0 \frac{d}{dy} L^{-1}_{b_0}[\bar{\phi}]\|_{B^1_{1\infty}} \lesssim \|\mu_0\|_{Lip} \sup_{\|\phi\|_{B^0_{1\infty}} \leq 1} \|L^{-1}_{b_0}[\bar{\phi}]\|_{B^2_{1\infty}} < \infty,$$

which will be used in the following estimate. For $\bar{\phi} = \phi - \int \phi d\mu_{b_0}$, and since $v_{b_0,h} \in L_0^2(\mathbb{T}) \cap H^2 \subset C(\mathbb{T})$ in view of Lemma 12 below,

$$
\begin{aligned}
\|v_{b_0,h}\|_\infty \lesssim \|v_{b_0,h}\|_{B_{\infty 1}^0} &\lesssim \sup_{\phi \in C(\mathbb{T}):\|\phi\|_{B_{1\infty}^0} \leq 1} \left| \int v_{b_0,h}\phi \right| \\
&= \sup_{\phi \in C(\mathbb{T}):\|\phi\|_{B_{1\infty}^0} \leq 1} \left| \int (L_{b_0}^*)^{-1}[f_{h,b_0}] L_{b_0} L_{b_0}^{-1}[\bar{\phi}] \right| \\
&= \sup_{\phi \in C(\mathbb{T}):\|\phi\|_{B_{1\infty}^0} \leq 1} \left| \int f_{h,b_0} L_{b_0}^{-1}[\bar{\phi}] \right| \\
&\leq \sup_{\|\phi\|_{B_{1\infty}^0} \leq 1} \|\mu_0 \frac{d}{dy} L_{b_0}^{-1}[\bar{\phi}]\|_{B_{1\infty}^1} \|h\|_{(B_{1\infty}^1)^*} \lesssim \|h\|_{(B_{1\infty}^1)^*}.
\end{aligned}
$$

The covariance of the limiting Gaussian process is obtained as follows: since $G_{b_0}(x, y)$ is the periodic Green kernel of $L_{b_0}^{-1}$, the Green kernel of $(L_{b_0}^*)^{-1}$ is $G_{b_0}(y, x)$, and thus by the definitions and integration by parts,

$$
(L_{b_0}^*)^{-1}[f_{h,b_0}] = -\int_{\mathbb{T}} G_{b_0}(y, \cdot) \frac{d}{dy}[h\mu_0](y)dy = \int_{\mathbb{T}} \frac{d}{dy} G_{b_0}(y, \cdot) h(y)\mu_0(y)dy.
$$

Inserting for h the limit $\mathbb{W}_0 \sim \mathcal{N}_{b_0}$ of $\sqrt{T}(b - \hat{b}_T)$ gives the desired form of the limiting covariance. Finally, the limit distribution of the MAP estimate follows from the same (in fact simpler) arguments and Theorem 4.

6. Appendix: Some basic facts on the elliptic PDEs involved. We record here some basic facts about elliptic PDEs and refer to, e.g., Chapter II.3 in [6] as a reference for standard background material in the periodic setting considered here. The generator $L = L_b$ of the diffusion process given in (5) is a strongly elliptic second order partial differential operator. We will suppress the dependence on b in most of what follows; all that is required is that b is 'smooth enough', and $b \in V_J^{\otimes d}$ for a S-regular wavelet basis with S large enough will be sufficient throughout. The maximum principle for elliptic operators (see [6, 23]) implies that any (strong and then also weak) periodic solution of the Laplace equation

$$
(64) \qquad\qquad\qquad Lu = 0 \text{ on } \mathbb{T}^d
$$

equals a constant. The adjoint operator $L^* = L_b^*$ was defined in (6), and in the periodic setting considered here the operators (L, L^*) form a Fredholm pair on $L^2(\mathbb{T}^d)$, see p.175f. in [6]. As a consequence, the inhomogeneous equation

$$
(65) \qquad\qquad\qquad Lu = f, \quad f \in L^2(\mathbb{T}^d),
$$

has a solution u if and only if $\langle f, m \rangle_{L^2} = 0$ for every solution $m \in L^2(\mathbb{T}^d)$ of

$$(66) \qquad\qquad L^* m = 0 \text{ on } \mathbb{T}^d.$$

By the Fredholm property the kernel of L^* has the same dimension as the kernel of L and inspection of the form of L^* shows that the solutions $m \in L^2(\mathbb{T}^d)$ to (66) are determined up to a normalising constant. It follows that

$$(67) \qquad\qquad L^* m = 0 \iff m \in \mathcal{K} = \{c\mu : c \in \mathbb{R}\},$$

where $\mu > 0$ is the unique solution m ('invariant measure') satisfying $\int_{\mathbb{T}^d} m = 1$. Positivity of μ can be deduced from appropriate heat kernel estimates: in fact (arguing, e.g., as on p.167f. in [39]) the solution μ can be seen to be Lipschitz continuous and bounded away from zero on $\mathbb{T}^d$, and $\|\mu\|_{Lip}$ is bounded by a fixed constant that only depends on d and on an upper bound for $\|b\|_\infty$, proving in particular Proposition 1.

We can now state the following basic result for the PDE (65).

LEMMA 11.   *Let $t \geq 2$ and assume $b \in C^{t-2}(\mathbb{T}^d)$. For any $f \in L^2_\mu(\mathbb{T}^d)$, there exists a unique solution $L_b^{-1}[f] \in L^2_0(\mathbb{T}^d)$ of equation (65) satisfying $L_b L_b^{-1}[f] = f$ almost everywhere. Moreover,*

$$\|L_b^{-1}[f]\|_{H^t} \lesssim \|f\|_{H^{t-2}},$$

with constants depending on t, d and on an upper bound B for $\|b\|_{B^{t-2}_{\infty\infty}}$.

PROOF. By standard Sobolev space theory and definition of the Laplacian we have for any $u \in \mathcal{H} \equiv H^t \cap \{u : \langle u, 1 \rangle_{L^2} = 0\}$ the inequality

$$(68) \qquad\qquad \|u\|_{H^t} \lesssim \|\Delta u\|_{H^{t-2}}.$$

Indeed, for $\{e_k : k = (k_1, \ldots, k_d) \in \mathbb{Z}^d\}$ the usual trigonometric basis of $L^2(\mathbb{T}^d)$ we have $\langle u, e_0 \rangle_{L^2} = \langle u, 1 \rangle_{L^2} = 0$, $\langle \Delta u, e_k \rangle = -(2\pi)^2 \sum_j k_j^2 \langle u, e_k \rangle$, and $\sup_{k \neq 0}(1 + \|k\|^2)/\|k\|^2 < \infty$, which gives the result using the characterisation of Sobolev norms in the basis $\{e_k\}$. We then also have, by the triangle inequality and (3),

$$(69) \quad \|u\|_{H^t} \lesssim \|Lu\|_{H^{t-2}} + \|b.\nabla u\|_{H^{t-2}} \lesssim \|Lu\|_{H^{t-2}} + \|b\|_{B^{t-2}_{\infty\infty}} \|u\|_{H^{t-1}}$$

for all $u \in \mathcal{H}$, with constants depending on t, d. We now deduce from this the inequality

$$(70) \qquad\qquad \|u\|_{H^t} \lesssim \|Lu\|_{H^{t-2}} \quad \forall u \in \mathcal{H}.$$

Indeed, if the latter inequality does not hold true, then there exists a sequence $u_m \in \mathcal{H}$ such that $\|u_m\|_{H^t} = 1$ for all m but $\|Lu_m\|_{H^{t-2}} \to 0$ as $m \to \infty$. At the same time, by compactness, u_m converges in $\|\cdot\|_{H^{t-1}}$-norm (if necessary along a subsequence) to some $u \in \mathcal{H}$ satisfying $Lu = 0$. Using (69) with fixed constant depending only on B, t, d, we see that u_m is also Cauchy in H^t, and its limit must necessarily satisfy $\|u\|_{H^t} = 1$. However, as remarked after (64), the only solution $u \in \mathcal{H}$ to $Lu = 0$ on $\mathbb{T}^d$ equals $u = const = 0$, a contradiction to $\|u\|_{H^t} = 1$, proving (70).

By the Fredholm property and (67), a solution u_f to (65) exists whenever $\int f d\mu = 0$, and for $f \in H^{t-2}(\mathbb{T}^d)$ any such solution belongs to $H^t(\mathbb{T}^d)$ (see Theorem 3.5.3 in [6], which is proved for smooth b, but the proof remains valid for $b \in C^{t-2}(\mathbb{T}^d)$). The weak maximum principle (p.179 in [23]) now implies that u_f is unique up to an additive constant, and applying (70) to the unique selection $u_f = L^{-1}[f] \in \mathcal{H}$ completes the proof. □

We next obtain corresponding results for the adjoint PDE. It follows from (67) that the unique element $m \in \mathcal{K}$ satisfying $\int_{\mathbb{T}^d} m = 0$ must necessarily vanish identically, and we can study the solution operator $(L^*)^{-1}$ of the inhomogeneous adjoint PDE

$$(71) \qquad\qquad L^* u = f \text{ on } \mathbb{T}^d,$$

which assigns to any $f \in L_0^2(\mathbb{T}^d)$ the unique solution $u = (L^*)^{-1}[f] \in L_0^2(\mathbb{T}^d)$. Indeed, using the Fredholm property from Section 3.6 in [6] in a reverse way (with L equal to our L^* so that the new L^* is our $(L^*)^* = L$), we see that solutions $u = u_f$ to (71) exist for any periodic f for which $\int f = 0$ (since solutions to $Lu = 0$ equal constants), and if u_1, u_2 are two such solutions, so that $L^*(u_1 - u_2) = 0$ and $\int u_1 = \int u_2$, then necessarily $u_1 = u_2$ by what precedes.

LEMMA 12.   *Let $t \geq 2$ and assume $b \in C^{t-1}(\mathbb{T}^d)$. Then for any $f \in H^{t-2}(\mathbb{T}^d) \cap L_0^2(\mathbb{T}^d)$, we have $(L_b^*)^{-1}[f] \in H^t(\mathbb{T}^d)$ and*

$$\|(L_b^*)^{-1}[f]\|_{H^t} \lesssim \|f\|_{H^{t-2}},$$

with constants depending on t, d and on an upper bound B for $\|b\|_{B^{t-1}_{\infty\infty}}$.

PROOF.   The proof is similar to the one of Lemma 11 after deriving the basic inequality

$$(72) \qquad \begin{aligned} \|u\|_{H^t} &\lesssim \|L^* u\|_{H^{t-2}} + \|b.\nabla u + div(b)u\|_{H^{t-2}} \\ &\lesssim \|L^* u\|_{H^{t-2}} + \|b\|_{B^{t-2}_{\infty\infty}}\|u\|_{H^{t-1}} + \|b\|_{B^{t-1}_{\infty\infty}}\|u\|_{H^{t-2}} \end{aligned}$$

in analogy to (69).   □

We can also give a version of Lemma 12 with $t = 0$. Since $(L_b^*)^{-1}[f] \in L_0^2(\mathbb{T}^d)$ we have for all $f \in L_0^2(\mathbb{T}^d)$ and $\bar{\phi} = \phi - \int \phi d\mu_b$ the estimate

$$\|(L_b^*)^{-1}[f]\|_{L^2} = \sup_{\|\phi\|_{L^2} \le 1} \left| \int (L_b^*)^{-1}[f] L_b L_b^{-1}[\bar{\phi}] \right| = \sup_{\|\phi\|_{L^2} \le 1} \left| \int f L_b^{-1}[\bar{\phi}] \right|$$

$$(73) \qquad\qquad \le \|f\|_{H^{-2}} \sup_{\|\phi\|_{L^2} \le 1} \|L_b^{-1}[\bar{\phi}]\|_{H^2} \lesssim \|f\|_{H^{-2}},$$

where we have used Lemma 11 with $t = 2$ in the last inequality and where the constants in the last inequality depend only on d and on bounds for $\|b\|_{B^1_{\infty\infty}}$ and $\|\mu_b\|_{L^2}$.

6.0.1. *Refinements on the Besov scale.* For the the proofs of Theorems 5 and 6 we need more refined regularity estimates for the solutions of the PDE involved, replacing the Sobolev norms in Lemma 11 by appropriate Besov norms. The inequality

$$(74) \qquad\qquad \|L_b^{-1}[f]\|_{B^t_{1\infty}} \lesssim \|f\|_{B^{t-2}_{1\infty}}, \ t - 2 \ge 0, \forall f \in L^2_{\mu_b}(\mathbb{T}^d),$$

with constants depending on b only via a bound B for $\|b\|_{B^{t-1}_{\infty\infty}}$, is proved in the same way as Lemma 11, replacing the basic inequality (68) by its analogue for Besov norms

$$(75) \qquad\qquad \|u\|_{B^t_{1\infty}} \lesssim \|\Delta u\|_{B^{t-2}_{1\infty}} \quad \forall u \in B^t_{1\infty} \cap \{u : \langle u, e_0 \rangle = 0\},$$

which is proved as follows: for all u such that $\langle u, e_0 \rangle = 0$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, an equivalent Littlewood-Paley norm on any Besov space $B^r_{1\infty}$ is given by

$$\|u\|_{B^r_{1\infty}} = \sup_{j \in \mathbb{N}_0} 2^{jr} \big\| \sum_{k \in \mathbb{Z}, k \ne 0} \psi_j(k) \langle u, e_k \rangle e_k \big\|_{L^1(\mathbb{T}^d)},$$

where the $\psi_j = \psi(\cdot/2^j), \mathrm{supp}(\psi) \in (1/2, 2)^d$ form a Littlewood-Paley resolution of unity, see p.162f. in [50]. Then as after (68),

$$\|u\|_{B^t_{1\infty}} = \sup_{j \in \mathbb{N}_0} 2^{jt} \big\| \sum_{k \in \mathbb{Z}, k \ne 0} \frac{1}{4\pi^2 \|k\|^2} \psi_j(k) \langle \Delta u, e_k \rangle e_k \big\|_{L^1(\mathbb{T}^d)}$$

$$= \sup_{j \in \mathbb{N}_0} 2^{j(t-2)} \big\| \sum_{k \in \mathbb{Z}, k \ne 0} M_j(k) \psi_j(k) \langle \Delta u, e_k \rangle e_k \big\|_{L^1(\mathbb{T}^d)}$$

where $M_j = M(\cdot/2^j)$ and $M = \Phi/(4\pi^2 \|\cdot\|^2)$ with Φ a smooth function supported in $(1/4, 9/4)^d$ such that $\Phi = 1$ on $(1/2, 2)^d$. By a standard Fourier

multiplier inequality (e.g., Lemma 4.3.27 in [26], which easily generalises to $d > 1$) the last norm can be estimated by

$$\sup_{j\in\mathbb{N}_0} 2^{j(t-2)} \big\| \sum_{k\in\mathbb{Z}, k\neq 0} \psi_j(k)\langle \Delta u, e_k\rangle e_k \big\|_{L^1(\mathbb{T}^d)} \times \|F^{-1}M_j\|_{L^1(\mathbb{R}^d)},$$

where F^{-1} is the inverse Fourier transform. Since Φ is smooth and supported in $(-1/4, 3/4)^d$, both M and $F^{-1}M$ belong to the Schwartz-class $\mathcal{S}$, so that (75) follows from

$$\sup_j \|F^{-1}[M_j]\|_{L^1(\mathbb{R}^d)} = \|F^{-1}[M]\|_{L^1(\mathbb{R}^d)} < \infty.$$

References.

[1] ABRAHAM, K. Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. *Bernoulli, to appear, arXiv:1802.05635* (2018).

[2] AECKERLE-WILLEMS, C., AND STRAUCH, C. Concentration of scalar ergodic diffusions and some statistical implications. *arXiv:1807.11331* (2018).

[3] AECKERLE-WILLEMS, C., AND STRAUCH, C. Sup-norm adaptive simultaneous drift estimation for ergodic diffusions. *arXiv:1808.10660* (2018).

[4] AGAPIOU, S., LARSSON, S., AND STUART, A. M. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Process. Appl. 123*, 10 (2013), 3828–3860.

[5] BASS, R. F. *Stochastic processes.* Cambridge Univ. Press, Cambridge, 2011.

[6] BERS, L., JOHN, F., AND SCHECHTER, M. *Partial differential equations.* Lectures in Applied Mathematics, Vol. III. Wiley & Sons, Inc. New York-London-Sydney, 1964.

[7] BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O., AND FEARNHEAD, P. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol. 68*, 3 (2006), 333–382.

[8] BLADT, M., FINCH, S., AND SØRENSEN, M. Simulation of multivariate diffusion bridges. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 78*, 2 (2016), 343–369.

[9] CASTILLO, I. On Bayesian supremum norm contraction rates. *Ann. Statist. 42*, 5 (2014), 2058–2091.

[10] CASTILLO, I. Pólya tree posterior distributions on densities. *Ann. Inst. Henri Poincaré Probab. Stat. 53*, 4 (2017), 2074–2102.

[11] CASTILLO, I., AND NICKL, R. Nonparametric Bernstein–von Mises Theorems in Gaussian white noise. *Ann. Statist. 41*, 4 (2013), 1999–2028.

[12] CASTILLO, I., AND NICKL, R. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist. 42*, 5 (2014), 1941–1969.

[13] CASTILLO, I., AND ROUSSEAU, J. A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist. 43*, 6 (2015), 2353–2383.

[14] DALALYAN, A. Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist. 33*, 6 (2005), 2507–2528.

[15] Dalalyan, A., and Reiss, M. Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Related Fields 134*, 2 (2006), 248–282.

[16] Dalalyan, A., and Reiss, M. Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probab. Theory Related Fields 137*, 1-2 (2007), 25–47.

[17] Dashti, M., Law, K. J. H., Stuart, A. M., and Voss, J. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems 29*, 9 (2013), 095017, 27.

[18] Dudley, R. M. *Real analysis and probability.* Cambridge University Press, Cambridge, 2002.

[19] Dudley, R. M. *Uniform central limit theorems*, second ed. Cambridge University Press, New York, 2014.

[20] Ethier, S. N., and Kurtz, T. G. *Markov processes, Characterization and convergence.* John Wiley & Sons, Inc., New York, 1986.

[21] Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. Convergence rates of posterior distributions. *Ann. Statist. 28*, 2 (2000), 500–531.

[22] Ghosal, S., and van der Vaart, A. W. *Fundamentals of Nonparametric Bayesian Inference.* Cambridge University Press, New York, 2017.

[23] Gilbarg, D., and Trudinger, N. S. *Elliptic partial differential equations of second order.* Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[24] Giné, E., and Nickl, R. Uniform central limit theorems for kernel density estimators. *Probab. Theory Related Fields 141*, 3-4 (2008), 333–387.

[25] Giné, E., and Nickl, R. An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields 143*, 3-4 (2009), 569–596.

[26] Giné, E., and Nickl, R. *Mathematical foundations of infinite-dimensional statistical models.* Cambridge University Press, New York, 2016.

[27] Giordano, M., and Kekkonen, H. Bernstein-von Mises theorems and uncertainty quantification for linear inverse problems. *arXiv:1811.04058* (2018).

[28] Gugushvili, S., and Spreij, P. Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. *Lith. Math. J. 54*, 2 (2014), 127–141.

[29] Horn, R. A., and Johnson, C. R. *Matrix analysis*, second ed. Cambridge University Press, Cambridge, 2013.

[30] Knapik, B., and Salomond, J.-B. A general approach to posterior contraction in nonparametric inverse problems. *Bernoulli 24*, 3 (2018), 2091–2121.

[31] Knapik, B., van der Vaart, A. W., and van Zanten, J. H. Bayesian inverse problems with Gaussian priors. *Ann. Statist. 39*, 5 (2011), 2626–2657.

[32] Kutoyants, Y. A. *Statistical inference for ergodic diffusion processes.* Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2004.

[33] Löcherbach, E., Loukianova, D., and Loukianov, O. Penalized nonparametric drift estimation for a continuously observed one-dimensional diffusion process. *ESAIM Probab. Stat. 15* (2011), 197–216.

[34] Monard, F., Nickl, R., and Paternain, G. P. Efficient nonparametric Bayesian inference for X-ray transforms. *Ann. Statist. 47*, 2 (2019), 1113–1147.

[35] Nickl, R. Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation. *Journal of the European Mathematical Society, to appear; arXiv:1707.01764* (2017).

[36] Nickl, R., and Söhl, J. Bernstein-von Mises theorems for statistical inverse problems II: compound Poisson processes. *arXiv:1709.07752* (2017).

[37] NICKL, R., AND SÖHL, J. Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist. 45*, 4 (2017), 1664–1693.

[38] NICKL, R., VAN DE GEER, S., AND WANG, S. Convergence rates for penalised least squares estimators in PDE-constrained regression problems. *arXiv:1809.08818* (2018).

[39] NORRIS, J. R. Long-time behaviour of heat flow: global estimates and exact asymptotics. *Arch. Rational Mech. Anal. 140*, 2 (1997), 161–195.

[40] PAPASPILIOPOULOS, O., POKERN, Y., ROBERTS, G. O., AND STUART, A. M. Nonparametric estimation of diffusions: a differential equations approach. *Biometrika 99*, 3 (2012), 511–531.

[41] PARDOUX, E., AND VERETENNIKOV, A. Y. On the Poisson equation and diffusion approximation. I. *Ann. Probab. 29*, 3 (2001), 1061–1085.

[42] POKERN, Y., STUART, A. M., AND VAN ZANTEN, J. H. Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl. 123*, 2 (2013), 603–628.

[43] RAY, K. Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat. 7* (2013), 2516–2549.

[44] RAY, K. *Asymptotic theory for Bayesian nonparametric procedures in inverse problems*. PhD thesis, University of Cambridge, 2014.

[45] RAY, K. Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist. 45*, 6 (2017), 2511–2536.

[46] REVUZ, D., AND YOR, M. *Continuous martingales and Brownian motion*, third ed., vol. 293. Springer-Verlag, Berlin, 1999.

[47] RIVOIRARD, V., AND ROUSSEAU, J. Bernstein-von Mises theorem for linear functionals of the density. *Ann. Statist. 40*, 3 (2012), 1489–1523.

[48] RUNST, T., AND SICKEL, W. *Sobolev spaces of fractional order, Nemytskij operators, and nonlinear partial differential equations*. Walter de Gruyter & Co., Berlin, 1996.

[49] SCHAUER, M., VAN DER MEULEN, F., AND VAN ZANTEN, H. Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli 23*, 4A (2017), 2917–2950.

[50] SCHMEISSER, H.-J., AND TRIEBEL, H. *Topics in Fourier analysis and function spaces*. A Wiley-Interscience Publication. John Wiley & Sons, Ltd., Chichester, 1987.

[51] SCHMISSER, E. Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics 47*, 1 (2013), 61–84.

[52] STRAUCH, C. Sharp adaptive drift estimation for ergodic diffusions: the multivariate case. *Stochastic Process. Appl. 125*, 7 (2015), 2562–2602.

[53] STRAUCH, C. Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions. *Probab. Theory Related Fields 164*, 1-2 (2016), 361–400.

[54] STRAUCH, C. Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist. 46*, 6B (2018), 3451–3480.

[55] VAN DER MEULEN, F., AND SCHAUER, M. Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat. 11*, 1 (2017), 2358–2396.

[56] VAN DER MEULEN, F., AND VAN ZANTEN, H. Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli 19*, 1 (2013), 44–63.

[57] VAN DER MEULEN, F. H., VAN DER VAART, A. W., AND VAN ZANTEN, J. H. Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli 12*, 5 (2006), 863–888.

[58] VAN DER VAART, A. W. *Asymptotic statistics*. Cambridge Univ. Press, Cambridge, 1998.

[59] VAN DER VAART, A. W., AND VAN ZANTEN, J. H. Rates of contraction of posterior

distributions based on Gaussian process priors. *Ann. Statist. 36*, 3 (2008), 1435–1463.

[60] van Waaij, J., and van Zanten, H. Gaussian process methods for one-dimensional diffusions: optimal rates and adaptation. *Electron. J. Stat. 10*, 1 (2016), 628–645.

[61] Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. *arXiv:1011.3027* (2010).

Statistical Laboratory
Department of Pure Mathematics
 and Mathematical Statistics
University of Cambridge
CB3 0WB Cambridge
United Kingdom
E-mail: r.nickl@statslab.cam.ac.uk

Department of Mathematics
King's College London
Strand
London WC2R 2LS
United Kingdom
E-mail: kolyan.ray@kcl.ac.uk