

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

BY RICHARD NICKL<sup>1</sup> AND KOLYAN RAY<sup>2</sup>

<sup>1</sup>Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, [rnickl@statslab.cam.ac.uk](mailto:rnickl@statslab.cam.ac.uk)

<sup>2</sup>Department of Mathematics, King's College London, [kolyan.ray@kcl.ac.uk](mailto:kolyan.ray@kcl.ac.uk)

The problem of determining a periodic Lipschitz vector field  $b = (b_1, \dots, b_d)$  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  $\mu_b$ . The limiting Gaussian process distributions have a covariance structure that is asymptotically optimal from an information-theoretic point of view.

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

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

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  $\mu_b$  of the diffusion, which in the *ergodic* case describes the probabilities

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

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., [1–3, 13, 14, 29, 30, 39, 52, 55]), 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 [15] first obtained convergence rates of multivariate

---

Received October 2018; revised March 2019.

*MSC2010 subject classifications.* Primary 62G20; secondary 62F15, 65N21.

*Key words and phrases.* Penalised least squares estimator, asymptotics of nonparametric Bayes procedures, Bernstein–von Mises theorem, uncertainty quantification.

nonparametric kernel-type estimators. Schmisser [48] established adaptive  $L^2$ -convergence rates of certain model selection based projection estimators and Strauch [49–51] obtained adaptive convergence rate results for  $b$  in pointwise and  $L^2$ -loss, and for  $\mu_b$  in  $\|\cdot\|_\infty$ -loss; more discussion can be found below.

For observations  $(X_t : 0 \leq t \leq T)$ , 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 [6, 7, 37, 46, 53] 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  $d \leq 4$ . When  $d \leq 3$ , we further prove nonparametric Bernstein–von Mises theorems that establish asymptotic normality of the re-centred and scaled posterior distributions  $\sqrt{T}(b - \hat{b}_T)|(X_t : 0 \leq t \leq T)$  in a (large enough) function space. From this, we deduce central limit theorems for the implied plug-in estimators for the invariant density  $\mu_b$ . The proofs imply that the limiting covariances obtained coincide with the semiparametric information lower bounds for these estimation problems. We exploit that the nonlinear identification map  $b \mapsto \mu_b$  can be shown to be "one-smoothing". As inference on  $b$  is approximately a nonparametric regression problem [15], this offers an analytical explanation for why the invariant density  $\mu_b$  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  $d \geq 2$  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, 14, 39, 52, 55], 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  $d \geq 2$  the theory of elliptic partial differential equations (PDEs) is required. PDE techniques are an effective alternative to the functional inequalities used in [15, 50, 51], in particular the requirement that  $b$  be a gradient vector field  $\nabla B$  for some  $B : \mathbb{R}^d \rightarrow \mathbb{R}$ , and thus of *reversibility* of  $(X_t)$ —used in the references [15, 48, 50, 51]—can be avoided this way; neither does  $X_0 \sim \mu_b$  have to be started in equilibrium as in [48]. 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 [15, 50, 51]. The techniques of the present paper extend in principle, albeit at the expense of considerable technicalities, to the nonperiodic case if  $b$  is known outside of a compact subset of  $\mathbb{R}^d$  and upon employing assumptions on  $b$  as in [38]. 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  $\mu_b$  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., [1, 34]), Bayesian inference thus cannot be based on a prior assigned directly to the invariant measure  $\mu_b$ . 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  $\mu_b$ .

Our proofs employ techniques from Bayesian nonparametric and semiparametric statistics, specifically [8, 10–12], and are related to recent results on Bayesian inverse problems [4, 16, 27, 28, 36, 40], Bernstein–von Mises theorems [24, 31, 33, 35, 41, 42, 44] and diffusion models [1, 25, 34, 39, 52, 54, 55].

## 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  $\mu$  be a probability measure on  $\mathbb{T}^d$ . If its Lebesgue density, also denoted by  $\mu$ , 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  $\langle f, g \rangle_\mu = \int fg d\mu$  for  $f, g \in L^2(\mathbb{T}^d)$ . The symbol  $L^2_0(\mathbb{T}^d)$  denotes the subspace of functions  $f$  for which  $\int_{\mathbb{T}^d} f(x) dx = 0$ , and  $L^2_\mu(\mathbb{T}^d)$  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 [47] 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  $f = (f_1, \dots, f_d)$ . 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  $\|f\|_{H^s}^2 = \sum_{j=1}^d \|f_j\|_{H^s}^2$ . We shall repeatedly use multiplication inequalities for Besov–Sobolev norms,

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

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

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

(see Section 4.3 of [23]), where the base Daubechies wavelets are taken “ $S$ -regular”,  $S \in \mathbb{N}$ . The dimension of  $V_J$  is  $O(2^{Jd})$  as  $J \rightarrow \infty$ , 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  $H^s(\mathbb{T}^d)$  (page 370f in [23]).

If  $\mu$  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  $\mu$ , also called the law  $\mathcal{L}(Z) = \mu$  of  $Z$ . We write  $Z_T \rightarrow^d Z$ , or  $Z_T \rightarrow^d \mathcal{L}(Z)$  when no confusion can arise, to denote the usual notion of weak convergence of the laws  $\mathcal{L}(Z_T) \rightarrow \mathcal{L}(Z)$  as  $T \rightarrow \infty$ ; see, for example, Chapter 11 in [17].

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

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

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 nonnegative 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  $X_t = (X_t^1, \dots, X_t^d)$ . We denote by  $P_b = P_b^x$  the cylindrical probability measure describing the law of  $(X_t)$  in path space  $C([0, \infty) \rightarrow \mathbb{R}^d)$  when  $X_0 = x$ ; its restriction  $P_b^T = P_b^{T,x}$  to the separable space  $C([0, T] \rightarrow \mathbb{R}^d)$  describes the law of the process  $X^T \equiv (X_t : t \in [0, T])$  until time  $T$ ; see, for example, 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  $b : \mathbb{T}^d \rightarrow \mathbb{R}^d$  from an observed trajectory  $X^T$ . The periodic model (which has also been used in [39, 52] 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),

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

where  $\mu_b$  is a uniquely defined probability measure on  $\mathbb{T}^d$  and where we identify  $\varphi$  with its periodic extension to  $\mathbb{R}^d$  on the left-hand side. The measure  $\mu_b$  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  $\mu$  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  $\varphi$ . 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  $\varphi$ , and hence necessarily  $L^* \mu = 0$  (in the weak sense), which can be used to identify  $\mu$  via the adjoint generator  $L^*$ .

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

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

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

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

so that  $\mu_b$  can be identified as the (weak) solution of the PDE

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

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  $\mu_b$  to (7) satisfying  $\int_{\mathbb{T}^d} d\mu_b = 1$  exists. Moreover,  $\mu_b$  is Lipschitz continuous and bounded away from zero on  $\mathbb{T}^d$ , with  $\|1/\mu_b\|_\infty$  and the Lipschitz constant  $\|\mu_b\|_{\text{Lip}}$  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  $\mu_b$  a weak solution suffices.

If  $b$  arises as a gradient vector field  $\nabla B$  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  $\mu_b$  via  $b = (1/2)\nabla \log \mu_b$ . However, the invariant measure  $\mu_b$  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  $\text{div}(v\mu) = 0$ ), 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  $\phi$ , and as a consequence  $\mu_b$  is also the invariant measure for  $L_{b+v}$ . Thus any statistical approach to recover  $b$  via first estimating  $\mu_b$  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  $b : \mathbb{T}^d \rightarrow \mathbb{R}^d$ ,

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

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

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

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

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

Our proofs imply that the denominator in the last expression is finite and nonzero with probability approaching one under the law of  $X^T$  as  $T \rightarrow \infty$ . The map  $(b, c) \mapsto \int_0^T b(X_t) \cdot c(X_t) dt + \langle b, c \rangle_{\mathbb{H}}$  induces an inverse covariance  $D_H^{-1}$  on some linear subspace  $H \subset V_J^{\otimes d}$ . [Since  $1 \in V_J^{\otimes d}$ , we have  $\dim H \neq 0$  a.s., and our proofs imply in fact that  $H = V_J^{\otimes d}$  with probability approaching one as  $T \rightarrow \infty$ .] By characterisations of Gaussian laws (e.g., Theorem 9.5.7 in [17]) and linearity of  $b \mapsto \int_0^T b(X_t) \cdot dX_t$ , 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  $b = (b_1, \dots, b_d)$  of the form:

$$(11) \quad b_j = \sum_{l \leq J} \sum_{r=0}^{2^{ld}-1} \sigma_{lgl,r,j} \Phi_{l,r}, \quad gl,r,j \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), \quad j = 1, \dots, d,$$

where the  $\Phi_{l,r}$  form a  $S$ -regular periodised wavelet basis of  $L^2(\mathbb{T}^d)$  (cf. (4)), where  $J = J_T \rightarrow \infty$  as  $T \rightarrow \infty$  in a way to be chosen below, and where the weights  $\sigma_l$  govern the regularisation prescribed by the penalty functional. We will tacitly assume throughout that  $S$  is large enough (depending on parameters  $s, a, \alpha$  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 (page 75 in [23]) 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 [20, 55]. 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 [39, 52, 55], 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  $\Pi_T$  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  $M_T \rightarrow \infty$ , as  $T \rightarrow \infty$ ,*

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

*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  $\alpha$ , 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., [1, 40]) and is left to the reader.

From the previous theorem, and imposing slightly stronger conditions on  $b_0$  and  $\Pi_T$ , 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 [8] (see also [9, 11, 12, 33]), 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  $\Pi_T$  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  $\delta > 5/2$ ,

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

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

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

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

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

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

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

**PROOF.** Consider the function

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

The posterior is a Gaussian measure on the finite-dimensional space  $V_J^{\otimes d}$ , centred 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 [23]) yields that  $\hat{b}_T$  is a maximizer of  $H$ . Using Theorem 2.5 in [19] 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  $T \rightarrow \infty$ . 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, for example, via the asymptotic equivalence results in [15], see also [49, 50]). The  $\|\cdot\|_\infty$ -rates are then also optimal whenever  $d \leq 4$ , 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 [10, 11]; see also the recent contributions [9, 31, 33, 35, 42]. 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  $\sqrt{T}(b - \hat{b}_T) \mid X^T$  as (conditionally on  $X^T$ ) random vector fields acting linearly on test functions  $\phi = (\phi_1, \dots, \phi_d)$  by integration

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

and show that a Bernstein–von Mises theorem holds true uniformly in  $\phi$  belonging to any bounded subset of the Besov space  $B_{1\infty}^\rho$ ,  $\rho > d/2$ ,  $d \leq 3$ . 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  $\mu_b$  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  $\tau, \tau'$  on a metric space  $(S, e)$ , define the bounded Lipschitz (BL) metric for weak convergence (page 157 in [18]) by

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

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

**THEOREM 3.** *Let  $1 \leq d \leq 3$ ,  $\rho > d/2$ ,  $a > \max(3d/2 - 1, 1)$  and let  $s \geq a$  be such that  $s > a - 1 + d/2$ . Suppose  $b_0 \in C^s(\mathbb{T}^d) \cap H^s(\mathbb{T}^d)$ . Let  $\Pi_T$  be the Gaussian prior from (11) with  $\sigma_l = 2^{-l(\alpha+d/2)}$ ,  $0 \leq \alpha < a \wedge s - d/2$  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  $T \rightarrow \infty$ ,*

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

As in related situations in [10, 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  $T \rightarrow \infty$ ,*

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

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 [10, 11] 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 semiparametric 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 [22]).

**2.5. Bayesian inference on the invariant measure.** We now turn to the problem of making inference on the invariant measure  $\mu_b$ . Frequentist estimators of  $\mu_b$  can be suggested directly based on (2), for example [51]. For the Bayesian statistician, modelling  $\mu_b$  directly by a prior is not coherent since  $\mu_b$  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  $\mu_b$  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*

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

as  $T \rightarrow \infty$ , 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

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

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  $H^{-r}(\mathbb{T}^d) = (H^r(\mathbb{T}^d))^*$ :

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

**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, [21]) 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 [21]). 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

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

as  $T \rightarrow \infty$ , 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 [23] for a precise definition of  $\beta_S$  for nonseparable  $S$ ). One further deduces that the estimated invariant probabilities induced by the MAP estimate  $\hat{b}_T$  obey the limit law

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

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  $L_b^{-1}[g] = \int_{\mathbb{T}} G_b(\cdot, y)g(y) dy$ ,  $g \in L^2_{\mu_b}(\mathbb{T})$ , with periodic Green kernel  $G_b : \mathbb{T} \times \mathbb{T} \rightarrow \mathbb{R}$  such that  $G_b(\cdot, x) \in H^1(\mathbb{T})$  for all  $x \in \mathbb{T}$ . This follows, for example, 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  $T \rightarrow \infty$ ,*

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

$$\sqrt{T}(\mu_{\hat{b}_T} - \mu_0) \xrightarrow{d} \bar{N}_{b_0} \quad \text{in } C(\mathbb{T}),$$

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

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

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 [22].

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 ([56], 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 nonlinear 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 peri-

odic 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  $\mathcal{F}_T \subset L^2_{\mu_0}(\mathbb{T}^d) \cap H^{d/2+\kappa}(\mathbb{T}^d)$  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

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

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

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

and for any  $x > 0$ ,

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

PROOF. By Lemma 11 and the Sobolev embedding theorem, the Poisson equation  $Lu = L_{b_0}u = f$  has a unique solution  $L^{-1}[f] \in H^{d/2+\kappa+2} \cap L^2_0 \subset C^2$  satisfying  $LL^{-1}[f] = f$  for any  $f \in \mathcal{F}_T$ . We may therefore define for  $f \in \mathcal{F}_T$ ,

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

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

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

Recall Bernstein's inequality for continuous local martingales (page 153 of [43]): if  $M$  is a continuous local martingale vanishing at 0 with quadratic variation  $[M]$ , then for any stopping time  $T$  and any  $y, K > 0$ ,

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

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

$$\begin{aligned} P_{b_0}(|Z_T(f)| \geq \sqrt{T}x) &= P_{b_0}(|Z_T(f)| \geq \sqrt{T}x, [Z_*(f)]_T \leq Td_L^2(f, 0)) \\ &\leq 2e^{-\frac{x^2}{2d_L^2(f, 0)}}. \end{aligned}$$

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

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

a nonasymptotic inequality. The process  $(T^{-1/2}Z_T(f) : f \in \mathcal{F}_T)$  is thus mean-zero and sub-Gaussian with respect to  $d_L$ . From this, we deduce the expectation bound  $E_{b_0} \sup_{f \in \mathcal{F}_T} T^{-1/2}|Z_T(f)| \leq 4\sqrt{2}J_{\mathcal{F}_T}$  by the usual chaining method for sub-Gaussian processes (e.g., Theorem 2.3.7 of [23]—the factor 6 scales the sub-Gaussian constant, see after Definition 2.3.5 of [23]). This chaining bound extends to exponential  $\psi_2$ -Orlicz norms  $\|\cdot\|_{\psi_2}$  (see Exercise 2.3.1 of [23]), so one further has  $\|\sup_{f \in \mathcal{F}_T} T^{-1/2}|Z_T(f)|\|_{\psi_2} \leq 16\sqrt{6}J_{\mathcal{F}_T}$ . Using Lemma 2.3.1 of [23] and that for any random variable  $X$ ,  $\|X - \mathbb{E}X\|_{\psi_2} \leq 2\|X\|_{\psi_2}$ , we obtain for any  $x > 0$ ,

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

Using the expectation bound just derived, the above inequality yields

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

Combining the above gives the required sub-Gaussian inequality.  $\square$

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

LEMMA 2 ([45], page 345). For  $t > 0$  and any bounded  $f, g \in H^t(\mathbb{T}^d)$ ,

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

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  $C = C(p, d, \Phi, \|\mu_0\|_\infty)$ ,

$$\|\gamma\rho - \langle \gamma, \rho \rangle_{\mu_0}\|_{H^p} \leq C2^{J(p+d/2)}\|\gamma\|_{L^2}\|\rho\|_{L^2}.$$

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

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

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,

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

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

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

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

3.2. *A restricted isometry inequality for  $h_T^2(\cdot, \cdot)$  on  $V_J^{\otimes d}$ .* We next consider 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 [58]), 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  $x > 0$ ,*

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

PROOF. Let  $b_j, \bar{b}_j \in V_J$  and write  $b = (b_1, \dots, b_d)$ ,  $\bar{b} = (\bar{b}_1, \dots, \bar{b}_d)$  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

$$\begin{aligned} h_T^2(b, \bar{b}) &= \frac{1}{T} \sum_{j=1}^d \int_0^T \left( \sum_{l,r} (\theta_{l,r,j} - \bar{\theta}_{l,r,j}) \Phi_{l,r}(X_s) \right)^2 ds \\ &= \sum_{j=1}^d \sum_{l,r} \sum_{l',r'} (\theta_{l,r,j} - \bar{\theta}_{l,r,j}) (\theta_{l',r',j} - \bar{\theta}_{l',r',j}) \\ &\quad \times \frac{1}{T} \int_0^T \Phi_{l,r}(X_s) \Phi_{l',r'}(X_s) ds \\ &= \sum_{j=1}^d (\theta_{j,\cdot} - \bar{\theta}_{j,\cdot})^T \hat{\Gamma} (\theta_{j,\cdot} - \bar{\theta}_{j,\cdot}), \end{aligned}$$

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  $v_J \times v_J$  symmetric 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 right-hand side of 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

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

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  $\delta$ -covering of  $\Theta$  in  $\|\cdot\|_{\Gamma}$ -distance. For every  $u \in \Theta$ , let  $u^l = u^l(u)$  denote the closest point in this  $\delta$ -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,  $\|u\|_{\mathbb{R}^{v_J}} = \|g_u\|_{L^2} \leq \|1/\mu_0\|_{\infty}^{1/2} \times \|g_u\|_{\mu_0} = \|1/\mu_0\|_{\infty}^{1/2} \|u\|_{\Gamma}$ . For  $(\lambda_i)_{i=1}^{v_J}$  the eigenvalues of the symmetric matrix  $\Lambda$  and  $\lambda_{\max} = \max_i |\lambda_i|$ , applying the Cauchy–Schwarz inequality gives

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

where the last inequality follows from page 234 of [26]. 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  $\delta_0^2 + 2\|1/\mu_0\|_{\infty} \|\mu_0\|_{\infty} \delta_0 \leq 1/2$ ,

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

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 [23]).

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

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

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  $\kappa$  as above and any  $u \in \Theta$ . Substituting this into (20) gives

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

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

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 [20], and formulated in the Brownian semimartingale setting relevant here by van der Meulen et al. [55]), 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 [55], restated in the present context. The proof relies on martingale arguments which generalize to the multi-dimensional 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  $\sigma$ -algebras satisfying mild measurability conditions; see Section 2 of [55]. In particular, these are satisfied by the finite-dimensional spaces considered in Theorem 1.

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

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

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

$$(23) \quad \Pi_T(b : \|b - b_0\|_{\mu_0} \leq \varepsilon_T) \geq e^{-C_3 T \varepsilon_T^2}.$$

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

$$(24) \quad \begin{aligned} \liminf_{T \rightarrow \infty} P_{b_0}(c_\gamma \|b - b_0\|_{\mu_0} \leq h_T(b, b_0), \forall b \in \mathfrak{B}_T \\ \text{s.t. } h_T(b, b_0) \geq D_\gamma \varepsilon_T, \text{ and} \\ h_T(b_1, b_2) \leq C_\gamma \|b_1 - b_2\|_{\mu_0}, \forall b_1, b_2 \in \mathfrak{B}_T \\ \text{s.t. } h_T(b_1, b_2) \geq D_\gamma \varepsilon_T) \geq 1 - \gamma. \end{aligned}$$

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

The proof of the theorem implies in particular that the denominator in (10) is nonzero 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 [57] (see also Sections 11.3 and 11.4.5 in [20]), there exist measurable sets  $B_T \subset V_J$  such that for  $\varepsilon_T = T^{-\frac{a \wedge s}{2a+d}} (\log T)$ :

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

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

$$\begin{aligned} \Pi_T(\|b - b_0\|_{\mu_0} \leq 4\sqrt{d}\|\mu_0\|_{\infty}^{1/2}\varepsilon_T) &\geq \Pi_T\left(\sup_{j=1,\dots,d} \|b_j - b_{0,j}\|_{L^2} \leq 4\varepsilon_T\right) \\ &\geq \prod_{j=1}^d \Pi_T(\|b_j - b_{0,j}\|_{\infty} \leq 4\varepsilon_T) \geq e^{-dT\varepsilon_T^2}, \end{aligned}$$

thereby verifying (23) for (a constant multiple of)  $\varepsilon_T$ .

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  $b_{0,J} = (b_{0,1,J}, \dots, b_{0,d,J})$ . 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

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

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

$$\begin{aligned} &\left\{2c_{\gamma}\|b - b_{0,J}\|_{\mu_0} \leq h_T(b, b_{0,J}) \leq \frac{1}{2}C_{\gamma}\|b - b_{0,J}\|_{\mu_0}, \right. \\ &\quad \left. \forall b \in V_J^{\otimes d} \text{ with } h_T(b, b_{0,J}) \geq (D_{\gamma} - r)\varepsilon_T\right\} \\ &\cap \{h_T(b_1, b_2) \leq C_{\gamma}\|b_1 - b_2\|_{\mu_0}, \forall b_1, b_2 \in V_J^{\otimes d} \text{ with } h_T(b_1, b_2) \geq D_{\gamma}\varepsilon_T\} \\ &\supset \left\{2c_{\gamma}\|b_1 - b_2\|_{\mu_0} \leq h_T(b_1, b_2) \leq \frac{1}{2}C_{\gamma}\|b_1 - b_2\|_{\mu_0}, \forall b_1, b_2 \in V_J^{\otimes d}\right\} \end{aligned}$$

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

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



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  $x = \sqrt{2}M_02^{Jd/2} \rightarrow \infty$  and  $M_0 > 1$  large enough yields

$$P_{b_0} \left( \sup_{b_1, b_2 \in V_J^{\otimes d}: b_1 \neq b_2} \left| \frac{h_T^2(b_1, b_2) - \|b_1 - b_2\|_{\mu_0}^2}{\|b_1 - b_2\|_{\mu_0}^2} \right| \geq \frac{CM_0}{\sqrt{T}} 2^{J[d+(d/2+\kappa-1)+]} \right) \leq de^{(c'_0 - M_0^2)2^{Jd}} \rightarrow 0,$$

where  $0 < \kappa < s - d/2 + 1$  (or  $\kappa = 0$  if  $d = 1$ ). Since  $T^{-1/2}2^{J[d+(d/2+\kappa-1)+]} \rightarrow 0$  as  $T \rightarrow \infty$  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  $T \rightarrow \infty$ . 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  $\varepsilon_T$  in the Hellinger distance  $h_T$ .

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

$$P_{b_0} \left( 2c_\gamma \|b_1 - b_2\|_{\mu_0} \leq h_T(b_1, b_2) \leq \frac{1}{2}C_\gamma \|b_1 - b_2\|_{\mu_0}, \forall b_1, b_2 \in V_J^{\otimes d} \right) \rightarrow 1.$$

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

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

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 “quantitative” 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  $T \rightarrow \infty$ , 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 ([56], page 142),

$$(26) \quad \sup_{A \text{ measurable}} |\Pi(A | X^T) - \Pi^{\mathcal{D}_T}(A | X^T)| \leq 2\Pi(\mathcal{D}_T^c | X^T) \xrightarrow{P_{b_0}}_{T \rightarrow \infty} 0.$$

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

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

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

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

where  $M_T \rightarrow \infty$  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  $\tilde{\Phi}_{\lambda,k,j} = (\tilde{\Phi}_{\lambda,k,j,1}, \dots, \tilde{\Phi}_{\lambda,k,j,d}) : \mathbb{T}^d \rightarrow \mathbb{R}^d$  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  $\Gamma_T$  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  $\Gamma_T$  as in (30), envelope  $\sigma_{\Gamma_T}$  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  $T \rightarrow \infty$ . Moreover, if  $b \sim \Pi^{D_T}(\cdot | X^T)$  then for all  $\lambda \leq J$ , as  $T \rightarrow \infty$ ,

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

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

(ii) Assume the conditions of Theorem 3 and let  $\bar{D}_T$  be the set from (29) with  $\Gamma_T$  as in (30) with  $a_\lambda = 1$  for all  $\lambda$  and envelope  $\sigma_{\Gamma_T}$  as in (31). Then  $\Pi(\bar{D}_T | X^T) = 1 - o_{P_{b_0}}(1)$  as  $T \rightarrow \infty$  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  $\Gamma_T, \sigma_{\Gamma_T}$  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,

$$\begin{aligned}
 & \Pi^{D_T} \left( b : \sum_{j=1}^d \|b_j - b_{0,j}\|_\infty \geq \tilde{\varepsilon}_T \mid X^T \right) \\
 & \leq \tilde{\varepsilon}_T^{-1} \sum_{j=1}^d E^{\Pi^{D_T}} [\|b_j - b_{0,j}\|_\infty \mid X^T] \\
 (34) \quad & \leq \tilde{\varepsilon}_T^{-1} \sum_{j=1}^d E^{\Pi^{D_T}} [\|b_j - P_{V_j}[b_{0,j}]\|_\infty \mid X^T] \\
 & \quad + \tilde{\varepsilon}_T^{-1} \sum_{j=1}^d \|P_{V_j}[b_{0,j}] - b_{0,j}\|_\infty.
 \end{aligned}$$

Since  $b_0 \in C^s$ , the second sum is of order  $O(2^{-J^s}) = O(T^{-\frac{s}{2a+d}}) = o(\tilde{\varepsilon}_T)$  by standard results in approximation theory (cf. after (4) above). Suppose first that  $d \leq 4$  and let  $a_\lambda = 2^{\lambda d/2} 2^{-J^{d/2}} (\log T)^{-\eta}$  for some  $\eta > 1$ . The standard inequality  $\sup_x \sum_k |\Phi_{\lambda,k}(x)| \lesssim 2^{\lambda d/2}$  (Section 4.3 in [23]) now gives

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

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

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

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  $T \rightarrow \infty$ , the convergence to zero of

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

3.4.3. *Proof of Theorems 3 and 4.* Let  $b \sim \Pi^{\bar{D}_T}(\cdot \mid X^T)$  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  $\Pi^{\bar{D}_T}(\cdot \mid X^T)$  in place of  $\Pi(\cdot \mid X^T)$ .

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  $\eta = (\eta_1, \dots, \eta_d) \in \mathcal{B}^\rho(1)$ . For projections

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

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

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

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

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

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

$$\begin{aligned} \beta(\bar{\Pi}^{\bar{D}_T}, \mathcal{N}_{b_0}) &\leq \beta(\bar{\Pi}_\kappa^{\bar{D}_T}, \mathcal{N}_{b_0, \kappa}) + \beta(\bar{\Pi}^{\bar{D}_T}, \bar{\Pi}_\kappa^{\bar{D}_T}) + \beta(\mathcal{N}_{b_0}, \mathcal{N}_{b_0, \kappa}) \\ &= \beta_{V_\kappa^{\otimes d}}(\bar{\Pi}_\kappa^{\bar{D}_T}, \mathcal{N}_{b_0, \kappa}) \\ &\quad + \sum_{i=1}^2 \sup_{\|F\|_{\text{Lip}} \leq 1} |E[F(Z_i) - F(P_{(\kappa)}(Z_i))]| \\ (36) \quad &\leq \beta_{V_\kappa^{\otimes d}}(\bar{\Pi}_\kappa^{\bar{D}_T}, \mathcal{N}_{b_0, \kappa}) + \sum_{i=1}^2 E\|Z_i - P_{(\kappa)}(Z_i)\|_{(B_{1\infty}^\rho)^*} \\ &= A + B + C. \end{aligned}$$

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

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

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

$$\begin{aligned} &E\|Z_1 - P_{(\kappa)}(Z_1)\|_{(B_{1\infty}^\rho)^*} \\ &= E \sup_{\|\eta\|_{B_{1\infty}^\rho} \leq 1} \sqrt{T} |\langle b - \hat{G}_J, \eta - P_{V_\kappa^{\otimes d}}[\eta] \rangle_{L^2}| \\ (38) \quad &\lesssim \sum_{\kappa < \lambda} 2^{-\lambda(\rho-d/2)} E \max_{k,j} \sqrt{T} |\langle b_j - \hat{G}_{J,j}, \Phi_{\lambda,k} \rangle_{L^2}| \\ &\lesssim \sum_{\kappa < \lambda} 2^{-\lambda(\rho-d/2)} E \max_{k,j} \sqrt{T} |\langle b_j - b_{0,j}, \Phi_{\lambda,k} \rangle_{L^2}| \\ &\quad + \sum_{\kappa < \lambda} 2^{-\lambda(\rho-d/2)} \max_{k,j} \sqrt{T} |\langle \hat{G}_{J,j} - b_{0,j}, \Phi_{\lambda,k} \rangle_{L^2}|. \end{aligned}$$

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  $T \rightarrow \infty$  and  $\kappa \rightarrow \infty$ , 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 (60) below, Bernstein’s inequality (15) implies that these variables are sub-Gaussian 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  $\lambda, k$  for our choice of  $J, s$ . Thus by the usual sub-Gaussian maximal inequality (Lemma 2.3.4 in [23]), 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  $\kappa, T \rightarrow \infty$ .

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 [23]), and arguing as in Theorem 3.7.28 in [23] 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  $\Phi_{\lambda,k,j} = (0, \dots, 0, \Phi_{\lambda,k}, 0, \dots, 0) : \mathbb{T}^d \rightarrow \mathbb{R}^d$ , where the nonzero coordinate occurs in the  $j$ th entry. Then, recalling the definition (35) of  $P_{(\kappa)}(Z_2)$  and using again (37) and the standard sub-Gaussian maximal inequality, now for the variables  $(Z_2(\Phi_{\lambda,k,j}) \sim N(0, \|\Phi_{\lambda,k}\|_{1/\mu_0}^2))$ ,

$$\begin{aligned} E \|Z_2 - P_{(\kappa)}(Z_2)\|_{(B_{1\infty}^\rho)^*} &= E \sup_{\|\eta\|_{B_{1\infty}^\rho} \leq 1} |Z_2(\eta - P_{V_\kappa^{\otimes d}}[\eta])| \\ (39) \qquad \qquad \qquad &\lesssim \sum_{\kappa < \lambda} 2^{-\lambda(\rho-d/2)} E \max_{k,j} |Z_2(\Phi_{\lambda,k,j})| \\ &\lesssim \sum_{\kappa < \lambda} 2^{-\lambda(\rho-d/2)} \sqrt{\lambda} =_{\kappa \rightarrow \infty} o(1). \end{aligned}$$

For term  $A$ , we show convergence of the finite-dimensional distributions by the semiparametric techniques from Section 4: consider the basis  $(\Phi_{\lambda,k,j} : k, j; \lambda \leq \kappa)$  of  $V_\kappa^{\otimes d}$  for  $\kappa$  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

$$\begin{aligned} E \Pi^{\bar{D}_T} \left[ e^{u\sqrt{T} \langle (b-b_0, \Phi_{\lambda,k,j}/\mu_0) \rangle_{\mu_0} - (u/\sqrt{T}) \int_0^T P_{V_j^{\otimes d}}[\Phi_{\lambda,k,j}/\mu_0](X_t) dW_t} \Big|_{X^T} \right] \\ = C_T \exp \left\{ \frac{u^2}{2} \int_{\mathbb{T}^d} \|P_{V_j^{\otimes d}}[\Phi_{\lambda,k,j}/\mu_0]\|^2 d\mu_0 \right\}, \end{aligned}$$

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  $T \rightarrow \infty$  for fixed  $u \in \mathbb{R}$ . We also have  $\|P_{V_j^{\otimes d}}[\Phi_{\lambda,k,j}/\mu_0]\|_{\mu_0} \rightarrow \|\Phi_{\lambda,k}/\mu_0\|_{\mu_0} = \|\Phi_{\lambda,k}\|_{1/\mu_0}$  as  $J \rightarrow \infty$  since  $P_{V_j}$  are  $L^2$ -projections. The same is true if  $\Phi_{\lambda,k,j}$  is replaced by arbitrary finite linear combinations

$\sum_j \sum_{\lambda \leq \kappa, k} a_{\lambda, k, j} \Phi_{\lambda, k, j}$ ,  $\kappa$  fixed, and thus by Proposition 29 in [35] we deduce joint weak convergence of the finite-dimensional distributions. In particular, for every fixed  $\kappa \in \mathbb{N}$ ,

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

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  $\kappa$ , 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 [23] and arguing as in the proof of Theorem 2.7 in [31]. Since  $\mathcal{N}_{b_0}$  has Bochner-mean zero, we deduce that

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

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  $\nu_T$  the law of the latter variables, then arguing just as in (36) we have

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

where  $\tilde{Z}_2 = \mathcal{L} Z_2$  from above,  $\tilde{Z}_1$  has law  $\nu_T$  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  $\kappa$ , by applying the martingale central limit theorem as in (46) to  $(1/\sqrt{T}) \int_0^T (\Phi_{\lambda, r, j}/\mu_0)(X_t).dW_t$ ,  $\lambda \leq \kappa$  fixed, and using (15) to show that the term

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

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

3.4.4. *Proofs of Theorems 5 and 6.* The proofs for Section 2.5 are given in Section 5 below and 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 (45) below is the key to these proofs and can be derived from perturbation arguments for the PDE (7) as follows: let  $\mu_b$  and  $\mu_{b+h}$  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

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

which is the same as

$$\begin{aligned} & \frac{\Delta}{2} (\mu_b - \mu_{b+h}) - b.\nabla (\mu_b - \mu_{b+h}) - \text{div}(b)(\mu_b - \mu_{b+h}) \\ & = -h.\nabla \mu_{b+h} - \text{div}(h)\mu_{b+h}. \end{aligned}$$

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

$$(43) \quad L_b^* u = -h \cdot \nabla \mu_{b+h} - \operatorname{div}(h) \mu_{b+h}.$$

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

$$(44) \quad L_b^* v_h = -h \cdot \nabla \mu_b - \operatorname{div}(h) \mu_b = - \sum_{j=1}^d \frac{\partial}{\partial x_j} (h_j \mu_b) \equiv f_h$$

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

$$\begin{aligned} \int_{\mathbb{T}^d} f_h(x) dx &= \sum_{j=1}^d \int_0^1 \cdots \int_0^1 \frac{\partial}{\partial x_j} (h_j(x) \mu_b(x)) dx \\ &= \sum_j \int_{\mathbb{T}^{d-1}} [(h_j \mu_b)(x_1, \dots, x_{j-1}, 1, x_{j+1}, \dots, x_d) \\ &\quad - (h_j \mu_b)(x_1, \dots, x_{j-1}, 0, x_{j+1}, \dots, x_d)] dx^{(j)} = 0, \end{aligned}$$

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 (43) and (44) to see that the differences  $w_{b,h} = \mu_b - \mu_{b+h} - v_h$  are the unique (periodic) integral-zero solutions of

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

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

$$(45) \quad \mu_b - \mu_{b+h} = v_{b,h} + w_{b,h} = (L_b^*)^{-1}[f_h] + (L_{b+h}^*)^{-1}[\bar{f}_h],$$

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  $\mathbb{T}^d$ ).

**Acknowledgments.** We thank James Norris for helpful discussions and the Associate Editor and referees for providing comments and several references.

The first author was supported by ERC Grant no. 647812 (UQMSI).

### SUPPLEMENTARY MATERIAL

**Supplement to “Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions”** (DOI: [10.1214/19-AOS1851SUPP](https://doi.org/10.1214/19-AOS1851SUPP); .pdf). The remaining proofs and technical results, along with a review of the relevant PDE material, are given in the supplement [32]. We linearly continue the numbering scheme for sections, equations, lemmas, etc., from the main document in the supplement, and items referred to which do not appear in the main article can be found in the supplement (e.g., Lemma 6).

### REFERENCES

[1] ABRAHAM, K. (2018). Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. *Bernoulli*. To appear. Available at [arXiv:1802.05635](https://arxiv.org/abs/1802.05635).  
 [2] AECKERLE-WILLEMS, C. and STRAUCH, C. (2018). Concentration of scalar ergodic diffusions and some statistical implications. Preprint. Available at [arXiv:1807.11331](https://arxiv.org/abs/1807.11331).  
 [3] AECKERLE-WILLEMS, C. and STRAUCH, C. (2018). Sup-norm adaptive simultaneous drift estimation for ergodic diffusions. Preprint. Available at [arXiv:1808.10660](https://arxiv.org/abs/1808.10660).

- [4] AGAPIOU, S., LARSSON, S. and STUART, A. M. (2013). Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Process. Appl.* **123** 3828–3860. MR3084161 <https://doi.org/10.1016/j.spa.2013.05.001>
- [5] BASS, R. F. (2011). *Stochastic Processes. Cambridge Series in Statistical and Probabilistic Mathematics* **33**. Cambridge Univ. Press, Cambridge. MR2856623 <https://doi.org/10.1017/CBO9780511997044>
- [6] BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHEAD, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 333–382. MR2278331 <https://doi.org/10.1111/j.1467-9868.2006.00552.x>
- [7] BLADT, M., FINCH, S. and SØRENSEN, M. (2016). Simulation of multivariate diffusion bridges. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 343–369. MR3454200 <https://doi.org/10.1111/rssb.12118>
- [8] CASTILLO, I. (2014). On Bayesian supremum norm contraction rates. *Ann. Statist.* **42** 2058–2091. MR3262477 <https://doi.org/10.1214/14-AOS1253>
- [9] CASTILLO, I. (2017). Pólya tree posterior distributions on densities. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 2074–2102. MR3729648 <https://doi.org/10.1214/16-AIHP784>
- [10] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856 <https://doi.org/10.1214/13-AOS1133>
- [11] CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. MR3262473 <https://doi.org/10.1214/14-AOS1246>
- [12] CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semi-parametric models. *Ann. Statist.* **43** 2353–2383. MR3405597 <https://doi.org/10.1214/15-AOS1336>
- [13] DALALYAN, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33** 2507–2528. MR2253093 <https://doi.org/10.1214/009053605000000615>
- [14] DALALYAN, A. and REISS, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Related Fields* **134** 248–282. MR2222384 <https://doi.org/10.1007/s00440-004-0416-1>
- [15] DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab. Theory Related Fields* **137** 25–47. MR2278451 <https://doi.org/10.1007/s00440-006-0502-7>
- [16] DASHTI, M., LAW, K. J. H., STUART, A. M. and VOSS, J. (2013). MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Probl.* **29** 095017, 27. MR3104933 <https://doi.org/10.1088/0266-5611/29/9/095017>
- [17] DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 <https://doi.org/10.1017/CBO9780511755347>
- [18] DUDLEY, R. M. (2014). *Uniform Central Limit Theorems*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **142**. Cambridge Univ. Press, New York. MR3445285
- [19] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007 <https://doi.org/10.1214/aos/1016218228>
- [20] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 <https://doi.org/10.1017/9781139029834>
- [21] GINÉ, E. and NICKL, R. (2008). Uniform central limit theorems for kernel density estimators. *Probab. Theory Related Fields* **141** 333–387. MR2391158 <https://doi.org/10.1007/s00440-007-0087-9>
- [22] GINÉ, E. and NICKL, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569–596. MR2475673 <https://doi.org/10.1007/s00440-008-0137-y>
- [23] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, New York. MR3588285 <https://doi.org/10.1017/CBO9781107337862>
- [24] GIORDANO, M. and KEKKONEN, H. (2018). Bernstein–von Mises theorems and uncertainty quantification for linear inverse problems. Preprint. Available at [arXiv:1811.04058](https://arxiv.org/abs/1811.04058).
- [25] GUGUSHVILI, S. and SPREIJ, P. (2014). Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. *Lith. Math. J.* **54** 127–141. MR3212631 <https://doi.org/10.1007/s10986-014-9232-1>
- [26] HORN, R. A. and JOHNSON, C. R. (2013). *Matrix Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2978290
- [27] KNAPIK, B. and SALOMOND, J.-B. (2018). A general approach to posterior contraction in nonparametric inverse problems. *Bernoulli* **24** 2091–2121. MR3757524 <https://doi.org/10.3150/16-BEJ921>
- [28] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881 <https://doi.org/10.1214/11-AOS920>



- [29] KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer Series in Statistics. Springer, London. MR2144185 <https://doi.org/10.1007/978-1-4471-3866-2>
- [30] LÖCHERBACH, E., LOUKIANOVA, D. and LOUKIANOV, O. (2011). Penalized nonparametric drift estimation for a continuously observed one-dimensional diffusion process. *ESAIM Probab. Stat.* **15** 197–216. MR2870512 <https://doi.org/10.1051/ps/2009016>
- [31] MONARD, F., NICKL, R. and PATERNAIN, G. P. (2019). Efficient nonparametric Bayesian inference for X-ray transforms. *Ann. Statist.* **47** 1113–1147. MR3909962 <https://doi.org/10.1214/18-AOS1708>
- [32] NICKL, R. and RAY, K. (2019). Supplement to “Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions.” <https://doi.org/10.1214/19-AOS1851SUPP>.
- [33] NICKL, R. and SÖHL, J. (2017). Bernstein–von Mises theorems for statistical inverse problems II: Compound Poisson processes. Preprint. Available at [arXiv:1709.07752](https://arxiv.org/abs/1709.07752).
- [34] NICKL, R. and SÖHL, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.* **45** 1664–1693. MR3670192 <https://doi.org/10.1214/16-AOS1504>
- [35] NICKL, R. and SÖHL, J. (2019). Bernstein–von Mises theorems for statistical inverse problems II: Compound Poisson processes. *Electron. J. Stat.* **13** 3513–3571. MR4013745 <https://doi.org/10.1214/19-ejs1609>
- [36] NICKL, R., VAN DE GEER, S. and WANG, S. (2018). Convergence rates for penalised least squares estimators in PDE-constrained regression problems. Preprint. Available at [arXiv:1809.08818](https://arxiv.org/abs/1809.08818).
- [37] PAPASPILIOPOULOS, O., POKERN, Y., ROBERTS, G. O. and STUART, A. M. (2012). Nonparametric estimation of diffusions: A differential equations approach. *Biometrika* **99** 511–531. MR2966767 <https://doi.org/10.1093/biomet/ass034>
- [38] PARDOUX, E. and VERETENNIKOV, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736 <https://doi.org/10.1214/aop/1015345596>
- [39] POKERN, Y., STUART, A. M. and VAN ZANTEN, J. H. (2013). Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.* **123** 603–628. MR3003365 <https://doi.org/10.1016/j.spa.2012.08.010>
- [40] RAY, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* **7** 2516–2549. MR3117105 <https://doi.org/10.1214/13-EJS851>
- [41] RAY, K. (2014). Asymptotic theory for Bayesian nonparametric procedures in inverse problems. Ph.D. thesis, University of Cambridge.
- [42] RAY, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **45** 2511–2536. MR3737900 <https://doi.org/10.1214/16-AOS1533>
- [43] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **293**. Springer, Berlin. MR1725357 <https://doi.org/10.1007/978-3-662-06400-9>
- [44] RIVOIRARD, V. and ROUSSEAU, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.* **40** 1489–1523. MR3015033 <https://doi.org/10.1214/12-AOS1004>
- [45] RUNST, T. and SICKEL, W. (1996). *Sobolev Spaces of Fractional Order, Nemytskij Operators, and Non-linear Partial Differential Equations*. *De Gruyter Series in Nonlinear Analysis and Applications* **3**. de Gruyter, Berlin. MR1419319 <https://doi.org/10.1515/9783110812411>
- [46] SCHAUER, M., VAN DER MEULEN, F. and VAN ZANTEN, H. (2017). Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli* **23** 2917–2950. MR3648050 <https://doi.org/10.3150/16-BEJ833>
- [47] SCHMEISSER, H.-J. and TRIEBEL, H. (1987). *Topics in Fourier Analysis and Function Spaces*. A Wiley-Interscience Publication. Wiley, Chichester. MR0891189
- [48] SCHMISSER, E. (2013). Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics* **47** 61–84. MR3023015 <https://doi.org/10.1080/02331888.2011.591931>
- [49] STRAUCH, C. (2015). Sharp adaptive drift estimation for ergodic diffusions: The multivariate case. *Stochastic Process. Appl.* **125** 2562–2602. MR3332848 <https://doi.org/10.1016/j.spa.2015.02.003>
- [50] STRAUCH, C. (2016). Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions. *Probab. Theory Related Fields* **164** 361–400. MR3449393 <https://doi.org/10.1007/s00440-014-0614-4>
- [51] STRAUCH, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist.* **46** 3451–3480. MR3852658 <https://doi.org/10.1214/17-AOS1664>
- [52] VAN WAAIJ, J. and VAN ZANTEN, H. (2016). Gaussian process methods for one-dimensional diffusions: Optimal rates and adaptation. *Electron. J. Stat.* **10** 628–645. MR3471991 <https://doi.org/10.1214/16-EJS1117>
- [53] VAN DER MEULEN, F. and SCHAUER, M. (2017). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.* **11** 2358–2396. MR3656495 <https://doi.org/10.1214/17-EJS1290>

- [54] VAN DER MEULEN, F. and VAN ZANTEN, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli* **19** 44–63. MR3019485 <https://doi.org/10.3150/11-BEJ385>
- [55] VAN DER MEULEN, F. H., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2006). Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli* **12** 863–888. MR2265666 <https://doi.org/10.3150/bj/1161614950>
- [56] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 <https://doi.org/10.1017/CBO9780511802256>
- [57] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** 1435–1463. MR2418663 <https://doi.org/10.1214/009053607000000613>
- [58] VERSHYNIN, R. (2010). Introduction to the non-asymptotic analysis of random matrices. Preprint. Available at [arXiv:1011.3027](https://arxiv.org/abs/1011.3027).