

SUPPLEMENTARY MATERIAL TO “NONPARAMETRIC BAYESIAN INFERENCE FOR REVERSIBLE MULTI-DIMENSIONAL DIFFUSIONS”

BY MATTEO GIORDANO¹ AND KOLYAN RAY²

¹Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, mg846@cam.ac.uk

²Department of Mathematics, Imperial College London, kolyan.ray@imperial.ac.uk

In this supplement, we provide additional background material and all the technical results and proofs not included in the main article. For the convenience of the reader, we repeat here the statements of results in the main article whose proofs appear in this supplement. We linearly continue the equation numbering scheme from the main document to the supplement.

CONTENTS

A	Additional material and technical results	1
A.1	Properties of Gaussian priors and proof of Lemma 2.3	1
A.2	Proof of Theorem 4.1	4
A.3	Proofs of convergence rates for posterior means	5
A.4	A PDE estimate	7
A.5	The Gaussian correlation inequality	7
B	Proofs of Lemmas 5.1, 5.2 and 5.4	8
B.1	Proof of Lemma 5.1	8
B.2	Proof of Lemma 5.2	11
B.3	Proof of Lemma 5.4	13
	References	14

APPENDIX A: ADDITIONAL MATERIAL AND TECHNICAL RESULTS

A.1. Properties of Gaussian priors and proof of Lemma 2.3.

A.1.1. *Periodic Matérn processes.* The Matérn process on $\mathbb{R}^d$ with smoothness parameter $s + 1 - d/2 > 0$ is a stationary Gaussian process with covariance kernel (Example 11.8 in [7])

$$(27) \quad K(x, y) \equiv K(x - y) = \int_{\mathbb{R}^d} e^{-i(x-y) \cdot \xi} (1 + \|\xi\|^2)^{-s-1} d\xi, \quad x, y \in \mathbb{R}^d.$$

Using a standard approach to periodization [17], one can construct a corresponding stationary periodic kernel via the Poisson summation formula:

$$K_{\text{per}}(x, y) \equiv K_{\text{per}}(x - y) = \sum_{m \in \mathbb{Z}^d} K(x - y + m), \quad x, y \in \mathbb{R}^d.$$

Since the Poisson summation formula preserves the Fourier transform (Theorem 8.31 of [5]), K_{per} has Fourier coefficients

$$\int_{\mathbb{T}^d} e^{-2\pi i k \cdot u} K_{\text{per}}(u) du = \int_{\mathbb{R}^d} e^{-2\pi i k \cdot u} K(u) du = \frac{(2\pi)^d}{(1 + 4\pi^2 \|k\|^2)^{s+1}}, \quad k \in \mathbb{Z}^d,$$

the last equality following from (27) and the Fourier inversion formula (e.g., Section 8.3 in [5]). Thus, using the Fourier series of $K_{\text{per}}(u)$, the periodized Matérn kernel has series

representation

$$K_{\text{per}}(x, y) = (2\pi)^d \sum_{k \in \mathbb{Z}^d} \frac{1}{(1 + 4\pi^2 \|k\|^2)^{s+1}} e_k(x) \overline{e_k(y)},$$

where $\{e_k, k \in \mathbb{Z}^d\}$ is the Fourier basis of $L^2(\mathbb{T}^d)$, i.e. $e_k(x) = e^{2\pi i k \cdot x}$. Theorem I.21 in [7] then implies that the centred Gaussian process $W = \{W(x) : x \in \mathbb{T}^d\}$ with covariance kernel K_{per} has RKHS

$$\mathbb{H} = \left\{ h = \sum_{k \in \mathbb{Z}^d} h_k e_k : \|h\|_{\mathbb{H}}^2 = (2\pi)^d \sum_{k \in \mathbb{Z}^d} h_k^2 (1 + 4\pi^2 \|k\|^2)^{s+1} < \infty \right\}.$$

By the Fourier series characterisation of Sobolev spaces, we thus have $\mathbb{H} = H^{s+1}(\mathbb{T}^d)$ and $\|\cdot\|_{\mathbb{H}}$ is an equivalent norm to $\|\cdot\|_{H^{s+1}}$. The above computation also implies that the periodic Matérn process has Karhunen-Loève series expansion given by (10), equivalent (up to constants) to the mean-zero Gaussian processes with inverse Laplacian covariance appearing in [14, 15, 24, 23].

For the regularity of the sample paths, by Proposition I.4 in [7], which applies also to the periodized kernel K_{per} , the periodic Matérn process has a version W whose sample paths are in $C^{s+1-\frac{d}{2}-\eta}(\mathbb{T}^d)$ for any $\eta > 0$. By Lemma I.7 in [7], this version then defines a Borel random element in $C^r(\mathbb{T}^d)$ for all $r < s + 1 - d/2$. In particular, $C^r(\mathbb{T}^d)$ is a separable linear subspace of $C^{(d/2+\kappa) \vee 2}(\mathbb{T}^d)$ for some $\kappa > 0$ provided that $s + 1 > d/2 + (d/2) \vee 2$ and $r > (d/2 + \kappa) \vee 2$. In conclusion, the periodic Matérn process satisfies Condition 2.1 for $s + 1 > d/2 + (d/2) \vee 2$.

A.1.2. Gaussian conjugacy formulae and proof of Lemma 2.3. We first provide details for the derivation of the conjugate formula (12). For basis functions $(h_k, k \in \mathbb{N}) \subset C^2 \cap \dot{L}^2$, and fixed $K \in \mathbb{N}$, identifying a function $B = \sum_{k=1}^K B_k h_k$ with its coefficient vector $\mathbf{B} = (B_1, \dots, B_K)^T \in \mathbb{R}^K$, we can write the log-likelihood (7) in quadratic form as

$$\begin{aligned} \ell_T(\mathbf{B}) &= -\frac{1}{2} \int_0^T \sum_{j=1}^d \sum_{k,k'} B_k B_{k'} \partial_{x_j} h_k(X_t) \partial_{x_j} h_{k'}(X_t) dt + \sum_{j=1}^d \int_0^T \sum_{k=1}^K B_k \partial_{x_j} h_k(X_t) dX_t^j \\ &= -\frac{1}{2} \sum_{k,k'} B_k B_{k'} \left[\int_0^T \nabla h_k(X_t) \cdot \nabla h_{k'}(X_t) dt \right] + \sum_{k=1}^K B_k \left[\int_0^T \nabla h_k(X_t) \cdot dX_t \right] \\ &= -\frac{1}{2} \mathbf{B}^T \Sigma \mathbf{B} + \mathbf{B}^T \mathbf{H}, \end{aligned}$$

for Σ and $\mathbf{H}$ defined as after (12). Under the same identification, we may regard the prior (11) as a multivariate normal distribution on $\mathbb{R}^K$ with diagonal covariance matrix $\Upsilon = \text{diag}(v_1^2, \dots, v_K^2)$. Then, by Bayes' formula, the posterior density is given by

$$d\Pi(\mathbf{B} | X^T) \propto e^{\ell_T(\mathbf{B})} e^{-\frac{1}{2} \mathbf{B}^T \Upsilon^{-1} \mathbf{B}} = e^{-\frac{1}{2} \mathbf{B}^T \Sigma \mathbf{B} + \mathbf{B}^T \mathbf{H} - \frac{1}{2} \mathbf{B}^T \Upsilon^{-1} \mathbf{B}}.$$

Completing the squares then gives

$$d\Pi(\mathbf{B} | X^T) \propto e^{-\frac{1}{2} [\mathbf{B} - (\Sigma + \Upsilon^{-1})^{-1} \mathbf{H}]^T (\Sigma + \Upsilon^{-1}) [\mathbf{B} - (\Sigma + \Upsilon^{-1})^{-1} \mathbf{H}]},$$

which completes the derivation of (12). We next provide the proof of Lemma 2.3.

LEMMA 2.3. *Let Π be a centred Gaussian Borel probability measure on $L^2(\mathbb{T}^d)$ that is supported on $C^2(\mathbb{T}^d) \cap \dot{L}^2(\mathbb{T}^d)$. Then the posterior distribution (8) is almost surely (under the law of the data X^T) Gaussian on $L^2(\mathbb{T}^d)$.*

PROOF. We follow ideas in [15]. For $K \in \mathbb{N}$, let $P_K : L^2 \rightarrow L^2$ be the projection onto the Fourier approximation space $E_K = \text{span}\{e_k, |k| \leq K\}$, $e_k(x) = e^{2\pi i k \cdot x}$. Consider the approximate posterior

$$\Pi_K(A|X^T) = \frac{\int_A e^{\ell_T(P_K B)} d\Pi(B)}{\int_{C^2} e^{\ell_T(P_K B')} d\Pi(B')}, \quad A \subset C^2 \cap \dot{L}^2 \text{ measurable.}$$

Since $\ell_T(P_K B)$ only depends on $P_K B$, $\Pi_K(\cdot|X^T)$ can be decomposed as the product of a Gaussian measure on E_K (satisfying the conjugate formula (12)) and a Gaussian measure obtained as the push-forward of Π under the projection operator onto the orthogonal complement of E_K in L^2 . It follows that $\Pi_K(\cdot|X^T)$ is Gaussian in L^2 . We now show that, a.s. under the law of the data X^T , $\Pi_K(\cdot|X^T) \rightarrow \Pi(\cdot|X^T)$ weakly in L^2 as $K \rightarrow \infty$, which in turn will conclude the proof since the class of Gaussian measures is closed with respect to weak convergence.

For $B \in C^2$, we have $\|B - P_K B\|_{H^2} \rightarrow 0$ as $K \rightarrow \infty$. It follows that $e^{\ell_T(P_K B)} \rightarrow e^{\ell_T(B)}$ a.s., since, given the data X^T , the function $B \mapsto \ell_T(B)$ in (7) is continuous with respect to the H^2 -norm. We next show that the limiting function $e^{\ell_T(B)}$, $B \in C^2$, is dominated by a Π -integrable function. Under P_{B_0} , using Itô's lemma (Theorem 39.3 of [2]),

$$\begin{aligned} \ell_T(B) &\leq \int_0^T \nabla B(X_t) \cdot dX_t \\ &= B(X_T) - B(X_0) - \frac{1}{2} \int_0^T \Delta B(X_t) dt \\ &\leq 2\|B\|_\infty + \frac{1}{2} \sum_{j=1}^d \int_0^T \left| \frac{\partial^2}{\partial x_j^2} B(X_t) \right| dt, \end{aligned}$$

and using Young's inequality with ε and $p = q = 2$, we obtain that for all $\eta > 0$,

$$\begin{aligned} \ell_T(B) &\leq \frac{2}{\eta} + \frac{\eta}{2} \|B\|_\infty^2 + \frac{1}{2} \sum_{j=1}^d \int_0^T \left[\frac{1}{2\eta} + \frac{\eta}{2} \left| \frac{\partial^2}{\partial x_j^2} B(X_t) \right|^2 \right] dt \\ &\leq \frac{2}{\eta} + \frac{\eta \|B\|_\infty^2}{2} + \frac{dT}{4\eta} + \frac{dT\eta}{4} \|B\|_{C^2}^2 = \frac{dT+8}{4\eta} + \frac{\eta(dT+2)}{4} \|B\|_{C^2}^2. \end{aligned}$$

In conclusion, for all $\eta > 0$,

$$e^{\ell_T(B)} \leq e^{(dT+8)/(4\eta)} e^{\eta(dT+2)\|B\|_{C^2}^2/4}.$$

Taking $\eta > 0$ small enough, since Π can be regarded as a Gaussian measure on C^2 , Fernique's theorem [4, Theorem 2.8.5] implies that the right hand side in the last display is Π -integrable. By the dominated convergence theorem, we then conclude that a.s. under the law of the data X^T , as $K \rightarrow \infty$,

$$\int_{C^2} e^{\ell_T(P_K B')} d\Pi(B') \rightarrow \int_{C^2} e^{\ell_T(B')} d\Pi(B'),$$

and likewise, for all measurable $A \subset C^2 \cap \dot{L}^2$, by boundedness of the indicator function $1_A(\cdot)$,

$$\int_A e^{\ell_T(P_K B)} d\Pi(B) = \int_{C^2} e^{\ell_T(P_K B)} 1_A(B) d\Pi(B) \rightarrow \int_A e^{\ell_T(B)} d\Pi(B),$$

showing as required that $\Pi_K(\cdot|X^T) \rightarrow \Pi(\cdot|X^T)$ weakly in C^2 (and also in L^2). $\square$

A.2. Proof of Theorem 4.1.

THEOREM 4.1. *Let d_T be a semimetric on the parameter space $\mathcal{H} \subseteq C^1(\mathbb{T}^d)$ for the drift b and let $\Pi = \Pi_T$ be priors for b . Let $\varepsilon_T \rightarrow 0$ satisfy $\sqrt{T}\varepsilon_T \rightarrow \infty$, let $\xi_T \rightarrow 0$, $\mathcal{H}_T \subseteq \mathcal{H}$ and let ϕ_T be a sequence of test functions satisfying*

$$P_{b_0}\phi_T \rightarrow 0, \quad \sup_{b \in \mathcal{H}_T: d_T(b, b_0) \geq D\xi_T} P_b(1 - \phi_T) \leq Le^{-(C+4)T\varepsilon_T^2}$$

for some $C, D, L > 0$ with $\Pi(\mathcal{H}_T^c) \leq e^{-(C+4)T\varepsilon_T^2}$. Suppose further that there exist deterministic sets $\mathcal{SB}_T \subseteq \mathcal{H}$ with $\Pi(\mathcal{SB}_T) \geq e^{-CT\varepsilon_T^2}$ and

$$P_{b_0} \left(\sup_{b \in \mathcal{SB}_T} \int_0^T \|b(X_s) - b_0(X_s)\|^2 ds \leq T\varepsilon_T^2 \right) \rightarrow 1,$$

where b_0 is the true drift function. Then for $M > 0$ large enough, as $T \rightarrow \infty$,

$$P_{b_0} \Pi(b : d_T(b, b_0) \geq M\xi_T | X^T) \rightarrow 0.$$

PROOF. The proof follows by the standard arguments for test-based contraction rates, see e.g. Theorem 2.1 of [6], with the only difference being how we control the denominator in the Bayes formula. We now detail this argument, which is a modification of Lemma 4.2 of van der Meulen et al. [21] adapted to our setting. By (7), the log-likelihood ratio equals (replacing ∇B with b), under P_{b_0} ,

$$\log \frac{dP_b}{dP_{b_0}}(X^T) = \int_0^T [b(X_s) - b_0(X_s)] \cdot dW_s - \frac{1}{2} \int_0^T \|b(X_s) - b_0(X_s)\|^2 ds =: M_T^b - \frac{1}{2} [M^b]_T.$$

Set $\bar{\Pi} = \Pi(\cdot \cap \mathcal{SB}_T) / \Pi(\mathcal{SB}_T)$ to be the normalized prior restricted to the set $\mathcal{SB}_T$. Since $P_{b_0}(\sup_{b \in \mathcal{SB}_T} [M^b]_T \leq T\varepsilon_T^2) \rightarrow 1$ by assumption, we henceforth work on this event. By Jensen's inequality,

$$\log \int \frac{dP_b}{dP_{b_0}}(X_T) \frac{d\Pi(b)}{\Pi(\mathcal{SB}_T)} \geq \int_{\mathcal{SB}_T} \log \frac{dP_b}{dP_{b_0}}(X_T) d\bar{\Pi}(b) = \int_{\mathcal{SB}_T} M_T^b - \frac{1}{2} [M^b]_T d\bar{\Pi}(b).$$

For fixed $T > 0$, define

$$Z_t^T = \int_{\mathcal{SB}_T} M_t^b d\bar{\Pi}(b) = \sum_{i=1}^d \int_0^t \int_{\mathcal{SB}_T} b_i(X_s) - b_{0,i}(X_s) d\bar{\Pi}(b) dW_s^i,$$

where the last equality follows from stochastic Fubini's theorem (Theorem 64 of Chapter IV in [16]) since $\mathcal{SB}_T$ is deterministic. The process $(Z_t^T : t \geq 0)$ is a continuous local martingale under P_{b_0} with quadratic variation

$$[Z^T]_t = \int_0^t \sum_{i=1}^d \left(\int_{\mathcal{SB}_T} b_i(X_s) - b_{0,i}(X_s) d\bar{\Pi}(b) \right)^2 ds.$$

Using Jensen's inequality, Fubini's theorem and the definition of $\mathcal{SB}_T$,

$$\begin{aligned} [Z^T]_T &\leq \int_0^T \sum_{i=1}^d \int_{\mathcal{SB}_T} (b_i(X_s) - b_{0,i}(X_s))^2 d\bar{\Pi}(b) ds \\ &= \int_{\mathcal{SB}_T} \int_0^T \|b(X_s) - b_0(X_s)\|^2 ds d\bar{\Pi}(b) = \int_{\mathcal{SB}_T} [M^b]_T d\bar{\Pi}(b) \leq T\varepsilon_T^2. \end{aligned}$$

By Bernstein's inequality (20), for any $x > 0$,

$$P_{b_0}(|Z_T^T| \geq x) = P_{b_0}(|Z_T^T| \geq x, [Z^T]_T \leq T\varepsilon_T^2) \leq 2\exp\left(-\frac{x^2}{2T\varepsilon_T^2}\right).$$

Setting $x = LT\varepsilon_T^2$ gives $P_{b_0}(|Z_T^T| \geq LT\varepsilon_T^2) \rightarrow 0$ for any $L > 0$. On the event $\{|Z_T^T| \leq LT\varepsilon_T^2\}$, which has P_{b_0} -probability tending to one, and using the second to last display,

$$\int_{\mathcal{SB}_T} M_T^b - \frac{1}{2}[M^b]_T d\bar{\Pi}(b) = Z_T^T - \frac{1}{2} \int_{\mathcal{SB}_T} [M^b]_T d\bar{\Pi}(b) \geq -(L + 1/2)T\varepsilon_T^2.$$

In conclusion, we have shown $P_{b_0}(\int dP_b/dP_{b_0} d\bar{\Pi}(b) \geq e^{-(L+1/2)T\varepsilon_T^2} \Pi(\mathcal{SB}_T)) \rightarrow 1$ for any $L > 0$. Substituting in $\Pi(\mathcal{SB}_T) \geq e^{-CT\varepsilon_T^2}$ gives $P_{b_0}(\int dP_b/dP_{b_0} d\bar{\Pi}(b) \geq e^{-(C+L+1/2)T\varepsilon_T^2}) \rightarrow 1$ for any $L > 0$. $\square$

A.3. Proofs of convergence rates for posterior means.

COROLLARY 2.2. *Let $\hat{B}_T = E^{\Pi_T}[B|X^T]$ be the posterior mean/MAP estimate. Under the conditions of Theorem 2.1, as $T \rightarrow \infty$,*

$$\|\nabla \hat{B}_T - \nabla B_0\|_2 = O_{P_{B_0}}(T^{-s/(2s+d)}).$$

PROOF. By Lemma 2.3, the posterior $\Pi(\cdot|X^T)$ of B is Gaussian, P_{B_0} -almost surely. Since the set $\{B \in \dot{L}^2 : \|\nabla B\|_2 \leq MT^{-s/(2s+d)}\}$ is closed, convex and symmetric, Theorem 2.1 and then Anderson's inequality imply that

$$\begin{aligned} 3/4 &< \Pi(B : \|\nabla B - \nabla B_0\|_2 \leq MT^{-s/(2s+d)} | X^T) \\ &\leq \Pi(B : \|\nabla B - \nabla \hat{B}_T\|_2 \leq MT^{-s/(2s+d)} | X^T) \end{aligned}$$

for $M, T > 0$ large enough with P_{B_0} -probability tending to one. In particular, the posterior probability of the intersection of the above two sets is at least $1/2$, and so there exists $B \in \dot{L}^2$ in both sets. For such a B ,

$$\|\nabla \hat{B}_T - \nabla B_0\|_2 \leq \|\nabla \hat{B}_T - \nabla B\|_2 + \|\nabla B - \nabla B_0\|_2 \leq 2MT^{-s/(2s+d)},$$

which thus holds with P_{B_0} -probability tending to one. $\square$

The proof for (the non-conjugate) p -exponential priors follows from uniform integrability arguments developed in [12].

COROLLARY 2.5. *Let $\hat{B}_T = E^{\Pi_T}[B|X^T]$ be the posterior mean. Under the conditions of Theorem 2.4, as $T \rightarrow \infty$,*

$$\|\nabla \hat{B}_T - \nabla B_0\|_p = O_{P_{B_0}}(T^{-s/(2s+d)}).$$

PROOF. Inspection of the proof of Theorem 4.1 shows that the assertion of Theorem 2.4 can be strengthened as follows: under the same assumptions, for any $A > 0$ we can find large enough $M > 0$ such that for $\varepsilon_T = T^{-s/(2s+d)}$,

$$(28) \quad P_{B_0}\left(\Pi(B : \|\nabla B - \nabla B_0\|_p > M\varepsilon_T | X^T) > e^{-AT\varepsilon_T^2}\right) \rightarrow 0$$

as $T \rightarrow \infty$. Using Jensen's inequality,

$$\|\nabla \hat{B}_T - \nabla B_0\|_p \lesssim \|E^{\Pi}[B|X^T] - B_0\|_{W^{1,p}} \leq E^{\Pi}[\|B - B_0\|_{W^{1,p}} | X^T].$$

For any $M > 0$, the last quantity is bounded by

$$\begin{aligned} & M\varepsilon_T + E^\Pi \left[\|B - B_0\|_{W^{1,p}} \mathbf{1}_{\{\|B - B_0\|_{W^{1,p}} > M\varepsilon_T\}} | X^T \right] \\ & \leq M\varepsilon_T + E^\Pi \left[\|B - B_0\|_{W^{1,p}}^p | X^T \right]^{1/p} \Pi(B : \|B - B_0\|_{W^{1,p}} > M\varepsilon_T | X^T)^{1/q}, \end{aligned}$$

where we have used Hölder's inequality with $1/p + 1/q = 1$. Under the identifiability assumption $B, B' \in \dot{L}^2(\mathbb{T}^d)$, Poincaré's inequality implies that the norm $\|\nabla B - \nabla B'\|_p$ is equivalent to the Sobolev norm $\|B - B'\|_{W^{1,p}}$. Using the above and (28), for any $A > 0$, we can find $M > 0$ large enough such that

$$\begin{aligned} & P_{B_0}(\|\nabla \hat{B}_T - \nabla B_0\|_p > 2M\varepsilon_T) \\ & \leq P_{B_0} \left(E^\Pi \left[\|B - B_0\|_{W^{1,p}}^p | X^T \right]^{1/p} \Pi(B : \|B - B_0\|_{W^{1,p}} > M\varepsilon_T | X^T)^{1/q} > M\varepsilon_T \right) \\ & \leq P_{B_0} \left(E^\Pi \left[\|B - B_0\|_{W^{1,p}}^p | X^T \right] e^{-pAT\varepsilon_T^2/q} > M^p \varepsilon_T^p \right) + o(1). \end{aligned}$$

For $C > 0$ the constant in the small ball estimate (25), define the event

$$A_T = \left\{ \int_{C^2 \cap \dot{L}^2} e^{\ell_T(B) - \ell_T(B_0)} d\Pi(B) \geq e^{-(C+1)T\varepsilon_T^2} \right\},$$

which is shown to satisfy $P_{B_0}(A_T) \rightarrow 1$ as $T \rightarrow \infty$ in the proof of Theorem 4.1. The second last display is then upper bounded by

$$\begin{aligned} & P_{B_0} \left(\left\{ \frac{\int_{C^2 \cap \dot{L}^2} \|B - B_0\|_{W^{1,p}}^p e^{\ell_T(B) - \ell_T(B_0)} d\Pi(B)}{\int_{C^2 \cap \dot{L}^2} e^{\ell_T(B) - \ell_T(B_0)} d\Pi(B)} > M^p \varepsilon_T^p e^{pAT\varepsilon_T^2/q} \right\} \cap A_T \right) + o(1) \\ & \leq P_{B_0} \left(\int_{C^2 \cap \dot{L}^2} \|B - B_0\|_{W^{1,p}}^p e^{\ell_T(B) - \ell_T(B_0)} d\Pi(B) > M^p \varepsilon_T^p e^{(pA/q - C - 1)T\varepsilon_T^2} \right) + o(1). \end{aligned}$$

By Markov's inequality and Fubini's theorem, the latter probability is smaller than

$$\begin{aligned} & \frac{e^{-(pA/q - C - 1)T\varepsilon_T^2}}{M^p \varepsilon_T^p} \int_{C^2 \cap \dot{L}^2} \|B - B_0\|_{W^{1,p}}^p E_{B_0} \left[e^{\ell_T(B) - \ell_T(B_0)} \right] d\Pi(B) \\ & = \frac{e^{-(pA/q - C - 1)T\varepsilon_T^2}}{M^p \varepsilon_T^p} E^\Pi \|B - B_0\|_{W^{1,p}}^p, \end{aligned}$$

since $E_{B_0} [e^{\ell_T(B) - \ell_T(B_0)}] = E_B[1] = 1$. Using the definition (14) of the p -exponential prior Π , the continuous embedding $B_{p1}^1 \subset W^{1,p}$ and that $|\sum_{i=1}^m a_i|^p \leq m^{p-1} \sum_{i=1}^m |a_i|^p$,

$$\begin{aligned} E^\Pi \|B\|_{W^{1,p}}^p & \lesssim E^\Pi \|B\|_{B_{p1}^1}^p = T^{-\frac{d}{2s+d}} E \left[\sum_{l=0}^J 2^{l(1+\frac{d}{2}-\frac{d}{p})} \left(\sum_r 2^{-pl(s+1+\frac{d}{2}-\frac{d}{p})} |\rho_{lr}|^p \right)^{\frac{1}{p}} \right]^p \\ & \leq T^{-\frac{d}{2s+d}} (J+1)^{p-1} \sum_{l=0}^J 2^{pl(1+\frac{d}{2}-\frac{d}{p})} \sum_r 2^{-pl(s+1+\frac{d}{2}-\frac{d}{p})} E |\rho_{lr}|^p \\ & \lesssim T^{-\frac{d}{2s+d}} J^{p-1} \sum_{l=0}^J 2^{-l(ps-d)} \lesssim T^{-\frac{d}{2s+d}} J^{p-1} \lesssim 1, \end{aligned}$$

since $2^J \simeq T^{1/(2s+d)}$ and $s > d/p + (d/2) \vee 2 > d/p$ by assumption. Combining the last three displays then yields that for any $A > 0$, we can find $M > 0$ large enough such that

$$P_{B_0}(\|\nabla \hat{B}_T - \nabla B_0\|_p > 2M\varepsilon_T) \lesssim \frac{e^{-(pA/q - C - 1)T\varepsilon_T^2}}{M^p \varepsilon_T^p} + o(1)$$

since $B_0 \in W^{1,p}$. Taking $A > (C + 1)q/p$ and using that $\varepsilon_T = T^{-s/(2s+d)}$, the right-hand side tends to zero as $T \rightarrow \infty$. $\square$

A.4. A PDE estimate. In the proofs we used a regularity estimate for solutions to the Poisson equation $L_b u = f$, where the generator L_b of the diffusion is the strongly elliptic second order partial differential operator given in (5). For some basic facts about this elliptic PDE, see Section 6 in [13], while for more general theory for periodic elliptic PDEs, see Chapter II.3 in [3]. The following estimate is only proved for $t \geq 2$ in Lemma 11 of [13], but the proof can be extended to general $t \in \mathbb{R}$.

LEMMA A.1. *Let $t \in \mathbb{R}$ and assume $b \in C^{|t-2|}(\mathbb{T}^d)$. For any $f \in \dot{L}_{\mu_b}^2(\mathbb{T}^d)$, there exists a unique solution $L_b^{-1}[f] \in L^2(\mathbb{T}^d)$ of the equation*

$$L_b u = f, \quad f \in L^2(\mathbb{T}^d),$$

satisfying $L_b L_b^{-1}[f] = f$ almost everywhere and $\int_{\mathbb{T}^d} L_b^{-1}[f](x) dx = 0$. Moreover,

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

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

PROOF. Recall the multiplication inequality for Besov-Sobolev norms with $s \geq 0$ ([20], p. 143),

$$\|fg\|_{B_{pq}^s} \leq c(s, p, q, d) \|f\|_{B_{pq}^s} \|g\|_{B_{\infty\infty}^s} \leq c'(s, p, q, d) \|f\|_{B_{pq}^s} \|g\|_{C^s},$$

which by duality implies that for $s \geq 0$,

$$\|fg\|_{B_{pq}^{-s}} = \sup_{\|\varphi\|_{B_{p'q'}^s} \leq 1} \left| \int_{\mathbb{T}^d} f \frac{g\varphi}{\|g\varphi\|_{B_{p'q'}^s}} \right| \|g\varphi\|_{B_{p'q'}^s} \leq c(s, p, q, d) \|f\|_{B_{pq}^{-s}} \|g\|_{B_{\infty\infty}^s},$$

where $1/p + 1/p' = 1/q + 1/q' = 1$. Combining the last displays yields the multiplier inequality

$$(29) \quad \|fg\|_{B_{pq}^s} \leq c(|s|, p, q, d) \|f\|_{B_{pq}^s} \|g\|_{B_{\infty\infty}^{|s|}}, \quad s \in \mathbb{R}.$$

The proof of the result then follows as in Lemma 11 of [13] upon using the multiplier inequality (29) instead of (3) when establishing (69) in that paper. We omit the details here. $\square$

A.5. The Gaussian correlation inequality. The Gaussian correlation inequality states that for any closed, symmetric, convex sets K, L in $\mathbb{R}^d$ and any centred Gaussian measure μ in $\mathbb{R}^d$, we have

$$\mu(K \cap L) \geq \mu(K)\mu(L).$$

This was proved by Royen [18]; see also [10] for a self-contained presentation and proof. In the proof of Lemma 5.2 we have used the following extension to Gaussian measures on separable Banach spaces, which was referred to, without full details, in Remark 3 (i) in [10]. We include a proof for completeness.

LEMMA A.2. *Let μ be a centred Borel Gaussian measure on a separable Banach space $\mathbb{B}$. Let K, L be closed (with respect to the topology induced by $\|\cdot\|_{\mathbb{B}}$), convex, symmetric subsets of $\mathbb{B}$. Then,*

$$\mu(K \cap L) \geq \mu(K)\mu(L).$$

PROOF. For $X \sim \mu$, the Karhunen-Loève expansion ([8], Theorem 2.6.10) of X implies that there exists a complete orthonormal system $\{x_n, n \geq 1\}$ of the RKHS of X and i.i.d. standard normal random variables $(\xi_n)_{n \geq 1}$, such that $X_n = \sum_{i=1}^n \xi_i x_i$ converges almost surely to X in the norm of $\mathbb{B}$. We first show that for any closed, convex, symmetric set $K \subseteq \mathbb{B}$,

$$(30) \quad \mu(K) = \Pr \left(\sum_{i=1}^{\infty} \xi_i x_i \in K \right) = \lim_{n \rightarrow \infty} \Pr \left(\sum_{i=1}^n \xi_i x_i \in K \right).$$

Since K is closed and $X_n \rightarrow X$ almost surely, and hence also in law, the Portmanteau lemma ([22], Lemma 2.2) implies

$$\limsup_{n \rightarrow \infty} \Pr(X_n \in K) \leq \Pr(X \in K).$$

On the other hand, since K is convex and symmetric, Anderson's inequality ([8], p. 49) yields $\Pr(X \in K) \leq \Pr(X_n \in K)$ for all $n \geq 1$. Together, these establish (30).

Now let $K, L \subseteq \mathbb{B}$ be arbitrary closed, convex, symmetric sets and define

$$K_n = \left\{ z = (z_1, \dots, z_n) \in \mathbb{R}^n : \sum_{i=1}^n z_i x_i \in K \right\} \subseteq \mathbb{R}^n$$

and analogously L_n . It is straightforward to check that K_n is a convex and symmetric subset of $\mathbb{R}^n$. Furthermore, if $\{z^{(k)}\}_{k \geq 1} \subset K_n$ converges to some $z = (z_1, \dots, z_n) \in \mathbb{R}^n$, then

$$\left\| \sum_{i=1}^n z_i^{(k)} x_i - \sum_{i=1}^n z_i x_i \right\|_{\mathbb{B}} \leq \max_{i=1, \dots, n} \|x_i\|_{\mathbb{B}} \sum_{i=1}^n |z_i^{(k)} - z_i| \rightarrow 0, \quad k \rightarrow \infty,$$

so that, K being closed in $\mathbb{B}$, $\sum_{i=1}^n z_i x_i \in K$ and $z \in K_n$. This shows that K_n is closed, convex and symmetric in $\mathbb{R}^n$. Denoting by $\gamma_n = \mathcal{L}(\xi_1, \dots, \xi_n)$ the standard Gaussian measure on $\mathbb{R}^n$, (30) then implies $\mu(K) = \lim_{n \rightarrow \infty} \gamma_n(K_n)$. Note that identical considerations apply to L and L_n and, since $K \cap L$ is also closed, convex and symmetric in $\mathbb{B}$, also to $K \cap L$ and $K_n \cap L_n$.

Applying this and the Gaussian correlation inequality for the finite-dimensional Gaussian measure γ_n ,

$$\mu(K \cap L) = \lim_{n \rightarrow \infty} \gamma_n(K_n \cap L_n) \geq \lim_{n \rightarrow \infty} \gamma_n(K_n) \gamma_n(L_n) = \mu(K) \mu(L).$$

□

APPENDIX B: PROOFS OF LEMMAS 5.1, 5.2 AND 5.4

B.1. Proof of Lemma 5.1.

LEMMA 5.1. Suppose $B_0 \in H^{s+1}(\mathbb{T}^d) \cap \dot{L}^2(\mathbb{T}^d)$ for $s > d \vee (3/2)$. If $\varepsilon_T = T^{-s/(2s+d)}$, $J \in \mathbb{N}$ satisfies $2^J \simeq T^{1/(2s+d)}$ and $M > 0$, then as $T \rightarrow \infty$,

$$P_{B_0} \left(\sup_{B \in V_J \cap \dot{L}^2 : \|\nabla B - \nabla B_0\|_{L^2(\mu_0)} \leq M \varepsilon_T} \frac{1}{T} \int_0^T \|\nabla B(X_s) - \nabla B_0(X_s)\|^2 ds \leq M^2 \varepsilon_T^2 + o(\varepsilon_T^2) \right) \rightarrow 1.$$

PROOF. We first show that $\frac{1}{T} \int_0^T \|\nabla B(X_s) - \nabla B_0(X_s)\|^2 ds$ is close to its ergodic average $\|\nabla B - \nabla B_0\|_{L^2(\mu_0)}^2$, uniformly over $\mathcal{SB}_T = \{B \in V_J \cap \dot{L}^2 : \|\nabla B - \nabla B_0\|_{L^2(\mu_0)} \leq M\varepsilon_T\}$. Define

$$\mathcal{F}_T = \{f_B(x) := \|\nabla B(x) - \nabla B_0(x)\|^2 - \|\nabla B - \nabla B_0\|_{L^2(\mu_0)}^2 : B \in \mathcal{SB}_T \cup \{B_0\}\}.$$

Since $x \mapsto \|x\|^2$ is a smooth function, $\mathcal{SB}_T \subset V_J$ and $B_0 \in H^{s+1}$, it holds that $\mathcal{F}_T \subset \dot{L}_{\mu_0}^2 \cap H^s$. By Lemma A.1, the Poisson equation $L_{B_0} u = f_B$ (writing L_{B_0} for $L_{\nabla B_0}$) has a unique solution $L_{B_0}^{-1}[f_B] \in \dot{L}^2$ for any $B \in \mathcal{SB}_T$. Applying Lemma 1 of [13], since $s > d \vee (3/2) > (d/2 + \kappa) \vee 1$ for $\kappa > 0$ small enough, gives that for any $x \geq 0$,

$$(31) \quad P_{B_0} \left(\sup_{f_B \in \mathcal{F}_T} |\mathbb{G}_T[f_B]| \geq \sup_{f \in \mathcal{F}_T} \frac{2\|L_{B_0}^{-1}[f_B]\|_\infty}{\sqrt{T}} + J_{\mathcal{F}_T}(4\sqrt{2} + 192x) \right) \leq e^{-x^2/2},$$

where $J_{\mathcal{F}_T} = \int_0^{D_{\mathcal{F}_T}} \sqrt{\log 2N(\mathcal{F}_T, 6d_L, \tau)} d\tau$, $d_L^2(f, g) = \sum_{i=1}^d \|\partial_{x_i} L_{B_0}^{-1}[f - g]\|_\infty^2$ and $D_{\mathcal{F}_T}$ is the d_L -diameter of $\mathcal{F}_T$. We now proceed to bound $J_{\mathcal{F}_T}$.

For $B \in \mathcal{SB}_T$, write $h_i = \partial_{x_i}(B - B_0)$ so that $f_B = \sum_{i=1}^d h_i^2 - \|h_i\|_{L^2(\mu_0)}^2$. Using the Sobolev embedding theorem, Lemma A.1 and the Runst-Sickel lemma ([13], Lemma 2), for any $\kappa > 0$,

$$\begin{aligned} d_L(f_B, f_{\bar{B}}) &\lesssim \|f_B - f_{\bar{B}}\|_{H^{d/2+\kappa-1}} \\ &\lesssim \sum_{i=1}^d \left\| h_i^2 - \bar{h}_i^2 - \int_{\mathbb{T}^d} (h_i^2 - \bar{h}_i^2) d\mu_0 \right\|_{H^{d/2+\kappa-1}} \\ &\lesssim \sum_{i=1}^d \|h_i - \bar{h}_i\|_\infty \|h_i + \bar{h}_i\|_{H^{(d/2+\kappa-1)_+}} + \|h_i - \bar{h}_i\|_{H^{(d/2+\kappa-1)_+}} \|h_i + \bar{h}_i\|_\infty \\ &\quad + \|\mu_0\|_\infty \|h_i - \bar{h}_i\|_2 \|h_i + \bar{h}_i\|_2 \|1\|_{H^{d/2+\kappa-1}}, \end{aligned}$$

where the constants depend only on d, κ and $\|\nabla B_0\|_{B_{\infty\infty}^{[d/2+\kappa-1]}} \lesssim \|B_0\|_{H^{[d/2+\kappa-1]+d/2+1}} \lesssim \|B_0\|_{H^{s+1}}$, and we note that $\|\mu_0\|_\infty \leq e^{4\|B_0\|_\infty} < \infty$ using (6). Since $B, \bar{B} \in V_J$, we have $\|h_i - \bar{h}_i\|_{H^u} \lesssim \|B - \bar{B}\|_{H^{u+1}} \leq 2^{Ju} \|B - \bar{B}\|_{H^1}$ for $u \geq 0$ and $\|h_i - \bar{h}_i\|_\infty \lesssim \|B - \bar{B}\|_{B_{1\infty}^1} \lesssim 2^{Jd/2} \|B - \bar{B}\|_{H^1}$. Furthermore,

$$\begin{aligned} \|h_i + \bar{h}_i\|_{H^u} &\lesssim \sup_{B \in \mathcal{SB}_T} \|B - B_0\|_{H^{u+1}} \lesssim \sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{H^{u+1}} + \|B_0 - P_J B_0\|_{H^{u+1}} \\ &\lesssim 2^{Ju} \sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{H^1} + 2^{-J(s-u)} \|B_0\|_{H^{s+1}}. \end{aligned}$$

Note that $\|\nabla B - \nabla B_0\|_{L^2(\mu_0)} \simeq \|B - B_0\|_{H^1}$ since μ_0 is both bounded and bounded away from zero and $B, B_0 \in \dot{L}^2$. Therefore, $\sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{H^1} + 2^{-Js} \|B_0\|_{H^{s+1}} \leq C(M+1)\varepsilon_T$, so that $\|h_i + \bar{h}_i\|_{H^u} \leq C(M+1)2^{Ju}\varepsilon_T$. Similarly, using the Sobolev embedding theorem,

$$\begin{aligned} \|h_i + \bar{h}_i\|_\infty &\lesssim \sup_{B \in \mathcal{SB}_T} \|B - B_0\|_{B_{1\infty}^1} \lesssim \sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{B_{1\infty}^1} + \|B_0 - P_J B_0\|_{B_{1\infty}^1} \\ &\lesssim 2^{Jd/2} \sup_{B \in \mathcal{SB}_T} \|B - P_J B_0\|_{H^1} + 2^{-J(s-d/2)} \|B_0\|_{H^{s+1}} \\ &\leq C(M+1)2^{Jd/2}\varepsilon_T. \end{aligned}$$

Combining these bounds yields

$$d_L(f_B, f_{\bar{B}}) \leq C(M+1)2^{J[d/2+(d/2+\kappa-1)_+]\varepsilon_T} \|B - \bar{B}\|_{H^1}.$$

Since $\mathcal{SB}_T \subseteq (V_J, \|\cdot\|_{H^1})$ is finite dimensional, using the last display and the usual covering argument for balls in finite dimensional spaces (e.g. [8], Proposition 4.3.34),

$$\begin{aligned} \log N(\mathcal{F}_T, 6d_L, \tau) &\leq \log N(\mathcal{SB}_T, C(M+1)2^{J[d/2+(d/2+\kappa-1)_+]\varepsilon_T} \|\cdot\|_{H^1}, \tau) \\ &\lesssim \dim(V_J) \log(C(M+1)2^{J[d/2+(d/2+\kappa-1)_+]\varepsilon_T} \sup_{B, \bar{B} \in \mathcal{SB}_T} \|B - \bar{B}\|_{H^1} / \tau) \\ &\lesssim \dim(V_J) \log(CR_T / \tau), \end{aligned}$$

where $C > 0$ and

$$R_T := (M+1)^2 2^{J[d/2+(d/2+\kappa-1)_+]\varepsilon_T^2} \rightarrow 0$$

under the current assumption on s for $\kappa > 0$ small enough. Recall the inequality $\int_0^a \sqrt{\log(A/x)} dx \leq 4a\sqrt{\log(A/a)}$ for any $A \geq 2$ and $0 < a \leq 1$ ([8], p. 190). Using this inequality, the last display with $\dim(V_J) = O(2^{Jd})$ and that $\mathcal{F}_T$ has d_L -diameter $D_{\mathcal{F}_T} \lesssim R_T \rightarrow 0$,

$$J_{\mathcal{F}_T} \lesssim \dim(V_J) \int_0^{D_{\mathcal{F}_T}} \sqrt{\log([CR_T] \vee 2/\tau)} d\tau \lesssim 2^{Jd/2} D_{\mathcal{F}_T} \sqrt{\log([CR_T] \vee 2/D_{\mathcal{F}_T})}.$$

Taking $D_{\mathcal{F}_T} \simeq R_T$ in the last display gives $J_{\mathcal{F}_T} \lesssim 2^{Jd/2} R_T \left(1 + \sqrt{\log(1/R_T)}\right)$. Arguing as for the bound for $d_L(f_B, f_{\bar{B}})$ above, one has for all $B \in \mathcal{SB}_T$,

$$\begin{aligned} \|L_{B_0}^{-1}[f_B]\|_\infty &\lesssim \|f_B\|_{H^{d/2+\kappa-1}} \lesssim \sum_{i=1}^d \left\| h_i^2 - \int_{\mathbb{T}^d} h_i^2 d\mu_0 \right\|_{H^{d/2+\kappa-1}} \\ &\lesssim \sum_{i=1}^d \|h_i\|_\infty \|h_i\|_{H^{(d/2+\kappa-1)_+}} \lesssim 2^{J[d/2+(d/2+\kappa-1)_+]} (M+1)^2 \varepsilon_T^2 \end{aligned}$$

where the constants depend only on $d, \kappa, \|\mu_0\|_\infty$ and $\|\nabla B_0\|_{B_\infty^{[d/2+\kappa-1]}} \lesssim \|B_0\|_{H^{[d/2+\kappa-1]+d/2+1}} \lesssim \|B_0\|_{H^{s+1}}$. Substituting this bound and $J_{\mathcal{F}_T} \lesssim 2^{Jd/2} R_T \left(1 + \sqrt{\log(1/R_T)}\right)$ into (31),

$$P_{B_0} \left(\sup_{f_B \in \mathcal{F}_T} |\mathbb{G}_T[f_B]| \geq C 2^{Jd/2} R_T \left(1 + \sqrt{\log(1/R_T)}\right) (1+x) \right) \leq e^{-x^2/2}$$

for any $x \geq 0$. Set

$$\zeta_T = M_T T^{-1/2} 2^{Jd/2} R_T (1 + \sqrt{\log(1/R_T)}) = O(M_T \sqrt{\log T} T^{-\frac{s-d/2-(d/2+\kappa-1)_+}{2s+d}} \varepsilon_T^2),$$

which satisfies $\zeta_T = o(\varepsilon_T^2)$ for $M_T \rightarrow \infty$ growing slow enough, since $s > d \vee (3/2) > d/2 + (d/2 + \kappa - 1)_+$ for $\kappa > 0$ small enough. Then using the definition of the empirical process $\mathbb{G}_T[f_B]$, for any $M_T \rightarrow \infty$,

$$P_{B_0} \left(\sup_{B \in \mathcal{SB}_T} \left| \frac{1}{T} \int_0^T \|\nabla B(X_s) - \nabla B_0(X_s)\|^2 ds - \|\nabla B - \nabla B_0\|_{L^2(\mu_0)}^2 \right| \geq \zeta_T \right) \rightarrow 0$$

as $T \rightarrow \infty$. The result then follows because on the complement of the event in the last display,

$$\begin{aligned} \sup_{B \in \mathcal{SB}_T} \frac{1}{T} \int_0^T \|\nabla B(X_s) - \nabla B_0(X_s)\|^2 ds &\leq \sup_{B \in \mathcal{SB}_T} \|\nabla B - \nabla B_0\|_{L^2(\mu_0)}^2 + \zeta_T \\ &\leq M^2 \varepsilon_T^2 + o(\varepsilon_T^2). \end{aligned}$$

□

B.2. Proof of Lemma 5.2.

LEMMA 5.2. For $s, M, \kappa > 0$, $p \in [1, 2]$ and sequences $\varepsilon_T = T^{-s/(2s+d)}$, $\bar{\varepsilon}_T = T^{-(s+1)/(2s+d)}$, define the sets

$$\mathcal{B}_T = \{B = B_1 + B_2 : \|B_1\|_\infty \leq \bar{\varepsilon}_T, \|B_1\|_{C^1} \leq \varepsilon_T, \|B_2\|_{B_{pp}^{s+1}} \leq M, \|B\|_{C^{(d/2+\kappa) \vee 2}} \leq M\}.$$

Assume either:

- (i) $p = 2$ and $B = W/(\sqrt{T}\varepsilon_T)$ for $W \sim \Pi_W$ a Gaussian process satisfying Condition 2.1;
- (ii) $B = W/(T\varepsilon_T^2)^{\frac{1}{p}}$ for $W \sim \Pi_W$ a p -exponential random element as in (13) with $s > (d/2 + \kappa) \vee 2 + d/p - 1$.

Let $\Pi = \Pi_T$ be the law of B . Then, for every $K > 0$, there exists $M > 0$ large enough such that $\Pi(\mathcal{B}_T^c) \leq e^{-KT\varepsilon_T^2}$.

PROOF. (i) In the Gaussian case, define the sets

$$\mathcal{B}_{T,1} := \{B = B_1 + B_2 : \|B_1\|_\infty \leq \bar{\varepsilon}_T, \|B_1\|_{C^1} \leq \varepsilon_T, \|B_2\|_{B_{22}^{s+1}} \leq M\},$$

$$\mathcal{B}_{T,2} := \{B : \|B\|_{C^{(d/2+\kappa) \vee 2}} \leq M\}.$$

To upper bound $\Pi(\mathcal{B}_T^c)$ it thus suffices to upper bound $\Pi(\mathcal{B}_{T,1}^c)$ and $\Pi(\mathcal{B}_{T,2}^c)$. Since $\|g\|_{B_{22}^{s+1}} = \|g\|_{H^{s+1}} \leq c_0 \|g\|_{\mathbb{H}}$ for all $g \in \mathbb{H}$ under Condition 2.1, Borell's isoperimetric inequality ([8], Theorem 2.6.12) gives

$$\begin{aligned} \Pi(\mathcal{B}_{T,1}) &= \Pi_W(W = W_1 + W_2 : \|W_1\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T, \|W_1\|_{C^1} \leq \sqrt{T}\varepsilon_T^2, \|W_2\|_{B_{22}^{s+1}} \leq M\sqrt{T}\varepsilon_T) \\ (32) \quad &\geq \Phi(\Phi^{-1}(\Pi_W(W : \|W\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T, \|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2)) + M\sqrt{T}\varepsilon_T/c_0), \end{aligned}$$

where Φ is the standard normal cumulative distribution function. Now for $\mathbb{H}_1$ and H_1^{s+1} the unit balls of $\mathbb{H}$ and H^{s+1} respectively, we have under Condition 2.1 that

$$\log N(\mathbb{H}_1, \|\cdot\|_{C^1}, \tau) \leq \log N(c_0 H_1^{s+1}, \|\cdot\|_{C^1}, \tau) \lesssim \tau^{-d/s},$$

where the last inequality follows by arguing as in Theorem 4.3.36 of [8]. By Theorem 1.2 of Li and Linde [11], this yields

$$(33) \quad \Pi_W(\|W\|_{C^1} \leq \eta) \geq e^{-c_1^2 \eta^{-2d/(2s-d)}} \quad \text{as } \eta \rightarrow 0,$$

for any $d/2 < s < \infty$ and some $c_1 = c_1(d, s, c_0) > 0$, which implies, since $\sqrt{T}\varepsilon_T^2 \rightarrow 0$,

$$\Pi_W(\|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2) \geq e^{-c_1^2 (\sqrt{T}\varepsilon_T^2)^{-2d/(2s-d)}} = e^{-c_1^2 T\varepsilon_T^2}.$$

Using the same argument, now with the bound

$$\log N(\mathbb{H}_1, \|\cdot\|_\infty, \tau) \leq \log N(c_0 H_1^{s+1}, \|\cdot\|_\infty, \tau) \lesssim \tau^{-d/(s+1)},$$

it follows for some $c_2 > 0$ that $\Pi_W(\|W\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T) \geq e^{-c_2^2 T\varepsilon_T^2}$. The Gaussian correlation inequality (which holds for Gaussian measures in separable Banach spaces, see Lemma A.2 below) then gives for $c_3 = c_1^2 + c_2^2$

$$\begin{aligned} \Pi_W(\|W\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T, \|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2) &\geq \Pi_W(\|W\|_\infty \leq \sqrt{T}\varepsilon_T\bar{\varepsilon}_T) \Pi_W(\|W\|_{C^1} \leq \sqrt{T}\varepsilon_T^2) \\ &\geq e^{-c_3 T\varepsilon_T^2}. \end{aligned}$$

Using the standard inequality $\Phi^{-1}(y) \geq -\sqrt{2\log(1/y)}$ for $0 < y < 1$, the right hand side of (32) is thus lower bounded by

$$\Phi((M/c_0 - \sqrt{2c_3})\sqrt{T}\varepsilon_T).$$

Defining $m_T = -\Phi^{-1}(e^{-KT\varepsilon_T^2}/2)$, this further gives $m_T \leq \sqrt{2\log 2} + \sqrt{2KT}\varepsilon_T$, which can be made smaller than $(M/c_0 - \sqrt{2c_3})\sqrt{T}\varepsilon_T$ by taking $M = M(K, c_0, c_3)$ large enough. For such M , the last display is lower bounded by $\Phi(m_T) = 1 - \Phi(\Phi^{-1}(e^{-KT\varepsilon_T^2}/2)) = 1 - \frac{1}{2}e^{-KT\varepsilon_T^2}$.

To bound $\Pi(\mathcal{B}_{T,2}^c)$, recall that by Condition 2.1, W defines a Gaussian Borel random element in a separable linear subspace $\mathcal{S}$ of $C^{(d/2+\kappa)\vee 2}$. Using the Hahn-Banach theorem, we may thus represent its norm as

$$\|W\|_{C^{(d/2+\kappa)\vee 2}} = \sup_{L \in \mathcal{L}} |L(W)|,$$

where $\mathcal{L}$ is a countable set of bounded linear functionals on $(\mathcal{S}, \|\cdot\|_{C^{(d/2+\kappa)\vee 2}})$. Applying Fernique's theorem [8, Theorem 2.1.20] to the centred Gaussian process $(X(L) = L(W) : L \in \mathcal{L})$, we have $E\|W\|_{C^{(d/2+\kappa)\vee 2}} = E \sup_{L \in \mathcal{L}} |X(L)| \leq D < \infty$, and for $M = M(D) > 0$ large enough and since $\sqrt{T}\varepsilon_T \rightarrow \infty$,

$$\Pi(\mathcal{B}_{T,2}^c) \leq \Pi_W \left(W : \|W\|_{C^{(d/2+\kappa)\vee 2}} - E\|W\|_{C^{(d/2+\kappa)\vee 2}} \geq M\sqrt{T}\varepsilon_T \right) \leq 2e^{-D'M^2T\varepsilon_T^2}$$

for some fixed constant $D' > 0$. Taking $M > 0$ large enough, the right-hand side can be made less than $\frac{1}{2}e^{-KT\varepsilon_T^2}$, concluding the proof.

(ii). Turning to p -exponential priors, define the set

$$\mathcal{B}'_T = \{B = B'_1 + B'_2 + B'_3 : B'_i \in V_J \cap \dot{L}^2, \|B'_1\|_\infty \leq \bar{\varepsilon}_T, \\ \|B'_2\|_{H^{s+1+d/2-d/p}} \leq M^{\frac{p}{2}}(T\varepsilon_T^2)^{\frac{1}{2}-\frac{1}{p}}, \|B'_3\|_{B_{pp}^{s+1}} \leq M\}.$$

We lower bound the prior probability of $\mathcal{B}'_T$ using the generalization of Borell's inequality to p -exponential measures. The space of admissible shifts (cfr. Proposition 2.7 in [1]) of the scaled p -exponential random element $B = W/(T\varepsilon_T^2)^{\frac{1}{p}}$ is $\mathcal{Q} = V_J \cap \dot{L}^2$, with norm

$$\|h\|_{\mathcal{Q}} = (T\varepsilon_T^2)^{\frac{1}{p}} \left(\sum_{l=0}^J \sum_r 2^{2l(s+1+\frac{d}{2}-\frac{d}{p})} |\langle h, \Phi_{lr} \rangle_2|^2 \right)^{\frac{1}{2}} = (T\varepsilon_T^2)^{\frac{1}{p}} \|h\|_{H^{s+1+d/2-d/p}}, \quad h \in \mathcal{Q}.$$

Then, recalling the $\mathcal{Z}$ -norm defined in (26), Proposition 2.15 in [1] implies

$$\Pi(\mathcal{B}'_T) = \Pi(B = B'_1 + B'_2 + B'_3 : \|B'_1\|_\infty \leq \bar{\varepsilon}_T, \|B'_2\|_{\mathcal{Q}} \leq (M^p T\varepsilon_T^2)^{\frac{1}{2}}, \|B'_3\|_{\mathcal{Z}} \leq M(T\varepsilon_T^2)^{\frac{1}{p}}) \\ \geq 1 - \frac{1}{\Pi(\|B\|_\infty \leq \bar{\varepsilon}_T)} \exp(-(M^p/k)T\varepsilon_T^2)$$

for some $k = k(p) > 0$. By an analogous small ball computation as in the proof of Proposition 6.3 in [1], it follows that as $T \rightarrow \infty$

$$-\log \Pi(\|B\|_\infty \leq \bar{\varepsilon}_T) = -\log \Pi_W(\|W\|_\infty \leq \bar{\varepsilon}_T(T\varepsilon_T^2)^{\frac{1}{p}}) \\ \simeq [\bar{\varepsilon}_T(T\varepsilon_T^2)^{\frac{1}{p}}]^{\frac{d}{s+1-d/p}} = T\varepsilon_T^2.$$

Thus, for some constant $c_1 = c_1(s, p, d) > 0$, we have $\Pi(\mathcal{B}'_T) \geq 1 - e^{-[(M^p/k)-c_1]T\varepsilon_T^2}$, so that for any $K > 0$ we can choose $M = M(K, c_1, k) = M(K, s, p, d) > 0$ large enough to obtain $\Pi(\mathcal{B}'_T) \geq 1 - e^{-KT\varepsilon_T^2}$.

We conclude the proof by showing that $B'_T \subset \mathcal{B}_T$. First, since $B'_1 \in V_J$ we have $\|B'_1\|_{C^1} \leq 2^J \|B'_1\|_\infty \lesssim T^{1/(2s+d)} \bar{\varepsilon}_T = \varepsilon_T$, and therefore both norm bounds on $B_1 = B'_1$ in (27) are satisfied. Also note that

$$(34) \quad \|B'_1\|_{C^{(d/2+\kappa)\vee 2}} \leq 2^{J(d/2+\kappa)\vee 2} \|B'_1\|_\infty \lesssim T^{\frac{(d/2+\kappa)\vee 2}{2s+d}} \bar{\varepsilon}_T = o(1)$$

since by assumption $s+1 > (d/2+\kappa) \vee 2 + d/p$. Next, since also $B'_2 \in V_J$, using the continuous embedding $H^{s+1}(\mathbb{T}^d) \subseteq B_{pp}^{s+1}(\mathbb{T}^d)$, $p \leq 2$ (p.33 in [9]),

$$\begin{aligned} \|B'_2\|_{B_{pp}^{s+1}} &\lesssim \|B'_2\|_{H^{s+1}} \leq 2^{Jd(\frac{1}{p}-\frac{1}{2})} \|B'_2\|_{H^{s+1+d/2-d/p}} \\ &\leq 2^{Jd(\frac{1}{p}-\frac{1}{2})} M^{\frac{p}{2}} (T\varepsilon_T^2)^{\frac{1}{2}-\frac{1}{p}} \simeq M^{\frac{p}{2}}. \end{aligned}$$

Thus, taking $B_2 = B'_2 + B'_3$ implies $\|B_2\|_{B_{pp}^{s+1}} \lesssim M^{p/2} + M \lesssim M$ for $M \geq 1$ as required. Finally, using (34) and the continuous embedding $B_{pp}^{s+1} \subset C^{(d/2+\kappa)\vee 2}$ holding for all $s+1 > (d/2+\kappa) \vee 2 + d/p$ (p.170 in [19], using the embedding $B_{pp}^{s+1} \subset B_{p1}^{s'} \subset C^{(d/2+\kappa)\vee 2}$ holding for $s+1 > s' > (d/2+\kappa) \vee 2 + d/p$) if $d \leq 3$, in which in case $(d/2+\kappa) \vee 2 = 2 \in \mathbb{N}$ for small κ),

$$\|B\|_{C^{(d/2+\kappa)\vee 2}} \leq \|B'_1\|_{C^{(d/2+\kappa)\vee 2}} + \|B_2\|_{B_{pp}^{s+1}} \lesssim M,$$

concluding the proof. $\square$

B.3. Proof of Lemma 5.4.

LEMMA 5.4. *For $s, M, \kappa > 0$ and $p \in [1, 2]$, let $\mathcal{B}_T$ be the set in (27), with $\varepsilon_T, \bar{\varepsilon}_T$ as in Lemma 5.2. If $2^J \simeq T^{1/(2s+d)}$, then there exists a finite constant C depending on s, d, m and the wavelet basis $\{\Phi_{lr}\}$ such that*

$$\left\{ \mu_B = \frac{e^{2B}}{\int_{\mathbb{T}^d} e^{2B(x)} dx} : B \in \mathcal{B}_T \right\} \subset \{ \mu : \|\mu - P_J \mu\|_{W^{1,p}} \leq C\varepsilon_T \}.$$

PROOF. Since $\|B\|_\infty \leq \|B\|_{C^2} \leq M$ for every $B \in \mathcal{B}_T$, this implies $e^{-2M} \leq \int_{\mathbb{T}^d} e^{2B} dx \leq e^{2M}$ and hence $\|\mu_B - P_J \mu_B\|_{W^{1,p}} \leq e^{2M} \|e^{2B} - P_J e^{2B}\|_{W^{1,p}}$, so it suffices to bound the last quantity. For a function f on the torus $\mathbb{T}^d$, denote by $\bar{f}$ its periodic extension to $\mathbb{R}^d$. Recall that the periodic projection satisfies $P_J f(x) = \int_{\mathbb{R}^d} K_J(x, y) \bar{f}(y) dy$ for all $x \in (0, 1]^d$, where $K_J(x, y) = 2^{Jd} \sum_{k \in \mathbb{Z}^d} \phi(2^J x - k) \phi(2^J y - k)$ is the unperiodized wavelet kernel and ϕ is the unperiodized father wavelet used in the construction of the periodized wavelet basis (see (4.127) in [8]). Using that $\int_{\mathbb{R}^d} K_J(x, y) dy = 1$ for all $x \in (0, 1]^d$ and writing $B = B_1 + B_2$ as in (27),

$$\begin{aligned} |\partial_{x_i}(e^{2B} - P_J e^{2B})(x)| &= \left| \partial_{x_i} \int_{\mathbb{R}^d} K_J(x, y) (e^{2B_1(x)+2B_2(x)} - e^{2\bar{B}_1(y)+2\bar{B}_2(y)}) dy \right| \\ &\leq \left| \partial_{x_i} \left(e^{2B_1(x)} \int_{\mathbb{R}^d} K_J(x, y) (e^{2B_2(x)} - e^{2\bar{B}_2(y)}) dy \right) \right| \\ &\quad + \left| \partial_{x_i} \left(\int_{\mathbb{R}^d} K_J(x, y) (e^{2B_1(x)} - e^{2\bar{B}_1(y)}) e^{2\bar{B}_2(y)} dy \right) \right| \\ &\leq e^{2B_1(x)} \left| 2\partial_{x_i} B_1(x) [e^{2B_2(x)} - P_J e^{2B_2(x)}] + \partial_{x_i} [e^{2B_2(x)} - P_J e^{2B_2(x)}] \right| \\ &\quad + \left| \int_{\mathbb{R}^d} \partial_{x_i} K_J(x, y) (e^{2B_1(x)} - e^{2\bar{B}_1(y)}) e^{2\bar{B}_2(y)} dy \right| \\ &\quad + e^{2B_1(x)} \left| 2\partial_{x_i} B_1(x) \int_{\mathbb{R}^d} K_J(x, y) e^{2\bar{B}_2(y)} dy \right|. \end{aligned}$$

Using that $\int_{\mathbb{R}^d} \partial_{x_i} K_J(x, y) dy \lesssim 2^J$ by the localization property of wavelets and that $|e^x - 1| \lesssim x$ for small $|x|$, the last display is bounded by a multiple of

$$e^{2\|B_1\|_\infty} \|B_1\|_{C^1} \left| (e^{2B_2} - P_J e^{2B_2})(x) \right| + e^{2\|B_1\|_\infty} \left| \partial_{x_i} (e^{2B_2} - P_J e^{2B_2})(x) \right| \\ + 2^J e^{2\|B_2\|_\infty} \|B_1\|_\infty + e^{2\|B_1\|_\infty} \|B_1\|_{C^1} |P_J e^{2B_2}(x)|.$$

Taking p^{th} powers and integrating, and using the embedding $B_{p1}^0(\mathbb{T}^d) \subset L^p(\mathbb{T}^d)$ (Theorem 1, p.163 in [19]), then yields

$$\|\partial_{x_i} (e^{2B} - P_J e^{2B})\|_p \lesssim e^{2\|B_1\|_\infty} \|B_1\|_{C^1} \|e^{2B_2} - P_J e^{2B_2}\|_{B_{p1}^0} + e^{2\|B_1\|_\infty} \|e^{2B_2} - P_J e^{2B_2}\|_{B_{p1}^1} \\ + 2^J e^{2\|B_2\|_\infty} \|B_1\|_\infty + e^{2\|B_1\|_\infty} \|B_1\|_{C^1} \|P_J e^{2B_2}\|_{B_{p1}^0}.$$

By Lemma 5.3, $\|e^{2B_2}\|_{B_{pp}^{s+1}} \lesssim 1 + \|B_2\|_{B_{pp}^{s+1}} + \|B_2\|_{B_{pp}^{s+1}}^{s+1} \leq c(m, s, p)$, which implies

$$\|e^{2B_2} - P_J e^{2B_2}\|_{B_{p1}^1} = \sum_{l>J} 2^{l(1+d/2-d/p)} \left(\sum_r |\langle e^{2B_2}, \Phi_{lr} \rangle_2|^p \right)^{1/p} \\ = \sum_{l>J} 2^{-ls} \left(2^{pl(s+1+d/2-d/p)} \sum_r |\langle e^{2B_2}, \Phi_{lr} \rangle_2|^p \right)^{1/p} \\ \leq \|e^{2B_2}\|_{B_{pp}^{s+1}} \sum_{l>J} 2^{-ls} \lesssim 2^{-Js}.$$

By a similar computation, $\|e^{2B_2} - P_J e^{2B_2}\|_{B_{p1}^0} \lesssim 2^{-J(s+1)}$, while by Hölder's inequality with exponents $(p/(p-1), p)$,

$$\|P_J e^{2B_2}\|_{B_{p1}^0} = \sum_{l \leq J} 2^{l(d/2-d/p)} \left(\sum_r |\langle e^{2B_2}, \Phi_{lr} \rangle_2|^p \right)^{1/p} \\ \leq \left(\sum_{l \leq J} [2^{-l(s+1)}]^{p-1} \right)^{\frac{p-1}{p}} \left(\sum_{l \leq J} 2^{pl(s+1+d/2-d/p)} \sum_r |\langle e^{2B_2}, \Phi_{lr} \rangle_2|^p \right)^{1/p} \\ \lesssim \|e^{2B_2}\|_{B_{pp}^{s+1}} \lesssim 1.$$

Combining the above bounds, using the definition of $\varepsilon_T, \bar{\varepsilon}_T$ in Lemma 5.2 and that $2^J \simeq T^{1/(2s+d)}$, we thus obtain that for any $B \in \mathcal{B}_T$,

$$\|\partial_{x_i} (e^{2B} - P_J e^{2B})\|_p \lesssim e^{2\bar{\varepsilon}_T} \varepsilon_T 2^{-J(s+1)} + e^{2\bar{\varepsilon}_T} 2^{-Js} + 2^J e^{2M} \bar{\varepsilon}_T + e^{2\bar{\varepsilon}_T} \varepsilon_T \lesssim \varepsilon_T.$$

By a similar, in fact easier, computation, we also obtain

$$\|e^{2B} - P_J e^{2B}\|_p \lesssim e^{2\bar{\varepsilon}_T} 2^{-J(s+1)} + e^{2M} \bar{\varepsilon}_T = o(\varepsilon_T).$$

The required bias bound then follows from the last two displays. $\square$

REFERENCES

- [1] AGAPIOU, S., DASHTI, M. and HELIN, T. (2021). Rates of contraction of posterior distributions based on p -exponential priors. *Bernoulli* **27** 1616–1642. [MR4278794](#)
- [2] BASS, R. F. (2011). *Stochastic processes*. Cambridge Univ. Press, Cambridge. [MR2856623](#)
- [3] BERS, L., JOHN, F. and SCHECHTER, M. (1964). *Partial differential equations. Lectures in Applied Mathematics, Vol. III*. Wiley & Sons, Inc. New York-London-Sydney.
- [4] BOGACHEV, V. I. (1998). *Gaussian measures. Mathematical Surveys and Monographs* **62**. American Mathematical Society, Providence, RI. [MR1642391](#)

- [5] FOLLAND, G. B. (1999). *Real analysis*, second ed. *Pure and Applied Mathematics (New York)*. John Wiley & Sons, Inc., New York Modern techniques and their applications, A Wiley-Interscience Publication. [MR1681462](#)
- [6] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. [MR1790007](#)
- [7] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of nonparametric Bayesian inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge University Press, Cambridge. [MR3587782](#)
- [8] GINÉ, E. and NICKL, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press, New York. [MR3588285](#)
- [9] LASSAS, M., SAKSMAN, E. and SILTANEN, S. (2009). Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl. Imaging* **3** 87–122. [MR2558305](#)
- [10] LATALA, R. and MATLAK, D. (2017). Royen’s proof of the Gaussian correlation inequality. In *Geometric aspects of functional analysis. Lecture Notes in Math.* **2169** 265–275. Springer, Cham. [MR3645127](#)
- [11] LI, W. V. and LINDE, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.* **27** 1556–1578. [MR1733160](#)
- [12] MONARD, F., NICKL, R. and PATERNAIN, G. P. (2021). Consistent inversion of noisy non-Abelian X-ray transforms. *Comm. Pure Appl. Math.* **74** 1045–1099. [MR4230066](#)
- [13] NICKL, R. and RAY, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist.* **48** 1383–1408. [MR4124327](#)
- [14] PAPASILIOPOULOS, O., POKERN, Y., ROBERTS, G. O. and STUART, A. M. (2012). Nonparametric estimation of diffusions: a differential equations approach. *Biometrika* **99** 511–531.
- [15] 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](#)
- [16] PROTTER, P. E. (2004). *Stochastic integration and differential equations*, second ed. *Applications of Mathematics (New York)* **21**. Springer-Verlag, Berlin Stochastic Modelling and Applied Probability. [MR2020294](#)
- [17] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian processes for machine learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. [MR2514435](#)
- [18] ROYEN, T. (2014). A simple proof of the Gaussian correlation conjecture extended to some multivariate gamma distributions. *Far East J. Theor. Stat.* **48** 139–145. [MR3289621](#)
- [19] SCHMEISSER, H.-J. and TRIEBEL, H. (1987). *Topics in Fourier analysis and function spaces. A Wiley-Interscience Publication*. John Wiley & Sons, Ltd., Chichester. [MR891189](#)
- [20] TRIEBEL, H. (2010). *Theory of function spaces. Modern Birkhäuser Classics*. Birkhäuser/Springer Basel AG, Basel Reprint of 1983 edition [MR0730762], Also published in 1983 by Birkhäuser Verlag [MR0781540]. [MR3024598](#)
- [21] 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](#)
- [22] VAN DER VAART, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. [MR1652247](#)
- [23] VAN WAAIJ, J. (2019). Adaptive posterior contraction rates for empirical Bayesian drift estimation of a diffusion. *arXiv e-prints* arXiv:1909.12710.
- [24] 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](#)