

NONPARAMETRIC BAYESIAN POSTERIOR CONTRACTION RATES FOR DISCRETELY OBSERVED SCALAR DIFFUSIONS¹

BY RICHARD NICKL AND JAKOB SÖHL

University of Cambridge

We consider nonparametric Bayesian inference in a reflected diffusion model $dX_t = b(X_t) dt + \sigma(X_t) dW_t$, with discretely sampled observations $X_0, X_\Delta, \dots, X_{n\Delta}$. We analyse the nonlinear inverse problem corresponding to the “low frequency sampling” regime where $\Delta > 0$ is fixed and $n \rightarrow \infty$. A general theorem is proved that gives conditions for prior distributions Π on the diffusion coefficient σ and the drift function b that ensure minimax optimal contraction rates of the posterior distribution over Hölder–Sobolev smoothness classes. These conditions are verified for natural examples of nonparametric random wavelet series priors. For the proofs, we derive new concentration inequalities for empirical processes arising from discretely observed diffusions that are of independent interest.

1. Introduction. Many fundamental models for dynamic stochastic phenomena in continuous time are based on the concept of a *diffusion*, whose evolution is described mathematically by

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

where W_t is a standard Brownian motion. Diffusions occur naturally in the physical and biological sciences, in economics and elsewhere, and their deep relationship to stochastic and partial differential equations makes them a central object of study in modern mathematics. Various specifications of the *drift function* b and the *diffusion coefficient* σ lead to a flexible class of random continuous motions. In scientific applications, a key challenge is to recover the parameters b, σ from some form of observations of the diffusion. Unless specific knowledge is available, the resulting statistical models for the parameters σ, b and the probability laws $\mathbb{P}_{\sigma b}$ of the Markov process $(X_t : t \geq 0)$ are naturally *infinite-dimensional* (= “nonparametric”).

Statistical observations in the real world usually are collected in a discrete fashion, say in form of observed increments $X_0, X_\Delta, \dots, X_{n\Delta}$ of the diffusion, where $1/\Delta$ is the sampling frequency. We are interested in the possibly most realistic scenario where $\Delta > 0$ is fixed and more information accrues in form of an increasing

Received October 2015; revised July 2016.

¹Supported by European Research Council (ERC) under Grant No. 647812.

MSC2010 subject classifications. Primary 62G05; secondary 60J60, 62F15, 62G20.

Key words and phrases. Nonlinear inverse problem, Bayesian inference, diffusion model.

sampling horizon $n\Delta \rightarrow \infty$. As revealed in the seminal paper by Gobet, Hoffmann and Reiß [13], this “low frequency” sampling regime implies that inference on (σ, b) constitutes a nonlinear nonparametric inverse problem, and the authors solve this problem in a minimax way by a delicate estimation technique based on ideas from spectral theory.

Alternative methodology for nonparametric inference in diffusion models has been put forward recently, notably of a Bayesian flavour; see Roberts and Stramer [23], Papaspiliopoulos et al. [19], Pokern et al. [21], van der Meulen et al. [31], van Waaij and van Zanten [29] and references therein. While such Bayesian methods are attractive in applications [14, 26, 30], particularly since they provide associated uncertainty quantification procedures (“credible regions”), our understanding of their frequentist sampling performance is extremely limited. This is particularly so in the “low frequency” regime when $\Delta > 0$ is thought to be fixed: the only references we are aware of are the consistency results in [15, 16, 32], which only hold under the very restrictive assumption that σ is constant and known, and only in a weak topology. As pointed out by Stuart [26] and van Zanten [30], obtaining theoretical performance guarantees for Bayesian algorithms in nonlinear inverse problems is, however, of key importance if such methods are to be used in scientific applications. Only very few rigorous results are currently available.

In this paper, we give the first proof of the fact that nonparametric prior distributions on the diffusion parameters (σ, b) give rise to posterior distributions that contract at the (minimax) frequentist optimal convergence rates over natural regularity classes, in the low frequency sampling regime. This is achieved by using the generic “testing approach” introduced in the landmark paper Ghosal et al. [9]; see also [10] and [34]—but the adaptation to the diffusion case requires the resolution of two major mathematical obstacles to obtain satisfactory results:

- The “*small ball probability conditions*” need to reflect the inverse problem nature of the discrete sampling scheme, and the resulting perturbation of the “information theoretic distance” (KL-divergence) associated to the statistical experiment needs to be precisely quantified. For linear inverse problems, this has already been noted in the paper by Ray [22]. In the nonlinear diffusion setting here, however, the situation is much more complicated, and requires a fine analysis of the inverse operator of the infinitesimal generator of the diffusion (which could be viewed as the linearisation of the nonlinear inverse operator). In the case of high frequency ($\Delta \rightarrow 0$) or continuous observations, small ball probabilities may be computed for an information distance closely related to the L^2 -distance on b and σ (see [33]), but to obtain optimal results in the low frequency setting one has to show that instead small ball probabilities may be computed in a weaker norm—precisely, as we show, in a certain negative order Besov norm. Simultaneously, one has to ensure that the invariant measure μ is correctly modelled by the (induced) prior too—note that the smoothness degree of μ is generally not identified by the regularity of (σ, b) . This last fact also should guide practitioners who often devise

priors for σ and b without paying attention to the implied model for the invariant measure.

- The *construction of frequentist tests* with sufficiently good exponential error bounds for type two errors in a large enough support set of the prior in [9, 10] relies on properties of the likelihood ratio test and the associated Hellinger distance between experiments. In the setting of diffusions, this approach appears difficult to implement—instead we use the “concentration of measure” approach of Giné and Nickl [11] to the construction of such tests. To do this, we prove a Bernstein-type inequality for empirical processes driven by discretely sampled diffusions, relying on work of Adamczak [1], and use it to derive sharp concentration bounds for the estimators (and resulting plug-in tests) put forward in Gobet et al. [13]. These concentration results, which are of independent interest, are derived in Section 3.

We demonstrate that our general conditions are verified for natural nonparametric priors on (σ, b) . It is convenient to give a hierarchical prior specification that first models the inverse diffusion coefficient σ^{-2} , and then, conditional on σ^2 , the drift function b , which explicitly generates a prior for the invariant measure μ too. The individual prior choices are quite flexible and allow for general random series priors, as we show, with one technically vital restriction that they are constrained to a fixed regularity class that ensures sufficient smoothness of the individual parameters. This is necessary to deduce various probabilistic properties of the diffusion that our proofs rely on. In particular, following [13], we rely on the assumption that the diffusion considered is a reflected one, and hence lives in a compact interval of \mathbb{R} . This corresponds to the usual von Neumann boundary conditions required for the infinitesimal generator L to be injective and to have a discrete spectrum. As a consequence, to cope with these boundary conditions, we model b and σ only in the interior of the given interval (this also has some deeper mathematical reasons since the second eigenfunction of L identifies b, σ only in the interior of the domain, and since our approach to construct tests is based on first estimating this eigenfunction). Again, these are concessions to the mathematical intricacies of the problem at hand, and further hard work will be required to alleviate those. We discuss possible extensions and limitations of our approach in Section 2.3.3 below.

2. Main results.

2.1. *A nonparametric model for diffusions on $[0, 1]$.* Consider a scalar diffusion process $(X_t : t \geq 0)$ on $[0, 1]$ starting at $X_0 = x_0$, and whose evolution is described by the stochastic differential equation (SDE):

$$(1) \quad dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad t \geq 0,$$

where the process is reflected at the boundary points $\{0, 1\}$ (for a precise definition see Section 3 below). For the pair $\vartheta = (\sigma, b) \in C([0, 1]) \times C([0, 1])$, we maintain

the following model:

$$\Theta := \left\{ \vartheta = (\sigma, b) : b(0) = b(1) = \sigma'(0) = \sigma'(1) = 0, b', \sigma', \sigma'' \text{ exist}, \right. \\ \left. \max(\|b\|_\infty, \|b'\|_\infty, \|\sigma\|_\infty, \|\sigma'\|_\infty, \|\sigma''\|_\infty) \leq D, \inf_{x \in [0,1]} \sigma^2(x) \geq d \right\},$$

where D, d are arbitrary fixed positive constants. For $(\sigma, b) \in \Theta$, the SDE (1) has a pathwise solution described by the Markov process $(X_t : t \geq 0)$ with invariant measure $\mu = \mu_{\sigma b}$, whose law we denote by $\mathbb{P}_{\sigma b}$ whenever $X_0 \sim \mu$. The observation scheme considered is such that increments $X_0, X_\Delta, \dots, X_{n\Delta}$ are sampled at distance $\Delta > 0$, and we study statistical inference on σ and b when n tends to infinity as $\Delta > 0$ remains fixed (the “low frequency sampling” regime). Thus, the $X_{i\Delta}$ ’s form an ergodic Markov chain and we write $p_{\sigma b}(\Delta, x, y)$ for the associated transition probability density functions with respect to Lebesgue measure dy . We shall also—in abuse of notation—write μ both for the invariant measure and its density.

Let Π be a (prior) probability distribution on some σ -field \mathcal{S} of subsets of Θ , and given $(\sigma, b) \sim \Pi$ assume that the law of $(X_t : t \geq 0) | (\sigma, b)$ is described by the diffusion (1) started in the invariant measure $\mu_{\sigma b}$. If the mapping $(\sigma, b) \mapsto p_{\sigma b}(\Delta, x, y)$ is \mathcal{S} - $\mathcal{B}_{\mathbb{R}}$ measurable for all x, y , then by standard arguments (as in Chapter 7.3 in [12]), the posterior distribution given the discrete sample from the diffusion is

$$(2) \quad (\sigma, b) | X_0, X_\Delta, \dots, X_{n\Delta} \\ \sim \frac{\mu_{\sigma b}(X_0) \prod_{i=1}^n p_{\sigma b}(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) d\Pi((\sigma, b))}{\int_{\Theta} \mu_{\sigma b}(X_0) \prod_{i=1}^n p_{\sigma b}(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) d\Pi((\sigma, b))}.$$

We wish to devise natural conditions on the prior Π that imply that the posterior distribution contracts at the optimal convergence rate δ_n in some distance function d about any fixed “true” parameter pair $\vartheta_0 = (\sigma_0, b_0) \in \Theta$. More precisely, we wish to prove that, as $n \rightarrow \infty$,

$$\Pi(\vartheta : d(\vartheta_0, \vartheta) > \delta_n | X_0, X_\Delta, \dots, X_{n\Delta}) \rightarrow 0 \quad \text{in } \mathbb{P}_{\sigma_0 b_0}\text{-probability,}$$

under the “frequentist” assumption that $(X_t : t \geq 0) \sim \mathbb{P}_{\sigma_0 b_0}$. The rate δ_n will depend on regularity properties of (σ_0, b_0) that we describe now.

2.2. *Contraction theorem.* In [13], it was shown that the frequentist minimax rates for estimating the parameter (σ, b) in $L^2([A, B])$ -loss ($0 < A < B < 1$) are given by

$$(3) \quad n^{-s/(2s+3)} \quad \text{for } \sigma^2 \quad \text{and} \quad n^{-(s-1)/(2s+3)} \quad \text{for } b$$

whenever $(\sigma, b) \in \Theta_s$, where the regularity classes $\Theta_s \subseteq \Theta$ are defined as

$$\Theta_s := \left\{ \vartheta = (\sigma, b) \in \Theta : \|\sigma\|_{H^s} \leq D, \|b\|_{H^{s-1}} \leq D \right\}, \quad s \geq 2,$$

with H^s the usual L^2 -Sobolev space over $[0, 1]$. This particular coupling of the regularities s of σ and $s - 1$ of b is natural; see Remark 7 below.

The above rates reflect the recovery complexity of an ill-posed problem of order one and two, respectively. For the Bayesian posterior distribution to have good frequentist properties, it is well known [9] that the prior should charge small neighbourhoods of the “true pair” (σ_0, b_0) with sufficient probability, where “neighbourhood” is understood with respect to the information distance induced by the observations. As noted by [22] for linear inverse problems,

$$Y = Af + \varepsilon, \quad A : L^2 \rightarrow L^2 \quad \text{linear,}$$

with unknown parameter $f \in L^2$ and Gaussian white noise ε , a key point is to take advantage of the fact that the usual information distance $\|f\|_{L^2}$ (when $A = \text{Id}$) is transformed to $\|Af\|_{L^2}$, which typically corresponds to a negative (or dual) Sobolev norm $\|f\|_{H^{-w}} = \|f\|_{(H^w)^*}$ induced by the eigen-basis of A , and where w denotes the level of ill-posedness.

One of the key contributions of this paper is to produce a similar result in our nonlinear and non-Gaussian setting of discretely sampling a diffusion. To obtain sharp results, the Hilbert scale of Sobolev norms will have to be replaced by more flexible Besov norms. To this end, for $s > 0$ denote by $(B_{1\infty}^s)^*$ the dual space of the Besov space $B_{1\infty}^s = B_{1\infty}^s([0, 1])$, equipped with the usual dual norm [see (33) below]. We refer to [17, 28] or Chapter 4.3 in [12] for the usual definitions and basic properties of Besov spaces. We further note that any prior distribution on (σ, b) induces a prior distribution on the invariant measure μ of the diffusion [see (10) below], and the following result implicitly requires this induced prior to correctly model the parameter μ , also. See Remark 7 for discussion.

In what follows, for two sequences $(a_m : m \in \mathbb{N})$, $(b_m : m \in \mathbb{N})$, we write $a_m \lesssim b_m$ whenever $a_m \leq Cb_m$ for all $m \in \mathbb{N}$ and some fixed constant $C > 0$, and we write $a_m \simeq b_m$ whenever both $a_m \lesssim b_m$ and $b_m \lesssim a_m$ hold.

THEOREM 1. *Let $\Pi = \Pi_n$ be a sequence of prior distributions on Θ , suppose that $X_0, \dots, X_{n\Delta}$ are discrete observations of a diffusion process (1) started in the stationary distribution μ , and let $\Pi(\cdot | X_0, \dots, X_{n\Delta})$ be the resulting posterior distribution (2) on Θ .*

Assume Π satisfies for some $(\sigma_0, b_0) \in \Theta_s$ and $\mu_0 \in L^2$, that:

- (i) $\Pi(\Theta_s) = 1$ for some $s \geq 2$ (and for some $d > 0, D > 0$),
- (ii) there exists a constant $C > 0$ and a sequence ε_n satisfying

$$n^{-(s+1)/(2s+3)} \lesssim \varepsilon_n \lesssim n^{-3/8}(\log n)^{-1/2}$$

such that for all n large enough

$$(4) \quad \begin{aligned} & \Pi((\sigma, b) \in \Theta : \|\mu - \mu_0\|_{L^2([0,1])} + \|\sigma^{-2} - \sigma_0^{-2}\|_{(B_{1\infty}^1)^*} \\ & + \|b - b_0\|_{(B_{1\infty}^2)^*} < \varepsilon_n) \\ & \geq e^{-Cn\varepsilon_n^2}. \end{aligned}$$

Then if $(X_t : t \geq 0) \sim \mathbb{P}_{\sigma_0 b_0}$ where $X_0 \sim \mu_0$, for true parameters (σ_0, b_0) with associated invariant measure $\mu_0 = \mu_{\sigma_0 b_0}$, the posterior distribution contracts about (σ_0^2, b_0) in $L^2 \equiv L^2([A, B])$ for any $0 < A < B < 1$, at rates

$$\delta_n \equiv n\varepsilon_n^3 \quad \text{and} \quad \delta'_n \equiv n^2\varepsilon_n^5;$$

that is, for some fixed constant M , as $n \rightarrow \infty$ and in $\mathbb{P}_{\sigma_0 b_0}$ -probability,

$$\Pi((\sigma, b) : \|\sigma^2 - \sigma_0^2\|_{L^2} > M\delta_n \text{ or } \|b - b_0\|_{L^2} > M\delta'_n | X_0, \dots, X_{n\Delta}) \rightarrow 0.$$

REMARK 2. *Optimal rates.* The optimal choice of ε_n is of the order

$$\varepsilon_n \simeq n^{-(s+1)/(2s+3)}$$

in which case $\varepsilon_n = O(n^{-3/8}(\log n)^{-1/2})$ is always satisfied since $s \geq 2$, and the resulting contraction rates are of the desired minimax order:

$$\delta_n \simeq n^{-s/(2s+3)}, \quad \delta'_n \simeq n^{-(s-1)/(2s+3)}.$$

2.3. Examples of prior distributions.

2.3.1. *Some preliminaries on function spaces.* We now show how Theorem 1 applies to some concrete prior distributions. To do so, we need to define the following Hölder-type function spaces \mathcal{C}^t :

DEFINITION 3. For $t > 0$ and $[t]$, the largest integer $k \leq t$ we define

$$\mathcal{C}^t([0, 1]) := \{f \in C([0, 1]) : \|f\|_{\mathcal{C}^t} < \infty\},$$

$$\text{where } \|f\|_{\mathcal{C}^t} := \sum_{k=0}^{[t]} \|D^k f\|_{\infty} + \sup_{0 \leq x < x+h \leq 1} \frac{|D^{[t]} f(x+h) - D^{[t]} f(x)|}{h^{t-[t]} \log(1/h)^{-2}}.$$

The additional logarithmic factor in the Hölder condition is convenient as then

$$f \in \mathcal{C}^t \quad \Rightarrow \quad f \in H^t \cap B_{\infty 1}^t$$

follows, which allows to combine knowledge of spectral properties of the diffusion expressed in terms of Sobolev H^t -norms with wavelet characterisations of Hölder and Besov spaces. Note that the continuous imbeddings of \mathcal{C}^t into H^t and into $B_{\infty 1}^t$ follow easily from wavelet characterisations of the norms of these spaces (see [17] or [12], Chapter 4.3, where we note that $H^t = B_{22}^t$), and these also imply that $B_{\infty 1}^t, t \in \mathbb{N}$, is continuously imbedded into the classical spaces \mathcal{C}^t of t -times continuously differentiable functions.

In fact, we will work with the equivalent wavelet norm of \mathcal{C}^t given by

$$(5) \quad \|f\|_{\mathcal{C}^t} \equiv \|f\|_{\mathcal{C}^{t,2}}, \quad \|f\|_{\mathcal{C}^{t,\gamma}} = \sup_{l,k} 2^{l(t+1/2)\gamma} |\langle f, \psi_{lk} \rangle|,$$

where $\{\psi_{lk} : k = 0, \dots, 2^l - 1, l \geq J_0 - 1\}$, $J_0 \geq 2$, $J_0 \in \mathbb{N}$, is a sufficiently regular boundary-adapted Daubechies wavelet basis of $L^2([0, 1])$ that also generates H^t, \mathcal{C}^t as well as the Besov spaces B^t_{pq} (see Chapter 4.3.5 in [12]).

We shall assume that the true pair (σ_0, b_0) lies in $\Theta_s \cap (\mathcal{C}^s \times \mathcal{C}^{s-1})$. Moreover, to avoid tedious technicalities about boundary conditions, we assume that b_0 is supported in the interior of $[0, 1]$, and that $\log \sigma_0^{-2}$ and $\log \mu_0$ have expansions into wavelet series supported in the interior of $[0, 1]$. Such functions can be modelled by infinite Daubechies wavelet series ψ_{lk} that are supported in a given fixed interval $[A, B]$. The minimax estimation rates over functions satisfying these constraints are the same as those in (3) up to $\log n$ factors (see Remark 5 below), and the minor loss in generality comes at the gain of substantial technical simplifications. Thus, we adopt the following condition, formulated in terms of μ_0 and σ_0 (implicitly defining b_0).

ASSUMPTION 4. For $0 < A < B < 1$ given, let \mathcal{I} be the maximal set of double indices (l, k) such that the Daubechies wavelet functions ψ_{lk} , $(l, k) \in \mathcal{I}$, are all supported in $[A, B]$.

We assume that the invariant density $\mu_0 \in \mathcal{C}^{s+1}$ is of the form

$$\log \mu_0(x) = \sum_{l,k \in \mathcal{I}} \beta_{lk} \psi_{lk}(x), \quad x \in [0, 1],$$

with $2^{l(s+3/2)} l^2 |\beta_{lk}| \leq B$ for some $B > 0$.

We further assume that the diffusion coefficient $\sigma_0 \in \mathcal{C}^s$ has the form

$$\log \sigma_0^{-2}(x) = \sum_{l,k \in \mathcal{I}} \tau_{lk} \psi_{lk}(x), \quad x \in [0, 1],$$

where $2^{l(s+1/2)} l^2 |\tau_{lk}| \leq B$.

Note that the assumptions ensure that σ_0^{-2} and μ_0 are bounded and bounded away from zero on $[0, 1]$, and that as a consequence, so is $\sigma_0^2 \in \mathcal{C}^s$. It also implies that $2b_0 = (\sigma_0^2 \mu_0)' / \mu_0$ is contained in \mathcal{C}^{s-1} and supported in $[A, B]$ (since $\sigma_0^2 \mu_0$ is constant outside of that interval).

REMARK 5 (Minimax rates over \mathcal{C}^s -classes). Assumption 4 is restricting (σ_0, b_0) beyond having to lie in Θ_s . The lower bound proofs in [13] imply that these further restrictions do not change the minimax rates, except for a $(\log n)^\gamma$, $\gamma > 0$, factor induced by the weighting with the factor l^2 in the wavelet norm. Note that the lower bound in [13] is also based on wavelets that are supported in the interior of $[0, 1]$, and works with constant invariant density $\mu_0 = 1 \in \mathcal{C}^{s+1}$, which means that $2b$ just equals $(\sigma^2)'$.

2.3.2. *Random wavelet series prior.* We consider the following hierarchical prior specification. Let $s \geq 2$. Wavelet coefficients will be constructed from random variables drawn i.i.d. from probability density

$$(6) \quad \varphi : [-\tilde{B}, \tilde{B}] \rightarrow [0, \infty), \quad B \leq \tilde{B} < \infty, \quad \inf_{x \in [-B, B]} \varphi(x) \geq \zeta, \quad \zeta > 0.$$

This in particular includes the cases where φ is the density of a uniform $U[-B, B]$ random variable [so that $\zeta = 1/(2B)$], or the case where φ equals the density of the truncated normal distribution given by $\varphi(x) \simeq e^{-x^2/2}$ for $x \in [-B, B]$ and $\varphi(x) = 0$ otherwise [so that $\zeta \geq (2\pi)^{-1/2}e^{-B^2/2}$].

We first model $\log \sigma^{-2}$ as a wavelet series, that is,

$$\sigma^{-2}(x) = \exp \left\{ \sum_{l,k \in \mathcal{I}, l \leq L_n} 2^{-l(s+1/2)} l^{-2} u_{lk} \psi_{lk}(x) \right\}, \quad x \in [0, 1],$$

where u_{lk} are drawn i.i.d. from density φ satisfying (6). Here, we can take $L_n = \infty$ (so that the prior is independent of n) but in our result below we also allow for L_n to equal a sequence of integers diverging with n .

Conditional on σ we use the identity

$$(7) \quad 2b = \frac{(\sigma^2 \mu)'}{\mu} = (\sigma^2)' + \sigma^2 (\log \mu)',$$

and the law of $b|\sigma^2$ is modelled by taking a wavelet prior

$$H(x) = \sum_{l,k \in \mathcal{I}, l \leq \bar{L}_n} 2^{-l(s+3/2)} l^{-2} \bar{u}_{lk} \psi_{lk}(x), \quad x \in [0, 1], \quad \bar{L}_n \in \mathbb{N} \cup \{\infty\},$$

and the resulting prior $e^H / \int e^H$ on the parameter μ , where the \bar{u}_{lk} are drawn i.i.d. from density $\bar{\varphi}$ satisfying (6), independent of the u_{lk} 's from above. Concretely,

$$b|\sigma^2 = ((\sigma^2)' + \sigma^2 H')/2,$$

and the resulting prior distribution induced on $(\sigma^2, b) = (\sigma^2, ((\sigma^2)' + \sigma^2 H')/2)$ is denoted by $\Pi = \Pi_{L_n, \bar{L}_n}$.

PROPOSITION 6. *Let σ_0, μ_0 satisfy Assumption 4 for some $s \geq 2, B > 0$ and choose*

$$\varepsilon_n = n^{-(s+1)/(2s+3)} (\log n)^\eta, \quad \eta = \frac{s-1}{2s+3}.$$

Let $\Pi = \Pi_{L_n, \bar{L}}$ be the preceding prior and, if $l_n = \min(L_n, \bar{L}_n) < \infty$, assume $2^{-l_n(s+1)} \lesssim \varepsilon_n$. Then Π satisfies the hypotheses of Theorem 1 for this choice of ε_n , all D large and all $d > 0$ small enough. As a consequence, the resulting posterior distribution contracts about the true parameter (σ_0^2, b_0) at the minimax optimal rate within $\log n$ factors (Remark 2).

REMARK 7 (Coupling of smoothness indices and ill-posed inverse problems). In contrast to estimation of σ^2 and b , estimation of μ is not ill-posed and possible at the standard nonparametric rate $n^{-\alpha/(2\alpha+1)}$, when μ is α -smooth. These estimation problems are, however, interacting with each other. On the one hand σ , b and μ are closely related by classical identities [e.g., (7) and (10)], making it natural that σ is modelled one degree smoother than b . On the other hand, the smoothness of μ is *not* identified by the smoothness of b and σ (e.g., even for nonregular b , σ the invariant density μ can be very smooth, e.g., constant on $[0, 1]$). Our results rely on the assumption that μ_0 is at least $s + 1$ smooth [whenever $(\sigma, b) \in \Theta_s$], *and* that the prior implicitly models the regularity of μ correctly. This is related to the fact, made explicit in the proofs that follow, that the information distance between samples $X_0, X_\Delta, \dots, X_{n\Delta}$ from parameters (σ, b) and (σ_0, b_0) *does* necessarily involve the L^2 -distance $\|\mu_{\sigma b} - \mu_{\sigma_0 b_0}\|_{L^2}$, and the hierarchical prior from above takes this into account.

REMARK 8 (Credible sets). While our results imply that Bayesian recovery algorithms can be expected to work in principle in (scalar) diffusion models, we emphasise that mere contraction theorems as those obtained here do not yet justify the use of Bayesian posterior inference (“credible regions”) in scientific practice. This problem is more involved; see the recent paper [27] and its discussion. An interesting topic for future research in this direction would be to obtain nonparametric Bernstein–von Mises theorems as in [7, 8] for the diffusion model considered here. While the contraction results obtained here are useful for this too, obtaining exact posterior asymptotics will require a more elaborate analysis.

2.3.3. *Choice of the prior: Extensions and perspectives.* In computational practice, the methodology closest to the one considered here is described in Section 5.1 of [19], where a certain Gaussian prior is chosen for the drift function b , while the diffusion coefficient is modelled parametrically. A data augmentation method is devised that allows to sample from the posterior distribution (2) in this situation. The random wavelet series priors on b from the previous subsection allow for truncated Gaussian priors—by choosing B large enough our theory can approximate the case of a Gaussian prior on the drift function at least in practice. From a rigorous point of view, however, our proofs rely fundamentally on the technical restriction that the prior for (σ, b) concentrates on a fixed smoothness ball in $C^2 \times C^1$, a condition *not* satisfied by Gaussian priors. Whether it can be relaxed is not clear: for instance, the constant C in the Gaussian-type tails $e^{-Cne_n^2}$ of our tests scales unfavourably as $C \approx e^{-\|b\|_\infty}$. This is not an artefact of our concentration inequalities but corresponds precisely to the connection between mixing times of Markov chains and their spectral gap (see also [20]).

The wavelet priors used in the present paper are convenient in our proofs. They could be replaced by B -spline basis priors with random coefficients. In

fact, B -spline bases generate wavelet bases by a simple Gram–Schmidt orthonormalisation step (see [17], Section 1.3 and page 74), so that proofs would go through with only formal (but notationally cumbersome) changes.

Another extension of interest would be to allow for “adaptive” priors that select the generally unknown smoothness degree s by a hyper-prior. While in principle such results should be within the scope of our techniques, they would require significant modification of the spectral bias estimates from [13], and this is left for future research.

3. Proofs I: Concentration inequalities for reflected diffusions.

3.1. *Definitions and transition densities.* Let $b : [0, 1] \rightarrow \mathbb{R}$ be measurable and bounded, let $\sigma : [0, 1] \rightarrow (0, \infty)$ be continuous and let $\nu : [0, 1] \rightarrow \mathbb{R}$ satisfy $\nu(0) = 1, \nu(1) = -1$. Consider the reflected diffusion on $[0, 1]$

$$(8) \quad dX_t = b(X_t) dt + \sigma(X_t) dW_t + \nu(X_t) dL_t(X).$$

Here, $(W_t : t \geq 0)$ is a standard Brownian motion and $(L_t(X) : t \geq 0)$ is a nonanticipative continuous nondecreasing process which increases only for $X_t \in \{0, 1\}$. This model is considered in [13], and under the above conditions there exists a weak solution $(X_t : t \geq 0)$ of the SDE; see [25]. An in our setting equivalent construction of this reflected diffusion is by extending b and σ to be defined on \mathbb{R} as follows: First, we extend the functions to $(-1, 1]$ by $\sigma(x) = \sigma(-x)$ and $b(x) = -b(-x)$ for $x \in (-1, 0)$ and second to \mathbb{R} by $\sigma(x) = \sigma(x + 2k)$ and $b(x) = b(x + 2k)$ for all $x \in \mathbb{R}$ and $k \in \mathbb{Z}$ such that $x + 2k \in (-1, 1]$. If $(\sigma, b) \in \Theta$, then the so extended functions $\bar{\sigma}$ and \bar{b} are bounded Lipschitz functions on \mathbb{R} and we can define the strong Markov process $(Y_t : t \geq 0)$ as the pathwise solution of the equation

$$(9) \quad dY_t = \bar{b}(Y_t) dt + \bar{\sigma}(Y_t) d\bar{W}_t$$

on the whole of \mathbb{R} (see Theorems 24.2 and 39.2 in [5]), where \bar{W}_t is another Brownian motion. A version of the process $(X_t : t \geq 0)$ can then be obtained from $(Y_t : t \geq 0)$ by a simple projection described in the proof of the following proposition.

By standard results for one-dimensional diffusions (e.g., [4], Chapter 4), the invariant density of the Markov process $(X_t : t \geq 0)$ is given by

$$(10) \quad \mu(x) = \mu_{\sigma b}(x) = \frac{1}{G\sigma^2(x)} \exp\left(\int_0^x \frac{2b(y)}{\sigma^2(y)} dy\right), \quad x \in [0, 1],$$

with normalising constant

$$(11) \quad G := G_{\sigma b} = \int_0^1 \frac{1}{\sigma^2(y)} \exp\left(\int_0^y \frac{2b(z)}{\sigma^2(z)} dz\right) dy.$$

We see that whenever b and σ are bounded and σ is bounded away from zero, the invariant density is bounded and bounded away from zero. Under the stronger assumption $(\sigma, b) \in \Theta$, we can obtain a similar result also for the transition densities $p_{\sigma b}(\Delta, x, y)$ of the corresponding Markov process $(X_t : t \geq 0)$. The proof is given in the supplement [18].

PROPOSITION 9. *Let $(\sigma, b) \in \Theta$. Then there are constants $0 < K' < K < \infty$ depending only on D, d such that $K' \leq p_{\sigma b}(\Delta, x, y) \leq K$ for all $x, y \in [0, 1]$.*

3.2. A Bernstein type inequality.

THEOREM 10. *Let the time difference between observations $\Delta > 0$ and constants $D, d > 0$ in the definition of Θ be given. Then there exists $\kappa > 0$ depending only on $\Delta, D, d > 0$ such that for all reflected diffusions (8) with $(\sigma, b) \in \Theta$ and arbitrary initial distribution, for all bounded functions $f : [0, 1] \rightarrow \mathbb{R}$, all $r > 0$, all $n \in \mathbb{N}$, and $Z = \sum_{j=0}^{n-1} (f(X_{j\Delta}) - \mathbb{E}_\mu[f(X_0)])$,*

$$\mathbb{P}(|Z| > r) \leq \kappa \exp\left(-\frac{1}{\kappa} \min\left(\frac{r^2}{n\|f\|_{L^2(\mu)}^2}, \frac{r}{\log(n)\|f\|_\infty}\right)\right).$$

PROOF. We make use of the concentration inequality given in Theorem 6 in [1] with $m = 1$ and verify the assumptions by using results in [6]. Let X_0, X_1, \dots be a Markov chain with values in (S, \mathcal{B}) . For $x \in S$ and $A \in \mathcal{B}$, we introduce the transition kernels $P(x, A) = \mathbb{P}(X_1 \in A | X_0 = x)$ and $P^n(x, A) = \mathbb{P}(X_n \in A | X_0 = x)$. For a measurable function $V : S \rightarrow \mathbb{R}$, we define $PV(x) = \mathbb{E}[V(X_1) | X_0 = x]$. The following three assumptions are assumed in [6], where we slightly strengthen the minorization condition to be compatible with the assumption in [1].

(A1) *Minorization condition.* There exists $C \in \mathcal{B}$, $\tilde{\beta} > 0$ and a probability measure ν on (S, \mathcal{B}) such that for all $x \in C$ and $A \in \mathcal{B}$

$$P(x, A) \geq \tilde{\beta}\nu(A),$$

as well as for all $x \in S$ there exists $n \in \mathbb{N}$ such that $P^n(x, C) > 0$.

(A2) *Drift condition.* There exist a measurable function $V : S \rightarrow [1, \infty)$ and constants $\lambda < 1$ and $K < \infty$ satisfying

$$PV(x) \leq \begin{cases} \lambda V(x), & \text{if } x \notin C, \\ K, & \text{if } x \in C. \end{cases}$$

(A3) *Strong aperiodicity condition.* There exists $\beta > 0$ such that $\tilde{\beta}\nu(C) \geq \beta$.

The conditions (A1)–(A3) are verified for the reflected diffusion as follows: Let $C = [0, 1]$, $\tilde{\beta}$ the uniform lower bound on the transition density given by Proposition 9 and ν be the uniform distribution on $[0, 1]$. Then (A1) is satisfied. For (A2), we can take V constant to one and $K = 1$. And (A3) is satisfied with $\beta = \tilde{\beta}$.

By Proposition 4.1(ii) and Proposition 4.4, equation (21) in [6] the constant τ in Theorem 6 in [1] is finite. Under conditions (A1)–(A3), there exists a unique invariant measure μ (Theorem 1 in [6]). Using Corollary 6.1 in [6] and that the Markov chain is reversible, we obtain that for all $f \in L^2(\mu)$

$$(12) \quad \left\| P^n f - \int f d\mu \right\|_{L^2(\mu)} \leq \rho^n \left\| f - \int f d\mu \right\|_{L^2(\mu)}$$

for some $\rho < 1$.

Since we are in the case $m = 1$, the quantity $(\mathbb{E}[T_2])^{-1} \text{Var} Z_1$ in [1] is equal to the asymptotic variance, see the third remark after Theorem 6 there. We bound the asymptotic variance, using (12) and the Cauchy–Schwarz inequality (see also (3) in [18]), by

$$\lim_{n \rightarrow \infty} n^{-1} \text{Var}_\mu \left(\sum_{j=0}^{n-1} f(X_{j\Delta}) \right) \leq \frac{1 + \rho}{1 - \rho} \text{Var}(f(X_0)) \leq \frac{1 + \rho}{1 - \rho} \|f\|_{L^2(\mu)}^2.$$

This establishes the concentration inequality for a fixed pair $(\sigma, b) \in \Theta$. The constants τ and ρ can be chosen uniformly for the class Θ since there is a common lower bound $\tilde{\beta}$ on the transition densities. \square

Note that the above proof can be generalised to diffusions on \mathbb{R} , arguing along the lines of [24].

The following generalisation of the previous theorem to bivariate Markov chains $(X_{\Delta j}, X_{\Delta(j+1)}) : j \in \mathbb{N}$ with invariant measure $\mu_2(x, y) = p(\Delta, x, y)\mu(x)$ is obtained in a similar way; see the supplement [18] for a proof.

THEOREM 11. *Let $\Delta, D, d > 0$ be given. Then there exists $\kappa > 0$ depending only on $\Delta, D, d > 0$ such that for all reflected diffusions (8) with $(\sigma, b) \in \Theta$ and arbitrary initial distributions, for all bounded functions $f : [0, 1]^2 \rightarrow \mathbb{R}$, all $r > 0, n \in \mathbb{N}$, and $Z = \sum_{j=0}^{n-1} (f(X_{j\Delta}, X_{(j+1)\Delta}) - \mathbb{E}_{\mu_2}[f(X_0, X_\Delta)])$,*

$$\mathbb{P}(|Z| > r) \leq \kappa \exp\left(-\frac{1}{\kappa} \min\left(\frac{r^2}{n\|f\|_{L^2(\mu_2)}^2}, \frac{r}{\log(n)\|f\|_\infty}\right)\right).$$

3.3. Concentration inequality for suprema of empirical processes. Let \mathcal{F} be a class of functions. For $f \in \mathcal{F}$ let either $Z(f) = \sum_{j=0}^{n-1} (f(X_{j\Delta}) - \mathbb{E}_\mu[f(X_0)])$ or $Z(f) = \sum_{j=0}^{n-1} (f(X_{j\Delta}, X_{(j+1)\Delta}) - \mathbb{E}_{\mu_2}[f(X_0, X_\Delta)])$. By a change of variables, we can rewrite the previous concentration inequalities as

$$(13) \quad \mathbb{P}(|Z(f)| > \max(\sqrt{v^2 x}, ux)) \leq \kappa e^{-x},$$

where $u = \kappa \log(n)\|f\|_\infty$ and $v^2 = \kappa n\|f\|_{L^2(\mu)}^2$ or $v^2 = \kappa n\|f\|_{L^2(\mu_2)}^2$. Let I be a subset of a linear space of finite dimension d , and consider a class of bounded

measurable functions $\mathcal{F} = \{f_i : i \in I\}$ indexed by I , and such that $0 \in \mathcal{F}$. We define $V^2 = \sup_{f \in \mathcal{F}} v^2$, $U = \kappa \log n \sup_{f \in \mathcal{F}} \|f\|_\infty$ to obtain the following functional concentration inequality.

THEOREM 12. *For $\tilde{\kappa} = 18$ and for all $x \geq 0$, we have*

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |Z(f)| \geq \tilde{\kappa}(\sqrt{V^2(d+x)} + U(d+x))\right) \leq 2\kappa e^{-x}.$$

Given (13), Theorem 12 follows by the usual chaining argument for empirical processes, given for instance in the form of Theorem 2.1 in Baraud [3]. Baraud’s proof applies directly in our setting, where we notice that his Assumption 2.1 can be replaced by (13), since that assumption is only used to apply Bernstein’s inequality in the form (13).

4. Proofs II: The main contraction Theorem 1. Our strategy to prove the main theorem of this article is as follows: In the spirit of [9, 10], we first derive a general contraction theorem for discretely sampled diffusions that requires the prior to charge small neighbourhoods of the true parameter measured in the information distance (a version of the KL-divergence), and that admits the existence of certain frequentist tests uniformly in the parameter space. We then show how the information distance can be controlled by suitable dual Besov norms, and use the concentration inequalities from the previous subsection to construct suitable tests.

4.1. General contraction theorem with tests. We denote by $\mathbf{K}(\mathbb{P}, \mathbb{Q}) := \mathbb{E}_{\mathbb{P}}[\log \frac{d\mathbb{P}}{d\mathbb{Q}}]$ the Kullback–Leibler divergence between two probability measures \mathbb{P} and \mathbb{Q} defined on the same σ -algebra. We write $\mathbb{P}_{\sigma b}$ and $\mathbb{E}_{\sigma b}$ for the probability and the expectation with respect to the reflected diffusion started in the invariant distribution. We also introduce the notation

$$\mathbf{KL}((\sigma_0, b_0), (\sigma, b)) := \mathbb{E}_{\sigma_0 b_0} \left[\log \left(\frac{p_{\sigma_0 b_0}(\Delta, X_0, X_\Delta)}{p_{\sigma b}(\Delta, X_0, X_\Delta)} \right) \right],$$

and for every $\varepsilon, \kappa > 0$ we define

$$\begin{aligned} B_{\varepsilon, \kappa} &= \left\{ \vartheta = (\sigma, b) \in \Theta : \mathbf{KL}((\sigma_0, b_0), (\sigma, b)) \leq \varepsilon^2, \right. \\ (14) \quad &\text{Var}_{\sigma_0 b_0} \left(\log \frac{p_{\sigma b}(\Delta, X_0, X_\Delta)}{p_{\sigma_0 b_0}(\Delta, X_0, X_\Delta)} \right) \leq 2\varepsilon^2, \\ &\left. \mathbf{K}(\mu_{\sigma_0 b_0}, \mu_{\sigma b}) \leq \kappa, \text{Var}_{\sigma_0 b_0} \left(\log \frac{\mu_{\sigma b}(X_0)}{\mu_{\sigma_0 b_0}(X_0)} \right) \leq 2\kappa \right\}. \end{aligned}$$

THEOREM 13. *Let $\Pi = \Pi_n$ be a sequence of prior distributions on a σ -field \mathcal{S} of subsets of Θ and suppose that $X_0, \dots, X_{n\Delta}$ are discrete observations*

of a reflected diffusion process (8), started in the stationary distribution μ . Let $\Pi(\cdot|X_0, \dots, X_{n\Delta})$ be the resulting posterior distribution (2). For $\vartheta_0 = (\sigma_0, b_0) \in \Theta$, ε_n , a sequence of positive real numbers such that $\varepsilon_n \rightarrow 0$, $\sqrt{n}\varepsilon_n \rightarrow \infty$, and C, κ fixed positive constants, suppose Π satisfies for all n large enough

$$(15) \quad \Pi(B_{\varepsilon_n, \kappa}) \geq e^{-Cn\varepsilon_n^2}.$$

Assume moreover that there exists $0 < \bar{L} < \infty$ such that $\Pi(\Theta \setminus \mathcal{B}_n) \leq \bar{L}e^{-(C+4)n\varepsilon_n^2}$ for some sequence $\mathcal{B}_n \subseteq \Theta$ for which we can find a sequence of tests (indicator functions) $\Psi_n \equiv \Psi(X_0, \dots, X_{n\Delta})$ and of distance functions d_n such that for every $n \in \mathbb{N}$, $M > 0$ large enough,

$$\mathbb{E}_{\sigma_0 b_0}[\Psi_n] \xrightarrow{n \rightarrow \infty} 0, \quad \sup_{(\sigma, b) \in \mathcal{B}_n : d_n((\sigma, b), (\sigma_0, b_0)) \geq M\varepsilon_n} \mathbb{E}_{\sigma b}[1 - \Psi_n] \leq \bar{L}e^{-(C+4)n\varepsilon_n^2}.$$

Then the posterior distribution $\Pi(\cdot|X_0, \dots, X_{n\Delta})$ contracts about (σ_0, b_0) at rate ε_n in the distance d_n , that is, in $\mathbb{P}_{\sigma_0 b_0}$ -probability, as $n \rightarrow \infty$,

$$\Pi((\sigma, b) : d_n((\sigma, b), (\sigma_0, b_0)) > M\varepsilon_n | X_0, \dots, X_{n\Delta}) \rightarrow 0.$$

The proof of this theorem follows the standard pattern from [9, 10] (see also Section 7.3.1 in [12]), and is given in the supplement [18].

4.2. *Small ball lemma.* Recall the sets $B_{\varepsilon, \kappa}$ defined in (14).

LEMMA 14. *There exists a constant $\bar{C} > 0$ such that for every $\kappa > 0$ and for all $\varepsilon > 0$ small enough*

$$\left\{ (\sigma, b) \in \Theta : \|\mu - \mu_0\|_{L^2([0,1])} + \|\sigma^{-2} - \sigma_0^{-2}\|_{(B^1_{1\infty})^*} + \|b - b_0\|_{(B^2_{1\infty})^*} < \frac{\varepsilon}{\bar{C}} \right\} \subseteq B_{\varepsilon, \kappa}.$$

A crucial step in the proof of this key lemma is the observation, partly borrowed from [13], that the $L^2([0, 1]^2)$ -distance between the transitions densities $p_{\sigma b}, p_{\sigma_0 b_0}$ is related to a suitable Hilbert–Schmidt (HS) norm of the difference between the corresponding transition operators. Using the semigroup representation $P_\Delta = e^{\Delta L}$ of the transition operators P_Δ , we can then approximate the information distance on the underlying experiment by the HS-distance between the corresponding inverse operators of the infinitesimal generators L of the underlying diffusions. In turn, we can obtain analytic expressions for the Green’s function of the inverses of these generators, which ultimately gives the reduction to the dual Besov norms appearing above. We split the proof into several steps, given in the following subsections.

4.2.1. *The infinitesimal generator L and its inverse.* We begin by defining the function $S(\cdot) = 1/s'(\cdot)$, derived from the scale function $s(\cdot)$,

$$S(x) := \frac{1}{2} \sigma^2(x) \mu(x) = \frac{1}{2G} \exp\left(\int_0^x \frac{2b(y)}{\sigma^2(y)} dy\right),$$

with G the normalising constant of the invariant density as in (11). The infinitesimal generator $L = L_{\sigma b}$ of the diffusion (8) is given by the action

$$(16) \quad Lf(x) = \frac{1}{2} \sigma^2(x) f''(x) + b(x) f'(x) = \frac{1}{\mu(x)} (S(x) f'(x))',$$

where the domain of this unbounded operator on $L^2(\mu)$ is the subspace of the L^2 -Sobolev space H^2 with Neumann boundary conditions

$$\text{dom}(L) = \{f \in H^2([0, 1]) : f'(0) = f'(1) = 0\}.$$

We fix the invariant measure μ_0 belonging to σ_0 and b_0 and consider $L = L_{\sigma b}$ on $L^2(\mu_0)$, which by the bound from above and away from zero of the invariant densities is the same set of functions as $L^2(\mu)$. We introduce

$$V := \left\{ f \in L^2(\mu_0) : \int_0^1 f d\mu_0 = 0 \right\} \quad \text{and} \quad V^\perp := \{f \in L^2(\mu_0) : f \text{ constant}\}.$$

We denote by $\mu_0(L)$ the operator that sends f to the constant function $\mu_0(Lf) = \int Lf(x) \mu_0(x) dx$. We observe that the operator $L - \mu_0(L)$ leaves the space V invariant, and denote by $(L - \mu_0(L))|_V$ its restriction to V . Next, we introduce an integral operator J and show that J is an explicit representation of the inverse of $(L - \mu_0(L))|_V$. We define

$$Jf(x) = \int_0^1 K(x, z) f(z) \mu_0(z) dz, \quad f \in V,$$

with kernel $K = K_{\sigma b}$ defined as

$$K(x, z) = 2G \left(H(x, z) - \frac{\mu(z)}{\mu_0(z)} \int_0^1 H(x, y) \mu_0(y) dy \right),$$

where

$$\begin{aligned} H(x, z) &= H_{\sigma b}(x, z) \\ &= \int_0^1 \left(\mathbb{1}_{[z, x]}(y) - \mathbb{1}_{[z, 1]}(y) \int_y^1 \mu_0(x) dx \right) \exp\left(-\int_0^y \frac{2b(v)}{\sigma^2(v)} dv\right) dy \frac{\mu(z)}{\mu_0(z)}. \end{aligned}$$

Here, $\mathbb{1}_{[z, x]} = 0$ if $x < z$. Writing $\mathbb{1}_{[z, x]}(y) = \mathbb{1}_{[0, x]}(y) \mathbb{1}_{[z, 1]}(y)$ and using $\mathbb{1}_{[z, 1]}(y) = \mathbb{1}_{[0, y]}(z)$ as well as Fubini's theorem, an alternative representation of J

is given by

$$Jf(x) = 2G \int_0^1 \int_0^y \left(f(z) - \int_0^1 f d\mu \right) \mu(z) dz \left(\mathbb{1}_{[0,x]}(y) - \int_y^1 d\mu_0 \right) \times \exp \left(- \int_0^y \frac{2b(v)}{\sigma^2(v)} dv \right) dy.$$

We compute the first two derivatives:

$$\begin{aligned} \frac{d}{dx}(Jf)(x) &= 2G \int_0^x \left(f(z) - \int f d\mu \right) \mu(z) dz \exp \left(- \int_0^x \frac{2b(v)}{\sigma^2(v)} dv \right), \\ (17) \quad \frac{d^2}{dx^2}(Jf)(x) &= 2G \left(f(x) - \int f d\mu \right) \mu(x) \exp \left(- \int_0^x \frac{2b(v)}{\sigma^2(v)} dv \right) \\ &\quad + 2G \int_0^x \left(f(z) - \int f d\mu \right) \mu(z) dz \left(- \frac{2b(x)}{\sigma^2(x)} \right) \\ &\quad \times \exp \left(- \int_0^x \frac{2b(v)}{\sigma^2(v)} dv \right). \end{aligned}$$

It follows

$$\begin{aligned} LJf(x) &= \frac{1}{2} \sigma^2(x) \frac{d^2}{dx^2}(Jf)(x) + b(x) \frac{d}{dx}(Jf)(x) \\ &= \frac{1}{2} \sigma^2(x) 2G \left(f(x) - \int f d\mu \right) \mu(x) \exp \left(- \int_0^x \frac{2b(v)}{\sigma^2(v)} dv \right) \\ &= f(x) - \int f d\mu \end{aligned}$$

and thus $(L - \mu_0(L))Jf(x) = f(x) - \int f d\mu_0$. Consequently, $(L - \mu_0(L))J$ is the identity operator on V . We see from the first derivative in (17) that $Jf \in \text{dom}(L)$. Using $\mathbb{1}_{[0,x]}(y) = \mathbb{1}_{[y,1]}(x)$ and Fubini's theorem, one also sees that $\int_0^1 Jf(x)\mu_0(x) dx = 0$, and consequently $Jf \in \text{dom}(L) \cap V$.

To see that $(L - \mu_0(L))|_V$ is injective, suppose that for $f, f_1 \in \text{dom}(L) \cap V$ we have $(L - \mu_0(L))f = (L - \mu_0(L))f_1$ or equivalently $Lf = Lf_1 + c_0$. This implies by integration with respect to $d\mu(x)$ that $S(x)f'(x) = S(x)f_1'(x) + c_0 \int_0^x d\mu + c_1$ with $c_1 = c_0 = 0$ since $f'(0) = f_1'(0) = f'(1) = f_1'(1) = 0$. Another integration gives $f(x) = f_1(x) + c_2$ and $c_2 = 0$ by $f, f_1 \in V$. We conclude that $f = f_1$ showing that $(L - \mu_0(L))|_V$ is injective, and by what precedes the inverse mapping $(L - \mu_0(L))|_V^{-1} : V \rightarrow \text{dom}(L) \cap V$ exists, and has integral representation J . Note that when $L = L_{\sigma_0 b_0}$ then in view of (16) we have $\mu_0(L_{\sigma_0 b_0})(f) = 0$ for all $f \in \text{dom}(L)$, and hence the same integral representation follows for $(L_{\sigma_0 b_0})|_V^{-1} = (L_{\sigma_0 b_0} - \mu_0(L_{\sigma_0 b_0}))|_V^{-1}$.

The following lemma bounds the HS-norm distance between the Green kernels and is proved in the supplement [18].

LEMMA 15. *There exists $\tilde{C} > 0$ such that for all $(\sigma, b), (\sigma_0, b_0) \in \Theta$*

$$\begin{aligned} & \left(\int_0^1 \int_0^1 (K_{\sigma b} - K_{\sigma_0 b_0})^2(x, z) \mu_0(x) \mu_0(z) \, dx \, dz \right)^{1/2} \\ & \leq \tilde{C} \|\mu_{\sigma b} - \mu_0\|_{L^2([0,1])} + \tilde{C} \left\| \frac{1}{\sigma^2} - \frac{1}{\sigma_0^2} \right\|_{(B_{1^\infty}^1)^*} + \tilde{C} \|b - b_0\|_{(B_{1^\infty}^2)^*}. \end{aligned}$$

4.2.2. *Proof of Lemma 14.* We have

$$\text{KL}((\sigma_0, b_0), (\sigma, b)) = \mathbb{E}_{\sigma_0 b_0} \left[\log \left(\frac{\mu_0(X_0) p_{\sigma_0 b_0}(\Delta, X_0, X_\Delta)}{\mu_0(X_0) p_{\sigma b}(\Delta, X_0, X_\Delta)} \right) \right].$$

We see that this is the Kullback–Leibler divergence between the probability measures corresponding to the densities $\mu_0 p_{\sigma_0 b_0} = \mu_0(x) p_{\sigma_0 b_0}(\Delta, x, y)$ and $\mu_0 p_{\sigma b} = \mu_0(x) p_{\sigma b}(\Delta, x, y)$ with respect to the Lebesgue measure on $[0, 1]^2$. By Lemma 8.2 in [9], we have

$$\text{KL}((\sigma_0, b_0), (\sigma, b)) \leq 2h^2(\mu_0 p_{\sigma b}, \mu_0 p_{\sigma_0 b_0}) \left\| \frac{p_{\sigma_0 b_0}}{p_{\sigma b}} \right\|_\infty,$$

where $h^2(p, q) = \int (\sqrt{p} - \sqrt{q})^2$ is the usual Hellinger distance between two densities p, q . The transition densities are bounded from above and from below in view of Proposition 9. Thus, the Hellinger distance can be bounded by the L^2 -norm of the difference between the densities

$$h^2(\mu_0 p_{\sigma b}, \mu_0 p_{\sigma_0 b_0}) \leq \left\| \frac{1}{\mu_0 p_{\sigma_0 b_0}} \right\|_\infty \|\mu_0 p_{\sigma b} - \mu_0 p_{\sigma_0 b_0}\|_{L^2([0,1]^2)}^2.$$

We want to bound the last quantity in terms of the Hilbert–Schmidt norm distance $\|P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0}\|_{\text{HS}}$ between the transition operators of the respective diffusions acting on $L^2(\mu_0)$. We have the integral representation

$$\begin{aligned} (P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0})f &= \int (p_{\sigma b}(x, y) - p_{\sigma_0 b_0}(x, y)) f(y) \, dy \\ &= \int \frac{p_{\sigma b}(x, y) - p_{\sigma_0 b_0}(x, y)}{\mu_0(y)} f(y) \mu_0(y) \, dy \end{aligned}$$

and thus the Hilbert–Schmidt norm is given by

$$\begin{aligned} \|P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0}\|_{\text{HS}}^2 &= \int \int \left(\frac{p_{\sigma b}(x, y) - p_{\sigma_0 b_0}(x, y)}{\mu_0(y)} \right)^2 \mu_0(x) \mu_0(y) \, dx \, dy \\ &= \int \int (p_{\sigma b}(x, y) - p_{\sigma_0 b_0}(x, y))^2 \frac{\mu_0(x)}{\mu_0(y)} \, dx \, dy. \end{aligned}$$

In summary, we can bound $\text{KL}((\sigma_0, b_0), (\sigma, b))$ by a constant multiple of

$$\|\mu_0 p_{\sigma b} - \mu_0 p_{\sigma_0 b_0}\|_{L^2([0,1]^2)}^2 \leq \|\mu_0\|_\infty^2 \|P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0}\|_{\text{HS}}^2.$$

Let $(e_k)_{k \in \mathbb{N}}$ be any orthonormal basis of $L^2(\mu_0)$ such that $e_0 = 1$ (for instance, we can take the eigen-basis of the operator $P_\Delta^{\sigma_0 b_0}$). Then

$$\|P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0}\|_{\text{HS}}^2 = \sum_{k=0}^\infty \|(P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0})e_k\|_{L^2(\mu_0)}^2 = \sum_{k=1}^\infty \|(P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0})e_k\|_{L^2(\mu_0)}^2.$$

We denote by $\mu_0(P_\Delta^{\sigma b})$ the operator that sends f to the constant function $\mu_0(P_\Delta^{\sigma b} f) = \int P_\Delta^{\sigma b} f(x) \mu_0(x) dx$ and write the Hilbert–Schmidt norm as

$$\begin{aligned} \|P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0}\|_{\text{HS}}^2 &= \sum_{k=1}^\infty \|(P_\Delta^{\sigma b} - \mu_0(P_\Delta^{\sigma b}) - P_\Delta^{\sigma_0 b_0} + \mu_0(P_\Delta^{\sigma b}))e_k\|_{L^2(\mu_0)}^2 \\ &= \sum_{k=1}^\infty \|(P_\Delta^{\sigma b} - \mu_0(P_\Delta^{\sigma b}) - P_\Delta^{\sigma_0 b_0})e_k\|_{L^2(\mu_0)}^2 + \sum_{k=1}^\infty \mu_0(P_\Delta^{\sigma b} e_k)^2, \end{aligned}$$

where we have used that all e_k with $k \geq 1$ are orthogonal to e_0 , and thus to all constant functions, and that $P_\Delta^{\sigma b} - \mu_0(P_\Delta^{\sigma b}), P_\Delta^{\sigma_0 b_0}$ leave the space $V = \{f \in L^2(\mu_0) : \int f d\mu_0 = 0\}$ invariant (note that μ_0 is the invariant measure for $P_\Delta^{\sigma_0 b_0}$). By the last observation, we can write

$$\begin{aligned} (18) \quad &\|P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0}\|_{\text{HS}}^2 \\ &= \|(P_\Delta^{\sigma b} - \mu_0(P_\Delta^{\sigma b}))|_V - P_\Delta^{\sigma_0 b_0}|_V\|_{\text{HS}}^2 + \sum_{k=1}^\infty \mu_0(P_\Delta^{\sigma b} e_k)^2. \end{aligned}$$

We represent $P_\Delta^{\sigma_0 b_0}|_V = \exp(\Delta L_{\sigma_0 b_0}|_V)$ and

$$(P_\Delta^{\sigma b} - \mu_0(P_\Delta^{\sigma b}))|_V = \exp(\Delta(L_{\sigma b} - \mu_0(L_{\sigma b}))|_V),$$

possible since, by standard properties of Markov semigroups, the derivatives

$$\frac{d}{d\Delta}(P_\Delta^{\sigma b} - \mu_0(P_\Delta^{\sigma b}))|_V(f) = \frac{d}{d\Delta} \exp(\Delta(L_{\sigma b} - \mu_0(L_{\sigma b}))|_V)(f)$$

coincide for all $f \in V \cap \text{dom}(L)$, and setting $\Delta = 0$ gives the identity $\text{id}|_V$ on both sides in the last but one display.

For every (σ, b) , the operators $L_{\sigma b}$ have a discrete nonpositive spectrum $\{\lambda_k : k \in \mathbb{N}\}$ (e.g., pages 97–100 in [2]), with eigenfunctions u_k orthonormal in $L^2(\mu)$. One checks directly that the spectra of $L_{\sigma_0 b_0}|_V, (L_{\sigma b} - \mu_0(L_{\sigma b}))|_V$ equal the ones of $L_{\sigma_0 b_0}, L_{\sigma b}$ but with eigenvalue $\{0\}$ removed (corresponding to $u_0 = 1$), and, in the second case, for different eigenfunctions $u_k - \mu_0(u_k)$, orthonormal for the scalar product $\langle \cdot, \cdot \rangle_{V, \mu} \equiv \langle \cdot - \int(\cdot) d\mu, \cdot - \int(\cdot) d\mu \rangle_\mu$ on V . Their inverses $L_{\sigma_0 b_0}|_V^{-1}, (L_{\sigma b} - \mu_0(L_{\sigma b}))|_V^{-1}$ derived in Section 4.2.1 are bounded linear operators on V (in view of their integral representations), and in view of their spectral representation they are also symmetric, and thus self-adjoint for the scalar products $\langle \cdot, \cdot \rangle_{\mu_0}$ and $\langle \cdot, \cdot \rangle_{V, \mu}$ on V , respectively.

Using the functional calculus and what precedes, we can hence represent $(P_{\Delta}^{\sigma b} - \mu_0(P_{\Delta}^{\sigma b}))|_V$ and $(P_{\Delta}^{\sigma_0 b_0})|_V$ as the composition of the inverses $(L_{\sigma b} - \mu_0(L_{\sigma b}))|_V^{-1}$, $L_{\sigma_0 b_0}|_V^{-1}$ and the function $f : z \rightarrow \exp(\Delta z^{-1})$, which is Lipschitz continuous on $(-\infty, 0)$, with Lipschitz constant Λ . Proposition 3 in [18] states that if f is Lipschitz continuous with Lipschitz constant Λ on the union of the spectra of two operators T_1 and T_2 , which are self-adjoint with respect to different scalar products $\langle \cdot, \cdot \rangle_{V, \mu}$ and $\langle \cdot, \cdot \rangle_{\mu_0}$ then we have

$$\|f(T_1) - f(T_2)\|_{\text{HS}} \leq \Lambda \|T_1 - T_2\|_{\text{HS}},$$

where the HS-norm is for operators from $(V, \langle \cdot, \cdot \rangle_{\mu_0})$ to $(V, \langle \cdot, \cdot \rangle_{V, \mu})$ (and where, strictly speaking, the operators in the preceding display are composed with suitable identity operators between these spaces). Since $\langle \cdot, \cdot \rangle_{V, \mu}$ and $\langle \cdot, \cdot \rangle_{\mu_0}$ induce equivalent norms on V , up to constants the same holds when the Hilbert–Schmidt norm is interpreted for operators from $(V, \langle \cdot, \cdot \rangle_{\mu_0})$ to $(V, \langle \cdot, \cdot \rangle_{\mu_0})$. We thus obtain from the results in Section 4.2.1 that

$$\begin{aligned} & \| (P_{\Delta}^{\sigma b} - \mu_0(P_{\Delta}^{\sigma b}))|_V - P_{\Delta}^{\sigma_0 b_0}|_V \|_{\text{HS}}^2 \\ & \lesssim \| (L_{\sigma b} - \mu_0(L_{\sigma b}))|_V^{-1} - L_{\sigma_0 b_0}|_V^{-1} \|_{\text{HS}}^2 \\ & = \int_0^1 \int_0^1 (K_{\sigma b} - K_{\sigma_0 b_0})^2(x, z) \mu_0(x) \mu_0(z) dx dz \equiv A. \end{aligned}$$

Before we bound A further, let us next consider the second term in (18). Using $\int_0^1 \mu_{\sigma b}(x) p_{\sigma b}(\Delta, x, y) dx = \mu_{\sigma b}(y)$ and Parseval’s identity, we obtain

$$\begin{aligned} & \sum_{k=1}^{\infty} \mu_0(P_{\Delta}^{\sigma b} e_k)^2 \\ & = \sum_{k=1}^{\infty} [\mu_0(P_{\Delta}^{\sigma b} e_k - e_k)]^2 \\ & = \sum_{k=1}^{\infty} \left(\int_0^1 \left[\int_0^1 \mu_0(x) p_{\sigma b}(\Delta, x, y) e_k(y) dy - \mu_0(x) e_k(x) \right] dx \right)^2 \\ & \leq 2 \sum_{k=1}^{\infty} \left(\int_0^1 \int_0^1 (\mu_0(x) - \mu_{\sigma b}(x)) p_{\sigma b}(\Delta, x, y) e_k(y) dy dx \right)^2 \\ & \quad + 2 \sum_{k=1}^{\infty} \left(\int_0^1 (\mu_{\sigma b}(x) - \mu_0(x)) e_k(x) dx \right)^2 \\ & = 2 \sum_{k=1}^{\infty} \left\langle \frac{\int_0^1 (\mu_{\sigma b} - \mu_0)(x) p_{\sigma b}(\Delta, x, \cdot) dx}{\mu_0}, e_k \right\rangle_{\mu_0}^2 + 2 \sum_{k=1}^{\infty} \left\langle \frac{\mu_{\sigma b} - \mu_0}{\mu_0}, e_k \right\rangle_{\mu_0}^2 \end{aligned}$$

$$= 2 \int_0^1 \left(\frac{\int_0^1 (\mu_{\sigma b} - \mu_0)(x) p_{\sigma b}(\Delta, x, y) dx}{\mu_0(y)} \right)^2 \mu_0(y) dy + 2 \int_0^1 \left(\frac{\mu_{\sigma b}(y) - \mu_0(y)}{\mu_0(y)} \right)^2 \mu_0(y) dy,$$

which can be bounded by $B \approx \|\mu_{\sigma b} - \mu_0\|_{L^2([0,1])}^2$, using that the transition density is bounded from above and μ_0 away from zero. Using Lemma 15 for term A, we obtain, for some constant $C > 0$ and \bar{C} large enough, the overall bound

$$KL((\sigma_0, b_0), (\sigma, b)) \leq C(A + B) \leq \varepsilon^2.$$

In order to see

$$(19) \quad \text{Var}_{\sigma_0 b_0} \left(\log \frac{p_{\sigma b}(\Delta, X_0, X_\Delta)}{p_{\sigma_0 b_0}(\Delta, X_0, X_\Delta)} \right) \leq 2\varepsilon^2$$

we bound the variance by the second moment and use Lemma 8.3 in [9], which implies that

$$\mathbb{E}_{\sigma_0 b_0} \left[\left| \log \frac{\mu_0 p_{\sigma b}}{\mu_0 p_{\sigma_0 b_0}} \right|^2 \right] \leq 4h^2(\mu_0 p_{\sigma b}, \mu_0 p_{\sigma_0 b_0}) \left\| \frac{p_{\sigma_0 b_0}}{p_{\sigma b}} \right\|_\infty.$$

Proceeding as for the Kullback–Leibler divergence above shows (19).

It remains to show that

$$(20) \quad K(\mu_{\sigma_0 b_0}, \mu_{\sigma b}) \leq \kappa \quad \text{and} \quad \text{Var}_{\sigma_0 b_0} \left(\log \frac{\mu_{\sigma b}(X_0)}{\mu_{\sigma_0 b_0}(X_0)} \right) \leq 2\kappa.$$

By Lemmas 8.2 and 8.3 in [9], it suffices to bound $h^2(\mu_{\sigma b}, \mu_{\sigma_0 b_0}) \|\mu_{\sigma_0 b_0} / \mu_{\sigma b}\|_\infty$. Using that $\mu_{\sigma b}$ and $\mu_{\sigma_0 b_0}$ are bounded from above and from below and bounding the Hellinger distance by the L^2 -norm, the Hellinger distance is bounded by a multiple of $\|\mu_{\sigma b} - \mu_{\sigma_0 b_0}\|_{L^2([0,1])}^2$. So for $\varepsilon > 0$ small enough we have (20).

4.3. *Construction of tests.* We will now construct the tests needed in Theorem 13. The tests are based on the spectral estimators constructed in [13], which are defined by

$$(21) \quad \hat{\sigma}^2(x) := \frac{2\Delta^{-1} \log(\hat{\kappa}_1) \int_0^x \hat{u}_1(y) \hat{\mu}(y) dy}{\hat{u}'_1(x) \hat{\mu}(x)},$$

$$(22) \quad \hat{b}(x) := \Delta^{-1} \log(\hat{\kappa}_1) \frac{\hat{u}_1(x) \hat{u}'_1(x) \hat{\mu}(x) - \hat{u}_1'' \int_0^x \hat{u}_1(y) \hat{\mu}(y) dy}{\hat{u}'_1(x)^2 \hat{\mu}(x)},$$

where $\hat{\kappa}_1, \hat{u}_1$ are estimates of the second largest eigenvalue and associated eigenfunction of the operator $P_{\Delta}^{\sigma b}$, and where $\hat{\mu}$ is defined in (30) below. Using the concentration inequality Theorem 12, we can prove the following for these estimators.

THEOREM 16. *Let $s > 1$ and let ε_n be such that $n^{-(s+1)/(2s+3)} \lesssim \varepsilon_n \lesssim n^{-3/8}(\log n)^{-1/2}$. For all $D > 0$, there exists $R > 0$ such that for n large enough we have uniformly in $\vartheta = (\sigma, b) \in \Theta_s$*

$$\mathbb{P}_{\vartheta}(\|\widehat{\sigma}^2 - \sigma^2\|_{L^2([A,B])} \geq Rn\varepsilon_n^3 \text{ or } \|\widehat{b} - b\|_{L^2([A,B])} \geq Rn^2\varepsilon_n^5) \leq e^{-Dn\varepsilon_n^2}.$$

We postpone the proof of Theorem 16 to the end of the subsection, but record how it can be used to construct tests for the separation metric

$$(23) \quad d_n(\vartheta, \vartheta_0) = n^{-1}\varepsilon_n^{-2}\|\sigma^2 - \sigma_0^2\|_{L^2([A,B])} + n^{-2}\varepsilon_n^{-4}\|b - b_0\|_{L^2([A,B])}.$$

Given Theorem 16 the proof of the following result is elementary (see the supplement [18]).

THEOREM 17. *For $\vartheta_0 \in \Theta_s$, there exists a sequence of tests (indicator functions) $\Psi_n \equiv \Psi(X_0, \dots, X_{n\Delta})$ such that for every $n \in \mathbb{N}$, $C > 0$, there exists $M = M(C) > 0$ large enough such that*

$$\mathbb{E}_{\vartheta_0}[\Psi_n] \xrightarrow{n \rightarrow \infty} 0, \quad \sup_{\vartheta \in \Theta_s: d_n(\vartheta, \vartheta_0) \geq M\varepsilon_n} \mathbb{E}_{\vartheta}[1 - \Psi_n] \leq e^{-(C+4)n\varepsilon_n^2}.$$

In the remaining part of this subsection, we derive concentration inequalities for the successive steps in the estimation procedure and at the end of the section we prove Theorem 16. We denote by ψ_λ with $\lambda = (l, k)$, $|\lambda| = l$, a compactly supported L^2 -orthonormal wavelet basis of $L^2([0, 1])$ as after (5). Let V_J be the L^2 -closed linear spaces spanned by the wavelets up to level $|\lambda| \leq J$. We define π_J to be the L^2 -orthogonal projection onto V_J and π_J^μ to be the $L^2(\mu)$ -orthogonal projection onto V_J . We construct estimators as in [13]. We estimate the action of the transition operator on the wavelet spaces $(\mathbf{P}_\Delta^J)_{\lambda, \lambda'} := \langle P_\Delta^{\sigma b} \psi_\lambda, \psi_{\lambda'} \rangle_\mu$ by

$$(\widehat{\mathbf{P}}_\Delta)_{\lambda, \lambda'} := \frac{1}{2n} \sum_{l=1}^n (\psi_\lambda(X_{(l-1)\Delta})\psi_{\lambda'}(X_{l\Delta}) + \psi_{\lambda'}(X_{(l-1)\Delta})\psi_\lambda(X_{l\Delta}))$$

and the $\dim(V_J) \times \dim(V_J)$ -dimensional Gram matrix \mathbf{G} with entries $\mathbf{G}_{\lambda, \lambda'} = \langle \psi_\lambda, \psi_{\lambda'} \rangle_\mu$ by

$$\widehat{\mathbf{G}}_{\lambda, \lambda'} := \frac{1}{n} \left(\frac{\psi_\lambda(X_0)\psi_{\lambda'}(X_0)}{2} + \frac{\psi_\lambda(X_{n\Delta})\psi_{\lambda'}(X_{n\Delta})}{2} + \sum_{l=1}^{n-1} \psi_\lambda(X_{l\Delta})\psi_{\lambda'}(X_{l\Delta}) \right).$$

Let u_1 be the eigenfunction of $P_\Delta^{\sigma b}$ corresponding to the second largest eigenvalue κ_1 . Let u_1^J be the eigenfunction belonging to the second largest eigenvalue κ_1^J of the operator $\pi_J^\mu P_\Delta^{\sigma b}$.

LEMMA 18. $\|u_1^J\|_\infty$ is bounded uniformly in $(\sigma, b) \in \Theta_s$ and $J \in \mathbb{N}$.

PROOF. By Lemma 6.6 in [13], $\|u_1\|_{H^{s+1}}$ is uniformly bounded in Θ_s . By Corollary 4.6 in [13], this implies that $\|u_1^J\|_{H^1}$ is uniformly bounded, and the Sobolev imbedding implies the result. \square

Subsequently, $\|\cdot\|_{\ell^2 \rightarrow \ell^2}$ denotes the usual norm of an operator on ℓ^2 .

LEMMA 19. *Let u_1^J be the vector associated with the normalised eigenfunction u_1^J of $\pi_J^\mu P_\Delta^{\sigma b}$ with eigenvalue κ_1^J . Let $J = J_n \rightarrow \infty$ as $n \rightarrow \infty$ be a sequence of integers and let n be such that $2^J \leq cn^{1/2}/\log n$ for some $c > 0$. For (26), assume that also $2^{-sJ} \leq c\sqrt{2^J/n}$. Then for all $D > 0$ there exists $C > 0, \kappa > 0$ such that uniformly over Θ_s*

$$(24) \quad \mathbb{P}\left(\|(\hat{P}_\Delta - P_\Delta^J)u_1^J\|_{\ell^2} < C\sqrt{\frac{2^J}{n}}\right) \geq 1 - 2\kappa e^{-D2^J},$$

$$(25) \quad \mathbb{P}\left(\|(\hat{G} - G)u_1^J\|_{\ell^2} < C\sqrt{\frac{2^J}{n}}\right) \geq 1 - 2\kappa e^{-D2^J},$$

$$(26) \quad \mathbb{P}\left(\|\hat{\mu} - \mu\|_{L^2} < C\sqrt{\frac{2^J}{n}}\right) \geq 1 - 2\kappa e^{-D2^J},$$

$$(27) \quad \mathbb{P}\left(\|\hat{G} - G\|_{\ell^2 \rightarrow \ell^2} < C\frac{2^{2J}}{\sqrt{n}}\right) \geq 1 - \kappa 2^{J+2} e^{-D2^J},$$

$$(28) \quad \mathbb{P}\left(\|\hat{P}_\Delta - P_\Delta^J\|_{\ell^2 \rightarrow \ell^2} < C\frac{2^{2J}}{\sqrt{n}}\right) \geq 1 - \kappa 2^{J+2} e^{-D2^J}.$$

Moreover, for all $\delta, D > 0$ there exists n_0 such that we have for all $n \geq n_0$,

$$(29) \quad \mathbb{P}(\|\hat{\mu} - \mu\|_{L^\infty([0,1])} < \delta) \geq 1 - 2\kappa e^{-D2^J}.$$

PROOF. We have

$$\begin{aligned} [(\hat{P}_\Delta - P_\Delta^J)u_1^J]_\lambda &= \frac{1}{2n} \sum_{l=1}^n (\psi_\lambda(X_{(l-1)\Delta})u_1^J(X_{l\Delta}) + u_1^J(X_{(l-1)\Delta})\psi_\lambda(X_{l\Delta})) \\ &\quad - \mathbb{E}[\psi_\lambda(X_0)u_1^J(X_\Delta) + u_1^J(X_0)\psi_\lambda(X_\Delta)]. \end{aligned}$$

We express the ℓ^2 -norm $\|(\hat{P}_\Delta - P_\Delta^J)u_1^J\|_{\ell^2}$ by its dual representation

$$\begin{aligned} \sup_{\|v\|_{\ell^2} \leq 1} \left| \sum_{|\lambda| \leq J} v_\lambda \frac{1}{2n} \sum_{l=1}^n (\psi_\lambda(X_{(l-1)\Delta})u_1^J(X_{l\Delta}) + u_1^J(X_{(l-1)\Delta})\psi_\lambda(X_{l\Delta})) \right. \\ \left. - \mathbb{E}[\psi_\lambda(X_0)u_1^J(X_\Delta) + u_1^J(X_0)\psi_\lambda(X_\Delta)] \right| \end{aligned}$$

$$= \sup_{\|v\|_{L^2} \leq 1, v \in V_J} \left| \frac{1}{2n} \sum_{l=1}^n (v(X_{(l-1)\Delta}) u_1^J(X_{l\Delta}) + u_1^J(X_{(l-1)\Delta}) v(X_{l\Delta})) \right. \\ \left. - \mathbb{E}[v(X_0) u_1^J(X_\Delta) + u_1^J(X_0) v(X_\Delta)] \right|.$$

We consider the class $\mathcal{F} = \{f(x, y) = (v(x)u_1^J(y) + u_1^J(x)v(y))/2, \|v\|_{L^2} \leq 1, v \in V_J\}$ with $\dim(V_J) = 2^{J+1}$. In order to determine V^2 and U in Theorem 12, we use Lemma 18 and calculate for $f \in \mathcal{F}$

$$\|f\|_{L^2(\mu_2)}^2 = \frac{1}{4} \|v(x)u_1^J(y) + u_1^J(x)v(y)\|_{L^2(\mu_2)}^2 \\ \leq \|u_1^J\|_\infty^2 \|v\|_{L^2(\mu)}^2 \leq \|u_1^J\|_\infty^2 \|\mu\|_\infty \|v\|_{L^2}^2 \leq C, \\ \|f\|_\infty = \|u_1^J\|_\infty \left\| \sum_{|\lambda| \leq J} \langle v, \psi_\lambda \rangle \psi_\lambda \right\|_\infty \leq C \left\| \sum_{|\lambda| \leq J} |\psi_\lambda| \right\|_\infty \leq C2^{J/2}.$$

We obtain the bounds $V^2 \leq \tilde{C}n$ and $U \leq \tilde{C} \log(n)2^{J/2}$ for some constant \tilde{C} . Applying Theorem 12 yields

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} \frac{|Z(f)|}{n} \geq \tilde{\kappa} \left(\sqrt{\tilde{C}(D+2)} \frac{2^J}{n} + \tilde{C}(D+2) \frac{\log(n)2^{3J/2}}{n} \right) \right) \leq 2\kappa e^{-D2^J}.$$

By choice of $J = J_n$, the first term with $\sqrt{2^J/n}$ dominates the second term for large n and this implies (24). The bound (25) for $\hat{\mathbf{G}}$ follows in the same way by considering the class of functions $\mathcal{F} = \{f(x, y) = (v(x)u_1^J(x) + v(y)u_1^J(y))/2, \|v\|_{L^2} \leq 1, v \in V_J\}$.

We denote the empirical measure by $\mu_n = \frac{1}{n+1} \sum_{l=0}^n \delta_{X_{l\Delta}}$, and define

$$(30) \quad \hat{\mu} = \sum_{|\lambda| \leq J} \frac{1}{n+1} \sum_{l=0}^n \psi_\lambda(X_{l\Delta}) \psi_\lambda,$$

$K_J(x, y) = \sum_{|\lambda| \leq J} \psi_\lambda(x) \psi_\lambda(y)$ and $K_J(\mu) = \int K_J(\cdot, y) \mu(y) dy$. We consider the variance term $\hat{\mu} - \pi_J \mu$ and represent, for B_0 a countable subset of the unit ball B of $L^2([0, 1])$,

$$\|H\|_{L^2} = \sup_{f \in B_0} \left| \int_{[0,1]} H(t) f(t) dt \right|, \quad H \in L^2([0, 1]).$$

Then $\|\hat{\mu} - \pi_J \mu\|_{L^2} = \|\mu_n - \mu\|_{\mathcal{K}}$ with $\|H\|_{\mathcal{K}} := \sup_{k \in \mathcal{K}} |H(k)|$ and

$$\mathcal{K} := \left\{ x \mapsto \int_{[0,1]} f(t) K_J(t, x) dt - \int_{[0,1]} f(t) K_J(\mu)(t) dt : f \in B_0 \right\}.$$

We apply the concentration inequality Theorem 12 to the class $\mathcal{K} \subseteq V_J$. As in (20) and (22) in [11], we bound $\sup_{k \in \mathcal{K}} \|k\|_\infty \leq C2^{J/2}$ and $\sup_{k \in \mathcal{K}} \|k\|_{L^2(\mu)}^2 \leq C$. We

obtain

$$\mathbb{P}\left(\|\widehat{\mu} - \pi_J \mu\|_{L^2} \geq \tilde{\kappa} \left(\sqrt{C(D+2) \frac{2^J}{n}} + C(D+2) \frac{\log(n)2^{3J/2}}{n} \right)\right) \leq 2\kappa e^{-D2^J}.$$

By choice of $J = J_n$, the term $\sqrt{2^J/n}$ dominates the second term and we have for $C > 0$ large enough

$$\mathbb{P}\left(\|\widehat{\mu} - \pi_J \mu\|_{L^2} \geq C\sqrt{\frac{2^J}{n}}\right) \leq 2\kappa e^{-D2^J}.$$

Since $\|\mu\|_{H^s}$ is uniformly bounded over Θ_s [cf. (10)], we have $\|\mu - \pi_J \mu\|_{L^2} \leq C2^{-Js}$. By the triangle inequality and the assumption $2^{-Js} \leq c\sqrt{2^J/n}$, we obtain (26) by possibly increasing the constant C . Claim (29) follows by a similar empirical process type bound for $\mu_n - \mu$, corresponding to the case $r = \infty$ in Section 3.1.2 in [11], with $\delta_n = \sqrt{n}\varepsilon_n^2 \rightarrow 0$ there eventually less than any $\delta > 0$ (and using Theorem 12 in place of Talagand’s inequality). Details are left to the reader.

Next, we use the bound

$$(31) \quad \|\widehat{\mathbf{G}} - \mathbf{G}\|_{\ell^2 \rightarrow \ell^2} \leq \sum_{|\lambda| \leq J} \|(\widehat{\mathbf{G}} - \mathbf{G})e_\lambda\|_{\ell^2},$$

where e_λ are orthonormal vectors of $(V, \|\cdot\|_{\ell^2})$. We represent

$$\begin{aligned} \|(\widehat{\mathbf{G}} - \mathbf{G})e_\lambda\|_{\ell^2} &= \sup_{\|v\|_{L^2} \leq 1, v \in V_J} \left| \frac{1}{n} \sum_{l=1}^n \frac{1}{2} (v(X_{(l-1)\Delta})\psi_\lambda(X_{(l-1)\Delta}) \right. \\ &\quad \left. + v(X_{l\Delta})\psi_\lambda(X_{l\Delta})) - \mathbb{E}[v(X_0)\psi_\lambda(X_0)] \right|. \end{aligned}$$

Similar as before, we consider functions of the form $f(x, y) = (v(x)\psi_\lambda(x) + v(y)\psi_\lambda(y))/2$. Using $\|\psi_\lambda\|_\infty \leq C2^{J/2}$ we calculate $\|f\|_{L^2(\mu_2)}^2 \leq C2^J$ and $\|f\|_\infty \leq C2^J$. For fixed λ , Theorem 12 yields the concentration inequality

$$\mathbb{P}\left(\|(\widehat{\mathbf{G}} - \mathbf{G})e_\lambda\|_{L^2} \geq \tilde{\kappa} \left(\sqrt{C(D+2) \frac{2^{2J}}{n}} + C(D+2) \frac{\log(n)2^{2J}}{n} \right)\right) \leq 2\kappa e^{-D2^J}.$$

The first term in the sum dominates for large n . Upon choosing a larger constant $C > 0$, we obtain

$$\mathbb{P}\left(\|(\widehat{\mathbf{G}} - \mathbf{G})e_\lambda\|_{L^2} \geq C\frac{2^J}{\sqrt{n}}\right) \leq 2\kappa e^{-D2^J}.$$

By observing that the sum in (31) is over 2^{J+1} summands, we obtain (27), by enlarging the constant C if necessary.

The final bound (28) for $\|\widehat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta^J\|_{\ell^2 \rightarrow \ell^2}$ follows similarly by considering the functions $f(x, y) = (v(x)\psi_\lambda(y) + v(y)\psi_\lambda(x))/2$. \square

In the following, we assume $2^J \leq cn^{1/4}/\log n$ and will say that an event A occurs with *sufficiently high probability* if for all $D > 0$ there exists $C > 0$ such that $\mathbb{P}(A)$ is at least as large as the probability of the intersection of the events in the previous theorem. Then for n large enough the events in (27)–(29) include the events that $\|\widehat{\mathbf{G}} - \mathbf{G}\|_{\ell^2 \rightarrow \ell^2} \leq \frac{1}{2} \|\mathbf{G}^{-1}\|_{\ell^2 \rightarrow \ell^2}^{-1}$, $\|\widehat{\mu} - \mu\|_{L^\infty([0,1])} \leq \frac{1}{2} \inf_{x \in [0,1]} \mu(x)$. This implies that $\widehat{\mathbf{G}}$ is invertible with $\|\widehat{\mathbf{G}}^{-1}\|_{\ell^2 \rightarrow \ell^2} \leq 2\|\mathbf{G}^{-1}\|_{\ell^2 \rightarrow \ell^2}$ and that $\widehat{\mu}$ is bounded away from zero on $[0, 1]$.

LEMMA 20. *Assume $2^J \leq cn^{1/4}/\log n$. For n large enough, we have with sufficiently high probability and uniformly over Θ_s*

$$\|(\widehat{\mathbf{G}}^{-1} \widehat{\mathbf{P}}_\Delta - \mathbf{G}^{-1} \mathbf{P}_\Delta^J) \mathbf{u}_1^J\|_{\ell^2} < C \sqrt{\frac{2^J}{n}}.$$

PROOF. We decompose

$$\begin{aligned} \widehat{\mathbf{G}}^{-1} \widehat{\mathbf{P}}_\Delta - \mathbf{G}^{-1} \mathbf{P}_\Delta^J &= \widehat{\mathbf{G}}^{-1} (\widehat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta^J) + (\widehat{\mathbf{G}}^{-1} - \mathbf{G}^{-1}) \mathbf{P}_\Delta^J \\ (32) \qquad \qquad \qquad &= \widehat{\mathbf{G}}^{-1} ((\widehat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta^J) + (\mathbf{G} - \widehat{\mathbf{G}}) \mathbf{G}^{-1} \mathbf{P}_\Delta^J). \end{aligned}$$

Using that $\mathbf{G}^{-1} \mathbf{P}_\Delta^J \mathbf{u}_1^J = \kappa_1^J \mathbf{u}_1^J$ and $\|\widehat{\mathbf{G}}^{-1}\| \leq 2\|\mathbf{G}^{-1}\|$, we obtain

$$\begin{aligned} &\|(\widehat{\mathbf{G}}^{-1} \widehat{\mathbf{P}}_\Delta - \mathbf{G}^{-1} \mathbf{P}_\Delta^J) \mathbf{u}_1^J\|_{\ell^2} \\ &\leq 2\|\mathbf{G}^{-1}\| (\|(\widehat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta^J) \mathbf{u}_1^J\|_{\ell^2} + \|(\mathbf{G} - \widehat{\mathbf{G}}) \kappa_1^J \mathbf{u}_1^J\|_{\ell^2}). \end{aligned}$$

The results follows from this and (24), (25). \square

LEMMA 21. *Assume $2^J \leq cn^{1/4}/\log n$. For n large enough, we have with sufficiently high probability and uniformly over Θ_s*

$$\|\widehat{\mathbf{G}}^{-1} \widehat{\mathbf{P}}_\Delta - \mathbf{G}^{-1} \mathbf{P}_\Delta^J\|_{\ell^2 \rightarrow \ell^2} < C \frac{2^{2J}}{\sqrt{n}}.$$

PROOF. From (32), we deduce

$$\begin{aligned} \|\widehat{\mathbf{G}}^{-1} \widehat{\mathbf{P}}_\Delta - \mathbf{G}^{-1} \mathbf{P}_\Delta^J\| &\leq 2\|\mathbf{G}^{-1}\| (\|\widehat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta^J\| + \|\mathbf{G} - \widehat{\mathbf{G}}\| \|\mathbf{G}^{-1}\| \|\mathbf{P}_\Delta^J\|) \\ &\leq C (\|\widehat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta^J\| + \|\mathbf{G} - \widehat{\mathbf{G}}\|). \end{aligned}$$

The result follows by applying the concentration from (27) and (28). \square

LEMMA 22. *Assume $2^J \leq cn^{1/4}/\log n$. Let $\widehat{\kappa}_1$ be the second largest eigenvalue of the matrix $\widehat{\mathbf{G}}^{-1} \widehat{\mathbf{P}}_\Delta$ with corresponding eigenvector $\widehat{\mathbf{u}}_1$ and eigenfunction*

$\widehat{u}_1 = \sum_{\lambda} (\widehat{u}_1)_{\lambda} \psi_{\lambda} \in V_J$. For n large enough, we have with sufficiently high probability and uniformly over Θ_s

$$|\widehat{\kappa}_1 - \kappa_1^J| + \|\widehat{u}_1 - u_1^J\|_{\ell^2} < C\sqrt{\frac{2^J}{n}},$$

$$\|\widehat{u}_1 - u_1^J\|_{H^1} < C\sqrt{\frac{2^{3J}}{n}}, \quad \|\widehat{u}_1 - u_1^J\|_{H^2} < C\sqrt{\frac{2^{5J}}{n}}.$$

PROOF. By Lemma 21, we have that $\|\widehat{G}^{-1}\widehat{P}_{\Delta} - G^{-1}P_{\Delta}^J\|_{\ell^2 \rightarrow \ell^2}$ converges to zero. Thus, the concentration in Lemma 20 carries over to concentration of $\widehat{\kappa}_1$ and \widehat{u}_1 by Proposition 4.2 and Corollary 4.3 in [13]. The uniform choice of ρ and R is possible as in the proof of their Corollary 4.15. The second and third claim are consequences of the first by the usual Bernstein inequalities for functions in V_J : $\|\widehat{u}_1 - u_1^J\|_{H^1} \leq C2^J \|\widehat{u}_1 - u_1^J\|_{L^2}$ and $\|\widehat{u}_1 - u_1^J\|_{H^2} \leq C2^{2J} \|\widehat{u}_1 - u_1^J\|_{L^2}$ (arguing, e.g., as in Proposition 4.2.8 in [12]). \square

PROOF OF THEOREM 16. By starting with a slightly larger constant $\widetilde{D} > D$, the factor in front of the exponential function can be removed and events of sufficiently high probability are seen to have probability at least $1 - e^{-D2^J}$. We then choose $2^J = n\varepsilon_n^2$. By Lemma 22, the corresponding bias estimates in [13] and the Sobolev imbedding, we have with sufficiently high probability

$$|\widehat{\kappa}_1 - \kappa_1| + \|\widehat{u}_1 - u_1\|_{H^1} < Cn\varepsilon_n^3,$$

$$\|\widehat{u}'_1 - u'_1\|_{L^{\infty}([A, B])} \leq \|\widehat{u}_1 - u_1\|_{H^2} < Cn^2\varepsilon_n^5,$$

where we used that the bias term is dominated by the variance term since $\varepsilon_n \gtrsim n^{-(s+1)/(2s+3)}$. For the estimation of $\widehat{\mu}$, we choose $\bar{J} \geq J$ differently such that $2^{\bar{J}} \sim n^{1/(2s+1)}$ and obtain

$$\|\widehat{\mu} - \mu\|_{L^2} < Cn^{-s/(2s+1)} = o(n\varepsilon_n^3).$$

In addition, the event can be chosen such that $\widehat{\mu}$ and \widehat{u}'_1 are bounded from below on $[A, B]$ uniformly over Θ_s , since μ and u'_1 are (Proposition 6.5 in [13]). By Lemma 6.6 in [13], we have that $\|u_1\|_{H^{s+1}}$, $s \geq 2$, is bounded uniformly over Θ_s . This implies in particular uniform bounds for $\|u_1\|_{L^2}$, $\|u'_1\|_{L^2}$, $\|u''_1\|_{L^2}$, $\|u_1\|_{\infty}$ and $\|u'_1\|_{\infty}$. By the convergence of \widehat{u}_1 in H^2 , these bounds carry over to bounds on \widehat{u}_1 . From the expressions (21) for $\widehat{\sigma}$ and (22) for \widehat{b} and the above bounds, we deduce Theorem 16. \square

4.4. Conclusion of the proof of Theorem 1. Theorem 1 follows from Theorem 13: We choose $\mathcal{B}_n = \Theta_s$ for all n . By Lemma 14, there exists \bar{C} such that

$$\left\{ (\sigma, b) \in \Theta : \|\mu - \mu_0\|_{L^2((0,1))} + \|\sigma^{-2} - \sigma_0^{-2}\|_{(B^1_{\infty})^*} + \|b - b_0\|_{(B^2_{\infty})^*} < \frac{\varepsilon}{\bar{C}} \right\}$$

$$\subseteq B_{\varepsilon, \kappa}.$$

By assumption, we have (4) and by dividing C by \bar{C}^2 we ensure (15) with a possibly different constant C . We define d_n as in (23), so that the existence of tests is guaranteed by Theorem 17. The result follows.

5. Proofs III: Wavelet series priors. We record the following technical lemma whose proof is given in the supplement [18]. Define the dual norm

$$(33) \quad \|f\|_{(B_{1\infty}^s)^*} := \sup_{g: \|g\|_{B_{1\infty}^s} \leq 1} \left| \int_0^1 f(x)g(x) dx \right|, \quad s \geq 0,$$

where the norm of $B_{1\infty}^s$ is defined as in (4.79) or equivalently (4.149) in [12].

LEMMA 23. (a) *Let f, g have $B_{\infty\infty}^1$ -norm at most B' . Then there exists a constant $c(B')$ such that*

$$\|e^f - e^g\|_{(B_{1\infty}^1)^*} \leq c(B') \|f - g\|_{(B_{1\infty}^1)^*}.$$

(b) *For all $f \in L^\infty$, we have, for $s > 0$,*

$$\|f\|_{(B_{1\infty}^s)^*} \leq \|f\|_{B_{\infty 1}^{-s}} \equiv \sum_l 2^{-l(s-1/2)} \max_k |\langle f, \psi_{lk} \rangle_{L^2([0,1])}|.$$

(c) *For all $(\sigma, b), (\sigma_0, b_0) \in \Theta$ with corresponding invariant measures μ, μ_0 , assuming also that $\sigma, \sigma_0, \mu, \mu_0$ are all periodic on $[0, 1]$, we have*

$$\|b - b_0\|_{(B_{1\infty}^2)^*} \lesssim \|\mu - \mu_0\|_{L^2} + \|\sigma^{-2} - \sigma_0^{-2}\|_{(B_{1\infty}^1)^*}.$$

PROOF OF PROPOSITION 6. We first show that $\Pi(\Theta_s) = 1$: By construction of the priors $(\log(\sigma^{-2}), \log \mu)$ is almost surely norm-bounded in $\mathcal{C}^s \times \mathcal{C}^{s+1}$ by \tilde{B} , and this bound carries over to (σ^2, μ) up to constants. By (7), we thus have $\|b\|_{\mathcal{C}^{s-1}} \lesssim \|\sigma^2\|_{\mathcal{C}^s} + \|\mu\|_{\mathcal{C}^s} \lesssim \tilde{B}$. Then by (5) and the remarks before it, we have the continuous imbeddings $(\sigma, b) \in (\mathcal{C}^s \times \mathcal{C}^{s-1}) \subseteq (H^s \times H^{s-1}) \cap (B_{\infty 1}^s \times B_{\infty 1}^{s-1}) \subseteq \mathcal{C}^2 \times \mathcal{C}^1$. Summarising, given \tilde{B} , $(\sigma, b) \in \Theta_s$ is true Π almost surely for suitable $D = D(\tilde{B})$ and $d = d(\tilde{B})$.

To verify the small ball estimate, note that by Lemma 23(c) and independence of the priors,

$$\begin{aligned} & \Pi(\vartheta = (\sigma, b) \in \Theta : \|\mu - \mu_0\|_{L^2} + \|\sigma^{-2} - \sigma_0^{-2}\|_{(B_{1\infty}^1)^*} + \|b - b_0\|_{(B_{1\infty}^2)^*} < \varepsilon_n) \\ & \geq \Pi\left(\|\sigma^{-2} - \sigma_0^{-2}\|_{(B_{1\infty}^1)^*} + \|\mu - \mu_0\|_{L^2} < \frac{2\varepsilon_n}{c}\right) \\ & \geq \Pi\left(\|\sigma^{-2} - \sigma_0^{-2}\|_{(B_{1\infty}^1)^*} < \frac{\varepsilon_n}{c}\right) \mathbb{P}\left(\|\mu - \mu_0\|_{L^2} < \frac{\varepsilon_n}{c}\right) \end{aligned}$$

for some constant $c > 0$. Examining the first factor, we can use Lemma 23(a), (b) and the definition of the Besov norm to obtain the lower bound:

$$\begin{aligned} & \Pi\left(\|\log \sigma^{-2} - \log \sigma_0^{-2}\|_{B_{\infty^1}^{-1}} < \frac{\varepsilon_n}{c'(\tilde{B})}\right) \\ &= \mathbb{P}\left(\sum_l 2^{-l/2} \max_k |\tau_{lk} - 2^{-l(s+1/2)} l^{-2} u_{lk}| < \frac{\varepsilon_n}{c'(\tilde{B})}\right), \end{aligned}$$

where $u_{lk} = 0$ for all $l > L_n$ (when $L_n < \infty$). We define $t_{lk} = 2^{l(s+1/2)} l^2 \tau_{lk}$ such that $|t_{lk}| \leq \tilde{B}$, and $M(J) = \sum_{l=J_0}^J \sum_{k=0}^{2^l-1} 1 \leq 2 \cdot 2^J$. We choose $J = J_n$ of order $\varepsilon_n \sim 2^{-J(s+1)}/J^2$ but such that $\tilde{c}\varepsilon_n \geq 2^{-J(s+1)}/J^2$ for some constant $\tilde{c} > 0$ to be determined later. By choice of $L = L_n$, we have $2^{-L(s+1)} \lesssim 2^{-J(s+1)}/J^2$ so that L is eventually larger than J . By choosing $\tilde{c} > 0$ small enough, the last probability is bounded below by [all indices (l, k) are tacitly assumed to lie in \mathcal{I} only]

$$\begin{aligned} & \mathbb{P}\left(\sum_{l \leq J} 2^{-l(s+1)} l^{-2} \max_k |t_{lk} - u_{lk}| < \frac{\varepsilon_n}{c'(\tilde{B})} - \tilde{c}2^{-J(s+1)}/J^2\right) \\ & \geq \mathbb{P}\left(\max_{l \leq J} \max_k |t_{lk} - u_{lk}| < c' \varepsilon_n\right) = \prod_{l \leq J} \prod_k \mathbb{P}(|t_{lk} - u_{lk}| < c' \varepsilon_n) \\ & \geq (\zeta c' \varepsilon_n)^{M(J)} \geq e^{-c''(\log n)^{1-2/(s+1)}/\varepsilon_n^{1/(s+1)}} \geq e^{-Cn\varepsilon_n^2/2} \end{aligned}$$

for some constant $C > 0$, completing the treatment of this term. For the second term, notice that since $H, H_0 = \log \mu_0$ are bounded functions the exponential map is Lipschitz on the union of their ranges, and thus $\|\mu - \mu_0\|_2 \lesssim \|H - H_0\|_\infty$. Then one proves, using $\|h\|_\infty \lesssim \sum_l 2^{l/2} \max_k |\langle h, \psi_{lk} \rangle|$ and proceeding just as above with $\bar{u}_{lk} = 0$ for $l > \bar{L}_n$, that (again all indices are tacitly assumed to lie in \mathcal{I} only)

$$\begin{aligned} & \mathbb{P}(\|H - H_0\|_\infty < c' \varepsilon_n) \\ & \geq \mathbb{P}\left(\sum_l 2^{l/2} \max_k |\beta_{lk} - 2^{-l(s+3/2)} l^{-2} \bar{u}_{lk}| < c'' \varepsilon_n\right) \end{aligned}$$

is lower bounded by $e^{-Cn\varepsilon_n^2/2}$. We conclude overall that for n large enough,

$$\Pi((\sigma, b) \in \Theta : \|\sigma^{-2} - \sigma_0^{-2}\|_{(B_{1^\infty}^1)^*} + \|\mu - \mu_0\|_{L^2} < \varepsilon_n) \geq e^{-Cn\varepsilon_n^2}. \quad \square$$

Acknowledgements. The authors thank an Associate Editor and two referees for careful reading of and critical remarks on the manuscript, Radosław Adamczak for helpful discussions on concentration inequalities for Markov chains.

SUPPLEMENTARY MATERIAL

The supplement to “Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions” (DOI: [10.1214/16-AOS1504SUPP](https://doi.org/10.1214/16-AOS1504SUPP.pdf); .pdf). This supplement contains several proofs of results in the main paper and states and proves a proposition on Lipschitz properties of self-adjoint operators.

REFERENCES

- [1] ADAMCZAK, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. [MR2424985](#)
- [2] BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften* **348**. Springer, Cham. [MR3155209](#)
- [3] BARAUD, Y. (2010). A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression. *Bernoulli* **16** 1064–1085. [MR2759169](#)
- [4] BASS, R. F. (1998). *Diffusions and Elliptic Operators*. Springer, New York. [MR1483890](#)
- [5] BASS, R. F. (2011). *Stochastic Processes. Cambridge Series in Statistical and Probabilistic Mathematics* **33**. Cambridge Univ. Press, Cambridge. [MR2856623](#)
- [6] BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. [MR2114987](#)
- [7] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. [MR3127856](#)
- [8] CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. [MR3262473](#)
- [9] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. [MR1790007](#)
- [10] GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** 192–223. [MR2332274](#)
- [11] GINÉ, E. and NICKL, R. (2011). Rates on contraction for posterior distributions in L^r -metrics, $1 \leq r \leq \infty$. *Ann. Statist.* **39** 2883–2911. [MR3012395](#)
- [12] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge University Press, Cambridge.
- [13] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. [MR2102509](#)
- [14] GOLIGHTLY, A. and WILKINSON, D. J. (2005). Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* **61** 781–788. [MR2196166](#)
- [15] GUGUSHVILI, S. and SPREIJ, P. (2014). Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. *Lith. Math. J.* **54** 127–141. [MR3212631](#)
- [16] KOSKELA, J., SPANO, D. and JENKINS, P. A. (2015). Consistency of Bayesian nonparametric inference for discretely observed jump diffusions. Preprint. Available at [arXiv:1506.04709](#).
- [17] MEYER, Y. (1992). *Wavelets and Operators. Cambridge Studies in Advanced Mathematics* **37**. Cambridge Univ. Press, Cambridge. [MR1228209](#)
- [18] NICKL, R. and SÖHL, J. (2017). Supplement to “Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions.” DOI:[10.1214/16-AOS1504SUPP](#).
- [19] 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](#)
- [20] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** 1–32. [MR3383563](#)
- [21] 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](#)
- [22] RAY, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* **7** 2516–2549. [MR3117105](#)

- [23] ROBERTS, G. O. and STRAMER, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika* **88** 603–621. [MR1859397](#)
- [24] SÖHL, J. and TRABS, M. (2016). Adaptive confidence bands for Markov chains and diffusions: Estimating the invariant measure and the drift. *ESAIM Probab. Stat.* **20** 432–462.
- [25] STROOCK, D. W. and VARADHAN, S. R. S. (1971). Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.* **24** 147–225. [MR0277037](#)
- [26] STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. [MR2652785](#)
- [27] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. [MR3357861](#)
- [28] TRIEBEL, H. (2010). *Theory of Function Spaces*. Birkhäuser/Springer Basel AG, Basel. [MR3024598](#)
- [29] 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](#)
- [30] VAN ZANTEN, H. (2013). Nonparametric Bayesian methods for one-dimensional diffusion models. *Math. Biosci.* **243** 215–222. [MR3065207](#)
- [31] VAN DER MEULEN, F., SCHAUER, M. and VAN ZANTEN, H. (2014). Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Comput. Statist. Data Anal.* **71** 615–632. [MR3131993](#)
- [32] VAN DER MEULEN, F. and VAN ZANTEN, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli* **19** 44–63. [MR3019485](#)
- [33] 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](#)
- [34] 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](#)

STATISTICAL LABORATORY
DEPARTMENT OF PURE MATHEMATICS
AND MATHEMATICAL STATISTICS
UNIVERSITY OF CAMBRIDGE
CB3 0WB, CAMBRIDGE
UNITED KINGDOM
E-MAIL: r.nickl@statslab.cam.ac.uk
j.soehl@statslab.cam.ac.uk