

## Adaptive nonparametric drift estimation for diffusion processes using Faber–Schauder expansions

van der Meulen, Frank; Schauer, Moritz; van Waaij, Jan

**DOI**

[10.1007/s11203-017-9163-7](https://doi.org/10.1007/s11203-017-9163-7)

**Publication date**

2017

**Document Version**

Final published version

**Published in**

Statistical Inference for Stochastic Processes

**Citation (APA)**

van der Meulen, F., Schauer, M., & van Waaij, J. (2017). Adaptive nonparametric drift estimation for diffusion processes using Faber–Schauder expansions. *Statistical Inference for Stochastic Processes*, 1-26. <https://doi.org/10.1007/s11203-017-9163-7>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# Adaptive nonparametric drift estimation for diffusion processes using Faber–Schauder expansions

Frank van der Meulen<sup>1</sup> · Moritz Schauer<sup>2</sup> ·  
Jan van Waaij<sup>3</sup>

Received: 15 December 2016 / Accepted: 14 June 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** We consider the problem of nonparametric estimation of the drift of a continuously observed one-dimensional diffusion with periodic drift. Motivated by computational considerations, van der Meulen et al. (Comput Stat Data Anal 71:615–632, 2014) defined a prior on the drift as a randomly truncated and randomly scaled Faber–Schauder series expansion with Gaussian coefficients. We study the behaviour of the posterior obtained from this prior from a frequentist asymptotic point of view. If the true data generating drift is smooth, it is proved that the posterior is adaptive with posterior contraction rates for the  $L_2$ -norm that are optimal up to a log factor. Contraction rates in  $L_p$ -norms with  $p \in (2, \infty]$  are derived as well.

## 1 Introduction

Assume continuous time observations  $X^T = \{X_t, : t \in [0, T]\}$  from a diffusion process  $X$  defined as (weak) solution to the stochastic differential equation (sde)

$$dX_t = b_0(X_t) dt + dW_t, \quad X_0 = x_0. \quad (1)$$

Here  $W$  is a Brownian Motion and the drift  $b_0$  is assumed to be a real-valued measurable function on the real line that is 1-periodic and square integrable on  $[0, 1]$ . The assumed

---

✉ Jan van Waaij  
j.vanwaaij@uva.nl

Frank van der Meulen  
f.h.vandermeulen@tudelft.nl

Moritz Schauer  
m.r.schauer@math.leidenuniv.nl

<sup>1</sup> TU Delft, Mekelweg 4, 2628 CD Delft, The Netherlands

<sup>2</sup> Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

<sup>3</sup> Korteweg-de Vries Institute for Mathematics, Science Park 107, 1098 XG Amsterdam, The Netherlands

periodicity implies that we can alternatively view the process  $X$  as a diffusion on the circle. This model has been used for dynamic modelling of angles, see for instance [Pokern \(2007\)](#) and [Hindriks \(2011\)](#).

We are interested in nonparametric adaptive estimation of the drift. This problem has recently been studied by multiple authors. [Spokoiny \(2000\)](#) proposed a locally linear smoother with a data-driven bandwidth choice that is rate adaptive with respect to  $|b''(x)|$  for all  $x$  and optimal up to a log factors. Interestingly, the result is non-asymptotic and does not require ergodicity. [Dalalyan and Kutoyants \(2002\)](#) and [Dalalyan \(2005\)](#) consider ergodic diffusions and construct estimators that are asymptotically minimax and adaptive under Sobolev smoothness of the drift. Their results were extended to the multidimensional case by [Strauch \(2015\)](#).

In this paper we focus on Bayesian nonparametric estimation, a paradigm that has become increasingly popular over the past two decades. An overview of some advances of Bayesian nonparametric estimation for diffusion processes is given in [van Zanten \(2013\)](#).

The Bayesian approach requires the specification of a prior. Ideally, the prior on the drift is chosen such that drawing from the posterior is computationally efficient while at the same time ensuring that the resulting inference has good theoretical properties, which is quantified by a contraction rate. This is a rate for which we can shrink balls around the true parameter value, while maintaining most of the posterior mass. More formally, if  $d$  is a semimetric on the space of drift functions, a contraction rate  $\varepsilon_T$  is a sequence of positive numbers  $\varepsilon_T \downarrow 0$  for which the posterior mass of the balls  $\{b : d(b, b_0) \leq \varepsilon_T\}$  converges in probability to 1 as  $T \rightarrow \infty$ , under the law of  $X$  with drift  $b_0$ . For a general discussion on contraction rates, see for instance [Ghosal et al. \(2000\)](#) and [Ghosal and van der Vaart \(2007\)](#).

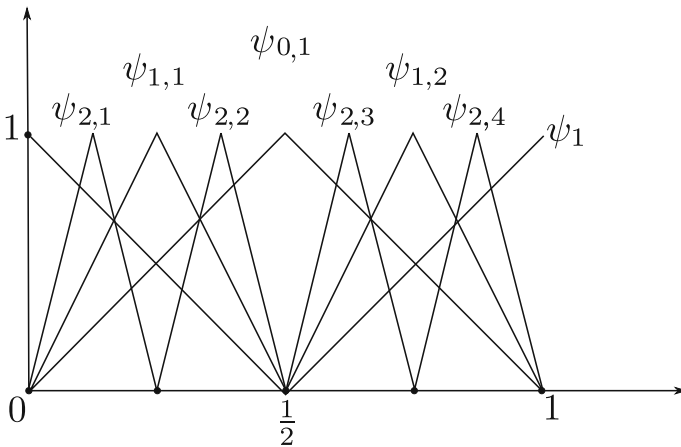
For diffusions, the problem of deriving optimal posterior convergence rates has been studied recently under the additional assumption that the drift integrates to zero,  $\int_0^1 b_0(x) dx = 0$ . In [Papaspiliopoulos et al. \(2012\)](#) a mean zero Gaussian process prior is proposed together with an algorithm to sample from the posterior. The precision operator (inverse covariance operator) of the proposed Gaussian process is given by  $\eta ((-\Delta)^{\alpha+1/2} + \kappa I)$ , where  $\Delta$  is the one-dimensional Laplacian,  $I$  is the identity operator,  $\eta, \kappa > 0$  and  $\alpha + 1/2 \in \{2, 3, \dots\}$ . A first consistency result was shown in [Pokern et al. \(2013\)](#).

In [van Waaij and van Zanten \(2016\)](#) it was shown that this rate result can be improved upon for a slightly more general class of priors on the drift. More specifically, in this paper the authors consider a prior which is defined as

$$b = L \sum_{k=1}^{\infty} k^{-1/2-\alpha} \varphi_k Z_k, \quad (2)$$

where  $\varphi_{2k}(x) = \sqrt{2} \cos(2\pi kx)$ ,  $\varphi_{2k-1}(x) = \sqrt{2} \sin(2\pi kx)$  are the standard Fourier series basis functions,  $\{Z_k\}$  is a sequence of independent standard normally distributed random variables and  $\alpha$  is positive. It is shown that when  $L$  and  $\alpha$  are fixed and  $b_0$  is assumed to be  $\alpha$ -Sobolev smooth, then the optimal posterior rate of contraction,  $T^{-\alpha/(1+2\alpha)}$ , is obtained. Note that this result is nonadaptive, as the regularity of the prior must match the regularity of  $b_0$ . For obtaining optimal posterior contraction rates for the full range of possible regularities of the drift, two options are investigated: endowing either  $L$  or  $\alpha$  with a hyperprior. Only the second option results in the desired adaptivity over all possible regularities.

While the prior in (2) (with additional prior on  $\alpha$ ) has good asymptotic properties, from a computational point of view the infinite series expansion is inconvenient. Clearly, in any implementation this expansion needs to be truncated. Random truncation of a series expansion is a well known method for defining priors in Bayesian nonparametrics, see for instance [Shen](#)



**Fig. 1** Elements  $\psi_1$  and  $\psi_{j,k}$ ,  $0 \leq j \leq 2$  of the Faber–Schauder basis

and Ghosal (2015). Exactly this idea was exploited in van der Meulen et al. (2014), where the prior is defined as the law of the random function

$$b^{R,S} = SZ_1\psi_1 + S \sum_{j=0}^R \sum_{k=1}^{2^j} Z_{jk}\psi_{jk}, \tag{3}$$

where the functions  $\psi_{jk}$  constitute the Faber–Schauder basis (see Fig. 1).

These functions feature prominently in the Lévy-Ciesielski construction of Brownian motion (see for instance (Bhattacharya and Waymire 2007, paragraph 10.1)).

The prior coefficients  $Z_{jk}$  are equipped with a Gaussian distribution, and the truncation level  $R$  and the scaling factor  $S$  are equipped with independent priors. Truncation in absence of scaling increases the apparent smoothness of the prior (as illustrated for deterministic truncation by example 4.5 in van der Vaart and van Zanten (2008)), whereas scaling by a number  $\geq 1$  decreases the apparent smoothness. (Scaling with a number  $\leq 1$  only increases the apparent smoothness to a limited extent, see for example Knapik et al. (2011).)

The simplest type of prior is obtained by taking the coefficients  $Z_{jk}$  independent. We do however also consider the prior that is obtained by first expanding a periodic Ornstein–Uhlenbeck process into the Faber–Schauder basis, followed by random scaling and truncation. We will explain that specific stationarity properties of this prior make it a natural choice.

Draws from the posterior can be computed using a reversible jump Markov Chain Monte Carlo (MCMC) algorithm (cf. van der Meulen et al. (2014)). For both types of priors, fast computation is facilitated by leveraging inherent sparsity properties stemming from the compact support of the functions  $\psi_{jk}$ . In the discussion of van der Meulen et al. (2014) it was argued that inclusion of both the scaling and random truncation in the prior is beneficial. However, this claim was only supported by simulations results.

*In this paper we support this claim theoretically by proving adaptive contraction rates of the posterior distribution in case the prior (3) is used.* We start from a general result in van der Meulen et al. (2006) on Brownian semimartingale models, which we adapt to our setting. Here we take into account that as the drift is assumed to be one-periodic, information accumulates in a different way compared to (general) ergodic diffusions. Subsequently we verify that the resulting prior mass, remaining mass and entropy conditions appearing in this

adapted result are satisfied for the prior defined in Eq. (3). An application of our results shows that if the true drift function is  $B_{\infty, \infty}^{\beta}$ -Besov smooth,  $\beta \in (0, 2)$ , then by appropriate choice of the variances of  $Z_{jk}$ , as well as the priors on  $R$  and  $S$ , the posterior for the drift  $b$  contracts at the rate  $(T/\log T)^{-\beta/(1+2\beta)}$  around the true drift in the  $L_2$ -norm. Up to the log factor this rate is minimax-optimal (See for instance [Kutoyants 2004](#), Theorem 4.48). Moreover, it is adaptive: the prior does not depend on  $\beta$ . In case the true drift has Besov-smoothness greater than or equal to 2, our method guarantees contraction rates equal to essentially  $T^{-2/5}$  (corresponding to  $\beta = 2$ ). A further application of our results shows that for  $L_p$ -norms we obtain contraction rate  $T^{-(\beta-1/2+1/p)/(1+2\beta)}$ , up to log-factors.

The paper is organised as follows. In the next section we give a precise definition of the prior. In Sect. 3 a general contraction result for the class of diffusion processes considered here is derived. Our main result on posterior contraction for  $L^p$ -norms with  $p \geq 2$  is presented in Sect. 4. Many results of this paper concern general properties of the prior and their application is not confined to drift estimation of diffusion processes. To illustrate this, we show in Sect. 5 how these results can easily be adapted to nonparametric regression and nonparametric density estimation. Proofs are gathered in Sect. 6. The appendix contains a couple of technical results.

## 2 Prior construction

### 2.1 Model and posterior

Let

$$L^2(\mathbb{T}) = \left\{ b : \mathbb{R} \rightarrow \mathbb{R} \mid \int_0^1 b(x)^2 dx < \infty \text{ and } b \text{ is } 1\text{-periodic} \right\}$$

be the space of square integrable 1-periodic functions.

**Lemma 1** *If  $b_0 \in L^2(\mathbb{T})$ , then the SDE Eq. (1) has a unique weak solution.*

The proof is in Sect. 6.1.

For  $b \in L^2(\mathbb{T})$ , let  $P^b = P^{b,T}$  denote the law of the process  $X^T$  generated by Eq. (1) when  $b_0$  is replaced by  $b$ . If  $P^0$  denotes the law of  $X^T$  when the drift is zero, then  $P^b$  is absolutely continuous with respect to  $P^0$  with Radon-Nikodym density

$$p_b(X^T) = \exp\left(\int_0^T b(X_t) dX_t - \frac{1}{2} \int_0^T b^2(X_t) dt\right). \quad (4)$$

Given a prior  $\Pi$  on  $L^2(\mathbb{T})$  and path  $X^T$  from (1), the posterior is given by

$$\Pi(b \in A \mid X^T) = \frac{\int_A p_b(X^T) \Pi(db)}{\int p_b(X^T) \Pi(db)}, \quad (5)$$

where  $A$  is Borel set of  $L^2(\mathbb{T})$ . These assertions are verified as part of the proof of Theorem 3.

### 2.2 Motivating the choice of prior

We are interested in randomly truncated, scaled series priors that simultaneously enable a fast algorithm for obtaining draws from the posterior and enjoy good contraction rates.

To explain what we mean by the first item, consider first a prior that is a *finite series* prior. Let  $\{\psi_1, \dots, \psi_r\}$  denote basis functions and  $Z = (Z_1, \dots, Z_r)$  a mean zero Gaussian random vector with precision matrix  $\Gamma$ . Assume that the prior for  $b$  is given by  $b = \sum_{i=1}^r Z_i \psi_i$ . By conjugacy, it follows that  $Z \mid X^T \sim N(W^{-1}\mu, W^{-1})$ , where  $W = G + \Gamma$ ,

$$\mu_i = \int_0^T \psi_i(X_t) dX_t \quad \text{and} \quad G_{i,i'} = \int_0^T \psi_i(X_t)\psi_{i'}(X_t) dt \tag{6}$$

for  $i, i' \in \{1, \dots, r\}$ , cf. (van der Meulen et al. 2014, Lemma 1). The matrix  $G$  is referred to as the Grammian. From these expressions it follows that it is computationally advantageous to exploit *compactly supported* basis functions. Whenever  $\psi_i$  and  $\psi_{i'}$  have nonoverlapping supports, we have  $G_{i,i'} = 0$ . Depending on the choice of such basis functions, the Grammian  $G$  will have a specific sparsity structure (a set of index pairs  $(i, i')$  such that  $G_{i,i'} = 0$ , independently of  $X^T$ .) This sparsity structure is inherited by  $W$  as long as the sparsity structure of the prior precision matrix matches that of  $G$ .

In the next section we make a specific choice for the basis functions and the prior precision matrix  $\Gamma$ .

### 2.3 Definition of the prior

Define the ‘‘hat’’ function  $\Lambda$  by  $\Lambda(x) = (2x)\mathbf{1}_{[0, \frac{1}{2})}(x) + 2(1 - x)\mathbf{1}_{[\frac{1}{2}, 1]}(x)$ . The Faber–Schauder basis functions are given by

$$\psi_{j,k}(x) = \Lambda\left(2^j x - k + 1\right), \quad j \geq 0, k = 1, \dots, 2^j$$

Let

$$\psi_1 = \left(\psi_{0,1}\left(x - \frac{1}{2}\right) + \psi_{0,1}\left(x + \frac{1}{2}\right)\right) \mathbf{I}_{[0,1]}(x).$$

In Fig. 1 we have plotted  $\psi_1$  together with  $\psi_{j,k}$  where  $j \in \{0, 1, 2\}$ .

We define our prior as in (3) with Gaussian coefficients  $Z_1$  and  $Z_{jk}$ , where the truncation level  $R$  and the scaling factor  $S$  are equipped with (hyper)priors. We extend  $b$  periodically if we want to consider  $b$  as function on the real line. If we identify the double index  $(j, k)$  in (3) with the single index  $i = 2^j + k$ , then we can write  $b^{R,S} = S \sum_{i=1}^{2^{R+1}} \psi_i Z_i$ . Let

$$\ell(i) = \begin{cases} 0 & \text{if } i \in \{1, 2\} \\ j & \text{if } i \in \{2^j + 1, \dots, 2^{j+1}\} \quad \text{and } j \geq 1 \end{cases}$$

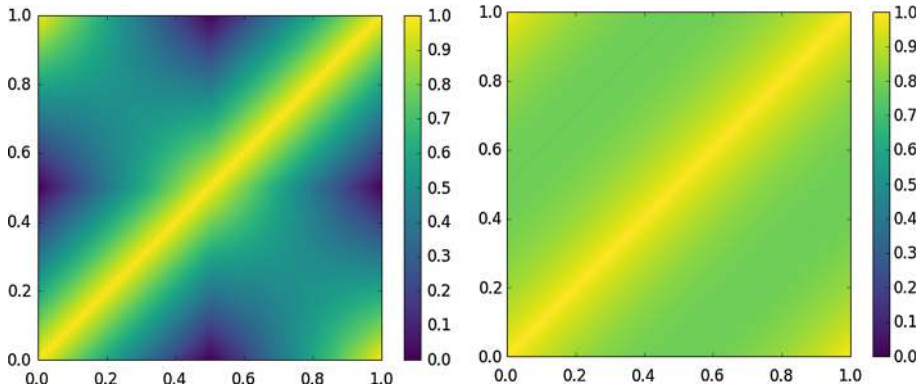
We say that  $\psi_i$  belongs to level  $j \geq 0$  if  $\ell(i) = j$ . Thus both  $\psi_1$  and  $\psi_{0,1}$  belong to level 0, which is convenient for notational purposes. For levels  $j \geq 1$  the basis functions are per level orthogonal with essentially disjoint support. Define for  $r \in \{0, 1, \dots\}$

$$\mathcal{I}_r = \{i : \ell(i) \leq r\} = \{1, 2, \dots, 2^{r+1}\}.$$

Let  $A = (\text{Cov}(Z_i, Z_{i'}))_{i,i' \in \mathbb{N}}$  and define its finite-dimensional restriction by  $A^r = (A_{i,i'})_{i,i' \in \mathcal{I}_r}$ . If we denote  $Z^r = \{Z_i, i \in \mathcal{I}_r\}$ , and assume that  $Z^r$  is multivariate normally distributed with mean zero and covariance matrix  $A^r$ , then the prior has the following hierarchy

$$b \mid R, S, Z^R = S \sum_{i \in \mathcal{I}_R} Z_i \psi_i \tag{7}$$

$$Z^R \mid R \sim N(0, A^R) \tag{8}$$



**Fig. 2** Heat maps of  $(s, t) \mapsto \text{Cov}(b(s), b(t))$ , in case  $S = 1$  and  $R = \infty$ . *Left* Brownian bridge plus the random function  $Z_1 \psi_1$ . *Right* periodic Ornstein–Uhlenbeck process with parameter  $\gamma = 1.48$  and  $\sigma^2$  chosen such that  $\text{Var}(b(s)) = 1$

$$(R, S) \sim \Pi(\cdot). \tag{9}$$

Here, we use  $\Pi$  to denote the joint distribution of  $(R, S)$ .

We will consider two choices of priors for the sequence  $Z_1, Z_2, \dots$ . Our first choice consists of taking independent Gaussian random variables. If the coefficients  $Z_i$  are independent with standard deviation  $2^{-\ell(i)}/2$ , the random draws from this prior are scaled piecewise linear interpolations on a dyadic grid of a Brownian bridge on  $[0, 1]$  plus the random function  $Z_1 \psi_1$ . The choice of  $\psi_1$  is motivated by the fact that in this case  $\text{Var}(b(t)|S = s, R = \infty) = s^2$  is independent of  $t$ .

We construct this second type of prior as follows. For  $\gamma, \sigma^2 > 0$ , define  $V \equiv (V_t, t \in [0, 1])$  to be the cyclically stationary and centred Ornstein–Uhlenbeck process. This is a periodic Gaussian process with covariance kernel

$$\text{Cov}(V(s), V(t)) = \frac{\sigma^2}{2\gamma} \frac{e^{-\gamma h} + e^{-\gamma(1-h)}}{1 - e^{-\gamma}}, \quad h = t - s \geq 0. \tag{10}$$

This process is cyclically stationary, that is, the covariance only depends on  $|t - s|$  and  $1 - |t - s|$ . It is the unique Gaussian and Markovian prior with continuous periodic paths with this property. This makes the cyclically stationary Ornstein–Uhlenbeck prior an appealing choice which respects the symmetries of the problem.

Each realisation of  $V$  is continuous and can be extended to a periodic function on  $\mathbb{R}$ . Then  $V$  can be represented as an infinite series expansion in the Faber–Schauder basis:

$$V_t = \sum_{i \geq 1} Z_i \psi_i(t) = Z_1 \psi_1(t) + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} Z_{j,k} \psi_{j,k}(t) \tag{11}$$

Finally by scaling by  $S$  and truncating at  $R$  we obtain from  $V$  the second choice of prior on the drift function  $b$ . Visualisations of the covariance kernels  $\text{Cov}(b(s), b(t))$  for first prior (Brownian bridge type) and for the second prior (periodic Ornstein–Uhlenbeck process prior with parameter  $\gamma = 1.48$ ) are shown in Fig. 2 (for  $S = 1$  and  $R = \infty$ ).

### 2.4 Sparsity structure induced by choice of $Z_i$

Conditional on  $R$  and  $S$ , the posterior of  $Z^R$  is Gaussian with precision matrix  $G^R + \Gamma^R$  (here  $G^R$  is the Gramian corresponding to using all basis functions up to and including level  $R$ ).

If the coefficients are independent it is trivial to see that the precision matrix  $\Gamma$  does not destroy the sparsity structure of  $G$ , as defined in (6). This is convenient for numerical computations. The next lemma details the situation for periodic Ornstein–Uhlenbeck processes.

**Lemma 2** *Let  $V$  be defined as in Eq. (10)*

1. *The sparsity structure of the precision matrix of the infinite stochastic vector  $Z$  (appearing in the series representation (11)) equals the sparsity structure of  $G$ , as defined in (6).*
2. *The entries of the covariance matrix of the random Gaussian coefficients  $Z_i$  and  $Z_{i'}$ ,  $A_{i,i'} = \mathbb{E}Z_i Z_{i'}$ , satisfy the following bounds:  $A_{11} = A_{22} = \frac{\sigma^2}{2\gamma} \coth(\gamma/2)$  and for  $\gamma \leq 1.5$  and  $i \geq 3$ ,*

$$0.95 \cdot 2^{-\ell(i)} \sigma^2 / 4 \leq A_{ii} \leq 2^{-\ell(i)} \sigma^2 / 4$$

and  $A_{12} = A_{21} = \frac{\sigma^2}{2\gamma} \sinh^{-1}(\gamma/2)$  and for  $i \neq i'$

$$|A_{ii'}| \leq \begin{cases} 0.20\sigma^2 2^{-1.5(\ell(i) \vee \ell(i'))} & i \wedge i' \leq 2 < i \vee i', \\ 0.37\sigma^2 2^{-1.5(\ell(i) + \ell(i'))} & \text{otherwise.} \end{cases}$$

The proof is given in Sect. 6.2. By the first part of the lemma, also this prior does not destroy the sparsity structure of the  $G$ . The second part asserts that while the off-diagonal entries of  $A^r$  are not zero, they are of smaller order than the diagonal entries, quantifying that the covariance matrix of the coefficients in the Schauder expansion is close to a diagonal matrix.

### 3 Posterior contraction for diffusion processes

The main result in [van der Meulen et al. \(2006\)](#) gives sufficient conditions for deriving posterior contraction rates in Brownian semimartingale models. The following theorem is an adaptation and refinement of Theorem 2.1 and Lemma 2.2 of [van der Meulen et al. \(2006\)](#) for diffusions defined on the circle. We assume observations  $X^T$ , where  $T \rightarrow \infty$ . Let  $\Pi^T$  be a prior on  $L^2(\mathbb{T})$  (which henceforth may depend on  $T$ ) and choose measurable subsets (sieves)  $\mathcal{B}_T \subset L^2(\mathbb{T})$ . Define the balls

$$B^T(b_0, \varepsilon) = \{b \in \mathcal{B}_T : \|b_0 - b\|_2 < \varepsilon\}.$$

The  $\varepsilon$ -covering number of a set  $A$  for a semimetric  $\rho$ , denoted by  $N(\varepsilon, A, \rho)$ , is defined as the minimal number of  $\rho$ -balls of radius  $\varepsilon$  needed to cover the set  $A$ . The logarithm of the covering number is referred to as the entropy.

The following theorem characterises the rate of posterior contraction for diffusions on the circle in terms of properties of the prior.

**Theorem 3** *Suppose  $\{\varepsilon_T\}$  is a sequence of positive numbers such that  $T\varepsilon_T^2$  is bounded away from zero. Assume that there is a constant  $\xi > 0$  such that for every  $K > 0$  there is a*



measurable set  $\mathcal{B}_T \subseteq L^2(\mathbb{T})$  and for every  $a > 0$  there is a constant  $C > 0$  such that for  $T$  big enough

$$\log N(a\varepsilon_T, B^T(b_0, \varepsilon_T), \|\cdot\|_2) \leq CT\varepsilon_T^2, \tag{12}$$

$$\Pi^T(B^T(b_0, \varepsilon_T)) \geq e^{-\xi T\varepsilon_T^2}, \tag{13}$$

and

$$\Pi^T(L^2(\mathbb{T}) \setminus \mathcal{B}_T) \leq e^{-KT\varepsilon_T^2}. \tag{14}$$

Then for every  $M_T \rightarrow \infty$

$$P_{b_0} \Pi^T(b \in L^2(\mathbb{T}) : \|b - b_0\|_2 \geq M_T\varepsilon_T \mid X^T) \rightarrow 0$$

and for  $K$  big enough,

$$\Pi^T(L^2(\mathbb{T}) \setminus \mathcal{B}_T \mid X^T) \rightarrow 0. \tag{15}$$

Equations (12), (13) and (14) are referred to as the entropy condition, small ball condition and remaining mass condition of Theorem 3 respectively. The proof of this theorem is in Sect. 6.3.

### 4 Theorems on posterior contraction rates

The main result of this section, Theorem 9 characterises the frequentist rate of contraction of the posterior probability around a fixed parameter  $b_0$  of unknown smoothness using the truncated series prior from Sect. 2.

We make the following assumption on the true drift function.

**Assumption 4** The true drift  $b_0$  can be expanded in the Faber–Schauder basis,  $b_0 = z_1\psi_1 + \sum_{j=0}^\infty \sum_{k=1}^{2^j} z_{jk}\psi_{jk} = \sum_{i \geq 1} z_i\psi_i$  and there exists a  $\beta \in (0, \infty)$  such that

$$\llbracket b_0 \rrbracket_\beta := \sup_{i \geq 1} 2^{\beta \ell(i)} |z_i| < \infty. \tag{16}$$

Note that we use a slightly different symbol for the norm, as we denote the  $L^2$ -norm by  $\|\cdot\|_2$ .

*Remark 5* If  $\beta \in (0, 2)$ , then Assumption 4 on  $b_0$  is equivalent to assuming  $b_0$  to be  $B_{\infty, \infty}^\beta$ -Besov smooth. It follows from the definition of the basis functions that

$$z_{jk} = b_0 \left( (2k - 1)2^{-(j+2)} \right) - \frac{1}{2} b_0 \left( 2^{-(j+2)}(2k - 2) \right) - \frac{1}{2} b_0 \left( 2^{-(j-2)}2k \right).$$

Therefore it follows from equations (4.72) (with  $r = 2$ ) and (4.73) (with  $p = \infty$ ) in combination with equation (4.79) (with  $q = \infty$ ) in Giné and Nickl (2016), Section 4.3, that  $\|b_0\|_\infty + \llbracket b_0 \rrbracket_\beta$  is equivalent to the  $B_{\infty, \infty}^\beta$ -norm of  $b_0$  for  $\beta \in (0, 2)$ .

If  $\beta \in (0, 1)$ , then  $\beta$ -Hölder smoothness and  $B_{\infty, \infty}^\beta$ -smoothness coincide (cf. Proposition 4.3.23 in Giné and Nickl (2016)).

For the prior defined in Eqs. (7)–(9) we make the following assumptions.

**Assumption 6** The covariance matrix  $A$  satisfies one of the following conditions:

(A) For fixed  $\alpha > 0$ ,  $A_{ii} = 2^{-2\alpha \ell(i)}$  and  $A_{ii'} = 0$  for  $i \neq i'$ .

(B) There exists  $0 < c_1 < c_2$  and  $0 < c_3$  with  $3c_3 < c_1$  independent from  $r$ , such that for all  $i, i' \in \mathcal{I}_r$

$$c_1 2^{-\ell(i)} \leq A_{ii} \leq c_2 2^{-\ell(i)},$$

$$|A_{i i'}| \leq c_3 2^{-1.5(\ell(i) + \ell(i'))} \quad \text{if } i \neq i'.$$

In particular the second assumption is fulfilled by the prior defined by Eq. (10) if  $0 < \gamma \leq 3/2$  and any  $\sigma^2 > 0$ .

**Assumption 7** The prior on the truncation level satisfies for some positive constants  $c_1, c_2$ ,

$$P(R > r) \leq \exp(-c_1 2^r r),$$

$$P(R = r) \geq \exp(-c_2 2^r r). \tag{17}$$

For the prior on the scaling we assume existence of constants  $0 < p_1 < p_2, q > 0$  and  $C > 1$  with  $p_1 > q|\alpha - \beta|$  such that

$$P(S \in [x^{p_1}, x^{p_2}]) \geq \exp(-x^q) \quad \text{for all } x \geq C. \tag{18}$$

The prior on  $R$  can be defined as  $R = \lceil 2 \log Y \rceil$ , where  $Y$  is Poisson distributed. Equation (18) is satisfied for a whole range of distributions, including the popular family of inverse gamma distributions. Since the inverse gamma prior on  $S^2$  decays polynomially (Lemma 17), condition (A2) of Shen and Ghosal (2015) is not satisfied and hence their posterior contraction results cannot be applied to our prior. We obtain the following result for our prior.

**Theorem 8** Assume  $b_0$  satisfies Assumption 4. Suppose the prior satisfies assumptions 6 and 7. Let  $\{\varepsilon_n\}_{n=1}^\infty$  be a sequence of positive numbers that converges to zero. There is a constant  $C_1 > 0$  such that for any  $C_2 > 0$  there is a measurable set  $\mathcal{B}_n \subseteq L^2(\mathbb{T})$  such that for every  $a > 0$  there is a positive constant  $C_3$  such that for  $n$  sufficiently large

$$\log P\left(\|b^{R,S} - b_0\|_\infty < \varepsilon_n\right) \geq -C_1 \varepsilon_n^{-1/\beta} |\log \varepsilon_n| \tag{19}$$

$$\log P\left(b^{R,S} \notin \mathcal{B}_n\right) \leq -C_2 \varepsilon_n^{-1/\beta} |\log \varepsilon_n| \tag{20}$$

$$\log N(a\varepsilon, \{b \in \mathcal{B}_n : \|b - b_0\|_2 \leq \varepsilon_n\}, \|\cdot\|_\infty) \leq C_3 \varepsilon_n^{-1/\beta} |\log \varepsilon_n|. \tag{21}$$

The following theorem is obtained by applying these bounds to Theorem 3 after taking  $\varepsilon_n = (T/\log T)^{-\beta/(1+2\beta)}$ .

**Theorem 9** Assume  $b_0$  satisfies Assumption 4. Suppose the prior satisfies assumptions 6 and 7. Then for all  $M_T \rightarrow \infty$

$$P_{b_0} \Pi^n \left( b : \|b - b_0\|_2 \geq M_T \left( \frac{T}{\log T} \right)^{-\frac{\beta}{1+2\beta}} \mid X^T \right) \rightarrow 0$$

as  $T \rightarrow \infty$ .

This means that when the true parameter is from  $B_{\infty,\infty}^\beta[0, 1], \beta < 2$  a rate is obtained that is optimal possibly up to a log factor. When  $\beta \geq 2$  then  $b_0$  is in particular in the space  $B_{\infty,\infty}^{2-\delta}[0, 1]$ , for every small positive  $\delta$ , and therefore converges with rate essentially  $T^{-2/5}$ .

When a different function  $\Lambda$  is used, defined on a compact interval of  $\mathbb{R}$ , and the basis elements are defined by  $\psi_{jk} = \sum_{m \in \mathbb{Z}} \Lambda(2^j(x - m) + k - 1)$ ; forcing them to be 1-periodic. Then Theorem 9 and derived results for applications still holds provided  $\|\psi_{jk}\|_\infty = 1$  and

$\psi_{j,k} \cdot \psi_{j,l} \equiv 0$  when  $|k - l| \geq d$  for a fixed  $d \in \mathbb{N}$  and the smoothness assumptions on  $b_0$  are changed accordingly. A finite number of basis elements can be added or redefined as long as they are 1-periodic.

It is easy to see that our results imply posterior convergences rates in weaker  $L^p$ -norms,  $1 \leq p < 2$ , with the same rate. When  $p \in (2, \infty]$  the  $L^p$ -norm is stronger than the  $L^2$ -norm. We apply ideas of [Knapik and Salomond \(2014\)](#) to obtain rates for stronger  $L^p$ -norms.

**Theorem 10** *Assume the true drift  $b_0$  satisfies assumption 4. Suppose the prior satisfies assumptions 6 and 7. Let  $p \in (2, \infty]$ . Then for all  $M_T \rightarrow \infty$*

$$P_{b_0} \Pi^n \left( b : \|b - b_0\|_p \geq M_T T^{-\frac{\beta-1/2+1/p}{1+2\beta}} (\log T)^{\frac{2\beta-2\beta/p}{1+2\beta}} \mid X^T \right) \rightarrow 0$$

as  $T \rightarrow \infty$ .

These rates are similar to the rates obtained for the density estimation in [Giné and Nickl \(2011\)](#). However our proof is less involved. Note that we have only consistency for  $\beta > 1/2 - 1/p$ .

## 5 Applications to nonparametric regression and density estimation

Our general results also apply to other models. The following results are obtained for  $b_0$  satisfying Assumption 4 and the prior satisfying assumptions 6 and 7.

### 5.1 Nonparametric regression model

As a direct application of the properties of the prior shown in the previous section, we obtain the following result for a nonparametric regression problem. Assume

$$X_i^n = b_0(i/n) + \eta_i, \quad 0 \leq i \leq n, \quad (22)$$

with independent Gaussian observation errors  $\eta_i \sim N(0, \sigma^2)$ . When we apply [Ghosal and van der Vaart \(2007\)](#), example 7.7 to Theorem 8 we obtain, for every  $M_n \rightarrow \infty$ ,

$$\Pi \left( b : \|b - b_0\|_2 \geq M_n \left( \frac{n}{\log n} \right)^{-\frac{\beta}{1+2\beta}} \mid X^n \right) \xrightarrow{P^{b_0}} 0$$

as  $n \rightarrow \infty$  and (in a similar way as in Theorem 10) for every  $p \in (2, \infty]$ ,

$$\Pi \left( b : \|b - b_0\|_2 \geq M_n n^{-\frac{\beta-1/2+1/p}{1+2\beta}} (\log n)^{\frac{2\beta-2\beta/p}{1+2\beta}} \mid X^n \right) \xrightarrow{P^{b_0}} 0$$

as  $n \rightarrow \infty$ .

### 5.2 Density estimation

Let us consider  $n$  independent observations  $X^n := (X_1, \dots, X_n)$  with  $X_i \sim p_0$  where  $p_0$  is an unknown density on  $[0, 1]$  relative to the Lebesgue measure. Let  $\mathcal{P}$  denote the space of densities on  $[0, 1]$  relative to the Lebesgue measure. The natural distance for densities is the Hellinger distance  $h$  defined by

$$h(p, q)^2 = \int_0^1 \left( \sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx.$$

Define the prior on  $\mathcal{P}$  by  $p = \frac{e^b}{\|e^b\|_1}$ , where  $b$  is endowed with the prior of Theorem 9 or its non-periodic version. Assume that  $\log p_0$  is  $\beta$ -smooth in the sense of Assumption 4. Applying Ghosal et al. (2000), theorem 2.1 and van der Vaart and van Zanten (2008), lemma 3.1 to Theorem 8, we obtain for a big enough constant  $M > 0$

$$\Pi \left( p \in \mathcal{P} : h(p, p_0) \geq M \left( \frac{n}{\log n} \right)^{-\frac{\beta}{1+2\beta}} \mid X^n \right) \xrightarrow{P_0} 0,$$

as  $n \rightarrow \infty$ .

## 6 Proofs

### 6.1 Proof of lemma 1

Since conditions (ND) and (LI) of (Karatzas and Shreve 1991, theorem 5.15) hold, the SDE Eq. (1) has a unique weak solution up to an explosion time.

Assume without loss of generality that  $X_0 = 0$ . Define  $\tau_0 = 0$  and for  $i \geq 1$  the random times

$$\tau_i = \inf\{t \geq \tau_{i-1} : |X_t - X_{\tau_{i-1}}| = 1\}.$$

By periodicity of drift and the Markov property the random variables  $U_i = \tau_i - \tau_{i-1}$  are independent and identically distributed.

Note that

$$\inf\{t : X_t = \pm n\} \geq \sum_{i=1}^n U_i$$

and hence non-explosion follows from  $\lim_{n \rightarrow \infty} \sum_{i=1}^n U_i = \infty$  almost surely. The latter holds true since  $U_1 > 0$  with positive probability, which is clear from the continuity of diffusion paths.

### 6.2 Proof of lemma 2

*Proof of the first part.* For the proof we introduce some notation: for any  $(j, k), (j', k')$  we write  $(j, k) < (j', k')$  if  $\text{supp } \psi_{j',k'} \subset \text{supp } \psi_{j,k}$ . The set of indices become a lattice with partial order  $<$ , and by  $(j, k) \vee (j', k')$  we denote the supremum. Identify  $i$  with  $(j, k)$  and similarly  $i'$  with  $(j', k')$ .

For  $i > 1$ , denote by  $t_i$  the time points in  $[0, 1]$  corresponding to the maxima of  $\psi_i$ . Without loss of generality assume  $t_i < t_{i'}$ . We have  $G_{i,i'} = 0$  if and only if the interiors of the supports of  $\psi_i$  and  $\psi_{i'}$  are disjoint. In that case

$$\max \text{supp } \psi_{j,k} \leq t_{(j,k) \vee (j',k')} \leq \min \text{supp } \psi_{j',k'}. \tag{23}$$

□

The values of  $Z_i$  can be found by the midpoint displacement technique. The coefficients are given by  $Z_1 = V_0, Z_2 = V_{\frac{1}{2}}$  and for  $j \geq 1$

$$Z_{j,k} = V_{2^{-j(k-1/2)}} - \frac{1}{2} (V_{2^{-j(k-1)}} + V_{2^{-jk}}).$$

As  $V$  is a Gaussian process, the vector  $Z$  is mean-zero Gaussian, say with (infinite) precision matrix  $\Gamma$ . Now  $\Gamma_{i,i'} = 0$  if there exists a set  $\mathcal{L} \subset \mathbb{N}$  such that  $\mathcal{L} \cap \{i, i'\} = \emptyset$  for which conditional on  $\{Z_{i^*}, i^* \in \mathcal{L}\}$ ,  $Z_i$  and  $Z_{i'}$  are independent.

Define  $(j^*, k^*) = (j, k) \vee (j', k')$  and

$$\mathcal{L} = \{i^* \in \mathbb{N} : i^* = 2^j + k, \text{ with } j \leq j^*\}.$$

The set  $\{Z_{i^*}, i^* \in \mathcal{L}\}$  determine the process  $V$  at all times  $k2^{-j^*-1}, k = 0 \dots, 2^{j^*+1}$ . Now  $Z_i$  and  $Z_{i'}$  are conditionally independent given  $\{V_t, t = k2^{-j^*-1}, k = 0 \dots, 2^{j^*+1}\}$  by (23) and the Markov property of the nonperiodic Ornstein–Uhlenbeck process. The result follows since  $\sigma(\{Z_{i^*}, i^* \in \mathcal{L}\}) = \sigma(\{V_t, t = k2^{-j^*-1}, k = 0 \dots, 2^{j^*+1}\})$ .

**Lemma 11** Let  $K(s, t) = \mathbb{E}V_s V_t = \frac{\sigma^2}{2\gamma} \frac{1}{1-e^{-\gamma}} (e^{-\gamma|t-s|} + e^{-\gamma(1-|t-s|)})$ . If  $x \notin (s, t)$

$$\frac{1}{2}K(s, x) - K\left(\frac{s+t}{2}, x\right) + \frac{1}{2}K(t, x) = 2 \sinh^2\left(\gamma \frac{t-s}{4}\right) K\left(\frac{t+s}{2}, x\right)$$

*Proof* Without loss of generality assume that  $t \leq x \leq 1$ . With  $m = (t + s)/2$  and  $\delta = (t - s)/2$

$$\begin{aligned} & \left( e^{-\gamma|s-x|} + e^{-\gamma(1-|s-x|)} \right) - 2 \left( e^{-\gamma|m-x|} + e^{-\gamma(1-|m-x|)} \right) + \left( e^{-\gamma|t-x|} + e^{-\gamma(1-|t-x|)} \right) \\ &= e^{-\gamma|t-x|} e^{-2\gamma\delta} - 2e^{-\gamma|t-x|} e^{-\gamma\delta} + e^{-\gamma|t-x|} + e^{-\gamma(1-|s-x|)} - 2e^{-\gamma(1-|s-x|)} e^{-\gamma\delta} \\ & \quad + e^{-\gamma(1-|s-x|)} e^{-2\gamma\delta} = (1 - e^{-\gamma\delta})^2 (e^{-\gamma|t-x|} + e^{-\gamma(1-|s-x|)}) \\ &= (1 - e^{-\gamma\delta})^2 e^{\gamma\delta} \left( e^{-\gamma|m-x|} + e^{-\gamma(1-|m-x|)} \right) \end{aligned}$$

The result follows from  $(1 - e^{-\gamma\delta})^2 e^{\gamma\delta} = 4 \sinh^2(\gamma\delta/2)$  and scaling both sides with  $\frac{1}{2} \frac{\sigma^2}{2\gamma} \frac{1}{1-e^{-\gamma}}$ . □

*Proof of the second part* Denote by  $[a, b], [c, d]$  the support of  $\psi_i$  and  $\psi_{i'}$  respectively and let  $m = (b + a)/2$  and  $n = (d + c)/2$  but for  $i = 1$ , let  $m = 0$ .  $Z_1 = V(0)$ ,  $Z_2 = V_{1/2}$  and  $\text{Var}(Z_1) = \text{Var}(Z_2) = \frac{\sigma^2}{2\gamma} \coth(\gamma/2)$ , and  $\text{Cov}(Z_1, Z_2) = \frac{\sigma^2}{2\gamma} \sinh^{-1}(\gamma/2)$ . Note that the  $2 \times 2$  covariance matrix of  $Z_1$  and  $Z_2$  has eigenvalues  $\frac{\sigma^2}{2\gamma} \tanh(\gamma/4)$  and  $\frac{\sigma^2}{2\gamma} \coth(\gamma/4)$  and is strictly positive definite. □

By midpoint displacement,  $2Z_i = 2V_m - V_a - V_b, i > 2$  and  $K(s, t) = \mathbb{E}V_s V_t = \frac{\sigma^2}{2\gamma} \frac{1}{1-e^{-\gamma}} (e^{-\gamma|t-s|} + e^{-\gamma(1-|t-s|)})$ .

Assume without loss of generality  $b - a \geq d - c$ . Define  $\delta$  to be the halfwidth of the smaller interval, so that  $\delta := (d - c)/2 = 2^{-j'-1}$ . Then

$$(b - a)/2 = 2^{-j-1} = h\delta, \quad \text{with } h = 2^{j'-j}.$$

Consider three cases:

1. The entries on diagonal,  $i = i'$ ;
2. The interiors of the supports of  $\psi_i$  and  $\psi_{i'}$  are non-overlapping;
3. The support of  $\psi_{i'}$  is contained in the support of  $\psi_i$ .

*Case 1.* By elementary computations for  $i > 2$ ,

$$\begin{aligned} 4 \frac{2\gamma}{\sigma^2} (1 - e^{-\gamma}) A_{ii} &= 6(1 + e^{-\gamma}) + 2(e^{-\gamma 2\delta} + e^{-\gamma(1-2\delta)}) - 8(e^{-\gamma\delta} + e^{-\gamma(1-\delta)}) \\ &= 2(1 - e^{-\gamma\delta})(3 - e^{-\gamma\delta}) + 2e^{-\gamma} (1 - e^{\gamma\delta})(3 - e^{\gamma\delta}). \end{aligned}$$

As  $\delta \leq \frac{1}{4}$  and under the assumption  $\gamma \leq 3/2$  the last display can be bounded by

$$0.9715 \cdot 4\gamma\delta(1 - e^{-\gamma}) \leq 4\frac{2\gamma}{\sigma^2}(1 - e^{-\gamma})A_{ii} \leq 4\gamma\delta(1 - e^{-\gamma}).$$

Hence  $0.9715 \cdot 2^{-j}\sigma^2/4 \leq A_{ii} \leq 2^{-j}\sigma^2/4$ .

Case 2. Necessarily  $i, i' > 2$ . By twofold application of lemma 11

$$\begin{aligned} A_{ij} &= (K(c, b) - 2K(n, b) + K(d, b))/4 \\ &\quad - 2(K(c, m) - 2K(n, m) + K(d, m))/4 \\ &\quad + (K(c, a) - 2K(n, a) + K(d, a))/4 \\ &= 2 \sinh^2(\gamma \frac{d-c}{4})(K(n, b) - 2K(n, m) + K(n, a))/2 \\ &= 4 \sinh^2(\gamma \frac{b-a}{4}) \sinh^2(\gamma \frac{d-c}{4})K(n, m). \end{aligned} \tag{24}$$

Using the convexity of  $\sinh$  we obtain the bound

$$2 \sinh^2(x/2) \leq 0.55x^2 \tag{25}$$

for  $0 \leq x \leq 1$ . Note that  $f(x) = e^{-\gamma x} + e^{-\gamma(1-x)}$  is convex on  $[0, 1]$ , from which we derive  $f(x) \leq 1 + e^{-\gamma}$ . Using this bound, and the fact that for  $\gamma \leq 3/2$ ,

$$\gamma^2 K(n, m) \leq \frac{\sigma^2}{2} \gamma \coth(\gamma/2) \leq \sigma^2(1 + \gamma/2), \tag{26}$$

which can be easily seen from a plot, that

$$\begin{aligned} |A_{ii'}| &\leq 0.55^2 \gamma^4 \cdot 2^{-2j-2} \cdot 2^{-2j'-2} |K(n, m)| \\ &\leq 0.0095 \sigma^2 \gamma^2 (1 + \gamma/2) 2^{-1.5(j+j')}. \end{aligned}$$

Case 3.

For  $i' > 2, i = 1$  with  $m = 0$  or  $i = 2$  with  $m = \frac{1}{2}$ , using Eq. (26), we obtain

$$\begin{aligned} |A_{ii'}| &= |K(m, n) - \frac{1}{2}K(m, c) - \frac{1}{2}K(m, d)| \\ &\leq 2 \sinh^2(\gamma \frac{d-c}{4})K(m, n) \\ &\leq 0.55\gamma^2 2^{-2j'-2} K(m, n) \\ &\leq 0.098\sigma^2(1 + \gamma/2)2^{-1.5j}. \end{aligned} \tag{27}$$

When  $i, i' > 2$  then, using the calculation Eq. (24) and Lemma 11 noting that  $a, b$  and  $m$  are not in  $(c, d)$ , we obtain

$$A_{ii'} = 2 \sinh^2(\gamma \frac{d-c}{4})(K(n, b) - 2K(n, m) + K(n, a))/2.$$

Write  $x = \gamma|a - m| = \gamma|b - m| = \gamma h\delta$  and  $\alpha = \frac{|m-n|}{|b-m|} \in (0, 1)$ . A simple computation then shows

$$e^{-\gamma|b-n|} - 2e^{-\gamma|m-n|} + e^{-\gamma|a-n|} = e^{-(1+\alpha)x} - 2e^{-\alpha x} + e^{-(1-\alpha)x}.$$

The derivative of  $f(\alpha) := e^{-(1+\alpha)x} - 2e^{-\alpha x} + e^{-(1-\alpha)x}$  is nonnegative, for  $\alpha, x > 0$  hence  $f(\alpha)$  is increasing and so  $f(0) \leq f(\alpha) \leq f(1)$ . Note that  $f(0) = 2e^{-x} - 2 \geq -2x$ , for  $x > 0$  and  $f(1) = e^{-2x} - 2e^{-x} + 1 =: g(x)$ . Maximising  $g'(x)$  over  $x > 0$  gives  $g'(x) \leq 1/2$  and  $g(0) = 0$  and therefore  $f(1) = g(x) \leq x/2$ .

It follows that

$$-2\gamma h\delta \leq e^{-\gamma|b-n|} - 2e^{-\gamma|m-n|} + e^{-\gamma|a-n|} \leq \gamma h\delta/2.$$

For the other terms we derive the following bounds. Write

$$\begin{aligned} & e^{-\gamma(1-|b-n|)} - 2e^{-\gamma(1-|m-n|)} + e^{-\gamma(1-|a-n|)} \\ &= e^{-\gamma+(1+\alpha)x} - 2e^{-\gamma+\alpha x} + e^{-\gamma+(1-\alpha)x} =: h(\alpha). \end{aligned}$$

Now  $h(\alpha)$  is decreasing for  $x \leq \log 2$  and convex and positive for  $x \geq \log 2$ . In both case we can bound  $h(\alpha)$  by its value at the endpoints  $\alpha = 0$  and  $\alpha = 1$ . Using that  $2x \leq \gamma$  we obtain  $0 \leq h(0) = e^{-\gamma}(2e^x - 2) \leq 2x$  and  $0 \leq h(1) = e^{-\gamma}(e^{2x} - 2e^x + 1) \leq 2x$ . So  $0 \leq h(\alpha) \leq 2\gamma h\delta$ .

Using the bound Eq. (25) and  $x/(1 - \exp(-x)) \leq (1 + x)$  we obtain

$$|A_{ii'}| \leq 0.061\sigma^2\gamma(1 + \gamma)2^{-1.5(j+j')}.$$

### 6.3 Proof of theorem 3

A general result for deriving contraction rates for Brownian semi-martingale models was proved in [van der Meulen et al. \(2006\)](#). Theorem 3 follows upon verifying the assumptions of this result for the diffusion on the circle. These assumptions are easily seen to boil down to:

1. For every  $T > 0$  and  $b_1, b_2 \in L^2(\mathbb{T})$  the measures  $P^{b_1, T}$  and  $P^{b_2, T}$  are equivalent.
2. The posterior as defined in equation Eq. (5) is well defined.
3. Define the (random) *Hellinger semimetric*  $h_T$  on  $L^2(\mathbb{T})$  by

$$h_T^2(b_1, b_2) := \int_0^T (b_1 - b_2)^2(X_t) dt, \quad b_1, b_2 \in L^2(\mathbb{T}). \tag{28}$$

There are constants  $0 < c < C$  for which

$$\lim_{T \rightarrow \infty} P^{\theta_0, T} \left( c\sqrt{T}\|b_1 - b_2\|_2 \leq h_T(b_1, b_2) \leq C\sqrt{T}\|b_1 - b_2\|_2, \forall b_1, b_2 \in L^2(\mathbb{T}) \right) = 1.$$

We start by verifying the third condition. Recall that the local time of the process  $X^T$  is defined as the random process  $L_T(x)$  which satisfies

$$\int_0^T f(X_t) dt = \int_{\mathbb{R}} f(x)L_T(x) dx.$$

For every measurable function  $f$  for which the above integrals are defined. Since we are working with 1-periodic functions, we define the periodic local time by

$$\mathring{L}_T(x) = \sum_{k \in \mathbb{Z}} L_T(x + k).$$

Note that  $t \mapsto X_t$  is continuous with probability one. Hence the support of  $t \mapsto X_t$  is compact with probability one. Since  $x \mapsto L_T(x)$  is only positive on the support of  $t \mapsto X_t$ , it follows that the sum in the definition of  $\mathring{L}_T(x)$  has only finitely many nonzero terms and is therefore well defined. For a one-periodic function  $f$  we have

$$\int_0^T f(X_t) dt = \int_0^1 f(x)\mathring{L}_T(x) dx,$$

provided the involved integrals exists. It follows from (Schauer and van Zanten 2017, Theorem 5.3) that  $\dot{L}_T(x)/T$  converges to a positive deterministic function only depending only on  $b_0$  and which is bounded away from zero and infinity. Since the Hellinger distance can be written as

$$h_T(b_1, b_2) = \sqrt{T} \sqrt{\int_0^1 (b_1(x) - b_2(x))^2 \frac{\dot{L}_T(x)}{T} dt}$$

it follows that the third assumption is satisfied with  $d_T(b_1, b_2) = \sqrt{T} \|b_1 - b_2\|_2$ .

Conditions 1 and 2 now follow by arguing precisely as in lemmas A.2 and 3.1 of van Waaij and van Zanten (2016) respectively (the key observation being that the convergence result of  $\dot{L}_T(x)/T$  also holds when  $\int_0^1 b(x) dx$  is nonzero, which is assumed in that paper).

The stated result follows from Theorem 2.1 in van der Meulen et al. (2006) (taking  $\mu_T = \sqrt{T} \varepsilon_T$  in their paper).

### 6.4 Proof of theorem 8 with Assumption 6 (A)

The proof proceeds by verifying the conditions of theorem 3. By Assumption 4 the true drift can be represented as  $b_0 = z_1 \psi_1 + \sum_{j=0}^\infty \sum_{k=1}^{2^j} z_{jk} \psi_{jk}$ . For  $r \geq 0$ , define its truncated version by

$$b_0^r = z_1 \psi_1 + \sum_{j=0}^r \sum_{k=1}^{2^j} z_{jk} \psi_{jk}.$$

#### 6.4.1 Small ball probability

For  $\varepsilon > 0$  choose an integer  $r_\varepsilon$  with

$$C_\beta \varepsilon^{-1/\beta} \leq 2^{r_\varepsilon} \leq 2C_\beta \varepsilon^{-1/\beta} \quad \text{where} \quad C_\beta = \frac{\|b_0\|_\beta^{1/\beta}}{(2^\beta - 1)^{1/\beta}}. \tag{29}$$

For notational convenience we will write  $r$  instead of  $r_\varepsilon$  in the remainder of the proof. By lemma 16 we have  $\|b_0^r - b_0\|_\infty \leq \varepsilon$ . Therefore

$$\|b^{r,S} - b_0\|_2 \leq \|b^{r,S} - b_0^r\|_2 + \|b_0^r - b_0\|_2 \leq \|b^{r,S} - b_0^r\|_\infty + \varepsilon$$

which implies

$$P(\|b^{r,S} - b_0\|_2 < 2\varepsilon) \geq P(\|b^{r,S} - b_0^r\|_\infty < \varepsilon).$$

Let  $f_S$  denotes the probability density of  $S$ . For any  $x > 0$ , we have

$$\begin{aligned} P(\|b^{R,S} - b_0\|_2 < 2\varepsilon) &= \sum_{r \geq 1} P(R = r) \int_0^\infty P(\|b^{r,S} - b_0\|_2 < 2\varepsilon) f_S(s) ds \\ &\geq P(R = r) \inf_{s \in [L_\varepsilon, U_\varepsilon]} P(\|b^{r,S} - b_0^r\|_\infty < \varepsilon) \int_{L_\varepsilon}^{U_\varepsilon} f_S(s) ds, \tag{30} \end{aligned}$$

where

$$L_\varepsilon = \varepsilon^{-\frac{p_1}{q\beta}} \quad \text{and} \quad U_\varepsilon = \varepsilon^{-\frac{p_2}{q\beta}}$$



and  $p_1, p_2$  and  $q$  are taken from Assumption 7. For  $\varepsilon$  sufficiently small, we have by the second part of Assumption 7

$$\int_{L_\varepsilon}^{U_\varepsilon} f_S(s) ds \geq \exp\left(-\varepsilon^{-\frac{1}{\beta}}\right)$$

By choice of  $r$  and the first part of Assumption 7, there exists a positive constant  $C$  such that

$$P(R = r) \geq \exp\left(-c_2 2^r r\right) \geq \exp\left(-C\varepsilon^{-\frac{1}{\beta}}|\log \varepsilon|\right),$$

for  $\varepsilon$  sufficiently small.

For lower bounding the middle term in Eq. (30), we write

$$b^{r,s} - b'_0 = (sZ_1 - z_1)\psi_1 + \sum_{j=0}^r \sum_{k=1}^{2^j} (sZ_{jk} - z_{jk})\psi_{jk}$$

which implies

$$\|b^{r,s} - b'_0\|_\infty \leq |sZ_1 - z_1| + \sum_{j=0}^r \max_{1 \leq k \leq 2^j} |sZ_{jk} - z_{jk}| \leq (r + 2) \max_{i \in \mathcal{J}_r} |sZ_i - z_i|.$$

This gives the bound

$$P(\|b^{r,s} - b'_0\|_\infty < \varepsilon) \geq \prod_{i \in \mathcal{J}_r} P\left(|sZ_i - z_i| < \frac{\varepsilon}{r + 2}\right).$$

By choice of the  $Z_i$ , we have for all  $i \in \{1, 2, \dots\}$ ,  $2^{\alpha\ell(i)} Z_i$  is standard normally distributed and hence

$$\begin{aligned} \log P\left(|sZ_i - z_i| < \frac{\varepsilon}{r + 2}\right) &= \log P\left(\left|2^{\alpha\ell(i)} Z_i - 2^{\alpha\ell(i)} z_i/s\right| < \frac{2^{\alpha\ell(i)} \varepsilon}{(r + 2)s}\right) \\ &\geq \log\left(\frac{2^{\alpha\ell(i)} \varepsilon}{(r + 2)s}\right) - \frac{2^{2\alpha\ell(i)} \varepsilon^2}{(r + 2)^2 s^2} - \frac{2^{2\alpha\ell(i)} z_i^2}{s^2} + \frac{1}{2} \log\left(\frac{2}{\pi}\right), \end{aligned}$$

where the inequality follows from lemma 18. The third term can be further bounded as we have

$$2^{2\alpha\ell(i)} z_i^2 = 2^{2(\alpha-\beta)\ell(i)} 2^{2\beta\ell(i)} z_i^2 \leq 2^{2(\alpha-\beta)\ell(i)} \llbracket b_0 \rrbracket_\beta^2.$$

Hence

$$\begin{aligned} \log P\left(|sZ_i - z_i| < \frac{\varepsilon}{r + 2}\right) &\geq \log\left(\frac{2^{\alpha\ell(i)} \varepsilon}{(r + 2)s}\right) - \frac{2^{2\alpha\ell(i)} \varepsilon^2}{(r + 2)^2 s^2} \\ &\quad - \frac{2^{2(\alpha-\beta)\ell(i)} \llbracket b_0 \rrbracket_\beta^2}{s^2} + \frac{1}{2} \log\left(\frac{2}{\pi}\right). \end{aligned} \tag{31}$$

For  $s \in [L_\varepsilon, U_\varepsilon]$  and  $i \in \mathcal{J}_r$  we will now derive bounds on the first three terms on the right of Eq. (31). For  $\varepsilon$  sufficiently small we have  $r \leq r + 2 \leq 2r$  and then inequality (29) implies

$$\log C_\beta \leq r + 2 \leq 2 \log(4C_\beta) + \frac{2}{\beta} |\log \varepsilon|.$$

*Bounding the first term on the RHS of (31).* For  $\varepsilon$  sufficiently small, we have

$$\begin{aligned} \log \left( \frac{(r+2)s}{2^{\alpha \ell(i)} \varepsilon} \right) &\leq \log \left( \frac{(r+2)U_\varepsilon}{\varepsilon} \right) = \log \left( (r+2)\varepsilon^{-\left(1+\frac{p_2}{q\beta}\right)} \right) \\ &\leq \log \left\{ 2 \log(4C_\beta) + \frac{2}{\beta} |\log \varepsilon| \right\} + \left( 1 + \frac{p_2}{q\beta} \right) |\log \varepsilon| \\ &\leq \tilde{C}_{p_2,q,\beta} |\log \varepsilon|, \end{aligned}$$

where  $\tilde{C}_{p_2,q,\beta}$  is a positive constant.

*Bounding the second term on the RHS of (31).* For  $\varepsilon$  sufficiently small, we have

$$\frac{2^{2\alpha \ell(i)} \varepsilon^2}{(r+2)^2 s^2} \leq \frac{2^{2\alpha r} \varepsilon^2}{(\log C_\beta)^2 L_\varepsilon^2} \leq \frac{(2C_\beta)^{2\alpha}}{(\log C_\beta)^2} \varepsilon^{\frac{2}{\beta}(-\alpha+\beta+p_1/q)} \leq 1.$$

The final inequality is immediate in case  $\alpha = \beta$ , else it suffices to verify that the exponent is non-negative under the assumption  $p_1 > q|\alpha - \beta|$ .

*Bounding the third term on the RHS of (31).* For  $\varepsilon$  sufficiently small, in case  $\beta \geq \alpha$  we have

$$\frac{2^{2(\alpha-\beta)\ell(i)} \llbracket b_0 \rrbracket_\beta^2}{s^2} \leq \llbracket b_0 \rrbracket_\beta^2 L_\varepsilon^{-2} \leq 1.$$

In case  $\beta < \alpha$  we have

$$\frac{2^{2(\alpha-\beta)\ell(i)} \llbracket b_0 \rrbracket_\beta^2}{s^2} \leq \frac{2^{2(\alpha-\beta)r} \llbracket b_0 \rrbracket_\beta^2}{L_\varepsilon^2} \leq (2C_\beta)^{2(\alpha-\beta)} \varepsilon^{\frac{2}{\beta}(p_1/q-\alpha+\beta)} \leq 1$$

as the exponent of  $\varepsilon$  is positive under the assumption  $p_1 > q|\alpha - \beta|$ .

Hence for  $\varepsilon$  small enough, we have

$$\log P \left( |sZ_i - z_i| < \frac{\varepsilon}{r+2} \right) \geq -\tilde{C}_{p_2,q,\beta} |\log \varepsilon| - 3.$$

As  $-2^{r+1} \geq -4C_\beta \varepsilon^{-1/\beta}$  we get

$$\begin{aligned} \log \inf_{s \in [x^{p_1}, x^{p_2}]} P \left( \|b^{r,s} - b_0^r\|_\infty < \varepsilon \right) &\geq -4C_\beta \varepsilon^{-1/\beta} \left( \tilde{C}_{p_2,q,\beta} |\log \varepsilon| + 3 \right) \\ &\gtrsim -\varepsilon^{-1/\beta} |\log \varepsilon|. \end{aligned}$$

We conclude that the right hand side of Eq. (30) is bounded below by  $\exp(-C_1 \varepsilon^{-1/\beta} |\log \varepsilon|)$ , for some positive constant  $C_1$  and sufficiently small  $\varepsilon$ .

### 6.4.2 Entropy and remaining mass conditions

For  $r \in \{0, 1, \dots\}$  denote by  $\mathcal{C}_r$  the linear space spanned by  $\psi_1$  and  $\psi_{jk}$ ,  $0 \leq j \leq r$ ,  $k \in 1, \dots, 2^j$ , and define

$$\mathcal{C}_{r,t} := \{b \in \mathcal{C}_r, \llbracket b \rrbracket_\alpha \leq t\}.$$

**Proposition 12** For any  $\varepsilon > 0$

$$\log N(\varepsilon, \mathcal{C}_{r,t}, \|\cdot\|_\infty) \leq 2^{r+1} \log(3A_\alpha t \varepsilon^{-1}),$$

where  $A_\alpha = \sum_{k=0}^\infty 2^{-k\alpha}$ .

*Proof* We follow (van der Meulen et al. 2006, §3.2.2). Choose  $\varepsilon_0, \dots, \varepsilon_r > 0$  such that  $\sum_{j=0}^r \varepsilon_j \leq \varepsilon$ . Define

$$U_j = \begin{cases} [-2^{-\alpha j}t, 2^{-\alpha j}t]^{2^j} & \text{if } j \in \{1, \dots, r\} \\ [-t, t]^2 & \text{if } j = 0 \end{cases}.$$

For each  $j \in \{1, \dots, r\}$ , let  $E_j$  be a minimal  $\varepsilon_j$ -net with respect to the max-distance on  $\mathbb{R}^{2^j}$  and let  $E_0$  be a minimal  $\varepsilon_0$ -net with respect to the max-distance on  $\mathbb{R}^2$ . Hence, if  $x \in U_j$ , then there exists a  $e_j \in E_j$  such that  $\max_k |x_k - e_k| \leq \varepsilon_j$ .

Take  $b \in \mathcal{C}_{r,t}$  arbitrary:  $b = z_1 \psi_1 + \sum_{j=0}^r \sum_{k=1}^{2^j} z_{jk} \psi_{jk}$ . Let  $\tilde{b} = e_1 \psi_1 + \sum_{j=0}^r \sum_{k=1}^{2^j} e_{jk} \psi_{jk}$ , where  $(e_1, e_{0,1}) \in E_0$  and  $(e_{j,1}, \dots, e_{j,2^j}) \in E_j$  (for  $j = 1, \dots, r$ ). We have

$$\begin{aligned} \|b - \tilde{b}\|_\infty &\leq |z_1 - e_1| \|\psi_1\|_\infty + \sum_{j=0}^r \max_{1 \leq k \leq 2^j} |z_{jk} - e_{jk}| \|\psi_{jk}\|_\infty \\ &\leq |z_1 - e_1| + \sum_{j=0}^r \max_{1 \leq k \leq 2^j} 2^{j\alpha} |2^{-j\alpha} z_{jk} - 2^{-j\alpha} e_{jk}|. \end{aligned}$$

This can be bounded by  $\sum_{j=0}^r \varepsilon_j$  by an appropriate choice of the coefficients in  $\tilde{b}$ . In that case we obtain that  $\|b - \tilde{b}\|_\infty \leq \varepsilon$ . This implies

$$\log N(\varepsilon, \mathcal{C}_{r,t}, \|\cdot\|_\infty) \leq \sum_{j=0}^r \log |E_j| \leq \sum_{j=0}^r 2^j \log \left( \frac{3 \cdot 2^{-\alpha j} t}{\varepsilon_j} \right).$$

The asserted bound now follows upon choosing  $\varepsilon_j = \varepsilon 2^{-j\alpha} / A_\alpha$ . □

**Proposition 13** *There exists a constant a positive constant  $K$  such that*

$$\log N(a\varepsilon, \{b \in \mathcal{C}_r : \|b - b_0\|_2 \leq \varepsilon\}, \|\cdot\|_2) \leq 2^{r+1} \log(6A_\alpha K 2^{ar}).$$

*Proof* There exists a positive  $K$  such that

$$\{b \in \mathcal{C}_r : \|b - b_0\|_2 \leq a\varepsilon\} \subset \{b \in \mathcal{C}_r : \|b\|_2 \leq K\}.$$

By lemma 21, this set is included in the set

$$\left\{ b \in \mathcal{C}_r : \|b\|_\infty \leq \sqrt{32}^{(r+1)/2} K \right\}. \tag{32}$$

By lemma 20, for any  $b = z_1 \psi_1 + \sum_{j=0}^r \sum_{k=1}^{2^j} z_{jk} \psi_{jk}$  in this set we have

$$\max \left\{ |z_1|, |z_{jk}|, j = 0, \dots, r, k = 1, \dots, 2^j \right\} \leq 2 \|b\|_\infty \sqrt{32}^{(r+1)/2} K.$$

Hence, the set Eq. (32) is included in the set  $\{b \in \mathcal{C}_r : \|b\|_\alpha \leq a(r, \varepsilon)\} = \mathcal{C}_{r,a(r,\varepsilon)}$ , where  $a(r, \varepsilon) = 2^{1+ar} \sqrt{32}^{(r+1)/2} K$ .

Hence,

$$N(a\varepsilon, \{b \in \mathcal{C}_r : \|b - b_0\|_2 \leq \varepsilon\}, \|\cdot\|_2) \leq N(\varepsilon, \mathcal{C}_{r,a(r,\varepsilon)}, \|\cdot\|_2).$$

Using Lemma 21 again the latter can be bounded by

$$N\left(\varepsilon \sqrt{32}^{(r+1)/2}, \mathcal{C}_{r,a(r,\varepsilon)}, \|\cdot\|_\infty\right)$$

The result follows upon applying Proposition 12. □

We can now finish the proof for the entropy and remaining mass conditions. Choose  $r_n$  to be the smallest integer so that  $2^{r_n} \geq L\varepsilon_n^{-\frac{1}{\beta}}$ , where  $L$  is a constant, and set  $\mathcal{B}_n = \mathcal{C}_{r_n}$ . The entropy bound then follows directly from Proposition 13.

For the remaining mass condition, using Assumption 7, we obtain

$$P\left(b^{R,S} \notin \mathcal{B}_n\right) = P(R > r_n) \leq \exp\left(-c_1 2^{r_n} r_n\right) \leq \exp\left(-C_3 \varepsilon_n^{-\frac{1}{\beta}} |\log \varepsilon_n|\right),$$

and note that the constant  $C_3$  can be made arbitrarily big by choosing  $L$  big enough.

### 6.5 Proof of theorem 8 under assumption 6 (B)

We start with a lemma.

**Lemma 14** *Assume there exists  $0 < c_1 < c_2$  and  $0 < c_3$  with  $c_3 < c_1$  independent from  $r$ , such that for all  $i, i', 2 \leq \ell(i), \ell(i') \leq r$ ,*

$$c_1 2^{-\ell(i)} \leq A_{ii} \leq c_2 2^{-\ell(i)}, \tag{33}$$

$$|A_{ii'}| \leq c_3 2^{-1.5(\ell(i)+\ell(i'))} \quad \text{if } i \neq i'. \tag{34}$$

Let  $\tilde{A} = (A_{ii'})_{2 \leq \ell(i), \ell(i') \leq r}$  (so the right-lower submatrix of  $A^r$ ). Then for all  $x \in \mathbb{R}^{|\mathcal{I}_r|-2}$

$$(c_1 - c_3)x' \tilde{\Lambda} x \leq x' \tilde{A} x \leq 2c_2 x' \tilde{\Lambda} x.$$

where  $\tilde{\Lambda} = (\tilde{\Lambda}_{ii'})_{2 \leq \ell(i), \ell(i') \leq r}$  is the diagonal matrix with  $\tilde{\Lambda}_{ii} = 2^{-\ell(i)}$ ,

*Proof* In the following the summation are over  $i, i', 2 \leq \ell(i), \ell(i') \leq r$ . Trivially,  $x' A^r x = \sum_i x_i^2 A_{ii} + \sum_{i \neq j} x_i A_{ij} x_j$ . By the first inequality

$$c_1 x' \Lambda^{(r)} x = c_1 \sum_i x_i^2 2^{-\ell(i)} < \sum_i x_i^2 A_{ii} < c_2 \sum_i x_i^2 2^{-\ell(i)} = c_2 x' \Lambda^{(r)} x.$$

On the other hand

$$\left| \sum_{i \neq i'} x_i A_{ii'} x_{i'} \right| \leq c_3 \sum_{i \neq i'} |x_i| 2^{-1.5\ell(i)} |x_{i'}| 2^{-1.5\ell(i')} \leq c_3 \left( \sum_i |x_i| 2^{-1.5\ell(i)} \right)^2.$$

At the first inequality we used the second part of of (33). The second inequality follows upon including the diagonal. By Cauchy-Schwarz, this can be further bounded by

$$c_3 \left( \sum_i x_i^2 2^{-\ell(i)} \right) \left( \sum_i 2^{-2\ell(i)} \right) \leq c_3 x' \Lambda x,$$

where the final inequality follows from  $\sum_i 2^{-2\ell(i)} \leq \sum_{i=3}^\infty 2^{-2\ell(i)} = \sum_{j=1}^\infty 2^j 2^{-2j} = 1$ . The result follows by combining the derived inequalities.  $\square$

We continue with the proof of Theorem 8. Write  $A$  as block matrix

$$A = \begin{bmatrix} A_1 & B' \\ B & A_2 \end{bmatrix},$$

with  $A_1$  a  $2 \times 2$ -matrix, and  $B, A_2$  defined accordingly. By lemma 2

$$A_1 = \frac{\sigma^2}{2} \gamma \begin{bmatrix} \coth(\gamma/2) & \sinh^{-1}(\gamma/2) \\ \sinh^{-1}(\gamma/2) & \coth(\gamma/2) \end{bmatrix}.$$

Define the  $2 \times 2$ -matrix

$$\Lambda_1 = c \frac{\sigma^2}{2} \gamma \tanh(\gamma/4) \mathbf{I}, \quad c \in (0, 1).$$

where  $\mathbf{I}$  is the  $2 \times 2$ -identity matrix. It is easy to see that  $A_1 - \Lambda_1$  is positive definite.

When  $A_2 - \Lambda_2 - B(A_1 - \Lambda_1)^{-1}B'$  is positive definite, then it follows from the Cholesky decomposition that  $A - \Lambda$  is positive definite, where  $\Lambda = \text{diag}(\Lambda_1, \Lambda_2)$  positive definite. Note

$$(BA_1^{-1}B')_{i,i'} = \sum_{k,k'} B_{ik}(A_1)_{kk'}^{-1}B_{i'k'} \leq \left( \sum_{k,k'} (A_1)_{kk'}^{-1} \right) (B_{i,1} \vee B_{i,2})(B_{i',1} \vee B_{i',2})$$

where

$$\left( \sum_{k,k'} (A_1)_{kk'}^{-1} \right) = \frac{2}{\sigma^2 \gamma} \frac{2}{\sinh^{-1}(\gamma/2) + \coth(\gamma/2)} \leq \frac{2}{\sigma^2(1 + \gamma)}.$$

Therefore

$$|(BA_1^{-1}B')_{ii'}| \leq 0.020\sigma^2(1 + \gamma/4)2^{-1.5(\ell(i)+\ell(i'))}$$

Now consider  $\tilde{A} = A_2 - \Lambda_2 - B(A_1 - \Lambda_1)^{-1}B'$ . By lemma 2 and the bound on  $|(BA_1^{-1}B')_{ii'}|$  and choosing  $c > 0$  in the definition of  $\Lambda_1$  small enough, under the assumption that  $\gamma \leq 1.5$ ,

$$0.945 \cdot 2^{-\ell(i)}\sigma^2/4 < \tilde{A}_{ii} < 1.03 \cdot 2^{-\ell(i)}\sigma^2/4.$$

and for  $i \neq i'$   $|\tilde{A}_{ii'}| \leq 0.9415 \frac{\sigma^2}{4} 2^{-1.5(\ell(i)+\ell(i'))}$ . Therefore by lemma 14  $\tilde{A} - \Lambda_2$  is positive definite with diagonal matrix  $\Lambda_2$  with diagonal entries  $2^{-\ell(i)}$ .

It follows that  $x' \Lambda x \asymp x' A x$ . This implies that the small ball probabilities and the mass outside a sieve behave similar under Assumption 6(B) as when the  $Z_i$  are independent normally distributed with zero mean and variance  $\xi_i^2 = \Lambda_{ii}$ . As this case corresponds to Assumption 6(A) with  $\alpha = \frac{1}{2}$  for which posterior contraction has already been established, the stated contraction rate under Assumption 6(B) follows from Anderson’s lemma (lemma 19).

### 6.6 Proof of theorem 10: convergence in stronger norms

The linear embedding operator  $T : L^p(\mathbb{T}) \rightarrow L^2(\mathbb{T}), x \mapsto x$  is a well-defined injective continuous operator for all  $p \in (2, \infty]$ . Its inverse is easily seen to be a densely defined, closed unbounded linear operator. Following Knapik and Salomond (2014) we define the modulus of continuity  $m$  as

$$m(\mathcal{B}_n, \varepsilon) := \sup \{ \|f - f_0\|_p : f \in \mathcal{B}_n, \|f - f_0\|_2 \leq \varepsilon \}.$$

Theorem 2.1 of Knapik and Salomond (2014) adapted to our case is

**Theorem 15** (Knapik and Salomond (2014)) *Let  $\varepsilon_n \downarrow 0, T_n \uparrow \infty$  and  $\Pi$  be a prior on  $L_p(\mathbb{T})$  such that*

$$\mathbb{E}_0 \Pi \left( \mathcal{B}_n^c \mid X^{T_n} \right) \rightarrow 0,$$

for measurable sets  $\mathcal{B}_n \subset L^p(\mathbb{T})$ . Assume that for any positive sequence  $M_n$

$$\mathbb{E}_0 \Pi \left( b \in \mathcal{B}_n : \|b - b_0\|_2 \geq M_n \varepsilon_n \mid X^{T_n} \right) \rightarrow 0,$$

then

$$\mathbb{E}_0 \Pi \left( b \in L^p(\mathbb{T}) : \|b - b_0\|_p \geq m(\mathcal{B}_n, M_n \varepsilon_n) \mid X^{T_n} \right) \rightarrow 0.$$

Note that the sieves  $\mathcal{C}_{r,t}$  which we define in Sect. 6.4.2 have by Eq. (15) the property  $\Pi(\mathcal{C}_{r,t}^c \mid X^T) \rightarrow 0$ . By lemmas 21 and 23, the modulus of continuity satisfies  $m(\mathcal{C}_{r,u}, \varepsilon_n) \lesssim 2^{r(1/2-1/p)} \varepsilon_n$ , for all  $p \in (2, \infty]$ , (assume  $1/\infty = 0$ ), and the result follows.

**Acknowledgements** This work was partly supported by the Netherlands Organisation for Scientific Research (NWO) under the research programme ‘‘Foundations of nonparametric Bayes procedures’’, 639.033.110 and by the ERC Advanced Grant ‘‘Bayesian Statistics in Infinite Dimensions’’, 320637.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### A Lemmas used in the proofs

**Lemma 16** *Suppose  $z$  has Faber–Schauder expansion*

$$z = z_1 \psi_1 + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} z_{jk} \psi_{jk}.$$

If  $\|z\|_{\beta} < \infty$  (with the norm defined in (16)), then for  $r \geq 1$

$$\left\| z - \sum_{i \in \mathcal{I}_r} z_i \psi_i \right\|_{\infty} \leq \|z\|_{\beta} \frac{2^{-r\beta}}{2^{\beta} - 1}. \tag{35}$$

*Proof* This follows from

$$\begin{aligned} \left\| z - \sum_{i \in \mathcal{I}_r} z_i \psi_i \right\|_{\infty} &\leq \sum_{j=r+1}^{\infty} \left\| \sum_{k=1}^{2^j} z_{jk} \psi_{jk} \right\|_{\infty} \\ &= \sum_{j=r+1}^{\infty} 2^{-j\beta} \max_{1 \leq k \leq 2^j} 2^{j\beta} |z_{jk}| \leq \|z\|_{\beta} \sum_{j=r+1}^{\infty} 2^{-j\beta}. \end{aligned}$$

**Lemma 17** *If  $X \sim \text{IG}(A, B)$  then for any  $M > 0$ ,*

$$P(X \geq M) \leq \frac{B^A}{\Gamma(A)} M^{-A}.$$

*Proof* This follows from

$$\begin{aligned} P(X \geq M) &\leq \frac{B^A}{\Gamma(A)} \int_M^{\infty} x^{-\alpha-1} dx \\ &= -\frac{B^A}{\Gamma(A)} [x^{-\alpha}]_{x=M}^{\infty} = \frac{B^A}{\Gamma(A)} M^{-A}. \end{aligned}$$

□

**Lemma 18** Let  $X \sim N(0, 1)$ ,  $\theta \in \mathbb{R}$  and  $\varepsilon > 0$ . Then

$$P(|X - \theta| \leq \varepsilon) \geq \frac{e^{-\theta^2}}{\sqrt{2}} P(|X| \leq \sqrt{2}\varepsilon) \geq e^{\log \varepsilon - \varepsilon^2 - \theta^2 + \log \sqrt{\frac{2}{\pi}}}.$$

*Proof* Note that

$$\int_{\theta-\varepsilon}^{\theta+\varepsilon} e^{-\frac{1}{2}x^2} dx = \int_{-\varepsilon}^{\varepsilon} e^{-\frac{1}{2}(x+\theta)^2} dx$$

and

$$\frac{e^{-\frac{1}{2}(x+\theta)^2}}{e^{-\theta^2} e^{-\frac{1}{2}(\sqrt{2}x)^2}} = e^{\theta^2 - \frac{1}{2}(x+\theta)^2 + x^2} = e^{\frac{1}{2}(x-\theta)^2} \geq 1,$$

thus  $e^{-\frac{1}{2}(x+\theta)^2} \geq e^{-\theta^2} e^{-\frac{1}{2}(\sqrt{2}x)^2}$ , hence

$$\int_{\theta-\varepsilon}^{\theta+\varepsilon} e^{-\frac{1}{2}x^2} dx \geq e^{-\theta^2} \int_{-\varepsilon}^{\varepsilon} e^{-\frac{1}{2}(\sqrt{2}x)^2} dx = \frac{e^{-\theta^2}}{\sqrt{2}} \int_{-\sqrt{2}\varepsilon}^{\sqrt{2}\varepsilon} e^{-\frac{1}{2}u^2} du.$$

Now the elementary bound  $\int_{-y}^y e^{-\frac{1}{2}x^2} \geq 2ye^{-\frac{1}{2}y^2}$  gives

$$\begin{aligned} P(|X - \theta| \leq \varepsilon) &= \frac{1}{\sqrt{2\pi}} \int_{\theta-\varepsilon}^{\theta+\varepsilon} e^{-\frac{1}{2}x^2} dx \geq \frac{1}{\sqrt{2\pi}} \frac{e^{-\theta^2}}{\sqrt{2}} \int_{-\sqrt{2}\varepsilon}^{\sqrt{2}\varepsilon} e^{-\frac{1}{2}u^2} du \\ &= \frac{e^{-\theta^2}}{\sqrt{2}} P(|X| \leq \sqrt{2}\varepsilon) \geq \frac{1}{\sqrt{2\pi}} \frac{e^{-\theta^2}}{\sqrt{2}} 2\sqrt{2}\varepsilon e^{-\varepsilon^2} = \sqrt{\frac{2}{\pi}} e^{\log \varepsilon - \theta^2 - \varepsilon^2} \end{aligned}$$

□

**Lemma 19** (Anderson’s lemma) Define a partial order on the space of  $n \times n$ -matrices ( $n \in \mathbb{N} \cup \{\infty\}$ ) by setting  $A \leq B$ , when  $B - A$  is positive definite. If  $X \sim N(0, \Sigma_X)$  and  $Y \sim N(0, \Sigma_Y)$  independently with  $\Sigma_X \leq \Sigma_Y$ , then for all symmetric convex sets  $C$

$$P(Y \in C) \leq P(X \in C).$$

*Proof* See Anderson (1955).

□

**Lemma 20** Let

$$f = z_1 \psi_1 + \sum_{j=1}^r \sum_{k=1}^{2^j} z_{j,k} \psi_{j,k}.$$

Then

$$\sup_{i:\ell(i) \leq r} |z_i| \leq 2\|f\|_\infty.$$

*Proof* Note that  $|z_1| = |f(0)| \leq 2\|f\|_\infty$ , and  $|z_{0,1}| = |f(1/2)| \leq 2\|f\|_\infty$  and inductively, for  $j \geq 1$ ,  $z_{j,k} = f((2k - 1)2^{-(j+2)}) - \frac{1}{2}f(2^{-(j+1)}(k - 1)) - \frac{1}{2}f(2^{-(j-1)}k)$ , hence  $|z_{j,k}| \leq 2\|f\|_\infty$ . □

**Lemma 21** Let  $\mathcal{C}_r$  as in Sect. 6.4.2. Then

$$\sup_{0 \neq f \in \mathcal{C}_r} \frac{\|f\|_\infty}{\|f\|_2} \leq \sqrt{3} \cdot 2^{(r+1)/2}.$$

*Proof* Let  $f \in \mathcal{C}_r$  be nonzero. Note that for any constant  $c > 0$ ,

$$\frac{\|cf\|_\infty}{\|cf\|_2} = \frac{\|f\|_\infty}{\|f\|_2}.$$

Hence, we may and do assume that  $\|f\|_\infty = 1$ . Furthermore, since the  $L^2$  and  $L^\infty$  norm of  $f$  and  $|f|$  are the same, we also assume that  $f$  is nonnegative.

Let  $x_0$  be a global maximum of  $f$ . Clearly  $f(x_0) = 1$ . Since  $f$  is a linear interpolation between the points  $\{k2^{-j-1} : k = 0, 1, \dots, 2^{r+1}\}$ , we may also assume that  $x_0$  is of the form  $x_0 = k2^{-j-1}$ . We consider two cases

- (i)  $0 \leq k < 2^{r+1}$ ,
- (ii)  $k = 2^{r+1}$ .

In case (i) we have that  $f(x) \geq (1 - 2^{r+1}(x - k2^{-r-1}))I_{[k2^{-r-1}, (k+1)2^{-r-1}]}(x)$ , for all  $x \in [k2^{-r-1}, (k + 1)2^{-r-1}]$ . In case (ii)  $f(x) \geq 2^{r+1}(x - 1 + 2^{-r-1})I_{[1-2^{-r-1}, 1]}(x)$ , for all  $x \in [1 - 2^{-r-1}, 1]$ . Hence, in both cases,

$$\|f\|_2^2 \geq 2^{2r+2} \int_0^{2^{-r-1}} x^2 dx = \frac{1}{3}2^{2r+2}2^{-3r-3} = \frac{1}{3}2^{-r-1}.$$

Thus

$$\frac{\|f\|_\infty}{\|f\|_2} \leq \frac{1}{\frac{1}{\sqrt{3}}2^{-(r+1)/2}} = \sqrt{3} \cdot 2^{(r+1)/2},$$

uniformly over all nonzero  $f \in \mathcal{C}_{r,s}$ . □

**Lemma 22** *Let  $a_1, a_2, x_1, x_2$  be positive numbers. Then*

$$\frac{a_1 + a_2}{x_1 + x_2} \leq \frac{a_1}{x_1} \vee \frac{a_2}{x_2}.$$

*Proof* Suppose that the lemma is not true, so there are positive  $a_1, a_2, x_1, x_2$  such that,

$$\begin{aligned} \frac{a_1}{x_1} \vee \frac{a_2}{x_2} - \frac{a_1 + a_2}{x_1 + x_2} &= \left[ \frac{a_1}{x_1} - \frac{a_1 + a_2}{x_1 + x_2} \right] \vee \left[ \frac{a_2}{x_2} - \frac{a_1 + a_2}{x_1 + x_2} \right] \\ &= \left[ \frac{\frac{a_1}{x_1}(x_1 + x_2) - (a_1 + a_2)}{x_1 + x_2} \right] \vee \left[ \frac{\frac{a_2}{x_2}(x_1 + x_2) - (a_1 + a_2)}{x_1 + x_2} \right] < 0. \end{aligned}$$

Hence, both terms on the right-hand-side are negative. In particular, this means for the first term that  $x_2/x_1 < a_2/a_1$ . For the second term this gives  $x_1/x_2 < a_1/a_2$ . These two inequalities cannot hold simultaneously and we have reached a contradiction. □

**Lemma 23** *Let  $\mathcal{C}_r$  and  $\mathcal{C}_{r,s}$  as in Sect. 6.4.2. Then for  $p \in [2, \infty)$ ,*

$$\sup_{0 \neq f \in \mathcal{C}_r} \frac{\|f\|_p}{\|f\|_2} \leq \frac{3^{1/2}}{(p + 1)^{1/p}} 2^{(r+1)(1/2-1/p)}.$$

*Proof* Let  $f \in \mathcal{C}_r$ . Just as in proof of lemma 21 we may assume that  $f$  is nonnegative and  $\|f\|_2 = 1$ . Hence

$$\sup_{0 \neq f \in \mathcal{C}_r} \frac{\|f\|_p}{\|f\|_2} = \left( \sup_{0 \neq f \in \mathcal{C}_r, \|f\|_2=1} \|f\|_p^p \right)^{1/p} = \left( \sup_{0 \neq f \in \mathcal{C}_r, \|f\|_2=1} \frac{\|f\|_p^p}{\|f\|_2^p} \right)^{1/p}.$$



Note that

$$\|f\|_p^p = \sum_{k=0}^{2^{r+1}-1} \int_{k2^{r-1}}^{(k+1)2^{r-1}} f(x)^p dx.$$

Hence, by repeatedly applying lemma 22

$$\frac{\sum_{k=0}^{2^{r+1}} \int_{k2^{r-1}}^{(k+1)2^{r-1}} f(x)^p dx}{\sum_{k=0}^{2^{r+1}-1} \int_{k2^{r-1}}^{(k+1)2^{r-1}} f(x)^2 dx} \leq \bigvee_{\substack{k \in \{0, \dots, 2^{r+1}-1\} \\ \exists x \in (k2^{r-1}, (k+1)2^{r-1}): f(x) \neq 0}} \frac{\int_{k2^{r-1}}^{(k+1)2^{r-1}} f(x)^p dx}{\int_{k2^{r-1}}^{(k+1)2^{r-1}} f(x)^2 dx}.$$

Note that  $f$  is a linear interpolation between the points  $k2^{r-1}, k \in \{0, 1, \dots, 2^{r+1}\}$ .

Now study affine functions  $g : [0, 2^{r-1}] \rightarrow \mathbb{R}$  which are positive. A maximum of  $g$  is attained in either 0 or  $2^{r-1}$ . Without lose of generality it is attained in 0. Using scaling in a later stadium of the proof, we assume for the moment that  $g(0) = 1$ . Hence  $a := g(2^{r-1}) \in [0, 1]$ . Note that

$$g(x) = 1 - (1 - a)2^{r+1}x.$$

When  $a = 1, \|g\|_p = \|g\|_2 = 1$ . Now consider  $a < 1$ ,

$$\int_0^{2^{r-1}} g(x)^p dx = \int_0^{2^{r-1}} [1 - (1 - a)2^{r+1}x]^p dx.$$

Let  $y = -x + \frac{2^{r-1}}{1-a}$  then  $x = -y + \frac{2^{r-1}}{1-a}$  and  $dx = -dy$ . Hence

$$\begin{aligned} \int_0^{2^{r-1}} g(x)^p dx &= \int_{-2^{r-1} + \frac{2^{r-1}}{1-a}}^{\frac{2^{r-1}}{1-a}} (1 - a)^p 2^{rp+p} y^p dy \\ &= 2^{-r-1} \frac{1}{p+1} \left[ \frac{1 - a^{p+1}}{1 - a} \right]. \end{aligned}$$

Note that for a constant  $c > 0$  and a function  $h$ ,

$$\frac{\|ch\|_p^p}{\|ch\|_2^2} = \frac{c^p \|h\|_p^p}{c^2 \|h\|_2^2} = c^{p-2} \frac{\|h\|_p^p}{\|h\|_2^2}.$$

Let

$$c^2 = 3 \cdot 2^{r+1} \frac{1 - a}{1 - a^3}.$$

Hence  $cg$  has  $L^2$ -norm one and

$$\begin{aligned} \|cg\|_p^p &= c^p \|g\|_p^p \\ &= \left( 3 \cdot 2^{r+1} \frac{1 - a}{1 - a^3} \right)^{\frac{p}{2}} 2^{-r-1} \frac{1}{p+1} \left[ \frac{1 - a^{p+1}}{1 - a} \right] \\ &= \frac{3^{p/2}}{p+1} 2^{(r+1)(p/2-1)} (1 - a)^{p/2-1} (1 - a^3)^{-p/2} (1 - a^{p+1}). \end{aligned}$$

The maximum is attained for  $a = 0$ , then

$$\|cg\|_p^p = \frac{3^{p/2}}{p+1} 2^{(r+1)(p/2-1)}$$

Hence

$$\|cg\|_p = \frac{3^{1/2}}{(p+1)^{1/p}} 2^{(r+1)(1/2-1/p)}$$

and the result follows, using that  $\|f\mathbf{I}_{(k2^{-r-1}, (k+1)2^{-r-1})}\|_2^2 \leq \|f\|_2^2$  and that for  $0 < c' < c$ ,

$$\frac{\|c'g\|_p^p}{\|c'g\|_2^2} = \frac{(c'/c)^p \|cg\|_p^p}{(c'/c)^2 \|cg\|_2^2} \leq \frac{\|cg\|_p^p}{\|cg\|_2^2}.$$

□

## References

- Anderson TW (1955) The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc Am Math Soc* 6:170–176
- Bhattacharya R, Waymire E (2007) A basic course in probability theory. Universitext, Springer, New York
- Dalalyan A (2005) Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann Stat* 33(6):2507–2528
- Dalalyan AS, Kutoyants YA (2002) Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math Methods Stat* 11(4):402–427
- Ghosal S, van der Vaart AW (2007) Convergence rates of posterior distributions for noniid observations. *Ann Stat* 35(1):192–223
- Ghosal S, Ghosh JK, van der Vaart AW (2000) Convergence rates of posterior distributions. *Ann Stat* 28(2):500–531
- Giné E, Nickl R (2011) Rates of contraction for posterior distributions in  $L^r$ -metrics,  $1 \leq r \leq \infty$ . *Ann Stat* 39(6):2883–2911
- Giné E, Nickl R (2016) Mathematical foundations of infinite-dimensional statistical models. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge
- Hindriks R (2011) Empirical dynamics of neuronal rhythms. PhD thesis, Vrije Universiteit Amsterdam
- Karatzas I, Shreve SE (1991) Brownian motion and stochastic calculus, volume 113 of graduate texts in mathematics, 2nd edn. Springer, New York
- Knapik BT, van der Vaart AW, van Zanten JH (2011) Bayesian inverse problems with Gaussian priors. *Ann Stat* 39(5):2626–2657
- Knapik B, Salomond J-B (2014) A general approach to posterior contraction in nonparametric inverse problems. *Bernoulli*
- Kutoyants YA (2004) Statistical inference for ergodic diffusion processes. Springer, New York
- Papaspiliopoulos O, Pokern Y, Roberts GO, Stuart AM (2012) Nonparametric estimation of diffusions: a differential equations approach. *Biometrika* 99(3):511
- Pokern Y (2007) Fitting Stochastic Differential Equations to Molecular Dynamics Data. PhD thesis, University of Warwick
- Pokern Y, Stuart AM, van Zanten JH (2013) Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stoch Process Appl* 123(2):603–628
- Schauer M, van Zanten JH (2017) Uniform central limit theorems for additive functionals of diffusions on the circle. In preparation
- Shen W, Ghosal S (2015) Adaptive Bayesian procedures using random series priors. *Scand J Stat* 42(4):1194–1213
- Spokoiny VG (2000) Adaptive drift estimation for nonparametric diffusion model. *Ann Stat* 28(3):815–836
- Strauch C (2015) Sharp adaptive drift estimation for ergodic diffusions: the multivariate case. *Stoch Process Appl* 125(7):2562–2602
- van der Meulen FH, van der Vaart AW, van Zanten JH (2006) Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli* 12(5):863–888
- van der Meulen FH, Schauer M, van Zanten JH (2014) Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Comput Stat Data Anal* 71:615–632
- van der Vaart AW, van Zanten JH (2008) Rates of contraction of posterior distributions based on Gaussian process priors. *Ann Stat* 36(3):1435–1463

- van Waaij J, van Zanten H (2016) Gaussian process methods for one-dimensional diffusions: optimal rates and adaptation. *Electron J Stat* 10(1):628–645
- van Zanten JH (2013) Nonparametric Bayesian methods for one-dimensional diffusion models. *Math Biosci* 243(2):215–222