Nonparametric Bayesian inference for ergodic diffusions

Laura Panzar

ORTEC Finance
Orly Plaza
Orlyplein 145c
1043 DV Amsterdam
The Netherlands

Harry van Zanten*

Department of Mathematics
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands
j.h.v.zanten@tue.nl

May 28, 2009 (second revised version)

Abstract

The problem of nonparametric drift estimation for ergodic diffusions is studied from a Bayesian perspective. In particular, Gaussian process priors are exhibited that yield optimal contraction rates if the drift function belongs to a smoothness class.

1 Introduction

In this paper we study nonparametric Bayes procedures for the problem of making inference about the drift function θ of an ergodic diffusion process X satisfying the stochastic differential equation

$$dX_t = \theta(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad X_0 = x, \tag{1.1}$$

where W is a standard Brownian motion. We assume we observe a continuous record $X^{T_n} = (X_t : t \in [0, T_n])$ for some time horizon $T_n > 0$ and

*Research partially funded by the Netherlands Organization for Scientific Research (NWO).

consider an asymptotic framework in which $T_n \to \infty$ as $n \to \infty$. Since the diffusion function σ can essentially be recovered without error from the quadratic variation in this setting, we assume that it is known.

For parametric ergodic diffusion models the asymptotic behaviour of Bayes procedures is essentially known. Under regularity conditions such a model is locally asymptotically normal and Bayesian estimators converge at the usual parametric rate ${T_n}^{-1/2}$ and are asymptotically efficient, cf. Kutoyants (2004). Various nonparametric methods for continuous observations from the model (1.1) have been considered in the literature as well. In particular kernel-type estimators have been widely studied for θ belonging to a smoothness class. See, among others, Banon (1978), Tuan (1981), Van Zanten (2001), Galtchouk and Pergamenshchikov (2001), Kutoyants (2004). Optimal adaptive methods were developed by Spokoiny (2000) and Dalalyan (2005).

Nonparametric Bayesian methods for diffusions have recently been considered in the paper Van der Meulen et al. (2006). For continuously observed ergodic diffusions the latter paper provides a general result for obtaining contraction rates for nonparametric priors, in the spirit of the results of Ghosal et al. (2000) for i.i.d. data and Ghosal and Van der Vaart (2007) for several non-i.i.d. settings. The only known concrete asymptotic result for a specific prior deals with ergodic diffusion models of the type (1.1) with a decreasing drift function θ. It is proved in Van der Meulen et al. (2006) that the posterior based on a Dirichlet-type prior contracts in that case at the rate $T_n^{-1/3}$ (up to a logarithmic factor), which is optimal for that setting.

The focus in the present paper is on drift functions satisfying a smoothness condition rather than a shape restriction. It is by now well known that in Bayesian nonparametrics, prior distributions have to be carefully chosen. Naive choices may lead to inconsistent procedures (cf. e.g. Diaconis and Freedman (1986)) or suboptimal convergence rates (Castillo (2008)). Our main aim is to propose priors that perform well when the drift function belongs to a smoothness class. If the true drift function is α-smooth in an appropriate sense, the optimal minimax rate of a nonparametric estimator is $T_n^{-\alpha/(1+2\alpha)}$, cf. Kutoyants (2004). We will exhibit certain Gaussian process priors for the drift which yield contraction of the posterior at this optimal rate, up to a logarithmic factor.

The priors are constructed from so-called Riemann-Liouville processes, which are essentially (fractionally) integrated Brownian motions. The use of integrated Brownian motion priors goes back at least to Wahba (1978), who employed them in a nonparametric regression setting. In the paper Van der Vaart and Van Zanten (2008a) they were shown to have good asymptotic properties in several statistical settings, including nonparametric regression and density estimation. A complicating factor in the ergodic diffusion setting is the fact that the unknown function of interest is defined on the whole real line. We deal with this situation by employing a sequence of priors constructed from Riemann-Liouville processes indexed by increasing compact intervals. The rate at which these intervals grow influences the rate of contraction of the posterior.

The Gaussian process priors we propose are not adaptive, in the sense that the optimal prior for α-regular drift functions depends on the regularity α. We expect that adaptation can be achieved by treating α as a hyperparameter, endowing it with an additional prior distribution. Recent studies have shown that such hierarchical Bayes procedures can indeed be rate adaptive when the prior on the hyperparameter is carefully chosen (see for instance Belitser and Ghosal (2003), Huang (2004), Ghosal et al. (2008), Van der Vaart and Van Zanten (2009)). The results of this paper can serve as a first step in the construction of such adaptive Bayesian methods for ergodic diffusions.

The remainder of the paper is organized as follows. In the next section we present the precise assumptions on the diffusion model, the construction of the Riemann-Liouville prior, and the main result on the rate of contraction of the corresponding posterior distribution. The proof of the theorem relies on general rate of contraction results for ergodic diffusions given in Section 3. The proof itself is presented in the final Section 4.

2 Main results

2.1 Model assumptions

We assume we have observations $X^{T_n} = (X_t : t \in [0, T_n])$ from the diffusion model

$$dX_t = \theta_0(X_t)\, dt + \sigma(X_t) dW_t, \qquad X_0 = x. \tag{2.1}$$

Here W is a Brownian motion, $x \in \mathbb{R}$ is a fixed initial point (this is not essential) and T_n is a time horizon that increases to infinity as $n \to \infty$. The drift and diffusion functions θ_0 and σ are assumed to verify the usual Lipshitz and linear growth conditions that ensure the existence of a unique (strong) solution of the SDE (2.1), see for instance Karatzas and Shreve (1991). The aim is to estimate the unknown drift function θ_0. The positive function σ is considered to be known, as it can be reconstructed from the observations without error, at least on the range of the observations. The diffusion function σ only enters the likelihood through the quantity $\sigma^2(X_t)$, cf. (2.2) in the following subsection. In our setting the quadratic variation process $(\langle X\rangle_t : t \in [0, T_n])$ of the diffusion is fully observed and it almost surely holds that $\sigma^2(X_t) = d\langle X\rangle_t /dt$.

In addition to the Lipschitz and linear growth conditions we assume that σ is a positive function bounded away from zero and infinity, and that the true drift function satisfies

$$\limsup_{|x|\to\infty} \theta_0(x)\mathrm{sign}(x) < 0.$$

These assumptions are standard in this setting and ensure that the SDE (2.1) generates a recurrent, ergodic diffusion on the whole real line with invariant probability measure μ_0 satisfying

$$\mu_0(dx) = C_0 \frac{e^{2\int_0^x \frac{\theta_0(y)}{\sigma^2(y)}\,dy}}{\sigma^2(x)}\,dx,$$

with C_0 a normalizing constant. Moreover, the assumptions imply that the measure μ_0 has subexponential tails, i.e. there exist constants $C_1, C_2 > 0$ such that

$$\mu_0(x : |x| > a) \leq C_1 e^{-C_2 a}$$

for $a > 0$ large enough.

2.2 Construction of the prior

The Bayesian approach to making inference about θ_0 is to endow it with a prior distribution Π_n and base the inference on the posterior distribution, which is given by

$$\Pi_n(B \mid X^{T_n}) = \frac{\int_B L(\theta, X^{T_n})\,\Pi_n(d\theta)}{\int L(\theta, X^{T_n})\,\Pi_n(d\theta)}.$$

Here $L(\theta, X^{T_n})$ is the likelihood,

$$L(\theta, X^{T_n}) = \exp\Big(\int_0^{T_n} \frac{\theta(X_t)}{\sigma^2(X_t)}\,dX_t - \frac{1}{2}\int_0^{T_n} \frac{\theta^2(X_t)}{\sigma^2(X_t)}\,dt\Big), \qquad (2.2)$$

cf. e.g. Liptser and Shiryayev (1977). The priors Π_n we will propose below give full mass to the function space $L^2(\mu_0) = L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mu_0)$. For $\theta \in L^2(\mu_0)$ we have

$$\int_0^{T_n} \theta^2(X_t)\,dt \leq \|\theta\|_{2,\mu_0}^2 \sup_{y\in\mathbb{R}} l_{T_n}(y),$$

where $\|f\|_{2,\mu_0}^2 = \int f^2\,d\mu_0$ and $(l_T(y) : T \geq 0, y \in \mathbb{R})$ is the local time of the diffusion relative to the normalized speed measure μ_0 (cf. e.g. Itô and McKean (1965)). In particular the integral on the left is almost surely finite. Together with the assumption that σ is bounded away from 0 this ensures that $L(\theta, X^{T_n})$ is well defined for all θ in the support of Π_n.

Recall that the Riemann-Liouville process (RL-process) with parameter $\alpha > 0$ is defined by

$$R_x^\alpha = \int_0^x (x-y)^{\alpha-1/2}\, dB_y, \qquad x \geq 0,$$

where B is a standard Brownian motion. Integration by parts shows that R^α can be viewed as an α-fold (fractional) integral of the Brownian motion B. It is clear that the RL-process is a centered, Gaussian process. The scaling property of Brownian motion implies that it is α-self-similar in the sense that for all $c > 0$, the process $c^{-\alpha} R_{cx}$ is a RL-process with parameter α as well. For $M > 0$ we define the modified RL-process with parameter α, indexed by $[0, M]$, by

$$S_x^{\alpha,M} = M^\alpha \sum_{k=0}^{\underline{\alpha}+1} Z_k \Big(\frac{x}{M}\Big)^k + R_x^\alpha, \qquad x \in [0, M],$$

where $\underline{\alpha}$ is the biggest integer strictly smaller than α and $Z_1, Z_2, \ldots$ are independent standard Gaussian variables, independent of R^α.

As prior on θ_0 we employ the law Π_n of the process

$$S_{x+M_n}^{\alpha,2M_n} 1_{[-M_n,M_n]}(x), \qquad x \in \mathbb{R}, \tag{2.3}$$

where M_n is an increasing sequence that will be specified below. Note that the sample paths of the process (2.3) vanish outside the interval $[-M_n, M_n]$. Since we are attempting to estimate the drift function θ_0 which is supported on the whole real line we will have to choose the constants M_n such that $M_n \to \infty$ to ensure posterior consistency.

2.3 Contraction rates

In the statistical setting under consideration the natural global statistical metric is the random Hellinger distance h_n defined by

$$h_n^2(\theta, \psi) = \frac{1}{T_n} \int_0^{T_n} \Big(\frac{\theta(X_t) - \psi(X_t)}{\sigma(X_t)}\Big)^2\, dt, \tag{2.4}$$

cf. also Van der Meulen et al. (2006). Note that by ergodicity, we have that

$$h_n(\theta, \psi) \to \Big\| \frac{\theta - \psi}{\sigma} \Big\|_{2,\mu_0}$$

P_{θ_0}-almost surely, where $\|\cdot\|_{2,\mu_0}$ is the L^2-norm associated to the invariant measure μ_0, i.e. $\|f\|_{2,\mu_0}^2 = \int f^2\, d\mu_0$. Moreover, since σ is bounded away from zero and infinity, the norm $\|\cdot/\sigma\|_{2,\mu_0}$ is equivalent to the $L^2(\mu_0)$-norm.

We say that the posterior associated to the prior Π_n contracts around θ_0 at the rate ε_n if

$$\Pi_n\Big(\theta : h_n(\theta,\theta_0) > L\varepsilon_n \,|\, X^{T_n}\Big) \xrightarrow{P_{\theta_0}} 0$$

for L large enough, as $n \to \infty$. The convergence is convergence in probability under the true model. In other words, posterior contraction at the rate ε_n means that asymptotically, the posterior mass is concentrated on h_n-balls around the true drift function θ_0 with radius of the order ε_n. We remark that it follows from the results of Van Zanten (2003) that if the posterior contracts at the rate ε_n in this sense, then for all compact subsets $J \subset \mathbb{R}$,

$$\Pi_n\Big(\theta : \Big\|(\theta-\theta_0)1_J\Big\|_{2,\mu_0} > L\varepsilon_n \,|\, X^{T_n}\Big) \xrightarrow{P_{\theta_0}} 0$$

for L large enough, as $n \to \infty$.

If a prior Π_n yields a posterior contraction rate ε_n in this sense, there exist for every compact $J \subset \mathbb{R}$ nonparametric estimators of the restriction of θ_0 to J with attain the rate of convergence ε_n relative to the $L^2(\mu_0)$-norm, cf. Ghosal et al. (2000). This implies that the contraction rate of the posterior as defined above is limited by the optimal convergence rate of estimators relative to this norm. The main result of this paper asserts that if the true drift function is α-smooth, a properly chosen modified RL-process prior yields the optimal rate $T_n^{-\alpha/(1+2\alpha)}$, up to a logarithmic factor.

For $\alpha > 0$, we denote by $C^\alpha(\mathbb{R})$ be the space of α-Hölder continuous functions. For $\alpha \in (0,1]$ this is the space of function f for which there exists a constant C such that $|f(x)-f(y)| \le C|x-y|^\alpha$ for $x,y \in \mathbb{R}$. For general $\alpha > 0$ a function f belongs to $C^\alpha(\mathbb{R})$ if it has $\underline{\alpha}$ continuous, bounded derivatives and highest derivative $f^{(\underline{\alpha})}$ belongs to $C^{\alpha-\underline{\alpha}}(\mathbb{R})$.

Theorem 2.1. *Let the model assumptions of Section 2.1 be satisfied. Let the prior Π_n be the law of the shifted modified Riemann-Liouville process defined by (2.3) in Section 2.2, with $\alpha > 0$ and $M_n = \log^p T_n$ for some $p > 1$. Then if $\theta_0 \in C^\alpha(\mathbb{R})$, the posterior contracts around θ_0 at the rate*

$$\varepsilon_n = \Big(\frac{T_n}{\log^p T_n}\Big)^{-\alpha/(1+2\alpha)}.$$

The presence of the logarithmic factor is due to the fact that we are using priors that give full prior mass to classes of functions supported on the increasing interval $[-M_n, M_n]$. We need this to ensure that we are eventually close to the true drift function θ_0 on any interval $J \subset \mathbb{R}$. We conjecture that if a fixed prior Π is used, with M_n replaced by a fixed M, we have contraction at the optimal rate without the extra logarithmic factor, but only on the interval $J = [-M, M]$, not globally. Moreover, it is likely that the global Hölder condition we impose can then be replaced by a smoothness condition on the interval $[-M, M]$ only. The use of such a fixed prior can be seen as misspecifying the model. Techniques analogous to those of Kleijn and Van der Vaart (2006) for i.i.d. data may be necessary to prove (or disprove) this conjecture.

The proof of Theorem 2.1 is deferred to Section 4. In the next section we first prepare some general results on posterior contraction rates for the model we are investigating.

3 Some general posterior contraction results for ergodic diffusion models

3.1 Rates of posterior contraction in ergodic diffusion models

A set of sufficient conditions for having a certain given rate of posterior contraction in the ergodic diffusion model is given by Theorem 3.3 and Lemma 2.2 of Van der Meulen et al. (2006). The conditions as stated in the cited paper are however not flexible enough to be directly useful for the proof of Theorem 2.1. In this section we adapt the results to make them suitable for our purposes.

The result below gives conditions that are quite common in rate of contraction results in Bayesian nonparametrics. Roughly speaking, it requires that sufficient prior mass is put near the true parameter and that there exists certain subsets of the parameter space, so-called *sieves*, of which we can control the metric entropy and that contain most of the prior mass. The key difference with the results of Van der Meulen et al. (2006) is that the theorem presented here puts less restrictive conditions on the amount of prior mass that is allowed to lie outside a certain sieve. We need this refinement to be able to deal with the Gaussian process priors we are interested in.

For a metric space (A, d) and $\varepsilon > 0$ we denote by $N(\varepsilon, A, d)$ the ε-covering number of (A, d), i.e. the minimal number of balls of d-radius ε that are needed to cover A. Recall that h_n is the random Hellinger metric defined by (2.4).

Theorem 3.1. *Let Π_n be a sequence of prior distributions on $L^2(\mu_0)$, ε_n a sequence of positive numbers such that $\varepsilon_n \downarrow 0$ and $T_n\varepsilon_n^2 \to \infty$. Suppose that there exists a $C_0 > 0$ such that for every $C > C_0$, there exist subsets $\Theta_{C,n} \subset L^2(\mu_0)$ and a constant $K > 0$ such that*

$$\log N(\varepsilon_n, \Theta_{C,n}, \|\cdot\|_{2,\mu_0}) \leq KT_n\varepsilon_n^2, \tag{3.1}$$

$$\Pi_n(\theta : \|\theta - \theta_0\|_{2,\mu_0} < \varepsilon_n) \geq e^{-T_n\varepsilon_n^2}, \tag{3.2}$$

$$\Pi_n(\Theta_{C,n}^c) \leq e^{-CT_n\varepsilon_n^2}. \tag{3.3}$$

Then for all $L > 0$ sufficiently large,

$$\Pi_n\Big(\theta : h_n(\theta, \theta_0) > L\varepsilon_n \,|\, X^{T_n}\Big) \xrightarrow{P_{\theta_0}} 0$$

as $n \to \infty$.

Proof. Let $U_n = \{\theta : h_n(\theta, \theta_0) > L\varepsilon_n\}$ be the set of interest. Fix $\gamma > 0$ and a nonempty, compact interval $J \subset \mathbb{R}$. According to the proof of Theorem 3.3 of Van der Meulen et al. (2006) there exist constants $c_1, c_2 > 0$ (depending on γ and J) and events A_n such that $\mathbb{P}_{\theta_0}(A_n^c) \leq \gamma$ for n large enough and on A_n,

$$c_1\|(\theta - \psi)1_J\|_{2,\mu_0}^2 \leq h_n^2(\theta, \psi) \leq c_2\|\theta - \psi\|_{2,\mu_0}^2$$

for all $\theta, \psi \in \Theta$ (this follows from local time considerations, cf. Van Zanten (2003)). For all $C > C_0$,

$$\Pi_n(U_n \,|\, X^{T_n}) \leq \Pi_n(U_n \cap \Theta_{C,n} \,|\, X^{T_n})1_{A_n} + \Pi_n(\Theta_{C,n}^c \,|\, X^{T_n})1_{A_n} + 1_{A_n^c}.$$

To prove the statement of the theorem it thus suffices to show that $C > C_0$ can be chosen such that

$$\mathbb{E}_{\theta_0}\Pi_n(U_n \cap \Theta_{C,n} \,|\, X^{T_n})1_{A_n} \to 0,$$

$$\mathbb{E}_{\theta_0}\Pi_n(\Theta_{C,n}^c \,|\, X^{T_n})1_{A_n} \to 0.$$

The proof of Theorem 2.1 of Van der Meulen et al. (2006) shows that for *any* $C > C_0$, the first of these convergence statements holds. It remains to prove that the second one holds from some choice of $C > C_0$. Set $c = c_1 \vee c_2$. By the proof of Lemma 2.2 of Van der Meulen et al. (2006),

$$\mathbb{E}_{\theta_0}\Pi_n(\Theta_{C,n}^c \,|\, X^{T_n})1_{A_n} \leq \frac{\Pi_n(\Theta_{C,n}^c)e^{(1+\frac{1}{2}c^2)T_n\varepsilon_n^2}}{\Pi_n(\theta : \|\theta - \theta_0\|_{2,\mu_0} < \varepsilon_n)} + e^{-\frac{1}{2c^2}T_n\varepsilon_n^2}.$$

By (3.2) and (3.3) this is further bounded by

$$e^{(2+\frac{1}{2}c^2-C)T_n\varepsilon_n^2} + e^{-\frac{1}{2c^2}T_n\varepsilon_n^2}.$$

Choosing C large enough ensures that this expression convergence to 0 as $n \to \infty$. $\square$

3.2 Rates of contraction for Gaussian priors in ergodic diffusion models

In this section we present a general theorem giving conditions under which a sequence of Gaussian process priors yields a certain rate of posterior contraction for the ergodic diffusion model described in Section 2.1. The theorem complements the results of Van der Vaart and Van Zanten (2008a), who consider rates of contraction for Gaussian process priors in several other statistical settings.

Let positive numbers M_n be given and let $S^n = (S^n_x)_{x\in I_n}$ be a sequence of centered, continuous Gaussian processes indexed by $I_n = [-M_n, M_n]$. Assume that almost surely

$$\int_{I_n} (S^n_x)^2\, \mu_0(dx) < \infty$$

for all n, so that we may view S^n as a centered Gaussian random element in the Hilbert space $L^2(\mu_0)$. We implicitly view S^n as indexed by the whole real line by setting it equal to 0 outside I_n. The reproducing kernel Hilbert space (RKHS) of the process S^n is denoted by $\mathbb{H}^n$, i.e. $\mathbb{H}^n$ is the closure of the linear span of the collection of functions $\{x \mapsto \mathbb{E}S^n_x S^n_y : y \in I_n\}$ relative to the inner product

$$\big\langle x \mapsto \mathbb{E}S^n_x S^n_{y_1}, x \mapsto \mathbb{E}S^n_x S^n_{y_2}\big\rangle_{\mathbb{H}^n} = \mathbb{E}S^n_{y_1} S^n_{y_2}.$$

As prior distribution on the drift function we employ the law that the process Π_n of the process W^n generates on $L^2(\mu_0)$. We say that *the posterior contracts around θ_0 at the rate ε_n* if

$$\Pi_n\Big(\theta : h_n(\theta, \theta_0) > L\varepsilon_n \,|\, X^{T_n}\Big) \xrightarrow{P_{\theta_0}} 0$$

for L large enough, as $n \to \infty$. The following theorem gives sufficient conditions for posterior contraction at the rate ε_n.

Theorem 3.2. *Suppose the positive numbers $\varepsilon_n \downarrow 0$ satisfy $T_n\varepsilon_n^2 \to \infty$,*

$$\inf_{h\in\mathbb{H}^n : \|(h-\theta_0)1_{I_n}\|_{2,\mu_0} < \varepsilon_n} \|h\|^2_{\mathbb{H}^n} \le T_n\varepsilon_n^2, \tag{3.4}$$

$$-\log \mathbb{P}\Big(\Big\|S^n 1_{I_n}\Big\|_{2,\mu_0} < \varepsilon_n\Big) \le T_n\varepsilon_n^2 \tag{3.5}$$

and

$$\int_{I_n^c} \theta_0^2\, d\mu_0 \le \varepsilon_n^2. \tag{3.6}$$

Then the posterior contracts around θ_0 at the rate ε_n.

Proof. Consider the concentration function $\varphi^n_{\theta_0}$ associated to the process S^n (viewed as a random element in $L^2(\mu_0)$) and θ_0, which is defined by

$$\varphi^n_{\theta_0}(\varepsilon) = \inf_{h\in\mathbb{H}^n:\|h-\theta_0\|_{2,\mu_0}<\varepsilon} \|h\|^2_{\mathbb{H}^n} - \log \mathbb{P}(\|S^n\|_{2,\mu_0} < \varepsilon)$$

(cf. Van der Vaart and Van Zanten (2008a)). Let n be fixed for the moment. By (3.4) there exists an $h \in \mathbb{H}^n$ such that $\|(h - \theta_0)1_{I_n}\|_{2,\mu_0} < \varepsilon_n$ and $\|h\|^2_{\mathbb{H}^n} \leq 2T_n\varepsilon_n^2$. By (3.6), we have

$$\|h - \theta_0\|^2_{2,\mu_0} = \|(h - \theta_0)1_{I_n}\|^2_{2,\mu_0} + \|\theta_0 1_{I_n^c}\|^2_{2,\mu_0} < 2\varepsilon_n^2.$$

Combined with (3.5) this shows that

$$\varphi^n_{\theta_0}(\sqrt{2}\varepsilon_n) \leq 3T_n\varepsilon_n^2.$$

By Theorem 2.1 of Van der Vaart and Van Zanten (2008a) (applied with $2\varepsilon_n$ in the place of ε_n) this implies that for every large enough positive constant C there exist measurable subsets $\Theta_{C,n}$ of $L^2(\mu_0)$ such that

$$\log N(6\varepsilon_n, \Theta_{C,n}, \|\cdot\|_{2,\mu_0}) \leq 12CT_n\varepsilon_n^2, \tag{3.7}$$

$$\Pi_n(\theta : \|\theta - \theta_0\|_{2,\mu_0} < 4\varepsilon_n) \geq e^{-4T_n\varepsilon_n^2}, \tag{3.8}$$

$$\Pi_n(\Theta^c_{C,n}) \leq e^{-4CT_n\varepsilon_n^2}. \tag{3.9}$$

The result now follows from Theorem 3.1. □

4 Proof of the main result

4.1 Auxiliary result for the Riemann-Liouville prior

Consider the modified RL-process $(S^{\alpha,M}_x)_{x\in[0,M]}$ defined in Section 2.2 and let $\mathbb{H}^{\alpha,M}$ be the RKHS associated to the process. For the proof of Theorem 2.1 we need upper bounds for the quantities

$$-\log \mathbb{P}\Big(\|S^{\alpha,M}\|_{\infty,M} < \varepsilon\Big)$$

and, for $w \in C^\alpha[0, M]$,

$$\inf_{h\in\mathbb{H}^{\alpha,M}:\|h-w\|_{\infty,M}<\varepsilon} \|h\|^2_{\mathbb{H}^{\alpha,M}},$$

where $\|f\|_{\infty,M} = \sup_{x\in[0,M]} |f(x)|$. For $M = 1$, Theorem 4.3 of Van der Vaart and Van Zanten (2008a) implies that both quantities are bounded by a constant times $\varepsilon^{-1/\alpha}$. The following lemma shows how de upper bounds depend on M.

Lemma 4.1. *For $w \in C^\alpha[0,M]$*

$$\inf_{h\in\mathbb{H}^{\alpha,M}:\|h-w\|_{\infty,M}<\varepsilon} \|h\|^2_{\mathbb{H}^{\alpha,M}} - \log \mathbb{P}\Big(\|S^{\alpha,M}\|_{\infty,M} < \varepsilon\Big) \le KM\varepsilon^{-1/\alpha},$$

where K is a constant that only depends on the C^α-norm of w.

Proof. By Lemma 5.3 of Van der Vaart and Van Zanten (2008b), the quantity on the left-hand side in the statement of the lemma is bounded by

$$-\log \mathbb{P}(\|S^{\alpha,M} - w\|_{\infty,M} < \varepsilon).$$

Clearly, this equals

$$-\log \mathbb{P}\Big(\sup_{x\in[0,1]} |M^{-\alpha}S^{\alpha,M}_{Mx} - M^{-\alpha}w(Mx)| < M^{-\alpha}\varepsilon\Big).$$

By the self-similarity of the RL-process, the rescaled process $(M^{-\alpha}S^{\alpha,M}_{Mx})_{x\in[0,1]}$ has the same distribution as $(S^{\alpha,1}_x)_{x\in[0,1]}$. Moreover, the function $x \mapsto M^{-\alpha}w(Mx)$ belongs to $C^\alpha[0,1]$ and its C^α-norm is equal to that of w. Hence, by Lemma 5.3 of Van der Vaart and Van Zanten (2008b) again, it suffices to prove the result for $M = 1$. As mentioned above, this is given by Theorem 4.3 of Van der Vaart and Van Zanten (2008a). □

4.2 Proof of Theorem 2.1

We apply Theorem 3.2 to the sequence of priors Π_n constructed in Section 2.2. By Lemma 4.1, conditions (3.4) and (3.5) of Theorem 3.2 are fulfilled if

$$M_n\varepsilon_n^{-1/\alpha} \lesssim T_n\varepsilon_n^2. \tag{4.1}$$

By assumption, the drift function θ_0 grows at most linearly. Moreover, the invariant measure μ_0 has subexponential tails, cf. Section 2.1. It follows that

$$\int_{x:|x|>M_n} \theta_0^2(x)\,\mu_0(dx) \lesssim \int_{M_n}^\infty x^2e^{-Cx}\,dx$$

for some constant $C > 0$. The right-hand side is bounded by a constant times $\exp(-C'M_n)$ for another constant $C' > 0$. Hence, condition (3.6) of Theorem 3.2 is fulfilled if

$$e^{-cM_n} \lesssim \varepsilon_n^2 \tag{4.2}$$

for all $c > 0$. The proof is completed by observing that the sequences M_n and ε_n defined in the statement of the theorem satisfy (4.1) and (4.2). □

References

Banon, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control Optim.* **16**(3), 380–395.

Belitser, E. and Ghosal, S. (2003). Adaptive Bayesian inference on the mean of an infinite-dimensional normal distribution. *Ann. Statist.* **31**(2), 536–559.

Castillo, I. (2008). Lower bounds for posterior rates with Gaussian process priors. *Preprint* .

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

Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14**(1), 1–67.

Galtchouk, L. and Pergamenshchikov, S. (2001). Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes. *Math. Methods Statist.* **10**(3), 316–330.

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

Ghosal, S., Lember, J. and Van der Vaart, A. W. (2008). Nonparametric Bayesian model selection and averaging. *Electron. J. Stat.* **2**, 63–89.

Ghosal, S. and Van der Vaart, A. W. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35**(1), 192–223.

Huang, T.-M. (2004). Convergence rates for posterior distributions and adaptive estimation. *Ann. Statist.* **32**(4), 1556–1593.

Itô, K. and McKean, H. P., Jr. (1965). *Diffusion processes and their sample paths*. Springer-Verlag, Berlin.

Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. Springer-Verlag, New York.

Kleijn, B. J. K. and Van der Vaart, A. W. (2006). Misspecification in infinite-dimensional Bayesian statistics. *Ann. Statist.* **34**(2), 837–877.

Kutoyants, Y. A. (2004). *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag, London.

Liptser, R. S. and Shiryayev, A. N. (1977). *Statistics of random processes I*. Springer-Verlag, New York.

Spokoiny, V. G. (2000). Adaptive drift estimation for nonparametric diffusion model. *Ann. Statist.* **28**(3), 815–836.

Tuan, P. D. (1981). Nonparametric estimation of the drift coefficient in the diffusion equation. *Math. Operationsforsch. Statist. Ser. Statist.* **12**(1), 61–73.

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**(5), 863–888.

Van der Vaart, A. W. and Van Zanten, J. H. (2008a). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36**(3), 1435–1463.

Van der Vaart, A. W. and Van Zanten, J. H. (2008b). Reproducing kernel Hilbert spaces of Gaussian priors. *IMS Collections* **3**, 200–222.

Van der Vaart, A. W. and Van Zanten, J. H. (2009). Adapative Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth. *Ann. Statist., to appear* .

Van Zanten, J. H. (2001). Rates of convergence and asymptotic normality of kernel estimators for ergodic diffusion processes. *J. Nonparametr. Statist.* **13**(6), 833–850.

Van Zanten, J. H. (2003). On uniform laws of large numbers for ergodic diffusions and consistency of estimators. *Stat. Inference Stoch. Process.* **6**(2), 199–213.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40**(3), 364–372.